National Library of Canada
Canadian Theses Service
Ottawa, Canada
K1A 0N4

Bibliothèque nationale du Canada
Services des thèses canadiennes

CANADIAN THESES

THÈSES CANADIENNES

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30.

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30.

THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED

LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE

NL-339 (r. 86/06)

Canada

ISOLATION AND CHARACTERIZATION OF *DROSOPHILA MELANOGASTER* GENOMIC DNA SEQUENCES CONTAINING BACTERIAL INSERTION SEQUENCES

by

Penelope A. Johnson

A thesis presented to the University of Ottawa
in partial fulfillment of the requirements
for the degree of
Master of Science
in Biology

Ottawa, Ontario

© Penelope A. Johnson, Ottawa, Canada, 1987.

Permission has been granted to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film.

The author (copyright owner) has reserved other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without his/her written permission.

L'autorisation a été accordée à la Bibliothèque nationale du Canada de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur (titulaire du droit d'auteur) se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation écrite.

ISBN 0-315-36515-3

UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

The University of Ottawa requires the signature of all persons using or photocopying this thesis. Please sign below, and give address and date.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my gratitude to Dr. Pierre Moreau who has been my supervisor, mentor and friend during this research. The debt I owe him is beyond repaying.

To Drs. Vern Seligy and Donal Hickey, the members of my advisory committee, many thanks for fruitful discussion and guidance during this work.

Technical assitance during this research was tirelessly and willingly given by Nancy (Tourville) Delcellier; I would like to especially thank her for her support and friendship during the last months of work.

Thanks to the other students who have shared in and contributed to the fun of 'doing science' for the last two years: especially to Mary, Janice, Benoit, Steve, and Sharon, thank you!

To Mr. George Ben-T and Mr. Paul Brunon, thank you for your patient and excellent work on the photographic figures. Likewise, to Jacques Hélie, thank you for the drawn figures and labelling of photographic plates.

To the unseen people who have always encouraged my education and scientific endeavours to whom I owe a special debt of gratitude: To my mother/my friend, Pamela, thanks for your encouragement, enduring wisdom, sensitivity and inspiration! You will always be my role model. To Griff, thank you for your infinite patience, your sense of humour and your ability to make me laugh at myself. Lastly, I

thank my father who has shown me the meaning of courage and perseverance in the face of insurmountable odds.

Financial support for this research was provided by Ontario Graduate Scholarships, NSERC Scholarships, University of Ottawa Entrance Scholarships, and Medical Research Council operating grants.

TABLE OF CONTENTS

Chapter III

RESULTS

PART 1 Isolation of the Drosophila α-Amylase Gene

PART 2 Identifying Sequences Comprising L1231 and L1232

ABSTRACT

The intent of this thesis was to isolate and characterize the α-amylase gene of *Drosophila melanogaster*. This primary step would be a starting point for studies on the role of *cis*-acting factors in tissue-specific expression of the gene.

A *Drosophila melanogaster* genomic library constructed in the lambda bacteriophage vector, EMBL4, was first probed using a murine pancreatic α-amylase cDNA, pMPa21 (Schibler, et al, 1980). Probing under low stringency hybridization conditions, eight recombinant clones were isolated and characterized. However, these clones are not related and do not encode α-amylase. In order to increase the hybridization stringency, a second probe (pUCa1), believed to contain a Drosophila amylase cDNA was next used to screen the library.

Using pUCa1, two recombinant clones were isolated from the Drosophila genomic library and were characterized by restriction endonuclease digestion. Although, pUCa1 expressed a starch-hydrolyzing activity, hybridized to an RNA of the same size as α-amylase, and hybridized to pMPa21, it was subsequently found not to contain α-amylase coding sequences , but the bacterial insertion sequence IS1 (M. Whitely & P. Moreau, personal communication).

The intensity of hybridization of the probe to the two isolated recombinants suggested that both recombinants contained IS1. Characterization of these two clones continued and restriction and hybridization data confirmed that they harbour IS1. The two isolated recombinants also hybridize to discrete Drosophila genomic DNA sequences which are not homologous to IS1.

This thesis describes the isolation of eukaryotic genomic sequences which have served as targets for the insertion of the prokaryotic mobile element, IS1. The conclusions presented serve as a warning to molecular biologists using prokaryotic systems to propagate eukaryotic genes.

RESUME

L'objet de cette thèse était d'isoler et de caractériser le gène α-amylase chez *Drosophila melanogaster*. Cette première étape devait être le point de départ d'une étude sur le rôle des facteurs responsables de l'expression des gènes dans des tissus spécifiques.

Une librairie de séquences génomiques de *Drosophila melanogaster*, construite dans lambda EMBL4, a été sondée en utilisant le cDNA du gène α-amylase pancréatique de souris, pMPa21 (Schibler, et al, 1980). En utilisant des conditions d'hybridation non stringentes, huit recombinants ont été isolés et caractérisés. Cependant, aucun de ces clones ne sont identiques. Afin d'augmenter la stringence d'hybridation, nous avons utilisé une deuxième sonde croyant qu'elle contenait le cDNA α-amylase de Drosophile (pUCa1).

En utilisant pUCa1 comme sonde, deux clones recombinants furent isolés de la librairie génomique de Drosophile et furent caractérisés par des digestes d'enzymes de restriction. Bien que pUCa1 exprimait une activité capable d'hydrolyser l'amidon, s'hybridait à un ARN de même taille que celui de l'α-amylase, et s'hybridait au pMPa21, il a été démontré qu'il ne contenait pas de séquences α-amylase, mais plutôt, une séquence d'insertion, IS1 d'origine bactérienne (M. Whitely & P. Moreau, communication personnelle).

L'intensite d'hybridation de la sonde aux deux recombinants isolés suggère que ces deux clones contiennent IS1. Pour cette raison nous avons poursuivi la caractérisation de ces clones. Nos résultats des digestes et hybridations ont par la suite confirmé qu'ils contenaient effectivement IS1. De plus, les deux recombinants isolés s'hybrident à des séquences spécifiques d' ADN génomique de Drosophile qui ne sont pas homologues au IS1.

Cette thèse décrit comment des séquences eucaryotes génomiques ayant servi de cibles pour l'insertion de l'élément mobile procaryote IS1, ont été isolées. De plus, les conclusions présentées devraient servir d'avertissement aux biologistes moléculaires qui utilisent des systèmes procaryotes pour la propagation de gènes eucaryotes.

LIST OF ABBREVIATIONS

Ad-2	adenovirus Type 2
ADH	alcohol dehydrogenase
AMG	anterior midgut
Amy	amylase gene locus
bp	base pairs
BSA	bovine serum albumin
cDNA	complementary DNA
$CHCl_3$	chloroform
CIP	calf intestinal phosphatase
DNA	Deoxyribonucleic Acid
E. coli	*Escherichia coli*
E1A	early gene 1A of Ad-virus
EDTA	ethylene diamine tetraacetic acid
HTLV III	human T-cell lymphoma virus III
IS1	insertion sequence 1
Kb	kilobase pairs
L-Broth	Luria Broth
map	midgut activity pattern gene locus
m.o.i	multiplicity of infection
mRNA	messenger RNA
Mt-II	metallothionein II gene
nt	nucleotide
NZCYM	NZ-amine-casamino acid-yeast media
PAF	polyvinyl-pyrrylodine-albumin-ficoll
PEG	polyethylene glycol
pfu	plaque forming units
phage	bacteriophage
PMG	posterior midgut
Poly(A)+	polyadenylated RNA
rDNA	ribosomal DNA
RNA	Ribonucleic Acid
rRNA	ribosomal DNA
SDS	sodium dodecyl sulfate
SM	sodium-magnesium buffer
SV40	simian virus 40
TAE	tris-acetate-EDTA buffer
TBE	tris-borate-EDTA buffer
TE	tris-EDTA buffer
TES	tris-NaCl-EDTA buffer
µCi	micro Curie
µg	micro gram
µL	micro Litre

LIST OF TABLES

LIST OF FIGURES

Chapter I

INTRODUCTION

The function of a particular tissue is defined by expression of specified sets of genes within the cells that compose that tissue. One of the fundamental questions of modern biology is how selective expression of genes in specific tissues is regulated. We were interested in studying tissue-specific control of gene expression using *Drosophila melanogaster* as a model system.

The α-amylase gene of *Drosophila melanogaster* encodes a representative tissue-specific enzyme expressed primarily in midgut cells of Drosophila. The original purpose of this thesis was to isolate and characterize the α-amylase gene of *Drosophila melanogaster*. Isolation of the gene was to be the primary step in a larger study in which the gene would be used to examine control of tissue-specific gene expression *in vivo* by methods developed by Rubin and Spradling (1982).

In order to isolate the α-amylase gene, a *Drosophila melanogaster* genomic library was first screened using a mouse α-amylase cDNA probe under low stringency hybridization conditions. The low stringency of the hybridization conditions and the minimal homology of the mouse cDNA to the Drosophila gene resulted in isolation of several unrelated clones.

Subsequently, a cDNA made in our laboratory by M. Whitely from Drosophila messenger RNA, was used to probe the

library. This cDNA (pUCa1) expresses a protein capable of hydrolyzing starch. In addition, pUCa1 hybridizes on Northern blots to a 1.6 Kb RNA, the correct size to encode α-amylase. From these data we concluded that pUCa1 encoded *Drosophila melanogaster* α-amylase. Using pUCa1 as a probe, two recombinant clones were isolated from the genomic library. These clones were characterized by restriction endonuclease digestion and hybridization to the probe. Simultaneously, pUCa1 was sequenced and, to our astonishment, was found not to contain Drosophila α-amylase coding sequences but a bacterial insertion sequence closely related to IS1 (M. Whitely & P. Moreau, manuscript in preparation).

Further characterization of the recombinant clones isolated using pUCa1 suggests that they also contain IS1 which has integrated into Drosophila DNA. This event has never been documented previously and raises intriguing and pertinent questions: Is there homology between prokaryotic and Drosophila transposable elements? Are there specific target sites for insertion of bacterial sequences into Drosophila DNA? Are these target sites similar to ones in prokaryotic DNA? Most importantly, what is the broader significance of insertion of prokaryotic sequences into eukaryotic DNA? Answers to these questions may have implications regarding the evolution of mobile elements and the context in which molecular biology is studied.

The first part of this introduction will explain the research project within the context of tissue-specific gene expression of α-amylase and the strategies employed to isolate the *Drosophila melanogaster* α-amylase gene. In the second part, the involvement of bacterial Insertion Sequences in the research project will be explored.

I.1 *Drosophila melanogaster* as a model system for the study of control of tissue-specific gene expression

The fruit fly, *Drosophila melanogaster*, is a perfect tool for the study of tissue-specific gene expression *in vivo*. The genetics of the fruit fly have been intensively studied by classical geneticists and most structural genes have been mapped to chromosomal loci. Selective breeding has created fly strains which are mutant for many tissue-specific genes. In addition, genetic manipulation of embryos has been made possible within the last few years (Rubin & Spradling, 1982; Spradling & Rubin, 1982) marrying the techniques of genetics with those of molecular biology. This technology, called germline transformation, allows exogenous genes to be introduced into germline DNA of Drosophila embryos (Rubin & Spradling, 1982). The exogenous gene thus becomes an inherited trait of the transformed line. Insertion of the gene into the germline is facilitated by using P-element transposable elements as vectors (Spradling & Rubin, 1982).

P-elements are a family of transposable elements of *Drosophila melanogaster* which transpose with high

frequency in fly strains which carry no endogenous P-elements (M-strains). Further, P-elements transpose preferentially into germline chromosomes (Engels, 1980; Kidwell, et al, 1982; Spradling & Rubin, 1983). Hence, an exogenous gene is incorporated into the genome by cloning it into a transposable P-element and microinjecting the cloned gene directly into the primordial germline cells of M-strain embryos (Rubin & Spradling, 1983; Spradling & Rubin, 1983). The P-element carries the cloned gene into the germline DNA of the recipients resulting in microinjected embryos which carry an exogenous gene stably integrated into their genome. Thereafter the new gene is a heritable trait of the transformed strain.

Genes microinjected thus far are expressed in appropriate tissues, at the correct developmental stage, and, for the most part, independently of the insertion site of the gene into the genome. For example, microinjection of the genes for Xanthine Dehydrogenase (Spradling & Rubin, 1983; Rubin & Spradling, 1983), Alcohol Dehydrogenase (Goldberg, et al., 1983), Dopa Decarboxylase (Scholnick, et al., 1983), and SGS3 'glue' (Richards, et al, 1983) all reconstitute some level of tissue-specific enzyme activity in fly strains mutant at these loci. Conversely, injection of mutated genes can cause anomolous activity in recipient flies (Hiromi, et al, 1986). Although the integration site of the microinjected gene is random, enzyme activity is limited to those

tissues where the native gene would normally be expressed. Therefore, it appears that control of tissue-specific expression in Drosophila is contiguous with the gene and its flanking sequences.

Germline transformation of *Drosophila melanogaster* using P-elements provides an ideal system for *in vivo* study of control of tissue-specific gene expression.

1.2 DNA sequences controlling gene expression

Already there is much evidence that control of tissue-specific gene expression is to a large extent exerted at the level of transcription. Germline transformation has demonstrated that *cis*-acting sequences control tissue-specific gene expression. DNA sequences flanking or within the coding sequences dictate the efficiency of gene transcription via specific *trans*-acting cellular factors.

Transcription of most eukaryotic genes transcribed by RNA polymerase II is maintained at a basal level by promoter sequences which flank the gene on the 5'(upstream) side. The promoter acts as a focus for RNA polymerase II and thus potentiates accurate and efficient transcription from the correct start site, the cap site, defined on the gene sequence as nucleotide +1 (Benoist & Chambon, 1981). The following elements comprise the typical eukaryotic promoter: the Goldberg-Hogness sequence which lies approximately 30 base pairs (bp) 5' to the cap site of most eukaryotic genes and conforms to the consensus sequence, TATAAA, and a second region 70-90 bp upstream

which may have canonical sequences GGCAATCT or GGGCGG (Benoist, et al., 1980; Benoist & Chambon, 1981; Mathis & Chambon, 1981; Dierks, et al, 1983; Efstratiadis, et al, 1980).

Other elements called enhancers are found flanking many viral genes (Moreau, et al, 1981; Hen, et al, 1983; Fuhrman, et al, 1981; Hearing & Shenk, 1986; Hynes, et al, 1983) and some cellular genes (Voss, et al, 1986; Goodbourn, et al,1985; Gillies, et al, 1983). Enhancers are usually found far upstream of the cap site, (more than 100 bp) but are occasionally found within the gene coding sequences (Khoury & Gruss, 1983; Banerji, et al, 1983; Queen & Baltimore, 1983; Gillies, et al, 1983; Osborne, et al, 1984). First characterized in Simian Virus 40 (SV40), these elements were called enhancers because they increased transcription from the associated promoter (Moreau, et al, 1981; Banerji, et al, 1981). As with promoters, enhancers act in *cis* (Moreau, et al, 1981). However, enhancers are typically distinguished from promoters by several characteristics. Unlike promoters, enhancers increase transcriptional efficiency even when placed at a distance from the gene (Moreau, et al, 1981; Atchison & Perry, 1986; Fromm & Berg, 1983). Additionally, enhancers can exert their effects from 3' or 5' positions relative to the promoter and can act independently of orientation relative to the direction of transcription (Fromm & Berg, 1983; Wasylyk et al, 1983). A 'core' sequence

(Weiher, et al, 1983) and repeated sequence motifs (Khoury & Gruss, 1983; Serfling, et al, 1985), are common features of many enhancers. However, unlike the promoter, no single sequence feature is characteristic of all enhancers. As well, promoters usually function in all cell types whereas viral enhancers function maximally within the range of hosts that the virus infects and, within the host, the enhancer usually functions best in particular cell types (Berg, et al, 1983; Yoshimura, et al, 1985; Laimins, et al, 1982; Laimins,et al, 1984; Zaret & Yamamoto, 1984; Hynes, et al, 1983; de Villiers, et al, 1982); cellular enhancers direct transcription within specific cell types (Hayday, et al, 1984; Gerlinger, et al, 1986; Karin, et al, 1984; Voss, et al, 1986; Ciliberto, et al, 1985; Adams, et al, 1985) or are inducible by specific physiological circumstance (Goodbourn, et al, 1985; Bienz & Pelham, 1986). Yet, as analysis of various upstream sequences continues, the distinction between enhancers and promoters becomes increasingly blurred (Serfling, et al, 1986). Thus, promoters associated with some cellular genes direct tissue-specific transcription initiation from heterologous, non-tissue-specific enhancers (Mason, et al, 1985; Grosschedl & Baltimore, 1985; Gillies, et al, 1983).

Other *cis*-acting sequences associated with cellular genes display some characteristics of enhancers and elicit expression of cloned heterologous genes only in those tissues or analogous cell systems where the native genes

would be expressed (Grosschedl, et al, 1984; Chada, et al, 1985; Townes, et al, 1985, Hanahan, et al, 1985; Swift,et al, 1984; Ornitz, et al, 1985; Hammer, et al, 1984; Palmiter et al, 1982; Walker, et al, 1983; Bodary, et al, 1983; Zinn, et al, 1983; D'Onofrio, et al, 1985; Hiyashi, et al, 1985; Kosche, et al, 1985; Charnay, et al, 1985; Hagenbuchle, et al, 1985; Fujita, et al, 1986).

Interestingly, DNA sequences associated with some genes have the characteristics of enhancers except that they act to repress transcription. Thus, upstream sequences associated with rat insulin I gene (Laimins, et al, 1986), yeast silent mating type cassette (Brand, et al, 1985), and HTLV III viral genes (Rosen, et al, 1985), can act in *cis*, at a distance, 3' or 5' to the gene, and regardless of orientation to stop transcriptional activity. These 'silencer' sequences shut down activity of an associated enhancer but this can presumably be reversed once the correct physiological conditions for transcription initiation prevail.

Maxims for gene expression in other eukaryotic systems apply to control of gene expression in *Drosophila melanogaster* as well (Mattox, et al, 1984; Benyajati, et al, 1984; Heberlein, et al, 1985; Skuse & Sullivan, 1985; Karin, et al, 1984; Hoffman & Corces, 1986; Klemenz & Gehring, 1986; DiNocero, et al, 1983; Rio & Rubin, 1985; Amin, et al, 1985). The control of tissue-specific gene expression in *D. melanogaster* is particularily well

illustrated by the 'nested' pupal cuticle protein gene which lies in the intron of a gene encoding an enzyme of the purine pathway. These two genes are encoded on opposite strands of the same DNA sequence but are unrelated by function. Each is expressed at different developmental stages in distinct tissues with no apparent overlap of expression (Henikoff, et al, 1986).

1.3 *Trans*-acting factors and gene expression

Cis-acting sequences are targets for *trans*-acting factors which modulate gene expression by binding the DNA (Davison, et al, 1983; Dynan & Tjian, 1985; Sassone-Corsi & Borrelli, 1986; Gariglio, et al, 1981). Unlike prokaryotic RNA polymerase, eukaryotic RNA polymerase II does not have an intrinsic ability to recognize the promoter *in vitro*. Assembly of *trans*-acting factors into initiation complexes allows RNA polymerase II to recognize and bind the promoter and commence accurate transcription (Davison, et al, 1983; Dynan & Tjian, 1985; Takahasi, et al, 1986). Crude cellular extracts such as, S100 (Manley, et al, 1980), whole cell extracts (Weil, et al, 1979), and mammalian nuclear extracts (Dignam, et al, 1983), can reconstitute faithful *in vitro* transcription by RNA polymerase II.

As well, *in vitro* competition experiments between wild-type and mutant upstream sequences for limited cellular factors and DNAase I footprinting of sequences bound by these factors (Wu, 1985), have enabled the isolation of

trans-acting factors which are required for transcription of particular genes (Dean, et al, 1983; Ucker, et al, 1983; Payvar, et al, 1983; Chandler, et al, 1983; Sawadogo & Roeder, 1985; Hayward, et al, 1982; Miyamoto, et al, 1985; Ohlsson, et al, 1986; Renkawitz, et al, 1984; Kovesdi, et al, 1986). These factors evoke transcription in specific cell-types or under particular physiological circumstances. For example, Scholer and Gruss (1983) showed that extracts of B and T lymphocytes and HeLa cells cause SV40 enhancer activity but only the lymphocyte extract allows IgH enhancer activity. Some lymphocyte specific factors have recently been identified (Maeda, et al, 1986; Church, et al, 1985; Ephrussi, et al, 1985; Mercola, et al, 1985; Henninghausen, 1985; Weinberger, et al, 1986). As well, *trans*-acting factors which repress transcriptional activity from enhancers have also been characterized. (Borrelli, et al, 1984; Gorman, et al, 1985).

Drosophila melanogaster genes also respond to a variety of *trans*-acting cellular factors *in vivo* and *in vitro* (Parker & Topol, 1984a; Parker & Topol, 1984b; Heberlein, et al, 1985; Desplan, et al, 1985; Benyajati & Davis, 1984; Chao & Guild, 1986; Gilmour & Lis, 1985; Gilmour, et al, 1986). Most importantly to the study undertaken for this thesis, the pattern of α-amylase expression in the posterior midgut of *Drosophila melanogaster* is regulated by the product of a linked gene called *map* (midgut activity pattern) (Abraham & Doane, 1978); another gene is postulated

to direct expression of α-amylase in the anterior midgut (Doane, et al, 1983).

In summary, cell-type specific gene expression in eukaryotes can be achieved by the binding of *trans*-acting factors to particular *cis*-acting DNA sequences. Tissue-specific gene expression of α-amylase in *Drosophila melanogaster* is expected to be directed by means similar to those already discussed. Isolating and characterizing the amylase gene would be the first step towards confirming this hypothesis and would enable us to explore other aspects of tissue-specific gene expression. For instance, by mutating upstream sequences of the gene, would it be possible to destroy tissue-specificity of expression without eliminating expression altogether?

1.4 The α-amylase gene of *Drosophila melanogaster*

The intent of this research project was to isolate and characterize the α-amylase gene from *D. melanogaster*. Subsequently, the gene would be microinjected into embryos of a Drosophila strain which do not produce viable α-amylase -- an α-amylase null strain (Haj-Ahmed & Hickey, 1982) -- to study the various levels of control of α-amylase gene expression *in vivo*.

We chose to study the α-amylase gene in *Drosophila melanogaster* because it provides a complex and interesting system of tissue-specific gene expression. α-amylase activity is expressed mainly in the midgut and hemolymph

of *D. melanogaster* (Abraham & Doane, 1978). As well, expression of the gene is controlled at various levels by *cis*- and *trans*-acting factors: it is subject to changes in its temporal and spatial expression during development (Doane, et al, 1983) and the level of expression can be manipulated through diet composition (Hickey, 1979; Hoorn & Scharloo, 1981).

The structural gene for α-amylase in *Drosophila melanogaster* has been mapped to a locus on chromosome 2R (Bahn, 1971). Among different strains of *D. melanogaster* there exist a number of polymorphic forms of α-amylase which can be distinguished on the basis of differences in electrophoretic mobility (Doane, 1969). Expression of two different isozyme variants within the same *D. melanogaster* strain led to genetic recombination studies by Bahn (1980) who determined that the *Amy* gene is actually comprised of tightly linked duplicate loci which are separated by 0.008 crossover units. In mammals, the amylase gene is also duplicated and evolution of the two copies has resulted in distinct pancreatic and salivary gland α-amylases (Mackay, et al, 1985).

Distribution of α-amylase expression within the posterior midgut (PMG) is genetically determined (Doane, et al, 1983) by different alleles of the linked gene designated *map* (Abraham & Doane, 1978). *map* activity is conspicuous in strains which produce identical amylase isozymes yet display distinct patterns of expression in the

PMG. Abraham and Doane (1978) suggest that *map* regulates expression of α-amylase via a *trans*-acting factor which binds to *cis*-elements on *Amy* within the PMG (Abraham & Doane, 1978; Treat-Clemons & Doane, 1980). Indeed, *in vivo* transcription and translation of amylase mRNA injected into Xenopus oocytes (LaBarca and Paigen, 1977) demonstrated that *map* controls *Amy* expression at the level of translatable mRNA (Doane, et al, 1983). The *map* gene also exerts a strain specific temporal 'switch' from larval to adult patterns of expression in the PMG (Hickey, 1981; Doane, et al, 1983). A similar temporal switch from neonatal to adult transcription patterns of the murine α-amylase gene , leads to tissue-specific promoter utilization in parotid and liver cells (Harding, et al, 1978; Shaw, et al, 1985). *map* may modulate amylase expression in *D. melanogaster* by switching on an 'adult' specific promoter. The *map* gene exerts its control only in the PMG and is, therefore, itself a tissue-specific gene. Strain specific activity patterns in the anterior midgut (AMG) suggest that there is probably a gene similar to *map* which controls the pattern of expression of α-amylase in the AMG (Doane, et al, 1983).

Additionally, α-amylase activity can be manipulated through the diet; high levels of amylase activity are characteristic of flies reared on complex carbohydrate media whereas flies raised on simple sugar media have low levels of α-amylase activity (Hickey, 1979; Hoorn & Scharloo,

1980). Correspondingly, the chromosome 'puffs' in the area to which the α-amylase gene has been mapped increase in size in flies fed on a starch carbohydrate source (Doane, et al, 1983). In *Drosophila melanogaster* 'puffing' of the chromosome is associated with increased transcription of genes in the region of the puff (Daneholt, et al, 1975). Accordingly, α-amylase mRNA levels are higher in flies reared on the starch diet as shown by translation of mRNA injected into oocytes (B. Benkel, personal communication). The phenomenon of diet-controlled expression is also seen for the rat α-amylase gene and is reflected in a corresponding concentration of amylase mRNA in the pancreas (Giorgi, et al, 1984).

Additional considerations made α-amylase an attractive system for study. We had available a mouse α-amylase cDNA to use as a probe for the Drosophila gene (a gift from U. Schibler, Swiss Institute for Experimental Cancer Research). A strain of α-amylase null flies (Donal Hickey) was available for the microinjection experiments. As well, the enzyme is easily isolated and the test for activity is uncomplicated. The enzyme is purified as a monomer with approximate molecular weight of 54,000 daltons (Doane, et al, 1983). It is a digestive enzyme which randomly hydrolyzes the α-(1-4) glycosidic bonds of starch to yield glucose and maltose molecules. Therefore, simple tests with I_2-KI reagent are sufficient to show α-amylase activity.

The α-amylase gene of *D. melanogaster* includes many aspects which make it interesting to study. Expression of the gene is controlled by *cis*- and *trans*-acting factors which affect its tissue-specific expression, developmental aspects of its expression and diet-regulated expression.

1.5 Approach to isolating the α-amylase gene of *D. melanogaster*

Originally, it was intended that the α-amylase gene of *D. melanogaster* would be isolated using a murine α-amylase cDNA as probe. Murine, human, and Xenopus homeobox genes have been isolated by virtue of homology with Drosophila sequences used as probes (Colberg-Poley, et al, 1985; Carrasco, et al, 1984; McGinnis, et al, 1984). Conversely, Drosophila genes have been isolated using mammalian genes or cDNAs as probes (Cowman, et al, 1985; Caggese, et al, 1986). Therefore, there is reason to expect that the α-amylase gene for *D. melanogaster* could be isolated by homology with the α-amylase gene from another animal. α-amylase has much the same digestive function in all eukaryotes and one might predict, therefore, a similarity in amino-acid sequence and possibly, to a lesser extent, in the mRNA among species. Indeed, for eukaryotic α-amylases characterized to date, there are four conserved domains in the protein which may be reflected in conserved DNA sequence domains (Mackay, et al, 1985).

We chose to use a complementary DNA for mouse pancreatic α-amylase (a gift from U. Schibler) to probe the

Drosophila melanogaster genome for the corresponding gene in Drosophila. Doane, et al, (1983) had conducted pilot experiments which established that a cDNA for mouse salivary α-amylase (Schibler, et al, 1980; Hagenbuchle, et al, 1980) had homology with discrete sequences in the *D. melanogaster* genome. Similarily, P. Moreau (unpublished data) showed that the pancreatic α-amylase cDNA from mouse had homology to discrete sequences in DNA from Oregon R strain of *D. melanogaster* and to DNA sequences from Drosophila tissue culture lines. These preliminary data indicated that it should be possible to isolate the α-amylase gene from *Drosophila melanogaster* using the mouse α-amylase pancreatic cDNA. In fact, several recombinant clones were isolated during this study using the murine probe. However, the low stringency of the screening conditions resulted in the isolation of clones, which appeared to be unrelated to each other. A more homologous probe was required to distinguish which, if any, of these clones encoded α-amylase. A second cDNA, pUCa1, which expressed a starch-hydrolyzing activity much like α-amylase, was made in our laboratory by M. Whitely from *Drosophila melanogaster* mRNA. The enzymatic activity expressed by pUCa1 from an expression vector and its hybridization to an mRNA species on Northern blots in the correct size range for α-amylase was convincing evidence that the cDNA encoded Drosophila α-amylase. Therefore, pUCa1 was used to screen the clones isolated

with the murine probe and also to re-screen the genomic library.

We were unaware that pUCa1 harboured a sequence almost identical to prokaryotic Insertion Sequence IS1. Yet, pUCa1 expresses an enzymatic activity akin to α-amylase which has never been described for IS1. Because the library of genomic DNA is propagated in a prokaryote host, by using the pUCa1 probe we were unknowingly looking for Insertion Sequences in the genomic library. Previously, two occurances of IS1 insertion into genomic libraries had been documented (Devos, et al, 1979; Fernandez, et al, 1986). However, the event documented in this thesis is distinct because IS1 has transposed into eukaryotic DNA.

We know now that the original premise of using the mouse pancreatic cDNA would have permitted the isolation of the Drosophila α-amylase gene because in 1985, Gemmill, et al, isolated and characterized the α-amylase gene from *Drosophila melanogaster* strain Canton S using a mouse salivary α-amylase cDNA as probe. The final proof that the gene is α-amylase is that the isolated sequence expresses α-amylase in Xenopus oocyte translation system (Levy, et al, 1985). Subsequently, α-amylase genes from six additional strains, including the amylase-null strain, were isolated and characterized (Gemmill, et al, 1986).

It is not within the scope of the Introduction to discuss all aspects of this project. However, these particular facts are pertinent because the surprising

results presented in this thesis changed the direction of the research. The use of pUCa1 as a probe to screen the Drosophila genomic library resulted in isolation of two clones which harbour a bacterial insertion sequence. In addition, these clones contain Drosophila genomic sequences. Therefore, this thesis documents the isolation of eukaryotic genomic sequences into which bacterial insertion sequences have integrated.

1.6 Prokaryotic Insertion Sequence IS1

Small DNA sequences called insertion sequences (IS) comprise approximately 1 to 2 percent of the genome of the bacterium, *Escherichia coli*. Insertion sequences are transposable elements that can move to new locations on the chromosome and also to extra-chromosomal elements such as plasmids and bacteriophage. Insertion sequences range in size from 768 to 5,000 bp and are flanked by terminal inverse repeats of 20 to 50 bp (Kleckner, 1977; Iida, et al, 1983). The inverted repeats are essential for transposition and probably serve as recognition points for the transposition mechanism by defining the ends of the element (Gamas, et al, 1985).

Larger prokaryotic transposable elements, some of which include IS at their termini, usually encode antibiotic or heavy metal resistance, making their presence easily identifiable in prokaryotes. IS, however, do not have marker phenotypes. IS were detected because of their strong polar

effects on target genes, that is, their ability to inactivate not only the gene into which they insert, but also genes downstream in the operon (Iida, et al, 1983). This polar effect is due to nonsense codons in most reading frames of the insertion sequence and served to distinguish IS mediated mutations from simple point mutations which usually only inactivate the mutated gene (Calos & Miller, 1980). That these polar mutations could revert to wild-type with concommitant loss of discrete DNA sequences at the reverted gene pointed to the mobile nature of the insertion mutation. Occasionally, insertion of an IS can cause activation of a flanking gene by readthrough from promoter sequences in the IS (Calos & Miller, 1980).

Models of IS transposition propose that insertion at the target site requires duplication of the element. The discovery of cointegrate structures, the proposed intermediate of transposition, argues for the replication model of transposition (Kleckner, 1977; Bukhari, 1981; Galas & Chandler, 1982; Braedt, 1985). Transposition proceeds, according to the models, by 1) introduction of a staggered break in both DNA strands at the target sequence, 2) fusion of the donor and recipient replicons by the insertion sequence, 3) replication through the insertion sequence and staggered cuts at each end of the insertion, hence generating the duplication in target sequences characteristic of each insertion sequence and, finally, 4) resolution of the cointegrate structure by deletion of

the donor replicon and one copy of the insertion sequence. Functions required for transposition and resolution are most likely encoded on the sequence itself (Bukhari, 1981; Kleckner, 1977).

Some insertion sequences have a stringent requirement for particular DNA sequences at the target site. These 'hotspots' are frequent sites for insertion of an element. For example, IS4 displays the most stringent sequence requirement and has only ever been shown to insert at one site on the gal operon (Chadwell, et al, 1977).

Insertion sequences can cause rearrangements, insertions and deletions within the host genome. As proposed by the transposition models, IS can fuse independently replicating genomes. These cointegrates are unstable and break down causing resolution of the genomes, each of which then carry an IS. IS also serve as portable regions of homology and can catalyze homologous recombination between unrelated DNA sequences by recombination of homologous IS within the host sequences. In addition, two IS flanking an unrelated DNA sequence, can cause inversion or deletion of the sequence by homologous recombination of the two IS and resolution of the recombined structure (Reif & Saedler, 1973). Large composite elements are actually comprised of DNA sequences flanked by IS copies which cause mobilization of the intervening DNA (Kleckner, 1977; Cornelis & Saedler, 1980; Galas & Chandler, 1982). Occasionally, an insertion sequence leaves its target site altogether. Excision is

independent of transposition (Calos & Miller, 1980; Machida, et al, 1982) although precise excision is probably also dependent upon exact recognition of the inverted repeats of the IS (Gamas, et al, 1985). Usually, excision is accurate and only the IS is deleted but, occasionally, excision of IS from a target site can cause deletion of some flanking host DNA (Kleckner, 1977).

IS1 is probably the best characterized of the IS of *E. coli* and related enteric bacteria. Sequencing data established that a modified copy of IS1 is present in pUCal (M. Whitely & P. Moreau, manuscript in preparation). IS1 is 768 bp long and includes inverted repeats of 20 to 23 bp at each end of the insertion sequence. IS1 generates a characteristic 8 or 9 base pair duplication of target DNA upon insertion (Johnsrud, 1979; Ohtsubo & Ohtsubo, 1978). Although IS1 does not have a stringent sequence requirement for its target site, it does have a preference for insertion into AT-rich DNA sequences and cleavage at the target site usually falls at a GC couplet (Ohtsubo & Ohtsubo, 1978; Kleckner, 1977; Iida, 1983).

IS1 was originally isolated by virtue of its insertion into the lactose operon of *E. coli* where it created a revertible mutation (Starlinger & Saedler, 1976). IS1, or related variants, have also been isolated from *Shigella dysenteria*, and from other operons of *E. coli* (Iida, et al, 1983). Approximately five to eight dispersed

copies of IS1 normally reside on the chromosome of *E. coli* strain K12 (Saedlar and Heiss, 1973).

IS1 has two long open reading frames on one DNA strand and several smaller open reading frames on both DNA strands (Machida, et al, 1984). It has been demonstrated that the two long open reading frames encode proteins (insA and insB) which are required for plasmid cointegration mediated by IS1. InsA and insB are transcribed from the same promoter in the same orientation. They are basic proteins which may have DNA binding potential (Machida, et al, 1984). The possible protein products of the other reading frames have not been determined but deletion studies indicate that these putative proteins would not be required for transposition functions (Machida, et al, 1984). The modified IS1 in pUCa1 is disrupted by a a one base pair deletion in one reading frame; the reading frame is restored by a one base pair insertion further downstream in the sequence (M. Whitely, unpublished data).

Characterization of the two recombinant clones isolated using pUCa1 as a probe points to the conclusion that each one also harbours a bacterial insertion sequence closely related to IS1. These two clones are interesting because the target sites for insertion of the prokaryotic sequences are *Drosophila melanogaster* genomic sequences. This raises questions as to how the selection of target site for insertion occurs and whether prokaryotic insertion sequences preferentially integrate at specific eukaryotic targets.

Are the targets in both cases the same? Were the two insertions independent events? Do the eukaryotic targets for these prokaryotic sequences share homology with targets for Drosophila transposable elements? These provoking questions may elucidate aspects of the evolution of all mobile genetic elements.

In addition, did insertion of IS1 into Drosophila sequences cause rearrangement of the target DNA? Could this type of rearrangement have evolutionary significance for Drosophila? For instance, the proximity of common bacterial strains, such as *E. coli*, and the gut cells of most eukaryotes, suggests that blurring of the boundaries between eukaryotic and prokaryotic DNA sequences may occur *in vivo*.

This event also has implications for the study of molecular biology. It is possible that some well-characterized genes isolated from genomic libraries of eukaryotic DNA also contain bacterial insertion sequences. Can we assume that all eukaryotic cDNA and genomic libraries contain insertion sequences? What is their frequency? Do bacterial strains commonly used to propagate these libraries contain insertion sequences? This work, therefore, poses some interesting questions which may have gone unanswered had not these surprising events taken place. In addition, it is a cautionary tale for those of us working with the common tools of modern molecular biology.

1.7 Thesis Outline

The next chapter outlines the methods used in this thesis. These methods were used in order to isolate the α-amylase gene of *Drosophila melanogaster*. Instead, bacterial insertion sequences which had inserted into Drosophila DNA were isolated. Nevertheless, the same methods for isolation and characterization of the DNA sequence are used despite the origin of the sequences. Chapter III is divided into two parts; the first part details the results of screening the genomic library; the second part describes data that led to identification of IS1 in the two recombinants isolated probing with pUCa1, and experiments which led to the conclusion that IS1 had integrated into Drosophila genomic sequences. Chapter IV is a discussion of the implications of this work and and the possible direction of further research based on the results obtained. Finally, Chapter V closes the thesis with concluding remarks.

Chapter II

METHODS AND MATERIALS

2.1 SCREENING DROSOPHILA GENOMIC LIBRARY

Techniques to isolate genes from total genomic DNA were developed by Benton and Davis (1977) and Maniatis, et al, (1978). Genomic DNA is isolated and fractionated, either by random exonuclease digestion or by partial digestion with a restriction endonuclease. The digestion strategy produces large random pieces of DNA rather than a biased population of specific fragments generated by limit digestion. The fractionated DNA is cloned into a suitable cloning vector (eg. plasmid, bacteriophage or cosmid DNA). The recombinant vector is propagated in an appropriate prokaryotic host and screened for sequences homologous to a chosen probe.

The Drosophila genomic DNA was fractionated following partial Sau3AI digest and cloned into bacteriophage lambda EMBL4 (Frischauf, et al, 1983). Sau3AI has a four base pair recognition sequence, G-A-T-C which is compatible for cloning into the BamHI sites in the lambda vector. Sau3AI sites should occur every 256 bp in a genome with 50 percent G+C content and only a fraction of these sites need be cleaved to produce fragments 20 Kb long (Karn, et al, 1980). The digested Drosophila DNA is enriched in the 14-20 Kb size range to avoid multiple insertions of small unrelated pieces of DNA. The entire Drosophila genome of $4x10^7$ bp can be represented in the equivalent of $1x10^4$ phage if the average insert size is 17 Kb (Maniatis, et al, 1978).

DNA was isolated from *D. melanogaster* strain Oregon R and cloned into bacteriophage lambda EMBL4 (European Molecular Biology Laboratories, Heidleberg, FDR). The phage genome was altered such that a substantial proportion of its gene complement was removed. The genes which were removed lie on an internal DNA fragment and are not required for phage propagation. This 'stuffer' fragment was replaced by the Drosophila DNA which was ligated to the left and right 'arms' of the phage genome; these remaining pieces of phage DNA encode all information essential for phage propagation. The recombinant DNA molecules were packaged *in vitro*.

Phage lambda cannot package genome sizes less than 80% or more than 120% of wild-type so that only those phage which have incorporated Drosophila DNA of sufficient size will be packaged (Karn, et al, 1980; Karn, et al, 1983). In addition, phage which may have been reconstituted with the parental stuffer fragment can be selected against by *Spi* selection (Karn, et al, 1983; Frischauf, et al, 1983). This selection system is based on growth restriction of the parental phage in a specific bacterial host strain. The restrictive host, Q359, carries a P_2 lysogen on the bacterial chromosome which does not allow growth of lambda phage carrying the stuffer fragment. *Spi* selection and genome size limitation make EMBL4 an ideal choice for the propagation of genomic libraries.

2.1.1 Plaque Lifts

The recombinant *Drosophila melanogaster* genomic library constructed in the vector EMBL4, was screened by the method of Benton and Davis (1977) and Maniatis et al (1982). Briefly, on the initial screening, 1x10⁴ pfu/150mmx15mm petri plate were plated on *Escherichia coli* strain Q358 or LE392 in 0.7% top NZCYM-agar on 1% agar-NZCYM plates (NZCYM medium is 10g/L N-Z Amine A, 5 g/L NaCl, 5 g/L Yeast Extract, 1 g/L Casamino Acids, and 2 g/L $MgSO_4$). Plaques were grown at 37°C for 12-16 hours until they were 2-3mm in diameter. The plates were transferred to 4°C to allow the top agar to harden. The plaques were replicated onto Biodyne A filters (Pall Co.) by laying a filter onto the surface of the plate for 1 minute. The filter and plate were marked to allow the plate to be aligned with the autoradiograph of the filter after hybridization. The filter was lifted from the agar surface and treated sequentially for 1 minute in 0.5 M NaOH, 1.5 M NaCl; for 5 minutes in 0.5 M Tris-HCl, pH 7.5, 1.5 M NaCl; then briefly in 2xSSPE (20xSSPE is 3.0 M NaCl, 0.2 M $NaPO_4$ pH 7.7, 0.02 M EDTA). The filters were air dried for 30 minutes then baked in a vacuum oven at 80°C for 1 hour. These filters were hybridized as outlined in Section 2.4.4.

2.1.2 Screening Positive Plaques

Plaques which hybridized the probe above background level on the initial screening were isolated and replated at lower density (50-200 pfu/100mmx15mm petri dish). Putative

positive plaques were isolated from the initial screening plates by placing the plate over an autoradiograph of the hybridized nylon filter and aligning the ink marks. The plaque corresponding to the hybridization signal was drawn through a sterile paster pipette. The plaque was resuspended in 1 mL SM buffer (0.1 M NaCl, 8.3 mM $MgSO_4$, 50 mM Tris, pH 7.5) and either replated at lower density for second or third screening or the phage DNA was isolated for analysis by restriction endonuclease digestion.

2.2 BACTERIOPHAGE LAMBDA ISOLATION

2.2.1 Stocks of Lambda Bacteriophage Library

A stock of the *Drosophila melanogaster* genomic library constructed in EMBL4 was made by plating $2x10^4$ pfu/150mmx15mm petri plate. The plates were incubated overnight at 37°C then placed at 4°C for 1 hour to harden the top agar. The plates were then overlayed with 10 mL of cold SM and agitated gently for at least 1 hour. The liquid was pipetted from the plates and $CHCl_3$ was added. Cellular debris and agar were removed from the supernatant by centrifugation. The titre of the library was determined (Maniatis, et al, 1980).

For stocks of purified recombinant bacteriophage, one purified plaque was picked from a plate and resuspended in 1 mL SM buffer with a few drops of $CHCl_3$ then stored airtight at 4°C (Davis, et al, 1980).

2.2.2 Small Lysates

Bacteriophage lambda were grown on NZCYM-1% agar media (Maniatis, et al, 1980) plated at a density of approximately 100 pfu/100mmx15mm petri plate. Individual plaques were picked and eluted into 1 mL of SM Buffer plus 3 drops $CHCl_3$. These were agitated or allowed to stand at room temperature until phage dissipated into solution.

100 µL of the phage-infused SM was added to 100 µL of stationary phase host *Escherichia coli* strain and to this 4 mL of NZCYM broth were added. This was then agitated with aeration at 37°C for approximately 6 hours. Once lysis was apparent, 100 µL of $CHCl_3$ was added to lyse any remaining infected intact cells and agitated for an additional 15 minutes. The cellular debris was centrifuged at 4°C at 4000xg for 10 minutes. At this point the small lysate can be titred for phage and can be used to infect a large preparation of cells or the DNA can be extracted. This is also the method of choice for keeping stocks of individual recombinant bacteriophage lambda clones.

2.2.3 Large Scale Lambda Isolation

The small lysate can be used to do a large-scale infection of cells. $1x10^{10}$ cells/500 mL NZCYM broth were infected to a multiplicity of infection (m.o.i.) of 0.5. The culture was agitated for 6 hours at 37°C at

which time lysis was apparent. $CHCl_3$ (2.5 mL) and approximately 1 µg/mL DNAase I were added to the lysing cells and agitated for a further 15 minutes. Cell debris was removed by centrifugation at 8000xg for 10 minutes at 4°C. NaCl and PEG 8000 (Fisher Scientific Co.) were added sequentially to final concentrations of 0.5 M and 10%, respectively and left to stand for at least 1 hour at 4°C to precipitate the phage particles. The precipitated phage were collected by centrifugation at 4°C at 8000xg for 10 minutes and then resuspended by rotation in 10 mL cold SM buffer. Cesium chloride (IBI, Boeringher-Mannhfem) was added to 0.75 g/mL the phage were then centrifuged to equilibration on the CsCl gradient at 4-15°C using a Beckman L8-70 ultracentrifuge (Beckman, Quik-Seal Polyallomar tubes) for 4-16 hours at 45,000-25,000 rpm, respectively. The band of bacteriophage particles was removed from the gradient through a syringe and dialyzed against a large volume of SM buffer at 4°C to remove Cesium Chloride.

2.3 DNA ISOLATION

2.3.1 Small Scale Lambda DNA Preparation

Small DNA preparations from recombinant phage were made for preliminary analysis of insert DNA. DNAase I was added to small lysates of the recombinant phage (Section 2.2.2) to a final concentration of 10 µg/mL and incubated at 37°C for 1 hour. NaCl and PEG 8000 (Fisher Scientific Co.) were

added sequentially to a final concentration of 0.5 mM and 10%, respectively, and incubated 1 hour at 4°C. The PEG precipitated phage were collected by centrifugation and the pellet was resuspended in 0.2 mL cold SM buffer. This was extracted three times with an equal volume of $CHCl_3$. EDTA was then added to a final concentration of 10 mM. This was then extracted sequentially with equal volumes of phenol, phenol/chloroform, and chloroform. The DNA was precipitated in 0.3 M NaCl, 2 volumes absolute ethanol at -20°C and collected by centrifugation. The pellet was dried and resuspended in 100 µL TE (10 mM Tris, pH 7.5-7.9, 1 mM EDTA, pH 8.0). This crude DNA was stored at -20°C.

2.3.2 Large Scale Lambda DNA Preparation

DNA of better quality was isolated from phage taken from CsCl gradient (Section 2.2.3) Phage dialyzed against SM were treated with 10 mM EDTA, pH 8.0, 0.3% SDS and 100 mg/mL Proteinase K (Sigma Co.) and incubated at 37°C for 1 hour to destroy the phage protein coat. The DNA was then extracted three times with an equal volume of phenol. The first phenol interface was re-extracted with 200 µL TES (10 mM Tris, pH 7.5-7.9, 1 mM EDTA, 100 mM NaCl). The resulting final aqueous phase was dialyzed against a large volume of TES for 3-12 hours at 4°C. The concentration of DNA was estimated by spectrophotometry at 260 nm.

2.3.3 Large Scale Plasmid DNA Preparation

Plasmid DNA was isolated from host cells (C600, HB101) by the method of Kahn, et al, (1979). Competent bacterial cells transfected with the plasmid to be isolated were grown overnight in 500 mL L-broth (10 g/L Tryptone, 5 g/L Yeast Extract, 10 g/L NaCl) at 37°C with agitation in the presence of the appropriate selective antibiotic. The cells were centrifuged to a pellet at 4,000xg for 15 minutes. The cell pellet was resuspended in 10 mL 20% sucrose, 40 mM Tris, pH 7.5. The resuspended cells were treated with 2ml (30 mg/mL) of lysozyme (Sigma Chemical Corp.) for 10 minutes on ice; 10 mL 0.2 M EDTA, pH 8.0 for 10 minutes; and finally, with 10 mL 2% Triton. The resulting lysate was centrifuged for 45 minutes at 35,000 rpm in a 42.1 Ti rotor (Beckman) to pellet the cellular debris. Cesium chloride was added to the supernatant to 1.03 g/mL and ethidium bromide to 1.3 mg/mL. The plasmid DNA was equilibrated on the CsCl gradient at 44,000 rpm for 14-16 hours in a Beckman 50vTi rotor in Beckman 40 mL Quik-Seal ultracentrifuge tubes in a Beckman L8-70 ultracentrifuge.

The (supercoiled) plasmid DNA was removed from the gradient through an 18 gauge needle into a syringe. Ethidium bromide was removed by four extractions with equal volume of isopropanol equilibrated with TE and saturated with CsCl. The crude DNA was dialyzed against 2 L of TE for 2 hours. The DNA was further purified by treating

with 10 µg/mL RNAase A (Sigma Co.) for 1 hour at 65°C followed by 100 µg/mL Proteinase K (Sigma Co.) and 0.2% SDS for 2 hours at 37°C. The DNA was finally phenol extracted twice and extracted once with $CHCl_3$. The purified DNA was precipitated in 100 mM NaCl, 2 volumes absolute ethanol at -20°C for 1-12 hours. The precipitate was collected by centrifugation, lyophilized and resuspended in 0.5 mL TE. The concentration of DNA was estimated by spectrophotometry at 260 nm.

2.3.4 Rapid Plasmid DNA Preparation

For quick analysis of plasmids, rapid DNA isolation by the alkaline-lysis method of Birnboim and Doly (1979) was carried out.

2.3.5 Drosophila Genomic DNA

500 mg live weight flies (Strain Oregon R) were homogenized in a sterile glass Dounce homogenizer in 1 mL Buffer A (10 mM Tris, pH 7.5, 60 mM NaCl, 10 mM EDTA, 0.2 mg/mL Proteinase K). The buffer was equilibrated to 37°C for 1 hour before use. After homogenization the Dounce was rinsed with an equal volume of Buffer B (0.2 M Tris, pH 9.0, 30 mM EDTA, 2% SDS, 0.2 mg/mL Proteinase K). The two extracts were combined and incubated at 37°C for a minimum of 1 hour.

The homogenate was then extracted with an equal volume of phenol equilibrated with TE. The interface of the first phenol extraction was re-extracted with 200 µL TE.

Following phenol extraction, the aqueous phases were combined and extracted with an equal volume of $CHCl_3$. The crude nucleic acids were precipitated from the final aqueous phase at -20°C for 1-12 hours by addition of NaCl to 100 mM and 2 volumes of cold (-20°C) absolute ethanol.

Precipitate was collected by centrifugation at 10,000 rpm at 4°C in a Sorvall RC2-B then dried under vacuum. The precipitate was resuspended in 2 mL TE, NaCl was added to 100 mM and RNAase A (Sigma Co.) to 10 µg/mL and the solution was incubated at 37°C for 1 hour. Next, the DNA was treated with 50 µg/mL Proteinase K at 37°C for 1 hour. The DNA was phenol and $CHCl_3$ extracted and reprecipitated in 100 mM NaCl, 2 volumes cold absolute ethanol. The precipitate was resuspended in 1 mL TE. DNA concentration was estimated by spectrophotometry at 260 nm.

2.3.6 Bacterial DNA

Bacterial DNA was isolated from stationary phase cells of strains C600, Q358, or LE392. The method follows that of Sato and Miura (1963) with modifications. 1.5 mL of cells were centrifuged and the pellet was resuspended in 50 µL of a solution of 2 mg/mL lysozyme, 0.15 M NaCl, 0.1 M EDTA, pH 8.0. The cells were then incubated at 37°C until lysis was apparent (approximately 10 minutes). The lysate was quick frozen in a dry ice-ethanol bath. 125 µL of 0.1% SDS, 0.1 M NaCl, 0.1 M Tris-HCl, pH 9.0, was added to the frozen cells. The cells were then allowed to thaw ~~slowly~~ on ice.

Once thawed, the lysed cells were extracted 1-4 times with phenol or until the aqueous phase became cleared of debris. Then the aqueous phase was extracted with $CHCl_3$. The nucleic acids were precipitated by addition of 100 mM NaCl and 2 volumes cold absolute ethanol. The nucleic acids were collected by centrifugation and resuspended in 100 µL TES. To this, 5 µL 20xSSC and 3 µL RNAase at 5 mg/mL were added. The crude DNA was then incubated at 37°C for 30 minutes. The DNA was extracted twice with phenol, once with $CHCl_3$, and precipitated with addition of 2 volumes cold absolute ethanol. The final precipitate was collected and resuspended in 100 µL TE.

2.4 DNA ANALYSIS

2.4.1 Restriction Enzyme Digests

Purified DNA was cleaved with various restriction endonucleases by the method of Davis, et al, (1980) and Maniatis, et al, (1980). Briefly, an appropriate amount of DNA (0.1-1.0 µg) dissolved in TE or TES, a required amount of 5x restriction buffer of the approriate salt composition and 3-5 units of enzyme/µg of DNA (New England Biolabs, Boeringher-Mannhiem, Bethesda Research Labs), with the volume completed with sterile, double-distilled H_2O, was digested at 37°C for 1-1.5 hours. Restriction enzyme buffers were typically made with high salt (100 mM NaCl), medium salt (50 mM NaCl), or low salt (0 mM NaCl) in 50 mM Tris, pH

7.5, 10 mM $MgCl_2$. The enzyme was inactivated by incubating at 65°C for 15 minutes or by addition of an equal volume of phenol followed by $CHCl_3$ extraction and ethanol precipitation. For digests of lambda DNA, treatment at 65°C for 10 minutes also denatures the cohesive (cos) ends of the lambda DNA.

Genomic DNA was digested with 6-12 units of enzyme/µg of DNA for 12-18 hours. An internal standard of EMBL4 wild-type DNA was included in genomic DNA digests to indicate when the digest was complete.

A combination of pBR322 and rabbit ß-globin gene DNAs digested with HinfI, EcoRI, HindIII, BamHI serve as molecular weight markers. The molecular weights of the marker fragments are listed (bp): 15,000, 8,600, 7,600, 6,600, 6,000, 4,365 (pBR322 cleaved at single EcoRI site), 2,900, 2,150, 1,630, 1,260-1,200 (doublet), 516-506 (doublet), 396, 344, 298, 221-220 (doublet), 154, 75.

2.4.2 Agarose Gel Electrophoresis

DNA cleaved with restriction endonucleases was separated on agarose gels (0.7% - 1.4%) electrophoresed in TAE (0.04 M Tris-acetate, 0.001 M EDTA) or TBE (0.08 M Tris-phosphate, 0.002 M EDTA) buffer. Higher percentage gels were used for analysis of small DNA fragments of less than 1 Kb while low percentage gels were used to separate higher molecular weight DNA. Gels were stained with ethidium bromide (1 µg/mL of gel) and visualized under shortwave ultraviolet light.

2.4.3 Southern Transfer

DNA was digested and electrophoresed as described in the previous section usually on 0.8% agarose gels. Transfer of DNA to membrane was carried out according to the methods of Southern (1975), Meinkoth and Wahl (1982), and Reed and Mann (1985), using nylon (Amersham Hybond-N), modified nitrocellulose (Bio-Rad Z-probe), or nitrocellulose membranes (Schleicher and Schuell). Transfer to nylon or modified nitrocellulose was carried out in 0.025 M NaPO4, pH 6.5 or 20xSSC and to nitrocellulose in 20xSSC. Genomic DNA was transferred to nitrocellulose only.

2.4.4 DNA Hybridization

The first probe used on the Drosophila library was the α-amylase cDNA from mouse pancreas. The evolutionary distance between mouse and Drosophila suggests that there is a minimum of homology between coding sequences for α-amylase in each organism. Therefore, it was necessary to maximize hybridization parameters which would allow these two regions of presumed low homology to form stable hybrids.

Several parameters affect the hybridization of two strands of DNA. These parameters affect the rate of hybrid formation or hybrid stability. The cardinal parameter of hybrid formation is, obviously, the degree of homology between the two sequences. The other parameters can be adjusted to allow sequences with low homology to form stable hybrids.

The rate of hybrid formation is maximized with appropriate salt concentration (>0.4 M NaCl) (Dove & Davidson, 1962), optimum pH (in the range from 7.0-9.0), and by addition of anionic dextran polymers (Wahl, et al, 1979) which decrease actual volume of hybridization. The optimal length of the probe is between 500 and 1,500 bp. The degree of complexity of both the probe and gene also affects the rate of hybridization, but for any two pieces of DNA this is constant.

Hybridization of two DNA sequences is a reversible process and by understanding the factors which affect hybrid stability one can optimize conditions to allow formation of stable hybrids between two sequences with a high percentage of mismatched base pairs (Meinkoth & Wahl, 1984). We expected the mouse cDNA to form many mismatched base pairs with the Drosophila sequence we were seeking. The equation below (Meinkoth & Wahl, 1984) describes the effects of salt concentration, base composition (which is constant for any given probe and gene), the length of hybridizing regions, and the concentration of helix destabilizing agents such as formamide on the melting temperature (T_m) of duplexes:

$$T_m = 81.5^oC + 16.6\log M + 0.41(\%G+C) - 500/n - 0.61(\%\text{formamide})$$

M = salt concentration (mol/liter)
G + C = base composition
n = length of the shortest chain in the duplex

The T_m is approximately 25°C above the optimal temperature for hybrid formation for any given pair of sequences. The T_m decreases by approximately 1°C for every

mismatched base pair between the two strands of the duplex (Bonner, et al, 1973).

The stringency of hybridization is, therefore, a measure of the degree of homology expected between two strands of DNA. In the hybridization of the mouse α-amylase cDNA to the corresponding gene in *Drosophila melanogaster*, the stringency was lowered to allow for low homology of the two sequences. Stringency can be altered during hybridization by changing the formamide concentration or temperature, or during post-hybridization washes by changing salt concentration and temperature.

Hybridization was performed as described by Wahl, et al, (1979), Thomas (1980), and Maniatis, et al, (1982) in a sealed plastic bag. Prehybridization of the Southern transfer or plaque lift filters was for 4 hours at either 37°C or 42°C (depending on the stringency requirements) in the following conditions: 40%-50% formamide (depending on stringency) (BDH Co.), 5xPAF (50xPAF or Denhardt's Solution is 1% Ficoll (Sigma Chemical Co.), 1% Polyvinylpyrrolidone, 1% Bovine Serum Albumin), 5xSSPE, 0.1% SDS, 0.2 mg/mL denatured salmon sperm DNA. Hybridization solution conditions were as follows: 50%-40% formamide, 2xPAF, 5xSSPE, 0.1% SDS, 0.2 mg/mL denatured salmon sperm DNA, 6%-10% dextran sulfate (Wahl 1979), and 10^7-10^9 cpm/ug of radiolabelled probe (Maniatis, et al, 1980) for 12-18 hours at 37-42°C.

Post-hybridization washes were in the following conditions: twice for 15 minutes at room temperature in 2x SSC (20xSSC is 3.0 M NaCl, 0.3 M Sodium Citrate buffered to a final pH of 7.0), 0.1% SDS, followed by 2 washes for 30 minutes at 50°C or 65°C (depending on stringency) in a solution of 0.1xSSC to 2xSSC (again, depending on stringency requirements), 0.1% SDS.

The Southern transfer or plaque lifts were then autoradiographed for 4 hours to 7 days at -70°C with a Dupont Cronex Lightning Plus intensifying screen on Kodak XAR-5 X-Ray film.

2.5 RADIOLABELLING OF DNA

2.5.1 Nick Translation

Approximately 0.1 - 1 µg purified DNA was labelled in a nick translation reaction according to the method of Rigby et al (1977) and Kelly, et al, (1970).

100 µCi of α^{32}P-dCTP (sometimes with 100 µCi α^{32}P-dATP, as well) (stabilized aqueous solution, 3000 Ci/mMol, Amersham Corp.) were used in a total reaction volume of 100 µL. The reaction conditions were as follows: 25 mM KH_2PO_4, pH 7.4, 5 mM $MgCl_2$, 10 µM each dTTP, dATP (unless used as a second label), dGTP, 50 µg/mL BSA (Pentax Fraction V, Boeringher-Mannheim) and double-distilled sterile H_2O to make up the volume. Activated (1 pg) DNAase I, (Sigma Chemical Corp.) was added for 1 minute at 15°C followed by addition of 5 units of DNA

polymerase I (Boeringher-Mannheim). The reaction was then incubated at 15°C for 1 hour. The reaction was terminated by addition of 0.5 mL Stop Buffer (100 mg/mL denatured salmon sperm DNA in HENS Buffer (0.1 M HEPES, pH 7.2, 1 mM EDTA, 0.01 M NaCl, 0.1% SDS)) and heating the reaction to 65°C for 10 minutes. Unincorporated nucleotides were separated from incorporated products by passing the reaction through a 10 mL Sephadex G50 medium column equilibrated with HENS Buffer. Just prior to use in hybridization reactions, the probe was denatured by boiling for three minutes.

Alternately, double strand probes were made to high specific activity (10^9 cpm/µg) using the Multiprime System (Amersham Corp.). This procedure involves denaturation of the DNA to be labelled, hybridization of a random hexamer primer to the DNA, synthesis of DNA from the primer using *E. coli* DNA polymerase I (Klenow fragment) and incorporating α^{32}P-dCTP (3000 Ci/mmol, Amersham Corp.). The radiolabelled probe was denatured from the unlabelled template by boiling for 3 minutes just prior to use. With this labelling procedure there is no requirement for removal of unincorporated label as incorporation is typically above 60%.

2.5.2 Single Strand Probe Preparation

Small preparations of single-strand M13 recombinant bacteriophage were made for the purpose of preparing single strand radiolabelled probes. Single plaques of infected cells were grown with host JM 101 for 6.5 hours in 5 mL L-broth. The infected cells were separated from phage in the supernatant by centrifugation. 200 µL of 40% PEG 8000 (Fisher Scientific Co.) and 2.5 M NaCl were added to the supernatant then incubated at room temperature for 15 minutes. The phage were pelleted by centrifugation. Phage proteins were removed by extracting twice with phenol, once with phenol/chloroform, and once with chloroform. The resulting single-strand DNA was precipitated in 100 mM NaCl, 2 volumes cold absolute ethanol, at -20°C. The final pellet was resuspended in 40 µL TES.

100 ng single strand phage DNA was annealled to 0.6 ng 17-mer sequencing primer (Amersham Corp.) The primer was extended in a reaction mixture comprised of 10xHIN Buffer (100 mM Tris, pH 7.5, 500 mM NaCl, 50 mM $MgCl_2$, 10 mM Dithioerythritol) 50 µM dTTP,dATP,dGTP, and 20 µCi of α-^{32}P-dCTP (3000 Ci/mMol, Amersham Corp.) in the presence of 0.5 units DNA polymerase I large (Klenow) fragment (Boeringher-Mannheim) at 15°C for 30 minutes. The unincorporated nucleotides were separated from incorporated products as described for nick translation products in Section 2.5.1. Before use for hybridization, the probe was denatured by boiling for 3 minutes.

2.6 SUBCLONING

2.6.1 Isolation of DNA fragments for subcloning

Fragments to be subcloned were isolated from 1% agarose gels. The plasmid or phage DNA containing the fragment of interest was digested with the appropriate restriction enzyme to isolate the fragment. The digest was run through a 1% agarose gel in TAE buffer at 100 volts for 1-2 hours or at 20 volts for 12 hours. The DNA fragment was isolated by cutting a section of agarose from the gel just in front of the fragment. The fragment was electrophoresed into the TAE filling this cavity. The TAE in the cavity was collected and passed through an Elutip (Schleicher and Schuell) to remove impurities. The DNA eluted from the Elutip was then precipitated in 100 mM NaCl, 2 volumes of cold ethanol. The precipitate was collected by centrifigation. An aliquot of the isolated DNA fragment was run through an agarose gel to estimate the concentration of the fragment.

2.6.2 Ligation of fragment to vector

The fragment was ligated to plasmid vector by the cohesive ends of the DNA molecules. The vector DNA was dephosphorylated with Calf Alkaline Phosphatase (Boeringher-Mannheim) at 25 units/µL at 37°C for 1 hour. Dephosphorylation was carried out simultaneously with restriction enzyme digestion in restriction enzyme buffer. The reactions were terminated by addition of phenol followed by chloroform extraction and precipitation of the DNA.

DNA at a ratio of approximately 2 fragment molecules to one vector molecule was ligated in Eppendorf tubes in a volume of 20 µL. The reaction was carried out in 0.6 mM ATP (Sigma Corp.), 20 mM Tris, pH 7.5, 10 mM $MgCl_2$, 10 mM dithioerythritol with the addition of 1-2 units T4 DNA Ligase (Boeringher-Mannheim) at room temperature for 2-4 hours or for 12 hours at 15°C. The ligation products were transfected into competent *E. coli* cells and plated on appropriate selective antibiotic supplemented L-broth agar plates.

Chapter III

RESULTS

PART 1 Isolation of the Drosophila α-Amylase Gene

3.1 SERIES 100

3.1.1 Probing with a Mouse Amylase cDNA Series 100

The library of *Drosophila melanogaster* genomic DNA (strain Oregon R) cloned into the lambda vector, EMBL4, was constructed in the European Molecular Biology Laboratories (Heidleberg, FDR). For the first plating (Series 100), $5x10^4$ pfu ($1x10^4$ pfu/150mmx15mm petrie plate) were used to infect host strain Q358. Considering that the average insert size in EMBL4 is 15 Kb (Frischauf, et al, 1983; Karn, et al, 1983) and that the haploid genome of *D. melanogaster* is $4x10^7$ bp, this represents the equivalent of five *Drosophila melanogaster* genomes. Series 100 was screened with the mouse pancreatic α-amylase cDNA (U. Schibler, Swiss Institute for Experimental Cancer Research). The α-amylase cDNA was excised from a plasmid constructed in our laboratory, pHGCaR. This construct allows excision of the complete cDNA via BamHI sites at either end.

The first screening yielded approximately 50 plaques which gave signals of varying intensity above background. 50 plaques represent 0.1% of the total number of plated plaques. Assuming that the library is composed of single copy sequences (Maniatis, et al, 1980), one plaque per $1x10^4$ or five plaques from this plating, should be α-amylase

clones. Thus, not all of the 50 hybridizing clones should represent α-amylase sequences.

Because a low homology murine probe was used, Series 100 was hybridized under low stringency conditions (hybridization was in 40% formamide at 37°C). Post-hybridization washes were as follows: twice in 2xSSC at room temperature, followed by two washes in 2xSSC at 50°C for 15 minutes. Using the equation from Section 2.4.4, we estimate that hybridization and stringency wash conditions impose a minimum duplex length of 12 bp assuming a 50% G+C composition of the probe.

The plaques which hybridized above background did so with varying intensity. At this point in the screening procedure it would have been misleading to select plaques for further screening on the basis of the relative intensity of the hybridization signal above background. The plating density of the library results in the plaques being almost confluent which may prevent strong signals from some potential positive clones. Therefore, the first potentially positive plaques were isolated on the basis of a higher than background signal but randomly as to the strength of that signal. Fourteen plaques were picked for further screening (designated 101 to 114).

Plaques were picked as outlined (Section 2.1.1) but, invariably, because of the high plating density, potentially positive plaques were still contaminated with phage from surrounding plaques. In fact, the contaminating negative

phage served as a negative control for positive phage. The second screening was plated at lower density (50-100 pfu/100mmx15mm petrie dish) to isolate plaques from each other, positive from negative. Twelve of the original fourteen plaques still gave a positive hybridization signal. The signal of each positive plaque was more intense than the signal seen in the first screening.

The third screening served to purify the positive plaques further. Eight of the original fourteen remained positive through this screening. The four eliminated between the second and third screenings were re-screened to ensure that they were negative.

3.1.2 Isolation and characterization of DNA from Series 100 recombinant clones

DNA from small lysates of the eight positive clones (101, 105, 107, 109, 110, 111, 113, 114) was isolated for preliminary characterization. DNA was digested with EcoRI which should excise the insert DNA from the vector. There are two EcoRI sites in the vector and these flank the two BamHI sites used to clone the Drosophila DNA. However, the BamHI sites are not necessarily regenerated upon ligation of the vector to the Drosophila DNA.

The fragments generated by EcoRI were run through a 0.8% agarose gel, transferred to Z-probe membrane (Biorad), and hybridized to the murine -α-amylase cDNA under low stringency conditions. Table 3.1a lists the sizes of the

EcoRI fragments of each clone with the hybridizing fragment underlined.

Each of these eight clones gave a distinct restriction pattern (Figure 3.1a) and the probe hybridized to a single EcoRI fragment in each of the eight clones (Figure 3.1b). The EMBL4 vector digested with EcoRI generates three fragments (Figure 3.1a, lane 9); the left arm (20.4 Kb), right arm (9.2 Kb) and stuffer fragment (13.4 Kb). There was no hybridization of the probe to the lambda vector (Figure 3.1b, lane 9). However, the probe hybridized a 9.2 Kb fragment in two of the clones (105 and 110) which co-migrates with the shorter of the vector arms (lanes 2 and 5, respectively, Figure 3.1a,b). If these two clones continued to generate identical restriction and hybridization patterns during subsequent analysis, it might have indicated that they are overlapping clones.

As an additional control, the same blot was dehybridized and re-probed with pBR322 which did not hybridize to any of the Drosophila fragments under the same hybridization conditions (data not shown).

There is considerable variation in the strength of hybridization signal for each hybridizing fragment. This can be explained, for example, for clones 109 and 111, by the corresponding amount of DNA on the gel. However, for clones 105 and 114, for example, approximately the same amount of DNA was loaded on the gel, yet 105 gives a

TABLE 3.1 SERIES 100
ECORI & KPNI/ECORI FRAGMENT SIZES
(Kb)

a) EcoRI

101	105	107	109	110	111	113	114	EMBL 4
20.5	20.5	20.5	20.5	20.5	20.5	20.5	20.5	20.5
							12.2	13.4
9.20	9.20	9.20	9.20	9.20	9.20	9.20	9.20	9.20
		6.6						
		5.3	5.4					
		4.8	4.8					
				4.2		4.3		
					3.9		3.7	
3.4			3.3	3.4	3.6			
3.3								
	3.0							
2.6								
2.2	2.1				2.2	2.2		
						2.0		
			1.8	1.8	1.9	1.8		
			1.6	1.6		1.2		
		1.3						
						*		

* small fragment not visible on gel hybridizes the probe at 1 Kb

b) KpnI/EcoRI

101	105	107	109	110	111	113	114	EMBL 4
20.5	20.5	20.5	20.5	20.5	20.5	20.5	20.5	17.5
							12.2	13.4
9.20	9.20	9.20	9.20	9.20	9.20	9.20	9.20	9.20
		6.6						
			5.4					
	4.2		4.8					
				4.2		4.3		
					3.9		3.7	
3.4			3.3		3.6			
3.2	3.0	3.0						
2.6		2.6d		2.5				
2.2					2.2	2.2		
	2.0					2.0		
			1.8		1.8	1.8		
			1.6	1.6				
1.2	1.2	1.2	1.2	1.2d	1.2d	1.2	1.2	1.2
						*		

* small fragment not visible on gel hybridizes the probe at 1 Kb

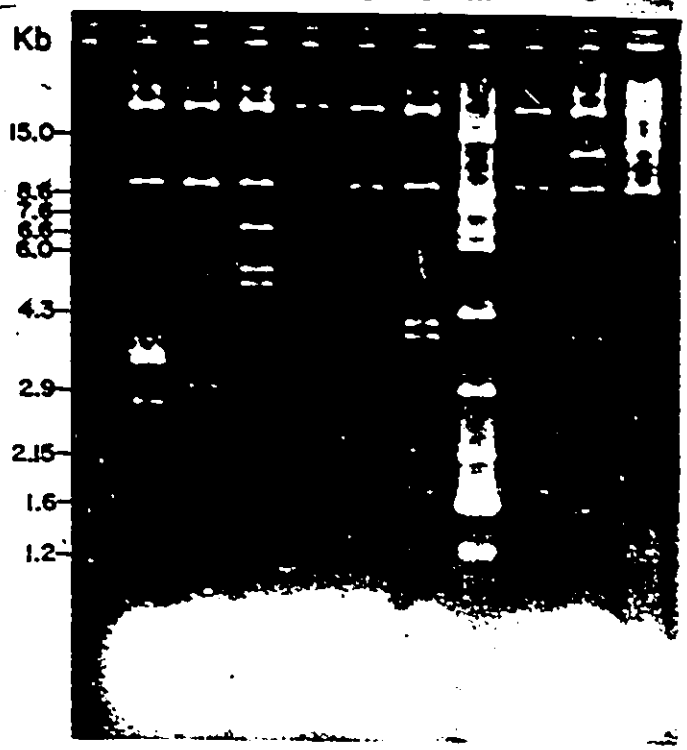
FIGURE 3.1

DNA isolated from Series 100 lambda recombinant clones was EcoRI digested, fractionated on a 0.8% agarose gel in TAE, blotted on a nylon filter and hybridized for 12 hours at 37°C in the presence of 40% formamide with pMPa21 nick-translated plasmid (specific activity 1x10⁸ cpm/µg).

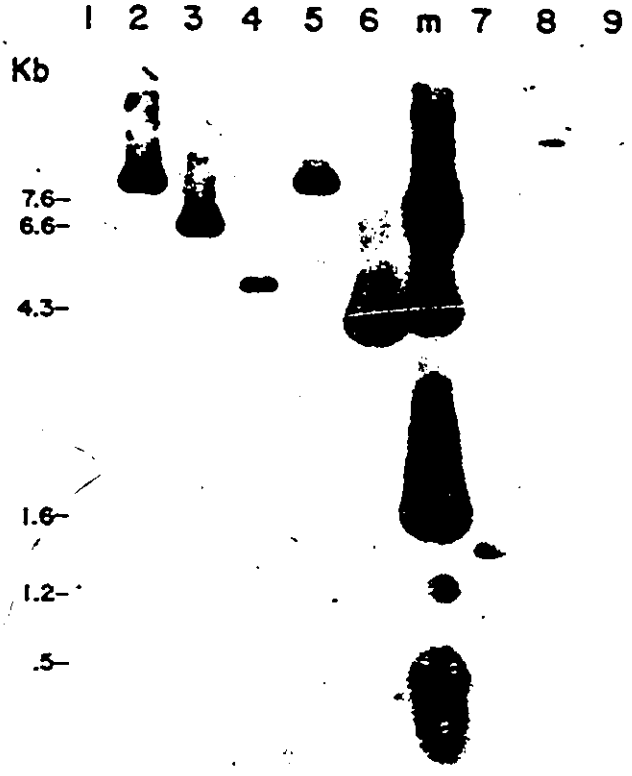
Panel a is the ethidium bromide stained gel with marker fragments (lane m) indicated. Panel b is the autoradiograph of the hybridized filter exposed with an intensifying screen for 24 hours with the marker fragments (lane m) indicated.

Lane 1) 101, 2) 105, 3) 107, 4) 109, 5) 110, 6) 111 m) 1 µg Marker, 7) 113, 8) 114, 9) EMBL4

a.



b.



qualitatively stronger signal than 114. This may be due to more homology of 105 to the probe.

Subsequently, the eight recombinant clones were digested with KpnI and HindIII and with these enzymes in combination with EcoRI. Tables 3.1b, 3.2a, and 3.2b, give the respective sizes of fragments from these digests and the hybridizing fragment for each clone is underlined.

The two 9.2 Kb fragments generated in the EcoRI digests of 105 and 110 generate fragments with different sizes and hybridization patterns in the double digest with KpnI; 105 gives a 4.2 Kb hybridizing fragment in the double digest, whereas the 9.2 Kb fragment generated from 110 by EcoRI is not cleaved by KpnI (lanes 2 and 5, respectively, Figure 3.2). It is expected that identical sequences from the same DNA strain would have identical restriction patterns (Maniatis, et al, 1980; Gemmill, et al, 1986).

Because of the expected low homology of the murine probe to the Drosophila α-amylase gene, we had used low stringency to probe the genomic library. We anticipated that we would isolate some clones which hybridized to the murine probe but which were unrelated to *D. melanogaster* α-amylase. However, as well, we expected to isolate overlapping clones comprising the α-amylase gene. The dissimilar restriction patterns of the eight isolated clones suggested that they did not overlap.

TABLE 3.2

SERIES 100
HINDIII & HINDIII/ECORI FRAGMENT SIZES
(Kb)

a) HindIII

101	105	107	109	110	111	113	114	EMBL 4
23	25	23	21	21	23d	20	23	23.
		11.0	10.0				18	
7.6		7.6	6.6			8.6		
6.0	6.0							5.7
				5.3				5.5
5.0				5.0	5.0			
4.5	4.5	4.5	4.5	4.5d	4.5	4.5	4.5	4.5
	4.3			4.3				
			3.3	2.3		2.5		

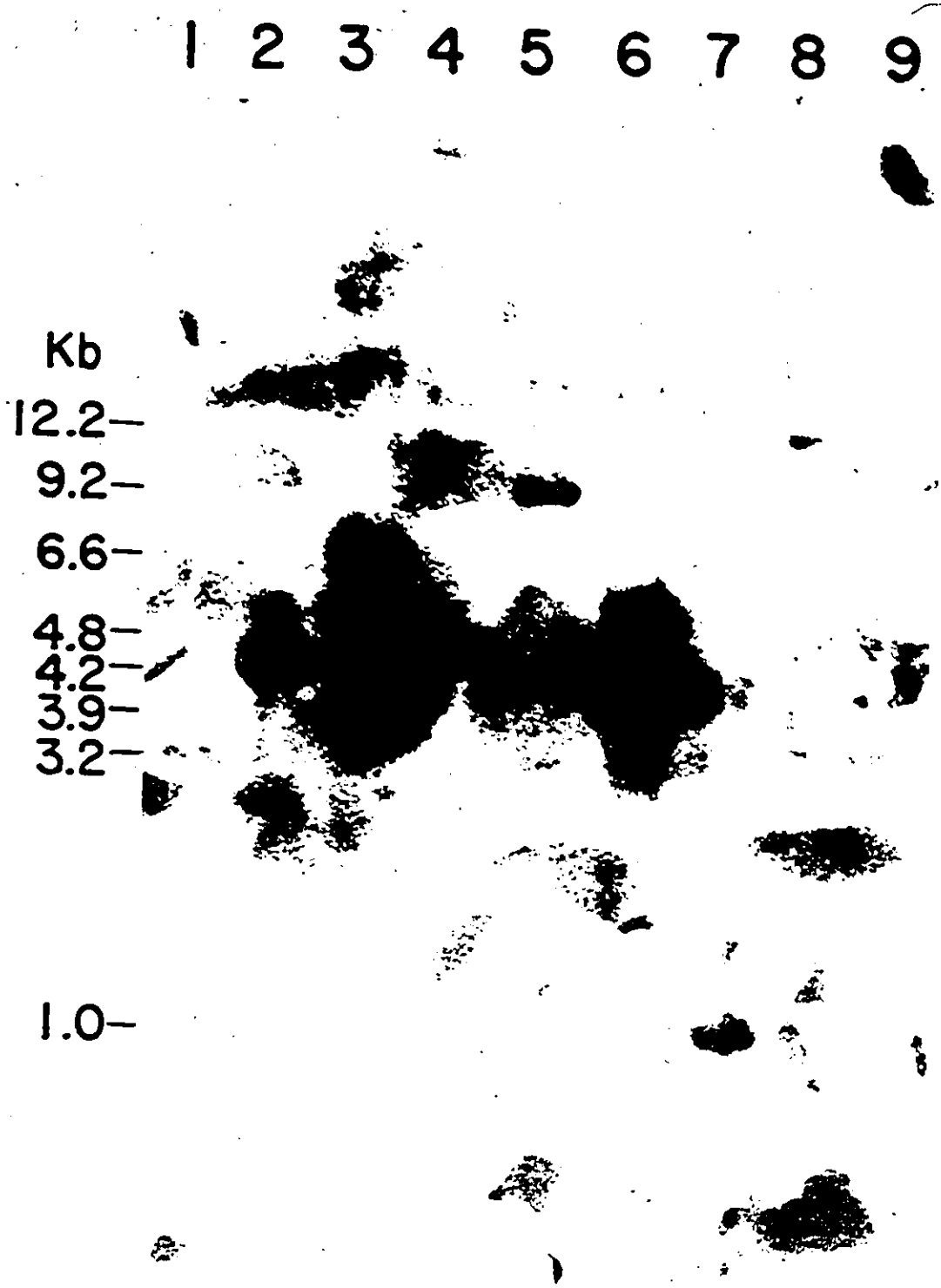
b) HindIII/EcoRI

20	20	20	20	20	20	20.5	20	20
							11.0	7.4
		5.5						5.7
		5.3	4.9					
4.8	4.8	4.8	4.8	4.8	4.8	4.8	4.8	4.8
4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5
	4.3	3.5			3.9			
3.3				3.3			3.8	
	2.9							
2.6	2.8		2.8	2.5	2.5	2.8		
2.3	2.1		2.0d	2.3	2.2			
2.2	2.0		1.8	1.8	1.9			
1.8			1.6	1.6		1.8		
1.2		1.4	1.4	1.2	1.3	1.0		
				1.0				

FIGURE 3.2

EcoRI/KpnI digested DNA, isolated from Series 100 lambda recombinant clones, were fractionated on a 0.8% agarose gel in TAE, blotted on a nylon filter. The Southern transfer was hybridized for 12 hours at 37°C in the presence of 40% formamide with pMPa21 nick-translated plasmid (2x10⁸ cpm/µg).

The figure shows the autoradiograph of the hybridized filter exposed for 24 hours in the presence of an intensifying screen. The sizes of the hybridizing bands are indicated: lane 1) 101, 2) 105, 3) 107, 4) 109, 5) 110, 6) 111, 7) 113, 8) 114.


1 2 3 4 5 6 7 8 9
Kb
12.2
9.2
6.6
4.8
4.2
3.9
3.2
1.0

At this point we had the choice of two strategies for isolating the α-amylase gene: we could re-screen the library using the murine amylase probe until we isolated overlapping clones, or we could find a probe which shared more homology with *Drosophila melanogaster* α-amylase. We chose to follow the second route. To this end, a cDNA (pUCa1) had been isolated in our laboratory which was synthesized from *Drosophila melanogaster* polyA+ mRNA and which displayed many characteristics of a Drosophila cDNA encoding α-amylase (M. Whitely and P. Moreau, unpublished data). The following observations led to the conclusion that pUCa1 encoded Drosophila α-amylase: 1) pUCa1 was isolated by virtue of expression, from an expression vector, of an enzyme capable of hydrolyzing starch 2) pUCa1 hybridized to a Drosophila RNA on Northern blots at approximately 1.6 Kb, the estimated size for the α-amylase mRNA 3) pMPa21, the murine α-amylase cDNA, hybridized to fragments generated from pUCa1 digested with various enzymes. Therefore, pUCa1 appeared to encode the *D. melanogaster* α-amylase cDNA. This was the homologous probe with which we would isolate the α-amylase gene of *Drosophila melanogaster*.

pUCa1 was first used to probe the DNA of the eight clones which had been isolated using the mouse pancreatic α-amylase cDNA to determine if any of these clones contained the α-amylase gene.

3.1.3 Series 100 probed with pUCa1

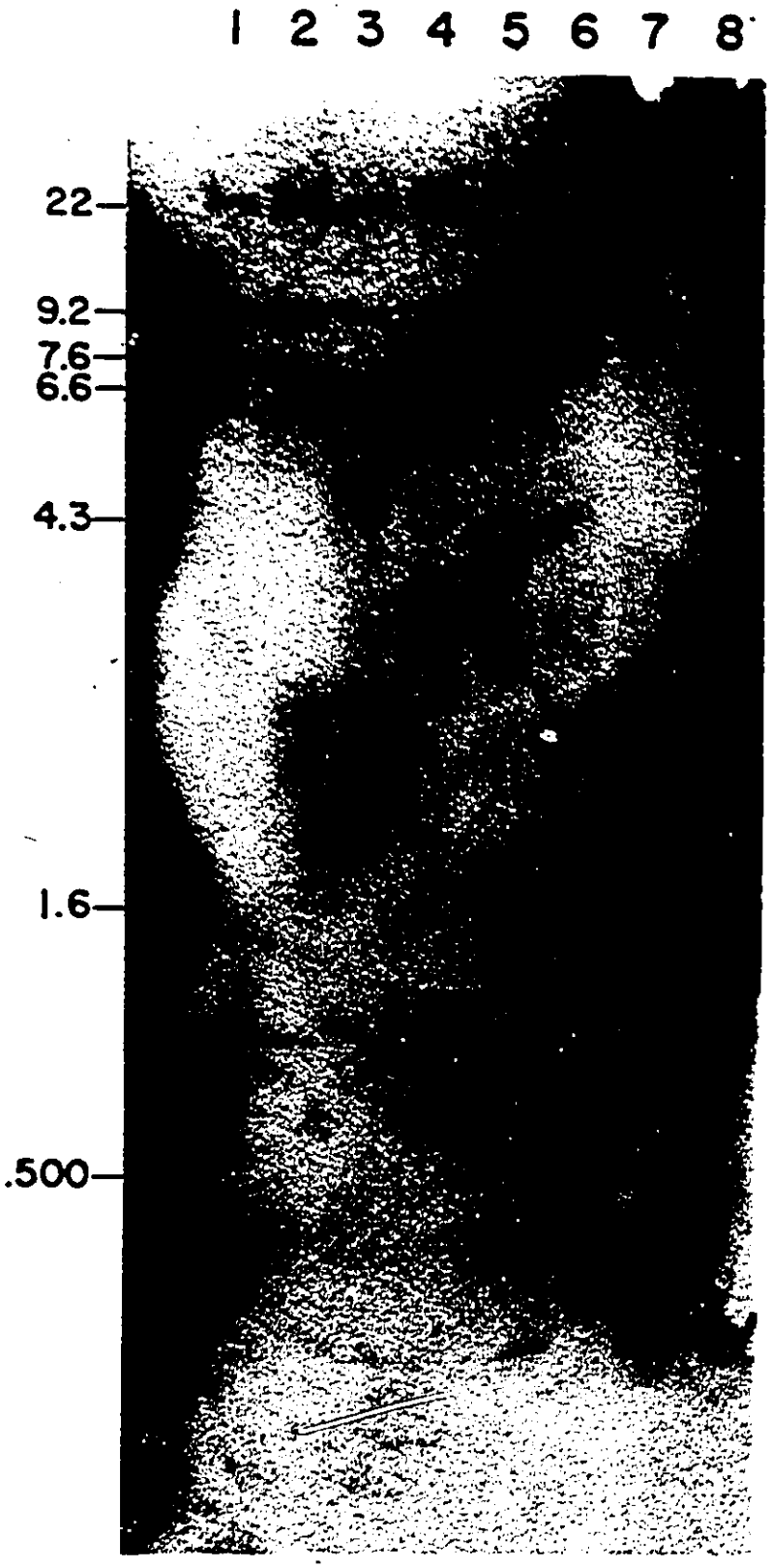
Hybridization of pUCa1 to the eight recombinant clones from Series 100 was performed under more stringent conditions (hybridization in 50% formamide at 42°C, post-hybridization washes twice in 0.2xSSC at 65°C for 30 minutes) because perfect homology was expected between a Drosophila cDNA and the transcribed regions of the corresponding Drosophila gene. In addition, more stringent conditions would eliminate clones isolated by virtue of homology to pMPa21 under low stringency conditions, but which are unrelated to α-amylase. Higher stringency requires the formation of longer duplexes, in other words, more homology.

Hybridization of EcoRI digested DNA from the eight clones of Series 100 to pUCa1 shows that none have homology with pUCa1 (Figure 3.3). Hybridization of pUCa1 to the lambda vector was evident after long exposure of the blot (48 hours) but not to the insert DNA. It was not obvious at this point why pUCa1 should hybridize to lambda, but as analysis of pUCa1 progressed this aspect of the probe was clarified and will be discussed later.

FIGURE 3.3

EcoRI digested DNA isolated from Series 100 lambda recombinant clones was fractionated on a 0.8% agarose gel in TAE, blotted on a nylon filter and hybridized for 16 hours at 42°C in the presence of 50% formamide with pUCa1 nick-translated plasmid (specific activity 5×10^7 cpm/µg).

The figure shows the autoradiograph of the hybridized filter exposed for 48 hours in the presence of an intensifying screen. The sizes of the Marker bands are indicated although the marker is not included on the figure: lane 1) 101, 2) 105, 3) 107, 4) 109, 5) 110, 6) 111, 7) 113, 8) 114.


1 2 3 4 5 6 7 8
22
9.2
7.6
6.6
4.3
1.6
.500

3.2 SERIES 200-1200

Because none of the clones isolated using the mouse amylase cDNA probe hybridized to pUCa1, the entire genomic library was rescreened. The next set of experiments involved hybridization of pUCa1 (or subclones derived from pUCa1) to the EMBL4 *Drosophila melanogaster* genomic library.

3.2.1 Probing with pUCa1--Series 200

The cDNA in pUCa1 had been cloned into the expression vector pUC8 which contains genes from the lactose operon of *E. coli*. The strain used to propagate the genomic library, Q358, also contains genes of the lactose operon. Therefore, it was not possible to probe with the entire pUCa1 plasmid as this would result in interfering background hybridization of the pUC8 vector to Q358 chromosomal DNA. For this reason, the cDNA was excised from pUCa1 by digestion with EcoRI which yields the 2.7 Kb pUC8 vector, and fragments from the cDNA of 3.2 Kb and 0.3 Kb. The 3.2 Kb fragment was extracted from a 1% agarose gel and nick translated incorporating α-^{32}P-dCTP. Nick translation typically produced probes with a specific activity of 10^7-10^9 cpm/μg.

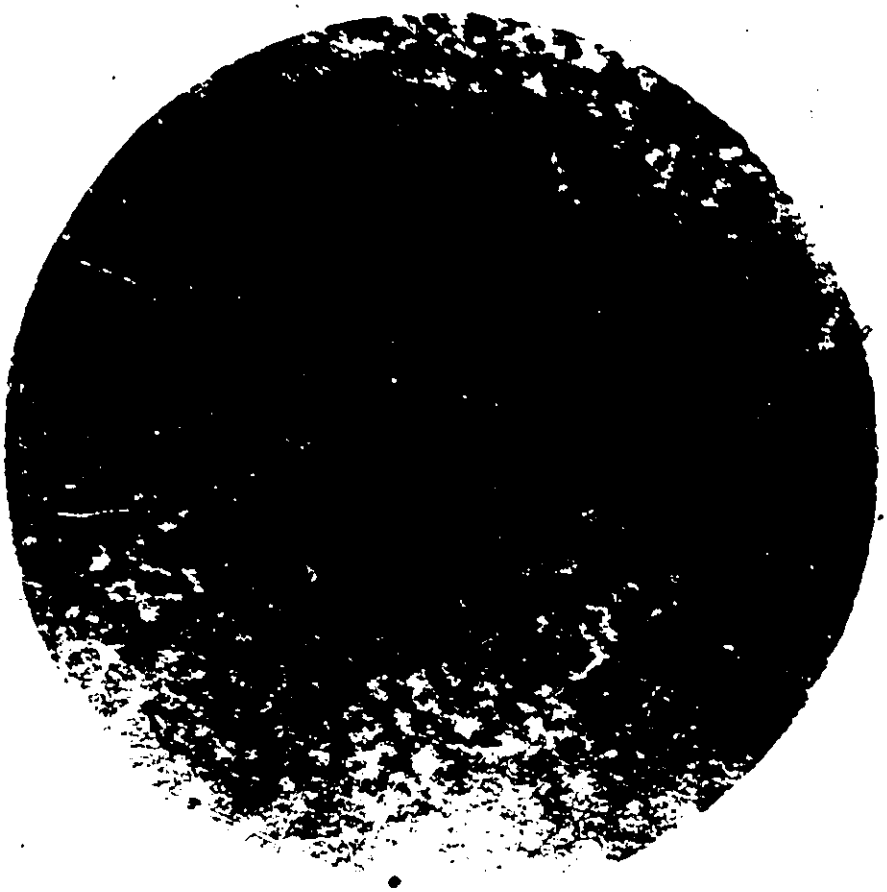
The stringency of hybridization conditions for screening the library was increased to 50% formamide at 42°C and the post-hybridization washes were done in 0.2xSSC at 60°C twice for 30 minutes (compare to low stringency conditions: 40% formamide at 37°C for hybridization and 2xSSC at 50°C for post-hybridization washes). The filters

were hybridized with the radiolabelled pUCa1 3.2 Kb fragment probe. Hybridized filters were subsequently exposed to X-ray film for 24 hours at -70°C with an intensifying screen. The resulting autoradiograph shows that the probe hybridized to all plaques and resulted in interfering background hybridization (Figure 3.4). There were approximately twelve plaques which gave hybridization signals above the background level and these were picked for subsequent screenings (Series 200).

The twelve plaques picked from the first screening of Series 200 were plated at lower density and screened again with the pUCa1 3.2 Kb fragment cDNA. On this second screening only four plates had plaques which gave higher than background hybridization signals and the signal was not as intense as those seen with the positives picked in Series 100. A total of six plaques were picked for a third screening. The hybridization signal from the third screening was again very weak. Nonetheless, DNA was isolated from small lysates of these plaques and probed with pUCa1 (the entire plasmid). Only the lambda vector hybridized the probe and this was obvious only after long exposure of the Southern transfer (data not shown). We concluded that these 'false positives' were picked due to hybridization of the probe to lambda DNA or to the host strain used to propagate the library. The level of background hybridization is expected to be equal for each

FIGURE 3.4

Autoradiograph of one plaque lift filter from the first screening of Series 200 probed with the 3.2 Kb EcoRI fragment from pUCa1 nick translated (specific activity 1x10⁹ cpm/µg). Each of 5 plates contained 1x10⁴ pfu/150mmx15mm petrie plate, plated on *E. coli* strain Q358. The plaques were replicated onto nylon filters and hybridized in the presence of 50% formamide at 42°C. The hybridized filters were exposed for 24 hours with an intensifying screen.



plaque, yet the plaques picked had a higher than background hybridization signal on the first screening. Therefore, on subsequent screenings of the library, putative positive plaques were compared to a negative control plaque.

3.2.2 Series 300

Series 300 was a rescreening of the library with pUCa1. The library was plated in a new host strain, *Escherichia coli* strain LE392, which has all genes of the lactose operon deleted. The change to a new plating strain was made to allow pUCa1, the entire plasmid, to be used as probe and to reduce background hybridization of vector sequences to the host strain, Q358, used until now to propagate the library.

On the first screening there was hybridization of all plaques to the probe, however, as before, some plaques gave a signal above this background signal. Plaques were picked from the first screening of Series 300, but the second screening of Series 300 clearly showed that none of these plaques was a true positive by comparison to a negative control.

Series 200 and 300 demonstrated that sequences present in pUCa1 hybridized to lambda DNA causing interfering background signal. Therefore, the cDNA from pUCa1 was cloned into pBR322 to alleviate any background hybridization of pUC8 vector sequences to the genomic library vector, EMBL4.

3.2.3 Probing with 3.5/pBR--Series 400

The 3.5 Kb cDNA insert of pUCa1 was cloned into the EcoRI site of pBR322 (3.5/pBR). This subclone was made to screen the library and alleviate any background which may be due to hybridization of the pUC8 vector. Series 400 was probed with 3.5/pBR and duplicate filters were probed with pBR322 as a control. Both probes were nick translated to a similar specific activity, however, the filters probed with 3.5/pBR gave very high uniformly distributed background hybridization signal whereas the duplicate filters probed with pBR322 gave only a low level of background signal.

It was becoming increasingly evident with each screening of the library that the pUCa1 cDNA itself was causing interfering background hybridization signal. The major part of the cDNA is 3.2 Kb long which is twice the required coding capacity for α-amylase mRNA (1.6 Kb). The cDNA must contain extraneous sequences unrelated to α-amylase which were causing background hybridization. Therefore, we believed that using a smaller probe derived from pUCa1 might eliminate these interfering sequences but would still retain coding sequences for α-amylase.

Subsequently, the 0.3 Kb EcoRI fragment from pUCa1 was sequenced (M. Whitely, personal communication) and was identified as a cDNA of Drosophila mitochondrial rRNA. This particular sequence occurs at high frequency in transcribed populations of *Drosophila melanogaster* genes. Thus, it is

not surprising that such an RNA species would be cloned inadvertantly because of its small size and relative abundance compared to other RNA sequences. We felt that this small cDNA was responsible for the high level of interfering background hybridization signal seen on previous screenings of the genomic library. The next set of screenings used smaller probes derived from pUCa1 which did not include the 300 bp rRNA.

3.2.4 Probing with OriI and OriII--Series 500-600

To facilitate the sequencing of pUCa1, the 3.2 Kb EcoRI fragment had been cloned into bacteriophage M13mp8 and partially deleted using the protocol of Dale, et al, (1983) (M. Whitely, personal communication).

The subclone designated OriI contained 3' sequences from the cDNA and its reverse complement, OriII contained 5' sequences from the cDNA. The 5' end was mapped to the EcoRI site distal to the 0.3 Kb small EcoRI fragment because transcription initiated from this end relative to the small fragment.

We used two of these deletion mutants to probe the genomic library. With some of the extraneous sequence deleted from the cDNA, it was hoped that the interfering background hybridization would disappear. The replicative form (double stranded) DNA of OriI and OriII deletion mutants was used to screen the library (Series 500 and 600, respectively). The background hybridization was much reduced over that seen with 3.5/pBR and several putative

positive plaques were picked for second screening. However, upon second screening, none of these plaques yielded a positive hybridization signal.

The library was rescreened with OriI and OriII and the stringency of post-hybridization washes was reduced to 1xSSC at 60°C, twice, for 30 minutes which allowed hybrids of minimum length 20 bp to be retained. This change was effected because it was possible that a weak positive signal from a first screening might be neglected and that this signal could be increased by reducing the washing stringency. For instance, if the probe hybridized to a part of the gene which contained an intron, the duplexes formed may not be stable at higher stringency. However, using OriI and OriII to probe the library under these new conditions did not yield any positive plaques.

3.2.5 Probing with pMPa21 and pUCa1--Series 700-800

Series 700 and 800 were probed with pMPa21 and pUCa1, respectively, under lower stringency conditions in strain LE392. These re-screenings did not yield any positive plaques beyond the second screening stage.

3.2.6 Probing with p5--Series 900

Sequencing data showed that one of the deletion mutants derived from pUCa1 contained at least one long open reading frame (M. Whitely, personal communication). This mutant, p5, was derived from OriI by further deletion from the 3' end of the insert. The insert in p5 is about 200 bp

long and is near the 3' end of the large EcoRI fragment of the cDNA.

Replicative form DNA from p5 was used to probe the library for Series 900. Screening with p5 did not yield positive clones beyond the second screening.

3.2.7 Probing with single-strand probes p5 and p2 Series 1000-1200

Because single strand probes have been found to give stronger hybridization signals and less interfering background (Meinkoth & Wahl, 1984), the next strategy for screening the library was to use single-strand DNA from p5 and its reverse complement p2 as probes. p5 is the coding strand from the open reading frame and p2 is the reverse complement of p5.

The positive strand DNA of each deletion mutant was labelled by primer extension from a 17-mer sequencing primer annealed at the 3' end of the multiple cloning site of M13mp8. Series 1000 was probed with p5 and 1100 and 1200 were probed with p2. Each of these series yielded putative positive signals on the first screening (Figure 3.5).

On a second screening of the putative positive plaques from Series 1200 (probed with p2), six plaques on one plate and one plaque on another plate gave positive hybridization signals (Figure 3.6). All of the positively hybridizing plaques were picked as well as a negative control from each plate. A third screening of the seven putative positive plaques confirmed that they contained sequences which

FIGURE 3.5

Autoradiograph of one plaque lift filter from the first screening of Series 1200. This filter was probed with p2 M13mp8 single strand DNA, primer extended from a 17-mer sequencing primer annealed at the 3' multiple cloning site (specific activity $1x10^8$ cpm/µg). Each of 5 plates contained $1x10^4$ pfu/150mmx15mm petrie plate, plated on *E. coli* strain LE392. The plaques were replicated onto nylon filters and hybridized in the presence of 50% formamide at 42°C. The hybridized filters were exposed for 24 hours with an intensifying screen.

Arrows indicate the two putative positive plaques picked for subsequent screening.

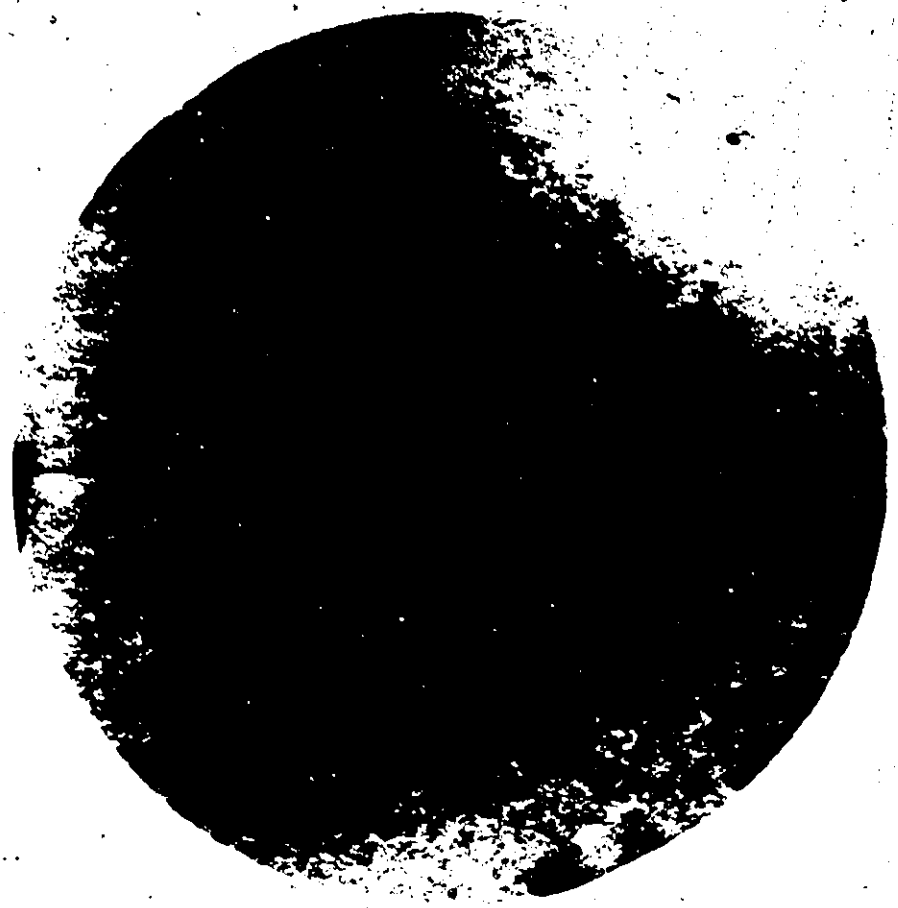
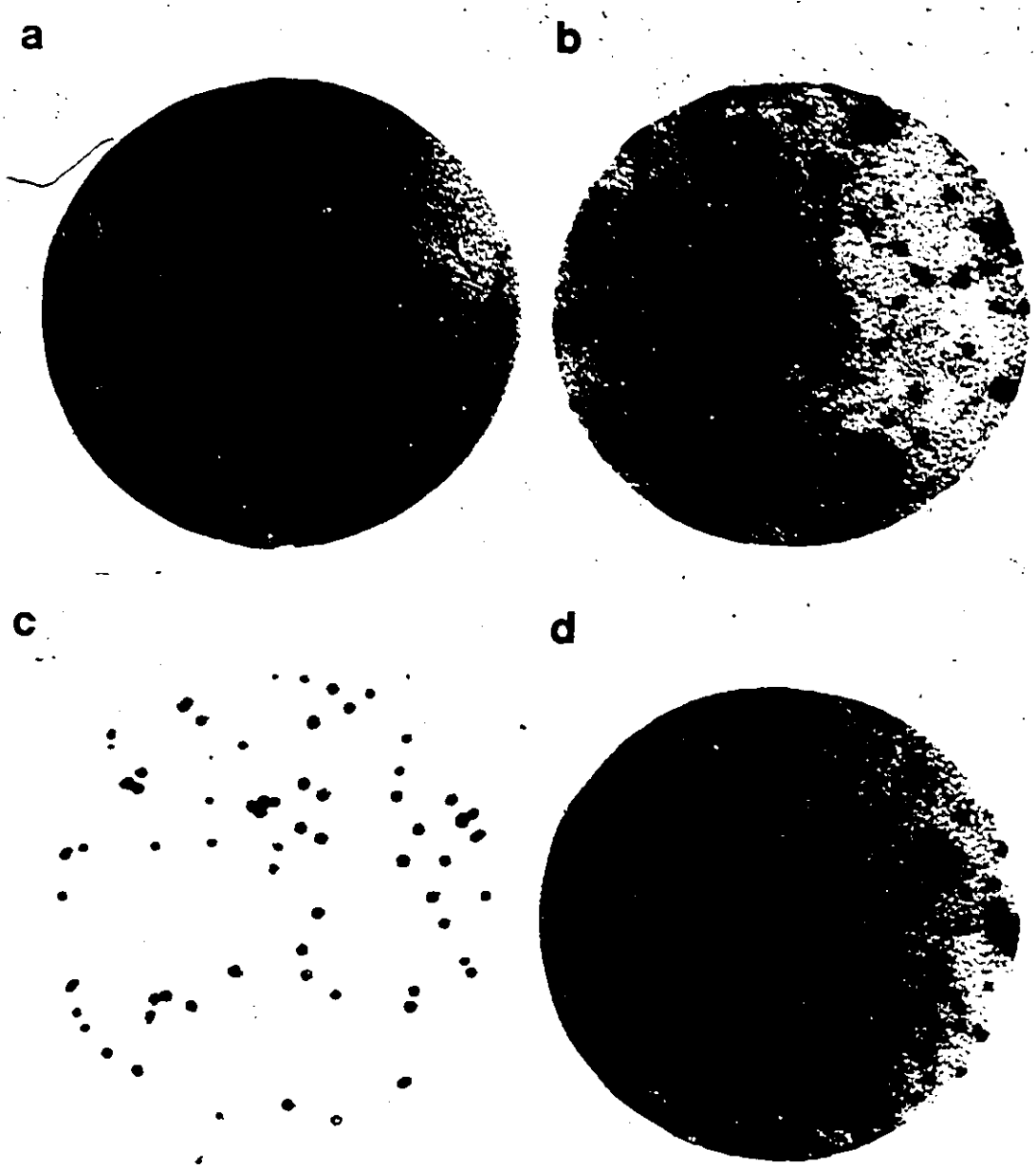


FIGURE 3.6

Second and third screening of putative positive plaques picked from first plating of Series 1200. Putative positives were replated in *E. coli* strain LE392 at a density of 50-100 pfu/100mmx15mm petrie plate. The filters were hybridized at 42°C in 50% formamide with primer extended p2 M13mp8 (specific activity $5x10^7$ cpm/µg).

Arrows indicate the positive plaques on the second screening. Panel a shows second screening and the six positive plaques later designated L1231 a-f; Panel b shows the second screening and the positive plaque later designated L1232; Panel c shows the third screening of L1231c; and Panel d shows the third screening of L1232.


a
b
c
d

hybridized the p2 probe whereas negative plaques did not (Figure 3.6). The six positive plaques picked from one plate on the second screening were expected to be identical having been derived from one plaque from the first screening. Therefore, there were two potentially independent positive plaques. These two plaques were designated L1231(a-f) and L1232. Both were re-screened with p5, the reverse complement of p2, and also gave a positive signal with this probe, indicating that the two recombinant clones are homologous to this particular sequence.

Finally, p5 and p2 were used to probe the library again but no new positive clones were isolated. A summary of the various screenings of the *Drosophila melanogaster* genomic library in EMBL4 is presented in Table 3.3.

The results which follow outline the characterization of L1231 and L1232 and various subclones derived from these original recombinants. Characterization eventually led to the conclusion that these two recombinants do not encode α-amylase. However, at this point in the research this fact was not evident and characterization proceeded as if the clones contained genomic sequences from *Drosophila melanogaster*. In fact, they do contain Drosophila sequences but they also contain a sequence closely related or identical to Insertion Sequence 1 (IS1) of *E. coli*. These clones were isolated because pUCa1 itself contains IS1.

TABLE 3.3

SUMMARY OF
SCREENINGS OF *DROSOPHILA MELANOGASTER*
GENOMIC LIBRARY

Series No.	Host Strain	Probe
100	Q358	pMPa21
200	Q358	pUCa1
300	LE392	pUCa1
400	LE392	3.5/pBR
500	LE392	OriI
600	LE392	OriII
700	LE392	pMPa21
800	LE392	pUCa1
900	LE392	p5
1000	LE392	ss-p5
1100	LE392	ss-p2
1200	LE392	ss-p2

3.3 CHARACTERIZATION OF L1231 AND L1232

3.3.1 Mapping L1231 and L1232

DNA from L1231(a-f) and L1232 was isolated and digested with EcoRI. As expected, the EcoRI digests from L1231(a-f) were identical, indicating that the first six clones contain the same insert DNA; the EcoRI restriction pattern of L1232 was different from the L1231 clones, suggesting that L1232 is an independent clone. The approximate size of insert is 15 Kb for both recombinant clones.

One of the identical L1231 clones and clone L1232 were digested with EcoRI (Figure 3.7 a,b, lanes 2 and 3,), HindIII (Figure 3.7 a,b, lanes 5 and 6) and XhoI (Figure 3.7 a,b, lanes 8 and 9) and were hybridized to p2. In the EcoRI digest the probe hybridized an 8.5 Kb fragment in L1231 and a 7.1 Kb fragment in L1232. No hybridization to EMBL4 DNA was seen (Figure 3.7 a,b, lanes 4, 7). Subsequently, the recombinant clones were digested with EcoRI/HindIII, EcoRI/XhoI, and HindIII/XhoI and probed with p2 (Figure 3.8 a,b; lanes 1 and 2, 3 and 4, 5 and 6, respectively). Table 3.4 lists the fragments generated by these digests and the hybridizing fragment in each case.

These digests narrowed the region of homology of the probe to the Drosophila DNA insert. The location of the hybridizing region for each clone was deduced by HindIII and XhoI digests (Figure 3.11).

TABLE 3.4

L1231 and L1232
DIGEST FRAGMENT SIZES
(Kb)

a) L1231

EcoRI	HindIII	XhoI	EcoRI/HindIII	EcoRI/XhoI	XhoI/HindIII
20.0	22.0	22.0	20.0	20.0	22.0
	11.3	13.0			
9.2				9.2	8.8
8.5		7.9			
			5.0		
	4.8		4.8		4.8
	4.5		4.5		4.5
4.0				4.0	
				3.2	
				2.7	
	2.4	2.2	2.4d		2.4
				2.2	2.1
1.4			1.4		
			0.8	1.4	
0.6	0.8		0.6	0.6	0.8

b) L1232

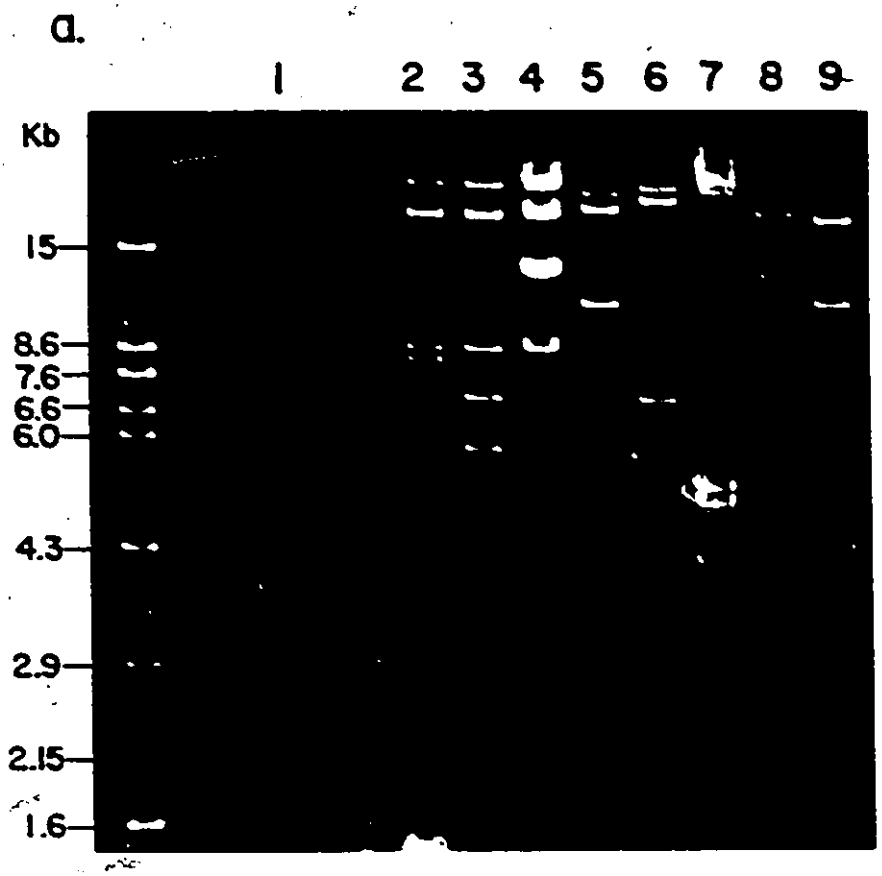
EcoRI	HindIII	XhoI	EcoRI/HindIII	EcoRI/XhoI	XhoI/HindIII
20.0	26.0	20.0	20.0	20.0	20.0
9.2		11.0		9.2	
7.1	7.0				6.4
5.8					
			4.8		
	4.5	5.2	4.5d		4.5
		3.3		3.3d	3.3
	3.2	2.6		2.5	2.8
	2.0				
		2.0	2.0	2.0	
1.9			1.9		
	1.8		1.8	1.8	1.7
	1.1				1.4
0.6			1.2d		1.2d
			0.5		

FIGURE 3.7

L1231 and L1232 (1µg) digested with EcoRI, HindIII and XhoI, was fractionated on a 0.8% agarose gel, blotted on nylon and hybridized at 42°C in 50% formamide with p2 primer extended single strand M13mp8 (specific activity $1x10^8$ cpm/µg).

Panel a shows the ethidium bromide stained gel with the marker bands (lane m) indicated. Panel b shows the autoradiograph exposed for 4 hours with an intensifying screen and the hybridizing fragment sizes indicated.

Lane 1) $1x10^{-3}$ µg pUCa1 digested with EcoRI, 2) L1231 EcoRI, 3) L1232 EcoRI, 4) EMBL4 EcoRI, 5) L1231 HindIII, 6) L1232 HindIII, 7) EMBL4 HindIII, 8) L1231 XhoI, 9) L1232 XhoI


a.
1 2 3 4 5 6 7 8 9
Kb
15
8.6
7.6
6.6
6.0
4.3
2.9
2.15
1.6

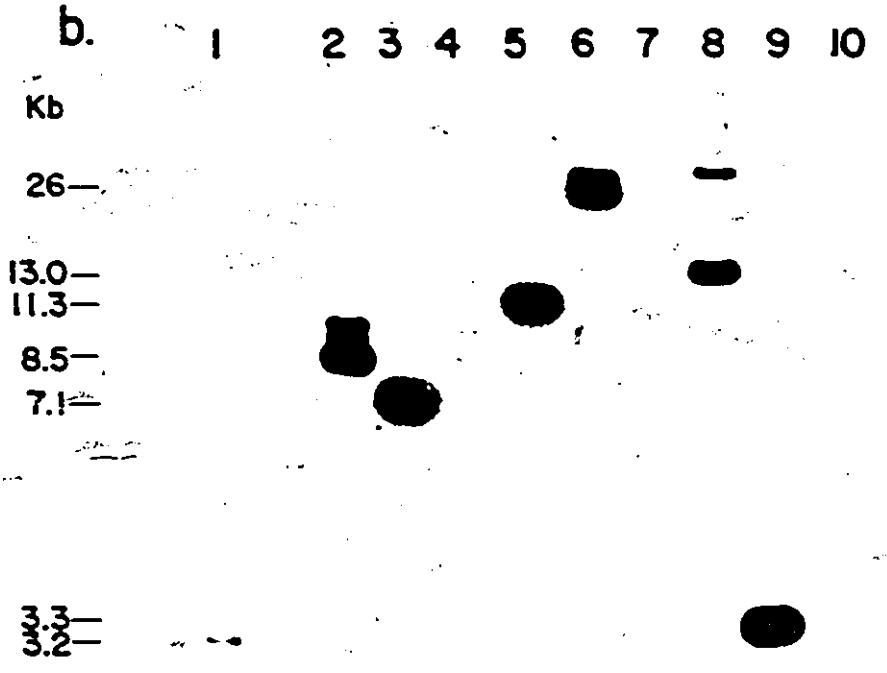

b.
1 2 3 4 5 6 7 8 9 10
Kb
26
13.0
11.3
8.5
7.1
3.3
3.2

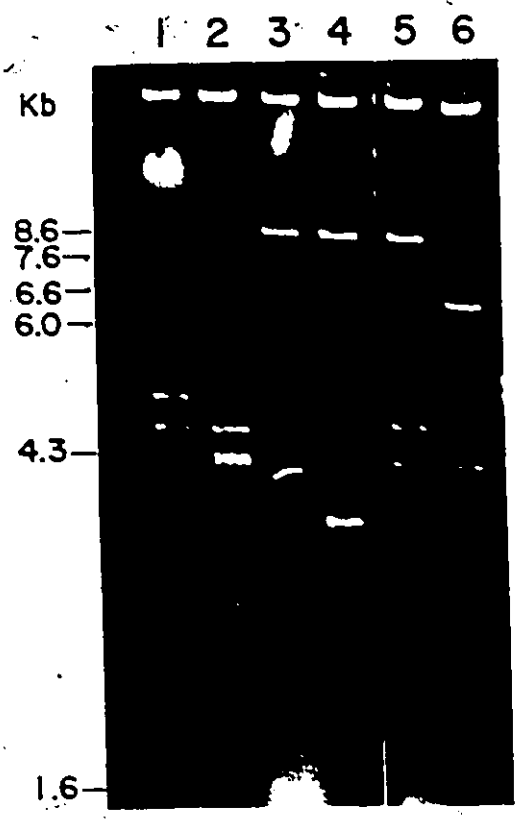
FIGURE 3.8

L1231 and L1232 (1µg) digested with EcoRI/HindIII, EcoRI/XhoI and HindIII/XhoI, fractionated on a 0.8% agarose gel, blotted on nylon and hybridized at 42°C in 50% formamide with p2 primer extended single strand M13mp8 (specific activity $5x10^7$ cpm/µg).

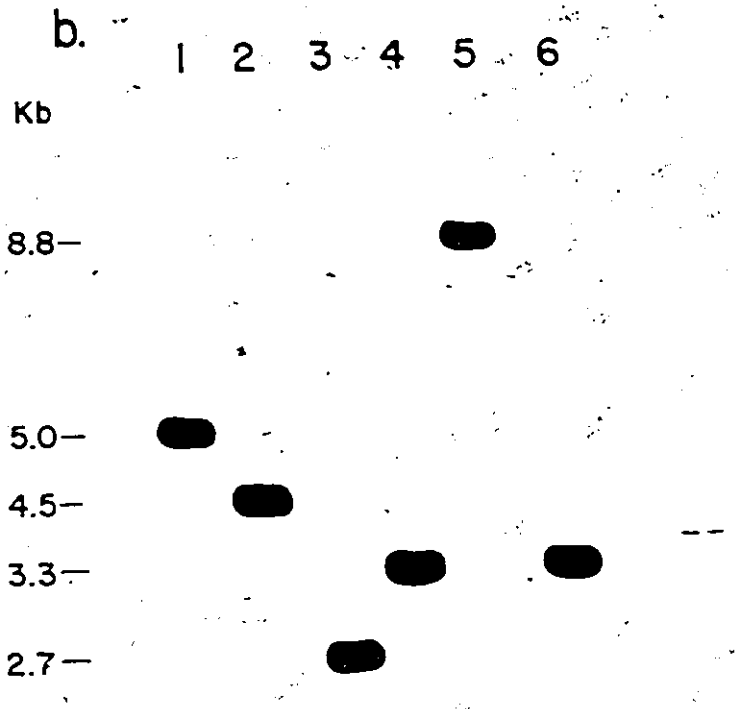
Panel a shows the ethidium bromide stained gel with the marker bands (lane m) indicated although the marker is not included on the figure. Panel b shows the autoradiograph exposed for 4 hours with an intensifying screen and the hybridizing fragment sizes indicated.

Lane 1) L1231 EcoRI/HindIII, 2) L1232 EcoRI/HindIII, 3) L1231 EcoRI/XhoI, 4) L1232 EcoRI/XhoI, 5) L1231 HindIII/XhoI, 6) L1232 HindIII/XhoI.

a.



b.



EMBL 4 is not cleaved by XhoI, therefore, the vector 'arms' must be included in two fragments which also contain insert DNA cleaved at internal XhoI sites. The largest fragments generated by XhoI from L1231 are 22 Kb and 13 Kb; the 22 Kb fragment includes the left arm and the 13 Kb fragment includes the 9.2 Kb right arm. Since the 13 Kb fragment also hybridizes the probe, 3.8 Kb of DNA, some of which is homologous to the probe, is from insert DNA adjacent to the right arm of the lambda vector.

For L1232, the XhoI hybridizing fragment is 3.3 Kb long. Thus, in L1232, the hybridizing fragment lies internal to the vector arms and is completely excised by XhoI.

The left arm of the vector has no HindIII sites and therefore must generate a fragment larger than 20 Kb. The right arm is cleaved once by HindIII, generating a 4.5 Kb fragment from the extreme right end of lambda. The other 4.8 Kb of lambda must be included in a fragment which also contains insert DNA.

In L1231 the probe hybridizes an 11 Kb HindIII fragment which contains the remaining 4.8 Kb of the right lambda arm. The HindIII and XhoI digest data suggest that the hybridizing fragment in L1231 is adjacent to the right arm of the vector. Thus, the hybridizing EcoRI fragment of L1231 would be generated by an EcoRI site at least 8.5 Kb from the vector EcoRI site in the right arm.

The largest fragment from the HindIII digest of L1232 hybridizes the probe. This suggests that the hybridizing region of L1232 lies close to the left arm. However, the XhoI digest precludes the hybridizing fragment being directly apposed to the vector since XhoI excises a 3.3 Kb hybridizing region.

Digests with pairs of the enzymes EcoRI, XhoI and HindIII served to refine the map further (Figure 3.11). From L1231, EcoRI/HindIII yields a 5.0 Kb hybridizing fragment by cleaving the 4.8 Kb of the lambda vector from the 11 Kb HindIII fragment. However, addition of these two fragments leaves approximately 1 Kb unaccounted for which cannot be generated by HindIII. Therefore, it must be generated by EcoRI from the vector cloning site and another site 1 Kb within the insert. There are two fragments, approximately 0.6 and 1.4 Kb long, generated by EcoRI single and double digest with HindIII. EcoRI/XhoI gives a 2.7 Kb hybridizing fragment, cleaving the 9.2 Kb lambda right arm from the 13.0 Kb XhoI fragment. Again, 1 Kb cannot be accounted for by addition of these fragments and must be generated by EcoRI sites from the vector and the insert. XhoI/HindIII, gives an 8.8 Kb fragment which hybridizes the probe and which includes 4.8 Kb of lambda DNA from the vector HindIII site. Therefore, XhoI cleaves 2.1 Kb from the HindIII hybridizing fragment.

From L1232, EcoRI/HindIII generates a 4.5 Kb hybridizing fragment co-migrating with the 4.5 Kb fragment

generated by HindIII from lambda. The EcoRI site of the vector cleaves the 20 Kb left lambda arm from the hybridizing insert fragment generated by HindIII. EcoRI/XhoI and XhoI/HindIII digests both yield a 3.3 Kb hybridizing fragment which is the same size as the hybridizing fragment from the XhoI single digest. Therefore, the XhoI sites flanking the hybridizing region are internal to the EcoRI and HindIII generated fragments.

Further mapping of insert fragments was only possible if the vector DNA remained more or less intact. Table 3.5 gives the sizes of the fragments for BamHI and Asp718 digests of the two clones (Figure 3.9, 3.10). From L1231, BamHI generates a 10.8 Kb hybridizing fragment (Figure 3.9a, lane 10; 3.9b, lane 5); one BamHI site was regenerated at the cloning site internal and adjacent to the vector EcoRI site and the other is 11 Kb from this site. Asp718 generates a 20 Kb hybridizing fragment which includes the right arm of the vector and extends close to the most internal BamHI site (Figure 3.9a, lane 7; 3.9b, lane 2).

For L1232, Asp718 gives an 11 Kb hybridizing fragment which includes sequences (approximately 1 Kb) from the left arm of the vector (Figure 3.10a, lane 7; 3.10b, lane 1). BamHI generates a 23 Kb hybridizing fragment which includes 20 Kb from the left arm of the vector (Figure 3.10a, lane 10; 3.10b, lane 4); BamHI/EcoRI double digest yields a 3.1 Kb hybridizing fragment (Figure 3.10a, lane 11; 3.10b, lane 5). Additional digests were made choosing enzymes on

TABLE 3.5

L1231 and L1232
DIGEST FRAGMENT SIZES
(Kb)

a)

L1231			
Asp718	Asp718/EcoRI	BamHI	BamHI/EcoRI
21	18.0	21	20
17.5		10.8	
	9.2	9.2	9.2
7.0	8.5		8.5
	3.9		
		3.1	3.1
1.3	1.3d	1.25	1.2
	0.6		1.1

b)

L1232			
17	17	23	20
14			
11.0			
	9.2	9.2	9.2
	7.1	8.6	
			6.2
	3.2	3.2	3.2
			3.1
	1.8	1.2	1.8
1.6	1.6		
			0.9

FIGURE 3.9

L1231 (1µg) was digested with various enzymes, fractionated on 0.8% agarose gel and transferred to nylon. The Southern blot was hybridized to M13mp8_p2 single strand DNA primer extended from a 17-mer sequencing primer (specific activity 1x10⁹ cpm/µg). Hybridization was in 50% formamide at 42°C for 16 hours and the filter was subsequently exposed for 4 hours with an intensifying screen.

Panel a shows the ethidium bromide stained gel and the marker fragment sizes (lane m) are indicated. Lane 1) EcoRI, 2) HindIII, 3) EcoRI/HindIII, 4) XhoI, 5) XhoI/EcoRI, 6) XhoI/HindIII, 7) Asp718, 8) Asp718/EcoRI, 9) BglII, 10) BamHI, 11) BamHI/EcoRI, 12) BamHI/XhoI

Panel b shows the autoradiograph of some of the digests in Panel a hybridized to the probe. Lane 1) XhoI/HindIII, 2) Asp718, 3) Asp718/EcoRI, 4) BglII, 5) BamHI, 6) BamHI/EcoRI, 7) BamHI/XhoI

a.


m 1 2 3 4 5 6 7 8 9 10 11 12
Kb
15.0
8.6
7.6
6.6
6.0
4.3
2.9
2.1
1.6
.5

b.

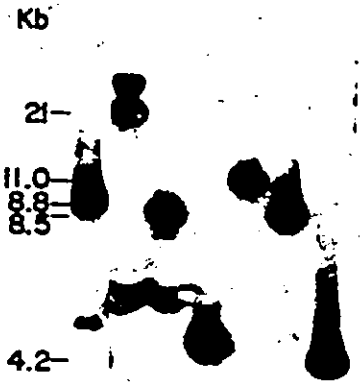

1 2 3 4 5 6 7
Kb
21
11.0
8.8
8.5
4.2

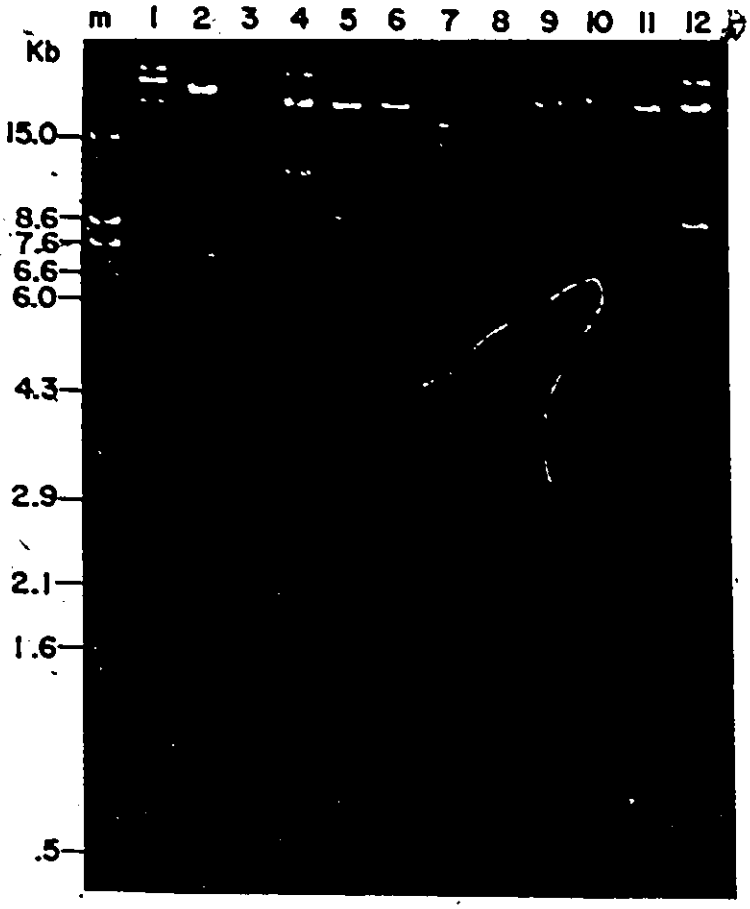
FIGURE 3.10

L1232 (1µg) was digested with various enzymes, fractionated on 0.8% agarose gel and transferred to nylon. The Southern blot was hybridized to M13mp8 p2 single strand DNA, primer extended from a 17-mer sequencing primer (specific activity 1x10⁹). Hybridization was in 50% formamide at 42°C for 16 hours and the filter was subsequently exposed for 4 hours with an intensifying screen.

Panel a shows the ethidium bromide stained gel and the marker fragment sizes (lane m) are indicated. Lane 1) EcoRI, 2) HindIII, 3) EcoRI/HindIII, 4) XhoI, 5) XhoI/EcoRI, 6) XhoI/HindIII, 7) Asp718 , 8) Asp718/EcoRI, 9) BglII, 10) BamHI, 11) BamHI/EcoRI, 12) BamHI/XhoI

Panel b shows the autoradiograph of some of the digests in Panel a hybridized to the probe. The sizes of the hybridizing fragments are indicated. Lane 1) Asp718, 2) Asp718/EcoRI, 3) BglII, 4) BamHI, 5) BamHI/EcoRI, 6) BamHI/XhoI

a.


m 1 2 3 4 5 6 7 8 9 10 11 12
Kb
15.0
8.6
7.6
6.6
6.0
4.3
2.9
2.1
1.6
.5

b.

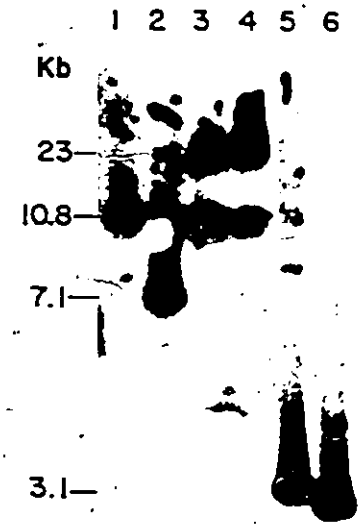
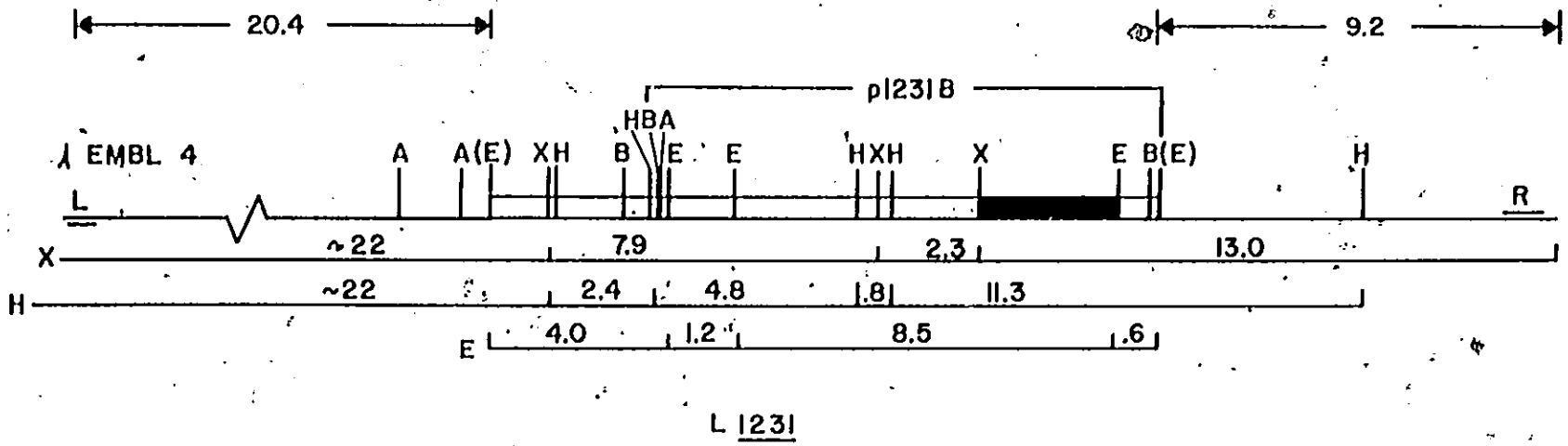
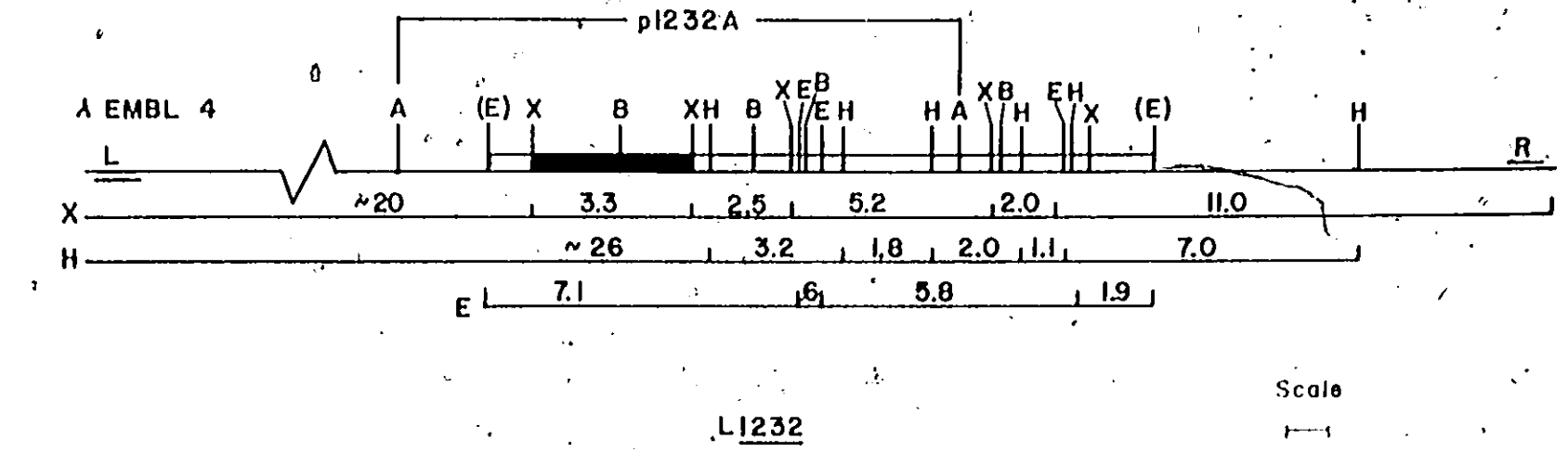

1 2 3 4 5 6
Kb
23
10.8
7.1
3.1

FIGURE 3.11

Restriction enzyme site map of L1231 (upper) and L1232 (lower)showing the EcoRI (E), HindIII (H), XhoI (X), Asp718 (A), and BamHI (B) sites on each recombinant clone. The organization and sizes of the EcoRI, HindIII, and XhoI fragments are indicated beneath each map. The fragments subcloned into p1231B (11 Kb BamHI fragment) and p1232A (10.8 Kb Asp718 fragment) are indicated above each map. The shaded area represents the smallest hybridizing restriction fragment for each clone. The EcoRI sites shown in brackets are the cloning sites in lambda.



L 1231



L1232

Scale

1 kb

the basis of frequency and site of cleavage in the EMBL4 DNA. Figure 3.11 summarizes mapping of the two lambda clones. The two clones were also digested with the following enzymes: SmaI, SmaI/EcoRI, BglII, PvuI, SalI, XbaI, XbaI/EcoRI, SstI, SstI/EcoRI, NcoI (data not shown).

3.3.2 Subclones p1231B and p1232A

The task of mapping the EcoRI, HindIII and XhoI sites in the inserts of L1231 and L1232 was not facilitated by the large size of the lambda vector. This was especially apparent in cases where one of the lambda arms remained attached to insert DNA. Resulting size estimates were highly inaccurate due to the high molecular weight of the DNA. In order to locate these sites within the insert DNA it was necessary to subclone some of the insert into a smaller more manageable vector. Therefore, the 11 Kb hybridizing BamHI fragment of L1231 and the 11 Kb hybridizing Asp718 fragment of L1232 were subcloned into compatible sites in the multiple cloning site of the vector pUC18. These two large fragments were chosen for subcloning in order to include the hybridizing region and to preserve DNA flanking the hybridizing region. The resulting subclones were called p1231B and p1232A, respectively.

The first digest of the subclones with EcoRI confirmed that the same (8.5 Kb and 7.1 Kb) hybridizing fragments were retained (Lane 1, Figure 3.12 a and b, respectively).

FIGURE 3.12

p1231B and p1232A (2µg) digested with various enzymes, fractionated on a 0.8% agarose gel, and hybridized with primer extended single strand M13mp8 p2 DNA (specific activity $3x10^8$ cpm/µg) in 50% formamide at 42°C for 16 hours.

The figure shows the autoradiographs of digested p1231B (Panel a) Lane 1) EcoRI, 2) EcoRI/BamHI, 3) BamHI, 4) HindIII, 5) XhoI, 6) Asp718. p1232A (Panel b) Lane 1) EcoRI, 2) EcoRI/Asp718, 3) Asp718, 4) HindIII, 5) XhoI, 6) BamHI. Both autoradiographs were exposed 4 hours with an intensifying screen. The size of the hybridizing fragment from each digest is indicated.

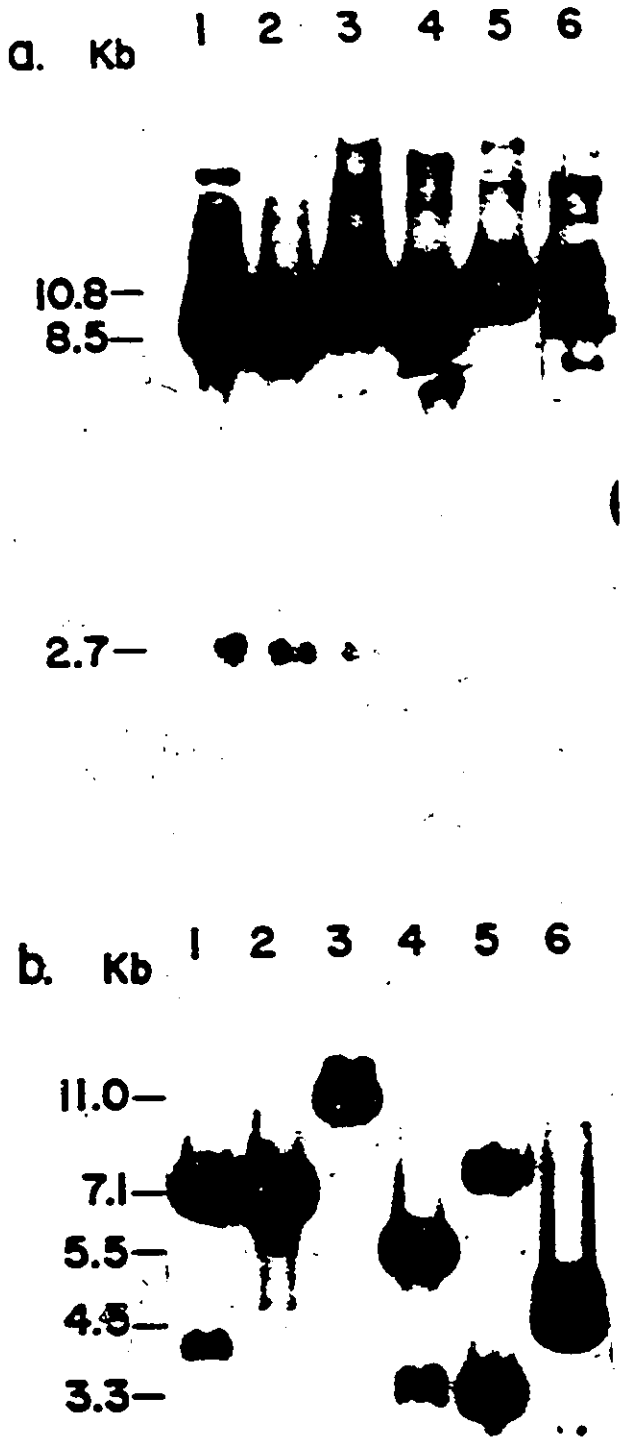

a. Kb
1 2 3 4 5 6
10.8—
8.5—
2.7—
b. Kb
1 2 3 4 5 6
11.0—
7.1—
5.5—
4.5—
3.3—

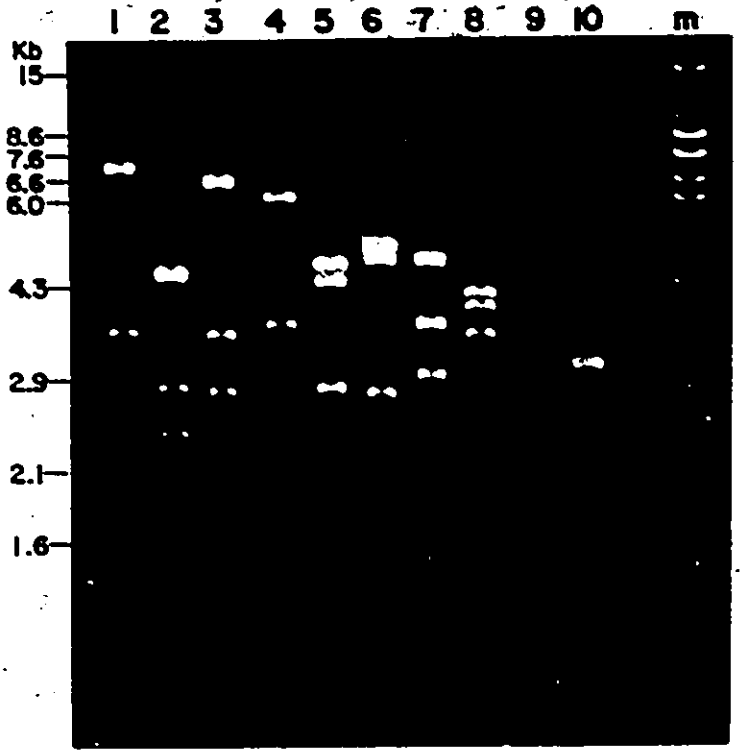
FIGURE 3.13

p1231B (1µg) was digested with various enzymes, fractionated on a 0.8% agarose gel, and hybridized with primer extended single strand M13mp8 p2 DNA (specific activity 4x10^7 cpm/µg) in 50% formamide at 42°C for 16 hours.

Panel a shows the ethidium bromide stained gel and the marker fragment sizes (lane m) are indicated. Panel b is the autoradiograph exposed for 4 hours with an intensifying screen and the sizes of the hybridizing fragments are indicated.

Lane 1) XhoI/HindIII, 2) BamHI/XhoI, 3) BamHI/HindIII, 4) PstI, 5) BglII/EcoRI, 6) BglII/BamHI, 7) BglII/ HindIII, 8) BglII/XhoI, 9) Sau3AI, 10) RsaI.

a.


1 2 3 4 5 6 7 8 9 10 m
Kb
15
8.6
7.8
6.6
6.0
4.3
2.9
2.1
1.6

b.

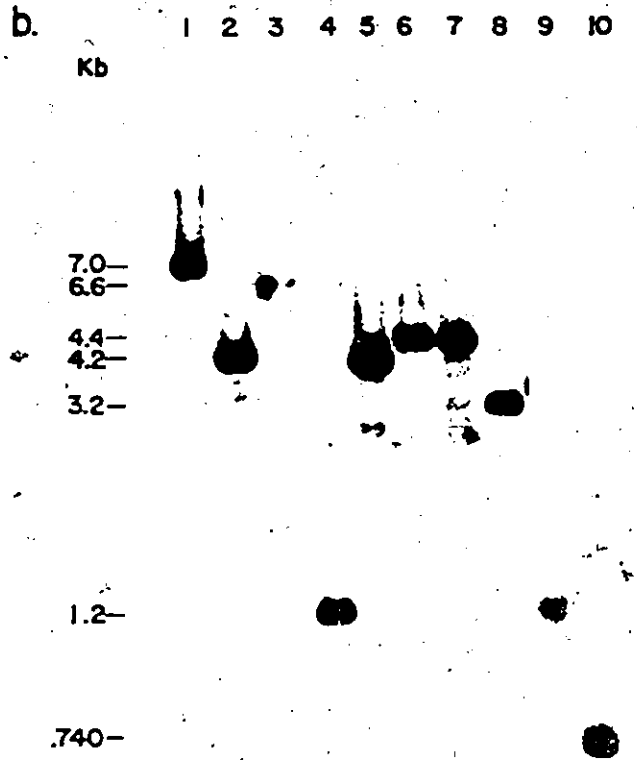
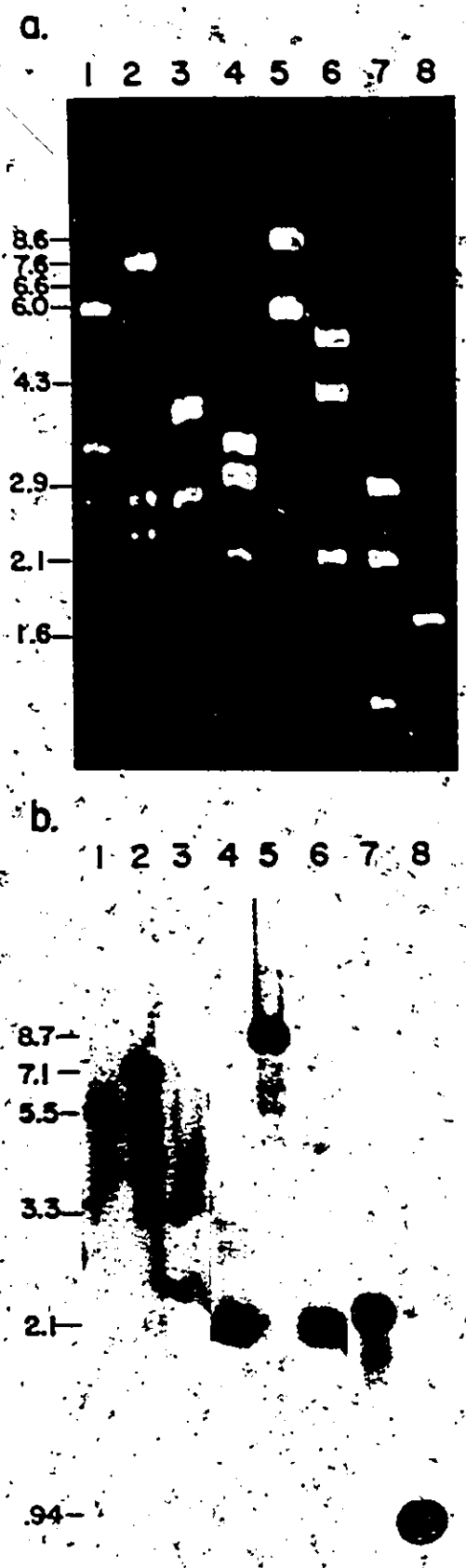

1 2 3 4 5 6 7 8 9 10
Kb
7.0
6.6
4.4
4.2
3.2
1.2
.740

FIGURE 3.14

p1232A (1µg) was digested with various enzymes and fractionated on a 0.8% agarose gel, and hybridized with primer extended single strand M13mp8 p2 DNA (specific activity $4x10^7$ cpm/µg) in 50% formamide at 42°C for 16 hours.

Panel a shows the ethidium bromide stained gel and the marker fragment sizes are indicated although the marker is not included on the figure. Panel b is the autoradiograph exposed for 4 hours with an intensifying screen and the sizes of the hybridizing fragments are indicated.

Lane 1) Asp718/HindIII, 2) Asp718/EcoRI, 3) Asp718/XhoI, 4) PstI, 5) BglII, 6) BglII/EcoRI, 7) Sau3AI, 8) RsaI


a.
1 2 3 4 5 6 7 8
8.6
7.6
6.6
6.0
4.3
2.9
2.1
1.6
b.
1 2 3 4 5 6 7 8
8.7
7.1
5.5
3.3
2.1
.94

As well the subclones were digested with HindIII, XhoI, Asp718, and BamHI (Figure 3.12 a,b). Subsequently, p1231B was digested with BamHI in combination with EcoRI, HindIII, and XhoI (Figure 3.13), and p1232A was digested with Asp718 in combination with EcoRI, HindIII, and XhoI (Figure 3.14).

Table 3.6 lists the fragments generated by these digests and the hybridizing fragment in each case. These digests served firstly, to orient the subcloned insert and secondly, to place with finality, the XhoI, EcoRI, and HindIII sites shown on the map of the lambda clones (Figure 3.11) and on the map of the two subclones (Figure 3.15).

Other restriction enzyme digests were made whose sites are not shown because they do not add pertinent information to the map. The following enzymes were used to digest p1231B and p1232A: AccI, BglI, BglI/BamHI, BglII, BglII/BamHI, BglII/EcoRI, BglII/HindIII; BglII/XhoI, EcoRI/PvuI, HindII, NcoI, PstI, PvuI, PvuII, RsaI, SalI, Sau3AI, SmaI, SphI, SstI, XbaI, XhoI/PvuI, XmaI/BamHI (data not shown).

Important information was obtained from digests with PstI, RsaI, and Sau3AI (Figures 3.13 and 3.14) which yield numerous fragments from each subclone. Only some of the PstI sites are shown in Figure 3.15. PstI was later used to subclone the 1.3 Kb hybridizing fragment from p1231B (Figure 3.13, lane 4) and the 2.1 Kb hybridizing fragment from p1232A (Figure 3.14, lane 4) into pBR322. RsaI and Sau3AI digests yield many restriction fragments from the

TABLE 3.6

p1231B and p1232A
DIGEST FRAGMENT SIZES
(Kb)

a) p1231B

EcoRI	HindIII	XhoI	EcoRI/ HindIII	EcoRI/ XhoI	XhoI/ HindIII	EcoRI/ BamHI	BamHI/ HindIII	BamHI/ XhoI
8.5	8.5	10.9				8.5		
					7.0		6.6	
			5.0		3.3			4.2
	3.3			3.2			3.3	
2.7			2.7	2.8		2.7	2.7	2.7
			2.5	2.7				
		2.3		2.3	2.2			2.3
1.4			1.4	1.4		1.4		
	0.9		0.9		0.9		0.8	
0.6			0.6	0.6		0.6		

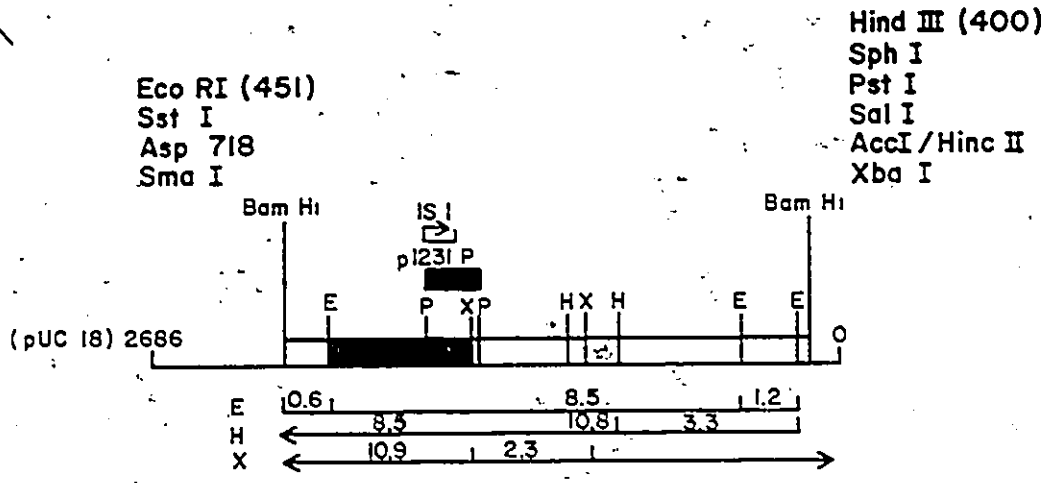
b) p1232A

EcoRI	HindIII	XhoI	EcoRI/ HindIII	EcoRI/ XhoI	BamHI	Asp718/ EcoRI	Asp718/ HindIII	Asp718/ XhoI
7.1		7.6				7.1		
	5.5				5.2		5.5	
			4.5	4.3	4.3			
4.3								3.5
	3.2	3.3	—	3.3	3.5			3.3
	3.0						3.0	
			2.7	2.5		2.7	2.7	2.7
2.4		2.5	2.1	2.4		2.3		2.5
	1.8		1.8				1.8	1.8
			1.5					
			1.3		1.2	1.4		
			0.9	0.7			0.9	
0.6			0.6	0.6		0.6		

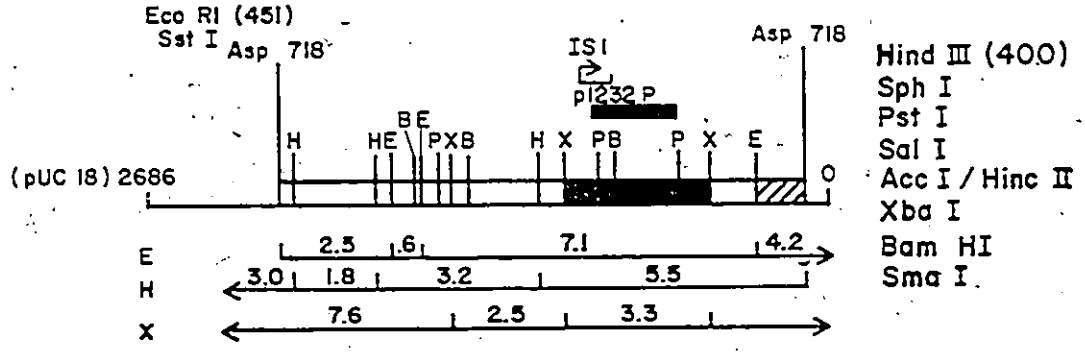
FIGURE 3.15

Restriction enzyme site map of p1231B (upper) and p1232A (lower) showing the EcoRI (E), HindIII (H), XhoI (X), Asp718 (A), and BamHI (B) sites on each recombinant clone. As well, the pertinent PstI sites (P) are indicated. The organization and sizes of the EcoRI, HindIII, and XhoI fragments are indicated beneath each map.

The shaded area corresponds to the shaded area in Figure 3.11. The IS1 and its orientation are indicated above the map as is the PstI fragment subcloned into p1231P and p1232P. The sites in the multiple cloning site on either side of the inserts are indicated for reference.



p1231 B



p1232 A

Scale
1 kb

subclones yet in each digest only one small fragment from each subclone hybridized the probe (Figures 3.13 and 3.14). The small size of these hybridizing fragments pointed out that the hybridizing region of both subclones is not larger than approximately 1 Kb (Figure 3.13b, lanes 9 and 10, respectively, for p1231B; Figure 3.14b, lanes 7 and 8, respectively, for p1232A). This observation was very disturbing because the coding region for the α-amylase gene must be at least 1.6 Kb (without introns) to encode the 54,000 dalton protein. One explanation of this observation would be that the gene present in the lambda clones may have been truncated by the Sau3AI digest used to construct the Oregon R library. However, as determined later, this was not the case.

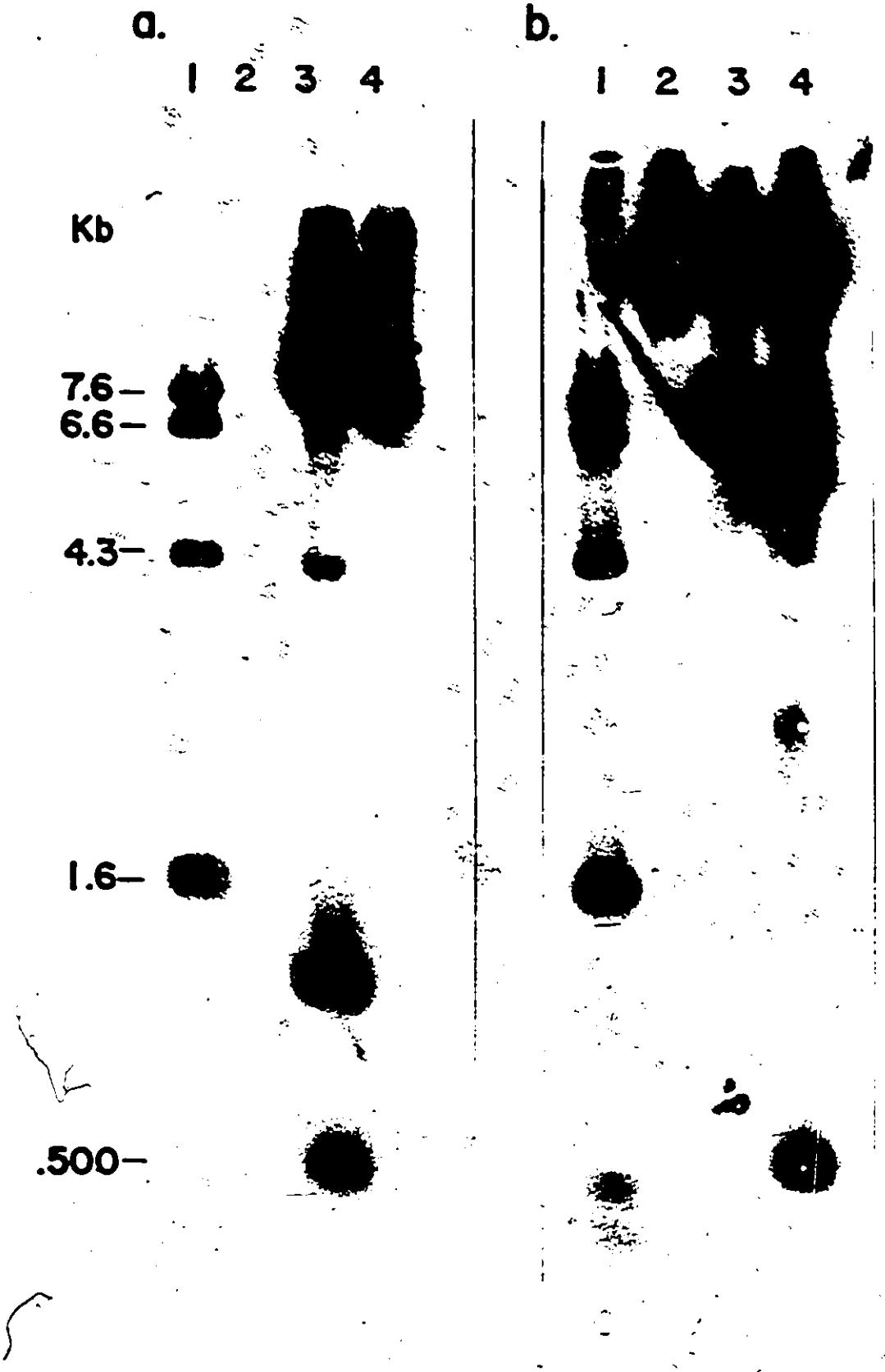
3.3.3 Cross-hybridization of p1231B and p1232A

Because the two original lambda clones were isolated using the same probe, it was probable that they shared common sequences. To test this possibility, L1231 and L1232, the original lambda clones, were digested with EcoRI and each was probed with the subclone of the other phage. Therefore, p1231B was used to probe L1232 and p1232A was used to probe L1231. Figure 3.16 demonstrates that p1231B hybridizes to the 7.1 Kb band generated by EcoRI in L1232 (Figure 3.16a, lane 4). Conversely, p1232A hybridizes to the 8.5 Kb band of EcoRI digested L1231 (Figure 3.16b, lane 4). In addition, p1232A hybridizes the left arm of the lambda vector. This was expected because one of the Asp718

FIGURE 3.16

This figure shows results of crosshybridization of p1231B to L1232 (Panel a) and p1232A to L1231 (Panel b). The two plasmids (p1231B and p1232A) were nick translated to a specific activity of $3x10^8$ cpm/µg and were hybridized to the EcoRI digested lambda recombinant clones (L1231 and L1232) immobilized on nylon filters. Hybridization was done in 50% formamide at 42°C for 16 hours. Exposure of the autoradiographs shown in the figure was for 4 hours with an intensifying screen.

The marker fragments are indicated (lane 1). Lane 2) EMBL4 digested with EcoRI (1µg), 3) L1231 (1µg) digested with EcoRI, 4) L1232 (1µg) digested with EcoRI.


a.
b.
1 2 3 4
1 2 3 4
Kb
7.6–
6.6–
4.3–
1.6–
.500–

sites used to subclone the L1232 fragment into pUC18 is within the left lambda arm.

This experiment demonstrated that L1231 and L1232 are homologous at some part of their sequence and these homologous regions correspond to their homology to the p2 probe.

p1231B and p1232A were also probed with M13mp8 which did not hybridize to the insert DNA (data not shown) but did hybridize to the pUC18 vector.

3.3.4 Hybridization of p1231B and p1232A to Drosophila DNA

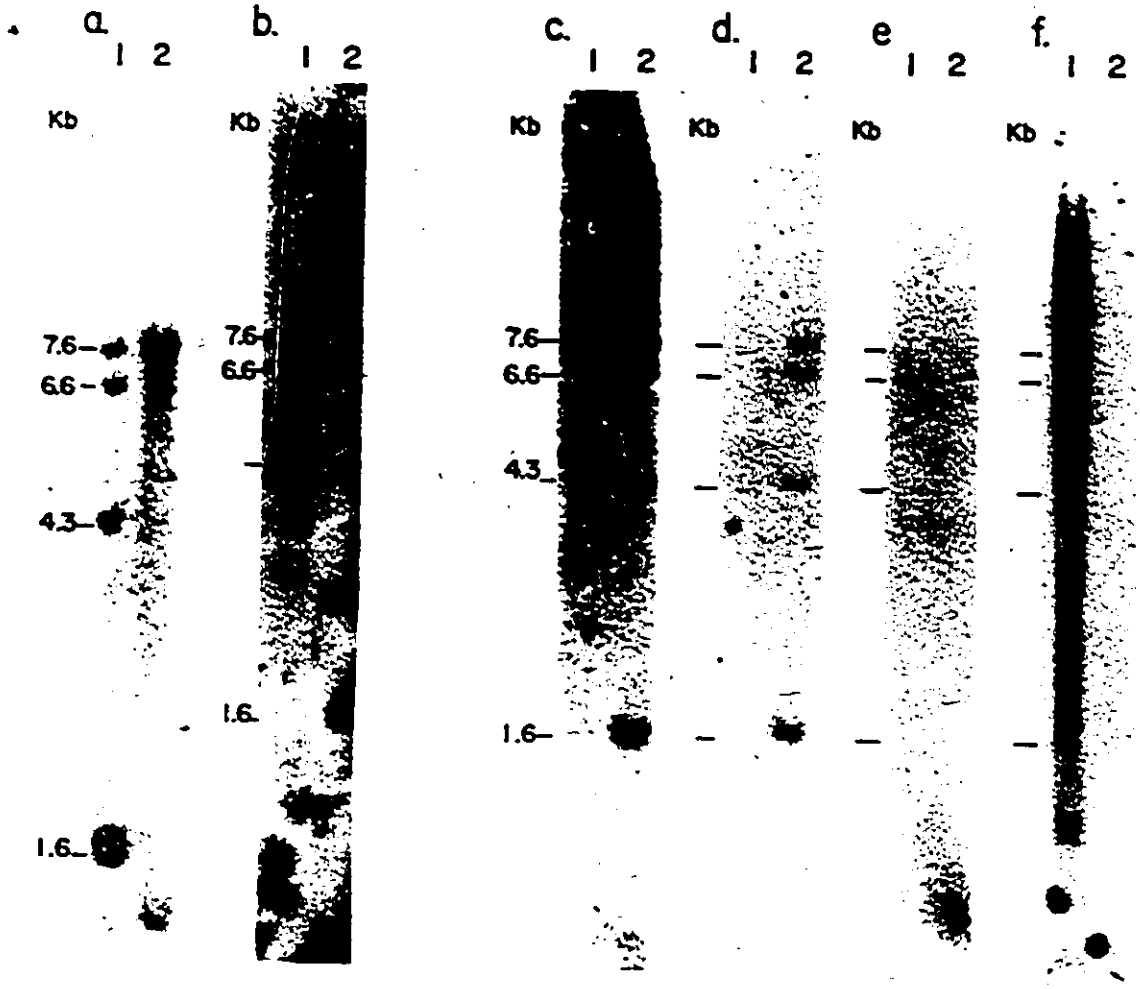
The two subclones were next hybridized to Oregon R genomic DNA in order to see if sequences in the isolated clones corresponded to unique sequences in the Drosophila genome.

Subclones p1231B and p1232A were nick translated and hybridized to EcoRI digested *D. melanogaster* strain Oregon R DNA transferred to nitrocellulose (Schleicher & Schuell). Both probes hybridized a genomic fragment at approximately 7.6 Kb (Figure 3.17 a,b). In addition, p1231B hybridized a band at 1.0 Kb and clone p1232A hybridized a band at 6.0 Kb. The plasmid pPa-1 containing the *Drosophila melanogaster* Alcohol Dehydrogenase Gene (ADH) was used as a positive control and hybridized to the predicted fragment at 4.4 Kb (Figure 3.17c). pBR322 and EMBL4 served as negative controls and did not hybridize to Oregon R DNA (Figure 3.17 d,e, respectively). As an additional control, a randomly

FIGURE 3.17

Oregon R *Drosophila melanogaster* DNA was digested with EcoRI, fractionated on a 0.8% agarose gel in TAE, transferred to nitrocellulose and hybridized in 50% formamide at 42°C to various probes nick translated to similar specific activity ($1-3x10^8$ cpm/µg). The hybridizing marker fragment sizes are indicated. The hybridized filters were exposed to X-ray film for 72 hours with an intensifying screen.

Panel a probed with p1231B; Lane 1) 0.1 ng marker, 2) Oregon R DNA. Panel b probed with p1232A; Lane 1) Oregon R DNA, 2) 0.1 ng marker. Panel c probed with *D. melanogaster* alcohol dehydrogenase gene cloned into pBR322, panel d probed with pBR322, panel e probed with EMBL4, panel f probed with random clone from Drosophila Oregon R genomic library; Lane 1) Oregon R DNA, 2) 0.1 ng marker. The marker fragments are indicated.


a.
1 2
Kb
7.6
6.6
4.3
1.6
b.
1 2
Kb
7.6
6.6
1.6
c.
1 2
Kb
7.6
6.6
4.3
1.6
d.
1 2
Kb
e
1 2
Kb
f.
1 2
Kb

isolated clone from the Drosophila genomic library was also used as probe. The hybridization pattern suggests that this clone may contain repetitive sequences from the Drosophila genome (Figure 3.17f).

The 7.6 Kb band, which may be common to both clones, indicates that the fragment in p1231B is larger than its corresponding genomic fragment by approximately 1 Kb. The genomic fragment hybridizing p1232A is approximately 600 bp larger than the 7.1 Kb band from the lambda clone but this observation is misleading because L1232 is truncated at the EcoRI site of the lambda cloning site. The observation that the genomic fragment hybridizing L1231 is smaller than the corresponding fragment from the clone, coincides with data presented later leading to the conclusion that a prokaryotic insertion sequence had integrated into *D. melanogaster* DNA. p1231B generates EcoRI fragments of 0.6 Kb and 1.4 Kb either (or both) of which could be hybridizing to the genomic DNA in the 1 Kb range. Clone L1232 yields a 5.8 Kb EcoRI fragment, a part of which is present in p1232A and which corresponds to the hybridizing 6.0 Kb band in genomic Oregon R DNA.

These observations are important because they established that the two isolated subclones contain unique sequences which hybridize to discrete sequences in the *Drosophila melonogaster* genome. The hybridization is specific to the insert DNA of the subclones because lambda and pBR322 DNA do not hybridize to Oregon R DNA.

It was interesting that both subclones hybridized to a genomic restriction fragment of similar size. This could indicate that the subclones contain the same DNA sequence. However, this is not obvious from the restriction digest data of the subclones and to confirm these observations it was necessary to carry out more refined mapping. The results of these experiments are presented later.

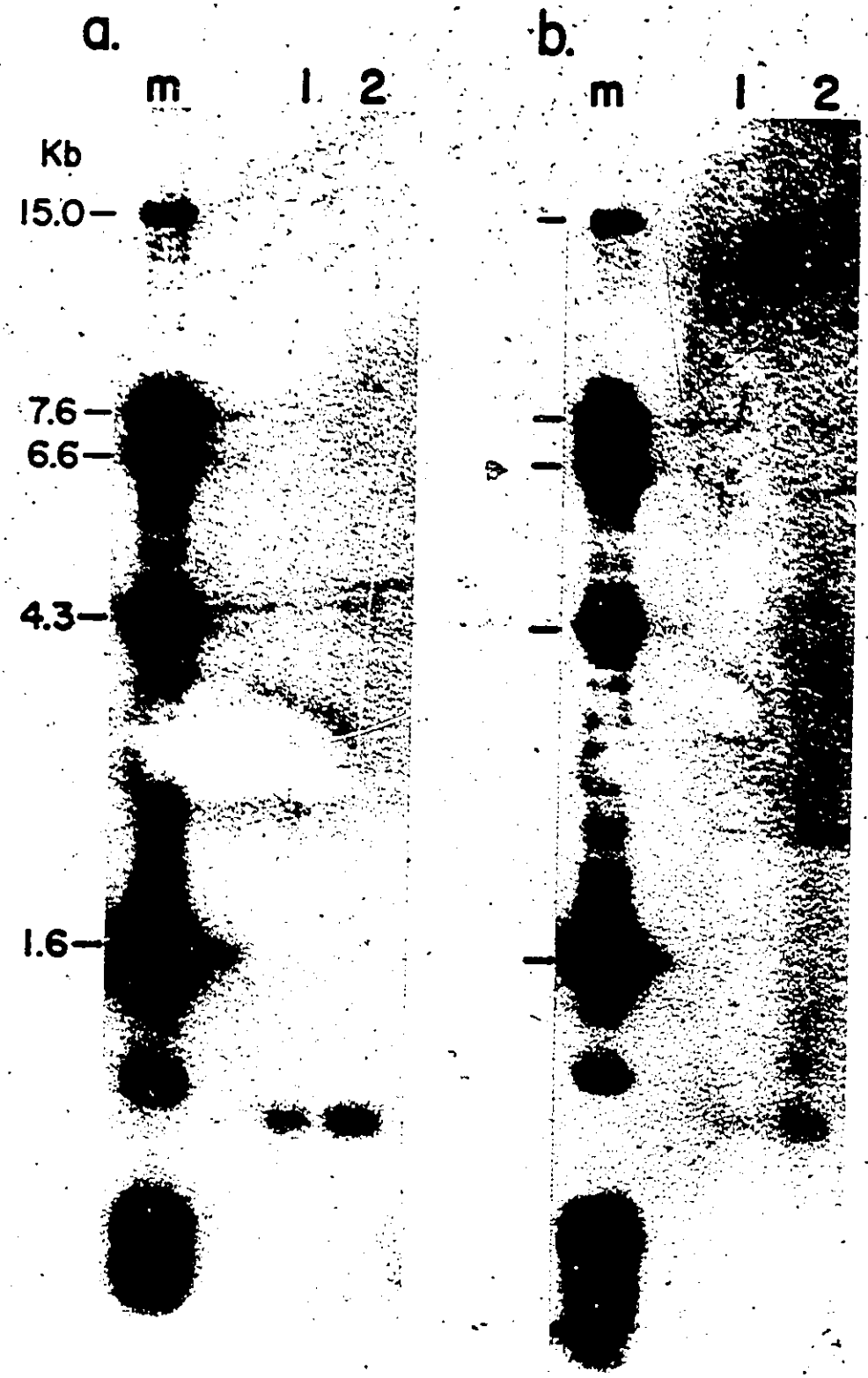
3.3.5 Hybridization of pUCa1 to Drosophila genomic DNA

To ensure that pUCa1, the probe used to isolate L1231 and L1232, did indeed encode Drosophila sequences, Oregon R DNA digested with EcoRI was probed with pUCa1 (Figure 3.18a). The 300 bp Drosophila mitochondrial ribosomal DNA from pUCa1 was cloned into pBR322 (pBR/300) and served as a control (Figure 3.18b). Both probes hybridized to a 1 Kb species (Figure 3.18a,b). However, pUCa1 did not hybridize any other Drosophila genomic DNA. Thus, the major part (3.2 Kb EcoRI fragment) of the pUCa1 cDNA did not hybridize to corresponding Drosophila sequences. This observation could have no other explanation than that, apart from the mitochondrial DNA, pUCa1 does not encode a Drosophila gene.

FIGURE 3.18

Oregon R and Canton S genomic DNA (5 µg) were digested with EcoRI, fractionated on 0.8% agarose in TAE, transferred to nitrocellulose and hybridized in 50% formamide at 42°C for 16 hours. The plasmids were nick-translated to a specific activity of 1x10⁸ cpm/µg. The autoradiographs were exposed to X-ray film for 72 hours with an intensifying screen.

Panel a was hybridized with pUCa1 and panel b was hybridized with the 300 bp Drosophila mitochondrial ribosomal cDNA cloned into pBR322. The marker (.1 µg) fragments are indicated. Lane 1) Oregon R DNA, Lane 2) Canton S DNA.


a.
m
1
2
Kb
15.0
7.6
6.6
4.3
1.6
b.
m
1
2

Simultaneously, detailed characterization of pUCa1 had been carried out by sequence analysis (M. Whitely, unpublished data). This data made clear that pUCa1 does not encode Drosophila α-amylase but a totally unrelated DNA sequence. The next part of this thesis will describe the data which enabled us to identify the sequences in the two clones isolated from the *Drosophila melanogaster* genomic library using pUCa1.

PART 2 Identifying Sequences Comprising L1231 and L1232

3.4 IDENTITY OF THE P2 PROBE

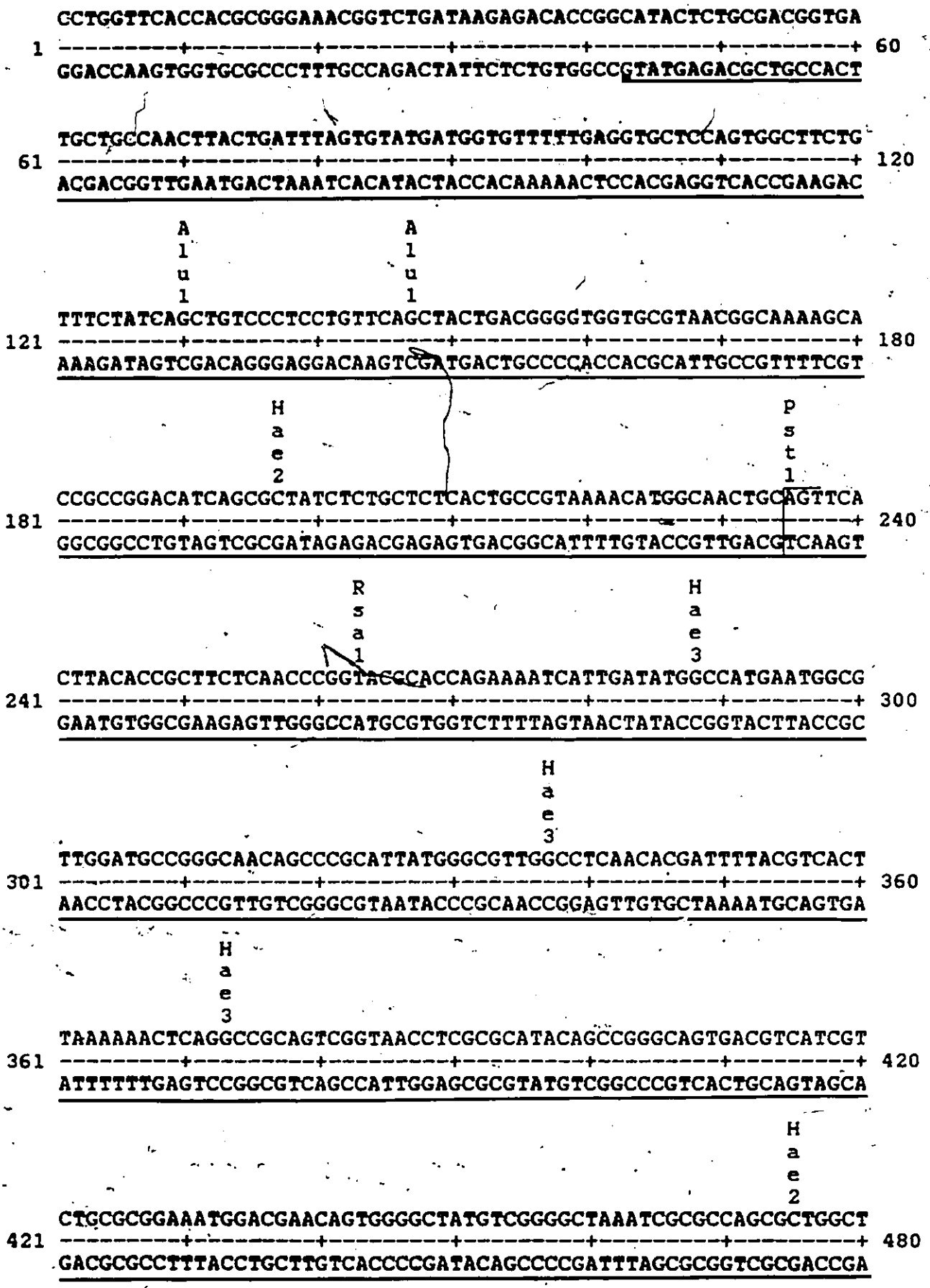
By the time we arrived at this point in the research, data had accumulated which indicated that pUCa1 does not encode Drosophila α-amylase. However, because of its starch hydrolyzing activity, we felt pUCa1 must still encode another amylase. Finally, the sequence data for pUCa1 had been completed. We compared the sequence data to computer gene banks expecting to find some homology for pUCa1 with published α-amylase sequences. Startlingly, none of the published gene sequences of starch-hydrolyzing enzymes shared extensive homology with pUCa1. To our even greater stupefaction we found that pUCa1, the probe we had used on the genomic library, shared almost perfect homology (98%) with insertion sequence IS1 of *E. coli*. Yet, pUCa1 produces an amylase-like enzymatic activity which has never been demonstrated for IS1.

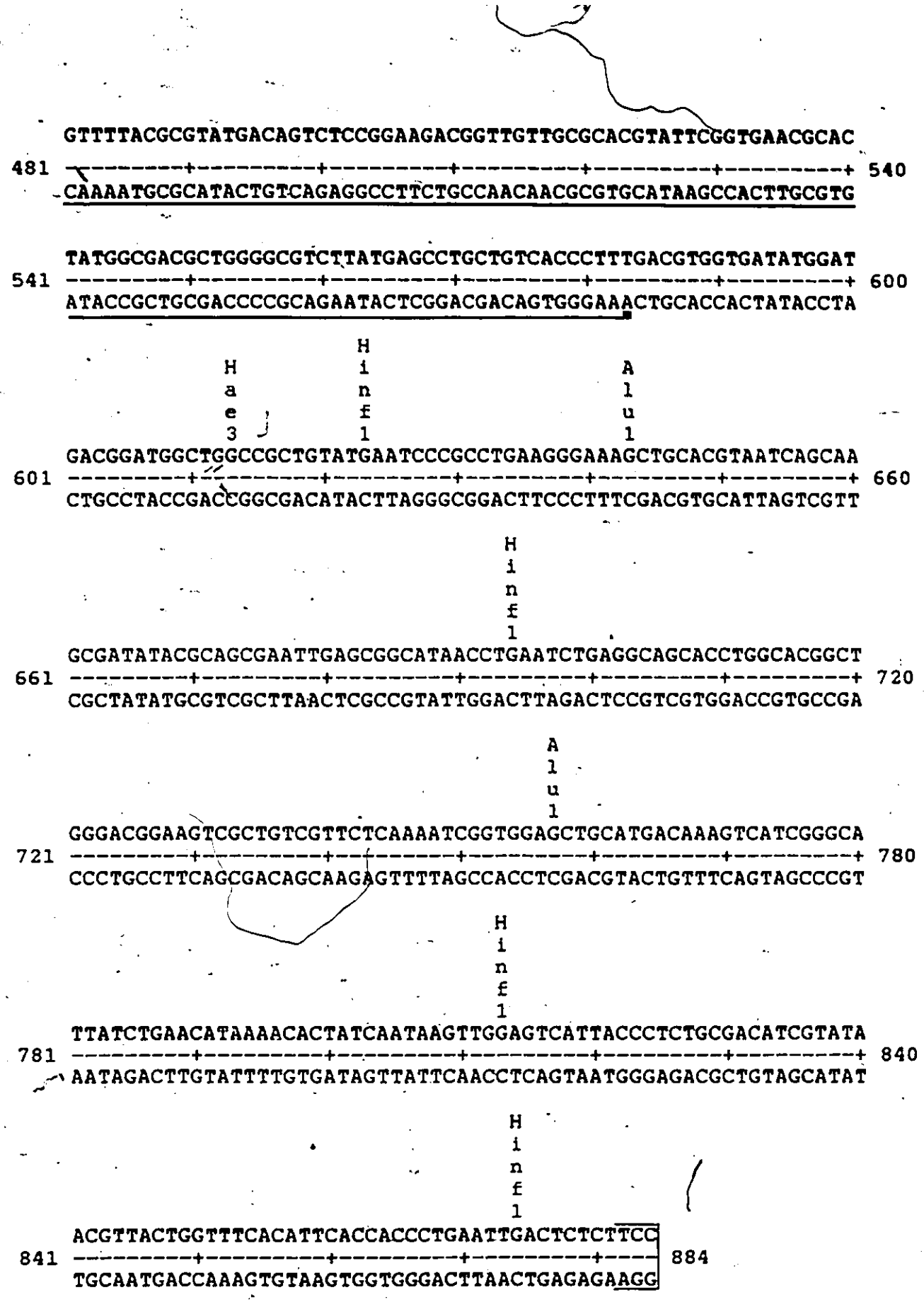
Homology of the probe (p2) with IS1 starts at nucleotide 43 and ends at nucleotide 583 in IS1. The corresponding region in p2 is from nucleotide 541 to nucleotide 1, respectively, in the reverse orientation to which p2 was sequenced (Figure 3.19). The two sequences differ at only six nucleotides. Four of these are base pair mutations within the last ten nucleotides of p2; the fifth mutation is a one base pair deletion in p2 at nucleotide 67, the sixth change is a one base pair insertion at nucleotide 119 in p2. These last two mutations disrupt and then restore the open reading frame in p2.

FIGURE 3.19

Sequence of IS1 (Johnsrud, 1979; Ohtsubo & Ohtsubo, 1978) with homology to p2 shown below (Whitely & Moreau, unpublished data) and homology to the two PstI subclones (described later) within the box brackets.

The restriction enzyme sites pertinent to mapping the two subclones are indicated: AluI, HaeII, HaeIII, HinfI, RsaI, PstI.


CCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCATACTCTGCGACGGTGA
GGACCAAGTGGTGCGCCCTTTGCCAGACTATTCTCTGTGGCCGTATGAGACGCTGCCACT
TGCTGCCAACTTACTGATTTAGTGTATGATGGTGTTTTTGAGGTGCTCCAGTGGCTTCTG
ACGACGGTTGAATGACTAAATCACATACTACCACAAAAACTCCACGAGGTCACCGAAGAC
Alu1
Alu1
TTTCTATCAGCTGTCCCTCCTGTTCAGCTACTGACGGGGTGGTGCGTAACGGCAAAAGCA
AAAGATAGTCGACAGGGAGGACAAGTCGATGACTGCCCCACCACGCATTGCCGTTTTCGT
Hae2
Pst1
CCGCCGGACATCAGCGCTATCTCTGCTCTCACTGCCGTAAAACATGGCAACTGCAGTTCA
GGCGGCCTGTAGTCGCGATAGAGACGAGAGTGACGGCATTTTGTACCGTTGACGTCAAGT
Rsa1
Hae3
CTTACACCGCTTCTCAACCCGGTACCCACCAGAAAATCATTGATATGGCCATGAATGGCG
GAATGTGGCGAAGAGTTGGGCCATGCGTGGTCTTTTAGTAACTATACCGGTACTTACCGC
Hae3
TTGGATGCCGGGCAACAGCCCGCATTATGGGCGTTGGCCTCAACACGATTTTACGTCACT
AACCTACGGCCCGTTGTCGGGCGTAATACCCGCAACCGGAGTTGTGCTAAAATGCAGTGA
Hae3
TAAAAAACTCAGGCCGCAGTCGGTAACCTCGCGCATACAGCCGGGCAGTGACGTCATCGT
ATTTTTTGAGTCCGGCGTCAGCCATTGGAGCGCGTATGTCGGCCCGTCACTGCAGTAGCA
Hae2
CTGCGCGGAAATGGACGAACAGTGGGGCTATGTCGGGGCTAAATCGCGCCAGCGCTGGCT
GACGCGCCTTTACCTGCTTGTCACCCCGATACAGCCCCGATTTAGCGCGGTCGCGACCGA


GTTTTACGCGTATGACAGTCTCCGGAAGACGGTTGTTGCGCACGTATTCGGTGAACGCAC
481
CAAAATGCGCATACTGTCAGAGGCCTTCTGCCAACAACGCGTGCATAAGCCACTTGCGTG
540
TATGGCGACGCTGGGGCGTCTTATGAGCCTGCTGTCACCCTTTGACGTGGTGATATGGAT
541
ATACCGCTGCGACCCCGCAGAATACTCGGACGACAGTGGGAAACTGCACCACTATACCTA
600
Hae3
Hinf1
Alu1
GACGGATGGCTGGCCGCTGTATGAATCCCGCCTGAAGGGAAAGCTGCACGTAATCAGCAA
601
CTGCCTACCGACCGGCGACATACTTAGGGCGGACTTCCCTTTCGACGTGCATTAGTCGTT
660
Hinf1
GCGATATACGCAGCGAATTGAGCGGCATAACCTGAATCTGAGGCAGCACCTGGCACGGCT
661
CGCTATATGCGTCGCTTAACTCGCCGTATTGGACTTAGACTCCGTCGTGGACCGTGCCGA
720
Alu1
GGGACGGAAGTCGCTGTCGTTCTCAAAATCGGTGGAGCTGCATGACAAAGTCATCGGGCA
721
CCCTGCCTTCAGCGACAGCAAGAGTTTTAGCCACCTCGACGTACTGTTTCAGTAGCCCGT
780
Hinf1
TTATCTGAACATAAAACACTATCAATAAGTTGGAGTCATTACCCTCTGCGACATCGTATA
781
AATAGACTTGTATTTTGTGATAGTTATTCAACCTCAGTAATGGGAGACGCTGTAGCATAT
840
Hinf1
ACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCC
841
TGCAATGACCAAAGTGTAAGTGGTGGGACTTAACTGAGAGAAGG
884

The only other sequence sharing extensive homology to p2 is IS1 from *Shigella dysenteriae* (95% homology), which is a close relative of *E. coli lacI* IS1. The remainder of the insert of pUCa1 is comprised of the 300 bp mitochondrial rRNA and rearranged pUC8 sequences including the *lacI* gene.

Thus, the two recombinant clones isolated from the Drosophila genomic library had been isolated, not by virtue of homology to an α-amylase cDNA, but by homology to a sequence now shown to be a close variant of IS1. The region hybridizing the probe in the recombinant clones gives a qualitatively intense hybridization signal, indicating a high degree of homology with the probe. Indeed, the stringency used during hybridization and post-hybridization washes requires a minimum duplex length of 60 bp given that the G+C content of IS1 is 53%. It was then apparent that the hybridizing sequences in the recombinant clones probably contained IS1 as well. Accordingly, the small size of the hybridizing fragments from digests of p1231B and p1232A with RsaI and Sau3AI, also suggested that these two clones contained bacterial insertion sequences.

The clones isolated from the genomic library hybridize to *D. melanogaster* DNA but they also hybridize a sequence shown to be IS1. Together, these observations led us to surmise that some of the DNA in these genomic clones shares almost perfect homology with IS1. Although the original intent of this research was not to isolate such sequences, intriguing questions are raised by these results. Two

possibilities exist for this situation: Is the region hybridizing to IS1 a *Drosophila melanogaster* sequence with extensive homology to IS1? Or, is it possible that IS1 transposed into Drosophila DNA during the preparation and/or propagation of the genomic library? The hybridization signal from Oregon R DNA probed with the two subclones is comparable to that of a single copy gene in Drosophila. Yet, from the stringency conditions employed, we estimate a substantial duplex of perfect homology. It is improbable that a Drosophila sequence would share such extensive homology to bacterial DNA. We decided to pursue these conjectures and the next set of experiments were performed to test the two possibilities.

3.4.1 Subcloning of the hybridizing PstI fragments from p1231B and p1232A

The region homologous to IS1 in the two subclones had been pared down to single RsaI fragments which are less than 1 Kb in length. In order to preserve some flanking sequences, the larger hybridizing PstI fragments of p1231B (1.3 Kb) and p1232A (2.1 Kb) were excised and subcloned into pBR322. These two new subclones (designated p1231P and p1232P) were used to characterize the IS1-homologous sequences and the flanking Drosophila sequences of the recombinant clones.

3.4.2 Hybridization of p1231P and p1232P to *D.melanogaster* DNA, *E. coli* DNA, and Bacillus DNA

Chromosomal DNA from *D. melanogaster* strain Oregon R and from *E. coli* strains C600, LE392 and Q358 were digested with EcoRI. Samples of each DNA were electrophoresed through four 0.8% agarose gels. The gels were then probed with 1) p1231P , 2) p1232P, 3) pPa-1, containing the ADH gene of *D. melanogaster*, and 4) pBR322. *D. melanogaster* ADH clone pPA-1 served as a positive control for a single copy Drosophila gene. pBR322 served as a negative control.

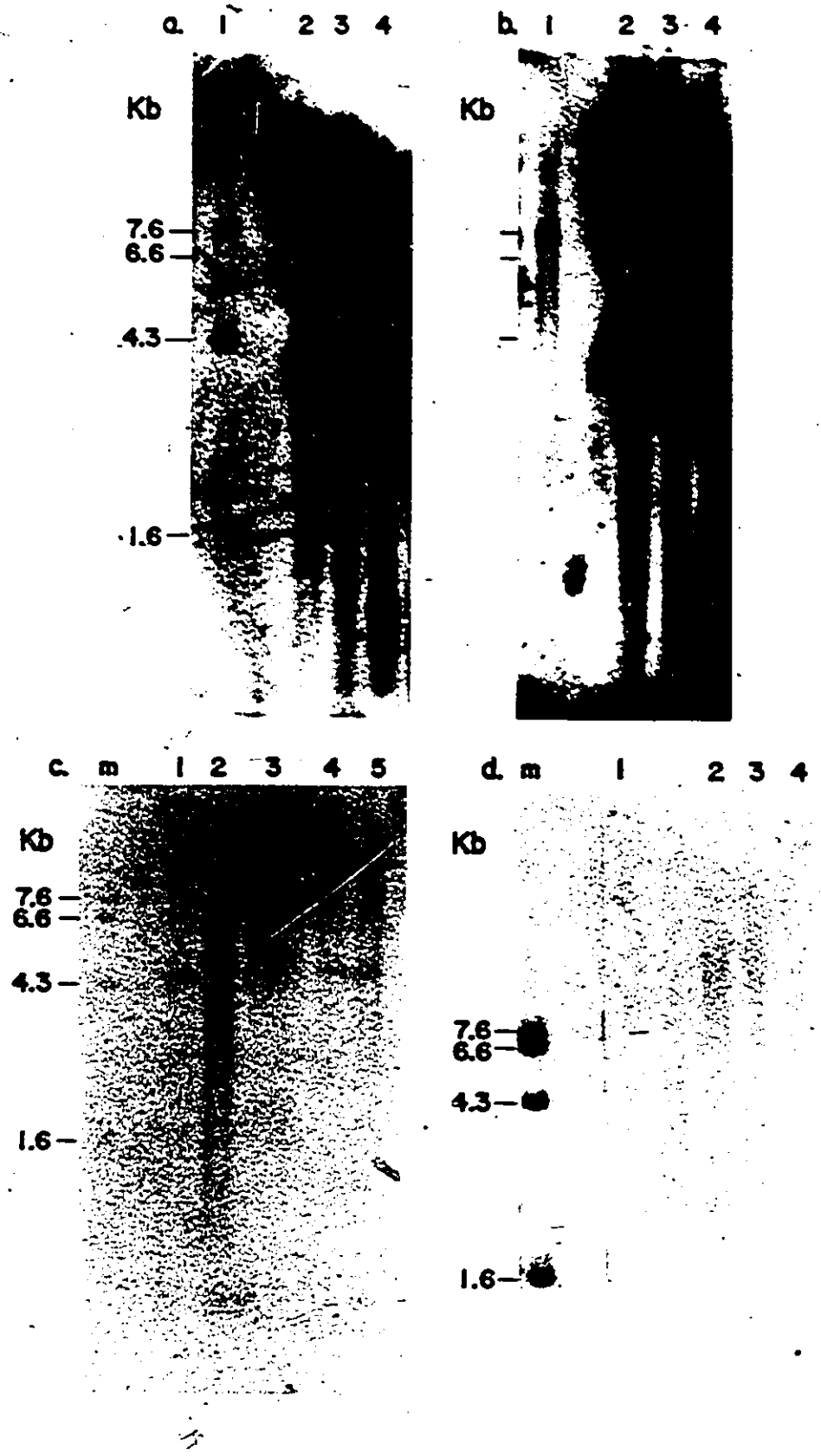
Both p1231P and p1232P hybridized very strongly to *E. coli* DNA from all three strains (Figure 3.20a,b, respectively). Several discrete bands are discernible in the prokaryotic genomic DNA and the pattern of hybridization suggests that the probe hybridizes to a dispersed, repetitive element in *E. coli* DNA.

As well, both Pst subclones hybridize under the same hybridization conditions to Oregon R DNA (Figure 3.20a,b, lanes 2, 3, 4). Both probes again hybridized to a 7.6 Kb fragment in Oregon R DNA. The ADH positive control hybridized a specific band in Oregon R DNA at 4.4 Kb (Figure 3.20c). The ADH serves as a positive control showing the hybridization intensity of a single copy gene to Drosophila DNA. The hybridization signal from ADH and from the insert DNA are of similar intensity and thus

FIGURE 3.20

Oregon R genomic DNA and *Escherichia coli* strains C600, LE392, and Q358 chromosomal DNA digested with EcoRI. The genomic DNA was fractionated on a 0.8% agarose gel in TAE, transferred to nitrocellulose and hybridized in 50% formamide at 42°C for 16 hours. The probes were labelled by random primer extension (Multiprime -- Amersham Corp.) to similar specific activity (1x10^9 cpm/µg). The autoradiographs were exposed for 12 hours with an intensifying screen. The marker fragments are indicated although the marker is not shown on the figure.

Panel a was probed with p1231P; panel b was probed with p1232P; panel c was probed with *D. melanogaster* alcohol dehydrogehase gene; panel d was probed with pBR322. Lane 1) 5 µg Oregon R DNA, 2) 5µg C600 DNA, 3) 5µg LE392 DNA, 4) 5µg Q358 DNA.


a. 1 2 3 4
Kb
7.6
6.6
4.3
1.6
b. 1 2 3 4
Kb
c. m 1 2 3 4 5
Kb
7.6
6.6
4.3
1.6
d. m 1 2 3 4
Kb
7.6
6.6
4.3
1.6

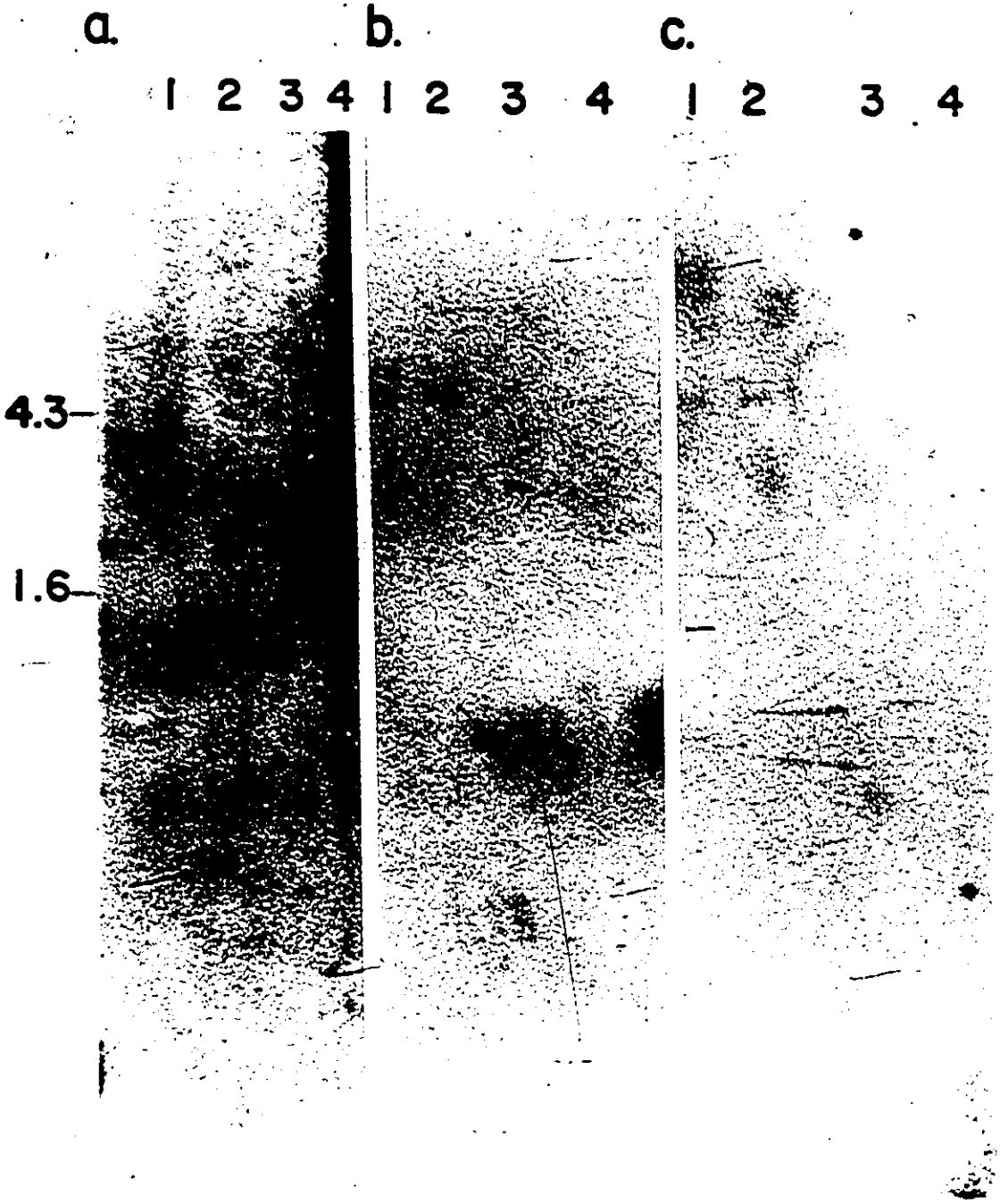
are similar in homology to their respective genomic sequences. The ADH plasmid hybridized slightly to *E. coli* DNA but, given the low complexity of *E. coli* DNA the hybridization signal cannot be directly compared to ADH hybridization to Oregon R DNA. The ADH control hybridized to one band in the C600 lane at the origin of the gel. This band may be contaminating protein or undigested DNA and no other bands are present that would suggest specific hybridization to sequences in *E. coli* DNA by the *D. melanogaster* ADH gene. pBR322 used as a negative control does not hybridize to either *E. coli* or Oregon R DNA (Figure 3.20d).

Because p1231P and p1232P are PstI subclones from the original lambda clones, we next used the Pst subclones to probe Oregon R DNA digested with PstI. We reasoned that if the sequence present in the recombinant clones was comprised solely of Drosophila sequences, p1231P should hybridize a 1.3 Kb fragment and p1232P should hybridize a 2.1 Kb fragment from Oregon R DNA. On the other hand, if IS1 sequences had inserted into Drosophila DNA, the sizes of hybridizing PstI fragments in Oregon R DNA would not correspond to the sizes of the hybridizing PstI fragments from the recombinant clones. PstI fragments from the recombinants should be larger than corresponding genomic fragments unless there are internal PstI sites in the insertion. Accordingly, PstI digested Oregon R DNA was probed with the excised PstI fragments of p1231P and p1232P

FIGURE 3.21

Oregon R DNA digested with PstI, fractionated on 0.8% agarose gel in TAE, transferred to nitrocellulose and hybridized in 50% formamide at 42°C to multiprimed plasmids ($5x10^7$ cpm/µg). The autoradiographs were exposed to X-ray film for 72 hours with an intensifying screen.

Panel a was hybridized to pBR322 DNA, panel b to p1231P and panel c to p1232P. Marker fragment sizes are indicated. Lane 1) 5 µg Oregon R DNA, 2) as lane 1, 3) .1 ng marker DNA, 4) 1 µg marker DNA.


a.
b.
c.
1 2 3 4 1 2 3 4 1 2 3 4
4.3
1.6

(1.3 Kb and 2.1 Kb, respectively) and with pBR322. Interestingly, p1231P and p1232P both hybridized a fragment at 5.0 Kb in the genomic DNA (Figure 3.21b,c).

Firstly, these results indicate that neither subclone contains a hybridizing PstI fragment corresponding to one of similar size in Oregon R DNA. Secondly, the fact that both subclones hybridized to PstI fragments of identical size, as they had in EcoRI digested DNA, suggests the possibility, with even more probability, that both subclones contain the same Drosophila sequences. Experiments described later tested this possibility further. Again, pBR322 did not hybridize to Oregon R DNA.

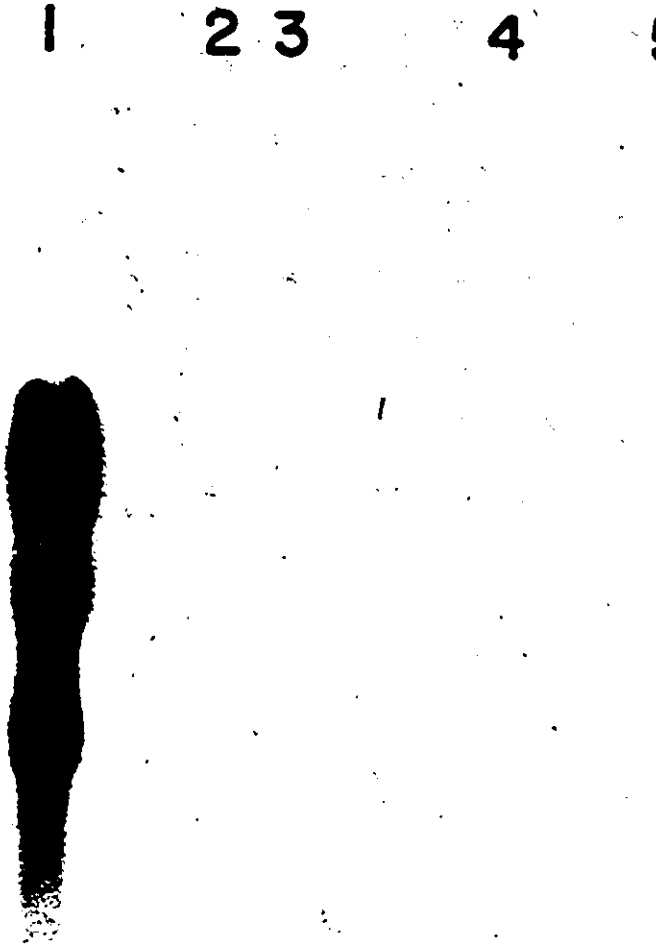
To test the specificity with which the subclones hybridized to prokaryotic DNA, p1231P, p1232P, and p2 were used to probe chromosomal DNA from two different strains of Bacillus. In parallel, Oregon R DNA and *E. coli* DNA were probed to ensure that p2 hybridized to *E. coli* DNA but not to Oregon R DNA.

The two subclones hybridized the same 7.6 Kb band in Oregon R DNA and also hybridized to *E. coli* DNA as reported previously (data not shown), whereas p2 hybridized only to *E. coli* DNA under the same hybridization conditions (Figure 3.22). None of the probes hybridized to Bacillus DNA indicating that the prokaryotic element present in all three probes is specific to *E. coli* DNA.

FIGURE 3.22

Oregon R genomic DNA, C600 DNA, *Bacillus subtilis*, and *Bacillus amyloliquifaciens* digested with EcoRI and fractionated on 0.8% agarose gel in TAE. The DNA was transferred to nitrocellulose, hybridized in 50% formamide at 42°C for 16 hours, and exposed to X-ray film for 12 hours.

The figure shows the autoradiograph of the filter hybridized to primer extended M13mp8 p2 single stranded DNA (specific activity $3x10^8$ cpm/µg). Lane 1) 5 µg C600 DNA 2) 5 µg *B. amyloliquifaciens* DNA, 3) 5 µg *B. subtilis* DNA, 4) 5 µg Oregon R DNA, 5) .1 µg marker DNA.


1
2 3
4
5

In summary, the recombinants contain Drosophila sequences as shown by hybridization to Oregon R DNA. The Drosophila sequences hybridizing to the two PstI subclones do not display homology to IS1 as demonstrated by the non-hybridization of p2 to these sequences. As well, the PstI subclones hybridized to *E. coli* sequences which also hybridize to IS1. None of the probes used hybridized to Bacillus DNA. Thus, we had demonstrated that the recombinant clones isolated from a Drosophila genomic library contained sequences from *E. coli* DNA as well as discrete sequences from *Drosophila melanogaster*.

Next, it was important to establish whether the *E. coli* sequences in the two subclones could be any sequence other than IS1.

3.4.3 Homology of *E. coli* sequences in p1231P and p1232P to other insertion sequences

We needed now to establish that the *E. coli* sequences present in L1231 and L1232 were in fact IS1. GENBANK was searched for sequences sharing homology with the IS1 probe used to isolate these recombinants. In particular, other insertion sequences (besides IS1) and transposable elements were scanned for homology.

Our experiments had been performed under high stringency conditions and only one published sequence was presented which had the extensive homology required to form a duplex of sufficient length under these conditions; that

is, IS1 isolated from *E. coli lacI* gene of the lactose operon which shares 98% homology with the probe.

3.4.4 Mapping p1231P and p1232P

It was almost certain that L1231 and L1232 contained Drosophila sequences into which IS1 had integrated. Certainly it did not appear as if the region of homology was due to Drosophila sequences which themselves are homologous to IS1. To test the premise that p1231P and p1232P each contained IS1, the two subclones were digested with several restriction enzymes based on sequence data from IS1. Fragments generated from digests of the subclones which would hybridize the probe could be predicted if IS1 was present in the two clones. Figure 3.24 outlines the fragments hybridizing the IS1 probe (p2) and the corresponding sites in IS1. A description follows of the reasoning by which we arrived at the summarizing restriction site map for each PstI subclone (Figure 3.25).

There is one PstI site at nucleotide 235 in IS1. If p1231B and p1232A contained IS1, by subcloning the 1.3 Kb and 2.1 Kb fragments, respectively, 235 bp of sequence had been eliminated from the 5' end of IS1 in the PstI subclones. Additionally, this should mean that more than one fragment from p1231B and p1232A digested with PstI should hybridize to the probe. Indeed , for p1231B another fragment at 6.0 Kb and for p1232A two fragments at 3.0 Kb and 2.7 Kb hybridized the probe, but with less intensity than the 1.3 and 2.1 Kb fragments which had been subcloned

(Figure 3.13 and 3.14, respectively). One interpretation of this data could be that pUC18 sequences hybridize the M13-derived probe to illuminate the other bands. However, since PstI does not cut within pUC18 except at the multiple cloning site, this explanation does not account for three hybridizing bands for p1232A.

Therefore, the extra hybridizing band has homology to the insert in p2 which would be due to 235 bp of IS1 to the internal PstI site. In p1231B, since there is only one lesser band, the 235 bp from IS1 would be attached directly to the pUC18 sequences. Therefore, the orientation of the proposed IS1 sequence was established in p1231B and p1232A (Figure 3.15).

Additional digests with HinfI, RsaI, AluI, HaeII, and Sau3AI (Figure 3.23) tested the hypothesis that p1231P and p1232P contained one copy each of IS1 or a closely related sequence.

The orientation of the fragment in each subclone was given by the HinfI digest (Figure 3.23, lanes 3,4). The fragment which hybridizes the probe must contain the sequence from nucleotide 235 (PstI) to a HinfI site at nucleotide 623 in the putative IS1. Fragments generated by HinfI sites past this point will not hybridize the probe because p2 homology with IS1 ends at nucleotide 584 within IS1. The insert in both subclones is flanked by HinfI sites in pBR322 at nucleotides 633 and 3364, which generate fragments of 1,382 bp and 250 bp, respectively, to the PstI

FIGURE 3.23

p1231.P and p1232.P digested with various enzymes, fractionated on 1% agarose in TAE, transferred to nylon and hybridized in 50% formamide at 42°C for 16 hours. The filters were hybridized to primer extended single strand M13mp8 p2 DNA (specific activity 3x10⁸) and were subsequently exposed for 4 hours without an intensifying screen.

Panel a (i) shows the ethidium bromide gel of Sau3al (lanes 1,2), HinfI (lanes 3,4), HaeII (lanes 5,6), and PstI (lanes 7,8); panel a (ii) shows the ethidium bromide gel of AluI (lanes 1,2) and RsaI (lanes 3,4). Panels b (i) and (ii) show the correspondig autoradiographs. In panel a the marker fragment sizes are indicated and in panel b the hybridizing fragment sizes are indicated. Each odd numbered lane is p1231P and each even numbered lane is p1232P.

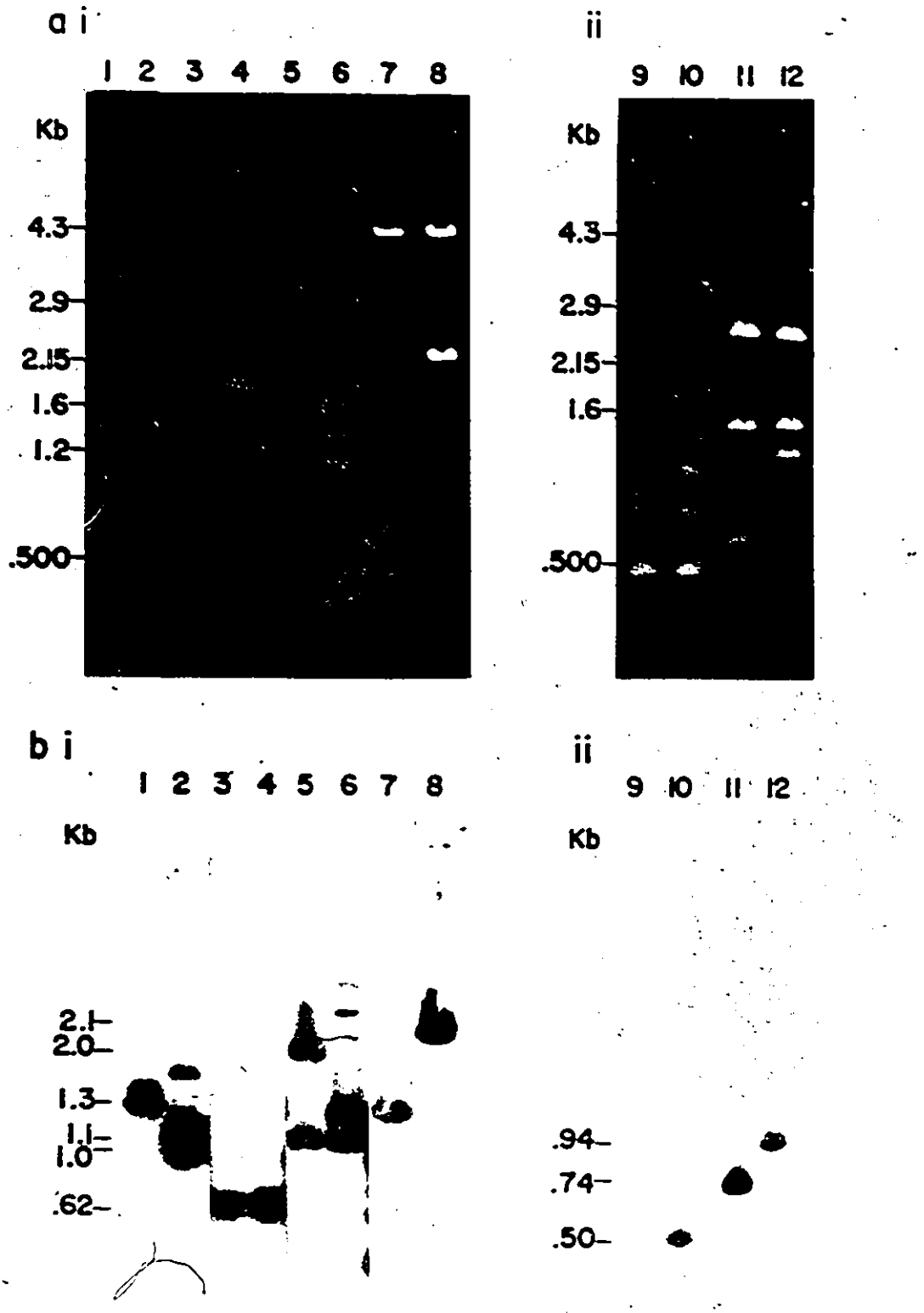
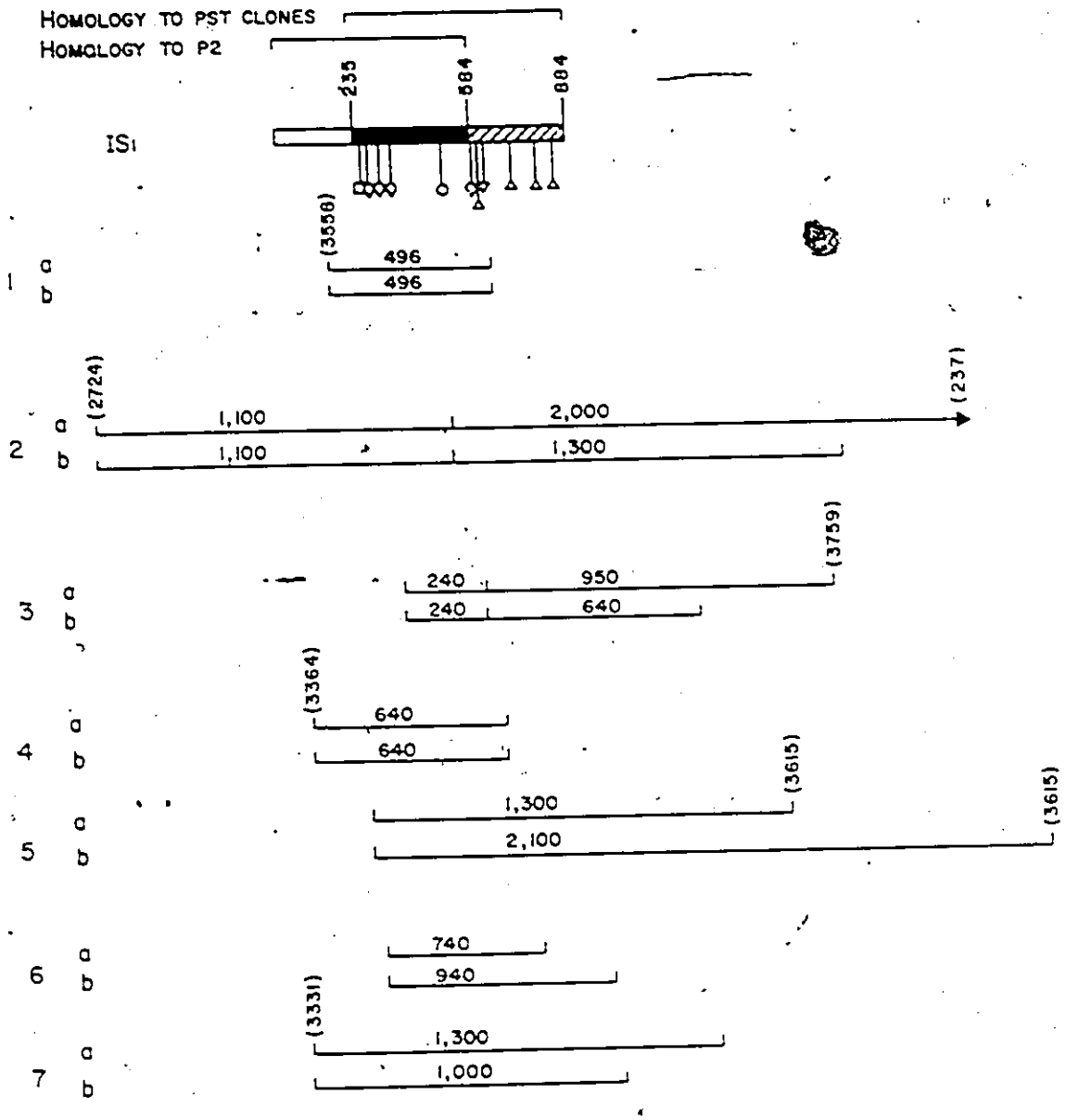

a i
1 2 3 4 5 6 7 8
Kb
4.3
2.9
2.15
1.6
1.2
.500
ii
9 10 11 12
Kb
4.3
2.9
2.15
1.6
.500
b i
1 2 3 4 5 6 7 8
Kb
2.1
2.0
1.3
1.1
1.0
.62
ii
9 10 11 12
Kb
.94
.74
.50

FIGURE 3.24

Restriction map of IS1 with homology to PstI subclones and homology to p2 shown above the map. The corresponding hybridizing fragments and their sizes from each PstI subclone are shown below the map.

The shaded area represents the extent of homology with p2 and the crosshatched area further represents the extent of IS1 in the subclones. Each (a) diagram represents the hybridizing fragments from p1231P and each (b) diagram represents the hybridizing fragments from p1232P. Diagram 1) AluI digest (AluI site △), 2) HaeII digest (HaeII site ○), 3) HaeIII digest (HaeIII site ◇), 4) HinfI digest (HinfI site ▽), 5) PstI digest, 6) RsaI digest (RsaI site □), 6) Sau3AI digest (Sau3AI site ✕). The numbers in brackets are the pBR322 sites (not shown on the map) which generate the particular fragment.


HOMOLOGY TO PST CLONES
HOMOLOGY TO P2
IS1
235
584
884
(3558)
1
a
b
496
496
(2724)
(237)
2
a
b
1,100
2,000
1,100
1,300
(3759)
3
a
b
240
950
240
640
(3364)
4
a
b
640
640
(3615)
(3615)
5
a
b
1,300
2,100
6
a
b
740
940
(3331)
7
a
b
1,300
1,000

site in pBR322 at 3,614. Therefore in one orientation the hybridizing fragment would be 1,770 bp and in the other orientation the hybridizing fragment would be 638 bp, if IS1 is present. In p1231P and p1232P, the HinfI generated hybridizing fragment is 640 bp (Figure 3.23bi, lanes 3 and 4, respectively). Therefore, the orientation is dictated by HinfI sites and in both clones, 5' and 3' designated ends would be in the same orientation as transcription of the Tetracycline resistance gene but opposite that of the β-lactamase gene into which they are inserted.

The RsaI hybridizing fragment in each case is generated from an internal RsaI site in IS1 and should be at least 621 bp long because a single RsaI site in IS1 falls at nucleotide 263. This would leave 28 bp of IS1 attached to a 1.3 Kb fragment generated from a site in pBR322 at nucleotide 2,284. This fragment would not hybridize strongly, if at all, under the stringency conditions used. The site which determines the size of the RsaI hybridizing fragment would be within Drosophila DNA as there is only one RsaI site in IS1. The hybridizing RsaI fragment of p1231P is 740 bp long (Figure 3.23bii, lane 3); in p1232P the hybridizing fragment is 940 bp long (Figure 3.23bii, lane 4).

AluI should generate a hybridizing fragment from each clone of 490 bp generated from an internal AluI site at 643 bp in IS1 and a site in pBR322 at 3,558. Each clone gives such a hybridizing fragment (Figure 3.23bii, lanes 1 and 2).

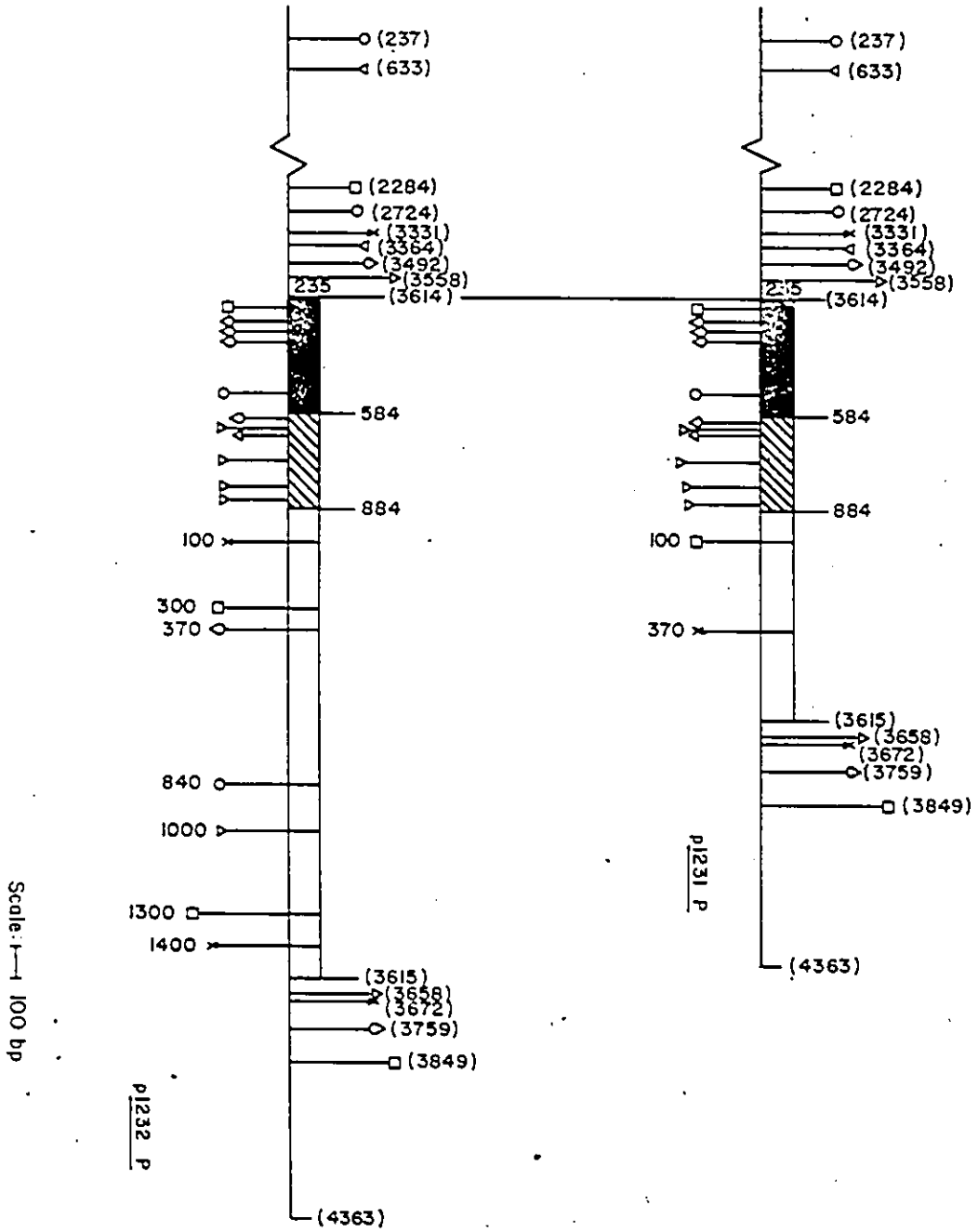
The other fragments generated from AluI, one internal fragment of 114 bp and one fragment extending into *D. melanogaster* DNA of at least 127 bp, are not hybridized by the probe because p2 homology with IS1 ends at nucleotide 584 in IS1.

HaeII cleaves IS1 once at nucleotide 475, therefore, Hae II should give two hybridizing fragments from each clone extending in either direction from this site. Given the orientation, one hybridizing fragment includes an 890 bp pBR322 fragment (from a HaeII site at 2,724 to PstI at 3614), attached to the IS1 fragment from nucleotide 235 to 475, to give a total length of 1,130 bp. Both subclones hybridize one fragment approximately 1,100 bp long (Figure 3.23bi, lanes 5 and 6). The other hybridizing fragment would have to include IS1 sequence from the HaeII site at nucleotide 475, to the end of IS1 at nucleotide 884, attached to a HaeII fragment from Drosophila sequences. In p1231P there do not appear to be any HaeII sites in the Drosophila DNA, as the second hybridizing fragment is 2,047 bp long which would include the 409 bp from IS1, approximately 650 bp of Drosophila DNA, and 987 bp from the pBR322 HaeII site at nucleotide 237 (Figure 3.23bi, lane 5).

If the IS1 of p1232P was duplicated, as might be predicted by the larger size of the insert in this subclone, we would expect to have at least three hybridizing fragments from HaeII. Such is not the case, as each subclone hybridizes two fragments from the HaeII digest. The second

FIGURE 3.25

Summarizing restriction enzyme map of each PstI subclones. The shaded area represents the extent of homology with p2 (nucleotide 235 to 584 on this map) and the crosshatched area represents the extent of IS1 in the subclones. AluI (Δ), HaeII (O), Hae III (◇), HinfI (▽), RsaI (□), Sau3A1 (X); these sites in IS1 are shown below the diagram in the shaded and crosshatched areas. The pertinent IS1 coordinates are shown above the cross-hatched and shaded areas. pBR322 sites are shown above the map and in brackets. The sites mapped to Drosophila DNA flanking the IS1 insertion in each subclone are shown below the map. Nucleotide 884 is the end of IS1; coordinates in Drosophila DNA are measured from this point (Johnsrud, 1979; Ohtsubo & Ohtsubo, 1978).


(237)
(633)
(2284)
(2724)
(3331)
(3364)
(3492)
(3558)
235
(3614)
584
884
100
300
370
840
1000
1300
1400
(3615)
(3658)
(3672)
(3759)
(3849)
(4363)
p1232 P
(237)
(633)
(2284)
(2724)
(3331)
(3364)
(3492)
(3558)
235
(3614)
584
884
100
370
(3615)
(3658)
(3672)
(3759)
(3849)
(4363)
p1231 P
Scale: 100 bp

hybridizing fragment of p1232P is approximately 1,300 bp long (Figure 3.23bi, lane 6). This hybridizing band must be an internal fragment generated from nucleotide 475 in IS1 to a HaeII site within Drosophila sequences, approximately 840 bp from the end of IS1. Another fragment, 1.5 Kb in length, does not hybridize (Figure 3.23bi, lane 6) and would be comprised of sequences from an internal HaeII site in Drosophila DNA to the flanking pBR322 HaeII site at nucleotide 237.

Sau3AI does not cleave IS1 so the hybridizing fragment should be at least 932 bp long, one site being at 3,331 in pBR322 and the other being past the end of IS1 (nucleotide 884 in IS1) within Drosophila sequences. p1231P hybridizes a fragment of 1,300 bp and p1232P hybridizes a 1 Kb fragment (Figure 3.23bi, lanes 1 and 2, respectively).

In summary, the restriction data indicates that the hybridizing sequences present in each subclone is IS1. Based on the premise that there is one IS1 present in each subclone, it would appear that p1231P has, in addition to IS1, 650 bp of *Drosophila melanogaster* DNA and that p1232P carries 1,450 bp of *D. melanogaster* DNA (Figure 3.25).

Further characterization of the subcloned PstI fragments was carried out by crosshybridization to each other and to p2 (the IS1 probe).

3.4.5 Crosshybridization of excised fragments

To demonstrate that the same *E. coli* sequence was present in both subclones the PstI fragments from each subclone and the IS1 sequence from double stranded p2 were

crosshybridized. The PstI fragments from p1231P and p1232P were excised from pBR322 and digested with RsaI and HaeIII. p2 was excised from mp8 via the EcoRI site at the end of the insert and a PvuI site 156 bp from the other end of the insert within mp8 sequences. Duplicate blots with samples of each digested fragment were probed with the excised PstI fragments of p1231P and p1232P.

Only the RsaI fragments proposed to contain the 621 bp of IS1 sequence crosshybridized with IS1 or the other subclone used as probe (Figure 3.26). The hybridizing fragment from p1231P is approximately 750 bp (Figure 3.26 b,c,d, lane 1) and from p1232P is 940 bp (Figure 3.26, b,c,d, lane 2). The hybridizing RsaI fragment from p2 is approximately 800 bp, and includes 156 bp from the EcoRI site to the PvuI site in M13mp8 as well as 620 bp of IS1 sequence (Figure 3.26, b,c,d, lane 3).

The remaining RsaI fragments from the subclones are surmised to contain only Drosophila DNA and do not hybridize to IS1. Nor do the Drosophila sequences from one subclone hybridize to Drosophila DNA from the other subclone.

HaeIII digested DNA from the excised fragments of p2, p1231P and p1232P all crosshybridized a fragment of 240 bp (Figure 3.27 b,c) as predicted by the IS1 sequence. In addition, a 950 bp fragment from p1231P (Figure 3.27,b,c, lane 3) and a 640 bp fragment from p1232P (Figure 3.27, b,c, lane 2) hybridized when self-probed or when probed with the excised fragment of the other subclone. A corresponding hybridizing fragment probing the p2 fragment with the

FIGURE 3.26

The PstI fragments were excised from subclones p1231P and p1232P (1.3 Kb and 2.1 Kb, respectively) and the EcoRI/PvuI fragment from double stranded p2 (800 bp). These were digested with RsaI and fractionated on a 1.4% agarose gel in triplicate then transferred to nylon and the triplicate blots were cut apart. Each excised fragment was also multiprimed to a specific activity of $1x10^9$ cpm/µg and used as probe on one of the three identical filters. The filters were hybridized in 50% formamide at 42°C for 16 hours and exposed to X-ray film for 4 hours without an intensifying screen.

Panel a shows one of the triplicate ethidium bromide stained gels with the marker fragments (lanes m) indicated. Panel b shows the filter hybridized with the 1.3 Kb p1231P fragment. Panel c shows the filter hybridized with the 2.1 Kb p1232P fragment. Panel d shows the filter hybridized with the 800 bp p2 fragment. Lane 1) 1.3 Kb p1231P fragment, 2) 2.1 Kb p1232P fragment, 3) 800 bp p2 fragment. The sizes of the marker fragments and hybridizing fragments are indicated for the filters.

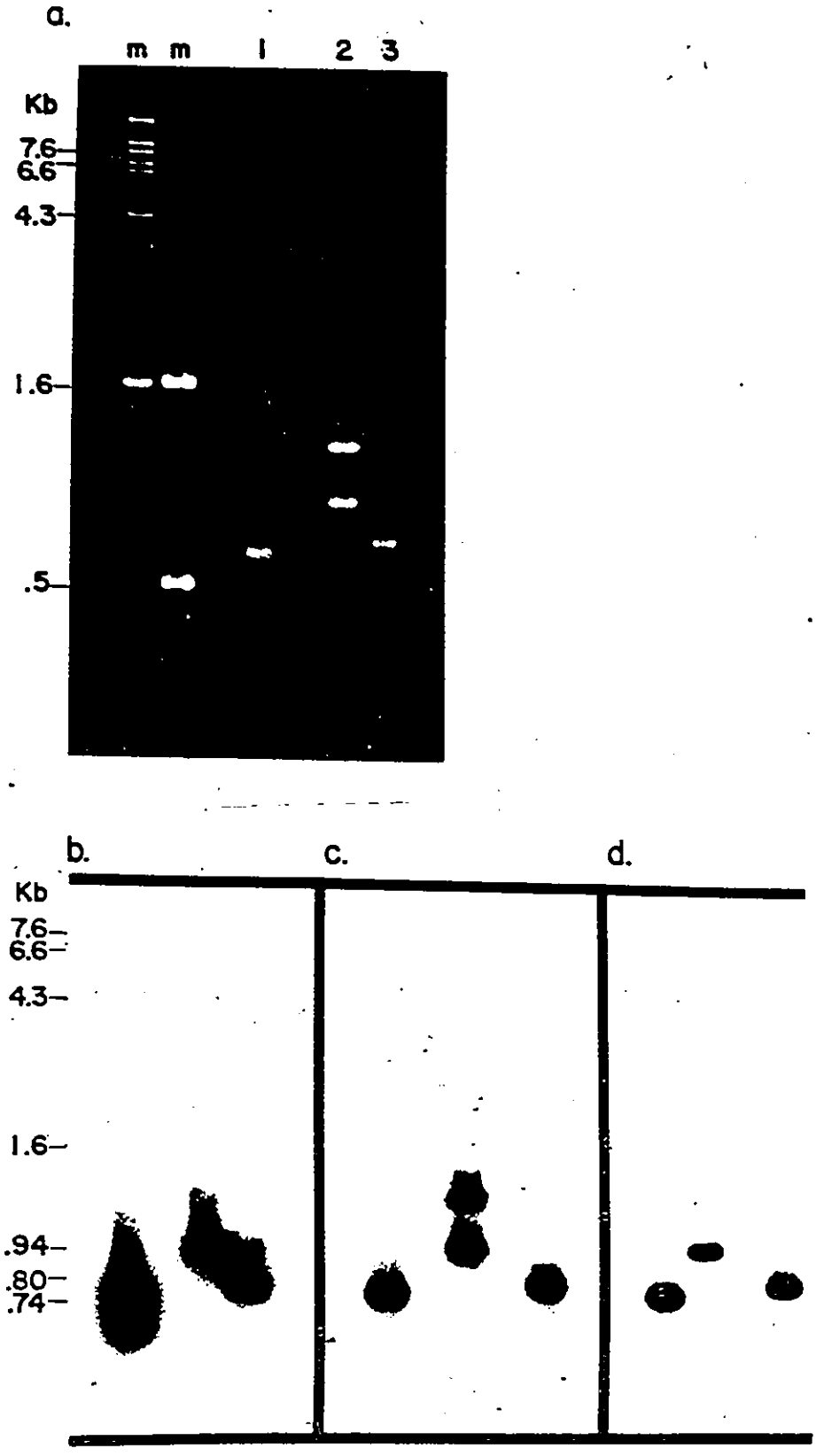
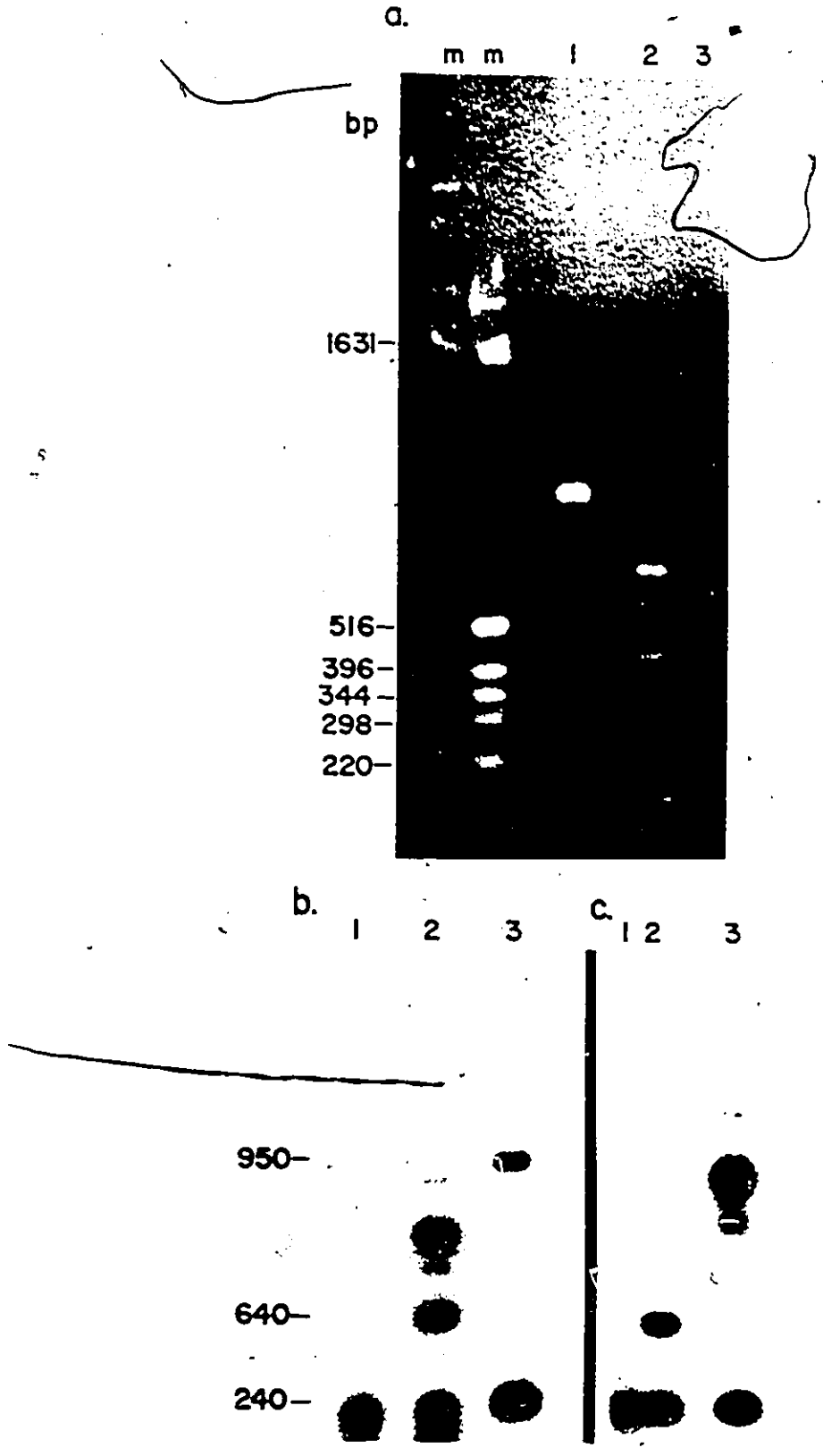

a.
m m 1 2 3
Kb
7.6
6.6
4.3
1.6
.5
b.
c.
d.
Kb
7.6
6.6
4.3
1.6
.94
.80
.74

FIGURE 3.27

The PstI fragments from subclones p1231P and p1232P (1.3 Kb and 2.1 Kb, respectively) and the EcoRI/PvuI fragment from double stranded p2 (800 bp) were excised and digested with Hae III. These were fractionated on a 1.4% agarose gel in duplicate in TAE then transferred to nylon. The duplicate blots were cut apart. Each PstI fragment was also multiprimed to a specific activity of $1x10^9$ cpm/µg and used as probe on one of the two identical filters. The filters were hybridized in 50% formamide at 42°C for 16 hours and exposed to X-ray film for 4 hours without an intensifying screen.

Panel a shows one of the duplicate ethidium bromide stained gels with the marker fragments (lanes m) indicated; Lane 1) 1.3 Kb p1231P fragments, 2) 2.1 Kb p1232P fragments, 3) 800 bp p2 fragments. Panel b shows the filter hybridized with the 2.1 Kb p1232P fragment. Panel c shows the filter hybridized with the 1.3 Kb p1231P fragment. Lane 1) 800 bp p2 fragment digested with HaeIII, 2) 2.1 Kb p1232P fragment digested with HaeIII, 3) 1.3 Kb p1231P fragment digested with HaeIII. The sizes of the hybridizing fragments are indicated for the filters.


a.
m m 1 2 3
bp
1631–
516–
396–
344–
298–
220–
b.
1 2 3
c.
1 2 3
950–
640–
240–

subclone fragments, is conspicuously absent (Figure 3.27b,c, lane 1). The cross-hybridizing fragments from the subclones are comprised of sequences from the HaeIII site in IS1 at nucleotide 613 to the end of IS1 (nucleotide 884) attached to Drosophila sequences to the first HaeIII site. Thus, the subclone-derived probes hybridize to sequences from the other clone but not to p2 sequences because p2 homology with IS1 does not extend beyond nucleotide 584 in IS1.

The Drosophila DNA flanking IS1 does not hybridize the probe unless it comprises part of a restriction fragment including IS1. In addition, because fragments generated by sites in the flanking Drosophila DNA are different in each subclone, the immediately flanking DNA of each subclone must be different. However, Oregon R DNA digested with EcoRI and PstI hybridized identical fragments when probed with both subclones, suggesting that the recombinants could contain the same Drosophila genomic sequences.

The opposing viewpoints presented by these observations were resolved by crosshybridization of Drosophila sequences from the Pst subclone of one recombinant to the larger, pUC18, subclone of the other recombinant.

3.4.6 Crosshybridization of Drosophila sequences from p1232P to p1231B

Oregon R DNA digested with PstI and EcoRI and probed with p1231P and p1232P suggested that the same PstI fragment of genomic DNA had served as a target for insertion of IS1.

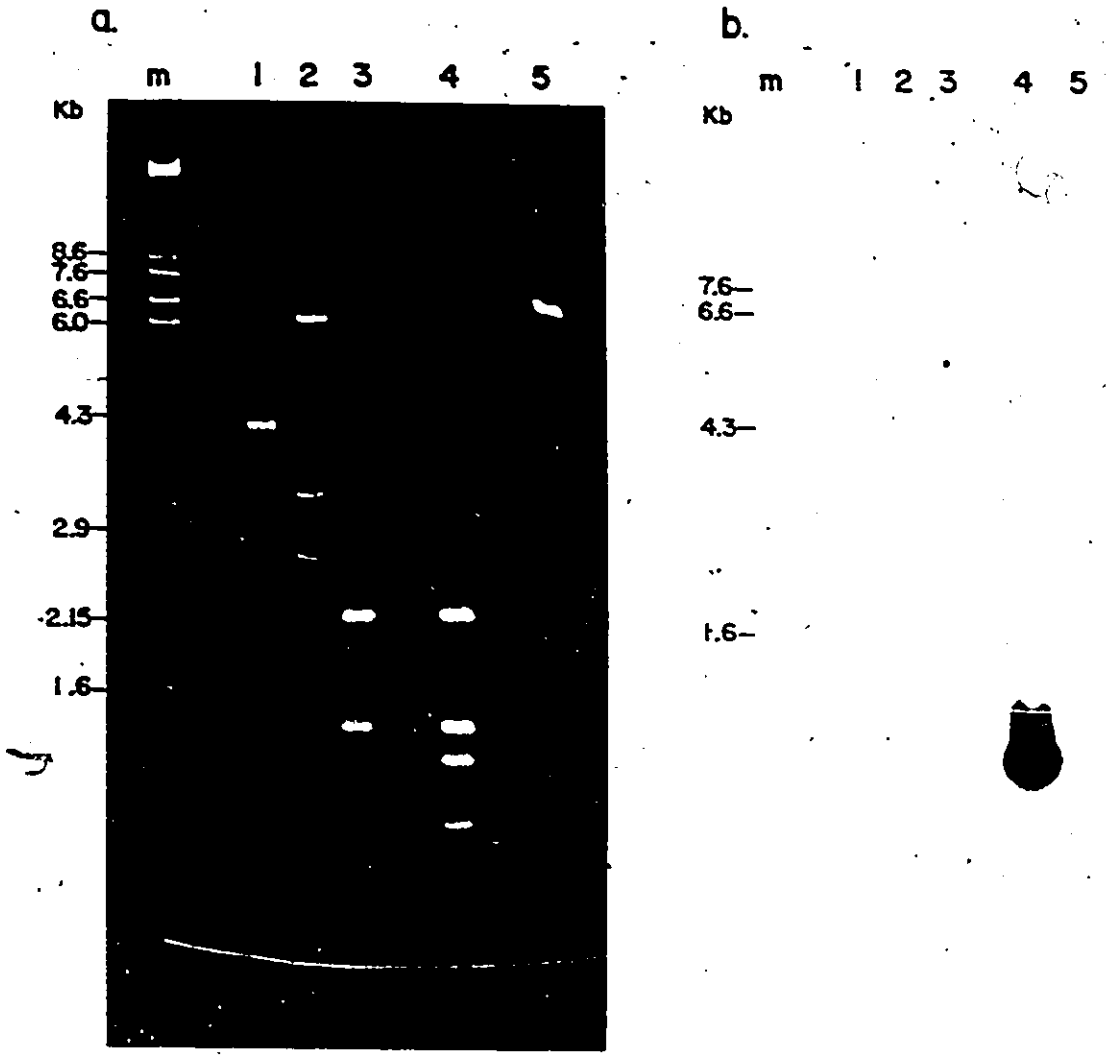
Both probes hybridized a genomic fragment of similar size (Section 3.4.2). However, the restriction data for p1231B and p1232A do not confirm this hypothesis as the restriction maps do not appear to coincide. Accordingly, p1231P and p1232P appear to contain different Drosophila sequences flanking IS1. It is possible that IS1 inserted at a different point on the same sequence of Drosophila DNA in each subclone. To determine if there is a common fragment of Drosophila DNA in both subclones, the Drosophila DNA from p1232P was excised and used to probe p1231B.

p1232P was digested with RsaI and the 1.1 Kb internal fragment, which contains only Drosophila DNA, was excised from the subclone. This fragment was radiolabeled and used to probe XhoI/BamHI and HindIII/BamHI digested p1231B. p1231P and p1232P were digested with RsaI to serve as negative and positive controls, respectively. The p1232P 1.1 Kb RsaI fragment did not hybridize to p1231B or p1231P (Figure 3.28). Therefore, it appears that the Drosophila DNA present in p1232P is not derived from the same genomic sequence as that present in p1231B. Thus, in two separate digests of genomic Oregon R DNA, two different probes hybridized fragments of similar size but unrelated sequence.

FIGURE 3.28

The 1.1 Kb Drosophila DNA RsaI fragment from p1232P was excised from the plasmid and multiprimed to a specific activity of $1x10^9$ cpm/µg. This was used to probe p1231B digested with XhoI/BamHI (lane 1) and HindIII/BamHI (lane 2), p1231P digested with RsaI (lane 3), p1232P digested with RsaI (lane 4), and double strand M13mp8 p2 (lane 5) digested with EcoRI/PvuI. The digested DNA was fractionated on a 0.8% agarose gel in TAE, transferred to nylon and hybridized to the probe in 50% formamide at 42°C for 16 hours. The resulting autoradiograph was exposed for 12 hours with an intensifying screen.

Panel a shows the ethidium bromide gel with the marker fragment sizes indicated. Panel b shows the autoradiograph with marker fragment sizes indicated.


a.
m 1 2 3 4 5
Kb
8.6
7.6
6.6
6.0
4.3
2.9
2.15
1.6
b.
m 1 2 3 4 5
Kb
7.6
6.6
4.3
1.6

Alternatively, we could digest Oregon R DNA with other restriction enzymes to ensure that the Drosophila sequences present in each subclone are different. The two clones should probably hybridize to fragments of different sizes generated by other enzymes if the Drosophila sequences are different.

3.4.7 Additional genomic data

Oregon R DNA was digested with BamHI and XhoI then hybridized with p1231P, p1232P and pBR322. The size of fragments hybridizing each probe in the digested DNA are different for each subclone (data not shown). p1231P hybridizes a BamHI fragment of 4.3 Kb and an XhoI fragment of 2.5 Kb. p1232P hybridizes a BamHI fragment of 2.5 Kb and an XhoI fragment of 2.5 Kb. Thus, it is confirmed that different DNA sequences are present in the recombinant clones and IS1 probably integrated separately into two different target sites.

3.5 SUMMARY OF RESULTS

We began by searching a genomic library for the gene which encodes α-amylase in *Drosophila melanogaster*. Our first approach was to probe the Drosophila genomic library under low stringency conditions with murine pancreatic α-amylase cDNA. This probe was expected to share minimum homology with the corresponding Drosophila gene. Accordingly, eight recombinant clones were isolated which

were unrelated to each other and probably unrelated to the α-amylase gene.

Next, we probed the Drosophila genomic library with pUCa1, a cDNA with characteristics in common with Drosophila α-amylase cDNA. Foremost among these is the enzymatic ability to hydrolyze starch encoded by pUCa1.

Probing the Drosophila genomic library using a sequence derived from pUCa1, two hybridizing recombinant phage were isolated and characterization of the two was undertaken. However, unknown to us at that time, pUCa1 contained, not α-amylase coding sequences, but IS1 of *E. coli*. When this fact was evident, we decided that characterization of the two recombinant clones should continue because observations about these clones raised interesting questions.

These recombinants hybridized to discrete Drosophila sequences but also to *E. coli* sequences. IS1 did not hybridize to Drosophila DNA. These considerations together presented us with the unequivocal conclusion that IS1 had integrated into Drosophila sequences. That a prokaryotic insertion sequence should find suitable targets in Drosophila sequences appeared to us to be an interesting phenomenon. We decided to pursue this anomaly and characterize the insertion sequence and flanking Drosophila sequences.

The experiments which followed allowed us to orient IS1 with respect to the flanking Drosophila DNA. In addition,

we determined from the degree of homology with the probe and by restriction enzyme analysis, that the prokaryotic sequence present in the two clones is IS1. Finally, it was demonstrated that the Drosophila sequences in each clone are different presenting the probability that at least two targets for IS1 insertion exist in the Drosophila genome.

Chapter IV

DISCUSSION

The α-amylase gene of *Drosophila melanogaster* presented an interesting opportunity to study a complex system of tissue-specific control of gene expression. The gene is controlled by *cis* and *trans*-acting factors at all levels of its expression: 1) α-amylase expression is limited mainly to the midgut tissues, 2) within the posterior midgut, the pattern of expression is subject to control by the linked *map* gene, 3) expression of the gene is developmentally regulated by a temporal 'switch', 4) the level of expression of α-amylase within the midgut appears to be dictated by carbohydrate source in the diet.

The object of this thesis was to isolate and characterize the α-amylase gene of *Drosophila melanogaster*. Subsequently, the isolated gene would be used for *in vivo* studies on the nature of control of tissue-specific gene expression. To accomplish the primary objective of isolating the gene, we first used a murine pancreatic α-amylase cDNA to probe a Drosophila genomic library under low stringency conditions. This probe did not immediately yield the α-amylase gene and we sought another approach. A new cDNA probe, pUCa1, synthesized from Drosophila RNA, was isolated in our laboratory by expression of a starch-hydrolyzing activity (M. Whitely and P. Moreau, unpublished data). This cDNA also hybridized to the mouse amylase cDNA

and to a 1.6 Kb Drosophila RNA. We concluded that this cDNA encoded *Drosophila melanogaster* α-amylase and it was used to screen the genomic library.

After various screenings of the library, two recombinant clones were isolated using a probe derived from pUCa1 (p2). It is estimated that these two clones share extensive homology to the probe. Characterization of the two clones proceeded.

Much to our amazement, pUCa1 was found to contain IS1 of *E. coli* and, despite the starch-hydrolyzing capacity encoded by pUCa1, we could find no homology with sequences encoding starch-hydrolyzing enzymes. The two clones isolated probing with pUCa1 sequences also appear to harbour sequences closely related, if not identical, to IS1. In addition to the IS1, the two recombinants were shown to contain *D. melanogaster* sequences. Characterization of the two clones continued because we had witnessed a unique event warranting explanation and discussion: the integration of a prokaryotic insertion sequence into eukaryotic genomic sequences.

The following discussion continues in three parts. The first part outlines the problems encountered searching for the α-amylase gene. The second part provides a synopsis of observations which led to the conclusion that IS1 has integrated into *D. melanogaster* genomic sequences. Finally, the third part provides a general discussion speculating on

possiblities and implications raised by results presented in this thesis.

4.1 Probing a Drosophila genomic library for α-amylase genes

As stated in the Introduction, the original intent of this thesis was the isolation of the α-amylase gene of *Drosophila melanogaster*. A mouse pancreatic α-amylase cDNA, pMPa21 (Schibler, et al, 1980), was employed to probe a *Drosophila melanogaster* genomic library under low stringency conditions. When screening a genomic library representing four to five total genomes, several independent clones are expected to share common hybridizing fragments. We also anticipated isolating many unrelated clones because we were using a probe with low homology to the gene of interest under low stringency conditions. Thus, isolating related clones under these conditions would be especially reassuring.

Using this strategy, eight recombinant clones which hybridized the murine probe were isolated. These were characterized by restriction endonuclease digestion and hybridization with the probe. As restriction analysis progressed, however, it became obvious that the eight clones were not related. Fragments of similar size in restriction digests were not all hybridizing the probe. In particular, the two fragments of identical size generated from two clones (105 and 110) by digestion with EcoRI which had looked so promising, did not yield identical hybridizing

fragments when digested with other enzymes. We were forced to conclude that the low stringency conditions used for the isolation of Drosophila sequences homologous to pMPa21 resulted in the isolation of sequences unrelated to each other and probably unrelated to α-amylase. Increasing the stringency resulted in the probe not hybridizing to any of the recombinant clones.

The α-amylase gene is not highly conserved between species as evolution has not imposed a stringent requirement for the enzyme. Genes encoding fundamental structural proteins such as actin and histones or developmental protein 'switches' like the homeobox genes, characteristically share a high degree of homology between species. The amylase gene has a less conspicuous role in the development or structural integrity of the organism and so is subject to more mutations than the more fundamental genes. For this reason, it is expected that the degree of homology between sequences coding for amylases should be minimal. However, we predicted that the basic enzymatic function of all eukaryotic amylases would provide a common 'core' sequence encoding the active site of the enzyme. This reasoning led us to believe that the mouse pancreatic cDNA could be used to isolate the core sequence of the Drosophila amylase gene. Indeed, we were later proven correct when the *Drosophila melanogaster* α-amylase gene was isolated using a mouse salivary α-amylase cDNA as probe (Gemmill, et al, 1985; Levy, et al, 1985).

The isolation of unrelated sequences probing with the mouse cDNA was a turning point as we considered two options to isolate the amylase gene: we could continue probing the library with the murine cDNA until related clones were isolated which, we hoped, would encode α-amylase; or, we could isolate a more homologous probe. The first strategy was rejected because at this time pUCa1 had been isolated. pUCa1, a cDNA made from Drosophila mRNA, was isolated because it expressed a starch-hydrolyzing activity from an expression vector. Bacterial colonies harbouring pUCa1 and induced by IPTG in the growth medium were surrounded by 'haloes' of hydrolyzed starch when the plates were stained with an Iodine reagent. Screening putative positive clones using a cDNA homologous to the gene has successfully led to isolation of several genes including the mouse α-amylase gene family (Schibler et al, 1982). As already described, detailed characterization of pUCa1 led to the inevitable conclusion that the cDNA did not encode α-amylase. Instead, the pUCa1 insert is comprised of rearranged pUC8 sequences, a 300 bp cDNA of Drosophila mitochondrial rRNA sequences, and an almost identical variant of *E. coli* IS1. Obviously, when we decided to use pUCa1 as a probe, this data was not yet available to us.

pUCa1 was first used to screen the eight recombinants isolated using the murine probe. None of them hybridized pUCa1 and the clones were rejected as not encoding the gene for α-amylase. pUCa1 was then used to re-screen the genomic

library to isolate the α-amylase gene. Screening the *Drosophila melanogaster* genomic library using pUCa1 and its subcloned derivatives as probes yielded the two positive recombinant clones already described.

The equivalent of 6.5×10^5 plaques were screened to isolate these two positive clones. Interestingly, both clones were isolated using the pUCa1 deletion mutant, p2, as probe. During previous screenings, positive clones may have been present but may not have been detected due to consistently high background hybridization levels. Not surprisingly, pUCa1 and the larger deletion mutants contain rearranged pUC8 sequences including the *lacI* gene which contributed to the background problem. In p2 all pUC8 sequences have been deleted and only the modified IS1 sequence remains, hence, it gives lower background. It is apparent now why it was so difficult to overcome the background problem that plagued every screening of the library. Firstly, pUC8 sequences present in most of the deletion mutants, except p2, hybridized to the host strain or to lambda sequences. As well, the *lacI* gene present in all probes except p2 would cause interfering hybridization to Q358, the first strain used to propagate the library. Lastly, LE392, the strain used most often for plating the library, contains IS1 sequences, as demonstrated in the Southern blots of *E. coli* DNA. Every probe used, with the exception of pMPa21, harboured IS1 sequences. p2, which contains only IS1 sequences, hybridized to the two isolated

clones above the background of IS1 hybridizing to the host strain. This stronger relative hybridization signal allowed the isolation of the two plaques because they harboured more than the background number of IS1 homologous sequences. Accordingly, even on the two plates with positive plaques, the negative plaques are still visible, probably as a result of chromosomal copies of IS1 hybridizing the p2 probe. The combination of hybridizing pUC8 sequences and IS1 signal from the larger probes would have been sufficient to obscure signals from positive clones on previous screenings.

Inevitably, the isolation of these two clones leads us away from discussion of the original object of isolating the Drosophila α-amylase gene because these two clones do not contain the gene. Nevertheless, they were isolated by methods which are commonly employed to isolate genes using a probe capable of encoding an enzymatic activity akin to α-amylase. It is only in the characterization of an isolated sequence that the nature and origin of that sequence is truly determined. Regardless of the origin of the sequences we have isolated, they may have as easily been the amylase gene, and we would have proceeded in an identical manner.

Duplicated α-amylase genes have now been isolated from several strains of *Drosophila melanogaster*, including the amylase-null strain (Gemmill, et al, 1986). In each of these strains the gene has been characterized by

restriction endonuclease digestion and hybridization to the probe, the mouse salivary α-amylase cDNA (Schibler, et al, 1980). Undoubtedly, we would also have isolated the gene using the mouse pancreatic cDNA as probe.

We do not know how many clones Gemmill, et al, isolated and characterized before the α-amylase gene was found. To narrow the putative positive clones to a few possibilities, Gemmill's group hybridized them in situ to Drosophila polytene salivary chromosomes. The positive clones hybridized to the region on chromosome 2R to which the α-amylase genes had been mapped (Bahn, 1971). The positive clones were confirmed as α-amylase genes by expression of a protein hydrolyzing starch and giving the isozyme pattern expected from the Drosophila strain used to make their library (Levy, et al, 1985). We isolated and characterized only eight clones of 50 that hybridized the murine pancreatic cDNA on the first screening of the library. To narrow our search for the gene, we used a probe which was believed to encode a cDNA homologous to the α-amylase gene. Ironically, had we not been looking for the *Drosophila melanogaster* α-amylase cDNA by screening for the enzymatic activity, we would never have isolated pUCa1. The fact that this cDNA actually produced an enzymatic activity like α-amylase and hybridized a 1.6 Kb band on Northern blots, prompted us to use it as probe.

Thus, instead of probing for the α-amylase gene we were actually probing for insertion sequence IS1 in the

Drosophila genomic library. Even so, we have characterized something unique, an event which has not been previously documented and which raises provoking questions.

4.2 Characterization of sequences in the positive clones

Restriction endonuclease digests and hybridization of Southern transfers to the p2 probe immediately narrowed the homologous region of the two positive clones. These data are summarized in Figures 3.11 and 3.15. At this point in the characterization, it was not yet evident that the sequences which had been isolated were not related to α-amylase.

The first real indication that L1231 and L1232 did not contain the α-amylase gene came from restriction enzyme digests with Sau3AI and RsaI. Both of these enzymes yielded one small band from each clone which hybridized the probe. RsaI in particular gave a very small hybridizing fragment from each clone, less than 1 Kb, which is not sufficient to encode α-amylase.

Simultaneously, sequencing data established that pUCa1 did not contain the four conserved domains characteristic of α-amylases (Mackay, et al, 1985). As well, the cDNA was too large (3.5 Kb) to encode only the α-amylase messenger, and sequencing data did not reveal any duplication that would account for its large size. Finally, the major part of the cDNA (3.2 Kb EcoRI fragment) did not hybridize to Drosophila genomic sequences. Comparison of the DNA sequence of pUCa1 to published sequences in GENEBANK provided conclusive

evidence that pUCa1 did not contain α-amylase. Instead, pUCa1 had "picked up" a mutated bacterial insertion sequence, a close variant of IS1. The remainder of pUCa1 was comprised of rearranged pUC8 sequences and 300 bp of Drosophila mitochondrial ribosomal DNA. The species hybridizing to pUCa1 on the Northern blot is actually the mitochondrial ribosomal RNA corresponding to the 300 bp EcoRI ribosomal DNA fragment from pUCa1.

Thus, facts about pUCa1 together with observations from the two isolated genomic clones presented the indisputible conclusion that none of these sequences encode *Drosophila melanogaster* α-amylase.

The search of GENBANK for sequences homologous to IS1 yielded information that was important to identifying sequences in the two positive clones hybridizing to IS1. The only published sequences with enough homology to account for the very intense hybridization signal observed were two closely related variants of IS1. Thus, restriction digest data from the two PstI subclones were analyzed from the viewpoint that IS1 was present in each clone. As summarized in Figure 3.25, these data demonstrate that the subclones have the restriction and hybridization patterns expected if IS1 is present in the inserts. Accordingly, the two PstI subclones hybridized strongly to three strains of *E. coli* DNA on Southern blots. Comparison of the Drosophila ADH clone hybridized to the single copy Drosophila gene demonstrates that the subclones do not hybridize in a

similar manner. The subclones hybridize to several bands of the prokaryotic DNA; this is characteristic of a dispersed, repetitive sequence being present in the genome. The modified IS1 in p2 hybridizes to C600 genomic DNA giving a similar pattern. As well, the region of homology between the two clones which is demonstrated by cross-hybridization of the 8.5 Kb and 7.1 Kb EcoRI fragments, coincides with their homology to the probe. Taken together, these observations point to IS1 sequences being present in the two recombinant clones isolated from a *Drosophila melanogaster* genomic library.

DNA flanking the IS1 sequences in the two recombinant clones could have been bacterial sequences inadvertently cloned into the vector and propagated along with Drosophila recombinant clones. However, this does not appear to be the case since both recombinants hybridize to discrete Drosophila genomic sequences. The difference in molecular weight between the genomic fragments and the fragments isolated from the recombinant clones can be accounted for by the size of the insertion sequence. For instance, the 8.5 Kb EcoRI fragment from L1231 is approximately 1 Kb larger than the corresponding fragment hybridizing in EcoRI digested Oregon R DNA.

The relative intensities of hybridization signal between prokaryotic DNA and eukaryotic DNA are not compatable because the genome of *E. coli* is several orders of magnitude less complex than the genome of Drosophila and

thus favours the kinetics of hybridization. However, the intensity of hybridization to Drosophila DNA does compare to the hybridization signal from the single copy ADH gene of *Drosophila melanogaster*.

The observation that both subclones hybridized fragments of similar size in EcoRI and PstI digested Oregon R genomic DNA may have indicated that the same genomic sequence served as target in both clones. However, Drosophila sequences flanking the insertion sequence in the clones do not cross-hybridize. Accordingly, the enzyme digest data for each clone is unique. The restriction enzyme sites immediately flanking the insertion may be different in each clone due to DNA rearrangement caused by the insertion, but outside of this region, the two clones should share common sites if the target sequence in both clones are the same.

Sequence requirements at the target site are more stringent for IS1 than for some drug resistance transposable elements and for phage Mu (Kleckner, 1977). However, the DNA sequence requirement at the target site is not absolute. The insertion sites of IS1 are usually AT-rich regions. The enzymatic machinery by which IS1 integrates at a target site must have recognized similar target sites present on the eukaryotic DNA. Confirmation of this would require sequencing of the eukaryotic target site in both cases or DNA hybrid melting experiments.

4.3 Implications and possibilities

Drosophila DNA is not the usual target for bacterial insertion sequences, as in their normal contexts the two genomes do not meet. However, by propagating eukaryotic genomic libraries in bacterial hosts we force the blending of genetic elements between prokaryotes and eukaryotes. To date, this is the first documented isolation of a prokaryotic mobile genetic element from eukaryotic genomic DNA.

Insertion sequences do not encode selectable marker phenotypes and were originally isolated from the bacterial genome because they can cause revertible, polar mutations of genes into which they integrate. Insertion of IS into genomic eukaryotic DNA is difficult to detect because isolation of eukaryotic genomic sequences is not based on selection by gene product. It is probable that the nonsense codons present in most reading frames of IS would cause mutations preventing expression of eukaryotic genes.

Within a prokaryotic host, selection of eukaryotic genes into which prokaryotic drug or metal resistance transposons have integrated could be facilitated. It is possible that investigators have found such insertions in their genomic libraries, but this has never been documented. Identification of insertion sequences in eukaryotic genes would go virtually undetected because IS do not have marker phenotypes. Well characterized eukaryotic genes isolated from prokaryote-propagated libraries may contain insertion

sequences which have not been detected because expression of the gene is not required.

Recently, IS1 has been isolated from a human genomic library constructed in a cosmid vector (Fernandez, et al, 1986). The insertion sequence was present in the vector, not in the cloned human DNA. Insertion of IS1 in several independently isolated clones had caused repeated deletions of large segments of the human DNA insert. Indeed, some isolated clones contained only the vector and IS1, the human DNA had all been deleted.

In another study, IS1 was isolated from three independent recombinant clones of double stranded copy DNA of bacteriophage MS2 cloned into pBR322 (Devos, et al, 1979). In each case, IS1 had inserted into the poly(dA)-poly(dT) linker used to clone the phage DNA into the vector. Surprisingly, in successive transfer experiments, the IS1-containing molecules became the predominant species in cultures inoculated with recombinant clones with and without the IS1 sequence present. IS1 presence in the plasmid conferred a selective advantage over plasmids which did not contain IS1.

In our study, the insertion sequence has integrated into Drosophila DNA and not into the vector. IS1 is often responsible for illegitimate recombination events such as deletions and translocation of DNA sequences near the target site (Reif & Saedler, 1975). Hot spots for deletions may not always coincide with ends of the insertion element

(Ohtsubo and Ohtsubo, 1978). In our recombinants the presence of IS1 does not appear to have precipitated large deletions of DNA. Therefore, it would appear that in this case insertion of IS1 has not disturbed the stability of the replicon. Presumably, however, most IS1 insertions into lambda sequences would cause mutations resulting in loss of these bacteriophage to the library, hence, only insertions into cloned DNA were isolated. The presence of IS1 does not appear to confer a selective advantage to the phage or the host cell since many plaques were screened before these two were isolated.

It is not clear whether transposition of the two elements isolated during this study were the result of genetic exchange from one clone to the other or if these were independent insertion events once bacteriophage DNA was introduced into the cell. In the former case, one of the isolated recombinants originally had two copies of the insertion. Insertion of one of these copies into the other clone could be described by recombination between Drosophila sequences common to both clones. Further, homologous recombination of Drosophila sequences resulted in exchange of Drosophila sequences and loss of one copy of IS1 from the first clone. In such a case, Drosophila sequences should be common to both clones. However, the data do not support this view; the clones have dissimilar restriction digest patterns and their respective Drosophila genomic sequences do not cross-hybridize. In the latter case, in which the

two clones are the result of independent insertion events, the insertion element recognized a common target site on two independent clones. The DNA flanking the insertion in each clone would then be dissimilar except for general homology of the target sites. This possibility seems to be borne out by the evidence presented.

It is not possible to test whether the integrity of the Drosophila sequences surrounding the insert is preserved until these sequences are identified. However, sequencing of the termini of the insertion sequence would establish whether they are intact and whether duplication of 8-9 bp at the target site occurred upon insertion. Identification of this characteristic duplication of target sequences would prove that insertion of the element was an active event rather than the result of inadvertant cloning of bacterial sequences into the genomic library.

The proposed pathways of transposition in prokaryotes and eukaryotes are descriptively similar (Calos & Miller, 1980; Bukhari, 1981). In both systems, insertion is believed to occur by duplication of the mobile element, cutting of DNA at the target site, insertion of one copy of the element into the break, and repair of the cuts made at each end of the insertion. Alternatively, it is believed that IS1 can transpose directly to a new target site. The enzymes required for cointegrate formation by IS1, namely insA and insB (Machida, et al, 1984), must have recognized a suitable target for integration of IS1 in eukaryotic DNA for

the insertions described in this thesis to take place. IS1 does not have an absolute sequence requirement for a target site, therefore DNA sequence cannot be the only indication of a suitable insertion point. Sequences rich in AT-residues are usual target sites in prokaryotes. In addition, IS1 has been shown to prefer integration into target sites that display homology with the inverted repeats of the element (Calos & Miller, 1980). As well, integrity of the terminal repeats of IS1 is essential for efficient transposition (Gamas, et al, 1985). Sequencing of the insertion sites in Drosophila DNA would determine if there is homology to IS1 at these targets. However, regardless of how IS1 recognizes suitable insertion sites, those sites must be present in the two recombinant clones.

Interestingly, IS1 insertions into operons frequently occur into sequences conforming to the Pribnow consensus, a defining element of the prokaryotic promoter. The Pribnow box, like the eukaryotic Goldberg-Hogness sequence, is involved in RNA polymerase binding. Ohtsubo and Ohtsubo (1978) suggest that RNA polymerase may play a role in IS1-mediated recombinantion by altering the conformation of the DNA strands at the target site. The Goldberg-Hogness and Pribnow consensus sequences are strikingly similar; perhaps the sites of integration of IS1 in this study are Drosophila promoter sites. Probing a Northern blot of *D. melanogaster* messenger RNA would elucidate whether the the two clones contain transcribed regions flanking the IS1. Another

possibility is that IS1 inserted into ribosomal RNA gene sequences in the Drosophila genomic library. These abundant sequences are AT-rich and, in pUCa1, one mitochondrial ribosomal gene has already served as a target for IS1 insertion (P. Moreau, personal communication). However, this mitochondrial ribosomal cDNA from pUCa1 and the two recombinant subclones do not hybridize to similar fragments of EcoRI digested genomic DNA (1 Kb and 7.6 Kb, respectively). This observation would suggest that the two subclones do not contain mitochondrial ribosomal DNA sequences. Accordingly, the ribosomal mitochondrial DNA from pUCa1 does not hybridize to sequences within the subclones under high stringency conditions (P. Moreau, personal communication).

Drosophila transposable elements also display a range of stringency requirements for target site. For example, the *white* locus is a 'hot' spot for P-element activity in germ-line chromosomes of Drosophila (Engels, 1980; Rubin & O'Hare, 1984). In addition, P-elements frequently sponsor deletion and rearrangement mutations at their target sites. Another characteristic common to bacterial insertion sequences and Drosophila transposable elements is that neither carry selectable genes. Identification of the target sites for the two IS1 insertions into *Drosophila melanogaster* DNA isolated during this study, may elucidate the evolutionary origin of a family of Drosophila mobile genetic elements. If the IS1 target sites in Drosophila are

similar in sequence or are derived from the same chromosomal locus, they may coincide with a site of frequent insertion for a particular mobile element of *Drosophila melanogaster*. The requirement for the same target site for prokaryotic and eukaryotic mobile elements may indicate a common evolutionary ancestor for both.

Eukaryotic mobile elements may be able to integrate into the prokaryotic genome but this could only occur if prokaryotic enzymes required for transposition could function on eukaryotic transposons. Functions required for transposition are encoded on the elements themselves and in eukaryotes coding regions usually contain introns. It would be interesting to test the possibility of converse transposition and such experiments could uncover clues as to mechanisms by which transposition occurs and the evolution of all mobile genetic elements.

The origin of the bacterial insertion sequence found in both recombinant Drosophila genomic clones is assumed to be the bacterial host in which the library was made or propagated. Southern blots demonstrate that three *E. coli* strains commonly used to propagate plasmids, bacteriophage, and cosmids contain elements highly homologous to IS1. Then, shouldn't the isolation of prokaryotic insertion sequences from eukaryotic DNA occur frequently? This thesis also serves as a warning to all of us who utilize the common tools and methods of molecular biology to be

aware of the inherent pitfalls using prokaryotic vectors and cells to propagate eukaryotic genes.

Alternatively, the insertion sequences may have originated in the eukaryotic DNA and may have already been present when the library was made. This hypothesis cannot be tested directly, however, the IS1 present in p2 does not hybridize to Oregon R sequences on our Southern transfers. This does not exclude the possibility that the Oregon R DNA source used to make the library harbours prokaryotic insertion-like sequences.

As unrealistic as this conjecture may appear, there are examples of prokaryotic and eukaryotic DNA being closely juxtaposed in which such a scenario may conceivably occur. For instance, nitrogen-fixing plants establish symbiotic relationships with particular bacterial strains in the soil enabling them to carry out nitrogen fixation. One such bacterial species, *Rhizobium japonicum*, harbours insertion-like elements clustered around the *nif* region of its genome (Kaluza, et al, 1985). Could these elements not establish themselves in plant DNA juxtaposed to the bacteroid? Similarily, many animal species rely upon prokaryotic populations within the gut for digestion of plant matter. It is not outside the realm of possibility that mobile elements from gut inhabitants could occasionally take up residence in eukaryotic DNA.

The frequency of IS1 translocation in *E. coli* is equivalent to the spontaneous rate of mutation (Starlinger &

Saedler, 1976). However, transposition frequency may be a characteristic of individual IS1 copies as some appear to be more active than others (Machida, et al, 1982). It is generally accepted that an equilibrium must exist for transposition between integration events and loss of insertion sequences already present in the genome. This equilibrium is maintained by the elements themselves as a control against their dominating the genome and thereby destroying the host organism and themselves as well (Iida, et al, 1983). This equilibrium may be disturbed by physiological factors or by introduction of a genome devoid of insertion elements into the cell. The new replicon may be colonized by mobile elements already present in the cell, hence plasmids and bacteriophage have been isolated which carry transposable elements once they have passed through the cell (Calos & Miller, 1978; Arber, et al, 1981; Campbell, 1981). Judging by the frequency of isolation of prokaryotic mobile elements from genomic DNA of eukaryotes, prokaryotic insertion elements enter eukaryotic DNA rarely or else go undetected. The latter case is probably more accurate.

Physiological conditions contribute to the rate of transposition. For instance, high temperature generally results in an increased rate of transposition perhaps because of the presence of a cold-sensitive repressor encoded by the IS element (Iida, et al, 1983). However, transposition of IS elements from the *E. coli* chromosome

appears to be induced in stationary phase cells especially when these are left standing in the refrigerator (Iida, et al, 1983). Refrigeration of a stock of cells for plating genomic or cDNA libraries is a common laboratory practise and lends credence to the proposal that more IS elements are present in eukaryotic genes isolated from libraries than is documented. In addition, changing the host strain used to propagate the library may lead to anomolous insertions of mobile prokaryotic elements (V. Seligy, personal communication). Generally, it appears that increased transposition of mobile elements is a response to stress in the cells.

This thesis began with the task of isolating the *Drosophila melanogaster* amylase gene. The results presented detail instead the isolation of IS1 from Drosophila genomic sequences. The isolation of the α-amylase gene is now a 'fait accompli' (Gemmill, et al, 1985; Levy, et al, 1985), while the insertion of IS1 into Drosophila genomic DNA has never been described. Whether characterization of this unique event continues or not, we have provided valuable information to the study of molecular biology. Firstly, this may not be an isolated incident and we should never take for granted that eukaryotic and prokaryotic genetic elements are necessarily separate, and distinguishible entities. By propagating eukaryotic genes in prokaryotic hosts, we take the risk of blurring the distinctions between

these two systems. Secondly, the transposition of IS1 into eukaryotic genomic sequences may provide a look back in evolution; from this point it may be possible to gain new understanding of the orlgins of all mobile genetic elements.

Chapter V

CONCLUSIONS

Two recombinant clones, L1231 and L1232, have been isolated from a *Drosophila melanogaster* genomic library constructed in the lambda vector EMBL4. These recombinant clones were isolated from the library using a probe harbouring bacterial insertion sequence IS1. The isolation and characterization of the sequences within the isolated recombinant clones are described in this thesis.

Hybridization of PstI subclones of the original lambda recombinants to *E. coli* genomic DNA demonstrates that a dispersed repetitive element of the *E. coli* genome is present in both subclones. Data resulting from restriction endonuclease digestion of the PstI subclones of L1231 and L1232 indicate that this *E. coli* genomic sequence is IS1. Further, one copy of IS1 is present in each of these clones. Crosshybridization of isolated fragments from each PstI subclone confirmed that the sequence generates restriction fragments and hybridization patterns predicted for IS1.

Fragments generated by restriction enzyme digests of the original lambda clones, L1231 and L1232, their subclones in pUC18, p1231B and p1232A, and smaller PstI fragment subclones in pBR322, p1231P and p1232P, resulted in the restriction maps in Figures 3.11, 3.15, and 3.25, respectively. These maps represent the best fit of the data

generated and coincide with the published restriction maps of IS1.

Using subcloned fragments from the two recombinant clones as probe, it was demonstrated that other sequences within the clones have homology to discrete Drosophila genomic sequences. Hybridization of the IS1 containing probe, p2, to Drosophila DNA shows that these genomic sequences are not homologous to IS1. In addition, pBR322 and EMBL4 DNA do not hybridize to Drosophila DNA. Therefore, we conclude that the recombinant clones, which we have shown to include IS1, also contain Drosophila genomic sequences. Comparison of restriction digest data generated from the two recombinant lambda clones and their subclones to digested Drosophila DNA, demonstrates that IS1 has integrated into Drosophila genomic sequences. Further, the IS1 element has inserted into two different target sites in each clone as demonstrated by distinct restriction enzyme maps and the fact that Drosophila sequences in each subclone do not cross-hybridize.

We have not identified the *Drosophila melanogaster* target sequences in the two recombinant clones. However, we can conclude that the Drosophila sequences are not mitochondrial ribosomal rDNA. However, it is possible that these sequences are transcribed regions. We feel that continuing investigations will identify these Drosophila sequences.

The origin of the these insertion sequences isolated from *Drosophila melanogaster* DNA is unknown. However, we have shown that our source of Oregon R *Drosophila melanogaster* DNA does not contain sequences homologous to IS1. In addition, we have demonstrated that three common strains of *E. coli*, C600, LE392, and Q358, contain sequences homologous to IS1. Therefore, indirect evidence indicates that the source of IS1 in both clones is one of these prokaryotic hosts. The number of library screenings required to isolate the two lambda recombinants further indicates that insertion of IS1 into Drosophila DNA is a rare event.

Abraham, I. and Doane, W.W. (1978) Genetic regulation of tissue-specific expression of Amylase structural genes in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA 75:4446-4450

Adams, J.M., Harris, A.W., Pinkert, C.A., Corcoran, L.M., Alexander, W.S., Cory, S., Palmiter, R.D. and Brinster, R.L. (1985) the c-myc oncogene driven by immunoglobulin enhancers induces lymphoid malignancy in transgenic mice. Nature (London) 318:533-538.

Amin, J., Mestril, R., Lawson, R., Klapper, H., and Voellmy, R. (1985) The Heat Shock Consensus Sequence Is Not Sufficient for hsp70 Gene Expression in *Drosophila melanogaster*. Mol. & Cell Biol. 5:197-203.

Arber, W., Humberlin, M., Caspers, P., Reif, H.J., Iida, S., and Meyer, J. (1981) Spontaneous Mutations in the *Escherichia coli* Prophage P1 and IS-Mediated Processes. Cold Spring Harbor Symp. Quant. Biol. 45:38-40.

Atchison, M.L. and Perry, R.P. (1986) Tandem Kappa Immunoglobulin Promoters Are Equally Active in the Presence of the Kappa Enhancer: Implications for Models of Enhancer Function. Cell 46:253-262.

Bahn, E. (1971) Cytogenetical localization of the Amylase region in *Drosophila melanogaster* by means of translocations. Hereditas 67:75-78

Bahn, E. (1980) Chromosomal region determining Amylase Isozymes in *Drosophila melanogaster*. Hereditas 58:1-12

Banerji, J., Olson, L., and Schaffner, W. (1983) A Lymphocyte-Specific Cellular Enhancer Is Located Downstream of the Joining Region in Immunoglobulin Heavy Chain Genes. Cell 33:729-740.

Banerji, J., Rusconi, S., and Schaffner, W. (1981) Expression of a β-Globin Gene Is Enhanced by Remote SV40 DNA Sequences. Cell 27:299-308

Benoist, C. and Chambon, P. (1981) In vivo sequence requirements of the SV40 early promoter region. Nature (London) 290:304-310

Benoist, C., O'Hare, K., Breathnach, R., and Chambon, P. (1980) The ovalbumin gene-sequence of putative control regions. Nucl. Acids Res. 8:127-142.

Benton, W.D. and Davis, R.W. (1977) Screening gt Recombinant Clones by Hybridization to Single Plaques in situ. Science 196:180-182

Benyajati, C. and Dray, J.F. (1984) Cloned Drosophila alcohol dehydrogenase genes are correctly expressed after transfection into Drosophila cells in culture. Proc. Natl. Acad. Sci. USA 81:1701-1705.

Benyajati, C., Spoerel, N., Haymerle, H., and Ashburner, M. (1984) The Messenger RNA for Alcohol Dehydrogenase in *Drosophila melanogaster* Differs in Its 5' End in Different Developmental Stages. Cell 38:125-133.

Berg, P.E., Yu, J.-K., Popovic, Z., Schumperli, D., Johansen, H., Rosenberg, M., and Anderson W.F. (1983) Differential Activation of the Mouse β-Globin Promoter by Enhancers. Molec. Cell. Biol. 3:1246-1254.

Bienz, M., and Pelham, H.R.B. (1986) Heat Shock Regulatory Elements Function as an Inducible Enhancer in the Xenopus hsp70 Gene and when Linked to a Heterlogous Promoter. Cell 45:753-760.

Bingham, P.M., Kidwell, M.G., and Rubin, G.M. (1982) The Molecular Basis of P-M Hybrid Dysgensis: The Role of the P Element, a P-Strain-Specific Transposon Family: Cell 29:995-1004

Birnboim, H.C. and Doly J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl. Acids Res. 7:1513-1523.

Bodary, S., Grossi, G., Hagenbuchle, O., and Wellauer, P. (1985) Members of the Amy-2 Alpha-amylase Gene Family of Mouse Strain CE/J Contain Duplicated 5' Termini. J. Mol. Bio. 182:1-10.

Bonner, T.I., Brenner, D.J., Neufeld, B.R., and Britten, R.J. (1973) Reduction in the Rate of DNA Reassociation by Sequence Divergence. J. Mol. Biol. 81:123-135.

Borrelli, E., Hen, R., and Chambon, P. (1984) Adenovirus-2 E1A products repress enhancer-induced stimulation of transcription. Nature (London) 312:608-612.

Braedt, G. (1985) Recombination in recA Cells Between Direct Repeats of Insertion Element IS1. J. Bacteriology, 162:529-534.

Brand, A.H., Breeden, L., Abraham, J., Sternglanz, R., and Nasmyth, K. (1985) Characterization of a "Silencer" in Yeast: A DNA Sequence with Properties Opposite to Those Of A Transcriptional Enhancer. Cell 41:41-48.

Bukhari, A.I. (1981) Models of DNA transposition. Trends in Biochemical Science. 6:56-60.

Caggese, C., Caizzi, R., Grieco, F., Bozzetti, M.R., and Ritossa, F. (1986) Genetic determinants of glutamine synthetase in *Drosophila melanogaster*: role of the 10B8-11 region. Molec. Gen. Genet. 204:208-213

Calos, M.P. and Miller, J.H. (1980) Transposable Elements. Cell 20:579-595.

Campbell, A. (1981) Some General Questions about Movable Elements and Their Implications. Cold Spring Harbor Symp. Quant. Biol. 45:1-9.

Carrasco, A.E., McGinnis, W., Gehring, W.J., and De Robertis, E. (1985) Cloning of an X. laevis Gene Expressed during Early Embryogenesis Coding for a Peptide Region Homologous to Drosophila Homeotic Genes. Cell 37:409-414

Chada, K., Magram, J., Raphael, K., Radice, G., Lacy, E., and Costantini, F. (1985) Specific expression of a foreign β-globin gene in erythroid cells of transgenic mice. Nature (London) 314:377-380.

Chadwell, H., Fritz, H.-J., Habermann, P., Klaer, R., Kuhn, S., and Starlinger, P. (1979) Studies with IS1 and IS4. Cold Spring Harbor Symp. Quant. Biol. 43:1187-1192.

Chandler, V.L., Maler, B.A., and Yamamoto, K.R. (1983) DNA Sequences Bound Specifically by Glucocorticoid Receptor In Vitro Render A Heterologous Promoter Hormone Responsive In Vivo. Cell 33:489-499.

Chao, A.T. and Guild, G.M. (1986) Molecular analysis of the ecdysterone-inducible 2B5 'early' puff in *Drosophila melanogaster*. EMBO J. 5:143-150.

Charnay, P., Mellon, P., and Maniatis, T. (1985) Linker Scanning Mutagenesis of the 5'-Flanking Region of the Mouse β-Major-Globin Gene: Sequence Requirements for Transcription in Erythroid and Nonerythroid Cells. Molec. Cell. Biol. 5:1498-1511.

Church, G.M., Ephrussi, A., Gilbert, W., and Tonegawa, S. (1985) Cell-type-specific contacts to immunoglobulin enhancers in nuclei. Nature (London) 313:798-801.

Ciliberto, G., Dente, L., and Cortese, R. (1985) Cell-Specific Expression of a Transfected Human α1-Antitrypsin Gene. Cell 41:531-540.

Colberg-Poley, A.M., Voss, S.D., Chowdhury, K., and Gruss, P. (1985) Structural analysis of murine genes containing homoeo box sequences and their expression in embryonal carcinoma cells. Nature (London) 314:713-718.

Cornelis, G. and Saedler, H. (1980) Deletions and an Inversion Induced by a Resident IS1 of the Lactose Transposon Tn951. Molec. Gen. Genet. 178:367-374.

Cowman, A.F., Zuker, C.S., and Rubin, G.M. (1986) An Opsin Gene Expressed in Only One Photoreceptor Cell Type of the Drosophila Eye. Cell 44:705-710

D'Onofrio, C., Colantuoni, V., Cortese, R. (1985) Structure and cell-specific expression of a cloned human retinol binding protein gene: the 5'-flanking region contains hepatoma specific transcriptional signals. EMBO J. 4:1981-1989.

Daneholt, B. (1975) Transcription in Polytene Chromosomes. Cell 4:1-9.

Davis, R.W., Botstein, D., and Rothe, J.R. (1980) A Manual for Genetic Engineering: Advanced Bacterial Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 254 pages.

Davison, B.L., Egly, J.-M., Mulvihill, E.R., and Chambon, P. (1983) Formation of stable preinitiation complexes between eukaryotic class B transcription factors and promoter sequences. Nature (London) 301:680-686

Dean, D.C., Knoll, B.J., Riser, M.E., and O'Malley, B.W. (1983) A 5'-flanking sequence essential for progesterone regulation of an ovalbumin fusion gene. Nature (London) 305:551-554.

Desplan, C., Theri, J., and O'Farrell, P.H. (1985) The Drosophila developmental gene, engrailed, encodes a sequence-specific DNA binding activity. Nature (London) 318:630-635.

deVilliers, J., Olson, L., Tyndall, C., and Schaffner, W., (1982) Transcriptional "enhancers" from SV40 and polyoma virus show a cell-type preference. Nucl. Acids Res. 10:7965-7976.

Devos, R., Contreras, R., van Emmelo, J., and Fiers, W. (1979) Identification of the Translocatable Element IS1 in a Molecular Chimera Constructed with Plasmid pBR322 into which a Bacteriophage MS2 DNA Copy was Inserted by the Poly(dA)-Poly(dT) Linker Method. J. Mol. Bio. 128:621-632.

Dierks, P., van Ooyen, A., Cochran, M.D., Dobkin, C., Reiser, J., and Weissman, C. (1983) Three Regions Upstream of the Cap Site Are Required for Efficient and Accurate Transcription of the Rabbit β-Globin Gene in Mouse 3T6 Cells. Cell 32:695-706

Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucl. Acids Res. 11:1475-1489.

DiNocero, P.P. and Dawid, I.B. (1983) Transient expression of genes introduced into cultured cells of Drosophila. Proc. Natl. Acad. Sci. USA 80:7095-7098.

Doane, W.W. (1969) Amylase Variants in *Drosophila melanogaster*: Linkage Studies and Characterization of Enzyme Extracts. J. Exp. Zool. 171:321-342.

Doane, W.W., Treat-Clemons, L.G., Gemmill, R.M., Levy, J.N., Hawley, S.A., Buchberg, A.M., and Paigen, K. (1983) Genetic Mechanism for Tissue-Specific Control of Alpha-Amylase Expression in *Drosophila melanogaster*. In Isozymes: Current Topics in Biological and Medical Research 9: Gene Expression and Development pp. 63-90 Alan R. Liss Inc., N.Y.

Dove, W.F. and Davidson, N. (1962) Cation Effects on the Denaturation of DNA. J. Mol. Biol. 5:467-478.

Dynan, W.S. and Tjian, R. (1985) Control of eukaryotic mesenger RNA synthesis by sequence-specific DNA-binding proteins. Nature (London) 316:774-778

Efstratiadis, A., Posakony, J.W., Maniatis, T., Lawn, R.M., O'Connell, C., Spritz, R.A., DeReil, J.K., Forget, B.G., Weissman, S.M., Slightom, J.L., Blechl, A.E., Smithies, O., Baralle, F.E., Shoulders, C.C., and Proudfoot, N.J. (1980) The Structure and Evolution of the Human β-Globin Gene Family. Cell 21:653-668.

Engels, W.R. (1983) The P family of transposable elements in Drosophila. Ann. Rev. Genet. 17:315-344.

Ephrussi, A., Church, G.M., Tonegawa, S., and Gilbert, W. (1985) B Lineage-Specific Interactions of an Immunoglobulin Enhancer with Cellular Factors in Vivo. Science 227:134-140.

Fernandez, C., Lanhammar, D., Servenius, B., Rask, L., and Peterson, P.A. (1986) Spontaneous insertions into cosmid vector: a warning. Gene 42:215-219.

Frischauf, A.-M., Lehrach, H., Poustka, A., and Murray, N. (1983) Lambda Replacement Vectors Carrying Polylinker Sequences. J. Mol. Biol. 170:827-842.

Fromm, M. and Berg, P. (1983) Simian Virus 40 Early- and Late-Region Promoter Functions are Enhanced by the 72-Base-Pair Repeat Inserted at Distant Locations and Inverted Orientations. Molec. Cell. Biol. 3:991-999.

Fuhrman, S.A., Van Beveren, C. and Verma, I.M. (1981) Identification of a RNA polymerase II initiation site in the long terminal repeat of Moloney murine leukemia viral DNA, Proc. Natl. Acad. Sci. USA 78:5411-5415

Fujita, Takashi, Shibuya, H., Ohashi, T., Yamanishi, K., and Taniguchi, T. (1986) Regulation of Human Interleukin-2 Gene: Functional DNA Sequences in the 5' Flanking Region for the Gene Expression in Activated T-Lymphocytes. Cell 46:401-407.

Galas, D.J. and Chandler, M. (1982) Structure and Stability of Tn9-mediated Cointegrates: Evidence for Two Pathways of Transposition. J. Mol. Bio. 154:245-272.

Gamas, P., Galas, D., and Chandler, M. (1985) DNA sequence at the end of IS1 required for transposition. Nature (London) 317:458-461.

Gariglio, P., Bellard, M., and Chambon, P. (1981) Clustering of RNA polymerase B molecules in the 5' moiety of the adult β-globin gene of hen erythrocytes. Nucl. Acids Res. 9:2589-2600.

Gemmill, R.M., Levy, J.N., and Doane, W.W. (1985) Molecular cloning of α-amylase genes from *Drosophila melanogaster*. I. Clone isolation by use of a mouse probe, Genetics 110:299-312

Gemmill, R.M., Schwartz, P.E., and Doane, W.W. (1986) Structural organization of the Amy locus in seven strains of *Drosophila melanogaster*. Nucl. Acids Res., 14:5337-5349.

Gerlinger, P., LeMeur, M., Irrmann, C., Renard, P., Wasylyk, C., and Wasylyk, B. (1986) B-lymphocyte targeting of gene expression in transgenic mice with the immunoblobulin heavy chain enhancer. Nucl. Acids Res. 14:6565-6575.

Gillies, S.D., Morrison, S.L., Oi, V.T., and Tonegawa, S. (1983) A Tissue-specific Transcription Enhancer Element Is Located in the Major Intron of a Rearranged Immunoglobulin Heavy Chain Gene. Cell 33:717--728.

Gilmour, D.S. and Lis, J.T. (1985) In Vivo Interactions of RNA Polymerase II with Genes of *Drosophila melanogaster*. Molec. Cell. Biol. 5:2009-2018.

Gilmour, D.S., Pflugfelder, G., Wang, J.C., and Lis, J.T. (1986) Topoisomerase I Interacts with Transcribed Regions in Drosophila Cells. Cell 44:401-407.

Giorgi, D., Bernard, J.P., Lapointe, R., and Dagorn, J.C. (1984) Regulation of amylase messenger RNA concentration in rat pancreas by food content. The EMBO Journal 3:1521-1524.

Goldberg, D.A., Posakony, J.W., and Maniatis, T. (1983) Correct Developmental Expression of a Cloned Alcohol Dehydrogenase Gene Transduced into the Drosophila Germ Line. Cell 34:59-73

Goodbourn, S., Burstein, H., and Maniatis, T. (1986) The Human ß-Interferon Gene Enhancer Is under Negative Control. Cell 45:601-610.

Goodbourn, S., Zinn, K., and Maniatis, T. (1985) Human ß-Interferon Gene Expression Is Regulated by and Inducible Enhancer Element. Cell 41:509-520.

Gorman, C.M., Rigby, P.W.J., and Lane, D.P. (1985) Negative Regulation of Viral Enhancers in Undifferentiated Embryonic Stem Cells. Cell 42:519-526.

Grosschedl, R., Weaver, D., Baltimore, D., and Costantini, F. (1984) Introduction of a µ Immunoglobulin Gene into the Mouse Germ Line: Specific Expression in Lymphoid Cells and Synthesis of Functional Antibody. Cell 38:647-658.

Grosschedl,R. and Baltimore, D. (1985) Cell-Type Specificity of Immunoglobulin Gene Expression is Regulated by at Least Three DNA Sequence Elements. Cell 41:885-897.

Hagenbuchle, O., Bovey, R., and Young, R.A. (1980) Tissue-Specific Expression of Mouse α-Amylase Genes: Nucleotide Sequence of Isoenzyme mRNAs from Pancreas and Salivary Gland. Cell 21:179-187.

Hagenbuchle, O., Schibler, U., Petrucco, S., Van Tuyle, G.C., and Wellauer, P.K. (1985) Expression of Mouse Amy-2a Alpha-amylase Genes is Regulated by Strong Pancreas-specific Promoters. J. Mol. Bio. 185:285-293.

Haj-Ahmed, Y. and Hickey, D.A. (1982) A molecular explanation of frequency-dependent selection in Drosophila. Nature (London) 299:350-352

Hammer, R.E., Palmiter, R.D., and Brinster, R.L. (1984) Partial correction of murine hereditary growth disorder by germ-line incorporation of a new gene. Nature (London) 311:65-67.

Hanahan, D. (1985) Heritable formation of pancreatic ß-cell tumours in transgenic mice expressing recombinant insulin/simian virus 40 oncogenes. Nature (London) 315:115-122.

Harding, J.D. and Rutter, W.J. (1978) Rat Pancreatic Amylase mRNA: Tissue Specificity and Accumulation During Embryonic Development. J. Biol. Chem. 253:8736-8740.

Hayday, A.C., Gillies, S.D., Saito, H., Wood, C., Wiman, K., Hayward, W.S., and Tonegawa, S. (1984) Activation of a translocated human c-*myc* gene by an enhancer in the immunoglobulin heavy-chain locus. Nature (London) 307:334-340.

Hayward, M.A., Brock, M.L., and Shapiro, D.J. (1982) Activation of vitellogenin gene transcription is a direct response to estrogen in *Xenopus laevis* liver. Nucl. Acids Res. 10:8273-8284.

Hearing, P. and Shenk, T. (1986) The Adenovirus Type 5 E1A Enhancer Contains Two Functionall Distinct Domains: One Is Specific for E1A and the Other Modulates All Early Units in *Cis*. Cell 45:229-236.

Heberlein, U., England, B., and Tjian, R. (1985) Characterization of Drosophila Transcription Factors That Activate the Tandem Promoters of the Alcohol Dehydrogenase Gene. Cell 41:965-977.

Hen, R., Borrelli, E., Sassone-Corsi, P., and Chambon, P. (1983) An enhancer element is located 3450 base paires upstream from the adenovirus-2 E1A capsite. Nucl. Acids Res. 11:8747-8760.

Henikoff, S., Keene, M.A., Fechtel, K., and Fristrom, J.W. (1986) Gene within a Gene: Nested Drosophila Genes Encode Unrelated Proteins on Opposite DNA Strands. Cell 44:33-42.

Hennighausen, L., Siebenlist, U., Danner, D., Leder, P., Rawlins, D., Rosenfeld, P., and Kelly, T., Jr. (1985) High-affinity binding site for a specific nuclear protein in the human IgM gene. Nature (London) 314:289-292.

Hickey (1981) Regulation of Amylase Activity in Drosophila melanogaster: Variation in the Number of Enzyme Molecules Produced by Different Amylase Genotypes. Biochemical Genetics 19:783-796.

Hickey, D.A. (1977) Selection for amylase allozymes in *Drosophila melanogaster*. Evolution 31:800-804.

Hiromi, Y., Okamoto, H., Gehring, W.J., and Hotta, Y.(1986) Germline Transformation with Drosophila Mutant Actin Genes Induces Constitutive Expression of Heat Shock Genes. Cell 44:293-301.

Hiyashi, S., Kondoh, H., Yasuda, K., Soma, G.-i., Ikawa, Y., and Okada, T.S. (1985) Tissue-specific regulation of a chicken δ-crystallin gne in mouse cells: involvement of the 5' end region. EMBO J. 4:2201-2207.

Hoffman, E. and Corces, V. (1986) Sequences Involved in Temperature and Ecydysterone-Induced Transcription Are Located in Separate Regions of a *Drosophila melanogaster* Heat Shock Gene. Mol. & Cell Biol. 6:663-673.

Hoorn, A.J.W., and Scharloo, W. (1981) Functional Significance of *Amylase* Polymorphism in *Drosophila melanogaster*.III. Ontogeny of *Amylase* and Some α-Glucosidases. Biochem. Genet. 18:51-63

Hultmark, D., Klemenz, R., and Gehring, W.J. (1986) Translational and Transcriptional Control Elements in the Untranslated Leader of the Heat-Shock Gene *hsp22*. Cell 44:429-438.

Hynes, N., van Ooyen, A.J.J., Kennedy, N., Herrlick, P., Ponta, H., and Groner, B. (1983) Subfragments of the large terminal repeat cause glucocorticoid responsive expression of mouse mammary tumor virus and of an adjacent gene. Proc. Natl. Acad. Sci. USA 80:3637-3641.

Iida, S., Meyer, J., and Arber, W. (1983) Prokaryotic IS Elements. In Mobile Genetic Elements. ed. J.A. Shapiro. Academic Press, N.Y. pp.159-221.

Johnsrud, L. (1979) DNA Sequence of the Transposable Element IS1. Molec. Gen. Genet. 169:213-218.

Kahn, M., Kolter, R., Thomas, C., Figurski, D., Meyer, R., Remaut, E., and Helinski, D.R. (1979) Plasmid Cloning Vehicles Derived from Plasmids ColE1, F, R6K, and RK2. *In Methods in Enzymology*. Ed. R. Wu, v. 68:268-269.

Kaluza, K., Hahn, M., and Hennecke, H. (1985) Repeated Sequences Similar to Insertion Elements Clustered Around the *nif* Region of the *Rhizobium japonicum* Genome. J. of Bacteriology 162:535-542.

Karin, M., Haslinger, A., Holtgreve, H., Richards, R.I., Krauter, P., Westphal, H.M., and Beato, M. (1984) Characterization of DNA sequences through which cadmium and glucocorticoid hormones induce human metallothionein-I. gene. Nature (London) 308:513-519.

Karn, J., Brenner, S., and Barnett, L. (1983) [1] New Bacteriophage Lambda Vectors with Positive Selection for Cloned Inserts. In Methods in Enzymology 101:3-19.

Karn, J., Brenner, S., Barnett, L., and Cesareni, G. (1980) Novel bacteriophage cloning vector. Proc. Natl. Acad. Sci. USA 77:5172-5176.

Kelly, R.B., Cozzarelli, N.R., Deutscher, M.P., Lehman, I.R., and Kornberg, A. (1970) Enzymatic Synthesis of Deoxyribonucleic Acid XXXII. Replication of Duplex Deoxyribonucleic Acid by Polymerase at a Single Strand Break. J. Biol. Chem. 245:39-45.

Khoury, G. and Gruss, P. (1983) Enhancer Elements. Cell 33:313-314

Kleckner, N. (1977) Translocatable Elements in Procaryotes. Cell 11:11-23.

Klemenz, R. and Gehring, W.J. (1986) Sequence Requirement for Expression of the *Drosophila melanogaster* Heat Shock Protein *hsp22* Gene during Heat Shock and Normal Development. Mol. & Cell Biol. 6:2011-2019.

Klemenz, R., Hultmark, D., and Gehring, W.J. (1985) Selective translation of heat shock mRNA in *Drosophila melanogaster* depends on sequence information in the leader. The EMBO Journal 4:2053-2060.

Kosche, K.A., Dobkin, C., and Bank, A. (1985) DNA sequences regulating human β globin gene expression. Nucl. Acids Res. 13:7781-7790.

Kovesdi, I., Reichel, R., and Nevins, J.R. (1986) Identification of a Cellular Transcription Factor Involved in E1A *Trans*-Activation. Cell 45:219-228.

Labarca, C. and Paigen, K. (1977) mRNA-directed synthesis of catalytically active mouse β-glucuronidase in *Xenopus* oocytes. Proc. Natl. Acad. Sci. USA 74:4462-4465.

Laimins, L., Homgren-Konig, M., and Khoury, G. (1986) Transcriptional "silencer" element in rat repetitive sequences associated with the rat insulin 1 gene locus. Proc. Natl. Acad. Sci. USA 83:3151-3155.

Laimins, L., Khoury, G., Gorman, C., Howard, B., and Gruss, P. (1982) Host-specific activation of transcription by tandem repeats from simian virus 40 and Moloney murine sarcoma virus. Proc. Natl. Acad. Sci. USA, 79:6453-6457.

Laimins, L., Tsichlis, P., and Khoury, G. (1984) Multiple enhancer domains in the 3' terminus of the Prague strain of Rous sarcoma virus. Nucl. Acids Res. 12:6527-6528.

Lawn, R.M., Fritsch, E.F., Parker, R.C., Blake, G., and Maniatis, T. (1978) The Isolation and Characterization of Linked δ- and β-Globin Genes from a Cloned Library of Human DNA. Cell 15:1157-1174.

Lawson, R., Mestril, R., Schiller, P., and Voellmy, R. (1984) Expression of heat shock-β-galactosidase hybrid genes in cultured Drosophila cells. Molec. Gen. Genet. 198:116-124.

Levy, J.N., Gemmill, R.M., and Doane, W.W. (1985) Molecular cloning of α-amylase genes from *Drosophila melanogaster*. II. Clone organization and verification. Genetics 110:313-324

Machida, Y., Machida, C., and Ohtsubo, E. (1984) Insertion Element IS1 Encodes Two Structural Genes Required for its Transposition. J. Mol. Bio. 177:229-245.

Machida, Y., Machida, C., Ohtsubo, H., and Ohtsubo, E. (1982) Factors determining frequency of plasmid cointegraction mediated by insertion sequence IS1. Proc. Natl. Acad. Sci. USA, 79:277-281.

Mackay, R.M., Baird, S., Dove, M.J., Erratt, J.A., Gines, M., Moranelli, F., Nasim, A., Willick, G.E., Yaguchi, M., and Seligy, V.L. (1985) Glucanase Gene Diversity in Prokaryotic and Eukaryotic Organisms. BioSystems 18:279-292.

Maeda, H., Kitamura, D., Kudo, A., Araki, K., and Watanabe, T. (1986) *Trans*-acting Nuclear Protein Responsible for Induction of Rearranged Human Immunoglobulin Heavy Chain Gene. Cell 45:25-33.

Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) Molecular cloning--A laboratory manual. Cold Spring Harbor Laboratory. Cold Spring Harbor, N.Y.

Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C., and Quon, D. (1978) The Isolation of Structural Genes from Libraries of Eucaryotic DNA. Cell 15:687-701.

Manley, J.L., Fire, A., Cano, A., Sharp, P.A., and Gefter, M.L. (1980) DNA-dependent transcription of adenovirus genes in a soluble whole-cell extract. Proc. Natl. Acad. Sci. USA 77:3855-3859

Mason, J.O., Williams, G.T., and Neuberger, M.S. (1985) Transcription Cell Type Specificity Is Conferred by an Immunoglobulin V_κ Gene Promoter That Includes a Functional Consensus Sequence. Cell 41:479-487.

Mathis, D.J. and Chambon, P. (1981) The SV40 early region TATA box is required for accurate in vitro initiation of transcription. Nature (London) 290:310-315

Mattox, W.W. and Davidson, N. (1984) Isolation and Characterization of the *Beadex* Locus of *Drosophila melanogaster*: a Putative *cis*-acting Negative Regulatory Element for the *heldup-a* Gene. Mol. & Cell Biol. 4:1343-1353.

McGinnis, W., Garber, R.L., Wirz, J., Kuroiwa, A., and Gehring, W.J. (1984) A Homologous Protein-Coding Sequence in Drosophila Homeotic Genes and Its Conservation in Other Metazoans. Cell 37:403-408

Meinkoth , J. and Wahl, G. (1984) Hybridization of Nucleic Acids Immobilized on Solid Supports. Anal. Biochem. 138:267-284.

Mercola, M., Goverman, J., Mirell, C., and Calame, K. (1985) Immunoglobulin Heavy-Chain Enhancer Requires One or More Tissue-Specific Factors. Science 227:266-270.

Miyamoto, N.G., Moncollin, V., Egly, J.M., and Chambon, P. (1985) Specific interaction between a transcription factor and the upstream element of the adenovirus-2 major late promoter. EMBO J. 4:3563-3570.

Mohun, T.J., Brennan, S., Dathan, N., Fairman, S., and Gurdon, J.B. (1984) Cell type-specific activation of actin genes in the early amphibian embryo. Nature (London) 311:716-721.

Moreau, P., Hen, R., Wasylyk, B., Gaub, M.P., and Chambon, P. (1981) The SV40 72 base pair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. Nucl. Acids Res. 9:6047-6068

O'Hare, K. and Rubin, G.M. (1983) Stuctures of P Transposable Elements and Their Sites of Insertion and Excision in the *Drosophila melanogaster* Genome. Cell 34:25-35.

Ohlsson, H. and Edlund, T. (1986) Sequence-Specific Interactions of Nuclear Factors with the Insulin Gene Enhancer. Cell 45:35-44.

Ohtsubo, H. and Ohtsubo, E. (1978) Nucleotide sequence of an insertion element, IS1. Proc. Natl. Acad. Sci. USA 75:615-619.

Ornitz, D.M., Palmiter, R.D., Hammer, R.E., Brinster, R.L., Swift, G.H., and MacDonald, R.J. (1985) Specific expression of an elastase-human growth hormone fusion gene in prancreatic acinar cells of transgenic mice. Nature (London) 313:600-602.

Osborne, T.F., Arvidson, D.N., Tyau, E.S., Dunsworth-Browne, M., and Berk, A.J. (1984) Transcription Control Region Within the Protein-Coding Portion of Adenovirus E1A Genes. Molec. Cell. Biol. 4:1293-1305

Palmiter , R.D., Brinster, R.L., Hammer, R.E., Trumbauer, M.E., Rosenfeld, M.G., Birnberg, N.C., and Evans R M. (1982) Dramatic growth of mice that develop from eggs microinjected with metallothionein-growth hormone fusion genes. Nature (London) 300:611-615.

Parker, C.S. and Topol, J. (1984a) A Drosophila RNA Polymerase II Transcription Factor Contains a Promoter-Region-Specific DNA-Binding Activity. Cell 36:357-369.

Parker, C.S. and Topol, J. (1984b) A Drosophila RNA Polymerase II Transcription Factor Binds to the Regulatory Site of an hsp 70 Gene. Cell 37:273-283.

Payvar, F., DeFranco, D., Firestone, G.L., Edgar, B., Wrange, O., Okret, S., Gustafsson, J.-A., and Yamamoto, K.R. (1983) Sequence-Specific Binding of Glucocorticoid Receptor to MTV DNA at Sites within and Upstream of the Transcribed Region. Cell 35:381-392.

Picard, D. and Schaffner, W. (1984) A lymphocyte-specific enhancer in the mouse immunoglobulin K gene. Nature (London) 307:80-82.

Prentki, P., Karch, F., Iida, S., and Meyer, J. (1981) The plasmid cloning vector pBR325 contains a 482 base-pair-long inverted duplication. Gene 14:289-299.

Queen, C. and Baltimore, D. (1983) Immunoglobulin Gene Transcription Is Activated by Downstream Sequence Elements. Cell 33:741-748.

Reed, K.C. and Mann, D.A. (1985) Rapid transfer of DNA from agarose gels to nylon membranes. Nucl. Acids Res. 13:7207-7220.

Reif, H.J. and Arber, W. (1980) Analysis of Transposition of IS1-*kan* and Its Relatives. Cold Spring Harbor Symp. Quant. Biol. 45:40-43.

Reif, H.J. and Saedler, H. (1975) IS1 is Involved in Deletion Formation in the *gal* Region of E. coli K12. Molec. Gen. Genet. 137:17-28.

Renkawitz, R., Schutz, G., von der Ahe, D., and Beato, M. (1984) Sequences in the Promoter Region of the Chicken Lysozyme Gene Required for Steriod Regulation and Receptor Binding. Cell 37:503-510

Richards, G., Cassab, A., Bourouis, M., Jarry, B., and Dissous, C. (1983). The normal developmental degulation of a cloned *sgs3* 'glue' gene chromosomally integrated in *Drosophila melanogaster* by P element transformation. EMBO J. 2:2137-2142

Rigby, P.W.J., Dieckmann, M., Rhodes, C., and Berg, P. (1977) Labeling Deoxyribonucleic Acid to High Specific Activity *in Vitro* by Nick Translation with DNA Polymerase I. J. Molec. Biol. 113:237-251.

Rio, D.C. and Rubin, G.M. (1985) Transformation of Cultured *Drosophila melanogaster* Cells with a Dominant Selectable Marker. Mol. & Cell Biol. 5:1833-1838.

Rosen, C.A., Sodoroski, J.G., and Haseltine, W.A. (1985) the Location of *Cis*-Acting Regulatory Sequences in the Human T Cell Lymphotropic Virus Type III (HTLV-III/LAV) Long Terminal Repeat. Cell 41:813-823.

Rubin, G.M. and Spradling, A.C. (1982) Genetic Transformation of *Drosophila* with Transposable Element Vectors. Science 218:348-353.

Rubin, G.M. and Spradling, A.C. (1983) Vectors for P element-mediated gene transfer in *Drosophila*. Nucl. Acids Res. 11:6341-6351

Saedler, H. and Heiss, M. (1973) Multiple Copies of the Insertion-DNA Sequences IS1 and IS2 in the Chromosome of *Escherichia coli* K12. Mol. Gen. Genet. 122:267-277.

Saedler, H., Kubai, D.F., Nomura, M., and Jaskunas, S.R. (1975) IS1 and IS2 Mutations in the Ribosomal Protein Genes of E. coli K-12. Molec. Gen. Genet. 141:85-89.

Sassone-Corsi, P. and Borrelli, E. (1986) Transcriptional regulation by *trans*-acting factors. Trends in Genetics 2:215-219

Sato, G. and Miura, Y. (1963) Preparation of transforming deoxyribonucleic acid by phenol treatment. Biochim. Biophys. Acta 72:619-629.

Sawadogo, M. and Roeder, R.G. (1985) Interaction of a Gene-Specific Transcription Factor with the Adenovirus Major Late Promoter Upstream of the TATA Box Region. Cell 43:165-175.

Schibler, U., Pittet, A.-C., Young, R.A., Hagenbuchle, O. Tosi, M., Gellman, S., and Wellauer, P.K. (1982) The Mouse α-Amylase Multigene Family: Sequence Organization of Members Expressed in the Pancreas, Salivary Gland, and Liver. J. Mol. Bio. 155:247-266.

Schibler, U., Tosi, M., Pittet, A.-C., Fabiani, L., and Wellauer, P.K. (1980) Tissue-specific Expression of Mouse α-Amylase Genes. J. Mol. Bio. 142:93-116.

Schmidt, F., Besemer, J., and Starlinger, P. (1976) The Isolation of IS1 and IS2 DNA. Molec. Gen. Genet. 145:145-154.

Scholer, H.R. and Gruss, P. (1983) Specific Interaction between Enhancer-Containing Molecules and Cellular Components. Cell 36:403-411.

Scholnick, S.B., Morgan, B.A., and Hirsh, J. (1983) The Cloned Dopa Decarboxylase Gene Is Developmenatlly Regulated when Reintegrated into the Drosophila Genome. Cell 34:37-45.

Serfling, E., Jasin, M., and Schaffner, W. (1985) Enhancers and eukaryotic gene transcription. Trends in Genetics 1:224-230.

Shaw, P., Sordat, B., and Schibler, U. (1985) The Two Promoters of the Mouse α-Amylase Gene *Amy-1* Are Differentially Activated during Parotid Gland Differentiation. Cell 40:907-912.

Skuse, G.R. and Sullivan, D.T. (1985) Developmentally regulated alternated modes of expression of the *Gpdh* locus of Drosophila. The EMBO Journal 4:2275-2280.

Southern, E.M. (1975) Detection of Specific Sequences Among DNA Fragmenst Separated by Gel Electrophoresis. J. Molec. Biol. 98:503-517.

Spradling, A.C. and Rubin, G.M. (1983) The Effect of Chromosomal Position on the Expression of the Drosophila Xanthine Dehydrogenase Gene. Cell 34:47-57

Spradling, A.C. and Rubin, G.M. (1982) Transpostion of Cloned P Elements into *Drosophila* Germ Line Chromosomes. Science 218:341-347

Starlinger, K. and Saedler, H. (1976) IS-elements in Micro-organisms. Current Top. Microbiol. Immunol. 75:111-153.

Swift, G.H., Hammer, R.E., MacDonald, R.J., and Brinster, R.L. (1984) Tissue-Specific Expression of the Rat Pancreatic Elastase I Gene in Transgenic Mice. Cell 38:639-646.

Takahashi, K., Vigneron, M., Matthes, H., Wildeman, A., Zenke, M., and Chambon, P. (1986) Requirement of stereospecific alignments for initiation from the simian virus 40 early promoter. Nature (London) 319:121-126

Thomas, P.S. (1980) Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc.Natl. Acad. Sci. USA 77:5201-5205.

Townes, T.M., Chen, H.Y., Lingrel, J.B., Palmiter, R.D., and Brinster, R.L. (1985) Expression of Human β-Globin Genes in Transgenic Mice: Effects of a Flanking Metallothionein-Human Growth Hormone Fusion Gene. Molec. Cell. Biol. 5:1977-1983.

Treat-Clemons, L.G. and Doane, W.W. (1980) Control of amylase levels by the *trans*-acting, tissue-specific *map* locus in *Drosophila melanogaster*. Genetics 94:s105.

Ucker, D.S., Firestone, G.L., and Yamamaoto, K.R. (1983) Glucocorticoids and Chromosomal Position Modulate Murine Mammary Tumor virus Transcription by Affecting Efficiency of Promoter Utilfzation. Molec. Cell. Biol. 3:551-561.

Voss, S.D., Schlokat, U., and Gruss, P. (1986) The role of enhancers in the regulation of cell-type-specific transcriptional control. Trends in Biochem. Sci. 11:287-290.

Wahl, G.M., Stern, M., and Stark, G.R. (1979) Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl paper and rapid hybridization by using dextran sulfate. Proc. Natl. Acad. Sci. USA 76:3683-3689.

Walker , M.D., Edlund, T., Boulet, A.M., and Rutter, W.J. (1983) Cell-specific expression controlled by the 5'-flanking region of insulin and chymotrypsin genes. Nature (London) 306:557-561.

Wasylyk, B., Wasylyk, C., and Chambon, P. (1984) Short and long range activation by the SV40 enhancer. Nucl. Acids Res. 12:5589-5608.

Wasylyk, B., Wasylyk, C., Augereau, P., and Chambon, P. (1983) The SV40 72 bp Repeat Preferentially Potentiates Transcription Starting from Proximal Natural or Substitute Promoer Elements. Cell 32:503-514.

Weiher, H., Konig, M., and Gruss, P. (1983) Multiple Point Mutations Affecting the Simian Virus 40 Enhancer. Science 219:626-6631.

Weil, P.A., Segall, J., Harris, B., Ng, S.-Y., and Roeder, R.G. (1979) Faithful Transcription of Eukaryotic Genes by RNA Polymerase III in Systems Reconstituted with Purified DNA Templates. J. Biol. Chem. 254:6163-6173

Weinberger, J., Baltimore, D., and Sharp, P.A. (1986) Distinct factors bind to apparently homologous sequences in the immunoglobulin heavy-chain enhancer. Nature (London) 322:846-848.

Wu, C. (1985) An exonuclease protection assay reveals heat-shock element and TATA box DNA-binding proteins in crude nuclear extracts. Nature (London) 317:84-87.

Yoshimura, F.K., Davison, B., and Chaffin, K. (1985) Murine Leukemia Virus Long Terminal Repeat Sequences Can Enhance Gene Activity in a Cell-type-Specific Manner. Molec. Cell. Biol. 5:2832-2835. Young, K., Donovan-Peluso, M., Cubbon, R., and Bank, A. (1985) Trans acting regulation of ß--globin gene expression in erythroleukemia (K562) cells. Nucl. Acids Res. 14:5203-5210.

Zinn, D., DiMaio, D., and Maniatis, T. (1983) Identification of Two Distinct Regulatory Regions Adjacent to the Human ß-Interferon Gene. Cell 34:865-879.