EXPRESSION OF THE YOLK PROTEIN 3 GENE OF *DROSOPHILA MELANOGASTER*

SUSAN LIDDELL

Thesis presented for the Degree of Doctor of Philosophy

University of Edinburgh

1989

DEDICATION

For Derek

ABSTRACT

Yolk protein 3 is a member of a small developmentally regulated, co-ordinately expressed gene family in *Drosophila melanogaster*. The yolk proteins (YP s) are synthesised in two tissues of the adult female - in the fat body cells and the ovarian follicle cells. Thus, the yolk proteins are expressed in a tissue-, sex-, and stage-specific manner and represent a good model system in which to study many aspects of gene expression in a eukaryotic organism. This thesis is concerned with the further molecular characterisation of yolk protein 3, and to this end two approaches are described.

The investigation of a female sterile mutant fs(1)A1526, noted for the absence of YP3 in the haemolymph, shows that a YP3 polypeptide is produced in the fat body cells of the female, but fails to be secreted into the haemolymph. The yp3 transcript level is shown to be reduced in both sites of synthesis in the mutant females. Genetic analysis maps the lesion causing these mutant YP3 phenotypes to the region of the yp3 locus, thereby separating the female sterile effect from the mutant YP3 phenotype. Finally, sequence determination of the yp3 gene from fs(1)A1526 flies reveals seven nucleotide variations in the 5′ region of the gene, and two within the coding region. One of the mutations in the coding sequence causes a single amino acid substitution at position 10 of the polypeptide, within the signal sequence of the protein. This is a replacement of an alanine by an aspartic acid residue and represents a likely candidate to explain the secretion defect of the YP3 protein in fs(1)A1526 flies.

Another approach to the molecular characterisation of YP3 is described which involves the use of the *Drosophila* P-element mediated transformation system to determine *cis*-regulatory regions of the yp3 gene. In this, the *cis*-acting sequences defining the tissue-specific and sex-limited regulation of yp3 expression are shown to be contained within a 3.1 kb genomic DNA fragment comprising the yp3 coding sequence with 0.7 kb of 5′ sequences, and 0.8 kb of 3′ sequences flanking the gene.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Mary Bownes, for all the advice and encouragement she has given me throughout the time I have spent in her group.

I also want to thank all of the past and present members of Mary's group, especially: Mairead Blair, Brenda Grimes, Alberto Martinez, Alan Shirras, Gregor Reid, Katrin Lineruth, Roger Slee, Jackie Forrest, Ann Scott, and Robert Kozma.

Thank you to all of the media and washing up team, particularly Margaret Anderson, Ruby Docherty, and Joan Smail.

A very warm thanks to Fiona Govan for her typing and calming influence.

Many thanks to Frank Johnston and Graham Brown for their superb photographic service.

Thanks to: Alberto, Nora, Brenda and Fiona - for proofreading; Mairead for diagrams and Kathleen for photocopying.

Thanks to many others: David, Elaine, Melanie, Michael, Nora, Patrick, Moira, Heather, Fiona, Chris, Jane, Erica, Nik, Anjam, Liz, Marys, Simon.

Finally, a big thanks to Mum and Reg, and Christine for their love and support.

ABBREVIATIONS

amp	Ampicillin
APS	Ammonium persulphate
ATP	Adenosine-5′-triphosphate
bp	Base pair
BCIG	(x-gal) 5-bromo-4-chloro-3-indolyl β-D-galactoside
BSA	Bovine serum albumin
butyl-PBD	2-(4′-tert-butylphenyl)-5-(4″-diphenyly)-1,3,4-oxadiazole
°C	Degrees centigrade
cDNA	Complementary deoxyribonucleic acid
Ci	Curies
cm	centimetre(s)
CPM	Counts per minute
CTP	Cytidine-5′-triphosphate
(d)dATP	2′(3′-di) deoxyadenosine-5′-triphosphate
(d)dCTP	2′(3′-di) deoxycytidine-5′-triphosphate
(d)dGTP	2′(3′-di) deoxyguanosine-5′-triphosphate
(d)dTTP	2′(3′-di) deoxythymidine-5′-triphosphate
(d)dNTP	2′(3′-di) deoxynucleotide-5′-triphosphate
dH_2O	Distilled water
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
DTT	Dithiothreitol
EDTA	Diaminoethanetetra-acetic acid
g	Gram(mes)
glycine	Amino acetic acid
GTP	Guanosine-5′-triphosphate
HEPES	N-2-hydroxyethylpiperazine-N′-2-ethanesulphonic acid
IPTG	Isopropyl-β-D-thiogalactoside
kb	Kilobase
kDa	KiloDalton
Klenow	Large fragment of DNA polymerase I
K rpm	kilorevolutions per minute

M	Molar
mA	Milliampere(s)
mCi	Millicurie(s)
mg	Milligram(s)
ml	Millilitre(s)
mm	Millimetre(s)
mM	Millimolar
mmol	Millimole(s)
min	Minute(s)
MOPS	Morpholinopropanesulphonic acid
mRNA	Messenger ribonucleic acid
m wt/ml wt	molecular weight
NAD	β-nictoninamide-adenine dinucleotide
neo^R	neomycin resistance gene
ng	Nanogram(s)
nm	Nanometre(s)
nmol	Nanomole(s)
NP4O	Nonidet-P40
OD	Optical density
^{32}P	β-emitting isotope of phosphorous
PEG	Polyethylene glycol
%	Percentage
pg	Picogram(s)
pH	$-1\log_{10}$ (hydrogen ion concentration)
PMSF	Phenylmethylsulphonyl fluoride
polyA	Polyadenylic acid
polyU	Polyuridylic acid
psi	Pounds per square inch
RNA	Ribonucleic acid
RNase	Ribonuclease
rRNA	Ribosomal ribonucleic acid
rpm	Revolutions per minute
S	Svedberg unit
^{35}S	β-emitting isotope of sulphur

SAM	S-adenosyl-1-methionine
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
TEMED	NNN'N'-tetra-methyl-1,2-diamino-ethane
Tris	Tris(hydroxymethyl)-amino-methane
Triton-X-100	Octylphenoxypolyethoxyethanol
tRNA	Transfer ribonucleic acid
TCA	Trichloroacetic acid
U	Unit(s)
UTP	Uridine-5'-triphosphate
UV	Ultraviolet
μCi	Microcurie(s)
μg	Microgram(s)
μl	Microlitre(s)
μM	Micromolar
μmol	Micromole(s)
V	Volts
v/v	volume per volume
w/v	weight per volume
yp1	Yolk protein gene 1
yp2	Yolk protein gene 2
yp3	Yolk protein gene 3
YP1	Yolk Protein 1
YP2	Yolk Protein 2
YP3	Yolk Protein 3
YPF1	Yolk Protein 1 Binding Factor

CONTENTS

CHAPTER 1

INTRODUCTION

INTRODUCTION

This thesis describes a molecular investigation of yp3, a member of the small yolk protein gene family in *Drosophila melanogaster*. Two approaches were taken: the first involved the analysis of a female sterile strain, fs(1)A1526, in which YP3 is absent from the haemolymph and eggs, and the second was concerned with the identification of *cis*-acting DNA sequences involved in the normal expression of yp3.

This chapter presents relevant background information concerning the system used for this study: an overview of oogenesis and vitellogenesis in *Drosophila melanogaster* is given; the yolk proteins and the yolk protein genes are described; and the current understanding of how regulation of yolk protein gene expression is achieved is discussed. Signal sequences and their involvement in protein secretion are described to provide an understanding of this type of targeting within cells and this is necessary to interpret aspects of the YP3 phenotype in the mutant fs(1)A1526 that are presented in this thesis.

SECTION 1

THE YOLK PROTEINS

OOGENESIS

Drosophila melanogaster females have one pair of ovaries joined at a common oviduct (see Fig. 1.1a), each ovary is composed of 15-20 ovarioles lying in parallel. The oocytes mature (Fig. 1.1b) as they travel from the germarium at the anterior tip of the ovariole, through the vitellarium, towards the oviduct. Each ovariole therefore contains egg chambers in different stages of development. In the germarium, division of stem cells generates clusters of sixteen cells, one of which becomes the oocyte while the other fifteen become the nurse cells and remain in close connection with the oocyte as it matures. Before exit from the germarium, the oocyte/nurse cell cluster is surrounded by small somatic cells (the profollicle cells) and the whole assembly of cells is known as the egg chamber.

Each oocyte and its fifteen nurse cells are germline cells derived from an embryonic pole cell by a series of four mitoses with incomplete cell division, which results in the production of a cluster of cells interconnected by intercellular bridges (Mahowald and Strassheim, 1970). During oogenesis the nurse cells undergo DNA replication and become highly polytenised (Jacob and Stirling, 1959). This presumably reflects the function of these cells which is to provide the developing oocyte with essential components such as lipid droplets (King, 1970) and RNA, the bulk of which is composed of the ribosomal RNAs (Mahowald and Tiefert, 1970). However some of the low abundance RNA species have been shown to be crucial to embryo development, for example in the establishment of egg polarity (see review by Akam, 1987).

As the oocytes pass through the vitellarium, they increase dramatically in volume, largely due to the accumulation of yolk protein components. The process by which the yolk is synthesised and accumulated is known as vitellogenesis. As vitellogenesis proceeds, the oocyte grows to occupy increasingly more of the egg chamber and the nurse cells gradually degenerate.

The process of oogenesis has been extensively studied and a convenient staging system developed (King, 1970) based on the size, shape and appearance of the egg chambers (see Fig. 1.1b). Detailed descriptions of the morphology and development of the ovary can also be found in Mahowald and Kambysellis (1980) and Bownes and Dale (1982). Vitellogenesis proceeds from stage 8 until late in the 10th stage. During the later stages of oogenesis, the follicle cells that surround the egg chamber become polypoid and display a diverse set of movements so that they eventually surround the

Figure 1.1

a) *Drosophila melanogaster* female reproductive system

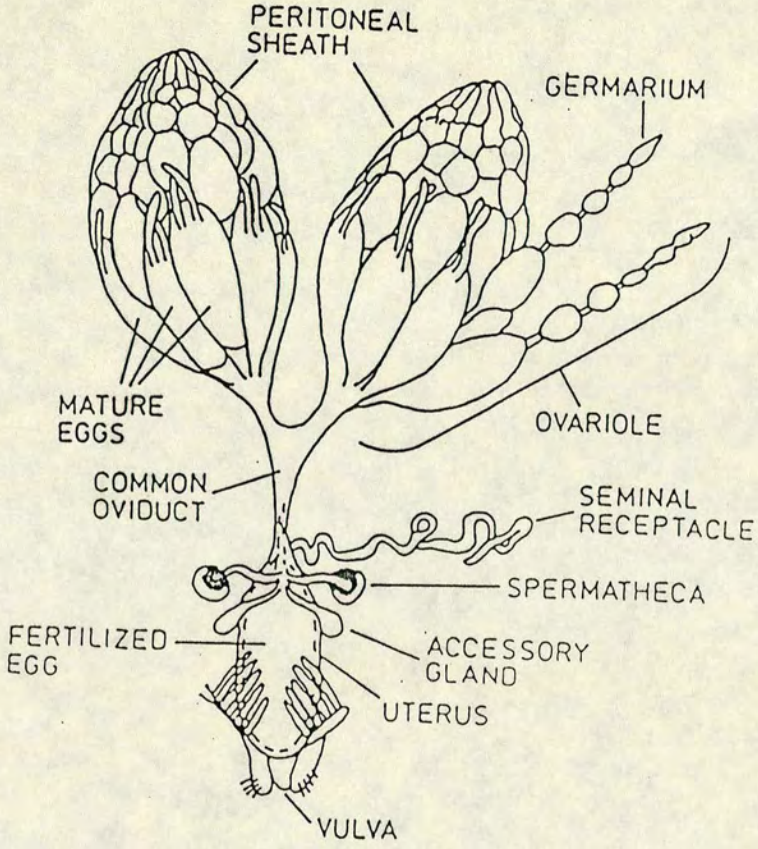
This diagram shows the paired ovaries of a single female fly, along with the common oviduct and uterus and their accessory structures.

b) Stages of oogenesis

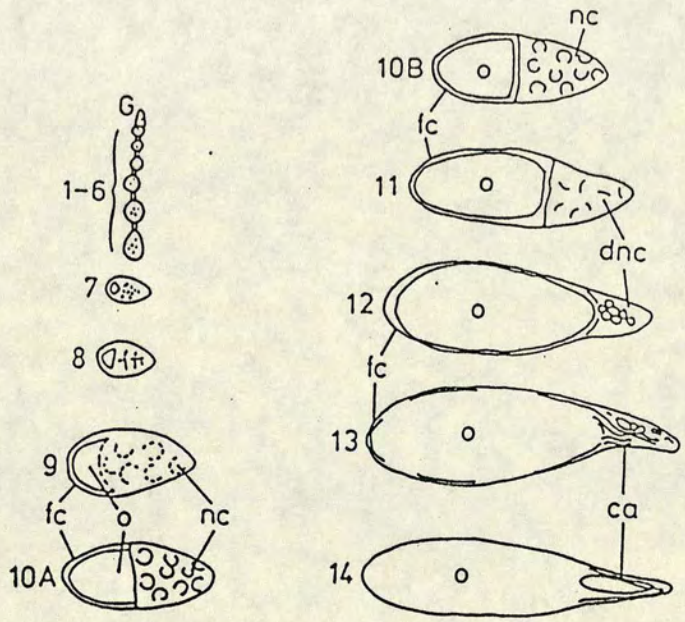
Schematic illustration of the 14 developmental stages (as in King, 1970) of the *Drosophila melanogaster* egg chamber.

g - germarium
fc - follicle cells
o - oocyte
nc - nurse cells
dnc - degenerating nurse cells
ca - chorionic appendages

A


PERITONEAL SHEATH
GERMARIUM
MATURE EGGS
OVARIOLE
COMMON OVIDUCT
SEMINAL RECEPTACLE
SPERMATHECA
FERTILIZED EGG
ACCESSORY GLAND
UTERUS
VULVA

B


G
1-6
7
8
9
fc
o
nc
10A
nc
10B
o
fc
11
o
dnc
12
o
fc
13
o
ca
14
o

oocyte and exclude the degenerating nurse cells. From stage 11 onwards, the follicle cells secrete various protective layers around the oocyte, including the vitelline membrane, a waxy layer, the endochorion and the exochorion (King, 1970; Mahowald and Kambysellis, 1980). A great deal of information has been elucidated concerning the organisation and expression of the genes encoding the chorion proteins (Martinez-Cruzado *et al.* 1988; Mariani *et al.*, 1988). At stage 14 of oogenesis synthesis of the chorion is completed and the follicle cells degenerate. The mature oocytes are fertilised as they pass into the uterus.

VITELLOGENESIS

Vitellogenesis is the process by which yolk is synthesised, and accumulated by the developing oocyte. Two types of yolk granules appear in the oocyte, α-spheres and β-spheres. The α yolk spheres are composed of protein and appear during the early stages of vitellogenesis (stages 8-10 of oogenesis), the β-spheres appear in the later stages (13-14) and contain glycogen (Bownes, 1982a). The majority of the proteinaceous yolk in *Drosophila melanogaster* is composed of three polypeptides, YP1, YP2 and YP3 (Bownes and Hames, 1977; Warren and Mahowald, 1979), and these are the only proteins which are considered here.

In organisms such as *Xenopus* and other insects such as *Locusta migratoria*, the yolk polypeptides are synthesised in only one location (the liver and the fat body, respectively) as large precursors known as vitellogenins (Tata, 1976; Kunkel and Nordin, 1985). After they have been proteolytically processed and taken up by the oocytes, the mature polypeptides are known as vitellins, to distinguish them from their primary, unprocessed precursors. However, in *Drosophila melanogaster*, since no major differences are found between vitellogenin and vitellin, and the same set of yolk polypeptides are synthesised in more than one location, they are known simply as the yolk proteins (or YP s).

In *D. melanogaster* the three major yolk proteins are synthesised in the adult female fat body and in the ovarian follicle cells that surround the developing oocytes (Bownes and Hames, 1978; Brennan *et al.*, 1982). The proteins are secreted from the fat body cells into the haemolymph, and from the follicle cells (probably into the interfollicular spaces) (Brennan *et al.*, 1982), and are selectively taken into the oocyte by receptor mediated endocytosis (Mahowald, 1972; Giorgi and Jacob, 1977).

On emergence from the pupal case there are no fully developed egg chambers in the ovaries. Transcripts from the yp genes are first seen soon after eclosion and their levels build up before reaching a plateau after 24 hours (Barnett and Wensink, 1981). The appearance of yolk proteins in the haemolymph mirrors the pattern of the transcripts and yolky oocytes are first apparent 10 hours after eclosion (Bownes, 1982a). Mated females normally begin to lay eggs when they reach 2 days of age.

The YP s synthesised in the two tissues appear to be identical, however differences in the relative amounts produced at the different sites have been reported. *In vivo* pulse labelling of intact flies and 2D-PAGE analysis of the proteins from the two tissues indicated that in the fat body cells all three yolk polypeptides are equally abundant. However, in the ovary, YP1 and 2 are produced at roughly equal levels but YP3 is synthesised at only approximately 25% of the level of YP1 or YP2 (Brennan *et al.*, 1982). PAGE analysis of the products of cell free translations of RNA extracted from fat bodies and ovaries showed that YP3 is considerably under-represented in the ovary samples, inferring that the ovaries contain reduced levels of yp3 mRNA relative to the amounts of yp1 and yp2 transcripts (Isaac and Bownes, 1982). RNA dot blot hybridisations using DNA probes for yp1 and yp3 indicated that the ovary contributes 35% of the total mRNA s coding for YP1 and YP2, but only ~10% of the transcripts for YP3 (Brennan *et al.*, 1982). Northern blot analysis confirmed that the ovary has a lower level of yp3 mRNA compared to the numbers of yp1 and 2 mRNA s. An attempt made to determine the cause of the different molar ratios of the yp transcripts in the two sites of synthesis illustrated that no apparent amplification of any of the yolk protein genes occurs and therefore inferred that either less yp3 mRNA is produced or this transcript has a faster rate of turnover in the ovary (Williams and Bownes, 1986). This study indicated that the level of yp3 transcripts in the ovary is not due to a lower rate of transcription and therefore may be the result of reduced stability. It should be noted that this is a difficult query to resolve *in vivo* in *Drosophila* since it is not feasible to perform pulse-chase experiments (as it is not possible to ensure that all unincorporated nucleotides are removed) to determine transcript half-lives, so in this case indirect means were employed to attempt an estimation of the transcription rate of the yp genes. However, that the two tissues produce different molar ratios of the YP s indicates the yp genes are subject to different controls at each site of synthesis.

REGULATION OF VITELLOGENESIS

The process of vitellogenesis is clearly critical to oocyte development and its regulation has been extensively investigated. The yolk proteins are expressed in a sex-stage-, and tissue-specific pattern and have therefore been studied as a model system for gene regulation in *Drosophila*. Molecular aspects of the yolk protein genes and their regulation are discussed later in this chapter.

Various factors have been found to affect vitellogenesis, including the hormonal condition of the animal, its nutritional status, and its genetic constitution with respect to sexual phenotype.

Nutrition

It is well known that *Drosophila* females lay eggs in relation to their environmental conditions such as temperature, humidity, and light. Nutrition is also important since starvation, or removal of specific dietary components leads to a cessation in egg production (Bownes and Blair, 1986).

Starvation of females results in a drop in vitellogenesis and egg production within one day, this occurs while the haemolymph level of YP's is still high indicating that there is a mechanism which stops yolk protein uptake or oocyte maturation when the diet is deficient. If starvation lasts for two days or more, the level of yp transcripts in the fat body decreases (this reduction is also seen in the ovary but by this time no egg chambers of the stages that normally synthesise YP's are present) (Bownes and Blair, 1986). These effects can be gradually reversed if a normal diet is restored. The effect of starvation on message levels is specific to the yp's and does not appear to be the result of failure of fat body development (Bownes *et al.*, 1988a).

The data indicates that in *Drosophila* a system exists for modulating yp gene expression and egg production in relation to diet. This kind of effect has been reported in other insects such as the mosquito, *Aedes aegypti*, where vitellogenesis is induced in response to a blood meal (van Handel and Lea, 1984).

Sex Genes

Yolk protein synthesis in *Drosophila* is also controlled by the regulatory genes governing somatic sexual differentiation. The primary determinant of sex in *Drosophila melanogaster* is the X:A ratio, the number of X-chromosomes relative to sets of autosomes in a cell. Flies with one X-chromosome and two sets of autosomes are male, those with equal numbers of X-chromosomes and sets of autosomes are female. The X:A ratio determines the sex of a fly by initiating a cascade of regulatory genes (for review see Baker, 1989), which controls the expression of the terminal differentiation genes whose products are responsible for the sexually dimorphic characteristics of adult flies. Mutations in the hierarchy of regulatory genes can interfere with or alter the sexual pathway. Thus, chromosomally female flies (X/X) that are homozygous for a mutation of the *doublesex* (*dsx*) locus are intersexual and display a variety of male and female characteristics.

The relationship between sexual phenotype and ability to synthesise yolk proteins has been investigated. Postlethwaite *et al.* (1980) and Bownes and Nöthiger (1981) analysed triploid intersexes (flies with XX:3 sets of autosomes) and sets of sex mutants causing intersexual and pseudomale phenotypes. No yolk proteins were detected in haemolymph samples from true males or pseudomales (e.g. X/X; *tra-2/tra-2*). Yolk proteins were found in intersexual flies (e.g. X/X; *dsx/dsx*). Some intersexes showed low levels of YPs; those with the smallest amounts were flies with predominantly male characteristics. Overall, these studies revealed a strong correlation with the ability of a fly to synthesise YPs and its sexual morphological phenotype, regardless of the sex chromosome constitution.

A temperature sensitive mutant was used to investigate how at least one member of the sex determination regulatory hierarchy, *transformer-2* (*tra-2*), affects the expression of the yolk proteins (Belote *et al.*, 1985). Chromosomal females can be transformed into males (pseudomales) when homozygous for a null mutation at the *tra-2* locus; reared at the permissive temperature (16°C) *tra-2*^{ts2} flies develop as phenotypic females and synthesise yolk proteins, but develop as males if reared at the restrictive temperature (29°C) and correspondingly no YPs are detected. Belote *et al.* (1985) performed temperature shift experiments with *tra-2*^{ts2} flies and examined them for the presence of yolk proteins. The results showed that the *tra-2* gene product is required by the adult for initiation of YP synthesis and is likely to be necessary for its

maintenance. That it is required for maintenance of YP synthesis was confirmed (Bownes *et al.*, 1987) by shifting *tra-2*^{ts1} flies to 29°C then, when YP synthesis was greatly reduced, returning them to 18°C and showing that YP synthesis was resumed in these flies. Investigating the yp mRNA levels indicated that the control of YP synthesis by the *tra-2* gene product is at the level of transcription or transcript stability (Belote *et al.*, 1985).

YP synthesis was further examined to elucidate whether the sex-specific control of YP synthesis operates at a transcriptional or a post-transcriptional level, or both. Kraus *et al.* (1988) effectively removed the yp1 gene from sex-specific transcriptional control by expressing the YP1 coding sequence under the control of the hsp70 promoter in transformed flies. The efficiency of the expression of the yp1 gene at the post-transcriptional levels of RNA splicing, translation, and protein secretion in transformed males was compared with that in normal females. This study showed that the yp1 transcript in *Drosophila* males is stable, efficiently spliced and translated, proving that there are no post-transcriptional controls involved in the sex-specific control of YP1 expression in male *Drosophila* and that this regulation in males is achieved by limiting the transcription of the YP genes.

Hormones

The hormones ecdysone and juvenile hormone (JH) have key roles in regulating many aspects of insect development, such as moulting and metamorphosis. Both hormones are also involved in adult reproduction and as part of this system vitellogenesis is one of the processes affected. The precise roles of JH and ecdysone vary between insects (e.g. see Kelly *et al.*, 1987) making it difficult to perceive general features or produce a unifying model. Generally, JH is secreted from the corpora allata and ecdysteroids are produced by the ring gland in the larva and a variety of tissues in the adult such as the ovaries and oenocytes. A wide diversity of experimental techniques have been used to examine the roles of JH and ecdysone in insects. Many *Drosophila* mutants have been involved in examining the regulation of vitellogenesis; such as *ecd*^{1} (*ecdysoneless*) (Redfern and Bownes, 1983), a temperature sensitive mutant which has very low levels of active ecdysone at the restrictive temperature, and *ap*^{4} (*apterous*), a mutant that shows defective vitellogenesis (Redfern and Bownes, 1982) and has reduced levels of JH (Bownes, 1989). Another approach has involved the use of

isolated abdomens. When female abdomens are ligated shortly after eclosion, the synthesis of yolk proteins is considerably reduced. Such studies have implicated JH and ecdysone in the regulation of vitellogenesis. Note that the experiments in *Drosophila* do not use JH itself, but the synthetic analogue, ZR515.

Juvenile hormone has been shown to stimulate the uptake of yolk proteins by oocytes (Postlethwaite and Handler, 1979) and affects the yolk protein levels. JH administered to isolated abdomens stimulates synthesis of the YP s (Postlethwaite and Handler, 1979). This effect is seen for both the fat body and the ovary (Jowett and Postlethwaite, 1980). When adult females were starved to reduce yolk protein levels, treatment with JH caused a slight increase in yp transcript accumulation (Bownes and Blair, 1986; Bownes *et al.*, 1988a). From these studies alone it is clear that JH can affect the level of synthesis of the yolk proteins. However, the amount of hormone required to achieve this may be high in comparison to the levels that are detectable in the fly. The precise *in vivo* role of JH with respect to yp expression, and how this is carried out remains unclear. The titre of JH in *D. melanogaster* is high at eclosion and drops off shortly afterwards, and in the mutant ap^4 some yp synthesis is maintained even although the JH level is very reduced, so it is possible that JH acts to initiate ovarian and fat body maturation to generate a state of competence for YP synthesis in these tissues, and thereafter is only required at low levels to maintain vitellogenesis and yolk protein production. Another mutant, isolated recently, shows some similarities to the mutant phenotype of ap^4. The mutant *cricklet* (*clt*) (Shirras and Bownes, 1989) is defective in yp synthesis and shows defective histolysis of larval fat body and rudimentary ovarian development. These features suggest that clt is defective somewhere in the JH response pathway, such as a mutant JH carrier protein or receptor. If this proves to be the case the mutant *clt* provides further support for the theory of JH being most important to vitellogenesis during the early stages of adult development.

The steroid hormone ecdysone has also been shown to affect the synthesis of the yolk proteins. Addition of 20-hydroxyecdysone to isolated abdomens causes an increase in YP levels (Jowett and Postlethwaite, 1980). Isolated abdominal body walls (which also produce virtually no YP s) treated with 20-hydroxyecdysone leads to a secondary induction of YP synthesis (Wu *et al.*, 1987); in mature females that have been starved to reduce YP levels, ecdysone causes an increase in yp mRNA accumulation (Bownes and Blair, 1986); and in normally fed females the hormone may cause an increase in the transcript levels of the yp's (Bownes *et al.*, 1983). Ecdysone

does not appear to affect ovarian YP synthesis since levels of yp transcripts in the ovary are not altered when females are injected with 20-hydroxyecdysone (Bownes *et al.*, 1983).

Injection of *D. melanogaster* males with ecdysone leads to the appearance of yolk polypeptides in the haemolymph (Postlethwaite *et al.*, 1980) and of yp transcripts in the fat body (Bownes *et al.*, 1983). This effect of ecdysone on yp synthesis in males is transient (Bownes, 1982b) and transcription of the yp genes can be prevented using cycloheximide, an inhibitor of protein synthesis (Bownes *et al.*, 1987). The effect of cycloheximide appears to be specific to yp gene transcription and suggests that other gene products are necessary for the ecdysone induced activation of transcription of the yolk protein genes in males.

Although expression of the yolk protein genes can be affected by ecdysone under all these experimental conditions, the actual *in vivo* role of ecdysone in the regulation of yolk protein synthesis is not clear. The levels of hormone required to induce synthesis in males or increase levels in females are usually several orders of magnitude greater than the normal hormone titre in adults (with the exception of one set of isolated abdomen experiments, Wu *et al.*, 1987). The actual level of ecdysterones in *Drosophila* is currently an issue of debate. Some assays indicate levels in males and females are similar and that the titre does not alter throughout the adult lifespan (Bownes and Rembold, 1987; Bownes, 1989), while other data suggest that variations in ecdysone levels do occur (Schwartz *et al.*, 1985).

Transcription of the yp genes can be induced with ecdysone in sex gene mutants, such as *transformer* and *doublesex*, which do not normally synthesise the YP's (Bownes *et al.*, 1987; Bownes and Nothinger, 1981), indicating that ecdysone can somehow override the normal sex-limiting controls imposed on the yolk protein genes in males. When Shirras and Bownes (1987) investigated the *cis*-sequences of yp1 involved in regulating yp gene expression, they found that while normal female synthesis was obtained, induction of male expression with ecdysone could not be achieved with the same DNA sequences. This provides some support for the idea that induction in males does not operate via the normal system of expression in females.

The issue of hormonal regulation of vitellogenesis has been intensively investigated in insects other than *D. melanogaster*. In series of experiments similar to the types performed in *Drosophila*, the vitellogenin genes of other species (such as *Aedes aegypti*) can be affected by *in vivo* levels of JH and ecdysone (e.g. see Hagedorn,

1989) and their roles in regulating vitellogenin synthesis have been more clearly established.

THE YOLK PROTEINS OF *DROSOPHILA MELANOGASTER*

General Aspects and Characteristics of the Yolk Proteins

Adult female *Drosophila melanogaster* synthesise three yolk polypeptides of similar sizes and characteristics, in two tissues, the fat body cells and the ovarian follicle cells (Bownes and Hames, 1977; Warren and Mahowald, 1979; Isaac and Bownes, 1982; Brennan *et al.*, 1982). The three yolk proteins are encoded by three single copy genes which are found on the X-chromosome (Barnett *et al.*, 1980). The three polypeptides, YP1, YP2 and YP3, are of apparent molecular weights 47kDa, 46kDa and 45kDa respectively, as determined by SDS-polyacrylamide gel electrophoresis. These values are similar to molecular weight estimates calculated from the predicted amino acid sequences of YP1, 2 and 3 of 48300Da 49682Da and 46090Da respectively (Hung and Wensink, 1981, 1983; Garabedian *et al.*, 1987; Yan *et al.*, 1987). The gene sequences predict polypeptide lengths of 439 amino acids for YP1, 442 for YP2, and 420 for YP3. Comparison of these predicted sequences reveals considerable similarities in the primary sequence of the three proteins with YP1 and 2 showing 53% identity, YP2 and 3 48%, and YP1 and 3 53%. The proteins show the highest degree of similarity to one another in the carboxyl two fifths where up to 70% identity is reached. It is highly likely that this degree of similarity reflects structural features necessary to the common functions of the YPs, such as transport in the haemolymph, uptake by oocytes and packaging into yolk granules.

It is not known how the three yolk proteins are organised in the native form. In Gingeras *et al.* (1973) they appear to have a molecular weight of approximately 190kDa, however Isaac (1982) suggests that the YPs associate as hexamers.

The yolk proteins have long been assumed to play a nutritional role in embryogenesis. In *D. melanogaster* the YPs begin to be broken down and utilised only during the later stages of embryogenesis, when midgut formation is occurring (Bownes, 1982). Recently, however, another function of the YPs was proposed. Bownes *et al.* (1988b) compared the sequences of the YPs with those of other proteins and identified a sequence showing significant similarity to the substrate binding site of triacylglycerol

lipase from the domestic pig. The authors suggest that this similarity reflects binding site similarities and that the function of this region in the YP's is to bind apolar conjugates of ecdysteroids so that when the YP's are proteolytically cleaved during embryogenesis, the conjugates become available for metabolism and subsequent release of hormone. In this way the timed breakdown of the yolk proteins would lead to a timed release of hormone, triggering other events in embryogenesis, such as cuticle deposition.

The conservation of the yolk proteins within other species of *Drosophila* and in other dipteran species (*Lucilia, Protophormia, Calliphora* and *Sarcophaga*) was demonstrated by the ability of the yolk proteins from distantly as well as closely related species to be selectively taken up by *D. melanogaster* and *D. funebris* ovaries (Martinez and Bownes, submitted 1989). The authors suggest that not only the domains of the YP's involved in selective endocytosis are conserved, but other features may also be retained (such as the hypothesised fatty acid ecdysteroid conjugate binding, described above) and that the YP's may therefore be functionally more important than an amino acid supply.

Modifications of the Yolk Proteins

Two major processing steps are involved in the maturation of the yolk proteins in *D. melanogaster* (Fig. 1.2a) (Warren *et al.*, 1979).

YP1, YP2 and YP3 each possess a typical leader peptide (Fig. 1.2b) that is characteristic of secreted proteins. The first processing event is the cleavage of this amino-terminal peptide from each of the nascent yolk polypeptide chains. The cleavage event is seen *in vitro* as a cotranslational reduction of 1kD in apparent molecular weight when dog pancreas microsomes are added to *in vitro* translations (Brennan *et al.*, 1980). The unprocessed forms of the YP's are normally never seen *in vivo*. Minoo and Postlethwaite (1985) attempted to locate the site of signal peptidase cleavage by *in vivo* labelling the yolk proteins with ^{35}S-cysteine. In the amino acid sequence predicted from the DNA sequence of the three genes, YP1 has only one cysteine residue at position 12, YP2 has two at positions 8 and 12, and YP3 has a total of four at positions 7, 12, 103 and 312. Of the three proteins only YP3 was found to be labelled with ^{35}S-cysteine, indicating that at least 12 residues are removed from YP1 and YP2 upon signal peptide cleavage. From the predicted amino acid sequences, Hung and Wensink (1983)

Figure 1.2

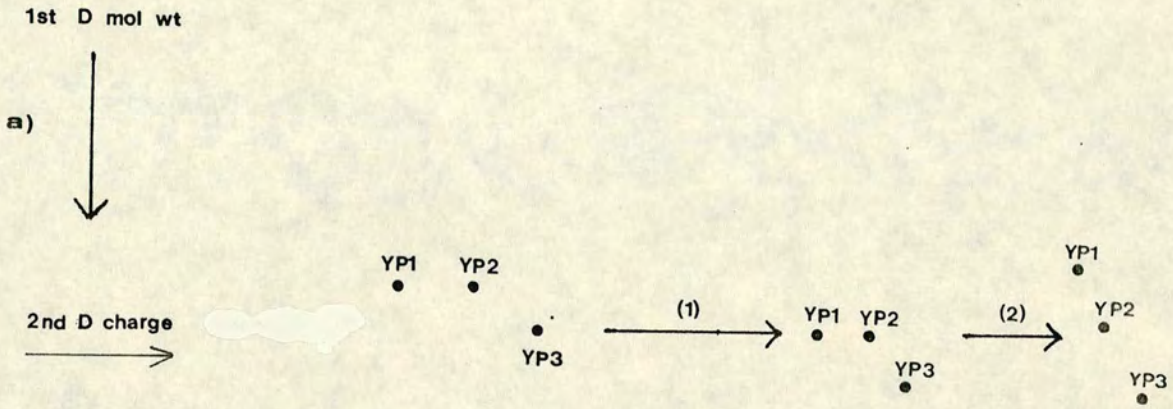
a) The effects of the two major modifications of the yolk proteins, represented diagrammatically as they are observed on 2-dimensional gels.

(i) signal peptide cleavage

(ii) increase in molecular weight of YP1

(b) The sequence of the first 25 amino acids of the yolk proteins predicted from the DNA sequence of the yp1, yp2 and yp3 genes. The postulated signal sequence peptidase cleavage site is indicated by an arrow.

Figure 1.2



b)

YP1 MNPMRVLSLLACLAVAALAKPNGRM

YP2 MNPLRTLCVMACLLAVAMGNPQSGN

YP3 MMSLRICLLATCLLVAAHASKDASN

suggest that the leader peptides of YP1 and YP2 are 20 amino acids long. However, this proposal would place a charged residue (Lys) in the case of YP1, and a large polar residue (Asn) in the case of YP2, at the -1 position relative to the site of cleavage, two classes of residue that are generally not found in this position of signal peptides (von Heijne, 1983, 1984b). It is more likely that signal peptidase cleaves the nascent YP1 and YP2 polypeptides after amino acid 19 since processing at this site in both cases better fulfils the requirements for signal peptidase substrates. Similarly, from the predicted amino acid sequence of YP3 (Garabedian *et al.*, 1987) and using the rules for signal peptide cleavage outlined by von Heijne (1983, 1984b) and Perlman and Halvorson (1983), the most likely site of signal peptidase processing of YP3 is between residues 19 and 20 (see Fig. 1.2b). This is based on the following points:

- The residues at positions 17 and 19 (both Ala) are the only set of amino acids in this region of the YP3 polypeptide which conform to the (-3,-1) rule (von Heijne, 1983).
- His is a favourable residue for the -2 position (von Heijne, 1983).
- The increased incidence of charged residues in the region just after amino acid 19 is typical of mature protein sequences (von Heijne, 1984b).

Yan *et al.* (1987) also predict the site of signal peptidase cleavage for all three yolk polypeptides to be after residue 19. The YP signal sequences do not show the turn promoting residues that are commonly found at the -5/-6 position (von Heijne and Abrahamsén, 1989) and the hydrophobic residues continue to be present into the c-region. This means that there is no clearly defined boundary between the c-terminal end of the h-region. and the beginning of the c-region, a characteristic proposed to be involved in defining a "window" for cleavage (von Heijne, 1984b). However, this is sometimes observed, more particularly for the shorter signal peptides where the different regions are less well defined and appear to overlap (von Heijne, 1985).

The second major processing event affecting the yolk proteins results in an increase in the apparent molecular weight of YP1 from 45kD to 46kD (Fig. 1.2a) (Warren *et al.*, 1979). This takes place after signal peptide cleavage: the modification causing it has not been identified.

In addition to the two major processing steps all three yolk proteins undergo other modifications as indicated by the presence of several charge variations seen on 2-dimensional polyacrylamide gels (Warren *et al.*, 1979; Minoo and Postlethwaite, 1985). Separation of the YP s using iso-electric focusing in the first dimension and SDS-PAGE

in the second, results in a distinctive pattern where the yolk polypeptides focus over a narrow pH range of 7.0-7.2, YP1 being the most acidic and YP3 the most basic (Warren and Mahowald, 1979). A number of charge variants for each YP are discernible on these gels.

Two other modifications that may affect the YP s in *D. melanogaster* are phosphorylation and glycosylation.

In vivo labelling with ^{32}P-inorganic phosphate demonstrated that all three of the yolk proteins contain covalently attached phosphate moieties (Brennan and Mahowald, 1982; Minoo and Postlethwaite, 1985). Partial proteolysis of ^{32}P-phosphorylated YP s (Brennan and Mahowald, 1982) indicated that, by the methods used, the yolk proteins are phosphorylated in a manner that is indistinguishable in the two tissues of synthesis. It is not clear whether this modification alone is sufficient to account for the apparent charge heterogeneity noted for the YP s. At what stage during maturation of the YP s the process of phosphorylation occurs has not been elucidated, and the actual sites of phosphate addition have not been established.

Purified yolk proteins can be phosphorylated *in vitro* his was performed using protein kinase and [gamma-^{32}P] ATP (Di Mario *et al.*, 1987). *In vivo*, the three proteins are equivalently labelled by ^{32}P phosphate (Brennan and Mahowald, 1982), however, *in vitro*, YP2 is preferentially labelled above the level seen for YP1 and YP3. Di Mario *et al.* (1987) suggest that phosphorylation of previously unphosphorylated sites of YP2 is occurring, rather than phosphate exchange.

Some evidence exists which indicates that the yolk proteins are subject to glycosylation. Injection of females with ^{14}C-mannose or ^{14}C-glucosamine resulted in the incorporation of both sugars into the mature YP s (Minoo and Postlethwaite, 1985). However, when Mintzas and Kambysellis (1982) prepared purified individual YP s from *D. melanogaster* egg extracts and estimated glycosylation by chemical methods, they found that only YP2 contained substantial amounts of carbohydrates. The sequence Asn-X-Thr/Ser signals a potential site of N-linked glycosylation. Searching the predicted amino acid sequences of the yolk proteins (Hung and Wensink, 1983; Hoveman *et al.*, 1982; Yan *et al.*, 1987) reveals none of these sites in the YP1 or YP3 polypeptide sequences, and one in the YP2 sequence at residue 25 (Asn^{25}-Arg-Ser). This indicates that the carbohydrates that are attached to the yolk proteins of *D. melanogaster* are not N-linked and are perhaps O-linked. Further support for the YP s not bearing N-linked oligosaccharides comes from Minoo and Postlethwaite (1985) who mention that

endoglycosidase-H, which cleaves N-linked sugars, does not cleave the *D. melanogaster* yolk protein carbohydrates.

As so little is known about the glycosylation of the YP's of *D. melanogaster*, its importance with respect to their secretion has not been determined. In other insects investigations concerning the glycosylation of the vitellins and vitellogenins have been more successful and in general, it appears that their oligosaccharides are of the simple, high mannose type (Kunkel and Nordin, 1985). In *Blatella germanica* (the cockroach), glycosylation seems to be required for normal cleavage of a major pro-vitellogenin peptide, this was shown by treating the insect with tunicamycin (a potent inhibotor of N-glycosylation) and indicates that the attachment of oligosaccharides is required for the recognition by, or transit to, compartmentalised fat body proteases (Wojchowski *et al.*, 1986). The same study showed that when glycosylation is inhibited the vitellogenins are not secreted. However, whether secretion of vitellogenins in *B. germanica* depends only upon glycosylation or also involves the subsequent cleavage events, is not clear. In insects in general, the role of glycosylation of the vitellins and vitellogenins has not been elucidated; Kunkel and Nordin (1985) have postulated that perhaps the oligosaccharide chains on these proteins are required for correct conformation upon assembly.

Another post-translational modification that the YP's of *D. melanogaster* undergo is the sulphation of tyrosine residues. When female flies were treated with inorganic ^{35}S-sulphate, the YP's were prominently labelled (Baeuerle and Huttner, 1985). Since the charge variants of each YP incorporated similar ratios of ^{35}S-sulphate this modification is unlikely to be responsible for the different isoelectric forms observed. The site of tyrosine sulphation *in vivo* and *in vitro* (in mouse fibroblast cells) has been identified in YP2 as tyrosine 172 (Baeuerle *et al.*, 1988; Friederich *et al.*, 1988a). The effect of sulphation at this residue on the intracellular transport and secretion of YP2 has been investigated *in vitro*. The inhibition of tyrosine sulphation, either by replacement of the tyrosine residue by phenylalanine (by oligonucleotide-directed mutagenesis), or by chlorate treatment of the cells, was shown to retard transport to the cell surface (Friederich *et al.*, 1988b). The effect of tyrosine sulphation at residue 172 on the transport and secretion of YP2 in *D. melanogaster* is being examined (in a collaborative effort by Bownes, Martinez and Huttner) by P-element mediated germ line transformation using the site directed mutagenised YP2 gene encoding phenylalanine at residue 17).

Although lipids have been shown to be associated with the vitellogenins in many other insects (Kunkel and Nordin, 1985), the linkage of lipids into the yolk proteins of *D. melanogaster* has not been investigated to any large degree. However, Bownes *et al.* (1988b) have shown that treatment of purified YP s with esterase and proteases causes the release of ecdysteroid hormone indicating that the YP s bind fatty acid ecdysteroid conjugates. They propose another function for the YP s, in addition to their assumed role as a nutritional source where their degradation leads to the release of an ecdysteroid hormone that can trigger key events in embryogenesis.

THE YOLK PROTEIN GENES

The three yolk polypeptides of *D. melanogaster* are encoded by three single copy genes located on the X-chromosome (Fig. 1.3.a) (Postlethwaite and Jowett, 1980; Barnett *et al.*, 1980). All three genes have been cloned (Barnett *et al.*, 1980) and sequenced (Hung and Wensink, 1981, 1983; Hovemann *et al.*, 1981; Garabedian *et al.*, 1987; Yan *et al.*, 1987). The yp1 and yp2 genes are very closely linked, being separated by only 1.2kb, and are divergently transcribed (Hung *et al.*, 1982). The yp3 gene is located nearer to the centromere, approximately 1000kb away from yp1 and yp2. *In situ* hybridisations to larval salivary gland chromosomes localised the yp1 and yp2 genes to the 8F-9A region of the X-chromosome, and yp3 to the 12B/C region (Barnett *et al.*, 1980; Bownes, 1982).

Southern hybridisation experiments (Barnett *et al.*, 1980) and heteroduplex analyses (Hung *et al.*, 1982) indicated that while yp1 and yp2 share considerable sequence homology, the yp3 gene seemed quite dissimilar to the other two. However, the nucleotide sequences of the genes showed that all three genes have a high level of sequence similarity which is most extensive in the translated regions of the genes, particularly at the 3′ ends. The length of the mRNA's and the number and length of the introns of the genes determined by the DNA sequence data are in agreement with previous estimates (Hung *et al.*, 1982). The gene structure varies very little between the three genes (Fig. 1.3.b), yp3 has two introns while yp1 and yp2 each have one. All of the yp introns are short in length, the yp1 intron at 76nts is the longest. The first of the introns in yp3 and those in yp1 and yp2 are located at a very similar position in each of the genes, after a short first exon, at nucleotide 279 in yp1, 286 in yp2 and 262 in yp3 (relative to the transcriptional start site). A single *in vivo* transcript was detected

Figure 1.3

a) Positions of the yp genes on the X-chromosome (Barnett *et al.*, 1982).

The upper part of the diagram represents portions of the *D. melanogaster* X-chromosome (in polytenised form) and shows the banding pattern in the regions where the yp1 and yp2, and the yp3 loci are found.

In the lower section the open boxes represent the yp genes, the arrows indicate the direction of transcription, and the numbers are lengths of DNA in kbp.

b) Diagram showing various features of the yp genes and their mRNA's.

The single lines represent 5′ and 3′ flanking sequences (the dotted lines indicate areas where the full length is not represented).

The hatched boxes represent the exons of the genes, and the small open boxes the introns. The lines above represent the transcripts, the caret symbols (^) show the positions of the introns.

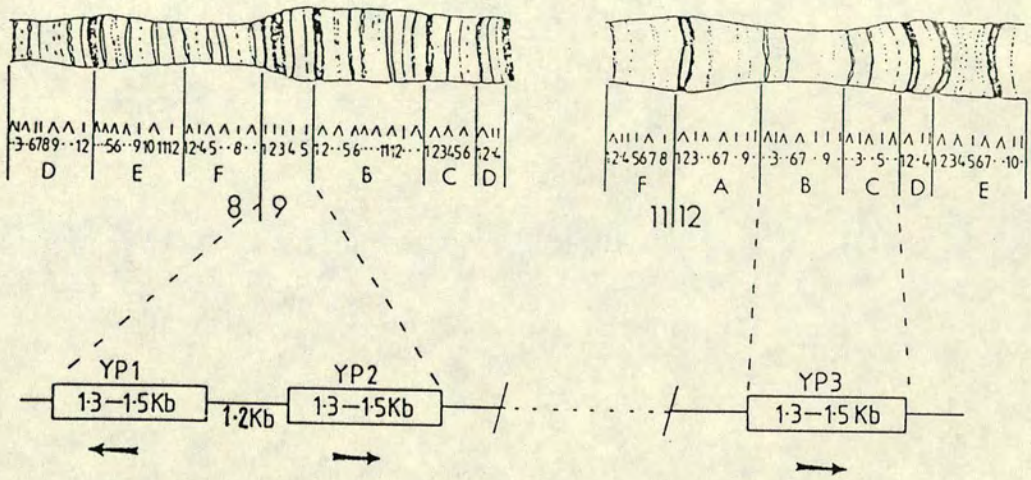
The arrows indicate the direction of transcription and the sites of polyA addition.

o = TATA box sequence

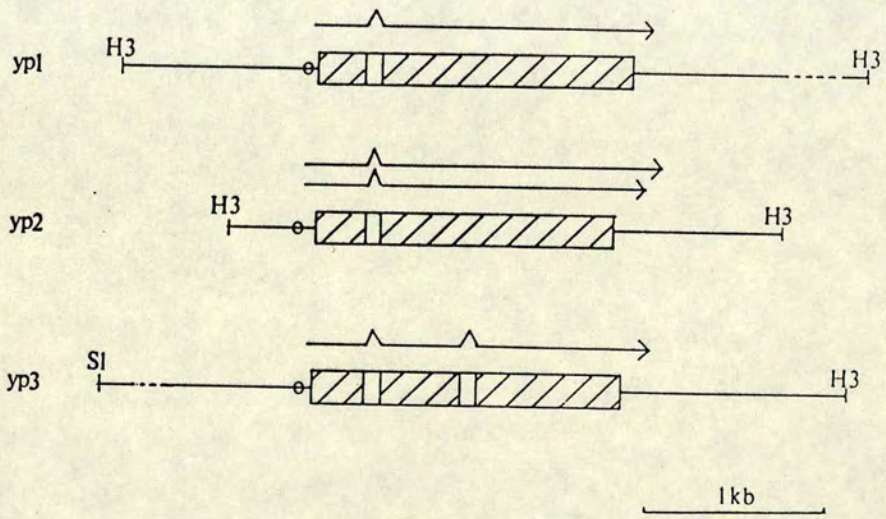
S1 = SalI recognition site

H3 = HindIII recognition site

A


D
E
F
B
C
D
8
9
F
A
B
C
D
E
11
12
YP1
1·3—1·5Kb
1·2Kb
YP2
1·3—1·5Kb
YP3
1·3—1·5 Kb

B


ypl
H3
H3
yp2
H3
H3
yp3
SI
H3
1kb

for each of the yp1 and yp3 genes and two were found for the yp2 gene (Hung *et al.*, 1982). The transcript lengths were estimated, by Northern blot experiments and RNA protection studies, to be 1.60kb for yp1, 1.60kb and 1.67kb for yp2 and 1.54kb for yp3. These lengths are in accord with those estimated from the DNA sequences of the genes of 1559nts for the yp1 mRNA, 1614nts and 1698nts for the yp2 mRNA's and 1499nts for the yp3 transcript.

In many other aspects of their structural organisation and arrangement of consensus sequences, the yp genes are very similar. Much of this, including the position of 5′ consensus sequences (such as the TATA box, capping site, etc) is indicated in figure 1.3.b.

The high degree of homology shown by the DNA and peptide sequences, and the locations and numbers of introns, are indicative of an evolutionary relationship between the three yolk protein genes.

Comparison of the *D. melanogaster* yolk protein gene family with those of other *Drosophila* species revealed considerable similarities within the *D. melanogaster* sibling species group. All of these species have three yolk proteins of similar sizes to those of *D. melanogaster* that are encoded by single copy X-chromosomal genes, two of which are closely linked (Kozma and Bownes, 1986). The yolk protein gene of a distantly related *Drosophila* species, *D. funebris*, shows homology to the *D. melanogaster* yp DNA sequences (on Southern blots) and the yolk polypeptide is recognised by antibodies raised to the *D. melanogaster* YP s. That the *D. funebris* yolk protein is more similar to YP3 than it is to YP1 and YP2 provides circumstantial evidence to support the speculation that, in *D. melanogaster*, YP3 was the original precursor gene that gave rise to YP1 and YP2 by gene duplication events. Yan *et al.* (1987) continue to propagate this idea and discuss the intron/exon organisation of the yp genes in the light of this proposal in an attempt to substantiate it. However, the idea that YP3 represents the ancestral gene seems less likely due to evidence from another *Drosophila* species, *D. grimshawi*. This is a member of the Hawaiian species of *Drosophila*, which are quite distantly related to *D. melanogaster*. The vitellogenin genes from *D. grimshawi* have been characterised (Hatzopoulos and Kambysellis, 1987) and show a great deal of similarity to the organisation of the *D. melanogaster* gene family. In *D. grimshawi* there are three vitellogenin genes, two of which are closely linked (with an intergenic spacer of 1.75kb) and divergently transcribed. All three genes have either 1 or 2 short introns and show the highest degree of homology at their 3′ ends. The degree of similarity

shown in the organisation and number of the *D. grimshawi* vitellogenin genes and the *D. melanogaster* yp genes makes it seem very unlikely that the single yp gene in *D. funebris* represents the original *Drosophila* yp gene system since the Hawaiian *Drosophila* are thought to be more distantly related to *D. melanogaster* than *D. funebrisis* (see Fig. 1, Chapter 17, Throckmorton, 1975) and so at the time of divergence of the Drosophilidae into Drosophila and Saphophora, the yp gene family with three members was already established. Evidence from other dipteran species support this idea. Martinez and Bownes (1989) note that several other diptera, namely *Calliphora erythrocephala* (the blowfly), *Sarcophaga argyrostoma* (the fleshfly), *Musca domestica* (the housefly), *Lucilia sericata* (the green bottlefly) and *Protophormia terrae-novae* (the blue bottle fly), all have three vitellogenins that are very similarly sized to YP1, YP2 and YP3 of *D. melanogaster*, and which are recognised by antisera raised to the *D. melanogaster* YP s. This indicates that the three member yp gene family existed in this form before the divergence of the Muscamorpha into the Drosophilidae and the Muscoidae, and overall the yolk protein gene system has been conserved in these groups (with a few exceptions such as *D. funebris* and *D. virilis*, and it is plausible that these species have lost one or two yp genes).

However, the dipteran vitellogenin gene family seems to be different to the gene family that is used by many other organisms that have yolk proteins (including *Xenopus*, chicken, nematodes and sea urchins), which all use distantly related members of what appears to be the same family of genes to encode their vitellogenins (Bluementhal and Zucker-Aprison, 1987).

REGULATION OF YOLK PROTEIN GENE EXPRESSION

Yolk protein gene expression involves complex developmental and genetic signals to ensure that the polypeptides are synthesised with the correct specificities, namely: sex, tissue, developmental stage of the fly, and stage within the lifetime of a single cell type (the ovarian follicle cells). Regulation of expression at the level of transcript initiation depends upon interactions of signals defining these specificities (i.e. *trans*-acting protein factors) with sequences in or around the yp genes. Several *cis*-acting elements responsible for particular aspects of the transcription pattern of the yp1 and yp2 genes have been identified, and attention is now also being focused upon identifying

proteins which interact with these *cis*-elements and with other sequences that are not yet identified.

cis-Acting Sequences

Examination of the *in vivo* expression pattern of deletion and substitution variants of the yp genes using P-element mediated germ line transformation (Spradling and Rubin, 1982; Rubin and Spradling, 1982) studies has identified some of the important *cis*-acting sequences involved in the regulation of yp1 and yp2. Expression of the sequences transformed into flies has been distinguished from endogenous yp expression in a variety of ways: by tagging the yp transcripts with a fragment of M13 DNA (Garabedian *et al.*, 1985), truncating the yp1 gene (Tamura *et al.*, 1985), fusing various promoter sequences to reporter genes such as a hsp70/lacZ fusion gene (Garabedian *et al.*, 1986), and using a mutant strain of *Drosophila* in which YP2 is not synthesised (Tamura *et al.*, 1985).

Garabedian *et al.* (1985) separated the yp1 and yp2 genes at the HindIII site of the intergenic spacer and inserted a portion of bacteriophage M13 DNA into the second exon of the yp1 and yp2 genes (Fig. 1.4, constructs a and b), which allowed them to determine the expression pattern of these constructs by Northern blot analysis. Transcripts from both of the introduced genes were found in flies of the correct sex and age, transcripts from only the yp1-M13 gene were found in fat bodies and transcripts from only the yp2-M13 gene were detected in ovaries. This means that both constructs bear the DNA sequences necessary for the correct sex- and time-specific pattern of yp synthesis and that there are two tissue specifying *cis*-elements acting on the yp1 and yp2 genes, one that is necessary for expression in fat body cells, and the other for expression in the ovaries.

The results of Tamura (1985) provided supporting evidence for the majority of the conclusions of Garabedian *et al.* (1985). In this case a 5.0kb genomic fragment of DNA containing a 3′ truncated yp1 gene and the complete yp2 coding sequences (Fig. 1.4, construct c) was used for transformations. A mutant fly strain that does not produce any detectable YP2 polypeptide allowed synthesis from the introduced yp2 gene to be detected by SDS-PAGE of protein extracts from transformed lines. This clearly showed that YP2 was made in the fat body cells and exported into the haemolymph of transformed female flies. Contrary to the claim by the authors, the analysis performed

Figure 1.4

A selection of the yp gene constructs that have been used in transformation studies to identify the important regulatory regions of the yp1 and yp2 genes. The results of these experiments are discussed in the text.

Notes:

—— sequences that are 5′ or 3′ to the yp1 or yp2 genes. The numbers refer to the distance in bp from the yp1 cap site.

yp1 and yp2 coding sequences. The arrows indicate the direction of transcription.

yp1 and yp2 introns.

M13 sequence

hsp70/lacZ sequences: consisting of 195 bp, or 43 bp, of sequence upstream of the *Drosophila* hsp70 cap site fused to 8 kbp of the *E.coli lacZ* operon.

Adh coding sequences (a total of 1.75 kb of Adh DNA sequence is included in this construction).

// The slashed lines indicate that only a portion of the sequence is included in the diagram.

E Indicates that these constructs have been tested for ecdysone inducibility - in no case was induction observed.

f-g f has slightly higher β-galactosidase activity than g.

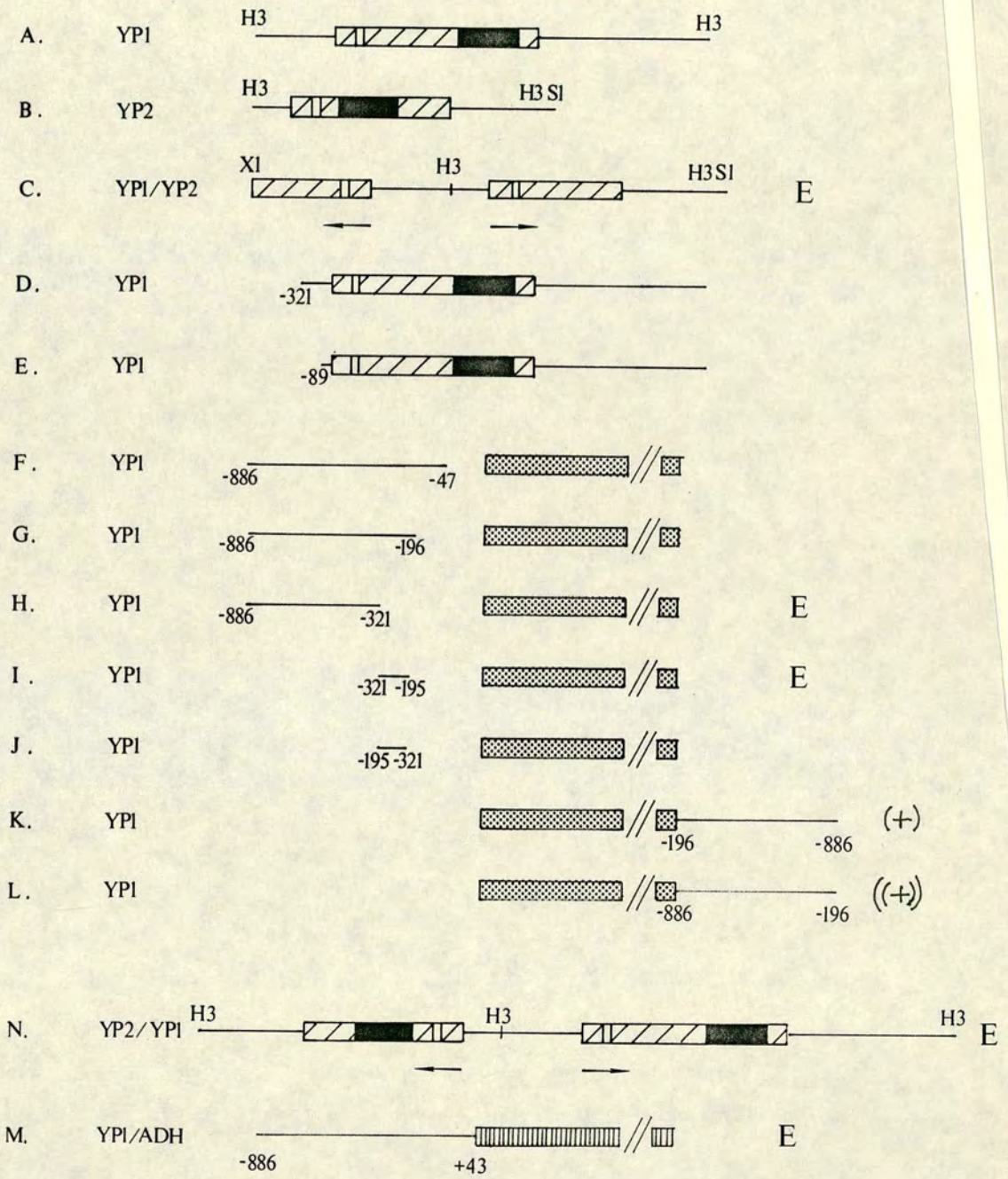
k-l In these constructs the yp1 sequences were placed 8 kbp downstream of the hsp70/lacZ cap site.

(+) Indicates that expression is weak, as measured by β-galactosidase activity, and ((+)) indicates that expression is weaker still.

H3 HindIII recognition site

X1 XbaI recognition site

S1 SalI recognition site


A.
YP1
H3
H3
B.
YP2
H3
H3 S1
C.
YP1/YP2
X1
H3
H3 S1
E
D.
YP1
-321
E.
YP1
-89
F.
YP1
-886
-47
G.
YP1
-886
-196
H.
YP1
-886
-321
E
I.
YP1
-321
-195
E
J.
YP1
-195
-321
K.
YP1
-196
-886
(+)
L.
YP1
-886
-196
((+))
N.
YP2/YP1
H3
H3
H3
E
M.
YP1/ADH
-886
+43
E

does not appear to be sufficient to show that any of the YP2 found in the ovaries of the transformed flies was the result of YP2 synthesis by the ovarian follicle cells. Since they did not look for *de novo* synthesis the YP2 found could all be fat body derived. However, the study did indicate that sufficient signals were present in the transformed DNA sequences to ensure the proper sex and temporal expression of YP2 in the fat body cells. Since transcription from the introduced truncated yp1 gene continues into adjacent vector sequences, DNA probes specific to these sequences allowed the presence of transcripts originating from the transformed yp1 sequences to be determined by Northern blot experiments. Transcripts from the introduced yp1 were detected only in female fat body cells and ovarian follicle cells, at roughly the same levels as the transcripts from the native yp1 gene, these transcripts also followed the temporal pattern normally expected of the yp's. This demonstrates that the 5.0kb DNA fragment contains all the *cis*-acting signals required for normal expression of the yp1 gene. As fewer yp1 sequences are present in the construct used in this study than are found in that used by Garabedian *et al.* (1985), this further defined the region necessary for yp1 fat body expression to a DNA fragment 2.2kb in length, between the upstream HindIII site and the XhoI site within the yp1 sequences.

The DNA sequences that signal expression of yp1 in female fat body cells were further localised and defined in another set of experiments performed by Garabedian *et al.* (1986). The authors took their original M13 tagged yp1 gene and created a series of 5′ deletions in the upstream region (Fig. 1.4, constructs d and e), examination of the transcripts from these constructs in transformed flies demonstrated that the sequences between nucleotides -321 and -89 upstream of the yp1 mRNA capping site are sufficient to determine the sex-, stage- and fat body-specific expression of the yp1-M13 gene. A second set of deletions was constructed to further define the essential *cis*-sequences for this pattern of expression. In this, fragments of yp1 DNA were joined to a *Drosophila* heat shock gene (hsp70) fused in frame to the *E. coli* β-galactosidase (lacZ) gene. Expression from these constructs in transformed lines was rapidly assayed by histochemical staining for β-gal activity, and showed that a 125bp fragment of DNA located 196bp upstream of the yp1 mRNA capping site is sufficient to direct the proper sex-, stage- and tissue-specificities of the yp1 fat body expression pattern (Fig. 1.4, constructs f to i), and that this region can confer these transcriptional specificities on a heterologous promoter. The authors went on to establish that the tissue- and sex-specific regulation by the 125bp region does not depend on its orientation (Fig. 1.4,

constructs, i and j), and that when placed 8kb downstream of the hsp70/lacZ fusion gene, it can still operate in the same manner albeit with a weaker influence (Fig. 1.4, constructs k and l).

Overall then, the 125bp region found upstream from the yp1 gene has been shown to have a positive effect on the transcription of nearby sequences. This effect is regardless of the orientation of the region, and is reduced when the region is moved several kilobases from the transcriptional start site and reorientated. In these respects this 125bp region has similar characteristics of some of the viral transcription enhancers that have been described (e.g. Maniatis *et al.*, 1987).

The initial transformation experiments by Garabedian *et al.* (1985) indicated that a 2.8kb DNA fragment including the yp2 gene contains sequences specifying transcription in the ovaries. Further experiments have characterised two ovarian regulatory regions located within the 2.8kb fragment (Logan *et al.*, 1989). One region located upstream of the yp2 cap site (nt -159 → -343) has a positive tissue-specific influence on the ovarian transcription of yp1 and yp2. However, this region alone is not sufficient for high levels of yp1 ovarian transcription. The second regulatory element was identified within the first 105bp of the yp2 gene that is necessary for high and apparently normal levels of (at least) yp1 expression in the ovaries.

A DNA fragment containing the upstream enhancer region (-43 → -343) confers the correct stage and cell specificity of yp ovarian expression on a heterologous promoter (the hsp70/β-gal) fusion. The stage specific effect appears to be influenced by the orientation of the enhancer region. If the region is inverted (at the same or a longer distance from the promoter) its effect on ovarian transcription is seen to vary between transformed lines with different subsets of follicle cells expressing the hsp70/β-gal fusion protein at different times (the pattern is always constant within any one line). The cause of the orientation dependence of the ability of this region to influence transcription of a heterologous gene is not clear, the altered specificities may be due to increased sensitivity to regulatory elements at different chromosomal locations. However, the ovarian enhancer region clearly specifically influences expression in the follicle cells and limits the expression to a portion of the lifetime of those cells. It is possible that the mode of regulation of yp synthesis in ovarian follicle cell is related to the regulatory system proposed to account for the timing of chorion gene expression in the follicle cells (Mariani *et al.*, 1988) where sets of closely linked positive and negative

cis regulatory elements seem to dictate the precise temporal specificity of chorion gene expression.

Shirras and Bownes (1987) fused the 929bp HindIII-AvaII yp1 fragment to the *Drosophila* Adh gene (so that the genes were joined within their 5' non-translated leader sequences)(Fig. 1.4, construct m) and examined the expression pattern in flies transformed with this construct by measuring Adh activity. Only female flies showed any Adh activity and the fat body was the only tissue which stained significantly. The authors used this system to address the question of whether the *cis*-sequences responsible for the sex-limited transcription of yp1 are included in the 929bp fragment of yp1 DNA. To do this the transformed yp1/Adh fusion gene was obtained in a dsx^{D}/dsx genetic background and assayed for Adh activity. Flies mutant in *dsx* were used as the *dsx* gene product is considered to act at the end of the sex-determination pathway. In dsx^{D}/dsx flies (which are pseudomales) the yolk proteins are not synthesised and so if the yp1/Adh gene contains DNA sequences that respond to regulation by the sex-determination genes no Adh activity should be detected (since the mutant *dsx* protein will not be able to interact in a direct or indirect manner with these sequences). This was found to be the case, no Adh activity was detected in the pseudomales carrying the yp1/Adh gene and this experiment therefore shows that the yp1 DNA fragment used in this study contain the site where the *dsx* gene product exerts its effect.

Male *Drosophila* can be induced to synthesise yp transcripts and polypeptides on treatment with 20-hydroxyecdysone (Bownes *et al.*, 1983; Postlethwaite *et al.*, 1980) and in sugar-fed females (Bownes and Blair, 1986). Attempts have been made to determine the *cis*-acting sequences of the yp's that mediate this response, but so far none of the constructs transformed that have been examined (Bownes and Shirras, unpublished results with constructs h, i, a and n in figure 1.4, Postlethwaite, personal communication, construct c in figure 1.4) have given rise to ecdysone induced transcripts in males. Shirras and Bownes (1987) examined Adh activity in male flies transformed with their yp1/Adh fusion gene (Fig. 1.4, construct m). When male flies were injected with 20-hydroxyecdysone, synthesis of the native yolk proteins was induced (as demonstrated by their presence on western blots), but no Adh activity or yp1-Adh fusion transcripts were detectable indicating that the *cis*-elements required for the male response to ecdysone treatment are not present within the 929bp fragment of yp1 used in this study.

Transformation studies of the type discussed above demonstrate *cis*-sequences necessary in the normal expression of the yp genes. Another approach, comparison of the DNA sequences surrounding the genes, can indicate other regions of sequence that may be involved in regulating yp synthesis, but that have not been identified so far by transformation experiments.

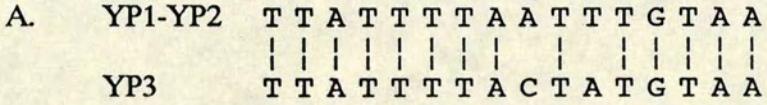
The sequence of the yp3 gene was searched by computer by homologies with the 125bp yp1-yp2-fat body enhancer region (Garabedian *et al.*, 1987). Two regions were found in the yp3 sequence that display a high degree of similarity to different sections of the enhancer (Fig. 1.5). One of these is found upstream of the yp3 gene at nt -909 →-924 and shows 87% homology to the sequence -254 → -269 of the fat body enhancer. The other region of homology is found 457nt downstream of the transcribed sequences of yp3 at nt 2081 → 2107, it shows 70% homology to nt -283 →-309 of the yp1-yp2 fat body enhancer. Whether these sequences actually are involved in regulating yp3 expression in the fat body has not been established.

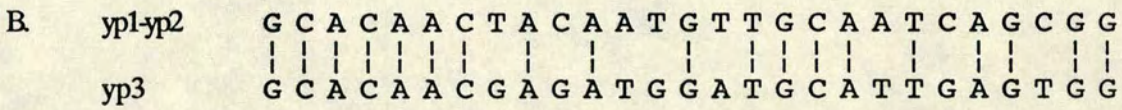
Yan *et al.* (1987) studied matrix plots to search for regions of homology in the 5′ untranscribed regions of the yp1, yp2 and yp3 genes. They discuss several regions showing varying degrees of sequence homology between the three genes regions, however it should be noted that the relevance of this kind of data without supporting experimental evidence is open to some debate. In this case the authors discuss a region which they call the H-box and predict that features of it may be important in controlling aspects of yp expression. This H-box is conserved, repeated and localised in the yp genes (H-box similar sequences occur three times within 300nts upstream of yp3, twice between nt -150 and -190 in yp2, and four times between -100 and -450 in yp1) and shows homology to the binding sites of some vertebrate steroid hormones and to the ecdysone response element of the hsp27 gene of *Drosophila*. However, it seems unlikely that these sequences alone are involved in mediating any yp gene response to ecdysone in the fly since the six copies of this H-box found in the intergenic region of yp1 and yp2 are not sufficient to direct the ecdysone induced expression of yp1 and yp2 in male flies transformed with a 7.6kb DNA fragment consisting of the yp1 and yp2 genes and the entire intergenic region (Fig. 1.4, construct n, Bownes and Shirras, unpublished observation).

Figure 1.5 Homologies between yp3 flanking sequences and the yp1-yp2 fat body enhancer region (taken from Garabedian *et al.*, 1987).

A. -909 → -924 of yp3
-254 → -269 of yp1-yp2

B. +2081 → +2107 of yp3
-283 → -309 of yp1-yp2





trans-Acting Sequences

In the case of the yolk protein genes of *D. melanogaster*, correct expression is known to be influenced by several *trans*-acting components such as the genes of the somatic sex-determination pathway and the insect hormones ecdysone and juvenile hormone (JH). Attempts are being made to identify other *trans*-acting sequences involved in the yolk protein gene expression pattern.

One approach has identified a DNA binding protein (YPF1) which binds specifically to a sequence in the yp1 gene (Mitsis and Wensink, 1989a). YPF1 was found in crude KcO cell extracts using a gel binding assay to detect protein/DNA interactions. The factor binds with high affinity to a 31bp sequence in the translated region of the yp1 gene, beginning at nt -148 downstream of the transcription initiation site. Transformation experiments with M13-tagged yp1 and yp2 genes in which this sequence was deleted showed that the YPF1 binding site in the yp1 gene is necessary for normal steady state *in vivo* levels of yp1 mRNA. YPF1 activity was detected in ovaries (in late stages of oogenesis) and early stages of embryogenesis, but in no other tissues of males, females or larvae. This distribution of YPF1 activity differs from the developmental pattern of yp expression in female fat body cells and follicle cells of mid-stage egg chambers and it is therefore difficult to imagine, as the authors suggest, that the protein could play a positive, transcriptional role in yolk protein synthesis. The protein has been purified from KcO cells and consists of a heterodimer of two sub-units with molecular weights of 85,000 and 69,000Da(Mitsis and Wensink, 1989b). The function of YPF1 has not been established.

The other approach attempting to identify *trans*-factors involved in yp regulation takes the form of a genetic screen for mutations that abolish or greatly reduce yolk protein synthesis in adult female flies. In this, a construction of a fusion gene consisting of 890bp of DNA upstream of the yp1 gene (the 929bp HindIII-AvaII fragment, which is sufficient to confer fat body expression on a heterologous promoter; see Fig. 1.4, construct m) fused to a truncated Adh gene (the construct is fully described in Shirras and Bownes, 1987) was introduced into the germ line by P-element transformation. Expression of the yp1-Adh fusion can be detected in flies of an Adh null background by Adh activity assays and so provides an easily scoreable marker for yp gene expression. This allows a screen to be carried out for EMS-induced recessive mutations that abolish or greatly reduce Adh activity (and YP synthesis) and so lead to the discovery of *trans*-

acting genes regulating yolk protein synthesis. So far, one mutant resulting from such a screen has been described (Shirras and Bownes, 1989). The phenotypic characteristics of this mutant, cricklet (clt) are varied: the flies are defective in yp synthesis, vitellogenesis does not occur, other apparent effects include defective synthesis of larval serum protein 2 and defective histolysis of the larval fat body. These characteristics and the similarities clt mutant flies show to mutants deficient in Juvenile Hormone lead the authors to favour a hypothesis in which the clt encodes a protein essential for mediating the response of the adult tissues to JH.

To date these are the only studies published which attempt to identify and characterise *trans*-sequences of the yolk proteins in *D. melanogaster.*

CONCLUDING REMARKS

The yolk proteins of *D. melanogaster* have proved to be, and continue to be, a good model system for the study of a family of genes that are sex-limited, tissue-specific and developmentally regulated. The abundance of the yolk proteins and their transcripts greatly facilitates the molecular analysis of this co-ordinately expressed gene family. The use of standard genetics, made possible by the vast array of mutants of *D. melanogaster*, and the implementation of more recent developments, particularly the P-element mediated transformation of modified sequences into germline cells, have made it possible to test various hypotheses concerning the regulation of vitellogenesis and yp gene expression.

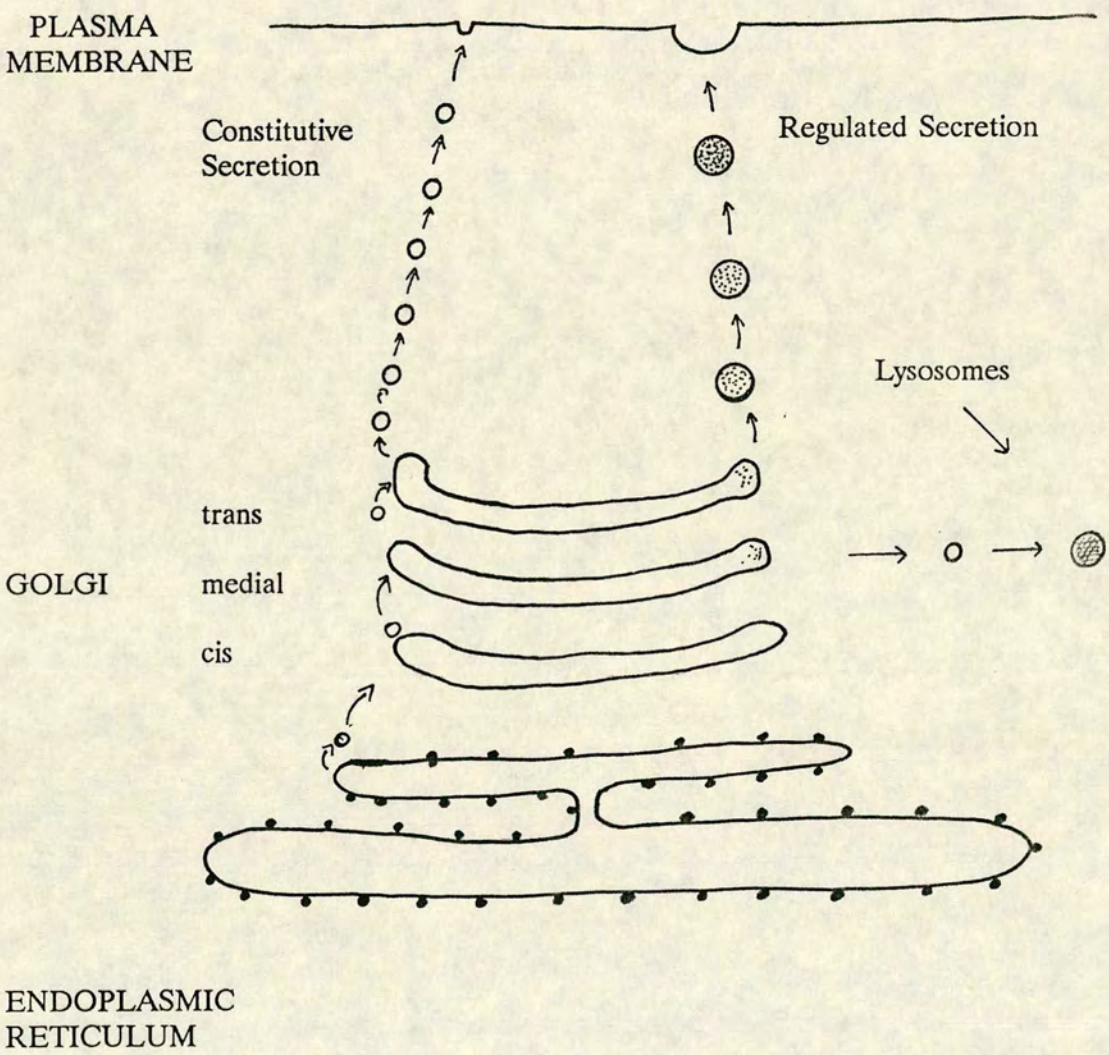
The emphasis which has recently been placed on the understanding of genetic regulation has meant that, for at least the yolk proteins, the characterisation of the polypeptides themselves has been somewhat superficial. However, in general, more research is now being concentrated on protein characterisation in an attempt to assign functional roles to particular gene products and better understand their activities, such as the importance of glycosylation. This should result in a more comprehensive knowledge of genetic regulation.

SECTION 2

PROTEIN SECRETION

Figure 1.6

Generalised diagram of the pathway of protein secretion in a eukaryotic cell (adapted from Verner and Schatz, 1988).



INTRODUCTION

All eukaryotic cells produce proteins that are destined for sites other than the cytoplasm, the compartment in which they are normally synthesised. These proteins can be secreted from the cell, or become incorporated into the lumen of various organelles (such as mitochondria and chloroplasts) or the cellular membranes.

Similarly, bacterial cells produce proteins that become localised in the inner or outer membrane, the periplasmic space between these membranes, or uncommonly, excreted into the growth medium.

In order for these non-cellular or subcellular locations to be realised the proteins must cross at least one membrane barrier. Thus protein targeting to and translocation across membranes is a fundamental cellular process. The processes involved in the movement of proteins to and across various membrane structures are selective and efficient, a particular protein will travel only to its proper location and only those proteins destined for a particular site are transported there.

This section describes basic aspects of protein secretion that are clearly important to yolk proteins since they are amongst the most abundant polypeptides secreted from the *Drosophila* fat body.

A. PATHWAY AND MODELS OF SECRETION

Early work on protein secretion particularly focused on the pathway followed by plasma membrane and secreted proteins from their synthesis to release into the extracellular fluid (Palade, 1975). Proteins that are exported are synthesised on membrane bound ribosomes of the rough endoplasmic reticulum (RER), sequestered into the lumen of the ER, and subsequently transported to the Golgi apparatus and to secretory vesicles from which they are released into the environment surrounding the cell (see Fig. 1.6). During the passage from the ER through the Golgi apparatus and secretory vesicles to the final exit site, many proteins undergo various forms of modification, such as glycosylation and phosphorylation. The specificity of protein transport necessitates mechanisms whereby the cell can distinguish between those proteins which are destined to remain in the cytoplasm, those that are targeted to the Golgi apparatus, lysosymes or other cellular compartments or membranes of the cell, and those that are secreted.

It has long been known that proteins which are exported from bacteria or secreted from eukaryotic cells transiently carry a stretch of amino acids, usually at the amino-terminus. These N-terminal extensions (known as signal sequences, signal peptides, leader sequences or presequences) play a critical role in directing proteins to their target membranes. In 1975 Blobel and Dobberstein proposed a model whereby the signal peptide directs the synthesising ribosome to the membrane of the ER. This is known as the signal hypothesis and is perhaps the most popular model invoked to account for protein secretion. Other models have been proposed, such as the membrane trigger hypothesis (Wickner, 1980), and the loop model (Inouye and Halegoua, 1980).

The essential tenets of the signal hypothesis are that the information for localisation of nascent proteins into the lumen of the ER is encoded by the signal sequence which interacts with various receptors in the cytoplasm and the ER to result in targeting to and translocation across the membrane (Walter *et al.*, 1984; Walter and Lingappa, 1986).

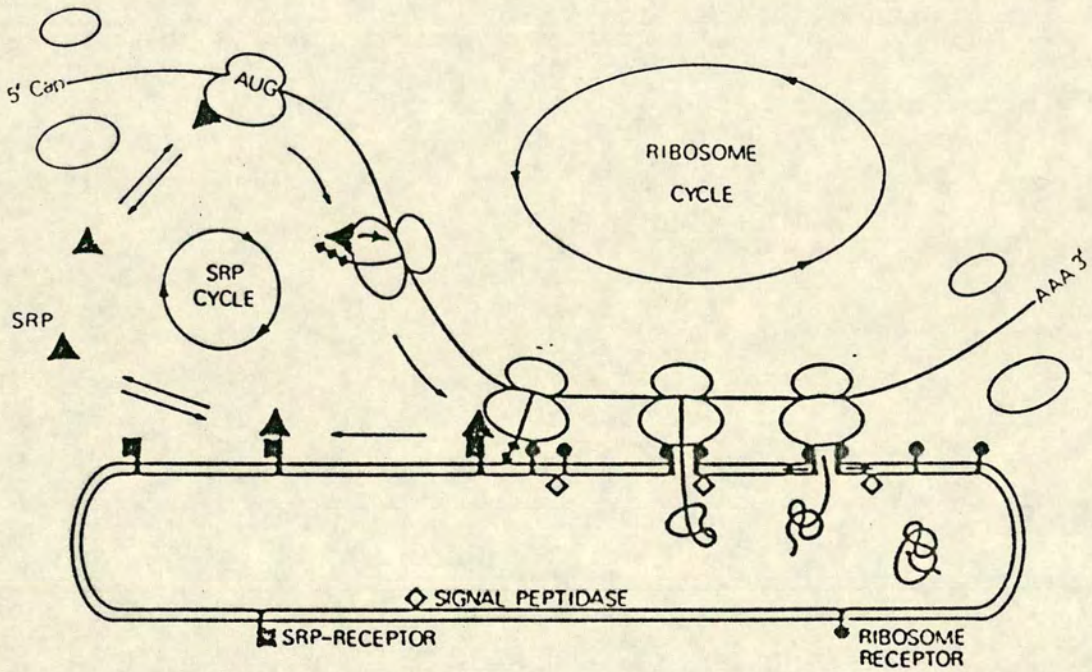
In contrast to the signal hypothesis, the membrane trigger hypothesis emphasises the importance of protein conformation in protein secretion (Wickner, 1980). In this model the leader sequence interacts with the remainder of (or a domain of) the precursor protein to confer a conformation which is competent for partition across the hydrophobic part of the membrane. The membrane trigger model does not require the aid of a proteinaceous translocation apparatus.

Another hypothesis, the loop model of protein secretion proposes that the signal sequences might be inserted as a loop (or helical hairpin) into the membrane of the ER with the N-terminus of the signal sequence remaining on the cytoplasmic side (Inouye and Halegona, 1980). In this model the positively charged residues at the amino-terminus of the leader peptide interact with the negative charges of the phosphatidyl-glycerol head groups of the membrane surface and the nascent chain is threaded through the membrane and appears as a loop on the other side.

No one model seems to completely account for all the data so far accumulated in all of the different protein translocation systems studied. It seems likely that components of different models are necessary to explain the mechanism(s) of the types of export that occur.

Figure 1.7

Diagram showing model for protein targetting to and translocation across the ER membrane (taken from Walter *et al.*, 1984). Most of the key features of the model are discussed in the text, for a more extensive description see Walter *et al.*, 1984.



PROCESSES IN PROTEIN SECRETION

The essential events involved in protein secretion in eukaryotes, with reference to the signal peptide and largely according to the signal hypothesis, can be viewed essentially as three steps: entry into the pathway, where the ribosome complex synthesising the nascent protein is directed to the ER; interaction of the complex with the membrane to form a translocation-competent species; and translocation of the nascent protein across the membrane and its release into the lumen of the ER (see Fig. 1.7).

(i) Targeting, Delivery and Initial Interaction with Membrane

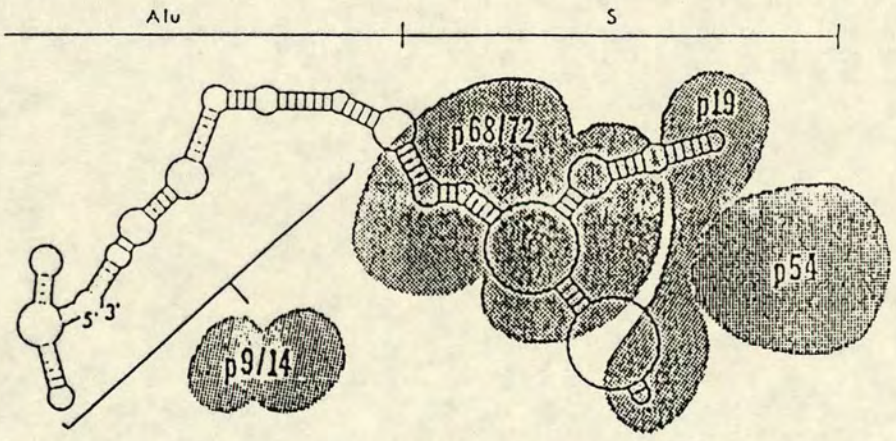
The initial event in protein secretion is the recognition and binding of the signal sequence of a nascent peptide chain as it emerges from the synthesising ribosome by the signal recognition particle (SRP) (Walter *et al.*, 1981). The binding of the SRP to the signal peptide blocks further elongation of the nascent chain (Walter and Blobel, 1981). The mRNA/ribosome/nascent polypeptide complex is delivered to the ER membrane by the SRP, this is mediated by the SRP receptor, an integral membrane protein of the ER (Meyer *et al.*, 1982). The interaction of the SRP with the SRP receptor on the surface of the ER releases the translational arrest, synthesis is resumed, and the SRP is released and is free to begin the cycle again (Walter and Blobel, 1981). It is apparently the SRP-SRP receptor interaction which provides the specificity of the targeting reaction of the ribosome/nascent chain/SRP complex to the correct intracellular membrane (Walter, 1987).

The SRP, purified from the canine pancreas, has been extensively studied (Walter *et al.*, 1984) and fairly well characterised (Siegel and Walter, 1988a). Three distinct activities have been attributed to it:

- recognition of signal peptides as they are extruded from synthesising ribosomes (Walter *et al.*, 1981)
- blocking of further elongation of the presecretory nascent chain (Walter and Blobel, 1981)
- promotion of translocation across the ER membrane by its interaction with the SRP receptor (Walter and Blobel, 1980).

Figure 1.8

Map of the relative positions of the SRP protein domains with respect to the secondary structure of the SRP RNA (taken from Siegel and Walter, 1988b). The Alu and S domains of the SRP RNA, and the activities assigned to the protein domains are indicated. The positions where p68/72 and p19 bind to the RNA have been determined by footprinting analysis; p9/14 has not been precisely determined, but is thought to bind to much of the Alu end of the SRP RNA.



In addition, a recent report (Sanz and Meyer, 1988) indicates that the SRP stabilises the translocation competent conformation of secretory proteins *in vitro*.

The mammalian SRP is a small, elongated, 11S cytoplasmic ribonucleoprotein which is composed of a 300 nucleotide long RNA molecule and six non-identical polypeptide chains organised into four SRP proteins. The proteins consist of two monomers, a 19kDa polypeptide and a 54kDa polypeptide, and two heterodimers, one composed of a 9kDa and a 14kDa polypeptide and the other of a 68kDa and a 72kDa polypeptide (Siegel and Walter, 1985). The RNA molecule (referred to as 7SLRNA and SRP RNA) has two regions of homology to the human Alu right monomer sequence, 100nt at its 5′ end and 45nt at its 3′ end (Ullu *et al.*, 1982). Little is known about the function of the SRP RNA in the SRP other than it seems to have a structural role as the backbone to the particle, holding the proteins in a specific geometry (Walter and Blobel, 1983). Nuclease protection studies have mapped the binding sites of the 19kDa monomer and the 68/72kDa heterodimer on the SRP RNA (see Fig. 1.8) (Siegel and Walter, 1988b). Other functions may be attributable to the RNA, such as involvement in the elongation arrest and the RNA engaging in base pairing interactions with other nucleic acids during the active cycle of the SRP (Walter and Lingappa, 1986), however these are still at the level of speculation.

Several studies have attempted to assign functional roles to the different protein domains of the SRP by creating biochemically the equivalent of mutations in the particle. These have established that signal recognition, elongation arrest and translocation promotion are independent activities (Siegel and Walter, 1988a); the p68/72kDa heterodimer interacts with the SRP receptor to target secretory proteins to the ER membrane and the 54kDa monomer is required for signal recognition.

Until recently no candidate for a yeast SRP-like molecule had been identified, however small ribonucleoproteins similar to the mammalian SRP have now been purified from two strains of yeast (Poritz *et al.*, 1988). Each RNA is of a similar size to the canine SRP RNA and can bind the canine SRP proteins. The protein components of the mammalian SRP recognise homologous elements of the mammalian and yeast RNAs.

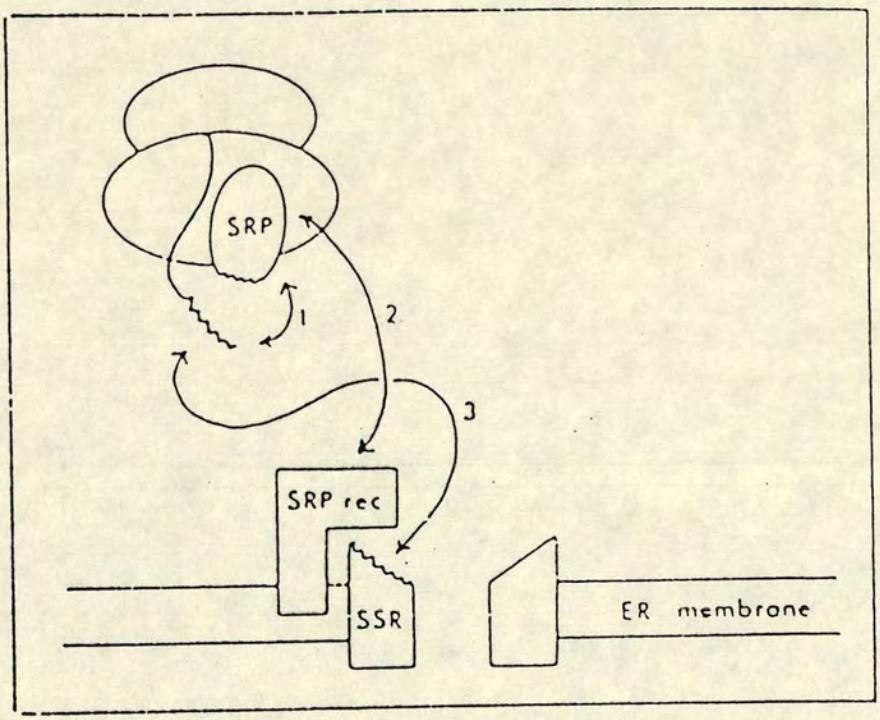
The other fairly well studied component of the eukaryotic translocation system, the SRP receptor, also known as the "docking protein" (Meyer *et al.*, 1982), is an integral membrane protein of the ER with a mass of 72kDa (Gilmore *et al.*, 1982a, 1982b). One function of the SRP receptor appears to be to relieve the SRP-imposed

Figure 1.9

Diagram illustrating the sequence of information flow during targeting to the mammalian ER membrane:

1. signal sequence interactions with the SRP
2. SRP interacts with SRP receptor
3. signal sequence and SSR interact.

(Taken from Walter, 1987).



translational arrest (Walter and Blobel, 1981), however Siegel and Walter (1985) have proposed that its primary function may be to direct the SRP bound nascent polypeptide complex to the ER where the ribosome can be bound to the membrane to allow translocation to occur.

The primary structure of the SRP receptor has been determined and a model of its disposition in the ER membrane developed (Lauffer *et al.*, 1985). The amino-terminus, which appears to have an uncleaved signal sequence, is the membrane region with two hydrophobic stretches, that may insert into the lipid bilayer as a double "helical hairpin". The cytoplasmic domain contains regions with a high proportion of charged residues and shows homology to nucleic acid binding proteins. It is possible that this domain interacts directly with the 7SLRNA of the SRP (Lauffer *et al.*, 1985).

(ii) **Interaction with the ER Membrane**

The steps following the initial targeting events described above are still poorly understood. Once targeting has taken place the ribosome complex /SRP/SRP receptor interaction is likely to be replaced by a direct interaction of the ribosome complex with the surface of the ER membrane. It has been postulated that this interaction is mediated by two other integral RER membrane glycoproteins, ribophorins I and II (Briggs and Gierasch, 1986). Binding of ribosomes by microsomal subfractions correlates with the presence of these proteins and they can be chemically cross-linked to ribosomes. Microsomal subfractions bearing the ribophorins can be stripped of the native ribosomes and can then bind other ribosomes which need not be carrying a nascent secretory peptide, indicating that the signal sequence is not involved in any interaction of the ribosomes with the ribophorins (or other membrane proteins).

The release of the SRP from the ribosome is catalysed by the SRP receptor in a process that requires GTP (Connolly and Gilmore, 1989). The SRP and SRP receptor detach from the ribosome complex (Gilmore and Blobel, 1983) and are free to recycle, thus both molecules are thought to act catalytically in the targeting process (Walter *et al.*, 1984).

After its release from the SRP and without the need for elongation of the nascent chain, the signal peptide interacts with another membrane protein, the signal sequence receptor (SSR) (Fig. 1.9) (Wiedmann *et al.*, 1987). The SSR is a glycosylated protein of the rough ER that spans the lipid bilayer and has a mass of ~35kDa. A

glycosylated membrane spanning protein of ~34kDa has recently been purified from canine microsomes (Hartmann *et al.*, 1989). It has not yet been established whether this protein is identical with the SSR, however it may be essential for protein translocation across the ER membrane. The existence of an SSR in the ER which recognises signal peptides during the normal translocation process does not necessarily indicate that the SRP is functionally redundant and it has been proposed (Walter, 1987) that a dual recognition system may result in a higher specificity and selectivity in the recognition event (if the binding of signal peptides by the SRP and SSR is distinguished by emphasising different aspects or subtleties of their structure). Additionally, the SRP can be thought of as an adapter molecule between the cytoplasmic translocation and the membrane bound translation machines (Walter, 1987).

(iii) Translocation across the Membrane and Release into the Lumen of the ER

Subsequent steps in the pathway of translocation are unclear. Wiedmann *et al.* (1987) have proposed that the SSR may mediate the interaction of the signal peptide with the membrane lipids or with other proteins that may constitute a "translocation pore".

One of the most controversial areas in protein export is the nature of the translocation site. The signal hypothesis favours a mechanism in which the exported protein crosses the membrane through a proteinaceous pore. Experiments which tested the accessibility of the nascent chain to aqueous solutes (Gilmore and Blobel, 1985) indicated that integral membrane proteins are not only involved in the initial attachment, but also in the translocation of the exported protein through the membrane. Although Gilmore and Blobel (1985) suggest that the nascent chain does not interact with the membrane lipids, there is some indirect evidence in prokaryotes that the exported proteins may contact the lipids during translocation (Briggs and Gierasch, 1986). Also, the degree of interaction of synthetic signal sequences with lipids *in vitro* has been found to correlate with their activity *in vivo* (Briggs and Gierasch, 1986). It is possible that the translocation site consists of both proteins and the lipids within the membrane.

Early studies had indicated that translocation across the ER membrane was obligatorily cotranslational (i.e. that translocation occurred only while the polypeptide chain was being actively synthesised by the ribosome) (Blobel and Dobberstein, 1975). However, a fairly large body of evidence now indicates that this is not the case and it

is possible for full length, newly synthesised polypeptides to be transported across membranes. *In vitro* experiments in *Saccharomyces cerevisiae* have demonstrated that the translocation of some secretory proteins can occur post-translationally, examples include prepro-α-factor (Waters and Blobel, 1986) and prepro-carboxypeptidase Y (Hansen and Walter, 1988). In *E. coli* there is evidence that bacterial proteins can be secreted post-translationally, both *in vivo* and *in vitro* (Verner and Schatz, 1988). Import into cellular organelles such as mitochondria and chloroplasts can also occur post-translationally (Verner and Schatz, 1988). Finally, an SRP subparticle lacking the protein and RNA domains required for translational arrest was found to still be competent in promoting secretory protein translocation *in vitro* (Siegel and Walter, 1986).

Whether a post-translational or a co-translational mode of secretion is utilised *in vivo*, or whether both are used in different instances perhaps depending upon the particular protein, remains unresolved. Silhavy *et al.* (1983) suggested that the coupling between translation and translocation is not as tight for prokaryotes as it is for eukaryotic secretion, since no SRP-like translation arrest activity has been noted for bacterial systems.

There are in principle several modes by which a polypeptide chain could be translocated across a membrane (Singer *et al.*, 1987):

- the polypeptide could be continuously threaded through the membrane
- the entire polypeptide could be translocated at once
- successive segments (or domains) could be translocated as units one after another.

Various lines of evidence indicate that complete polypeptides are not translocated all at once. Tightly folded protein precursors cannot be efficiently translocated across any membrane system studied (Meyer, 1988; and see p.36). Also, experiments have shown that modifications which occur during or after translocation across the ER membrane (such as N-linked glycosylation) can take place at the amino-terminal end of the nascent polypeptide chains before synthesis of the chain is completed (Walter *et al.*, 1984). This indicates that polypeptides do not cross the membrane all at once and supports the idea that proteins traverse the membrane in an extended conformation. All of this data is inconsistent with the idea that any major degree of protein folding of the precursor need occur on the cytoplasmic side of the membrane.

When translocation is complete and the leader peptide has been cleaved, the nascent polypeptide is released into the lumen of the ER. This is accompanied by the folding of the protein into its mature conformation. The permeability barrier of the membrane is restored and the ribosome is released (it is now free to re-enter the pool of ribosomes in the cytosol).

MODIFICATION OF PROTEINS IN THE ER

Various modifications of nascent polypeptides are known to occur on the luminal side of the ER. Amongst these are the cleavage of the signal peptide from the remainder of the nascent chain by signal peptidase, and the core glycosylation of certain asparagine residues by oligosaccharyl transferase.

Signal Peptide Cleavage

The signal peptides are cleaved by proteases that are integral membrane proteins sometime after the nascent peptide has begun to be extruded through the membrane (Gilmore and Blobel, 1985).

E. coli has two different leader peptidases, one of these, the product of the lsp gene, signal peptidase II, has been found to cleave only the prolipoprotein signal sequence. The "general" signal peptidase in *E. coli*, signal peptidase I, is encoded by the *lep* gene and has a molecular mass of 36kD*a*. The cleavage event appears to take place on the outside of the inner membrane and signal peptidase I is thought to be orientated such that the bulk of the protein is at the outer surface of this membrane. *E. coli* possesses a signal peptide peptidase activity (Briggs and Gierasch, 1986) which is also found on the inner membrane and degrades the signal peptide only after it is cleaved from the mature protein (this activity was purified and identified as protease IV) (Briggs and Gierasch, 1986).

Signal peptidase has also been isolated from dogs, chickens, and yeast (Verner and Schatz, 1988). The canine signal peptidase was purified from pancreas microsomes as a complex of six polypeptides, present in stoichiometric amounts to one another, two of which are glycoproteins (the two glycoprotein subunits are differently glycosylated forms of the same protein, the cDNA encoding this has been cloned and sequenced (Shelness *et al.*, 1988)). It is conceivable that the signal peptidase complex marks the

translocation site in the ER and may constitute all or part of the postulated proteinaceous pore (Blobel, 1987).

The yeast signal peptidase has a molecular weight of 18.8kDa and is thought to be the product of the *sec*11 gene (although this is still tentative) (Böhni *et al.*, 1988). It is not yet known whether yeast signal peptidase is a single or multisubunit enzyme (Deshaies *et al.*, 1989).

Although the prokaryotic leader peptidases and the eukaryotic signal peptidases display structural differences, the substrate specificities are very similar since either enzyme can cleave prokaryotic or eukaryotic substrates (Deshaies *et al.*, 1989). This may indicate that some of the other proteins associated with the vertebrate signal peptidases may perform other functions such as increasing the efficiency or specificity of cleavage, or other aspects of translocation such as formation of the translocation site in the membrane.

Asparagine-Linked Glycosylation

The ER is the site where asparagine-linked (N-linked) glycosylation occurs in which core oligosaccharides are transferred from a lipid donor to certain asparagine residues on the nascent polypeptides. The sequence of these sites on both yeast and mammalian polypeptides is Asn-X-Ser/Thr, where X can be any residue except proline. Subsequently, some of the core residues are trimmed by other ER membrane residing enzymes. Core glycosylation and trimming often occur before all of the polypeptide chain has been translocated across the membrane and even before its synthesis is complete (Rose and Doms, 1988).

In addition to signal peptide cleavage and N-linked glycosylation, polypeptides translocated into the lumen of the ER are subject to other post-translational covalent modifications such as O-linked glycosylation and disulphide bond formation.

CONFORMATION OF PREPROTEINS

Many studies have recently demonstrated that in order to be efficiently translocated across a membrane protein precursors must have the proper conformation. Most evidence indicates that this is an unfolded conformation which is correlated with an "openness or looseness" of structure characterised by an increased sensitivity to attack

by proteases (Meyer, 1988). That preproteins cannot be translocated in a tightly folded form has been demonstrated for transport across the bacterial plasma membrane, the ER membrane, and for the import of nuclear encoded proteins into mitochondria (Meyer, 1988).

A translocation competent conformation could be achieved in a number of ways (Verner and Schatz, 1988):

- the presequence itself may interfere with the tight folding of the precursor
- the nascent precursor may interact with the translocation machinery, or other proteins, to prevent tight folding
- the nascent polypeptide may be partially unfolded in the cytosol or at the target membrane by an energy-dependent process.

Factors Involved in Preprotein Conformation

Studies including the use of genetics and refined *in vitro* assays in which translocation can be uncoupled from translation, have identified a number of proteins which are involved in promoting the translocation of preproteins across membranes, some of these appear to play a role in maintaining a translocation competent conformation (Deshaies *et al.*, 1988a, 1989; Verner and Schatz, 1988).

In *E. coli*, genetic experiments have identified many genes in which mutations cause defective protein export (Benson *et al.*, 1985; Oliver, 1985). Some of the products of these genes have been fairly well studied.

In vitro translocation experiments have identified a cytosolic factor which is involved in protein export in *E. coli*. The active factor was purified from a lysate fraction competent in promoting translocation of proOmpA (the precursor of the *E. coli* outer membrane protein A) (Crooke *et al.*, 1988; Crooke and Wickner, 1987). This factor is known as trigger factor and is a single polypeptide with a molecular mass of ~63kDa. Further investigations indicated that trigger factor acts to stabilise a translocation competent formation (Crooke *et al.*, 1988), at least for some preproteins, and it has been speculated that a precursor/trigger factor complex forms co-translationally (Meyer, 1988).

Studies *in vivo* have provided evidence that *E. coli* has another factor, the product of the *secB* locus, which seems to function in an analogous manner to trigger factor, but for a different set of preproteins (Meyer, 1988). Experiments utilising

preMBP (maltose binding protein, an *E. coli* periplasmic protein) demonstrated that the SecB protein is capable of maintaining the translocation competent formation and protease sensitive state of preMBP and showed that the SecB protein has to be present cotranslationally in order to be able to do this (Collier *et al.*, 1988).

Another approach, one of photo-crosslinking, has been used to identify a further cytosolic factor in *E. coli* which was inferred to maintain the unfolded, translocation competent state of preproteins. The factor, a heat shock protein, GroEL, was shown to form a complex with unfolded, newly-synthesised pre-β-lactamase in *E. coli* (Bochareva *et al.*, 1988).

A number of genes that are required at various stages in protein secretion have been identified in yeast (Deshaies *et al.*, 1989). The involvement of the 70kDa heat-shock like proteins in yeast protein translocation has been investigated by genetical and biochemical approaches. The translocation of prepro-α-factor (the precursor form of the yeast mating factor) *in vitro*, across yeast microsomal membranes, is stimulated by the addition of fractions of yeast cytosol. This observation allowed the purification of a 70kDa heat shock protein as an active factor in promoting translocation (Deshaies *et al.*, 1988a) (other proteins were also noted, but were not identified). Also, experiments using yeast strains with mutations in the genes coding for a subset of the 70kDa heatshock-like protein family in yeast (*SSA1, 2, 3* and *4*) have allowed the products of the *SSA1* and *SSA2* genes to be implicated in the translocation of presursor proteins (Deshaies *et al.*, 1988b). Mutant yeast cells that are depleted in this subset of the hps70 polypeptides were shown to accumulate precursor forms of proteins normally destined for import into the ER and mitochondria. A clue to the mechanism of function of these proteins in promoting the translocation of precursor proteins across membranes may be gained from work showing that their effect can be transiently mimicked *in vitro* by the protein denaturant urea (Chirico *et al.*, 1988) indicating that precursor proteins must be unfolded to be translocated, so perhaps the 70kDa heatshock-like proteins use the energy derived from ATP hydrolysis to unfold the precursor molecules, or to inhibit folding of nascent chains into a translocation incompetent conformation.

Note that the hydrolysis of ATP is necessary for post-translational translocation of proteins across all known membrane systems (Verner and Schatz, 1988). Also worthy of mention is the observation that cytosolic factors from yeast and *E. coli* can substitute

for one another in reconstituted systems (Weng *et al.*, 1988; Müller and Blobel, 1984) indicating that at least some of the factors in the eukaryotic and prokaryotic translocation systems are functionally similar.

The correct conformation of preproteins seems to be a requirement for their translocation across mammalian microsomes. It is likely that the majority of proteins are translocated across the ER membrane cotranslationally as a result of the elongation arrest activity of the SRP. This may function to retain the length of time the nascent polypeptide remains translocation-competent as the preprotein would not have the opportunity to adopt a highly folded conformation (Walter, 1987). However, the translocation competence of preproteins in a mammalian system has been addressed. This was performed using the bacteriophage M13 procoat protein which can be translocated post-translationally across pancreatic microsomal membranes. In this system the addition of purified mammalian hsp70 was found to increase the level of translocation of the procoat protein. In addition, an ATPase activity was found that is a translocation-stimulatory component of reticulocyte lysate (Meyer, 1988).

The observation of the ability of hsp70 to promote translocation in a mammalian system, taken together with the evidence that the *E. coli* hsp GroEL and some of the yeast hsp70-like family can act in a similar manner, perhaps illustrates the conservation of the role of these particular hsp's in protein translocation across membranes. It has been proposed that the preservation or alteration of the conformation of precursor proteins may be one of the functions of the proteins of the widely distributed and well conserved heatshock protein family (Meyer, 1988).

It is possible that in addition to the hsp's playing such a role in protein translocation, the mammalian SRP is involved in a similar way. The SRP has been shown *in vitro* to stabilise the bacterial proOMP-A (outer membrane protein A) for translocation across *E. coli* inner membrane vesicles, yeast microsomes and pancreatic microsomes (Crooke *et al.*, 1988; Sanz and Meyer, 1988). This indicates that SRP has the ability to prevent proteins from folding into a translocation incomplete transformation.

CONCLUDING REMARKS

In this section an overview of the important features of protein secretion has been presented. The level of understanding of the pathway of secretion (and

translocation across cellular membranes in general) has increased dramatically in the last few years. The elucidation of details of some of the components involved (and the characteristics of presecretory proteins important for secretion) has resulted from combining biochemical, molecular biological and genetical approaches. The use of genetic screens and selections have allowed the identification of a large collection of genes required for the transport and processing of secretory proteins in yeast (reviewed: Schekman, 1985; Deshaies *et al.*, 1989) and bacteria (reviewed: Benson *et al.*, 1985; Oliver, 1985). The involvement of genetically engineered secretory substrates and increasingly refined *in vitro* translation and translocation assays has lead to the identification and purification of various components active in the processes of protein secretion (discussed: Walter *et al.*, 1984; Meyer, 1988; Verner and Schatz, 1988).

In the following section the characteristics and role of the signal sequence are discussed.

B. SIGNAL SEQUENCES

INTRODUCTION

The signal sequence plays a central role in the process of protein secretion. The leader peptide is a stretch of amino acids that is normally found at the amino-terminus of nascent secretory proteins and is usually cleaved from the mature portion of the protein during translocation across the membrane.

A few secreted proteins have internal signal sequences which are required for export, but are not cleaved, an example is the signal sequence of ovalbumin which has been localised to amino-acids 22-41 (Briggs and Gierasch, 1986). Some membrane proteins have internal, uncleaved signal sequences, this is the case for the *E. coli* signal peptidase where the signal sequence has been identified as residues 62-76 (von Heijne *et al.*, 1988). Some other membrane proteins have N-terminal signal sequences which act as membrane anchors, such as the *Rhodopseudomonas viridis* reaction centre H-subunit (von Heijne, 1986a). In these cases the signal sequence acts to direct the nascent protein to the membrane and to halt translocation so that the anchored polypeptide is assembled in the membrane.

Targeting signals exist to direct newly synthesised proteins to particular locations other than the ER membrane for secretion or integration into the cell membrane.

Some of these are found at the N-terminus of the polypeptide, such as the mitochondrial targeting sequences, whereas others are found at the carboxy-terminus, such as the peroxisomal targeting signals (Verner and Schatz, 1988).

The requirement for the signal sequence in protein secretion cannot be overstressed. In the absence of a functional signal sequence secretion does not occur, as was clearly demonstrated with yeast invertase. This enzyme has a cytoplasmic and a secreted form, and the mRNAs encoding the alternative forms differ only in the presence (in the secreted form) or absence (in the cytoplasmic form) of the sequence coding for the signal peptide (Schauer *et al.* 1988).

INTERCHANGEABILITY OF SIGNAL SEQUENCES

Leader peptides display no strong sequence homology, so much so that even the signal sequences of otherwise highly homologous proteins bear little similarity to one another at the primary sequence level. For instance, although the *E. coli* proteins OmpF and PhoE are highly similar, their leader peptides show little similarity to one another. However, the general pattern to which they all conform results in signal sequences being almost entirely interchangeable from one protein to another, even amongst widely different organisms (Briggs and Gierasch, 1986).

Genes encoding secretory proteins from sources as diverse as *E. coli*, plants and insects when injected into *Xenopus laevis* oocytes, can be correctly expressed and the products properly secreted. Similarly, both the bacterium *E. coli* and the *in vitro* translocation system derived from the dog pancreas are able to properly translocate and cleave secretory proteins from a wide range of species.

Experiments using gene fusions have also served to demonstrate the interchangeability of signal sequences. The sequences encoding the leader peptide and the first five amino acids of the mature domain of *E. coli* β-lactamase fused to the gene for globin (normally a cytosolic protein) cause the resulting fusion protein to be transported across the *E. coli* inner membrane, or translocated across the membranes of the dog pancreas microsome system (Lingappa *et al.*, 1984). In another set of experiments, the sequences coding for chimp α-globin were fused to the 5′ end of the gene encoding bovine preprolactin (a secretory protein that has an N-terminal signal sequence), so that the resulting fusion protein contained the preprolactin signal peptide at an internal position (Perera and Lingappa, 1985). This internal signal sequence

directed the translocation of both proteins into the vesicle lumen and the signal sequence was cleaved by signal peptidase.

The phenomenon of interchangeability of signal sequences shows that the secretion apparatus in one organism can recognise (and often correctly process) the signal peptide from virtually any other organism. This is regardless of the diversity of the organisms, and therefore indicates that some aspects of the secretion pathway have been conserved during evolution.

CHARACTERISTICS OF SIGNAL SEQUENCES

Determination of the sequence of a large number of signal peptides from both eukaryotes and prokaryotes has allowed the comparison and analysis of these sequences (von Heijne, 1983, 1985; Perlman and Halvorsen, 1983).

All signal sequences appear to have evolved along the same general lines with three distinctive regions discernible: a positively charged N-terminal region (the n-region), a central hydrophobic segment (the h-region), and a more polar carboxy-terminal region that defines the cleavage site (the c-region). The overall length of signal peptides directed to the ER membrane varies between 13 and 36 amino-acids (Perlman and Halvorsen, 1983; von Heijne, 1985).

(i) The n-region

The amino-terminal region of the signal sequence usually has at least one charged residue, commonly lysine and/or arginine. Prokaryotic leader peptides often have basic amino acids in this region (von Heijne, 1984a; Perlman and Halvorsen, 1983). The n-region is typically 1-5 residues long and the length varies strongly with the overall signal sequence length (von Heijne, 1985).

Statistical analysis (von Heijne, 1984a) of the n-region demonstrates that prokaryotes and eukaryotes display a similar pattern of amino acids in this region in terms of net charge and position of charged residues, indicating that the n-region is under similar selective pressure in eukaryotes and prokaryotes. It has been proposed that the charged residues are involved in the initial interaction of the signal sequence with the negatively charged head groups of the lipids at the surface of the membrane (von Heijne, 1984a).

The function and importance of the charge of the n-region has been investigated.

The *E. coli* outer membrane lipoprotein (lpp) normally has two positively charged lysine residues in the n-region of its signal sequence (resulting in a net charge of +2). When these residues were removed and replaced with neutral or acidic amino acids both the secretion and the translation of lpp were affected (Inouye *et al.*, 1982). Those mutants of the n-region with a net charge of +1 or 0 showed reduced synthesis of lpp and the protein was transported and secreted at the same rate as wild type lpp. Mutants with a net negative charge showed a reduction in lpp synthesis but also displayed very slow rates of secretion and processing so that most of the protein was found accumulated in the cytoplasm in the precursor form. These experiments clearly demonstrate that, at least for *E. coli* lpp, the presence of the net positive charge (or the absence of net negative charge) is important both for the proper secretion and the level of translation of the protein.

The charge of the n-region has been implicated in the level of transcription of another bacterial protein. A mutant was isolated where the positively charged arginine residue of the n-region of the LamB signal sequence was altered to a serine. This mutation was shown to have the result of reducing the synthesis of the protein (although its secretion was normal), indicating that the charged residues of the n-region may be functioning to couple protein synthesis with export (Briggs and Gierasch, 1986).

Other studies, however, have indicated that mutations in the charged n-region of signal peptides have no effect on translocation.

The signal peptide and the first nine amino acids of *E. coli* lipoprotein (OMLP) was shown to be capable of promoting translocation of β-lactamase across mammalian microsomes *in vitro* with as high an efficiency as seen for mammalian secretory proteins. Various mutants in the n-region of the signal sequence of pre-OMLP/β-lactamase fusions were assayed for their ability to promote translocation (Garcia *et al.*, 1987). None of those mutants with net n-region charges of +1, 0, or -1 were found to have any effect on the translocation of the fusion protein across mammalian microsomal membranes or on the ability of the SRP to recognise the leader peptide. Although some of the mutants assayed in this study affect the kinetics of secretion in *E. coli in vivo* (as described above) (Inouye *et al.*, 1982), these results argue that the net positive charge of the amino-terminus of the signal sequence is not a requirement of translocation, nor does it increase the efficiency of the process *in vitro*.

Mutations at the amino-terminus of the signal sequence of yeast invertase were examined (see Table 1.1, such as substitutions of charged and polar residues, e.g. *suc2-301* and *suc2-331*, or the addition of a positive charge with the deletion of two hydrophobic residues, e.g. *suc2-236*) (Kaiser and Botstein, 1986). Most of the mutations limited to the n-region of the signal sequence showed normal levels of expression and did not prevent normal function of the leader peptide, suggesting that the amino-terminal region of the signal sequence is not specifically recognised by secretory factors (this is supported by the study mentioned above, by Garcia *et al.*, 1987).

Some eukaryotic signal sequences have been found which do not conform to the "consensus" of a positively charged n-region (von Heijne, 1984a) further indicating that the requirement for a positive charge in this region is not absolute. It is possible that in eukaryotes, translation and translocation are more tightly coupled than in prokaryotes due to the SRP/SRP receptor interaction, and that at least the mammalian translocation machinery therefore overcomes the requirement for a net positive charge of the amino-terminus of the leader peptide.

The importance of the conformation of the n-region to signal peptide function has also been investigated. Brown *et al.* (1984) constructed a set of mutations in the yeast invertase signal sequence that are predicted to cause significant alterations in the secondary structure of the n-region. The introduction of various amino acid insertions 3-4 residues long were found to have no effect on the secretion of the invertase. This study indicates that quite severe alterations at the amino-terminus of the signal sequence can be accommodated and have no effect on signal sequence function.

Overall, it is clear that while the n-region of signal sequences usually displays a general pattern of amino-acids, the constraints imposing this are not inflexible since examples of signal sequences have been found where the n-region does not entirely conform to the standard pattern, also it is possible to tamper with the "consensus" sequence quite severely without dramatically affecting the function of the signal peptide.

(ii) The h-region

The most distinctive region of the signal sequence is the stretch of uncharged hydrophobic residues known as the h-region or hydrophobic core. This follows the charged n-region and precedes the more polar c-region.

The h-region is rich in the hydrophobic residues Phe, Ile, Leu, Met, Val and Trp

Table 1.1

The amino acid sequence of the wild type (wt) yeast invertase signal sequence and a number of mutations in this region that have been investigated. The effect (if any) of these mutations is discussed in the text.

Amino acid substitutions are underlined and deletions are represented by dashed lines.

⟶ = normal signal peptidase cleavage site.

→ = site where cleavage occurs in the mutant *suc2-s1*.

Table 1.1

																		-1 ↓	+1		Reference
Wt	Met	Leu	Leu	Gln	Ala	Phe	Leu	Phe	Leu	Leu	Ala	Gly	Phe	Ala	Ala	Lys	Ile	Ser	Ala	Ser	Met	
suc2-s1	Met	Leu	Leu	Gln	Ala	Phe	Leu	Phe	Leu	Leu	Ala	Gly	Phe	Ala	Ala	Lys	Ile	Ser	Val	Ser	↓ Met	Schauer et al. (1985)
suc2-301	Met	Leu	Leu	Arg	Ser	Thr	---	Phe	Leu	Leu	Ala	Gly	Phe	Ala	Ala	Lys	Ile	Ser	Ala	Ser	Met	Kaiser & Botstein (1986)
suc2-331	Met	Leu	Leu	Arg	Ser	Thr	Leu	Phe	Leu	Leu	Ala	Gly	Phe	Ala	Ala	Lys	Ile	Ser	Ala	Ser	Met	" " "
suc2-236	Met	Leu	Leu	Arg	---	---	Leu	Phe	Leu	Leu	Ala	Gly	Phe	Ala	Ala	Lys	Ile	Ser	Ala	Ser	Met	" " "
suc2-211	Met	Leu	Leu	Arg	---	---	---	---	Leu	Leu	Ala	Gly	Phe	Ala	Ala	Lys	Ile	Ser	Ala	Ser	Met	" " "
suc2-437	Met	Leu	Leu	---	---	---	---	---	---	---	---	---	---	---	Ala	Lys	Ile	Ser	Ala	Ser	Met	" " "
D7	Met	Leu	Leu	Gln	Ala	Pro	Arg	Gly	---	---	---	Gly	Phe	Ala	Ala	Lys	Ile	Ser	Ala	Ser	Met	Perlman et al. (1986)

(von Heinje, 1985; Perlman and Halvorsen, 1983). The charged and the polar residues - Asp, Gln, Arg, Lys, His, Gly, Pro, Glu, Asn, Ser, Thr, Tyr - are virtually absent from the h-region, although they tend to dominate the c-region. Von Heinje (1982) noted that the more hydrophobic residues tend to be concentrated around the mid-points of the h-region of secretory proteins, other than this he notes no convincing patterns or consensus sequences apparent in the distribution of amino acids in the hydrophobic core (von Heijne, 1985). However, others (Perlman and Halvorsen, 1983) suggest that a non-random fine structure exists in this area.

The hydrophobic core is variable in length, ranging from 7-20 amino acids with most signal sequences having an h-region of 10-15 residues (von Heijne, 1985). Eukaryotic signal sequences generally have more hydrophobic h-regions (more Leu and less Ala) that are generally shorter than those of bacteria (von Heijne and Abrahamsén, 1989).

It has been proposed that the h-region partitions into the membrane and spans the lipid bilayer (von Heijne, 1985). The length variations observed for signal sequences may indicate that the hydrophobic core achieves this as a structure composed partly of α-helix, partly extended chain, depending on its length (von Heijne, 1985). The membrane interior is 25-30Å, an extended chain of 8 amino acids and a 20 residue long helix are both estimated to be around 27Å in length.

The maximal length of the h-region may function as a label that distinguishes between the signal sequences of exported proteins and the transmembrane segments of membrane proteins. The average h-region of one class of N-terminal anchor segment of membrane proteins is generally a little more hydrophobic, more evenly hydrophobic (von Heijne, 1982), and longer (usually around 25 residues) than that of the average signal peptide of secretory proteins. Thus some N-terminal anchor sequences are like signal sequences with an extra long h-region (von Heijne, 1986).

A stretch of 8 hydrophobic residues are thought to be the minimum capable of forming an α-helix or β-strand structure (Perlman and Halvorsen, 1983), and this minimum length of 8 amino acids is a conserved feature of the h-region of all the signal peptides studied. von Heijne (1981) suggested that sufficient free energy must be released in the signal peptide/membrane interaction for successful secretion to occur. It is possible that the free energy liberated when 8 hydrophobic amino acids partition across the membrane may represent the functional lower limit for secretion to occur. Another possibility is that a secondary structure, such as an α-helix or a β-strand, must

be of sufficient length to allow binding of the signal sequence to the SRP and the minimum length of 8 hydrophobic residues represents the lower limit of this (Perlman *et al.*, 1986).

Mutations of the h-region have illustrated how crucial an intact hydrophobic core is to the function of the signal sequence.

Only one eukaryotic signal peptide, that of the secreted form of yeast invertase, has been studied in any great detail with the aim of demonstrating its role in export and improving the understanding of the mechanism of its action.

A number of mutations of the invertase signal sequence have been constructed *in vitro*, many of these are deletion and substitution mutations within or around the region of the hydrophobic core (Perlman *et al.*, 1986; Kaiser and Botstein, 1986). These were reintroduced into yeast, either as autonomously replicating plasmids or were integrated into the chromosomal locus, and analysed. The 19 amino acid long signal sequence of yeast invertase normally has a hydrophobic core of 11 contiguous, non-polar residues (Table 1.1). The studies mentioned here conclude that a stretch of 8 (Perlman *et al.*, 1986) or 9 (mutant *suc2-236*, Table 1.1) hydrophobic residues is sufficient for secretion of invertase. A 7 residue long h-region (mutant *suc2-211*, fig 1ss1) allows retention of partial function of membrane translocation, but with a 5 residue (*D7*, Table 1.1) (or less, *suc2-437*, Table 1.1) long hydrophobic core the resultant enzyme remains uncleaved, unglycosylated and cytoplasmic (or only weakly attached to the ER membrane).

A recent study supports the data accumulated so far for the h-region of the yeast invertase signal sequence. Revertants were obtained of secretion defective variants which were originally created by fusing random fragments to the 5′ end of the invertase gene (Preuss and Botstein, 1989). Most of the revertants characterised were found to be point mutations and duplications in the upstream region of the gene, each of which introduced hydrophobic residues into the non-functional leader sequence, suggesting that the defective random leader peptides might simply lack adequate hydrophobicity to be effective signal sequences. It is worth noting that the revertants were capable of targeting the protein to the ER membrane but were not cleaved like wild type signal sequences and some of the fusion proteins accumulated at intermediate stages of the secretory pathway.

In *E. coli* more detailed information regarding the importance of the h-region in protein export has been obtained. Detailed studies of mutants of the *E. coli* outer

membrane bacteriophage λ receptor protein (LamB) and the periplasmic maltose binding protein (MBP), have demonstrated that disruption of the hydrophobic core usually results in defective secretion of the protein and accumulation of the protein precursors within the cytoplasm and cytoplasmic membrane.

Each of five unique mutational alterations of the *MalE* signal sequence that cause defective secretion of MBP are located within the hydrophobic core (Bedouelle *et al.*, 1980). These mutants (denoted *MalE* 10-1, 14-1, 16-1, 18-1 and 19-1, shown in Table 1.2) cause the MBP precursor to accumulate in the cytoplasm. In the mutant *MalE* 10-1 a proline substitutes for a leucine at residue 10 (or -17 with respect to the cleavage site); although this represents an exchange of one hydrophobic residue for another, proline is a destabiliser of α-helices so it is conceivable that it causes a secretory defect by altering the secondary structure of the signal sequence. The other four mutations introduce a charged residue into the central hydrophobic core. Of these, the two with the strongest effect on MBP export are *MalE* 19-1 and *MalE* 18-1, both substitute an arginine residue for a methionine, this represents the most radical alteration with respect to hydrophobicity in this group of mutants. The substitution of lysine for the neutral amino acid threonine in *MalE* 16-1 is a more moderate alteration in terms of hydrophobicity and has the least effect on MBP export.

Additional *MalE* signal sequence mutations have been obtained (Bankaitis *et al.*, 1985). Of these, two mutations (*MalE* 10-2, *MalE* 11-3, see Table 1.2) introduce a charged residue early in the hydrophobic core. In *MalE* 10-2 an arginine residue substitutes for a leucine, and in *MalE* 11-3 a glutamic acid replaces an alanine residue. Both of these residues cause a much less dramatic export block than the introduction of a charged residue to the more central region of the hydrophobic core, as seen in *MalE* 19-1 and 18-1. It would seem that the strength of the export block resulting from introduction of a charged residue into the hydrophobic core is a function of position. Those that are introduced to the centre of the core cause the strongest export defects, whereas similar changes toward the amino-terminal boundary of the core show much weaker defects.

In an attempt to elucidate further information concerning the functional determinants of the MBP signal sequence, Ryan *et al.* (1986) isolated intragenic revertants of *MalE* signal sequence mutants. The mutations isolated and characterised had second site mutations within the signal sequence and suppressed the export block to varying extents. The results support a length requirement for the hydrophobic core.

Table 1.2 MalE signal sequence mutants that affect the export of MBP.

Wild type sequence of the MBP signal peptide:

```
                5                  10                  15                  20                  25        +1
Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys
                                                                                                    ↑
                                                                                          signal peptide
                                                                                          cleavage site
```

Mutant	Amino Acid Alteration	Effect on Export	Reference
10-1	Leu_{10} → Pro	Show a less dramatic reduction of MBP export than 18-1, 19-1	Bedouelle *et al.* (1980)
14-1	Ala_{14} → Glu		
16-1	Thr_{16} → Lys	Least effect on MBP export (of this group)	" "
18-1	Met_{18} → Arg	Severe block on MBP export (~99%)	" "
19-1	Met_{19} → Arg		
10-2	Leu_{10} → Arg	Slight reduction in MBP export efficiency	Bankaitis *et al.* (1985)
11-3	Ala_{11} → Glu	50% reduction in MBP export	" "
Δ12-18	Amino acids 12-18 deleted	Strongest signal sequence mutation of MalE, export is reduced to less than 1% of synthesised MBP	Bankaitis *et al.* (1985)

In addition, this study showed that interruptions in the core by charged residues can to some extent be compensated for by increasing overall hydrophobicity of the signal peptide without an accompanying increase in core length.

The most severe block of MBP secretion is seen in the mutant *MalE* Δ12-18 where a 21bp deletion results in the removal of 7 residues from the hydrophobic core (Bankaitis *et al.*, 1985). Less than 1% of the MBP synthesised in *MalE* Δ12-18 cells is secreted and processed. The export block can be almost completely reversed by a second mutation where 3 hydrophobic residues are inserted into the core region. Other second site mutations in the *MalE*Δ12-18 signal sequence relieve the export deficiency by various degrees, and there is a good correlation between the length of the hydrophobic core in these suppressor mutants and MBP export proficiency. The authors of this study conclude that there is a threshold level for the length of the hydrophobic core below which export is not possible.

Of all the mutant *MalE* signal sequences that severely affect secretion of MBP, a high proportion are altered at residue 14, 16, 18 or 19. This observation suggests that these residues play a crucial role in the initiation of MBP export.

A similar subset of four key residues has been identified in the hydrophobic core of the LamB signal sequence. Each of four different point mutations in the lamB signal sequence (Emr *et al.*, 1980) introduces a charged residue into the hydrophobic core at position 14, 15, 16 and 19 (shown in Table 1.3), and results in a strong export block where more than 95% of the LamB synthesised accumulates in the cytoplasm.

The introduction of a charged residue into the hydrophobic core does not always interfere with the export function of the signal peptide. This is the case for two other LamB signal sequence mutations, LamBS96 and LamBS73 (see Table 1.3), which introduce either a basic or an acidic residue (respectively) at residue 17 (Bankaitis *et al.*, 1985). The effect of the mutations in both cases is only a small decrease in LamB export, less than 2% of the total LamB synthesised accumulates in the cytoplasm. These mutants provide a striking example of the importance of the position in which a charged residue is introduced into the hydrophobic core and how severely it impedes the secretion of the protein.

Two mutations which cause a severe block of LamB export remove 4 (LamBS78) and 12 (LamBS60) residues of the hydrophobic core, both of these affect the subgroup of critical amino acids in the LamB h-region. The larger of these deletion mutants LamBS60, results in the total block of LamB export.

Table 1.3 Mutations in the h-region of the LamB signal sequence.

Wild type sequence of the LamB signal peptide:

```
                5                  10                  15                  20                  25  +1  +2
Met Met Ile Thr Leu Arg Lys Leu Pro Leu Ala Val Ala Val Ala Ala Gly Val Met Ser Ala Gln Ala Met Ala Val Asp Phe His
                                                                                                 ↑
                                                                                    signal peptidase
                                                                                    cleavage site
```

Mutant	Amino Acid Alteration	Effect on Export	Reference
S71	Val_{14} → Asp		Emr *et al.* (1980)
S70	Ala_{15} → Glu	95% of the synthesised LamB remains cytoplasmic	
S99	Ala_{16} → Glu		
S69	Met_{19} → Arg		
S78	Δ10-13	>95% cytoplasmic	Emr *et al.* (1980)
S60	Δ10-21	total export block	Emr *et al.* (1980)
S96	Gly_{17} → Arg	No major export defect, only 2% of the LamB synthesised accumulates in the cytoplasm	Bankaitis *et al.* (1985)
S73	Gly_{17} → Asp		

These subsets of amino acids in the h-region of the LamB and MBP signal peptides have been implicated as key residues in protein export since mutations showing the introduction of a charged residue at these positions have been isolated on multiple, independent occasions, and similar mutations at other residues within the h-core do not result in such major signal sequence dysfunction. Additionally, all other strong export-defective mutations (such as the deletion *MalE*Δ12-18) in some way affect some or all of these key residues.

Bankaitis *et al.* (1985) discuss two proposals that attempt to explain how alterations at these subsets of residues can drastically affect protein export.

One idea is that the amino acids at these positions of the h-region constitute some kind of recognition site for the cellular protein secretion machinery. The strong export defective mutations, such as *MalE*Δ12-18 and LamBS60, can be suppressed by second site mutations that are extragenic and identify genes that encode components of the protein export apparatus in *E. coli*, such as prl (protein localisation) A, B and C (e.g. Emr *et al.*, 1981; Trun and Silhavy, 1989). This provides strong evidence that the key residues interact with components of the *E. coli* protein secretion machinery.

Another proposal concerning the role of the critical amino acids of the hydrophobic core attributes a primarily structural role to these residues. Hydrophobicity appears to be the major feature disrupted by the point mutations that display the strongest effect on export. Based on an inspection of mutant and wild type signal peptides, the correlation between the length of the h-region and the efficiency of initiation of protein export led to the proposal of a parameter called the hydrophobic axis length (HAL) to provide a measure of signal peptide functionality. The threshold HAL (tHAL) is the minimum length a periodic structure (α-helix or β-strand) of the h-region must display to be functional. Bacterial signal peptides with a tHAL of 18Å or greater would be expected to be functional. Note that the number of residues needed to reach the tHAL depends upon the conformation they adopt, 12 residues in an α-helix or 5 residues in an extended β-strand would be around 18Å in length. In a model where, after the initial interaction of the charged n-region residues with the membrane surface, the signal peptide loops into one membrane the importance of tHAL to signal function could be explicable. The stringent requirement for a minimum hydrophobic core length may reflect the distance that the signal peptide must be inserted into the lipid bilayer in order to successfully initiate protein translocation across the membrane.

In a recent report (Thom and Randal, 1988), the mutant *MalE* 18-1 was used in an attempt to elucidate where in the pathway of secretion the h-region is important. The *MalE* 18-1 mutation is a substitution of an arginine residue at position 18 of the leader peptide where a methionine is normally found (Table 1.2) and results in a strong block on the export of MBP (Bedouelle *et al.*, 1980). The mutated leader carried by the *MalE* 18-1 precursor was found to be active in mediating entry to the export pathway since newly synthesised precursor *MalE* 18-1 was found associated with the membrane. Thus the defect in export seen for *MalE* 18-1 occurs at a step following initial association with the membrane, possibly the translocation step. The authors suggest that this data supports the proposal that the hydrophobicity of the leader peptide is important during the direct interaction of the precursor with the phospholipid bilayer (Briggs *et al.*, 1986) and that the results they present indicate that such interactions are crucial for translocation and that other parameters render the leader functional during the entry stage.

In conclusion, extensive studies of mutants in the h-region of signal peptides in prokaryotes and eukaryotes, *in vitro* and *in vivo*, have clearly illustrated how essential an intact hydrophobic core is to the functioning of signal sequences. Sufficient length and hydrophobicity have been shown to be the crucial parameters governing effective function of the h-region. Mutations where charged residues are introduced to the hydrophobic core generally result in significant interference with the export function of the signal peptide, although a few examples exist where this effect is not striking.

It has been proposed that the hydrophobic core is the most important region of the signal peptide for recognition by cellular factors involved in protein secretion such as the SRP. Another possible function has been suggested where the h-region is the important area of the signal peptide for interaction with and translocation across the membrane. Precisely at what point(s) in the secretory pathway the h-region is important is under investigation. Studies of the type performed by Garcia *et al.* (1987) on n-region mutations, where the efficiency of recognition by the SRP of various n-region mutants and interaction with the membrane were examined, should assist in determining such aspects of the hydrophobic core.

(iii) **The c-region**

The last 5 (in eukaryotes) and 6 (in prokaryotes) residues of the signal sequence define the site of signal peptidase cleavage. This segment is known as the c-region and the residues found here are more polar than those found in the h-region (von Heijne, 1985). There is very little variation observed in the length of the c-region and it is usually between 4-7 residues long.

In contrast to the variability in the primary structure of the signal sequence at the n- and h-regions, leader peptides display marked similarities to one another in the c-region, particularly in the last 3 residues before the cleavage site. This observation led to the postulation of the "-3, -1" rule (von Heijne, 1983; Perlman and Halvorsen, 1983) in which the residues that occupy the positions -1 and -3 (relative to the site of cleavage, between -1 and +1 of the mature protein sequence) must have small neutral side chains. Alanine is the residue found most frequently at the -1 position and Ser, Gly, Cys, Thr and Gln are also found. At the -3 position, the same group of amino acids are the most commonly found with the exception of Gln. Whole classes of residues are not found at the -3 and -1 positions of signal sequences; these are the aromatic (Phe, His, Tyr, Trp), charged (Asp, Glu, Lys, Arg), or large polar (Asn at -1 and -3, and Gln at -3) residues. Proline is not found in the region -3 to +1. Conversely, the small neutral residues are quite rare at position -2 where bulky aromatic amino acids seem to be selected. The secondary structure disrupting residues Gly and Pro are found predominantly at positions -4 and -5, respectively. The hydrophobic residues seem to be tolerated in positions +1, -2 to -4 and then become dominant from position -6 onwards into the central h-region. Charged residues are tolerated at -2, -4 and -5, but are much more common in the mature sequences from position +1 onwards (von Heijne, 1983, 1984).

Another feature usually observed in the c-region is the presence of residues predicted to be associated with β-turns (Ser, Gly, Pro). These frequently occur just after the end of the hydrophobic core and could therefore be important for signal peptidase access to the cleavage site (see Fig. 1.10) (Perlman and Halvorsen, 1983).

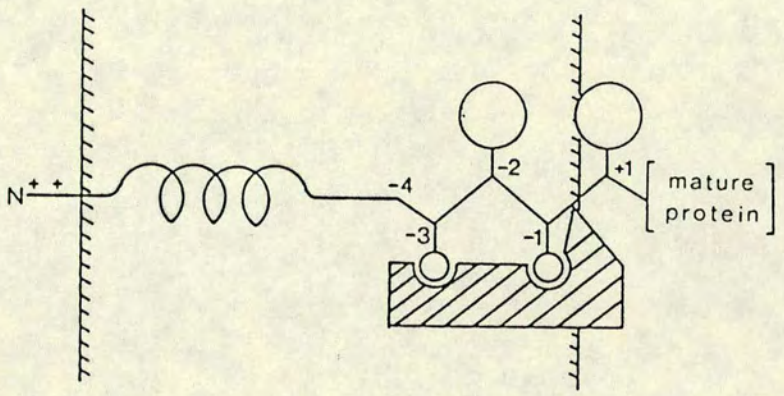
The pattern or "consensus" sequence of amino acids observed in the c-region is perhaps most easily understood in terms of particular cleavage requirements of the signal peptidase. Whether the residues represent a recognition site or if the preference for small, neutral amino acids is simply a reflection of steric considerations is unclear.

Figure 1.10

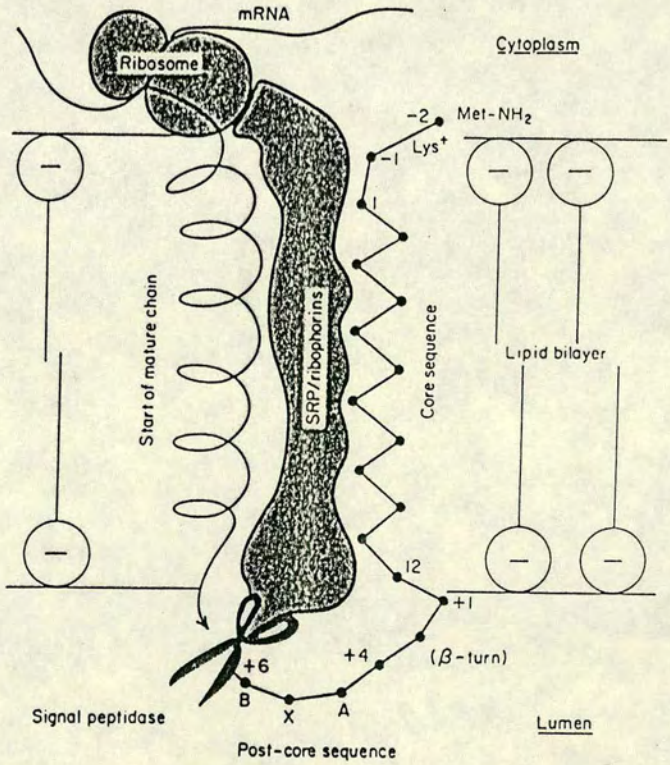
A. Illustration of signal sequence-protease complex proposed by von Heijne (1983) in which the signal peptide spans the membrane as a "helix and sheet" structure and the small, neutral residues at positions -1 and -3 fit into a pocket in the peptidase thereby defining the cleavage site as between -1 and +1.

B. Model proposed by Perlman and Halvorson (1983) showing the positioning of the signal peptide in the membrane at the time of signal peptide cleavage. It shows the β-turn at position +1/+2 (i.e. -5/-6) which the authors believe help to correctly position the signal sequence at the active site of the protease. (The nomenclature in this diagram differs from that used in the text; B = -1, X = -2, A = -3 etc. relative to the site of signal peptidase cleavage).

A


N+ +
-4
-3
-2
-1
+1
mature protein

B


mRNA
Ribosome
Cytoplasm
-2
Met-NH2
-1
Lys+
1
Start of mature chain
SRP/ribophorins
Core sequence
Lipid bilayer
12
+1
+4
(β-turn)
+6
B
X
A
Signal peptidase
Lumen
Post-core sequence

Various speculations have been made (von Heijne, 1983; Perlman and Halvorsen, 1983) as to the nature of the structure of the signal sequence/signal peptidase complex, one possibility being that of an alternating sheet structure near the cleavage site adjacent to a largely hydrophobic α-helix (see Fig. 1.10) (von Heijne, 1983b).

von Heijne (1984b) proposes that cleavage of the signal sequence is influenced by the position of the c-terminus of the hydrophobic core, which defines a "window for processing", and that potential cleavage sites within this window compete for access to the signal peptidase cleavage site. This model is derived primarily from statistical observations with supporting evidence from mutations in these regions which behave as the model predicts. For example, a secretion mutant of yeast invertase (*suc2-s1*) has been characterised (Schauer *et al.*, 1985) where the residue at position -1 has been changed from an Ala to a Val, and therefore no longer constitutes a signal processing site. The mutant preprotein is however cleaved at an adjacent site (and at a slower rate) that now represents the best candidate for a signal peptidase processing site in the region (see Table 1.1). This example of the secondary cleavage site being utilised follows the predictions set out in the von Heijne (1984b) "cleavage window" model.

When a secretory protein from one organism is expressed in another organism, correct cleavage of the signal peptide does not always occur. For instance, when α-amylase from *Bacillus stearothermophilus* was expressed and secreted by *E. coli*, a large fraction (40%) was found to be cleaved at an alternative site. The aberrant cleavage took place 3 residues (to the N-terminal side) before the normal site of processing. This type of "incorrect" cleavage apparently reflects differences in the specificity of signal peptidases from different organisms (von Heijne and Abrahamsén, 1989). This is clear in data of signal sequences from different species compiled by von Heijne and Abrahamsén (1989) which indicates that the length of the c-region (and therefore site of cleavage) varies between different species, with eukaryotes having shorter c-regions than bacterial signal peptides and gram positive bacteria having slightly longer c-regions than the gram negative *E. coli* (the mean c-region lengths are 5, 6 and 8 residues respectively).

Cleavage of the signal peptide from the remainder of the mature protein occurs either as the polypeptide crosses the membrane or very shortly afterwards (Gilmore and Blobel, 1985).

Cleavage of the signal sequence is not a requirement for the translocation of precursor proteins across a membrane as has been amply demonstrated in a variety of

Table 1.4

This is a list of some mutations found in the c-region of signal sequences of secreted or exported proteins. In all these cases, although no signal peptide cleavage occurs (or the rate of cleavage is drastically reduced) the preprotein has been shown to have translocated across the membrane. In some cases the uncleaved proteins finally reach their normal locations.

Table 1.4

Protein	Normal Location	Mutation	Cleavage	Location of Mutant Protein	Reference
α-amylase/β-lactamase	α-amylase signal sequence promotes secretion of β-lactamase into culture medium	Last 6 residues of the signal sequence are deleted	NONE	The fusion remains membrane bound	Palva *et al.* (1982)
Lipoprotein (lpp/OMLP)	Outer membrane	Gly_{-7} → Asp	NONE	50% reaches the outer membrane, the other 50% remains on the outer surface of the inner membrane	Lin *et al.* (1980)
Coat protein φM13	Plasma membrane (en route to surface of virus particle)	Pro_{-6} → Ser Ser_{-3} → Phe Ala_{-1} → Thr	NONE NONE NONE	Plasma membrane	Kuhn and Wickner (1985)
Maltose Binding Protein (MBP)	Periplasm	Ala_{-3} → Asp	NONE	Membrane anchored	Fikes and Bassford Jr (1987)
		Ala_{-3} → Asp, with deletion of 7 residues of h-core	NONE	Eventually secreted to periplasm	

Protein	Normal Location	Mutation	Cleavage	Location of Mutant Protein	Reference
β-lactamase	Periplasm	$Pro_{-4} \rightarrow Leu$	< 5% is cleaved	Membrane associated	Koshland *et al.* (1982)
		Substitution of residues -6 → -3 (fs (18,21))			
		Substitution of residues -10 → -3 (fs (14,21))			
Acid Phosphatase (PH05) in *Saccharomyces cerevisiae*	Final location unclear, periplasm or cell wall, some is secreted	Δ-3 → -10	NONE	50% remains in the cells, 50% is secreted to the cell surface (but cannot cross the cell wall)	Haguenauer-Tsapis and Hinnen (1984)
Invertase in *Saccharomyces cerevisiae*	Secreted	$Ala_{-1} \rightarrow Val$	Some processing (at an adjacent residue) at a slow rate	Very little is secreted. Aggregates of the uncleaved form found in the ER lumen in association with ER membrane, transport is delayed. The cleaved form may remain by an association with the unprocessed sub-units.	Schauer *et al.* (1985)

systems. One example of an uncleaved signal peptide where the protein is successfully translocated is seen for the product of a fusion of the DNA sequences encoding an α-amylase signal sequence and the structural gene for *E. coli* β-lactamase (devoid of its own signal sequence) (Palva *et al.*, 1982). The complete α-amylase signal sequence successfully directs the secretion of β-lactamase from *B. subtilis*, into the culture medium. However, if the last six residues of the signal sequence are absent, the β-lactamase is translocated across the membrane, but remains unprocessed and membrane bound. Other examples where non-cleaved signal sequences are able to direct translocation are detailed in Table 1.4.

Although not essential for the process of translocation, signal peptidase activity is a requirement for normal cellular growth, as was established in a set of experiments where the goal was to define the role(s) of the leader peptidase (Dalbey and Wickner, 1985). The translocation and secretion of several precursor proteins were studied in an *E. coli* strain where the level of active signal peptidase could be controlled. When signal peptidase expression was repressed, after a few cell doubling times the growth rate slowed, and if its synthesis was then induced the growth rate abnormality was rescued.

Gene disruption experiments of the SEC11 locus in *Saccharomyces cerevisiae* (Böhni *et al.*, 1988) also illustrate the necessity of the signal peptidase for growth. The SEC11 locus is thought to encode the yeast signal peptidase (or a component of a signal peptidase complex), disruption of the chromosomal locus by gene displacement results in lethality of the cells, confirming the essential role of the SEC11 locus in cell growth.

When cleavage of the signal peptide does not accompany translocation, the precursor protein usually either remains associated with the membrane, or its transport from the lumen of the ER to subsequent compartments is dramatically delayed.

In the experiments of Dalbey and Wickner (1985) the precursors of the *E. coli* proteins, MBP (normally a periplasmic protein) and OMPA (normally an outer membrane protein), were retained at the outer surface of the inner membrane when the expression of signal peptidase was repressed, and released only when signal peptidase synthesis was allowed to occur.

Other mutations of signal peptides in the c-region which prevent cleavage have been found where the precursor protein apparently remains associated with the membrane. This is the case for several signal sequence cleavage mutants of *E. coli* β-lactamase expressed in *Salmonella typhimurium* (Koshland *et al.*, 1982) that are

translocated but are not cleaved and remain associated with the outer surface of the plasma membrane.

In some instances of cleavage site mutations, subsequent transport of the mutant precursor protein is not completely prevented, but significantly delayed. This is seen for a mutant of the bacterial outer membrane protein lipoprotein (lpp) in which amino acid 14 of the 20 residue long signal sequence is changed from glycine to aspartic acid. This mutation results in non-processing of the pro-lpp signal peptide, however the precursor is found in both the inner and outer membranes (Lin *et al.*,1980) indicating that signal peptide cleavage is not essential for assembly into the outer membrane.

The export of unprocessed MBP to the periplasm has also been noted (Fikes and Bassford, 1987). In the mutant responsible (*CC15*) 7 amino acids are deleted from the h-region of the signal sequence (residues 12-18 inclusive), and the alanine found at position -3 has been altered to an aspartic acid. The *CC15* mutation results in the slow export of precursor MBP to the periplasm. Mutating the wild type signal sequence so that the alanine residue normally found at position -3 is substituted for an aspartic acid results in the protein remaining uncleaved and anchored to the membrane. The authors propose that the reason the uncleaved protein in the mutant *CC15* can be secreted into the periplasm, while the mutant altered only at residue -3 produces an uncleaved protein which remains membrane associated, is that the *CC15* h-region is insufficient to serve as a membrane anchor so that the unprocessed MBP is released into the periplasm in this mutant.

Several strains of yeast that are mutant in the secretion of invertase accumulate aggregates of the active enzyme in the lumen of the ER (Schauer *et al.*, 1985). Characterisation of one of these mutants, *suc2-s1*, showed the presence of two forms of invertase which differed in size by 2.5kD. The *suc2-s1* gene was found to be mutated in the signal sequence at position -1, resulting in a valine residue substituting for the alanine normally found at this position. This has the effect of destroying the signal peptidase cleavage site and the enzyme now cleaves at an adjacent site (see Table 1.1). Processing occurs at a slower rate than is usual for the wild type precursor invertase. The *suc2-s1* mutation reduces by 50-fold the rate of invertase transport from the lumen of the ER to the Golgi body, but thereafter transport appears to proceed at a normal rate. The delayed secretion due to the signal sequence mutation could be accounted for either by aberrant sub-unit interaction or aggregation within the lumen of the ER to the point where proper packaging and transport are hindered, or by the *suc2-s1*

invertase remaining in association with a stable ER component. Since the secretion of other proteins is not affected, the authors favour the idea that the low level and delay of invertase secretion is due to the defective signal peptide cleavage resulting in association of the invertase with a component of the ER membrane.

Finally, there is at least one example of a secretory protein where leader peptide cleavage is not required for normal secretion. This is the case in the yeast mating pheromone, α-factor. It is synthesised as several repeat units of the 13 amino acid long pheromone with some intervening residues. This precursor form is known as prepro-α-factor, the signal peptide of which is not removed upon its translocation into the ER, although it does possess a leader peptide cleavage site (Julius *et al.*, 1984). However, it is possible that cleavage of the signal peptide of prepro-α-factor is not necessary since, as it proceeds along the transport path, the prepro-region is effectively removed when the pheromone repeat units are split up by another type of endoproteolytic scission. In the model proposed by the authors the signal sequence and a few of the first residues of the prepro-α-factor remain attached to the ER/Golgi membrane when the mature pheromone α-factor units are released.

In summary, the c-region of signal peptides shows a more conserved, distinctive sequence of amino acids than the n- or h-regions. This is attributed to the substrate requirements of the signal peptidase for recognition and cleavage. Once again, the study of mutations in this region has been crucial in understanding the important parametres of the c-region in signal peptide function. Mutations in the c-region do not interfere with the translocation of presecretory proteins, but in the cases where non-cleavage results the protein is either not secreted or is exported at a signficiantly reduced rate.

(iv) Conformation of Signal Sequences

Little is known about the physical characteristics of signal sequences that influence their participation in the process of protein secretion. The observations of the interchangeability of signal sequences from one protein to another and the ability of a signal peptide to promote the translocation of an otherwise cytoplasmic protein, coupled with the absence of primary sequence homology indicate that signal sequences must share some secondary or tertiary structural similarities.

Recent work has served to further illustrate the low information content of the

primary structure of signal peptides. Kaiser *et al.* (1987) performed gene fusion experiments where the yeast secretory invertase signal sequence was replaced with essentially random peptide sequences encoded by unselected fragments derived from human genomic DNA. A surprisingly large proportion - around 20% - of these random sequences were found to be able to function as export signals and this activity appeared to correlate with overall hydrophobicity. This suggests that some gross property of signal sequences is recognised as the initiator of export signal.

Methods used to predict the preferred conformations of polypeptides (e.g. Chou and Fassman, 1974a and b) indicate that signal sequences have a high probability of adopting α-helix and β-sheet structures in the hydrophobic region. Most signal sequences are predicted to form a β-turn near the signal peptidase cleavage site (Perlman and Halvorson, 1983), whereas the n-region displays no consistent conformational preference.

Some studies have attempted to elucidate the importance of conformation in signal sequence function by changing the signal sequence in such a way that predicts the conformational tendencies will be altered. Various insertions into the n-region of the yeast invertase signal sequence that were predicted to stabilise an α-helix, favour a β-turn or destabilise α-helices or β-sheets, were found to have no effect upon the secretion of the protein (Brown *et al.*, 1984). The n-region therefore appears to have few constraints with respect to conformation. Similar substitution experiments in the c-region indicated that the absence of a β-turn in this area can inhibit the cleavage of the signal sequence (Briggs and Gierasch, 1986). In the case of the h-region, changes that are predicted to alter the conformation of the hydrophobic core can drastically affect signal sequence function so that no export occurs. This was demonstrated with the *E. coli* LamB signal sequence which is predicted to adopt an α-helical structure, even though two α-helix breaking residues are present in the hydrophobic core. As these two amino acids, proline and glycine, are separated by seven residues they are not thought to interfere with the helical potential of the remainder of the sequence. In a mutant strain where LamB export is defective due to a deletion of four of the amino acids between the proline and glycine residues, the mutant signal sequence was predicted to have a random conformation in the h-region. Pseudorevertants of this export defective mutant were found to have substituted the proline or glycine residue for one which would allow the h-region to adopt an α-helical structure (Briggs and Gierasch, 1986).

Various biophysical studies of isolated signal sequences (in the form of synthetic peptides, signal sequence fragments and peptides resembling signal sequences) indicate that signal sequences adopt different conformations in polar and apolar environments. Generally, signal peptides are thought to adopt a random conformation in an aqueous environment and an α-helical conformation in an apolar environment (Briggs and Gierasch, 1986). Briggs *et al.* (1986) demonstrated a correlation between the ability of signal sequences to adopt these various conformations and their ability to initiate protein export *in vivo*. The authors propose that these conformational interconversions in different conditions may facilitate the initiation of the export process, and suggest a model for the events that occur when a signal sequence first encounters the membrane. This is mainly based on the signal sequence, largely unstructured in the cytoplasm, adopting a β-structure upon approach to a lipid surface and subsequently, an α-helical structure when inserted into the membrane.

INVOLVEMENT OF MATURE SEQUENCES IN SECRETION

The interchangeability of signal sequences from one protein to another demonstrates that this stretch of amino acids is usually sufficient to promote protein secretion. However, not all signal sequence/structural gene fusions, or secretory proteins expressed in a foreign host are correctly targeted, cleaved or secreted, and attaching a signal peptide to a cytoplasmic protein does not always ensure its translocation across a membrane. An example of the latter category is described for yeast invertase/β-galactosidase hybrids (Emr *et al.*, 1984) where even when 90% of the *suc2* gene is fused to *E. coli* β-galactosidase sequences, the hybrid protein does not get secreted and is found in the ER. Such observations indicate that the signal sequence does not always supply all the requirements for correct targeting and processing. Various studies have addressed the question of the role of mature sequences in protein secretion in both eukaryotic and prokaryotic systems.

One of the first instances of a mutation in the mature protein sequence affecting cleavage of the signal peptide was noted for the bacteriophage M13 coat protein. A change of residue from glutamic acid to leucine at reside +2 in the mature protein sequence resulted in the procoat being a poor substrate for signal peptidase *in vivo* and *in vitro* (Russel and Model, 1981).

A set of deletion mutants were constructed in the passenger domain of bovine

preprolactin, which carried the normal signal sequence (Andrews *et al.*, 1988). These mutants were assayed by translation (of the products from a cell free transcription system) and translocation across canine pancreatic microsomal membranes. Deletions in the mature prolactin domain resulted in diminished "passenger" translocation and therefore show that the sequence of a secreted protein domain can influence the ability of the signal peptide to direct its translocation across the ER membrane. The authors suggest that the reduced passenger translocation for the deletion mutants reflects a secondary effect of the deletion on protein-folding which either masks the signal directly or generates a substrate less easily transported. The effect of deletions in the prolactin passenger gradually diminish as they are made further into the mature sequence and deletions from amino acid 11 onwards do not drastically affect the efficiency of translocation. This may be a reflection on the coupling of translation and translocation in eukaryotic systems so that only shorter range folding interactions are able to influence the efficiency of the process of secretion. Indeed, in *E. coli* an alteration of the mature sequence of the maltose binding protein (MBP) at amino acid 283 (Tyrosine - Aspartic Acid) has been shown to partially rescue an export defect of MBP resulting from an amino acid substitution in the signal sequence (Cover *et al.*, 1987).

Alteration of the amino -terminus of the mature sequence of the *E. coli* alkaline phosphatase (PhoA) has been found to severely affect its translocation across the cytoplasmic membrane into the periplasm (Li *et al.*, 1988). When the amino-terminus of the mature PhoA is altered by introducing two positively charged amino acids very close to the beginning of the sequence, the export of PhoA is reduced by a factor of 50. It appears that the positive charge of these amino acids is interfering with the translocation process, since when one of the positively charged residues is eliminated or replaced by a negatively charged residue, high levels of export are restored. While other structural features have not been ruled out, the authors favour an explanation based on the charge of the region, and this is supported by data from von Heijne (1986b) who notes that the absence of net positive charge in the first residues of the mature sequences of secreted proteins is a conserved feature of bacterial exported proteins.

In an extension of the studies noting the importance of mature sequences in protein secretion, the interaction of the mature sequences of an exported protein with components of the secretion apparatus has been demonstrated in *E. coli.* A recent study (Gannon *et al.*, 1989) analysed the export of fusions of maltose-binding protein

(which requires the SecB protein for export) and alkaline phosphatase (which does not need SecB for export) in *secB*-carrying mutant cells to show that the first third of the mature sequences of *E. coli* MBP determine its requirement for the product of the SecB locus for efficient export. This substantiates the report by Collier *et al.* (1988) which indicated that amino acids 151-186 of MBP are necessary and sufficient for interaction with the SecB protein, and that this interaction prevents premature folding of the newly synthesised polypeptide into an export incompetent form.

OBJECTIVES OF THE PROJECT

The aims of this project were two fold; firstly, the characterisation of the mutant fs(1)1526 was undertaken, and secondly, a derivative of this mutant strain was used in a study to determine some of the *cis*-acting sequences involved in regulating the expression of the yp3 gene in *Drosophila melanogaster.*

The strain fs(1)A1526 was generated in a large screen for first chromosomal female sterile mutants by Gans *et al.* (1975). Females that are homozygous for fs(1)A1526 lay apparently normal eggs which generally fail to develop beyond the first embryonic cleavage (Zalokar *et al.*, 1975). It was noted that adult females homozygous for fs(1)A1526 display normal levels of YP1 and YP2 but lack YP3 in their haemolymph (Bownes and Williams, unpublished observation). Such a characteristic made this mutant valuable. It represents a tool that can be used to investigate the *cis*-acting sequences involved in controlling the expression of YP3. The characterisation of the mutant strain itself is pertinent to the study of the YP s generally. For example, the obvious query arose as to whether the lack of YP3 was directly responsible for the female sterile phenotype of fs(1)A1526 flies. The lesion causing the lack of YP3 was of interest since no other YP mutant displaying this phenotype has been extensively characterised, and the information gained from investigating the mutant could provide additional insight into the yolk protein system in *D. melanogaster.*

It is possible to use the mutant strain fs(1)A1526 to investigate the *cis*-sequences of yp3 as the complete gene can be transformed, using the P-element mediated germ line transformation system developed by Rubin and Spradling (1982), and obtained in the fs(1)A1526 background where any YP3 produced will be the result of expression from the introduced yp3 sequences. This study was regarded as worthwhile for a number of reasons. Prior to this study, very few of the transformation studies

investigating YP1 and YP2 regulatory sequences involved using the complete, unaltered genes and it is possible that disruption of the gene sequences (such as inserting a segment of M13 DNA) may give rise to aberrant results. Using the unmodified wild type gene should avoid such pitfalls and ensure that more of the sequences important in the regulation of yolk protein synthesis are identified. The results from the transformation studies with yp1 and yp2 were becoming increasingly difficult to interpret since their proximity results in the two genes sharing regulatory elements. As the yp3 gene is situated much further along the X-chromosome, it is entirely independent of the other yp genes and possesses all the sequence elements necessary to correctly modulate its expression, thereby providing a simpler system to investigate regulation of YP synthesis.

CHAPTER 2

MATERIALS AND METHODS

2.1 SOLUTIONS

Chemicals were obtained from Sigma, BDH, Aldrich.

Enzymes (restriction endonucleases, T4 DNA ligase, T4 DNA polymerase, Klenow fragment of DNA polymerase, DNA polymerase, DNase I, RNase A, proteinase K) were obtained from Pharmacia and BRL).

Radioisotopes were obtained from Amersham.

Nitrocellulose membranes were obtained from Schleicher and Schüll.

Hybond-N nylon membranes were obtained from Amersham.

Standard solutions were made using distilled water in sterile, baked glassware and were generally made sterile by autoclaving (15 psi/15 min) or by passing through a 0.45μm nitrocellulose filter. Solutions not detailed in the text are described below.

TE	10mM Tris-Cl pH 7.4-8.0, 1mM EDTA
SEQ TE	10mM Tris-Cl pH8.0, 0.1mM EDTA
0.5M EDTA	0.5M diaminoethanetetra-acetic acid, pH8.0
10x High Salt Restriction Buffer	1M NaCl, 0.5M Tris-Cl pH7.5, 0.1M $MgCl_2$, 10mM DTT
10x Medium Salt Restriction Buffer	0.5M NaCl, 0.1M Tris-Cl pH 7.5, 0.1M $MgCl_2$, 10mM DTT
10x Low Salt Restriction Buffer	0.1M Tris-Cl pH 7.5, 0.1M $MgCl_2$, 10mM DTT
10x TBE gel buffer	0.89M Tris-Borate, 0.89M Boric acid, 10mM EDTA
10x DNA gel loading sample buffer	50% (v/v) glycerol, 0.25% (w/v) Bromophenol blue

10x MOPS gel buffer	0.2M Na-MOPS pH 7.0, 50mM Sodium Acetate, 10mM EDTA
RNA Sample buffers	
FSB (formaldehyde sample buffer)	50% (v/v) formamide, 25% (v/v) formaldehyde (at 4% (w/v)), 25% (v/v) 10x MOPS buffer
FDE (Ficoll, dye, EDTA)	50% (v/v) 0.2M EDTA pH 7.0, 3% (w/v) Ficoll (type 400), 0.25% (w/v) Bromophenol blue
RNA extraction buffer	100mM Tris-Cl pH 7.5, 10mM EDTA, 150mM Lithium Chloride, 1% (w/v) SDS
OLB (oligo labelling buffer)	Solution 0: 0.125M $MgCl_2$, 1.25M Tris-Cl pH 8.0 Solution A: 0.95ml Solution 0, 18ml 2-mercaptoethanol, 25μl 20mM dATP, 25μl 20mM dTTP, 25μl 20mM dGTP Solution B: 2M Hepes pH 6.6 Solution C: Hexadeoxyribonucleotides (Pharmacia) suspended in TE at 90 OD units/ml OLB is made by mixing Solution A, B and C in the ratio 2:5:3 (and is stored at -20°C).
20x SSC	3M NaCl, 0.3M Na-Citrate
20x SSPE	3.6M NaCl, 20mM $NaHPO_4$ pH 7.4, 20mM EDTA pH 7.4

Salmon Sperm DNA	Made up as a 5mg/ml stock and sonicated until the viscosity is reduced. The solution is then phenol extracted (2.4.1.1) and stored at -20°C.
10x Denhardts Solution	0.2% (w/v) BSA, 0.2% (w/v) polyvinyl pyrrolidine, 0.2% (w/v) Ficoll
Ringers Solution	3.2gNaCl, 3.0gKCl, 1.8g $MgSO_4$, 0.69g $CaCl_2.2H_2O$, 1.79g Tricine, 3.6g glucose, 17.1g sucrose. Made up to 1 litre with dH_2O, pH adjusted to 6.95, filter sterilised and stored at 4°C.
TM buffer	0.1M Tris-Cl pH8.0, 50mM $MgCl_2$
X-gal (BC1G)	20mg/ml 5-bromo-4-chloro-3-indoyl-ß-galactoside, made up in dimethyl formamide and stored at 4°C.
IPTG	10mg/ml Isopropyl-β-D-thiogalactoside, stored at -20°C.
APS	15% or 25% (w/v) ammonium persulphate
30% Acrylamide stock (for DNA sequencing)	30% (w/v) acrylamide, 1% (w/v) N-N-methylene bis-acrylamide, deionised with MB-1 (BDH), filtered and stored in the dark at 4°C.

Protein Gel Solutions:

2x Polypeptide Sample Buffer	50mM Tris-Cl pH6.8, 1% (w/v) SDS, 20% (v/v) glycerol, 0.1% (v/v) β-mercaptoethanol, 0.01% (w/v) Bromophenol blue

Separating Gel Solutions:

	20%	7%	10%
3M Tris-Cl pH8.8	3ml	3ml	5ml
10% (w/v) SDS	0.2ml	0.2ml	0.4ml
30% : 0.8% Acrylamide: Bis-Acrylamide	13.3ml	4.6ml	13.3ml
80% (w/v) sucrose	3.4ml	-	-
dH_2O	0.6ml	12.7ml	21.3ml
15% APS	35µl	35µl	133µl
TEMED	5.5µl	5.5µl	25µl

Stacking Gel Solution:

0.5M Tris-Cl pH 6.8	1.25ml
30% : 0.8% acrylamide Bis-Acrylamide	1.25ml
10% (w/v) SDS	0.1ml
dH_2O	7.3ml
15% APS	75µl
TEMED	7µl

10x Gel Running Buffer

Glycine	144g	Made up to 1 litre with dH_2O
Tris-base	30.2g	
SDS	10g	

Western Blotting Solutions

Transfer Buffer	57.6g Glycine, 12g Tris base, made to 3.2 litres with dH_2O, then add 0.8 litre methanol
TS Buffer	10mM Tris-Cl pH7.4, 0.9% (w/v) NaCl

Developing Solution	120mg 4-chloro-1-napthol dissolved in 40mls methanol, then add 200mls TS buffer and H_2O_2 to 0.01% (v/v).

2.2 *DROSOPHILA* STOCKS AND STRAINS

Drosophila strains used are listed in Lindsley and Grell (1968) and Lindsley and Zimm (1985). Table 2.1 presents a list of stocks and balancer chromosomes used with brief descriptions of their relevant features.

Fly stocks were maintained in vials and bottles at 18°C or 25°C on various media as listed in Table 2.2. A fungicide, nipagin, was added to a final concentration of 4.5g/L and on occasion antibiotics, such as Gentamycin (to 40mg/L), were added. Where necessary (see 2.4.9) the antibiotic Geneticin G418 (BRL or Sigma) was used to supplement the medium at a final concentration of 0.75-1.0mg/ml. G418 was stored as an aqueous solution at -20°C for periods of up to 4 weeks.

2.3 MICROBIOLOGICAL STRAINS AND MEDIA

2.3.1 Microbiological strains

All bacterial strains, plasmid vectors, bacteriophage vectors and other plasmids used are listed in Table 2.3.

Bacterial stocks were maintained on the appropriate plates at 4°C and also as stabs at room temperature in the dark.

Lambda bacteriophage were stored at 4°C in phage buffer with a few drops of chloroform added.

M13 bacteriophage were stored as DNA in SEQTE buffer at -20°C.

Plasmid DNA samples were also stored suspended in TE buffer at -20°C.

2.3.2 Media

All media (listed in Table 2.4) were sterilised by autoclaving. Antibiotics, vitamins and sugars were made up in distilled water and filter sterilised. Where appropriate, antibiotics were added to plates and media, e.g. ampicillin to a final concentration of 50μg/ml.

Table 2.1 *Drosophila* Stocks

STOCKS	FEATURES	REFERENCE
Oregon R	A wild type strain	Lindsley and Grell (1968)
fs(1)A1526/FM3	A female sterile strain which lacks YP3 in the haemolymph and eggs.	Gans *et al.* (1975)
Df(1)HA92/FM7	A deficiency strain in which the X-chromosome is deleted from 12A6/7-12D3	Lindsley and Zimm (1985)
y w ct^6 m f	A strain bearing a multiply marked X-chromosome	UMEÅ Drosophila stock centre, UMEÅ, Sweden

Balancer Chromosomes

FM3 In(1)FM3, l(1), y^{31d} sc^{8} dm B l(1)

FM7 In(1)FM7, y^{31d} sc^{8} w^{a} sn^{x2} v^{of} g^{4} B

CyO In(2LR)O, Cy dp^{lv1} pr cn^{2}

TM3 In(3LR)TM3, y^{+} ri p^{p} sep Sb bx^{34e} e^{s} Ser

Symbols used are listed in Lindsley and Grell (1967), and Lindsley and Zimm (1985).

Table 2.2 Media Used for Propagating *Drosophila*

Adh Food	Flake yeast 100g, Sugar 100g, Agar 16g, dH_2O 1L
Cornmeal Food	Cornflour 250g, Sugar 500g, Yeast pellets 175g, Agar 100g, dH_2O 10L
Treacle Egglaying Medium	Black Treacle 22.5ml, Agar 6.6g, dH_2O 140ml. Autoclave and dilute 1:1 with dH_2O before pouring on to suitable Petri dishes.

Table 2.3

Hosts	Genotype	Comments	Reference
HB101	F^-, hsdS20, recA13, ara14, proA2, lacY1, galK2, rpsL20(Sm^r), xy15, mtl1, supE44, λ^-	Used as general plasmid host	Boyer (1969)
NM522	hsdΔ($M^-S^-R^-$)Δlac, Δ pro, supE, thi, /F'proA$^+$B$^+$, laqI^q, lacZ ΔM15, traD36	Used as a host for M13 bacteriophage and plasmids (e.g. puchsneo)	Gough and Murray (1983)
TG1	Δ(lac,pro), supE, thi, hsdD5/F'traD36, proA$^+$B$^+$, laqI^q, lacZΔ M15	Used as a host for M13 bacteriophage	Amersham International (unpublished)

Plasmids	Genotype	Comments	Reference
pGEM1/2	Amp^r	Vector used for subcloning *Drosophila* sequences	Promega Biotechnology (Melton, 1984)
pUChsneo	Amp^r	Vector for *Drosophila* P-element germ line transformation	Steller and Pirrotta (1985)
phsπ	Amp^r	Plasmid containing a helper element for *Drosophila* transformations	Steller and Pirrotta (1985)
pYP1/2/3	Amp^r	pBR322 (Sutcliffe, 1978) containing a HIII (yp1 and yp2) or SalI-HIII (yp3) fragment encoding *Drosophila* yp1, 2 and 3	Barnett *et al.* (1980)

Table 2.3 (continued)

pGemAdh	Amp^r	Gemini-1 containing an XbaI fragment of the *Drosophila* alcohol dehydrogenase gene	Source: Alan Shirras, University of Edinburgh (see Shirras and Bownes, 1987)

Bacteriophage	Comment	Reference
M13mp18/19	Vectors used for the production of sequencing templates	Yanish-Perron (1983)

Table 2.4

L broth	10g Difco Bacto Tryptone, 5g Difco Bacto Yeast extract, 5g NaCl per litre, pH 7.2
2 x TY broth	16g Difco Bacto Tryptone, 10g Difco Bacto Yeast extract, 5g NaCl per litre
L agar	As L broth with the addition of 15g Difco agar per litre
BBL agar	10g Baltimore Biological Laboratories trypticase, 5g NaCl, 10g Difco agar per litre
BBL top agar	As BBL agar except only 6.5g Difco agar per litre
Spitzen minimal salts	10g $(NH_4)_2SO_4$, 20g K_2HPO_4, 30g KH_2PO_4, 5g tri-sodium citrate, 1g $MgSO_4$ per litre
Minimal medium	80ml spitzen minimal salts, 4ml 20% (w/v) glucose, 0.2ml 1mg/ml vitamin B1 per 400ml
Minimal agar	As minimal medium with the addition of 6g Difco Bacto agar per 400ml
Phage buffer	3g KH_2PO_4, 7g $Na_2H\ PO_4$, 5g NaCl, 10ml 0.1M $MgSO_4$, 10ml 0.1M $CaCl_2$, 1ml 1% (w/v) gelatin per litre

2.4 METHODS

2.4.1 General Methods

2.4.1.1 Phenol Extraction

Phenol was redistilled, saturated with water, equilibrated with TE buffer and 8-hydroxyquinoline was added to 0.1% (w/v). Solutions of nucleic acid were de-proteinised by shaking, or vortexing, with an equal volume of phenol or phenol/ chloroform (1:1). The phases were then separated under centrifugation in a Sorval RC-5B centrifuge (5K rpm, 10 minutes) or a microfuge (12K rpm, 5 minutes) and the aqueous layer carefully removed to a fresh tube.

2.4.1.2 Ether Extraction

Di-ethyl ether was equilibrated with TE buffer. Solutions were shaken, or vortexed, with an equal volume of di-ethyl ether to remove any remaining phenol after de-proteinisation. The phases were allowed to settle before the ether phase was removed. Any remaining ether was removed by heating the sample to 65°C for 2 minutes.

2.4.1.3 Precipitation of Nucleic Acids

The solution was adjusted to 0.3M sodium acetate (pH 5.5) then 2.5 volumes of ethanol or 0.6 volumes of propan-2-ol were added.

The samples were incubated at -20°C for a minimum of 60 minutes, for small quantities of nucleic acids the precipitation was left at -20°C overnight. The precipitate was recovered by centrifugation in a Sorval centrifuge (10K rpm, 20 minutes, 4°C) or a microfuge (12K rpm, 10 minutes). The pellet was washed in 70% ethanol, dried under vacuum and re-dissolved in TE buffer or sterile, distilled water.

DNA was also precipitated from solutions containing nucleoside triphosphates by the addition of an equal volume of 4M ammonium acetate (pH 6.5) and 4 initial volumes of ethanol. After incubation in a dry ice/methanol bath for 15 minutes, the DNA was recovered as above. This precipitation carried out two times is successful in removing >90% of unincorporated labelled triphosphates from reactions (Okayama and Berg, 1982).

2.4.1.4 Restriction Endonuclease Digestion of DNA

In general, DNA was digested with a 2-4 fold excess of the enzyme for a minimum of 60 minutes in reaction conditions recommended by the manufacturer. After digestion, restriction enzymes were removed by phenol extraction (2.4.1.1).

2.4.1.5 Extraction of DNA from Agarose Gels

a) Electroelution: Seachem agarose gels were run (see 2.4.4.1) until the fragments were sufficiently separated and a small trough was cut in front of the desired fragment. The fragment was then electrophoresed into the trough (using a UV illuminator to visualise the ethidium bromide stained DNA) and drawn into a Gilson tip. The DNA solution was then purified using an Elutip (2.4.1.6).

b) Low melting point agarose gels: DNA fragments were fractionated on low melting point agarose gels and the desired fragment was cut from the gel. An equal volume of TE buffer was added and the sample heated to 65°C to melt the agarose. The sample was then phenol extracted (2.4.1.1) and the aqueous phase was further purified using an Elutip column (2.4.1.6).

2.4.1.6 Elutip-D column Purification of DNA

For some purposes (such as oligolabelling and cloning) it was necessary to further purify DNA samples. Elutip-d columns were used as recommended by the manufacturers (Schleicher and Schüll). Columns were washed in high salt buffer (1M NaCl, 20mM Tris-Cl pH 7.5, 1mM EDTA) and then low salt buffer (0.2M NaCl, 20mM Tris-Cl pH 7.5, 1mM EDTA). The DNA sample, in low salt buffer, was applied to the column and then eluted in high salt buffer. Following this, the sample was ethanol precipitated (2.4.1.3).

2.4.1.7 Ligation of DNA

The vector and insert DNA were cut to completion with the appropriate endonucleases. Vector molecules were prevented from self ligation by removing the terminal 5'-phosphate groups by the addition of 0.01 units of calf intestinal phosphatase 5-15 minutes before the digestion time was complete. After restriction, where necessary, DNA fragments were end-filled by the addition of all four dNTP's to 0.25mM and 1-2 units of crude T4 DNA polymerase and incubated at 25°C for 15 minutes. Prior to ligation the vector and insert DNA solutions were usually deproteinised (2.4.1.1) and

ethanol precipitated (2.4.1.3). Typically, between 1-15ng insert DNA were ligated in a reaction with vector to insert DNA concentration at a 3:1 molar ratio. Ligations were carried out in 10-20μl reaction volumes containing 20mM Tris-Cl pH 7.6, 5mM $MgCl_2$ and 5mM DTT, supplemented with 1mM ATP, and incubated overnight at 4°C. Blunt end ligations required 1 unit of T4 DNA ligase and cohesive end ligations 0.01 units. The ligation products were then transformed into *E. coli* (2.4.1.8).

2.4.1.8 Transformation into *E. coli*

L-broth (100ml) was inoculated with 1ml of an overnight culture of the relevant bacterial strain and grown at 37°C with shaking until the OD_{650} reached 0.45-0.55 (or 0.25-0.35 if the strain was NM522). The culture was chilled on ice for 10 minutes and the cells were then pelleted by centrifugation (2K rpm, 10 minutes, 4°C) and resuspended in 50ml of ice-cold, sterile 50mM $CaCl_2$. The suspension was left on ice for 15 minutes and the cells were re-pelleted and resuspended in 2.5ml of ice-cold 50mM $CaCl_2$. One 0.1ml aliquot of this suspension was removed, into a 5ml glass tube, for each transformation, mixed with the ligation mixture and incubated on ice for 30 minutes. The cells were then heat-shocked at 42°C for 2 minutes and returned to ice for 1 minute.

In the case of plasmid DNA transformations, 900μl of L-broth was added to each tube and the cells incubated at 37°C for 1 hour. Samples of 100-200μl were then plated out on to L-agar plates supplemented by the appropriate antibiotic. The plates were inverted and incubated at 37°C overnight.

In the case of bacteriophage M13 DNA transformations, 3ml of molten BBL top agar at 42°C, 25μl of BCIG (25mg/ml), 25μl of IPTG (25mg/ml) and 0.2ml of stationary NM522 cells were added to each tube. The complete mixture was then poured on to BBL plates and allowed to solidify before being inverted and incubated at 37°C overnight.

2.4.2 **Preparation of Samples from *Drosophila melanogaster***

2.4.2.1 Preparation of Genomic DNA from Adult Flies

A rapid method for DNA extraction from single flies was used, essentially this method is that of Marcus (1985). Care was taken throughout to prevent shearing of genomic DNA.

One or two flies were etherised, frozen in liquid nitrogen and homogenised (see 2.4.2.3b) in 50μl of 0.15M NaCl, 50mM Tris.HCl (pH 8.0), 15mM EDTA. A further 50μl of the same buffer containing 0.04% SDS was added. Proteinase K was added to 50μg/ml and the homogenate incubated at 65°C for 30-60 minutes. The mixture was then phenol extracted twice and ethanol precipitated for at least 3-4 hours. The nucleic acid pellet was left to resuspend the TE buffer at 4°C overnight.

Genomic DNA was digested for 2 hours with 10 units of the chosen restriction enzyme, in the appropriate buffer conditions, and RNase A (to 0.1μg/ml) was also added to the reaction mixture.

2.4.2.2 Preparation of RNA from *Drosophila* Adults and Tissues

Usually between 10-20 adult flies, or tissues such as ovaries, were frozen in liquid nitrogen and homogenised (see 2.4.2.3b) in 100-500μl RNA extraction buffer (100mM Tris-Cl pH 7.5, 10mM EDTA, 150mM LiCl , 1% (w/v) SDS).

The homogenate was extracted 3 times with a 1:1 ratio of phenol:chloroform and then ethanol precipitated for at least 4 hours at -20°C. The pellet was washed in 70% ethanol before being resuspended in dH_2O and 2 volumes of 3M Lithium Chloride added. The sample was incubated at 4°C overnight, centrifuged and the pellet washed in 70% ethanol, dried and resuspended in dH_2O.

2.4.2.3 Extraction of Protein Samples from *Drosophila*

a) Collection of haemolymph samples from adults: The flies were attached by their wings to a slide using double-sided adhesive tape. Glass needles were prepared by pulling 50μl glass capillaries in a vertical pipette puller (Model 700C, David Kopf Instruments, Tujunga, California). The haemolymph was collected with a glass needle through the anterior ventral surface of the thorax, placed into polypeptide sample buffer and frozen immediately in liquid nitrogen before storing at -20°C.

b) Preparation of protein samples from adults, tissues and eggs: Adult females 3-4 days old were etherised and placed in a few drops of Ringer's solution in a cavity slide. The ovaries were dissected out using watchmaker's forceps, placed into polypeptide sample buffer and frozen in liquid nitrogen. The ovaries were homogenised in a 1.5ml eppendorf tube with a pasteur pipette the tip of which had been melted into a ball to fit snugly in the bottom of the tube.

Whole flies and fat body samples (carcasses with the ovaries and most of the gut dissected away) were frozen in liquid nitrogen and homogenised in polypeptide sample buffer as for ovaries.

Eggs were collected from adult females raised on yeasty fly flood for 3 days and allowed to lay on egg collection plates for periods of 1-3 hours. Eggs were removed from the plates with a dampened fine-haired paint brush and placed in polypeptide sample buffer, frozen in liquid nitrogen and homogenised.

Samples were stored at -20°C until loaded on to acrylamide gels (2.4.3.3).

c) Ovary cultures: Ovaries (2-12 pairs) were dissected out (as b. above) and cultured in eppendorf tubes in 50-100μl Ringers solution for 1-3 hours at room temperature. The ovaries were then pelleted in a microfuge (12K rpm, 5 min). the supernantant was carefully removed to a fresh tube and an equal volume of polypeptide sample buffer was added. Samples were loaded immediately on to polyacrylamide gels (2.4.3.3).

2.4.3 **Preparation of Plasmid DNA**

2.4.3.1 Small Scale

Small scale plasmid DNA was prepared by modification of the rapid extraction method of Birnboim and Doly (1974).

A single bacterial colony was used to inoculate 5ml of L-broth and grown with shaking, at 37°C overnight. A 1.5ml aliquot of this culture was pelleted in a microfuge (12K rpm, 1 min) and the supernatant discarded. The cells were resuspended in 100μl of 50mM glucose, 10mM EDTA, 25mM Tris.HCl pH 8.0 and 10μl of lysosyme (10mg/ml) were added, vortexed and incubated on ice. After 5 minutes 150μl of 0.2M NaOH, 1% SDS were added, vortexed, and the mixture was incubated on ice for a further 2 minutes before the addition of 150μl of 3M sodium acetate pH 5.5 and vortexing. The sample was then phenol extracted (2.4.1.1), the DNA precipitated by ethanol (2.4.1.3) and the DNA pellet resuspended in 20μl TE buffer. Aliquots of 2μl were used in restriction digests (2.4.1.4) when RNase-A (to 0.1μg/ml) was also added to the reaction mixture.

2.4.3.2 Large Scale

A fresh overnight culture (as small scale) of bacteria containing the plasmid required was used to inoculate 500ml L-broth supplemented with the appropriate antibiotic and grown overnight, with shaking, at 37°C. The cells were pelleted in a Sorval RC-5B centrifuge (5K rpm, 5 minutes) and resuspended in 10ml of ice cold 50mM glucose, 10mM EDTA, 25mM Tris.HCl pH 8.0 with the addition of 200μl lysosyme (5mg/ml). After the suspension was incubated on ice for 5 minutes, 20ml of 0.5M NaOH, 1% SDS were added, with vigorous mixing, and incubated on ice for 5 minutes after which 15ml of 3M sodium acetate pH 5.5 were added and the mixture incubated on ice for 60 minutes. After centrifugation (12K rpm, 20 minutes), the supernatant was phenol extracted (2.4.1.1) and the DNA precipitated by the addition of 0.6 volumes of propan-2-ol and incubation at -20°C for 60 minutes. The DNA was pelleted (12K rpm, 20 minutes), resuspended in 10ml TE buffer and 10g caesium chloride and 1ml of ethidium bromide (10mg/ml) were added. The resulting solution was centrifuged in a Sorval OTD 50B ultracentrifuge at 38K rpm and 18°C, for 48 hours. The DNA bands were visualised using UV illumination and the plasmid band collected. Ethidium bromide was removed by extraction with caesium chloride saturated butan-2-ol, the DNA precipitated by ethanol (2.4.1.3) and resuspended in TE buffer.

2.4.4 **Gel Electrophoresis**

2.4.4.1 Agarose Gel Electrophoresis of DNA

Agarose gels were made and run in 1 x TBE gel buffer. Ethidium bromide was added at a concentration of 0.5mg/ml to both the gel and the running buffer. DNA gel loading sample buffer was added to the DNA samples to 10% (v/v).

Mini-gels were cast and run in a Cambridge Biotechnology model CB1000 mini-gel apparatus. These gels were run at 50-100V for 1-2 hours. Larger agarose gels (16 x 10 x 1cm) were run submerged in home-made gel kits at 100-150V for 1-4 hours.

After electrophoresis DNA bands were visualised on a chromatovue C-70G UV viewing system (254nm short, 265 long).

2.4.4.2 Agarose Gel Electrophoresis of RNA

RNA samples (typically 5-10μg total RNA in dH_2O) were added to an equal volume of formaldehyde sample buffer (FSB) and heated to 60°C for 5 minutes, then

immediately cooled on ice before adding 0.25 volumes of Ficoll-dye-EDTA (FDE). The samples were immediately loaded on to 1.3% (w/v) agarose gels which were made and run in 1 x MOPS buffer. The gels also had formaldehyde added to 17.3% (v/v). The gels (16 x 10 x 1cm) were run submerged in home-made gel kits at 100V for 3-4 hours.

2.4.4.3 SDS-Polyacrylamide Gel Electrophoresis of Proteins (Laemmli, 1970)

Samples for SDS-PAGE analysis (see 2.4.2.3) were collected into 30μl of polypeptide sample buffer, immediately frozen in liquid nitrogen and then stored at -20°C. Immediately before loading on to gels, the samples were boiled for 2 minutes, cooled on ice, vortexed briefly and centrifuged (12K rpm, 1 minute).

Acrylamide gels used for protein separation were linear 10% (w/v) acrylamide gels or gradient (usually 7-20% (w/v)) acrylamide gels. Gradient gels were poured using a two-chambered gradient pouring device via a peristaltic pump (P3-Pharmacia). Gels were run at 140V for 16-18 hours.

After electrophoresis, gels to be stained were fixed in 20% (w/v) TCA for 30 minutes, then stained in 0.1% (w/v) Coomassie Blue in a 5:5:1 mixture of acetic acid, methanol and dH_2O for 1-4 hours. Gels were destained in a 5:5:1 mixture of acetic acid, methanol and dH_2O for 1-6 hours.

2.4.4.4 Polyacrylamide Gels for DNA Sequencing

Sequencing gels (40cm x 20cm x 0.4mm) and (40cm x 30cm x 0.4mm) were 0.6% (w/v) acrylamide (19:1 acrylamide:bis-acrylamide) in TBE gel buffer containing 7.7M urea. Non-gradient gels were poured with 1 x TBE acrylamide gel mix. In buffer gradient gels the top two thirds of the gel contained 0.5 x TBE acrylamide gel mix and the bottom third a gradient of 0.5-2.5 x TBE acrylamide gel mix which was created using a 10ml pipette. The tables below list the volumes of acrylamide mix used and the volumes of APS and TEMED which were required to initiate polymerisation of the mixture.

1 x TBE gels

Gel Dimensions	1 x TBE acrylamide mix	APS (25% (w/v))	TEMED
20cm x 40xm	35ml	80μl	80μl
30cm x 40cm	75ml	140μl	140μl

0.5-2.5 x TBE gradient gels

Gel Dimensions	0.5 x TBE acrylamide mix	2.5 x TBE acrylamide mix	APS (25% (w/v))	TEMED
20cm x 40cm	30ml	-	70μl	70μl
	-	10ml	14μl	14μl
30cm x 40cm	60ml	-	140μl	140μl
	-	20ml	30μl	30μl

Sequencing gels were run at a constant power of 40 watts for times ranging between 1.5-7.5 hours. After electrophoresis, gels were fixed in 10% (v/v) acetic acid, 10% (v/v) methanol for 20-30 minutes, fixed to blotting paper (Ford Goldmedal Blotting Paper), dried down on a BioRad Model 583 Gel Drier, and autoradiographed (2.4.6.3) for between 16 and 18 hours at room temperature.

2.4.5 **Transfers to Membrane Filters**

2.4.5.1 Southern Blotting (Southern, 1975)

Gels for Southern transfers were prepared with Miles Seachem agarose. *Drosophila* genomic DNA samples were run in 1% (w/v) agarose gels (2.4.4.1). When electrophoresis was complete, and after photographing, gels were gently shaken for 30 minutes in 0.5M NaOH, 1.5M NaCl, then soaked for 30 minutes in 1M Tris.HCl pH 5.5, 3M NaCl and washed in 2 x SSC.

A blotting paper wick was placed on a platform supported above a reservoir of 20 x SSC with the ends of the wick immersed in the buffer. The wick was thoroughly wet with 20 x SSC and the gel laid on top. A sheet of Hybond-N membrane filter (Amersham) was cut to the same dimensions as the gel and was wetted in 2 x SSC before being laid on top of the gel. Three sheets of blotting paper the same size as the gel were wetted with 2 x SSC and laid on top of the membrane. Care was taken during these steps to avoid air bubbles being trapped. Further pieces of blotting paper (again with the same dimensions as the gel) were laid, dry, on top of the wet paper to a

thickness of 3-5cm, these were followed by a stack of paper towels to a thickness of 5cm. A sheet of perspex was laid on top of the paper towels and a weight placed upon this. Transfer was allowed to occur for 16-24 hours after which the membrane filter was removed, rinsed briefly in 2 x SSC and wrapped in clingwrap (Vitafilm, Goodyear Tyre and Tubber Co, Ltd, Staffordshire, GB). The DNA was fixed to the filter by UV irradiation in a UV light box (ChromatoVue C-70G, UV viewing system), with the DNA side towards the source of UV light (254nm short and 365nm long), for 5 minutes. The membrane was sealed in a plastic bag until probed (for hybridisation conditions, see 2.4.7.1).

2.4.5.2 Northern Blotting (Thomas, 1980)

Gels for Northern transfers were prepared (2.4.4.2) with Miles Seachem agarose. After electrophoresis, and without further treatment, the gels were transferred to a platform with a blotting paper wick in a reservoir of 20 x SSC. The remainder of the transfer procedure was carried out as described for Southern blotting (2.4.5.1).

2.4.5.3 Western Blotting

Protein gels were run as described (2.4.4.3) and transfer of the electrophoresed proteins to a nitrocellulose membrane was performed electrophoretically in a BioRad Trans-Blot Cell. The nitrocellulose was cut to the same dimensions as the gel and a sandwich of blotting paper/nitrocellulose filter/gel/blotting paper was placed between two scotchbrite pads, all prewetted in transfer buffer, and inserted into the cell such that the nitrocellulose filter was positioned between the gel and the anode. The proteins were electrophoresed on to the nitrocellulose filter at 60V for 4 hours. After transfer, the proteins on the filter were visualised by staining with a solution of 0.3% (w/v) Ponceau S in 3% (w/v) TCA, the filter was rinsed with dH_2O after the effectiveness of transfer had been established.

The filter was shaken, overnight, at 4°C in a solution of 3% (w/v) BSA in TS buffer. The filter was rinsed in five changes of TS over a period of 30 minutes and then sealed in a bag containing a solution of 3% (w/v) BSA in TS and 40μl of rabbit YP3-β-galactosidase antibodies (see Chapter 3) and shaken at room temperature for 4 hours. After rinsing, as before, the filter was sealed in a bag containing 10ml 3% (w/v) BSA in TS and 50μl of goat anti-rabbit IgG-peroxidase conjugate and shaken at room temperature for 1-2 hours. The filter was again rinsed in 5 changes of TS over 30

minutes before developing in 200ml of developing solution (2.1). The reaction was stopped with dH_2O and the filter was dried and stored in the dark.

2.4.6 **Radioactive Labelling of DNA**

2.4.6.1 Radioactive Labelling of DNA by Random Priming (Feinberg and Vogelstein, 1983, 1984)

The DNA fragment to be labelled was denatured by boiling for 2 minutes followed immediately by rapid cooling on ice. Between 50-200ng of DNA was labelled in the reaction in a total volume of 50μl containing 10μl OLB (2.1), 1μl 20mg/ml BSA, 20-50μCi ^{32}P-dCTP, 1μl (5 units/μl) Klenow fragment of DNA polymerase I. The reaction was incubated for 4-16 hours at room temperature.

2.4.6.2 Measurement of Radioactivity Incorporated into DNA

The efficiency of incorporated radiolabel was measured by precipitation with tri-chloroacetic acid (TCA). One aliquot from the reaction mixture was added to 200μl ice cold dH_2O containing 10μg BSA and incubated on ice for 10 minutes. The sample was made 10% (w/v) TCA and the precipitate collected on 2.5cm Whatman GF/C glass fibre circles using a Buchner funnel equipped with a water pump. The discs were washed with 5% (w/v) TCA and dried. Another sample was spotted directly on to a Whatman GF/C glass filter and dried. Each filter was inserted into a scintillation vial, immersed in scintillant (6g/L butyl-PBD in toluene) and the radioactivity measured by counting in a liquid scintillation spectrometer (Intertechnique SL 3000). The percentage incorporation of label into nucleic acid was estimated by comparing the two values and usually probes used achieved 50-90% incorporation and were 10^7-10^9CPM/μl DNA.

2.4.6.3 Autoradiography

Autoradiography of dried DNA sequencing gels, dried protein gels and nitrocellulose or nylon membrane filters was performed using Cronex 4 (Dupont) X-ray film and cassettes. Films were developed in an Agfa 1 automatic film processor.

2.4.7 Hybridisation of Membrane Filters with Radiolabelled Nucleic Acid Probes

2.4.7.1 Hybridisation of Southern Blot Filters

Filters were prehybridised in heat sealable polythene bags (Krups Vacupac) for a minimum of 1 hour at 65°C in 10ml hybridisation solution. Hybridisation solution is 0.5M sodium phosphate pH 7.0, 7% (w/v) SDS with the addition of sonicated salmon sperm DNA to 10μg/ml freshly denatured by boiling for 5 minutes and cooled rapidly on ice. The prehybridisation solution was removed and replaced with 10ml of hybridisation solution and the ^{32}P-labelled probe added. The hybridisation was allowed to proceed at 65°C for 18-24 hours after which the solution was removed and the filter rinsed in 2 x SSC, 0.5% (w/v) SDS, washed in this solution for 60 minutes at 65°C and then in 0.1 x SSC, 0.5 (w/v) SDS for 60 minutes at 65°C. The filter was partially dried with blotting paper and sealed in a fresh polythene bag and autoradiographed (2.4.6.3).

2.4.7.2 Hybridisation of Northern Blot Filters

The procedure for hybridisation of Northern blot filters with a radiolabelled DNA probe is performed in the same manner as the procedure for Southern blot hybridisation with a different hybridisation solution and at a different temperature.

The solution used for prehybridisation and hybridisation of Northern blot filters is 0.4M sodium phosphate pH 7.0, 7% (w/v) SDS, 50% (v/v) formamide with the addition of sonicated salmon sperm DNA to 10μg/ml which has been newly denatured by boiling for 5 minutes and cooled rapidly on ice. Northern blot membrane filters were prehybridised and hybridised at 42°C.

2.4.8 DNA Sequencing by the Dideoxynucleotide Chain Termination Method (Sanger *et al.*, 1977)

2.4.8.1 Preparation of Template DNA

Single recombinant plaques were inoculated into 1.5ml of 2 x TY broth with 50μl from an overnight culture of TG1 cells. The cultures were grown, with rigorous shaking, for 5½ hours and then transferred into 1.5ml microcentrifuge tubes and the cells pelleted by centrifugation (12K rpm, 10 minutes). The supernatants were transferred to fresh tubes and 150μl of 20% (w/v) polyethylene glycol 6000 (PEG 6000), 2.5M NaCl was added to each tube to precipitate the phage. After mixing, the tubes

Table 2.5 Composition of ddNTP Solutions (all volumes in microlitres)

	T mix	C mix	G mix	A mix
0·500mM dTTP	25	500	500	500
0·500mM dCTP	500	25	500	500
0·500mM dGTP	500	500	25	500
10mM ddTTP	40	-	-	-
10mM ddCTP	-	6	-	-
10mM ddGTP	-	-	16	-
10mM ddATP	-	-	-	0.6
SEQTE	935	969	959	500

were left for 15 minutes at room temperature and then centrifuged (12K rpm, 15 min) to pellet the phage. The supernatant was completely removed by aspiration and the pellet resuspended in 100μl SEQTE buffer. The phage suspension was phenol extracted (2.4.1.1), the aqueous phase was ether extracted (2.4.1.2) and finally, ethanol precipitated (2.4.1.3). The pellet was dissolved in 30μl SEQTE buffer and stored at -20°C.

2.4.8.2 Sequencing Reactions

Annealing: Template DNA and primer were annealed in a 10μl volume containing 8μl template DNA solution, 1μl TM buffer, 1μl 17-mer primer (0.2 mol/μl) at 65°C for 30 minutes and allowed to cool to room temperature slowly. The solutions of primed templates were centrifuged briefly to drive any condensation to the bottom of the tube.

Primed synthesis: All reactions were carried out in 1.5ml uncapped Sarstedt tubes in plastic 10-hole centrifuge racks. Components were mixed by brief centrifugation. Aliquots of each annealed template and primer (2μl) were dispensed into the 4 sequence reaction tubes containing 2μl of dideoxy A/G/C/T nucleotide mixes respectively (see Table 2.5). Finally, to each tube 2μl of a solution containing 0.25 units/μl Klenow fragment of DNA polymerase I, 0.5μCi/μl α-^{35}S-dATP, 10mM DTT was added and the different components mixed. After 20 minutes at room temperature, 2μl of chase mix (0.25mM of each dNTP in SEQTE buffer) was added to each tube and mixed by brief centrifugation. After 20 minutes at room temperature the reactions were stopped by the addition of 4μl of formamide loading buffer to each tube (0.3% (w/v) Xylene cyanol, 0.3% (w/v) Bromophenol blue, 10mM EDTA, in 100% deionised formamide). The DNA strands of the polymerised reactions were denatured by boiling for 2 minutes, immediately before loading on to polyacrylamide sequencing gels (2.4.3.4).

2.4.9 **P-Element Mediated Germ-Line Transformation of *Drosophila***

P-element transformation of Oregon-R *Drosophila* embryos was carried out essentially as described by Karess (1985). DNA fragments were subcloned into the transformation vector pUChsneo (Steller and Pirotta, 1985) and the resulting plasmid construct was mixed with a helper plasmid phsπ (Steller and Pirotta, 1985) in a mass ratio of 5:1. Ethanol precipitation (2.4.1.3) was performed and the DNA pellet was thoroughly washed with 70% ethanol, dried and resuspended in 5mM potassium chloride, 100mM sodium phosphate pH 6.8 at a concentration of 1mg/ml. Needles for

injections were pulled from 25μl Drummond microcapillaries using a vertical needle puller (as 2.4.3.3).

OrR embryos were collected from adults which were transferred to yeasty fly food immediately on eclosion and aged for 3 days at 25°C. The flies were transferred to fresh bottles and allowed to lay on lightly yeasted, treacle egg laying plates (Table 2.2) at 18°C for 3 hours and these plates were discarded. Subsequent 1 hour collections were found to contain mainly embryos of stages 1 and 2 (Bownes, 1975). These embryos were manually dechorionated with a damp, finely bristled paintbrush, on double-sided adhesive tape mounted on a microscope slide and aligned on a narrow strip of double-sided adhesive tape mounted on a cover slip. The embryos were dessicated by placing the cover slip in a petri dish containing silica gel dessicant for 5-15 minutes. The cover slip was then mounted on a slide and the embryos immediately covered with Kel-F halocarbon oil.

Once the needle was mounted in the micromanipulator (Leitz) and a sharp point produced by fracturing the tip against the edge of the slide, the needle was filled through the point with the prepared DNA solution of pUChsneo-yp3 construct and helper plasmid. Approximately 1-5% of egg volume was injected into the posterior end of the embryo. Embryos of later than stages 1 and 2 (Bownes, 1975) were discarded and the remaining injected embryos were incubated in a humid chamber for 24-48 hours at 18°C. Hatched larvae were placed in vials of yeasty fly food, incubated at 25°C and when the adults eclosed they were individually crossed with 3 OrR flies, in vials of fly food which was supplemented with Geneticin G418 at 0.5-1.5mg/ml. These vials of flies were heat shocked at 37°C every 48 hours. Any emerging progeny were used to start transformed lines by crossing individually with 3 OrR flies. Stably transformed lines were established and were characterised by Southern blot analysis and genetic analysis.

CHAPTER 3

ANALYSIS OF A YP3 MUTANT OF *DROSOPHILA MELANOGASTER*

INTRODUCTION

The mutant fs(1)A1526 was generated by Gans *et al.* (1975) in a large screen for X-chromosome genes functioning mainly or exclusively during oogenesis. Females homozygous for fs(1)A1526 lay apparently normal eggs which fail to develop beyond the first embryonic cleavage (Zalokar *et al.*, 1975). YP3 is absent from the haemolymph of these females (Bownes and Williams, unpublished observation) and the eggs they lay (see Fig. 3.1a). This chapter describes the characterisation and analysis of the YP3 phenotype in the mutant fs(1)A1526 by genetical and molecular methods.

DOES fs(1)A1526 HAVE MORE THAN ONE MUTATION?

The lesion causing the female sterile phenotype of fs(1)A1526 was mapped to an area around the cut locus (ct^6) at 7B3/4 (20.0 map units) on the X-chromosome (Gans *et al.*, 1975), while the yp3 locus is situated some distance away at 12BC (which is at 44 map units approximately) (see Fig. 3.2). It was unclear whether the mutation causing the female sterility also directly causes the YP3 phenotype seen in fs(1)A1526 females or whether another mutation exists on the same chromosome that is responsible, perhaps at the yp3 locus itself.

It is unlikely a *trans*-acting mutation would affect the synthesis of only one of the YP s and not the entire family since it is reasonable to expect that the three genes respond in a similar manner to any regulatory molecules. This is borne out by the observation that the three members of the yp gene family are co-ordinately expressed in the fat body (Isaac and Bownes, 1982). Further support for this idea comes from a study undertaken to identify *trans*-regulators of the yp s where only genes affecting the synthesis of all the YP s have so far been identified (Shirras and Bownes, 1989). Similarly, it is difficult to envisage how a mutation of any molecule involved in processes such as mRNA stability or protein transport could result in altered regulation of yp3 whilst functioning normally for yp1 and 2.

Few other strains have been identified where one of the YP's has been found to have an altered profile while the other two are apparently normal. Two of these strains, fs(1)1163 (Gans *et al.*, 1975) and fs(1)K313 (Komitopoulou *et al.*, 1983), are female sterile and were noted as defective in secretion of YP1 and YP2, respectively, from the fat body (Bownes and Hodson, 1980; Williams *et al.*, 1987). However, unlike

Figure 3.1

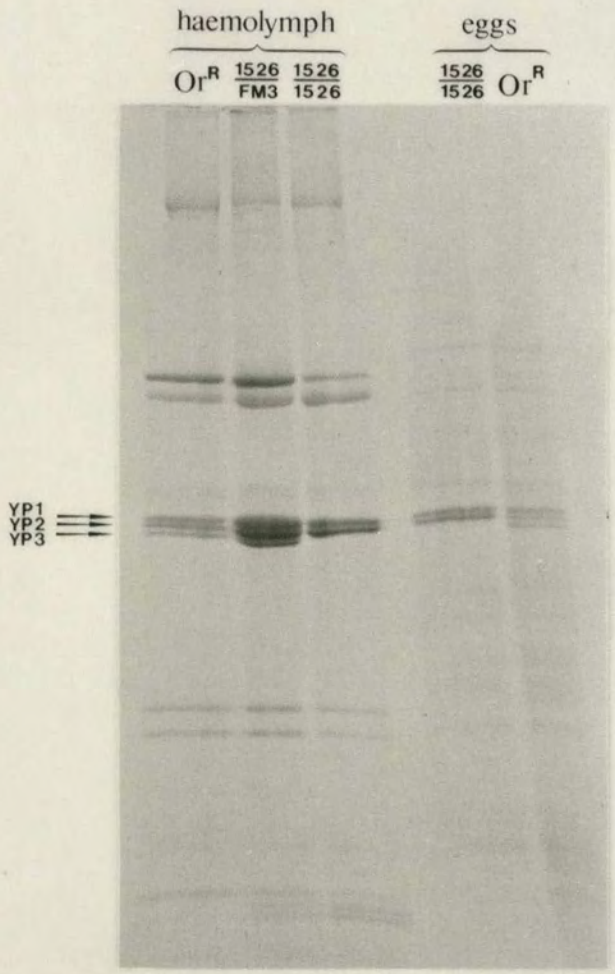
A. Coomassie blue stained 7-20% acrylamide gradient protein gel of haemolymph and egg samples from wild type (Or^R) and fs(1)A1526 females.

B. Coomassie blue stained 7-20% acrylamide gradient protein gel of female *Drosophila melanogaster* haemolymph and egg samples, from wild type flies (Or^R), homozygous fs(1)A1526 flies ($^{1526}/_{1526}$) and heterozygous flies carrying the fs(1)A1526 mutation over the Df(1)HA92 deletion chromosome ($^{HA92}/_{1526}$).

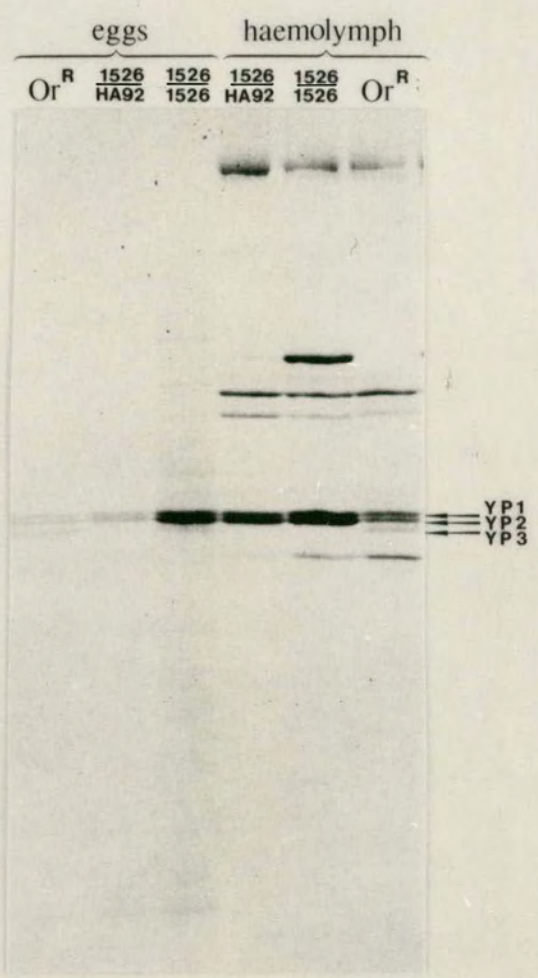
The genotype of the flies are indicated at the top of the gel. The positions of the YP s are indicated.

Each haemolymph lane represents pooled samples from 4-5 individuals and each egg lane shows the proteins from ~10 eggs.

A


haemolymph
eggs
Or^R
1526/FM3
1526/1526
1526/1526
Or^R
YP1
YP2
YP3

B


eggs
haemolymph
Or^R
1526/HA92
1526/1526
1526/HA92
1526/1526
Or^R
YP1
YP2
YP3

fs(1)A1526, in both these cases the mutation responsible for the female sterility maps to the region of the yp1 and 2 loci. Further analysis of fs(1)1163 and fs(1)K313 suggested that DNA sequence variations within the coding regions of the yp1 and 2 genes which result in amino acid substitutions may be responsible for the inefficient secretion of these proteins from the fat body cells (Saunders and Bownes, 1986; Williams *et al.*, 1987). It should be noted that these DNA sequence changes have not been conclusively proven to be responsible for the mutant phenotypes and the possibility exists of another, closely situated mutation on the fs(1)1163 and fs(1)K313 chromosomes that is responsible for yolk protein secretion defects observed in these mutant strains. However, the isolation of such double mutations twice independently makes this seem quite unlikely.

A further question concerning the fs(1)A1526 strain is whether or not the absence of YP3 is itself responsible for the observed female sterility. Only one other strain has been reported which is lacking in one of the YP's. The absence of YP2 was noted in a female sterile strain and although the two effects were reported as having been separated from one another (Tamura *et al.*, 1985), no detailed information has been published. The secretory defect in fs(1)1163 and fs(1)K313 flies results in a reduced level of YP1 or YP2, respectively, in the eggs laid by the mutant females. It has not been established whether or not the reduction in the amount of one of the YP's in the eggs causes or contributes to the female sterility. When this study of fs(1)A1526 was undertaken it was not clear if all three YP's are essential for normal development of *Drosophila* eggs.

In an attempt to resolve these queries concerning fs(1)A1526, a cross was performed to obtain females which possessed one fs(1)A1526 X-chromosome and another X-chromosome that carries a deletion which includes the yp3 locus. The chromosome chosen for this was Df(1)HA92, which is deleted for the region 12A6/7 to 12D3 (and is s^+ g^- ty^- na^+) (see Fig. 3.2).

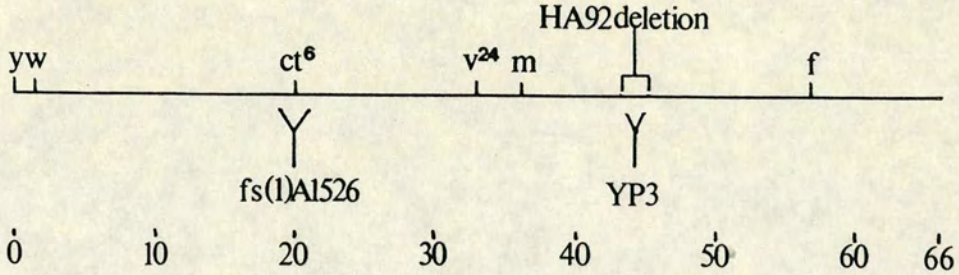
Since the female sterility of fs(1)A1526 was mapped near the 7B region, there should be rescue of this phenotype by the Df(1)HA92 X-chromosome. If the mutated locus at 20 codes for a factor which regulates the expression of yp3 and the fs(1)A1526 yp3 itself is normal, then the product of the wild type copy of this locus (supplied by the Df(1)HA92 chromsome), will be able to exert its effect on $yp3^{1526}$ (whether at the level of transcription, or mRNA stability etc.) and YP3 will be present in the haemolymph of females that are fs(1)A1526/Df(1)HA92. Alternatively, if an additional

Figure 3.2

Diagram (a) and Table (b) of the X-chromosome of *D. melanogaster* showing the positions of various loci mentioned in the text.

A. Schematic representation of the *D. melanogaster* X-chromosome. The positions of various loci and the deleted region of Df(1)HA92 are indicated. the numbering indicates map units by recombination.

A



B

LOCUS	SITE (by recombination mapping)	CYTOLOGICAL POSITION
y (yellow)	0.0	1A5/8
w (white)	1.5	3C2/3
ct (cut)	20.0	7B3/4
m (miniature)	36.1	10E1/2
f (forked)	56.7	15F1/2
v (vermillion)	33.0	10A1/2
fs(1)A1526	19-21	-
yp3		12B/C
Df(1)HA92	-	12A6/7 - 12D3
s (sable)	43.0	
g (garnet)	44.4	12B6/7
ty (tiny)	44.5	
na (narrow abdomen)	45.2	

Figure 3.3

Crossing scheme used to generate:-

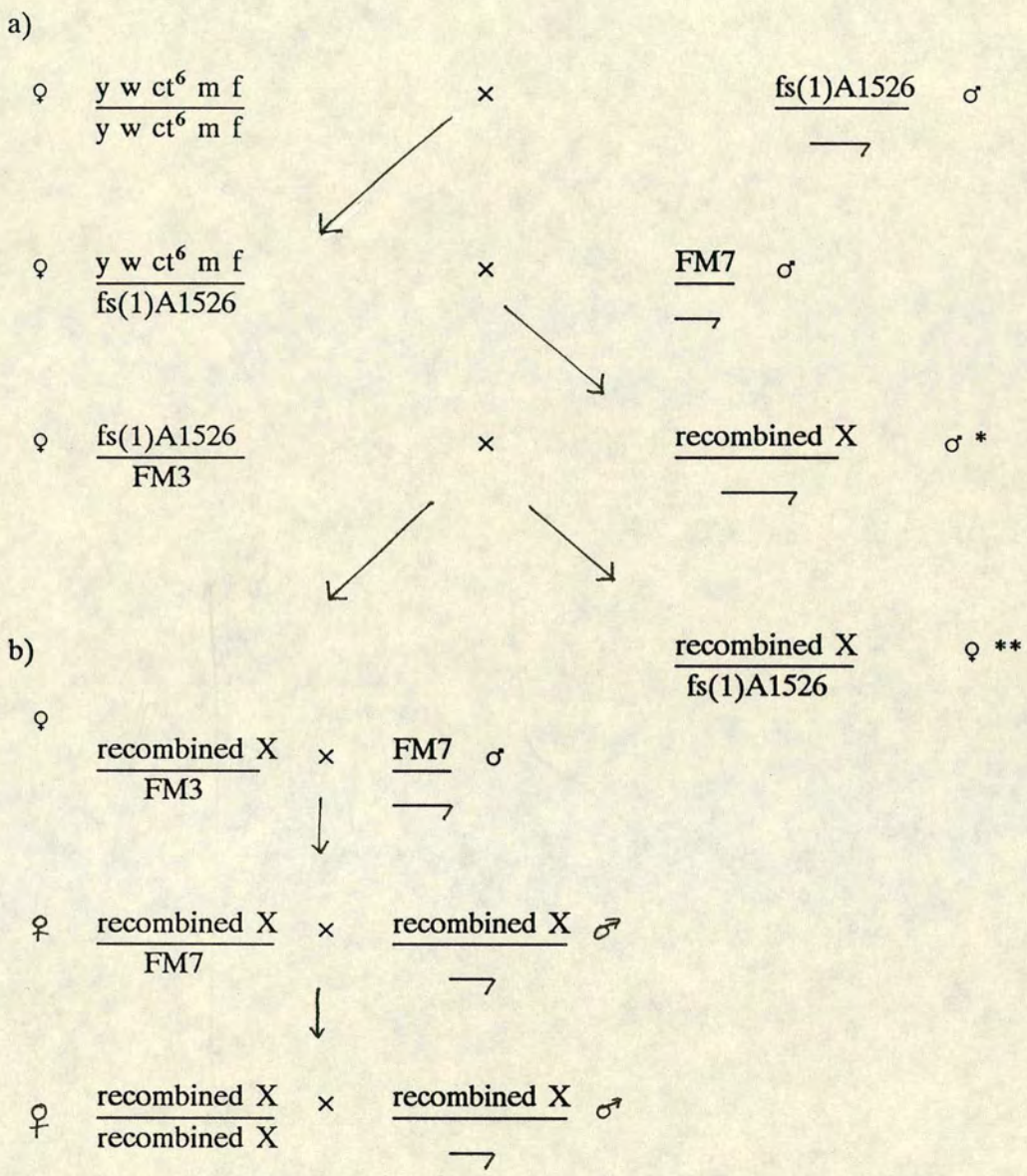
a) flies recombinant for the X-chromosome, to separate the two mutations on the fs(1)A1526 chromosome, and

b) suitable homozygous recombinants (i.e. fertile and lacking YP3 in the haemolymph).

* at this stage individual males with a recombined X-chromosome were mated with fs(1)A1526 /FM3 females.

** these females were tested for fertility and their YP3 profile determined; where suitable, females from the same line, balanced with the FM3 chromosome, were used to generate a strain homozygous for the recombined X-chromosome.

Figure 3.3


a)
♀ y w ct⁶ m f / y w ct⁶ m f × fs(1)A1526 ♂
♀ y w ct⁶ m f / fs(1)A1526 × FM7 ♂
♀ fs(1)A1526 / FM3 × recombined X ♂ *
b)
recombined X / fs(1)A1526 ♀ **
♀ recombined X / FM3 × FM7 ♂
♀ recombined X / FM7 × recombined X ♂
♀ recombined X / recombined X × recombined X ♂

mutation exists on the fs(1)A1526 X-chromosome at the yp3 locus, and it is responsible for the absence of YP3, then this phenotype will not be rescued since this region is deleted from the Df(1)HA92 X-chromosome.

Male flies which were fs(1)A1526 were mated with females which were Df(1)HA92/FM7. Female progeny from this cross that were fs(1)A1526/Df(1)HA92 were found to lack YP3 in their haemolymph (Fig. 3.1b) and were at the same time completely fertile. Since the fertility is rescued and YP3 is not found in the haemolymph and eggs of these flies, this demonstrates that the lesion responsible for the observed female sterility has no effect on the synthesis, processing or stability of the yp3 mRNA or protein. A further mutation must therefore lie on the fs(1)A1526 chromosome, within the area deleted in the deficiency chromosome, that is responsible for the absence of YP3 in the haemolymph of these flies.

In addition, this result clearly indicates that YP3 is not an absolute requirement for fertility and that *Drosophila melanogaster* eggs are viable in the absence of one of the three yolk proteins.

CONSTRUCTION OF A YP3 MUTANT STRAIN THAT IS FERTILE

Since the lesion responsible for the mutant YP3 phenotype was located within the region deleted in Df(1)HA92 and is therefore over 20 recombination map units distant to that causing the female sterility observed in fs(1)A1526 flies, it was possible to obtain, by recombination, a fertile strain which lacks YP3 in the haemolymph.

The fs(1)A1526 chromosome was recombined with a multiply-marked X-chromosome, the strain chosen for this carries markers for yellow (1, 0.0), white (1, 1.5), ct^6 (1, 20.0), miniature (1, 36.1) and forked (1, 56.7) (Stock No. 334000 from Umea stock centre). Females with a recombined X-chromosome and the fs(1)A1526 X-chromosome were obtained (as shown in Fig. 3.3a) and investigated. Haemolymph samples removed from those females that displayed normal fertility were subjected to SDS-PAGE to determine whether YP3 was absent from the haemolymph of any of the lines. Only a small number of fertile lines had to be analysed in this way before several were found where YP3 was not present in the haemolymph (e.g. Fig. 3.4). Two of these lines were chosen to be propagated as homozygous stocks (as shown in Fig. 3.3b), one of these lines carries the markers y wt ct^6 v^{24} $yp3^{1526}$ and the other ct^6 v^{24} $yp3^{1526}$.

Figure 3.4

Coomassie blue stained 7-20% gradient acrylamide gel of haemolymph samples from fertile females. Two of these strains (14 and 19) are homozygous for a recombined X-chromosome. The genotypes of the strains of flies are as follows:-

Or^R: Oregon R, a wild type strain.

14 - y ct^6 v^{24}

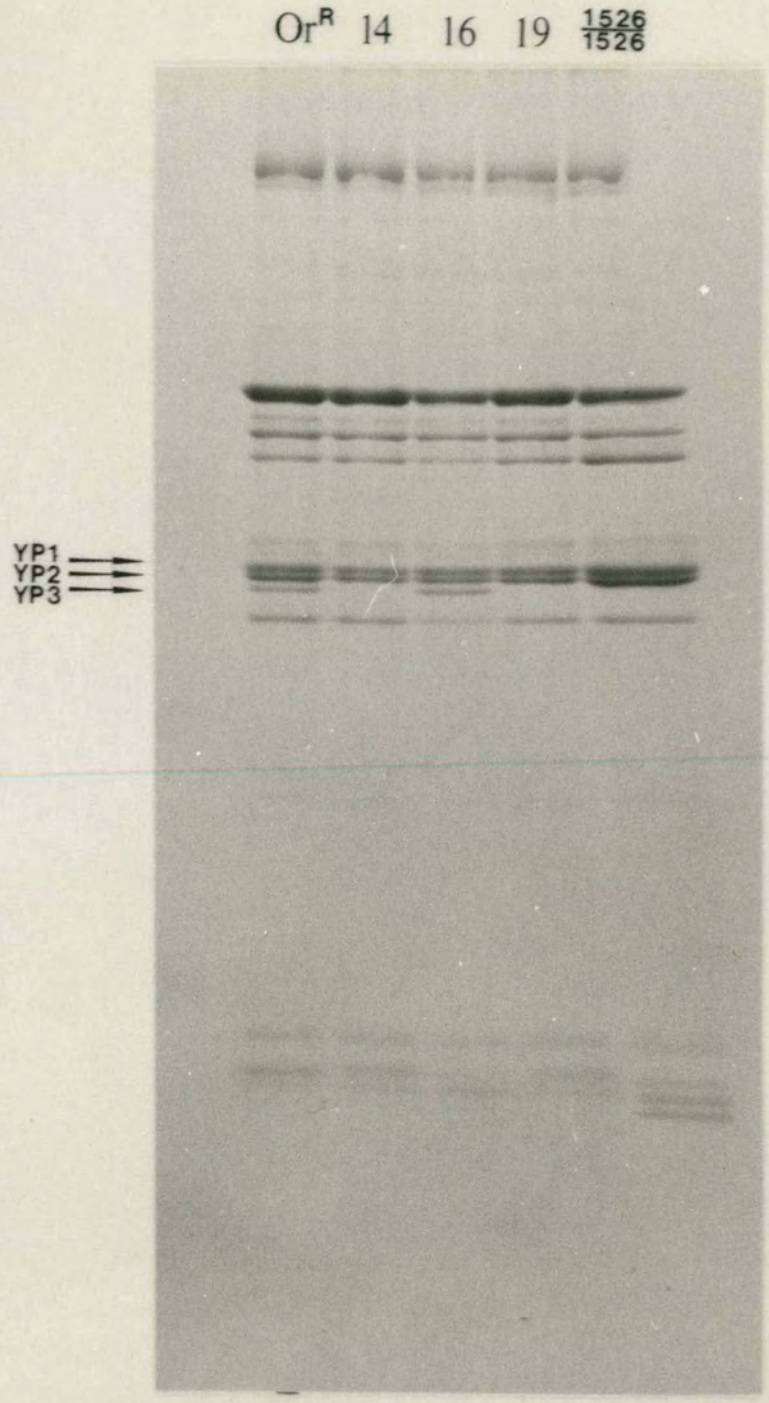
16 - y w ct^6 m f

19 - y w ct^6

1526 : Homozygous fs(1)A1526 flies.

The position of the YP s are indicated.

Each track represents the proteins in the haemolymph of 4-5 females.


Or^R 14 16 19 1526/1526
YP1
YP2
YP3

INVESTIGATION OF YP3 TRANSCRIPTS IN fs(1)A1526

In an attempt to determine at what level the expression of yp3 is affected in fs(1)A1526 flies, northern blot analysis was carried out. In this, RNA samples were probed with a fragment of yp3 DNA which consists of some 5' flanking sequences and the first exon of the gene. This represents the area of the yp3 gene that displays least homology to the yp1 and yp2 sequences and when hybridised against a southern blot of plasmids containing the genes for yp1, 2 and 3, it hybridises only to its own sequences (Fig. 3.5b). The southern blot was reprobed with an equivalent fragment from the 5' end of the yp1 gene and this hybridised to only fragments containing yp1 sequences on the filter (Fig. 3.5c). The fragments used as yp1 and yp3 specific probes are shown in Fig. 3.5.

Total RNA was extracted from adult females homozygous for fs(1)A1526 and from Oregon R females, blotted and probed with the yp3 specific fragment. The autoradiograph (Fig. 3.6a) clearly shows that in the fs(1)A1526 homozygous females there is a reduced level of yp3 transcript relative to the wild type levels seen in the OrR females. To ensure that this is due to a genuine decrease in the number of yp3 transripts in fs(1)A1526 females and not as a result of the flies being non-vitellogenic, or simply due to a difference in the loading of RNA in the tracks, the filter was subsequently reprobed with the yp1 specific fragment. The autoradiograph of this northern (Fig. 3.6a) indicates that the fs(1)A1526 females were vitellogenic since yp1 transcripts are present, and that the tracks are equally loaded, and so demonstrates that less yp3 mRNA is present in the mutant females compared to the level in wild type flies.

Further Northern blot analysis was performed to determine whether both of the tissues normally involved in YP synthesis are producing the yp3 transcripts that are found in fs(1)A1526 females. In order to do this RNA was extracted from female fat body (carcass with the ovaries and most of the gut dissected away) and ovary samples from fs(1)A1526 and OrR adult females. Duplicate nothern blot filters were hybridised with the yp3 specific probe and an alcohol dehydrogenase (Adh) probe. The purpose of using the Adh probe in this instance was to check the loading of the RNA on the filter to allow a quantitation of the level of yp3 message relative to a transcript that should be equally abundant in the mutant and wild type flies. This was carried out since it is not known whether the reduction in the transcript or protein level of one of

Figure 3.5

Southern blot of yolk protein gene containing plasmids digested with HindIII.

A. Agarose gel showing the plasmids stained with Ethidium Bromide.

B. The same gel blotted and hybridised with the yp3 5′ probe.

C. The same filter stripped and hybridised with the yp1 5′ probe.

The tracks are as follows:-

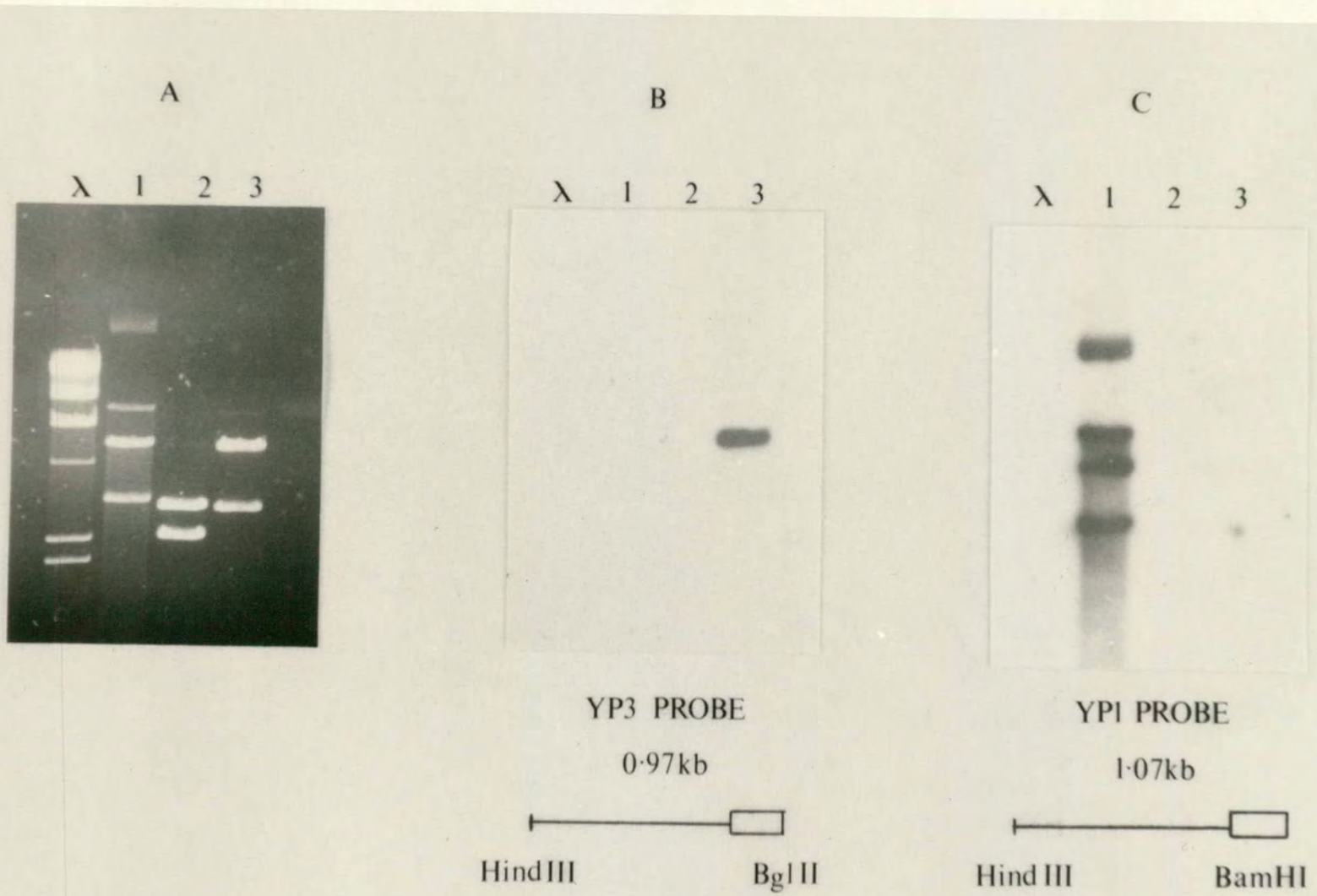
λ = λDNA digested with HindIII to provide size markers (sizes , in kbp, are 23.1, 9.4, 6.6, 4.4, 2.3, 2.0 and 0.5).

1 = plasmid containing yp1 sequences; all 4 bands contain some yp1 sequences so the yp1 probe hybridises to all 4 bands (c).

2 = plasmid containing yp2 sequences. The upper (2.9kb) band represents the plasmid sequence and the lower (2.1kb) band represents yp2 sequences.

3 = plasmid containing yp3 gene. The upper (4.3kb) band contains yp3 sequences and the lower (2.9kb) band encodes plasmid sequences.

The diagram below B and C shows the DNA fragments that were used as the yp3 5′-probe and the yp1 5′-probe. The open boxes indicate coding sequences and the single lines represent 5′ upstream sequences.


A
λ 1 2 3
B
λ 1 2 3
YP3 PROBE
0·97kb
Hind III
Bgl II
C
λ 1 2 3
YP1 PROBE
1·07kb
Hind III
BamHI

the yolk proteins can affect the transcription of the other two. The results of this experiment (Fig. 3.6b and c) reveal that both the fat body cells and the ovarian follicle cells of adult female fs(1)A1526 flies are synthesising yp3 mRNA.

The autoradiographs were subjected to densitometry in order to quantitate the reduction in level of yp3 transcript found in fs(1)A1526 females relative to that found in wild type flies. Densitometric scans, using the yp1 and Adh transcript levels to calibrate the RNA loading, indicated that the mutant females have between 30-50% of the wild type level of yp3 transcripts.

The northern blot experiments presented here illustrate that in adult female flies of the mutant fs(1)A1526, the level of yp3 transcript is considerably reduced compared to that found in wild type females, indicating that either fewer transcripts are synthesised from yp3 in fs(1)A1526 females or the yp3 mRNA produced has a higher rate of turnover than normal. This is the first mutant which has been reported to have a reduced transcript level for only one of the yp s. The level of transcription from the yolk protein genes has been shown to be alterable by, for example, treatment with the hormone ecdysone (Bownes *et al.*, 1983), or starvation (Bownes and Blair, 1986; Bownes *et al.*, 1988). However, in these circumstances, all three yolk proteins are affected. In fs(1)1163 mutant flies where the export defect results in a higher level of YP1 in the fat body cells than seen in wild type flies (Bownes and Hodson, 1980), the yp1 mRNA level does not appear to be affected (Minoo, 1982).

ECDYSONE INDUCED SYNTHESIS OF YP3 mRNA IN fs(1)A1526 MALES

Males of *D. melanogaster* do not normally synthesise YP s. However, injection with 20-hydroxyecdysone results in the appearance of YP s in the haemolymph (Postlethwaite *et al.*, 1980), and of YP transcripts in the fat body (Bownes *et al.*, 1983). Since ecdysone also induces a response in female flies resulting in an increase in the level of yolk protein transcripts (Bownes and Blair, 1983), it is possible that the reduced level of yp3 transcripts in fs(1)A1526 females is due to a mutation in a *cis*-acting sequence of the yp3 gene that is normally involved in the response to ecdysone.

To determine whether the yp3 gene in fs(1)A1526 flies is capable of responding to 20-hydroxyecdysone, RNA was prepared from fs(1)A1526 and wild type males which had been injected with the hormone. The RNA samples were northern blotted and hybridised with the yp3 probe. The result of this experiment (Shown in Fig. 3.6d)

Figure 3.6

A. Northern blot of RNA samples extracted from whole wild type (Wt) females and homozygous fs(1)A1526 females (M) as indicated at the top of the lanes. The blot was hybridised with the yp3 specific probe and subsequently stripped and rehybridised with the yp1 probe. Each lane represents approximately 8μg of total RNA.

B, C. Duplicate Northern blots of RNA samples extracted from female fat body tissues (B) and ovaries (C). The filters were hybridised with the yp3 probe or an Adh probe. Each lane represents approximately 8μg of total RNA.

D. Northern blot of RNA extracted from 20-HO-ecdysone induced (♂E) and non-induced males (♂), and from ovaries of Or^R females as a control. Both fs1(A)1526 and Or^R males were tested, as indicated at the top of the blot. The filter was hybridised with the yp3 5′-probe. Each lane represents approximately 8μg of total RNA.

Wt = wild type

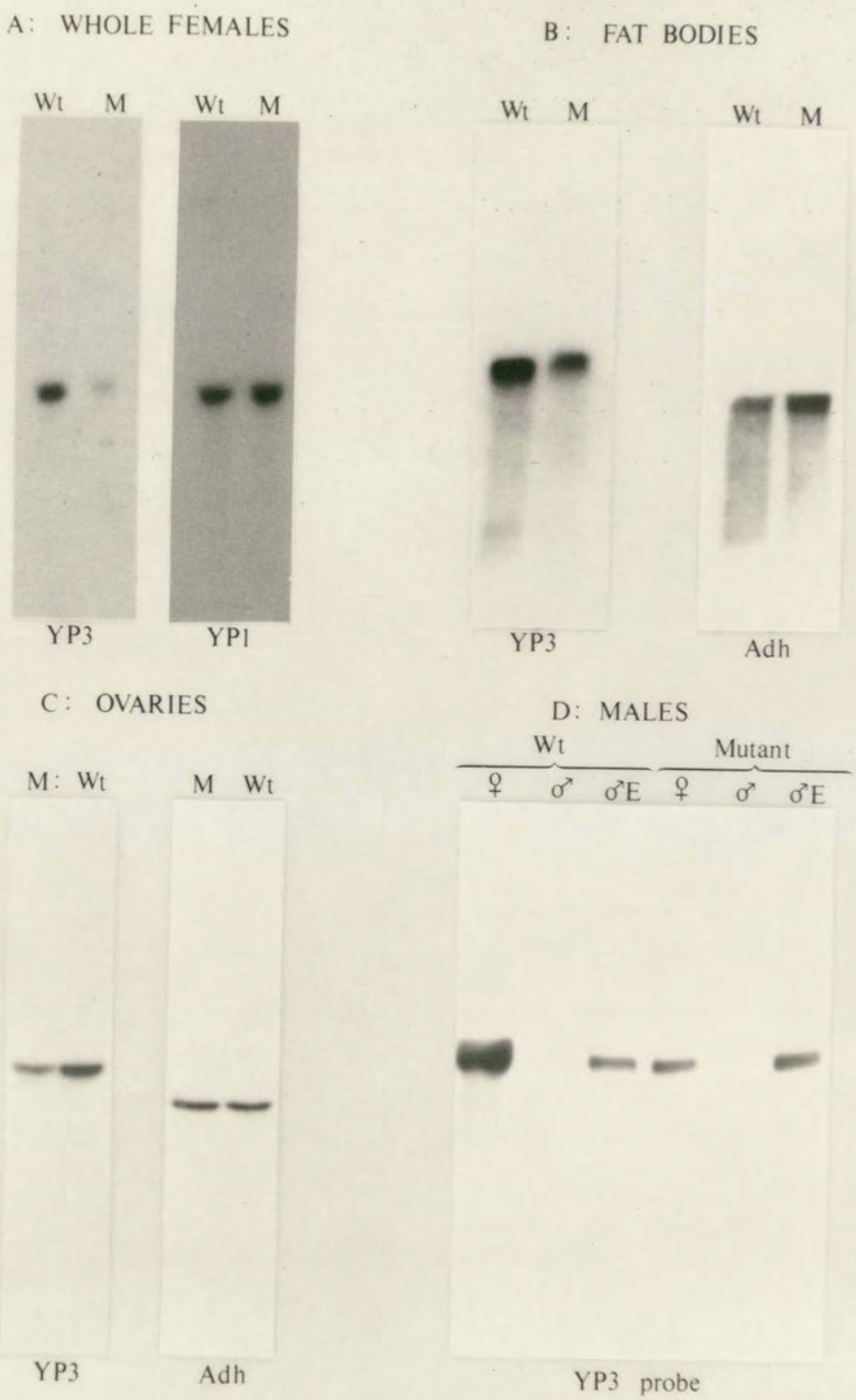
M = fs(1)A1526 mutant

♀ = total RNA extracted from adult females.

♂ = total RNA extracted from adult males.

♂E = total RNA extracted from adult males 24hrs after treatment with 20-hydroxyecdysone (and maintained at 18°C).

The Adh transcript is 1.15kb in length.


A: WHOLE FEMALES
Wt M Wt M
YP3 YP1
B: FAT BODIES
Wt M Wt M
YP3 Adh
C: OVARIES
M: Wt M Wt
YP3 Adh
D: MALES
Wt Mutant
♀ ♂ ♂E ♀ ♂ ♂E
YP3 probe

clearly shows the presence of similar amounts of yp3 mRNA in the mutant and wild type ecdysone treated males, while the untreated males show no yp3 transcripts. This indicates that the DNA sequences involved in at least the male response to the steroid hormone ecdysone are apparently normal in the mutant flies.

INVESTIGATION OF YOLK PROTEIN 3 IN fs(1)A1526

The use of SDS-PAGE revealed that in homozygous fs(1)A1526 females YP3 is absent from the haemolymph and the eggs (Fig. 3.1). An antiserum raised to a YP3-β-galactosidase fusion protein was obtained (generous gift of J.M. Axton and D. Glover) and used to further characterise the YP3 phenotype in the mutant females. The antiserum was raised in rabbits to a fusion polypeptide produced in bacterial cell culture from an expression construct bearing a fragment of yp3 cDNA fused to the *E. coli* β-Galactosidase coding sequences. The yp3 DNA fragment is approximately 600nt long and encodes 76 residues of exon 3 and all (or almost all) of exon 2 ($\geq$124 residues). This means that the antiserum is not actually specific to YP3 but has some cross-reaction to YP1 and YP2 since some of the sequences it was raised against are common to all three YPs.

A western blot was carried out using samples of haemolymphs and homogenised female carcasses (i.e. with the ovaries and some gut removed) from Oregon R and fs(1)A1526 homozygotes. This western (Fig. 3.7a) reaffirms that YP3 is absent from the haemolymph of fs(1) A1526 females. However, whilst the carcasses of OrR females show only a small amount of YP3, the antiserum reacts with an abundant protein present in the carcasses of fs(1)A1526 females. This protein is of a higher molecular weight, by approximately 0.5-1kD, than the YP3 in wild type flies. It is unlikely that this protein is YP2, YP1 or any protein other than YP3 not only because the antiserum was raised against YP3, but because none of the other samples on the western (which do contain YP1 and 2) show such a strong cross-reaction to a similarly sized protein. It is therefore most probable that the protein in the carcasses of fs(1)A1526 females is YP3 (henceforth, this protein is designated $YP3^{1526}$).

In wild type flies the YPs are synthesised and secreted from the fat body cells and there is no apparent storage of these proteins in this tissue. This is illustrated in the OrR female carcass track in Figure 3.7a and c where only a small amount of YP3 is present. However, the $YP3^{1526}$ found in the carcasses of the mutant flies is quite

Figure 3.7

A. Western blot of haemolymph samples from 3-4 females (the two left hand side tracks) and carcasses of 2 females (the two right hand side tracks) of OrR and fs(1)A1526 females. The antiserum used was the YP3-β-galactosidase antiserum, described in the text.

B. Haemolymph samples from wild type and fs(1)A1526 females run on the same gel as (A). This time an antiserum raised against all three yolk polypeptides (Bownes *et al.*, 1987) was used.

C. Western blot of samples from fs(1)A1526 and OrR females as indicated. The YP3-β-galactosidase antiserum was used.

In each case:

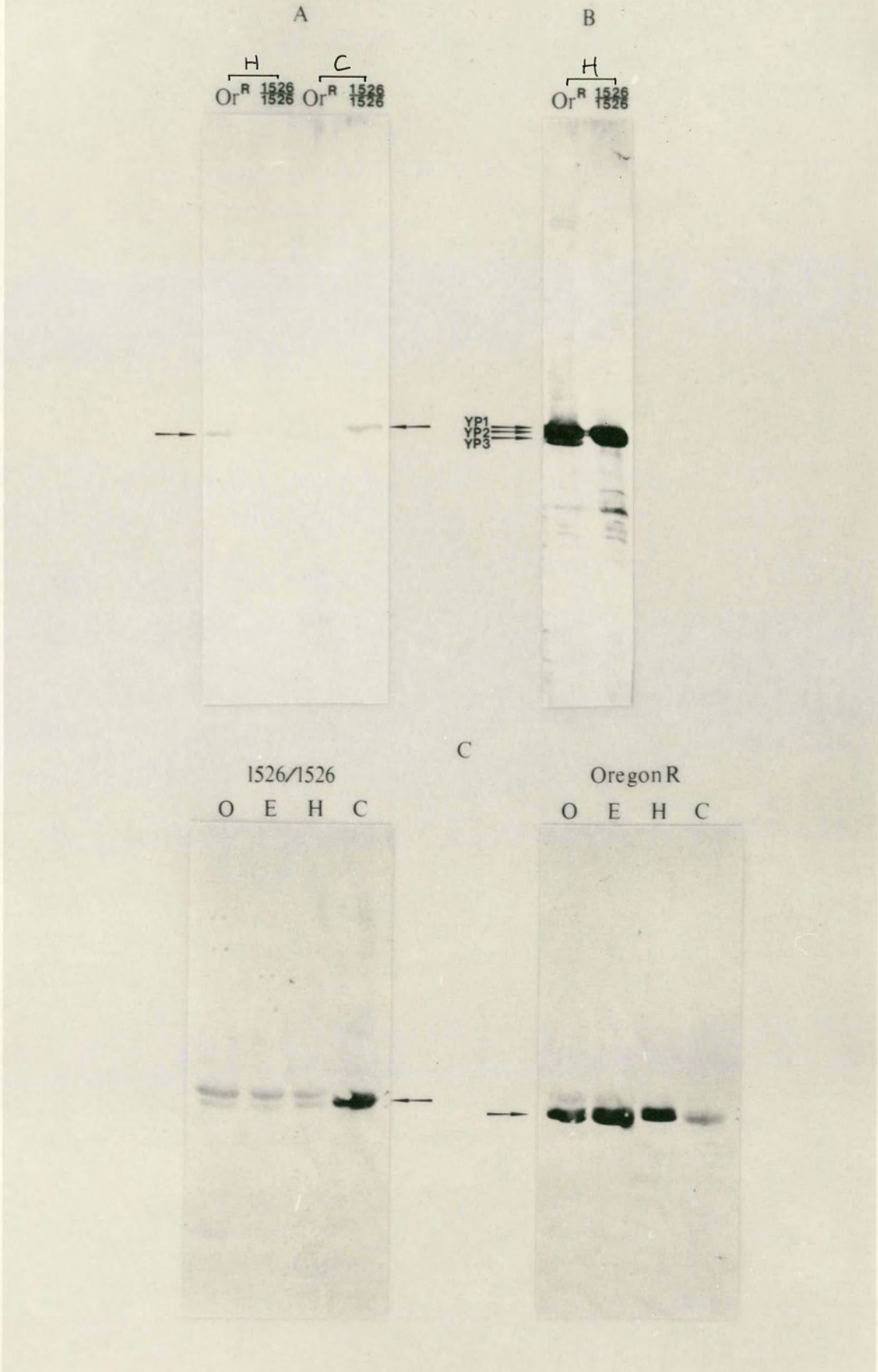
E - the egg lane; represents 40 eggs

O - the ovary lane; represents 1 pair of ovaries

H - the haemolymph lane; represents the haemolymph from 3-4 females

C - the carcass lane; represents 1 fly carcass (i.e. ovaries and gut dissected away)

The positions of the YPs are indicated on (B) and the arrows of (A) and (C) indicate the position of the YP3 band.


A
B
H
C
Or^R 1526/1526 Or^R 1526/1526
H
Or^R 1526/1526
YP1
YP2
YP3
C
1526/1526
O E H C
Oregon R
O E H C

abundant so it is possible that in these flies the $YP3^{1526}$ produced is not secreted and accumulates within the fat body cells.

In an attempt to determine if the ovarian follicle cells of the mutant females also synthesise $YP3^{1526}$, a western blot was carried out using various samples, including ovaries and fat bodies from OrR and fs(1)A1526 females. The result of this (Fig. 3.7c) confirms that no YP3 is present in the eggs laid by fs(1)A1526 females. It would appear from this western that the ovaries of fs(1)A1526 females do not synthesise either wild type YP3 or $YP3^{1526}$. However, it should be remembered that only a small subset of cells in the ovaries actually synthesise YP's, only for a relatively limited amount of time (stages 8-11 of oogenesis), and degenerate shortly afterwards. Also, in wild type ovaries YP3 is synthesised at a reduced level compared to the amount of YP1 and 2 (Isaac and Bownes, 1982; Brennan *et al.* 1982). Therefore, it is possible that the ovaries of the mutant females synthesise $YP3^{1526}$ which is retained in the follicle cells (none is found in the eggs) and since these cells degenerate soon after YP synthesis, it is entirely possible that any $YP3^{1526}$ produced would remain undetected by this method. This would seem to be fairly likely as yp3 mRNA is synthesised in the ovaries of fs(1)A1526 flies (Figs. 3.6).

SUMMARY

This chapter has been concerned with the investigation and characterisation of the YP3 mutant phenotype observed initially in the strain fs(1)A1526.

Rescue of the female sterile phenotype by a chromosome deficient at the yp3 locus revealed the existence of a second mutation, in the region between 12A6/7 and 12D3, on the fs(1)A1526 chromosome that is responsible for the YP3 mutant phenotype. This showed that the mutation causing the female sterility which maps nearby the ct^6 locus is not responsible for the absence of YP3 from the haemolymph of these flies. That the second mutation was located approximately 20 map units away from the ct^6 locus allowed the separation of the two mutant phenotypes and construction of a fertile strain which does not possess any haemolymph or egg YP3. This strain was used as reported in chapter 5 for *Drosophila* transformation experiments with the complete Canton-S yp3 gene.

Northern blot analysis was used to demonstrate that yp3 mRNA is present at reduced levels in the fat body and ovarian follicle cells in fs(1)A1526 females indicating

that a mutation results either in less synthesis of $yp3^{1526}$ transcripts, or an increased rate of turnover of the $yp3^{1526}$ message. Detection of yp3 mRNA in fs(1)A1526 males treated with the steroid hormone ecdysone indicated that the mutation does not affect the yp3 DNA sequences that are involved in the male response to the hormone.

Western blotting revealed that while no YP3 is found in haemolymph or egg samples, a protein cross-reacting with antisera raised to a yp3-β-galactosidase fusion polypeptide is found in the carcasses of the mutant flies. This protein is presumably located in the fat body cells and ovarian follicle cells, the normal sites of YP synthesis. This indicates that YP3 is produced in fs(1)A1526 females but is not exported. Since the protein detected is of a higher molecular weight (by around 0.5-1kD) than is normally observed for YP3, this suggests that the polypeptide is not being properly modified.

In conclusion the data presented in this chapter - that the mutation causing the YP3 phenotype is situated in the region of the yp3 locus, that the abundance of the yp3 mRNA is affected, and that the polypeptide produced is not properly processed or secreted into the haemolymph or oocytes - taken together, indicate that the mutation responsible lies within the yp3 gene itself. The following chapter discusses the identification, by DNA sequence analysis, of a lesion highly likely to cause some or all of the mutant phenotypes of YP3 first seen in fs(1)A1526 adult females.

CHAPTER 4

MOLECULAR CLONING AND SEQUENCE ANALYSIS OF $yp3^{1526}$

INTRODUCTION

In fs(1)A1526 female flies the mutation causing the YP3 mutant phenotype lies within the area of the X-chromosome deleted in Df(1)HA92, the level of YP3 mRNA is reduced, and the polypeptide is produced but fails to be secreted. It therefore seems likely that the molecular lesion responsible for the mutant phenotype of YP3 in fs(1)A1526 lies within, or very near to, the yp3 gene. Such a lesion could be identified by comparison of the mutant and wild type yp3 DNA sequences. This chapter describes the analysis of the $yp3^{1526}$ gene by molecular cloning and sequencing.

SEQUENCE ANALYSIS OF $yp3^{1526}$

A recombinant clone containing to the $yp3^{1526}$ had already been isolated (by J. Williams, in this laboratory) from a genomic library made with DNA obtained from homozygous fs(1)A1526 females and using the immunity insertion vector λNM1149 (Murray, 1983).

The recombinant lambda $yp3^{1526}$ clone was used as a source of $yp3^{1526}$ DNA sequences for subcloning into the Gemini 1 plasmid vector (Melton, 1984) and into the bacteriophage M13 (M13mp18 and M13mp19) (Yannish-Perron, 1983) for use in DNA sequencing.

The $yp3^{1526}$ DNA sequence was determined using the di-deoxynucleotide chain termination method (Sanger *et al.*, 1977; Biggin *et al.*, 1983), the strategy used is outlined in figure 4.1.

The DNA sequence obtained was compiled and compared to the published Canton-S yp3 sequences (Garabedian *et al.*, 1987) using the University of Wisconsin Genetics Computer Group Sequence Analysis Software Package Version 5.2.

The DNA sequence of $yp3^{1526}$ is presented in Appendix I alongside the Canton-S yp3 sequence, areas where the 1526 sequence diverges from the wild type are indicated. The sequence changes identified in the $yp3^{1526}$ gene compared with the Canton-S sequence are presented in figures 4.2 - 4.4.

A total of 9 bases differ between a 2.45kb fragment of $yp3^{1526}$ and the equivalent sequence of the Canton S yp3 gene. Of these changes, 7 are found in the 5′ non-transcribed region and the other 2 are located within the coding sequence of the gene at codons 10 and 149 (see Table 4.1). A C → A transversion occurs at the

Figure 4.1

Strategy for sequencing $yp3^{1526}$.

The top line represents the yp3 gene. The introns are represented by the two open boxes. The hatched boxes represent the exons. Transcription of the gene proceeds from the left to the right hand side.

The lines below represent $yp3^{1526}$ subclones and the arrows indicate the direction in which they were sequenced.

o = indicates the location of primers synthesised.

The sites where the restriction enzymes SacI (an isochizomer of SstI) BglII and BamHI cleave the YP3 sequence are indicated.

One subclone is marked on one region with a dotted line to indicate that sequence in this region was not determined from this clone.

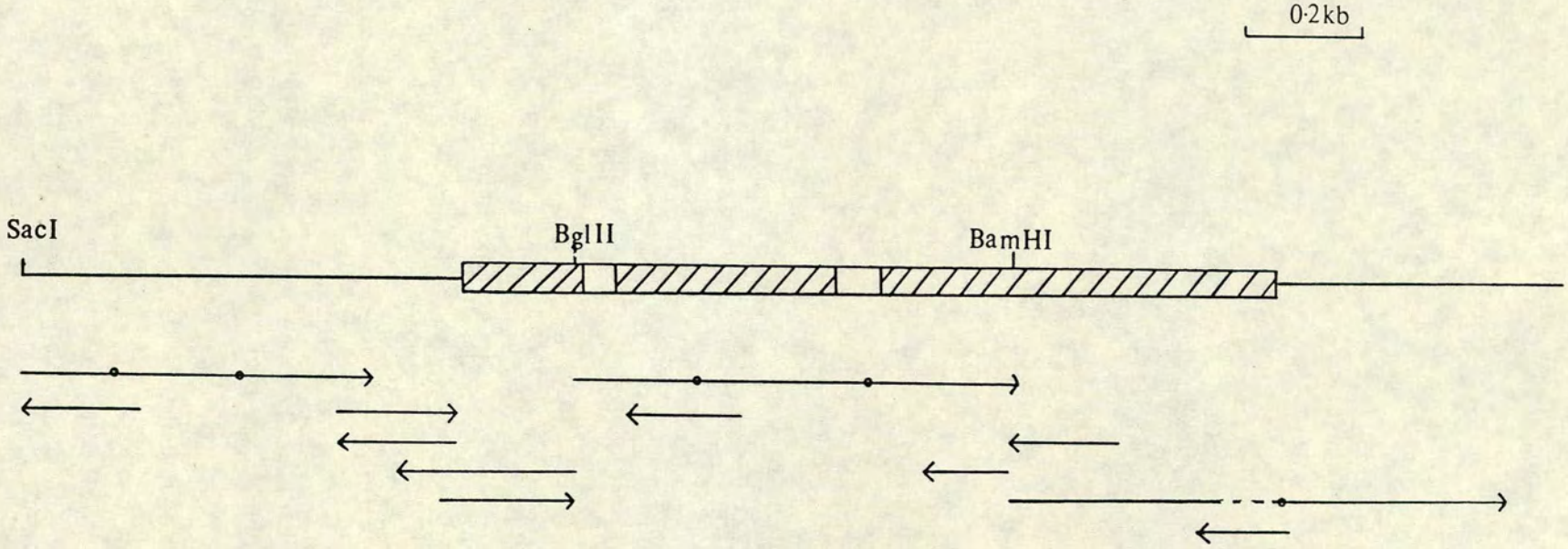

0·2kb
SacI
BglII
BamHI

Figure 4.2

Regions of sequencing gels showing 4 of the sequence changes in the upstream region of $yp3^{1526}$. The wild type sequences on the left hand side are compared with the mutant sequences on the right, the sequence read from each gel is presented alongside.

A. The C → T transition at nucleotide -580 and the T → C transition at nucleotide -547.

B. The A → C transversion at nucleotide -308 and the T → C transition at nucleotide -332.

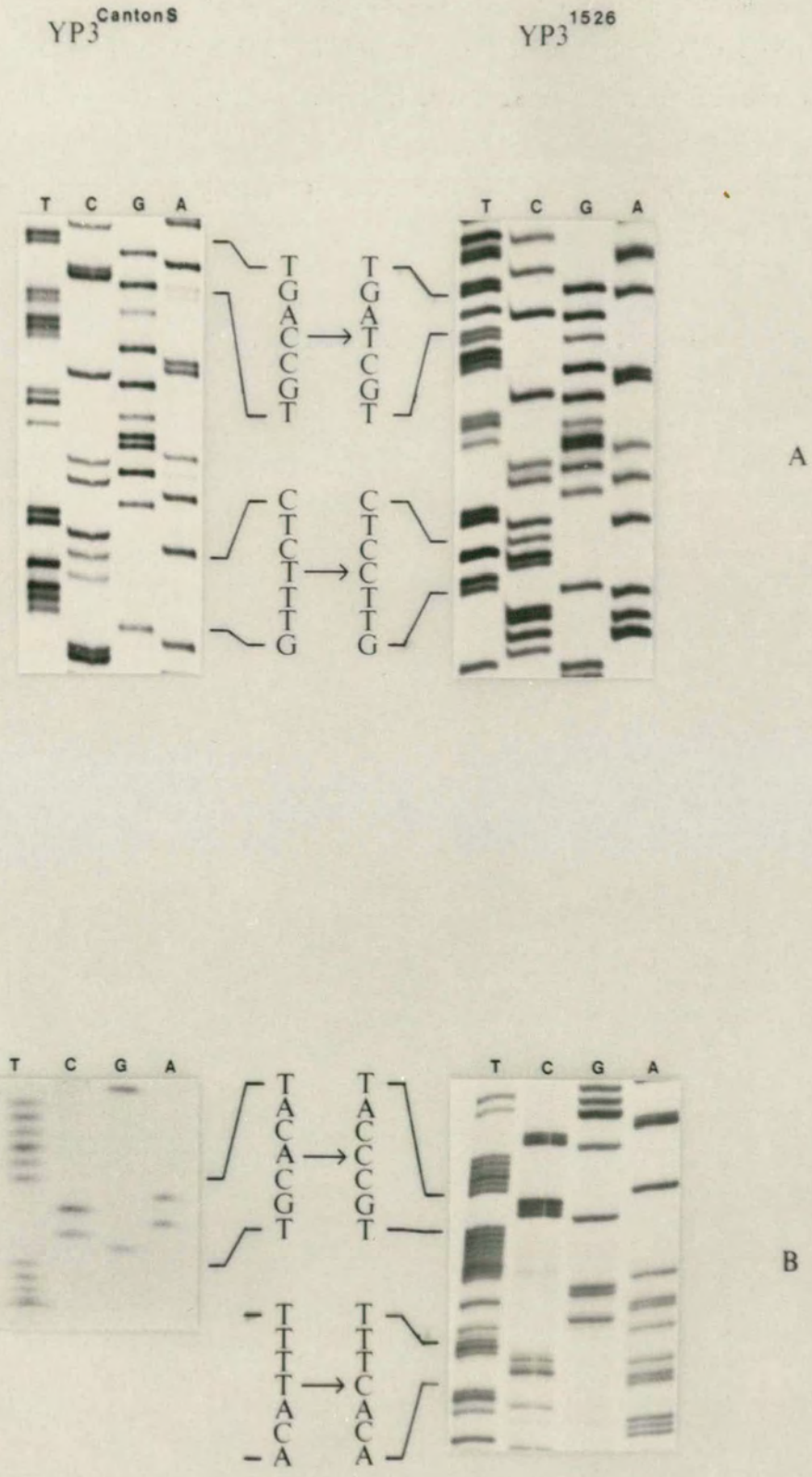

YP3CantonS
YP3 1526
T C G A
TGACCGT
TGATCGT
CTCTTTG
CTCCTTG
T C G A
A
T C G A
TACACGT
TACCCGT
TTTTACA
TTTCACA
T C G A
B

Figure 4.3

Regions of sequencing gels showing 3 of the sequence changes in the upstream region of $yp3^{1526}$ and sequence at the transcript start site. The wild type sequences on the left hand side are compared with the mutant sequences on the right, the sequence read from each gel is presented alongside.

A. The insertion of 1 nucleotide at position -79.

B. Two C → A transversions, at positions -44 and -37.

C. The sequence around the transcription initiation site (the first residue corresponds to the first nucleotide of the mRNA and is marked with an asterisk) illustrating that this is normal in the mutant sequence.

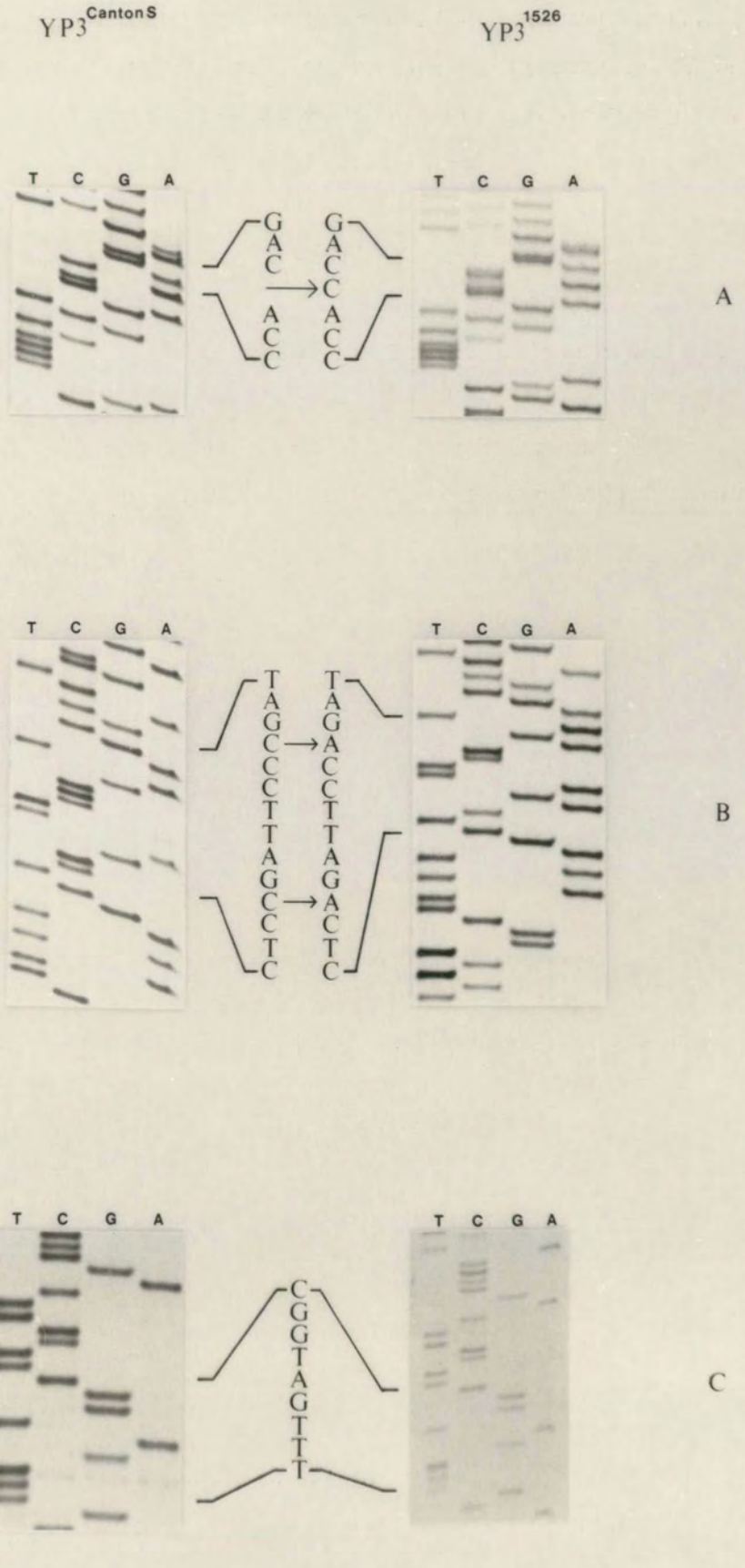

YP3Canton S
YP3 1526
T C G A
GAC → ACC
GACCACC
A
TAGCCCTTAGCCTC → TAGACCTTAGACTC
B
CGGTAGTTT
C

Figure 4.4

Regions from sequencing gels showing the mutations in $yp3^{1526}$ within the coding regions. The wild type sequences on the left hand side are compared to the mutant sequences on the right and the sequence read from the gel is presented alongside.

A. The C → A transversion in the 2nd nucleotide of the 10th codon which results in an aspartic acid residue replacing an alanine residue.

B. The C → T transition in the 3rd position of codon 149 in the second exon. This is a silent mutation.

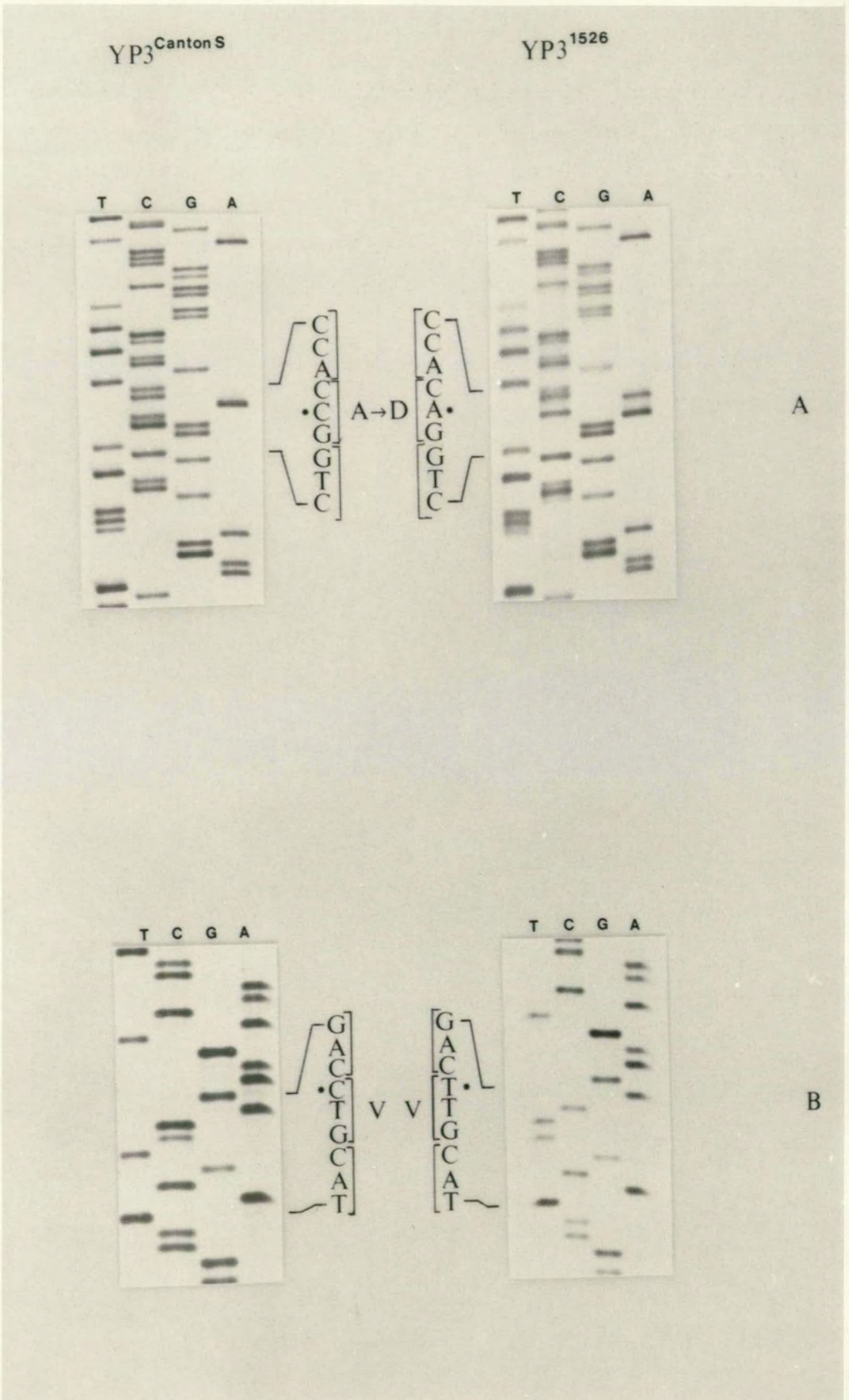

YP3Canton S
YP3 1526
T C G A
CCACCGGTC
A→D
CCACAGGTC
T C G A
A
T C G A
GACCTGCAT
V V
GACTTGCAT
T C G A
B

Table 4.1

Table of the sequence variations identified between wild type yp3 (Canton S) and $yp3^{1526}$. The nucleotides listed are for the coding strand of the gene.

POSITION (relative to transcript start)	SEQUENCE ALTERATION		
-580	T → C		
-547	C → T		
-322	T → C		
-308	A → C		
-79	G insertion		
-44	G → T		
-37	G → T		
+86	C → A codon 10	GCC → GAC	
		Ala	AspAc
+ 566	C → T codon 149	GTC → GTT	
		Val	Val

Figure 4.5

The amino acid sequence of the wild type cantonS) ¯YP3 signal peptide (upper line) aligned with the sequence of the $YP3^{1526}$ mutant signal peptide (lower line). The residue that is altered in the mutant leader peptide is underlined.

$YP3^{CantonS}$: Met Met Ser Leu Arg Ile Cys Leu Leu Ala Thr Cys Leu Leu Val Ala Ala His Ala

$YP3^{1526}$: Met Met Ser Leu Arg Ile Cys Leu Leu Asp Thr Cys Leu Leu Val Ala Ala His Ala

second nucleotide of the 10th codon (GCC → GAC) and the results in the substitution of an alanine residue by an aspartic acid residue. (see Fig 4.5) This change in sequence disrupts the recognition site for the restriction endonucleases Pal I, Hae III and Eae I. The sequence change in the 3rd base of codon 149 (GCC → GCT), a C → T transition, is a silent mutation since it does not cause an amino acid substitution and valine is the amino acid incorporated in both cases.

To ensure that the sequence changes noted for the $yp3^{1526}$ gene are real, the wild type gene was resequenced in the relevant regions (these are shown in figures 4.2 - 4.4 along with the mutant sequence). During this, the sequence published by Garabedian *et al.* (1987) was found to be incorrect in the region where initiation of transcription occurs. Two nucleotides presented in the published sequence are not present so that where the sequence reads AATCGAAA, it actually reads ATCAAA (Fig. 4.3). Note that this is in agreement with the sequence published by Yan *et al.* (1987). The first A in this sequence is the first nucleotide of the yp3 transcript.

DISCUSSION

The $yp3^{1526}$ gene is compared in this study with the yp3 gene from Canton-S, it should be noted that this was not the strain of flies used in the mutagenesis experiment by Gans *et al.* (1975). In this, flies bearing an X-chromosome marked with v^{24} (vermillion) were subjected to the mutagen ethyl methyl sulphonate (EMS). It is therefore possible that some of the sequence changes noted here are polymorphisms between the Canton S and v^{24} strains and to be able to definitively state which sequence variations represent those induced in the EMS mutation experiment it would be necessary to sequence these regions of the $yp3^{v24}$ gene. However, it remains valid to study the sequence variations between the yp3 gene in fs(1)A1526 and Canton-S and to assess which of these may be responsible for the phenotype of YP3 in the mutant flies. It was hoped that several characteristics of YP3 in fs(1)A1526 females would be explained by the DNA sequence alterations in the $yp3^{1526}$ gene: the increase in molecular weight of the $YP3^{1526}$ polypeptide and its failure to be secreted, and the reduction in level of $yp3^{1526}$ mRNA.

A protein which is normally secreted could fail to be exported for a number of reasons. Both the signal sequence and the mature sequence of secreted polypeptides have been shown to be important for proper progress through the secretory pathway.

Mature Sequences and Secretion

Some alterations in the mature domain of secretory proteins (i.e. the remainder of the polypeptide following signal peptide cleavage) have been found to affect their secretion. Some secretory mutants have been identified which are mutated in the region immediately following the signal peptide cleavage site. For instance, a net positive charge of +2 very early in the mature sequence of *E. coli* alkaline phosphatase (where the net charge is normally 0, interferes with protein export (Li *et al.*, 1988). It has been speculated that in such mutations the polypeptide is no longer a good substrate for the signal peptidase or is unsuitable for translocation across the membrane. Examples where alterations of residues occurring further into the mature domain of polypeptides affect secretion have also been reported. For instance, a mutation *suc2-s2* that converts a threonine to a isoleucine at residue 64 in the mature sequence of yeast invertase, results in a sevenfold reduction in the rate of its secretion (Schauer *et al.*, 1985). In cases such as suc2-s2 it is possible that the substituted residue affects the conformation of the protein or other post-translational modification which may be important for efficient transport through the ER and Golgi systems.

Proteins could be retained in the ER and Golgi membranes systems because a mutation causes the protein to have a prolonged interaction with ER/Golgi residing protein(s), or by causing aberrant self-aggregation so that proper packaging for transport is hindered. Another possibility is that a mutation causing an amino acid substitution may result in the creation of a sequence which somehow signals retention of the protein in the ER. An example of an ER retention signal has been described by Munro and Pelham (1987). The tetra-peptide Lys-Asp-Glu-Leu (KDEL) was found at the carboxyl termini of several soluble proteins that remain within the lumen of the ER, and adding this peptide to the carboxyl end of a normally secreted protein resulted in its retention in the ER.

The regulation of intracellular traffic is an area of research that is currently attracting much attention (for review see Klausner, 1989). However, the present molecular understanding of the mechanisms of trafficking and sorting of proteins through the cellular network of membrane bound organelles, for which the ER acts as a port of entry, is still fairly elementary. One model proposed to account for how cells sort and correctly localise particular polypeptides is the bulk flow and retention

hypothesis. In this model, once a protein is targeted to the ER, it will travel through the vacuolar system bypassing various organelles, unless it reaches a point for which it possesses a unique signal that labels the protein for retention at this site (e.g. KDEL). The existence of positive transport signals has also been postulated since different proteins can be secreted at different rates. It has been suggested that N-linked glycosylation can serve as a direct transport signal in promoting protein export, certainly it has been found to be crucial in the proper export of some proteins, however, its requirement for secretion is not universal (Rose and Doms, 1988). The major problem in assessing the effect of an alteration in the primary structure of a protein or modifications such as glycosylation is that it is difficult to determine whether an alteration affects protein targeting or sorting due to the disruption of a specific transport signal, or because it has a general effect on protein folding and conformation.

Signal Peptides and Secretion

Alterations occurring in the signal peptide can effect the secretion of proteins (examples have been discussed in the Signal Sequences section of the Introduction). A mutation in the c-region of the signal sequence may result in non-cleavage of the leader peptide from a secretory protein (see Table 1.4 in Introduction). In many cases where this has occurred the mutant protein is successfully translocated but remains anchored to the membrane and so fails to be secreted (e.g. a Pro → Leu change at position -4 in the signal sequence of β-lactamase; Koshland *et al.*, 1982). In other examples of c-region mutations the uncleaved protein is secreted from the cells, but at a very reduced rate, e.g. *suc2-s1*, an alteration of Ala → Val at position -1 in the signal sequence of yeast invertase; Schauer *et al.*, 1985).

Mutations in the hydrophobic core of signal sequences of secretory proteins have been shown to have drastic consequences on their export. The length and hydrophobic nature of the h-region appears to be crucial to the effective functioning of the signal sequence so that most cases where the h-core is disrupted, for example by insertion of a charged residue (e.g. S71, Val → Asp at -12 position of LamB signal sequence), or partial deletion of the h-region (e.g. S78, a deletion of 4 residues of the h-region of LamB; Emr *et al.*, 1980), the protein is not secreted. In these cases the secretion of the mutant protein seems to be affected at an early stage of the secretory pathway. The mutant proteins do not undergo any post-translational modifications, such as signal

peptide cleavage, glycosylation or phosphorylation, that are associated with entry of secretory proteins into the lumen of the ER, and so do not appear to have been translocated across the membrane. The mutant proteins can often be isolated with the cytoplasmic fractions of the cell indicating that they do not reach the membrane, or are only weakly associated so that the gentle disruption used to isolate the various cell fractions is sufficient to remove them from the membrane.

In the mutant strain with which this study has been concerned, fs(1)A1526 flies, the mutation causing both the non-cleavage of the signal peptide and the failure to be secreted of the YP3 polypeptide is most likely to be the transversion of C → A in the 10th codon, which is at position -9 if the signal sequence is 19 residues long as suggested in the yolk protein section of the Introduction). This is the only alteration within the coding sequence of the gene which results in a variation of the amino acid sequence. The only other mutation in the coding region, at residue 149, is silent. The effect of the mutation at amino acid 10 is the replacement of an alanine residue with an aspartic acid residue. (see Fig 4.5) Amino acid 10 lies within the hydrophobic core of the YP3 signal sequence and, as has been discussed, the introduction of a charged residue into the hydrophobic region of a signal peptide can severely disrupt the secretion of the protein.

This mutation could render the $YP3^{1526}$ signal peptide unrecognisable by the translocation apparatus (e.g. the SRP) (or any interaction is somehow defective), or interfere with the event of translocation across the ER membrane. This would explain the increased molecular weight of the non-secreted $YP3^{1526}$ since if translocation into the lumen of the ER does not occur, the signal peptidase cannot cleave the signal sequence, and so the protein has a higher molecular weight due to the retention of the extra amino acids.

Another possibility is that the presence of the aspartic acid residue at amino acid 10 somehow interferes with recognition and cleavage of the $YP3^{1526}$ signal sequence by the signal peptidase. However, the amino acid alteration in the $YP3^{1526}$ signal peptide is probably not close enough to the postulated site of cleavage, after residue 19, to have this effect. The c-region, which appears to define the site of peptidase cleavage in eukaryotic signal sequences is generally 6 residues long, the aspartic acid residue at position -9 relative to the site of cleavage is probably too distant to effect this process.

It is possible that the mutation in $YP3^{1526}$ causes prolonged interaction with components of the translocation apparatus (or internal ER residing membrane proteins

involved in transport after translocation). This could result in $YP3^{1526}$ saturating the translocation sites or proteins at the membrane (e.g. SSR) which would be likely to interfere in the secretion of other proteins. This occurs for a secretion mutant of invertase in yeast, where the product of *suc-s2* seems to affect proper secretion of other proteins by "clogging up" the pathway (see Introduction).

Female fs(1)A1526 flies show a reduction in the steady state level of the yp3 transcript in both site of synthesis (see Fig. 3.6B and C). A number of factors, acting at different levels of gene expression have been implicated in governing transcript levels.

At the level of transcript initiation, a variety of *cis*-acting sequences have been shown to be involved in specifying the expression. Most genes that are transcribed by RNA polymerase II have in common a combination of "general" *cis*-acting sequences around the promoter region (defined as the sequence in the immediate vicinity of the transcription start site) that are believed to be essential for the basal transcription (e.g. the TATA box and CCAAT box)(Mitchell and Tjian, 1989). In addition to these ubiquitous regulatory regions, other *cis*-regulatory elements have been identified in eukaryotic promoters. These *cis*-acting sequences are specific to particular groups of genes (promoter-specific) and modulate particular characteristics of their expression such as tissue specificity and developmental expression (Biggin and Tjian, 1989).

The yolk protein genes have been shown to contain several *cis*-regions which regulate their tissue-specific synthesis in *Drosophila melanogaster* females (see Regulation of YP Gene Expression in Introduction). Mutations in any of the promoter-specific or general *cis*-regulatory regions could reduce or abolish the ability of these sequences in directing or modulating transcription of the genes. Any such lesions occurring in the sequence of $yp3^{1526}$ could explain the reduction in level of yp3 transcript seen in fs(1)A1526 females. Therefore, the seven sequence alterations upstream of the transcript start site identified in $yp3^{1526}$ are candidates for mutations causing this particular phenotype. Unfortunately, no details concerning the promoter-specific *cis*-regulatory regions in the yp3 gene has been established (see Chapter 5). None of the sequence changes in upstream region of $yp3^{1526}$ affect any of the putative *cis*-regulatory regions identified by a computer search comparing the yp3 sequence with regions known to affect expression of the yp1 and yp2 genes (Fig. 1.5 in Introduction, and Figs. 5.7, 5.8 and 5.9 in Chapter 5).

The steady state level of mRNA's is not determined solely by the level at which transcription is initiated. In recent years more emphasis has been placed on

understanding the mechanisms of post-transcriptional regulation processes which affect steady state mRNA levels (Ross, 1988). The primary structure of an mRNA is important in determining its turnover rate, and a few examples have been described where the polypeptide it encodes also affects the half-life of a mRNA.

The transcript half-life can be determined by the primary mRNA sequence in several ways: sequences may contain cleavage sites for ribonucleases; some regions may function as stabilisation sequences; and others may act in an indirect way by providing binding sites for proteins. Sequences in the 3′ untranslated regions of eukaryotic mRNA's have been determined which affect their turnover rates. Amongst these is a set of AU-rich sequences found in the 3′ region of some short-lived mRNA's which have been shown to be responsible for de-stabilising these transcripts (Brawerman, 1989).

In transcripts which are unspliced due to mutations in sequences at or around the consensus splice junctions, the steady state levels can be affected. Such effects have been observed for two mutant alleles of the alcohol dehydrogenase gene in *Drosophila melanogaster* that are deleted around the splice junctions of the intron. The *Adh* transcripts detected *in vivo* in these mutants, Adh^{fn4} and Adh^{fn6}, are only 5-10% of the wild type level (Benyajati *et al.*, 1982). The authors suggest that this may be caused by the unspliced mutant transcripts being more unstable than the wild type transcripts, or perhaps transcriptional efficiency is somehow reduced.

Sequences located in the 5′ untranslated region of mRNA's have also been implicated in decay rates. An example of one of these is the 5′ sequences of the cellular oncogene *c-myc*; if the first section of the transcript is deleted, the turnover rate decreases (Ross, 1988).

Recently an apparent link between translation and mRNA decay has been noted. Some studies indicate that cells are able to recognise and degrade mRNA's where the mutant protein has been translated. One example of such an effect is seen in the human gene encoding triosephosphate isomerase (TPI). An anaemia-inducing mutation in the TPI gene that causes premature termination of the protein also causes a reduction in the half-life of the transcript (Daar and Maquat, 1988). In this study the authors mutagenised TPI alleles *in vitro*, transformed them into cultured cells and examined the levels of TPI mRNA. Their results indicated that nonsense codons or frameshift mutations result in a decrease of TPI mRNA stability. In another system, nonsense mutations within the yeast *ura-3* gene have been shown to cause an

abnormally short half-life of the corresponding transcript. Such mutations found at the amino terminus of the coding region destabilise the transcript more efficiently than those near the carboxy-terminus. A correlation between nonsense codons and reduced mRNA levels has been established for several other eukaryotic mRNAs (e.g. see references in Ross, 1989).

One of the most interesting examples of the coding sequence of a gene affecting the transcript stability is β-tubulin in animal cells (reviewed by Cleveland, 1988). It was realised several years ago that the concentration of β-tubulin subunits in the cell autoregulates the stability of the transcript. In this system actively translating β-tubulin mRNA's are marked for degradation by the co-translational binding (possibly of the unpolymerised tubulin subunits themselves) to the amino-terminus of the nascent β-tubulin shortly after the polypeptide chain emerges from the ribosome. The recognition element to which the autoregulatory factor(s) bind is the amino-terminal tubulin tetrapeptide (Met-Arg-Glu-Ile). How the protein/protein interaction is transduced through the translation apparatus to result in degradation of the transcript is not known. One possibility proposed by Yen *et al.* (1988) is the activation of an RNase localised to actively translating ribosomes.

In each of the cases described above, where the coding region has some effect upon the stability of the mRNA, the common factor is that the mRNA stability is associated with ongoing protein synthesis. This suggests the existence of a mechanism whereby the degradation of mRNAs is integrally linked to ribosome attachment and translation. Little information is currently available on how this type of nuclease activity is regulated and on how cellular factors affect the susceptibility of transcripts to decay. Further insights into this type of regulation of mRNA levels will require the identification and characterisation of ribonucleases whose function is associated with the process of protein synthesis.

The question which now has to be addressed is which, if any, of the processes described above, could best explain the reduction in the level of yp3 transcripts observed in fs(1)A1526 females.

If the turnover rate of $yp3^{1526}$ is altered (rather than initiation of transcription being affected), this could be due to either recognition of the altered primary mRNA sequence resulting in degradation, or translation of the mutant polypeptide resulting in the decay of the mRNA. Firstly, how could any of the primary sequence determinants

of RNA stability affect yp3 mRNA? The sequence determination of $yp3^{1526}$ indicates that the only nucleotide alteration likely to have an impact on the transcript is the C → A transversion at nucleotide +86. The other sequence variation, which occurs at nucleotide +566, is not likely to be involved since it has no effect on the codon specification, nor does it cause premature termination of the polypeptide chain or insertion of a nonsense codon (both of which can affect the stability of the transcript). However, the mutation at nucleotide +86 may affect the turnover rate of the transcript if it disrupts a region important for mRNA stability, or if it creates an "unstable" sequence, or if, like the Adh^{fn4} and Adh^{fn6} alleles, the consensus splice sequences are affected.

The mutation at nucleotide +86 need not affect the half-life by altering the primary structure of the $yp3^{1526}$ transcript, the polypeptide it encodes may be involved. The sequence alteration results in the incorporation of a different amino acid at residue 10 of the polypeptide (Ala → Asp). As already discussed, this amino acid substitution is highly likely to be the cause of the secretion defect and increased molecular weight of the $YP3^{1526}$ protein. Bearing in mind the cases where the transcript turnover rate is related to ongoing protein synthesis, it is plausible that a similar mechanism occurs in fs(1)A1526 flies and causes the increased rate of decay of the $yp3^{1526}$ transcript. This would involve some (currently unidentified) cellular machinery recognising the faulty $YP3^{1526}$ polypeptide and transducing this information to result in the degradation of $yp3^{1526}$ mRNA. That is, this possibility would include, like the other examples described above, linking the degradation of the $yp3^{1526}$ mRNA with ribosome attachment and translation.

Other explanations are possible, for instance, there may be some other kind of feedback regulation depending simply on the unusually high level of $YP3^{1526}$ polypeptide in the cells. Again, however, this postulates the involvement of other cellular factors recognising the alteration of $YP3^{1526}$ (either in itself, or in the level at which it is present).

Transformation experiments using yp3 genes where only the upstream regions have been altered should indicate whether any of these sequence changes causes the mutation of a *cis*-regulatory sequence. Note that if this is the case the region affected must be involved in determining the level of transcript initiation, as the reduction in $yp3^{1526}$ mRNA is seen in the fat body cells and the follicle cells of the ovary. Unless,

of course, a fat body regulatory element and an ovarian regulatory element have both been mutated.

These are examples of the type of experiments that could be performed in *Drosophila*. Taking the mutant yp3^{1526} gene into another system may give some insights into the nature of the secretion problem of the YP3^{1526} polypeptide. One way to investigate whether the YP3^{1526} signal sequence is competent for translocating across the endoplasmic reticulum would be to study this in an *in vitro* system. It is already known that all of the yolk proteins are correctly translocated and cleaved in the dog pancreas microsome *in vitro* translation/translocation assay. It would therefore only be necessary to perform a similar experiment using the transcript from a yp3 gene engineered to encode aspartic acid at residue 10 (and from a control wild type yp3), and determine whether the mutant YP3 can cross the microsomal membrane and be cleaved by the signal peptidase.

Experiments performed in yeast with the gene for the secreted form of invertase (discussed in the Introduction) have been successful in indicating some of the points in the secretion pathway where the transport of proteins with mutant signal sequences is aborted. The major advantage in using yeast is the array of mutants that are identified which affect transport out of the cell (see relevant section in the Introduction) and which are providing further insights into the secretion pathway. It would be interesting to use this system with the mutant YP3^{1526}, perhaps fusing sequences encoding the signal sequence to another, easily assayable gene such as invertase. This kind of study may indicate where in the process of secretion from the cells, the YP3^{1526} signal sequence is defective.

FINAL COMMENTS

Of course, all these are speculatory ideas at present and the queries regarding the mutant characteristics of the YP3 polypeptide and transcript level in fs(1)A1526 flies (or other strains carrying the mutated yp3^{1526} gene) can only be answered by experimental investigations.

The first step in solving the questions posed here should involve experiments to determine if the mutation that results in failure of the polypeptide to be secreted and its increase in molecular weight also causes the reduction in level of yp3 mRNA. The solution to this may be obtained in a transformation experiment (in *Drosophila*) using

an engineered yp3 identical to the wild type gene in every respect except at codon 10 (where it should code for the aspartic acid residue of $YP3^{1526}$). Examination of the protein and the transcript level of yp3 in the transformed flies should (1) prove whether the aspartic acid at codon 10 actually does cause the secretion defect (and the increase in molecular weight) of $YP3^{1526}$, and (2) (if this is the case) whether the $yp3^{1526}$ transcript level is the result of the same mutation or is caused by one of the other nucleotide variations seen in the $yp3^{1526}$ gene. Determining the level of yp3 mRNA would be simpler in a mutant in which yp3 is not transcribed or the transcript size is detectably different to the wild type. However, if the codon alteration is unrelated to the mRNA level, wild type amounts of yp3 transcript may be achieved, which would be easily distinguished from the reduced level in the mutant.

CHAPTER 5

INVESTIGATION OF *cis*-ACTING SEQUENCES REGULATING THE EXPRESSION OF yp3 IN *DROSOPHILA MELANOGASTER*

INTRODUCTION

Several components are known to be involved in regulating the expression of eukaryotic genes. The initiation of mRNA synthesis (and how frequently this occurs) depends upon proteins that interact with specific elements in and around a gene to activate or repress its transcription. One essential step in unravelling the interactions leading to tissue specific, and developmentally and hormonally regulated gene expression is the determination and characterisation of the *cis*-sequences that are involved in specifying the expression pattern of particular genes.

Some of the *cis*-acting regions necessary for the correct developmental specificity of yp1 and yp2 expression in *Drosophila melanogaster* have been identified (see relevant section in the Introduction). The first experiments showed that separate *cis*-sequences specify transcription of yp1 and yp2 in the fat body cells and the ovarian follicle cells (Garabedian *et al.*, 1985). A recent paper (Logan *et al.*, 1989) reported that two regions of DNA are required for expression of yp1 and yp2 in the ovary. One of the objectives of this study was to identify some of the *cis*-sequences involved in regulating the expression of yp3.

To achieve this, the wild type yp3 gene from Canton-S flies was introduced into the germline of a mutant strain by P-element mediated transformation of *Drosophila* embryos. The strain chosen for this was fs(1)A1526 which is described in Chapter 3 of this thesis. This strain does not secrete YP3 from the sites where it is synthesised so no YP3 is found in the haemolymph or eggs from homozygous fs(1)A1526 females. This meant that the wild type yp3 gene could be introduced into the fs(1)A1526 background and its expression in the fat body detected by the presence of YP3 in the haemolymph.

TRANSFORMATIONS

The yp3 gene from the wild type strain Canton-S was inserted into the *Drosophila* transformation vector, pUChsneo (Steller and Pirrotta, 1985). The advantage of this vector is that it provides a positive selection for transformants since it contains the bacterial neomycin resistance gene under the control of the *Drosophila* hsp70 promoter. Flies transformed with pUChsneo are resistant to the neomycin analogue Geneticin G418. Other features of pUChsneo include:- the inverted terminal repeat

sequences of the P-element, these allow the transposition event into the *Drosophila* genome; a polylinker (containing multiple restriction endonuclease sites) situated within a region encoding one of the β-galactosidase domains - this allows the identification of recombinant plasmids on the basis of the blue/white colour change assay; and the β-lactamase gene conferring ampicillin resistance in *E. coli* cells. The transformation vector pUChsneo does not contain sequences encoding the P-element transposase, therefore to ensure transposition of the P-sequences, a helper plasmid is required. The plasmid chosen for this was phsπ (Steller and Pirrotta, 1985), which contains the transposase gene fused to the hsp70 promoter (to generate sufficiently high levels of transposase) but does not itself transpose into the genome as it has only one of the P-element terminal repeats. Using a helper to transiently supply the transposase activity ensures that no further transposition events occur once the transposon has integrated into the genomic DNA (unless the fly strain used has an endogenous source of P-element transposase). To ensure that the fs(1)A1526 strain used here does not carry any P sequences, genomic DNA was extracted from fs(1)A1526 flies and subjected to Southern blotting analysis. The probe used was the internal 840bp HindIII fragment from pπ25.1 (O'Hare and Rubin, 1983) which contains some of the left hand terminal repeat and sequences internal to the P-element. This fragment was radioactively labelled by the random priming method (2.4.6.1). Another wild type *Drosophila* strain was used as a positive control, this line (Luminy) contains apparently complete P-elements, but displays a weak P-cytotype (Bregliano and Kidwell, 1983).

The Southern blot in figure 5.1 shows that homozygous and heterozygous (with an FM3 balancer chromosome) fs(1)A1526 flies do not contain any sequences that are recognised by the P-element probe. Flies that are fs(1)A1526 have no P-elements within their genome, therefore this strain is likely to be of M-cytotype and so is suitable for using in these transformation experiments. DNA from two other strains on the blot (Z1 and Z2, which have mutant alleles for β-galactosidase) show many differently sized fragments that are recognised by the sequences in the P-element probe. This shows that these strains contain P-sequences, some of which could be complete P-elements and the flies may therefore be P-cytotype and would be of no use in a transformation study of this kind.

The complete, unaltered $yp3^{Canton\text{-}s}$ gene contained within a SalI-EcoRI 4.2kb fragment of DNA from pYP3 (Barnett *et al.*, 1980) was ligated into the SalI-EcoRI sites

Figure 5.1

A. Southern blot of HindIII digested genomic DNA extracted from various Drosophila strains, as indicated at the top of the tracks. The probe used is an 840bp HindIII fragment (from pπ 25.1) containing only P-element sequences. DNA was extracted from the following strains:-

L: Luminy - a wild type strain

$\frac{1526}{1526}$: Homozygous fs(1)A1526 flies

$\frac{1526}{1526}$: Heterozygous fs(1)A1526 flies

Z1: Strain carrying mutant allele of β-galactosidase (Lindsley and Zimm, 1985)

Z2: " " " "

Or^R: Oregon R - a wild type strain

Each lane represents DNA extracted from two individuals.

The size markers were λDNA digested with HindIII.

The arrow indicates an 840bp P-element fragment present in the genome of Luminy, Z1 and Z2 flies.

B. Southern blot of EcoRI digested genomic DNA extracted from transformant flies. The probe used was a 4.2kb SalI-EcoRI fragment containing the yp3 gene. The lanes represent DNA extracted from the following strains:

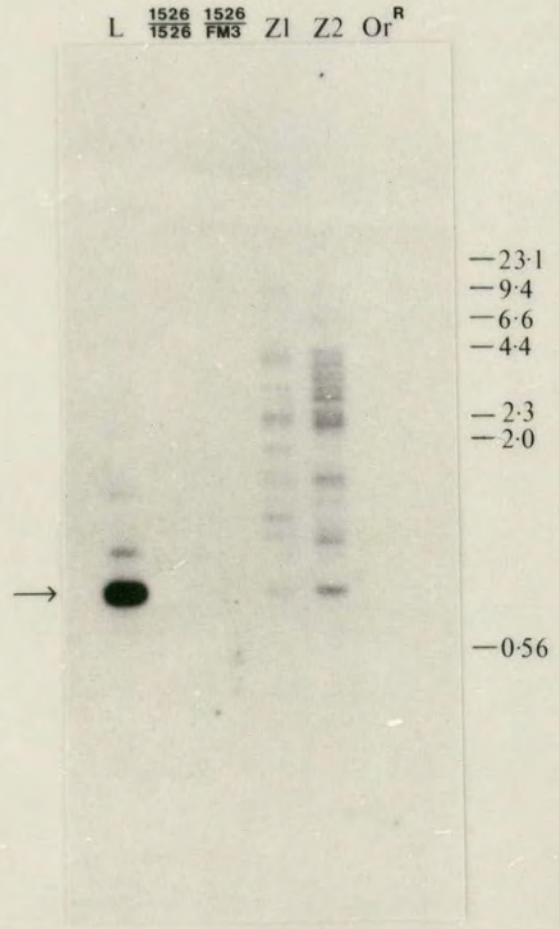
Or^R: Oregon R - a wild type strain

1B: Flies transformed with the transposon contained within pUChsneoyp3.4 • 2.

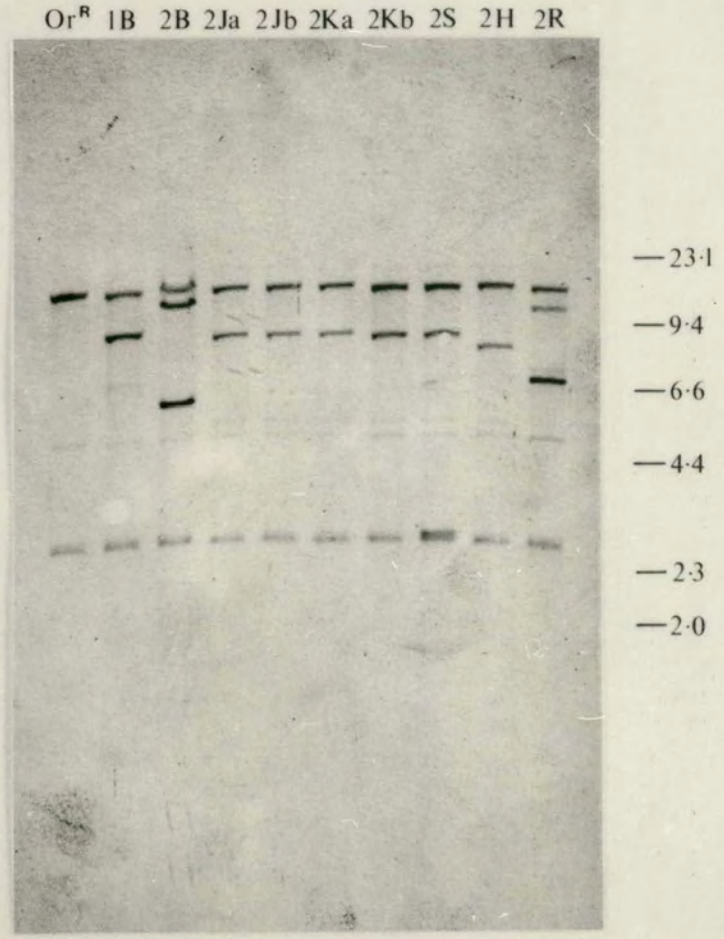
2B, 2Ja, 2Jb, 2Ka, 2Kb, 2S, 2H, 2R: Strains transformed with the transposon from pUChsneoyp3.3 • 1.

The size markers were λDNA digested with HindIII.

A


L
1526/1526
1526/FM3
Z1
Z2
Or^R
23·1
9·4
6·6
4·4
2·3
2·0
0·56

B


Or^R
1B
2B
2Ja
2Jb
2Ka
2Kb
2S
2H
2R
23·1
9·4
6·6
4·4
2·3
2·0

in the polylinker of pUChsneo (see Fig. 5.2A), the resulting plasmid was named pUChsneoyp3/4.2 (where the 4.2 refers to the size, in kb, of the yp3 fragment). To delete some of the DNA sequences upstream of the coding region of the yp3 gene, pUChsneoyp3/4.2 was reopened at the SalI and SacI sites. T4 DNA polymerase was used to fill in the SalI cohesive end and remove the SacI 3′ overhang, and the blunt ends were then ligated (Fig. 5.2B). This resulted in the removal of 1.1kb of DNA sequences from the 5′ flanking region of the yp3 gene, the plasmid generated was named pUChsneoyp3/3.1.

Oregon R embryos were injected, as described in Section 2.4.9, with a mixture of transposon (pUChsneoyp3/4.2 or pUChsneoyp3/3.1) and helper (phsπ) plasmids in a 5:1 ratio. Table 5.1 lists the numbers of embryos injected with each of the two constructs and details the number of survivors at each stage. A total of 4 transformant lines were obtained from the pUChsneoyp3/4.2 injections, two of which were maintained. From the injections with pUChsneoyp3/3.1, a total of 8 transformant lines were obtained, 6 of these lines were maintained.

ANALYSIS OF TRANSFORMED LINES

The fly lines transformed with the wild type yp3 gene were characterised by genetic analysis, to find into which chromosome the transposon had inserted in each line, and by Southern blot analysis, to determine the number of insertions in each line.

An attempt was made to show by linkage analysis into which chromosome the transposon had inserted in each transformed line. To do this a series of crosses was performed as in figure 5.3. The basis of the chromosomal assignment is outlined in this figure. However, the results of these crosses proved to be difficult to interpret, in many cases the progeny emerging from the back cross to cn/cn;e/e females did not always fall into the expected classes, e.g. in a batch of offspring which were largely cn/cn, some e/e would appear. It seems most likely that this problem was associated with flies emerging from the G418 selection that were not truly neomycin resistant. This could happen because different batches of G418 supplied appear to vary in toxicity to *Drosophila* larvae, and this can result in neomycin sensitive flies surviving, or alternatively, neomycin resistant flies being killed.

Figure 5.2

The two plasmids used in the transformation experiments. The construction of these plasmids is described in the text. The lines with arrows represent the yp3 transcript (the carets represent the introns).

A. The yp3 fragment contained in pUChsneoyp3/4.2.

B. The yp3 fragment contained in pUChsneoyp3/3.1.

P = P-element terminal repeats

lacZ = β-galactosidase sequences

hs = the promoter from the *hsp70* gene

neo = the neomycin resistant gene

white = sequences from the *Drosophila white* locus (note that these are not to scale).

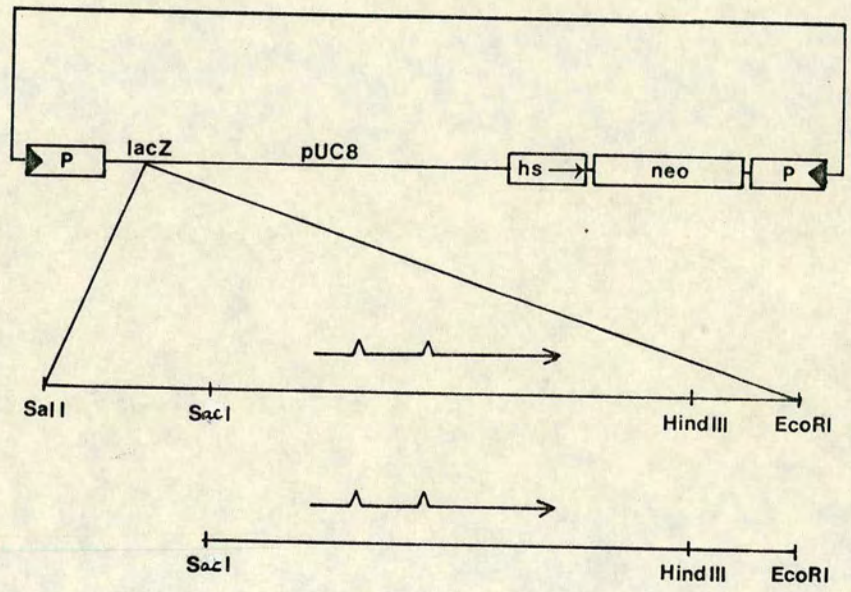

P
lacZ
pUC8
hs
neo
P
Sal I
Sac I
Hind III
EcoRI
A
Sac I
Hind III
EcoRI
B
1kb

Table 5.1

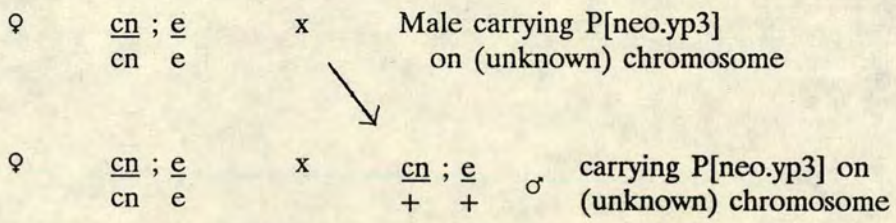
Statistics of embryo injections and number of survivors at each stage. The figures in brackets are percentages of the number of embryos injected surviving to that stage.

Construct	pUChsneoyp3.4.2		pUChsneoyp3.3.1	
Embryos injected	704		230	
Number hatched	142	(20)	53	(23)
Number eclosed	75	(11)	28	(12)
Number of fertile Go adults	45	(6)	17	(7)
Number of transformed lines	4 ↓ 2		9 ↓ 6	

Figure 5.3

Genetic analysis of transformed lines. These crosses were performed for several individuals from each transformed line to determine which chromosome the transposon goes into. Food was supplemented with Geneticin G418 to select for neomycin resistant offspring. These flies must therefore carry the transposon bearing the neomycin resistant gene (and the yp3 gene) P[neo.yp3].

e (ebony) is a recessive marker for chromosome 3.
cn (cinnibar) " " " " 2.



Scoring the phenotypes of the emerging offspring indicates which chromosome the transposon is found on:-

(1) If only females emerge from the first cross, then the insertion must be on the X-chromosome.

(2) Lines where cn/cn flies emerge do not have an insertion on the 2nd chromosome.

(3) Lines where e/e flies emerge do not have an insertion on the 3rd chromosome.

(4) Lines where both e/e flies and cn/cn flies emerge have an insertion on more than one chromosome.

(5) Lines where cn/cn ; e/e flies emerge have an insertion on chromosome 4 (or 1 if the first cross has not already ruled this out).

The tentative chromosome assignments derived from this analysis are given in Table 5.2 along with the number of insertions in each line identified by Southern analysis described below.

For the Southern blotting experiment genomic DNA was extracted (as Section 2.4.2.1) from four individuals of each transformed line digested with the restriction endonuclease EcoRI and the fragments separated by electrophoresis on a 1% agarose gel. The blot was probed with the complete yp3 gene (the SalI-EcoRI fragment from pYP3) labelled by the non-radioactive DNA labelling method (Mannheim Boehringer). This uses a random priming method to incorporate digoxigenin (DIG) labelled dUTP into the DNA which is hybridised to the blot. After washing the filter as normal, an antibody to DIG conjugated with alkaline phosphatase is added. This enzyme catalyses the colour reaction with X-phosphate and nitroblue tetrazolium salt, and in this way regions on the blot which hybridise the DIG-dUTP labelled probe are detected. Figure 5.1B shows the result of this experiment. The uppermost band visible in each lane represents the 15kb EcoRI fragment of the endogenous yp3 gene, and the lower molecular weight band(s) represent DNA fragments containing the transposed yp3 gene. EcoRI cleaves the transposed sequences only once and so the si e of fragment liberated depends on how near the next EcoRI site is to the site of the insertion.

Some additional, faint bands are visible on this Southern blot. These are likely to be the result of cross-hybridisation to other *Drosophila* sequences. The yp1 gene has two closely situated recognition sites for EcoRI (~110bp apart), the yp2 gene has none. Cleavage at these sites generates a 3.7kb yp1 fragment and a 9.6kb yp1/yp2 fragment. A fragment of approximately 3.5kb present in all samples on the blot in figure 5.1B is probably due to cross-hybridisation to the yp1 EcoRI fragment. A fainter band of approximately 5.3kb visible on the blot is highly likely to be cross-hybridisation to a *Drosophila* lipase gene (known to show a 5.5kb EcoRI genomic fragment from another study; M. Bownes, A. Shirras and J. Swinbourne, personal communication). A few other faint bands that are visible are likely to represent partial digestion products.

The results of this Southern blot are listed in Table 5.2. Two of the lines contain 2 insertions and all of the others have 1 insertion. It should be borne in mind that the DNA samples in each track were extracted from more than one fly and it is impossible to state for the lines that have more than one insertion, whether two insertions are linked, or if each is present in a different individual. Note also that due to the small number of flies sampled for each line, it is possible that some lines carry

Table 5.2

Chromosome assignment and number of insertions in each of the transformed lines. The brackets indicate very tentative assignment (and ND indicates not done).

LINE	INSERTION	POSSIBLE INSERT LOCATION (by linkage analysis)	NUMBER OF INSERTIONS (by Southern analysis)
1A	P [neo.yp3.4-2]	1	ND
1B	P [neo.yp3.4-2]	3	1
2A	P [neo.yp3.3-1]	2	ND
2B	"	2/(4)	2
2J	"	3	1
2S	"	3	1
2K	"	(3 or 4)	1
2R	"	2/(2)	2
2H	"	ND	1

more insertions than have been detected. Separate samples of some lines(i.e. 2Ja/2Jb and 2Ka/2Kb) were examined by Southern blotting where the genetic data had indicated the possibility that more than one insertion was present. In both of the 2J and 2K samples, the size of the DNA fragment generated by EcoRI digestion appears to be the same (or it is very similarly sized), if this is the case, it is unlikely that the a and b sublines of 2J and 2K were derived from more than one transposition event. It should also be noted that for lines 2J, 2K and 2S, the EcoRI fragment containing the transposon appears to be of the same size (approximately 9.4kb). This may simply be due to low resolution of similarly sized fragments in this gel, but if the fragments are the same size then it is possible that these three lines are derived from the same transposon insertion event. Repeating the Southern blot using a different restriction endonuclease (SmaI, for example) would resolve the problem. If the 2J, 2K and 2S lines share an equivalently sized SmaI fragment containing the transposed yp3 then it is more likely that they all represent the same transposon insertion (and are really the same transformed line).

To more completely and accurately establish the number and chromosomal location of the insertions, each line should be subjected to further characterisation by *in situ* hybridisations to larval salivary gland polytene chromosomes.

YP3 EXPRESSION IN THE TRANSFORMED FLIES

To analyse the expression of the yp3 sequences introduced into the germ line of the transformed flies, it was first necessary to obtain the chromosome carrying the insertion in the $YP3^{1526}$ background. This was performed in the crossing scheme shown in Figure 5.4. Two types of $YP3^{1526}$ flies were used; the fs(1)A1526 strain, and the fertile $YP3^{1526}$ strains described in chapter 3.

Fat Body Expression

Expression of the transformed yp3 sequences in the fat body cells was examined initially by analysing the haemolymph protein samples by SDS-PAGE and staining the proteins in the gel with coomassie blue. Figure 5.5A shows the result from a gradient polyacrylamide-SDS gel. From this gel it is possible to see that a band with the same apparent molecular weight as YP3 is present in the haemolymph samples of at least

Figure 5.5

A. Coomassie blue stained 7-20% gradient polyacrylamide gel with haemolymph samples extracted from transformed females. The tracks are as follows:

Or^R: Oregon R, a wild type strain.

2B, 2R, 2S, 2Ka, 2Ja, 2Jb: Lines transformed with the P[UChsneoyp3.3•1] transposon.

$\frac{1526}{1526}$: Homozygous fs(1)A1526 flies.

Each track represents the proteins in the haemolymph of 4-5 females.
The positions of the yolk proteins are indicated.

B. Western blot of haemolymph samples from transformed females. The strain or transformed lines are indicated at the top of each track and are as follows:

Or^R: Oregon R, a wild type strain.

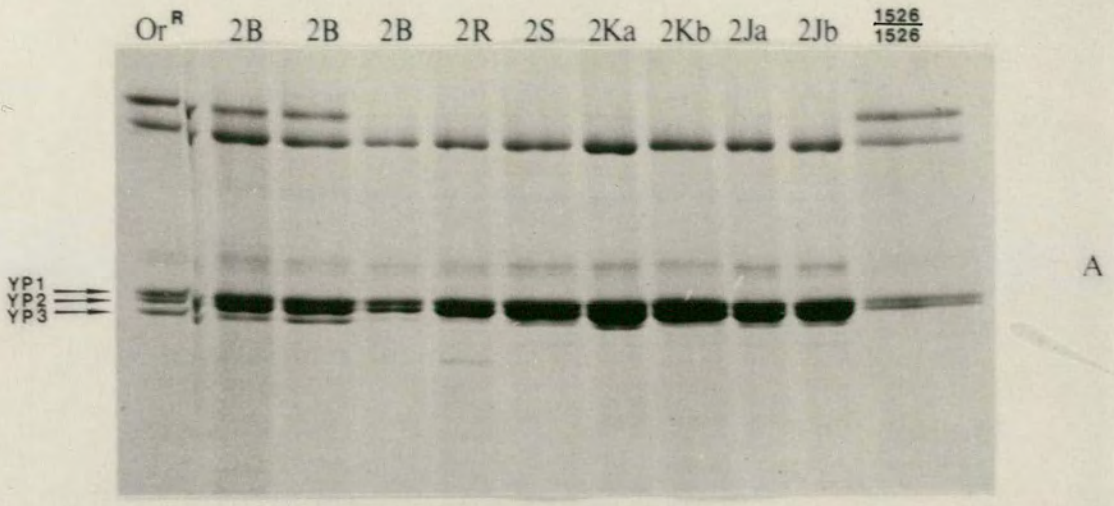
$\frac{1526}{FM3}$: Heterozygous fs(1)A1526.

$\frac{1526}{1526}$: Homozygous fs(1)A1526.

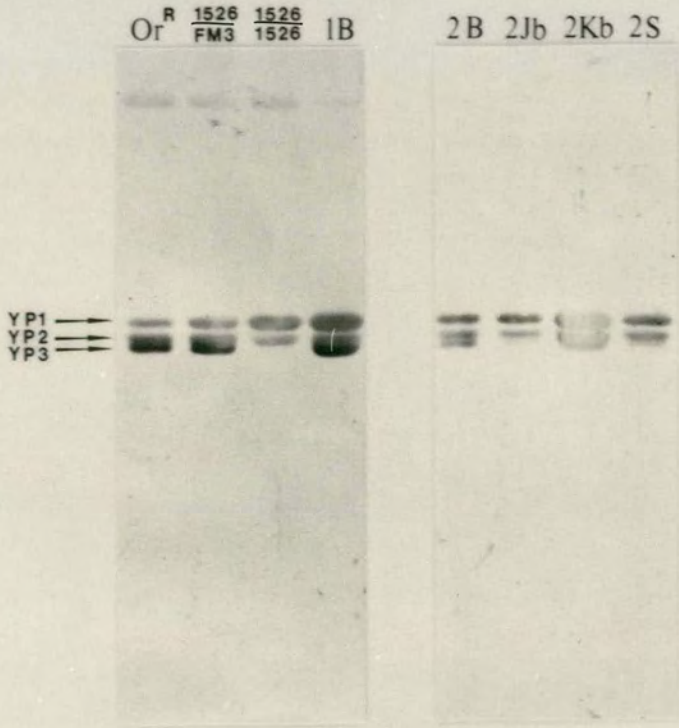
1B: A line of flies transformed with the P[UChsneoyp3.4•2] transposon.

2B, 2Jb, 2Kb, 2S: Lines of flies transformed with the P[UChsneoyp3.3•1] transposon.

Each track represents the proteins in the haemolymph of 2-3 females.
The YP3-β-galactosidase antiserum was used. The positions of the yolk proteins are indicated.
The gel used for the Western blot was a 10% linear polyacrylamide gel.


Or R
2B
2B
2B
2R
2S
2Ka
2Kb
2Ja
2Jb
1526/1526
YP1
YP2
YP3

A


Or R
1526/FM3
1526/1526
1B
2B
2Jb
2Kb
2S
YP1
YP2
YP3

B

some of the lines transformed with pUChsneoyp3/3.1. The level at which this protein is present differs between the transformed lines. Any difference in the level of YP3 present in the three samples of haemolymph detected from line 2B females is probably due to some of the flies sampled carrying one insertion per genome and others carrying two (line 2B was shown by Southern blotting to have 2 insertions of the transposon in its genome).

To prove conclusively that the faint band seen on such coomassie blue stained protein gels is YP3, the YP3-β-galactosidase antiserum (see chapter 3) was used in Western blotting analysis. The result of one Western blot of haemolymph samples from transformed flies is shown in figure 5.5B. It is clear from this blot that the band seen on coomassie gels is YP3. This proves that the line transformed with both pUChsneoyp3/4.2 and pUChsneoyp3/3.1 synthesise YP3 in the fat body cells which is secreted into the haemolymph. It is possible that line 1B (transformed with pUChsneoyp3/4.2) has synthesised larger amounts of YP3 than the shorter construct transformants (i.e. pUChsneoyp3/3.1). However, these experiments do not prove this for reasons to be discussed at the end of this chapter.

From the results presented it is apparent that both of the yp3 constructs introduced into the genome of the transformed flies contain the *cis*-sequences necessary for expression of yp3 in the fat body cells.

Ovary Expression

The transformed lines were examined to establish whether the yp3 sequences contained in the two constructs were sufficient to direct the expression of yp3 in the ovarian follicle cells.

It had been expected, on the basis of the yp3 transcript level initially detected in $YP3^{1526}$ females (see Fig. 3. A, in chapter 3), that transcription from the introduced yp3 gene would be detectable by an increase in the level of yp3 mRNA in the ovaries of the transformed lines. Therefore, Northern blot analysis was used to try to determine if ovarian transcription was occurring from the introduced YP3 genes. However, this approach proved to be unsuccessful since, although the yp3 mRNA level is reduced in $YP3^{1526}$ females, enough yp3 transcripts are present to mask low levels of transcription from an introduced yp3 gene. Therefore, YP3 protein was investigated.

Figure 5.6

Western blots of proteins present in the supernatant of ovary cultures, electrophoresed on linear 10% polyacrylamide gels. The strain/transformed lines sampled are indicated at the top of each lane. The YP3-β-galactosidase antiserum was used. The positions of the yolk proteins are indicated.

The samples in the lanes are as follows:

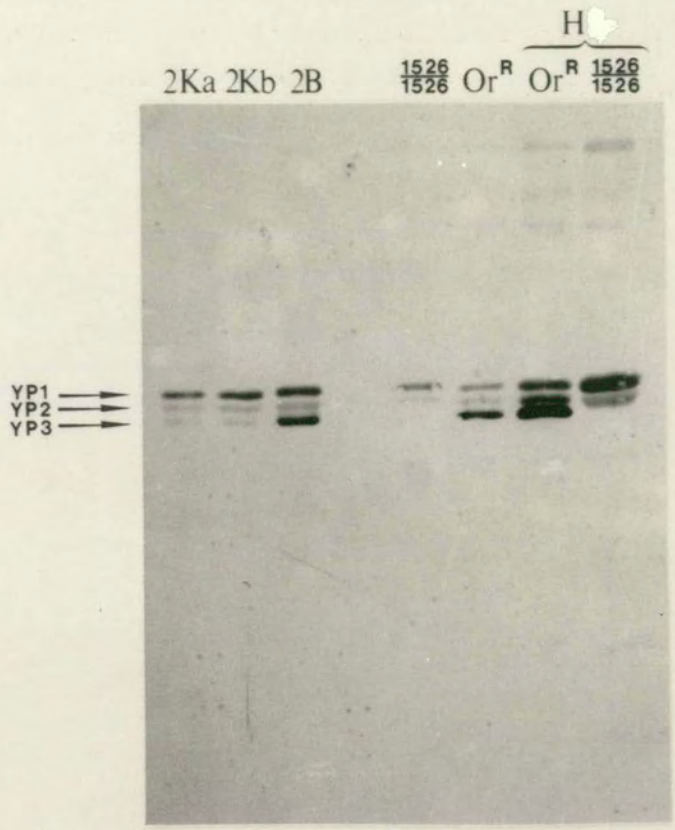
2Ka, 2Kb, 2B, 2S, 2Jb: Lines transformed with P[UChsneoyp3.3 • 1] transposon.

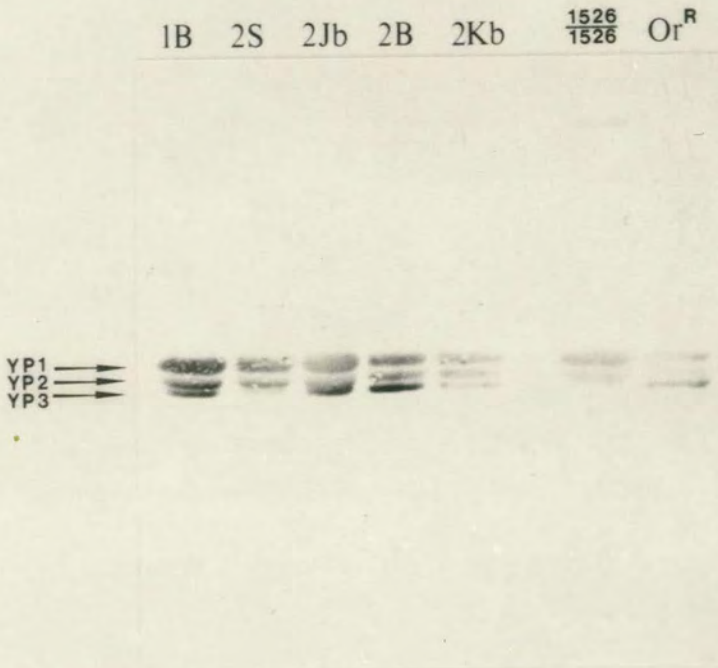
1B: Line transformed with the transposon P[UChsneoyp3.4 • 2].

1526 / 1526 : Homozygous fs(1)A1526

Or^R: Oregon R, a wild type strain.

H = Haemolymph samples


H
2Ka 2Kb 2B
1526/1526 Or^R Or^R 1526/1526
YP1
YP2
YP3


1B 2S 2Jb 2B 2Kb 1526/1526 Or^R
YP1
YP2
YP3

When examining the YP3 polypeptide in the ovaries of the transformed lines, it is necessary to look at *de novo* synthesis to ensure that the YP3 detected is synthesised by the ovarian follicle cells and not derived solely from the fat body cells. To do this use was made of the observation that *Drosophila* ovaries from adult females can be cultured *in vitro* and synthesise YP's which are exported into the culture medium (Hames and Bownes, 1978; Isaac, 1982). The ovaries from transformed females were cultured (as 2.4.2.3) and the proteins in the culture medium separated on 10% polyacrylamide-SDS gels which were Western blotted and probed with the YP3-β-galactosidase antiserum. Examples of two such blots are presented in figure 5.6. These Westerns show that YP3 is clearly present in the culture medium of ovaries from the transformant females and Oregon R females, but not the $YP3^{1526}$ control females. The conclusion that can be drawn here is that YP3 is synthesised in the ovaries of females transformed with either the 3.1kb or the 4.2kb yp3 gene fragment. Therefore, the *cis*-sequences necessary to specify yp3 transcription are contained within a 3.1kb DNA fragment.

Ecdysone Induction in Transformant Males

Drosophila males treated with large doses of the steroid hormone ecdysone have been shown to transiently synthesise the yolk proteins (Postlethwaite *et al.*, 1980; Bownes *et al.*, 1983).

Males from the transformed lines generated in this study were examined in an attempt to determine if the *cis*-acting sequences that mediate the response of ecdysone induced YP3 synthesise are present in either of the neo.yp3 constructs.

A Northern blot analysis of ecdysone induced YP3 synthesis in males proved to be inconclusive. The synthesis of $YP3^{1526}$ mRNA in males induced by 20-hydroxyecdysone attains a level similar to that achieved by wild type flies. Elevation above this level in yp3 transformant males was not detected, while this did not show that transcription was occurring, neither did it prove that transcription from the transformed yp3 genes was not taking place.

Since the Northern analysis was unsuccessful another approach was followed. Haemolymph samples collected from transformant males injected with 20-hydroxyecdysone were subjected to Western blot analysis using the YP3-β-galactosidase antiserum (whole male carcasses of hormone treated males were also examined).

Although the experiment was repeated many times, it was not successful in showing whether or not yp3 was synthesised in the ecdysone-induced transformed males.

Identification of Putative YP3 *cis*-Acting Regulatory Sequences by Homology

A region of DNA sequence 182bp in length, located upstream (nt -341 → -159) of the yp2 gene, has been identified which has a positive tissue-specific influence on the transcription of yp1 and yp2 in the ovary (Logan *et al.*, 1989). A second region 107bp long, within the first exon of yp2 (nt -2 → +105) has been shown to be necessary for transcription of yp1 in the ovarian follicle cells (Logan *et al.*, 1989). Since yp3 is also expressed in the ovary in adult females, it too must contain sequences which specify expression in these cells. These sequences may share some homology with the yp1/yp2 ovarian regulatory regions. A computer search for such regions of yp3 was carried out using the University of Wisconsin Genetics Computer Group Sequence Analysis Software Package, Version 5.2. This analysis has been performed for the yp1/yp2 fat body enhancer element (Garabedian *et al.*, 1987) and is discussed in the Introduction (in section on regulation of yp gene expression).

The output from the Bestfit program search of the yp3 sequence for homology with the 182bp long upstream ovarian regulatory element is shown in figure 5.7A. When a particularly A + T rich region of the regulatory sequence was ignored and the search repeated, some other regions were identified (Fig. 5.7B and C). The upstream ovarian element has a high A + T content, so it is possible that the identification of these homologous regions may simply be due to the relatively high A + T content of the sequences. However, this does not rule out the possibility that these sequences may be important in the regulation of yp3 transcription in the ovary.

Some areas of the yp3 sequence that show homology to the other yp1/yp2 regulatory element found in the first 105bp of the yp2 transcript are shown in figure 5.8. Sequence A is identified as the best fitting sequence, this region is located within the first exon of yp3 in an almost identical position to that taken yp by the yp1/yp2 regulatory sequences in the yp2 gene. However, the same region is identified in the yp1 gene (when its sequence is compared to the yp2 regulatory region) and it did not specify yp1 ovarian expression in transformation assays (Garabedian *et al.*, 1985). It seems most likely that this region in yp3 shows such a high level of homology (70-75%)

Figure 5.7

Regions of the yp3 gene which display similarity to the upstream ovarian enhancer identified in the yp2 gene (Logan *et al.*, 1989).

The upper line in each case is the yp2 sequence and the lower line is the yp3 sequence. The numbering alongside the sequences show where the regions are located relative to the transcript start sites of the genes.

Figure 5.7

YP2 upstream element

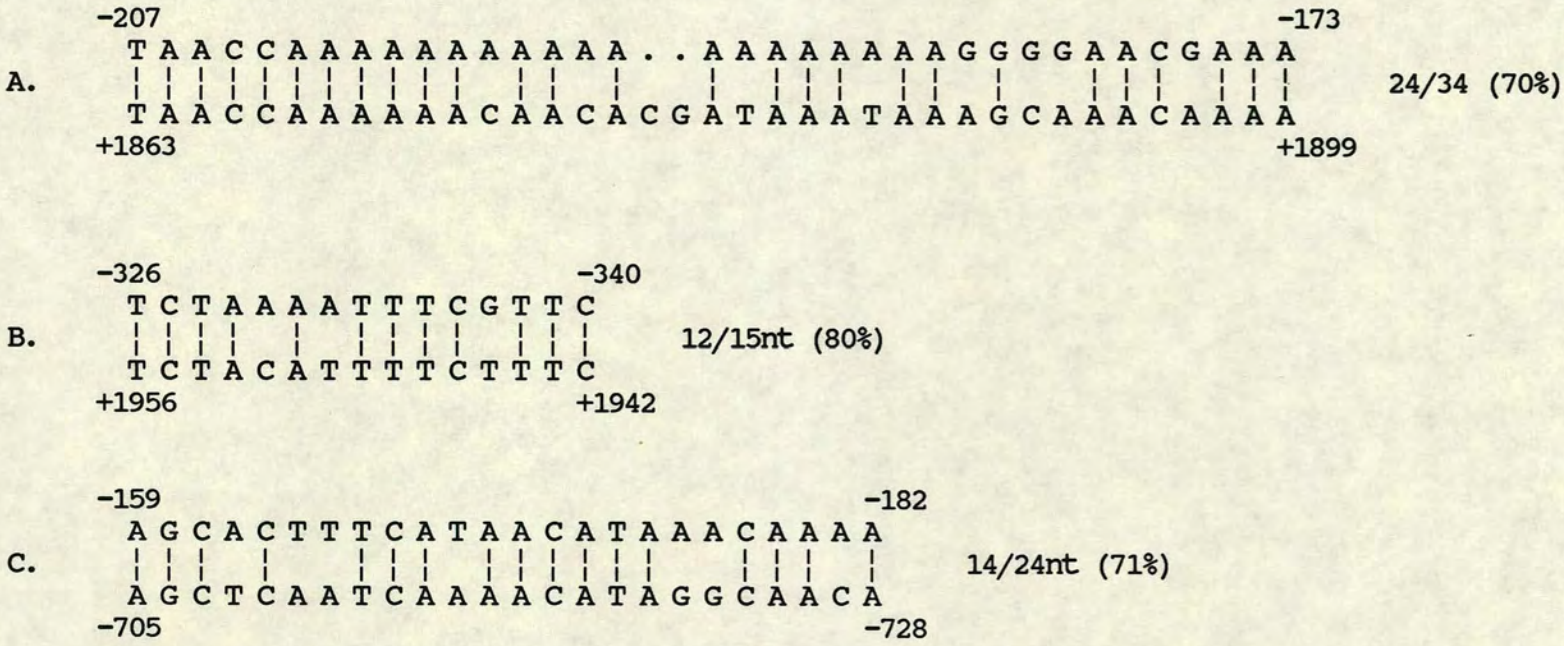


Figure 5.8

Regions of the yp3 gene which display similarity to the intragenic ovarian enhancer element identified in yp2 (Logan *et al.*, 1989).

The upper line in each case is the yp2 sequence and the lower line is the yp3 sequence. The numbering alongside the sequences show where the regions are located relative to the transcript start sites of the genes.

Figure 5.8

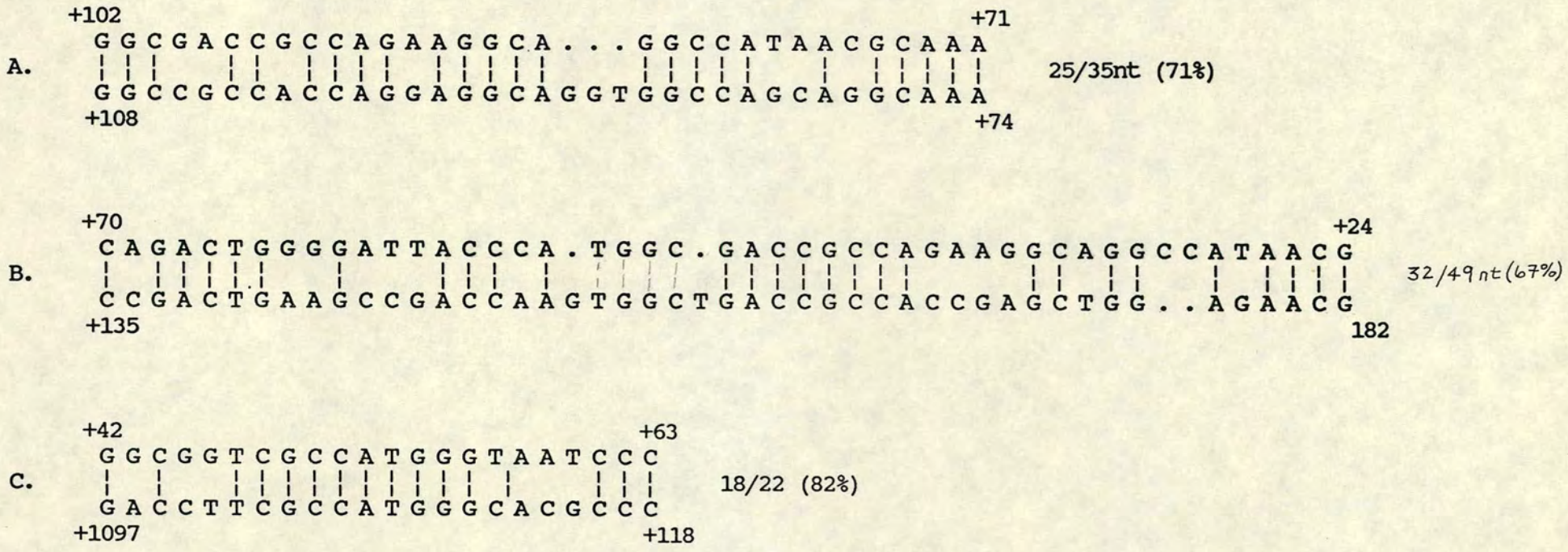

A.
+102
+71
G G C G A C C G C C A G A A G G C A . . . G G C C A T A A C G C A A A
G G C C G C C A C C A G G A G G C A G G T G G C C A G C A G G C A A A
+108
+74
25/35nt (71%)
B.
+70
+24
C A G A C T G G G G A T T A C C C A . T G G C . G A C C G C C A G A A G G C A G G C C A T A A C G
C C G A C T G A A G C C G A C C A A G T G G C T G A C C G C C A C C G A G C T G G . . A G A A C G
+135
182
32/49 nt (67%)
C.
+42
+63
G G C G G T C G C C A T G G G T A A T C C C
G A C C T T C G C C A T G G G C A C G C C C
+1097
+118
18/22 (82%)

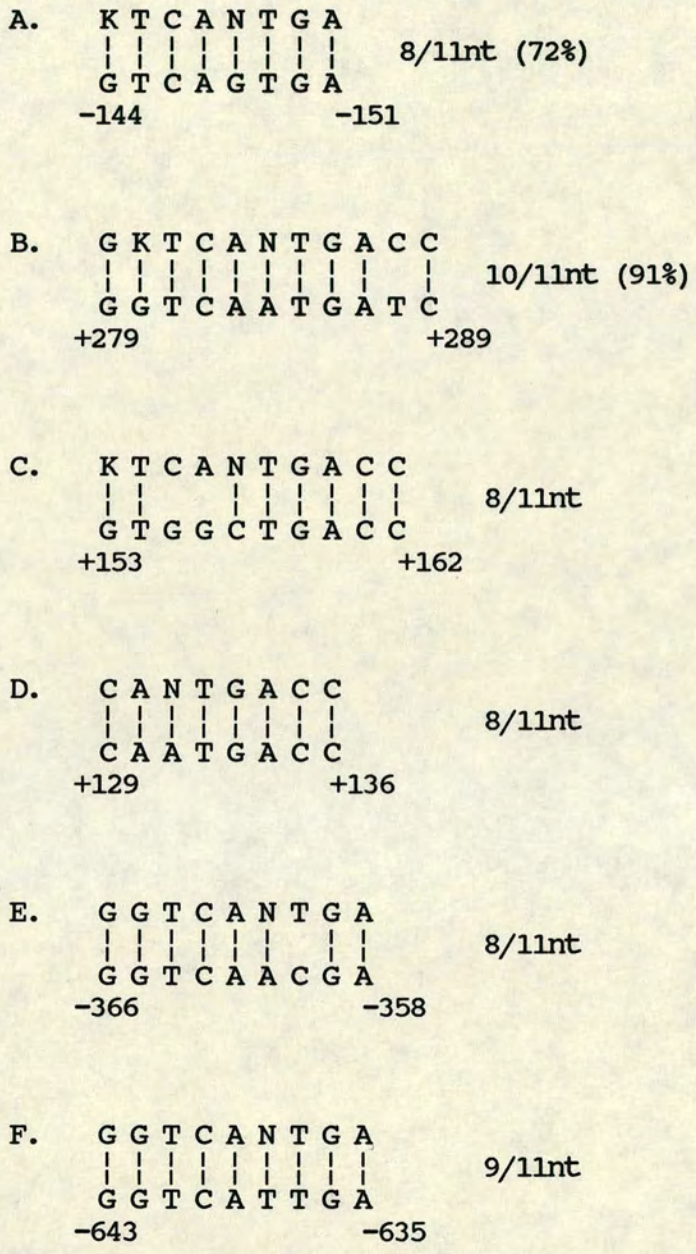
Figure 5.9

Regions of the yp1/yp2 genes (A) and the yp3 gene (B → F) which display similarity to the ecdysone inducible consensus sequence element.

The upper line in each case is the consensus ecdysone response element and the lower line is yp1/yp2 (A) or yp3 (B → F) sequences. The numbers alongside the sequence show where the region is located relative to the transcript start sites of the genes.

N = any nucleotide
K = G or T

Figure 5.9


A.
K T C A N T G A
G T C A G T G A
8/11nt (72%)
-144
-151
B.
G K T C A N T G A C C
G G T C A A T G A T C
10/11nt (91%)
+279
+289
C.
K T C A N T G A C C
G T G G C T G A C C
8/11nt
+153
+162
D.
C A N T G A C C
C A A T G A C C
8/11nt
+129
+136
E.
G G T C A N T G A
G G T C A A C G A
8/11nt
-366
-358
F.
G G T C A N T G A
G G T C A T T G A
9/11nt
-643
-635

with the yp2 ovarian regulatory element simply due to the level of conservation between the three yp genes (even though the 5′ end of the coding sequence is the area of yp1, 2 and 3 which, as a whole, shows least similarity). However, it is possible that subregions of the areas of notable homology (Fig. 5.8B and C) in yp3 are important in the ovarian regulation.

A consensus sequence, 11bp in length, that specifies an ecdysone-induced response, has been derived from studies with the ecdysone inducible genes *eip* 28/29 and one heat shock gene (L. and P. Cherbas, personal communication). A computer search was performed with this ecdysone response element sequence to detect any similar sequences within the yp1, yp2 and yp3 sequences. Only one sequence was detected in the yp1 and yp2 sequences that shows homology (longer than 5 bp) (Fig. 5.9A). However, in yp3 several sequences showing similarities to the ecdysone response consensus sequence were identified (Fig. 5.9B → F). Of these, sequence B (Fig. 5.9) is most similar with 10/11 nts matching those in the ecdysone element, the sequence is located at the beginning of exon 3 in the yp3 sequence.

It should be noted that two of the ecdysone response elements identified for the *eip* 28/29 genes are situated approximately 2kb and 4kb downstream of the coding sequences of the genes. Since only 0.8kb of the sequence 3′ to the coding regions of yp3 has been determined, any ecdysone response elements homologous to the identified consensus sequence that are located beyond this boundary cannot be identified in this search. If any of the elements involved in regulating the yp3 response to ecdysone are further than 1.8kb upstream of the transcription start site, these too will be missed in the search for homologous sequences. Similarly, only 0.8kb of sequence 3′ of the yp2 gene have been determined, and again any ecdysone regulatory elements beyond these sequences will not be detected in this search.

The relevance of the YP3 regions identified in this computer search is included in the discussion at the end of this chapter.

DISCUSSION

The complete unaltered yp3 gene from a wild type strain of *Drosophila melanogaster* (Canton S) was cloned into the P-element transformation vector pUChsneo. The recombinant plasmid, pUChsneoyp3/4.2, carries a 4.2kb HindIII-EcoRI fragment (taken from pYP3, Barnett *et al.*, 1980) which contains all of the coding

sequences of yp3, 1.8kb of 5′ flanking DNA and 0.8kb of 3′ flanking DNA (from the second consensus site for polyA addition). A second recombinant plasmid, pUChsneoyp3/3.1, was created by removing a 1.1kb fragment from the 5′ end of the yp3 upstream region (leaving 0.7kb of 5′ sequences). These two plasmid constructions, bearing different lengths of sequence 5′ to the coding region of yp3, were transformed into wild type (Oregon R) embryos. Two lines of flies transformed with construct pUChsneo yp3/4.2 and six lines transformed with pUChsneoyp3/3.1 were established.

To study expression from the injected yp3 sequences, crosses were performed to obtain the transformed constructs in the $YP3^{1526}$ genetic background. Flies homozygous for the mutant yp3 gene from the strain fs(1)A1526 produce YP3 that is of a higher molecular weight than wild type YP3, and which is not secreted from the cells where it is synthesised (Chapter 3). Since no YP3 is found in the haemolymph or oocytes, strains that are homozygous for the $yp3^{1526}$ gene were used in this transformation study to assay expression from the wild type YP3 sequences.

Samples of haemolymph proteins from transformed females were analysed using coomassie blue staining of SDS-polyacrylamide gels and by Western blotting using an antibody raised against a YP3-β-galactosidase fusion protein (described in chapter 3). The results of these experiments illustrate that YP3 is synthesised and secreted from the fat body cells of adult females transformed with both the 4.2kb and 3.1kb yp3 gene fragments. Thus, the *cis*-acting sequences necessary to specify expression of YP3 in the cells of the adult female fat body are contained with in a 3.1kb fragment of DNA.

Ovaries dissected from adult female flies were cultured *in vitro* in order to determine whether expression from the transformed yp3 sequences was occurring in the other site of yp synthesis - the ovarian follicle cells. Proteins excreted from the ovaries into the culture medium were examined by Western blotting of SDS-polyacrylamide gels, again using the YP3-β-galactosidase antiserum. These experiments proved that YP3 is synthesised by the ovary in the females transformed with the 4.2kb and the 3.1kb yp3 gene fragments. This indicates that, in addition to containing the DNA-sequences necessary for fat body specific expression of yp3, the 3.1kb fragment contains the *cis*-acting regulatory sequences required to specify transcription of yp3 in the ovary.

Thus, it has been shown that the *cis*-acting regions conferring tissue specific expression of yp3 in *D. melanogaster* females are found within a 3.1kb fragment of DNA. Exactly where the tissue-regulatory regions are located within this sequence is unknown. To determine where the regulatory regions are situated within the yp3

sequences, more transformation experiments will have to be performed using other constructs bearing various amounts of yp3 sequence. Some putative *cis* regulatory sequences of yp3 have been identified by virtue of their homology with regulatory elements that have been defined by transformation studies with the yp1 and yp2 genes (see below for this). Although very little can be deduced regarding any actual role of these sequences in the regulation of yp3, they do indicate useful starting points in the design of future plasmid constructions to further identify the *cis*-regulatory regions of yp3 in transformation experiments.

The sequence of yp3 was compared with the sequence elements which were identified in transformation studies with the yp1 and yp2 genes (Garabedian *et al.*, 1986; Logan *et al.*, 1989). The yp3 sequences identified by a computer search, that show homology with the ovarian regulatory regions, are detailed in figures 5.7 and 5.8 (the yp3 sequences similar to the yp1/yp2 fat body enhancer are presented in figure 1.5 of the introduction).

A DNA sequence located upstream of the yp3 transcription start site shows homology with one of the regions important for the expression of yp1/yp2 in the ovarian follicle (Fig. 5.7) (Logan *et al.*, 1989). This is found at nt -728 → -706 in the 5′ flanking sequence of yp3, and shows 71% homology with nt -182 → -159 of the upstream yp1/yp2 ovarian regulatory element. It could be speculated that this region is involved in specifying expression of yp3 in the ovarian follicle cells. However, the results of the transformations performed with yp3 presented here show that sequences located more than 705bp upstream of the yp3 gene are not necessary for the expression of yp3 in the ovary, i.e. the upstream yp3 region identified by the computer search is not included in the 3.1kb yp3 fragment which is sufficient to confer yp3 transcription in the ovary). However, the possibility that this region has some involvement in regulating the synthesis of yp3 in the ovary, for example it may act to increase the levels synthesised, has not been disproven by these transformation experiments.

The sequence upstream of yp3 (nt -924 → -909) identified as sharing homology with the yp1/yp2 fat body enhancer sequence (Garabedian *et al.*, 1987), is also deleted from the 3.1kb fragment of yp3. The yp3 sequences within the 3.1kb fragment have been shown in this study to be sufficient to specify yp3 expression in the cells of the adult female fat body, therefore it can be concluded that the upstream region similar to the yp1/yp2 fat body element is not necessary for initiating the transcription of yp3 in this tissue. Again, it is still possible that this region of yp3 is in some way involved

with the expression of yp3 in the fat body cells. Further analysis of the YP3 levels obtained in lines transformed with the two constructs described here will elucidate whether this region is involved in regulating the level of yp3 transcription.

Of the other yp3 sequences showing similarity to the yp1/yp2 ovarian regulatory elements that were identified in the computer search, no strong conclusions can be drawn concerning whether any of these actually are involved in regulating yp3 expression in the ovary. YP3 is present at a lower level than YP1 and YP2 in the ovarian follicle cells (Isaac and Bownes, 1982; Brennan *et al.*, 1982), thus it is possible that the region(s) regulating yp3 expression in these cells will not be identical and may in fact be fairly dissimilar. Although this may not be the case if the yp3 transcript is less stable in the ovarian follicle cells (Williams and Bownes, 1986).

Most of the transformation studies with yp1 and yp2 have involved altering the genes in some way (e.g. the M13-tagged yp1 and yp2 genes in Garabedian *et al.*, 1985, and the truncated yp1 gene in Tamura *et al.*, 1985). Such modifications can affect the stability of the transcripts and alter their steady state levels and so equating the levels of expression of modified sequences with the levels of the normal transcript is not possible. Transformation studies which involve altering the gene cannot therefore be used to determine which sequences are necessary for normal levels of expression. Transcript stability proved to be a problem in the early stages of the transformation studies using M13-tagged yp genes to identify tissue regulatory elements. Although the problem was overcome for the yp1 and yp2 genes (Garabedian *et al.*, 1985) no success has been achieved in creating an M13-tagged yp3 gene that gives rise to a stable yp3 transcript (P. Wensink, personal communication).

It is also possible that such gene modifications may disrupt some of the *cis*-acting regulatory sequences. This is a particularly valid point to bear in mind when considering the yolk protein genes since an ovarian regulatory region has now been identified within the coding sequence of the yp2 gene (Logan *et al.*, 1989).

There are also disadvantages associated with using reporter genes, such as alcohol dehydrogenase (adh) and β-galactosidase, in transformation studies. The point to be noted here is that there is no effective way to relate the level of activity measured in, for example, an adh activity determination, with the actual level of the protein normally produced from the promoter region being studied. Similarly, it is not possible to equate the transcript level achieved using a reporter sequence with the level of the actual transcript normally produced. This is due to different mRNA species having

different and characteristic rates of turnover, resulting in different steady state levels. Reporter genes are, however, invaluable in determining the relative strengths of different promoter or enhancer regions, and in identifying tissue-specific regulatory elements.

Noting the points discussed above, perhaps the only proper way to identify the *cis*-acting sequences that are necessary to achieve the normal level of synthesis from a particular gene is to use the unaltered gene itself. In order to do this in *Drosophila* it is necessary to have some way of distinguishing expression from the transformed gene and the endogenous gene, therefore mutants have to be used.

Only one other transformation study (involving the yolk proteins) has made use of such an approach. A YP2 mutant was used by Tamura *et al.* (1985) when they investigated the *cis*-regulatory elements of the yp1/yp2 region (for some details see p. 22 in the Introduction). Their experiments indicated that the *cis*-acting region responsible for synthesis of YP2 in the fat body cells is contained within a 5.0kb genomic yp1/yp2 fragment. The transformation experiments described in the present study provide a system where it will be possible to elucidate at least some *cis*-acting sequences that are involved in regulating the level at which one member of the yolk protein gene family is synthesised. In a system of this type, however, variables are still present. It has recently been demonstrated that the DNA sequences around (but not necessarily nearby) the site of insertion into the chromosome can have an effect on the transcription of the transformed sequences (Wilson *et al.*, 1989). Use is being made of this to identify genes of interest, for instance genes that are expressed in a tissue specific or temporally controlled manner (e.g. Bellen *et al.*, 1989). However, in a study aiming to determine *cis*-acting regulatory regions of transformed genes, such position-dependent effects can alter the pattern of expression obtained from a sequence inserted into the genome. For this reason, in order to deduce the *cis*-controlling sequences of a transformed sequence it is necessary to study its expression pattern in a number of different transformed lines where the insertion is known to be in different genomic locations. One other drawback in using the complete, unaltered gene is that it is not possible to identify regulatory elements that lie within the transcription unit. Again, this is because modification of the transcript may affect the turnover rate, and alterations in the coding regions may affect the viability of the polypeptide, e.g. some changes in the amino acid sequence could result in incorrect folding of the protein which may then be recognised as being faulty and become a target for degradation by proteases.

A recombinant chromosome bearing the $yp3^{1526}$ gene and the introduced yp3 sequences would be useful in the investigation of the YP3 *cis*-acting sequences. For example, with the transformed yp3 sequences on the same chromosome as the defective yp3 gene both can more easily be obtained in another genetic background, such as *tra-*2^{ts1} or ecd^1. This would facilitate the study of the effects of the products of other loci on the expression of the transformed yp3 gene and allow identification of the *cis*-sequences that interact with such *trans*-regulatory loci.

One of the transformed lines created in this study bearing a pUChsneoyp3/4.2 insertion (line 1A) was not examined by Southern blotting. The genetic analysis had shown that it carries an X-chromosomal insertion and to use this strain to study expression from the introduced yp3 gene it is necessary to recombine the insertion on to an X-chromosome with the mutant yp3 locus (i.e. $yp3^{1526}$). This should be possible since a preliminary linkage map indicated that the insertion is situated to the left of the miniature locus, somewhere between 10.5 and 27 recombination map units on the X-chromosome. Since the yp3 locus is at approximately 44 map units, recombinant chromosomes resulting from a crossover between the neo^R insertion and yp3 locus should be fairly easily generated. These can be identified by selection for the insertion by neomycin, and Western blotting of carcass samples (as in Fig. 3.7a, Chapter 3) to determine whether the yp3 at the normal locus is the wild type copy or the mutant copy $yp3^{1526}$.

Sex Specific Regulation

The results discussed here provide the basis of a system which can be used to assess the regulation of the yp genes in *Drosophila* males. Only females normally synthesise the yolk proteins, no yolk polypeptides or yp transcripts are detected in males. Regulation of yp expression males is achieved at the level of yp gene transcription. This is governed by the regulatory genes determining somatic sexual differentiation in *Drosophila* (Belote *et al.*, 1985; Kraus *et al.*, 1988). It would be interesting to obtain the transformed yp3 sequences in a genetic background that contains mutant alleles for one of the genes involved in the sex-determination pathway (e.g. *tra-*2^{ts2}, *dsx*) and determine if yp3 expression occurs. This would reveal whether the *cis*-sequences responsible for the regulation of yp synthesis by a particular sex-gene product are present within the transformed fragment of yp3 DNA. Another study using

a yp1/Adh fusion gene has established that the sequences responsible for mediating the sex-specific response of yp1 in the fat body cells to the product of the *doublesex* locus is located within a 939bp yp1 gene fragment.

The yp3 fragments studied here are sufficient to modulate the correct sex-limited expression of yp3. No wild type YP3 has been detected in male haemolymph or carcass samples from any of the transformed lines. It is theoretically possible that YP3 is synthesised in these males, however, it would need to be at a low level and be degraded rapidly (however, it is known that YP's produced in ecdysone treated males are fairly stable, Kozma and Bownes, 1986b).

Ecdysone Induction of yp3 in Males

Drosophila melanogaster males treated with the hormone 20-hydroxyecdysone can be induced to express the yolk protein genes (Postlethwaite *et al.*, 1980; Bownes *et al.*, 1983). Attempts were made to ascertain if the yp3 fragments analysed in the transformation study presented here contain the DNA sequences that mediate this response to ecdysone. As reported in the results section of this chapter, these experiments were not successful in elucidating whether or not ecdysone induced synthesis of YP3 occurs in the transformed males. Problems which are noted with respect to why these experiments were inconclusive are outlined below.

Induction of any YP synthesis in male flies reared on food supplemented with the antibiotic geneticin was generally very difficult to achieve, even when the flies were transferred to normal, yeast-supplemented food at eclosion. This may be due to detrimental effects of the antibiotic on even neomycin-resistant flies, for example, the time taken for neo^R eggs laid on G418 food to eclose from the pupal case can be twice that taken by wild type embryos on normal food.

Ecdysone induction of YP synthesis in $YP3^{1526}$ stains was also more difficult to obtain than for Oregon R males. The reason for this is unclear, it may be that the presence of other mutations in the flies (such as yellow and white, or other lesions generated in the fs(1)A1526 chromosome by the EMS induced mutagenesis by Gans *et al.*, 1975) is detrimental to the flies in such a way that it interferes with the hormone induction. One of the major problems is that males treated with ecdysone become quite dehydrated, all of the flies used here became very dehydrated sooner than the wild type males, and so collection of haemolymph samples from these flies was very difficult.

yp Transcription - Ecdysone Regulation

Ecdysone has been shown to affect the synthesis of the yp genes in male and female *D. melanogaster* in a variety of experimental systems (see Regulation of Vitellogenesis - Hormones; in Introduction). However, the precise role of ecdysone in the normal regulation of yp synthesis in the fat body cells of the adult female is still unknown. Even in other insects such as the *tse-tse* fly and mosquito, where the role of ecdysteroids in modulating the expression of the vitellogenin genes has been more clearly established (e.g. Hagedorn, 1989), the mode of action by which this is achieved is not understood in detail.

A computer-assisted search was undertaken to determine whether the yp genes contain any ecdysone regulatory-like sequences. Regions of the yp genes that display similarity with a consensus sequence shown to be involved in the ecdysone induction of one of the hsp genes and the *eip* 28/29 genes are shown in figure 5.9. One region in the yp1/yp2 genes is presented, this is the only sequence identified that shares more than 6bp similarity with the ecdysone regulatory sequence. This region is located 500bp upstream of the yp2 transcript initiation site, however ecdysone induction has not been achieved in males transformed with several different yp1/yp2 constructs which include this region. It is possible that this sequence is involved in mediating the ecdysone induced transcription of the yp1 and yp2 genes, but requires additional sequences located 3′ of one or both of the genes, as in the case in the *eip* 28/29 genes. These genes have one element upstream and two downstream of the coding region, however, only the downstream elements are able to confer ecdysone inducibility in transformed tissue culture cells (L. and P. Cherbas, personal communication).

Several regions of yp3 which display notable similarity with the consensus ecdysone response sequence element were noted in the computer search (Fig. 5.9). Whether these sequences are involved in the yp3 response to ecdysone treatment can be established by transformation experiments where these sequences are modified or deleted. However, most of the currently available evidence indicates that ecdysone is not acting directly to effect induction of yp transcription and that other factors are likely to be involved. A study using inhibitor of protein synthesis, indicated that some other protein factor is synthesised in ecdysone-treated flies that is required to mediate transcription of the yp genes (Bownes *et al.*, 1987).

The yolk protein genes would therefore appear to be in a class of ecdysone regulated genes where induction requires other gene products and whose transcripts increase only slightly on treatment with the hormone. It is unclear whether such genes contain sequences which interact directly with the hormone receptor complex or are subject to indirect effects of the hormone on general cellular components. Thus, the putative *cis*-ecdysone regulatory sequences present in the yolk proteins may have been identified on a purely random manner, and that some other DNA sequences are responsible for mediating the yp ecdysone response via a *trans*-acting polypeptide factor. None of the sequences identified in the computer search are actually identical to the consensus ecdysone responsive element. It is even possible that these regions of the yp's represent redundant sequences left over from a time when an "ancestral *Drosophila*" species used ecdysteroids to directly regulate yp synthesis.

Regardless of the identity of the *trans*-acting factor which mediates the ecdysone induction of the yp's, transformation studies should identify the yp *cis*-sequences involved. A transformation study using a yp1/Adh fusion gene to measure yp transcription has already shown that the *cis*-acting sequences which mediate the yp gene response in males is outwith the normal female developmental controls (Shirras and Bownes, 1987).

FINAL COMMENTS

This study describes a system whereby examination of *cis*-regulatory regions specifying transcription levels can be identified. The experiments performed here have shown that a 3.1kb yp3 fragment contains the DNA sequences required to specify transcription in the ovary and the fat body cells of the adult female. More precise determination of the important sequences requires further transformation experiments using deleted yp3 fragments. To enable determination of *cis*-sequences regulating the actual levels of yp3 produced in these tissues it would be necessary to (1) examine a larger number of transformed lines (particularly for the longer construct pUChsneoyp3/4.2, since expression was examined in only one line transformed with this plasmid), and (2) to use a known copy number of the gene for which expression is being measured (thus the transformed lines have to be made isogenic and homozygous). Ideally, transformation experiments investigating *cis*-acting regulatory regions would involve using both approaches, so that results obtained with the wild type unaltered

genes can be supported by studies using reporter genes or "tagged" genes to positively identify important regulatory elements. Another transformation study which has been set up to investigate YP secretion, and uptake by oocytes, uses selection of transformants based on eye colour (K. Lineruth and M Bownes). It is possible that in this study it will be feasible to determine whether the ecdysone inducible elements are present within the 3.1kb yp3 fragment as neomycin supplemented food is not used, therefore induction of yp expression should not be a problem. When the *cis*-regulatory sequences have been further defined it will be possible to begin to investigate the trans-factors concerned and analyse the interactions which govern the regulation of yp3 expression.

CONCLUSIONS AND PERSPECTIVES

CONCLUSIONS AND PERSPECTIVES

The experiments described in this investigation of yolk protein 3 in *Drosophila melanogaster* have illustrated a number of points.

Genetic analysis has shown that in fs(1)A1526 flies, the mutant aspects of the YP3 phenotype (the non-secretion and increased molecular weight of the polypeptide, and the reduction in the transcript level) are not related to the female sterile effect. I separated the two lesions by recombination and established fertile lines which contain the $yp3^{1526}$ gene. These flies are fertile even although no YP3 is present in the eggs, proving that all three YP's are not essential for normal egg development and that embryos containing only YP1 and YP2 are viable.

In the molecular analysis of yolk protein 3 in fs(1)A1526 I showed that the YP3 polypeptide is synthesised in the fat body cells but fails to be exported to the haemolymph. The YP3 produced shows an increase in molecular weight by approximately 1.0-1.5kDa as estimated from SDS-PAGE. These aspects of the $YP3^{1526}$ polypeptide could be explicable if the signal sequence fails to direct the protein into the lumen of the ER. I determined the $yp3^{1526}$ nucleotide sequence and identified a mutation which results in the substitution of an aspartic acid residue in place of an alanine residue at position 10 in the signal peptide region of the protein. This lesion is likely to be the cause of the secretion defect of $YP3^{1526}$ and could be responsible for the increase in molecular weight of $YP3^{1526}$ compared to the wild type protein. It is possible that this same lesion also causes the reduction in the steady state level of $yp3^{1526}$ mRNA, however, this has not been proven and could be caused by one or more of the altered bases identified in the upstream region of the $yp3^{1526}$ gene. These sequence alterations may be the result of polymorphisms between strains, or ethyl methyl sulphonate (EMS) induced mutations.

This leaves several questions regarding $yp3^{1526}$, such as whether one mutation causes all three mutant aspects of yolk protein 3. These queries can be resolved by transformation experiments. That the alteration in the $YP3^{1526}$ signal sequence may cause the secretion defect can also be proved by similar transformation experiments in *Drosophila*; and the stage during the process of secretion at which the mutation interferes can be investigated in a cell-free *in vitro* translation/translocation system or in yeast.

For the second part of this molecular investigation into yp3, I used the P-

element mediated germline transformation system to successfully determine *cis*-regulatory regions specifying the sex-limited and tissue-specific expression of yp3. The sequences responsible for regulating the sex- and tissue-specific synthesis of yp3 are contained within a 3.1kb genomic yp3 fragment, more precise determination of these sequences can now be investigated using this system with fragments of yp3 varying in the length of 5′ and 3′ flanking sequences.

REFERENCES

Akam, M. (1987) Development **101**: 1-22.

Andrews, D.W., Perera, E., Lesser, C., and Lingappa, V. (1988) J. Biol. Chem. 263: 15791-15798

Baeuerle, P.A. and Huttner, W.B. (1985) J. Biol. Chem. **260**: 6434-6439.

Baeuerle, P.A., Lottspeich, F. and Huttner, W.B. (1988) J. Biol. Chem. **263**: 14925-19429.

Baker, B.S. (1989) Nature **340**: 521-524.

Bankaitis, V.A., Ryan, P.J., Rasmussen, B.A. and Bassford Jr., P.J. (1985) Current Topics in Membranes and Transport **24**: 105-150.

Barnett, T., Pachl, C., Gergen, J.P. and Wensink, P.C. (1980) Cell **21**, 729-738.

Barnett, T. and Wensink, P.C. (1981) In: ULCA-ICN Symposium, vol. 23 (Eds. Brown, D.D. and Fox, C.F.), pp. 97-106.

Bedouelle, H., Bassford Jr., P.J., Fowler, A.V., Zabin, I., Beckwith, J. and Hofnung, M. (1980) Nature **285**: 78-81.

Bellen, H.J., O'Kane, C.J., Wilson, C., Grossniklaus, U., Pearson, R.K. and Gehring, W.J. (1989) Genes Dev. **3**: 1288-1300.

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

Benson, S.A., Hall, M.N. and Silhavy, T.J. (1985) Ann. Rev. Biochem. **54**: 101-134.

Benyajati, C., Place, A.R., Wang, N., Pentz, E. and Sofer, W. (1982) Nucl. Acids Res. **10**: 7261-7272.

Bernstein, P. and Ross, J. (1989) T.I.B.S. **14**: 373-377.

Biggin, M.D. and Tjian, R. (1989) T.I.G.S. **5**: 377-383.

Blobel, G. (1987) Hepatology **7**: 26-29.

Blumenthal, T. and Zucker-Aprison, E. (1987) In: Molecular Biology of Invertebrate Development. UCLA Symposium on Molecular and Cellular Biology. New Series, vol. 66 (Ed. O'Connor, J.D.), Alan R. Liss Inc., New York, pp. 3-19.

Bochkareva, E.S., Lissin, N.M. and Girshovich, A.S. (1988) Nature **336**: 254-257.

Böhni, P.C., Deshaies, R.J. and Schekman, R. (1988) J. Cell. Biol. **106**: 1035-1042.

Bownes, M. (1975) J. Embryol. Exp. Morph. **33**: 789-801.

Bownes, M. (1982a) Quart. Rev. Biol. **57**: 247-274.

Bownes, M. (1982b) J. Insect Physiol. **28**: 317-328.

Bownes, M. (1989) J. Insect Physiol. **35**: 409-413.

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

Bownes, M. and Blair, M. (1986) J. Insect Physiol. **32**: 493-501.

Bownes, M. and Dale, L. (1982) Gametogenesis. In: A Handbook of *Drosophila* development. (Ed. Ransom, R.), Elsevier Biomedical Press, Amsterdam, pp. 31-66/

Bownes, M. and Hames, B.D. (1977) J. Exp. Zool. **200**: 149-156.

Bownes, M. and Hames, B.D. (1978) FEBS Lett. **96**: 327-330.

Bownes, M. and Hodson, B.A. (1980) Mol. Gen. Genetic. **180**: 411-418.

Bownes, M. and Nöthiger, R. (1981) Mol. Gen. Genet. **182**: 222-228.

Bownes, M. and Rembold, H. (1987) FEBS **164**: 709-712.

Bownes, M. Scott, A. and Blair, M.G. (1987) Development **101**: 931-941.

Bownes, M., Scott, A. and Shirras, A. (1988a) Development **103**: 119-128.

Bownes, M., Shirras, A.D., Blair, M.G., Collins, J. and Coulson, A. (1988b) Proc. Natl. Acad. Sci. USA **85**: 1554-1557.

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

Brawerman, G. (1989) Cell **57**: 9-10.

Bregliano, J.-C. and Kidwell, M.G. (1983) Hybrid Dysgenesis Determinants. In: Mobile Genetic Elements (Ed. Shapiro), Academic Press, New York.

Brennan, M.D., Warren, T.G. and Mahowald, A.P. (1980) J. Cell Biol. **87**: 516-520.

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

Brennan, M.D. and Mahowald, A.P. (1982) Insect Biochem. **12**: 669-673.

Briggs, M.S., Cornell, D.G., Dluhy, R.A. and Gierasch, L.M. (1986) Science **233**: 206-208.

Briggs, M.S. and Gierasch, L.M. (1986) Adv. Protein Chem. **38**: 109-180.

Chirico, W., Waters, and Blobel, G. (1988) Nature **332**: 805-810.

Chou, P.Y. and Fassman, G.D. (1978a) Adv. Enzymol. **47**: 45-148.

Chou, P.Y and Fassman, G.D. (1978b) Ann. Rev. Biochem. **47**: 251-276.

Cleveland, D.W. (1988) T.I.B.S. **13**: 339-343.

Collier, D.N., Bankaitis, V.A., Weiss, J.B. and Bassford Jr., P.J. (1988) Cell **53**: 273-283.

Connolly, T. and Gilmore, G. (1989) Cell **57**: 599-610.

Cover, W.H., Ryan, J.P., Bassford Jr., P.J., Walsh, K.A., Bollinger, J. and Randall, L.L. (1987) J. Bact. **169**: 1794-1800.

Crooke, E., Guthrie, B., Leckers, S., Lill, R. and Wickner, W. (1988) Cell **54**: 1003-1011.

Crooke, E. and Wickner W. (1987) Proc. Natl. Acad. Sci. USA **84**: 5216-5220.

Dalbey, R.E. and Wickner, W. (1985) J. Biol. Chem. **260**: 15925-15931.

Daar, I.O. and Maquat, L.E. (1988) Mol. Cell. Biol. **8**: 802-813.

Deshaies, R.J., Kepes, F. and Böhni, P.C. (1989) T.I.G.S. **5**: 87-93.

Deshaies, R.J., Koch, B.D. and Schekman, R. (1988a) T.I.B.S. **13**: 384-388.

Deshaies, R.J., Koch, B.D., Werner-Washburne, M., Craig, E.A. and Schekman, R. (1988b) Nature **332**: 800-805.

Di Mario, P.J., Warren, T.G. and Mahowald, A.P. (1987) Insect Biochem. **17**: 1187-1192.

Eilers, M., Hwang, S. and Schatz, G. (1988) EMBO J. **7**: 1139-1145.

Emr, S.D., Hanley-Way, S. and Silhavy, T.J. (1981) Cell **23**: 79-88.

Emr, S.D., Hedgpeth, J., Clément, J-M., Silhavy, T.J. and Hofnung, M. (1980) Nature **285**: 82-85.

Emr, S.D., Schauer, I., Hansen, W., Esmon, P. and Schekman, R. (1984) Mol. Cell. Biol. **4**: 2347-2355.

Feinberg and Vogelstein (1983) Anal. Biochem. **132**: 6-13.

Feinberg and Vogelstein (1984) Anal. Biochem. **137**: 266-267.

Fikes, J.D. and Bassford Jr., P.J. (1987) J. Bact. **169**: 2352-2359.

Friederich, E., Baeuerle, P.A., Garoff, H., Hovemann, B. and Huttner, W.B. (1988a) J. Biol. Chem. **263**: 14930-14938.

Friederich, E., Frietz, H.-J. and Huttner, W.B. (1988b) J. Cell. Biol. **107**: 1655-1667.

Gannon, P.M., Li, P., Kumamoto, C.A. (1989) J. Bact. **171**: 813-818.

Gans, M., Audit, C., and Masson, M. (1975) Genetics **81**: 634-704.

Garabedian, M.J., Hung, M.-C. and Wensink, P.C. (1985) Proc. Natl. Acad. Sci. USA **83**: 1396-1400.

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

Garabedian, M.J., Shirras, A.D., Bownes, M. and Wensink, P.C. (1987) Gene **55**: 1-8.

Garcia, P.D., Ghrayeb, J., Inouye, M. and Walter, P. (1987) J. Biol. Chem. **262**: 9463-9468.

Gilmore, R., Blobel, G. and Walter, P. (1982a) J. Cell Biol. **95**: 463-469.

Gilmore, R., Blobel, G. and Walter, P. (1982b) J. Cell Biol. **95**: 470-477.

Gilmore, R. and Blobel, G. (1983) Cell **35**: 677-685.

Gilmore, R. and Blobel, G. (1985) Cell **42**: 497-505.

Gingeras, T.R., Gelti-Douka, H. and Kambysellis, M.P. (1973) Dros. Inf. Serv. **50**: 58.

Giorgi, F. and Jacob, J. (1977) J. Embryol. Exp. Morph. **39**: 45-57.

Gough, J.A. and Murray, N.E. (1983) J. Mol. Biol. **166**: 1-19.

Hagendorn, H.H. (1989) In: Ecdysone. From Chemistry to Mode of Action. (Ed. Koolman, J.), Thieme Medical Publishers Inc., New York, pp. 279-289.

Haguenauer-Tsapis, R. and Hinnen, A. (1984) Mol. Cell. Biol. **4**: 2668-2675.

Hansen, W. and Walter, P. (1988) J. Cell. Biol. **106**: 1075-1087.

Hartmann, E., Wiedmann, M. and Rapoport, T.A. (1989) EMBO J. **8**: 2225-2229.

Hatzopoulos, P. and Kambysellis, M.P. (1987) Mol. Gen. Genet. **206**: 475-484.

Hovemann, B., Galler, R., Walldorf, U., Küpper, N. and Bautz, E.K.F. (1981) Nucl. Acids Res. **9**: 4721-4734.

Hung, M.C. and Wensink, P.C. (1981) Nucl. Acids Res. **9**: 6407-6419.

Hung, M.-C., Barnett, T., Woolford, C. and Wensink, P.C. (1982) J. Mol. Biol. **154**: 581-602.

Hung, M.C. and Wensink, P.C. (1983) J. Mol. Biol. **164**: 481-492.

Inouye, M., and Halegoua, S. (1980) CRC *Crit. Rev. Biochem.* **7**: 339-371.

Inouye, S., Soberon, X., Franceschini, T., Nakamura, K., Itakura, K and Inouye M. (1982) Proc. Natl. Acad. Sci. USA **79**: 3438-3441.

Isaac, P.G. (1982) Molecular Studies of the Vitellogenins of *D. melanogaster.* Ph.D. Thesis, University of Edinburgh, U.K.

Isaac, P.G. and Bownes, M. (1982) Eur. J. Biochem. **123**: 527-534.

Jacob, J. and Stirling, J.L. (1959) Chromosoma **10**: 210-228.

Jowett, T. and Postlethwait, J. (1980) Dev. Biol. **80**: 225-234.

Julius, D., Schekman, R. and Thorner, J. (1984) Cell **36**: 309-318.

Kaiser, C.A. and Botstein, D. (1986) Mol. Cell. Biol. **6**: 2382-2391.

Kaiser, C.A. Preuss, D., Grisafi, P. and Botstein, D. (1987) Science **235**: 312-317.

Karess, R.E. (1985) P. Element Mediated Germ Line Transformation of *Drosophila* in DNA Cloning. (Ed. Glover, D.), IRL Press, Oxford.

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

Kelly, T.J., Adams, T.S., Schwartz, M.B., Birbaum, M.J., Rubenstein, E.C. and Imberski, R.B. (1987) Insect Biochem. **17**: 1089-1093.

King, R.C. (1970) Ovarian Development in *D. melanogaster.* Academic Press, New York.

Klausner, R.D. (1989) Cell **57**: 703-706.

Koreaki, I., Akiyama, Y., Yura, T. and Shiba, K. (1986) J. Bact. **167**: 201-204.

Koshland, D., Sauer, R.T. and Botstein, D. (1982) Cell **30**: 903-914.

Kozma, R. and Bownes, M. (1986a) Mol. Gen. Genet. **204**: 302-309.

Kozma, R. and Bownes, M. (1986b) Insect Biochem. **16**: 263-271

Kraus, K.W., Lee, Y., Lis, J.T. and Wolfner, M.F. (1988) Mol. Cell. Biol. **8**: 4756-4764.

Kuhn, A. and Wickner, W. (1985) J. Biol. Chem. **260**: 15914-15918.

Kunkel, J.G. and Nordin, J.H. (1985) Yolk Proteins. In: Comprehensive Insect Physiology, Biochemistry and Pharmacology, vol. 8. (Ed. Gilbert, L.I. and Kerkut, G.A.), Pergamon Press, New York/Oxford/Paris.

Lauffer, L., Garcia, P.D., Harkins, R.N., Coussens, L., Ullrich, A. and Walter, P. (1985) Nature **318**: 334-338.

Li, P., Beckwith, J. and Inouye, H. (1988) Proc. Natl. Acad. Sci. USA **85**: 7685-7689.

Lin, J.J.C., Kanazawa, H. and Wu, H.C. (1980) J. Bact. **141**: 550-557.

Lindsley, D.L. and Grell, E.N. (1968) Genetic Variations of *D. melanogaster*. Carnegie Institute of Washington, Publication no. 627.

Lindsley, D.L. and Zimm, G. (1985) Dros. Inf. Serv. **62**.

Lingappa, V.R., Chaidez, J., Yost, C.S. and Hedgepeth, J. (1984) Proc. Natl. Acad. Sci. USA **81**: 456-460.

Logan, S.K., Garabedian, M.J. and Wensink, P.C. (1989) Genes Dev. **3**: 1453-1461.

Mahowald, A.P. (1972) J. Morphol. **137**: 29-48.

Mahowald, A.P. and Kambysellis, M.P. (1980) Oogenesis. In: The Genetics and Biology of *Drosophila*, vol. 2d. (Eds. Ashburner, M. and Wrights, T.R.F.), Academic Press, London, pp. 141-224.

Mahowald, A.P. and Strassheim, J.M. (1970) J. Cell. Biol. **45**: 306-320.

Mahowald, A.P. and Tiefert, M. (1970) W. Roux. Arch. **165**: 8-25.

Maniatis, T., Goodbourn, S. and Fischer, J.A. (1987) Science **236**: 1237-1244.

Marcus, C.H. (1985) Dros. Info. Serv. **61**: 193-194.

Mariani, B.D., Lingappa, J.R. and Kafatos, F.C. (1988) Proc. Natl. Acad. Sci. USA 85: 3029-3033.

Martinez-Cruzado, J.C., Swimmer, C., Fenerjian, M.G. and Kafatos, F. (1988) Genetics **119**: 663-677.

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

Meyer, D.I. (1988) T.I.B.S. **13**: 471-474.

Meyer, D.I., Krause, E. and Dobberstein, B. (1982) Nature **297**: 647-650.

Minoo, P. (1982) Molecular Analysis of a Quantitative Yolk Polypeptide Variant in *D. melanogaster*. Ph.D. Thesis, Eugene, Oregon.

Minoo, P. and Postlethwaite, J.H. (1985) Arch. Insect Biochem. Physiol. **2**: 7-27.

Mitchell, P.J. and Tjian, R. (1989) Science **245**: 371-378.

Mitsis, P.G. and Wensink, P.C. (1989a) J. Biol. Chem. **264**: 5188-5194.

Mitsis, P.G. and Wensink, P.C. (1989b) J. Biol. Chem. **264**: 5195-5202.

Mintzas, A.C. and Kambysellis, M.P. (1982) Insect Biochem. **12**: 25-33.

Müller, M. and Blobel, G. (1984) Proc. Natl. Acad. Sci. USA **81**: 7737-7741.

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

Oliver, D. (1985): Ann. Rev. Microbiol. **39**: 615-648.

Palade, G. (1975) Science **189**: 347-358.

Palva, I., Sarvas, M. Lehtovaara, P., Sibakov, M. and Kääriänen, L. (1982) Proc. Natl. Acad. Sci. USA **79**: 5582-5586.

Perera, E. and Lingappa, V.R. (1985) J. Cell Biol. **101**: 2292-2301.

Perlman, D. and Halvorson, H.O. (1983) J. Mol. Biol. **167**: 391-409.

Perlman, D. Raney, P. and Halvorson, H.O. (1986) Proc. Natl. Acad. Sci. USA 83: 5033-5037.

Poritz, M.A., Siegel, V., Hansen, W., and Walter, P. (1988) Proc. Natl. Acad. Sci. USA **85**: 4315-4319.

Postlethwaite, J.H., Bownes, M. and Jowett, T. (1980) Dev. Biol. 79: 379-387.

Postlethwaite, J.G. and Handler, A.M. (1979) J. Insect Physiol. **25**: 455-460.

Postlethwaite, J.H. and Jowett, T. (1980) Cell **20**: 671-678.

Preuss, D. and Botstein, D. (1989) Mol. Cell. Biol. **9**: 1452-1464.

Redfern, C.P.F and Bownes, M. (1982) Mol. Gen. Genet. **185**: 181-183.

Redfern, C.P.F. and Bownes, M. (1983) Mol. Gen. Genet. **189**: 432-440.

Rose, J.K. and Doms, R.W. (1988) Ann. Rev. Cell. Biol. **4**: 257-288.

Ross, J. (1988) Mol. Biol. Med. **5**: 1-4.

Rothman, J.H., Yamashiro, C.T., Kane, P.M. and Stevens, T.H. (1989) TIBS **14**: 347-350.

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

Russel, M. and Model, P. (1981) Proc. Natl. Acad. Sci. USA **78**: 1717-1721.

Ryan, J.P., Duncan, M.C., Bankaitis, V.A. and Bassford Jr., P.J. (1986) J. Biol. Chem. **261**: 3389-3395.

Sanz, P. and Meyer, D.I. (1988) EMBO J. **7**: 3553-3557.

Saunders, R.D.C. and Bownes, M. (1986) Mol. Gen. Genet. **205**: 557-560.

Schauer, I., Emr, S., Gross, C. and Schekman, R. (1985) J. Cell Biol. **100**: 1664-1675.

Schekman, R. (1985) Ann. Rev. Cell. Biol. **1**: 115-143.

Schwartz, M.B., Kelly, T.J., Imberski, R.B. and Rubenstein, E.C. (1985) J. Insect Physiol. **31**: 947-959.

Shelness, G.S., Kanwar, Y.S. and Blobel, G. (1988) J. Biol. Chem. **263**: 17063-17070.

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

Shirras, A.D. and Bownes, M. (1989) Proc. Natl. Acad. Sci. USA 86: 4559-4563.

Siegel, V. and Walter, P. (1985) J. Cell Biol. **100**: 1913-1921.

Siegel, V. and Walter, P. (1986) Nature **320**: 81-84.

Siegel, V. and Walter, P. (1988a) Cell **52**: 39-49.

Siegel, V. and Walter, P. (1988b) Proc. Natl. Acad. Sci. USA **85**: 1801-1805.

Silhavy, T.J., Benson, S.A. and Emr, S.D. (1983) Microbiol. Rev. **47**: 313-344.

Singer, S.J., Maher, P.A. and Yaffe, M.P. (1987) Proc. Natl. Acad. Sci. USA **84**: 1015-1019.

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

Steller, H. and Pirrotta, V. (1985) EMBO J. **4**: 167-171.

Tamura, T., Kuner, T.C. and Postlethwaite, J. (1985) Proc. Natl. Acad. Sci. USA **82**: 7000-7004.

Tata, J.R. (1976) Cell **9**: 1-4.

Thom, J.R. and Randall, L.L. (1988) J. Bact. (1988) **170**: 5654-5661.

Throckmoreton, L.H. (1975) The Phylogeny, Ecology and Geography of *Drosophila*. In: Handbook of Genetics, vol. 3 (Ed. King, R.C.), pp. 421-469.

Trun, N.J. and Silhavy, T.J. (1989) J. Mol. Biol. **205**: 665-676.

Ullu, E., Murphy, S. and Melli, M. (1982) Cell **29**: 195-202.

van Handel, E. and Lea, A.O. (1984) J. Insect Physiol. **30**: 871-875.

Verner, K. and Schatz, G. (1988) Science **241**: 1307-1313.

von Heijne, G. (1981) Eur. J. Biochem. **116**: 419-422.

von Heijne, G. (1982) J. Mol. Biol. **159**: 537-541.

von Heijne, G.(1983a) Eur. J. Biochem. **133**: 17-21.

von Heijne, G. (1984a) EMBO J. **3**: 2315-2318.

von Heijne, G. (1984b) J. Mol. Biol. **173**: 243-251.

von Heijne, G. (1985) J. Mol. Biol. **184**: 99-105.

von Heijne, G. (1986a) J. Mol. Biol. **189**: 239-246.

von Heijne, G. (1986b) J. Mol. Biol. **192**: 287-290.

von Heijne, G. and Abrahamsén, L. (1989) FEBS Letts. **244**: 439-446.

von Heijne, G., Wickner, W., and Dalbey, R.E. (1988) Proc. Natl. Acad. Sci. USA **85**: 3363-3366.

Walter, P. (1987) Nature **328**: 763-764.

Walter, P. and Blobel, G. (1980) Proc. Natl. Acad. Sci. USA **77**: 7112-7116.

Walter, P. and Blobel, G. (1981) J. Cell Biol. **91**: 557-561.

Walter, P. and Blobel, G. (1983) Cell **34**: 525-533.

Walter, P. Gilmore, R. and Blobel, G. (1984) Cell **38**: 5-8.

Walter, P., Ibrahimi, I. and Blobel, G. (1981) J. Cell Biol. **91**: 545-550.

Walter, P. and Lingappa, V.R. (1986) Ann. Rev. Cell. Biol. **2**: 499-516.

Warren, T.G., Brennan, M.D. and Mahowald, A.P. (1979) Proc. Natl. Acad. Sci. USA **76**: 2848-2852.

Warren, T.G. and Mahowald, A.P. (1979) Dev. Biol. **68**: 130-139.

Waters, M.G. and Blobel, G. (1986) J. Cell. Biol. **102**: 1543-1550.

Weng, Q.,, Chen, L. and Tai, P.C. (1988) J. Bact. **170**: 126-131.

Wickner, W. (1980) Science **210**: 861-868.

Wiedmann, M., Kurzchalia, T.V., Bielka, H. and Rapoport, T. (1987) J. Cell Biol. 104: 201-209.

Wiedmann, M., Kurzchalia, T.V., Hartmann, E. and Rapoport, T.A. (1987) Nature 328: 830-833.

Williams, J.L. and Bownes, M. (1986) Eur. J. Biochem. **161**: 95-101.

Williams, J.L., Saunders, R.D.C., Bownes, M. and Scott, A. (1987) Mol. Gen. Genet. **209**: 360-365.

Wilson, C., Pearson, R.K., Bellen, H.J., O'Kane, C.J., Grossniklaus, U. and Gehring, W.J. (1989) Genes Dev. **3**: 1301-1313.

Wojchowski, D.M., Parsons, P., Nordin, J.H. and Kunkel, J.G. (1986) Dev. Biol. **116**: 422-430.

Wu, S., Zhang, J. and Ma, M. (1987) Physiol. Ent. **12**: 355-361.

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

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

Yen, T.J., Machlin, P.S. and Cleveland, D.W. (1988) Nature **334**: 580-585.

Zalokar, M., Audit, C. and Erk, I. (1975) Dev. Biol. **47**: 419-432.

APPENDIX I

The sequence of $yp3^{1526}$ aligned with the published sequence (Garabedian *et al.* 1987; Yan *et al.* (1987).

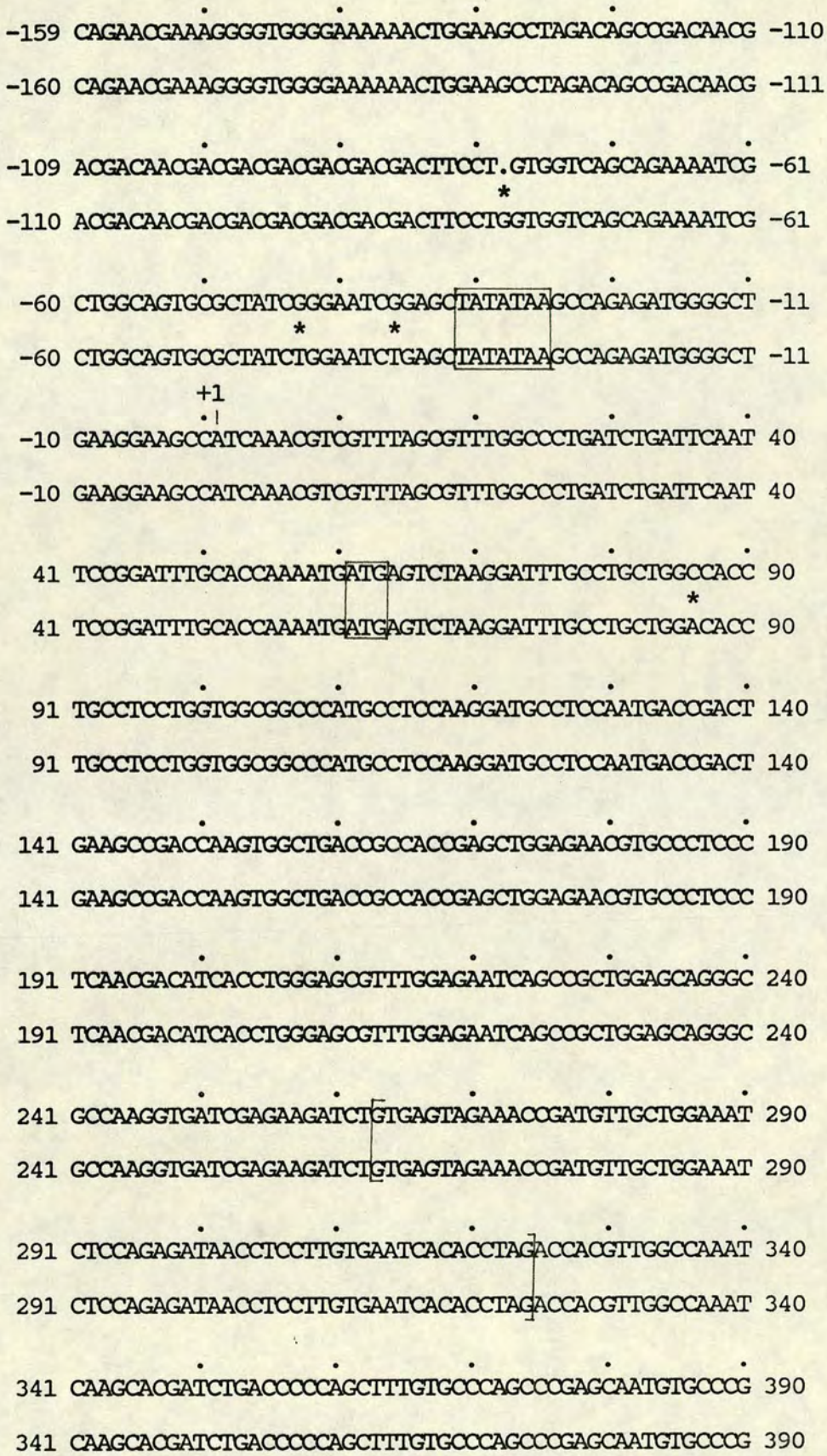
The upper sequence is yp3 and the lower sequence is yp3, mismatches due to sequence variation are represented by asterisks.

The following features are indicated in 5′ to 3′ order:

TATA box (TATATAA), transcript initiation site (cap site) (nt +1), start codon (ATG), introns ([ ]), termination codon (TAG), and two polyA recognition sequences AATAAA.

Numbering of nucleotides is relative to the transcript initiation site.

```
                .         .         .         .         .
-709 GAGCTCATCATTTTAGCAGTTGCTATGCTTTTGCATATATAAATATAATG -660

-710 GAGCTCATCATTTTAGCAGTTGCTATGCTTTTGCATATATAAATATAATG -661

                .         .         .         .         .
-659 CATTCACCTGGCGGCTGGTCATTGATTCCAATTTGGCCGGCTTCCAATCG -610

-660 CATTCACCTGGCGGCTGGTCATTGATTCCAATTTGGCCGGCTTCCAATCG -611

                .         .         .         .         .
-609 CTGGAGGTCAATGCCGGGTCACACCAGTTTCTCACTTGACGCAGGTGTTG -560
                                    *
-610 CTGGAGGTCAATGCCGGGTCACACCAGTTCCTCACTTGACGCAGGTGTTG -561

                .         .         .         .         .
-559 CAAGTTTGTTGCCAGTTCAATTCTAATCAAGGGATCTGCACAAGTTGTTT -510
                *
-560 CAAGTTTGTTGCTAGTTCAATTCTAATCAAGGGATCTGCACAAGTTGTTT -511

                .         .         .         .         .
-509 CAATCAATCCGTACTAGAATACATTTTAAGTGCAGAGAACAAAAATTTGC -460

-510 CAATCAATCCGTACTAGAATACATTTTAAGTGCAGAGAACAAAAATTTGC -461

                .         .         .         .         .
-459 ATTACTTTGGGAATTATATGCATAAATCTGTAAGTGTCGTTAAAACCAAA -410

-460 ATTACTTTGGGAATTATATGCATAAATCTGTAAGTGTCGTTAAAACCAAA -411

                .         .         .         .         .
-409 TGATAGTGATGATACAAATATATCACGATGCAATACTACTAGTGGTCAAC -360

-410 TGATAGTGATGATACAAATATATCACGATGCAATACTACTAGTGGTCAAC -361

                .         .         .         .         .
-359 GATTTTCCAATAATCTAAATCTTAACATTTTATGAATGGATTTTTTTTTG -310
                                    *
-360 GATTTTCCAATAATCTAAATCTTAACACTTTATGAATGGATTTTTTTTTG -311

                .         .         .         .         .
-309 CACATTTTTTGCCAAGTGTGAAGAGGTTCAAAAACCTTAGTGCGATAAGA -260
     *
-310 CCCATTTTTTGCCAAGTGTGAAGAGGTTCAAAAACCTTAGTGCGATAAGA -261

                .         .         .         .         .
-259 GAACTAAATGGTTGGCAAACACACACACATGTGAAATAAATCCGGCTATT -210

-260 GAACTAAATGGTTGGCAAACACACACACATGTGAAATAAATCCGGCTATT -211

                .         .         .         .         .
-209 TGCAATCAATTTTCCCTTGACTTGCACTTTATACACCGGCGACAGATCAG -160

-210 TGCAATCAATTTTCCCTTGACTTGCACTTTATACACCGGCGACAGATCAG -161
```


-159 CAGAACGAAAGGGGTGGGGAAAAAACTGGAAGCCTAGACAGCCGACAACG -110
-160 CAGAACGAAAGGGGTGGGGAAAAAACTGGAAGCCTAGACAGCCGACAACG -111
-109 ACGACAACGACGACGACGACGACGACTTCCT.GTGGTCAGCAGAAAATCG -61
-110 ACGACAACGACGACGACGACGACGACTTCCTGGTGGTCAGCAGAAAATCG -61
-60 CTGGCAGTGCGCTATCGGGAATCGGAGCTATATAAGCCAGAGATGGGGCT -11
-60 CTGGCAGTGCGCTATCTGGAATCTGAGCTATATAAGCCAGAGATGGGGCT -11
+1
-10 GAAGGAAGCCATCAAACGTCGTTTAGCGTTTGGCCCTGATCTGATTCAAT 40
-10 GAAGGAAGCCATCAAACGTCGTTTAGCGTTTGGCCCTGATCTGATTCAAT 40
41 TCCGGATTTGCACCAAAATGATGAGTCTAAGGATTTGCCTGCTGGCCACC 90
41 TCCGGATTTGCACCAAAATGATGAGTCTAAGGATTTGCCTGCTGGACACC 90
91 TGCCTCCTGGTGGCGGCCCATGCCTCCAAGGATGCCTCCAATGACCGACT 140
91 TGCCTCCTGGTGGCGGCCCATGCCTCCAAGGATGCCTCCAATGACCGACT 140
141 GAAGCCGACCAAGTGGCTGACCGCCACCGAGCTGGAGAACGTGCCCTCCC 190
141 GAAGCCGACCAAGTGGCTGACCGCCACCGAGCTGGAGAACGTGCCCTCCC 190
191 TCAACGACATCACCTGGGAGCGTTTGGAGAATCAGCCGCTGGAGCAGGGC 240
191 TCAACGACATCACCTGGGAGCGTTTGGAGAATCAGCCGCTGGAGCAGGGC 240
241 GCCAAGGTGATCGAGAAGATCTGTGAGTAGAAACCGATGTTGCTGGAAAT 290
241 GCCAAGGTGATCGAGAAGATCTGTGAGTAGAAACCGATGTTGCTGGAAAT 290
291 CTCCAGAGATAACCTCCTTGTGAATCACACCTAGACCACGTTGGCCAAAT 340
291 CTCCAGAGATAACCTCCTTGTGAATCACACCTAGACCACGTTGGCCAAAT 340
341 CAAGCACGATCTGACCCCCAGCTTTGTGCCCAGCCCGAGCAATGTGCCCG 390
341 CAAGCACGATCTGACCCCCAGCTTTGTGCCCAGCCCGAGCAATGTGCCCG 390

391 TCTGGATTATCAAGTCCAATGGACAGAAGGTTGAGTGCAAGTTGAACAAC 440

391 TCTGGATTATCAAGTCCAATGGACAGAAGGTTGAGTGCAAGTTGAACAAC 440

441 TATGTGGAGACGGCCAAGGCACAGCCCGGATTCGGCGAGGATGAGGTCAC 490

441 TATGTGGAGACGGCCAAGGCACAGCCCGGATTCGGCGAGGATGAGGTCAC 490

491 CATTGTCCTGACTGGTCTGCCCAAGACCAGCCCCGCTCAGCAGAAGGCCA 540

491 CATTGTCCTGACTGGTCTGCCCAAGACCAGCCCCGCTCAGCAGAAGGCCA 540

541 TGCGCAGGTTGATCCAGGCCTACGTCCAGAAGTACAACCTCCAGCAGCTG 590

*

541 TGCGCAGGTTGATCCAGGCCTACGTTCAGAAGTACAACCTCCAGCAGCTG 590

591 CAGAAGAACGCCCAGGAGCAGCAGCAGCAGCTCAAGAGCAGCGACTACGA 640

591 CAGAAGAACGCCCAGGAGCAGCAGCAGCAGCTCAAGAGCAGCGACTACGA 640

641 CTACACCAGCAGCGAGGAGGCCGCTGACCAATGGAAATCCGCCAAGGCTG 690

641 CTACACCAGCAGCGAGGAGGCCGCTGACCAATGGAAATCCGCCAAGGCTG 690

691 CCAGCGGCGATTTGATCGTAAGTTGGTCGCATTCCTATATTTCATAATTA 740

691 CCAGCGGCGATTTGATCGTAAGTTGGTCGCATTCCTATATTTCATAATTA 740

741 AACGTGTACATATGGATATTTATGAAATTCAAATTGCAGATCATTGACCT 790

741 AACGTGTACATATGGATATTTATGAAATTCAAATTGCAGATCATTGACCT 790

791 CGGCTCCACCCTGACCAACTTCAAACGCTACGCGATGCTGGATGTTCTGA 840

791 CGGCTCCACCCTGACCAACTTCAAACGCTACGCGATGCTGGATGTTCTGA 840

841 ACACCGGCGCCATGATCGGCCAGACCCTGATCGATCTGACCAACAAGGGT 890

841 ACACCGGCGCCATGATCGGCCAGACCCTGATCGATCTGACCAACAAGGGT 890

891 GTGCCCCAGGAGATCATCCATCTGATCGGCCAGGGAATCAGCGCCCATGT 940

891 GTGCCCCAGGAGATCATCCATCTGATCGGCCAGGGAATCAGCGCCCATGT 940

941 GGCCGGAGCTGCTGGCAACAAGTACACCGCCCAAACCGGACACAAGCTGC 990

941 GGCCGGAGCTGCTGGCAACAAGTACACCGCCCAAACCGGACACAAGCTGC 990

991 GCCGCATCACCGGTCTGGATCCCGCCAAGGTGCTGTCCAAGCGTCCCCAG 1040

991 GCCGCATCACCGGTCTGGATCCCGCCAAGGTGCTGTCCAAGCGTCCCCAG 1040

1041 ATCCTGGGTGGTCTGTCCCGCGGCGATGCTGACTTCGTTGATGCCATTCA 1090

1041 ATCCTGGGTGGTCTGTCCCGCGGCGATGCTGACTTCGTTGATGCCATTCA 1090

1091 CACATCGACCTTCGCCATGGGCACGCCCATCCGTTGCGGCGATGTTGACT 1140

1091 CACATCGACCTTCGCCATGGGCACGCCCATCCGTTGCGGCGATGTTGACT 1140

1141 TCTACCCCAACGGACCGTCCACCGGTGTTCCCGGCTCCGAGAATGTGATC 1190

1141 TCTACCCCAACGGACCGTCCACCGGTGTTCCCGGCTCCGAGAATGTGATC 1190

1191 GAGGCTGTGGCCCGTGCCACCCGTTACTTTGCCGAGTCTGTGCGTCCCGG 1240

1191 GAGGCTGTGGCCCGTGCCACCCGTTACTTTGCCGAGTCTGTGCGTCCCGG 1240

1241 TAGCGAGCGCAATTTCCCCGCCGTTCCGGCCAACTCGCTGAAGCAGTACA 1290

1241 TAGCGAGCGCAATTTCCCCGCCGTTCCGGCCAACTCGCTGAAGCAGTACA 1290

1291 AGGAGCAGGATGGCTTTGGCAAGCGCGCCTACATGGGTCTCCAGATCGAC 1340

1291 AGGAGCAGGATGGCTTTGGCAAGCGCGCCTACATGGGTCTCCAGATCGAC 1340

1341 TACGATCTGCGCGGTGACTACATCTTGGAGGTCAACGCCAAGAGCCCCTT 1390

1341 TACGATCTGCGCGGTGACTACATCTTGGAGGTCAACGCCAAGAGCCCCTT 1390

1391 CGGTCAGCGCAGCCCTGCCCACAAGCAGGCCGCCTACCATGGCATGCACC 1440

1391 CGGTCAGCGCAGCCCTGCCCACAAGCAGGCCGCCTACCATGGCATGCACC 1440

1441 ACGCCCAGAACTAGAGCGCCCATGGCCACGCCCCCTGGTTACCAGGGACG 1490

1441 ACGCCCAGAACTAGAGCGCCCATGGCCACGCCCCCTGGTTACCAGGGACG 1490

1491 TTCGATCGTCACGCACTTTCTGATAATCAGAAAATAAAAACCCGGAATGC 1540

1491 TTCGATCGTCACGCACTTTCTGATAATCAGAAAATAAAAACCCGGAATGC 1540

1541 GTAGTTTAGCTTAGAAGTTTCATCAAACAATCAAAAAAGAAAAATCTATA 1590

1541 GTAGTTTAGCTTAGAAGTTTCATCAAACAATCAAAAAAGAAAAATCTATA 1590

1591 AAATCCCATAAAAATAAAAGCTGCAAATTTTCGAAAAGTCAAGTTTTTTA 1640

1591 AAATCCCATAAAAATAAAAGCTGCAAATTTTCGAAAAGTCAAGTTTTTTA 1640

1641 ATAGCAATAGCAATGGTTATTCTGGATTGGATTCTAACTTTTATGGTATT 1690

1641 ATAGCAATAGCAATGGTTATTCTGGATTGGATTCTAACTTTTATGGTATT 1690

1691 AAAAAACACACACAAGAATTTGCTGGGCACATTTTTAGGCACCCCTTCTG 1740

1691 AAAAAACACACACAAGAATTTGCTGGGCACATTTTTAGGCACCCCTTCTG 1740

APPENDIX II

One-letter symbols and three-letter abbreviations for amino acids.

A	Ala	Alanine
B	Asx	Aspargine or Aspartic Acid
C	Cys	Cystine
D	Asp	Aspartic Acid
E	Glu	Glutamic Acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Aspargine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine
Z	Glx	Glutamine or Glutamic Acid