Published as: *Evol Dev.* 2008 ; 10(5): 537–545.

Evolution of an insect-specific GROUCHO-interaction motif in the ENGRAILED selector protein

Chris Todd Hittinger¹ and Sean B. Carroll^{*}

Howard Hughes Medical Institute, Laboratory of Genetics, University of Wisconsin-Madison, 1525 Linden Dr., Madison, WI 53706-1534, USA

Abstract

Animal morphology evolves through alterations in the genetic regulatory networks that control development. Regulatory connections are commonly added, subtracted, or modified via mutations in *cis*-regulatory elements, but several cases are also known where transcription factors have gained or lost activity-modulating peptide motifs. In order to better assess the role of novel transcription factor peptide motifs in evolution, we searched for synapomorphic motifs in the homeotic selectors of *Drosophila melanogaster* and related insects. Here, we describe an evolutionarily novel GROUCHO (GRO)-interaction motif in the ENGRAILED (EN) selector protein. This “ehIFRPF” motif is not homologous to the previously characterized “engrailed homology 1” (eh1) GRO-interaction motif of EN. This second motif is an insect-specific “WRPW”-type motif that has been maintained by purifying selection in at least the dipteran/lepidopteran lineage. We demonstrate that this motif contributes to *in vivo* repression of the *wingless* (*wg*) target gene and to interaction with GRO *in vitro*. The acquisition and conservation of this auxiliary peptide motif shows how the number and activity of short peptide motifs can evolve in transcription factors while existing regulatory functions are maintained.

INTRODUCTION

The transcriptional regulatory networks that control development are comprised of interacting proteins that regulate the expression of downstream target genes to promote pattern formation and cellular differentiation (Levine and Tjian 2003; Levine and Davidson 2005). Many transcription factors bind directly to DNA sequences located in the *cis*-regulatory elements of target genes. Once bound, transcription factors can either activate or repress transcription directly or by recruiting co-factors that possess regulatory properties (Orphanides and Reinberg 2002).

During evolution, the transcriptional readout of developmental networks can be modified either by altering specific *cis*-regulatory elements of individual genes, by altering the protein-coding regions of the regulatory proteins, or both (Davidson 2001; Levine and Tjian 2003; Carroll 2005; Carroll et al. 2005). The evolution of regulatory protein function is generally viewed as rare and constrained due to the attendant pleiotropic effects on the expression of its suite of downstream regulatory targets, especially for transcription factors deployed in multiple tissues (Stern 2000; Mann and Carroll 2002; Wray et al. 2003; Carroll 2005; Carroll et al. 2005; Wray 2007). In contrast, altering a *cis*-regulatory element can allow the evolution of the expression of a single gene in a single tissue at a single developmental stage.

^{*}Author for correspondence: sbcarroll@wisc.edu, (608) 265-2004, FAX: (608) 262-9343.

¹Current address: Center for Genome Sciences, Department of Genetics, Washington University in St. Louis School of Medicine, 4444 Forest Park Ave., St. Louis, MO 63108-2212, USA

Nevertheless, the coding sequences of some developmental regulators clearly have evolved in functionally and evolutionarily important ways (Mann and Carroll 2002; Hsia and McGinnis 2003; Carroll 2005; Carroll et al. 2005; Wagner 2007; Wagner and Pyle 2007). Detailed studies of several transcription factors have identified both gains or losses of short activity-modulating peptide motifs outside of their DNA-binding domains (Grenier and Carroll 2000; Alonso et al. 2001; Lohr et al. 2001; Galant and Carroll 2002; Ronshaugen et al. 2002; Shiga et al. 2002; Lamb and Irish 2003; Hittinger et al. 2005; Lohr and Pick 2005). Some evolutionary changes appear to be qualitative, such as the acquisition of limb-repression function by ULTRABITHORAX (UBX) (Grenier and Carroll 2000) or the abandonment of the ancestral homeotic function and the adoption of its derived role in segmentation by FUSHI TARAZU (FTZ) (Alonso et al. 2001; Lohr et al. 2001). However, closer examination of the functions of specific peptide motifs suggests that even these qualitative changes may have been achieved through the evolution of multiple contributing peptide motifs. For example, the “QA” motif of insect UBX was capable of conferring novel regulatory potential in experiments that tested its evolved sufficiency in modern *Drosophila melanogaster* (Galant and Carroll 2002), but it displayed quantitative contributions in tests of its evolved necessity (Gebelein et al. 2002; Ronshaugen et al. 2002; Hittinger et al. 2005), suggesting that other motifs also play quantitative roles in protein activity (Tour et al. 2005). Similarly, the “YPWM” and the “LXXLL” motifs of FTZ have well-defined homeotic and segmentation roles, respectively, but displayed quantitative and/or epistatic contributions in some assays (Lohr and Pick 2005). Taken together, these case studies suggest that quantitative changes frequently underlie changes in regulatory protein activity.

We searched several arthropod selector proteins for additional examples of peptide motifs that might have contributed to the evolution of their transcriptional regulatory activity. We focused on synapomorphic (conserved, derived) peptide motifs possessed by *D. melanogaster* and related insects that might have novel activities, as observed for the QA and LXXLL motifs. One evolutionarily novel motif in the ENGRAILED (EN) selector protein resembled previously characterized GROUCHO (GRO)-interaction motifs. Here, we show that this motif has been conserved by purifying selection in dipterans and lepidopterans, but it is not present in any non-insect EN homologs. We further demonstrate that the ehIFRPF motif contributes to *in vitro* interaction with GRO and *in vivo* repression ability. The EN-GRO interaction is ancient and is generally thought to depend on the previously characterized “engrailed homology 1” (eh1) peptide motif (Logan et al. 1992; Smith and Jaynes 1996; Tolkunova et al. 1998) conserved in all bilaterians sampled. Our new results suggest that EN acquired a second GRO-interaction motif during insect evolution that quantitatively modulates the strength of the EN-GRO interaction. The evolution of short, degenerate motifs affecting existing functions may be common among transcription factors.

RESULTS

Identification of an evolutionarily novel GRO-interaction motif in *Drosophila* EN

Like most selector proteins, insect EN sequences contain a number of uncharacterized synapomorphic peptide motifs outside of its DNA-binding homeodomain, some of which have been noted before (Fig. 1) (Hui et al. 1992; Logan et al. 1992). These motifs may have evolved to impart novel regulatory capacity or to modulate existing regulatory functions, but most lack sequence similarity to functionally characterized motifs. However, one previously unnoted and uncharacterized motif perfectly matches the consensus of the “WRPW”-type motif within the dipteran and lepidopteran sequences (Fig. 1). We call this motif the “ehIFRPF” motif after its sequence in *D. melanogaster*. This ehIFRPF motif is located between the well-characterized eh1 motif and the “eh2” motif that interacts with

EXTRADENTICLE (EXD) (Peltenburg and Murre 1996). Below, we show that the ehIFRPF constitutes an evolutionarily novel, second GRO-interaction motif.

The co-repressor GRO interacts with transcription factors via two distinct types of short peptide motifs (Jimenez et al. 1997; Courey and Jia 2001). The eh1-type motif first described in EN contains a loose 14 residue-consensus that is shared with many related homeobox-containing proteins, including its paralog INVECTED (INV) (Smith and Jaynes 1996). The shorter WRPW-type motif was first described in HAIRY (Paroush et al. 1994) and has a $\psi\Omega(K/R)P\Omega$ consensus. [Peptide motif abbreviations are according to standard nomenclature where ψ = aliphatic residue (I, L, M, or V) and Ω = aromatic residue (F, W, or Y) (Aasland et al. 2002).] Note that the first aliphatic residue is not necessarily present when both aromatic residues are tryptophan, as in the case of HAIRY. Tryptophan is believed to provide a more hydrophobic pocket and stronger protein-protein interaction that may not require the additional aliphatic residue (Kobayashi et al. 2001). Both types of GRO-interaction motifs physically interact with the same three-dimensional region of GRO but have different contact residues and conformations (Jennings et al. 2006).

The peptide sequence of the ehIFRPF motif is perfectly conserved among all 12 *Drosophila* species for which genome sequences are available (Adams et al. 2000; Richards et al. 2005; Clark et al. 2007) and varies within the known WRPW-type consensus in all other available dipteran and lepidopteran sequences (Keys et al. 1999; Holt et al. 2002; Xia et al. 2004; Nene et al. 2007) (Fig. 1). A putative ehIFRPF motif (GFKPY) is present in a similar position in *Tribolium castaneum* EN (Richards et al. 2008) and may be functional, but instances where the first aliphatic residue of the motif is missing have only been described when tryptophan is the aromatic residue in both cases (Paroush et al. 1994). The next closest matches are in *Periplaneta americana* EN1 (Marie and Bacon 2000) and *Schistocerca gregaria* EN-1 (Peel et al. 2006), but histidine has never been known to substitute for an aromatic residue in WRPW-type motifs, making it less likely that these sequences function as GRO-interaction motifs. All other EN homologs lack any sequence resembling the ehIFRPF motif [*i. e.* allowing at most one mismatch against the $\psi\Omega(K/R)P\Omega$ consensus], including all insect INV, *Apis mellifera* EN (Honeybee Genome Sequencing Consortium 2006), and all non-insect EN/INV homologs. The absence from *A. mellifera* is particularly striking because it suggests either that the ehIFRPF motif has been lost in this lineage [perhaps through gene conversion; see a recent comprehensive phylogenetic analysis of insect EN/INV homologs (Peel et al. 2006)] or that the putative ehIFRPF motif in *T. castaneum* evolved independently in parallel or is non-functional.

The ehIFRPF motif has been under purifying selection

The conservation of the ehIFRPF motif suggests that it is important to EN function and may be under purifying selection. In order to test the latter hypothesis, we developed a statistical method tailored to address special analytical challenges posed by short motifs. Coding sequences are generally assumed to be under purifying selection, but most tests have focused on large, globular protein domains that are easily alignable (Nei and Kumar 2000). However, x-ray crystallography suggests that animal homeodomain and many other transcription factors only form reliably ordered structures in their DNA binding domains. For example, the two of the best-known homeodomain structures of EN and UBX/EXD contain only the homeodomain and immediately adjacent sequences (Kissinger et al. 1990; Passner et al. 1999). The UBX/EXD structure is particularly informative because the evolutionarily labile linker region between the homeodomain and the EXD-interaction motif is disordered (Passner et al. 1999). Sequences outside of the homeodomain may play important roles that include protein-protein interactions and may be only conditionally ordered in the presence of interaction partners. An important evolutionary consequence of these differences in protein structure and function is that much of the sequence of these

proteins is unalignable, except for short and often degenerate peptide motifs (Fig. 1). Intuitively, we suspect such peptide motifs are under purifying selection, but efforts to illustrate this have focused on the development of sliding window scanning tools that have not incorporated statistical tests because of the absence of a widely agreed upon model of gap evolution (Simon et al. 2002).

The availability of genome sequence data for many insect species allows us to use the phylogenetic depth of conservation to overcome the limited length of conservation for short peptide motifs. Similar to a previously described test for positive selection (Zhang et al. 1997; Nei and Kumar 2000), we applied a parsimony-based phylogenetic distance framework that summed the synonymous and non-synonymous distances across the entire tree of interest. However, we inverted the test to ask whether there is a statistical deviation from neutrality that favored synonymous substitutions (see MATERIALS AND METHODS for details). Application of this method to the ehIFRPF motif conclusively demonstrated that this motif has been under purifying selection within the *Drosophila* genus ($P < 10^{-3}$), within the dipterans ($P < 10^{-4}$), and within the dipteran/lepidopteran clade ($P < 10^{-4}$).

The ehIFRPF motif contributes to repression by EN

The action of purifying selection on the ehIFRPF motif predicts that it is important for *D. melanogaster* EN function. We tested this hypothesis by ectopically expressing wild-type EN⁺ and a version of EN with a mutated ehIFRPF (EN^{IFRPF-}) using a previously described *paired* (*prd*)-*GAL4* assay for the repression of *wingless* (*wg*) in *D. melanogaster* embryos (Kobayashi et al. 2003) (with modifications to allow for more precise quantification; see MATERIALS AND METHODS for details). In this assay, embryos expressing EN⁺ efficiently repressed alternating stripes of *wg* where *en* is driven by the *prd* regulatory elements (Fig. 2A). At the same stage and level of protein expression, EN^{IFRPF-} repressed *wg* expression in embryos only about half as well (Fig. 2B). This decreased repression efficiency was significant ($P < 10^{-4}$), even though there was substantial embryo-to-embryo variability and the effect was limited to the narrow timepoint when WG first ceases to be present in regions ectopically expressing EN⁺ (Stage 10). By comparison, a version of EN with a mutated eh1 motif (EN^{eh1-}) and a version of EN with both motifs mutated (EN^{eh1- ehIFRPF-}) displayed more limited repression ability in this assay, even when expressed at higher levels than EN⁺ or EN^{ehIFRPF-} (Fig. 2, C and D). Thus, the ehIFRPF motif contributes to the ability of EN to repress the transcription of target genes.

The ehIFRPF motif physically interacts with GRO

In order to determine whether the ehIFRPF motif performs its role in repression by the predicted interaction with the co-repressor GRO, we tested the ability of the same panel of mutant EN proteins to interact with GST-GRO and M2-GRO *in vitro* in pulldown assays. EN^{IFRPF-} and EN^{eh1-} both displayed strong reductions in their ability to interact with GRO ($P < 10^{-4}$; Fig. 3), suggesting that both motifs contribute equally to the interaction of EN with GRO, at least in the *in vitro* context tested here. EN^{eh1- ehIFRPF-} displayed an additional slight reduction in its GRO-binding affinity, suggesting each motif was able to interact with GRO to some extent on its own ($P = 0.014$; Fig. 3). Interestingly, even EN^{eh1- ehIFRPF-} still interacted fairly well with GST-GRO, well above the level of GST controls with excess quantities of GST (data not shown). Previous work with truncated and mutated versions of EN have also failed to obliterate the EN-GRO interaction (Jimenez et al. 1997; Tolkunova et al. 1998), suggesting that the interaction may be distributed across additional motifs or interacting residues throughout EN in addition to the discrete eh1 and ehIFRPF motifs.

DISCUSSION

By searching selector proteins for synapomorphic peptide motifs that might contribute to their regulatory activity, we found a second GRO-interaction motif in the EN transcription factor. This ehIFRPF motif is insect-specific and has been under purifying selection in the dipteran/lepidopteran lineage. We have shown that the ehIFRPF motif contributes to interaction with GRO *in vitro* and repression of a target gene *in vivo*. The role of this second GRO-interaction motif in *D. melanogaster* EN suggests that it has been retained during evolution to modulate the strength of the EN-GRO interaction and EN repression activity. These observations provide a clear example of a transcription factor peptide motif evolving to modulate a pre-existing function.

EN contains at least two distinct GRO-interaction motifs

Our results demonstrate that EN contains two distinct GRO-interaction motifs that participate in repression (one of each known type). Given the extensive previous characterization of EN, it is somewhat surprising that the ehIFRPF motif remained uncharacterized. However, closer examination of previous studies reveals that all truncated versions of EN and chimeric proteins that grafted the eh1 motif onto heterologous proteins to confer repression ability also contained the ehIFRPF motif because of their close proximity in the linear sequence (Jimenez et al. 1997; Tolkunova et al. 1998). Moreover, WRPW-type motifs have typically been found toward the C-terminus of proteins and their consensus definition was only extended to include motifs that lack tryptophan residues after the eh1 motif had been characterized (Kobayashi et al. 2001). This study highlights the capacity for phylogenetic footprinting and evolutionary analysis to inform more detailed functional analysis of even well-characterized proteins.

Role of the ehIFRPF motif

The sites and conformations of the protein-protein interactions between GRO and both types of GRO-interaction motifs have recently been solved (Jennings et al. 2006). Both interactions depend heavily on hydrophobic interactions at the same general location in the GRO-tetramer but adopt different conformations. This model suggests that simultaneous interaction with both types of motifs by a single GRO-tetramer is probably impossible because both motifs would be required to occupy some of the same space.

How, then, can two motifs possibly promote a seemingly stronger interaction than one, as our data suggest? The simplest explanation is that each motif may recruit its own GRO-tetramer *in vivo*. Alternatively, multiple motifs may raise the effective local concentration of GRO by providing alternate interaction faces that limit diffusion away from EN when it is bound to DNA. Under this model, possession of multiple interaction sites by EN may increase the likelihood that a GRO-tetramer will interact with another site. More provocatively, the eh1 and ehIFRPF peptide motifs might also bind to other regions of GRO in some *in vivo* contexts. Indeed, the apparent discrepancy between the relative contributions of the eh1 and ehIFRPF motifs to the *in vitro* interaction with GRO and to *wg* repression suggests that the eh1 motif may have a stronger effect in some cellular contexts, perhaps due to the presence of DNA or additional co-factors not present in our *in vitro* assay.

Toward a non-adaptive model of peptide motif turnover

The appearance, retention, and refinement of novel peptide motifs in transcription factors are governed by a delicate interplay between the evolutionary forces of selection and drift. It may seem tempting to argue that novel peptide motifs must impart novel functions, but this is not necessarily the case. The evolution of a peptide motif that contributes quantitatively to

the function of a transcription factor alters the fitness landscape of all peptide motifs performing similar functions. Some previously deleterious mutations may become effectively neutral and, through an evolutionary process analogous to models of transcription factor binding site turnover (Ludwig et al. 2000) or models of degeneration and complementation following gene duplication (Force et al. 1999), two motifs may ultimately perform the tasks previously performed by a single motif. The emergence of a new peptide motif would thus result in the molecular evolution of sequence without the functional evolution of the network or its phenotypic readout. In the specific case here, the interaction between EN and GRO via the eh1 motif is conserved in all bilaterians, and the role of EN in patterning the posterior segments is conserved among arthropods (Davis and Patel 1999). Thus, it appears that the ehIFRPF motif evolved in a subset of insects without conferring any new biochemical or developmental function.

The degeneracy of WRPW-type and similar peptide motifs suggests that they can easily appear in random protein sequences and/or be refined via cumulative selection on weak, non-consensus motifs. The consensus $\psi\Omega(K/R)P\Omega$ is probably too strict because the strong WRPW motif of *D. melanogaster* HAIRY does not have an N-terminal aliphatic residue, and the GFKPY motif of *T. castaneum* EN may even be functional. Thus, it seems reasonable to assume that many tetrapeptide $\Omega(K/R)P\Omega$ motifs may provide weak GRO-interaction. How common are such motifs? They should be very common. If one built a 400-residue protein randomly from the genetic code, such $\Omega(K/R)P\Omega$ motifs would exist in one in 44 proteins in the absence of selection $\{1/[(5/61)(8/61)(4/61)(5/61)]*(400 - 3)\} = 44\}$. Once a motif has any quantitative contribution to overall function that is strong enough to respond to selection, the motif could be refined gradually via cumulative selection to create a higher (or lower) affinity motif. With the steady turnover of protein sequence during evolution, most proteins have had many opportunities to evolve any particular short, degenerate peptide motif many times during their history.

The insect-specific ehIFRPF GRO-interaction motif evolved in a transcription factor that already possessed at least one GRO-interaction motif that has been maintained for greater than 500 million years. The acquisition of this second motif provides a clear example of a peptide motif that has been acquired in a selector protein, not to evolve a novel function, but to modulate an existing function. The effects of peptide motifs with quantitative functions are difficult to detect and measure, but they may be common in transcription factors (Galant et al. 2002; Gebelein et al. 2002; Ronshaugen et al. 2002; Merabet et al. 2003; Hittinger et al. 2005; Lohr and Pick 2005; Tour et al. 2005; Merabet et al. 2007). Even our double mutant of both the eh1 and ehIFRPF motifs failed to completely obliterate the EN-GRO interaction, so we suspect that there are still other weak interaction motifs to be discovered. Based on these findings, we anticipate that many evolutionarily novel peptide motifs in transcription factors quantitatively contribute to previously existing functions.

MATERIALS AND METHODS

Sequence collection and analysis

When complete or draft genome sequences covering this portion of *en* were available, these sequences were used: *D. melanogaster* (Adams et al. 2000), *Drosophila pseudoobscura* (Richards et al. 2005), ten other *Drosophila* spp. (Clark et al. 2007), *Anopheles gambiae* (Holt et al. 2002), *Aedes aegypti* (Nene et al. 2007), *Bombyx mori* (Xia et al. 2004), *A. mellifera* (Honeybee Genome Sequencing Consortium 2006), and *T. castaneum* (Richards et al. 2008) from class Insecta. When not available, individually published sequences were used: *Junonia coenia* (Keys et al. 1999), *P. americana* (Marie and Bacon 2000), *S. gregaria* (Peel et al. 2006) from class Insecta; *Artemia franciscana* (Manzanares et al. 1993) and *Sacculina carcini* (Gibert et al. 2000) from class Crustacea, *Strigamia maritima* (Kettle et al.

2003) from class Myriapoda, *Parasteatoda tepidariorum* (syn. *Achaearanea tepidariorum*) (Akiyama-Oda and Oda, GenBank AB125741) and *Cupiennius salei* (Damen et al. 1998; Damen 2002) from class Chelicerata. A previous phylogenetic study of many of these sequences identified new insect EN/INV homologs, suggested widespread lineage-specific gene conversion, and reassigned some homologs as EN or INV orthologs (Peel et al. 2006). All EN/INV sequences were searched for any sequence matching the $\psi\Omega(K/R)P\Omega$ consensus of the ehIFRPF motif between their eh1 and eh2 motifs. However, since no INV orthologs contained an ehIFRPF motif, only EN orthologs (Peel et al. 2006) are shown in Figure 1 within the insect lineage.

Purifying selection test

The 15-bp DNA sequence encoding the ehIFRPF motif was collected from all available dipteran and lepidopteran EN orthologs, including all 12 available *Drosophila* genomes, except *Drosophila willistoni* whose EN sequence could not be located at the time of analysis (February 2006). With the aid of PAUP* version 4.0b10 (Swofford 2002) and the known phylogenetic tree (Clark et al. 2007), we inferred the ancestral sequence at each node by maximum parsimony. Since the tree is known and the sequences are closely related, there was generally little ambiguity for these ancestral sequence inferences. For one position (bp 12), two ancestral dipteran sequences were equally parsimonious (C or G). Both inferences were analyzed separately below and gave identical numerical results. To simplify calculations, identical sequences were excluded from further analysis, leaving seven distinct modern *Drosophila* sequences, two mosquito (family: Culicidae) sequences, two lepidopteran sequences, and nine inferred ancestral sequences (Fig. S1).

Inverting a conceptually similar and widely used test for positive selection that was designed for sequences with limited variability but aligned over many taxa (Zhang et al. 1997; Nei and Kumar 2000), we calculated N (non-synonymous sites), S (synonymous sites), n (non-synonymous substitutions), and s (synonymous substitutions) along each branch using the Modified Nei-Gojobori (Nei and Gojobori 1986; Nei and Kumar 2000) methods as implemented by MEGA version 3.0 (Kumar et al. 2004) to make pairwise comparisons between adjacent inferred ancestral and modern sequences or between adjacent inferred ancestral sequences. These values were summed across the relevant branches of the tree, and Fisher's Exact Test was used to test for a fewer than expected (under neutrality) number of non-synonymous substitutions.

Construction of mutants

The eh1 and ehIFRPF motifs were mutagenized and cloned into full-length EN cDNAs by PCR (singly and in combination) using an EN cDNA plasmid template (Poole et al. 1985). The mutant version of the eh1 motif had a peptide sequence of LAASASAAASDRAA (instead of LAFSISNILSDRFG) encoded by 5'-CTGGCCGCTTCCGCCTCCGCCGCGCGAGCGATCGTGCCGCA-3' (mutations are underlined). The mutant version of the ehIFRPF motif had a peptide sequence of IAAAA (instead of IFRPF) encoded by 5'-ATAGCCGCGCCGCCG-3'. The complete coding sequences and cloning junctions for all constructs were verified by DNA sequencing.

Immunohistochemistry

Embryos were prepared, stained, and destained as described previously (Carroll and Scott 1985). EN was detected using a 1:500 dilution of a polyclonal antibody raised in rabbits and affinity-purified using the N-terminal 150 residues of EN [C. H. Girdham and P. H. O'Farrell, unpublished but previously used reagent (Teodoro and O'Farrell 2003)]. Thus, this antibody was purified using a region of EN that excludes the eh1 and ehIFRPF motifs and should interact equivalently with all mutant proteins. EN was visualized by confocal

microscopy with a donkey anti-rabbit secondary antibody conjugated to rhodamine red-X (Jackson ImmunoResearch Laboratories, West Grove, PA). WG was detected using the monoclonal mouse anti-WG antibody 4D4 (Brook and Cohen 1996) obtained from the Developmental Studies Hybridoma Bank at the University of Iowa (Iowa City, IA) and visualized by confocal microscopy with a donkey anti-mouse secondary antibody conjugated to Cy5 (embryos for quantification) or goat anti-mouse secondary antibody conjugated to FITC (embryos in Fig. 2).

WG repression assay

Wild-type and mutated versions of EN cDNAs were cloned into the *EcoRI* and *XbaI* sites of pUAST (Brand and Perrimon 1993) with a 5'-GGCGCC-3' Kozak sequence added immediately upstream of the start site. A panel of about 10 independent transformant lines was created from each pUAST-*en* construct. *UAS-en* lines were crossed to a *prd-Gal4* line (Brand and Perrimon 1993) and assayed for EN and WG expression as described above and as previously used in an assay for EN repression ability (Kobayashi et al. 2003). Preliminary experiments indicated that substantial defects in WG repression ability were observed in embryos ectopically expressing EN^{eh1-} and EN^{eh1- ehIFRPF-} (Fig. 2). However, quantitative differences in WG repression between embryos expressing EN⁺ and EN^{IFRPF-} were difficult to resolve due to wide variation in the levels of ectopic EN expression, which led to varying degrees of WG repression.

Two *UAS-en*⁺ lines and one *UAS-en*^{IFRPF-} line were selected for detailed quantitative analysis of WG repression. These lines were selected because their EN ectopic expression levels were similarly moderate (2.22 ± 0.33 EN⁺ expression index $< 2.57 \pm 0.51$ EN^{IFRPF-} expression index; $P < 10^{-2}$; which is conservative for our conclusion that EN^{IFRPF-} has diminished WG repression ability) but sufficient for EN⁺ to cause robust repression. Stage 10 embryos ectopically expressing EN in the *prd* expression domain (even-numbered stripes) were selected in double-blind fashion, and confocal Z-sections were collected for both the EN and WG channel. Earlier stages produced more sporadic WG repression, presumably because EN had insufficient time to repress WG, while WG was generally completely repressed at later stages. These Z-sections were combined by maximum intensity projection with IMAGEJ version 1.33u (Rasband WS, <http://rsb.info.nih.gov/ij/>). EN and WG expression values for each stripe were calculated for stripes three through 13 by measuring the maximum intensity of each stripe and subtracting the average of the background signals from the interstripes on either side (except for stripe three where only the background signal from posterior to the stripe was recorded). EN and WG expression values for the even-numbered stripes (*i. e.* the *prd* ectopic expression domain) four through 12 were divided by the average of the adjacent expression values for the odd-numbered stripes (*i. e.* no ectopic expression). These five expression values from each stripe for both EN and WG were then averaged to calculate the EN and WG expression indices for each embryo. The EN expression index was treated as a measure of ectopic EN expression, while the WG expression index was treated as a measure of repression ability.

in vitro pulldown assays

Wild-type and mutated versions of EN cDNAs were cloned into the *NcoI* and *EcoRI* sites of pT7βplink (Dalton and Treisman 1992). EN was produced and labeled with [³⁵S]-methionine using these vectors as templates and the TNT® Quick Coupled Transcription/Translation system according to the manufacturers instructions (Promega, Madison, WI). Epitope-tagged GST-GRO and M2-GRO were expressed and isolated as described previously (Dubnicoff et al. 1997). Pulldown assays were performed as described previously (Zhang et al. 1996). For a given experiment, GST-GRO or M2-GRO was isolated in a single batch of glutathione or anti-M2 beads (Sigma-Aldrich, St. Louis, MO), respectively, which

then was evenly divided between TNT® products. After parallel incubation and washes, bound TNT® products were resolved by SDS/polyacrylamide gel electrophoresis, and the EN band was quantified using a phosphorimager. Raw intensities were normalized for slight differences in TNT® product input and normalized to the EN⁺ signal.

Statistics

All statistical tests were performed using MSTAT version 4.01 (Drinkwater N, <http://mcardle.oncology.wisc.edu/mstat/>) and are reported as one-tailed *P* values. The WG repression assay deployed a Wilcoxon rank sum test. Pulldown assays deployed Lehman's test, a multiple experiment permutation test derived from the Wilcoxon rank sum test.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Craig Nelson, Vicky Kassner, Diccon Fiore, and Carroll Lab members for technical advice; Antonis Rokas for assistance with the double-blind experiment; Albert Courey for GST-GRO and M2-GRO constructs; Charles Girdham and Pat O'Farrell for anti-EN; the Developmental Studies Hybridoma Bank for anti-WG (4D4); Tom Kornberg for the *en* cDNA template; and all those involved in genome sequencing projects that made their data publicly available. This research was supported by a Howard Hughes Medical Institute Investigatorship (SBC) and a HHMI Predoctoral Fellowship (CTH). Additional support to write the manuscript was provided to CTH by a Washington University in St. Louis Department of Genetics Fellowship and a NIH-NHGRI Postdoctoral NRSA Traineeship (5T32HG00045).

REFERENCES

- Aasland R, Abrams C, Ampe C, Ball LJ, Bedford MT, Cesareni G, Gimona M, Hurley JH, Jarchau T, Lehto VP, et al. Normalization of nomenclature for peptide motifs as ligands of modular protein domains. *FEBS Lett.* 2002; 513:141–4. [PubMed: 11911894]
- Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, et al. The genome sequence of *Drosophila melanogaster*. *Science.* 2000; 287:2185–95. [PubMed: 10731132]
- Alonso CR, Maxton-Kuechenmeister J, Akam M. Evolution of Ftz protein function in insects. *Curr Biol.* 2001; 11:1473–8. [PubMed: 11566109]
- Brand AH, Perrimon N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development.* 1993; 118:401–15. [PubMed: 8223268]
- Brook WJ, Cohen SM. Antagonistic interactions between wingless and decapentaplegic responsible for dorsal-ventral pattern in the *Drosophila* Leg. *Science.* 1996; 273:1373–7. [PubMed: 8703069]
- Carroll SB. Evolution at two levels: on genes and form. *PLoS Biol.* 2005; 3:e245. [PubMed: 16000021]
- Carroll SB, Grenier, JK.; Weatherbee, SD. *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design.* 2nd ed.. Blackwell Science; Malden, MA: 2005.
- Carroll SB, Scott MP. Localization of the fushi tarazu protein during *Drosophila* embryogenesis. *Cell.* 1985; 43:47–57. [PubMed: 3000605]
- Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC, Kellis M, Gelbart W, Iyer VN, et al. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature.* 2007; 450:203–18. [PubMed: 17994087]
- Courey AJ, Jia S. Transcriptional repression: the long and the short of it. *Genes Dev.* 2001; 15:2786–96. [PubMed: 11691830]
- Dalton S, Treisman R. Characterization of SAP-1, a protein recruited by serum response factor to the c-fos serum response element. *Cell.* 1992; 68:597–612. [PubMed: 1339307]

- Damen WG. Parasegmental organization of the spider embryo implies that the parasegment is an evolutionary conserved entity in arthropod embryogenesis. *Development*. 2002; 129:1239–50. [PubMed: 11874919]
- Damen WG, Hausdorf M, Seyfarth EA, Tautz D. A conserved mode of head segmentation in arthropods revealed by the expression pattern of Hox genes in a spider. *Proc Natl Acad Sci U S A*. 1998; 95:10665–70. [PubMed: 9724761]
- Davidson, EH. *Genomic Regulatory Systems: Development and Evolution*. Academic Press; San Diego, CA: 2001.
- Davis GK, Patel NH. The origin and evolution of segmentation. *Trends Cell Biol*. 2000; 9:M68–72. [PubMed: 10611687]
- Dubnicoff T, Valentine SA, Chen G, Shi T, Lengyel JA, Paroush Z, Courey AJ. Conversion of dorsal from an activator to a repressor by the global corepressor Groucho. *Genes Dev*. 1997; 11:2952–7. [PubMed: 9367978]
- Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*. 1999; 151:1531–45. [PubMed: 10101175]
- Galant R, Carroll SB. Evolution of a transcriptional repression domain in an insect Hox protein. *Nature*. 2002; 415:910–3. [PubMed: 11859369]
- Galant R, Walsh CM, Carroll SB. Hox repression of a target gene: extradenticle-independent, additive action through multiple monomer binding sites. *Development*. 2002; 129:3115–26. [PubMed: 12070087]
- Gebelein B, Culi J, Ryoo HD, Zhang W, Mann RS. Specificity of Distalless repression and limb primordia development by abdominal Hox proteins. *Dev Cell*. 2002; 3:487–98. [PubMed: 12408801]
- Gibert JM, Mouchel-Vielh E, Queinnec E, Deutsch JS. Barnacle duplicate engrailed genes: divergent expression patterns and evidence for a vestigial abdomen. *Evol Dev*. 2000; 2:194–202. [PubMed: 11252562]
- Grenier JK, Carroll SB. Functional evolution of the Ultrabithorax protein. *Proc Natl Acad Sci U S A*. 2000; 97:704–9. [PubMed: 10639143]
- Hittinger CT, Stern DL, Carroll SB. Pleiotropic functions of a conserved insect-specific Hox peptide motif. *Development*. 2005; 132:5261–70. [PubMed: 16267091]
- Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, Wincker P, Clark AG, Ribeiro JM, Wides R, et al. The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science*. 2002; 298:129–49. [PubMed: 12364791]
- Honeybee Genome Sequencing Consortium. Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature*. 2006; 443:931–49. [PubMed: 17073008]
- Hsia CC, McGinnis W. Evolution of transcription factor function. *Curr Opin Genet Dev*. 2003; 13:199–206. [PubMed: 12672498]
- Hui CC, Matsuno K, Ueno K, Suzuki Y. Molecular characterization and silk gland expression of *Bombyx* engrailed and invected genes. *Proc Natl Acad Sci U S A*. 1992; 89:167–71. [PubMed: 1346065]
- Jennings BH, Pickles LM, Wainwright SM, Roe SM, Pearl LH, Ish-Horowicz D. Molecular recognition of transcriptional repressor motifs by the WD domain of the Groucho/TLE corepressor. *Mol Cell*. 2006; 22:645–55. [PubMed: 16762837]
- Jimenez G, Paroush Z, Ish-Horowicz D. Groucho acts as a corepressor for a subset of negative regulators, including Hairy and Engrailed. *Genes Dev*. 1997; 11:3072–82. [PubMed: 9367988]
- Kettle C, Johnstone J, Jowett T, Arthur H, Arthur W. The pattern of segment formation, as revealed by engrailed expression, in a centipede with a variable number of segments. *Evol Dev*. 2003; 5:198–207. [PubMed: 12622737]
- Keys DN, Lewis DL, Selegue JE, Pearson BJ, Goodrich LV, Johnson RL, Gates J, Scott MP, Carroll SB. Recruitment of a hedgehog regulatory circuit in butterfly eyespot evolution. *Science*. 1999; 283:532–4. [PubMed: 9915699]
- Kissinger CR, Liu BS, Martin-Blanco E, Kornberg TB, Pabo CO. Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions. *Cell*. 1990; 63:579–90. [PubMed: 1977522]

- Kobayashi M, Fujioka M, Tolkunova EN, Deka D, Abu-Shaar M, Mann RS, Jaynes JB. Engrailed cooperates with extradenticle and homothorax to repress target genes in *Drosophila*. *Development*. 2003; 130:741–51. [PubMed: 12506004]
- Kobayashi M, Goldstein RE, Fujioka M, Paroush Z, Jaynes JB. Groucho augments the repression of multiple Even skipped target genes in establishing parasegment boundaries. *Development*. 2001; 128:1805–15. [PubMed: 11311161]
- Kumar S, Tamura K, Nei M. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform*. 2004; 5:150–63. [PubMed: 15260895]
- Lamb RS, Irish VF. Functional divergence within the APETALA3/PISTILLATA floral homeotic gene lineages. *Proc Natl Acad Sci U S A*. 2003; 100:6558–63. [PubMed: 12746493]
- Levine M, Davidson EH. Gene regulatory networks for development. *Proc Natl Acad Sci U S A*. 2005; 102:4936–42. [PubMed: 15788537]
- Levine M, Tjian R. Transcription regulation and animal diversity. *Nature*. 2003; 424:147–51. [PubMed: 12853946]
- Logan C, Hanks MC, Noble-Topham S, Nallainathan D, Provart NJ, Joyner AL. Cloning and sequence comparison of the mouse, human, and chicken engrailed genes reveal potential functional domains and regulatory regions. *Dev Genet*. 1992; 13:345–58. [PubMed: 1363401]
- Lohr U, Pick L. Cofactor-interaction motifs and the cooption of a homeotic Hox protein into the segmentation pathway of *Drosophila melanogaster*. *Curr Biol*. 2005; 15:643–9. [PubMed: 15823536]
- Lohr U, Yussa M, Pick L. *Drosophila fushi tarazu*, a gene on the border of homeotic function. *Curr Biol*. 2001; 11:1403–12. [PubMed: 11566098]
- Ludwig MZ, Bergman C, Patel NH, Kreitman M. Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature*. 2000; 403:564–7. [PubMed: 10676967]
- Mann RS, Carroll SB. Molecular mechanisms of selector gene function and evolution. *Curr Opin Genet Dev*. 2002; 12:592–600. [PubMed: 12200165]
- Manzanares M, Marco R, Garesse R. Genomic organization and developmental pattern of expression of the engrailed gene from the brine shrimp *Artemia*. *Development*. 1993; 118:1209–19. [PubMed: 7903633]
- Marie B, Bacon JP. Two engrailed-related genes in the cockroach: cloning, phylogenetic analysis, expression and isolation of splice variants. *Dev Genes Evol*. 2000; 210:436–48. [PubMed: 11180849]
- Merabet S, Kambris Z, Capovilla M, Berenger H, Pradel J, Graba Y. The hexapeptide and linker regions of the AbdA Hox protein regulate its activating and repressive functions. *Dev Cell*. 2003; 4:761–8. [PubMed: 12737810]
- Merabet S, Saadaoui M, Sambrani N, Hudry B, Pradel J, Affolter M, Graba Y. A unique Extradenticle recruitment mode in the *Drosophila* Hox protein Ultrabithorax. *Proc Natl Acad Sci U S A*. 2007; 104:16946–51. [PubMed: 17942685]
- Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol*. 1986; 3:418–26. [PubMed: 3444411]
- Nei, M.; Kumar, S. *Molecular Evolution and Phylogenetics*. Oxford University Press; Oxford, UK: 2000.
- Nene V, Wortman JR, Lawson D, Haas B, Kodira C, Tu ZJ, Loftus B, Xi Z, Megy K, Grabherr M, et al. Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science*. 2007; 316:1718–23. [PubMed: 17510324]
- Orphanides G, Reinberg D. A unified theory of gene expression. *Cell*. 2002; 108:439–51. [PubMed: 11909516]
- Paroush Z, Finley RL Jr. Kidd T, Wainwright SM, Ingham PW, Brent R, Ish-Horowicz D. Groucho is required for *Drosophila* neurogenesis, segmentation, and sex determination and interacts directly with hairy-related bHLH proteins. *Cell*. 1994; 79:805–15. [PubMed: 8001118]
- Passner JM, Ryoo HD, Shen L, Mann RS, Aggarwal AK. Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex. *Nature*. 1999; 397:714–9. [PubMed: 10067897]
- Peel AD, Telford MJ, Akam M. The evolution of hexapod engrailed-family genes: evidence for conservation and concerted evolution. *Proc Biol Sci*. 2006; 273:1733–42. [PubMed: 16790405]

- Peltenburg LT, Murre C. Engrailed and Hox homeodomain proteins contain a related Pbx interaction motif that recognizes a common structure present in Pbx. *Embo J.* 1996; 15:3385–93. [PubMed: 8698039]
- Poole SJ, Kauvar LM, Drees B, Kornberg T. The engrailed locus of *Drosophila*: structural analysis of an embryonic transcript. *Cell.* 1985; 40:37–43. [PubMed: 3917855]
- Richards S, Gibbs RA, Weinstock GM, Brown SJ, Denell R, Beeman RW, Gibbs R, Beeman RW, Brown SJ, Bucher G, et al. The genome of the model beetle and pest *Tribolium castaneum*. *Nature.* 2008; 452:949–55. [PubMed: 18362917]
- Richards S, Liu Y, Bettencourt BR, Hradecky P, Letovsky S, Nielsen R, Thornton K, Hubisz MJ, Chen R, Meisel RP, et al. Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and cis-element evolution. *Genome Res.* 2005; 15:1–18. [PubMed: 15632085]
- Ronschaugen M, McGinnis N, McGinnis W. Hox protein mutation and macroevolution of the insect body plan. *Nature.* 2002; 415:914–7. [PubMed: 11859370]
- Shiga Y, Yasumoto R, Yamagata H, Hayashi S. Evolving role of Antennapedia protein in arthropod limb patterning. *Development.* 2002; 129:3555–61. [PubMed: 12117806]
- Simon AL, Stone EA, Sidow A. Inference of functional regions in proteins by quantification of evolutionary constraints. *Proc Natl Acad Sci U S A.* 2002; 99:2912–7. [PubMed: 11880638]
- Smith ST, Jaynes JB. A conserved region of engrailed, shared among all en-, gsc-, Nk1-, Nk2- and msh-class homeoproteins, mediates active transcriptional repression in vivo. *Development.* 1996; 122:3141–50. [PubMed: 8898227]
- Stern DL. Evolutionary developmental biology and the problem of variation. *Evolution Int J Org Evolution.* 2000; 54:1079–91.
- Swofford, DL. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer; Sunderland, MA: 2002. Version 4.0b10
- Teodoro RO, O'Farrell PH. Nitric oxide-induced suspended animation promotes survival during hypoxia. *Embo J.* 2003; 22:580–7. [PubMed: 12554658]
- Tolkunova EN, Fujioka M, Kobayashi M, Deka D, Jaynes JB. Two distinct types of repression domain in engrailed: one interacts with the groucho corepressor and is preferentially active on integrated target genes. *Mol Cell Biol.* 1998; 18:2804–14. [PubMed: 9566899]
- Tour E, Hittinger CT, McGinnis W. Evolutionarily conserved domains required for activation and repression functions of the *Drosophila* Hox protein Ultrabithorax. *Development.* 2005; 132:5271–81. [PubMed: 16284118]
- Wagner GP. The developmental genetics of homology. *Nat Rev Genet.* 2007; 8:473–9. [PubMed: 17486120]
- Wagner GP, Pyle AM. Tinkering with transcription factor proteins: the role of transcription factor adaptation in developmental evolution. *Novartis Found Symp.* 2007; 284:116–25. discussion 125–9, 158–63. [PubMed: 17710850]
- Wray GA. The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet.* 2007; 8:206–16. [PubMed: 17304246]
- Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, Romano LA. The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol.* 2003; 20:1377–419. [PubMed: 12777501]
- Xia Q, Zhou Z, Lu C, Cheng D, Dai F, Li B, Zhao P, Zha X, Cheng T, Chai C, et al. A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). *Science.* 2004; 306:1937–40. [PubMed: 15591204]
- Zhang H, Catron KM, Abate-Shen C. A role for the Msx-1 homeodomain in transcriptional regulation: residues in the N-terminal arm mediate TATA binding protein interaction and transcriptional repression. *Proc Natl Acad Sci U S A.* 1996; 93:1764–9. [PubMed: 8700832]
- Zhang J, Kumar S, Nei M. Small-sample tests of episodic adaptive evolution: a case study of primate lysozymes. *Mol Biol Evol.* 1997; 14:1335–8. [PubMed: 9402743]

Fig. 1.

The eh1FRPF motif is an evolutionarily novel, conserved insect-specific GRO-interaction motif. Location of the eh1FRPF motif relative to the previously characterized eh1 and eh2 sequence motifs, which physically interact with GRO and EXD, respectively. The *D. melanogaster* sequence of partially conserved peptide motifs between the eh1 and eh2 motifs are shown. ..., non-conserved (dipteran/lepidopteran clade) bases not shown. No function is known for the motif whose sequence in *D. melanogaster* is “LGSLCKAVSQIG”. Possible eh1FRPF motifs with one mismatch (lowercase) are also shown for *T. castaneum*, *P. americana*, and *S. gregaria*. No putative eh1FRPF motifs that match the consensus $\Psi\Omega(K/R)P\Omega$ (or contain a single mismatch) were found in any other EN/INV homologs (see MATERIALS AND METHODS for organisms and sequences searched).

Fig. 2. The $ehIFRPF$ motif contributes to *in vivo* repression of WG by ectopic expression of EN. A, WG expression in the presence of ectopic EN^+ (A', ectopic EN^+ expression; A'', merge with EN^+ expression in red, WG expression in green); B, ectopic expression of EN^{IFRPF-} ; C, ectopic expression of EN^{eh1-} ; D, ectopic expression of $EN^{eh1-}ehIFRPF-$. Note that ectopic EN^+ strongly repressed the even numbered WG stripes (A). However, WG levels in the even stripes were only about half as repressed when EN^{IFRPF-} was ectopically expressed, instead of EN^+ [B; 0.225 ± 0.088 WG expression index for $EN^{ehIFRPF-}$ ($N = 30$) versus 0.119 ± 0.092 for EN^+ ($N = 23$); $P < 10^{-4}$]. Note that the mutant EN proteins lacking the $eh1$ motif (C and D) had a more limited ability to repress WG, even at higher expression levels.

Fig. 3.

The ehIFRPF motif physically interacts with GRO *in vitro*. Lanes 1-4 show relative amounts of labeled proteins pulled down by GST-GRO. Lane 1, EN⁺; lane 2, EN^{eh1-}; lane 3, EN^{IFRPF-}; lane 4, EN^{eh1-ehIFRPF-}. M2-GRO produced from an insect cell line gave similar results, while interactions with excess quantities of GST were much weaker (data not shown). Note that removing either the eh1 motif or the ehIFRPF motif reduced the interaction to $49 \pm 10\%$ ($N = 11$) or $48 \pm 9\%$ ($N = 11$), respectively, of the full-strength EN⁺-GRO interaction (both comparisons, $P < 10^{-4}$). Removing both motifs resulted in a slight additional reduction to $40 \pm 7\%$ ($N = 8$; $P < 10^{-3}$ compared with EN⁺; $P = 0.014$ compared with either EN^{IFRPF-} and EN^{eh1-}).