ASSESSING MOLECULAR FUNCTION IN THE DROSOPHILA NERVOUS SYSTEM: A REVERSE GENETIC APPROACH

Thesis by

Bruce A. Hamilton

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California
1993

(Submitted February 19, 1993)

Acknowledgements

I begin by thanking Frank Diaz for having sparked my interest in biology while I attended Canyon High School. Adelaide Carpenter, Dan Lindsley, and especially Rick Firtel gave me further encouragement while I was at UCSD. My time at Caltech would have been the lesser without the efforts of many people. I thank the members of my thesis committee, David Anderson, Scott Emr, Howard Lipshitz, and especially my research advisors, Elliot Meyerowitz and Kai Zinn, for their encouragement and generous support during my stay at Caltech. I am also grateful to Norman Davidson for sound advice and many productive conversations. I am indebted to Mike Palazzolo for having taken an active interest in my education and befriending me at an early stage in my graduate career. I have been fortunate to have had Joo (Esther) Chang, Chris Martin, Pete Mathers, Carol Mayeda, Mike Strathmann, Shin-Shay Tian, and K. VijayRaghavan as both substantial collaborators and good friends. I have also enjoyed the advice, expertise, and company of Sue Celniker, Barry Condron, Paul Mueller, Jeff Stack. I appreciate the diligent efforts of the undergraduate students who haved worked with me: Joe Liao of Harvard University and Albert Ho, Sacha Malin, and Sarah Yoder of Caltech.

Finally, I could never have completed, nor should ever have started, this work without the love and enthusiastic support of my family and dear friends. I must especially thank Marianne Woo for her love and forbearance.

Thank you all very much.

ABSTRACT

I describe the application of new methods for reverse genetic analyses in *Drosophila melanogaster* to genes expressed in the nervous system. The methods can be devided into two classes: tools for molecular analysis and tools for genetic analysis.

The first set of methods is designed to facilitate rapid, large-scale molecular analysis of cDNA clones. This set of tools begins with a family of bacteriophage λ cDNA cloning vectors and *E. coli* host cell strains that allow automatic plasmid subcloning by *in vivo* site-specific recombination. A high density filter hybridization and diagnostic PCR assay technique applied to size-selected libraries then substantially simplifies the isolation of full-length cDNA clones in these vectors. Next, a transposon γδ-facilitated DNA sequencing procedure minimizes the labor required to isolate a nested set of initiation sites for chain termination sequencing of each strand of a cloned DNA segment. I also describe the characterization of 250 cDNA clones isolated from adult heads on the basis of gross expression patterns in the embryonic ventral nerve cord and larval fat body.

The second set of procedures facilitates the isolation of mutations in the chromosomal genes that correspond to isolated cDNA clones. I describe three such experiments. A plasmid rescue and hybridization strategy allowed the isolation of *PlacW* elements inserted adjacent to 4 cloned genes in an array of nearly 700. Modifications of this procedure to take advantage of *P*-element local transposition allowed the isolation and characterization of an apparent null mutation in the receptor-linked protein tyrosine phosphatase gene *DPTP99A*. In a pilot study, I demonstrate the feasibility of isolating chemically induced mutations that remove targeted restriction enzyme cleavage sites with a PCR-based assay. In addition, I describe the serendipitous isolation of a *PlacW*-induced mutation, *encumbered*, that affects the morphogenesis of imaginal wing discs as well as adult longevity, activity, and fertility.

Table of Contents

Chapter 1
Introduction

A brain and its genome

Many features of neural development are shared by most nervous systems, implying that the underlying molecular mechanisms are likely to be conserved. As for any differentiated cell type, a multipotent precursor cell must first acquire a specific cell fate. This appears to occur, in both vertebrates and invertebrates neuronal precursors, by a series of choices that result in progressive fate restrictions (Anderson 1989; Doe 1992; Greenspan 1992) and is often accompanied by a defined pattern of cell migration. The developing neuron then typically extends an axon along a prescribed pathway to reach its intended target (Ramón y Cajal 1990). Once at its target cell, the neurite gives rise to a stereotypic pattern of synaptic specializations (Goodman et al. 1984). Maturation and modification of these synaptic connections may continue throughout the life of the organism. Furthermore, these developmental steps appear to be intertwined. Fate determination events, even at the level of single regulatory molecules, can affect choice of axonal pathway (Doe 1988; Doe 1992). Conversely, retrograde signals from innervation targets can affect other differentiation events, such as choice of neurotransmitter phenotype by rat sympathetic neurons (Schotzinger and Landis 1988; Landis 1990). Such observations argue that the molecular mechanisms underlying these events may be interdependent.

The complexity of cell types in the nervous system and the diversity of extracellular signals to which they must respond, both during development and in mature function, argues that construction of a nervous system involves a large number of molecules. While this may not require an unusual complexity of molecules to be expressed in any one cell , such molecular complexity is likely to be required within the population of cells to establish a properly wired and functioning nervous system. This provokes two questions. How many molecules are required for development, both at the

level of the individual neuron and at the level of a complete nervous system? What are these molecules and what does each one do?

I have based my approach to the above questions on the premise the events of neuronal development require the concerted activities of a large number of gene products. Thus, I believe that a detailed understanding of these processes will require the characterization of a large number of molecules. Therefore, rather than selecting one or two neuronal molecules for immediate analysis by existing methods, I began my studies by trying to develop methods that would allow extensive analyses of relatively large panels of cloned genes. I chose to work in Drosophila because of the accessibility of its genome to both molecular and genetic techniques and because of its relatively small genome size. Although I feel that the interrelation of developmental processes justifies a broad approach to neuronal development, I have been particularly interested in axon pathfinding and synaptic specification and this is reflected in some of the later experiments described here.

Molecular complexity underlying neural development and function: how many players in the game?

Estimates of gene number in the fly: The issue of gene number in Drosophila has received considerable attention from several authors. More than 4000 characterized genes have been catalogued by Lindsley and Zimm (1992), and new genes continue to be reported at an unprecedented rate. Intensive genetic and cytological analyses of sample genomic regions have produced gene number estimates in the range of 5,000-10,000 (Muller and Prokofyeva 1935; Judd, Shen and Kaufman 1972; Hochman 1974; Sorsa 1988). These are necessarily minimum estimates, as they include only genes whose mutant phenotypes are readily observed in laboratory culture; disruptions of many structural genes for which no phenotype is evident have been reported (discussed further

in Chapter 4). RNA-driven reassociation kinetics from several tissue preparations led Izquierdo and Bishop (1979) to estimate the existence of more than 7000 distinct polyadenylated transcripts in Drosophila. However, this type of analysis is also prone to understatement, owing to both cell-type specificity and message prevalence. Genes whose transcripts are expressed (or polyadenylated) at low levels, in rare cell types, or transiently during brief developmental windows would be overlooked in these experiments (Palazzolo et al. 1989). Better data should accrue from genomic sequencing efforts now in progress; as large stretches of contiguous sequence are generated and analyzed, it should become feasible to estimate an average density of protein-encoding genes in the genome. Such data for chromosome III of *Saccharomyces cerevisiae* have been recently reported and suggest a 5-fold higher gene density than had been represented on the genetic map of that intensively studied organism (Oliver et al. 1992).

Genes expressed in the nervous system: Estimating the number of genes used to construct the nervous system is a more troublesome task for two reasons. First, one must decide what such an estimate will include. Although general cell-lethal genes and most housekeeping genes (and any neural-specific isozymes of them) are required for the development of the nervous system, these are of comparatively little interest to most neuroscientists. I make this point because it matters in the types of data used to count the genes. Most of the experiments that bear on this issue assay only the expression patterns of genes. Second, one must consider what such an estimate will mean. Although the nervous system may express a larger number of genes than other tissues, the nervous system also comprises more cell types than any other tissue. Thus, it is not necessarily (and probably not) the case that each individual neuron expresses more of its genome than any other differentiated cell type. On the other hand, it is precisely the diversity of component cell types and the complexity of their interactions that allow the cells of the

brain to perform their collective functions, to create networks with emergent properties, and to fascinate people like me.

Direct molecular data for Drosophila on this subject are limited. One experimental difficulty in assessing the number of genes expressed in the fly nervous system has been in isolating sufficient quantities of appropriate tissues for reassociation kinetics experiments such as have been done for whole animals. However, some informative data can be derived from cDNA cloning experiments. Palazzolo and coworkers used a pool sorting strategy to isolate 436 cDNA clones that represent mRNA expressed in the adult head but not in the undifferentiated cells of the early embryo (Palazzolo et al. 1989). Because the adult head is largely neural tissue and is thought to contain relatively few other cell types (predominantly fat body and muscle), this clone set should be enriched for messages that are expressed in neural tissue. Of these cloned genes, 250 are expressed by late embryogenesis. These 250 clones were examined for expression in the embryonic CNS and larval fat body by bulk filter hybridization with cDNA probes (Appendix 3). Among these 250 clones, 109 hybridized to labeled cDNA prepared from purified embryonic nerve cords, 67 to cDNA from larval fat body, and 35 to both. Thus, 52% (74 out of 141) of the detected clones, or 68% of clones detected by nerve cord cDNA, showed some specificity for neural tissue.

A separate line of evidence for heavy genomic commitment to the nervous system comes from so-called enhancer trap experiments with modified *P* elements. O'Kane and Gehring examined 49 *lacZ*-expressing *P* element lines; of these 23 (46%) expressed detectable β-galactosidase in CNS or PNS (O'Kane and Gehring 1987). In a similar experiment, Bier and coworkers screened 3768 independent *PlacW* lines for *lacZ* expression. 31% of their lines stain in the nervous system (Bier et al. 1989). Since only 64% of these lines stained any embryonic cells, this is 49% of staining lines. Bellen et al. reported similar results for 536 independent *P1ArB* insertion lines: 48% of these lines

stain CNS and 20% stain PNS (Bellen et al. 1989). One potential difficulty in drawing strict conclusions from this data is that the artificial enhancer-promoter combinations generated by insertion of the modified *P* element may not reflect an endogenous gene expression pattern selected through evolution. However, from the insertion strains analyzed in detail to date, this does not appear to generate frequent discrepancies. A more serious problem in drawing quantitative conclusions from this type of analysis is the non-random distribution of *P* element insertion sites. Like most transposons, *P* elements are prone to both hot spots and cold spots for insertion and the basis for these preferences is not entirely clear. Nonetheless, these experiments provide further evidence that a large fraction of the expressed fly genome is expressed in the nervous system.

Recent evidence for this has also been presented from a purely genetic study. Somatic mosaic analysis of the fly visual system for 65 randomly chosen X-linked lethal mutations revealed 26 cell-lethal genes and 21 genes for which mutant cell clones produced abnormal eye or laminar structures (Thaker and Kankel 1992). Thus, 72% of these lethal genes appear to be required for the development of the adult visual system, 32% with some apparent specificity. Although limited in several respects, this experiment is significant in that it does not strictly rely on detected expression patterns of gene products, but rather on a morphogenetic requirement for each gene.

Deriving hypotheses of molecular functions from cDNA sequence and expression pattern

Assessing neuronal functions of gene products from molecular analyses: Sequence analysis of a cloned molecule may suggest a biochemical or cellular function for the corresponding protein (or RNA) by comparison to sequences of known molecules. Knowing the expression pattern may further suggest an organismal function by delimiting potential sites of action. Such data provide hypotheses that can be tested in situ in the fly by altering the sequence or expression of the gene product.

Meaningful large-scale analysis of expressed genes requires each of the following steps for each gene: identify a (full-length) cDNA clone to be sequenced, generate and analyze the DNA sequence of the clone, and determine the expression pattern of the gene. Each of these steps has been tedious and performed only on a gene-by gene basis. I have tried to improve the rate-limiting steps in each of these procedures to enable a small group of workers (or eventually, hopefully, machines) to analyze large numbers of genes in a rapid fashion. The results of these efforts are detailed in Chapter 2. Automation and scale-up for at least some of these methods are currently in progress by others as part of the Human Genome Initiative (M. Palazzolo, personal communication).

Methods for rapid cDNA analysis: Sorting cDNA clone collections by cross-hybridization or equivalent strategies minimizes redundant isolations that would occur due to both dissimilar expression levels in a chosen tissue (which can vary by several orders of magnitude in complex tissues) and random sampling of normalized libraries (Palazzolo et al. 1989). The uniqueness of each clone justifies the investment of greater effort in characterizing each clone and its gene than would be feasible in a random clone selection strategy such as described by Ventor and colleagues (Adams et al. 1992). In addition, a high percentage of the sorted clone collection can be assayed for gross expression pattern by hybridizing labeled cDNA from defined sources to an array of clones (Appendix 3). Such hybridization experiments are useful in prioritizing members of the clone array for further analyses. These data increase the information value of each selected clone as a unique tag for a gene that is expressed in a tissue of interest. Once the initial clone set has been sorted and prioritized, a clone or set of clones that represent each corresponding message can be sequenced and subjected to other detailed analyses with a minimum of duplication.

In Chapter 2, I describe a set of tools and techniques designed to minimize and simplify the manipulations required for the steps between selection of a clone from the

sorted pool and generation of full coding sequence. In many cases, the clone used for sorting may be too short to include the full coding sequence. A rapid and simple method for obtaining long cDNA clones from libraries is described Section 2.3. Cloning vectors and host cells that permit automatic (*in vivo*) subcloning from phage to plasmid are described in Sections 2.2 and 2.5. Transposon facilitated DNA sequencing (Section 2.4) allows rapid data collection by minimizing the amount of labor required to generate nested sets sequencing templates.

Coding sequence analysis of cDNA, combined with expression pattern analysis should allow hypotheses of cellular and developmental functions to be made for a significant fraction of molecules. The availability of sophisticated genetic methods should allow testing of these hypotheses in the fruit fly. In addition, such methods can be applied to assess the developmental roles of molecules whose cellular or biochemical activity is known.

Adhesion molecules, tyrosine phosphorylation, and axon pathfinding

One of the most striking features of neuronal development is the extension of neurites along stereotyped pathways to establish an elaborate pattern of specific synaptic connections. Although detailed cellular descriptions of extending neurites and growth cones date back more than 100 years to Ramón y Cajal's first description in 1890 (Ramón y Cajal 1990), the molecular mechanisms that drive these events remain poorly understood. Pathfinding by growth cones and selective fasciculation of developing axons requires both extrinsic cues in the extracellular space and a correct interpretation of them by the neuron. Evidence has been presented for diffusible factors, cell surface molecules, and extracellular matrix molecules functioning as guidance cues and for a variety of mechanisms to transduce these signals across the membrane(Bixby and Harris 1991).

Negative cues that cause the inhibition or collapse of growth cones have also been described (Patterson 1988; Walter, Allsopp and Bonhoeffer 1990).

Much attention has been given to potential roles for cell adhesion molecules in the development of axonal pattern (for example, see Goodman et al. 1984; Edelman 1986; Hortsch and Goodman 1991). In vivo ligands or binding partners (such as the Notch and Delta proteins in Drosophila) are not known for the majority of these molecules and precise developmental roles of their adhesive properties remain largely unclear. Moreover, adhesion is unlikely to be the only role such molecules can play. For example, the genetic interaction in Drosophila between mutations in the *abl* homologue and the gene encoding the fasciclin I homophilic adhesion molecule argues that the fasciclin I protein might be involved in a signal transduction pathway parallel to that of the *abl* tyrosine kinase (Elkins et al. 1990). Furthermore, exposure of rat PC12 cells to antibodies against the adhesion molecules L1 and N-CAM has been reported to trigger changes in intracellular second messenger levels, possibly through a G-protein coupled mechanism (Schuch, Lohse and Schachner 1989). In addition, a number of presumed signal transduction molecules such as protein kinases (Hunter 1987) and protein phosphatases (Fischer, Charbonneau and Tonks 1991; Charbonneau and Tonks 1992) have been cloned that include adhesion molecule-like extracellular domains in addition to their catalytic domains. These results raise the question of whether a significant fraction of presumed adhesion molecules might actually be more important in reporting on the extracellular milieu than in adhering to it.

Four experiments have suggested a specific role for the regulation of protein phosphotyrosine in events related to neurite outgrowth. Maness et al. (Maness et al. 1988) observed high concentration of the *src* tyrosine kinase in vertebrate growth cones, however they were unable to directly assess its function in the context of axon growth and guidance. Gertler et al. (Gertler et al. 1989) and Elkins et al. (Elkins et al. 1990)

provided stronger evidence in Drosophila by showing that two different double mutant combinations that include a null allele of the *D-abl* tyrosine kinase suffer severe and specific disruptions in the axonal array of the embryonic CNS. In addition, three groups have reported the isolation of axon-specific receptor-linked protein tyrosine phosphatases (R-PTPs) in Drosophila (Hariharan, Chuang and Rubin 1991; Tian, Tsoulfas and Zinn 1991; Yang et al. 1991). A first functional test of one of these molecules in presented in Chapter 4.

Elucidating the molecular details of neural development is likely to require not only the identification of diverse classes of molecules, but also the ability to test hypotheses about their functions in the nervous system. The variety of genetic approaches available in Drosophila make this a desirable organism in which to test such hypotheses.

Testing hypotheses of function with reverse genetics

Concepts and limitations: Perhaps, in a strict sense, the term reverse genetics should only be applied when germline mutations are isolated in the endogenous gene of a known molecule. However, the current technical repertoire includes a variety of methods for perturbing gene function that fall outside this narrow definition. I will expand the discussion (but not the definition) here to also include germline and somatic interference methods such as antisense expression, dominant interfering product expression, and laser inactivation of proteins.

The chief advantage of a germline mutational approach is also its chief disadvantage: it affects every copy of the gene product in every cell. The advantage to this is that the level and type of mutant gene product is reproducible from experiment to experiment and cell-autonomy is not a confounding problem in the interpretation. On the other hand, because many genes are known to function in several different developmental

pathways, the potential exists for early phenotypes to mask later effects. For example, the effects of *ftz* and *eve* on fate determination in the nervous system were not seen initially because of earlier defects in segmentation. However, as this example suggests, such problems may eventually be overcome by introduction of an appropriate transgene (Doe 1988).

Gene product interference methods offer a different set of trade-offs (Herskowitz 1987). Two advantages of interfering transgenes, such as antisense genes or genes encoding poison product monomers, are flexibility and speed. A large variety of potential interfering transgenes can be tried in parallel experiments, by either germline or somatic transformation. The experimenter could potentially target different splice variants of the gene product or different parts of a complex expression pattern. On the other hand, it is often difficult to determine the level to which the targeted gene function has been reduced and whether this reduction is uniform in every cell type of interest. Thus, it may be difficult to determine whether a given result corresponds to a loss, reduction, or misregulation of function. Negative results in such experiments are frequent and difficult to interpret. Similar arguments can be applied to methods aimed directly at protein products, such as CALI. However, such methods have proven successful and enlightening in a number of instances (Jay and Keshishian 1990; Amaya, Musci and Kirschner 1991).

The crucial differences between these methods for my purposes have been scale and generality. While transgenes and CALI reagents must be independently produced and tested for each targeted gene product, molecular screening methods for reverse genetics have the potential to be scaled up to encompass large numbers of targets (Chapter 3). Furthermore, the existence of a null mutation is easily proven with molecular probes.

Mutagenesis methods for reverse genetics in Drosophila: A variety of methods have been used to introduce and identify new mutations into cloned Drosophila genes. Reported successes with chemical mutagenesis have requirements that are untenable for general application. Classical genetic screens over mapped deficiencies require prior knowledge (or a good guess) of the phenotype. Screens based on loss of antibody binding are labor intensive and require an antibody for each gene being targeted. Nucleic acid screens for chemically induced mutations in the fly have not been reported, although a test case is presented here in Appendix 2.

Mutagenesis by *P* element transposon tagging has several distinct advantages. The rate of mutagenesis can be tightly controlled and precisely followed using genetically marked, non autonomous elements as mutagens under the direction of an immobile source of transposase. The isolation of one allele provides a direct route to isolating subsequent alleles through either imprecise excision, transposon recruitment, or local transposition mechanisms. Engels and coworkers have also reported the use of *P* element insertions to generated targeted mutations through an excision-dependent gene conversion mechanism (Engels et al. 1990). This method requires the prior isolation of a *P* element within a few kb of the targeted site. Molecular screening methods have been reported that allow the isolation of mutations in the absence of a known phenotype.

Three types of molecular screening methods for *P* element insertions have been reported. PCR-based screens have been successful in several laboratories (Ballinger and Benzer 1989; Kaiser and Goodwin 1990; Pereira et al. 1992). I have developed screens based on plasmid rescue to enable simultaneous screening of large numbers of target genes (Chapter 3) and to facilitate screening the entire length of a single, moderately large gene (Chapter 4). In addition, I have described the use of direct DNA gel blot hybridization as an assay for highly efficient mutagenesis based on the local transposition of *P* elements.

I have developed these tools not for their own sake, but to analyze the molecules at hand. For example, the cDNA reagents and methods that I developed have been applied not only to large scale analyses but also to the more directed analysis of genes encoding R-PTPs expressed on developing CNS axons (Tian et al. 1991) and a candidate gene for a mutation that affects morphogenesis and viability in the fly (Appendix 1). The mutagenesis strategies described in Chapter 3 were adapted to directly test the role of one R-PTP in elaborating axon pathways during CNS development (Chapter 4) and these methods are currently being used for the mutagenesis of the remaining axonal R-PTPs to allow a more thorough genetic analysis of these molecules. Furthermore, I firmly believe that the pace of new technologies and new techniques can only compound the rate of new discoveries and new understandings of neural development.

Literature cited

Adams, M. D., M. Dubnick, A. R. Kerlavage, R. Moreno, J. M. Kelley, T. R. Utterback, J. W. Nagle, C. Fields and J. C. Venter, 1992. Sequence identification of 2,375 human brain genes. *Nature* **355:** 632-634.

Amaya, E., T. J. Musci and M. W. Kirschner, 1991. Expression of a dominant negative mutant of the FGF receptor disrupts mesoderm formation in Xenopus embryos. *Cell* **66:** 257-270.

Anderson, D. J., 1989. The neural crest cell lineage problem: neuropoiesis? *Neuron* **3:** 1-12.

Ballinger, D. G. and S. Benzer, 1989. Targeted gene mutations in Drosophila. *Proc. Nat'l. Acad. Sci.* **86:** 9402-9406.

Bellen, H., C. J. O'Kane, C. Wilson, U. Grossniklaus, R. Kurth-Pearson and W. J. Gehring, 1989. P-element-mediated enhancer detection: a versatile method to study development in Drosophila. *Genes Dev.* **3:** 1288-1300.

Bier, E., H. Vaessin, S. Shepherd, K. Lee, K. McCall, S. Barbel, L. Ackerman, R. Carretto, T. Uemura, E. Grell, L. Y. Jan and Y. N. Jan, 1989. Searching for pattern and mutation in the Drosophila genome with a P-*lacZ* vector. *Genes Dev.* **3:** 1273-1287.

Bixby, J. L. and W. A. Harris, 1991. Molecular mechanisms of axon growth and guidance. *Ann. Rev. Cell Biol.* **7:** 117-159.

Charbonneau, H. and N. K. Tonks, 1992. 1002 protein phosphatases. *Ann. Rev. Cell Biol.* **8:** 463-493.

Doe, C. Q., 1988. Expression and function of the segmentation gene *fushi tarazu* during Drosophila neurogenesis. *Science* **239:** 170-175.

Doe, C. Q., 1992. The generation of neuronal diversity in the Drosophila embryonic central nervous system. pp. 119-154. In: Determinants of neuronal identity, Edited by M. Shankland and E. R. Macagno. Academic Press, Inc., San Diego, CA.

Edelman, G. M., 1986. Cell adhesion molecules in the regulation of animal form and tissue pattern. *Ann. Rev. Cell Biol.* **2:** 81-116.

Elkins, T., K. Zinn, L. McAllister, F. M. Hoffmann and C. S. Goodman, 1990. Genetic analysis of a Drosophila neural cell adhesion molecule: interaction of *fasciclin I* and Abelson tyrosine kinase mutations. *Cell* **60:** 565-575.

Engels, W. R., D. M. Johnson-Schlitz, W. B. Eggleston and J. Sved, 1990. High-frequency P element loss in Drosophila is homolog dependent. *Cell* **62:** 515-525.

Fischer, E. H., H. Charbonneau and N. K. Tonks, 1991. Protein tyrosine phosphatases: a diverse family of intracellular and transmembrane enzymes. *Science* **253:** 401-406.

Gertler, F. B., R. L. Bennett, M. J. Clark and F. M. Hoffmann, 1989. Drosophila *abl* tyrosine kinase in embryonic CNS axons: a role in axonogenesis is revealed through dosage-sensitive interactions with *disabled*. *Cell* **58:** 103-113.

Goodman, C. S., M. J. Bastiani, C. Q. Doe, S. du Lac, S. L. Helfand, J. Y. Kuwada and J. B. Thomas, 1984. Cell recognition during neuronal development. *Science* **225:** 1271-1279.

Greenspan, R. J., 1992 Initial determination of the neurectoderm in Drosophila. pp. 155-188. In: Determinants of neuronal identity, Edited by M. Shankland and E. R. Macagno. Academic Press, Inc., San Diego, CA.

Herskowitz, I., 1987 Functional inactivation of genes by dominant negative mutations. *Nature* **329:** 219-222.

Hariharan, I., P.-T. Chuang and G. M. Rubin, 1991 Cloning and characterization of a receptor-class phosphotyrosine phosphatase gene expressed on central nervous system axons in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci.* **88:** 11266-11270.

Hochman, B., 1974 Analysis of a whole chromosome in Drosophila. *Cold Spring Harbor Symp. Quant. Biol.* **38:** 581-589.

Hortsch, M. and C. S. Goodman, 1991 Cell and substrate adhesion molecules in Drosophila. *Ann. Rev. Cell Biol.* **7:** 505-557.

Hunter, T., 1987 A thousand and one protein kinases. *Cell* **50:** 823-829.

Izquierdo, M. and J. O. Bishop, 1979 An analysis of cytoplasmic RNA populations in *Drosophila melanogaster*. Oregon R. *Biochem. Genet.* **17:** 473-497.

Jay, D. G. and H. Keshishian, 1990 Laser inactivation of fasciclin I disrupts axon adhesion of grasshopper pioneer neurons. *Nature* **348:** 548-550.

Judd, B. H., M. W. Shen and T. C. Kaufman, 1972 The anatomy and function of a segment of the X chromosome of *Drosophila melanogaster*. *Genetics* **71:** 139-156.

Kaiser, K. and S. F. Goodwin, 1990 "Site-selected" transposon mutagenesis of Drosophila. *Proc. Nat'l. Acad. Sci.* **87:** 1686-1690.

Landis, S. C., 1990 Target regulation of neurotransmitter phenotype. *Trends Neurosci.* **13:** 344-350.

Lindsley, D. L. and G. G. Zimm, 1992 *The Genome of Drosophila melanogaster*. Academic Press, Inc.: San Diego, CA. pp.

Maness, P. F., M. Aubury, C. G. Shores, L. Frame and K. H. Pfenninger, 1988 *c-src* gene product in developing rat brain is enriched in nerve growth cone membranes. *Proc. Natl. Acad. Sci.* **85:** 5001-5005.

Muller, H. J. and A. A. Prokofyeva, 1935 The individual gene in relation to the chromomere and the chromosome. *Proc. Nat'l. Acad. Sci.* **21:** 16-26.

O'Kane, C. J. and W. J. Gehring, 1987 Detection *in situ* of genomic regulatory elements in Drosophila. *Proc. Nat'l. Acad. Sci.* **84:** 9123-9127.

Oliver, S. G., Q. J. M. Vanderaart, M. L. Agostonicarbone, M. Aigle, L. Alberghina, D. Alexandraki, G. Antoine, R. Anwar, J. P. G. Ballesta et al, 1992 The complete DNA sequence of yeast chromosome III. *Nature* **357:** 38-46.

Palazzolo, M. J., D. R. Hyde, K. VijayRaghavan, K. Mecklenberg, S. Benzer and E. Meyerowitz, 1989 Use of a new strategy to isolate and characterize 436 Drosophila cDNA clones corresponding to RNAs detected in adult heads but not in early embryos. *Neuron* **3:** 527-539.

Patterson, P. H., 1988 On the importance of being inhibited, or saying no to growth cones. *Neuron* **1:** 263-267.

Pereira, A., J. Doshen, E. Tanaka and L. S. B. Goldstein, 1992 Genetic analysis of a Drosophila microtubule-associated protein. *J. Cell Biol.* **116:** 377-383.

Ramón y Cajal, S., 1990 New ideas on the structure of the nervous system in man and vertebrates. MIT Press: Cambridge, MA. 195 pp.

Schotzinger, R. J. and S. C. Landis, 1988 Cholinergic phenotype developed by noradrenergic sympathetic neurons after innervation of a novel cholinergic target in vivo. *Nature* **335:** 637-639.

Schuch, U., M. J. Lohse and M. Schachner, 1989 Neural cell adhesion molecules influence second messenger systems. *Neuron* **3:** 13-20.

Sorsa, V., 1988 *Chromosome maps of Drosophila*. CRC Press, Inc.: Boca Raton, FL. 149 pp.

Thaker, H. M. and D. R. Kankel, 1992 Mosaic analysis gives an estimate of the extent of genomic involvement in the development of the visual system in *Drosophila melanogaster*. *Genetics* **131:** 883-894.

Tian, S.-S., P. Tsoulfas and K. Zinn, 1991 Three receptor-linked protein-tyrosine phosphatases are selectively expressed on central nervous system axons in the Drosophila embryo. *Cell* **67:** 675-685.

Walter, J., T. E. Allsopp and F. Bonhoeffer, 1990 A common denominator of growth cone guidance and collapse? *Trends Neurosci.* **13:** 447-452.

Yang, X., K. T. Seow, S. M. Bahri, S. H. Oon and W. Chia, 1991 Two Drosophila receptor-like tyrosine phosphatase genes are expressed in a subset of developing axons and pioneer neurons in the embryonic CNS. *Cell* **67:** 661-673.

Chapter 2
Tools for large scale cDNA analysis

§2.1: Overview

Overview of Chapter 2

I first decided to pursue the molecular genetics of neural development in about 1988. I decided that, for me, a satisfactory understanding of such complex developmental phenomena would eventually require the identification and analysis of a large number of distinct gene products. Toward this end I began a collaborative effort with the several colleagues, whose names appear as co-authors on the following subsections, to develop a set of tools that would allow a rapid analysis of a large number of protein-encoding sequences expressed in neural tissue. This chapter consists of a series of technical papers that came out of this collaborative effort, presented in chronological order. Together these methods simplify each major step required for cDNA sequence analysis, from isolating appropriate phage clones to generating a nested set of sequencing templates.

One step in cDNA analysis that had required considerable effort for each clone under consideration is subcloning cDNA inserts from a library phage into a more versatile plasmid vector for further analysis. To circumvent this, my collaborators and I developed a series of phage vectors that allow "automatic" subcloning mediated by in vivo site-specific recombination, described in section 2.2. My specific contributions to this effort were in the design and construction of the λLOX arms, the plasmids embedded in the λEXLX vectors, and the P1 lysogen strains used as recombinagenic hosts for in vivo plasmid subcloning and in the functional testing of these features. I also constructed and characterized the libraries in λEXLX described in this paper. The development of improved host strains, a later extension of this work, is described in section 2.5.

Another limiting step had been the isolation of sufficiently large cDNA clones to allow analysis of coding regions. In early 1989, Mike Palazzolo and I began to discuss and experiment with different ways to isolate long cDNA molecules, starting from either cDNA libraries or polyadenylated RNA. Out of this work, I constructed a series of size-

selected and unselected cDNA libraries from Drosophila head RNA and demonstrated and refined the high-density plaque screen and diagnostic PCR screening method described in section 2.3.

The remaining roadblock that we all felt needed to be addressed was the sequence analysis itself. Existing methods for generating overlapping templates for sequencing required significant in vitro manipulations for each clone that did not seem likely to ever become automatable. Mike Strathmann, Mike Palazzolo and I began to discuss the use of bacterial transposons as mobile priming sites for chain termination sequencing. Initially, we each set out to test different transposition systems, including derivatives of Tn5 and bacteriophage mu. It was soon clear, however, that the γδ experiments initiated by Strathmann would be by far the superior method. Carol Mayeda and I tested the system further with subclones of the near full-length cDNA clones isolated in the experiments described in section 2.3. These results form a significant portion of the data presented in the Strathmann et al. paper included here as section 2.4.

Together, these methods provide a fluent and relatively labor-free transition from the isolation of a prototype cDNA clone to the sequence analysis of a corresponding full-length clone.

§2.2: Phage lambda cDNA cloning vectors for subtrative hybridization, fusion-protein synthesis and Cre-*loxP* automatic plasmid subcloning

Gene, 88 (1990) 25–36
Elsevier

GENE 03454

Phage lambda cDNA cloning vectors for subtractive hybridization, fusion-protein synthesis and Cre-*loxP* automatic plasmid subcloning

(Recombinant DNA; bacteriophages λ, T7, f1 and P1; *E. coli* RNA polymerase; promoters; site-specific recombination)

Michael J. Palazzolo [a], Bruce A. Hamilton [a], Dali Ding [a], Christopher H. Martin [b], David A. Mead [c], Robert C. Mierendorf [d], K. Vijay Raghavan [a,*], Elliot M. Meyerowitz [a] and Howard D. Lipshitz [a]

[a] Division of Biology 156-29, California Institute of Technology, Pasadena, CA 91125 (U.S.A.); [b] Department of Biological Sciences, Columbia University, New York, NY 10027 (U.S.A.) Tel. (212)854-3066; [c] Department of Chemistry, University of Wisconsin, Madison, WI 53706 (U.S.A.) Tel. (608)262-2021 and [d] Novagen, Inc., Madison, WI 53711 (U.S.A.) Tel. (608)238-6110

Received by A.D. Riggs: 30 July 1989
Revised: 2 October 1989
Accepted: 5 October 1989

SUMMARY

We describe the construction and use of two classes of cDNA cloning vectors. The first class comprises the λEXLX(+) and λEXLX(−) vectors that can be used for the expression in *Escherichia coli* of proteins encoded by cDNA inserts. This is achieved by the fusion of cDNA open reading frames to the T7 gene *10* promoter and protein-coding sequences. The second class, the λSHLX vectors, allows the generation of large amounts of single-stranded DNA or synthetic cRNA that can be used in subtractive hybridization procedures. Both classes of vectors are designed to allow directional cDNA cloning with non-enzymatic protection of internal restriction sites. In addition, they are designed to facilitate conversion from phage λ to plasmid clones using a genetic method based on the bacteriophage P1 site-specific recombination system; we refer to this as automatic Cre-*loxP* plasmid subcloning. The phage λ arms, λLOX, used in the construction of these vectors have unique restriction sites positioned between the two *loxP* sites. Insertion of a specialized plasmid between these sites will convert it into a phage λ cDNA cloning vector with automatic plasmid subcloning capability.

INTRODUCTION

We have constructed a new series of phage λ cDNA cloning vectors with a number of features designed to facilitate and simplify the construction of directionally cloned cDNA libraries, their subtractive hybridization, the rapid analysis of cDNA inserts by nt sequence analysis and the expression of fusion proteins. Both the λEXLX and the λSHLX classes of vectors are designed to allow conventional, directional cDNA cloning (Palazzolo and Meyerowitz, 1987; Meissner et al., 1987) or cDNA synthesis with non-enzymatic protection of internal restriction sites (M. Strathmann and M. Simon, unpublished) followed by directional cloning. In addition, they are designed to

Correspondence to: Dr. H.D. Lipshitz, Division of Biology 156-29, California Institute of Technology, Pasadena, CA 91125 (U.S.A.) Tel. (818)356-6446; Fax (818)564-8709; or: Dr. E.M. Meyerowitz, Division of Biology 156-29, California Institute of Technology, Pasadena, CA 91125 (U.S.A.) Tel. (818)356-6889; Fax (818)449-0756.

* Present address: Molecular Biology Unit, Tata Institute of Fundamental Research, Bombay 400005 (India) Tel. (22)495-2411.

Abbreviations: aa, amino acid(s); Ap, ampicillin; bp, base pair(s); Cm, chloramphenicol; ds, double strand(ed); IPTG, isopropyl-β-D-thiogalactopyranoside; kb, kilobase(s) or 1000 bp; Km, kanamycin; nt, nucleotide(s); oligo, oligodeoxyribonucleotide; ORF, open reading frame; *ori*, origin of DNA replication; pfu, plaque-forming units; PolIk, Klenow (large) fragment of *E. coli* DNA polymerase I; PCR, polymerase chain reaction; ^{R}, resistant/resistance; ^{S}, sensitive/sensitivity; SDS, sodium dodecyl sulfate; ss, single strand(ed); Tc, tetracyclin; wt, wild type; [ ], denotes plasmid- or P1 phage-carrying state.

facilitate conversion from λ to plasmid clones in vivo using a method we refer to as automatic Cre-*loxP* plasmid subcloning. Automatic subcloning makes it possible to combine the high efficiency of cloning afforded by phage λ-derived vectors with rapid and simple conversion of each clone to a specialized plasmid (Brenner et al., 1982; Sauer and Henderson, 1988; Short et al., 1988). The conversion strategy used here is based on the site-specific recombination system of the phage P1 (Sternberg et al., 1983; Sauer and Henderson, 1988). Several different specialized plasmids have been inserted into λ arms in such a way that the plasmid sequences are flanked by 34 bp direct repeats known as *loxP*. The *loxP* sites are the *cis*-acting genetic elements recognized by the P1 recombinase, known as Cre. Plasmids are automatically subcloned when the *loxP*-containing λ derivatives are infected into *E. coli* strains expressing Cre.

The plasmid portions of the λEXLX expression class of vectors are derivatives of the T7 expression system developed by Studier and colleagues (Studier and Moffat, 1986; Rosenberg et al., 1987; Lin et al., 1987). cDNA is directionally cloned downstream from the T7 RNA polymerase promoter between the N-terminal portion of the T7 gene *10* and a T7 transcription terminator. A large quantity of fusion protein is produced when cDNA plasmid clones are introduced into cells that express T7 RNA polymerase. Furthermore, these fusion proteins are often insoluble, allowing simple preparation for immunization (Lin et al., 1987). The original T7 expression vectors were based on pBR322 (Rosenberg et al., 1987). In contrast, our plasmids contain the pUC and the f1 origins of replication and they are expected to have a higher copy number and ss copies of the cDNA clones can easily be generated. It is thus relatively straightforward to select a cDNA clone from a λEXLX library, convert the clone to a plasmid by Cre-*loxP* automatic subcloning, sequence the insert by the dideoxy chain-termination method, alter the reading frame if necessary and induce the expression of large amounts of fusion protein.

At least two strategies have been proposed to facilitate subtractive hybridization procedures in situations where large amounts of RNA are difficult to obtain (Palazzolo and Meyerowitz, 1987; Pruitt, 1988). Both approaches involve the construction of cDNA libraries from the tissues of interest, followed by the purification of DNA from these libraries to generate ss DNA or cRNA copies that can be used in the subtraction procedures. The second class of vectors that we describe here, the λSHLX vectors, have been constructed to facilitate either strategy.

It is sometimes useful to build cDNA cloning vectors for specialized reasons. To simplify construction of other vectors, we have built our vectors in a modular fashion. Specifically, we have developed a set of 38-kb λ arms in which unique restriction sites have been positioned between two *loxP* sites. After cloning a specialized plasmid between the *loxP* sites, it is straightforward to convert it into a λ cDNA cloning vector with automatic plasmid subcloning capability.

RESULTS AND DISCUSSION

(a) Initial test of Cre-*loxP* automatic plasmid subcloning

Our initial experiments asked whether the Cre-*loxP* site-specific recombination system of phage P1 would be useful for automatic subcloning in λ-based cloning vectors. This was done by constructing a λgt10 (Huynh et al., 1985) derivative (λGLX) that contained plasmid sequences flanked by *loxP* sequences. To be useful for automatic subcloning, the *loxP* sites should recombine in a precise fashion to generate an autonomously replicating plasmid. Furthermore, the conversion from phage to plasmid should occur frequently enough to be useful for efficient library or clone conversion. The construction of λGLX is diagrammed in Fig. 1.

To test for conversion from phage to plasmid, *E. coli* cells that express the Cre recombinase were required. We generated a variety of P1 lysogens that express Cre; these are listed in Table I and discussed in section *ei*, below. In our initial tests, the first of these strains to be constructed, DM103[P1] (Table I), was infected with λGLX and plated on Ap plates. pGLX1 was found to be excised precisely, recircularized in vivo and replicated (Fig. 1). Detailed restriction mapping of the pGLX1 plasmid as well as nt sequence analysis of the pGLX1 region containing the *loxP* site, confirmed that pGLX1 was identical to pGLX2 except that it contained only a single *loxP* site.

(b) Universal λ arms containing *loxP* sites: construction of λLOX

Several characteristics are desirable for a set of λ arms that can accept any plasmid for Cre-*loxP* automatic subcloning. First, restriction sites that are to be placed in the multiple cloning site of the finished vector must be removed. Second, the acceptor arms must contain directly repeated *loxP* sites. Third, unique restriction sites must be positioned between the *loxP* sites so the arms can be opened for plasmid insertion. The construction of λLOX, an acceptor meeting these three criteria, is outlined in Figs. 2 and 3.

(c) Construction of the λEXLX T7 gene *10* expression vectors

The plasmid portions of the λEXLX vectors consist of: (1) the T7 elements of pET3xa (the T7 promoter, the coding sequences for the N-terminal end of the product of T7 gene *10* and the T7 transcription terminator) (Rosenberg

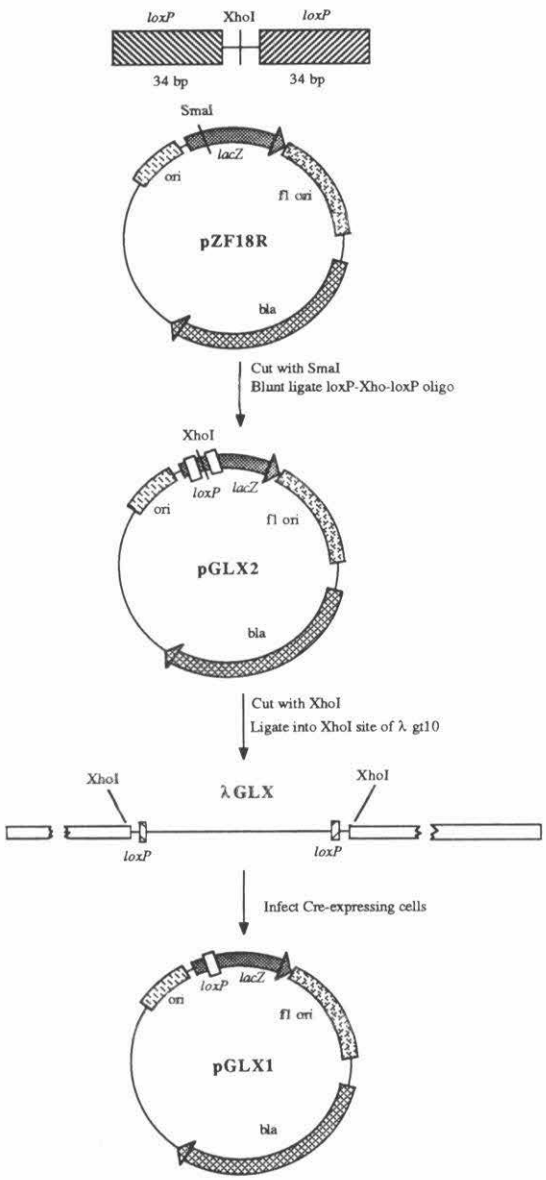


Fig. 1. Construction and test of λGLX. The *loxP* plasmid, called pGLX2, was generated first and then inserted into the λgt10 phage arms. pGLX2 was constructed by inserting two complementary oligos that contain a *Xho*I site flanked by directly repeated *loxP* sites into the *Sma*I site of pZF18R (Mead et al., 1986). pGLX2 thus contains two directly repeated *loxP* sites that are flanked by a variety of unique restriction sites and are separated by a unique *Xho*I site. This plasmid was then digested with *Xho*I and ligated into the *Xho*I site of λgt10 to generate λGLX. The phage contains an *Xho*I site flanked by *loxP* sites. A P1 lysogenic strain (DM103[P1]) was then infected with this phage and Ap^R plasmids, pGLX1, were recovered post-conversion and found to contain a single *loxP* site. The directions of transcription are represented by arrows.

et al., 1987); (2) the essential plasmid elements of pGEM3Zf (the pUC *ori*, the f1 *ori* and the *E. coli bla* gene) (Promega, Madison, WI); (3) a multiple cloning site that includes *Eco*RI, *Apa*I, *Hin*dIII and *Sac*I restriction sites; and (4) an SP6 RNA polymerase promoter. The construction of these plasmids, pMP3(+) and pMP3(−), that differ solely in the orientation of the f1 *ori*, is diagrammed in Fig. 4. These plasmids were linearized by digestion with *Sma*I and ligated between the *loxP* sites in the λLOX acceptor arms, completing the construction of λEXLX(+) and λEXLX(−) (Figs. 4 and 6A).

(d) Construction of the λSHLX subtractive hybridization vectors

The plasmid portions of the λSHLX vectors, pDD1 and pDD2, were designed to have the following features: (1) a multiple cloning site that allows the construction of directionally cloned cDNAs by a number of different strategies; (2) T7 and SP6 bacteriophage RNA polymerase promoters facing each other from opposite ends of the cloning sites; (3) *Not*I and *Sfi*I restriction sites that are placed just outside of the RNA polymerase promoters; and (4) the essential plasmid elements of pGEM3Zf (pUC *ori*, the f1 *ori* and the *E. coli bla* gene). Since these vectors are designed for the subtraction of one cDNA library from another, two vectors, λSHLX1 and λSHLX2, have been built; these differ solely in the relative orientations of the multiple cloning sites. A more detailed explanation of this last point is given in section **h**, below. The construction of the pDD1 and pDD2 plasmids is outlined in Fig. 5. These plasmids were linearized by digestion with *Hpa*I and separately inserted between the *loxP* sites in the λLOX arms to give the final vectors, λSHLX1 and λSHLX2 (Figs. 5 and 6B).

(e) Tests of the functional elements of the vectors

Each of the four cDNA cloning vectors contains a variety of functional elements. These include: *loxP* sites, the pUC *ori*, the f1 *ori*, the RNA polymerase promoters, the multiple cloning sites and a drug resistance marker. Each of these was tested in each of the constructs to ensure that they are intact and functional. In addition, test cDNA libraries have been constructed in both the λEXLX and λSHLX vectors.

*(i) Cre-*loxP *automatic subcloning*

Cre-*loxP*-mediated in vivo subcloning requires a specialized host-cell strain that produces the Cre protein. Conversion from phage to plasmid clones in vivo is accomplished by infecting such cells with phage and selecting for Ap^R colonies on agar plates. We have made several common cloning strains of *E. coli* lysogenic for a recombinant P1 phage ($P1Cm^Rr^-m^-$; Yarmolinsky et al., 1989) that carries a selectable marker (Cm^R) and lacks the P1

28

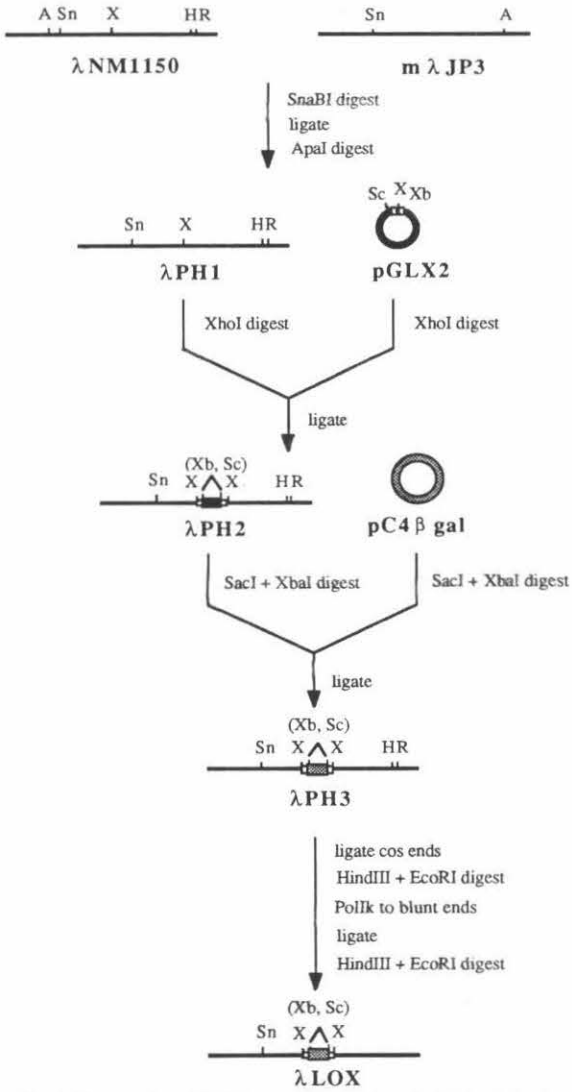


Fig. 2. Construction of λLOX, arms with *loxP* sites. λLOX is a derivative of λNM1150 (Murray, 1983; Pouwels et al., 1985) in which the *Apa*I, *Hin*dIII and *Eco*RI sites have been removed and two *loxP* sites have been inserted into the *Xho*I site. Phage mλJP3 (M. Strathmann and M. Simon, unpublished) is a derivative of λNM1150 in which the *Apa*I site in the left arm has been removed and a new *Apa*I site has been introduced in the right arm (M. Strathmann and M. Simon, pers. commun.). mλJP3 and λNM1150 DNAs were mixed and digested with *Sna*BI. The digested mixture was ligated and then digested with *Apa*I to eliminate the undesirable ligation products. DNA from the resulting phage (λPH1) was digested with *Xho*I and ligated to *Xho*I-digested pGLX2 DNA (see Fig. 1 for structure of pGLX2), to introduce *loxP* sites into the λ vector. DNA from this phage (λPH2) was digested with *Sac*I + *Xba*I to remove the plasmid sequences and ligated to an 850 bp stuffer fragment of SV40 DNA taken from pC4βgal (Thummel et al., 1988) to maintain minimum phage genome size for efficient packaging. The *cos* ends of the resulting phage (λPH3) DNA were ligated. To eliminate the *Eco*RI and *Hin*dIII sites, digestion was carried out with these two enzymes (hence deleting DNA between the sites), followed by end-filling with PolIk and religation. This was followed by a second *Hin*dIII + *Eco*RI digestion to reduce background. The resultant phage (λLOX) gave large, clear plaques, did not hybridize to plasmid DNA probe and did not contain restriction sites for *Apa*I, *Eco*RI or *Hin*dIII. After digestion of λLOX with *Xba*I + *Sac*I (and blunting the ends with T4 DNA polymerase if necessary), any linearized plasmid with appropriate ends can be ligated between the *loxP* sites. A detailed map of λLOX is given in Fig. 3. Restriction sites are marked: A, *Apa*I; H, *Hin*dIII; R, *Eco*RI; Sc, *Sac*I; Sn, *Sna*BI; X, *Xho*I; Xb, *Xba*I. Caret symbol, λLOX vectors with either orientation of the *Xba*I and *Sac*I sites are available. Figures are not to scale.

restriction-modification system (Table I). To test the frequency of conversion from phage to plasmid, both these P1 lysogens and their parental non-P1 lysogens were infected with each of the four vectors. The infected P1 lysogens were plated on agar plates containing Ap and the number of resistant colonies scored. The infected non-P1 lysogens were plated on nonselective medium and the resulting phage were counted. We define the apparent conversion frequency from phage to plasmid operationally as the ratio of the number of Ap^R colonies when plated on a P1 lysogen to the number of plaques when plated on its parental non-P1 lysogen. Representative results are shown in Table II; the apparent conversion frequency varied between 5 and 30%. Since infection of a λ lysogenic strain should favor the plasmid excision pathway over the lytic λ pathway, we tested the conversion frequency in *E. coli* strains that were lysogenic for both P1 and λ (see Table II for representative results). In the doubly lysogenic strains the conversion of phage to plasmid is more efficient; typical apparent conversion frequencies of 40 to 50% were observed for vectors alone, while for individual cDNA clones frequencies of from 80 to 140% were observed (Table II).

(ii) Propagation of automatically subcloned plasmids

Yields of purified automatically subcloned plasmid DNA from our P1 lysogens (Table I) were often an order of magnitude lower than yields we obtained when the same DNA was transformed into the parental non-Cre-expressing *E. coli* strains (data not shown). In addition, plasmids propagated with different efficiencies on the various host strains; for example, pEXLX(+) propagated well on strain BM25.5 (Table I) but pEXLX(−) and pSHLX1 and 2 propagated poorly. Our results suggest that there is a negative correlation between the presence of Cre recombinase in a cell and the propagation of *loxP*-containing plasmids, particularly when the plasmid sequences are present at high copy number. Similar poor plasmid DNA yields were previously observed in cells in which the gene that encodes Cre was cloned and introduced into *E. coli* as part of a λ lysogen (Sauer and Henderson, 1988; although not

TABLE I

Bacterial strains used to perform and test in vivo subcloning

Strain[a]	Based on (Ref.)[b]	Prophages[c] (Ref.)[b]	Additional properties[c]	Comments
DM103[P1]	JM103 (1)	P1Cm^{R}r$^-$m$^-$ (5)	F', Tc^S, *recA*$^+$	
BH23	KW251 (2)	*λimm*434*kan* (4)	F$^-$, Tc^R, *recA*$^+$, *mcrA*, *mcrB* (6)	
KW251[P1]	KW251 (2)	P1Cm^{R}r$^-$m$^-$ (5)	F$^-$, Tc^R, *recA*$^+$, *mcrA*, *mcrB* (6)	
BM25.5	KW251 (2)	*λimm*434*kan* (4), P1Cm^{R}r$^-$m$^-$ (5)	F$^-$, Tc^R, *recA*$^+$, *mcrA*, *mcrB* (6)	Robust colonies. Good DNA yields for pEXLX(+)
BH24	XL1-blue (3)	*λimm*434*kan* (4)	F', Tc^R, *recA*$^-$	
BH42	XL1-blue (3)	P1Cm^{R}r$^-$m$^-$ (5)	F', Tc^R, *recA*$^-$	
BM25.6	XL1-blue (3)	*λimm*434*kan* (4), P1Cm^{R}r$^-$m$^-$ (5)	F', Tc^R, *recA*$^-$	Minute colonies. Can mobilize plasmid to new host by two-step f1 rescue
BH28	JM101 (1)	*λimm*434*kan* (4)	F', Tc^S, *recA*$^+$	
BH82	JM101 (1)	P1Cm^{R}r$^-$m$^-$ (5)	F', Tc^S, *recA*$^+$	
BM25.8	JM101 (1)	*λimm*434*kan* (4), P1Cm^{R}r$^-$m$^-$ (5)	F', Tc^S, *recA*$^+$	Can mobilize plasmid to new host by one-step f1 rescue

[a] Lysogens for *λimm*434*kan* were obtained by infecting 0.5 to 1.5 ml of fresh saturated culture with 10^7 to 10^8 pfu for 20 min at 37°C. Infections were spread on plates containing 50 μg Km/ml to select lysogens. To construct P1Cm^{R}r$^-$m$^-$ lysogens, 0.05 to 0.5 ml of fresh saturated culture was made 5 to 20 μM for $CaCl_2$ and then infected with 10^5 to 10^7 pfu for 20 min at 37°C. P1 lysogens were selected on plates containing 30 μg Cm/ml. Strains lysogenic for both phage were constructed serially, by infecting the *λimm*434*kan* lysogen with P1 phage.

[b] (1) Yanisch-Perron et al. (1985); (2) Promega, Madison, WI (U.S.A.); (3) Bullock et al. (1987); (4) M. Strathmann and M. Simon (unpub.); (5) Yarmolinsky et al. (1989); (6) Raleigh and Wilson (1986).

[c] *kan*, gene encoding Km resistance; r$^-$m$^-$, mutant in both the restriction and modification functions of P1 (5); *mcr*, mutant in methylcytosine-specific restriction systems (6); F' markers: JM101 and 103 derivatives [F', *traD*36, *proAB*, *lacI*^{q}ZΔM15] (1); XL1-blue derivatives [F', *traD*36, *proAB*$^+$, *lacI*^{q}ZΔM15, Tn*10*, (Tc^R)] (3). For *λimm*434 see Fig. 3.

discussed in this paper, it appears from their Fig. 3 that plasmid DNA yields were reduced in their Cre-expressing strain). S. Elledge (pers. commun.) has constructed a series of Cre-*loxP* cDNA cloning vectors related to those described here; however, his automatically subcloned plasmids contain a pBR322 (rather than a pUC) *ori* and thus are not present at high copy number. In this case, he has found that there are good DNA yields from these plasmids when automatically subcloned in a λ-lysogenic host strain in which the *cre* gene is present in the lysogenized λ sequences. Should our reasoning be correct, then the production of Cre recombinase under control of an inducible promoter, rather than constitutively, should eliminate the depression of plasmid growth. We are currently exploring such possibilities.

In the meantime, we have devised a strategy by which the site specific recombination is conducted in doubly lysogenic cells (λ and P1) and the resulting plasmid is rapidly transferred to a second host strain that does not express the Cre recombinase. The cDNA vector is introduced into an *E. coli* strain that is a P1 and λ lysogen and also contains an F' episome (strain BM25.8; see Table I) making the cell susceptible to infection by phage M13. Simultaneously, the cells are infected with M13KO7 helper phage (Yanisch-

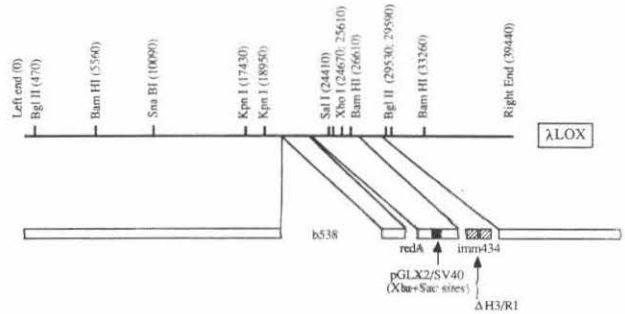


Fig. 3. Structure of λLOX. A detailed map of λLOX is shown, with the names and locations (± 10 bp) of the restriction sites above. Below are shown the locations of deletions and insertions relative to wt; open boxes, wt λDNA; blackened box, pGLX2/SV40 DNA that contains *Sac*I and *Xba*I sites *not* shown on the map of λLOX (see Fig. 2); hatched boxes, *c*I immunity (*imm*) region of phage 434. Since λLOX is derived from λNM1150, it too is Red$^-$ (see Pouwels et al., 1985 for a detailed map and genotype of λNM1150). ΔH3/RI, deletion of DNA between the *Eco*RI and *Hin*dIII sites (see Fig. 2 legend).

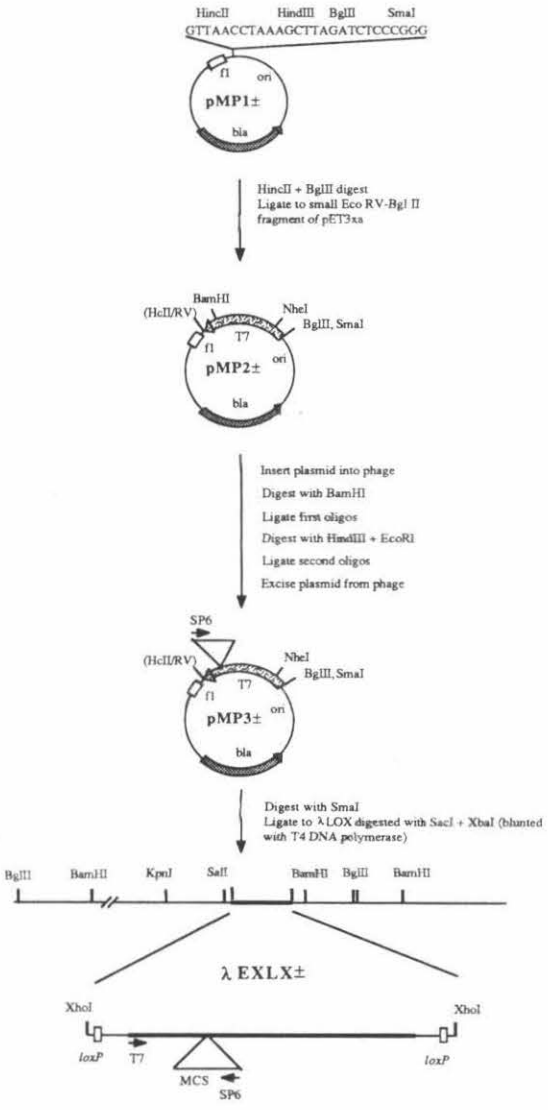


Fig. 4. Construction of T7 gene *10*-cDNA fusion protein expression vectors λEXLX. The expression plasmids in λEXLX(+) and λEXLX(−) contain a plasmid backbone derived from pGEM3Zf (Promega, Madison, WI), an inducible promoter and protein-coding region derived from pET3xa (Rosenberg et al., 1987) and a multiple cloning site and an SP6 RNA polymerase promoter (antisense to the protein-coding region) constructed from synthetic oligos. The two plasmids differ from each other in whether the plasmid backbone was derived from pGEM3Zf(+) in the case of λEXLX(+) or pGEM3Zf(−) in the case of λEXLX(−), and thus in the relative orientation of the f1 *ori*. The pMP1 plasmids were constructed as follows: pGEM3Zf plasmids were digested with *Pvu*II to remove the *lacZ* gene and cloning site, and were then ligated to a pair of complementary synthetic oligos containing *Hin*cII, *Hin*dIII, *Bgl*II and *Sma*I restriction sites. The resulting plasmids, pMP1(+) and pMP1(−), were digested with *Bgl*II + *Hin*cII and ligated to the small *Bgl*II-*Eco*RV fragment of pET3xa to create pMP2(+) and pMP2(−). This fragment of pET3xa contains the T7 gene *10* promoter, the coding sequence for the first 260 aa of the T7 gene *10* product, a *Bam*HI restriction site, and a T7 transcription terminator (Rosenberg et al., 1987). The pMP2(+) and pMP2(−) plasmids were inserted into phage mλJ (provided by M. Strathmann and M. Simon) by homologous recombination with a fragment of the *E. coli bla* gene in the mλJ phage. mλJ is a λNM1150 derivative that contains 650 bp of the *E. coli bla* gene inserted between the *Eco*RI and *Hin*dIII sites, and also lacks the *Apa*I site present in λNM1150. Integration was detected by growth on a λ-lysogen; this requires replication of the phage DNA off the plasmid *ori*. DNAs from the resulting phages, mλMP2(+) and mλMP2(−), were digested with *Bam*HI and ligated to a pair of oligos to insert *Hin*dIII, *Xba*I and *Eco*RI restriction sites and destroy the *Bam*HI site. The DNA from these phages, mλMP2P and mλMP2A, was digested with *Hin*dIII + *Eco*RI and ligated to a second pair of complementary oligos that introduced *Sac*I, *Hin*dIII, *Apa*I and SP6 RNA polymerase promoter sites, while destroying the original *Hin*dIII site. The plasmids, pMP3(+) and pMP3(−), were recovered in a λ-lysogenic strain [MP23, NM522 (Gough and Murray, 1983) with a λ*imm*434 lysogen. M. Strathmann and M. Simon, unpublished]. {Genotype: Δ(*lac-proAB*, *thi*$^-$, *hsd*Δ5, *sup*E, [F′, *proAB*, *lacI*^{q}, ZΔM15], λ*imm*434*kan*.} by homologous recombination between the partial *bla* genes, under selection for Ap^R. The pMP3(+) and pMP3(−) plasmids were digested with *Sma*I and inserted into the λLOX vector arms that had been digested with *Sac*I + *Xba*I and blunt-ended with T4 DNA polymerase, to create the final cloning vectors λEXLX(+) and λEXLX(−) (see also Fig. 6A). HcII, *Hin*cII; RV, *Eco*RV. The directions of transcription are represented by arrows.

Perron et al., 1985). In these cells, site-specific recombination occurs and the λ clone is converted into a plasmid clone, which is then replicated from the f1 *ori* to yield ss DNA that is packaged into infectious particles. Cells from a second host strain (XL1-blue; Bullock et al., 1987) that contains an F′ episome with a Tc^R gene are added to the mixture which is then plated on selective medium containing both Ap and Tc. After incubation, colonies appear that contain the converted plasmid but do not express the recombinase. Yields of plasmid DNA from these colonies are equivalent to those obtained for pGEM3Zf in non-Cre-expressing strains (not shown).

The above method for rescue of automatically subcloned plasmids occasionally results in deletion of part of the cDNA inserts, particularly when these are > 2 kb in length. In this case, it is useful to carry out the automatic plasmid subcloning, followed by purification of plasmid DNA from the Cre-expressing cells. Enough DNA can be obtained to transform a non-Cre-expressing strain (XL1-blue). Plasmids can then be purified from these cells with excellent yields and with their cDNA inserts intact (not shown).

(iii) Function of the f1 ori

We have tested the function of the f1 *ori* in all four vectors. Plasmids were recombined out of the phage λ vectors and transformed into *E. coli* cells containing an F′ episome (XL1-blue). These cells were then infected with the

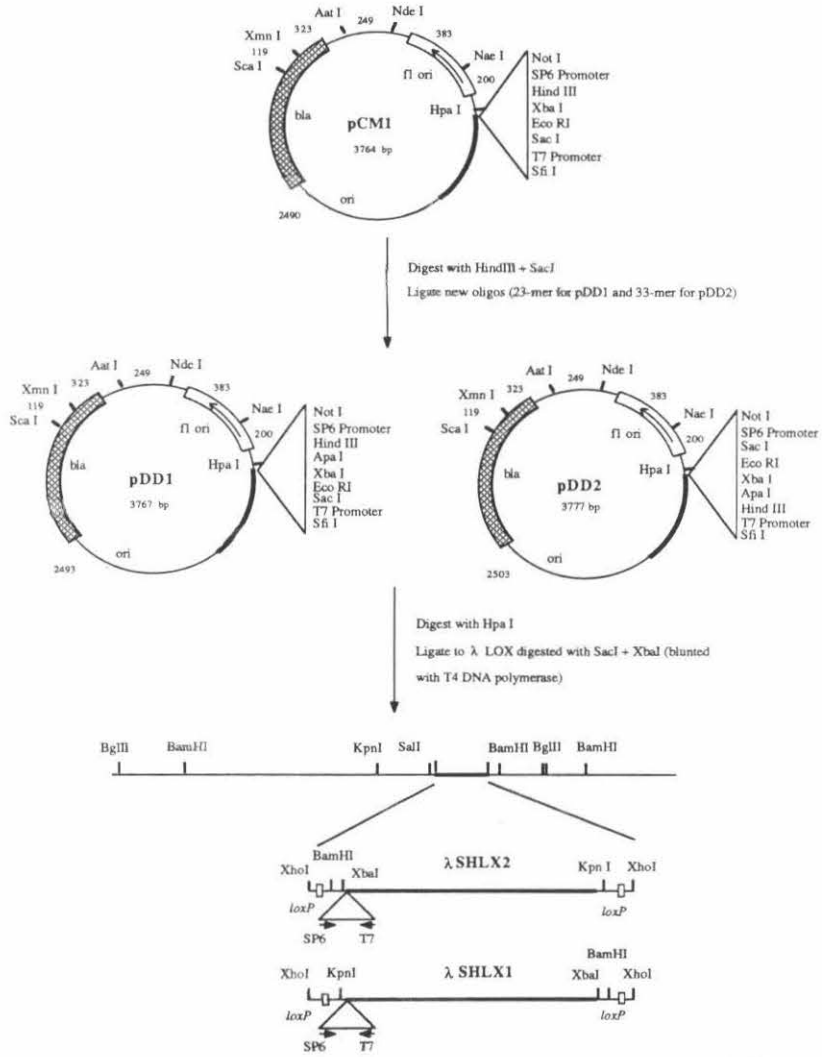


Fig. 5. Construction of the λSHLX vectors. pCM1 (C.H.M., C. Mayeda and M.J.P., unpub.) was chosen to build the λSHLX vectors because it already contained most of the desirable plasmid elements and only required alteration of the polylinker. pCM1 contains the pUC *ori*, the f1 *ori* and the Ap^R marker derived from the *Pvu*II fragment of pGEM3Zf(+). It also contains a 1-kb fragment from pATH (Dieckmann and Tzagoloff, 1985). This pATH fragment is vestigial from an earlier construction and has no relevance to the λSHLX vectors. To construct the λSHLX vectors, pCM1 DNA was digested with *Hin*dIII + *Sac*I. Two pairs of complementary oligos were separately ligated to generate new polylinkers with unique *Sac*I, *Eco*RI, *Xba*I, *Apa*I and *Hin*dIII restriction sites between the SP6 and T7 RNA polymerase promoters. The resulting plasmid was named *pDD1*. A second version, called pDD2, with the orientation of the multiple cloning sites reversed relative to pDD1 was also constructed. The pDD plasmids were linearized at the *Hpa*I site and ligated to λLOX DNA that had been digested with *Sac*I + *Xba*I and blunted with T4 DNA polymerase. The resultant phage vectors were named λSHLX1 and λSHLX2 (see also Fig. 6B). The *Xba*I site of λLOX was regenerated upon ligation of the p*DD* plasmids, thus is not unique to the multiple cloning site.

M13 helper phage K07 and grown overnight (Yanisch-Perron et al., 1985). Phage particles containing ss DNA were purified and the DNA was extracted. The yield of ss DNA was approximately the same as that from control pGEM3Zf plasmids (not shown).

(iv) Function of the SP6 and T7 RNA polymerase promoters

Chain-termination nt sequence analysis verified the presence of the consensus sequence for the SP6 and T7 RNA polymerase promoters in each of the vectors (Fig. 6). To test the function of these promoters, DNA prepared

32

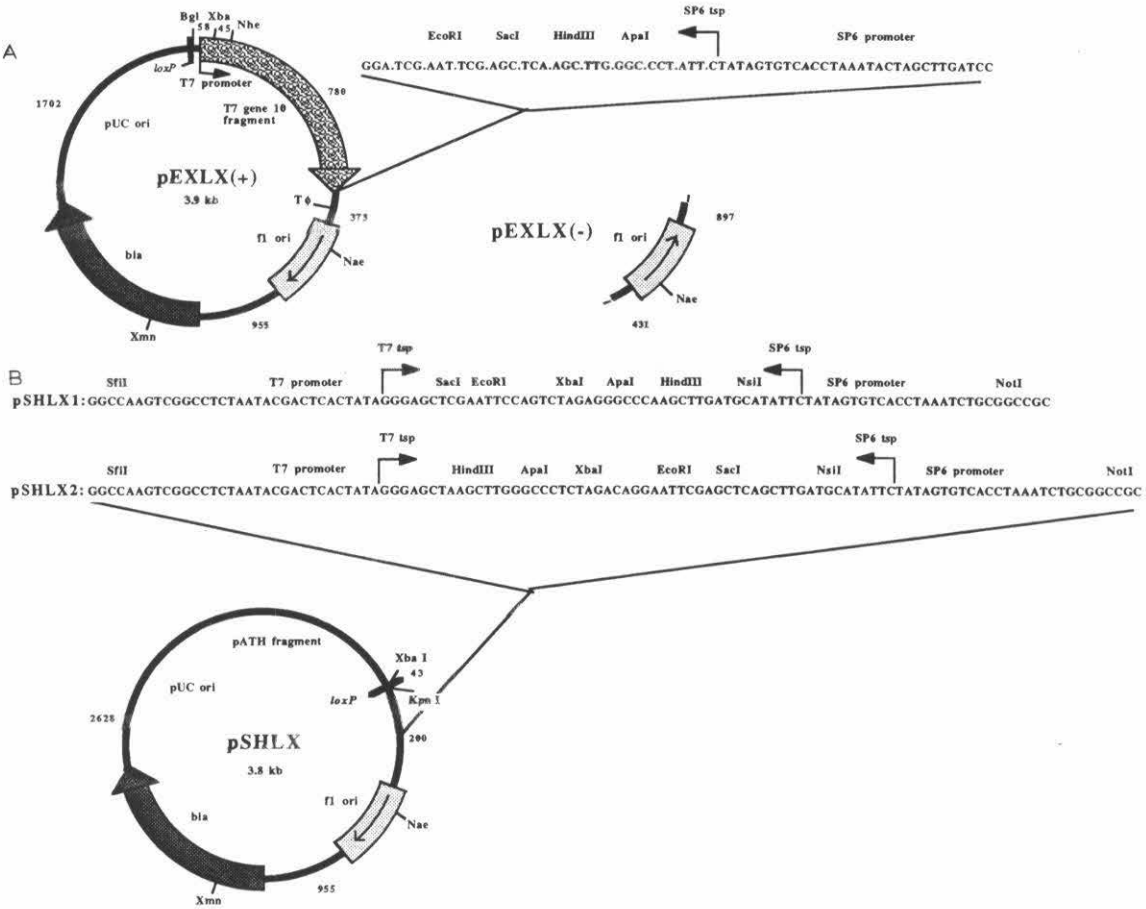


Fig. 6. Structures of the automatically subcloned plasmid portions of the (A) λEXLX and (B) λSHLX vectors. The pUC *ori*, β-lactamase gene (*bla*) and f1 *ori* are from pGEM3Zf(+) in pEXLX(+) and pSHLX, or from pGEM3Zf(−) in pEXLX(−). The T7 promoter, gene *10* protein coding fragment, and T7 transcription terminator in pEXLX plasmids are from pET3xa (Rosenberg et al., 1987). Multiple cloning sites and additional bacteriophage promoters were introduced with oligos as described in Figs. 2, 4 and 5. The directions of transcription and the orientation of the f1 *ori* are represented by arrows. The gene *10* reading frame in the pEXLX polylinker is indicated by the periods. Tϕ, T7 transcription terminator; *tsp*, transcription start point. Figures are not to scale; approximate distances between restriction sites are given in bp.

from each of the plasmids was linearized with an appropriate restriction enzyme and used as template in T7 and SP6 RNA polymerase reactions containing [α-^{32}P]CTP. The amount of cRNA synthesized was assayed by measuring the retention of radioactivity on DE-81 filters (Maniatis et al., 1982). These experiments indicated that all four plasmids contained SP6 and T7 promoters that functioned roughly as efficiently as they did in pGEM3Zf (not shown).

(v) Expression of T7 gene 10 *protein*

To test the ability of pEXLX(+) and pEXLX(−) to produce large amounts of the T7 gene *10* product in cells that express T7 RNA polymerase, the plasmids were introduced into a pLysS host strain (Moffat and Studier, 1987; Studier et al., 1990). This is a strain of *E. coli* that contains an IPTG-inducible T7 RNA polymerase gene on a λ lysogen. Since low levels of this polymerase are expressed even without IPTG-induction (Studier and Moffat, 1986), we were able to compare the expression of gene *10* protein from pEXLX and pET3xa simply by lysing uninduced overnight cultures in an SDS-urea buffer followed by SDS-polyacrylamide gel electrophoresis. Our results demonstrated that the amounts of the gene *10* protein produced by pEXLX(+) and pEXLX(−) are approximately equivalent to those from pET3xa (data not shown).

TABLE II

Apparent conversion frequencies of λEXLX(+) vector and λEXLX(+) cDNA clones in a P1 lysogen (KW251[P1]) and in a P1-λ double-lysogen (BM25.5)

Vector/cDNA[a]	Non-P1 lysogen[b]	Number of plaques[b]	P1 lysogen[b]	Number of colonies[b]	Conversion[c] (%)
λEXLX(+)	KW251	282	KW251[P1]	75	27
			BM25.5	117	41
λEXLX(+)	KW251	413	KW251[P1]	22	5
			BM25.5	190	46
λEXLX(+)cDNA1	KW251	315	BM25.5	372	118
λEXLX(+)cDNA2	KW251	171	BM25.5	220	129
λEXLX(+)cDNA3	KW251	171	BM25.5	144	84
λEXLX(+)cDNA4	KW251	110	BM25.5	140	127
λEXLX(+)cDNA5	KW251	364	BM25.5	332	91
λEXLX(+)cDNA6	KW251	140	BM25.5	148	106
λEXLX(+)cDNA7	KW251	392	BM25.5	396	101
λEXLX(+)cDNA8	KW251	91	BM25.5	128	141
λEXLX(+)cDNA9	KW251	290	BM25.5	0	0
λEXLX(+)cDNA10	KW251	170	BM25.5	212	125
λEXLX(+)cDNA11	KW251	104	BM25.5	88	85
λEXLX(+)cDNA12	KW251	349	BM25.5	344	99

[a] The results of two representative experiments with the λEXLX(+) vector alone are shown. The random cDNA clones were isolated from the *Drosophila* λEXLX(+) cDNA library described in section *evii*.

[b] In each experiment, equal volumes of appropriate phage suspension were used to infect equal volumes of saturated cultures of each *E. coli* strain. Infections were carried out for 20 min at 37°C. Infections for phage titer were plated in NZCYM top agarose. Infections for plasmid conversion were spread on plates containing Ap or carbenicillin (with or without Km and Tc; the presence of these two drugs did not influence the number of colonies). That these colonies resulted from plasmid excision was verified by growing cells from several individual colonies overnight in liquid culture, preparing rapid plasmid DNA preparations and analyzing the restriction fragments generated by digestion with diagnostic restriction enzymes (*Hin*fI for the vector alone, *Apa*I + *Sac*I for cDNA clones). In all cases tested, appropriate restriction patterns were observed (data not shown). cDNA 9 did not convert upon retesting; however, it appears to be unique in that we have seen no other such nonconverting clones in several subsequent experiments. Similar apparent conversion frequencies were found for λSHLX vectors and cDNA clones as well as for other lysogenic host strains (data not shown). In all cases, fresh cultures of BM25 cells gave better apparent conversion frequencies than did either old cultures or cultures grown beyond saturation.

[c] Apparent conversion frequency is defined in section *ei* and is listed here as a percentage.

(vi) Construction of a cDNA library in λSHLX

To test the ability of the λSHLX vectors to accept cDNA molecules, a library was constructed in λSHLX2. Poly(A)$^+$ RNA was isolated from a mixed-stage liquid culture of the N2 strain of *Caenorhabditis elegans*. cDNA was prepared according to a method developed by M. Strathmann and M. Simon (pers. commun.) in which the cDNA can be directionally cloned with all of the internal sites protected (see section **g**, below for details). In the construction of our library, 5 μg of poly(A)$^+$ RNA was converted into ds cDNA. This material was then fractionated on a 1% agarose gel and four individual size fractions were recovered: 0.5–1.0 kb (approx. 420 ng recovered), 1–2 kb (450 ng), 2–3 kb (330 ng) and greater than 3 kb (360 ng). Material from each of these fractions was then ligated into *Apa*I + *Sac*I-digested λSHLX2 and the cloning mixture was packaged. Yields were between 10^6 and 10^7 pfu (e.g., the 1–2-kb material yielded 7.5×10^6 pfu).

When the 2–3-kb fraction library was screened at low density with probe synthesized from the *C. elegans* poly(A)$^+$ RNA, almost half the clones gave positive autoradiographic signals. Over 100 individual plaques that did not give a detectable hybridization signal were isolated and screened by the PCR using oligo primers that flank the cloning sites (Saiki et al., 1988). Over 48% contained inserts longer than 1 kb, while 26% had shorter inserts, and 26% had no detectable insert by this assay. The selection of nonhybridizing plaques biases towards clones that are relatively rare in the mRNA population, or that have no inserts (Palazzolo and Meyerowitz, 1987; Palazzolo et al., 1989). Of these λSHLX2 cDNA clones from the 2–3-kb fraction, 22 were converted to plasmids by the method described in section *eii*. Restriction mapping of DNA prepared from the individual clones showed *Apa*I-*Sac*I inserts that matched the size of the inserts detected by the PCR test for 20 of the 22 selected clones. The remaining two converted clones showed, in one case, a significantly smaller insert (0.57 kb vs. 1.75 kb by PCR) with a normal vector band and, in the other case, aberrant insert and vector bands. The reason for these discrepancies has not been determined. A set of five additional clones, selected on the

basis of showing a hybridization signal with cDNA probe derived from *C. elegans* poly(A)$^+$ RNA, and thus likely to contain cDNAs representing relatively abundant RNAs, were analyzed by chain-termination nt sequence analysis. One of the clones was found to be identical to the nematode *hsp70A* gene (Snutch et al., 1988) over the 187 nt of sequence that was determined. A second clone showed a 74 out of 78 aa match to the yeast elongation factor 1-α A chain predicted protein sequence (Nagashima et al., 1986).

(vii) Construction of cDNA libraries in λEXLX

We have constructed *Drosophila* cDNA libraries in the λEXLX vectors. cDNA synthesized from poly(A)$^+$ RNA purified from 0–24-h *Drosophila* embryos was cloned into λEXLX(+) and cDNA synthesized from poly(A)$^+$ RNA purified from adult *Drosophila* was cloned into λEXLX(−) using the Strathmann–Simon method outlined in section **g** below. The cloning gave between 1.2×10^6 and 4.6×10^7 clones/μg poly(A)$^+$ RNA. Each library was shown to contain cDNA inserts by λ plaque filter hybridization screens with ^{32}P-labelled cDNA synthesized from the same source of RNA that was used in its construction. After a 2-h exposure, about 40% of the clones gave positive hybridization signals in this screen. These libraries have been screened with probes generated from selected clones isolated by Palazzolo et al. (1989) from a subtracted *Drosophila* library in which adult-head cDNA was subtracted with 0–1-h embryo RNA. These probes identified a small number of homologous clones that were plaque-purified, automatically subcloned and are currently under detailed analysis.

(f) Applications of λLOX

Almost any plasmid can be linearized and inserted between the *loxP* sites of λLOX to generate a specialized cloning vector with Cre-*loxP* automatic subcloning. Here we have constructed and tested two sets of vectors: one for expression of cDNA clones in *E. coli* and a second for use in subtractive hybridization procedures. However, the λLOX arms can be used to construct specialized vectors containing other types of plasmids, such as shuttle vectors for yeast or different types of eukaryotic expression vectors. To use λLOX for these applications it is important first to introduce the appropriate cloning sites (such as *Hin*dIII, *Eco*RI, *Apa*I, and *Sac*I) into the plasmid of interest. This plasmid can then be linearized with a different restriction enzyme and ligated into the *Xba*I and *Sac*I sites (blunt-ended with T4 DNA polymerase) of λLOX.

(g) Methods for directional cDNA library construction

At least two strategies have been devised recently to facilitate directional cDNA cloning with the protection of internal endogenous restriction sites. Meissner et al. (1987) synthesized ds cDNA that was primed initially with oligo(dT). This was then methylated with M·*Bam*HI + M·*Alu*I methyltransferases to protect internal *Bam*HI and *Hin*dIII restriction sites. A specialized linker was then blunt-end ligated to this cDNA; this linker contains a *Bam*HI site flanked by the last 4 nt of the *Hin*dIII recognition site. When ligated to the cDNA, the first 2 nt of the *Hin*dIII site were generated at the 3′ end by the poly(dA) sequences and a *Bam*HI site was positioned at the 5′ end of the cDNA. In the case of λEXLX and λSHLX, one would ligate a linker containing an *Eco*RI site flanked by the last 4 nt of the *Hin*dIII recognition site. The cDNA would then be digested with *Eco*RI + *Hin*dIII and cloned into the unique *Eco*RI and *Hin*dIII sites in any of the four vectors that we have constructed. We have successfully constructed 0–1-h *Drosophila* embryo cDNA libraries in λSHLX2 using this strategy.

The second protocol was devised by M. Strathmann and M. Simon (pers. commun.). Briefly, an *Apa*I site is generated at the 3′-end of the cDNA by priming first strand cDNA synthesis with an oligo(dT)-*Apa*I primer-adapter. First strand cDNA synthesis is carried out using 5-methyl dCTP in place of dCTP, while second strand synthesis utilizes unmethylated dCTP. *Sac*I linkers are ligated to the ds cDNA and the cDNA is digested with *Apa*I + *Sac*I. All internal *Apa*I + *Sac*I sites are hemimethylated and cannot be cleaved by *Apa*I or *Sac*I, while the restriction sites at the ends can be cleaved because they are unmethylated.

(h) Applications of λEXLX(+) and λEXLX(−) vectors

It is often desirable to obtain clones in which the cDNA of interest is inserted into an expression vector so that the protein encoded by the cDNA can easily be purified for use as an immunogen. The plasmid expression vectors developed by Studier and colleagues (Rosenberg et al., 1987) offer a number of attractive features. Expression is completely silent in cells that do not contain the T7 RNA polymerase gene and is abundant in cells that do (Studier and Moffat, 1986). Thus, control of gene expression is very tight, limiting problems that occur when the foreign product is lethal in *E. coli*. In addition, their vectors provide termination of transcription as well as initiation. Finally, the fusion protein is often insoluble (Lin et al., 1987), allowing a rapid and simple one-step purification. Our λEXLX(+) and λEXLX(−) vectors are λ-based versions of the pET expression vectors that additionally feature directional cDNA cloning with the protection of all the internal sites, Cre-*loxP* automatic subcloning and the ability to produce ss DNA by induction of an f1 *ori*. Recently, we have generated a monoclonal antibody to the T7 gene *10* protein product (C. Mayeda and M.J.P., unpublished) and are currently pursuing the use of this antibody in the immunopurification of expressed gene *10* fusion proteins.

The use of the λEXLX vectors for gene *10* fusion protein expression requires that the gene *10* and cDNA ORFs be in the same frame. In previous expression vectors (Young and Davis, 1983) it was important to ensure that the cDNA clone was inserted into the vector in the appropriate reading frame. In the case of the λEXLX vectors, ss DNA (purified as outlined in sections e*ii* and e*iii*) can be used to alter the reading frame of any individual clone by a variety of in vitro mutagenesis methods (Taylor et al., 1985; Kunkel et al., 1986) that involve priming of DNA synthesis with oligos containing small mismatches designed either to insert or to delete nt.

(i) Applications of the λSHLX vectors

Two methods have been developed for using cDNA libraries to generate the large amounts of the ss DNA or cRNA required for subtractive hybridization procedures (Palazzolo and Meyerowitz, 1987; Pruitt, 1988). The λSHLX vectors have been designed to facilitate either strategy.

One strategy follows the hybridization selection methods of Pruitt (1988). Briefly, a cDNA library representing sample X is prepared in λSHLX1, and one representing Y is constructed in λSHLX2. Each library is then amplified and aliquots are converted to ss DNA as outlined in sections e*ii* and e*iii*. Since the multiple cloning sites in the SHLX1 and SHLX2 vectors are in opposite orientations, the ss DNA recovered from the converted pSHLX1 library will represent the 'sense' orientation and that from the pSHLX2 library the 'antisense' orientation. The ss Y-DNA can then be covalently attached to an insoluble substrate and the X-DNA can be hybridized to an excess of the Y-DNA. The soluble fraction, that should be enriched in X-specific sequences, can then be transformed into *E. coli* to give an 'X – Y' plasmid library (Pruitt, 1988).

In the second strategy (modified from Palazzolo and Meyerowitz, 1987; Palazzolo et al., 1989), X- and Y-cRNA can be synthesized in vitro using as template either plasmid or phage DNA prepared from each library. Specifically, DNA from the X-library in automatically subcloned pSHLX1 can be linearized at the *Not*I site (since this enzyme has an 8-bp recognition sequence, this reduces the probability of digestion within the cloned cDNA sequences) and then used as template for a T7 RNA polymerase reaction to make 'sense' X-cRNA. Similarly, DNA from the Y-library in pSHLX2 can be linearized at the *Sfi*I site and used as template for a SP6 RNA polymerase reaction to make 'sense' Y-cRNA. The X-cRNA can be reverse-transcribed to make 'antisense' cDNA; this is accomplished by priming cDNA synthesis with an oligo complementary to the SP6 promoter sequences present in the X-cRNA (this is preferable to the use of an oligo comprised partly of the *Not*I recognition sequence since the latter is G + C-rich). This 'antisense' X-cDNA can be hybridized to an excess of 'sense' Y-cRNA. After subtractive hybridization the 'X – Y'-cDNA sequences can be made ds and recloned.

ACKNOWLEDGEMENTS

We thank F.W. Studier (Brookhaven National Laboratory) for providing vectors and host strains, M. Strathmann (Caltech) and S. Elledge (Baylor College of Medicine) for providing vectors, host strains and advice, and S. Lewis for assistance with DNA sequencing. Technical assistance was provided by C. Mayeda and M. Whitney. C.H.M. thanks Martin Chalfie for providing laboratory space. M.J.P. was a postdoctoral research fellow of the Helen Hay Whitney Foundation and is a Lucille P. Markey Scholar; B.A.H. is supported in part by a USPHS Predoctoral National Research Service Award (T32 GM07616); D.D. is supported in part by a predoctoral fellowship from the California Foundation for Biochemical Research; C.H.M. is supported by a postdoctoral fellowship in Molecular Studies of Evolution from the Alfred P. Sloane Foundation and USPHS grant GM30997 to Martin Chalfie; K.V.R. was supported by a Proctor & Gamble postdoctoral fellowship and a Rockefeller Foundation Biotechnology Career Fellowship. This research was supported by a scholar's grant from the Lucille P. Markey Charitable Trust (M.J.P.), USPHS Program Project Grant GM40499 (H.D.L. and E.M.M.), Developmental Biology Grants from the Lucille P. Markey Charitable Trust (H.D.L. and E.M.M.) and the Searle Scholars Program of the Chicago Community Trust (H.D.L.).

REFERENCES

Brenner, S., Cesarini, G. and Karn, J.: Phasmids: hybrids between ColE1 plasmids and *E. coli* bacteriophage lambda. Gene 17 (1982) 27–44.

Bullock, W.O., Fernandez, J.M. and Short, J.M.: XL-1 blue: a high efficiency plasmid transforming *recA Escherichia coli* strain with beta-galactosidase selection. Biotechniques 5 (1987) 376–379.

Dieckmann, C.L. and Tzagoloff, A.: Assembly of the mitochondrial membrane system: *CBP6*, a yeast nuclear gene necessary for synthesis of cytochrome *b*. J. Biol. Chem. 260 (1985) 1513–1520.

Gough, J.A. and Murray, N.E. (1983) Sequence diversity among related genes for recognition of specific targets in DNA molecules. J. Mol. Biol. 166 (1983) 1–20.

Huynh, T.V., Young, R.A. and Davis, R.W.: Constructing and screening cDNA libraries in λgt10 and λgt11. In Glover, D.M. (Ed.), DNA Cloning, Vol. 1. *IRL* Press, Oxford, 1985, pp. 49–78.

Kunkel, T.A., Roberts, J.D. and Zakour, R.A.: Rapid and efficient site-specific mutagenesis without phenotypic selection. Methods Enzymol. 154 (1986) 367–382.

Lin, S., Dunn, J.J., Studier, F.W. and Stafford, D.W.: Expression of human factor IX and its subfragments in *Escherichia coli* and gener-

ation of antibodies to the subfragments. Biochemistry 26 (1987) 5267–5274.

Maniatis, T., Fritsch, E.F. and Sambrook, J.: Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982.

Mead, D.A., Szczesna-Skorupa, E. and Kemper, B.: Single-stranded DNA 'blue' T7 promoter plasmids: a versatile tandem promoter system for cloning and protein engineering. Protein Engin. 1 (1986) 67–74.

Meissner, P.S., Sisk, W.P. and Berman, M.L.: Bacteriophage λ cloning system for the construction of directional cDNA libraries. Proc. Natl. Acad. Sci. USA 84 (1987) 4171–4175.

Moffatt, B.A. and Studier, F.W.: T7 lysozyme inhibits transcription by T7 RNA polymerase. Cell 49 (1987) 221–227.

Murray, N.: Lambda vectors. In Hendrix, R.W., Roberts, J.W., Stahl, F.W. and Weisberg, R.A. (Eds.), Lambda II. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1983, pp. 677–684.

Nagashima, K., Akasai, M., Nagata, S. and Kaziro, Y.: Structure of the two genes coding for polypeptide chain elongation factor 1-alpha (EF-1-alpha) from *Saccharomyces cerevisiae*. Gene 45 (1986) 265–273.

Palazzolo, M.J. and Meyerowitz, E.M.: A family of lambda phage cDNA cloning vectors, λSWAJ, allowing the amplification of RNA sequences. Gene 52 (1987) 197–206.

Palazzolo, M.J., Hyde, D., Raghavan, K.V., Mecklenberg, K., Benzer, S. and Meyerowitz, E.M.: Isolation and characterization of 436 *Drosophila* cDNA clones corresponding to RNAs expressed in the adult head but not detected in the early embryo using a new cloning strategy. Neuron 3 (1989) 527–539.

Pouwels, P.H., Enger-Valk, B.E. and Brammar, W.J.: Cloning Vectors. A Laboratory Manual. Elsevier, Amsterdam, 1985, p. 1-A-vi-11.

Pruitt, S.C.: Expression vectors permitting cDNA cloning and enrichment for specific sequences by hybridization selection. Gene 66 (1988) 121–134.

Raleigh, E.A. and Wilson, G.: *Escherichia coli* K-12 restricts DNA containing 5-methylcytosine. Proc. Natl. Acad. Sci. USA 83 (1986) 9070–9074.

Rosenberg, A.H., Lade, B.N., Chui, D., Lin, S., Dunn, J.J. and Studier, F.W.: Vectors for selective expression of cloned DNAs by T7 RNA polymerase. Gene 56 (1987) 125–135.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A.: Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239 (1988) 487–491.

Sauer, B. and Henderson, N.: The cyclization of linear DNA in *Escherichia coli* by site-specific recombination. Gene 70 (1988) 331–341.

Short, J.M., Fernandez, J.M., Sorge, J.A. and Huse, W.D.: λZAP: a bacteriophage λ expression vector with in vivo excision properties. Nucleic Acids Res. 16 (1988) 7583–7600.

Snutch, T.P., Heschl, M.F.P. and Baillie, D.L.: The *Caenorhabditis elegans hsp70* gene family: a molecular genetic characterization. Gene 64 (1988) 241–256.

Sternberg, N., Hoess, R. and Abremski, K.: The P1 *lox*-Cre site specific recombination system: properties of *lox* sites and biochemistry of *lox*-Cre interactions. In Cozarelli, N.R. (Ed.), Mechanisms of DNA Replication and Recombination. UCLA Symposia on Molecular and Cellular Biology. Vol. 10. Liss, New York, 1983, pp. 671–684.

Studier, F.W. and Moffat, B.: Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. 189 (1986) 113–130.

Studier, F.W., Rosenberg, A.H. and Dunn, J.J.: Use of T7 RNA polymerase to direct the expression of cloned genes. Methods Enzymol. 185 (1990) in press.

Taylor, J.W., Ott, J. and Eckstein, F.: The rapid generation of oligodeoxyribonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA. Nucleic Acids Res. 13 (1985) 8764–8785.

Thummel, C.S., Boulet, A.M. and Lipshitz, H.D.: Vectors for *Drosophila* P-element-mediated transformation and tissue culture transfection. Gene 74 (1988) 445–456.

Yanisch-Perron, C., Vieira, J. and Messing, J.: Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33 (1985) 103–119.

Yarmolinsky, M.B., Hansan, E.B., Jafri, S. and Chattoraj, D.K.: Participation of the lytic replicon in P1 plasmid maintenance. J. Bacteriol. 171 (1989) 4785–4791.

Young, R.A. and Davis, R.W.: Efficient isolation of genes by using antibody probes. Proc. Natl. Acad. Sci. USA 80 (1983) 1194–1198.

§2.3: Rapid Isolation of long cDNA clones from existing libraries

Nucleic Acids Research, Vol. 19, No. 8 **1951**

Rapid isolation of long cDNA clones from existing libraries

Bruce A.Hamilton, Michael J.Palazzolo+ and Elliot M.Meyerowitz*
Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA

Submitted February 5, 1991

Obtaining full-length or even near full-length cDNA clones has been a time-consuming and labor-intensive step in the analysis of cloned genes. Recent methods which use PCR as a preparative tool for cloning (1, 2) have made this step considerably more rapid, but may introduce sequence errors in the resulting clones due to numerous sequential rounds of in vitro replication. We describe a method for identifying long cDNA clones from existing cDNA libraries using PCR purely as a diagnostic tool.

The method is shown schematically in Figure 1. Sequence information from the 5′ end of an existing clone is used to design a gene-specific PCR primer. This is used in combination with a primer that abuts the 5′ cloning site in the vector to amplify a directionally cloned cDNA insert and then to reamplify the insert in the presence of ^{32}P and the absence of significant vector sequences to generate a radiolabeled 5′ end probe for screening libraries by plaque filter hybridization (3). For efficiency, we screen size-selected cDNA libraries at very high phage densities (200,000 to 1,000,000 pfu per 150 mm diameter plate). This density allows an entire library to be screened on relatively few plates. Crude phage plugs are picked into SM buffer. Each plug lysate contains a few hundred to a few thousand phage clones, one of which corresponds to the probe. Each lysate is then subjected to PCR between the gene-specific primer and a 5′ vector primer. The size of the amplification product shows the distance from the gene-specific primer to the 5′ end of the desired cDNA clone from each crude plug. Only cDNA clones of a desired length are purified and analyzed further. The principle limitations in this strategy are the quality of available libraries and the size range of reliable PCR amplification.

Results from two typical experiments are shown in Figure 2. 1C12 and 2F8 are clones from a large collection of cDNAs that represent RNAs expressed in the head but not in the early embryo of *D. melanogaster* (4). The initial clone of 1C12 is 1 kb in length

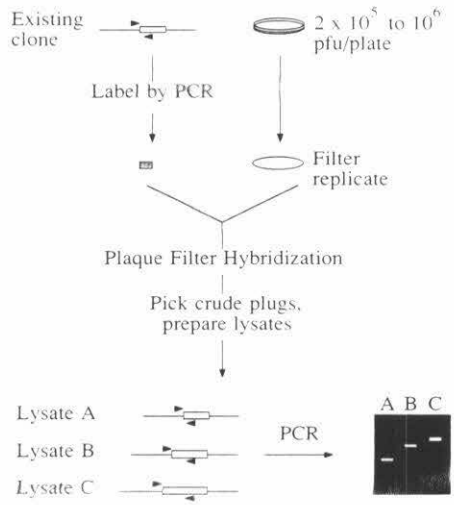


Figure 1. Flow diagram of screening procedure. PCR primers are represented as arrowheads. In each lysate only one phage clone, shown schematically, should contain sequences corresponding to the gene-specific primer. See text for full description. Detailed protocols available on request.

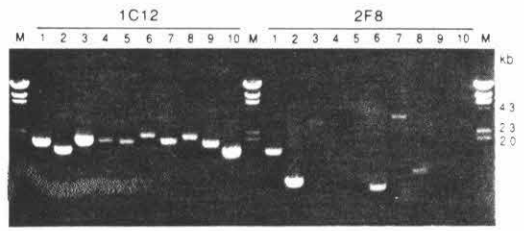


Figure 2. PCR products from two typical experiments. *Drosophila* adult head cDNA libraries in λEXLX (5), size-selected for cDNAs larger than 1.2 kb and 2.0 kb, respectively, were plated at 250,000 pfu/150 mm plate. Filter replicates were hybridized to ^{32}P-labeled cDNA inserts ($\sim 5\times10^9$ cpm/μg) from λSWAJ3 clones representing RNAs expressed in the adult head (4). Hybridization conditions were 50% formamide/5×SSPE/1×Denhardt's solution/1% SDS/100 μg ml^{-1} salmon sperm DNA at 42°C. Probe was added to $1-3\times10^6$ cpm/ml after 45 min. incubation in this buffer. Filters were washed 3 times in 0.1×SSPE/0.3% SDS at 50–55°C for 15–20 min. Plugs of phage corresponding to autoradiographic signals were picked with a baked Pasteur pipette or a broken 1 ml disposable serological pipette (outer diameter: 5 mm) into 200–300 μl SM buffer and allowed to resuspend for several hours. PCR was performed on 1 μl of each lysate in 40 μl reaction volume. Primers were used at 250 nM each. Amplifications were done with 40 cycles of 94°C, 45 sec.; 55°C, 45 sec.; 72°C, 3 min. PCR primers: SWAJ3.5, 5′-ATTTAGGTGACACTATAGAATACAC-3′; AG.5 (λEXLX), 5′-GCTGGTACCGGATCGAATTC-3′; 1C12, 5′- ATATTCCATGCCGTTTCGTG-3′; 2F8, 5′-CTTTGT*CTC*GTCATGCGTTG-3′.

* To whom correspondence should be addressed

+Present address: Department of Genetics, Washington University School of Medicine, St Louis, MO 63110, USA

and hybridizes to a single poly(A)$^+$ RNA of approximately 5200 nt. Ten phage plug lysates from size-selected cDNA libraries were analyzed by PCR. The longest PCR product, from lysate 6, is 2.1 kb. The full length of the corresponding cDNA clone is 3.0 kb. The original 2F8 clone is 0.4 kb and hybridizes to several poly(A)$^+$ RNAs, ranging from 2500 to 6700 nt. The amplification product from lysate 7 is 3.4 kb and corresponds to a cDNA clone of 3.7 kb. The longest cDNA clone identified for each probe was subsequently plaque-purified, converted to plasmid in vivo (5), restriction mapped, and partially sequenced to verify its identity.

We have obtained comparable results with many other cDNA probes and with phage densities as high as 2×10^6 pfu per plate. Exposures of 24 hr or longer with an intensifying screen are usually necessary to detect hybridization signals on filters containing 10^6 or more pfu. While in our hands the efficiency of detection (hybridization signals/phage density) decreases with increasing phage density above 10^5 pfu per plate, the efficiency of screening (hybridization signals/filter) increases.

Abbreviations: kb, kilobase pairs; nt, nucleotide; PCR, polymerase chain reaction; pfu, plaque forming unit.

ACKNOWLEDGEMENTS

We thank M. Strathmann, K.Vijay Raghavan, and T.Wilke for constructive discussions and C.Mayeda and M.Whitney for excellent technical assistance. B.A.H. was supported in part by a USPHS Predoctoral National Research Service Award (T32 GM07616). This work was supported by a scholar's grant from the Lucille P.Markey Charitable Trust (M.J.P.) and USPHS Program Project Grant GM40499 (E.M.M.).

REFERENCES

1. Frohman,M.A., Dush,M.K. and Martin,G.R. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 8998–9002.
2. Loh,E.Y., Elliott,J.F., Cwirla,S., Lanier,L.L. and Davis,M.M. (1989) *Science* **243**, 217–220.
3. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*, Second Edition. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
4. Palazzolo,M.J., Hyde,D., Raghavan,K.V., Mecklenberg,K., Benzer,S. and Meyerowitz,E.M. (1989) *Neuron* **3**, 527–539.
5. Palazzolo,M.J., Hamilton,B.A., Ding,D., Martin,C.M., Mead,D.A., Mierendorf,R.C., Raghavan,K.V., Meyerowitz,E.M. and Lipshitz,H.D. (1990) *Gene* **88**, 25–36.

§2.4: Transposon-facilitated DNA sequencing

Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 1247–1250, February 1991
Genetics

Transposon-facilitated DNA sequencing

($\gamma\delta$/polymerase chain reaction)

MICHAEL STRATHMANN, BRUCE A. HAMILTON, CAROL A. MAYEDA, MELVIN I. SIMON, ELLIOT M. MEYEROWITZ, AND MICHAEL J. PALAZZOLO

California Institute of Technology, Division of Biology, 147-75, Pasadena, CA 91125

Communicated by E. B. Lewis, September 26, 1990

ABSTRACT **We describe here a transposon-based DNA sequencing strategy that allows the introduction of sequencing priming sites throughout a target sequence by bacterial mating. A miniplasmid was designed to select against transposon insertions into the vector. Sites of transposon insertion are mapped by the polymerase chain reaction with bacterial overnight cultures providing the templates. A small set of plasmids with transposons spaced several hundred base pairs apart can then be sequenced. Sequencing primers corresponding to the transposon ends allow sequencing in both directions. Thus, the entire sequence of both strands can be easily determined.**

One of the major problems in DNA sequence analysis of large or even moderately sized fragments is how to position unsequenced regions next to known priming sites. A variety of techniques have been developed for this purpose including random shotgun subcloning, unidirectional deletions and subcloning, and the continued synthesis of additional oligodeoxynucleotide primers (1–4). These methods are expensive or require many molecular manipulations.

A number of strategies employ bacterial transposons to generate priming sites within a target DNA sequence (5–10). Several criteria exist for an efficient transposon-based sequencing strategy: (*i*) Mobilization of the transposon must be relatively simple. (*ii*) Selection for transposon insertions into the plasmid as opposed to the bacterial chromosome must be efficient. (*iii*) The transposon must insert into the target sequence and not into the plasmid vector. (*iv*) The transposition sites must be easily mapped to minimize the number of required sequencing reactions. In this paper we describe a transposon-based strategy that meets these criteria.

We employ $\gamma\delta$, which belongs to the Tn*3* family of transposons (11) and which has been used previously in transposon-facilitated strategies (8, 20). The members of this family contain 38-base-pair (bp) terminal inverted repeats and transpose by a replicative mechanism. Donor and target sequences are joined in an intermediate structure termed a cointegrate. The cointegrate, which contains two copies of the transposon, is rapidly resolved by a site-specific recombination system. The resolvase is encoded by the transposon and acts at the 120-bp *res* site located within the mobile element.

$\gamma\delta$ is present on the F factor. Consequently, transposition to a plasmid transiently fuses the F factor and plasmid in a cointegrated structure. This cointegrate can be transferred to a recipient cell by conjugation. Resolution of this structure in the recipient yields the F factor and the plasmid each with a single $\gamma\delta$ insertion (Fig. 1).

This paper describes the use of conjugal transfer of a plasmid to introduce $\gamma\delta$ insertions into a target sequence (12). The target DNA fragment has been subcloned into a minimal plasmid in which nearly all the plasmid sequences are selectable. Under these conditions recovered transpositions almost always contain a $\gamma\delta$ transposon inserted into the target sequences. The sites of insertion can be readily mapped by polymerase chain reaction (PCR) (13). Finally, orientation-specific sequencing primers allow sequence analysis in both directions from the insertion point.

METHODS

Bacterial Mating. Two *Escherichia coli* strains are grown overnight under appropriate antibiotic selection. The donor strain, DPWC (*supE42* Δ*recA*[*Sst II–Eco*RI] *srl*::Tn*10*-[Tet^s], F$^+$), contains the target plasmid, which confers resistance to ampicillin. The recipient, strain JGM, is strain MC1061 (14) that carries Tn*5*seq1 (9) and is F$^-$ and kanamycin resistant. One-tenth milliliter of each overnight culture and 2 ml of LB medium are combined in a sterile 15-ml tube and incubated on a rotary wheel (30 rpm) for 3 hr or longer at 37°C. One-tenth milliliter of a 100-fold dilution of the mating mixture is plated on an LB-agar plate containing both ampicillin at 100 μg/ml and kanamycin at 50 μg/ml. This plate is incubated overnight at 37°C. Generally, this procedure yields 50–500 Amp^{r}Kan^r colonies. Plating nondiluted and 10-fold dilutions of the mating mixture results in bacterial lawns that consist primarily of bacteria resistant only to kanamycin.

Construction of the Miniplasmid Vector. The miniplasmid vector, used as a transposon target, was constructed by the PCR (14) followed by standard recombinant DNA techniques (15). (*i*) Two oligonucleotides were synthesized that would PCR-amplify the β-lactamase gene and the replication origin of the pUC plasmids. One oligonucleotide, N-AMP (5′-ATGAGACAATAACCCTGA-3′), hybridizes just upstream of the β-lactamase gene (near position 4210 of the pBR322 map) (16). The second oligonucleotide, ORI F (5′-GCCCGGGCGTTGCTGGCGTTTTTCC), is located around pBR322 position 2520 and contains a *Sma* I site. A PCR was performed by using this oligonucleotide pair as primers and pBluescriptKS2 (Stratagene) as template. The PCR product was phosphorylated with T4 polynucleotide kinase, ligated, and introduced into *E. coli* to generate plasmid pOAS. The polylinker from pBluescriptKS2 was then introduced into pOAS to generate pMOB (Fig. 2). To accomplish this construction, the polylinker was first PCR-amplified by using the reverse and universal sequencing oligonucleotides that are commercially available (New England Biolabs). Plasmid pOAS was linearized with *Sma* I and ligated to the polylinker.

DNA Sequence Analysis. Templates for DNA sequence analysis were prepared by using rapid-boil plasmid preparations (17). The sequencing reactions were done with the United States Biochemical T7 Sequenase version 2.0 kit according to the enclosed protocols. The sequencing primers are oligonucleotides complementary to transposon sequences adjacent to the inverted repeat ends. GD1 (5′-CAACGAATTATCTCCTT-3′) will sequence outward from the $\gamma\delta$ end

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: PCR, polymerase chain reaction.

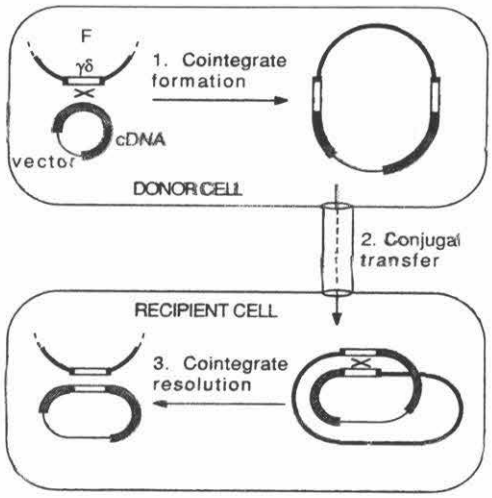


FIG. 1. Transposon mobilization into target DNA sequences. This diagrams the probable mechanism of γδ-transposon insertion into the target sequences by bacterial mating. According to this model (5), a cointegrate is formed in the donor cell between the F factor and the target plasmid. In this cointegrate, both plasmids are flanked by γδ transposons. After transfer, the cointegrate is resolved, leaving a transposon copy in each plasmid.

closest to the *Sac* I site in the transposon (12). GD2 (5′-TCAATAAGTTATACCAT-3′) will sequence outward from the opposite end.

PCR Conditions. All PCRs were done in 40-μl total volume with a light mineral oil overlay. For the construction steps, a three-step protocol was used that included the following: 94°C for 45 sec, 55°C for 45 sec, and 72°C for 2 min. This protocol was then repeated for 10 cycles. For the transposon mapping experiments a two-step PCR protocol was used: 92°C for 20 sec and 72°C for 2 min. One-half microliter of a bacterial overnight culture was added to the 40-μl reaction mixture and subjected to the PCR for 40 cycles. For each template two PCRs were done. Both PCRs used an oligonucleotide (GDIR; 5′-TTTCGTTCCATTGGCCCTCAAACCCC-3′) complementary to the inverted repeat of the γδ transposon. The second oligonucleotide primer used in one PCR (LREV; 5′-AACAGCTATGACCATGATTACGCCAAG-3′) was complementary to a sequence just upstream of the T7 promoter (see Fig. 2). The second oligonucleotide primer used in the other PCR mapping (LUNI; 5′-GTAAAACGACGGCCAGTGAGCGCG-3′) is complementary to a region immediately adjacent to the T7 RNA promoter. The PCR buffer was purchased from Cetus.

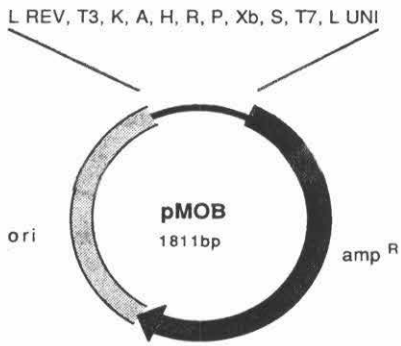


FIG. 2. Map of pMOB, the miniplasmid used for transposon-targeted DNA sequence analysis. This plasmid contains (*i*) a pUC plasmid origin of replication (pUC ori), (*ii*) β-lactamase gene (Amp^r), (*iii*) a multiple cloning site (MCS), (*iv*) bacteriophage RNA polymerase promoters (T3 and T7), (*v*) primer sites for PCR mapping of the transposon insertion sites (LREV, LUNI). The polylinker was derived from a PCR from the plasmid pBluescriptKS2; many restriction sites occur in this polylinker. We have tested for and used only the *Apa* I (A), *Eco*RI (R), *Kpn* I (K), *Pst* I (P), *Hin*dIII (H), *Sac* I (S) and *Xba* I (Xb) sites (these sites are all unique and found only in the MCS region). Because this plasmid was constructed from DNA fragments amplified by PCRs, the DNA sequence may vary somewhat from sequence of the parent plasmids.

RESULTS AND DISCUSSION

Efficiency of γδ Transposition into Target Plasmids. Transposition of γδ from an F factor to a plasmid is believed to produce a cointegrate that can be conjugally transferred. After conjugation the cointegrate is resolved in the recipient cell (Fig. 1).

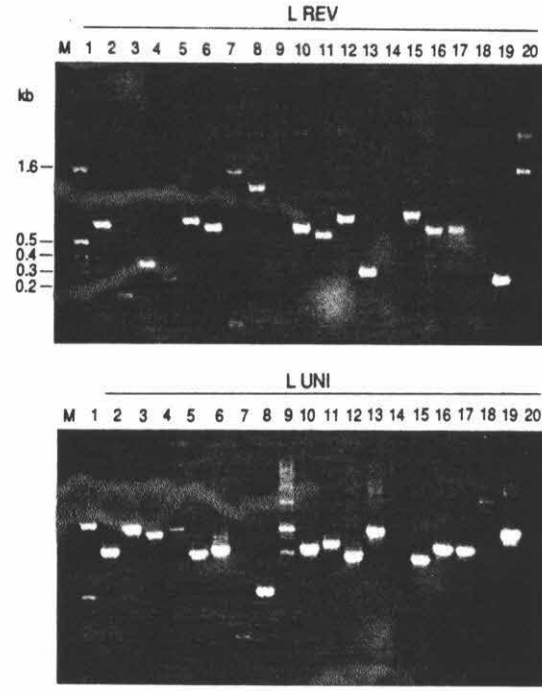


FIG. 3. PCR mapping of transposon-insertion sites. γδ transposons were introduced by bacterial mating into 1.7-kb cDNA (1H2) that was subcloned into pMOB. Bacterial overnight cultures from each of 20 individual transposition events provided the templates in two separate PCRs. In the first set of PCRs LREV and GDIR were the primers; PCR products were then analyzed by agarose gel electrophoresis (*Upper*, labeled LREV). The second set of PCRs were identical, except that LUNI and GDIR oligonucleotides were the primers (*Lower*); results of these PCRs are labeled LUNI. Lanes in each of these gels are matched so that the upper gel shows the PCR products generated by using the first set of primers, and the bottom gels show the analogous products with the second set of primers. Exact sequences of the oligonucleotides and PCR conditions are indicated in text; interpretation of this data is presented in Fig. 4. As mentioned in text, some PCRs gave ambiguous results: for example, multiple PCR products can be seen in lane 9 (*Lower*), whereas no apparent products can be seen in either reaction in lane 14. Size markers (M) are *Hin*fI digests of pBR322.

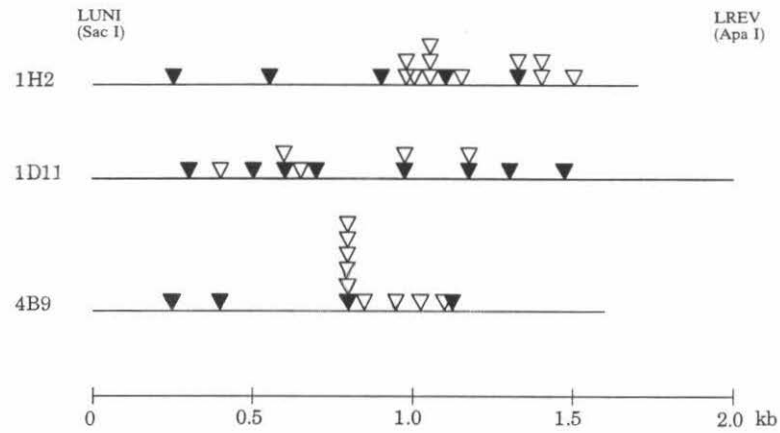


FIG. 4. PCR mapping analysis allows the generation of maps of the γδ-transposon-insertion points. Insertions into three different cDNAs (1H2, 1.7 kb; 1D11, 2.0 kb; and 4B9, 1.6 kb) are shown. These plasmids are all longer versions of cDNA clones originally isolated by Palazzolo *et al.* (19). Each triangle represents the insertion of a single γδ transposon into a given cDNA. PCR mappings that gave ambiguous results were excluded from the map. The filled triangles represent clones selected for DNA sequence analysis. A complete double-stranded analysis was possible from just these selected clones.

Although the mobilized plasmid is expected to contain a single copy of γδ, some experiments (18) indicated that an alternative method of plasmid transfer is possible in certain *E. coli* strains. Examination of transferred plasmids in these strains showed that only 30% of the plasmids contained transposons after transfer. To determine the fraction of mobilized plasmids that contain γδ, *E. coli* DPWC (a donor strain; see *Methods*) was transformed with a 3-kilobase (kb) plasmid vector conferring ampicillin resistance, pBluescriptKS2 (Stratagene), into which a 1.5-kb *Drosophila* cDNA fragment (4B9; ref. 19) had been subcloned. These cells were then mated to JGM (a kanamycin-resistant recipient strain, see *Methods*). Recipient cells that received the plasmid were selected on plates containing ampicillin and kanamycin. Restriction enzyme analysis of 20 mobilized plasmids indicated that, in each instance, the plasmid contained a γδ insertion. Furthermore, restriction fragment length polymorphisms in the restriction enzyme digests suggested that the transposons had inserted at different sites.

Construction of a Miniplasmid Sequencing Vector. An important criterion for the successful application of transposition to DNA sequence analysis is that the transposon be forced into the target DNA sequences and not the plasmid. Restriction analysis of the γδ insertions described above suggested that most insertion events occurred in the vector and not in the insert. We used dideoxynucleotide chain-termination DNA sequencing to identify the transposon-insertion sites in more detail. Of eight clones chosen at random and sequenced, all transposition sites occurred at different locations in the vector.

These experiments suggest that a plasmid, in which insertions into the vector sequences can be selected against, is useful, as then only the transposon insertions in the target can be recovered. Ideally, the vector should contain only an origin of replication, a drug-resistance gene, and a multiple cloning site [the construction of such a miniplasmid, pMOB (Fig. 2), has been described].

To test where transposons insert in this 1.8-kb construct, three different *Drosophila* cDNA molecules (1H2, 1D11, 4B9) (19) were subcloned into this plasmid, and each subclone was separately used as target for γδ transposition. Twenty clones from each of the mating mixtures were selected and analyzed by restriction mapping. All 60 clones contained a γδ transposon, and most insertion sites were in or near the target cDNA fragments.

Analysis of Transposon-Insertion Sites by PCR. The ability to rapidly and simply map the sites of insertion is important for minimizing the labor required to sequence a given target DNA fragment. PCR promised to allow such an identification. For this purpose, we synthesized three oligodeoxynucleotides. One (GDIR) matches the inverted repeat found at each end of γδ. The other two oligonucleotides (LREV and LUNI) flank the cloning site of the miniplasmid (see Fig. 2). Two separate PCRs can be used to determine the point of insertion of a given transposon. In one reaction LREV and GDIR are used as primers, whereas LUNI and GDIR are used in the second reaction. In both cases the same plasmid containing a γδ transposon is the template. Size of the LUNI–GDIR PCR product allows determination of the distance from the transposon-insertion site to the LUNI site at one end of the target fragment, whereas the size of the LREV–GDIR PCR product allows a similar determination of the distance from transposition site to the opposite end of the subclone. Furthermore, the two PCR products should add up to approximately the same size as the fragment subcloned into the miniplasmid.

Such an analysis was performed on the 60 plasmids isolated in the transposon experiments described above. Bacterial overnight cultures provided the templates in two separate PCRs. These reactions were subsequently analyzed by agarose gel electrophoresis. Analysis of 20 transposition events into one clone is shown in Fig. 3. The size of the fragment in each lane delimits the distance of the transposition site from the fixed points in the plasmid.

These experiments allowed us to map the γδ insertion sites for most of the 60 plasmids containing transposons (Fig. 4). Three conclusions can be drawn from these results. (*i*) Forty-two of the 60 transposition events occurred within the cDNA inserts and could be rapidly localized. (*ii*) The insertion sites were sufficiently dispersed within the target to be useful for DNA sequencing. (*iii*) Eighteen of the PCRs gave apparently anomalous results, including multiple PCR products or the apparent absence of PCR products. The transposon insertions that gave such results could not be placed on the transposition maps by this technique and were not further characterized.

DNA Sequence Analysis By Using Transposon Sequences as Priming Sites. Several plasmids were then chosen for DNA sequence analysis of the different cDNA clones. The plasmids that were sequenced were selected because the transposons were spaced ≈300–400 bp apart (Fig. 4). Because end-specific sequences are found immediately adjacent to the inverted repeats, oligonucleotides complementary to these regions can be used to sequence outward from each end of the transposon (see *Methods* for primer sequences). Thus, it is straightforward to simultaneously obtain complete sequence information from both strands of the target sequence from a relatively small number of plasmid–transposon templates. All three plasmids presented in this paper were completely sequenced by a small number of transposon-containing templates (Fig. 4). In addition, we have sequenced six other cDNA clones with inserts from 1.2 to 2 kb. For each clone, a screen of 20–30 γδ insertions was sufficient to obtain a subset of transposon insertions that were spaced every 300–400 bp along the cDNA insert.

Some Limitations to this Sequencing Strategy. One major limitation to this strategy is the inability of the PCR to map transpositional events that are relatively distant from the fixed points of the plasmid (LREV and LUNI). Specifically, PCRs typically yield anomalous products on templates in which transposons have inserted >3 kb away from the fixed plasmid point. One potential response to this limitation is the use of strand-switching PCR to map the position of unknown transposon insertions relative to known transposon-insertion points. In other words, two plasmids that contain the same initial insert but have transposons in different locations can be mapped relative to each other in a single PCR. This reaction contains both plasmids as templates but uses only the inverted-repeat oligonucleotide (GDIR) as primer. The PCR product should be the DNA sequence between the two transpositional events, and its size will map the position of the unknown site relative to the known one.

To test this strategy we used various combinations of the plasmids containing the 1H2 cDNA and different transpositional events. PCRs containing different pairwise combinations of 1H2 plasmids as templates and only GDIR as primer resulted in PCR products of the sizes predicted by the results presented in Figs. 3 and 4.

We thank K. VijayRaghavan, Susan Celniker, and Christopher Martin for many helpful discussions. We also thank Douglas Berg for λTnSeq5 and Randy Reed for the sequence of GD1 and GD2. M.J.P. is a Lucille P. Markey Scholar of the Lucille P. Markey Charitable Trust. B.A.H. is supported, in part, by a U.S. Public Health Service Predoctoral National Research Service Award (T32 GM07616). This work was also supported by U.S. Public Health Service Program Project Grant GM40499 (E.M.M.) and by U.S. Department of Energy Grant FG03-89ER60891 (M.I.S.).

1. Barnes, W. M., Bevan, M. & Sons, P. H. (1980) *Methods Enzymol.* **65,** 98–122.
2. Messing, J. (1983) *Methods Enzymol.* **101,** 20–78.
3. Henikoff, S. (1984) *Gene* **28,** 351–359.
4. Strauss, E. C., Kobori, J. A., Siu, G. & Hood, L. E. (1986) *Anal. Biochem.* **154,** 353–360.
5. Guyer, M. (1983) *Methods Enzymol.* **101,** 362–369.
6. Ahmed, A. (1985) *Gene* **39,** 305–310.
7. Adachi, T., Mizuchi, M., Robinson, E. A., Appella, E., O'Dea, M. H., Gellert, M. & Mizuchi, K. (1987) *Nucleic Acids Res.* **15,** 771–784.
8. Liu, L., Whalen, W., Das, A. & Berg, C. M. (1987) *Nucleic Acids Res.* **15,** 9461–9469.
9. Nag, D. K., Huang, H. V. & Berg, D. E. (1987) *Gene* **64,** 135–145.
10. Barrett, B. K. & Berget, P. B. (1989) *DNA* **8,** 287–295.
11. Grindley, N. F. (1988) *The Recombination of Genetic Material,* ed. Low, K. B. (Academic, New York), pp. 283–360.
12. Guyer, M. (1978) *J. Mol. Biol.* **126,** 347–365.
13. Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A. & Arnheim, N. (1985) *Science* **230,** 1350–1354.
14. Casadaban, M. J. & Cohen, S. N. (1980) *J. Mol. Biol.* **138,** 179–207.
15. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
16. Sutcliffe, J. G. (1978) *Cold Spring Harbor Symp. Quant. Biol.* **43,** 77–90.
17. Holmes, D. S. & Quigley, M. (1981) *Anal. Biochem.* **114,** 193–197.
18. Goto, N., Shoji, A., Horiuchi, S. & Nakaya, R. (1984) *J. Bacteriol.* **159,** 590–596.
19. Palazzolo, M. J., Hyde, D. R., VijayRaghavan, K., Mecklenburg, K., Benzer, S. & Meyerowitz, E. M. (1989) *Neuron* **3,** 527–539.
20. Strausbaugh, L. D., Bourke, M. T., Sommer, M. T., Coon, M. E. & Berg, C. M. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 6213–6217.

§2.5: A removeable source of Cre recombinase for simplified in vivo subcloning

A removeable source of Cre recombinase for simplified in vivo subcloning

Bruce Hamilton and Kai Zinn

Several λ cloning vectors are currently available that allow automatic plasmid subcloning by Cre/loxP site specific recombination *in vivo* (1,2). However, it has been noted that high copy number plasmids that contain a loxP target site grow poorly in the presence of constitutive Cre expression (1). To circumvent this problem, we have made a medium copy number, Cre-producing plasmid that can be lost by incompatibility with the excised plasmid and selected against in sucrose-containing media.

The plasmid, pSacCre, was constructed by cloning a PCR-amplified *cre* gene from bacteriophage P1 and the *B. subtilis sacB* gene from plasmid pUCD800 (3) into the β-lactamase gene of pBR322 (Fig. 1). The resulting plasmid expresses Cre recombinase and confers resistance to tetracycline (but not ampicillin). The *sacB* gene provides a negative selection against the plasmid by making the cell intolerant to ≥5% sucrose in LB media.

Automatic subcloning is accomplished in pSacCre strains by a simple protocol. An appropriate bacteriophage clone is infected into an *E. coli* strain that harbors pSacCre. The infected cells are spread out and grown on LB agar containing ampicillin or carbenicillin to select for the excised plasmid subclone. Conversion frequencies (pfu/cfu) vary from 10^{-3} to 10^{-5} depending on the host strain. For DNA minipreparations, an isolated colony is regrown in liquid LB media supplemented with 5-8% sucrose to select against cells that have retained the pSacCre plasmid. Sample digests from several such preparations are shown in Fig. 2.

Acknowledgements We thank Dr. C. I. Kado for the gift of pUCD800 and Drs. B. Condron and M. Strathmann for helpful discussions.

References

1. Palazzolo, M. J., Hamilton, B. A., Ding, D., Martin, C. H., Mead, D. A., Mierendorf, R. C., Raghavan, K. V., Meyerowitz, E. M., and Lipshitz, H. D. (1990) Phage lambda cDNA cloning vectors for subtractive hybridization, fusion-protein synthesis and Cre-loxP automatic plasmid subcloning. *Gene* **88**: 25-36

2. Elledge, S. J., Mulligan, J. T., Ramer, S. W., Spottswood, M., and Davis, R. W. (1991) LambdaYES - A multifunctional cDNA expression vector for the isolation of genes by complementation of yeast and *Escherichia coli* mutations. *Proc. Nat'l. Acad. Sci.*, **88**: 1731 -1735

3. Gay, P., Le Coq, D., Steinmetz, M., Berkelman, T., and Kado, C. I. (1985) Positive selection procedure for entrapment of insertion sequence elements in Gram-negative bacteria. *J. Bacteriol.* **164**: 918-921

Figure Legends

Figure 1. Plasmid pSacCre. The *cre* gene from P1cmR was amplifed by PCR with the primers 5'–TGTCTGCAGAAATATCAACTAATTATAGCA–3' and 5'–CTGGATATCAAAAGGGTTTGATCGTGATAG–3', to include the pR2 and pR3 promoters (4) and cloned into pBR322 digested with Pst I and Ssp I to create pBCre32. *B. subtilis sacB* was isolated from pUCD800 (3) by digestion with Bam HI and Pst I and cloned into pBCre32 digested with Ase I and Pst I.

Figure 2. Results of pSacCre-mediated subcloning of a λEXLX clone. Single colonies were picked into 2.5 ml Super Broth or 2.5 ml LB supplemented with carbenicillin and shaken at 37°C. After 1 hr., 0.67 ml 24% Sucrose was added (to 5% final concentration) to the LB cultures. Plasmid DNA was isolated from cultures after overnight growth and digested with Sac I and Apa I to release the insert. Lane 1 is a size marker and lane 2 is a

plasmid control. Lanes A-E are plasmid preparartions from Super Broth cultures, lanes F-J are preparations from LB/sucrose cultures.

Figure 1

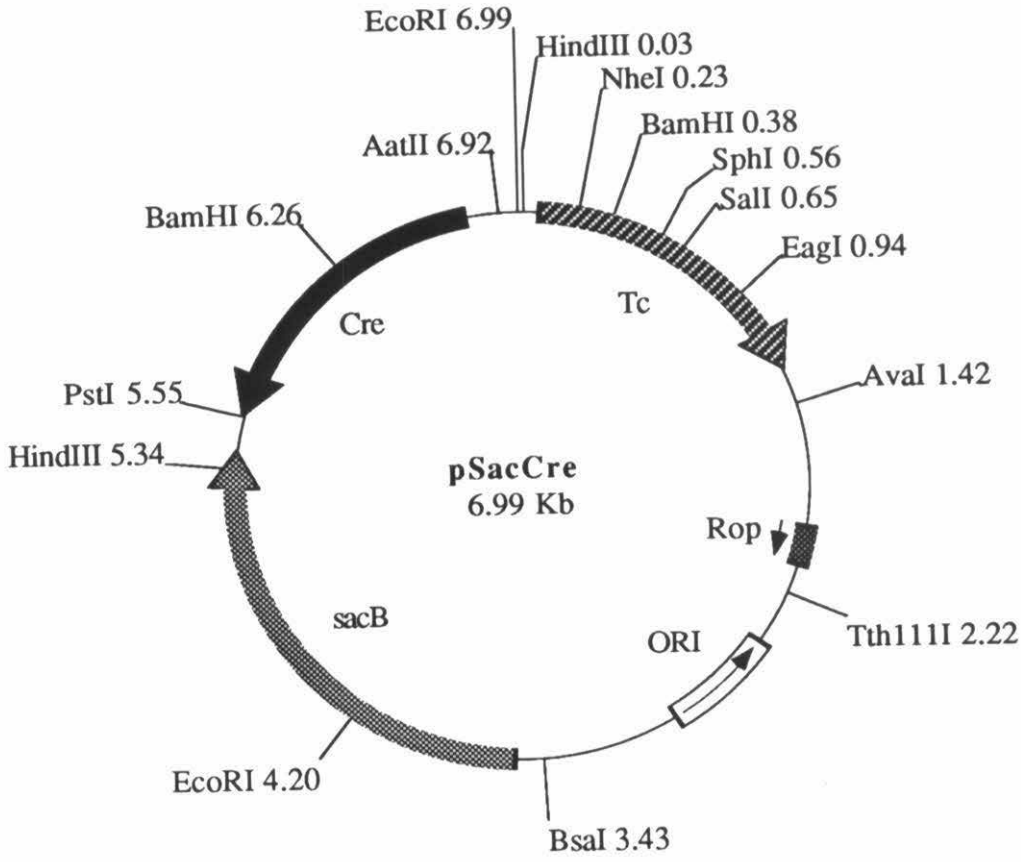
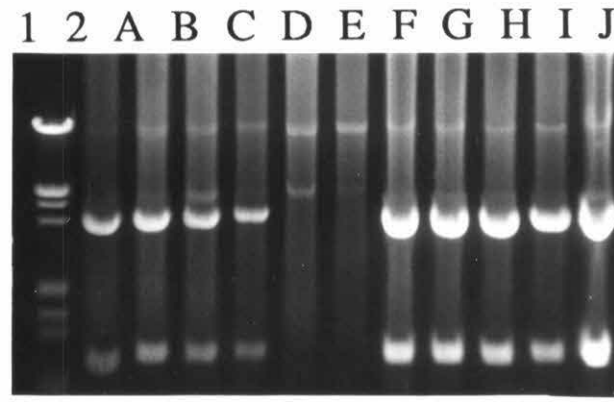


Figure 2



Chapter 3
Large scale screen for transposon insertions into cloned genes

Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 2731–2735, April 1991
Genetics

Large scale screen for transposon insertions into cloned genes

(*Drosophila*/*P* element/reverse genetics/plasmid rescue)

BRUCE A. HAMILTON, MICHAEL J. PALAZZOLO*, JOO H. CHANG, K. VIJAYRAGHAVAN†, CAROL A. MAYEDA*, MICHAEL A. WHITNEY‡, AND ELLIOT M. MEYEROWITZ

Division of Biology 156-29, California Institute of Technology, Pasadena, CA 91125

Communicated by Norman H. Horowitz, December 14, 1990 (received for review August 31, 1990)

ABSTRACT **We describe a method of screening for transposon insertions in or near *Drosophila* loci that correspond to cloned DNA sequences. We mobilize a modified *P* element transposon that carries a bacterial plasmid origin of replication and a drug-resistance marker. The genomic sequences flanking each transposon insertion site can then be rescued as a plasmid in *Escherichia coli*. Libraries of such plasmids, representing pools of transposon-mutagenized individuals, are used as hybridization probes against cloned sequences to determine whether a transposon has inserted next to a particular site in the genome. The number of loci that can be screened simultaneously by this procedure is quite large. We have screened an array of cDNA clones representing almost 700 distinct loci against libraries representing 760 mutagenized flies, and we obtained hybridization signals to 7 different cDNAs. Three of these events have been analyzed in detail and represent genuine insertions near genomic sequences that correspond to the cDNAs.**

To understand the development and physiology of a nervous system in molecular detail, we have initiated a large scale analysis of RNAs expressed in the *Drosophila* brain. Our study began with the isolation of a large number of cDNA clones that represent messages expressed in the adult head but not in the preblastoderm embryo (1). Combining molecular and genetic analyses of these molecules should lead to testable models of their neuronal functions. An important requirement of this approach is the ability to identify mutations in the genetic loci that encode the cloned molecules. Such a genetic screen must meet several criteria: the screen must be simple and compatible with studying many loci simultaneously, it must have high sensitivity and low background, and it must not require any assumptions about mutant phenotypes. This paper describes a screening procedure that meets these criteria.

Although other methods to disrupt cloned *Drosophila* genes have been reported, none meets all of the criteria listed above. While classical genetic screens have been successful, they require a prediction of a mutant phenotype. Screening for loss of antibody binding (2, 3) requires unique reagents and assays for each locus. Methods based on PCR (4–8) require DNA sequence data and at least one unique oligodeoxynucleotide primer for each locus considered. Each locus must also be screened in a separate biochemical reaction. In addition, distinguishing true insertion events from spurious amplification products (ref. 7; B.A.H., M.A.W., C.A.M., and M.J.P., unpublished data) requires considerable effort in larger experiments.

We describe a strategy for reverse genetics that circumvents these problems. The method is based on plasmid rescue of genomic DNA (9, 10) that flanks the insertion sites of a modified *P* element transposon, PlacW (11). By using a library of rescued sequences as a hybridization probe against an array of cDNA clones, we are able to assay for insertions in or near a large number of cloned loci in a single experiment. This requires no assumptions of expected mutant phenotypes and requires neither antisera nor DNA sequence data for PCR primers. Because PlacW carries a w^+ minigene as a dominant genetic marker, it should be possible by standard mutagenesis methods to obtain deletion alleles of loci that are tagged but not functionally disrupted by the transposon. Enough probe could be synthesized from a single plasmid library to screen an array of cloned DNA representing the estimated sequence complexity of the *Drosophila* genome.

MATERIALS AND METHODS

Fly Strains and Crosses. We obtained new PlacW insertions by genetic mobilization. As a source of transposons we used C(1)RM, *y w* 4[PlacW], which carries four copies of PlacW on each homolog and was kindly provided by Dan Lindsley (University of California, San Diego). Transposase activity from the stable source P[ry^+ Δ2-3](99B) (12) was supplied on a derivative of TM2, *Ubx ry*, that carries this insertion, kindly made available by John Merriam (University of California, Los Angeles). Other genetic elements have been described (13, 14). Flies were raised on standard medium (15) at 22°C. C(1)RM, *y w* 4[PlacW] virgins were mated to *shi*; TM2, *Ubx ry* P[ry^+ Δ2–3](99B)/*Sb ry* P[ry^+ Δ2–3](99B) males (12 females and 4 males per pint bottle) for 3–5 days and transferred or discarded. Each mating was transferred no more than twice. Dysgenic virgin females of the genotype C(1)RM, *y w* 4[PlacW]; TM2 *Ubx ry* P[ry^+ Δ2–3](99B) were collected by heat treatment as described (16), and mated as described above to males of the genotype *w*; CyO; TM6, *Ubx/T(2;3)Xa* or *w*; CyO; TM6, *Hu/T(2;3)Xa*. Progeny males with pigmented eyes (indicating a PlacW transposition to an autosome) were individually mated to two or three females of an appropriate balancer strain to establish temporary lines. Males were removed for DNA preparations after 3–5 days.

cDNA Gel Blots. cDNA inserts from λSWAJ3 clones (1, 17) were isolated by performing PCR directly on high-titer lysates of cloned bacteriophage with primers that abut the cloning site (SWAJ3.1, 5′-ATTTAGGTGACACTATAGAATACAC-3′; SWAJ3.2, 5′-CGGAAGCTTGGGCTGCAGGTCGACT-3′). λSWAJ3 contains no plasmid sequences that could hybridize to the plasmid probes in the screening procedure. DNA gel blots of PCR products were prepared by standard methods (18).

DNA Preparations. We prepared DNA from 10–20 flies by a standard method (19) with modifications. Flies were col-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

*Present address: Department of Genetics, Washington University School of Medicine, St. Louis, MO 63110.
†Permanent address: Molecular Biology Unit, Tata Institute of Fundamental Research, Bombay 400005, India.
‡Present address: Department of Biology and Center for Molecular Medicine M-049, University of California, San Diego, La Jolla, CA 92023.

lected on dry ice in a 1.7-ml microcentrifuge tube. Frozen flies were ground in 100 μl of 5% sucrose/80 mM NaCl/0.1 M Tris·HCl, pH 8.5/0.5% SDS/50 mM EDTA with a baked glass rod; the rod was rinsed with another 100 μl of this solution and the sample was refrozen. After 30 min at 70°C, each sample was made 160 mM in KOAc and placed on ice for 30 min. The precipitate was removed by centrifugation. DNA was precipitated from the supernatant with 150 μl of isopropanol and was collected by centrifugation. Recovered DNA was resuspended in 100 μl of 10 mM Tris·HCl, pH 7.5/1 mM EDTA.

Plasmid Rescue. To plasmid rescue transposon-flanking sequences from the mutagenized chromosomes of either 140 or 160 flies, 50 μl from each of seven or eight DNA preparations was pooled. Four 30-μl aliquots (≈3 μg each) of pooled DNA were digested to completion with *Sac* II in separate 40-μl reaction mixtures. Four similar aliquots were each partially digested by incubating with 5 units of *Eco*RI at 37°C for 9 min in a 40-μl reaction mixture. Digests were stopped by heating to 70°C for 15 min. DNA in each reaction mixture was cyclized in a 0.5-ml vol by adding 0.46 ml of a 1.09× ligase reaction mixture [1× is 50 mM Tris·HCl, pH 7.8/10 mM $MgCl_2$/20 mM dithiothreitol/1 mM ATP/1000 units of T4 DNA ligase per ml (New England Biolabs)] and incubating at room temperature for 4–20 hr. DNA was then precipitated by adding LiCl to 160 mM and 0.5 ml of isopropanol and centrifuged to pellet DNA. Pellets were resuspended in 30 μl of water. Similar aliquots were pooled, precipitated with EtOH, and resuspended in 8–9 μl of water. Two 4-μl aliquots from each resuspended pellet were used to transform *Escherichia coli* (strain LE392) by electroporation (20) and selected on LB agarose plates (18) containing carbenicillin (200 μg/ml).

Screening Libraries of Rescued Plasmids. Radiolabeled plasmid DNA from each plasmid-rescue library was hybridized to DNA gel blots containing the array of cDNA inserts (see above). Plasmid DNA was prepared from primary transformant colonies by the rapid boiling method (21). Approximately 100 ng of restriction-digested DNA from each library was labeled by random priming (22) with a commercially available kit (Prime-A-Gene, Promega). Membranes were incubated at 42°C in 50% formamide/5× SSPE (1× SSPE = 0.18 M NaCl/10 mM phosphate, pH 7.4/1 mM EDTA)/1% (wt/vol) SDS/1× Denhardt's reagent/100 μg of sheared salmon testes DNA per ml (18) for at least 15 min before adding probe to a final concentration of 1–4 × 10^6 cpm/ml. Hybridizations were for 12–20 hr at 42°C. Membranes were washed three times in 0.1× SSPE/0.3–0.5% SDS at 50–55°C for 20–60 min.

PCR Amplification of Fly DNA. PCRs were performed on 4 μl of fly DNA solution, in 50 mM Tris·HCl (pH 8.9 at 1 M)/2 mM $MgCl_2$/0.1% Triton X-100/200 μM each dATP, dCTP, dGTP, and dTTP/0.25 μM each oligodeoxynucleotide in a 40-μl reaction volume with 2–3 units of *Taq* DNA Polymerase (Promega). The *P* element terminal sequence oligodeoxynucleotide (7) was the gift of Dennis Ballinger (Sloan–Kettering Research Institute). Other primers were as follows: c3B7, 5′-GGCCGATTGTTAAGTTGCTGTAGTG-3′; c4D12, 5′-TGGTCCCGATAAAGTACCAAACCTC-3′; c8H9, 5′-CACACTTTCACTGCTGCGATATCGG-3′.

Other Methods. Radiolabeled cDNA inserts were prepared by sequential PCR amplifications: a standard amplification from a phage lysate as described above followed by a labeling amplification that is 50–150 μM in each dATP, dGTP, and dTTP, 400–800 nM in [α-^{32}P]dCTP (3000 Ci/mmol; 1 Ci = 37 GBq), and 2–5% (vol/vol) of the first amplification reaction. *In situ* hybridizations to polytene chromosomes were as described (1). Other methods were essentially as described (18).

RESULTS

Identification of Inserted Loci by Flanking Sequence Rescue. Our screening procedure, diagrammed in Fig. 1, consists of genetic crosses to mobilize transposons, plasmid rescue of genomic DNA that flanks the new transposon insertions, and hybridization of radiolabeled rescued plasmids to a collection of cDNA clones bound on nylon membranes. The genetic crosses allow recovery of stable PlacW insertions on the autosomes of F_2 males, usually one insertion per male. These males are mated to establish lines and then used to prepare DNA. This DNA is digested with a restriction enzyme and then circularized. Only genomic sequences flanking PlacW insertion sites remain physically linked to the plasmid sequences of the transposon. These are recovered by transformation into *E. coli*. Radiolabeled plasmid DNA from drug-resistant colonies is used as a hybridization probe against an array of ≈680 cloned cDNA inserts (1).

Our first probe against the cDNA filters represented a pool of 140 F_2 male flies. Genomic DNA was prepared from 7

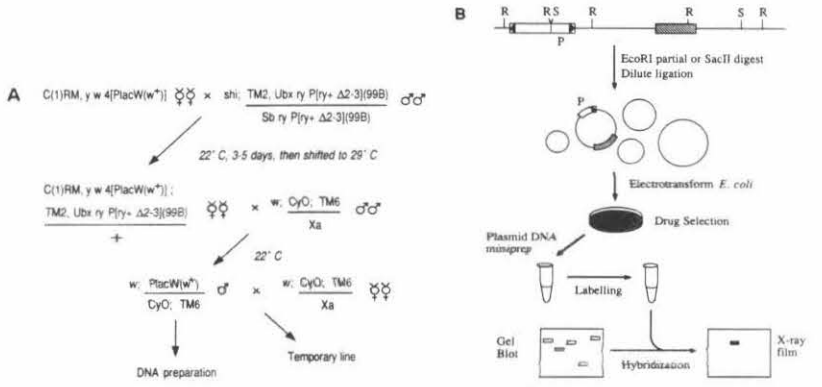


FIG. 1. Overview of screening procedure. (*A*) Genetic crosses used to generate PlacW insertion lines. (*B*) Steps in screening the lines. PlacW transposon (open box with arrowheads) integrated near a hypothetical cloned cDNA sequence (hatched box). R, *Eco*RI recognition site; S, *Sac* II recognition site. The genomic DNA from a number of flies bearing different PlacW insertions is digested with a restriction enzyme and then ligated into closed circles. Only a small fraction of the circles will contain plasmid sequences, P, from a transposon; only these are recovered as plasmids after transformation and drug selection. Recovered plasmids are isotopically labeled and used as a hybridization probe to filter-bound cDNA clones.

groups of 20 flies each. Half of the DNA from each preparation was pooled and aliquots were either partially digested with *Eco*RI or digested to completion with *Sac* II. Digested fly DNA was circularized in the presence of T4 DNA ligase. Transformation of *E. coli* produced 908 drug-resistant colonies from the *Eco*RI-digested DNA and 563 colonies from the *Sac* II-digested DNA. Plasmid DNA was prepared directly from colonies, yielding ≈8 μg per library. DNA from each library was labeled to high specific activity ($7–9 \times 10^8$ cpm/μg) and equal amounts of radioactivity from each were combined as a hybridization probe against the filter array of cDNA targets. Autoradiography revealed hybridization to two cDNA clones: c3B7 and c4D12 (Fig. 2). Similar screens, representing PlacW insertions in 620 additional males, revealed hybridization signals to clones c1C2, c1F4, c4B9, c4E10, c6F9, and c8H9 (Table 1). Two additional clones in our collection gave a hybridization signal with every probe tested; these signals appear to be nonspecific and we have neither counted them as positive signals nor pursued them further.

Characterization of Rescued Plasmids. A rescued plasmid that corresponds to a PlacW insertion near a gene of interest should contain both a single contiguous piece of the fly genome that hybridizes to the cDNA and a *P* element terminal repeat sequence from the transposon. We have tested this for the first three presumed insertion events (c3B7, c4D12, and c8H9), as described below.

We isolated the relevant plasmids by colony filter hybridization to cDNA probes, using both the original plasmid-rescue libraries and primary transformant colonies representing either pools of 20 fly lines or single fly lines. A c3B7 cDNA probe allowed the isolation of nine independent plasmids from *Eco*RI-rescued libraries, but none from *Sac* II-rescued libraries. Each plasmid contained a 4-kilobase (kb) genomic fragment with no internal *Eco*RI sites. It is unlikely that these plasmids could have arisen by ligation of c3B7 homologous sequences to noncontiguous plasmid sequences, as this would generate an internal *Eco*RI site and a new rescued fragment size in each instance. Similarly, independent plasmids for c4D12 were isolated from *Eco*RI (partial digest), but not *Sac* II, libraries. Each of these contained either a 1.2-kb genomic fragment with no internal *Eco*RI sites or a 2-kb genomic fragment with one *Eco*RI site 1.2 kb from the *P* element end. Several independent c8H9 homologous plasmids were rescued by *Sac* II but not by *Eco*RI. Each of these contained a 7-kb genomic fragment lacking internal *Sac* II sites.

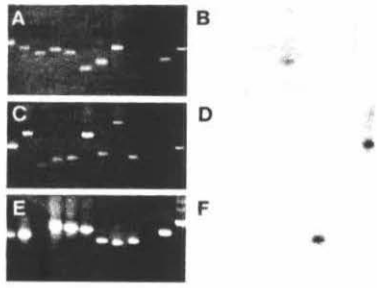


FIG. 2. Hybridization of rescued plasmids to cDNA gel blots. DNA gel blots of PCR-amplified cDNA inserts were hybridized to probes from rescued plasmids representing 140–160 mutagenized flies. (*A*) Agarose gel of some amplified cDNAs, including c3B7, stained with ethidium bromide and visualized by UV fluorescence. (*B*) Autoradiogram of the gel blot corresponding to *A*, showing hybridization to cDNA c3B7. (*C*) Gel including c4D12. (*D*) Autoradiogram of blot of *C*, showing hybridization to c4D12. (*E*) Gel including c8H9. (*F*) Autoradiogram corresponding to gel in *E* showing hybridization to c8H9.

Table 1. Plasmid rescue and hybridization results

Probe	No. of lines	No. of colonies	cDNA clone(s)
1	140	1471	c3B7, c4D12
2	140	776	—
3	160	390	c8H9
4	160	500	c1C2, c1F4, c4B9
5	160	454	c1F4, c4E10

A rescued plasmid and the original cDNA clone for each locus were used to probe duplicate gel blots of genomic DNA digested with a variety of restriction enzymes. For each locus, the rescued plasmid and the cDNA identify overlapping patterns of bands and each probe appears to be single copy in the genome. This further demonstrates that the recovered sequences represent discrete, tagged sites rather than ligations of disjoint sequences (or tagged dispersed repeat sequences).

A genuine rescue product should also contain one of the *P* element terminal repeat sequences from the PlacW transposon. We subjected one of the plasmids rescued from each locus to chain-termination sequencing (23) by using a synthetic primer that corresponds to the *P* element terminus (7). From each of these reactions, we obtained a unique sequence ladder, indicating that each of these plasmids contains a single *P* element terminus.

Isolation of Fly Lines Carrying the Identified Insertions. We have used two different methods to identify the single lines of flies that carry the desired PlacW insertions. The first method is based on PCR; the second is based on plasmid rescue.

We used sequence information from the rescued plasmids to design locus-specific PCR primers that lie adjacent to the PlacW terminal repeat; we used these primers in combination with the *P* element terminal repeat primer for PCR on fly genomic DNA templates. Amplification between the c3B7 and *P* element primers was used to detect a PCR product from DNA from pools of fly lines and then from individual lines. Only 1 of the original 20-fly DNA preparations in the pool that hybridized to c3B7 supports an amplification product of the size predicted from DNA sequence and illustrated by PCR on the rescued plasmid (≈220 base pairs; Fig. 3). Two of the 20 lines in this pool had failed to propagate. We prepared DNA from each of the remaining 18 lines for PCR. Nine of the lines gave amplification products of the appropriate size. This result was repeated with a second set of DNA preparations and has also been partially verified by plasmid rescue experiments and *in situ* hybridizations to polytene chromosomes (see below).

In addition to PCR sorting experiments, we assayed for the c3B7 and c4D12 insertions by plasmid rescue. PCR results indicated that a single pool contained both insertions. We

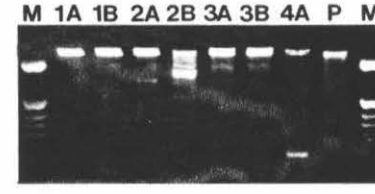


FIG. 3. PCR amplification between c3B7 and *P* element primers. Lanes M, size standards. Alphanumerically labeled lanes are amplification products from each of the pools of 20 flies used to make the first rescued-plasmid library. Lane P, control amplification from the rescued plasmid.

used *Eco*RI partial digests of new fly DNA preparations to rescue transposon-flanking sequences from each of 13 lines from this pool (2 lines were propagating poorly and so were not used to prepare DNA and 3 more lines did not yield drug-resistant colonies in this experiment). Minipreparations of plasmid DNA from 1–3 individual colonies for each line rescued were analyzed by DNA gel blot hybridization. A c3B7 probe hybridizes to plasmids from each of the lines implicated by PCR from which rescued plasmids were recovered: lines 4.01, 4.03, 4.08, 4.10, and 4.13. The filter was stripped of the c3B7 probe and hybridized to a c4D12 probe. The c4D12 probe identified one line, 4.09. Since the 13 plasmid-rescued lines were sufficient to identify one line bearing the c4D12 insertion and several bearing the c3B7 insertion, we have not tried further to rescue plasmids from any of the remaining 5 lines.

We failed to identify an insertion line for c8H9 by either method. From the pool of 20 lines that produced the c8H9 *Sac* II rescue plasmid, 6 lines failed to propagate and could not be tested. This is anomalous: we usually recover offspring from >90% of single-male matings. It appears from the data that the initial c8H9 hybridization signal was due to a genuine insertion, but that the corresponding line was lost.

***In Situ* Hybridization to Polytene Chromosomes.** As a final confirmation of the first two identified insertions, we tested whether the cDNAs for which we isolated insertions and a PlacW element in the identified line are associated with the same cytological location on polytene chromosomes (Fig. 4). Biotin-labeled probes prepared from cDNA clones in pEXLX (24) and from rescued genomic sequences were individually hybridized to larval salivary gland polytene chromosomes from a wild-type strain (Oregon-R). PlacW insertions were localized by hybridizing a biotinylated pBR322 probe to the plasmid sequences of the transposon in polytene chromosomes from identified insertion lines. Each probe identifies a single site in the genome and for each locus the cDNA, the rescued genomic sequence, and the PlacW element all map to the same cytological location: 45D for clone c4D12 and 90D for c3B7. In addition to the results shown in Fig. 4, we have tested the location of PlacW elements in lines 4.01 and 4.08; these also indicate single PlacW insertions at 90D. The rescued fragment identified by c8H9 hybridizes to a single site at 71F.

DISCUSSION

With the advent of modified transposons for enhancer traps (11, 25, 26), it has become possible to clone genes selected for either mutant phenotypes or expression patterns directed by associated regulatory elements. We have used such modified transposons to take a different approach: isolating lesions in genes that correspond to cloned sequences.

Comparative Advantages of This Method. The screen described here offers several advantages not shared by previous methods. Any presumed insertion can be rapidly verified by characterizing the rescued plasmid before investing significant time in attempting to isolate and characterize a line of flies. Moreover, each of the presumed insertions we have analyzed appears to reflect a genuine correspondence between a contiguous rescued genomic fragment and a target cDNA, suggesting that the incidence of false positive signals should generally be low. Since the modified transposon carries a dominant visible marker (w^+), isolation of subsequent deletion alleles by imprecise excision of the transposon (27–29), radiation mutagenesis, or exposure to chemical mutagens known to cause deletions (30) should be straightforward. In addition, since very little extra effort is required to screen for insertions into additional loci and sufficient mass of rescued plasmid DNA is recovered from minipreparations, projects representing up to 10^4 unique target sequences (or more) should be feasible. This may prove useful for genome mapping and related studies.

Fidelity of the Screen. From probes representing 760 PlacW insertion lines, we obtained hybridization signals to eight clones in our array. We have characterized three of these insertions in detail to demonstrate that they are genuine, rather than artifacts of the screening procedure. Chain-termination sequencing shows that each rescued plasmid contains a single *P* element terminus, as predicted. Gel blot hybridization to restriction-digested genomic DNA shows that each rescued sequence represents a single continuous site in the genome, which encodes the corresponding cDNA. For c4D12 and c3B7, we have also shown that a PlacW element is present in the identified insertion line at the same cytological location as the rescued sequences and the cDNA. Similar criteria have now been used to verify and obtain lines for the c4B9 and c1C2 insertions (although the latter appears to be a transposon; ref. 31 and B.A.H. and J. Liao, unpublished data). Since all of the insertions analyzed appear genuine, we conclude that the screening procedure is reliable.

Screening Efficiency. We have used a collection of nearly 850 cDNA clones (all but the eye-specific clones in ref. 1) as hybridization targets in this screen. Cross-hybridization data suggest that these represent 682 discrete loci (1). However, preliminary DNA sequence data from a subset of almost 250 of these clones suggest that 2–5% of the collection comprise related or identical clones (M.A.W., C.A.M., B.A.H., and M.J.P., unpublished data). This is consistent with a report on

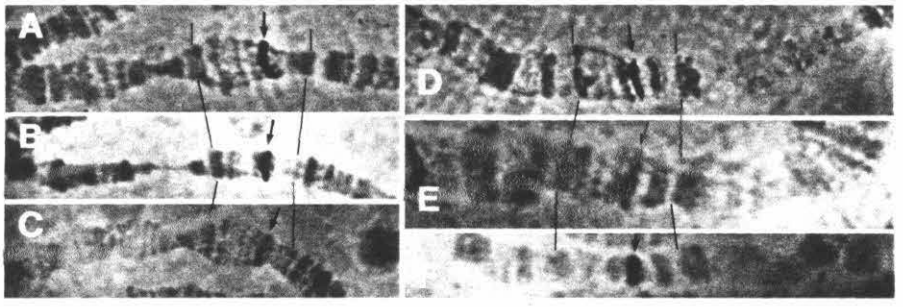


FIG. 4. *In situ* hybridization to polytene chromosomes. Arrows indicate the sites of hybridization signals (90D in *A–C*; 45D in *D–F*). Lines delimit the numbered division into which each signal falls. (*A*) Hybridization of cDNA clone c3B7-15 probe to polytene chromosomes from wild-type strain Oregon-R. (*B*) Rescued plasmid c3B7.R2 hybridized to Oregon-R chromosomes. (*C*) pBR322 probe hybridized to polytene chromosomes from the insertion line 4.03. Hybridization is to the plasmid sequences of the PlacW transposon. (*D*) cDNA clone c4D12-11, Oregon-R chromosomes. (*E*) Rescued plasmid 4D12R4.1, Oregon-R chromosomes. (*F*) pBR322 probe to insertion line 4.09.

eye-specific cDNAs from this collection (31). Thus, the number of distinct loci we have screened may be as low as 650.

We assayed 760 fly lines in this screen, predominantly carrying one PlacW element each. However, our dysgenesis scheme allows redundant isolations of single transposition events that occur in premeiotic germ cells or their progenitors; for example, 9 individual lines bear the c3B7 insertion. Such clustering of identical insertions is not surprising. While the transposase should be active quite early in development, the overall rate of transposition is low (15–25% of male progeny from the mass-mated dysgenic females show w^+ function) and any insertion that occurred very early in the germ lineage of one female would comprise a large fraction of the w^+ progeny in a given bottle (see *Materials and Methods*). Thus, the actual number of independent insertions represented by these flies is <760.

Modifications to the screening procedure should allow screening of the large numbers of mutagenized flies needed to find insertions near single target sites. Mating flies en masse and hybridizing DNA gel blots of rescued libraries to cloned probes could save considerable labor. Alternatively, rescued plasmid libraries could be screened by PCR using two locus-specific primers a known distance apart. Having rescued only genomic sequences that flank an insertion obviates the need for a transposon-specific primer and knowing the size of legitimate amplification products reduces background caused by spurious amplification products.

Sorting Lines. We described two methods for finding the appropriate insertion line among a pool of lines known to contain the insertion: PCR and flanking sequence plasmid rescue. PCR is simple, rapid, and extremely efficient; however, this high sensitivity makes it prone to contamination artifacts. Plasmid rescue assays are less sensitive to trace contamination, but they require more starting material.

The PCR sorting strategy we describe differs from those described by Ballinger and Benzer (7) and Kaiser and Goodwin (8) in two important respects. First, the veracity required of the PCR to determine which line carries the identified insertion is much lower than that required to ask whether any line carries any such insertion. Second, we can predict the size of the genuine amplification product because our gene-specific primer is derived by sequencing the rescued genomic fragment from the *P* element primer. This more closely resembles the method of Kim and Smithies (6) and ought to alleviate the signal/noise problems caused by spurious amplification products.

We thank D. L. Lindsley and J. Merriam for fly stocks; E. B. Lewis and S. E. Celniker for fly stocks, useful discussions, and comments on the manuscript; D. G. Ballinger for the *P* element oligodeoxynucleotide, fly stocks, and discussions; U. VijayRaghavan, S. Halsell, and T. Jack for practical advice; A. van der Bliek for *shi*; and H. D. Lipshitz, C. H. Martin, and members of the Meyerowitz laboratory for comments on drafts of this paper. B.A.H. was supported in part by a U.S. Public Health Service Predoctoral National Research Service Award (T32 GM07616); M.J.P. is a Lucille P. Markey Scholar; K.V.R. was supported by a Rockefeller Foundation Biotechnology Career Fellowship. This work was supported by a scholar's grant from the Lucille P. Markey Charitable Trust (M.J.P.) and U.S. Public Health Service Program Project Grant GM40499 (E.M.M.).

1. Palazzolo, M. J., Hyde, D., Raghavan, K. V., Mecklenberg, K., Benzer, S. & Meyerowitz, E. M. (1989) *Neuron* **3,** 527–539.
2. Van Vactor, D., Jr., Krantz, D. E., Reinke, R. & Zipursky, S. L. (1988) *Cell* **52,** 281–290.
3. Katz, R., Moats, W. & Jan, Y. N. (1988) *EMBO J.* **7,** 3471–3477.
4. Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A. & Arnheim, N. (1985) *Science* **230,** 1350–1354.
5. Saiki, R. K., Gelfand, D. H., Stoffel, B., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988) *Science* **239,** 487–491.
6. Kim, H.-S. & Smithies, O. (1988) *Nucleic Acids Res.* **16,** 8887–8903.
7. Ballinger, D. G. & Benzer, S. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 9402–9406.
8. Kaiser, K. & Goodwin, S. F. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 1686–1690.
9. Perucho, M., Hanahan, D. & Wigler, M. (1980) *Cell* **22,** 309–317.
10. Steller, H. & Pirrotta, V. (1986) *Mol. Cell. Biol.* **6,** 1640–1649.
11. Bier, E., Vaessin, H., Shepherd, S., Lee, K., McCall, K., Barbel, S., Ackerman, L., Carretto, R., Uemura, T., Grell, E., Jan, L. Y. & Jan, Y. N. (1989) *Genes Dev.* **3,** 1273–1287.
12. Robertson, H. M., Preston, C. R., Phillis, R. W., Johnson-Schlitz, D. M. Benz, W. K. & Engels, W. R. (1988) *Genetics* **118,** 461–470.
13. Lindsley, D. L. & Grell, E. H. (1968) *Genetic Variations of Drosophila melanogaster* (Carnegie Institute of Washington, Washington, DC).
14. Lindsley, D. L. & Zimm, G. (1990) *Drosoph. Inf. Serv.* **68,** 1–382.
15. Lewis, E. B., (1960) *Drosoph. Inf. Serv.* **34,** 117–118.
16. Grigliatti, T. (1986) in *Drosophila: A Practical Approach*, ed. Roberts. D. B. (IRL, Oxford, England), pp. 39–58.
17. Palazzolo, M. J. & Meyerowitz, E. M. (1987) *Gene* **52,** 197–206.
18. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), 2nd Ed.
19. Bender, W., Speierer, P. & Hogness (1983) *J. Mol. Biol.* **168,** 17–33.
20. Dower, W. J., Miller, J. F. & Ragsdale, C. W. (1988) *Nucleic Acids Res.* **16,** 6127–6145.
21. Holmes, D. S. & Quigley, M. (1981) *Anal. Biochem.* **114,** 193–197.
22. Feinberg, A. P. & Vogelstein, B. (1984) *Anal. Biochem.* **137,** 266–267.
23. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
24. Palazzolo, M. J., Hamilton, B. A., Ding, D., Martin, C. M., Mead, D. A., Mierendorf, R. C., Raghavan, K. V., Meyerowitz, E. M. & Lipshitz, H. D. (1990) *Gene* **88,** 25–36.
25. O'Kane, C. J. & Gehring, W. J. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 9123–9127.
26. Wilson, C., Pearson, R. K., Bellen, H. J., O'Kane, C. J., Grossniklaus, U. & Gehring, W. J. (1989) *Genes. Dev.* **3,** 1301–1313.
27. Voelker, R. A., Greenleaf, A. L., Gyurkovics, H., Wisely, G. B., Huang, S.-M. & Searles, L. L. (1984) *Genetics* **107,** 279–294.
28. Daniels, S. B., McCarron, M., Love, C. & Chovnick, A. (1985) *Genetics* **109,** 95–117.
29. Salz, H. K., Cline, T. W. & Schedl, P. (1987) *Genetics* **117,** 221–231.
30. Ashburner, M. (1989) *Drosophila: A Laboratory Handbook* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
31. Hyde, D. R., Mecklenburg, K. L., Pollock, J. A., Vihtelic, T. S. & Benzer, S. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 1008–1012.

Chapter 4
Directed mutagenesis of a Drosophila gene encoding a protein tyrosine phosphatase by *P* element local transposition and molecular screening

Directed mutagenesis of a Drosophila protein tyrosine phosphatase gene by *P*-element local transposition and molecular screening

Bruce A. Hamilton, Albert Ho*, and Kai Zinn

Division of Biology, 216-76

California Institute of Technology,

Pasadena, CA 91125

*Present address:

Washington University School of Medicine

Box 8033

660 S. Euclid Avenue

St. Louis, MO 63110

corresponding author:

Dr. Kai Zinn

Division of Biology, 216-76

California Institute of Technology

Pasadena, CA 91125

Running head: Directed mutagenesis of *DPTP99A*

ABSTRACT

Drosophila *P* elements have been recently shown to transpose preferentially into nearby sites. Here we describe the use of *P* element local transposition for directed mutagenesis. Among transposants from a small pilot screen we isolated a new locus, *encumbered*, based on its visible phenotypes. This mutation is associated with a transposition from 45D1-2 to 45C1-2. In a second set of experiments, we used a molecular screen to identify four different *PlacW* insertions near the *DPTP99A* gene among 938 independent lines derived by mobilizing an element at 99B1-3. *DPTP99A* encodes an axonal receptor-linked protein tyrosine phosphatase and is located at 99A7-8. All four insertions were isolated from a pool of 170 lines enriched for new transpositions by scoring a visible marker carried by the transposon. Imprecise excision of one inserted element produced a small deletion that removes most of the *DPTP99A* coding sequence. Embryos homozygous for this deletion lack detectable DPTP99A protein, but do not appear morphologically abnormal. We discuss the implications of these results for the roles of PTPs in neural development and for the design of efficient directed mutagenesis experiments based on local transposition.

Receptor-linked protein tyrosine phosphatases (R-PTPs) are transmembrane glycoproteins that may provide a direct link between selective cell recognition and signal transduction *via* control of tyrosine phosphorylation (reviewed by (Fischer, Charbonneau and Tonks 1991; Charbonneau and Tonks 1992). A specific role for phosphotyrosine regulation in axonal patterning is suggested by the finding of high concentrations of c-*src* in mammalian growth cones (Maness *et al.* 1988) and by the finding in Drosophila that certain double mutant combinations that include lesions in the *D-abl* tyrosine kinase gene produce severe disruption of central nervous system (CNS) axon pathways (Gertler *et al.* 1989; Elkins *et al.* 1990). More recently, three different R-PTPs (DLAR, DPTP10D, and DPTP99A) have been shown to be selectively expressed on CNS axons during embryonic development in Drosophila (Yang *et al.* 1991; Tian, Tsoulfas and Zinn 1991; Hariharan, Chuang and Rubin 1991). Each of these proteins has an adhesion molecule-like extracellular region composed of fibronectin type III (FN) repeats, with or without additional immunoglobulin-like (Ig) domains, linked by a single transmembrane segment to one or two cytoplasmic PTP domains. To address the roles that R-PTPs play in the assembly of the CNS axon array, we have begun to search for mutations in the genes encoding these molecules.

Because we were unsure whether a mutation in a single R-PTP gene would result in lethality or in any other predictable defect, we chose to develop mutagenesis strategies that would allow us to isolate lesions in these genes without making assumptions about the resulting phenotypes. The three axonal R-PTPs have similar structures and expression patterns, suggesting that they may have partially overlapping functions. Furthermore, the signal transduced by any one R-PTP may be functionally redundant at the cellular level with a signal transduced by a different biochemical mechanism. Mutations in other Drosophila genes encoding neuronal surface proteins often have no apparent embryonic phenotype and most are viable (Bieber *et al.* 1989; Elkins *et al.*

1990; Grenningloh *et al.* 1990; Nose, Mahajan and Goodman 1992). In at least one case this appears to be due to a redundancy of signalling pathways (Elkins *et al.* 1990). For *DPTP99A* , the lack of existing deficiency mutations spanning the gene also makes conventional strategies to identify lethal complementation groups in this region difficult.

We have adapted a molecular screen for *P* element insertions (Hamilton *et al.* 1991) to identify new insertions adjacent to *DPTP99A* among local transposition events from an element located in 99B. The utility of *P* elements as mutagens is well known (Bingham, Kidwell and Rubin 1982; Rubin, Kidwell and Bingham 1982) and molecular screens for *P* element insertions have allowed mutations to be isolated in cloned DNA segments without regard to their phenotype (Ballinger and Benzer 1989; Kaiser and Goodwin 1990; Hamilton *et al.* 1991). *P* elements have recently been shown to preferentially transpose to new sites in the vicinity of their original location (Hawley *et al.* 1988; Tower *et al.* 1993; Zhang and Spradling 1993) and an elevated transposition frequency can still be observed at distances of the order of a hundred kb (Tower *et al.* 1993). These data suggested that an efficient directed mutagenesis strategy could be developed by using a *P* element cytologically mapped near the target gene as a starting point for local transposition. Collections of mapped single insert lines have now been reported by several research groups (O'Kane and Gehring 1987; Cooley, Kelley and Spradling 1988; Bier *et al.* 1989; Bellen *et al.* 1989) and many are widely available from either individual laboratories or stock centers.

An additional advantage of *P* elements as mutagens is that flanking sequence deletions can be recovered following mobilization of the transposon (Voelker *et al.* 1984; Daniels *et al.* 1985; Salz, Cline and Schedl 1987; Cooley, Thompson and Spradling 1990; Engels *et al.* 1990). Such imprecise excisions of a nearby *P* element can provide a means to mutate genes refractory to direct *P* element insertion. An efficient mutagenesis strategy based on local transposition should thus include the screening of a large region

around the target gene for insertion events, so that insertions flanking the gene can be used to generate excision mutations.

In this paper, we provide two examples of mutagenesis by local transposition. First, we describe the isolation of *encumbered* (*enc*), a new locus affecting morphogenesis of the wing imaginal disc that was uncovered by a local transposition from 45D1-2 to 45C1-2. Second, we describe the directed mutagenesis of *DPTP99A* using a stepwise molecular screen. *P* element insertions adjacent to the gene were generated by local transposition from an element in 99B and identified by screening 35 kb of genomic sequence from the locus. An apparent protein-null deletion was generated by imprecise excision of one of these elements.

MATERIALS AND METHODS

Drosophila stocks: Single-insertion lines of *PlacW* , a *P* element derivative that contains a partially complementing w^+ minigene and bacterial plasmid sequences (Bier *et al.* 1989), were maintained as homozygous stocks in a w^- background. Line 4.09 carries a *PlacW* element at 45D1-2 on the second chromosome (Hamilton *et al.* 1991). Two equivalent isolates of a *PlacW* insertion at 99B1-3, lines P1445 and P1446, were obtained from the Bloomington Stock Center as starting sites for local transpositions into *DPTP99A*. The immobile element *P[Δ2-3, ry*$^+$*]*(99B), referred to below as *Δ2-3* (Robertson *et al.* 1988),was used as the transposase source in all experiments. The chromosome *TM2, Ubx Δ2-3* was provided by J. Merriam. Other chromosomes and variants are described by Lindsley and Zimm (1992).

Local transposition screen on the second chromosome: The single *PlacW* element in line 4.09 was mobilized by mating *w; PlacW* females to *S / In(2L+2R)Cy, Cy E(S) Kpn; Sb Δ2-3 / TM2, Ubx Δ2-3* males. Dysgenic male offspring with the genotype *w; PlacW / In(2L+2R)Cy, Cy E(S) Kpn; + / Sb Δ2-3* were mated to *w; S / CyO*.

Phenotypically w^+; S^+ F2 males were crossed back to *w; S / CyO* and their *w; PlacW / CyO* progeny were inbred to establish stable lines.

Local transposition screens on the third chromosome: Lines P1445 and P1446 are equivalent and were used interchangeably. The *PlacW* element was mobilized in male germlines by mating females to *shi; Sb Δ2-3 / TM2, Ubx Δ2-3* males or by mating *PlacW* animals of either sex to *w; Sb Δ2-3 / TM2, Ubx Δ2-3*. Dysgenic males were mated 1-2 to a vial with *w; TM3 / TM6B, Tb Hu* females. Selected F2 animals were crossed back to *w; TM3 / TM6B, Tb Hu* to create stable lines. Dysgenesis products from female germlines were established similarly, except that only *w; Sb Δ2-3 / TM2, Ubx Δ2-3* was used to introduce the transposease source. Plasmid rescues of transposon-flanking sequences were as described (Hamilton *et al.* 1991), except that libraries were made by complete digestion with Eco RI.

Generating deletions by imprecise excision: Transposons in the double insert lines CB24 and CB27 (described in text) were mobilized by matings to Δ2-3 as described for the third chromosome screen above, except that all dysgeneses were in male germlines and balanced F2 animals were selected for reduction or loss of w^+ function. Lines established from selected F2 males were analyzed for deletions into the DPTP99A gene by genomic DNA gel blot hybridization against a cDNA probe. The breakpoint of the HA64 deletion that lies within the gene was further mapped by gel blot hybridizations of genomic DNAs, digested with a panel of restriction enzymes, to a variety of genomic and cDNA probes. The other HA64 breakpoint apparently lies outside of the cloned region. We do not detect a cytological deficiency associated with HA64.

Immunohistochemistry and other methods: Immunohistochemistry on whole mount Drosophila embryos using HRP-conjugated secondary antibodies was performed as described (Elkins *et al.* 1990) except that signal detection was enhanced in most experiments by adding NiCl to 0.03% in the staining reaction. Monoclonal antibody

6C8C3 is directed against an amino terminal portion of DPTP99A and has been described previously (Tian *et al.* 1991). The antibody used to visualize CNS axons is BP102 (expression pattern shown in Elkins *et al.* 1990). PNS axons were visualized using MAb 22C10 (Fujita *et al.* 1982). The anti-fasciclin MAbs were: 6D8 (fasciclin I; Hortsch and Goodman 1990); 1D4 (fasciclin II); 2D5 (fasciclin III; Patel, Snow and Goodman 1987). All MAbs were obtained from C.S. Goodman and colleagues. Molecular biology procedures were by standard methods or as described in Sambrook, Fritsch and Maniatis (1989).

RESULTS

Mobilization of P4.09 and isolation of *encumbered*: In an attempt to isolate intracistronic transpositions for a locus at 45D1-2 (Hamilton *et al.* 1991), we generated a series of w^+ transposition derivatives of the *PlacW* strain 4.09 (Fig 1: cross scheme). F2 males were only selected qualitatively for w^+ function, not for altered levels of complementation. Genomic DNA was prepared from 50 w^+ lines, digested with EcoRI, and used for gel blot experiments. Hybridization with a pBR322 probe (corresponding to one end of the transposon) and a *lacZY* probe (derived from the other end) identified new insertions in 20 of the 50 lines (DNA of such lines retains one or both ends of the donor element plus one new band detected with either probe). Of the remaining 30 lines, 28 show no change from the parental line with these probes and two lines contain deletions of one end of the element. Correlation of these gel blot hybridization results with a score of eye color intensity indicated that new transpositions could often be visually identified by an increased level of w^+ complementation. Similar results were obtained in a screen involving transpositions from a lethal *PlacW* insertion at 90D (data not shown).

Although none of the lines examined contained a new insertion within the region of interest at 45D1-2, one line was saved for further analysis because of its visible defects

and adult lethality. Homozygous adults have crumpled wings, often with incised margins, and a variable number of missing thoracic bristles. These adults are hypoactive and infertile, and die prematurely. We call this new locus *encumbered* (*enc*), reflecting the appearance of adults during locomotion. *In situ* hybridization to polytene chromosomes reveals a new *PlacW* insertion at 45C1-2. It is not yet clear whether the phenotypes are due to the new insertion or a small rearrangement of the donor site at 45D1-2, as we have been unable to separate these sites by recombination. However, we have been able to isolate a high frequency of w^-, transposase-induced reversions of the morphological and viability defects, strongly suggesting that these phenotypes are associated with at least one of the insertion sites. This allele of *enc* complements at least one class of lethal excisions generated from the 45D transposon, implying that *enc* is located at 45C. *enc* also complements at least one allele of each of three nearby genes that have similar phenotypes: *apterous*, *bloated* and *vestigial* .

Plasmid rescue screens for *PlacW* insertions near *DPTP99A*: We began screening for *P*-element insertions at *DPTP99A* using a plasmid rescue and hybridization strategy (Hamilton *et al.* 1991) . To identify insertions within a region of interest, genomic DNA from pools of *PlacW* transposant lines is digested with a restriction endonuclease, ligated to form covalent circles, and transformed into *E. coli*. Colonies recovered after drug selection comprise a plasmid library of Drosophila sequences that flank *PlacW* insertion sites. The library is then screened by hybridization to identify pools in which one or more lines contain transposons located within the target region. The crosses used to generate transposant lines for the *DPTP99A* experiments are shown in Figure 2. Transposants were generated from line P1445, which contains a single *PlacW* element that we have mapped to 99B1-3 by *in situ* hybridization to polytene chromosomes. Plasmid libraries rescued from pools of these lines were radiolabeled and hybridized to filters containing a set of genomic and cDNA clones that span 35 kb at the

DPTP99A locus (Figure 3). Probes from each end of this 35 kb region and from the cDNA hybridize to the 99A7-8 doublet of salivary gland polytene chromosomes.

In one experiment, we established and screened stable lines from the darkest-eyed F2 male from each of 766 F1 mating vials (Table 1). Rescued-plasmid libraries from all of these lines (pool size 20 – 140) were hybridized to the *DPTP99A* clones. In this screen, we were unable to identify any insertions in the region spanned by the clones, although two pools exhibited hybridization to clone λg1 that could not be reproduced with subpools or single lines.

To improve the efficiency of screening, we tried to generate F2 populations that were enriched for new transposition events. From a separate set of F1 crosses, we established lines from 170 independent F2 progeny whose eye color appeared distinct from that of age-matched same-sex controls carrying the parental P1445 chromosome over *TM3, Sb* or *TM6B, Tb Hu* (Table 2). The rate of transposition appeared low as measured by scoring eye color changes, and we collected only an average of one transposant per 6.25 vials in this experiment. Rescued-plasmid libraries from 98 of these lines were hybridized to the *DPTP99A* clone set. One library pool of 40 lines hybridized to clone λg1. From this pool we recovered two independent lines, CB24 and CB27, which bear insertions adjacent to the 3' end of the gene. These insertions are both within a 4.3 kb EcoR1 fragment, one end of which is 4 kb 3' to the end of the *DPTP99A* coding sequence (Figure 3). The remaining plasmid rescue pools failed to hybridize to the clone set.

Genomic DNA gel blot screen for insertions near *DPTP99A*: To ensure that we were not missing any insertions by the plasmid rescue strategy, 60 of the negative lines (98 minus a subpool of 18 that had been rescreened to isolate CB24 and CB27 and 20 for which either w^+ was not balanced on the third chromosome or the eye color closely resembled that of the P1445 stock upon rescoring), as well as the remaining 72

untested lines, were screened individually by genomic DNA gel blot hybridization. DNA samples from 132 insertion lines, as well as from P1445 and *w; TM3 / TM6B* controls, were each digested with Eco RI and used for gel blot experiments. Filters were hybridized to several probes spanning 36 kb of the *DPTP99A* locus. These included a cDNA containing the entire protein-coding sequence, and genomic fragments corresponding to 6 kb of 5' flanking and 4.3 kb of 3' flanking sequences (see Fig. 3). Polymorphisms were only detected using the 3' flanking probe, which corresponds to the 4.3 kb genomic Eco RI fragment into which the CB24 and CB27 elements are inserted. This probe detected a polymorphic band in each of two new lines, FA30 and FB15, that were among the 72 lines that had not been included in the plasmid rescue pools (Fig. 4). None of the lines that had tested negative in the plasmid rescue experiment showed polymorphisms with any of the probes.

We also used the gel blots to estimate the number of transposition events represented in the enriched lines by hybridization to a pBR322 probe (Table 3). This detects the Eco RI fragment adjacent to the right end of each intact transposon. Hybridization of the gel blots to labeled pBR322 detects new bands in 109 of the 132 lines examined. Of these 109 lines, 101 appear to be double inserts, having both the parental band (corresponding to the *PlacW* at 99B1-3) and a second, polymorphic band. One additional line shows hybridization to at least three distinct bands. The remaining seven lines do not show the parental band and could result from either single new insertions with loss or rearrangements of the starting transposon.

Characteristics of insertion lines: Two of the double insertion chromosomes, CB24 and FA30, are recessive lethal and a third, CB27, is male sterile. The FB15 double insertion chromosome is homozygous viable and fertile in both sexes. We note that the severity of the insertion phenotype is inversely proportional to the distance between the inserted element and *DPTP99A*. Furthermore, homozygotes for the lethal CB24

chromosome are well stained by a monoclonal antibody (MAb 6C8C3) against the extracellular domain of DPTP99A (Tian *et al.* 1991), suggesting that loss of DPTP99A expression is not the cause of lethality in this line.

We have also assessed the state of the donor element at 99B in the insertion lines. Plasmid rescue of sequences on each side of the transposon allowed us to design PCR primers on each side of the *PlacW* element that produce a unique product in combination with a primer complementary to the *P* terminal repeat. DNA prepared from each line supports the amplification of a product the same size as that from parental line P1445 for the right (plasmid) side of the transposon. In addition, hybridization of a pBR322 probe to genomic DNA from each line shows the same size fragment as the parental line. In contrast, none of the four lines yields a parental-like product from the left (*lacZ*) end: FB15 DNA supports the amplification of an abnormally large product (whose origin we have not determined), while none of the other lines yield any product. Furthermore, hybridization of pBR322 to Bam HI-digested genomic DNAs reveals only a new band larger than the parental band in all cases (data not shown). This suggests that the right end of the donor element has remained intact while the left end or a sequence flanking it has been deleted or modified on each chromosome. It is also worth noting that each of the 99A secondary insertions is in the same orientation, although they have different locations within the Eco RI fragment. This orientation selectivity is consistent with other recent reports on local transposition events (Tower *et al.* 1993; Zhang and Spradling 1993).

Generation of an apparent protein-null allele by imprecise excision: To generate a mutation that would eliminate *DPTP99A* expression, we remobilized the *PlacW* elements in the first two insertion lines, CB24 and CB27. We established 112 lines that had visibly reduced *w* function, 63 lines from CB24 and 49 from CB27. Gel blots of Eco RI-digested genomic DNA were serially hybridized to a cDNA probe and a

5' flanking sequence probe (see Fig. 3) to detect breakpoints of deletions that would remove DPTP99A protein-coding sequences. One line, HA64, was isolated on this basis (Fig. 5). The 5' breakpoint of this deletion lies in the large 5th intron, removing the cytoplasmic PTP domains but leaving the exons encoding the N-terminal extracellular and transmembrane domains intact.

To examine whether *DPTP99A*^{HA64} homozygotes might express a truncated DPTP99A protein, we stained embryo collections with MAb 6C8C3, which recognizes the extracellular domain. Virgin females of the genotype *w*; *HA64* / + were collected and mated to males of the same genotype. Embryos were aged to 12-15 hrs at 25°C before fixation and stained with the MAb using nickel-enhanced HRP immunohistochemistry. Approximately one-quarter of the embryos collected from HA64 / + parents failed to stain with 6C8C3 (Table 4). In contrast, parallel experiments with collections from either *w*; CB24 / *TM3* or *w*; + / + showed strong staining in every embryo. The distinction in staining is illustrated in Fig. 6.

We have examined the morphology of the embryonic CNS in HA64 homozygotes by staining embryos with several MAbs that detect epitopes expressed on neuronal cell surfaces. Dissected ventral nerve cords and body walls from stained whole-mount HA64 homozygous embryos were compared to those from their heterozygous and wild-type sibs. We saw no obvious defect in homozygous HA64 embryos stained for epitopes expressed on most or all CNS (MAb BP102) or PNS (22C10) axons. For CNS axons, which are very densely packed, such an assay can only detect fairly severe disruptions in the axon array. We also examined specific CNS axon bundles using MAbs against the three fasciclin proteins, which are expressed on different axon subsets (Patel *et al.* 1987; Zinn, McAllister and Goodman 1988; Grenningloh, Rehn and Goodman 1991; McAllister, Goodman and Zinn 1992). The fasciclin II MAb allows visualization of most

motor axon bundles in the periphery. No alterations were observed in HA64 embryos using any of these markers.

DISCUSSION

Tower *et al.* (1993) have recently shown that *P* elements preferentially transpose to nearby chromosomal sites. In this paper, we demonstrate that local transposition can be used as a tool for efficient directed mutagenesis by tagging a new genetic locus based on its visible phenotypes and by isolating a mutation in *DPTP99A* solely on the basis of molecular criteria. Our results show that genes located as far as one letter division of the standard polytene chromosome map away from a *P* element insertion site can be targeted by local transposition. By extrapolation to other regions of the genome, this indicates that the current density of mapped *P* element single insertions (approximately 1500 are currently available from the Bloomington Stock Center) is sufficient to attempt directed mutageneses of most Drosophila genes. These results also suggest that a small array of selected *P* element lines could be used to saturate the *P* element-mutable sites within a defined genomic region in an efficient manner.

Isolation of *encumbered*: In a small local transposition screen of 50 F2 lines, we isolated *encumbered*, a new genetic locus associated with a *PlacW* transposition from 45D1-2 to 45C1-2. *Encumbered* is required for proper morphogenesis in the wing imaginal disc. The structures primarily affected by *enc* at the gross anatomical level are the wings themselves and thoracic bristles derived from the wing disc. Bristles derived from other imaginal cells are not affected. The adult lethality is probably due to other sites of action. The phenotypes present in *enc* are reminiscent of *apterous* and *vestigial*, both of which complement *enc*, and it will be of interest to assess possible genetic interactions among these loci.

Directed mutagenesis of *DPTP99A* : We established and screened a total of 938 *PlacW* lines derived from an insertion at 99B1-3. Of these, 766 lines were established from males that had the darkest eye color in their respective vials, but were not compared to controls. In similar but separate crosses, an additional 170 lines were established from F2 animals whose eye color appeared to differ significantly from that of age-, sex-, and genotype- matched animals carrying the parental chromosome. All four of the insertions isolated near the *DPTP99A* target gene were found in this smaller collection. This highlights the importance of following changes in expression of a visible marker on the transposon to the efficiency of the screen. The 170 selected P1445 transposants were likely enriched at least 6-fold for new insertions relative to the more naive screen, containing new transposon sites in approximately 80% of the lines examined. In a similar experiment with a *P1wB* element at 99A5-6, transposant lines could not be enriched for additional insertions by scoring eye color due to the high level of w^+ function provided by a single copy of the original insert. Hybridization of genomic DNA gel blots to pBR322 revealed only 11 new transposon sites among 60 lines, or 18% (B. A. H., unpublished results.) The ability to effectively discriminate between zero, one, and more copies makes the partially-complementing w^+ minigene in the *PlacW* element a valuable marker for local transposition experiments (Bier *et al.* 1989; Golic and Lindquist 1989).

Since all four of the new *PlacW* insertions that we obtained near *DPTP99A* are located in a 4.3 kb EcoR1 fragment 3' to the coding sequence, we induced an excision of one of these elements that generated a null deletion mutant. Interpretation of the phenotype of this null allele, $DPTP99A^{HA64}$, is complicated by the recessive lethality conferred by the original insertion, CB24. Our results suggest that a portion of one or more loci mutable to lethality (by the CB24 and FA30 insertions) or to male sterility (by the CB27 insertion) may lie within the 4.3 kb EcoR1 fragment. Thus, we cannot determine from our data whether a mutation affecting only *DPTP99A* would confer

lethality. Nevertheless, we have shown that there is no gross visible defect in the embryonic CNS associated with a null mutation in *DPTP99A*.

General considerations for directed mutagenesis screens. The ability to select new transposition events by visually scoring changes in w^+ function conferred by an increase in *PlacW* copy number was essential for the success of our screen. Our results also suggest that because the coding sequences of a gene may prove refractory to transposon insertion, intronic and flanking region DNA should also be used as a target in local transposition screens. Excision derivatives of the initial insertions can then be generated to create null alleles.

Several methods allow identification of lines bearing insertions in a desired region. Genomic DNA gel blots are useful for monitoring the rate of transposant recovery and for screening small numbers of transposant lines for desired insertions. We isolated two insertions adjacent to *DPTP99A* by analyzing DNA gel blots of 132 lines. Mass-screening strategies based on plasmid rescue or PCR, however, allow the recovery of lower probability events from larger pools of insertions (Ballinger and Benzer 1989; Kaiser and Goodwin 1990; Hamilton *et al.* 1991). We used plasmid rescue from a total of 866 lines to isolate two additional insertions near *DPTP99A*.

Zhang and Spradling (1993) have reported that the distribution of transposon insertion sites differs between those generated in male germline and those generated in female germlines, with males being more prone to hot and cold spots. This suggests that while male germline transpositions may be used to quickly saturate relative hot spots, female germline transposition should also be employed in order to obtain a broader distribution of insertion sites. In our screen, however, the only insertion generated by transposition in the female germline was localized to the same 4.3 kb EcoR1 fragment as the three insertions generated in male germlines.

Our isolation of four insertions near the target gene from among 170 selected lines (936 total lines) represents a large increase in efficiency relative to molecular screens that rely on interchromosomal transpositions (Ballinger and Benzer 1989; Kaiser and Goodwin 1990; Hamilton *et al.* 1991). For example, in an earlier plasmid rescue screen for interchromosomal transpositions into sequences corresponding to a collection of nearly 700 cDNA clones, insertions were isolated for only about 1% of the target sequences among 760 transposant lines carrying autosomal insertions (Hamilton *et al.* 1991). The 4.3 kb EcoR1 fragment containing the four insertion sites is apparently a local hot spot for *P* element insertion within the 35 kb of screened genomic *DPTP99A* DNA. Our ability to identify the insertions does not mean that this DNA segment is a prominent third chromosome hot spot, however, since no *P* elements had been identified at this site among the collections to which we had access.

Our identification of the *encumbered* locus demonstrates the utility of local transposition for transposon tagging genetically defined sites prior to cloning. In addition, this suggests that local transposition screening may allow efficient saturation of *P*-mutable sites within a defined region of the genome. This may facilitate the genetic analysis of genomic regions refractory to other approaches (Kongsuwan, Dellavalle and Merriam 1986).

Possible functions of DPTP99A in Drosophila embryonic development.

The DPTP99A proteins have an extracellular domain composed of three FN repeats, and two cytoplasmic PTP homology domains (Yang *et al.* 1991; Tian *et al.* 1991; Hariharan, Chuang and Rubin 1991). It is closely related (63% identity in the first PTP domain) to a human R-PTP, HPTPγ (Kreuger, Streuli and Saito 1990). DPTP99A protein is predominantly localized to CNS axons in the embryo, and is apparently present on all axon bundles. It is expressed from the beginning of axonogenesis (stage 12/5 of Klämbt, Jacobs and Goodman 1991) and this expression is maintained until at least stage 16.

We do not detect a CNS phenotype in embryos homozygous for the null *DPTP99A*^{HA64} mutation. However, our ability to detect a phenotype is limited by the available antibodies, and MAbs recognizing all CNS axons can only reveal major defects. Although DPTP99A protein is expressed on most or all axons, its activity may be uniquely required in only a subset of axon bundles that is not visualized with the anti-fasciclin MAbs. Alternatively, perhaps DPTP99A activity is required only for later events such as synaptogenesis rather than for axon outgrowth and guidance.

The lack of a CNS phenotype for HA64 could also suggest that the axonal R-PTPs may be functionally redundant. However, the Drosophila R-PTPs are more closely related to human cognates than to each other (Strueli *et al.* 1989; Tian *et al.* 1991) and this evolutionary conservation argues against a complete redundancy for both ligand and substrate specificity. Perhaps DPTP99A transduces a signal to a unique set of cytoplasmic substrates , but the ultimate effect of this signal in any given cell is redundant with a signal produced by another pathway.

If DPTP99A activity is functionally redundant, analysis of double mutant embryos in which a null *DPTP99A* mutation is combined with mutations in other genes may help to elucidate its role in CNS development. We are currently attempting to use local transposition to generate mutations in two other Drosophila R-PTP genes. Mutations in genes encoding other molecules that interact with phosphotyrosine signaling pathways might also produce interesting double mutant phenotypes when combined with *DPTP99A*^{HA64}. The *fasciclin I* mutation has no known embryonic phenotype as a single mutation, but causes severe disruptions of CNS axon pathways when combined with a mutation in the *D-abl* tyrosine kinase gene (Elkins *et al.* 1990). We are currently combining the *DPTP99A*^{HA64} mutation with *fasciclin I, II, III,* and *D-abl* mutations in order to determine whether such a conditional phenotype can be observed.

We thank K. Golic and A. Spradling for discussing their data on local transposition of *P* elements prior to publication, W. Chia and colleagues for the sequences of *DPTP99A* intron/exon boundaries, K. Matthews and the Bloomington Stock Center for *PlacW* insertion lines P1445 and P1446, G. Rubin and colleagues for *P1wB* line A70, C. S. Goodman and colleagues for monoclonal antibodies, D. Smoller for P1 clone 2-1, and E. B. Lewis and the Mid-America Drosophila stock center for *ap*, *blo*, and *vg* alleles. We thank S. Celniker and members of our laboratory for comments on this manuscript. This work was supported by NIH grant #NS28182 to K. Z., as well as by Basil O'Connor Starter Scholar Research Award #5-816 from the March of Dimes Birth Defects Foundation, a Pew Scholars Award, and a McKnight Scholars Award. B. A. H. was supported in part by National Research Service Award #5 T32 HG00021-02 from the National Center for Human Genome Research.

LITERATURE CITED

Ballinger, D. G. and S. Benzer, 1989 Targeted gene mutations in Drosophila. Proc. Natl. Acad. Sci. USA **86:** 9402-9406.

Bellen, H., C. J. O'Kane, C. Wilson, U. Grossniklaus, R. Kurth-Pearson and W. J. Gehring, 1989 P-element-mediated enhancer detection: a versatile method to study development in Drosophila. Genes Dev. **3:** 1288-1300.

Bieber, A. J., P. M. Snow, M. Hortsch, N. H. Patel, J. R. Jacobs, Z. R. Traquina, J. Schilling and C. S. Goodman, 1989 Drosophila neuroglian: a member of the immunoglobulin superfamily with extensive homology to the vertebrate neural adhesion molecule L1. Cell **59:** 447-460.

Bier, E., H. Vaessin, S. Shepherd, K. Lee, K. McCall, S. Barbel, L. Ackerman, R. Carretto, T. Uemura, E. Grell, L. Y. Jan and Y. N. Jan, 1989 Searching for pattern and mutation in the Drosophila genome with a P-*lacZ* vector. Genes Dev. **3:** 1273-1287.

Bingham, P. M., M. G. Kidwell and G. M. Rubin, 1982 The molecular basis of P-M hybrid dysgenesis: the role of the P element, a P strain-specific transposon family. Cell **29:** 995-1004.

Charbonneau, H. and N. K. Tonks, 1992 1002 protein phosphatases. Annu. Rev. Cell Biol. **8:** 463-493.

Cooley, L., R. Kelley and A. Spradling, 1988 Insertional mutagenesis of the Drosophila genome with single P elements. Science **239:** 1121-1128.

Cooley, L., D. Thompson and A. C. Spradling, 1990 Constructing deletions with defined endpoints in Drosophila. Proc. Natl. Acad. Sci. USA **87:** 3170-3173.

Daniels, S. B., M. McCarron, C. Love and A. Chovnick, 1985 Dysgenesis-induced instability of rosy locus transformation in Drosophila melanogaster: analysis of

excision events and the selective recovery of control element deletions. Genetics **109:** 95-117.

Elkins, T., K. Zinn, L. McAllister, F. M. Hoffmann and C. S. Goodman, 1990 Genetic analysis of a Drosophila neural cell adhesion molecule: interaction of *fasciclin I* and Abelson tyrosine kinase mutations. Cell **60:** 565-575.

Engels, W. R., D. M. Johnson-Schlitz, W. B. Eggleston and J. Sved, 1990 High-frequency P element loss in Drosophila is homolog dependent. Cell **62:** 515-525.

Fischer, E. H., H. Charbonneau and N. K. Tonks, 1991 Protein tyrosine phosphatases: a diverse family of intracellular and transmembrane enzymes. Science **253:** 401-406.

Fujita, S. C., S. L. Zipursky, S. Benzer, A. Ferrus and S. C. Shotwell, 1982 Monoclonal antibodies against the Drosophila nervous system. Proc. Natl. Acad. Sci. USA **79:** 7929-7933.

Gertler, F. B., R. L. Bennett, M. J. Clark and F. M. Hoffmann, 1989 Drosophila *abl* tyrosine kinase in embryonic CNS axons: a role in axonogenesis is revealed through dosage-sensitive interactions with *disabled*. Cell **58:** 103-113.

Golic, K. G. and S. Lindquist, 1989 The FLP recombinase of yeast catalyzes site-specific recombination in the Drosophila genome. Cell **59:** 499-509.

Grenningloh, G., A. J. Bieber, E. J. Rehm, P. M. Snow, Z. Traquina, M. Hortsch, N. H. Patel and C. S. Goodman, 1990 Molecular genetics of neuronal recognition in Drosophila: evolution and function of immunoglobulin superfamily cell adhesion molecules. Cold Spring Harbor Symp. Quant. Biol. **55:** 327-340.

Grenningloh, G., E. J. Rehn and C. S. Goodman, 1991 Genetic analysis of growth cone guidance in Drosophila: fasciclin II functions as a neuronal recognition molecule. Cell **67:** 45-57.

Hamilton, B. A., M. J. Palazzolo, J. H. Chang, K. V. Raghavan, C. A. Mayeda, M. A. Whitney and E. M. Meyerowitz, 1991 Large scale screen for transposon insertions into cloned genes. Proc. Natl. Acad. Sci. USA **88:** 2731-2735.

Hariharan, I., P.-T. Chuang and G. M. Rubin, 1991 Cloning and characterization of a receptor-class phosphotyrosine phosphatase gene expressed on central nervous system axons in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **88:** 11266-11270.

Hawley, R. S., R. A. Steuber, C. H. Marcus, R. Sohn, D. M. Baronas, M. L. Cameron, A. E. Zitron and J. W. Chase, 1988 Molecular analysis of an unstable P element insertion at the singed locus of *Drosophila melanogaster*: evidence for intracistronic transposition of a P element. Genetics **119:** 85-94.

Hortsch, M. and C. S. Goodman, 1990 Drosophila fasciclin I, a neural cell adhesion molecule, has a phosphatidylinositol lipid membrane anchor that is developmentally regulated. J. Biol. Chem. **265:** 15104-15109.

Kaiser, K. and S. F. Goodwin, 1990 "Site-selected" transposon mutagenesis of Drosophila. Proc. Natl. Acad. Sci. USA **87:** 1686-1690.

Klämbt, C., J. R. Jacobs and C. S. Goodman, 1991 The midline of the Drosophila central nervous system: a model for the genetic analysis of cell fate, cell migration and growth cone guidance. Cell **64:** 801-816.

Kongsuwan, K., R. P. Dellavalle and J. R. Merriam, 1986 Deficiency analysis of the tip of chromosome 3R in *Drosophila melangaster*. Genetics **112:** 539-550.

Kreuger, N. X., M. Streuli and H. Saito, 1990 Structural diversity and evolution of human receptor-like protein tyrosine phosphatases. EMBO J. **9:** 3241-3252.

Lindsley, D. L. and G. G. Zimm, 1992 *The Genome of Drosophila melanogaster*. Academic Press, Inc.: San Diego, CA. pp.

Maness, P. F., M. Aubury, C. G. Shores, L. Frame and K. H. Pfenninger, 1988 *c-src* gene product in developing rat brain is enriched in nerve growth cone membranes. Proc. Natl. Acad. Sci. USA **85:** 5001-5005.

McAllister, L., C. S. Goodman and K. Zinn, 1992 Dynamic expression of the cell adhesion molecule fasciclin I during embryonic development in Drosophila. Development **115:** 267-276.

Nose, A., V. B. Mahajan and C. S. Goodman, 1992 Connectin: a homophilic cell adhesion molecule expressed on a subset of muscles and the motoneurons that innervate them in Drosophila. Cell **70:** 553-567.

O'Kane, C. J. and W. J. Gehring, 1987 Detection *in situ* of genomic regulatory elements in Drosophila. Proc. Natl. Acad. Sci. USA **84:** 9123-9127.

Patel, N. H., P. M. Snow and C. S. Goodman, 1987 Characterization and cloning of Fasciclin III: a glycoprotein expressed on a subset of neurons and axon pathways in Drosophila. Cell **48:** 975-988.

Robertson, H. M., C. R. Preston, R. W. Phillis, D. M. Johnson-Schlitz, W. K. Benz and W. R. Engels, 1988 A stable genomic source of P element transposase in *Drosophila melanogaster*. Genetics **118:** 461-470.

Rubin, G. M., M. G. Kidwell and P. M. Bingham, 1982 The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations. Cell **29:** 987-994.

Salz, H. K., T. W. Cline and P. Schedl, 1987 Functional changes associated with structural alterations induced by mobilization of a P element inserted in the *Sex-lethal* gene of Drosophila. Genetics **117:** 221-231.

Sambrook, J., E. F. Fritsch and T. Maniatis, 1989 *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory: Cold Spring Harbor, New York. pp.

Strueli, M., N. X. Kreuger, A. Y. M. Tsai and H. Saito, 1989 A family of receptor-linked protein tyrosine phosphatases in humans and Drosophila. Proc. Natl. Acad. Sci. USA **86:** 8698-8702.

Tian, S.-S., P. Tsoulfas and K. Zinn, 1991 Three receptor-linked protein-tyrosine phosphatases are selectively expressed on central nervous system axons in the Drosophila embryo. Cell **67:** 675-685.

Tower, J., G. H. Karpen, N. Craig and A. C. Spradling, 1993 Preferential transposition of Drosophila P elements to nearby chromosomal sites. Genetics **133:** 347-359.

Voelker, R. A., A. L. Greenleaf, H. Gyurkovics, G. B. Wisely, S.-M. Huang and L. L. Searles, 1984 Frequent imprecise excision among reversions of a P element-caused lethal mutations in Drosophila. Genetics **107:** 279-294.

Yang, X., K. T. Seow, S. M. Bahri, S. H. Oon and W. Chia, 1991 Two Drosophila receptor-like tyrosine phosphatase genes are expressed in a subset of developing axons and pioneer neurons in the embryonic CNS. Cell **67:** 661-673.

Zhang, P. and A. C. Spradling, 1993 Efficient and dispersed local P element transposition from Drosophila females. Genetics **133:** 361-373.

Zinn, K., L. McAllister and C. S. Goodman, 1988 Sequence analysis and neuronal expression of fasciclin I in grasshopper and Drosophila. Cell **53:** 577-587.

Table 1: Naive w^+ F2 screens

Germline sex, *Δ2-3* source	F1 matings	#F1 / vial	F2 mated	Plasmid screen
male, *Sb Δ2-3*	182	1	157	157
male, *TM2, Ubx Δ2-3*	37	1	22	22
male subtotal	219	-	179	179
female, *Sb Δ2-3*	31	1	31	30
female, *TM2, Ubx Δ2-3*	750	1	641	559
female subtotal	781	-	672	589
TOTAL	1000	-	859	768

Table 2: F2 screens enriched by eye color scoring

Germline sex, *Δ2-3* source	F1 matings	Average# F1 / vial	Fertile F1 matings	F2 selected	Plasmid screen	Gel blot screen	Total screened	Lines isolated
male, *Sb Δ2-3*	753	1.9	702	116	57	76	87	1
male, *TM2, Ubx Δ2-3*	≥200	1.9	≥160	54	18	6	24	2
male subtotal	≥953	1.9	≥862	170	57	82	111	3
female, *Sb Δ2-3*	327	1.4	310	56	28	39	45	
female, *TM2, Ubx Δ2-3*	215	1.3	206	19	13	11	14	
female subtotal	542	1.3	516	75	41	50	59	1
TOTAL	≥1495	-	≥1378	245	98	132	170	4

Table 3: Genomic DNA gel blot hybridization to pBR322.

Germline sex	Δ2-3 source	#F2 scored	No change	Single insert	Double	Triple
male	*Sb Δ2-3*	76	9	4	60	3
male	*TM2, Ubx Δ2-3*	6	1	1	3	1
female	*Sb Δ2-3*	39	7	2	30	0
female	*TM2, Ubx Δ2-3*	11	8	1	2	0
TOTAL		132	25	8	95	4

TABLE 4. Staining of whole mount embryos collected from HA64/+ parents

Batch	Parental genotype	#stained	#unstained	%unstained
1	*w; HA64 / +*	363	114	23.9
2	*w; HA64 / +*	245	65	21.0
3	*w; + / +*	138	0	0.0

FIGURE LEGENDS

FIGURE 1. Second chromosome mobilization scheme.

FIGURE 2. Third chromosome mobilization scheme.

FIGURE 3. Map of *DPTP99A* locus. Half-filled triangles represent the insertion site and orientation of isolated *PlacW* insertions: 1, CB24; 2, CB27; 3, FB15. The filled side represents the plasmid end of the element. The open bar shows the sequences removed by the HA64 deletion. The 3' end of the deletion has not been mapped. EcoRI sites are indicated by short vertical lines on the unlabeled bar. The site marked with * is not present on TM3. The cDNA clone used in these experiments is truncated at a SnaBI site in the 3' untranslated region. Positions of exons relative to the genomic map are from (Yang *et al.* 1991). Genomic clones in λGEM-11 were isolated by hybridization with fragments of the cDNA. Probe A is derived from a 6.5 kb SalI-ScaI fragment from a bacteriophage P1 clone of the region. Probe B is a 4.3 kb Eco RI fragment isolated from λg1.

FIGURE 4. Screening insertions with DNA gel blots: isolation of FA30 and FB15. Gel blots of Eco RI-digested genomic DNA from each of 132 independent PlacW insertion lines generated from a single insertion at 99B were hybridized with probes from the DPTP99A locus. Several lanes of one gel hybridized to a labeled 4.3 kb Eco RI fragment flanking the 3' end of the gene (probe B in figure 3) are shown. The lane marked "a" corresponds to line FA30 and "b" to line FB15.

FIGURE 5. Excision line HA64 contains a deletion into the 5th intron of *DPTP99A*. ***A.*** Gel blots of Eco RI-digested DNA from several excision lines hybridized to the cDNA probe shown in figure 4. Lane 2 is line HA64. ***B.*** The HA64 deletion breakpoint lies within the 5th intron. DNA from lines CB24 / TM3 (lanes 1 and 4), HA64 / TM3 (2 and

5), and P1445 (3 and 6) was digested with Hind III (lanes 1-3) or Eco RI (lanes 4-6) and used to prepare gel blots. Filters were hybridized with a PCR product from cDNA that corresponds to the transmembrane segment (amino acids 323 to 406 from Tian *et al.* 1991).

FIGURE 6. Detection of DPTP99A protein in whole mount embryos stained with MAb 6C8C3. The embryos shown in (A) and (B) are from the same 12-15 hr. old embryo collection from a *DPTP99A*^{HA64}/+ stock, and are of approximately the same age. ***A.*** DPTP99A expression is detected by MAb 6C8C3. ***B.*** The lack of 6C8C3 staining shows that DPTP99A is not expressed in homozygous HA64 embryos.

Figure 1

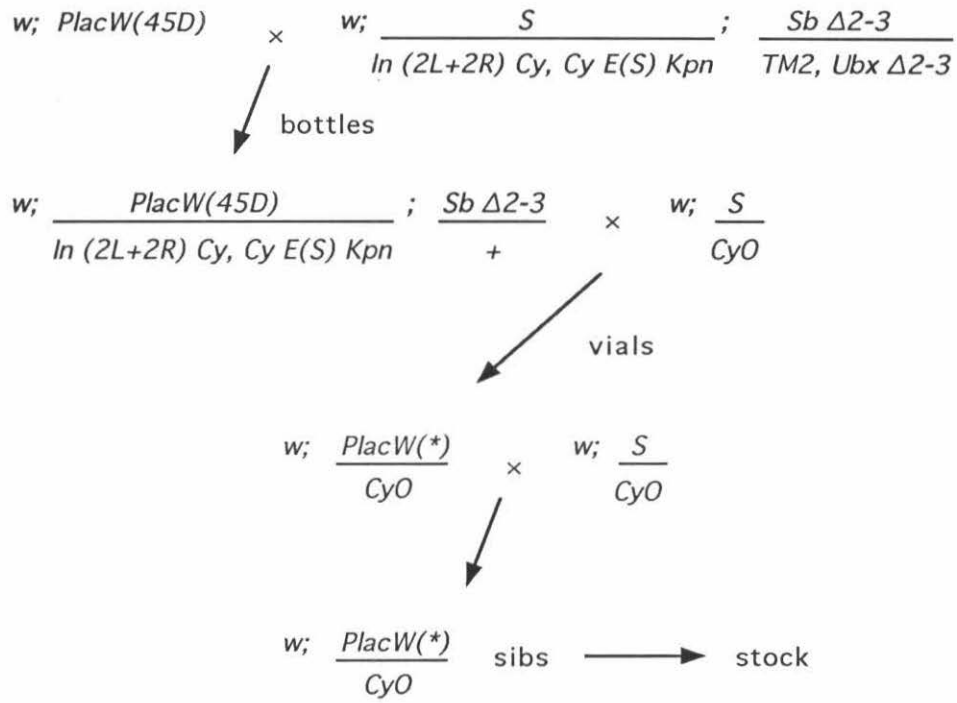

w; PlacW(45D) × w; S / In (2L+2R) Cy, Cy E(S) Kpn ; Sb Δ2-3 / TM2, Ubx Δ2-3
bottles
w; PlacW(45D) / In (2L+2R) Cy, Cy E(S) Kpn ; Sb Δ2-3 / + × w; S / CyO
vials
w; PlacW(*) / CyO × w; S / CyO
w; PlacW(*) / CyO sibs → stock

Figure 2

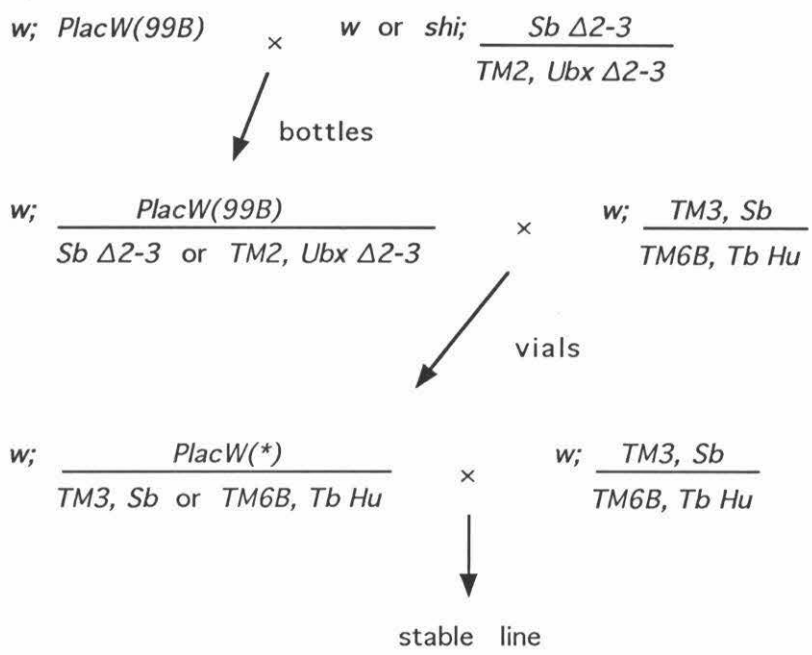
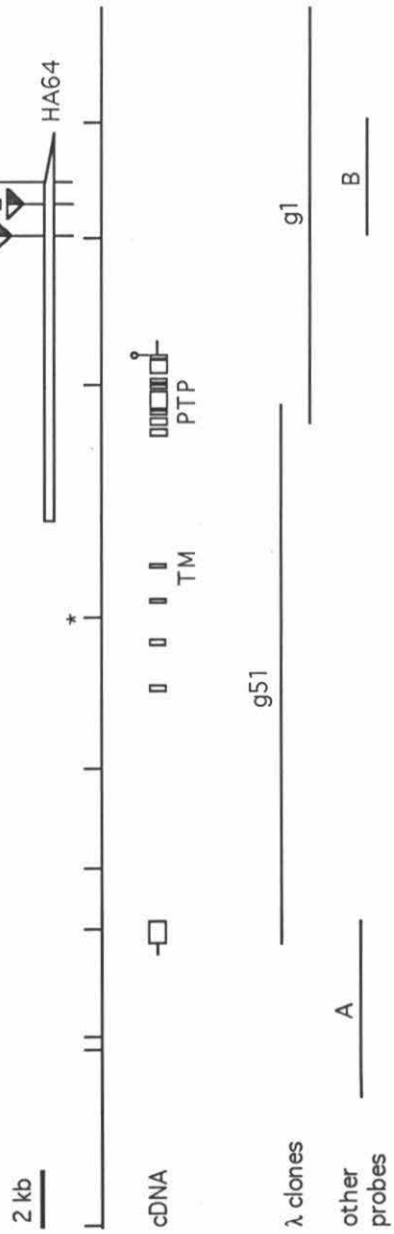
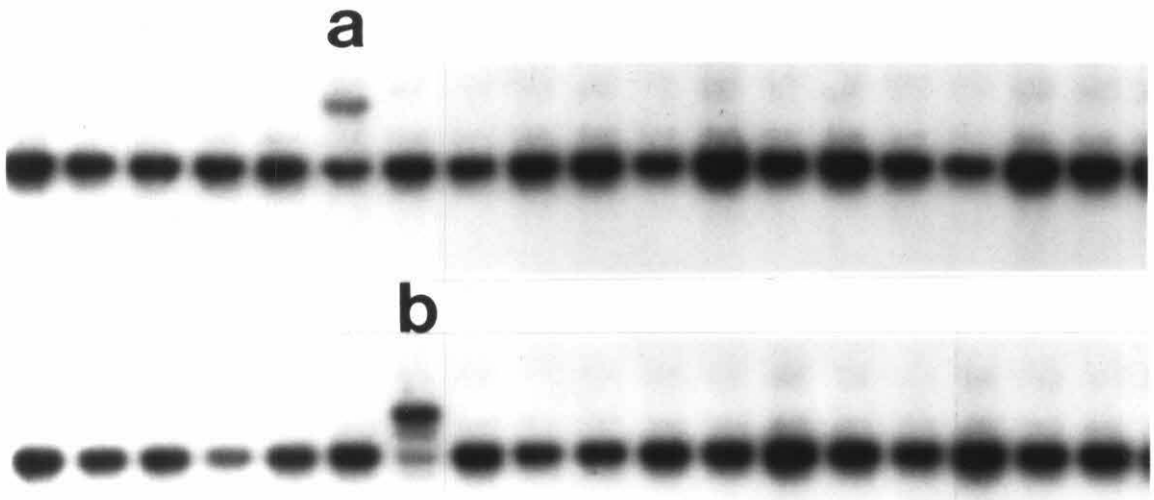
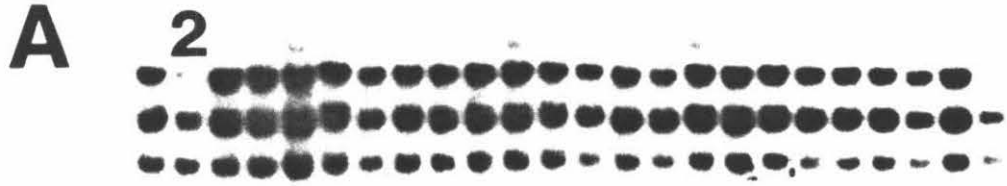
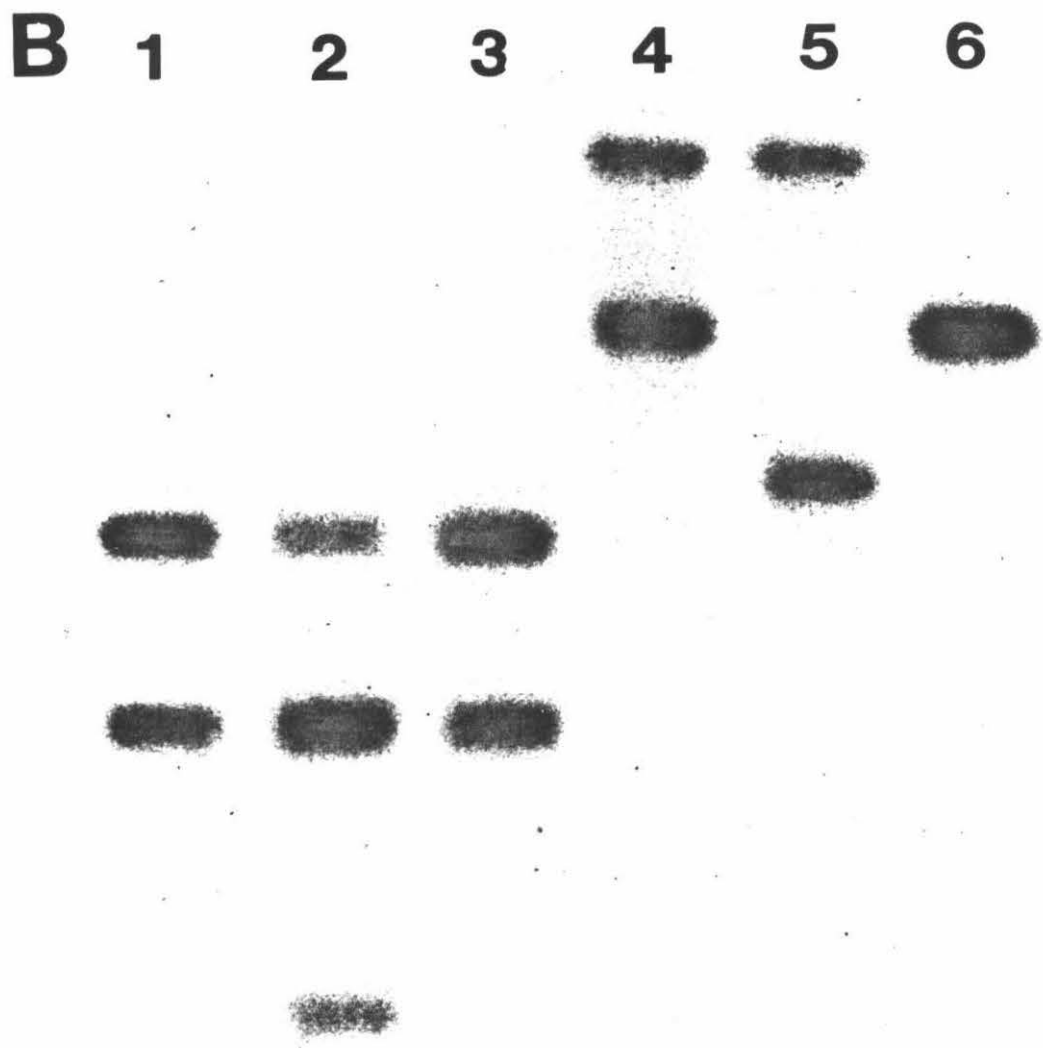

w; PlacW(99B) × w or shi; Sb Δ2-3 / TM2, Ubx Δ2-3
bottles
w; PlacW(99B) / Sb Δ2-3 or TM2, Ubx Δ2-3 × w; TM3, Sb / TM6B, Tb Hu
vials
w; PlacW(*) / TM3, Sb or TM6B, Tb Hu × w; TM3, Sb / TM6B, Tb Hu
stable line

Figure 3




a
b


A
2


B
1
2
3
4
5
6


A
B

Appendix 1

Encumbered is required for proper morphogenesis in the Drosophila wing disc

Encumbered is required for proper morphogenesis in the Drosophila imaginal wing disc

Bruce Hamilton

Summary

I describe in more detail the *encumbered* (*enc*) mutation, which was briefly described in chapter 4. This mutation affects adult structures derived from imaginal wing discs as well as adult longevity and fertility. I propose a candidate cDNA as the affected locus.

Materials and Methods

Drosophila stocks and variants have been described previously (Hamilton et al. 1991; Lindsley and Zimm 1992). All stocks were reared at 22-25°C on standard food. Plasmid rescue of transposon flanking sequences and subsequnt isolation and sequencing of cDNA clones were carried out by described methods (Hamilton et al. 1991; Hamilton, Palazzolo and Meyerowitz 1991; Strathmann et al. 1991). *PlacW* mobilizatin crosses to generate excisions used the same crossing scheme described for the second chromosome screen in Chapter 4, except that w^- progeny were selected in the F2 generation.

Results

Isolation of the encubered^1 allele: The *enc*^{1} mutation was recovered as an exceptional progeny after genetic mobilization of a *PlacW* element on the second chromosome in a screen for intrachromosomal transposants (Chapter 4). The striking phenotype in anesthetized homozygous animals is the tortuously crumpled appearance of the wings (Figure 1). In awake animals, the held-up, distorted wings contribute to locomotive difficulties and relative inactivity that are partially relieved by their removal.

With or without their wings, the animals all but cease to move early in adulthood, clinging to the side of the culture vial until they eventually and prematurely expire. These phenotypes led to the characterization of these flies as being *encumbered* and this has been formalized as a name for the mutant.

The mutant chromosome carries both the donor element at 45D1-2 and a newly inserted element at 45C1-2. Although the parental insertion line shows no phenotype, DNA gel blotting experiments suggest a rearrangement of sequences flanking the donor element. Efforts to separate the two insertion sites by meiotic recombination have been so far unsuccessful. Thus, it is not clear which of these inserion sites is responsible for the phenotypes.

Morphological defects: The degree of wing deformity in enc^1 homozygotes varies from a crumpled umbrella-like appearance, as shown in Figure 1, to a nubbish appearance similar to that described for severe alleles of *vestigial* or *apterous* (Lindsley and Zimm 1992). The milder phenotypes are the more common. In addition to altered topography of the wing blade, incisions of the wing margin are frequent among enc^1 homozygotes. Such incisions also have been in at least some alleles of a variety of Drosophila genes that are mutable to lethality and play a role in neural development, including *cut*, *Notch*, *scalloped*, and *Serrate*.

Thoracic structures derived from the wing discs are also affected by the enc^1 mutation. Bristles, less frequently their sockets, and much less frequently patches of scutelum are deleted or deformed in homozygotes. The bristles are not affected uniformly, ranging from 6% of posterior dorso-central bristles missing up to 73% of anterior notopluerals, as illustrated in Figure 2.

Lethality and sterility: Homozygous adults are rare in the balanced stock. When observed at low population density in vials, these animals tend to become stuck by their deformed wings to the food or to the side of the vial and soon perish. However,

homozygous adults whose wings have been removed with spring scissors under CO_2 anesthesia have significanty higher mortality rates than heterozygous or wild-type controls (Figure 3), indicating that the wing deformity and subsequent immobilization are not the primary cause of the premature adult deaths in these animals.

Matings between homozygous adults are invariably sterile, even when the adults survive for several days. Homozygous females are occasionally fertile when mated to nonmutant males, but their fecundity remains low. Homozygous males appear to be completely sterile. The basis for this sterility has not been investigated.

Excision alleles and revertants: Mutations caused by the insertion of a *P* element can often be reverted at high frequency by remobilization of the element. I generated 40 w^- derivatives of the double insertion chromosome by mating in a stable source of transposase (Robertson et al. 1988) for one generation. Thirty-four of these excision lines were back-crossed to the enc^1 allele. Heterozygous progeny from 25 excision lines showed no wing defect, one produced a less severe phenotype than enc^1 homozygotes, and seven produced progeny indistinguishable from enc^1 homozygotes. The remaining line failed to produce any enc^1 / excision progeny in this cross.

Candidate cDNAs: The genomic sequences flanking each insertion site have been recovered by plasmid rescue and used to isolate cDNA clones from a variety of libraries, including libraries constructed from embryos, adult heads, antennae, and larval imaginal discs. Sequence from cDNAs representing several transcription units flanking the 45D insertion site reveals no open reading frames (B. H. and J. Liao, unpublished results). A single transcribed sequence, which contains a long open reading frame, is detected by RNA gel blots or cDNA library screening with probes derived from the 45C insertion site.

The nucleotide sequence and predicted open reading frame of the 45C transcript are shown in Figure 4. The site of the *PlacW* insertion within 3' untranslated sequence is indicated. This sequence does not show significant similarity to any sequences reported in

current databases. The predicted novel protein is unusual in its high content of so-called PEST sequences and in its high content of charged residues except histidine.

Discussion

Because the *enc*^{1} chromosome contains new lesions at both 45D1-2 and 45C1-2, it is not yet clear which of these sites is responsible for the mutant phenotypes or even whether all the phenotypes are attributable to one locus. However, the recovery of a high frequency of excision-mediated revertants strongly argues for an association between the observed phenotypes and at least one of the inserted *PlacW* elements. One argument in favor of the 45C site is that the suspected 45D rearrangement nevertheless leaves both ends of the donor element intact and does not seem to involve a third insertion site. A rearrangement with these characteristics is unlikely to be functionally reversed by remobilization of the transposons. Furthermore, the variability in morphological phenotype and ability to recover apparent non-complementing lethal alleles are that the *enc*^{1} allele is a hypomorph. This would be consistent with an insertion into 3' untranslated sequence, which could lead to alterations in RNA splicing, stability, or translation efficiency, but probably not to its total blockage. Further elucidation of the molecular biology underlying the *enc* locus will require germline transformation and complementation experiments with cloned DNA segments.

Acknowledgements: I thank J. Liao for sequence analysis of several 45D cDNAs; Ed Lewis and Sue Celniker for assistance with the photograph in figure 1; and G. Rubin, J. Fristrom, and M. McKenna for larval and antennal cDNA libraries.

Literture cited

Hamilton, B. A., M. J. Palazzolo, J. H. Chang, K. V. Raghavan, C. A. Mayeda, M. A. Whitney and E. M. Meyerowitz, 1991 Large scale screen for transposon insertions into cloned genes. Proc. Natl. Acad. Sci. USA **88:** 2731-2735.

Hamilton, B. A., M. J. Palazzolo and E. M. Meyerowitz, 1991 Rapid isolation of long cDNA clones from existing libraries. Nucl. Acids Res. **19:** 1951-1952.

Lindsley, D. L. and G. G. Zimm, 1992 *The Genome of Drosophila melanogaster.* Academic Press, Inc.: San Diego, CA. pp.

Robertson, H. M., C. R. Preston, R. W. Phillis, D. M. Johnson-Schlitz, W. K. Benz and W. R. Engels, 1988 A stable genomic source of P element transposase in *Drosophila melanogaster*. Genetics **118:** 461-470.

Strathmann, M., B. A. Hamilton, C. A. Mayeda, M. I. Simon, E. M. Meyerowitz and M. J. Palazzolo, 1991 Transposon-facilitated DNA sequencing. Proc. Natl. Acad. Sci. USA **88:** 1247-1250.

Figure legends

Figure 1: Photograph of a typical *enc*^{1} homozygote (left) and a homozygote of the parental single *PlacW* line, 4.09 (right).

Figure 2: Thoracic bristles derived from wing discs are not uniformly affected by the *enc*^{1} mutation. Percentages give the frequency with which the indicated bristle is missing in homozygotes.

Figure 3: Survival curve. Homozygous male or virgin female adults and heterozygous male adults were collected from a *w*; *enc*^{1} / *CyO* stock 0-24 hr. after eclosion. Wings were removed from each animal under CO_2 anesthesia. Animals were maintained at 25°C in glass culture vials with standard food. Each vial was started with either 10 males (heterozygotes and 2 homozygote vials) or 10 virgin females (1 homozygote vial). For each genotype, 30 animals were observed periodically for 10 days. No discrepancy in survival rate was noted between male and female homozygotes.

Figure 4: Sequence analysis of a candidate cDNA for *enc*. Consensus nucleotide sequence of three cDNA clones isolated from adult head libraries and conceptual translation

of the long open reading frame. Neither the nucloetide sequence nor the translated amino acid sequence detects significant similarity to entries in the public databases. The integration site of the *PlacW* element is indicated by the arrow.

Figure 2

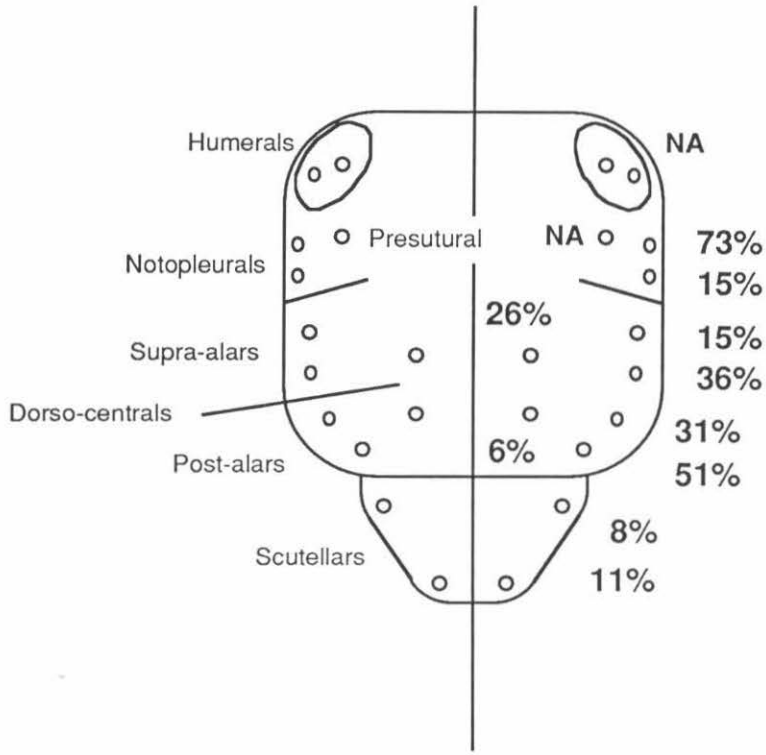

Humerals
NA
Presutural
NA
Notopleurals
73%
15%
26%
Supra-alars
15%
36%
Dorso-centrals
31%
6%
Post-alars
51%
8%
Scutellars
11%

Figure 3

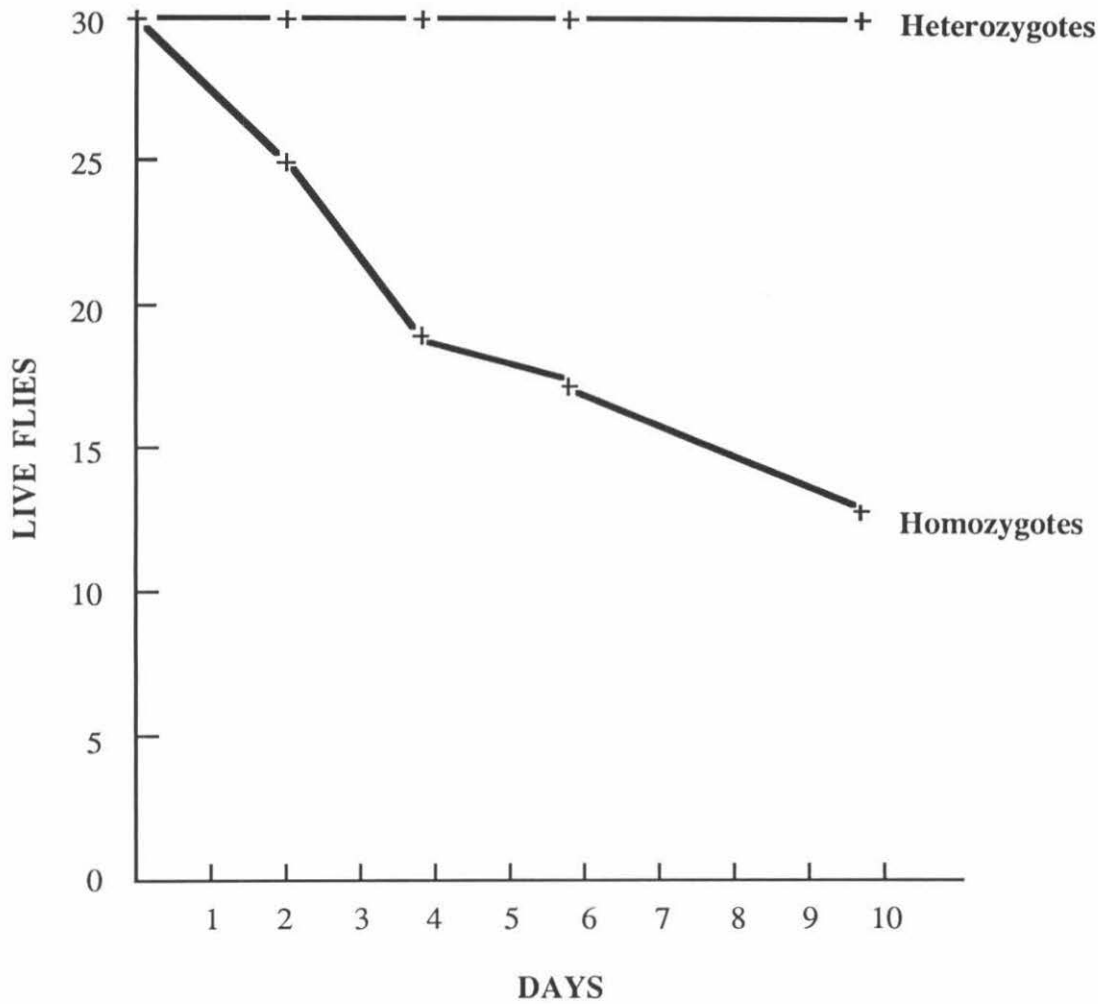
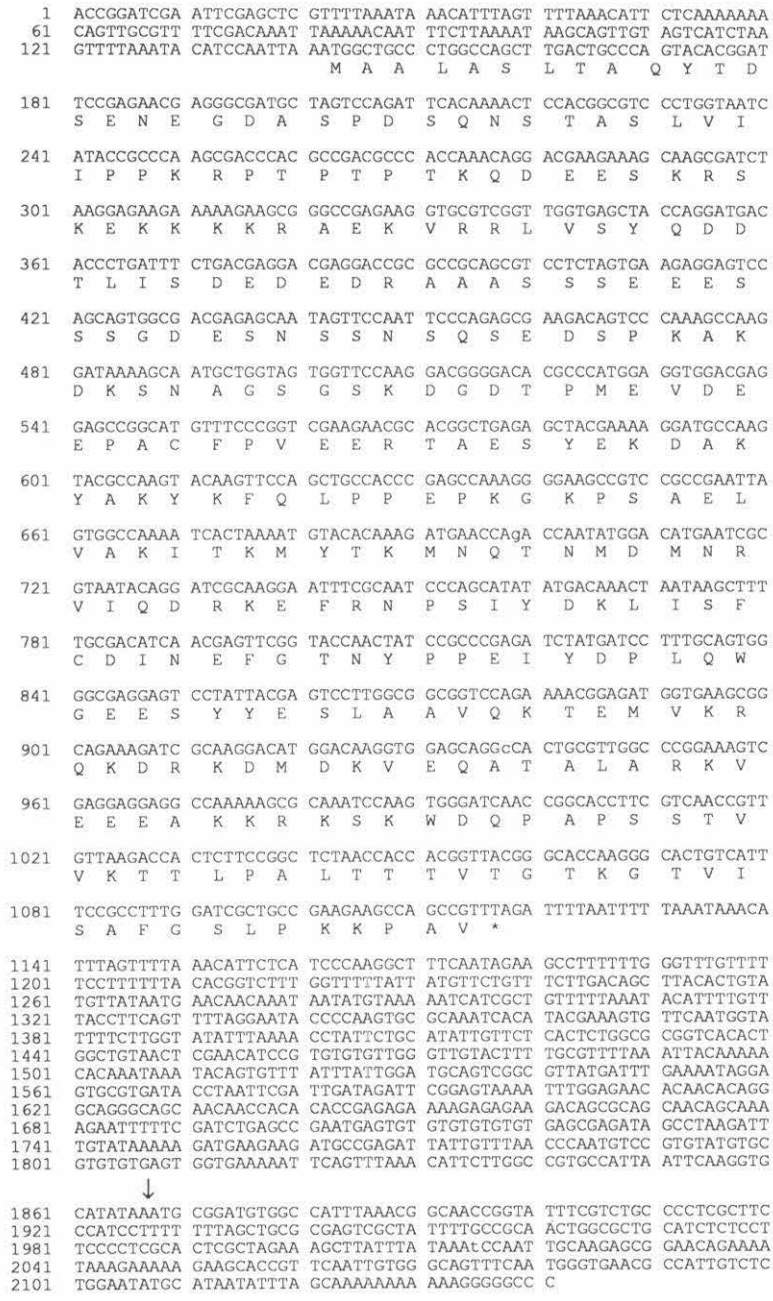

30
25
20
15
10
5
0
LIVE FLIES
1
2
3
4
5
6
7
8
9
10
DAYS
Heterozygotes
Homozygotes

Figure 4


1 ACCGGATCGA ATTCGAGCTC GTTTTAAATA AACATTTAGT TTTAAACATT CTCAAAAAAA
61 CAGTTGCGTT TTCGACAAAT TAAAAACAAT TTCTTAAAAT AAGCAGTTGT AGTCATCTAA
121 GTTTTAAATA CATCCAATTA AATGGCTGCC CTGGCCAGCT TGACTGCCCA GTACACGGAT
M A A L A S L T A Q Y T D
181 TCCGAGAACG AGGGCGATGC TAGTCCAGAT TCACAAAACT CCACGGCGTC CCTGGTAATC
S E N E G D A S P D S Q N S T A S L V I
241 ATACCGCCCA AGCGACCCAC GCCGACGCCC ACCAAACAGG ACGAAGAAAG CAAGCGATCT
I P P K R P T P T P T K Q D E E S K R S
301 AAGGAGAAGA AAAAGAAGCG GGCCGAGAAG GTGCGTCGGT TGGTGAGCTA CCAGGATGAC
K E K K K K R A E K V R R L V S Y Q D D
361 ACCCTGATTT CTGACGAGGA CGAGGACCGC GCCGCAGCGT CCTCTAGTGA AGAGGAGTCC
T L I S D E D E D R A A A S S S E E E S
421 AGCAGTGGCG ACGAGAGCAA TAGTTCCAAT TCCCAGAGCG AAGACAGTCC CAAAGCCAAG
S S G D E S N S S N S Q S E D S P K A K
481 GATAAAAGCA ATGCTGGTAG TGGTTCCAAG GACGGGGACA CGCCCATGGA GGTGGACGAG
D K S N A G S G S K D G D T P M E V D E
541 GAGCCGGCAT GTTTCCCGGT CGAAGAACGC ACGGCTGAGA GCTACGAAAA GGATGCCAAG
E P A C F P V E E R T A E S Y E K D A K
601 TACGCCAAGT ACAAGTTCCA GCTGCCACCC GAGCCAAAGG GGAAGCCGTC CGCCGAATTA
Y A K Y K F Q L P P E P K G K P S A E L
661 GTGGCCAAAA TCACTAAAAT GTACACAAAG ATGAACCAgA CCAATATGGA CATGAATCGC
V A K I T K M Y T K M N Q T N M D M N R
721 GTAATACAGG ATCGCAAGGA ATTTCGCAAT CCCAGCATAT ATGACAAACT AATAAGCTTT
V I Q D R K E F R N P S I Y D K L I S F
781 TGCGACATCA ACGAGTTCGG TACCAACTAT CCGCCCGAGA TCTATGATCC TTTGCAGTGG
C D I N E F G T N Y P P E I Y D P L Q W
841 GGCGAGGAGT CCTATTACGA GTCCTTGGCG GCGGTCCAGA AAACGGAGAT GGTGAAGCGG
G E E S Y Y E S L A A V Q K T E M V K R
901 CAGAAAGATC GCAAGGACAT GGACAAGGTG GAGCAGGcCA CTGCGTTGGC CCGGAAAGTC
Q K D R K D M D K V E Q A T A L A R K V
961 GAGGAGGAGG CCAAAAAGCG CAAATCCAAG TGGGATCAAC CGGCACCTTC GTCAACCGTT
E E E A K K R K S K W D Q P A P S S T V
1021 GTTAAGACCA CTCTTCCGGC TCTAACCACC ACGGTTACGG GCACCAAGGG CACTGTCATT
V K T T L P A L T T T V T G T K G T V I
1081 TCCGCCTTTG GATCGCTGCC GAAGAAGCCA GCCGTTTAGA TTTTAATTTT TAAATAAACA
S A F G S L P K K P A V *
1141 TTTAGTTTTA AACATTCTCA TCCCAAGGCT TTCAATAGAA GCCTTTTTTG GGTTTGTTTT
1201 TCCTTTTTTA CACGGTCTTT GGTTTTTATT ATGTTCTGTT TCTTGACAGC TTACACTGTA
1261 TGTTATAATG AACAACAAAT AATATGTAAA AATCATCGCT GTTTTTAAAT ACATTTTGTT
1321 TACCTTCAGT TTTAGGAATA CCCCAAGTGC GCAAATCACA TACGAAAGTG TTCAATGGTA
1381 TTTTCTTGGT ATATTTAAAA CCTATTCTGC ATATTGTTCT CACTCTGGCG CGGTCACACT
1441 GGCTGTAACT CGAACATCCG TGTGTGTTGG GTTGTACTTT TGCGTTTTAA ATTACAAAAA
1501 CACAAATAAA TACAGTGTTT ATTTATTGGA TGCAGTCGGC GTTATGATTT GAAAATAGGA
1561 GTGCGTGATA CCTAATTCGA TTGATAGATT CGGAGTAAAA TTTGGAGAAC ACAACACAGG
1621 GCAGGGCAGC AACAACCACA CACCGAGAGA AAAGAGAGAA GACAGCGCAG CAACAGCAAA
1681 AGAATTTTTC GATCTGAGCC GAATGAGTGT GTGTGTGTGT GAGCGAGATA GCCTAAGATT
1741 TGTATAAAAA GATGAAGAAG ATGCCGAGAT TATTGTTTAA CCCAATGTCC GTGTATGTGC
1801 GTGTGTGAGT GGTGAAAAAT TCAGTTTAAA CATTCTTGGC CGTGCCATTA ATTCAAGGTG
1861 CATATAAATG CGGATGTGGC CATTTAAACG GCAACCGGTA TTTCGTCTGC CCCTCGCTTC
1921 CCATCCTTTT TTTAGCTGCG CGAGTCGCTA TTTTGCCGCA ACTGGCGCTG CATCTCTCCT
1981 TCCCCTCGCA CTCGCTAGAA AGCTTATTTA TAAAtCCAAT TGCAAGAGCG GAACAGAAAA
2041 TAAAGAAAAA GAAGCACCGT TCAATTGTGG GCAGTTTCAA TGGGTGAACG CCATTGTCTC
2101 TGGAATATGC ATAATATTTA GCAAAAAAAA AAAGGGGGCC C

Appendix 2

Screening mixed pools of genomic DNA for small deletions: a PCR-based assay for chemical mutagenesis of cloned target genes

Detection of small genomic deletions in mixed DNA samples by PCR and implications for chemical mutagenesis in reverse genetics of *Drosophila melanogaster*

Bruce Hamilton

Abstract

Previous molecular screens for mutations in cloned Drosophila genes have relied on detection of *P* element insertions. While effective, these methods are limited by the insertion site bias of the transposon and consequently by the variety of lesions that can be generated. Efficient detection of chemically induced mutations would therefore be of considerable value. I demonstrate here a method to detect small deletions, such as those induced by exposure to diepoxides. This method is based on detecting the loss of a restriction site that is flanked by PCR primers. Pooled DNA samples are digested with the appropriate restriction endonuclease and then used as template for PCR amplification. DNA samples that lack the restriction site are detected by providing a nuclease-resistant amplification product smaller than that supportied by parental DNA. I show detection of the ry^1 allele, a deletion of ~100 bp that removes a Sac I site, in dilutions of up to one part in 400.

Introduction

The object of molecular genetics is to correlate observable phenotypes with molecular functions or dysfunctions. This can be done in either direction: mutations that produce interesting phenotypes can be cloned and cloned DNA segments of interest can be mutagenized *in situ*. These approaches have been applied, with varying degrees of difficulty, in a variety of organisms ranging from bacteriophage to rodents.

Mutagenizing a cloned gene without a prediction of the expected phenotype requires a physical assay for lesions in the DNA. The most general assays for detecting mutations in cloned Drosophila genes have required the use of *P* elements as the mutagen (Ballinger and Benzer 1989; Kaiser and Goodwin 1990; Hamilton et al. 1991). While transposon mutagenesis has a number of advantages, it also suffers in several respects from the non-randomness of recovered insertion sites. Developing methods to detect chemically induced lesions in cloned genes is therefore of interest.

One class of mutation that should be readily detectable is small deletions. Mutations induced at the *rosy* (*ry*)locus by diepoxybutane (DEB) treatment have been examined by Reardon et al. to determine the molecular nature of each lesion. These authors show that nearly 80% of the *ry* mutations they recovered were deletions smaller than 250 bp (Reardon et al. 1987). This size is sufficiently small that one could expect to assay such deletions by PCR (Saiki et al. 1988). In the experiment reported here, I use the a small deletion in the *ry* gene to test a PCR-based method for detecting such lesions. I chose to work with the *ry* locus because of the number of available mutant alleles that have been analyzed at the molecular level (Clark et al. 1986; Coté et al. 1986; Reardon et al. 1987; Lee et al. 1987; Keith et al. 1987).

Materials and Methods

The ry^1 allele has been described in detail (Clark et al. 1986; Coté et al. 1986); other genetic variants of Drosphila are described in (Lindsley and Zimm 1992). DNA preparations and PCRs were performed essentially as described (Hamilton et al. 1991). Oligonucleotide primers ry1A (5'-ACATTGTGGCGGAGATCTCGATGGC-3') and ry 1B (5'- ATCACAGATGGGTTGATCGCTGCAG-3') flank the ry^1 deletion and amplify a 372 bp product by PCR on ry^+ DNA.

Results

The strategy described here is designed to detect small deletions that remove a restriction site between two PCR primer sites. DNA from pools of fly lines under consideration is digested with a restriction enzyme to cleave the targeted site. The digested DNA is then used as a template for PCR to detect restriction resistant molecules. Amplification products from deletion mutants are distinguished from those of uncleaved non-mutant DNA by size fractionation through agarose gels.

To test the sensitivity of detection, PCR reactions were performed on Sac I-digested DNA from a range of dilutions of ry^1 with ry^+. DNA samples from 20 ry^1 *su(Hw)*^{2} *jvl* bx^{34e} / *TM1* adult males, from 20 ry^+ males, and from one ry^1 *su(Hw)*^{2} *jvl* bx^{34e} / *TM1* male mixed with 19 ry^+ males were prepared An aliquot of DNA from each of these samples was then digested with Sac I for 3.5 hr.s at 37°C. DNA from the mixed sample was then serially diluted ten-fold with DNA from the ry^+ stock to create effective dilutions of the heterzygous ry^1 fly of 1/200 and 1/2000 in the new samples. Approximately 200 ng DNA was then used as template in each PCR amplification with primers ry1A and ry1B. As shown in figure 1, the mutant product can be detected when the heterozygous mutant fly contributes as little as one two-hundredth of the DNA sample (hence the mutant allele is diluted 1/400), but not one two-thousandth (1/4000).

Discussion

A simple implementation of the method described here might fail to detect a large fraction of potentially informative deletions, either because mutation did not remove the restriction site or because the mutation removes or falls outside of a primer binding site. However, because so little material is needed in each PCR assay, the same pools of mutagenized fly DNA could be for several assays. Thus, multiple primer pairs could be used to assay a larger region of the gene. Furthermore, since the assay only requires the

targeted resriction site be unique within the amplified region, primer pairs and restriction sites could probably be chosen to allow detection of most DEB-induced mutations within a well-characterized gene.

The sensitivity of detection with this method will depend on the efficiency of cleavage by the restriction enzyme and on the amplification disparity between the wild-type and mutant PCR templates. I have detected the equivalent of one heterozygote out of 200 flies using Sac I and a small template from the *rosy* gene. This is already a useful sensitivity for mutagenesis; Ballinger and Benzer (1989) successfully used pools of 200 fly lines to detect *P* element insertoins by PCR.

Acknowledgements: I thank C. H. Martin for productive discussions and E. B. Lewis for the *rosy* mutant stock.

Literature cited

Ballinger, D. G. and S. Benzer, 1989 Targeted gene mutations in Drosophila. Proc. Natl. Acad. Sci. USA **86:** 9402-9406.

Clark, S. H., M. McCarron, C. Love and A. Chovnik, 1986 On the identification of the *rosy* locus DNA in *Drosophila melanogaster*: intragenic recombination mapping of mutations associated with insertions and deletions. Genetics **112:** 755-767.

Coté, B., W. Bender, D. Curtis and A. Chovnick, 1986 Molecular analysis of the rosy locus in *Drosophila melanogaster*. Genetics **112:** 769-783.

Hamilton, B. A., M. J. Palazzolo, J. H. Chang, K. V. Raghavan, C. A. Mayeda, M. A. Whitney and E. M. Meyerowitz, 1991 Large scale screen for transposon insertions into cloned genes. Proc. Natl. Acad. Sci. USA **88:** 2731-2735.

Kaiser, K. and S. F. Goodwin, 1990 "Site-selected" transposon mutagenesis of Drosophila. Proc. Natl. Acad. Sci. USA **87:** 1686-1690.

Keith, T. P., M. A. Riley, M. Kreitman, R. C. Lewontin, D. Curtis and G. Chambers, 1987 Sequence of the structural gene for xanthine dehydrogenase (rosy locus) in *Drosophila melanogaster*. Genetics **116:** 67-73.

Lee, C. S., D. Curtis, M. McCarron, C. Love, M. Gray, W. Bender and A. Chovnick, 1987 Mutations affecting expression of the rosy locus in *Drosophila melanogaster*. Genetics **116:** 55-66.

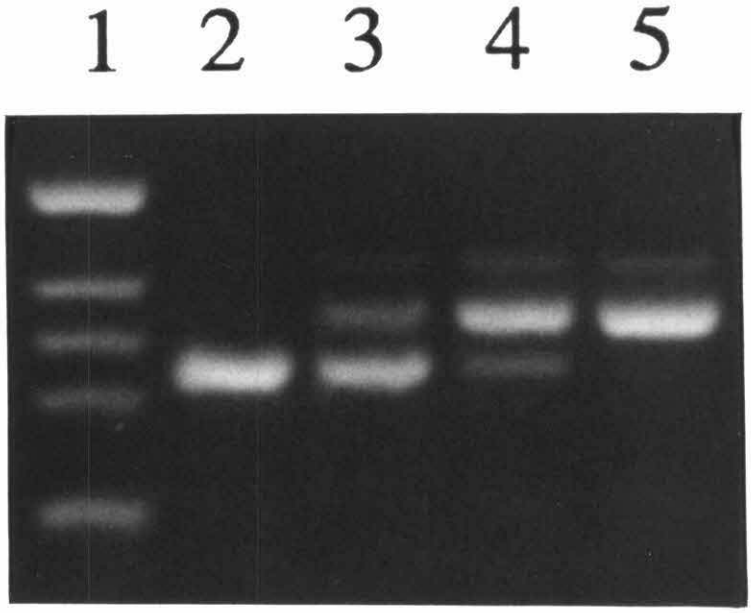
Lindsley, D. L. and G. G. Zimm, 1992 *The Genome of Drosophila melanogaster*. Academic Press, Inc.: San Diego, CA. pp.

Reardon, J. T., C. A. Liljestrand-Golden, R. L. Dusenbury and P. D. Smith, 1987 Molecular analysis of diepoxybutane-induced mutations at the rosy locus of *Drosophila melanogaster*. Genetics **115:** 323-331.

Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis and H. A. Erlich, 1988 Primer-directed amplification of DNA with a thermostable DNA polymerase. Science **239:** 487-491.

Figure Legend

Figure 1 PCR reaction products size fractionated by electrophoresis through agarose and visualized by ethidium bromide fluorescence. The lanes are: 1, pBR322 digested with Hinf I as a size marker; 2, amplification product from DNA of the heterozygous *ry* stock; 3, amplification from DNA of 1 *ry* heterozygote among 20 flies (1/20); 4, 1/200; 5, 1/2000.


1
2
3
4
5

Appendix 3

Rapid assessment of gross expression patterns by bulk filter hybridization: genes selectively expressed in the Drosophila embryonic nerve cord

Rapid assessment of RNA gross expression patterns by bulk filter hybridization: genes selectively expressed in the Drosophila embryonic nerve cord

Bruce Hamilton

Summary

A panel of 250 cDNA clones representing mRNAs expressed in the Drosophila adult head was analyzed by hybridization to radiolabeled cDNA probes derived from embryonic ventral nerve cord and larval fat body. This procedure allowed the array of clones to be sorted by gross expression pattern. Each clone could be categorized as being predominantly expressed in one tissue, expressed in both tissues, or not expressed in either tissue above a threshold level for detection. Such data allow the panel of clones to be organized such that priority in future experiments is given to cloned sequences expressed in a desired tissue, such as the central nervous system, without excluding clones for sequences expressed at low levels. This experiment has broader implications for the design of large scale cDNA sequence analysis projects. The data presented here also suggest that a high percentage of genes expressed in the nervous system are expressed in patterns with at least some tissue specificity.

Materials and Methods

The cDNA array. Drosophila adult head cDNA clones in λSWAJ (Palazzolo et al. 1989).were amplified out of phage lysates by PCR with primers that abut the cloning site, as described (Hamilton et al. 1991). Amplification products were electrophoresed through agarose gels, photographed, transfered to Hybond-N nylon filters, and cross-linked by UV irradiation.

cDNA probes. Radiolabeled cDNA was prepared from two source tissues: embryonic nerve cord and larval fat body. Nerve cords were bulk isolated from embryos as described (Goodman et al. 1984). Body wall fat body was hand dissected from mid to late third instar larvae in cold Ringer's. Polyadenylated RNA was prepared from each tissue by serial organic extractions followed by oligo(dT) cellulose chromatography. cDNA was synthesized by standard methods from an oligo(dT) primer with ^{32}P-dCTP (3000 Ci / mmole) as the only source of dCTP in the reaction. Two mg poly(A)$_n$ was added to each probe prior to use to inhibit hybridization of probe to poly(dA) tracts in the amplified cDNA clones. Hybridizations and washes were essentially as described (Hamilton, et al. 1991).

Results

Sequences in an array of 250 sorted Drosophila cDNA clones whose corresponding RNAs were previously shown to be expressed in adult head and in late (24 hr.) but not early (0-1 hr.) whole embryos (Palazzolo et al. 1989) were assayed for expression in embryonic nerve cord and larval fat body. Duplicate sets of DNA gel blots of PCR amplified cDNA inserts were hybridized to high specific activity cDNA synthesized from poly(A)$^+$ RNA from each tissue. After washing the filters, hybridization signals were detected by autoradiography for up to one week with an intensifying screen. Sample data are shown in Figure 1.

Of the 250 head cDNA clones, 141 were detected by hybridization of at least one of the probes: 74 clones were detected only by the nerve cord probe, 32 only by the fat body probe, and 35 by bith probes. The remaining 109 clones were not detected by either probe.

Discussion

The adult head of Drosophila is predominantly composed of three tissue types: nervous system, fat body and muscle. Heads are easily isolated in bulk and therefore are an easily obtained source for neurally-enriched tissue. In addition, molecules that are important in neural development often continue to be expressed in the adult head. For example, fasciclin I, which plays an important role in construction of the embryonic CNS axon scaffold (Elkins et al. 1990) is represented in this collection of cDNAs. Thus, a collection of cDNA sequences expressed in both the adult head and the late embryo but not in the very early embryo might be enriched for molecules that act in the development or functioning of the nervous system.

I have presented hybridization data for a set of 250 cDNAs that represent messages in the head and late embryo. More than half of these clones (56%) were detected by at least one of the probes. Of 109 clones detected by the nerve cord probe, 74 were not detected by the fat body probe. This suggests that ~70% of the polyadenylated RNAs expressed in both the adult head and embryonic nerve cord are expressed with at least some tissue specificity.

The rapidity and ease with which large numbers of clones can be sorted make this method of analysis a useful step in spite of its limited sensitivity. Although some additional percentage of the clones should also correspond to tissue-specific messages from cell types not represented in the probes, such as muscle, epidermis, and PNS, a significant fraction of the clones probably represent messages too low in abundance in any readily prepared tissue to be detected by this method. This is the chief limitation of this strategy. On the other hand, this procedure has several valuable features to recommend it. The principle advantage is scale. RNA from a very small amount of tissue is sufficient to screen even much larger arrays of clones than I have described here. One corollary of increasing scale is decreasing labor per cloned gene. While a thousand clones can easily be analyzed in a

single experiment of the kind described here, performing and analyzing a thousand RNA gel blots or in situ hybridization experiments is far less tenable.

The information generated by this strategy allows a quick and rational triage of clones for more intensive analyses. For experiments aimed at neural development, the data here would be taken to give highest priority to the 74 clones expressed specifically in neural tissue. Lower priority might be given to the 109 clones for which no information was generated and 35 clones that were detected by both probes. The 32 clones detected only by the fat body probe would be excluded from further analysis. Rapid coding sequence analysis (Hamilton, Palazzolo and Meyerowitz 1991; Strathmann et al. 1991) on clones of high potential interest is a rational and methodical alternative to the expressed sequence tag approach to identifying and classifying sequences expressed in the brain (Adams et al. 1992).

Acknowledgements: I thank M. J. Palazzolo for the λSWAJ cDNA clones and advice and K. G. Zinn for teaching and assisting with the mass isolation of embryonic nerve cords.

Literature cited

Adams, M. D., M. Dubnick, A. R. Kerlavage, R. Moreno, J. M. Kelley, T. R. Utterback, J. W. Nagle, C. Fields and J. C. Venter, 1992 Sequence identification of 2,375 human brain genes. Nature **355:** 632-634.

Elkins, T., K. Zinn, L. McAllister, F. M. Hoffmann and C. S. Goodman, 1990 Genetic analysis of a Drosophila neural cell adhesion molecule: interaction of *fasciclin I* and Abelson tyrosine kinase mutations. Cell **60:** 565-575.

Goodman, C. S., M. J. Bastiani, C. Q. Doe, S. du Lac, S. L. Helfand, J. Y. Kuwada and J. B. Thomas, 1984 Cell recognition during neuronal development. Science **225:** 1271-1279.

Hamilton, B. A., M. J. Palazzolo, J. H. Chang, K. V. Raghavan, C. A. Mayeda, M. A. Whitney and E. M. Meyerowitz, 1991 Large scale screen for transposon insertions into cloned genes. Proc. Natl. Acad. Sci. USA **88:** 2731-2735.

Hamilton, B. A., M. J. Palazzolo and E. M. Meyerowitz, 1991 Rapid isolation of long cDNA clones from existing libraries. Nucl. Acids Res. **19:** 1951-1952.

Palazzolo, M. J., D. R. Hyde, K. VijayRaghavan, K. Mecklenberg, S. Benzer and E. Meyerowitz, 1989 Use of a new strategy to isolate and characterize 436 Drosophila cDNA clones corresponding to RNAs detected in adult heads but not in early embryos. Neuron **3:** 527-539.

Strathmann, M., B. A. Hamilton, C. A. Mayeda, M. I. Simon, E. M. Meyerowitz and M. J. Palazzolo, 1991 Transposon-facilitated DNA sequencing. Proc. Natl. Acad. Sci. USA **88:** 1247-1250.

Figure Legends

Figure 1: Sample data. ***A.*** Photograph of one gel of the PCR amplified clones. ***B.*** Autoradiogram of the corresponding filter hybridized to the nerve cord cDNA probe.

Figure 2: Summary of data from all filters in this experiment.

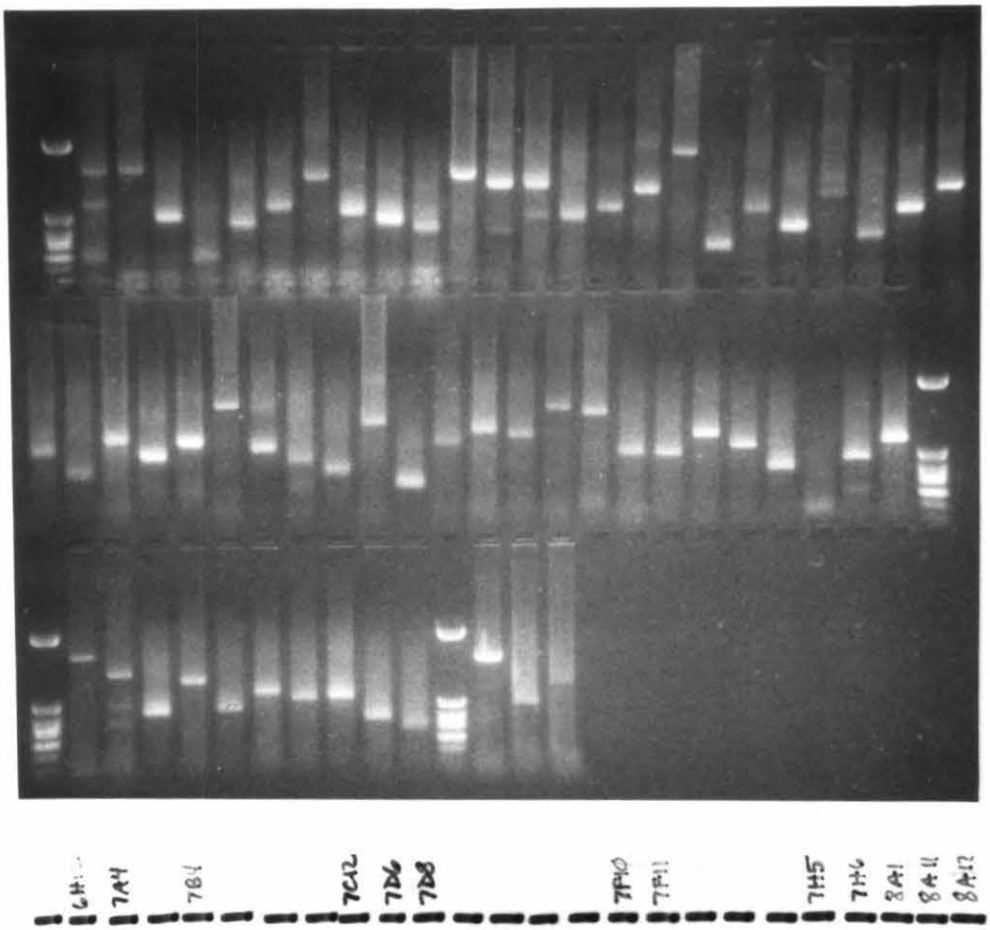
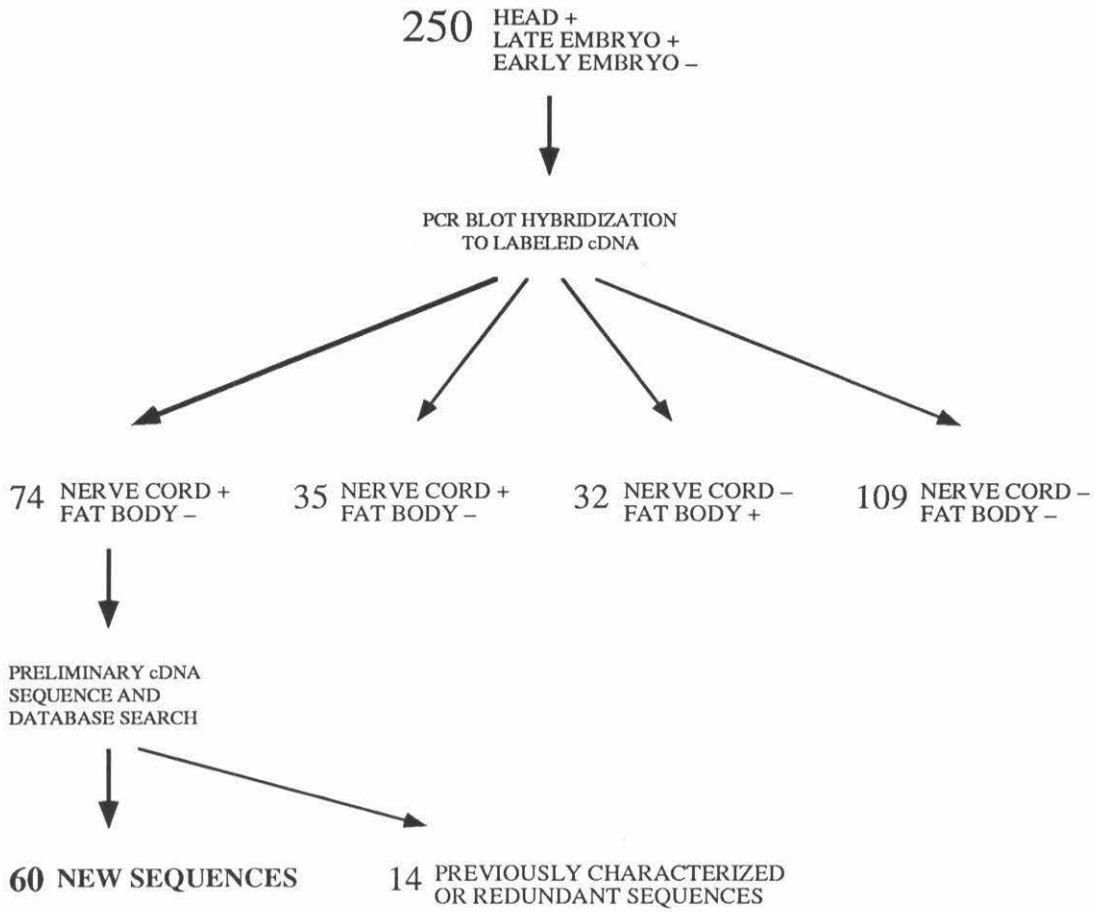


Figure 2


250 HEAD +
LATE EMBRYO +
EARLY EMBRYO –
PCR BLOT HYBRIDIZATION
TO LABELED cDNA
74 NERVE CORD +
FAT BODY –
35 NERVE CORD +
FAT BODY –
32 NERVE CORD –
FAT BODY +
109 NERVE CORD –
FAT BODY –
PRELIMINARY cDNA
SEQUENCE AND
DATABASE SEARCH
60 NEW SEQUENCES
14 PREVIOUSLY CHARACTERIZED
OR REDUNDANT SEQUENCES