Development 128, 395-405 (2001)
Printed in Great Britain © The Company of Biologists Limited
DEV5417

Multiple protein functions of Paired in *Drosophila* development and their conservation in the Gooseberry and Pax3 homologs

Lei Xue*, Xuelin Li‡ and Markus Noll§

Institute for Molecular Biology, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland
*Present address: Department of Genetics, Yale University School of Medicine, New Haven, Connecticut 06510, USA
‡Present address: Department of Biology, University of California at San Diego, 9500 Gilman Drive, La Jolla, Calfornia 92093, USA
§Author for correspondence (e-mail: noll@molbio.unizh.ch)

Accepted 20 November 2000; published on WWW 11 January 2001

SUMMARY

The *Drosophila* segmentation gene *paired*, whose product is homologous to the *Drosophila* Gooseberry and mammalian Pax3 proteins, has three general functions: proper development of the larval cuticle, survival to adulthood and male fertility. Both DNA-binding domains, the conserved N-terminal paired-domain and *prd*-type homeodomain, are required within the same molecule for all general *paired* functions, whereas a conserved His-Pro repeat located near its C terminus is a transactivation domain potentiating these functions. The C-terminal moiety of Paired includes two additional functional motifs: one, also present in Gooseberry and Pax3, is required for segmentation and cuticle development; the other, retained only in Gooseberry, is necessary for survival. The male fertility function, which cannot be replaced by Gooseberry and Pax3, is specified by the conserved N-terminal rather than the divergent C-terminal moiety of Paired. We conclude that the functional diversification of *paired*, *gooseberry* and *Pax3*, primarily determined by variations in their enhancers, is modified by adaptations of their coding regions as a necessary consequence of their newly acquired spatiotemporal expression.

Key words: *Drosophila melanogaster*, *paired*, paired-domain, *prd*-type homeodomain, conservation of protein function

INTRODUCTION

Position along the anteroposterior axis of the developing *Drosophila* embryo is initially defined by the sequential activities of four classes of segmentation genes – maternal coordinate genes and zygotic gap, pair-rule and segment-polarity genes – and the homeotic genes, which form a hierarchical network (Nüsslein-Volhard and Wieschaus, 1980; Baumgartner and Noll, 1990; Small and Levine, 1991; Peifer and Bejsovec, 1992; St Johnston and Nüsslein-Volhard, 1992). Genes within this hierarchy are expressed in progressively refined domains and thus define the position along this axis with increasing precision.

The *paired* (*prd*) gene, which belongs to the pair-rule class of segmentation genes (Nüsslein-Volhard and Wieschaus, 1980), encodes a transcription factor containing a paired-domain (PD) and an extended *prd*-type homeodomain (HD) in its N-terminal half (Bopp et al., 1986) and a His-Pro (PRD) repeat near its C-terminal end (Frigerio et al., 1986). Prd protein is initially expressed in a broad anterior stripe at the end of the thirteenth nuclear division of syncytial blastoderm (Gutjahr et al., 1993a). By mid-cellularization, Prd appears in an anterior dorsal patch and in a characteristic pair-rule pattern of seven stripes, which by cellular blastoderm are converted into 14 stripes spanning each parasegment boundary. During germ band extension, Prd expression decreases in the epidermal stripes but later accumulates in a few specific cells of the central nervous system (CNS) and certain head regions (Gutjahr et al., 1993a). Together with the other pair-rule genes, *prd* specifies position along each double-segmental repeat and activates the segment-polarity genes, including *gooseberry* (*gsb*), *wingless* (*wg*) and *engrailed* (*en*), which are expressed at a single-segment periodicity (DiNardo and O'Farrell, 1987; Ingham et al., 1988; Bopp et al., 1989). In *prd* mutant embryos, every other stripe of Gsb, Wg and En protein is abolished, which results in the loss of the posterior part of even-numbered parasegments and of the adjacent anterior part of odd-numbered parasegments (Nüsslein-Volhard and Wieschaus, 1980; Bopp et al., 1989; Ingham and Martinez Arias, 1992).

As a transcription factor, Prd is of particular interest because it possesses in its N-terminal portion two DNA binding domains, a PD and a HD (Bopp et al., 1986; Treisman et al., 1991), while most transcription factors contain only one DNA binding domain. The coexistence of a PD and HD was also observed in several other members of the *Pax* gene family (Noll, 1993). Although in vitro experiments suggest that these two domains can function either independently or cooperatively when present in the same molecule (Treisman et al., 1991; Underhill et al., 1995; Jun and Desplan, 1996), the situation in vivo remains unknown. To understand the biological significance of the coexistence of the PD and HD in Prd, their role had to be examined in vivo.

We have previously shown that Prd is required in vivo not only for the expression of segment-polarity genes and normal development of the larval cuticle but also for the survival of the embryo to adulthood and for male fertility (Xue and Noll, 1996; Xue and Noll, 2000). Two Prd homologs, the *Drosophila* Gsb and murine Pax3 proteins, which share with Prd a highly conserved N-terminal moiety including the PD and HD, but have divergent C-terminal portions, are able to perform some of these functions of Prd when placed under the control of the entire *prd cis*-regulatory region (Xue and Noll, 1996). It follows that the acquisition of new *cis*-regulatory elements rather than changes in the coding region is the major evolutionary drive for the functional diversification among these three genes. However, both *prd*-Gsb and *prd*-Pax3 perform the cuticle function of Prd at low efficiency, and only *prd*-Gsb is able to rescue the *prd* mutant embryos to adulthood when it is present in two copies (Xue and Noll, 1996). These results indicate that the coding region also plays an important role in further modification of protein functions.

Here we show which protein domains of Prd are required for each of its in vivo functions. By constructing a series of *prd* transgenes expressing wild-type or mutated Prd proteins under the control of the entire *prd cis*-regulatory region and introducing them into *prd* mutants, we determined which of the Prd functions can be rescued by the respective transgenes. Our results demonstrate that both the PD and HD of Prd have to be present in *cis* for the activation of segment-polarity genes, wild-type cuticle formation and viability. In addition to the PRD repeat, which constitutes an important activation domain facilitating all functions of Prd, its C terminus contains at least two essential functional motifs. One motif, required for its function in larval cuticle development, is present in the C termini of both Gsb and Pax3, whereas another motif, needed for its role in viability, is present only in the Gsb C terminus. Finally, the determinant for the male fertility function, which cannot be replaced by the *prd*-Gsb and *prd*-Pax3 transgenes, resides in the conserved N-terminal rather than the divergent C-terminal moiety of Prd. This observation strongly challenges the classic view using *percentage* amino acid identity as a measure of functional equivalence between homologous proteins.

MATERIALS AND METHODS

Construction of rescue and expression plasmids

All transgenes whose coding portions are illustrated in Fig. 1 were generated in two steps. First, wild-type cDNAs or their mutated versions were subcloned into pKSpL5, a derivative of Bluescript pKS$^+$ (Xue and Noll, 1996), or into the appropriate pAR vector (Studier and Moffat, 1986). Subsequently, these sequences were recovered as *Xba*I fragments from pKSpL5 or *Xba*I-*Nhe*I fragments from pAR subclones and inserted in the correct orientation into the *Spe*I site of the *prd*-0 P element construct (Xue and Noll, 1996).

To produce the pKSpL5-Gsb, pKSpL5-Prd and pKSpL5-Pax3 subclones, the *gsb*-cDNA BSH9c2 (Baumgartner et al., 1987), *prd*-cDNA c7340.6 (Frigerio et al., 1986) and *Pax3*-cDNA (Goulding et al., 1991), respectively, were cloned into the unique *Eco*RI site of the polylinker of pKSpL5 in the required orientation. To obtain pKSpL5-GsbP17L, the 1.4 kb *Eag*I-*Hin*dIII fragment in pKSpL5-Gsb was replaced by the corresponding PCR product, generated with the primers gsbP17L (5′-TTC ATC AAC GGC CGT CTG TTG-3′) and T3 (5′-ATT AAC CCT CAC TAA AG-3′). To generate pKSpL5-*int*-Gsb, the *prd* intron was amplified by PCR from *prd*-SN20 with the following primers (mismatches were introduced to generate the underlined *Spe*I sites): prdint1 (5′- GAT ATT CTA CTA GTC AAG GTG AG-3′) and prdint2 (5′- GCC GCT GTA CTA GTC TGG AAT GA-3′). Subsequently, the PCR product was inserted into the unique *Spe*I site of the polylinker of pKSpL5-Gsb in the appropriate orientation. The pKSpL5-PrdΔPRD construct was produced by inserting the blunt-ended 2.2 kb *Hin*dIII-*Eco*RI fragment from PrdΔPR (Cai et al., 1994) between the two *Sma*I sites in the pKSpL5 polylinker. To obtain pKSpL5-PrdN+GsbC, the 850 bp *Eco*RI-*Pvu*II fragment of c7340.1 *prd*-cDNA (Frigerio et al., 1986) and the 750 bp *Fsp*I-*Eco*RI fragment of pKSpL5-Gsb were ligated into the *Eco*RI site of pKSpL5. To generate pKSpL5-PrdN, the 2.4 kb *Eco*RI-*Spe*I fragment in pKSpL5-Prd was replaced with the 1.0 kb *Eco*RI-*Avr*II fragment obtained by PCR from pKSpL5-Prd with the use of the primers T7 (5′-AAT ACG ACT CAC TAT AG-3′) and prd-4 (5′-GCC TGA GAC CTA GGT GTG CTG-3′).

pAR-GsbΔP, pAR-GsbΔH, pAR-GsbC and pAR-GsbN were produced by PCR mutagenesis and subcloned into the pAR3040 vector (X. Li, L. X. and M. N., unpublished). To generate pAR-Gsb+PRD, the 650 bp *Sac*I-*Nhe*I fragment from pAR-gsb.fl (Gutjahr et al., 1993b) was replaced with the corresponding 950 bp fragment from a *prd*-cDNA subcloned in pAR3038 (Gutjahr et al., 1993a), pAR-prd. To construct pAR-GsbN+PoxnC, the *Nde*I/*Spe*I-digested GsbN PCR product (amplified from pAR-Gsb with the primers T7 and gsbres8 [5′-CCT GCT GGG TGA CTA GTT GCT TGC GCA-3′]) was ligated with the *Spe*I/*Bcl*I-digested PoxnC PCR product (amplified from the *poxn*-cDNA P4c6 [Dambly-Chaudière et al., 1992] with the primers poxnres3 [5′-TCA AAA CTT GAT CAG TGG CGA GA-3′] and poxnres4 [5′-GCG CAA CAG CGG ACT AGT GAC CGA TGA GAT-3′]) between the *Nde*I and *Bam*HI sites of the pAR3040 polylinker. To generate pAR-GsbN+Pax3C, the *Nde*I/*Spe*I-digested GsbN PCR product and the *Spe*I/*Bam*HI-digested Pax3C PCR fragment (amplified from pKSpL5-Pax3 with the primers pax3res2 [5′-CCA GGA GGA TCC ACC CCC TAG AAC GT-3′] and pax3res3 [5′-TGG AGG AAA CTA GTT GGA GCC AA-3′]) were ligated between the *Nde*I and *Bam*HI sites of pAR3040, whereas pAR-GsbN+PrdC was obtained by ligating the *Nde*I/*Spe*I-digested GsbN fragment and the *Nhe*I/*Bcl*I-digested PrdC PCR product (amplified from pKSpL5-Prd with the primers prdC-1 [5′-CGC AAG CAG CTA GCC TCG GTC TC-3′] and prdC-2 [5′-GTA GGT GGT TGA TCA GTG TCT CT-3′]) between the *Nde*I and *Bam*HI sites of pAR3040. Finally, the pAR-Pax3N+GsbC was generated by inserting the *Nde*I/*Bam*HI-digested Pax3N PCR product (amplified from pKSpL5-Pax3 with the primers pax3res4 [5′-GCT GCC CCC CAT ATG ACC ACG CT-3′] and pax3res5 [5′-AGT TGA TTG GAT CCA GCT TGT-3′]) and the *Bam*HI-*Nhe*I fragment of pAR-GsbC between the *Nde*I and *Nhe*I sites of the pAR3040 vector.

Generation of transgenic flies

Rescue constructs were injected together with pUChspΔ2-3 P-element helper plasmid (D. Rio, personal communication) into ry^{506} embryos, and ry^+ transformants were selected (Rubin and Spradling, 1982).

Transgenic prd^- embryos carrying one or two copies of the specified transgenes were obtained as follows. Two types of stocks were established for all transgenes except *prd*-GsbN, *prd*-GsbN+PoxnC and *prd*-PrdN, which exhibit a dominant-negative effect on Prd functions, namely *Df(2L)Prl/SM1*; *P/P* and $prd^{2.45}$*/SM1*; *P/P* (*P* stands for the P elements that contain the transgenes). To rescue the cuticle functions of Prd with one or two copies of the transgenes, *prd* mutant embryos were collected from crosses between $prd^{2.45}$*/SM1*; *P/P* and $prd^{2.45}$*/SM1* flies or from crosses of $prd^{2.45}$*/SM1*; *P/P* flies inter se. To rescue the viability and male fertility functions of Prd with one or two copies of the transgenes, $prd^{2.45}$*/SM1*; *P/P* flies were crossed with *Df(2L)Prl/SM1* or *Df(2L)Prl/SM1*; *P/P* flies. When the transgene failed to rescue the cuticle functions, its ability to provide the viability functions was assayed by supplying the cuticle

functions with two copies of the *prd*-Pax3 transgene. To rescue the male fertility function of Prd with one or two copies of the *prd*-PrdN (or *prd*-GsbN) transgene, male flies of the genotype *Df(2L)Prl prd*-GsbN+PrdC/*prd*$^{2.45}$ *prd*-GsbN+PrdC; *prd*-PrdN (or *prd*-GsbN)/*TM3, Sb* or *Df(2L)Prl prd*-GsbN+PrdC/*prd*$^{2.45}$ *prd*-GsbN+PrdC; *prd*-PrdN (or *prd*-GsbN)/*prd*-PrdN (or *prd*-GsbN) were obtained from crosses between *Df(2L)Prl prd*-GsbN+PrdC/*SM1*; *prd*-PrdN (or *prd*-GsbN)/*TM3, Sb* and *prd*$^{2.45}$ *prd*-GsbN+PrdC/*SM1*; *prd*-PrdN (or *prd*-GsbN)/*TM3 Sb* flies and subsequently tested for fertility. To rescue the male fertility function of Prd with one or two copies of the *prd*-GsbN+Pax3C transgene, one copy of *prd*-Gsb was supplied to rescue the viability functions by crossing *Df(2L)Prl prd*-GsbN+Pax3C/*SM1*; *prd*-Gsb flies with *prd*$^{2.45}$/*SM1* or *prd*$^{2.45}$ *prd*-GsbN+Pax3C/*SM1* flies. The *prd*-GsbN+PoxnC transgene could not be tested for the viability and male fertility functions of Prd because of its strong dominant-negative effect, which results in lethality in the presence of one copy of the wild-type *prd* gene.

Expression and purification of proteins and band-shift assays

Prd, Gsb, GsbΔP, GsbΔH, GsbC, GsbN, GsbN+PoxnC, GsbN+Pax3C and GsbN+PrdC proteins were expressed in *E. coli* BL21(DE3) cells, transformed with the corresponding pAR subclones and purified to about 50% purity as described previously (Gutjahr et al., 1993a). Band-shift assays were performed as described (Xue and Noll, 1996).

Immunostaining of embryos and preparation of cuticles

Immunostaining of embryos with the anti-Gsb antiserum was carried out as described (Gutjahr et al., 1993b). Double-staining of embryos for β-gal and Gsb, Wg or En protein was performed according to Lawrence et al. (Lawrence et al., 1987). Cuticles were prepared essentially as described (Wieschaus and Nüsslein-Volhard, 1986).

RESULTS

Both PD and HD are strictly required in *cis* for the functions of Prd

The independent contributions of the PD and HD to the in vivo functions of the Prd protein were investigated with the aid of flies carrying the *prd*-Gsb transgene, which expresses the Gsb protein under the control of the entire *prd cis*-regulatory region and performs all Prd functions required for survival of the embryos to adults (Xue and Noll, 1996). Three rescue constructs, *prd*-GsbΔP, *prd*-GsbΔH and *prd*-GsbC, were derived from *prd*-Gsb by deleting the PD, the extended *prd*-type HD, or both (Fig. 1). More than 10 independent transgenic lines were obtained for each construct. Each transgene was subsequently introduced into *prd* mutants and tested for its ability to rescue the following functions of Prd: (i) activation of segment-polarity genes downstream of *prd*, such as *gsb*, *wg* and *en*, (ii) development of wild-type larval cuticle, and (iii) survival to adulthood after rescue of the cuticular phenotype by *prd*-Pax3. The first two tests assay for what we will henceforth call the cuticle function of Prd, the third for its viability function (Xue and Noll, 1996). The *prd* embryos used in these and all subsequent experiments were either homozygous for the *prd*$^{2.45}$ allele or hemizygous when this allele was combined with the deficiency *Df(2L)Prl*. The *prd*$^{2.45}$ allele does not produce any functional Prd protein because it carries a 1.1 kb insertion (Kilchherr et al., 1986) after amino acid 45 of the PD (Frigerio et al., 1986). The tests showed that none of the three transgenes is able to perform any function of Prd, even when two copies are present (Figs 1-4).

In discussing these results the following findings are relevant. First, all three transgenes express truncated Gsb proteins at a level comparable to that expressed by the *prd*-Gsb transgene and detected by anti-Gsb specific for the C-terminal moiety (Gutjahr et al., 1993b). Second, homozygous *prd*-GsbΔP exerts a weak dominant-negative effect on the cuticle function of wild-type Prd, i.e., a few embryos show a weak pair-rule phenotype while most embryos are wild-type and survive to adulthood. Third, GsbΔH and GsbC proteins are able to perform the Gsb function in cuticle development when expressed under the control of *gsb cis*-regulatory elements (X. L., L. X. and M. N., unpublished). Consequently, the products of the transgenes are made and are partially functional, but they cannot rescue any of the Prd functions. Evidently, both the PD and HD are required for the in vivo functions of Prd.

To minimize the possible disruption of the overall protein structure that might result from the deletion of entire domains, we constructed *prd*-GsbP17L, which contained a PD in which Pro at position 17 was replaced by Leu (Fig. 1), identical to the amino acid substitution resulting from point mutations in human PAX3 and PAX5. The first human mutation produces Waardenburg's syndrome I (Baldwin et al., 1992; Tassabehji et al., 1992), the second results in a complete loss of its DNA-binding activity in vitro (Czerny et al., 1993). The *prd*-GsbP17L, like *prd*-GsbΔP, fails to rescue any Prd functions when similarly tested (Figs 1 and 3D,G) and hence suggests that DNA binding of the PD is essential for the rescue of Prd functions.

When the PD, HD or both are deleted in Gsb, it can no longer bind to GEE1 (Xue and Noll, 1996), a Prd target site located in the *gsb* early enhancer GEE (Li et al., 1993), as evident from band-shift assays (Fig. 5, lanes 8-10). This is in agreement with the in vivo results and implies that both domains recognize binding sites in GEE1 with which they probably interact cooperatively.

To test whether *prd*-GsbΔP and *prd*-GsbΔH can complement each other to rescue some Prd functions, we introduced both transgenes into the same *prd* mutants in four different combinations: one copy or two copies of each transgene or two copies of one combined with one copy of the other. In all four combinations, no rescue is observed of segment-polarity gene activation (Fig. 2P-R), cuticular phenotype (Fig. 3H) and viability (when the cuticle function was provided by *prd*-Pax3; not shown). It follows that *prd*-GsbΔP and *prd*-GsbΔH cannot complement for any Prd functions. Consistent with this result, GsbΔP and GsbΔH proteins when present together failed to interact with GEE1 in a band-shift assay (Fig. 5, lane 11). We conclude that both the PD and HD are strictly required in *cis* for the in vivo functions of Prd, most probably because of their cooperative DNA-binding activities.

The *prd* intron is dispensable for the normal functions of Prd

From the original *prd* rescue construct *prd*-SN20, capable of performing all Prd functions (Gutjahr et al., 1994), the *prd*-Gsb transgene was derived by replacing a genomic fragment including the end of the leader, most of the coding region and the 356 bp intron of *prd* with a *gsb*-cDNA (Xue and Noll, 1996). As indicated by our previous results, activation of Prd target genes and the cuticular phenotype of *prd*$^{-}$ embryos could

be fully rescued with two and partially with one copy of the *prd*-Gsb transgene. In addition, the *prd*-Gsb transgene was able to rescue some *prd* mutants to adulthood when present in two copies. These results demonstrated that the *prd*-Gsb transgene, albeit at a lower efficiency than *prd*-SN20, is able to perform all the Prd functions that are required by the animal to survive to adulthood. Further studies showed that all male flies rescued by the *prd*-Gsb transgene are sterile, presumably because the severely reduced accessory glands have lost their functions (Xue and Noll, 2000). Therefore, a new function of Prd, henceforth called the male fertility function, was uncovered by the presence of the *prd*-Gsb transgene in *prd* mutants. This conclusion was confirmed by the observation that a *prd* rescue construct, differing from *prd*-SN20 by the lack of 5 kb of the downstream regulatory sequences, is unable to confer fertility on *prd* mutant males rescued to viable adults since these lack accessory glands (Bertuccioli et al., 1996; Xue and Noll, 2000).

These results left open the question of whether the reduced ability of *prd*-Gsb to replace the *prd* functions described above are attributable to the protein coding region or to the presence of enhancers in the missing *prd* intron. To distinguish between these possibilities, we constructed two transgenes (Fig. 1). The *prd*-Prd transgene was derived from *prd*-SN20 by deleting the intron. In the other transgene, *prd-int*-Gsb, the *prd* intron was inserted into *prd*-Gsb between the *prd* leader and the *gsb*-cDNA sequences. We found that one copy of the *prd*-Prd transgene is sufficient to fully rescue all Prd functions in a prd^{-} background, namely cuticle development (Fig. 6A), viability (Fig. 4) and male fertility (Fig. 1). On the other hand, the *prd-int*-Gsb transgene is functionally indistinguishable from *prd*-Gsb as in its presence the cuticular phenotype of *prd* mutants is rescued partially by one copy (Fig. 6B) and completely by two copies (Fig. 1). In addition, two copies of the *prd-int*-Gsb transgene can also rescue about 10% of *prd* mutants to adulthood (Fig. 4), an efficiency comparable to that of *prd*-Gsb (Fig. 4; Xue and Noll, 1996). By contrast, all rescued males are sterile (Fig. 1). From these results, we conclude that the *prd* intron is dispensable for the normal functions of *prd*.

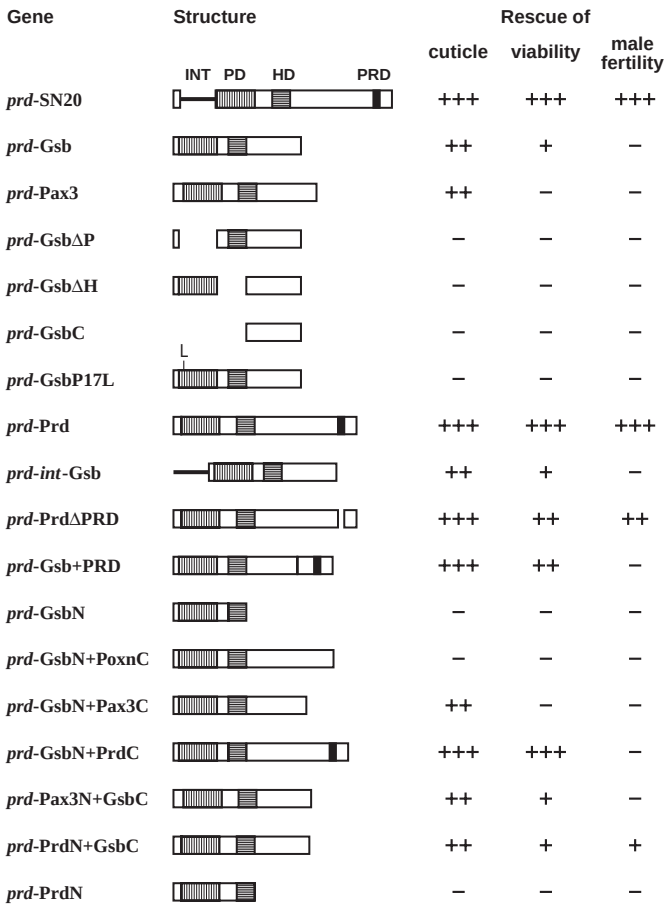

Fig. 1. Rescue of Prd functions by *prd* transgenes. The portions of the coding regions present in the *prd* transgenes listed in the left column are displayed next to them as vertically hatched (paired-domain, PD), horizontally hatched (*prd*-type homeodomain, HD), filled (His-Pro domain, PRD) or open bars (remaining coding regions, mainly encoding C termini of proteins), while the presence of the *prd* intron (INT) in two transgenes is indicated as a horizontal black line. All structures of the coding regions and the *prd* intron are drawn to scale; the full-length Prd protein consists of 613 amino acids. The ability of the transgenes to rescue the Prd functions by restoring a wild-type cuticular phenotype (cuticle), viability and male fertility is quantified on the right by +++ if one copy of the transgene is sufficient for full rescue; ++ if two copies of the transgene are sufficient for full rescue, while one copy results in only a partial rescue; + if two copies of the transgene result in a partial rescue, while one copy produces no rescue; and – if no rescue is obtained with either one or two copies of the transgenes. Transgenic prd^{-} embryos carrying one or two copies of the specified transgenes were obtained as described in Materials and methods.

The PRD repeat constitutes an important transactivation domain

In addition to the PD and HD, the *prd* gene encodes a third domain, the 21-amino acid His-Pro (or PRD) repeat, located near its C-terminal end (Frigerio et al., 1986). To investigate the in vivo function of the PRD repeat, we constructed two transgenes: *prd*-PrdΔPRD and *prd*-Gsb+PRD (Fig. 1). In *prd*-PrdΔPRD, the PRD repeat was removed by PCR mutagenesis, while in *prd*-Gsb+PRD the last 12 amino acids of Gsb were replaced by the 118-amino acid C-terminal end of Prd, which includes the PRD repeat. Although one copy of the *prd*-PrdΔPRD transgene can fully rescue the cuticular phenotype of *prd* mutants (Fig. 6C), its rescue efficiencies for viability (Fig. 4) and male fertility (Fig. 1) are dramatically reduced. Thus, one copy of *prd*-PrdΔPRD rescues less than 20% of *prd* mutants to viable adults, an efficiency that is 4- to 6-fold lower than that seen with *prd*-Prd (Fig. 4), and the fertility of the rescued males is strongly reduced (Fig. 1). However, two copies of *prd*-PrdΔPRD completely rescue both the embryonic lethality (Fig. 4) and male sterility of *prd* mutants (Fig. 1).

Moreover, *prd*-Gsb+PRD displays considerably enhanced rescue efficiencies of the cuticle phenotype and embryonic viability as compared to those observed with *prd*-Gsb. One copy of *prd*-Gsb+PRD completely rescues the cuticular phenotype of *prd* mutants (Fig. 6D). It can also rescue the embryonic lethality, partially with one copy and fully with two copies, an efficiency that is about 9-fold higher than that of *prd*-Gsb (Fig. 4). However, neither one nor two copies of *prd*-Gsb+PRD are able to rescue the male sterility of *prd* mutants (Fig. 1).

Evidently, the PRD repeat, though not absolutely required,

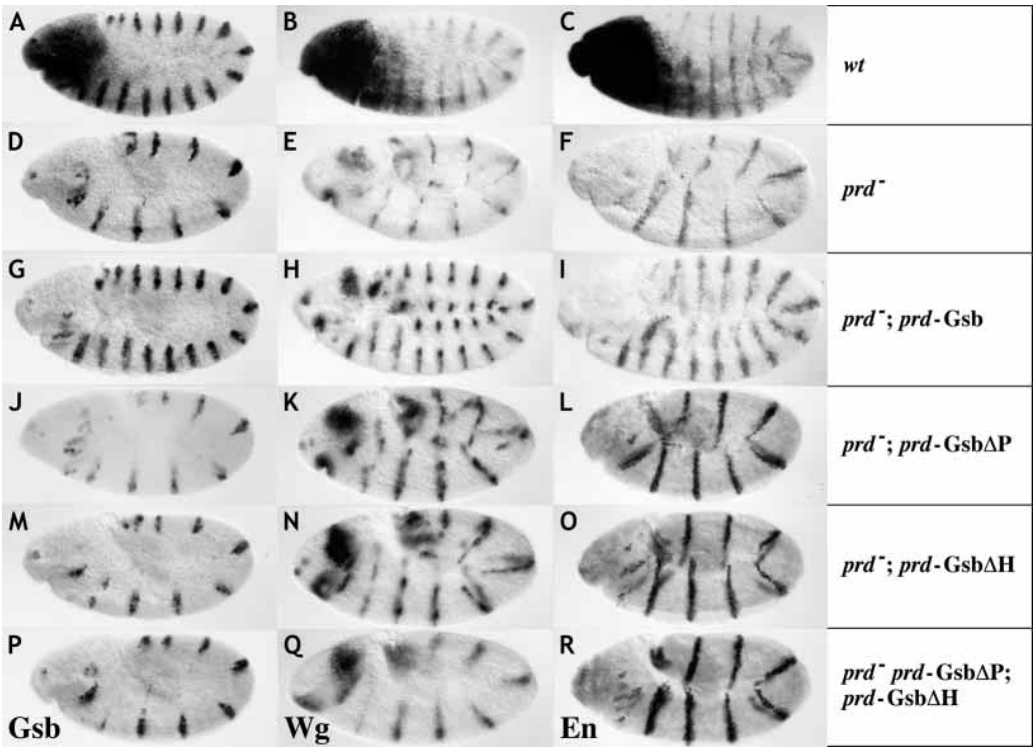

Fig. 2. Rescue of Gsb, Wg and En expression patterns in *prd*$^{-}$ embryos depends on the presence in *cis* of the paired- and homeodomain in the *prd*-Gsb transgene. Expression patterns of Gsb (A,D,G,J,M and P), Wg (B,E,H,K,N and Q) and En (C,F,I,L, O and R) in wild-type (A-C) and homozygous *prd*$^{2.45}$ embryos carrying no (D-F) or two copies of *prd*-Gsb (G-I), *prd*-GsbΔP (J-L), *prd*-GsbΔH (M-O) or *prd*-GsbΔP and *prd*-GsbΔH (P-R) are shown during the extended germ band stage of embryogenesis. Embryos are oriented with their anterior to the left and dorsal side up. Expression of Gsb, Wg and En is fully rescued to the wild-type pattern (A-C) by two copies of the *prd*-Gsb transgene (G-I; Xue and Noll, 1996), but displays the pair-rule pattern of *prd*$^{-}$ embryos (D-F) when 'rescued' by two copies of *prd*-GsbΔP (J-L), *prd*-GsbΔH (M-O) or *prd*-GsbΔP and *prd*-GsbΔH (P-R). Embryos were collected from *prd*$^{2.45}$*/CyO, hb-lacZ* flies (A-F), from *prd*$^{2.45}$*/CyO, hb-lacZ* flies homozygous for *prd*-Gsb (G-I), *prd*-GsbΔP (J-L) or *prd*-GsbΔH (M-O) on the third chromosome or from *prd*$^{2.45}$ *prd*-GsbΔP/*CyO, hb-lacZ*; *prd*-GsbΔH flies (P-R). The *lacZ*-marked *CyO* balancer chromosome expresses β-galactosidase under the control of the *hunchback* (*hb*) promoter (Driever et al., 1989). These embryos were doubly stained for β-galactosidase and Gsb, Wg or En protein by the use of mouse monoclonal antibodies against En and rabbit antisera against β-galactosidase, Gsb or Wg. Three quarters of the embryos showed β-galactosidase expression in the anterior region (A-C) and thus contained one or two copies of the wild-type *prd* gene, whereas the remaining quarter did not stain for β-galactosidase (D-R), as expected for homozygous *prd*$^{2.45}$ mutants.

potentiates all functions of Prd, and thus functions as an important transcriptional activation domain.

The C-terminal portion of Prd includes at least three functional motifs

Our previous work showed that two evolutionary alleles of *prd*, *prd*-Gsb and *prd*-Pax3, which share the same *cis*-regulatory region but not identical coding sequences with *prd*, have conserved most of the functions of Prd (Xue and Noll, 1996). However, neither Gsb (Xue and Noll, 2000) nor Pax3 are able to replace the male fertility function of Prd, as all *prd*$^{-}$ males rescued by two copies of *prd*-Gsb or combinations of *prd*-Gsb with *prd*-Pax3 are sterile (Fig. 1). These results demonstrated that the Prd, Gsb and Pax3 proteins are functionally nearly equivalent. Although no significant similarity has been found among their C-terminal sequences, it is probable that some functional motifs in their 3-D structures have been conserved during evolution (Xue and Noll, 1996). Thus, the C terminus of Prd may contain, in addition to the PRD repeat, three motifs ('motif' defined as a single feature of a domain that permits a specific interaction of the domain with another protein or nucleic acid, and thus may consist in extreme cases of a single specific amino acid within a domain, such as position 50 of the homeodomain) or domains that are necessary for the cuticle, viability and male fertility functions of Prd. While both *prd*-Gsb and *prd*-Pax3 can support normal cuticle development, only *prd*-Gsb is able to rescue the lethality of *prd* mutants (Xue and Noll, 1996). Hence, we assume that the motif required for the cuticle function is conserved among the C termini of Prd, Gsb and Pax3, whereas the motif essential for viability is expected to be present only in the C termini of Prd and Gsb. In addition, as both Gsb and Pax3 are unable to replace Prd in promoting male fertility, the motif performing this function is apparently missing in both Gsb and Pax3.

To test this hypothesis, we constructed a series of transgenes, *prd*-GsbN, *prd*-GsbN+PoxnC, *prd*-GsbN+Pax3C and *prd*-GsbN+PrdC, all derived from *prd*-Gsb by either deleting its C-terminal moiety or replacing it with that of Pox neuro (Poxn) (Dambly-Chaudière et al., 1992), Pax3 or Prd (Fig. 1). All these proteins bind, in band-shift assays, to the Prd target site GEE1 (Xue and Noll, 1996) with affinities similar to those of Prd and Gsb (Fig. 5, lanes 2-7), presumably by their two DNA-binding domains of GsbN. If the functional divergence observed for *prd*-Pax3, *prd*-Gsb and *prd*-Prd resides in the C-terminal portions of their coding regions, we expect (i) that *prd*-GsbN and *prd*-GsbN+PoxnC lack the hypothetical motifs required for Prd-dependent cuticle formation, viability and male fertility, and thus are unable to perform any functions of Prd; (ii) that *prd*-GsbN+Pax3C, which is assumed to contain the motif for the cuticle function, can rescue the cuticle phenotype, but not the lethality of *prd* mutants; and (iii) that *prd*-GsbN+PrdC, which includes all functional motifs present in the C terminus of Prd, is able to execute all Prd functions. Most of these expectations were borne out by the following

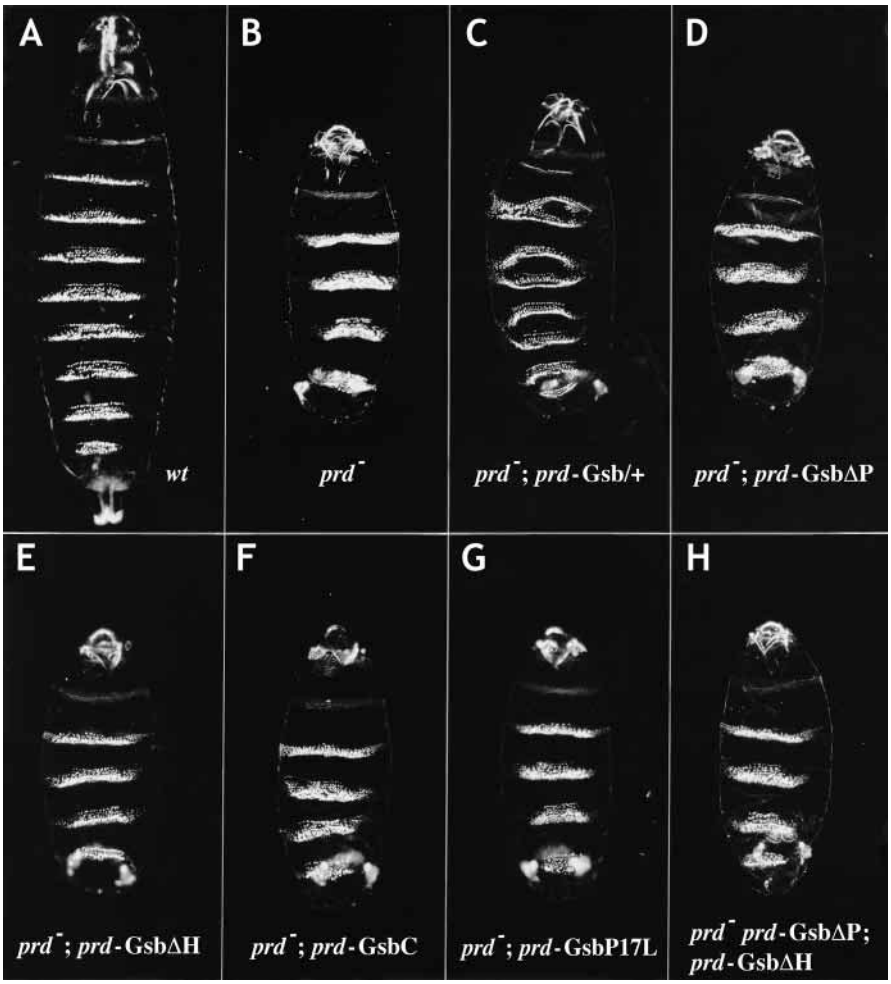

Fig. 3. Rescue of the cuticular phenotype of *prd*⁻ embryos depends on the presence in *cis* of a functional paired- and homeodomain in the *prd*-Gsb transgene. Ventral views of a cuticle of a wild-type (A) or $prd^{2.45}/prd^{2.45}$ embryo (B) or of a $prd^{2.45}/prd^{2.45}$ embryo carrying one copy of *prd*-Gsb (C) or two copies of *prd*-GsbΔP (D), *prd*-GsbΔH (E), *prd*-GsbC (F), *prd*-GsbP17L (G) or of both *prd*-GsbΔP and *prd*-GsbΔH (H) are shown under dark-field illumination (anterior is up). *prd*⁻ embryos carrying one or two copies of the specified transgenes were obtained as described in Materials and Methods and the legend to Fig. 2. Embryos were collected and allowed to develop for 24 hours at 25°C before cuticle preparation.

findings. First, either one or two copies of *prd*-GsbN or *prd*-GsbN+PoxnC are unable to provide the cuticle function (Fig. 1). Moreover, both *prd*-GsbN and *prd*-GsbN+PoxnC exert a dominant-negative effect on the cuticle function of Prd since one copy of *prd*-GsbN+PoxnC or two copies of *prd*-GsbN are lethal and produce a pair-rule phenotype when only one wild-type allele of *prd* is present (not shown). This effect might result from formation of inactive heterodimers of these proteins

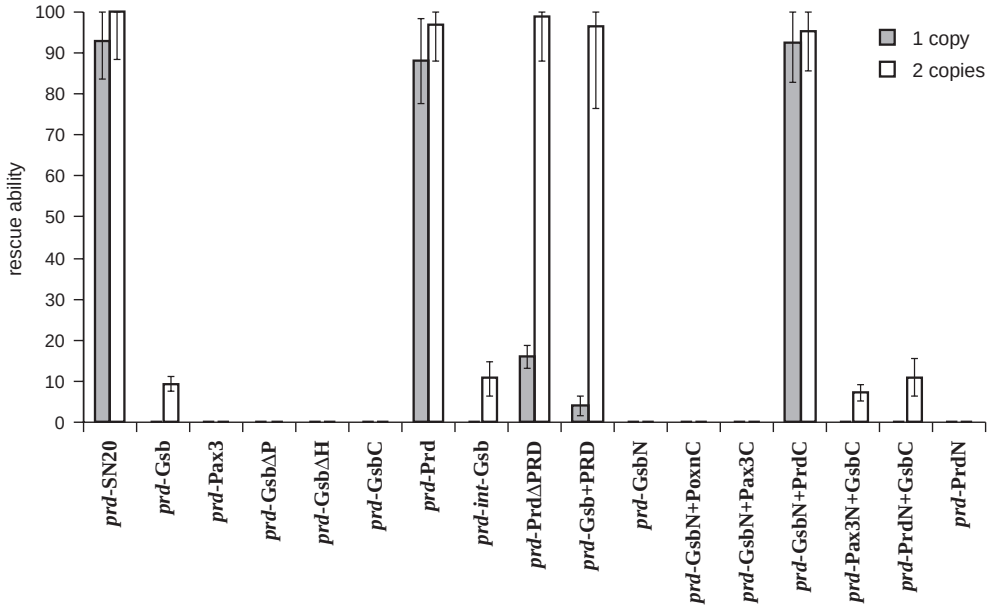

Fig. 4. Rescue of *prd* mutants to viable adults with one or two copies of *prd* transgenes. The percentage of *prd*⁻ embryos rescued to adulthood by one (stippled bars) or two copies (open bars) of the *prd* transgenes indicated on the abscissa, is shown on the ordinate. Transgenic embryos were produced by the crosses described in Materials and Methods.

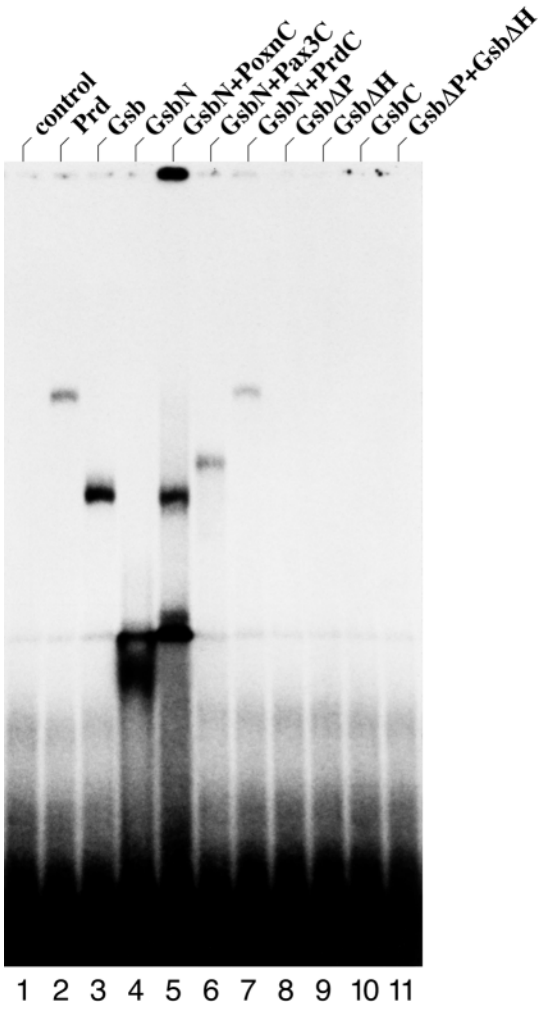

Fig. 5. Specific recognition of the Prd target site GEE1 of *gsb* by wild-type and mutant Prd proteins. Band-shift assays of the oligonucleotide GEE1 were performed with 1 μg of protein from crude extracts of *E. coli* BL21 (DE3) cells, in which expression of the empty vector pAR3040 was induced (control), with 30 ng (lanes 2, 6, 7) or 100 ng (lanes 3-5, 8-10) of protein from partially purified extracts of *E. coli* BL21 (DE3) cells, in which expression of Prd, Gsb, GsbN, GsbN+PoxnC, GsbN+Pax3C, GsbN+PrdC, GsbΔP, GsbΔH or GsbC had been induced (see Materials and Methods). The band-shift assay in the presence of both GsbΔP and GsbΔH (lane 11) was performed with 100 ng each of partially purified extracts.

with Prd through their HDs or from competition for Prd DNA target sites of GsbN or GsbN+PoxnC, as these proteins possess similar DNA-binding, yet not the same transcriptional activation capabilities as Prd. Second, *prd*-GsbN+Pax3C, like *prd*-Pax3, is able to rescue, partially with one copy (Fig. 6E) and completely with two copies (Fig. 1), the cuticular phenotype of *prd* mutants, yet fails to rescue lethality (Fig. 4) and male sterility (Fig. 1) of these mutants. Finally, as expected, one copy of *prd*-GsbN+PrdC can fully rescue all Prd functions that are required for embryonic survival to adulthood (Figs 4, 6F). However, surprisingly *prd*-GsbN+PrdC fails to rescue the male sterility of *prd* mutants (Fig. 1), which implies that at least part of the motif required to promote male fertility resides in the highly conserved N-terminal moiety of Prd.

We conclude that, in addition to a transactivation domain, the C-terminal portion of Prd includes sequences important for its cuticle and viability functions. Furthermore, the motif required for cuticle formation is retained in the C termini of both Gsb and Pax3, the viability function only in that of Gsb.

Mapping of viability and cuticular functions to the C-terminal moiety of Prd

An approximate location in Prd of the motif required for its cuticle function may be determined from the prd^{IIN} mutant (Tearle and Nüsslein-Volhard, 1987). Since this *prd* allele produces a weak cuticular phenotype and truncates the 126 C-terminal amino acids of Prd (Bertuccioli et al., 1996), part of the motif required for the cuticle function is located in this C-terminal portion of Prd (Fig. 7). In addition, another truncated Prd protein, $Prd^{\Delta PRT}$, lacking the last 74 amino acids including the PRD repeat, is still able to promote normal cuticle formation (Bertuccioli et al., 1996). Therefore, the motif required for cuticle function must be located around residue 487 and before residue 540 of Prd (Fig. 7).

To map the motif necessary for the viability function, we introduced either one or two copies of the *prd*-Pax3 transgene into $Df(2L)Prl/prd^{IIN}$ mutants. While one copy of the *prd*-Pax3 transgene is unable to rescue any mutants to adulthood, two copies of *prd*-Pax3 do rescue some of these mutants to viable adults (not shown), yet cannot rescue $Df(2L)Prl/prd^{2.45}$ embryos, which lack any functional Prd protein (Xue and Noll, 1996). This suggests that at least part of the sequences needed for viability must reside between the beginning of the C-terminal portion of Prd at residue 273 and the end of Prd^{IIN} at residue 487 (Fig. 7).

The male fertility function is determined by the N-terminal portion of Prd

Although the N-terminal portions of Prd, Gsb and Pax3 are highly conserved and exhibit the same DNA-binding ability in vitro (Xue and Noll, 1996), the failure of *prd*-GsbN+PrdC to rescue the male sterility of *prd* mutants implies that they are functionally different. To corroborate this conclusion, two transgenes were constructed, *prd*-Pax3N+GsbC and *prd*-PrdN+GsbC (Fig. 1), in which the N-terminal portion of Gsb in *prd*-Gsb is replaced by that of Pax3 or Prd. In *prd* mutants, *prd*-Pax3N+GsbC behaves like *prd*-Gsb. It is able to rescue the cuticular phenotype, partially with one copy (Fig. 6G) and fully with two copies (Fig. 1). In particular, two copies of *prd*-Pax3N+GsbC can rescue some *prd* mutants to viable adults with an efficiency comparable to that of *prd*-Gsb (Fig. 4), but fail to rescue the sterility of male flies (Fig. 1). These results indicate that the N-terminal portions of Gsb and Pax3 are equivalent with respect to the functions of Prd. By contrast, *prd*-PrdN+GsbC is not only able to rescue the cuticular phenotype (Figs 1, 6H) and lethality (Fig. 4) of *prd* mutants with similar efficiencies as *prd*-Gsb, but also the male fertility function of Prd (Fig. 1). It follows that the N-terminal portion of Prd includes the crucial determinant for its role in male fertility (Fig. 7), presumably through its recognition of specific DNA target sites. The fact that *prd*-PrdN+GsbC needs two copies to rescue male sterility is consistent with the finding from *prd* mutants, rescued by *prd*-PrdΔPRD, that the PRD repeat is required to enhance the efficiency of the protein in promoting male fertility (Fig. 1).

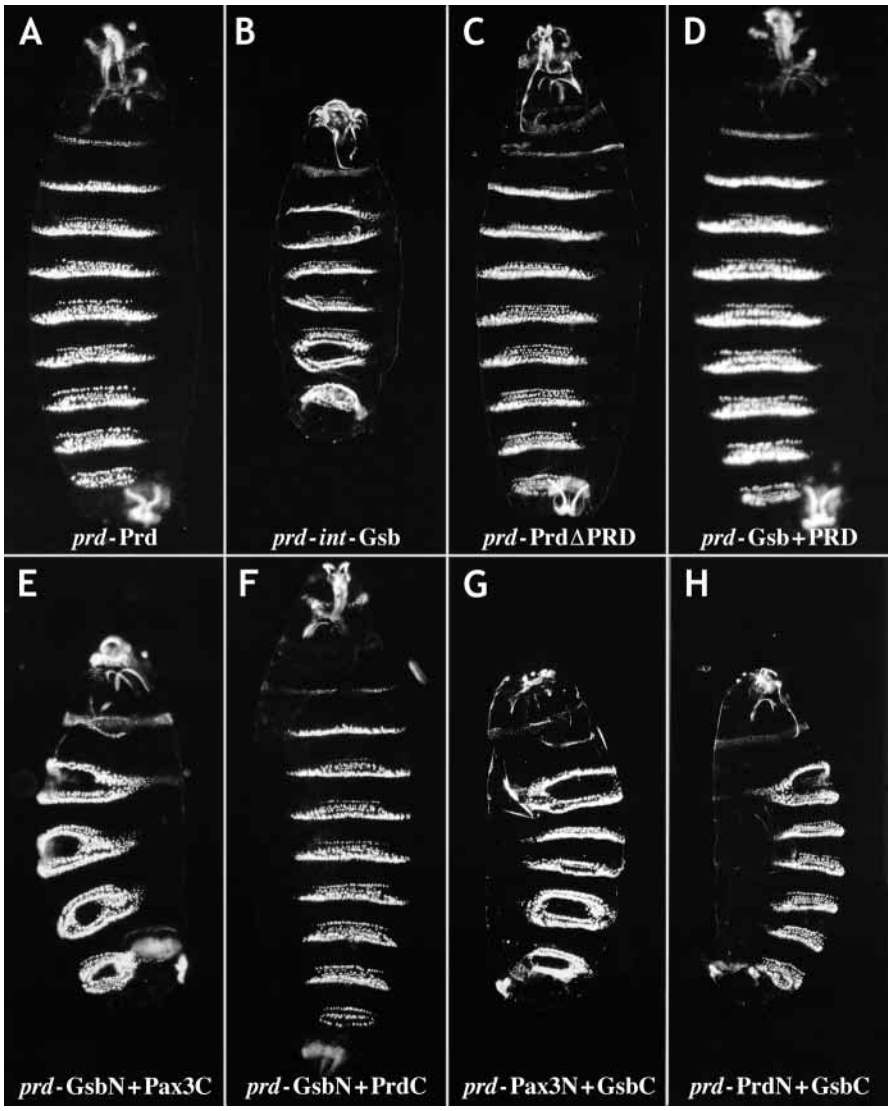

Fig. 6. Rescue of the cuticular phenotype of *prd*$^{-}$ embryos by *prd* transgenes. Ventral views of cuticles of *prd*$^{2.45}$/*prd*$^{2.45}$ embryos carrying one copy of *prd*-Prd (A), *prd*-*int*-Gsb (B), *prd*-PrdΔPRD (C), *prd*-Gsb+PRD (D), *prd*-GsbN+Pax3C (E), *prd*-GsbN+PrdC (F), *prd*-Pax3N+GsbC (G) or *prd*-PrdN+GsbC (H) are shown under dark-field illumination (anterior is up). Homozygous *prd*$^{2.45}$ embryos carrying two copies of the *prd*-*int*-Gsb, *prd*-GsbN+Pax3C, *prd*-Pax3N+GsbC or *prd*-PrdN+GsbC transgene exhibit fully rescued cuticles (not shown). *prd*$^{-}$ embryos carrying one or two copies of the transgenes were obtained as described in Materials and Methods. Embryos were collected and allowed to develop for 24 hours at 25°C before cuticle preparation.

To investigate if the N-terminal portion of Prd suffices to rescue male sterility, the *prd*-PrdN transgene was constructed by deleting the C-terminal portion of *prd*-Prd (Fig. 1). The cuticle and viability functions of *prd*-PrdN were examined in *prd* mutants, its male fertility function in *prd* mutants rescued by two copies of *prd*-GsbN+PrdC. Not surprisingly, *prd*-PrdN is unable to rescue any Prd functions (Figs 1, 4). On the contrary, *prd*-PrdN exerts a dominant-negative effect on *prd* since two copies of *prd*-PrdN are lethal and produce a pair-rule phenotype in the presence of only one wild-type *prd* allele (not shown), presumably because PrdN inhibits Prd function by competing with Prd for the same DNA target sites or by forming inactive heterodimers with Prd.

In summary, these results show that (i) the function of Prd in male fertility is determined by its N-terminal portion, which can bind specifically to DNA targets not recognized by Gsb and Pax3; (ii) the N-terminal portion of Prd is by itself not sufficient to perform any Prd functions, presumably because of the lack of an activation domain; (iii) the N-terminal portions of Prd, Gsb and Pax3 are functionally divergent despite their high sequence conservation; and (iv) the C-terminal portions of Prd and Gsb are qualitatively similar in functions even though their primary sequences have widely diverged.

DISCUSSION

The *Drosophila prd* gene is of interest because it encodes a multidomain transcription factor with multiple functions and, in form of its paralogs *gsb* and *gsbn* and its orthologs *Pax3* and *Pax7* (Fitch, 2000), plays a key role in both invertebrate and vertebrate development (Noll, 1993). Here we discuss the results of an extensive analysis aimed at elucidating the contributions of the various domains of the Prd protein to its in vivo functions. The *prd* gene is particularly well suited for such studies because (i) its entire *cis*-regulatory region has been identified (Gutjahr et al., 1994); (ii) its in vivo functions have been characterized (Nüsslein-Volhard and Wieschaus, 1980; Xue and Noll, 1996, 2000); (iii) various *prd* mutants are available (Nüsslein-Volhard and Wieschaus, 1980; Nüsslein-Volhard et al., 1984; Kilchherr et al., 1986); and (iv) the complete rescue of its in vivo functions has been achieved (Gutjahr et al., 1994).

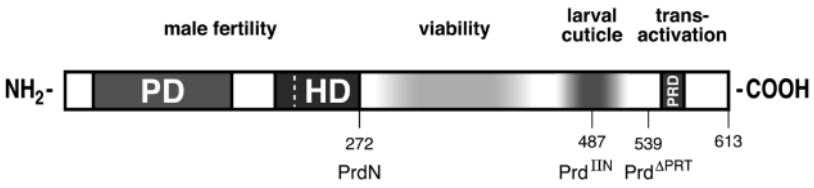

Fig. 7. Map of known domains and hypothetical motifs or domains of Prd specific for different functions during development. The functions of the Prd protein, including the paired-domain (PD), *prd*-type homeodomain (HD), and the PRD repeat (Frigerio et al., 1986; Bopp et al., 1986) as well as hypothetical domains or motifs responsible for the viability and larval cuticle functions, are specified above its linear map. The C-terminal ends of full-length Prd (residue 613) and the truncated proteins PrdN (272), Prd^{IIN} (487) and $Prd^{\Delta PRT}$ (539) are indicated. For details, see text.

In contrast to previous work performed in vitro (Treisman et al., 1991; Underhill et al., 1995; Jun and Desplan, 1996) or in cell culture (Han et al., 1989; Cai et al., 1994), the present study is aimed at establishing all wild-type in vivo functions of the Prd protein, a task that requires expression under the original control elements rather than by the artificial means of a heat-shock promoter (Miskiewicz et al., 1996) or incomplete *prd* enhancer elements (Bertuccioli et al., 1996).

With the aid of two evolutionary alleles of *prd*, *prd*-Gsb and *prd*-Pax3, in which the *gsb* and *Pax3* coding regions were placed under the control of the entire *prd cis*-regulatory region, it was shown previously that Prd activity is required in vivo during at least three distinct developmental stages to ensure proper segmentation of the larval cuticle, postembryonic viability and male fertility (Xue and Noll, 1996; Xue and Noll, 2000). Here, we constructed a series of *prd* transgenes which express various versions of the Prd protein, including truncations or chimeras of Prd, Gsb and Pax3 under the control of the complete *prd cis*-regulatory region. All transgenes were tested for their ability to rescue any of these Prd functions. Thus, this report is the first example of a complete functional analysis of the Prd protein under natural conditions.

Cooperativity between PD and HD

The presence of two DNA-binding domains, PD and HD, in Prd and some other members of the *Pax* gene family raises the question of whether the regulation of any of its target genes requires the binding of both or only one of its two DNA-binding domains. Both mechanisms are compatible with in vitro results (Treisman et al., 1991; Underhill et al., 1995; Jun and Desplan, 1996). Our in vivo studies show that both PD and HD are absolutely required for Prd function because deletion of either or both of these domains from the *prd*-Gsb transgene result in the complete loss of its ability to rescue the segment-polarity gene activation, cuticular phenotype and lethality of *prd* mutants. Moreover, since a point mutation in the PD, *prd*-GsbP17L, eliminates all Prd functions, the DNA-binding ability of the PD is necessary for the normal functions of Prd. An analogous mutation abolishes DNA binding of the human PAX5 protein (Czerny et al., 1993) and causes Waardenburg's syndrom I when present in PAX3 (Baldwin et al., 1992). Our observation that *prd*-GsbΔP and *prd*-GsbΔH cannot complement for any function of Prd implies that the PD and HD must be present in the same Prd molecule, presumably because each Prd function requires the recognition of at least one composite DNA target site.

In agreement with our findings, Prd proteins unable to bind DNA as a result of single amino acid substitutions in either the PD or HD can no longer activate the ectopic expression of Prd-target genes when expressed ubiquitously under the control of the heat-shock promoter (Miskiewicz et al., 1996) nor will they perform any Prd in vivo function when expressed under the control of some of its own enhancers (Bertuccioli et al., 1996). In addition, a composite Prd target site has been identified in the *even-skipped* enhancer whose mutation in either the PD or HD binding portion dramatically reduces Prd binding activity both in vitro and in vivo (Fujioka et al., 1996). Our finding that the PD and HD cannot complement in *trans* for any function of Prd agrees with some observations obtained with mutant transgenes in vivo (Bertuccioli et al., 1996), but contradicts results obtained in vitro (Jun and Desplan, 1996), and in vivo when the two Prd mutant proteins are expressed under heat-shock control (Miskiewicz et al., 1996). Taken together, these results imply that the PD and HD of Prd may interact with their DNA targets cooperatively and that this cooperativity can occur in *trans* only if the proteins are produced at concentrations much higher than those occurring naturally.

The role of the C-terminal PRD repeat of Prd

The PRD repeat, which encodes a 20-30 amino acid His-Pro repeat, was discovered in an attempt to verify predictions of the gene network hypothesis in a search for protein-coding domains of *prd* (Frigerio et al., 1986; Noll, 1993). It was found in a number of *Drosophila* early developmental genes, including *bicoid* (*bcd*) and *daughterless* (*da*) (Frigerio et al., 1986; Berleth et al., 1988; Cronmiller et al., 1988), but its in vivo function remained unknown. Previous experiments in cell culture systems showed that the PRD repeat is part of a transactivation domain (Han et al., 1989; Cai et al., 1994) that is necessary to drive ectopic expression of Prd-target genes under the control of ubiquitously expressed Prd (Cai et al., 1994). Other studies, however, suggested that the PRD repeat is not essential for in vivo functions of Prd (Bertuccioli et al., 1996).

Our results demonstrate that the Prd protein whose PRD repeat has been deleted in *prd*-PrdΔPRD is still able to perform all in vivo functions of Prd, which implies that the PRD repeat is not absolutely required for Prd function. However, the fact that one copy of *prd*-PrdΔPRD exhibits significantly reduced efficiency in its ability to rescue the lethality and male sterility of *prd* mutants indicates that the PRD repeat greatly facilitates these Prd functions. This conclusion is corroborated and extended by the results obtained with *prd*-Gsb+PRD transgenes, which demonstrate that the PRD repeat enhances the viability as well as the cuticle function of Prd. Thus, the PRD repeat is an important transactivation domain that facilitates all functions of Prd.

Conservation of functional motifs in the divergent C-terminal moieties of Prd, Gsb and Pax3

Previous work has demonstrated that Prd, Gsb and Pax3 proteins are, at least partially, functionally equivalent (Li and

Noll, 1994; Xue and Noll, 1996). When expressed under the control of the entire *cis*-regulatory region of *prd*, both Gsb and Pax3 can activate Prd-target genes necessary for the generation of wild-type cuticle, while Gsb is able to rescue *prd* mutants to adulthood. These results strongly suggested that the acquisition of *cis*-regulatory regions rather than the divergence of their coding regions is the primary evolutionary mechanism responsible for the functional diversification of *prd, gsb* and *Pax3* genes. However, although Gsb and Pax3 can substitute for most Prd functions, they do so at considerably reduced efficiencies, which indicates that these proteins had to adapt their new functions for optimal performance by subsequent mutations producing the observed divergence of the Prd, Gsb and Pax3 proteins. Here, we have studied the result of this process of adaptation by examining the functional differences between these proteins when expressed as evolutionary alleles under the same *cis*-regulatory region.

Our results lead us to postulate that, in addition to the PRD repeat, two motifs or domains are present in the C-terminal portion of Prd on whose functions the formation of wild-type larval cuticle and survival to adulthood depend. Although no significant similarity has been found among the primary sequences of the C-terminal moieties of Prd, Gsb and Pax3, the motif required for implementing wild-type cuticle is shared by all three proteins. In contrast, the motif necessary for Prd's viability function is retained only in Gsb, presumably as secondary or tertiary protein structure. It should be stressed that at least two independent functions of Prd are required for viability, one of which Pax3 is able to perform even better than Gsb (Xue and Noll, 1996). However, Pax3 is unable to substitute for one of the viablity functions of Prd and even exerts a dominant-negative effect on it (Xue and Noll, 1996). In agreement with our postulate, combining our results with those obtained with two weak *prd* alleles encoding truncated Prd proteins (Tearle and Nüsslein-Volhard, 1987; Bertuccioli et al., 1996), allowed us to map the motifs for the cuticle and viability functions within the C terminus of Prd.

The male fertility function of Prd

Although *prd*-Gsb, an evolutionary allele of *prd*, rescues *prd* mutants to viable adults, all males are sterile. Since wild-type males transgenic for two copies of *prd*-Gsb are fertile, we conclude that *prd* has a function required for male fertility. Moreover, as *prd*-Gsb includes the entire *cis*-regulatory region of *prd* (Xue and Noll, 1996), its failure to rescue male fertility must be caused by the inability of Gsb to replace this function of the Prd protein. Since Prd and Gsb share a highly conserved N-terminal portion consisting of two DNA-binding domains, the PD and HD, it seemed plausible to map this functional difference to their divergent C termini. Surprisingly, however, the protein-domain-swapping experiments indicate that the conserved N-terminal rather than the divergent C-terminal portion is the determinant for this particular function of Prd. Therefore, we suggest that at least one specific Prd target site, recognized by Prd but not Gsb, is involved in male fertility.

The male fertility function of Prd is controlled by a specific *prd* enhancer uncovered in *prd* mutants by a *prd* rescue construct that lacks 5 kb of the downstream regulatory region (Bertuccioli et al., 1996). Consistent with this interpretation, a *prd* transgene that expresses Prd merely under the control of this 5 kb regulatory region is able to confer fertility to *prd*-Gsb males mutant for *prd* (L. X. and M. N., unpublished). Males completely deficient for this fertility function of *prd* have no accessory glands, while accessory glands begin to form in *prd* mutant males rescued by *prd*-Gsb, but stop development at a severely reduced size (Bertuccioli et al., 1996; Xue and Noll, 2000). These findings are in agreement with our hypothesis that new functions evolve primarily through the acquisition of new enhancers during gene duplication (Li and Noll, 1994; Xue and Noll, 1996) and that the adaptation of the protein is secondary and a necessary consequence of its expression in the newly acquired context of this function.

Our results further imply that the C-terminal portions of Prd and Gsb, though divergent in their primary sequences, are still qualitatively the same. Hence, we questioned the validity of amino acid similarity as a general measure of functional equivalence in homologous proteins (Xue and Noll, 1996). Instead, Yockey (1992) proposed to replace this measure of functional equivalence by calculations of the mutual entropy between two protein sequences, a more precise statistical measure that takes into account the probability by which certain amino acids are replaced by others.

We are indebted to T. Kornberg for anti-En monoclonal antibodies, R. Nusse for anti-Wg antiserum, E. Hafen for the pUChspΔ2-3 P-element helper plasmid, P. Gruss for a *Pax3*-cDNA, M. Weir for the PrdΔPR plasmid and C. Nüsslein-Volhard for the *prd* mutant stocks. We thank Fritz Ochsenbein for expert art work, E. Frei for stimulating discussions and J. Alcedo, E. Frei and H. Noll for critical comments on the manuscript. This work was supported by Grant 31-40874.94 from the Swiss National Science Foundation and by the Kanton Zürich.

REFERENCES

Baldwin, C. T., Hoth, C. F., Amos, J. A., Da-Silva, E. O. and Milunsky, A. (1992). An exonic mutation in the *HuP2* paired domain gene causes Waardenburg's syndrome. *Nature* **355**, 637-638.

Baumgartner, S., Bopp, D., Burri, M. and Noll, M. (1987). Structure of two genes at the *gooseberry* locus related to the *paired* gene and their spatial expression during *Drosophila* embryogenesis. *Genes Dev.* **1**, 1247-1267.

Baumgartner, S. and Noll, M. (1990). Network of interactions among pair-rule genes regulating *paired* expression during primordial segmentation of *Drosophila*. *Mech. Dev.* **33**, 1-18.

Berleth, T., Burri, M., Thoma, G., Bopp, D., Richstein, S., Frigerio, G., Noll, M. and Nüsslein-Volhard, C. (1988). The role of localization of *bicoid* RNA in organizing the anterior pattern of the *Drosophila* embryo. *EMBO J.* **7**, 1749-1756.

Bertuccioli, C., Fasano, L., Jun, S., Wang, S., Sheng, G. and Desplan, C. (1996). In vivo requirement for the paired domain and homeodomain of the *paired* segmentation gene product. *Development* **122**, 2673-2685.

Bopp, D., Burri, M., Baumgartner, S., Frigerio, G. and Noll, M. (1986). Conservation of a large protein domain in the segmentation gene *paired* and in functionally related genes of *Drosophila*. *Cell* **47**, 1033-1040.

Bopp, D., Jamet, E., Baumgartner, S., Burri, M. and Noll, M. (1989). Isolation of two tissue-specific *Drosophila* paired box genes, *pox meso* and *pox neuro. EMBO J.* **8**, 3447-3457.

Cai, J., Lan, Y., Appel, L. F. and Weir, M. (1994). Dissection of the Drosophila Paired protein: Functional requirements for conserved motifs. *Mech. Dev.* **47**, 139-150.

Cronmiller, C., Schedl, P. and Cline, T. W. (1988). Molecular characterization of *daughterless*, a *Drosophila* sex determination gene with multiple roles in development. *Genes Dev.* **2**, 1666-1676.

Czerny, T., Schaffner, G. and Busslinger, M. (1993). DNA sequence recognition by Pax proteins: bipartite structure of the paired domain and its binding site. *Genes Dev.* **7**, 2048-2061.

Dambly-Chaudière, C., Jamet, E., Burri, M., Bopp, D., Basler, K., Hafen, E., Dumont, N., Spielmann, P., Ghysen, A. and Noll, M. (1992). The

paired box gene *pox neuro*: a determinant of poly-innervated sense organs in *Drosophila*. *Cell* **69**, 159-172.

DiNardo, S. and O'Farrell, P. H. (1987). Establishment and refinement of segmental pattern in the *Drosophila* embryo: spatial control of *engrailed* expression by pair-rule genes. *Genes Dev*. **1**, 1212-1225.

Driever, W., Thoma, G. and Nüsslein-Volhard, C. (1989). Determination of spatial domains of zygotic gene expression in the *Drosophila* embryo by the affinity of binding sites for the bicoid morphogen. *Nature* **340**, 363-367.

Fitch, W. M. (2000). Homology, a personal view on some of the problems. *Trends Genet.* **16**, 227-231.

Frigerio, G., Burri, M., Bopp, D., Baumgartner, S. and Noll, M. (1986). Structure of the segmentation gene *paired* and the *Drosophila* PRD gene set as part of a gene network. *Cell* **47**, 735-746.

Fujioka, M., Miskiewicz, P., Raj, L., Gulledge, A. A., Weir, M. and Goto, T. (1996). *Drosophila* Paired regulates late *even-skipped* expression through a composite binding site for the paired domain and the homeodomain. *Development* **122**, 2697-2707.

Goulding, M. D., Chalepakis, G., Deutsch, U., Erselius, J. R. and Gruss, P. (1991). Pax-3, a novel murine DNA binding protein expressed during early neurogenesis. *EMBO J*. **10**, 1135-1147.

Gutjahr, T., Frei, E. and Noll, M. (1993a). Complex regulation of early *paired* expression: intial activation by gap genes and pattern modulation by pair-rule genes. *Development* **117**, 609-623.

Gutjahr, T., Patel, N. H., Li, X., Goodman, C. S. and Noll, M. (1993b). Analysis of the *gooseberry* locus in *Drosophila* embryos: *gooseberry* determines the cuticular pattern and activates *gooseberry neuro*. *Development* **118**, 21-31.

Gutjahr, T., Vanario-Alonso, C. E., Pick, L. and Noll, M. (1994). Multiple regulatory elements direct the complex expression pattern of the *Drosophila* segmentation gene *paired*. *Mech. Dev*. **48**, 119-128.

Han, K., Levine, M. S. and Manley, J. L. (1989). Synergistic activation and repression of transcription by Drosophila homeobox proteins. *Cell* **56**, 573-583.

Ingham, P. W., Baker, N. E. and Martinez-Arias, A. (1988). Regulation of segment polarity genes in the *Drosophila* blastoderm by *fushi tarazu* and *even-skipped*. *Nature* **331**, 73-75.

Ingham, P. W. and Martinez Arias, A. (1992). Boundaries and fields in early embryos. *Cell* **68**, 221-235.

Jun, S. and Desplan, C. (1996). Cooperative interactions between paired domain and homeodomain. *Development* **122**, 2639-2650.

Kilchherr, F., Baumgartner, S., Bopp, D., Frei, E. and Noll, M. (1986). Isolation of the *paired* gene of *Drosophila* and its spatial expression during early embryogenesis. *Nature* **321**, 493-499.

Lawrence, P. A., Johnston, P., Macdonald, P. and Struhl, G. (1987). Borders of parasegments in *Drosophila* embryos are delimited by the *fushi tarazu* and *even-skipped* genes. *Nature* **328**, 440-442.

Li, X., Gutjahr, T. and Noll, M. (1993). Separable regulatory elements mediate the establishment and maintenance of cell states by the *Drosophila* segment-polarity gene *gooseberry*. *EMBO J.* **12**, 1427-1436.

Li, X. and Noll, M. (1994). Evolution of distinct developmental functions of three *Drosophila* genes by acquisition of different cis-regulatory regions. *Nature* **367**, 83-87.

Miskiewicz, P., Morrissey, D., Lan, Y., Raj, L., Kessler, S., Fujioka, M., Goto, T. and Weir, M. (1996). Both the paired domain and homeodomain are required for in vivo function of *Drosophila* Paired. *Development* **122**, 2709-2718.

Noll, M. (1993). Evolution and role of *Pax* genes. *Curr. Opin. Genet. Dev*. **3**, 595-605.

Nüsslein-Volhard, C. and Wieschaus, E. (1980). Mutations affecting segment number and polarity in *Drosophila*. *Nature* **287**, 795-801.

Nüsslein-Vohard, C., Wieschaus, E. and Kluding, H. (1984). Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster*. I. Zygotic loci on the second chromosome. *Roux's Arch. Dev. Biol*. **193**, 267-282.

Peifer, M. and Bejsovec, A. (1992). Knowing your neighbors: Cell interactions determine intrasegmental patterning in *Drosophila*. *Trends Genet*. **8**, 243-249.

Rubin, G. M. and Spradling, A. C. (1982). Genetic transformation of *Drosophila* with transposable element vectors. *Science* **218**, 348-353.

Small, S. and Levine, M. (1991). The initiation of pair-rule stripes in the *Drosophila* blastoderm. *Curr. Opin. Genet. Dev*. **1**, 255-260.

St Johnston, D. and Nüsslein-Volhard, C. (1992). The origin of pattern and polarity in the *Drosophila* embryo. *Cell* **68**, 201-219.

Studier, F. W. and Moffat, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol*. **198**, 113-130.

Tassabehji, M., Read, A. P., Newton, V. E., Harris, R., Balling, R., Gruss, P. and Strachan, T. (1992). Waardenburg's syndrome patients have mutations in the human homologue of the *Pax-3* paired box gene. *Nature* **355**, 635-636.

Tearle, R. and Nüsslein-Volhard, C. (1987). Tübingen mutants and stocklist. *Dros. Inform. Serv*. **66**, 209-269.

Treisman, J., Harris, E. and Desplan, C. (1991). The paired box encodes a second DNA-binding domain in the *paired* homeo domain protein. *Genes Dev*. **5**, 594-604.

Underhill, D. A., Vogan, K. J. and Gros, P. (1995). Analysis of the mouse Splotch-delayed mutation indicates that the Pax-3 paired domain can influence homeodomain DNA-binding activity. *Proc. Natl. Acad. Sci. USA* **92**, 3692-3696.

Wieschaus, E. and Nüsslein-Volhard, C. (1986). Looking at embryos. In *Drosophila, A Practical Approach* (ed. D. B. Roberts), pp. 199-227. Oxford, UK: IRL Press.

Xue, L. and Noll, M. (1996). The functional conservation of proteins in evolutionary alleles and the dominant role of enhancers in evolution. *EMBO J*. **15**, 3722-3731.

Xue, L. and Noll, M. (2000). *Drosophila* female sexual behavior induced by sterile males showing copulation complementation. *Proc. Natl. Acad. Sci. USA* **97**, 3272-3275.

Yockey, H. P. (1992). *Information Theory and Molecular Biology*. Cambridge, UK: Cambridge University Press.