Molecular Structure and Transformation of the Glucose Dehydrogenase Gene in Drosophila melanogaster

Ross Whetten, Edward Organ, Philip Krasney, Diana Cox-Foster and Douglas Cavener

Department of Molecular Biology, Vanderbilt University, Nashville, Tennessee 37235

ABSTRACT

We have precisely mapped and sequenced the three 5' exons of the Drosophila melanogaster Gld gene and have identified the start sites for transcription and translation. The first exon is composed of 335 nucleotides and does not contain any putative translation start codons. The second exon is separated from the first exon by 8 kb and contains the Gld translation start codon. The inferred amino acid sequence of the amino terminus contains two unusual features: three tandem repeats of serine-alanine, and a relatively high density of cysteine residues. P element-mediated transformation experiments demonstrated that a 17.5-kb genomic fragment contains the functional and regulatory components of the Gld gene.

The glucose dehydrogenase gene (Gld) in Drosophila melanogaster is required at a single stage in development for the modification of the puparium. Gld mutants fail to close at the termination of metamorphosis but can be easily rescued by excising the anterior end of the puparium case (Cavener and Macintyre 1983). Despite the simple mutant phenotype, the GLD enzyme and mRNA are transiently expressed at every major stage of development (Cavener et al. 1986a; Cavener 1987a). The temporal pattern of Gld mRNA accumulation is highly correlated with accumulation of the major molting hormone ecdysterone and has been demonstrated to be regulated by this hormone during the third larval instar (M. Murtha and D. Cavener, unpublished data). Gld mRNA is expressed in a variety of ectodermal tissues including the hypophysis and antennal-maxillary complex (embryos); the anterior spiracular gland cells and the epidermis (third instar larvae); wings, legs, antennae, cibarium, epidermis, rectal papillae, neck, trachea, and some components of the reproductive tract of both male and female (pharate adults) (D. Foster-Cox, C. Schonbaum and D. Cavener, unpublished results). During the adult stage Gld expression is almost entirely limited to the male anterior ejaculatory duct (Cavener and Macintyre 1983; Cavener et al. 1986a). The GLD enzyme is apparently secreted since it can be recovered from the molting fluid of pupae and is transferred from adult males to females during copulation.

Inasmuch as Gld regulation involves steroid hormone control, a germ layer lineage restriction, and sexual differentiation, it is an excellent paradigm for developmentally regulated genes. In order to elucidate the cis-acting elements which are responsible for the various aspects of its regulation, we have engaged in a detailed molecular analysis of the Gld gene. Genomic DNA clones of the Gld gene were isolated by the method of chromosome walking, and identified on the basis of the localization of three independent Gld mutations (Cavener et al. 1986a). These three mutations were found in a 7-kb region at 84C8 on the right arm of the third chromosome. Northern hybridization analysis identified a 2.8-kb poly(A) RNA derived from this region which is highly correlated with the pattern of expression exhibited by the GLD enzyme throughout development. This correlation includes the virtual restriction during the adult stage to the male ejaculatory duct. We describe fine-scale mapping of the 5' half of the gene including the start site of transcription and translation. Proof that a 17.5-kb restriction fragment including the transcription unit contains the entire Gld gene is provided by P element-mediated gene transformation experiments.

MATERIALS AND METHODS

Subcloning and DNA sequencing: For the analysis of exon 1 genomic restriction fragments from lambda clone E14b (Cavener et al. 1986a) were inserted into Bluescript KS(+) or KS(-) phagemids (Stratagene, Inc.) which contain T7 and T3 phage promoters and can be propagated as a double-stranded DNA plasmid or, upon superinfection with helper phage, as a single-stranded DNA phage. The genomic subclones used for the analysis of exons II and III were previously described (Cavener et al. 1986a). These subclones are in the SP64/65 vectors (Promega Biotec, Inc.), which contain the SP6 phage promoter, or in pEMBLS/9 phagemid vectors (Dente, Cesareni and Cortese 1983).
A series of terminal deletions were constructed from the subclones in the phagemid vectors using the HENIKOFF (1984, 1987) EndIII method. These deletions were used for DNA sequencing and transcript mapping experiments. DNA sequencing of single strand templates from the deletion mutants was performed by the chain termination method (SANGER and COULSON 1975) using the Klenow fragment of Escherichia coli PolI and [α-32P]ATP. In some cases double stranded templates were sequenced using the alkaline denaturation method of CHEN and SEEBURG (1985). The DNA sequences of exons I, II and III were verified by sequencing each nucleotide from two independent DNA templates or (in most cases) by sequencing both strands.

**Northern hybridization:** Total RNA and poly(A)*RNA* were isolated after previously published procedures (CAVENGER et al. 1986a). The RNAs were fractionated on 2.2 m formaldehyde/1.2% agarose gels (SEED 1982). RNA gels were blotted to nitrocellulose or nylon membranes. The cRNA probes were prepared following the procedures of Promega Biotech. Hybridizations were performed at 58°C in standard hybridization buffer containing 50% (v/v) formamide. Filters were washed at room temperature with 2× SSC-0.2% (w/v) SDS and then at 65°C with 0.2× SSC-0.2% (w/v) SDS.

**RNase protection experiments:** The nuclease protection procedure of ZINN, DIAMO and MANIATIS (1983) was initially used to map exons II and III. Poly(A)*RNA* (1–5 μg) precipitated with ethanol was redisolved in 28 μl of hybridization buffer (80% (v/v) formamide, 400 mM NaCl, 20 mM PIPES, pH 6.4). RNA probes were synthesized using [γ-32P]ATP. After synthesis of the cRNA probe, the DNA template was removed by digestion with DNase I, the reaction continued with phenol and chloroform, and the probes were recovered by ethanol precipitation. The probes were redissolved in 20–50 μl of hybridization buffer, and 2 μl were then added to the RNA solution. The hybridization mixture was heated to 80°C for 3 min, then hybridized overnight at 40–50°C. The hybrids were then digested in 30 μl of an RNaseA solution (40 μg/ml RNase A, 2 μg/ml RNase T1, 300 mM NaCl, 10 mM Tris pH 7.5, 5 μM EDTA) for 1 hr at 23°C. The digests were stopped by phenol/chloroform extraction and the hybrids precipitated by the addition of 5 μg of carrier RNA and 2 volumes of cold ethanol. The samples were electrophoresed on sequencing gels for maximum resolution. FISCHER and MANIATIS (1985) had noted that high concentrations of RNase A can lead to undesired cleavage of poly(U)-poly(A) hybrid tracts. Since exon I contains such a sequence we found that it was necessary to reduce the concentration of RNase A in the reaction by 400-fold (i.e., to 0.1 μg/ml).

**Primer extension experiments:** The method used for primer extension experiments was modified from a protocol obtained from ROBERT THOMPSON (personal communication). Oligonucleotides were synthesized with a Biosearch DNA synthesizer in the laboratory of STEVEN LLOYD (Dartmouth University). The oligonucleotide primers were 5‘- end-labeled using [γ-32P]ATP (MANIATIS, FRITSCH and SAMBROOK 1982). Poly(A)*RNA* (1–5 μg) was mixed with 5 pmol of 5‘-end-labeled oligonucleotide primer in a total volume of 10 μl of distilled, pyrococaric acid diethyl ester-treated water. The mixture was heated 3 min at 65°C and then incubated for 30–60 min at 45°C or 50°C. An equal volume of 2X reaction buffer was then added to start the extension reaction. The 2X reaction buffer contained 100 mM Tris (pH 8.3), 20 mM DTT, 20 mM MgCl2, 2 mM of each of the four dNTPs, 100 units/ml plasmid RNase inhibitor and 1000 units/ml avian myeloblastosis virus reverse transcriptase. The extension reaction was allowed to proceed 30 min at 43°C, then stopped by the addition of EDTA to a final concentration of 20 mM. Sodium acetate was added to a final concentration of 0.3 M and the reaction products precipitated with ethanol. The redissolved samples were electrophoresed on sequencing gels.

**P element-mediated transformation:** The pWG67 transformation clone was constructed from a 9.7-kb KpnI-Sall fragment containing the 5’ half of *Gld*, a 7.8-kb SalI-Kpnl fragment containing the 3’ half of *Gld*, and the pW5 P element-translation vector of KLEMMENZ, WEBER and GERSHING (1987). The pWG67 Gld clone was co-injected with the ps25.7 wc helper P element into Δγ w (yellow; white) preblastoderm embryos using the general procedures of SPRADLING and RUBIN (1982). Survivors were backcrossed to the Δγ w host strain and putative transformants among the progeny were detected as adult flies with red eyes and yellow bodies. As noted by KLEMMENZ et al. (1987) transformants using the pW5 vector do not typically exhibit wild-type red eye color; instead they display colors similar to Δγ w hypomorphs. Two independently transformed strains, T-7.1 and T-14.3, were found to carry pWG67 inserts on chromosome II. T-7.1 and T-14.3 were separately crossed to a Gld null strain (Δγ w; Gld®*/Gld®cu/cu). F1 males were backcrossed to the Gld null parent strain and their progeny analyzed. As expected red eye-curved wing (w cu) flies self-eclosed whereas white eye-curved wing (Δγ cu) flies died in the “head-jammed” state typical of the Gld lethal phenotype. Approximately equal numbers of these two phenotypes were observed, consistent with independent assortment of the rescuing factors (i.e., the pWG67 insertions) on chromosome II from the curled mutation (tightly linked to Gld) on chromosome III. The T-14.3 insertion induced a recessive lethal mutation so it has not been possible to create a homozygous strain. Genomic Southern analysis of the transformants was performed after previously published procedures (CAVENGER et al. 1986a) to verify the integration of the transforming DNA and to determine copy number.

**RESULTS**

**Low resolution mapping of the Gld 2.8-kb mRNA:** A series of 22 restriction fragments from a 27-kb region containing the Gld gene was subcloned into one of three vectors which support the synthesis of cRNA probes. Single stranded cRNA probes representing both strands were used to probe Northern blots containing total RNA from pharate adults and female and male adults. A summary of these results is presented in Figure 1. Probes spanning a region of 14 kb of genomic DNA detected the Gld 2.8-kb mRNA. The positive probes were clustered in three groups separated by two putative intronic regions. The data germane to the 3’ half of the Gld transcription unit have been published (CAVENGER et al. 1986a). Only one other RNA species, 1.5 kb in length, has been consistently observed to hybridize to probes from this region. The 1.5-kb RNA is identified by a few of the 3’ probes which also identify the 2.8-kb mRNA. The temporal pattern of expression of the 1.5-kb RNA is not correlated with the pattern of GLD enzyme expression. Two RNAs complementary to the strand containing Gld mRNA were detected using
Molecular Units of Drosophila Gld

Figure 1.—Molecular map of the Gld gene. (A) Restriction map derived from genomic clones and partially confirmed by whole genome Southern analysis. (B) Transcript map of the 2.8-kb Gld mRNA. Thick lines and numbers indicate the location and order of exons. Stippled line denotes the region of the 3' exons (IV-?) which have not been precisely mapped. Thin lines denote introns. (C) Composite 7-kb deletion Dfa41/Dfr29 and the T2;3 XaD5 reciprocal translocation breakpoint. As predicted, flies bearing these mutations lack GLD enzyme activity, lack the Gld 2.8-kb mRNA, and exhibit the noneclosion mutant phenotype characteristic of Gld mutants. (D) The 17.5-kb Kpnl fragment within the pWG67 clone which is able to provide Gld functions upon transformation.

probes from this region. However, these antisense RNAs are not transcribed from the Gld locus (CAVENHER et al. 1986a).

The 5' most genomic probes which hybridize to the Gld 2.8-kb RNA map more than 5 kb upstream from the composite A41/R29 deletion which genetically localizes Gld. In order to provide further evidence that the 2.8-kb RNA detected by the 5' extreme probes corresponds to the Gld mRNA, a Northern blot containing RNA from the A41/R29 deletion was probed with a complementary radiolabeled cRNA from this region (Figure 2). As predicted the 2.8-kb RNA is not detected in the A41/R29 deletion, as was previously shown for probes corresponding to more 3' exonic regions (CAVENHER et al. 1986a). The Northern blot was subsequently hybridized with an Adh (alcohol dehydrogenase) probe to confirm the presence and integrity of the RNA in the A41/R29 sample.

High resolution mapping: Exon I—The genomic subclone pCG4, containing a 4.2-kb XbaI/EcoRI fragment, was determined to correspond to the putative 5' end of the Gld gene (coordinates 147.4-151.6 of Figure 1) Two sets of terminal deletion mutations of pCG4 (Figure 3) were constructed using the HENIKOFF (1984) ExoIII method. One set contained deletions from the XbaI end while the other set contained deletions from the EcoRI end (Figure 3). Single-stranded cRNA probes were made from representatives from each of the two sets and used to probe Northern blots of poly(A+) or total RNA from pharate adults. These experiments lead to the delineation of a region between the deletion breakpoints of subclones 25a and 31a which hybridize with the Gld 2.8-kb RNA.

RNase protection experiments were used to precisely map the position of exon I (Figure 4). Three radiolabeled cRNA probes complementary to the Gld mRNA were individually hybridized with pharate adult poly(A+) RNA, subjected to RNase A and T1 digestion, and fractionated on urea-PAGE gels to

Figure 2.—Confirmation of the identity of the Gld 2.8-kb RNA. (A) A Northern blot was hybridized with a cRNA probe corresponding to the 5' most region in Figure 1 which hybridizes with a 2.8-kb RNA. As expected the 2.8-kb RNA is not detected in A41/R29 males. This blot was reprobed with an Adh-specific radiolabeled probe to demonstrate the integrity of the A41/R29 RNA sample. The 1.1-kb RNA corresponds to the Adh mRNA. (B) A Northern blot hybridized with the 32P-end-labeled 24-mer oligonucleotide used in the primer extension experiments (Figure 6).
determine the size of the protected fragments. All three experiments yielded single protected fragments. The difference in the length of the fragments protected by probes B and C is approximately equal to the difference in the lengths of the two probes suggesting that the two probes both protect the 5' end of exon I but neither protects the 3' end. Furthermore, this suggests a precise position of the 5' end of exon I (i.e., 313 and 192 nucleotides (nt) from the deletion breakpoints of probes B and C, respectively). Probe A is thought to protect the entire exon as implied by the deletion subclone/Northern blot analysis described above (Figure 3). Thus, the probe A-protected fragment, 335 nt, is the estimated size for exon I. Because the three probes have one common end we interpret these results as indicating the presence and position of a single 335 nt exon. It should be noted that it was necessary to modify the RNase protection protocol of ZINN, DIMAIO and MANIATIS (1983) as suggested by FISCHER and MANIATIS (1985) in order to prevent cleavage at a poly(A) tract within the Gld mRNA when hybridized to the complementary probes.

The DNA sequence of an 1143-bp region beginning at the Xba end of the fragment was determined (Figure 5) using the method of SANGER and COULSON (1975), as described in MATERIALS AND METHODS. A sequence (ATC/GTAAGT) similar to the 5' splice junction consensus (MOUNT 1982) was found at the predicted position based upon the RNase protection experiments. At the predicted 5' end of exon I is the sequence TGAGTCGG which is very similar to the Drosophila transcription start site consensus sequence (SNYDER et al. 1982; CHERBAS et al. 1986).

In order to confirm that the 5' end of exon I corresponds to the 5' end of the Gld mRNA and the putative start site of transcription, primer extension experiments were conducted. The 24-mer oligonucleotide primer used for this experiment should bind to the Gld mRNA approximately 100 nt downstream of the 5' end of exon I predicted by the RNase protection experiments. Using this primer, adult male and pharate adult poly(A)+ RNA yielded a major 100 ± 1 nt primer extension product (Figure 6). As predicted adult female poly(A)+ RNA yields very little of this extension product. These results were consistently obtained in three separate experiments. Raising the temperature of the primer extension reaction substantially reduces the amount of other primer extension products without decreasing the signal from the major 100-nt product. Therefore, we speculate that these other primer extension products are the result of random binding of the primer to other RNAs under
Molecular Units of Drosophila Gld

FIGURE 5.—DNA sequence of the Gld promoter region and exon I. The sequence begins at the PstI site and ends at the XbaI site shown in Figure 4A and in Figure 1 (coordinates 147.5–148.6). Numbers are relative to the start site of transcription (+1). Double underlined: four direct repeats of the TAGACCA motif. Dash underline: a 13 nt palindrome (starting at −73) and the TATA box (starting at −31). Solid underline: exon I. Over line (+77 to +100): oligo-233 complementary sequence used for primer extension experiments.

the rather low stringency conditions dictated by the primer extension reaction. It is important to note that this primer detects only the Gld 2.8 mRNA in Northern hybridization experiments which are conducted under much higher stringency conditions (Figure 2B).

Exons II and III—Eight kilobases downstream of exon I, probes from two small adjacent restriction fragments (PvuII-BamHI, 450 bp; BamHI-HindIII, 340 bp; see Figure 1) detect the Gld 2.8 mRNA (Caveney et al. 1986a). The DNA sequence of this region was determined and analyzed for potential RNA splice sites and coding regions (Figure 7). From this analysis emerged a model for two small coding exons (177 nt and 121 nt) separated by a 73-nt intron (Figure 8). The details of this model were tested by RNase protection experiments. The probes used and the predicted products are shown in Figure 8A and the results of the experiments to confirm the model are shown in Figure 8, B–D. Probe A protects two RNA fragments with the predicted lengths of the two exons (Figure 8B). A cRNA probe derived from a partially digested template at the BamHI site yields the 121-nt product, a small amount of the 177-nt and a 27-nt product (not shown) which corresponds to the 3' terminus of exon II (Figure 8C). Since probe B is predicted to terminate in exon II it should yield a single product which should confirm the position of exon II in the DNA sequence. As shown in Figure 8D the probe B experiment gives a 155-nt protected fragment. Although similar experiments were not done to determine the precise position of exon III, we are confident that our assignment of its position in Figure 7 is correct since the predicted 40 amino acids encoded in exon III are perfectly conserved within three divergent Drosophila species whereas the putative intronic sequences immediately flanking exon III
are poorly conserved (P. Krasney and D. Cavenor, unpublished data).

To confirm that no Gld exons exist between exons I and II, a primer extension experiment was performed using a primer corresponding to exon II (at the BamHI site). If no other exonic sequences lie between exons I and II, this experiment should yield a 490-nt fragment. Three primer extension products were observed: 180, 185 and 485 nt (data not shown). The latter is very close to the predicted fragment. The smaller primer extension products are most likely the result of the primer binding to partially complementary sites in exon I.

The Pustell-IBI codon bias method was used to search for a putative protein coding region in exons I, II and III (Figure 9). It is almost certain that exon I is entirely untranslated since it does not contain any start codons and does not exhibit a codon bias typical of Drosophila genes. A 96 amino acid open reading frame was identified which begins 10 nt from the 5' end of exon II and continues through exon III (Figure 7). The putative start codon is flanked by sequences which are similar to the Drosophila consensus sequence (C/A A A A/C A U G) for translation initiation.
Molecular Units of Drosophila Gld

Figure 8.—RNase protection mapping of exon II, intron II, and exon III. (A) The final derived map of exon II and exon III (open boxes) and intron II (solid line between the two exons). The complementary cRNA probes used to determine the size and boundaries of exon II and exon III are below the map. The thick portion of each probe map represents the protected fragment after RNase digestion. (B) Probe A. Lanes 1 and 2, molecular weight markers; Protecting RNAs: lane 3, male; lane 4, pharate adult; lane 5, female; and lane 6, yeast. (C) Probe A, partially linearized at the BamHI site. This yields two cRNA probes: a full length PvuII-HindIII probe and a BamHI-HindIII probe. Protecting RNAs: lane 1, yeast; lane 2, female; lane 3, pharate adult; and lane 4, adult male. Molecular weight markers in lane 5. In addition, a 27-nt fragment was also observed in lanes 3 and 4 (not shown) which corresponds to a region in exon II from the BamHI site to the 3' end. (D) Probe B. Molecular weight markers—1; Protecting RNAs: 2—adult male.

Figure 9.—Codon bias analysis of exons I, II and III. The DNA sequence file of exon I and the first 307 nt of intron I was fused to a sequence file containing the last 529 nt of intron I, exon II, intron II, exon III and the first 215 nt of intron III. This composite sequence was analyzed using the Pustell-IBI Protein Coding Region Locator program on a Compaq 286 computer. The vertical dashed lines represent the value of the C-statistic calculated for successive 40 nt steps. At the top of each line is a number (1–3) indicating the reading frame. Values which extend above the line labelled M indicate regions which display significant bias in codon usage when compared with a Drosophila codon bias table. This analysis identifies two putative coding regions located in exons II and III. The “T”s below the C-statistic values denote the positions of termination codons in reading frames 1, 2 and 3.
completely rescue Gld mutants from their non-eclosion lethal phenotype. Moreover, these transformants exhibit the normal temporal pattern of Gld expression: low expression in feeding third instars, high expression in wandering third instars (i.e., immediately before pupariation), high expression during metamorphosis and in adult males and very low expression in adult females (Table 1). The quantitative levels of GLD expression are quite similar between the host and transformant lines, although GLD activity in T-14.3 adult males is significantly lower than what we have observed for a variety of wild-type strains. The latter effect is probably due to the influence of the local genomic environment of this transformant.

**DISCUSSION**

Gene structures in Drosophila have often been dichotomized between small genes (ca. 1-7 kb) which encode enzymes or other nonregulatory proteins and large genes (ca. 50-100 kb) which encode proteins which function to regulate development (LEWIN 1987). However, the Gld gene (ca. 18 kb) joins a growing list of intermediate sized genes encoding enzymes (Dunce—CHEN, DENOME and DAVIS 1986; Gart—HENIKOFF et al. 1986; Ace—HALL and SPIERER 1986) which obviate this dichotomy. One structural feature which is generally shared by Drosophila genes of all lengths is the presence of a small 5' exon usually containing the untranslated leader sequence and a very small portion of the coding region or, in few cases such as Gld, containing only the leader sequence. The first exon is then followed by what is typically the largest intron of the gene (e.g., yellow—CHIA et al. 1986; large subunit of RNA polIII—BIGGS, SEARLES and GREENLEAF 1985; Gart—HENIKOFF et al. 1986; α1, α2 and α4 tubulin genes—THEURKAUF et al. 1986). In the case of Gld the first intron is unusually large (8 kb). We speculate that these structural features may be the result of independent origins of the regulatory and coding regions consistent with the exon shuffling model of gene evolution (GILBERT 1978) or that such gene structures are the result of random acquisition of regulatory elements and coding
sequences along a contiguous linear sequence. The latter idea is a simple extension of the model proposed by SENAPATHY (1986) for the evolution of coding exons.

Upstream of the Gld transcription start site is a somewhat unusual TATA sequence (~31, TTTAAAAA) similar to that found for the Drosophila dopa decarboxylase gene (HIRSCH, MORGAN and SCHOLNICK 1986). Two interesting sequence elements found upstream of the TATA box are: (1) a 13-bp palindrome (at -78) separated by a single base pair at the axis of dyad symmetry and (2) four copies of a 7 bp dispersed repeat (at -248, -154, -144 and -106) based upon the sequence motif TAGACCA. A search in the promoter regions of a number of Drosophila genes in the Genbank data base and recent publications failed to detect these particular sequences in other known and putative promoters. In addition to the sequence elements immediately upstream of the start site of transcription, a 72-bp tetranucleotide tandem repeat element was found in the middle of intron 1 (CAVENER et al. 1988). This sequence element, named the YYRR box, is conserved in the Gld gene of three divergent Drosophila species indicating that it may serve some function. The requirement of these sequence elements for Gld expression is currently under investigation using the techniques of in vitro mutagenesis and P-element mediated transformation. The transformation experiments reported herein indicate that an 17.5-kb KpnI genomic fragment, which includes 2.3 kb of sequence to the 5' side of the Gld transcription start site, is sufficient for normal quantitative and qualitative expression.

Another unusual feature of the Gld gene is the presence of a large (344 nt) untranslated leader sequence. Although long leader sequences were once thought to be unique to heat shock genes (SOUTHGATE, AYME and VOELLMY 1983), the number of other Drosophila genes which have been reported to contain long leader sequences (i.e., >200 nt) has dramatically increased in the past few years (e.g., yellow—CHIA et al. 1986; Antennapedia—LAUGHON et al. 1986 and STROEGER, JORGENSEN and GARBER 1986; Ace—HALL and SPIERER 1986; Kruppel—ROSENBERG et al. 1986; Notch—WHARTON et al. 1985 and KIDD, KELLY and YOUNG 1986; large subunit of RNA pol II—BIGGS, SEARLES and GREENLEAF 1985; Ultrabithorax—WILDE and AKAM 1987; Dint I—RIJSEWIK et al. 1987). The discovery of long leader sequences among these genes has raised the question of translational control particularly since virtually all such leaders contain multiple short upstream open reading frames (uORFs). The Gld leader sequence is thus somewhat unique among long leaders in being devoid of uORFs.

The Gld enzyme is secreted into the molting fluid of pupae and into the seminal fluid of adult males. Thus we expected to find a hydrophobic signal sequence at the inferred amino terminus of Gld. Although some hydrophobic residues are located among the first 27 amino acids, a highly hydrophobic region is not found until residues 28 through 40. However, analysis of this putative signal sequence (Figure 7) using the weight-matrix procedure of HEIJNE (1986) indicated that it did not conform particularly well to the eukaryotic signal sequence consensus (analysis not shown). In addition, no other region in the first fifty residues of the inferred Gld preprotein was found to conform significantly better. Obviously, the location of the signal peptide and the cleavage site must await direct sequence analysis of the mature Gld protein. Two other unusual features of the inferred amino terminus are the serine-alanine triplet repeat (residues 2–7) and the presence of five cysteine residues.

We thank SUSAN SCHLITZ for technical assistance and ROBERT THOMPSON for advice on the primer extension experiments. This work was supported by grants from the National Institutes of Health and the National Science Foundation to D.C.

LITERATURE CITED

CAVENER, D., G. CORNETT, D. COX and R. WHERTEN, 1986a Isolation of the eclosion gene cluster and the developmental


Communicating editor: V. G. Finnerty