MOLECULAR AND BIOLOGICAL ANALYSIS OF THE PUTATIVE GOLGI MANNOSIDASE II AND OTHER ENZYMES INVOLVED IN *N*-LINKED GLYCOSYLATION IN *DROSOPHILA MELANOGASTER.*

ANNE LOCKYER
LADY MARGARET HALL, OXFORD

Thesis submitted for the Degree of Doctor of Philosophy

University of Oxford

Genetics Laboratory

Department of Biochemistry

Oxford

RADCLIFFE SCIENCE LIBRARY OXFORD
DEPOSITED
10 OCT 1996
THESIS

December, 1995

ABSTRACT

Glycosylation is a fundamental post-translational modification of proteins that occurs in all eukaryotes, but the biological significance of specific glycosylation structures is still largely unknown. We have chosen to study this process by cloning and characterizing the genes involved in the *N*-linked glycosylation pathway in *Drosophila melanogaster* and analysing mutants of these genes Since this biochemical pathway for the production of *N*-linked glycans is considerably conserved, studying it in *Drosophila*, a useful organism for genetic research, should yield results applicable to other eukaryotes.

This thesis describes the attempt to identify a *Drosophila* homologue of rat ER mannosidase and how this investigation resulted in the identification of a *Drosophila* cDNA with considerable homology to calnexin, an enzyme involved in the folding of *N*-glycosylated proteins. It also describes the cloning and characterization of the genomic sequence for the putative Golgi mannosidase II (*GmII*)* in *Drosophila*. The expression and control of expression of *GmII* have been investigated further, and have suggested expression of this gene throughout *Drosophila* development. Analysis of the *GmII* promoter region has shown expression from a TATA-less promoter contained within 288bp of sequence 5' to the start of transcription.

---

* Where-ever *GmII* is referred to throughout this thesis it should be noted that *GmII* is the *Drosophila* structural homologue of Golgi mannosidase II. Functional homology for this enzyme is still to be demonstrated.

ACKNOWLEDGEMENTS

I would like to thank Dr D. B. Roberts, my supervisor, for suggesting this project and for his continued help and encouragement during the three years I worked with him. I would also like to thank the other members of the *Drosophila* laboratory: Dr Jeremy Foster, for giving me Golgi mannosidase II, Dr Karen Carpenter, Dr Stefanos Christodoulou, Dr Graeme Standen and Ben Yudkin for their help and advice and Mrs Wendy Mulvany and Mrs Jean Matthews for ensuring the smooth running of the laboratory. I also acknowledge the help of other members of the Genetics Laboratory, in particular Mr David Hunter and Miss Gillian Oldfield.

I acknowledge the support of the SERC/BBSRC for funding me for three years and for Lady Margaret Hall for providing funding for conference expenses.

I would like to thank my family for their support during my years as a student - my mother and father for their help and Jane and Nigel for housing me during the last few months. I would also like to thank John Sinclair for the use of his Apple Macintosh PowerBook.

Thanks are also due to my friends at LMH - the Greeblie network, for making my time in Oxford such an entertaining one.

ABBREVIATIONS

A	adenine
AMPS	ammonium persulphate
bp	base pair
BSA	bovine serum albumin
C	cytosine
cDNA	complementary DNA
DAB	3,3'-diaminobenzidine
DNA	deoxyribonucleic acid
DTT	dithiothreitol
ER	endoplasmic reticulum
Fuc	fucose
G	guanine
g	gram(s)
Gal	galactose
GCG	Genetics computer group, University of Wisconsin
Glc	glucose
GlcNAc	N-acetylglucosamine
GmII	*Drosophila* homologue of Golgi mannosidase II
hr	hour(s)
HRP	horse radish peroxidase
IAA	isoamyl alcohol
IPTG	isopropylthiogalactoside
kb	kilobase pair(s)
l	litre(s)
lacZ	*E. coli* β-galactosidase gene

M	molar (moles/litre)
Man	mannose
mg	milligram(s)
ml	millilitre(s)
mM	millimolar
mRNA	messenger RNA
ng	nanogram(s)
OD	optical density
PBS	phosphate buffered saline
PCR	polymerase chain reaction
RE	restriction enzyme
RNA	ribonucleic acid
SDS	sodium dodecyl sulphate
SSC	standard saline citrate
T	thymine
UV	ultra violet
µg	microgram(s)
µl	microlitre(s)
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside

TABLE OF CONTENTS

CHAPTER ONE

GENERAL INTRODUCTION 1

1.1 Glycosylation 1

1.1.1 *O*-linked glycosylation 1

1.1.2 *N*-linked glycosylation 2

1.1.3 Glycoforms 2

1.2 The biological roles of glycosylation 4

1.2.2 Physical roles of protein glycosylation 5

1.2.3 Cell/protein recognition 6

1.2.4 Glycosylation and neural development 6

1.2.5 Disease and glycosylation 7

1.2.6 Viruses and glycosylation 8

1.2.7 General conclusions about glycosylation 9

1.3 The *N*-linked glycosylation pathway 10

1.4 Enzymes in the *N*-linked glycosylation trimming pathway 10

1.4.1 Early ER trimming 11

1.4.2 Class I mannosidases 12

1.4.3 Class II mannosidases 12

1.4.4 Golgi mannosidase II 13

1.4.5 Complex oligosaccharides 13

1.5 *N*-linked glycosylation in *Drosophila* 14

1.5.1 *Drosophila* as a model organism 15

1.5.2 Glycosylation in insects 16

1.6 Cloning *N*-linked glycosylation genes in *Drosophila* 18

1.7 Analysis of *N*-linked glycosylation genes 19

1.7.1 Saturation mutagenesis 19

1.7.2 Disruption of normal expression 19

1.7.3 Investigation of normal expression 20

1.8 Research presented here 20

CHAPTER TWO

MATERIALS AND METHODS 22

2.1 Solutions, buffers and media 22

2.2 Extraction and purification of plasmid DNA 23

2.2.1 Alkaline lysis plasmid mini-prep 23
2.2.2 Smaller scale plasmid DNA purification 24
2.2.3 P1 clone DNA purification 25
2.2.4 Qiagen Midi prep 25
2.3 DNA extraction from flies 26
2.4 Digestion of DNA by restriction endonucleases 27
2.4.1 Plasmid digests 27
2.4.2 Genomic digests 28
2.5 Gel electrophoresis 28
2.5.1 DNA markers (sizes in bp) 28
2.5.2 Standard gel electrophoresis 28
2.6 DNA purification from agarose gels 28
2.6.1 Use of geneclean IIR (BIO 101 inc) 28
2.6.2 Using Low Melting Point (LMP) agarose 29
2.7 Southern blotting and radioactive hybridization 30
2.7.1 Southern blotting 30
2.7.2 Radioactive labelling of probes 30
2.7.3 Spin Dialysis 31
2.7.4 Hybridization 32
2.7.5 Autoradiography 32
2.7.6 Reuse of filters by stripping 33
2.8 Radioactive end-labelling of DNA 33
2.8.1 λ ladder 33
2.8.2 "Genecleaned" inserts 34
2.9 *In situ* hybridization 35
2.9.1 Chromosome Squashes 35
2.9.2 Biotin Nick Translation 35
2.9.3 Pretreatment of chromosome slides 36
2.9.4 Hybridization 36
2.9.5 Peroxidase Staining 37
2.10 Sequencing 38
2.10.1 Preparation of oligonucleotides for use as primers 38
2.10.2 Preparation of template DNA by lithium chloride precipitation to remove RNA 39
2.10.3 Sequencing Reaction 39
2.10.4 Preparation of gel 41

2.10.5 Sequencing gel run 41
2.11 Sub-cloning 42
2.11.1 Preparation of competent cells 42
2.11.2 Ligation 43
2.11.3 Cell transformation by electroporation 44
2.12 Screening a lambda phage library 44
2.12.1 Preparation of bacteria 44
2.12.2 Phage Titre 45
2.12.3 Library plate 45
2.12.4 Repeat screening procedures to check identifications 46
2.13 DNA extraction from phage 47
2.14 RNA extraction from flies 48
2.14.1 Total RNA extraction 48
2.14.2 mRNA extraction. Dynal - DynabeadsR mRNA direct 49
2.14.3 mRNA extraction - Pharmacia Biotech QuickprepR Micro mRNA purification kit 50
2.15 RNA gel and northern blotting 51
2.16 First strand cDNA synthesis from mRNA 51
2.17 Polymerase Chain Reaction 52
2.17.1 PCR primer design 52
2.17.2 Basic reaction mix 52
2.17.3 PCR programs 53
2.17.4 PCR electrophoresis gel 54
2.17.5 Contamination problems 54
2.18 Microinjection of *Drosophila* embryos 54
2.18.1 Apple agar plates 54
2.18.2 Microinjection 55
2.19 Fly culture 56
2.20 Staining for β-galactosidase activity 56
2.20.1 Embryos 56
2.20.2 Imaginal discs 57

CHAPTER THREE
ENDOPLASMIC RETICULUM MANNOSIDASE 58
3.1 Introduction 58
3.2 Identifying *Drosophila* cDNA clones for ER mannosidase 59

3.2.1 Library screening 59
3.2.2 Initial investigation of the cDNA clones 59
3.3 Restriction enzyme analysis 60
3.3.1 Clones to RE map 60
3.3.2 Restriction digests of cDNAs 60
3.3.3 End labelling inserts 60
3.3.4 Restriction enzyme maps 61
3.4 Selection of cDNA 4 for sequencing 61
3.4.1 Sub-cloning 61
3.4.2 *In situ* hybridization 62
3.5 Sequence Analysis 62
3.5.1 Refined RE map of cDNA 4 62
3.5.2 cDNA 4 sequence compared to rat ER mannosidase 62
3.5.3 FASTA database search (1993, 1995) 63
3.5.6 BLAST database search (1995) 63
3.6 Is cDNA 4 the *Drosophila* homologue of Calnexin? 64
3.6.1 Calnexin 64
3.6.2 Calnexin in other species 66
3.6.3 cDNA 4 compared to calnexin in other species 66
3.6.4 A calreticulin homologue in *D. melanogaster* 66
3.7 Discussion and conclusions 67
3.7.1 ER mannosidase 67
3.7.2 Calnexin 68
3.7.3 Further work 70

CHAPTER FOUR
GOLGI MANNOSIDASE II : GENE STRUCTURE 71
4.1 Aims and introduction 71
4.2 Identifying the 5' end of *GmII* 72
4.3 *In situ* hybridization 72
4.4 *GmII* genomic clone 73
4.4.1 P1 Clones 73
4.4.2 Genomic sequence 74
4.5 *GmII* intron analysis 75
4.5.1 Intron splicing 75
4.5.2 Intron consensus sequences 76

4.5.3 Intron base pair composition 77
4.5.4 Nested genes 77
4.5.5 Intron size 78
4.5.6 Intron position 78
4.6 Sequences 5' and 3' to the *GmII* transcribed sequence 79
4.6.1 Discrepancy between phage and P1 5' sequences 79
4.6.2 Analysis of 5' region 80
4.6.3 *GmII* 3' sequence 80
4.7 Discrepancies between *GmII* cDNA and genomic DNA 81
4.8 Spatial and temporal expression of *GmII* 83
4.8.1 Northern blotting 84
4.8.2 Analysis of first strand cDNA by PCR 84
4.8.3 Where and when is *GmII* expressed? 84
4.8.4 Is *GmII* differentially spliced? 85
4.9 Discussion and conclusions 85

CHAPTER FIVE
GOLGI MANNOSIDASE II : FLY TRANSFORMATION 87
5.1 Aims and introduction 87
5.2 P-element mediated transformation of flies 87
5.2.1 Theory 87
5.2.2 Vectors 88
5.2.3 Helper plasmids 89
5.3 *GmII* construct for injection 89
5.3.1 3' end modified 89
5.3.2 The 5' end of the construct 90
5.4 Transformation of flies with *GmII* 91
5.4.1 Microinjection 91
5.4.2 Transformation efficiency 92
5.5 Localization of inserts 93
5.5.1 Sex chromosome insertion 93
5.5.2 Autosomal insertions 93
5.5.3 Summary of stocks 94
5.5.4 *In situ* hybridization 94
5.5.5 Position effects and dosage compensation 95
5.6 Test for expression 96

5.7 Comparison between two transformed stocks and *yw* flies 97
5.8 Rescue of *GmII* mutants 97
5.9 Discussion and conclusions 97

CHAPTER SIX
GOLGI MANNOSIDASE II : PROMOTER ANALYSIS 99
6.1 Introduction 99
6.2 Analysis of *GmII* promoter sequence 99
6.2.1 TATA box 99
6.2.2 Initiator sequence 100
6.2.3 Upstream elements 101
6.2.4 Downstream elements 102
6.3 Transforming flies with the *GmII* promoter 102
6.3.1 pCaSpeR - AUG - β-gal 103
6.3.2 *GmII* promoter fragment by PCR 103
6.4 Transformation 104
6.4.1 Transformed stocks carrying the *lacZ* constructs 104
6.4.2 *In situ* hybridization 105
6.5 β-galactosidase expression from 700bp promoter 105
6.5.1 Embryos 105
6.5.2 Imaginal discs 106
6.6 β-galactosidase expression of deleted constructs 107
6.7 Discussion and conclusions 107
6.7.1 Expression of *GmII* 107
6.7.2 *GmII* promoter 108
6.7.3 Further work 108

CHAPTER SEVEN
GENERAL DISCUSSION 109
7.1 Calnexin 109
7.2 Golgi mannosidase II 109
7.3 Future work 110

APPENDIX I : PRIMER CATALOGUE 112
APPENDIX II : INBRED STOCKS 117
APPENDIX III : DNA AND AMINO ACID CODES 118
REFERENCES 119

CHAPTER ONE
GENERAL INTRODUCTION

1.1 Glycosylation

The process of adding sugar to proteins and lipids to produce glycoproteins, proteoglycans and glycolipids is known as glycosylation. This form of post-translational modification occurs in the endoplasmic reticulum (ER) and Golgi body. The structures of added sugars have been elucidated, but the biological function of these highly specific complex structures remains largely unknown. They have, however, been shown to be significant in many aspects of the biology of organisms, for example cell-cell interaction, protein targeting and recognition processes (for review see Varki, 1993).

There are two major different types of protein glycosylation, *O*-linked and *N*-linked that differ primarily in the core structures while peripheral sequences exhibit similarities, especially in the larger sugar chains (for review see Paulson [1989]). The attached sugars are highly branched, structurally complex molecules possessing secondary and tertiary structure, able to undergo precise conformational changes [Homans *et al.* 1987; Balaji *et al.* 1994]. As such they have a considerable effect, not only on the physical properties of the glycoprotein [Olden *et al.* 1982] but also influence its bioactivity, bio-distribution, immunogenicity and circulatory life-time [Ashwell *et al.* 1982; Rademacher *et al.* 1988a].

1.1.1 *O*-linked glycosylation

O-linked glycosylation joins sugars by glycosidic bonds to hydroxylated side chains of serine, threonine, lysine and proline. Five core structures are found, defined by the sugars directly attached to the initial GalNAc residue [Kornfeld *et al.* 1985]. In contrast to *N*-linked glycosylation (see section 1.1.2) the oligosaccharides are formed

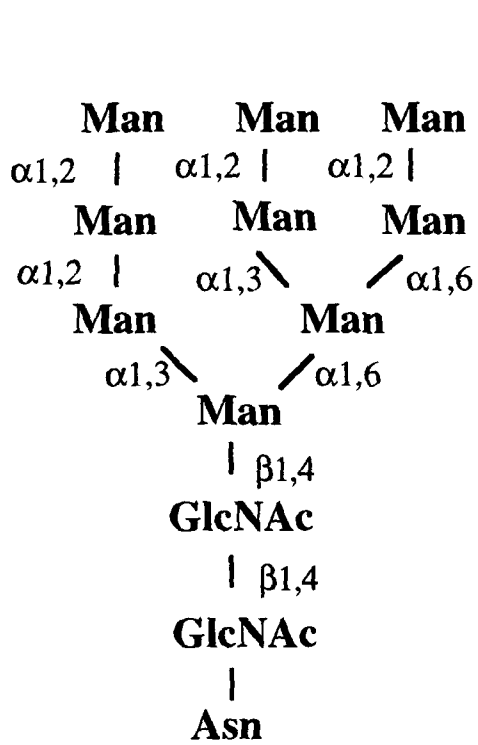
by the direct addition of the sugars to the protein and are simpler and less branched. Such glycoproteins are often secreted for lubricant or protection purposes, such as saliva, bronchial and intestinal secretions [Sheckman *et al.* 1982; Kornfeld *et al.* 1985; Tanner *et al.* 1987].

1.1.2 *N*-linked glycosylation

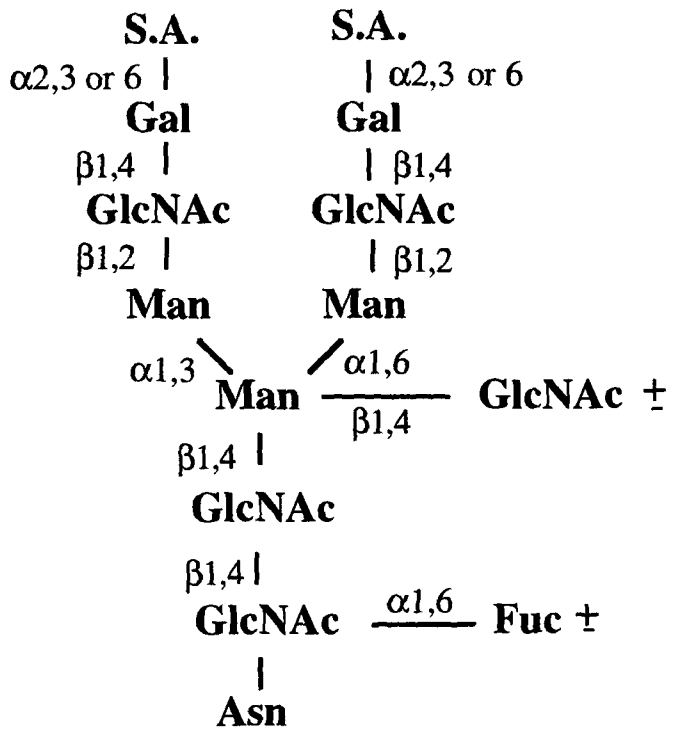
N-linked glycosylation involves the addition of a common branched oligosaccharide to the asparagine residue of a polypeptide through an *N*-glycosidic bond [Wagh *et al.* 1981; Montreuil 1982]. The oligosaccharide structures found in mammals are complex, hybrid and high mannose types of oligosaccharide (Fig. 1.1) [Kornfeld *et al.* 1985]. These contain a common chitobiose structure of -$(GlcNAc)_2$ with three attached mannoses, but have different sugar chains attached to this core structure. The high mannose structures have the chitobiose core and many mannose residues. Complex structures contain the chitobiose core, three mannose residues, further *N*-acetylglucosamines, galactose, sialic acid and sometimes fucose residues (Fig. 1.2). Hybrid structures show features from both the complex and high mannose structures. The extent of trimming and modification of *N*-glycans is determined by a) the polypeptide backbone which affects the access of glycosyltransferases by i) steric hindrance and ii) restraining oligosaccharides in particular formations that distort them sufficiently to prevent recognition by the enzyme [Rademacher *et al.* 1988a]; b) the expression of processing enzymes such as the glycosyltransferases which determine the glycosylation capacity of the tissue [Rademacher *et al.* 1988a] and c) competition for substrates [Marth 1994].

1.1.3 Glycoforms

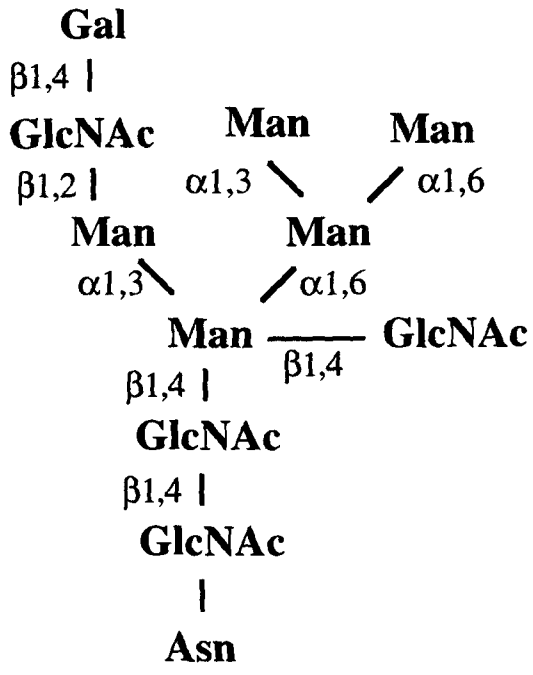
Many different forms of a glycosylated protein will be present in the organism with the same polypeptide chain, but different sugar moieties. These different



a) High mannose



b) Complex



c) Hybrid

Figure 1.1. Structures in *N*-linked glycosylation. Man - mannose, Gal - galactose, S.A. - Sialic acid, Asn - asparagine. (After Kornfeld, 1985).

glycoproteins are known as glycoforms [Rademacher *et al.* 1988a]. They are defined by the following criteria. i) Different glycoproteins from the same cell may have different oligosaccharide structures. ii) Individual glycoproteins may have several different structures at several different sites and the glycosylation of each may be non-independent. iii) The glycoprotein may have different oligosaccharide structures at the same site (known as site heterogeneity). The pattern of glycosylation is reproducible and not random and it is cell type dependent, for example the rat T-cell marker (Thy-1) in thymocytes and brain, has the same polypeptide sequence but different attached oligosaccharides [Parekh *et al.* 1987]. Since the glycosylation contributes to or is an intrinsic part of the physiological activity of the glycoprotein, different glycoforms produce a range of biological activity. The activity of any individual glycoprotein is a composite of its glycoform activities.

That the range of glycoproteins produced is dependent on the cell or tissue type has implications for the *in vitro* expression of cloned genes coding for proteins which are normally glycosylated. These "recombinant glycoproteins" [Parekh *et al.* 1989a] may not be produced with the same spectrum of glycoforms when synthesized in non-native tissues and this may affect the function of the protein. Recombinant glycoproteins for therapeutic use (e.g. erythropoietin, tissue plasminogen activator (tPA)) do not show the native range of glycoforms when produced in foreign cell lines [Takeuchi *et al.* 1988; Tsuda *et al.* 1988; Parekh *et al.* 1989b], which may affect their biological activity. It has been suggested that in cell culture systems the culture environment also affects glycosylation [Goochee *et al.* 1990]. In a controlled way, changing the glycosylation of a protein or "designer glycosylation" [Rademacher 1992] could be one method of producing new glycoforms, to meet specific needs by increasing the half-life or functional spectrum of a glycoprotein or targeting specific tissues.

1.2 The biological roles of glycosylation

There have been many suggestions as to the significance of glycosylation (for reviews see [Rademacher *et al.* 1988a; Varki 1993; Lis *et al.* 1993]). The recent review by Varki [1993] provides an extensive survey of over 1000 references and discusses the biological functions of glycosylation. Varki concludes that i) the biological roles of oligosaccharides have been shown to vary from trivial in some instances, to crucial for development, growth, function and survival of an organism in other cases; ii) it is difficult to predict whether the oligosaccharide is significant for any given glycoprotein, or what function it may have; iii) the same oligosaccharide sequence may have different functions in different locations or during different developmental stages and iv) specific biological roles may be carried out by unusual oligosaccharide sequences. These unusual structures are also more likely to become targets for pathogenic organisms, since they provide specific recognition sites, and are therefore subject to more variation from evolutionary pressures.

The complexity of carbohydrate structures suggests some highly significant function. The cellular and extracellular enzymes involved in glycosylation as well as the energy requirements for processing would be extremely inefficient if the carbohydrate structures were redundant [Fredieu *et al.* 1994]. It has also been estimated that at least 1% of *Drosophila* genes are involved in coding for the enzymes required for glycosylation processes [Williams *et al.* 1991]. This figure is based on an investigation into the *N*-linked glycosylation structures in *Drosophila* which showed these to consist mostly of the oligomannose series (discussed further in section 1.5.2) [Williams *et al.* 1991]. A minimum of about 25 genes code for the enzymes required to synthesize these structures and a similar number may be needed in both *O*-linked and lipid glycosylation. Taken together, and estimating the size of the *Drosophila* genome to be approximately 10,000 genes, this means that 1% of the *Drosophila* genome codes for enzymes involved in glycosylation.

Due to the variability and complexity of the roles of glycosylation in biological events I have chosen to discuss a few examples that demonstrate this variety and why there is much current interest in this subject [Geisow 1991]. I wish to concentrate particularly on the functions of oligosaccharides attached to proteins. For a broader review see [Varki 1993]

1.2.1 Proteoglycans

Polysaccharides and proteoglycans are involved in i) the construction of barriers for limiting cell migration [Wight *et al.* 1992], ii) filtering and retaining biopolymers and metabolites, as basement membranes and the glycocalyx [Ruoslahti 1988; Kjellen *et al.* 1991], and iii) in building bone cartilage and other structural tissues [Hook *et al.* 1984; Hardingham *et al.* 1992]. Proteoglycans carry one or more glycosaminoglycan chains [Ruoslahti 1989].

1.2.2 Physical roles of protein glycosylation

Oligosaccharides may have a purely structural role, as an aid in the folding and conformation of glycoproteins (e.g. Riederer *et al.* 1991). They may also provide information as to the folding *status* of glycoproteins [Hammond *et al.* 1994]. The attached sugars affect the physical properties of the protein such as its stability, rate of proteolysis and solubility and therefore affect the half-life of proteins and cells [Rademacher *et al.* 1988a; Parekh *et al.* 1989a]. The associated oligosaccharides make up a considerable proportion of the glycoprotein, often on the surface. For example rat Thy-1 has almost a shell of surrounding carbohydrate [Parekh *et al.* 1987], leading to the suggestion that the original purpose of the attached sugars to some proteins was as a flexible protective coating that has now been modified for other purposes. However these roles for glycosylation do not explain the huge diversity of structures found attached to glycoproteins.

1.2.3 Cell/protein recognition

Where oligosaccharides have been found to be significant their role is often in mediating specific recognition events such as protein targeting (e.g. lysozymal enzymes [Kornfeld 1987]) or modulating biological processes, for example cell-matrix or cell-cell interactions (e.g. selectins in the immune system [Lasky 1992] or sperm-egg binding [Miller *et al.* 1992]). They may be responsible for providing cellular identity as any individual cell will present a surface of simple, complex and hybrid oligosaccharides attached to membrane spanning proteins or glycolipids, which provide ligands for complementary protein receptors [Lis *et al.* 1993].

Cell-cell interactions are particularly important during development since cell identity and co-ordination is required to organize structures. Changes in glycosylation patterns during development, identified by monoclonal antibody recognition of cell surface carbohydrate antigens [Feizi 1985], suggest carbohydrate involvement in these processes.

Perturbing the production of complex oligosaccharides has also been demonstrated to be lethal during late embryogenesis in mice [Ioffe *et al.* 1994; Metzler *et al.* 1994] (see section 1.4.5) and altering the terminal sialic acids also leads to developmental abnormalities in the early embryo [Varki *et al.* 1991].

1.2.4 Glycosylation and neural development

Cell-cell recognition is also likely to be particularly crucial in neural development. Neural glycoproteins have been identified carrying galactose or GalNAc in growth cones and neurites of invertebrates (references in Fredieu *et al.* 1994), including neurones and axons in the eye discs of *Drosophila* [Fristrom *et al.* 1982]. Antibodies to horseradish peroxidase (HRP) recognize a carbohydrate component of the insect nervous system [Jan *et al.* 1982; Snow *et al.* 1987] which has been demonstrated to be the fucosylated chitobiose core [Kurosaka *et al.* 1991].

Glycosylated proteins that express this epitope are also found unglycosylated, showing that the carbohydrate is not present for protein function and stability at least in this instance [Snow *et al.* 1987]. These observations add weight to the suggestion that glycosylation could be highly significant in axonal guidance and the development of the nervous system [Katz *et al.* 1988]. *Drosophila* mutants with abnormal anti-HRP staining and therefore presumably with masked or altered carbohydrate epitopes have been identified. One of these is *neurally altered carbohydrate* (*nac*) which leads to no anti-HRP staining in the imaginal tissue or in the neural tissue of the adult fly [Katz *et al.* 1988]. Despite this, these flies show no abnormal phenotype under normal laboratory conditions, and show only some abnormal behaviour under conditions of stress. Katz [1988] suggests that this may be due to i) the oligosaccharides providing redundant information only completely uncovered when multiple proteins are simultaneously disturbed or ii) the anti-HRP carbohydrate in the mutant is partially masked, but not completely disabled.

1.2.5 Disease and glycosylation

Mutants affecting genes involved in glycosylation seem to be rare [Varki 1993] which could be explained either by i) such mutations having little biological effect, ii) their being lethal and preventing completion of embryogenesis or iii) the mutations being undetected due to alternative glycosylation pathways.

However, several human genetic diseases have been attributed to defective glycosylation enzymes (for review see Koscielak 1995). For example, a glycosylation defect is thought to cause congenital dyserythropoietin anaemia type II [Fukuda *et al.* 1987] and has been attributed to a mutation in Golgi mannosidase II [Fukuda 1990; Fukuda *et al.* 1990] (see section 1.4.4). Disfunctioning phospho-N-acetylglycosamyl transferase causes I-cell disease and pseudo-Hurler polydystrophy [Kornfeld 1986] and

GlaNAc transferase II deficiency causes carbohydrate-deficient syndrome (CDGS) type II [Charuk *et al.* 1995].

Furthermore, changes in the glycosylation of serum antibody IgG have been found in patients with rheumatoid arthritis [Parekh *et al.* 1985]. This disease is characterized by inflammations in the synovial tissues in conjunction with bone erosion and joint distruction. The inflammation is thought to be associated with immune complexes, aggregates of IgG caused by its defective glycosylation [Parekh *et al.* 1988; Rademacher *et al.* 1988b].

1.2.6 Viruses and glycosylation

There is evidence that oligosaccharides form target structures for micro-organisms, toxins and antibodies *and* that they can mask such target structures [Hansson *et al.* 1984; Markwell *et al.* 1981]. Species-specific glycosylation is significant when considering viruses. Viruses do not code for their own glycosylation enzymes, but must make use of the host's own system and are therefore glycosylated in the manner of the host's own proteins. This may make the virus specific for that host as the carbohydrate side chains provide protection for the virus from the host's immune response during primary infection. To the host organism there is little advantage in providing these traitorous signposts for pathogenic micro-organisms, but it does provide a mechanism for blocking infection by changing glycosylation [Varki 1993]. The interactions of pathogens and oligosaccharides may have resulted in an evolutionary "arms race" and a diversification of the enzymes producing the complex types of oligosaccharide [Marth 1994]. Viral specificity could also be turned to the host's advantage if a soluble glycoconjugate decoy can be used to trap the virus, such as the mucins in mucosal membranes [Varki 1993].

The HIV glycoprotein gp120 is affected by changes in glycosylation. Prevention of glycosylation or prevention of glucose unit removal results in a markedly

lowered fusion activity and decreased infectivity (Feizi *et al.* 1991 and others - see Varki 1993).

1.2.7 General conclusions about glycosylation

The same oligosaccharide structure can have many roles on different proteins, for example the oligosaccharides with mannose-6-phosphate markers (mannoses phosphorylated at position 6) which are responsible for targeting proteins to the lysosome are also found on non-lysosomal proteins (e.g. Lee *et al.* 1988). This indicates that it is only vital in some instances, but these cases are enough to ensure that the machinery for producing this structure is maintained and so it appears on many proteins. The branched nature of oligosaccharides due to their ability to form multiple linkages means that their information carrying ability is large. Glycosylation generates some of the diversity required in recognition events from a limited set of genes. The more complex structures may have co-evolved with pathogens and parasites [Marth 1994], the increasing elaboration due to existing structures already having a vital function. Homologous proteins in different species with different attached sugars suggest that the oligosaccharides are not crucial for their function, however glycosylation could be involved in generating the observed morphological and developmental differences between organisms apparently consisting of similar proteins [Varki 1993]. Within a species, differences in glycosylation can be a form of polymorphism such as that found in blood groups, which, as with protein polymorphism, do not mean that the sequence is unimportant, but that it can withstand small differences while retaining its function [Varki 1993]. The maintenance of this polymorphism may be a response to pathogens.

1.3 The *N*-linked glycosylation pathway

Many of the proteins that are transported through the ER and Golgi apparatus are glycosylated (for review see Kornfeld *et al.* 1985). In *N*-linked glycosylation a common branched oligosaccharide is added to proteins in the ER. This structure is built up on the lipid carrier molecule, dolichol phosphate, by the stepwise addition of sugars (Fig. 1.3) and then transferred to an asparagine residue on the nascent polypetide chain. The peptide glycosylation sequence is Asn-X-Ser/Thr where X can be any amino acid except proline [Marshall 1972]. This transfer requires a membrane bound oligosaccharyltransferase, with its active site in the lumen of the ER.

Once this core structure has been transferred it then undergoes subsequent modification during the protein's transport through the ER and Golgi body (Fig. 1.4) The trimming in the ER immediately removes three glucose residues and one mannose. Then as the glycoprotein is transported through the Golgi body, further mannosidases remove mannose units and other glycosyltranferases add sugars. The trimming pathway involves a sequence of enzymes that are discussed in section 1.4. These modifications form the complex, high-mannose and hybrid oligosaccharides that are found in mature glycans (see section 1.1.2).

1.4 Enzymes in the *N*-linked glycosylation trimming pathway

The conservation in eukaryotes of such a complex system involving a large number of enzymes both for the construction and degradation of the oligosaccharide moieties suggests it to be of fundamental importance to organisms as a whole, even if unimportant for some individual glycoproteins. Furthermore it is conceivable that the control of the synthesis of these enzymes, which has been proposed to be one method of generating the diversity of oligosaccharide structures (see below) will also be conserved.

Figure 1.3. Build up of oligosaccharide structure on dolichol phosphate (After Alberts, *et al*, 1989).

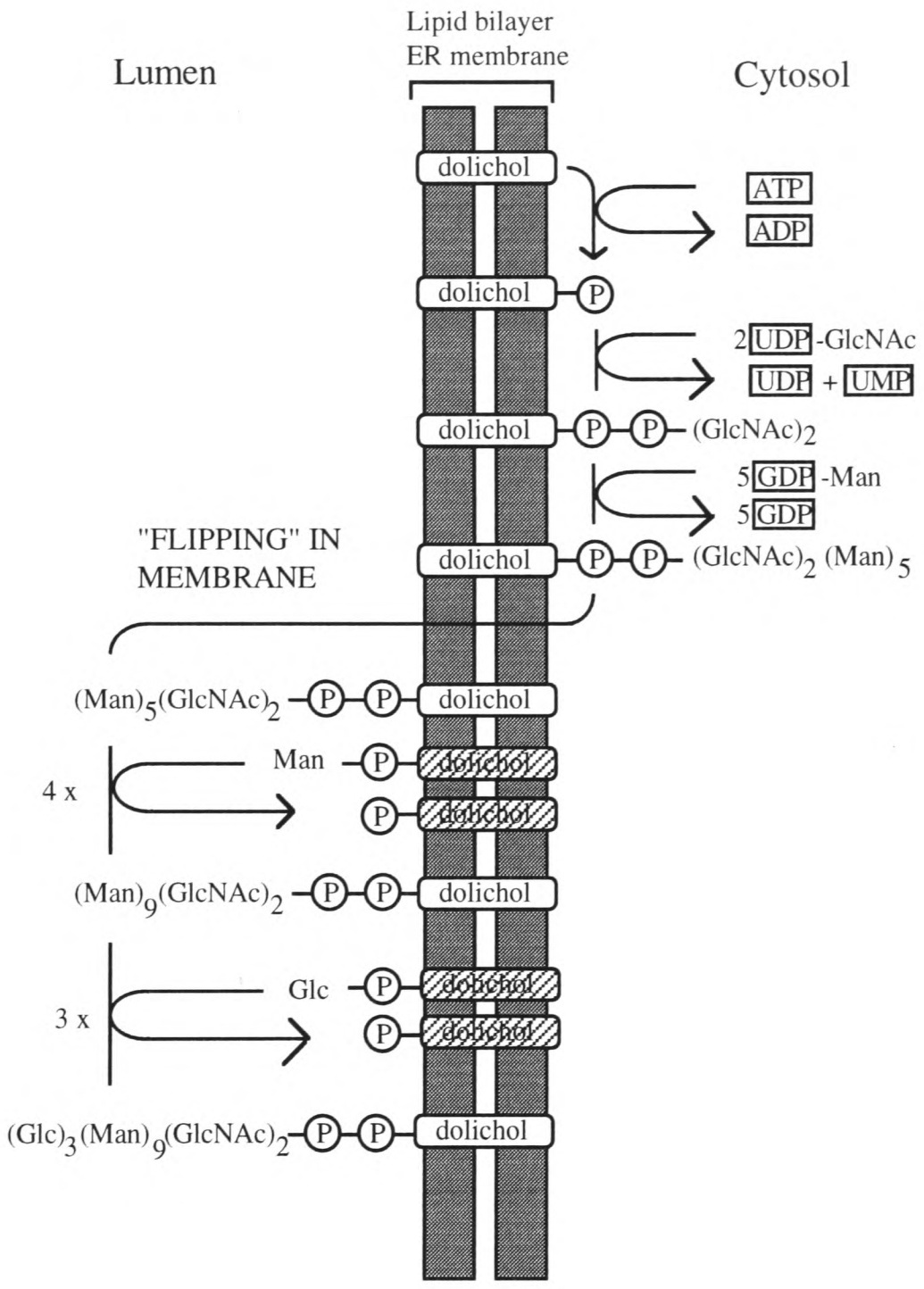


Fig 1.4 Processing of oligosaccharides in the ER and Golgi body

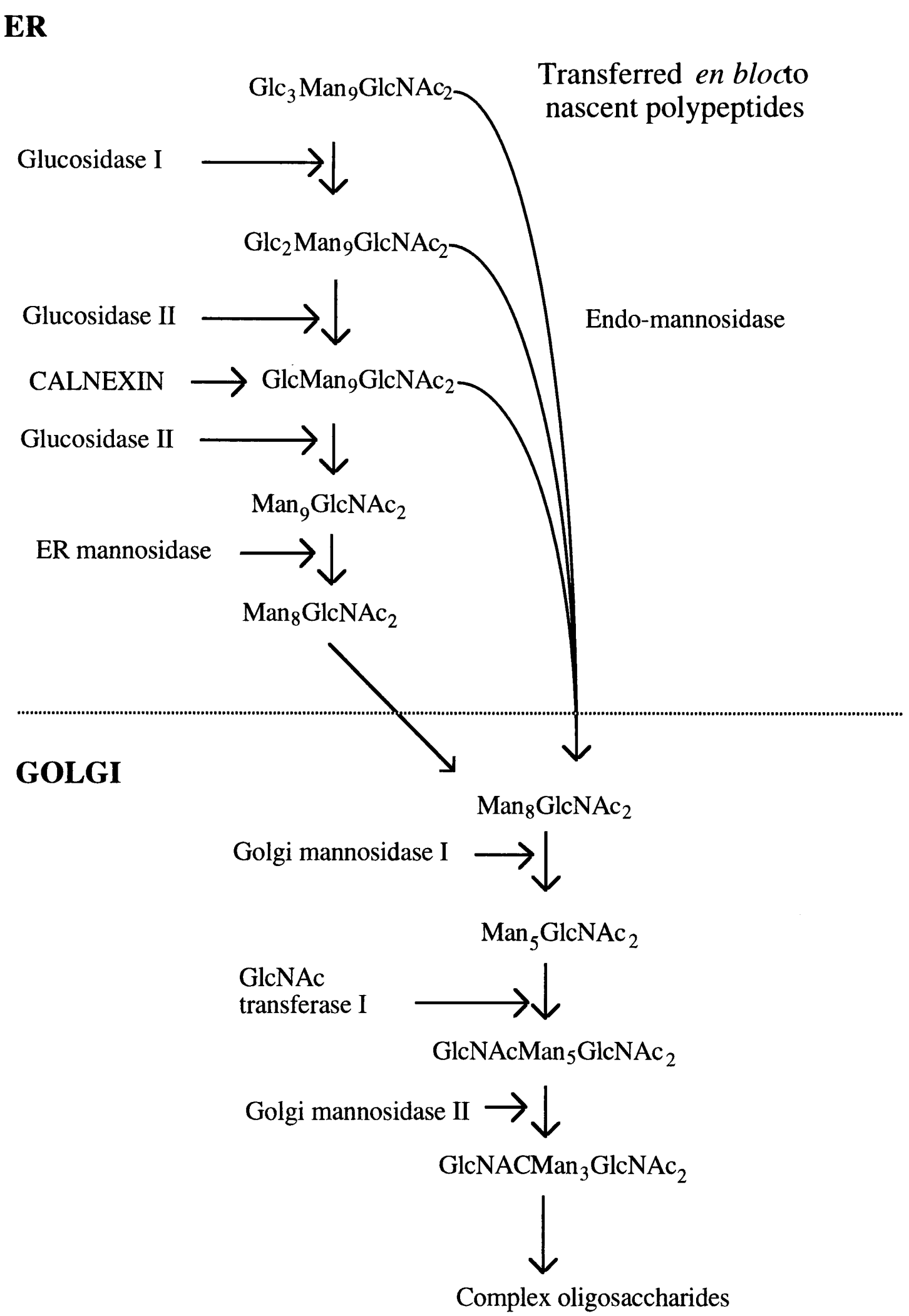


Figure 1.4 shows some of the enzymes that have been identified and play a part in the trimming pathway. A recent review [Moremen *et al.* 1994] suggests that this trimming may be more involved and require more enzymes than originally suggested [Kornfeld *et al.* 1985], particularly since a recycling system transferring glycoproteins back to the ER from the Golgi body has been discovered [Pelham 1991] which confuses the order of processing. Moremen [1994] divides the identified mannosidases into two classes defined by sequence (Fig 1.5). These classes are discussed separately.

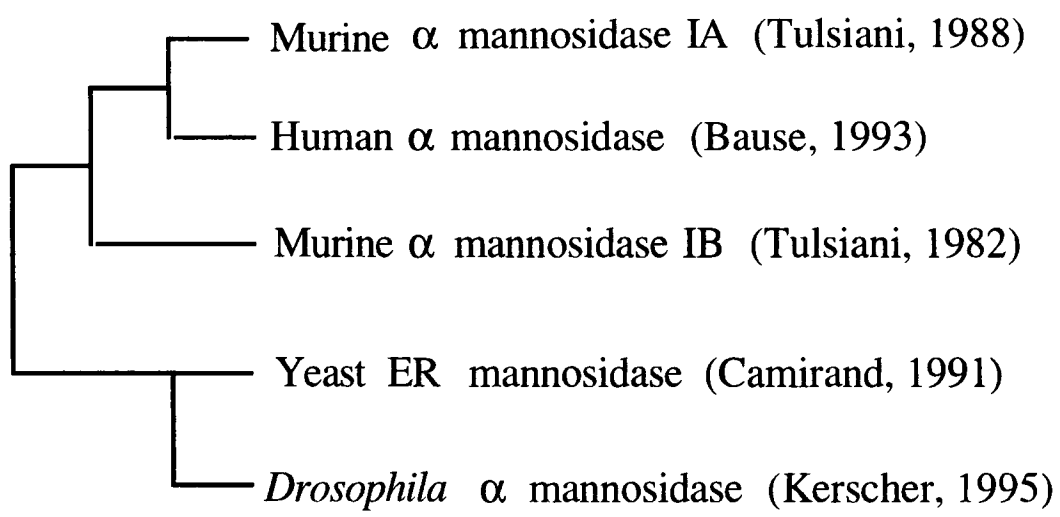
1.4.1 Early ER trimming

After the *en bloc* transfer of $Glc_3Man_9GlcNAc_2$ oligosaccharide to the glycosylation site of a polypeptide in the ER [Kornfeld *et al.* 1985] two glucosidase enzymes remove the terminal glucose units. A specific α1,2 glucosidase I trims the first glucose unit, then an α1,3 specific glucosidase II removes the next two sequentially. Both enzymes have been identified in several organisms including calf liver [Hettkamp *et al.* 1984; Michael *et al.* 1980] and yeast [Bause *et al.* 1986; Saunier *et al.* 1982]. Glucosidase I has recently been cloned [Kalz-Fuller *et al.* 1995] from a human hippocampus cDNA library by PCR using primers derived from the amino-acid sequence of pig liver glucosidase I [Bause *et al.* 1989].

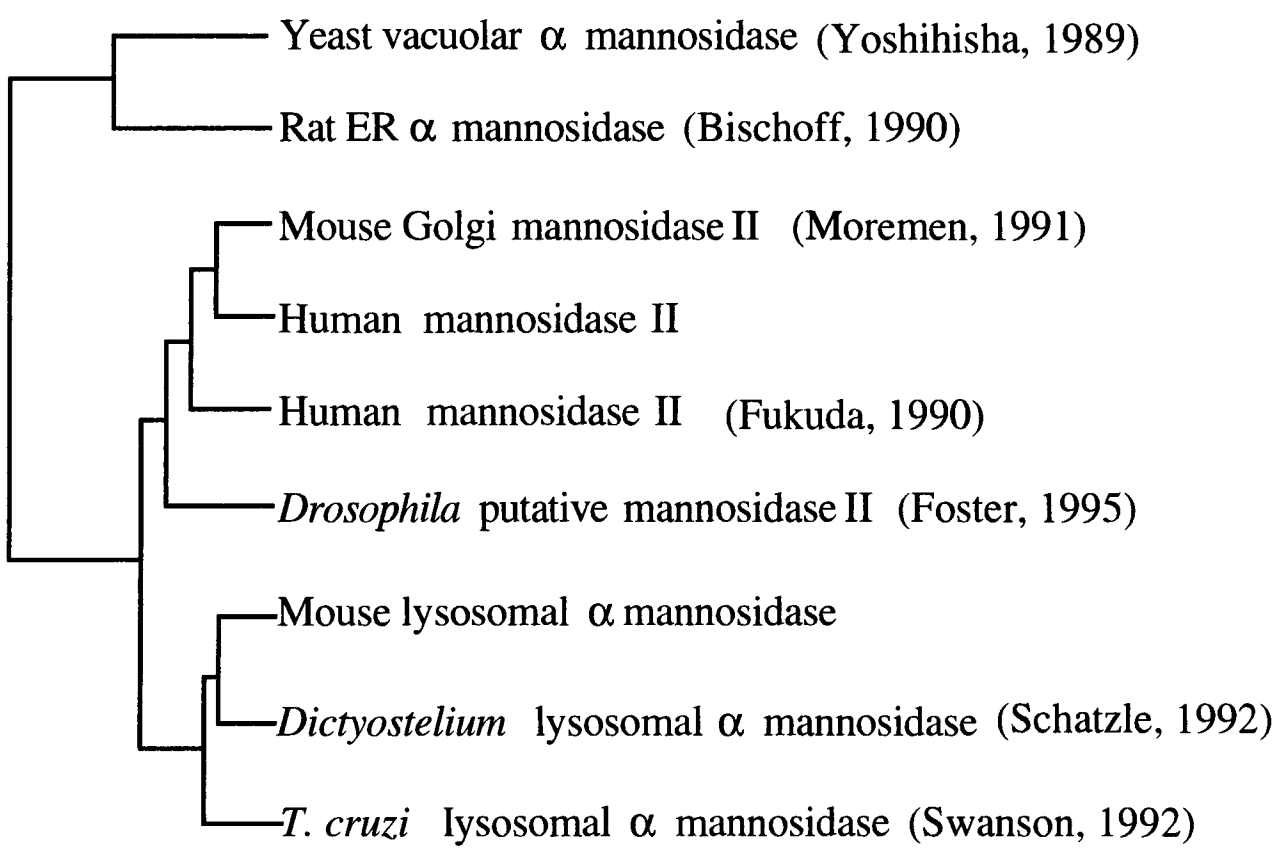
The production of $GlcMan_9GlcNAc_2$ is significant since this substrate is recognised by calnexin, an ER molecular chaperone involved in glycoprotein folding [Ware *et al.* 1995]. This substrate is also produced by the action of UDP-glucose: glycoprotein glucosyltransferase which adds a terminal glucose residue to unfolded glycoproteins with an attached $Man_9GlcNAc_2$ oligosaccharide [Sousa *et al.* 1992]. The folded glycoprotein's $Man_9GlcNAc_2$ carbohydrate moiety is subject to mannose trimming as it passes through the ER and into the Golgi body. An alternative endo α-mannosidase pathway has been identified which by-passes the trimming by the glucosidases [Fujimoto *et al.* 1991]. A mannosidase from rat liver has been identified

Figure 1.5. Relationships between mannosidases (K. Moremen, pers. comm.).

a) Class I α1,2 mannosidases. 63-73kD, inhibited by dMNJ and Ca^{2+}dependent.



b) Class II α-mannosidases. 100-125kD and inhibited by indolizidine, swainsonine and mannostatin.



which is capable of trimming $Glc_{3-1}Man_{9-4}GlcNAc_2$ to $Man_{8-3}GlcNAc_2$ by the removal of $Glc_{1-3}Man$ [Lubas *et al.* 1987; Lubas *et al.* 1988; Moore *et al.* 1990].

1.4.2 Class I mannosidases

Class I mannosidases are 63-73kDa Ca^{2+} dependent enzymes. Mannosidases in this class include a yeast specific (Man_9) ER α1,2 mannosidase [Camirand *et al.* 1991], rat Golgi Mannosidase I [Tulsiani *et al.* 1982] and rabbit [Lal *et al.* 1994] and human [Bause *et al.* 1993] α1,2 mannosidases which remove all four α1,2 linked mannoses from $Man_9GlcNAc_2$. This human mannosidase is 88% identical to the murine protein and 35% identical to the yeast protein. A comparison of these processing mannosidases (Fig 1.5a) showed that they are conserved throughout eukaryotic evolution [Herscovics *et al.* 1994]. A *Drosophila* α-mannosidase has been also been cloned [Kerscher *et al.* 1995] which has 68% similarity to the human mannosidase. This gene is likely to be the *Drosophila* homologue of rat Golgi mannosidase I since it was independently discovered by PCR amplification of fly genomic DNA using primers based on the rat sequence (S. Christodoulou, pers. comm.). The redundancy of the *N*-linked glycosylation pathway at this point (see Fig. 1.4) means that flies mutant for this gene show only small phenotypic differences from wild-type flies [Kerscher *et al.* 1995].

1.4.3 Class II mannosidases

The second class of mannosidases includes Golgi mannosidase II, which has been identified in mice [Moremen *et al.* 1991]. This enzyme recognizes the substrate $GlcNAcMan_5GlcNAc_2$. The mouse Golgi mannosidase II is similar to lysosomal α-mannosidase identified in *Dictyostelium discoidium* [Schatzle *et al.* 1992], which can however trim the $Man_5GlcNAc_2$ substrate without the addition of N-acetylglucosamine. It also has some similarity to ER/cytosolic mannosidases several of which have been

identified in mammals. An ER mannosidase identified in the rat [Bischoff *et al.* 1990] has 33% identity and 54% similarity to a yeast vacuolar mannosidase [Yoshihisha *et al.* 1989] both of which fall into this class of mannosidases. We have identified a *Drosophila* structural homologue of Golgi mannosidase II [Foster *et al.* 1995] which also falls into this class of mannosidase.

1.4.4 Golgi mannosidase II

Golgi mannosidase II (EC 3.2.1.114) trims two mannose units from the structure $GlcNAcMan_5GlcNAc_2$ to form $GlcNAcMan_3GlcNAc_2$ [Tulsiani *et al.* 1982] and circumvents the production of hybrid structures. The enzyme has high specificity for this substrate which is made by the addition of a *N*-acetylglucosamine unit by GlcNAc transferase I (see section 1.4.5). The inhibition of Golgi mannosidase II by swainsonine leads to the accumulation of pentamannose oligosaccharides [Tulsiani *et al.* 1983; Kang *et al.* 1993] suggesting that there is no alternative pathway at this stage. Reduced α-mannosidase II activity has been found in the human disorder, hereditary erythroblastic multinuclearity with a positive acidified-serum-lysis test (HEMPAS) [Fukuda 1990; Fukuda *et al.* 1990] which is a debilitating, but not lethal disease. The defect in this disease is probably in the 5' control region since there are reduced levels of mRNA leading to less than 10% of the normal mannosidase activity, although this could also be due to unstable mRNA [Fukuda *et al.* 1990].

1.4.5 Complex oligosaccharides

Glycosyltransferases transfer sugar residues from an activated donor substrate to a growing carbohydrate group (for review see Paulson *et al.* 1989) to form the more elaborate complex oligosaccharide structures. Some of these enzymes have been identified and include *N*-acetylglucosaminetransferases, galactosytransferases, sialyltransferases and fucosyltransferases. Most of the 100 or more

glycosyltransferases, which it has been estimated are needed to produce the range of oligosaccharide structures found, are involved in elaborating the diverse terminal sequences to form oligosaccharides in the complex group. Several glycosyltransferase cDNAs have been cloned by PCR on the basis of their amino-acid sequence, for example GlcNAc transferase I in humans [Kumar *et al.* 1990] and rabbits [Sarker *et al.* 1991] and rat GlcNAc transferase II [D'Agostaro *et al.* 1995] and III [Nishikawa *et al.* 1992]

The enzyme, GlcNAc transferase I, commits the oligosaccharide to the complex pathway by the addition of a single *N*-acetyglucosamine to the $Man_5GlcNAc_2$ structure. This enzyme has been identified and mutated in mice which die between 9.5 and 10.5 days of development [Metzler *et al.* 1994; Ioffe *et al.* 1994]. Analyses of these mutant mice show that the embryos are developmentally retarded, most noticeably in neural tissue and the left/right body plan asymmetry was randomizied in the absence of complex glygans [Ioffe *et al.* 1994]. The complex oligosaccharides provide vital information required during mouse embryogenesis, including morphogenic events during post implantation development [Metzler *et al.* 1994].

1.5 *N*-linked glycosylation in *Drosophila*

Perturbations in glycosylation can be produced using chemical inhibitors of the enzymes involved [Elbein 1987; Elbein 1991b] and this approach has demonstrated some significant roles for the oligosaccharides in for example viral envelope glycoprotein function [Elbein 1991a]. Cellular interaction has also been disturbed using structurally defined oligosaccharides as inhibitors although not available for all classes of enzyme. The significance of removing oligosaccharides from glycoproteins has also been investigated by the chemical or enzymatic removal of attached sugars or protein engineering to remove glycosylation sites [Varki 1993]. While these methods have suggested glycosylation is highly significant for some processes, they can have

other biological effects confusing the part glycosylation plays in the glycoprotein function [Elbein 1991b; Shur 1994]. A more controlled and reliable method to discover the function of any enzyme is to study natural gene variants or genetic mutants of genes coding for the enzymes involved [Stanley 1987]. In this way it is possible to eliminate, enhance or alter the expression of the enzyme. This has occurred to a certain extent in the mouse especially with the use of the Cre/loxP genetic recombination system to induce mutants [Shur 1994; Marth 1994]. However another organism with much potential for genetic manipulation of glycosylation is *Drosophila* [Williams *et al.* 1991].

The long term aim of our laboratory is to produce and study mutant *Drosophila* strains deficient in specific enzymes involved in protein glycosylation. These mutants should be recessive lethals, but by selecting conditional mutants (for example, temperature sensitive mutants) it will be possible to study the consequences of blocking glycosylation at different developmental stages and on adult behaviour.

1.5.1 *Drosophila* as a model organism

Drosophila melanogaster is a much studied organism with an eighty year history of research. The reasons for this lie in its usefulness as an organism for studying genetics. Fruit flies are i) small, and easily grown and manipulated in laboratory conditions; ii) have a short generation time of about 2 weeks allowing the results of breeding experiments to be obtained quickly; and iii) have many described phenotypic mutations that can be used as markers in genetic experiments. Fruitflies also provide a useful developmental model, with a complex nervous system, but with a genome approximately one twentieth that of mammals allowing the rapid assimilation of molecular information. There is, therefore, an extensive background of previous research in *Drosophila*. There is also a large community of scientists working on *Drosophila*, creating and investigating clones and mutants of *Drosophila* genes and exchanging the information, flies and clones discovered. All these factors make

Drosophila ideal to study the genetics of any system, including *N*-linked glycosylation, provided this pathway is sufficiently similar in other eukaryotes to allow the results of such an investigation to be considered universal.

1.5.2 Glycosylation in insects

Several analyses of oligosaccharides in insect cells have shown only high mannose *N*-glycans in endogenous proteins [Butters *et al.* 1981; Ryan *et al.* 1985; Nagao *et al.* 1987] and in cloned glycoproteins expressed in insect cells [Kuroda *et al.* 1986; Wojchowski *et al.* 1987; Lucklow *et al.* 1988; Jarvis *et al.* 1989]. Sialic acid has been identified during development in *Drosophila* [Roth *et al.* 1992] although the type of oligosaccharide linkage is unknown. Complex oligosaccharides have been identified in lepidopteran cell culture [Davidson *et al.* 1990]. *Spodoptera frugiperda* cells were used to express recombinant human plasminogen protein which demonstrated 40% complex oligosaccharides [Davidson *et al.* 1990]. Further analysis showed these species of complex oligosaccharides to occur in a time dependent manner after baculovirus infection [Davidson *et al.* 1991a]. Incubating cell extracts with $Man_9GlcNAc_2$ substrate [Davidson *et al.* 1991b] showed that baculovirus infection "induces" the expression of an α-mannosidase that prevents the usual accumulation of $Man_6GlcNAc_2$ by cleaving it to $Man_5GlcNAc_2$, the substrate for GlcNAc transferase I [Ren *et al.* 1995]. This is the first enzyme in the committed complex N-glycan pathway, after which other glycosyltranferases can act [Tabas *et al.* 1978]. This implies that the cells possess the ability to synthesize these proteins, but that they do not act under normal circumstances.

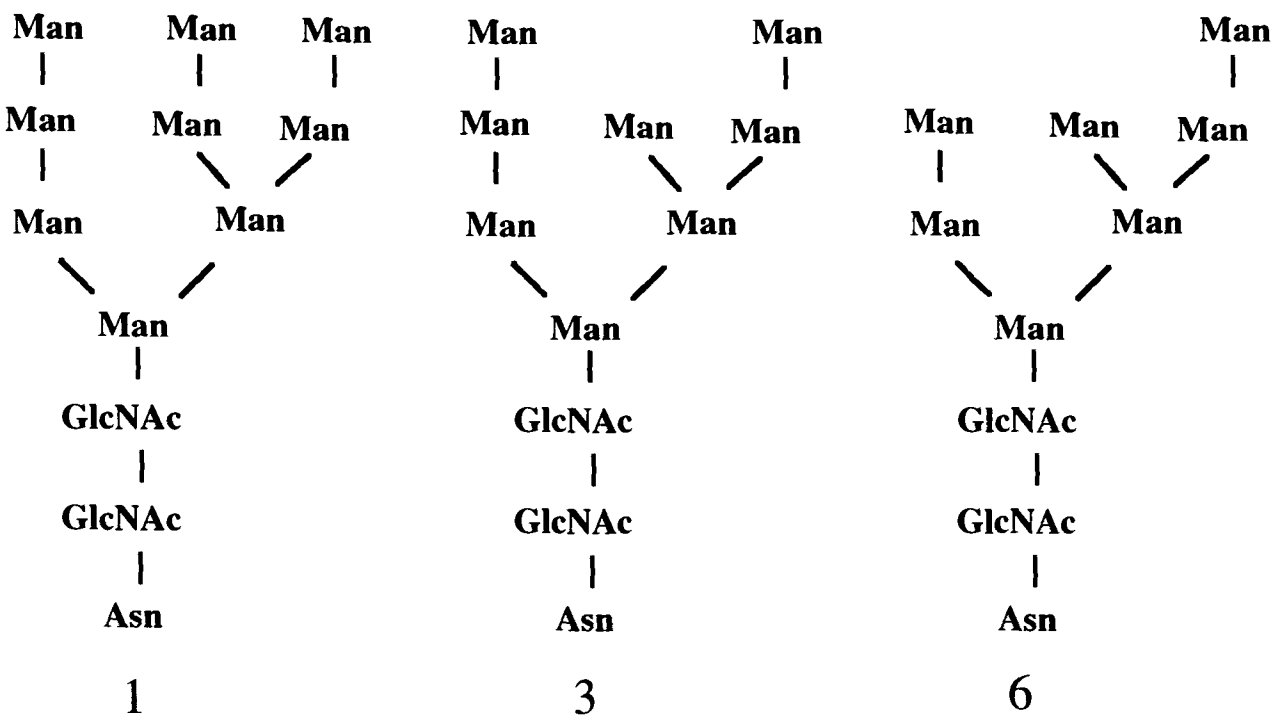
The branched precursor that is transferred *en bloc* to the protein in the ER has been identified in *Drosophila* [Parker *et al.* 1991] showing that the build-up pathway is likely to be the same in *Drosophila* as in other eukaryotes. Analysis of oligosaccharides of *D. melanogaster* [Williams *et al.* 1991] has shown that the major *N*-linked species

are of the oligomannose series (Fig. 1.6) with attached α1-6 fucoses in some instances. At first sight this series seems too limited to provide the diversity of structures required for specific recognition thought to be provided by attached oligosaccharides. However due to fact that sugars can form several linkages, producing the branched oligosaccharide structure, and these structures form different conformations [Balaji *et al.* 1994], this series is more extensive than it appears. The number of ways each collection of sugars can be arranged is given beneath each structure (Fig. 1.6). The analysis of *Drosophila* oligosaccharides investigated 95% quantitatively of the sugar structures found associated with membrane bound proteins in third instar larvae and on the larval serum protein 2. The remaining 5% may contain the more complex structures thought to be required for the recognition roles provided by glycosylated proteins. Small amounts of unusual structures may have been hidden from the analysis technique.

There seems to be some confusion therefore about whether complex oligosaccharides are found in *Drosophila* glycoproteins. They have not been found by analysis of sugars, but seem to be made under certain conditions. The presence of complex oligosaccharides is suggested by the presence of sialic acid (see above) and by the binding of plant lectins (carbohydrate binding proteins) [Callaerts *et al.* 1995] although the oligosaccharide linkage in both these cases is unknown. However, since only the build-up pathway oligosaccharides and oligomannose glycans have been definitely identified in *Drosophila* and there is some doubt as to the finding of complex *N*-glycans in insects under normal conditions, we have chosen to study the earlier steps in the *N*-linked glycosylation pathway including mannosidases in the ER and Golgi body.

The discovery of Golgi mannosidase II in *Drosophila* [Foster *et al.* 1995] provides indirect evidence to suggest the formation of complex oligosaccharides in this organism if this homologue is shown to have the same substrate specificity as its mouse

Figure 1.6. The oligomannose series found in *Drosophila*. The number beneath each figure is the number of isomers which it can form.



homologue and an α-mannosidase II recently identified in cultured insect cells (Bm-N) [Altmann *et al.* 1995]. If this is the case, it suggests the presence of GlcNAc transferase I activity, considered the first step in the complex oligosaccharide pathway (see section 1.4.5).

1.6 Cloning *N*-linked glycosylation genes in *Drosophila*

The investigation of glycosylation in *Drosophila* has shown it to be consistent with glycosylation found in all eukaryotes although some doubt remains as to the presence of complex oligosaccharides. However, since the basic synthesis pathway and early trimming appear to be similar, discoveries made pertaining to genes involved in *Drosophila* glycosylation can be extrapolated to all eukaryotes. Our aims are to investigate the genes coding for enzymes in the *N*-linked glycosylation pathway by identifying and cloning them. Several different approaches could be used to identify the genes involved. One is to isolate the enzyme, partially sequence and design redundant PCR primers from this sequence. The resulting PCR fragment could then be used to screen a *Drosophila* genomic or cDNA library. However since cDNA clones for some of the enzymes had already been identified in other organisms, these were used as a starting point. The observed similarity in the *N*-linked glycosylation structures found in all eukaryotes suggests the pathway to be of ancient origin. Since, at least in part, the *N*-linked glycosylation pathway is common to yeast and mammals it is likely to be an 'ancient' pathway and the genes coding for the enzymes involved may be conserved. Therefore our approach was to use rodent cDNAs to probe a *Drosophila* cDNA library to identify clones with similar sequence, likely to code for proteins with a similar function. This approach was successful in identifying a *Drosophila* homologue of Golgi mannosidase II [Foster *et al.* 1995].

1.7 Analysis of *N*-linked glycosylation genes

1.7.1 Saturation mutagenesis

The cloned coding sequence can be used to map the gene cytogenetically. Once mapped, classical *Drosophila* genetics allows the selection of mutants in the region. Transformation rescue experiments will allow the identification of mutants of the appropriate gene.

1.7.2 Disruption of normal expression

Another method for investigating the effect of disrupting a gene is to introduce a reverse complimented coding sequence either by injection [Rosenberg *et al.* 1985; Schuh *et al.* 1989] or by transformation with a P-element construct that will express an anti-sense mRNA [Patel *et al.* 1992b]. These techniques block expression by the anti-sense RNA hybridizing to native mRNA preventing translation and produce mutant phenocopies [Rosenberg *et al.* 1985; Qian *et al.* 1988; Patel *et al.* 1992b].

Ectopic expression can be induced by introducing extra copies of the gene under the control of a heat shock or tissue specific promoter that allows the gene to be over-expressed at different stages of development or in different tissues (e.g.[Struhl 1985; Zucker *et al.* 1988]). The drawbacks to these techniques are that with heat shock promoters, although the stage at which ectopic expression is induced can be controlled, the expression is then ubiquitous. This type of construct has also been shown to be "leaky" with a basal level of expression even before heat-shocking and heat shocking could also affect the fly in other ways [Petersen 1990]. Tissue specific expression is limited by the availability of cloned and characterized promoters. Some of these problems can be overcome using the GAL4 activation system [Brand *et al.* 1993]. The coding sequence of the gene of interest is linked to a promoter carrying the GAL4 binding sites also known as upstream activating sequences (UASs). Crossing a fly stock carrying this construct to one carrying specific genomic enhancers or heat shock

promoters linked to the yeast transcriptional activator GAL4 can then induce expression allowing altered expression to be investigated in a range of tissues and developmental stages.

1.7.3 Investigation of normal expression

Analysis of the expression of the gene, and its control can provide some insight as to the function of the gene. Transcripts can be detected by i) northern analysis [Sambrook *et al.* 1989]; ii) *in situ* hybridization of biotin/digoxygenin labelled probe to mRNA [Tautz *et al.* 1989; Gergen 1992; Patel *et al.* 1992a; Tautz 1992]; iii) PCR of first strand cDNA [Gubler *et al.* 1983] and iv) linking 5' sequences of identified genes to a reporter gene whose expression can be assayed [Lis *et al.* 1983; Thummel *et al.* 1988]. Control sequences can be investigated for characterized promoter and enhancer elements. Mutating control sequences either by deletion or point mutation and investigating the effect this has on the expression of the gene can also be used to identify important sequences in the 5' region of the gene.

1.8 Research presented here

This thesis describes the use of a rat ER mannosidase cDNA clone in an attempt to discover the equivalent gene in *Drosophila.* This approach was unsuccessful in identifying a *Drosophila* homologue of a mannosidase but did result in the identification of a putative cDNA for another protein, calnexin, which is closely involved with the *N*-linked glycosylation pathway [Hammond *et al.* 1994]. These results are discussed in Chapter 3. The analysis of a genomic clone for the putative *Drosophila* homologue of Golgi mannosidase II is also described here. The cDNA for this gene [Foster *et al.* 1995] has been used to clone the genomic gene and the intron/exon structure analysed, which is described in Chapter 4. The gene has been transformed into flies for use in a rescue experiment. This transformation is discussed in Chapter 5. The 5' region of the

gene has also been analysed for promoter sequences and by linking to a reporter gene the expression of the gene observed. These experiments are described in Chapter 6.

CHAPTER TWO
MATERIALS AND METHODS

The methods described in this chapter are referred to in the text using the notation {2.} to indicate in which section the practical details are described. Unless otherwize indicated the techniques described in this chapter are standard laboratory procedures developed from [Roberts 1986; Ashburner 1989b; Sambrook *et al.* 1989] and kit protocols provided by the manufacturer.

2.1 Solutions, buffers and media

1) Church hybridization buffer: 0.5M $NaHPO_4$, pH 7.2, 1% bovine serum albumin (BSA), 1mM ethylene diamine tetraacetate (EDTA), 7% sodium dodecyl sulphate(SDS).

2) Filter stripper: 2mM Tris-HCl, pH 7.5, 1mM EDTA, 0.1% SDS.

3) *In situ* hybridization buffer (6ml): 2xstandard saline citrate (SSC) (600µl), formamide (3ml), 5M NaCl (720µl), 1M TRIS pH7.5 (74µl), Denhardts 50x (600µl), 0.5M EDTA (12µl), dextran sulphate (0.6g), H_2O (440µl).

4) 6x Loading buffer II: 0.25% bromophenol blue, 0.25% xylene cyanol, 15% Ficoll type 400.

   2 x RNA loading buffer : 0.15ml formamide (deionized), 0.18ml formaldehyde (37%), 0.2ml glycerol, 50 µl 20x MOPS

5) Luria bertani (LB) medium : 1% Bacto-tryptone, 0.5% Bacto-yeast extract, 1% NaCl pH 7.5.

6) 20 x MOPS (100ml): 0.4M MOPS (Morpholinopropaniesalferic acid) (8.37g), 0.1M NaAc (0.82g), 0.02M EDTA (0.94g), made up to 100ml with DEPC treated water. Adjust pH to 7.0 with 5M NaOH.

7) NZY (500ml): 2.5g NaCl, 2g $MgSO_4$, 2.5g yeast extract, 5g casein hydrosylate. Adjust pH to 7.5.

8) Phosphate buffered saline (PBS): 2.6mM KH_2PO_4, 26mM Na_2HPO_4, 145mM NaCl. Adjust pH to 7.2.

9) Post blot wash: 0.2M Tris-HCl, pH 7.2, 2xSSC.

10) Ringers solution (1000ml): 6.5g NaCl, 0.14g KCl, 0.2 $NaHCO_3$, 0.12g $CaCl_2$, 0.01g NaH_2PO_4.

11) SM buffer (1000ml): 5.8g NaCl, 2g $MgSO_4$, 50ml 1M Tris.HCl, pH 7.5, 5ml 2% gelatine.

12) 20 x standard saline citrate (SSC): 3M NaCl, 0.3M $Na_3C_6H_5O.2H_2O$. Adjust pH to 7.0.

13) 50 x TAE: 2M Tris-acetate, 0.05M EDTA.

14) 10 x TBE: 0.89M Tris-borate, 2mM EDTA, pH8.3.

15) TE: 10mM Tris-HCl, pH 8.0, 1mM EDTA, pH 8.0

2.2 Extraction and purification of plasmid DNA

2.2.1 Alkaline lysis plasmid mini-prep

1) Inoculate 50ml Falcon tubes with 10 ml LB + Ampicillin (75µg/ml) with single colonies from agar plates, and incubate overnight, shaking at 37°C.

2) Take 1ml samples from tubes with well grown colonies, freeze in liquid nitrogen and store at -20°C.

3) Centrifuge Falcon tubes at 3000g to precipitate bacteria (5 mins).

4) Remove medium and suck out excess liquid.

5) Add 300µl lysozyme in GTE to each tube (lysozyme 5mg/ml GTE: 50mM glucose (90mg in 10ml); 10mM EDTA; 25mM Tris-HCl, pH8.0). Resuspend bacteria by vigorous vortexing and leave at room temp for 5mins.

6) Add 600µl NaOH/SDS (fresh, chilled), shake and incubate the tubes on ice for 5mins. (NaOH/SDS : 5M NaOH 200µl, 10% SDS 500µl, H_2O 4.3ml).

7) Add 450µl KAc, invert the tubes and leave on ice for 20mins (KAc must be fresh: 5M KAc (49.7g/100ml) 3ml; Glacial acetic acid 575µl; H_2O 1425µl) Transfer to 2ml microfuge tubes.

8) Centrifuge tubes at 12,000g for 15mins at 4°C and retain supernatant.

9) Extract protein and lipid from the samples with equal volumes of phenol (equilibrated with TE), 1:1 phenol:chloroform (24:1 with isoamyl alcohol) and chloroform. In each extraction, centrifuge the tubes for 2mins at 4°C and retain supernatant.

10) Add one volume chilled propan-2-ol and leave for 15mins on ice.

11) Pellet the DNA by centrifuging for 12,000g for 5mins at 4°C and remove the supernatant.

12) Wash with 70% ethanol (-20°C) and spin as before.

13) Remove the supernatant, dry the pellet and dissolve in 30µl TE + RNAse (25µg/ml) at room temperature.

2.2.2 Smaller scale plasmid DNA purification

Very clean DNA is not required for checking colonies for plasmid inserts, for example after bacterial transformations, and therefore to speed up the process and allow more colonies to be tested, the following modifications to the technique are made.

1) Inoculate 3ml LB in 15ml Falcon tubes. Leave horizontal shaking at 37°C overnight.

2) Pour 2ml of culture into 2ml microfuge tubes and spin at 12,000g in a microfuge.

3) Resuspend cells in 150µl GTE (without lysozyme) and use 300µl NaOH/SDS solution and 225µl KAc solution in steps 6 and 7.

4) Omit phenol/ chloroform extractions (step 9.)

5) Resuspend in 30µl TE.

6) To minimise degradation due to contaminating nucleases digest with restriction enzymes for a maximum of 2hrs.

2.2.3 P1 clone DNA purification

P1 plasmid DNA can be extracted by the alkaline lysis mini-prep method with the following modifications

1) Grow colonies in LB with kanamycin (25μg/ml) and isopropylthio-β-D-galactoside (IPTG) (1mM).

2) Pipette DNA solution carefully to prevent shearing of DNA.

2.2.4 Qiagen Midi prep

This Qiagen kit is used to prepare the large, pure amounts of DNA required for transformation of flies.

1) Inoculate 3ml LB/Ampicillin (75μg/ml) with bacterial colony from plate. Grow shaking at 37°C for ~9hrs.

2) Use 0.5ml of culture to inoculate 50ml LB. Grow overnight, shaking at 37°C.

3) Harvest cells by centrifuging for 15mins, 6,000g at 4°C. Drain tubes.

4) Add 4ml resuspension buffer P1 (100μg/ml RNAse A, 50mM Tris/HCl, 10mM EDTA, pH8.0) and resuspend cells completely.

5) Add 4ml lysis buffer P2 (200mM NaOH, 1% SDS). Mix by inverting 4-6 times. Leave for 5mins.

6) Add 4ml neutralization buffer P3 (3.0M KAc, pH 5.5). Mix by inverting 5-6 times. Incubate on ice for 15mins.

7) Mix samples again and centrifuge for 30mins, 30,000g at 4°C.

8) Recentrifuge if necessary for 15mins.

9) Equilibrate Qiagen tip by adding 4ml equilibration buffer QBT (750mM NaCl, 50mM MOPS, 0.15% Triton X-100, 15% ethanol, pH7.0). Allow column to empty by gravity flow.

10) Apply supernatant to tip and allow to flow through. The plasmid DNA should remain bound in the column.

11) Wash 2 x 10ml with wash buffer QC (1.0M NaCl, 50mM MOPS, 15% ethanol, pH 7.0.)

12) Elute with 5ml elution buffer QF (1.25M NaCl, 50mM Tris/ HCl, 15% ethanol, pH 8.5.) Collect in 10ml tube.

13) Precipitate DNA with 0.7 volumes propan-2-ol at room temperature. Centrifuge immediately for 30min, 15,000g at 4^{o}C. Remove supernatant.

14) Wash with 5ml cold 70% ethanol and recentrifuge.

15) Resuspend in 5µl TE.

2.3 DNA extraction from flies

1) Use 2ml Extraction buffer (100mM Tris HCl, pH 8.5; 50mM NaCl; 50mM EDTA; 1% SDS; 0.15mM spermine; 0.5mM spermidine) for 50 flies.

2) Tap flies into screw-topped vial and plunge into liquid nitrogen.

3) Place 50 flies into homogenizer and crush.

4) Add 1ml extraction buffer and transfer to 15ml Falcon tube.

5) Rinse homogenizer with a further 1ml extraction buffer.

6) Incubate vials with proteinase K (100µg/ml) for 2 hr at 37^{o}C.

7) Extract once with an equal volume of phenol equilibrated with TE, twice with phenol and chloroform:isoamyl alcohol (24:1) and once with chloroform:IAA alone. Shake the tubes, centrifuge at 3000g for 5mins and transfer the supernatant to another vial each time.

8) After the last centrifugation transfer the supernatant to large (30ml) corex tubes. Add 200µl of 3M NaAc and 4ml 100% ethanol (-20°C) to each tube to precipitate the DNA.

9) Leave overnight at -40°C.

10) Centrifuge the tubes at 10,000rpm for 15min to pellet the DNA. Pour ethanol away and replace with 70% ethanol (-20°C)

11) Centrifuge again, pour off liquid and dry tubes.

12) Resuspend the DNA in 50µl TE and rinse the tube with 25µl TE. Store at -40°C.

2.4 Digestion of DNA by restriction endonucleases

Restriction enzymes (REs) cut DNA at specific recognition sequences. Each RE cleaves within a different, often palindromic, sequence of base pairs. Enzymes were supplied with 10x reaction buffers by GibcoBRL unless otherwise stated.

2.4.1 Plasmid digests

For each gel lane use 2µg DNA, 1µl enzyme, 1µl 10 x Buffer and make up to 10µl with deionized water. NB. 1µl of enzyme is sufficient even for larger digests. Leave 2-16hrs at 37°C.

Cutting plasmids with more than one enzyme is sometimes required to remove the plasmid vector from the insert. When the two enzymes require the same 10x Buffer this can be done simultaneously; when needing different buffers it may be possible to do this sequentially. Use the enzyme requiring the lowest salt concentration first. After the usual period of digestion change the buffer concentration and add the second enzyme. Where buffers are not compatible it is necessary to ethanol precipitate the DNA and resuspend between digests.

2.4.2 Genomic digests

For each lane: 5µg DNA, 1µl enzyme, 2µl 10x buffer, 2µl spermidine (25mM), 1µl RNAse, make up to 20µl with deionized water. Leave overnight at 37°C.

2.5 Gel electrophoresis

2.5.1 DNA markers (sizes in bp)

1) λ HindIII ladder (Gibco BRL): 23130, 9416, 6557, 4361, 2322, 2027, 564, 125.

2) 1kb ladder (Gibco BRL): 12216, 11198, 10180, 9162, 8144, 7126, 6108, 5090, 4072, 3054, 2036, 1636, 1018, 517, 506, 396, 344, 298, 220, 201, 154, 134, 75 (Hartly, 1980)

3) 123bp ladder (Gibco BRL): 1476, 1353, 1230, 1107, 984, 861, 738, 615, 492, 369, 246, 123.

2.5.2 Standard gel electrophoresis

1) Pour 0.8% agarose in TBE into a gel mould held at 4°C.

2) Load wells with samples and 1xloading buffer (final) and run gel in TBE at 100V for 1-1.5 hr.

3) Remove gel to a dish with ethidium bromide (0.5mg/ml) and leave shaking for 10-15min.

4) View under UV light and photograph.

2.6 DNA purification from agarose gels

2.6.1 Use of geneclean II^{R} (BIO 101 inc)

1) Run DNA on an electrophoresis gel {2.5}, 0.8% agarose in TAE.

2) Stain, photograph and excise band.

3) Divide the excised fragment into approximately 200mg slices and chop into small pieces. Place agarose slices into Eppendorf tubes, vortex briefly and spin down.

4) Add at least 3 volumes NaI solution (300µl/100mg agarose), incubate at 45-55^{o}C for 2min, mix and incubate further until agarose had dissolved (approximately 5mins)

5) Chill the tubes on ice and keep cold for the following stages.

6) Add Glassmilk™ (previously vortexed for 1min) using 5µl for 5µg or less of DNA. Mix and incubate on ice for 5-15min. The Glassmilk™ beads adhere to the DNA allowing other impurities to be removed. The tubes must not be vortexed in the following processes.

7) Centrifuge tubes and discard the supernatant.

8) Wash pellet 3x with NEW wash using 50-250µl for each tube. Centrifuge between washes and discard supernatant.

9) After the third wash dry the tubes.

10) Add 7µl TE to each tube and incubate at 45-55^{o}C for 2-3 min. Centrifuge for 30sec and transfer the TE/DNA to a fresh tube.

11) Repeat elution from Glassmilk™ with a further 7µl TE/tube.

12) Combine washes and check on a minigel {2.5}.

2.6.2 Using Low Melting Point (LMP) agarose

1) Run 0.8% LMP agarose gel {2.5} at 4^{o}C. Excise DNA band in gel slice. Melt slice at 65^{o}C for 20 minutes.

2) Add TE to give a final volume of 1ml. Extract DNA solution with an equal volume of phenol; phenol/chloroform and 2x chloroform.

3) Add 100µl 3M NaAc and 2-5 volumes of ethanol (-20^{o}C). Leave overnight at -20^{o}C or 1hr at -40^{o}C

4) Spin for 10min in a microfuge.

5) Wash pellet in 70% ethanol and respin. Remove supernatant and dry pellet.

6) Resuspend in 10µl TE

2.7 Southern blotting and radioactive hybridization

2.7.1 Southern blotting [Southern 1975] [Botchan *et al.* 1976]

1) After viewing and photographing, place gel on the Southern blotting apparatus (a platform covered with two perpendicular strips of 3MM paper soaked in 0.4M NaOH surrounded by a reservoir of 0.4M NaOH)

2) Cut a Hybond N^+ nylon filter (Amersham) to size and place on top (remove all air bubbles).

3) Place strips of cling film around the sides.

4) On top of the gel and filter place two sheets of 3MM paper (pre-soaked in 0.4M NaOH) and a stack of paper towels. Apply a small weight evenly to the top of the paper towels.

5) Leave overnight.

6) Dismantle blot. Treat the filter by washing in Post Blot Wash for 15min followed by 2min in 2xSSC. If not being probed immediately, blot with 3MM paper, wrap in cling film, heat seal in polythene and store at -20^{o}C.

8) If used straight away, leave in 2xSSC and transfer, rolled in a nylon mesh to hybridization bottle.

2.7.2 Radioactive labelling of probes

Using rapid Multiprime DNA labelling system (Amersham International plc).

1) Place 25ng DNA probe and 25µl TE in pierced Eppendorf tube and boil in a water bath for 10min. Stand on ice.

2) Add 10µl labelling mix (dATP, dGTP, dTTP in buffer solution containing Tris-HCl, pH 7.8, $MgCl_2$ and 2-mercaptoethanol) and 5µl primer solution (random hexanucleotides in an aqueous solution containing nuclease free BSA.)

3) Add 1µl Klenow (1u/µl DNA polymerase I in 50mM KPO_4 (pH 6.5), 10mM 2-mercaptoethanol and 50% glycerol) and 2µl (~ 20µCi) ^{32}P dCTP (Amersham).

4) Leave for 2 hr at 37°C.

2.7.3 Spin Dialysis

1) To produce a column of Sephadex G-50 in Eppendorf tube pierce the bottom and top of the tube with 21-gauge syringe needle. Pack a small amount of glass wool in the bottom. Fill the tube with a slurry of Sephadex G-50 in TE and partly dry the column by spinning for 7secs in a bench centrifuge with a Nunc tube beneath to collect the eluate. Add more of the mixture to the tube and spin as before.

2) Replace the lower Nunc tube to remove the TE and leave it empty to collect the labelled probe.

3) Remove the reaction mixture from 37°C. Add 50µl TE and transfer 100µl TE/Probe mix to the Sephadex G-50 column. Spin down (7secs).

4) Add 100µl TE to the probe reaction tube. Transfer 100µl TE/Probe mix to the column and spin as before.

5) Remove the column tube and the Nunc tube containing eluted probe. The approximate ratio of incorporated to unincorporated radioactivity can be estimated by comparing the count rate of column (containing labelled nucleotides) to Nunc tube (containing labelled probe.)

6) Place the labelled probe tube in a boiling water bath for 10min to denature the probe.

2.7.4 Hybridization [Church *et al.* 1984]

The stringency of hybridization can be altered both by changing the temperature at which hybridzation occurs (42-65^{o}C range, higher temperature, more stringent) and the concentration of SSC in which the filters are washed (6-0.1xSSC, lower concentration, more stringent)

1) Roll the filter up in nylon mesh and unwind in a hybridization bottle.

2) Add 10ml of filtered Church hybridization buffer and place the bottle in the hybridization oven at 65^{o}C. Leave for at least 1hr.

3) Remove the hybridization probe from the boiling water bath and add 800µl Church buffer at 65^{o}C.

4) Transfer the probe mixture to the hybridization tube.

5) Mix and leave in rotating oven at 65^{o}C overnight.

6) Remove hybridization bottle from the oven and carefully pour the probe mix away. Add 1-2 inch depth of SSC, preheated to 65^{o}C and replace the bottle in the oven for 10min.

8) Remove the bottle from the oven and pour the solution away. Place the filter and mesh in a plastic box, unwind and remove the mesh. Cover the filter with SSC and leave for 15min in a shaking water bath at 65^{o}C.

9) Remove the box, drain and monitor filter. Wash further at 65^{o}C if necessary, to reduce counts.

10) Blot filter between sheets of 3MM paper, wrap in cling film and mount in autoradiograph case.

2.7.5 Autoradiography

1) Place x-ray film (Blue-sensitive X-ray film, Genetic Research Instrumentation, Dunmow, Essex) into autoradiograph case under safe light.

2) Leave at -70°C for appropriate length of time (depending on radioactive counts).

3) Develop in dark-room. Remove film under safe light and wash in developer for 3mins (less if very black)

5) Rinse in water for 1min.

6) Wash in fixer for 3 min.

7) Wash in running water for 5min and hang up to dry.

2.7.6 Reuse of filters by stripping

1) Heat filter stripper to 65°C.

2) Place filters in a plastic container and leave with filter stripper shaking at 65°C for 15mins.

3) Monitor filters and wash further with fresh filter stripper if still registering radioactive counts.

2.8 Radioactive end-labelling of DNA

2.8.1 λ ladder

1) Mix in an Eppendorf tube :

DNA	2µg
Water	15µl
Mixed Nucleotide buffer	5µl
^{32}P dCTP (Amersham)	1µl (~10µCi)
Klenow fragment	1µl

2) Incubate at room temperature for 10mins.

3) Make up to 100µl with TE.

4) Remove unincorporated nucleotides with spin dialysis through Sephadex G25 {2.7.3}

5) Wash column with 100µl TE.

6) Store eluted labelled DNA at -70°C.

2.8.2 "Genecleaned" inserts

1) In an Eppendorf mix:

3µl "genecleaned" insert.

2µl 5x Polynucleotide Kinase (PNK) buffer (0.7M Tris-HCl, pH 7.6, 10mM $MgCl_2$, 50mM DTT).

4µl Water.

2) Add 1µl [γ-^{32}P] -ATP (Amersham).

3) Incubate at 70°C for 2mins

4) Stand on ice for 2min.

5) Add 0.5µl PNK. Incubate at 37°C for 30min.

6) Incubate 70°C for 2min.

7) Stand on ice for 2min.

8) Add 0.5µl PNK. Incubate at 37°C for 30min.

9) Spin dialyse {2.7.3} through Sephadex G-50 to remove unincorporated nucleotides. Transfer eluted DNA to Eppendorf tube.

10) Add 20µl 3M NaAc and 200µl chilled propan-2-ol.

11) Store tube overnight at -70°C.

12) Spin tubes at 4°C for 15min.

13) Remove supernatant and add 400µl 70% ethanol

14) Microfuge for 15min at 4°C.

15) Remove supernatant and resuspend DNA in 3µl TE.

2.9 *In situ* hybridization

2.9.1 Chromosome Squashes

1) Select wandering third star larvae from fly bottle and wash in water.

2) Dissect out salivary glands in 100mM NaCl using fine-tipped forceps.

3) Dip the glands briefly in 45% acetic acid and transfer to clean siliconized cover slip. Remove excess acetic acid with tissue.

4) Add 7µl of 3:2:1 acetic acid/water/lactic acid (made fresh from 500µl 60% acetic acid/ 100µl lactic acid) Leave for a few minutes until glands become transparent.

5) Place a slide on the drop to pick up the coverslip, turn upright and label.

6) Place the slide between layers of tissue paper and hit several times with the blunt end of a pencil. Examine under phase contrast on the light microscope, looking for well spread but intact chromosomes. Hit further if necessary.

7) Hold coverslip still under tissue and squash onto the slide with the thumb.

8) Dip slide into liquid nitrogen until the bubbling stops. Blow onto the surface to produce a frost and flip the coverslip off using a clean razor blade.

9) Dry the slide in ethanol 3 x 10min.

10) Store all the slides at room temperature in airtight container with silica gel.

2.9.2 Biotin Nick Translation (Nick translation kit, Amersham)

For about 10 slides:

1) In an Eppendorf tube place: 5µl 10x Buffer (300µM each dNTP in Tris-HCl, pH 7.8, MgCl and 2-mercaptoethanol), 2µl Biotin-21-dUTP (Clontech), 1µg DNA, 5µl Enzyme solution (0.5u/µl DNA pol, 10pg/µlDNAse I in Tris-HCl (pH 7.5) MgCl, glycerol and BSA) sterile distilled water up to 50µl.

Mix gently and incubate at room temperature for 3hr.

2) Add 5µl Stop solution (kit) and 15 µl 2% Blue Dextran and heat at 65^{O}C for 5min.

3) Add 1/10 volumes 5M NaCl (7µl) and 2 volumes ethanol (140µl). Stand at -70°C for 15mins to precipitate the DNA.

4) Centrifuge at 12000xg for 5mins and remove the supernatant.

5) Resuspend the pellet in 10µl preheated TE (65°C).

6) Heat at 65°C for 10min.

7) Store at -20°C.

2.9.3 Pretreatment of chromosome slides

1) Wash the slides (held in a rack) in a series of baths.

a)	2xSSC at 65°C	30min
b)	2xSSC at RT.	10min
c)	70mM NaOH	2min (21ml 1M NaOH made up to 300ml with water)
d)	2xSSC	5min
e)	70% Ethanol	5min
f)	70% Ethanol	5min
g)	95% Ethanol	5min
h)	95% Ethanol	5min

2) Allow to air dry.

3) Hybridize slides the same day.

2.9.4 Hybridization

1) Denature the probe (3µl biotinylated probe plus 7µl TE) in a boiling water bath for 10mins, then stand on ice.

3) Add 3µl carrier DNA (Nick Translation Kit) and 150µl *in situ* Hybridization mix.

4) Boil for 10mins then chill on ice.

6) Pipette 15µl of the probe mix onto a siliconized cover-slip (ensuring no bubbles)

7) Pick up the cover-slip by inverting the slide onto it so that the squash is covered. Squeeze a layer of 'puncture repair kit' glue around the edge of the cover-slip.

8) Incubate all the slides overnight at 45°C in a humid box (2xSSC moist towels laid flat)

2.9.5 Peroxidase Staining

1) Cut the glue with a razor blade and float the cover-slips off in 2xSSC.

2) Series of washes:-

a)	2xSSC at 53°C	3 x 20min
b)	PBS at RT.	2 x 5min
c)	PBS -TX	2min (0.1% Triton X-100; 300µl in 300ml)
d)	Rinse in PBS	

3) Pipette 100µl of 1:250 dilution of SA-HRP onto each slide and lower on a coverslip. Incubate the slides at 37°C for 30min in a humid box (PBS moist towels laid flat.)

Streptavadin Horseradish Peroxidase dilution:- 5µl SA-HRP : 1245µl 1x dilution buffer (20ml PBS; 20µl 0.5M EDTA; 0.2g BSA. Stored at 4°C)

4) Remove the coverslips from the slides in a beaker of PBS.

5) Series of washes:-

a)	PBS	2 x 5min
b)	PBS-TX	2min
c)	Stand in PBS	

6) Mix 0.5ml Diamino-benzidene (DAB 10mg/ml PBS) with 9.5ml PBS and 17µl 6% H_2O_2

7) Line up the slides flat in a slide box and pipette DAB solution over the squashes. Incubate in the dark for 45min, checking every 15min that slides are still covered by DAB solution.

8) Drain the DAB off and rinse slides in H_2O and 3 x PBS (Coplin jars). Rinse well as DAB leaves a messy deposit. Leave the slides to dry.

9) Stain with 5% Giemsa (1ml Giemsa, 19ml H_2O, 0.2ml 1M Sodium Phosphate pH 6.8, filtered through Whatman No.1) and rinse well in water. Length of time for staining varies: Stain for 2min and check under microscope.

11) Mount the stained squashes by placing a small drop of warmed XAM on the slide using Pasteur pipette and lowering on a No.0 cover-slip.

12) Store in dark.

13) Examined under 20x and 63x light objective lenses. Band localization determined on the basis of [Lefevre 1976].

2.10 Sequencing

2.10.1 Preparation of oligonucleotides for use as primers

All primers were made by the Val Cooper of the Oxford University synthesis service, Dyson Perrins Laboratory.

1) 17mer oligonucleotides used as primers for sequencing reactions.

2) Heat in sealed vial overnight (15hrs) at 55^{o}C.

3) Cool at -70^{o}C for 30min.

4) Remove cap carefully in fume cupboard and take 1ml to store at -20^{o}C. Place vial in freeze-dryer.

5) Remove from freeze-dryer when all liquid has evaporated. Resuspend in 200µl double distilled water.

6) Transfer to Eppendorf tube. Extract 3x with butan-1-ol, centrifuging at 12,000g for 2mins each time.

7) After butanol removed for last time, add 20µl 3M NaAc and 800µl 100% ethanol. A visible precipitate forms at room temperature. Centrifuge for 10mins, 12,000g and remove supernatant.

8) Redissolve pellet in 200µl double distilled water and reprecipitate by adding 20µl NaAc and 800µl ethanol. Spin down as before.

9) Dry pellet and dissolve in 50µl TE. Store at -40^{o}C.

10) Determine Optical Density at 260nm (1/200 dilution) and use OD_{260} = 1 for 20µg/ml to calculate concentration. For a sequencing primer use at 3ng/µl.

2.10.2 Preparation of template DNA by lithium chloride precipitation to remove RNA

1) After the last ethanol precipitation for a plasmid mini prep, resuspend the pellet in 100µl TE.

2) Add 300µl ice-cold 4M LiCl (3.2g in 20ml) and leave on ice for 30mins.

3) Centrifuge the tubes at 12,000g for 10mins at 4^{o}C.

4) Transfer the supernatants to fresh tubes and add 600µl propan-2-ol. Leave at room temperature for 30mins.

5) Centrifuge at 12,000g for 10mins at 4^{o}C.

6) Discard supernatants and add 1ml 70% ethanol.

7) Vortex briefly and centrifuge for 5mins at 4^{o}C.

8) Remove supernatants and leave the tubes to dry at room temperature.

9) Resuspend DNA pellets in 10µl TE.

2.10.3 Sequencing Reaction [Sanger *et al.* 1977]

Using Sequenase^{R} Version 2.0 (United States Biochemical Corporation)

1) Use 5µg of plasmid DNA to be sequenced for each sequencing reaction. Make the volume up to 25µl with TE.

2) Denature in 0.2M NaOH; 2mM EDTA for 30mins at 37^{o}C.

3) Neutralize by adding 3.25µl 3M NaAc pH 4.5-5.5.

4) Precipitate the DNA with 3-4 volumes of ethanol at -70°C for 15mins

5) Centrifuge for 5mins at 4°C and 12,000g.

6) Wash the pellet with 70% ethanol and centrifuge as before.

7) Discard the supernatant and resuspend the pellet in 7µl water.

Taking no more than 5 tubes at a time:-

8) To each reaction add:

2µl Sequenase reaction buffer (200mM Tris-HCl, pH 7.5, 100mM MgCl, 250mM NaCl))

1µl primer (3ng/µl)

Incubate at 37°C for 30mins.

9) Meanwhile:-

a) Dilute labelling mix for dGTP (7.5µM dGTP, 7.5µM dCTP, 7.5µM dTTP) 1 in 5 in water.

b) Place 2.5µl of ddNTP for dGTP in each of four tubes labelled A G C T for each reaction.

c) Dilute Sequenase™ (13u/µl in 20mM KPO_4, pH 7.5 1mM DTT, 0.1mM EDTA, 50% glycerol) 1 to 8 in ice-cold Sequenase enzyme dilution buffer (10mM Tris-HCl, pH 7.5, 5mM DTT, 0.5mg/ml BSA.)

10) To the 10µl annealed primers add:

1µl DTT

2µl diluted labelling mix

0.5µl (~10µCi) α ^{35}S ATP (Amersham)

2µl diluted Sequenase™.

Mix by pipetting gently up and down. (Avoid introducing air bubbles).

11) Incubate at room temperature for no more than 5mins.

12) Place labelled nucleotide tubes at 37°C for at least 1min.

13) Add 3.5µl of incubated mixture to each nucleotide tube and continue incubation at 37°C for no more than 5mins.

14) Add 4µl of stop solution (95% formamide, 20mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol FF) to each tube and mix. Store tubes at -20°C until ready to load gel.

2.10.4 Preparation of gel

1) Clean siliconized gel plates with Teepol, rinse in hot water and demin. water. Leave to dry.

2) Place 0.4mm spacers between gel plates and tape the edges carefully, with particular care for the bottom corners. Add bulldog clips to press the plates together.

3) Urea/acrylamide solution. Dissolve 56g urea in a 90ml water, add 22.8ml 40% acrylamide (38% acrylamide, 2% bis-acrylamide, Scotlab) and make up to 117ml with water. Add Ambolite MB-3 and filter through Whatman No. 1

4) To make gel:-

10ml 10x TBE

610µl 10% AMPS (fresh)

90ml filtered urea/acrylamide solution

92.3µl TEMED

Mix after adding each component.

10) Pour gel over tray, maintaining constant gentle flow, ensuring no air bubbles. Insert comb and leave to set. (1hr)

11) Clean gel; remove tape and comb. Mark well positions. Assemble apparatus using 1x TBE as buffer in top and bottom tanks. Wash urea out of wells with syringe and needle.

2.10.5 Sequencing gel run

1) Thaw sequencing reaction.

2) Wash out wells again.

3) Heat samples sequentially to 75-80^{o}C for 2mins before loading 2µl with a Hamilton syringe in the order A G C T.

4) Tighten backing plate and run gel at 2kV, 80mA and 38W.

5) After appropriate length of time (3 or 6hrs) switch off gel apparatus and dismantle. Place gel in tray and prise plates apart leaving the gel on one plate.

6) Fix gel and remove the urea by pouring over a mix of 10% methanol, 10% acetic acid in water. Leave for 20-50mins.

5) Lift plate and gel out of fix and drain liquid off. Place 3MM paper over gel and peel off with gel attached. Cut paper to fit gel dryer and cover with cling film. Assemble gel on dryer with a second layer of paper beneath the gel and a layer of plastic over it. Dry under vacuum at 60-70^{o}C for 1.5hrs

6) Expose gel to X-ray film (Kodak BioMax MR film) overnight at room temperature.

2.11 Sub-cloning

2.11.1 Preparation of competent cells

1) Inoculate 3ml LB with *E. coli* DH5α. Leave at 37^{o}C shaking overnight.

2) Use 1ml of overnight culture to inoculate 300-400ml LB. Leave shaking at 37^{o}C until OD_{650} = 0.6.

3) Aliquot 2 x 125ml into sterile pots. Centrifuge 1000g, 10mins at 4oC.

4) Resuspend cells in 2 x 125ml cold sterile water. Pellet as before.

5) Resuspend cells in 2 x 75ml cold sterile water. Pellet as before.

6) Resuspend cells in 2 x 1ml cold sterile water, transfer to Eppendorf tubes and spin in microfuge, 12,000g, 4^{o}C for 1min.

7) Resuspend all cells together in 0.5ml cold water or 10% glycerol, if freezing. Freeze in liquid nitrogen and store at -70^{o}C.

2.11.2 Ligation

1) Prepare vector and insert by digesting with appropriate enzymes overnight {2.4.1}.

2) Where necessary separate the fragments of DNA to be ligated on a electrophoresis gel and extract DNA {2.6}. Where the plasmid vector has only been cut with one enzyme this is unnecessary. However in this case it is often advisable to treat the vector with calf intestinal alkaline phosphatase (1µl CIAP (Boehringer Mannheim GmbH, Germany), 10µl 10x Buffer, water to 100µl) at 37^{o}C for 30mins followed by phenol/chloroform extraction and ethanol precipitation. This reduces recircularization during ligation.

3) The molar ratio of insert to vector required is 10 : 1 for same enzyme restriction site at both ends; 1 : 10 for directional cloning (different termini) and 1 : 1 for blunt ended termini.

4) For the ligation reaction mix in an Eppendorf tube ~200ng vector (2kb), the appropriate amount of insert fragment and water to 15µl.

Heat at 45^{o}C for 5mins to disrupt any recircularized vector.

6) Chill on ice. Add 4µl 5x T4 Buffer (250mM Tris-HCl, pH 7.6, 50mM $MgCl_2$, 5mM ATP, 5mM DTT, 25% (w/v) polyethylene glycol-8000) and 1µl T4 ligase (1 U/µl, Gibco BRL).

7) Incubate overnight at 16^{o}C.

8) Ethanol precipitate by adding 1/10 volume of 3.0M NaAc, pH5.2, mix, then add 2 volumes of ethanol (-20^{o}C). Hold mixture at -40^{o}C for 1 hr. Centrifuge at 12000g for 15min at 4^{o}C. Remove supernatant, dry pellet and wash in 750µl of 70%

ethanol (-20°C). Centrifuge for 5min at 4°C as before. Remove supernatant and dry pellet at room temperature.

9) Resuspend in 5µl water.

2.11.3 Cell transformation by electroporation

1) Use 2µl of the ligation mix and 3µl water. Set up control with 5µl water.

2) Add 40µl of competent cells {2.11.1} and mix, avoiding introducing air bubbles. Place this mixture in an electroporation cuvette, ensuring no air bubbles. Pulse at 2500V, 200Ω and 25µF capacitance. Immediately add 1ml LB to the cuvette.

3) Transfer the cells to plastic vials and allow to recover, shaking at 37°C for 1hr.

4) Plate cells out on selective media. (For pUC vectors use LB/Amp (0.75µl/ml) X-gal (400µg/ml in Dimethyl formamide) IPGT (200µg/ml)). Plate out 100µl aliquot then spin down the remaining cells and resuspend in 100µl and plate out. Dry the plates in a laminar flow cabinet and leave overnight at 37°C.

5) Examine plates for colonies. For pUC vectors white colonies have been transformed by plasmid and insert. Blue colonies contain the plasmid (with Ampicillin resistance gene) but no insert since the galactose gene on vector plasmid is functioning.

6) Inoculate 6 x 10ml LB amp with white colonies and grow up overnight, shaking at 37°C. Mini prep to extract plasmid DNA {2.2.2}.

2.12 Screening a lambda phage library

2.12.1 Preparation of bacteria

1) Grow plating bacteria (C600hfl) in 10ml LB medium, supplemented with 0.2% maltose and 10mM $MgSO_4$, overnight, shaking at 37°C.

2) Pellet bacteria by spinning at 2000rpm, room temperature IEC bench centrifuge; resuspend in 6ml sterile 10mM $MgSO_4$.

2.12.2 Phage Titre

3) Determine phage titre:

i) Serially dilute bacteriophage λ library between 10^{-3} and 10^{-6} using SM buffer.

ii) Mix 100µl of phage with 200µl prepared bacteria and incubate at 37°C for 20min, to allow the bacteriophage to adsorb to the bacterial cells.

iii) Melt NZY-top agar 0.7% and aliquot into 3ml samples. Hold at 47°C. Add the phage/bacteria mix to this and quickly pour onto prewarmed (37°C) NZY agar plates (1.5% NZY, 85mm round Petri plates). Once set incubate the plates, inverted, at 37°C overnight.

iv) Using a plate with countable numbers of plaques, and taking the dilution into consideration, calculate the phage titre. Assume the surface area of one plaque to be $0.5mm^2$, and calculate the amount required for confluence on a 20x20cm plate. Dilute accordingly.

2.12.3 Library plate

4) Make two 20x20cm square Petri plates. Incubate the appropriate volume of phage suspension with 1ml prepared host bacteria at 37°C for 20mins.

5) Add this to a 25ml aliquot of melted NZY top agar held at 47°C, and quickly pour onto a prewarmed (37°C) 20x20cm NZY-agar plate. When set incubate at 37°C overnight. Inversion to prevent condensation is not possible due to the weight of media in a 20x20cm plate.

6) Continue incubation until almost confluent lysis was achieved. Chill the plate at 4°C to allow the agar to become firm.

7) For each plate label three 20x20cm Hybond N filters (Amersham.) Mark the underside of the plates with dots to allow orientation of the filters. Place the first filter

on top of the agar for 2 minutes. Use a 21G needle to pierce the filter above the dots on the plate.

8) Remove the filter and place DNA side up on 2 sheets of Whatman 3MM paper soaked in denaturing solution (3M NaCl: 1M NaOH 1:1 v/v) for 5mins.

9) Transfer filter to 2 sheets of Whatman 3MM paper soaked in neutralising solution (3M NaCl: 1M Tris pH 7.2, 1:1 v/v 1mM EDTA) for 5 minutes

10) Rinse filter in 2xSSC for 2-10 minutes.

11) Consecutively place the second and third filter onto the plate for 5 and 10 minutes respectively, and treat in the same way. Store the plates at $4^{o}C$.

12) Cross-link the DNA to the filters by UV-irradiating, DNA-side down for 3 minutes. If not used immediately, store at -20oC wrapped in cling-film.

13) Hybridize to appropriate hybridization mix {2.7}.

14) Align the autorads using the holes on the filters. Mark positives which coincide, and pick off the plate, using the blunt end of a sterile Pasteur pipette by aligning the autorads to the dots on the plates. Transfer the plug to an Eppendorf tube containing 1ml SM and 0.1ml chloroform, which prevents any bacterial growth.

2.12.4 . Repeat screening procedures to check identifications

15) Round 1: Leave plugs at $4^{o}C$ overnight to elute the phage. Repeat steps 1-3iii, plating the positives onto small 85mm round Petri dishes. Use dilutions in the range 10^{-3} - 10^{-7}.

16) Use plates with confluent lysis for lifts. Repeat steps 7-12 using round 82mm Hybond-N membranes. Take two lifts for each plate.

17) Hybridize the filters to the appropriate probe, as before and pick off positives with the blunt end of a Pasteur pipette.

18) Round 2: Repeat as for round 1. Take lifts from plates with well spaced plaques. After hybridization, pick off positives using the sharp end of a Pasteur pipette.

19) Round 3: Repeat the plating, lifts and hybridization processes, until all the plaques showed positive. Select plaques at random to check.

20) Round 4: Check positives to make sure there is no contamination, by plating out a final time and hybridizing. Select those plugs from the previous round for which all the plaques showed positive for DNA extraction.

2.13 DNA extraction from phage

1) Grow plating bacteria (C600hfl) in 10ml LB medium, supplemented with 0.2% maltose and 10mM $MgSO_4$, overnight shaking at 37^{o}C.

2) Pellet bacteria by spinning at 2000g, room temperature, IEC bench centrifuge; and resuspend in 6ml sterile 10mM $MgSO_4$.

3) Plate out serial dilutions(10^{-1}-10^{-4}) of the selected clone. NZY agarose, rather than agar must be used for both top and bottom of the plated as agar contains ingredients which inhibit restriction enzymes.

4) Add 5ml SM buffer to plates with almost confluent lysis, and shake at room temperature for 4-6 hours to elute the phage.

5) Harvest the SM from the plates and centrifuge at 3000g, room temperature for 10mins.

6) To the supernatant add RNAse A and DNAse I, each to a final concentration of 1mg/ml, and incubate at 37^{o}C for 30mins.

7) Transfer 1ml samples to Eppendorf tubes and add 0.5ml 25% PEG (polyethylene glycol) in 2.5ml NaCl. Vortex the tubes and incubate overnight at 4^{o}C.

8) Pellet the phage particles by spinning at 10,000g, 4^{o}C for 10mins.

9) Discard the supernatant and drain the tubes. Resuspend the phage in 0.5ml SM by vortexing.

10) Add 5ml 10% SDS and 5ml 0.5M EDTA , and incubate the tubes at 68^{o}C for 15mins.

11) Carry out one extraction with an equal volume of phenol/TE and one with phenol/chloroform {as for 2.2}. Precipitate the DNA by adding an equal volume of propan-2-ol, and stand at -70^{o}C for 20mins.

12) Recover the DNA by spinning at 10,000g, 4oC for 15mins. Wash the pellet in 70^{o}C, respin and air-dry.

13) Dissolve in 30ml TE and store at -40^{o}C.

2.14 RNA extraction from flies

NB: All solutions, plastic and glass must be RNAse free. Solutions are treated with DEPC (Diethyl pyrocarbonate) where possible and autoclaved. Solutions which are not suitable for DEPC treatment (e.g.TRIS) are made up from DEPC treated water. All plasticware is baked at 120^{o}C overnight. Where glassware needs to be used it is soaked in water with DEPC prior to autoclaving.

2.14.1 Total RNA extraction

1) Homogenize 0.2ml of frozen tissue (eggs, larvae, males and females) in 2ml phenol saturated in 0.2M NaAc, pH5.0 at 65^{o}C. Remove the liquid to a Falcon tube and rinse the homogenizer with a further 2ml of phenol.

2) Add 3ml NaAc, pH5.0 and 0.8ml 10% SDS.

3) Leave at 65^{o}C for 5min.

4) Vortex, cool to room temperature and add 4ml chloroform.

5) Vortex and centrifuge at 3500g for 5min.

6) Remove upper aqueous phase to fresh tube.

7) Add 2ml phenol (NaAc saturated) and 2ml chloroform. Vortex and centrifuge as before.

8) Remove aqueous phase to a fresh tube.

9) Add 4ml chloroform, vortex and centrifuge.

10) Remove aqueous phase, leaving behind organic and interphases.

11) Add 2.5 volumes (10ml) Ethanol (-20^{O}C).

12) Leave overnight at -20^{O}C (or 1hr at -70^{O}C)

13) Centrifuge at 10,000 x g for 10min. Pour off ethanol.

14) Wash precipitate 70% ethanol, and centrifuge again.

15) Dry pellet and redissolve in 30µl DEPC-treated water.

2.14.2 mRNA extraction. Dynal - Dynabeads^{R} mRNA direct

1) Homogenize 20-50mg of frozen tissue in 1.0ml Lysis/binding buffer (100mM Tris-HCl, pH 8.0, 500mM LiCl, 10mM EDTA, pH 8.0, 1% LiDS (SDS), 5mM DTT). Transfer to Eppendorf tubes.

2) Centrifuge for 30-60secs in a microfuge to remove debris. Store on ice.

3) Prepare Dynabeads Oligo $(dT)_{25}$ (5mg/ml PBS, 0.02% NaN_3) by removing 250µl to Eppendorf tubes in MPC (magnetic particle concentrator). Remove supernatant removed and wash beads with 200µl lysis /binding buffer. Remove supernatant again.

4) Add lysate to the tubes with beads, not in MPC. Mix and leave at room temperature for 3-5 minutes.

5) Replace in MPC and remove supernatant after 2mins (leave longer if solution viscous)

7) Wash 3x with 0.5-1.0ml Washing buffer with LiDS (10mM Tris-HCl, pH 8.0, 0.15M LiCl, 1mM EDTA, 0.1% LiDS) at room temperature.

8) Elute with 8µl elution buffer (2mM EDTA, pH 8.0, placing at 65°C for 2min, repeat with 7µl elution buffer. Store at -40°C

9) Regenerate the Dynabeads :-

a) Resuspend in 200µl Reconditioning solution (0.1M NaOH.)

b) Transfer to new tube, shake and incubate at 65°C for 2mins

c) Place in MPC for 30sec and remove supernatant. Repeat steps 1-3 twice.

d) Resuspend in 200µl Storage buffer (250mM Tris-HCl, pH 8.0, 20mM EDTA, 0.1% Tween-20, 0.02% sodium azide), mix and place in MPC. Remove and discard supernatant. Repeat three times.

e) Resuspend in 250µl Storage buffer.

2.14.3 mRNA extraction - Pharmacia Biotech Quickprep^{R} *Micro* mRNA purification kit

1) Leave solutions to warm to room temperature for 20-30mins. Place extraction buffer at 37°C to dissolve any precipitate, then cool to room temperature. Mix Oligo(dT) cellulose slurry (25mg/ml in storage buffer with 0.15% Kathon CG^{R})and aliquot 1ml into Eppendorf tubes. Place 0.5ml Elution buffer (10mM Tris-HCl, pH7.5, 1mM EDTA) at 65°C.

2) Homogenize ~50mg of tissue in 0.4ml Extraction buffer (buffered aqueous solution of guanidium thiosulphate and N-lauroyl sarcosine)

3) Dilute with 0.8ml Elution buffer. Homogenise then transfer to Eppendorf tubes.

4) Centrifuge homogenate and oligo(dT)-cellulose tubes for 1min.

5) Remove buffer from oligo dT cellulose and replace with cleared cellular homogenate. Resuspend cellulose and mix gently by inversion for 3mins. The mRNA binds to the oligo(dT)-cellulose.

6) Centrifuge 10 secs in a microfuge. Remove supernatant.

7) Wash 5x with high salt buffer (10mM Tris-HCl, pH 7.5, 1mM EDTA, 0.5M NaCl) Add 1ml, resuspend cellulose, spin for 10secs and remove liquid.

8) Wash 2x with low salt buffer (10mM Tris-HCl, pH 7.5, 1ml EDTA, 0.1M NaCl) in the same way.

9) Add 0.3ml low salt buffer, resuspend and transfer to microspin column in a 2ml lid-less microfuge tube. Spin for 5secs.

10) Wash 2x with 0.5ml low salt buffer, discarding the effluent from the lower tube in between 5sec spins.

11) Place column in clean tube. Add 0.2ml prewarmed Elution buffer, spin 5secs and retain eluate.

2.15 RNA gel and northern blotting

1) RNA 1% agarose gel: 1g agarose, 2.5ml 20x MOPS, 44.5ml DEPC treated water. Melt and cool to less than 60°C. In fume cupboard, add 6ml 37% Formaldehyde, mix and pour. Run gel in 1x MOPS, 70V.

2) For each lane: 15µl RNA solution, 7µl loading buffer, 1.5µl EtBr 0.2mg/ml, 1.5µl BrBP (1% Bromophenol Blue)

RNA marker - 5µl of RNA standard I.

3) Blot as for southern blotting, but using Hybond N (Amersham) with 20xSSC instead of 0.4M NaOH. Leave overnight. UV irradiate filter for 3mins, store at -20°C.

2.16 First strand cDNA synthesis from mRNA

Pharmacia Biotech - First-strand cDNA Synthesis Kit.

1) Use 20-150ng of mRNA.

2) Heat to 65°C for 10mins. Chill on ice.

3) Put 5µl Bulk mix (FPLC*pure*^{R} Murine Reverse Transcriptase, RNAguard, RNase/DNase-free BSA, dNTPs in aqueous buffer), 1µl DTT (200mM), and 1µl Not 1 dT(18) primer (1:25 dilution of 5'-d[AAC TGG AAG AAT TCG CGG CCG CAG GAA T_{18})-3' at 5µg/µl) in an Eppendorf tube. Add mRNA and mix by pipetting.

4) Incubate at 37°C for 1hr.

5) Heat to 90°C for 5mins to inactivate enzymes.

6) Store at -40°C. Use 1-2µl for PCR.

2.17 Polymerase Chain Reaction

The polymerase chain reaction (PCR) amplifies DNA fragments between specific oligonucleotide primers which anneal to template DNA.

2.17.1 PCR primer design

The annealing temperature of primers is calculated according to the formula

Tm (°C) = 4 (G+C) + 2 (A+T). [Lathe 1990]

For specific priming an annealing temperature of 58-62°C is best. The primer should contain approximately equal numbers of A/Ts to G/Cs and therefore needs to be in the region of 21-25bp. Longer primers may be required if incorporating restriction sites. A 3bp run of G and C at the 3' and a G at the 5' end of the primer also helps to stabilise its annealing to the template DNA.

2.17.2 Basic reaction mix

For each reaction use:

5µl 10xReaction buffer (500mM KCl, 100mM Tris-HCl, pH 9.0, 1% Triton X-100)

3µl 25mM $MgCl_2$

5µl 2mM dNTPs

2.5µl each primer (200ng/µl)

0.5µl Taq polymerase (Promega)

30.5µl sterile distilled water

1µl template DNA (concentration depends on template)

Set up in 1.5 or 0.5ml Eppendorf and cover with 60µl mineral oil to prevent evaporation during the temperature cycling steps.

The reaction mix can be made up in bulk and aliquoted before the addition of the polymerase and template DNA. This reduces the number of pipetting steps which reduces the risk of contamination. A negative control (no template DNA) should be run with each set of reactions, to check if DNA contamination is present in any of the solutions. If possible the appropriate positive controls should also be run, to allow any reaction problems to be resolved easily.

2.17.3 PCR programs

For most reactions use:

Step	denaturing (94°C)	annealing (58°C)*	elongation (75°C)
1-3	4mins	1min	1.5min
4-6	1min	1min	1.5min
7	repeat 4-6 32 times		
8-10	1min	1min	5min
11	hold at 4°C		

*The annealing temperature was varied according to the calculated primer annealing temperature and previous experimentation.

PCR machines i) Intelligent Heating Block, Genesys Instruments, Cambridge.

ii) Minicycler™, MJ Research, USA, supplied by Genetic Research Instrumentation Ltd, Essex.

2.17.4 PCR electrophoresis gel

Separate PCR products by running 10µl of the reaction sample with 2µl loading buffer on a 1% agarose gel {2.5}

2.17.5 Contamination problems

Contamination problems can be overcome by using new Gilson pipettes used only for pipetting the components of the reaction mix (water, dNTPs, buffer, MgCl, primers and polymerase) except the template DNA. Use the pipettes for no other solutions and use no other pipettes for these solutions. Residual contamination problems can be removed by UV irradiating the reaction mix for 2mins prior to adding the polymerase and the template. These measures should effectively remove contamination.

2.18 Microinjection of *Drosophila* embryos

2.18.1 Apple agar plates

For 2 litres (~120 plates) prepare:-

1) Mix 17.5g agar and 500ml deionized water in each of two 2L flasks. Autoclave at 120psi for 20mins.

2) Dissolve 25g sucrose in deionised water by microwaving briefly. Make up to 500ml and add 500ml clear apple juice.

3) Dissolve 4g Nipagin in 20ml 95% ethanol.

4) Mix together all of the above and stir carefully with a glass rod. Avoid introducing air bubbles. Pour into 8cm Petri dishes. When set place a spot of yeast at the centre of each plate.

2.18.2 Microinjection

1) Place an apple juice agar plate in cage and allow flies to lay eggs for 50mins. Then remove plate and replace with a fresh apple agar plate. Leave flies to lay for 30-50mins.

2) Wash embryos from the plate into a small wire mesh strainer using a spatula and distilled water. Wash embryos and dry briefly by placing strainer onto paper towel.

3) Dechorionate embryos by placing in a watch glass of diluted bleach (70%) for 40secs. Rinse thoroughly with distilled water for 1min, and damp dry again.

4) Transfer embryos to apple agar plate and align 50.

5) Spread a small drop of glue (from soaking sellotape in n-heptane) onto the surface of a coverslip. Allow to dry, then invert and touch aligned embryos in the middle of glued area. They should stick easily with pressing.

6) Dry the embryos by placing in plastic container containing silica gel for 10 mins.

7) Cover embryos with Voltelef oil (10S) to prevent further drying. Place cover slip onto slide with a drop of water.

8) Pull needle from a thin glass tube (1x90mm) using a pipette puller (Narishige) heater No2 set at 7.00. Break open needle by touching to a glass slide (under the microscope.) DNA solution contains construct for injection (150-300µg/µl) and helper plasmid pWC (50-150µg/µl). Spin down tube with DNA solution to remove particles and fill needle by gentle suction with 1ml syringe.

9) Inject DNA into the posterior of embryos, using a micromanipulator.

10) Place coverslip with injected embryos onto an apple agar plate with a yeast spot and leave plate in a plastic box with damp tissues.

11) Examine plate 24-30hrs later and remove larvae from oil into yeast spot. Allow to pupate and then remove to a small food vial to continue development.

2.19 **Fly culture** [Roberts 1986]

1) Stocks are maintained at 25^{o}C and flies transferred to fresh medium every two weeks. Medium is dispensed into 7.5x2.25cm glass or plastic vials or 230ml bottles, and paper and yeast added, and will store at 4^{o}C for several weeks. Bungs are made of polyurethane foam which is reusable after autoclaving.

2) Medium. (Made by Mrs J. Matthews).

i) Cornmeal-treacle-agar medium. For 30 vials - 200ml water, 5g agar, 25g wholemeal flour and treacle. Combine agar with the minimum amount of water and mix flour with 55ml water to make a smooth paste. Boil remaining water and add agar and treacle. Reboil and slowly add flour suspension whilst stirring to avoid lumps. Simmer for 20ml and add Nipagin M (anti-fungal agent). Pour into sterile vials, tilt vials to increase surface area, and seed with live yeast.

ii) Yeast-glucose-agar medium. (Used for producing fat third-instar larvae for dissection). 10g glucose, 10g dried yeast, 2g agar and 100ml water. Make agar and yeast into a paste and add to remaining boiling water. Simmer for 10mins, dispense and seed with live yeast.

3) Fly manipulation. Flies are examined when unconscious on CO_2 diffusers under a dissecting microscope.

2.20 **Staining for β-galactosidase activity** [Ghysen *et al.* 1989]

2.20.1 Embryos

1) Collect embryos on an apple agar plate o/n. Collect and dechorionate in bleach {2.18.2}

2) Transfer embryos to a 1:1 mixture of heptane and fixative (1:10 37% formaldehyde: phosphate-citrate buffer). Shake vigorously, let rest for 5mins, then repeat shaking and resting twice more.

3) Discard heptane (upper phase) and replace fixative twice with phosphate-citrate buffer, pH 7.6 (9:1:10 - 0.2M Na_2HPO_4 (freshly made): 0.1M citric acid: water.)

4) Transfer embryos to a watch glass and blow gently to remove traces of heptane.

5) Replace phosphate-citrate buffer with incubation buffer (5mM potassium ferri- and ferro-cyanides and 0.02% Triton X-100 in freshly made phosphate-citrate buffer). Sink embryos and transfer to multiple well plates.

6) Add a saturating amount of X-gal powder and incubate in a moist chamber at 30^{o}C until stained.

2.20.2 Imaginal discs

1) Dissect imaginal discs from third instar larvae in Ringers solution.

2) Transfer immediately to fixative {2.20.1} in multiple well plate and fix for 15 mins.

3) Remove fix, rinse in phosphate-citrate buffer {2.20.1} and add incubation buffer {2.20.1}. Add a saturating amount of X-gal powder and leave at 30^{o}C overnight.

CHAPTER THREE
ENDOPLASMIC RETICULUM MANNOSIDASE

3.1 Introduction

The approach to cloning a *Drosophila* homologue for ER mannosidase was to i) identify putative cDNA clones from an embryonic library using a previously cloned rat cDNA homologue for the enzyme; ii) analyse these by hybridization after cutting with restriction enzymes; and iii) sequence the most strongly hybridizing RE fragment to demonstrate sequence homology with the rat clone. This approach was successful for finding a Golgi mannosidase II *Drosophila* homologue [Foster *et al.* 1995] (discussed in Chapter 4).

A cDNA clone for ER mannosidase has been identified from rat liver tissue [Bischoff *et al.* 1990], by using PCR (Polymerase Chain Reaction) to amplify fragments between redundant primers designed from the amino acid sequence. A 524bp PCR product was used to screen a cDNA λgt11 rat liver library. Six clones were identified. The longest was sequenced, and contained a long open reading frame. This clone was used to identify a second, over-lapping clone. Together these produced a cDNA clone of 3392bp that contained a single open reading frame of 1040 amino-acids.

This cDNA was expressed in COS cells, producing a 400-fold increase in α-mannosidase activity. Database searches showed portions of the protein sequence to have significant similarity to vacuolar α-mannosidases from *Saccharomyces cerevisiae* [Yoshihisha *et al.* 1989].

The cDNA for rat ER α-mannosidase cloned into Bluescript plasmid was obtained from Joyce Bischoff. A truncated version (lacking the poly adenylated tail characteristic of cDNAs) of ~3.3kb was cloned into the HindIII site of plasmid pUC9 (J. Foster) and named pER man (-).

3.2 Identifying *Drosophila* cDNA clones for ER mannosidase

3.2.1 Library screening

A *D. melanogaster* embryonic cDNA library [Hoheisel *et al.* 1991] was screened (J. Foster) with the truncated version of the rat ER mannosidase clone. The library is arrayed on a high-density *in situ* filter, containing 9216 colonies. Clones were identified by radioactive hybridization of rat ER mannosidase single stranded DNA probe {2.7}. The screening identified several cDNA clones which were supplied by Jörg Hoheisel, Deutsches Krebsforschungszentrum, Heidelberg, Germany (Fig. 3.1).

3.2.2 Initial investigation of the cDNA clones

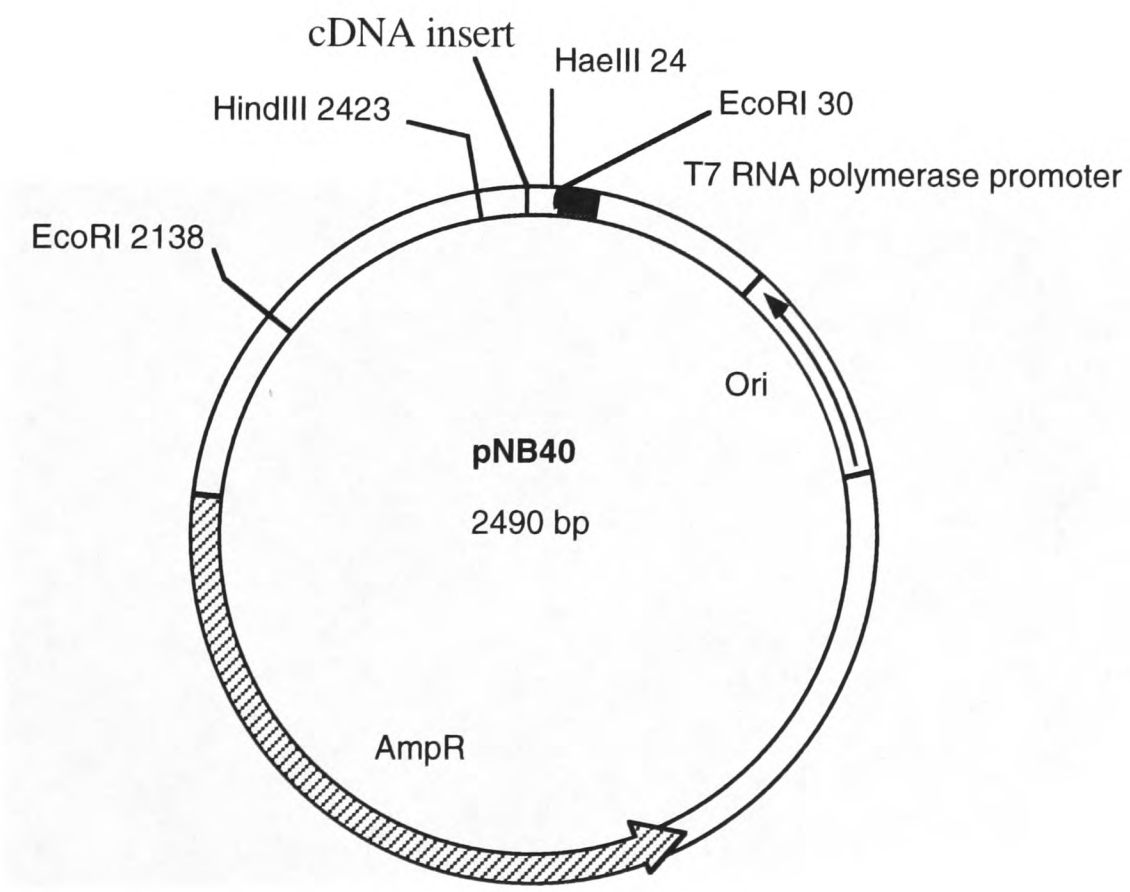
The cDNAs are cloned into the Ampicillin resistant vector pNB40 [Brown *et al.* 1988] but cannot be cleanly cut out (Fig. 3.2). Digestion with HindIII and EcoRI removes most of the vector, leaving 67bp at the 5' end and 30bp at the 3' end of the insert. The clones were initially investigated by digesting with HindIII and EcoRI and separating the fragments by electrophoresis {2.5} (Fig. 3.3a). Six of the seven clones have the same insert fragment sizes (~1.7kb) and one has a slightly shorter insert (~1.4kb). Clone 33 and 36 were chosen for further investigation.

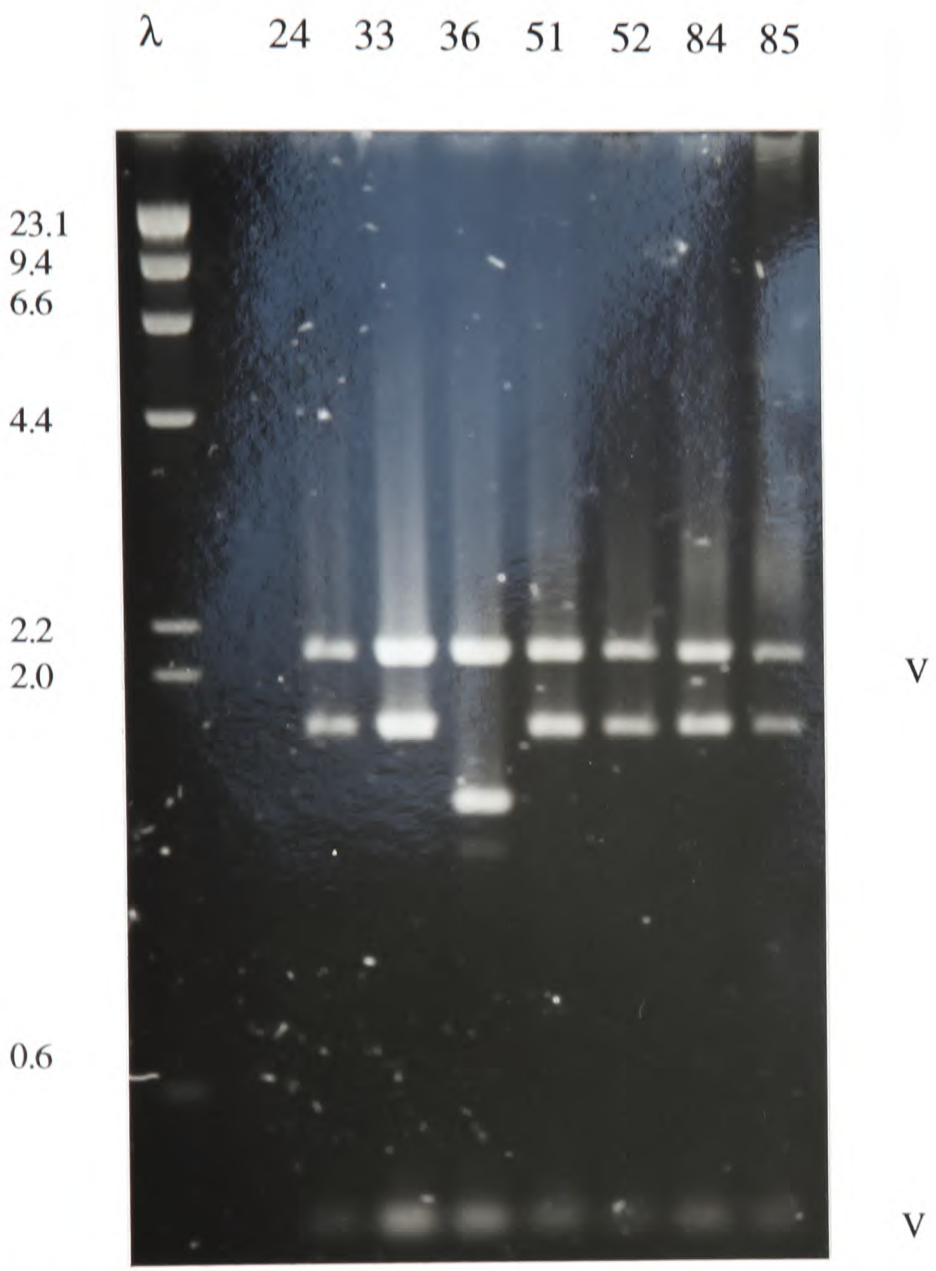
The gel was blotted and probed {2.7} with rat pER Man (-) insert, prepared by digesting with HindIII and purifying by gel electrophoresis {2.6.1}. It was important to use only the insert as a probe since pUC9 (the rat pER Man (-) vector) and pNB40 have sufficient similarity for cross hybridization. The insert was further purified by a second digest with DraI (which does not cut the insert but digests the vector into small fragments) and a second gel purification. Even after this process some vector contamination remained which hybridized to the vector bands (see Fig. 3.3b). The rat pER man (-) probe hybridized to all the cDNA inserts.

Figure 3.1. cDNA clones. These clones were identified with rat ER mannosidase and supplied by J. Hoheisel.

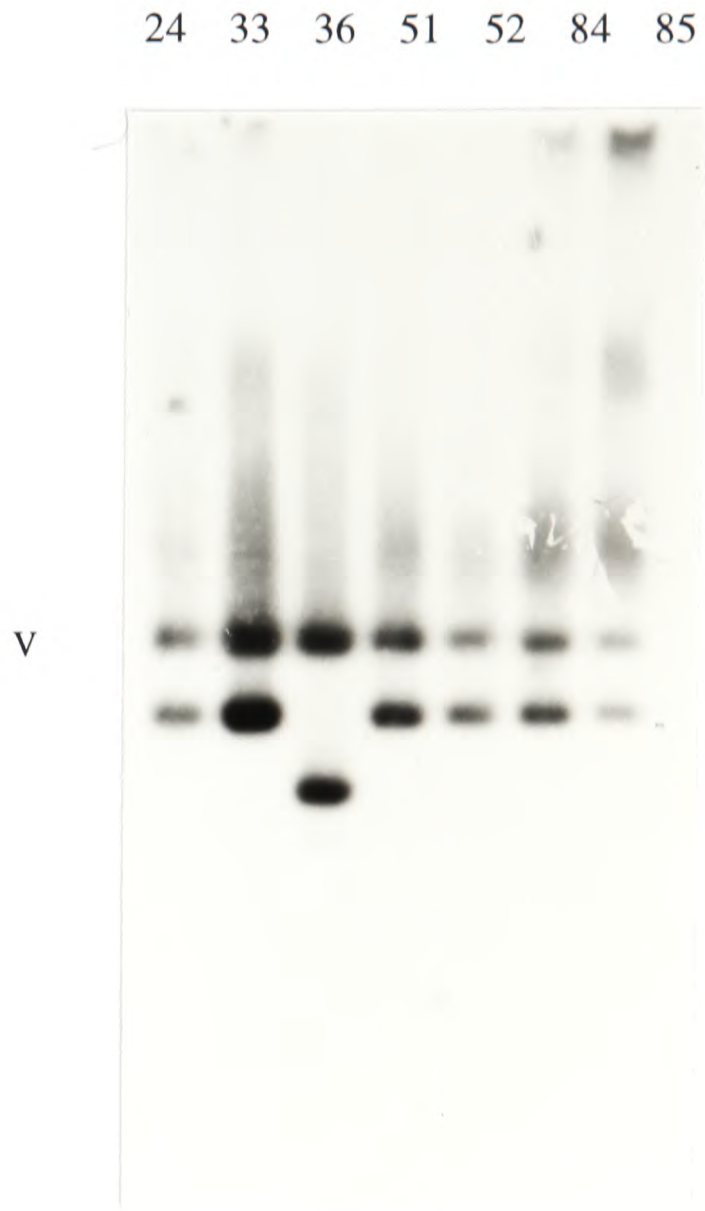
CODE	CLONE NAME
cDNA 24	cDNA1: 24 E 2
cDNA 33	cDNA1: 33 H 12
cDNA 36	cDNA1: 36 E 11
cDNA 51	cDNA1: 51 G 1
cDNA 52	cDNA1: 52 F 10
cDNA 84	cDNA1: 84 H 10
cDNA 85	cDNA1: 85 C 4

Figure 3.2. pNB40 - plasmid vector for cDNA clones.





a) cDNA clones digested with HindIII and EcoRI.



b) Probed with rat ER man (-) insert.

(Hybridization conditions 65°C. Washed 0.2xSSC, 65°C. Autoradiograph exposed 48hrs)

Figure 3.3 Initial investigation of cDNA clones (V= vector).

3.3 Restriction enzyme analysis

3.3.1 Clones to RE map

The two different clones obtained from this initial investigation (cDNA 33 & cDNA 36) were combined with two other clones (cDNA 4 and cDNA 6), which had been identified from a previous screen of the same embryonic library by J. Foster. These were investigated simultaneously by restriction enzyme analysis to see if any of the clones were the same or had overlapping regions. Preliminary investigation showed that the clones were cut by three restriction endonucleases; AvaI, PvuII and HaeIII.

3.3.2 Restriction digests of cDNAs

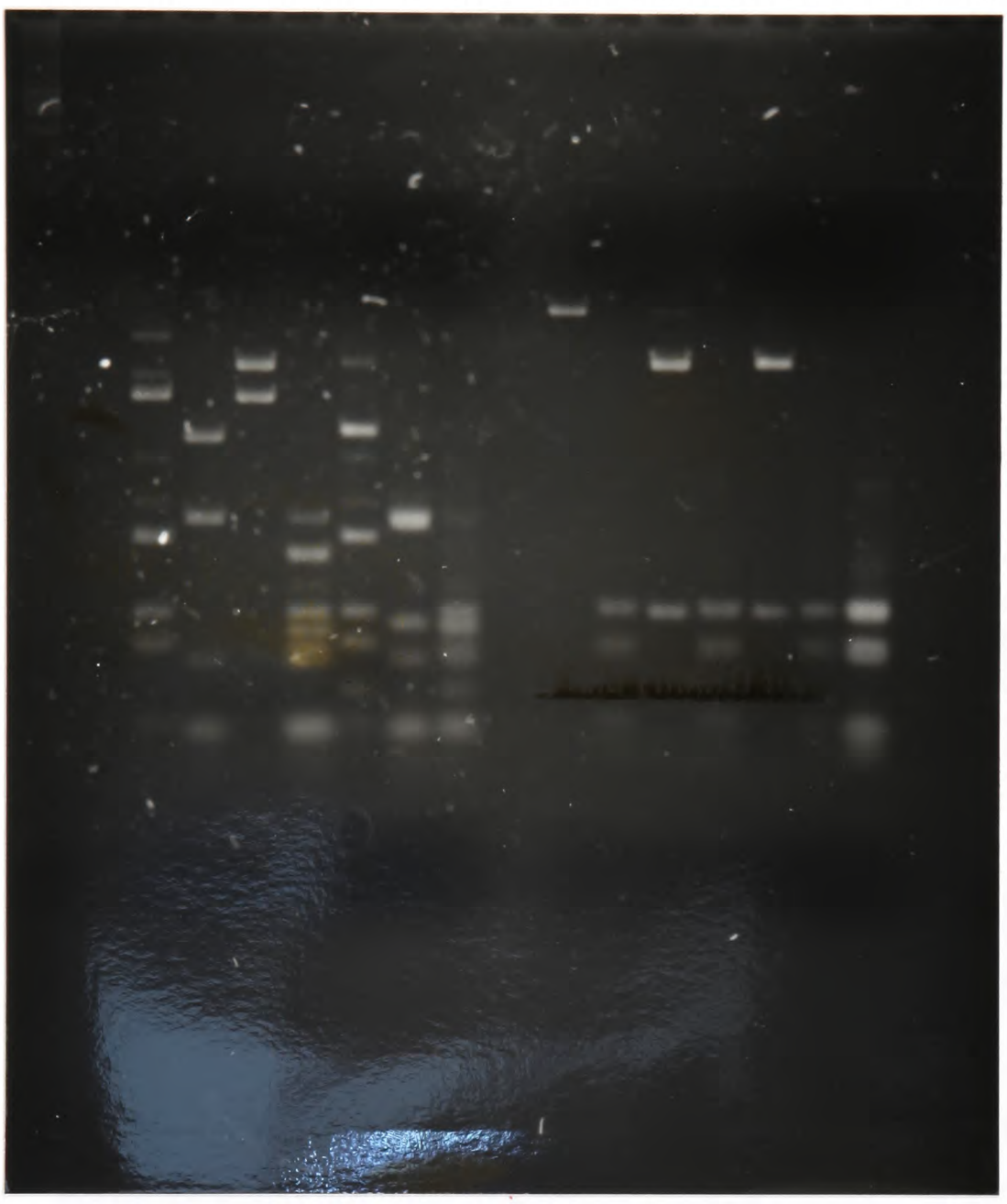
Purified inserts were prepared of all four clones {2.6.1}. The vector was removed by double digests with EcoRI and HindIII for 33, 36 and 6. cDNA 4 had a HindIII cut site within the insert (J. Foster) and was therefore digested with EcoRI and DraI which do not cut the insert. The purified inserts were digested with AvaI, PvuII and HaeIII and double and triple digests {2.4} were also carried out to produce restriction enzyme maps. The digests were run on 2.0% agarose gels {2.5}, to resolve the small fragments (Fig. 3.4).

3.3.3 End labelling inserts

To clarify the restriction enzyme map the inserts were cut out with HindIII and/or EcoRI and the ends radioactively labelled {2.8.2}. The labelled inserts were digested with HaeIII and cDNA 33 and cDNA 36 also with PvuII. HaeIII cuts at the 3' end of the insert in the small vector sequence that remained after EcoRI digestion (see Fig. 3.2) so that after HaeIII digestion a tiny 6bp fragment would be labelled and would disappear from the gel. This allowed the remaining labelled fragment from a HaeIII digestion to be identified as the 5' end of the insert. The digested fragments

Figure 3.4. Restriction digests. (A - AvaI, H - HaeIII, P - PvuII).

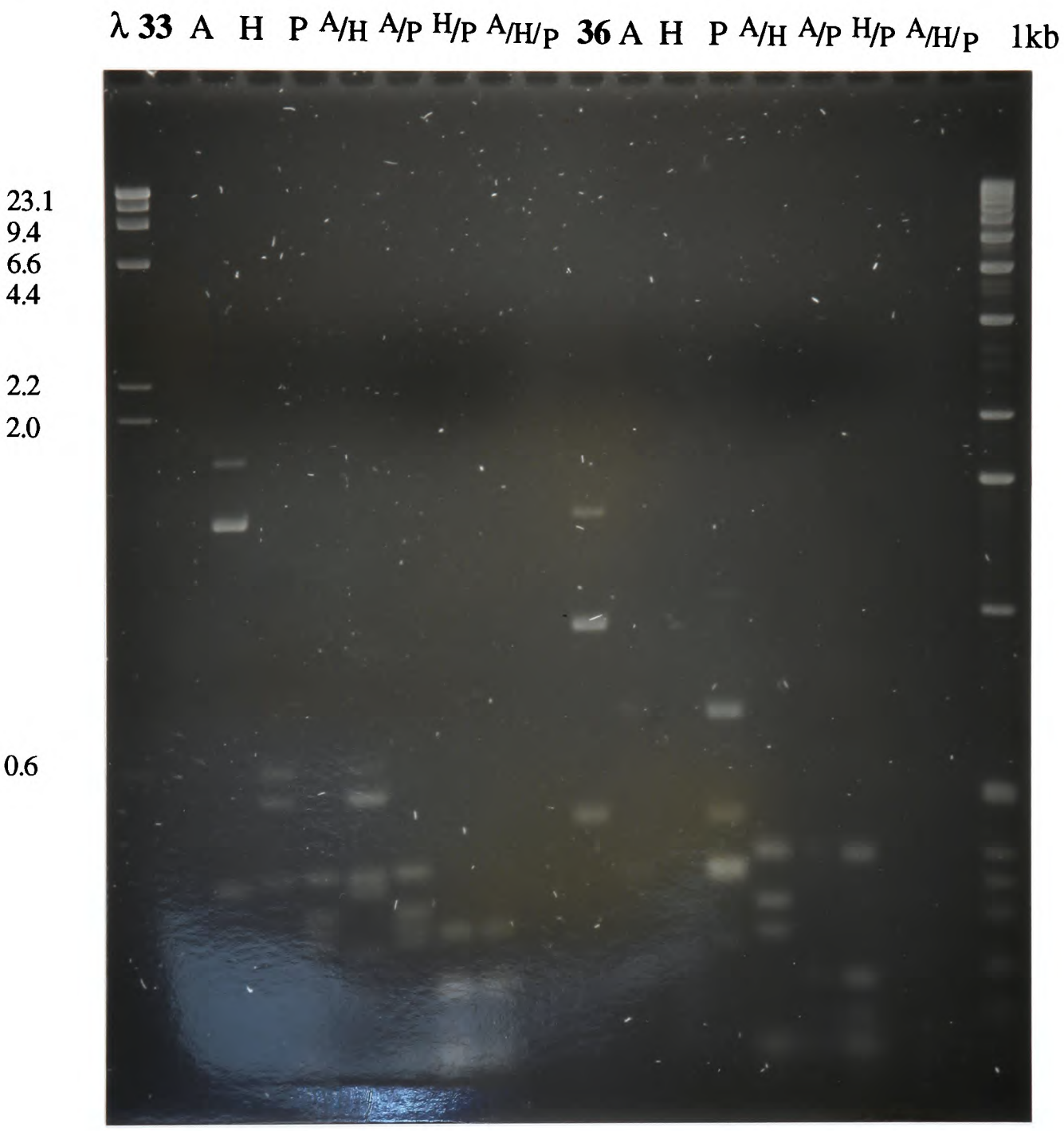
a) cDNA 4 and cDNA 6.



Gel run at 60V for 6hrs. Fragment sizes calculated on the basis of previously run gels.

Figure 3.4 (cont). Restriction digests. (A - AvaI, H - HaeIII, P - PvuII).

b) cDNA 33 and cDNA 36.



Gel run at 60V for 8hrs.

were separated on a gel (Fig. 3.5a) This was blotted and the labelled fragments identified by autoradiography {2.7.4} (Fig 3.5b). A partial HaeIII digest for cDNA 36 meant both ends remained labelled and two bands are seen on the autoradiograph. A similar gel with unlabelled digested inserts was blotted and probed with rat ER mannosidase {2.7} (Fig. 3.5c).

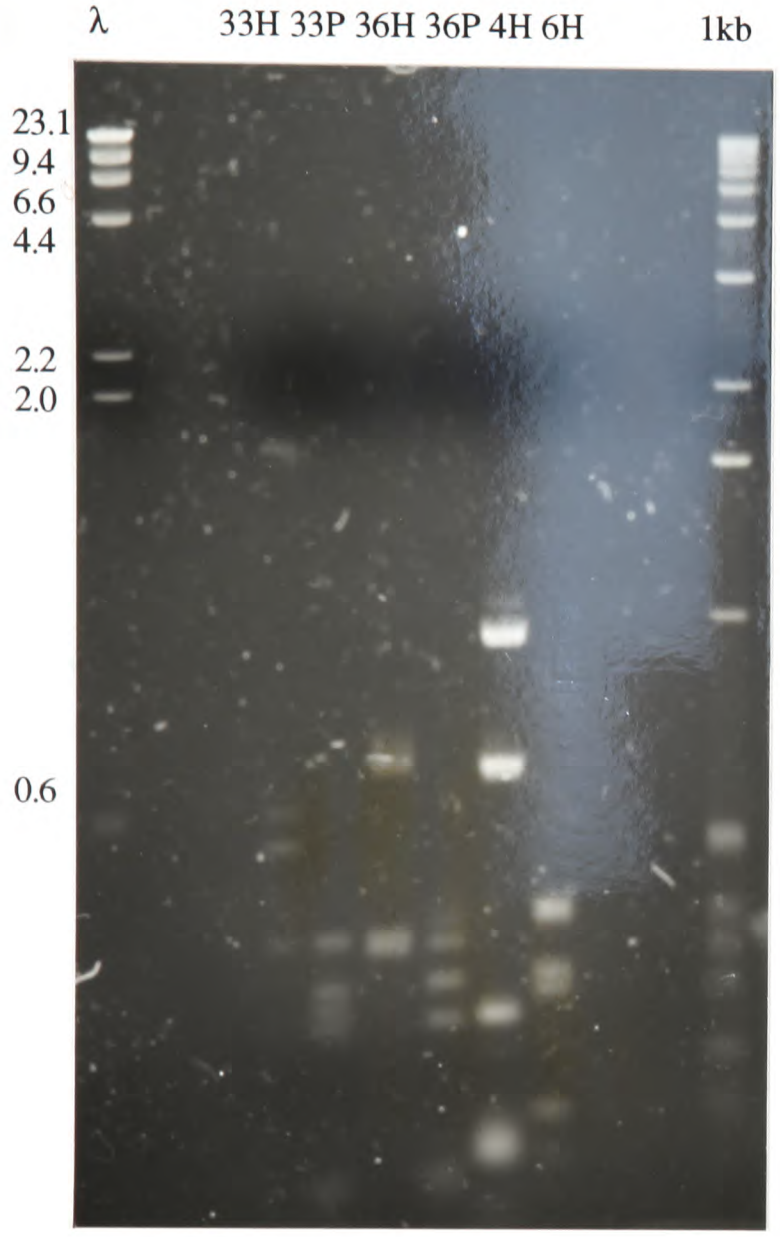
3.3.4 Restriction enzyme maps

Figure 3.6 summarises the fragment data obtained from the restriction enzyme digests. Restriction enzyme maps were calculated on the basis of these data (Fig. 3.7). The cDNAs 33 and 36 had many fragments in common suggesting that these clones overlap. There was some difficulty with AvaI digests which meant that these data were not available for some of the clones. A restriction map could not be drawn for cDNA 6. The use of enzymes that cut so frequently led to the production of many small fragments, which were hard to resolve on the electrophoresis gels and difficult to position on the map.

3.4 Selection of cDNA 4 for sequencing

3.4.1 Sub-cloning

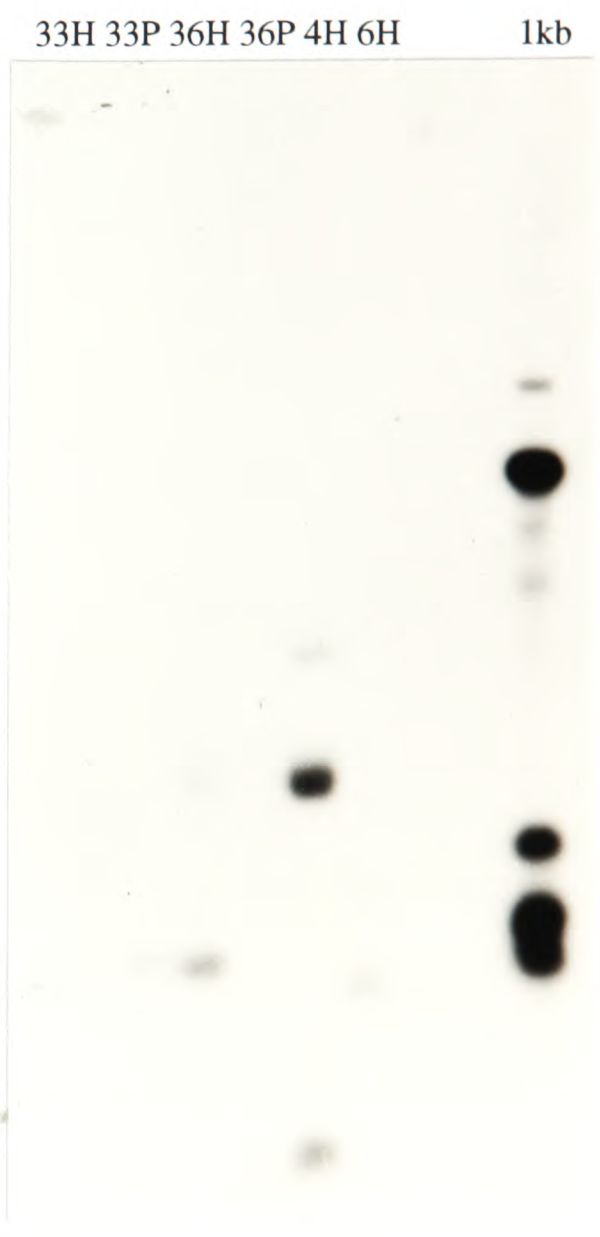
To decide which clone to investigate further the inserts were cut out of the plasmid with EcoRI and HindIII and run on three duplicate gels. These were blotted and probed with rat ER mannosidase and washed with increasing stringency between autoradiographs to test which clone hybridized most strongly (Fig. 3.8). Apart from the vector sequence, the rat ER mannosidase hybridized strongly to a ~0.78kb HindIII-HindIII fragment of cDNA 4 even when the filter was washed at 0.2 x SSC, 65^{o}C. The HindIII fragment is at the 5' end of the clone, a fragment that was also picked up when a HaeIII digest was probed with the rat ER mannosidase insert (Fig. 3.5c).



a) Electrophoresis gel.



b) end labelled fragments.



c) Digest probed with rat ER man (-).
(Filter washed at 2xSSC, 65oC.
Autoradiograph exposed 72hrs)

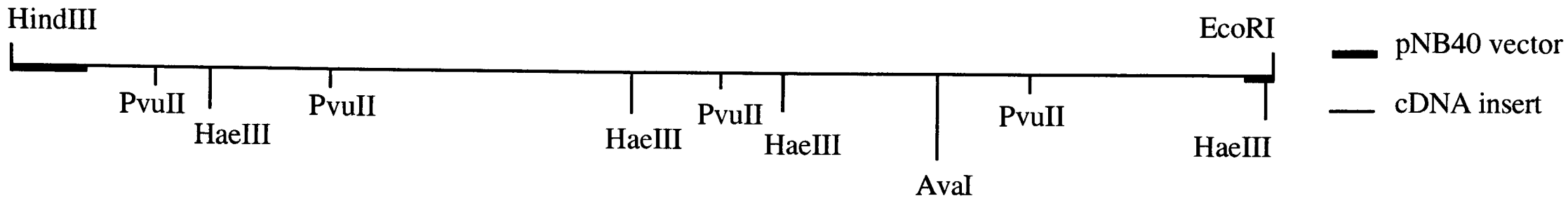
Figure 3.5. Clarifying the RE maps.

Enzyme	AvaI	HaeIII	PvuII	AvaI/HaeIII	AvaI/PvuII	HaeIII/PvuII	HaeIII/ AvaI/ PvuII
33	1445 335	545 520 355 130 (5' end)	355 230 190 (end) 150 90 (end)	520 355 325 150 130	355 230 190 150 90 *	180 135 90	280 225 170
36	970 450	660 300 150	355 (end) 230 150 100 (end)	660 300 150 *	355 230 150 100 *	300 130 100 90	300 130 100 90 *
4	1175 490 335 265 100	930 570 (5' end) 200 90	1410 1175	570 (5' end) 335 305 265 200 100	965 505 335 275 200 110	600 570 (5' end) 315 200 100	335 305 265 200 100
6	1995	350 275 (5' end) 220 135 120	1415 335	350 275 220 135 120	1415 335 *	350 275 180	350 275 180 *

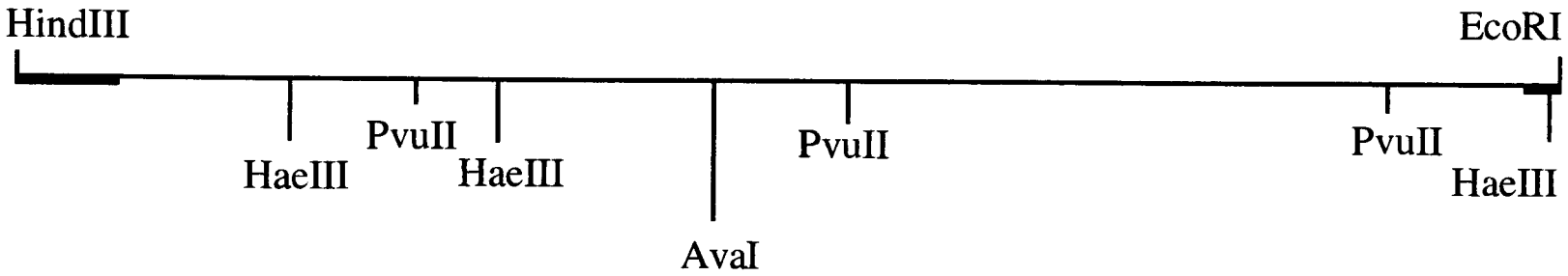
Figure 3.6. Restriction fragment data for the for cDNAs (see Figs. 3.4 and 3.5).

(* indicates where AvaI has not cut the cDNA).

cDNA 33



cDNA 36



cDNA 4

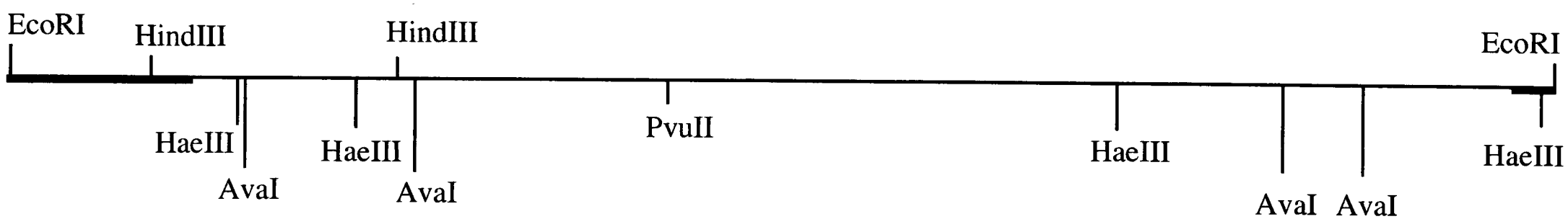
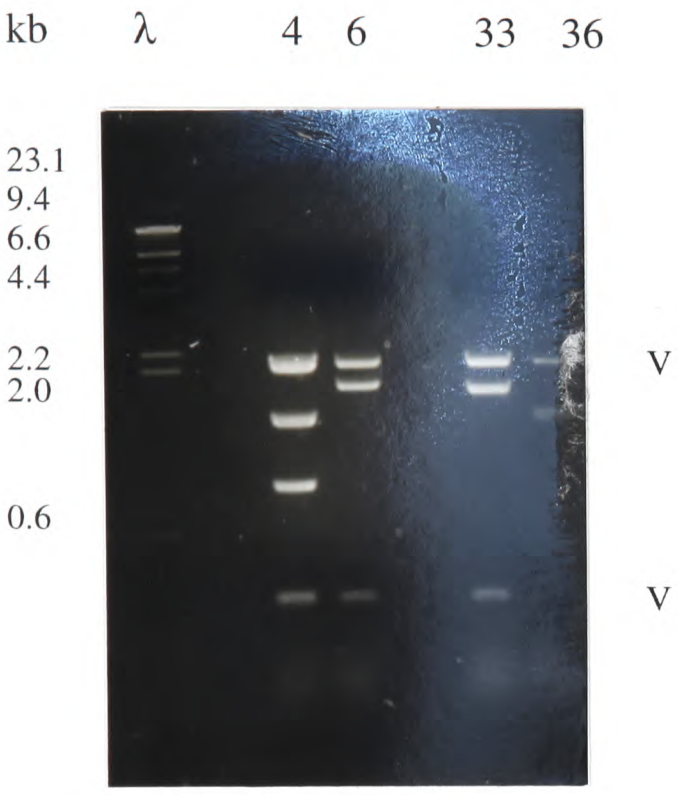
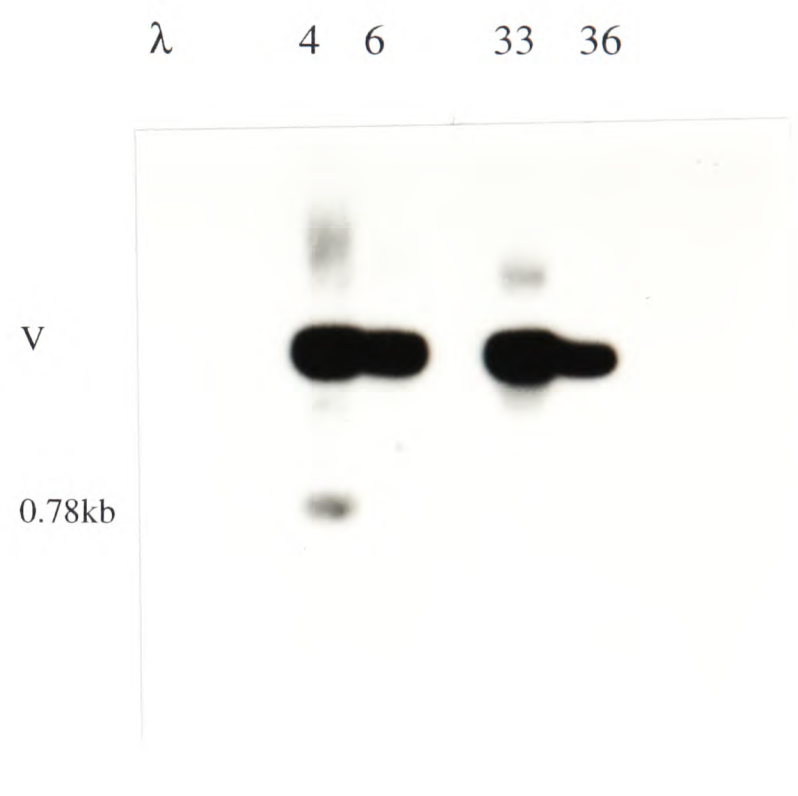


Figure 3.7. Restriction maps of cDNA 33, 36 and 4. Scale: 100mm = 1kb.



a) Electrophoresis gel showing all four cDNAs digested with HindIII and EcoRI.



b) Autoradiograph of blotted gel probed with rat ER mannosidase insert. (Filter washed to 0.2xSSC, 65^{o}C).

(V= vector band).

Figure 3.8. Which cDNA clone to investigate further.

The 0.78kb fragment was subcloned into the HindIII site of pUC18 {2.11} (Fig. 3.9).

3.4.2 *In situ* hybridization

The cDNA 4 clone was used as a probe for *in situ* hybridization to polytene chromosomes (Fig. 3.10). It mapped to 57D/E.

3.5 Sequence Analysis

The 0.78kb subclone of cDNA 4 was sequenced {2.10}. Four primers were used; universal and reverse primers from pUC vectors and two synthesized 17mer primers on the normal and reverse strands (Fig. 3.11). The sequence obtained is shown in Figure 3.12.

3.5.1 Refined RE map of cDNA 4

From the sequence obtained the restriction enzyme sites were checked (Fig. 3.13). The sites shown in bold are definite - the other sites are possible given the presence of ambiguous nucleotides in the sequence. Since the analysis of the data showed no homology with the rat ER mannosidase (see section 3.5.2) it was not thought worthwhile to resolve these ambiguities.

3.5.2 cDNA 4 sequence compared to rat ER mannosidase

This was compared directly to the rat ER mannosidase sequence at the nucleotide level, using the BESTFIT sequence alignment program (P. Haeberli, GCG version 1.0, Philip Delaquess, GCG version 4.0) of the University of Wisconsin Genetics Computer Group (Fig. 3.14). This showed 63.1% similarity (because some of the bases were unresolved) and 53.8% identity over only 65 base pairs. If the sequences were homologous a much longer region of similarity would be expected. It

Figure 3.9. pUC18. This plasmid vector contains ORI - origin of replication, *Amp* - ampicillin resistance gene, allowing selection of bacteria containing the plasmid and *lacZ*, coding for β-galactosidase, interupted by the mutiple cloning site with the restriction sites shown in the diagram, allowing the selection of recombinant plasmids (Yanisch-Perron, 1985).

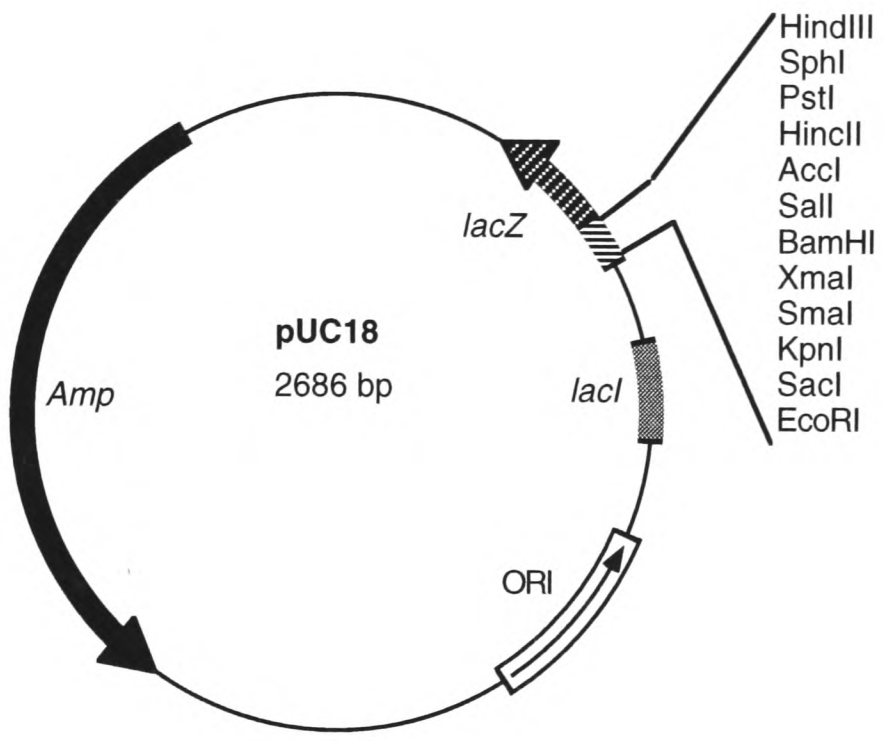
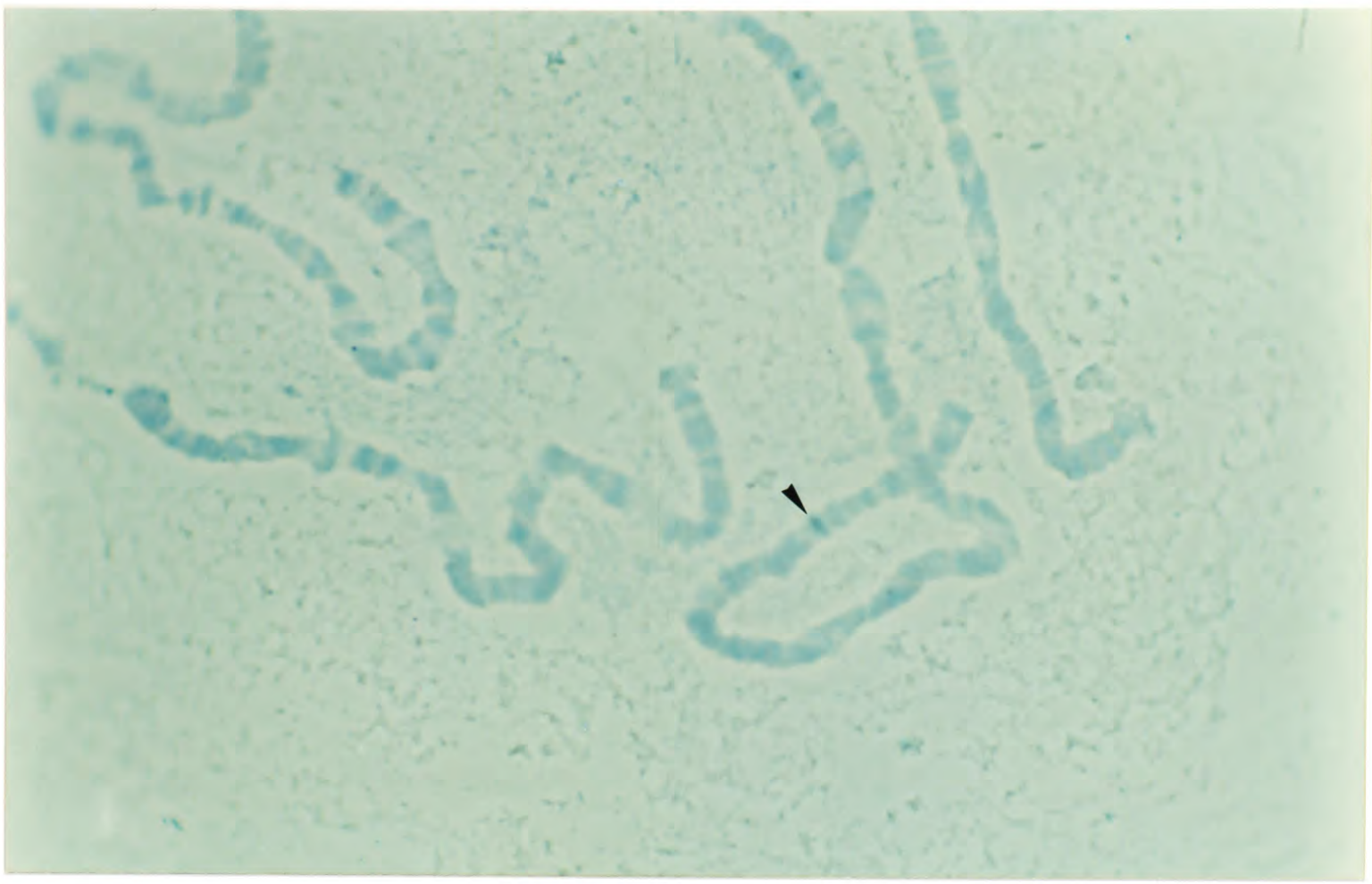


Fig 3.10. *In situ* hybridization of cDNA 4.



The hybridized band is clearer in this, and other cDNA 4 *in situ* hybridized squashes, when viewed down the microscope (63x objective.)

Figure 3.11. Primers for sequencing the 0.78kb HindIII fragment of cDNA 4.

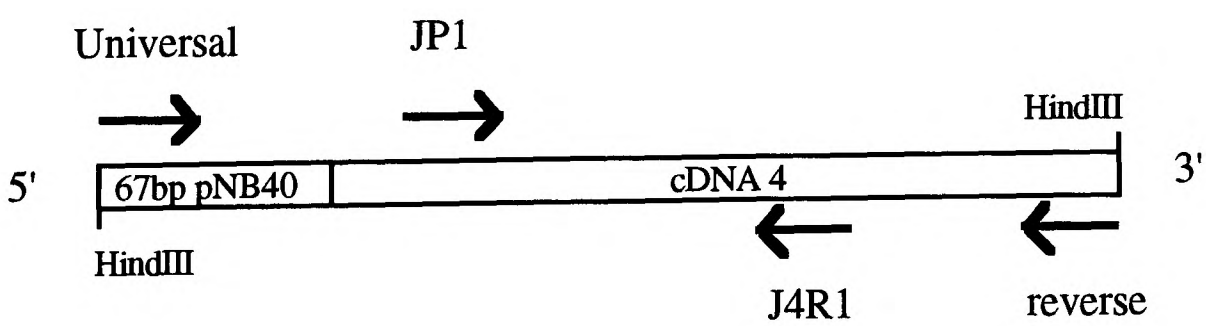


Figure 3.12. Sequence of cDNA 4 HindIII fragment (pNB40 sequence not shown). (One strand only sequenced)

```
0     GAGGCGAAAA GGAAGCACAG AACACAGTAG GCAGCGCAGC AGAGAAAATG
50    TATAATTTAC ATTAATTTAA TTACAAATAG TTCGTCGTGC CAGATCTCGA
100   TTCTTGTTAT TCCGCAACAG ACCGCAGTGG GGGAAAACTG TTGAAAAGCA
150   CAGCGCGCAC AAGTGAGGTC ACCACCTGGT TTTCCGCGGA TACTAAATCC
200   CAGCCAGGCA GCCAGCGAGN GAGCGAAAAG TGGGTGAAAA TGGCGTGGAA
250   AATGGGAGGA AACCGGGCCG CAACGCTTGC NNNNCTCTTC GCCAGCAGTT
300   TGCTGCTGCT GAGTGCNGCC AATGCCGCCG ATNTGGACAC GGAGTCGGAC
350   GACTTCGAGG ATGGTTATGT GGAGGATGTG AGGAGGAGCN NNNNTTCATC
400   GGTGGCGATG AGAAGCTGGC CTACGAGAGC CCGGTTATCG ATGCCAAGAA
450   GTTCCACTTC GCGNATCACT TCGACGACGT GGAGGAGTCC AGGAAGAGGT
500   GGGTGCTGTC GCAGGCCAAG AAGGACGAAA TCGCCGAGGA GATCTCCAAG
550   TACGACGGCA TCTGGAACTG GGAGTCGCCG AAGCGNATCG TTTGGGCTAC
600   ACTTGGGTCT GGTGCTCAAG TCGAAGGCNA AACACGNNGC MATAGCCGCC
650   CCMMTCMMCA AACCGTTCGA GTTCAAGTCG GACAAGCCGC TGGTGGTGCA
700   GTACGAGGTC ACGTTGCAGG AGGGTCAGGA GTGCGGTGNC TCGTATCTGA
```

Figure 3.13. RE map of cDNA 4 750bp fragment

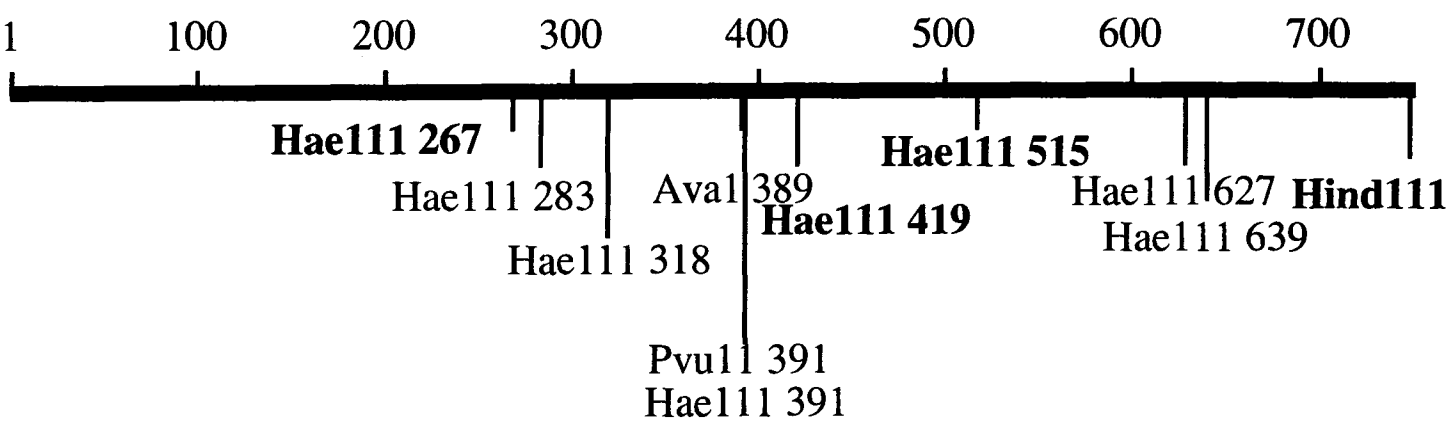
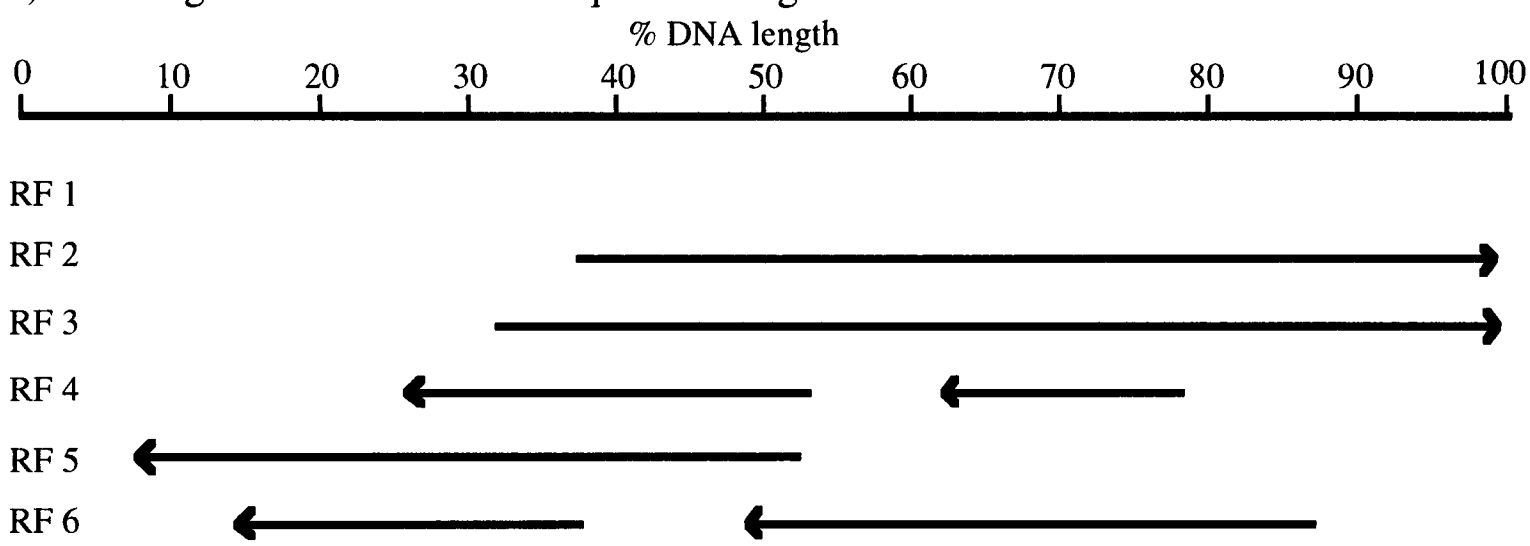


Figure 3.14. cDNA 4 HindIII fragment nucleotide sequence compared to the total rat ER mannosidase nucleotide sequence (RNMANER RNA; ROD; 3360bp. Accession No. M57547; J05632). (The values generated were from the University of Wisconsin, Genetics Computer Group BESTFIT program).

Quality: 19.4	Length: 65	Percent Similarity: 63.077
Ratio: 0.298	Gaps: 0	Percent Identity: 53.846

Figure 3.15. cDNA 4 sequence translated and compared to Rat ER mannosidase.

a) Reading frames for cDNA4 sequenced fragment.



b) Peptide sequence. cDNA 4 sequence translated in all reading frames (see text) and compared to the predicted rat ER mannosidase peptide sequence (MAN1-RAT PRT; 1040aa. Accession No. P21139).

Reading frame	1	2	3	4	5	6
Quality	75.5	79.1	81.4	72.5	72.9	82.6
Length	294	283	298	267	330	284
Ratio	0.306	0.320	0.328	0.291	0.298	0.334
Gaps	6	8	10	8	7	3
% Similarity	38.528	38.430	39.184	34.410	36.249	35.581
% Identity	17.316	18.595	16.735	16.192	18.331	13.763

was also investigated for open reading frames (Fig. 3.15a) and translated in all six, since the sequence ambiguities may have lost or gained start and stop codon, or led to frame shifts within the sequence. These predicted peptide sequences were compared with the ER α-mannosidase protein sequence (Fig. 3.15b). Again there was no match since less than 25% identity is not considered significant.*

3.5.3 FASTA database search (1993, 1995)

A search of the GenEMBL database, using the FASTA program, [Pearson *et al.* 1988] in GCG, failed to identify the rat ER mannosidase or any other mannosidase. The FASTA search was repeated in 1995 while preparing the data for this thesis. The 10 most closely matching sequences are shown in Figure 3.16a. Searching the SwissProt database with the translated peptide sequences also identified no mannosidases (Fig. 3.16b). There were also no other significant matches which suggests that this cDNA has not been previously sequenced. The RF2 peptide does identify human and dog calnexin and RF3 the *S. pombe* homologue of calnexin with higher optimised scores than the other identified protein sequences. This score is calculated on the basis of joining together matching sections of sequence and indicates a region of homology. The human and dog calnexin sequences were present in the database in 1993, but this homology was not considered significant or of immediate interest as the function of calnexin as a carbohydrate binding protein (see section 1.4.1) was only recently identified [Jackson *et al.* 1994].

3.5.6 BLAST database search (1995)

While preparing the sequence data for this thesis in 1995, the BLAST database search program [Gish *et al.* 1990; Gish *et al.* 1993] was used which had not been available on the Oxford Pathology molecular biology VAX GCG version 7.0 in 1993. This program has the advantage of allowing a protein database search with a nucleotide

---

* The BESTFIT program optimises the alignment between sequences by leaving gaps. Since all of the translated sequences have similar scores, none can be considered a significant match.

Figure 3.16. FASTA results using the 0.75kb sequence of cDNA to search databases.

a) Nucleotide sequence. The GenEMBL database was searched with cDNA 4 sequenced fragment.

			init1	initn	opt
Gbnew:Scchrix_2	Z47047	S. cerevisiae chromosome IX comp.	64	178	65
Emfun:Scchrix-2		Continuation of SCCHRIX from base 350001.	64	178	65
Emnew:Scchrix-2	Z38125	S. cerevisiae chromosome IX comp.	64	178	65
Empro:Ecuw82 /rev	L10328	E.coli genomic region from 81.5.	74	151	74
Emnew:Ecuw82 /rev	L10328	E. coli; the region from 81.5..	74	151	74
Empln:Hvcrh2a	L27349	Hordeum vulgare calreticulin (CRH2.	96	144	99
Empln:Hvcrh1a	L27348	Hordeum vulgare calreticulin (CRH1.	96	144	99
Emnew:Cer144	U23515	Caenorhabditis elegans cosmid R144..	66	143	75
Emrod:Mmcy03 /rev	X01682	Mouse gene - cytochrome P3-450..	66	142	81
Emvrl:Hehcmvcg	X17403	Human Cytomegalovirus Stra. AD16....	67	142	67

b) Peptide sequence. The swissProt database was searched with the predicted peptide sequence in all reading frames.

i) Reading frame 1.

			init1	initn	opt
Swissprot31:Plec_Rat	P30427	PLECTIN	44	118	75
Swissprot31:Arga_Psepu	P32042	AMINO-ACID ACETYLTRANSFE	46	108	49
Swnew:Polg_Ppvra	P17767	GENOME POLYPROTEIN	63	96	63
Swnew:Polg_Ppvra	P17767	GENOME POLYPROTEIN	63	96	63
Swnew:Polg_Ppvd	P13529	GENOME POLYPROTEIN	61	94	61
Swnew:Polg_Ppvd	P13529	GENOME POLYPROTEIN	61	94	61
Swissprot31:Ocs1_Maize	P24068	OCS-ELEMENT BINDING FACT	64	92	65
Swnew:Polg_Ppvna	P17766	GENOME POLYPROTEIN	59	92	61
Swnew:Polg_Ppvna	P17766	GENOME POLYPROTEIN	59	92	61
Swissprot31:Vnua_Prvka	P33485	PROBABLE NUCLEAR ANTIGEN	51	92	59

ii) Reading frame 2.

			init1	initn	opt
Swissprot31:Trhy_Human	Q07283	TRICHOHYALIN	67	109	70
Swissprot31:U108_Hsv11	P10192	PROTEIN UL8	58	106	73
Swissprot31:Pol_Eiavy	P03371	POL POLYPROTEIN (PROTEASE	63	105	63
Swissprot31:Pol_Eiav9	P11204	POL POLYPROTEIN (PROTEASE	63	103	63
Swissprot31:Pol_Siva1	P27973	POL POLYPROTEIN (PROTEASE	65	103	67
Swissprot31:Pol_Eiavc	P32542	POL POLYPROTEIN (PROTEASE	63	103	63
Swissprot31:Calx_Human	P27824	CALNEXIN PRECURSOR (MAJOR	103	103	125
Swissprot31:Yk23_Yeast	P36136	HYPOTHETICAL 31.0 KD PROT	61	103	62
Swissprot31:Polg_Hcvj7	P27961	GENOME POLYPROTEIN (CONTA	50	102	51
Swissprot31:Calx_Canfa	P24643	CALNEXIN PRECURSOR (PP90)	102	102	124

Figure 3.16 (cont) FASTA results.

iii) Reading frame 3

			init1	initn	opt
Swissprot31:Polg_Bvdvs	Q01499	GENOME POLYPROTEIN	51	118	51
Swissprot31:Tegu_Hsv11	P10220	LARGE TEGUMENT PROTEIN	63	108	63
Swissprot31:Calx_Schpo	P36581	CALNEXIN HOMOLOG PRECURSo	81	104	116
Swissprot31:Mapa_Rat	P34926	MICROTUBULE-ASSOCIATED PR	53	104	65
Swissprot31:Prpx_Horvu	P16273	PATHOGEN-RELATED PROTEIN	100	100	101
Swissprot31:Plec_Rat	P30427	PLECTIN	55	100	74
Swissprot31:Angr_Viban	P19828	ANGR PROTEIN	48	100	48
Swissprot31:Mosa_Maize	P15268	AUTONOMOUS TRANSPOSABLE E	60	98	62
Swissprot31:Cpsf_Chick	P16419	C-PROTEIN, SKELETAL MUSCL	42	97	46
Swissprot31:Ppol_Bovin	P18493	NAD(+) ADP-RIBOSYLTRANSFE	48	95	49

iv) Reading frame 4.

			init1	initn	opt
Swnew:Rpol_Eav	P19811	POL POLYPROTEIN	44	104	44
Swnew:Rpol_Eav	P19811	POL POLYPROTEIN	44	104	44
Swissprot31:Ygy3_Halsq	P21561	HYPOTHETICAL 50.6 KD PROT	61	104	71
Swissprot31:X169_Human	P41229	XE169 PROTEIN	38	100	39
Swissprot31:Ntch_Rat	P28570	SODIUM-DEPENDENT CHOLINE	41	98	41
Swissprot31:Ntcr_Rabit	P31661	SODIUM- AND CHLORIDE-DEPE	41	98	41
Swissprot31:Poln_Rubvt	P13889	NONSTRUCTURAL POLYPROTEIN	37	91	52
Swissprot31:Yj12_Yeast	P39526	HYPOTHETICAL 229.9 KD PRO	45	90	54
Swissprot31:Yks5_Yeast	P34233	HYPOTHETICAL 65.7 KD PROT	44	90	48
Swissprot31:Per_Artra	P28313	PEROXIDASE (EC 1.11.1.7)	48	85	49

v) Reading frame 5.

			init1	initn	opt
Swnew:Kr2_Hsveb	P28966	PROBABLE SERINE/THREONINE	66	95	75
Swnew:Kr2_Hsveb	P28966	PROBABLE SERINE/THREONINE	66	95	75
Swissprot31:Yihn_Ecoli	P32135	HYPOTHETICAL 46.3 KD PROT	66	91	68
Swissprot31:Cpd6_Human	P10635	CYTOCHROME P450 IID6 (EC	45	90	50
Swissprot31:Ylh8_Caeel	P34362	HYPOTHETICAL 22.7 KD PROT	68	85	71
Swnew:Acvs_Nocla	P27743	DELTA-(L-ALPHA-AMINOADIPY	45	82	45
Swissprot31:Cina_Eleel	P02719	SODIUM CHANNEL PROTEIN	50	82	50
Swnew:Acvs_Nocla	P27743	DELTA-(L-ALPHA-AMINOADIPY	45	82	45
Swissprot31:Ul09_Hsveb	P28947	ORIGIN OF REPLICATION BIN	37	81	38
Swissprot31:Bind_Arbpu	P24608	BINDIN PRECURSOR	42	81	51

vi) Reading frame 6.

			init1	initn	opt
Swnew:Cc2_Caeel	P34556	CELL DIVISION CONTROL PRO	62	91	62
Swnew:Cc2_Caeel	P34556	CELL DIVISION CONTROL PRO	62	91	62
Swissprot31:Olf7_Rat	P23270	OLFACTORY RECEPTOR-LIKE P	56	90	69
Swissprot31:Tra_Pseae	P06695	TRANSPOSASE FOR TRANSPOSO	46	88	46
Swissprot31:Nu5c_Vicfa	P15958	NADH-PLASTOQUINONE OXIDOR	37	84	42
Swissprot31:Pol_Htlv2	P03363	POL POLYPROTEIN (REVERSE	82	82	91
Swissprot31:Citb_Ecoli	P05853	CITB PROTEIN	42	81	57
Swnew:7les_Drome	P13368	SEVENLESS PROTEIN (EC 2.7	42	81	53
Swnew:7les_Drome	P13368	SEVENLESS PROTEIN (EC 2.7	42	81	53
Swnew:Acvt_Pench	P26046	DELTA-(L-ALPHA-AMINOADIPY	41	80	46

sequence; it automatically translates the sequence in all six reading frames and looks for similarity with amino-acid sequences in the data base. This allows it to combine different reading frames from the query sequence in comparison to any single database sequence, particularly important where ambiguities in the query sequence may have shifted the reading frame.

The results from the search of the SwissProt protein database with the 750bp sequence data from cDNA 4 are shown in Figure 3.17a. cDNA 4 shows highly significant similarity with six calnexin or calnexin homologue proteins. The sequence alignment, which the BLAST program provides, shows that all the calnexin amino acid sequences align to the same amino-acid regions of cDNA 4. Figure 3.17b is an example of the BLAST output showing these amino-acids sequences aligned to human calnexin.

3.6 Is cDNA 4 the *Drosophila* homologue of Calnexin?

3.6.1 Calnexin

Calnexin is a membrane bound protein of the ER. It is thought to act as a molecular chaperone in glycoprotein folding. Calnexin was identified as a calcium binding protein in the ER membrane [Wada *et al.* 1991]. It was originally shown to bind MHC I molecules in the ER [Degen *et al.* 1991], but it has since been demonstrated to associate transiently with many different newly synthesized glycoproteins [Ou *et al.* 1993; David *et al.* 1993]. One of its functions is to retain incorrectly or incompletely folded glycoproteins as part of the ER quality control mechanism [Bergeron *et al.* 1994]. It is associated with at least three other membrane proteins [Wada *et al.* 1991] the roles of which are not known.

Calnexin initially binds to the carbohydrate chains of glycoproteins possibly to allow the polypeptide conformational and steric freedom to fold and interact with other chaperones in the ER [Hammond *et al.* 1994]. Evidence for carbohydrate binding lies

Figure 3.17. BLAST results. The Swissprot database was searched using the GCG BLAST program.

a) Table of identified proteins.

Sequences producing High-scoring Segment Pairs:		Frame	Score	P(N)	N
sp|P27824|	CALX_HUMAN CALNEXIN PRECURSOR.....	+2	91	2.1e-16	3
sp|P35564|	CALX_MOUSE CALNEXIN PRECURSOR.	+2	88	5.6e-16	3
sp|P35565|	CALX_RAT CALNEXIN PRECURSOR.	+2	88	5.6e-16	3
sp|P24643|	CALX_CANFA CALNEXIN PRECURSOR (PP90).	+2	88	5.7e-16	3
sp|P34652|	CALX_CAEEL CALNEXIN HOMOLOG PRECURSOR	+2	79	9.8e-14	3
sp|P36581|	CALX_SCHPO CALNEXIN HOMOLOG PRECURSOR	+3	73	0.00011	3
sp|P16273|	PRPX_HORVU PATHOGEN-RELATED PROTEIN.	+3	86	0.0033	1
sp|Q02817|	MUC2_HUMAN MUCIN 2 (INTESTINAL MUCIN	-1	63	0.066	3
sp|P13055|	E75C_DROME ECDYSONE-INDUCIBLE PROTEIN	-1	62	0.12	2
sp|P32323|	AGA1_YEAST A-AGGLUTININ ATTACHMENT SU	-3	56	0.20	2
sp|P13983|	EXTN_TOBAC EXTENSIN PRECURSOR (CELL W	-1	55	0.24	3
sp|P24152|	EXTN_SORVU EXTENSIN PRECURSOR (PROLIN	-1	47	0.36	3
sp|P29402|	CALX_ARATH CALNEXIN HOMOLOG PRECURSOR	+3	71	0.44	1
sp|P36783|	VE2_HPV14 E2 PROTEIN.	+3	52	0.53	3
sp|P13729|	SGS3_DROSI SALIVARY GLUE PROTEIN SGS-	+3	68	0.73	1

b) BLAST significant alignments. The significant human calnexin alignment is given as typical of the next 5 calnexin homologues.

```
P27824:CALX_HUMAN CALNEXIN PRECURSOR (MAJOR HISTOCOMPATIBILITY
COMPLEX).  Length = 592

  Plus Strand HSPs:

 Score = 91 (41.9 bits), Expect = 2.1e-16, Sum P(3) = 2.1e-16
 Identities = 16/28 (57%), Positives = 21/28 (75%), Frame = +2

Query:   500 WVLSQAKKDEIAEEISKYDGIWNWESPK 583
             W+LS+AKKD+  +EI+KYDG W  E  K
Sbjct:    83 WILSKAKKDDTDDEIAKYDGKWEVEEMK 110

 Score = 80 (36.8 bits), Expect = 2.1e-16, Sum P(3) = 2.1e-16
 Identities = 14/22 (63%), Positives = 17/22 (77%), Frame = +3

Query:   684 KPLVVQYEVTLQEGQECGXSYL 749
             KPL+VQYEV  Q G ECG +Y+
Sbjct:   144 KPLIVQYEVNFQNGIECGGAYV 165

 Score = 71 (32.7 bits), Expect = 2.1e-16, Sum P(3) = 2.1e-16
 Identities = 16/30 (53%), Positives = 20/30 (66%), Frame = +3

Query:   606 GLVLKSKAKHXAIAAPXXKPFEFKSDKPLV 695
             GLVL S+AKH AI+A   KPF F +   +V
Sbjct:   119 GLVLMSRAKHHAISAKLNKPFLFDTKPLIV 148
```

in blocking of binding by tumicamycin, a glycosylation inhibitor [Jackson *et al.* 1994]. The protein binds to $Glc_1Man_9GlcNAc_2$ [Ware *et al.* 1995]. Hammond *et al.* [1994] suggest a mechanism that involves the previously identified enzyme UDP-glucose: glycoprotein glucosyltransferase which recognises $Man_{9-7}GlcNAc_2$ oligosaccharide chains only when associated with incorrectly folded proteins [Sousa *et al.* 1992]. It adds a glucose to the end of the chain thus forming the required substrate for calnexin. This allows the N-linked oligosaccharides to serve as signals for the glycopolypeptide's folding status (Fig. 3.18).

In some cases preventing the binding of a glycoprotein to calnexin with glycosylation inhibitors (tunicamycin or castenospermine) appears to have no effect and the protein is capable of folding without attached oligosaccharides [Elbein 1991b] and therefore without the aid of calnexin [Hammond *et al.* 1994]. However the *S. pombe* calnexin homologue has been shown to be essential for yeast viability [Jannatipour *et al.* 1995]. After the initial binding, calnexin interacts with the polypeptide and the subsequent removal of associated oligosaccharides does not disrupt this binding [Zhang *et al.* 1995].

The structure of calnexin has been elucidated to a certain extent [Wada *et al.* 1991] and appears similar in most of the species homologues so far identified. The first 20 amino-acids make up a signal sequence. The next 570 amino-acids make up a potential luminal sequence containing four subdomains with sequence similarities to another calcium dependent ER protein, calreticulin. A potential transmembrane domain has been identified between amino-acids 482-502 and the remaining 90 amino-acids make up the cytosolic sequence including an ER retention motif (RKPRRE) [Jackson *et al.* 1990].

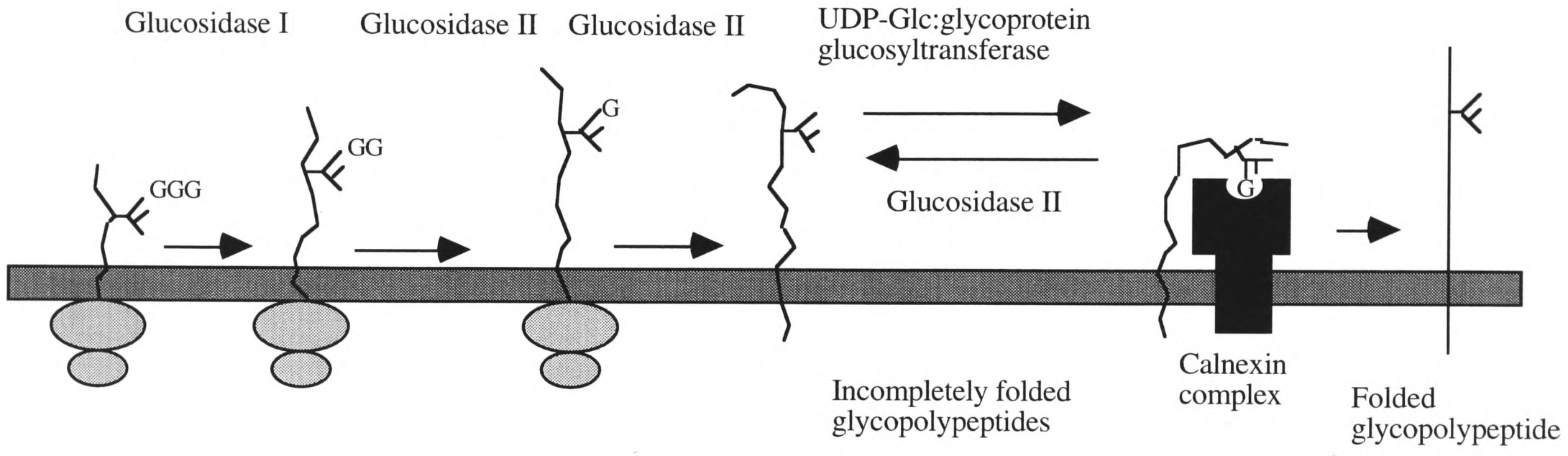


Figure 3.18. Mechanism for detecting the folding status of glycoproteins and their interaction with calnexin (from Hammond, 1994). (G = Glucose).

3.6.2 Calnexin in other species

Calnexin or calnexin homologues have been identified in several mammalian species (humans, mice and rats [Tjoelker *et al.* 1994], dogs [Wada *et al.* 1991]), *X. laevis* [Treves *et al.* 1992], *C. elegans* [Wilson *et al.* 1994], *S. mansoni* [Hawn *et al.* 1993], *A.thaliana* [Huang *et al.* 1993; Boyce *et al.* 1994], *S. cerevisiae* [de Virgilio *et al.* 1993; Parlati *et al.* 1995] and *S. pombe* [Jannatipour *et al.* 1995]. There has been no previous report of a *Drosophila* homologue of calnexin.

3.6.3 cDNA 4 compared to calnexin in other species

The cDNA 4 nucleotide sequence was altered (by removing an N at 586) to combine the reading frames 2 and 3 that identified calnexin. Aligning this altered sequence with the human calnexin sequence (Fig 3.19) showed 39.9% identity between the sequences and 54.4% similarity. These figures are significant, particularly as most of the similarity is towards the end of the sequenced fragment where the conserved sequences are found. The sequenced fragment of cDNA 4 corresponds to part of the first calnexin specific sequence section and the start of the calreticulin/calnexin common sequence. A multiple amino-acid sequence alignment using the PILEUP program from GCG (a simplification of the progressive alignment method of [Feng *et al.* 1987] and similar to that used by [Higgins *et al.* 1989]) is shown in Figure 3.20. Interestingly although cDNA 4 is missing some of the amino-acids found at the start of the homologous region of the protein in mammals, the start of homology with cDNA 4 coincides directly with that found for *C. elegans* and *S. pombe*.

3.6.4 A calreticulin homologue in *D. melanogaster*

Calnexin shares a protein domain with a protein called calreticulin. Calreticulin is a second ER localized Ca^{2+} binding protein, which is, however, not membrane spanning. A *Drosophila* homologue of calreticulin has been identified [Smith 1992]. It

Figure 3.19. BESTFIT sequence alignment of the modified cDNA 4 amino acid sequence and human calnexin (SwissProt: P27824).

```
           Quality:   96.6          Length:    171
             Ratio:  0.589           Gaps:      6
Percent Similarity: 54.430  Percent Identity: 39.873


 86 RKPGRNACXXLRQQFAAAECXQCRRXGHGVGRLRGWLCGGCEE......E 129
    ... :  |  |  . | .|. :.:  :.:|  : : | :..||      :
  1 MEGKWLLCMLLVLGTAIVEAHDGH..DDDVIDIEDDLDDVIEEVEDSKPD 48


130 XXFIGGDEKLAYESPVIDAKKFHFAXHFDDVEESRKRWVLSQAKKDEIAE 179
        .:..|:.|..|| .:. : ||  || .. |   |:||.||||:..:
 49 TTAPPSSPKVTYKAPVPTGEVY.FADSFDRGTLS..GWILSKAKKDDTDD 95


180 EISKYDGIWNWESPKRS.FGLHLGLVLKSKAKHXAIAAPXXKPFEFKSDK 228
    ||.|||| |: |. | | :. . ||||.|:||| ||.|.  ||| |.. |
 96 EIAKYDGKWEVEEMKESKLPGDKGLVLMSRAKHHAISAKLNKPFLFDT.K 144


229 PLVVQYEVTLQEGQECGXSYL 249
    ||:|||||.:|:| ||| .|:
145 PLIVQYEVNFQNGIECGGAYV 165
```

```
     Mouse  ...DLDDVIE EVEDSKSKSD ASTPPSPKVT YKAPVPTGEV Y.FADSFDRG SLSG..WILS KAKKDDTDDE IAKYDGKWEV DEMKE.TKLP
       Rat  ...DLDDVIE EVEDSKSKSD TSTPPSPKVT YKAPVPTGEV Y.FADSFDRG SLSG..WILS KAKKDDTDDE IAKYDGKWEV DEMKE.TKLP
       Dog  ...DLDDVIE EVEDSKSKPD TSAPTSPKVT YKAPVPSGEV Y.FADSFDRG TLSG..WILS KAKKDDTDDE IAKYDGKWEV DEMKE.TKLP
     Human  ...DLDDVIE EVEDSKPDT. TAPPSSPKVT YKAPVPTGEV Y.FADSFDRG TLSG..WILS KAKKDDTDDE IAKYDGKWEV EEMKE.SKLP
 C.elegans  ...DDEEEVT KGSDDKEEFV PSLFVAPKLS DK.....STP N.FFDYFPVG SKIGLTWIKS LAKKDDVDSD IAKYNGEWSI GAPTK.VSIE
Arabdopsis  ...SFDEPFD G......... .......... .......... .......... .....RWIVS KNSDYEGVWK HAKSEG.... .........H
  S.pombe   ...VKAPLVE QFQGAWSE.. .......... .......... .......... .....RWIPS HAKRFVNGIE EMSYVGEWTV EESSGPGALK
    Yeast   ...FWEHF.. .......... .......... .......... ....QAYTNT KHLNQEWITS EAVNNEGSK. ..IYGAQWRL SQGRLQGSAW
    cDNA4   XGHGVGRLRG WLCGGCEEEX XFIGGDEKLA YESPVIDAKK FHFAXHFDDV EESRKRWVLS QAKKDEIAEE ISKYDGIWNW ESPKRSFGL.

(cont)
     Mouse  GDKGLVLMSR AKHHAISAKL NKPFLF.DTK PLIVQYEVNF QNGIECGGAY VKLL.SKTAE LSLDQF.HDK TPYTIMFGPD KCGEDYK.LH
       Rat  GDKGLVLMSR AKHHAISAKL NKPFLF.DTK PLIVQYEVNF QNGIECGGAY VKLL.SKTSE LNLDQF.HDK TPYTIMFGPD KCGEDYK.LH
       Dog  GDKGLVLMSR AKHHAISAKL NKPFLF.DTK PLIVQYEVNF QNGIECGGAY VKLL.SKTPE LNLDQF.HDK TPYTIMFGPD KCGEDYK.LH
     Human  GDKGLVLMSR AKHHAISAKL NKPFLF.DTK PLIVQYEVNF QNGIECGGAY VKLL.SKTPE LNLDQF.HDK TPYTIMFGPD KCGEDYK.LH
 C elegans  GDLGLIVKTK ARHHAIAAKL NTPFAF.DAN TFVVQYDIKF EEGQECGGGY LKLL.SEGAE KDLANF.QDK TAYTIMFGPD KCGATGK.VH
Arabdopsis  EDYGLLVSEK ARKYGIVKEL DEPLNL.KEG TVVLQYEVRF QEGLECGGAY LKYLRPQEAG WTPQGF.DSE SPYSIMFGPD KCGGTNK.VH
  S.pombe   GEAGLVMKDE AAHHAISYEF DEPINE.PEK DLVVQYEVNP EEGLNCGGAY LKLL....AE PTHGEM.SNS IDYRIMFGPD KCGVNDR.VH
    Yeast   .DKGIAVRTG NAAAMIGHLL ETPINVSETD TLVVQYEIKL DNSLTCGGAF IKLMSGFMNV EALKHYAPDT EGVELVFGPD YCAPEINGVQ
    cDNA4   .HLGLVLKSK AKHXAIAAPX XKPFEFKSDK PLVVQYEVTL QEGQECGXSY L......... .......... .......... ..........
```

Figure 3.20. Sequence alignment of the identified calnexin sequences in different species (using GCG PILEUP). The blue indicates where an amino acid is conserved in five or more of the species. Residues in cDNA 4 which are found in any of the other sequences are underlined.

(dots indicate gaps in the sequenced introduced by the PILE-UP program to optimise alignments)

was possible that the similarity found between cDNA 4 and calnexin was due entirely to the calreticulin domain. Therefore the cDNA 4 sequence was compared with *Drosophila* calreticulin using BESTFIT and at the nucleotide level shows a 79bp region with 63% homology at the end of the sequenced (Fig. 3.21a) cDNA 4 fragment. Comparing the amino acid sequences of the two (Fig 3.21b) showed 12.9% identity and 36.9% similarity which is not considered significant. These results are consistent with cDNA 4 having a calreticulin-like domain, but show that cDNA 4 is not *Drosophila* calreticulin. This is also emphasized by including *Drosophila* calreticulin in the comparison between species (Fig. 3.22). This diagram shows that cDNA 4 has a domain in common with the calnexin proteins that is missing in calreticulin and only shares a small region of homology with calreticulin which is also shared by the other calnexin homologues.

The SwissProt program Prodom release 28 (http://protein.toulouse.inra.fr) shows an alignment of calnexin and calreticulin sequences indicating common sequence domains (Fig 3.23). Adding cDNA 4 to this alignment shows that the 750bp sequence is equivalent to the first two sequence domains. A 3' section of cDNA 4 has also been sequenced by J. Foster (sequence not shown) and this corresponds to part of the fifth sequence domain.

3.7 Discussion and conclusions

3.7.1 ER mannosidase

The cDNA clones selected with the ER mannosidase rat gene have been investigated by restriction enzyme mapping and radioactive hybridization. The analysis of these four clones showed that they were similar, in as much as the same restriction enzymes cut them. In particular it was seen from the start that cDNA 33 and cDNA 36 were very similar, since cutting with the same restriction enzymes gave some of the same sized fragments. Clone 4, seemed to be different from the rest, having many

Figure 3.21. *Drosophila* calreticulin compared to cDNA 4.

a) Nucleotide sequence BESTFIT of *Drosophila* calreticulin (Gen EMBL: X64461) and cDNA 4.

```
           Quality:   25.8              Length:     79
             Ratio:  0.327                Gaps:      0
Percent Similarity: 64.557   Percent Identity: 63.291


1276 TCCAACGAGGACAAGCCCCTCGTGGTGCAGTTCTCCGTGAAGCACGAGCA 1325
     | |||   |||||||||| || |||||||||| |   || | |    || |
 672 TTCAAGTCGGACAAGCCGCTGGTGGTGCAGTACGAGGTCACGTTGCAGGA 721
               .          .
1326 GAATATCGACTGCGGCGGCGGCTATGTGA 1354
     |  |   || ||||| |:|   ||| |||
 722 GGGTCAGGAGTGCGGTGNCTCGTATCTGA 750
```

b) Amino acid sequence BESTFIT of *Drosophila* calreticulin (SwissProt: P29413) and cDNA 4.

```
           Quality:   71.9              Length:    264
             Ratio:  0.295                Gaps:      5
Percent Similarity: 36.929   Percent Identity: 12.863


   4 GSTEHSRQRSRENV*FTLI*LQIVRRARSRFLLFRNRPQWGKTVEKHSAH 53
     ::. .  :.  |..  .::  :|. .: .:. :: . ..... :.|.
 114 DCSLDQTDMHGESPYEIMFGPDICGPGTKKVHVIFSYKGKNHLISKDIRC 163
          .         .         .         .         .
  54 K*GHHLVFRGY*IPARQP....AXERKVGENGVENGR.....KPGRNACX 94
     | : .  | .. :.: ..      : ||:...:|:::     |. ::::.
 164 KDDVYTHFYTLIVRPDNTYEVLIDNEKVESGNLEDDWDFLAPKKIKDPTA 213
          .         .         .         .         .
  95 XLRQQFAAAECXQCRRXGHGVGRLRGWLCGGCEEEXXFIGGDEKLA.YES 143
       .:::.. .. . .   .. :: :.  ..:.:..     :||. : :|.
 214 TKPEDWDDRATIPDPDDKKPEDWDKPEHIPDPDATKPEDWDDEMDGEWEP 263
          .         .         .         .         .
 144 PVIDAKKFHFAXHFDDVEESRKRWVLSQAKKDEIAEEISKYDGIWNWESP 193
     |:||...|. . : .:::::.. : .:.::   |||:.    |: : : ..
 264 PMIDNPEFKGEWQPKQLDNPNYKGAWEHP...EIANPEYVPDDKLYLRKE 310
          .         .         .         .         .
 194 KRSFGLHL..........GLVLKSKAKHXAIAAPXXKPFEFKSDKPLVVQ 233
       .:|:.|          .:::.....  | ||:  |  :  ..|   .|
 311 ICTLGFDLWQVKSGTIFDNVLITDDVELAAKAAAEVKNTQAGEKKMKEAQ 360
          .
 234 YEVTLQEGQECGXS 247
      ||  ..::| : .
 361 DEVQRKKDEEEAKK 374
```

Figure 3.22. PILEUP of calnexin species homologues (see Fig. 3.20) and *Drosophila* calreticulin (Calr). The underlined residues in *Drosophila* calreticulin and cDNA 4 are those which they share with any of the calnexin sequences.

```
                151                                                 200
      Mouse  PVPTGEVYFA DSFDRGSLSG ..WILSKAKK DDTDDEIAKY DGKWEVDEMK
        Rat  PVPTGEVYFA DSFDRGSLSG ..WILSKAKK DDTDDEIAKY DGKWEVDEMK
        Dog  PVPSGEVYFA DSFDRGTLSG ..WILSKAKK DDTDDEIAKY DGKWEVDEMK
      Human  PVPTGEVYFA DSFDRGTLSG ..WILSKAKK DDTDDEIAKY DGKWEVEEMK
  C.elegans  ....STPNFF DYFPVGSKIG LTWIKSLAKK DDVDSDIAKY NGEWSIGAPT
 Arabdopsis  .......... .......... .RWIVSKNSD YEGVWKHAKS EG........
    S.pombe  .......... .......... .RWIPSHAKR FVNGIEEMSY VGEWTVEESS
       Calr  GFISAEVYLK ENFDNENWED .TWIYSK... .....HPGKE FGKFVLTPGT
      Yeast  .......... QAYTNTKHLN QEWITSEAVN NEGSK...IY GAQWRLSQGR
      cDNA4  .....KFHFA XHFDDVEESR KRWVLSQAKK DEIAEEISKY DGIWNWESPK

                201                                                 250
      Mouse  E.TKLPGDKG LVLMSRAKHH AISAKLNKPF LF.DTKPLIV QYEVNFQNGI
        Rat  E.TKLPGDKG LVLMSRAKHH AISAKLNKPF LF.DTKPLIV QYEVNFQNGI
        Dog  E.TKLPGDKG LVLMSRAKHH AISAKLNKPF LF.DTKPLIV QYEVNFQNGI
      Human  E.SKLPGDKG LVLMSRAKHH AISAKLNKPF LF.DTKPLIV QYEVNFQNGI
  C.elegans  K.VSIEGDLG LIVKTKARHH AIAAKLNTPF AF.DANTFVV QYDIKFEEGQ
 Arabdopsis  .....HEDYG LLVSEKARKY GIVKELDEPL NL.KEGTVVL QYEVRFQEGL
    S.pombe  GPGALKGEAG LVMKDEAAHH AISYEFDEPI NE.PEKDLVV QYEVNPEEGL
       Calr  FYNDAEADKG IQTSQDARFY AASRKFD.GF SN.EDKPLVV QFSVKHEQNI
      Yeast  LQGSAW.DKG IAVRTGNAAA MIGHLLETPI NVSETDTLVV QYEIKLDNSL
      cDNA4  R..SFGLHLG LVLKSKAKHX AIAAPXXKPF EFKSDKPLVV QYEVTLQEGQ

                251                                                 300
      Mouse  ECGGAYVKLL .SKTAELSLD QF.HDKTPYT IMFGPDKCG. EDYKLHFIFR
        Rat  ECGGAYVKLL .SKTSELNLD QF.HDKTPYT IMFGPDKCG. EDYKLHFIFR
        Dog  ECGGAYVKLL .SKTPELNLD QF.HDKTPYT IMFGPDKCG. EDYKLHFIFR
      Human  ECGGAYVKLL .SKTPELNLD QF.HDKTPYT IMFGPDKCG. EDYKLHFIFR
  C.elegans  ECGGGYLKLL .SEGAEKDLA NF.QDKTAYT IMFGPDKCG. ATGKVHLIFR
 Arabdopsis  ECGGAYLKYL RPQEAGWTPQ GF.DSESPYS IMFGPDKCG. GTNKVHFILK
    S.pombe  NCGGAYLKLL ....AEPTHG EM.SNSIDYR IMFGPDKCG. VNDRVHFIFK
       Calr  DCGGGYVKLF ...DCSLDQT DM.HGESPYE IMFGPDICGP GTKKVHVIFS
      Yeast  TCGGAFIKLM SGFMNVEALK HYAPDTEGVE LVFGPDYCAP EINGVQFAIN
      cDNA4  ECGXSYL... .......... .......... .......... ..........
```

(dots indicate gaps in the sequenced introduced by the PILE-UP program to optimise alignments)

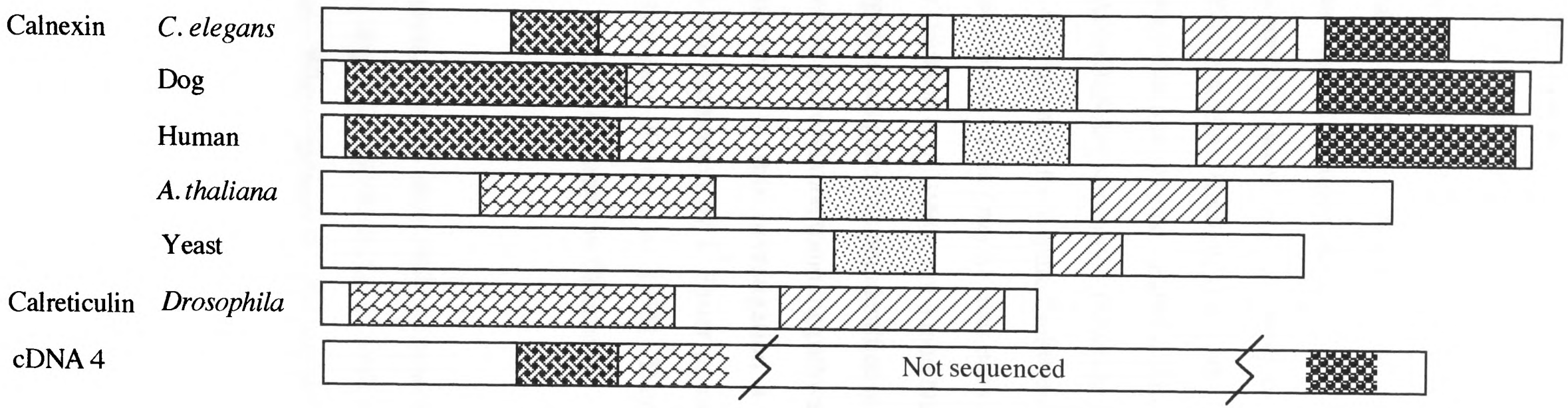


Figure 3.23. Calnexin and calreticulin sequence domains (after Prodom Release 28, SwissProt domain program). The cDNA 4 sequence obtained so far corresponds to the first sequence domain in calnexin, part of the calnexin/calreticulin common domain and the fifth calnexin domain.

AvaI restriction sites, and showed one fragment that hybridized to rat ER mannosidase even under stringent conditions. However when this strongly hybridizing fragment was sub-cloned and sequenced, it showed no significant homology to the sequenced rat ER mannosidase cDNA either with the nucleotide sequence or the predicted protein sequence using the BESTFIT program of the University of Wisconsin, Genetics Computer Group. Therefore I was forced to conclude that cDNA 4 was not the *Drosophila* homologue of ER mannosidase.

A recent review [Moremen *et al.* 1994] has cast doubt upon the role of this rat mannosidase. Earlier studies [Bischoff *et al.* 1983; Bischoff *et al.* 1986; Bischoff *et al.* 1990] all suggested that this enzyme catalysed the cleavage of an α1,2 linked mannose unit from the $Man_9GlcNAc_2$ substrate. However it has now been suggested that this α-mannosidase does not have the required substrate specificity and may be involved in glycoprotein catabolism. This enzyme also does not have the required ER locating signal sequence [Jackson *et al.* 1990], nor a membrane spanning domain. It is possible that the homologue of this gene was not identified in *Drosophila* because it is not part of the highly conserved glycosylation trimming pathway.

Simultaneously to this investigation a mouse cDNA clone for Golgi mannosidase II had been used by J. Foster to successfully identify a *Drosophila* homologue for that gene [Foster *et al.* 1995]. Therefore my attention turned to Golgi mannosidase II, which is discussed in Chapter 4.

3.7.2 Calnexin

The sequences were also used to search the GenEMBL and SwissProt Databases. The original FASTA search in 1993 failed to identify any significantly similar sequences. When repeated in 1995 and taken in conjunction with the BLAST results, a significant similarity to the oligosaccharide-binding molecular chaperone, calnexin, was discovered. That this similarity was not found originally could have

been due to few calnexin sequences being present in the database. Much of the work on calnexin has been carried out in the last few years and at the time of the initial search in April 1993 three sequences were already in the database. They have since been updated and others added. The use of a different database search program may have been more important in identifying calnexin. FASTA did identify calnexin among the top 40 similar sequences to the translations in reading frames 2 and 3 in 1995, but since FASTA only makes use of the single reading frame translation with which it is supplied, shifts in reading frame cannot be taken into account and therefore the similarity scores are reduced to a non-significant level. This makes BLAST the better search program to use with fragmented sequences.

Why did the rat ER mannosidase identify cDNA 4 which may be a calnexin homologue? Calnexin is an oligosaccharide-binding protein located in the ER as is the ER mannosidase. However sequences for retaining proteins in the ER are located at the carboxy-terminal of the protein and therefore could not be responsible for the strong hybridization of rat ER mannosidase to the 0.78kb fragment at the 5' end of cDNA 4 which was subcloned. Additionally the rat ER mannosidase does not have this ER retention signal (see section 3.7.1). ER mannosidase binds to the substrate $GlcNAc_2Man_9$ while the calnexin substrate is the same with the addition of a single glucose. Could there be some similarity in binding site that would lead to the hybridization? No region of significant similarity could be identified. However it seems at least possible that a short region of similarity between cDNA 4 and the rat ER mannosidase sequence could be responsible for the hybridization results seen. This is a situation not without precedent as other members of the laboratory have previously identified clones by hybridization that have subsequently failed to show any similarity to the sequence with which they were identified [Mellor 1992].

It may be however just chance that has led to the identification of a protein with any connection to the one that we were searching for. Using a cDNA library means

that sequencing any clone may identify a protein previously identified either in *Drosophila* or in other organisms.

3.7.3 Further work

The similarity of cDNA 4 to calnexin was only discovered in 1995, using ~~availability of~~ the BLAST program on the Oxford Pathology Molecular Biology VAX. Had this similarity been discovered in 1993 it is unlikely that we would have considered this protein of interest to our group as its function as a general oligosaccharide-binding protein was then unknown. However now there is much interest in the significance of this protein as part of the *N*-linked glycosylation pathway. It has been suggested that the "intimate association of calnexin with N-linked oligosaccharides ... may be a major factor responsible for preserving the *en bloc* mode of glycosylation that originated in early eukaryotes" [Ware *et al.* 1995]. We intend to study cDNA 4 initially by completing and confirming the sequence data already obtained. Further investigation would involve transformation rescue of mutants of the gene in *Drosophila,* similar to that discussed for the putative *Drosophila* Golgi mannosidase II in Chapters 4 and 5.

CHAPTER FOUR
GOLGI MANNOSIDASE II : GENE STRUCTURE

4.1 Aims and introduction

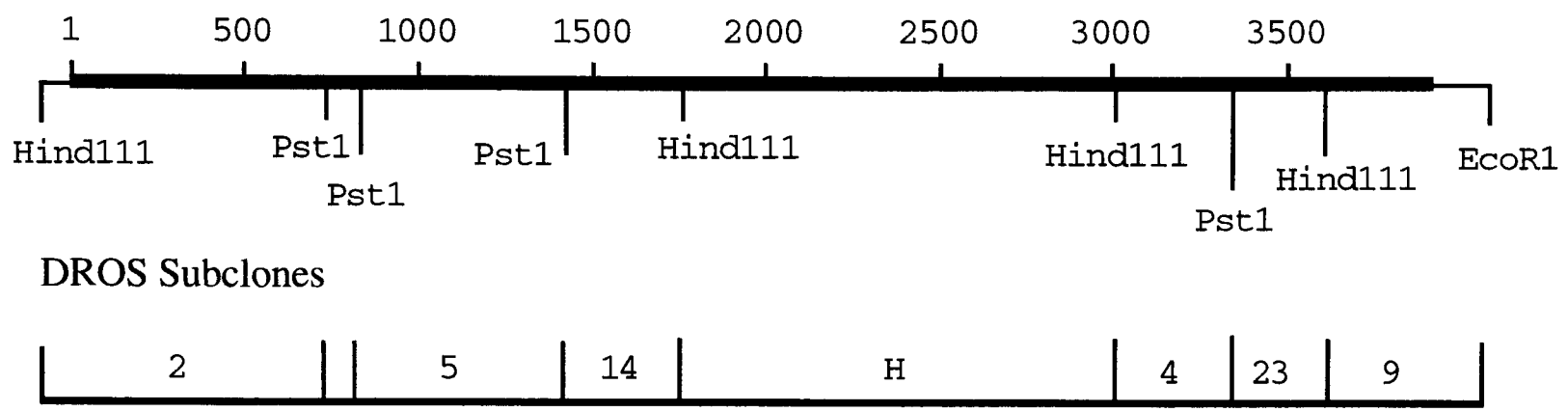
Murine α-mannosidase II cDNA clones have been isolated, characterized and expressed [Moremen 1989; Moremen *et al.* 1991]. A partial cDNA clone was isolated using a probe determined by mixed oligonucleotide-primed amplification of cDNA by PCR. This was used to identify further overlapping clones from a cDNA library, which were sequenced to obtain the full open reading frame for Golgi mannosidase II. We were given the full mouse Golgi mannosidase II cDNA clone by Kelly Moremen, University of Georgia, USA.

The mouse cDNA clone was used to screen a *Drosophila* cDNA library arrayed on a filter [Hoheisel *et al.* 1991] as for the identification of putative ER mannosidase cDNA clones. Two colonies were positive and carried the same insert. One of these (~4kb) was sequenced and compared to the original mouse sequence with which it was identified. At the nucleotide level there is 60% identity between the sequences and there is greater than 60% similarity with the peptide sequence predicted from the mouse cDNA [Foster *et al.* 1995]. This was called *Drosophila* Golgi mannosidase II (*GmII*) since it shows considerable structural homology with mouse clone, although functional homology still remains to be demonstrated.

The sequencing of the *Drosophila* cDNA provided seven different subclones (Fig. 4.1) and many primers for use in identifying and sequencing the genomic gene. (Refer to Appendix I for primer catalogue.)

The cDNA for the *Drosophila* homologue of Golgi mannosidase II had therefore been identified and cloned. My aim was to clone the genomic sequence in order to investigate the structure, expression and control of this gene which would give

a) *Drosophila GmII* cDNA



b) DROS Subclones

Figure 4.1. a) RE map of *GmII* cDNA showing HindIII, PstI, and EcoRI sites. The end HindIII and EcoRI sites are in the vector pNB40 (see Chapter 3).

b) HindIII/PstI DROS subclones made by J. Foster.

some insight into its natural function and significance. A genomic clone was also required to identify mutants of *GmII* generated from a saturation mutagenesis experiment in the region of the gene by transformation rescue [Yudkin 1996].

4.2 Identifying the 5' end of *GmII*

The 5' end of the gene containing controlling sequences was identified by screening a HindIII λ phage library of genomic *Novosibirsk Drosophila* DNA (D. Coates) with DROS 2, the subclone from the 5' end of the *GmII* cDNA {2.12: hybridization conditions - 65^{o}C, 2xSSC.} This initially identified 7 clones, which were subsequently reduced to two by further rounds of hybridization. Phage DNA was extracted from one clone {2.11} and partially sequenced up-stream of the start of transcription using primer D2U4 (Appendix I) from the 5' end of the cDNA (Fig. 4.2). This showed the sequence to overlap the mannosidase cDNA clone (underlined) with complete identity and to lead into the 5' controlling region. The phage clone was obtained from a HindIII library, an enzyme that cuts within the *GmII* cDNA sequence, so this clone does not contain the full coding sequence, required for the transformation rescue experiment. Therefore it was necessary to try a different approach to obtain a full length clone of the gene. This was achieved firstly by physically mapping the cDNA in the *D.melanogaster* genome and then using this information to obtain existing genomic P1 clones from the region (*Drosophila* Genome Centre).

4.3 *In situ* hybridization

In situ hybridization {2.9} of biotinylated *GmII* cDNA to polytene chromosomes localized the gene to the region 85D on 3R. This result was later confirmed by B. Yudkin (Fig. 4.3) and narrowed to the region 85D 12-16.

-340 AAAATATACA AATTATATTC AATTTAACAA GTTCCTTGAA

-300 CAAAATTTTA AGAATTAAAT ACCAAAAATT ATATACTGGG CTCAACCGGG

-250 ATTTGAACCC GGGACCTCTC GCACCCAAAG CGAGAATCAT ACCCCTAGAC

-200 CATTGAGCCT CTTAGTTTGG GACGGCCAAA TAGCCCAAAC GAGGGAAAAT

-150 TGTTGCCCAA TTACATTTTC TTTAATTGCC TTTAAGTTTT GTCTTTCTAA

-100 GCCATTTTAA CAATTACATT TTGCAGAGTT AGCAGCGCCG ATAACTGACA

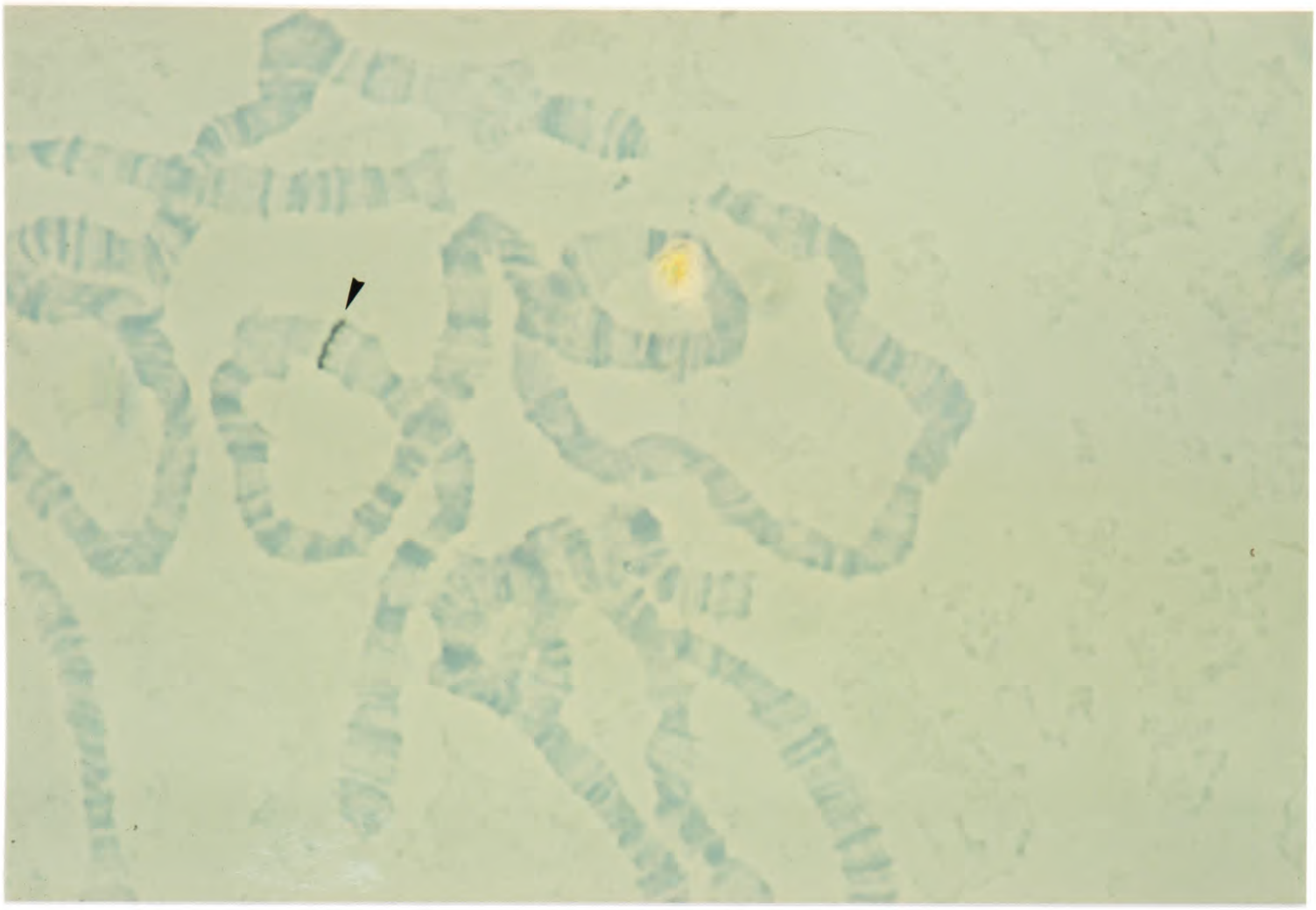
-50 GCCTATCGAT TTTCGGGAAA AGCAAACATA TCGCCAGATC TGATTTGTTT

0 ATGCTAAAAA AAAACCACGG AAGTTCCAAC CGATTGGATT GTTGCAGTTG

+50 GCTGCTGTGT CTCTGACGTC

Figure 4.2. Sequence of phage clone corresponding to the 5' end of *GmII*. The underlined sequence overlaps the previously sequenced *GmII* cDNA.

Figure 4.3. *In situ* hybridization of *GmII* to a squash of polytene chromosomes (B. Yudkin).



4.4 *GmII* genomic clone

4.4.1 P1 Clones

The bacteriophage P1 cloning system [Sternberg 1990] has been used to clone fragments up to 100kb in length. A bacteriophage P1 library of *D. melanogaster* has been constructed [Smoller *et al.* 1991]. 3840 individual clones were made from a partial Sau3A digest of adult fly DNA from the isogenic strain *y; cn bw sp*, in the BamHI digested vector pNS582tet14Ad10 (Fig. 4.4a). This library has subsequently been supplemented using a second vector pAd10sacBII [Pierce *et al.* 1992] (Fig. 4.4b). This second vector has the advantage of a positive selection system for clones containing inserts. Clones without inserts express the *sacB* gene which kills *E. coli* cultured in sucrose solution. The *Drosophila* Genome Centre now maintains a library of 9,216 arrayed clones with an average insert size of 80kb and 5-fold coverage of the *D. melanogaster* genome. These clones have been mapped by *in situ* hybridization to squashes of polytene chromosomes from Oregon R flies [Lozovskaya *et al.* 1993].

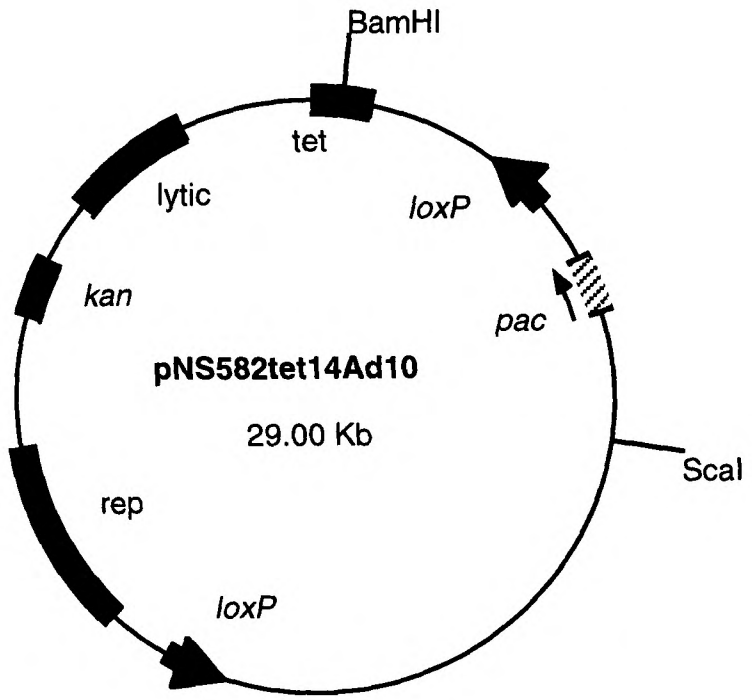
The localisation of the *GmII* cDNA to 85D by *in situ* hybridization was used to select P1 clones in this region. The table shows information on the three P1 clones provided by Spyros Artavanis-Tsakonas, Yale University.

Clone No.	Clone name	Grid position	Vector*	Mapped to:
1	DS05016	53-24/53B12	a	85D11-17
2	DS01769	19-41/19D5	b	85D18-27
3	DS06620	69-92/69H8	b	85D1-2

* See Fig. 4.4.

DNA prepared from the P1 clones {2.2.3} was digested with BamHI and separated by gel electrophoresis {2.4, 2.5}. Southern blotting {2.7} identified P1 clone **1** (DS05016), mapping to 85 D11-17 as containing *GmII* (Fig. 4.5a-b) and more specifically identified only a single ~6kb band in an EcoRI digest of the P1 DNA which

a) pNS582tet14Ad10.



b) pAD10-SacBII.

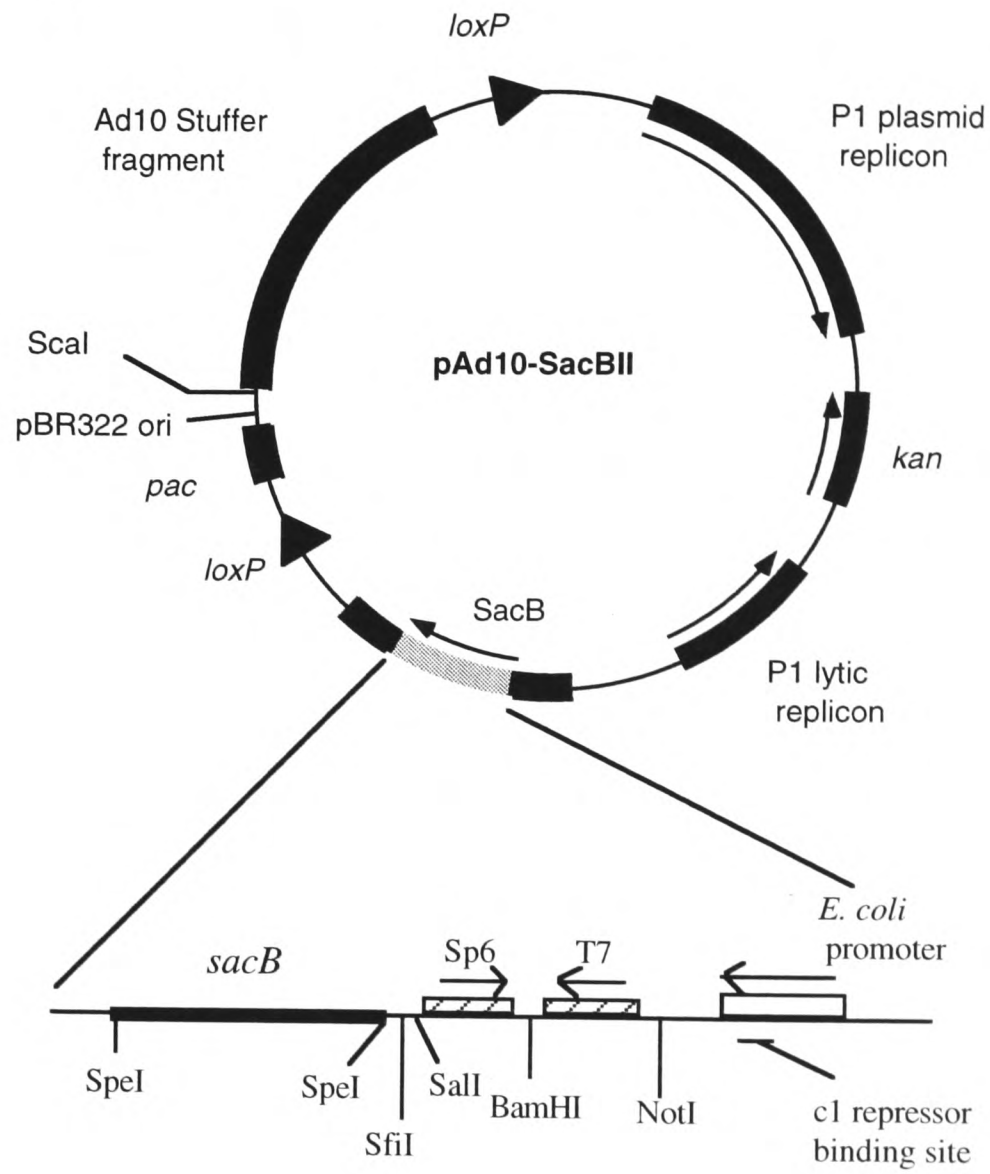


Figure 4.4. P1 vectors used to make a *D. melanogaster* genomic library (Smoller, 1991 and Pierce, 1992). Such vectors contain a P1 packaging site (*pac*) for packing ligated vector and inserting into phage particles; two P1 *loxP* recombination sites which recircularize the DNA upon infecting an *E. coli* strain with P1 Cre recombinase; bacterial kanamycin resistance (kan^r) for selection of clones containing the plasmid; a P1 plasmid replicon to maintain the plasmid in E. coli and a *lac* promoter-regulated P1 lytic replicon to allow amplification of the plasmid upon induction with IPTG in cells with the $lacI^q$ repressor.

must contain the whole gene (Fig. 4.5c-d). This 6kb EcoRI-EcoRI fragment was subcloned into the EcoRI site of pUC18 (see Fig. 3.9) by excising the band from the gel, purifying the DNA, ligating into pUC18 treated with alkaline phosphatase to reduce recircularization, and transforming *E. coli.* with the resulting construct {2.6, 2.11}. The presence of the correct insert was checked by partial sequencing of several positive (white) colonies. The plasmid was called subclone pP1 (E-E).

4.4.2 Genomic sequence

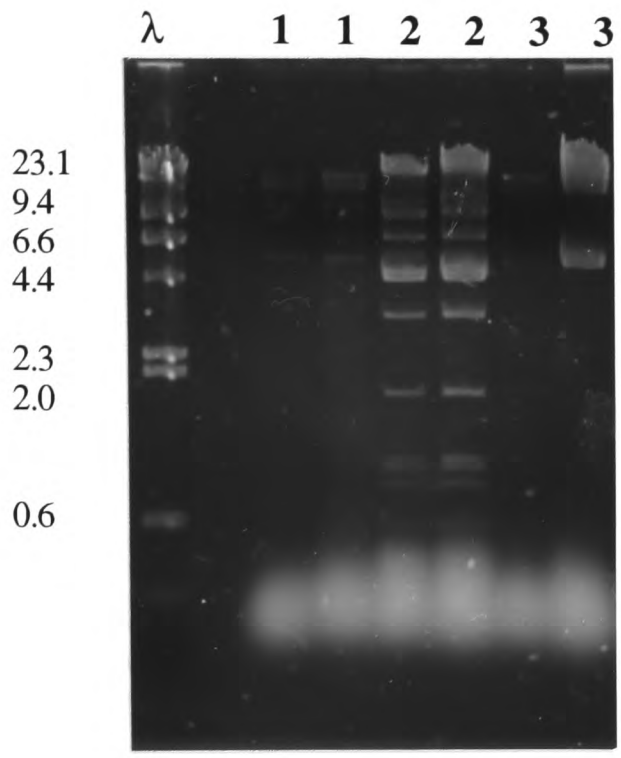
The genomic P1 clone **1** was initially investigated for introns [Padgett *et al.* 1986]. Pairs of sequencing primers were used for PCR {2.17} on the P1 genomic DNA and *GmII* cDNA to see if the genomic DNA gave the same size band or if a larger band was produced suggesting the presence of an intron between the primers. The PCR products were separated by electrophoresis. Figure 4.6 shows the electrophoresis gels, the pairs of primers used and the expected and actual fragment sizes.

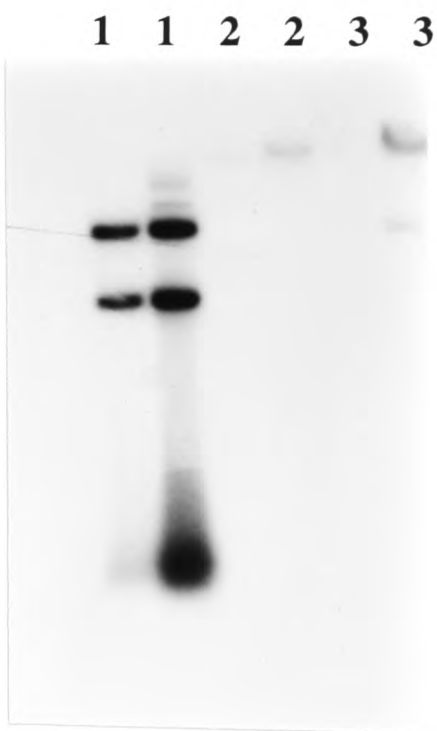
The observed discrepancies in PCR fragment size between cDNA and genomic DNA templates suggested the presence of introns in the genomic sequence. The entire pP1 (E-E) subclone was sequenced {2.10} (Fig. 4.7) and compared to the cDNA sequence to confirm these results; to obtain intron positions and sequences (underlined); and to check that there were no further introns too small to be identified by PCR. Sequencing primers from sequencing the cDNA were used as far as possible. New primers required to sequence the introns on both strands are shown in appendix I. The 5' and 3' regions were also sequenced on both strands (italics) and are discussed in section 4.6.

Figure 4.5. P1 clones.

a) Gel of P1 clones **1-3** digested with BamHI.

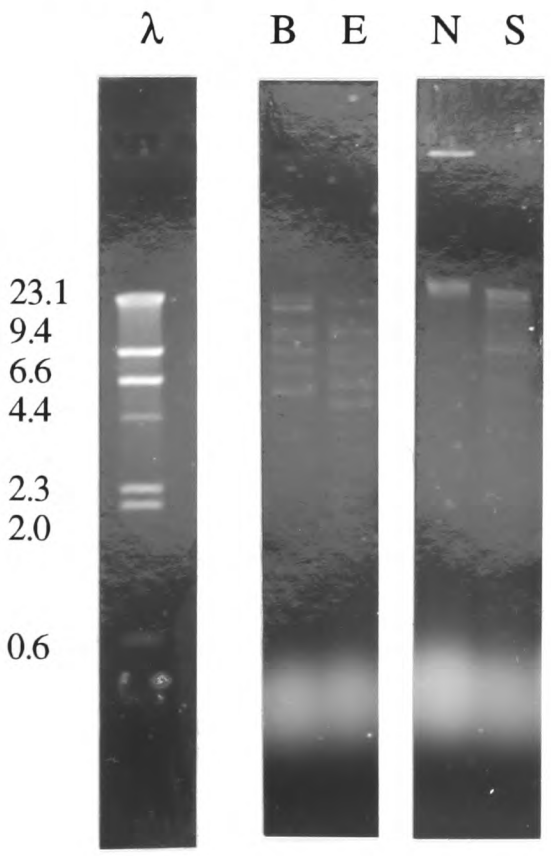
b) Autoradiograph of P1 clones **1-3** probed with *GmII* cDNA (conditions: 65^{o}C, 2xSSC).





c) Gel of P1 clone **1** digested with various REs (B - BamHI, E - EcoRI, N - NotI, S - SmaI).

d) Autoradiograph of P1 clone **1** probed with *GmII* cDNA (conditions: 65^{o}C, 2xSSC).



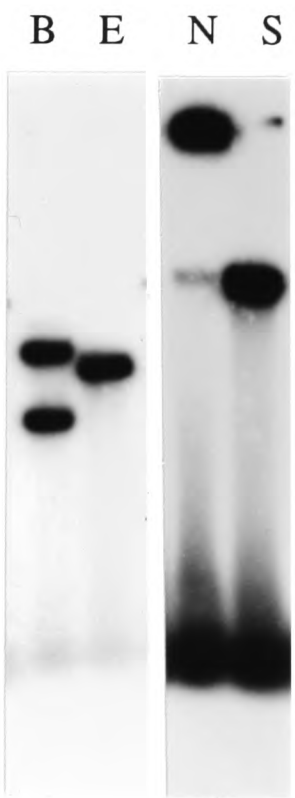
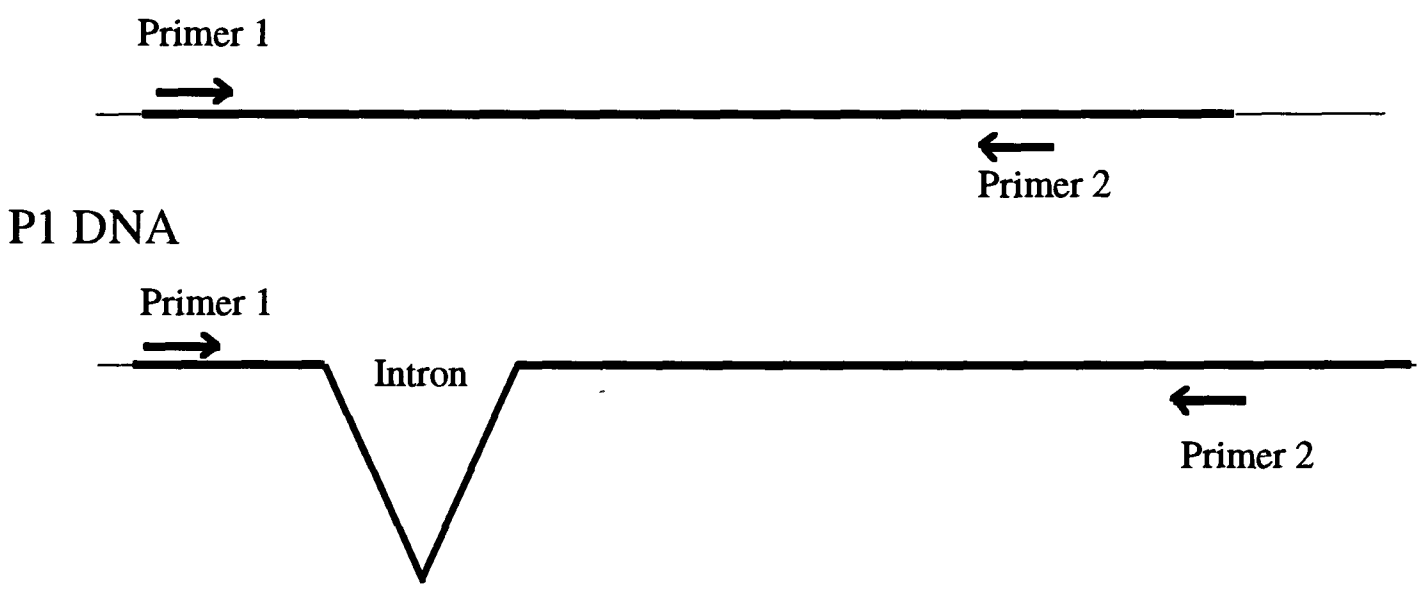


Figure 4.6. PCR intron experiment.

a) Pairs of primers were used to test if they spanned an intron.



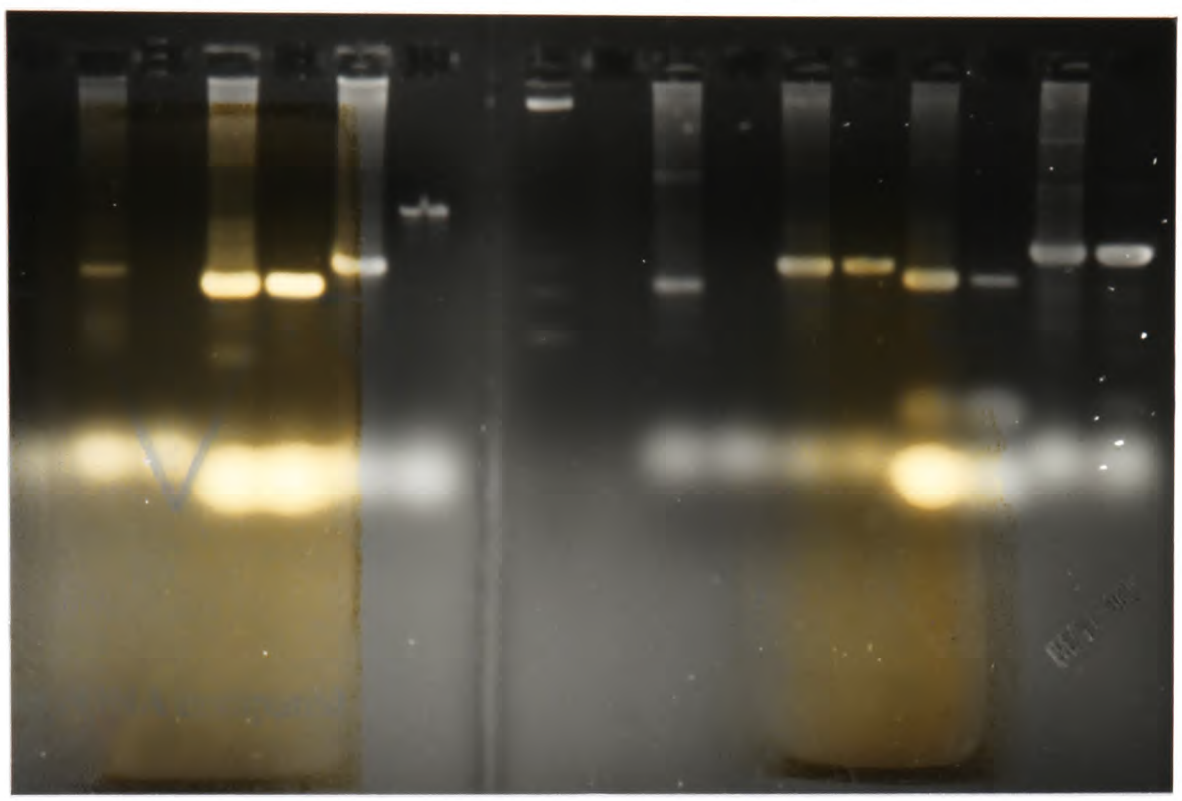
b) Genomic and cDNA compared - PCR results.

	Primers. (see appendix)	cDNA band size*	Genomic DNA band size	Notes
1	D2R2/D2U2	325bp	no band	?
2	D2R3/D2U1	292bp	292bp	No obvious intron
3	D2R4/D5/D2	379bp	600+bp	221+bp intron
4	D2/D5/D5U3	288bp	no band	?
5	D5R1/D5U1	372bp	372bp	No obvious intron
6	D5R3/D14U1	314bp	314bp	No obvious intron
7	D14R1/DHU2	648bp	648bp	No obvious intron
8	D23U1/DHU5	672bp	~792bp	~120bp intron
9	DHU2/DHR3	329bp	~593bp	~274bp intron

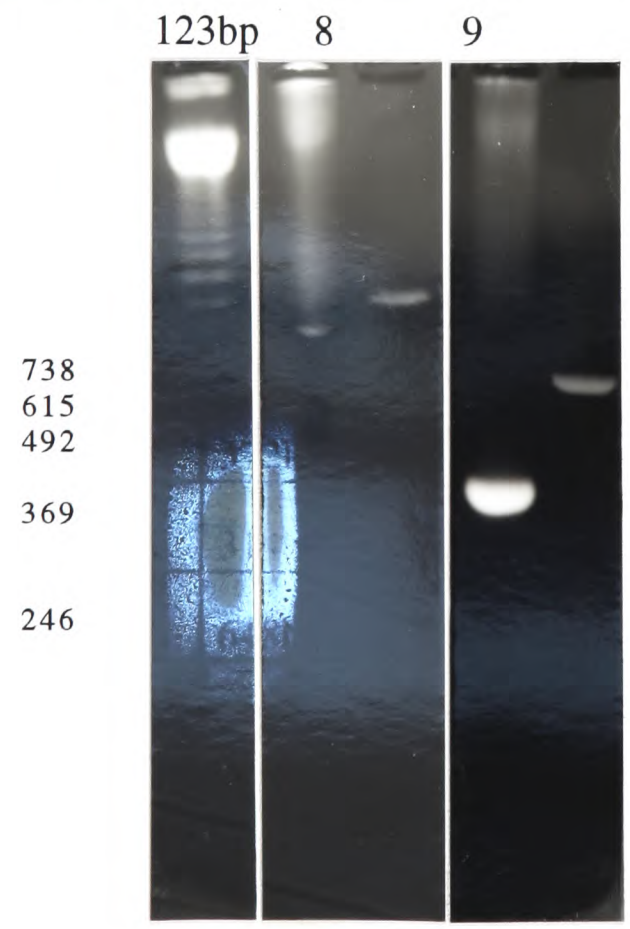
*Calculated from sequence.

Figure 4.6 (cont). PCR intron experiment (PCR annealing temperature: 58^{o}C).

c) i) Gel photograph of PCR samples 1-7 (for sizes see Fig. 4.6b). (0-no DNA).



ii) Gel photograph of PCR samples 8-9 (D.B. Roberts).



-700 *AATTCCCATC TTTTGGCCTC CGTTGGTGAC TCCAGGGAAT CGAACACGGC GGCTCCAATT AAAAGATAGG TGAAAGTGCA TACCACCAGG GAAAGGGTAC*

-600 *GGACATTCTG TCGCTTCATC ATATAGTCTC AAATATAATC ACTACGATCT ATAATCAATT AATATATTTT GCACCACGCG CCTTTTTCGT TCAGTTAAAG*

-500 *ATGCTGCGTC TGACAAAATC GATAGTTGTC AGCAAAACAA ACACACAGCA TCCGCCAAAA ATGGTTGATT TCGGTTACAG TGGCCGAATA ACATATGATA*

-400 *CAATGAATTT TAAAAATTTC ATCATTCTTT AAATATGTAA TAAGTTTTCT TTTATCTATT GATTTGACCG TTAATCTCAA TAAAGGATTG TGATTTTACC*

-300 *GTTAGTTTAT CACAACAGAA GCAACAGAAT TTGATGGCGT TGCCACTTGG CGGTATCTGT TAACTCTGTA CAGCTAAAAA GCTGTAATGT CACGATAACT*

-200 *AAGGATAATA ATAATTTTTT TAAAGGATCA ATATACACAA ACAACATAAG AGTTTTATTA CATTTTCTTT AATTGCCTTT AAAGTTTTTG TCTTTCTAAG*

-100 *CCATTTTAAC AATTTAACAT TTTGCAGTTA GCAGCGCCGA TAACTTGACA GCCTATCGAT TTTCGGGAAA AGCAAACATA TCGCCAGATC TGATTTGTTT*

0 ATGCTAAAAA AAAACCACGC AAGTTCCAAC CGATTGGATT GTTGCAGTTG GCTGCTGTGT CTCTGACGTC GACGTTTCCA TGCTTAATTA ACCGGAAA**AT**

100 **G**TTGCGAATA CGTCGGCGGT TCGCTTTGGT AATTTGCTCC GGCTGCCTGC TGGTTTTCCT CAGCCTGTAC ATAATCCTCA ATTTTGCGGC GCCGGCAGCC

200 ACCCAGATAA AGGTAAATAA ATGCTTCTAA ATCAATTCGA ATGTATAATA ACTTCTTTAT TTTAGCAGCC CAACTATGAG AACATTGAGA ACAAGCTGCA

300 TGAGCTGGAA AATGGTTTGC AGGAGCACGG GGAGGAGATG CGGAATCTCA GGGCGCGTCT TGCCGAAACA TCCAATCGCG ACGATCCAAT AAGACCTCCA

400 CTTAAAGTGG CTCGTTCCCC GAGGCCAGGG CAATGCCAAG ATGTGGTCCA AGACGTGCCC AATGTGGATG TACAGATGCT GGAGCTATAC GATCGCATGT

500 CCTTCAAGGA CATAGATGGA GGCGTGTGGA AACAGGGCTG GAACATTAAG TACGATCCAC TGAAGTACAA CGCCCATCAC AAACTAAAAG TCTTCGTTGT

600 GCCGCACTCG CACAACGATC CTGGATGGAT TCAGACGTTT GAGGAATACT ACCAGCACGA CACCAAGCAC ATCCTGTCCA ATGCACTACG GCATCTGCAC

700 GACAATCCCG AGATGAAGTT CATCTGGGCG GAAATCTCCT ACTTTGCTCG GTTCTATCAC GATTTGGGAG AGAACAAAAA GCTGCAGATG AAGTCGTAAG

800 CATGCTTTCC CTATATAGTA TTGATACAAT AATGAATAAC GGTAGTTAAC CCACAAAACA AAGTGCAATG CCATTTTGCT ATTGCATTCC ATAAACCTTA

900 CACAGTTGTT ATGTGCAGCA GTGAAATATT TTACTTCGTT ATTGCGTCAA AAGCGTTTAG ATAAACAAAA TGCATGATCA CGCTGCTTGT TTTCCGAAAC

1000 AGCTTTGTAC TTATCTACCT AATATTTATA ACCCAAAGCA GAAATAGGAA TTGATAAGCT TTGCAGTCTC AACAACTACA ATATTTAGTT TTTAACAAAA

1100 AATAATAAAT TATTTACTTA TTTACAGCAT TGTAAAGAAT GGACAGTTGG AATTTGTGAC TGGAGGATGG GTAATGCCGG ACGAGGCCAA CTCCCACTGG

1200 CGAAACGTAC TGCTGCAGCT GACCGAAGGG CAAACATGGT TGAAGCAATT CATGAATGTC ACACCCACTG CTTCCTGGGC CATCGATCCC TTCGGACACA

Figure 4.7. pP1 (E-E) sequence. Introns are underlined and *5' and 3' sequences* to the transcript in italics. The **start** and **stop** codons in the long ORF are in bold and the poly adenylation signal underlined.

1300 GTCCCACTAT GCCGTACATT TTGCAGAAGA GTGGTTTCAA GAATATGCTT ATCCAAAGGA CGCACTATTC GGTTAAGAAG GAACTGGCCC AACAGCGACA

1400 GCTTGAGTTC CTGTGGCGCC AGATCTGGGA CAACAAAGGG GACACAGCTC TCTTCACCCA CATGATGCCC TTCTACTCGT ACGACATTCC TCATACCTGT

1500 GGTCCAGATC CCAAGGTTTG CTGTCAGTTC GATTTCAAAC GAATGGGCTC CTTCGGTTTG AGTTGTCCAT GGAAGGTGCC GCCGCGTACA ATCAGTGATC

1600 AAAATGTGGC AGCACGCTCA GATCTGCTGG TTGATCAGTG GAAGAAGAAG GCCGAGCTGT ATCGCACGAA CGTGCTGCTG ATTCCGTTGG GTGACGACTT

1700 CCGCTTCAAG CAGAACACCG AGTGGGATGT GCAGCGCGTG AACTACGAAA GGCTGTTCGA ACACATCAAC AGCCAGGCCC ACTTCAATGT CCAGGCGCAG

1800 TTCGGCACAC TGCAGGAATA CTTTGATGCA GTGCACCAGG CGGAAAGGGC GGGACAAGCC GAGTTTCCCA CGCTAAGCGG TGACTTTTTC ACATACGCCG

1900 ATCGATCGGA TAACTATTGG AGTGGCTACT ACACATCCCG CCCGTATCAT AAGCGCATGG ACCGCGTCCT GATGCACTAT GTACGTGCAG CAGAAATGCT

2000 TTCCGCCTGG CACTCCTGGG ACGGTATGGC CCGCATCGAG GAACGTCTGG AGCAGGCCCG CAGGGAGCTG TCATTGTTCC AGCACCACGA CGGTATAACT

2100 GGCACAGCAA AAACGCACGT AGTCGTCGAC TACGAGCAAC GCATGCAGGA AGCTTTAAAA GCCTGTCAAA TGGTAATGCA ACAGTCGGTC TACCGATTGC

2200 TGACAAAGCC CTCCATCTAC AGTCCGGACT TCAGTTTCTC GTACTTTACG CTCGACGACT CCCGCTGGCC AGGATCTGGT GTGGAGGACA GTCGAACCAC

2300 CATAATACTG GGCGAGGATA TACTGCCCTC CAAGCATGTG GTGATGCACA ACACCCTGCC CCACTGGCGG GAGCAGCTGG TGGACTTTTA TGTATCCAGT

2400 CCGTTTGTAA GCGTTACCGA CTTGGCAAAC AATCCGGTGG AGGCTCAGGT GTCCCCGGTG TGGAGCTGGC ACCACGACAC ACTCACAAAG ACTATCCACC

2500 CACAAGGCTC CACCACCAAG TACCGCATCA TCTTCAAGGC TCGGGTGCCG CCCATGGGCT TGGCCACCTA CGTTTTAACC ATCTCCGATT CCAAGCCAGA

2600 GTGAGTAGCC AGTTAAAGAT TACCCATTTA TCAAAGGAAC AAGAATCCTT TCCGTACTTT CTGTATTAAT TGTACATTTT CGTAACACGT TTCGCTATGC

2700 CGTTTCCTTA AACCATTATA GAGTGCAGGT TTTATTTATT GATATTAATA TTTCATTAAC CAGATGCCGT AATTGTCTTC TATTTATACT ATTTATTTAT

2800 ACTTTTTCAA TAAGCATTTC AATGTTACTT ACAGGCACAC CTCGTATGCA TCGAATCTCT TGCTCCGTAA AAACCCGACT TCGTTACCAT TGGGCCAATA

2900 TCCGGAGGAT GTGAAGTTTG GCGATCCTCG AGAGATCTCA TTGCGGGTTG GTAACGGACC CACCTTGGCC TTTTCGGAGC AGGGTCTCCT TAAGTCCATT

3000 CAGCTTACTC AGGATAGCCC ACATGTACCG GTGCACTTCA AGTTCCTCAA GTATGGCGTT CGATCGCATG GCGATAGATC CGGTGCCTAT CTGTTCCTGC

3100 CCAATGGACC AGCTTCGCCA GTCGAGCTTG GCCAGCCAGT GGTCCTGGTG ACTAAGGGCA AACTGGAGTC GTCCGTGAGC GTGGGACTTC CGAGCGTGGT

3200 GCACCAGACG ATAATGCGCG GTGGTGCACC TGAGATTCGC AATCTGGTGG ATATAGGCTC ACTGGACAAC ACGGAGATCG TGATGCGCTT GGAGACGCAT

3300 ATCGACAGCG GCGATATCTT CTACACGGAT CTCAATGGAT TGCAATTTAT CAAGAGGCGG CGTTTGGACA AATTACCTTT GCAGGCCAAC TATTATCCCA

Figure 4.7 (cont). Full pP1 (E-E) sequence.

3400 TACCTTCTGG TATGTTCATT GAGGATGCCA ATACGCGACT CACTCTCCTC ACGGGTCAAC CGCTGGGTGG ATCTTCTCTG GCCTCGGGCG AGCTAGAGAT

3500 TATGCAAGAT CGTCGCCTGG CCAGCGATGA TGAACGCGGC CTGGGACAGG GTGTTTTGGA CAACAAGCCG GTGCTGCATA TTTATCGGCT GGTGCTGGAG

3600 AAGGTTAACA ACTGTGTCCG ACCGTCAGAG CTTCATCCTG CCGGCTATTT GACAAGTGCC GCTCACAAAG CATCGCAGTC ACTGCTGGAT CCACTGGACA

3700 AGTTTATATT CGCTGAAAAT GAGTGGATCG GGGCACAGGG GCAATTTGGT GGCGATCATC CTTCGGCTCG TGAGGATCTC GATGTGTCGG TGATGAGACG

3800 CTTAACCAAG AGCTCGGCCA AAACCCAGCG AGTAGGCTAC GTTCTGCACC GCACCAATCT GATGCAATGC GGCACTCCAG AGGAGCATAC ACAGAAGGTG

3900 AGCTTAGAAG TTTATTATAT ATTAATATCA AAGAAAGTGA AATGGTATTT ATTTATAATT AACAATATTG TTTGACTAAT TGTTCGTTCG TCAGCTGGAT

4000 GTGTGCCACC TACTGCCGAA TGTGGCGAGA TGCGAGCGCA CGACGCTGAC TTTCCTGCAG AATTTGGAGC ACTTGGATGG CATGGTGGCG CCGGAAGTGT

4100 GCCCCATGGA AACCGCCGCT TATGTGAGCA GTCACTCAAG C**TGA**TCGCAA CGTGTTATGG CAAAGGATAA AGCAGAAACT GAAGCCAAGC GACCATCTCA

4200 CAACCTATGA ATCTCGGACG CTAATCCACT GCAATACAGG CAATCTTCGT TTAATGTTCG ACTTAAGCAC TGACTGCACC GAAAGACGTA ATGCGTAATC

4300 CATACTACGT ATCATTTATT TATTCAAGCT TCCACACGAG GGTTTGGAAA TTCAATCCTA TGACGCAGCG GTGTATTTTT CGCACCCCCG TGTAACTACA

4400 CGTTTTACTC TGTTTCCCCT AGTTTTTTAC GTTTTTTTGT GATATCCATC CAGTCCCACA GCAAGGCAAT ACATGTCTGT ACTCGATATC CCCAATGGGG

4500 TTTTATTGCC TATAACAATG AATCGTAACT ATAACTAACG TTTAACTGTA AAGTAGGCTT AGTGATAATG ACGCAACTGC GACTGCAATG ATATTAATCT

4600 CGACGTTTGT TTTTTACAAA TTGAATAAAA TTGAGTCTCA TTTAG*AATCG ACTTTTACGT TGCTATTCTA AACTAAAAAC ATAACGTCTC TTTTTATTCG*

4700 *AAAAAGGAAA GGATTTGTAT TACTATTCAA AATTCGTCAT TTTGGCTTTA GTTAGAATCG TGTTGTTGGT AAGTTATGAA TTGCACTTTC TATTGAATAT*

4800 *TTAATTTTTC TCAGTAAATC TAAAATAAAT ACATTTTAAT TTAATTTTTA AGATTAGATT TAATTTTTTT TTTAATATTG TAAGTAATGA ATGTTAGTAA*

4900 *ACAGTTGATC TTTGTATTAT TAAGTTTTGA ATTTATGTGG ATTATGAAAC GAGCTTTTGA GACGACTAAT TAAAGCCATA ATCATGAGTT ATAGTTTTAG*

5000 *TGGATGAAAA ACAGAATGCA ATCTTATCAG CTCCAGTGCG GGGTACTATT ATTGCCAGAA GACCTACAGA GATGAGATAC TAAGAGCGCG GAATTGGAAA*

5100 *ATGCACAATG GAAATGGAAA TCCGGATTGA GTGTAAAGTG GTTGTTTTGG ATAGAGAGCG AATCGTGCGC AGTTCGCTCT CTTCAAAATC CGAAACGAGC*

5200 *AACGAAAGCA AAACCGAATC CGAATCGGAA TCCGAAA*

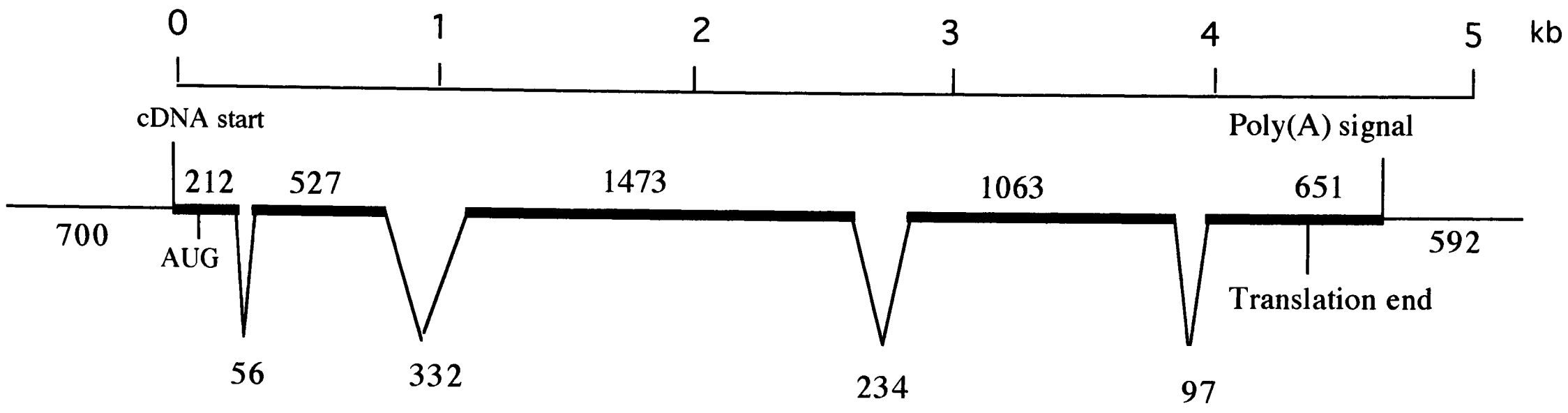
Figure 4.7 (cont). Full pP1 (E-E) sequence.

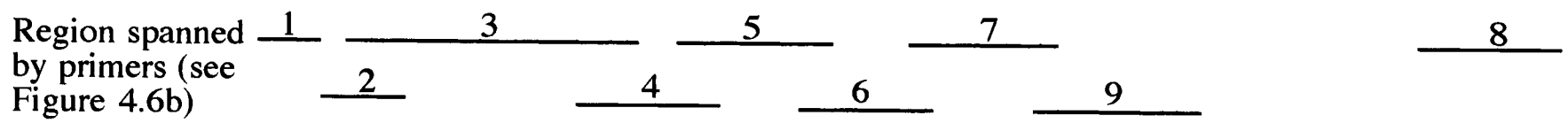
4.5 *GmII* intron analysis

Sequencing identified four introns (underlined in Fig. 4.7). The positions of the introns were found by comparing the cDNA with the genomic DNA. Figure 4.8 shows the structure of the gene.

4.5.1 Intron splicing

Introns are removed from the mRNA precursor in the nucleus by splicing - the removal of the intron sequence and subsequent rejoining of exon sequences (for reviews see Padgett *et al.* 1986; Sharp 1987). The pre-mRNA becomes associated with proteins, simultaneously with transcription, to form hnRNP particles (heterogeneous nuclear ribonucleoprotein particles). The proteins form stable particles at intron/exon boundaries that coalesce into pairs and form the spliceosome to catalyse splicing. The proteins involved are snRNPs (small nuclear ribonucleoproteins) which consist of proteins complexed to stable RNA molecules. The proteins U1, U2, U5, U4/U6 make up the spliceosome and are thought to act by RNA-RNA base pair complementarity [Maniatis *et al.* 1987]. Specific sequences in the intron are therefore necessary for the recognition by these proteins and subsequent removal of introns. The U1 snRNP binds to the 5' end of the intron [Padgett *et al.* 1983] and the U2 snRNP to an internal consensus sequence [Keller *et al.* 1984]. The 5' end of the intron is cut and forms a 2'-5' phosphodiester linkage to an adenine base in an internal consensus sequence [Ruskin *et al.* 1984] to form a lariat intermediate. The spliceosome assembly holds the RNA pieces in place until the 3' end is also cleaved and the exons joined to form the spliced mRNA. This splicing mechanism is conserved for all nuclear introns - those in genes housed in the nucleus. Other types of introns exist in mitochondrial and chloroplast genes, which have different splicing mechanisms [Fedorov *et al.* 1992].





Primers	1	2	3	4	5	6	7	8	9
PCR Result	-	no intron	221+ intron	no intron	no intron	no intron	no intron	120 intron	274 intron

Figure 4.8. *Drosophila GmII* gene structure.

4.5.2 Intron consensus sequences

Intron splicing consensus sequences have been identified by analysis of introns of published sequences in the database [Csank *et al.* 1990; Mount *et al.* 1992]. The general sequence for all eukaryotes for the 5' end is [A/C A G | **G T** A/G A G T] and more specifically examining *Drosophila* is [A/C A G | **G T** A/G A G T A/T]. At the 3' end all organisms show the sequence [T Y_9 N C **A G** | G] and in *Drosophila* [T_5 Y_3 T T N C **A G** | A/G T]. Figure 4.9 shows the expected intron consensus sequences and those found in *GmII* introns I-IV. The sequences found are very similar to those expected and contain the GT at the 5' end and AG at the 3' end that are conserved in all eukaryotes. (Some non-conforming splice sites in nuclear genes have been identified [Senapathy *et al.* 1990] and other classes of introns in organelle genes have no consensus sequence [Fedorov *et al.* 1992].)

An internal consensus sequence required for the splicing of introns has also been identified located between 20 and 55 bp upstream from the 3' end of the intron [Keller *et al.* 1984]. This sequence is less rigidly conserved with the exception of a single A to which the 5' end joins to form a lariat intermediate during splicing [Ruskin *et al.* 1984]. The *Drosophila* consensus sequence is [C/T T A/G **A** T/C] [Keller *et al.* 1985; Mount *et al.* 1992] and in *Drosophila* is located 18 to 35bp upstream from the 3' splice site [Keller *et al.* 1985] and at least 38bp downstream from the 5' splice site [Mount *et al.* 1992]. Mount *et al* [1992] have also reduced the stated distance required from the 3' end to 15bp. Figure 4.10 shows the 3' ends of the four *GmII* introns and marks where the internal site may be. In particular the limits of the sequence lying more than 38bp downstream from the 5' end and 15bp upstream of the 3' end [Mount *et al.* 1992] allows the position in the 56bp intron (intron I) to be identified within 3bp and this sequence differs from the consensus sequence by one mismatch. Intron II has one match to the consensus site within the 15 - 35bp window and two sites with one mismatch, intron III has one mismatched site and intron IV three sites that fit the

5' end consensus sequence.

	-3	-2	-1	1	2	3	4	5	6	7
All	A/C	A	G	G	T	A/G	A	G	T	A/T
Dros	A/C	A	G	G	T	A/G	A	G	T	A/T
I	**A**	**A**	**G**	**G**	**T**	**A**	**A**	A	**T**	**A**
II	G	T	C	**G**	**T**	**A**	**A**	**G**	C	**A**
III	**A**	G	A	**G**	**T**	**G**	**A**	**G**	**T**	**A**
IV	**A**	**A**	**G**	**G**	**T**	**G**	**A**	**G**	C	**T**

3' end consensus sequence.

	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	1	2
All	T	C/T	C/T	C/T	C/T	C/T	C/T	C/T	C/T	C/T		C	A	G	G	
Dros	T	T	T	T	T	C/T	C/T	C/T	T	T		C	A	G	A/G	T
I	C	**T**	**T**	**T**	A	**T**	**T**	**T**	**T**	A	**G**	**C**	**A**	**G**	C	C
II	**T**	**T**	A	C	**T**	**T**	A	**T**	**T**	**T**	**A**	**C**	**A**	**G**	C	A
III	A	A	**T**	G	**T**	**T**	A	**C**	**T**	**T**	**A**	**C**	**A**	**G**	**G**	C
IV	**T**	G	**T**	**T**	C	G	**T**	**T**	C	G	**T**	**C**	**A**	**G**	C	**T**

Figure 4.9. *GmII* intron sequences compared to expected sequences.

	-35	-15	0 (3' splice site)
Intron I	AAATGCTTCT AAATCAATTC GAATGTATAA TAACTTCTTT ATTTTAGCAG		
		(+38 from 5' splice site)	
Intron II	ACAATATTTA GTTTTTAACA AAAAATAATA AATTATTTAC TTATTTACAG		
Intron III	TATACTATTT ATTTATACTT TTTCAATAAG CATTTCAATG TTACTTACAG		
Intron IV	GTATTTATTT ATAATTAACA ATATTGTTTG ACTAATTGTT CGTTCGTCAG		

Figure 4.10. 3' ends of *GmII* introns. The *Drosophila* intron internal splicing consensus sequence is C/T T A/G **A** C/T. Sequences which conform or have one mismatch are underlined.

consensus sequence and one site with one mismatch. Therefore the introns found in *GmII* fit the expected splice mechanism consensus sequences found in other *Drosophila* genes.

4.5.3 Intron base pair composition

An analysis of the base pair composition of the introns and exons is shown in Figure 4.11a. In any randomly selected nucleotide sequence the proportions of bases would be expected to be equal. A Chi-squared test was performed to see if the ratios of AT bases to GC bases in introns and exons conform to this expected ratio or if it is significantly higher in introns. The results of the test show that indeed the number of ATs in exons is not significantly different from the number of GCs, while in introns the number of ATs is significantly higher than GCs (see Fig. 4.11b). This is in accordance with other *Drosophila* introns [Mount *et al.* 1992] and in other organisms such as *Dictyostelium discoidium*, *Caenorhabditis elegans* and *Tetrahymena thermophilia* [Csank *et al.* 1990]. This may account for the rigidity with which the intron splicing consensus sequences are conserved in yeast and mammals, while in the other organisms the base composition of introns may also be important in identifying intron sequences during splicing [Csank *et al.* 1990; Mount *et al.* 1992].

4.5.4 Nested genes

The sequences of the four introns were investigated for the presence of open reading frames, which may indicate nested genes or alternatively spliced exons. The introns in *GmII* are not very long and I, III, and IV had many stop codons in all reading frames. Intron II contained two short reading frames in frames 1 and 2, but without the appropriate splicing ends or with enough sequence to allow splicing (under

Figure 4.11. *GmII* intron/exon base composition.

a) Frequencies and percentages of bases found in *GmII* introns and exons.

Base	Exon	%	Intron	%
A	988	25.1	245	34.1
T	917	23.4	265	36.9
G	1001	25.5	97	13.5
C	1020	26.0	112	15.6
Total	3926	100%	719	100.1%

b) Chi-squared test. Is the number of A/T bases in introns significantly higher than the number of G/C bases?

	Exon	Intron
A/T	1905 (48.5%)	510 (71%)
G/T	2021 (51.5%)	209 (29%)
Chi^2 value*	3.43	126
Conclusion	Not significantly different	Significantly different

* Chi^2 calculated according to the formula :

$$\frac{[A/T - (A/T + G/C)/2]^2}{(A/T + G/C)/2} + \frac{[G/C - (A/T + G/C)/2]^2}{(A/T + G/C)/2}$$

A/T - Number of A + T bases in sequence

G/C - Number of G + C bases in sequence

the criteria defined in section 4.5.2). It was therefore concluded that there are no additional coding sequences contained within the gene.

4.5.5 Intron size

Intron sizes in *Drosophila* have been analysed [Mount *et al.* 1992] and show a median size of approximately 79bp, in contrast to mammals where it is rare to have introns less than 70bp. Mount *et al* [1992] investigates *Drosophila* introns which range in size from 51bp to 5392bp and appear to fall into two distribution groups, depending on size. Larger introns have been identified in some *Drosophila* genes (e.g. the 79kb intron of the *dunce* gene [Chen *et al.* 1987]). The lower limit of the range may be constrained due to the stoichiometry of splicing since the internal consensus sequence is usually located a minimum of 38 bp downstream from the 5' end of the intron and 15bp upstream of the 3' end. The introns found in the *GmII* sequence range from 56bp to 332bp.

4.5.6 Intron position

The "exon shuffling" theory of protein evolution suggests that exons code for functional protein domains [Gilbert 1978; Blake 1978] and implies that intron position is non random - that in a gene sequence the introns separate exons coding for independent protein units. *GmII* has a type II transmembrane domain [Foster *et al.* 1995] contained within the first exon, which is consistent with this theory. This first exon also contains a further 11 amino-acid residues before the intron. The protein structure of *GmII* is not known either in *Drosophila* or the mouse, nor has the catalytic site been identified, so the correlation of the other exons with protein domains is not yet possible.

One way of determining if the positions of introns are non random is to compare the position of introns in a protein conserved between species. The mouse

GmII genomic structure has not been elucidated. A PCR experiment was designed to test mouse genomic DNA for introns. Using the two aligned cDNA sequences [Foster *et al.* 1995] the positions of the *Drosophila GmII* introns could be superimposed on the mouse sequence. Primers were designed from the mouse cDNA sequence to amplify a fragment across the estimated positions of introns II and III in the most conserved regions of the gene (Fig. 4.12a). Mouse genomic DNA and the Golgi mannosidase II cDNA [Moremen *et al.* 1991] were used as templates (Fig. 4.12b). Interestingly, although the expected fragment size was amplified from the cDNA showing the PCR conditions were suitable for primer annealing, no strong band was obtained with the genomic DNA, even though the presence of faint bands indicated no degradation of the template. This suggests the presence of long introns, expected in mammalian genes (see section 4.5.5), too long for amplification under the conditions suitable for the cDNA fragment. The mouse gene may therefore contain long introns in similar positions to *Drosophila GmII* introns II and III.

Studies are also underway to investigate the position and sequence of the first intron in *GmII* in different *Drosophila* species (D. B. Roberts. pers. comm.)

4.6 Sequences 5' and 3' to the *GmII* transcribed sequence

Gene expression is often controlled or regulated by sequences upstream and less often downstream of the start of transcription. Specific recognition sequences are required for binding of the proteins involved in transcription [Bucher 1990].

4.6.1 Discrepancy between phage and P1 5' sequences

Sequencing upstream of the coding region in pP1 (E-E) revealed a discrepancy between the sequence of the phage clone (section 4.2) and the pP1 (E-E) subclone sequence (Fig. 4.13). The sequences are the same up to 144bp upstream from the start of transcription, but then they diverge. This may be a cloning artefact. Even though

Figure 4.12. PCR on mouse genomic DNA and mannosidase cDNA.

a) PCR primers. (5' -> 3').

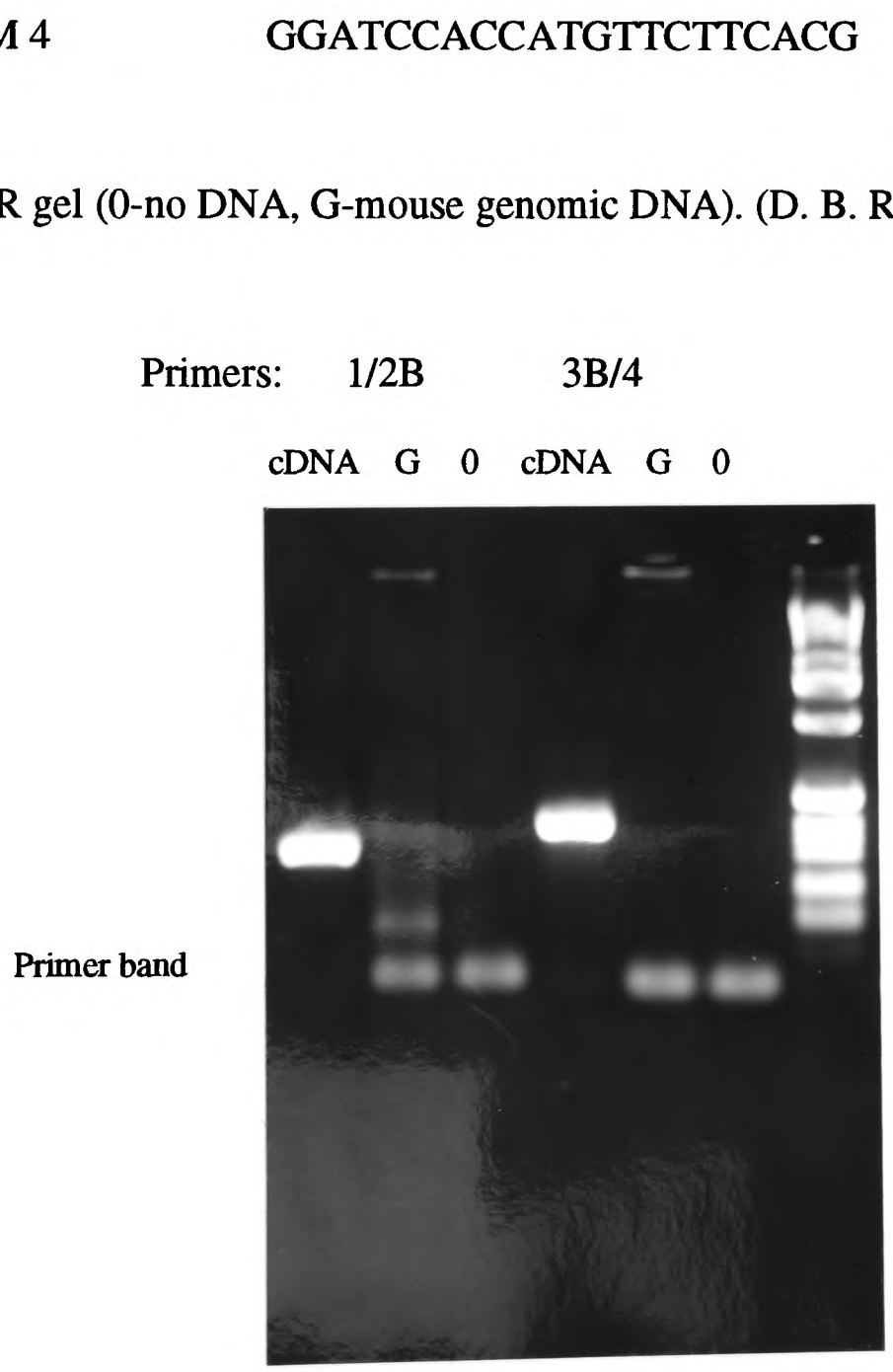
Spanning Intron III

MANM 1	GATGGTGGAGTTTGGAAGCAAGG
MANM 2B	GGCTTCATCAGGCATAACCCAG

Spanning Intron II

MANM 3B:	GCCTTCTTAGAGATGGATACG
MANM 4	GGATCCACCATGTTCTTCACG

b) PCR gel (0-no DNA, G-mouse genomic DNA). (D. B. Roberts).



NB: The amount of mouse genomic DNA used took into account the size of genome. An appropriate control to confirm these results would be to test primers designed from exon sequences on the genomic DNA.

the divergence begins more than 140bp upstream from the start of transcription one or other of the sequences may contain regions important for controlling the expression of the gene. As the sequence is required to transform flies it was important to discover which was the correct sequence. The experiment involved designing two unique primers one in each of the diverged sequences (P15 and PH5) and a common primer within the coding sequence (J5B) (see Fig. 4.13). The primers were used for PCR on genomic DNA, with the appropriate positive and negative controls. A second primer within the coding sequence (D2R3) was used to check that the reaction was working correctly with all three template DNAs. Control reactions with the three sets of primers and no template DNA controlled for contamination. The P1 unique primer produced a band with the genomic DNA, while the phage primer produced none indicating that the P1 clone contains the correct 5' sequence (Fig. 4.14).

4.6.2 Analysis of 5' region

The sequenced 5' region of *GmII* was investigated for promoter sequences. These sequences, TATA, Inr sequence, CAAT and the GC box are required for positioning the transcriptional machinery - RNA polymerase II and associated transcription factor proteins [Bucher 1990]. The sequences are usually located in the first 100bp upstream of the start of transcription. The TATA box is not present in the *GmII* sequence at about 25 bp upstream, the expected site, so it was concluded that *GmII* falls into the class of genes expressed from TATA-less promoters. The detailed sequence analysis of the *GmII* promoter region and further investigations in the absence of obvious promoter sequences are discussed in Chapter 6.

4.6.3 *GmII* 3' sequence

After transcription the mRNA is cleaved and polyadenylated at the 3' end [Birnstiel *et al.* 1985]. Transcription termination usually occurs considerably further

Figure 4.13. Phage and pP1 (E-E) sequences compared.

```
-340    Phage.....  ATATACAAAT TATATTCAAT TTAACAAGTT CCTTGAACAA
                    ...|..|...  |..||...|.  |.||..|.|.  ...|.....
        pP1(E-E).  ...GATTTGACCG TTAATCTCAA TAAAGGATTG TGATTTTACC

-300    AAATTTAAGA ATTAAATACC AAAAATTATA TACTGGGCTC AACCGGGATT
        .....|....  ...||.....  ..||...|..  |....|||..  ..||.....
        GTTAGTTTAT CACAACAGAA GCAACAGAAT TTGATGGCGT TGCCACTTGG
P15                                                 GCCACTTGG
PH5                                     GCGA  GAATCATACC  CCTAG>
-250    TGAACCCGGG ACCTCTCGCA CCCAAAGCGA GAATCATACC CCTAGACCAT
        .|....|.|.  .......|.|  |....|...|  |....|....  |.........
        CGGTATCTGT TAACTCTGTA CAGCTAAAAA GCTGTAATGT CACGATAACT
        CGGTATCTG>

-200    TGAGCCTCTT AGTTTGGGGA CGGCCAAATA GCCCAACGAG GGAAAATTGT
        ...|....|.  |...|.....  ......|..|  ....|...|.  ..||.||...
        AAGGATAATA ATAATTTTTT TAAAGGATCA ATATACACAA ACAACATAAG

-150    TGCCCAATTA CATTTTCTTT AATTGCCTTT AAAGTTTTTG TCTTTCTAAG
        .|....||||  ||||||||||  ||||||||||  ||||||||||  ||||||||||
        AGTTTTATTA CATTTTCTTT AATTGCCTTT AAAGTTTTTG TCTTTCTAAG

-100    CCATTTTAAC AATTTAACAT TTTGGAGTTA GCAGCGCCGA TAACTTGACA
        ||||||||||  ||||||||||  ||||||||||  ||||||||||  ||||||||||
        CCATTTTAAC AATTTAACAT TTTGCAGTTA GCAGCGCCGA TAACTTGACA

-50     GCCTATCGAT TTTCGGGAAA AGCAAACATA TCGCCAGATC TGATTTGTTT
        ||||||||||  ||||||||||  ||||||||||  ||||||||||  ||||||||||
        GCCTATCGAT TTTCGGGAAA AGCAAACATA TCGCCAGATC TGATTTGTTT

0       ATGCTAAAAA AAAACCACGG AAGTTCCAAC CGATTGGATT GTTGCAGTTG
        ||||||||||  |||||||||.  ||||||||||  ||||||||||  ||||||||||
        ATGCTAAAAA AAAACCACGC AAGTTCCAAC CGATTGGATT GTTGCAGTTG

+50     GCTGCTGTGT CTCTGACGTC ..............................
        ||||||||||  ||||||||||
        GCTGCTGTGT CTCTGACGTC..............................

            <GATGA   AGTTCATCTG   GGCG                    J5B
+650    CCCGAGATGA AGTTCATCTG GGCGGAAATC TCCTACTTTG CTCGGTTCTA
        ||||||||||  ||||||||||  ||||||||||  ||||||||||  ||||||||||
        CCCGAGATGA AGTTCATCTG GGCGGAAATC TCCTACTTTG CTCGGTTCTA
```

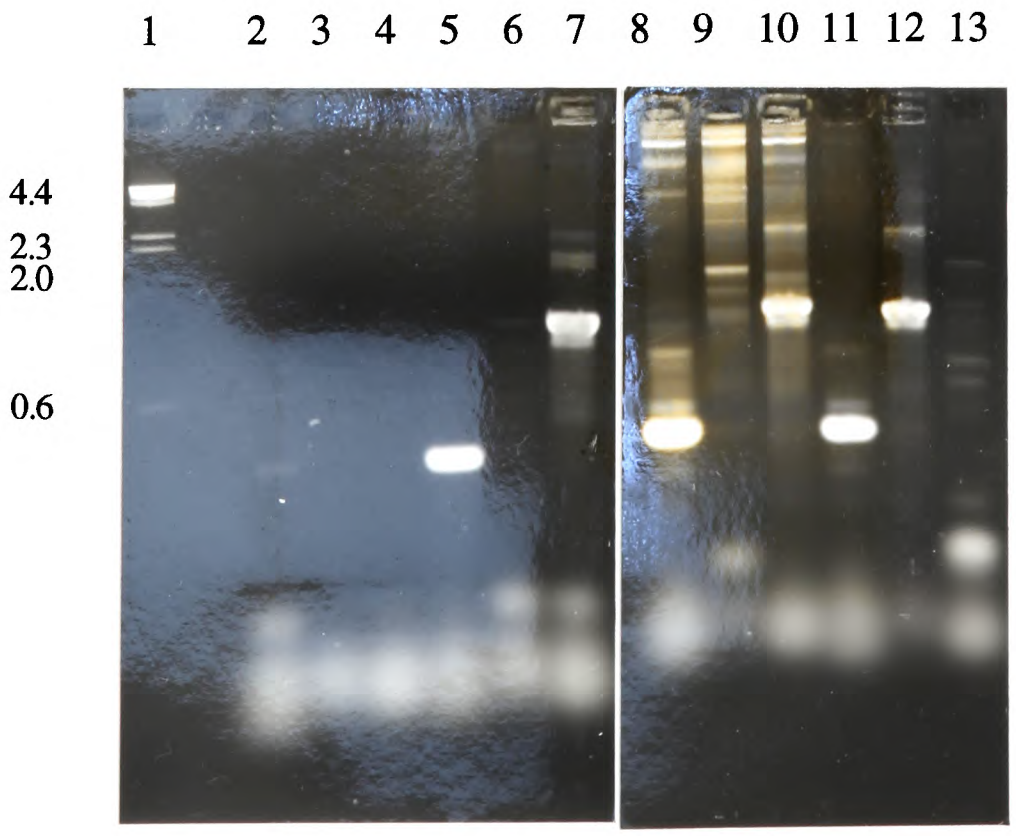
Figure 4.14. PCR experiment to test which was the correct 5' sequence.

a) Expected sizes.

Lane	DNA	Primer 1	Primer 2	Fragment size
1	λ HindIII ladder	-	-	{2.5.1}
2	No DNA	J5B	D2R3	-
3	No DNA	J5B	PH5	-
4	No DNA	J5B	P15	-
5	Genomic	J5B	D2R3	325
6	Genomic	J5B	PH5	923*
7	Genomic	J5B	P15	987*
8	P1	J5B	D2R3	325
9	P1	J5B	PH5	-
10	P1	J5B	P15	987
11	Phage clone	J5B	D2R3	325
12	Phage clone	J5B	PH5	923
13	Phage clone	J5B	P15	-

* Either the Phage or the P1 primer will work on the genomic DNA indicating which contains the correct *Drosophila* 5' sequence.

b) PCR gels (PCR annealing temperature 62 °C).



downstream than the extent of the fully processed mRNA. The end of transcription from RNA polymerase II may only be loosely specified by weak serially repeated termination sequences, and in some cases may involve the U_4-U_5 runs in a GC-rich region required for termination of RNA polymerase III transcription. The 3' modification requires an endonuclease, a poly(A) polymerase and a specificity component that recognises the conserved polyadenylation AAUAAA sequence [Proudfoot *et al.* 1976]. This sequence is usually found 11-30 nucleotides upstream from the start of polyadenylation [Wahle *et al.* 1992]. Mutations in the sequence reduce the efficiency of both cleavage and addition of the poly(A) tail [Sheets *et al.* 1990]. *GmII* has a polyadenylation site at -16 to -24bp upstream from the start of polyadenylation. The start of polyadenylation is usually (70%) at a conserved A nucleotide, often (59%) preceded by a C [Sheets *et al.* 1990]. *GmII* is polyadenylated after G A (see Fig. 4.7).

4.7 Discrepancies between *GmII* cDNA and genomic DNA

Comparing the *GmII* genomic and cDNA sequences revealed a number of discrepancies at the DNA level and 5 changes at the amino acid level (Fig. 4.15a).

The P1 sequence was obtained by only sequencing on one strand so the cDNA sequence was taken to be correct and only the differences that would alter the predicted amino-acid sequence were investigated further, sequencing on the other strand and using dITP where necessary. The results from this additional investigation are shown in the table in Figure 4.15a. Some of the differences were eliminated by this further study, but two positions (365 and 3663) show amino acid differences between the two sequences (Fig. 4.15b). The change in each case is a G in the pP1(E-E) sequence to an A in the cDNA sequence which produces the same predicted amino acid change of a glutamic acid in the predicted peptide sequence translated from the pP1(E-E) sequence

pP1 (E-E)	cDNA	Position (P1)	Amino-acid change	Result
T	G	361	No change	Polymorphism T/G
G	A	365	Glu/Lys	Polymorphism G/A
CG	GC	1401/2	His/Gln; Val/Leu	GC correct in both
C	T	1494	No change	Not resolved
C	T	1563	No change	Not resolved
G	A	1668	No change	G correct in both
CG	GC	1674/5	Val/Leu	GC correct in both
G	A	3628	Glu/Lys. HindIII site in cDNA	Polymorphism G/A
T	A	3663	No change	Not resolved
T	G	4423	No change	Not resolved

Figure 4.15. Discrepancies between pP1 (E-E) sequence and the *GmII* cDNA.

a) Table of differences.

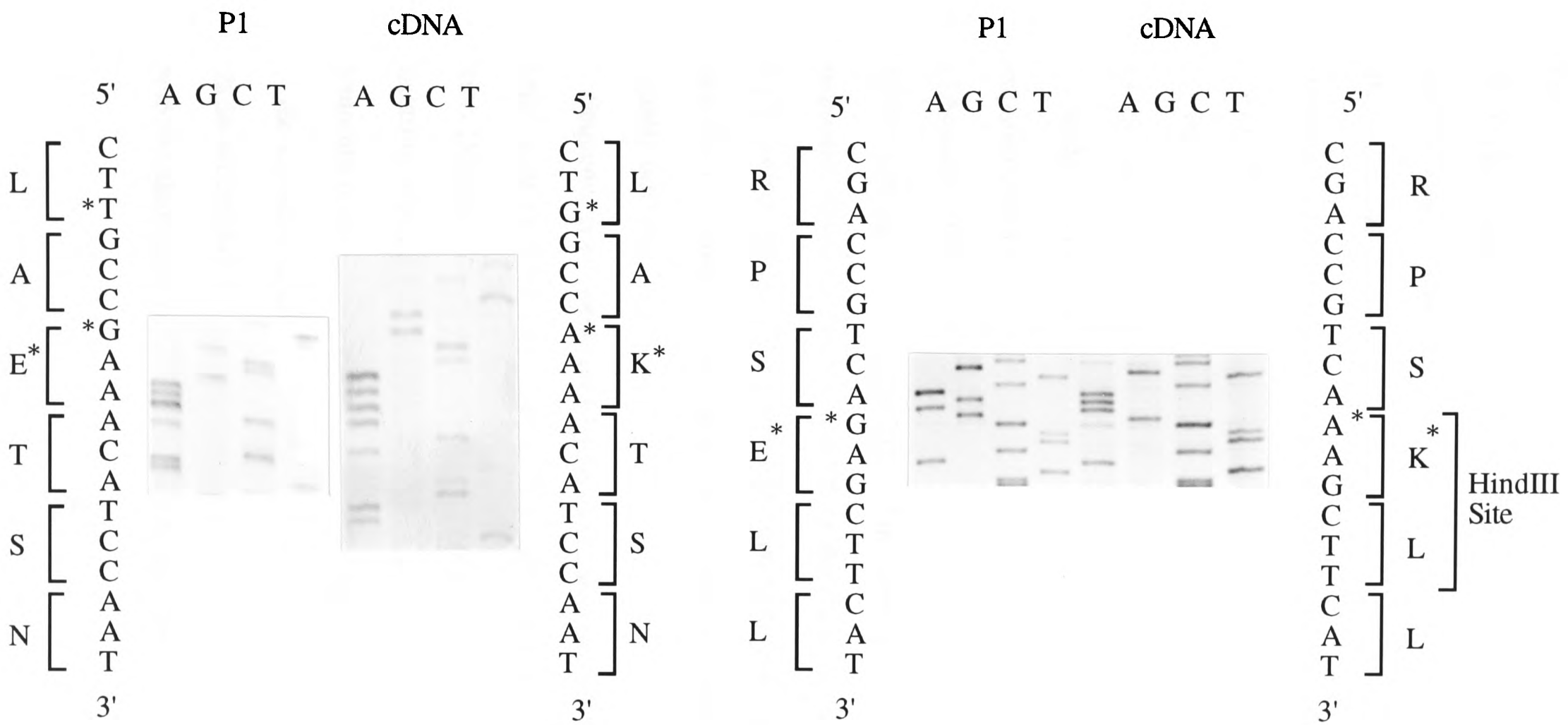


Figure 4.15 (cont).

b) Sequencing autoradiographs showing differences between the *GmII* cDNA and pP1 (E-E) sequences. Amino-acid sequence also shown.

to a lysine in the predicted peptide sequence from the cDNA sequence. This is a semi-conservative change, from acidic to basic, both residues being polar.

The second base pair change (position 3663) removes a HindIII site present in the cDNA sequence (Fig. 4.16a). This polymorphism was also identified by screening inbred stocks. Genomic DNA was prepared {2.3} from different inbred stocks (Appendix II) and digested with HindIII. Screening a Southern blot of this digest with ^{32}P-labelled *GmII* cDNA, {2.7} identified this HindIII restriction fragment length polymorphism. The *GmII* cDNA identified 4 bands in the *BrA* stock and 5 in the *OrR* stock, with the largest *BrA* band (~2.2kb) being cut into ~1.5kb and ~700bp bands (Fig. 4.16b). The pP1(E-E) sequence is of the *BrA* type while the cDNA sequence is of the *OrR* type.

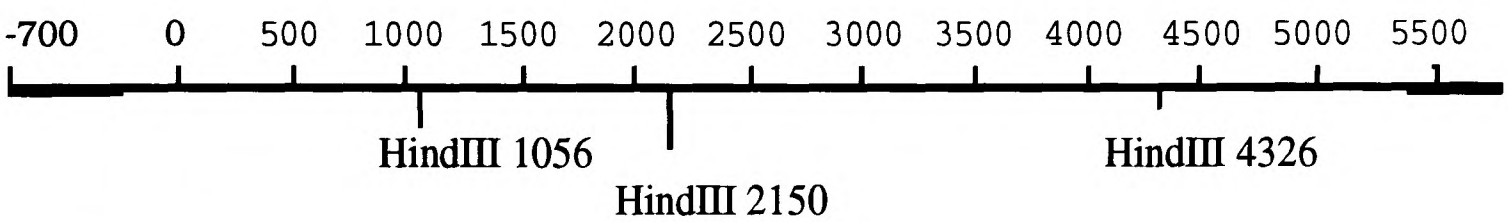
Nucleotide diversity is calculated from the number of base changes per kilobase of compared coding sequence. Estimates of the nucleotide diversity of *Drosophila* genes [Aquadro 1992] is limited since nucleotide diversity can only be investigated by extensive cloning and sequencing. The average nucleotide diversity for *D. melanogaster*, based on data from 0.5% of the euchromatic portion of the genome is 0.004 [Aquadro 1993]. The nucleotide diversity calculated from the differences between the two *GmII* coding sequences is 0.002, less than the average, but not significantly different.

Sequence polymorphism may be due to genetic drift or natural selection, but the contribution of each may be hard to distinguish. The neutral theory of molecular evolution [Kimura 1983] suggests genetic drift to be the most significant with the accumulation of neutral or nearly neutral changes while deleterious mutations are kept to a minimum by natural selection. Selection models take into account a hitch-hiking effect - the co-inheritance of linked nucleotides with the selected allele. This effect predicts an accumulation of variation around sites selected by balanced selection, since these selected alternative alleles are present in the population longer than expected under

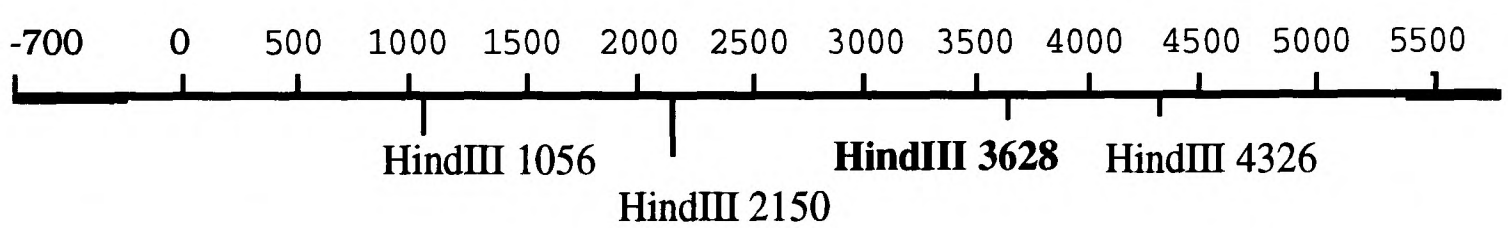
Figure 4.16. HindIII polymorphism in *GmII*.

a) HindIII restriction maps of pP1(E-E) and the cDNA sequence with introns.

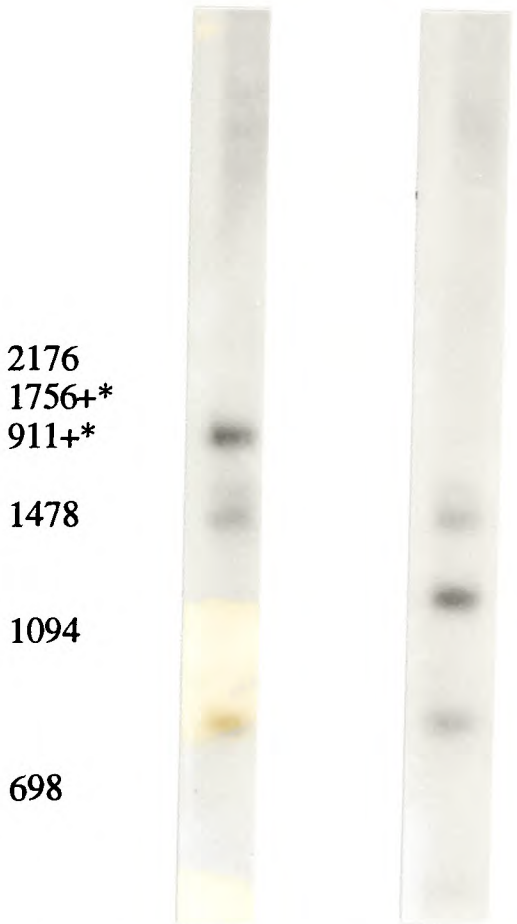
i) pP1(E-E) subclone



ii)cDNA type



b) Genomic DNA from inbred stocks digested with HindIII and probed with *GmII*.



*These fragments sizes are calculated from the EcoRI fragment that was subcloned into pUC18. These fragments would be longer in a HindIII genomic digest. The 5' and 3' fragments of *GmII* cannot be determined from this HindIII digest since the actual size of these end HindIII fragments is unknown.

genetic drift alone, leading to an accumulation of variants at linked neutral sites. Directed selection models predict lower levels of accumulated variation around strongly selected sites as the favoured allele sweeps through the population to fixation, along with its linked variants. The extent of this hitch-hiking effect depends on the level of recombination. The variation in *GmII* appears to be neutral; there are only two base pair changes that affect the amino-acid sequence and the change that these cause is semi-conservative and unlikely to affect the function of the protein.

The distribution of the HindIII restriction fragment length polymorphism in *GmII* can be traced to a certain extent. Among the 8 inbred stocks maintained in our laboratory, six (*S12A, S11B, S11C, BrA, 1AO, IgB*) showed the absence of the HindIII site (cDNA type) and two (*BrB* and *OrR*) contained the extra site (pP1(E-E) type) (results not shown). The origin of the cDNA was from an embryonic cDNA library made from the isogenic second chromosome strain *dp cn bw* [Brown *et al.* 1988] while the pP1(E-E) came from the P1 genomic library made from the isogenic strain of genotype *y; cn bw sp* [Smoller *et al.* 1991] (see section 4.4.1). That these widely dispersed stocks contain the same polymorphism suggests that it is widely spread in wild populations. This is also particularly likely given that the *BrA* and *BrB* stocks, started from two flies caught at the same location (Brightwell-cum-Sotwell, Oxfordshire), contain one allele each. These small inbred populations have been cultured in the laboratory and contain one or other allele, either by fixation by random genetic drift or by a founder effect in which only a single allele was represented in the starting group of a few or single flies.

4.8 Spatial and temporal expression of *GmII*

GmII is expected to be ubiquitously expressed as its role in carrying out *N*-glycosylation trimming is thought to be required in all cells. Expression of genes can be investigated by analysis of mRNA transcripts although this does not necessarily

result in immediate translation into protein (for example in many maternally inherited mRNAs).

4.8.1 Northern blotting

Initial investigations of the expression of *GmII* involved northern blotting. Total RNA extracted from different developmental stages {2.14.1} was separated by electrophoresis, blotted to a nylon filter and probed with *GmII* cDNA {2.15, 2.7}. This revealed the presence of a faint band at 4kb (the expected size) but also showed considerable degradation of the mRNA indicated by large smears in all the lanes (results not shown). The result could not be repeated. An alternative stategy was used.

4.8.2 Analysis of first strand cDNA by PCR

The presence of mRNA was investigated by extracting {2.14.2, 2.14.3} and immediately converting to first strand cDNA {2.16} which could then be used in PCR reactions. One limitation of this approach is that it is not quantitative so levels of expression were not investigated, only the presence or absence of the transcript. The use of different primers in the PCR reaction allows close analysis of differential splicing of transcripts, but again a limitation of this approach is that if differently spliced mRNAs are present simultaneously, perhaps one with an exon missing, this effect would be masked. Despite these drawbacks, the use of first strand cDNA in PCR allowed a rudimentary investigation of the expression of *GmII*.

4.8.3 Where and when is *GmII* expressed?

Primers (DHR3B and DHU2b) were made in exons III and IV, spanning the third intron (Fig. 4.17). The primers were used for PCR {2.17} to investigate first strand cDNA made from embryos (0-22hrs), 3rd instar larvae, pupae, and male and female flies. In addition to this developmental profile, cDNA was made from adult

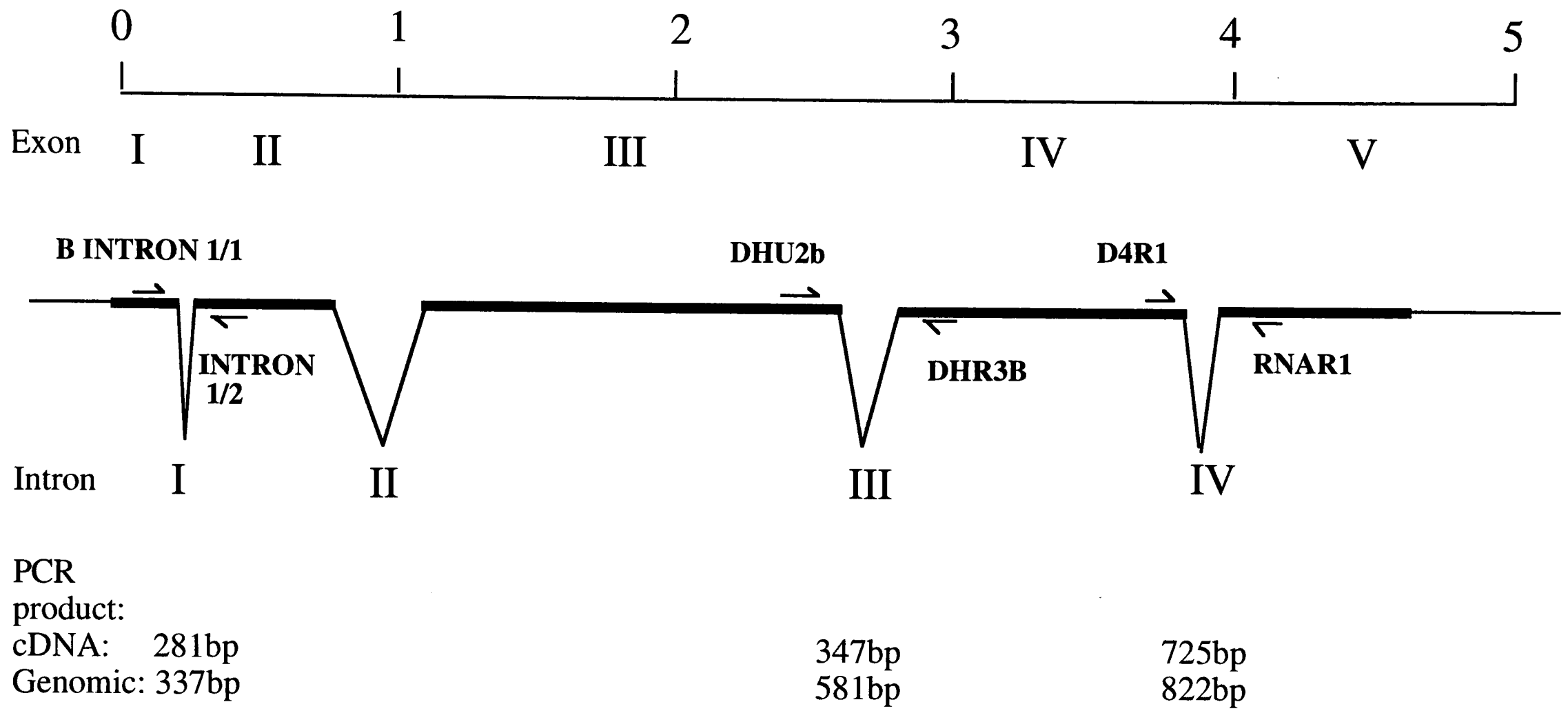


Figure 4.17. PCR primers used to investigate exon expression of *GmII*. Refer to Appendix I for sequences.

heads and bodies separately, to see if the heads, containing a high proportion of neural tissue, showed any difference in expression of *GmII*. The PCR products were separated by gel electrophoresis {2.17.4} (Fig. 4.18a). All the reactions produced the 347bp fragment expected from the cDNA. All the developmental stages show the gene to be expressed and both the head and body cDNA also show *GmII* transcripts.

4.8.4 Is *GmII* differentially spliced?

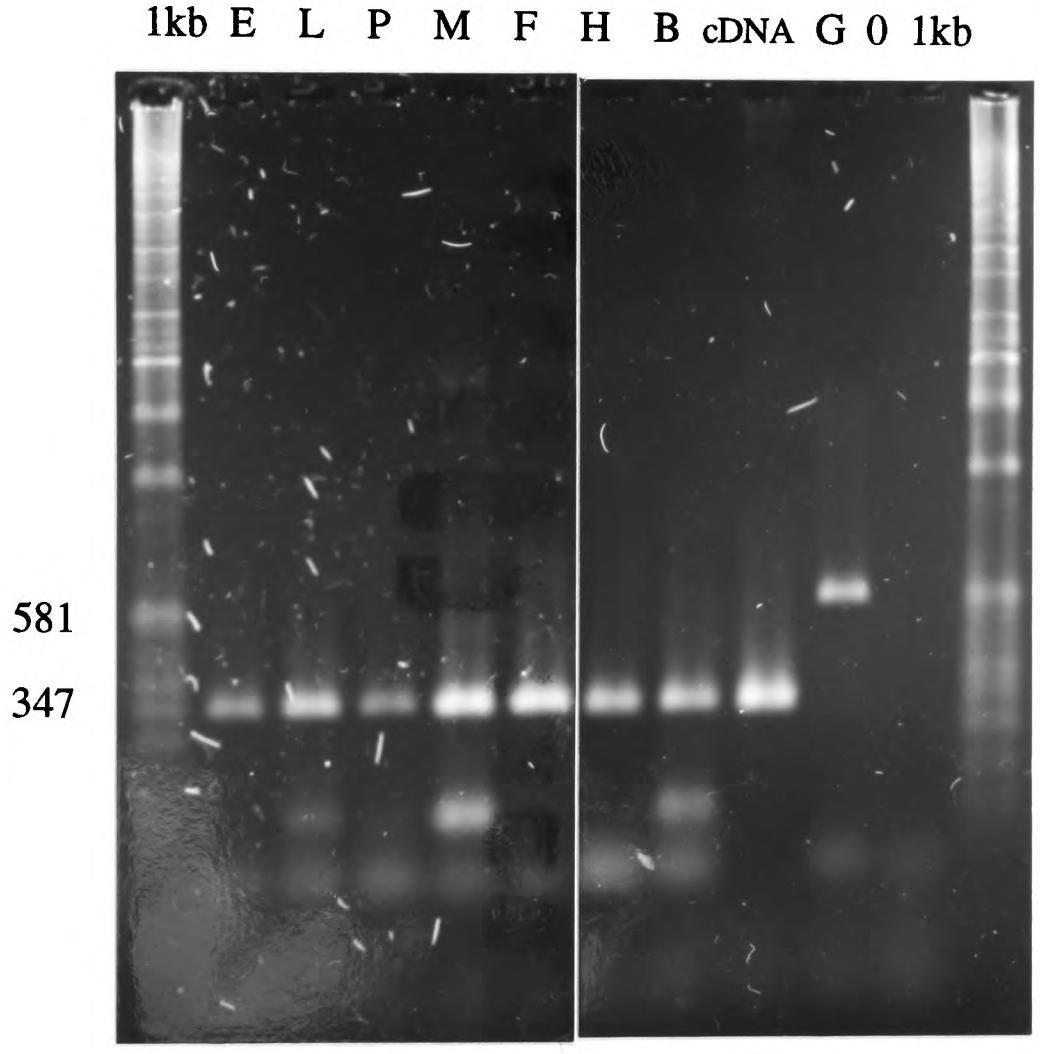
Many *Drosophila* genes show differential splicing during development. This seemed unlikely for *GmII* considering: i) its function as a membrane bound catalytic enzyme is unlikely to be carried out by different domains; ii) only one positive cDNA had been identified after analysing the clones from the embryonic library; iii) the presence of only one band on the northern blot (although unrepeated) and iv) the sequenced *GmII* contained relatively few exons and appears to have no other possible exons either contained in the introns or in the 3' sequence. However differential splicing was investigated by designing primers in each exon. Since exons III and IV had already been investigated and shown to be positive in each stage and tissue investigated, primers were made spanning intron I (B INTRON 1/1 and INTRON 1/2) and intron IV (D4R1 and RNAR1) (see Fig. 4.17). PCR on the same cDNA demonstrated that all exons are expressed at all developmental stages and tissues investigated (Fig 4.18b-c). The faint band seen in Figure 4.18b is not due to contaminating genomic DNA as it is absent from the other PCRs on the same cDNA samples and is probably due to non-specific primer binding. There appears to be no differential splicing detectable within the limitations of this method.

4.9 Discussion and conclusions

The *GmII* cDNA clone has been used to identify the genomic sequence, which has been cloned and investigated. The results of this investigation show that *GmII* has

Figure 4.18. PCR gels of first strand cDNA from different developmental stages and tissues (E-embryos, L-3rd instar larvae, P-pupae, M-male adult flies, F-female adult flies, H-heads, B-bodies), *GmII* cDNA (cDNA) genomic DNA (G) and no DNA control (0). (PCR annealing temperature 62^{o}C). For primers see Fig. 4.17.

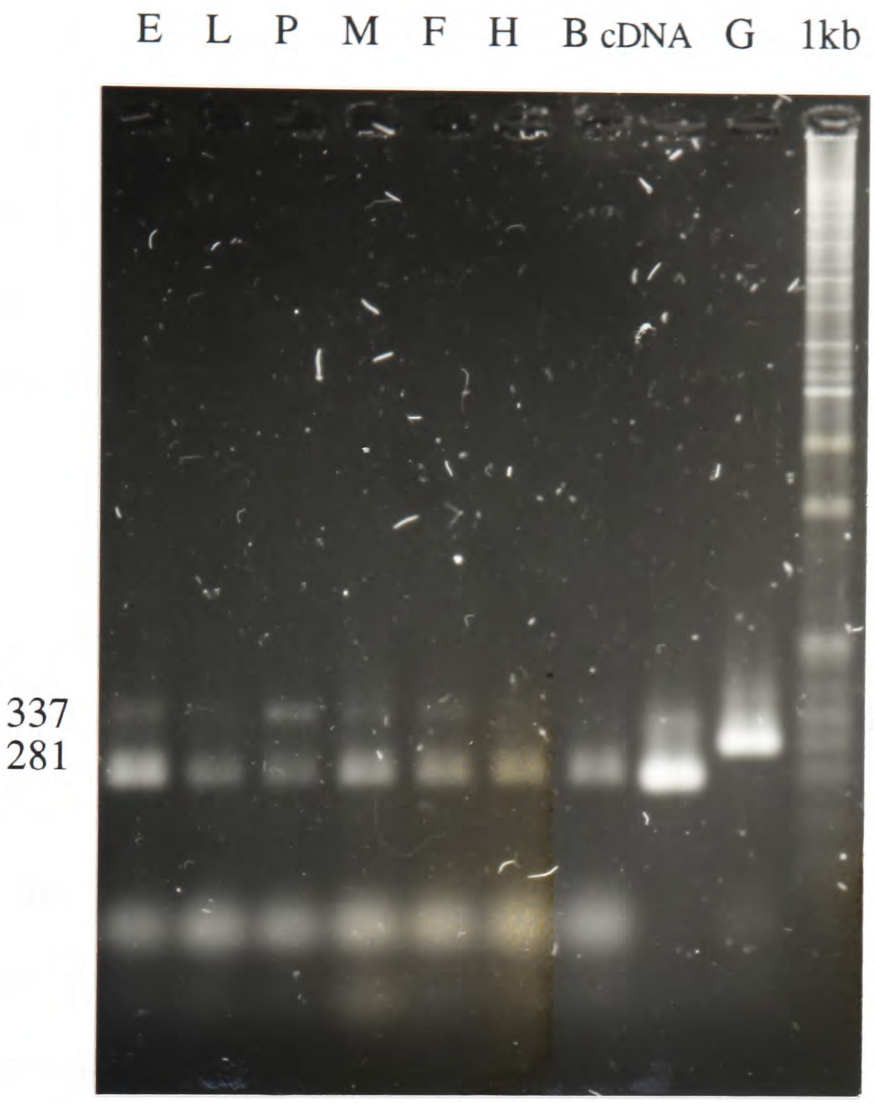
a) Primers in exons III and IV.



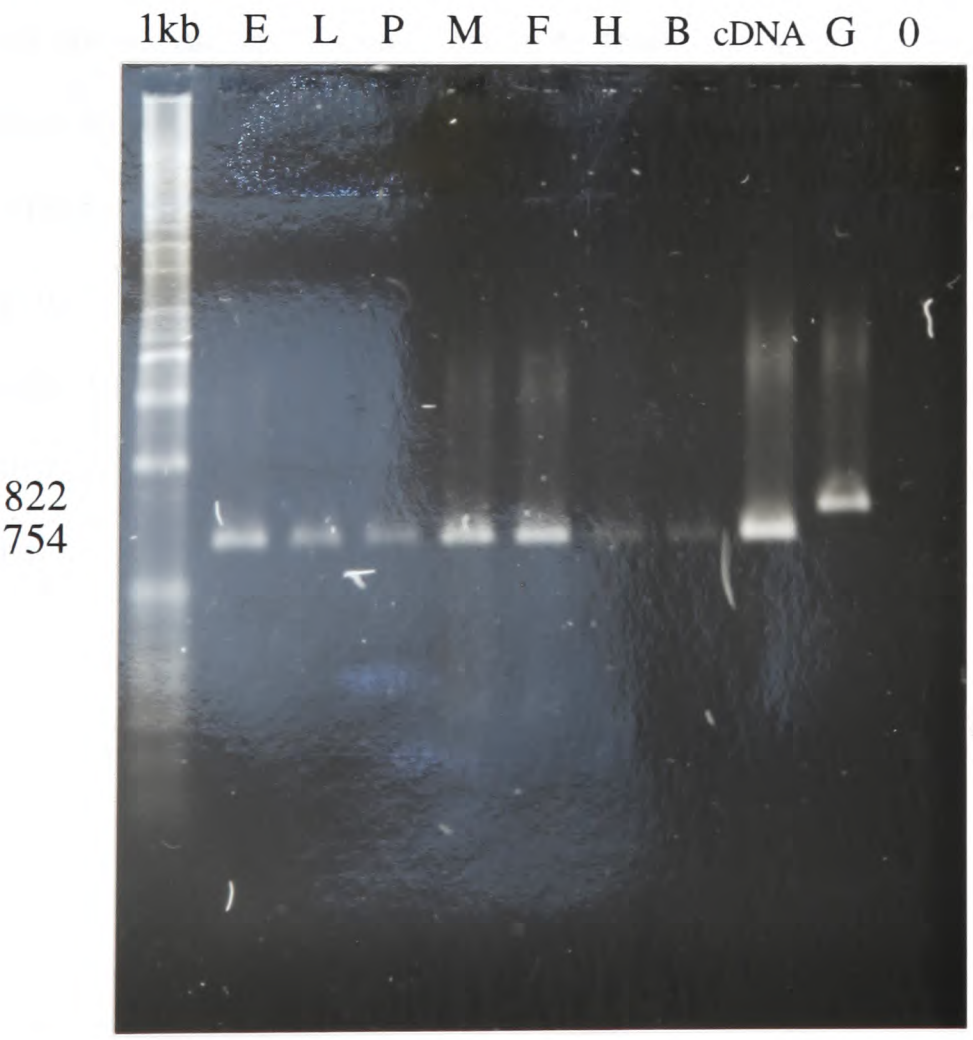
NB: An appropriate control to confirm the results shown in Fig. 4.18a-c would be to perform PCR on a first strand cDNA reaction prepared at the same time with no mRNA. This would control for cloned *GmII* cDNA sequences contaminating the first strand cDNA reaction.

Figure 4.18 (cont). PCR gels of first strand DNA.

b) Primers in exons I and II.



c) Primers in exons IV and V.



a simple gene structure, with the coding region split into 5 exons. The 4 identified introns demonstrate the expected splicing consensus sequences and do not contain any open reading frames. The 5' sequence in the pP1(E-E) subclone has been shown to be correct, using PCR on genomic DNA, but does not demonstrate any obvious expression control sequences. Further investigation is needed to examine how the expression of *GmII* is controlled (see Chapter 6). *GmII* transcripts containing all exons have been shown to be expressed at all developmental stages and tissues investigated.

The gene structure of *GmII*, while interesting in its own right, has not revealed any clues to the biological significance of the protein. The detected expression of the gene appears to be ubiquitous both spatially and temporally at the level of the resolution to which it has been investigated. Two further lines of investigation were undertaken to provide more information about the biological significance of the *GmII* and of *N*-linked glycosylation; the identification and rescue of a *GmII* mutant and an investigation of the promoter of the gene in conjunction with examining the expression of the gene at greater resolution.

The genomic clone of *GmII* discussed in this chapter was required for the transformation rescue of *GmII* mutants from a saturation mutagenesis experiment undertaken by other members of the laboratory to produce mutants in the 85D region. This is discussed in Chapter 5. The promoter region identified in this chapter was also investigated further by the fusion of various lengths of the sequence 5' to the *GmII* transcribed region to a reporter gene. The use of a reporter gene also allowed the expression of *GmII* to be visualised directly. These experiments are discussed in Chapter 6.

CHAPTER FIVE
GOLGI MANNOSIDASE II : FLY TRANSFORMATION

5.1 Aims and introduction

As stated before one of the approaches taken by this laboratory to investigate *N*-linked glycosylation is to identify mutants of the genes involved and to investigate their phenotype. A saturation mutagenesis experiment was undertaken by B. Yudkin and W. Mulvaney to find mutants of *GmII* [Yudkin 1996]. Mutations were induced by ethyl methane sulphate (EMS). Exposed flies were crossed to a stock of flies carrying a deficiency in 85D balanced by a duplication chromosome (Fig. 5.1a). The deficiency was made by recombining two chromosomes carrying inversions in the region. This was necessary since the deficiency uncovers a haplo-insufficient region resulting in sterility [Kemphues *et al.* 1983]. Lethal mutants were identified (Fig. 5.1b). A mutation in *GmII* was assumed to be lethal due to the fundamental level at which this enzyme is thought to act. This view is also reinforced by the discovery that mice mutant for *N*-acetylglucosaminyltransferase I, the enzyme immediately before *GmII* in the *N*-linked glycosylation degradation pathway (see section 1.4.5) die after 10 days of development [Ioffe *et al.* 1994; Metzler *et al.* 1994].

Mutants of *GmII* can be identified by their rescue with a chromosome carrying the transformed *GmII* gene. For this reason, and to allow other investigations into the expression of the gene, a transformed line carrying *GmII* was made.

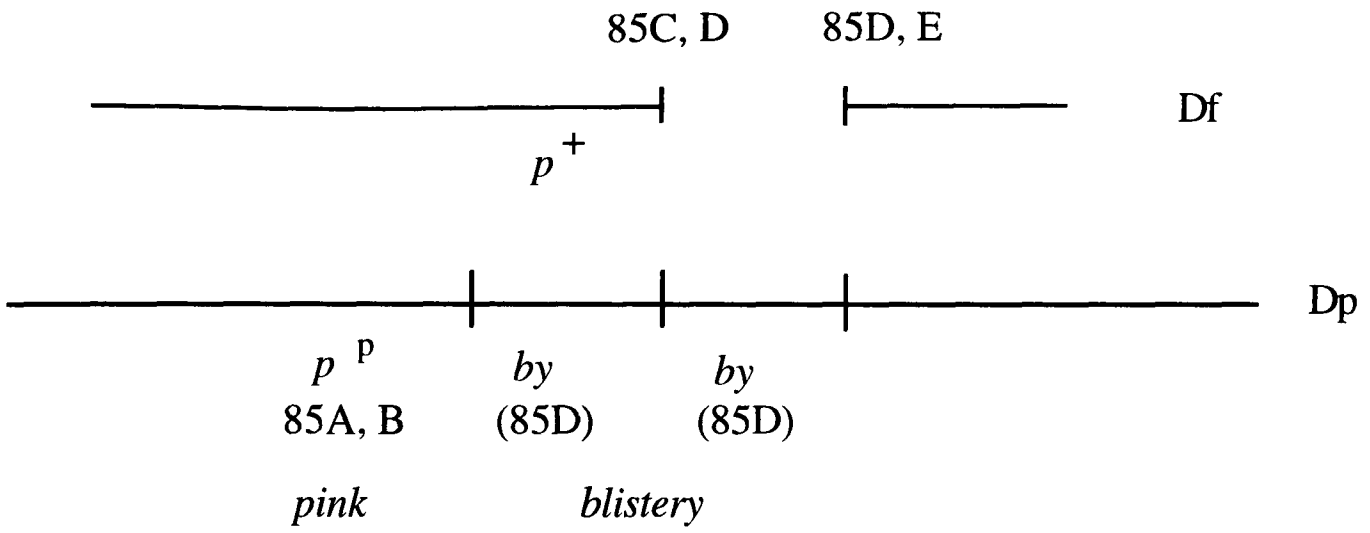
5.2 P-element mediated transformation of flies

5.2.1 Theory

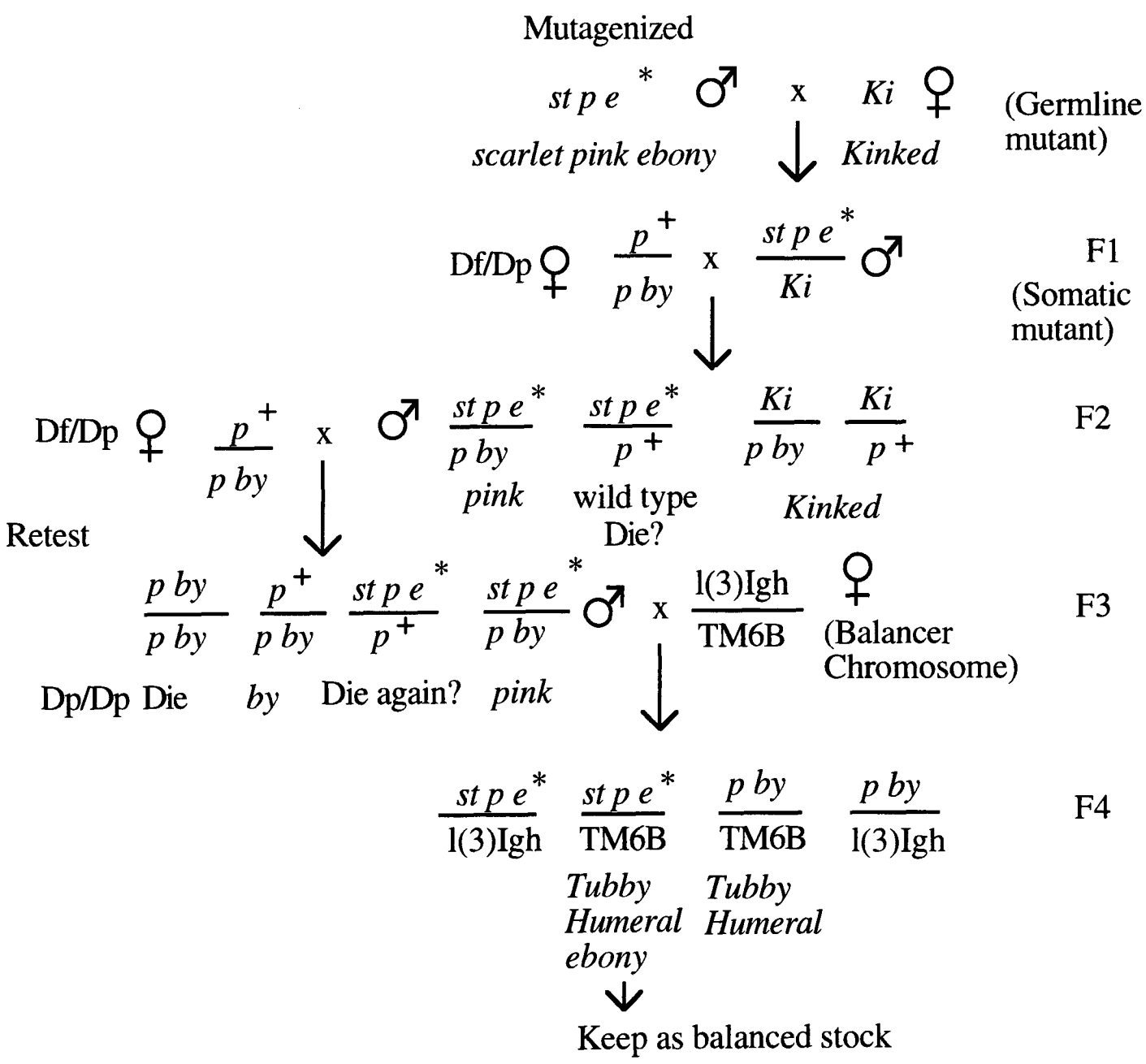
Drosophila transformation is the stable integration of exogenous DNA into the germline [Ashburner 1989a]. The application of P-elements for transformation makes use of a naturally occurring integration mechanism. P-elements are 3kb transposable

Figure 5.1. Df/Dp (Deficiency/Duplication) and mating scheme to identify mutants.

a) A section of the third chromosome of the Df/Dp stock.



b) Identifying mutants using Df/Dp.



elements with short (31bp) inverted terminal repeats [Rubin *et al.* 1982]. Mobile P-elements encode for transposase functions that allow excision or integration of the element into the genome. This probably requires a target sequence similar to GGCCAGAC, which is duplicated upon insertion [O'Hare *et al.* 1983].

P-elements are also found in natural P strains of *Drosophila* with deletions that prevent the production of their own transposase, although capable of using transposase functions coded in other intact elements. This allows the construction of vectors retaining the P-element arms required for integration but without the transposase coding function. Exogenous DNA can be inserted in its place in the plasmid and thence into the genome upon integration [Rubin *et al.* 1982].

5.2.2 Vectors

Vectors for introducing exogenous DNA into the genome have been constructed from native P-elements. One example is pCaSpeR (Fig. 5.2) which has the P-element flanking sequences to allow its insertion into the genome. Such vectors require a minimum of 50bp flanking sequence on either side of the P-element arms and a minimum of 540bp and 240bp from the 5' and 3' ends respectively of P-element sequence [Pirrotta 1988]. The vector also has a multiple cloning site containing an EcoRI site into which exogenous DNA can be cloned. The vector has some pUC8 sequence containing the bacterial resistance gene for Ampicillin which allows the growth and selection of the recombinant plasmid in preparation for its microinjection into embryos. This sequence is not integrated into the host's DNA.

In order to select transformants the vector also contains a *white*$^{+}$ "mini gene" which when transformed into *white*$^{-}$ hosts gives pigmented eyes. This is a modified form of the *white* gene that has had a HindIII-XbaI region removed from the long 5' intron. An SpeI fragment (4.14kb) of this construct containing ~300bp 5' and ~630bp 3' to the *white* gene has been inserted into the EcoRI site of the Carnegie 4 P-element

Figure 5.2. P-element vector pCaSpeR (Pirrotta, 1988).

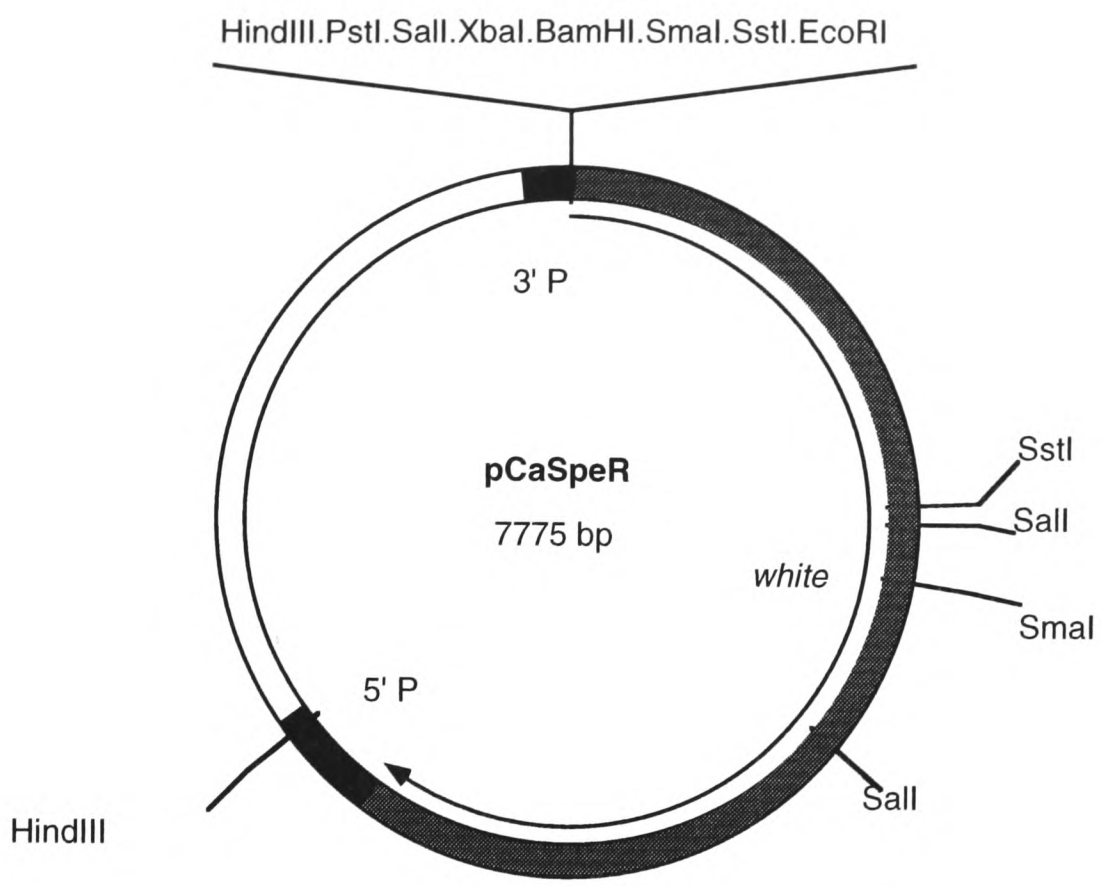
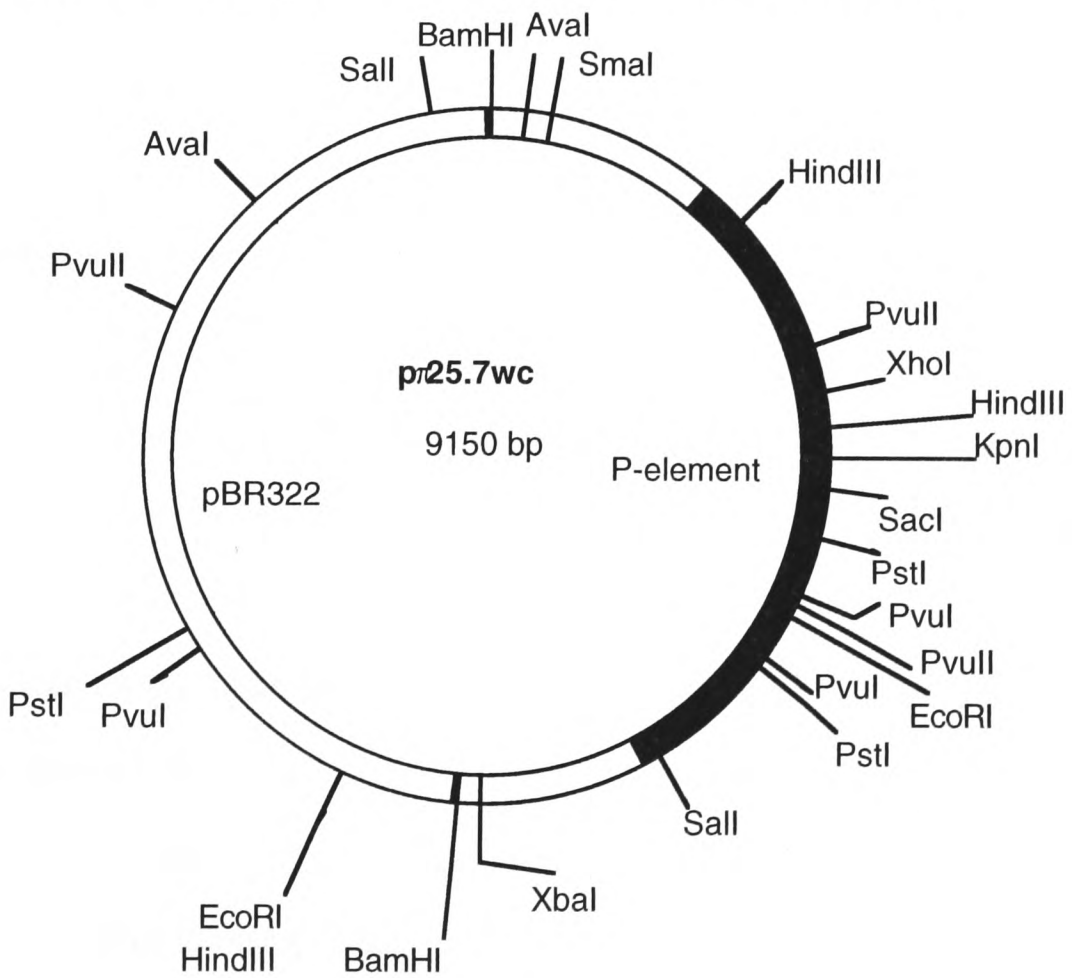


Figure 5.3. pπ25.7wc - helper P-element plasmid (Karess, 1984).



vector [Rubin *et al.* 1983]. This has the effect of both reducing the length of the construct and in reducing the expression of *white*. The first is an advantage in increasing transformation efficiency which is dependent on construct length [Spradling 1986]. The second means that transformed flies often have only pale pigmented eyes, which, although it may make them harder to identify, allows an estimation of insert number distinguishing hemizygotes from homozygotes [Klemenz *et al.* 1987].

5.2.3 Helper plasmids

The integration of the P-element requires transposase, normally encoded by the P-element sequence. Where this has been removed the transposase function must come from a second, transposase functioning element, simultaneously introduced. One such helper plasmid is pπ25.1 [O'Hare *et al.* 1983]. One problem with this plasmid is that it may also become integrated and continue to produce transposase destabilizing the transformed line. To over come this problem pπ25.7wc ("wings clipped") has had the last 23bp of the 3' inverted repeat sequence deleted to prevent insertion (Fig 5.3) [Karess *et al.* 1984].

5.3 *GmII* construct for injection

5.3.1 3' end modified

The transcript from the transformed construct needed to be distinguished from the transcript of the native gene so that expression of the transformed gene could be demonstrated. The untranslated 3' end of the gene was altered to produce a truncated transcript that could then be identified by PCR analysis of first strand cDNA from extracted mRNA. The enzyme BsaBI cuts twice in the 3' end, but nowhere else in the gene (Fig. 5.4). The pP1(E-E) subclone was digested with BsaBI, and religated to produce a plasmid lacking 151bp (pP1(E-E(Δ151))). The plasmid was sequenced to check it was missing the fragment (results not shown). This manipulation assumes that

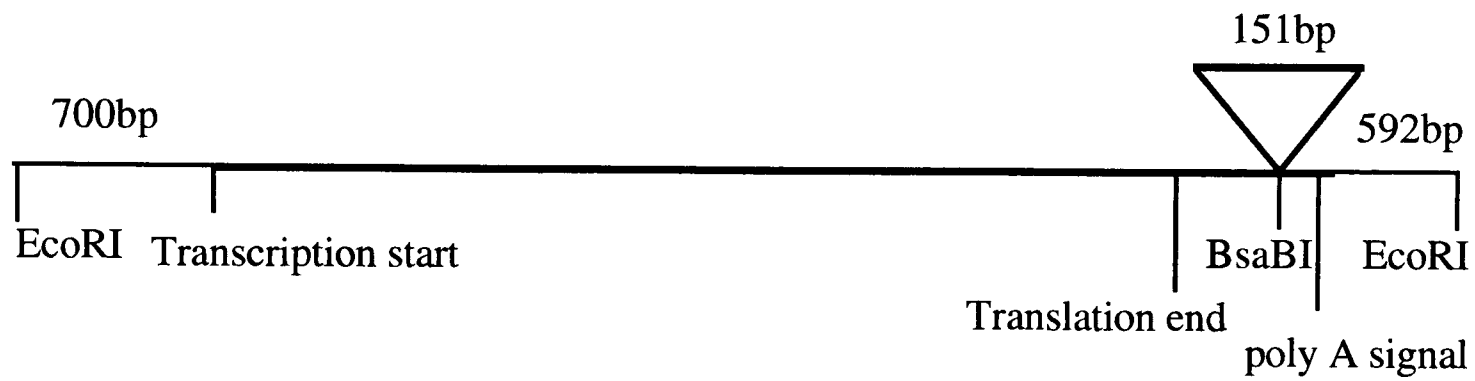


Figure 5.4. pP1(E-E(Δ151)) - modified pP1 (E-E) used for transformation.

altering the 3' untranslated region would have no effect on the expression of the gene. The 3' end of genes is not significant in splicing since splicing can occur simultaneously to transcription before the 3' end is complete [Krainer *et al.* 1984; Beyer *et al.* 1988]. The poly-adenylation consensus sequence and site (see section 4.6.3) were not affected by the modification.

The entire gene was contained in the pP1(E-E) subclone, with 700bp upstream and ~600bp downstream flanking sequences. This insert with the modified 3' end, was removed from pUC18 by digestion with EcoRI, and purified from a gel {2.6}. It was then ligated into the EcoRI site of the P-element transformation vector pCaSpeR (see section 5.2.2) and used to transform *E. coli* {2.11}. The construct was called pP[*white*$^+$ *GmII*(E-E(Δ151))].

5.3.2 The 5' end of the construct

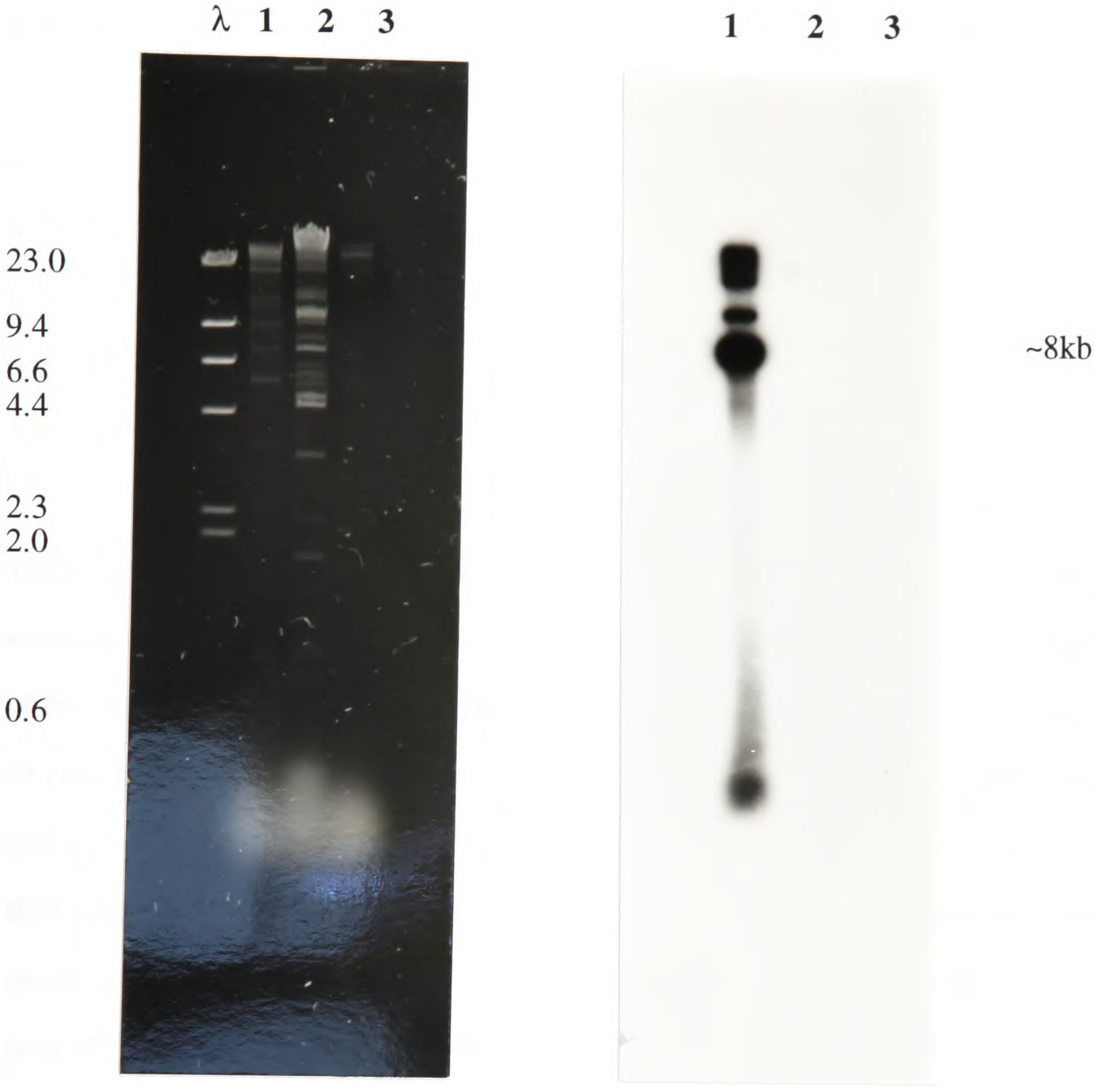
The pP[*white*$^+$ *GmII*(E-E(Δ151))] construct contained 700bp of sequence 5' to the start of *GmII* transcription. This length of DNA should be enough to contain the promoter sequences which are normally found in the first 100bp [Bucher 1990]. However there could be other essential upstream sequences, such as the tissue specific enhancer sequences of the *white* gene [Pirrotta *et al.* 1985]. Was 700bp enough? Only by injecting and testing the expression of the construct would we know. In the meantime it seemed sensible to make a construct with a longer 5' region.

The initial investigation of the P1 clone **1** (section 4.4.1) identified two BamHI fragments when probed with *GmII*. This result is confirmed by the presence of only a single BamHI site (at 4387) in the pP1(E-E) sequence. BamHI digests of P1 **1** were probed with DROS 2, the *GmII* 5' subclone (see Fig. 4.1). This identified a ~8kb fragment at the 5' end of *GmII* (Fig. 5.5), which would contain ~4kb of 5' sequence. Plans were made to clone this into pP[*white*$^+$ *GmII*(E-E(Δ151))]. For this the orientation of the P1(E-E) insert in pCaSpeR needed to be checked since the cloning of

Figure 5.5. P1 clones digested with BamHI.

a) Electrophoresis gel.

b) Probed with DROS 2.
(washing conditions:2xSSC 65°C)



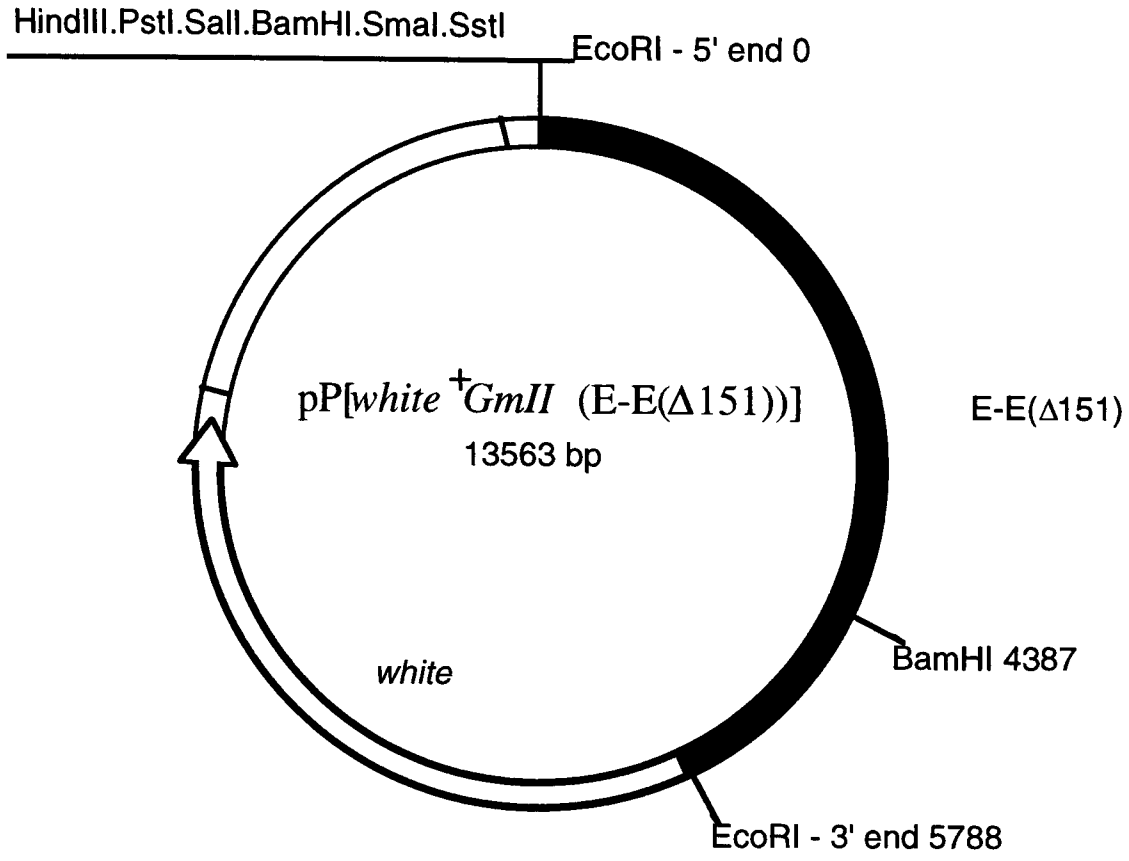
an insert with identical termini meant that it could have ligated into the vector either way. The orientation of the P1 (E-E) fragment in pCaSpeR was demonstrated by digestion with BamHI (Fig. 5.6a) and running the digests on a 0.8% electrophoresis gel (Fig. 5.6b). Three out of the four clones tested were in the correct orientation (Fig. 5.6ai) to insert the BamHI P1 **1** fragment. The orientation of this fragment would subsequently need to be investigated in a similar way.

While this construct was being made, pP[*white*$^+$ *GmII*(E-E(Δ151))] was transformed into flies (section 5.4) and its expression demonstrated (section 5.6 and 5.7). This larger construct was set aside.

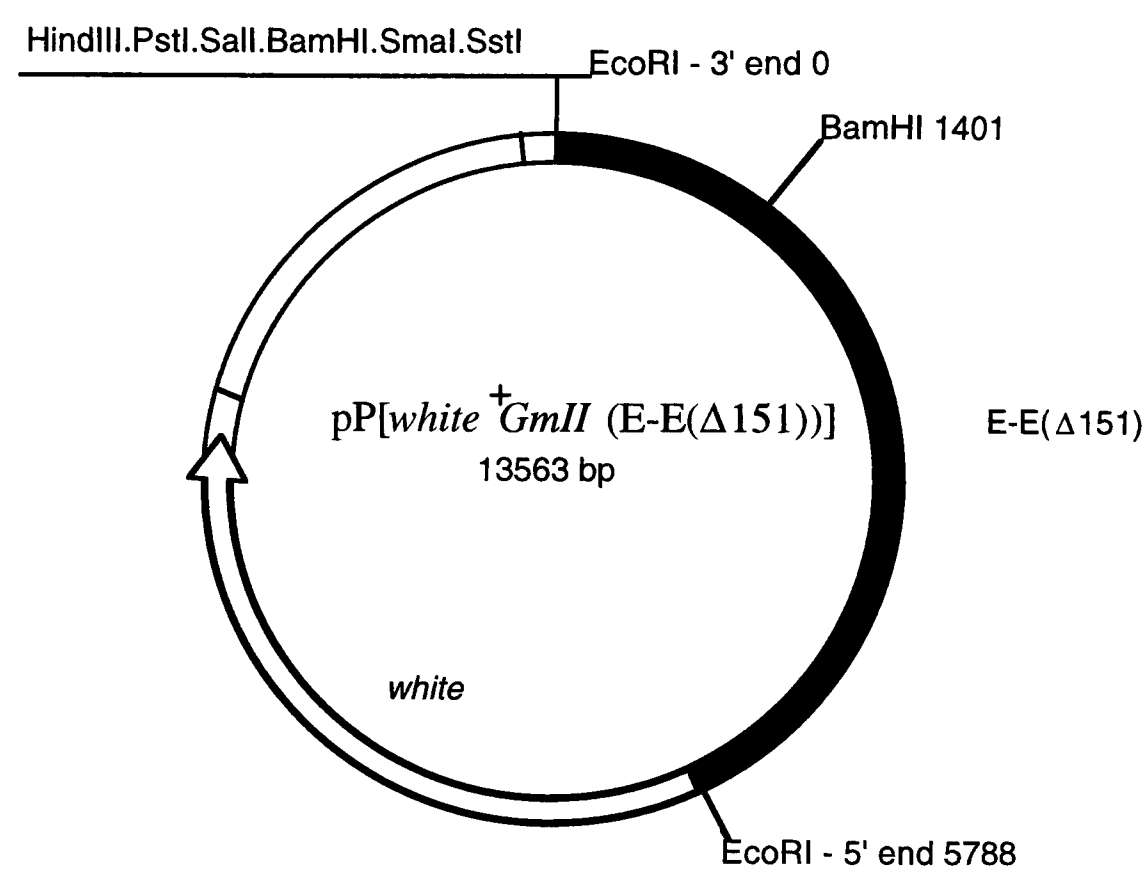
5.4 Transformation of flies with *GmII*

5.4.1 Microinjection

Plasmid DNA was microinjected into *Drosophila* embryos [Spradling *et al.* 1982; Spradling 1986]. The strain of flies used was *yellow white*. These flies have white eyes which allow the identification of pigmented transformants by the expression of the mini*white*+ gene from pCaSpeR. The early embryo in *Drosophila* is a syncytium of rapidly dividing nuclei that do not become enclosed in cellular membranes until 3 hours of development at 25^{o}C. The germ cells in *Drosophila* consist of a dozen nuclei that migrate to the posterior pole and enclose in cellular membranes after about 90 minutes of development at 25^{o}C. The exogenous DNA must be introduced prior to this pole cell formation. Therefore embryos are injected 1.5-2hrs after laying {2.18} with a solution containing the construct (300-1000μg/ml) and pπ25.7wc (50-150μg/ml), [Spradling 1986]. The flies were allowed to continue development. Survivors were crossed with virgin *yellow white* flies and their progeny screened for pigmented eyes.



i) BamHI digest : 9158bp, 4405bp.

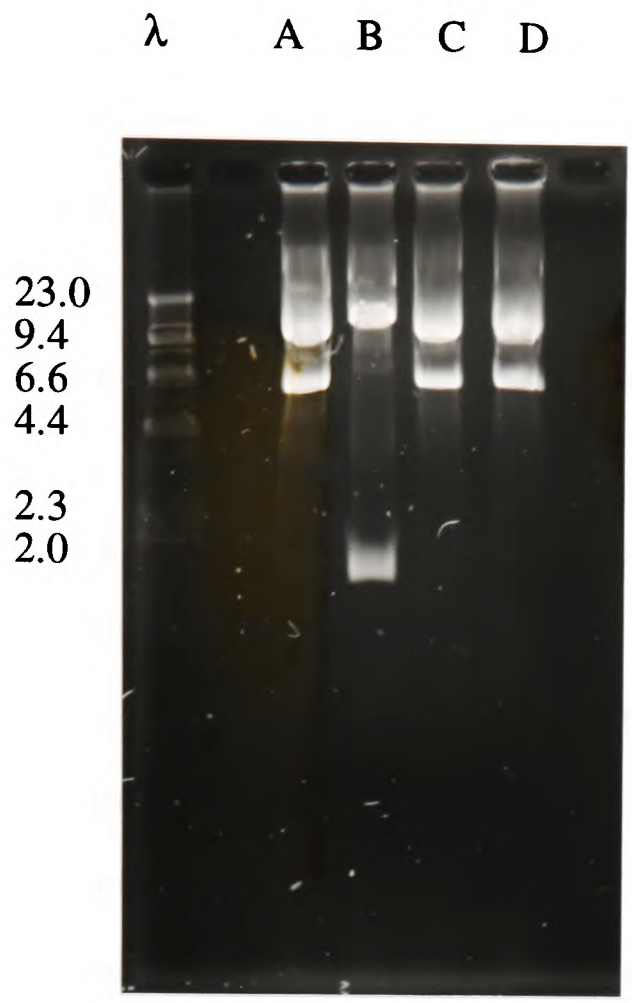


ii) BamHI digest : 12144bp, 1419bp.

Figure 5.6. a) RE map of 2 possible orientations of the P1 (E-E(Δ151)) insert in pCaSpeR.

Figure 5.6 (cont). Checking the orientation of the P1 EcoRI fragment in pCaSpeR.

b) Gel showing separate clones digested with BamHI.



5.4.2 Transformation efficiency

Figure 5.7 shows the survival data for injecting flies with pP[*white*$^+$ *GmII*(E-E(Δ151))]; the numbers (Fig. 5.7a) and percentages (Fig. 5.7b) surviving from each stage. In this first experiment the survival rate of injected embryos was very poor. A number of measures were undertaken to improve this. The length of time that the embryos are dried is very important - too much and the embryos shrivel and die, too little and they burst upon injection. The length of time for desiccation must be empirically determined. The amount of oil covering the embryos must also be judged carefully; too much and the embryos lack oxygen and die, too little and they dry out. The needle shape is also critical, it must be pulled sufficiently thin to make only a small hole upon entry, but if too long, it makes manipulation difficult. Again, only by altering the settings on the needle puller by trial and error was the correct setting found. The temperature of injection may also be important. Many laboratories inject at 18^{o}C to slow the development of the embryo and reduce leakage of cytoplasm [Ashburner 1989a]. Injections were carried out both at room temperature (20-25^{o}C) and at 18^{o}C but the survival rate did not improve at the lower temperature. Lastly the skill of the injector is critical and patience and practice is required.

The percentage of fertile crosses was also quite low. Sterility is caused by damage to genital and abdominal structures. This disruption of the pole region of embryos occurs during injection and is also dependent upon the skill of the injector.

The efficiency of transformation decreases with increasing length of construct [Spradling 1986]. The pP[*white*$^+$ *GmII*(E-E(Δ151))] construct contains ~10.6kb of exogenous DNA between the P-element arms, which is a relatively long transformation construct. The percentage of transformed flies in the fertile crosses was just over 20%. The expected transformation efficiency for this length, according to Spradling (1986), is in the region of 5-35% of fertile off-spring.

Figure 5.7. Tables of transformation efficiency

a) Survival data.

P-element construct	pP[*white*$^+$ *GmII*(E-E(Δ151))]
Embryos injected	2049
Larvae	95
Pupae	69
Flies (crossed to *yw*)	66
Fertile crosses	23
Transformed vials	5

b) Survival percentages.

Injected embryo survival to larvae	4.6%
Larval survival to flies	69.5%
Fertile crosses	34.8%
Fertile vials transformed	21.7%

5.5 Localization of inserts

The transformation experiment produced 5 vials with red-eyed transformants. As many flies as possible from these vials were crossed individually to *yw* flies {2.19} (Fig. 5.8).

5.5.1 Sex chromosome insertion

Several lines could be immediately identified as having the P-element insertion on the X chromosome since from parental red-eyed males in G1 it is possible to deduce that the insertion is on the X chromosome when the female progeny is pigmented and all the males are white eyed (Fig. 5.8a). In some cases for female red-eyed parents a difference in expression between the sexes in their progeny also indicated if the insertion was sex-linked (see section 5.5.6). Stocks were set up according to the scheme shown in Figure 5.8a.

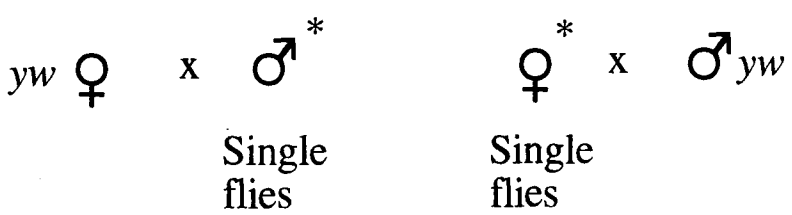
5.5.2 Autosomal insertions

Some lines were identified as carrying autosomal insertions, by mixed pigmented progeny from a pigmented male, either immediately or after further crosses. Flies were crossed to CyO/Gla virgin females and TMI/TM3 virgin females in two separate crosses per strain according to the scheme shown in Figure 5.8bi and ii to discover on which autosome the gene was inserted. These stocks have marker mutations for each chromosome. CyO/Gla flies have a curly wings mutation on one chromosome II and glazed eyes mutation the other second chromosome. TM1/TM3 flies have moiré eyes (phenotypically shiny like shot silk) and serrate wings mutation one on each third chromosome.

Figure 5.8. Establishing lines of transformed flies. Not all the progeny from each cross are shown. Pigmented flies are marked * hemizygous and ** homozygous for a transformed chromosome.

G0 - injected flies crossed to yw

G1 - Pigmented flies* crossed to yw



a) X chromosome insertions.

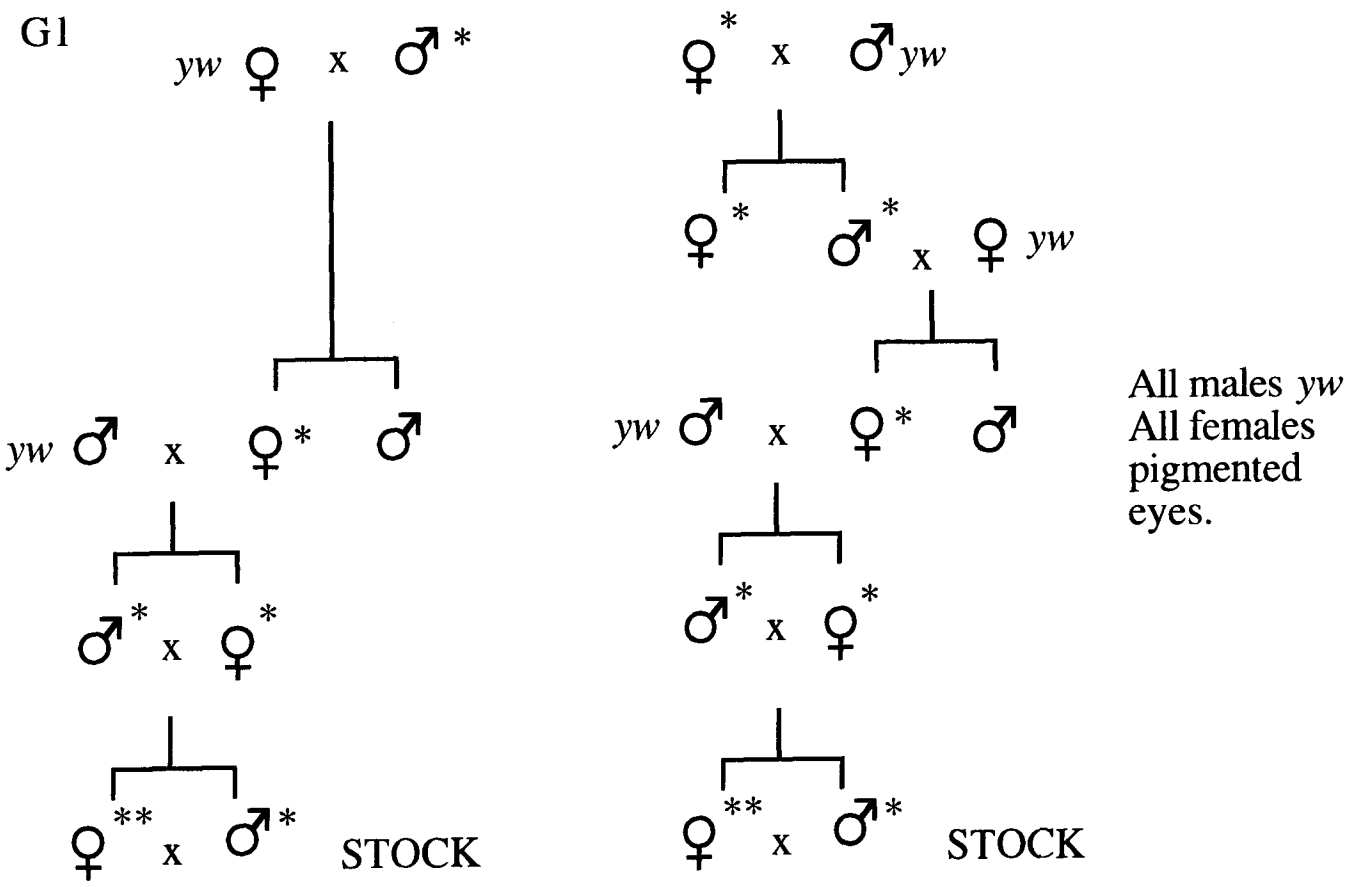
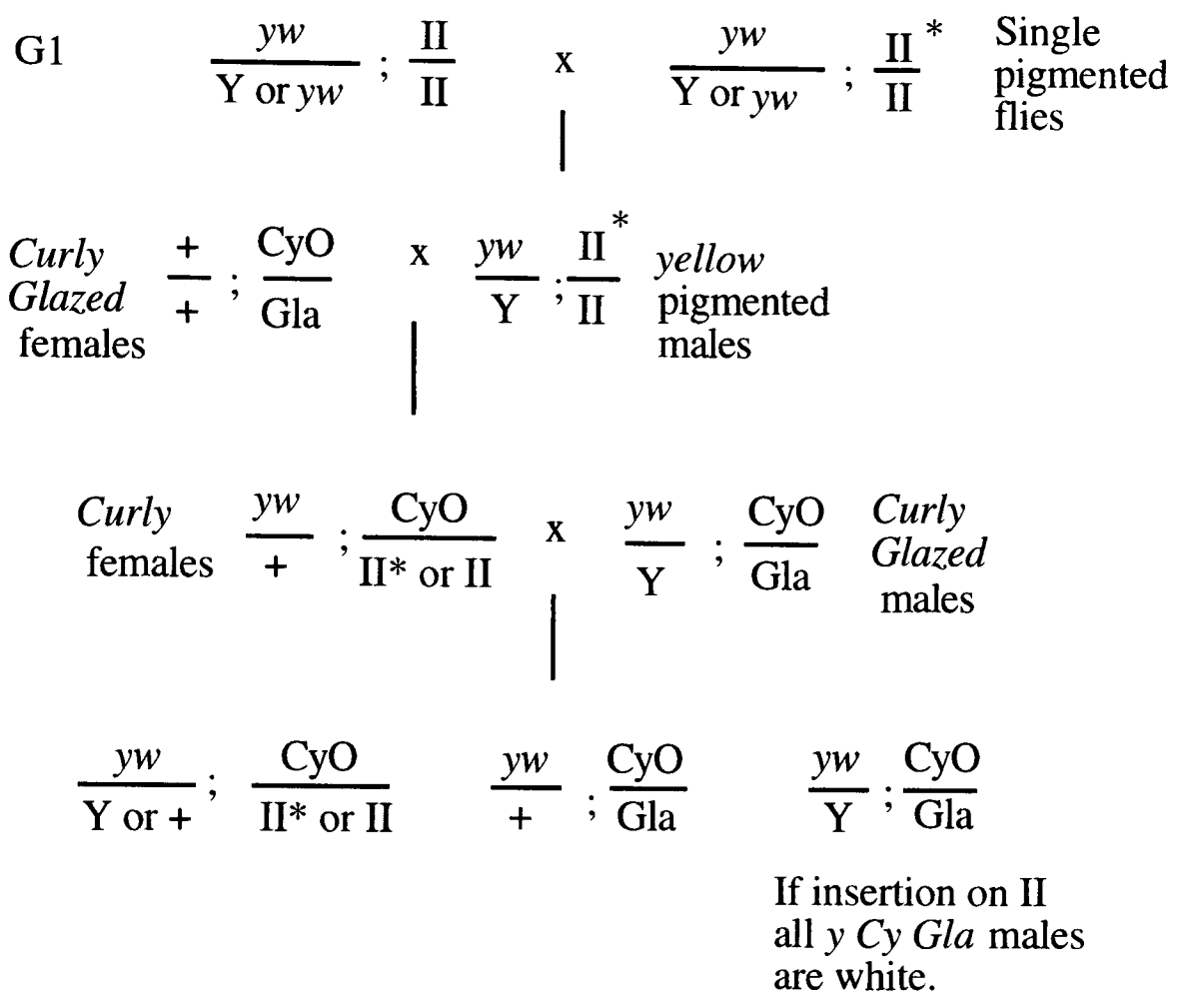


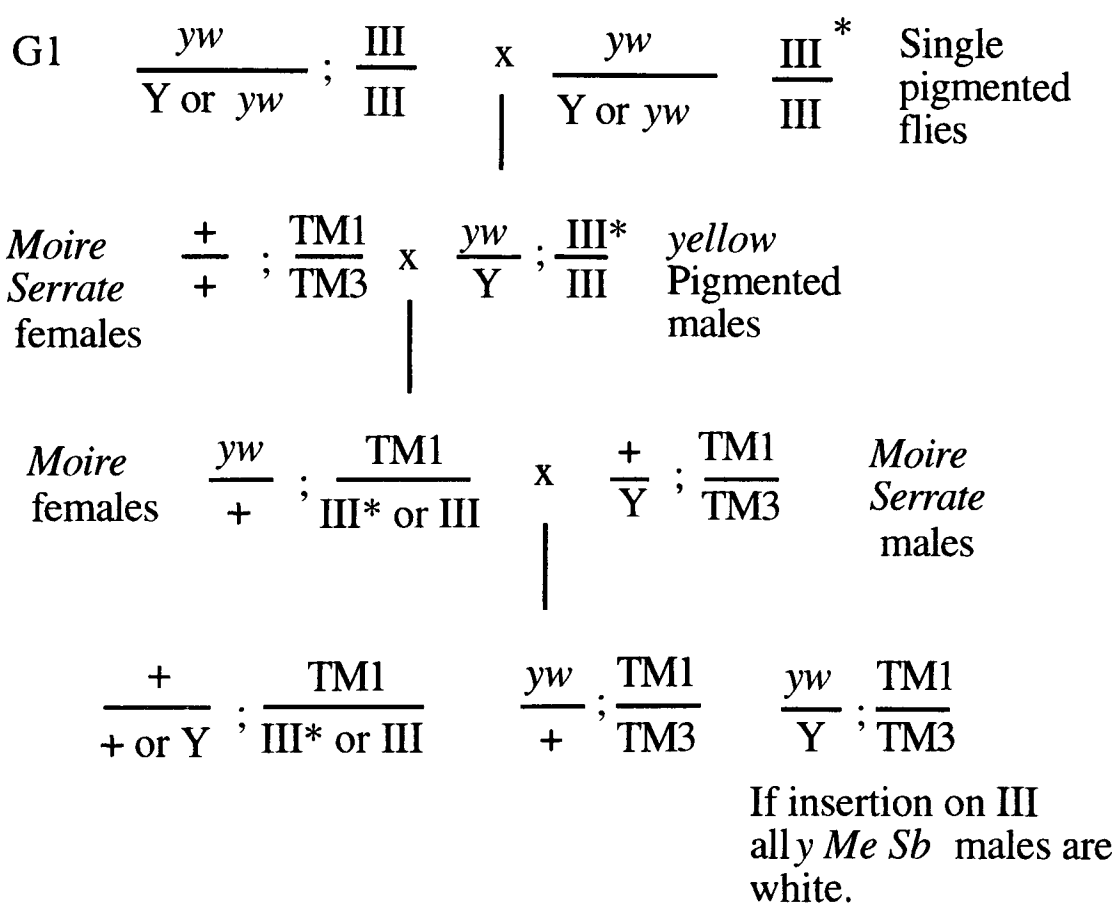
Figure 5.8 (cont).

b) Autosomal insertions.

i) Cross to CyO/Gla.



ii) Cross to TM1/TM3.



5.5.3 Summary of stocks

Stocks characterized for transformants were set up (Fig. 5.9). The stocks are named *yw* (;) P[*white*$^{+}$ *GmII*(E-E(Δ151))] X/II/III (depending on insert chromosome location) and alphabetically. They are hereafter referred to by their short name given in Figure 5.9. Where possible all the stocks were maintained as homozygous for the chromosome carrying the P-element. These were set up by crossing deeply pigmented males and females, the presence of two copies of the mini*white*+ gene being distinguishable from one. Only one stock from each independent transformed line was maintained. Transformants from the same original G0 vial could have the same P-element insertion site; only insertions on different chromosome could be guaranteed to be independent, as could insertions from different G0 vials.

Two stocks carried homozygous lethal insertions, presumably due to the insertion interrupting some vital gene. The stock **II B** was maintained over Gla and **III A** over TM1.

The eye colour of the different transformed lines varies between bright red and pale orange (Fig 5.10). This variation could be due to the number of P-element insertions or due to position effects (see section 5.5.5). The eye colour for the homozygous lethal insertion stocks is reduced by their only carrying the insert on one chromosome and distorted for **II B** by the *Gla* eye mutation on the other chromosome 2.

5.5.4 *In situ* hybridization

The location of the insert was refined by *in situ* hybridization of *GmII* {2.9} for three strains of transformed flies. Stocks **X A**, **II A** and **III A** were used. The *in situ* hybridization confirmed the insertion chromosome, demonstrated that each of the investigated stocks had one insertion each, and allowed the insertion to be mapped.

Stock Name *yw*; P[*white*$^{+}$ *GmII*(E-E(Δ151))]	G0 vial	Insertion chromosome	Eye colour	Notes
X A	7	X	Bright red	Used for rescue
X B	8	X	Yellow/orange	
X C	14	X	Pale orange	
X D	40	X	Orange/red	
II A	7	II	Red	
II B	8	II	-	Balanced with Gla*
II C	40	II	Orange/red	
III A	16	III	Yellow/pale orange	Balanced with TM1*
III B	40	III	Orange/red	Variable colour

*These stocks carry homozygous lethal insertions and are therefore maintained with Gla and TM1 balancer chromosomes.

Fig 5.9. Transformed fly stock summary. Each stock carries at least one, independent P-element insertion of pP[*white*$^{+}$ *GmII*(E-E(Δ151))], the construct carrying the *GmII* gene.

Figure 5.10. Pigmented eyes of transformed flies. The eyes of XA, XB, IIC and IIIB males are clearly darker than the females

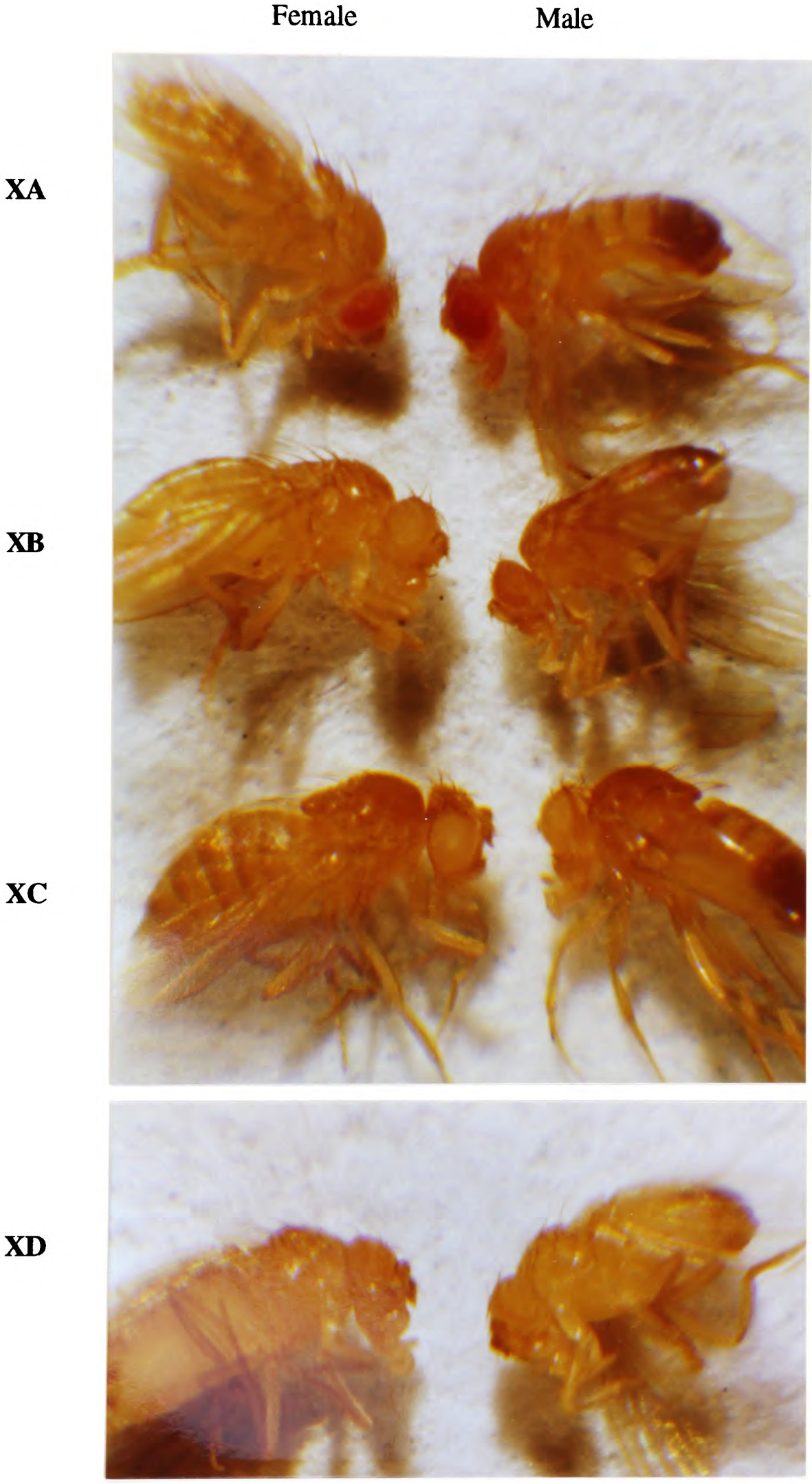
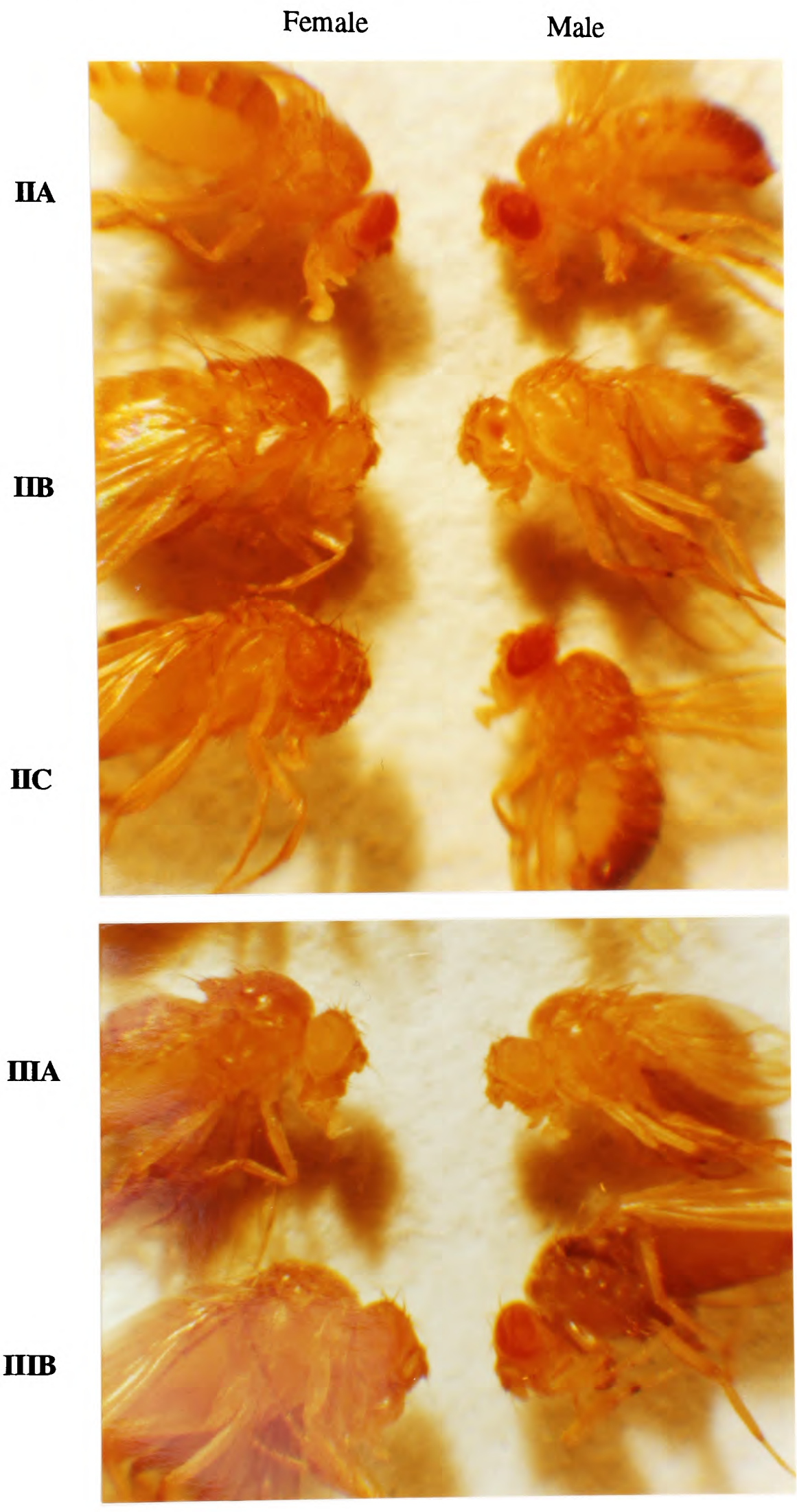


Figure 5.10 (cont). Pigmented eyes of transformed flies.



Stock **X A** had one insertion in 9E/F, the **II A** insertion mapped to 46C and the insertion in **III A** mapped to 64E (Fig. 5.11).

5.5.5 Position effects and dosage compensation

Variation in expression of the *white+* gene carried by the P-element vector is due to insert copy number, the position of the insert, dosage compensation or a combination of all three.

The level of expression of integrated DNA is sensitive to the local sequence environment [Spradling *et al.* 1983; Wakimoto *et al.* 1986] due to integration into passive or active gene regions. Different expression due to genomic position has also been found by [Hazelrigg *et al.* 1984]. In this case the position effect inactivation of expression is due to an insertion in the proximity of telomeric or centromeric heterochromatin [Levis *et al.* 1985]. Very few transposon insertions have been found to map in heterochromatin, a fact that could be due to i) the satellite-type DNA found there not carrying the required target sequence (see section 5.2.1), ii) the condensation in the region preventing insertion or iii) the difficulty of detecting transformants if the *white* gene has been inactivated.

In *Drosophila* dosage compensation is achieved by the hypertranscription of genes on the male's X chromosome. This allowed the identification of some X-linked integrations of pP[*white*$^{+}$ *GmII*(E-E(Δ151))] immediately since the males showed much darker eyes than the females, an effect also noted by [Spradling *et al.* 1983]. Autosomal insertions of X-linked genes still show dosage compensation in some cases [Hazelrigg *et al.* 1984]. Sequences required for dosage compensation have been identified between -216bp and the transcription start site of *white* [Pirrotta *et al.* 1985] and are therefore carried with the *white* gene in the pCaSpeR vector. This dosage compensation of the *white* gene in the pP[*white*$^{+}$ *GmII*(E-E(Δ151))] transposon was

Figure 5.11. *In situ* hybridization of *GmII* to three transformed fly stocks. *GmII* is localized at 85D on the third chromosome. Inserted pP[*white*$^+$ *GmII*(E-E(D151))] constructs are arrowed.

a) XA

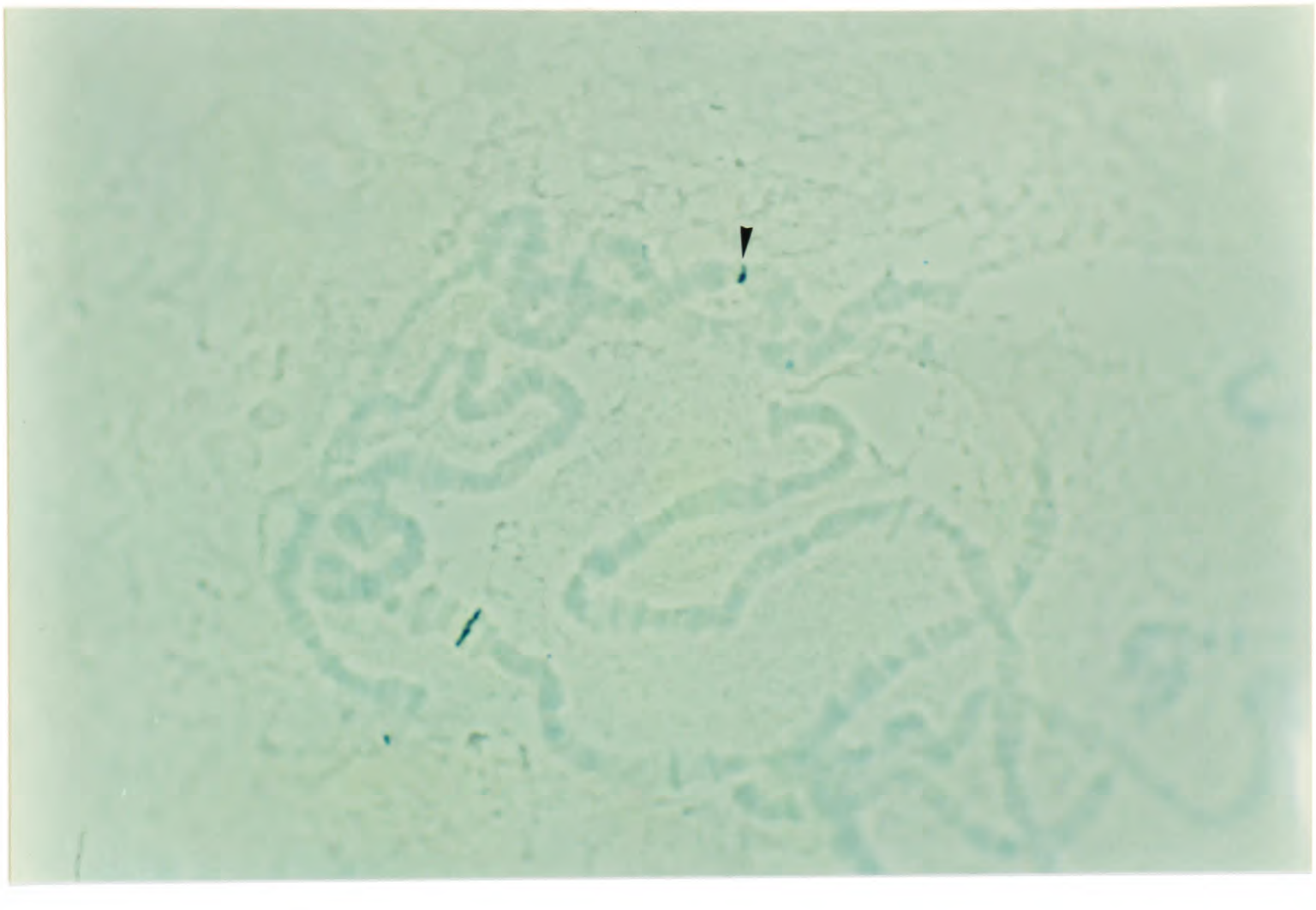


b) IIA



Figure 5.11 (cont). *In situ* hybridization.

c) IIIA



NB: The second (not arrowed) band is the endogenous *GmII* sequence localized at 85D.

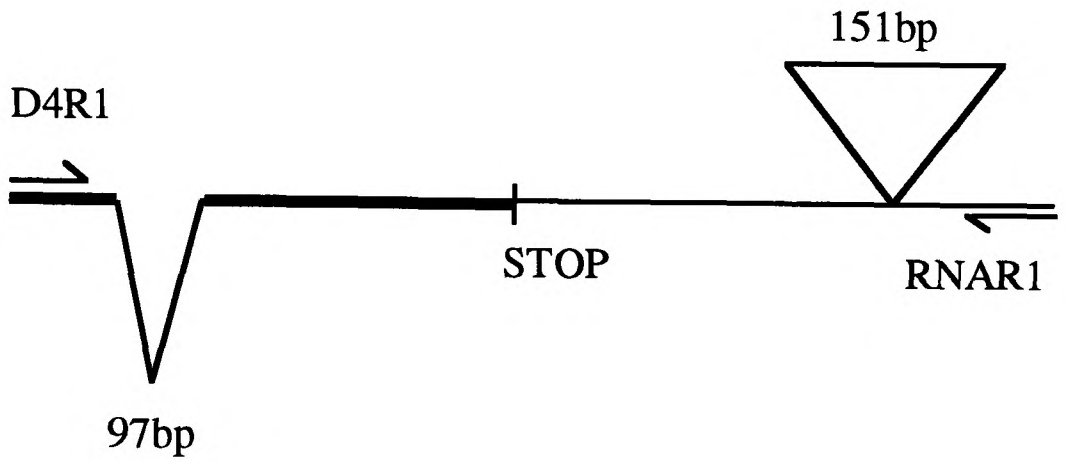
noticeable where males carrying autosomal insertions showed darker eyes than females in the same transformed stock.

The eye colour of the different transformed lines carrying *GmII* varied from bright red to pale orange (see Fig 5.10). The stocks **X A** and **II A** that have been investigated by *in situ* hybridization show considerable difference in eye colour. This is not due to copy number, since they are each homozygous for only one insertion of the same P-element construct. A combination of dosage compensation and position effects may have led to the levels of expression seen in **X A** which have bright red eyes. There may be some cumulative effect of having an X-linked gene integrated onto the X-chromosome, since these flies have much brighter red eyes than autosomal integrations. However, since these flies also showed brighter red eyes than other X-linked insertions of the same construct, a position effect must also be contributing to the levels of expression seen. It is assumed that the level of *GmII* expression is correlated with the level of *white* expression with respect to dosage compensation (of X chromosome integrated genes) and position effect.

5.6 Test for expression

Progeny from individual crosses of red-eyed flies to *yw* (see Fig 5.8) were collected for mRNA extraction {2.14.3} and first strand cDNA synthesis {2.16}. This was used for PCR {2.17}with two primers, one upstream from the last intron and the other downstream from the truncated region (Fig. 5.12a). The use of primers spanning the intron as well as spanning the truncated region controls for the possibility of genomic DNA contaminating the RNA and therefore cDNA sample. Figure 5.12b shows the PCR result. Positive controls showed the genomic DNA to produce a 822bp fragment which contains both the 151bp deleted in the modified *GmII* insert and the 97bp in intron IV. The injected construct DNA contains the intron, but not the BsaBI fragment, which gives a 673bp fragment. The native *GmII* cDNA does not contain the

a) 3' end of *GmII* - primers used for expression test by PCR.



Expected fragment sizes:

P1 genomic DNA (P1)	822bp
cDNA (without intron) (cDNA)	725bp
pP[*white*$^{+}$ *GmII*(E-E(Δ151))] (C)	671bp
Transcript of inserted gene	574bp

b) PCR gel showing fragments.

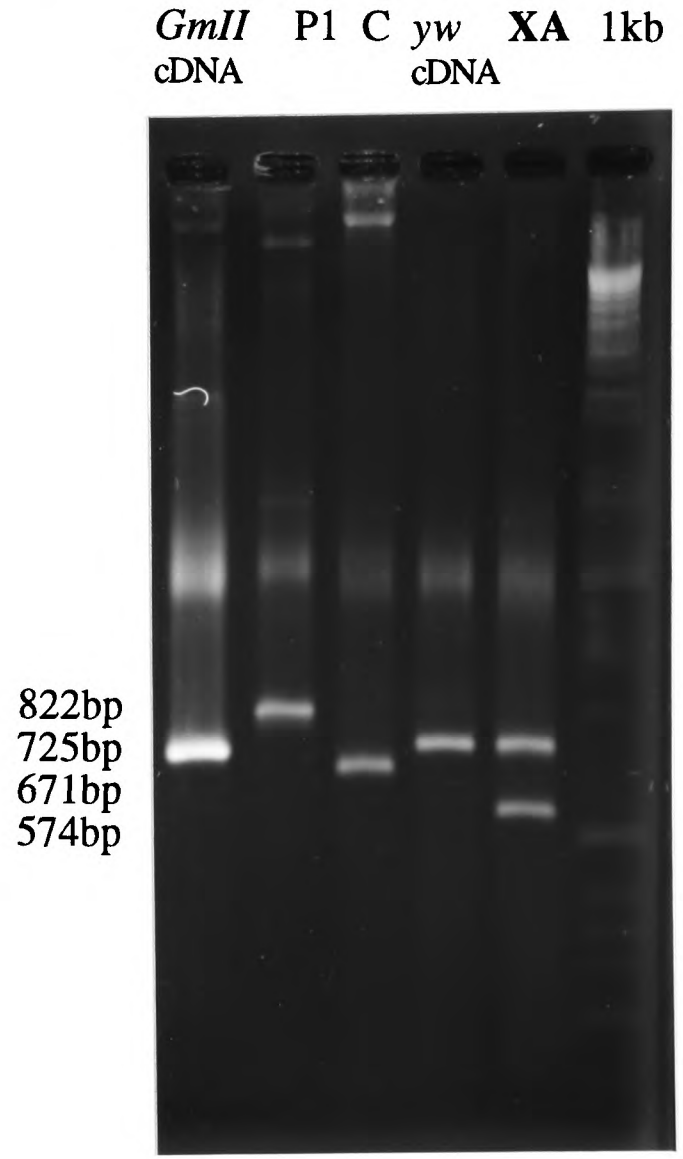


Figure 5.12. Expression test for stock **XA**.

intron and produces a 725bp fragment. The cDNA from the transformed gene does not contain the intron and has the fragment deleted from the 3' untranslated end and should give a 576bp fragment. This was found in the transformed flies. There is no reason to suppose that the transcript, once expressed normally is not also translated in the usual way.

5.7 Comparison between two transformed stocks and *yw* flies

The expression of the transformed gene in two transformed stocks (**X A** and **III B**) was compared with *yw* flies. The same PCR test of first strand cDNA with primers spanning the deleted region and the last intron was performed on cDNA from mRNA from different developmental stages (embryos 0-22hrs, 3rd instar larvae, pupae and adult flies) and from heads and bodies separately. The resulting PCR gels (Fig. 5.13) show that the inserted gene in these two transformed lines is expressed as the normal gene is at all stages and in all tissues investigated.

5.8 Rescue of *GmII* mutants

This was undertaken by B. Yudkin, who generated the mutants. He used the transformed strain **X A**. This was crossed to the Df/Dp line which had been used in identifying mutants in the 85D region and the identified lethal mutants were tested with it (Fig. 5.14). Currently, out of 93 lethal mutants analysed, none have been rescued using transformants carrying *GmII*. The search for a *GmII* mutant continues.

5.9 Discussion and conclusions

The previously cloned genomic sequence of *GmII* (Chapter 4) was required to identify mutants of *GmII*. The genomic sequence was used as it would have the native *Drosophila* promoter and gene structure that should produce wild type expression. A P-element construct was made containing *GmII* (pP[*white*$^{+}$ *GmII*(E-E(Δ151))]). This

Figure 5.13. PCR gels. Test of the expression of *GmII* from the integrated construct of two transformed stocks and from wild-type. (E-embryos, L-larvae, P-pupae, F-adult flies, H-heads, B-bodies, C-construct, G-genomic DNA, cDNA-*GmII* clone). Fragment sizes as for Fig. 5.12.

a) *yw.*

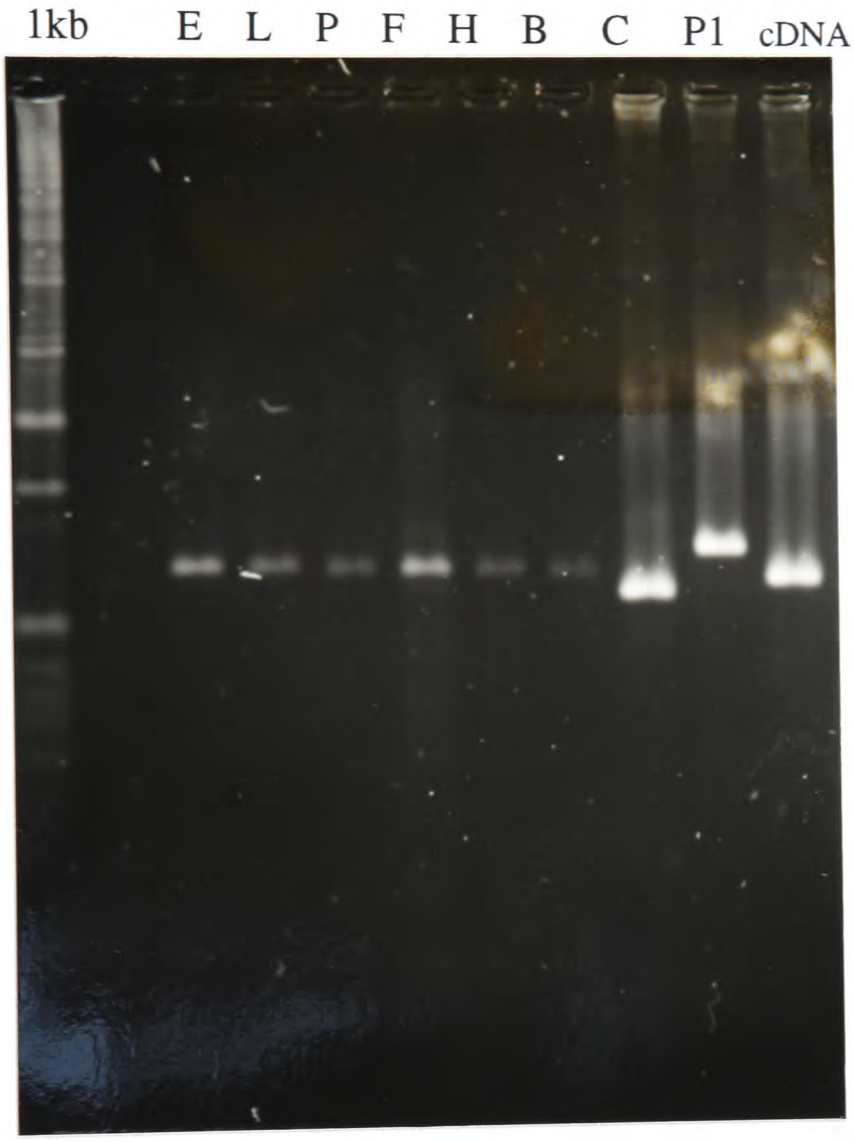
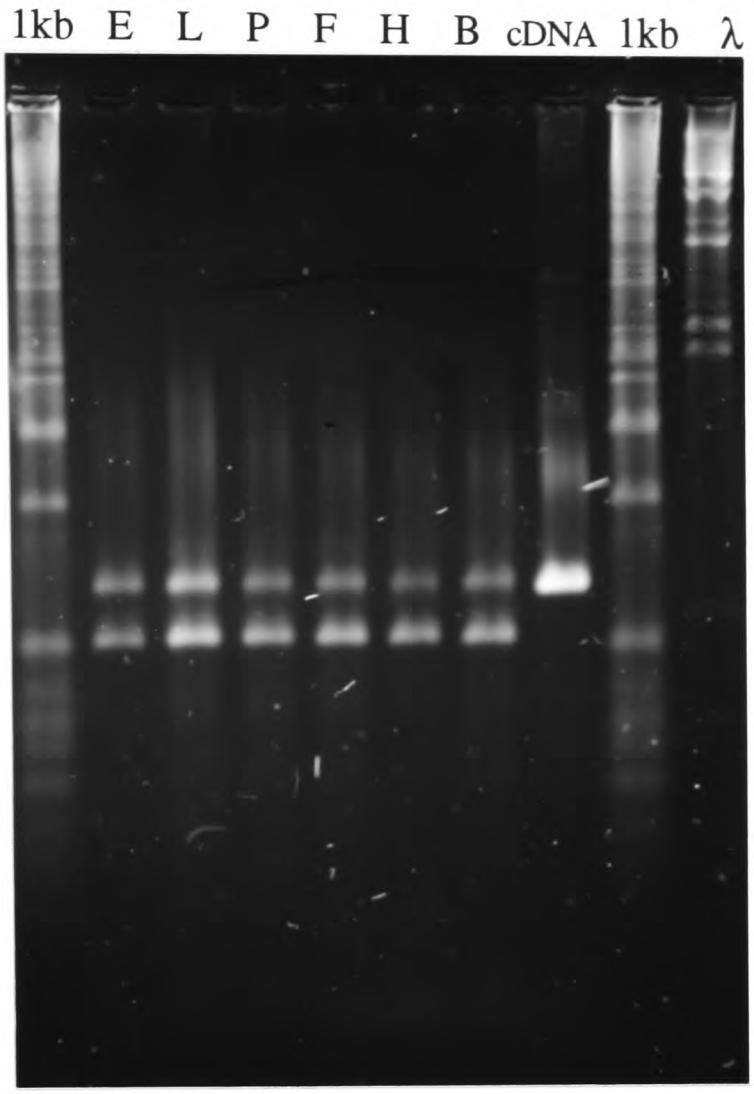
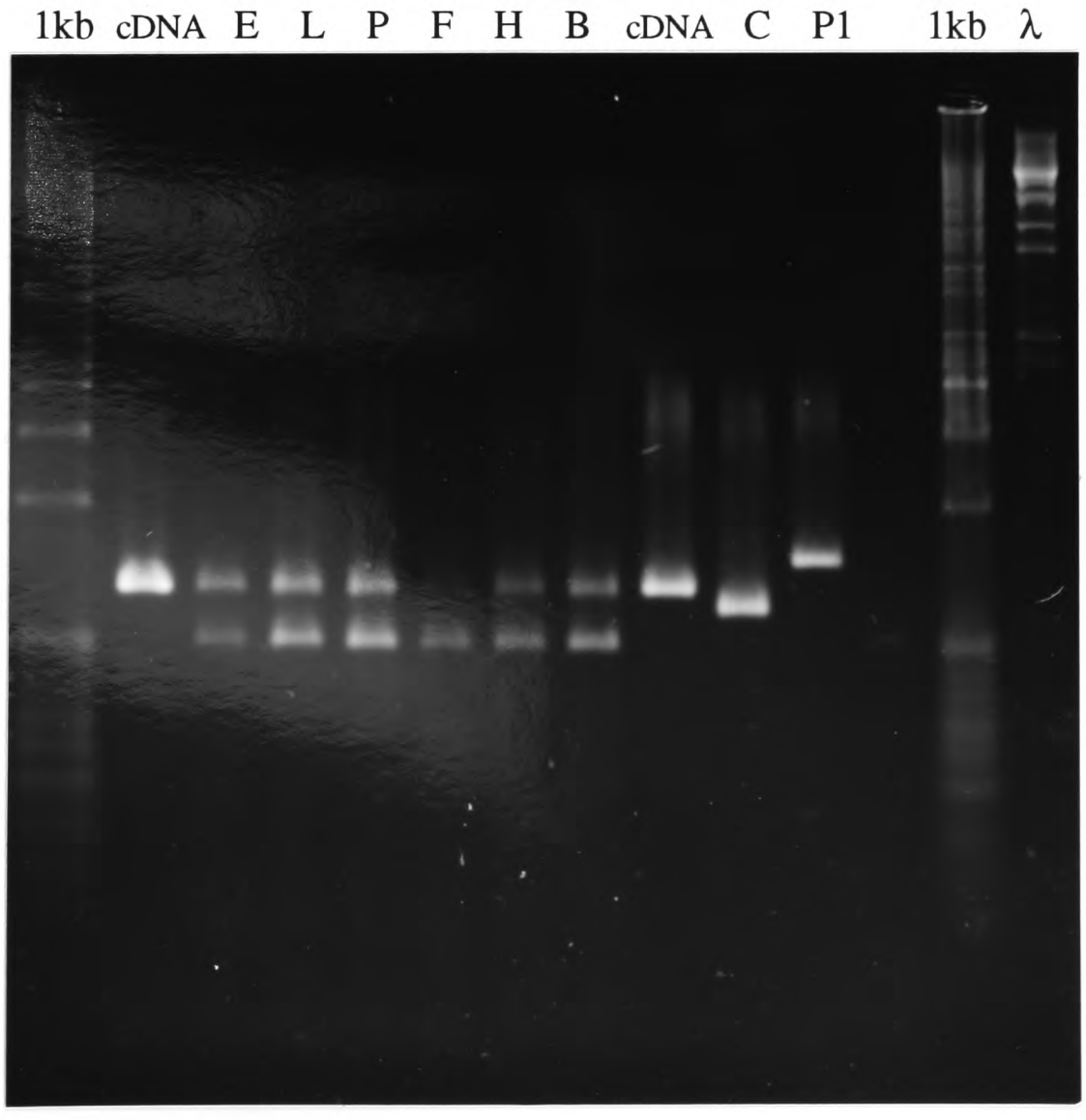


Figure 5.13 (cont). PCR gels testing expression of *GmII* in transformed stocks.

b) **XA.**



c) **IIIB.**



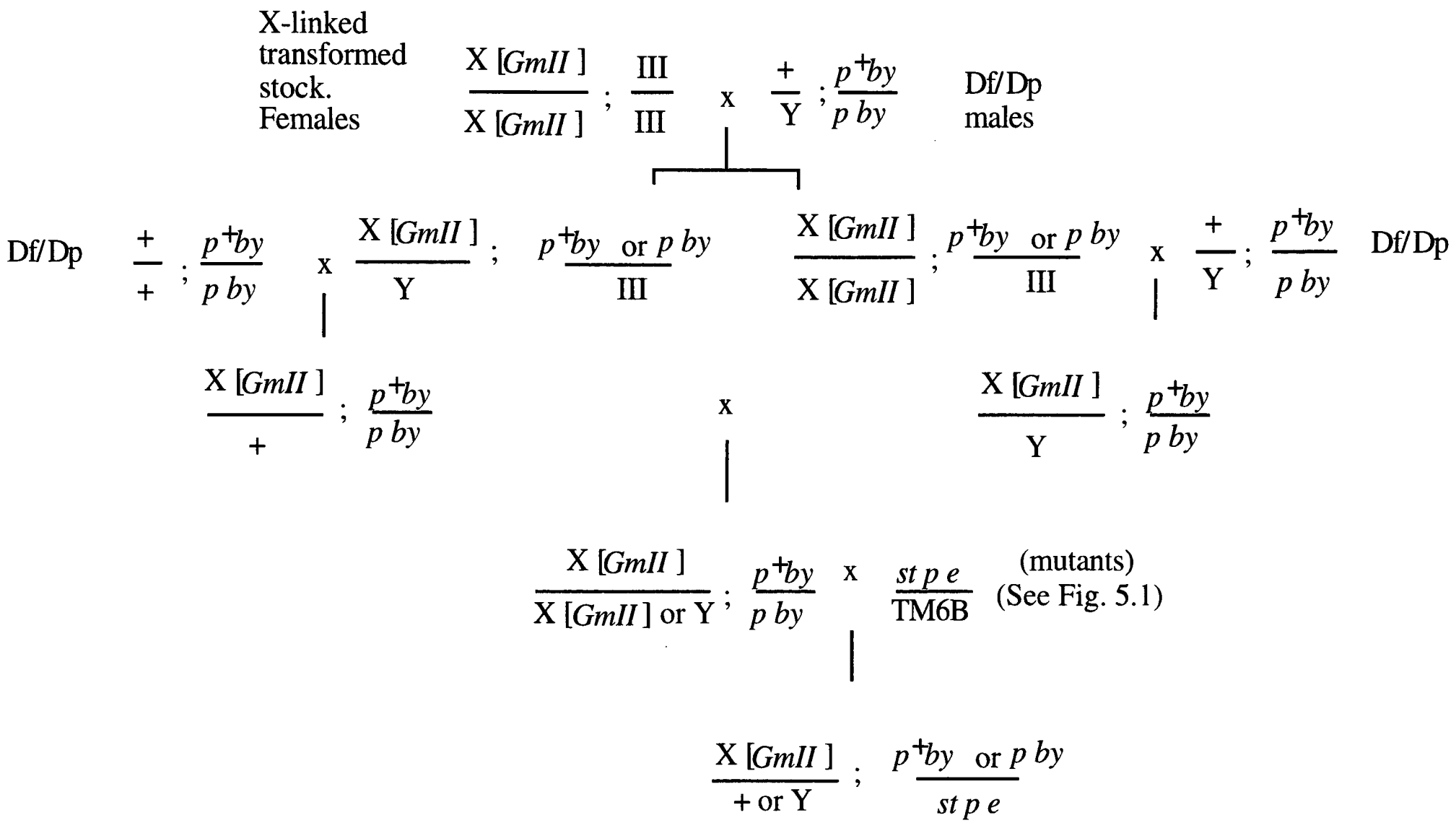


Figure 5.14. Mating scheme to identify mutants of *GmII* using a homozygous transformed line carrying an inserted *GmII* construct on the X chromosome. (X [*GmII*] is used to denote *yw* P[*white*$^{+}$ *GmII*(E-E(Δ151))] **XA**).

version of *GmII* had a modified 3' end that allowed its transcript to be detected and monitored in different developmental stages and tissues. Fly transformation with this construct demonstrated that the gene was expressed from 700bp of 5' sequence. Investigation of this expression in embryos, larvae, pupae, whole adult flies and heads and bodies separately showed no differences of expression between the native and modified genes.

The transformation demonstrated that the promoter elements for *GmII* were all contained within 700bp 5' from the start of transcription. Could the gene be expressed from less? This question was addressed by linking shorter lengths of this region to a reporter gene and investigating the expression of these constructs in *Drosophila.* The full length (700bp) 5' sequence was also linked to *lacZ* to investigate further the expression of *GmII*. These experiments are discussed in Chapter 6.

The search for a *GmII* mutant with this construct continues. The lack of success so far could be due to a number of reasons. The formal possibilities are that i) a mutant in *GmII* has not yet been made; ii) mutants of *GmII* are not lethal; iii) a lethal mutant of *GmII* has not been rescued by the transformed construct due to its chromosome location or iv) such a lethal mutant has not been rescued by the transformed construct due to incorrect expression of the *GmII* protein. Although no differences in expression could be detected, more subtle differences in expression could be responsible. Alternatively the modification of the transcript could have affected its translation or stability. Now that the expression has been detected from this construct it would be possible to transform flies with a construct with an unmodified 3' end to ensure no disruption of the transcript. It would also be possible to use the longer 5' end provided from the BamHI fragment from the original P1 **1** clone (section 5.3.2) in another transformation experiment. It is hoped that the identification and investigation of a mutant of *GmII* will yield informative results as to the significance of *GmII*.

CHAPTER SIX
GOLGI MANNOSIDASE II : PROMOTER ANALYSIS

6.1 Introduction

The transformation experiment described in Chapter 5 demonstrated that 700bp 5' sequence is sufficient for the transcription of *GmII*. Here I consider which regions or bases in this sequence are vital for transcription, by analysing the *GmII* 5' sequence for well characterised promoter elements (section 6.2) and the expression of a reporter gene (*lacZ*) regulated by different lengths of *GmII* sequence.

However it is possible that the expression of *GmII* could be fine tuned by enhancer elements, sequences that are usually located away from the promoter region, either upstream; within the gene (or an intron); or downstream. Such enhancer elements may be involved in tissue specific expression [Pirrotta *et al.* 1985][Hazelrigg 1987] and are less easy to analyse.

6.2 Analysis of *GmII* promoter sequence

6.2.1 TATA box

Many eukaryotic promoters have a TATA box located at -25/-30bp from the transcription start site [Bucher 1990]. This *cis*-acting sequence is one of the best characterised basal eukaryotic promoter elements. It is recognised by the TATA binding protein (TBP) which induces DNA bending and participates in transcription by all three RNA polymerases [Hernandez 1993]. A TBP homologue called TRF (TATA recognition factor) has been identified in *Drosophila* [Crowley *et al.* 1993], but has been shown to have a restricted expression pattern only in the embryo. Mutations in the TATA box do not prevent transcription initiation, but alter the transcription start point, suggesting a role in positioning the transcription machinery. An analysis of

Drosophila promoter elements [Arkhipova 1995] yielded the consensus sequence $^{A}/G/_{C}$ **T A T A A A** $^{G}/_{A}$ $^{G}/_{A}$.

The 5' region of *GmII* shows no TATA consensus sequence at -25/-30bp, and therefore falls into the TATA-less subset of *Drosophila* promoter sequences, a set containing about one half of *Drosophila* promoter sequences. The mechanism for transcription of TATA-less promoters is not understood, although TBP has been shown to participate in some cases *in vitro* [Zhou *et al.* 1992]. This may be due to association with TBP associated factors (TAFs) which form a multiprotein complex with TBP and may attach to other upstream sequences such as the GC box (see section 6.2.3) [Pugh *et al.* 1991].

TATA-less promoters have been divided into two types [Smale *et al.* 1989]; firstly GC-rich constitutively expressed housekeeping genes (for review see Sehgal *et al.* 1988) and secondly regulated genes. There appears to be no specific TATA box substitutes in -25/-30bp region in *Drosophila* [Arkhipova 1995]. Transcription from TATA-less promoters may be more flexible; transcription may occur at any one of a cluster of start points, or in the absence of upstream sequences the cap-site sequence or initiator (Inr) may supply the positioning element.

6.2.2 Initiator sequence

The cap-site or initiator sequence (Inr) is present in the immediate site of RNA transcription start [Smale *et al.* 1989; Gill 1994]. In vertebrates the sequence has the general form $^{A}/_{T}$ $^{A}/_{T}$ C A $(^{A}/_{T})_{5}$ located between -3 and +5, while in arthropods the sequence T C A G T was identified [Cherbas *et al.* 1993]. The sequence A T C A $^{G}/_{T}$ T $^{T}/_{C}$ was identified in insects [Kozak 1984] and extended to $^{T}/A/_{G}$ T C A $^{G}/_{T}$ T $^{T}/_{C}$ G in *Drosophila* [Arkhipova 1995] and found in approximately one third of *Drosophila* promoters. In most cases the 5' end of the mRNAs maps to the first nucleotide in this sequence, although in some cases a gap must be introduced after the second nucleotide

for maximum homology [Hultmark *et al.* 1986]. The presence of this sequence has no obvious correlation with the presence or absence of the TATA box. The *Drosophila* Inr consensus sequence compared to the start of *GmII* cDNA clone has many mismatches.

T/A/G	T	C	A	G/T	T	T/C	G	Initiator sequence
A	T	G	C	T	A	A	A	*GmII* sequence

GmII mRNA Start

The start site suggested by the beginning of the cDNA could be an artefact from cloning the cDNA. However searching the *GmII* 5' region for better matches finds no complete matches, but one base pair difference mismatch at four sites. There are also sites downstream. These suggest that either individual bases in the sequence are more significant than others or that other sequences in the region are responsible for locating the transcriptional machinery. Alternatively several mRNAs with different lengths of untranslated leader sequence could be transcribed. There does not seem to be enough evidence to relocate the start of transcription, but the possibility of different start sites remains.

TATA-less promoters also tend to be enriched by short runs of T in the initiator region [Arkhipova 1995]. *GmII* contains two runs of 3 T's in the 7bp 5' to the mRNA start site (see Fig 4.10).

6.2.3 Upstream elements

Important upstream elements have been identified by testing the effect on transcription of individual base substitutions. One such sequence is located at -80/-100 and called the CAAT box, with a consensus sequence G G C **C A A T** C T. This sequence can function at variable distances from the start point, and more surprisingly, in either orientation. Its function seems to be in determining efficiency of the promoter. Although prominent in mammalian promoters [Bucher 1990], this sequence is not over-represented in *Drosophila* promoters [Arkhipova 1995]. A search of the 5' *GmII*

sequence did not find this sequence in either orientation; although several CAAT and TAAC tetramers were found in the region -50 - -110, none had the required surrounding sequences.

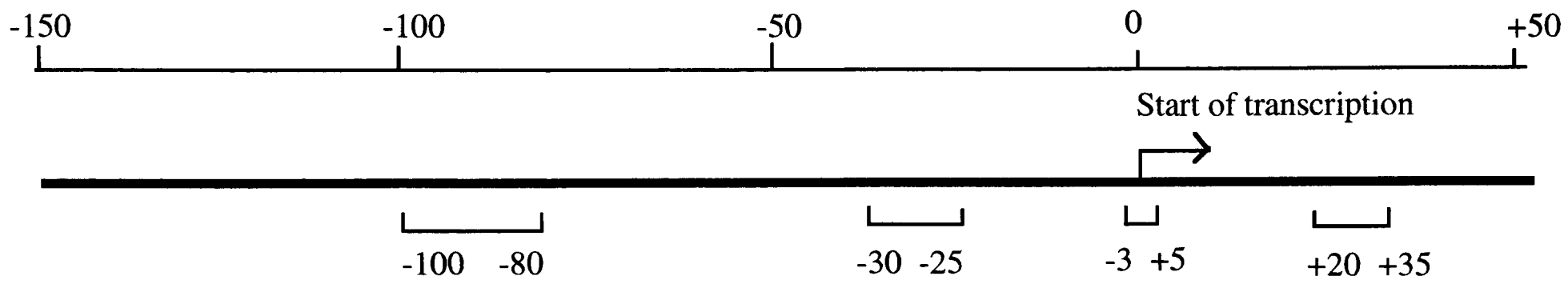
Other promoter elements have been characterised. One relatively common promoter element is the GC-box at around -90. This has the consensus sequence of G G G C G G, which can occur in either orientation [Bucher 1990]. This is the recognition sequence for the transcription factor Sp1. This sequence again is not over-represented in *Drosophila* promoters [Arkhipova 1995]. Another promoter element sequence is the octamer A T T T G C A T [Parslow *et al.* 1987]. Neither sequence is present in the *GmII* 5' sequence.

6.2.4 Downstream elements

Downstream elements located +20/+35 bp from the start of transcription, have also been identified in *Drosophila* genes by searching for over-represented sequences in a data set of *Drosophila* promoter sequences by Arkhipova [1995], who estimates more than half of *Drosophila* genes may contain them. They have been shown to be essential for transcription and may be associated with the binding of nuclear proteins. These downstream elements are strongly over represented in TATA-less promoters. They consist of several shorter sequences whose positions are less rigidly fixed, compared with TATA and the initiator site. *GmII* does contain one of these sequences (ACGC) at +16-+20bp downstream, however the shortness and multiplicity of these sequences make it hard to tell if this is significant or merely chance.

6.3 Transforming flies with the *GmII* promoter

The analysis of the *GmII* 5' region revealed few of the characterized promoter elements (summarized in Fig. 6.1). Which regions therefore are required for transcription? The initial investigation was to shorten the amount of 5' sequence used.



Position	-100 -80	-30 -25	-3 +5	+20 +35
General consensus sequence	GGGCGG GGCCAATCT	TATAAAA	ATCA(G/T)T(C/T)	Various

Name	GC- box	CAAT-box	TATA-box	Inr	Downstream elements
GmII?	No	No	No	?	?

Figure 6.1. Characterized promoter elements.

The lack of convenient restriction sites in the *GmII* 5' sequence required shortened 5' lengths to be made by PCR. These could be linked to the *lacZ* gene of *E. coli* coding for the enzyme β-galactosidase, and used to transform flies, allowing visualisation of spatial and temporal expression of *GmII*.

6.3.1 pCaSpeR - AUG - β-gal

The P-element vector chosen to use for cloning the 5' sequences and transforming flies was pCaSpeR - AUG - β-gal (Fig. 6.2). This was constructed using *Adh, lacZ* and SV40 sequences from the pC4-AUG-βgal vector and inserting them into the EcoRI and PstI sites in the pCaSpeR polylinker (see Fig. 5.2) [Thummel *et al.* 1988]. It therefore contains not only the P-element arms required for transformation and the mini-*white+* gene as a marker, but also the *lacZ* gene to act as a reporter whose expression can be detected by staining. The AUG sequence from *Adh* allows the gene to be expressed from a promoter not in frame. This vector (now containing a unique SflI site in the multiple cloning site 3' to the *lacZ* gene, unimportant for this investigation) was supplied by Jean-Antoine Lepesant, Jaques-Monod Institute, Paris. The 5' cloning site into which the promoter sequences can be inserted contains three unique restriction sites - EcoRI, BamHI and KpnI.

6.3.2 *GmII* promoter fragment by PCR

The restriction enzymes chosen for insertion were EcoRI and BamHI. These have unique sites in the pCaSpeR-AUG-βGal SflI vector and cut nowhere in the 5' region of *GmII*. PCR primers were designed which incorporated the cut site sequences for these enzymes (Fig 6.3). The primers require about 20bp of homologous sequence, the restriction sites and 2bp additional sequence required for efficient cutting of the site. The amount of extra sequence required depends on the enzyme used (New England Biolabs, 1993). The 3' BamHI site was chosen to be at +80bp in the

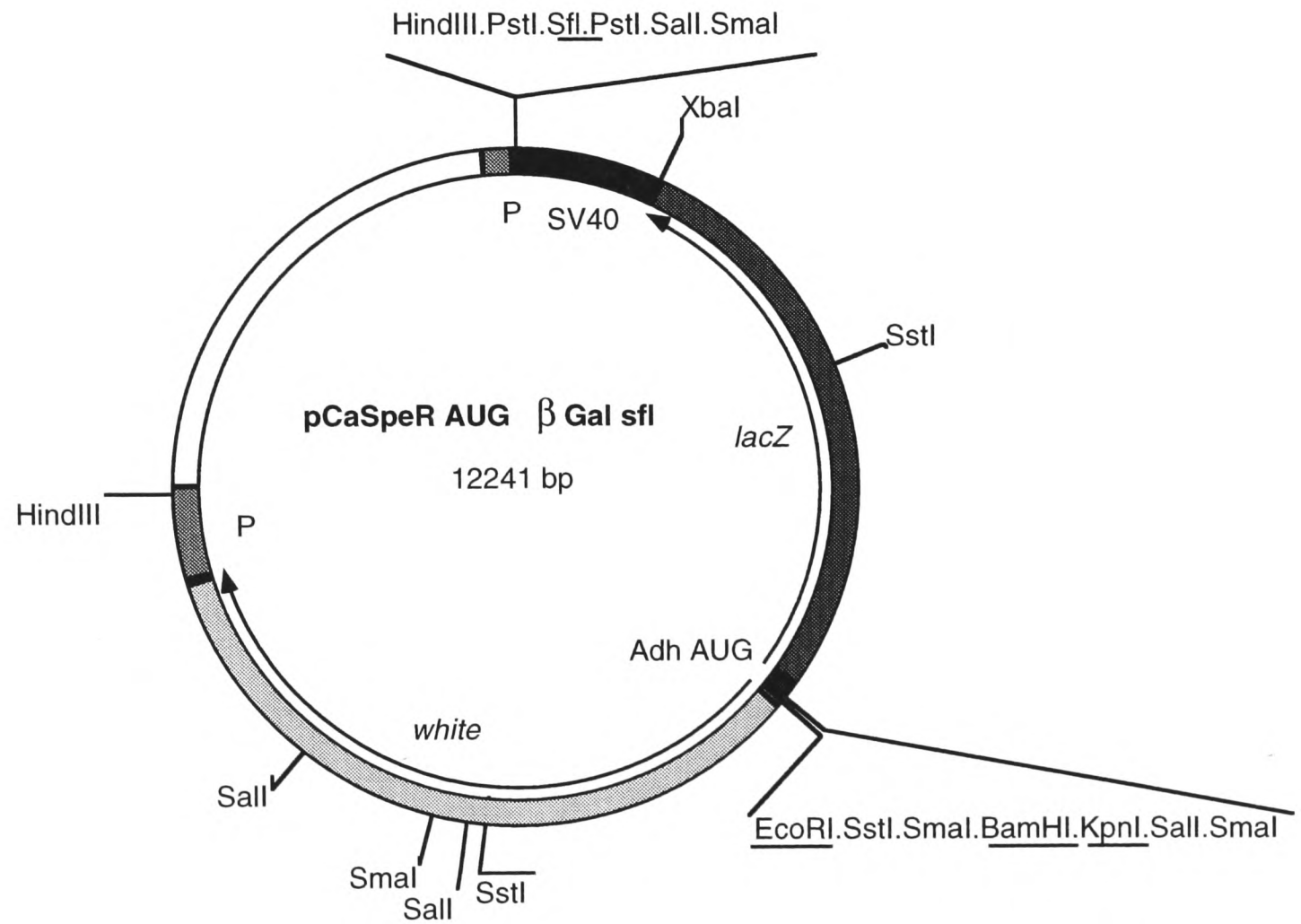


Figure 6.2. pCaSpeR AUG βGal sfl. A P-element vector carrying the *lacZ* reporter gene, *white* + and four unique cloning sites (underlined).

Figure 6.3. PCR primers for the 5' region of *GmII*. The restriction site in each primer is underlined.

```
   GCGAATTCCCATC  TTTTGG>                                          EcoI
-700  AATTCCCATC TTTTGGCCTC CGTTGGTGAC TCCAGGGAAT CGAACACGGC

-650  GGCTCCAATT AAAAGATAGG TGAAAGTGCA TACCACCAGG GAAAGGGTAC

-600  GGACATTCTG TCGCTTCATC ATATAGTCTC AAATATAATC ACTACGATCT

-550  ATAATCAATT AATATATTTT GCACCACGCG CCTTTTTCGT TCAGTTAAAG

   GCGAATTCGCGTC  TGACAAAATC  GATAG >                          EcoII
-500  ATGCTGCGTC TGACAAAATC GATAGTTGTC AGCAAAACAA ACACACAGCA

-450  TCCGCCAAAA ATGGTTGATT TCGGTTACAG TGGCCGAATA ACATATGATA

-400  CAATGAATTT TAAAAATTTC ATCATTCTTT AAATATGTAA TAAGTTTTCT

-350  TTTATCTATT GATTTGACCG TTAATCTCAA TAAAGGATTG TGATTTTACC

                G CGAATTCGA-  GCAACAGA-T  TTGATGGCG  >     EcoIII
-300  GTTAGTTTAT CACAACAGAA GCAACAGAAT TTGATGGCGT TGCCACTTGG

-250  CGGTATCTGT TAACTCTGTA CAGCTAAAAA GCTGTAATGT CACGATAACT

-200  AAGGATAATA ATAATTTTTT TAAAGGATCA ATATACACAA ACAACATAAG

-150  AGTTTTATTA CATTTTCTTT AATTGCCTTT AAAGTTTTTG TCTTTCTAAG

-100  CCATTTTAAC AATTTAACAT TTTGCAGTTA GCAGCGCCGA TAACTTGACA

-50   GCCTATCGAT TTTCGGGAAA AGCAAACATA TCGCCAGATC TGATTTGTTT

      cDNA  ->
0     ATGCTAAAAA AAAACCACGC AAGTTCCAAC CGATTGGATT GTTGCAGTTG

                  <  CTCTGACGTC  GACGTTTCCG  GATCCGC              BamR
50    GCTGCTGTGT CTCTGACGTC GACGTTTCCA TGCTTAATTA ACCGGAAAAT

100   GTTGCGAATA CG.....
```

untranslated leader sequence. This would retain any important downstream elements (see section 6.2.4). Three 5' EcoRI primers were designed: i) EcoI, which would produce a full length fragment from the pP1 E-E subclone retaining the 700bp sequence 5' to the start of transcription used in the first transformation experiment; ii) EcoII positioned 200bp closer to the start of transcription, producing a fragment with 500bp 5' sequence and iii) EcoIII, which would produce a fragment with 288bp 5' sequence.

Initially the full length fragment was produced using PCR {2.17} with Taq polymerase (Promega). However when cloned and the sequence checked, mistakes were discovered in the amplified fragment (results not shown). Taq polymerase has no proof-reading function. Two mistakes in approximately 100bp of sequence were found, which is a much higher error rate than claimed by the manufacturer. The other two fragments and the first repeated were amplified using *Pfu* (Strategene), a thermostable enzyme that has some 3' to 5' exonuclease proof-reading activity. The conditions for *Pfu* were determined empirically and required a much reduced annealing temperature (47^{o}C) and the addition of glycerol (1% final) in the reaction mixture.

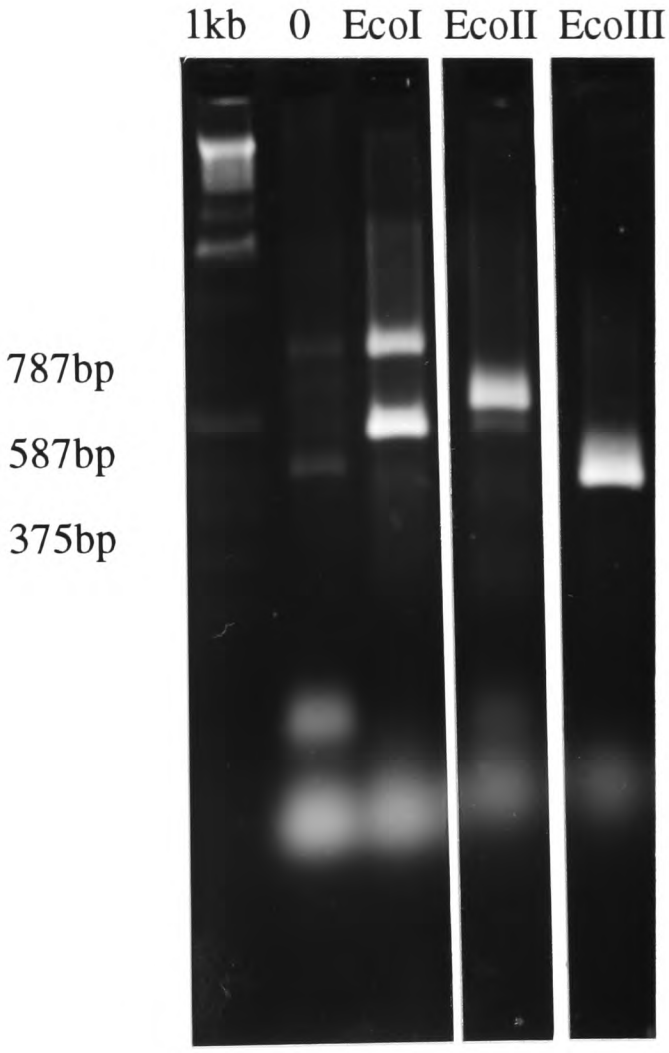
The fragments produced by PCR were checked by running an aliquot on an electrophoresis gel (Fig. 6.4) and then precipitating the remainder of the reaction volume. This was digested with EcoRI and BamHI and cloned into pCaSpeR - AUG - β-gal {2.11} to give the three constructs shown in Figure 6.5. The sequences of the PCR fragments were checked and found to be correct (results not shown). The three plasmids were called pP[*white*$^{+}$ P1(700) *lacZ*], pP[*white*$^{+}$ P1(500) *lacZ*] and pP[*white*$^{+}$ P1(288) *lacZ*].

6.4 Transformation

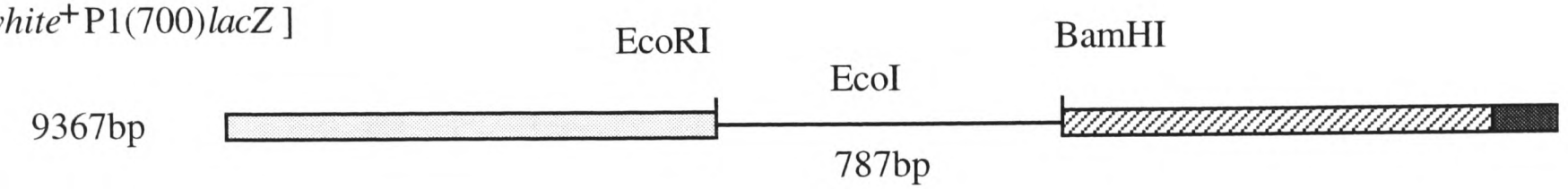
6.4.1 Transformed stocks carrying the *lacZ* constructs

Each of the three constructs was microinjected into flies {2.18}. The microinjection survival data are given in Figure 6.6. Survivors were crossed to *yw* and

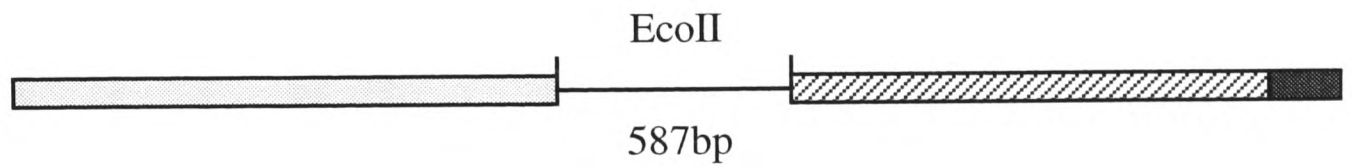
Figure 6.4. PCR fragments using the primers shown in Fig 6.3. (Conditions: using *Pfu*, 47°C annealing temperature, 1% glycerol (final)). The shorter band produced with EcoI was a spurious fragment probably amplified due to the low annealing temperature required for *Pfu*. and the lower specificity of annealing this produces.



i) pP[*white*$^{+}$P1(700)*lacZ*]



ii) pP[*white* $^{+}$P1(500)*lacZ*]



iii) pP[*white* $^{+}$P1(288)*lacZ*]

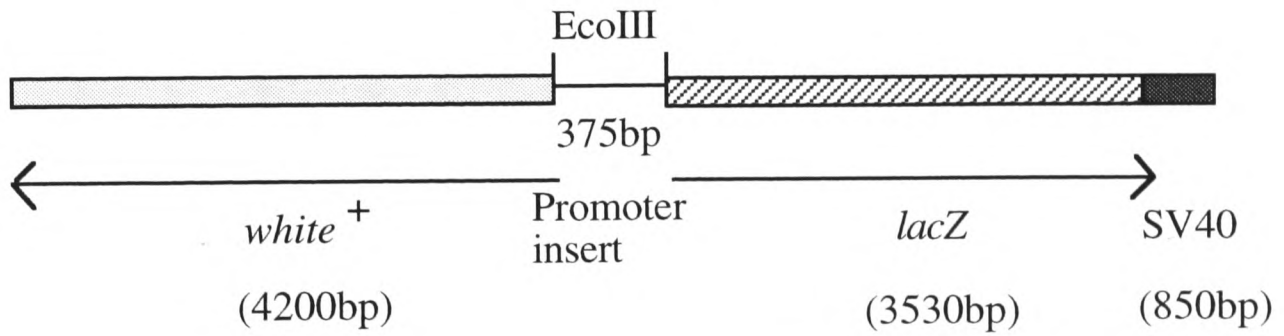


Figure 6.5. P-element constructs with different lengths of *GmII* 5' region. The P-element arms and pUC vector sequences are omitted.

Figure 6.6. Table of transformation efficiency for the three *lacZ* constructs.

a) Survival data.

P-element construct	pP[*white*+ P1 (700) *lacZ*]	pP[*white*+ P1 (500) *lacZ*]	pP[*white*+ P1 (288) *lacZ*]
Length of *GmII* 5' DNA	700bp	500bp	288bp
Embryos injected	941	891	1660
Larvae	39	53	89
Flies (crossed to *yw*)	18	24	43
Fertile crosses	17	23	35
Transformed vials	3	2	1

b) Survival percentages.

Construct	pP[*white*+ P1 (700) *lacZ*]	pP[*white*+ P1 (500) *lacZ*]	pP[*white*+ P1 (288) *lacZ*]
Injected embryo survival to larvae	4.1%	6.0%	5.4%
Larval survival to flies	46.2%	45.2%	48.3%
Fertile crosses	94.4%	95.8%	81.4%
Fertile vials transformed	17.6%	8.7%	2.86%

screened for red-eyed progeny (see section 5.5). Due to time constraints the transformed flies were not crossed to CyO/Gla and TM1/TM3 to determine which chromosome carried the insertion. The stocks were made homozygous by crossing red eyed males to females (see Fig. 5.8a). The mini *white* gene allows homozygotes to be distinguished from hemizygotes [Klemenz *et al.* 1987]. The independent stocks maintained are shown in Figure 6.7. Two independent lines carrying the 700bp construct (*yw*;P[*white*$^+$ P1(700) *lacZ*] A and *yw*; P[*white*$^+$ P1(700) *lacZ*] B) and the 500bp construct (*yw* P[*white*$^+$ P1(500) *lacZ*] A and *yw* P[*white*$^+$ P1(500) *lacZ*] B)were maintained. The one insertion of the shortest construct reduced the viability of the stock and so was maintained with a balancer construct for chromosomes 2 and 3 (P[*white*$^+$ P1(288) *lacZ*]/*Tu* (2;3) *pr cn*/CyO/TM6, *mwh ry*^{506} *e*).

6.4.2 *In situ* hybridization

Squashes of chromosomes from three stocks carrying each of the three constructs were probed with the *lacZ* gene {2.9} (Fig 6.8). The probe was prepared by digesting pCaSpeR-AUG-βGal with BamHI and PstI which cut in the two polylinkers of this construct (see Fig 6.2). The *lacZ* fragment was purified from an electrophoresis gel {2.6.2}. The *in situ* hybridization showed each stock (700 A, 500 A and 288) to carry one copy each of the constructs, at 72A and 9E/F respectively for the longer constructs. The balancer construct over which the transformed chromosome was maintained in the 288 stock, distorted the chromosome arms in the squash to such an extent that the location of the shortest contruct insertion could not be identified.

6.5 β-galactosidase expression from 700bp promoter

6.5.1 Embryos

The expression of the *lacZ* gene in embryos of a stock (700 A) carrying the full length promoter construct was investigated by staining for β-galactosidase activity

Stock name	Stock code	G0 vial	Insertion chromosome.	Eye colour	Notes
yw; P[*white*$^{+}$ P1(700) *lacZ*]A	**700 A**	5	Autosome	Orange (females) / red (males)	Used for β-gal staining and squashes.
yw; P[*white*$^{+}$ P1(700) *lacZ*]B	**700 B**	14	Autosome	Orange	
yw P[*white*$^{+}$ P1(500) *lacZ*]A	**500 A**	5	X chromosome	yellow/orange variable	Used for β-gal staining and squashes.
yw P[*white*$^{+}$ P1(500) *lacZ*]B	**500 B**	9	X chromosome	Orange	
P[*white*$^{+}$ P1(288) *lacZ*]/*Tu* (2;3) *pr cn* / CyO / TM6, *mwh ry*^{506} *e*	**288**	26	Autosome	-	Homozygous lethal. Used for β-gal staining and squashes.

Figure 6.7. Stocks carrying the three different *GmII* - *lacZ* constructs.

Figure 6.8. *In situ* hybridization of *lacZ* to the transformed stocks.

a) *yw*;P[*white*$^+$ P1(700) *lacZ*] A



b) *yw* P[*white*$^+$ P1(500) *lacZ*] A

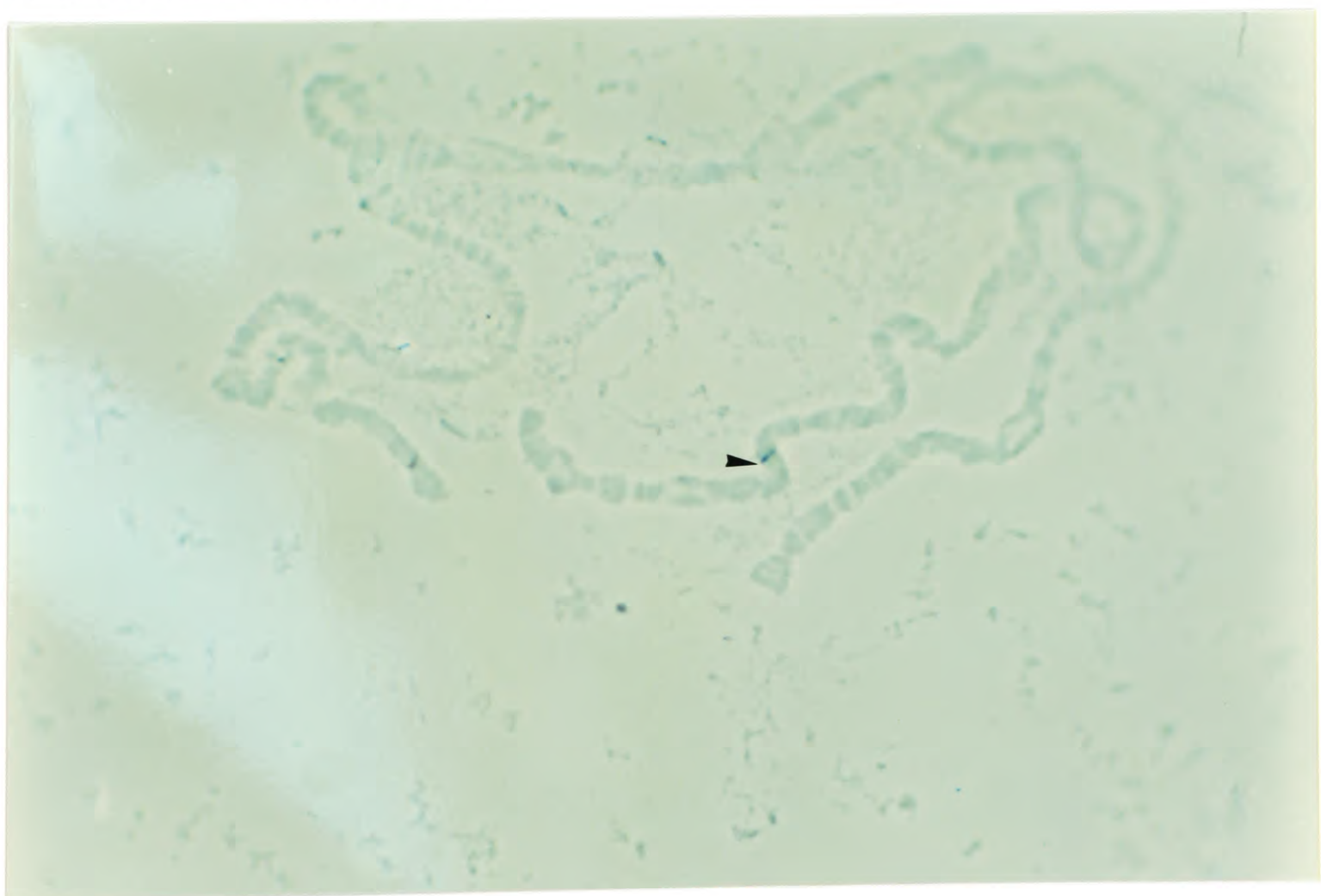
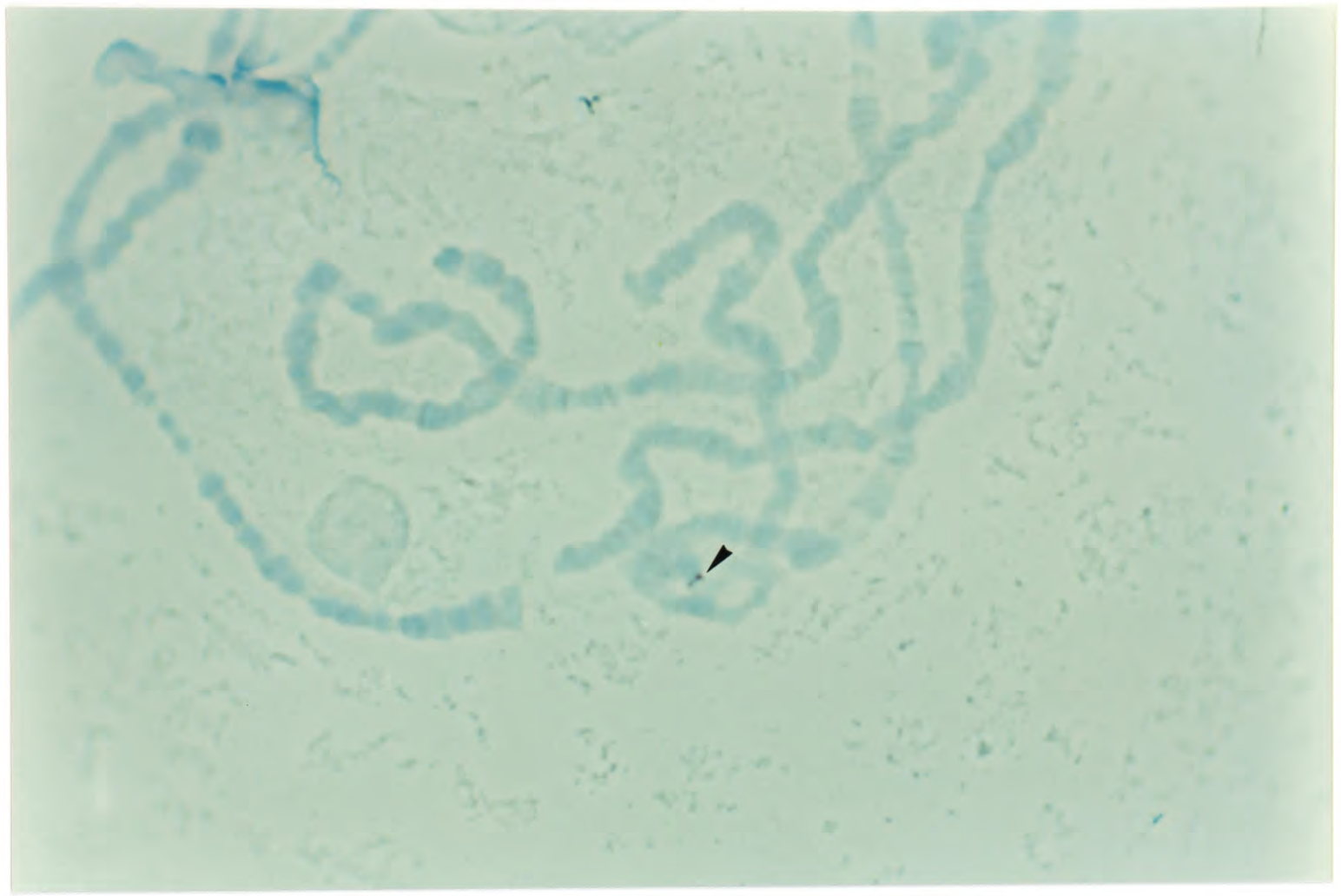


Figure 6.8 (cont). *In situ* hybridization to the transformed stocks.

c) P[*white*$^+$ P1(288) *lacZ*]/*Tu* (2;3) *pr cn*/CyO/TM6, *mwh* ry^{506} *e*.



{2.20.1}. Collections of aged embryos were made to allow differences of expression throughout larval development to be investigated. Figure 6.9 shows stained embryos from each stage investigated. Not all the embryos in each collection stained, probably due to differences in permeability depending on the efficiency of the fixation step. Figure 6.10 shows a table of numbers of embryos stained. It is clear that the transformed embryos show considerably more β-galactosidase activity than the low endogenous background level seen in the untransformed *yw* embryos.

The *lacZ* gene is linked to the *GmII* promoter and the staining seen can therefore be used to describe the expression of *GmII*. Expression in embryos was expected from the PCR experiments using first strand cDNA from *yw* embryos (see section 4.8.3) and from embryos carrying the transformed gene (see section 5.7). *GmII* seems to be expressed ubiquitously throughout embryo development and in all areas of the embryo although in the later stages of embryonic development some tissues stained darker blue.

6.5.2 Imaginal discs

Imaginal discs were dissected from third instar larvae of the transformed line carrying the 700bp promoter (700 A), *yw* and neu-*lacZ*/TM3, a stock that does show β-galactosidase expression in the imaginal discs (kindly donated by G. Jiminez). The discs were stained for β-galactosidase activity {2.20.2}. Figure 6.11 shows wing imaginal discs from neu-*lacZ*/TM3 and *yw*, and wing, eye-antennal and leg discs from 700A. The neu-*lacZ*/TM3 stock stained blue for certain cells in the imaginal discs as expected [Cubas *et al.* 1991]. The imaginal discs from *yw* larvae did not stain, nor did the discs from the transformed stock. This suggests that *GmII* is not expressed in the imaginal discs.

Imaginal discs are used to construct the adult fly and the cells in the imaginal discs divide during larval development. Carbohydrate-recognizing anti-HRP normally stains imaginal tissue [Katz *et al.* 1988] and peanut agglutinin which recognizes

Figure 6.9. **700 A** embryos (full length promoter) stained for β-galactosidase activity.

a) *yw* (control) at different developmental stages.



b) 0-2hrs

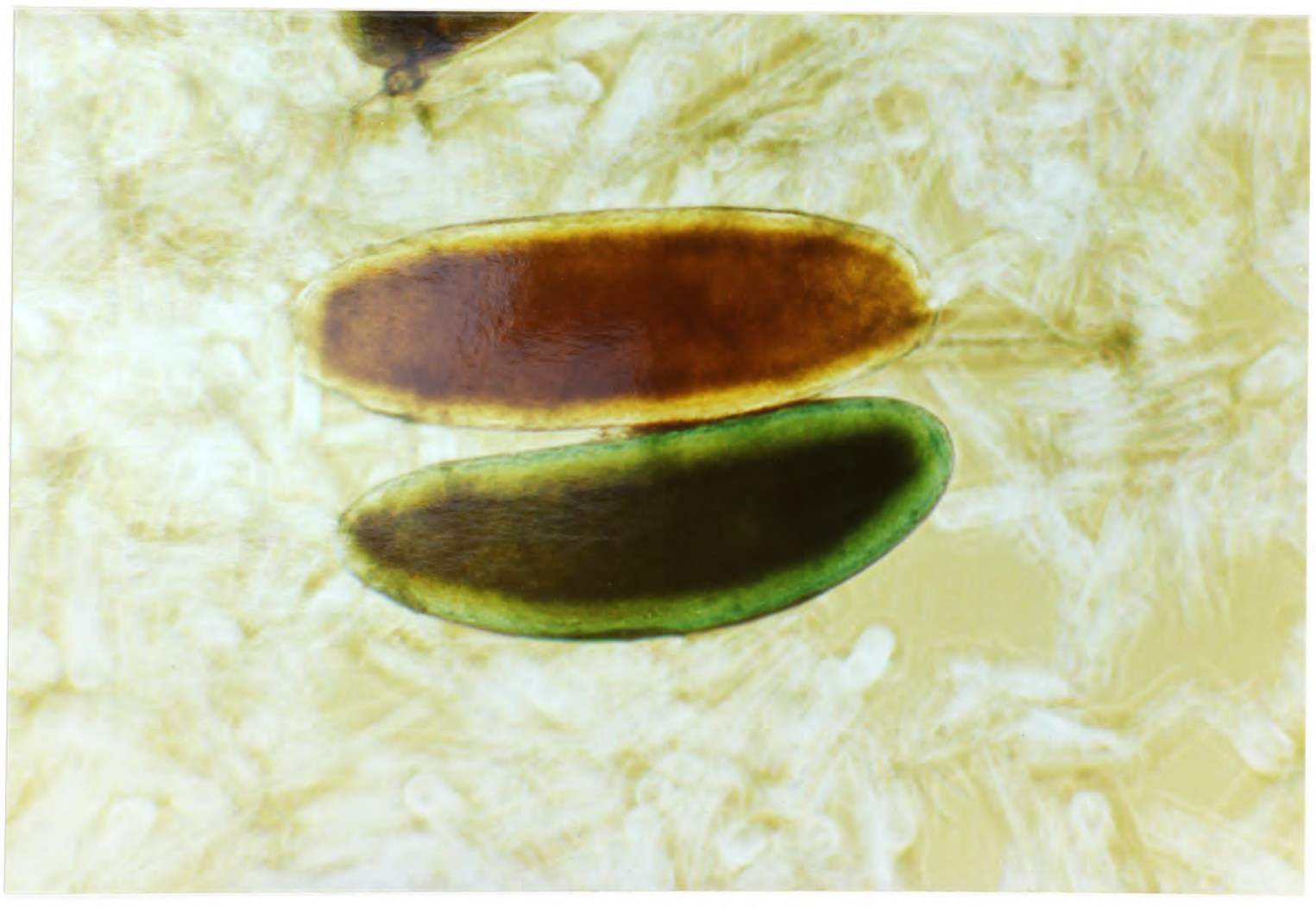


Figure 6.9 (cont). **700 A** embryos stained for β-galactosidase activity.

c) 2-4hrs



d) 4-6hrs



e) 6-9hrs

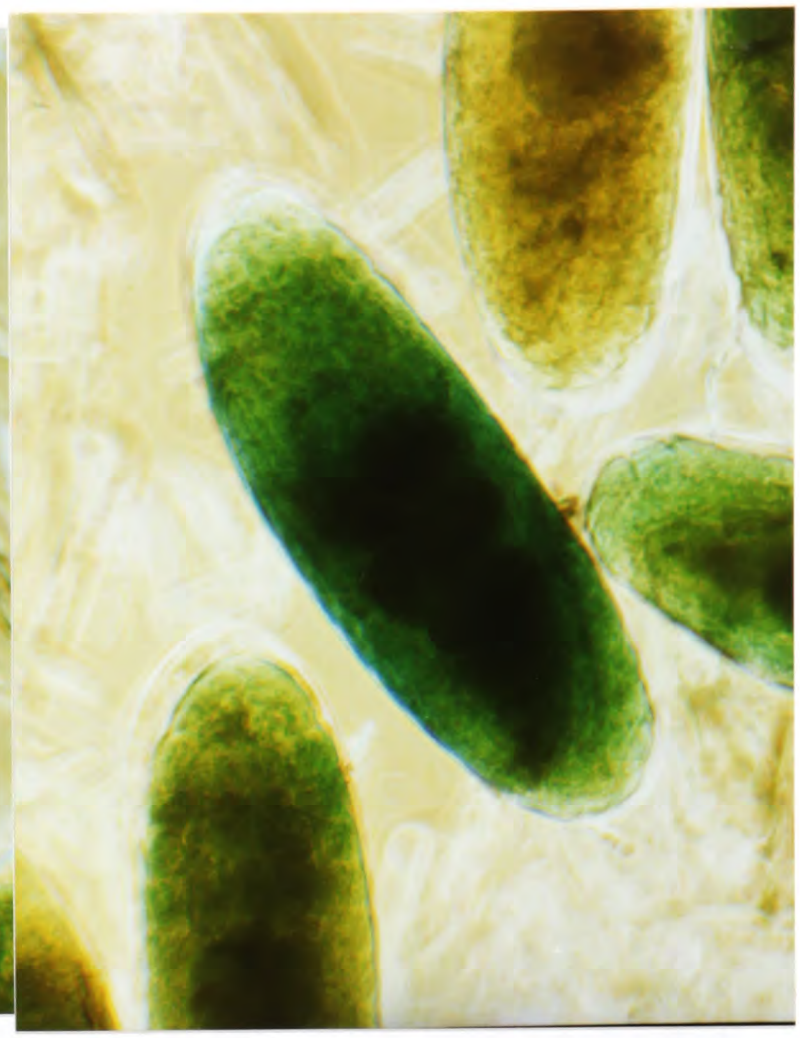


Figure 6.9 (cont). **700 A** embryos stained for β-galactosidase activity.

f) 10-13hrs



g) 13-16hrs



h) 16-18hrs

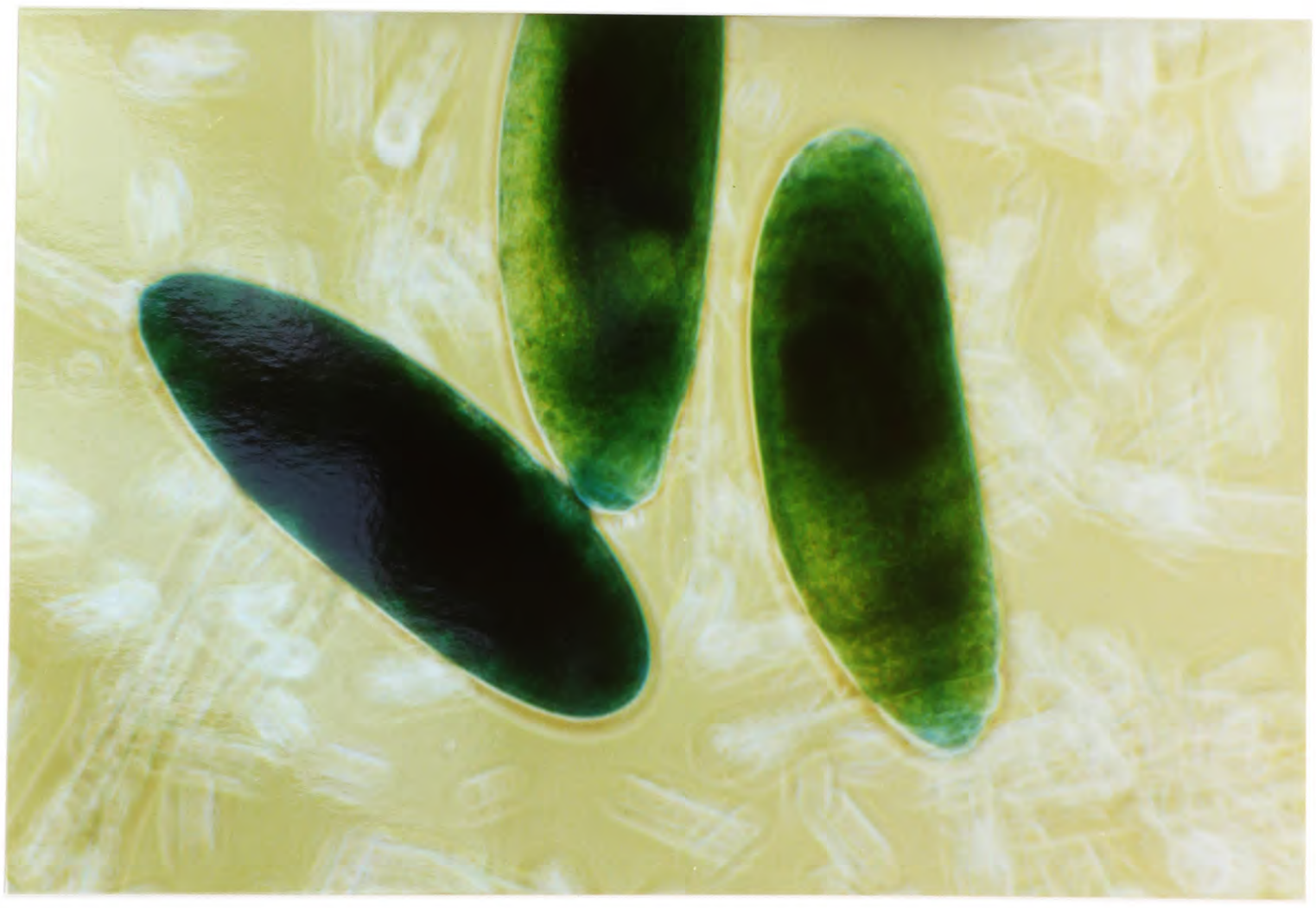


Figure 6.9 (cont). **700 A** embryos stained for β-galactosidase activity.

i) 18-20hrs

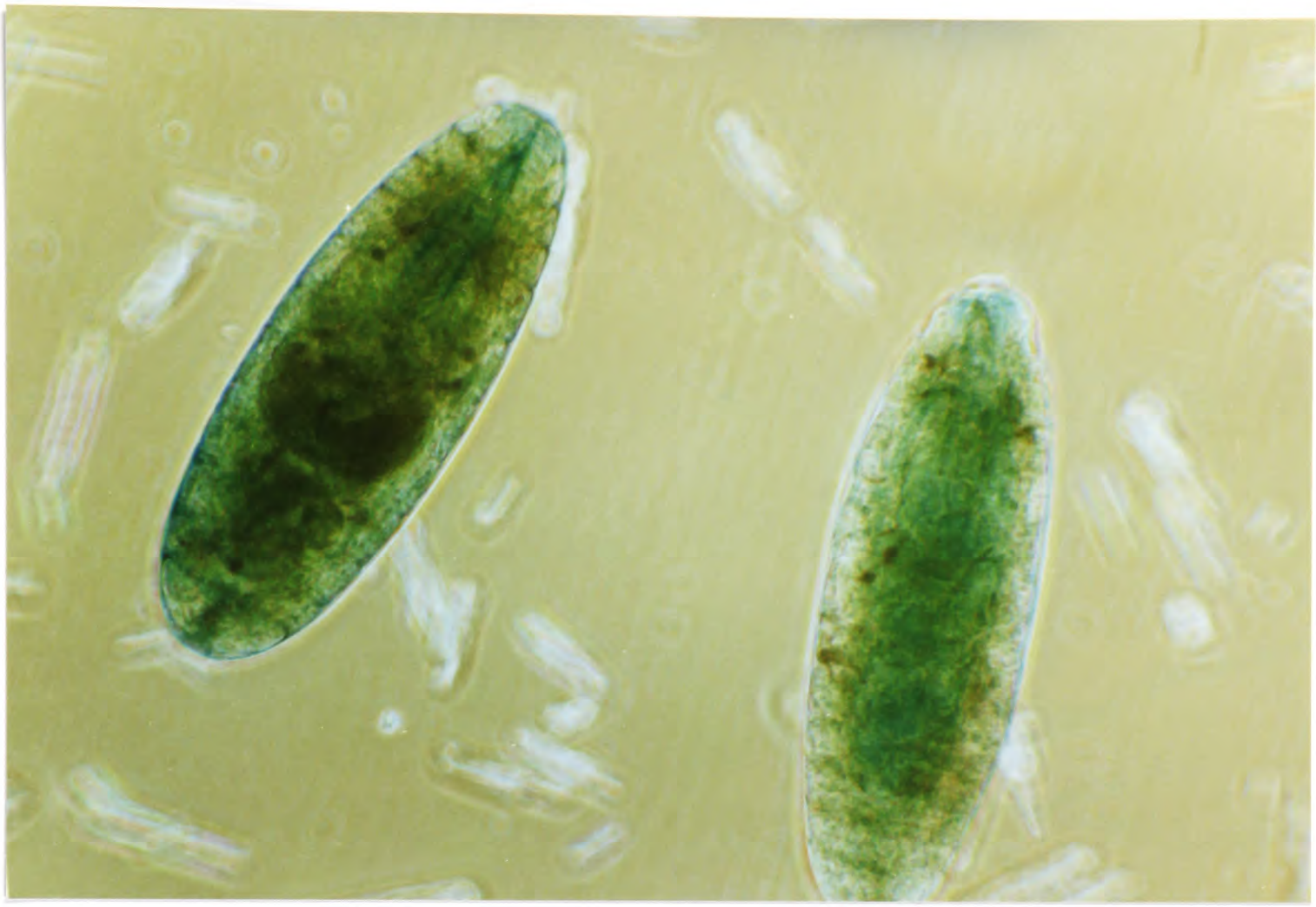


Figure 6.10. Summary of β-galactosidase activity for the full length promoter. Table of embryo staining.

Embryos	Nº. prepared	Nº. stained blue	Percentage
yw	217	6	2.8%
0-2hrs	22	16	72.7%
2-4hrs	173	101	58.4%
4-6hrs	54	28	51.9%
6-9hrs	62	39	62.9%
10-13hrs	352	330	93.8%
13-16hrs	225	172	76.4%
16-18hrs	205	163	79.5%
18-20hrs	82	62	75.6%
20-22hrs	88	40	45.5%

Figure 6.11. Imaginal discs stained for β-galactosidase activity.

a) *yw* wing imaginal disc.



b) neu-*lacZ* wing imaginal disc.



Figure 6.11 (cont). Imaginal discs stained for β-galactosidase activity.

c) **700A** imaginal discs.

i) wing disc.



ii) eye-antenal disc.



iii) leg disc.



galactose, identifies the developing neurons and axons of eye discs in *Drosophila* [Fristrom *et al.* 1982]. We might have expected that glycosylation was important during this process, for cell to cell recognition and co-ordination, and that enzymes involved would be expressed. That *GmII* is not expressed in imaginal discs is therefore surprising and it would be interesting to confirm this result by extracting mRNA from imaginal discs alone and using PCR to check the expression of both the native gene and the modified gene transcript (see section 5.6/7)

6.6 β-galactosidase expression of deleted constructs

Bulk collection of embryos from stocks carrying the two shorter promoter constructs were made. These embryos stained blue for β-galactosidase expression {2.20.1}. Some examples of stained embryos are shown in Figure 6.12a. This demonstrated that these lengths still provide enough 5' sequence for expression. Again some of the embryos did not stain. The percentages of embryos which stained blue are given in Figure 6.12b. There do not appear to be any differences in the staining patterns between embryos of stocks carrying the three different constructs. It appears that 288bp of 5' sequence is still sufficient for normal *GmII* expression.

6.7 Discussion and conclusions

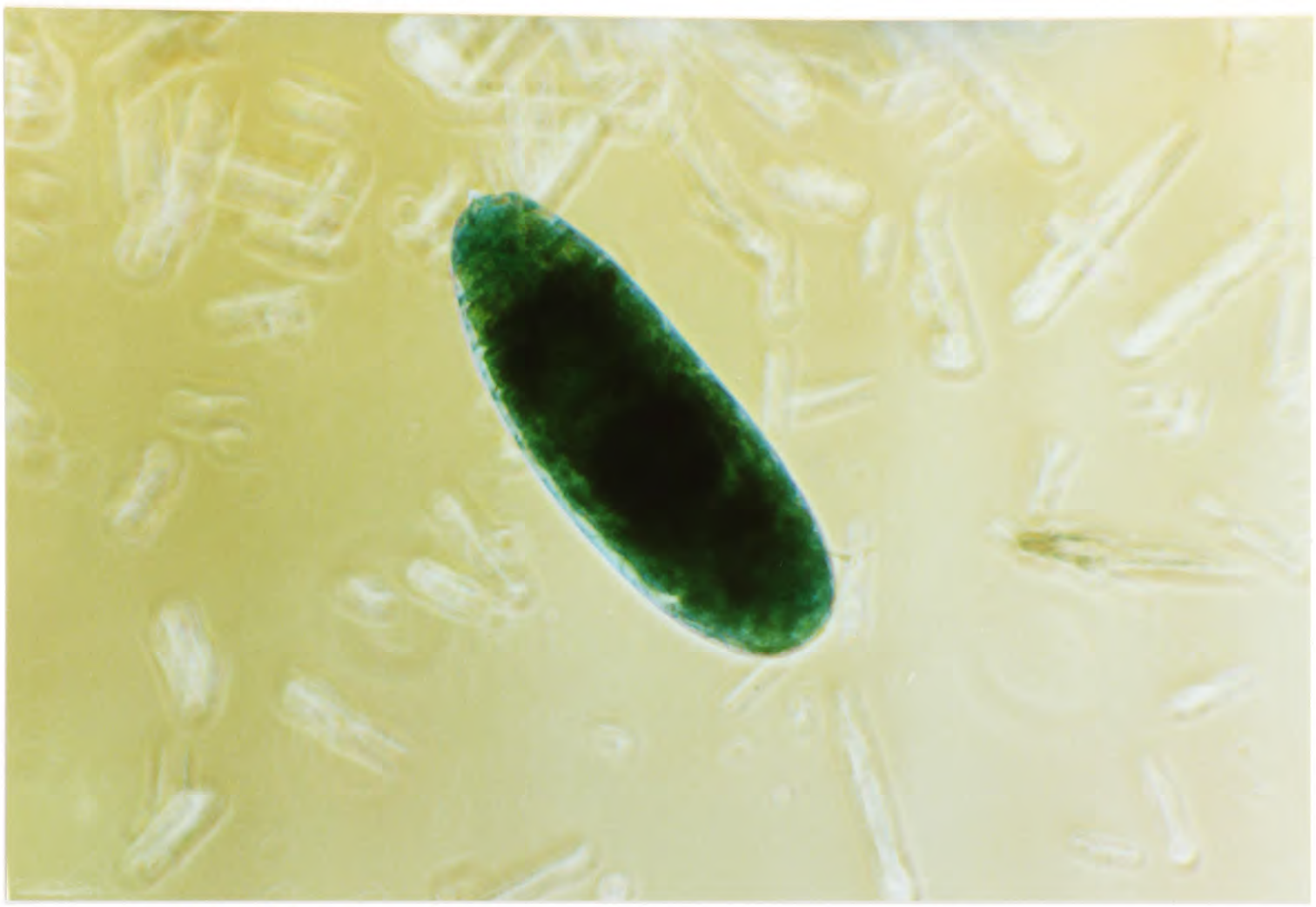
6.7.1 Expression of *GmII*

Investigating the expression of *GmII* by linking its functional promoter to *lacZ* suggested that *GmII* is expressed throughout embryo development. Staining whole embryos did not discriminate any differential expression. However imaginal discs from third instar larvae did not stain indicating that transformed *GmII* is not expressed in these discs. The results so far have been limited to embryos and imaginal discs and further investigation will be required to confirm these results and extend the information obtained (see section 6.7.3). However the ubiquitous expression in embryos suggests

Figure 6.12. Analysis of shorter constructs.

a) Embryos stained for β-galactosidase activity.

i) *yw* P[*white*$^+$ P1(500) *lacZ*]A.



ii) P[*white*$^+$ P1(288) *lacZ*]/*Tu* (2;3) *pr cn*/CyO/TM6, *mwh ry*^{506} *e.*



Figure 6.12 (cont). Analysis of shorter constructs.

ii) P[*white*$^+$ P1(288) *lacZ*]/*Tu* (2;3) *pr cn*/CyO/TM6, *mwh* ry^{506} *e* (cont).



b) Table of embryo staining percentages.

Embryos	N^{o}. prepared	N^{o}. stained blue	Percentage.
EcoII	163	101	62.0%
EcoIII	136	68	50.0%

that *GmII* is a vital protein required for development and maintains the assumption that mutants of *GmII* could be lethal. The lack of expression in imaginal discs seems somewhat surprising, but could indicate that the types of oligosaccharides required at this stage do not need Golgi mannosidase II expression and activity for their synthesis and so could be of the high mannose or hybrid types (see section 1.1.2).

6.7.2 *GmII* promoter

Shortening the length of 5' region attached to the *lacZ* gene demonstrated that 288bp of 5' sequence gave the same expression of *lacZ* as the 700bp of 5' sequence construct. The embryos from the shorter constructs did not show any differences in staining pattern. This suggests that there are no upstream enhancers in the region deleted required for embryo specific expression. Further work will be required to test that the expression of *GmII* at other stages of fly development is also unaffected.

6.7.3 Further work

There is still much potential to investigate the expression of *lacZ* from the *GmII* promoter in these transformed stocks. Unfortunately due to time constraints the investigations carried out were limited to whole embryos from all the *lacZ* stocks and imaginal discs from the 700bp sequence construct only. However future work could include an investigation of the expression in other developmental stages: larvae, pupae and adult flies. This would also include a closer examination of specific tissues and staining sections of flies. Investigating these areas in all three *lacZ* stocks would allow both the native expression to be elucidated and any aberrations of expression due to the shorter constructs to be demonstrated. It would also be interesting to compare the staining pattern seen with the 700bp construct with *in situ* hybridization of a digoxygenin (for example) labelled *GmII* cDNA probe to mRNA in whole embryos. This would test that normal expression was obtained from the 700bp construct.

CHAPTER SEVEN
GENERAL DISCUSSION

7.1 Calnexin

The discovery of a *Drosophila* calnexin homologue was a surprising and fortuitous result of a search for *Drosophila* ER mannosidase. There does not seem to be any reason why the rat mannosidase cDNA probe should have identified this *Drosophila* gene since no regions of sequence similarity were discovered. However now that this gene has been discovered in *Drosophila* it will be interesting to study it further. There seems to be considerable conservation of amino-acid sequence between the short region now sequenced in *Drosophila* and the other species in which the gene has been sequenced.

A *Drosophila* ER mannosidase has not been discovered. The reason for this may lie in the rat and *Drosophila* genes not being sufficiently conserved to allow cross hybridization between the species. Now that some doubt has been cast upon the role of the rat mannosidase this may be sufficient explanation. If the rat gene is not part of this highly conserved pathway, but an enzyme for glycoside degradation, it may not be sufficiently conserved in different species. A *Drosophila* ER mannosidase homologue to the more specific yeast mannosidase (see section 1.4.2) has yet to be identified.

7.2 Golgi mannosidase II

The *Drosophila* structural cDNA homologue of the mouse Golgi mannosidase II gene was used to clone the genomic copy of the gene. The investigations into the structure and expression of this gene were designed to uncover its biological significance. An investigation of the expression of *GmII* showed it to be ubiquitously expressed at the level of resolution investigated. Analysis of the 5' region showed

none of the usual promoter elements such as TATA, CAAT or a GC box. This was investigated further by introducing 5' lengths linked to the *lacZ* reporter gene and analysing the expression of this gene. Analysis so far has identified no differences in expression between the longest 700bp length of 5' sequence and the shortest construct of 288bp. This suggests no tissue specific enhancers in the developmental stages investigated.

The apparently ubiquitous expression of *GmII* may seem strange given that differences in glycosylation are required for the recognition role assigned to oligosaccharides. However the fact that a gene is expressed does not necessarily mean that it is translated or that the enzyme is active, has access to the oligosaccharides or has the correct substrate. Golgi mannosidase II is part of the pathway towards complex oligosaccharides and, assuming these are synthesized in *Drosophila*, the terminal differentiation producing a range of complex oligosaccharides may provide the specificity of structure required for recognition events.

Apart from the molecular analysis of *GmII*, we are also interested in the consequences to the fly of losing this gene. Its universal expression would suggest this effect to be dramatic and most probably lethal as there appears to be no alternative pathway at this point in the *N*-linked glycosylation pathway. To this end we are trying to identify a mutant of the gene from a saturation mutagenesis experiment using the entire *GmII* reintroduced into flies by P-element mediated transformation. This has as yet not identified a *GmII* mutant.

7.3 Future work

The aim of the laboratory is to investigate genes involved in glycosylation with a view to discovering the significance of oligosaccharides attached to proteins.

This a long term aim and requires in the interim considerable cloning and sequence and expression analysis.

The investigation for *Drosophila* Golgi mannosidase II is considerably advanced, but as yet has not revealed the significance of this gene as part of the *N*-linked glycosylation pathway. If it becomes apparent that the genes involved are not lethal, different approaches to saturation mutagenesis experiments to identify lethal mutants will have to be used.

The investigations into calnexin are just beginning. The presence of calnexin may explain the conservation of build-up pathway and *en bloc* transfer of oligosaccharide structure, to allow the correct folding of glycoproteins. This conservation suggests this to be a crucial step, at least for some proteins, although there are alternative processing enzymes at this point in the pathway. The analysis of this enzyme as part of the glycosylation pathway may provide some interesting answers as to the function of glycosylation of proteins.

7.4 Critical assessment of results and future direction for this work.

The aim of this project was to investigate which particular glycosylation structures are required for specific biological events by investigating the enzymes required for N-linked glycosylation. That specific oligosaccharides are involved has been demonstrated many times as discussed in Chapter one, but little is known about the criteria which determine whether any particular oligosaccharide is significant. One way to investigate which oligosaccharide structures are important is to investigate the enzymes involved in creating such structures. This is now being investigated in several mammalian species including man and the mouse. However, our aim is to investigate these enzymes in *Drosophila*, which is a useful genetic model.

The identification of enzymes in the *N*-linked glycosylation pathway in *Drosophila* is so far limited to Golgi mannosidase I [Kerscher *et al*, 1995] and the putative *GmII* discussed in this project. The *Drosophila* Golgi mannosidase I gene was discovered by chance by a group interested in the Drosophila *optomotor blind* mutation, and has since been passed to our group for further investigation of phenotypic effects. The Golgi mannosidase II investigation which is discussed in this thesis, was to understand the expression and control of expression of *GmII* to allow the significance of the oligosaccharide structures it creates to be assessed.

The genomic sequence of *GmII* contains four introns, but none contain any coding sequence and all are present in the cDNA. Analysis of first strand cDNA has shown this cDNA to be expressed throughout development, however a control for contaminating cDNA should have been included in this experiment. The full length sequence was used to transform *yw* flies with a view to rescuing mutants generated by Ben Yudkin. This construct had been altered by deleting a 151bp fragment from the untranslated 3' region. This construct has so far failed to identify a *GmII* mutant and although expression from the truncated construct has been demonstrated at the mRNA level, protein expression has not been demonstrated. The use of antibodies raised against the *GmII* protein or an assay for enzyme activity would confirm expression

from this construct, but such methods have so far been unsuccessful in identifying native enzyme activity (D. B. Roberts, pers. comm.).

Such methods would also be useful in assaying expression in different developmental stages and tissues which have so far only been investigated by RT-PCR and by linking the 5' region to the *lacZ* reporter gene. The *lacZ* expression was assayed directly by staining for its activity, but better pictures may be obtained by using antibodies against *lacZ*. The results of these analyses suggest that *GmII* is ubiquitously expressed throughout development and in most tissues, although there may be differences in the levels of expression. These stocks transformed with the three *lacZ* constructs still have much potential for investigating the expression of *GmII*.

The ultimate aim of our work is to identify the enzymes in the *N*-linked glycosylation pathway with a view to identifying the specific oligosaccharide structures required for the biological processes in which glycosylation has been shown to be required. Thus far the ubiquitous expression of *GmII*, an enzyme involved in the construction of complex oligosaccharides, may suggest that such structures are required. However we are still looking for more glycosylation enzymes by a variety of methods, including PCR using primers designed on the basis of previously cloned cDNAs in other organisms.

The identification of a *Drosophila* homologue of calnexin, a molecular chaperone involved in the folding of glycosylated proteins, is a separate investigation, but still involved in *N*-linked glycosylation. The *Drosophila* gene for calreticulin , a second ER localised molecular chaperone, has been previously identified and we intend further investigation of both these genes, to discover the part they play in interacting with glycosylated proteins. This again is with the aim of dissecting the process of *N*-linked glycosylation in *Drosophila* with a view to extrapolating the knowledge obtained to glycosylation in other organisms.

APPENDIX I

PRIMER CATALOGUE

This appendix shows the complete pP1 (E-E) subclone sequence with the primers used for sequencing both the genomic and cDNA clones. In addition PCR primers are also marked. Sequencing primers were usually 17bp in length and were also used for PCR where possible. Primers made specifically for PCR were usually longer (in the order of 21-25bp)

```
-700   AATTCCCATC TTTTGGCCTC CGTTGGTGAC TCCAGGGAAT CGAACACGGC

-650   GGCTCCAATT AAAAGATAGG TGAAAGTGCA TACCACCAGG GAAAGGGTAC

-600   GGACATTCTG TCGCTTCATC ATATAGTCTC AAATATAATC ACTACGATCT
                              P15U2 >
-550   ATAATCAATT AATATATTTT GCACCACGCG CCTTTTTCGT TCAGTTAAAG
                    < 3GJ5

-500   ATGCTGCGTC TGACAAAATC GATAGTTGTC AGCAAAACAA ACACACAGCA

-450   TCCGCCAAAA ATGGTTGATT TCGGTTACAG TGGCCGAATA ACATATGATA

-400   CAATGAATTT TAAAAATTTC ATCATTCTTT AAATATGTAA TAAGTTTTCT
         P15U3 >
-350   TTTATCTATT GATTTGACCG TTAATCTCAA TAAAGGATTG TGATTTTACC
                                                  < 2GJ5b
-300   GTTAGTTTAT CACAACAGAA GCAACAGAAT TTGATGGCGT TGCCACTTGG
                                                  P15 >
-250   CGGTATCTGT TAACTCTGTA CAGCTAAAAA GCTGTAATGT CACGATAACT

-200   AAGGATAATA ATAATTTTTT TAAAGGATCA ATATACACAA ACAACATAAG

-150   AGTTTTATTA CATTTTCTTT AATTGCCTTT AAAGTTTTTG TCTTTCTAAG
                                        < 1GJ5
-100   CCATTTTAAC AATTTAACAT TTTGCAGTTA GCAGCGCCGA TAACTTGACA

-50    GCCTATCGAT TTTCGGGAAA AGCAAACATA TCGCCAGATC TGATTTGTTT
       D2RIA >
0      ATGCTAAAAA AAAACCACGC AAGTTCCAAC CGATTGGATT GTTGCAGTTG
                   B-INTRON 1/1 >                  D2R6 >
50     GCTGCTGTGT CTCTGACGTC GACGTTTCCA TGCTTAATTA ACCGGAAAAT
                                                  < D2U4
100    GTTGCGAATA CGTCGGCGGT TCGCTTTGGT AATTTGCTCC GGCTGCCTGC
                             D2R2 >
```

150 TGGTTTTCCT CAGCCTGTAC ATAATCCTCA ATTTTGCGGC GCCGGCAGCC

200 ACCCAGATAA AGGTAAATAA ATGCTTCTAA ATCAATTCGA ATGTATAATA

250 ACTTCTTTAT TTTAGCAGCC CAACTATGAG AACATTGAGA ACAAGCTGCA

300 TGAGCTGGAA AATGGTTTGC AGGAGCACGG GGAGGAGATG CGGAATCTCA

< D2U3

350 GGGCGCGTCT TGCCGAAACA TCCAATCGCG ACGATCCAAT AAGACCTCCA

D2R3 > < INTRON 1/2

400 CTTAAAGTGG CTCGTTCCCC GAGGCCAGGG CAATGCCAAG ATGTGGTCCA

450 AGACGTGCCC AATGTGGATG TACAGATGCT GGAGCTATAC GATCGCATGT

< D2U2

500 CCTTCAAGGA CATAGATGGA GGCGTGTGGA AACAGGGCTG GAACATTAAG

D2R4 >

550 TACGATCCAC TGAAGTACAA CGCCCATCAC AAACTAAAAG TCTTCGTTGT

600 GCCGCACTCG CACAACGATC CTGGATGGAT TCAGACGTTT GAGGAATACT

B INTRON 2/1 > D2R5 > < D2U1

650 ACCAGCACGA CACCAAGCAC ATCCTGTCCA ATGCACTACG GCATCTGCAC

< J5B D2/D5 >

700 GACAATCCCG AGATGAAGTT CATCTGGGCG GAAATCTCCT ACTTTGCTCG

750 GTTCTATCAC GATTTGGGAG AGAACAAAAA GCTGCAGATG AAGTCGTAAG

800 CATGCTTTCC CTATATAGTA TTGATACAAT AATGAATAAC GGTAGTTAAC

850 CCACAAAACA AAGTGCAATG CCATTTTGCT ATTGCATTCC ATAAACCTTA

P1UC >

900 CACAGTTGTT ATGTGCAGCA GTGAAATATT TTACTTCGTT ATTGCGTCAA

950 AAGCGTTTAG ATAAACAAAA TGCATGATCA CGCTGCTTGT TTTCCGAAAC

1000 AGCTTTGTAC TTATCTACCT AATATTTATA ACCCAAAGCA GAAATAGGAA

1050 TTGATAAGCT TTGCAGTCTC AACAACTACA ATATTTAGTT TTTAACAAAA

1100 AATAATAAAT TATTTACTTA TTTACAGCAT TGTAAAGAAT GGACAGTTGG

< P1I1

1150 AATTTGTGAC TGGAGGATGG GTAATGCCGG ACGAGGCCAA CTCCCACTGG

< INTRON 2/2

1200 CGAAACGTAC TGCTGCAGCT GACCGAAGGG CAAACATGGT TGAAGCAATT

< D5/D2

1250 CATGAATGTC ACACCCACTG CTTCCTGGGC CATCGATCCC TTCGGACACA

D5R1 >

1300 GTCCCACTAT GCCGTACATT TTGCAGAAGA GTGGTTTCAA GAATATGCTT

< D5U3

1350 ATCCAAAGGA CGCACTATTC GGTTAAGAAG GAACTGGCCC AACAGCGACA

1400 GCTTGAGTTC CTGTGGCGCC AGATCTGGGA CAACAAAGGG GACACAGCTC

< D5U2

1450 TCTTCACCCA CATGATGCCC TTCTACTCGT ACGACATTCC TCATACCTGT

D5R2 >

1500 GGTCCAGATC CCAAGGTTTG CTGTCAGTTC GATTTCAAAC GAATGGGCTC

1550 CTTCGGTTTG AGTTGTCCAT GGAAGGTGCC GCCGCGTACA ATCAGTGATC

D5R3 >

1600 AAAATGTGGC AGCACGCTCA GATCTGCTGG TTGATCAGTG GAAGAAGAAG

< D5U1

1650 GCCGAGCTGT ATCGCACGAA CGTGCTGCTG ATTCCGTTGG GTGACGACTT

1700 CCGCTTCAAG CAGAACACCG AGTGGGATGT GCAGCGCGTG AACTACGAAA

1750 GGCTGTTCGA ACACATCAAC AGCCAGGCCC ACTTCAATGT CCAGGCGCAG

1800 TTCGGCACAC TGCAGGAATA CTTTGATGCA GTGCACCAGG CGGAAAGGGC

1850 GGGACAAGCC GAGTTTCCCA CGCTAAGCGG TGACTTTTTC ACATACGCCG

< D14U1

1900 ATCGATCGGA TAACTATTGG AGTGGCTACT ACACATCCCG CCCGTATCAT

D14R1 >

1950 AAGCGCATGG ACCGCGTCCT GATGCACTAT GTACGTGCAG CAGAAATGCT

2000 TTCCGCCTGG CACTCCTGGG ACGGTATGGC CCGCATCGAG GAACGTCTGG

2050 AGCAGGCCCG CAGGGAGCTG TCATTGTTCC AGCACCACGA CGGTATAACT

2100 GGCACAGCAA AAACGCACGT AGTCGTCGAC TACGAGCAAC GCATGCAGGA

2150 AGCTTTAAAA GCCTGTCAAA TGGTAATGCA ACAGTCGGTC TACCGATTGC

< DHR6

2200 TGACAAAGCC CTCCATCTAC AGTCCGGACT TCAGTTTCTC GTACTTTACG

2250 CTCGACGACT CCCGCTGGCC AGGATCTGGT GTGGAGGACA GTCGAACCAC

2300 CATAATACTG GGCGAGGATA TACTGCCCTC CAAGCATGTG GTGATGCACA

DHU4 > < DHR5 < DHR4A

2350 ACACCCTGCC CCACTGGCGG GAGCAGCTGG TGGACTTTTA TGTATCCAGT

2400 CCGTTTGTAA GCGTTACCGA CTTGGCAAAC AATCCGGTGG AGGCTCAGGT

2450 GTCCCCGGTG TGGAGCTGGC ACCACGACAC ACTCACAAAG ACTATCCACC

< P1RA DHU2 >

2500 CACAAGGCTC CACCACCAAG TACCGCATCA TCTTCAAGGC TCGGGTGCCG

DHU2b >

2550 CCCATGGGCT TGGCCACCTA CGTTTTAACC ATCTCCGATT CCAAGCCAGA

2600 *GTGAGTAGCC AGTTAAAGAT TACCCATTTA TCAAAGGAAC AAGAATCCTT*

2650 *TCCGTACTTT CTGTATTAAT TGTACATTTT CGTAACACGT TTCGCTATGC*

2700 *CGTTTCCTTA AACCATTATA GAGTGCAGGT TTTATTTATT GATATTAATA*

2750 *TTTCATTAAC CAGATGCCGT AATTGTCTTC TATTTATACT ATTTATTTAT*

< DHR4

2800 *ACTTTTTCAA TAAGCATTTC AATGTTACTT ACAG*GCACAC CTCGTATGCA

< P1I2

2850 TCGAATCTCT TGCTCCGTAA AAACCCGACT TCGTTACCAT TGGGCCAATA

2900 TCCGGAGGAT GTGAAGTTTG GCGATCCTCG AGAGATCTCA TTGCGGGTTG

2950 GTAACGGACC CACCTTGGCC TTTTCGGAGC AGGGTCTCCT TAAGTCCATT

DHU3 >

3000 CAGCTTACTC AGGATAGCCC ACATGTACCG GTGCACTTCA AGTTCCTCAA

< DHR3

3050 GTATGGCGTT CGATCGCATG GCGATAGATC CGGTGCCTAT CTGTTCCTGC

< DHR3B

3100 CCAATGGACC AGCTTCGCCA GTCGAGCTTG GCCAGCCAGT GGTCCTGGTG

DHU4 >

3150 ACTAAGGGCA AACTGGAGTC GTCCGTGAGC GTGGGACTTC CGAGCGTGGT

3200 GCACCAGACG ATAATGCGCG GTGGTGCACC TGAGATTCGC AATCTGGTGG

< DHR2

3250 ATATAGGCTC ACTGGACAAC ACGGAGATCG TGATGCGCTT GGAGACGCAT

3300 ATCGACAGCG GCGATATCTT CTACACGGAT CTCAATGGAT TGCAATTTAT

DHU5 >

3350 CAAGAGGCGG CGTTTGGACA AATTACCTTT GCAGGCCAAC TATTATCCCA

3400 TACCTTCTGG TATGTTCATT GAGGATGCCA ATACGCGACT CACTCTCCTC

< DHR1

3450 ACGGGTCAAC CGCTGGGTGG ATCTTCTCTG GCCTCGGGCG AGCTAGAGAT

3500 TATGCAAGAT CGTCGCCTGG CCAGCGATGA TGAACGCGGC CTGGGACAGG

3550 GTGTTTTGGA CAACAAGCCG GTGCTGCATA TTTATCGGCT GGTGCTGGAG

3600 AAGGTTAACA ACTGTGTCCG ACCGTCAGAG CTTCATCCTG CCGGCTATTT

P1UB >

3650 GACAAGTGCC GCTCACAAAG CATCGCAGTC ACTGCTGGAT CCACTGGACA

3700 AGTTTATATT CGCTGAAAAT GAGTGGATCG GGGCACAGGG GCAATTTGGT

< D4U1 B INTRON 4/1 > D4R1>

3750 GGCGATCATC CTTCGGCTCG TGAGGATCTC GATGTGTCGG TGATGAGACG

3800 CTTAACCAAG AGCTCGGCCA AAACCCAGCG AGTAGGCTAC GTTCTGCACC

3850 GCACCAATCT GATGCAATGC GGCACTCCAG AGGAGCATAC ACAGAAG*GTG*

3900 *AGCTTAGAAG TTTATTATAT ATTAATATCA AAGAAAGTGA AATGGTATTT*

3950 *ATTTATAATT AACAATATTG TTTGACTAAT TGTTCGTTCG TCAG*CTGGAT

4000 GTGTGCCACC TACTGCCGAA TGTGGCGAGA TGCGAGCGCA CGACGCTGAC

4050 TTTCCTGCAG AATTTGGAGC ACTTGGATGG CATGGTGGCG CCGGAAGTGT

< INTRON 4/2 < SC 'B'

4100 GCCCCATGGA AACCGCCGCT TATGTGAGCA GTCACTCAAG CTGATCGCAA

< D23U1 D23R1 >

4150 CGTGTTATGG CAAAGGATAA AGCAGAAACT GAAGCCAAGC GACCATCTCA

4200 CAACCTATGA ATCTCGGACG CTAATCCACT GCAATACAGG CAATCTTCGT

4250 TTAATGTTCG ACTTAAGCAC TGACTGCACC GAAAGACGTA ATGCGTAATC

4300 CATACTACGT ATCATTTATT TATTCAAGCT TCCACACGAG GGTTTGGAAA

4350 TTCAATCCTA TGACGCAGCG GTGTATTTTT CGCACCCCCG TGTAACTACA

D9U1 >

4400 CGTTTTACTC TGTTTCCCCT AGTTTTTTAC GTTTTTTTGT GATATCCATC

4450 CAGTCCCACA GCAAGGCAAT ACATGTCTGT ACTCGATATC CCCAATGGGG

< J5A1

4500 TTTTATTGCC TATAACAATG AATCGTAACT ATAACTAACG TTTAACTGTA

P13U1 >

4550 AAGTAGGCTT AGTGATAATG ACGCAACTGC GACTGCAATG ATATTAATCT

< RNAR1

4600 CGACGTTTGT TTTTTACAAA TTGAATAAAA TTGAGTCTCA TTTAGAATCG

4650 ACTTTTACGT TGCTATTCTA AACTAAAAAC ATAACGTCTC TTTTTATTCG

4700 AAAAAGGAAA GGATTTGTAT TACTATTCAA AATTCGTCAT TTTGGCTTTA

3P1U2 >

4750 GTTAGAATCG TGTTGTTGGT AAGTTATGAA TTGCACTTTC TATTGAATAT

4800 TTAATTTTTC TCAGTAAATC TAAAATAAAT ACATTTTAAT TTAATTTTTA

4850 AGATTAGATT TAATTTTTTT TTTAATATTG TAAGTAATGA ATGTTAGTAA

4900 ACAGTTGATC TTTGTATTAT TAAGTTTTGA ATTTATGTGG ATTATGAAAC

4950 GAGCTTTTGA GACGACTAAT TAAAGCCATA ATCATGAGTT ATAGTTTTAG

< 3P1R1

5000 TGGATGAAAA ACAGAATGCA ATCTTATCAG CTCCAGTGCG GGGTACTATT

5050 ATTGCCAGAA GACCTACAGA GATGAGATAC TAAGAGCGCG GAATTGGAAA

5100 ATGCACAATG GAAATGGAAA TCCGGATTGA GTGTAAAGTG GTTGTTTTGG

5150 ATAGAGAGCG AATCGTGCGC AGTTCGCTCT CTTCAAAATC CGAAACGAGC

5200 AACGAAAGCA AAACCGAATC CGAATCGGAA TCCGAAA

APPENDIX II

DROSOPHILA INBRED STOCKS

S12A, S11B, S11C: Three separate isogenic lines established from single flies caught in Strömsvreten (Sweden)

1AO: (Lsp-1 αo βo γo.) Isogenic line carrying null alleles of Lsp-1 (Larval serum protein)

OrR: (Oregon-R.) Isogenic line from a strain caught originally in Oregon, USA (1926)

BrA, BrB: Two separate isogenic lines established from single flies caught in Brightwell, Oxfordshire.

IgB: Isogenic line established from a female caught in Ightham, Kent.

APPENDIX III

DNA AND AMINO ACID CODES

a) IUPAC codes for DNA

A or C = M	G or T = K
A or G = R	C, G or T = B
A or T = W	A, G or T = D
C or G = S	A, C or T = H
C or T = Y	A, C or G = V
A, G, C or T = N	

b) Amino acid codes

Alanine	Ala	A	Leucine	Leu	L
Arginine	Arg	R	Lysine	Lys	K
Asparagine	Asn	N	Methionine	Met	M
Aspartic acid	Asp	D	Phenylalanine	Phe	F
Cysteine	Cys	C	Proline	Pro	P
Glutamine	Gln	Q	Serine	Ser	S
Glutamic acid	Glu	E	Threonine	Thr	T
Glycine	Gly	G	Tryptophan,	Trp	W
Histidine	His	H	Tyrosine	Tyr	Y
Isoleucine	Ile	I	Valine	Val	V

REFERENCES

Altmann, F. and Marz, L. [1995] Processing of asparagine-linked oligosaccharides in insect cells: evidence for α-mannosidase II. *Glycoconj. J.*, **12,** 150-5.

Aquadro, C. F. [1992] Why is the genome variable? Insights from *Drosophila. TIG,* **8,** 355-362.

Aquadro, C. F. [1993] in Molecular approaches to fundamental and applied entomology. J. Oakeshott and M. Whitten Springer-Verlag.

Arkhipova, I. R. [1995] Promoter elements in *Drosophila melanogaster* revealed by sequence analysis. *Genetics,* **139,** 1359-1369.

Ashburner, M. [1989a] in Drosophila - A laboratory handbook. Cold Spring Harbour Laboratory Press.

Ashburner, M. [1989b] Drosophila - A laboratory mannual. Cold Spring Harbor Laboratory Press.

Ashwell, G. and Harford, J. [1982] Carbohydrate-specific receptors of the liver. *Annu. Rev. Biochem.,* **51,** 531-554.

Balaji, P. V., Qasba, P. K. and Rao, V. S. [1994] Molecular dynamics simulations of high-mannose oligosaccharides. *Glycobiology,* **4,** 497-515.

Bause, E., Bieberich, E., Rolfs, A., Volker, C. and Schmidt, B. [1993] Molecular cloning and primary structure of Man_9-mannosidase from human kidney. *Eur. J. Biochem.,* **217,** 535-540.

Bause, E., Erkins, R., Schweden, J. and Jaenicke, L. [1986] Purification and characterization of trimming glucosidase I from *Saccharomyces cerevisiae. FEBS. Lett.,* **206,** 208-212.

Bause, E., J, S., Gross, A. and Orthen, B. [1989] Purification and characterization of trimming glucosidase I from pig liver. *Eur. J. Biochem.*, **183,** 661-9.

Bergeron, J. J. M., Brenner, M. B., Thomas, D. Y. and Williams, D. B. [1994] Calnexin: a membrane-bound chaperone of the endoplasmic reticulum. *Trends Biochem. Sci.,* **19,** 124-128.

Beyer, A. B. and Osheim, Y. N. [1988] Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes and Development,* **2,** 754-765.

Birnstiel, M. L., Busslinger, M. and Strub, K. [1985] Transcription termination and 3' processing: the end is in site. *Cell,* **41,** 349-359.

Bischoff, J. and Kornfeld, R. [1983] Evidence for an α-mannosidase in endoplasmic reticulum of rat liver *J. Biol. Chem,* **258,** 7907 - 7910.

Bischoff, J. and Kornfeld, R. [1986] The soluble form of rat liver α-mannosidase is immunologically related to the endoplasmic reticulum membrane α-mannosidase. *J. Biol. Chem,* **261,** 4758-4765.

Bischoff, J., Moremen, K. and Lodish, H. F. [1990] Isolation, characterization and expression of cDNA encoding a rat liver endoplasmic reticulum α-mannosidase *J. Biol. Chem.*, **265,** 17110 - 17117.

Blake, C. C. F. [1978] Do genes-in-pieces imply proteins-in-pieces? *Nature,* **273,** 267.

Botchan, M., Topp, W. and Sambrook, J. [1976] The arrangement of simian virus 40 sequences in the DNA of transformed cells. *Cell,* **9,** 269-287.

Boyce, J. M., Coates, D., Fricker, M. D. and Evans, D. E. [1994] Genomic sequence of a calnexin homologue from *Arabdopsis thaliana*. *Plant Physiol.*, **106,** 1691.

Brand, A. H. and Perrimon, N. [1993] Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development,* **118,** 401-415.

Brown, N. H. and Kafatos, F. C. [1988] Functional cDNA libraries from *Drosophila* embryos. *J. Mol. Biol,* **203,** 423-437.

Bucher, P. [1990] Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J. Mol. Biol.,* **212,** 563-578.

Butters, T. D., Hughes, R. C. and Vischer, P. [1981] Steps in the biosynthesis of mosquito cell membrane glycoproteins and the effects of tunicamycin. *Biochim. Biophys. Acta,* **640,** 672-686.

Callaerts, P., Vulsteke, V., Peumans, W. and De Loof, A. [1995] Lectin binding sites during *Drosophila* embryogenesis. *Arch. Dev. Biol.,* **204,** 229-243.

Camirand, A., Heysen, A., Grondin, B. and Herscovics, A. [1991] Glycoprotein biosynthesis in *Saccharomyces cerevisiae. J. Biol. Chem.,* **266,** 15120-15127.

Charuk, J. H., Tan, J., Berardini, M., Haddad, S., Reithmeier, R. A., Haeken, J. and Schachter, H. [1995] Carbohydrate-deficient glycoprotein syndrome type II. An autosomal recessive N-acetylglucosaminyltransfarase II deficiency different from typical hereditary erythroblastic multinuclearity, with a positive acidified-serum lysis test (HEMPAS). *Eur. J. Biochem.,* **230,** 797-805.

Chen, C. N., Malone, T., Beckendorf, S. K. and Davis, R. L. [1987] At least two genes reside within a large intron of the *dunce* gene of *Drosophila*. *Nature,* **329,** 721-4.

Cherbas, L. and Cherbas, P. [1993] The arthropod initiator: the capsite consensus plays an important role in transcription. *Insect Biochem. Mol. Biol.,* **189,** 617-631.

Church, G. M. and Gilbert, W. [1984] Genomic sequencing. *PNAS USA,* **81,** 1991-1995.

Crowley, T. E., Hoey, T., Liu, J.-K., Jan, Y. N., Jan, L. Y. and Tjian, R. [1993] A new factor related to TATA-binding protein has highly restricted expression pattern restricted expression pattern in *Drosophila*. *Nature,* **361,** 557-61.

Csank, C., Taylor, F. M. and Martindale, D. W. [1990] Nuclear pre mRNA introns : analysis and comparison of intron sequences from *Tetrahymena thermophilia* and other eukaryotes. *Nuc. Acids. Res.,* **18,** 5133-41.

Cubas, P., de Celis, J.-F., Campuzano, S. and Modolell, J. [1991] Proneural clusters of *achaete-scute* expression and the generation of sensory organs in the *Drosophila* imaginal wing disc. *Genes & Dev.,* **5,** 996-1008.

D'Agostaro, G. A., Zingoni, A., Moritz, R. L., Simpson, R. J., Schachter, H. and Bendiak, B. [1995] Molecular cloning and expression of cDNA encoding the rat UDP-N-acetylglucosamine:a-6-D mannsise b-1,2-N-acetylglucosamyltransferase II. *J. Biol. Chem.,* **270,** 15211-21.

David, V., Hochstenbach, F., Rajagopalan, S. and Brenner, M. B. [1993] Interaction with newly synthesized and retained proteins in the endoplasmic reticulum suggests a chaperone function for human integral membrane protein IP90 (calnexin) *J. Biol. Chem.,* **268,** 9585-9592.

Davidson, D. J. and Castellino, F. J. [1990] Oligosaccharide processing in the expression of Human Plasminogen cDNA by Lepidopteran Insect (*Spodoptera frugiperda*) cells. *Biochemistry,* **29,** 5584-90.

Davidson, D. J. and Castellino, F. J. [1991a] Asparagine-linked oligosaccharide processing in lepidopteran insect cells. Temporal dependence of the nature of the oligosaccharides assembled on Asparagine-289 of recombinant human plasminogen produced in baculovirus vector infected *Spodoptera frugiperda* (IPLB-SF-21AE) cells. *Biochemistry,* **30,** 6165-74.

Davidson, D. J. and Castellino, F. J. [1991b] α-Mannosidase-catalysed trimming of high-mannose glycans in noninfected and baculovirus-infected *Spodoptera frugiperda* cells (IPLB-SF-21AE). A possible contributing regulatory mechanism for assembly of complex-type oligosaccharides in infected cells. *Biochemistry,* **30,** 9811-5.

de Virgilio, C., Buerckert, N., Neuham, J.-M., Boller, T. and Wiemken, A. [1993] CNE1, a *Saccharomyces cerevisiae* homologue of the genes encoding mammalian calnexin and calreticulin. *Yeast,* **9,** 185-188.

Degen, E. and Williams, D. B. [1991] Participation of a novel 88kD protein in the biogenesis of murine class I histocompatibility molecules. *J. Cell. Biol,* **112,** 1099-1115.

Elbein, A. D. [1987] Inhibitors of the biosynthesis and processing of *N*-linked oligosaccharide chains. *Annu. Rev. Biochem.,* **56,** 497-534.

Elbein, A. D. [1991a] Glycosidase inhibitors as antiviral and/or antitumor agents. *Semin. Cell. Biol.,* **2,** 309-17.

Elbein, A. D. [1991b] Glycosidase inhibitors: Inhibitors of *N*-linked oligosaccharide processing. *FASEB J.,* **5,** 3055-3063.

Fedorov, A., Suboch, G., Bujakov, M. and Fedorova, L. [1992] Analysis of nonuniformity in intron phase distribution. *Nuc. Acids. Res.,* **20,** 2553-7.

Feizi, T. [1985] Demonstration by monoclonal antibodies that carbohydrate structures of glycoproteins and glycolipids are onco-developmental antigens. *Nature,* **314,** 53-57.

Feizi, T. and Larkin, M. [1991] AIDS and glycosylation. *Glycobiology,* **1,** 17-24.

Feng, D. F. and Doolittle, R. F. [1987] Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J. Mol. Evol.,* **25,** 351-360.

Foster, J. M., Yudkin, B., Lockyer, A. E. and Roberts, D. B. [1995] Cloning and sequence analysis of *GmII*, a *Drosophila melanogaster* homologue of the cDNA encoding murine Golgi α-mannosidase II. *Gene,* **154,** 183-186.

Fredieu, J. R. and Mahowald, A. P. [1994] Glycoconjugate expression during *Drosophila* embryogenesis. *Acta Anat.,* **149,** 89-99.

Fristrom, D. K. and Fristrom, J. W. [1982] Cell surface binding sites for peanut agglutinin in the differentiating eye disc of *Drosophila*. *Dev. Biol.,* **92,** 418-427.

Fujimoto, K. and Kornfeld, R. [1991] α-Glucosidase II-deficient cells use endo α-mannosidase as a by-pass route for *N*-linked oligosaccharide processing. *J. Biol. Chem.,* **266,** 3571-3578.

Fukuda, M. N. [1990] HEMPAS disease: genetic defect of gycosylation. *Glycobiology,* **1,** 9-15.

Fukuda, M. N., Dell, A. and Scartezzini, P. [1987] Primary defect of congenital dyserythropoietic anaemia Type II *J. Biol. Chem,* **262,** 7195-7206.

Fukuda, M. N., Masri, K. A., Dell, A., Luzzatto, L. and Moremen, K. W. [1990] Incomplete synthesis of *N*-glycans in congenital dyserythropoietic anemia type II caused by a defect in the gene encoding α-mannosidase II. *PNAS USA,* **87,** 7443-7447.

Geisow, M. J. [1991] The coming of age of glycobiology. *TIBTech,* **9,** 221-225.

Ghysen, A. and O'Kane, C. [1989] Neural enhancer-like elements as specific cell markers in *Drosophila*. *Development,* **105,** 35-52.

Gilbert, W. [1978] Why genes in pieces? *Nature,* **271,** 501.

Gill, G. [1994] Taking the initiative. *Curr. Biol.,* **4,** 374-376.

Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. [1990] Basic Local Alignment Search Tool. *J. Mol. Biol.,* **215,** 403-410.

Gish, W. and States, J. D. [1993] Identification of protein regions by database similarity search. *Nat. Genet.,* **3,** 266-72.

Goochee, C. F. and Monica, T. [1990] Environmental effects on protein glycosylation. *Bio/Technology,* **8,** 421-427.

Gubler, V. and Hoffman, B. J. [1983] A simple and very efficient method for generating cDNA libraries. *Gene,* **25,** 263-269.

Hammond, C., Braakman, I. and Helenius, A. [1994] Role of *N*-linked oligosaccharide recognition, glucose trimming, and calnexin in glycoprotein folding and quality control. *PNAS USA,* **91,** 913-917.

Hansson, G. C., Karlsson, K.-A., Larson, G., Stromberg, N., Thurin, J., Orvell, C. and Norrby, E. [1984] A novel approach to the study of glycolipid receptors in viruses. *FEBS lett.,* **170,** 15-18.

Hardingham, T. E. and Fosang, A. J. [1992] Proteoglycans: many forms and many functions. *FASEB J.,* **6,** 861-870.

Hawn, T. R., Tom, T. D. and Strand, M. [1993] Molecular cloning and expression of SmIrV1, a *Schistosoma mansoni* antigen with similarity to calnexin, calreticulin and OvRaII. *J. Biol. Chem.,* **268,** 7692-7698.

Hazelrigg, T. [1987] The *Drosophila white* gene: a molecular update. *TIG,* **3,** 43-47.

Hazelrigg, T., Levis, R. and Rubin, G. M. [1984] Transformation of *white* locus DNA in *Drosophila*: dosage compensation, *zeste* interaction and position effects. *Cell,* **36,** 469-481.

Hernandez, N. [1993] TBP, a universal eukaryotic transcription factor? *Genes Dev.,* **7,** 1291-1308.

Herscovics, A., Schneikert, J., Athanassiadis, A. and Moremen, K. [1994] Isolation of a mouse Golgi mannosidase cDNA, a member of a gene family conserved from yeast to mammals. *J. Biol. Chem.*, **269,** 9864-9871.

Hettkamp, H., Legler, G. and Bause, E. [1984] Purification by affinity chromatography of glucosidase I, an endoplasmic reticulum hydrolase involved in the processing of asparagine-linked oligosaccharides. *Eur. J. Biochem.*, **142,** 85-90.

Higgins, D. G. and Sharp, P. M. [1989] Fast and sensitive multiple sequence alignments on a microcomputer. *Comp. Appl. Biosci.*, **5,** 151-153.

Hoheisel, J. D., Lennon, G. G., Zehetner, G. and Lehrach, H. [1991] Use of high coverage reference libraries of *Drosophila melanogaster* for relational data analysis. *J. Mol. Biol.*, **220,** 903 - 914.

Homans, S. W., Dwek, R. A. and Rademacher, T. W. [1987] Solution conformations of *N*-linked oligosaccharides. *Biochemistry,* **26,** 6571-6578.

Hook, M., Kjellen, L. and Johansson, S. [1984] Cell-surface glycosaminoglycans. *Ann. Rev. Biochem.*, **53,** 847-869.

Huang, L., Franklin, A. E. and Hoffman, N. E. [1993] Primary structure and characterization of an *Arabidopsis thaliana* calnexin-like protein. *J. Biol. Chem.*, **268,** 6560-6566.

Hultmark, D., Klemenz, R. and Gehring, W. J. [1986] Translational and transcriptional control elements in the untranslated leader of the heat-shock gene hsp22. *Cell,* **44,** 429-438.

Ioffe, E. and Stanley, P. [1994] Mice lacking *N*-acetylglucosaminyltransferase I activity die at mid-gestation, revealing an essential role for complex or hybrid *N*-linked carbohydrates. *PNAS USA,* **91,** 728-732.

Jackson, M. R., Cohen-Doyle, M. F., Peterson, P. A. and Williams, D. B. [1994] Regulation of MHC class I transport by the molecular chaperone calnexin (p88, IP90). *Science,* **263,** 384-387.

Jackson, M. R., Nilsson, T. and Peterson, P. A. [1990] Identification of a consensus motif for retention of transmembrane proteins in the endoplasmic reticulum. *EMBO Journal,* **9,** 3153-3162.

Jan, L. J. and Jan, Y. N. [1982] Antibodies to horseradish peroxidase as specific neural markers in *Drosophila* and in grasshopper embryos. *PNAS USA,* **79,** 2700-2704.

Jannatipour, M. and Rokeach, L. A. [1995] The *Schizosaccharomyces pombe* homologue of the chaperone calnexin is essential for viability. *J. Biol. Chem.,* **270,** 4845-4853.

Jarvis, D. L. and Summers, M. D. [1989] Glycosylation and secretion of human tissue plasminogen activator in recombinant baculovirus-infected insect cells. *Mol. cell. Biol.,* **9,** 214-223.

Kalz-Fuller, B., Bieberich, E. and Bause, E. [1995] Cloning and expression of glucosidase I from human hippocampus. *Eur. J. Biochem.,* **231,** 344-351.

Kang, M. S., Bowlin, T. L., Vijay, I. K. and Sunkara, S. P. [1993] Accumulation of pentamannose oligosaccharides in human mononuclear leukocytes by action of swainsonine, an inhibitor of glycoprotein processing. *Carbohydr. Res.,* **248,** 327-37.

Karess, R. E. and Rubin, G. M. [1984] Analysis of P transposable element functions in *Drosophila*. *Cell,* **38,** 135-146.

Katz, F., Moats, W. and Jan, Y. N. [1988] A carbohydrate epitope expressed uniquely on the cell surface of *Drosophila* neurons in the mutant *nac* (neurally altered carbohydrate). *EMBO J.,* **7,** 3471-3477.

Keller, E. B. and Noon, W. A. [1984] Intron splicing: a conserved internal signal in introns of animal pre-mRNAs. *PNAS USA,* **81,** 7417-20.

Keller, E. B. and Noon, W. A. [1985] Intron splicing: a conserved signal of *Drosophila* pre-mRNAs. *Nuc. Acid. Res.,* **13,** 4971-81.

Kemphues, K. J., Raff, E. C. and Kaufman, T. C. [1983] Genetic analysis of *B2t*, the structural gene for a testis-specific β-Tubulin subunit in *Drosophila melanogaster. Genetics,* **105,** 345-56.

Kerscher, S., Albert, S., Wucherpfennig, D., Heisenberg, M. and Schneuwly, S. [1995] Molecular and genetic analysis of the *Drosophila mas-1* (mannosidase-1) gene which encodes a glycoprotein processing α1,2-mannosidase. *Developmental Biology,* **168,** 613-626.

Kimura, M. [1983] The neutral theory of molecular evolution. Cambridge University Press.

Kjellen, L. and Lindahl, U. [1991] Proteoglycans: structures and interactions. *Ann. Rev. Biochem.,* **60,** 443-475.

Klemenz, R., Weber, U. and Gehring, W. J. [1987] The *white* gene as a marker in a new P-element vector for gene transfer in *Drosophila. Nuc. Acid. Res.,* **15,** 3947-59.

Kornfeld, R. and Kornfeld, S. [1985] Assembly of asparagine-linked oligosaccharides. *Ann. Rev. Biochem.,* **54,** 631 - 664.

Kornfeld, S. [1986] Trafficking of lysosomal enzymes in normal and disease states. *J. Clin. Invest.,* **77,** 1-6.

Kornfeld, S. [1987] Trafficking of lysosomal enzymes. *FASEB J.,* **1,** 462-468.

Koscielak, J. [1995] Diseases of aberant glycosylation. *Acta Biochim. Pol.,* **42,** 1-10.

Kozak, M. [1984] Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nuc. Acid Res*, **12**, 857-72.

Krainer, A. R., Maniatis, T., Ruskin, B. and Green, M. R. [1984] Normal and mutant human β-Globin pre-mRNAs are faithfully and efficiently spliced *in vitro*. *Cell,* **36,** 993-1005.

Kumar, R., Yang, J., Larsen, R. D. and Stanley, P. [1990] Cloning and expression of *N*-acetylglucosaminyltransferase I, the medial Golgi transferase that initiates complex *N*-linked carbohydrate formation. *PNAS USA,* **87,** 9948-52.

Kuroda, K., Hauser, C., Rott, R., Klenk, H.-D. and Doerfler, W. [1986] Expression of the influenza virus haemagglutinin in insect cells by a baculovirus vector. *EMBO J.,* **5,** 1359-1365.

Kurosaka, A., Yano, A., Itoh, N., Kuroda, Y., Nakagawa, T. and Kawasaki, T. [1991] The structure of a neural specific carbohydrate epitope of horseradish peroxidase recognized by anti horseradish peroxidase antiserum. *J. Biol. Chem.,* **266,** 4168-72.

Lal, A., Schutzbach, J. S., Forsee, W. T., Neame, P. J. and Moremen, K. W. [1994] Isolation and expression of murine and rabbit cDNAs encoding an α1,2-mannosidase involved in the processing of asparagine-linked oligosaccharides. *J. Biol. Chem.,* **269,** 9872-9881.

Lasky, L. A. [1992] Selectins: interpreters of cell-specific carbohydrate information during inflammation. *Science,* **258,** 964-969.

Lee, S. J. and Nathans, D. [1988] Proliferin secreted by cultured cells binds to mannose-6-phosphate receptors. *J. Bio. Chem.,* **263,** 3521-3527.

Lefevre, G. J. [1976] in The Genetics and Biology of *Drosophila* M. Ashburner and E. Novitski New York, Academic Press.

Levis, R., Hazelrigg, T. and Rubin, G. M. [1985] Effects of genomic position on the expression of transduced copies of the *white* gene of *Drosophila*. *Science*, **229,** 558-561.

Lis, H. and Sharon, N. [1993] Protein Glycosylation - structural and functional aspects. *Eur. J. Biochem.*, **218,** 1-27.

Lis, J. T., Simon, J. A. and Sutton, C. A. [1983] New heat shock puffs and β-galactosidase activity resulting from transformation of *Drosophila* with an hsp70-*lacZ* hybrid gene. *Cell,* **35,** 403-410.

Lozovskaya, E. R., Petrov, D. A. and Hartl, D. L. [1993] A combined molecular and cytogenetic approach to genome evolution in *Drosophila* using large-fragment DNA cloning *Chromosoma,* **102,** 253-266.

Lubas, W. A. and Spiro, R. G. [1987] Golgi endo α-D-mannosidase from rat liver, a novel *N*-linked carbohydrate unit processing enzyme. *J. Biol. Chem,* **262,** 3775-3781.

Lubas, W. A. and Spiro, R. G. [1988] Evaluation of the role of rat liver Golgi endo-α-D-mannosidase in processing *N*-linked oligosaccharides. *J. Biol. Chem.,* **263,** 3990-8.

Lucklow, V. A. and Summers, M. D. [1988] Trends in the development of baculovirus expression vectors. *Bio/Technology,* **6,** 47-55.

Maniatis, T. and Reed, R. [1987] The role of small nuclear ribonucleoprotein particles in pre-mRNA splicing. *Nature,* **325,** 673-678.

Markwell, M. A. K., Svennerholm, L. and Paulson, J. C. [1981] Specific gangliosides function as host cell receptors for Sendai virus. *PNAS USA,* **78,** 5406-5410.

Marshall, R. D. [1972] Glycoproteins. *Ann. Rev. Biochem.,* **41,** 673-702.

Marth, J. D. [1994] Will the transgenic mouse serve as a Rosetta Stone to glycoconjugate function? *Glycoconjugate Journal,* **11,** 3-8.

Mellor, J. [1992] Molecular analysis of genes thought to be involved in the early stages of *N*-linked glycosylation in *Drosophila melanogaster*. D. Phil, Oxford.

Metzler, M., Gertz, A., Sarkar, M., Schachter, H., Schrader, J. and Marth, J. D. [1994] Complex asparagine-linked oligosaccharides are required for morphogenic events during post-implantational development. *EMBO J,* **13,** 2056-2065.

Michael, J. M. and Kornfeld, S. [1980] Partial purification and characterization of the glucosidases involved in the processing of asparagine-linked oligosaccharides. *Arch. Biochem. Biophys.,* **199,** 249-258.

Miller, D. J., Macek, M. B. and Shur, B. D. [1992] Complementarity between sperm surface β1,4-galactosyltransferase and egg coat ZP3 mediates sperm-egg binding. *Nature,* **357,** 589-593.

Montreuil, J. [1982] in Comprehensive biochemistry. M. Florkin and E. H. Stotz Elsivier.

Moore, S. E. and Spiro, R. G. [1990] Demonstration that Golgi endo-α-D-mannosidase provides a glucosidase independent pathway for the formation of complex *N*-linked oligosaccharides of glycoproteins. *J. Biol. Chem.,* **265,** 13104-13112.

Moremen, K. and Robbins, P. [1991] Isolation, characterization, and expression of cDNAs encoding murine α-mannosidase II, a Golgi enzyme that controls conversion of high mannose to complex N-glycans. *J. Cell Biol.,* **115,** 1521 - 1534.

Moremen, K. W. [1989] Isolation of a rat liver Golgi mannosidase II clone by mixed oligonucleotide-primed amplification of cDNA. *Proc. Natl. Acad. Sci. USA,* **86,** 5276 - 5280.

Moremen, K. W., Trimble, R. B. and Herscovics, A. [1994] Glycosidases of the asparagine-linked oligosaccharide processing pathway. *Glycobiology,* **4,** 113-125.

Mount, S. M., Burks, C., Hertz, G., Stormo, G. D., White, O. and Fields, C. [1992] Splicing signals in *Drosophila*: Intron size, information content and consensus sequences. *Nucl. Acids. Res.,* **20,** 4255-4262.

Nagao, E., Takahashi, N. and Chino, H. [1987] Asparagine-linked oligosaccharides of locust liphorin. *Insect Biochem.,* **17,** 531-538.

Nishikawa, A., Ihara, Y., Hatakeyama, M., Kangawa, K. and Taniguchi, N. [1992] Purification, cDNA cloning and expression of UDP-*N*-acetylglucosamine:β-D-mannoside β-1,4-*N*-acetylglucosaminyltransferase III from rat kidney. *J. Biol. Chem.,* **267,** 18199-204.

O'Hare, K. and Rubin, G. M. [1983] Structures of P transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. *Cell,* **34,** 25-35.

Olden, K., Parent, J. B. and L, W. S. [1982] Carbohydrate moieties of glycoproteins. A re-evaluation of their function. *Biochim. Biophys. Acta,* **650,** 209-232.

Ou, W.-J., Cameron, P. H., Thomas, D. Y. and Bergeron, J. J. M. [1993] Association of folding intermediates of glycoproteins with calnexin during protein maturation. *Nature,* **364,** 771-776.

Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S. and Sharp, P. A. [1986] Splicing of mRNA precusors. *Ann. Rev. Biochem.,* **55,** 1119-1150.

Padgett, R. A., Mount, S. M., Steitz, J. A. and Sharp, P. A. [1983] Splicing of mRNA precusors is inhibited by antisera to small nuclear ribonucleoproteins. *Cell,* **35,** 101-107.

Parekh, R. B., Dwek, R. A., Edge, C. J. and Rademacher, T. W. [1989a] *N*-glycosylation and the production of recombinant glycoproteins. *TIBTECH,* **7,** 117-122.

Parekh, R. B., Dwek, R. A., Rudd, P. M., Thomas, J. R., Rademacher, T. W., Warren, T., Wun, T. C., Hebert, B., Reitz, B. and Palmier, M. [1989b] *N*-glycosylation and *in vitro* enzymatic activity of human recombinant tissue pasminogen activator expressed in Chinese hamster ovary cells and a murine cell line. *Biochemistry,* **28,** 7650.

Parekh, R. B., Dwek, R. A., Sutton, B. J., Fernandes, D. L., Leung, A., Stanworth, D., Rademacher, T. W., Mizuochi, T., Taniguchi, T., Matsuta, M., Takeuchi, F., Nagano, Y., Miyamoto, T. and Kobata, A. [1985] Association of rheumatoid arthritis and primary osteoarthritis with changes in the glycosylation pattern of total serum IgG. *Nature,* **316,** 452-57.

Parekh, R. B., Isenberg, D. A., Ansell, B. M., Roitt, I. M., Dwek, R. A. and Rademacher, T. W. [1988] Galactosylation of IgG associated oligosaccharides: reduction in patients with adult and juvenile onset rheumatoid arthritis and relation to disease activity. *Lancet,* **1,** 966-969.

Parekh, R. B., Tse, A. G. D., Dwek, R. A., Williams, A. F. and Rademacher, T. W. [1987] Tissue specific *N*-glycosylation, site specific oligosaccharide patterns and lentil lectin recognition of rat Thy-1. *EMBO J.,* **6,** 1233-1244.

Parker, G. F., Williams, P. J., Butters, T. D. and B., R. D. [1991] Detection of the lipid-linked precusor oligosaccharide of *N*-linked protein glycosylation in *Drosophila melanogaster. FEBS Lett,* **290,** 58-60.

Parlati, F., Dominguez, M., Bergeron, J. J. M. and Thomas, D. Y. [1995] *Saccharomyces cerevisiae CNE1* encodes an endoplasmic reticulum (ER) membrane protein with sequence similarity to calnexin and calreticulin and functions as a constituent of the ER quality control apparatus. *J. Biol. Chem.,* **270,** 244-253.

Parslow, T. G., Jones, S. D. and Bond, B. [1987] The immunoglobulin octanucleotide independent activity and selective interaction with enhancers. *Science,* **235,** 1498-1501.

Patel, R. and Jacobs-Lorena, M. [1992b] Generation of *minute* phenotypes by a transformed antisense ribosomal protein gene. *Dev. Genet.,* **13,** 256-63.

Paulson, J. C. [1989] Glycoproteins: what are the sugar chains for? *TIBS,* **14,** 272 - 276.

Paulson, J. C. and Colley, K. J. [1989] Glycosyltransferases. Structure, localization and control of cell type specific glycosylation. *J. Biol. Chem,* **264,** 17615-17618.

Pearson, W. R. and Lipman, D. J. [1988] Improved tools for biological comparison. *PNAS USA,* **85,** 2444-2448.

Pelham, H. R. [1991] Recycling of proteins between the endoplasmic reticulum and the Golgi complex. *Curr. Opin. Cell. Biol.,* **3,** 585-591.

Petersen, N. S. [1990] The effects of heat and chemical stress on development. *Adv. Genet.,* **28,** 275-296.

Pierce, J. C., Sauer, B. and Sternberg, N. [1992] A positive selection vector for cloning high molecular weight DNA by the bacteriophage P1 system: Improved cloning efficacy. *PNAS USA,* **89,** 2056 - 2060.

Pirrotta, V. [1988] in Vectors, a survey of molecular cloning vecors and their uses. R. L. Rodrigues and D. T. Denhardt Boston, Butterworth.

Pirrotta, V., Steller, H. and Bozzetti, M. P. [1985] Multiple upstream regulatory elements control the expression of the *Drosophila white* gene. *EMBO J.,* **4,** 3501-3508.

Proudfoot, N. J. and Brownlee, G. G. [1976] 3' Non-coding region sequences in eukaryotic messenger RNA. *Nature,* **263,** 211-214.

Pugh, B. F. and Tjian, R. [1991] Transcription from a TATA-less promoter requires a multisubunit TFIID complex. *Genes Dev.*, **5,** 1935-1945.

Qian, S., Hongo, S. and Jacobs-Lorena, M. [1988] Antisense ribosomal protein gene expression specifically disrupts oogenesis in *Drosophila melanogaster*. *PNAS USA,* **85,** 9601-5.

Rademacher, T. W. [1992] Therapeutic challenges: does Glycobiology have a role? *TIBTech,* **10,** 227-230.

Rademacher, T. W., Parekh, R. B. and Dwek, R. A. [1988a] Glycobiology *Ann. Rev. Biochem,* **57,** 785-838.

Rademacher, T. W., Parekh, R. B., Dwek, R. A., Isenberg, D., Rook, G., Axford, J. S. and Roitt, I. [1988b] The role of IgG glycoforms in the pathogenesis of rheumatoid arthritis. *Springer Semin. Immunopath,* **10,** 231-249.

Ren, J., Bretthauer, R. K. and Castellino, F. J. [1995] Purification and properties of a Golgi-derived (a1,2)-mannosidase I from baculovirus-infected lepidopteran insect cells (IPLB-SF21AE) with preferential activity toward mannose6-N-acetylucosamine2. *Biochemistry,* **34,** 2489-95.

Riederer, M. A. and Hinnen, A. [1991] Removal of *N*-glycosylation sites of the yeast acid phosphatase severely affects protein folding. *J. Bacteriol.,* **173,** 3539-46.

Rosenberg, U. B., Preiss, A., Seifert, E., Jackle, H. and Knipple, D. C. [1985] Production of phenocopies by *kruppel* antisense RNA injection into *Drosophila* embryos. *Nature,* **313,** 703-6.

Roth, J., Kempf, A., Reuter, G., Schauer, R. and Gehring, W. J. [1992] Occurance of sialic acids in *Drosophila melanogaster*. *Science,* **256,** 673-5.

Rubin, G. M. and Spradling, A. C. [1982] Genetic transformation of *Drosophila* with transposable element vectors. *Science,* **218,** 348-353.

Rubin, G. M. and Spradling, A. C. [1983] Vectors for P element-mediated gene transfer in *Drosophila*. *Nuc. Acid. Res.*, **11,** 6341-6351.

Ruoslahti, E. [1988] Structure and biology of proteoglycans. *Ann. Rev. Cell. Biol.*, **4,** 229-255.

Ruoslahti, E. [1989] Proteoglycans in cell regulation. *J. Biol. Chem,* **264,** 13369-13372.

Ruskin, B., Krainer, A. R., Maniatis, T. and Green, M. R. [1984] Excision of an intact intron as a novel lariat structure during pre-mRNA splicing *in vitro*. *Cell,* **38,** 317-331.

Ryan, R. O., Anderson, D. R., Grimes, W. J. and Law, J. H. [1985] Arylphorin from *Manduca sexta*: carbohydrate structure and immunological studies. *Arch. Biochem. Biophys,* **243,** 115-124.

Sambrook, J., Fritsch, E. F. and Maniatis, T. [1989] Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory Press.

Sanger, F. S., Nicklen, S. and Coulson, A. R. [1977] DNA sequencing with chain-terminating inhibitors. *PNAS USA,* **74,** 5463-5467.

Sarker, M., Hull, E., Nishikawa, Y., Simpson, R. J., Moritz, R. L., Dunn, R. and Schachter, H. [1991] Molecular cloning and expression of cDNA encoding the enzyme that controls conversion of high-mannose to hybrid and complex *N*-glycans: UDP-*N*-acetylglucosamine:α-3-D-mannoside β-1,2-*N*-acetylglucosaminyltransferase I. *PNAS USA,* **88,** 234-8.

Saunier, B., Kilker, R. D., Tkacz, J. S., Quaroni, A. and Hercovics, A. [1982] In hibition of *N*-linked oligosaccharide formation by 1-deoxynojirimycin, an inhibitor of processing glucosidases. *J. Biol. Chem.*, **257,** 14155-14161.

Schatzle, J., Bush, J. and Cardelli, J. [1992] Molecular cloning and characterization of the structural gene coding for the developmentally regulated

lysosomal enzyme, α-mannosidase, in *Dictyostelium discoidium. J. Biol. Chem,* **267,** 4000-4007.

Schuh, R. and Jackle, H. [1989] Probing *Drosophila* gene function by antisense RNA. *Genome,* **31,** 422-25.

Sehgal, A., Patil, N. and Chao, M. [1988] A constituative promoter directs expression of the nerve growth factor receptor gene. *Mol. Cell. Biol.,* **8,** 3160-3167.

Senapathy, P., Shapiro, M. B. and Harris, N. L. [1990] Splice junctions, branch point sites, and exons: Sequence statistics, identification, and applications to genome project. *Methods enzymol.,* **183,** 252-278.

Sharp, P. A. [1987] Splicing of mRNA precusors. *Science,* **235,** 766-771.

Sheckman, R. and Novick, P. [1982] in The molecular biology of yeast, *Saccharomyces*. J. N. Strathern, E. W. Jones and J. R. Broach Cold Spring Harbor Laboratory Press.

Sheets, M. D., Ogg, S. C. and Wickens, M. P. [1990] Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and poly adenylation *in vitro*. *Nuc. Acid. Res.,* **18,** 5799-5815.

Shur, B. D. [1994] The beginning of a sweet tale. *Current Biology,* **4,** 996-999.

Smale, S. T. and Baltimore, D. [1989] The "initiator" as a transcription control element. *Cell,* **57,** 103-113.

Smith, M. J. [1992] Nucleotide sequence of a *Drosophila melanogaster* gene encoding a calreticulin homologue. *DNA seq.,* **3,** 247-50.

Smoller, D. A., Petrov, D. and Hartl, D. L. [1991] Characterization of bacteriophage P1 library containing inserts of *Drosophila* DNA of 75-100 kilobase pairs. *Chromosoma,* **100,** 487 - 494.

Snow, P. M., Patel, N. H., Harrelson, A. L. and Goodman, C. S. [1987] Neural-specific carbohydrate moiety shared by many surface glycoproteins in *Drosophila* and grasshopper embryos. *J. Neurosci.*, **7**, 4137-4144.

Sousa, M. C., Ferrero-Garcia, M. A. and Parodi, A. J. [1992] Recognition of the oligosaccharide and protein moieties of glycoproteins by UDP-Glc:Glycoprotein glucosyltransferase. *Biochemistry,* **31,** 97-105.

Southern, E. [1975] Detection of specific sequences amoung DNA separated by gel electrophoresis. *J. Mol. Biol.,* **98,** 503 - 517.

Spradling, A. C. [1986] in *Drosophila*: A practical approach D. B. Roberts Oxford, IRL Press.

Spradling, A. C. and Rubin, G. M. [1982] Transposition of cloned P-elements into *Drosophila* germ line chromosomes. *Science,* **218,** 341-347.

Spradling, A. C. and Rubin, G. M. [1983] The effect of chromosomal position on the expression of the *Drosophila* xanthine dehydrogenase gene. *Cell,* **34,** 47-57.

Stanley, P. [1987] Glycosylation mutants and the functions of mammalian carbohydrates *TIG,* **3,** 77 - 81.

Sternberg, N. [1990] Bacteriophage P1 cloning system for the isolation, amplification and recovery of DNA fragments as large as 100 kilobase pairs. *PNAS USA,* **87,** 103 - 107.

Struhl, G. [1985] Near-reciprocal phenotypes caused by inactivation or indescriminate expression of the *Drosophila* segmentation gene *ftz*. *Nature,* **318,** 677-680.

Swanson, P. M., Carter, C. E., Hager, C., Kim, W. J., Obermeier, S. and Oeltmann, T. N. [1992] Identification and characterization of an alpha-mannosidase from *Trypanosoma cruzi*. *Glycobiology,* **2,** 563-9.

Tabas, I. and Kornfeld, S. [1978] The synthesis of complex oligosaccharides. *J. Biol. Chem.*, **253,** 7779-7786.

Takeuchi, M., Takasaki, S., Miyazaki, H., Kata, T., Hoshi, S., Kochibe, N. and Kobata, A. [1988] Comparative study of the asparagine-linked sugar chains of human erythropoietins purified from urine and the culture medium of recombinant Chinese hamster ovary cells. *J. Bio. Chem.*, **263,** 3657-3663.

Tanner, W. and Lehle, L. [1987] Protein glycosylation in yeast. *Biochim. Biophys. Acta.*, **906,** 81-99.

Tautz, D. and Pfeifle, C. [1989] A non- radioactive *in situ* hybridization method for the localization of specific RNAs in *Drosophila* embryos reveals translational control of the segmentation gene *hunchback*. *Chromosoma,* **98,** 81-85.

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

Tjoelker, L. W., Seyfried, C. E., Eddy, R. L., Byers, M. G., Shows, T. B., Calderon, J., Schreiber, R. B. and Gray, P. W. [1994] Human, mouse, and rat calnexin cDNA cloning: Identification of potential calcium binding motifs and gene localization to human chromosomes. *Biochemistry,* **33,** 3229-3236.

Treves, S., Zorzato, F. and Pozzan, T. [1992] Identification of calreticulin isoforms in the central nervous system. *Biochem. J.,* **287,** 579-81.

Tsuda, E., Goto, M., Murkami, A., Akai, K., Veda, M., Kawanishi, G., Takahashi, N., Sasaki, R., Chiba, H., Ishihara, H., Mori, M., Tejimi, S., Endo, S. and Arata, Y. [1988] Comparative structural study of *N*-linked oligosaccharides of urinary and recombinant erythropoietins. *Biochemistry,* **27,** 5646-5654.

Tulsiani, D. R. P., Hubbard, S. C., Robbins, P. W. and Touster, O. [1982] α-D-mannosidases of rat liver Golgi membranes. *J. Biol. Chem.*, **257,** 3660 - 3668.

Tulsiani, D. R. P. and Touster, O. [1983] Swainsonine, a potent mannosidase inhibitor, elevates rat liver and brain lysosomal a-D-mannosidase, decreases Golgi a-D-mannosidase II, and increases the plasma levels of several acid hydrolases. *Arch. Biochem. Biophys.*, **224,** 594-600.

Varki, A. [1993] Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology,* **3,** 97-130.

Varki, A., Hooshmand, F., Diaz, S., M, V. N. and Hedrick, S. M. [1991] Developmental abnormalities in transgenic mice expressing a sialic acid-specific 9-*O*-acetylesterase. *Cell,* **65,** 65-74.

Wada, I., Rindress, D., Cameron, P. H., Ou, W.-J., Doherty II, J. J., Louvard, D., Bell, A. W., Dignard, D., Thomas, D. Y. and Bergeron, J. J. M. [1991] SSRα and associated calnexin are major calcium binding proteins of the endoplasmic reticulum membrane. *J. Biol. Chem.*, **266,** 19599-19610.

Wagh, P. V. and Bahl, O. P. [1981] Sugar residues on proteins. *CRC Crit. Rev. Biochem.*, **10,** 307-377.

Wahle, E. and Keller, W. [1992] The biochemistry of 3' end cleavage and polyadenylation of messenger RNA precursors. *Ann. Rev. Biochem.*, **61,** 419-440.

Wakimoto, B. T., Kalfayan, L. J. and Spradling, A. C. [1986] Developmentally regulated expression of *Drosophila* chorion genes introduced at diverse chromosomal positions. *J. Mol. Biol.*, **187,** 33-45.

Ware, F. E., Vassilakos, A., Peterson, P. A., Jackson, M. R., Lehrman, M. A. and Williams, D. B. [1995] The molecular chaperone calnexin binds $Glc_1Man_9GlcNAc_2$ oligosaccharide as an initial step in recognizing unfolded glycoproteins. *J. Biol. Chem.*, **270,** 4697-4704.

Wight, T. N., Kinsella, M. G. and Qwarnstrom, E. E. [1992] The role of proteoglycans in cell adhesion, migration and proliferation. *Curr. Opin. Cell Biol.*, **4,** 793-801.

Williams, J. W., Wormald, M. R., Dwek, R. A., Rademaker, T. W., Parker, G. F. and Roberts, D. B. [1991] Characterisation of oligosaccharides from *Drosophila melanogaster* glycoproteins. *Biochim. Biophys. Acta,* **1075,** 146 - 153.

Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fraser, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnson, L., Jones, M., Kershaw, J., Kirsten, J., Laisster, N., Latreille, P., Lightning, J., Lloyd, C., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A., Saunders, D., Showkeen, R., Sims, M., Smaldon, N., Smith, A., Smith, M., Sonnhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterson, R., Watson, A., Weinstock, L., Wilkinson-Sproat, J. and Wohldman, P. [1994] 2.2Mb of contiguous nucleotide sequence from chromosome II of *C. elegans*. *Nature,* **368,** 32-8.

Wojchowski, D. M., Orkin, S. H. and Sytowski, A. J. [1987] Active human erythropoietin expressed in insect cells using a baculovirus vector: a role for *N*-linked oligosaccharide. *Biochim. Biophys. Acta.,* **910,** 224-232.

Yanisch-Perron, C., Vieira, J. and Messing, J. [1985] Improved M13 phage cloning vectors and host strains. Nucleotide sequences for the M13mp18 and pUC19 vectors. *Gene,* **33,** 103-119.

Yoshihisha, T. and Anraku, Y. [1989] Nucleic acid sequence of *AMSI*, the structural gene of vacuolar α-mannosidase of *Saccharomyces cerevisiae. Biochem. Biophys. Res. Commun,* **163,** 908-915.

Yudkin, B. [1996] A genetical investigation of *N*-linked glycosylation in *Drosophila melanogaster*. D. Phil, Oxford.

Zhang, Q., Tector, M. and Salter, R. D. [1995] Calnexin recognises carbohydrate and protein determinants of class I major histocompatibility complex molecules. *J. Biol. Chem,* **270,** 3944-2948.

Zhou, Q., Lieberman, P. M., Boyer, Y., G and Berk, A. J. [1992] Holo-TFIID supports transcriptional activation by diverse activators and from a TATA-less promoter. *Genes Dev.*, **6,** 1964-1974.

Zucker, C. S., Mismer, D., Hardy, R. and Rubin, G. M. [1988] Ectopic expression of a minor *Drosophila* opsin in the major photoreceptor cell class: Distinguishing the role of a primary receptor and cellular context. *Cell,* **53,** 475-482.