ELSEVIER

Mechanisms of Development 53 (1995) 185–195

Selection and characterization of sequences with high affinity for the engrailed protein of *Drosophila*

Maria Teresa Saenz-Robles[a,b,1], Florence Maschat[a,2], Tetsuya Tabata[a], Matthew P. Scott[b], Thomas B. Kornberg[a,*]

[a]*Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143, USA*
[b]*Departments of Developmental Biology and Genetics, Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA 94305, USA*

Received 1 June 1995; accepted 12 July 1995

Abstract

The *engrailed* gene helps to direct *Drosophila melanogaster* development by encoding a homeodomain-containing DNA binding protein. To identify genes whose transcription *engrailed* regulates, we developed a method to isolate genomic sequences to which engrailed protein binds with high affinity. Fragments of genomic DNA were fractionated on an engrailed protein affinity column, and fragments that were retained in the presence of 0.4–1.0 M KCl were isolated and cloned. The isolated fragments include regions of the *engrailed* and *cubitus interruptus* gene promoters, both of which are candidate targets of *engrailed*, and most fragments contain regions that engrailed protein protects from DNaseI digestion. Chromosomal deletions that remove some of the engrailed binding sites (located either at 64D, 96B or 99D) interact genetically with *engrailed*. Characterization of a transcript encoded in region 64D revealed its dependence on engrailed protein.

Keywords: Engrailed; Sequence specific binding; Developmental regulation

1. Introduction

The body pattern of the *Drosophila* embryo is generated by a process that creates a repetitive series of segments along the anterior/posterior axis. This process is controlled by regulatory genes that uniquely specify each segmental region by their activity (or inactivity) in precisely defined domains (reviewed by Kornberg and Tabata, 1993). An example is the *engrailed* gene, which is expressed in only the posterior part (compartment) of each segment. *Engrailed* function in posterior compartment cells is essential for proper development and for maintaining compartment and segment borders (Garcia-Bellido and Santamaria, 1972; Lawrence and Morata, 1976; Kornberg, 1981).

Engrailed encodes a nuclear protein (EN) with 552 residues (DiNardo et al., 1985; Poole et al., 1985). EN has a 60 residue homeodomain that binds DNA with high affinity and that is structurally similar to the Antennapedia homeodomain (Otting et al., 1988; Qian et al., 1989; Kissinger et al., 1990). EN is similar in size and sequence to INV, the protein encoded by the *invected* gene. *Invected* is a non-essential gene that is expressed in patterns that are almost identical with *engrailed*, and that encodes a protein that includes a homeodomain of almost identical sequence (Coleman et al., 1987). EN and INV are thought to provide partially redundant functions, and, except where noted below, are not differentiated in the discussion that follows. Although it is likely that EN and INV function as a transcription regulators (Jaynes and O'Farrell, 1991), their mechanisms of action are poorly understood and their regulatory targets have yet to be definitively identified.

Several approaches have been taken to identify regulatory targets of genes such as *engrailed*. For instance, genetic studies have identified *cubitus interruptus* (*ci*; Eaton

* Corresponding author, Tel./Fax: +1 415 4768821.
[1] Current address: Division of Immunogenetics, Department of Pediatrics, Rangos Research Center, Children's Hospital of Pittsburgh, Pittsburgh, PA 15213, USA.
[2] Current address: Institut Jacques Monod, CNRS et Université Paris 7, F-75251 Paris Cedex 05, France.

0925-4773/95/$09.50
SSDI 0925-4773(95)00433-2

and Kornberg, 1990), *patched* (*ptc*; Nakano et al., 1989; Hooper and Scott, 1989), *Deformed* (*Dfd*; Jack et al., 1988), *decapentaplegic* (*dpp*; Raftery et al., 1991) and *hedgehog* (*hh*; Tabata et al., 1992) as well *engrailed* itself (Heemskerk et al., 1991) as genes whose normal patterns of expression depend upon *engrailed* function. Expression of *ci*, *ptc*, *Dfd*, and *dpp* appears to be limited to anterior compartments by *engrailed* function in adjacent posterior compartments; in posterior compartment cells, *engrailed* function is needed for stable expression of *hh* and *engrailed* itself. Thus, through direct or indirect mechanisms, EN acts to both positively and negatively regulate these targets. Molecular approaches to identify the direct targets of EN and other homeodomain proteins in *Drosophila* have included immune precipitation of protein bound to chromatin (Gould et al., 1990; Tomotsune et al., 1993), immune precipitation and footprinting of homeodomain proteins bound to cloned gene fragments (for examples, see Desplan et al., 1985; Driever and Nusslein-Volhard, 1989; Qian et al., 1993), and in situ localization by immunohistochemistry (Serrano et al., 1995). Although these studies have identified genomic sequences to which these proteins putatively bind in vivo, the lack of specificity with which these homeodomain proteins recognize DNA sequences (Hoey and Levine, 1988) and the inability of these methods to distinguish between sites with higher or lower affinity, means that such DNA binding studies are unlikely to offer a definitive way to identify direct regulatory targets. Most importantly, the possibility that the proteins bind to artifactual, lower affinity sites during the isolation procedures cannot be ruled out.

In this paper we describe a new method to identify targets of EN regulation. This method involves affinity chromatography with an EN-containing matrix to select genomic DNA fragments that bind EN tightly. This method yields pools of DNA fragments that are enriched with sequences having high affinity for EN, and its ability to discriminate against sequences with moderate or low affinity binding sites distinguishes it from immunoprecipitation and footprinting methods. We describe a genetic and molecular analysis of these binding sequences, and describe the isolation and characterization of a gene (*Msr-110*) that is juxtaposed to one of them. We show that the normal expression pattern of *Msr-110* is dependent upon *engrailed* function.

2. Results

2.1. Isolation of DNA sequences with high affinity for EN

To prepare an EN affinity matrix, EN protein isolated from *Drosophila* cells was coupled indirectly to protein A-sepharose beads. Coupling was through a polyclonal antibody directed against the N-terminal, non-homeodomain-containing portion of EN. Total *Drosophila* genomic DNA was digested to an approximate size of 150–450 bp, and 1 mg was applied to the matrix. After extensive washes, the bound DNA was eluted with steps of increasing concentration of KCl (Fig. 1A). DNA that eluted with 0.3 M, 0.4 M, and 1.0 M KCl represented 3%, 0.6%, and 0.1% of the total DNA applied, respectively. Upon re-application, more than 93% of the DNA pooled from the 0.3 M, 0.4 M or 1.0 M eluates was retained on the resin, and each pool eluted with a profile

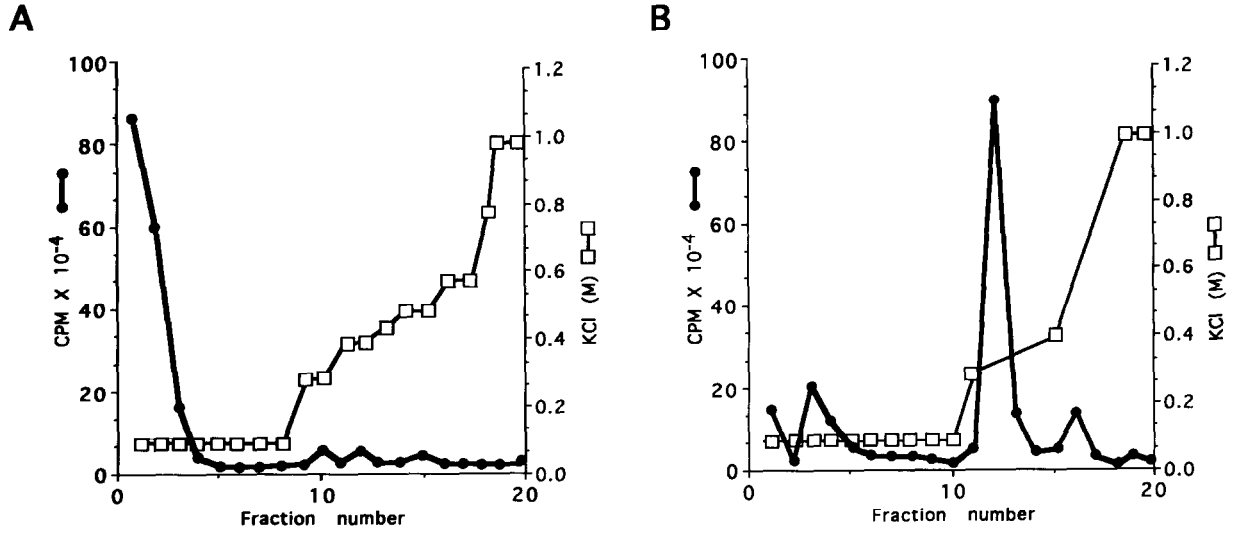


Fig. 1. Isolation of genomic DNA with high affinity for EN. Genomic DNA was digested with *Mbo*I digested and applied to an EN-affinity column after a portion was end-labeled with ^{32}P. DNA adhering to the column was eluted with successive steps of increasing concentrations of KCl. DNA was monitored by ^{32}P (closed circles) and ionic strength was monitored by conductivity (open squares). (A) initial application of genomic DNA (B) re-application of DNA pooled from fractions containing DNA in the 0.3, 0.4 and 1.0 M eluates of the initial column run and labeled with ^{32}P a second time.

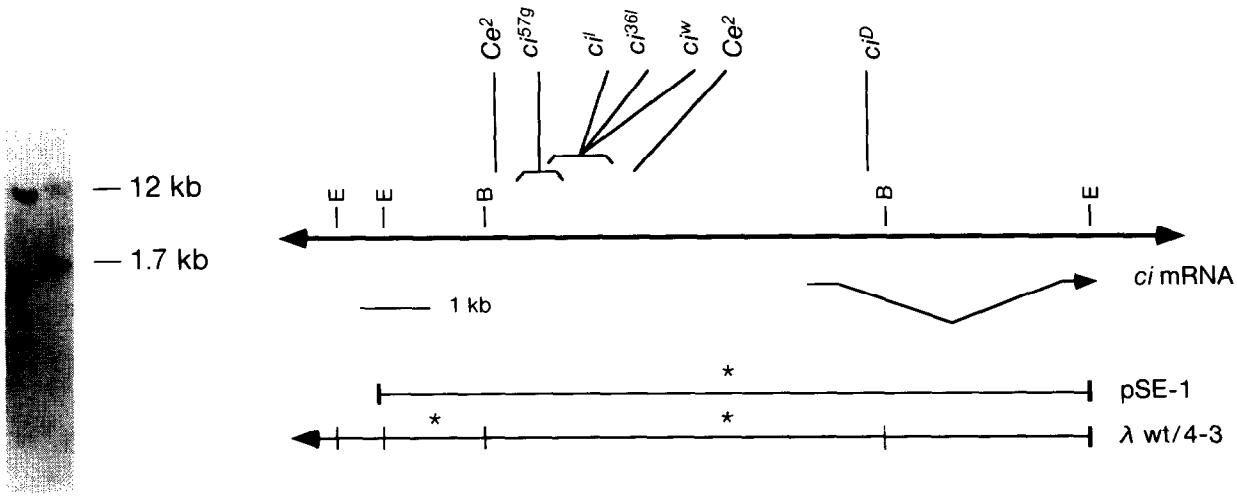


Fig. 2. *ci* DNA has high affinity EN-binding sequences. (Left) Plasmid (pSE-1, left lane) and phage (λwt/4–3, right lane) DNA containing portions of the *ci* gene were digested with *Eco*RI or *Eco*RI + *Bgl*II, respectively, and a Southern blot was hybridized with a probe consisting of the fragments recovered in the 1.0 M fraction of the EN-affinity column (100 ng) that had been labeled by random priming in the presence of four labeled [α-^{32}P]dNTPs (300 μCi each; 400 Ci/mmol). (Right) The *ci* genomic region (thick horizontal line) is shown along with *Eco*RI and *Bgl*II restriction sites and the location of 7 *ci* mutants (Orenic et al., 1987). The 5′ end of the *ci* mRNA, pSE-1 plasmid and λwt/4–3 are drawn below and the fragments of the plasmid and phage that hybridize with the probe are indicated (*).

similar to that observed with the initial application (Fig. 1B).

2.2. *Characterization of sequences with high affinity for EN*

DNA fragments that were recovered in the 0.4 M and 1.0 M eluates were cloned into a plasmid vector. To assess the identity of these cloned DNA fragments, an aliquot of the eluted fragments was labeled and used to probe a Southern blot containing DNA from putative gene targets of *engrailed*. The pool of fragments includes DNA from the regulatory region upstream of *ci* (Fig. 2) and *engrailed* (not shown).

Southern analysis using isolated inserts as probes indicated that some sequences were represented multiple times in the clone library (not shown). Twenty-two clones with different sequences were randomly selected for further characterization. Sequence analysis revealed that one fragment (clone 106) contains a portion of the *engrailed* regulatory region that includes the K1, K2, and K3 EN binding sites (Desplan et al., 1985; Hoey and Levine, 1988). DNase digestion studies with the 22 fragments

Table 1
Sequences in 15 fragments protected from DNase digestion by EN

106	--TTAAATGTCAATTAAATATCAATCAATTTCG--(30N)--TTAATTGAAGGCC--
110	--TTAATTGCATAA--(9IN)--TGATTAGTCAAT--
116	--TGGATGG--(21N)--TGAATGAGTCAATGAATCACACAAATGAATTGTTTGGTTG--
117	--CCCTGTTAGCATGCTATTGCCCACTTTGCTGGAT--(54N)-- TTCAGATGATAAGAGGTACTTCTATCAGACAGT--
407	--TTTGCATTTCATTACAAATGACATTTGAGGGGaaTAATAATCTATTTTA--(23N)-- TGATATCATTTGTTTC-
410	--TTCATTTGCCTAATTT--
417	--TTCCCGTTTCaACGTCTAATTAA--(21N)--GTTTATCAAG--
419	--ATaTTAATTAATTGC--
1.2	--CTTAATTGAAC--
11A5	--TCATTTGTTGATTAGTAAATCAAAGAAGCTCATTAATTTTCCTGAAAAGTGAC--(13N)-- AGAGAATAAGAGTCATTGCAGGTCNG--(20N)-- GGTNGGAATGAGATGAAGTNGGGCCCTAAATCTTTAGCCTTG
11B11	--GCTACTTAACTGCT--(9N)--ATGGATTGCTGATGA--
11B12	--AGGGGAA--(41N)--AATTGCGCGC--
11F6	--ATCGGGTTCTGCACCGTCAGTCGG--
11G4	--AAAATCGATTTCAAGTTCAGCGATATG--(6N)-- CATAATAAATTCGCGGGCCACAATAAATAATAGATCAAATAC--
11G5	--CTCACCT--(18N)--CATCCAATTAATCAAC--

Lower case letters and "N" refer to sequences not protected from DNase digestion.

revealed that 17 contain sequences that are protected in the presence of EN. Since both strands of each of the 22 fragments were not tested, the 77% of fragments with protected sites represents a minimum estimate. Fifteen of the fragments had regions of protection that varied in size from 10–15 bp to almost 30 bp (Table 1); other portions of these fragments and the two other fragments were characterized by strong enhancements and partial protection.

2.3. Genetic analysis of fragments with high affinity binding sites

To map the genomic origin of several random clones, six clones were selected and used to isolate larger genomic fragments from an appropriate library, and in situ hybridization was performed to polytene chromosomes. As expected, clone 106 hybridized to 48A, the cytological location of *engrailed*. The other clones did not localize to regions with known putative targets (clone 106-48A; clone 110-64D; clone 117-2B–D; clone 407.3-99D; clone 407.4-24D1–4; clone 417-96E; clone 419-96B). To ask if there is any functional relationship between *engrailed* and genes linked to the high affinity binding sequences, genetic interactions between *engrailed* and regions for which deficiencies were available (64D, 96B, and 99D) were examined. Flies carrying both an *engrailed* deficiency and a deficiency covering the identified region were generated, and the phenotype of these transheterozygotes was assessed. The *engrailed* deficiencies used either remove the entire coding region of *engrailed* and extend through the adjacent invected transcription unit or remove the regulatory region upstream of the *engrailed* transcription unit that is required for both *engrailed* and *invected* expression (Ali, Z. and T.B.K., unpublished; Gustavson, 1993). Therefore, these deficiencies are likely to affect both *invected* and *engrailed* expression. Flies hemizygous for engrailed and *invected* have a wild type appearance; animals lacking *engrailed* function or homozygous for the *engrailed/invected* deficiency die as embryos with segment deletions and fusions. In contrast to *engrailed*, *invected* mutants are viable and develop normally (Gustavson, 1993).

Regions 64D, 96B, and 99D are haplo-sufficient in an otherwise wild type background, but produced flies with abnormal wings when transheterozygous with deletions of *engrailed/invected*. In each genotype, the affected region of the fly was the posterior part of the wing, a phenotype that is consistent with the possible role of EN as a regulator in the posterior compartment. Each transheterozygote has a characteristic wing vein phenotype, as illustrated in Fig. 3. Where possible, several deficiencies in the proximity of the binding sites that did not delete the EN binding sequence were tested for interaction with the *engrailed/invected* deficiency; and in each case tested, the phenotype of these transheterozygotes was normal. For instance, other deficiencies in region 64 (e.g. *Df(3L)X37* (63F-64C_{1-5}) and *Df(3R)10H* (64B_{10-12}-64C_{5-7})) produced normal flies when transheterozygous with en^C and

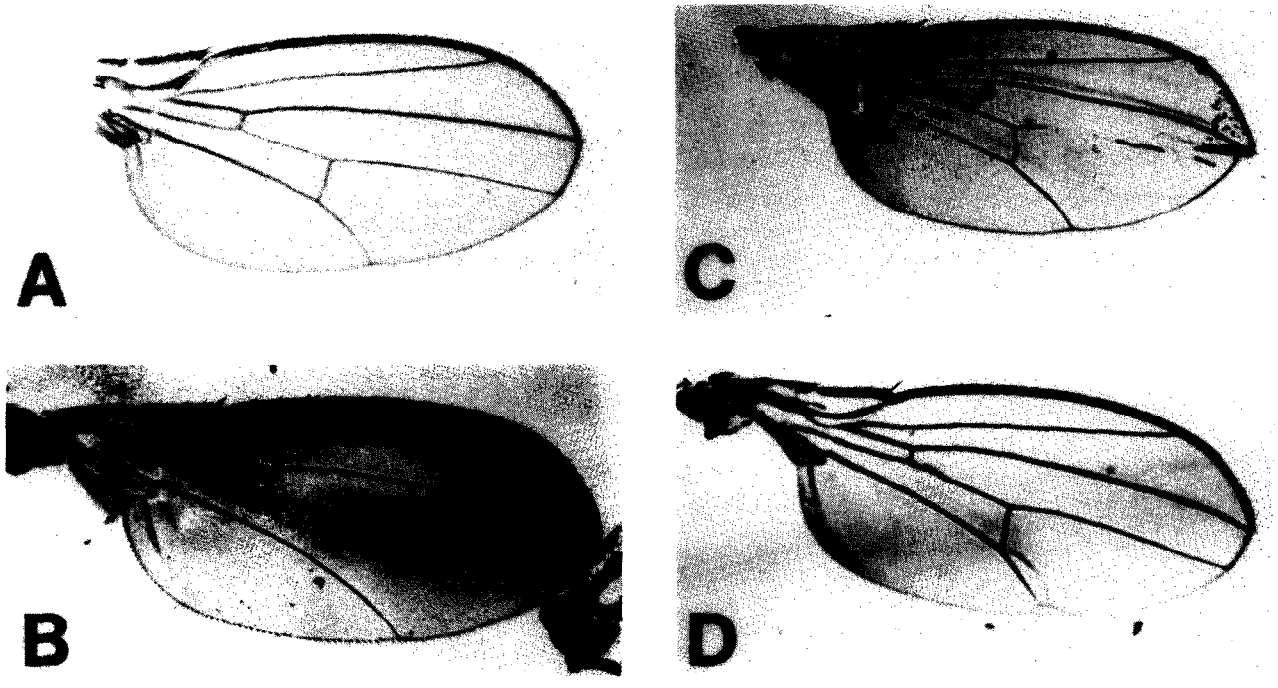


Fig. 3. Flies transheterozygous for deficiencies of engrailed and other selected regions have characteristic abnormalities. Wings were isolated from wild type flies (A) and from flies simultaneously hemizygous for engrailed/invected (*Df(2R)en^C*/+) and (B) region 64B_{7-9}-64D (*Df(3L) e^{13} st in ri p^p*/+); (C) region 96A_{2-9}-96D_{2-4} (*Df(3R)slo^8*/+); or (D) region 99D_{1-2}-99E (*Df(3R) X3F1*/+). Malformations seen in the transheterozygotes are characteristic of these genotypes.

en^D, whereas $Df(3R)e^{13}$ ($64B_{7-9}$-64D) produced flies with abnormal wings in approximately 50% of flies carrying either *engrailed* deficiency, or with the cytologically normal lethal *engrailed* allele, en^{LA4}. Approximately 80% and 50%, respectively, of the flies carrying en^D and en^C had wing abnormalities when hemizygous for the region $96A_{2-9}$-$96D_{2-4}$ ($Df(3R)slo^8$). Approximately 75% and 40%, respectively, of slo^8 flies that were heterozygous for en^{CX1} or en^{LA4} also had abnormal wings. In contrast, these *engrailed* heterozygotes developed normally when simultaneously hemizygous for regions $96A_{1-7}$-$96A_{21-25}$ ($Df(3R)XS$), $95E_{6-7}$-$96A_{18}$ ($Df(3R)slo^3$), or $95E_{6-7}$-$96A_{18}$ ($Df(3R)S87-5$). The single available deficiency of region 99D ($Df(3R)$ *X3F1* ($99D_{1-2}$-99E)) produced flies with abnormal wings with en^D (80%), en^C (58%), and en^1 (23%).

2.4. Characterization of Msr-110

Clone 110 was characterized further. It contains two short and well-defined regions that EN protected from DNase I digestion (Fig. 4A). When clone 110 was placed

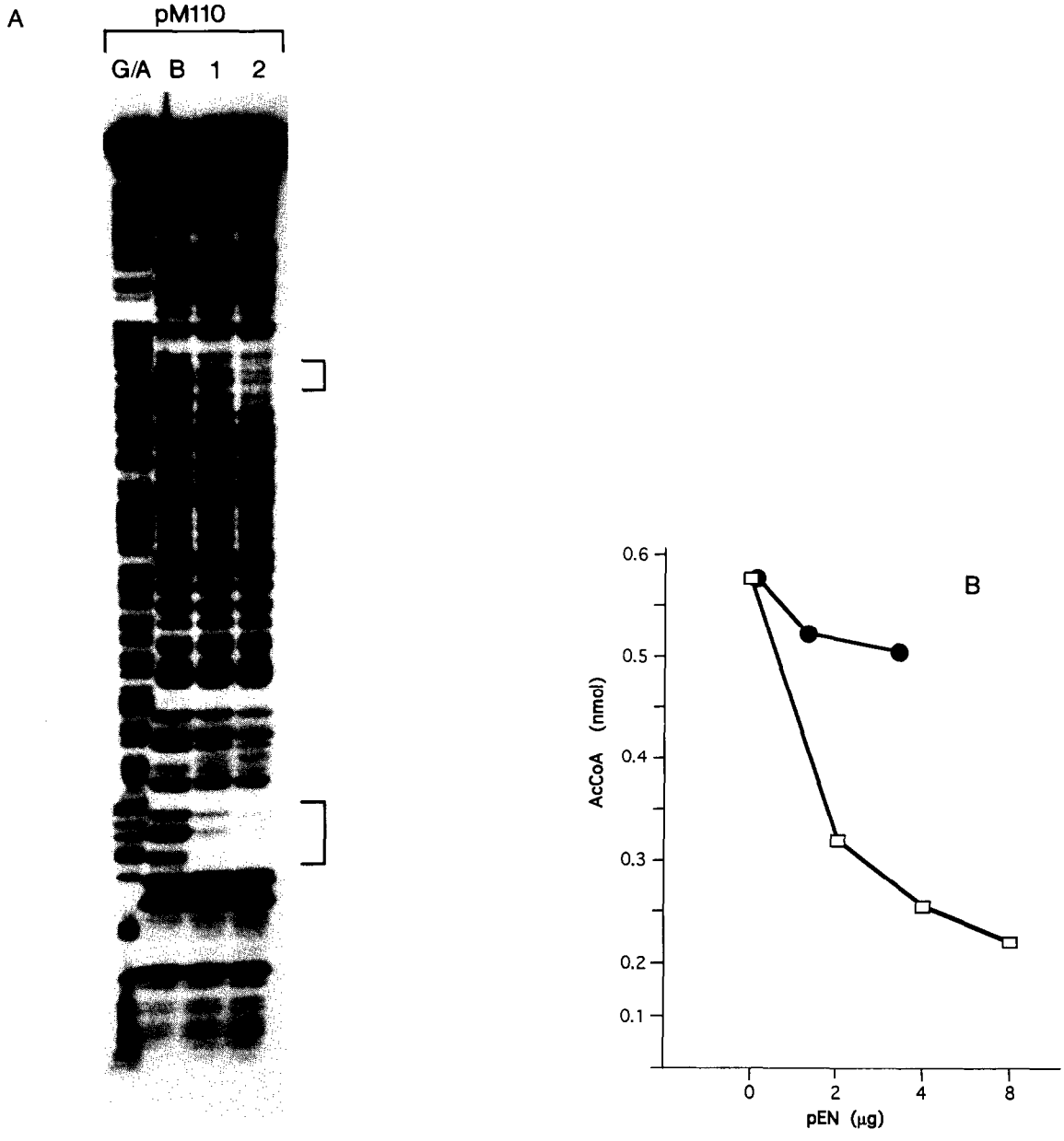


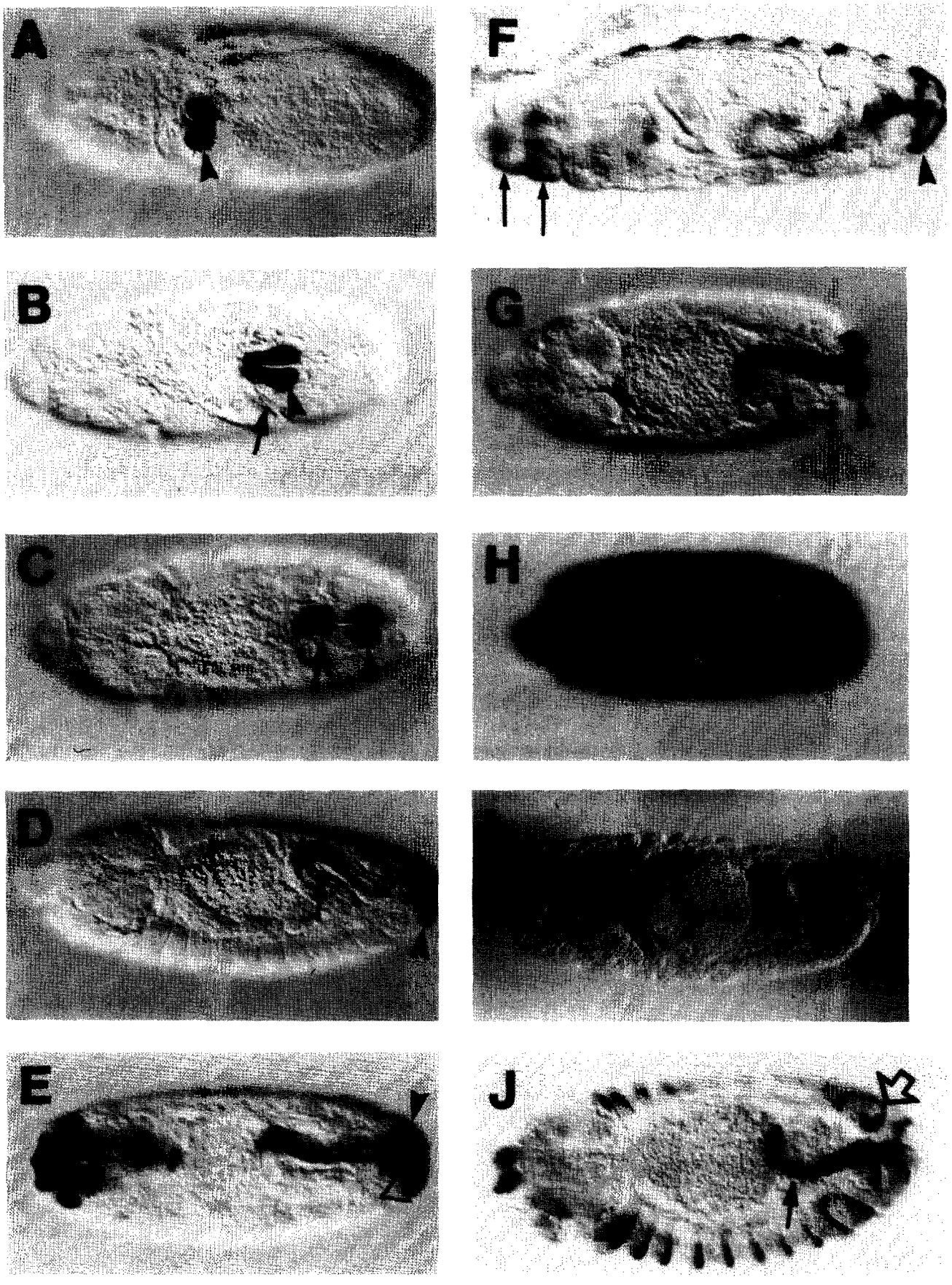
Fig. 4. Identification and functional characterization of the EN binding site in clone 110. (A) Four lanes from an autoradiogram showing a G/A ladder (left lane) and digestion by DNaseI in the absence of EN (–) or in the presence of increasing amounts of EN (+ and ++). (B) Co-transfection of a plasmid containing the clone 110 fragment placed in the regulatory region upstream of the CAT coding sequence with either pAc-EN (open squares), a plasmid containing the actin 5C promoter driving expression of the engrailed cDNA or pPAc-enM-STOP (closed circles), which is identical to pAc-EN except for the presence of a termination codon that stops translation of EN after amino acid 406 (Jaynes and O'Farrell, 1988). Data is plotted as nmol acetylCoA per unit of β-galactosidase. Cells transfected with a control plasmid vector yielded a value of 0.09 in this assay.

```
CCCCCCCCCCCGAGACGGCGAGTGGCAGTCGTACAACAAAGACGTAGTAATCCAAAGTTTCCAGAGGCCCAAAAGTCTCCCTGCTGAAAAGTTGGTCAGGC
AACTTTTTCAAAAACCAACAAGACCGCTGAAAAATAAATCGTACAACGAAATTCCTCTTACTTTTTGTCGAATTTTTGGAAAATCCTCTACTGCTATAGA

TGACCGAGAAGGATACCGTCTCGATGGCTACCGTGCAAATGAAGCCCGACTATGCGGCTAGCGAGGTCTACAGCACCGCCAGCGAACCGCCACCGGCCTA      300
M  T  E  K  D  T  V  S  M  A  T  V  Q  M  K  P  D  Y  A  A  S  E  V  Y  S  T  A  S  E  P  P  P  A  Y

CAAGCGCCAGGCGAATTCGGTGAAAATCGCCAAGATCACCGCCTTCACCATCATCGTGTCAGCCTTCATCCTGGGATCCTTCATCCTGGCCTCCTCCTAT      400
  K  R  Q  A  N  S  V  K  I  A  K  I  T  A  F  T  I  I  V  S  A  F  I  L  G  S  F  I  L  A  S  S  Y

CTGCAGGCCAAGGCCTCCTGCGATCAGGTCCAGGCCCTCGACTCGGTGCTGGAAAAGGAGCTAATGCTTGAGACTTTGCAGCAGGTGGGCAAGGAACTTC      500
  L  Q  A  K  A  S  C  D  Q  V  Q  A  L  D  S  V  L  E  K  E  L  M  L  E  T  L  Q  Q  V  G  K  E  L

CCCGTGCTGAACCCCTGCTGGGAGGAGCTGCTGGAGCTGCCGATGACTCCGAACTGCAGAGCTTGGAGCCGGAAAGTCGTAAGACTGAAGCCACCCAGGA      600
P  R  A  E  P  L  L  G  G  A  A  G  A  A  D  D  S  E  L  Q  S  L  E  P  E  S  R  K  T  E  A  T  Q  D

TGCCGAGGAGCACCCCGACAAGGACAACTCCTACTCCGACAGCGATGAGACCGATGAGCTGCAGAAGTTCCCCGGCAAGATGCCACTCGAGCTGGACTTG      700
  A  E  E  H  P  D  K  D  N  S  Y  S  D  S  D  E  T  D  E  L  Q  K  F  P  G  K  M  P  L  E  L  D  L

AGTGACCTGGCTGCCGCCATTCTGCGCAACAACAAGAAATCCCGCATGAACTGTGTGGTGGAGCGCAAGCACGCCGAGGAGATTGTCGACTCGCCCTCGA      800
  S  D  L  A  A  A  I  L  R  N  N  K  K  S  R  M  N  C  V  V  E  R  K  H  A  E  E  I  V  D  S  P  S

AGACGGTAGCCCTGCCCTTCGGAGTCAACCTGACCACGGATCCCAAGAAGGCACGCATCACCGGCGAGCGGATCTCGATCTTCTGCGACGGCGGCGACGA      900
K  T  V  A  L  P  F  G  V  N  L  T  T  D  P  K  K  A  R  I  T  G  E  R  I  S  I  F  C  D  G  G  D  D

CAAGGATAAGGAGGAGAAGGAGGCCAAGCGCCAGCGGGAGCAGGAGGAGGAGGACTCCGATGAGGATGCCGAGCCGCTGAGGCCCATGTTCATTCCCTTC     1000
  K  D  K  E  E  K  E  A  K  R  Q  R  E  Q  E  E  E  D  S  D  E  D  A  E  P  L  R  P  M  F  I  P  F

CAGCGAGTGCCCATTCCCTTCGGACGCATTCCCGACCAGATGCCGCTGACCCATATGCCCCAGAGGATGATGCCACCAGCCATGCAGATGCAACAGCAAT     1100
Q  R  V  P  I  P  F  G  R  I  P  D  Q  M  P  L  T  H  M  P  Q  R  M  M  P  P  A  M  Q  M  Q  Q  Q

TGCCCCAGCAGCAACAGCAGGGACCCATCATCATCCGCCAGCTGCCACCTCCGTTCCAGCACCTGCCCATGCGTCCTCCCATGCCGCCACAGGTGATGCA     1200
L  P  Q  Q  Q  Q  Q  G  P  I  I  I  R  Q  L  P  P  P  F  Q  H  L  P  M  R  P  P  M  P  P  Q  V  M  Q

GGCGCCTCGCATGGAGTCCTCCGAGGAAATGCAGATGCCCAAGGTCCAGACCGTGCGCATTCACATGCAGCAGATCCCCTTCCGTGAGATCCATGTGGCC     1300
  A  P  R  M  E  S  S  E  E  M  Q  M  P  K  V  Q  T  V  R  I  H  M  Q  Q  I  P  F  R  E  I  H  V  A

GACGATGTGCCCGTCCAGATCCCCGCCCAGCAGCAACGTCTGGAGATGCAGCAACGTCAGGAGATGCAACAGCGCCAGGAGATGCAGCAGCGCCATGAGA     1400
D  D  V  P  V  Q  I  P  A  Q  Q  Q  R  L  E  M  Q  Q  R  Q  E  M  Q  Q  R  Q  E  M  Q  Q  R  H  E

TGCAGCAGCGCCAGGAGATGCAGCAGCGTCATGAGATGCAGCAGCGCCAGGAGCAGCGTGAGATCCCCCAGATTCAGATCCAGCCCATGCCCTTCGGAAT     1500
M  Q  Q  R  Q  E  M  Q  Q  R  H  E  M  Q  Q  R  Q  E  Q  R  E  I  P  Q  I  Q  I  Q  P  M  P  F  G  M

GGCCCTGCAGCGCGTTGGCATCACCGCCGAGGATCTGAGGAACATCCAGCGCATGGCCGAGGATCGCATCACCGACGAGCTGCGTCGTCTGGCCGCCGAG     1600
  A  L  Q  R  V  G  I  T  A  E  D  L  R  N  I  Q  R  M  A  E  D  R  I  T  D  E  L  R  R  L  A  A  E

GAGGGTGCGGAATCCAGTGAGGGCAGCGACGAAGATGTCTCCAGCTCCGCCGAGCAGGCCACCACCGAGGCACAGACCACGGTGCCCCAGCAGCAGCAGC     1700
  E  G  A  E  S  S  E  G  S  D  E  D  V  S  S  S  A  E  Q  A  T  T  E  A  Q  T  T  V  P  Q  Q  Q  Q

AACAGACCGTCCAGCCTGAGCAACAGCATCCAGAGGAACAGCACCAGGAAGTCAGCCAGGAGCGACAGGAGCAGCAGCAGCAGGTGCCCCAGATGCG     1800
Q  Q  T  V  Q  P  E  Q  Q  H  P  E  E  Q  H  Q  E  V  S  Q  E  R  Q  E  Q  Q  Q  Q  Q  V  P  Q  M  R

CATTCAGCGTTTGATCCTGCAGCCCCTGCCTGAGCAGCCCCGTGACACCATGACCCCGCCGATGCCCGACCAGCCGCAGCAGCTGCCCTTCGGCCGCATG     1900
  I  Q  R  L  I  L  Q  P  L  P  E  Q  P  R  D  T  M  T  P  P  M  P  D  Q  P  Q  Q  L  P  F  G  R  M

GTCTACGGACGCAGTCTCGCTCAGCCCGTCCGCATCCCGGTGCCCATGATGCAGGCCATGGAGGGAGGAGCTGCTGCCCCCGAAGAGTCCCAGCGACCAC     2000
  V  Y  G  R  S  L  A  Q  P  V  R  I  P  V  P  M  M  Q  A  M  E  G  G  A  A  A  P  E  E  S  Q  R  P

ACTTCGTCCAGCCTCGTTCCGTTTAAGTAACAAGAAGACCACGCCTGTGATTGCTGCTGTGCGAATCGAATCTCGACGAGGGTGTAGACCGAGAGGAACT     2100
H  F  V  Q  P  R  S  V
AGAAGAAGGAGTCATCATGTACCCAACTAAACCCGAAGAATCTAAGATTGTAACCGAATCGTCTCGATTTGTTTTGTAATTAAAGCAGCTTAAAAGAAAA
AGTAAAAACCAAAACTAATCAACAAGTTTTAAAACATGCCTTTCTACCATATTTGGCCTTCTTCAAAAAACGTATTCTCTATATATCATTTTTGTATATT
TTATATTAATGCGTGTATAACATTCCAACCAACCAACCAACAACAACAACTAGAACACAGAAAGCAATCGCTTGTGAAAAAGAAAATTGTGAAAGTTTTC
CAAGTATTTTTCTTTCAGTTTTTTGAATTGTCTACAGTTTTGTTTTTCACAAATACACACAAATAAATCCGATGGAAATTTTTGTACATAAAAAAAAAAA
AAAAAAACGAATGCTGCGGCCGCCGAATTCCTGCAGCCCGGGGGATCCACTAGTTCTAGAGCGGCCGCCACCGCGGTGGAGCTCCAATTCGCGCTATAGT
GAGTCGTAGGGCAA
```

Fig. 5. Sequence of *Msr-110* cDNA and its conceptual translation. Several possible translation starts are located between nucleotides 300 and 339; the protein sequence shown utilizes the first. The double line underscores the putative transmembrane region. The single lines identify the 6 tandem repeats related to the sequence EMQQRQ.

upstream of the gene encoding chloramphenicol acetyl transferase (CAT), expression of CAT in a co-transfection assay was repressed in the presence of EN (Fig. 4B). Southern analysis indicated that the clone 110 sequence is present in one copy in the genome and Northern analysis of RNA extracted from embryos and imaginal discs located a 2.8 kb transcription unit within approximately 1 kb of the 110 sequence. Two cDNA copies of this

Fig. 6. Patterns of expression of *Msr-110* in developing embryos. In situ hybridization with a *Msr-110* cDNA probe reveals *Msr-110* transcripts in the presumptive anal plates prior to germ band shortening (arrowhead; A) and in a contiguous strip in the presumptive hindgut (arrow) during germ band shortening (B). Hindgut (arrow) and anal plate (arrowhead) expression separate as the process of germ band completes (C,D); hindgut expression is localized to the inner rim of the curved hindgut. During dorsal closure, expression appears in the head region and in the center of the primordia for the posterior spiracles (open arrowhead), and continues in the anal plate (E). Stage 16 embryo with expression in the anal plate confined to the periphery of its primordia (arrowhead), and expression in the epidermis greatest in the anterior part of each segment (thin arrows; F). Mutant embryos lacking engrailed/invected function ($Df(2R)en^{C}$) have heightened and uniform expression in the hindgut (G) and no apparent periodic expression in the epidermis (H). Ectopic expression of EN reduces *Msr-110* expression to negligible levels (I). (J) invected is expressed in a complementary pattern in the hindgut (where expression is limited to the outer half of the curve; arrow) and to a circumferential rim of the posterior spiracle primordia (open arrowhead).


A
B
C
D
E
F
G
H
J

2.8 kb transcript were isolated from an imaginal disc cDNA library (Brown et al., 1989). Both cDNAs hybridized to Northern blots of RNA extracted from embryos or imaginal discs. The longest cDNA was sequenced. It contains a single large potential open reading frame in its 2614 bp (Fig. 5), and two putative initiator methionines (at nucleotides 200 and 239 of the cDNA) match 2 of the 4 positions of the consensus for *Drosophila* translational starts (Cavener, 1987). Conceptual translations with these putative starts produce proteins with an M_r of approximately 69 k and a P_I of approximately 4.7. Near the N-terminus of the putative proteins is a hydrophobic sequence with 27 residues that could constitute a membrane spanning domain, and near the middle of the protein is a hexapeptide sequence (EMQQRQ) that repeats tandemly six times without interruption. With the exception of the third and fifth repeats, which have the sequence EMQQRH, the repeats have identical sequences. However, since the side chains of histidine and glutamine can be isostructural with respect to the side chain nitrogen, the conformation of these hexapeptide sequences is likely to be repeated without variation. Other than several stretches of polyglutamine, no significant homology to other known protein sequences was noted in BLAST searches of Genbank (Altschul et al., 1990).

Use of the *Msr-110* cDNA as a probe of poly(A)$^+$ RNA that had been extracted from embryos and larval imaginal discs revealed that the 2.8 kb *Msr-110* transcript was present during the later stages of embryonic development (12–18 h) as well as in larval imaginal discs (not shown). Use of this probe for in situ hybridization studies detected *Msr-110* RNA in embryos and discs of the same stages. Hybridization was detected first uniformly in the region of early stage 11 embryos that will give rise to the anal plates (Fig. 6A). In older embryos, the transcript was also detected in the primordia of the hindgut (stage 11; Fig. 6B–D), where the hybridization signal was greater in one longitudinal half, the antennal-maxillary complex (stage 12; Fig. 6E), two small circular patches that will contribute to the posterior spiracles (stage 13; Fig. 6E), in the median tooth, the pharynx and the proventriculus (stages 13–16; Fig. 6E), and, at a lower level, throughout the presumptive epidermis (stage 14; Fig. 6F). Expression in the cells of the presumptive epidermis was not uniform, but was greater in the anterior part of each segment.

To assess the role of EN in the elaboration of the *Msr-110* patterns of expression, the distribution of *Msr-110* RNA was also monitored in embryos that carry an *engrailed* transgene which is controlled by a heat shock promoter (HS-EN3; Poole and Kornberg, 1988). These studies showed that *Msr-110* RNA was reduced or was completely absent in HS-EN3 embryos that had been subjected to heat shock between 4h-9.5h AEL (stages 11–15; Fig. 6I), but that the RNA was unaffected by similar treatments of wild type embryos. Heat shocking younger HS-EN3 embryos was without effect. *Msr-110* RNA expression was also altered in mutant embryos that carry a deletion that removes the *engrailed/invected* gene complex (*Df(2R)en^C*). Hindgut expression was greatly enhanced in stage 11 mutant embryos (Fig. 6G), and in older embryos, expression in the epidermis did not have its characteristic segmental periodicity (Fig. 6H). These phenotypes suggest that in these tissues of the embryo, *engrailed/invected* may repress Msr-110 expression. Although *engrailed* is not expressed in the hindgut of the embryo, *invected* is expressed in the primordia of the hindgut and posterior spiracles in a pattern that is complementary to the pattern of Msr-110 expression (Fig. 6J).

3. Discussion

We have described the use of a new method that fractionates *Drosophila* genomic DNA based on its affinity for the homeodomain-containing transcription factor, EN. With this method, genomic DNA was bound to a column containing an EN-affinity matrix, and washes with increasing ionic strength produced pools of sequences with increasing affinity for EN. Approximately 0.1% of the DNA that had been applied to the column was recovered in the 1.0 M KCl eluate. Assuming that the DNA that was applied to the column represents mostly euchromatic, non-satellite sequences (Peacock et al., 1978), this fraction of the haploid genome has approximately 1.1×10^8 bp/1000 = 1.1×10^5 bp. DNA in this fraction was present as fragments with an average length of 300 bp, so approximately 367 different sequences were in the high affinity (1 M) pool. This level of complexity compares reasonably well with the approximately 50–65 polytene chromosome sites to which EN binds in vivo (Serrano et al., 1995).

Identifying regulatory targets of transcription factors can be straightforward if the protein can efficiently select its binding sites in vitro. For homeodomain proteins of the Antennapedia class, however, it has proven difficult to develop systems in which these proteins can discriminate among sequences that have a core sequence which includes TAAT (for instance, see Hoey and Levine, 1988; Ekker et al., 1991), although a promising system that selects and identifies sequences that respond to Ubx function in yeast has recently been described (Mastick et al., 1995). The method we used has the advantage that the EN on the affinity matrix had been produced in *Drosophila* cells where it was modified appropriately (Gay et al., 1988; Bourbon et al., 1995) and the protein was isolated without denaturation. Since, in addition, EN was coupled to the affinity matrix with an antibody preparation directed against the N-terminal, non-homeodomain-containing portion of the protein, the presentation of EN in the solid state was optimal for DNA binding. The high frequency of isolated fragments that contain binding sites that could be footprinted (77%), the presence in the high affinity fraction of fragments from the *engrailed* and *ci*

genes, and the genetic interactions between *engrailed* and the genomic regions from which several of the fragments derive together provide support for the effectiveness of the selection protocol. Most telling, perhaps, are the properties of the *Msr-110* gene.

Msr-110 derives from region 64D, deficiencies of which have a dominant phenotype in the posterior compartment of the wing blade when present in flies that also carry a deficiency for *engrailed*. This behavior indicates that the simultaneous reduction of function from both regions alters morphogenesis of the wing posterior compartment. The pattern of expression of Msr-110 is also suggestive of a functional connection between *engrailed* and *Msr-110*.

Msr-110 is expressed in embryos, first in the region of the presumptive anal pads of stage 11 embryos, and later in the primordia for the hindgut, posterior spiracles, proventriculus, various structures in the head and nasopharyngeal apparatus, and the epidermis. Several observations suggest that *engrailed* might be involved in limiting expression of *Msr-110* to these regions by restricting it from adjacent posterior compartments. First, expression of Msr-110 is completely suppressed upon ubiquitous induction of *engrailed* expression. Moreover, repression of *Msr-110* was observed only if EN was induced immediately before *Msr-110* would normally appear, and not if EN was ectopically expressed earlier. Consistent with these observations and the implication that EN directly regulates *Msr-110*, transcription in cultured cells of a reporter CAT gene fused to the *Msr-110* genomic fragment was sensitive in a similar manner to the presence of EN. Second, *Msr-110* expression in the anal pads, posterior spiracles, hindgut, and epidermis correlates inversely with patterns of *engrailed/invected* expression and with *engrailed/invected* function. Based upon patterns of *engrailed/invected* expression and the phenotypes of various homeotic mutants, the anal pads and posterior spiracles are thought to derive, respectively, from anterior abdominal segment 10 (aA10; Kuhn et al., 1992) and aA8 + pA8 (the inner lining and most distal aspects of the posterior spiracle are produced by a doughnut-like ring of cells that express engrailed, while the base and outer lining are produced by the encircled cells that do not express engrailed (Kuhn et al., 1992)). *Msr-110* expression is robust in the anal pads and in a circular patch that appears to lie inside of the ring of *engrailed*-expressing cells in the region of the presumptive posterior spiracle. The hindgut has an uncertain segmental identity, but it is presumably comprised of both anterior and posterior compartment cells (*invected* and *engrailed* are normally expressed in only one longitudinal half (Hama et al., 1990)). *Msr-110* is expressed more abundantly in the other longitudinal half of the hindgut, and in *engrailed/invected* mutant embryos, it is expressed uniformly and at significantly higher levels. Lastly, the epidermis is comprised of an alternating series of anterior and posterior compartments (Kornberg et al., 1985). *Msr-110* expression in the epidermis of normal embryos is also periodic, with more abundant expression in the anterior parts of each segment; in *engrailed/invected* mutant embryos, expression in the epidermis is uniform.

EN has been previously been implicated as positive or negative regulator of a number of different genes, and these observations strongly suggest that it negatively regulates at least some aspects of the pattern of *Msr-110* expression in the embryo. Although the presence of two strong binding sites in the region upstream of the *Msr-110* transcription unit suggests that its regulation of *Msr-110* is direct, a definitive conclusion awaits analysis of mutants with altered binding sites. We note that when EN was transfected into Schneider S2 cells, it activated a reporter construct that was linked to one of the other sequences identified in this study (clone 117; M. Saenz-Robles and T.B. Kornberg, unpublished), and that the genetic interaction between the region to which *Msr-110* maps (64D) and *engrailed* suggests that EN might have a positive role in *Msr-110* regulation in wing imaginal discs. These observations therefore underscore the complexity of the mechanisms through which EN functions and the rudimentary state of our understanding.

4. Experimental procedures

4.1. Fly culture and fly strains

Wild type refers to an Oregon R line. Deficiencies were obtained from the Indiana Drosophila Stock Center (*Df(3R) X3F1/TM3 Sb e* (deletion $99D_{1-2}$-99E) and *Df(3R)XS, Dp(3R)XS asp ats* p^p*/TM6B;y/y*$^{(+)}$*Y* (deletion $96A_{1-7}$-$96A_{21-25}$)), M. Simon (*Df(3L)* e^{13} *st in ri* p^p*/TM6* (deletion $64B_{7-9}$-64D), *Df(3L)X37 st in ri* p^p*/TM1* (deletion 63F-$64C_{1-5}$), and *Df(3L)10H st e/TM6* (deletion 60C ↔ $64B_{10-12}$;81F ↔ $64C_{5-7}$;81A → 100), and B. Ganetzky (*Df(3R)slo*^{8}*/Dp(3;3)Su8* (deletion $96A_{2-9}$-$96D_{2-4}$) and *Df(3R)slo*^{3}*/TM6*; *S87–5/TM3* (deletion $95E_{6-7}$-$96A_{18}$). *Df(2R)en*^{C} and *Df(2R)en*^{D} were isolated after X-irradiation of the transgenic strain *ryXho25* (Hama et al., 1990); they have small deletions that remove, respectively, most of the *engrailed* and *invected* genes and the regulatory region 5′ of *engrailed* transcription unit (Ali, Z. and T.B.K., unpublished). en^1 is a viable mutant with an insertion in the regulatory region (Kuner et al., 1985); *en*^{LA4} is a lethal allele with a mutation that causes premature termination of EN translation and *en*^{CX1} is an inversion with a breakpoint within the engrailed transcription unit (E. Gustavson and T.B.K, unpublished). The HS-EN3 line is a transformed strain with an *engrailed* cDNA under the control of the hsp70 promoter (Poole and Kornberg, 1988).

To ectopically induce EN, *Drosophila* embryos (HS-EN (Poole and Kornberg, 1988)) 0–6 h old were aged for 1 h at 24°C before placing the vial in which they had been collected in a 37°C water bath. Heat shock was for

30 min, followed by a 2 h recovery at 24°C. The heat shock regimen was repeated three times. At 9.5 h following the third heat shock, the embryos were dechorionated, fixed, and processed for in situ hybridization as previously described (Eaton and Kornberg, 1990).

4.2. Preparation of EN-enriched nuclear extracts

Schneider S2 cells (5 l) transfected with an *engrailed* cDNA controlled by a hsp-70 promoter (Gay et al., 1988) were raised to 37°C for 30 min to induce EN. After 2 h of recovery at 24°C, the nuclei were isolated as described previously (Soeller et al., 1988). EN was extracted from the nuclei in 0.2 M $(NH_4)_2SO_4$, and was applied to a column of heparin sepharose previously equilibrated with Buffer A (Hepes 25 mM pH 7.6, KCl 0.2 M, 10% glycerol, 0.1% NP40, 10 mM EDTA, PMSF 100 μg/ml and 1 μg/ml leupeptin). EN was eluted from the resin with 0.35 M KCl.

4.3 Chromatography of genomic fragments

A protein A-sepharose column (Beckman, 1 ml total resin) was allowed to bind a goat anti-EN polyclonal antibody preparation which had been generated against the amino-terminal portion of the protein that does not include the homeodomain. The antibody was cleared of lipids with the Beckman lipid clearing solution and bound to the resin following the manufacturer's instructions. About 450 μg of immunoglobulin were retained by the protein A-column and the resin was then equilibrated with Buffer A (10 ml). Two milliliters (2.4 mg) of EN (see above) was applied, followed by extensive washes with Buffer A; approximately 65 μg of EN bound to the resin.

Genomic DNA was digested with *Mbo*I, and 40 μg was terminally labeled with [γ-^{32}P]ATP. One milligram of a mixture of labeled and unlabeled DNA was applied to the EN-resin in 1.3 ml of Buffer A. After extensive washes with buffer A, the retained DNA was eluted batchwise with 0.3, 0.4 and 1.0 M KCl (1.5 ml each). Pooled fractions were concentrated by precipitation with 0.6 vol. isopropanol. The total DNA recovered with the salt elutions was 56 μg. It was pooled, a portion was labeled as before and re-applied to the EN column. Ninety-three percent bound to the resin, and its profile of elution with 0.3, 0.4 and 1.0 M KCl closely resembled the behavior observed after the initial application. The second elution yielded 30 μg (0.3 M KCl), 6.3 μg (0.4 M KCl) and 1.06 μg (1.0 M KCl). DNA (7.5 μg) from the second 0.3 M KCl step elution was re-loaded again in the EN-specific column. All DNA bound to the resin, and 76% was recovered in a 0.3 M KCl wash. Higher concentrations of KCl of up to 1.5 M failed to recover additional DNA.

4.4. Northern analysis

RNA was extracted with RNAzol™ following manufacturer's (Cinna-Biotech) instructions. PolyA$^+$ RNA was purified in oligo-dT resin and 20 μg of each sample was electrophoresed in a 1.2% HGT agarose gel (TAE buffer, 6.7% formaldehyde). After transfer to HYBOND paper (Amersham) and UV crosslinking, probe generated with hexamer random primers and [α-^{32}P]dCTP using *Eco*RI fragments from phage (phages 110.1 and 110.3) or cDNA clones (clone 110.3.4) as template were applied.

4.5. DNase footprinting

Cloned fragments were isolated after digestion with *Eco*RI and *Xba*I, labeled by treating successively with phosphatase and polynucleotide kinase, and footprinted with DNaseI. Increasing amounts of EN-protein were mixed with 2–5 ng DNA in 50 μl of 50 mM Hepes (pH 7.6), 0.1 mM EDTA, 0.1 M KCl, 10 mM $MgCl_2$, 10% glycerol, 4% polyvinyl alcohol (PVA) and 4 ng/μl poly-dI-dC. After 15 min on ice, the samples were transferred to room temperature. Fifty microliters of digestion buffer (10 mM $MgCl_2$, 5 mM $CaCl_2$) were added, followed with 1 μl of DNaseI (50 μg/ml in 10 mM HEPES, pH 7.9, 5 mM $MgCl_2$ and 1 mM $CaCl_2$). The reaction was stopped after 1 min with 100 μl of stop mix (20 mM EDTA, 0.2 M NaCl, 1% SDS, 250 μg/ml yeast tRNA) extracted twice with 200 μl phenol/chloroform (1:1) and once with chloroform, and precipitated with 2 vols. of ethanol. DNA was resolved on a 6% sequencing gel, containing 6 M urea.

4.6. Isolation of genomic and cDNA clones

Approximately 50 ng of cloned fragment, isolated as above, was labeled in a PCR reaction containing PCR buffer (Perkin Elmer), 100 ng of primers, 0.2 mM dATP, GTP, and dTTP, 0.5 mM DTT and 2.5 units of Taq polymerase, and 50 μCi of [α-^{32}P]dCTP (3000 Ci/mmol) in 50 μl using a regimen of 3 cycles. The product was used to probe a λ-DASH *Drosophila* genomic library that had been transferred to nitrocellulose filters. Genomic fragments that included transcribed regions were labeled in a reaction containing 2 μl of hexamer random primers (Pharmacia) and [α-^{32}P]dCTP (3000 Ci/mmol), and were used to probe approximately 10^5 phage corresponding to an imaginal disc cDNA library (Brown et al., 1989).

4.7. Transfection

Transient transfection assays were performed Schneider S2 cells using the $CaCl_2$ procedure as previously described (Soeller et al., 1988). Cultures were transfected with a mixture of DNAs, that included carrier (1 μg pUC18) and 1 μg of a plasmid encoding β-galactosidase under the control of the copia promoter in a total of 10–15 μg DNA per experiment. DNA was precipitated as a fine white slurry and applied to cells for 48–72 h at 24°C. Cells were collected by centrifugation, washed twice and disrupted by two rounds of freeze-thawing. Debris was collected by centrifugation, and the supernatants were assayed for CAT and β-galactosidase activity as previously described (Soeller et al., 1988).

Acknowledgments

We thank Z. Ali for providing of *engrailed* mutant strains and D. Rosen and H. Bourbon for their contributions of purified EN that was used for the footprinting analysis, and to M. Simon and B. Ganetsky for deficiency strains. This work was supported by a postdoctoral fellowship to M.-T. S.-R. from the Fundacion Juan March and by a grant from the National Institutes of Health to M.P.S. and T.B.K.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) J. Mol. Biol. 215, 403–410.

Bourbon, H.-M., Martin-Blanco, E. and Kornberg, T.B. (1995) J. Biol. Chem., in press.

Brown, N.H., King, D.L., Wilcox, M. and Kafatos, F.C. (1989) Cell 59, 185–195.

Cavener, D.R. (1987) Nucleic Acids Res. 15, 1353–61.

Coleman, K.G., Poole, S.J., Weir, M.P., Soeller, W.C. and Kornberg, T.B. (1987) Genes Dev. 1, 19–28.

Desplan, C., Theis, J. and O'Farrell, P.H. (1985) Nature 318, 630–635.

DiNardo, S., Kuner, J., Theis, J. and O'Farrell, P. (1985) Cell 43, 59–69.

Driever, W. and Nusslein-Volhard, C. (1989) Nature 337, 138–143.

Eaton, S. and Kornberg, T.B. (1990) Genes Dev. 4, 1068–1077.

Ekker, S.C., Young, K.E., von, K.D. and Beachy, P.A. (1991) EMBO J. 10, 1179–1186.

Garcia-Bellido, A. and Santamaria, P. (1972) Genetics 72, 87–104.

Gay, N.J., Poole, S.J. and Kornberg, T.B. (1988) Nucleic Acids Res. 16, 6637–6647.

Gould, A.P., Brookman, J.J., Strutt, D.I. and White, R.A. (1990) Nature 348, 308–312.

Gustavson, E. (1993) Ph.D. Thesis, University of California, San Francisco, CA.

Hama, C., Ali, Z. and Kornberg, T.B. (1990) Genes Dev. 4, 1079–1093.

Heemskerk, J., DiNardo, S., Kostriken, R. and O'Farrell, P.H. (1991) Nature 352, 404–410.

Hoey, T. and Levine, M. (1988) Nature 332, 858–861.

Hooper, J.E. and Scott, M.P. (1989) Cell 59, 751–765.

Jack, T., Regulski, M. and McGinnis, W. (1988) Genes Dev. 2, 635–631.

Jaynes, J.B. and O'Farrell, P.H. (1988) Nature 336, 744–749.

Jaynes, J.B. and O'Farrell, P.H. (1991) EMBO J. 10, 1427–1433.

Kissinger, C.R., Liu, B., Martin-Blanco, E., Kornberg, T.B. and Pabo, C.O. (1990) Cell 63, 579–590.

Kornberg, T. (1981) Proc. Natl. Acad. Sci. USA, 78, 1095–1099.

Kornberg, T., Siden, I., O'Farrell, P. and Simon, M. (1985) Cell 40, 45–53.

Kornberg, T.B. and Tabata, T. (1993) Curr. Opin. Genet. Dev. 3, 585–593.

Kuhn, D.T., Sawyer, M., Packert, G., Turenchalk, G., Mack, J.A., Sprey, T.E., Gustavson, E. and Kornberg, T.B. (1992) Development 116, 11–20.

Kuner, J., Nakanishi, M., Ali, Z., Drees, B., Gustavson, E., Theis, J., Kauvar, L., Kornberg, T. and O'Farrell, P. (1985) Cell 42, 309–316.

Lawrence, P. and Morata, G. (1976) Dev. Biol. 50, 321–337.

Mastick, G.S., McKay, R., Oligino, T., Donovan, K. and López, A.J. (1995) Genetics 139, 349–363.

Nakano, Y., Guerrero, I., Hidalgo, A., Taylor, A., Whittle, J.R.S. and Ingham, P.W. (1989) Nature 341, 508–513.

Orenic, T., Chidsey, J. and Holmgren, R. (1987) Dev. Biol. 124, 50–56.

Otting, G., Qian, Y., Müller, M., Affolter, M., Gehring, W. and Wüthrich, K. (1988) EMBO J. 7, 4305–4309.

Peacock, W.J., Lohe, A.R., Gerlach, W.L., Dunsmuir, P., Dennis, E.S. and Appels, R. (1978) Cold Spring Harbor Symp. Quant. Biol. 42 (Pt. 2), 1121–1135.

Poole, S.J., Kauvar, L., Drees, B. and Kornberg, T. (1985) Cell 40, 37–43.

Poole, S.J. and Kornberg, T.B. (1988) In V. French, P. Ingham, J. Cooke and J. Smith (eds.), Mechanisms of Segmentation, The Company of Biologists, Cambridge. Vol. 104 Supplement, pp. 85–93.

Qian, S., Capovilla, M. and Pirrotta, V. (1993) EMBO J. 12, 3865–3877.

Qian, Y.Q., Billeter, M., Otting, G., Müller, M., Gehring, W.J. and Wütrich, K. (1989) Cell 59, 573–580.

Raftery, L.A., Sanicola, M., Blackman, R.K. and Gelbart, W.M. (1991) Development 113, 27–33.

Serrano, N., Brock, H.W., Demeret, C., Dura, J.-M., Randsholt, N.B., Kornberg, T.B. and Maschat, F. (1995) Development, in press.

Soeller, W.C., Poole, S.J. and Kornberg, T.B. (1988) Genes Dev. 2, 68–81.

Tabata, T., Eaton, S.E. and Kornberg, T.B. (1992) Genes Dev. 6, 2635–2645.

Tomotsune, D., Shoji, H., Wakamatsu, Y., Kondoh, H. and Takahashi, N. (1993) Nature 365, 69–72.