In vitro transcription of the *Drosophila engrailed* gene

Walter C. Soeller, Stephen J. Poole, and Thomas Kornberg

Department of Biochemistry, University of California, San Francisco, California 94143 USA

An enzyme system that accurately initiates transcription of the *engrailed* gene has been prepared from *Drosophila* embryos. The system has been separated chromatographically into two fractions, both of which are required for specific *engrailed* transcription. DNase footprint and competition analysis detected at least two sequence-specific DNA-binding proteins in one of these two fractions. Together, these proteins bind to eight regions within 400 bp of the transcription initiation sites. Most of the regions containing these binding sites are required for manimal *engrailed* transcription in vitro. In addition, a region downstream from the initiation sites and within the first 40 residues of the transcription unit is essential for transcription. Transient in vivo expression assays indicated that these same upstream and downstream sequences are required for transcription in *Drosophila* tissue culture cells.

The structure of the *Drosophila engrailed* gene is highly unusual. Mutations in the gene are spread over a 70 kb region of the genome, yet its single transcription unit is less than 4 kb in length (Drees et al. 1987). Flanking its transcription unit are transcriptionally silent regions that extend for 20 kb downstream and 45 kb upstream, suggesting complex and perhaps novel regulatory mechanisms. To understand the role of these flanking regions, and learn how expression of the *engrailed* gene is regulated, we have developed a system that faithfully initiates *engrailed* transcripts in vitro.

The central importance of transcriptional regulation in generating the temporal and spatial patterns of *engrailed* gene expression has been suggested by several observations. Mutations that alter the structure of the gene's untranscribed flanking regions affect both the quantity and disposition of *engrailed* transcripts (Weir and Kornberg 1985; and unpubl.). Also, the distinctive expression in the early embryo, in which single-cell-wide stripes of cells express the *engrailed* gene in each of the segmental primordia, is the result of an extremely rapid stepwise progression in a temporal pathway (Weir and Kornberg 1985). This progression initiates with 3 broad stripes, proceeds to a pattern with multiple stripes that alternate in intensity, and eventually results in 14 equally spaced zebra-like bands in blastoderm stage embryos (T. Karr, M. Weir, and T. Kornberg, in prep.). At this developmental stage, *engrailed* RNA has a half-life of approximately 15 min (M. Weir and T. Kornberg, unpubl.), and the progression of patterns is completed within 90 min. Such rapid and dramatic changes are likely to be under transcriptional control.

Regulation of *engrailed* gene expression has been investigated genetically. The *engrailed* gene is not expressed in *even-skipped* mutant embryos (Harding et al. 1986), and *engrailed* gene expression is also affected in *hairy* and *fushi tarazu* mutant embryos (Howard and Ingham 1986). Thus, the *even-skipped, hairy,* and *fushi tarazu* genes are required for the normal expression patterns of the *engrailed* gene. Although these studies do not indicate how this requirement is manifested at a molecular level, it is possible that the protein products of these genes directly regulate *engrailed* gene transcription.

An alternate approach to probing the mechanisms that control expression of the *engrailed* gene is to reconstruct the transcriptional regulation in vitro. In vitro systems from many cell and animal sources have been described that accurately reconstruct RNA polymerase II initiation (Dignam et al. 1983; Manley 1983). Several systems derived from mammalian sources have been shown to exhibit regulation in response to tissue-specific factors (Mizushima-Sugano and Roeder 1986; Gorski et al. 1986). In *Drosophila* systems, extracts of *Drosophila* tissue culture cell nuclei have been shown to initiate accurately from the promoters of *Drosophila* heat shock, alcohol dehydrogenase, actin, and histone genes (Parker and Topol 1984a,b; Heberlein et al. 1985). Such tissue-culture-derived systems unfortunately yield only small amounts of enzymes for biochemical characterization, and they cannot take advantage of either the rich genetic heritage of *Drosophila* or of its readily accessible developmentally staged populations. In this study, we have developed a method to prepare extracts from *Drosophila*

embryos, and have shown that these extracts can accurately initiate transcription of several *Drosophila* genes. We use this system to investigate transcription of the *engrailed* gene with the aim of elucidating the complex mechanisms that regulate this transcription.

Results

In vitro transcription in Drosophila *embryo extracts*

A cell-free extract of *Drosophila* embryos was prepared and assayed for its ability to transcribe a number of templates containing cloned *Drosophila* promoters. These templates were cleaved at sites downstream from the in vivo initiation sites and were used in run-off assays (Fig. 1A). The major run-off RNA product, 1.1 kb, 1.2 kb, and 1.4 kb in length for the *fushi tarazu*, alcohol dehydrogenase (Adh), and actin 5C genes, respectively, in each case corresponds to a transcript initiating at the in vivo start sites (Fyrberg et al. 1980, Benyajati et al. 1983; Laughon and Scott 1984). The size of the run-off transcript from an *engrailed* template (1.3 kb) was also consistent with the in vivo initiation sites (Drees et al. 1987). Under these conditions of template and extract concentrations, the relative efficiencies of the *fushi tarazu*, Adh, actin 5C, and *engrailed* templates were approximately 10 : 5 : 4 : 1, respectively. For the *engrailed* template, the rate of product accumulation was constant for a 30-

min incubation (data not shown), and the amount of RNA synthesized was proportional to the template concentration. Maximal incorporation was at 40 μg/ml DNA in the presence of 125 μg crude extract (Fig. 1B).

A primer extension assay was used to compare the in vitro initiation sites with the sites utilized in vivo. A 5'-end-labeled oligonucleotide was hybridized to RNA that had been synthesized in vitro or isolated from *Drosophila* embryos, these were elongated with reverse transcriptase, and the products were fractionated by electrophoresis (Fig. 2A). Transcription initiates in vitro at a cluster of closely spaced sites identical to those used in vivo. These are designated by their nucleotide spacing in the direction of transcription, with the major sites of initiation at residues $+1$, $+2$, $+5$, and $+6$. The *Drosophila* embryo extracts therefore accurately initiate transcription from the *engrailed* promoter.

Fractionation of the embryo extract

The embryo extract was fractionated by a procedure that has been used previously to identify transcription components of HeLa cells (Dynan and Tjian 1983) and *Drosophila* tissue culture K_c cells (Heberlein et al. 1985). After dialysis, the crude extract was applied to a heparin-Sepharose column and eluted with 0.4 M KCl (fraction H.4; nomenclature of Heberlein et al. 1985). The H.4 fraction was applied to a DEAE-Sepharose column from which a break-through fraction (D.1) and a
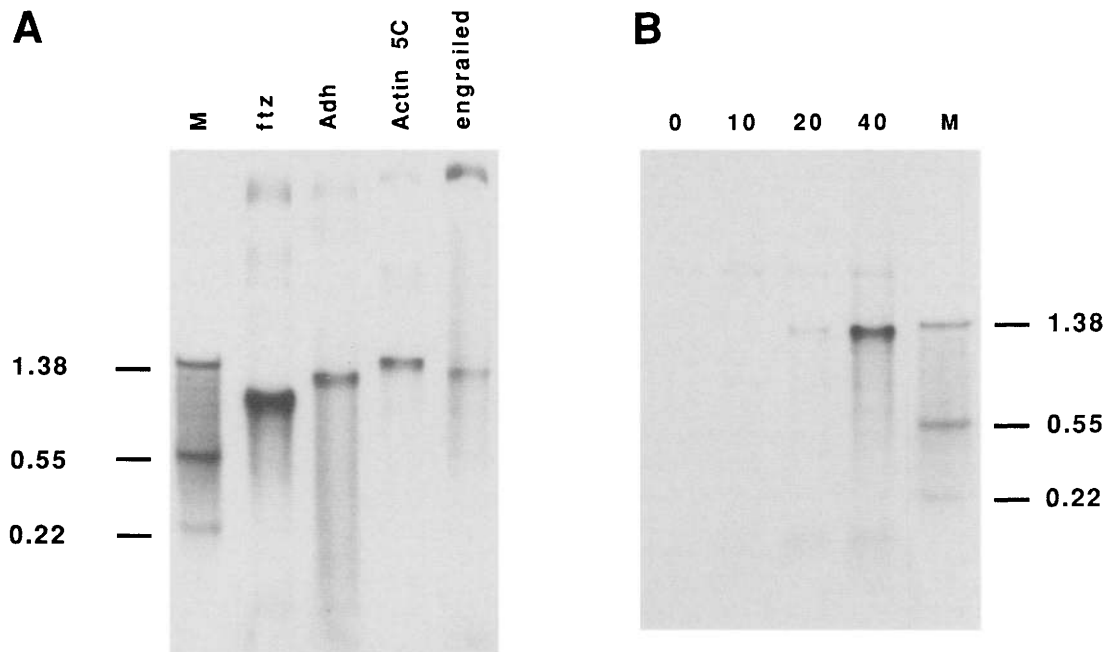

Figure 1. Run-off transcripts synthesized by embryo extracts. (*A*) Assay of various templates bearing *Drosophila* promoters. The indicated template DNAs (20 μg/ml) were each incubated with extract from embryos (5 mg/ml final protein concentration) as described in Experimental procedures. (Lane M) Marker RNAs generated by T7 RNA polymerase whose lengths are indicated in kilobases (kb). Template DNAs were digested as follows: (lane ftz) *Ava*I; (lane Adh (p13E-3)) *Bam*HI; (lane Actin 5C (pDmA₂)) *Hind*III; (lane *engrailed* plasmid (p6:4.7)) *Xho*I. The expected run-off transcripts are 1.1, 1.2, 1.4, and 1.3 kb, respectively. (*B*) Titration of *engrailed* DNA template. Increasing concentrations of plasmid p6:4.7 digested with *Xho*I were incubated with 125 μg crude nuclear extract from embryos. (Lane M) Markers as in *A*.
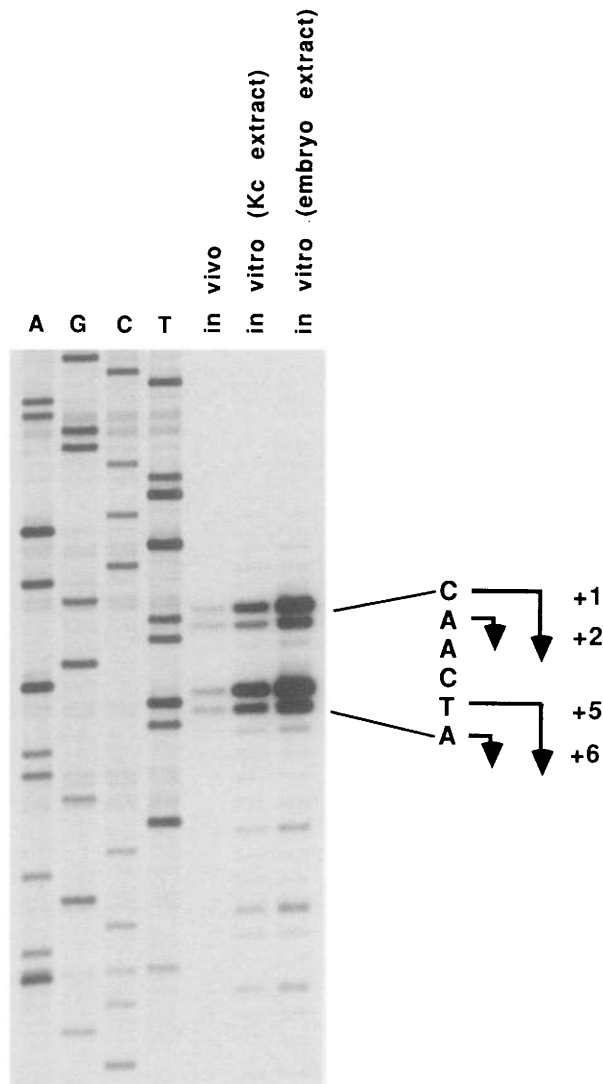
Figure 2. Identification of in vivo and in vitro initiation sites. A primer extension assay with primer 1 (which hybridizes to the coding strand from nucleotide positions +58 to +39, inclusive) was utilized to map the exact sites of initiation as described in Experimental procedures. Sequencing reactions were by the dideoxynucleotide termination method and used the coding strand as template. The sequence printed next to the primer extension products is that of the coding strand. The position to which the longest primer extension product maps is defined as +1; the adjacent upstream position is defined as −1. The complete sequence of this region is in Fig. 5C.

fraction that contained proteins that eluted with 0.25 M KCl (D.25 fraction) were collected.

Each of these protein fractions was tested for transcriptional activity on an *engrailed* template using a primer extension assay (Fig. 3). The crude extract and H.4 fractions alone were active, whereas the D.1 and D.25 fractions were not. However, transcription activity was restored when the D.1 and D.25 fractions were combined. Transcription activity was sensitive to 0.5 μg/ml α-amanitin, indicating that RNA polymerase II is an essential component. Although the identities of the

other components of these DEAE fractions have yet to be established, RNA polymerase II activity was unevenly distributed between them. On a template of single-stranded calf thymus DNA, 90% of the RNA polymerase II activity was present in the D.25 fraction. These results suggest that both DEAE fractions contain essential components: the D.25 has RNA polymerase II activity and the D.1 has some component(s) that enhance(s) specific transcriptional initiation from the *engrailed* promoter. Since the D.25 fraction could not be replaced in the reaction by a preparation of RNA polymerase II purified from *Drosophila* embryos (Coulter and Greenleaf 1982), either a distinct RNA polymerase II subspecies in the D.25 fraction is involved, or the RNA polymerase purified by an assay using a nonselective calf thymus DNA template is inactive in this initiation reaction (Kim and Dahmus 1986), or the required D.25 component(s) include additional activities.
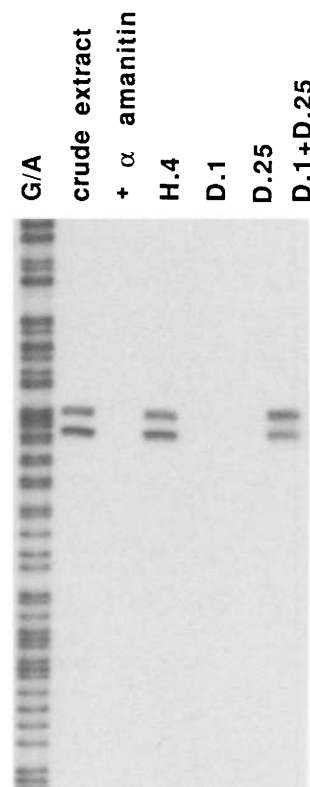

Figure 3. Reconstruction of *engrailed* transcription with partially purified protein fractions. The *engrailed* template (0.5 μg of p6:4.7) was incubated with various protein fractions (crude extract and H.4, 80 μg protein; D.1 alone, 30 μg; D.25 alone, 24 μg; D.1 and D.25, 15 and 12 μg protein, respectively) as described in the text. Primer 2 (which hybridizes to the coding strand from nucleotide position +167 to +148, inclusive) was used to reverse transcribe the RNA products. A Maxam and Gilbert sequencing reaction (G/A) was electrophoresed in an adjacent lane for size reference. Because of the large size of the extension products, they are resolved as two bands, although further electrophoresis would resolve them as four bands (see Fig. 2). Where indicated, α-amanitin was added to a final concentration of 0.5 μg/ml.

+231 TO -305 ENGRAILED PROMOTER FRAGMENT

NONCODING STRAND
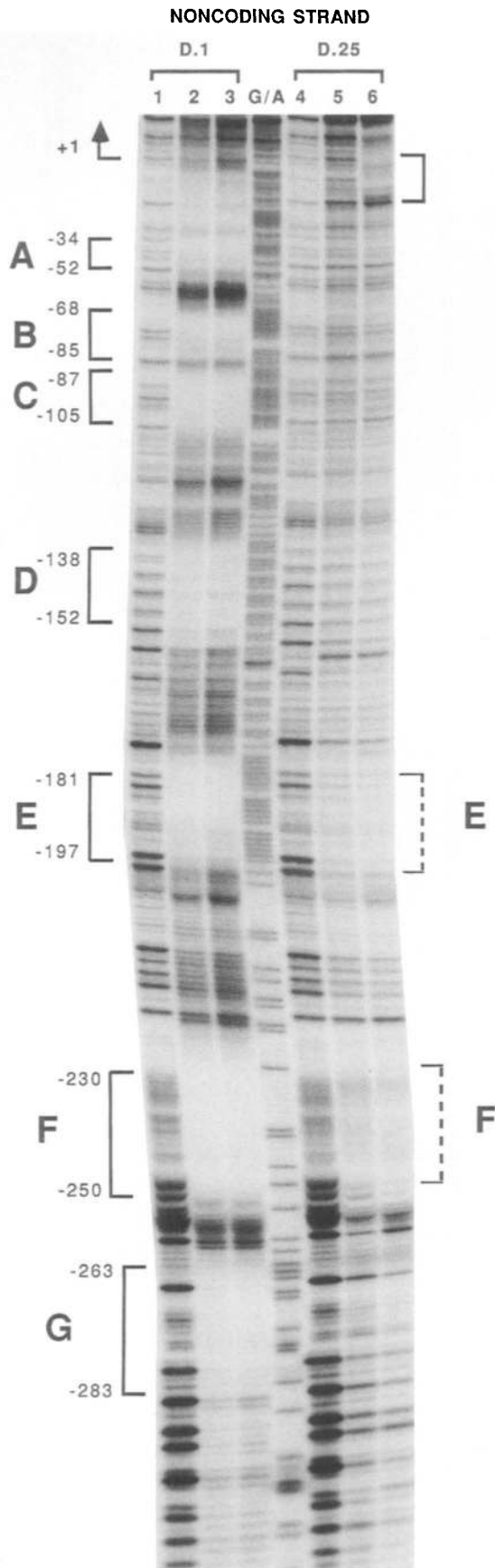
DNase I footprint analysis of the engrailed *promoter*

To analyze the activity conferring specificity on initiation of *engrailed* transcription, the capacity of the D.1 and D.25 fractions to protect the promoter region from DNase I digestion was determined. Several probes were used. The first was designed to investigate the first several hundred nucleotides of sequence upstream of the start sites. It extended from position -305 to $+231$ and was labeled at position -305. Increasing amounts of D.1 were incubated with this probe prior to partial degradation with DNase I (Fig. 4). Seven distinct regions of protection were detected (labeled A–G), the most proximal site (A) mapping to position -34 to -52, and the most distal site (G) mapping between positions -263 and -283. The lengths of the protected regions varied from 14 to 22 residues, and the distances between any two adjacent sites (other than B and C) varied from 11 to 38 residues. The extent of cleavage in the region between sites A and B was strongly enhanced in the presence of some preparations of the D.1 (or H.4, not shown) protein fractions (Fig. 4) but not others (Fig. 5). The source of this variability is not known.

In contrast, D.25 retained little sequence-specific binding activity. With D.25, regions E and F were only partially protected, and no protection was observed for the other regions. This experiment did reveal the presence of a new binding activity whose target sequence $(-9$ to $+10)$ includes the initiation sites. This activity could represent a subspecies of RNA polymerase II, or it could be one of its accessory factors. Purified RNA polymerase II did not bind to this region in these footprint reactions (data not shown).

Using a DNA fragment that extended further upstream, from residues -38 to -419, a protected sequence, called region H, was found to extend from nucleotides -305 to -348 (Fig. 5A). The length of this region is more than twice that of any of the seven other binding sites, and it contains four copies of the pentamer sequence GAGAG (CTCTC). This sequence is also in the center of regions A, B, C, E, and F. Of the protected regions in which this sequence was not found, region D has repeats of the sequence GAGTG (or CACTC). Re-

Figure 4. Footprint analysis of the *engrailed* promoter upstream of its initiation sites. Reactions were carried out as described in Experimental procedures. The probe extended from position $+231$ to -305 and was labeled at the -305 position by a 3'-end-filling reaction. Shown is the region from $+20$ to -300. (Lanes *1* and *4*) DNase I digests of the probe in the absence of any embryo protein fraction. (Lanes *2* and *3*) Binding reactions and DNase digests in the presence of 30 and 60 μg fraction D.1 protein, respectively. (Lanes *5* and *6*) Reactions were carried out in the presence of 15 and 30 μg fraction D.25 protein. Regions of reduced or absent DNase cleavage sites are bracketed and labeled *A–G.* The upturned arrow marks the position of the first of the four initiation sites. Dashed brackets refer to regions of weak protection from DNase cleavage. The unlabeled bracket adjacent to lane *6* denotes the target sequence of a new binding activity present in the D.25 fraction and described in the text.
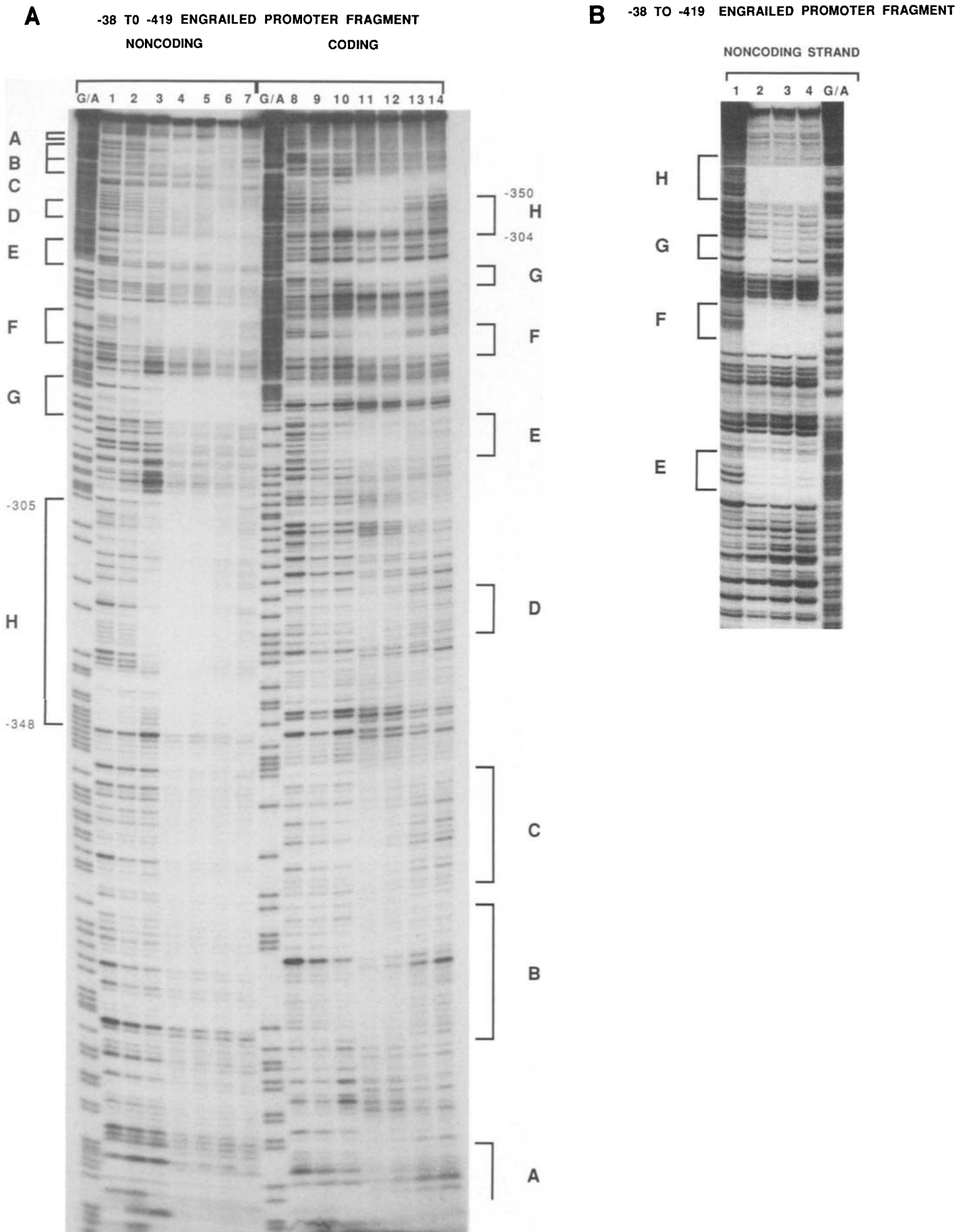
A -38 TO -419 ENGRAILED PROMOTER FRAGMENT

B -38 TO -419 ENGRAILED PROMOTER FRAGMENT

Figure 5. (See facing page for legend.)

gion G, which has imperfect dyad symmetry, has no similar sequence. The nucleotide sequence from −422 to +216 and the footprinted regions are shown in Figure 5C.

The number of distinct binding activities responsible for the eight protected regions was established by competition binding assays. A double-stranded oligonucleotide consisting of the 19 residues of region C was incubated in the presence of the −38/−419 probe and fraction H.4 (Fig. 5A). As the ratio of unlabeled competitor DNA to probe DNA was increased, protection at the six binding sites that contain the GAGAG pentamer disappeared. Region D contains the related pentanucleotide GAGTG, and its protection was consistently weak relative to the other regions, suggesting that its affinity for the binding protein is weak. The footprint of region D was also competed by the region C oligonucleotide. Only region G protection was unaffected in the presence of the region C competitor. In contrast, when a 60-bp *AvaI/DdeI* DNA fragment containing region G was used as a competitor, it suppressed protection only at region G (Fig. 5B). Thus, there appear to be two distinct binding activities specific for the *engrailed* promoter: one binds to sites containing the pentanucleotide sequence GAGAG (or CTCTC); the other recognizes a single site, region G, whose sequence contains dyad symmetry.

Upstream sequences are required for
engrailed *transcription*

If the DNA-binding proteins that recognize sequences upstream of the *engrailed* initiation sites are involved in promoting transcription, then removing these upstream sequences should affect the fidelity or efficiency of RNA polymerase II initiation. To evaluate the role of these sequences, 18 templates were constructed that retained various lengths of *engrailed* DNA upstream of the initiation sites (Fig. 6A). These templates ranged from Δ−887 (retaining 887 bp upstream of the first initiation site and 231 bp downstream) to Δ+15 (retaining only *engrailed* DNA from the downstream positions +15 to +231). Relative transcription activities of these DNAs were determined. Each reaction mixture also contained a control template consisting of *engrailed* promoter DNA with a 15-bp insertion downstream from the initiation sites. Transcripts from this internal control template could be readily distinguished by their increased length, permitting a direct comparison of the template efficiencies of the different promoter deletions. Template efficiencies of 18 promoter fragments were compared qualitatively by inspection of autoradiograms after electrophoretic separation of their primer extension products. Quantitative comparisons were made from densitometric scans of the autoradiograms. Densito-
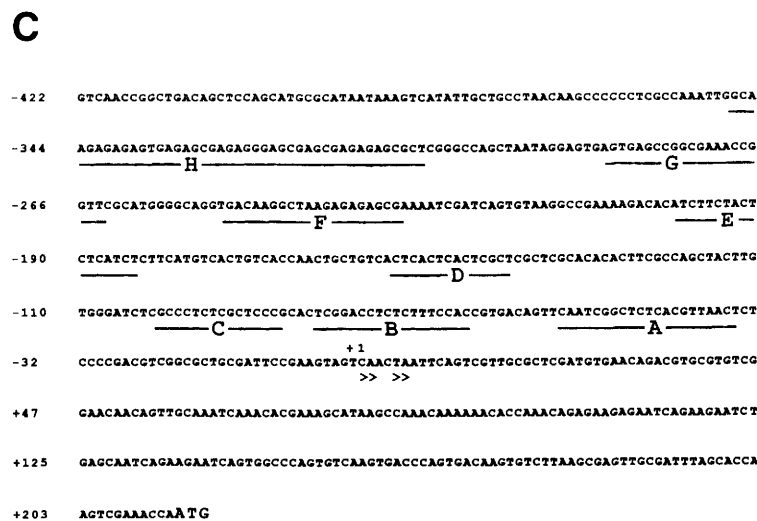
C

```
-422   GTCAACCGGCTGACAGCTCCAGCATGCGCATAATAAAGTCATATTGCTGCCTAACAAGCCCCCCTCGCCAAATTGGCA
                                                                            ——

-344   AGAGAGAGTGAGAGCGAGAGGGAGCGAGCGAGAGAGCGCTCGGGCCAGCTAATAGGAGTGAGTGAGCCGGCGAAACCG
               —————————H—————————————————————————          ———G———

-266   GTTCGCATGGGGCAGGTGACAAGGCTAAGAGAGAGCGAAAATCGATCAGTGTAAGGCCGAAAAGACACATCTTCTACT
       —                  —————————F—————————                          ———E—

-190   CTCATCTCTTCATGTCACTGTCACCAACTGCTGTCACTCACTCACTCGCTCGCTCGCACACACTTCGCCAGCTACTTG
       ————————                                          ————————D————

-110   TGGGATCTCGCCCTCTCGCTCCCGCACTCGGACCTCTCTTTCCACCGTGACAGTTCAATCGGCTCTCACGTTAACTCT
                   ————C————      ————B————                    ————A————
                                          +1
-32    CCCCGACGTCGGCGCTGCGATTCCGAAGTAGTCAACTAATTCAGTCGTTGCGCTCGATGTGAACAGACGTGCGTGTCG
                                       >>    >>

+47    GAACAACAGTTGCAAATCAAACACGAAAGCATAAGCCAAACAAAAAACACCAAACAGAGAAGAGAATCAGAAGAATCT

+125   GAGCAATCAGAAGAATCAGTGGCCCAGTGTCAAGTGACCCAGTGACAAGTGTCTTAAGCGAGTTGCGATTTAGCACCA

+203   AGTCGAAACCAATG
```

Figure 5. Footprint analysis of the *engrailed* promoter in the presence of competitor DNAs. (*A*) A probe extending from position −38 to −419 was labeled at either the downstream (coding) end or the downstream noncoding end by a 3′-end-filling reaction. Size markers were from Maxam and Gilbert G/A sequencing reactions. DNase digestions were carried out as described in Experimental procedures with the following additions: (lanes *1* and *8*) no added protein fraction; (lanes *2* and *9*) 1.4 μg fraction H.4; (lanes *3* and *10*) 14 μg fraction H.4; (lanes *4* and *11*) 14 μg of H.4 and 1.0 μg poly(dI/dC) (as a nonspecific competitor); (lanes *5–7*, *12–14*) identical to lanes *4* and *11* except for increasing amounts of region C oligonucleotide as competitor (4-, 40-, and 400-fold molar excess of sites, respectively; see text). (*B*) Same fragment and procedure as in *A* with only the noncoding strand labeled with polynucleotide kinase: (lane *1*), no added extract protein; (lane *2*) 14 μg fraction H.4 and 1.0 μg poly(dI/dC); (lanes *3* and *4*) same as lane *2* except that a 4- and 40-fold molar excess of a specific region G competitor (see text) was added, respectively. (*C*) The DNA sequence of the coding strand surrounding the *engrailed* initiation sites is shown and the location of the eight footprinted regions on the *engrailed* promoter are indicated. The four initiation sites are represented by > and the first is labeled as +1. The *engrailed* translational initation codon is presented in larger and bold type.
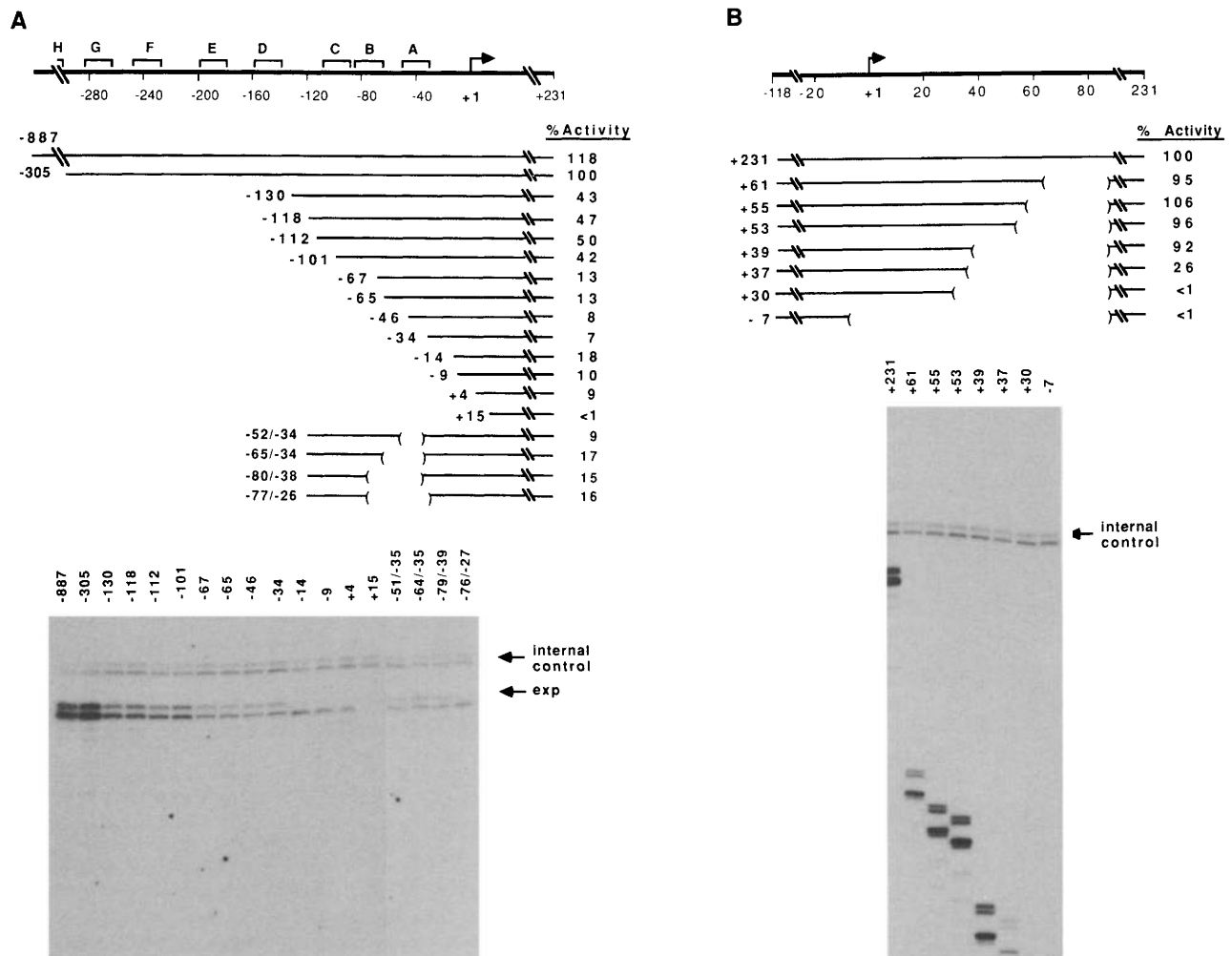
Figure 6. Transcription of promoter deletion mutants. (A) Upstream deletions. The top line represents *engrailed* genomic DNA to which the first initiation site and eight foot-printed regions have been mapped. The mutant templates have been labeled with respect to the upstream end points of the retained upstream sequence (prefixed with Δ in the text). The lines represent *engrailed* genomic DNA left intact in the constructs. Each template was assayed for transcriptional activity by primer extension analysis using primer 2. Each transcription reaction contained 8 μg/ml mutant template, 2 μg/ml internal control template, and 2.4 μg/ml crude extract. (B) Downstream deletions. Templates are labeled as described in the text. Reaction conditions were as in A.

metric values were compared to that of template Δ − 305, which was arbitrarily assigned a value of 100% (Fig. 6A).

A direct correlation between promoter length and template efficiency was observed: As the length of upstream sequence decreased, transcriptional activity decreased in discrete steps. Activity decreased twofold with template Δ − 130, decreased little further with templates Δ − 118, Δ − 112, and Δ − 101, and decreased an additional threefold with template Δ − 67. These quantitative changes correlate with the presence of protected regions B and C between positions − 67 and − 105, and regions D through G between positions − 138 and − 283. No change in the relative yield from the four start sites was observed. These results suggest that the sequence-

specific DNA-binding activities specific for the *engrailed* promoter play a major role in the efficiency of the initiation reaction.

A similar conclusion was suggested by the behavior of mutant templates with a internal deletion. These deletion mutants lack between 16 and 49 bp within the region bounded by positions − 80 and − 26, and are designated by the borders of their deletion end points. Templates Δ − 80/ − 38, and Δ − 77/ − 26 remove sequences that include both regions A and B; their transcription was reduced by approximately 85%. Templates Δ − 51/ − 35 and Δ − 65/ − 34 remove only sequences within and/or around region A; their activities were reduced by 83–91%. The fact that deletion of region A alone affects transcription as extensively as deletion of

region A and B argues either that spacing of region A relative to the initiation sites affects template efficiency or that this most proximal binding site plays a special role in this promoter. However, these results are still consistent with the hypothesis that the sequence-specific DNA-binding activities detected in this region are intimately involved with *engrailed* transcription.

All templates that retained 14 or fewer residues upstream of position $+1$ supported initiation only from sites $+5$ and $+6$. The inability of the $\Delta-14$ template or the $\Delta-77/-26$ internal deletion mutant to utilize initiation sites $+1$ and $+2$ suggests that *engrailed* DNA sequences between position -14 and -34 (the most proximal residue of DNA-binding region A) are used by the transcription apparatus in scanning the promoter sequence for the normal initiation sites. Interestingly, transcription was detectable from template $\Delta+4$, even though all upstream sequences, including initiation sites $+1$ and $+2$, are deleted, and only a single base pair upstream of initiation site $+5$ is retained. The capacity of templates lacking all of the upstream promoter sequences to support transcription, albeit at a low level, implies that sequences within the transcribed region are recognized by the transcriptional machinery that initiates RNA synthesis.

A region within the engrailed transcription unit is essential for transcription in vitro

To determine if sequences within the nontranslated leader of the *engrailed* transcription unit are involved in the initiation reaction, templates were constructed that deleted downstream sequences. These leader sequence deletion mutants had varying amounts of DNA deleted between positions -7 and $+127$, but all retained in common sequences upstream from -7 and downstream from $+127$ (Fig. 6B; designation of mutants is by their deletion endpoint closest to residue $+1$). Transcriptional activity of these DNAs was measured using a primer extension assay in the presence of the control insertion template and the crude extract. Since these deletion mutants remove sequences from within the transcription unit, the lengths of the transcripts and of the primer extension products were correspondingly shortened. Removal of sequences between $+39$ and $+124$ did not affect initiation quantitatively or qualitatively. In contrast, the transcription efficiency of the $\Delta+37$ DNA (which lacks sequences between $+37$ and $+124$) was lowered at all four sites by approximately 75%. The $\Delta+30$ mutant was essentially inactive. We conclude that there is a region within the first 38 residues downstream of site $+1$ that is essential for transcription.

Region C competitor DNA specifically blocks high level transcription of the engrailed promoter

To further evaluate the role of the DNA-binding proteins in promoting transcription, the activities of templates $\Delta-118$ and $\Delta-34$ were determined in the presence of region C oligonucleotide at concentrations pre-

viously shown (Fig. 5A) to suppress protection from DNase I digestion in regions A through F and region H. In these experiments, the Adh proximal promoter served as the control template since this promoter does not contain the GAGAG (or CTCTC) pentamer. Again, template efficiencies were determined from densitometric scans of the autoradiograms (Fig. 7). Transcription of template $\Delta-118$, which contains regions A, B, and C, decreased to 22% of its maximal level as the ratio of competitor DNA to template was increased. In contrast, the competitor had no appreciable effect on transcription of template $\Delta-34$. This template has all upstream binding sites deleted; its activity in the absence of competitor is comparable to template $\Delta-118$ in the presence of competitor. These results confirm the notion that the proteins that bind to the GAGAG (or CTCTC) pentamer sites play a major role in modulating this gene's transcription.

Similar experiments were carried out using the 60-bp *Ava*I/*Dde*I DNA fragment containing region G as the competitor (data not shown). No decrease in transcription was detected with templates containing region G. Thus, it is possible that the DNA-binding activity that recognizes region G plays no role in *engrailed* transcription per se. However, given the sizable distance (263 bp) of region G from the initiation sites, the failure to detect an effect on transcription is more likely due to the insensitivity of in vitro transcription assays to far upstream sequences. Efforts to address this question by moving the region G sequence closer to the initiation sites are currently under way.

In vivo transcription has similar upstream and downstream sequence requirements

To evaluate the importance of the sequences implicated by in vitro initiation assays, transcription was also quantitated in *Drosophila* tissue culture cells. Various lengths of upstream and downstream promoter sequences were fused with sequences encoding the enzyme chloramphenicol acetyl transferase (CAT), and plasmids with these fusion gene constructs were transfected into *Drosophila* Schneider 2 cells. To control for transfection efficiency, a plasmid containing a fusion of approximately 3.5 kb of the upstream *engrailed* promoter with the coding sequences of *E. coli lac Z* was transfected simultaneously. Forty-eight hours after transfection, extracts from the transfected cells were prepared, and CAT and β-galactosidase activities were measured.

CAT fusion genes with various lengths of upstream promoter sequences retained the first 127 bp downstream from the first initiation site. As with the in vitro primer extension experiments, template $\Delta-305$DM26 was arbitrarily assigned a transcriptional value of 100% (Fig. 8A). CAT activity decreased 3-fold with template $\Delta-118$DM26, and 11-fold with template $\Delta-38$DM26. Template $\Delta+2$DM26 reproducibly retained only 3% of the maximal transcription activity.

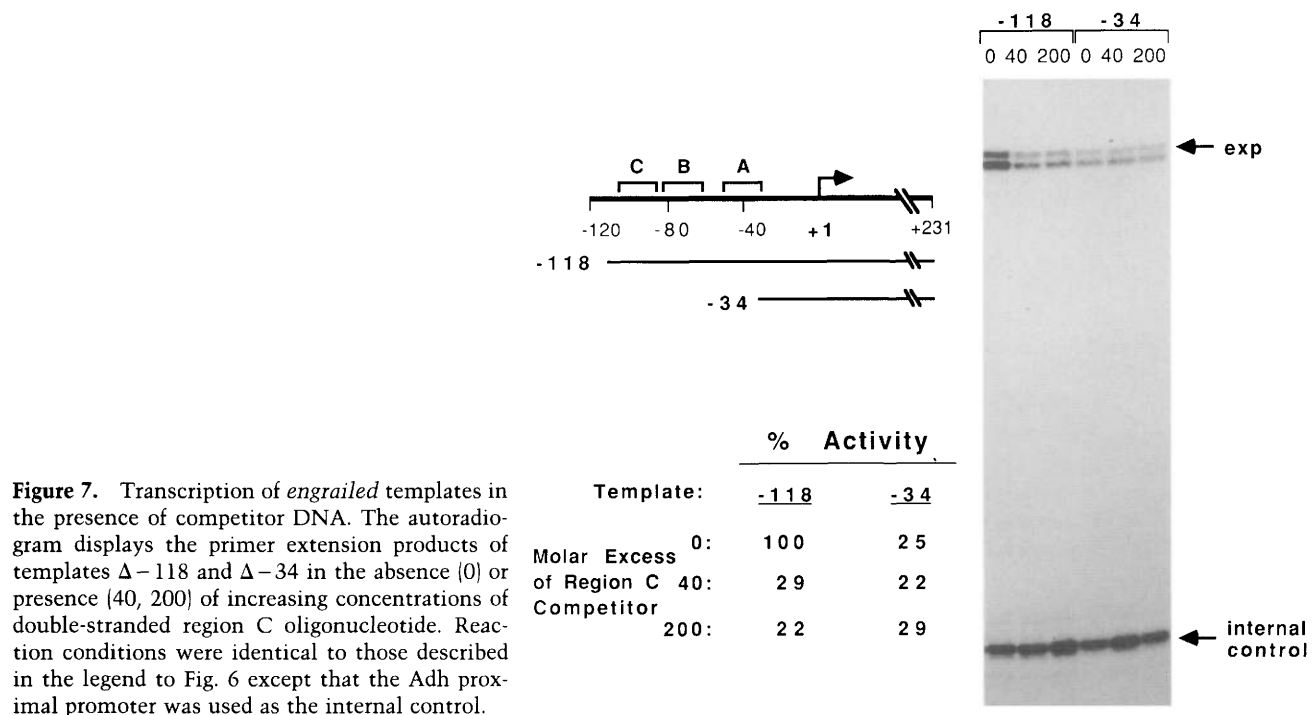Four CAT fusion genes with various lengths of the

Figure 7. Transcription of *engrailed* templates in the presence of competitor DNA. The autoradiogram displays the primer extension products of templates Δ−118 and Δ−34 in the absence (0) or presence (40, 200) of increasing concentrations of double-stranded region C oligonucleotide. Reaction conditions were identical to those described in the legend to Fig. 6 except that the Adh proximal promoter was used as the internal control.

		% Activity	
Template:		-118	-34
Molar Excess of Region C Competitor	0:	100	25
	40:	29	22
	200:	22	29

nontranslated leader were examined. These templates retained 118 bp of upstream sequence, and either 127, 39, 37, or 30 bp of downstream sequence. Template Δ+127DM26 was assigned a transcriptional value of 100% (Fig. 8B). CAT activity in extracts of cells transfected with template Δ+39DM26 was reduced by 54% (Compared to a 10% decrease that was observed in vitro, see Fig. 6B). The relative activity of template Δ+37DM26 (23%), was virtually identical to the in vitro efficiency. Template Δ+30DM26 had only 3% activity relative to template Δ+127DM26, thus confirming that sequences within the transcribed region of the *engrailed* gene are important for transcription. We conclude that both in vivo transcription activity and in vitro initiation efficiency depend on the same upstream and downstream sequences, and that the in vitro assays portray the initiation reaction faithfully.

Discussion

In this first investigation into the molecular mechanisms that directly regulate expression of the *engrailed* gene, we found that the *engrailed* promoter is unusual in several respects. It has neither a TATA box nor a CAAT element, sequences found in many eukaryotic promoters and thought to be important for RNA polymerase II recognition (Shenk 1981). Perhaps the absence of these sequences accounts for the multiple sites of initiation we observed for the *engrailed* gene in vitro and in vivo, and for the relatively low activity of the *engrailed* promoter in the in vitro transcription reaction (Fig. 1). The other promoters assayed in the in vitro system (Adh, Actin 5C, and *fushi tarazu*) have TATA homologies and are all approximately an order of magnitude

more efficient. This suggests that some sequence and protein components that confer specificity to RNA polymerase are different for the *engrailed* gene than for these other genes.

Features of the *engrailed* promoter that were unexpected in view of the extraordinary size of the *engrailed* regulatory region were the high density of functional sequence elements within the first 400 bp upstream of the sites of initiation, and the presence of a critical downstream sequence element. Eight distinct regions in this upstream region were recognized by sequence-specific DNA-binding activities. Interestingly, competition DNA-binding studies suggested that seven of the eight binding sites are recognized by the same protein, implying that several copies of this protein are involved in initiating *engrailed* transcription. Since binding was observed in the absence of RNA polymerase II, the function of the DNA binding is likely to be in the early steps of the initiation process (Hawley and Roeder 1985).

The identified downstream sequence is necessary and, for two of the four sites of initiation, sufficient for at least low levels of transcription. Deletion mutants that lack all of the upstream sequences and retain only the residues four nucleotides downstream from the first initiation site are nevertheless active for the two remaining sites of initiation. The overall template efficiency of such mutants was reduced 10-fold relative to templates with 300 upstream residues. The nature of the role of downstream sequences is unclear, although the close proximity of the essential region to the initiation sites suggests that the RNA polymerase or proteins with which it is directly associated interact with the sequence. A similar requirement for downstream sequences has been observed previously for the *Drosophila*

Ultrabithorax (M. Biggin, pers. comm.) and 74E ecdysone-responsive (C. Thummel, pers. comm.) genes.

Although we have not yet identified the proteins that bind to sequences within 350 bp upstream of the sites of initiation, it is unlikely that they are previously isolated transcription factors. For example, no detectable homology was found between these regions of the *engrailed* promoter and the sequences recognized by either HSTF, a heat shock factor (Parker and Topol 1984b), or Adf1, an *Adh* gene binding protein (Heberlein et al. 1985). Several aspects of the binding sequences within the *engrailed* promoter are, however, noteworthy.

Six of the eight upstream regions (regions A, B, C, E, F, and H) have in common the sequence GAGAG in the middle of each site, but little additional homology. Region D contains the related sequence GAGTG. Given that competition studies suggested that these sites are recognized by the same binding activity, the pentamer may form a core recognition sequence for the same DNA-binding protein. The importance of region A is also suggested by the existence of a homologous sequence in another *Drosophila* gene. Eleven of thirteen residues (−39 to −52) in region A are homologous with a sequence in the *invected* gene (K. Coleman and T. Kornberg, unpubl.). The region of homology in the *invected* sequence is in a similar location relative to the site of transcription initiation (from positions −23 to −37). The *invected* sequence is also protected from DNase I cleavage by DNA-binding activities in the H.4 and D.1 fractions (W. Soeller, K. Coleman, and T. Kornberg, unpubl.), as is region A of *engrailed*. Since the *invected* gene is closely related to the *engrailed* gene in sequence and pattern of expression [it has a homeobox that is virtually identical to that of *engrailed* (Poole et al. 1985) and it is also expressed exclusively in posterior compartment cells (Coleman et al. 1987)], we speculate that homologous sequences in the promoter regions of these genes help to orchestrate their coordinate regulation.

We have noted other homologies in the regions protected from DNase I digestion. Region B contains 9 of 13 base pairs identical to an upstream region of the *Ultra-*
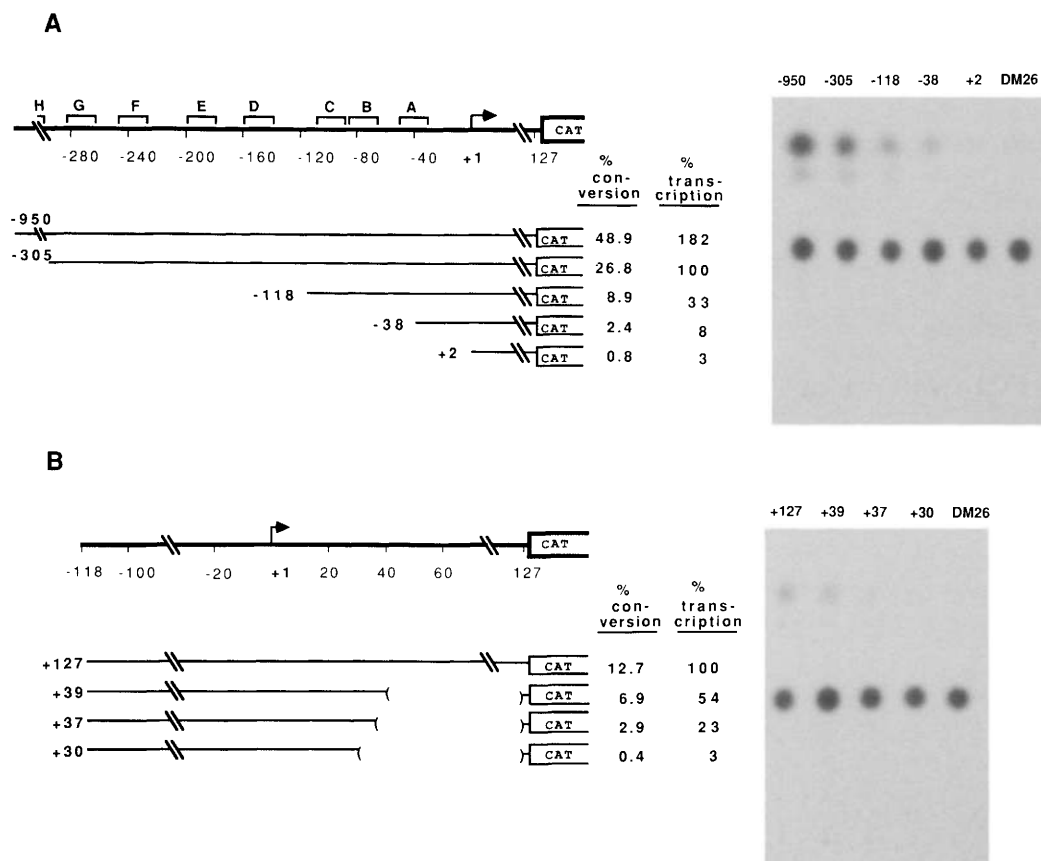

Figure 8. Transient expression assays of upstream and downstream deletion mutant templates. (*A*) *engrailed* templates labeled Δ-DM26 and containing various lengths of upstream sequence were fused to a reporter gene coding for CAT as described in Experimental procedures and were transfected with a β-galactosidase control template into Schneider 2 cells. (% conversion) The fraction of chloramphenicol substrate converted to acetylated forms. DM26 is the CAT plasmid vector with no inserted promoter present. Results reflect values obtained after normalizing β-galactosidase activity measured in each sample and subtracting background levels determined by transfection with pDM26 alone. CAT assays were carried out as described by Gorman et al. (1982) at 37°C for 5 min using 50 μg protein extract per assay. (*B*) Corresponding *engrailed* templates with downstream deletions fused to the CAT reporter gene.

bithorax promoter that can be protected from DNase I digestion (M. Biggin, pers. comm.). Likewise, region C has a significant degree of identity with an additional upstream protected region of the *Ultrabithorax* gene (M. Biggin, pers. comm.), and an upstream region of the *fushi tarazu* gene (−102 to −114 relative to the start site reported by Laughon and Scott 1984). Region G shares no obvious homology with the other six *engrailed* upstream sites. However, it does possess a significant degree of dyad symmetry. Such symmetry in a DNA-binding site is characteristic of DNA sequences bound by certain transcription factors (Topol et al. 1985; Ptashne 1986). In addition, this region shares a high degree of sequence similarity with an element discovered in the dopa decarboxylase (Ddc) promoter of *Drosophila melanogaster* and *D. virilis*. This element, situated between positions −56 and −73 in the *D. melanogaster* promoter, is identical to *engrailed* region G at 10 of 13 nucleotide residues (overlapping but not entirely coincident with the stretch of dyad symmetry). It has been shown to be essential for expression of Ddc in both the larval and adult nervous system (Scholnick et al. 1986). Since *engrailed* function is required in the adult nervous system (Lawrence and Johnston 1984) and since *engrailed* is expressed in the embryonic nervous system (DiNardo et al. 1985; C. Hama and T. Kornberg, in prep.), one may speculate that region G and its binding protein are involved in central nervous system-specific expression. This notion is consistent with our preliminary results that indicate the region G binding activity to be present in embryo but not tissue culture extracts (data not shown.)

We have demonstrated that the activity responsible for protection of the GAGAG sites from DNase digestion also modulates maximal *engrailed* transcription in vitro. Whether this activity or the GAGAG sites are involved in conferring spatial or stage specificity on *engrailed* expression is not known. However, results of P-element-mediated transformation experiments utilizing *engrailed* promoter–β-galactosidase DNA fusion constructs argue that these sites are not sufficient for tissue or stage specificity. Transformants containing constructs with only 1 kb of upstream sequence do not display a clear striping pattern. For a wild-type *engrailed* expression pattern to develop in the epidermis and central nervous system of the embryo, at least 7.5 kb of upstream sequence are required (C. Hama and T. Kornberg, in prep.). Dissection and mapping of *cis*-acting elements within this 7.5-kb region by such transformation experiments will allow us to test promising candidates for their effect on *engrailed* transcription in vitro. Such an approach should allow us to identify, purify, and characterize *trans*-acting activities responsible for the normal tissue-specific gene regulation.

It will be interesting to determine whether some of the activities responsible for protecting the *engrailed* promoter from DNase I digestion are also responsible for modulating transcription of the *Ultrabithorax*, *fushi tarazu*, *even-skipped*, *hairy*, or any other of the array of genes responsible for segmentation and body pattern in *Drosophila*. As noted in the introductory section,

studies of various mutant embryos suggest that genes such as *fushi tarazu*, *even-skipped*, and *engrailed* have cross-regulatory interactions. In vitro transcription reactions should help us understand the nature of these interactions.

Experimental procedures

Preparation of the embryo extract

Oregon R wild-type flies were grown under standard conditions at 25°C in population cages (Kriegstein and Hogness 1974). Embryos of various ages, usually 2–12 hr after egg laying, were collected on corn meal–agar trays, harvested with a solution of 0.7% NaCl, 0.04% Triton X-100 (Sigma), and dechorionated by immersion in 2.25% sodium hypochlorite (1/2 dilution of commercial bleach) for 90 sec. After extensive rinsing with sodium chloride/Triton solution and a final rinse with H_2O, the embryos were quickly blotted dry and weighed. All subsequent steps were carried out at 4°C.

The initial steps for isolation of nuclei closely follows the procedure described in Kuo et al. 1982. The embryos were suspended in 4 ml/gram of embryos in buffer A [0.35 M sucrose, 15 mM HEPES (pH 7.6), 10 mM KCl, 5 mM $MgCl_2$, 0.1 mM EDTA, 0.5 mM EGTA, 1 mM DTT, 1 mM PMSF, 1 μg/ml leupeptin (Boehringer-Mannheim)], and were homogenized by six strokes with a motorized Teflon pestle (G.K. Heller, Floral Park, New York) followed by six strokes with a glass–glass dounce homogenizer (type B pestle). The milky-white homogenate was passed through three layers of pre-wetted Miracloth (Calbiochem) and centrifuged at 4000 rpm in a Beckmann JA20 rotor for 10 min. Pellets were resuspended gently in buffer A (2 ml/g embryos), and the nuclei in 6-ml aliquots were layered over an equal volume of buffer B [O.8 mM sucrose, 15 mM HEPES (ph 7.6), 10 mM KCl, 5 mM $MgCl_2$, 0.1 mM EDTA, 0.5 mM EGTA, 1 mM DTT, 1 mM PMSF, 10 μg/ml leupeptin] in 12-ml graduated polycarbonate tubes. Nuclei were pelleted by centrifugation in a Beckman JS-14 rotor at 5000 rpm for 10 min.

Subsequent steps closely followed the procedure for preparation of K_c nuclear extracts (Parker and Topol 1984a; Heberlein et al. 1985). After noting the volume of the nuclear pellet and removing the supernatant by aspiration, nuclei were resuspended in 4 volumes of buffer C [15 mM HEPES (pH 7.6), 110 mM KCl, 3 mM $MgCl_2$, 0.1 mM EDTA, 1 mM DTT, 1 mM PMSF, and 10 μg/ml leupeptin] and transferred to Beckmann 60Ti polycarbonate tubes. Ammonium sulfate 4.0 M, (pH 7.6) was added stepwise to a final concentration of 0.36 mM, and lysis of the nuclei was accomplished by gentle mixing on a nutator (Clay Adams) for 30 min. The extract was clarified by centrifugation at 35,000 rpm for 1 hr in a Beckman 60Ti rotor. Material floating on the supernatant surface was removed by aspiration. The supernatant was collected by decantation and the active protein fraction was precipitated by the stepwise addition of 0.3 grams of solid ammonium sulfate per milliliter of supernatant. After stirring for 15 min, the precipitate was collected by centrifugation at 15,000 rpm for 15 min in a JA20 rotor. The pellet was gently resuspended with a glass rod in a volume of buffer HEMG [25 mM HEPES (pH 7.6), 40 mM KCl, 12.5 mM $MgCl_2$, 0.1 mM EDTA, 1 mM DTT, 1 mM PMSF, 10 μg/ml leupeptin, 10% glycerol (see Heberlein et al. 1985)]equal to one-third the volume of the nuclear pellet. The resulting suspension was clarified by centrifugation (10,000 rpm/JA20 rotor/10 min) and dialyzed against buffer HEMG until the conductivity was equivalent to a reading obtained with HEMG, 0.12 M KCl (approximately 3 hr). Precipitates were removed by centrifugation

(10,000 rpm/JA20 rotor/10 min) and the clarified nuclear extract was either frozen in small aliquots and stored at −80°C or loaded directly onto a heparin–Sepharose column. The protein concentrations of the extracts varied from 20 to 30 mg/ml as determined by the method of Bradford (Bradford 1976).

Fractionation of the embryo nuclear extract

Nuclear extract was applied to a heparin–Sepharose CL-6B column (Pharmacia, 1.3 ml extract/ml bed volume), equilibrated with HEMG, 0.1 M KCl and washed with 3 column volumes of the same buffer. Protein was eluted with 2 column volumes of HEMG, 0.4 M KCl, and the pooled fractions were dialyzed against HEMG, 40 mM KCl to a conductivity equivalent to that obtained with HEMG, 0.12 M KCl. The dialyzed protein fraction (H.4, 8 mg/ml protein) was applied to a column of DEAE–Sepharose CL6B (Pharmacia, 1.5 ml H.4/ml bed volume), and the column was washed with 3 volumes of HEMG, 0.1 M KCl, and eluted with 2 volumes of HEMG, 0.25 M KCl. Proteins flowing through the DEAE-Sepharose (D.1) or eluted with 0.25 M KCl (D.25) were pooled, aliquoted, frozen in liquid nitrogen, and stored at −80°C.

In vitro transcription

Run-off assays were performed essentially as described in Parker and Topol (1984a). In a final volume of 25 μl, 25 mM HEPES (pH 7.6), 6 mM MgCl$_2$, 50 mM KCl, 1 mM DTT, 0.05 mM EDTA, 3% glycerol, 500 μmM each of ATP, CTP, and UTP, 50 μM GTP, 6 μCi of 410 Ci/mole [^{32}PGTP] (Amersham), 4–20 μg/ml of linearized DNA template, and 40–160 μg crude nuclear extract were incubated for 30 min at 25°C. Reactions were stopped by addition of 4 μl of a stop mix consisting of 200 μg/ml yeast RNA and 40 mM EDTA. The volume was increased to 100 μl with 10 mM HEPES (pH 7.6), 1 mM EDTA, and unincorporated [^{32}PGTP] was removed by spin dialysis through a 1-ml G-50 column (Maniatis et al. 1982). Samples were extracted twice with phenol/chloroform and were precipitated with ethanol. Precipitated RNA product was loaded on a 1.4% formaldehyde–agarose (HGT; FMC) gel and electrophoresed for 3 hr at 6 V/cm. The gel was dried onto DE81 paper (Whatman), and the bands were visualized by exposure overnight to Kodak XAR film at −80°C with a Dupont Cronex screen.

Primer extension assays consisted of two steps. The first was identical to the run-off assay except that no labeled nucleoside triphosphate was present and the templates were not linearized with restriction endonucleases. In the second step, the unlabeled RNA product was resuspended in 3 μl H$_2$O to which was added 2 μl of an end-labeled, synthetic 20-nucleotide-long primer at a concentration of 0.5 μM and a sp. act. of 3 × 10^6 cpm/pmole primer. Two *engrailed* specific primers were used: primer 1 hybridized to the RNA from positions +58 to +39, inclusive; primer 2 hybridized from positions +167 to +148. Adh proximal promoter activity was monitored with a 27-base oligonucleotide specific for this gene (gift of V. Heberlein). The primer was hybridized to the RNA by incubation at 75°C for 2 min, at 55°C for 20 min, and at 25°C for 10 min. The hybridized primers were elongated with murine reverse transcriptase (200 units, BRL) in the presence of 50 mM Tris-HCl (pH 7.9), 50 mM KCl, 10 mM MgCl$_2$, 500 μmM each of dTTP, dATP, dGTP, and dCTP, 100 μg/ml actinomycin D, and 20 units of RNAs in (Promega) in a total volume of 10 μl. The reactions were incubated for 1 hr at 37°C followed by the addition of an equal volume of loading dye mix (bromophenol blue and xylene cyanol in 90% formamide). Aliquots were loaded onto 6% acrylamide (20 : 1 acrylamide/bisacrylamide)/7 M urea sequencing gels (60 cm × 20 cm × 0.4 mm) and run at 30 W for 2 hr before drying and exposure to film.

DNase I protection assays

Footprint reactions were performed in a volume of 20 μl, containing 25 mM HEPES (pH 7.6), 50 mM KCl, 6 mM MgCl$_2$, 0.1 mM EDTA, 10–20 ng of end-labeled DNA fragment, and up to 10 μl of the various protein fractions. Reactions were assembled on ice in the absence of the DNA probe and allowed to equilibrate for 10 min. DNA probe was added and the reactions were incubated on ice for an additional 15 min. The reactions were then placed at room temperature, diluted twofold with a solution of 10 mM CaCl$_2$ and 20 mM MgCl$_2$ and digested with an appropriate amount of DNase I (DPRF, Worthington; 1.6 to 20 ng) for 1 min. The reaction was stopped with 100 μl stop mix (0.1% SDS, 200 μg/ml RNA carrier, 20 mM EDTA, 0.2 M KCl). Reaction products were extracted with phenol/chloroform, precipitated with ethanol, and electrophoresed on 6% or 10% acrylamide (20 : 1 acrylamide/bisacrylamide)/7 M urea sequencing gels.

Cloned segments of Drosophila DNA used for in vitro transcription

DNA fragments used as templates for in vitro transcription included: a plasmid containing the *Drosophila* actin 5C gene in an 8.7-kb *Eco*RI–*Sal*I segment of the plasmid pDmA2 (Fyrberg et al. 1980); the distal and proximal promoters for alcohol dehydrogenase (Adh) in a 4.7-kb *Eco*RI fragment from plasmid p13E-3 cloned into pUC13 (Heberlein et al. 1985); and the promoter for *fushi tarazu* in a 3.5-kb *Hin*dIII fragment.

Templates for transcription of the *engrailed* gene were derived from plasmid p6:4.7, a 4.7-kb *Eco*RI fragment of λphage E7 (Kuner et al. 1985). The template Δ−887 is a *Pvu*II-*Pst*I fragment (extending from −887 to +231) of the 4.7-kb *Eco*RI fragment cloned into pEMBL 18. Upstream deletion templates Δ−130 and Δ−118 were generated by the DNase I/MnCl$_2$ method (Hong 1982); they retain 130 and 118 bp, respectively, of upstream sequence and 231 bp of the *engrailed* transcribed region. The remaining upstream deletion mutants were derived from Δ−118, which was linearized at the upstream end with *Hin*dIII, digested with exonuclease III and nuclease S1, treated with T4 DNA polymerase to fill in the ends, and recircularized with T4 DNA ligase (Henikoff 1984). With the exception of Δ+4 and Δ+15, each mutant retains various lengths of upstream regions and 231 bp of the downstream transcribed region. Δ+4 and Δ+15 retain no upstream sequences and only 228 and 216 bp of transcribed region, respectively.

For a control template, a 12-bp *Bam*HI linker was inserted at a filled *Dde*I site (position +127) in Δ−118. This control template, Δ−118Bam, was also used to generate deletions from the downstream end. To introduce deletions, Δ−118Bam was linearized with *Bam*HI, treated with exonuclease III and nuclease S1, and treated with T4 DNA polymerase to fill in the ends. These fragments were cleaved with *Pst*I (at position +231) and ligated to a purified *Bam*HI/*Pst*I fragment (*engrailed* sequences +127 to +231) of Δ−118Bam, followed by end filling and recircularization. This scheme reintroduced the transcribed region from +127 to +231, allowing use of primer 2 in the primer extension assays.

The internal deletion mutants Δ−52/−34, Δ−65/−34, Δ−80/−38, and Δ−77/−26 were derived from Δ−118. This plasmid was linearized with *Hpa*I at position −38 and was digested with exonuclease III and nuclease S1 prior to end filling and recircularization. This procedure resulted in the removal of between 16 and 49 bp within a region bounded by positions −80 and −26.

To construct the *engrailed*–CAT fusion templates, various lengths of the *engrailed* promoter were joined to the polylinker

directly upstream of the CAT coding sequences of plasmid pDM26 (Mismer and Rubin 1987) as follows: Δ−118DM26, with a *Bam*HI fragment of Δ−118Bam containing 118 bp of upstream and 127 bp of downstream *engrailed* sequences into the *Bam*HI site; Δ−960DM26, with a *ClaI/Bam*HI *engrailed* fragment (−960 to +127) into *SalI/Bam*HI-cut pDM26; Δ−305DM26, Δ−38DM26, and Δ+2DM26 from Δ−960DM26 after digestion with *Ava*I, *Hap*I, and *Hinc*II nucleases, respectively. Δ−118DM26 derivatives with 39, 37, or 30 bp of downstream sequences were generated with *Xba*I−*Bam*HI fragments from deletion mutants Δ+39, Δ+37, and Δ+30, respectively, and *Xba*I/*Bam*HI-cut pDM26.

Transfections were carried out as described by DiNocera and Dawid (1983) using 10 μg of template DNA and 1 μg of the β-galactosidase control DNA per 60-mm tissue culture plate. Cells were harvested after 48 hr and assayed for CAT (Gorman et al. 1982) and β-galactosidase (Miller 1972) activity.

Acknowledgments

Reference

Benyajati, C., N. Spoerel, H. Haymerle, and M. Ashburner. 1983. The messenger RNA for alcohol dehydrogenase in *Drosophila melanogaster* differs in its 5′ end in different development stages. *Cell* 33: 125–133.

Bradford, M.M. 1976. A rapid and sensitive methods for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72: 248–254.

Coleman, K., S. Poole, M. Weir, W. Soeller, and T. Kornberg. 1987. The *invected* gene of Drosophila: Sequence analysis and expression studies reveal a close kinship with the *engrailed* gene. *Genes Dev.* 1: 19–28.

Coulter, D.E. and A.L. Greenleaf. 1982. Properties of mutationally altered RNA polymerase II of Drosophila. *J. Biol. Chem.* 257: 1945–1952.

Dignam, J.D., R.M. Lebovitz, and R.G. Roeder. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* 11: 1475–1489.

DiNardo, S., M.M. Kuner, J. Theis, and P.H. O'Farrel. 1985. Development of embryonic pattern in *D. melanogaster* as revealed by accumulation of the nuclear *engrailed* protein. *Cell* 43: 59–69.

DiNocera, P.P. and I.B. Dawid. 1983. Transient expression of genes introduced in cultured cells of Drosophila. *Proc. Natl. Acad. Sci.* 80: 7095–7098.

Drees, B., Z. Ali, W.C. Soeller, K.G. Coleman, S.J. Poole, and T. Kornberg. 1987. The transcription unit of the Drosophila *engrailed* locus: An unusually small protein of a 70,000 base pair gene. *EMBO J.* 6: 2803–2809.

Dynan, W.S. and R. Tjian. 1983. Isolation of transcription factor that discriminate between different promoters recognized by RNA polymerase II. *Cell* 32: 669–680.

Fyrberg, E.A., K.L. Kindle, N. Davidson, and A. Sodja. 1980. The actin gene of Drosophila: A dispersed multigene family. *Cell* 19: 365–378.

Gorman, C.M., L.L. Moffat, and B.H. Howard. 1982. Recombi-

nant genomes which express chloramphenicol acetyltransferase in mammalian cells. *Mol. Cell. Biol.* 2: 1044–1051.

Gorski, K., M. Carneiro, and U. Schibler. 1986. Tissue-specific in vitro transcription from the mouse albumin promoter. *Cell* 47: 767–776.

Harding, K., C. Rushlow, H. Doyle, T. Hoey, and M. Levine. 1986. Cross-regulatory interactions among pair-rule genes in Drosophila. *Science* 233: 953–959.

Hawley, D.K. and R.C. Roeder. 1985. Separation and partial characterization of 3 functional steps in transcription initiation by human RNA polymerase II. *J. Biol. Chem.* 260: 8163–8172.

Heberlein, U., B. Englund, and R. Tjian. 1985. Characterization of Drosophila transcription factors that activate the tandem promoters of the alcohol dehydrogenase gene. *Cell* 41: 965–977.

Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequences. *Gene* 28: 351–359.

Hong, G. 1982. A systematic DNA sequencing strategy. *J. Mol. Biol.* 158: 539–549.

Howard, K. and P. Ingham. 1986. Regulatory interactions between the segmentation genes *fushi tarazu, hairy,* and *engrailed* in the Drosophila blastoderm. *Cell* 44: 949–957.

Kim, W.-Y., and M.E. Dahmus. 1986. Immunochemical analysis of mammalian RNA polymerase II subspecies. *J. Biol. Chem.* 261: 14219–14225.

Kriegstein, H.J. and D.S. Hogness. 1974. Mechanism of DNA replication in Drosophila chromosomes: Structures of replication forks and evidence for bidirectionality. *Proc. Natl. Acad. Sci.* 71: 135–139.

Kuner, J.M., M. Nakanishi, Z. Ali, B. Drees, E. Gustavson, J. Theis, L. Kauver, T. Kornberg, and P.H. O'Farrell. 1985. Molecular cloning of *engrailed*: A gene involved in the development of pattern in Drosophila melanogaster. *Cell* 42: 309–316.

Kuo, D.-H., H. Giloh, A. Blumenthal, and J.W. Sedat. 1982. A library of monoclonal antibodies to nuclear proteins from *Drosophila melanogaster* embryos. *Exp. Cell Res.* 142: 141–154.

Laughon, A. and M.P. Scott. 1984. Sequence of a Drosophila segmentation gene: Protein structure homology with DNA-binding proteins. *Nature* 310: 25–31.

Lawrence, P. and P. Johnston. 1984. On the role of the *engrailed* gene in the internal organs of Drosophila. *EMBO J.* 3: 2839–2844.

Maniatis, T., E.F. Fritsch, and J. Sambrook. 1982. *Molecular cloning: A laboratory manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Manley. J.L. 1983. Analysis of the expression of genes encoding animal mRNA by in vitro techniques. *Prog. Nucleic Acid Res. Mol. Biol.* 30: 195–244.

Miller, J. 1972. *Experiments in molecular genetics.* Cold Spring Harbor Laboratory, Cold Spring Habor, New York.

Mismer, D. and G.M. Rubin. 1987. Analysis of the promoter of the *ninaE* opsin gene in *Drosophila melanogaster. Genetics* 116: 565–578.

Mizushima-Sugano, J. and R. Roeder. 1986. Cell-type-specific transcription of an immunoglobulin κ light chain in vitro. *Proc. Natl. Acad. Sci.* 83: 8511–8151.

Parker, C.S. and J. Topol. 1984a. A Drosophila RNA polymerase II transcription factor contains a promoter region specific DNA binding activity. *Cell* 36: 357–369.

———. 1984b. A Drosophila RNA polymerase II transcription factor binds to the regulatory site of an hsp 70 gene. *Cell* 37: 273–283.

Poole, S., L. Kauvar, B. Drees, and T. Kornberg. 1985. The *engrailed* locus of Drosophila: Structural analysis of an embryonic transcript. *Cell* **40**: 37–43.

Ptashne, M. 1986. *A genetic switch: Gene control and phage*λ. Blackwell Scientific Publications, Boston.

Scholnick, S., S. Bray, B. Morgan, C. McCormick, and J. Hirsh. 1986. CNS and hypoderm regulatory elements of the *Drosophila melanogaster* dopa decarboxylase gene. *Science* **234**: 998–1002.

Shenk, T. 1981. Transcriptional control regions: Nucleotide sequence requirements for initiation by RNA polymerases II and III. *Curr. Top. Microbiol. Immunol.* **93**: 25–46.

Topol, J., D. Ruden, and C.S. Parker. 1985. Sequence required for in vitro transcriptional activation of a Drosophila hsp 70 gene. *Cell* **4**: 527–537.

Weir, M.P. and T. Kornberg. 1985. Pattern of *engrailed* and *fushi tarazu* transcripts reveal novel intermediate stages in Drosophila segmentation. *Nature* **318**: 433–445.

In vitro transcription of the Drosophila engrailed gene.

W C Soeller, S J Poole and T Kornberg

References	This article cites 34 articles, 10 of which can be accessed free at: **http://genesdev.cshlp.org/content/2/1/68.full.html#ref-list-1**
License	
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**