

Evolution of distinct developmental functions of three *Drosophila* genes by acquisition of different *cis*-regulatory regions

Xuelin Li & Markus Noll

Institute for Molecular Biology II, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

It is generally accepted that the specific function of a gene depends on its coding sequence. The three paired-box and homeobox genes *paired* (*prd*), *gooseberry* (*gsb*) and *gooseberry neuro* (*gsbn*) have distinct developmental functions in *Drosophila* embryogenesis^{1–5}. During the syncytial blastoderm stage, the pair-rule gene *prd*^{4,6} activates segment-polarity genes, such as *gsb*⁷, *wingless* (*wg*), and *engrailed* (*en*), in segmentally repeated stripes⁸. After germ-band extension, *gsb* maintains the expression of *wg*, which in turn specifies the denticle pattern by repressing a default state of ubiquitous denticle formation in the ventral epidermis⁹. In addition, *gsb* activates *gsbn*⁵, which is expressed mainly in the central nervous system^{2,3}, suggesting that *gsbn* is involved in neural development. Here we show that, despite the functional difference and the considerably diverged coding sequence of these genes, their proteins have conserved the same function. The finding that the essential difference between genes may reside in their *cis*-regulatory regions exemplifies an important evolutionary mechanism of how function diversifies after gene duplication.

The most conspicuous feature of the segmental organization of a *Drosophila* larva is its ventral denticle pattern (Fig. 1a). Recently, we have shown that *gsb* regulates this pattern through a *wg*-*gsb* autoregulatory loop that maintains the expression of *wg*, which represses denticle formation⁹. Thus, when *Hsgsb* embryos carrying a transgenic *gsb* gene under the control of the heat-inducible *hsp70* promoter were heat-shocked between 3 h 10 min and 6 h 20 min of development at 25 °C (early time interval), ubiquitous expression of *gsb* generated a naked larval cuticle (Fig. 1e). An earlier heat shock between 2 h 10 min and 3 h 10 min of development at 25 °C (early time interval), however, induced a pair-rule phenotype (Fig. 1b). This result is unexpected because it differs dramatically from the normal *gsb* gain-of-function phenotype (Fig. 1e). In wild-type embryos, *gsb* begins to be expressed only by the end of this early time interval, which coincides with the time of pair-rule gene rather than segment-polarity gene function. In fact, ubiquitous expression of pair-rule genes is known to result in pair-rule phenotypes that are nearly reciprocal to their loss-of-function phenotypes^{10–14}. Therefore, we suspected that activation of *Hsgsb* during the early time interval mimics the function of a ubiquitously expressed pair-rule protein. The most likely candidate was the Prd protein, as its N-terminal half consists of a paired-domain and a *prd*-

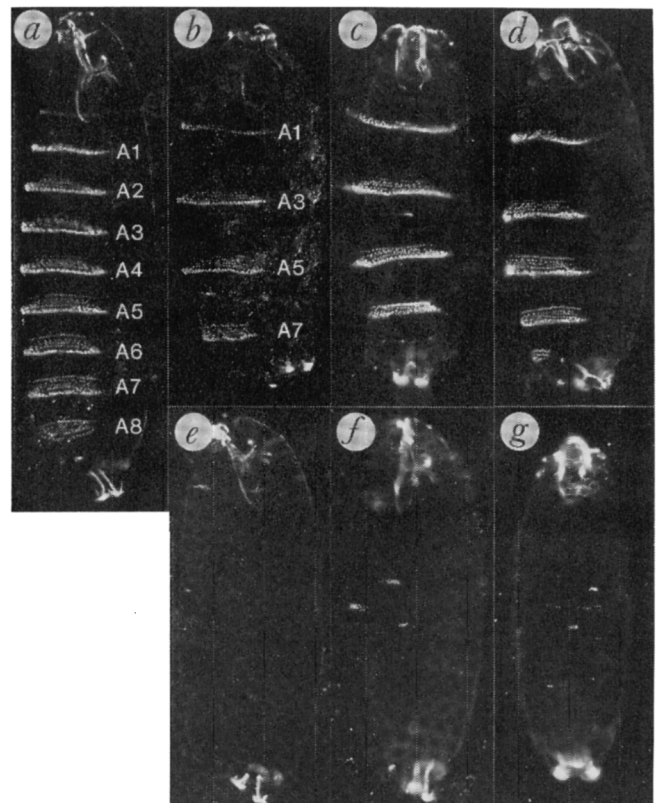


FIG. 1 Identical cuticular phenotypes induced by the ubiquitous expression of *gsb*, *prd* or *gsbn*. Cuticular preparations of wild-type (a), *Hsgsb* (b, e), *Hsprd* (c, f) and *Hsgsbn* (d, g) embryos heat-shocked during the early (a–d) or late (e–g) time interval are shown as ventral views under dark-field illumination. Ubiquitous activation of *Hsgsb*, *Hsprd* or *Hsgsbn* during the early time interval generates a pair-rule cuticular phenotype. In all cases, even-numbered abdominal denticle belts (A2, A4, A6, A8) and their anteriorly adjacent naked regions are lost, with the occasional exception of a few remaining denticles. This phenotype is nearly reciprocal to that of *prd*[–] embryos in which the odd-numbered denticle belts and their anteriorly neighbouring naked regions are deleted. Ubiquitous activation of *Hsgsb*, *Hsprd* and *Hsgsbn* by a heat shock during the late time interval produces a naked cuticular phenotype.

METHODS. Transgenic *Hsgsb*, *Hsprd* or *Hsgsbn* embryos, collected between 2 h 10 min and 3 h 10 min AEL (after egg laying) (early time interval) or between 3 h 10 min and 6 h 20 min AEL (late time interval), were heat-shocked for 15 min at 37 °C. After 24 h of development at 25 °C, cuticles were prepared essentially as described²⁸. Transgenic *Hsprd*, *Hsgsb* and *Hsgsbn* fly stocks were produced, as previously described²⁹, by cloning a *prd*-cDNA, c7340.4 (ref. 19), a *gsb*-cDNA, BSH9c2, or a *gsbn*-cDNA, BSH4c4 (ref. 3), into the P-element vector pKB255 (K. Basler and E. Hafen, unpublished) and subsequent germ-line transformation of *w¹¹¹⁸* embryos according to standard procedures³⁰.

type homeodomain and thus is highly homologous to the N-terminal half of Gsb². Indeed, early ubiquitous expression of *prd* in *Hsprd* embryos produced a phenotype¹² (Fig. 1c) indistinguishable from the *Hsgsb* pair-rule phenotype (Fig. 1b).

As *gsb* maintains the expression of *wg*⁹, we expect that the pair-rule phenotype of *Hsgsb* and *Hsprd* results from ectopic expression of the endogenous *gsb* and *wg* genes. Indeed, ubiquitous activation of either *gsb* or *prd* during the early time interval generates ectopic Gsb (Fig. 2b, c) and Wg stripes (Fig. 2f, g) anterior to the even-numbered wild-type Gsb (Fig. 2a) and corresponding Wg stripes (Fig. 2e). The observed pair-rule phenotype (Fig. 1b, c) is thus consistent with the ectopic *wg* expression and the resulting repression of denticle formation (Fig. 1b, c).

Activation of *Hsprd* has been shown to expand the odd-num-

bered *En* stripes posteriorly¹² (Fig. 2i, l). We now show that the same effect occurs in heat-shocked *Hsgsb* embryos (Fig. 2k). Note that the initial ectopic activation of *gsb*, *wg* and *en* by *Hsprd* or *Hsgsb* depends on the pair-rule function of *Prd*, which activates the segmental stripes of *Gsb*, *Wg* and *En*^{8,12}. In contrast, the subsequent maintenance of this ectopic expression of *gsb*, *wg* and *en* (Fig. 2a–m) is determined solely by their mutual activations and hence becomes independent of the product of the *Hsprd* or *Hsgsb* transgene.

Therefore, after heat shock within the early time interval, both *Hsgsb* and *Hsprd* embryos exhibit the same altered expression patterns of the endogenous *gsb*, *wg* and *en* genes and develop indistinguishable pair-rule phenotypes. We conclude that at this

time *Hsgsb* functions as *Hsprd*.

Just as *Gsb* can substitute for *Prd* during the early time interval, so might *Prd* replace *Gsb* function during the late time interval. Indeed, when we examined whether heat-induced *prd* expression during the late time interval repressed denticle formation in *Hsprd* embryos, we found that denticles are strongly repressed (Fig. 1f) as in the case of *Hsgsb* embryos treated similarly (Fig. 1e). Yet none of the embryos exhibit the pair-rule phenotype observed after heat shock during the early time interval.

Consistent with the altered phenotype observed after late activation of *Hsgsb* or *Hsprd*, expression of the endogenous *gsb*, *wg* and *en* genes shows a different response from that after early

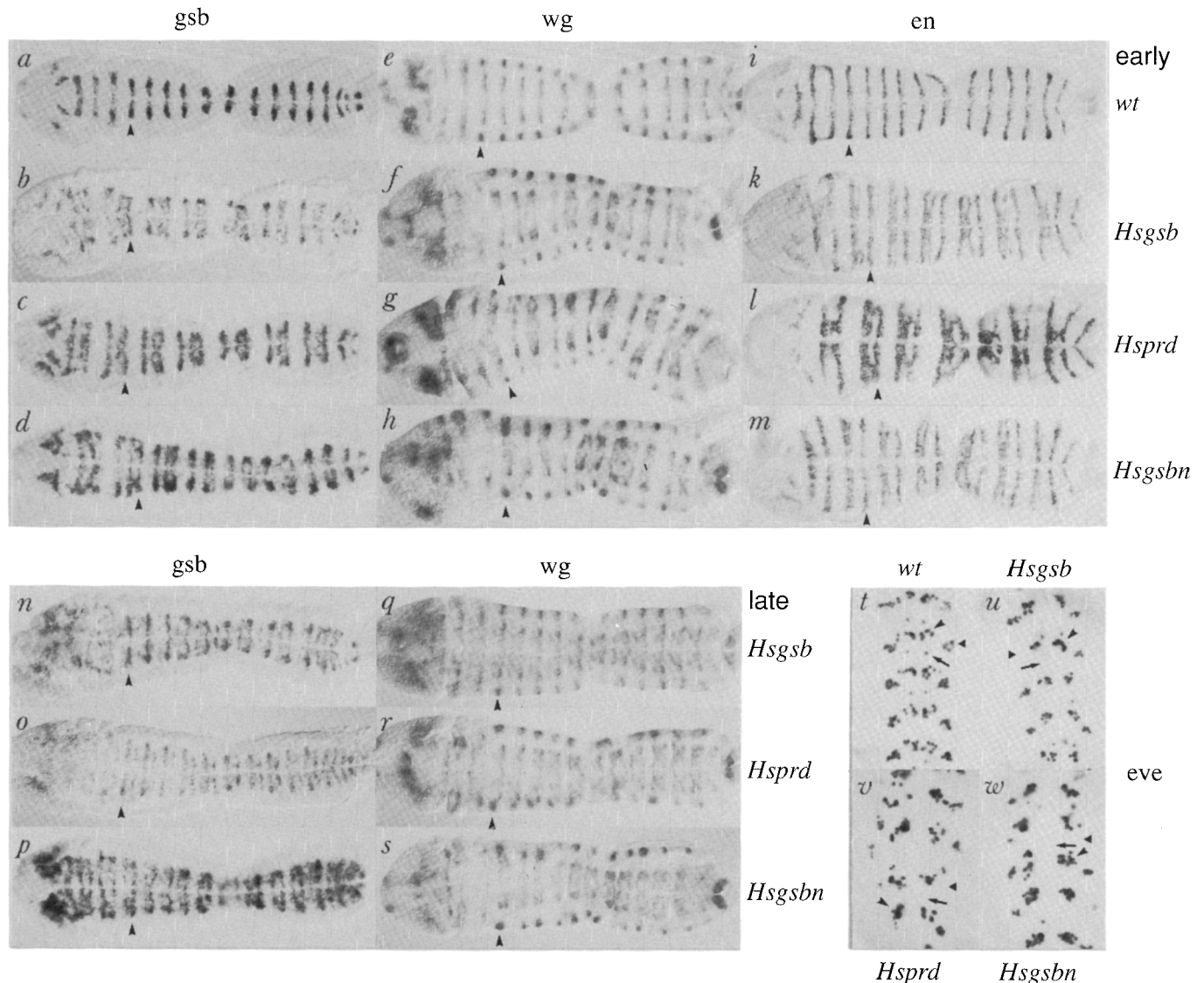


FIG. 2 Ubiquitous expression of *gsb*, *prd* or *gsbn* alters the expression of *gsb*, *wg*, *en* and *eve* in the same manner. a–s, Ectopic expression of *gsb*, *wg* and *en* induced by the ubiquitous expression of *gsb*, *prd* or *gsbn* during the early and late time interval. Expression patterns of *gsb* (a–d, n–p), *wg* (e–h, q–s) and *en* (i–m) are shown in wild-type (a, e, i), *Hsgsb* (b, f, k, n, q), *Hsprd* (c, g, l, o, r) and *Hsgsbn* (d, h, m, p, s) embryos 3–4 h after early (a–m) or late (n–s) heat-shock treatment. Embryos, oriented with their anterior to the left, are cut and unfolded along the amnioserosa to show the entire set of stripes. Arrowheads point to stripe 4 of *Gsb*, *Wg* or *En*. For convenience, the numbering of the *Wg* stripes follows that of the corresponding *Gsb* and *En* stripes^{3,31}. Thus, stripe 4 of *Wg* and *En* are adjacent to each other³² whereas stripe 4 of *gsb* overlaps with stripe 4 of both *Wg* and *En* (ref. 5). t–w, Change of *eve* expression in the CNS induced by the ubiquitous expression of

gsb, *prd* or *gsbn* during the late time interval. The expression in the CNS of *eve* is shown in wild-type (t), *Hsgsb* (u), *Hsprd* (v) and *Hsgsbn* (w) embryos 10 h after late heat shock. A trunk region of the CNS is shown with its anterior oriented up. Note the frequent loss of RP2 (arrows) and EL (triangles) neurons and the amplified CQ neurons (arrowheads) in *Hsgsb*, *Hsprd* or *Hsgsbn* compared with wild-type embryos.

METHODS. Transgenic *Hsgsb*, *Hsprd* or *Hsgsbn* embryos, collected between 2 h 30 min and 3 h 10 min AEL (a–m), between 3 h 40 min and 4 h 20 min AEL (n–s), or between 4 h 30 min and 6 h 30 min AEL (t–w) at 25 °C, were heat-shocked for 15 min at 37 °C and allowed to recover for 4 h (a–m), 3 h (n–s) or 10 h (t–w) at 25 °C before fixation and staining with anti-*Gsb*, anti-*Wg*, anti-*En* or anti-*Eve* antibodies as described⁸.

heat induction. Although heat shock within the late time interval induces an ectopic *gsb* and *wg* stripe anterior to each wild-type stripe in both *Hsgsb* (Fig. 2n, q) and *Hsprd* (Fig. 2o, r) embryos, ectopic *en* expression is no longer observed (not shown). As *gsb* does not affect *en* expression⁹, these observations imply that, during the late time interval, *prd* expressed ectopically carries out only the segment-polarity function of the ectopic Gsb protein.

The differences in cuticular phenotypes and expression of *gsb*, *wg* and *en* between the early and late activation of *Hsprd* or *Hsgsb* further suggest the existence of a dramatic transition from a 'pair-rule' stage to a 'segment-polarity' stage at about 3 h 10 min during *Drosophila* embryogenesis.

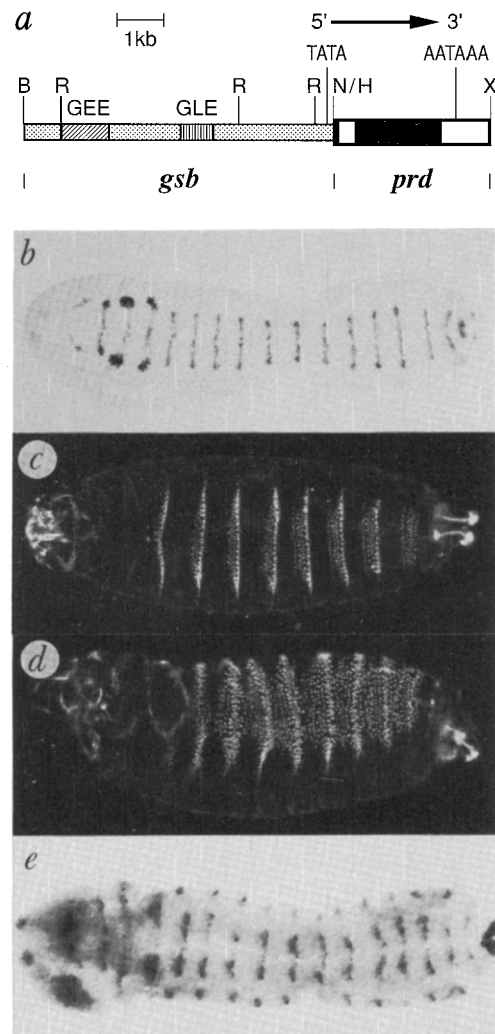
The *gsbn* gene encodes a protein whose N terminus, like that of the products of *gsb* and *prd*, consists of a paired-domain and a *prd*-type homeodomain but whose C terminus differs from those of the other two^{2,3}. Therefore, we investigated whether *Hsgsbn* could also substitute for *Hsgsb* or *Hsprd*, and found that indeed, heat-activated *Hsgsbn* produces the same pair-rule and naked cuticular phenotypes as *Hsgsb* and *Hsprd* (Fig. 1d, g). Moreover, activation of *Hsgsbn* during the early and late time interval again changes the expression patterns of *gsb*, *wg* and *en* in the same way as the similarly induced *Hsgsb* and *Hsprd* (Fig. 2a–s). Therefore Prd, Gsb and Gsbn can all replace each other with respect to these criteria.

To compare the effects of ubiquitous expression of *prd*, *gsb* or *gsbn* on central nervous system (CNS) development, *Hsprd*, *Hsgsb* or *Hsgsbn* embryos were heat-shocked during the late time interval and immunologically stained for the Even-skipped (Eve) protein. In addition to its role in segmentation, the pair-rule gene *eve* is expressed in and specifies the fate of certain neurons in the CNS^{15–17}. As expected, heat induction of *Hsprd*, *Hsgsb* or *Hsgsbn* disturbs the stereotype wild-type *eve* expression pattern in the CNS in a similar way (Fig. 2t–w). Particularly striking are the frequent loss of the RP2 and EL neurons and the increased number of the CQ neurons, suggesting that expression of *Hsprd*, *Hsgsb* and *Hsgsbn* interferes similarly with neural development.

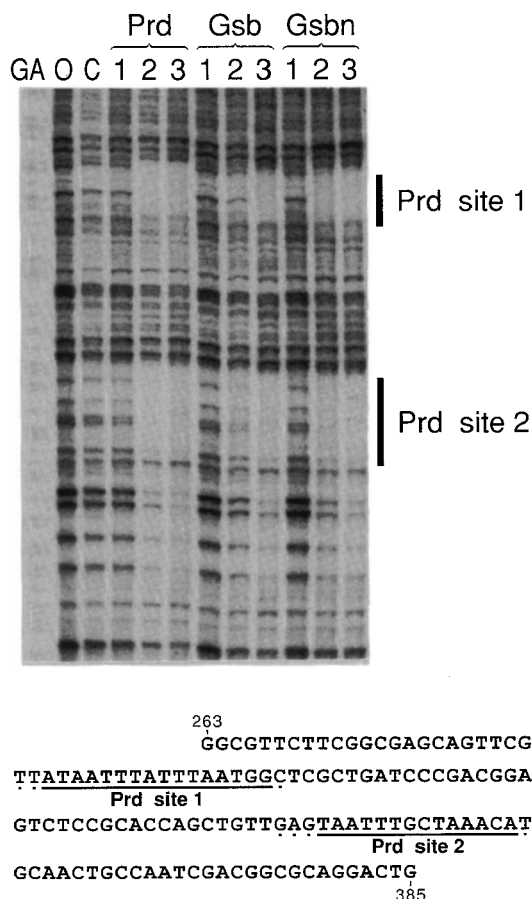
Although our results show that the Prd protein can replace Gsb in its ectopic function, it is important to test whether it can also substitute for normal Gsb function. Therefore, we attempted to rescue the cuticular phenotype of *gsb*[−] embryos by expressing a *prd* transgene in cells that normally express *gsb*. The *gsb* control region, including the *cis*-regulatory elements for the segmental stripes of *gsb*-ectodermal expression¹⁸, was fused to the *prd* gene comprising the entire coding region¹⁹ (Fig. 3a) and used to generate transgenic flies. As expected, embryos express this *gsb*-*prd* transgene in a *gsb*-like pattern (Fig. 3b). Moreover, the *gsb* cuticular phenotype of homozygous

FIG. 3 Rescue of the *gsb*[−] cuticular phenotype by Prd protein. **a**, Structure of the *gsb*-*prd* transgene. The 6.6-kilobase (kb) *Bam*HI–*Nru*I fragment of *gsb*, containing the upstream regulatory elements GEE and GLE, the TATA box and 104 base pairs (bp) of the *gsb* leader sequence¹⁸, was fused to the 3.2-kb *Hind*III–*Xba*I genomic fragment of *prd*, comprising 32 bp of the leader, the coding region, including the intron, and the entire trailer sequence¹⁹. GEE and GLE are the control elements responsible for the establishment and maintenance of *gsb* expression in the ectoderm¹⁸. The *prd* leader and coding region are shown as black boxes, and open boxes indicate the *prd* intron, the 3' trailer and part of the downstream region. Abbreviations of restriction sites: B, *Bam*HI; H, *Hind*III; N, *Nru*I; R, *Eco*RI; X, *Xba*I. **b**, Expression pattern of the *gsb*-*prd* transgene in embryos stained with anti-Prd antibodies. The embryo, oriented with its anterior to the left, is at stage 11 (5.5 h AEL) and has been unfolded to show the entire set of stripes. At this time, the endogenous *prd* protein is no longer detectable⁴. The stained Prd protein originates from the *gsb*-*prd* transgene and shows a similar pattern to that of the endogenous Gsb protein (see Fig. 2a). **c**, **d**, Rescue of the *gsb*[−] cuticular phenotype by the *gsb*-*prd* transgene. Ventral views of the cuticle preparations of homozygous *Df(2R)gsb*^{ix62} embryos with (c) or without (d) a *gsb*-*prd* transgene are shown under dark-field illumination. As the deficiency *Df(2R)gsb*^{ix62} deletes in addition to *gsb* and *gsbn* at least another lethal gene, *zipper*²⁰, lethality is not rescuable. Note that the head defects resulting from the *zipper* mutation are visible in both embryos (oriented with their anterior to the left). **e**, Activation of *wg* by the *gsb*-*prd* transgene in homozygous *Df(2R)gsb*^{ix62} embryos. Embryos have been stained with anti-Wg and anti-Gsbn antibodies to identify those embryos that fail to stain for Gsbn and thus are homozygous for *Df(2R)gsb*^{ix62}. The embryo shown, unfolded and oriented with its anterior to the left, is at the late stage 11 (7 h AEL). Note that at this stage *wg* expression has completely decayed in homozygous *Df(2R)gsb*^{ix62} embryos that carry no *gsb*-*prd* transgene^{9,33}.

METHODS. To prepare the *gsb*-*prd* rescue construct, the 1.5-kb *Nru*I–*Sma*I fragment (+104 bp to +1.6 kb) of the plasmid 9Z2', which contains the 8.1-kb *Bam*HI fragment of *gsb* (−6.5 kb to +1.6 kb) in pKSpL3 (ref. 18), was replaced by the blunt-ended 3.6-kb *Hind*III–*Xba*I fragment of *prd* from the genomic DNA clone D7.11 (ref. 6), generating the *gsb*-*prd* gene in pKSpL3. From this construct, the *gsb*-*prd* gene was removed as *Xba*I fragment (one *Xba*I site in polylinker, the other ~400 bp upstream of the *Xba*I site of *prd*) and cloned into the *Xba*I site of the pW6 vector containing the P-element and the mini-white gene³⁴. The resulting *gsb*-*prd* rescue construct was injected into *w*¹¹¹⁸ embryos as described³⁰, and 11 transgenic lines were obtained. The transgenic *gsb*-*prd* embryos were stained with anti-Prd antibodies⁴ to verify that the *gsb*-*prd* gene is expressed in a *gsb*-like pattern. To test whether *gsb*-*prd* can rescue the *Df(2R)gsb*^{ix62} cuticular phenotype, six transgenic lines, carrying the *gsb*-*prd* transgene on either the first or third chromosome, were crossed with *w*¹¹¹⁸, *Df(2R)gsb*^{ix62}/CyO flies to gen-



erate *Df(2R)gsb*^{ix62}/+; *gsb*-*prd*/(+ or *w*¹¹¹⁸) flies. Embryos from these flies were allowed to develop until 24 h AEL, when cuticles were prepared as described²⁸.



Df(2R)gsb^{IX62} embryos (Fig. 3d), in which both *gsb* and *gsbn* are deleted^{2,20}, is completely rescued by the *gsb-prd* transgene of all six independent transgenic lines tested (Fig. 3c). Because in wild-type embryos, *gsb* maintains the expression of *wg*, which represses denticle formation^{9,21}, we expect, and find, that the rescue of the *gsb*⁻ cuticular phenotype results from activation of *wg* by *gsb-prd* in *gsb*⁻ embryos (Fig. 3e). These results corroborate our previous conclusion that Prd can substitute for the function of Gsb.

The three proteins Prd, Gsb and Gsbn are transcription factors because they contain a homeo- and a paired-domain. Because they can replace each other's gene regulatory functions, we expect them to recognize the same DNA sequences. To examine their DNA-binding specificities, each protein was analysed by footprinting on a previously characterized *gsb* *cis*-regulatory element, GEE. This element activates *gsb* by pair-rule proteins including Prd¹⁸, and so should be a target for Prd protein binding *in vivo*. As shown in Fig. 4, all three proteins bind precisely to the same sites, demonstrating that they share the same DNA-binding specificity. It is unclear whether the paired-domain, homeodomain or both participate in the binding to these extremely (A+T)-rich sequences. Their similarity to previously defined homeodomain-binding sites²², however, suggests an involvement of the homeodomain in this interaction.

It was startling to discover that Prd, Gsb and Gsbn, very different in their C-terminal halves, are functionally equivalent, as this implies that the essential determinant of the function of the three genes is encoded in their specific *cis*-regulatory regions controlling their temporal and spatial expression. As the N-terminal paired- and homeodomains have been highly conserved in these genes, they must have evolved from the same ancestral gene by repeated duplication. Our results thus provide the first experimental example that, during evolution, duplicated genes

may acquire new functions by changes in their regulatory regions generating an altered expression, rather than by mutations in their coding sequences. Moreover, the highly diverged C-terminal halves of Prd, Gsb and Gsbn imply that considerable changes in the coding region may be tolerated without significantly altering the function of the protein. The idea that mutations in the *cis*-regulatory rather than coding regions drive functional diversification has already been suggested²³.

We propose that during duplication, the duplicated portion of a gene, including its coding region, is juxtaposed to new *cis*-regulatory elements which change its expression and thus give it a new function—ultimately the activation of a different set of genes. Subsequently, mutations accumulating in the coding region may further contribute to an altered function of the gene. This hypothesis provides a simple explanation for what seem to be redundant functions revealed in gene knock-outs in vertebrates^{24–27}. In these cases, elimination of a gene's function is partly compensated for by the overlapping expression of a homologous protein with an equivalent function. If during evolution a new gene arises by the combination of new *cis*-regulatory sequences with the duplicated coding portion and some of the *cis*-regulatory elements of an old gene, the expression of the new gene will overlap in time and space with that of the old gene and generate redundancy. □

Received 4 August; accepted 27 October 1993.

1. Noll, M. *Curr. Opin. Genet. Dev.* **3**, 595–605 (1993).
2. Bopp, D., Burri, M., Baumgartner, S., Frigerio, G. & Noll, M. *Cell* **47**, 1033–1040 (1986).
3. Baumgartner, S., Bopp, D., Burri, M. & Noll, M. *Genes Dev.* **1**, 1247–1267 (1987).
4. Gütjahr, T., Frei, E. & Noll, M. *Development* **117**, 609–623 (1993).
5. Gütjahr, T., Patel, N. H., Li, X., Goodman, C. S. & Noll, M. *Development* **118**, 21–31 (1993).
6. Kilchherr, F., Baumgartner, S., Bopp, D., Frei, E. & Noll, M. *Nature* **321**, 493–499 (1986).
7. Bopp, D., Jamet, E., Baumgartner, S., Burri, M. & Noll, M. *EMBO J.* **8**, 3447–3457 (1989).
8. Ingham, P. W. & Martinez-Arias, A. *Cell* **68**, 221–235 (1992).
9. Li, X. & Noll, M. *EMBO J.* **12**, 4499–4509 (1993).
10. Struhl, G. *Nature* **318**, 677–680 (1985).
11. Ish-Horowitz, D. & Pinchin, S. M. *Cell* **51**, 405–415 (1987).
12. Morrissey, D., Askew, D., Raj, L. & Weir, M. *Genes Dev.* **5**, 1684–1696 (1991).
13. Manoukian, A. S. & Krause, H. M. *Genes Dev.* **6**, 1740–1751 (1992).
14. Manoukian, A. S. & Krause, H. M. *Development* **118**, 785–796 (1993).
15. Frasch, M., Hoey, T., Rushlow, C., Doyle, H. & Levine, M. *EMBO J.* **6**, 749–759 (1987).
16. Doe, C. Q., Smouse, D. & Goodman, C. S. *Nature* **333**, 376–378 (1988).
17. Patel, N. H., Schafer, B., Goodman, C. S. & Holmgren, R. *Genes Dev.* **3**, 890–904 (1989).
18. Li, X., Gütjahr, T. & Noll, M. *EMBO J.* **12**, 1427–1436 (1993).

19. Frigerio, G., Burri, M., Bopp, D., Baumgartner, S. & Noll, M. *Cell* **47**, 735–746 (1986).
20. Côté, S. et al. *EMBO J.* **6**, 2793–2801 (1987).
21. Bejsovec, A. & Martínez-Arias, A. *Development* **113**, 471–485 (1991).
22. Hoey, T. & Levine, M. *Nature* **332**, 858–861 (1988).
23. Jacob, F. *Science* **196**, 1161–1166 (1977).
24. Thomas, K. R. & Capecci, M. R. *Nature* **346**, 847–850 (1990).
25. McMahon, A. P. & Bradley, A. *Cell* **62**, 1073–1085 (1990).
26. Joyner, A. L., Herrup, K., Auerbach, B. A., Davis, C. A. & Rossant, J. *Science* **251**, 1239–1243 (1991).
27. Lohnes, D. et al. *Cell* **73**, 643–658 (1993).
28. Wieschaus, E. & Nüsslein-Volhard, C. in *Drosophila, a Practical Approach* (ed. Roberts, D. B.), 199–227 (IRL Press, Oxford, 1986).
29. Dambly-Chaudière, C. et al. *Cell* **69**, 159–172 (1992).
30. Rubin, G. M. & Spradling, A. C. *Science* **218**, 348–353 (1982).
31. DiNardo, S. & O'Farrell, P. H. *Genes Dev.* **1**, 1212–1225 (1987).
32. van den Heuvel, M., Nusse, R., Johnston, P. & Lawrence, P. A. *Cell* **59**, 739–749 (1989).
33. Hidalgo, A. *Mech. Dev.* **35**, 77–87 (1991).
34. Klemenz, R., Weber, U. & Gehring, W. J. *Nucleic Acids Res.* **15**, 3947–3959 (1987).
35. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560–564 (1977).
36. Stanojevic, D., Hoey, T. & Levine, M. *Nature* **341**, 331–335 (1989).

ACKNOWLEDGEMENTS. We thank K. Basler and E. Hafen for generating the transgenic *Hsprd*, *Hsgsb* and *Hsgsbn* flies, F. Ochsenbein for photography and artwork, T. Gütjahr for the anti-Gsb antiserum, R. Smith, M. Wilcox, K. Coleman, T. Kornberg, M. van den Heuvel, R. Nusse and A. Martínez-Arias for gifts of anti-En monoclonal antibody and anti-Wg antisera, and H. Noll for discussions and comments on the manuscript. This work has been supported by the Swiss NSF and by the Kanton Zürich.