

Functional domains of the *Drosophila* Engrailed protein

Kyuhyung Han¹ and James L. Manley

Department of Biological Sciences, Columbia University, New York, NY 10027, USA

¹Present address: Department of Genetic Engineering, Hallym University, Chuncho, South Korea

Communicated by D. Ish-Horowicz

We have studied the transcriptional activity of the *Drosophila* homeodomain protein Engrailed (En) by using a transient expression assay employing Schneider L2 cells. En was found to very strongly repress promoters activated by a variety of different activator proteins. However, unlike another *Drosophila* homeodomain-containing repressor, Even-skipped (Eve), En was unable to repress the activity of several basal promoters in the absence of activator expression. These findings indicate that En is a specific repressor of activated transcription, and suggest that En may repress transcription by a different mechanism than Eve, perhaps by interfering with interactions between transcriptional activators and the general transcription machinery. By analyzing the properties of a variety of En mutants, we identified a minimal repression domain composed of 55 residues, which can function when fused to a heterologous DNA binding domain. Like repression domains identified in the *Drosophila* repressors Eve and Krüppel, the En repression domain is rich in alanine residues (26%), but unlike these other domains, is moderately charged (six arginine and three glutamic acid residues). Separate regions of En that may in some circumstances function in transcriptional activation were also identified.

Key words: *Drosophila*/Engrailed protein/functional domains/transcriptional repressor

Introduction

Intricate temporal and spatial regulation of gene expression is the basic principle underlying such processes as the generation of diverse cell types during development and the proper response of individual cells to changing environment. Differential gene expression is achieved primarily by the action of a multitude of transcription factors, which function in most instances to regulate initiation of target gene transcription. It is reasonable to assume that both activators and repressors of transcription are required, since many forms of regulation need both positive and negative modes of action to ensure accurate and timely modulation. In fact, both transcriptional activators and repressors have been identified and characterized during the last few years (for reviews, see Levine and Manley, 1989; Mitchell and Tjian, 1989; Johnson and McKnight, 1990; Ptashne and Gann, 1990; Renkawitz, 1990). However, perhaps reflecting the nature of the assays employed, many more activators have

been identified, and some general features of how they function have begun to emerge. For example, most transcriptional activators are composed of at least two distinct domains: one for sequence-specific DNA binding and the other, which is thought to interact, directly or indirectly, with the general transcription machinery, for transcriptional activation (e.g. Stringer *et al.*, 1990; Dynlacht *et al.*, 1991; Lin *et al.*, 1991; Colgan and Manley, 1993).

In contrast to activators, much less is known about transcriptional repressors. Where studied, these proteins, like activators, appear to contain distinct DNA binding and 'effector' domains. The *Drosophila* Krüppel protein, the first repressor whose repression domain was identified in any detail, contains zinc-finger motifs and a separate alanine-rich region (23 alanines out of 85 residues) as a DNA binding domain and a repression domain, respectively (Licht *et al.*, 1990; but see also Zuo *et al.*, 1991; Sauer and Jäckle, 1991). Other known repressors, such as the *Drosophila* Even-skipped (Han and Manley, 1993) and Engrailed (Jaynes and O'Farrell, 1991) proteins, the Wilms tumor gene product WT1 (Madden *et al.*, 1991) and the v-ErbA oncogene product (Baniahmad *et al.*, 1992), also contain distinct DNA binding and repression domains. A minimal repression domain in Eve has been characterized, and is composed of a small number of residues (57) and is rich in alanines and prolines (Han and Manley, 1993).

In *Drosophila*, it is clear that transcriptional repressors play essential roles in developmentally important processes. For example, the striped pattern of segmentation gene expression during early embryogenesis is a complex process involving numerous hierarchical and cross-regulatory interactions resulting in selective activation and/or repression of appropriate genes (for reviews, see Scott and Carroll, 1987; Carroll, 1990; Pankratz and Jäckle, 1990). One such early acting gene, *tramtrack*, appears to encode a transcriptional repressor that prevents inappropriate expression of several pair-rule genes (Brown *et al.*, 1991; Read *et al.*, 1992). Selective repression also appears to play an important role in the spatial regulation of at least several of the pair-rule genes (e.g. Pankratz *et al.*, 1990; Riddihough and Ish-Horowicz, 1991; Small *et al.*, 1991; Topol *et al.*, 1991). The product of the segment polarity gene *engrailed* (En) appears to repress another segment polarity gene, *ci-D*, such that the two are expressed in opposite segmental compartments (Eaton and Kornberg, 1990).

En is a homeodomain-containing protein (Poole *et al.*, 1985) that binds to DNA with a sequence specificity similar or identical to that of several other homeodomain proteins (Desplan *et al.*, 1988; Hoey and Levine, 1988). Previous transient expression experiments in cultured cells demonstrated that En can repress transcription activated by another homeodomain protein, Fushi tarazu (Ftz), when fragments containing homeodomain binding sites were inserted upstream of a basal promoter (Jaynes and O'Farrell, 1988; Han *et al.*, 1989). En was also found to repress

transcription activated by Ftz *in vitro* (Ohkuma *et al.*, 1990a), and could also repress transcription *in vitro* by competing with the general transcription factor TFIID for the TATA box (Ohkuma *et al.*, 1990b). Repression by En in these cases can be explained by DNA binding site competition. However, En also functions by an apparently distinct mechanism, as it was shown to repress transcription activated by either of two different activators bound to distinct sites separated from the En binding sites, and this repression activity was found to require protein sequences outside the homeodomain (Jaynes and O'Farrell, 1991). Repression in these cases may have been due to quenching, whereby the DNA-bound repressor interacts with the activator to prevent activation, or to direct repression, where the repressor interferes with the assembly or activity of the basal promoter (see Levine and Manley, 1989).

Here we examine further the requirements for repression by En, which suggest that En can efficiently repress activated but not basal transcription. We also precisely map the En repression domain, and in addition identify other regions of the protein that may participate in transcriptional regulation.

Results

Engrailed represses transcription activated by several different activators

En has been shown previously to repress transcription activated by both the yeast GAL4 activator and the mammalian glucocorticoid receptor under conditions where competition between the activator and the repressor for DNA binding sites was unlikely (Jaynes and O'Farrell, 1991). To extend these results, we asked if En can repress transcription brought about by a number of different transcriptional activators, using a co-transfection assay employed previously (Han *et al.*, 1989). Four different GAL4 fusion proteins, each containing a distinct type of activation domain, were used. GAL4-VP16 contains the acidic activation domain of the herpes simplex virus protein VP16 (Ma and Ptashne, 1987a,b); GAL4-ftzQ, a glutamine-rich region from the Ftz protein (Han and Manley, 1993); GAL4-zenST, a serine plus threonine-rich region from the Zen protein (Han and Manley, 1993); and GAL4-CTF1P, a proline-rich region from CTF1 (Mermod *et al.*, 1989; Tanese *et al.*, 1991). Expression vectors encoding each of the GAL4 fusion proteins were co-transfected into *Drosophila* Schneider cells together with a reporter plasmid (pSGNI-CAT; see Figure 1A), increasing amounts of an En expression vector (pAct-en) and a constant amount of a *lacZ* expression vector (pcopiaLTR-*lacZ*) as an internal control. The amount of each activator expression vector was adjusted to attain similar levels of CAT expression (~ 35 - to ~ 95 -fold activation; see Figure 1A legend) in the absence of En expression.

Transcription induced by all four of the activators was very efficiently repressed by En (Figure 1A). At the lowest concentration of pAct-en, expression was in each case reduced by a factor of 5–10, while at higher levels, CAT activities were reduced to near baseline levels. Very strong repression was also observed when 100-fold more pAct-GAL4-ftzQ was used in co-transfections (~ 2200 -fold activation was reduced to ~ 7 -fold activation), and repression was totally dependent on the presence of En binding sites in the reporter plasmid (data not shown). The fact that En could efficiently repress transcription induced by all four

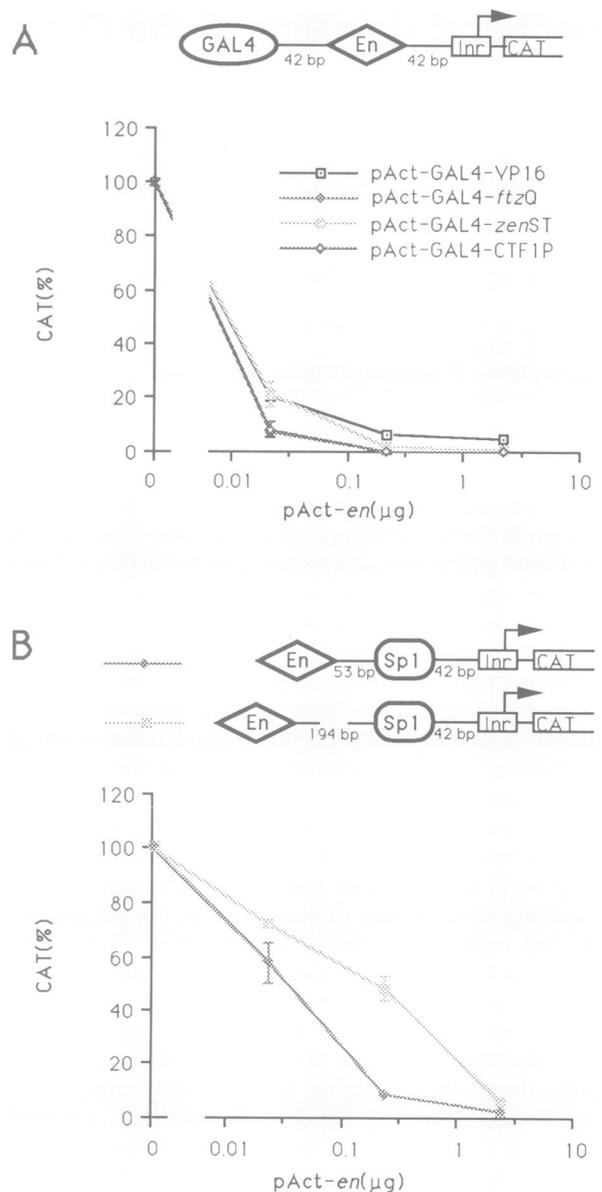


Fig. 1. En represses activated transcription. The indicated CAT reporter plasmids (Inr = TdT initiator element), diagrammed above each graph, were co-transfected into *Drosophila* Schneider L2 cells with a constant amount of the appropriate activator expression vector and the indicated amounts of En expression vector. Relative CAT activities, which are $100 \times (\text{CAT}_{\text{experiment}} - \text{CAT}_{\text{basal}}) / (\text{CAT}_{\text{activated}} - \text{CAT}_{\text{basal}})$, are shown. $\text{CAT}_{\text{activated}}$ is the normalized CAT activity (with respect to the internal *lacZ* control; see Materials and methods) in the presence of the activator, but in the absence of En, and $\text{CAT}_{\text{basal}}$ is the normalized CAT activity in the absence of both activator and En. Error bars indicate standard deviations. **(A)** En represses expression activated by different classes of activation domains. pSGNI-CAT (1 μg) was co-transfected with one of the four different GAL4 fusion expression vectors (1 ng of pAct-GAL4-VP16, 2 ng of pAct-GAL4-ftzQ, 5 ng of pAct-GAL4-zenST or 200 ng of pAct-GAL4-CTF1P), 0–2 μg of pAct-en, 2 μg of pcopiaLTR-*lacZ* and 5 μg of pGem1 carrier. $\text{CAT}_{\text{basal}}$ value was 3.7 ± 0.4 (arbitrary unit; mean \pm standard deviation). $\text{CAT}_{\text{activated}}$ values were 211 ± 13 , 347 ± 73 , 194 ± 59 and 130 ± 2 , respectively. **(B)** En represses expression activated by Sp1. A total of 4 μg of pNSI-CAT (top) or 4 μg of pNGSI-CAT (bottom) was co-transfected as in (A), except that 1 ng of pAct-Sp1 and 2 μg of pGem1 carrier were used. $\text{CAT}_{\text{basal}}$ values were 7.5 ± 0.7 and 8.2 ± 0.6 , respectively. $\text{CAT}_{\text{activated}}$ values were 170 ± 14 and 35 ± 3.3 , respectively.

different types of activators strongly suggests that no sequence-specific protein–protein interactions between the activators and En are required for repression. Although the spacing (42 bp) between binding sites for the activators and En was likely sufficient to prevent DNA binding site competition, to eliminate competition more definitively the GAL4 binding sites were moved further upstream. En was found to repress expression similarly when GAL4–VP16 was bound 209 or 42 bp upstream of En (data not shown), suggesting that short-range physical interference is not responsible for the observed repression activity of En.

We next asked whether En can repress transcription induced by an activator with a different DNA binding domain, i.e. Sp1, to answer two questions that could be raised from the above results. First, although En was able to repress expression induced by four different activation domains, it was possible that repression might have resulted from a specific interaction between the GAL4 DNA binding domain and the En protein. Second, in the experiments described above, En might have blocked the interaction between activators and the general transcription machinery simply by binding between the activator binding sites and the initiator element. To exclude these possibilities, we used two reporter plasmids containing Sp1 (GC boxes; Courey and Tjian, 1988) and En binding sites. The En sites were positioned 53 bp upstream of the Sp1 sites in one, while this distance was increased to 194 bp in the other. Figure 1B shows that En repressed CAT expression from both constructs. Although repression was somewhat less efficient than observed with the GAL4 fusions at low pAct-*en* concentrations, CAT activities were again reduced to near baseline levels at the highest concentration of En expression vector.

The fact that En could repress transcription induced by four different types of GAL4 fusion activators, by Sp1, and by the mammalian glucocorticoid receptor protein (Jaynes and O'Farrell, 1991) strongly suggests that repression by En does not require any specific protein–protein interaction between the activator and En proteins. Neither does short-range physical interference appear to be involved, since En repressed expression from sites located either 194 bp upstream of Sp1 or 209 bp downstream of GAL4–VP16.

Engrailed cannot repress basal transcription

The results described above are consistent with the idea that En can function as a direct repressor, i.e. that its effect on activated expression may reflect its ability to block basal transcription. Indeed, we and others have recently presented data that Eve can function in this manner (Biggin and Tjian, 1989; Johnson and Krasnow, 1992; Han and Manley, 1993). We therefore wished to test whether En can repress a minimal basal promoter in a binding site-dependent manner. Previously, En was shown to repress the TATA-containing promoter *Hsp70*, either containing or lacking En binding sites, by >10-fold *in vitro*, and this was shown to have resulted from competition with TFIID for the TATA box (Ohkuma *et al.*, 1990b). To avoid the possibility of such competition, we initially used the reporter plasmid containing the 17 bp TATA-less TdT initiator element (Smale and Baltimore, 1989) described above (pSGNI-CAT). Increasing amounts of pAct-*en* were co-transfected with pSGNI-CAT and the pcopiaLTR-*lacZ* internal control (Figure 2A). Surprisingly, En could not repress expression from the

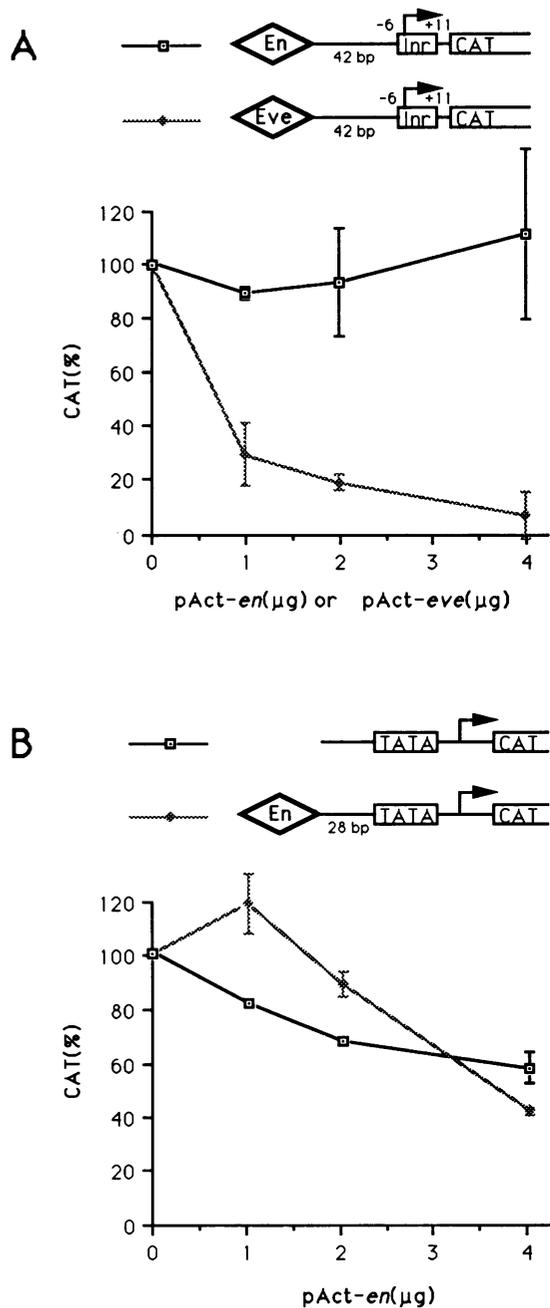


Fig. 2. En cannot repress basal transcription. Minimal promoter–CAT reporter plasmids with or without the En binding sites as diagrammed above each graph were co-transfected with increasing amounts of En expression vector. **(A)** En cannot repress basal transcription driven by the TATA-lacking TdT initiator. pSGNI-CAT (4 µg) was co-transfected with the indicated amounts of pAct-*en* or pAct-*eve* and 2 µg of an internal control, pcopiaLTR-*lacZ*. The total amount of pAct plasmid for each construction was adjusted to 4 µg by adding the vector pActPPA as necessary. Basal CAT activity (without En) was 8.5 ± 0.2 (mean \pm standard deviation). Relative CAT activities are shown in which the activity in the absence of En or Eve expression vectors was taken as 100%. **(B)** En also cannot repress basal transcription driven by the E1b TATA minimal promoter. A total of 4 µg of pE1bTATA-CAT (top) or pSGN-E1bTATA-CAT (bottom), instead of pSGNI-CAT, was co-transfected under the same conditions as in (A). Basal CAT activities were 17.2 ± 0.9 and 7.2 ± 0.1 , respectively.

TdT initiator at all, even at the highest amount (4 µg) of pAct-*en* tested. In contrast, as shown previously (Han and Manley, 1993), Eve effectively repressed expression from

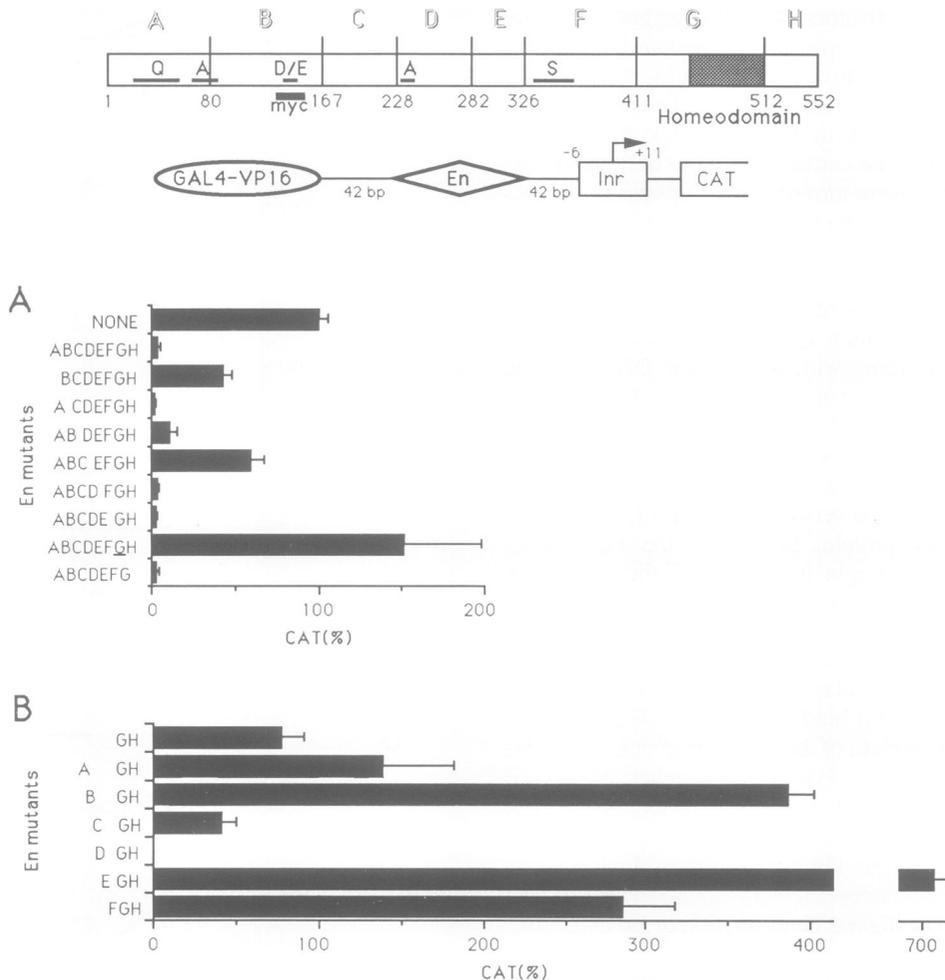


Fig. 3. En contains a single strong repression domain. The indicated truncation and deletion derivatives of En were tested for their ability to repress transcription activated by GAL4-VP16. Full-length En, arbitrarily divided into eight regions A-H (numbers indicate amino acid residues) and the reporter plasmid (pSGNI-CAT) are diagrammed above the graphs. Regions rich in a specific amino acid (denoted by single-letter code) are indicated by thick lines. myc: an 18 amino acid region showing sequence homology to the vertebrate and viral Myc proteins. pSGNI-CAT (1 μ g) was co-transfected with 1 ng of pAct-GAL4-VP16, 4 μ g of each En mutant expression vector (or 4 μ g of an empty Actin plasmid pActPPA for NONE), 2 μ g of pcopiaLTR-lacZ and 3 μ g of pGem1 carrier. Relative CAT activities were determined as in Figure 1. CAT_{basal} value was 2.8 ± 1 and $CAT_{\text{activated}}$ value was 122 ± 7 .

the same promoter. En was also unable to repress another TATA-lacking basal promoter (sequences -38 to -1 from the *en* promoter itself; Soeller *et al.*, 1988) containing En binding sites situated 22 bp upstream. In fact, En weakly activated (by ~3-fold) expression from the *en* basal promoter (results not shown).

It may have been possible that the inability of En to repress basal transcription was a property of the TATA-lacking promoters tested. Although TATA-less promoters appear to require all of the general transcription factors, including the multisubunit TFIID complex, for both basal and activated transcription (Pugh and Tjian, 1991; Zhou *et al.*, 1992), we also wished to test the response to En of a TATA-containing minimal promoter. For this, a reporter plasmid containing the 11 bp adenovirus E1b TATA (Martin *et al.*, 1990; Colgan and Manley, 1992) was used. When pAct-*en* was co-transfected with this plasmid (pE1bTATA-CAT, which lacks En binding sites), CAT expression was slightly decreased (less than a factor of 2), perhaps reflecting competition between En and TFIID for the TATA element (Figure 2B). When the reporter plasmid contained En binding sites, i.e. NP6, the repression detected was again

very weak (Figure 2B). In fact, expression from the NP6-containing plasmid was completely resistant to repression at low concentrations of pAct-*en*. This may reflect the fact that En binds to NP6 more strongly than the TATA element (Ohkuma *et al.*, 1990b), which could have resulted in reduced competition for the TATA element at low En concentrations. A similar weak repression of basal transcription was reported by Jaynes and O'Farrell (1991), which may also have resulted from possible competition with TFIID.

Engrailed contains a single strong repression domain

We next wished to determine which part(s) of the En protein are necessary for transcriptional repression. To map possible repression domain(s), a number of systematic truncations and internal deletions were made by dividing wild-type En into eight regions (69 residues on average). The start codon and consensus leader sequence for all N-terminally truncated En proteins were provided by a 9 amino acid long influenza (flu) hemagglutinin epitope tag (YPYDVPDYA; Wilson *et al.*, 1984; Field *et al.*, 1988), which did not affect the stability or repression activity of the proteins (data not

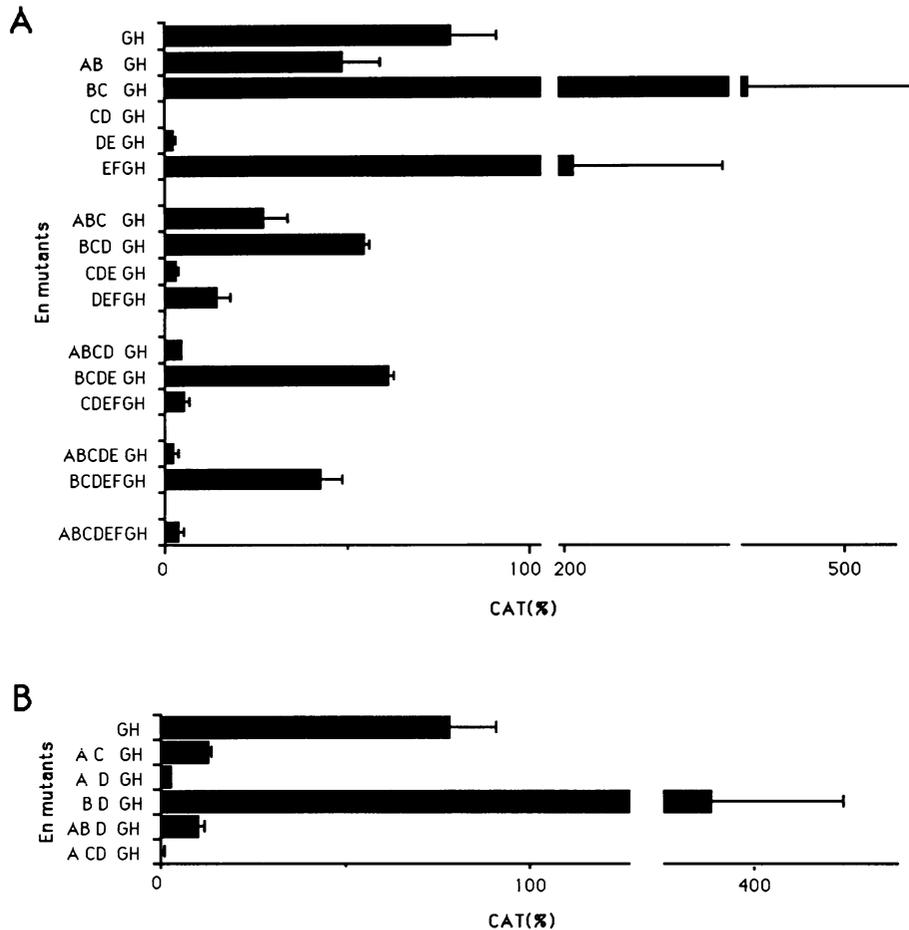


Fig. 4. En also contains a weak activation domain and a separate masking domain. Expression vectors encoding the indicated truncation and deletion derivatives of En were co-transfected with pSGNI-CAT, pAct-GAL4-VP16 and pcopiaLTR-*lacZ* exactly as described in Figure 3. CAT activities are expressed relative to the value obtained from a transfection lacking an En expression vector.

shown). To assay repression, GAL4-VP16 was used as an activator and pSGNI-CAT as the reporter. First, each of the eight regions (A, B, C, D, E, F, G and H) was deleted one at a time (except region G, where only a short segment comprising helix 3 of the homeodomain was deleted; see Materials and methods) to investigate which region(s) is(are) necessary for maximal repression. The results of co-transfections with the appropriate plasmids are shown in Figure 3A. Deleting regions B, E, F or H affected repression very little, if at all. However, removal of regions A, D or G resulted in substantial losses of repression activity. Deletion of region C also resulted in a significant loss of repression activity, although to a lesser degree. Region G contains the homeodomain and loss of repression activity when a portion of it was deleted reflects, at least in part, a loss of DNA binding (see below). Judging from the results of this series of mutants, regions A and D (and possibly C) could constitute repression domains.

Next, fusions of regions A, B, C, D, E and F one at a time to GH were constructed to determine if any of these regions alone is sufficient to convert the inactive GH to a repressor. [GH instead of G served as a minimal DNA binding domain because both were inactive, while GH was more stable than G in the nuclei of transfected cells (data not shown).] The results of co-transfections are shown in Figure 3B. Regions B, E and F could not convert GH to a repressor, as expected from the deletion analysis shown

above. In fact, all three were able to function as weak activation domains. Fusion of D or C to GH resulted in a potent and a weak repressor, respectively. In fact, DGH was a stronger repressor than wild-type En, completely blocking activated expression. Additionally, like native En, DGH was unable to repress basal expression (data not shown). On the other hand, the AGH fusion protein was completely inactive as a repressor, which contrasts with the requirement of region A established by the deletion analysis shown in Figure 3A. This could mean that region A must act together with another part of En to constitute a strong repression domain. A similar but not identical phenomenon was observed in the course of mapping the repression domain of Eve, where a combination of two weak repression regions in the same molecule resulted in a potent repressor (Han and Manley, 1993).

A series of fusion proteins was next constructed by adding an increasing number of contiguous regions (Figure 4A) or by adding combinations of domains (Figure 4B) back to GH to test the idea that En might have separate, weak repression domains that can work together to form a strong repression domain. The first thing that stands out from the results of co-transfections with these constructs is that the weak activation domain B was dominant over repression domains C and D. For example, CGH was a weak repressor (Figure 3B), while BCGH activated CAT expression by almost 5-fold; CDGH was a potent repressor, while BCDGH

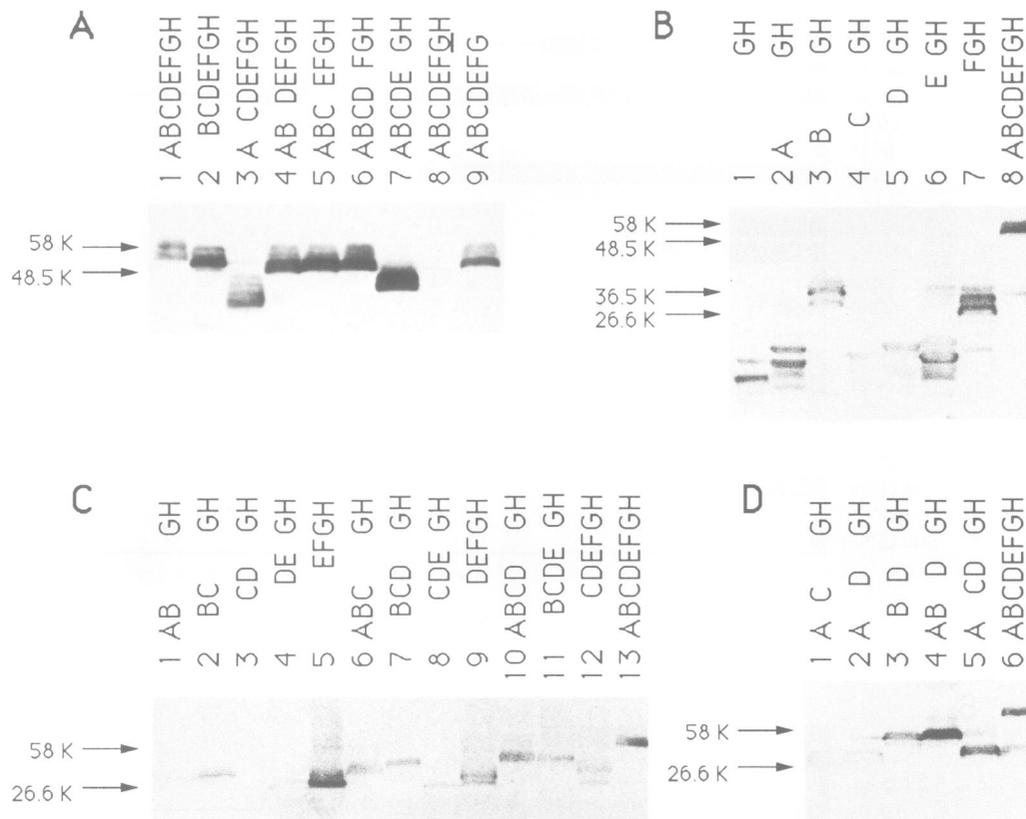


Fig. 5. Accumulation of En proteins in transfected cells. Nuclear extracts were prepared from cells transfected with plasmids encoding the indicated En derivatives and En proteins were detected by Western blotting, as described in Materials and methods. The numbers on the left indicate the mol. wts of protein markers. Some of the mutants, such as CDGH, DEGH and CDEGH in panel C, also ACGH and ADGH in panel D, were underestimated due to interfering protein bands at ~30 kDa (see the text). Wild-type En (designated ABCDEFGH) was included in each set of Western blots to facilitate comparisons between panels.

was almost neutral (Figure 4A). Region B could even convert the extremely strong repressor DGH (Figure 3B) into a weak activator (BDGH; Figure 4B). A second conclusion is that the neutral or weak activation domain A suppressed the activity of region B, but not other regions such as C or D, so that the activity of the strong repression domain D became prominent even when region B co-exists in the same molecule. For example, while BGH activated CAT expression by 4-fold (Figure 3B), ABGH was a weak repressor (Figure 4A); BDGH also activated expression by 4-fold, but ABDGH repressed by ~6-fold (Figure 4B). However, the activity of the very strong repressor CDGH was not affected by addition of region A (Figure 4B). Third, no combinations of weak domains were able to create a potent repression domain. For example, the repression activity of the fusion protein ACGH was similar to that of CGH (Figure 4B). Also, a combination of the weak repression domain C and the strong repression domain D did not produce a super-strong repression domain, as the repression activities of CDGH and DGH were comparable at all concentrations of expression vector (Figure 4A and results not shown). Taken together, these results suggest that there are regions of En with potentially different activities, that there is a hierarchy among different domains such that the activity of one can influence or override that of another, and there is only one strong repression domain (D).

Accumulation and DNA binding of En derivatives

When analyzing the results of experiments such as those described above, it must be shown that the mutations

introduced did not affect the protein in such a way that its activity was altered due to possibly trivial reasons, i.e. the differences in activities of the mutant proteins could conceivably have reflected one of several variables, not directly related to repression, such as protein stability, nuclear localization and/or DNA binding affinity. We addressed these possibilities for the En derivatives described above, and the results are shown in Figures 5 and 6.

The relative accumulation and nuclear localization of the proteins were measured by immunoblot analysis. To this end, cells were co-transfected with pcpiaLTR-*lacZ* and each of the expression vectors utilized in the experiments shown in Figures 3 and 4, small portions of the harvested cells were assayed for β -galactosidase to adjust for variations in transfection efficiencies, and nuclear extracts were prepared and analyzed by Western blotting using the monoclonal antibody 4D9, which recognizes an epitope in the homeodomain (Patel *et al.*, 1989). The results, shown in Figure 5, make several points. First, the amounts of most of the mutant proteins were similar, differing by at most several-fold, although some mutants, such as CDGH, DEGH and CDEGH in Figure 5C, and ACGH and ADGH in Figure 5D, were underestimated due to abundant co-migrating proteins of ~30 kDa that interfered with staining and/or the epitope-antibody interaction (see Materials and methods). An exception is ABCDEFGH, which gave rise to only a very faint signal. This was not due to a failure in nuclear targeting, since virtually no signal above background was detected inside transfected cells by indirect immunofluorescence (data not shown), nor was it due to

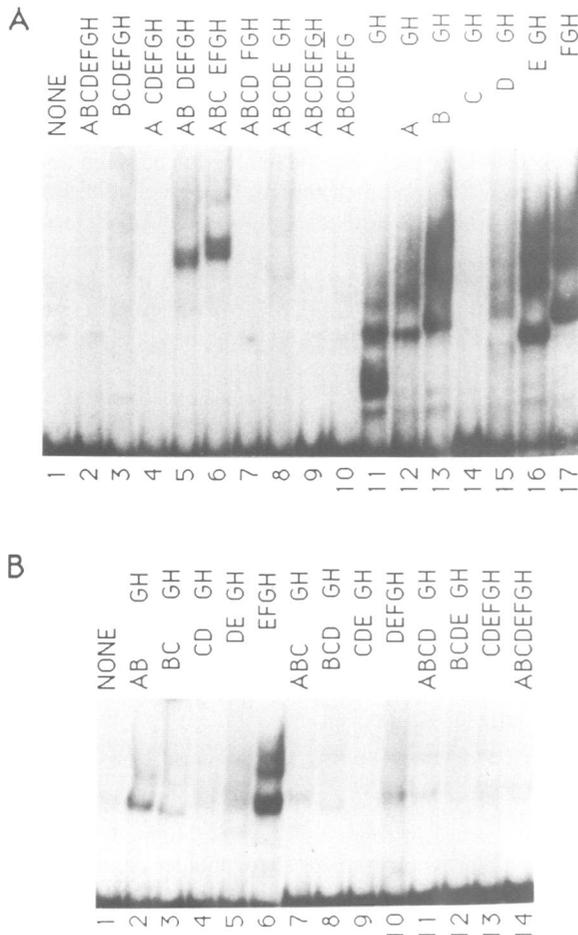


Fig. 6. DNA binding activities of En proteins isolated from transfected cells. The relative DNA binding activities of wild-type and mutant En proteins were measured by performing gel retention assays. The same nuclear extracts employed in the experiment shown in Figure 5 were used. Reaction mixtures contained 0.3 μ l of nuclear extract in a volume of 10 μ l and the DNA probe was an end-labeled 79 bp NP6 fragment. DNA-protein complexes were resolved by gel electrophoresis as described in Materials and methods.

disruption of the epitope since the deleted segment lies outside the epitope (Patel *et al.*, 1989). Therefore, it is likely that this protein was particularly unstable. However, other than this there was no correlation between the amount of protein detected and repression activity. For example, EFGH, which was one of the most stable proteins, was totally inactive, while the very strong repressor CDGH accumulated to relatively low levels. A second point is that wild-type En and most of the mutant derivatives gave rise to multiple bands during electrophoresis, suggesting that the protein is extensively modified post-translationally, probably by phosphorylation at serine residues (Gay *et al.*, 1988).

To estimate the relative DNA binding activities of the wild-type and mutant proteins, gel retention assays were performed with nuclear extracts prepared from transfected cells (see Materials and methods). For these assays, the same relative amounts of nuclear extracts that were used in the Western blot analysis were used in the binding reactions. In each sample, the volume of nuclear extract was adjusted as necessary with nuclear extract prepared from cells transfected with 'empty' expression vector, so that all samples had equivalent amounts of total protein and

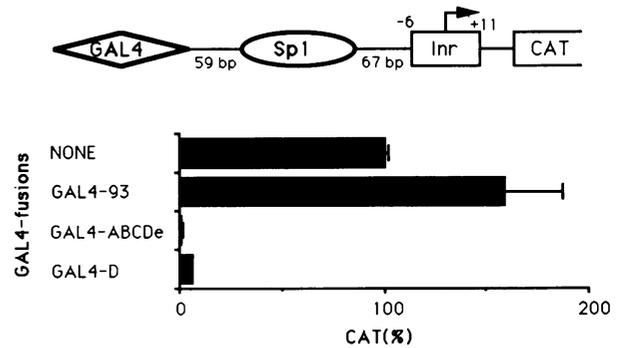


Fig. 7. Repression domain D functions when fused to a heterologous DNA binding domain. The reporter plasmid pGSI-CAT is diagrammed above the graph. The indicated GAL4-fusion proteins were tested for their ability to repress transcription activated by Sp1. GAL4-93 denotes the GAL4 DNA binding domain (residues 1-93). GAL4-ABCDE contains residues 2-298 of En fused to GAL4-93, while GAL4-D is a fusion with En domain D (see Figure 3). pGSI-CAT (1 μ g) was co-transfected with 5 ng of pAct-Sp1, 2 μ g of GAL4-fusion expression vector, 2 μ g of pcpialTR-lacZ and 5 μ g of pGem1 carrier. CAT_{basal} value was 0.5 \pm 0.1 and the CAT_{activated} value was 45 \pm 1.4.

transfected DNA. The DNA probe was an end-labeled 79 bp NP6 fragment, and the specificity of binding was verified with unlabeled competitors (data not shown).

The results of such gel retention assays, with the same mutants analyzed in Figures 3, 4 and 5, are shown in Figure 6, and two principal points emerge. First, in general, smaller proteins tended to have higher affinities for the NP6 DNA than larger proteins. For example, there were more intense shifted bands with the smaller 'addition mutants' (right side of Figure 6A) than with the larger 'deletion mutants' (left side). Second, no mutant protein repressed strongly just because it has a strong DNA binding affinity, and *vice versa*. In fact, there is an inverse correlation, due to domains C and D, between relative DNA binding affinity and repression activity. For example, all the deletion mutants containing both C and D bound to NP6 very weakly, while those lacking either of these two regions bound significantly stronger than the others (lanes 5 and 6). A similar trend was observed with the addition mutants shown on the right side of Figure 6A. The DNA binding affinities of all the addition mutants lacking regions C and D (lanes 12, 13, 16 and 17) were significantly higher than those of either CGH (lane 14) or DGH (lane 15). The potent repression domain D influenced DNA binding somewhat less than the weak repression domain C (compare lanes 14 and 15), suggesting that a very low apparent DNA binding affinity is not an absolute requirement of a strong repressor. The same general trend can also be observed with the mutants analyzed in Figure 6B. These results are similar to those obtained when wild-type and mutant derivatives of Eve were analyzed in gel retention assays, although with Eve the inverse correlation between strong repression and DNA binding was even more pronounced (Han and Manley, 1993).

Repression domain D functions when fused to a heterologous DNA binding domain

The results described above strongly suggest that region DGH of En is necessary and sufficient for very strong and specific repression of activated transcription. Most of En's repression activity therefore appears to reside in domain D,

which constitutes residues 228–282. However, it was conceivable from the experiments described so far that domain D must interact with region G (the homeodomain) and/or H to exert its repression activity. To resolve the question whether domain D is a repression domain independent of a specific DNA binding domain, fragments of En were fused to the GAL4 DNA binding domain (residues 1–92) and the resultant fusion proteins tested for their ability to block expression activated by Sp1 (Figure 7). A GAL4 fusion containing regions ABCD and part of E (GAL4–ABCDe) reduced CAT activity almost to the basal level, indicating that the remainder of En (i.e. the GH region) is dispensable for efficient repression (see also Jaynes and O’Farrell, 1991). As with intact En (see Figure 2), GAL4–ABCDe was unable to repress basal expression (results not shown). This contrasts with similar GAL4–Eve fusion proteins, which retained the ability to repress basal promoters (Han and Manley, 1993, and K.Han, unpublished data). These results support the idea that it is differences in the repression domains of the two proteins that are responsible for their differential abilities to repress basal transcription. A GAL4–domain D fusion protein (GAL4–D) also repressed activated (but not basal) expression, demonstrating that domain D can function efficiently with a heterologous DNA binding domain (Figure 7). It is possible that region D constitutes a minimal repression domain, as efforts to subdivide this region have been unsuccessful (results not shown).

Discussion

The data presented here indicate that Engrailed is a transcriptional regulator that can repress activated, but not basal, transcription. In the Introduction, we described several models for repression by transcription factors: competitive binding, quenching and direct repression. The findings described here argue against the first two models, and favor a form of the third, as the mechanism of En repression in these experiments. The strong repression we observed was almost certainly not due to competition for binding sites with either transcriptional activators or the general transcription machinery. En was able to efficiently repress activated transcription brought about not only by GAL4–VP16 bound 209 bp upstream of the En sites, but also by Sp1 bound 194 bp downstream. In addition, since a TATA-less minimal promoter was used, En did not function by competing with TFIID. The fact that En could repress activated, but not basal, expression from the TdT initiator element when En binding sites were situated 42 bp upstream demonstrates that this distance was sufficient to prevent any steric interference between the bound En protein and the general transcription machinery. These findings strongly suggest that repression by En was not due to DNA binding site competition, a conclusion consistent with previous studies (Jaynes and O’Farrell, 1991). However, we note that En can most likely also function by this mechanism to block the activity of several homeobox activator proteins (Jaynes and O’Farrell, 1988; Han *et al.*, 1989; Ohkuma *et al.*, 1990a). With respect to quenching, En was shown to repress transcription induced by a variety of different activators in a DNA binding site-dependent manner. This is difficult to explain by the quenching model, which would involve specific protein–protein interactions between the activator and the repressor.

Although we could not detect repression of basal transcription, our results are not incompatible with a variation of the direct repression model. It may be that En interacts physically with the general transcription machinery, but in such a way that it does not affect uninduced transcription, but instead interferes with the interaction between transcriptional activators and their target(s). For example, En might physically interact with the multisubunit TFIID complex without interfering with formation of the basal preinitiation complex, but resulting instead in masking, or perhaps release of, TATA binding protein (TBP)-associated factors (TAFs; Dynlacht *et al.*, 1991), which are required for transcription induced by most activators (Dynlacht *et al.*, 1991; Tanese *et al.*, 1991; Zhou *et al.*, 1992). Alternatively, En might be able to destabilize interactions between transcriptional activators and TBP itself, TFIIB, or another essential basal transcription factor.

Our data have identified a single strong repression domain in En, the 55 residue domain D. The fact that domain D functions when fused to heterologous DNA binding domain provides compelling evidence that En contains a functional repression domain separable from and independent of its native DNA binding domain. The repression domains of all three *Drosophila* transcriptional repressors that have been characterized, Krüppel (Licht *et al.*, 1990), Eve (Han and Manley, 1993) and En, function when fused to heterologous DNA binding domains, suggesting that, as with activators, separable DNA binding and repression domains may be a general feature of repressors. As highlighted in Figure 8, the three repression domains share another common feature, which is that they are all rich in alanine residues. In each case, alanines are by far the most abundant amino acid, representing at least 26% of the residues in the currently defined domains. In addition, it is noteworthy that the primary structure of domain D is almost identical to the same region in the *Drosophila virilis* En protein, although the remainder of the protein (with the exception of the region containing the homeodomain) is not well conserved (Kassir *et al.*, 1986). Together these findings suggest that alanine-rich domains may be an important feature of a class of repressor proteins.

Our results suggest that other domains of En might be functionally important. For example, region B contains a

| | |
|---------------|--|
| En D (55) | <u>R</u> QQQAAAAA <u>T</u> AAMML <u>E</u> RANFLNCFNP AA <u>Y</u> PRIHE <u>E</u> I <u>V</u> QSRLRRSAANAVIP <u>P</u> PM |
| Eve C2D2 (57) | PYPPYAPAAAAA <u>A</u> AVGHPSY <u>Q</u> Y RYTPYH <u>I</u> PARPAPPHPAGPHMHHPHMMGS |
| Kr (85) | VHL <u>D</u> RSMSLSPPMSANTSATSAAAIYPAM GLQQAAAA <u>S</u> AFGMLSPTQLLAANRQAAA FMAQLPMSTLANTLFP <u>H</u> NPAAALFGAWAA |
| En A (80) | MA <u>L</u> EDRCSPQSAPSPITLQM <u>Q</u> HLHHQQQ QQQQQQQM <u>Q</u> HLHQLQQLLHQQQL AAGVFHHPAMAF <u>D</u> AAAAA <u>A</u> AAAAA |

Fig. 8. Primary structures of repression domains. The numbers in parentheses indicate the number of amino acid residues in each domain. En D: residues 228–282 of En. Eve C2D2: residues 160–179 plus 209–246 of Eve. Kr: residues 26–110 of Krüppel. En A: residues 1–80 of En. The single-letter amino acid code is used. Alanines (A) and prolines (P) are in bold. Acidic residues, found only in En D and En A, are underlined.

weak activation domain that can under some conditions be dominant over the repression domains. Domain A, which is almost neutral by itself, suppresses the activity of domain B to allow repression. Intriguingly, domain A contains a long stretch (13) of alanine residues (Poole *et al.*, 1985; see Figure 8). These alanines appear to be necessary for suppression of region B activity, since an internal deletion derivative of En lacking the alanine stretch is no longer functional (K.Han, unpublished results). Domain B contains a region rich in glutamate and aspartate residues that has homology to an 18-amino acid sequence present in members of the *myc* family of proteins (Kassis *et al.*, 1986). This homology region may function as a weak transcriptional activation domain, although no particular biological function has yet been assigned to the corresponding region in Myc (reviewed by Prendergast and Ziff, 1992). The only instance where a strong case can be made for En functioning positively is in autoregulation. Intriguingly, support for this notion comes both from genetic experiments in the embryo (Heemskerk *et al.*, 1991) and from the transfection experiment described here.

En cannot effectively repress basal transcription in cultured cells, while another direct repressor, Eve, can. Similar results have been obtained *in vitro*, where Eve was also shown to repress basal transcription (Biggin and Tjian, 1989; Johnson and Krasnow, 1992), while En could repress the

Hsp 70 promoter *in vitro* only very weakly, after preincubation with TFIID to reduce binding of En to the TATA box (Ohkuma *et al.*, 1990b). These findings are consistent with the idea that these two proteins employ somewhat different molecular mechanisms to achieve transcriptional repression. This idea is further supported by the fact that although both En and Eve can repress transcription induced by all the GAL4 fusion proteins tested, their relative effectiveness with one activator was not identical. The serine plus threonine-rich activator GAL4-*zenST* was only moderately repressed by Eve (Han and Manley, 1993), but very strongly repressed by En.

The differences in activities displayed by En and Eve may reflect the fact that, although their repression domains are both alanine rich, they differ from each other in their overall amino acid composition and distribution of proline residues (see Figure 8). The smaller number of prolines in En domain D are clustered at the center and C terminal end, perhaps allowing other regions to form α -helical structures, which is in fact predicted from Chou-Fasman analysis (K.Han, unpublished data). In contrast, the proline residues in the Eve repression domain are distributed in such a frequent and seemingly random way that this domain possibly lacks significant α -helical structure (with the exception of the alanine stretch). These differences may imply that the molecular targets of the two repressors are distinct. By

Table I. En derivatives

| En protein | Primary structure | Calculated mol. wt |
|-----------------|---|--------------------|
| A B C D E F G H | En residues 1-552 (full length) | 59 382 |
| B C D E F G H | M + flu9 + LS + En 80-552 | 52 133 |
| A C D E F G H | En 1-80 + 167-552 | 50 369 |
| A B D E F G H | En 1-167 + 228-552 | 52 832 |
| A B C E F G H | En 1-227 + P + 285-552 | 53 309 |
| A B C D F G H | En 1-282 + 326-552 | 55 172 |
| A B C D E G H | En 1-326 + 410-552 | 51 561 |
| A B C D E F G H | En 1-500 + 512-552 | 57 972 |
| A B C D E F G | En 1-512 + RSRP | 55 433 |
| | M + flu9 + LSR + En 411-552 | 18 281 |
| A G H | En 1-80 + DR + 411-552 | 25 752 |
| B G H | M + flu9 + LS + En 80-167 + R + 411-552 | 27 488 |
| C G H | M + flu9 + En 167-228 + 411-552 | 24 754 |
| D G H | M + flu9 + LS + En 228-282 + DSR + 410-552 | 24 590 |
| E G H | M + flu9 + LST + En 285-326 + 410-552 | 22 468 |
| F G H | M + flu9 + En 326-552 | 25 934 |
| A B G H | En 1-167 + R + 411-552 | 34 737 |
| B C G H | M + flu9 + LS + En 80-228 + 411-552 | 34 056 |
| C D G H | M + flu9 + En 167-282 + DSR + 410-552 | 31 045 |
| D E G H | M + flu9 + LS + En 228-326 + 410-552 | 28 555 |
| E F G H | M + flu9 + LST + En 285-552 | 30 289 |
| A B C G H | En 1-228 + 411-552 | 41 305 |
| B C D G H | M + flu9 + LS + En 80-282 + DSR + 410-552 | 40 347 |
| C D E G H | M + flu9 + En 167-326 + 410-552 | 35 010 |
| D E F G H | M + flu9 + LS + En 228-552 | 36 376 |
| A B C D G H | En 1-282 + DSR + 410-552 | 47 596 |
| B C D E G H | M + flu9 + LS + En 80-326 + 410-552 | 44 312 |
| C D E F G H | M + flu9 + En 167-552 | 42 831 |
| A C G H | En 1-80 + 167-228 + 411-552 | 32 292 |
| A D G H | En 1-80 + D + 228-282 + DSR + 410-552 | 32 043 |
| B D G H | M + flu9 + LS + En 80-167 + 228-282 + DSR + 410-552 | 33 797 |
| A B D G H | En 1-167 + 228-282 + DSR + 410-552 | 41 046 |
| A C D G H | En 1-80 + 167-282 + DSR + 410-552 | 38 583 |

The single-letter amino acid code is used.

Table II. GAL4–En fusion proteins

| GAL4 fusion | Primary structure |
|-------------|-------------------------------------|
| GAL4–93 | GAL4(1–93) + IPI |
| GAL4–ABCDe | GAL4(1–92) + FALL + En(2–298) + RP |
| GAL4–D | GAL4(1–93) + IHRW + En(229–282) + I |

analogy with transcriptional activators, one might expect to detect cooperation between repressors containing distinct repression domains. However, we have failed to detect any synergistic repression by co-expression of Eve and En (K.Han, unpublished results), suggesting that the two proteins cannot function combinatorially. Further studies, especially *in vitro* transcription and analysis of potential protein–protein interactions, will help understand how these repressors function and what their molecular targets are, thereby shedding light on the molecular mechanisms employed by this class of repressor proteins.

Materials and methods

Recombinant plasmids

The plasmids pActPflu and pAct-en were described previously (Han *et al.*, 1989; Han and Manley, 1993). All *en* expression vectors were constructed by standard subcloning methods utilizing these plasmids. Portions of DNA around the junctions in all mutants were sequenced to confirm their identities. Descriptions of all the En derivatives used in this study are shown in Table I.

pAct-Sp1 was described by Courey and Tjian (1988). The GAL4–activation domain fusion protein expression vectors were described previously (Han and Manley, 1993). Details of the GAL4–En fusion proteins used in this study are shown in Table II.

A minimal CAT reporter plasmid, pGAP4-CAT (2.9 kb; Han and Manley, 1993) was used for construction of all the reporter plasmids used in this study. Detailed strategies used for constructing reporter plasmids are available on request.

DNA transfection and transient expression assay

Transient expression assays were performed in duplicate essentially as described by Han *et al.* (1989). In each set of experiments, an appropriate amount of an empty expression vector (pActPPA; Han *et al.*, 1989) was added as necessary to make the concentration of Actin promoter in each transfection the same. In the experiments shown in Figures 5 and 6, 150 mm tissue culture dishes were used for nuclear extract preparation. Each transfection contained 30 µg of En expression vector, 10 µg of pcpialTR-lacZ as an internal control, and 20 µg of pGem1 carrier. Three days after transfection, cells were harvested as described in Han *et al.* (1989). Five per cent of the harvested cells were resuspended in 0.05 ml of 0.25 M Tris–HCl (pH 7.8), and whole-cell extracts were prepared for β-galactosidase assays, while the remainder was used for nuclear extract preparation. Transfection efficiencies were determined and CAT assays were performed as described in Han *et al.* (1989). All transfection experiments were repeated between two and six times.

Western blot analysis

The relative accumulation of each En derivative in nuclear extracts was measured by performing Western blot analysis as described in Han and Manley (1993), except the monoclonal antibody 4D9 (kindly provided by J.Jaynes and P.O'Farrell; Patel *et al.*, 1989) was used as the primary antibody. Gels were electrophoresed for only a short time to facilitate quantitative estimates of the relative accumulation of En derivatives. However, in some cases this resulted in protein overloading, which interfered with detection. For these samples, estimates were made from gels electrophoresed for longer times, where this problem was not observed.

Gel retention assays

DNA binding reactions were performed as described in Han and Manley (1993), except that the reaction mixtures contained 0.3 µl of nuclear extract, 1 µg of poly(dI-dC), 0.2 µg of poly(dA-dT), 1.5 µg of sonicated salmon sperm DNA and 2 ng of labeled NP6 DNA. The same relative amounts of nuclear extracts that were used in the Western blot analysis were used again, although

the volume of nuclear extract for each binding reaction was adjusted to be the same by adding as necessary nuclear extract prepared from cells transfected with 'empty' expression vector, so that all samples had equivalent amounts of total protein. Samples were loaded on a prerun 4% polyacrylamide gel (80:1, acrylamide:bis) containing 1 × Tris–glycine–EDTA. Electrophoresis was at 180 V at 4°C for 1–2 h.

Acknowledgements

We are grateful to M.Green, K.Martin, T.Hoey, J.Jaynes, J.Colgan, N.Tanese and R.Tjian for providing plasmids, and to J.Jaynes and P.O'Farrell for providing the mAb 4D9. We thank D.Read, J.Jaynes and J.Colgan for technical advice and useful discussions, and M.Wang for excellent technical assistance. This work was supported by NIH grant GM 37971.

References

- Baniahmad, A., Kohne, A.C. and Renkawitz, R. (1992) *EMBO J.*, **11**, 1015–1023.
- Biggin, M.D. and Tjian, R. (1989) *Cell*, **58**, 433–440.
- Brown, J.L., Sonoda, S., Ueda, H., Scott, M. and Wu, C. (1991) *EMBO J.*, **10**, 665–674.
- Carroll, S.B. (1990) *Cell*, **60**, 9–16.
- Colgan, J. and Manley, J.L. (1992) *Genes Dev.*, **6**, 304–315.
- Colgan, J. and Manley, J.L. (1993) *Nature*, **362**, 549–553.
- Courey, A.J. and Tjian, R. (1988) *Cell*, **55**, 887–898.
- Desplan, C., Theis, J. and O'Farrell, P.H. (1988) *Cell*, **54**, 1081–1090.
- Dynlacht, B.D., Hoey, T. and Tjian, R. (1991) *Cell*, **66**, 563–576.
- Eaton, S. and Kornberg, T.B. (1990) *Genes Dev.*, **4**, 1068–1077.
- Field, J., Nikawa, J.-I., Broek, D., MacDonald, B., Rodgers, L., Wilson, I.A., Lerner, R. and Wigler, M. (1988) *Mol. Cell. Biol.*, **8**, 2159–2165.
- Fischer, J.A., Ginger, E., Maniatis, T. and Ptashne, M. (1988) *Nature*, **332**, 853–856.
- Gay, N.J., Poole, S.J. and Kornberg, T.B. (1988) *Nucleic Acids Res.*, **16**, 6637–6647.
- Han, K. and Manley, J.L. (1993) *Genes Dev.*, **7**, 491–503.
- Han, K., Levine, M.S. and Manley, J.L. (1989) *Cell*, **56**, 573–583.
- Heemskerk, J., DiNardo, J., Kostriken, R. and O'Farrell, P.H. (1991) *Nature*, **352**, 404–410.
- Hoey, T. and Levine, M. (1988) *Nature*, **332**, 858–861.
- 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.
- Johnson, F.B. and Krasnow, M.A. (1992) *Genes Dev.*, **6**, 2177–2189.
- Johnson, P.F. and McKnight, S.L. (1990) *Annu. Rev. Biochem.*, **58**, 799–839.
- Kassis, J.A., Poole, S.J., Wright, D.K. and O'Farrell, P.H. (1986) *EMBO J.*, **5**, 3583–3589.
- Levine, M. and Manley, J.L. (1989) *Cell*, **59**, 405–408.
- Licht, J.D., Gossel, M.J., Figge, J. and Hansen, U.M. (1990) *Nature*, **346**, 76–79.
- Lin, Y.-S., Ha, I., Maldonado, E., Reinberg, D. and Green, M.R. (1991) *Nature*, **353**, 569–571.
- Ma, J. and Ptashne, M. (1987a) *Cell*, **48**, 847–853.
- Ma, J. and Ptashne, M. (1987b) *Cell*, **51**, 113–119.
- Madden, S.L., Cook, D.M., Morris, J.F., Gashler, A., Sukhatme, V.P. and Rauscher, F.J., III (1991) *Science*, **253**, 1550–1553.
- Martin, K.J., Lillie, J.W. and Green, M.R. (1990) *Nature*, **346**, 147–152.
- Mermod, N., O'Neill, E.A., Kelly, T.J. and Tjian, R. (1989) *Cell*, **58**, 741–753.
- Mitchell, P.J. and Tjian, R. (1989) *Science*, **245**, 371–378.
- Ohkuma, Y., Horikoshi, M., Roeder, R.G. and Desplan, C. (1990a) *Cell*, **61**, 475–484.
- Ohkuma, Y., Horikoshi, M., Roeder, R.G. and Desplan, C. (1990b) *Proc. Natl. Acad. Sci. USA*, **87**, 2289–2293.
- Pankratz, M.J. and Jäckle, H. (1990) *Trends Genet.*, **6**, 287–292.
- Pankratz, M.J., Seifert, E., Gerwin, N., Bill, B., Nauber, U. and Jäckle, H. (1990) *Cell*, **61**, 309–317.
- Patel, N.H., Martin-Blanco, E., Coleman, K.G., Poole, S.J., Ellis, M.C., Kornberg, T.B. and Goodman, C.S. (1989) *Cell*, **58**, 955–968.
- Poole, S.J., Kauvar, L.M., Drees, B. and Kornberg, T. (1985) *Cell*, **40**, 37–43.
- Prendergast, G.C. and Ziff, E.B. (1992) *Trends Genet.*, **8**, 91–97.
- Ptashne, M. and Gann, A.A.F. (1990) *Nature*, **346**, 329–331.
- Pugh, B.F. and Tjian, R. (1990) *Cell*, **61**, 1187–1197.
- Read, D., Levine, M. and Manley, J.L. (1992) *Mech. Dev.*, **38**, 183–196.

- Renkawitz,R. (1990) *Trends Genet.*, **6**, 192–197.
- Riddihough,G. and Ish-Horowicz,D. (1991) *Genes Dev.*, **5**, 840–854.
- Sauer,F. and Jäckle,H. (1991) *Nature*, **353**, 563–566.
- Scott,M.P. and Carroll,S.B. (1987) *Cell*, **51**, 689–698.
- Smale,S. and Baltimore,D. (1989) *Cell*, **57**, 103–113.
- Small,S., Kraut,R., Hoey,T., Warrior,R. and Levine,M. (1991) *Genes Dev.*, **5**, 827–839.
- Soller,W.C., Poole,S.J. and Kornberg,T. (1988) *Genes Dev.*, **2**, 68–81.
- Stringer,K.F., Ingles,C.J. and Greenblatt,J. (1990) *Nature*, **345**, 783–786.
- Tanese,N., Pugh,B.F. and Tjian,R. (1991) *Genes Dev.*, **5**, 2212–2224.
- Topol,J., Dearolf,C.R., Prakash,K. and Parker,C.S. (1991) *Genes Dev.*, **5**, 855–867.
- Wilson,I.A., Niman,H.L., Houghten,R.A., Cherenon,A.R., Connolly,M.L. and Lerner,R. (1984) *Cell*, **37**, 767–778.
- Zhou,Q., Lieberman,P.M., Boyer,T.G. and Berk,A.J. (1992) *Genes Dev.*, **6**, 1964–1974.
- Zuo,P., Stanojevic,D., Colgan,J., Han,K., Levine,M. and Manley,J.L. (1991) *Genes Dev.*, **5**, 254–264.

Received on February 16, 1993; revised on April 5, 1993