Structure and Expression of the *singed* locus in *Drosophila melanogaster*.

Jamie Paterson

Dept of Biochemistry,
Imperial College of Science, Technology and Medicine,
London. SW7 2AZ

Submitted for the Degree of Ph.D.

April 1990

Abstract

Mutations at the singed (sn) locus of Drosophila melanogaster result in gnarled, twisted bristles and kinked, wavy hairs. In addition to this cuticular phenotype some alleles disrupt oogenesis resulting in female sterility.

Using DNA blotting techniques lesions in singed mutants were mapped onto the interval cloned by Roiha et al Genetics 119:75-83,(1988). These changes suggested that at least part of the gene was encoded by this interval. The region of cloned DNA was extended and the gene localised using cDNA hybridisation and RNA blotting techniques. Genomic and cDNA sequence data were then used to define the precise structure of the gene. Searching sequence data banks with both nucleic acid and predicted protein sequences found no homologous entries.

RNA blotting experiments were used to show the developmental profile of singed transcription. A combination of genetics and RNA blotting was then used to demonstrate the ovarian restriction of two particular transcripts. Similar methods were employed to prove maternal inheritance of singed transcripts.

Tissue specific expression of singed was investigated by generating P[lacZ] transposon insertions into the gene. Reporter gene B-galactosidase staining is restricted to later stages of oogenesis and persists into embryogenesis. This pattern of staining is consistent with the phenotypic disruptions seen in mutant ovaries.

Expression of singed RNAs has been examined in the following alleles : sn^{X2}, sn^{3}, sn^{w}, sn^{e} and $sn^{(+)}$. The allele sn^{3} has a deletion which removes the wild-type transcription start site. Transcripts found in this mutant start in the upstream interval now juxtaposed to the truncated first exon and appear to be directed by a cryptic promoter active only in this allele. Examination of sn^{X2} shows that this allele produces abnormally large sn transcripts. The three related P-element insertion mutants, sn^{w}, sn^{e} and $sn^{(+)}$ show a complex pattern of transcription. RNA molecules representing insertion readthrough transcripts, truncated transcripts as well as wildtype sized transcripts are detected. The wildtype sized RNAs are thought to be a product of splicing events which have removed P-element sequences from the longer readthrough transcripts. These P-element alleles also show a P-M cytotype dependence to their level of transcription.

Acknowledgements

I would like to thank my supervisor, Dr. Kevin O'Hare, for help and guidance during the course of this Ph.D. I would also like to thank the following for the technical assistance Alan Driver, Andy Mitchelson, Mark Tudor and Will Whitfield. Finally thanks are due to all members of the six floor both past and present for providing a stimulating and enjoyable work environment.

Contents

	Page
Abstract	2
Acknowledgements	3
Contents	4
List of Figures	8
List of Abbreviations	11
Chapter 1 - Introduction	12
1.0 *Drosophila* and genetics	13
1.1 *Drosophila* and molecular genetics	13
1.2 The lifecycle of *Drosophila melanogaster*	14
1.3 *Drosophila* oogenesis	16
1.4 The *singed* locus of *Drosophila melanogaster*	22
1.5 P-M hybrid dysgenesis and P-elements	26
1.6 P-element insertion and rearrangements at *singed*	28
1.7 Aims and objectives	31
Chapter 2 - Materials and Methods	33
2.1 Bacterial strains	34
2.2 Bacterial media	34
2.3 Preparation of DNA	35
2.3.1 Small scale preparation of plasmid DNA	35
2.3.2 Large scale preparation of plasmid DNA	35
2.3.3 Large scale preparation of lambda DNA	36
2.3.4 Preparation of *Drosophila* DNA	37
2.4 Preparation of *Drosophila* poly-A^+ RNA	37
2.5 Ethanol precipitation of nucleic acids	38
2.6 Phenol extraction of DNA	39
2.7 Enzymatic digestion of DNA	39
2.7.1 Endonuclease digestion of DNA	39
2.7.2 Exonuclease digestion of DNA	39
2.8 Ligation of DNA	40

2.9 Agarose gel electrophoresis 40

2.9.1 Neutral gels 40
2.9.2 Formaldehyde gels 41

2.10 Purification of DNA fragments from agarose gels 41

2.11 Nucleic acid transfer and filter hybridisation 42

2.11.1 DNA transfer and filter hybridisation 42
2.11.2 RNA transfer and filter hybridisation 43

2.12 Radioactive labelling of DNA 43

2.12.1 Labelling duplex DNA 43
2.12.2 Labelling single stranded DNA 44

2.13 Construction of genomic lambda libraries 44

2.14 Subcloning DNA fragments 46

2.14.1 DNA vectors 46
2.14.2 Identification of subclones 46
2.14.3 Plaque hydridisation 46
2.14.4 Colony hybridisation 47

2.15 Dideoxy sequencing 47

2.16 Computer analysis of sequence data 47

2.17 Detection of β-galactosidase activity in Drosophila tissue 48

2.17.1 β-galactosidase staining in whole mount embryos 48
2.17.2 β-galactosidase staining in pupae 48
2.17.3 β-galactosidase staining in ovarioles 48

2.18 Microscopy 49

2.19 Drosophila media and stocks 49

2.19.1 Drosophila media 49
2.19.2 Oregon-R population cages 49
2.19.3 Drosophila stocks 50

Chapter 3 - Molecular characterisation of the singed gene 53

3.0 Introduction 54

3.1 DNA blotting analysis of the singed region 54

3.2 Transcription of the singed region 65

3.3 Extent of the singed transcription unit - RNA analysis 68

3.4 Extent of the singed transcription unit -DNA analysis 71

3.5 Structural analysis of the singed gene 78

3.6 Concluding remarks 87

Chapter 4 - Tissue specific expression of singed 89

4.0 Introduction 90

4.1 Adult female 3.0kb and 3.3kb RNAs are ovary specific 91

4.2 Maternal inheritance of singed transcripts 94

4.3 Enhancer detector systems: P[lacZ] mutagenesis 97

4.4 Enhancer detector systems: Pupal P[lacZ] staining 103

4.5 Enhancer detector systems: Ovarian P[lacZ] staining 108

4.6 Enhancer detector systems: Maternally inherited $sn^{P[lacZ]}$ 117

4.7 Enhancer detector systems: Paternally inherited $sn^{P[lacZ]}$ 117

4.8 Concluding remarks 120

Chapter 5 - Molecular analysis of mutant alleles of singed 124

5.0 Introduction 125

5.1 Molecular characterisation of sn^{X2} and sn^{3} 129

5.1.1 Analysis of the sn^{X2} mutation 130
5.1.2 Analysis of the sn^{3} mutation 138

5.2 Introduction to mutant RNA blotting experiments 141

5.3 RNA blotting analysis of sn^{3} 148

5.4 RNA blotting analysis of sn^{e} 156

5.5 RNA blotting analysis of $sn^{(+)}$ 166

5.6 RNA blotting analysis of sn^{w} 172

5.7 Effect of cytotype on P-element alleles of sn 178

5.8 Concluding remarks 181

Chapter 6 - General Discussion 186

6.0 Introduction 187

6.1 What is the role of sn in germline and somatic cells ? 187

6.2 How do mutant alleles disrupt sn gene function ? 190

6.3 How does cytotype effect sn gene expression ? 191

6.4 Approaches to further work 191

Appendix 194

Figure A1 195
Figure A2 197
Figure A3 198

Literature cited 199

List of Figures

Page

Figure 1.3.1 - Oogenesis : early stages 18

Figure 1.3.2 - Oogenesis : late stages 20

Figure 1.4.1 - *sn*$^{-}$ phenotypes 24

Figure 1.6.1 - DNA map of the *sn* interval. 30

Figure 3.1.1 - Genetic and DNA map of the *singed* region. 56

Figure 3.1.2 - Table of *sn* mutants analysed by DNA blotting 57

Figure 3.1.3 - DNA blotting analysis of Dp(1:3)*sn*^{13a1} 59

Figure 3.1.4 - Detection of an DNA lesion in *sn*^{2} 61

Figure 3.1.5 - Identification of an RFLP within the *sn* locus 64

Figure 3.2.1 - Expression of *singed* transcripts during development. 67

Figure 3.3.1 - Localisation of the 3'ends of *sn* transcripts 70

Figure 3.4.1 - Alignment of cDNA maps 74

Figure 3.4.2 - Intergenic region between *sn* and *olfE* 77

Figure 3.5.1 - Physical maps of the 3 *sn* RNAs 80

Figure 3.5.2 - Sequence analysis of the *singed* gene. 82

List of Figures (continued)

Page

Figure 4.1.1 - 3.0 and 3.3kb RNAs are ovary dependent. 93

Figure 4.2.1 - Maternal inheritance of singed RNAs. 96

Figure 4.3.1 - P[lacZ] mutagenesis crossing scheme 100

Figure 4.3.2 - Out crossing scheme for $sn^{P[lacZ]}$ mutants 102

Figure 4.4.1 - Pupal $sn^{P[lacZ]}$ staining : early stages 105

Figure 4.4.2 - Pupal $sn^{P[lacZ]}$ staining : late stages 107

Figure 4.5.1 - Ovarian $sn^{P[lacZ]}$ staining : whole mount 110

Figure 4.5.2 - Ovarian $sn^{P[lacZ]}$ staining : early stages 112

Figure 4.5.3 - Ovarian $sn^{P[lacZ]}$ staining : border cell migration 114

Figure 4.5.4 - Ovarian $sn^{P[lacZ]}$ staining : late stages 116

Figure 4.7.1 - Embryo $sn^{P[lacZ]}$ staining : paternal 119

List of Figures (continued)

Page

Figure 5.0.1 - DNA maps of mutant alleles of sn chosen for transcriptional analysis. 127

Figure 5.1.1 - Sequence alignment between sn^{X2} and sn^{+} 133

Figure 5.1.2 - RNA blotting analysis of sn^{X2} 135

Figure 5.1.3 - DNA blot of sn^{X2} and parental chromosome 137

Figure 5.1.4 - Sequence alignment between sn^{3} and sn^{+} 140

Figure 5.2.1 - RNA blots of pupal sn^{3}, sn^{W}, sn^{e} and $sn^{(+)}$ samples 143

Figure 5.2.2 - RNA blots of adult male sn^{3}, sn^{W}, sn^{e} and $sn^{(+)}$ samples 145

Figure 5.2.3 - RNA blots of adult female sn^{3}, sn^{W}, sn^{e} and $sn^{(+)}$ samples 147

Figure 5.3.1 - Summary diagram of sn^{3} RNA blotting data. 150

Figure 5.3.2 - Expression of sn^{3} transcripts during development. 155

Figure 5.4.1 - Summary diagram of sn^{e} RNA blotting data. 158

Figure 5.4.2 - sn^{+} splice donor and acceptor site analysis. 162

Figure 5.4.3 - sn^{e} splice donor and acceptor site analysis. 164

Figure 5.5.1 - Summary diagram of $sn^{(+)}$ RNA blotting data. 168

Figure 5.5.2 - $sn^{(+)}$ splice donor and acceptor site analysis. 170

Figure 5.6.1 - Summary diagram of sn^{W} RNA blotting data. 174

Figure 5.6.2 - sn^{W} splice donor and acceptor site analysis. 176

Figure 5.7.1 - Cytotype dependent transcription of sn^{W}, sn^{e} and $sn^{(+)}$. 180

LIST OF ABBREVIATIONS

AMPS	ammonium persulphate
bp	nucleotide base pair
cDNA	complementary DNA
Ci	Curie
DEPC	diethylpyrocarbonate
dpm	disintegrations per minute
DTT	dithiothreitol
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
EDTA	ethylenediamine tetra-acetic acid
F_1	first filial generation
F_2	second filial generation
IPTG	isopropyl β-D-thiogalactopyranoside
kb	kilobase
kD	kiloDalton
M	molar
l	litre
mol	mole
PEG	polyethylene glycol
pers.comm.	personal communication
pH	negative decimal lograthim of hydrogen ion concentration in moles per litre.
poly-A^+ RNA	polyadenylated RNA
RNA	ribonucleic acid
RNase	ribonuclease
rpm	revolutions per minute
SDS	sodium dodecyl sulphate
SSC	standard saline citrate
SET	standard EDTA Tris
TE(10:1)	Tris.HCl pH7.5(10mM),EDTA(1mM)
TEMED	Tetramethylethylene diamine
TM10	Tris.HCl pH7.5(10mM),$MgCl_2$(10mM)
Tris	tris(hydroxymethyl) methylamine
X-Gal	5-bromo-4-chloro-3-indolyl β-D-galactopyranoside
~	approximately
%	percent (weight per volume unless stated otherwise)

CHAPTER 1

Introduction

1.0 Drosophila and genetics

The basic principles of inheritance were described in 1866 by an Augustinian Monk, Gregor Johann Mendel. The experimental organism used in these studies (Mendel,1866) was a variety of pea, Pisum sativuum. This paper presented conclusions derived from eight years of plant breeding experiments. At the time of publication few people read or understood the paper. However by 1900 interest and comprehension of Mendelian genetics had grown and it became clear that more convenient experimental organisms were required. About 1911, a professor at Columbia University T.H.Morgan began genetic studies on a small fruit-fly, Drosophila melanogaster. This organism was particularly well suited for genetic studies as it bred rapidly, was easy to handle and produced many offspring. Drosophila also turned out to be genetically simple, having only 4 pairs of chromosomes. Over the next 79 years many scientists have studied this organism and Drosophila melanogaster is now one of the most thoroughly characterised multicellular organisms used in genetic research.

1.1 Drosophila and molecular genetics

By the late 1960s genetics had progressed to the stage where mechanisms of gene regulation were being elucidated and the gene itself was being subdivided into different functional regions (reviewed in Miller and Reznikoff, 1978). Genetics and biochemistry were being used in unison to investigate gene structure and function. DNA had been identified as the genetic material (Avery, M^{C}Leod and M^{C}Carthy, 1944) and its double helical structure had been defined (Watson and Crick,1953). Molecular geneticists also knew how information was stored within the structure of DNA and the code which encrypts this information was known (reviewed in Roth,1974). The

advent of recombinant DNA technology and DNA sequencing techniques in the late 1970s allowed molecular geneticists to address the question : how do genes work?

In Drosophila these techniques were used to investigate fundamental questions about gene regulation and gene function. These studies are still on-going but the combination of genetics and molecular biology has already proven itself to be an extremely powerful method of investigating gene action.

1.2 The lifecycle of Drosophila melanogaster

One of the initial attractions of Drosophila, as an experimental organism, was the fact that fruit flies take only 10 to 20 days to complete one lifecycle in laboratory conditions. This allowed geneticists to observe several generations over a relatively short period of time. Nearly all stages of the Drosophila lifecycle are open to study and biologists have described the structure of each developmental stage in detail. Mutations disrupt the morphology and function of many tissues in the fly thus geneticists regularly consult descriptions of the Drosophila lifecycle.

The egg of Drosophila melanogaster is cylindrical, about 0.5 millimetres long by 0.3 millimetres wide. The unfertilised oocyte consists of a female germ cell encased within a protective layer of chorion and follicle cells. Fertilisation by a spermatocyte results in a complex but characterised series of nuclear divisions, cell divisions and cell differentiation events. Embryogenesis lasts about 24 hours at 25^{o}C (Bodenstein, 1950) and results in a first instar larvae which hatches out of the chorion and is able to forage for food. The larvae grow, moulting their first cuticle after 48 hours at 25 ^{o}C. Second instar larva have characteristically altered foraging behaviour but continue to feed and increase in body mass. A final cuticle moult occurs after approximately 72 hours (25^{o}C) resulting in a third instar larva .

The larval stages of development increase the mass of individuals dramatically. Many larval cells become enlarged and polyploid. These cells facilitate the rapid growth of larvae but the majority of larval cells do not survive into the adult stage of development. Most adult tissues are derived from groups of diploid cells which do not take part in larval development. In larvae these cells are sequestered in small groups of tissue called as imaginal discs. Each imaginal disc gives rise to particular structures within the adult body plan. The differentiation of imaginal discs and disintegration of larval cells takes place during the pupal stage of development (5-6 days after hatching at 25^{o}C).

When third instar larvae are preparing to pupate they crawl out of the culture medium and adhere to a dry surface. The larvae pupate within the last larval skin, which at first is soft and white but gradually hardens and darkens. The transformation which takes place within the pupal case is called metamorphosis. During this complex series of events the majority of larval cells are destroyed and the diploid imaginal discs grow, divide and differentiate to form the adult fly. When the change is complete the imago (adult) emerges by forcing its way through the anterior end of the pupal case. At first the fly is elongated, with unexpanded wings. However within a short period of time the wings expand and the body takes on a more rounded appearance.

The main role of an adult fruitfly is to mate with a member of the opposite sex and thus complete the lifecycle of the species. Flies locate members of the opposite sex by a combination of sight, sound and pheromones. Mating requires specific behavioural interactions to ensure that flies breed with the correct sex and species. Finally the male donates some of its germ cells, in the form of spermatozoa, to the female. These may be stored within the female body before being used to fertilise oocytes.

The formation and development of germ cells in both males and females has been extensively characterised. The mutation studied in this thesis disrupts the formation of

female germ cells (Mohr,1922) and thus a brief summary of oogenesis is presented.

1.3 Drosophila oogenesis

The adult female contains a pair of ovaries located in the abdomen of the fly. The ovary consists of a cluster of parallel ovarioles which contain egg-chambers lying in a single file arrangement (Figure 1.3.1). Each egg chamber consists of 15 nurse cells and an oocyte surrounded by a single cell thick envelope of follicle cells. The oocyte and 15 nurse cells are descendants of a single germ cell, the germarial cystoblast, located at the anterior tip of the ovariole (the germarium). The follicle cells which form the outer layer of the egg-chamber are derived from somatic tissue. Thus egg-chambers contain both germline and somatically derived cells.

Each ovariole is differentiated into an anterior cylinder shaped germarium and a posterior vitellarium (Figure 1.3.1). The consecutive mitoses which produce the 15 nurse cells and oocyte occur within the germarium. During this process the daughter cells remain interlinked by cytoplasmic bridges. These "ring canals" persist throughout egg-chamber development and allow gene products to be transferred from nurse cells into the developing oocyte. Before leaving the germarium the cluster of germ cells becomes enveloped in a layer of follicle cells. The first egg-chamber in the vitellarium is connected to the germarium by a stalk of follicle cells. The second and subsequent chambers lie in single file along the length of the ovariole , each chamber at a more advanced stage of development than the previous (anterior egg-chamber). The majority of egg-chamber growth and differentiation occurs within the vitellarium.

The development of egg-chambers has been subdivided into 14 stages (King,1960). The criteria used to stage egg-chambers include size, shape and location of the chamber within the ovariole, relative size of oocyte with respect to the nurse cells, nuclear and subcellular

Figure 1.3.1 - Oogenesis: early stages

A diagram of a single ovariole containing stages 1-6. NC = nurse cell, FC = follicle cell, RC = ring canal. Later stage egg-chambers (S7 and S8) have been isolated from the vitellarium to improve diagram clarity (Taken from King,1970).

Figure 1.3.1

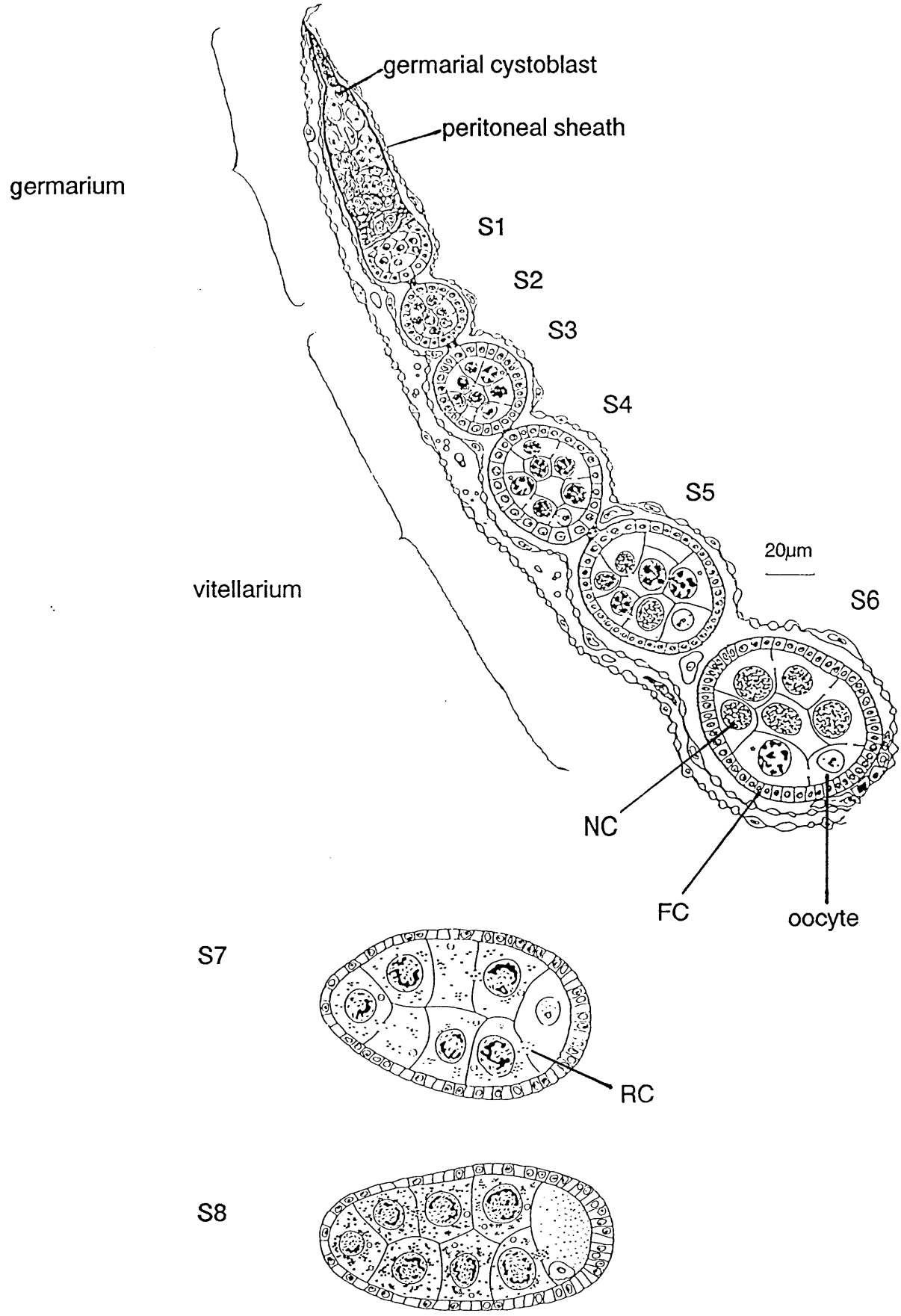
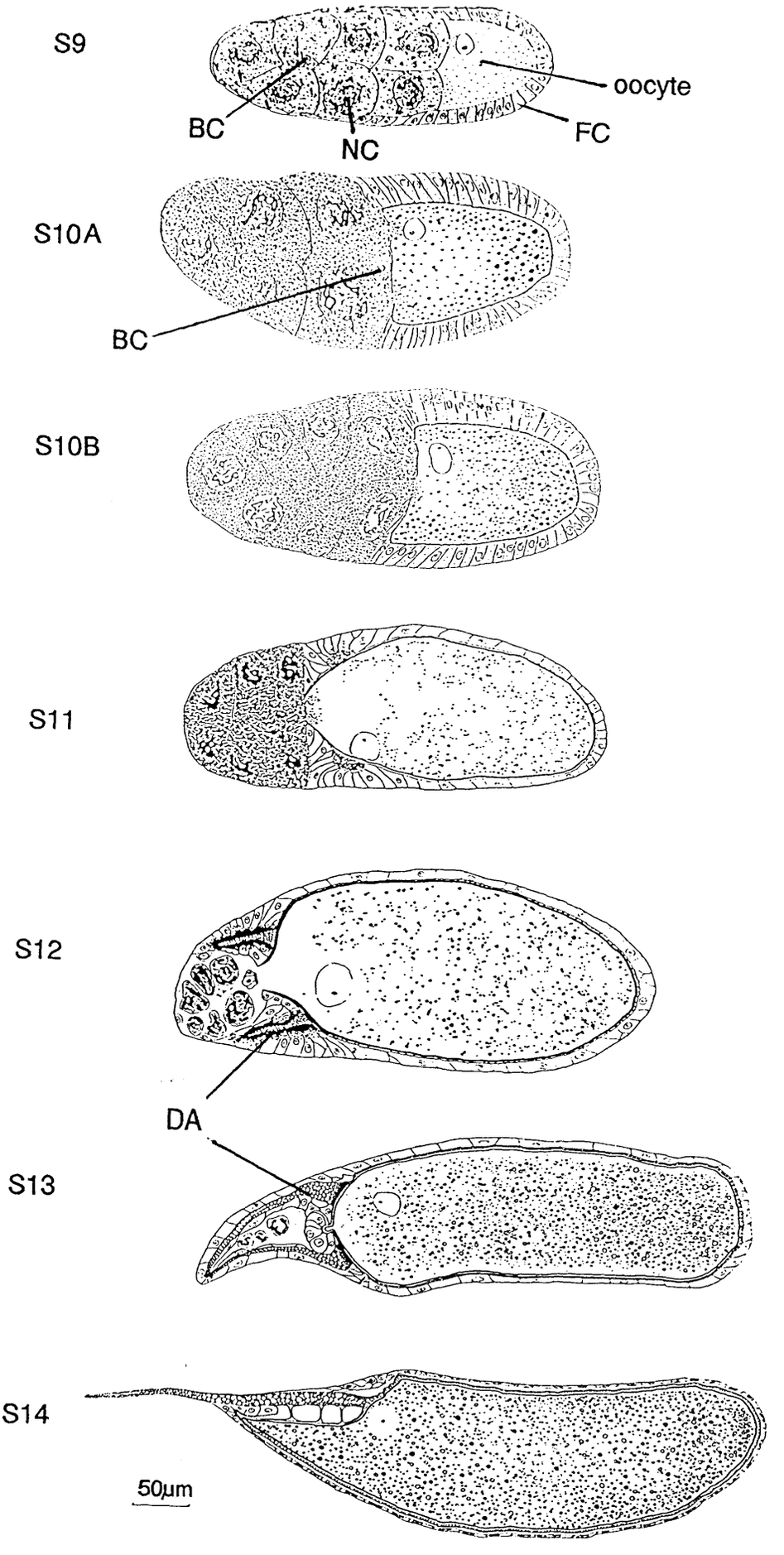

germarial cystoblast
peritoneal sheath
germarium
S1
S2
S3
S4
S5
20µm
vitellarium
S6
NC
FC
oocyte
S7
RC
S8

Figure 1.3.2 - Oogenesis: late stages

Drawings of longitudinal sections through late stage egg-chambers. The migration of border cells can be seen in stages 9 and 10A. Nurse cell degeneration is initiated at stage 10B and dorsal appendages start to from during late stage 11, early stage 12. BC = border cells, NC = nurse cells, FC = follicle cells, DA = dorsal appendages (Taken from King,1970).

Figure 1.3.2



morphology and the characteristic migratory behaviour of follicle cells.

Upon leaving the germarium the nurse cells of an egg-chamber begin replicating their DNA and become polyploid (endoreplication). During stages 1 to 4 the nurse cells increase their level of polyploidy, follicle cells grow and divide and the whole egg-chamber gradually increases in size. During these stages of development the oocyte nucleus is arrested in meiotic prophase, synaptonemal complexes form and genetic exchange occurs. By stage 5 the egg-chamber is spherical and surrounded by approximately 1100 cuboidal follicle cells. During stage 5 and 6 nurse cells continue to replicate their DNA but those cells closest to the oocyte undergo more replication cycles than those cells further away (Jacob and Sirlin, 1959).

By stage 7 the shape of egg-chambers has changed from roughly spherical (stages 1-6) to an elongated oval shape. Later stage egg-chambers and mature oocytes retain this cylindrical shape. Vitellogenesis, the process of yolk production and uptake, is initiated in stage 8 egg-chambers. Vitellogenesis continues throughout stages 9, 10 and 11.

The layer of follicle cells surrounding the oocyte differentiate at stage 8-9. Follicle cells migrate posteriorily and become columnar in shape. Cells remaining over the anterior half of the egg-chamber become flattened and enlarged in order to cover the entire nurse cell chamber.

Stage 9 egg-chambers can be distinguished by the position of various migrating follicle cells within the egg-chamber. Initially a group of 6 to 10 follicle cells which reside at the anterior pole of the egg-chamber pinch off the follicular epithelium and migrate through the interior of the nurse cell chamber. These follicle cells pass between the nurse cells and eventually reach the surface of the oocyte. They then migrate dorsally to a position directly opposite the oocyte nucleus. This group of specialised cells, called border cells, have a role in secreting vitelline membrane during later stages of

development (10B, 11 and 12).

A stage 10A oocyte occupies approximately half of the egg-chamber volume. Vitellogenesis is packing protein and lipid molecules into the oocyte so that it now appears translucent. The nurse cells have reached their maximal level of production and gene products are passed into the developing oocyte through the ring canals.

Nurse cell degradation begins in stage 10B. Initially this is seen as a change in nuclear morphology (dense to dispersed), which represents a reduction in DNA content. The border cells and columnar follicle cells surrounding the stage 10B oocyte begin to secrete vitelline membrane.

Nurse cell degeneration continues into stage 11 and the oocyte swells as it takes up water from its surroundings. The vitelline membrane forms around the oocyte, leaving only one gap, opposite the border cells, where lipid-rich cytoplasm is transferred from the nurse cells to the oocyte. By the end of stage 11 the inner endochorion has been formed and during stage 12 the outer endochorion is secreted by the columnar follicle cells surrounding the oocyte. Nurse cell degeneration is nearly complete by stage 12 and the dorsal appendages begin to form. These structures are built by follicle cells near the anterior end of the oocyte. During stage 13 the exochorion and micropylar complex are completed.

Stage 14 is used to complete the dorsal appendages and the follicle cells die off. The mature oocyte now passes down the lateral oviduct into the common oviduct in preparation for fertilisation and deposition.

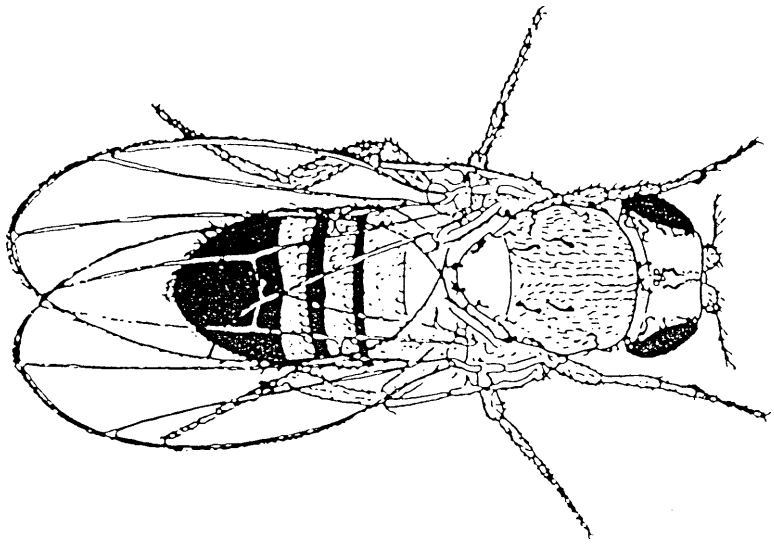
1.4 The singed locus of Drosophila melanogaster

Mutant alleles of the sex-linked singed (sn) locus alter the morphology of adult hairs and bristles when homozygous (Figure 1.4.1 Mohr,1922). In addition to this somatic phenotype some sn mutations disrupt oogenesis resulting in female sterility. The phenotype of these extreme alleles of sn suggest that the gene product(s) has

Figure 1.4.1 - $\underline{sn}^{-}$ phenotypes

An extreme $\underline{sn}^{-}$ bristle phenotype produces gnarled and twisted bristles and hairs on the cuticle of an adult fly (upper section, Taken from Lindsay and Grell,1968). In some mutants these bristle and hair abnormalities are associated with a female sterility phenotype. The disruption to oogenesis is shown diagramatically in the lower section of this figure. The $\underline{sn}^{-}$ ovary contains egg-chambers with normal morphology upto stage 6. Beyond this stage of development, $\underline{sn}^{-}$ egg-chambers start to disintegrate and appear grossly abnormal (Taken from King, 1970).

Figure 1.4.1



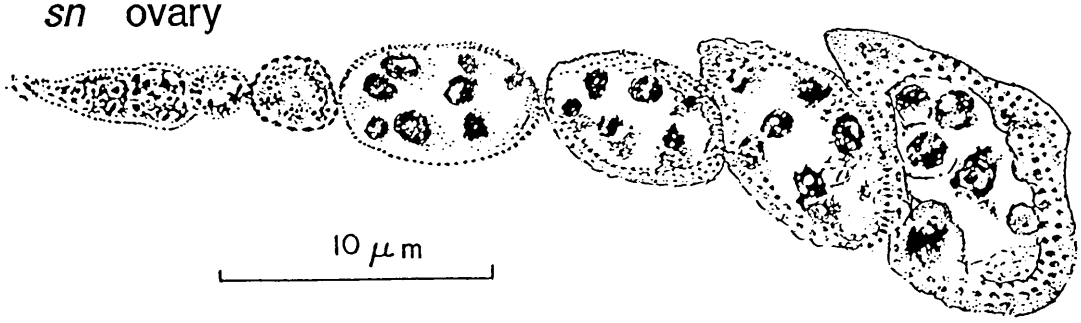
sn$^-$ adult

100µm

wildtype ovary



sn$^-$ ovary



10 µm

a role in both somatic and germ cell differentiation.

The somatic phenotype of sn can vary between two extremes. The most severe alleles of sn reduce the long slender bristles (macrochaete) located on the dorsal surface of the thorax to short, gnarled and twisted structures. Smaller bristles (microchaete) and surface hairs of the adult fly become kinked and bent. This bristle phenotype gives the fly an overall "burnt" appearance and hence the gene was named singed. Weaker alleles of sn only effect the morphology of macrochaete: they may become bent or kinked near their tip.

The bristle phenotype of sn^- individuals appears to be due to disruption of the bristle organ. Wildtype bristle organs consist of four cells, the trichogen which produces the bristle shaft, the tormogen which forms a socket around the base of the shaft and two sensory cells (the sensory cell and the neurilemma cell). In sn^- mutants the trichogen and tormogen are spacially displaced with respect to wildtype bristle organs (reviewed in Poodry,1980). Hairs are the secretory product of single cells (trichomes). To date no analysis of sn^- trichomes has been reported.

In general sn alleles with an extreme bristle phenotype are female sterile (eg. sn^1) while alleles with weak bristle phenotypes are female fertile (eg. sn^2). The one exception to this rule is sn^3: extreme bristle phenotype but female fertile. Molecular genetic analysis of this mutant (presented later in this thesis) shows that a most unusual mechanism acts to circumvent female sterility in this mutation.

The first sign of sn female sterility is the failure of vitellogenesis during the mid-stages of oogenesis (Bender, 1960). The lack of yolk deposition in the oocyte results in a flacid mis-shapen egg which never develops beyond a stage 13 egg-chamber (Figure 1.4.1). Other processes such as nurse cell endoreplication, dorsal appendage development and micropyle formation appear to be effected by lack of sn function. Flacid sn^- eggs may be deposited by the mother. Fertilisation by a sn^+ father

fails to rescue embryonic development. There appears to be a critical requirement for sn product during oogenesis.

Transplanting sn^- ovaries into a sn^+ host fails to rescue the oogenic phenotype (Clancy and Beadle,1937). This shows that the sn gene is expressed autonomously within ovaries. The site of sn gene expression has been investigated using genetically mosaic animals (Perrimon and Gans,1983). This has demonstrated a genetic requirement for sn expression in female germ cells. In contrast to females, mutations at the sn locus have no effect on male fertility.

The phenotype of sn mutants suggest that the sn gene provides a function required in somatic cells for hair and bristle formation and in female germ cells for successful oogenesis.

1.5 P-M hybrid dysgenesis and P-elements

Many sn mutations have been isolated from P-M dysgenic crosses. A group of three such mutations are analysed in Chapter 5 of this study. Thus in order to understand some of the genetics and biology of P-M hybrid dysgenesis a brief summary of this phenomenon is presented.

P-M hybrid dysgenesis is a term used to describe a series of genetic traits present in the F_1 progeny of a cross between P strain males and M strain females. The reciprocal cross of M strain males mated with P strain females produce normal offspring. P strains contain regulatory factors (described as P cytotype) which repress the process of P-M dysgenesis. M strains lack these repressor agents (M-cytotype) thus when male P strain chromosomes enter an M-cytotype environment de-repression occurs resulting in hybrid dysgenesis.

Genetic traits produced by P-M hybrid dysgenesis include sterility, male recombination and chromosomal rearrangements. The sterility phenotype of these crosses is highly temperature dependent; this is often referred to as gonadal dysgenesis. Within the F_1 generation P-M hybrid dysgenesis is limited to germ cells.

Genetically, P cytotype results from the presence of a family of transposable elements, P-elements, within the chromosomes of P strain flies. Copies of these mobile genetic elements are dispersed throughout the genome (Bingham et al.,1982).

P-elements can be subdivided into two classes; those capable of making transposase (autonomous) and those unable to manufacture transposase (defective). Non-defective P-elements have been defined as a 2.9kb interval of DNA flanked by 31bp inverted repeats (Spradling and Rubin,1982; O'Hare and Rubin,1983). Full length P-elements encode two proteins; one capable of catalysing P-element transposition (87kdal) and one thought to repress transposition (66kdal). Repressor protein (66kdal) is produced in all somatic cells. Transposase protein is only produced in germ cells thus P-element transposition is limited to the germline of F_1 P-M dysgenic individuals.

Defective P-elements are related to the full length 2.9kb P-element by various internal deletions. These elements are unable to produce transposase. However, provided at least 150bp of sequence at either end of the P-element remains intact, defective elements may be mobilised if supplied with transposase from a 2.9kb P-element.

To determine the cytotype of a particular fly stock geneticists use an assay for P-element mobility based on a hypermutable allele of sn; sn^W. During P-M hybrid dysgenesis sn^W mutates at a high frequency (40-60%) to either a more extreme allele, sn^e or an allele with wildtype appearance, $sn^{(+)}$ (Engels,1979). The instability of sn^W is highly sensitive to the number of intact P-elements in a genome.

Over the past ten years P-elements have been used for a variety of investigative purposes. Unmodified or *in vitro* modified P-elements can be used for transposon tagging (reviewed in Shapiro,1983). P-element alleles of the particular gene are isolated from dysgenic crosses and the P-element allele is cloned by virtue of its homology

to the 2.9kb P-element. The fact that P-element transposase can induce the transposition of sequences flanked on either side by the 31bp P-element inverted repeat and some associated P-element sequences has been used to stably transform flies (Rubin and Spradling,1982). These types of experiment allow the genetic function of defined regions of DNA to be determined. The control of P-element transposition has been investigated and somatic transposition can now be induced using the *in vitro* modified Δ2-3 P-element (Laski et al.,1986). More recently bacterial genes have been introduced into P-element constructs (eg.P[lacZ] elements; O'Kane and Gehring,1987). When inserted next to an active regulatory element these transposons express the bacterial reporter gene allowing cell specific gene expression to be studied.

1.6 P-element insertions and rearrangements at singed

New alleles of the *sn* locus are generated at remarkably high frequencies during P-M hybrid dysgenesis ($>10^{-3}$, reviewed by Engels,1983). This suggests that P-elements preferentially insert into the *sn* locus. To investigate this phenomenon and the basis of sn^{W} hypermutability Roiha et al.,(1988) cloned sequences from the *sn* locus.

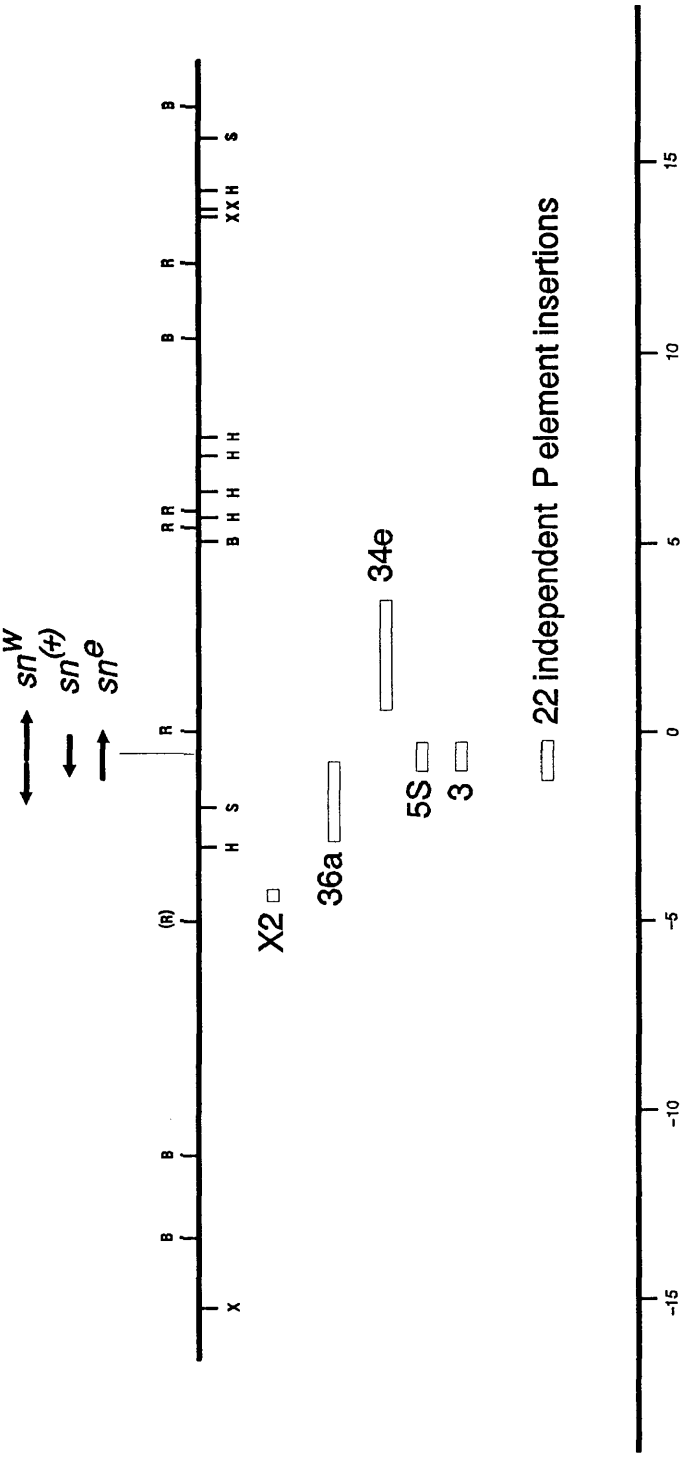
DNA from the *sn* locus (cytological location 7D1-2) was isolated using inversions between a previously cloned P-element (at cytological position 17C) and the hypermutable allele sn^{W} (Figure 1.6.1). To confirm the location of the *sn* gene nine non-P-element alleles of *sn* were examined for lesions in the DNA flanking the site of chromosome rearrangement. Five of these alleles showed alterations in their DNA maps confirming that sequences from the *sn* locus had been cloned (Figure 1.6.1).

To characterise P-element insertions at the *sn* locus twenty two alleles of *sn*, generated by P-M mutagenesis were studied for alterations in their DNA maps. All twenty two alleles contain P-element sequences inserted within a 700bp interval. At least four different insertion sites

Figure 1.6.1 - DNA map of the sn region

The DNA map is redrawn from Roiha et al.(1988). The coordinate system is in kilobases and uses the EcoRI site closest to the inversion breakpoints in B332.1 and B337.2 (chromosome rearrangements used to clone the sn region) as origin. The abbreviations for restriction enzymes are X = XhoI, B = BamHI, H = HinDIII, S = SalI. The site shown as (R) is a polymorphic EcoRI site. The intervals within which different sn mutations show changes in their DNA map are indicated as open boxes. X2 = sn^{X2}, 36a = sn^{36a}, 34e = sn^{34e}, 5S = sn^{5S}, 3 = sn^{3}. The site of insertion and structure of the 3 P-element alleles sn^{W}, sn^{e}, $sn^{(+)}$ are denoted as follows: vertical line marks site of insertion for all 3 mutations, arrows represent the size and orientaion of P-element sequences in these alleles (see Chapter 5).

Figure 1.6.1


sn^w
sn^(+)
sn^e
X2
36a
34e
5S
3
22 independent P element insertions

are used, so the clustering is thought to reflect chromatin structure within this region.

DNA probes from the sn locus were also used to investigate the hypermutable structure of sn^{W}, and its derivative alleles sn^{e} and $sn^{(+)}$ (Roiha et al.,1988; Nitasaka and Yamazaki,1988). The sn^{W} mutation contains two defective P-elements in a head to head orientation near the 5' end of the sn gene. These P-elements appear to be separated by only 8bp. Excision of the distal P-element results in a more extreme bristle phenotype (sn^{e}) while excision of the proximal element leads to a wildtype bristle phenotype ($sn^{(+)}$).

The three P-element alleles sn^{W}, sn^{e} and $sn^{(+)}$ were originally isolated in P cytotype backgrounds. Each chromosome was described as carrying a sterile allele of sn and thus were maintained in compound first chromosome stocks. During an attempt to clone the sn gene by transposon tagging the sn^{W}, sn^{e} and $sn^{(+)}$ chromosomes were crossed into a non-P-element containing genetic background. These M cytotype stocks lost their female sterile phenotype allowing sn^{W}, sn^{e} and $sn^{(+)}$ chromosomes to be maintained as homozygous stocks (K.O'Hare pers. comm.). M cytotype flies show exactly the same bristle phenotype as their P cytotype counterpart. Thus sn^{W}, sn^{e} and $sn^{(+)}$ display cytotype dependent sterility.

1.7 Aims and objectives

To date interest in the sn gene has concentrated on P-element mutations (Roiha et al.,1988; Nitasaka and Yamazaki,1988). However the phenotype of sn mutants suggest that the gene product(s) play a crucial role during oogenesis and also affect the formation of cuticular hairs and bristles. Sequences from the gene have been cloned (Roiha et al.,1988) thus molecular genetics can be used to investigate the function of sn product(s) in these apparently unconnected processes.

This thesis set out to define the structure and function of the sn gene and thereby specify the role of sn

product(s) in both somatic and germline cells. Investigations of gene function are often aided by analysing mutant alleles of the gene. Thus knowing how mutations disrupt sn gene function should help to define the role of the wildtype gene product(s). Finally a molecular investigation of the cytotype dependent sterility of sn^{w}, sn^{e} and $sn^{(+)}$ is presented.

CHAPTER 2

Materials and Methods

2.1 Bacterial strains

Plasmids and lambda bacteriophages were propagated in E.coli strain C600 :F^-, thi-1, thr-1, leu-B6, lac-Y1, ton-A21, supE44,λ^- (Appleyard, 1954), or K802 : $hsdR^-$, $hsdM^+$, gal^-, met^-, supE (Wood, 1966). For spi selection of λ red^-gam^- recombinants the strain Q359 was used: $hsdR^-$, $hsdM^+$, supE, ϕ80,(P2) (Karn et al.,1980). To grow recombinant λ gt10 bacteriophage, the strain C600 hflA was used (Hoyt et al., 1982). M13 bacteriophage and plasmids requiring blue-white selection for recombinants were propagated in E.coli JM101 : Δ(lac, pro), supE, thi^-, F'traD36, proAB, $lacI^q$, Z ΔM15(Messing et al., 1981).

2.2 Bacterial Media

Liquid cultures of E.coli were grown in Luria broth (L-broth) as described in Maniatis et al., (1982). L-plates were made by adding 15g per litre of Bacto-agar (Difco) to L-broth and autoclaving. M13 and λ bacteriophage were plated onto L-plates in a 5ml overlay of top agar (0.7%) or top agarose (0.7%) (Maniatis et al.,1982) depending on requirements. Blue-white selection of M13 recombinants due to the insertional inactivation of the lacZ gene was performed by adding 10μl 0.1M IPTG and 25μl 0.2% X-GAL in dimethylformamide to 5ml of top agar for each plate. For blue-white selection when cloning using the plasmid vector pBluescribe M13- (Stratagene), the same amounts of IPTG and X-GAL along with 4μl of 60mg.ml^{-1} ampicillin were added to a 5ml pre-poured layer of L-agar on L-plates containing ampicillin. Genomic libraries in λ vectors were plated in an overlay of L-agarose containing 10mM $MgCl_2$, onto a support of λ agar both prepared as described in Maniatis et al.,(1982). For dilution and storage of bacteriophage, phage buffer was used: 10mM tris.HCl pH7.5, 10mM $MgSO_4$, 50mM NaCl, 0.001% gelatin. Antibiotics were prepared and used as described by Maniatis et al., (1982).

2.3 Preparation of DNA

2.3.1 Small scale preparation of plasmid DNA

Bacterial colonies were inoculated into 2ml L-Broth containing 50-70μg.ml^{-1} ampicillin where appropriate and grown overnight with shaking at 37^{o}C. Small scale preparations of plasmid DNA were carried out using 1ml of the overnight cultures according to the method of Birnboim and Doly (1979), as adapted by Burke and Ish-Horowicz (1981).

2.3.2 Large scale preparation of plasmid DNA

Large scale plasmid DNA preparations were carried out by the above method as modified by D.Ish-Horowicz. Plasmid cultures (200-500ml) were grown overnight with shaking at 37^{o}C in L-broth containing 50-70μg.ml^{-1} ampicillin. Cells were harvested by centrifugation in a Sorvall GS-3 (Du Pont) rotor at 7,000rpm for 5 minutes. The pellet was resuspended in 38ml of 0.9% glucose, 25mM tris.HCl pH8, 10mM EDTA and then 40ml of 0.2M NaOH, 1% SDS added. After 5 minutes, 20ml of ice-cold 29.5% potassium acetate, 11.4% v/v glacial acetic acid was added and the mixture stored on ice for 10 minutes. After centrifugation at 8,000rpm for 5 minutes the supernatant was filtered through muslin and 0.6 volumes of isopropanol added and then centrifuged again at 8,000rpm for 5 minutes. After draining, the pellet was resuspended in 8.1ml T.E.(10,1), 0.9ml of 3M Na acetate added and transferred to a 30ml Corex tube (Du Pont) where 15ml of ethanol was added. The precipitate was pelleted at 5,000rpm for 5 minutes in a Sorvall SS-34 rotor and resuspended in 9ml T.E.(10,1); 12g CsCl and 1ml of 10mg.ml^{-1} ethidium bromide were then added and the sample centrifuged in Oakridge tubes at 10,000rpm for 30 minutes to pellet excess RNA. The supernatant was transferred to an ultracentrifuge tube, made up with 1.55g.ml^{-1} CsCl if necessary and then centrifuged in a Beckman Vti.50 or

50.Ti rotor at 45,000rpm for 24-48 hours. The plasmid band was visualised under UV illumination and harvested using a needle and syringe. The ethidium bromide was removed by repeated extractions with isopropanol saturated with CsCl-saturated T.E.(10,1) and then the plasmid sample was dialysed exhaustively in T.E.(10,1). Plasmid DNA concentrations were determined spectrophotometrically and adjusted to 200μg.ml^{-1}. Yields were typically 1-4mg plasmid DNA.

2.3.3 Large scale preparation of lambda DNA

An aliquot (2-3ml) of C600 grown overnight in L-broth containing 10mM $MgCl_2$ and 0.2% maltose was infected with bacteriophage at a multiplicity of 0.01-0.1 by incubating at 37^{o}C for 15 minutes. The infected cells were added to 250-400ml of pre-warmed L-broth containing 10mM $MgCl_2$ and shaken at 37^{o}C for 4-7 hours until lysis occurred. After the addition of 3ml chloroform and 30 minutes further shaking, the mixture was centrifuged in a Sorvall GS-3 rotor at 5,000rpm for 10 minutes. To the supernatant was added NaCl to 6% and PEG 6000 (BDH) to 7%, and the mixture left overnight at 4^{o}C. The PEG precipitate was pelleted by centrifugation at 5,000rpm for 30 minutes at 4^{o}C, resuspended in 3ml of phage buffer and transferred to a 15ml Corex tube to which was added 30μl of 1mg.ml^{-1} DNase I (Sigma) and 30μl of 1mg.ml^{-1} RNase (Sigma). After incubating at 37^{o}C for 1 hour the sample was extracted with 3ml chloroform and then loaded on the top of a CsCl density gradient consisting of 2ml layers of 1.6, 1.5, 1.4 and 1.3g.ml^{-1} CsCl made up in 10mM tris.HCl pH7.5, 10mM $MgCl_2$ (TM10). The sample was then centrifuged in a Beckman SW-41Ti rotor at 25,000rpm for 2.5-3 hours at 4^{o}C. The phage band was extracted using a needle and syringe and dialysed extensively against TM10. The phage sample was made 20mM with EDTA and 0.1% with SDS and then extracted once with phenol, once with 1:1 phenol/chloroform and once with chloroform. After ethanol precipitation, washing with 70% ethanol and drying, the

DNA pellet was resuspended in T.E.(10,1). Total yields were typically 100-500 μg DNA per 250ml of culture.

2.3.4 Preparation of Drosophila DNA

Fresh or frozen adult flies (100-500mg) were weighed and added to 5ml of homogenisation buffer: 0.1M NaCl, 30mM tris.HCl pH8, 10mM EDTA, 7.7mM 2-mercaptoethanol, 0.5% Triton X-100 and homogenised in an electrically driven 30ml Dounce homogeniser. The homogenate was filtered through muslin into a 30ml Corex tube. The homogeniser was rinsed with a further 2ml of homogenisation buffer and this was also filtered into the Corex tube. Nuclei were pelleted in a Sorvall HB-4 rotor at 4,000rpm for 10 minutes at 4^{o}C and the pellet resuspended in 1ml homogenisation buffer and transferred to an Eppendorf tube. After centrifugation for 1 minute in a microcentrifuge the pellet was resuspended in 0.3ml of lysis buffer: 30mM tris.HCl pH8, 100mM EDTA and 30μl of 1mg.ml^{-1} proteinase-K (Sigma) and 30μl of 10% N-lauryl sarcosinate added. After incubation at 45^{o}C for 2-3 hours, the sample was extracted once with 0.3ml 1:1 phenol/chloroform, once with 0.3ml chloroform and the DNA precipitated by the addition of 15μl 5M NaCl and 0.8ml ethanol. The pellet was washed twice with 70% ethanol, air-dried and resuspended in 0.8ml T.E.(10,1) per 1g of original weight of flies, giving a DNA concentration of ~200μg.ml^{-1}. To aid resuspension samples were incubated 65^{o}C for 10 minutes.

2.4 Preparation of Drosophila poly-A^{+} RNA

All aqueous solutions and automatic disposable pipette tips used were autoclaved before use. Glass pipettes and Corex tubes were baked overnight at 250^{o}C. Adult flies, pupae, larvae or embryos (2-4g) were placed in a 50ml sterile plastic Falcon tube containing 5ml 100:1 chloroform:isoamyl alcohol, 5ml neutralised phenol and 10ml homogenisation buffer: 0.15M Na-acetate, 5mM EDTA, 1%

SDS, 50mM tris.HCl pH9, 20μg.ml^{-1} polyvinyl sulphate (Sigma). The mix was homogenised using a tissue disrupter (Polytron Kinematica) for 2 bursts, half maximal setting, of 30 seconds each. After spinning at full speed in a bench-top centrifuge for 5 minutes, the lower phase was removed and 5ml phenol and 5ml 100:1 chloroform:isoamyl alcohol added. The mix was vortexed and then centrifuged and the lower phase removed as above. The upper aqueous phase was then extracted twice with 10ml 100:1 chloroform:isoamyl alcohol. The process of vortexing and centrifugation was repeated as described above. The upper phase was then transferred to a 30ml Corex tube and 20ml 96% ethanol added. After storage overnight at -20^{o}C, the precipitate was pelleted in a Sorvall HB-4 rotor at 10,000rpm for 15 minutes at 4^{o}C, then resuspended in 4ml SDS buffer: 10mM tris.HCl pH7.5, 100mM NaCl, 1mM EDTA, 0.5% SDS containing 0.3mg.ml^{-1} proteinase-K (Sigma) and incubated for 2 hours at 37^{o}C. The sample was extracted twice with 2ml phenol and 2ml 100:1 chloroform:isoamyl alcohol and then twice with 4ml 100:1 chloroform:isoamyl alcohol with vortexing and centrifugation at 5,000rpm for 10 minutes in a Sorvall HB-4 rotor. The aqueous phase was then transferred to a 15ml Corex tube and ethanol precipitated. The precipitate was pelleted as above and resuspended in 3ml T.E.(10,1) containing 0.5% SDS.

Poly-A^{+} RNA was selected by chromatography through oligo-dT cellulose (Collaborative Research) according to the method of Levis and Penman (1978). The poly-A^{+} RNA obtained was ethanol precipitated and resuspended in SDS buffer. Yields were between 50 and 100μg per gram of starting material.

2.5 Ethanol precipitation of nucleic acids

Nucleic acids were precipitated from solution by the addition of 0.1 volumes of 3M Na-acetate and either 2 volumes (DNA precipitation) or 2.5 volumes (RNA precipitation) of 96% ethanol. Precipitates were pelleted by centrifugation in a Sorvall HB-4 rotor or in a

microcentrifuge, depending on volume. Pellets were washed with 70% ethanol and either air-dried or dried in a vacuum dessicator.

2.6 Phenol extraction of DNA

DNA solutions were phenol extracted by adding an equal volume of neutralised phenol, vortexing and separating the layers by centrifugation. The aqueous layer was then extracted with an equal volume of 1:1 phenol:chloroform and then with an equal volume of chloroform in the same manner. Redistilled phenol was neutralised by vortexing with an equal volume of 0.3M tris base and separating the layers by centrifugation, after which the phenol was stored under a layer of T.E.(10,1) pH7.5.

2.7 Enzymatic digestion of DNA

2.7.1 Endonuclease digestion of DNA

Restriction enzyme digestion of DNA samples was carried out according to Maniatis et al.,(1982) in either the Low, Medium or High salt buffers described therein to facilitate simultaneous multiple digests. Enzymes were supplied by either BCL or New England Biolabs Inc. Digestions of cloned or genomic DNA were routinely carried out for greater than 4 hours or overnight respectively. DNA fragments for use as molecular weight standards were generated by digesting λcIts857 DNA with HindIII or BstE II (LambdaII, Appendix II.1983).

2.7.2 Exonuclease digestion of DNA

Fragments of duplex DNA were progressively shortened from both ends using Bal 31 nuclease (Boehringer Mannheim GmbH) as described in Ponz, et al.,(1982). In order to determine the kinetics of DNA shortening, small scale time course experiments were performed at several ratios of

enzyme to DNA. Digests were incubated at 25°C and samples were removed at specific time points during the reaction. Nuclease activity was quenched by immediately adding time point samples to an equal volume of 500mM EDTA. The extent of shortening was determined by eletrophoresing time point samples on a neutral agarose gel. DNA fragments were visualised by short-wave UV illumination. The small scale reaction with the most appropriate kinetics (e.g. 300bp difference between each time point) was scaled up several fold. Samples from the large scale reaction were electrophoresed on a low-gelling-temperature agarose gel and shortened DNA fragments were always purified as described in Section 2.10. To increase cloning efficiencies the samples of shortened DNA were digested with an endonuclease which cuts within the central region of the initial full length molecule. Fragments with one Bal 31 blunt end and one endonuclease cut sticky end were directionally cloned into suitably prepared M13 vectors.

2.8 Ligation of DNA

DNA samples were ligated in 50mM tris.HCl pH7.5, 10mM $MgCl_2$, 10mM DTT, 1mM rATP containing 0.5-5 units of T4 DNA ligase. The ligation volumes were 20μl containing 0.1-0.2μg total DNA and the reactions were carried out at 14°C for either 4 hours (cohesive end ligations) or for 20 hours (blunt end ligations). Vector to target ratios were in the range 1:2 to 1:4.

2.9 Agarose gel electrophoresis

2.9.1 Neutral gels

Agarose gel electrophoresis of DNA was carried out in slab gels of 0.6-1.5% (w/v) agarose (FMC BioProducts) in 89mM tris borate pH8, 89mM boric acid, 2mM EDTA (TBE) containing 0.5-1ug.ml^{-1} ethidium bromide. Before loading, DNA samples were combined with a 1/6th volume of sample buffer: 0.25% bromophenol blue, 30% glycerol in H_2O.

Large gels (15x22cm) were run at 2-4V.cm^{-1}; minigels (10x10cm) were run at up to 6V.cm^{-1}. DNA fragments were visualised and photographed under short-wave UV illumination. The sizes of DNA fragments were determined from a calibration curve of the migration distances vs. the fragment sizes of DNA size standards (section 2.7.1).

2.9.2 Formaldehyde agarose gels

Formaldehyde agarose gel electrophoresis of RNA was carried out in submerged slab gels of 1-1.5% agarose in 18mM Na_2HPO_4, 2mM NaH_2PO_4 (1x gel buffer) containing 6% formaldehyde. The gels were submerged in 2 litres of 1x gel buffer containing 3% formaldehyde and run at 1V.cm^{-1} for 16-18 hours. Poly-A$^+$ RNA samples were ethanol precipitated and resuspended in 20μl of sample buffer: 50% formamide, 6% formaldehyde, 1x gel buffer; then combined with 20μl dye buffer: 50% formamide, 50% glycerol, 1x gel buffer, 0.1% bromophenol blue, 0.1% xylene cyanol. The samples were heated at 65^{o}C for 5 minutes and cooled on ice before loading. Size standards were single stranded DNA fragments obtained by denaturing λcIts857 DNA digested with HindIII.

2.10 Purification of DNA fragments from agarose gels

DNA restriction enzyme digests were loaded on gels made with 0.7-0.9% low-gelling-temperature (LGT) agarose (FMC BioProducts). Under UV illumination, the fragments of interest were excised in as small a piece of gel as possible. The slab was then placed in an Eppendorf tube. An equal volume of T.E.(10,1) was added and the agarose melted in a 65^{o}C waterbath for 5 minutes, following which the sample was phenol extracted, ethanol precipitated and resuspended in T.E.(10,1).

2.11 Nucleic acid transfer and filter hybridisation

2.11.1 DNA transfer and filter hybridisation

Transfer of DNA fragments to filters was carried out following the method of Southern (1975), with the modification that directly following electrophoresis, the DNA in the gel was partially depurinated by shaking in 0.25M HCl for 15 minutes. Prior to transfer, filters (nitrocellulose from Schleicher and Schuell or Amersham Hybond-N) were prewetted in H_2O and then 2x SSC (1x SSC is 150mM NaCl, 15mM Na-citrate). After transfer, the filters were rinsed in 2x SSC and then baked at 80°C for 2 hours. The filters were pre-hybridised in a solution of 5x SSC, 0.1% SDS at 65°C for at least 2 hours and then hybridised in 5x SSC, 0.1% SDS, 2x Denhardt's solution (1x Denhardt's is 0.02% BSA, 0.02% Ficoll, 0.02% polyvinyl pyrrolidone, all Sigma), 100μg.ml^{-1} single-stranded salmon sperm DNA. The radioactive DNA probe was denatured in a boiling waterbath for 5 minutes before addition to the filter bag, to yield a probe DNA concentration of 10-50ng.ml^{-1}. After 16-48 hours hybridisation, the filters were washed by shaking in 2x SSC, 1x Denhardt's, 0.1% SDS for 2 x 30 minutes, followed by 2 x 30 minutes in 2x SSC, 0.1% SDS, finishing with 2 x 15 minutes in 3mM tris base, 0.1% SDS; all washes were at room temperature. Filters were autoradiographed using X-ray film (RX, Fuji or XAR-5, Kodak) either at room temperature or at -70°C with an intensifying screen (Du Pont Cronex Lightning Plus) (Laskey and Mills,1977).

For re-probing Southern blot filters, hybridisation probes were removed from nitrocellulose filters by boiling in water for three minutes. On occasions, the length of time elapsed between hybridisations obviated the need to remove the previous hybridisation probe. Hybridisation probes were removed from Hybond-N filters according to the manufacturers instructions (Amersham).

2.11.2 RNA transfer and filter hybridisation

RNA gels were transferred to Hybond-N filters in the same way as DNA gels except that following electrophoresis, the gels were equilibrated in 10x SSC for 20 minutes prior to setting up the transfer; and that the transfer solution used was 10x SSC. Filters were prehybridised overnight at 42^{o}C in hybridisation solution: 50% de-ionised formamide, 5x SET (1x SET is 150mM NaCl, 30mM tris.HCl pH8, 2mM EDTA), 5x Denhardt's, 400ug.ml^{-1} single-stranded salmon sperm DNA, 0.5% SDS, 2x RNA gel buffer (see section 2.9.2). Filters were hybridised in hybridisation solution at 42^{o}C for two nights, generally with radioactive single-stranded M13 DNA probes. The filters were washed at room temperature for 2 hours in four changes of 5x Denhardt's, 2x SET, 0.5% SDS and then for 30 minutes in two changes of 0.5x SET, 0.5% SDS. Autoradiography was performed as for DNA hybridisation filters. To control for the amount of poly-A^{+} sample loaded in each lane, the filters were later hybridised to an oligolabelled DNA probe of actin 5C DNA (Fyrberg et al.,1983).

2.12 Radioactive labelling of DNA

2.12.1 Labelling duplex DNA

Radioactive labelling of λ DNA, plasmid DNA or DNA fragments purified by agarose gel electrophoresis was carried out by either the nick translation method of Rigby et al.,(1977) or by the oligolabelling method of Feinberg and Vogelstein (1983) as adapted by Feinberg and Vogelstein (1984). Typically 0.05 to 0.5μg of DNA and 0.5 to 2μl of alpha-^{32}P-dCTP at ~3000Ci/mmol (Amersham) was used per reaction, and unincorporated nucleotides were removed by passage down a Sephadex G-50 column using the spun-column method (Maniatis et al.,1982). The specific activities obtained for the probes were typically 1-9x10^{8}dpm.μg^{-1} DNA.

2.12.2 Labelling single stranded DNA

M13 clones were used to generate radioactive single-stranded DNA probes by using T7 DNA polymerase (United States Biochemicals or Pharmacia) to synthesise a strand complementary to the insert in the presence of alpha-^{32}P-dCTP (adapted from Burke, 1984). The products of the extension reaction were digested with a restriction enzyme that cleaved either at the distal end of the inserted DNA fragment or at a site within the insert. The sample was denatured in 0.1M NaOH and then loaded onto a denaturing, 1.5% low-gelling-temperature agarose gel containing 30mM NaOH, 10mM EDTA. After electrophoresis at 3V.cm^{-1} for 16 hours or 12V.cm^{-1} for 3 to 4 hours, the position of the radioactive single-stranded DNA was determined by autoradiography and the gel slab excised. The slab was added directly to the required volume of hybridisation solution and melted by boiling for 2 minutes. This type of probe was used extensively in RNA blotting experiments.

2.13 Construction of genomic lambda libraries

Two separate libraries were made using DNA from sn^{X2}/sn^{3} heterozygotes. The first library was made using size selected DNA from a Sau3AI partial digest. After a series of test digests to determine a suitable enzyme concentration and time course, a set of time points from a large scale partial digest were electrophoresed on an LGT agarose gel. DNA in the size range from 15 to 25kb was excised in a gel slice and purified. The genomic DNA was then ethanol precipitated in the presence of λEMBL4 (Frischauf et al., 1983) DNA cut with BamH1 and Sal1. The mixture of bacteriophage and genomic DNA was resuspended in ligase buffer, 5 units of T4 DNA ligase added and the reaction incubated at 14^{o}C overnight. Aliquots of the ligation were packaged in vitro using the method of Hohn and Murray (1977). The number of recombinant bacteriophage was determined by titrating the packaging reactions on E.coli stain Q359. This showed that a total of 30,000

recombinant bacteriophage had been recovered.

The probability of recovering both mutations from the Sau3a1 library was low (Maniatis et al., 1982) thus a second library was constructed. The DNA map of the wildtype sn locus shows that interval mutated in sn^{X2} and sn^{3} lies within a 16kb BamH1 fragment (Figure 3.1.1, -11.0 to +5.0). This restriction fragment is present in both mutations and is suitable for cloning into the λ replacement vector λEMBL4. A sample of the sn^{X2}/sn^{3} DNA was digested to completion with BamH1. After size selection (12-20kb) and purification the genomic DNA was resuspended in ligation buffer containing λEMBL4 cut with BamH1 and Sal1. T4 DNA ligase was added and the reaction incubated at 14°C overnight. Samples of the ligation were *in vitro* packaged into bacteriophage particles and the number of recombinant bacteriophage titrated on *E.coli* strain Q359. The BamH1 library contains sufficient recombinant bacteriophage (120,000) to ensure that at least one λ clone carrying the mutagenised region of each allele should be isolated (Maniatis et al.,1982).

Libraries were plated out to a density of 15-30,000 plaques per 150mm petri dish using *E.coli* Q359 (λEMBL4 libraries) or C600 *hflA* (λgt10 libraries). Filter lifts were made as described in section 2.14.3, and following hybridisation of the filters, positive plaques were removed as cores and placed in a small volume of phage buffer. Two to three rounds of plaque hybridisation were sufficient to purify the positive plaques, after which they were amplified by the plate-lysate technique (Maniatis et al.,1982), followed by large scale preparation of DNA.

2.14 Subcloning DNA fragments

2.14.1 DNA vectors

Subcloning was generally carried out using gel-purified DNA fragments which were ligated to M13 or plasmid vectors as described above. Plasmid vectors used were pBluescribe M13- : a derivative of pUC19 containing bacteriophage T3 and T7 promoters flanking the cloning site polylinker (Stratagene Inc.). M13 vectors used were M13mp8, mp9 (Messing and Vieira, 1982), mp18, mp19 (Yanisch-Perron et al., 1985), tg130 and tg131 (Kieny et al., 1983). Approximately 100ng of ligated DNA was used to transform E.coli made competent by the method of Mandel and Higa (1970).

2.14.2 Identification of subclones

Plasmid recombinants were generally identified by small scale plasmid preparations followed by restriction enzyme digestion and agarose gel electrophoresis. The plasmid pBluescribe M13- facilitates the identification of recombinants through the blue-white colour selection associated with the insertional inactivation of the lacZ gene on the plasmid. In the case of M13 phage, single ddTTP/dTTP dideoxy sequencing reactions of several clones (T-tracking) was often used to identify the clone of interest. On occasions where the desired plaques or colonies formed a very small proportion of the total, the following hybridisation procedures were used.

2.14.3 Plaque hybridisation

M13 or λ plaques to be screened were plated using BBL top agar to a density of 50-300 plaques per 90mm petri dish. Filter lifts onto nitrocellulose filters (Schleicher and Schuell) were made and processed by the method of Benton and Davis (1977). After 2 hours baking at 80^{o}C, the filters were hybridised with a radioactive

probe, washed and autoradiographed as described previously.

2.14.4 Colony hybridisation

Colonies for screening were picked in duplicate into ordered arrays and one set transferred to Whatman 541 paper (Gergen et al., 1979). The filter-borne colonies were denatured and neutralised as if gels for Southern transfer and then taken directly to hybridisation without baking or prehybridisation.

2.15 Dideoxy sequencing

DNA sequencing was performed by the dideoxy chain terminating method of Sanger et al.,(1977). The preparation of single stranded M13 DNA for sequencing and duplex M13 DNA for subcloning were carried out as described in Bankier and Barrell (1983). Double stranded templates were prepared for sequencing using the method of Guo et al., (1983). The majority of clones were sequenced using T7 DNA polymerase (Tabor and Richardson.,1987; supplied by United States Biochemicals or Pharmacia) using the directions provided in "Sequenase" Kits (United States Biochemicals). The radioactive label used in the sequencing reactions was ^{35}S-dATP-alpha S at ~400 Ci.mmol^{-1} (Amersham) and nucleotides were from Pharmacia. On occasions, clones were sequenced with DNA polymerase Klenow fragment (Amersham or BCL) using the the directions provided in Bankier and Barrell (1983). Sequence reactions were fractionated on 6% polyacrylamide gels (57:3 mono:bis-acrylamide) in 1x TBE buffer, 7M urea.

2.16 Computer analysis of sequence data

Sequence data from individual clones was assembled and analysed using the programmes of Queen and Korn (1984) contained in the Microgenie software package (Beckman Instruments) on an IBM XT personal computer. Further

analysis of sequences was performed using programmes of the Staden-Plus software package (Staden.,1982, Staden.,1984); supplied by Amersham International Ltd, with an IBM AT personal computer. Searches of the EMBL, GENBANK, SWISSPROT and NBRF sequence databases were performed using programmes within the UWGCG package (University of Wisconsin Genetics Computer User Group) on the SERC Daresbury VAX mainframe computer.

2.17 Detection of β-galactosidase activity in Drosophila tissue

2.17.1 β-galactosidase staining in whole mount embryos

Collections of embryos (0-8 hours) were fixed and stained according to the protocol of Ghysen and O'Kane.,(1989). Stained embryos were mounted beneath coverslips in 0.7% NaCl and photographed using a Nikon Microphot FX microscope on Kodak print or colour reversal film.

2.17.2 β-galactosidase staining in pupae

Staining of pupal stages of Drosophila development was achieved by dissecting the pupa out of the pupal case and fixing for 10 minutes in 2.5% gluteraldehyde, 50mM PIPES pH7.5. The fixed pupae were washed in 0.7%NaCl, submerged in 500ul staining solution (Simon et al.,1985) and incubated overnight at 37^{o}C. After staining pupae were washed in 0.7%NaCl and photographed using a Nikon FG camera mounted on a Nikon SMZ-10 dissecting microscope with Kodak print or colour reversal film.

2.17.3 β-galactosidase staining in ovarioles

Ovaries were dissected from 3 day old females and stained as described in Grossniklaus et al.,(1989). Mounted ovaries were examined using a Nikon Microphot FX microscope and photographed on Kodak print or colour reversal film. Whole mounts were photographed as described

for pupae. To examine stages of oogenesis in more detail, ovarioles were teased apart and then mounted and photographed as described for embryos.

2.18 Microscopy

Three different microscopes were used during these studies. For general dissection and fly handling a Nikon SMZ-2 was used. Microphotography was carried out using either a Nikon Microphot FX or a Nikon SMZ-10 depending on magnification and optics required. Photographs were taken on Kodak EPT135-36 film processed by CETA Photographic Ltd. or Kodak Gold 200ASA print film processed at CETA (above) or Fastfoto Ltd. All microscope coverslips and slides were supplied by Chance Propper Ltd.

2.19 Drosophila media and stocks

2.19.1 Drosophila media

Drosophila stocks were maintained on cornmeal, yeast, treacle food: 40ml powdered brewer's yeast (ICN), 105ml cornmeal (ICN), 60ml black treacle (Tate and Lyle) and 9g agar per litre to which was added 15ml of 10% (w/v) methyl-4-hydroxybenzoate (Nipagin) and 2ml propionic acid. Plates for embryo collection contained grape juice agar medium: 62.5g glucose, 325g sucrose, 8.5g powdered brewer's yeast, 37.5ml grape juice concentrate (Waitrose own brand), 25g agar per litre also containing Nipagin and propionic acid as above. Stocks were maintained at 18^{o}C or 25^{o}C.

2.19.2 Oregon-R population cages

Developmentally staged individuals used to produce wildtype RNA were collected from Oregon-R population cages. Three cages each containing 5-10,000 flies are maintained in a humidified 25^{o}C room subject to a 12 hour on 12 hour off light cycle. A fresh cage is set up each

week using an overnight collection of embryos from the 3 current cages. Each population cages contains individuals of a different age (1 week, 2 weeks or 3 weeks) thus the fresh cage replaces the oldest population. The population cages are fed daily with a tray of grape juice media smeared with yeast paste. Timed egg collections made on grape juice/yeast plates were used to establish synchronous cultures of Drosophila. Specific stages of the life cycle were then harvested and used to prepare poly-A^+ RNA.

2.19.3 Drosophila stocks

The following list describes all stocks used in this work together with their origins. A description of most of the genetic markers and mutations listed can be found in Lindsley and Grell (1968).

Stock	Source	Comment
Canton-S (Ct-S)	BG	wildtype
Oregon-R (Or-R)	BG	wildtype
$ClB/w^e\ sn^1$	BG	sn^1 mutant
$w^{co}\ sn^2$	BG	sn^2 mutant
$y\ sn^3\ v$	KOH	sn^3 mutant
sn^4	BG	sn^4 mutant
$y^2\ w^a\ sn^5$ B/FM6	BG	sn^5 mutant
sn^{34e}	BG	sn^{34e} mutant
sn^{36a}/C(1)DX y f	BG	sn^{36a} mutant

In(1)dl-49, y Hw m^2 g^4/ sn^c	BG	sn^{X2} parental chr.
106/FM7	MS	source of sn^{X2} DNA
y w^{47c23} run^{AA33}/FM7c	KOH	FM7c used as source of homozygous sn^{X2} females for RNA work
A46	WG	2^{nd} chr. insertion of P[lacZ,ry^+]
w $sn^{P[lacZ,ry^+]1}$/ Y C(1)DX,y f; ry^{506}	JP	Section 4.3
$sn^{P[lacZ,ry^+]2}$/ Y C(1)DX,y f; ry^{506}	JP	Section 4.3
C(1)DX,y f/Y; ry^{506}	KOH	ry^{506} stock
Basc; Cy/bw^{V1}; Ubx/Sb	MGK	called H41-RM
w; Sb eΔ2-3/TM6	WG	Stable source of P-transposase
sn^W;ry^{506} [M]	KOH	sn^W M cytotype stock
y sn^W/ y^{+Y};bw;st [M]	KOH	sn^W M cytotype stock
4sn^e; ry^{506} [M]	KOH	sn^e M cytotype stock
4$sn^{(+)}$; ry^{506} [M]	KOH	$sn^{(+)}$ M cytotype stock

sn^{W}/C(1)DX,y f/Y [P]	KOH	sn^{W} P cytotype stock
$4sn^{e}$/C(1)DX,y f/Y [P]	KOH	sn^{e} P cytotype stock
$4sn^{(+)}$/C(1)DX,y f/Y [P]	KOH	$sn^{(+)}$ P cytotype stock
sn^{W}/Basc/Y [P]	KOH	sn^{W} [P] females
$4sn^{e}$/Basc/Y [P]	KOH	sn^{e} [P] females
$4sn^{(+)}$/Basc/Y [P]	KOH	$sn^{(+)}$ [P] females

Balancer chromosomes

First chromosome balancers

FM6 = In(1)FM6,y^{31d} sc^{8} dm B

FM7 = In(1)FM6,y^{31d} sn^{X2} sc^{8} dm B

ClB = In(1)C1,sc 1(C) t^{2} sl B

Basc = In(1) sc^{S1L} sc^{8R}+S, sc^{S1} sc^{8} w^{a} B

Third chromosome balancers

TM6 = In(3R)TM6 ss^{-} bx^{34e} e

Key

BG = Mid-America Stock Centre, Bowling Green, USA.
KOH = Kevin O'Hare, Imperial College, London, UK.
MS = Mike Simmons, University of Minnesota, USA.
WG = Walter Gehring, Biozentrum, Basel, Switzerland.
MGK = Margaret Kidwell, University of Arizona, USA.
JP = This work, see Chapter 4.

CHAPTER 3

Molecular characterisation of the *singed* gene.

3.0 Introduction

At least part of the sn locus had previously been cloned by Roiha et al.,(1988) - Chapter 1, Figure 1.6.1. Starting with this cloned interval the experiments in this chapter describe the localisation and molecular characterisation of the sn transcription unit.

3.1 DNA blotting analysis of the singed region

DNA blotting experiments to detect lesions in mutant alleles of sn have previously been carried out by Roiha et al.,(1988) - See Chapter 1. Of 9 non-P-element alleles studied by these investigators 5 were shown to contain alterations within the interval -2.0 to +4.0 Figure 1.6.1. Analysis of P-element induced breakpoint (2) and insertion (22) mutants locate lesions within the region -0.7 to 0.0 (See Chapter 1, Figure 1.6.1). This suggests that at least part of the singed gene is encoded within the interval -2.0 to +4.0 Figure 3.1.1.

Although the DNA blotting experiments of Roiha et al.,(1988) detected lesions in some sn mutants the analysis was incomplete. Potential lesions were only mapped within part of the cloned interval, -16.6 to +5.5 and +6.1 to +12.4. Thus the first experiments reported here extend the DNA blotting analysis to cover the entire cloned region.

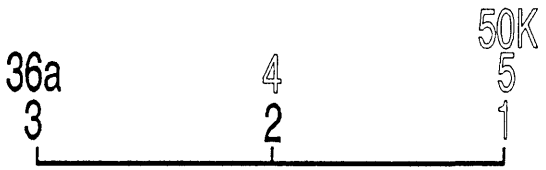
The original fly stocks analysed by Roiha et al.,(1988) are no longer available, thus replacement stocks were obtained from Drosophila Stock Centres. A list of mutant alleles analysed in this study is shown in Figure 3.1.2. Initially DNA blotting experiments concentrated on determining the size of Eco R1 fragments within the DNA of mutant alleles. Consecutive Eco R1 subclones covering the entire cloned region were used as probes in these DNA blotting experiments. When alteration were detected, other restriction endonucleases were used to map the site and nature of the lesion in further DNA blotting experiments.

Figure 3.1.1 - Genetic and DNA map of the singed region.

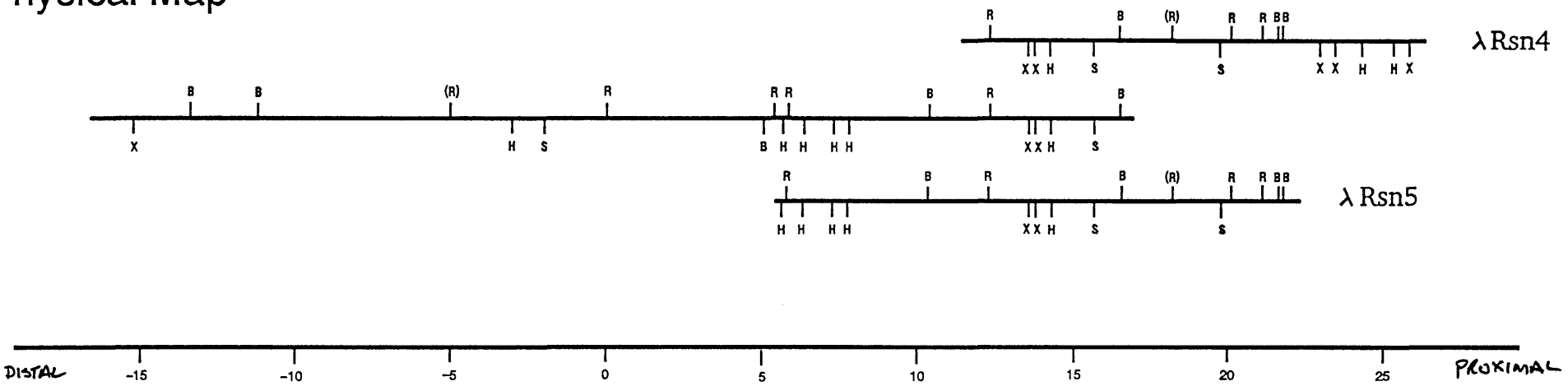
The genetic map of the singed locus is redrawn from Hexter,(1957). Alleles with detectable DNA lesions are shown in dark letters while those without detectable DNA lesions are in light letters. The position of the genetic map above the DNA map approximately matches the sites of DNA disturbances in individual alleles. The DNA map is redrawn from Roiha et al.,(1988) and was derived by restriction mapping overlapping phage containing inserts from Canton S. The co-ordinate system is in kilobases. Restriction enzyme sites are abbreviated as follows R= EcoRI, H= HindIII, S= SalI, B= BamHI, X= XhoI. Restriction enzyme sites shown in brackets are polymorphic. The two phage λRsn4 and λRsn5 were isolated from the wildtype (Oregon R) library of Mariani et al., 1985 using the region +12.5 to +17.5 as a probe.

Figure 3.1.1

Genetic Map



Physical Map



This analysis confirmed the site and structure of DNA lesions present in $\underline{sn}^{X2}$,$\underline{sn}^{3}$,$\underline{sn}^{34e}$ and $\underline{sn}^{36a}$ ant that no lesions were detectable in $\underline{sn}^{1}$, $\underline{sn}^{4}$ and $\underline{sn}^{5}$ (data not shown). Different or new results were obtained in the analysis of $\underline{sn}^{2}$ and Dp(1:3)$\underline{sn}^{13a}$.

Figure 3.1.2 - Table of sn mutants analysed by DNA blotting

Allele	Genotype	Source	Lesion
$\underline{sn}^{1}$	ClB/w^{e} sn^{1}	BG	-
$\underline{sn}^{2}$	w^{co} sn^{2}	BG	+5.8 to +12.2
$\underline{sn}^{3}$	y sn^{3} v	KOH	-0.9 to 0.0
$\underline{sn}^{4}$	sn^{4}	BG	-
$\underline{sn}^{5}$	$y^{2}w^{a}sn^{5}$B/FM6	BG	-
$\underline{sn}^{34e}$	sn^{34e}	BG	+2.4 to +4.3
$\underline{sn}^{36a}$	C(1)DX y f/sn^{36a}	BG	-1.2 to 0.0
$\underline{sn}^{X2}$	106/FM7	MS	-1.9 to -1.2
Dp(1:3)$\underline{sn}^{13a1}$	-	BG	0.0 to +5.5
Ct-S	wt	BG	-
Or-R	wt	BG	-

Key: BG = Mid-America Stock Centre, Bowling Green, USA.
MS = Mike Simmons, University of Minnesota, USA.
KOH = Kevin O'Hare, Imperial College, London, U.K.
A description of the genetic markers and mutations listed above can be found in Lindsley and Grell,(1968).

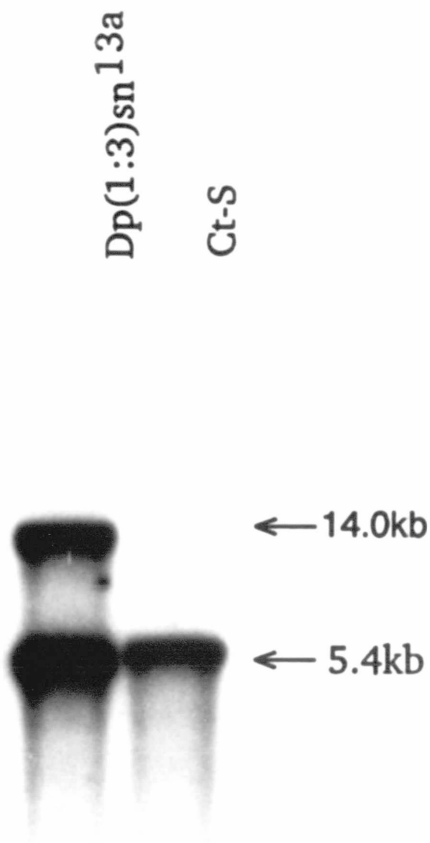
The stock Dp(1:3)$\underline{sn}^{13a1}$ is derived from the now extinct translocation stock T(1:3)$\underline{sn}^{13a1}$ (Valencia, 1966) The translocation in this stock disrupts the sn gene by transferring part of the 1^{st} chromosome onto the 3^{rd} chromosome. DNA blotting analysis of Dp(1:3)$\underline{sn}^{13a}$ locates the novel 1^{st}:3^{rd} chromosome junction within the interval 0.0 to +5.5 Figure 3.1.3. This is consistent with previous localisations of the sn gene.

The DNA blotting analysis of $\underline{sn}^{2}$ located a complex alteration within the region +5.8 to +12.2 Figure 3.1.4.

Figure 3.1.3 - DNA blotting analysis of Dp(1:3)$\underline{sn}^{13a1}$

The panel shows a DNA blot of <u>Eco</u> R1 digested Dp(1:3)$\underline{sn}^{13a1}$ and Ct-S DNA probed with the region 0.0 to +5.4. The additional band seen in the Dp(1:3)$\underline{sn}^{13a1}$ lane is the novel 1^{st}:3^{rd} chromosome junction present in this stock.

Figure 3.1.3



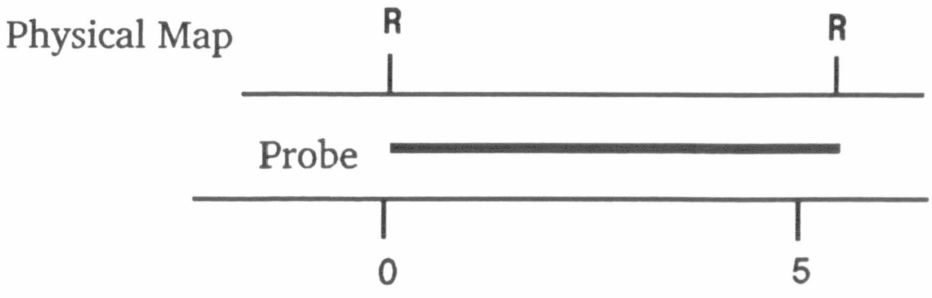
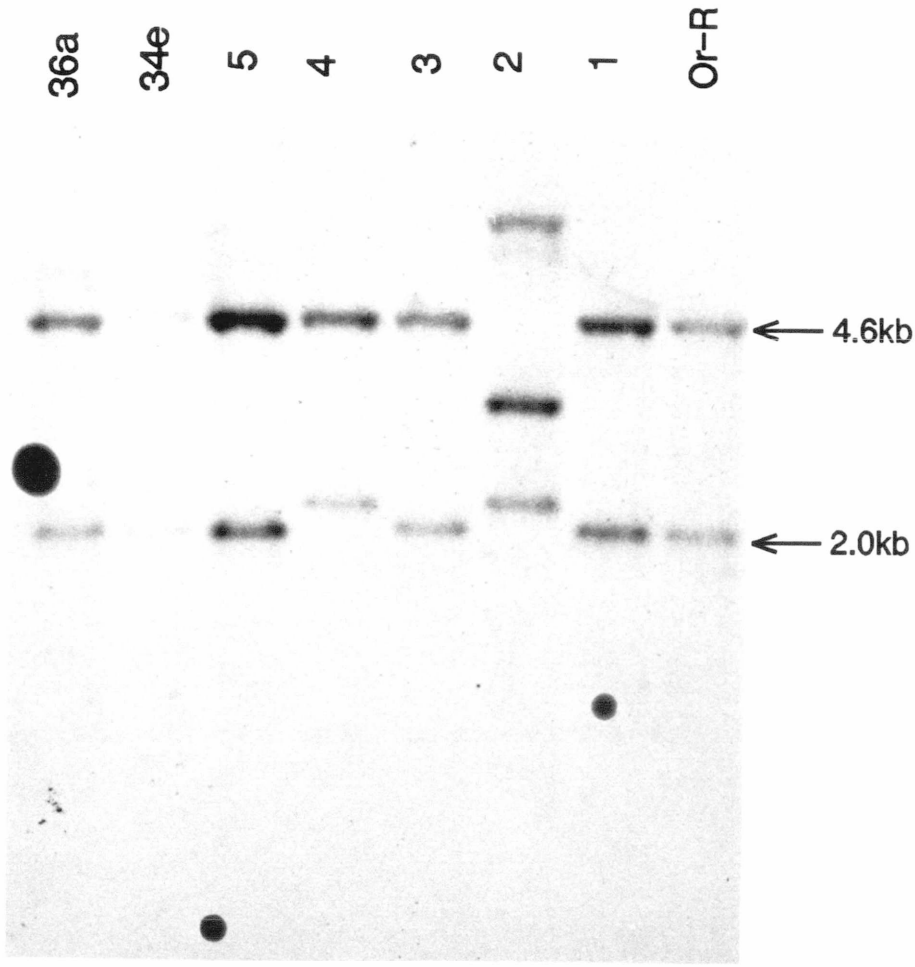
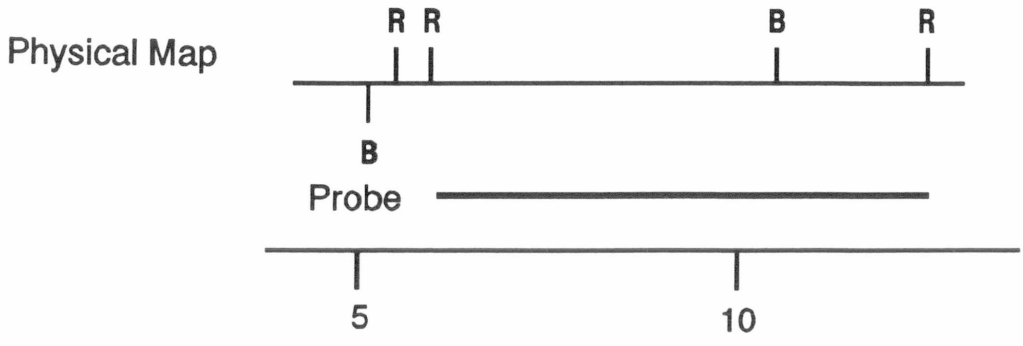


Figure 3.1.4 - Detection of a DNA lesion in $\underline{sn}^2$

The panel shows a DNA blot of 7 samples digested with both Eco R1 and Bam H1 restriction endonucleases. The probe (+6.1 to +12.7) detects a lesion within the interval +6.1 to +10.5 in $\underline{sn}^2$ DNA. The size alteration seen in $\underline{sn}^4$ is discussed in Section 3.1 and Figure 3.1.5. Allele names are abbreviated as follows: Or-R = Oregon-R, $\underline{sn}^1 = 1$, $\underline{sn}^2 = 2$, $\underline{sn}^3 = 3$ $\underline{sn}^4 = 4$, $\underline{sn}^5 = 5$, $\underline{sn}^{34e} = 34e$, $\underline{sn}^{36a} = 36a$. These abbreviations are also used in Figure 3.1.5.

Figure 3.1.4





Further DNA blotting experiments identified this change as either a large insertion or non-cytological rearrangement within the interval +8.0 to +10.5 (data not shown).

The detection and localisation of a lesion in $\underline{sn}^2$ is contrary to the results of Roiha et al.,(1988). To explain the detection of a lesion in one sample of $\underline{sn}^2$ and not in a second sample the possibility of genetic contamination must be considered. Examples of contaminated sn stocks are seen in Drosophila Stock Lists where some female sterile alleles of sn are available in apparently homozygous fertile stocks. The $\underline{sn}^2$ stock analysed in this study was checked for alterations in genetic markers present on the 1^{st} chromosome. No differences were found. The sn phenotype of this stock is identical to the description of $\underline{sn}^2$ in lindsley and Grell,1968. It is therefore assumed that the DNA alteration detected in the region +8.0 to +10.5 represents the true $\underline{sn}^2$ mutation.

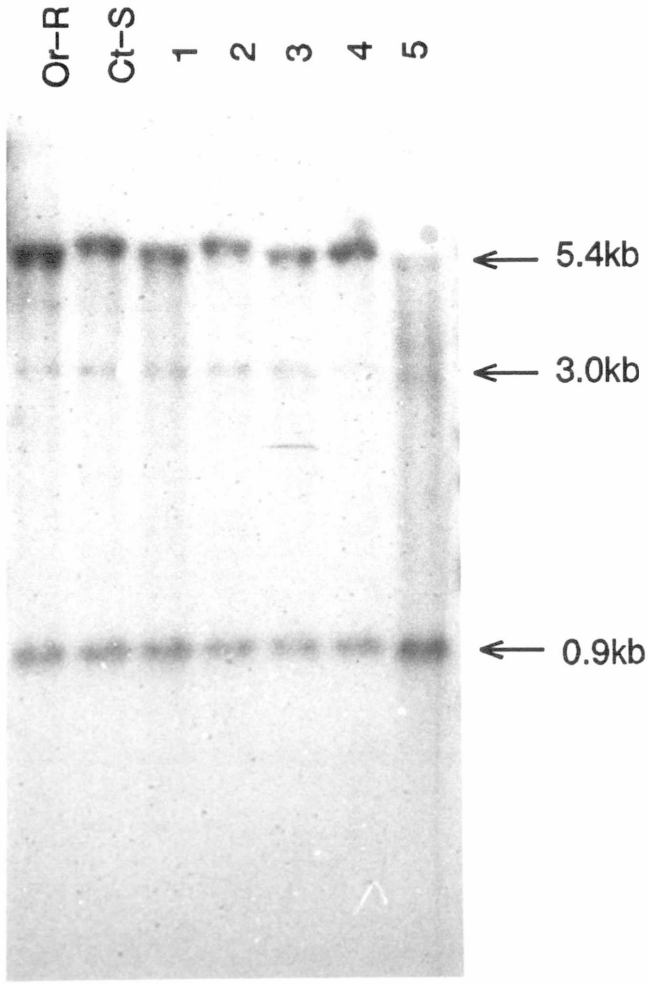
In addition to the $\underline{sn}^2$ lesion, Figure 3.1.4 shows the interval from +10.5 to +12.5 to be larger in $\underline{sn}^4$ than in wildtype Oregon-R (Or-R) DNA. Investigation of this size difference lead to the discovery of a fragment length polymorphism within this region. The interval from +10.5 to +15.8 varies between two discrete sizes (Figure 3.1.5). This variation can be detected using many different endonucleases and thus is not due to polymorphism of a particular restriction endonulease site (data not shown). The 500bp difference is even observed between wildtype strains (Figure 3.1.5) and thus must not affect the function of the sn gene. Subsequent analysis of the gene structure (see below) suggests that these polymorphisms lie within introns of the sn transcript.

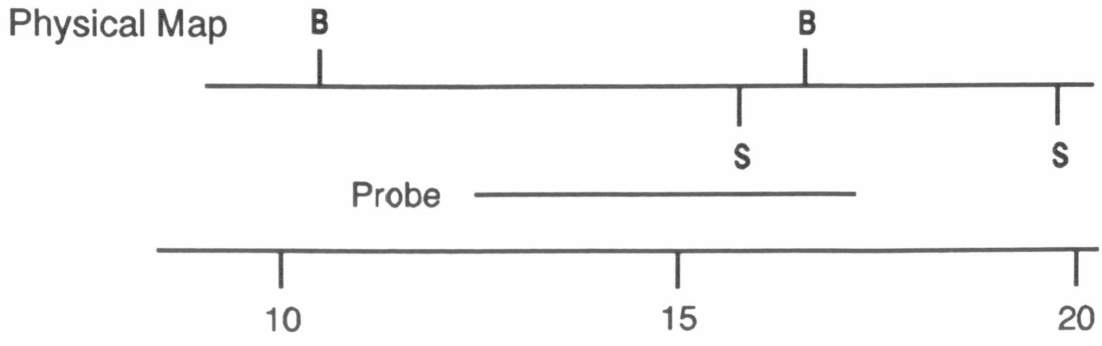
The discovery of a DNA lesion in $\underline{sn}^2$ allows a direct comparison between the site of molecular change and the position of the allele in the genetic map of sn (Hexter, 1957 ;Figure 3.1.1). The pair of distal alleles $\underline{sn}^{36a}$ and $\underline{sn}^3$ have DNA changes in the region -2.0 to +4.0. The central allele containing a detectable DNA lesion, $\underline{sn}^2$ has an alteration in the region +8.0 to +10.5. This suggests that the mutations in the proximal alleles, $\underline{sn}^1$, $\underline{sn}^5$ and

Figure 3.1.5 - Identification of an RFLP within the sn locus

The panel shows a DNA blot of 7 samples digested with both Bam H1 and Sal 1 restriction endonucleases. The probe (+12.7 to +17.5) detects a fragment length polymorphism of 500bp within the interval +10.5 to +15.8. Note the RFLP is present in the wildtype stain Ct-S but absent from the Or-R wildtype stain.

Figure 3.1.5





sn^{50k}, lie to the right of +8.0.

It now seemed possible that the gene extended beyond the right end of the cloned interval (+17.5). Thus the cloned region was extended by undertaking a chromosome walk in this direction. A probe from +12.5 to +17.5 was used to isolate homologous phage from a wildtype library (Mariani et al.,1985). Two independent phage λRsn4 and λRsn5 were isolated and mapped (Figure 3.1.1). They extend the originally cloned interval by +9.5kb, from +17.5 to +27.

To complete the DNA blotting analysis mutants shown in Figure 3.1.2 were tested for DNA lesions within interval +17.5 to +27. No alterations were found (data not shown).

3.2 Transcription of the singed region

The analysis of mutant sn alleles by DNA blotting and the subsequent alignment of the genetic map with the physical map suggests that the sn transcription unit lies in the interval -2.0 to beyond +8.0 (Figure 3.1.1). To identify transcripts from this interval the region from -2.0 to 0.0 was hybridised to an RNA blot of poly-A^+ RNA made from various stages of Drosophila development (K.O'Hare pers.comm.). The results (Figure 3.2.1) show 3 transcripts of 3.0, 3.3 and 3.6kb in adult females and embryos. The 3.6kb transcript is also present in adult males and is most abundant in pupae. These transcripts are derived from an interval disrupted in many sn mutants. The presence of transcripts during pupation and in adult females fits with a requirement for sn product during hair and bristle formation and oogenesis. It is concluded that these transcripts are derived from the sn gene.

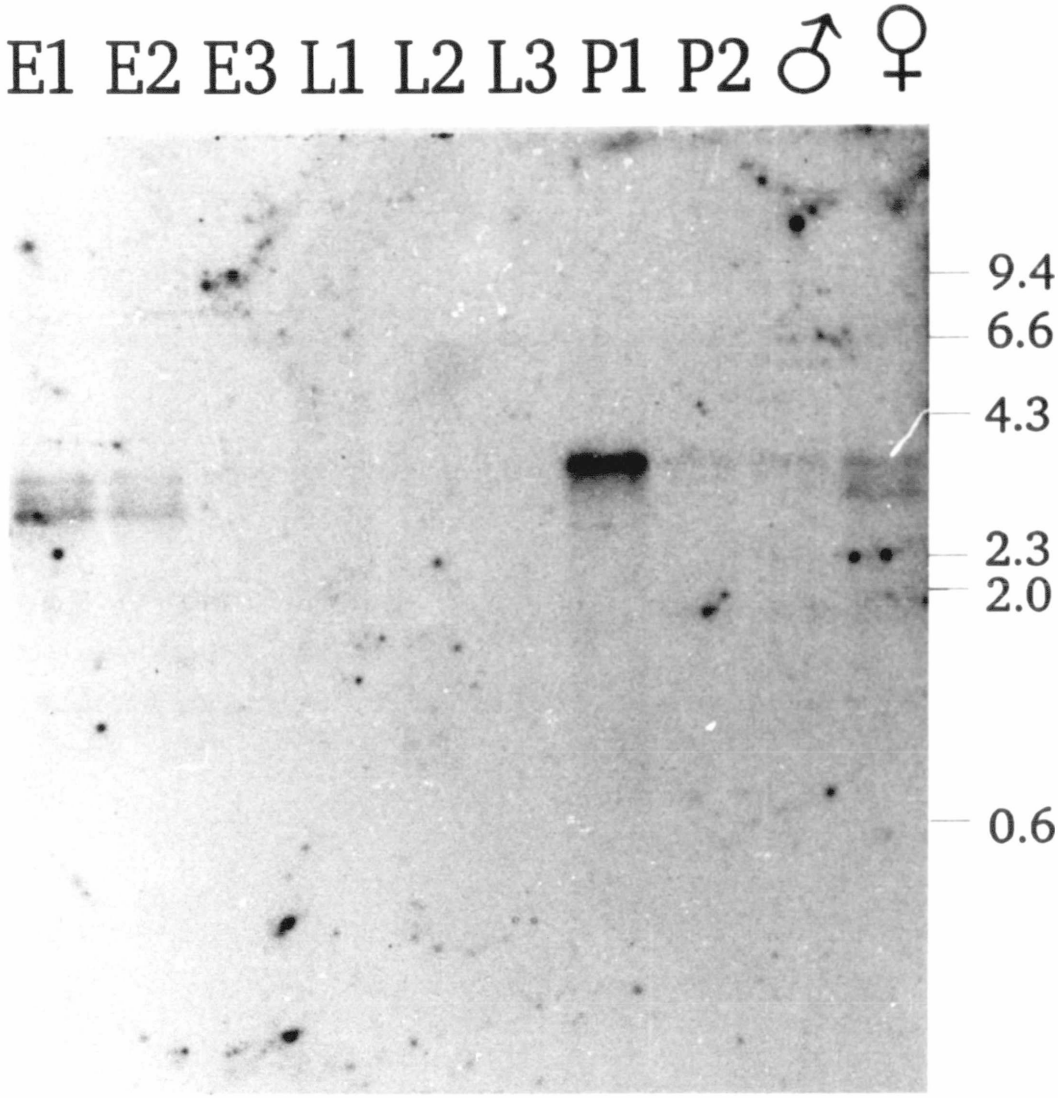
To demonstrate the direction of transcription strand specific probes from the region -2.0 to 0.0 were hybridised to a blot of adult male and female RNA. This showed all 3 sn RNAs to be transcribed from distal to proximal (left to right Figure 3.1.1,O'Hare pers.comm.).

Using a more distal DNA fragment , -5.0 to -2.0 as a

Figure 3.2.1 - Expression of singed transcripts during development.

Poly-A^+ RNA samples from different stages of development were used to produce this RNA blot probed with sequences from the sn locus (-2.0 to 0.0 Figure 1). E1= 0-4hr embryos, E2= 4-8hr embryos, E3= 0-16hr embryos, L1= 1^{st} instar larvae, L2= 2^{nd} instar larvae, L3= 3^{rd} instar larvae. EP= early pupae, LP= late pupae, ♂ = adult males, ♀ = adult females. Lambda HindIII molecular weight markers were used to estimate the size of transcripts. Equivalent quantities of RNA were loaded in each lane as judged by hybridisation to an actin probe (data not shown).

Figure 3.2.1



probe on RNA blots, detects a 5.4 kb RNA that is transcribed in the opposite direction to the 3.0, 3.3 and 3.6kb sn RNAs (K.O'Hare pers.comm.). This transcript is derived from the olfE gene (G. Hassan pers. comm.).

3.3 Extent of the singed transcription unit - RNA analysis

Apart from identifying sn RNAs the above RNA blotting experiments include some information about the extent of the sn transcription unit.

The probe used to detect sn RNAs is only 2.0kb yet the sn transcripts are 3.0, 3.3 and 3.6kb. The sn transcripts must therefore contain sequences from outwith the probe fragment. The adjacent distal genomic interval detects a divergent transcript in RNA blotting experiments. Unless this gene overlaps sn, the sn RNAs must incorporate sequences proximal to (to the right of) 0.0 (Figure 3.1.1).

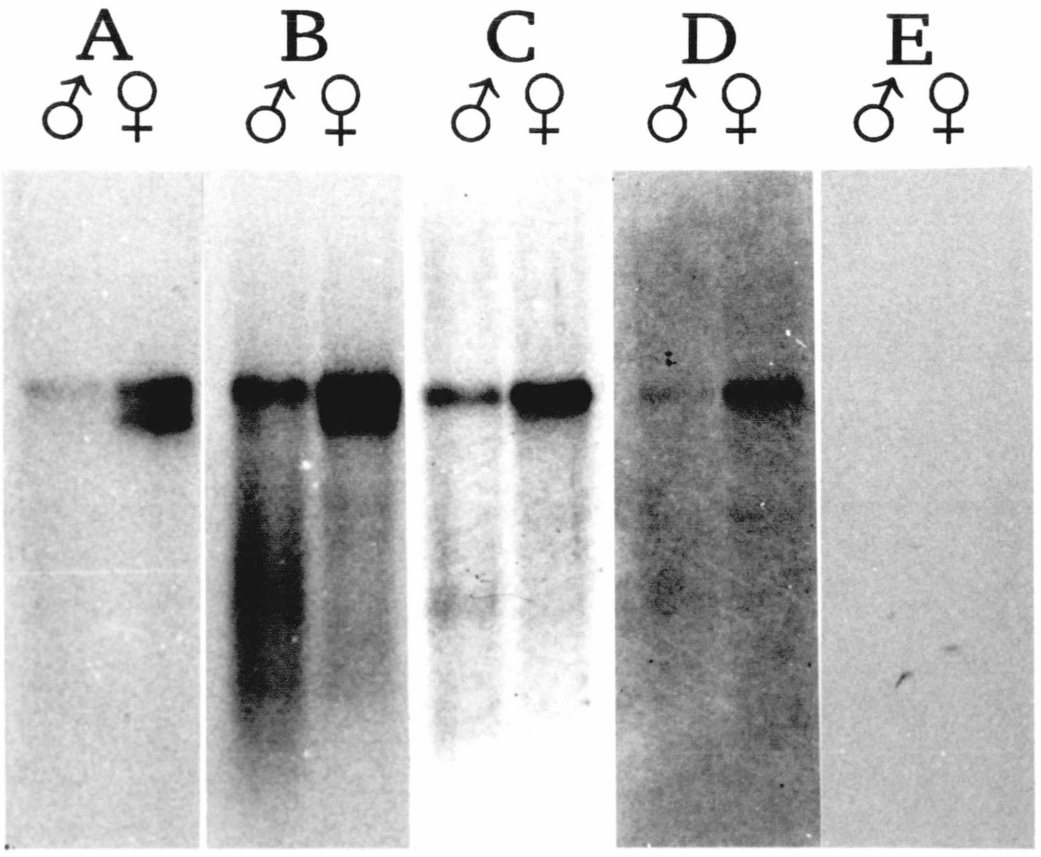
Some data locating the 5' end of the sn transcription unit has also been presented. The interval -2.0 to 0.0 detects all 3 sn RNAs but the region -5.0 to -2.0 detects no sn transcripts on RNA blots. Transcription of the sn RNAs is from distal to proximal. The 5' ends of all 3 sn RNAs probably lie within the interval -2.0 to 0.0.

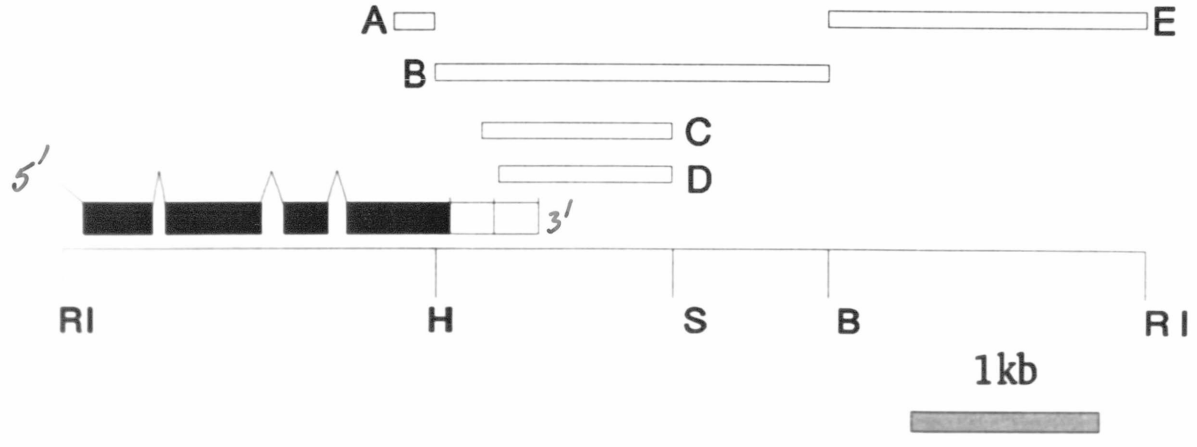
To map the 5' ends of the 3 sn RNAs a variety of short subclones from the -2.0 to 0.0 interval were hybridised to RNA blots carrying samples of male and female RNA. No probe was found which discriminated between the three RNAs. They either detected all three RNAs or none of them (K.O'Hare pers.comm.). The most interesting results were obtained by comparing the hybridisation pattern of two probes whose proximal limits differed by only 100 bp (K.O'Hare pers.comm.). The first probe (-1.8 to -0.6) detected all 3 RNAs, while the second probe (-2.7 to -0.7) did not detect any sn transcripts. These results suggest that the 5' ends of all three RNAs map to the interval -0.7 to -0.6. If the 5' ends of all three RNAs are within such a small region, it suggests that the

Figure 3.3.1 - Localisation of the 3'ends of sn transcripts

Panels A to E show RNA blots of male and female RNA hybridised to the corresponding strand specific probe shown in the diagram below (+12.3 to +18.3). Equivalent quantities of poly-A^+ RNA were loaded in each lane as judged by hydridisation of an actin probe (data not shown).

Figure 3.3.1





transcript size differences must be due to different 3' ends, and/or different splicing.

To map the 3' ends of the RNAs, a series of more proximal probes were used. Again, probes from within the interval -0.6 to +14.3 either hybridised to all three RNAs, or to none of them (data not shown). Differences were found using probes from +14.0 to +18.3 (Figure 3.3.1). Probe A from +14.0 to +14.3 hybridises to all 3 sn transcripts (Figure 3.3.1A). The adjacent genomic interval, probe B from +14.3 to +16.6, hybridises most strongly to the 3.6kb RNA, and less strongly to the 3.3kb RNA and least strongly to the 3.0kb RNA (Figure 3.3.1B). A probe from the next genomic interval, +16.6 to +18.3 (Figure 3.3.1E) does not hybridise to any sn RNAs.

To investigate the graduated hybridisation pattern seen with probe B, smaller probes were used. Probe D (+14.6 to +15.6) detects only the 3.6kb transcript on RNA blots (Figure 3.3.1D). Probe C (+14.5 to +15.6) detects the 3.6 kb RNA, does not detect the 3.0 kb and appears not to detect the 3.3 kb RNA (Figure 3.3.1C). The analysis of cDNAs and the genomic DNA sequence (see below), suggests that the 3' end of the 3.3kb RNA maps to +14.55, so probe C might be expected to hybridise to the 3.3kb RNA. Hybridisation of the probe to the more abundant 3.6 kb RNA (and any degraded 3.6kb RNA) may be obscuring detection of hybridisation to the 3.3kb RNA. Furthermore, the short region of homology (57bp) and the high A-T content (46/57) of the expected hybrid between probe C and the 3.3kb RNA make it impossible to draw an absolute conclusion. Nonetheless, these results suggest that the 3'ends of the RNAs map to different positions within the interval +14.3 and +16.6.

3.4 Extent of the singed transcription unit - cDNA analysis

Using RNA blotting the approximate limits of the sn transcription unit have been defined. To characterise the structure of the transcription unit a series of cDNAs were isolated and analysed.

Initially adult female, embryo and pupal cDNA libraries (Poole et al.,1985) were screened with a probe covering the proximal region of the gene -2.7kb to +5.7kb. Phage carrying cDNA inserts homologous to this probe were plaque purified and DNA prepared from the each phage isolate.

A total of 18 cDNA clones were isolated from four different libraries. Two of these clones were from the olfE gene and will be discussed later. Only 10 of the 16 sn cDNA inserts were longer than 2.0kb. Eight of these clones were mapped and where ever possible the terminal DNA sequences of each insert were determined. The restriction enzyme maps of these clones can be aligned with that of the largest clone, P5 (Figure 3.4.1). The maps of these cDNAs are colinear with the map of P5. Clones which are smaller than P5 appear to be terminally deleted at one or both ends of the insert with respect to P5.

The regions of genomic DNA corresponding to the ends of the P5 cDNA were mapped by hydridising it to DNA blots of the cloned genomic interval . This placed the 5' end of P5 at -0.7 and the 3' end at +14.4 (data not shown). Other workers in the laboratory had previously sequenced the interval from -3.0 to 0.0 (A.Driver and K.O'Hare pers. comm.). The 5' end of P5 and terminal sequences from other cDNAs were compared to this interval. Four independent cDNAs, 2 embryonic and 2 pupal, initiate at the same location as P5 (Figure 3.4.2). All other cDNAs have ends which map within P5 (Figure 3.5.2). This data suggests that all 3 sn RNAs have the same site for the initiation of transcription at -0.7. The identification of a single transcript initiation site is consistent with the RNA analysis.

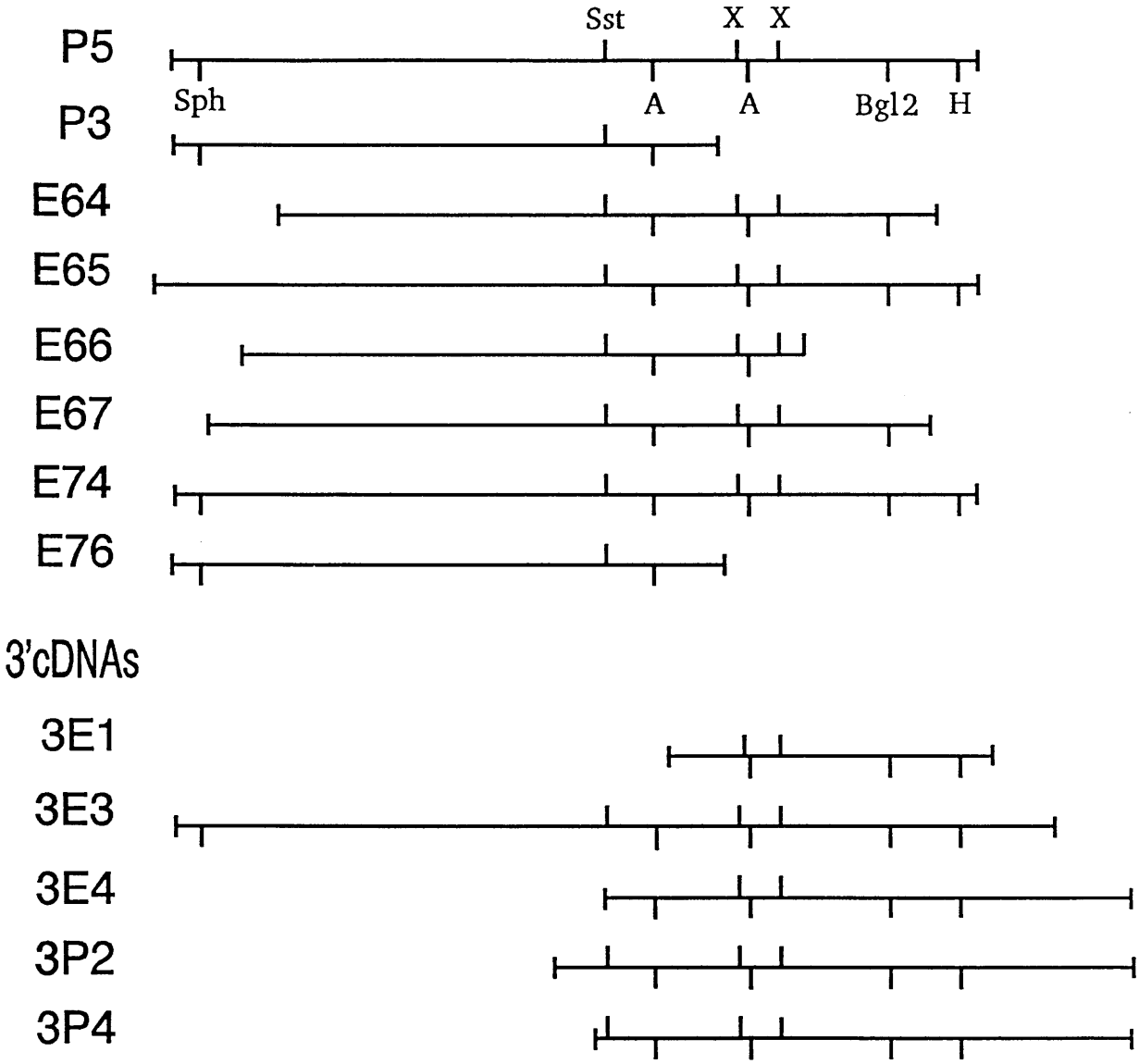
Of all the cDNA clones which start at -0.7, P5 is the only cDNA not to start with a mismatch between cDNA and genome (Figure 3.4.1). Clones P3, P7,E74 and E76 all start with a G mismatched to a C in genomic DNA. It is believed that this nucleotide is a copy of the capping G nucleotide enzymatically added to the 5' end of mRNA molecules. Quite

Figure 3.4.1 - Alignment of cDNA maps

Restriction maps of cDNAs are aligned with the physical map of clone P5. The 5′ end of each clone is to the left. Compared to P5 the upper group of cDNAs are terminally deleted from one or both ends of the insert. The 3′cDNAs have 3′ ends which extend beyond the end of P5. Their 5′ ends all map within the P5 clone. Apart from the terminal deletions or extensions all maps are collinear with P5. This suggests that alternate splicing does not occur within sn transcripts. The bands across the end of each clone represents an EcoR1 restriction site added during cDNA cloning. The restriction map of P5 uses the following abbreviations Sph = Sph1, Sst = Sst1, A = Apa1, X = Xho1, Bgl2 = Bgl2, H = Hind111. The vertical bars on other cDNAs are aligned with their corresponding restriction sites on the P5 cDNA map.

Note: The 190bp of sequence at the 5′ end of clone E65 is not homologous to any part of the cloned sn interval (data not shown). The addition of the 190bp mismatch sequence is believed to be an artifact of cDNA synthesis and cloning.

Figure 3.4.1


Sst
X
X
P5
Sph
A
A
Bgl2
H
P3
E64
E65
E66
E67
E74
E76
3'cDNAs
3E1
3E3
3E4
3P2
3P4
1kb

how this 5′ to 5′ linked nucleotide is copied by reverse transcriptase is unknown, but the presence of this G nucleotide in a cDNA is considered to be diagnostic of a true 5′end.

The probe used to isolate the singed cDNAs also contained sequences from within the olfE transcription unit. Two olfE cDNAs P4 and E75 were isolated and their 5′ ends mapped to -2.5. Terminal sequence information shows the two cDNAs to start within one base pair of each other (Figure 3.4.2). Clone E75 starts with a mismatched C nucleotide. Since the RNA is transcribed in the opposite direction this C nucleotide corresponds to a G on the coding strand. As described above it is assumed that this nucleotide represents the 5′ cap present on mRNA molecules. If the olfE transcription unit starts at -2.5 then the 5′ intergenic gap between singed and olfE is 1621bp (Figure 3.4.2).

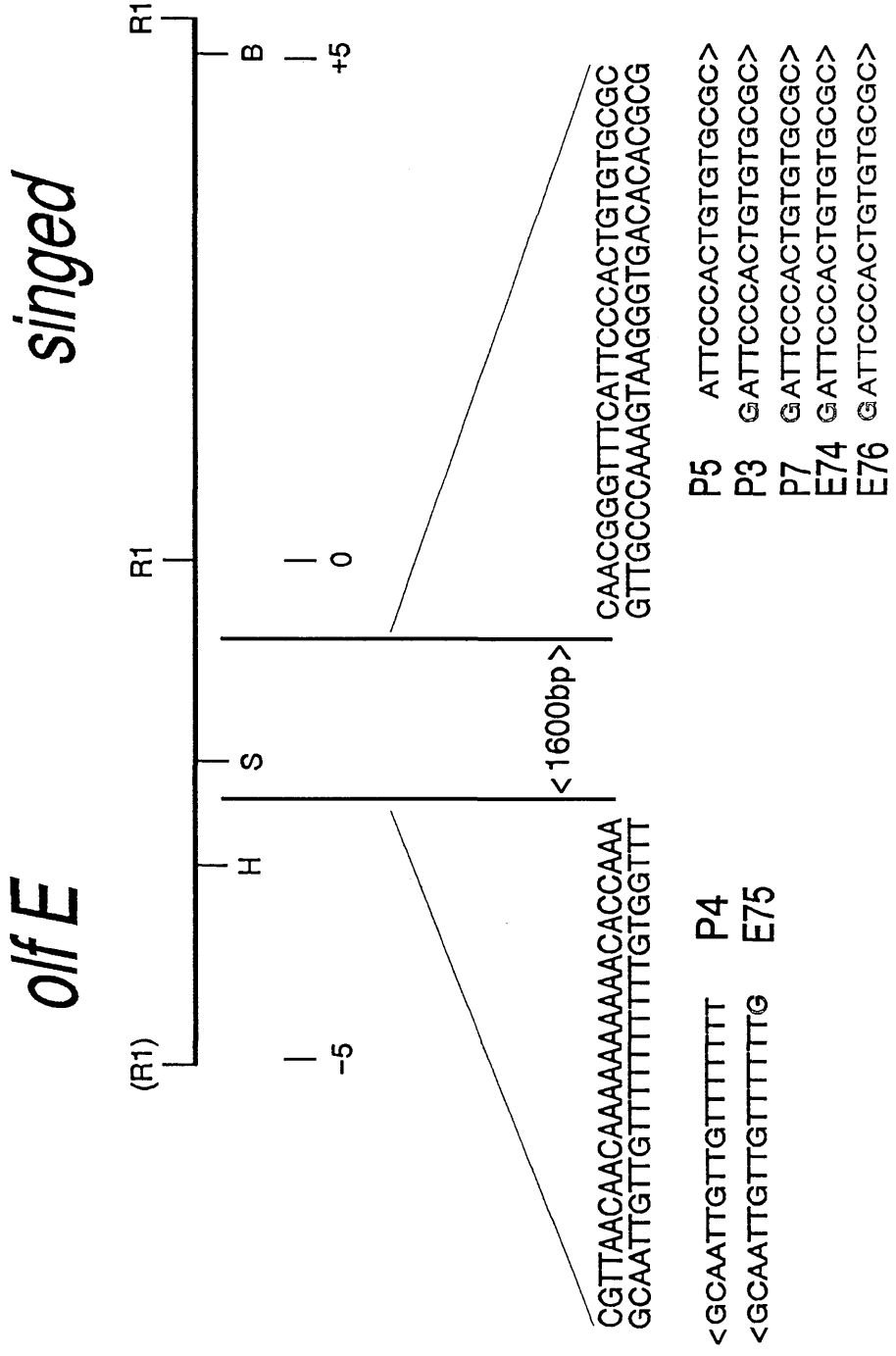
DNA blotting experiments show that the 3′ end of P5 maps to +14.4 (data not shown). As sequences from beyond +14.4 are present in all 3 sn RNAs (Figure 3.3.1) clone P5 must lack sequences from the 3′ end of the transcript. Subsequent analysis of P5 and the genomic sequence of this interval shows that P5 is terminated by a stretch of oligo-A (A_{16}) which is longer than the corresponding run of genomically encoded A residues (A_{14} - Figure 3.5.2C). One explanation for this anomalous 3′ end is that internal priming has occurred on the A rich region during cDNA synthesis rather than by priming on the 3′ poly-A.

To isolate cDNAs containing the 3′ end of the sn transcription unit, the cDNA libraries described above were screened using a probe from the genomic interval +14.3 to +15.7. Five clones were isolated which hybridise to both P5 and this genomic interval. These cDNA clones are collectively referred to as the 3′ cDNAs. The 5′ ends of the 3′ cDNAs fall within the sequence of P5, and their restriction enzyme maps show no differences to that of P5 except that they extend beyond the 3′ end of the P5 cDNA (Figure 3.4.1). The sequence analysis shows that the 3′ ends of the 3′ cDNA clones terminate in stretches of

Figure 3.4.2 - Intergenic region between sn and olfE

Figure 3.4.2 shows a DNA map of the intergenic region between sn and olfE. The transcriptional start site of each gene is marked and the DNA sequence around this point shown below the DNA map. Aligned with the genomic sequence are the 5′ sequences of cDNAs believed to indicate the start site of each transcription unit.

Figure 3.4.2


olf E
singed
(R1)
R1
H
S
R1
B
-5
0
+5
<1600bp>
CGTTAACAACAAAAAAAAAACACCAAA
GCAATTGTTGTTTTTTTTTTGTGGTTT
CAACGGGTTTCATTCCCACTGTGTGCGC
GTTGCCCAAAGTAAGGGTGACACACGCG
<GCAATTGTTGTTTTTTT P4
<GCAATTGTTGTTTTTTTG E75
P5 ATTCCCACTGTGTGCGC>
P3 GATTCCCACTGTGTGCGC>
P7 GATTCCCACTGTGTGCGC>
E74 GATTCCCACTGTGTGCGC>
E76 GATTCCCACTGTGTGCGC>

oligo-A and correspond to 3 sites in the genomic DNA sequence (Figure 3.5.2C). The separation of these 3 sites (~250 bp) is consistent with the process of polyadenylation being used to generate the 3 sizes of sn RNA.

From the cDNA and RNA analysis reported so far it can be concluded there is a single site for initiation of transcription of the 3 sn RNAs at -0.7. Transcripts terminate at one of three alternate polyadenylation sites at +14.4, +14.7 and +15.0. This produces the RNA size differences and transcripts do not otherwise differ in their intron-exon structures.

3.5 Structural analysis of the singed gene

The first step in analysing the intron-exon structure of sn at a nucleotide level was to determine the sequence of cDNA P5. To sequence this cDNA small fragments of the insert were subcloned into M13 sequencing vectors. Suitable fragments were generated by either restriction endonuclease digestion or by progressively shortening large fragments using Bal 31 exonuclease. Dideoxy-sequencing reactions were carried out on M13 template preparations and the nucleotide sequence of inserts were assembled into a contig using the Microgenie package of programs (Beckman). The identity of each nucleotide within the P5 contig was determined from at least one reading on each strand.

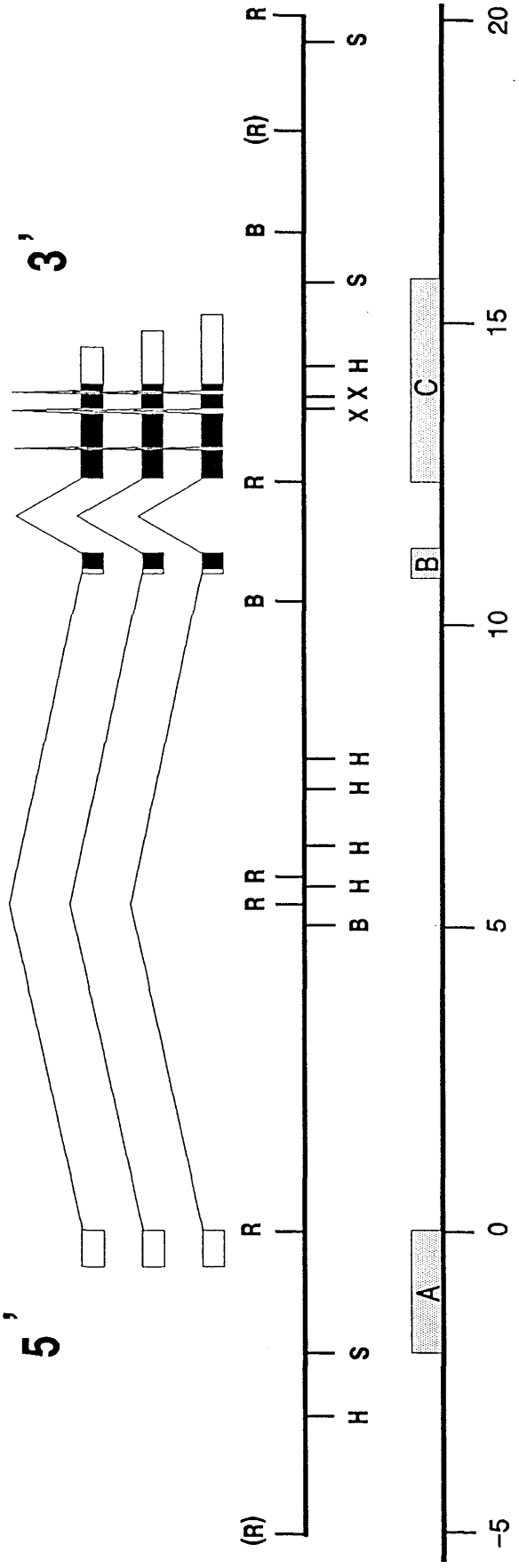
The second part of the structure analysis was to find the location of genomic sequences present in the P5 cDNA. Homologous regions were located by hybridising DNA blots of the cloned interval to a P5 probe. This produced an approximate intron-exon map (data not shown). Genomic intervals corresponding to P5, together with upstream and downstream were subcloned into M13 vectors and sequenced. The structural and sequence information generated is shown in Figure 3.5.1. and Figure 3.5.2.

The sn transcription unit covers approximately 16kb of DNA using 6 exons and developmentally regulated

Figure 3.5.1 - Physical maps of the 3 sn RNAs

The deduced intron-exon structure of the 3 sn RNAs is drawn above the physical map of the sn region (Figure 3.1.1). The 3 transcripts are identical except for the length of the final, exon. The upper transcripts terminates at +14.4 to produce an RNA of approximately 3.0kb. The middle transcript terminates at +14.7 resulting in an RNA of around 3.3kb. The lower transcript map represents the 3.6kb RNA which terminates at +15.0. Open boxes represent non-coding sequence and dark boxes denote protein coding regions. The shaded boxes marked A, B and C on the scale bar locate the sequence intervals in Figure 3.5.2 A, B and C respectively.

Figure 3.5.1



termination signals to produce 3 transcripts of 3.0kb, 3.3kb and 3.6kb Figure 3.5.1. Sequence data from the region upstream of the transcript initiation site shows that standard TATA (Breathnach and Chambon,1981) and CAAT (Benoist et al.,1980) eukaryotic promoter signals are not present (Figure 3.5.2A). The absence of these signal sequences may partly explain the low abundance of sn transcripts. At the proposed transcript initiation site (position 1269 of Figure 3.5.2A), a minor variant of the Drosophila transcript initiation consensus sequence is found (Busslinger et al.,1980). The first exon does have a single ATG codon (positions 1348-1350 in Figure 3.5.2A), but this is followed only 15 codons later by a stop codon. The 5′ exon is followed by a 10.8kb intron. The sequence of all the splice sites conform to consensus splice donor/acceptor sequences (Mount,1982). The most likely position for translation initiation is located in the second exon. A strong ribosome binding site (nucleotides 128-134 of Figure 3.5.2B; Cavener,1987) and a start codon (nucleotides 132-134 of Figure 3.5.2B) are located close to the beginning of the largest open reading frame (ORF) in the P5 sequence. This protein encoding region starts in the second exon and reaches a stop codon in the sixth exon (nucleotides 1625-1267 of Figure 3.5.2C). The 1773 nucleotide long ORF is present in all 3 transcripts and encodes a protein of 57 kdal.

The final exon of sn RNAs varies in length according to the position of polyadenylation. The use of the 3 polyadenylation sites is developmentally regulated in a tissue specific manner (Section 4.1). In most stages of development only the site at 2701 (Figure 3.5.2C) is used. This produces a 3.6kb transcript. In adult females two additional sites at 2457 and around 2219 (Figure 3.5.2C) are used. Distributing the polyadenylation events between the 3 sites produces the 3.0, 3.3 and 3.6kb transcripts seen in adult female RNA.

In an attempt to find a function for the singed gene the nucleic acid sequence of P5 was used to search sequence databases for homologous entries. The UWGCG

Figure 3.5.2 - Sequence analysis of the singed gene.

The interval shown in Section A starts at the SalI site at -2.0 and terminates 47bp beyond the EcoRI site at 0.0. Exon sequences are shown in upper case. The ends of specific cDNA molecules are denoted below the nucleic acid sequence. Section B shows the sequence of the second exon and surrounding DNA, located at +10.8. The start codon is underlined and the single letter code used to indicate the amino-acid sequence of the putative sn protein. Section C covers the interval from +12.5 to +16. The ORF stop codon is denoted by underlining and three stars. Potential polyadenylation signals (AATAAA) near the 3' ends of RNAs are underlined. The location of each sequence interval is shown beneath the physical map of the sn region in Figure 3.5.1.

Figure 3.5.2

Section A

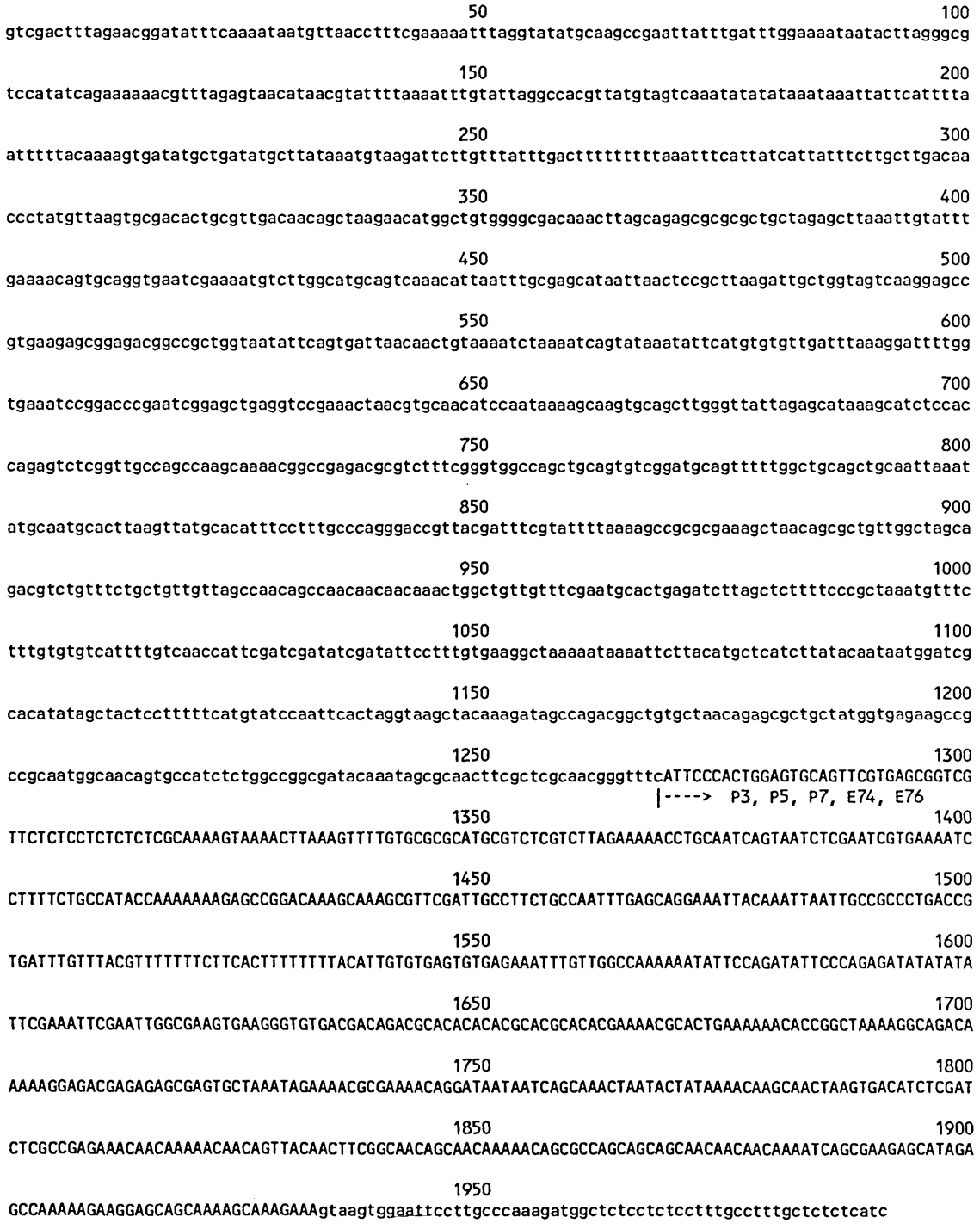


Figure 3.5.2

Section B

```
                                                  50                                                  100
acttaccatactttacccggaaactaatcgtcccactctcacatccttcattgcagCACCATAAGTTCGATCTTCCTTACGCAAAGAAGCCGGATTCCGG

                                                 150                                                  200
AGACAGCCTCAAGATTGCTCCCATCAGCACCATGAACGGCCAGGGCTGCGAGCTGGGCCACAGCAATGGCGACATCATCTCACAGAATCAACAGAAGGGA
                                  M  N  G  Q  G  C  E  L  G  H  S  N  G  D  I  I  S  Q  N  Q  Q  K  G

                                                 250                                                  300
TGGTGGACCATCGGCCTGATCAACGGCCAGCACAAGTACATGACCGCGGAGACCTTTGGGTTCAAGCTCAACGCCAATGGCGCCAGTCTGAAGAAGAAGC
 W  W  T  I  G  L  I  N  G  Q  H  K  Y  M  T  A  E  T  F  G  F  K  L  N  A  N  G  A  S  L  K  K  K

                                                 350                                                  400
AACTGTGGACGCTGGAACCCTCGAACACCGGTGAAAgtgagtagaggttcttaaacagatgggagaaatgttagttagttgggttcttgaaagatttaag
Q  L  W  T  L  E  P  S  N  T  G  E
```

Figure 3.5.2

Section C

```
                                                  50                                                100
gaattcttcgaaaaatgctttccttgcagcaatagctcgccaatgaagcgacttactaatccatggactttaccatatctgtttcagGTATTATCTACTT
                                                                                          S  I  I  Y  L

                                                 150                                               200
ACGATCTCATCTGAACAAGTACCTTTCGGTCGATCAGTTTGGCAACGTGCTGTGCGAGAGCGATGAGAGGGACGCGGGCAGCCGCTTCCAGATCAGCATC
 R  S  H  L  N  K  Y  L  S  V  D  Q  F  G  N  V  L  C  E  S  D  E  R  D  A  G  S  R  F  Q  I  S  I

                                                 250                                               300
AGTGAGGACGGCAGCGGACGTTGGGCGCTGAAGAACGAGTCGCGCGGCTACTTTCTGGGCGGCACTCCGGACAAACTGGTCTGCACGGCCAAGACGCCCG
 S  E  D  G  S  G  R  W  A  L  K  N  E  S  R  G  Y  F  L  G  G  T  P  D  K  L  V  C  T  A  K  T  P

                                                 350                                               400
GTGCCAGTGAGTTTTGGACGGTCCATTTGGCTGCCCGGCCGCAGGTGAATCTGCGCTCCATTGGACGCAAGCGATTCGCCCATTTGTCGGAGTCGCAGGA
G  A  S  E  F  W  T  V  H  L  A  A  R  P  Q  V  N  L  R  S  I  G  R  K  R  F  A  H  L  S  E  S  Q  D

                                                 450                                               500
CGAGATCCATGTGGACGCCAATATTCCTTGGGGCGAGGATACGCTCTTTACGCTGGAGTTCCGTGCCGAGGAGGGCGGTCGCTATGCTCTGCATACGTGC
 E  I  H  V  D  A  N  I  P  W  G  E  D  T  L  F  T  L  E  F  R  A  E  E  G  G  R  Y  A  L  H  T  C

                                                 550                                               600
AACAACAAgtgagtggactcattgccgtaactctacaatgcatcaaacttttttttttccccctttgcagATATCTGAACGCCAATGGAAAATTGCAGGTG
 N  N  K                                                                   Y  L  N  A  N  G  K  L  Q  V

                                                 650                                               700
GTGTGCAACGAGGATTGCCTGTTCAGCGCCGAATATCATGGTGGCCATCTGGCGCTGCGCGATCGTCAGGGTCAGTACTTGTCGCCCATTGGCTCCAAGG
 V  C  N  E  D  C  L  F  S  A  E  Y  H  G  G  H  L  A  L  R  D  R  Q  G  Q  Y  L  S  P  I  G  S  K

                                                 750                                               800
CGGTGTTGAAGTCCCGCTCGTCGTCAGTGACGCGGGATGAGCTCTTCTCGCTGGAGGATTCGCTGCCTCAGGCTTCCTTTATAGCCGGACACAATTTGCG
A  V  L  K  S  R  S  S  S  V  T  R  D  E  L  F  S  L  E  D  S  L  P  Q  A  S  F  I  A  G  L  N  L  R

                                                 850                                               900
ATATGTGAGCGTTAAGCAGGGCGTCGATGTGACGGCCAACCAGGACGAGGTCGGTGAGAACGAGACGTTCCAGCTGGAGTACGATTGGTCGGCGCACCGT
 Y  V  S  V  K  Q  G  V  D  V  T  A  N  Q  D  E  V  G  E  N  E  T  F  Q  L  E  Y  D  W  S  A  H  R

                                                 950                                              1000
TGGGCCCTACGCACCACCCAGGATCGCTACTGGTGTCTGTCGGCGGGCGGTGGCATCCAGGCCACCGGCAATCGTCGCTGTGCCGACGCTCTCTTCGAGC
 W  A  L  R  T  T  Q  D  R  Y  W  C  L  S  A  G  G  G  I  Q  A  T  G  N  R  R  C  A  D  A  L  F  E
    <---| P7
                                                1050                                              1100
TGATCTGGCACGGCGATGGCTCGCTCTCGTTCCGGGCTAACAACGGCAAGTTCTTGGCCACCAAGCGCTCTGGTCATCTGTTTGCCACCTCGGAGTCGAT
L  I  W  H  G  D  G  S  L  S  F  R  A  N  N  G  K  F  L  A  T  K  R  S  G  H  L  F  A  T  S  E  S  I

                                                1150                                              1200
CGAGGAGATAGCCAAGTTCTATTTCTACTTGATCAATAGgtaagggaattgtttacaccattctgtattttccgtttttaattgcattgtttttgttaaa
  E  E  I  A  K  F  Y  F  Y  L  I  N  R
<| P2
                                                1250                                              1300
ttaacagACCAATTCTTGTACTGAAGTGCGAGCAGGGATTCGTGGGCTATCGCACGCCCGGTAACCTGAAGCTCGAGTGCAATAAGGCCACCTACGAAAC
        P  I  L  V  L  K  C  E  Q  G  F  V  G  Y  R  T  P  G  N  L  K  L  E  C  N  K  A  T  Y  E  T
          <---| E76
                                                1350                                              1400
GATTCTGGTGGAGCGGGCCCAGAAGGGCCTGGTGCATCTGAAGGCGCATAGCGGCAAATACTGGCGCATCGAGGGCGAAAGCATCTCGGTGGATGCCGAT
  I  L  V  E  R  A  Q  K  G  L  V  H  L  K  A  H  S  G  K  Y  W  R  I  E  G  E  S  I  S  V  D  A  D

                                                1450                                              1500
GCGCCCAGCGATGGTTTCTTCCTCGAGCTTCGCGAGCCGACCAGGATCTGCATAAGgttggaccgcatttgctgtcatttaaaaacggctgtaaattatg
 A  P  S  D  G  F  F  L  E  L  R  E  P  T  R  I  C  I  R

                                                1550                                              1600
attgttaactaatgccgatcctctatattttagATCGCAGCAGGGCAAGTACCTGGGCGCTACGAAGAACGGGGCGTTCAAGCTGCTGGCTGATGGAACC
                                  S  Q  Q  G  K  Y  L  G  A  T  K  N  G  A  F  K  L  L  D  D  G  T

                                                1650                                              1700
GATTCGGCCACACAGTGGGAGTTCTAGGGCGTGACCACGACCATCCCGAATGGGTTGCTGCACGCGTCATCAGTATCACAATCACACACATACTTCCCAA
 D  S  A  T  Q  W  E  F ***

                                                1750                                              1800
CCAAGCGGGGATTCCCTAATTATTGATCCACGCAGCTGAAGCGCCAAGAAGAGTCGCACAAGAATTTGATATATCCTAAACAATTGCAAAATGAACGAAA
```

Figure 3.5.2

Section C continued

1850 1900
TTAAAATGATAGGAAGGAAACGTTGATAGACCCAAGGAGATACGAGTGAATGAAAAATGAAGAAACAAACTTATATATTTATATATAAAGATCTACGTAT

1950 2000
ACATAGAGAGAGATTTTTTTTAAACAACCTATTTAGCTTTTTATACGTTTTATACGATGAGAATATCTACAAACGAGAACAAGAAATCTTTTAGCGTAGA

2050 2100
CAACAAAACAAAACTTCTGTCGCGAGTCCTTCCGTTTTATCAAATATATCAGCAACTAGGCCAGCAAAAATGAAGAAAGACGTAATGGTTATACAAGCAC

2150 2200
ATCATATATACACCAATATATACATATTCGATTCGTAAAGCTTCACGGAAAAGAAACTAAAAAAAAAAGCGAAGGGAAATAAGAGCAAAAAAAAAAAAAAT
<---| P5, E67, E74
2250 2300
ATGAAAAAAATAAAAATAAATCAACACATATTCCTATAAATTTATTTATATATTTATATACATATTGATATGAACCTAATACATGAACACACACACATAC
<---| 3E1
2350 2400
ACACGACTCCACACGCTCACATAGACACGCATGTCCAAACACACACACACACACACACTCAAGCAAGCAATTTTAATACGCATAAACAAAATTTAGTCGC

2450 2500
TAAGCAAAAACTTAAAAAAAAATTAATCTGTAATAAAAAAACAAAGTCAAAAATTAAACATCAAATCTCTCCGTACGGCGATTTGCAATGGAATCAACTCC
<---| 3E3
2550 2600
ACTATATCTAGCATATGCATATATTGAATTAGTTTTTGTTATTTATTTGCGTCGATTTCTTAATTAATTAATTAGTAATCTTTTTTTTTTTTAATTTGTG

2650 2700
CACTTCAAAACGAAACAAAAACTCAATCCAATTTAGTTGCTAAATCTTAAAATGGAAATTTTGGTATGATACCATATGAAATATATACATATATACCAAT

2750 2800
Tatggttggttacttttatgggttaatgtacaacggaaagctgaatatgcgaagggttgcattattttaacggttacttttttgttcagcatttcaaagt
<| 3P2, 3P4, 3E4
2850 2900
tcaaacttaaatacatgcatctggcatccccatttccgctaactggcaactcggcgccgaatacaaccctgcaatgtactgccattttgaaaaggagtta

2950 3000
gaaattcaagaaataatttcaatataatttatttatcgtgaaataaaatgcaaatgaaccgaaggcataagtttaaatactgttattcataattttgtat

3050 3100
gaaaagcattaaagtgcaatcaatttagtgagcagcagtgaaataacgccatgtaatgcaataagtctaaaggtcgccaacttttgattgaccggctgac

3150 3200
aaatggttggtggtagaccccgccccccctcgcaattttccaacgcccacgggccatgggcacctgcaacgtgtgcggaaattggtgccaactgtcataac

3250 3300
aagatggagcacacttgcatacacacatacacgtacacctctaaatacatatgtacataaatgtggcagtggcaaccaggaacaatggggcgtggctgtg

3350 3400
gggatggaggggtgcgctcacttcgttggcgtccaactgcctccgtcagcaataataattcgccagtcattgacaatgtcggtgccataatggctctcc

3450 3500
actggctccggtggccagtattctctatctttggcttcgtccgcaagccattatatcccatcgccgccacccaattccatccaatccctggcccatttgg

tgcactgcgtcgac

WORDSEARCH programs running on the SERC Daresbury VAX computer were unable to find any significant matches to the P5 sequence in either the EMBL(FULL) version 20 or Genbank version 60 databases.

Computer analysis was also used to examine the putative protein coding region of cDNA P5. The encoded polypeptide has a molecular weight of 57kdal. Using AMERSHAM STADEN PLUS software on an IBM AT personal computer the polypeptide sequence of this protein was examined for hydrophobicity and charge distributions. No unusual structural features were revealed by this analysis. The sn polypeptide sequence was then compared to computer databases of protein sequence. UWGCG WORDSEARCH programs on the SERC Daresbury facility searched SWISSPROT version 11 and NBRF version 21 protein data bases; no obviously homologous entries were found.

3.6 Concluding Remarks

The sn transcription unit was identified using information derived from DNA and RNA blotting experiments. Analysing sn mutants by DNA blotting facilitated comparison between the physical and genetic maps of the sn locus (Section 3.1). DNA lesions were only detected in alleles belonging to the distal and central intra-allelic recombination sites. This is probably due to the asymmetric location of the protein coding sequences within the sn gene (Figure 3.5.1). Point mutations in protein coding regions will often be deleterious while similar lesions in non-coding regions are less likely to disrupt gene function. Mutations which map to non-coding regions are therefore more likely to be associated with large-scale changes to the structure of the gene, such as insertions and deletions. These large alterations in gene structure are easily detectable by DNA blotting (e.g. sn^3 and sn^{36a} at the distal recombination site and sn^2 at the central recombination site). Nearly all point mutations are undetectable by DNA blotting and often map to the protein coding region. Mutations which map to the proximal

recombination site ($\underline{sn}^1$, $\underline{sn}^5$ and $\underline{sn}^{50k}$) are likely to be point mutations which alter the sn protein sequence. This uneven distribution of large-scale lesions between coding and non-coding sequences within a gene was first pointed out for the white locus (Zachar and Bingham,1982).

RNA blotting analysis of sn shows that it produces transcripts of differing size. In adult males and pupae a single 3.6kb RNA is produced. In adult females and embryos 3 RNAs of 3.0, 3.3 and 3.6kb are detected. Further RNA blotting experiments and cDNA analysis shows that the three transcripts differ by using alternate polyadenylation sites to vary the length of their 3' untranslated sequence. Computer analysis of the putative sn polypeptide shows the molecule to contain no unusual structural features and no homologous sequences are found in protein data bases.

In addition to characterising the sn gene the neighbouring distal transcription unit olfE has been studied. Transcripts from the olfE gene initiate less than 1.7kb upstream of the sn RNA initiation site and are transcribed away from sn (K.O'Hare pers. comm.). The divergently transcribed gene produces a 5.4kb RNA present at all stages of development (K.O'Hare pers.comm.). DNA blotting experiments using cDNA clones for this transcript show that it has a short first exon followed by a large intron (data not shown). The developmental transcription pattern of sn and the 5.4kb RNA are different yet the intergenic gap is quite small. Since each gene has a large first intron relatively close to the point of transcript initiation, it is possible that not all of the sequences involved in controlling transcription are located in the upstream intergenic region; some may be present in the nearby introns. The yellow locus is an example of a Drosophila gene which has regulatory sequences within an introns (Geyer and Corses,1987).

CHAPTER 4

Tissue specific expression of *singed*.

4.0 Introduction

To extend the analysis of sn gene function, the tissue specific pattern of sn gene expression has been characterised. This information combined with the phenotypic description of sn mutants allows some assessment of sn function to be made.

There are several different methods for localising gene expression to particular cells. The most frequently used technique, in situ hybridisation identifies cells containing high concentrations of specific mRNA (Zalukar and Enh, 1976). Several attempts were made to localise sn RNA in embryos and ovaries but no conclusive results were obtained. This may have been due to the low abundance and wide dispersion of sn RNAs. An alternative to localising RNA is to detect the presence of gene specific protein(s). Antibodies raised against the gene product are used to locate the protein(s) in either tissue sections or whole mount samples (Wilcox, 1986). Part of the protein, encoded by cDNA P5, was fused to the 3' end of β-galactosidase. The fusion protein was then used to inoculate rabbits in an attempt to make polyclonal antisera which would recognise sn protein. Protein blotting experiments to test whether this polyclonal serum contains antibodies directed against sn protein were inconclusive. The absence of antibodies directed against sn protein prevents the use of this technique.

The methods described so far all detect actual gene products but indirect methods can be used to determine cell specific gene expression. Enhancer-detector systems localise gene expression by detecting the activity of cell specific enhancer sequences (O'Kane and Gehring, 1987). Applied to sn this technique has located cell specific expression in pupal, ovarian and embryonic tissue (see later). However the results obtained from enhancer-detector experiments require careful interpretation. The final method of detecting cell specific gene expression is a combination of genetics and RNA blotting. In these experiments genetic events are followed at a molecular

level revealing detailed gene expression data. Two of such molecular genetic experiments are described below.

4.1 Adult female 3.0kb and 3.3kb RNAs are ovary specific

The combination of genetics and RNA blotting is particularly well suited to the study of gene expression in ovarian tissue. Female sterile alleles of sn disrupt oogenesis thus it is likely that some or possibly all sn transcripts seen in adult females are derived from ovaries. To investigate this, the gonadal dysgenesis (GD) sterility which occurs during P-M hybrid dysgenesis was used to produce flies lacking ovaries (Kidwell and Novy, 1979). This temperature sensitive sterility is due to the genetic ablation of all germ cells at non-permissive temperatures. Thus P-M dysgenic adult females raised at 29°C suffer GD sterility and contain no ovaries, yet genetically identical females from the reciprocal non-dysgenic cross raised at 29°C are fertile with fully developed ovaries. Comparing the RNA samples derived from chromosomally identical flies with or without ovaries allows identification of any ovary specific transcript(s). For low abundance transcripts this genetic method for identifying ovary specific transcripts is significantly easier than collecting ovarian tissue by hand dissection and studying the transcripts present in this material.

Applying this technique to sn identified the 3.0 and 3.3kb RNAs as ovary dependent transcripts (K. O'Hare pers.comm.). RNAs from flies without ovaries, from chromosomally identical flies with ovaries, together with RNA from the male siblings of these flies was analysed for the presence of sn transcripts. An RNA blot showing these results is presented in Figure 4.1.1 (See legend to Figure 4.1.1 for experimental details). In dysgenic females lacking ovaries, the sn probe (-2.0 to 0) detects only the 3.6kb RNA. The other samples show wildtype patterns of sn transcripts. This shows that the 3.0kb and 3.3kb RNA are ovary dependent transcripts. The remaining 3.6kb RNA, present in the dysgenic females, is presumably from

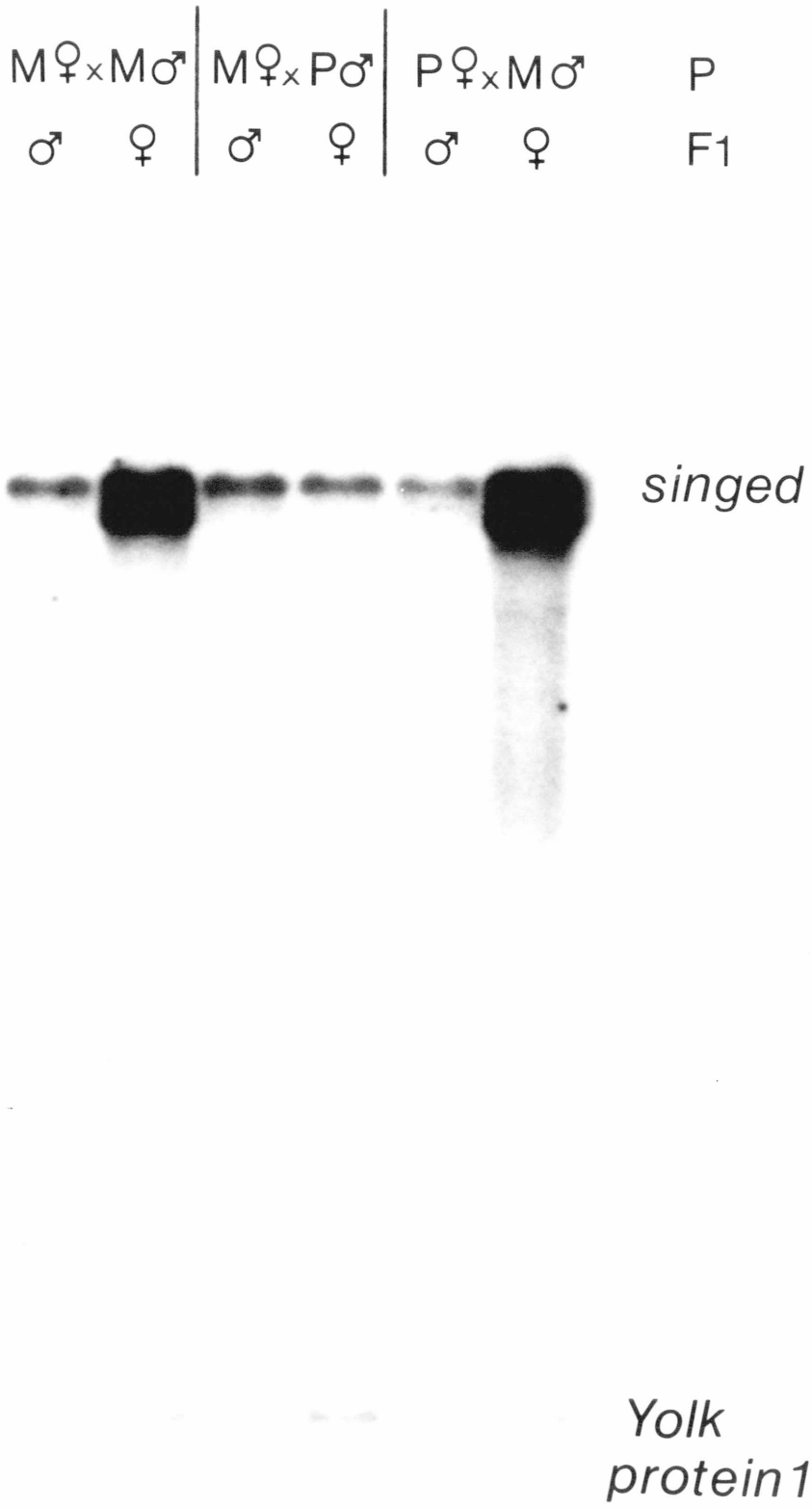
Figure 4.1.1 - 3.0 and 3.3kb RNAs are ovary dependent.

Poly(A)$^+$ RNAs from chromosomally identical females with or without ovaries were analysed. After hybridisation with a sn probe (-0.7 to -0.2 main panel), the blot was reprobed for the yolk protein 1 transcripts to confirm the sex of each sample.

Genetics

The reciprocal crosses between Canton S (wildtype M cytotype) and Harwich (wildtype P cytotype) were made at 29^{o}C. The progeny were allowed to mature at 25^{o}C in fresh bottles for 3 days before harvesting. Examination of these bottles showed that the F1 progeny from the cross of Canton S females with Harwich males were sterile, while the F1 progeny of the cross of Harwich females with Canton S males were fertile. Dissection of females showed that they lacked ovaries, as expected for this dysgenic cross.

Figure 4.1.1



somatic tissue. As a control, the RNA blot was reprobed for the yolk protein 1 RNA (Figure 4.1.1). This gene is not expressed in males, but is expressed in both somatic and germ cells of adult females (Hung and Wensink,1981).

4.2 Maternal inheritance of singed transcripts

The identification of the 3.0 and 3.3kb RNAs as ovary dependent transcripts permitted a re-evaluation of sn developmental expression data. The RNA blot in Figure 3.2.1 shows all 3 sn transcripts to be present in both adult females and early embryos. The presence of the 3.0 and 3.3kb ovary dependent transcripts in early embryos suggests that sn RNAs are laid down in developing oocytes and maternally inherited by early embryos. It is not clear if the sn RNAs found in embryos have a function. From mosaic studies, it is known that sn is required during oogenesis for the formation of viable eggs. What remains impossible to know is whether there is a role for sn after fertilisation. The few eggs laid by a mutant sn^- female sterile fly cannot be rescued by a paternal wildtype sn^+ gene.

To investigate maternal inheritance of sn RNAs, genetics and RNA blotting was used to follow the inheritance of specific sn transcripts. Cloning and sequencing studies show the sn^3 mutation to be a deletion of the region -0.5 to -0.8 (Figure 5.1.4). RNA blotting analysis has shown that normally upstream sequences (with respect to wildtype RNAs) are incorporated into mutant sn^3 transcripts. (Full details of this mutant and its pattern of transcription are described in section 5.1). The upstream sequences can be used as a specific tag for sn^3 RNAs. Using the region from -0.7 to -1.0 as a probe (probe A Figure 4.2.1), sn^3 RNAs can be specifically detected even in heterozygous sn^3/sn^+ adults (data not shown). Conversely using as a probe sequences from within the interval deleted in sn^3 (probe B Figure 4.2.1), wildtype sn^+ RNAs can be detected in sn^3/sn^+ individuals. Finally the fact that homozygous sn^3 females are fertile allows

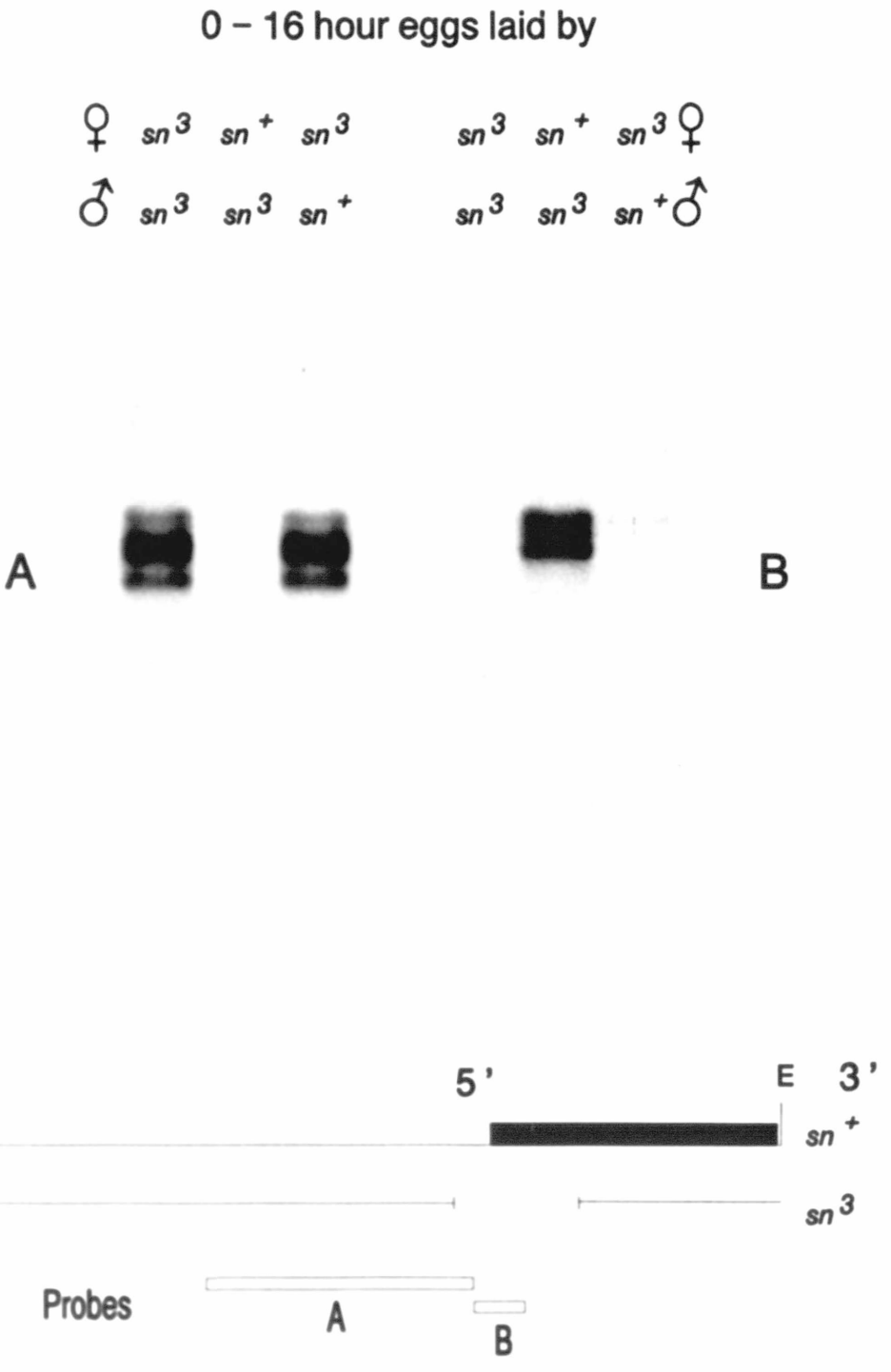
Figure 4.2.1 - Maternal inheritance of singed RNAs.

Poly-A$^+$ RNAs were prepared from 0-16hr embryos produced by the crosses indicated above each lane. In panel A, probe A only detects RNAs in embryos laid by sn^3 mothers, irrespective of the genotype of the fathers. In panel B the same RNA blot is reprobed with a sn$^+$ specific probe (probe B), the triplet of sn RNAs are only detected in embryos laid by sn$^+$ mothers irrespective of the genotype of the fathers. There is a faint band at 3.6kb in the lane of embryos derived from sn^3 females with sn$^+$ fathers. This RNA may either be a very low abundance zygotic sn transcript or some form of contamination.

Genetics

The reciprocal crosses between Canton S and y sn^3 v were made at 25^{o}C. Embryos were collected after overnight laying (0-16 hour embryos) and at the end of the day (0-8 hour embryos), and pooled. Some embryos were allowed to develop and the genotypes were as expected. The heterozygous sn$^+$/sn^3 and sn^3/sn$^+$ females were harvested and shown to have both wildtype and mutant RNAs (data not shown).

Figure 4.2.1


0 – 16 hour eggs laid by
♀ sn3 sn+ sn3
♂ sn3 sn3 sn+
sn3 sn+ sn3 ♀
sn3 sn3 sn+ ♂
A
B
S
5'
E
3'
sn+
sn3
Probes
A
B

the inheritance of $\underline{sn}^3$ transcripts to be followed from adult females to embryos.

RNA was extracted from embryos derived from the two reciprocal crosses of $\underline{sn}^+$ with $\underline{sn}^3$. Collection were made overnight (0-16hrs) or during the day (0-8hrs) and pooled. The RNA samples were then blotted and probed with first the $\underline{sn}^3$ specific probe and then the $\underline{sn}^+$ specific probe. The results Figure 4.2.1 show that sn RNAs found in the embryos always correspond to the maternal genotype. This proves that all three sn transcripts are maternally inherited.

A small amount of 3.6kb $\underline{sn}^+$ RNA is detected in eggs laid by $\underline{sn}^3$ mothers. The fact that the 3.0 and 3.3kb RNAs are not detected is consistent with this being somatic zygotic sn expression.

The fact that all 3 sn transcripts are maternally inherited shows that in adult females all 3 sn RNAs (the 3.6kb RNA in addition to the 3.3 and 3.0kb transcripts) are made in ovaries. Presumably they are made in the germline derived nurse cells and are contributed to the developing oocyte. Thus the 3.6kb RNA is made in both germline and somatic cells.

4.3 Enhancer detector systems : P[lacZ] mutagenesis

In order to identify individual cells as expressing sn, enhancer-detector P[lacZ] transposons (O'Kane and Gehring, 1987) were used to mutagenise the sn locus. The insertion specific pattern of β-galactosidase staining is thought to reflect the activity of enhancers in flanking sequences, and so be a guide to the transcription of nearby genes.

Certain aspects of this system make it particularly suitable for the study of sn expression. The sn locus contains a hotspot for P-element insertions (Roiha et al. 1988). Thus P[lacZ] insertions at sn can be obtained at a relatively high frequency. The mutagenesis scheme shown in Figure 4.3.1 uses the alteration of bristle phenotype to

detect P[lacZ] insertions at the sn locus. The efficiency of P-element transposition in and out of the sn gene is so high that care had to be taken not to produce secondary insertion/excision events in mutagenised $sn^{P[lacZ]}$ flies. This had apparently happened in a sn mutant generated by J.Merriam (UCLA) and analysed at Imperial College. To avoid this, the source of P-transposase (Δ 2-3) was removed after only one generation (Figure 4.2.1).

To investigate sn expression during hair and bristle formation it is necessary to study gene expression in pupae. The physical structure of pupae makes this analysis particularly difficult. The soft loosely structured tissue of the pupa is enclosed within a hard, impermeable pupal case. Most manipulations of the pupae result in structural damage and cut sections often have poor morphology. β-galactosidase staining of pupae requires only slight surgical manipulation thereby facilitating the study of pupal gene expression.

The final bonus of using the enhancer-detector system is that the β-galactosidase encoded by P[lacZ] transposons acts to enzymatically amplify the signal produced by reporter gene activation. Localisation of the P-element-β-galactosidase fusion protein to the nucleus concentrates the staining activity and increases the resolution of the system. If the low abundance of sn transcripts reflects weak sn enhancer activity, these aspects of the P[lacZ] system could increase the signal to a detectable level.

Using the crossing scheme outlined in Figure 4.3.1 8140 mutagenised males were visually screened for an alteration in bristle phenotype. Two independent P[lacZ] insertions into the sn gene were generated. Before studying these mutants each allele was subjected to an outcrossing scheme shown in Figure 4.3.2. This ensured the allele was present in a defined genetic background. Both mutants have extreme bristle phenotypes and fail to complement sn^3. The P[lacZ] insertion in each mutant was localised to the P element hotspot (-0.7 to 0.0) by DNA blotting (data not shown). In all tissues and at all times

Figure 4.3.1 - P[lacZ] mutagenesis crossing scheme

In the parental generation (P) virgin females carrying a stable source of P-transposase (Δ 2-3) were mated to males carrying a P[lacZ] transposons inserted on the 2nd chromosome (A46). Non-ebony i.e. Sb e Δ 2-3/ry^{506} virgin females of the first filial generation (F1) were selected for mating to ebony males i.e. TM6/rucuca. The females in this cross contain both the P[lacZ] element and the stable source of transposase (Δ2-3) facilitating transposition. Non-ebony (e), non-Stubble (Sb) male progeny of this cross were visually screened for altered bristle phenotypes. Only non-ebony males were scored in an attempt to limit the presence of P-transposase to only one generation. In this way it was hoped to avoid isolating flies in which an insertion event had been followed by an imprecise (mutagenic) excision event.

Notes

1. If recombination were to occur in the F1 females there might be an exchange of genetic markers on the 3rd chromosome. Segregation of stubble or ebony from Δ 2-3 would allow constitutive P-transposase to persist undetected into the F2 generation. This might have been avoided by conducting the reciprocal parental cross and mating the resulting F1 +/Y; A46/+; Sb e Δ 2-3/ ry^{506} males to compound first chromosome, ebony females. In this scheme the mutagenesis occurs in males where recombinational exchange of genetic markers cannot occur. Unfortunately no compound first chromosome, ebony females were available.

2. P-element insertions at sn are recovered at high frequency whenever P-elements are mobilised. Only half of the mutagenised F2 males can be scored for a sn because of the difficulty in scoring sn bristle phenotypes in a Sb background. The loss of half the mutagenised F2 males is only acceptable because of the greatly increased occurrence of P-element insertion at sn during element mobilisation.

Genetic Markers

w = white^{1} ry = rosy^{506}

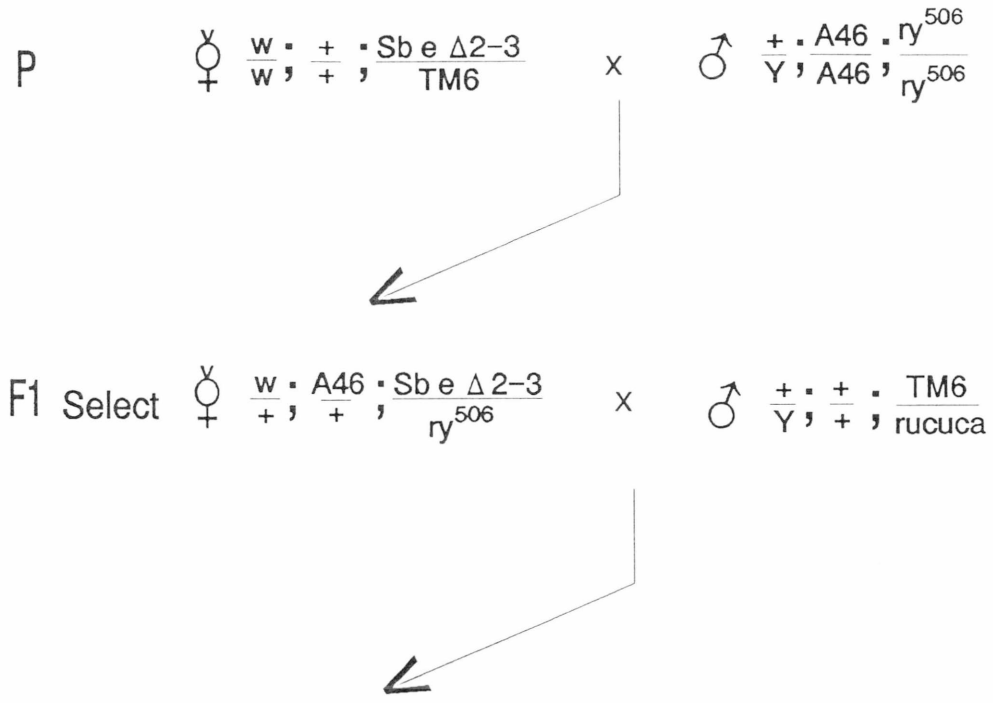
A46 = P[lacZ] insertion on the 2nd chromosome

Sb = Stubble e = ebony

TM6 = 3rd chromosome balancer; In(3R) TM6, ss^{-} bx^{34e} e

rucuca = multiply marked 3rd chromosome; ru h th st cu sr e ca

Figure 4.3.1



F2 Score $Sb^{+}e^{+}$ males for sn^{-} bristle phenotype

♂ +/Y ; A46/+ ; TM6/rucuca / ry506

♂ w/Y ; A46/+ ; TM6/rucuca / ry506

Figure 4.3.2 - Out crossing scheme for $\underline{sn}^{P[lacZ]}$ mutants

Males carrying potential sn mutations were recovered from the mutagenesis screen and subject to the outcrossing scheme shown in Figure 4.3.2. This was designed to replace all the autosomes in the mutant fly.

Genetic markers

Basc = 1^{st} chromosome balancer; In(1) sc^{S1L} sc^{8R}+S, sc^{S1} sc^{8} w^{a} B

Cy = Curly

bw^{V1} = $\underline{brown}^{V1}$

Sb = Stubble

Ubx = Ultrabithorax

sn^{-} = potential P[lacZ] allele of sn

Y = Y-chromosome

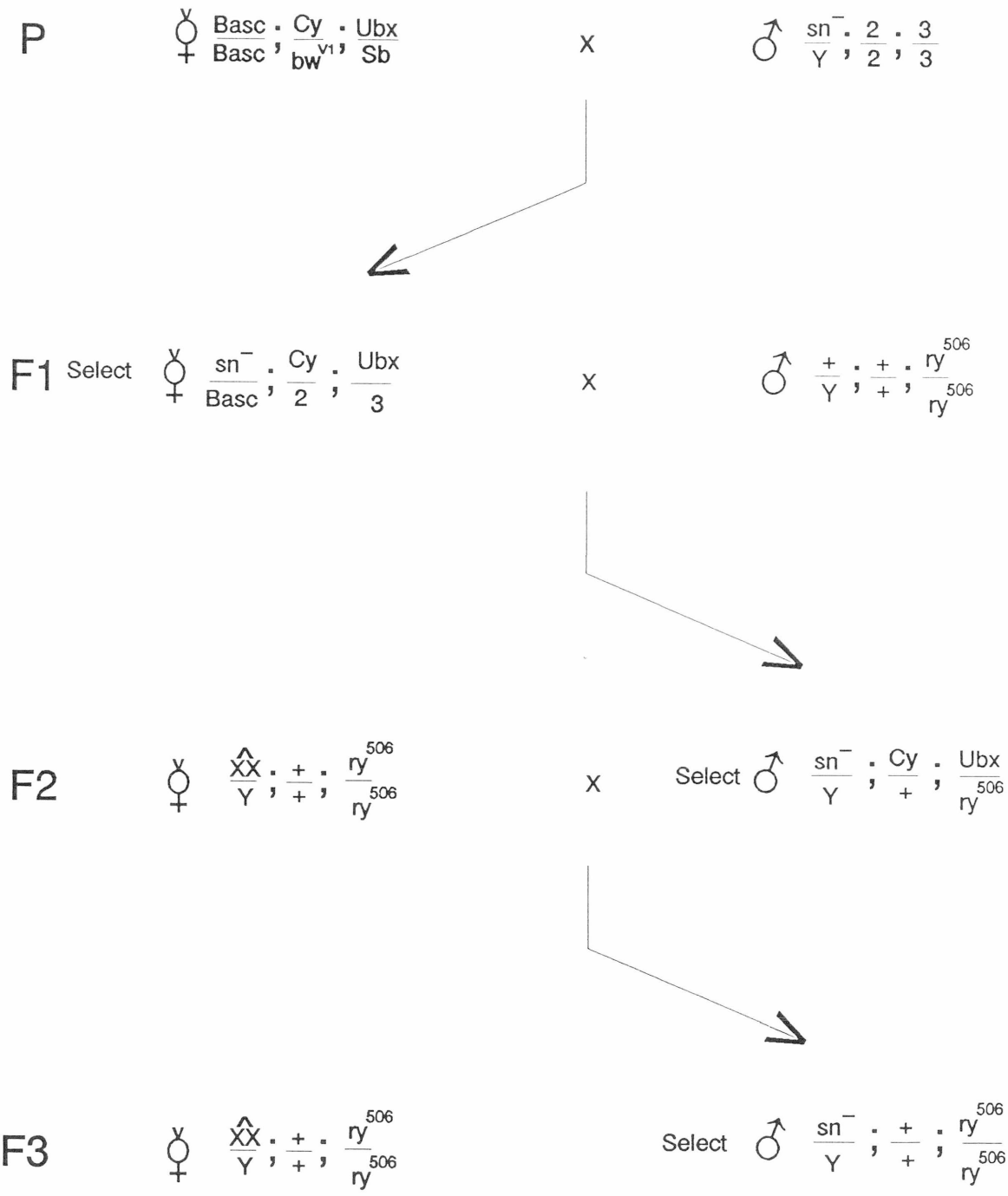
2 = any 2^{nd} chromosome present in the mutagenised stock

3 = any 3^{rd} chromosome present in the mutagenised stock

ry^{506} = $\underline{rosy}^{506}$

XX = compound 1^{st} chromosome

Figure 4.3.2


P
♀ Basc/Basc ; Cy/bw^V1 ; Ubx/Sb
x
♂ sn^-/Y ; 2/2 ; 3/3
F1 Select
♀ sn^-/Basc ; Cy/2 ; Ubx/3
x
♂ +/Y ; +/+ ; ry^506/ry^506
F2
♀ X̂X/Y ; +/+ ; ry^506/ry^506
x
Select ♂ sn^-/Y ; Cy/+ ; Ubx/ry^506
F3
♀ X̂X/Y ; +/+ ; ry^506/ry^506
Select ♂ sn^-/Y ; +/+ ; ry^506/ry^506

of development, the staining pattern of β-galactosidase from both alleles was identical. To exclude possible phenotypic effects of the $\underline{sn}^{P[lacZ]}$ insertion, staining patterns were always examined in heterozygous $\underline{sn}^{P[lacZ]}/\underline{sn}^{+}$ individuals.

4.4 Enhancer detector systems : Pupal $\underline{sn}^{P[lacZ]}$ staining

To examine sn expression during the formation of hairs and bristles, $\underline{sn}^{P[lacZ]}/\underline{sn}^{+}$ pupae were dissected out of their pupal cases and stained for β-galactosidase activity. The earliest pupal time point amenable to this type of dissection was stage P7 (Bainbridge and Bownes 1981), and staining is already present in the developing wings and lateral surfaces of the thorax (Figure 4.4.1). At a slightly later stage (P8), staining has extended across the entire upper layer of the thorax and is present throughout the legs and wings. As development continues, staining appears in the head. Thus a P10 pupa show β-galactosidase staining throughout the thorax, thorasic appendages and head. At a later stage of development P14-15 the abdomen has become stained and the imago has β-galactosidase staining across the entire surface of its body Figure 4.4.2. Crude hand dissection suggests that only epidermal layers are stained. The pattern and timing of the $\underline{sn}^{P[lacZ]}$ staining does not reflect the temporal and spatial development of hairs and bristles (Lees and Waddington, 1942). The β-galactosidase staining occurs earlier and in a different order to the formation of hairs and bristles. However the localisation $\underline{sn}^{P[lacZ]}$ staining to the epidermis is consistent with the role of sn in the formation of epidermal structures i.e. hairs and bristles.

Figure 4.4.1 - Pupal $\underline{sn}^{P[lacZ]}$ staining : early stages

Upper panel : The ventral side of $\underline{sn}^{P[lacZ]}/+$ pupae from stages P7 to P10 (Bainbridge and Bownes 1981).

Lower Panel : The dorsal side of $\underline{sn}^{P[lacZ]}/+$ pupae from stages P7 to P10 (Bainbridge and Bownes 1981).

Figure 4.4.1

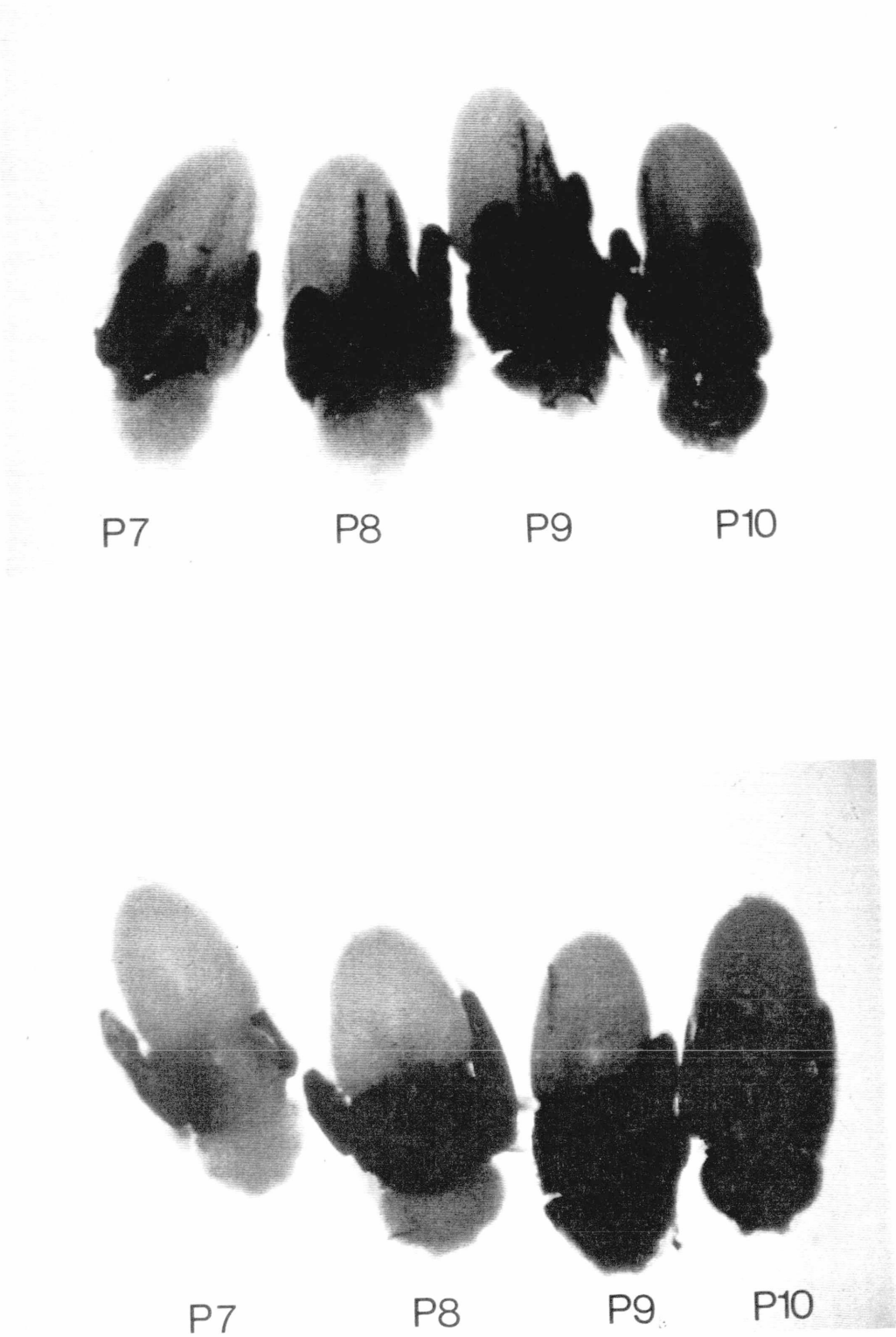
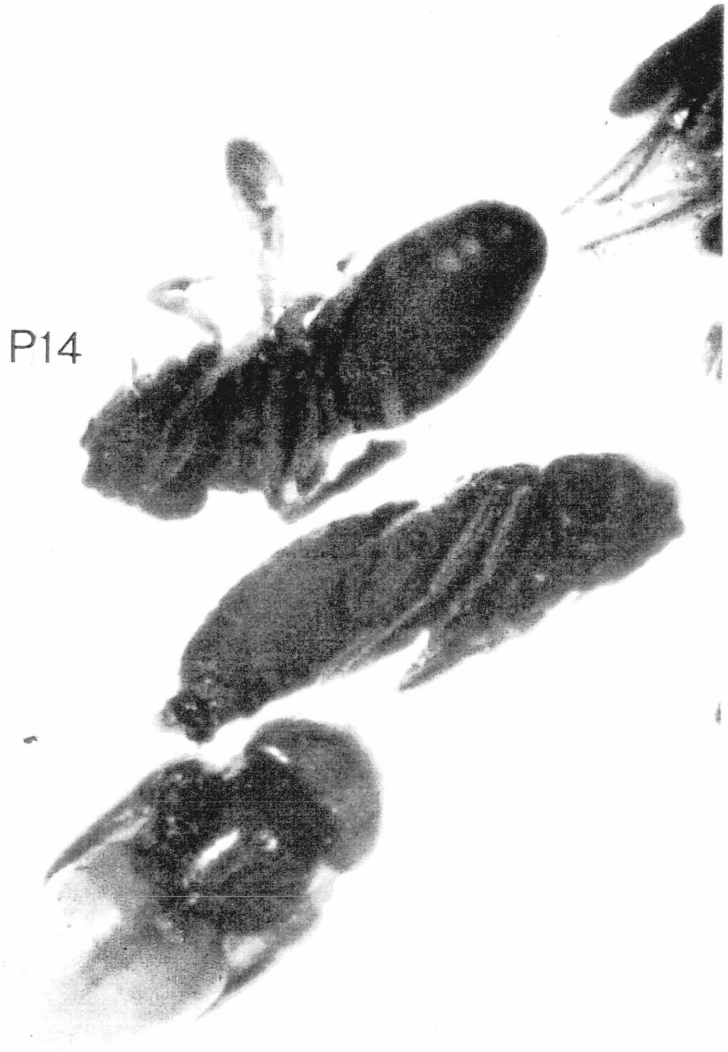


Figure 4.4.2 - Pupal $\underline{sn}^{P[lacZ]}$ staining : late stages

The two all blue pupae in the upper section of the photograph are $\underline{sn}^{P[lacZ]}/+$ stage P14 imagoes. The pupae at the base of the photograph is a dorsal view of a P10 $\underline{sn}^{P[lacZ]}$ pupae (Bainbridge and Bownes 1981).
x4

Figure 4.4.2



4.5 Enhancer detector systems : Ovarian $sn^{P[lacZ]}$ staining

The female sterility seen in some sn alleles is due to the disruption of the ovary development in homozygous sn^{-} flies. To understand the role of sn during oogenesis $sn^{P[lacZ]}/sn^{+}$ ovaries were stained for β-galactosidase activity. Ovaries were always stained as whole mounts (Figure 4.5.1) and subsequently dissected. No staining is seen in the germanium or in early egg-chambers, stages 1 to 7 (King 1970, Figure 4.5.1 and 4.5.2). The earliest staining is observed in stage 8 egg-chambers where progenitor border cells are seen to stain (Figure 4.5.2 and 4.5.3). These somatically derived cells are located at the anterior end of the egg-chamber. Six to ten of these cells pinch off the follicle cell layer surrounding the egg-chamber and start to migrate through the nurse cell chamber towards the oocyte. A few of these progenitor border cells are left behind at the anterior pole and quickly lose their β-galactosidase staining activity (Figure 4.5.2 lower panel). By stage 9 the border cells have started to migrate and the nurse cells begin to show β-galactosidase staining. All 15 nurse cells continue to express β-galactosidase throughout the rest of oogenesis. The oocyte nucleus does not stain. The intense blue staining in the border cells allows the migration of these cells to be followed from the anterior pole, through the nurse cell chamber, up to a position at the ring canals close to the oocyte nucleus (Figure 4.5.3 and 4.5.4 upper panel). In a stage 10 egg-chamber the border cells and nurse cells are clearly seen to be staining (Figure 4.5.4 upper panel). However close examination shows that in addition to the nurse cells and border cells, certain follicle cells surrounding the oocyte are also stained (Figure 4.5.4 upper panel). The staining cells are at the apical cap and the anterior corners of the oocyte. During stage 11 the nurse cells degenerate passing the majority of their contents into the oocyte. The oocyte stage 12 egg-chamber is intensely stained and the border cells are still visible in the degenerating nurse cell chamber

Figure 4.5.1 - Ovarian $\underline{sn}^{P[lacZ]}$ staining : whole mount

Whole mount $\underline{sn}^{P[lacZ]}/+$ ovary showing β-galactosidase staining. The germania are to the left and late stage egg-chambers are to the right. Note the absence of staining in early stages 1-7. Several nurse cell nuclei and some clusters of border cells are clearly visible.
x4

Figure 4.5.1

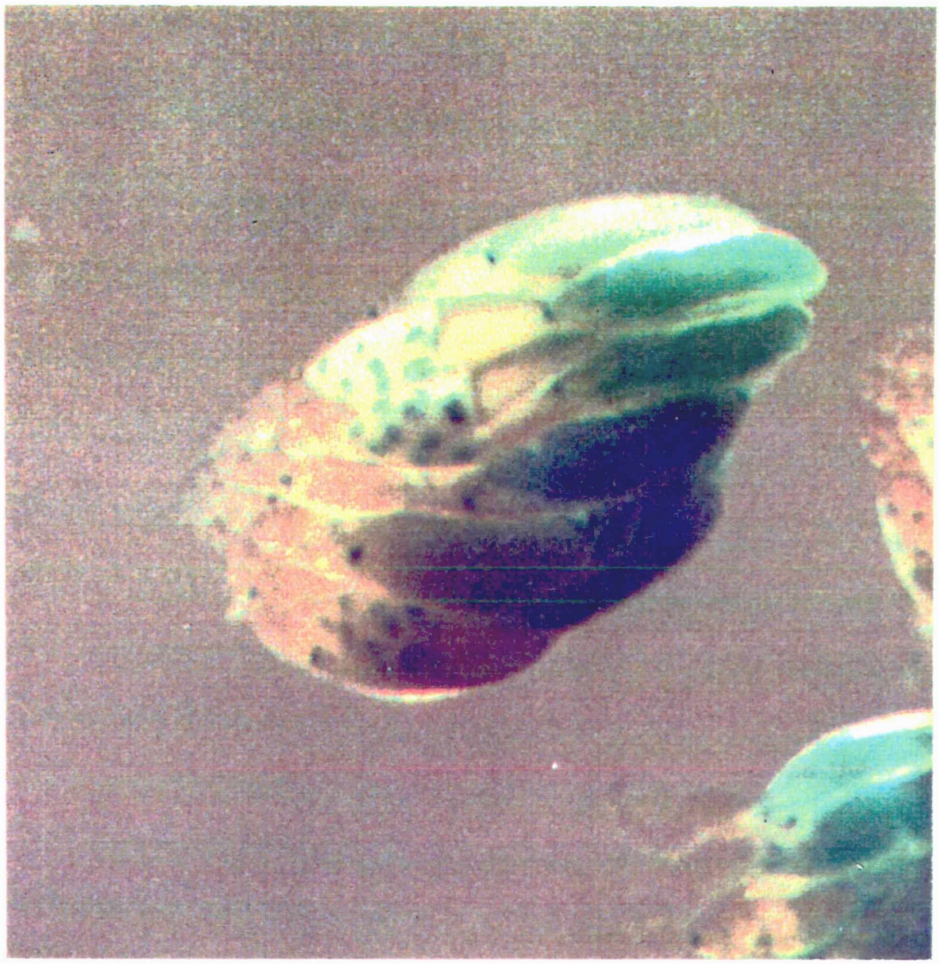
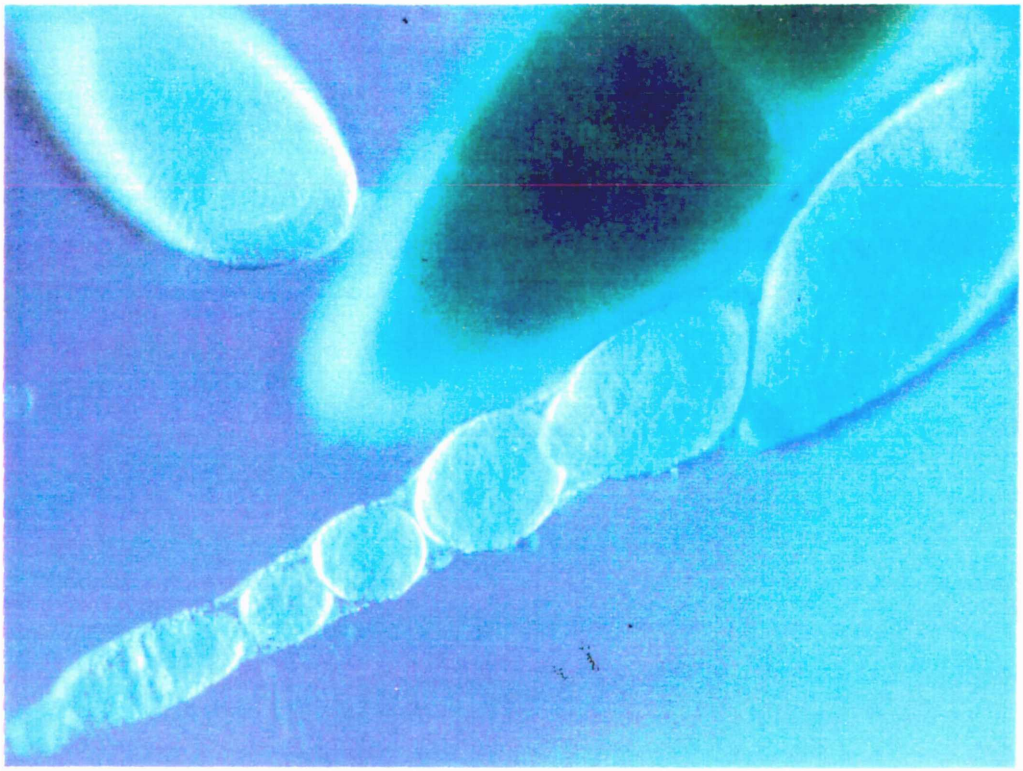


Figure 4.5.2 - Ovarian $\underline{sn}^{P[lacZ]}$ staining : early stages

Upper panel : An ovariole showing early stages of oogenesis; germanium to stage 8. No staining is seen in the germaniun or oocyst stages 1 to 7. The stage 8 egg-chamber at the right hand end of the ovariole has some progenitor border cell staining at to anterior tip of the egg-chamber (slightly below the plane of focus).
x20 Nomarski Optics

Lower panel : Stage 8 egg-chamber showing the border cells (BC) beginning to migrate through the nurse cell chamber. Staining of nurse cells and progenitor border cells left behind at the anterior pole is clearly visible.
x20 Nomarski Optics

Figure 4.5.2



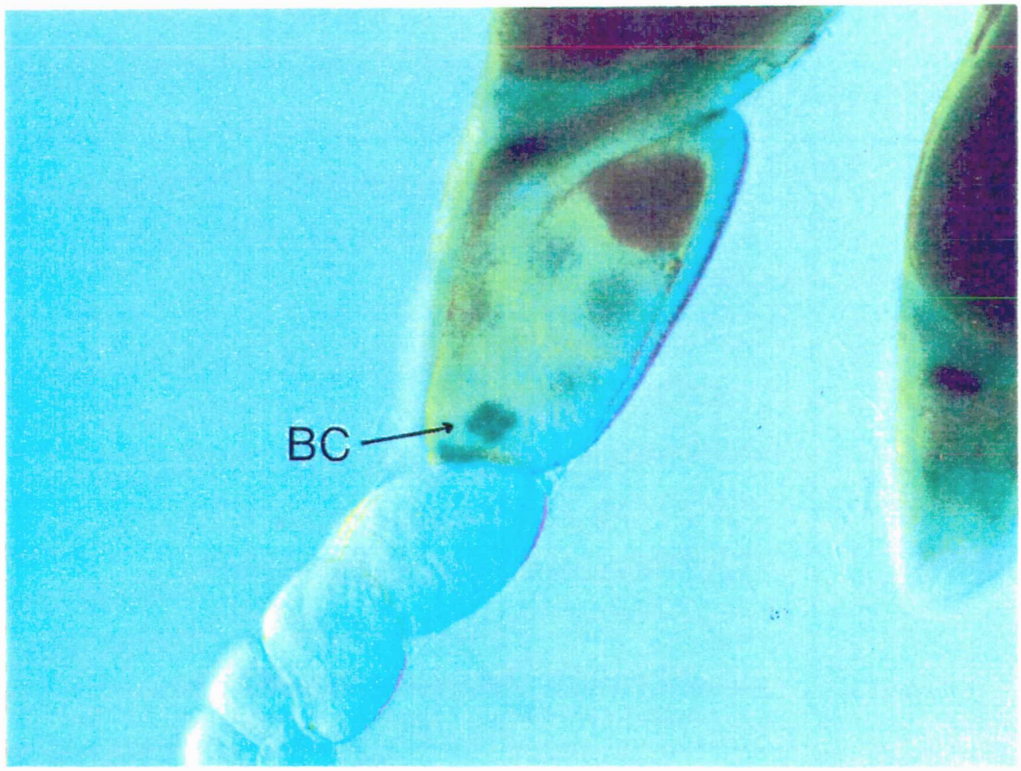


Figure 4.5.3 - Ovarian $\underline{sn}^{P[lacZ]}$ staining
: border cell migration

Upper panel : A stage 8 (centre left) and (centre right) stage 9 egg-chamber. In the stage 8 the border cells are still very close to the anterior pole and the nurse cells are only staining weakly. The border cells of the stage 9 egg-chamber are in the centre of the nurse cell chamber and nurse cell expression of β-galactosidase has increased (cf. stage 8). BC = border cells, NC = nurse cells.
x20 Nomarski Optics

Lower Panel : A stage 8 (centre left) and a stage 10a (centre right) egg-chamber. The border cells of the stage 10a egg-chamber are located against ring canals near the developing oocyte. The group of apical follicle cells (A) which express β-galactosidase activity are easily seen in this stage 10a.
x20 Nomarski Optics

Figure 4.5.3

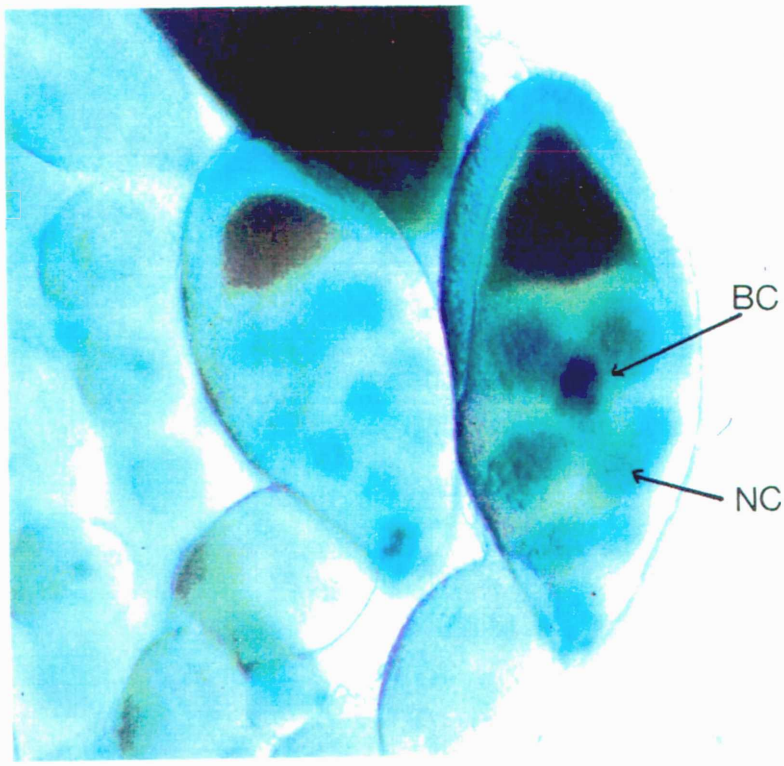


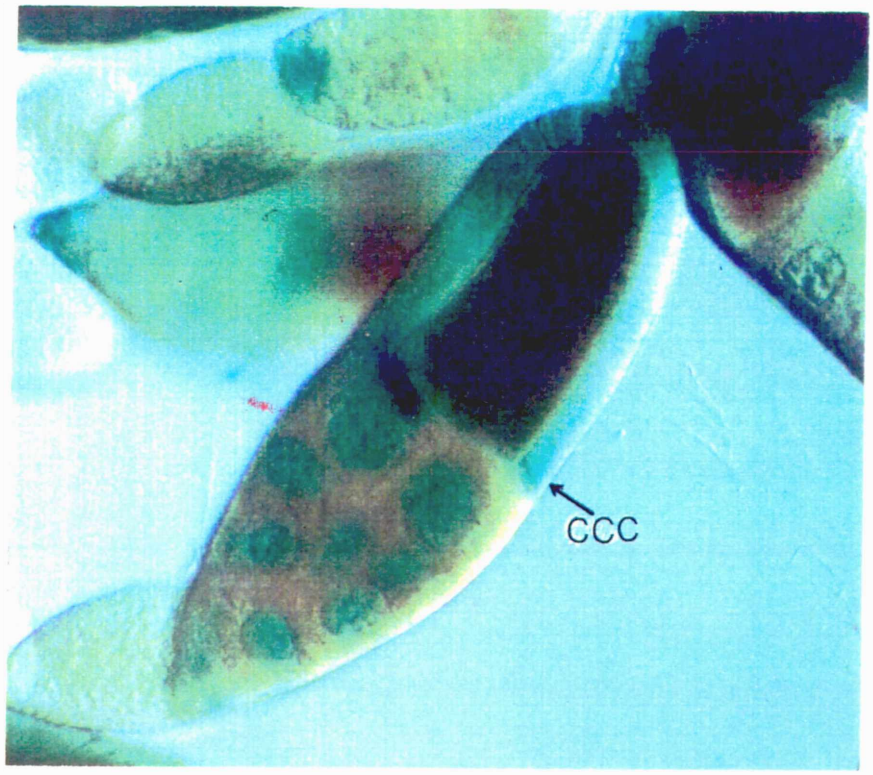


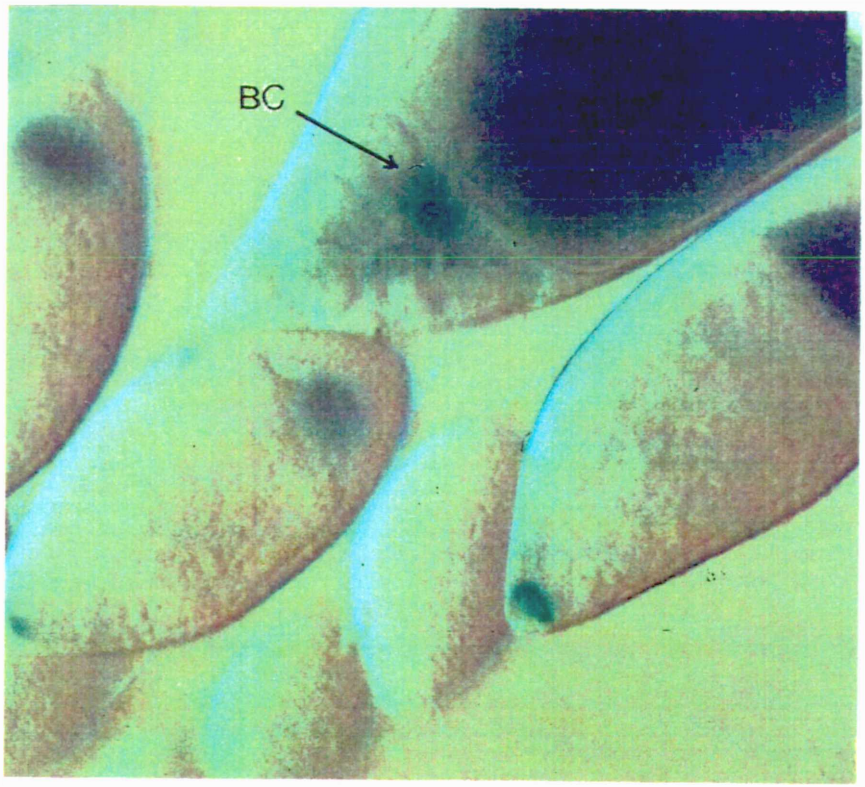
Figure 4.5.4 - Ovarian $\underline{sn}^{P[lacZ]}$ staining : late stages

Upper panel : Stage 10b egg-chamber. The border cells are now directly opposite the oocyte nucleus. Nurse cells and follicle cells at the apical cap and corner columnar cells (CCC) are all expressing β-galactosidase.
x20 Nomarski Optics

Lower panel : Bottom left and right two stage 8 egg-chambers showing progenitor border cells staining at the anterior tip of the egg-chamber. The upper central section of the photograph shows the degenerating nurse cell chamber of a stage 12 egg-chamber. Border cells are still strongly stained even though most of the nurse cell staining has been passed into the oocyte.
x20 Nomarski Optics

Figure 4.5.4





(Figure 4.5.4 lower panel). The β-galactosidase staining in the the oocyte renders the whole egg-chamber opaque preventing any study of the oocyte nucleus or continued follicle cell staining in the final stages. Mature egg-chambers (stage 14) retain this intense oocyte staining (Figure 4.5.1 right side of ovary).

4.6 Enhancer detector systems : Maternally inherited $sn^{P[lacZ]}$

Embryos derived from $sn^{P[lacZ]}/sn^{+}$ females were fertilised by wildtype males and stained for β-galactosidase activity. Note that these embryos are a mixture of $sn^{P[lac]}/sn^{+}$ and sn^{+}/sn^{+} genotypes but they all show the same pattern of staining (data not shown). Early embryos are intensely and evenly stained with β-galactosidase activity. As embryogenesis proceeds the staining is gradually lost. The uniform decay of the staining pattern eventually results in late stage embryos containing no detectable β-galactosidase activity. This staining pattern is consistent with a gradual loss of maternally inherited β-galactosidase activity, without significant zygotic expression.

4.7 Enhancer detector systems : Paternally inherited $sn^{P[lacZ]}$

RNA blotting (Section 4.2) showed that there was very little, if any, zygotic sn expression in 0-16hr embryos. Because of the high level of staining seen in embryos derived from $sn^{P[lacZ]}/sn^{+}$ mothers, it was possible that a low level of zygotic expression was being masked.

To investigate this possibility, wildtype virgin females were mated to $sn^{P[lacZ]}$ males. The resulting embryos were collected and stained for β-galactosidase activity. Early embryos show no staining (Figure 4.7.1 upper panel) because the zygotic genome is transcriptionally inactive. As embryogenesis continues and gastrulation occurs a horse shoe of punctate staining develops around the surface of the embryo (Figure 4.7.1). This staining is transient and does not appear to

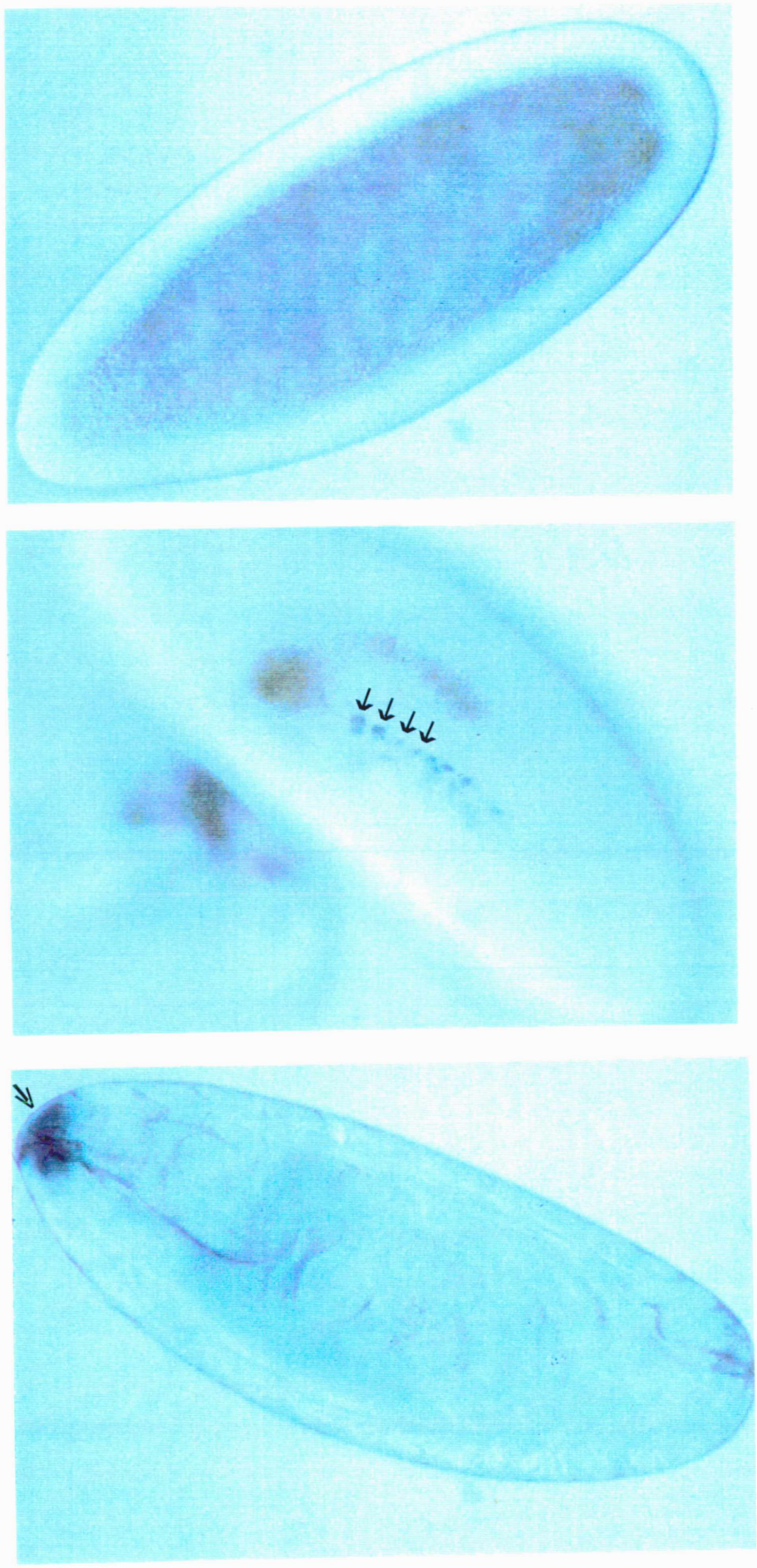
Figure 4.7.1 - Embryo $\underline{sn}^{P[lacZ]}$ staining : paternal

Upper panel : A 3hr embryo having just completed cellularisation. No β-galactosidase staining is detected. Compare with Figure 4.6.1 upper and middle panels.
x20 Nomarski Optics

Middle panel : A segmenting embryo (12hr). The punctate β-galactosidase staining is present in a horse shoe starting on the surface of on flank and traversing across the ventral surface to finish on the opposite flank. The staining is transient and the structure has yet to be identified.
x20 Nomarski Optics

Lower panel : Lateral view of an embryo just prior to hatching (24hr). β-galactosidase staining is concentrated into two cell clusters at the anterior tip of the embryo. Compare with lower panel of Figure 4.6.1.
x20 Nomarski Optics

Figure 4.7.1



correspond to any known structure. After the horse shoe of staining has disappeared β-galactosidase staining activity is seen to accumulate in two clusters of cells at the anterior tip of the embryo (Figure 4.7.1). This pattern of expression is hidden in embryos derived from $\underline{sn}^{P[lacZ]}/\underline{sn}^{+}$ mothers. At present the structures seen to stain in these embryos are not identified.

To explain these results the indirect nature of the method of detecting gene expression in P[lacZ] experiments must be considered. The only transcriptional control sequences on the P[lacZ] transposon are weak P-element promoter signals. Alone these sequences do not provide enough P[lacZ] RNA for β-galactosidase activity to be detected. Expression of β-galactosidase only occurs when the element is inserted next to an active enhancer sequence. The P[lacZ] elements being studied here are inserted into the sn gene, so that sn enhancer activity will result in β-galactosidase staining. Adjacent to sn gene is the olfE gene. This transcription unit is possibly close enough to the $\underline{sn}^{P[lacZ]}$ insertion to allow olfE enhancer activity to influence transcription of the P[lacZ] element. The RNA blotting experiment shown in Figure 4.2.1 shows that zygotic expression of sn during embryogenesis is barely detectable. Thus the most likely explanation for the specific β-galactosidase staining seen in embryos is that the olfE enhancer is directing expression of the P[lacZ] elements in sn. The staining of cells close to what is to become the mouth of the first instar larva is consistent with these cells having a role in olfaction. The barely detectable 3.6kb RNA seen in Figure 4.2.1 may be a result of sn expression due to the activation of the olfE gene.

4.8 Concluding remarks

The most striking sn phenotype is the alteration of bristle morphology in mutant alleles. This ranges from a slight bend or wave in the macrochaetes of weak mutants to all bristles being gnarled and twisted in extreme mutants.

In some extreme mutants hairs are wavy and kinked. Adult hairs and bristles are formed during metamorphosis at the pupal stage of the lifecycle. The presence of a 3.6kb sn transcript at this point in development is consistent with a requirement for sn during bristle and hair production. The staining pattern of β-galactosidase seen in the P[lacZ] insertion mutants of sn suggests that the gene is transcribed in a specific spacial and temporal pattern in the epidermis during pupation. β-galactosidase staining occurs before hair and bristle formation and does not follow the pattern of hair and bristle formation. This suggests that the sn product is not simply a component of bristles and hairs. In fact, the extent of staining suggests that sn may be expressed in more epidermal cells than just those making bristles and hairs.

The morphology of cuticular bristles and hairs is affected in sn mutants, but the ommatidial associated hairs in the compound eye are not affected. This observation is again consistent with the sn product having no direct role in the structure or formation of hairs in general. If sn has a more general role in epidermal development then it may be easier to correlate apparently unrelated bristle and oogenic phenotypes.

Transcription of sn in the female germline results in 3 RNA products, 3.0, 3.3, and 3.6kb in length. The transcripts differ only in their length of 3' untranslated sequence and thus all encode the same protein. Two transcripts, the 3.0 and 3.3kb RNAs, are ovary specific (Figure 4.1.1), but all 3 RNAs are maternally inherited showing that they are produced in the germline (Figure 4.2.1).

β-galactosidase staining from P[lacZ] insertions at sn identifies particular ovarian cells as actively expressing sn (Figures 4.5.1-4). The most obvious feature of this pattern is the restriction of β-galactosidase staining to mid and late stage egg-chambers (Figure 4.5.1). This suggests that the sn gene is not transcribed during early stages of oogenesis (stages 1-7). Consistent with the expression of sn in mid and late stage egg-

chambers is the disruption of these stages in ovaries of sn mutants (Bender, 1960).

The appearance of sn expression in middle stage egg-chambers and persistence in later stages could actually have been predicted from the analysis of sn transcripts across development (Figure 3.2.1). RNA blotting analysis shows that late pupae contain only the 3.6kb sn transcript (Figure 3.2.1). But the tissue sample from which this RNA was derived would certainly have included females whose ovaries contain early stage egg-chambers. The absence of the 3.0 and 3.3kb RNAs in late pupae shows that early egg- chambers do not produce the ovary dependent 3.3 and 3.0kb RNAs. These observations were expanded by analysing sn RNAs present in 0-4hr, 1 day, 2 day, and 3 day old adult females (O'Hare unpublished observation). The 3.0 and 3.3kb RNAs are first detected in 2 day old females. Newly eclosed and 1 day old females have ovaries with only early (stage 1-6) egg-chambers. This suggests that sn transcription is only activated in the germline at about stage 8, and is entirely consistent with the appearance of β-galactosidase staining seen in $sn^{P[lacZ]}$ mutants.

A striking correlation between β-galactosidase staining and the oogenic phenotype of sn is the localisation of expression to border cells (Figure 4.5.3). The function of the 6-10 border cells is to secrete vitelline membrane (King, 1970). Consistent with the $sn^{P[lacZ]}$ staining of border cells is the abnormality of the micropyle and vitelline membrane in eggs derived from mutant sn mothers (Bender, 1960). Note that staining in border cell occurs long before these cells are involved in micropylar formation or vitelline membrane secretion. This suggests that sn expression is require for other processes within the border cells.

In addition to the border cells, two other groups of follicle cells show β-galactosidase staining in $sn^{P[lacZ]}/sn^{+}$ ovarian tissue. These are a group of corner columnar cells at the anterior end of the oocyte and a few cells at the posterior apical cap of the oocyte. From the

oogenic phenotype of sn mutants the expression in the corner columnar cells may be related to the malformation of the dorsal filaments seen in sn eggs (Bender, 1960). No obvious phenotype can be associated with $sn^{P[lacZ]}$ expression at the apical cap.

In mutant sn ovaries, vitellogenesis is severely retarded. Nurse cells which normally produce and deposit yolk proteins show a reduced level of polyploidy and the whole process of vitellogenesis is disrupted. The timing of β-galactosidase staining seen in nurse cells of $sn^{P[lacz]}$ ovaries is consistent with a requirement for sn product during vitellogenesis.

Genetically it has been shown that sn product is required in the female germline (Perrimon and Gans, 1983). However, germ cell expression of sn does not exclude expression of sn in somatic ovary cells. In fact $sn^{P[lacZ]}$ staining suggests that sn expression first occurs in somatic tissue (border cells, stage 8) and subsequently germ cells express sn (nurse cells, stage 9). The expression of sn in somatic cells might have been anticipated from the disruption of the micropyle in sn mutants. This phenotype was originally interpreted as a secondary effect due to a lack of sn expression in germline nurse cells. Future analysis of sn must now consider a more direct role for sn in border cells, and also perhaps in other follicle cells.

CHAPTER 5

Molecular analysis of mutant alleles of *singed*.

5.0 Introduction

The results described in chapters 3 and 4 report the structure and expression of the wildtype sn gene. The experiments described in this chapter examine the structure and transcription of 5 sn mutants. Study of transcription in these mutants has led to a greater understanding of the molecular events responsible for each phenotype.

Two criteria were used to select alleles suitable for this analysis. Firstly, each allele must contain a lesion detectable by DNA blotting. This facilitated the identification and characterisation of mutagenic alterations. Secondly, the alleles chosen for this analysis must be representative of the entire range of sn phenotypes.

The alleles selected for study were sn^{X2}, sn^{3}, sn^{W}, sn^{e} and $sn^{(+)}$. DNA blotting experiments (Roiha et al.,1988) have characterised each allele as containing a DNA alteration within the interval -2.0 to 0.0 (Figure 5.0.1).

The sn^{3} mutation had been described as a small deletion (~300bp) within the interval -0.9 to 0.0. This results in an extreme bristle phenotype but the mutation has no effect on female fertility. To facilitate full interpretation of RNA blotting data the interval disrupted sn^{3} was cloned and sequenced (see below).

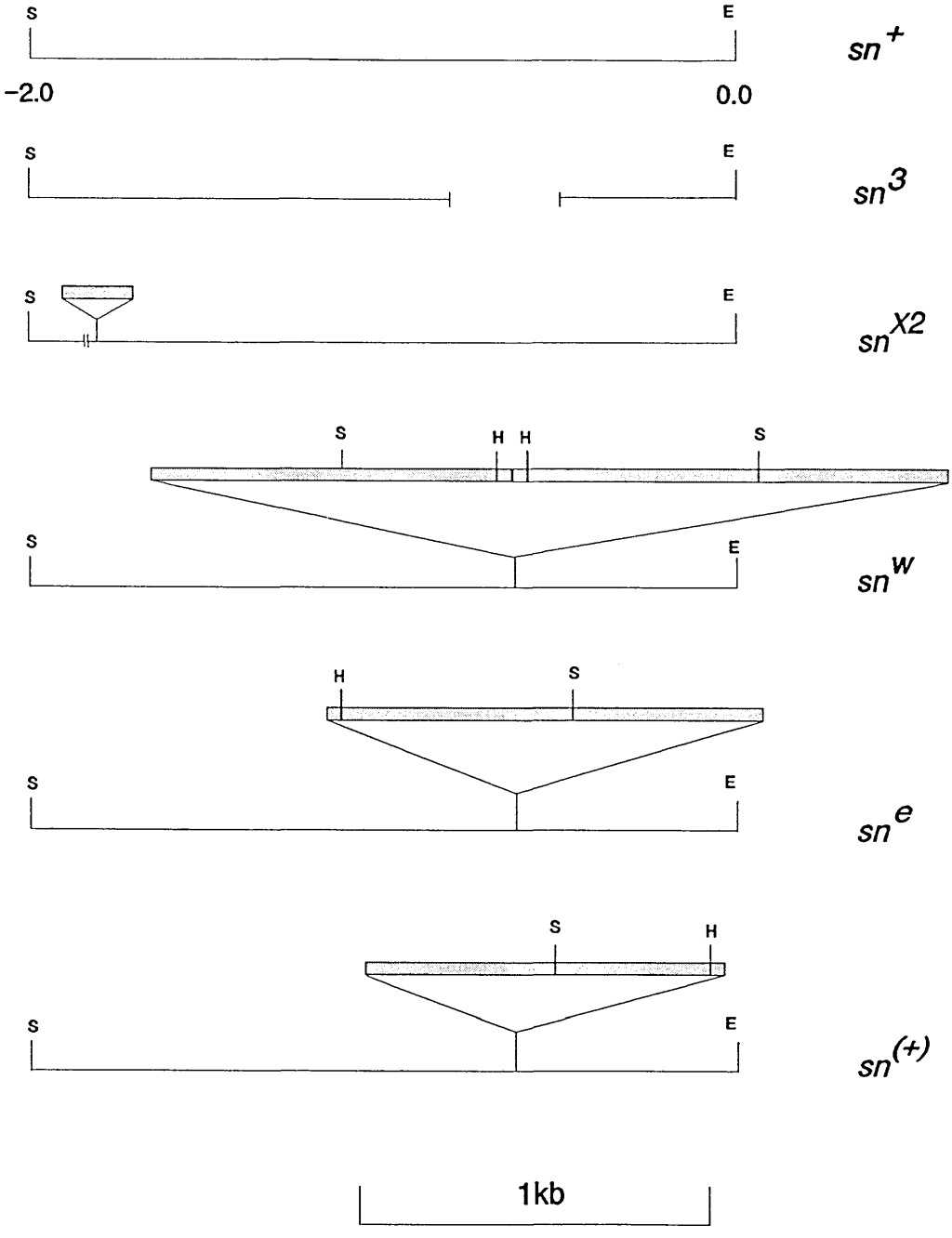
In contrast to sn^{3}, the extreme bristle phenotype of sn^{X2} is associated with female sterility. The DNA disruption in sn^{X2} is a ~100bp insertion within the interval -1.9 to -1.2. The first step in characterising this mutation was to clone and sequence the interval containing the insertion.

The remaining 3 alleles are P-M dysgenesis induced mutations. Each allele has P-element sequences inserted at the same site within the first exon of sn (Roiha et al., 1988; Chapter 1). This family of related insertions produce a range of bristle phenotypes. The P-element insertion in sn^{W} results in bent or kinked macrochaetes.

Figure 5.0.1 - DNA maps of mutant alleles of sn chosen for transcriptional analysis.

The DNA maps from -2.0 to 0.0 of sn^3, sn^{X2}, sn^W, sn^e and $sn^{(+)}$. The maps are derived from DNA blotting data (Roiha et al., 1988), and cloning and sequencing studies (K.O'Hare pers.comm. and this study). Insertions are denoted by shaded boxes above the wildtype maps, the boundaries of deletions are marked by vertical bars within DNA maps. Restriction enzyme sites are abbreviated as follows E = EcoR1, S = Sal1 H = HindIII.

Figure 5.0.1



The bristles of a $\underline{sn}^{e}$ fly are gnarled and twisted while the bristle phenotype of $\underline{sn}^{(+)}$ is indistinguishable from wildtype. The effect each of these alleles has on female sterility is dependent upon P-M cytotype (see Section 1.6). In the M cytotype all three alleles are female fertile but in the P cytotype each mutation becomes female sterile. The bristle phenotypes of these alleles are not affected by cytotype. The cytotype dependence of $\underline{sn}^{W}$, $\underline{sn}^{e}$, and $\underline{sn}^{(+)}$ female sterility is examined in Section 5.7.

The molecular structure of $\underline{sn}^{W}$ and its derivatives $\underline{sn}^{e}$ and $\underline{sn}^{(+)}$ have been characterised (Roiha et al., 1988; Nitasaka and Yamazaki 1988). The $\underline{sn}^{W}$ mutation contains two P-elements inserted adjacent to each other in a head to head orientation (Figure 5.0.1). The 0.95kb and 1.3kb P-elements inserted in $\underline{sn}^{W}$ are believed to be separated by only 8bp (Roiha et al.,1988). The 8bp gap is thought to represent the duplication of target sequences found bounding all P-element insertions (O'Hare and Rubin 1983).

The $\underline{sn}^{e}$ mutation is derived from $\underline{sn}^{W}$ by the precise excision of the 0.95kb element during P-element mobilisation (Figure 5.0.1). This leaves the 1.3kb P-element in $\underline{sn}^{e}$ inserted at the same nucleotide as the P-element sequences in $\underline{sn}^{W}$.

If the 1.3kb element is excised from $\underline{sn}^{W}$ then the remaining 0.95kb P-element produces no bristle phenotype. This allele is called $\underline{sn}^{(+)}$ (Figure 5.0.1). Because $\underline{sn}^{(+)}$, like $\underline{sn}^{e}$, is related to $\underline{sn}^{W}$ by an excision event the 0.95kb element in $\underline{sn}^{(+)}$ is inserted at the same position as the $\underline{sn}^{e}$ and $\underline{sn}^{W}$ elements (Roiha et al., 1988).

Further cloning and sequencing studies have been carried out on the $\underline{sn}^{e}$ and $\underline{sn}^{(+)}$ mutations. The 0.95kb P-element in $\underline{sn}^{(+)}$ and part of the 1.3kb element in $\underline{sn}^{e}$ have been cloned and sequenced (Roiha et al.,1988 and K.O'Hare unpublished observations). This shows both elements to duplicate the same 8bp $\underline{sn}$ target sequence (nucleotides 1352 - 1359 Figure 3.5.2A; O'Hare and Rubin 1983). The 0.95kb defective element is related to the 2.9kb autonomous P-element by a deletion of sequences from position 162 to 2113 within the full length P-element

(K.O'Hare unpublished and O'Hare and Rubin 1983). The 1.3kb $\underline{sn}^{e}$ element has been produced by a deletion of sequences from 697 to 2300 within the 2.9kb autonomous element.

The characterisation of $\underline{sn}^{e}$ and $\underline{sn}^{(+)}$ P-elements permitted the assembly of sequence files representing the structure of each allele within the interval -2.0 to 0.0 (Appendix Figures A2 and A3). Copies of the 2.9kb autonomous P-element sequence were modified to represent the structure of the $\underline{sn}^{e}$ and $\underline{sn}^{(+)}$ defective P-elements. These sequences were then inserted into copies of the genomic sn sequence from -2.0 to 0.0. Care was taken to ensure the P-element sequences were inserted in the correct orientation and that target site duplications were introduced into the final sequence files. These deduced sequences are consistent with all DNA blotting and sequence information available. A $\underline{sn}^{w}$ sequence file was assembled using both $\underline{sn}^{e}$ and $\underline{sn}^{(+)}$ P-element sequences inserted into the -2.0 to 0.0 genomic sn sequence. The assembled sequence data files are shown for $\underline{sn}^{w}$ in Appendix Figure A1, $\underline{sn}^{e}$ in Appendix Figure A2 and $\underline{sn}^{(+)}$ in Appendix Figure A3. This information will be used in the analysis of $\underline{sn}^{w}$, $\underline{sn}^{e}$, and $\underline{sn}^{(+)}$ transcription.

5.1 Molecular characterisation of $\underline{sn}^{X2}$ and $\underline{sn}^{3}$

The site and structure of P-element insertions in $\underline{sn}^{w}$, $\underline{sn}^{e}$ and $\underline{sn}^{(+)}$ have been defined at the nucleotide level. However the 2 other alleles chosen for study, $\underline{sn}^{3}$ and $\underline{sn}^{X2}$ were less well characterised. The detailed study of transcription planned for these alleles requires that both mutations should be cloned and sequenced.

In order to clone both $\underline{sn}^{X2}$ and $\underline{sn}^{3}$ sequences from a single set of experiments λ libraries were constructed with insert DNA isolated from heterozygous $\underline{sn}^{X2}/\underline{sn}^{3}$ individuals.

To isolate λ recombinants containing the region shown by DNA blotting to be disrupted in $\underline{sn}^{X2}$ and $\underline{sn}^{3}$, two libraries were made (see Chapter 2). A total of 30,000

Sau3AI recombinant phage and 120,000 BamH1 recombinant phage were plated out and screened with the genomic interval -5.0 to 0.0 (Figure 3.1.1). After three rounds of plaque purification 16 independent clones were identified as carrying sequences homologous to the interval -5.0 to 0.0 (Figure 3.1.1).

To identify clones carrying either the sn^{X2} insertion or the sn^{3} deletion, DNA was prepared from large scale bacteriophage lysates of each clone. The sn^{3} deletion had been shown to lack a Bgl1 site (1199 to 1210 Figure 3.5.2A) within the interval -0.9 to 0.0 (K.O'Hare unpublished observation). Thus DNA from each λ clone was digested with BamH1 and Bgl1 and the fragments separated by agarose gel electrophoresis. After DNA blotting the digests were probed with the interval -5.0 to 0.0 (data not shown). The absence of a Bgl1 site in sn^{3} DNA means that only one fragment in the BamH1 Bgl1 digest should hybridise to the probe. Several apparently identical phage were identified as carrying sn^{3} DNA, clone λB1 was chosen for further analysis. Phage which carry the sn^{X2} mutation have two fragments in a BamH1 Bgl1 digest which hybridise to the interval -5.0 to 0.0. One clone with this pattern of hybridisation (λS15) was selected for further study.

5.1.1 Analysis of the sn^{X2} mutation

Clone λS15 was isolated from the Sau3AI library and the restriction map of the 11.5kb insert is identical to the physical map of sn^{X2} (Roiha et al., 1988). The sn^{X2} lesion lies within the region -1.9 to -1.2. The simplest way to subclone this region from λS15 was to isolate the 2.1kb EcoR1-Sal1 interval equivalent to the wildtype -2.0 to 0.0 region. The 2.1kb EcoR1-Sal1 fragment carrying the sn^{X2} DNA alteration was subcloned into the pBluescribe-M13(-) and the new plasmid named psn9RS2-X2. Small restriction fragments from this insert were cloned into M13 vectors and then sequenced. The entire 2.1kb insert of psn9RS2-X2 was sequenced and compared to the wildtype

interval from -2.0 to 0.0 (Figure 5.1.1). The main difference between $\underline{sn}^{X2}$ and $\underline{sn}^{+}$ is the presence of an 80bp A/T rich insertion at -1.80 (Figure 5.1.1). Three smaller regions of mismatch between the $\underline{sn}^{X2}$ and $\underline{sn}^{+}$ sequences are located at positions 71-87, nucleotide 147 and nucleotide 157 (Figure 3.5.2A and Figure 5.1.1). The sequence analysis of psn9RS2-X2 is consistent with the structure of $\underline{sn}^{X2}$ derived from genomic DNA blotting experiments.

The differences between $\underline{sn}^{X2}$ and $\underline{sn}^{+}$ lie upstream of the transcription start site. This region typically contains sequence motifs important for the control of transcription. If the $\underline{sn}^{X2}$ lesions disrupt promoter or enhancer elements then transcription may be affected in this mutant.

To investigate the transcription of $\underline{sn}^{X2}$, RNA was prepared from $\underline{sn}^{X2}$ adult males and females. An RNA blot carrying these samples together with RNA derived from $\underline{sn}^{+}$ and $\underline{sn}^{3}$ males and females hybridised to the interval -0.5 to -0.35 is shown in Figure 5.1.2.

Transcription of $\underline{sn}^{X2}$ in adult males and females results in two high molecular weight RNAs (Figure 5.1.2 transcripts A and B). The abundance of these RNAs is extremely low and there appears to be no female specific transcripts. The low abundance, abnormal size and absence of female specific transcripts is consistent with the extreme bristle morphology and female sterile phenotype of this allele.

The presence of high molecular weight transcripts in $\underline{sn}^{X2}$ RNA (Figure 5.1.2) shows that some sn transcription is still occurring in these mutant flies. The $\underline{sn}^{X2}$ lesion does not abolish transcription but severely reduces the abundance of transcripts from the sn locus. This is consistent with a disruption of sn gene regulation signals by the $\underline{sn}^{X2}$ insertion as proposed above. However it is difficult to explain the increased size of $\underline{sn}^{X2}$ RNAs in terms of altered transcriptional control signals.

To ensure that the $\underline{sn}^{X2}$ lesion, detected by DNA blotting, was responsible for the phenotype associated

Figure 5.1.1 - Sequence alignment between $\underline{sn}^{X2}$ and $\underline{sn}^{+}$

The DNA sequence of $\underline{sn}^{+}$ (lower line) from the Sal1 site at 2.0 to the Sph1 site at -1.6, is compared to the corresponding region of $\underline{sn}^{X2}$ sequence (upper line). A perfect match between sequences is denoted by a star, gaps are introduced to maximise the number of matched nucleotides. The DNA maps below the sequence data show the DNA structure of $\underline{sn}^{X2}$ and $\underline{sn}^{+}$ within the region -2.0 to 0.0. The dark box below the $\underline{sn}^{X2}$ DNA map denotes the region of sequence compared to $\underline{sn}^{+}$ above. The lightly shaded box denotes the interval on $\underline{sn}^{X2}$ sequenced but not shown in this figure. (No differences between $\underline{sn}^{X2}$ and $\underline{sn}^{+}$ were found in this region). The $\underline{sn}^{+}$ sequence is numbered as in Figure 3.5.2A.

Figure 5.1.1

```
GTCGACTTTAGAACGGATATTTCAAAATAATGTTAACCTTTCGAAAAATTTAGGTGTATGCAAGCCGAATTATTTGATTTGGAAAATAATACTTAGGGCG
**********************************************************************************************************
GTCGACTTTAGAACGGATATTTCAAAATAATGTTAACCTTTCGAAAAATTTAGGTGTATGCAAGCCGAATTATTTGATTTGGAAAATAATACTTAGGGCG
 Sal 1
1                                                                                                    100

TCCATATCAGAAAAAACGTTT    GACCA      ATTTTAAAATGTGTGTTAGG CACGTTATGTAGTCAAATATATATAAATAAATTATTCTTATTA
**********************   ********** ********* ************************************
TCCATATCAGAAAAAACGTTTAGCGTAACATAACGTATTTTAAAATTTGTATTAGGCCACGTTATGTAGTCAAATATATATAAATAAATTATTC
101                                                                                          193

TTTATAAATATTATAATAATATTTAATAATAATAATATAATATATTCTTATTCTTATAAATTATTATTATTTTAATTTTTACAAAAGTGATATGCTGATA
                                                                  **********************************
                                                                  ATTTTAATTTTTACAAAAGTGATATGCTGATA
                                                                  194                          226

TGCTTATAAATGTAAGATTCTTGTTTATTTGACTTTTTTTTTAAATTTCATTATCATTATTTCTTGCTTGACAACCCTATGTTAAGTGCGACACTGCGTT
**********************************************************************************************************
TGCTTATAAATGTAAGATTCTTGTTTATTTGACTTTTTTTTTAAATTTCATTATCATTATTTCTTGCTTGACAACCCTATGTTAAGTGCGACACTGCGTT
227                                                                                                  327

GACAACAGCTAAGAACATGGCTGTGGGGCGACAAACTTAGCAGAGCGCGCGCTGCTAGAGCTTAAATTGTATTTGAAAACAGTGCAGGTGAATCGAAAAT
**********************************************************************************************************
GACAACAGCTAAGAACATGGCTGTGGGGCGACAAACTTAGCAGAGCGCGCGCTGCTAGAGCTTAAATTGTATTTGAAAACAGTGCAGGTGAATCGAAAAT
328                                                                                                  428

GTCTTGGCATGC
************
GTCTTGGCATGC
     Sph 1
429       438
```

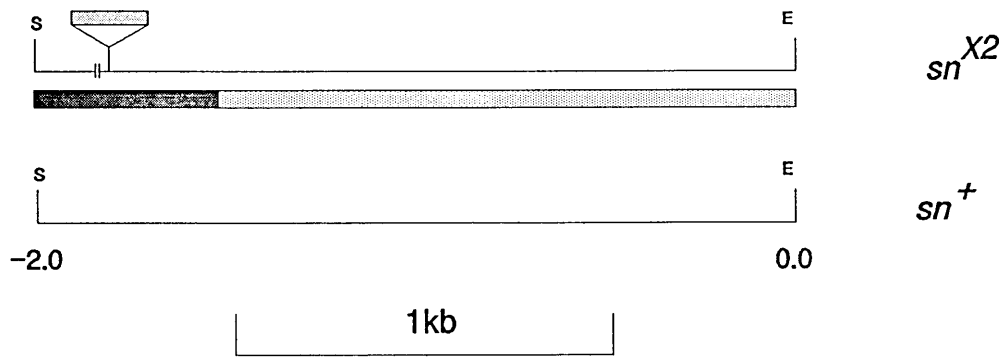


Figure 5.1.2 - RNA blotting analysis of $\underline{sn}^{X2}$

Poly-A^+ RNA was prepared from wildtype, $\underline{sn}^3$ and $\underline{sn}^{X2}$ adult males and adult females. Samples were then RNA blotted and hybridised to a sn probe (-0.5 to -0.35). The transcripts detected in $\underline{sn}^3$ RNA appear to be slightly smaller than their $\underline{sn}^+$ equivalents. Two high molecular weight transcripts are detected in $\underline{sn}^{X2}$ RNA: marked A and B.

Figure 5.1.2

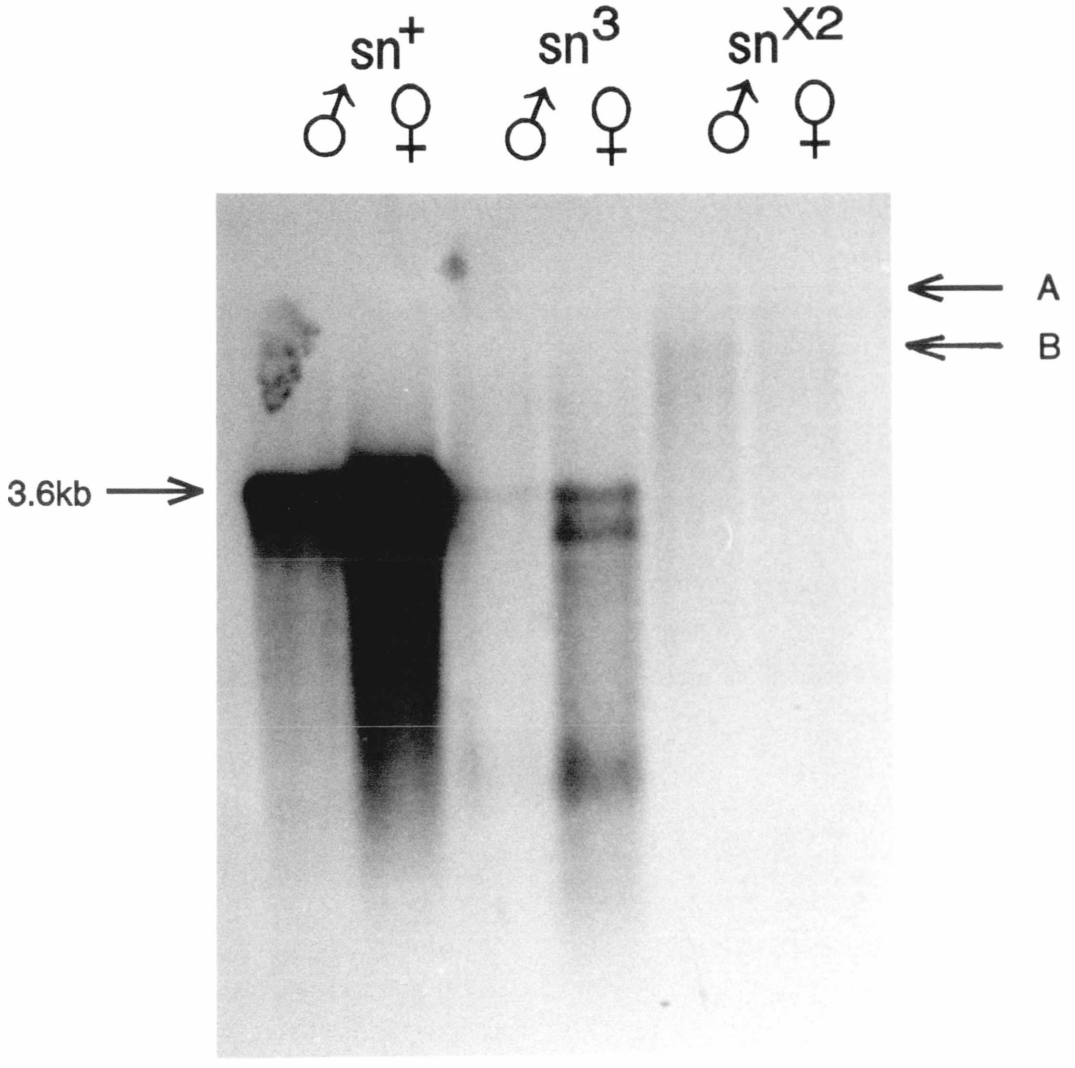
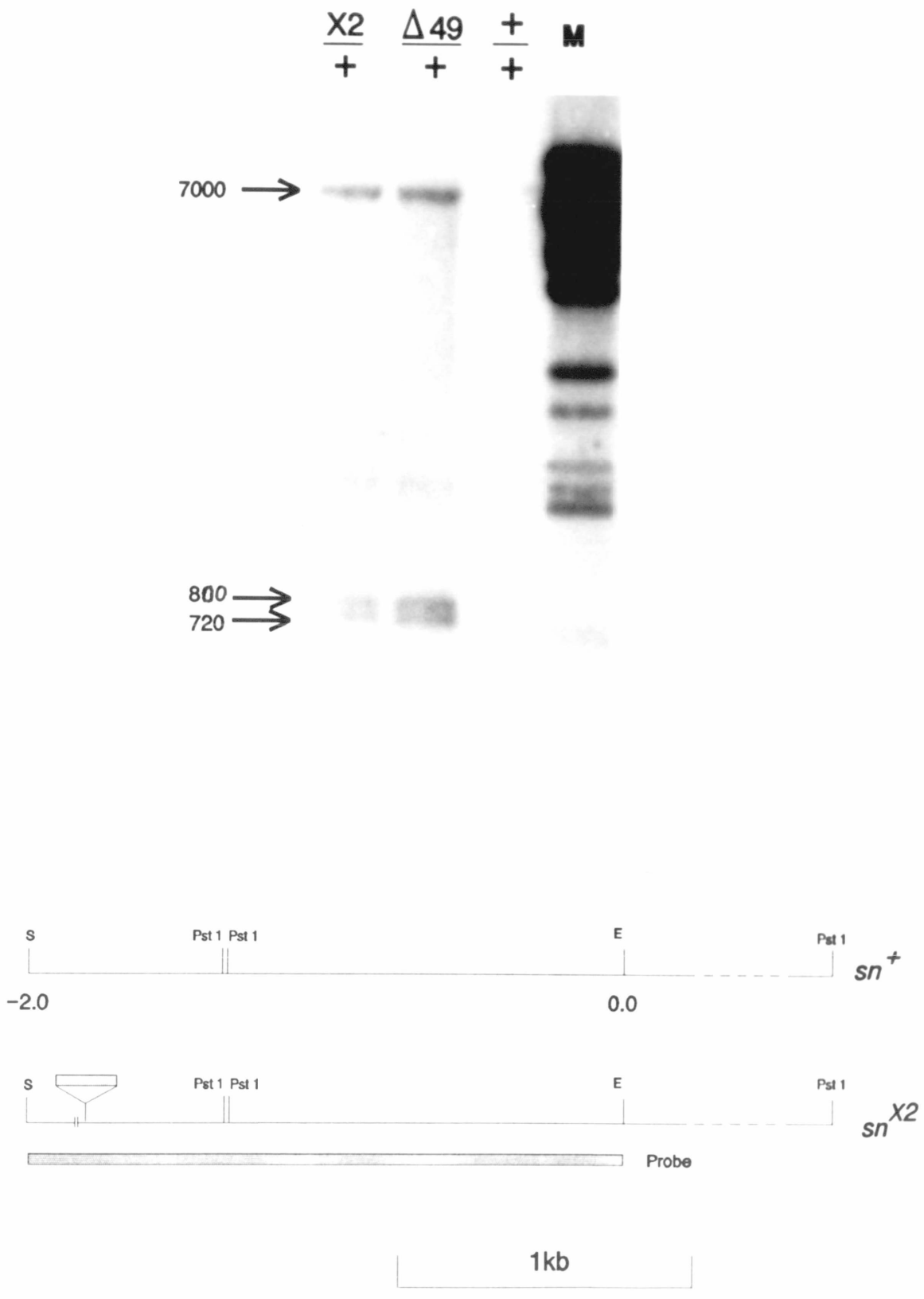


Figure 5.1.3 - DNA blot of $\underline{sn}^{X2}$ and parental chromosome

DNA samples of the genotype denoted above each lane were digested to completion with Pst1 and Sal1. Samples were then separated on an agarose gel and blotted. The DNA blot was hybridised to the interval -2.0 to 0.0. Both $\underline{sn}^{X2}$ and In(1)dl 49,y Hw m^2 g^4 (Δ49) contain an 80bp insertion within the wildtype region -2.0 to -1.3. The insertion increases the size of this Pst1 Sal1 band from 720bp to 800bp. The marker lane (M) contains λCI857 cut with BstEII. Below the blot are physical maps of $\underline{sn}^{X2}$ and $\underline{sn}^{+}$ within the region -2.0 to 0.0.

Figure 5.1.3


X2
+
Δ49
+
+
+
M
7000
800
720
S
Pst 1 Pst 1
E
Pst 1
sn+
-2.0
0.0
S
Pst 1 Pst 1
E
Pst 1
snX2
Probe
1kb

with this allele the DNA structure of the parental chromosome was examined. The $\underline{sn}^{X2}$ mutation arose by X-ray mutagenesis of the In(1)dl-49 chromosome (Bender,1960). The structures of $\underline{sn}^{X2}$ and In(1)dl-49 were compared using DNA from heterozygous $\underline{sn}^{X2}/\underline{sn}^{+}$ and In(1)dl-49,y Hw m^2 $g^4/\underline{sn}^{+}$ females. Figure 5.1.3 shows a DNA blot comparing the sn region of each chromosome. Both the $\underline{sn}^{X2}$ and In(1)dl-49 chromosomes contain the 80bp insertion at -1.80 (Figure 5.0.1). Since In(1)dl-49 carries a wildtype copy of the sn gene, the mutant phenotype of $\underline{sn}^{X2}$ cannot be due to the 80bp insertion at -1.80 (Figure 5.0.1). The insertion is presumably a polymorphism which does not effect sn expression.

Note the genetic location of the $\underline{sn}^{X2}$ lesion is not known and cannot be easily mapped because of the inversion present in the In(1)dl-49 chromosome. DNA blotting and sequencing analysis have failed to identify the mutagenic lesion in this allele, so $\underline{sn}^{X2}$ was excluded from further investigations.

5.1.2 Analysis of the sn^3 mutation

The $\underline{sn}^{3}$ clone λB1 was isolated from the BamH1 library and contains an 16.0kb insert. To facilitate detailed analysis of the $\underline{sn}^{3}$ DNA lesion, (-0.9 to -0.0) sequences homologous to the interval -2.0 to 0.0 were subcloned from λB1 into pBluescribe-M13(-) and sequenced. The $\underline{sn}^{3}$ sequence was compared to the $\underline{sn}^{+}$ sequence from -1.16 to 0.0 (Figure 5.1.4). This shows there is a deletion which removes nucleotides 1183 to 1475 (Figure 3.5.2 and Figure 5.1.4). This 292bp deletion removes 86bp upstream and 206bp downstream of the transcriptional start site (1269 Figure 3.5.2). In addition to removing the transcription start site the deletion is likely to eliminate promoter sequences used to initiate sn RNAs (Breathnach and Chambon, 1981). Deletion of promoter elements normally abolishes transcription. However the phenotype of $\underline{sn}^{3}$ suggests that some sn product is being produced in ovaries in order to prevent female sterility (See Section 5.3).

Figure 5.1.4 - Sequence alignment between $\underline{sn}^3$ and $\underline{sn}^+$

The DNA sequence of $\underline{sn}^+$ (lower line) from the Pst1 site at -1.2 to the EcoR1 site at 0.0 is compared to the corresponding region of $\underline{sn}^3$ sequence (upper line). A perfect match between sequences is denoted by a star, gaps are introduced into either sequence to maximise the number of matched nucleotides. DNA maps below the sequence data show the DNA structure of $\underline{sn}^3$ and $\underline{sn}^+$ within the region -2.0 to 0.0. The shaded box below the $\underline{sn}^3$ DNA map denotes the location of $\underline{sn}^3$ sequence compared to $\underline{sn}^+$ (above). The $\underline{sn}^+$ sequence is numbered as in Figure 3.5.2A.

Figure 5.1.4

```
CTGCAGCTGCAATTAAATATGCAATGCACTTAAGTTATGCACATTTCCTTTGCCCAGGGACCGTTACGATTTCGTATTTTAAAAGCCGCGCGAAAGCTAA
**********************************************************************************************************
CTGCAGCTGCAATTAAATATGCAATGCACTTAAGTTATGCACATTTCCTTTGCCCAGGGACCGTTACGATTTCGTATTTTAAAAGCCGCGCGAAAGCTAA
 Pst 1                                                                                                  892
742
CAGCGCTGTTGGCTAGCAGACGTCTGTTTCTGCTGTTGTTAGCCAACAGCCAACAACAACAAACTGGCTGTTGTTTCGAATGCACTGCGATCTTAGCTCT
**********************************************************************************************************
CAGCGCTGTTGGCTAGCAGACGTCTGTTTCTGCTGTTGTTAGCCAACAGCCAACAACAACAAACTGGCTGTTGTTTCGAATGCACTGCGATCTTAGCTCT
843                                                                                                    973

TTTCCCGCTAAATGTTTCTTTGTGTGTCATTTTGTCAACCATTCGATCGATATCGATATTCCTTTGTGAAGGCTAAAAATAAAATTCTTACATGCTCATC
**********************************************************************************************************
TTTCCCGCTAAATGTTTCTTTGTGTGTCATTTTGTCAACCATTCGATCGATATCGATATTCCTTTGTGAAGGCTAAAAATAAAATTCTTACATGCTCATC
974                                                                                                   1074

TTATACAATAATGGATCGCACATATAGCTACTCCTTTTTCATGTATCCAATTCACTAGGTAAGCTACAAAGATAGCCAGACGGCTGTGCTAACAGAGCGC
**********************************************************************************************************
TTATACAATAATGGATCGCACATATAGCTACTCCTTTTTCATGTATCCAATTCACTAGGTAAGCTACAAAGATAGCCAGACGGCTGTGCTAACAGAGCGC
1075                                                                                                  1175

T
TGCTATGGTGAGAAGCCGCCGCAATGGCAACAGTGCCATCTCTGGCCTTCGATGCAAATAGCGCAACTTCGCTCGCAACGGGTTTCATTCCCACTGGAGT
1176                                                                                                  1276


GCAGTTCGTGAGCGGTCGTTCGCGCCGCGCGCGCGCAAAAGTAAAAGTTAAAGTTTTGTGAGCGCATGCGTCTCGTCTTAGAAAAACCTGCAATCAGTAA
1277                                                                                                  1377

                                                                                                TACAAAT
                                                                                                *******
TCTCGAATCGTGAAAATCCTTTTCTGCCATACCAAAAAAAGAGCCGGACAAAGCAAAGCGTTCGATTGCCTTCTGCCAATTTGAGCAGGAAATTACAAAT
1378                                                                                                  1478

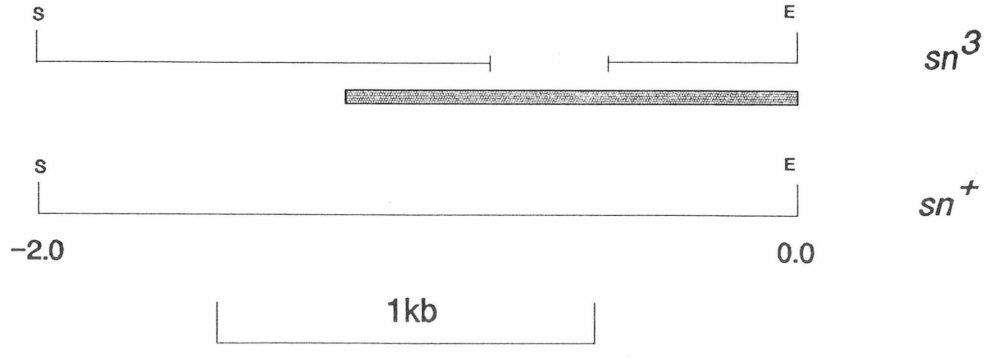
TATTGCCGCCCTGACCGTGATTTGTTTACGTTTTTTTCTTCACTTTTTTTTTACATTGTGTGAGTGTGAGAAATTTGTTGGCCAAAAAATATTCCAGATAT
**********************************************************************************************************
TATTGCCGCCCTGACCGTGATTTGTTTACGTTTTTTTCTTCACTTTTTTTTTACATTGTGTGAGTGTGAGAAATTTGTTGGCCAAAAAATATTCCAGATAT
1479                                                                                                  1579

TCCCAGAGATATATATATTCGAAATTCGAATTGGCGAAGTGAAGGGTGTGACGACAGACGCACACACACGCACGCACACGAAAACGCACTGAAAAAACAC
**********************************************************************************************************
TCCCAGAGATATATATATTCGAAATTCGAATTGGCGAAGTGAAGGGTGTGACGACAGACGCACACACACGCACGCACACGAAAACGCACTGAAAAAACAC
1580                                                                                                  1680

CGGCTAAAAGGCAGACAAAAAGGAGACGAGAGAGCGAGTGCTAAATAGAAAACGCGAAAACAGGATAATAATCAGCAAACTAATACTATAAAACAAGCAA
**********************************************************************************************************
CGGCTAAAAGGCAGACAAAAAGGAGACGAGAGAGCGAGTGCTAAATAGAAAACGCGAAAACAGGATAATAATCAGCAAACTAATACTATAAAACAAGCAA
1681                                                                                                  1781

CTAAGTGACATCTCGATCTCGCCGAGAAACAACAAAAACAACAGTTACAACTTCGGCAACAGCAACAAAAACAGCGCCAGCAGCAGCAACAACAACAAAA
**********************************************************************************************************
CTAAGTGACATCTCGATCTCGCCGAGAAACAACAAAAACAACAGTTACAACTTCGGCAACAGCAACAAAAACAGCGCCAGCAGCAGCAACAACAACAAAA
1782                                                                                                  1882

TCAGCGAAGAGCATAGAGCCAAAAAGAAGGAGCAGCAAAAGCAAAGAAAGTAAGTGGAATTC
**********************************************************************
TCAGCGAAGAGCATAGAGCCAAAAAGAAGGAGCAGCAAAAGCAAAGAAAGTAAGTGGAATTC
1883                                                    Eco R1
                                                        1945
```



5.2 Introduction to mutant RNA blotting experiments

The following series of experiments were designed to investigate the structure of RNA molecules produced in sn^3, sn^w, sn^e and $sn^{(+)}$ mutant alleles of *sn*. Describing the structure of these transcripts provides an insight into the mechanisms whereby each allele disrupts *sn* expression.

Transcription of these *sn* mutants was analysed at 3 stages of the *Drosophila* lifecycle. Adult bristle and hairs are made during metamorphosis thus RNA from homozygous pupae was analysed to investigate how mutant transcripts might result in the disruption of bristle formation. Somatic *sn* gene expression, initiated in pupae, continues into the adult stages of *Drosophila* development (Chapter 4). Transcription of *sn* in adult males appears to be limited to somatic tissue. Adult stage somatic cell expression of *sn* can therefore be studied by analysing transcripts detected in adult male RNA samples. In contrast to adult males, adult females require *sn* product for wildtype development in both somatic and germline tissue (Chapter 4). Germline transcription of *sn* results in 3 RNAs of 3.0, 3.3 and 3.6kb. The 3.0 and 3.3kb RNAs are specific to the adult female germline. Detection of these RNAs or their equivalents (see below) demonstrates that *sn* transcription is occurring within ovaries. To avoid complications of P-M cytotype, P-element alleles were studied in the M cytotype.

The structure of mutant transcripts was defined using a series of short DNA fragments as probes on RNA blots. The hybridisation pattern of 3 probes, adjacent to or within each DNA lesion was determined for each allele over the 3 stages of development chosen for study. Knowing the DNA sequence of each lesion, the location of the 3 probes and the approximate size of each transcript allows the gross structure of each mutant RNA to be determined.

Filters carrying RNA samples of wildtype and each

Figure 5.2.1 - RNA blots of pupal $\underline{sn}^3$, $\underline{sn}^w$, $\underline{sn}^e$ and $\underline{sn}^{(+)}$ samples

Poly-A^+ RNA was extracted from homozygous $\underline{sn}^+$, $\underline{sn}^3$, $\underline{sn}^w$, $\underline{sn}^e$ and $\underline{sn}^{(+)}$ pupae. RNA blots carrying aliquots of these samples were hybridised to 3 probes. The blot hybridised to probe A (-0.71 to -1.33) is shown in panel A, to probe B (-0.51 to -0.59) in panel B and to probe C (-0.35 to -0.5) in panel C. The location of each probe with respect to the DNA lesions in $\underline{sn}^3$, $\underline{sn}^w$, $\underline{sn}^e$ and $\underline{sn}^{(+)}$ is shown in the composite physical map shown below the RNA blots. This diagram covers the interval from -2.0 to 0.0 and the black box on the $\underline{sn}^+$ map denotes the location of the first sn exon. Equivalent quantities of RNA were loaded in each lane as judged by hybridisation to an actin probe (data not shown). The RNA in panel B appears to be slightly degraded.

Figure 5.2.1

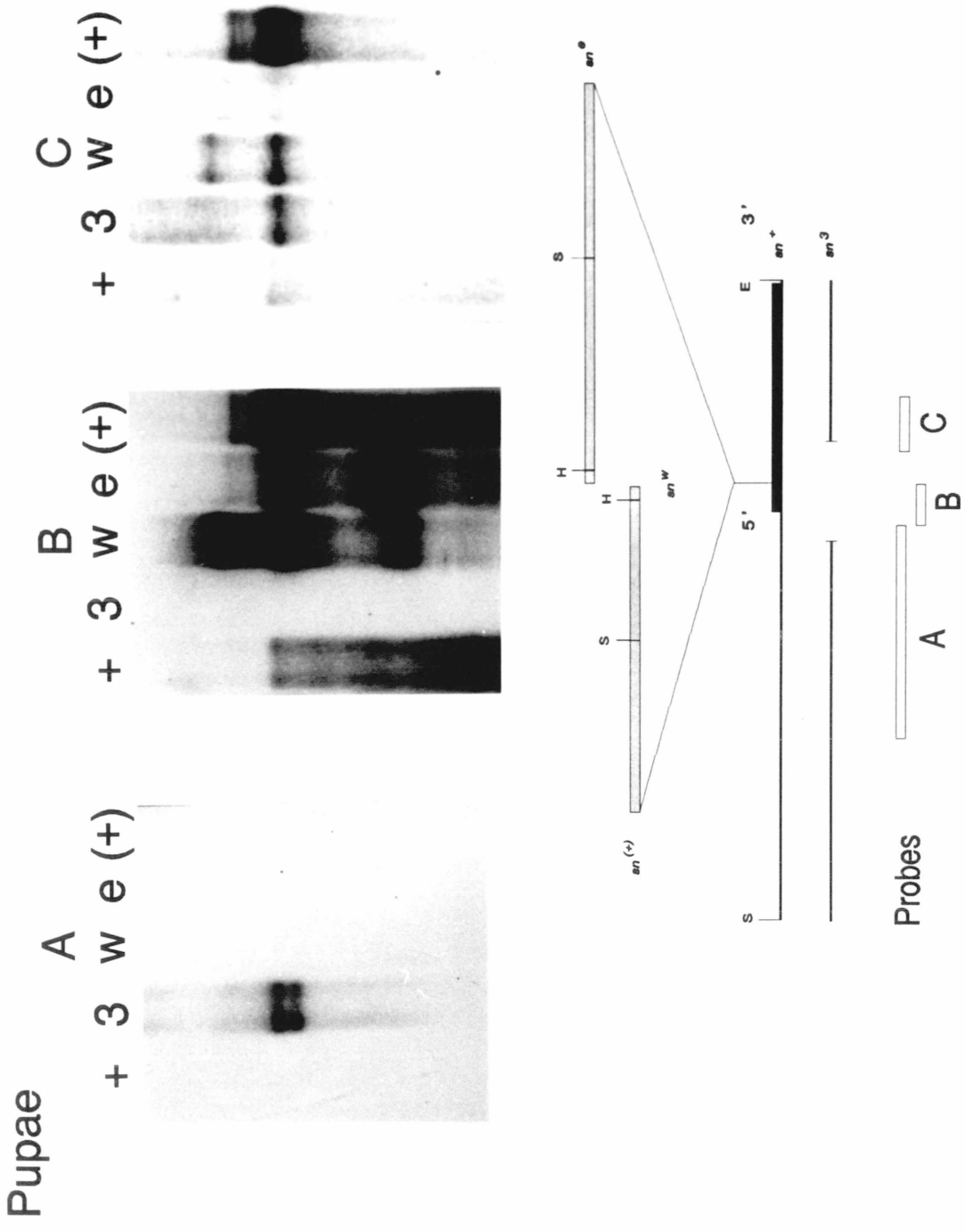


Figure 5.2.2 - RNA blots of adult male $\underline{sn}^3$, $\underline{sn}^W$, $\underline{sn}^e$ and $\underline{sn}^{(+)}$ samples

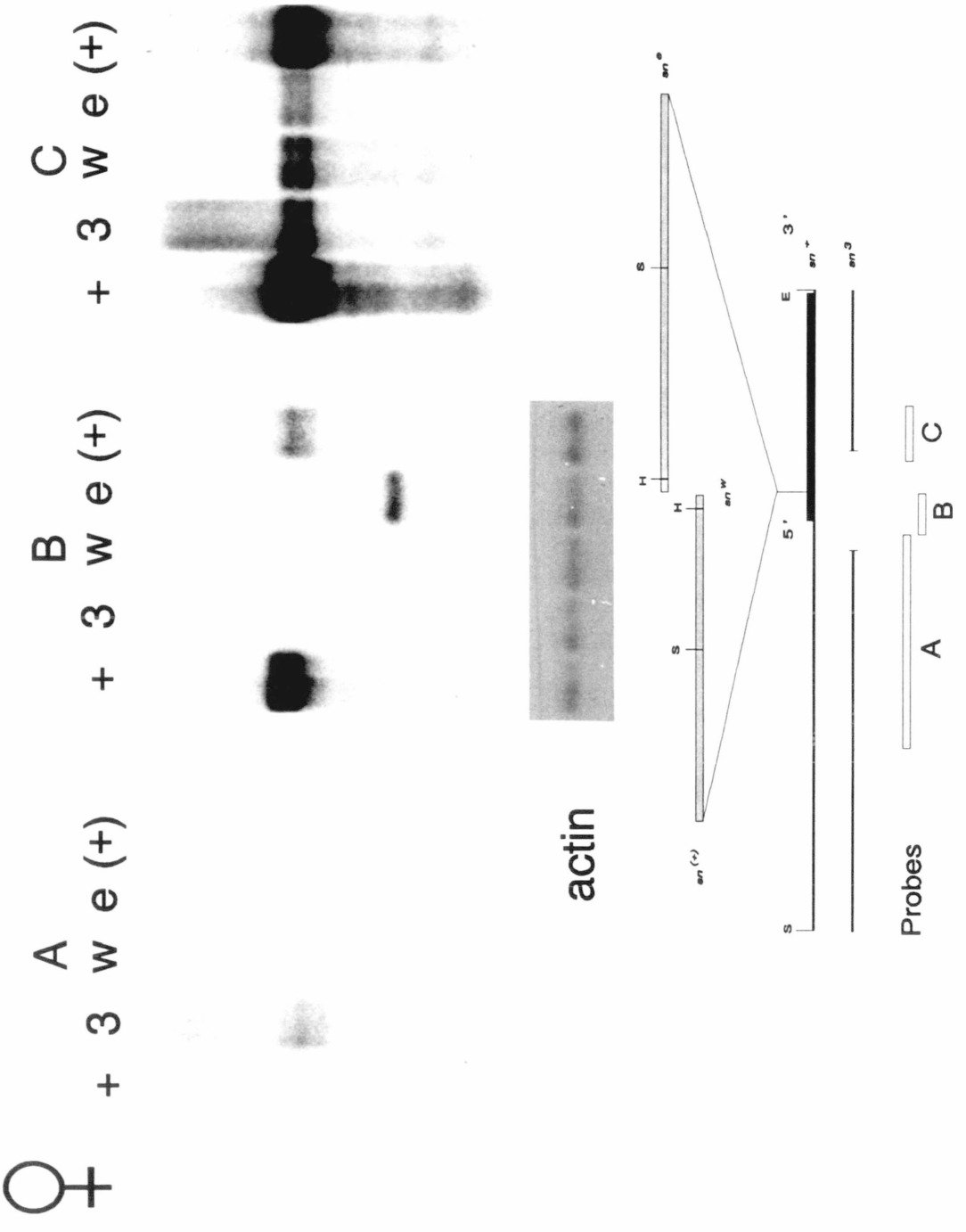
Poly-A^+ RNA was extracted from homozygous $\underline{sn}^+$, $\underline{sn}^3$, $\underline{sn}^W$, $\underline{sn}^e$ and $\underline{sn}^{(+)}$ adult males. RNA blots carrying aliquots of these samples were hybridised to 3 probes. The blot hybridised to probe A (-0.71 to -1.33) is shown in panel A, to probe B (-0.51 to -0.59) in panel B and to probe C (-0.35 to -0.5) in panel C. The location of each probe with respect to the DNA lesions in $\underline{sn}^3$, $\underline{sn}^W$, $\underline{sn}^e$ and $\underline{sn}^{(+)}$ is shown in the composite physical map shown below the RNA blots. This diagram covers the interval from -2.0 to 0.0 and the black box on the $\underline{sn}^+$ map denotes the location of the first $\underline{sn}$ exon. Equivalent quantities of RNA were loaded in each lane as judged by hybridisation to an actin probe (data not shown).

Figure 5.2.2

Figure 5.2.3 - RNA blots of adult female $\underline{sn}^3$, $\underline{sn}^W$, $\underline{sn}^e$ and $\underline{sn}^{(+)}$ samples

Poly-A^+ RNA was extracted from homozygous $\underline{sn}^+$, $\underline{sn}^3$, $\underline{sn}^W$, $\underline{sn}^e$ and $\underline{sn}^{(+)}$ adult females. RNA blots carrying aliquots of these samples were hybridised to 3 probes. The blot hybridised to probe A (-0.71 to -1.33) is shown in panel A, to probe B (-0.51 to -0.59) in panel B and to probe C (-0.35 to -0.5) in panel C. The location of each probe with respect to the DNA lesions in $\underline{sn}^3$, $\underline{sn}^W$, $\underline{sn}^e$ and $\underline{sn}^{(+)}$ is shown in the composite physical map shown below the RNA blots. This diagram covers the interval from -2.0 to 0.0 and the black box on the $\underline{sn}^+$ map denotes the location of the first $\underline{sn}$ exon. Equivalent quantities of RNA were loaded in each lane as judged by hybridisation to an actin probe.

Figure 5.2.3



mutant from a specific period of development were hybridised to each of the 3 probes. For pupae these results are shown in Figure 5.2.1, RNA blots of adult male samples are shown in Figure 5.2.1 and the adult female data is shown in Figure 5.2.3.

The format of these RNA blots (Figures 5.2.1-3) makes analysis of individual alleles rather difficult. To overcome this the hybridisation patterns of each allele were extracted from Figures 5.2.1-3 and diagramatically represented in a allele specific summary figures e.g. Figure 5.3.1.

5.3 RNA blotting analysis of sn^3

The DNA lesion in sn^3 has been characterised as a 292bp deletion which removes some putative promoter sequences, the transcriptional start site, and part of the first exon (Section 5.1). Transcription still occurs in this mutation but the resulting sn^3 RNAs are of slightly reduced molecular weight (Figure 5.1.2). Transcription of sn^3 is unable to provide sn function during metamorphosis but does supply sn product during oogenesis, thereby preventing female sterility.

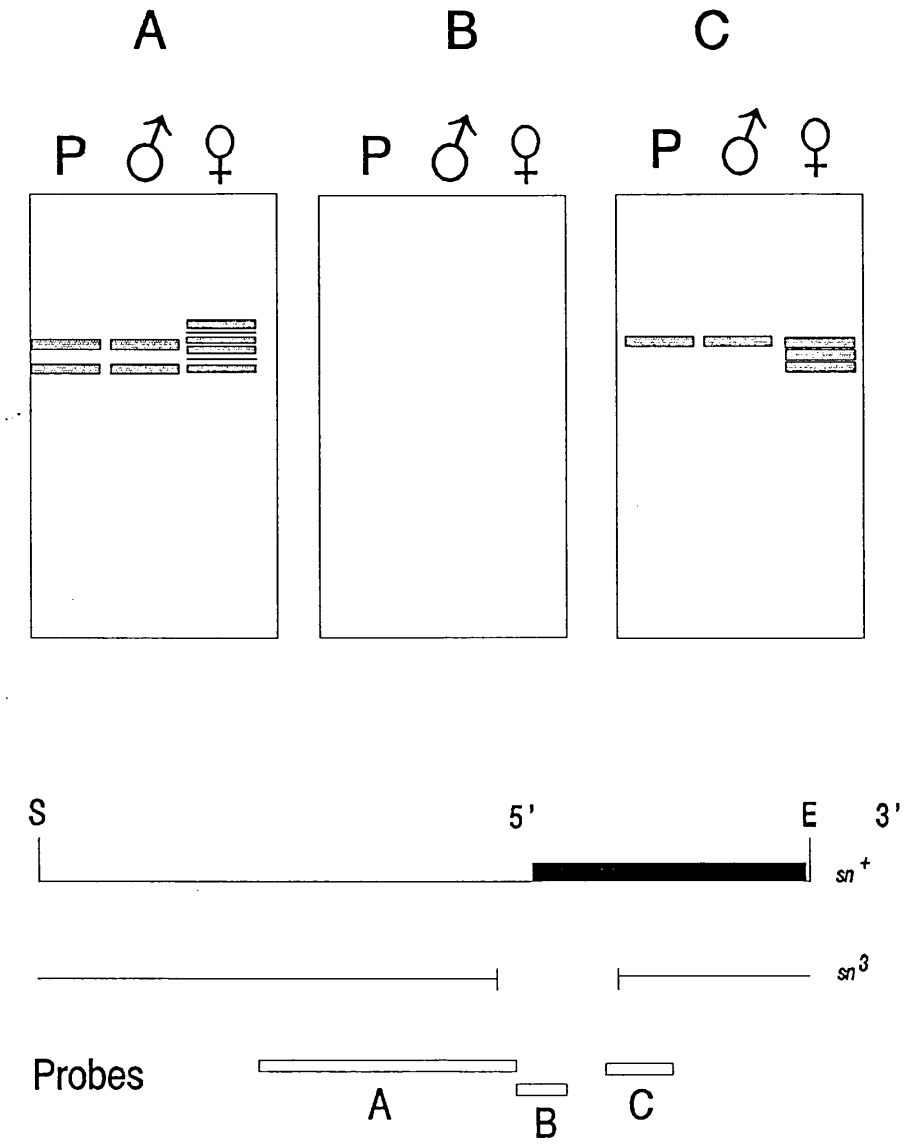
To investigate the structure and distribution of these mutant transcripts, poly-A^+ RNA was prepared from pupae, adult males and adult females. RNA blots carrying aliquots of these samples were hybridised to 3 probes. Probe A (607 to 1231 Figure 3.5.2A) includes sequences upstream of the sn^3 deletion and terminates within the deleted interval (Figure 5.3.1). The sequences in probe B (1346 to 1429 Figure 3.5.2A) lie completely within the sn^3 deletion and probe C (1444 to 1605 Figure 3.5.2A) starts within the sn^3 deletion and includes sequences downstream of the deletion (Figure 5.3.1). The original RNA blotting data are shown in Figures 5.2.1-3 and a diagramatical representation of sn^3 data is shown in Figure 5.3.1.

Probe A detects two transcripts of approximately 3.5 and 2.9kb in pupal and adult male RNA Figure 5.3.1. Adult female RNA contains at least 4 transcripts homologous to

Figure 5.3.1 - Summary diagram of $\underline{sn}^3$ RNA blotting data.

The $\underline{sn}^3$ RNA blotting data shown in Figures 5.2.1-3 has been collated and is summarised in 3 diagrams representing RNA blots. The intensity of transcript bands is represented by the thickness of boxes in this figure. Physical maps of the wildtype sn gene from -2.0 to 0.0, the $\underline{sn}^3$ lesion and the location and size of the 3 probes used in these investigations is shown below the diagrammatic RNA blots. P = pupal RNA, ♂ = adult male RNA, ♀ = adult female RNA.

Figure 5.3.1



probe A in the size range 2.9 to 4.0kb. Probe B sequences are deleted from $\underline{sn}^3$ DNA thus no transcripts are detected with this probe. Probe C detects a single 3.5kb transcript in pupae and adult males and a triplet (2.9 ,3.2, and 3.5kb) of RNAs in adult females.

The large number of $\underline{sn}$ transcripts detected by probes A and C in $\underline{sn}^3$ RNA can be split into two groups. Each transcript which hybridises to probe C is detected by probe A. Thus probe C detects a subset of probe A transcripts.

Transcripts detected by probe C are all approximately 50 to 100bp smaller then equivalent $\underline{sn}^+$ RNAs. The fact that all probe C RNAs contain a region homologous to probe A shows that these transcripts carry sequences derived from upstream of the $\underline{sn}^+$ transcript initiation site. It appears that all $\underline{sn}^3$ RNAs start at a novel transcript initiation site(s) within probe A. Thus normally upstream sequences are included in $\underline{sn}^3$ RNAs. The near wildtype size of transcripts detected by probe C suggests that RNA molecules initiated at the novel transcript initiation site(s), transcribe the entire $\underline{sn}$ gene and terminated at the $\underline{sn}^+$ polyadenylation sites. The suggestion that $\underline{sn}^3$ RNAs use $\underline{sn}^+$ polyadenylation sites is consistent with the wildtype pattern of transcript size differences seen in $\underline{sn}^3$ adult female RNA. This model assumes that these probe C hybridising transcripts initiate within a few base pairs of each other and that the wildtype splicing pattern is used to process these RNAs. The slight reduction in molecular weight reflects structural and size alterations in the first exon caused by the 292bp deletion and novel $\underline{sn}^3$ transcript initiation site(s). This explanation is consistent with the RNA blotting data, the site of the $\underline{sn}^3$ lesion and the structure of the DNA alteration in $\underline{sn}^3$.

Since the first exon of $\underline{sn}$ is non-coding these $\underline{sn}^3$ transcripts could be translated to produce functional $\underline{sn}$ protein.

There are several transcripts in $\underline{sn}^3$ RNA which do hybridise to probe A but not probe C. RNA from homozygous $\underline{sn}^3$ pupae and adult males contains a 2.9kb $\underline{sn}$ transcript

which is only detected by probe A. This RNA appears to initiate at a novel transcript initiation site(s). However sequences homologous to probe C have been lost from this transcript. Loss of probe C sequences suggests that a previously unused 5′ splice site is being used to excise RNA derived from region C. The size of this transcript (2.9kb) is consistent with the new intron ending at the normal second exon acceptor site. The rest of the molecule must then be subject to the wildtype pattern of RNA processing. Computer based sequence analysis of the $\underline{sn}^3$ sequence has identified several potential splice donor site within the interval covered by probe A (data not shown). Novel splicing from one of these 5′ splice sites to the second exon acceptor site appears to be the most logical explanation for this transcript. Primer extension, S1 protection and cDNA analysis would be required to locate the precise 5′ splice site used by these RNAs.

Note relatively few transcripts are processed in this manner: RNAs detected by probe C have not used this novel splice site(s).

RNA extracted from $\underline{sn}^3$ adult females contains two transcripts (3.6 and 4.0kb) which specifically hybridise to probe A. These transcripts appear to initiate at the same start site as the near wildtype $\underline{sn}^3$ RNAs and then splice out probe C sequences by using the novel donor site(s) described above. However the increased molecular weight of these transcripts suggests that further RNA processing abnormalities occur. Without additional RNA blotting, S1 fragment protection and primer extension experiments no assessment of transcript structure can be made.

There appears to be an inverse relationship between the presence of the 3.6 and 4.0kb transcripts and the mutant phenotype of $\underline{sn}^3$. The 3.6 and 4.0kb transcripts are absent in pupae where the $\underline{sn}^3$ bristle phenotype develops but are present in adult females where $\underline{sn}^3$ individuals are fertile. Assuming that all other $\underline{sn}^3$ RNAs fail to produce active sn protein then these transcripts may provide sn function in adult females to prevent sterility.

Transcripts which initiate at the novel sn^3 transcript start site(s) and then use the sn^+ intron-exon splice junctions are 2.9, 3.2 and 3.5kb in length. This suggests that the 3.6 and 4.0kb transcripts are subject to abnormal RNA processing. The additional sequences present in these transcripts may disrupt the sn ORF resulting in a non-functional protein. If these transcripts are non-functional then other explanations of the sn^3 phenotype are required.

The loss of the sn^+ transcript initiation site and promoter sequences in sn^3 individuals does not abolish transcription. The sn^3 deletion appears to activate a cryptic promoter and a novel transcript initiation site is used to produce sn^3 RNAs.

The concept of a cryptic promoter can be used in several ways to explain the phenotype of sn^3. The first explanation suggests that the sn^3 cryptic promoter fails to transcribe the sn gene in bristle and hair forming tissue. Transcripts detected in sn^3 pupae may be derived from ectopic sites of transcription.

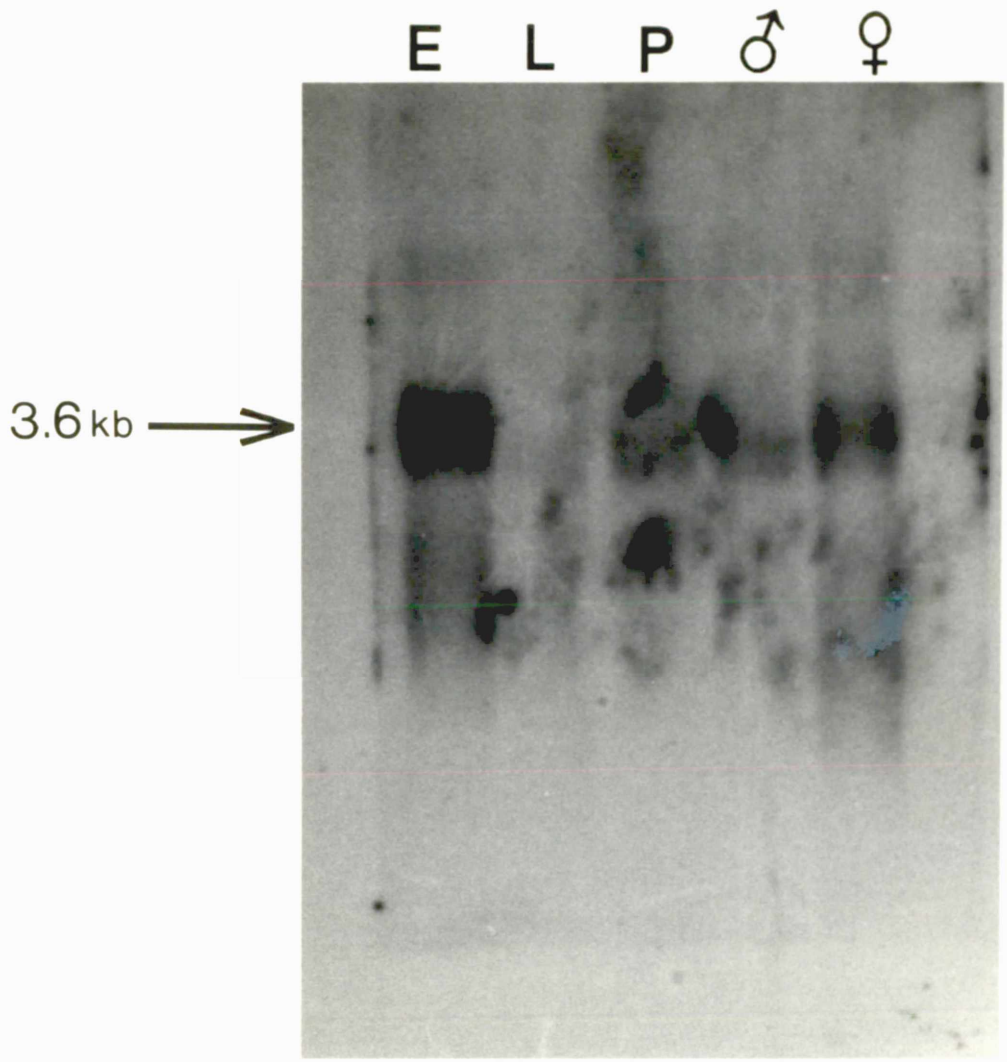
An alternative explanation is to suggest that the sn^3 cryptic promoter transcribes the sn gene in the correct tissue but it fails to produce sufficient transcripts capable of producing sn protein, to supply the demand for sn function in bristle and hair forming cells. Studies on the P-element alleles sn^w, sn^e and $sn^{(+)}$ (see below) show small reductions in pupal sn protein levels result in altered bristle morphology.

To investigate the possibility that the sn^3 cryptic promoter was ectopically active. Poly-A^+ RNA was prepared from sn^3 embryos, larvae, pupae, adult males and adult females. An RNA blot carrying samples of this material hybridised to the region -1.3 to -0.7 (probe A) is shown in Figure 5.3.2. In contrast to sn^+ (Figure 3.2.1), sn^3 transcripts are detected in larvae. The detection of sn^3 transcripts in larval RNA proves that the sn^3 cryptic promoter has altered tissue specificity with respect to the wildtype promoter.

Figure 5.3.2 - Expression of $\underline{sn}^3$ transcripts during development.

Poly-A^+ RNA was extracted from homozygous $\underline{sn}^3$ embryos, larvae, pupae, adult males and adult females. An RNA blot carrying equivalent quantities of each RNA was probed with sequences from the sn locus (-0.71 to -1.33 Probe A). Transcripts of the $\underline{sn}^3$ gene are detected during all stages of development examined. E = 0-24hr embryos, L = 1^{st}, 2^{nd} and 3^{rd} instar larvae, P = pupae, ♂ = adult male, ♀ = adult female.

Figure 5.3.2



5.4 RNA blotting analysis of sn^e

The sn^e mutation has been characterised as a 1.3kb P-element insertion in the first exon of the sn gene (Section 5.1). The insertion results in an extreme bristle bristle phenotype but has no effect on female sterility in the M-cytotype.

To investigate the phenotype of this allele, RNA was prepared from sn^e pupae, adult males and adult females. RNA blots carrying these samples were hybridised to 3 probes. Probe A (607 to 1231 Figure 3.5.2A) is just upstream of the the wildtype transcript initiation site. Probe B (1346 to 1429 Figure 3.5.2A) is upstream of the P-element insertion site but within the first exon of the wildtype sn gene. Probe C (1444 to 1605 Figure 3.5.2A) lies downstream of the sn^e insertion site and is part of the wildtype first exon. The original RNA blotting data are shown in Figures 5.2.1-3 and a diagramatical representation of the sn^e data is shown in Figure 5.4.1.

Probe A detects no transcripts in sn^e pupal, adult male and adult female RNA. Probes B and C both detect transcripts in sn^e RNA suggesting that transcription starts at the wildtype transcript initiation site in this mutation.

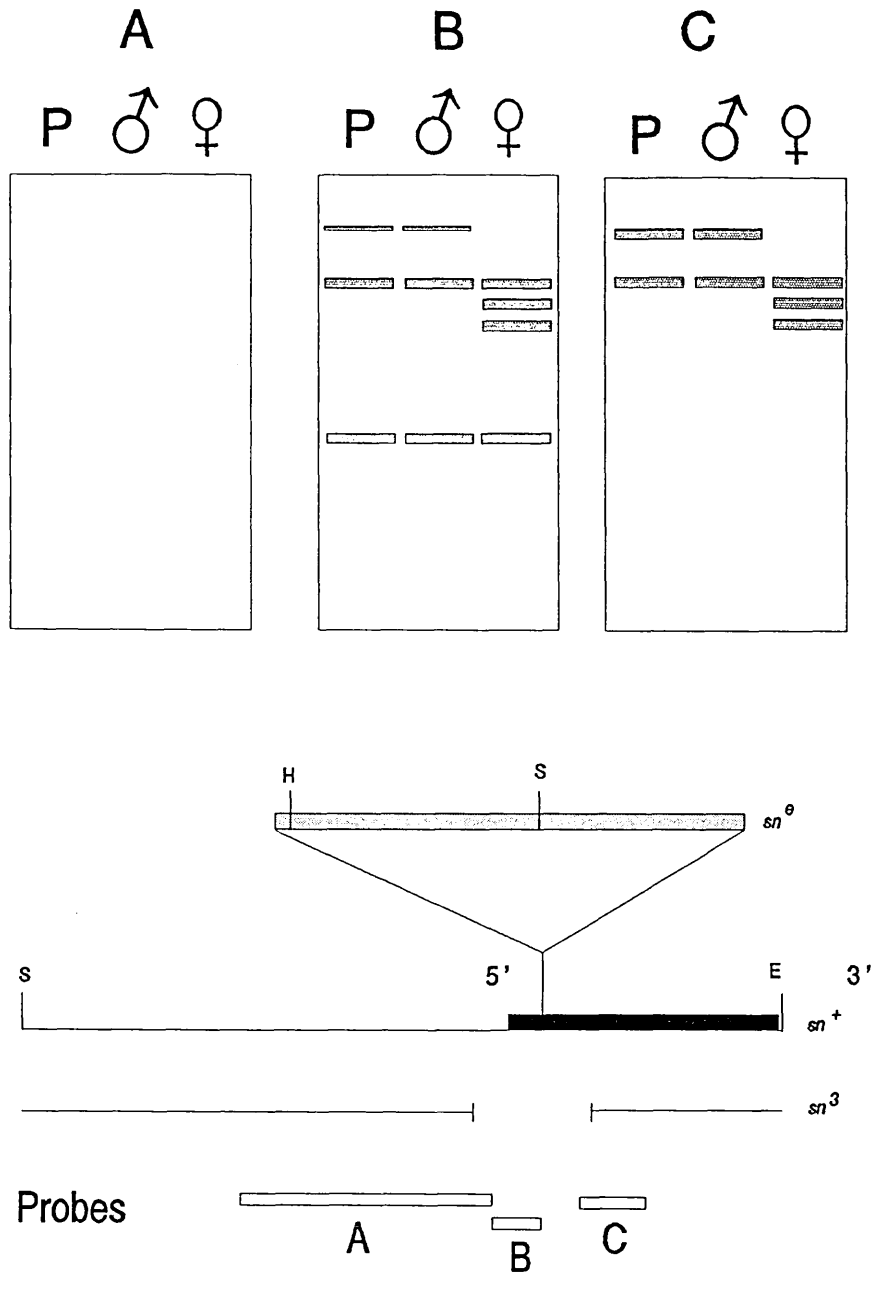
Probe B hybridises to 3 transcripts of 4.8, 3.6 and 1.5kb in sn^e pupal and adult male RNA. In sn^e adult females the probe detects transcripts of 1.5kb, 3.0, 3.3 and 3.6kb. Probe C detects 4.8kb and 3.6kb transcripts on blots of sn^e pupal and adult male RNA. Probe C detects 3 RNAs of approximately 3.0, 3.3 and 3.6kb in sn^e adult female RNA.

The simplest transcript to explain is the 1.5kb RNA detected in all lanes of sn^e RNA with probe B but not probe C (Figure 5.3.1). The sn^e P-element is inserted into the sn gene such that P-element transcription is in the same direction as sn transcription (left to right on Figure 5.3.1). This means that P-element splicing and polyadenylation signals incorporated into the transcripts

Figure 5.4.1 - Summary diagram of $\underline{sn}^{e}$ RNA blotting data.

The $\underline{sn}^{e}$ RNA blotting data shown in Figures 5.2.1-3 has been collated and is summarised in 3 diagrams representing RNA blots. The intensity of transcript bands is represented by the thickness of boxes in this figure. Physical maps of the wildtype $\underline{sn}$ gene from -2.0 to 0.0, the $\underline{sn}^{e}$ insertion and the location and size of the 3 probes used in these investigations is shown below the diagrammatic RNA blots. P = pupal RNA, ♂ = adult male RNA, ♀ = adult female RNA.

Figure 5.4.1



may be functional. Transcripts which initiate at the sn^+ start site and have transcribed the sn^e P-element may be terminated by the P-element polyadenylation site present at the 3' end of the sn^e insert. This would generate a 1.5kb RNA which would hybridise to probe B but not probe C.

The termination of transcripts within P-element insertions has also been described at the white (w) locus (Levis et al., 1984). In $w^{hd80k17}$ the inserted P-element polyadenylation site prematurely terminates approximately 70% of w transcripts. If sn^e transcripts are terminated in this way then significantly less than 70% of sn transcripts terminate at the P-element polyadenylation site (approximately 20-30%).

The inefficiency of the P-element polyadenylation site (above) allows transcription to continue through the entire sn gene. RNAs terminating at +15.0kb which have been normally spliced should be approximately 4.8kb in length (3.6kb of sn plus 1.3kb of P-element sequence). Pupal and adult male sn^e RNA contain transcripts of this size which hybridise to probes B and C. The size and hybridisation pattern of theses 4.8kb RNAs is consistent with the P-element readthrough structure proposed above. These transcripts would have the sn ORF intact, but have a much larger 5' untranslated region than sn^+ RNAs and so may not effectively produce sn protein when translated.

Note probes B and C do not detect readthrough transcripts in sn^e adult female RNA. This will be discussed in Section 5.8

Probes B and C also detect transcripts which are approximately the same size as sn^+ RNAs (Figure 5.3.1). In adult females the transcripts appear to use the wildtype multiple polyadenylation sites to produce 3.0, 3.3 and 3.6kb RNAs. The size of these RNAs suggests that in addition to using the sn^+ polyadenylation sites the normal pattern of splice donor/acceptor sites are used to process these transcripts. If this model is correct then the only way to generate wildtype sized RNAs is to the remove P-

element sequences from the initial transcript. A novel splicing event could remove the P-element sequences from the precursor readthrough RNA provided suitably located donor and acceptor sites exist. Splicing would then removes the intervening RNA. It appears that the majority of $\underline{sn}^{e}$ readthrough transcripts are processed in this manner.

To investigate the possibility of novel splicing, the wildtype sequence on either side of the $\underline{sn}^{e}$ insertion was searched for potential splice donor and acceptor sites. Computer based sequence analysis (Figure 5.4.2) shows that the $\underline{sn}^{e}$ insertion site lies between a donor site (D1, 1335-1336 Figure 3.5.2) and acceptor site (A1, 1362-1363 Figure 3.5.2). In $\underline{sn}^{+}$ RNA these sites (D1,A1) are separated by only 25bp. This interval is too small to be excised as an intron. (The smallest intron so far reported is a 31 nucleotide intron in the SV40 late transcript; Ghosk et al., 1978). In $\underline{sn}^{e}$ readthrough RNA (4.8kb Figure 4.5.1) D1 and A1 are sufficiently far apart that splicing enzymes may excise the intervening P-element sequences plus 25bp of $\underline{sn}$ sequence. This would result in a wildtype sized RNA.

The possibility that other splice sites, within the $\underline{sn}^{e}$ insertion, are used to remove P-element sequences was examined by analysing the $\underline{sn}^{e}$ sequence for splice donor/acceptor sites (Figure 5.4.3). In addition to D1/A1 there are several splices sites within the $\underline{sn}^{e}$ insertion which if used would produce approximately wildtype sized RNA. The most likely alternatives to D1/A1 are D2; a donor site located within the 31bp inverted repeats of P-elements and $A2^{e}$; an acceptor site close to the 3' end of the $\underline{sn}^{e}$ insertion. Combinations of splice sites such as $D1/A2^{e}$ and D2/A1 must also be considered as potential substrates for the novel splicing reaction. The analysis of $\underline{sn}^{(+)}$ and $\underline{sn}^{w}$ RNA (below) shows that wildtype sized transcripts are produced in both of these mutations. This suggests that insertions at the $\underline{sn}^{w}/\underline{sn}^{e}/\underline{sn}^{(+)}$ site are spliced out regardless of internal sequence composition. The only regions of shared sequence between these 3

Figure 5.4.2 - $\underline{sn}^{+}$ splice donor and acceptor site analysis.

The $\underline{sn}^{+}$ sequence from -2.0 to 0.0 (Figure 3.5.2A) was analysed for splice donor acceptor site using the Amersham Staden Plus sequence analysis software. The graphics output from this program has been annotated to mark the position of the first exon (shaded box above scale bar) and the site of $\underline{sn}^{w}$, $\underline{sn}^{e}$ and $\underline{sn}^{(+)}$ insertions. The axis above the scale bar represents the length of the sequence with the positions of consensus splice donor (D) and acceptor (A) sites (Staden,1984;Mount,1982) are marked in vertical lines along this axis. The height of donor/acceptor lines represents the degree of match each site has with its respective consensus sequence. The P-element insertions will separate the donor and acceptor sites D1 and A1. This may allow splicing enzymes to remove the intervening RNA. D^{sn} is the donor site used by wildtype transcripts to end the first exon. The scale bar is graduated at 50bp intervals.

Figure 5.4.2

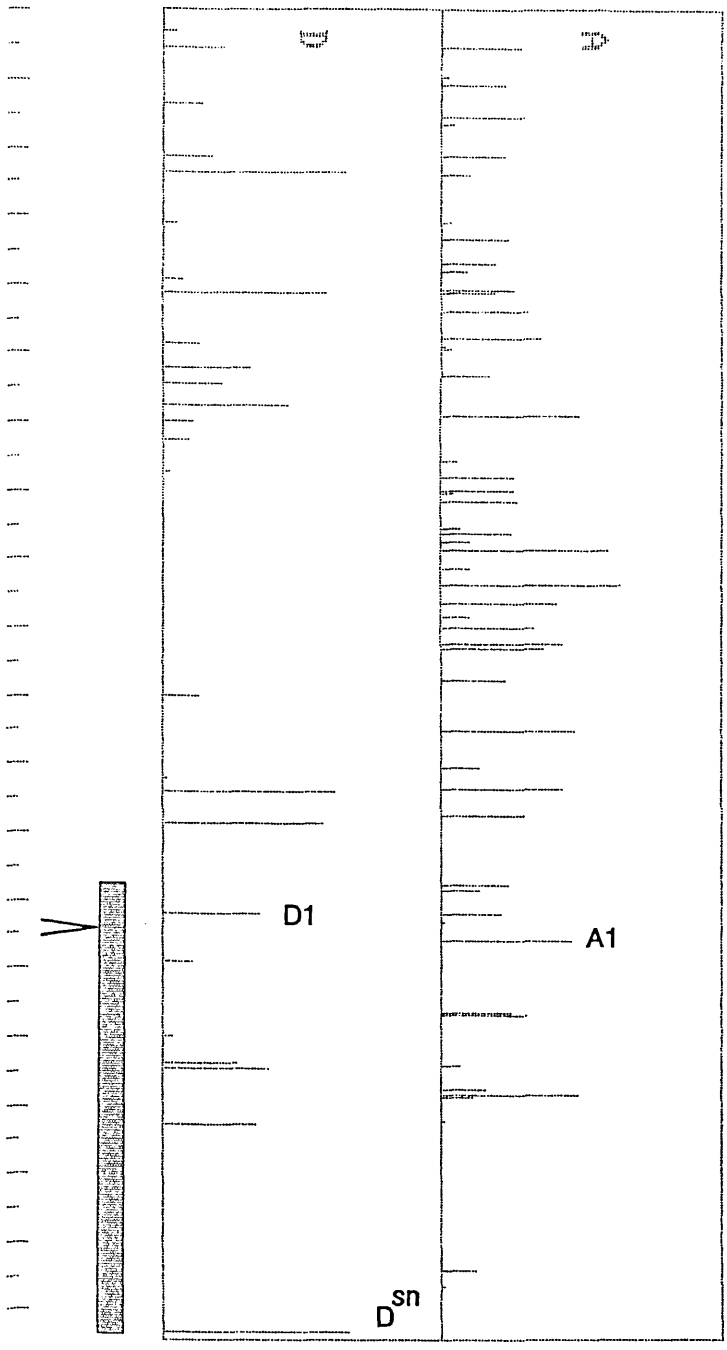
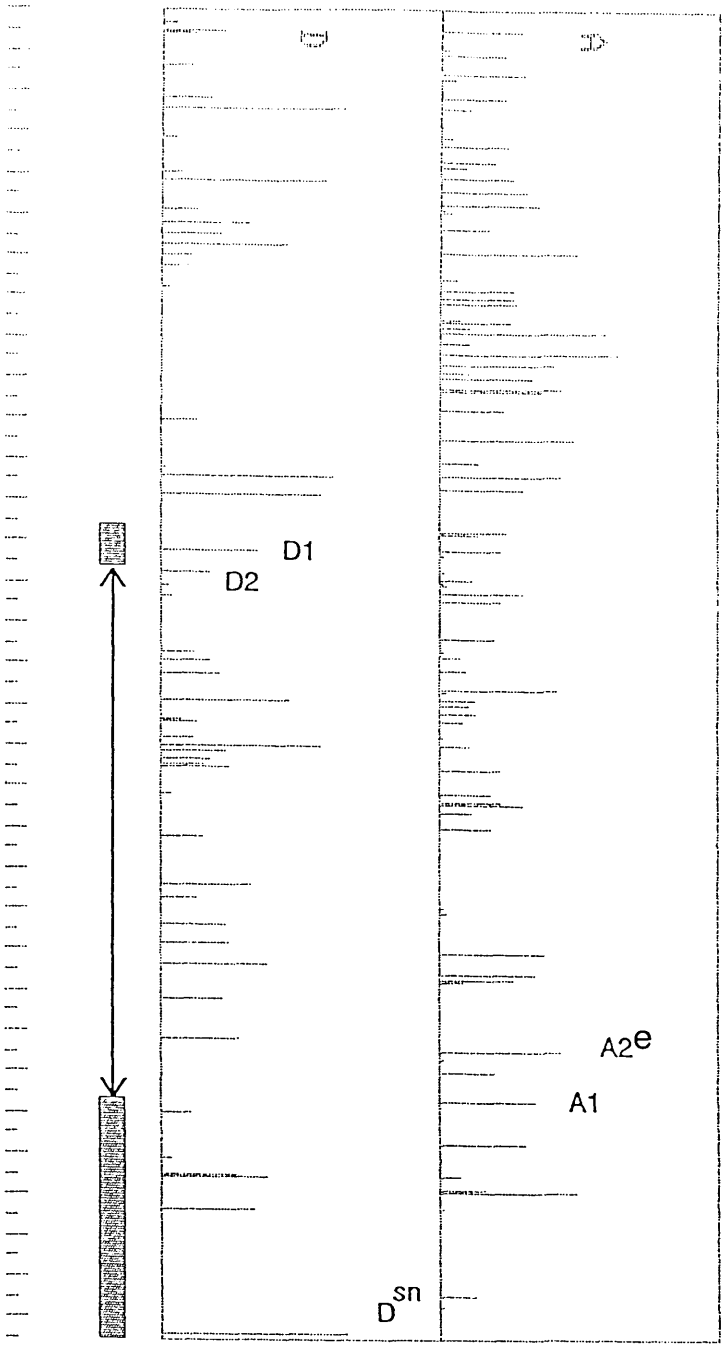


Figure 5.4.3 - $\underline{sn}^{e}$ splice donor and acceptor site analysis.

The $\underline{sn}^{e}$ sequence equivalent to the wildtype region from -2.0 to 0.0 but including the $\underline{sn}^{e}$ insertion (Figure A2) was analysed for consensus splice donor and acceptor sites using Amersham Staden Plus software. The graphics output of this program has been annotated to mark the position of the first exon (shaded box) and position and size of the $\underline{sn}^{e}$ P-element (double headed arrow). The $\underline{sn}^{e}$ insertion separates $\underline{sn}$ donor and acceptors sites D1, A1 (Figure 5.4.3) by 1.3kb thereby facilitating the removal of intervening P-element sequences. Two other suitable placed donor and acceptor sites are D2 and A2. The use of these sites in a splicing reaction would produce an approximately wildtype sized RNA. Because wildtype sized RNAs are detected in all 3 P-element alleles the novel splice is thought to occur between D1 and A1 (see text for details). D^{sn} is the donor site used by wildtype transcripts to end the first exon. The scale bar is graduated at 50bp intervals.

Figure 5.4.3



mutations are the 31bp P-element inverted repeats and flanking sn sequences. These observations favour the use of either D1/A1 and/or D2/A1 in the novel splicing event(s).

The type of computer based sequence analysis used to identify potential splice donor/acceptor sites is only capable of recognising sequence motifs which match consensus splice sites (Staden, 1984; Mount,1982). The identification of consensus splice sequences does not guarantee biological activity of the site. For instance, some of the splice sites used to produce P-element transcripts are barely detectable with this computer program.

The bristle phenotype of sn^e is the result of a lack of sn product during metamorphosis. Analysis of sn^e RNA shows that there are several transcripts in pupae which have the potential to produce wildtype sn function (3.5 and 4.8kb RNAs). Despite this sn^e individuals develop with gnarled and twisted bristles. Perhaps the level of sn protein supplied by these transcripts is below a threshold value required for normal bristle morphology. The lack of an oogenic phenotype is discussed in section 5.8.

5.5 RNA blotting analysis of $sn^{(+)}$

The $sn^{(+)}$ mutation is the result of a 0.95kb P-element insertion into the first exon of the sn gene (Section 5.1). No bristle phenotype is associated with this insertion. In the M-cytotype the allele is female fertile but in the P-cytotype $sn^{(+)}$ females are sterile. RNA blotting was used to examine $sn^{(+)}$ M-cytotype individuals.

RNA was extracted from $sn^{(+)}$ pupae, adult males and adult females. RNA blots carrying these samples were hybridised to 3 probes (Figures 5.2.1-3). A pictorial summary of this information is shown in Figure 5.5.1.

The upstream probe, probe A, detects no transcripts in the developmental stages of $sn^{(+)}$ examined (Figure 5.5.1). A pair of first exon probes (probes B and C) detect several transcripts in $sn^{(+)}$ RNA. The position of these probes (A,B and C) within the wildtype gene suggests that all $sn^{(+)}$ RNAs initiate at the wildtype transcription start site.

The transcripts detected in $sn^{(+)}$ can be subdivided into two groups: either wildtype sized transcripts (3.0, 3.3 and 3.6kb) or 4.5kb RNAs.

The size and hybridisation characteristics of the 4.5kb RNA suggests that it is a readthrough transcript composed of both sn and P-element sequences. The 0.95kb of $sn^{(+)}$ insertion sequence appears to be incorporated into transcripts of the $sn^{(+)}$ gene. Splicing removes sn introns from the primary transcript resulting in a 4.5kb RNA: 3.6kb of sn and 0.95kb of P-element sequences. In this readthrough transcript the entire sn protein coding interval is intact.

Note probe C detects a 4.5kb transcript in adult female RNA. This transcript is not detected by probe B and thus may originate from within the $sn^{(+)}$ P-element. However transcription from the $sn^{(+)}$ P-element promoter is in the opposite direction to sn transcription. Such P-element transcripts would not be detected by the single-stranded probes used in these investigations. The

Figure 5.5.1 - Summary diagram of $\underline{sn}^{(+)}$ RNA blotting data.

The $\underline{sn}^{(+)}$ RNA blotting data shown in Figures 5.2.1-3 has been collated and is summarised in 3 diagrams representing RNA blots. The intensity of transcript bands is represented by the thickness of boxes in this figure. Physical maps of the wildtype sn gene from -2.0 to 0.0, the $\underline{sn}^{(+)}$ insertion and the location and size of the 3 probes used in these investigations is shown below the diagrammatic RNA blots. P = pupal RNA, ♂ = adult male RNA, ♀ = adult female RNA.

Figure 5.5.1

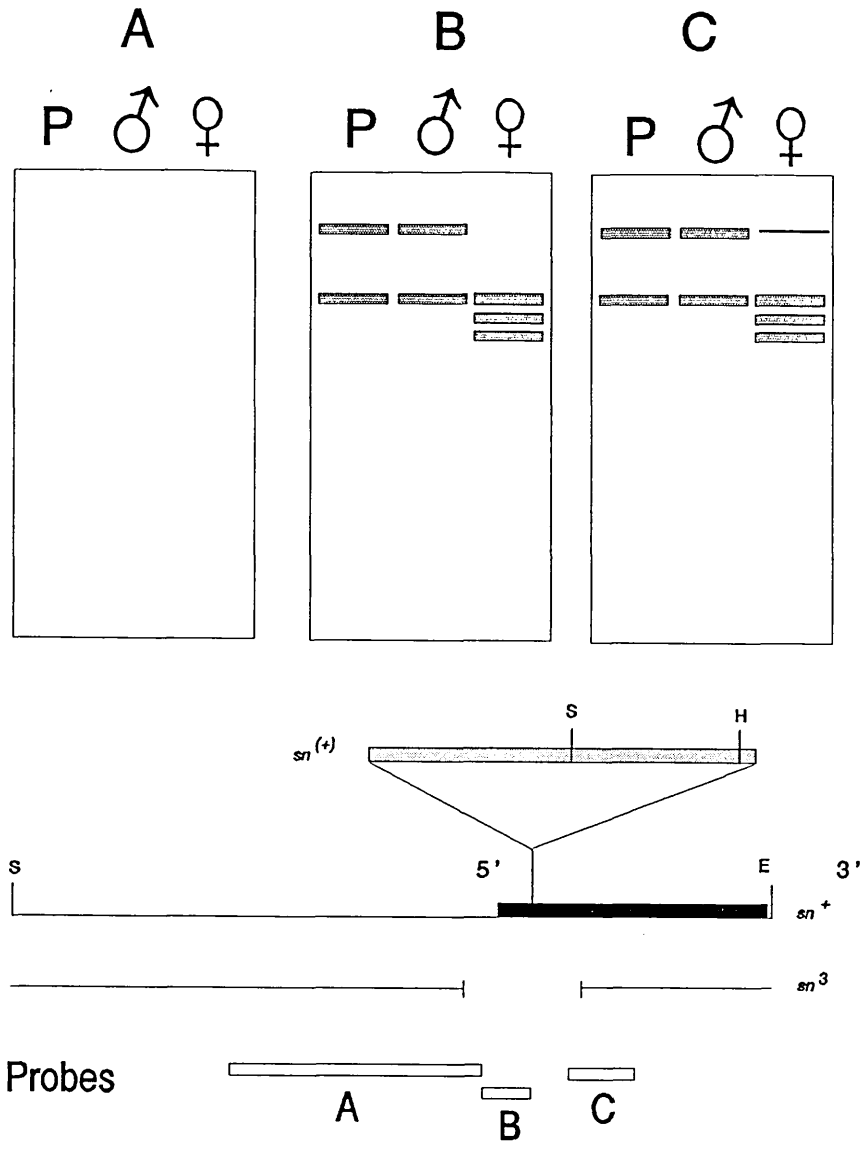
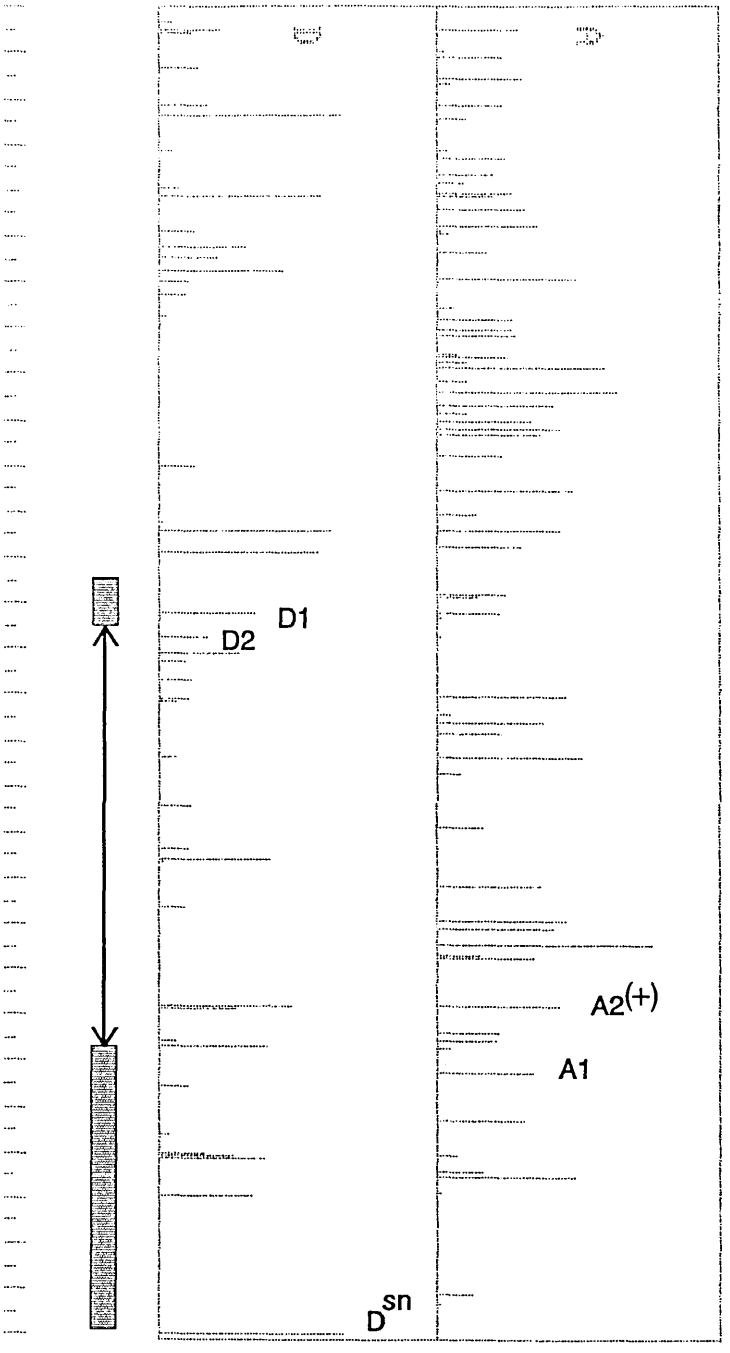


Figure 5.5.2 - $\underline{sn}^{(+)}$ splice donor and acceptor site analysis.

The $\underline{sn}^{(+)}$ sequence equivalent to the wildtype region from -2.0 to 0.0 but including the $\underline{sn}^{(+)}$ insertion (Figure A3) was analysed for consensus splice donor and acceptor sites using Amersham Staden Plus software. The graphics output of this program has been annotated to mark the position of the first exon (shaded box) and position and size of the $\underline{sn}^{(+)}$ P-element (double headed arrow). The $\underline{sn}^{(+)}$ insertion separates sn donor and acceptors sites D1, A1 (Figure 5.4.3) by 0.95kb thereby facilitating the removal of intervening P-element sequences. Two other suitable placed donor and acceptor sites are D2 and A2. The use of these sites in a splicing reaction would produce an approximately wildtype sized RNA. Because wildtype sized RNAs are detected in all 3 P-element alleles the novel splice is thought to occur between D1 and A1 (see text for details). D^{sn} is the donor site used by wildtype transcripts to end the first exon. The scale bar is graduated at 50bp intervals.

Figure 5.5.2



direction of transcription and size of the 4.5kb RNA suggests that these molecules are readthrough transcripts. The failure of probe B to detect this RNA might simply reflect differences in the specific activity or concentration of the probe used and the low abundance of these molecules. The fact that a single readthrough transcript is detected in $\underline{sn}^{(+)}$ adult female RNA is discussed in Section 5.8.

The hybridisation pattern of wildtype sized RNAs in $\underline{sn}^{(+)}$ are identical to those detected in $\underline{sn}^{e}$ RNA. This suggests that the splicing mechanism to remove P-element sequences discussed for $\underline{sn}^{e}$ also operates for $\underline{sn}^{(+)}$. The $\underline{sn}^{(+)}$ P-element separates the genomic splices sites D1 and A1 by 0.95kb (Figure 5.4.2). Increasing the spacing between D1 and A1 from 25bp to 0.95kb may allow splicing to remove the intervening RNA to give an approximately wildtype sized RNA. To check that there were no obvious alternative splice sites the $\underline{sn}^{(+)}$ sequence was analysed for potential splice donor and acceptor sites. The donor site within the 31bp P-element terminal repeat (D2) is again a possible alternative to D1 (Figure 5.5.2). There are a few strong acceptor sites within the $\underline{sn}^{(+)}$ P-element but the only site capable of producing an approximately wildtype sized RNA is $A2^{(+)}$ (Figure 5.5.2). As $\underline{sn}^{e}$, $\underline{sn}^{(+)}$ and $\underline{sn}^{W}$ all produce wildtype sized transcripts but have different inserts, it is likely that the novel splice(s) occurs between either D1 /A1 and/or D2/A1 (see above).

The bristle phenotype of $\underline{sn}^{(+)}$ individuals can now be explained in terms of RNA molecules. No prematurely terminated transcripts can be detected in $\underline{sn}^{(+)}$ RNA suggesting that an approximately wildtype number RNAs transcribe the $\underline{sn}^{(+)}$ protein coding interval. Most of these transcripts are spliced to produce a wildtype sized $\underline{sn}^{(+)}$ RNA. The deduced structure of these molecules suggests that they can be translated. With little or no reduction of $\underline{sn}$ product, $\underline{sn}^{(+)}$ individuals do not develop a bristle phenotype.

5.6 RNA blotting analysis of sn^W

The DNA lesion in sn^W has been characterised as two P-elements inserted in a head-to-head orientation within the first exon of sn (Roiha et al., 1988, Nitasaka and Yamazaki,1988). The structure of sn^W and relationship between sn^W, sn^e and $sn^{(+)}$ has been described above. RNA blotting was used to describe the structure of sn RNAs in sn^W M cytotype individuals.

RNA was extracted from sn^W pupae, adult males and adult females. RNA blots carrying these samples were hybridised to 3 probes Figures 5.2.1-3. A summary diagram of this data is shown in Figure 5.6.1.

Probe A (607 to 1231 Figure 3.5.2A) lies upstream of the wildtype transcript initiation start site and detects no transcripts in sn^W RNA. Probes B (1346 to 1429 Figure 3.5.2A) and C (1444 to 1605 Figure 3.5.2A) lie within the first exon of sn and detect several transcripts in sn^W RNA. This pattern of hybridisation suggests that sn^W transcripts are initiated by the wildtype promoter and start site within probe B.

The sn^W mutation is closely related to the sn^e and $sn^{(+)}$ insertions. Thus the experience of analysing sn^e and $sn^{(+)}$ transcripts can be used to classify sn^W RNAs into 3 groups: readthrough, wildtype sized and prematurely terminated transcripts.

Probes B and C detect a 5.7kb RNA in sn^W pupal and adult male RNA. The size and hybridisation characteristics of these molecules suggest that they are readthrough transcripts. The 2.25kb of sequence inserted in sn^W increases the size of transcripts terminated at +15.0kb from 3.6kb to 5.7kb, provided the wildtype pattern of sn intron-exon junctions are used. The absence of readthrough transcripts in sn^W adult female RNA is discussed in Section 5.8.

A 1.3kb RNA is detected in pupal and adult male RNA by probe B but not probe C. The size and fact that only the upstream first exon probe detects this RNA suggests that the transcript terminates within the sn^W P-

Figure 5.6.1 - Summary diagram of $\underline{sn}^{W}$ RNA blotting data.

The $\underline{sn}^{W}$ RNA blotting data shown in Figures 5.2.1-3 has been collated and is summarised in 3 diagrams representing RNA blots. The intensity of transcript bands is represented by the thickness of boxes in this figure. Physical maps of the wildtype $\underline{sn}$ gene from -2.0 to 0.0, the $\underline{sn}^{W}$ insertion and the location and size of the 3 probes used in these investigations is shown below the diagrammatic RNA blots. P = pupal RNA, ♂ = adult male RNA, ♀ = adult female RNA.

Figure 5.6.1

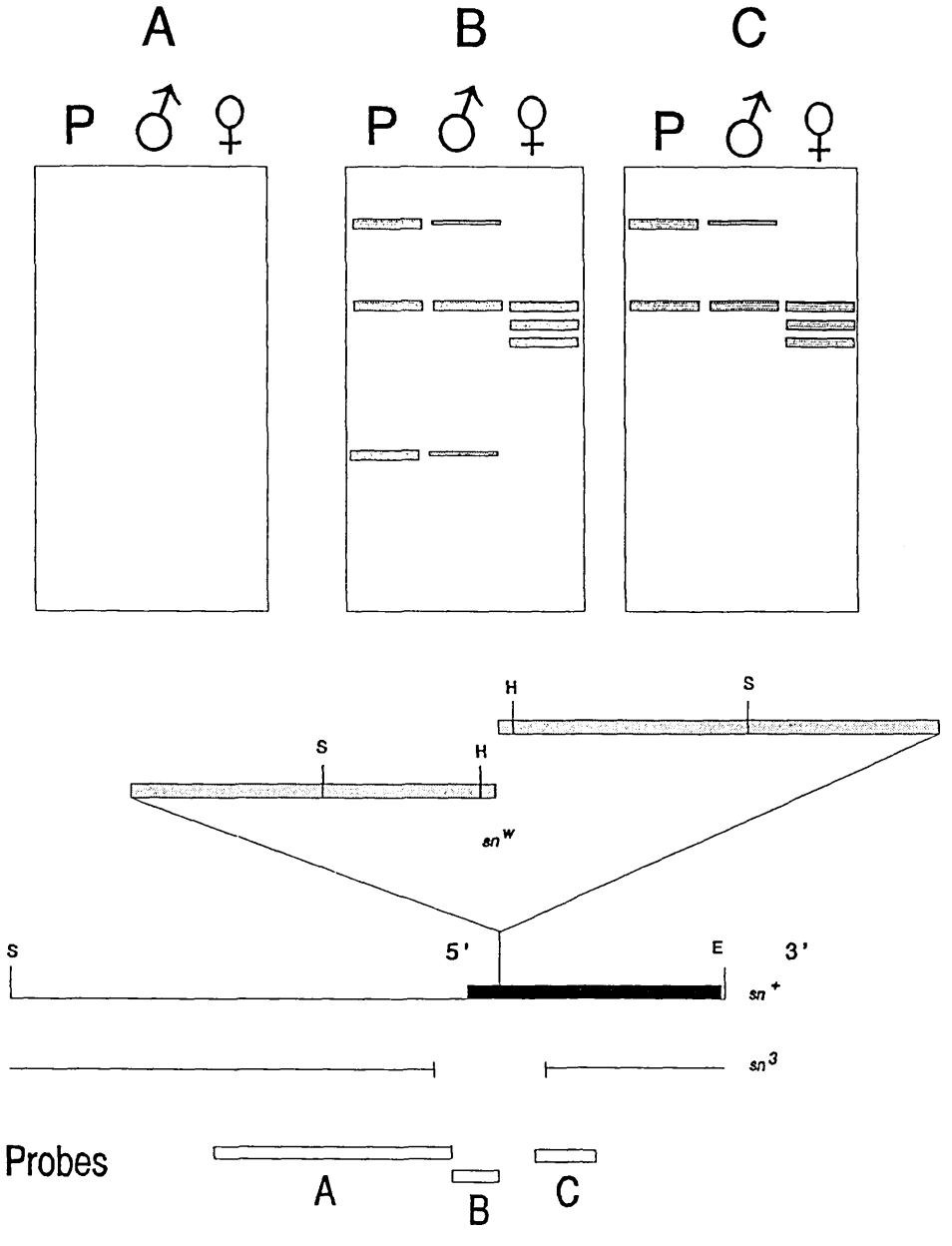
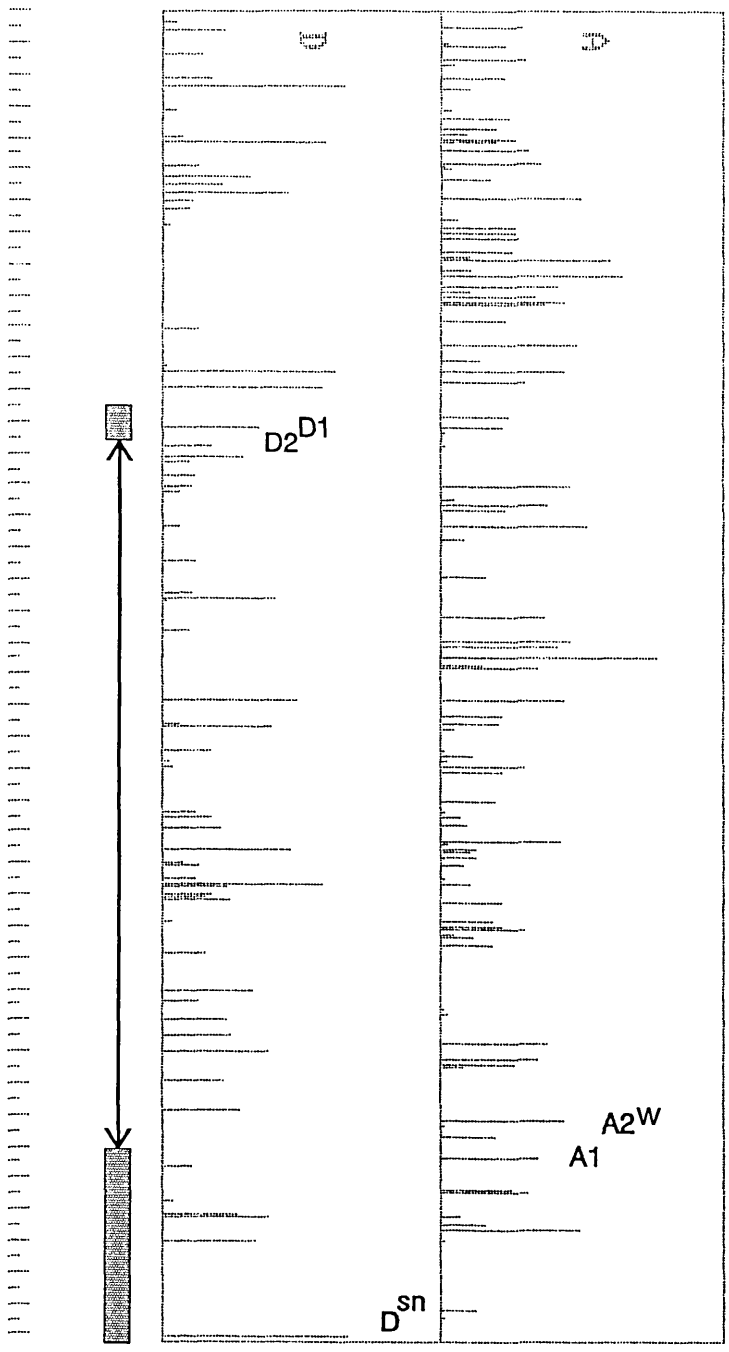


Figure 5.6.2 - $\underline{sn}^{W}$ splice donor and acceptor site analysis.

The $\underline{sn}^{W}$ sequence equivalent to the wildtype region from -2.0 to 0.0 but including the $\underline{sn}^{W}$ insertion (Figure A2) was analysed for consensus splice donor and acceptor sites using Amersham Staden Plus software. The graphics output of this program has been annotated to mark the position of the first exon (shaded box) and position and size of the $\underline{sn}^{W}$ P-elements (double headed arrow). The $\underline{sn}^{W}$ insertion separates $\underline{sn}$ donor and acceptors sites D1, A1 (Figure 5.4.3) by 2.25kb thereby facilitating the removal of intervening P-element sequences. Two other suitable placed donor and acceptor sites are D2 and A2. The use of these sites in a splicing reaction would produce an approximately wildtype sized RNA. Because wildtype sized RNAs are detected in all 3 P-element alleles the novel splice is thought to occur between D1 and A1 (see text for details). D^{Sn} is the donor site used by wildtype transcripts to end the first exon. The scale bar is graduated at 50bp intervals.

Figure 5.6.2



elements. The arrangement of P-element sequences in $\underline{sn}^W$ is such that the first P-element is transcribed in the opposite direction to sn and the second P-element is transcribed in the same direction as sn. Splicing and polyadenylation signals encoded by the second P-element are potentially active in $\underline{sn}^W$ transcripts. If the second P-element polyadenylation signal terminated $\underline{sn}^W$ readthrough transcripts then an RNA of approximately 2.3kb would be produced. The prematurely terminated $\underline{sn}^W$ RNA is smaller than this (1.3kb vs 2.3kb) suggesting that a more complex mechanism is used to generate this transcript.

There are two ways that the 1.3kb $\underline{sn}^W$ RNA could be produced. The first mechanism takes into account the fact that $\underline{sn}^W$ transcripts which include both P-element sequences have the potential to form a large (~0.9kb) stem loop structure. RNAs which initiate at the wildtype transcription start site and transcribe the $\underline{sn}^W$ gene as far as the end of the first P-element would be 1.1kb in length. As transcription continues into the second P-element a stem loop structure would start to form behind the RNA polymerase II complex. Very little is known about eukaryotic terminators but prokaryotic RNA polymerase terminators use stem loop structures to stall transcription and induce termination (reviewed by Adhyan and Gottesman,1978). It is possible that the formation of an approximately 200bp stem loop in $\underline{sn}^W$ RNA results in the premature termination of sn transcription. This would produce an RNA of approximately 1.3kb. As an alternative to stem loop terminators the 1.3kb RNA may be generated by a combination of polyadenylation at the second P-element polyadenylation site followed by splicing. RNAs which terminate at the P-element polyadenylation site would be approximately 2.3kb in length. Splice sites within this primary transcript could then be used to generate a 1.3kb RNA. The only way to distinguish between these two mechanisms would be to study $\underline{sn}^W$ cDNAs.

Which ever method is used to produce the 1.3kb RNA the process is inefficient. Readthrough and wildtype sized RNAs detected by probes B and C have failed to terminate

within the $\underline{sn}^{W}$ insertion.

Since the $\underline{sn}^{W}$ P-elements are located at the same site as the $\underline{sn}^{e}$ and $\underline{sn}^{(+)}$ insertions, the novel splicing events (described in Sections 5.4 and 5.5) may be used to produce the $\underline{sn}^{W}$ 3.0, 3.3 and 3.6kb RNAs (Figure 5.6.1). In addition to D1/A1, $\underline{sn}^{W}$ readthrough RNA has several splice donor/acceptor sites located within the P-element sequences (eg. D2 and $A2^{W}$). However P-element insertions at position 1359 appear to be spliced out regardless of sequence composition. This suggests that splice site located in sequences common to all 3 insertions are being used to excise the P-element RNA. This means that only D1/A1 or D2/A1 may be used in the novel splice.

The bristle phenotype of $\underline{sn}^{W}$ can be explained by examining the quantity of wildtype sized transcripts found in $\underline{sn}^{W}$ RNA compared to $\underline{sn}^{e}$ and $\underline{sn}^{(+)}$ RNA. Equal quantities of RNA were loaded in each lane of panel B Figure 5.2.1 it appears that $\underline{sn}^{W}$ has an intermediate number of wildtype sized RNAs in comparison to $\underline{sn}^{e}$ and $\underline{sn}^{(+)}$. This reflects the fact that fewer transcripts are terminated in $\underline{sn}^{W}$ than in $\underline{sn}^{e}$ and no termination transcripts are found in $\underline{sn}^{(+)}$. The slight reduction in wildtype sized $\underline{sn}^{W}$ RNAs results in a weak bristle phenotype.

5.7 Effect of cytotype on P-element alleles of sn

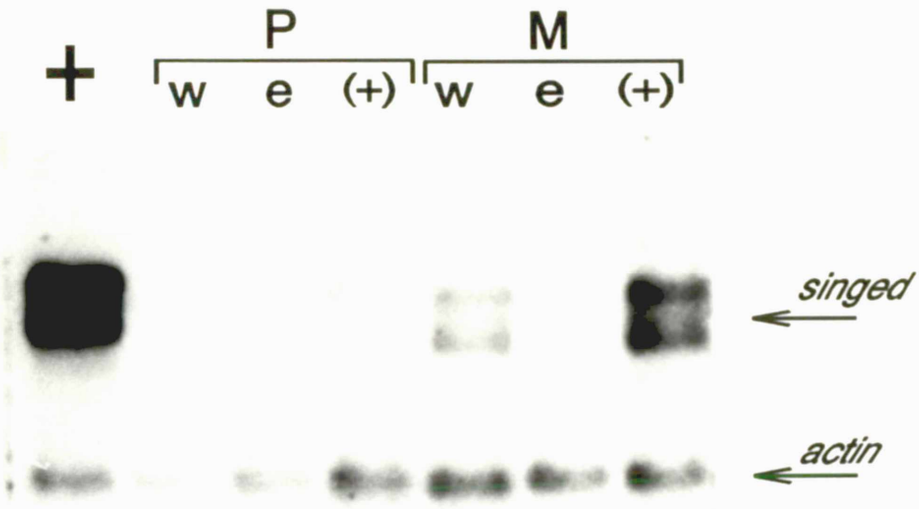
The fertility of $\underline{sn}^{W}$, $\underline{sn}^{e}$ and $\underline{sn}^{(+)}$ females is affected by P-M cytotype. In the M cytotype each allele is female fertile but in the P cytotype each allele is female sterile. Cytotype has no effect on the bristle morphology of these alleles. The modification of sn phenotypes by P-M cytotype appears to be restricted to only a few P-element alleles of sn (K.O'Hare pers. comm and H.M.Robertson pers. comm.).

To investigate the cytotype dependence of female sterility, RNA was prepared from homozygous adult female $\underline{sn}^{W}$, $\underline{sn}^{e}$ and $\underline{sn}^{(+)}$ individuals in either a P cytotype or M cytotype genetic background. Samples of these RNAs were blotted and probed with a region of the first exon (-0.35

Figure 5.7.1 - Cytotype dependent transcription of $\underline{sn}^{w}$, $\underline{sn}^{e}$ and $\underline{sn}^{(+)}$.

Poly-A^{+} RNA was prepared from homozygous $\underline{sn}^{w}$, $\underline{sn}^{e}$ and $\underline{sn}^{(+)}$ adult females in either the M or P cytotype. An RNA blot carrying samples of each RNA was probed with the region -0.35 to -0.50 (Probe C Figures 5.2.1-3). In the 1 week exposure shown here only the most abundant wildtype sized RNAs are detected. In the presence of P cytotype these RNAs are not detected. The loading of RNA in each lane was controlled by hybridisation to an actin probe. P = P cytotype, M = M cytotype, + = wildtype (M cytotype) w = $\underline{sn}^{w}$ e = $\underline{sn}^{e}$ (+) = $\underline{sn}^{(+)}$. Note if balancer chromosome breakdown occurs the $\underline{sn}^{(+)}$ flies will be indistinguishable from $\underline{sn}^{+}$ individuals. This appears to account for the $\underline{sn}$ transcripts detected in $\underline{sn}^{(+)}$ P cytotype RNA.

Figure 5.7.1



to -0.5, Probe C in Sections 5.2-6). The result of this experiment is shown in Figure 5.7.1.

In the P cytotype there are no sn transcripts detected in sn^{W}, sn^{e} and $sn^{(+)}$ adult female RNA (Figure 5.7.1). This result is consistent with the female sterile phenotype of these alleles in the P cytotype. The fact that all 3 wildtype sized transcripts in sn^{W}, sn^{e} and $sn^{(+)}$ RNA are abolished by P cytotype suggests that P cytotype represses transcription of these alleles in both somatic and germline tissue. Repression in the germline alone would leave a 3.6kb somatic transcript in each lane of P cytotype RNA (cf. Figure 4.2.1). If P cytotype represses somatic transcription of these alleles during metamorphosis, it is without affect on the bristle phenotype.

5.8 Concluding remarks

The lesion detected in sn^{X2} by DNA blotting (-0.9 to -0.7) has been identified as a non-mutagenic polymorphism. The mutagenic lesion in sn^{X2} cannot be detected by DNA blotting and thus must be either a small insertion/deletion (<50bp) or a point mutation. RNA blotting experiments show that transcription of the sn^{X2} gene results in 2 high molecular weight RNAs (Figure 5.1.3). This suggests that the mutagenic change in sn^{X2} alters the pattern of splice sites used to process sn RNAs. If intron sequences remain in sn^{X2} RNAs then this might account for the increase in transcript sizes. This hypothesis could be tested by using intron sequences as probes on blots carrying sn^{X2} RNA.

The sn^{3} deletion removes the wildtype sn transcript start site. Transcription of the sn^{3} gene is initiated by a cryptic promoter located upstream of the wildtype sn promoter. The cryptic promoter produces sn^{3} transcripts at all stages of development (Figure 5.3.2).

Homozygous sn^{3} flies develop with an extreme bristle phenotype even though sn^{3} RNAs are detected during metamorphosis. There are 3 ways of explaining this

observation. Firstly it is possible that the $\underline{sn}^3$ RNA detected in pupae cannot be translated to produce active sn protein. The absence or lack of sn protein during metamorphosis would then result in a bristle phenotype. The fertility of $\underline{sn}^3$ flies could then be rescued by the adult female specific 4.0 and 3.6kb RNAs.

The second way of explaining these observations is to suggest that the $\underline{sn}^3$ transcripts detected in pupae are not derived from hair and bristle forming tissue. If the cryptic promoter was active in the wrong tissue then sn protein would not be available to form wildtype hairs and bristles. The $\underline{sn}^3$ cryptic promoter has already been shown to produce sn RNA during larval stages of development when the $\underline{sn}^+$ promoter is inactive.

Finally the phenotype of $\underline{sn}^3$ may be explained in terms of threshold values for sn function. If all $\underline{sn}^3$ RNAs produce active sn protein then the bristle phenotype of $\underline{sn}^3$ may be due to a critical shortage of sn protein during metamorphosis. The cryptic promoter may fail to produce sufficient RNA for sn protein levels to break through the threshold level required for normal bristle morphology. If ovarian tissue has a lower requirement for sn protein than pupal tissue, then the cryptic promoter may supply sufficient sn RNA to satisfy the ovarian demand for sn function.

Transcriptional analysis of $\underline{sn}^w$, $\underline{sn}^e$ and $\underline{sn}^{(+)}$ identified 3 sorts of mutant RNA molecules. These are readthrough transcripts, wildtype sized RNAs and prematurely terminated transcripts. The readthrough transcripts appear to be precursors of the wildtype sized RNAs.

The wildtype sized RNAs detected in each P-element allele are produced by a novel splicing event which removes the inserted P-element sequences from the mutant transcript. This splicing event appears to be independent of the size, orientation or sequence composition of element inserted at this position (1359 Figure 3.5.2A). This suggests that splice sites within the sn sequence are

used to excise the inserted RNA. Two splice sites D1 and A1 (Figure 5.4.2) flank the site of P-element insertions. In sn wildtype RNA D1 and A1 are only 25bp apart and thus are too close for splicing enzymes to excise the RNA between them. Insertions into the 25bp gap between D1 and A1 increase the separation of the two sites and facilitate removal of intervening sequences either by a D1/A1 splice or from a splice site within the 31bp terminal repeats of the P-element (D2;Figures 5.4.3, 5.5.2, 5.6.2) to A1. This splice would remove the only ATG upstream of the translation start site. The absence of the first exon ATG may result in more efficient translation of the sn ORF from these transcripts.

The individual structures of readthrough transcripts have been described in Sections 5.4, 5.5, 5.6. However one result which has yet to be discussed is the presence of a single readthrough transcript in $sn^{(+)}$ adult female RNA but absence in sn^{W} and sn^{e} adult female RNA.

Alternate polyadenylation in adult females would be expected to produce 3 sizes of readthrough transcript. Analysis of the wildtype sized RNAs present in sn^{W}, sn^{e} and $sn^{(+)}$ has shown that alternate polyadenylation occurs within each of these mutants. The fact that only one size of readthrough transcript is detected in adult female RNA can be explained in terms of D1/A1 and/or D2/A1 splicing efficiencies. If the novel event, which removes P-element sequences, occurs more efficiently in female germ cells than in somatic tissues then all readthrough transcripts would be spliced down to wildtype sized RNAs. Somatic tissue would then be the source of a single adult female readthrough RNA equivalent to the 3.6kb transcript (ie 4.5kb in $sn^{(+)}$ adult female RNA). That no equivalent transcript is seen for sn^{e} and sn^{W} may be due to there being less readthrough RNA made than in $sn^{(+)}$ because of termination of transcription within the P-element inserts.

The two P-element alleles which produce bristle phenotypes, $\underline{sn}^{e}$ and $\underline{sn}^{w}$, both produce short prematurely terminated $\underline{sn}$ transcripts. There is a direct correlation between the abundance of these transcripts and the severity of bristle phenotype seen in these alleles. Pupae from homozygous $\underline{sn}^{e}$ individuals contain a large number of truncated $\underline{sn}$ transcripts and develop an extreme bristle phenotype. Pupae from homozygous $\underline{sn}^{w}$ individuals containing a few truncated $\underline{sn}$ transcripts and develop a weak bristle phenotype. Pupae from $\underline{sn}^{(+)}$ individuals do not contain truncated $\underline{sn}$ RNAs.

From the overall abundance of RNA, it seems that the insertions in $\underline{sn}^{w}$, $\underline{sn}^{e}$ and $\underline{sn}^{(+)}$ do not affect the rate of initiation of transcription. Thus the presence of many termination transcripts in $\underline{sn}^{e}$ significantly reduces the abundance of wildtype sized $\underline{sn}^{e}$ RNAs. In pupae this reduction in transcripts capable of producing $\underline{sn}$ protein results in an extreme bristle phenotype. There appears to be sufficient $\underline{sn}^{e}$ transcripts in adult females to prevent sterility.

Similarly the quantity of terminated transcripts in $\underline{sn}^{w}$ is small thus the number of wildtype sized RNAs is only slightly reduced in comparison to $\underline{sn}^{+}$. In pupae this small reduction results in the kinked bristle phenotype of a $\underline{sn}^{w}$ fly.

Note it is not possible to state with certainty that $\underline{sn}$ protein is produced by translation of readthrough RNAs. However, the proposed structure of the wildtype sized RNAs is very similar to $\underline{sn}^{+}$ RNAs. It seems likely, therefore, that they would be translated to give $\underline{sn}^{+}$ protein.

The bristle phenotypes of $\underline{sn}^{w}$, $\underline{sn}^{e}$ and $\underline{sn}^{(+)}$ have been explained in terms of the abundance of RNAs which might encode $\underline{sn}$ protein. The presence of a bristle phenotype but absence of a sterility phenotype in $\underline{sn}^{w}$ and $\underline{sn}^{e}$ individuals must now be explained.

Flies with reduced fertility compared to wildtype may be sufficiently fertile for homozygous stocks to be maintained. Thus although $\underline{sn}^{w}$, $\underline{sn}^{e}$ and $\underline{sn}^{(+)}$ are

classified as fertile minor disruptions to oogenesis may still occur within these individuals. The abundance of adult female wildtype sized transcripts in these mutants (Figure 5.7.1; M cytotype) suggests that $\underline{sn}^{(+)}$ individuals have near wildtype oogenesis, $\underline{sn}^{w}$ flies have slightly abnormal oogenesis and $\underline{sn}^{e}$ flies have more a serious disruption to oogenesis but are still classified as fertile. To give a full explanation of the $\underline{sn}^{w}$, $\underline{sn}^{e}$ and $\underline{sn}^{(+)}$ phenotypes a detailed description of oogenesis in each of these mutants would be required. However, a simplistic explanation based on the assumption that $\underline{sn}^{w}$, $\underline{sn}^{e}$ and $\underline{sn}^{(+)}$ flies are fully fertile uses the concept of threshold levels to explain the absence of sterility but presence of bristle phenotypes in these alleles. The highest levels of $\underline{sn}^{+}$ transcripts are seen in pupae. This suggests that metamorphosis has the largest demand for sn product. If wildtype transcription does not produce vast excess of sn function in pupae then slight reductions in transcripts levels will result in bristle abnormalities. Ovarian tissue may be less sensitive to reductions in sn transcript levels thus bristle phenotypes could develop in the absence of oogenic phenotypes.

The abundance of sn transcripts produced in $\underline{sn}^{w}$, $\underline{sn}^{e}$ and $\underline{sn}^{(+)}$ individuals depends on cytotype. In P cytotype female very few sn RNAs are detected (Figure 5.7.1). The lack of sn function in these individuals results in female sterility. Although the oogenic phenotype of these sn alleles is affected by cytotype the bristle phenotypes of $\underline{sn}^{w}$, $\underline{sn}^{e}$ and $\underline{sn}^{(+)}$ are not affected by cytotype. To fully understand the mechanism behind this genetic phenomenon a more detailed description of P cytotype is required.

CHAPTER 6

General Discussion

6.0 Introduction

The experiments described in this thesis were designed to address three fundamental questions about the sn gene.

1. What is the role of sn in germline and somatic cells?
2. How do mutant alleles disrupt sn gene function?
3. How does cytotype effect sn gene expression?

This discussion aims to answer these questions by taking an overall view of the gene structure, gene expression and phenotypic descriptions reported in this thesis.

6.1 What is the role of sn in germline and somatic cells ?

DNA from the sn locus had already been cloned (Roiha et al., 1988). This thesis describes the localisation, structural characterisation and partial nucleotide sequence of the sn gene. These investigations show the sn gene to produce 3 developmentally regulated RNAs which all encode the same 57kdal polypeptide.

Amino acid composition and secondary structure predictions suggest that the sn polypeptide is an intracellular protein with globular tertiary structure. This analysis is consistent with previous genetic studies which classified sn as a cell autonomous gene function.

To investigate the possibility that sn belonged to a class of previously identified proteins, the sn sequence was compared to databases containing the sequence of previously characterised genes and proteins. No highly homologous entries were found by this search.

To understand why sequence comparison studies failed the composition of databases and possible function(s) of sn protein should be considered.

The phenotype of sn^{-} individuals suggests that sn protein is required during metamorphosis and oogenesis. The sn gene product is not essential for viability but is required for species propagation (female sterility). Thus

although the sn gene is important for species survival it is not essential for cell survival. This opens up the possibility that sn function is only required in insects and thus the gene function may not be conserved throughout the animal kingdom. If the sn gene provides a function peculiar to insects then comparing the sequence of sn to databases composed mainly of bacterial and mammalian sequences is unlikely to find entries with overall homology.

In addition to searching for general homology computer programs can identify short regions of homology between two sequences. Short sequence motifs responsible for specific molecular interactions are often conserved between proteins of similar function (eg. ATP binding site). The sn sequence contains no conserved motifs. This suggests that the sn protein does not use the various molecular interactions characterised by these motifs.

The identification of cells specifically expressing the sn gene has helped to assign possible functions to the sn gene product. This analysis is most productive when the pattern of gene expression identifies cells with previously characterised functions. In these cases gene function and cell function may then be correlated.

The approximate timing and location of sn gene expression was determined by RNA blotting studies. Cells specifically expressing sn were identified by β-galactosidase staining in $sn^{P[lacZ]}/+$ individuals.

During metamorphosis the sn gene is expressed in the majority of epidermal cells. Although some of these cells are involved in producing hairs and bristles the majority of epidermal cells do not contribute to these specialised structures. This implies that sn has a more general function than simply making hairs and bristles.

Oogenic expression of sn occurs in both somatic and germline cells. Somatic cells which express sn during oogenesis are the border cells and 3 groups of follicle cells which surround the late stage oocyte. Germline expression of sn occurs in the nurse cells during mid and

late stage oogenesis.

Before starting to search for biological functions which may link these groups of cells it is worth considering how many functions the sn gene may have. Genetic data available at the start of these studies was consistent with the sn gene producing different products in somatic and germ cells (Bender,1960). The work presented in this thesis has shown that the sn gene produces 3 developmentally regulated RNAs which all encode the same 57kdal. polypeptide. Although it is possible that sn protein has different functions in somatic and germline tissue it is simpler to assume the protein has a single function which it executes in both types of tissue.

There appears to be only two processes which link all sites of sn gene expression. Each cell which expresses sn has undergone several cycles of DNA replication (without division) to become polyploid. Thus sn may be involved in the process of endoreplication. Unfortunately very little is known about the mechanistic process of endoreplication. However cell polyploidy is particularly important within Drosophila development, as the majority of larval and adult tissues contain polyploid cells. Of the many polyploid cells within an adult fly only a few display a sn^- phenotype. This observation suggests that sn protein does not have a role in endoreplication.

In addition to being polyploid all cells expressing sn are involved in the mass production of certain protein and lipid molecules. After synthesis these molecules are unidirectionally exported form the production cell. The sn protein may then have a role in these over production and export processes.

Many cells within the adult fly are required to produce large quantities of specific protein and/or lipid molecules but only a few express sn. This implies that sn protein does not play a major role in over production processes.

To identify the true role of sn protein in oogenesis and metamorphosis further studies on the subcellular localisation of sn protein may be required.

6.2 How do mutant alleles disrupt sn gene ?

Geneticists and biologists have always studied mutant alleles of a gene in order to gain insights into the function(s) performed by the wildtype gene. This approach was used to identify potential functions for the sn gene. Several mutant alleles of sn were studied at the molecular level but these investigations failed to provide additional information about the wildtype function of sn. However these experiments did characterise several interesting transcription and splicing events which help to explain the phenotypes of these alleles.

The lack of results relevant to the wildtype sn gene may be due to the fact that each allele chosen for study contained a defect outwith the protein coding interval and thus these mutations tended to alter the regulation of sn rather than the function of sn protein.

The answer to how mutations disrupt the sn gene was easily obtained by studying the molecular structure of DNA lesions in each allele and the resulting effects on mutant sn RNAs. These studies characterised a cryptic promoter in sn^3 and a novel splicing event in sn^w, sn^e and $sn^{(+)}$. They also demonstrated the sensitivity of metamorphosis to reductions in sn function. Slight reductions in sn transcript levels results in a bristle phenotype while oogenesis appears to be unaffected by small minor changes in sn protein levels.

6.3 How does cytotype effect sn gene expression?

This question arose out of the observation that sn^{W}, sn^{e} and $sn^{(+)}$ were fertile in M cytotype and female sterile in P cytotype. RNA blotting experiments show that P cytotype appears to suppress sn expression from these alleles in adult females resulting in sterility. P cytotype has no effect on the bristle phenotype of these alleles. Cytotype dependent sterility appears to be restricted to a few P-element alleles of the sn gene (K.O'Hare and H.M.Robertson pers. comm.). This phenomenon cannot be explained at a molecular level because the RNA and protein molecules responsible for P cytotype have yet to be defined and described.

Further experiments to define cytotype and investigate cytotype dependent transcription may utilise P[lacZ] insertions at sn. Initially these insertions will be tested for cytotype dependent expression of β-galactosidase. Having identified cytotype dependent P[lacZ] alleles these insertions may be used to quantify the suppression effect, define P-elements responsible for cytotype and examine the tissue specificity of cytotype dependent suppression.

6.4 Approaches to further work

The characterisation of the sn gene described in this thesis provides a starting point for further studies in a number of different areas.

Although the structure of the sn gene has been described it would be interesting to transform (Rubin and Spradling,1982) the region of DNA which encodes the sn transcription unit back into Drosophila. Having achieved full phenotypic rescue with a genomic fragment (eg. -2.0 to +15.5) a series of experiments could be conducted to identify regulatory elements. Upstream control sequences could be located by creating a series of 5′ deletions while potential internal regulatory elements within introns might be identified by constructing genomic::cDNA

chimeric genes which lack specific introns.

The success of germline transformation would also facilitate an analysis of the "hotspot" for P-element insertions located within the sn locus (Roiha et al., 1988). A high frequency of P-M dysgenesis induced mutations at the re-located sn gene would suggest that the "hotspot" was a result of sn sequence composition. Alternatively lack of P-element insertions into the transformed sn gene during P-M dysgenesis would suggest that the "hotspot" was due to the chromosomal location of the sn locus and not sequence composition of the sn gene.

In order to overcome the technical difficulties of localising low abundance sn transcripts an indirect method was used to locate cell specific sites of sn expression. The $sn^{P[lacZ]}$ results appear to correlate well with the phenotype of sn mutations but to directly localise sn gene products would confirm these results. Recent modifications to in situ hybridisation protocols (Tautz and Pfeiffe, 1989) may allow low abundance transcripts such as sn to be detected in whole mount tissue. Using optimal hydridisation and detection conditions sn RNA might be detected and thus confirm the $sn^{P[lacZ]}$ results.

As an alternative to detecting sn RNA molecules immunocytochemistry might be used to locate sn protein in tissue samples. Although previous efforts to produce polyclonal antisera which recognised sn protein were unsuccessful there is no obvious reason why this procedure should fail on a second attempt. In addition to producing rabbit polyclonal sera, mice should be injected with β-galactosidase::sn fusion protein to facilitate the production of monoclonal antibodies directed against the sn polypeptide. Either of these reagents would allow immunocytochemical localisation of sn protein in Drosophila tissues.

The analysis of transcription in mutant alleles of sn was originally conceived as a way of identifying methods of disrupting sn function. In fact the analysis appears to have characterised several mechanisms for circumventing mutagenic disruptions of the sn gene.

Investigation of the $\underline{sn}^3$ deletion showed that this mutation activated a cryptic promoter within upstream sequences of the sn gene. It appears that the novel promoter uses a single site of transcript initiation to start all $\underline{sn}^3$ RNAs. Normally primer extension or S1 fragment protection experiments would be used to locate this start site but the low abundance of sn transcripts makes these experiments technically difficult. An easier alternative might be to use polymerase chain reaction (Saiki et al.,1985) to amplify primer extension cDNA products (Frohmann et al.,1988). This type of analysis is particularly suitable for the study of low abundance transcripts and would define the precise nucleotide(s) used to initiate $\underline{sn}^3$ RNAs.

Analysis of RNAs from $\underline{sn}^w$, $\underline{sn}^e$ and $\underline{sn}^{(+)}$ suggest that a novel splicing event is removing P-element sequences from the majority of these transcripts. Although candidate donor and acceptor sites have been identified there is no physical evidence to suggest these sites are used in the novel splicing reaction. The best way to characterise this splicing event would be to use PCR to amplify mutant cDNAs made across the potential splice sites. Sequence analysis of cDNA molecules derived from $\underline{sn}^w$, $\underline{sn}^e$ and $\underline{sn}^{(+)}$ RNA would identify the precise position of the novel splice and hence identify the splice sites used.

APPENDIX

$\underline{sn}^{W}$, $\underline{sn}^{e}$ and $\underline{sn}^{(+)}$ sequence data.

The appendix contains 3 sequence files Figures A1-3 which represent the sequence of each P-element allele described in Chapter 5. Figure A1 displays the deduced sequence of $\underline{sn}^{W}$ from -2.0 to 0.0 (Figure 5.0.1), the sequence information in this file was analysed for splice donor and acceptor sites and the results shown in Figure 5.6.2. Figure A2 contains an assembled sequence representing the structure of $\underline{sn}^{e}$ from -2.0 to 0.0 (Figure 5.0.1). The splice donor/acceptor analysis of this sequence is shown in Figure 5.4.3. Figure A3 contains the dedeuced sequence of $\underline{sn}^{(+)}$ from -2.0 to 0.0 (Figure 5.0.1). Potential splice donor/acceptor sites were identified within this sequence using the Amersham Staden Plus set of programs. The result of this analysis is shown in Figure 5.5.2. In Figures A1-3 the 8bp target P-element duplication site is underline.

Figure A1


1
GTCGACTTTAGAACGGATATTTCAAAATAATGTTAACCTTTCGAAAAATTTAGGTGTATGCAAGCCGAATTATTTGATTT
101
GGAAAATAATACTTAGGGCGTCCATATCAGAAAAAACGTTTAGCGTAACATAACGTATTTTAAAATTTGTATTAGGCCAC
201
GTTATGTAGTCAAATATATATAAATAAATTATTCATTTTAATTTTTACAAAAGTGATATGCTGATATGCTTATAAATGTA
301
AGATTCTTGTTTATTTGACTTTTTTTTTAAATTTCATTATCATTATTTCTTGCTTGACAACCCTATGTTAAGTGCGACAC
401
TGCGTTGACAACAGCTAAGAACATGGCTGTGGGGCGACAAACTTAGCAGAGCGCGCGCTGCTAGAGCTTAAATTGTATTT
501
GAAAACAGTGCAGGTGAATCGAAAATGTCTTGGCATGCAGTCAAACATTAATTTGCGAGCATAATTAACTCCGCTTAAGA
601
TTGCTGGTAGTCAAGGAGCCGTGAAGAGCGGAGACGGCCGCTGGTAATATTCAGTGATTAACAACTGTAAAATCTAAAAT
701
CAGTATAAATATTCATGTGTGTTGATTTAAAGGATTTTGGTGAAATCCGGACCCGAATCGGAGCTGAGGTCCGAAACTAA
801
CGTGCAACATCCAATAAAAGCAAGTGCAGCTTGGGTTATTAGAGCATAAAGCATCTCCACCAGAGTCTCGGTTGCCAGCC
901
AAGCAAAACGGCCGAGACGCGTCTTTCGGGTGGCCAGCTGCAGTCTCGGATGCAGTTTTTGGCTGCAGCTGCAATTAAAT
1001
ATGCAATGCACTTAAGTTATGCACATTTCCTTTGCCCAGGGACCGTTACGATTTCGTATTTTAAAAGCCGCGCGAAAGCT
1101
AACAGCGCTGTTGGCTAGCAGACGTCTGTTTCTGCTGTTGTTAGCCAACAGCCAACAACAACAAACTGGCTGTTGTTTCG
1201
AATGCACTGCGATCTTAGCTCTTTTCCCGCTAAATGTTTCTTTGTGTGTCATTTTGTCAACCATTCGATCGATATCGATA
1301
TTCCTTTGTGAAGGCTAAAAATAAAATTCTTACATGCTCATCTTATACAATAATGGATCGCACATATAGCTACTCCTTTT
1401
TCATGTATCCAATTCACTAGGTAAGCTACAAAGATAGCCAGACGGCTGTGCTAACAGAGCGCTGCTATGGTGAGAAGCCG
1501
CCGCAATGGCAACAGTGCCATCTCTGGCCGGCGATACAAATAGCGCAACTTCGCTCGCAACGGGTTTCATTCCCACTGGA
1601
GTGCAGTTCGTGAGCGGTCGTTCTCTCCTCTCTCTCGCAAAAGTAAAACTTAAAGTTTTGTGCGCGCATGCGTCTCGTCC
1701
ATGATGAAATAACATAAGGTGGTCCCGTCGGCAAGAGACATCCACTTAACGTATGCTTGCAATAAGTGCGAGTGAAAGGA
1801
ATAGTATTCTGAGTGTCGTATTGAGTCTGAGTGAGACAGCGATATGATTGTTGATTAACCCTTAGCATGTCCGTGGGGTT
1901
TGAATTAACTCATAATATTAATTAGACGAAATTATTTTAAAGTTTTATTTTTAATAATTTGGAGTTTCCAATTAACTTT
2001
TGTTTTTGATTTTTAATTTCAATTTTTTTGTTGAAGTACTTAATTCTAAAATATATTCTAATTTTAAAATAAAAAGATTT
2101
TACTTGTTTATCAACATCGACGTTTCGCGCTGCTAATATTAATTTTTCCTTTACATTATTTGTAATTTCAAAATTATTAT
2201
TATTTGTATAATGCAAAAAAATACATTCTAGCTCTTTTAACTTCTCTTGAAATTTTTCGGACGGCTTAATAAGTCCGCCG
2301
TGAGACACCTCGTCGACGTATGTAAATGTCAAATTTTCTTTTACTTTGTCACTGATTCTCAATTTTTTAATGACATAGCC
2401
CGCGATATATTCCATCGCATCCTCCGTCAACTCATCCATTTCGGTGAAATCTATATTGCTTAACATCTCATCGACAGGCT
2501
CATCATCGACAGGCTCATCATCTTTACTTTTATTTTCGTTTTCTTTAGAACTGAAATCTAAATTAAGCCAGCTTTCAGAG
2601
TTGTCCTCTTCGATATTTCCCGAATTTCTTAACATTTCTGTATTCCTGGCTGGATAAGCAAAAGGATAATGATAATATTT
2701
CATTTTTTTTTATTCCACGTAAGGGTTAATGTTTTCAAAAAAAAATTCGTCCGCACACAACCTTTCCTCTCAACAAGCAA
2801
ACGTGCACTGAATTTAAGTGTATACTTCGGTAAGCTTCGGCTTTCGACGGGACCACCTTATGTTATTTCATCATGGTCTC
2901
GTCCATGATGAAATAACATAAGGTGGTCCCGTCGAAAGCCGAAGCTTACCGAAGTATACACTTAAATTCAGTGCACGTTT
3001
GCTTGTTGAGAGGAAAGGTTGTGTGCGGACGAATTTTTTTTTTGAAAACATTAACCCTTACGTGGAATAAAAAAAAATGAA
3101
ATATTGCAAATTTTGCTGCAAAGCTGTGACTGGAGTAAAATTAATTCACGTGCCGAAGTGTGCTATTAAGAGAAAATTGT
3201
GGGAGCAGAGCCTTGGGTGCAGCCTTGGTGAAAACTCCCAAATTTGTGATACCCACTTTAATGATTCGCAGTGGAAGGCT
3301
GCACCTGCAAAAGGTCAGACATTTAAAAGGAGGCGACTCAACGCAGATGCCGTACCTAGTAAAGTGATAGAGCCTGAACC
3401
AGAAAAGATAAAAGAAGGCTATACCAGTGGGAGTACACAAACAGAGTAAGTTTGAATAGTAAAAAAAATCATTTATGTAA
3501
ACAATAACGTGACTGTGCGTTAGGTCCTGTTCATTGTTTAATGAAAATAAGAGCTTGAGGGAAAAAATTCGTACTTTGGA
3601
GTACGAAATGCGTCGTTTAGAGCAGCAGCTGAGGGAGTCTCAACAGTTGGAGGAGTCTCTACGCAAAATCTTCACGGACA
3701
CGCAGATACGGATACTGAAGAATGGTGGACAAAGAGCTACGTTCAATTCCGACGACATTCACCGAAATGGATGAGTTGAC
3801
GGAGGATGCGATGGAATATATCGCGGGCTATGTCATTAAAAAATTGAGAATCAGTGACAAAGTAAAAGAAAATTTGACAT
3901
TTACATACGTCGACGAGGTGTCTCACGGCGGACTTATTAAGCCGTCCGAAAAATTTCAAGAGAAGTTAAAAGAGCTAGAA

Figure A1


4001
TGTATTTTTTTGCATTATACAAATAATAATAATTTTGAAATTACAAATAATGTAAAGGAAAAATTAATATTAGCAGCGCG
4101
AAACGTCGATGTTGATAAACAAGTAAAATCTTTTTATTTTAAAATTAGAATATATTTTAGAATTAAGTACTTCAACAAAA
4201
AAATTGAAATTAAAAATCAAAAACAAAAGTTAATTGGAAACTCCAAATTATTAAAAATAAAACTTTAAAAATAATTTCGT
4301
CTAATTAATATTATGAGTTAATTCAAACCCCACGGACATGCTAAGGGTTAATCAACAATCATATCGCTGTCTCACTCAGA
4401
CTCAATACGACACTCAGAATACTATTCCTTTCACTCGCACTTATTGCAAGCATACGTTAAGTGGATGTCTCTTGCCGACG
4501
GGACCACCTTATGTTATTTCATCATGGTCTCGTCTTAGAAAAACCTGCAATCAGTAATCTCGAATCGTGAAAATCCTTTT
4601
CTGCCATACCAAAAAAAGAGCCGGACAAAGCAAAGCGTTCGATTGCCTTCTGCCAATTTGAGCAGGAAATTACAAATTAA
4701
TTGCCGCCCTGACCGTGATTTGTTTACGTTTTTTTCTTCACTTTTTTTTACATTGTGTGAGTGTGAGAAATTTGTTGGCC
4801
AAAAAATATTCCAGATATTCCCAGAGATATATATATTCGAAATTCGAATTGGCGAAGTGAAGGGTGTGACGACAGACGCA
4901
CACACACGCACGCACACGAAAACGCACTGAAAAAACACCGGCTAAAAGGCAGACAAAAAGGAGACGAGAGAGCGAGTGCT
5001
AAATAGAAAACGCGAAAACAGGATAATAATCAGCAAACTAATACTATAAAACAAGCAACTAAGTGACATCTCGATCTCGC
5101
CGAGAAACAACAAAAACAACAGTTACAACTTCGGCAACAGCAACAAAAACAGCGCCAGCAGCAGCAACAACAACAAAATC
5201
AGCGAAGAGCATAGAGCCAAAAAGAAGGAGCAGCAAAAGCAAAGAAAGTAAGTGGAATTC

Figure A2

1
GTCGACTTTAGAACGGATATTTCAAAATAATGTTAACCTTTCGAAAAATTTAGGTGTATGCAAGCCGAATTATTTGATTT

101
GGAAAATAATACTTAGGGCGTCCATATCAGAAAAAACGTTTAGCGTAACATAACGTATTTTAAAATTTGTATTAGGCCAC

201
GTTATGTAGTCAAATATATATAAATAAATTATTCATTTTAATTTTTACAAAAGTGATATGCTGATATGCTTATAAATGTA

301
AGATTCTTGTTTATTTGACTTTTTTTTTAAATTTCATTATCATTATTTCTTGCTTGACAACCCTATGTTAAGTGCGACAC

401
TGCGTTGACAACAGCTAAGAACATGGCTGTGGGGCGACAAACTTAGCAGAGCGCGCGCTGCTAGAGCTTAAATTGTATTT

501
GAAAACAGTGCAGGTGAATCGAAAATGTCTTGGCATGCAGTCAAACATTAATTTGCGAGCATAATTAACTCCGCTTAAGA

601
TTGCTGGTAGTCAAGGAGCCGTGAAGAGCGGAGACGGCCGCTGGTAATATTCAGTGATTAACAACTGTAAAATCTAAAAT

701
CAGTATAAATATTCATGTGTGTTGATTTAAAGGATTTTGGTGAAATCCGGACCCGAATCGGAGCTGAGGTCCGAAACTAA

801
CGTGCAACATCCAATAAAAGCAAGTGCAGCTTGGGTTATTAGAGCATAAAGCATCTCCACCAGAGTCTCGGTTGCCAGCC

901
AAGCAAAACGGCCGAGACGCGTCTTTCGGGTGGCCAGCTGCAGTCTCGGATGCAGTTTTTGGCTGCAGCTGCAATTAAAT

1001
ATGCAATGCACTTAAGTTATGCACATTTCCTTTGCCCAGGGACCGTTACGATTTCGTATTTTAAAAGCCGCGCGAAAGCT

1101
AACAGCGCTGTTGGCTAGCAGACGTCTGTTTCTGCTGTTGTTAGCCAACAGCCAACAACAACAAACTGGCTGTTGTTTCG

1201
AATGCACTGCGATCTTAGCTCTTTTCCCGCTAAATGTTTCTTTGTGTGTCATTTTGTCAACCATTCGATCGATATCGATA

1301
TTCCTTTGTGAAGGCTAAAAATAAAATTCTTACATGCTCATCTTATACAATAATGGATCGCACATATAGCTACTCCTTTT

1401
TCATGTATCCAATTCACTAGGTAAGCTACAAAGATAGCCAGACGGCTGTGCTAACAGAGCGCTGCTATGGTGAGAAGCCG

1501
CCGCAATGGCAACAGTGCCATCTCTGGCCGGCGATACAAATAGCGCAACTTCGCTCGCAACGGGTTTCATTCCCACTGGA

1601
GTGCAGTTCGTGAGCGGTCGTTCTCTCCTCTCTCTCGCAAAAGTAAAACTTAAAGTTTTGTGCGCGCATGCGTCTCGTCC

1701
ATGATGAAATAACATAAGGTGGTCCCGTCGAAAGCCGAAGCTTACCGAAGTATACACTTAAATTCAGTGCACGTTTGCTT

1801
GTTGAGAGGAAAGGTTGTGTGCGGACGAATTTTTTTTTTGAAAACATTAACCCTTACGTGGAATAAAAAAAAATGAAATAT

1901
TGCAAATTTTGCTGCAAAGCTGTGACTGGAGTAAAATTAATTCACGTGCCGAAGTGTGCTATTAAGAGAAAATTGTGGGA

2001
GCAGAGCCTTGGGTGCAGCCTTGGTGAAAACTCCCAAATTTGTGATACCCACTTTAATGATTCGCAGTGGAAGGCTGCAC

2101
CTGCAAAAGGTCAGACATTTAAAAGGAGGCGACTCAACGCAGATGCCGTACCTAGTAAAGTGATAGAGCCTGAACCAGAA

2201
AAGATAAAAGAAGGCTATACCAGTGGGAGTACACAAACAGAGTAAGTTTGAATAGTAAAAAAAATCATTTATGTAAACAA

2301
TAACGTGACTGTGCGTTAGGTCCTGTTCATTGTTTAATGAAAATAAGAGCTTGAGGGAAAAAATTCGTACTTTGGAGTAC

2401
GAAATGCGTCGTTTAGAGCAGCAGCTGAGGGAGTCTCAACAGTTGGAGGAGTCTCTACGCAAAATCTTCACGGACACGCA

2501
GATACGGATACTGAAGAATGGTGGACAAAGAGCTACGTTCAATTCCGACGACATTCACCGAAATGGATGAGTTGACGGAG

2601
GATGCGATGGAATATATCGCGGGCTATGTCATTAAAAAATTGAGAATCAGTGACAAAGTAAAAGAAAATTTGACATTTAC

2701
ATACGTCGACGAGGTGTCTCACGGCGGACTTATTAAGCCGTCCGAAAAATTTCAAGAGAAGTTAAAAGAGCTAGAATGTA

2801
TTTTTTTGCATTATACAAATAATAATAATTTTGAAATTACAAATAATGTAAAGGAAAAATTAATATTAGCAGCGCGAAAC

2901
GTCGATGTTGATAAACAAGTAAAATCTTTTTATTTTAAAATTAGAATATATTTTAGAATTAAGTACTTCAACAAAAAAAT

3001
TGAAATTAAAAATCAAAAACAAAAGTTAATTGGAAACTCCAAATTATTAAAAATAAAACTTTAAAAATAATTTCGTCTAA

3101
TTAATATTATGAGTTAATTCAAACCCCACGGACATGCTAAGGGTTAATCAACAATCATATCGCTGTCTCACTCAGACTCA

3201
ATACGACACTCAGAATACTATTCCTTTCACTCGCACTTATTGCAAGCATACGTTAAGTGGATGTCTCTTGCCGACGGGAC

3301
CACCTTATGTTATTTCATCATGGTCTCGTCTTAGAAAAACCTGCAATCAGTAATCTCGAATCGTGAAAATCCTTTTCTGC

3401
CATACCAAAAAAAGAGCCGGACAAAGCAAAGCGTTCGATTGCCTTCTGCCAATTTGAGCAGGAAATTACAAATTAATTGC

3501
CGCCCTGACCGTGATTTGTTTACGTTTTTTTCTTCACTTTTTTTTACATTGTGTGAGTGTGAGAAATTTGTTGGCCAAAA

3601
AATATTCCAGATATTCCCAGAGATATATATATTCGAAATTCGAATTGGCGAAGTGAAGGGTGTGACGACAGACGCACACA

3701
CACGCACGCACACGAAAACGCACTGAAAAAACACCGGCTAAAAGGCAGACAAAAAGGAGACGAGAGAGCGAGTGCTAAAT

3801
AGAAAACGCGAAAACAGGATAATAATCAGCAAACTAATACTATAAAACAAGCAACTAAGTGACATCTCGATCTCGCCGAG

3901
AAACAACAAAAACAACAGTTACAACTTCGGCAACAGCAACAAAAACAGCGCCAGCAGCAGCAACAACAACAAAATCAGCG

4001
AAGAGCATAGAGCCAAAAAGAAGGAGCAGCAAAAGCAAAGAAAGTAAGTGGAATTC

Figure A3


1
GTCGACTTTAGAACGGATATTTCAAAATAATGTTAACCTTTCGAAAAATTTAGGTGTATGCAAGCCGAATTATTTGATTT
101
GGAAAATAATACTTAGGGCGTCCATATCAGAAAAAACGTTTAGCGTAACATAACGTATTTTAAAATTTGTATTAGGCCAC
201
GTTATGTAGTCAAATATATATAAATAAATTATTCATTTTAATTTTTACAAAAGTGATATGCTGATATGCTTATAAATGTA
301
AGATTCTTGTTTATTTGACTTTTTTTTTAAATTTCATTATCATTATTTCTTGCTTGACAACCCTATGTTAAGTGCGACAC
401
TGCGTTGACAACAGCTAAGAACATGGCTGTGGGGCGACAAACTTAGCAGAGCGCGCGCTGCTAGAGCTTAAATTGTATTT
501
GAAAACAGTGCAGGTGAATCGAAAATGTCTTGGCATGCAGTCAAACATTAATTTGCGAGCATAATTAACTCCGCTTAAGA
601
TTGCTGGTAGTCAAGGAGCCGTGAAGAGCGGAGACGGCCGCTGGTAATATTCAGTGATTAACAACTGTAAAATCTAAAAT
701
CAGTATAAATATTCATGTGTGTTGATTTAAAGGATTTTGGTGAAATCCGGACCCGAATCGGAGCTGAGGTCCGAAACTAA
801
CGTGCAACATCCAATAAAAGCAAGTGCAGCTTGGGTTATTAGAGCATAAAGCATCTCCACCAGAGTCTCGGTTGCCAGCC
901
AAGCAAAACGGCCGAGACGCGTCTTTCGGGTGGCCAGCTGCAGTCTCGGATGCAGTTTTTGGCTGCAGCTGCAATTAAAT
1001
ATGCAATGCACTTAAGTTATGCACATTTCCTTTGCCCAGGGACCGTTACGATTTCGTATTTTAAAAGCCGCGCGAAAGCT
1101
AACAGCGCTGTTGGCTAGCAGACGTCTGTTTCTGCTGTTGTTAGCCAACAGCCAACAACAACAAACTGGCTGTTGTTTCG
1201
AATGCACTGCGATCTTAGCTCTTTTCCCGCTAAATGTTTCTTTGTGTGTCATTTTGTCAACCATTCGATCGATATCGATA
1301
TTCCTTTGTGAAGGCTAAAAATAAAATTCTTACATGCTCATCTTATACAATAATGGATCGCACATATAGCTACTCCTTTT
1401
TCATGTATCCAATTCACTAGGTAAGCTACAAAGATAGCCAGACGGCTGTGCTAACAGAGCGCTGCTATGGTGAGAAGCCG
1501
CCGCAATGGCAACAGTGCCATCTCTGGCCGGCGATACAAATAGCGCAACTTCGCTCGCAACGGGTTTCATTCCCACTGGA
1601
GTGCAGTTCGTGAGCGGTCGTTCTCTCCTCTCTCTCGCAAAAGTAAAACTTAAAGTTTTGTGCGCGCATGCGTCTCGTCC
1701
ATGATGAAATAACATAAGGTGGTCCCGTCGGCAAGAGACATCCACTTAACGTATGCTTGCAATAAGTGCGAGTGAAAGGA
1801
ATAGTATTCTGAGTGTCGTATTGAGTCTGAGTGAGACAGCGATATGATTGTTGATTAACCCTTAGCATGTCCGTGGGGTT
1901
TGAATTAACTCATAATATTAATTAGACGAAATTATTTTTAAAGTTTTATTTTTAATAATTTGGAGTTTCCAATTAACTTT
2001
TGTTTTTGATTTTTAATTTCAATTTTTTTGTTGAAGTACTTAATTCTAAAATATATTCTAATTTTAAAATAAAAAGATTT
2101
TACTTGTTTATCAACATCGACGTTTCGCGCTGCTAATATTAATTTTTCCTTTACATTATTTGTAATTTCAAAATTATTAT
2201
TATTTGTATAATGCAAAAAAATACATTCTAGCTCTTTTAACTTCTCTTGAAATTTTTCGGACGGCTTAATAAGTCCGCCG
2301
TGAGACACCTCGTCGACGTATGTAAATGTCAAATTTTCTTTTACTTTGTCACTGATTCTCAATTTTTTAATGACATAGCC
2401
CGCGATATATTCCATCGCATCCTCCGTCAACTCATCCATTTCGGTGAAATCTATATTGCTTAACATCTCATCGACAGGCT
2501
CATCATCGACAGGCTCATCATCTTTACTTTTATTTTCGTTTTCTTTAGAACTGAAATCTAAATTAAGCCAGCTTTCAGAG
2601
TTGTCCTCTTCGATATTTCCCGAATTTCTTAACATTTCTGTATTCCTGGCTGGATAAGCAAAAGGATAATGATAATATTT
2701
CATTTTTTTTTATTCCACGTAAGGGTTAATGTTTTCAAAAAAAAATTCGTCCGCACACAACCTTTCCTCTCAACAAGCAA
2801
ACGTGCACTGAATTTAAGTGTATACTTCGGTAAGCTTCGGCTTTCGACGGGACCACCTTATGTTATTTCATCATGGTCTC
2901
GTCTTAGAAAAACCTGCAATCAGTAATCTCGAATCGTGAAAATCCTTTTCTGCCATACCAAAAAAAGAGCCGGACAAAGC
3001
AAAGCGTTCGATTGCCTTCTGCCAATTTGAGCAGGAAATTACAAATTAATTGCCGCCCTGACCGTGATTTGTTTACGTTT
3101
TTTTCTTCACTTTTTTTTACATTGTGTGAGTGTGAGAAATTTGTTGGCCAAAAAATATTCCAGATATTCCCAGAGATATA
3201
TATATTCGAAATTCGAATTGGCGAAGTGAAGGGTGTGACGACAGACGCACACACACGCACGCACACGAAAACGCACTGAA
3301
AAAACACCGGCTAAAAGGCAGACAAAAAGGAGACGAGAGAGCGAGTGCTAAATAGAAAACGCGAAAACAGGATAATAATC
3401
AGCAAACTAATACTATAAAACAAGCAACTAAGTGACATCTCGATCTCGCCGAGAAACAACAAAAACAACAGTTACAACTT
3501
CGGCAACAGCAACAAAAACAGCGCCAGCAGCAGCAACAACAACAAAATCAGCGAAGAGCATAGAGCCAAAAAGAAGGAGC
3601
AGCAAAAGCAAAGAAAGTAAGTGGAATTC

LITERATURE CITED

Adhyan,N. and S.Gottesman,(1978) Ann. Rev. Biochem. 47:967-996

Appleyard,R.C.,(1954) Genetics 39:400-452

Avery,O.T. C.M.M^{C}Leod and M.M^{C}Carthy,(1944) J. Exp. Med. 79:137-145

Bainbridge,P. and M.Bownes,(1981) J. Embryol. exp. Morph. 66:57-80

Bankier,A.T. and B.G.Barrell.(1983) In:Techniques in life sciences, Nucleic Acid Biochemistry. (Ed Favell,R.A) B5, Elsevier Scientific Publishers Ltd.,Ireland. pp1-34

Bender,H.A.(1960) Genetics 45:867-883

Benoist,C. K.O'Hare ,R.Breathnach and P.Chambon.(1980) Nuc. Acids Res. 8:127-142

Benton,W.D. and R.W.Davis.(1977) Science 196:180-192

Bingham,P.M., M.G.Kidwell and G.M.Rubin. (1982) Cell 29 997-1004

Birboim,H.C. and J.Doly.(1979) Nuc. Acids Res.7:1513-1523

Bodenstein,G.B. (1950) In:Bology of Drosophila. (Ed. Demerec,H.) Wiley, N.Y., U.S.A. pp 275-367

Breathnach,R. and P.Chambon. (1981) Ann. Rev. Biochem. 50:349-385

Burke,J.F. and D.Ish-Horowicz. (1981) Nuc. Acids. Res. 9:2989-2998

Burke,J.F. (1984) Gene 30:63-68

Busslinger,M. R.Portman, J.C. Irminger and M.L.Birnstiel. (1980) Nuc. Acids Res. 8:459-472

Cavener,D.R. (1987) Nuc. Acids Res. 15:1353-1361

Drosophila Stock list. (1982) Dros. Inf. Ser, 57

Engels,W.R. (1979) Proc.Natl.Acad.Sci.(USA) 76:4011-4015

Engels,W.R. (1983) Ann. Rev. Genet. 17:315-344

Feinberg,A.P. and B.Vogelstein. (1983) Ana. Biochem. 132:6-13

Feinberg,A.P. and B.Vogelstein. (1984) Ana. Biochem. 133:266-267

Frischauf,A-M. H.Lehrach, A.Poustka and N.Murray.(1983) J.Mol.Biol. 170 827-842

Frohman,M.A. M.K.Dush and G.R.Martin.(1988) Proc. Natl. Acad. Sci. (USA) 85:8998-9002

Fryberg,E.A. J.W. Mahaffey, B.J.Bond and N.Davidson.(1983) Cell 33:115-123

Gergen,J.P. R.H.Stern and P.C. Wensink. (1979) Nuc. Acids Res. 7:2115-2136

Geyer,P.K. and V.G.Corses. (1987) Genes and Dev. 1:996-1004

Ghosk,P.K. V.B. Reddy, J.Swinsioe, P.Lebovitz and S.M. Weissman. (1978) J.Mol.Biol. 126:813-822

Ghysen,A. and C.O'Kane. (1989) Development 105:35-52

Grossniklaus,U. H.J.Bellen, C.Wilson and W.J.Gehring. (1989) Development 107:189-200

Guo,L.H, R.C.A. Yang and R. Wu. (1983) Nuc. Acids Res. 11:5521-5539

Hexter,W.M. (1957) Genetics 42:376

Hohn,B. and K.Murray. (1977) Proc.Natl.Acad.Sci.(USA) 74:3259-3263

Hoyt,M.A., D.M.Knight, A.Das, H.I.Miller and H.Echols. (1982) Cell 31:565-573

Hung,M.C. and P.C. Wensink. (1981) Nuc. Acids. Res. 9:640-649

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

Kidwell,M.G. and J.B.Novy. (1979) Genetics 92:1127-1132

Kieng,M.P., R.Lathe, and J.P.Lecocq. (1983) Gene 26:91-99

King,R.C. (1960) Growth 24:265-323

King,R.C. (1970) Ovarian development in Drosophila melanogaster, Academic Press, N.Y., USA.

Lambda II. (1983) ; Appendix II: Complete annotated lambda sequence, (Eds. Hendrix,R.W., J.W.Roberts, F.W.Stahl, and R.A. Weisberg) Cold Spring Harbor Laboratory, N.Y., USA.

Laskey,R.A. and R.D.Mills. (1977) FEBS Letters 82:314-316

Laskey,F.A., D.C.Rio and G.M.Rubin. (1986) Cell 44:7-19

Lees,A.D. and C.H.Waddington,(1942) Proc. Roy.Soc.(London) Ser.B. 131:87-110

Levis,R., K.O'Hare and G.M.Rubin. (1984) Cell 38:471-481

Levis,R. and S.Penman. (1978) J.Mol.Biol. 120:487-515

Lindsley,D.L. and E.H.Gell. (1968) Carnegie Inst. Wash. Pub. 627

Mandel,M. and A.Higa. (1970) J.Mol.Biol. 53:159-162

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

Mariani,C., V.Pirrota and E.Manet. (1985) EMBO J. 42045-2052

Mendel,G.J. (1866) In:Proceedings of the Brunn Society for Natural History 1866

Messing,J.H., R.Crea and P.H.Seeburg. (1981) Nuc. Acids Res. 9:309-318

Messing,J. and J.Vieira. (1982) Gene 19:269-276

Miller,S. and R.C.Reznikoff. (1978) In: The Operon (Eds Miller,S. and R.C.Reznikoff) Cold Spring Harbor Laboratory N.Y.,USA.

Mohr,O.L. (1922) Z.Indukitive Abstannnungs Vererbungslehre 28:1-22

Mount,S.M. (1982) Nuc. Acids Res. 10:459-472

Nitasaka,E. and T.Yamazaki.(1988) The Japanese Journal of Genetics 63:303-312

O'Hare,K. and G.M.Rubin. (1983) Cell 34 25-35

O'Kane,C and W.J.Gehring. (1987) Proc. Natl. Acad. Sci.(USA) 84:9123-9127

Perrimon,N. and M.Gans. (1983) Dev.Bio. 100:365-373

Ponz,M., D.Solowiejezyk, M.Ballantine, E.Schwartz and S.Surrey. (1982) Proc.Natl.Acad.Sci.(USA) 79:4298-4302

Poodry,C.A. (1980) In: The genetics and biology of Drosophila Volume 2d (Eds.M.Asburnner and T.R.F.Wright) Academic Press, N.Y., USA.

Poole.S.J., M.K.Lawrence, B.Dress and T.Kornberg.(1985) Cell 40:37-42

Queen,C. and C.J.Korn. (1984) Nuc. Acids Res. 12:581-600

Rigby,P.W.J., M.Diekmann C.Rhodes and P.Berg. (1977) J.Mol.Biol. 113:237-251

Roiha,H., G.M.Rubin and K.O'Hare. (1988) Genetics 119:75-83

Roth,A. (1974) Ann.Rev.Genetics 8:319-346

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

Saiki, R.K., S.Scharf, F.Faloona, K.B.Mullis, G.T. Horn, H.A.Erlich and N.Arnheim. (1985) Science 230:1350-1355

Sanger,F., S.Nicklen and A.R.Coulson. (1977) Proc. Natl. Acad. Sci.(USA) 74:5463-5476

Shapiro, H. (1983) In: Mobile genetic elements (Ed. Shapiro, H.) Academic Press, N.Y.,USA.

Simon,J.A., C.A.Sutton, R.B.Lobell, R.L.Galser and J.T. Lis. (1985) Cell 40:805-817

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

Staden,R., (1982) Nuc. Acids Res. 10:2951-2961

Staden,R., (1984) Nuc. Acids Res. 12:551-567

Tabor,S. and C.C. Richardson. (1987) Proc. Natl. Acad. Sci. (USA) 84:4747-4771

Tautz,D. and C.Pfeiffe. (1989) Chromosoma 98 81-85

Valencia,G. (1966) Dros.Inf.Ser. 41:58

Watson,J.D. and F.H.C.Crick. (1953) Nature 171:740-741

Wilcox,M. (1986) In: Drosophila a practical approach (Ed. Roberts D.B.) IRL Press, Oxford, UK.

Wood,W.B. (1966) J.Mol.Biol. 16:118-133

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

Zachar,Z. and P.M.Bingham. (1982) Cell 30:529-541

Zalukar,M. and I.Enh. (1976) J.Mic.Biol.Cell 25:97-104