6. Davis, P. *Int. J. Osteoarchaeol.* **7,** 388–401 (1997).
7. Behrensmeyer, A. K. in *Taphonomy: Releasing the Data Locked in the Fossil Record* (eds Allison, P. A. & Briggs, D. E. G.) 291–335 (Plenum, New York, 1991).
8. Fisher, D. C. *Paleobiology* **7,** 262–275 (1981).
9. Chin, K., Tokaryk, T. T., Erickson, G. M. & Calk, L. C. *Nature* **393,** 680–682 (1998).
10. Andrews, P. & Fernández-Jalvo, Y. *Nature* **393,** 629–630 (1998).

Molecular biology

Protein encoding by both DNA strands

All the evidence so far points to a gene's protein-coding information being contained in only one of its two DNA strands, with this strand serving as a template for transcription of the precursor RNA that is eventually translated into protein[1]. Here we present structural evidence showing that the protein-coding information of the *modifier of mdg4* (*mod(mdg4)*) gene of the fruitfly *Drosophila* is provided by both of its complementary DNA strands, and not by just one. This novel organization means that RNA precursors generated from two DNA templates need to be joined subsequently into a single messenger RNA, a surprising feature that raises new questions regarding genome complexity and evolution.

The *mod(mdg4)* proteins of *Drosophila* are required during the fly's development and probably help to establish or maintain chromatin structure[2–4]. The genomic sequence of this locus has been determined[5] and Fig. 1a shows the arrangement of the *mod(mdg4)* region in the sequence (accession number AE003734), whose principal transcript is 2.2 kilobases (kb) long. Exons I to IV of the 2.2-kb RNA are transcribed from nucleotide positions 153,934 to 151,614; exons V and VI are transcribed from nucleotide positions 136,942 to 137,841 by using the complementary strand as template (Fig. 1a).

To verify the arrangement of sequence AE003734, we used the polymerase chain reaction to amplify DNA from two unrelated wild-type fly strains and an array of nested primers located in coding and non-coding sequences. The amplified DNA fragments for all combinations of primers and for the two strains of fly were of the expected sizes (results not shown). We conclude that the 2.2-kb RNA appears to be encoded by two separate sequences of the gene transcribed in opposite orientations.

We confirmed that the 2.2-kb mRNA was present in fly cells by using a ribonuclease (RNase) protection assay and an RNA probe spanning exons IV and V (Fig. 1b), which revealed a fragment of 598 nucleotides, the size expected in the event of protection by a 2.2-kb mRNA. Other fragments predicted to result from protection by additional *mod(mdg4)* RNAs containing exons IV (435 nucleotides) and V (168 nucleotides) were also detected.

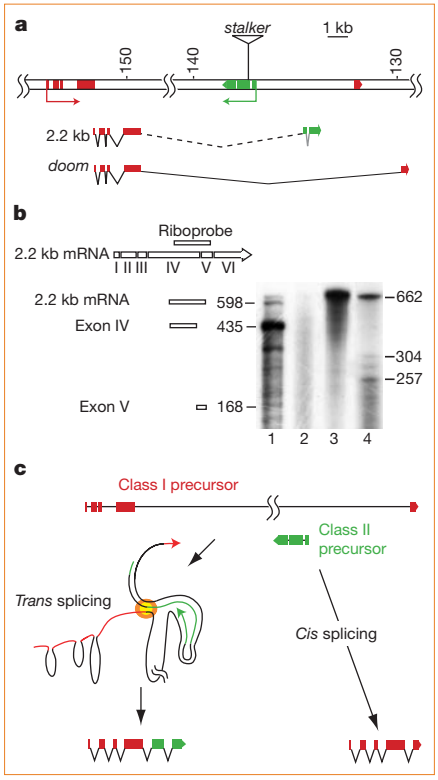

Figure 1 Organization of the *mod(mdg4)* locus. **a,** Top, genomic structure. Red and green boxes represent exons in different DNA strands. Arrows indicate the origin of transcription on each strand. A mutant, *mod(mdg4)*^{u1}, is shown here that contains an insertion of the *stalker* retrotransposon at position 137,206 (according to sequence AE003734). Bottom, alternative mRNAs produced by the *mod(mdg4)* locus. Lines connecting exons indicate *cis* splicing; dashed line connects exons transcribed in the opposite orientation. **b,** *Trans*-spliced 2.2-kb mRNA is present in *Drosophila* total RNA. A 598-nucleotide antisense RNA probe containing the last 435 nucleotides from exon IV and the first 163 from exon V of the 2.2-kb *mod(mdg4)* mRNA was synthesized by *in vitro* transcription with T7 polymerase in the presence of [^{32}P]CTP. The 598-nucleotide riboprobe used in the RNase protection assay is indicated. The 662-nucleotide fragment in lanes 3 and 4 corresponds to the riboprobe plus 64 extra nucleotides transcribed from the plasmid. Lane 1, 50 μg total *Drosophila* RNA; lanes 2 and 3, 50 μg yeast RNA; RNase was added only to lanes 1 and 2. Protected fragments of the predicted sizes are indicated on the left. On the basis of their size, additional bands are probably due to protection of other *mod(mdg4)* RNA species. Lane 4, size markers (nucleotides). **c,** RNA precursors and *cis* splicing or *trans* splicing at the *mod(mdg4)* locus (not to scale). Red and green boxes represent exons transcribed in opposite orientations. Pairing between both precursors facilitates *trans* splicing (circled) and the production of a functional mRNA. Further details are available from the authors.

The genomic organization of *mod(mdg4)* suggests that two RNA precursors are transcribed to yield the 2.2-kb mRNA. The RNA encoding the *doom* protein (Fig. 1a) is a candidate for one of these precursors[8]. It contains exons I to IV from the 2.2-kb mRNA and extends to nucleotide 131,723 ('class I' precursor in Fig. 1c). A second RNA precursor ('class II' precursor) containing exons V and VI probably originates at nucleotide 136,626 in a promoter region predicted by GENSCAN[9] (Fig. 1a, c).

Figure 1c shows a simplified model of *trans* splicing and *cis* splicing (donor and acceptor sites on the same molecule) of RNA precursors at the *mod(mdg4)* locus — two nested transcription units in opposite orientation produce two partially complementary RNAs that are able to form double-stranded RNA. It has been shown by using engineered DNAs that sequence complementarity between two RNA molecules can facilitate *trans* splicing in transfected cells[10].

Our results indicate that the evolutionary constraints imposed by a *cis* arrangement of the exons in a gene can be overcome by sequence complementarity between their encoded RNAs. Phylogenetic studies are needed to clarify the evolutionary process that shaped the genomic structure of *mod(mdg4)*. Genome sequences will reveal whether analogous structures are present in other genes and organisms.

Mariano Labrador, Fabien Mongelard, Piedad Plata-Rengifo, Ellen M. Baxter, Victor G. Corces, Tatiana I. Gerasimova
Department of Biology, The Johns Hopkins University, 3400 N Charles Street, Baltimore, Maryland 21218, USA
e-mail: corces@jhu.edu

1. Burge, C. B., Tuschl, T. & Sharp, P. A. in *The RNA World* 2nd edn (eds Gesteland, R. F., Cech, T. R. & Atkins, J. F.) 525–560 (Cold Spring Harbor Laboratory Press, New York, 1999).
2. Gerasimova, T. I., Gdula, D. A., Gerasimov, D. V., Simonova, O. & Corces, V. G. *Cell* **82**, 587–597 (1995).
3. Gerasimova, T. I. & Corces, V. G. *Cell* **92**, 511–521 (1998).
4. Buchner, K. *et al. Genetics* **155**, 141–157 (2000).
5. Adams, M. D. *et al. Science* **287**, 2185–2195 (2000).
6. Agabian, N. *Cell* **61**, 1157–1160 (1990).
7. Caudevilla, C. *et al. Proc. Natl Acad. Sci. USA* **95**, 12185–12190 (1998).
8. Harvey, A. J., Bidwai, A. P. & Miller, L. K. *Mol. Cell. Biol.* **17**, 2835–2843 (1997).
9. Burge, C. B. & Karlin, S. *Curr. Opin. Struct. Biol.* **8**, 346–354 (1998).
10. Solnick, D. *Cell* **42**, 157–164 (1985).