Overlapping genes of *Drosophila melanogaster*: organization of the z600–gonadal–Eip28/29 gene cluster

Robert A. Schulz and Barbara A. Butler

Department of Biochemistry and Molecular Biology, The University of Texas M.D. Anderson Cancer Center, Houston, Texas 77030 USA

The *Drosophila* gonadal (gd) gene is a member of a gene cluster that maps cytogenetically to the 71CD interval of chromosome 3. The gene is bordered distally by z600, a gene expressed predominantly during early embryogenesis, and proximally by Eip28/29, a gene regulated by ecdysone in *Drosophila* cell lines. gd can be expressed in either of two modes in adults: gdP expression leads to the transcripts 1300 and 1000, which are found in the ovaries, whereas gdlM expression leads to the transcripts 1500 and 1200, which are found in the testes. In situ hybridization analysis reveals that this expression occurs in the germ line during oogenesis and spermatogenesis. Structural studies identify an unusual gd sequence organization. The ovarian and testes transcripts differ at their 5' ends because of the utilization of different transcription initiation sites. This result indicates that alternative promoter usage is responsible for sex-specific gd expression. Within an expression mode, the two transcripts differ at their 3' ends as a result of multiple polyadenylation site usage; one of these sites resides within the 5' exon of the Eip28/29 gene. gd is overlapped by z600 as well because the z600 transcript is polyadenylated at position +91 of the gdl transcripts. An analysis of germ-line transformants reveals that gd can be expressed properly outside the overlapping gene environment because a 1.8-kb DNA region contains all the sequences necessary for gd sex-specific expression.

[Key Words: gonadal; alternative promoters; sex-specific expression; *Drosophila* germ line; overlapping genes; z600; Eip28/29]

Received October 27, 1988, revised version accepted December 23, 1988.

The Eip28/29 gene is regulated by the steroid hormone ecdysone in *Drosophila* cell lines (Savakis et al. 1984; Bieber 1986; Cherbas et al. 1986). As a step toward understanding this gene's transcriptional regulation in flies, we initiated an analysis of Eip28/29 gene expression throughout *Drosophila* development. These studies have revealed that the DNA flanking the gene is transcriptionally complex, with at least seven distinct transcripts being derived from transcription units that map entirely or in part to 2 kb of DNA (Schulz et al. 1989). On the basis of preliminary information on the structures of the RNAs and their differential accumulation during development, we have proposed that the transcripts are derived from three clustered genes: z600, gonadal (gd) and Eip28/29.

The z600 gene has been named in accordance with the zygotic appearance of a single 600-nucleotide gene transcript (Schulz et al. 1989). The gene is expressed abundantly during early embryogenesis and then again in adults at a much lower level. z600 is located upstream of gd and is transcribed in the same direction as gd and Eip28/29. Four transcripts are derived from the gd gene; these transcripts accumulate in pairs during development due to two modes of gd gene expression. gdP expression leads to a 1300-1000-nucleotide RNA pair present solely in embryos and adult female ovaries.

[These transcripts were identified previously as a 1200P/900-nucleotide pair (Schulz et al. 1989); however, results presented in this report indicate that both RNAs are ~100 nucleotides larger in size.] gdlM expression leads to a 1500-1200-nucleotide RNA pair that first becomes detectable during late larval development, accumulates to higher levels in pupae, and becomes extremely abundant in adult males. In adults there is a clear sex-specific expression of the gene because the gdP and gdlM transcripts accumulate exclusively in the ovaries and testes, respectively. Within an expression mode, the two transcripts differ at their 3' ends due to multiple polyadenylation site usage. S1 protection experiments have demonstrated that the 1300- and 1500-nucleotide transcripts extend through Eip28/29 flanking sequences into the first exon of Eip28/29, using a polyadenylation site located 2 bases before the initiating methionine codon of ecdysone-induced polypeptides (EIPs) 28 and 29 (Cherbas et al. 1986). Eip28/29 gene expression is quite different from that of the z600 or gd genes in that gene transcripts accumulate throughout development, with the exception of the one pupal period tested. Thus, the three genes show very different patterns of expression regardless of their organization in a densely packed cluster and the demonstrated overlap of the gd and Eip28/29 gene sequences.
An understanding of the diverse expression of these genes will require detailed information on individual gene structures and the mechanisms responsible for stage- and tissue-specific transcript accumulation. In this report we determine the structures of the z600 and gdl genes and initiate a preliminary analysis of gdl gene expression. For gdl, two possibilities exist for the distinct modes of gene expression: either the generation of gdlF and gdlM RNAs by transcription from alternative promoters or by alternative splicing of a common primary gene transcript. Our current studies identify several interesting characteristics of the gdl gene: (1) Alternative promoters are responsible for the sex-specific expression of the gene; (2) gdl expression occurs in the germ line during oogenesis and spermatogenesis; (3) gdl is overlapped by both z600 and Eip28/29 gene sequences; (4) gdl can be properly expressed outside of its normal overlapping gene environment.

**Results**

**Germ-line expression of the gdl gene**

The sex-specific expression of gdl in adults is evident in two distinct gene transcription modes [Schulz et al. 1989]. For convenience, these have been designated gdlF for the female mode of expression and gdlM for the male mode of expression. Figure 1 illustrates that gdlF expression leads to the transcripts 1300 and 1000, which are found in the ovaries, whereas gdlM expression leads to the transcripts 1500 and 1200, which are found in the testes. An analysis of gdl gene expression in the mutants tudor (tud) and transformer (tra) demonstrates that the specific pattern of transcript accumulation is dependent on a functional germ line being present and the specific type of gametogenesis that occurs in the animal. tud is a grandchildless mutant; progeny of homozygous tud females lack pole cells, the precursors of the germ line, but will form the normal somatic tissues of the gonads [Boswell and Mahowald 1985]. gdlF and gdlM transcripts are present in the appropriate gonads of animals containing a germ line (lanes 5 and 7) but are not detected in the gonads of animals lacking a germ line (lanes 6 and 8). The tra mutation transforms female somatic tissue into male somatic tissue in homozygous XX animals [Baker and Ridge 1980]. These pseudomales are sterile because their rudimentary testes contain no sperm, the mutation is soma specific and does not cause germ-line transformation. gdlM transcripts are not detected in the gonads of these somatically transformed animals (lane 2), suggesting that gdl is not expressed in somatic cells of the testes. In addition, gdlF expression is not observed, potentially as a result of the stunted development of the female germ line in this mutant. Proper gdl expression is observed in the tra genetic backgrounds which result in phenotypically normal gonads (lanes 1, 3, and 4).

This analysis of gdl expression suggests that the gene is expressed in the Drosophila germ line. To address the provenance of gdl gene expression in the gonads directly, we initiated in situ hybridization to RNAs in tissue sections. As demonstrated in Figure 2, gdlF and gdlM transcripts accumulate in the germ line during oogenesis and spermatogenesis, respectively. In female abdomens [Fig. 2A], grains are detected in egg chamber nurse cells and oocytes. In male abdomens [Fig. 2B], grains are detected in the lumen of the testes.

**Characterization of overlapping z600 and gdlM cDNAs**

The different modes of gdl gene expression in the germ line could be a result of transcription from alternative promoters or alternative splicing of a common primary gene transcript. A detailed structural analysis of the gdl gene was undertaken to distinguish between these two possibilities. Due to the close proximity of the z600 gene in the 5' flank of gdl, we decided to characterize this small upstream gene as well. Initially, gdlM and z600 cDNAs were isolated from adult male and from 0- to 5-hr embryonic cDNA libraries, respectively. The cDNAs were characterized by restriction enzyme mapping and DNA sequence analysis; an additional 1.6 kb of genomic sequence from coordinates -4001 to -2394 was determined as well. Based on these results and previously reported DNA sequence analysis [Cherbas et al. 1986], the z600 and gdlM cDNAs were aligned on the genomic DNA map, as illustrated in Figure 3. Both gdl cDNAs initiate at position -3401 of the genomic se-
**Figure 2.** In situ hybridization of a $^{35}$S-labeled $gdl$ cRNA probe to RNA in tissue sections of a $ry^{+5}$ female abdomen (A) and male abdomen (B). Grains accumulate predominantly in nurse cells (N) and oocytes (O) in the ovary and in the lumen of the testes (T).

**Figure 3.** Alignment of the $z600$ and $gdl$ cDNAs with the genomic DNA sequence. Boxed areas represent exon sequences, shaded regions indicate their location on the genomic sequence. The $z600$ cDNA is a nearly full-length cDNA, missing sequence from the 5' end of the transcript. The 1200 and 1500 $gdl$ cDNAs represent full-length copies of the $gdl$ transcripts. The overlap of the $z600$ and $gdl$ cDNAs is indicated by the crossing of shaded areas at their 3' and 5' ends, respectively, the overlap of the 1500 cDNA with the $Eip28/29$ transcripts is indicated by the extension of the $gdl$ 3' exon into the 5' exon of the $Eip28/29$ gene. Certain restriction enzyme cleavage sites are given and abbreviated as follows: [B] BamHI, [B] BclI, [E] EcoRI, [H] HindIII, [K] KpnI, [P] PstI, [S] SalI, and [V] PvuII. Data for the restriction map of the genomic sequence are taken from Cherbas et al. (1986); all restriction sites were confirmed by DNA sequence reported in this work (see Fig. 7).
quence, end in a short poly(A) sequence, and are contiguous up to the 3' end of the short cDNA. The length of the cDNAs and the sites of poly(A) addition demonstrate that the short and long cDNAs represent 1200 and 1500 cDNAs, respectively. The cDNAs are 979 and 1254 bp long with 3' ends at positions -1957 and -1682, consistent with the predicted polyadenylation sites of the 1200- and 1500-nucleotide RNAs based on sequence analysis and nuclease S1 protection experiments (Schulz et al. 1989). The latter site resides within the δ exon of the Eip28/29 gene, further demonstrating the overlap of the gdl 1500-nucleotide transcript with the Eip28/29 gene. The z600 cDNA is 424 bp long and ends in a poly(A) sequence. This analysis revealed, quite unexpectedly, that the gdlM transcripts overlap with the z600 gene as well, because the z600 transcript is polyadenylated at position +91 of both the 1200 and 1500 cDNAs.

z600 and gdl transcript analysis

The z600 and gdl transcripts were further characterized by primer extension and nuclease S1 protection analyses. Figure 4 shows the approach used to determine the 5' end of the z600 transcript. A 48-nucleotide HindIII–SalI genomic DNA primer from position -3744 to -3696 was 5'-end-labeled at the SalI site, hybridized with total RNA from 0- to 4-hr embryos, and then extended with reverse transcriptase. A 73-nucleotide extended fragment was observed following sizing on a denaturing gel and autoradiography (lane 2). In parallel, a 305-bp EcoRI–SalI genomic DNA fragment from position -4001 to -3696 was 5'-end-labeled at the same SalI site and hybridized to embryonic RNA. Following nuclease S1 treatment, protected bands of 73–75 nucleotides were observed (lane 3). These results demonstrate that the 5' end of the z600 RNA maps to approximately -3769 of the genomic sequence. This site is 35 bp upstream of the 5' end of the z600 cDNA, which shows that the z600 gene is 459 bp in length.

Although both gdlM cDNAs initiated at position -3401 of the genomic sequence, primer extension analysis was still required to demonstrate that this position corresponded to the 5' end of the gdlM transcripts. A synthetic primer identical to -3211 to -3195 of the genomic sequence (primer 1) was 5'-end-labeled, hybridized with total RNA from adult testes, and extended with reverse transcriptase. A single extended fragment of -207 nucleotides was observed, as shown in Figure 5 (lane 1). This size is consistent with the gdlM transcripts initiating at genomic position -3401 and demonstrates that the gdlM clones represent full-length cDNAs.

Figure 5 also shows the results of experiments designed to map the 5' end[s] of the gdlP transcripts. RNAs from adult ovaries and the 11-5-4 cell line were used in these experiments. The choice of ovaries is an obvious basis of Northern analysis, gdl is expressed at high levels in this line in the gdlP expression mode, resulting in an enriched source of gdlP transcripts. (R.A. Schulz, unpubl.) The right half of Figure 5 shows an initial nuclease S1 protection analysis of the gdlP transcripts. A 650-bp EcoRI–BamHI gdlM cDNA fragment was 5'-end-labeled at the BamHI site and hybridized with either total RNA from adult ovaries (lane 4) or 11-5-4 cells (lane 5). Multiple S1-protected bands were observed, ranging from 486 to 436 nucleotides. These results show that gdlP RNAs are homologous with the gdlM cDNA from the BamHI site 5' to position -3237 at the longest extent and to -3187 at the shortest extent. The different protected fragments probably represent multiple gdlP transcription initiation sites, as shown by primer extension analysis (lane 2). That is, multiple extended fragments are observed with the longest band of 160 nucleotides corresponding in size to the longest S1-pro-
Schulz and Butler

tected fragment. The observed heterogeneity may be a result of incomplete extension of the primer as well. Thus, although it appears that the gdIM transcripts initiate at a single site, there are multiple 5' ends for the gdIP RNAs with the longest transcripts initiating at genomic position -3237, 165 bp downstream of the gdIM initiation site.

Figure 5. Determination of the 5' ends of the gdIM and gdIP transcripts. The upper diagrams illustrate the experimental design for mapping the starts of the gd transcripts using primer extension and nuclease S1 analyses. In the primer extension experiment, a 17-nucleotide synthetic primer (primer 1) specific to the gdIM transcripts was hybridized with 10 μg of total RNA from testes and extended with reverse transcriptase (lane 1). Additionally, a 15-nucleotide synthetic primer (primer 2) homologous to the gdIM and gdIP transcripts was hybridized with 10 μg of total RNA from 11-5-4 cells and extended with reverse transcriptase (lane 2). A single extended fragment of 207 nucleotides was observed with testes RNA, whereas multiple extended fragments were observed with 11-5-4 cell RNA. The longest extended fragment from both of these experiments is diagramed at the top. In the nuclease S1 protection experiment, the cDNA probe was hybridized either with no RNA (lane 3), 10 μg of total RNA from ovaries (lane 4), or 10 μg of total RNA from 11-5-4 cells (lane 5) and then treated with nuclease S1. Multiple nuclease S1-resistant bands are observed with both ovary and 11-5-4 cell RNA; the longest protected fragment is diagramed at the top. Sizes of both the primer-extended and nuclease S1-resistant fragments were determined based on migration relative to end-labeled DNA size markers (lanes M).

Having identified structural differences at the 5' ends of the gdIM and gdIP transcripts, additional S1 protection experiments were undertaken to further characterize the gdIP RNAs (Fig. 6). In these experiments, 32P-labeled cRNAs were synthesized using the gdIP cDNAs as templates. The probes were then hybridized to total RNA from adult testes, adult ovaries, or 11-5-4 cells and then subjected to nuclease S1 treatment. The gdI sequence of the 1200 cRNA probe is protected in its entirety by testes RNA (lane 2), whereas ovarian RNA (lane 3) and 11-5-4 cell RNA (lane 4) protect ~825 nucleotides of the probe. This result is consistent with protection by gdIP RNAs that are identical in sequence to the 1200-nucleotide RNA, with the exception of the differences at the 5' ends. The use of the 1500 cRNA probe demonstrates the multiple 3' ends of both the gdIM and gdIP transcripts. Testes RNA generates S1-resistant fragments of ~1270 and 980 nucleotides, which correspond to protection of the probe by the 1500- and 1200-nucleotide RNAs, re-

Figure 6. gdIP transcript mapping by nuclease S1 protection analysis using gdIM cRNA probes. The 1200 cRNA probe was 1263 nucleotides long and contained 979 nucleotides of gdI sequence, a 25 nucleotide stretch of poly(A) sequence, and 259 nucleotides of vector sequence. Due to the presence of poly(A) sequence at the end of mRNAs, complete protection of this probe by the corresponding gdIM transcript should yield a nuclease S1-resistant fragment of 1004 nucleotides. The 1200 cRNA probe was hybridized with no RNA (lane 1), 10 μg of total RNA from testes (lane 2), ovaries (lane 3), and 11-5-4 cells (lane 4), followed by treatment with nuclease S1. In a parallel experiment, the 1500 cRNA probe was hybridized with no RNA (lane 5), 10 μg of total RNA from testes (lane 6), ovaries (lane 7), and 11-5-4 cells (lane 8), followed by treatment with nuclease S1. This probe contains 1254 nucleotides of gdI sequence, a 22-nucleotide stretch of poly(A) sequence, and 55 nucleotides of vector sequence; full protection of the gdI sequence would result in an S1-resistant band of ~1276 nucleotides.
ne of the overlapping genes and their transcripts

The sequence of the z600 and gd genes is presented in Figure 7, along with selected highlights of the different gene transcripts. Transcript initiation start sites are listed as +1 and are located at positions -3769 for z600, -3401 for gdM, -3337 for the 5' most gdL site, and -1743 for Eip28/29. Drosophila cap site consensus sequences, composed of 13 bp and often observed around transcription initiation sites, are located at positions -3311 for z600, -1957 for gdL (short 3' exon), and -1682 for gdM (long 3' exon). Polyadenylation signals are observed 15-19 nucleotides upstream of the polyadenylation sites in all three cases. The predicted amino acid sequences of putative z600 and gdL gene products are given under the DNA sequence.

Combining the z600 and gdL sequence and RNA structure analysis with that reported previously for the neighboring Eip28/29 gene [Cherbas et al. 1986], we have generated a schematic of the overlapping genes and transcription units in polytene region 71CD (Fig. 8). Seven distinct transcripts are derived from the three overlapping genes. The z600 gene is 459 bp in length and contains an open reading frame (ORF) of 90 codons. The z600 transcript is polyadenylated at a site within the first exon of the gdL gene and shares 91 nucleotides with the gdM transcripts. The gdL gene is 1720 bp in length, being composed of three exons of 453, 295, and 506 bp separated by introns of 56 and 410 bp. Four transcripts are derived from gdL, due to the utilization of different transcription initiation sites and multiple polyadenylation sites. The 1200- and 1000-nucleotide RNAs are derived from gd/, due to the utilization of different transcription initiation sites and multiple polyadenylation sites. The 1200- and 1000-nucleotide RNAs are designated as +1 and are located at positions -3769 for z600, -1957 for gdL (short 3' exon), and -1682 for gdM (long 3' exon). Polyadenylation signals are observed 15-19 nucleotides upstream of the polyadenylation sites in all three cases. The predicted amino acid sequences of putative z600 and gdL gene products are given under the DNA sequence.

Figure 7. Nucleotide sequence of the z600 and gonadal genes. The sequence is a composite of z600 and gdL cDNA and genomic DNA (coordinates -4001 to -2394) sequence determined in this study and gdL genomic DNA sequence (coordinates -3398 to -3337), 65 bp downstream of the 5' most gdL initiation site (3173 to -3162), and +1 for Eip28/29 (1746 to -1734). Polyadenylation sites are located at positions -3311 for z600, -1957 for gdL (short 3' exon), and -1682 for gdM (long 3' exon). Putative polyadenylation signals are observed 15-19 nucleotides upstream of the polyadenylation sites in all three cases. The predicted amino acid sequences of putative z600 and gdL gene products are given under the DNA sequence.
1500- and 1300-nucleotide RNAs are identical, except for the differences at the 5' ends. These differences have an effect on the potential coding capacity of the gene. A common ORF of 193 codons [ORF193] is present in all four transcripts, a consequence of the additional sequences at the 5' end of the gdlM transcripts is the presence of an additional ORF of 39 codons [ORF39]. The selection of different polyadenylation sites at the 3' ends of the transcripts has no effect on coding capacity but results in the 1500- and 1300-nucleotide transcripts sharing 62 nucleotides with the Eip28 and Eip29 transcripts. As reported previously, the latter two transcripts are derived from a primary Eip28/29 gene transcript by alternative splicing between the α and β exons [Schulz et al. 1986]. In all instances, the overlapping regions of the transcripts do not involve coding sequences but do include shared 3'-noncoding, 5'-flanking, and 5'-non-coding sequences. The comparison of z600 and gdl sequences with those of known genes present in sequence data banks has revealed no extensive homologies.

Expression of a truncated gdl gene in germ-line transformants

As a first step in the analysis of the regulation of gdl expression, we introduced a truncated gene into the Drosophila germ line and monitored transformants for normal gdl expression. The gdl sequences included a 1.8-kb BclI fragment (−3729 to −1958) containing 329/493 bp of 5'-flanking sequences relative to the gdlM/ gdlF transcription start sites and 1445 bp of genic sequences, including the first polyadenylation signal. This fragment was cloned into the CaSpER transformation vector, and the transposon was introduced into w− embryos, with eventual transformants identified by w+ selection. Nine independent transformants were obtained and assayed for gdl expression by Northern blot analysis. Because of the truncation just after the first polyadenylation signal, gdl expression could be monitored by the increase in the levels of the gdlM 1200-nucleotide and gdlF 1000-nucleotide RNAs. Figure 9 demonstrates the proper expression of the truncated gene in the B5 transformant line. The endogenous gene transcripts are observed in RNA from y w adult males (lane 2) and females (lane 4). A clear increase in the levels of 1200- and 1000-nucleotide RNAs are observed in B5 adult males (lane 1) and females (lane 3). A similar analysis shows that the expression of the introduced gdl gene occurs in the gonads of the transformant animals, based on the comparison of the 1200-nucleotide RNA levels in B5 (lane 5) and y w (lane 6) testes and 1000-nucleotide RNA levels in B5 (lane 8) and y w (lane 7) ovaries. These results are even more convincing when considering that the experiment is internally controlled due to the presence of the endogenous gdl transcripts in the transformant RNA samples. Overall, these results show that the 1.8-kb region is sufficient to direct proper sex- and tissue-specific expression of the gdl gene.

Figure 8. Schematic of the overlapping z600, gdl, and Eip28/29 genes and their transcripts. (Top) The genomic organization of the overlapping genes; (Bottom) different gene transcripts are diagramed. Exon sequences are diagramed as boxes and intron sequences as thin lines. Nonshaded regions in the exons correspond to 5' and 3'-noncoding sequences, whereas the shaded areas represent predicted coding regions. In the case of the gdlM transcripts, the small ORF [ORF39] specific to the gdlM transcripts is indicated by the crosshatched area. ORF193 common to the gdlM and gdlF transcripts is indicated by the shaded area. Only the longest gdlF transcripts are illustrated in the schematic.
noted that overlapping genes have been observed previously in eukaryotes (Schulz et al. 1989), the z600–gd–
Eip28/29 cluster clearly represents an unusual organiza-
tion of eukaryotic sequences.

Despite their presence in a densely packed cluster, the three genes exhibit quite diverse developmental, tissue, and cellular expression characteristics (Schulz et al. 1989). Abundant z600 gene expression is temporally re-
stricted to early embryogenesis; this expression is re-
stricted spatially as well, because gene transcripts are lo-
calized to dorsal and posterior regions at cellular blas-
stoderm (R.A. Schulz, J.L. Miksch, B. Sutton, and X. Xie, in prep.). gdl transcripts are present in embryos, are absent throughout most of development, and then reappear in adult females, specifically in developing egg chambers of the ovaries (see Fig. 2A). The gdl transcripts appear to be contributed maternally to the embryo, because the transcript titer is highest in unfertilized eggs and diminishes eventually in early embryonic stages (R.A. Schulz, unpubl.). However, these studies have not ruled out the possibility of low-level gdl expression during embryo-
genesis. gd transcripts are first detected in late larval development, accumulate to higher levels in pupae, and then become very abundant in adult males, specifically in germ cells of the testes (see Fig. 2B). Northern blot analysis of gd transcripts in dissected testes of these three developmental stages suggest that gd expression is limited to the male germ line during development (R.A. Schulz, unpubl.). In contrast to the other two genes of the cluster, Eip28/29 is expressed throughout embry-
onic, larval, and adult development. Eip28/29 gene trans-
cripts are present in, but not limited to, the ovary in adult females. Therefore, we can identify at least one tissue that expresses two of the overlapping genes. How-
ever, although we know that gdl is expressed in germ cells in the ovaries, the provenance of Eip28/29 expression in this tissue must still be ascertained.

The functional significance of overlapping genes is obscure at this time. However this unusual sequence organization raises questions concerning the related or independent regulation of genes in the cluster. One could envision at least three mechanisms by which the expression of a gene might be influenced as a result of this structural organization. First, overlapping genes may be subject to transcriptional interference in which transcriptional read through from an upstream promoter may negatively affect transcription from a downstream promoter (Adhya and Gottesman 1982). Such a phenom-
ennon has been implicated in the control of avian leuk-
Kosis virus transcription in which initiation from the promoter within the 5' long terminal repeat (LTR) inhibits transcription initiation from the promoter within the 3' LTR (Cullen et al. 1984). On the basis of our previous studies, which demonstrated gdl and Eip28/29 sequence overlap, we commented on the struct-
tural similarities to retrovirus organization and the pos-
sibility of transcriptional interference (Schulz et al. 1989). The documented overlap of the z600 gene with gdl reinforces the similarity of gdl gene structure with that of retroviruses; whether the diverse expression charac-

Discussion

We have completed the structural analysis of RNAs de-

erved from a densely transcribed segment of the 71CD interval of the Drosophila genome. In doing so, we have eluci-
dated the organization of an intriguing gene cluster. The finding is that there are three overlapping genes present within a 4.5-kb DNA region (see Fig. 8). The z600 gene generates a single 600-nucleotide transcript; an analysis of z600 and gd cDNAs reveals that the z600 transcript is polyadenylated at position +91 of the gdl gene (see Fig. 3). gdl is expressed in a sex-specific manner in adults, a result of the two distinct gene expression modes. We have used primer extension (see Fig. 5) and nuclease S1 (see Figs. 5 and 6) analyses to demon-
strate that the gdl and gd transcripts are identical save for the presence of additional sequences at the 5' ends of the gd transcripts. That is, different transcriptional initiation sites are employed in generating the two distinct RNA pairs. The two gd transcripts initiate at a single site located 30 bp downstream of a TATA box (A-T-A-A-T-A), a sequence motif that has been shown to be an important promoter component (Chambon et al. 1984). In contrast, there are multiple 5' ends for the gdl RNAs, with transcripts initiating at least 165 nucleo-
tides downstream of the gd initiation site. No obvious TATA sequence is observed near the gdl transcript 5' ends. The absence of such a sequence from the 5'-flanking regions of several eukaryotic genes has been corre-
lated with the generation of multiple gene transcripts with different 5' ends (Evans et al. 1988). Overall, the use of different gd and gdl transcription initiation sites indicates that alternative promoter usage, and not alternative splicing, is responsible for the sex-specific expression of gdl. The final observation from the structure analysis is that the gdl 1300- and 1500-nucleotide transcripts extend through the Eip28/29 flanking sequences into the δ exon of Eip28/29, being polyadenyl-
atated at position +62 of the gene. Although we have

Figure 9. Proper expression of a truncated gdl gene in germ-
line transformants. A Northern blot of RNA isolated from the y w transfor-
mation host and the B5 transformant line that harbors a gdl–CaSpeR transposon. Total RNAs from either five adults or five pairs of dissected gonads were assayed for gdl RNAs; expression of the introduced gdl gene can be monitored by the increase in the levels of the gd 1200- and gdl 1000-nucleotide RNAs. The RNA samples are as follows: B5 adult males (lane 1), y w adult males (lane 2), B5 adult females (lane 3), y w adult females (lane 4), B5 testes (lane 5), y w testes (lane 6), y w ovaries (lane 7), and B5 ovaries (lane 8).
Schulz and Butler

teristics of the three genes are a result of transcriptional interference in one or more cell types remains to be determined.

A second and quite different consequence of overlapping gene organization might be the use of a common regulatory element so as to coregulate the expression of two genes in a temporal or tissue-specific manner. In this case, it may be possible that both gdl and Eip28/29 are coordinately expressed during oogenesis. The coordinate regulation could be a result of a common ovary transcription element that affects transcription from both the gdl and Eip28/29 promoters. This does not appear to be a strategy used with other overlapping eukaryotic genes in which the tissue-specific expression characteristics of the genes are known. In other words, the overlapping genes at the Drosophila Gart complex are independently regulated by other genes of the somatic sex determination hierarchy [Nagoshi et al. 1988]. The finding of different transcription initiation sites for gdlM and gdlP transcripts confirms this hypothesis. An important question that must be addressed is whether separate tissue-specific regulatory elements exist that control expression in the testis versus the ovary or whether a common transcription element is responsible for expression in both the male and female germ lines. Although we have not addressed this point in our current work, we can say that gdl can be expressed properly outside of the overlapping gene environment and that a 1.8-kb DNA region contains all the sequences necessary for gdl sex- and tissue-specific expression [see Fig. 9]. Because this fragment includes 329/493 bp of 5′-flanking sequences relative to the gdlM/gdlP transcription start sites and because these sequences are located almost entirely within the transcribed region of z600, it is possible that essential gdl transcription control elements may reside internally within this upstream gene.

The use of alternative promoters for gdl expression in the male and female germ lines can be contrasted with the alternative splicing seen with at least two genes controlling somatic sexual differentiation. Differential processing of tra transcripts gives rise to a female-specific RNA and a non-sex-specific RNA [Boggs et al. 1987], whereas alternative splicing of doublesex (dsx) transcripts gives rise to both male- and female-specific RNAs [Nagoshi et al. 1988]. tra and dsx alternative splicing is essential for somatic differentiation and is regulated by other genes of the somatic sex determination hierarchy [Nagoshi et al. 1988]. Whether these different mechanisms of generating sex-specific transcripts represent a fundamental difference in the regulation of gene expression during sexual differentiation in the germ line versus soma remains an open question.

The function of the gdl gene in the male and female germ lines is currently unknown. However, our structural studies reveal that the gdlP and gdlM transcripts share an ORF whose translation could result in a gene product common to both germ lines [see Figs. 7 and 8]. One consequence of alternative gdl gene promoters is the presence of additional sequences at the 5′ end of the gdlM transcripts. These sequences contain an AUG codon generating an additional ORF upstream of the common coding region. The gdlM transcripts thus have the potential to serve as bicistronic mRNAs, both of the initiating AUG codons reside in a sequence context that can be considered translationally favorable [Kozak 1986, 1987]. Whether the gdlM transcripts produce multiple gene products, which have different functions during spermatogenesis, remains to be elucidated. Alternatively, the presence of the small ORF may somehow affect the translational efficiency of the downstream reading frame [Khalili et al. 1987; Kozak 1987]. The unusual gdl sequence organization may allow for both the transcriptional and translational regulation of gene expression during Drosophila gametogenesis.
Materials and methods

Drosophila strains

Flies were grown at 25°C on standard cornmeal–glucose–yeast agar medium containing Tegosept and supplemented with live yeast. ry + flies [McCarron et al. 1974] homozygous for chromosome 3 served as our wild-type stock. The fly stocks th st tra cp in ri p/TM2 and y, th st tra cp in ri p/TM6B were obtained from M. Wollner. All chromosomal markers and the TM2 balancer chromosome are described in Lindsay and Grell (1968), the TM6B balancer chromosome (Hf u Tc ca) is described in Craymer (1984). To obtain somatically transformed females and control flies, y/y, th st tra cp in ri p/TM2 males. XX flies were y +, XY flies were y, tra heterozygotes were Tb, and tra homozygotes were st. Gonads dissected from these animals all showed normal development with the exception of the gonads obtained from XXtra homozygotes, which exhibited a phenotype of rudimentary testes lacking sperm. The fly stock tud - bw sp/SMS (Boswell and Mahowald 1985) was also obtained from M. Wollner. To obtain germ-lineless flies, tud - bw sp females were crossed to ry + males. Gonads of the resulting progeny were greatly diminished in size and lacked egg chambers and sperm.

Nucleic acids, DNA sequencing, and transcript mapping

Total RNA was isolated from 0- to 4-hr embryos, sexed adults, and dissected gonads, following SDS–proteinase K treatment as described previously (Schulz et al. 1989). Total RNA from 11-5-4 cells was a gift of A. Bieber. gdl - cDNAs were isolated in a screen of an adult male cDNA library cloned in λgt10 (Poole et al. 1985). A z600 cDNA was isolated in a screen of a 0- to 5-hr embryonic cDNA library made in λgt10 by M. Goldschmidt-Clermont and D. Hognes. cDNA inserts were subcloned into the pUC8 vector [Vieira and Messing 1982] for restriction enzyme mapping or into M13mp8/9 vectors [Messing and Vieira 1982] for sequencing by the dideoxy method [Sanger et al. 1977]. The same method was used to sequence z600 and gdl genomic sequences, coordinates –4001 to –2394 based on the numbering system of Cherbas et al. [1986]. Northern, primer extension, and nuclease S1 protection analyses were performed as described previously (Cherbas et al. 1986; Schulz et al. 1986, 1989). Synthetic oligonucleotide primers were obtained from the M.D. Anderson Cancer Center core facility.

In situ hybridization

Female and male abdomens were frozen, fixed, embedded, and sectioned using the procedure of Liu et al. [1988]. RNA probes labeled with 32P, 35S were synthesized by T7 transcription of a linearized PstG-pBS template; this generates the PstG probe (coordinates –3729 to –198) cloned into the unique BamHI site of CaSpeR, the presence of a truncated white gene in the vector allows for the identification of germ-line transformants based on w + selection [V. Pirrotta, pers. comm.].

Germ-line transformation

Germ-line transformation was performed essentially as described by Rubin and Spradling [1982]. y w embryos were injected with 300 μg/ml of the gdl–CaSpeR transposon and 50 μg/ml of the helper plasmid wings-clipped [Karess and Rubin 1984]. The gdl sequences consisted of a 1.8-kb BclI fragment (coordinates –3729 to –198) cloned into the unique BamHI site of CaSpeR, the presence of a truncated white gene in the vector allows for the identification of germ-line transformants based on w + selection [V. Pirrotta, pers. comm.].

Acknowledgments

We are grateful to B. Sutton for technical assistance, to L. Cherbas for preliminary sequence information and the z600/gdl sequence homology searches, to M. Wollner for fly stocks, and to R. Davis and W. Klein for their helpful comments on the manuscript. This work was supported by grants to R.A.S. from the American Cancer Society (NP-596) and the National Science Foundation (DCB-8709846) and aided by grant RR511-23 from the National Institutes of Health.

References

Adelman, J.P., C.T. Bond, J. Douglass, and E. Herbert. 1987. Two mammalian genes transcribed from opposite strands of the same locus. Science 235: 1514–1517.

Adhya, S. and M. Gottesman. 1982. Promoter occlusion: Transcription through a promoter may inhibit its activity. Cell 29: 939–944.

Angerer, L.M., K.H. Cox, and R.C. Angerer. 1987. Demonstration of tissue-specific gene expression by in situ hybridization. Methods Enzymol. 152: 649–661.

Baker, E.S. and K.A. Ridge. 1980. Sex and the single cell: I. On the action of major loci affecting sex determination in D. melanogaster. Genetics 94: 383–423.

Bieber, A.J. 1986. Ecdysteroid inducible polypeptides in Drosophila Kc cells: Kinetics of mRNA induction and aspects of protein structure. Ph.D. thesis, Harvard University.

Boggs, R.T., P. Gregor, S. Idriess, J.M. Belote, and M. McKown. 1987. Regulation of sexual differentiation in D. melanogaster via alternative splicing of RNA from the transformer gene. Cell 50: 739–747.

Boswell, R.E. and A.P. Mahowald. 1985. tudor, a gene required for assembly of the germ plasm in Drosophila melanogaster. Cell 43: 97–104.

Chambon, P., A. Dierich, M.P. Gauthier, J. Jonstra, A. Kunst, J.-F. Le Pennec, P. Oudet, and T. Reudelhuber. 1984. Promoter elements of genes coding for proteins and modulation of transcription by estrogens and progesterone. Recent Prog. Horm. Res. 40: 1–42.

Chen, C., T. Malone, S.K. Beckendorf, and R.L. Davis. 1987. At least two genes reside within a large intron of the dunce gene of Drosophila. Nature 329: 721–724.

Cherbas, L., R.A. Schulz, M.M.D. Kochler, C. Savakis, and P. Cherbas. 1986. Structure of the Eip28/29 gene, an ecdysone-inducible gene from Drosophila. J. Mol. Biol. 189: 617–631.

Craymer, L. 1984. Techniques for manipulating chromosome rearrangements and their application to Drosophila melanogaster: II. Translocations. Genetics 107: 573–587.

Cullen, B.R., P.T. Lomedico, and G. Ju. 1984. Transcriptional interference in avian retroviruses—Implications for the promoter insertion model of leukaemogenesis. Nature 307: 241–245.

Evans, T., T. De Chiara, and A. Efstatiadis. 1988. A promoter of the rat insulin-like growth factor II gene consists of minimal control elements. J. Mol. Biol. 199: 61–81.

Garabedian, M.J., B.M. Shepard, and P.C. Wensink. 1986. A tissue-specific transcription enhancer from the Drosophila gene, an ecdysone-inducible gene from Drosophila. J. Mol. Biol. 199: 61–81.

Heinkoff, S., M.A. Keene, K. Fechtel, and J.W. Fristrom. 1986. Gene within a gene: Nested Drosophila genes encode unrelated proteins on opposite DNA strands. Cell 44: 33–42.

Hungr, M.-C. and P.C. Wensink. 1981. The sequence of the Drosophila melanogaster gene for yolk protein 1. Nucleic Acids Res. 9: 6407–6419.

---. 1983. Sequence and structure conservation in yolk pro-
teins and their genes. *J. Mol. Biol.* **164**: 481–492.

Karess, R.E. and G.M. Rubin. 1984. Analysis of P transposable element functions in *Drosophila*. *Cell* **38**: 135–146.

Khaliili, K., J. Brady, and G. Khoury. 1987. Translational regulation of SV40 early mRNA defines a new viral protein. *Cell* **48**: 639–645.

Kozak, M. 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**: 283–292.

——. 1987. An analysis of 5’-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15**: 8125–8148.

Lindsley, D.L. and E.H. Grell. 1968. Genetic variations in *Drosophila melanogaster*. *Carnegie Inst. of Washington Publ.* **627**.

Liu, X., L. Lorenz, Q. Yu, J.C. Hall, and M. Rosbash. 1988. Spatial and temporal expression of the *period* gene in *Drosophila melanogaster*. *Genes Dev.* **2**: 228–238.

McCallum, J., W. Gelbart, and A. Chovnick. 1974. Intracistronic mapping of electrophoretic sites in *Drosophila melanogaster*: Fidelity of information transfer by gene conversion. *Genetics* **76**: 289–299.

McKnight, S. and R. Tjian. 1986. Transcriptional selectivity of viral genes in mammalian cells. *Cell* **46**: 795–805.

Messing, J. and J. Vieira. 1982. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. *Gene* **19**: 269–276.

Nagoshi, R.N., M. McKeown, K.C. Burtis, J.M. Bealote, and B.S. Baker. 1988. The control of alternative splicing at genes regulating sexual differentiation in *D. melanogaster*. *Cell* **53**: 229–236.

Poole, S.J., L.M. Kauvar, B. Drees, and T. Kornberg. 1985. The *engrailed* locus of *Drosophila*: structural analysis of an embryonic transcript. *Cell* **40**: 37–43.

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

Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* **74**: 5463–5467.

Savakis, C., M.M.D. Koehler, and P. Cherbas. 1984. cDNA clones for the ecdysone-inducible polypeptides (EIP) mRNAs of *Drosophila Kc* cells. *EMBO J.* **3**: 235–243.

Schulz, R.A., L. Cherbas, and P. Cherbas. 1986. Alternative splicing generates two distinct Eip28/29 gene transcripts in *Drosophila Kc* cells. *Proc. Natl. Acad. Sci.* **83**: 9428–9432.

Schulz, R.A., W. Shimomchik, L. Cherbas, and P. Cherbas. 1989. Diverse expression of overlapping genes: The *Drosophila Eip28/29* gene and its upstream neighbors. *Dev. Biol.* **131**: 515–523.

Serfling, E., M. Jasim, and W. Schaffner. 1985. Enhancers and eukaryotic gene transcription. *Trends Gen.* **1**: 224–230.

Spencer, C.A., R.D. Gietz, and R.B. Hodgetts. 1986. Overlapping transcription units of the *dopa decarboxylase* region of *Drosophila*. *Nature* **322**: 279–281.

Vieira, J. and J. Messing. 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* **19**: 259–268.

Vincent, A., H.V. Colot, and M. Rosbash. 1986. Blastoderm-specific and read-through transcription of the *stry alpha* gene transformed into the *Drosophila* genome. *Dev. Biol.* **118**: 480–487.

Yamamoto, K.R. 1985. Steroid receptor regulated transcription of specific genes and gene networks. *Annu. Rev. Genet.* **19**: 209–252.
Overlapping genes of Drosophila melanogaster: organization of the z600-gonadal-Eip28/29 gene cluster.

R A Schulz and B A Butler

*Genes Dev.* 1989, 3:
Access the most recent version at doi:10.1101/gad.3.2.232

References
This article cites 36 articles, 7 of which can be accessed free at: http://genesdev.cshlp.org/content/3/2/232.full.html#ref-list-1

License

Email Alerting Service
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.