Structural and Functional Studies of the 68C Glue Protein Gene Cluster of *Drosophila melanogaster*

Thesis by

Mark David Garfinkel

Submitted in Partial Fulfillment of the Requirements

for the degree of

Doctor of Philosophy

in Molecular Biology

California Institute of Technology

Pasadena

1988

(Submitted February 24, 1988)

This dissertation is dedicated to my father, Charles Garfinkel, and to the memory of my mother, Lillian Allegra Gattengo Garfinkel. Words cannot express my gratitude for their love, support, and encouragement.

Acknowledgements

First, I must thank my advisor Elliot Meyerowitz for setting up the experimental system I have worked on, for allowing me opportunities to succeed (and sometimes to fail) on my own terms, and for being remarkably tolerant of my numerous idiosyncracies. Thanks also to my Thesis Committee members: Norman Davidson, E.B. Lewis, Carl Parker, and Barbara Wold.

My time at Caltech was enhanced by many people, friends and colleagues, members of Elliot Meyerowitz' and other laboratories. Space will not permit me to name them all. In particular, I would like to thank: Lynn Crosby, from whom I learned so very much about *Drosophila* and other organisms; Stewart Scherer and Stuart Kim, who always challenged me to think harder; Sue Celniker, Sarah Smolik-Utlaut, and Joanne Topol, for stimulating conversations about flies; Bill Mattox, for technical advice on science and on softball; Beverley Bond Matthews and James J. Lee, for helping me get through the worst night of my life; all of you who chauffeured me about town before I became autonomously mobile; Carol Mayeda, for microinjecting a truly vast number of *Drosophila* embryos for the experiments in Chapter 3; Caren Chang, for patiently listening to my all-too-frequent paranoid-delusional rantings; Margo Roark, for collaborating in the application of the ADH methods; Michael J. Palazzolo, M.D., Ph.D., for sharing his endless supply of bizarre stories, and for priceless medical advice; Pete Mathers, for reading early drafts of Chapters 1 and 3, and for being a source of flies and Fly Talk; Marty Yanofsky and Sherry Kempin, for free dinners, friendship, and the *Bal*31 deletion endpoint sequencing in Chapter 3; and especially K. Vijay Raghavan, for his bad puns, cheerful outlook, sound advice, and *many* grocery runs.

Special mention must be made of Geoffrey Duyk and of Mark Minie. Thank you both, for advice and friendship over the years and the miles.

Thank you, Boon Tan: none of this could have happened without you.

During the course of my graduate term, I was a predoctoral fellow of the National Science Foundation, and a trainee in molecular and cellular biology under the National Research Service Award 1 T32 GM07616. Part of the expense of preparing this dissertation was defrayed by the Jean Wiegle Memorial Fund.

Mark David Garfinkel

Abstract of Dissertation:
Structural and Functional Studies of the 68C Glue Protein Gene Cluster of *Drosophila melanogaster*

The 68C locus of the *Drosophila melanogaster* polytene chromosomes contains the structural genes for three glue polypeptides (sgs-3, sgs-7, and sgs-8) synthesized in the third instar larval salivary glands. The three 68C glue mRNAs are coded in a gene cluster of less than 5000 base-pairs, and are expressed coordinately under the control of the steroid hormone ecdysterone. Neither amplification nor DNA rearrangement of the locus occurs in the salivary gland. The nucleotide sequence of genomic DNA that includes the entire gene cluster was determined, as were the structures of each of the three glue protein mRNAs. Analysis of the sequences revealed that the three glue proteins form a diverged gene family. Each member of the gene family contains an amino-terminal hydrophobic block of amino acids, which is absent in the mature, secreted glue proteins, and a cysteine-rich carboxyl terminal module. sgs-3 differs from sgs-7 and sgs-8 by containing a third module between the other two, comprised largely of tandem repeats of the five amino acids Pro-Thr-Thr-Thr-Lys.

Two of the genes *Sgs-7* and *Sgs-8* are divergently transcribed with 475 base-pairs separating the two 5' ends. A transcriptional fusion gene was constructed by joining the 5' untranslated region of *Sgs-7* to the 5' untranslated region of the *D. melanogaster Adh* gene. A translational fusion gene was constructed by joining the *Sgs-8* gene to the *Escherichia coli lacZ* gene. When the fusion genes are placed in their normal divergently transcribed arrangement and reintroduced into *D. melanogaster* using P element gene transfer, third instar larval salivary gland expression of both alcohol dehydrogenase activity and β-galactosidase activity was observed. Expression of the two fusion genes requires the *l(1)npr-1*$^{+}$ gene product, which is known to regulate the

68C glue protein genes, supporting the proposal that this *trans*-acting factor affects glue protein gene transcription. Normal tissue, stage, and quantity of *Sgs-7—Adh* fusion gene expression is observed when 211 bp of the 5' flanking sequence is present. An *Sgs-7—Adh* fusion gene with 92 base-pairs upstream is non-functional. Third instar larval salivary gland expression of the *Sgs-8—lacZ* fusion gene is observed when 432 base-pairs of the intergenic region are present, while 415 base-pairs of 5' flanking sequence permits normal tissue and stage of expression at levels at least twentyfold reduced. The experiments suggest that a single region functioning bidirectionally, located closer to the *Sgs-7* gene, is required for expression of both genes.

Table of Contents

List of Illustrations

Page

List of Tables

Chapter 1:

General Introduction

The question of how organisms grow and develop from egg to embryo to maturity to senescence has fascinated people for thousands of years. But only in the last century have biologists gained information about development sufficient for paradigms to be proposed to answer this question. Two theories guide much of the research work in developmental biology: the concept of differential gene activity, and the concept that differential gene activity arises from the selective binding of proteins to specific DNA sequences.

The origin of the differential gene activity concept can be traced to the integration by Sutton and by Boveri in the early years of the 20th-Century of Mendel's laws of inheritance with late 19th-Century cytological studies of mitosis, meiosis, and fertilization (reviewed by Davidson, 1986). The realization that cells of different tissues contained identical chromosome sets eventually led to T.H. Morgan's statement in 1934 that "...different batteries of genes come into action as development proceeds..." (quotation from Davidson, 1986). Additional evidence that the nuclear genomes of different cell types are generally equivalent came from measurements of nuclear DNA content, from transplantation of nuclei into embryos to test for developmental potential, and from hybridization of specific cloned genes to gel blot filters of genomic DNAs from tissues that do or do not express these specific genes.

Incorporating the so-called "central dogma" of molecular biology with the concept of differential gene activity, the most elemental formulation of the differential gene activity concept states that the diverse tissues of a multicellular organism differ in the sets of DNA segments, genes, that are used as templates for the synthesis of RNA molecules. Messenger RNA is used as a template for the synthesis of protein molecules. These disparate subsets of protein molecules, some of which serve as

enzyme catalysts of metabolic reactions, others as components of macromolecular assemblies, are the sources of the disparate properties that distinguish the diverse tissues. By extension, differential gene activity also accounts for the differences within a tissue type that arise as a function of the organism's age.

A vast array of empirical data supports this version of the concept of differential gene activity. Numerous examples are known in a wide variety of organisms of a specific tissue type that contains protein products unique to that tissue type. In some cases, these are abundant amounts of protein components of intracellular structures, for example, the myofibrillar protein isoforms that assemble into the contractile apparatus in muscle cells. In other cases, these are enzymes that catalyze a particular metabolic pathway: various neurotransmitter biosynthetic enzymes in various specific subtypes of neurons. In still other cases, these are secreted proteins serving structural roles outside the cells that synthesize them: fibroin and sericin in the silk glands of moths. The specificity of protein synthesis and accumulation, as revealed by methods that allow for the identification of particular proteins, is paralleled by a corresponding specificity of RNA synthesis and accumulation, as revealed by nucleic acid hybridization methods. Thus, a variety of tissues express particular sets of genes at particular times during development.

Given the fact of differential gene activity, a question of paramount importance is: What is the mechanism by which gene activity is regulated? This question can be divided into three parts. First, if we adhere to the strictest definition of differential gene activity, we want to learn how different genes become available for transcription in different tissues. Second, if we accept the possibility that differential gene activity may be accomplished by means other than, or in addition to, the regulation of transcription,

we want to learn about these processes as well. Third, we want to learn the mechanism of differential *gene set* activity—how several genes are subjected together to a particular sort of developmental regulation.

One of the specific biochemical mechanisms for differential gene activity is the selective binding of protein molecules to specific DNA sequences in proximity to a gene to control mRNA transcription. This was first observed in the regulation of lactose catabolic enzyme biosynthesis in *Escherichia coli* by the biochemical-genetic experiments of Jacob and Monod and their colleagues, and by the protein-biochemical studies of Gilbert and his colleagues. In bacteria, coordination of the synthesis of a set of enzymes that comprise a particular metabolic pathway is generally accomplished by the "operon" organization, in which a single set of *cis*-acting DNA sequences regulate the transcription of several enzyme-coding regions, forming a polycistronic mRNA that is translated to each enzyme of the pathway. In contrast, eukaryotic transcription units that code for protein products are monocistronic, and coordinately regulated genes may be dispersed in the genome.

Different operons are subject to different regulatory strategies. These involve either positive control or negative control combined with either repression or induction, for a total of four kinds of control circuit (reviewed by Lewin, 1974). Several operons make use of regulatory mechanisms that affect processes other than transcription initiation. Despite the diversity of bacterial operons, the features common to all the regulatory schemes are: (i) the step subject to the most important quantitative control is transcription initiation; (ii) *trans*-regulatory proteins exert their effects by binding to (iii) specific targets, *cis*-acting DNA sequences; and (iv) allosteric regulation of the capacity of a *trans*-regulatory protein to bind its target is often mediated by relatively small

molecules, the intracellular concentrations of which serve as indicators of the cell's physiological state. The regulatory logic and the two macromolecular entities are the conceptual elements from bacterial gene systems almost universally applied to the problem of differential gene activity in metazoans.

A metazoan organism particularly well-suited for studying developmental gene regulation is the laboratory fly, *Drosophila melanogaster*. The organism is small, has a short generation time and is extremely prolific in laboratory culture conditions. Powerful methods of genetic analysis exist, including methods for testing *in vitro*-modified DNA segments for function by genetic transformation (Rubin and Spradling, 1982). The relatively small genome provides operational convenience in nucleic acids experiments and in mutational screens, and provides limits on the complexity of the problem. *Drosophila* has anatomically distinct larval and adult morphologies, which are separated by a complete metamorphosis, and which provide a rich diversity of tissues for study. At 25°C in laboratory culture conditions, embryogenesis requires 24 hours. The larval period is punctuated by two molts into three larval instars. Transient increases in the hemolymph concentration of the steroid hormone ecdysterone during the organism's life regulate numerous processes, including molting. The first and second larval instars each last about a day, during which the animal feeds and grows. The third larval instar lasts approximately two days, most of which time is also spent feeding and growing. During the final several hours of the third instar, the animal leaves the food in order to find a dry surface for pupal development. Eventually, the animal becomes immobile and molts without leaving the cuticle. This brief period may be considered a sessile fourth instar, and is followed by another molt. Inside these two layers of shed cuticle, metamorphosis takes place. Some larval tissues are histolyzed, others are remodeled for continued use during adulthood. The adult precursor tissues,

imaginal discs and abdominal histoblasts, undifferentiated since their formation during embryogenesis, are reorganized and differentiate into the adult exoskeleton. Metamorphosis takes four days, at the end of which the adult fly breaks open the pupal case and emerges. Within a day, sexual maturity is achieved, and the next generation soon begins.

Ecdysterone, which as stated triggers each cuticle molt, is synthesized by the ring gland under neuroendocrine control and secreted into the hemolymph that bathes the animal's tissues. Ecdysterone stimulates a variety of physiological changes. *In vivo* and *in vitro*, disparate tissues including the fat body, the salivary gland, the hypoderm, and the imaginal discs, respond to ecdysterone with tissue-specific changes in gene expression. Additionally, imaginal discs respond to ecdysterone exposure by undergoing disc eversion, beginning the process of metamorphosis. The classical view of insect endocrinology is that the relative concentrations of ecdysterone and the isoprenoid juvenile hormone together determine whether a molt is larval-to-larval, larval-to-pupal, or pupal-to-adult. In *D. melanogaster*, however, there is evidence that the hormonal control of development does not follow the classical view (see, for examples, Postlethwait, 1974; Richards, 1978).

The tissues that will eventually respond to ecdysterone in their characteristic manners arise very early during embryogenesis. Following fertilization and pronuclear fusion, the initial step of *Drosophila* development is a series of rapid, synchronous nuclear divisions in the absence of cytokinesis. Most of the nuclei migrate to the egg periphery, forming an orderly array beneath the egg plasma membrane. The first dozen or so nuclei to reach the extreme posterior pole of the embryo bud out with the posterior polar plasm and cleave away from the embryo to produce "pole cells"; this occurs 90

minutes after fertilization. The remaining peripheral nuclei undergo three or four further rounds of synchronous division before cell membranes surround them to form the cellular blastoderm. With gastrulation, which begins around three and a half hours after fertilization, the cellular blastoderm becomes organized into three germ layers. Following this, tissue and organ primordia become recognizable. In only a few hours, the *Drosophila* embryo is transformed from a single zygote nucleus to thousands of cells that are elaborating specific developmental fates.

The determination of cell fate in the early *Drosophila* embryo is a consequence of the interaction between cleavage nuclei and the particular regions of the egg cytoplasm they associate with during cellularization. Perhaps the most dramatic example of this is the determination of the germline. The posterior pole of the egg cytoplasm contains polar granules, particles that are seen by electron microscopy only in this region of the egg (Mahowald, 1962, 1968, 1971a, b). When the nuclei that occupy the posterior pole form pole cells they bud out with the polar granules. Embryos that lack polar granules, as a result of mutations in any one of a set of five genes that function during oogenesis, fail to form pole cells and develop into sterile, agametic, adults (Boswell and Mahowald, 1985; Lehmann and Nüsslein-Volhard, 1986; Schüpbach and Wieschaus, 1986). Conversely, when the posterior polar plasm including the polar granules is transplanted to ectopic locations, the nuclei at the transplantation site form cells that are determined for the germline fate (Illmensee and Mahowald, 1974; Niki, 1986). Thus, when the posterior pole plasm is experimentally situated at an ectopic location, germline-determined cells will form at the ectopic location, and when the posterior pole plasm is mutationally perturbed, germline-determined cells fail to form. The posterior polar plasm contains the determinant of the germline cell fate in *Drosophila.*

Other regions of the cellular blastoderm have been assigned characteristic cell fates with some precision by a number of techniques. These include observation of living embryos and of histological preparations (Poulson, 1950; Campos-Ortega and Hartenstein, 1985); selective destruction of cells using an ultraviolet laser microbeam (Lohs-Schardin *et al.*, 1979); statistical analysis of clone boundaries in genetic mosaics that arise from chromosome loss (Sturtevant, 1929; Garcia-Bellido and Merriam, 1969) or mitotic recombination; and transplantation of cells that are marked by microinjection with histochemical reagents such as horseradish peroxidase (Technau and Campos-Ortega, 1985; Hartenstein *et al.*, 1985).

One of the results of blastoderm fate mapping experiments is that eighty cells in the third cephalic segment are known to be committed to become the salivary glands: forty cells on each side of the embryo (Hartenstein *et al.*, 1985). As embryogenesis proceeds, the gland assumes its final form: it has two lobes formed by a single cell layer epithelium, the posterior ends of the lobes are closed off as simple sacks, and the anterior ends of the lobes are joined by a Y-shaped duct, the base of which empties into the pharynx.

The salivary gland is an organ specialized for the secretion of proteins, a specialization that first becomes apparent in ten-hour-old embryos. The cytoplasm of the salivary gland cells at this age can be seen in both the light microscope and in the electron microscope to be filled with secretion granules (Campos-Ortega and Hartenstein, 1985). During the third larval instar, the salivary gland, approximately 130 secretory cells in each lobe, is again engaged in the synthesis of secretory proteins that are first packaged into secretion granules (Korge, 1975, 1977; Zhimulev and Kolesnikov, 1975; Beckendorf and Kafatos, 1976). Late in the third larval instar, these

granules fuse with the salivary gland cell plasma membranes, resulting in the deposition of the proteins in the lumen of the gland. At the time of puparium formation, the contents of the lumen are expelled from the animal. As the prepupal period progresses, the salivary gland produces a third set of proteins that are secreted into the hemolymph of the sessile animal (Korge, 1977; Sarmiento and Mitchell, 1982). Of these three documented times of secretory protein synthesis, only the third larval instar secretion has a known function. This mixture of secreted proteins and glycoproteins, after expulsion from the larva, hardens to form a sticky mass that causes the animal to adhere to the surface it has crawled onto for the duration of pupal development (Fraenkel and Brooks, 1953; Lane *et al.*, 1972).

The *Salivary gland secretion* (*Sgs*), or glue protein, genes expressed in the salivary glands of *D. melanogaster* third instar larvae are an especially intriguing example of the developmental control of gene set activity. The set codes for proteins that are synthesized in only one tissue and at only one time in development (Korge, 1975, 1977; Beckendorf and Kafatos, 1976). The activation and the repression of gene set expression are under the control of ecdysterone. Salivary gland chromosomes undergo ecdysterone-regulated changes in chromosomal morphology whose relationship to ecdysterone-regulated changes in gene expression can be investigated.

The protein components of the third larval instar salivary gland secretion were identified by electrophoresis of total salivary gland proteins and of secretion masses isolated free of salivary gland tissue. Korge (1975) identified four protein components, and later showed that two of the components are glycosylated (Korge, 1977). Using a different separation method, Beckendorf and Kafatos (1976) resolved six protein components, four of which were glycosylated. Both of these prove to be

underestimates of the number of glue components. The glue proteins are synthesized only in the salivary gland and only during the latter half of the third instar (Korge, 1975, 1977; Beckendorf and Kafatos, 1976). Each glue protein is synthesized and accumulated during the third instar period to a prodigious amount: just before pupariation, one-third of the protein content of each salivary gland is present in the secretion mass in the lumen (Korge, 1977).

The structural genes for seven members of the glue protein gene set were identified by a combination of genetic means and nucleic acid molecular-biological means. The genetic approach exploited naturally occurring null alleles or electrophoretic variant alleles, meiotic recombination mapping of the variant alleles, and gene dosage studies using chromosome rearrangements. In this way, the *Sgs-1* gene was mapped to the left arm of the second chromosome, to the site 25B (Velissariou and Ashburner, 1980), the *Sgs-3* gene was mapped to the 68C site on the left arm of the third chromosome (Korge, 1975, 1977; Akam *et al.*, 1978), the *Sgs-4* gene was mapped to the X-chromosome location 3C (Korge, 1975, 1977), and *Sgs-6* was mapped to 71C, also on the left arm of chromosome three (Velissariou and Ashburner, 1981).

Over ten years ago, the then-new molecular cloning techniques were applied to the isolation of nucleic acid clones homologous with salivary gland-specific RNAs (Wolfner, 1980). Five homology classes of complementary DNA clone hybridized with salivary gland RNAs specific to the time of glue protein synthesis. Two of the five correspond to the genetically identified structural genes for *Sgs-4* (Muskavitch and Hogness, 1980) and *Sgs-3* (Meyerowitz and Hogness, 1982; Crowley *et al.*, 1983). Analysis of the other three homology classes led to the identification of the structural genes for *Sgs-5* (Guild and Shore, 1984), which is located at the third-chromosome

site 90BC, and those for *Sgs-7* and *Sgs-8* (Crowley *et al.*, 1983). The latter two glue protein genes form with *Sgs-3* a cluster of evolutionarily related genes, which occupies less than five thousand base-pairs of genomic DNA at the third-chromosome site 68C (Meyerowitz and Hogness, 1982; Garfinkel *et al.*, 1983). By both genetic and molecular criteria, the glue protein gene set exhibits both a dispersed and a clustered gene organization.

The facts about the glue protein gene set already mentioned—both dispersed and clustered gene organization, yet common tissue and time of expression—are sufficient to make the set an important tool for studying the basis of differential gene set activity. The biology of *Drosophila* makes these genes even more interesting. The salivary glands, like most other larval tissues, grow during larval life by increasing cell size. As the cell volume increases, nuclear DNA synthesis continues without mitosis. The result in the salivary gland is polytene chromosomes, chromosomes in which an average of 1000 chromatids per salivary gland cell nucleus come to be tightly apposed in register (reviewed by Berendes and Ashburner, 1978; Korge, 1987). Polytenized chromosome arms are marked by a pattern of alternating bands and interbands perpendicular to the axis of the chromosome. The banding pattern is characteristic of all polytene tissues in which the chromosome copy number is great enough to be analyzed with the light microscope. Upon this essentially invariant pattern of bands and interbands are superimposed local decondensations of chromatin, chromosome puffs, that are tissue-specific and developmental stage-specific (reviewed by Korge, 1987). Incubation of salivary glands with tritium-labelled ribonucleosides causes nascent RNA transcripts to become radioactive, and autoradiography of chromosome squash preparations of the treated tissue reveals the chromosome puffs to be sites at which transcription occurs (Pelling, 1964; Zhimulev and Belyaeva, 1975; Belyaeva and

Zhimulev, 1976). Chromosome puffing is taken to be a cytological manifestation of gene expression (Beermann, 1956).

Becker (1959) and Ashburner (1967, 1969) devised systems of nomenclature for the changing patterns of chromosome puffs in the *D. melanogaster* salivary glands. The Ashburner system recognizes twenty-one "puff stages" during the third instar and the prepupal period. Puff stage 1 involves a set of approximately ten puffs that can first be observed in the middle of the third larval instar, while the animal is still feeding, when the polytene chromosomes first become amenable to microscopic examination. The largest of these "intermolt puffs" are the five chromosomal sites (3C, 25B, 68C, 71C, and 90BC) that contain the glue protein structural genes already mentioned; glue protein synthesis occurs during a stage of salivary gland development when the corresponding chromosomal loci are in the puffed state. In puff stage 2, the intermolt puffs begin to regress, and several new puffs, among these 2B5 on the X-chromosome and 74EF and 75B on the left arm of the third chromosome, are induced. Subsequent puff stages are defined by characteristic sets of newly induced puffs and of regressing old puffs. The changes in the pattern of chromosome puffs are followed by changes in the pattern of protein synthesis (Tissières *et al.*, 1974; Zhimulev *et al.*, 1981; Poeting *et al.*, 1982). In addition to the question of how differential coordinate regulation of gene activity occurs, the glue protein gene set (and other uncharacterized salivary gland gene sets) present the question of the relationship between gene activity and chromosome morphology.

Several lines of evidence demonstrate that ecdysterone is the agent that regulates the puffing sequence in the larval and prepupal salivary glands. First, mutations have been isolated that are temperature-sensitive for the production of ecdysterone. Among

these mutations is *lethal(1)suppressor of forked*^{ts67g} (Dudick *et al.*, 1974). When larvae hemizygous for *l(1)su(f)*^{ts67g} are shifted to restrictive temperature late in the second instar or early in the third instar, chromosome puffing in the salivary glands is arrested at the intermolt stage (Hansson *et al.*, 1981). Second, the increase in hemolymph concentration of ecdysterone at the end of the third instar occurs at the time when the intermolt puffs begin to regress and the set of ecdysterone-induced early puffs are induced. After several hours, the ecdysterone-induced early puffs regress and are replaced by a complex series of ecdysterone-induced late puffs (Ashburner, 1967, 1969). Third, the temporal changes in chromosome puffs associated with the increase in ecdysterone concentration can be reproduced in explanted salivary glands that are cultured in the presence of the hormone (Ashburner, 1973). Finally, indirect immunofluorescence experiments that detect the location of ecdysterone, following photochemical activation of the hormone for crosslinking, show the hormone to be associated with several of the intermolt puffs when they are regressing, with early puffs when they are induced, and then with late puffs as they are induced and the early puffs regress (Gronemeyer and Pongs, 1980; Dworniczak *et al.*, 1983). Ecdysterone thus plays an important role in the regulation of salivary gland gene expression and chromosome activity, apparently by direct action on the genome, and presumably through the agency of a receptor protein specific for the hormone (Maroy *et al.*, 1978; Yund *et al.*, 1978).

Intermolt puff regression and the induction of the early puffs are "primary responses" to the application of the steroid. If inhibitors of protein synthesis are added to the culture medium along with ecdysterone, intermolt puff regression and early puff induction proceed. Therefore, the capacity of salivary glands to respond to the steroid in these ways already exists and does not require new gene expression (Ashburner,

1974). Drug inhibitor studies also showed that the regression of the early puffs and the induction of the late puffs both require ongoing RNA and protein synthesis; the inference is that early puff regression and late puff induction require a minimum of one early-puff gene product.

Additional evidence that early-induced gene products serve regulatory functions was obtained from segmental aneuploidy studies of the ecdysterone-induced early chromosome puffs at 74EF and 75B. Increased gene dosage resulted in an acceleration of early-puff regression and of late-puff induction; decreased gene dosage resulted in a delay of these changes in chromosome puffing (Walker and Ashburner, 1981). Genomic DNAs encompassing both puffs have been cloned, and the corresponding ecdysterone-induced transcription units identified and analyzed (K. Burtis, C.W. Jones, W. Segraves, C. Thummel and D.S. Hogness, unpublished). It has been suggested that protein products from both loci may be DNA-binding proteins that function in gene regulation by ecdysterone.

Ecdysterone-regulated gene expression requires the product of an X-chromosome locus as well. This conclusion comes from the study of the *lethal(1)non-pupariating-1* mutation (Kiss *et al.*, 1976, 1978) and other alleles of the overlapping complementation complex that maps to the ecdysterone-induced early puff at cytological location 2B5 (Belyaeva *et al.*, 1980). Animals carrying such mutations have salivary gland polytene chromosomes arrested in the intermolt puff stage, like *l(1)su(f)*^{ts67g} (Belyaeva *et al.*, 1981).

Aside from the intermolt puffs that harbor glue protein genes, and the heat-shock-induced puffs that contain genes coding for proteins synthesized in response to environmental stress (reviewed by Ashburner and Bonner, 1979), the apparent

regulatory functions of the 2B5, 74EF, and 75B loci are the only examples of genetic functions assigned to specific chromosome puffs.

The cytological evidence described above shows that ecdysterone is required for the regression of the intermolt puffs. That the hormone is also required for a concomitant cessation of glue protein gene transcription has been shown for the 68C intermolt puff. Salivary glands may be cultured *in vitro* in the presence of ecdysterone and ^{3}H-labelled ribonucleosides, and the newly synthesized radioactively labelled RNA recovered. The hormone-induced regression of the 68C intermolt puff is correlated with the rapid significant reduction of ^{3}H-ribonucleoside incorporation into 68C-homologous RNAs (Crowley and Meyerowitz, 1984). Beermann's view of changing polytene chromosome cytology as a manifestation of changing gene activity is supported by this experiment, as is the role of ecdysterone in the regulation of gene expression in the larval salivary gland.

That ecdysterone is required for the activation of glue protein gene expression was indicated by the phenotype of *l(1)su(f)*^{ts67g} larvae made hormone-deficient by growth at the restrictive temperature. These mutant larvae fail to make glue proteins even though the corresponding intermolt puffs are visible (Hansson *et al.*, 1981). In the salivary glands of *l(1)su(f)*^{ts67g} larvae at the restrictive temperature, RNA transcripts of *Sgs-3*, *Sgs-4*, *Sgs-7* and *Sgs-8* fail to accumulate, and glue protein gene transcription can be restored in these animals by feeding them ecdysterone (Hansson and Lambertsson, 1983).

A second agent required for glue protein gene activation is defined by the *l(1)npr-1* mutation. One aspect of the *l(1)npr-1* mutant phenotype is that intermolt puffs are present in the salivary gland chromosomes, but that expression of the three 68C

intermolt puff genes is blocked: neither RNA accumulation nor RNA synthesis assayed by radiolabelled nucleoside incorporation could be detected (Crowley *et al.*, 1984). Both *Sgs-4* and *Sgs-5* are expressed in *l(1)npr-1* larvae (Crowley *et al.*, 1984). Published reports indicate that ecdysterone is required for the activation of expression of all the glue protein genes tested, but only the 68C glue protein gene cluster is known to require the second regulatory factor coded by the *l(1)npr-1*$^{+}$ allele. Both the *l(1)su(f)*^{ts67g} and the *l(1)npr-1* phenotypes involve a dissociation of intermolt puff formation from glue protein gene transcription, and therefore the Beermann view of the relationship between chromosome puffing and gene expression is incomplete.

Two features of the 68C intermolt puff make it particularly interesting: the clustering of the three glue protein genes found there, and the dependence of their expression on the *l(1)npr-1*$^{+}$ gene product. The experiments reported in this dissertation address questions concerning the origin and functional significance of the clustering, the locations of the *cis*-acting sequences required for expression of two of the 68C glue protein genes, *Sgs-7* and *Sgs-8*, and indirectly address the mechanism of action of the *l(1)npr-1*$^{+}$ gene product required for expression of these genes.

Chapter 2 describes structural studies of the glue protein gene cluster in the 68C intermolt puff. The purposes of these studies were to obtain a comprehensive description of the genes and gene products of the cluster, and to try to identify potential *cis*-acting regulatory sequences that could account for the coordinated expression of these genes.

The sequence of 6751 base-pairs of genomic DNA that contain the three glue protein genes was determined. The RNA products were mapped by a combination of nuclease protection, primer extension sequence determination, and cDNA clone

sequence determination. The experiments precisely positioned the three transcription units on the genomic DNA sequence. *Sgs-8* is transcribed leftward, with a 69-nucleotide intervening sequence removed from the 422-nucleotide-long primary transcript. *Sgs-7* is transcribed rightward, with a 66-nucleotide intervening sequence removed from the 385-nucleotide-long transcript. *Sgs-3* is transcribed rightward, with a 73-nucleotide-long intervening sequence removed from the 1193-nucleotide-long primary transcript. The *Sgs-8* 5' end is separated from the *Sgs-7* 5' end by 475 base-pairs of non-transcribed intergenic DNA. The 3' end of *Sgs-7* is separated from the 5' end of *Sgs-3* by 1958 base-pairs of non-transcribed genomic DNA.

One of the results of this work was the discovery that the three glue protein genes at 68C are evolutionarily related to each other by gene duplication and subsequent sequence divergence, and form a diverged gene family. Each gene's single intervening sequence interrupts codon 10. All three predicted protein products share two structural features: a 23-amino-acid-long hydrophobic secretory leader peptide that is absent from each mature protein, and an approximately 50-amino-acid-long cysteine-rich carboxy-terminal region. These cysteine-rich segments are the mature sgs-7 and sgs-8 polypeptides. The *Sgs-3* gene differs from the other two in that it has a tandem-repetitious nucleotide region separating the hydrophobic-leader-coding region from the cysteine-rich-C-terminal-coding region. The tandem-repetitious nucleotides are translated to a tandem-repetitious peptide that is threonine-rich. The sgs-3 protein is known to be glycosylated (Korge, 1977), and the threonine residues are presumed to be the sugar attachment points.

One consequence of the evolutionary history of the *Sgs-3*, *-7*, *-8* gene family might be that each gene was duplicated along with its own set of *cis*-acting regulatory

sequences, and that such elements might retain sufficient sequence identity to be readily identified by inspection. The only readily recognizable sequence element upstream of all three 68C glue protein genes is the T-A-T-A box element located at approximately –30 base-pairs (Goldberg, 1979), an element that is found upstream of nearly all eukaryotic RNA polymerase II transcription units. Between the –44 and –91 base-pair positions relative to *Sgs-7* are sequences homologous to those located between –48 base-pairs and –93 base-pairs upstream of *Sgs-8*.

In Chapter 3, I describe experiments designed to identify *cis*-acting DNA sequences required for expression of the *Sgs-7* and *Sgs-8* genes. It had already been shown that the *cis*-acting sequences required for the expression of the *Sgs-3* gene do not include the *Sgs-7* and *Sgs-8* genes (Bourouis and Richards, 1985; Crosby and Meyerowitz, 1986; Vijay Raghavan *et al.*, 1986), supporting the view that the clustering of the three glue protein genes at 68C is a consequence of their evolutionary history only. By assaying fragments of the 68C cluster for *Sgs-7* and *Sgs-8* function, I determined whether or not these genes are regulated by separate *cis*-acting regulatory sequences. The specific hypothesis tested is whether or not the homologous sequences located within the first 100 base-pairs upstream of each of these genes represent conserved *cis*-acting regulatory elements that will allow each gene to be expressed independently of the other. The experiments show that this is not so. The interdigitated arrangement of *cis*-acting regulatory sequences observed suggests that the divergent transcription arrangement of the *Sgs-7*, *Sgs-8* gene pair is not simply an evolutionary vestige, but is functionally important for the regulation of these genes. In addition, the dependence of *Sgs-7* expression on the *l(1)npr-1*$^+$ gene product does not require sequences from the bulk of the *Sgs-7* transcription unit, a result that constrains the possible mechanism of action for the *l(1)npr-1*$^+$ gene product.

REFERENCES

Akam, M.E., Roberts, D.B., Richards, G.P and M. Ashburner (1978) *Cell* **13**: 215-225.

Ashburner, M. (1967) *Chromosoma* **21**: 398-428.

Ashburner, M. (1969) *Chromosoma* **27**: 47-63.

Ashburner, M. (1973) *Develop. Biol.* **34**: 47-61.

Ashburner, M. (1974) *Develop. Biol.* **39**: 141-157.

Ashburner, M. and J.J. Bonner (1979) *Cell* **17**: 241-254.

Beckendorf, S.K. and F.C. Kafatos (1976) *Cell* **9**: 365-373.

Becker, H.J. (1959) *Chromosoma* **10**: 654-678.

Beermann, W. (1956) *Cold Spring Harb. Symp. Quant. Biol.* **21**: 217-232.

Belyaeva, E.S. and I.F. Zhimulev (1976) *Cell Differen.* **4**: 415-427.

Belyaeva, E.S., Aizenon, M.G., Semeshin, V.F., Kiss, I.I., Koczka, K., Baritcheva, I.M., Gorelova, T.D. and I.F. Zhimulev (1980) *Chromosoma* **81**: 281-306.

Belyaeva, E.S., Vlassova, I.E., Biyasheva, Z.M., Kakpakov, V.T., Richards, G. and I.F. Zhimulev (1981) *Chromosoma* **84**: 207-219.

Berendes, H. and M. Ashburner (1978) In *The Genetics and Biology of Drosophila, 2B* (Ashburner, M. and T.R.F. Wright, eds.), pp. 453-498, Academic Press, New York.

Boswell, R.E. and A.P. Mahowald (1985) *Cell* **43**: 97-104.

Bourouis, M. and G. Richards (1985) *Cell* **40**: 349-357.

Campos-Ortega, J.A. and V. Hartenstein (1985) *The Embryonic Development of Drosophila melanogaster*, Springer-Verlag, Berlin, Heidelberg, New York, Tokyo.

Crosby, M.A. and E.M. Meyerowitz (1986) *Develop. Biol.* **118**: 593-607.

Crowley, T.E., Bond, M.W. and E.M. Meyerowitz (1983) *Mol. Cell. Biol.* **3**: 623-634.

Crowley, T.E., Mathers, P.H. and E.M. Meyerowitz (1984) *Cell* **39**: 149-156.

Crowley, T.E. and E.M. Meyerowitz (1984) *Develop. Biol.* **102**: 110-121.

Davidson, E.H. (1986) *Gene Activity in Early Development*, 3rd edit., Academic Press, New York.

Dudick, M.E., Wright, T.R.F. and L.E. Brothers (1974) *Genetics* **76**: 487-510.

Dworniczak, B., Seidel, R. and O. Pongs (1983) *EMBO J.* **2**: 1323-1330.

Fraenkel, G. and V.J. Brooks (1953) *Biol. Bull.* **105**: 442-449.

Garcia-Bellido, A. and J.R. Merriam (1969) *J. Exp. Zool.* **170**: 61-76.

Garfinkel, M.D., Pruitt, R.E. and E.M. Meyerowitz (1983) *J. Mol. Biol.* **168**: 765-789.

Goldberg, M.L. (1979) Ph.D. Dissertation, Stanford University, Stanford.

Gronemeyer, H. and O. Pongs (1980) *Proc. Natl. Acad. Sci., USA* **77**: 2108-2112.

Guild, G. M. and E.M. Shore (1984) *J. Mol. Biol.* **179**: 289-314.

Hansson, L. and A. Lambertsson (1983) *Mol. Gen. Genet.* **192**: 395-401.

Hansson, L., Lineruth, K. and A. Lambertsson (1981) *Wilhelm Roux' Arch.* **190**: 308-312.

Hartenstein, V., Technau, G.M. and J.A. Campos-Ortega (1985) *Wilhem Roux' Arch.* **194**: 213-216.

Illmensee, K. and A.P. Mahowald (1974) *Proc. Natl. Acad. Sci., USA* **71**: 1016-1020.

Kiss, I., Bencze, G., Fekete, E., Fodor, A., Gausz, J., Maroy, P., Szabad, J. and J. Szidonya (1976) *Theoret. Appl. Genet.* **48**: 217-226.

Kiss, I., Szabad, J. and J. Major (1978) *Molec. Gen. Genet.* **164**: 77-83.

Korge, G. (1975) *Proc. Natl. Acad. Sci., U.S.A.* **72**: 4550-4554.

Korge, G. (1977) *Develop. Biol.* **58**: 339-355.

Korge, G. (1987) In *Results and Problems in Cell Differentiation*, vol. 14, (W. Hennig, ed.), pp. 27-58, Springer-Verlag, Berlin, Heidelberg, New York, Tokyo.

Lane, N.J., Carter, Y.R. and M. Ashburner (1972) *Wilhelm Roux' Archiv.* **169**: 216-238.

Lehmann, R. and C. Nüsslein-Volhard (1986) *Cell* **47**: 141-152.

Lewin, B. (1974) *Gene Expression-1, Bacterial Genomes*, J. Wiley & Sons, New York.

Lohs-Schardin, M., Cremer, C. and C. Nüsslein-Volhard (1979) *Develop. Biol.* **73**: 239-255.

Mahowald, A.P. (1962) *J. Exp. Zool.* **151**: 201-216.

Mahowald, A.P. (1968) *J. Exp. Zool.* **167**: 237-262.

Mahowald, A.P. (1971a) *J. Exp. Zool.* **176**: 329-344.

Mahowald, A.P. (1971b) *J. Exp. Zool.* **176**: 345-352.

Maroy, P., Dennis, R., Beckers, C., Sage, B.A. and J.D. O'Connor (1978) *Proc. Natl. Acad. Sci., USA* **75**: 6035-6038.

Meyerowitz, E.M. and D.S. Hogness (1982) *Cell* **28**: 165-176.

Muskavitch, M.A.T. and D.S. Hogness (1980) *Proc. Natl. Acad. Sci., USA* **77**: 1041-1051.

Niki, Y. (1986) *Develop. Biol.* **113**: 255-258.

Pelling, C. (1964) *Chromosoma* **15**: 71-122

Poeting, A., Koerwer, W. and O. Pongs (1982) *Chromosoma* **87**: 89-102.

Postlethwait, J.H. (1974) *Biol. Bull.* **147**: 119-135.

Poulson, D.F. (1950) In *The Biology of Drosophila* (M. Demerec, ed.), pp. 168-274, J. Wiley & Sons, New York.

Richards, G. (1978) *Develop. Biol.* **66**: 32-42.

Sarmiento, L.A. and H.K. Mitchell (1982) *Develop. Genet.* **3**: 255-272.

Schüpbach, T. and E. Wieschaus (1986) *Wilhelm Roux' Arch.* **195**: 302-317.

Sturtevant, A.H. (1929) *Z. Wiss. Zool.* **135**: 323-356.

Technau, G.M. and J.A. Campos-Ortega (1985) *Wilhelm Roux' Arch.* **194**: 196-212.

Tissières, A., Mitchell, H.K. and U. Tracy (1974) *J. Mol. Biol.* **84**: 389-398.

Velissariou, V. and M. Ashburner (1980) *Chromosoma* **77**: 13-27.

Velissariou, V. and M. Ashburner (1981) *Chromosoma* **84**: 173-185.

Vijay Raghavan, K., Crosby, M.A., Mathers, P.H. and E.M. Meyerowitz (1986) *EMBO J.* **5**: 3321-3326.

Walker, V. and M. Ashburner (1981) *Cell* **26**: 269-277.

Wolfner, M.F. (1980) Ph.D. Dissertation, Stanford University, Stanford.

Yund, M.A., King, D.S. and J.W. Fristrom (1978) *Proc. Natl. Acad. Sci., USA* **75**: 6039-6043.

Zhimulev, I.F. and E.S. Belyaeva (1975) *Chromosoma* **49**: 219-231.

Zhimulev, I.F. and N.N. Kolesnikov (1975) *Wilhelm Roux' Archiv.* **178**: 15-28.

Zhimulev, I.F., Izquierdo, M.L., Lewis, M. and M. Ashburner (1981) *Wilhelm Roux' Archiv.* **190**: 351-357.

Chapter 2:

DNA Sequences, Gene Regulation and Modular Protein Evolution

in the *Drosophila* 68C Glue Gene Cluster

DNA Sequences, Gene Regulation and Modular Protein Evolution in the *Drosophila* 68C Glue Gene Cluster*

Mark D. Garfinkel, Robert E. Pruitt and Elliot M. Meyerowitz

Division of Biology, California Institute of Technology,
Pasadena, Calif. 91125, U.S.A.

(Received 16 February 1983, and in revised form 31 March 1983)
(Edited for the Journal of Molecular Biology *by I. Herskowitz)*

*The accompanying typescript is slightly modified from the version published in 1983 in the *Journal of Molecular Biology* **168**: 765-789. The data presented in Figure 4 can also be found in the National Institutes of Health DNA sequence database "GenBank" in the file named DROSGS378.

The 68C locus of the *Drosophila melanogaster* polytene chromosomes contains the structural genes for three glue polypeptides (sgs-3, sgs-7, and sgs-8) synthesized in the larval salivary glands during the third larval instar. When the messenger RNAs for the glue polypeptides are being synthesized, the locus is puffed; the puff regresses in response to the steroid hormone ecdysterone. The three 68C glue mRNAs are coded in a gene cluster of less than 5000 base-pairs, and are expressed coordinately. In the experiments described here we show that the coordinate expression of these RNAs does not result from amplification of the puff DNA, nor is it associated with puff DNA rearrangement. We also report the nucleotide sequence of 6751 base-pairs of genomic DNA that includes the entire gene cluster, and describe coding and non-coding sequences with possible regulatory roles. In addition, we deduce the amino acid sequences of the primary translation products of the glue mRNAs, and show that the glue proteins form a diverged gene family. The members of the family all contain an amino-terminal hydrophobic block of amino acids, which is absent in the mature, secreted glue proteins, and a cysteine-rich carboxyl terminal module. sgs-3 differs from sgs-7 and sgs-8 by containing a third module between the other two, comprised largely of tandem repeats of the five amino acids Pro-Thr-Thr-Thr-Lys.

1. Introduction

During the third larval instar, the major function of the *Drosophila melanogaster* salivary glands is the production of a set of secreted polypeptides (Zhimulev and Kolesnikov, 1975). There are at least eight of these (Crowley *et al.*, 1983), synthesized in the cytoplasm of the salivary gland cells throughout the instar and secreted into the lumen of the gland near the end of the larval stage. At the time of puparium formation the luminal contents are expelled, and they set to form a glue that fixes the puparium to its substrate for the duration of the pupal period (Lane *et al.*, 1972). A group of about ten puffs, or sites of highly active transcription, are present on the giant polytene chromosomes of the salivary gland cells when glue proteins are being synthesized; they disappear toward the end of the third larval instar, when glue synthesis terminates. These are known as the intermolt puffs (Ashburner, 1972). Genetic, cytogenetic and molecular mapping experiments have shown that at least four of these puffs contain structural genes for at least six of the glue polypeptides (Korge, 1975, 1977; Akam *et al.*, 1978; Velissariou and Ashburner, 1980, 1981; Crowley *et al.*, 1983). Other experiments have shown that the regression of the intermolt puffs is a consequence of an increase in the titer of the steroid hormone ecdysterone in the larval hemolymph several hours before pupariation (Ashburner, 1973). The regression of one of these puffs, that at 68C on the left arm of the third chromosome, appears to result directly from binding of the hormone (presumably through the mediation of a steroid receptor protein) to the puff (Gronemeyer and Pongs, 1980).

The molecular cloning of the 68C puff genomic DNA and of DNA complementary to the puff-encoded RNAs has been accomplished. Analysis using the cloned DNA has shown that one 5000 base-pair region of the puff DNA codes for three

different polyadenylated messenger RNAs, all found only in third larval instar salivary glands, and appearing and disappearing coordinately (Meyerowitz and Hogness, 1982). Each of the 68C RNAs, designated the group II, group III and group IV RNAs, is translated to a different salivary gland glue polypeptide, sgs-8, sgs-7 and sgs-3, respectively (Crowley *et al.*, 1983).

There are at least two features of the regulation of the 68C puff gene cluster that must be understood: the coordinate control of the three different RNAs, and the action of ecdysterone in puff regression. In the experiments described here, we test several hypotheses for the mechanism of coordinate regulation of the 68C glue RNAs, and in so doing find that the 68C glue proteins are evolutionarily related to each other in an unusual way. In addition, DNA sequence information obtained in these experiments constrains the possible types of regulatory DNA sequences that can be considered as important in coordinate regulation of the puff RNAs.

2. Materials and Methods

(a) *Insect culture*

Adult flies of the *D. melanogaster* third chromosome homozygous strain OR16f (Meyerowitz and Hogness, 1982) were reared in milk bottles or in population cages similar to the design of Elgin and Miller (1978) at 22°C. They were fed standard cornmeal-agar food that was supplemented with live yeast paste. Eggs were laid on food coated with yeast in plastic trays. The trays were covered with tight-fitting plastic boxes. Larvae were refed live yeast and were watered daily. Late third instar larvae were collected on days 5 and 6 after egg deposition.

(b) *Isolation of third instar salivary gland nucleic acids*

Third instar larvae were washed from the trays and boxes with cold distilled water. Food particles were removed by floating the larvae in 20% (w/v) sucrose. The clean larvae were then washed in Robb's (1969) PBS (phosphate-buffered saline), and crushed between metal rollers. Salivary glands and carcasses were collected on a fine-mesh Nitex (Tetko, Inc.) screen. Glands were separated from carcasses by filtration through a coarse-mesh Nitex screen. Salivary glands were collected in a plastic beaker. Fat bodies were removed from the glands by repeatedly allowing them to sediment at unit gravity. Gut, Malpighian tubules, and other tissues were removed from the glands by centrifugation through 32% Ficoll (Sigma) in Robb's PBS. Ficoll was removed by washing the glands with Robb's PBS. They were judged to be greater than 70% pure salivary glands.

RNA was extracted from the glands by dissolving the tissue in 0.1 M Tris-HCl (pH 8.0), 0.2 M NaCl, 0.1 M EDTA, 0.5% (w/v) sodium dodecyl sulfate. Repeated phenol, phenol/chloroform, and ether extractions were performed. The nucleic acids were precipitated with ethanol, washed and precipitated with ethanol again. Several mg of RNA were recovered from several hundred mg of tissue. RNA was stored in 10 mM sodium acetate (pH 5.0) at –80°C.

Poly $(A)^+$ RNA was obtained from salivary gland RNA by oligo(dT)-cellulose chromatography as described by Meyerowitz and Hogness (1982).

DNA was prepared from the glands by a modification of the procedure used by Meyerowitz and Hogness (1982) to obtain adult fly DNA. About 100 mg salivary gland tissue in Robb's PBS was treated with 1 ml 15% sucrose, 50 mM Tris-HCl (pH 8.0), 50 mM EDTA. The tissue was spun briefly in a hand-driven centrifuge and the supernatant discarded. One ml of 0.12 M sucrose, 150 mM Tris-HCl (pH 8.5), 75 mM EDTA, 0.75% sodium dodecyl sulfate was added to the tissue. Five µl 25% (v/v) diethyl pyrocarbonate in ethanol were added and the glands were lysed in a 2 ml Ten-Broeck homogenizer. The mixture was transferred to a 1.5 ml capped plastic tube, 65 µl 8 M potassium acetate were added, and the mixture allowed to stand in ice for 15 min. Precipitated debris and potassium dodecyl sulfate were removed by a 10-min. spin in a microcentrifuge. The supernatant was transferred to two 1.5 ml tubes and 1.1 ml ethanol were added to each tube. Nucleic acids were pelleted in the hand-driven centrifuge, rinsed twice with 70% (v/v) ethanol, and air-dried. Each pellet was resuspended in 20 µl 10 mM Tris-HCl (pH 8.0), 1 mM EDTA, and 100 ng pancreatic RNase A (a gift of D. Ridge) were added.

(c) *General DNA and recombinant DNA techniques*

Plasmids were grown in *Escherichia coli* HB101 using M9-Casamino acids supplemented with uridine as medium (Norgard *et al.*, 1979). Chloramphenicol amplification was sometimes used. Plasmid purification by CsCl/ethidium bromide gradient centrifugation was performed as described (Meyerowitz and Hogness, 1982).

Recombinant λ phage were propagated on *E. coli* K802 and were purified by standard methods (Maniatis *et al.*, 1978; Meyerowitz and Hogness, 1982). Phage DNA was extracted by the rapid formamide method "A" of Davis *et al.* (1980).

National Institutes of Health guidelines were followed for the P1-EK1 level containment of recombinant DNA-bearing organisms.

Preparation of *Drosophila* genome blot filters, nick-translation, and filter hybridization were performed as described by Meyerowitz and Hogness (1982).

(d) *DNA sequence determination by partial chemical cleavage*

(i) *End-labelling DNA*

After restriction enzyme digestion of plasmid DNA, one of three methods was used: for 5' protruding restriction site termini the 5' ends were labelled by dephosphorylation with calf intestinal alkaline phosphatase and subsequent rephosphorylation with [γ-^{32}P]ATP and T4 polynucleotide kinase (Maxam and Gilbert, 1980); 3' ends were labelled by incubating the DNA in 20 μM each of three unlabelled deoxynucleoside triphosphates, one [α-^{32}P] deoxynucleoside triphosphate, and *E. coli* DNA polymerase I Klenow fragment. For 3' protruding restriction sites the 3' ends

were labelled with [α-^{32}P]CTP and terminal deoxynucleotidyl transferase as described by Roychoudury and Wu (1980).

Fragments labelled at one end were obtained by digestion with a second restriction enzyme.

(ii) *Gel purification of labelled DNA*

The ^{32}P-labelled DNAs were resolved on polyacrylamide gels crosslinked with N,N'-bis-acrylyl-cystamine (Bio-Rad). Fragments were located by autoradiography, excised from the gel, and were released from the gel matrix by adding 2-mercaptoethanol to a concentration of 50% (v/v). The disulfide crosslinks are reduced after 0.5 to 2 h at room temperature. Nine ml of 0.1 M Tris-HCl (pH 7.5), 0.1 M NaCl were added and each gel slice was homogenized by thorough vortex mixing. The viscous mixes were incubated with 0.2 ml Whatman DE52 DEAE-cellulose resin, which was kept suspended by constant agitation. After several hours at room temperature, the radioactive resins were pelleted in a table top centrifuge, washed twice with 10 ml 0.1 M Tris-HCl (pH 7.5), 0.1 M NaCl to remove acrylamide residue, and were packed into small columns. DNA was eluted from the resin with three 1 ml volumes of 0.1 M Tris-HCl (pH 7.5), 1 M NaCl. Fine resin particles were pelleted by centrifugation, and the DNA precipitated from the supernatant by adding 10 μg yeast transfer RNA and 9 ml ethanol, followed by incubation overnight at –20°C. Recoveries generally exceeded 90%.

(iii) *Limited modification of bases*

The Maxam and Gilbert (1977, 1980) procedure modified by Smith and Calvo (1980) forms the basis of our sequence determination protocol. Base modification

conditions were chosen to enable us to read up to 650 nucleotides from a labelled end. The G+A reaction used only 2 µl 1 M pyridinium formate in a 22 µl volume at 37°C for 5 min. Hydrazine reactions (C, C+T) were performed in an ice water bath for 15 min. The dimethyl sulfate G reaction was done for 12 to 15 min. in an ice water bath using 0.125% (v/v) dimethyl sulfate. Stop solution, ethanol precipitation, and ethanol rinse steps were done as described (Maxam and Gilbert, 1980).

(iv) *Gel electrophoresis and autoradiography*

Sequence gels were 0.36 mm thick, and contained 100 mM-Tris/borate/EDTA (Maxam and Gilbert, 1980) and 50% (v/v) urea. To read nucleotides 1 to 50, a 40-cm long 20% polyacrylamide gel was run at 40 W constant power until the xylene cyanol marker had migrated 12 cm. To read nucleotides 35 to 650, multiple staggered runs on 80-cm long 5% polyacrylamide gels were performed. The gels were run at 2400 to 2800 V constant potential. Repeated loadings on one gel, or several gels run for different times, were used such that the xylene cyanol marker was allowed to migrate 30 cm, 90 cm, 150 cm, or 210 cm. This pattern of electrophoresis allowed for facile alignment of overlapping contiguous sequence.

Gels were transferred from the glass plates to sheets of Whatman 3MM paper or scrap X-ray film, covered with plastic wrap, and autoradiographed. Kodak XR-5 or XAR-5 film was used. Duplicate exposures of the 150 cm run and 210 cm run gels were done with or without DuPont Cronex Lightning-Plus intensifier screens. All autoradiographs were read independently by two persons. Discrepancies were resolved by reference to the original films, and by additional sequence determinations. Except for the leftmost 70 nucleotides, which were determined once, every position was

assigned on the basis of at least two independent experiments. See Fig. 3 for additional details.

(e) *Nuclease mapping of mRNAs*

^{32}P-labelled restriction fragments (about 50,000 cts/min.) were mixed with 20 µg yeast tRNA for mock hybridizations, or with 20 µg yeast tRNA and 2 µg salivary gland poly(A)$^+$ RNA. The nucleic acids were precipitated with ethanol, rinsed, and dried. Hybridizations were carried out by dissolving the nucleic acids in 100 µl of 70% deionized formamide, 10 mM PIPES (pH 6.4), 0.4 M NaCl, 0.1 mM EDTA (Casey and Davidson, 1977), heating to 70°C and annealing at 50.0°C for several hours. While leaving the hybridizations at 50.0°C, 15-µl portions were removed and diluted into 200-µl portions of nuclease assay buffer. The assay buffer tubes were pre-chilled in ice water baths, and enzyme was already added to the appropriate tubes. Rapid transfer of the hybridization portions, forceful pipetting, rapid vortex mixing, and immediate transfer to the digestion temperature all ensured that strand displacement was minimized.

The nuclease S_1 reaction was 0.3 M sodium acetate (pH 4.5), 0.4 M NaCl, 0.1 mM zinc acetate, 30 µg heat-denatured salmon sperm DNA/ml at 37°C for 15 min. The reaction was terminated by adding 600 µl ethanol and freezing on solid CO_2. The precipitated nucleic acids were pelleted, washed with 70% ethanol, dried and resuspended in Maxam and Gilbert (1980) sequence gel sample buffer. *Aspergillus oryzae* nuclease S_1 was obtained from Sigma Chemical Co.

Exonuclease VII digestions were carried out in 10 mM Tris-HCl (pH 7.5), 20 mM KCl, 10 mM EDTA (Berk and Sharp, 1978) at 45°C for 45 min. The reaction was

terminated by adding 20 µl 3 M sodium acetate (pH 5.0) and 600 µl ethanol. Nucleic acids were precipitated, washed, and resuspended as described for the nuclease S_1 samples. *E. coli* exonuclease VII was obtained from Bethesda Research Laboratories.

Portions of the initial ^{32}P-labelled restriction fragments were subjected to the partial chemical degradation sequence reactions. The nuclease digests were run alongside the sequence size standards on 5% polyacrylamide/urea gels which were 80 cm long.

(f) *Primer extension sequence determination*

These experiments were done according to the Ghosh *et al.* (1980) method as modified by Snyder *et al.* (1982). ^{32}P-labelled restriction fragment was mixed with 0.5 to 0.8 mg salivary gland RNA, precipitated with ethanol, rinsed, dried and hybridized in 200 µl of 70% formamide as described for nuclease mapping. The hybridization mix was then diluted by adding 1.5 ml oligo(dT) binding buffer. [^{32}P]DNA-poly(A)$^+$ RNA hybrids were recovered using oligo(dT)-cellulose chromatography. The hybrids were eluted, precipitated with ethanol, rinsed, dried and resuspended in reverse transcriptase buffer. Avian myeloblastosis virus reverse transcriptase (a gift from J. Beard) was added and the reaction incubated at 37°C for 3 h. NaOH was added to 0.1 M, and RNA hydrolyzed for 1 h at 37°C. The reaction was neutralized, extracted with phenol, extracted with chloroform, and precipitated twice with ethanol. The complementary DNA was rinsed with ethanol, dried and resuspended in water. The complementary DNA was divided into five batches, four for the sequence determination reactions and one as a standard.

(g) Bal*31 deletion construction*

Two µg of aDm2023 plasmid DNA were linearized by complete digestion with *Xho*I in a 20-µl reaction. After digestion, 12 µl water and 8 µl 5X *Bal*31 buffer (1X = 20 mM Tris-HCl, pH 8.1, 12 mM $CaCl_2$, 12 mM $MgCl_2$, 0.6 M NaCl, 1 mM EDTA) were added. 0.41 unit of nuclease *Bal*31 (Bethesda Research Laboratories) was added and allowed to react at 30°C for 10 min. The reaction was terminated by adding 13 µl 200 mM EGTA, followed by extractions with phenol and chloroform. The DNA was precipitated with ethanol and resuspended in 10 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 100 mM NaCl, 6 mM 2-mercaptoethanol, 100 µg gelatin/ml. Each deoxynucleoside triphosphate was added to 1 mM, 1.4 units *E. coli* DNA polymerase I Klenow fragment were added, and the reaction incubated at room temperature for 15 min. The enzyme was heat-inactivated, and the reaction diluted to 200 µl which included 10 mM dithiothreitol, 1 mM ATP, 550 ng *Eco*RI linkers (a gift from C. K. Itakura) and T4 DNA ligase (a gift from S. Scherer). The ligation reaction proceeded at room temperature overnight, and was then used to transform *E. coli* HB101 to ampicillin resistance. Transformants were colony purified, and retested for drug resistance. Small overnight cultures were grown and rapid plasmid isolations carried out. The resulting plasmid DNAs were digested with *Eco*RI and fractionated on a 1.2% agarose gel. The clone designated aDm2023Δ23 had about 950 base-pairs of *Drosophila* DNA, centered on the original *Xho*I site, removed. A large-scale preparation of this plasmid DNA was used for sequence determination.

3. Results

(a) *Three ways the 68C cluster is not regulated*

It is clear from the transcription map of the 68C gene cluster that the three mRNAs cannot derive from a common precursor, and thus that the tight coordination of expression of these RNAs does not result from their sharing a single promoter (Meyerowitz and Hogness, 1982).

A second way in which coordinate regulation of clustered genes can be accomplished is by amplification of the chromosomal region containing the genes at the time of their expression (Spradling and Mahowald, 1980; Spradling, 1981). A third mechanism by which coordinate expression might be triggered is by a DNA rearrangement in the chromosomal DNA of the expressing tissue (Brack *et al.*, 1978; Zieg *et al.*, 1978; Seidman *et al.*, 1979). To determine if the 68C glue gene cluster has undergone differential amplification or DNA rearrangement in the tissue of its expression in third instar larvae, high molecular weight DNA was isolated from adult flies and from third instar larval salivary glands. Equivalent amounts of each DNA preparation were digested with the restriction endonucleases *Eco*RI, *Hin*dIII and *Sal*I, subjected to electrophoresis in an agarose gel, and blotted to a nitrocellulose filter (Southern, 1975). The filter was hybridized with ^{32}P-labelled λaDm1501-10 DNA (Meyerowitz and Hogness, 1982); this phage contains 18.2 kb[1] of contiguous genomic DNA, including the 68C glue gene cluster (Fig. 1). The pattern of restriction fragment sizes and the autoradiographic intensities of hybridization of each fragment were identical in the adult and salivary gland lanes (Fig. 2); and the sizes of restriction

[1]Abbreviations used: kb, 10^3 base-pairs; cDNA, complementary DNA.

fragments found were the same as those measured in clones such as λaDm1501-10 that are derived from embryonic DNA. The same filter was washed of ^{32}P-labelled probe and sequentially rehybridized with two other ^{32}P-labelled probes: aDm2026, a plasmid containing the 1.65 kb *Hin*dIII fragment that includes the coding DNA for the two small, divergently transcribed RNAs II and III, and aDm2023, a plasmid with the 2.4 kb *Sal*I fragment that contains the coding DNA for the large RNA IV (Fig. 1). These hybridizations also showed no differences between adult DNA, either in restriction fragment size or extent of probe hybridization.

(b) *DNA sequence of the 68C cluster region*

One possible mechanism for the coordinate regulation of the three 68C RNAs is that each RNA has its own equivalent of a bacterial operator, and that the operators are identical, or nearly so, and therefore respond identically to the cellular signals that control puff transcription. To seek such DNA regions, the sequence of 6751 contiguous nucleotides of the *D. melanogaster* genomic DNA clone λbDm2002, containing the coding sequences of all three of the 68C glue RNAs, was determined. Figure 3 shows a restriction map of the sequenced DNA, the relative positions of the three glue RNAs in it, and the sequencing strategy used. All but the leftmost 70 nucleotides of the sequence were determined on each of the complementary DNA strands, and complementarity was observed everywhere, although at 11 *Eco*RII sites [C-C-(A/T)-G-G], in which methylation occurs at the inner cytosine residue in *dcm*$^+$ *E. coli*, there was a gap rather than a band in the C and C+T lanes of the sequencing gel (Ohmori *et al.*, 1978). Agreement between the sequence and the experimentally determined sites of digestion by 11 hexanucleotide-recognizing restriction endonucleases was perfect, except that three *Cla*I sites indicated by the sequence were

not cleaved by this enzyme. Each of the resistant sites overlaps the sequence G-A-T-C, where the *Cla*I site is A-T-C-G-A-T. G-A-T-C is a site for adenine methylation mediated by the *E. coli dam*$^+$ methylase (Geier and Modrich, 1979). Methylation of adenine in the *Cla*I recognition site appears to be sufficient to prevent the restriction activity of the enzyme, as predicted by Backman (1980). The 68C nucleotide sequence is presented in Figure 4. Analysis of chromosomal rearrangements with breakpoints near the sequenced DNA indicates that the sequence is presented in telomeric to centromeric order (E. M. Meyerowitz and M. A. Crosby, unpublished results).

The beginning of the sequence, nucleotides 1 through 463, is a sequence of a part of a transposable element of the repetitive *roo* family, that is present in our *D. melanogaster* Oregon-R strain just to the left of the glue-coding gene cluster (Meyerowitz and Hogness, 1982). That these nucleotides are in the transposable element was determined by comparison of the sequence of a clone from our 68C *roo*-containing strain (aDm2024, Fig. 1) to a homologous genomic clone (aDm2003, Fig.1, sequence not presented) from a wild-type chromosome which does not have a *roo* element adjacent to the 68C glue puff. The *roo* sequence presented here differs in only five positions from the similar sequence determined by Scherer *et al.* (1982) for their *B104* transposable element family which, as demonstrated in their paper, is equivalent to the *roo* family.

The nucleotides in positions 464 through 874 are DNA that is unique to the 68C puff region. Following this is one element of an inverted repeat sequence, from nucleotides 875 to 1159. The complementary element extends from positions 2853 to 2569. This pair of elements has been observed as a stem and loop structure in electron microscopic analysis of melted and reannealed DNA from λbDm2002 (Meyerowitz and

Hogness, 1982); the elements flank the 3' ends of the group II (sgs-8) and group III (sgs-7) RNAs. The sequence shows that each element is 285 base-pairs in length, and that the two elements are complementary at 93% (266) of their nucleotide positions (Fig. 5). The boundaries of the elements are distinct. Since no similar element appears adjacent to the 3' end of the third 68C RNA, the group IV sgs-3 RNA, it seems unlikely that this sequence is responsible for the coordinate expression of all three RNAs. This conclusion is supported by further observations on the inverted repeat elements: when DNA containing one or both of them is ^{32}P-labelled and used to probe plaque filters or colony filters containing DNA of λ or cosmid libraries of *D. melanogaster* genomic fragments, clones containing one region of the fly genome in addition to the 68C puff are obtained. This region contains three additional copies of the repeat element in 6 kb of contiguous sequence. The three copies are direct repeats, all in the same relative orientation. When clones containing this region are ^{3}H-labelled by nick-translation and hybridized to salivary gland polytene chromosomes (Pardue *et al.*, 1970), autoradiography shows the origin of the three additional elements to be in the 68C region, but clearly proximal to the 68C 3 to 5 position of the glue puff, and in a chromosomal area that is not puffed when the glue puff is present. ^{32}P-labelled probes containing the three proximal elements and the adjacent sequences do not give detectable signals when hybridized to RNA gel blots containing third larval instar salivary gland RNA. Likewise, ^{32}P-labelled cDNA made from third instar salivary gland polyadenylated RNA does not hybridize strongly to the DNA of the clones containing these elements in DNA gel blot experiments. Thus, not only does one 68C glue gene lack an adjacent repeat element, but three repeat elements exist in a chromosomal region that is unpuffed in third instar salivary glands, and that does not

contain coding sequences for any abundantly expressed third instar salivary gland RNA species.

The sequenced DNA between the two puff inverted repeat elements contains the coding DNA for the divergently transcribed sgs-8 and sgs-7 protein mRNAs. The region to the right of the repeat pair includes sequences 5' to, including, and 3' to the third puff mRNA, that coding for sgs-3 protein.

(c) *Localization of RNA-coding DNA sequences*

To analyze the functional relevance of the sequences found in the glue gene cluster, the 5' and 3' limits of mRNA coding regions, and the exact positions of any intervening sequences found within the RNAs were determined. These experiments began with the sequencing of cDNA clones representing each of the 68C puff mRNAs (the group II, group III, and group IV cDNA clones of Meyerowitz and Hogness, 1982). These sequences are presented in Figure 6. Nuclease protection experiments (Berk and Sharp, 1977, 1978) were then performed on the group II (sgs-8) and group III (sgs-7) mRNAs, to confirm and extend the sequence observations; while primer extension experiments (Ghosh *et al.*, 1980) were performed with the group IV, sgs-3 mRNA for the same purpose. The results for each of the mRNAs are presented in turn, starting with the leftmost.

(i) *sgs-8*

The DNA sequence of the group II cDNA clone is homologous with the genomic DNA sequence from positions 1215 at the 3' end of the RNA transcript, to position 1605 at the 5' end. Genomic nucleotides 1510 through 1578 are absent from

the cDNA clone, indicating a 69 nucleotide intervening sequence near the 5' end of the RNA transcript. The 3' end of the gene can only be localized to bases 1215 to 1218; the cDNA sequence is polydeoxyadenylate beginning at 1217 and the genomic positions 1215, 1216 and 1217 are A residues. Thus, RNA termination and poly(A) addition could occur after any of positions 1218 through 1215.

To confirm the presence of an intervening sequence in the sgs-8 mRNA, and to establish an approximate location for the 5' start of transcription, nuclease protection experiments were performed. The 404 base-pair *Xba*I-*Eco*RI fragment that includes nucleotides 1310 through 1713, and that extends from the middle of the *Sgs-8* gene to upstream of the 5' end, was labelled at the *Xba*I site on the strand complementary to the sgs-8 mRNA using [γ-^{32}P]ATP and T4 polynucleotide kinase, then annealed to total poly(A)$^+$ RNA from third instar larval salivary glands. The hybrid was treated with either the single strand-specific nuclease S_1, or with *E. coli* exonuclease VII (Fig. 7). After nuclease S_1 digestion, the labelled fragments that remained extended to positions 1510 to 1513, as determined by polyacrylamide gel electrophoresis adjacent to size standards made by performing sequencing reactions on the intact labelled *Xba*I-*Eco*RI fragment. After correcting for the different 3'-terminal moieties in the reaction products in the experimental and size-standard lanes (Sollner-Webb and Reeder, 1979), the 3' end of one of the major nuclease S_1 truncation products was seen to coincide with position 1510, thus confirming the presence of an intervening sequence. Exonuclease VII digests single-stranded DNA processively from 5' and 3' termini (Chase and Richardson, 1974). Treatment of the RNA-DNA hybrids with this nuclease reduces the labelled DNA fragment to lengths of 336, 337, and 338 nucleotides, the 3' ends of these fragments aligning with genomic sequence positions 1645 to 1647. Since exonuclease VII leaves undigested approximately five unpaired nucleotides extending

from RNA-DNA hybrids (Donahue *et al.*, 1982; Contreras *et al.*, 1982) the 5' terminus of the sgs-8 mRNA is very near position 1640. Controls for the nuclease digestion experiments included omitting salivary gland RNA from the hybridization reaction, in which case no labelled DNA fragment was protected from digestion, and omitting nuclease treatments, in which case the 404 base-pair starting DNA fragment was recovered intact.

(ii) *sgs-7*

The DNA sequence of the group III cDNA clone includes nucleotide positions 2164 to 2498 in the genomic clone sequence, with positions 2175 to 2240 missing. The absent sequence indicates that the sgs-7 mRNA contains a 66 nucleotide intervening sequence near its 5' end. There is no poly(A) tract at the end of the cDNA insert that represents the 3' end of the mRNA, but the genomic sequence from positions 2499 to 2516 is a tract of 18 consecutive A residues. It thus seems likely that the polyadenylation site of the RNA is coded between bases 2498 and 2516 of the DNA sequence, and the possibility exists that up to 18 residues of the poly(A) tail on this RNA are added transcriptionally, rather than post-transcriptionally.

Confirmation of the intervening sequence position, and establishment of the 5' end of the RNA were done in experiments similar to those used for the sgs-8 RNA. In this case a 683 base-pair *Msp*I-*Eco*RI fragment (positions 2400 to 1718), labelled at the *Msp*I site by use of [γ-^{32}P]ATP and T4 polynucleotide kinase, was used in hybridization to poly(A)$^+$ third instar larval salivary gland RNA. The results of the nuclease S_1 digestion experiments confirmed the intervening sequence location derived from the cDNA clone sequence; the exonuclease VII experiments placed the nucleotide coding for the 5' end of the RNA very near the *Kpn* I site at position 2112. It is worth

noting that all of the sequences between the 5' ends of the divergently transcribed *Sgs-8* and *Sgs-7* genes total less than 500 base-pairs.

(iii) *sgs-3*

The group IV cDNA clone, representing the sgs-3 RNA, was only partially sequenced. The sequence of the end derived from the 3' terminus of the RNA contained a poly(A) tract adjacent to genomic nucleotide position 5646, indicating that this is the poly(A) addition site of the RNA. The 5' end of the RNA is not represented in the cDNA clone. Nuclease protection experiments performed by K. Burtis and D. Hogness (personal communication) suggested that a small intervening sequence exists in the sgs-3 RNA near position 4550. To confirm this result, and to determine the precise size and location of this sequence, a primer extension experiment was done. An 111 base-pair *Hae*III-*Hha*I genomic DNA fragment, containing nucleotide positions 4725 through 4835 and ^{32}P-labelled at the *Hae*III site (4385) using T4 polynucleotide kinase, was prepared. This labelled DNA was melted and annealed to total third instar salivary gland RNA, and the poly(A)+ RNA-primer DNA hybrids collected by oligo(dT)-cellulose column chromatography. The primed RNA was then incubated in a reaction mixture containing deoxynucleoside triphosphates and reverse transcriptase, and the sequence of the resulting end-labelled cDNA determined. The sequence showed, first, that there is an intervening sequence present in genomic DNA but absent from sgs-3 mRNA, including 73 genomic nucleotides (positions 4514 to 4586). In addition, the strongest site of primer extension termination was genomic nucleotide position 4457, indicating that the 5'-terminal nucleotide of the mRNA is coded at or near this position.

(d) *Search for potential coordinate control sequences*

With the positions of mRNA coding sequences established, a search for duplicated sequences found in the same position relative to each of the RNAs was made. The only such sequences are found near and including the 5' ends of, and extending into the three genes. They are shown in Figure 8. Three alignment points are used in the Figure: the intervening sequence 5' boundary, the 3' boundary of this sequence, and the sequences near the 5' termini of the three RNA coding regions. In the vicinity of the transcription initiation region of each of the three genes is a conserved oligonucleotide (C/T)-A-T-C-(T/A)-G-(G/T) which has been observed at the 5' ends of other *Drosophila* genes (Snyder *et al.*, 1982). Approximately 30 nucleotides upstream of the 5' end of each gene is a T-A-T-A sequence (Goldberg, 1979), which has been shown to be required for correct initiation site selection by RNA polymerase II in other eukaryotic genes (Grosschedl and Birnstiel, 1980; Benoist and Chambon, 1981; McKnight *et al.*, 1981). In addition, the *Sgs-8* and *Sgs-7* gene regions share homology at two upstream locations, underlined in Figure 8, and extending almost 100 base-pairs 5' of the transcription initiation points. The *Sgs-3* gene region does not have either of these sequences. Within the RNA coding regions of the three genes there are considerable homologies. The 5' untranslated regions are all similar in their DNA sequence, as are the nucleotides flanking the translation initiation codons, and those coding for the first ten amino acids. The consensus splicing donor sequences (Lerner *et al.*, 1980; Sharp 1981) following the first ten codons are also homologous. The intervening sequences that follow show no detectable homology, until five nucleotides upstream of their 3' ends, where the 3' ends of these sequences and the consensus splicing acceptor signals are again similar. The *Sgs-8* and *Sgs-7* genes show about 40 additional nucleotides of sequence similarity downstream of their intervening

sequences; the *Sgs-3* gene is homologous to the other two for only the first 19 nucleotides of this region. The next substantial homology is in the 3' untranslated region, where about 20 nucleotides upstream of their 3' termini each of the genes possesses the A-A-T-A-A-A sequence implicated in polyadenylation or transcription termination in many eukaryotic sequences (Proudfoot and Brownlee, 1976). The inverted repeat elements begin 58 base-pairs downstream of the sgs-8 mRNA 3' end, and 70 base-pairs 3' of the sgs-7 mRNA terminus. It is clear that the *Sgs-8* and *Sgs-7* genes are partly homologous throughout much of their lengths, and that, along with some 5' sequences and their 3' inverted repeat elements, they comprise a large, inexact inverted repetition.

(e) *Protein products of the 68C glue genes*

The amino acid sequences of the proteins coded at the 68C glue puff have been determined from the mRNA coding sequences described above (Fig. 9). Translation of eukaryotic mRNAs usually begins at the methionine codon nearest the 5' end (Kozak, 1978); following this codon each of the 68C mRNAs has a long open reading frame. In all three cases, the intervening sequence occurs between the first and second nucleotides of codon ten. The first 23 amino acids of each reading frame contain a high proportion of hydrophobic residues; these are a signal peptide removed from the primary translation products before their secretion as glue proteins (Crowley *et al.*, 1983). The sgs-8 reading frame continues for another 52 amino acids, that of sgs-7 for 51 more residues. The sgs-3 protein contains 284 amino acids following the signal sequence, the carboxy-terminal 50 of which are similar in sequence to the secreted sgs-8 and sgs-7 proteins, which are quite similar to each other. In the region at the carboxy end of the proteins eight cysteine residues and 11 other amino acid positions are

identical in all three proteins. sgs-3 is different from the two small proteins by the presence of a 234 amino acid long segment between the leader peptide and the carboxyl-terminal amino acids. This extra segment contains an amino-terminal 49 amino acids of threonine-rich sequence, followed by 185 amino acids that are composed entirely of 37 tandem repeats of minor variants of a five amino acid unit (Fig. 10). The basic unit is Pro-Thr-Thr-Thr-Lys. The variations on this sequence, the different 15 base-pair sequences that code for the five amino acid repeats, and the higher-order repeats made of these 15 base-pair sequences are also shown in Figure 10.

4. Discussion

(a) *Possible regulatory sequences*

It is clear from the experiments described that the basis of coordinate control of the three 68C RNAs does not lie in the processing of the RNAs from a common precursor, or in DNA amplification. It is also clear that large-scale DNA rearrangements are unlikely to be associated with the coordinate expression of the 68C glue RNAs. DNA sequencing of the 68C puff was initiated as a means of testing the hypothesis that three coordinately regulated genes of similar function achieve their coordination through possession of identical stretches of regulatory sequences. The only sequences found to be shared by the three glue protein genes and their surrounding sequences are T-A-T-A sequences and regions of protein coding sequence within the genes. T-A-T-A regions are not candidates for specific control sequences, since they are a common feature of many eukaryotic genes transcribed by RNA polymerase II. It is possible that the highly conserved 5' translated regions of the three genes, which include the nucleotides surrounding each intervening sequence, are involved in coordinating the levels of the three glue protein RNAs. In particular, mechanisms of coordination involving common pathways of RNA splicing mediated by proteins that recognize shared sequences can be envisioned. It seems just as likely, though, that the nucleic acid sequence conservation in this region of the RNAs is a result of selection for the amino acid sequence of the hydrophobic signal peptide. No experiments have been done to differentiate between these two hypotheses.

No sequences were found which present themselves as obvious candidates for mediators of coordinate transcription. There are three possible reasons for this. One is

that the three genes are not truly coordinately expressed, but are regulated by different mechanisms that operate only in salivary glands, and at similar enough times that their activities have not yet been distinguished. A model in which *Sgs-8* and *Sgs-7* are controlled by nearly identical regulatory sequences, but *Sgs-3* by different sequences, would be consistent with our sequence observations, and would implicate the common upstream elements of the two small genes as regulatory sequences. The second possibility is that the three genes are controlled identically by control sequences of identical function, but that we cannot recognize these sequences. This failure of recognition could occur because control sequences of identical function do not necessarily have the same DNA sequence, because the regions are too small to appear significant to us, or because the DNA sequence that comprises the regions is not composed of contiguous nucleotides, but consists of required bases separated by other, inconsequential nucleotides. The final possible reason for our failure to find identical control sequences associated with each of the 68C glue genes is that all three genes are controlled by a single set of sequences in or near the gene cluster, that affects the transcribability of the entire region, perhaps by initiating puffing as a precondition to transcription. All of these possibilities are currently being tested in this laboratory, using DNA-mediated transformation of *Drosophila* embryos (Rubin and Spradling, 1982) to assay the function of various separate fragments of the 68C gene cluster.

(b) *Protein structure and evolution*

The sequence of the DNA coding for each of the three 68C mRNAs has allowed us to predict the amino acid sequences expected of the 68C protein products. These amino acid sequences have already enabled purification and identification of the 68C proteins, and the demonstration that they are all secreted glue polypeptides (Crowley *et*

al., 1983). The amino acid sequences show that the three 68C proteins are related to each other as a clustered gene family. The members of this family show modular construction: each member has a 23 amino acid amino-terminal portion composed largely of amino acids with hydrophobic side-chains, that is not present in the mature, secreted form of the protein. Each also has a cysteine-rich carboxy-terminal set of about 50 amino acids which show considerable sequence homology between the proteins, there being 19 positions in which all three have the same residue. The sgs-8 and sgs-7 polypeptides have only these two modules. sgs-3 contains a third module positioned between the other two, and consisting of 234 amino acids. A total of 128 of these are threonine, with much of the modules (185 residues) consisting of tandem repeats of the sequence Pro-Thr-Thr-Thr-Lys, or of sequences slightly diverged from this canonical one. The amino-terminal module probably serves as a signal peptide for protein secretion (Crowley *et al.*, 1983). The function of the carboxy-terminal module is unknown. One function of the threonine-rich module of the sgs-3 protein may be to provide a site for attachment of sugars. sgs-3 is extensively glycosylated *in vivo* (Beckendorf and Kafatos, 1976; Korge, 1977). It is unlikely that the site of carbohydrate attachment is asparagine, since the target sequence for asparagine glycosylation *via* the dolichol phosphate pathway is Asn-X-Thr or Asn-X-Ser (Staneloni and Leloir, 1982), and neither of these sequences appears in sgs-3. This leaves serine and threonine as possible sites of sugar attachment. There are only three serine residues in the processed sgs-3 polypeptide, while there are 128 threonine residues, all in the central module. An example of a protein extensively glycosylated by virtue of sugar attachment to numerous threonine residues is an antifreeze serum protein found in Antarctic fish (Feeney and Yeh, 1978), which is modified at almost every threonine by addition of a disaccharide to the threonine hydroxyl group.

The only other *Drosophila* glue polypeptide for which sequence information is published is sgs-4, transcribed from a locus found at 3C on the polytene chromosomes (Muskavitch and Hogness, 1982). Comparison of sgs-4 amino acid or nucleic acid sequences reveals no homology between this glue polypeptide and those coded at 68C. In one respect the sgs-4 protein is similar to sgs-3: both contain substantial regions comprised of tandem repeats of a small number of amino acids. In the case of sgs-4, the repeat unit contains the seven residues Thr-Cys-Lys-Thr-Glu-Pro-Pro. Similar periodic repeats are found in a number of proteins from different sources, including silk fibroin from the moth *Bombyx mori* (Sprague *et al.*, 1979; Gage and Manning, 1980; Manning and Gage, 1980), eggshell proteins of another moth, *Antherea polyphemus* (Jones *et al.*, 1979), and zein, the seed storage protein of maize (Geraghty *et al.*, 1981; Pedersen *et al.*, 1982).

The similarities in amino acid sequence between the three 68C gene products reflect similarities in the nucleotide sequences of the genes. Since the divergence at the DNA level is substantial enough to preclude cross-hybridization under our relatively non-stringent conditions of filter hybridization and washing, nucleotide sequencing experiments were necessary to discover the relation of the three glue genes. In fact, the three members of the the 68C glue gene family are so dissimilar that the order of the gene duplication events that presumably gave rise to the family cannot be deduced. As can be seen from Table 1, sgs-8 and sgs-7 appear more closely related to each other than either is to sgs-3 when the hydrophobic amino-terminal module is considered, but sgs-7 and sgs-3 are more closely related to each other than to sgs-8 when the carboxy-terminal cysteine-rich modules are compared.

The most striking feature of the evolution of this gene family is the appearance of the threonine-rich central module in sgs-3 (or its disappearance from sgs-7 and sgs-8). The gain (or loss) of this module is not mediated by intervening sequences at its termini, since the intervening sequences in the 68C genes are all in the middle of the amino-terminal module. Appearance (or disappearance) of the sgs-3 central module is also not due to tandem duplication, or deletion of tandem duplications: the module is unrelated to any of the other sequences of the gene. The possible importance of appearance and disappearance of modules in the evolution of structural proteins is pointed out by the employment of modular evolution in at least one other gene family, the *A. polyphemus* egg chorion proteins (Jones *et al.*, 1979). In this instance, two modules found in the middle, and at the carboxy-terminal end of the B class of chorion proteins are also found in members of the A chorion protein class. The sequences surrounding these modules are not shared between the two protein classes. In these proteins the shared regions are partly composed of tandem repeats of oligopeptides, exactly as is the threonine-rich central region of sgs-3. Studies of the 68C puff proteins in species of *Drosophila* other than *melanogaster* may help to answer some of the questions relating to evolutionary mechanisms raised by modular evolution of the 68C glue polypeptides.

We thank Dr. C. M. Rice for teaching us DNA sequencing methods, T. Hunkapiller for computer programs and instruction in their use, M. Douglas for aiding in preparing computer-generated Figures, and Dr. Leroy Hood for the use of his computer facility. We also thank Dr. G. Scherer, K. Burtis and H. Nick for communicating unpublished results, and Drs. S. Scherer and M. Snyder for discussion and advice. Two of us (M.D.G. and R.E.P.) are predoctoral fellows of the National Science Foundation. This work was supported by grant 1 RO1 GM 28075 awarded to E.M.M. by the National Institute of General Medical Sciences, National Institutes of Health, and by a National Research Service Award (1 T32 GM 07616), also from the National Institute of General Medical Sciences, National Institutes of Health.

REFERENCES

Akam, M.E., Roberts, D.B., Richards, G.P and M. Ashburner (1978) *Cell* **13**: 215-225.

Ashburner, M. (1972) *Chromosoma* **38**: 255-281.

Ashburner, M. (1973) *Develop Biol.* **35**: 47-61.

Backman, K. (1980) *Gene* **11**: 169-172.

Beckendorf, S.K. and F.C. Kafatos (1976) *Cell* **9**: 365-373.

Benoist, C. and P. Chambon (1981) *Nature (London)* **290**: 304-310.

Berk, A.J. and P.A. Sharp (1977) *Cell* **12**: 721-732.

Berk, A.J. and P.A. Sharp (1978) *Proc. Natl. Acad. Sci., U.S.A.* **75**: 1274-1278.

Brack, C., Hirama, M., Lenhard-Schuller, R. and S. Tonegawa (1978) *Cell* **15**: 1-14.

Casey, J. and N. Davidson (1977) *Nucl. Acids Res.* **4**: 1539-1552.

Chase, J.W. and C.C. Richardson (1974) *J. Biol. Chem.* **249**: 4545-4552.

Contreras, R., Gheysen, D., Knowland, J., van de Voorde, A. and W. Fiers (1982) *Nature (London)* **300**: 500-505.

Crowley, T.E., Bond, M.W. and E.M. Meyerowitz (1983) *Mol. Cell. Biol.* **3**: 623-634.

Davis, R.W., Botstein, D. and J.R. Roth (1980) *Advanced Bacterial Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Donahue, T.F., Farabaugh, P.J. and G.R. Fink (1982) *Gene* **18**: 47-59.

Elgin, S.C.R. and D.W. Miller (1978) In *The Genetics and Biology of Drosophila, 2A* (Ashburner, M. and T.R.F. Wright, eds.), pp. 112-121, Academic Press, New York.

Feeney, R.E. and Yeh, Y. (1978) *Advan. Protein Chem.* **32**: 191-282.

Gage, L.P. and R.F. Manning (1980) *J. Biol. Chem.* **255**: 9444-9450.

Geier, G.E. and P. Modrich (1979) *J. Biol. Chem.* **254**: 1408-1413.

Geraghty, D., Peifer, M.A., Rubenstein, I. and J. Messing (1981) *Nucl. Acids Res.* **9**: 5163-5174.

Ghosh, P.K., Reddy, V.B., Piatak, M., Lebowitz, P. and S.M. Weissman (1980) *Methods Enzymol.* **65**: 580-595.

Goldberg, M.L. (1979) Ph.D. Dissertation, Stanford University, Stanford.

Gronemeyer, H. and O. Pongs (1980) *Proc. Natl. Acad. Sci., U.S.A.* **77**: 2108-2112.

Grosschedl, R. and M.L. Birnstiel (1980) *Proc. Natl. Acad. Sci., U.S.A.* **77**: 1432-1436.

Jones, C.W., Rosenthal, N., Rodakis, G.C. and F.C. Kafatos (1979) *Cell* **18**: 1317-1332.

Korge, G. (1975) *Proc. Natl. Acad. Sci., U.S.A.* **72**: 4550-4554.

Korge, G. (1977) *Develop. Biol.* **58**: 339-355.

Kozak, M. (1978) *Cell* **15**: 1109-1123.

Lane, N.J., Carter, Y.R. and M. Ashburner (1972) *Wilhelm Roux' Archiv.* **169**: 216-238.

Lehninger, A.L. (1975) *Biochemistry*, 2nd edit., Worth Publishers, Inc., New York.

Lerner, M.R., Boyle, J.A., Mount, S.M., Wolin, S.L. and J.A. Steitz (1980) *Nature (London)* **283**: 220-224.

Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G.K. and A. Efstratiadis (1978) *Cell* **15**: 687-701.

Manning, R.F. and L.P. Gage (1980) *J. Biol. Chem.* **255**: 9451-9457.

Maxam, A.M and W. Gilbert (1977) *Proc. Natl. Acad. Sci., U.S.A.* **74**: 560-564.

Maxam, A.M and W. Gilbert (1980) *Methods Enzymol.* **65**: 499-560.

McKnight, S.L., Gavis, E.R., Kingsbury, R. and R. Axel (1981) *Cell* **25**: 385-398.

Meyerowitz, E.M. and D.S. Hogness (1982) *Cell* **28**: 165-176.

Muskavitch, M.A.T. and D.S. Hogness (1982) *Cell* **29**: 1041-1051.

Norgard, M.V., Emigholz, K. and J.J. Monahan (1979) *J. Bacteriol.* **138**: 270-272.

Ohmori, H., Tomizawa, J. and A.M. Maxam (1978) *Nucl. Acids Res.* **5**: 1479-1485.

Pardue, M.L., Gerbi, S.A., Eckhardt, R.A. and J.G. Gall (1970) *Chromosoma* **29**: 268-290.

Pedersen, K., Devereux, J., Wilson, D.R., Sheldon, E. and B.A. Larkins (1982) *Cell* **29**: 1015-1026.

Proudfoot, N.J. and G.G. Brownlee (1976) *Nature (London)* **263**: 211-214.

Robb, J.A. (1969) *J. Cell Biol.* **41**: 876-884.

Roychoudhury, R. and R. Wu (1980) *Methods Enzymol.* **65**: 43-62.

Rubin, G.M. and A.C. Spradling (1982) *Science* **218**: 348-353.

Scherer, G., Tschudi, C., Perera, J., Delius, H. and V. Pirrotta (1982) *J. Mol. Biol.* **157**: 435-454.

Seidman, J.G., Max, E.E. and P. Leder (1979) *Nature (London)* **280**: 370-375.

Sharp, P.A. (1981) *Cell* **23**: 643-646.

Smith, D.R. and J.M. Calvo (1980) *Nucl. Acids Res.* **8**: 2255-2274.

Snyder, M., Hunkapiller, M., Yuen, D., Silvert, D., Fristrom, J. and N. Davidson (1982) *Cell* **28**: 1027-1040.

Sollner-Webb, B. and R.H. Reeder (1979) *Cell* **18**: 485-499.

Southern, E.M. (1975) *J. Mol. Biol.* **98**: 503-517.

Spradling, A.C. (1981) *Cell* **27**: 193-201.

Spradling, A.C. and A.P. Mahowald (1980) *Proc. Natl. Acad. Sci., U.S.A.* **77**: 1096-1100.

Sprague, K.U., Roth, M.B., Manning, R.F. and L.P. Gage (1979) *Cell* **17**: 407-413.

Staneloni, R.J. and L.F. Leloir (1982) *CRC Crit. Rev. Biochem.* **12**: 289-326.

Velissariou, V. and M. Ashburner (1980) *Chromosoma* **77**: 13-27.

Velissariou, V. and M. Ashburner (1981) *Chromosoma* **84**: 173-185.

Zhimulev, I.F. and N.N. Kolesnikov (1975) *Wilhelm Roux' Archiv.* **178**: 15-28.

Zieg, J., Hilmen, M. and M. Simon (1978) *Cell* **15**: 237-244.

Table 1

Nucleotide and amino acid sequence homologies in the 68C glue polypeptide genes

Pairwise Comparison	Nucleotide identities[1]	Amino acid identities[1]	Amino acid similarity[2]
A. *Hydrophobic amino termini*			
sgs-8—sgs-7	53/69 (77%)	15/23 (65%)	20/23 (87%)
sgs-8—sgs-3	45/69 (65%)	11/23 (48%)	17/23 (74%)
sgs-7—sgs-3	44/69 (64%)	11/23 (48%)	17/23 (74%)
B. *Cysteine-rich carboxy termini*			
sgs-8—sgs-7	81/150 (54%)	20/50 (40%)	27/50 (54%)
sgs-8—sgs-3	88/150 (59%)	27/50 (54%)	38/50 (76%)
sgs-7—sgs-3	101/150 (67%)	28/50 (56%)	38/50 (76%)

[1]No gaps were introduced into either sets of alignments.

[2]Amino acid similarity includes both amino acid identities and "conservative" substitutions of amino acids with functionally similar side-chains (Lehninger, 1975).

Figure 1. λ clones and plasmid subclones used in this study. A restriction map of the studied region of the 68C puff is given, showing the position of the *roo* transposable element found in the OR16f chromosome. Genes II, III, and IV are indicated as filled boxes. Cleavage sites for the restriction enzymes *Eco*RI (R), *Hin*dIII (H) and *Sal*I (S) are shown. Extents of the genomic DNA clones are indicated. λaDm1501-10 and aDm2003 have been described (Meyerowitz and Hogness, 1982). aDm2023 contains the 2.4 kb *Sal*I fragment homologous to the *Sgs-3* gene inserted into the pBR322 *Sal*I site. aDm2024 contains the 5.7 kb *Sal*I fragment adjacent to that cloned in aDm2023. aDm2026 contains the 1.65 kb *Hin*dIII fragment that includes *Sgs-7* and *Sgs-8* inserted into the pBR322 *Hin*dIII site. aDm2027 contains the 0.53 kb *Hin*dIII fragment adjacent to that present in aDm2026. fDm9014 contains the 1.6 kb *Pvu*I fragment that includes *Sgs-3* inserted into the *Pvu*I site of pBR325. These five plasmid subclones were prepared from λbDm2002 (Meyerowitz and Hogness, 1982) by routine subcloning procedures.

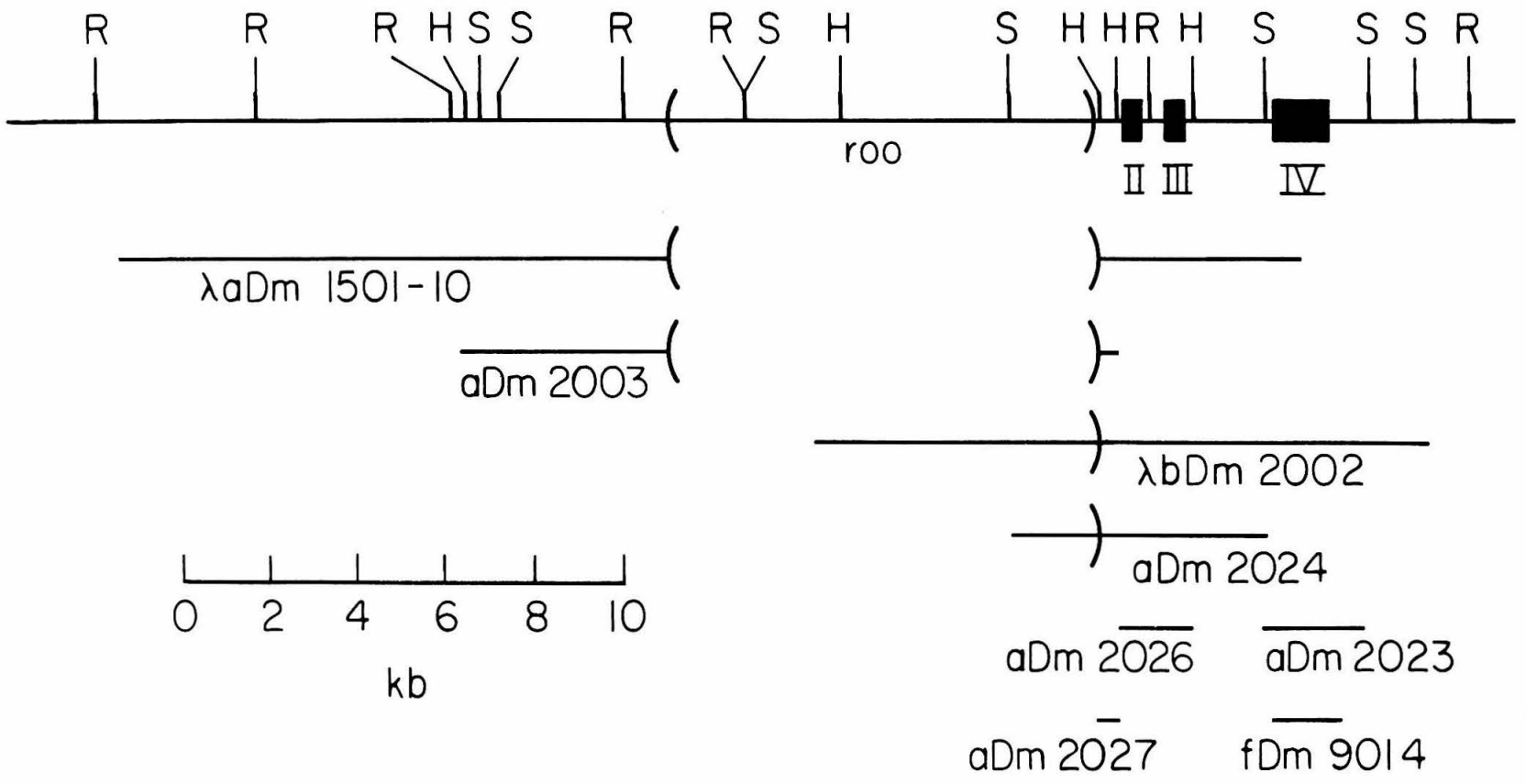

R R R H S S R R S H S H H R H S S S R
roo
II III IV
λaDm 1501-10
aDm 2003
λbDm 2002
aDm 2024
0 2 4 6 8 10
kb
aDm 2026
aDm 2023
aDm 2027
fDm 9014

Figure 2. 68C cluster region DNA is neither amplified nor rearranged in third instar larval salivary glands. High molecular weight DNAs from *D. melanogaster* strain OR16f adult flies (lanes 1, 3 and 5) or third instar larval salivary glands from the same strain (lanes 2, 4 and 6) were digested with *Eco*RI (lanes 1 and 2), *Sal*I (lanes 3 and 4), or *Hin*dIII (lanes 5 and 6). The digested DNAs were subjected to electrophoresis through a 0.9% agarose gel, blotted to nitrocellulose, and hybridized with ^{32}P-labelled λaDm1501-10 DNA. The size standard used was λ*c*I857 *S*7 DNA digested with *Hin*dIII. Autoradiographs of the genome blot filter were scanned with a Joyce-Loebl microdensitometer. Peaks from the *Eco*RI digest lanes were cut out and weighed. The ratio of adult fly DNA peaks to the corresponding salivary gland DNA peaks ranged from 0.9 to 1.1. To control for the preferential polytenization of euchromatic DNA in the salivary gland, the genome blot filter was washed and rehybridized with ^{32}P-labelled aDm2040 (E. M. Meyerowitz, unpublished results). This genomic clone derives from near 68C 10-11, is unique in the genome and is not detectably transcribed in third instar larval salivary glands. Microdensitometry of the resulting autoradiograph gave a ratio of adult fly DNA to salivary gland DNA of 0.8.

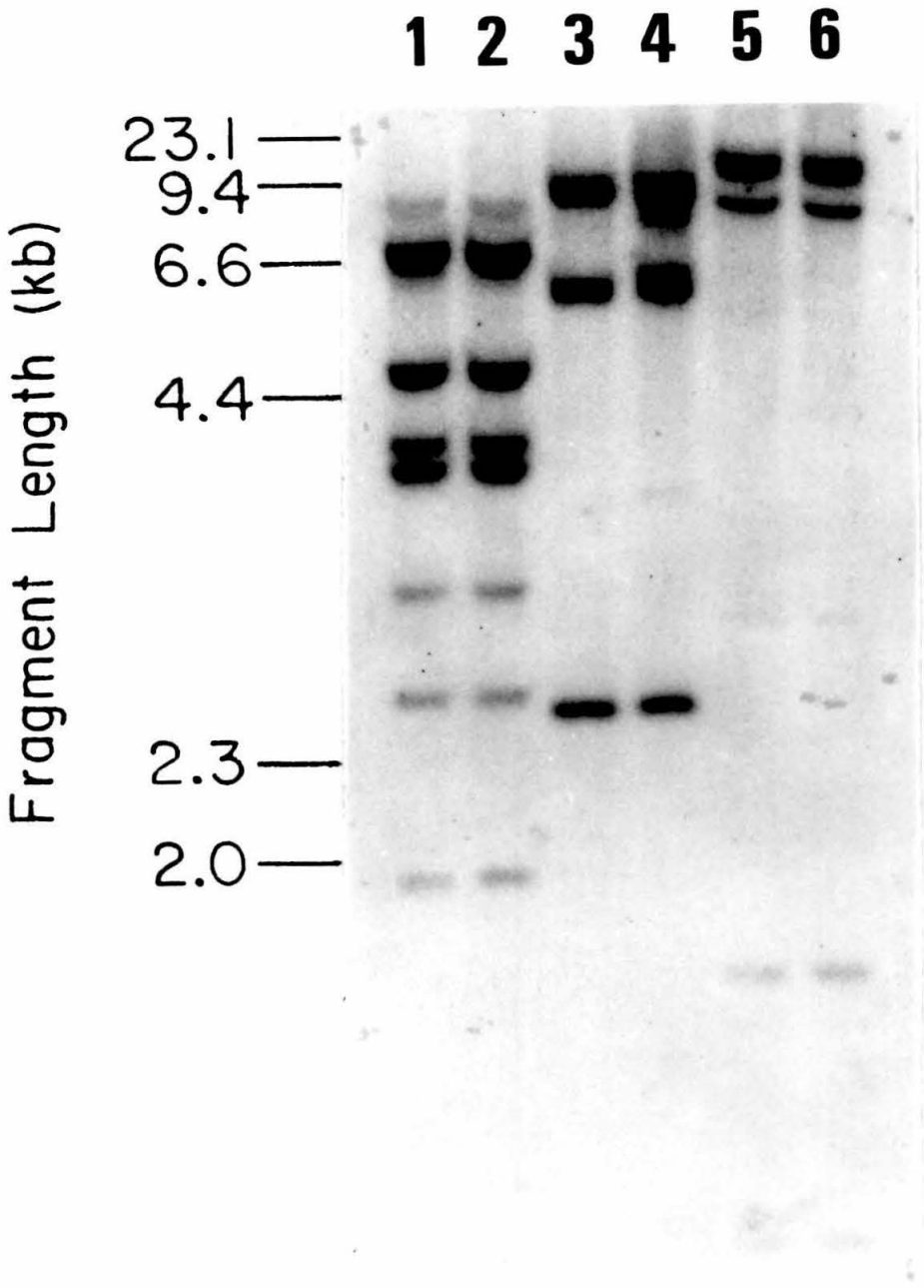

1 2 3 4 5 6
23.1
9.4
6.6
4.4
2.3
2.0
Fragment Length (kb)

Figure 3: Sequence determination strategy. Restriction enzyme cleavage sites are abbreviated as follows: B, *Bgl*II; C, *Cla*I; *Hin*dIII; K, *Kpn*I; P, *Pst*I; Pv, *Pvu*I; R, *Eco*RI; S, *Sal*I; Sc, *Sac*I; X, *Xba*I; Xh, *Xho*I. The two triangles indicate the extents of the inverted repeat elements. Close-parenthesis marks the right-hand boundary of the *roo* transposable element (Meyerowitz and Hogness, 1982) inserted at 68C in the fly stock used in these experiments. Below the map are indicated the directions and extents of transcription of the three 68C genes expressed in third instar larval salivary glands (Meyerowitz and Hogness, 1982). The small downward-pointing carats indicate the locations of the intervening sequences present in the genes. Below this is a summary of the sequence determination experiments. Arrowheads point in the directions sequence data were read. The base of each arrow is aligned with the labelled restriction site. Plain arrows are sequences read 3' to 5' using DNA polymerase I Klenow fragment-labelled DNA. Circles represent restriction sites labelled with T4 polynucleotide kinase; these sequences were read 5' to 3'. The arrow labelled 2003 indicates a sequence determined from the genomic subclone aDm2003 used to establish the point of insertion of the *roo* transposable element. The arrow labelled Δ23 represents sequence data obtained from the deletion clone aDm2023Δ23 (see Materials and Methods). The *Xho*I site was read across by sequence determination from a nearby *Hae*III site. The left *Sal*I site was read across from the adjacent *Pvu*I site. All restriction sites used as origins of sequence data were thus read across. Determinations were done on both strands everywhere except the leftmost 70 nucleotides of the *roo* transposable element. The scale is in base-pairs.

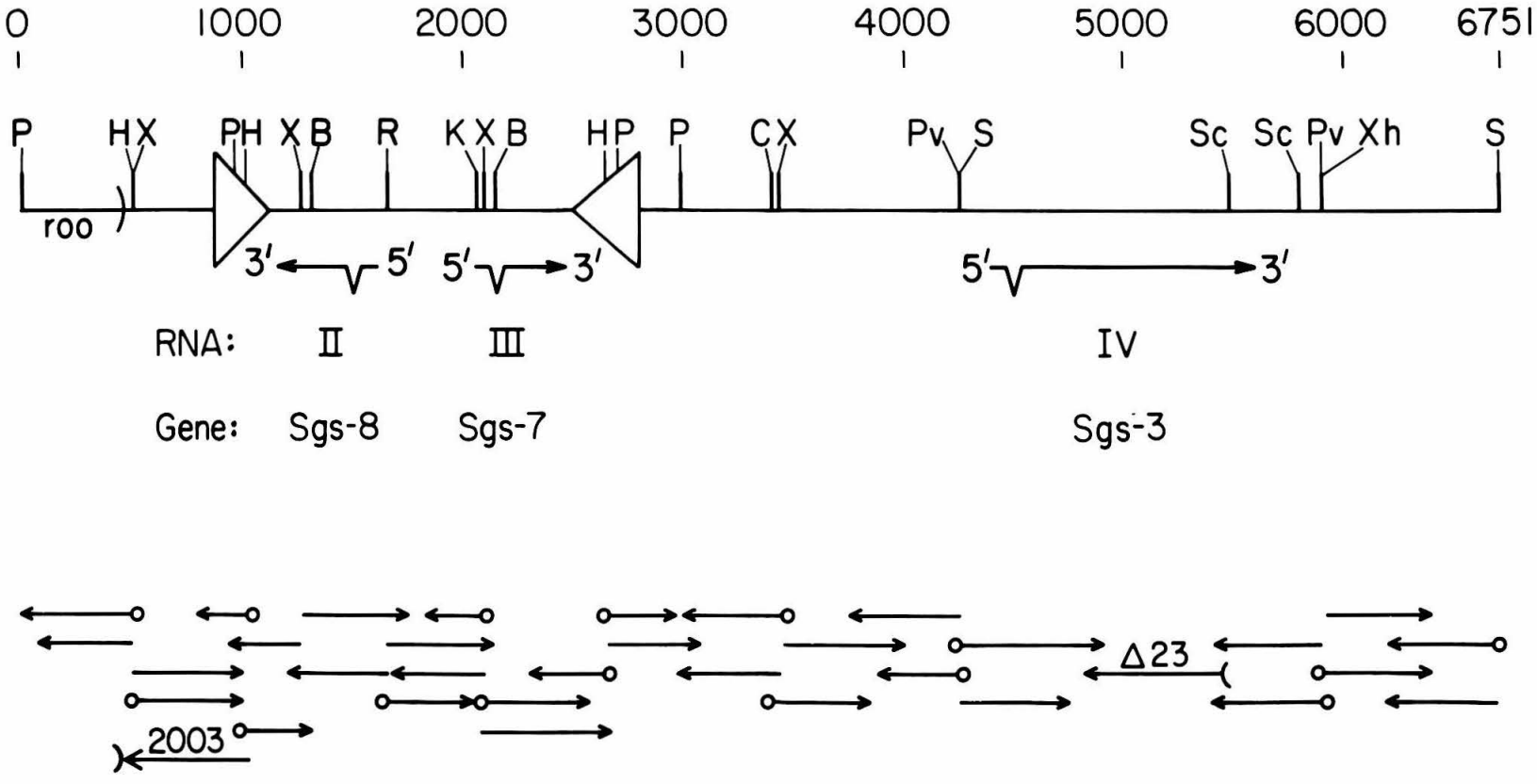

0
1000
2000
3000
4000
5000
6000
6751
P
HX
PH
XB
R
KXB
HP
P
CX
Pv
S
Sc
Sc
Pv
Xh
S
roo
3′
5′
5′
3′
5′
3′
RNA:
II
III
IV
Gene:
Sgs-8
Sgs-7
Sgs-3
Δ23
2003

Figure 4. The complete DNA sequence of the 68C cluster region. The sequence is written 5' to 3' left to right. The strand shown is the template for gene II transcription (positions from approximately 1645 to 1218), and is congruent with RNA III (approx. positions 2112 to 2498) and RNA IV (approx. positions 4457 to 5646). Protein-coding nucleotides are capitalized. The restriction sites shown are identical to the map in Fig. 3, with three additions: *Msp*I at 2399, *Hha*I at 4726, and *Hae*III at 4835. The *roo* transposable element begins at position 463 and continues leftward. The inverted repeat elements occupy positions 875 to 1159 and 2853 to 2569.


| Pst I |
1 ctgcagagtcgattaaaggctcagattagaccaaatgtaaaatcccagataagaagactttactcgttgnctttttgtaagaaactgattttatttggaa
101 atatcttcggtttaaataggtgacatgagaatcgcatcttaaagtaaatggcctacgcagaggcctaagtaaatagtccccgccttatcgaggtcccacg
201 ctcgggcacatctgcctatcttgagcggcgaggaccttatctgtggtctcccactaagggactattttaggaggcggggaacgatctcaagtgactgact
301 catgtagtgtgcacttaaattacatgtttttgagcaatgcacccatgtcgccttagataacaaaatcctaaatataatttatcgctctcgattcatttac
401 ataagatatgaacggagcccaaaattgtaagtctttaaatatattcgtgttcatgtgtgaacaaaatgtgaacaaatttaatgtgttttttttcttctggg
|Hind III| Xba I |
501 taatattcttccaaaagcttctagaatattttgaaatatttttttatttctgttgcagtgttttgacattgaaatacaatcgtcgtgggtaaacattattg
601 catacttttagggaatttttttaggaatagtaacccatttaattgttgatattacaaatatgtcaacaaaaccgtctttgcttactgttcaaacatttg
701 attattaaatgcttttgctattgtactcgctttttatgtgcttatgcgttatgggaagcaaaaacatagtaaagcgaatgcggaaatctgcttgttttgc
801 tataaaagcagtgggtatgattttaattaatcacaacacctgatcgttcaattccagtcgtatcgaaaaatctaatcttatttcaaaatgaagtatcttt
901 tcgtggttgcccttattgccctggccattcaggttgcgtcatctgctagtaccacaacaaccacggatgccaccaccacaacaacaaccactactgcagc
| Pst I |
| Hind III|
1001 atcaaccacaacaacaacaacggccgcttccactcacaaaaagctttgttggaggggcaacaattggtgccacacccgcattcccaagcggaagtgcaag
1101 aatccaaagaggtgccacaagaccatcgtgattgtgacacacagaaaagactagttctgacacttttaagcaatacatctgaaaataaataattaaaaa
1201 acaaagaaaaggagtttgggtcgtgcaggaatgataatagttttattagacagcagttttaaatgccagcagataaaaggtcagtgccgattaggttgag
| Xba I | | Bgl II |
1301 ggcacatctctaGAACAGCCAAACCTGCTCTCCGCAGGCGCACTGACGCACCTGCCGCTCAAGACCCTCCATAATGTTGATCAGATCTTTGCAGACGGGA
1401 ACCCGTGCGGAACACCCAGGACACGGCTCGCCACCAGGTCCACAAATCACGCATGAACAATCCTTGCAGCCCGAGGCAGGATCGGCGAATCCGATGAGCA
1501 TGATGCACGctgggtcagaatgaacggatggctaagtggatttgggtcacataaaaaatagtacttttggcacttacCAATGACGGCGACAACGAGCAG
1601 CTTCATggttgttgctttaacaaattaactttaccagatggtaaccgtttatgaacaccctaccccttttatagcaaaacaaatgtgttataggatcaat
|Eco RI |
1701 ggaaatttcattgaattcatccaaaaataaaatatataaccatttgtgcttaagcaaatagaaacacgatattaaacttcgccctttgttctcaccattt
1801 tctgtgtcatcgttcatactaatataatataacattttacatgccctttttactaaagaaagtattactcataaaatgaaatctaaattatatctgagta
1901 acaaatatattaaattaataagtatctataaaaagttaattctataaataaagcgcctgccgtataaaaagccaagtgtttggtgttttatttattttaa
2001 tacaattggtttgtccagtacttttattttttggatgtgctcactgaaatttccattgatccagctaacttttgcgctatataaaggtgttgctttcc
| Kpn I | | Xba I | | Bgl II |
2101 ttgagttggtaccatctggtaaagtagtctcaatctagatagaaccATGAAACTGATCGCAGTCACCATCATCGgtaactacataataagatctttaatc
2201 cacaaccaacttcaatatctcgcatcctcaatatccccagCTTGCATCCTGCTCATTGGATTCTCCGATCTAGCCCTGGGTGGTGCCTGTGAGTGCCAAC
| Msp
2301 CGTGTGGTCCTGGTGGAAAGGCCTGCACGGGCTGTCCCGAAAAGCCCCAACTTTGTCAGCAGCTCATTAGCGATATTCGCAATCTCCAGCAGAAGATCCG
I|
2401 GAAATGCGTCTGCGGAGAACCACAATGGATGATTtagacaccaatcacttttaaagatcacaaaaattcttccttaataaaattgttattactgcttcaa
2501 aaaaaaaaaaaaaaaatgtttgagttcttttttatcatttatttcagtatatatcgtccagaaaagaacaaaactagtttttcctgtgggtcacaatcac
| Hind III|
2601 gatggtcttgtggcacctcttgggattcttgcacttccgcttgggaatgcgggtgtggcaccagttgttgccccctccaacaaagctttttgtgagtggaa
| Pst I |
2701 gcggccgttgttgttgttgtggttgatgctgcagtggtggttgttgttgtggtggtggcatcagttgttgttgtgttactagcagaggaagccatctgga
2801 tggccagggcaataagggcaaccacgaaaaggtatttcattttgaaatttgatggaatttatctaagaagtccgcagtgaaataatcgaatttgctagat
2901 gctgtgttctgattttctggagttgcaattaagtcttttatagtggaatttctcttctgtttagttcctcgttttgtgctatcgagtacatttgccaaat
| Pst I |
3001 aataattccacaatgatttccttcctgcagacaaaataagtctcatgaactatattaaatatttgctatcaataaacgccgatccattgggttaccgacg
3101 acactaagacagctgtataaaggtttatgatattcatagcaatgtaccaaatcaaacatgataggaaaaataagccgagatcacaaataaaattgataaa
3201 aaatagcttaagtatttatgttcggattagatttttgttctacttttattatattcatatttgaacttataagagaatcggattatttcgagaaagtcg
3301 tagcaaatgcttgcaatgcattcttagaagttttccgtgagttttgttttataatctgccattttgaattatcactcaattatattgctttacatgttgt
| Cla I | | Xba I |
3401 tattacttttatttatattaaaatataatttttgatatacaatcgatttttaaaagcaattgtcaatataattctagatgatttgtgattataaaccttt


3501 tggttttcgtttataatgcaaatctaattataagtatatagtagtacgtctaatacatttcgtactaaacaattaaaaccttatcgtgaaatactgaagc
3601 aataaagaaccaagaaaatgttattgcctttgagtggtttggaattacatttagctagaggttggtgtatcggctaacaagtaaagaaggctgtatgtaa
3701 attcgttgaatcaatgtcaaattgcctgtcaaagtgcaaacgaagcccaaaatgtctatcctaattcgaacctaaaaatatatatttttttgaatatgcaa
3801 tactataagataattgaatagttttatggggcttatttgtaaagctaaattaagctaaatttaactgtccttatttatattattatatttactcagccta
3901 tattaaagacctattatttatagaatttaacgcagtttgtctgcaaaacatctctacacctttttctacccgttactcgtagagtaaaagggtatactcg
4001 tttcgctgagaagtaacaggcagaatataaagcatatatattcttgattagggtcaatagccgagtcgatctggccatgtccgtctgattctgtttgcca
4101 ctcccacatttttgaaaaatgttttataatttttttcatatttttattatctaaatctatcccttccacaccttagagcattaaatttaatttctttcccc
4201 caatttttaccgatattcgtgaaaaatgttatacattttccatttcacttgaactagctaagtaacgggtatctgttagtctcgttagcgttctctcttg
| PvuI || SalI |
4301 ttttaaaataaagtctaggcgatcgagtcgacccaaaagtatcaaacaaaggggagaaggcttgtgtttgcataatcgaaatactgactccatttttaga
4401 attgcagtttcagtgaaagcgtacctataaaaaggtgaggtatccgcaagaaaagtatcagtttgtggagaattaagtaaaaaacATGAAGCTGACCATT
4501 GCTACCGCCCTAGgtaggtttcaccgaatgctcttgttttcggtatttgagccactgatatattcatccgtttgccttctccacagCGAGCATCCTGCTT
4601 ATTGGCTCCGCTAATGTTGCCAACTGTTGCGATTGTGGATGCCCCACAACTACAACTACTTGTGCGCCACGTACCACGCAACCTCCGTGCACAACTACGA
|HhaI|
4701 CAACAACAACCACAACTACTTGTGCGCCACCCACACAACAATCTACCACGCAACCTCCATGCACGACATCTAAGCCCACCACACCTAAGCAAACTACCAC
|HaeIII|
4801 GCAACTTCCGTGCACAACACCCACCACCACTAAGGCCACCACCACGAAGCCCACCACCACTAAAGCCACCACCACTAAGGCCACCACCACTAAGCCCACC
4901 ACCACTAAGCAAACTACCACGCAACTTCCGTGCACAACACCCACCACCACTAAGCAAACTACCACGCAACTTCCGTGCACAACACCCACCACCACTAAGC
5001 CCACCACCACGAAGCCCACCACCACGAAGCCCACCACCACTAAGCCCACCACCACGAAGCCCACCACCACCAAGCCCACCACCACGAAGCCCACCACCAC
5101 TAAGCCCACCACCACGAAGCCCACCACCACTAAGCCCACCACCACGAAGCCCACCACCACGAAGCCCACCACCACTAAGCCCACCACCACGAAGCCCACC
5201 ACCACTAAGCCCACCACCACGAAGCCCACCACCACTAAGCCCACCACCACGAAGCCCACCACCACTAAGCCCACCACCACGAAGCCCACCACCACGAAGC
5301 CCACCACCACTAAGCCCACCACACCTAAGCCGTGCGGTTGCAAGAGCTGCGGTCCTGGAGGAGAGCCATGCAATGGATGTGCTAAGAGGGATGCACTGTG
5401 CCAGGATCTTAACGGCGTACTCCGCAATCTGGAGCGCAAGATCCGTCAATGCGTCTGCGGTGAACCGCAATGGTTGCTGtgaagcgtcgaagggagcgtct
| SacI |
5501 aatccactcccgtactgatcgatgtgactgcacccctgcgaaatatattctgtgggggagctcggccaggactttgactacgctttgtttttgttatcat
5601 caattgattttacgtgtaagaattaataaaattagttagactgcataaattttaaaagcatttattattattttacttgtattatttatgacaaattatt
5701 atttatctgttgggttttcgaaaatgttggttctaaattaagtttggccatcatttgatcgacttttttcgaatgtatctgttacttttaccaatgcgttg
| SacI |
5801 gctttggctcctagttctatgcgaagtcttaactatccgagctcttatgacttggtcaacttgtctcagctaactactgttggctcgggttcgaacttcg
| PvuI | | XhoI |
5901 gtttgggcccgactcgaatcggcggcttttacgatccgatccgccactcgagtgccgttgcttgtcgccggtggacgcgtttcgttggatatcaattgaaa
6001 tacctacatatgttgaaccggcggcgttttcttcgacccaacataagtgacaaataagtgcaactaaagacgtgaacatatttacaggatataaaacaaa
6101 cgtccgctatgtgcgatttataattggtgggtgatgccatataaacagcgggaatctactataatggctataactttaccgagtatctgtgcgcttgtgt
6201 gtgcgtgtgtgtgtgtgtacctgatgtcgagataaatctgcttgccttcaaattgtcaattgctggcttaatggctggtgcaatctcaagaatcggct
6301 cttttgttttgggctccttggcagtgagcagtgagtgcaatcgaattgaattgaataatatggatatggatgctaaatgactaaaaacgctctctcatgt
6401 agatgtctatcgcggagttgggcaaggcatttagtttccaggtgcactaatcgccggcagacaatcgtcggcggctaaggaacttcaggtcctcttaatt
6501 ggtgggacccccccccattcaggcaaaccctttgacaacaaccacaacgaagcattgtacttgtttgcctggcacgggggatcgatccacttatatgcccc
6601 ttgggtaagcggagttgccacagatttcgccactaatccagttggaatctctcggggaattgacagtgccaccgcatgctcgcatgcaaaatatcgatcc
| SalI |
6701 ctatcgaatcatatgcagcacatgtatgcaaagtgttccagcgaagtcgac

Figure 5. Comparison of the inverted repeat elements. The sequence labelled left is positions 791 to 1290, inclusive. The sequence labelled right is complementary to positions 2937 to 2438, inclusive. Both sequences are written 5' to 3'. Only nucleotides different from the left element are indicated for the right element. Brackets indicate the boundaries of the repeated elements.

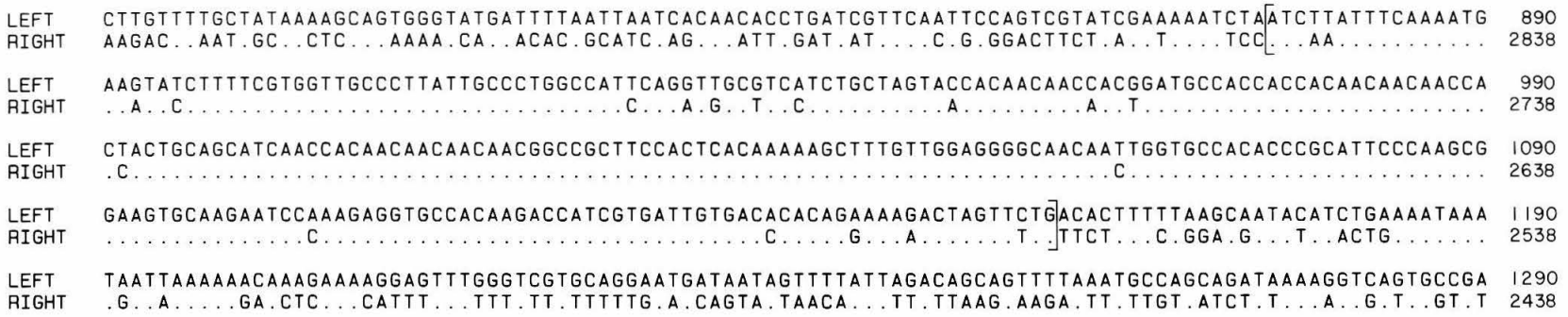

LEFT CTTGTTTTGCTATAAAAGCAGTGGGTATGATTTTAATTAATCACAACACCTGATCGTTCAATTCCAGTCGTATCGAAAAATCTAATCTTATTTCAAAATG 890
RIGHT AAGAC..AAT.GC..CTC...AAAA.CA..ACAC.GCATC.AG...ATT.GAT.AT....C.G.GGACTTCT.A..T....TCC...AA..... 2838
LEFT AAGTATCTTTTCGTGGTTGCCCTTATTGCCCTGGCCATTCAGGTTGCGTCATCTGCTAGTACCACAACAACCACGGATGCCACCACCACAACAACAACCA 990
RIGHT ..A..C.....C...A.G..T..C.....A.....A..T..... 2738
LEFT CTACTGCAGCATCAACCACAACAACAACAACGGCCGCTTCCACTCACAAAAAGCTTTGTTGGAGGGGCAACAATTGGTGCCACACCCGCATTCCCAAGCG 1090
RIGHT .C.....C..... 2638
LEFT GAAGTGCAAGAATCCAAAGAGGTGCCACAAGACCATCGTGATTGTGACACACAGAAAAGACTAGTTCTGACACTTTTTAAGCAATACATCTGAAAATAAA 1190
RIGHT .....C.....C.....G...A.....T..TTCT...C.GGA.G...T..ACTG..... 2538
LEFT TAATTAAAAAACAAAGAAAAGGAGTTTGGGTCGTGCAGGAATGATAATAGTTTTATTAGACAGCAGTTTTAAATGCCAGCAGATAAAAGGTCAGTGCCGA 1290
RIGHT .G..A.....GA.CTC...CATTT...TTT.TT.TTTTTG.A.CAGTA.TAACA...TT.TTAAG.AAGA.TT.TTGT.ATCT.T...A..G.T..GT.T 2438

Figure 6. Nucleotide sequences of cDNA clones homologous to the 68C genes. Each sequence is written 5' to 3' left to right and represents the RNA strand. Oligo(dG) and oligo(dC) joints created during cloning are not shown. For clones II and clones III, the locations of the intervening sequences are indicated by vertical lines. For clones II and IV, 3' polydeoxyadenylate is indicated by a subscript. Single base differences between the cDNA sequences and the corresponding genomic sequences are indicated by asterisks. The cDNA clones were isolated from a *Drosophila* strain heterozygous at the 68C locus (Meyerowitz and Hogness, 1982). The differences between the cDNA and genomic sequences are thus possibly genetic polymorphisms, though they may also represent errors in reverse transcription.

Group II cDNA Clone

TGAAGCTGCTCGTTGTCGCCGTCATTG|CGTGCATCATGCTCATCGGATTCGCCGATCCTGCCTCGGGCTGCAAGGATTGTTCATGCGTGATTTGTGGACC

TGGTGGCGAGCCGTGTCCTGGGTGTTCCGCACGGGTTCCCGTCTGCAAAGATCTGATCAACATTATGGTGGGTCTTGAGCGGCAGGTGCGTCAGTGCGCC

TGCGGGGAGCAGGTTTGGCTGTTCTAGAGATGTGCCCTCAACCTAATCGGCACTGACCTTTTATCTGCTGGCCTTTAAAACTGCTGTCTAATAAAACTAT

TATCATTCCTGCACGACCCA$_{31}$

Group III cDNA Clone

CACCATCATCG|CTTGCATCCTGCTCATTGGATTCTCCGATCTAGCCTTGGGTGGTACCTGTGAGTACCAACCGTGTGGTCCTGGTGGAAAGGCCTGCACG

GGCTGTCCCGAAAAGCCCCAACTTTGTCAGCAGCTCATTAGCGATATTCGCAATCTCCAGCAGAAGATCCGGAAATGCGTCTGCGGAGAACCACAATGGA

TGATTTAGACACCAATCACTTTTAAAGATCACAAAAATTCTTCCTTAATAAAATTGTTATTACTGCTTC

Group IV cDNA Clone

TTTGTTTTTGTCATCATCAATTGATTCTACGTGTAAGAATTAATAAAATTAGTTAGACTGCATA$_{18}$

Figure 7. Mapping of the gene II intervening sequence and 5' end. aDm2026 DNA was *Xba*I-cut, dephosphorylated, and labelled with [γ-^{32}P]ATP and T4 polynucleotide kinase. After *Eco*RI digestion and polyacrylamide gel electrophoresis, the purified 404 base-pair *Xba*I-*Eco*RI fragment (positions 1310 to 1713) was hybridized to third instar larval salivary gland poly(A)$^+$ RNA (lanes 1 to 6) or mock hybridized to yeast tRNA. Portions of the hybridization mixes were diluted into nuclease S_1 assay buffer (lanes 1 to 3) or into exonuclease VII assay buffer (lanes 4 to 6). Mock digestions without added enzyme: lanes 1 and 4. Nuclease S_1 at 350 units/ml: lane 2. Nuclease S_1 at 680 units/ml: lane 3. Exonuclease VII 4.4 units/ml: lane 5. Exonuclease VII 8.8 units/ml: lane 6. Portions of the 404 base-pair *Xba*I-*Eco*RI fragment were subjected to sequence reactions to make size standards (lanes C, C+T, G+A, G). The fragments were separated by electrophoresis through 0.36 mm-thick 5% polyacrylamide/50% urea gels and autoradiographed. This Figure is a composite of two gels that were run for different times in order to resolve both sets of protected fragments. The numbers at the left show the nucleotide positions of the bands indicated.

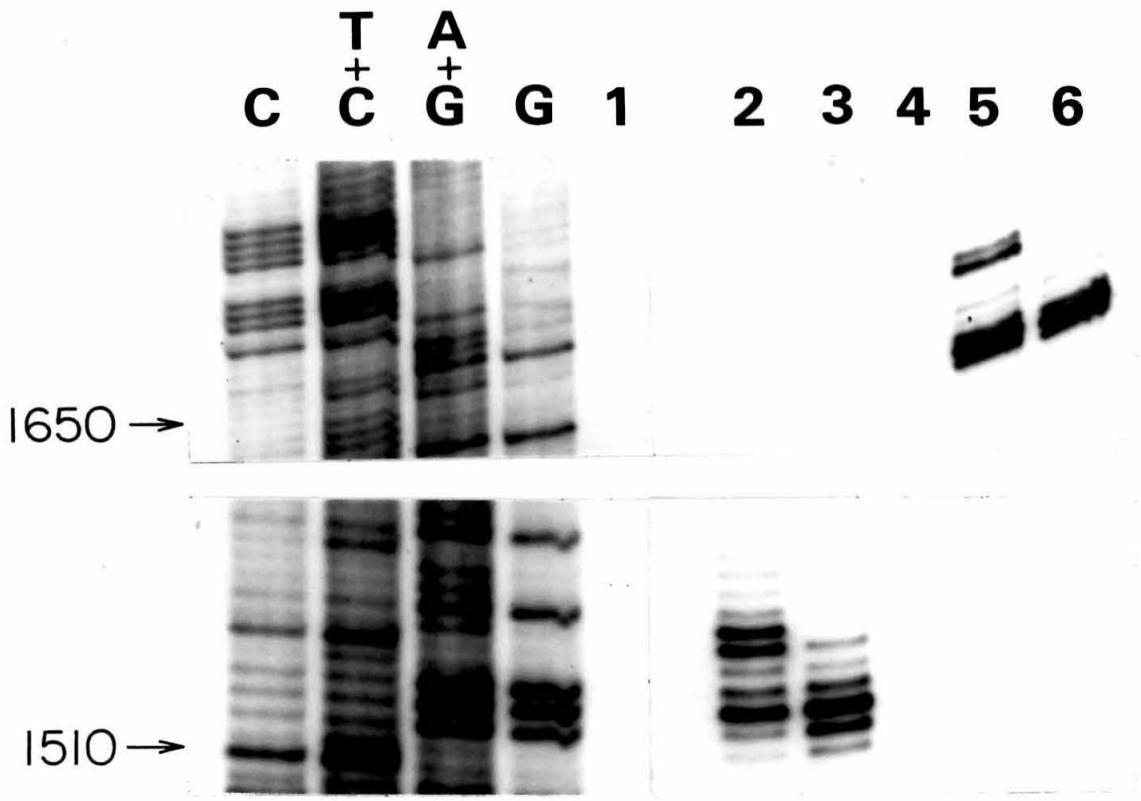

T + C
A + G
C
G
1
2
3
4
5
6
1650
1510

Figure 8. Comparison of the DNA sequences flanking the intervening sequences and 5' ends of the 68C genes. The 5' termini have been aligned by the common oligonucleotide observed around other *Drosophila* initiation points (Snyder *et al.*, 1982). Exonuclease VII protection points are marked by downward pointing arrows, cDNA extension endpoints are marked by upward pointing arrows. The T-A-T-A box sequences are marked by overlines and Gs. Highly conserved nucleotides flanking the splicing sites are boxed. Additional vertical lines mark the exon/intervening sequence (IVS) and intervening sequence/exon boundaries. Translation initiation codons are marked Met. Sequence homologies upstream of genes II and III are underlined.

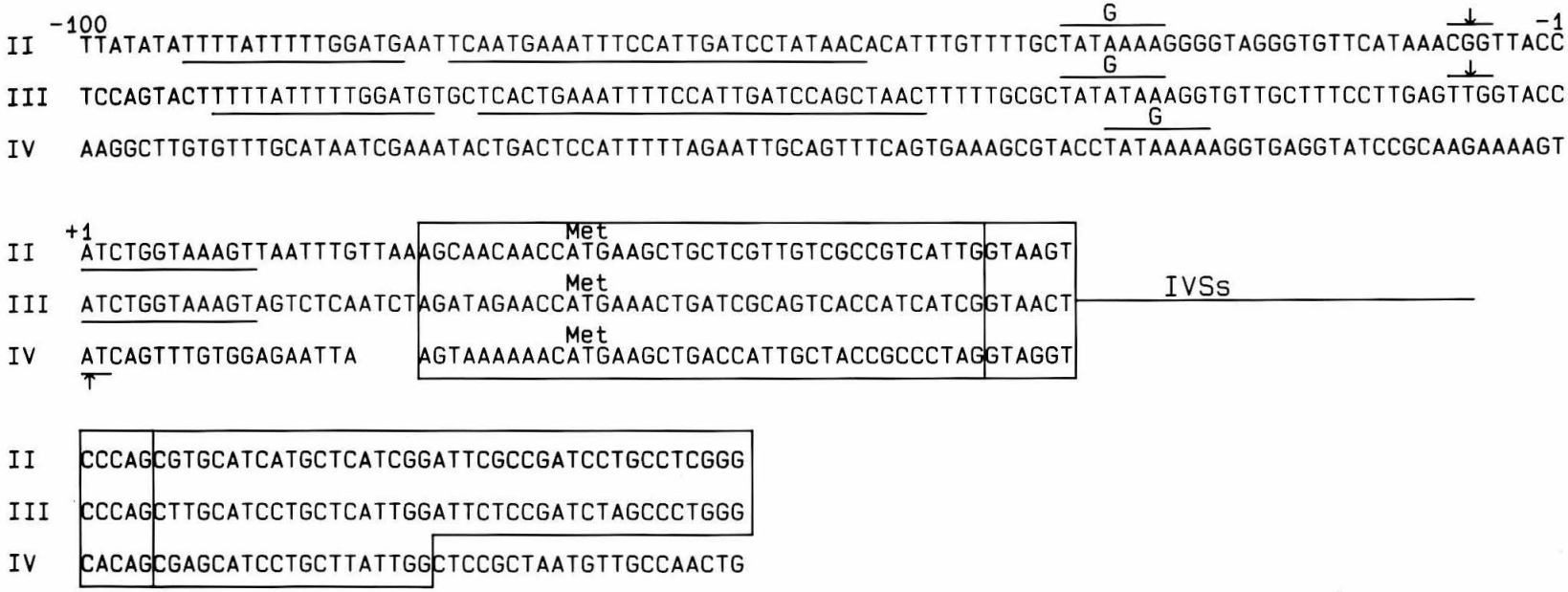
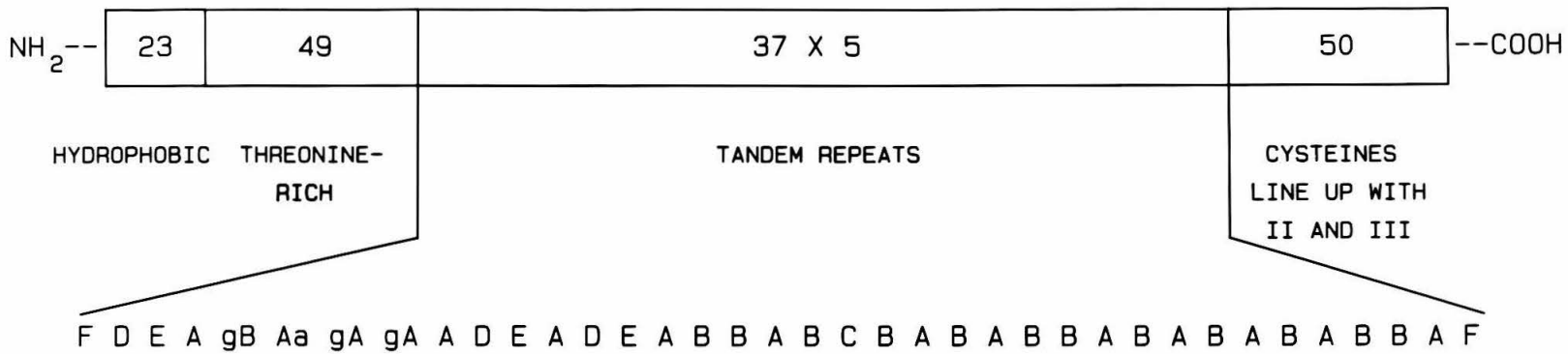

-100
-1
+1
Met
IVSs
II
III
IV

Figure 9. Complete amino acid sequences of the predicted protein products of the 68C genes. Each amino acid sequence is written in the standard three-letter code above the corresponding mRNA-congruent DNA strand. The sequences of genes II and III, and of their protein products sgs-8 and sgs-7, are interrupted so that the amino-terminal leaders of all three gene products may be aligned separately from the carboxy-terminal cysteine-rich modules. Nucleotides in *Sgs-7* and *Sgs-3* that are identical to those in *Sgs-8* are indicated by horizontal lines. The intervening sequence locations are indicated (IVS).


IVSs
sgs-8
MetLysLeuLeuValValAlaValIleAlaCysIleMetLeuIleGlyPheAlaAspProAlaSerGly
II
ATGAAGCTGCTCGTTGTCGCCGTCATTGCGTGCATCATGCTCATCGGATTCGCCGATCCTGCCTCGGGC
sgs-7
MetLysLeuIleAlaValThrIleIleAlaCysIleLeuLeuIleGlyPheSerAspLeuAlaLeuGly
III
sgs-3
MetLysLeuThrIleAlaThrAlaLeuAlaSerIleLeuLeuIleGlySerAlaAsnValAlaAsnCys
IV
sgs-3
CysAspCysGlyCysProThrThrThrThrThrCysAlaProArgThrThrGlnProProCysThrThrThrThrThrThrThrThrThrThrCysAlaProPro
IV
TGCGATTGTGGATGCCCCACAACTACAACTACTTGTGCGCCACGTACCACGCAACCTCCGTGCACAACTACGACAACAACAACCACAACTACTTGTGCGCCACCC
sgs-3
ThrGlnGlnSerThrThrGlnProProCysThrThrSerLysProThrThrProLysGlnThrThrThrGlnLeuProCysThrThrProThrThrThrLysAla
IV
ACACAACAATCTACCACGCAACCTCCATGCACGACATCTAAGCCCACCACACCTAAGCAAACTACCACGCAACTTCCGTGCACAACACCCACCACCACTAAGGCC
sgs-3
ThrThrThrLysProThrThrThrLysAlaThrThrThrLysAlaThrThrThrLysProThrThrThrLysGlnThrThrThrGlnLeuProCysThrThrPro
IV
ACCACCACGAAGCCCACCACCACTAAAGCCACCACCACTAAGGCCACCACCACTAAGCCCACCACCACTAAGCAAACTACCACGCAACTTCCGTGCACAACACCC
sgs-3
ThrThrThrLysGlnThrThrThrGlnLeuProCysThrThrProThrThrThrLysProThrThrThrLysProThrThrThrLysProThrThrThrLysPro
IV
ACCACCACTAAGCAAACTACCACGCAACTTCCGTGCACAACACCCACCACCACTAAGCCCACCACCACGAAGCCCACCACCACGAAGCCCACCACCACTAAGCCC
sgs-3
ThrThrThrLysProThrThrThrLysProThrThrThrLysProThrThrThrLysProThrThrThrLysProThrThrThrLysProThrThrThrLysPro
IV
ACCACCACGAAGCCCACCACCACCAAGCCCACCACCACGAAGCCCACCACCACTAAGCCCACCACCACGAAGCCCACCACCACTAAGCCCACCACCACGAAGCCC
sgs-3
ThrThrThrLysProThrThrThrLysProThrThrThrLysProThrThrThrLysProThrThrThrLysProThrThrThrLysProThrThrThrLysPro
IV
ACCACCACGAAGCCCACCACCACTAAGCCCACCACCACGAAGCCCACCACCACTAAGCCCACCACCACGAAGCCCACCACCACTAAGCCCACCACCACGAAGCCC
sgs-3
ThrThrThrLysProThrThrThrLysProThrThrThrLysProThrThrThrLysProThrThrProLys
IV
ACCACCACTAAGCCCACCACCACGAAGCCCACCACCACGAAGCCCACCACCACTAAGCCCACCACACCTAAG
sgs-8
CysLysAspCysSerCysValIleCysGlyProGlyGlyGluProCysProGlyCysSerAlaArgValProValCysLysAspLeuIleAsnIleMet
II
TGCAAGGATTGTTCATGCGTGATTTGTGGACCTGGTGGCGAGCCGTGTCCTGGGTGTTCCGCACGGGTTCCCGTCTGCAAAGATCTGATCAACATTATG
sgs-7
GlyAlaCysGluCysGlnProCysGlyProGlyGlyLysAlaCysThrGlyCysProGluLysProGlnLeuCysGlnGlnLeuIleSerAspIle
III
sgs-3
ProCysGlyCysLysSerCysGlyProGlyGlyGluProCysAsnGlyCysAlaLysArgAspAlaLeuCysGlnAspLeuAsnGlyValLeu
IV
sgs-8
GluGlyLeuGluArgGlnValArgGlnCysAlaCysGlyGluGlnValTrpLeuPhe
II
GAGGGTCTTGAGCGGCAGGTGCGTCAGTGCGCCTGCGGAGAGCAGGTTTGGCTGTTCTAG
sgs-7
ArgAsnLeuGlnGlnLysIleArgLysCysValCysGlyGluProGlnTrpMetIle
III
sgs-3
ArgAsnLeuGluArgLysIleArgGlnCysValCysGlyGluProGlnTrpLeuLeu
IV

Figure 10. *Sgs-3* contains a set of imperfect tandem repeats. The overall organization of the predicted *Sgs-3* protein product is indicated by the box diagram. The numbers in each box are the numbers of amino acids found in each segment of the protein. The central tandem repetitious region has been classified into several types of 15 nucleotide repeat units. Their order is shown under the box diagram. DNA sequences and corresponding amino acid sequences of the repeat unit types are shown.

TANDEM REPEATS IN THE SGS-3 GENE AND PROTEIN



REPEAT TYPE	DNA SEQUENCE	AMINO ACID SEQUENCE
A	CCC ACC ACC ACT AAG	Pro Thr Thr Thr Lys
gA	GCC ACC ACC ACT AAG	Ala Thr Thr Thr Lys
Aa	CCC ACC ACC ACT AAA	Pro Thr Thr Thr Lys
B	CCC ACC ACC ACG AAG	Pro Thr Thr Thr Lys
gB	GCC ACC ACC ACG AAG	Ala Thr Thr Thr Lys
C	CCC ACC ACC ACC AAG	Pro Thr Thr Thr Lys
D	CAA ACT ACC ACG CAA	Gln Thr Thr Thr Gln
E	CTT CCG TGC ACA ACA	Leu Pro Cys Thr Thr
F	CCC ACC ACA CCT AAG	Pro Thr Thr Pro Lys

Chapter 3:

Cis-Acting Sequences Required for Expression of the Divergently Transcribed *Drosophila melanogaster* *Sgs-7* and *Sgs-8* Glue Protein Genes

Cis-Acting Sequences Required for Expression of the Divergently Transcribed *Drosophila melanogaster* *Sgs-7* and *Sgs-8* Glue Protein Genes*

Mark D. Garfinkel and Elliot M. Meyerowitz

Division of Biology, California Institute of Technology, Pasadena, Calif. 91125, U.S.A.

*The accompanying typescript follows the format for publication by the *Journal of Molecular Biology*.

The 68C locus of the *Drosophila melanogaster* polytene chromosomes contains three glue protein structural genes (*Sgs-3*, *Sgs-7*, and *Sgs-8*) transcribed in the third larval instar salivary glands. Activation and repression of 68C glue gene expression requires the steroid hormone ecdysterone, while activation also requires the product of the unlinked gene *l(1)npr-1*$^+$. The *Sgs-7* and *Sgs-8* genes are divergently transcribed with 475 base-pairs separating the two 5' ends. A transcriptional fusion gene was constructed by joining the 5' untranslated region of *Sgs-7* to the 5' untranslated region of the *D. melanogaster Adh* gene. A translational fusion gene was constructed by joining the *Sgs-8* gene to the *Escherichia coli lacZ* gene. When the fusion genes are placed in their normal divergently transcribed arrangement and reintroduced into *D. melanogaster* using P element gene transfer, third instar larval salivary gland expression of both alcohol dehydrogenase activity and β-galactosidase activity was observed. Normal tissue, stage, and quantity of *Sgs-7—Adh* fusion gene expression is observed when 211 base-pairs of 5' flanking sequence are present. An *Sgs-7—Adh* fusion gene with 92 bp upstream is not functional. Normal tissue and stage of *Sgs-8—lacZ* fusion gene expression is observed when 432 bp of 5' flanking sequence are present, when 415 bp of 5' flanking sequence are present expression is reduced at least twentyfold. Expression of the fusion genes in the divergent arrangement requires the *l(1)npr-1*$^+$ gene product, supporting the proposal that this *trans*-acting factor affects glue protein gene transcription. The results are consistent with the hypothesis that a single region functioning bidirectionally, located closer to the *Sgs-7* gene, is required for expression of both genes.

1. Introduction

The salivary gland of *Drosophila melanogaster* is a tissue specialized for the secretion of proteins. Ten hours after fertilization, midway through embryogenesis, the salivary gland cells have secretory granules visible in their cytoplasm (Campos-Ortega and Hartenstein, 1985). During the third larval instar, about five days later, secretory granules again appear in the cytoplasm of the salivary gland cells. At the end of the third instar period, these proteins are exported to the lumen of the gland and then expelled from the animal (Korge, 1975, 1977; Zhimulev and Kolesnikov, 1975). During the prepupal period that immediately follows the third larval instar, the salivary glands are again active in the synthesis of secretory proteins (Korge, 1977; Sarmiento and Mitchell, 1982). Of the three documented times of secretory protein synthesis, only the third larval instar secretion has a known function. This mixture of secreted proteins and glycoproteins, after expulsion from the larva, hardens to form a sticky glue mass that affixes the animal to its substrate during metamorphosis (Fraenkel and Brooks, 1953; Lane *et al.*, 1972).

The *Salivary gland secretion* (*Sgs*), or glue protein, genes expressed in the salivary glands of *D. melanogaster* third instar larvae represent an intriguing example of the developmental control of gene set activity. The set codes for proteins that are synthesized in only one tissue and at only one time in development (Korge, 1975, 1977; Beckendorf and Kafatos, 1976). The activation and the repression of gene set expression are under the control of the steroid hormone ecdysterone. In the salivary gland cell nuclei are polytene chromosomes (reviewed by Korge, 1987) that have three types of cytologically visible structures. Bands and interbands are essentially constant features shared by the polytene chromosomes of every polytene tissue. Chromosome

puffs are localized regions of chromatin decondensation that vary with time in a single tissue and that differ between tissues. Expression of the glue protein gene set is correlated with the presence of chromosome puffs at the corresponding genetic loci, allowing the relationship between chromosomal morphology and gene expression to be investigated.

Seven genes are known to code for components of the glue (*Sgs-1*, Velissariou and Ashburner, 1980; *Sgs-3*, Korge, 1975, 1977, Akam *et al.*, 1978; *Sgs-4*, Korge, 1975, 1977; *Sgs-5*, Guild and Shore, 1984; *Sgs-6*, Velissariou and Ashburner, 1981; *Sgs-7* and *Sgs-8*, Crowley *et al.*, 1983). Five of the genes have been cloned (*Sgs-3, Sgs-7, Sgs-8*, Meyerowitz and Hogness, 1982, Crowley *et al.*, 1983; *Sgs-4*, Muskavitch and Hogness, 1980; and *Sgs-5*, Guild and Shore, 1984). The *Sgs-3*, *Sgs-7*, and *Sgs-8* genes are clustered within a five-thousand base-pair segment of the left arm of chromosome three at the site 68C, while the *Sgs-1*, *Sgs-4*, *Sgs-5*, and *Sgs-6* genes reside at dispersed locations in the genome. All of these chromosomal sites correspond to large intermolt chromosome puffs during the time of glue protein synthesis.

Late in the third instar, in response to rising hemolymph concentrations of the insect steroid hormone ecdysterone, the intermolt chromosome puffs regress and are replaced by a sequence of ecdysterone-induced puffs (Ashburner, 1973). When salivary glands are removed from third instar larvae prior to exposure to ecdysterone, and cultured *in vitro* in the presence of the steroid and protein synthesis inhibitors, regression of the intermolt puffs at 3C (*Sgs-4*), 25B (*Sgs-1*), and 68C (*Sgs-3*, *-7*, *-8*), and induction of the so-called early puffs will occur. Therefore, these cytological changes are primary responses to the hormone (Ashburner, 1972, 1974). The

regression of the 68C intermolt puff in salivary glands cultured in the presence of ecdysterone is associated with a rapid significant reduction in the incorporation of ^{3}H-ribonucleosides into 68C-homologous RNAs (Crowley and Meyerowitz, 1984). Both the intermolt puffs, at the time they are regressing, and the early puffs, as they are being induced, can be labelled by ecdysterone following photochemical activation (Gronemeyer and Pongs, 1980; Dworniczak *et al.*, 1983). Ecdysterone thus plays an important role in the cessation of glue protein gene expression at the end of larval life, apparently by direct action on the genes, presumably through the agency of a receptor protein specific for the hormone.

The existing evidence suggests that different glue protein genes are activated in different ways. Larvae hemizygous for the temperature-sensitive ecdysterone-deficient mutation *lethal(1)suppressor of forked*^{ts67g} fail to express *Sgs-3*, *Sgs-4*, *Sgs-7* and *Sgs-8* when shifted to non-permissive temperature early in the third instar. Transcription of the genes can be induced at restrictive temperature by feeding the mutant larvae ecdysterone (Hansson and Lambertsson, 1983). In contrast, the function eliminated by the *lethal(1)non-pupariating-1* mutation is required for expression of the three 68C genes *Sgs-3*, *Sgs-7* and *Sgs-8*, but not for expression of the *Sgs-4* and *Sgs-5* genes (Crowley *et al.*, 1984).

The 68C intermolt puff is particularly interesting because of the clustering of the three genes found there, and because of the dependence of their expression on the *l(1)npr-1*$^{+}$ gene product. The order of the genes is telomere, *Sgs-8*, *Sgs-7*, *Sgs-3*, centromere with respect to the left arm of the third chromosome (Crosby and Meyerowitz, 1986). Transcription of *Sgs-8* proceeds leftward, transcription of *Sgs-7* proceeds rightward, and transcription of *Sgs-3* proceeds rightward (Meyerowitz and

Hogness, 1982; Garfinkel *et al.*, 1983). The 5' end of *Sgs-8* is separated from the 5' end of *Sgs-7* by 475 base-pairs of non-transcribed intergenic spacer (Garfinkel *et al.*, 1983), and the two genes are surrounded by two 285-base-pair elements that form an inverted repeat (Meyerowitz and Hogness, 1982; Garfinkel *et al.*, 1983). The 3' end of *Sgs-7* is separated from the 5' end of *Sgs-3* by 1958 base-pairs.

The basis for the coordinate regulation of the 68C glue protein genes was sought in a structural study of the cluster (Garfinkel *et al.*, 1983). The sequence of 6751 base-pairs of genomic DNA was determined, as were sequences of complementary DNA clones representing each gene. Analysis of the sequences revealed unexpectedly that *Sgs-3*, *Sgs-7*, and *Sgs-8* are related to each other by gene duplication and divergence events. One consequence of their evolutionary history might be that each gene was duplicated along with its own set of *cis*-acting regulatory sequences, and that such elements might retain sufficient sequence identity to be identified by inspection. The only readily recognizable sequence element upstream of all three 68C glue protein genes is the T-A-T-A box element located at approximately –30 base-pairs (Goldberg, 1979), an element that is found upstream of nearly all eukaryotic RNA polymerase II transcription units. Between the –44 and –91 base-pair positions relative to *Sgs-7* are sequences homologous to those located between –48 base-pairs and –93 base-pairs upstream of *Sgs-8*. These conserved sequences will be referred to as the "right copy" and the "left copy," respectively, and contain at their cores sequences that are homologous to the ecdysterone regulatory element identified by Mestril *et al.* (1986) in their study of the *Hsp23* gene.

Various segments of the 68C gene cluster have been tested for function using *Drosophila* gene transfer methods. Vijay Raghavan *et al.* (1986), using P element-

mediated germline transformation (Rubin and Spradling, 1982), showed that the tissue and time of *Sgs-3* expression were regulated by DNA sequences within 130 base-pairs upstream of the 5' end of the gene, but that gene product abundance was substantially reduced compared to wild-type. Bourouis and Richards (1985) and Crosby and Meyerowitz (1986) showed that the *cis*-acting sequences required for full-level of *Sgs-3* expression requires more than 2 kb of additional 5' flanking sequences. Expression of the *Sgs-3* gene transferred *via* the P factor method retains the requirement for the *l(1)npr*$^+$ gene product (Crowley *et al.*, 1984). In these experiments, the *Sgs-3* gene functions normally in the absence of adjacent functional *Sgs-7* and *Sgs-8* genes; the clustering observed at 68C therefore appears to be a vestige of the evolutionary origin of the cluster, without consequence for gene expression.

The experiments described in this paper are the start of a similar analysis of the the *Sgs-7*, *Sgs-8* gene pair. The general question is to determine the arrangement of *cis*-acting sequences required for each gene to function. The specific question is whether the sequence elements conserved in the 5' flanking sequences adjacent to *Sgs-7* and *Sgs-8* are *cis*-acting regulatory sequences that are functionally conserved. In order to distinguish the gene products of newly introduced glue protein genes from the background of expression due to the chromosomal genes, we made two gene fusions. The protein-coding region of the *Sgs-8* gene was joined in the correct reading frame to the *E. coli lacZ* gene, resulting in a translational fusion gene that codes for a hybrid secretory protein with β-galactosidase enzyme activity. The 5' untranslated region of the *Sgs-7* gene was fused to the 5' untranslated region of the *D. melanogaster Adh* gene, creating a transcriptional fusion gene that directs the synthesis of a hybrid messenger RNA coding for alcohol dehydrogenase enzyme activity. These reporter gene choices were motivated by several considerations: first, to exploit sensitive

histochemical reactions for the *in situ* detection of the protein products coded by introduced glue gene constructions; second, to exploit sensitive quantitative enzyme activity assays for measurement of gene function; third, to enable us to test plasmid constructions simultaneously for *Sgs-7* gene function and for *Sgs-8* gene function; and fourth, to lay the foundation for efforts to use chemical selections on salivary gland-restricted alcohol dehydrogenase as a means of recovering mutations in *trans*-acting regulatory loci that affect glue protein gene expression.

The "somatic transformation" procedure (Martin *et al.*, 1986) provides a rapid means of qualitatively testing plasmid constructions. We used it as a prelude to P element-mediated germline transformation (Rubin and Spradling, 1982), which permits extended quantitative experimentation. The locations of *cis*-acting regulatory sequences revealed by the experiments suggest that the divergent transcription arrangement of the *Sgs-7*, *Sgs-8* gene pair is not simply an evolutionary vestige, but is functionally important for the expression of these genes.

2. Materials and Methods

(a) Drosophila *strains and methods*

(i) *Strains*

J.W Posakony (University of California, San Diego) provided flies of the genotype *Adh*^{fn6} *cn*; *ry*^{502}. Stocks were cultivated in half-pint bottles, or in shell vials, on standard medium at 22°C. The balancer chromosome, *In(3LR)TM3, Sb ry*^{RK} *Ser*, was kindly provided by B.T. Wakimoto (University of Washington, Seattle) and was first described by Karess and Rubin (1984). *T(2,3)Ata/CyO; TM3, Sb ry*^{RK} *Ser* was constructed by, and obtained from, M.A. Crosby. Markers and balancers are described by Lindsley and Grell (1968) and by Lindsley and Zimm (1985, 1986, 1987).

(ii) *Embryo manipulation*

For egg collections, adults that emerged over a one- or two-day period were collected and raised on standard medium supplemented with yeast paste. They were allowed to feed on the yeast for an additional two days before being transferred to containers suitable for egg collection. Eggs were recovered as per Crosby and Meyerowitz (1986), and were dechorionated, desiccated and microinjected essentially as described by Spradling and Rubin (1982).

(iii) *Germline transformation*

Carnegie 20 (Rubin and Spradling, 1983), which contains the *rosy*$^{+}$ (*ry*$^{+}$) eye color marker, was used as the P element vector for transposition of glue protein gene constructions. For microinjection, Carnegie 20-derived constructions were present at 500 µg/ml in 5 mM KCl, 0.1 mM sodium phosphate (pH 6.8) injection buffer (Spradling and Rubin, 1982). The injection buffer contained 100 µg/ml of the

transposase-producing, transposition-defective, P factor plasmid phsπ (H. Steller and V. Pirrotta, unpublished; provided by B. Butler and V. Pirrotta, Baylor College of Medicine, Houston, TX) that was used as the helper. Larvae that survived the injection procedure were cultivated singly in fly vials at 22°C; newly emerged adults were individually mated to virgin flies from the *Adh*^{fn6} *cn*; *ry*^{502} stock.

Progeny of the injected flies were surveyed for the presence of individuals carrying *ry*$^{+}$ transformation events. Sublines were established by mating each of these animals to virgin flies of the appropriate sex from the *Adh*^{fn6} *cn*; *ry*^{502} stock. The frequency with which *ry*$^{+}$ animals appeared in their progeny was noted, and virgin *ry*$^{+}$ siblings from each subline were mated *inter se*. The sex-linked insertion event *Tf(1)GLAX1.0-1*[1] was made homozygous by repeated brother-sister matings. Standard genetic techniques using the balancer strain *T(2,3)Ata /CyO; TM3, Sb ry*^{RK} *Ser* were applied to map, balance, and make homozygous the autosomal insertion events. When homozygous-lethal, a transformation event was made into a stock by selection of virgin flies possessing only the dominant visible marker necessary.

---

[1]Crosby and Meyerowitz (1986) first coined this nomenclature. *Tf()* refers to *T*rans*f*ormation event on chromosome () in a manner analogous to the naming of other kinds of *Drosophila* chromosome rearrangement. GLAX1.0 indicates the transformed segment: in this case, a composite P element that has *G*lue protein promoters driving the expression of the bacterial *LacZ* gene and the *Drosophila* *A*lcohol dehydrogenase gene, along with the *Drosophila* *X*anthine dehydrogenase gene (*rosy*$^{+}$) as selectable marker, and a total of 1.0 kilobase-pairs of 68C DNA. Different insertion events are distinguished by the hyphen-numeral that follows. The name of the transformed DNA segment is the same as that of the parental plasmid DNA which donated the segment. Such plasmids are designated with the generic "p" for plasmid. Therefore, pGLAX1.0 is the plasmid source of the composite P element that was incorporated into the X chromosome in the *Drosophila* strain *Tf(1)GLAX1.0-1*.

(iv) *Enzyme assays*

The glutaraldehyde-fixation method of Ursprung *et al.* (1970) was used to determine the tissue distribution of alcohol dehydrogenase activity. Alcohol dehydrogenase activity in soluble extracts of *Drosophila* salivary glands was measured using the method of Sofer and Ursprung (1968). Third instar larvae were dissected in chilled Ringer's solution, and the salivary glands were placed in 10 mM sodium phosphate (pH 6.8) inside a 500-µl centrifuge tube stored on ice (two salivary gland lobes from one animal per 10 µl sodium phosphate). Each centrifuge tube also contained 50-100 µl of packed 0.5-mm-diameter zirconium oxide beads (BioSpec Products, Bartlesville, OK). When all the salivary glands had been collected, the tissues were homogenized by one minute of vortex mixing. A 26-gauge needle hole was punched in the bottom of each 500 µl tube, and the extract collected in a 1.5-ml tube by a 1-min. spin in a 4°C microcentrifuge. If necessary, a second collection tube and spin were used to recover all the homogenate. Tissue debris was removed from each homogenate by a 5-min. centrifugation, followed by careful transfer of the homogenate into fresh tubes. The spectrophotometer cuvette contained 800 µl of assay buffer [50 mM sodium carbonate (pH 9.6), 1.5 mM NAD^+, 3% (v/v) 2-butanol], and the Hewlett-Packard Model 8451A spectrophotometer was blanked. Fifty microliters (five animal-equivalents of salivary gland tissue extract) were pipetted into the assay buffer, the reaction mixed by gently filling and emptying a Pasteur pipette, and automated one-second measurements taken every 30 seconds for fifteen minutes. The measurements were printed out and entered into the Cricket Graph software for the Apple Macintosh computer. Least-squares linear regression lines were calculated, using those initial values that were clearly linear with time, and the slope values converted

into Units of ADH[2] enzyme activity per animal-equivalent [One Unit is that amount of ADH enzyme that causes a change in absorbance at 340 nm of 0.001 per min. (Sofer and Ursprung, 1968)]. The first eleven measurements typically fell on a straight line that lacked any sign of saturation or other anomaly.

Controls for the soluble extract assay of ADH included the omission of any *Drosophila* tissue extract, which revealed that the rate of spontaneous reduction of NAD^+ was small; the addition of Adh^{fn6} *cn*; ry^{502} salivary gland extract, which showed a rate of reduction of NAD^+ equal to 0.07 ± 0.12 Units per animal-equivalent of extract ($n = 13$); and the omission of 2-butanol from *Tf()GLAX1.0* salivary gland extract reactions, which showed that the major NAD^+ reduction reaction was dependent upon the addition of a known ADH substrate. Our measurements therefore reflect the enzymatic activity of authentic alcohol dehydrogenase.

Histochemical determination of β-galactosidase activity made use of the chromogenic substrate X-Gal according to the method of Singh and Knox (1984), as modified by Vijay Raghavan *et al.* (1986). Soluble extract measurements made use of a hybrid of three procedures. First, we found that the same soluble extracts prepared for ADH activity measurements could be assayed for β-galactosidase activity. Second, we used the β-galactosidase assay buffer described by Miller (1972). Third, we used the fluorogenic substrate 4MUGal, and reaction and detection conditions similar to those

---

[2]Abbreviations used: ADH, alcohol dehydrogenase enzyme; β-gal, β-galactosidase enzyme; X-Gal, 4-bromo-5-chloro-3-indolyl-β-D-galactopyranoside; bp, base-pairs of DNA; 4MU, 4-methylumbelliferone; 4MUGal, 4-methylumbelliferyl-β-D-galactopyranoside; EDTA, ethylene diamine tetraacetic acid; kb, 10^3 base-pairs of double-stranded nucleic acid or 10^3 bases of single-stranded nucleic acid; EGTA, ethyleneglycol-bis (β-aminoethyl ether) *N, N'*-tetraacetic acid; SDS, sodium dodecyl sulfate.

used by Jefferson *et al.* (1987). Fifty microliters of extract were added to 550 µl "Z buffer" (Miller, 1972) adjusted to pH 7.5 and supplemented with 0.5 mM 4MUGal. A 100 µl portion of each reaction was removed at time zero and placed into 900 µl 0.2 M Na_2CO_3. The reactions were incubated at 37°C; 100 µl portions were removed every hour for four hours and stopped by adding them to 900 µl 0.2 M Na_2CO_3. The fluorescence of the reaction product, 4MU, was detected using an SLM8000C spectrofluorimeter. Slit widths were set at 4 nm, excitation wavelength was 365 nm, and emission wavelength was 455 nm. A standard curve was constructed using the above solutions containing four known concentrations of 4MU (0 µM, 0.01 µM, 0.1 µM, and 1 µM), and was used to convert the raw fluorescence counts-per-second into picomoles of 4MU. Since each timepoint represents one-sixth of the original reaction, each picomole value was multiplied by six. Cricket Graph was again used to calculate linear regression lines from the data. The slope values were converted into Units of sgs-8-β-galactosidase activity per animal-equivalent; we define one Unit of hybrid sgs-8-β-galactosidase enzyme as that amount of enzyme which releases one picomole 4MU per hour.

Controls for the soluble extract assay for β-galactosidase included these mock reactions: one to which no *Drosophila* tissue extract was added and 10 mM sodium phosphate was substituted, in which no increase in fluorescence was recorded; and one to which a salivary gland extract from the non-transformed Adh^{fn6} *cn*; ry^{502} strain was added, in which a rate of increase corresponding to 12.2 ± 7.1 pmol 4MU liberated per hour per animal-equivalent of extract ($n = 6$). One possible explanation for the increase in fluorescence in extracts that have no β-galactosidase demonstrated in histochemical reactions is that the 4MUGal substrate undergoes a conversion to a different form which is hydrolyzable by another enzyme (c.f., Jefferson *et al.*, 1987).

(v) *"Transient expression" assay*

Supercoiled plasmid DNA containing *Sgs* genes fused to histochemically assayable marker genes, lacking P factor sequences, were microinjected into the anterior end of syncytial cleavage stage embryos, rather than the posterior poleplasm, using the method of Spradling and Rubin (1982). The DNA concentrations were in the range of 400-600 μg/ml in the injection buffer described above. Survivors of the injection were retrieved, and were placed in apple juice agar plates (Nüsslein-Volhard *et al.*, 1984) supplemented with drops of live yeast paste. Third instar larvae were dissected and their salivary glands removed for histochemical staining.

(b) *General nucleic acids methods*

Chapter 2 contains the descriptions of, or gives references for, the following methods: preparation of genomic DNA from whole adult flies and of RNA from third instar larval salivary glands, cesium chloride gradient purification of plasmid DNA from *Escherichia coli*, nick translation labelling of DNA, whole genome Southern (1975) blot filters, and RNA gel blot filters.

(c) *Plasmid constructions*

(i) *General plasmid construction methods*

DNA fragments for clone constructions were isolated following agarose gel electrophoresis using Whatman DE81 paper.

The *E. coli* strains HB101 (Boyer and Roullard-Dussoix, 1969) and Bozo2.7 (R.E. Pruitt, unpublished experiments) were most frequently used as hosts for plasmid constructions. Bacterial cells were rendered competent to take up DNA using $CaCl_2$.

Plasmid DNAs were isolated using a slight modification of R. E. Pruitt's unpublished method. 0.5 ml of 0.2 M Tris-HCl (pH 8.0), 0.1 M EDTA, 0.1% (w/v) sodium N-lauroyl sarcosinate, 100 μg/ml proteinase K were added to the cell pellets. Cells were resuspended by vortex mixing and were incubated at 50°C for 15 minutes. Lysed-cell debris was sedimented by 15-minute centrifugation in a microcentrifuge. The top 400-450 μl of the supernatant were transferred to fresh tubes containing 5-μl portions of 100 mM phenylmethylsulfonyl fluoride in ethanol, and 800-900 μl ethanol were added to precipitate nucleic acids. After 15 minutes at room temperature the nucleic acids were recovered by a 15-minute centrifugation. The pellets were washed with 75% ethanol, dried, and resuspended in 50 μl of 10 mM Tris-HCl (pH 8.0), 1 mM EDTA, 10 ng/ml pancreatic RNase A. For restriction enzyme digest analysis of the miniprep DNA, 5-10 μl portions of DNA were digested in 15-20 μl reactions with around 10 Units of the appropriate enzyme for several hours at 37°C.

(ii) *Construction of the* Sgs-7—Adh *fusion gene*

The plasmid pHAP (Bonner *et al.*, 1984) was used as the source of the promoterless *D. melanogaster Adh* gene for fusion to *Sgs-7* promoter sequences. This plasmid contains a *D. melanogaster Adh* gene modified by the addition of a *Hin*dIII linker in the 5' untranslated region. The 1.8 kb *Xba*I-*Hin*dIII fragment encompassing the *Adh* gene was purified and ligated to *Hin*dIII and *Xba*I-digested pUC18 DNA. One transformant clone, designated nDm9035[3], possessing a single copy of the *Adh*

[3]Plasmid vectors are identified by a single letter: a is for pBR322 (Bolivar *et al.*, 1977), k is for pSP65 (Melton *et al.*, 1984), n is for pUC18 (Yanisch-Perron *et al.*, 1985), s is for DOA-3.8 (C. Chang, unpublished experiment). Clones designated Dm typically contain segments of *D. melanogaster* genomic DNA. When the vector backbone has a complicated and/or specialized history, the generic "p" for plasmid designation is used. pGAO refers to a plasmid containing a "glue promoter driving alcohol dehydrogenase only"; pGAZ refers to a plasmid containing "glue promoters driving alcohol dehydrogenase and lacZ," *etc.*

fragment, excisable only by double digestion with *Xba*I and *Hin*dIII, was picked for further work.

nDm9035 DNA was linearized with *Hin*dIII, and the restriction site termini were partially filled with the Klenow fragment of *E. coli* DNA polymerase I, dATP, and dGTP. The DNA was purified by adding EDTA, ammonium acetate and isopropanol. Likewise, aDm2026 DNA was digested with *Xba*I, and the termini partially filled by adding Klenow fragment, dCTP, and dTTP. This reaction was terminated by adding EDTA and SDS and heating the mixture for 10 minutes at 70°C. After agarose gel electrophoresis, the 824-base-pair fragment containing the *Sgs-7*, *Sgs-8* intergenic region and portions of the *Sgs-8* and *Sgs-7* transcription units, was purified. The partially filled nDm9035 DNA was mixed with the partially filled *Sgs-7*, *Sgs-8* fragment and their now-complementary two-nucleotide 5' extensions (Hung and Wensink, 1984) were allowed to anneal and ligate at 4°C. Several colonies were recovered, whose plasmid DNA showed a single copy of each fragment joined together in the desired orientation. Four of these colonies, designated pGAO-1 (shown in Figure 1) through pGAO-4, were purified by CsCl gradient centrifugation, and the sequences of the joints between *Sgs-7* and *Adh* were determined (Maxam and Gilbert, 1977; 1980). All four clones were identical, based on restriction mapping and partial-fill joint sequence; the partial-fill joint had the nucleotide sequence predicted, and the *Hin*dIII linker was precisely positioned with respect to the *Adh* sequence. The sequence is shown in Figure 2. The junction between the two genes is in their 5' untranslated regions. The *Sgs-7* contribution begins in the 5' flanking sequence and ends at +25 bp in the 5' untranslated region, while the *Adh* contribution begins in the 3' flanking sequence and ends at +13 bp in the 5' untranslated region [relative to the proximal

promoter (Benyajati *et al.*, 1983)]. Five nucleotides probably from the synthetic *Hin*dIII linker are sandwiched between the two *Drosophila* gene segments.

(iii) *Construction of pGLAX1.0*

Purified DNA of the plasmid pGAO-1 was digested to completion with *Xba*I. Two synthetic deoxytetradecamers (5'C-T-A-G-G-G-T-C-G-A-C-G-G-G3' and 5'A-A-T-T-C-C-C-G-T-C-G-A-C-C3') were obtained from the Microchemical Facility of the Division of Biology, California Institute of Technology, and were purified using acrylamide gel electrophoresis. The deoxytetradecamers were phosphorylated using ATP and T4 polynucleotide kinase, then mixed together with, and ligated to, the *Xba*I-digested pGAO-1. After heat-inactivation of the T4 DNA ligase, the material was digested exhaustively with *Eco*RI. The 2.2 kb fragment that contains the altered *Sgs-7—Adh* gene was purified and then ligated to *Eco*RI-digested pOX4 (E.M. Meyerowitz, unpublished experiment). The plasmid pOX4 contains *Sgs-8* sequences beginning in the 5' flanking sequence and ending at the *Bgl*II site at +245 bp joined to the *E. coli lacZ* gene. The first 51 codons of the *Sgs-8* protein-coding region are joined to a synthetic-linker-derived codon, followed by the *E. coli* β-galactosidase-coding region beginning at the ninth codon.

One *E. coli* HB101 transformant clone, designated pGAZ-1 (shown in Figure 1), was recovered with the modified *Sgs-7—Adh* fusion gene inserted in the correct orientation with respect to the *Sgs-8—lacZ* fusion gene present in the pOX4 vector fragment. The plasmid pGAZ-1 contains two non-contiguous regions of the 68C gene cluster. The first is the 755-base-pair *Bgl*II-*Xba*I restriction fragment spanning the 5' ends of the *Sgs-8* gene and of the *Sgs-7* gene, which have been fused to *E. coli lacZ* and to *D. melanogaster Adh*, respectively. The second is the 268-base-pair *Xba*I-

*Hin*dIII restriction fragment that contains the *Sgs-8* 3' untranslated region and polyadenylation site, which is located downstream of the *Sgs-8—lacZ* fusion gene and oriented in the direction of *Sgs-8—lacZ* transcription.

The transformation vector pGLAX1.0 was constructed by digesting pGAZ-1 with *Sal*I and gel-isolating the 9.2 kb *Sal*I fragment that contained the *Sgs-7—Adh* and *Sgs-8—lacZ* fusion genes. Carnegie 20 (Rubin and Spradling, 1983) was digested with *Sal*I and ligated to the fusion gene material. Four apparently identical clones were isolated following transformation of *E. coli* HB101, and one was chosen for subsequent work. The composite P element contained within the plasmid is diagrammed in Figure 1.

(iv) *Sgs-8—lacZ* Bal*31 deletion construction*

The 4.2 kb *Xba*I-*Kpn*I fragment containing the entire *Sgs-8—lacZ* fusion gene was gel-isolated from pGLAX1.0 and ligated to similarly digested pUC18 DNA. The resulting clone was called nDm9800A.

In order to remove DNA upstream of the 5' end of the *Sgs-8—lacZ* gene, nuclease *Bal*31 was used. nDm9800A DNA was linearized with *Kpn*I, and appropriate conditions of *Bal*31 digestion were determined by setting up replica reactions that varied either the time of reaction or the concentration of *Bal*31. The amount of DNA removed by *Bal*31 was determined by digesting the reactions with *Ava*I, followed by electrophoresis through a 2.5% agarose gel.

*Bal*31 reactions in which the DNA had been digested to the desired extent were pooled. The termini were end-filled using the Klenow fragment of *E. coli* DNA polymerase I. Phosphorylated deoxydodecamer linkers containing the *Bgl*II cleavage

site (sequence 5'G-G-A-A-G-A-T-C-T-T-C-C3'; New England Biolabs, Beverly, MA) were ligated overnight at room temperature to the *Bal*31-treated DNA. NaCl was added to 0.1 M, *Xba*I and *Bgl*II were added and the excess linkers were released by exhaustive digestion. After gel electrophoresis and DE81 paper purification, the modified *Sgs-8—lacZ* gene fragments were ligated to the *Xba*I-*Bgl*II vector fragment of pGAO-1. Following transformation of calcium-treated *E. coli* HB101, miniprep DNAs from 121 clones were analyzed. The amount of DNA removed from the 5' flanking sequence was estimated by digesting the miniprep DNAs with *Bgl*II and *Ava*I, followed by electrophoresis through 2.5% agarose. *Hin*fI fragments of pBR322 and *Hin*dIII-*Eco*RI double-digest fragments of λ *cI*857 *S*7 were used as the size standards. Twenty-two clones were chosen for colony purification and are designated pDm9801 through 9822.

(v) *Sgs-7—Adh* Bal*31 deletion construction*

The 2.2 kb *Eco*RI fragment containing the *Sgs-7—Adh* gene was gel-isolated from pGAZ-1 and recloned in the 3.1 kb *Eco*RI vector fragment of pGAO-1. The resulting clone was called pGAO-5.

In order to remove DNA upstream of the 5' end of the *Sgs-7—Adh* gene, nuclease *Bal*31 was again used. pGAO-5 DNA was linearized with *Bgl*II. *Bal*31 calibration reactions were performed in parallel with the reactions for nDm9800A, except that the second enzyme used for the deletion length measurement was *Kpn*I rather than *Ava*I. Following the Klenow reaction on the pooled DNA and *Bgl*II linker ligation, excess digestion was performed with *Sal*I and *Bgl*II, and the insert fragments were isolated as a broad band ≈2.3 kb long. The material was ligated to the *Bgl*II-*Sal*I vector fragment of pGAO-5 DNA. Following transformation of calcium-treated *E. coli*

HB101, miniprep DNAs from 60 clones were analyzed. The amounts of DNA removed from each clone's 5' flanking sequence was estimated by digesting each miniprep DNA with *Bgl*II and *Kpn*I, followed by electrophoresis through 2.5% agarose. The same size standards were used as for the *Sgs-8—lacZ* deletion measurements. Thirty clones were chosen for colony purification and are designated pDm9701 through 9730.

The directional cloning strategy placed both the *Sgs-8—lacZ* promoter deletion molecules and the *Sgs-7—Adh* promoter deletion molecules in a single orientation within the same vector backbone. This helps to ensure that differences in function between clones can be ascribed to the amount of glue protein gene 5' flanking sequence present.

(vi) *Sequence determination of* Bal*31 deletion breakpoints*

The endpoints of five pDm9700-series and four pDm9800-series *Bal*31 deletion molecules were determined. Relevant double-digest restriction fragments were isolated following gel electrophoresis, ligated to pUC118 (Vieria and Messing, 1987), and transformed into *E. coli* TB1 (T. Baldwin, personal communication). After determination of the structures of miniprep DNAs, suitable molecules were transformed into the host strain *E. coli* MV1193 (Vieria and Messing, 1987). Individual MV1193 transformants were used to inoculate small cultures. Following superinfection with bacteriophage M13K07, single-stranded DNAs were purified by standard procedures prior to use as templates in dideoxy sequence determination reactions (Sanger *et al.*, 1977; Vieria and Messing, 1987). [α-^{35}S]dATP was used to label the reaction products. Acrylamide gel electrophoresis followed typical procedures; the gels were fixed and dried before autoradiography. The positions of the *Bal*31 deletion endpoints were assigned by comparing the newly determined junction sequences with the

published sequence of the *Sgs-7, Sgs-8* intergenic region (Garfinkel *et al.*, 1983). The *Drosophila* sequences present in the pDm9700-series clones begin at the following positions in the published sequence and continue rightward: pDm9718, position 1781; pDm9730, position 1878; pDm9721, position 1903; pDm9719, position 1975; pDm9729, position 2022. These correspond to the following positions relative to the 5' end of the *Sgs-7—Adh* fusion gene: pDm9718, –333 base-pairs; pDm9730, –236 base-pairs; pDm9721, –211 base-pairs; pDm9719, –139 base-pairs; pDm9729, –92 base-pairs. The sequences present in the pDm9800-series clones begin at the following positions in the published sequence and continue leftward: pDm9812, position 2071; pDm9803, position 2054; pDm9802, position 1937; pDm9804, position 1865. These correspond to the following positions relative to the 5' end of the *Sgs-8—lacZ* fusion gene: pDm9812, –432 base-pairs; pDm9803, –415 base-pairs; pDm9802, –298 base-pairs; pDm9804, –226 base-pairs.

(vii) *Construction of pDm9800E*

A *Sgs-8—lacZ* fusion gene with the entire *Sgs-7, Sgs-8* intergenic region was placed into the same orientation as the *Sgs-8—lacZ* promoter deletion clones in the following way. Plasmid DNA from nDm9800A was digested with *Asp*718 and end-filled. The DNA was prepared for directional cloning by ligating deoxyoctamer *Bgl*II linkers (sequence 5'C-A-G-A-T-C-T-G3', obtained from New England Biolabs, Beverly, MA) to the flush ends of the *Asp*718 site, heat-inactivating the ligase, redigesting with a mixture of *Bgl*II and *Xba*I, and gel-isolating the fragment that contains all the 68C-derived DNA sequences and the *E. coli lacZ* coding region. The same *Bgl*II-*Xba*I vector fragment from pGAO-1 used for the *Sgs-8—lacZ Bal*31 deletion cloning was ligated to the gel-isolated *Sgs-8—lacZ* fragment. After *E. coli* HB101 was transformed with the ligation mixture, one clone with the desired structure

was isolated and named pDm9800E. As expected from the sequences of the *Bgl*II linker and the filled-in overhang due to *Asp*718 cleavage, the G-G-T-A-C-C site recognized by both *Asp*718 and *Kpn*I was reconstructed.

(viii) *Construction of additional germline transformation plasmids*

Selected *Sgs-7—Adh Bal*31 promoter deletion molecules and *Sgs-8—lacZ Bal*31 promoter deletion molecules were recloned into *Sal*I-digested Carnegie 20 for use in germline transformation. The pDm9700-series plasmids were linearized with *Bgl*II, the termini made flush, and ligated to *Sal*I deoxyoctamer linkers. Exhaustive digestion with *Sal*I removed the excess linkers and cleaved a second restriction site present in each clone, releasing the *Sgs-7—Adh* gene fragments, which were purified following agarose gel electrophoresis. The pDm9800-series plasmids were completely digested with *Bgl*II and *Asp*718 before Klenow treatment; the subsequent treatments to recover the approximately 4-kb insert fragments were identical to the pDm9700-series treatments. After ligation and bacterial transformation, vector-insert ligation products were identified by colony filter hybridization (Grunstein and Hogness, 1975). They were colony purified, grown in small cultures from which miniprep DNAs were isolated and characterized. The pDm9700-series P element transformation vector derivatives used in this report are called pGAX0.26, pGAX0.24, and pGAX0.12, in recognition of their sequence content. The pDm9800-series P element transformation vector derivatives used in this report are called pGLX0.68 and pGLX0.66, in recognition of their sequence content. The composite P elements contained in these plasmids are diagrammed in Figure 1.

3. Results

(a) *Recovery of germline transformants that express both alcohol dehydrogenase and β-galactosidase in third larval instar salivary glands*

A mixture of phsπ and pGLAX1.0 (shown in Figure 1) plasmid DNAs was injected into syncytial-cleavage-stage embryos of the strain Adh^{fn6} *cn*; ry^{502}. Out of the fifteen fertile adults recovered, two adults produced ry^+ progeny. One adult produced six ry^+ individuals; the other produced fifteen. Genetic analysis of the twenty-one sublines revealed a total of six distinct insertion events. The sublines derived from the first injected fly contain five different insertion events, based on both genetic and whole-genome Southern (1975) analyses. These are: (i) one line with an insertion located on the X-chromosome; (ii) one line with an insertion located on the second chromosome; (iii) one line that has a complex homozygous phenotype and has an insertion into chromosome 2; (iv) one line that has a homozygous-lethal insertion into the third chromosome; and (v) one insertion located on chromosome 3, which was isolated twice. The fifteen sublines descended from the second injection survivor were analyzed similarly. All of them mapped to the third chromosome and one subline had a homozygous-lethal third chromosome. The homozygous-lethal subline and one of the viable sublines were analyzed further by genomic DNA gel blot hybridization—both had exactly the same insertion event as defined by the junction fragments. We infer that all of these fifteen transformed lines represented a premeiotic cluster of P element insertion, and that the lethality in the one subline is of separate origin. These insertion events are designated *Tf()GLAX1.0*, following the convention of Crosby and Meyerowitz (1986), and are summarized in Table 1.

(i) *Histochemical assay of transformant third instar larvae*

Third instar larvae from each strain were dissected and processed for histochemical staining. Both ADH and β-gal enzyme activities were seen in the salivary glands. Glands from the *Tf(1)GLAX1.0-1* strain are shown in Figures 3A and 3B. In Figure 3A, one salivary gland lobe from a transformed animal was stained for alcohol dehydrogenase activity. The staining is purple-black, which is reproduced here as black. Two salivary gland lobes from another animal are stained for β-galactosidase activity in Figure 3B. The natural color of the staining is deep blue and is reproduced here as black as well. Evidence that the two fusion genes are transcribed and translated with tissue-specificity is shown by the absence of staining in the strip of fat body that adheres to each salivary gland, and by the absence of staining in the most-anterior cells of each salivary gland, cells that do not synthesize and secrete glue proteins as determined by ultrastructural examination (see Berendes and Ashburner, 1978). When third instar larval carcasses of the transformant strains are incubated with the alcohol dehydrogenase activity stain, no enzyme activity can be detected in any other tissue. Figures 3C and 3D show that the Adh^{fn6} *cn*; ry^{502} host strain fails to make either enzyme activity in the third instar salivary gland under our assay conditions; the ADH histochemical reagent reacts in no larval tissue, while a β-galactosidase presumed to be coded for in the *Drosophila* genome is apparent only in the midgut (data not shown; see also Farnogli *et al.*, 1987). In Adh^{+} larvae, the normal distribution of ADH enzyme includes a variety of tissues such as the fat body, Malpighian tubules, and midgut, most prominently (Ursprung *et al.*, 1970).

The salivary gland pattern of uniform, intense histochemical staining described for *Tf(1)GLAX1.0-1* is also seen in third instar larvae from four of the other insertion

events: *Tf(2)GLAX1.0-2*, *Tf(3)GLAX1.0-3*, *Tf(3)GLAX1.0-4/TM3, Sb* ry^{RK} *Ser*, and *Tf(3)GLAX1.0-7*.

A somewhat different result is observed with the sixth insertion, that present in the *Tf(2)GLAX1.0-5/CyO* strain: salivary gland histochemical staining reactions for both ADH and β-gal enzyme activities are weak and variegated. This insertion-bearing chromosome is essentially lethal, with occasional homozygous escapers that are apparently sterile.

(ii) *Enzyme activity assay of transformant third instar larval salivary glands*

Quantitative measurements of *Sgs-7—Adh* fusion gene expression in the homozygous-viable insertion strains were obtained by use of a soluble extract assay (Sofer and Ursprung, 1968). Late third instar larvae were dissected, and salivary glands that met the morphological criterion of approximately puff stage 4 (large glands with thin lumens along their length) were generally chosen. For the X-chromosome strain *Tf(1)GLAX1.0-1*, third instar larvae were not separated by sex before extirpation of the salivary glands and homogenization; we do not know if X-chromosome dosage compensation had been acquired by this insertion event. Portions of each extract were placed in a spectrophotometer cuvette that contained assay buffer (Sofer and Ursprung, 1968). Reduction of NAD^+ was monitored by absorbance at 340 nanometers. The rates of the change in absorbance were converted into Units of enzyme activity. Individual measurements of the reduction of NAD^+ in salivary gland extracts from the four homozygous-viable transformant lines ranged from 13 Units per animal-equivalent to about 60 Units per animal-equivalent (Table 2A). Each line was measured six independent times, and the average of the measurements is 32.2 Units per animal-equivalent, with a standard deviation of 11.3. Certain measurements might be

underestimates: if a given extract was prepared from a mixture of salivary glands that included younger stages, the accumulation of ADH activity would be expected to be less.

A representative of the ADH soluble extract assay procedure is shown in Figure 4A. Salivary gland extracts were prepared from *Tf(2)GLAX1.0-2* third instar larvae, and from Adh^{fn6} *cn*; ry^{502} third instar larvae. Each was reacted for 15 minutes, with the A_{340} measured automatically every 30 seconds. In the *Tf(2)GLAX1.0-2* reaction, partial saturation is evident, and the first eleven data points were used for the calculation of the line drawn in the figure.

We can estimate how much ADH protein was synthesized in the salivary glands of these transformant strains. Sofer and Ursprung (1968) purified *Drosophila melanogaster* ADH, assaying the activity using 2-butanol as the alcohol substrate, and obtained a value of 94,000 Units/mg protein. Lee (1982), using two purification schemes that each differ from the Sofer and Ursprung (1968) method, reported additional values of 225,000 Units/mg protein and 347,000 Units/mg protein for ADH enzyme assayed with 2-propanol. The variation among the three values may be partly due to the fact that 2-propanol is a better substrate that 2-butanol (Sofer and Ursprung, 1968). The mass of ADH protein present in each transformed animal's salivary gland tissue can be estimated to range from 13 Units divided by 94,000 U/mg protein (roughly 140 ng ADH protein) up to 60 Units divided by 94,000 U/ mg protein (roughly 640 ng ADH protein), with the average measurement of 32 Units converting to 340 ng ADH protein. Korge (1977) showed that the average salivary gland lobe contains about 1 μg of secreted protein in the lumen. There are two lobes per animal, or about 2 μg of salivary gland secretion protein. The sgs-3 protein has been estimated

to be about one-quarter to one-third of this mass, or about 600 ng per animal. This value cannot be simply compared to our estimates of ADH protein mass, because Korge (1977) measured the protein accumulation at the very end of third instar, using immobile larvae that were on the verge of pupariating, wheras we used animals that were several hours younger. Even so, these calculations serve to support our proposal that factors such as diminished RNA stability, diminished protein stability, or reduced promoter function as a consequence of the gene fusion breakpoints, are not large.

Quantitative measurement of *Sgs-8—lacZ* gene activity was obtained spectrofluorometrically. We were able to measure β-galactosidase activity in soluble extracts prepared in exactly the same way as for the *Sgs-7—Adh* measurements. Of the six independent extract preparations, only three were also used in the β-galactosidase assay. Figure 4B shows the results when 50 µl of the same *Tf(2)GLAX1.0-2* and Adh^{fn6} *cn*; ry^{502} salivary gland extracts used in Figure 4A were subjected to the fluorescence assay. From the slopes of the regression lines, we calculated unit activities in each extract. The β-galactosidase determinations are recorded in Table 2B. The average of the twelve measurements is 330 Units per animal-equivalent, with a standard deviation of 169.

Unlike the measurements for salivary gland ADH, the measurements of salivary gland sgs-8-β-galactosidase enzyme activity cannot be used to estimate the amount of hybrid protein product accumulated. Since the sgs-8-β-galactosidase fusion protein has not been purified, its kinetic properties for cleavage of any of the commonly used galactoside substrates are unknown. The standard β-galactosidase unit is defined in terms of micromoles of *o*-nitrophenol released per minute by cleavage of *o*-nitrophenyl-β-D-galactopyranoside (see Miller, 1972, for example), and it is known that different

substrates are recognized and cleaved by the native *E. coli* enzyme with different kinetics (Wallenfels, 1962). Extrapolating from our 4-methylumbelliferone release measurements of sgs-8-β-galactosidase activity to the *o*-nitrophenol release units of the natural *E. coli* enzyme would require too many assumptions.

In Figure 4C, we plot β-galactosidase activity versus alcohol dehydrogenase activity in the extracts that were subjected to both kinds of measurement. There is, as expected, a positive correlation between the two enzyme activities: extracts that contained more ADH contained more β-gal as well.

(iii) *RNA gel blot analysis of transformant third instar larvae*

As described above, histochemical reactions for β-gal and ADH showed that third instar larvae of the transformant strains express these enzymatic activities only in the salivary gland. In the case of β-galactosidase activity, this can be due only to the correct removal of the *Sgs-8* intervening sequence from the *Sgs-8—lacZ* fusion gene primary transcript, export of the transcript from the nucleus, and translation in salivary gland cells' cytoplasm. In the case of alcohol dehydrogenase activity, this requires the correct removal of both of the *Adh* intervening sequences from the *Sgs-7—Adh* fusion gene primary transcript, export of the resulting RNA species from the nucleus, and translation of the RNA in the cytoplasm of salivary gland cells.

RNA gel blots were performed using nucleic acids derived from third instar larval salivary glands and from the carcasses remaining following dissection. Figure 5 shows the results when salivary gland RNA from the third instar larvae of the *Tf(1)GLAX1.0-1* strain are compared with salivary gland RNA from Adh^{fn6} *cn*; ry^{502}; similar results were obtained with the other *Tf()GLAX1.0* strains. A nick-translated ^{32}P-labelled *E. coli* β-galactosidase gene hybridized with two transcripts of

approximately 3.6 kb length present in *Tf(1)GLAX1.0-1* transformant salivary gland RNA (Figure 5A). Figure 5B shows that the radiolabelled *Adh* gene probe hybridizes with a 1.1 kb RNA in the transformant salivary gland RNA. The *Sgs-7* probe hybridized only with the normal *Sgs-7* mRNA; apparently, the 25 nucleotides of homology present in the 5' untranslated region of the *Sgs-7—Adh* fusion transcript are too few to be detected (Figure 5C). Finally, an *Sgs-8* probe hybridizes with three transformant salivary gland transcripts, the normal *Sgs-8* mRNA and the two transcripts homologous with the *lacZ* probe, although the RNA gel blot in Figure 5D shows only the hybridization with the *Sgs-8* mRNA. In addition, the tissue specificity of enzyme accumulation revealed by the histochemical staining assay was verified by the tissue specificity of RNA accumulation revealed by RNA gel blot hybridization. RNA prepared from the carcasses of third instar larvae after the removal of the salivary glands did not hybridize with any of the radiolabelled DNA probes (data not shown).

(iv) *Appearance and disappearance of salivary gland enzyme activities*

The time of fusion gene expression in *Tf()GLAX1.0* salivary glands was inferred from observations of the protein products. Third instar larvae were dissected, and the salivary glands were removed and sorted into classes that roughly correspond to feeding early-to-mid third instar (glands small to medium in size, with spheroidal cell morphology, inferred to be puff stage 1); wandering third instar (glands large but with no or small lumens, inferred to be puff stages 2-4); and very late third instar (glands very large, with swollen, glue-filled, lumens, inferred to be approximately puff stage 8). Histochemical reactions of first category salivary glands revealed that ADH and β-gal activities were variable, ranging from uniform and darkly staining to nonuniform and lightly staining. Histochemical staining of second category salivary glands revealed greater amounts of enzymes that were uniformly distributed in the secretory cells of the

gland. In the oldest category, β-galactosidase activity was found inside the lumens and the ducts of the salivary glands, but not within the cells of the glands themselves. In contrast, the cells of very late salivary glands continue to stain for alcohol dehydrogenase, but at apparently reduced levels.

(v) *Analysis of integrated DNA*

DNA was isolated from adult flies of the transformed strains. The DNAs were digested with either *Bam*HI, *Eco*RI, *Pvu*I, or *Sac*I and whole-genome Southern (1975) blot filters were prepared. The filters were hybridized in succession with five ^{32}P-labelled cloned plasmid DNA probes (two from the *rosy* locus, one from the *Adh* locus, one from the *Sgs-7*, *Sgs-8* locus, and one of the *E. coli lacZ* protein-coding region), showing that all the transformant lines possess a single, unrearranged, insertion of the composite P element comprised of *ry*$^{+}$, *Sgs-7—Adh*, and *Sgs-8—lacZ*. Figure 6 shows representative results of this kind of experiment. Adult fly DNAs from all eight *Tf()GLAX1.0* strains and from *Adh*^{fn6} *cn*; *ry*^{502} were digested with *Bam*HI or with *Sac*I, separated by size using agarose gel electrophoresis, blotted to nitrocellulose filters, and hybridized with a ^{32}P-labelled genomic DNA subclone that detects the endogenous *rosy* locus and one junction fragment from each transformant strain. Data of this sort prove that the *Tf(3)GLAX1.0-3* insertion is identical to that present in *Tf(3)GLAX1.0-6* and that the insertions present in *Tf(3)GLAX1.0-7* and *Tf(3)GLAX1.0-8/TM3* are identical.

(vi) *Interaction with the* trans*-regulatory mutation,* l(1)npr-1

The *lethal(1)non-pupariating-1* mutation, *l(1)npr-1*, is a late-larval lethal mutation that fails to pupariate (Kiss *et al.*, 1976, 1978). It is one of the so-called "long" alleles of the overlapping complementation complex that maps to the X-

chromosome region 2B5 (Belyaeva *et al.*, 1980). In salivary glands of hemizygous mutant third instar larvae, several phenotypes are seen: (i) failure of intermolt puff regression and of ecdysterone-inducible puff formation (Belyaeva *et al.*, 1981); (ii) absence of accumulation of 68C glue protein gene transcripts (Crowley *et al.*, 1984); and (iii) absence of 68C glue gene RNA synthesis (Crowley *et al.*, 1984). We tested the effect this mutation has on the expression of the introduced glue protein fusion genes present in three autosomal, homozygous-viable transposition events.

Virgin females of the strain *y l(1)npr-1 w mal/Binsn* were mated to four different kinds of males: the non-transformed host strain Adh^{fn6} *cn*; ry^{502}, *Tf(2)GLAX1.0-2*, *Tf(3)GLAX1.0-3*, and *Tf(3)GLAX1.0-7*. Third-instar larval progeny from these crosses fall into three classes: (i) all females are ignored because they are heterozygous for a wild-type X chromosome derived from their fathers and one of the X-chromosomes from their mothers; (ii) males hemizygous for the *l(1)npr-1*$^+$ *Binsn*-balancer chromosome, which are recognizable by their *white*$^+$ Malpighian tubules and *yellow*$^+$ jaw hooks; and (iii) males hemizygous for the *y l(1)npr-1 w mal* chromosome, which are recognizable by their *white* Malpighian tubules and *yellow* jaw hooks. The class ii males constitute the wild-type control, and the class iii males constitute the *l(1)npr-1* mutant experimental.

The third instar larvae from these crosses were sorted by sex, and the males were sorted by the colors of the Malpighian tubules and of the mouth parts. Males were dissected, the salivary glands were removed with segments of fat body tissue and the mouth parts attached, and the tissue processed for histochemical staining. In Figure 7, panels B and F, experimental male larvae that carried one copy of the *Tf(2)GLAX1.0-2* chromosome and were hemizygous for *l(1)npr-1* showed no histochemical staining for

either β-galactosidase or alcohol dehydrogenase in their salivary glands, while salivary glands from control males that were hemizygous *l(1)npr-1*$^+$ and heterozygous for the *Tf(2)GLAX1.0-2* chromosome stained for both enzymes (Figure 7, panels A and E).Therefore, expression of both histochemically assayable *Sgs* fusion genes requires the *l(1)npr-1*$^+$ gene product. Salivary glands dissected from males descended from the non-transformed host strain cross of both X-chromosome genotypes failed to react with histochemical reagents (Figure 7, panels C, D, G, and H). No effect of hemizygosity for the *l(1)npr-1* mutation could be seen on the *Adh*$^+$ allele that is made heterozygous and that contributes ADH histochemical reaction in the fat body (shown in Figure 7B, D), the anterior midgut and the gastric caecae (data not shown).

(b) *A "transient expression" assay system*

The germline transformants of the *Tf()GLAX1.0* series showed that the 755 base-pairs of 68C DNA sequence residing between the *Bgl*II site of *Sgs-8*, inside that gene's protein-coding region, and the *Xba*I site of *Sgs-7*, inside that gene's 5' untranslated region, contain the *cis*-acting regulatory information necessary to direct bidirectional transcription with spatial and temporal specificity.

We set out to determine the functional organization of this interval. Our experimental design made use of the histochemical markers joined to the glue protein gene transcription units, *Bal*31 deletion derivatives that removed portions of the intergenic DNA sequences, the "transient expression" assay first described by Martin *et al.* (1986) for the rapid test of these deletion derivatives, and then used germline gene transfer to verify the transient expression assay results.

To establish a baseline of *Sgs-7—Adh* and *Sgs-8—lacZ* gene function in the transient expression assay, plasmid pGAZ-1 DNA was injected into the anterior regions of syncytial-cleavage-stage embryos of the Adh^{fn6} *cn*; ry^{502} strain. Third-instar larvae that had developed from the surviving embryos were dissected and their salivary glands were subjected to either the histochemical procedure that detects alcohol dehydrogenase enzyme activity, or that which detects β-galactosidase enzyme activity. As shown in Figure 8, mosaic patches of enzyme activity stain appear. A large fraction of the animals tested showed at least some expression; these varied in the number of cells that reacted and were often bilaterally asymmetric. Table 3 contains the data from these experiments. Both the fraction of salivary gland lobes that were scored as positive and the subjective estimation of patch size were used to assign the degree of gene function in these experiments. The application of the transient assay, and the verification of those results by germline gene transfer of the *Bal*31 deletion derivatives, will be described for both glue protein gene fusions, beginning with *Sgs-7—Adh.*

(c) *Deletion analysis of the* Sgs-7—Adh *fusion gene*

(i) *Transient expression assay of* Sgs-7—Adh *fusion gene deletions*

Five *Sgs-7—Adh* promoter deletion plasmids were tested in the transient expression assay. These are pDm9718, pDm9730, pDm9721, pDm9719, and pDm9729, possessing 333 base-pairs, 236 base-pairs, 211 base-pairs, 139 base-pairs, and 92 base-pairs, respectively, of the 5' flanking DNA sequence. The first two plasmids gave high levels of ADH enzymatic activity, that is, the injected animals that were mosaic had large patches of intensely-staining cells, while pDm9721 and pDm9719 apparently produced less ADH enzyme, based upon both the fraction of salivary glands that stained with the ADH histochemical reagent (Table 4), and the

subjective estimation of mosaic patch sizes and staining intensities. Thus, sequences between –211 bp and –236 bp, or perhaps spanning –211 bp, are the most distal boundary of a region that seems to elevate *Sgs-7* promoter function in this assay. The staining results with pDm9729 were negative, suggesting that sequences between 92 base-pairs and 139 base-pairs upstream of the *Sgs-7* start site (or perhaps spanning the –92 base-pair deletion endpoint) define a distal boundary for promoter function in the transient expression assay.

(ii) *Germline gene transfer assay of* Sgs-7—Adh *fusion gene deletions*

Based on these results, we transferred the *Sgs-7—Adh* gene segments from pDm9730, pDm9721, and pDm9729 into the Carnegie 20 P element vector. The resulting germline transformation vectors were all screened for the relative orientation of the *Sgs-7—Adh* gene, and a set of vectors with the same orientation was used. The resulting transformation vectors were called pGAX0.26, pGAX0.24, and pGAX0.12, respectively, in recognition of the content of each plasmid: glue protein gene fused to alcohol dehydrogenase with xanthine dehydrogenase as the selectable marker, and the total of 68C DNA in kb present (recalling that the *Sgs-7—Adh* fusion gene has 25 bp of *Sgs-7* 5' untranslated sequences).

Adh^{fn6} *cn*; *ry*^{502} syncytial-cleavage stage embryos were microinjected with a mixture of each transformation vector and the non-transposing helper plasmid phsπ. Germline transformants were recovered, genetically mapped by outcrossing to *T(2,3)Ata/CyO; TM3* and backcrossing the *Cy Sb ry*$^{+}$ progeny, and autosomal insertion events made homozygous by self-crossing sibling *Cy Sb ry*$^{+}$ individuals. For the –92 base-pair and the –211 base-pair *Sgs-7—Adh* fusion gene molecules, five autosomal homozygous-viable, homozygous-fertile transformant insertion lines were

chosen for subsequent experiments. These are called *Tf()GAX0.12* and *Tf()GAX0.24*, respectively. For the –236 base-pair *Sgs-7—Adh* fusion gene molecule, three autosomal homozygous-viable, homozygous-fertile transformant lines were obtained and used in subsequent experiments, and these are called *Tf()GAX0.26*.

In order to test for the copy number and integrity of the inserted sequences, adult fly DNA was prepared from each of these thirteen transformed lines. After digestion with either *Bam*HI or *Eco*RI, two-microgram portions were placed into wells of five different 0.5% agarose gels, separated by size, and each gel blotted to nitrocellulose. Each filter was hybridized with a single ^{32}P-labelled cloned plasmid DNA. These were two from the *rosy* locus, one from the *Adh* locus, one from the *Sgs-7, Sgs-8* locus, and the prototype P element plasmid pπ25.1 (Spradling and Rubin, 1982). Twelve of the lines possessed a single, unrearranged insertion event of the composite P element containing the *ry*$^{+}$ marker and the *Sgs-7—Adh* fusion gene; the thirteenth line, *Tf(3)GAX0.24-3*, showed clear evidence for a second insertion event that was either homozygous, or segregating in a fairly large fraction of the adult flies that contributed the DNA.

To test for tissue specificity and to derive a semiquantitative estimation of ADH activity, third instar larval salivary glands with adhering tissue were dissected, fixed in glutaraldehyde, washed extensively, and reacted with the ADH histochemical reagent. *Tf()GAX0.26* strains containing the –236 base-pair *Sgs-7—Adh* fusion gene gave a strong histochemical staining reaction, as predicted from the transient assay result with pDm9730; *Tf()GAX0.24* strains containing the –211 base-pair *Sgs-7—Adh* fusion gene also gave a strong histochemical reaction; *Tf()GAX0.12* strains bearing the –92 base-pair *Sgs-7—Adh* fusion gene failed to stain in the histochemical test, as predicted

from the transient assay of pDm9729. Histochemical staining of visceral tissue showed no ADH activity in the other internal organs of transformant larvae.

Soluble extract assays of third instar larval salivary glands gave the measurements in Table 5. Contrary to the prediction from the transient expression assays, the soluble extract measurements failed to reveal a significant difference between transformed lines carrying the –236 base-pair-containing *Sgs-7—Adh* fusion gene construction and transformed lines carrying the –211 base-pair containing *Sgs-7—Adh* construction. The average of all the measurements of the *Tf()GAX0.26* strains was 32 Units per animal-equivalent of salivary gland extract, with a standard deviation of 11. The average of all the measurements of the *Tf()GAX0.24* strains was 43 Units per animal-equivalent of salivary gland extract, with a standard deviation of 12. These ADH activity measurements are very similar to the values obtained from the *Tf()GLAX1.0* series. The average of the *Tf()GAX0.12* measurements was -4.2×10^{-3} Units per animal-equivalent of salivary gland extract, with a standard deviation of 0.5. They were judged to be lacking in alcohol dehydrogenase activity.

To test whether the absence of ADH enzyme activity in the –92-base-pair *Sgs-7—Adh* transformant fly lines was due to an absence of chimeric RNA accumulation, the experiment shown in Figure 9 was performed. Total nucleic acids from third instar larval salivary glands obtained from the five *Tf()GAX0.12* lines homozygous for this construction were subjected to RNA gel blot hybridization and failed to reveal any salivary gland RNA homologous with the nick-translation-labelled *Adh* gene probe used.

The results of these experiments show that the *Sgs-7* sequences required in *cis* for correct tissue and quantity of expression in the *Sgs-7—Adh* fusion gene test system

reside between –211 base-pairs (possibly –139 base-pairs, depending on the interpretation of the pDm9719 transient expression result) and –92 base-pairs relative to the start of *Sgs-7* gene transcription.

(d) *Deletion analysis of the* Sgs-8—lacZ *fusion gene*

(i) *Transient expression assay of* Sgs-8—lacZ *fusion gene deletions*

The 475 base-pair intergenic region attached to the *Sgs-8—lacZ* fusion gene, in the form of the plasmid pDm9800E, was tested by the transient expression assay. Strong staining for β-galactosidase activity, like that in the pGAZ-1 control, was observed in many salivary glands of third instar larvae that survived the microinjections. Four deletion derivatives were tested: pDm9812, with 432 base-pairs of 5' flanking sequence remaining; pDm9803, with 415 base-pairs remaining; pDm9802, with 298 base-pairs remaining; and pDm9804, with only 226 base-pairs remaining. The plasmid retaining the most 5' flanking DNA sequence, pDm9812, gave β-galactosidase activity that was comparable to the staining observed with the pGAZ-1 control plasmid. Neither pDm9803, pDm9802, nor pDm9804 exhibited any β-galactosidase activity. Like the pDm9700-series of plasmids, the plasmids pDm9800E, pDm9812, pDm9803, pDm9802 and pDm9804 are a consistent deletion set in that all five plasmid DNAs have the *Sgs-8—lacZ* fusion gene oriented in the same way, cloned between the same restriction sites, of the same vector fragment. These results indicated that *Sgs-8—lacZ* fusion gene expression in the somatic transformation assay requires a sequence element either between –432 bp and –415 bp relative to the *Sgs-8* transcription initiation site, or spanning the –415 bp position. The transient expression results are summarized in Table 6.

(ii) *Germline gene transfer assay of* Sgs-8—lacZ *fusion gene deletions*

Based on the pDm9800-series transient expression results, we transferred the *Sgs-8—lacZ* gene segments from pDm9812 and pDm9803 into Carnegie 20. The resulting germline transformation vectors were screened for the relative orientation of the *Sgs-8—lacZ* gene, and a pair of vectors with the same orientation was used. These plasmids are called pGLX0.68 and pGLX0.66, in recognition of their content and the quantity of 68C DNA present in them. Unlike the pGLAX1.0 plasmid, these plasmids lack the 268-base-pair fragment containing the *Sgs-8* 3' untranslated region and polyadenylation site.

Germline transformants were recovered and genetically mapped as before. Five autosomal homozygous-viable, homozygous-fertile transformant lines of each P element transformation vector [*Tf()GLX0.68* and *Tf()GLX0.66* lines] were chosen for subsequent experiments.

Copy number and integrity of the inserted sequences were tested by whole-genome Southern gel blot filter hybridizations of adult fly DNA prepared from the ten *Tf()GLX* lines. The hybridization probes used were two from the *rosy* locus, one from the *Sgs-7, Sgs-8* locus, one containing the *E. coli lacZ* protein-coding region, and the prototype P element plasmid pπ25.1. All ten lines appear to possess a single, unrearranged insertion event of the composite P element containing the ry^+ marker and the *Sgs-8—lacZ* fusion gene.

To test for tissue specificity and to derive a qualitative estimate of sgs-8—β-galactosidase enzyme activity, third instar larval salivary glands with adhering tissue were dissected and incubated in the X-Gal-containing histochemical reagent. *Tf()GLX0.68*-homozygous third instar larval salivary glands accumulated amounts of

β-galactosidase enzyme activity roughly similar to amounts found in the *Tf(1)GLAX1.0-1* third instar larval salivary glands (Figure 10A). This confirmed the transient expression result with the pDm9812 plasmid. The –415-base-pair-containing *Sgs-8—lacZ* fusion gene present in *Tf()GLX0.66*-homozygous transformant stocks gave very faint β-galactosidase staining confined to the third instar larval salivary gland (Figure 10B). This is also consistent with the transient expression assay result obtained with the pDm9803 plasmid, given that particularly low levels of enzyme activity derived from *Sgs* fusion genes are difficult to detect in the transient assay system (K. Vijay Raghavan, M. Roark, C. Mayeda and E.M. Meyerowitz, manuscript in preparation). Histochemical staining of visceral tissue showed β-galactosidase activity only in the midgut, which, again, could be attributed to the *Drosophila*-encoded enzyme.

Soluble extract assays of *Sgs-8—lacZ* expression in third instar larval salivary glands gave the measurements in Table 7. The mean Units per animal-equivalent and standard deviation of the fifteen measurements of the *Tf()GLX0.68* strains are 364 ± 280. The large standard deviation arises partially in the following way. Four of the five *Tf()GLX0.68* strains, measured three times each, gave a mean value of 235 β-gal Units per animal-equivalent with a standard deviation of 73. The fifth strain, *Tf(2)GLX0.68-4*, gave a substantially higher figure, 880 ± 142, which may reflect a *bona fide* example of quantitative position effect (Spradling and Rubin, 1983), or may possibly reflect a second insertion event that has a junction fragment undetected in the adult fly DNA gel blot hybridizations. The *Tf()GLX0.66* strains present a more difficult case of quantification. The X-gal staining unequivocally showed salivary gland β-galactosidase enzyme activity (Figure 10B). However, the fifteen soluble extract measurements (five lines measured three times each) gave a mean value of 24.7 β-gal Units per animal-

equivalent with a standard deviation of 9.1. If we subtract the 12.2 Units per animal-equivalent seen in the *Adh*^{fn6} *cn*; *ry*^{502} host strain, we get 12.5 Units per animal-equivalent for the *Tf()GLX0.66* strains, 352 (or 223) Units per animal-equivalent for the *Tf()GLX0.68* strains, and 318 Units per animal-equivalent for the *Tf()GLAX1.0* strains. Germline-transformant expression of the three constructions stand in the ratio of 100% for the *Tf()GLAX1.0* strains, 110% (or 70%) for the *Tf()GLX0.68* strains, and 4% for the *Tf()GLX0.66* strains. What seems certain is that the *Sgs-8—lacZ* gene expression seen in the *Tf()GLAX1.0* strains and the *Tf()GLX0.68* strains is at least twentyfold greater than the *Sgs-8—lacZ* expression seen in the *Tf()GLX0.66* strains, and that this difference is due to the 17 bp difference in the amount of *Sgs-8* 5' flanking sequence.

4. Discussion

(a) Sgs-7 *and* Sgs-8 *are regulated by sequences separate from those controlling* Sgs-3

Germline transformants of the *Tf()GLAX1.0* series demonstrated that the sequences required for tissue- and stage-specific expression of the *Sgs-8—lacZ* fusion gene and of the *Sgs-7—Adh* fusion gene reside between the *Bgl*II site in the *Sgs-8* protein-coding region and the *Xba*I in the *Sgs-7* 5' untranslated region. This restriction fragment of the 68C gene cluster does not overlap the restriction fragment shown by Crosby and Meyerowitz (1986) to support full-level of tissue- and stage-specific expression of the *Sgs-3* gene. Thus, the 68C glue protein gene cluster must contain a minimum of two tissue- and stage-specific regulatory elements, two quantitative control elements, and three RNA polymerase II transcription initiation elements.

The fact that the expression of the two glue protein fusion genes in the *Tf()GLAX1.0* transformant lines is dependent upon the *l(1)npr-1*$^+$ gene product suggests three conclusions. First, the process *l(1)npr-1*$^+$ regulates (either directly or indirectly) is promoter-specific since the *Alcohol dehydrogenase* gene, with its proximal promoter that functions in a variety of larval tissues, directs the accumulation of substantial amounts of ADH activity in the fat body, midgut, and gastric caecae of *l(1)npr-1*-mutant larvae. [A formal possibility that has not been tested is that the switch in *Adh* transcription from the proximal promoter to the distal promoter which occurs during the third larval instar (Benyajati *et al.*, 1983; Savakis *et al.*, 1986) is blocked by the *l(1)npr-1* mutation.] Second, the process of 68C glue protein gene expression controlled (either directly or indirectly) by the *l(1)npr-1*$^+$ gene product is most likely to be gene transcription. The expression of the *Sgs-7—Adh* fusion gene, which contains

only 25 nucleotides of the glue protein gene transcription unit, requires the *l(1)npr-1*$^+$ gene product. Any model for *l(1)npr-1*$^+$ function that requires a sequence-specific interaction with the body of the *Sgs-7* glue protein gene or of its primary transcript (the intervening sequence, for example) is untenable. Third, at 68C there is a minimum of two sites of action for *l(1)npr-1*$^+$, since the non-overlapping segment of the genome that allows *Sgs-3* expression at high level also retains the requirement for this genetic function (Crowley *et al.*, 1984).

(b) Sgs-7 *and* Sgs-8 *expression may depend upon the same regulatory elements*

The specific hypothesis tested by the experiments reported here is that the homologous sequences present in the 100 base-pairs immediately upstream of both the *Sgs-7* gene and the *Sgs-8* gene represented functionally conserved *cis*-acting regulatory elements. Our results refute this hypothesis. A deletion that removes part of the right copy causes a substantial reduction in the expression of the *Sgs-8—lacZ* fusion gene when assayed either by transient expression or by germline transformation. Deletions that remove the entire left copy have no effect on the expression of the *Sgs-7—Adh* fusion gene when assayed by germline transformation.

The non-equivalence of the sequence elements upstream of *Sgs-8* and *Sgs-7* may be interpreted in light of the sequence motif that Mestril *et al.* (1986) have identified as required for ecdysterone-induced *Hsp23* gene expression. The motif they observed between –228 bp and –192 bp relative to the start of *Hsp23* transcription is 5'A-T-T-T-T-C-C-A-T3' separated by 19 base-pairs from 5'A-T-G-G-C-A-G-A-T3'. The first portion of their sequence motif is an imperfect inverted repeat (seven out of the nine positions) of the second portion. Mestril *et al.* (1986) also noted that related

sequences, with a variable number of nucleotides separating the two components, could be found upstream of other members of the small heat shock protein gene cluster at 67B, among other ecdysterone-regulated genes. P.H. Mathers (personal communication) first noted that the conserved sequence regions upstream of *Sgs-7* and *Sgs-8* contained homology to the Mestril *et al.* motif. The right copy of the conserved sequence region upstream of *Sgs-7* contains a near-perfect match to the first portion of the Mestril *et al.* sequence; it is accompanied by an imperfect inverted repetition beginning one base-pair away, but the relative order (and therefore the location of the twofold symmetry axis) is reversed compared to the Mestril *et al.* sequence motif. Examination of the left copy, the one nearer to *Sgs-8*, reveals that the region homologous with the Mestril *et al.* motif has suffered two changes relative to the right copy nearer to *Sgs-7*. These are a single-base deletion at the proposed axis of symmetry and a single-base substitution in the imperfect inverse-repeat component six bases away from the deletion. The clustered changes might explain why the right copy elevates *Sgs-8—lacZ* gene function but that the left copy is dispensable for *Sgs-7—Adh* gene function.

Two new hypotheses are suggested by these results and are represented in Figure 11. The first hypothesis is that a bidirectional element resides in the *Sgs-7*, *Sgs-8* intergenic region to regulate tissue, stage, and quantity of expression of both glue protein genes. In this scheme, the left edge of the element is contained between the –92 bp deletion endpoint that inactivates the *Sgs-7—Adh* fusion gene and the –211 bp (or possibly the –139 bp) deletion endpoints that retain *Sgs-7—Adh* function. The deletion endpoint that partially inactivates the *Sgs-8—lacZ* fusion gene defines the right edge of the element. The second hypothesis, which is an elaboration of the first hypothesis, is that our experiments have identified two bidirectional elements within the

Sgs-7, *Sgs-8* intergenic region. One element, residing between –211 base-pairs (or possibly –139 base-pairs) and –92 base-pairs relative to the *Sgs-7* 5' end would determine the tissue and stage of bidirectional expression. When the *Sgs-7—Adh* fusion gene is deprived of this element, it is inert in both germline transformation and transient expression experiments. The second bidirectional element is a quantitative control element residing approximately 415 base-pairs upstream of the 5' end of *Sgs-8*. When this element is removed from the *Sgs-8—lacZ* fusion gene, only a low level of salivary gland expression is observed following germline transformation and that amount is below the level of detection in the transient expression assay. The two hypotheses are distinguished by the predictions they make concerning the effects of small interstitial deletions on the expression of the *Sgs-8—lacZ* fusion gene and on the expression of *Sgs-7—Adh* fusion gene. Three kinds of experiment will help to test the predictions. First, *Sgs-7—Adh* deletion molecules may be linked to *Sgs-8—lacZ* deletion molecules in a way that restores the divergently oriented transcription arrangement but with segments of the intergenic region removed. Such molecules could be tested simultaneously for the effects of the interstitial deletions on the expression of the *Sgs-8—lacZ* fusion gene and on expression of the *Sgs-7—Adh* fusion gene. Second, the putative *cis*-acting regulatory region may be tested for its capacity to confer the glue protein pattern of gene regulation on a heterologous promoter. In the third type of experiment, the putative *cis*-acting region may be tested for its capacity to restore normal quantity of expression on an *Sgs-3* gene construction lacking its own quantitative control sequences. Plasmid constructions of all three types exist (M.D. Garfinkel, unpublished experiments) and should prove useful in the further analysis of the *Sgs-7*, *Sgs-8* intergenic region.

(c) *Enhancements of the somatic transformation procedure*

Earlier experiments with the "somatic transformation" assay were performed by Martin *et al.* (1986) using the *D. melanogaster Adh* gene under the control of its own pair of promoters (Benyajati *et al.*, 1983). They made use of the same histochemical staining procedures mentioned earlier and observed mosaic patterns of staining consistent with the arrray of tissues that normally express *Adh.* The conclusions Martin *et al.* (1986) drew concerning the distribution of regulatory sequences were similar to those of Posakony *et al.* (1985) based on P factor-mediated gene transfer experiments. Shore and Guild (1987) reported transient expression results of promoter deletion experiments with the *Sgs-5* gene. They used an *Sgs-5* RNA-null variant host strain in their experiments, and each third instar larva that survived the microinjection procedure was subjected to RNA gel blot hybridization to detect salivary gland-specific expression of the injected *Sgs-5* genes. We believe that using histochemically assayable reporter genes reduces considerably the labor involved in the transient expression assay, possibly increases the sensitivity of the procedure since even a faint staining reaction within a single isolated cell of a salivary gland can be scored (unpublished observations of K. Vijay Raghavan, M. Roark, C. Mayeda, T. Todo, M.D. Garfinkel and E.M. Meyerowitz), and certainly renders the procedure more generally applicable.

(d) *Limitations of the somatic transformation procedure*

That supercoiled plasmid DNA could function following microinjection into *Drosophila* embryos was shown by the experiments of Spradling and Rubin (1982), in which a cloned P factor transposed from the microinjected plasmid molecules into the chromosomes of germline cells. Their pioneering work with the P transposable element

also showed that plasmid molecules bearing the *rosy*$^+$ gene could complement *ry*$^-$ mutations in the microinjection survivors independent of the occurrence of germline transposition events (Rubin and Spradling, 1982). Although it has not been demonstrated, one can reasonably infer that the basis of the eye color phenotype complementation in these experiments is the mosaic expression in somatic tissue of the microinjected *rosy*$^+$ gene. Indeed, *Drosophila* adults that are mosaic for the presence of a functional *ry*$^+$ gene by virtue of post-fertilization replicative repair of meiotic recombination heteroduplexes exhibit restoration of *ry*$^+$ eye color (Carpenter, 1982).

The speed of the transient expression procedure is obtained at the cost of quantitative data. Our results with both the *Sgs-7—Adh* fusion gene and with the *Sgs-8—lacZ* fusion gene illustrate this point. The risk of interpreting the frequencies of mosaic larval salivary glands as measures of the strengths of different promoter segments is shown by the results of the *Sgs-7—Adh* fusion gene with 236 base-pairs of 5' flanking sequence compared with only 211 base-pairs of 5' flanking sequence. Table 4 lists the transient expression results for such constructions: the –236 bp construction gave 29% of the tested glands exhibiting some expression, while the –211 bp construction gave only 5.4% mosaic salivary glands. Third instar larval salivary glands of the corresponding *Tf()GAX0.26* and *Tf()GAX0.24* germline transformant lines, when subjected to a quantitative soluble extract assay for ADH expression, gave no statistically significant difference in expression between the constructions (Table 5). And the possibility that the absence of β-galactosidase staining in somatically transformed salivary glands represents a threshold of detection is shown by the *Sgs-8—lacZ* germline transformation results. We tested by germline transformation only one *Sgs-8—lacZ* construction that was negative in the transient expression assay, and found it to be expressed to a degree greater than twentyfold reduced compared to

constructions with only slightly greater amounts of 5' flanking sequence. We must conclude that the detection of any ADH staining in microinjection survivors might represent wild-type or near wild-type quantity of expression, while the failure to detect β-gal staining in microinjection survivors does not exclude residual gene function.

(e) *Divergently transcribed gene pairs*

Here we will consider three kinds of divergently transcribed gene pairs, and what is known about the ways in which their *cis*-acting regulatory sequences are deployed. The first kind has both genes expressed in a single tissue at a single time during development. Our example is the *Sgs-7*, *Sgs-8* gene pair which, as we have seen, apparently depends on a single relatively small region that functions bidirectionally to specify tissue and stage of gene expression. The second kind has both genes expressed in two different tissues at a single time during development. Our example will be the *Yolk protein-1*, *Yolk protein-2* (*Yp-1*, *Yp-2*) gene pair. The third kind has both genes expressed within a single tissue but at different times during development. Our example for this will be the *Sgs-4*, *Pig-1* gene pair.

(i) *The Yolk protein-1, Yolk protein-2 (Yp-1, Yp-2) gene pair*

These genes are expressed in two tissues of the adult female: the ovarian follicle cells and the fat body (Brennan *et al.*, 1982). The *Yp-1*, *Yp-2* gene pair is divergently transcribed, with a 1225-base-pair intergenic spacer between the two 5' ends (Garabedian *et al.*, 1985). When the intergenic region is divided into two parts, one adjacent to a marked copy of the *Yp-1* gene, the other adjacent to a marked copy of the *Yp-2* gene, and each part tested for function following P element gene transfer, the *Yp-1* gene is expressed only in the fat body of adult females and the *Yp-2* gene is

expressed only in ovarian follicle cells of adult females (Garabedian *et al.*, 1985). The DNA sequence specifying expression in the adult female fat body functions in a bidirectional manner: a chimeric *Hsp70—lacZ* gene placed in varying relationships near a 126-bp segment of the 5' flanking sequences near *Yp-1* is expressed in female fat body following P element gene transfer (Garabedian *et al.*, 1986). Although the follicle cell specificity element has not been similarly localized and tested for function, it is reasonable to believe that it will function bidirectionally when positioned in proximity to a heterologous test gene. The *Sgs-7*, *Sgs-8* gene pair is a simplified example of the regulatory sequence arrangement seen in the *Yp-1*, *Yp-2* gene pair: expression in one tissue and one bidirectional regulatory element *versus* expression in two tissues and two bidirectional regulatory elements.

Yan *et al.* (1987), in their analysis of *Yp* gene sequences, noted that the Mestril *et al.* (1986) ecdysterone responsive element can be found in the female fat body tissue specificity element near the *Yp-1* gene (Garabedian *et al.*, 1986), and in the 5' flanking region of the *Yp-3* gene. The importance of the Mestril *et al.* sequence motif in the regulation of yolk protein gene expression is unclear for several reasons. First, the concentrations of ecdysterone in adult females and adult males are equal (Handler, 1982). Second, conditional mutations in the genetic regulatory hierarchy that governs somatic sex determination have conditional, reversible, effects on yolk protein gene expression (Belote *et al.*, 1985). Third, injection of ecdysterone into adult females induces transient increases in *Yp* gene expression in their fat body tissues and, when very high concentrations of hormone are injected, *de novo Yp* gene expression in the fat body of adult males (Postlethwait *et al.*, 1980; Bownes *et al.*, 1983). However, the *Yp-1—Adh* fusion gene tested by Shirras and Bownes (1987), which contains the

Mestril *et al.* motif, is not ecdysterone-inducible in the fat bodies of adult males under conditions that do induce the endogenous *Yp-1*, *Yp-2*, and *Yp-3* genes.

(ii) *The Sgs-4, Pig-1 gene pair*

The divergently transcribed *Sgs-4*, *Pig-1* gene pair possesses several interesting properties. First, the two genes' 5' ends are 840 base-pairs apart (Hofmann and Korge, 1987). Second, male larvae with a single X-chromosome accumulate the same amount of sgs-4 protein as female larvae with two X-chromosomes (Korge, 1975). Third, the disjunct timing of expression: the *Pig-1* RNA is present during larval life prior to appearance of *Sgs-4* RNA, and during the early part of the third instar *Pig-1* RNA can be demonstrated to be salivary gland-restricted (Hofmann and Korge, 1987; Chen *et al.*, 1987). Fourth, the two genes are located near the 5' end of an eighty-kilobase intervening sequence of the *dunce* transcription unit (Chen *et al.*, 1987).

Studies directed toward understanding the function of the *Sgs-4*, *Pig-1* intergenic region have concentrated on the regulation of *Sgs-4* expression. This is due partly to history (*Sgs-4* was discovered first) and partly to operational considerations, the large third instar larva being much more amenable to experimentation. *Sgs-4* regulatory sequences located in the 840-base-pair intergenic region were first identified by the analysis of naturally occurring variants in *Sgs-4* expression. A quantitative control region is defined by the hypomorphic *Sgs-4* allele present in the Japanese wild-type strain Hikone-R. This strain underproduces *Sgs-4* mRNA by around fiftyfold and is associated with the replacement of fifty-two base-pairs between positions –356 bp and –305 bp upstream of the 5' end by three base-pairs (Muskavitch and Hogness, 1982). Homology to the Mestril *et al.* (1986) ecdysterone responsive sequence motif appears within the region deleted from Hikone-R (noted by Hofmann and Korge,

1987). The *Sgs-4* allele in the Samarkand strain exhibits reduced levels of expression that are not subject to X-chromosome dosage compensation and is associated with a single nucleotide substitution at –344 bp (Hofmann and Korge, 1987), within the confines of the Hikone-R deletion, affecting the Mestril *et al.* homology. P factor transformation of *in vitro*-recombined fragments derived from the Oregon-R and the Samarkand alleles shows that a restoration of quantity of expression and of X-chromosome dosage compensation of *Sgs-4* depends upon which base-pair is present at the –344 bp location (Hofmann *et al.*, 1987).

The *Sgs-4* tissue and stage specificity element is defined by the Ber-1 allele, a null variant in which 92 base-pairs between positions –486 bp and –392 bp have been replaced by three base-pairs (Muskavitch and Hogness, 1982). An *Sgs-4* gene fragment possessing upstream sequences broken at –393 bp, and therefore lacking the region deleted from Ber-1 and further upstream sequences, fails to express following germline transformation (McNabb and Beckendorf, 1986). Neither the Ber-1 deletion nor the Hikone-R deletion has any effect on the expression of the *Pig-1* gene (P.H. Mathers, personal communication).

When the DNA sequences encompassing the regions deleted from the Ber-1 and Hikone-R alleles (–158 bp to –568 bp relative to the *Sgs-4* 5' end) are placed at the –380 bp position relative to the proximal promoter of the *Adh* gene, salivary gland expression of ADH activity is observed independent of the relative orientation of the *Sgs-4* fragment. This expression is not limited to the third instar period when *Sgs-4* is normally expressed; it is first observed in the second instar. Shermoen *et al.* (1987) attribute this to a combinatorial interaction between the *Adh* promoter segment (which is normally expressed in a number of second instar tissues) and the *Sgs-4* element (which

provides salivary gland specificity). An alternative view is that the DNA segment tested by Shermoen *et al.* (1987) contains two bidirectional *cis*-regulatory sequences, one that normally supports *Pig-1* expression and a second that normally supports *Sgs-4* expression, but the segment lacks sequences that serve to constrain the functions of the bidirectional elements to their appropriate transcription initiation sites. Supporting this proposal is an unpublished experiment mentioned in the discussion of Shermoen *et al.* (1987). Germline transformants that carry the sequences between –840 bp and +1 bp relative to the *Sgs-4* 5' end juxtaposed to the +2 position relative to the 5' end of the proximal *Adh* RNA, accumulate ADH activity only in the third instar salivary gland. Thus, in contrast to the *Sgs-7*, *Sgs-8* case and the *Yp-1*, *Yp-2* case where common patterns of tissue- and stage-specific expression depend upon *cis*-regulatory elements that function bidirectionally, the disjunct expression of the *Sgs-4*, *Pig-1* gene pair may depend on sequences that in some way serve to restrain the otherwise bidirectional function of *cis*-regulatory elements.

(f) *More on the Mestril* et al. *motif*

If we accept provisionally the possibility that the Mestril *et al.* sequence motif is, in fact, a DNA-binding site for the ecdysterone receptor (extrapolating from the vertebrate examples of steroid regulation of gene expression), sufficient for ecdysterone-regulated gene expression, the problem of how different tissues respond to changing ecdysterone concentrations in different ways would remain. This is because glue protein genes, yolk protein genes, and the small heat shock protein genes, all of which contain the Mestril *et al.* motif, are expressed in different tissues and at different times during development. A partial solution to this problem may lie in tissue-specific gene products derived from regulatory loci. The genomic DNA from at least part of the

overlapping complementation complex at 2B5, which contains the *l(1)npr-1*$^{+}$ function already implicated in such salivary gland functions as 68C glue protein gene transcription, has been cloned (Chao and Guild, 1986). Both radioactive cDNA hybridizations to cloned DNA gel blot filters and RNA gel blot filter hybridization gave evidence for long transcription units expressed in the salivary gland following ecdysterone induction (Chao and Guild, 1986). These long transcription units are utilized in tissue-specific ways since radioactive cDNA synthesized from hormone-induced salivary gland RNA and from hormone-induced imaginal disc RNA hybridized to unique regions of the chromosome walk (Chao and Guild, 1986). Similar complexity appears in the ecdysterone-induced early puffs at 74EF and 75B (K. Burtis, C.W. Jones, W. Segraves, C. Thummel and D.S. Hogness, unpublished), whose protein products have been suggested to be DNA-binding proteins.

(g) *Comparison with other glue protein genes*

(i) *Sgs-5*

Combining the deletion mapping data with the sequence determination of the *Sgs-5*^{CA2} RNA-null allele, Shore and Guild (1987) proposed that the *Sgs-5 cis*-acting regulatory elements occupy fewer than 109 base-pairs of 5' flanking sequence, and that one or more of the three nucleotide substitutions in the upstream region of the CA2 allele could account for its inactivity. Their proposal should be regarded as a provisional assignment of the *cis*-acting sequences specifying tissue and developmental stage of expression only, since the transient expression assay, as we have seen, is a qualitative one. As yet undiscovered quantitative control elements may lie upstream of this gene. Homology to the Mestril *et al.* (1986) sequence element can be seen between –96 bp and –118 bp upstream of the *Sgs-5* start point (M.D. Garfinkel, unpublished).

The homology is comprised of two seven-nucleotide regions that form a perfect inverted repetition, a 9-bp spacer separates the two halves of the repetition, and their relative order matches the Mestril *et al.* symmetry. The fact that the –109 bp deletion, which removes half of the inverted repeat, is capable of supporting *Sgs-5* expression is not inconsistent with the Mestril *et al.* motif playing a functional role, since the pBR322 sequences juxtaposed to the *Sgs-5* sequences at –109 bp partially reconstitute the inverted repetition (M.D. Garfinkel, unpublished).

(ii) *Sgs-4*

In addition to the complexities displayed by the *Sgs-4*, *Pig-1* intergenic region described above, quantitative control sequences for the *Sgs-4* gene appear to be distributed over more than 2.6 kb upstream. One quantitative control element is in the region deleted from the Hikone-R strain. A second quantitative control element is implied by the expression of *Sgs-4* constructions containing 840 base-pairs of 5' flanking sequence, which accumulate variable amounts of *Sgs-4* RNA averaging 50% of wild-type (McNabb and Beckendorf, 1986). Krumm *et al.* (1985) noted that P element transformants with 2.6 kb of 5' flanking sequence also fail to accumulate wild-type quantities of *Sgs-4* RNA.

(iii) *Sgs-3*

The available experimental evidence shows that tissue and stage specificity of *Sgs-3* expression depends on nucleotides located between –130 bp and +12 bp relative to the transcription initiation site, and that the quantitative control of *Sgs-3* expression is distributed among three elements located more than 130 bp upstream but less than 2760 base-pairs upstream. P.H. Mathers (personal communication) also noted homology to the Mestril *et al.* sequence motif between –90 bp and –76 bp relative to the *Sgs-3*

transcription site, but here, unlike the *Sgs-7, Sgs-8* gene pair, the two components of the inverted repetition are arranged in the same way as the Mestril *et al.* prototype.

The location of the tissue and stage specificity control element for *Sgs-3* was revealed by three kinds of construction tested by P element germline transformation. The first involved an *Sgs-3* allele coding for an 1120-nucleotide-long *Sgs-3* RNA, which was transferred into a strain that synthesizes an 800-nucleotide-long *Sgs-3* RNA. Quantification of salivary gland RNA gel blot filters revealed that the abundance of the 1120-nucleotide RNA was less than 10% that of the 800-nucleotide RNA (Vijay Raghavan *et al.*, 1986). The second involved a translational fusion joining *Sgs-3* to *E. coli lacZ*. Histochemical reaction of the salivary glands from germline transformants revealed faint blue X-gal staining (Vijay Raghavan *et al.*, 1986). A further limit on the tissue and stage specificity element was recently placed by analyzing germline transformants carrying the –130 bp to +12 bp segment of *Sgs-3* joined to a promoter-less *Adh* gene, which synthesize salivary gland-specific ADH enzyme activity (K. Vijay Raghavan, M. Roark, C. Mayeda and E.M. Meyerowitz, manuscript in preparation).

Most of the work on the upstream quantitative control elements of *Sgs-3* has involved the RNA length-variant alleles and quantification of RNA gel blot filters. Differences in the details of the methods used make it difficult to compare the results obtained by the two research groups studying the region. In spite of these differences, essentially wild-type *Sgs-3* mRNA abundance derived from test genes occurs in germline transformants of composite P elements that contain 68C DNA segments bearing the *Sgs-3* gene and the *Sgs-7* gene, for a total of 2.76 kb of 5' flanking sequence (Richards *et al.*, 1983; Bourouis and Richards, 1985; Crosby and

Meyerowitz, 1986). When the 68C segment is broken within the *Sgs-7* gene, leaving 2.27 kb of 5' flanking sequence, expression of *Sgs-3* is reduced slightly (Crosby and Meyerowitz, 1986). Consistent with the Crosby and Meyerowitz (1986) result is an experiment reported by Giangrande *et al.* (1987) that located an *Sgs-3* quantitative control element between –2.35 kb and –2.1 kb relative to the *Sgs-3* 5' end, within the *Sgs-7* gene. Giangrande *et al.* (1987) mapped a second quantitative control element to the 285-base-pair repeat element (Meyerowitz and Hogness, 1982) located to the right of the *Sgs-7* gene and therefore closer to *Sgs-3*.

Measurement of salivary gland ADH activity in germline transformants of the *Sgs-3—Adh* fusion gene with varying amounts of 5' flanking sequence provides evidence for an *Sgs-3* quantitative control element located between –983 bp and –130 bp (K. Vijay Raghavan, M. Roark, C. Mayeda and E.M. Meyerowitz, manuscript in preparation). Probably as a consequence of their methods of RNA preparation and gel blot quantification, Bourouis and Richards (1985) failed to detect *Sgs-3* RNA transcripts from constructions that had 983 bp or 130 bp of 5' flanking sequence. The *Sgs-3—Adh* fusion gene studies indicate another quantitative control element between 983 bp upstream and 2.76 kb upstream (K. Vijay Raghavan, M. Roark, C. Mayeda and E.M. Meyerowitz, manuscript in preparation), but unlike Richards' work has not subdivided the fragment further. Therefore, the sum of the available evidence is that three quantitative control elements are distributed in the 2.76 kb upstream of the *Sgs-3* gene; all three appear to function at least partially when inverted from their normal orientations with respect to *Sgs-3*, and they appear to act additively to regulate expression of this gene.

In conclusion, the *Drosophila* glue protein genes display a diversity of distributions of *cis*-acting regulatory sequences. The number and location of quantitative control elements can vary from one element fewer than 450 bp upstream of the transcription start site, as is the case of the *Sgs-8* region, to three such elements scattered over a 2.76 kilobase region upstream as in *Sgs-3*. Tissue and stage specificity control elements may reside between the closest quantitative control element and the transcription start point, as in the *Sgs-3* gene. The orientation independence, or bidirectional functionality, of upstream regulatory elements must be subject to different constraints in the different loci: *Sgs-7* and *Sgs-8* are divergently oriented and expressed simultaneously within a single tissue, while *Sgs-4* and its neighbor *Pig-1* are divergently oriented and expressed at disjunct times within a single tissue.

Further experiments are required to delimit the glue protein genes' *cis*-regulatory sequences with greater precision, to determine the mechanisms that constrain their functions along the chromosomes, and to identify the biochemical agents that recognize these sequences to accomplish coordinated gene expression.

M.D.G. would like to acknowledge the expert technical assistance of Carol A. Mayeda in the microinjection of *Drosophila* embryos, and of Martin F. Yanofsky in the sequence determination of *Bal*31 deletion endpoints. Peter H. Mathers critically commented on an early draft of the manuscript. We are grateful also to the following individuals: J.J. Bonner, V. Pirrotta, and G.M. Rubin for the timely provision of various cloned DNA molecules, in some cases prior to publication; J.W. Posakony, and B.T. Wakimoto for fly strains; and C. Hill and S. Horvath of the Microchemical Facility of the Division of Biology, California Institute of Technology, for the synthesis of the two deoxytetradecamers used in the construction of pGAZ-1. At the outset of this work, M.D.G. was supported by U.S. National Research Service Award 5 T32 GM07616, awarded by the National Institute of General Medical Sciences of the U.S. National Institutes of Health to the California Institute of Technology. E.M.M. is the recipient of grant 5 RO1 GM28075 from the National Institute of General Medical Sciences of the U.S. National Institutes of Health.

REFERENCES

Akam, M.E., Roberts, D.B., Richards, G.P and M. Ashburner (1978) *Cell* **13**: 215-225.

Ashburner, M. (1972) *Exptl. Cell Res.* **71**: 433-440.

Ashburner, M. (1973) *Develop. Biol.* **34**: 47-61.

Ashburner, M. (1974) *Develop. Biol.* **39**: 141-157.

Beckendorf, S.K. and F.C. Kafatos (1976) *Cell* **9**: 365-373.

Belote, J.M., Handler, A.M., Wolfner, M.F., Livak, K.J. and B.S. Baker (1985) *Cell* **40**: 339-348.

Belyaeva, E.S., Aizenon, M.G., Semeshin, V.F., Kiss, I.I., Koczka, K., Baritcheva, I.M., Gorelova, T.D. and I.F. Zhimulev (1980) *Chromosoma* **81**: 281-306.

Belyaeva, E.S., Vlassova, I.E., Biyasheva, Z.M., Kakpakov, V.T., Richards, G. and I.F. Zhimulev (1981) *Chromosoma* **84**: 207-219.

Benyajati, C., Spoerel, N., Haymerle, H. and M. Ashburner (1983) *Cell* **33**: 125-133.

Berendes, H. and M. Ashburner (1978) In *The Genetics and Biology of Drosophila, 2B* (Ashburner, M. and T.R.F. Wright, eds.), pp. 453-498, Academic Press, New York.

Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L. and H.W. Boyer (1977) *Gene* **2**: 95-113.

Bonner, J.J., Parks, C., Parker-Thornburg, J., Mortin, M.A. and H.R.B. Pelham (1984) *Cell* **37**: 979-991.

Bourouis, M. and G. Richards (1985) *Cell* **40**: 349-357.

Bownes, M., Blair, M., Kozma, R. and M. Dempster (1983) *J. Embryol. Exp. Morph.* **78**: 249-268.

Boyer, H.W. and D. Roullard-Dussoix (1969) *J. Mol. Biol.* **41**: 459-472.

Brennan, M.D., Weiner, A.J,. Goralski, T.J. and A.P. Mahowald (1982) *Develop. Biol.* **89**: 225-236.

Campos-Ortega, J.A. and V. Hartenstein (1985) *The Embryonic Development of Drosophila melanogaster*, Springer-Verlag, Berlin, Heidelberg, New York, Tokyo.

Carpenter, A.T.C. (1982) *Proc. Natl. Acad. Sci., USA* **79**: 5961-5965.

Chao, A.T. and G.M. Guild (1986) *EMBO J.* **5**: 143-150.

Chen, C-N., Malone, T., Beckendorf, S.K. and R.L. Davis (1987) *Nature (London)* **329**: 721-724.

Crosby, M.A. and E.M. Meyerowitz (1986) *Develop. Biol.* **118**: 593-607.

Crowley, T.E. and E.M. Meyerowitz (1984) *Develop. Biol.* **102**: 110-121.

Crowley, T.E., Bond, M.W. and E.M. Meyerowitz (1983) *Mol. Cell. Biol.* **3**: 623-634.

Crowley, T.E., Mathers, P.H. and E.M. Meyerowitz (1984) *Cell* **39**: 149-156.

Davis, R.W., Botstein, D. and J.R. Roth (1980) *Advanced Bacterial Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Dworniczak, B., Seidel, R. and O. Pongs (1983) *EMBO J.* **2**: 1323-1330.

Farnogli, J., Hyde, J. and W. Sofer (1987) *Biochem. Genet.* **25**: 327-333.

Fraenkel, G. and V.J. Brooks (1953) *Biol. Bull.* **105**: 442-449.

Garabedian, M.J., Hung, M-C. and P.C. Wensink (1985) *Proc. Natl. Acad. Sci., USA* **82**: 1396-1400.

Garabedian, M.J., Shepard, B.M. and P.C. Wensink (1986) *Cell* **45**: 859-867.

Garfinkel, M.D., Pruitt, R.E. and E.M. Meyerowitz (1983) *J. Mol. Biol.* **168**: 765-789.

Giangrande, A., Mettling, C. and G. Richards (1987) *EMBO J.* **6**: 3079-3084.

Goldberg, M.L. (1979) Ph.D. Dissertation, Stanford University, Stanford.

Gronemeyer, H. and O. Pongs (1980) *Proc. Natl. Acad. Sci., USA* **77**: 2108-2112.

Grunstein, M. and D.S. Hogness (1975) *Proc. Natl. Acad. Sci., U.S.A.* **72**: 3961-3965.

Guild, G. M. and E.M. Shore (1984) *J. Mol. Biol.* **179**: 289-314.

Handler, A.M. (1982) *Develop. Biol.* **93**: 73-82.

Hansson, L. and A. Lambertsson (1983) *Mol. Gen. Genet.* **192**: 395-401.

Hofmann, A. and G. Korge (1987) *Chromosoma* **96**: 1-7.

Hofmann, A., Keinhorst, A., Krumm, A. and G. Korge (1987) *Chromosoma* **96**: 8-17.

Hung, M-C. and P.C. Wensink (1984) *Nucl. Acids Res.* **12**: 1863-1874.

Jefferson, R.A., Kavanagh, T.A. and M.W. Bevan (1987) *EMBO J.* **6**: 3901-3907.

Karess, R.E. and G.M. Rubin (1984) *Cell* **38**: 135-146.

Kiss, I., Bencze, G., Fekete, E., Fodor, A., Gausz, J., Maroy, P., Szabad, J. and J. Szidonya (1976) *Theoret. Appl. Genet.* **48**: 217-226.

Kiss, I., Szabad, J. and J. Major (1978) *Molec. Gen. Genet.* **164**: 77-83.

Korge, G. (1975) *Proc. Natl. Acad. Sci., U.S.A.* **72**: 4550-4554.

Korge, G. (1977) *Develop. Biol.* **58**: 339-355.

Korge, G. (1987) In *Results and Problems in Cell Differentiation*, vol. 14, (W. Hennig, ed.), pp. 27-58, Springer-Verlag, Berlin Heidelberg New York Tokyo.

Krumm, A., Roth, G.E. and G. Korge (1985) *Proc. Natl. Acad. Sci., USA* **82**: 5055-5059.

Lane, N.J., Carter, Y.R. and M. Ashburner (1972) *Wilhelm Roux' Archiv.* **169**: 216-238.

Lee, C.Y. (1982) *Methods Enzymol.* **89**: 445-450.

Lindsley, D.L. and E.H. Grell (1968) *The genetic variations of Drosophila melanogaster*, Carnegie Institution of Washington, Baltimore, publication 627.

Lindsley, D. and G. Zimm (1985) In *Drosophila Information Service* 62.

Lindsley, D. and G. Zimm (1986) In *Drosophila Information Service* 64.

Lindsley, D. and G. Zimm (1987) In *Drosophila Information Service* 65.

Martin, P., Martin, A., Osmani, A. and W. Sofer (1986) *Develop. Biol.* **117**: 574-580.

Maxam, A.M and W. Gilbert (1977) *Proc. Natl. Acad. Sci., U.S.A.* **74**: 560-564.

Maxam, A.M and W. Gilbert (1980) *Methods Enzymol.* **65**: 499-560.

McNabb, S.L. and S.K. Beckendorf (1986) *EMBO J.* **5**: 2331-2340.

Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and M.R. Green (1984) *Nucl. Acids Res.* **12**: 7035-7056.

Mestril, R., Schiller, P., Amin, J., Klapper, H., Ananthan, J. and R. Voellmy (1986) *EMBO J.* 5: 1667-1673.

Meyerowitz, E.M. and D.S. Hogness (1982) *Cell* **28**: 165-176.

Miller, J.H. (1972) *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Muskavitch, M.A.T. and D.S. Hogness (1980) *Proc. Natl. Acad. Sci., USA* **77**: 7362-7366.

Muskavitch, M.A.T. and D.S. Hogness (1982) *Cell* **29**: 1041-1051.

Nüsslein-Volhard, C., Wieschaus, E. and H. Kluding (1984) *Wilhelm Roux' Archiv.* **193**: 267-282.

Posakony, J.W., Fischer, J.A. and T. Maniatis (1985) *Cold Spring Harb. Symp. Quant. Biol.* **50**: 515-520.

Postlethwait, J.H., Bownes, M. and T. Jowett (1980) *Develop. Biol.* **79**: 1205-1216.

Richards, G., Cassab, A., Bourouis, M., Jarry, B. and C. Dissous (1983) *EMBO J.* **2**: 2137-2141.

Rubin, G.M. and A.C. Spradling (1982) *Science* **218**: 348-353.

Rubin, G.M. and A.C. Spradling (1983) *Nucl. Acids Res.* **11**: 6341-6351.

Sanger, F., Nicklen, S. and A.R. Coulson (1977) *Proc. Natl. Acad. Sci., USA* **74**: 5463-5467.

Sarmiento, L.A. and H.K. Mitchell (1982) *Develop. Genet.* **3**: 255-272.

Savakis, C., Ashburner, M. and J.H. Willis (1986) *Develop. Biol.* **114**: 194-207.

Shermoen, A.W., Jongens, J., Barnett, S.W., Flynn, K. and S.K. Beckendorf (1987) *EMBO J.* **6**: 207-214.

Shirras, A.D. and M. Bownes (1987) *Mol. Gen. Genet.* **210**: 153-155.

Shore, E.M. and G.M. Guild (1987) *Genes & Develop.* **1**: 829-839.

Singh, M.B. and R.B. Knox (1984) *Histochem. J.* **16**: 1273-1296.

Sofer, W. and H. Ursprung (1968) *J. Biol. Chem.* **243**: 3110-3115.

Southern, E.M. (1975) *J. Mol. Biol.* **98**: 503-517.

Spradling, A.C. and G.M. Rubin (1982) *Science* **218**: 341-347.

Spradling, A.C. and G.M. Rubin (1983) *Cell* **34**: 47-57.

Ursprung, H., Sofer, W.H. and N. Burroughs (1970) *Wilhelm Roux' Archiv.* **164**: 201-208.

Velissariou, V. and M. Ashburner (1980) *Chromosoma* **77**: 13-27.

Velissariou, V. and M. Ashburner (1981) *Chromosoma* **84**: 173-185.

Vieria, J. and J. Messing (1987) *Methods Enzymol.* in the press.

Vijay Raghavan, K., Crosby, M.A., Mathers, P.H. and E.M. Meyerowitz (1986) *EMBO J.* **5**: 3321-3326.

Wallenfels, K. (1962) *Methods Enzymol.* **5**: 212-219.

Yan, Y.L., Kunert, C.J. and J.H. Postlethwait (1987) *Nucl. Acids Res.* **15**: 67-85.

Yanisch-Perron, C., Vieria, J. and J. Messing (1985) *Gene* **33**: 103-119.

Zhimulev, I. and N.N. Kolesnikov (1975) *Wilhelm Roux' Archiv.* **178**: 15-28.

Table 1

Summary of the Tf()GLAX1.0 *insertion events*

Transformation Event	Parent	Chromosome	Lethal?	Identical to Other Isolates?
Tf(1)GLAX1.0-1	no. 7	X	No	No
Tf(2)GLAX1.0-2	no. 7	2	No	No
Tf(3)GLAX1.0-3	no. 7	3	No	Yes, to -6
Tf(3)GLAX1.0-4	no. 7	3	Yes	No
Tf(2)GLAX1.0-5	no. 7	2	Yes	No
Tf(3)GLAX1.0-6	no. 7	3	No	Yes, to -3
Tf(3)GLAX1.0-7	no. 48	3	No	Yes, to -8
Tf(3)GLAX1.0-8	no. 48	3	Yes	Yes, to -7

Table 2

Third instar larval salivary gland enzyme activity measurements of Tf()GLAX1.0 *strains*

(A) Measurement of alcohol dehydrogenase activity

Strain	Individual Measurements[1]	Mean ± S.D.
Tf(1)GLAX1.0-1	30.5, 36.5, 39.5, 17.1, 12.6, 14.6	25.1 ± 11.8
Tf(2)GLAX1.0-2	33.2, 28.9, 23.4, 30.4, 27.9, 29.7	28.9 ± 3.2
Tf(3)GLAX1.0-3	18.0, 44.0, 42.0, 24.9, 28.1, 27.0	30.7 ± 10.2
Tf(3)GLAX1.0-7	60.1, 39.3, 40.1, 50.4, 39.9, 34.2	44.0 ± 9.5

[1]Numbers are Units ADH activity per animal-equivalent of extract. Unit definition of Sofer and Ursprung (1968).

(b) Measurement of β-galactosidase activity

Strain	Individual Measurements[2]	Mean ± S.D.
Tf(1)GLAX1.0-1	156.1, 28.0, 176.5	102.2 ± 75
Tf(2)GLAX1.0-2	279.8, 274.0, 337.0	296.9 ± 35
Tf(3)GLAX1.0-3	418.9, 469.8, 499.1	462.6 ± 40.5
Tf(3)GLAX1.0-7	644.9, 395.7, 278.7	439.7 ± 187

[2]Numbers are pmol 4MU released per hour per animal-equivalent of extract.

Table 3

Transient expression of Sgs-8—lacZ *and* Sgs-7—Adh *genes in the plasmid pGAZ-1*

Gene Tested	5' Flanking Sequence	(Animals) Lobes Tested	Lobes Positive
Sgs-7—Adh	721 bp	(80) 160	27[1]
Sgs-8—lacZ	499 bp	(33) 66	24

[1]Number of animals that stained, out of 80 animals tested.

Table 4

Transient expression of Sgs-7—Adh *in various* Bal*31 deletion plasmids*

Construction	5' Flanking Sequence	(Animals) Lobes Tested	Lobes Positive
pDm9718	333 bp	(11) 19	7
pDm9730 (A)[1]	236 bp	(8) 16	4
pDm9730 (B)	236 bp	(42) 84	25
pDm9721 (A)	211 bp	(20) 35	2
pDm9721 (B)	211 bp	(56) 112	6
pDm9719	139 bp	(6) 12	1
pDm9729 (A)	92 bp	(25) 50	0
pDm9729 (B)	92 bp	(39) 78	0

[1]Independent experiments are recorded sequentially and are distinguished by letters A and B.

Table 5

Salivary gland alcohol dehydrogenase measurements of germline transformants of Sgs-7—Adh *promoter deletion derivatives*

Strain	Amount upstream	Individual Measurements[1]	Mean ± S.D.
Tf(2)GAX0.26-1	236 bp	25.9, 31.7, 24.0	27.2 ± 4.0
Tf(3)GAX0.26-2	236 bp	42.3, 43.0, 35.5	40.3 ± 4.1
Tf(2)GAX0.26-3	236 bp	16.2, 49.5, 19.9	28.5 ± 18.3
Tf(2)GAX0.24-1	211 bp	46.4, 28.0, 31.5	35.3 ± 9.8
Tf(3)GAX0.24-2	211 bp	39.0, 22.4, 27.3	29.6 ± 8.5
Tf(3)GAX0.24-3[2]	211 bp	62.5, 54.5, 48.6	55.2 ± 7.0
Tf(3)GAX0.24-4	211 bp	51.8, 45.7, 50.8	49.4 ± 3.3
Tf(2)GAX0.24-5	211 bp	37.2, 52.7, 53.8	47.9 ± 9.3
Tf(2)GAX0.12-1	92 bp	–0.2, 0.3, 0.056	0.05 ± 0.25
Tf(2)GAX0.12-2	92 bp	–0.28, –0.34, 0.0035	–0.21 ± 0.18
Tf(3)GAX0.12-3	92 bp	–0.62, 0.28, 0.0048	–0.11 ± 0.46
Tf(2)GAX0.12-4[3]	92 bp	1.42, 0.36, –0.0073	0.59 ± 0.74
Tf(2)GAX0.12-5	92 bp	–0.88, –0.02, –0.14	–0.35 ± 0.47

[1]Numbers are Units ADH activity per animal-equivalent of extract. Unit definition of Sofer and Ursprung (1968).

[2]Adult fly DNA gel blot filter hybridization shows a second insertion event segregating.

[3]Stock contaminated by *Tf(2)GAX0.12-4/In(2LR)CyO* animals. Balancer-borne *Adh*$^+$ allele expression in a fat body adhering to a salivary gland probably accounts for the 1.42 U measurement.

Table 6

Transient expression of Sgs-8—lacZ *in various* Bal*31 deletion plasmids*

Construction	5' Flanking Sequence	(Animals) Lobes Tested	Lobes Positive
pDm9800E	475 bp	(74) 148	38
pDm9812	432 bp	(25) 50	18
pDm9803	415 bp	(22) 44	0
pDm9802	298 bp	(30) 60	0
pDm9804	226 bp	(57) 112	0

Table 7

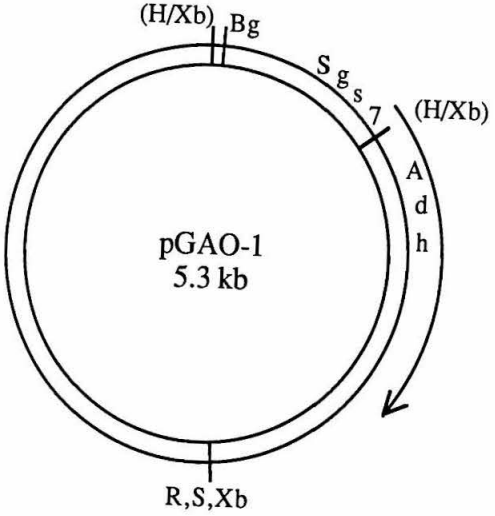
Salivary gland β-galactosidase measurements of germline transformants of Sgs-8—lacZ *promoter deletion derivatives*

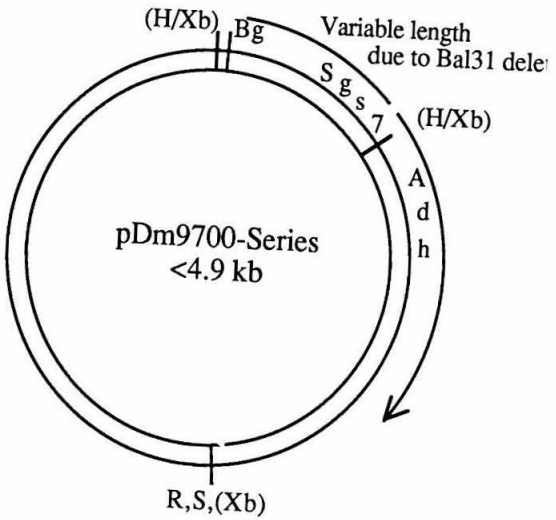
Strain	Amount upstream	Individual Measurements[1]	Mean ± S.D.
Tf(3)GLX0.68-1	432 bp	316.0, 228.7, 258.7	267.8 ± 44.4
Tf(2)GLX0.68-2	432 bp	155.0, 125.7, 109.9	130.2 ± 22.9
Tf(2)GLX0.68-3	432 bp	264.4, 328.5, 298.4	298.5 ± 30.0
Tf(2)GLX0.68-4	432 bp	822.4, 775.4, 1042.0	879.9 ± 142
Tf(3)GLX0.68-5	432 bp	230.7, 291.6, 209.8	244.0 ± 42.5
Tf(3)GLX0.66-1	415bp	15.1, 24.7, 35.5	25.1 ± 10.2
Tf(2)GLX0.66-2	415bp	10.2, 10.3, 15.5	12.0 ± 3.0
Tf(2)GLX0.66-3	415bp	43.8, 25.5, 33.1	34.1 ± 9.2
Tf(3)GLX0.66-4	415bp	27.5, 27.7, 25.5	26.9 ± 1.2
Tf(3)GLX0.66-5	415bp	25.8, 23.3, 26.7	25.3 ± 1.7

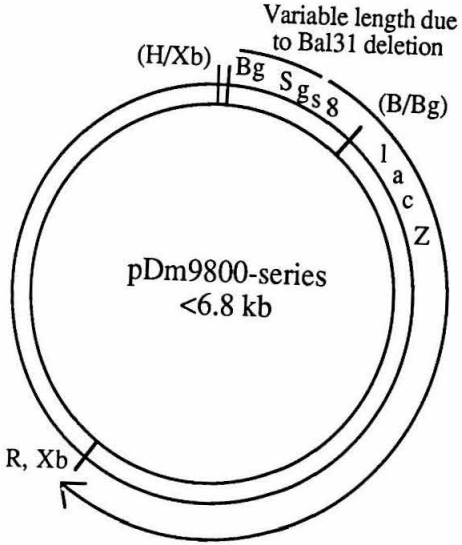
[1]Numbers are pmol 4MU released per hour per animal-equivalent of extract.

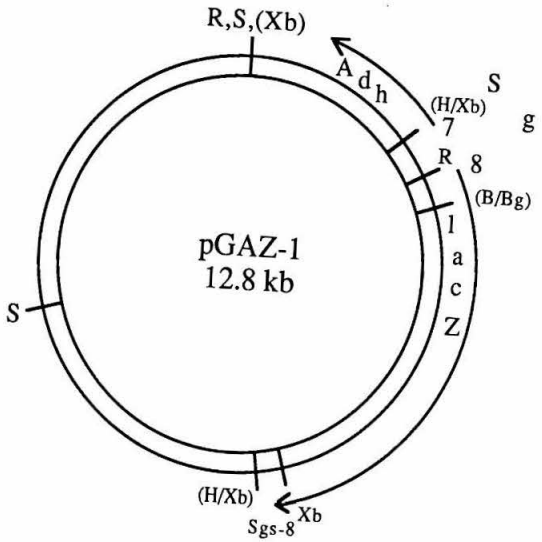
Figure 1. Plasmid maps. The first sheet contains circular maps of DNA molecules used in the somatic transformation experiments and as precursors for germline transformation vectors: pGAO-1 is the original *Sgs-7—Adh* gene fusion plasmid. The symbols (H/Xb) represent the junctions between the 824-bp *Xba*I fragment of 68C DNA that contains the 5' end of *Sgs-7* and the *Hin*dIII site of nDm9035. Other restriction sites are indicated by: Bg, *Bgl*II; R, *Eco*RI; S, *Sal*I; and Xb, *Xba*I. Transcription orientation of the *Sgs-7—Adh* fusion gene is shown by the arc a with clockwise-pointing arrowhead. pDm9700-series is a schematic diagram of the *Bal*31 deletion plasmids that have progressively shorter segments of the 5' flanking sequence of the *Sgs-7—Adh* fusion gene. Restriction site symbols are the same as for the pGAO-1 map with the addition of (Xb), which represents a recognition site for *Xba*I destroyed by the use of oligonucleotides. pDm9800-series is a schematic diagram of the *Bal*31 deletion plasmids that have progressively shorter segments of the 5' flanking sequence of the *Sgs-8—lacZ* fusion gene (the *Eco*RI site located 80 bp upstream of the *Sgs-8* 5' end is omitted for clarity). Transcription orientation of the *Sgs-8—lacZ* fusion gene is shown by the arc with a clockwise-pointing arrowhead. Restriction site symbols are the same as for the pGAO-1 map with the addition of (B/Bg), which represents the junction between the *Bgl*II site of the *Sgs-8* protein-coding region and the *Bam*HI site of the *lacZ* gene. The plasmid pGAZ-1 contains the *Sgs-7—Adh* fusion gene transcribed counterclockwise, the *Sgs-8—lacZ* fusion gene transcribed clockwise, and the *Sgs-7*, *Sgs-8* intergenic region reassembled. The 3' untranslated region of the sgs-8 mRNA is joined to the *Sgs-8—lacZ* fusion gene transcript by the presence of the 0.27 kb restriction fragment near the 6 o'clock position of the map. Note that the circular plasmid maps are drawn to different scales, with their contour lengths indicated. On the continuation sheet, the composite P elements contained in the pGLAX1.0 plasmid, the

three pGAX series plasmids, and the two pGLX series plasmids are drawn in linear form to the same scale. The origins of the various segments are shown, as are the transcription directions of the different genes. The overall length of the P element from pGLAX1.0 is 17.9 kb, while the pGAX P elements are nearer to 10.8 kb in length and the pGLX P elements are nearer to 12.7 kb. Restriction sites are indicated as for the circular maps, but with the addition of H for *Hin*dIII.

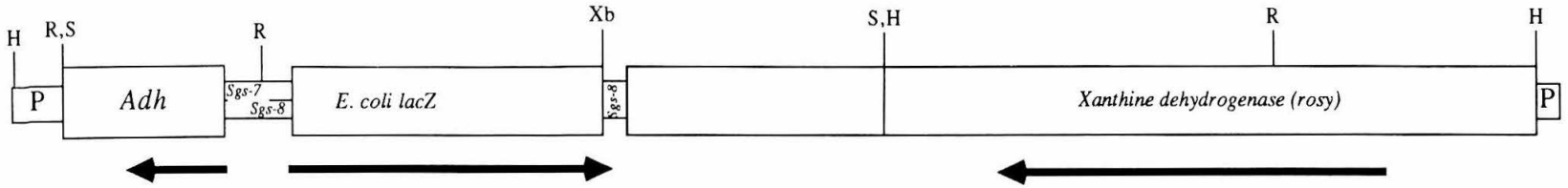

(H/Xb)
Bg
S g s 7
(H/Xb)
A d h
pGAO-1
5.3 kb
R,S,Xb


(H/Xb)
Bg
Variable length
due to Bal31 dele
S g s 7
(H/Xb)
A d h
pDm9700-Series
<4.9 kb
R,S,(Xb)


Variable length due
to Bal31 deletion
(H/Xb)
Bg S g s 8
(B/Bg)
l a c Z
pDm9800-series
<6.8 kb
R, Xb


R,S,(Xb)
A d h
(H/Xb)
S
g
7
R
8
(B/Bg)
l a c Z
pGAZ-1
12.8 kb
S
(H/Xb)
Sgs-8
Xb

pGLAX1.0



pGAX Series

pGLX Series

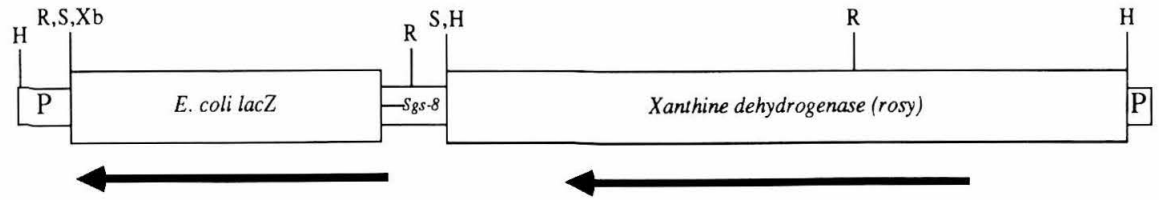


Figure 2. DNA sequences of the natural *Sgs-7* gene, the natural *Adh* gene and the *Sgs-7—Adh* fusion construction joint. The first forty-eight DNA nucleotides equivalent to *Sgs-7* mRNA are shown on the top line. The *Xba*I restriction site present in the DNA and the translation initiation codon are identified by overlines and labels (Garfinkel *et al.*, 1983). On the second line are shown the first forty nucleotides of the 5' untranslated region of *Adh*, relative to the proximal promoter; the translation initiation codon is not shown (Benyajati *et al.*, 1983). On the third line is the sequence of the junction between *Sgs-7* and the *Bal*31-deletion derivative of the *Adh* gene obtained from Bonner *et al.* (1984). The exact junction point and the origin of its component nucleotides are indicated. Note that the predicted RNA transcript of the chimeric gene would have the first twenty-five nucleotides of the natural sgs-7 messenger RNA and five nucleotides of synthetic-DNA origin replacing the first twelve nucleotides of the natural *Adh* proximal messenger RNA.

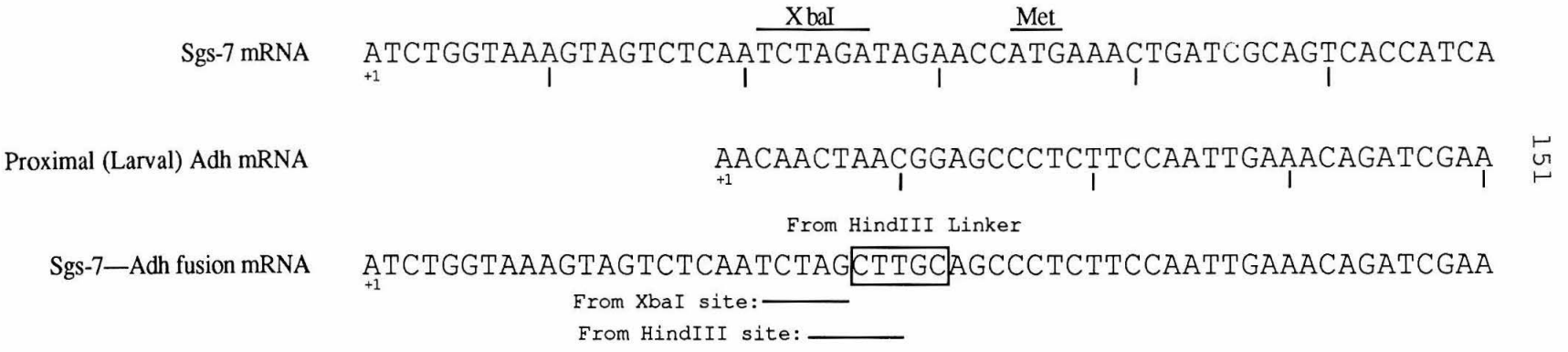

XbaI
Met
Sgs-7 mRNA
ATCTGGTAAAGTAGTCTCAATCTAGATAGAACCATGAAACTGATCGCAGTCACCATCA
+1
Proximal (Larval) Adh mRNA
AACAACTAACGGAGCCCTCTTCCAATTGAAACAGATCGAA
+1
From HindIII Linker
Sgs-7—Adh fusion mRNA
ATCTGGTAAAGTAGTCTCAATCTAGCTTGCAGCCCTCTTCCAATTGAAACAGATCGAA
+1
From XbaI site:
From HindIII site:

Figure 3. Histochemical staining of third-instar larvae. Mature third instar larvae of the *Tf(1)GLAX1.0-1* strain were dissected and the salivary glands removed along with some of the adjacent tissue. (A) One salivary gland lobe and its accompanying tissue were fixed in glutaraldehyde, washed extensively, and then transferred to the staining reagent for alcohol dehydrogenase activity, which is confined to the salivary gland. (B) Two salivary gland lobes and adhering fat body tissue were transferred to the staining reagent for β-galactosidase activity; such enzyme activity is confined to the salivary gland. As controls for the histochemical reactions, mature third instar larvae of the non-transformed host strain, Adh^{fn6} *cn*; ry^{502}, were dissected and the salivary glands were removed along with some of the adjacent tissue. (C) One salivary gland lobe and adhering tissue were transferred to the staining reagent for alcohol dehydrogenase activity; no staining is observed in the salivary gland. (D) Another salivary gland lobe and its accompanying tissue were prepared for β-galactosidase activity staining; no enzyme activity is detected.

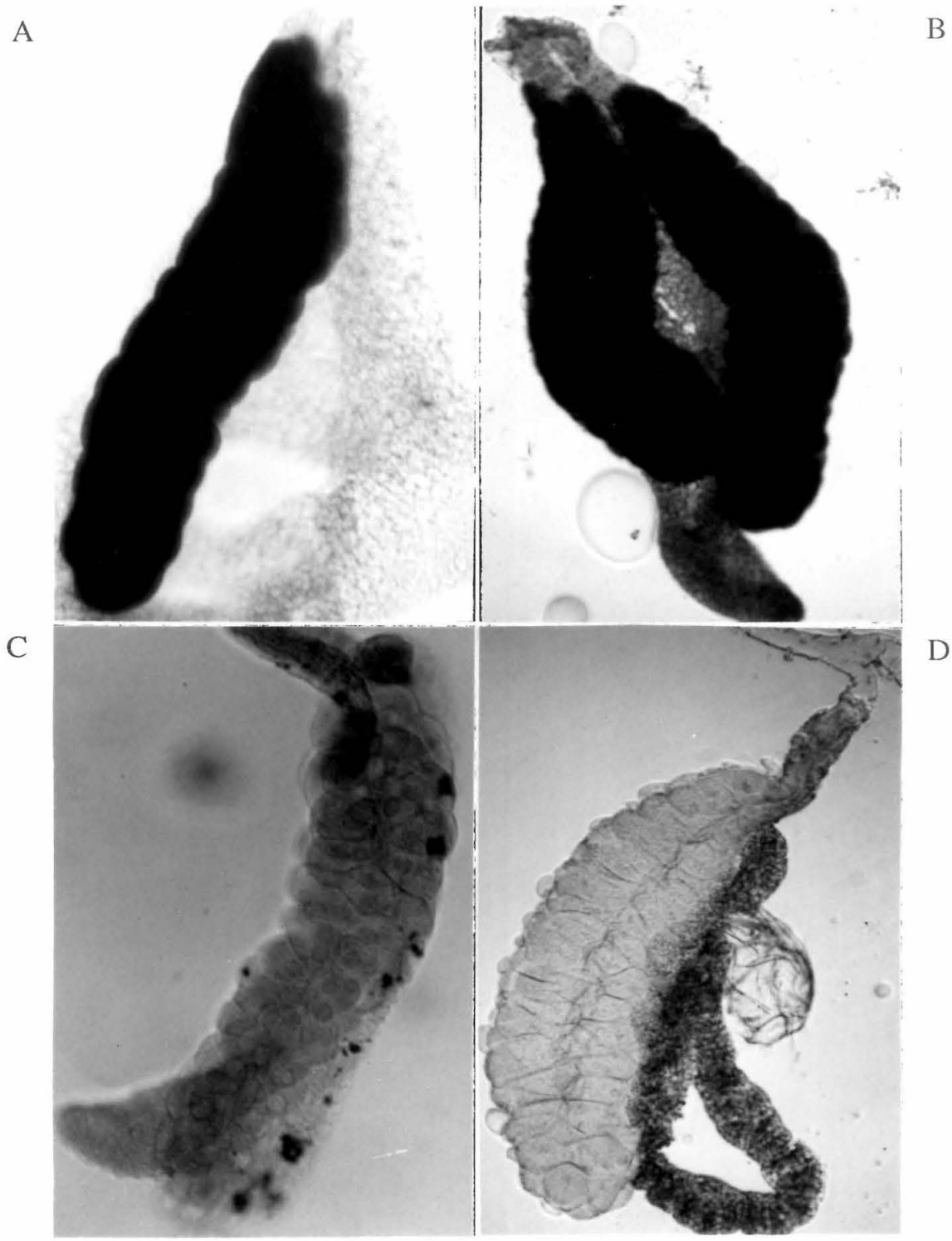
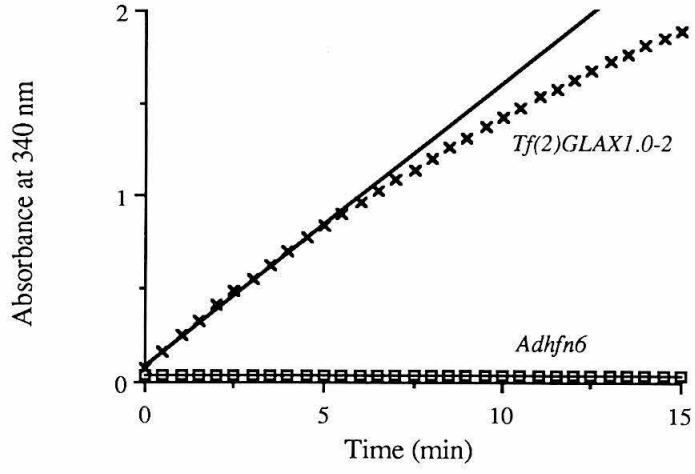
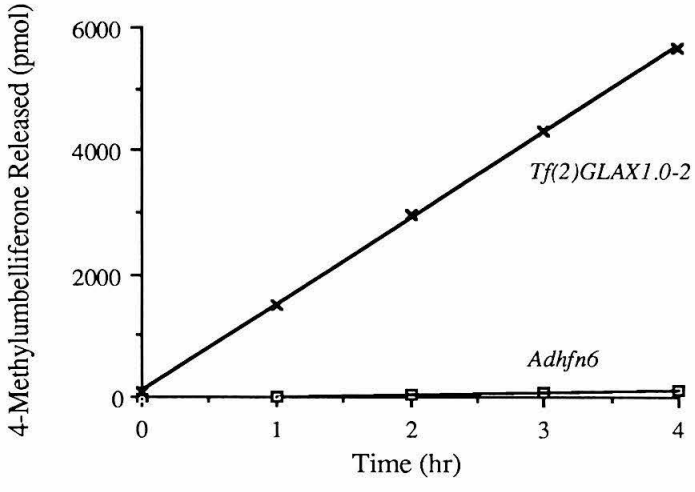

A
B
C
D

Figure 4. Soluble extract measurements of salivary gland enzyme activities. Fifteen third instar larvae from each of the strains *Tf(2)GLAX1.0-2* and *Adh*^{fn6} *cn*; *ry*^{502} were dissected and extracts prepared from the salivary glands. In panel A, five animal-equivalents, 50 µl, of each salivary gland extract were assayed for ADH enzyme activity. For the *Tf(2)GLAX1.0-2* strain, the linear regression line shown was calculated with the first eleven data points—zero through five minutes. The parameters of the calculated line are: Y-intercept, 0.0917; slope, 0.1522; correlation coefficient, 1.00. The ADH Units per animal-equivalent are given by (0.1522 divided by 5, then multiplied by 1000); the data shown here correspond to 30.4 Units per animal-equivalent. For the *Adh*^{fn6} *cn*; *ry*^{502} host strain, the linear regression line parameters were calculated from the full set of thirty-one data points: Y-intercept, 0.037; slope, 2.351 x10-4; correlation coefficient, 0.88. The *Adh*^{fn6} *cn*; *ry*^{502} background reaction corresponds to 0.05 Units per animal-equivalent. In panel B, five animal-equivalents of salivary gland extract, 50 µl, were assayed for β-galactosidase enzyme activity. For the *Tf(2)GLAX1.0-2* strain, the linear regression line shown was calculated with the five data points. The parameters of the calculated line are: Y-intercept, 105.98; slope, 1398.84; correlation coefficient, 1.00. The β-galactosidase Units per animal-equivalent are given by (1398.84, divided by 5); the data shown here correspond to 279 Units per animal-equivalent. For the *Adh*^{fn6} *cn*; *ry*^{502} host strain, the linear regression line parameters calculated are: Y-intercept, –24.4; slope, 33.38; correlation coefficient, 1.00. The *Adh*^{fn6} *cn*; *ry*^{502} background reaction corresponds to 6.67 Units per animal-equivalent. In panel C, each datapoint represents one β-galactosidase measurement and one ADH measurement made from an individual *Tf()GLAX1.0* salivary gland extract. Positive correlation (correlation coefficient 0.79) is seen.

A


Absorbance at 340 nm
2
1
0
0
5
10
15
Time (min)
Tf(2)GLAX1.0-2
Adhfn6

B


4-Methylumbelliferone Released (pmol)
6000
4000
2000
0
0
1
2
3
4
Time (hr)
Tf(2)GLAX1.0-2
Adhfn6

C

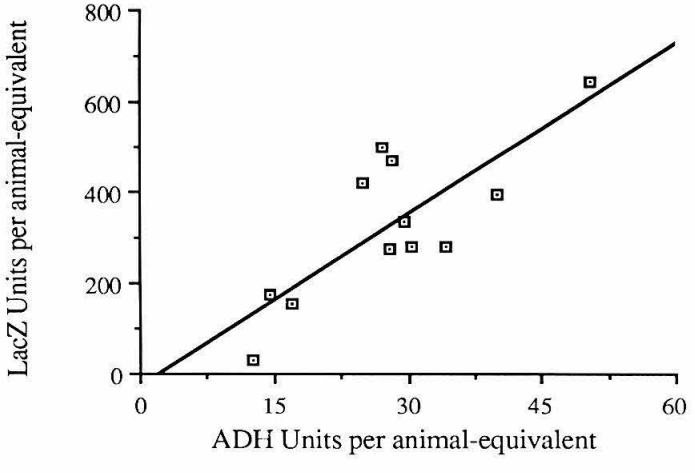

LacZ Units per animal-equivalent
800
600
400
200
0
0
15
30
45
60
ADH Units per animal-equivalent

Figure 5. Third instar larval salivary gland RNA gel blot hybridization. For each strain, ten third instar larvae were dissected and the salivary glands removed. RNA prepared using phenol and chloroform was resuspended in 80 μl of the HCHO- and $HCONH_2$-containing denaturation cocktail described by Crosby and Meyerowitz (1986). One-and-a-half animal-equivalents of each RNA sample were placed into separate wells of a 1.5% agarose gel containing ≈6% (v/v) formaldehyde. After electrophoresis, the RNA transferred to a nitrocellulose sheet using 20X SSPE (Davis *et al.*, 1980) as the transfer buffer. RNA gel blot filter sections were baked, then prehybridized and hybridized at 43°C using ^{32}P-nick-translation-labelled DNA probes. Filters were washed and autoradiographed. In each panel, lane 1 is salivary gland RNA from *Tf(1)GLAX1.0-1*, and lane 2 is salivary gland RNA from Adh^{fn6} *cn*; ry^{502}. (A) Hybridization with kEc001, a plasmid clone of the β-galactosidase-coding region of *E. coli lacZ* (E.M. Meyerowitz, unpublished experiment); (B) Hybridization with nDm9035, a plasmid subclone of *D. melanogaster Adh* genomic DNA including the greater part of the coding region; (C) Hybridization with sDm9039, a genomic DNA subclone containing the *D. melanogaster Sgs-7* gene as a *Pst*I-*Eco*RI fragment (M.D. Garfinkel, unpublished experiment); (D) Hybridization with sDm9040, a genomic DNA subclone containing the *D. melanogaster Sgs-8* gene as a *Pst*I-*Eco*RI fragment (M.D. Garfinkel, unpublished experiment).

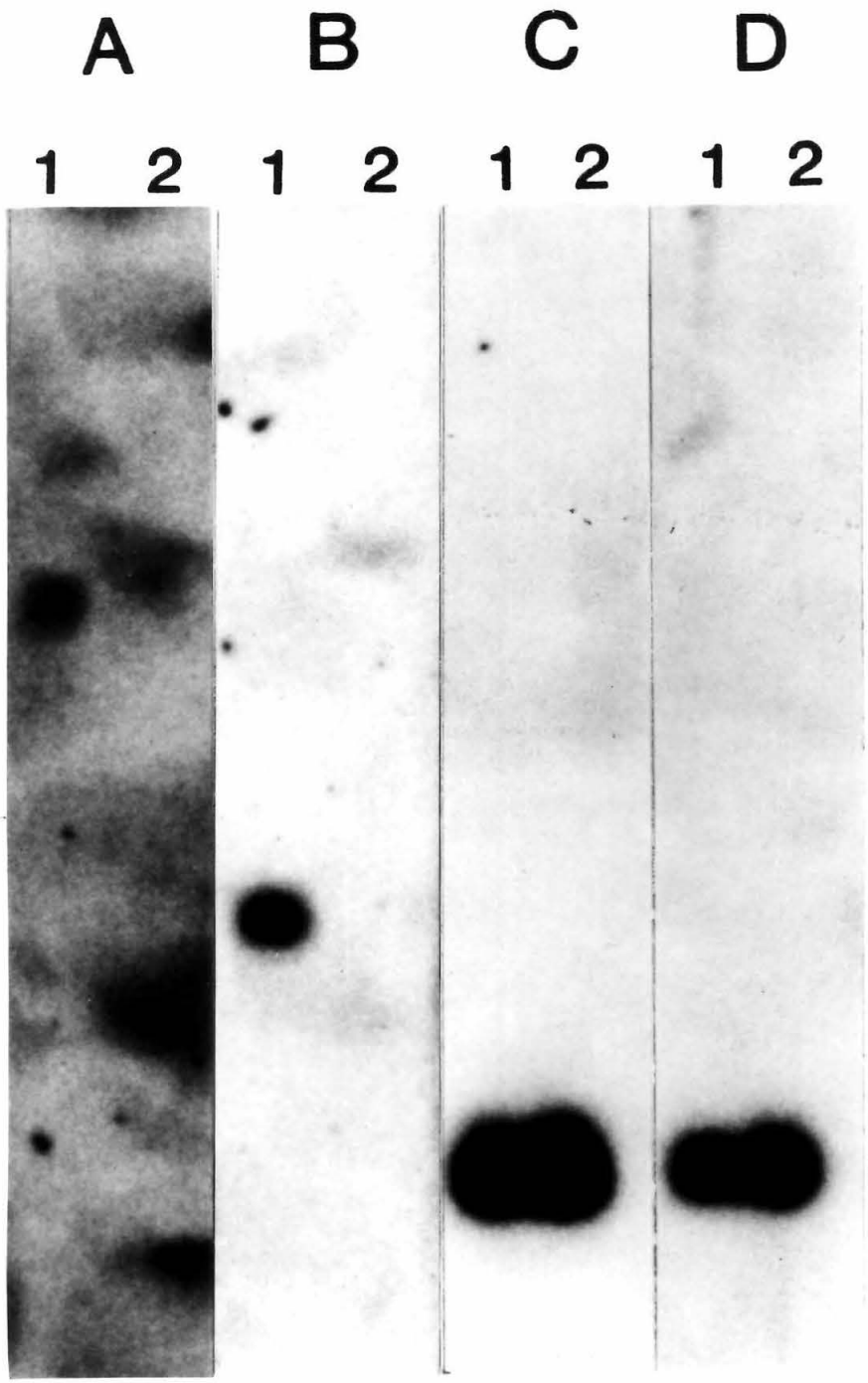
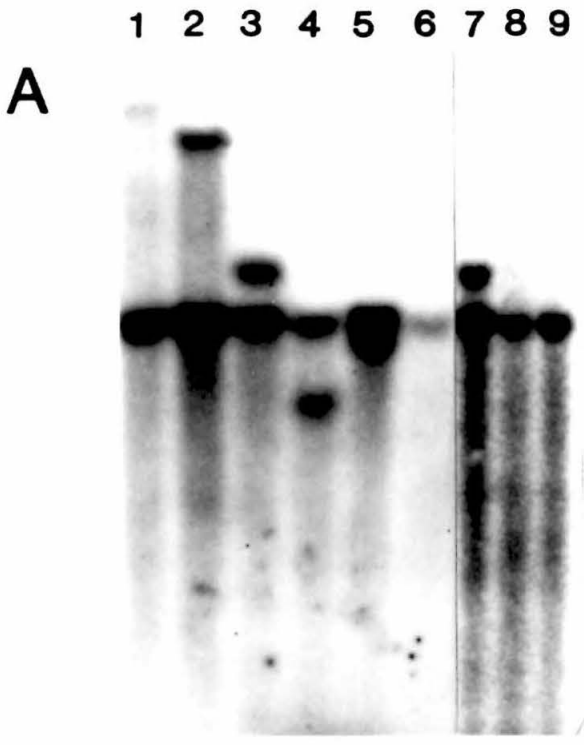

A
B
C
D
1 2 1 2 1 2 1 2

Figure 6. Adult fly DNA gel blot hybridization. Adult fly DNA was digested with either *Bam*HI (Panel A) or *Sac*I (Panel B) and electrophoretically separated by size in 0.6% agarose gels. The gel blot filters subsequently prepared were hybridized with ^{32}P-labelled aDm9030 (M.D. Garfinkel, unpublished experiment). This probe detects a 5.2 kb *Bam*HI fragment common to all the fly strains and two *Sac*I fragments of length 4.0 kb and 1.5 kb common to all the fly strains, which represent the centromere-proximal portion of the *rosy* locus at 87D14. Each transformant strain contains a single additional hybridizing fragment, which arises from integration of the P[*GLAX1.0*] element. Lane 1 is *Tf(1)GLAX1.0-1*; lane 2 is *Tf(2)GLAX1.0-2*; lane 3 is *Tf(3)GLAX1.0-3*; lane 4 is *Tf(3)GLAX1.0-4/TM3*; lane 5 is *Tf(2)GLAX1.0-5/CyO*; lane 6 is the non-transformed host strain Adh^{fn6} cn; ry^{502}; lane 7 is T*f(3)GLAX1.0-6*; lane 8 is *Tf(3)GLAX1.0-7*; and lane 9 is *Tf(3)GLAX1.0-8/TM3*.


1 2 3 4 5 6 7 8 9
A

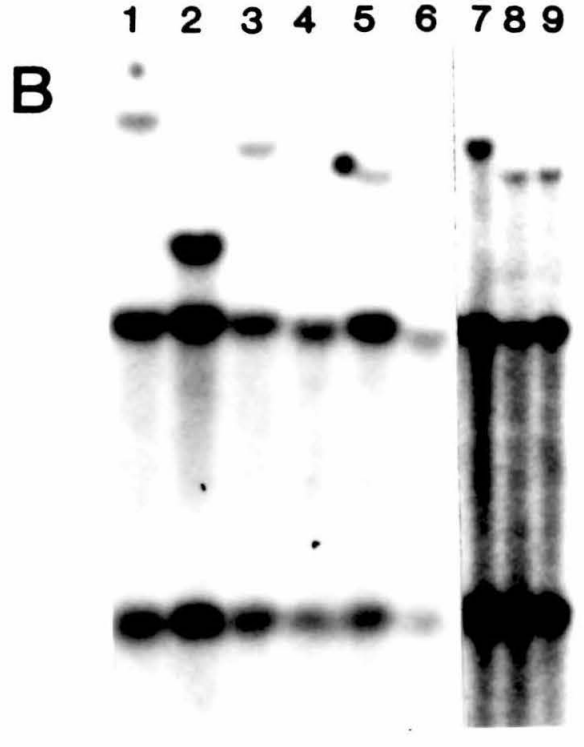
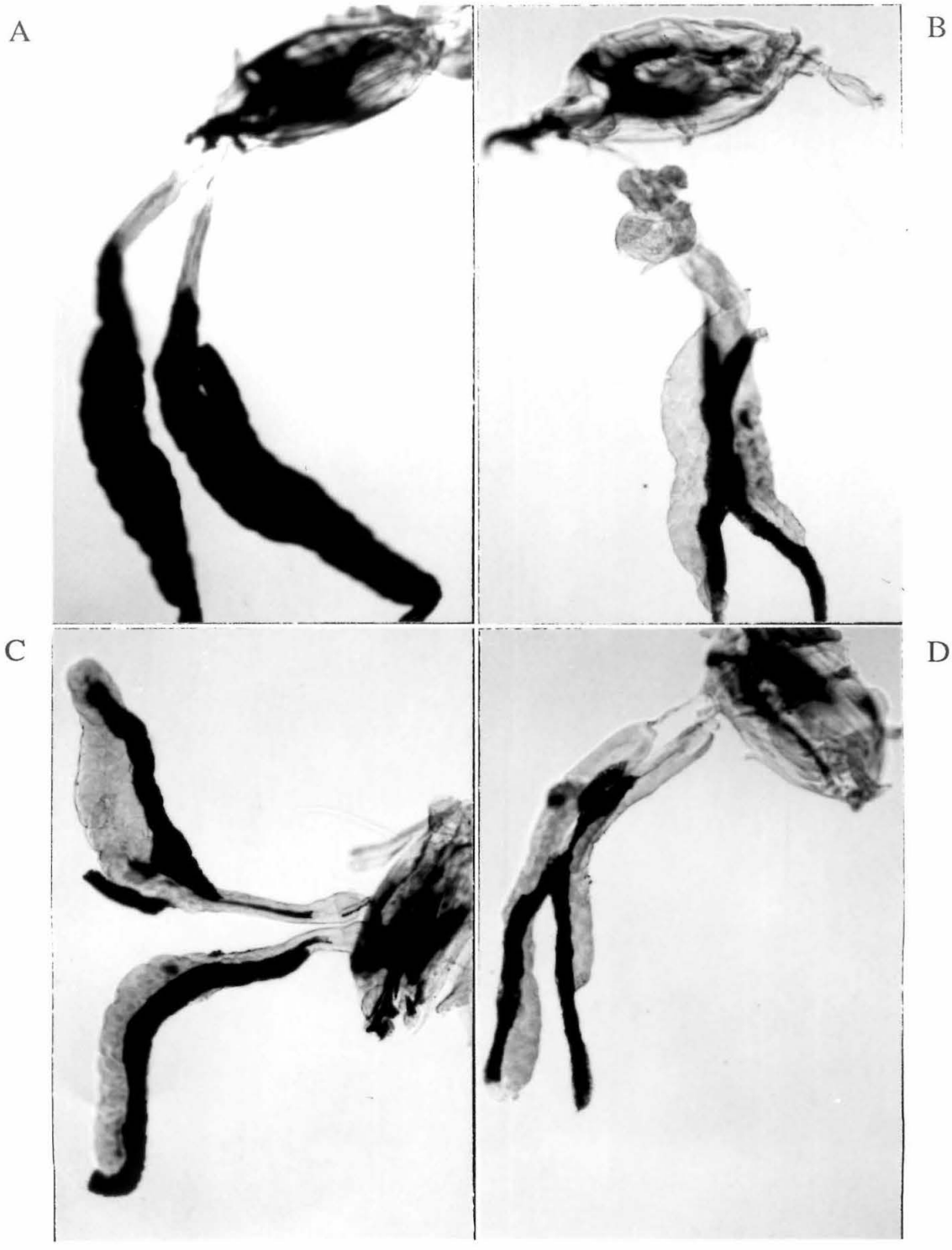

1 2 3 4 5 6 7 8 9
B

Figure 7. Effects of *l(1)npr-1* upon histochemical reactions. Virgin females from the balanced strain *y l(1)npr-1 w mal/Binsn* were mated to males from the transformant strain *Tf(2)GLAX1.0-2* and the male progeny dissected for histochemical staining (Panels A, B, E, F), or such virgin females were mated to Adh^{fn6} *cn*; ry^{502} males and their male progeny dissected for histochemical staining (Panels C, D, G, H). The salivary glands and adhering tissue in panels A through D were stained for alcohol dehydrogenase activity; in panels E through H the tissues were stained for β-galactosidase activity. The males in panels B, D, F, and H are hemizygous for *y l(1)npr-1 w mal*; the remaining males are hemizygous for the y^+ *l(1)npr-1*$^+$ w^+ mal^+ *Binsn* chromosome.


A
B
C
D


E
F
G
H

Figure 8. Representative mosaic patches due to transient expression of histochemically marked glue protein fusion genes. The plasmid pGAZ-1 was microinjected into syncytial embryos of the genotype Adh^{fn6} *cn*; ry^{502}, and third instar larvae that survived dissected. Salivary glands and adhering tissues were processed for either ADH histochemistry or for β-gal histochemistry. Panel A shows the alcohol dehydrogenase activity in one animal. Panel B shows the β-galactosidase activity in a different animal.

A

B

Figure 9. *Tf()GAX0.12* strains fail to accumulate *Sgs-7—Adh* RNA. For each strain, ten third instar larvae were dissected and salivary gland RNA was extracted. In each panel salivary gland RNA was obtained from: lane 1, *Tf(1)GLAX1.0-1*; lane 2, *Tf(2)GAX0.12-1*; lane 3, *Tf(2)GAX0.12-2*; lane 4, *Tf(3)GAX0.12-3*; lane 5, *Tf(2)GAX0.12-4*; lane 6, *Tf(3)GAX0.12-5*; and lane 7 is Adh^{fn6} *cn*; ry^{502}. Lanes 1 and 7 contain one-and-a-half animal-equivalents of RNA each; the rest contain three animal-equivalents. (A) Hybridization with the *D. melanogaster Adh* clone nDm9035; (B) Hybridization with the *D. melanogaster Sgs-7* subclone sDm9039.

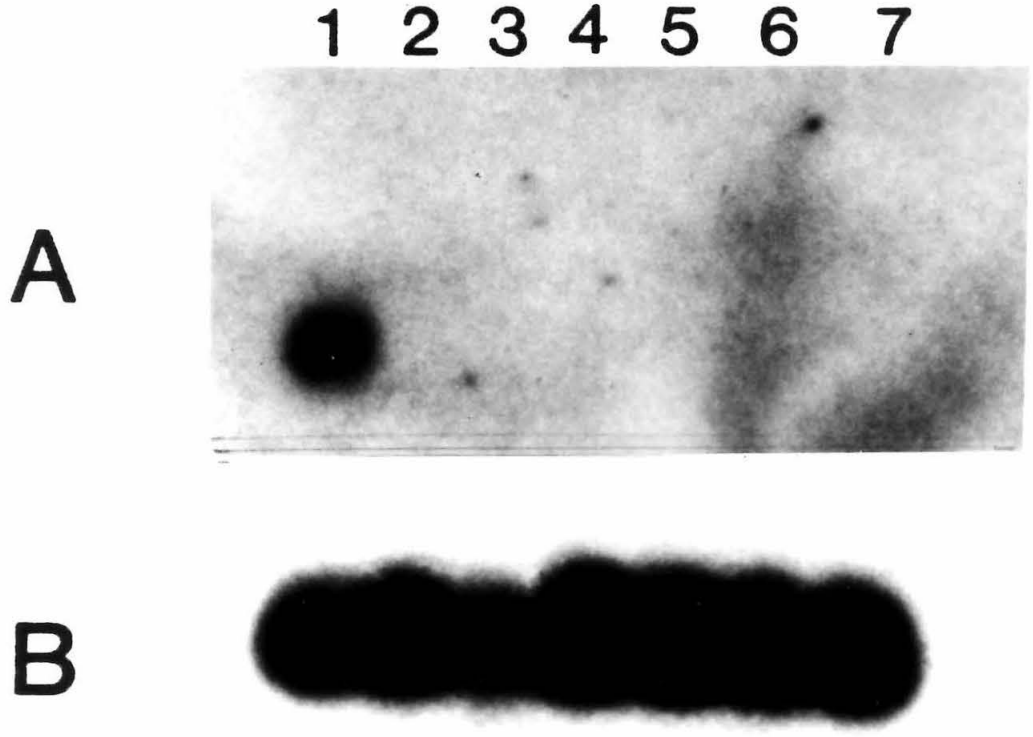
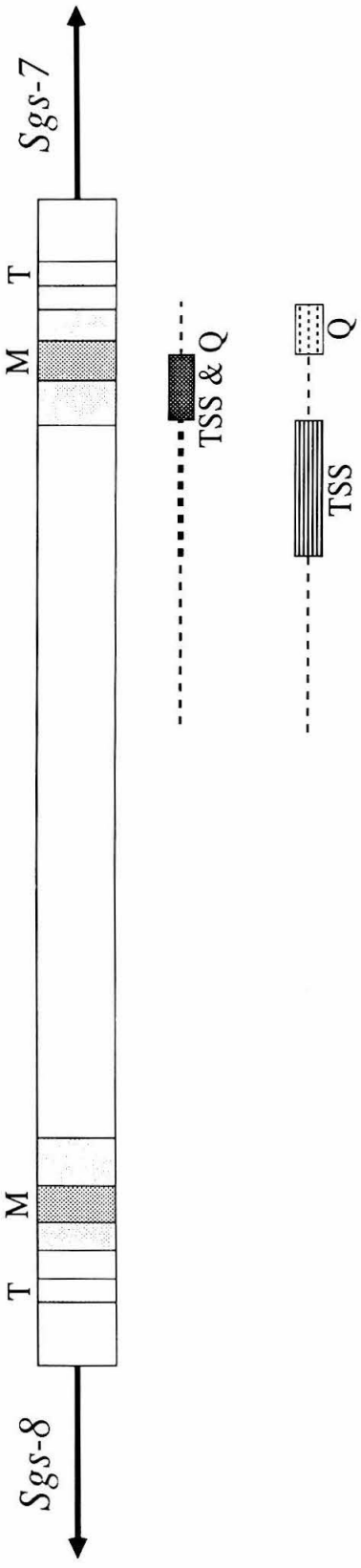

1 2 3 4 5 6 7
A
B

Figure 10. Histochemical staining of promoter-deletion derivatives of the *Sgs-8—lacZ* fusion gene. Panel A, salivary gland and adhering tissue from the strain *Tf(2)GLX0.68-3*, containing 432 base-pairs of 5' flanking sequence, were incubated in the X-Gal histochemical reaction mixture. Panel B, salivary gland and adhering tissue from the strain *Tf(2)GLX0.66-3*, containing seventeen fewer base-pairs of 5' flanking sequence, were incubated in the X-Gal histochemical reaction mixture.

A

B

Figure 11. Asymmetric location of regulatory elements in the *Sgs-7*, *Sgs-8* intergenic region. The divergently transcribed *Sgs-7*, *Sgs-8* gene pair are shown, with features within the intergenic region identified as follows: Boxes lightly shaded and labelled "T" are the Goldberg (1979) T-A-T-A motifs that precede eukaryotic RNA polymerase II transcription units. Boxes more heavily shaded are the left copy and the right copy of the conserved sequence elements in the 5' flanking regions identified by Garfinkel *et al.* (1983). The boxes most heavily shaded and labelled "M" are the segments homologous with the ecdysterone-responsive sequence identified by Mestril *et al.* (1986). In the first model of intergenic region function, tissue and stage specificity (TSS) and quantity of expression (Q) for both *Sgs-7* and *Sgs-8* are controlled by a single element that acts bidirectionally. The filled box represents the minimal length of the element as defined by the overlap of the sequences required for *Sgs-7—Adh* expression and those required for *Sgs-8—lacZ* expression. The uncertainty in assigning the left edge of the element is shown by broken lines (thick line represents the uncertainty from the transient expression results, thin line represents the uncertainty from the germline transformation results). The uncertainty in assigning the right edge of the element is shown by the thin broken line. In the second model of intergenic region function, tissue and stage specificity for both of these glue protein genes are controlled by a bidirectional element (striped box labelled TSS, uncertainty from the germline transformation results shown by broken line facing leftward) that is separable from a bidirectional element that regulates the quantity of expression of both glue protein genes (dotted box labelled Q, uncertainty from the germline transformation results shown by broken line facing leftward).


Sgs-7
T
M
TSS & Q
Q
TSS
M
T
Sgs-8