

Groucho: making its Marx as a transcriptional co-repressor

SUSAN M. PARKHURST
susanp@fred.fhrc.org

DIVISION OF BASIC SCIENCES, FRED HUTCHINSON CANCER RESEARCH CENTER, 1100 FAIRVIEW AVENUE NORTH, SEATTLE, WA 98109, USA.

An emerging theme in transcriptional repression systems is the use of sequence-specific DNA-binding proteins that also contain protein domains required for repression. While many repressor domains have been proposed to interact directly with the general transcriptional machinery or with activators (reviewed in Refs 1, 2), a subset of repressors appears to act indirectly by recruiting proteins, referred to as co-repressors, that bring about repression. The *Drosophila* gene *groucho* encodes such a transcriptional co-repressor that is required for proper embryogenesis (see Refs 3–5 and references therein). *groucho* takes its name from the phenotype of an adult viable allele where extra bristles above each eye resembled bushy eyebrows (Fig. 1). *groucho* was originally identified on the basis of its physical proximity to, and genetic interactions with, members of the *Enhancer of split* [*E(spl)*]-complex. Although these associations implicated Groucho in Notch signalling, more recent data show that Groucho interacts with a wide range of transcriptional regulatory systems. These studies are providing new insight into mechanisms of transcriptional repression, as well

as the significance of transcriptional repression in development.

Hairy and Groucho

Groucho is a maternally contributed protein that interacts genetically with the Hairy/*E(spl)* family of basic-helix–loop–helix (bHLH) transcription factors⁶. This family includes Hairy, Deadpan and seven members of the *E(spl)*-complex that act in segmentation, sex-determination and neurogenesis, respectively^{7–10}. Whereas the Hairy/*E(spl)* proteins show restricted expression patterns, Groucho is ubiquitously expressed during early development⁶. Embryos lacking maternally contributed Groucho show disruptions in all processes that are affected by the different members of the Hairy/*E(spl)* repressor bHLH protein class^{3,11}.

Consistent with its genetic interactions and loss-of-function phenotype, Groucho was recently shown to interact physically with all members of the Hairy/*E(spl)* bHLH family³. The conserved C-terminal WRPW tetrapeptide motif present in all Hairy/*E(spl)* family members was shown to be necessary for interaction with Groucho/TLEs using biochemical,

two-hybrid, and genetic approaches³ [TLEs (transducin-like *Enhancer of split*) are mammalian homologs of Groucho]. The WRPW motif is required for Hairy to act as a repressor: mutations in this motif weaken or eliminate repressor activity (Ref. 12 and references therein), whereas replacing it with the transcriptional activation domain of VP16 results in rapid transcriptional induction of promoters containing Hairy-binding sites¹³. The WRPW motif was also shown to confer repressor activity to the heterologous Gal4p DNA-binding protein¹⁴. Thus, the WRPW motif alone is sufficient for the recruitment of Groucho, and for transcriptional repression in assays using either endogenous or heterologous promoters.

Together, these results have led to the prevailing view that Hairy functions as a promoter-bound repressor: an intact bHLH region is required for Hairy to bind to specific DNA sites where it then uses its WRPW domain to recruit the Groucho co-repressor protein^{13,14}. Groucho, once recruited, can repress basal and activated transcription¹⁴.

Multi-component negative transcriptional regulators

Groucho and its mammalian homologs (TLE1–4), share structural and functional properties with Tup1p, a general transcriptional co-repressor in yeast^{5,15,16}. Their overall domain structure is similar, including an N-terminal Gln-rich region required for protein dimerization, internal Ser/Thr/Pro-rich sequences, and C-terminal tandem WD ('transducin-like') repeats implicated in protein–protein recognition. Tup1p, like Groucho, is unable to bind DNA on its own. Tup1p can mediate repression by its interaction with different classes of sequence-specific DNA-binding repressor proteins, such as Mat α 2p and Mig1p (Refs 16, 17). Groucho's broad expression, its ability to repress basal and activated transcription, and the pleiotropic effects of *groucho* mutations suggest that Groucho is also likely to interact with multiple classes

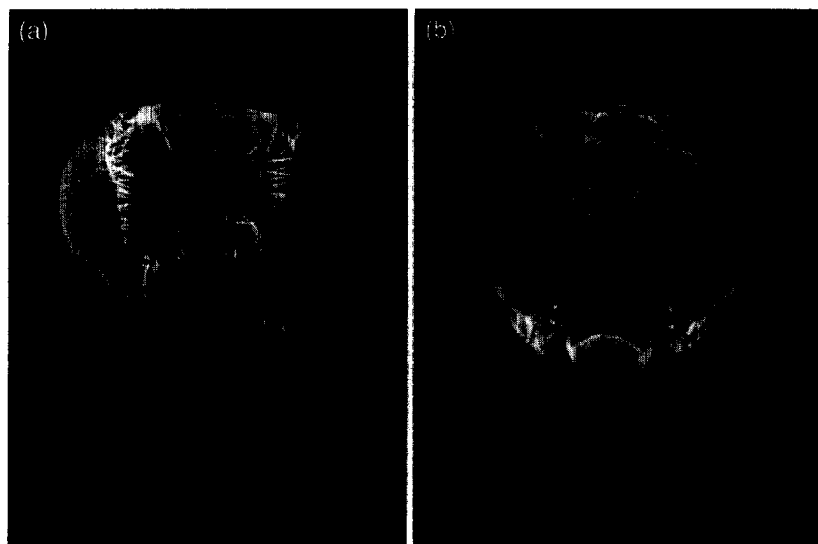


FIGURE 1. The *groucho* phenotype. Homozygous *groucho*¹ mutant flies (a) have extra bristles above each eye giving the appearance of bushy eyebrows compared with the wild type (b). In addition, the ocelli (simple eyes) present in the top center of the head are enlarged and not well separated.

of DNA-binding proteins that participate in transcriptional repression.

Groucho acts as a co-repressor for a subset of negative regulators

Several recent studies show that Groucho, like Tup1p, can indeed be recruited by a number of different transcriptional regulators. Once recruited, Groucho can even turn an activator into a repressor. A new family of DNA-binding proteins, referred to as Runt-domain proteins, include the products of the *Drosophila runt* and *lozenge* genes, as well as the mammalian PEBP2 α and AML1 genes, and are proposed to function as transcriptional activators¹⁸. The Runt domain is a 128 amino acid sequence necessary and sufficient for DNA binding. While the DNA-binding domain of Runt is unrelated to Hairy, Runt contains at its C-terminus a pentapeptide sequence reminiscent of Hairy's WRPW motif: VWRPY. The sequence similarity and C-terminal position of this motif led Aronson *et al.*¹⁸ to investigate whether Runt-domain proteins also interact with Groucho. Using biochemical, two-hybrid and genetic approaches, they demonstrated that Runt interacts directly with Groucho and that this interaction is dependent on Runt's C-terminal VWRPY motif¹⁸. While in most assays Runt has been shown to activate transcription, these studies show that in conjunction with Groucho, Runt now shows repressive properties.

The ability of Groucho to turn a transcriptional activator into a repressor has also been reported for the *Drosophila* Dorsal morphogen¹⁹. Dorsal is a REL family transcription factor that is required for dorsal-ventral patterning. Paroush *et al.*³ had reported disruptions to dorsal-ventral patterning in embryos lacking Groucho activity, suggesting a role for Groucho in this process. Using biochemical assays, Dubnicoff *et al.*¹⁹ have shown that Dorsal and Groucho interact physically, although Dorsal does not contain the WRP[W/Y] motif previously shown to interact with Groucho. While the Dorsal domain required for this interaction has not yet been reported, this interaction suggests that Groucho can be recruited by factors through unrelated protein domains.

When analyzing the phenotypes associated with loss of maternal Groucho, Paroush *et al.*³ also noticed

disruptions of segmentation gene expression patterns consistent with defects in the pathways required for patterning the termini of the *Drosophila* embryo, a process dependent on the Torso signalling pathway. By analyzing embryos lacking maternal *groucho* activity, Paroush *et al.*²⁰ showed that Groucho is required for proper terminal patterning. In particular, Groucho is required for repression of the terminal pathway target genes, *tailless* and *buckebain*, in regions outside the termini. In this case, the DNA-binding protein needed to recruit Groucho to repress *tailless* or *buckebain* has not yet been identified.

With the mounting evidence that transcriptional repressors function in concert with co-repressors, Jiménez *et al.*⁴ have taken a systematic approach to examine whether different known repressor activities require Groucho as a co-repressor. They substituted the WRPW domain in Hairy with several different known repression domains, and examined the effects of expressing these fusion proteins in embryos in the presence or absence of maternally contributed Groucho. Their results show that the repressor domain from the Engrailed homeodomain protein requires Groucho to function, whereas repressor domains from the Snail, Even-skipped, Krüppel and Knirps proteins act independently of Groucho. They went on to demonstrate that Engrailed interacts directly with Groucho, requiring the EH1 domain of Engrailed, which does not contain a WRP[W/Y] motif. Sequences within and outside Groucho's WD repeats are required for interaction with Hairy and Engrailed, but they show different requirements for the WD repeats⁴.

From these recent studies it is clear that Groucho is not a global co-repressor, but that it is required for a specific subset of repression systems. The motifs on the DNA-binding proteins required for association with Groucho can be unrelated structurally, as there are no significant similarities between Hairy, Dorsal and Engrailed, either within or outside their Groucho interaction domains. Several lines of evidence suggest that in repression systems where Groucho is recruited, it is required only in certain circumstances: Groucho appears to be dispensable for repression by Engrailed in imaginal discs (see Ref. 4 and references therein); Runt appears to

repress Engrailed independently of Groucho¹⁸; and in some developmental contexts, Hairy can function genetically as a repressor in the absence of the WRPW motif¹². Thus, while the WRP[W/Y] motif recruits Groucho, the presence of these motifs does not guarantee dependence on Groucho. Similarly, while proteins such as Engrailed and Even-skipped have structurally and functionally similar DNA-binding and repressor domains, only Engrailed shows Groucho-dependent repression⁴. The Groucho protein might also be modular. While the regions of Groucho required for interaction with the DNA-binding proteins have not yet been mapped in all cases, the regions of Groucho required for interaction with two of these, Hairy and Engrailed, are different⁴.

How does Groucho mediate repression?

Several models have been proposed for how co-repressors, such as Groucho, might mediate repression once recruited to target promoters by specific DNA-binding proteins (reviewed in Refs 1, 2). Groucho could repress transcription by interfering with interaction between activators and the general transcriptional machinery ('quenching'); by interacting (directly) with the general transcriptional machinery ('direct repression'); or by affecting chromatin organization. For the WD-containing protein Tup1p, each of these mechanisms has been implicated: blocking the activation domain of specific DNA binding proteins¹⁷; interaction with the general transcriptional machinery²¹; and organization of a repressive chromatin structure through direct interaction with the N-terminal regions of histones H3 and H4 (Ref. 22). While there are fewer data for Groucho, Palaparti *et al.*⁵ have shown that Groucho/TLEs also interact specifically with histone H3 in an N-terminal region shown genetically to be essential for H3 transcriptional silencing in yeast. It also remains a possibility that Groucho will participate both in quenching and in direct repression mechanisms at different developmental times, depending upon the presence and affinities of other interacting proteins recruited to the complexes.

Transcriptional repression clearly plays a major role in development. The studies described above are beginning to uncover the scope of protein

domains able to interact with one co-repressor, Groucho, as well as identifying which region(s) of Groucho is needed for each of these associations. A major challenge for the future will be to define the context-dependent activities of Groucho among different transcriptional repression systems.

References

- 1 Hanna-Rose, W. and Hansen, U. (1996) *Trends Genet.* 12, 229–234
- 2 Herschbach, B.M. and Johnson, A.D. (1993) *Annu. Rev. Cell Biol.* 9, 479–509
- 3 Paroush, Z. *et al.* (1994) *Cell* 79, 805–815
- 4 Jiménez, G., Paroush, Z. and Ish-Horowitz, D. (1997) *Genes Dev.* 11, 3072–3082
- 5 Palaparti, A., Baratz, A. and Stifani, S. (1997) *J. Biol. Chem.* 272, 26604–26610
- 6 Delidakis, C.A. *et al.* (1991) *Genetics* 129, 803–823
- 7 Delidakis, C.A. and Artavanis-Tsakonas, S. (1992) *Proc. Natl. Acad. Sci. U. S. A.* 89, 8731–8735
- 8 Knust, E. *et al.* (1992) *Genetics* 132, 505–518
- 9 Rushlow, C.A. *et al.* (1989) *EMBO J.* 8, 3095–3103
- 10 Younger-Shepherd, S. *et al.* (1992) *Cell* 70, 911–922
- 11 Schrons, H., Knust, E. and Campos-Ortega, J.A. (1992) *Genetics* 132, 481–503
- 12 Dawson, S.R. *et al.* (1995) *Mol. Cell. Biol.* 15, 6923–6931
- 13 Jiménez, G., Pinchin, S.M. and Ish-Horowitz, D. (1996) *EMBO J.* 15, 7088–7098
- 14 Fisher, A.L., Ohsako, S. and Caudy, M. (1996) *Mol. Cell. Biol.* 16, 2670–2677
- 15 Hartley, D.A., Priess, A. and Artavanis-Tsakonas, S. (1988) *Cell* 55, 785–795
- 16 Keleher, C.A. *et al.* (1992) *Cell* 68, 709–719
- 17 Treitel, M.A. and Carlson, M. (1995) *Proc. Natl. Acad. Sci. U. S. A.* 92, 3132–3136
- 18 Aronson, B.D. *et al.* (1997) *Mol. Cell. Biol.* 17, 5581–5587
- 19 Dubnicoff, T. *et al.* (1997) *Genes Dev.* 11, 2952–2957
- 20 Paroush, Z., Wainwright, S.M. and Ish-Horowitz, D. (1997) *Development* 124, 3827–3834
- 21 Redd, M.J., Stark, M.R. and Johnson, A.D. (1996) *Mol. Cell. Biol.* 16, 2865–2869
- 22 Edmondson, D.G., Smith, M. and Roth, S.Y. (1996) *Genes Dev.* 10, 1247–1259

Introns-early: slipping lately?

CLIVE N.A. TROTMAN

clive.trotman@stonebow.otago.ac.nz

DEPARTMENT OF BIOCHEMISTRY, UNIVERSITY OF OTAGO, DUNEDIN, NEW ZEALAND.

Introns are intervening sequences of noncoding DNA in the midst of a gene, often considerably longer than the coding sequences or exons. The mystery surrounding introns is how they originated – were they present in primitive coding sequences, or added later? Although introns are excised from primary transcript RNA (pre-mRNA) and are therefore untranslated, the notional location of the intron–exon boundary mapped on the protein might be the key to intron origin and function, and it is also convenient for describing the location of the intron. For example, haemoglobin genes commonly (but not always) have an intron at position B12-2 (B helix of the protein, 12th amino acid, following the 2nd nucleotide of the codon, or phase 2) and between G6 and G7, also a less common E helix intron (see Ref. 1 for a recent overview of globin introns; Fig. 1).

Early or late

The ‘introns-early’ (IE) view², essentially the exon theory of genes³, proposes that genes for some sophisticated proteins developed through the coalescence of primordial minigenes now represented by exons. Minigenes encoded modules of protein; traces of the noncoding linkers between

original minigenes survive as intron locations. The contrary ‘introns-late’ (IL) view⁴ is that fully developed genes had introns inserted into them later by a mechanism akin to reverse transcription. Because the debate revolves around how introns were acquired rather than when, ‘formative’ and ‘insertional’ might be better words than early or late. Different classes of intron are recognized⁵ and this discussion is limited to those removed by a spliceosome.

Introns are widespread in eukaryotic genes, many gene families offering statistically extensive datasets^{6,7}. Because introns are transmitted by inheritance, and some evidently have been inherited for hundreds of millions of years, it is unsurprising to find some consistency of intron location between homologous genes. However, unless divergence was recent the intron internal sequences will have largely mutated beyond recognition. The focus of attention is on intron positions that are subjectively related in different taxa but are, in fact, discordant in being anything from a single nucleotide to several codons discrepant; and also on positions where introns are present in some taxa but absent from others.

The IE interpretation of discordance is that an approximate alignment of introns in a family of genes could only reasonably be explained by perpetual inheritance, but that the picture has been corrupted by sporadic instances of slipping (sliding) and deletion of a mutational nature. In the globin example, the primordial gene would have had introns corresponding to loci in the B, E and G helices of the protein. These have occasionally slipped or have been deleted, particularly the E helix intron, which was predicted⁸ before discovery to satisfy a model of protein assembly from similar-sized modules.

The IL view is that base-wise slippage is implausible, partly because it is difficult to model the process without an intervening state of deranged reading frame that would surely inactivate the protein (many of the same difficulties might apply to alignment discordances between homologous proteins, which are not uncommon). Discordant location and spotty occurrence of introns are interpreted as indicative of independent events of insertion and loss, which have then been inherited. The mechanism that excises intron sequence from RNA could reverse and cause