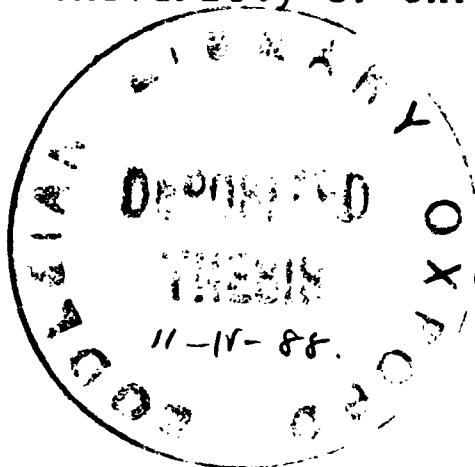


**MOLECULAR STUDIES OF THE HUMAN X AND Y
CHROMOSOMES**

Neil J. Fraser

**A Thesis Submitted for the Degree of
Doctor of Philosophy
in the University of Oxford**



**Genetics Laboratory
Department of Biochemistry
Oxford**

December 1987

ABSTRACT

MOLECULAR STUDIES OF THE HUMAN X AND Y CHROMOSOMES

Neil J. Fraser, St. Catherine's College, Oxford.

Submitted for the degree of Doctor of Philosophy,
Michaelmas 1987.

The isolation and characterisation of sequences from the X and Y chromosomes will give some insight into the evolutionary relationship between these chromosomes, and may be of use in the study of X-linked disorders.

The availability of cDNA and genomic sequences for the human *STS* locus (associated with the disorder, X-linked ichthyosis) has allowed a preliminary investigation of this locus in man and other species. The localisation of these sequences to Xp22.3, provides confirmation of the sub-regional assignment of the structural gene for *STS*. *STS* homologous sequences have been identified on the long arm of the Y chromosome. These sequences also appear present on the X and Y chromosomes of the chimpanzee. In other higher primates, they appear to be X-, but not Y-, linked, suggesting that the situation in man and chimpanzee is the result of a rearrangement between the X and Y chromosomes during the past 15 million years.

Another region of X-Y homology has been analysed. The locus *DXYS27* maps to Yp and Xq21. Restriction enzyme analysis and direct sequence comparison has shown the two loci to be $\approx 99\%$ homologous. Phylogenetic studies suggest that the locus is X-, but not Y-, linked in the chimpanzee, suggesting an evolutionarily recent transposition of material from the X to the Y chromosome. The mutations resulting in the X-Y differences appear to have occurred on both the X and Y chromosomes. It has been possible to demonstrate that the Y-specific locus is transferred to the X chromosome in many, but not all, aberrant X-Y interchanges resulting in XX maleness.

A sequence has been isolated that detects a hypervariable locus at Xp11.3-Xcen (*DXS255*). The hypervariability appears to be due to the presence of a tandemly repeated sequence of variable length. Attempts to clone this repeat have been unsuccessful, as it appears to be unstable in the vector/host systems employed. This sequence will be of value in linkage studies of disease loci known to be present in this region. Hypervariability at this locus has not been identified in other species, suggesting that the repeat sequence is an evolutionarily recent acquisition by the X chromosome.

Taken together, the results obtained suggest that the simple model predicting an ancient origin for the bulk of the Y chromosome will have to be reassessed.

ACKNOWLEDGEMENTS

This research was carried out while I was in receipt of an MRC grant. Most of the work was carried out in the Genetics Laboratory, Oxford, and I would like to thank Professor J.H. Edwards for full use of the facilities. Some of the work was carried out during a three month stay at the Murdoch Institute for Research into Birth Defects in Melbourne, and I thank Professor D. Danks and the other members of this department for making my stay there so enjoyable.

Many genomic DNA samples and clones have been generously provided by others, and I apologise here for not thanking everyone by name; however, I have tried to credit them all in the text.

Almost everyone in the Genetics Laboratory has helped me in some way or another and I apologise for not mentioning them all by name. Special thanks should go to the following:-

Yvonne Boyd for her part in the initial screening of the MOG cosmid bank and for her endless offers of help and advice.

Thomas Meitinger for spending two weeks on a 'Grand Tour' of Europe collecting blood samples for the retinitis pigmentosa linkage studies, and for many lively discussions.

Mark Ross and Mark Jobling for keeping me going when times looked bad.

Elaine Levy for developing my seemingly never ending rolls of film !

David Coates and Kevin O'Dell (when he wasn't examining his flies) for sharing their laboratory with me.

Thanks must go to my wife, Julie, for keeping me supplied with food, drink and love when I needed them most.

Last, but not least, I thank Ian Craig for being more than just a supervisor.

ABBREVIATIONS

STE	salt tris EDTA buffer
TE	tris EDTA buffer
PBS	phosphate buffered saline
TBE	tris borate EDTA
SSC	saline sodium citrate
SSPE	saline sodium phosphate EDTA buffer
SDS	sodium dodecyl sulphate
w/v	weight per volume
v/v	volume per volume
BSA	bovine serum albumin
PEG	polyethylene glycol
TEMED	NNN'N'-tetramethylethylenediamine
IPTG	isopropyl- β -D-thio-galactopyranoside
Xgal	5-bromo-4-chloro-3-indolyl- β -galactoside
DTT	dithiothreitol
DMSO	dimethylsulphoxide
cpm	counts per minute
rpm	revolutions per minute
bp	base pairs
kb	kilobase (one thousand base pairs or nucleotides)
dH ₂ O	distilled water
AMP ^R	ampicillin-resistant
TET ^R , ^S	tetracycline-resistant, sensitive
RFLP	restriction fragment length polymorphism
HVR	hypervariable repeat
dsDNA	double-stranded DNA
ssDNA	single-stranded DNA
♂	male
♀	female
Myr	million years

CONTENTS

CHAPTER 1 : INTRODUCTION

1.1 THE HUMAN X CHROMOSOME

1.1.1	Function	1
1.1.2	X-inactivation	2
1.1.3	X-linked loci	5
1.1.4	The uses of RFLPs in the study of the human X chromosome	7

1.2 THE RELATIONSHIP BETWEEN THE SEX CHROMOSOMES IN MAN

1.2.1	Evolution	9
1.2.2	Regions of X-Y homology	10
1.2.3	Abnormal interchanges between the X and Y chromosomes	15

1.3 OUTLINE OF RESEARCH 17

CHAPTER 2 : GENERAL MATERIALS AND METHODS

2.1 BUFFERS AND MEDIA 19

2.2 BACTERIAL STRAINS 20

2.3 CELL CULTURE 21

2.4 MAMMALIAN CELL DNA PREPARATION

2.4.1	Preparation of DNA from cell pellets and tissue	22
2.4.2	Preparation of DNA from blood	23

2.5 ISOLATION OF PLASMID AND COSMID DNA

2.5.1	Alkaline lysis mini prep. method	24
2.5.2	Large scale preparation of plasmid & cosmid DNA	25

2.6 TRANSFER OF BACTERIAL COLONIES TO NITROCELLULOSE AND HYBOND-N 27

2.7 ANALYSIS AND FRACTIONATION OF DNA USING AGAROSE GELS

2.7.1	Agarose gel electrophoresis	28
2.7.2	Southern blotting of agarose gels	29
2.7.3	Recovery of DNA from agarose gels	31

2.8 LABELLING OF DNA WITH ^{32}P

2.8.1	Nick translation of DNA	32
2.8.2	End labelling of λ / <i>Hind</i> III marker fragments with ^{32}P	32

2.9 REMOVAL OF UNINCORPORATED NUCLEOTIDES BY SPIN DIALYSIS 33

2.10 HYBRIDISATION TO DNA ON FILTERS 34

2.11	SUBCLONING OF RESTRICTION FRAGMENTS INTO PLASMID VECTORS	
2.11.1	Preparation of vector	36
2.11.2	Ligation of DNA fragments to vector DNA	37
2.12	PREPARATION OF COMPETENT CELLS AND TRANSFORMATION OF <i>E. coli</i> HB101/JM83/JM101	
2.12.1	Preparation of competent cells (CaCl ₂ technique)	37
2.12.2	Transformation of <i>E. coli</i> HB101/JM83 and selection for recombinants	38
2.12.3	Transformation of <i>E. coli</i> JM101 and selection for recombinant phage	39
2.13	PREPARATION OF SINGLE-STRANDED DNA TEMPLATES FROM M13mp18 CLONES	40
2.14	GENERAL DNA METHODS	
2.14.1	DNA storage	41
2.14.2	Estimation of DNA concentration	41
2.14.3	Miscellaneous	41

CHAPTER 3 : ISOLATION OF SEQUENCES SPECIFIC TO THE HUMAN X CHROMOSOME

3.1	INTRODUCTION	43
3.2	THE SCREENING OF A COSMID BANK FOR THE ISOLATION OF X-SPECIFIC SEQUENCES	
3.2.1	Materials and methods	47
3.2.2	Identification of cosmids containing human sequences	48
3.2.3	Isolation of cosmid DNA and identification of single-copy restriction fragments	51
3.2.4	Sub-regional localisation of single-copy sequences	54
3.3	DISCUSSION OF RESULTS	56

CHAPTER 4 : AN APPRAISAL OF THE TECHNIQUE OF 'DELETION ENRICHMENT' FOR THE CLONING OF SEQUENCES MAPPING TO DISTAL Xp

4.1	INTRODUCTION	61
4.2	MATERIALS AND METHODS	
4.2.1	General methods	63
4.2.2	Deletion enrichment cloning	64
4.2.3	Single-stranded hybridisation probe generation	65
4.3	AN EVALUATION OF THE TECHNIQUE BY ATTEMPTING TO CLONE Y-SPECIFIC DNA	67
4.4	GENERAL DISCUSSION OF RESULTS	70

CHAPTER 5 : AN ANALYSIS OF THE HUMAN STEROID SULPHATASE LOCUS USING PUTATIVE cDNA AND GENOMIC CLONES

5.1	INTRODUCTION	73
5.2	MATERIALS AND METHODS	77
5.3	THE SUB-REGIONAL ASSIGNMENT OF cDNA CLONES FOR STS	
5.3.1	An analysis of the clone p422	77
5.3.2	An analysis of the clone P2A7	79
5.3.3	An analysis of the genomic clones B2RI12 and B12RI5	82
5.4	SEQUENCE CONSERVATION OF <i>STS</i> HOMOLOGOUS SEQUENCES DURING EVOLUTION	
5.4.1	A search for <i>STS</i> homologous sequences in the mouse and rabbit	86
5.4.2	Conservation of <i>STS</i> homologous sequences during primate evolution	87
5.5	GENERAL DISCUSSION OF RESULTS	90

CHAPTER 6 : THE ANALYSIS OF AN X-Y HOMOLOGOUS SEQUENCE PRESENT ON Y_p AND X_q

6.1	INTRODUCTION	95
6.2	MATERIALS AND METHODS	
6.2.1	General	97
6.2.2	DNA sequencing	98
6.3	AN ANALYSIS OF THE DEGREE OF HOMOLOGY BETWEEN THE X AND Y COPIES OF <i>DXYS27</i>	
6.3.1	X-Y differences detected by restriction enzymes	101
6.3.2	The isolation of cosmids containing the X homologue of <i>DXYS27</i>	102
6.3.3	Sequence analysis of X59 γ and Y59 γ	105
6.4	THE ANALYSIS OF XX MALES AND AN XY FEMALE FOR THE PRESENCE OF THE Y HOMOLOGUE OF <i>DXYS27</i>	107
6.5	A SEARCH FOR SEQUENCES HOMOLOGOUS TO <i>DXYS27</i> IN PRIMATES	110
6.6	GENERAL DISCUSSION AND CONCLUSIONS	112

CHAPTER 7 : THE ANALYSIS OF A SEQUENCE (M27 β) DETECTING A HYPERVARIABLE LOCUS (*DXS255*) ON THE X CHROMOSOME

7.1	INTRODUCTION	116
7.2	MATERIALS AND METHODS	123
7.3	OBSERVATION OF MULTI-ALLELIC VARIATION DETECTED BY M27 β	123

7.4	REFINEMENT OF THE SUB-REGIONAL LOCALISATION OF THE LOCUS DETECTED BY M27 β	125
7.5	SUBCLONING OF M27 β INTO THE PLASMID VECTOR pUC9	125
7.6	DETERMINATION OF THE CAUSE OF THE MULTI-ALLELIC VARIATION	126
7.7	ANALYSIS OF THE DEGREE OF HETEROZYGOSITY AT THE LOCUS <i>DXS255</i>	127
7.8	INVESTIGATION OF THE INHERITANCE OF RFLP	129
7.9	THE POTENTIAL USE OF THE PROBE M27 β IN LINKAGE ANALYSIS	
7.9.1	Introduction	130
7.9.2	Investigation of the linkage between <i>DXS255</i> and the locus for retinitis pigmentosa	130
7.10	COMPARISON OF GENOMIC SEQUENCES IN THE REGION OF M27 β WITH THOSE PRESENT IN THE COSMID M27	134
7.11	IDENTIFICATION OF A 330bp RESTRICTION FRAGMENT OF HX α S2 FROM WHICH THE HYPERVARIABLE REPEAT (HVR) HAS BEEN DELETED	136
7.12	CONSERVATION OF SEQUENCES HOMOLOGOUS TO <i>DXS255</i> IN OTHER SPECIES	137
7.13	GENERAL DISCUSSION OF RESULTS	138
<u>CHAPTER 8 : AN ATTEMPT TO CLONE THE SEQUENCES RESPONSIBLE FOR THE OBSERVED HYPERVARIABILITY AT <i>DXS255</i></u>		
8.1	INTRODUCTION	143
8.2	MATERIALS AND METHODS	144
8.3	PREPARATION OF GENOMIC INSERTS	146
8.4	CONSTRUCTION AND SCREENING OF THE PLASMID LIBRARY	150
8.5	GENERAL DISCUSSION OF RESULTS	155
<u>CHAPTER 9 : SUMMARY AND GENERAL DISCUSSION</u>		158
APPENDIX		165
REFERENCES		166

CHAPTER 1

INTRODUCTION

1.1 THE HUMAN X CHROMOSOME

1.1.1 Function

Sexuality in mammals depends on the combination of the two sex chromosomes, X and Y, which differ from one another morphologically. Normally, the female has a homomorphic XX constitution and the male a heteromorphic XY constitution. Male development depends on the presence of a Y chromosome and is independent of the number of X chromosomes (Jacobs & Strong, 1959). Individuals with XO, XX, XXX, and XXXX chromosome constitutions are phenotypically female, and XY, XXY, XXXY, XXXXY and XYY individuals are male (reviewed in Craig & Goodfellow, 1986).

The human X and Y chromosomes are very different in size, genetic content and function. The presence of the Y chromosome in an embryo initiates the development of the primordial gonads into testes (Ohno, 1976). This chromosome is small and has few genetic loci other than those of sex determination and spermatogenesis (Goodfellow *et al.*, 1985). All secondary male sexual characteristics are the consequence of hormones produced by the testis (e.g. see Ohno, 1967). In contrast, the X chromosome has retained many genes unconnected with sex determination. The X-linkage of such genes has been conserved throughout mammalian evolution (Ohno, 1967). The simple and characteristic pattern of X-linked inheritance has meant that the human X chromosome has become the model chromosome for

studying human genetics and genetic diseases. More than 130 separate genetic loci of known function have been assigned to the X chromosome and for about an equal number, X-chromosomal location has been suggested but not proved (McKusick, 1987). Most of these loci have been placed on the X chromosome because of their pattern of inheritance and other characteristics of X-linked traits in families. A typical X-linked recessive disease will show male affliction and maternal inheritance. Some have been assigned to the X chromosome by the same methods used in autosome mapping: such as those based on interspecies somatic cell hybridisation, *in situ* hybridisation, or by the detection of small, microscopically visible deletions. Some methods unique to the X chromosome have corroborated X-linkage: Lyonisation (inactivation of one of the X chromosomes), Ohno's law of the evolutionary conservatism of the X chromosome in mammals (Ohno, 1967) and X-autosome translocations in females affected by X-linked recessive disorders. In addition to these loci, at least 222 random DNA sequences, many with known RFLPs have been regionally mapped on the X chromosome (Craig & Goodfellow, 1986).

1.1.2 X-inactivation

As females have two X chromosomes and males have only one, females will have twice the gene dosage as that of males for X-linked genes. This inequality of genetic content is remedied by genetically inactivating one of the X chromosomes in female cells (Lyon, 1972). In placental mammals, one of the X chromosomes is randomly inactivated early in embryonic development. This inactivation pattern is clonally maintained

and the female becomes a natural mosaic with two populations of cells that have inactivated one or the other X chromosomes. In humans, the X-inactivation pattern remains random in extraembryonic tissue, however in other species (e.g. marsupials) preferential paternal X-inactivation is the rule (reviewed in Martin, 1982).

It is still uncertain whether inactivation involves the alteration of X DNA or changes in X-chromosomal proteins or factors other than X DNA. If modification of the DNA is the means by which X-inactivation and maintenance of the inactive state are accomplished, then a possible mechanism by which this could occur involves changes in the pattern of DNA methylation. DNA methylation has been implicated in the control of transcriptional activity of autosomal genes (Cooper, 1983). Data in support of this idea come from experiments in which genes on the inactive X chromosome have been reactivated by treatment of cells with 5-azacytidine, a cytidine analogue that when incorporated into DNA leads to hypomethylation. Mohandas (1981) reported the reactivation of the hypoxanthine phosphoribosyltransferase (*HPRT*) locus on an inactive human X chromosome in a mouse-human somatic cell hybrid by using this drug. However, the reactivation was limited, and the expression of other genes on the inactive human X chromosome was detected in only a few cases. These and other data (Wolf & Migeon, 1982) were interpreted as indicating that hypomethylation caused by 5-azacytidine incorporation results in reactivation of individual genes on an inactive X chromosome, but not of the whole chromosome. Wolf et al. (1984) have studied the methylation of active, inactive and derepressed *HPRT* alleles

and have proposed that X chromosome dosage compensation may result from tight control of the level of methylation on the active, rather than the inactive, X chromosome. Kaslow & Migeon (1987) have suggested that DNA methylation of CpG clusters may be the final step that 'locks in' inactivity of X-linked genes.

The random pattern of X-inactivation can be disrupted by X-autosome translocations. It is thought that inactivation may spread to the autosomal material of a translocation chromosome. To maintain the gene dosage of loci present on the translocated autosome, in the majority of females with balanced X-autosome translocations, the normal X chromosome is preferentially inactivated (Mattei *et al.*, 1982). Studies in mice carrying X chromosome-autosome translocations have shown that only a single product of the translocation undergoes inactivation (reviewed in Mohandas *et al.*, 1987). This suggests that there is a single inactivation centre on the murine X chromosome from which inactivation spreads in both directions. A centre controlling inactivation in humans has been hypothesised to be localised to Xq13 (Mattei *et al.*, 1982). The observation of a random inactivation pattern for the cell lines TEL26 and CAT which contain interstitial deletions involving Xq13-Xq21.1 (Tabor *et al.*, 1983; R. Lindenbaum, personal communication) suggests that the location of such a centre must be proximal to this region. Three loci that are present on distal Xp and have been shown to escape from X-inactivation are *Xg* (Race & Sanger, 1975), *STS* (Shapiro *et al.*, 1979) and *MIC2* (Goodfellow *et al.*, 1984). Recent studies by Mohandas *et al.* (1987) have demonstrated that *STS* and *MIC2* sequences, even when flanked on either side by inactive X DNA, maintain their expression. This

suggests that the mechanism of spreading of inactivation operates in a sequence-specific fashion. Alternatively, in the case described, *STS* and *MIC2* may have undergone inactivation initially, but could not be maintained in this state. *MIC2* is the only known gene to be shared by the two human sex chromosomes (Goodfellow *et al.*, 1983). This escape from inactivation maintains male-female dosage. Although there is no evidence for expressed copies of *STS* and *Xg* on the Y chromosome, *STS* is present and expressed on the mouse Y chromosome (Keitges *et al.*, 1985).

1.1.3 X-linked loci

Until quite recently, the only two linkage groups assigned to the X chromosome were localised at the telomeres, due to a paucity of intervening markers. The *Xg* blood group and steroid sulphatase have been shown to be linked and assigned by deletion mapping to the terminal band of the X chromosome (reviewed in Polani, 1982). The gene *MIC2*, which codes for the expression of the cell surface antigen 12E7, has also been assigned to this region (Goodfellow *et al.*, 1983). Other loci in this linkage group include the *Xk* blood group, chronic granulomatous disease (*GCD*) and ocular albinism (*OA*) (Davies, 1985). These loci do not escape from X-inactivation. The other linkage group, localised in the telomeric region of the long arm includes loci for colour blindness, haemophilia A, glucose 6 phosphate dehydrogenase (*G6PD*) and adrenoleucodystrophy (Davies, 1985). Several gene specific probes have now been isolated from the X chromosome. These include the genes coding for *F8* (factor VIII), *F9* (factor IX), *G6PD*, *HPRT*, *MIC2* and *OTC*

(ornithine transcarbamylase) (Goodfellow *et al.*, 1985). More recently cDNA sequences for *DMD* (Duchenne muscular dystrophy) (reviewed in Monaco & Kunkel, 1987) and *STS* (Ballabio *et al.*, 1987; Yen *et al.*, 1987a) have been isolated. The isolation of these genes has allowed an in depth study of the defects at many of these loci, as well as a precise sub-regional localisation. However, for the majority of X-linked traits both the molecular nature and precise regional localisation remain unknown.

For X-linked genetic diseases that have not been defined at the biochemical level, a knowledge of the precise chromosomal localisation of the defect can be exploited to devise strategies for the cloning of the gene responsible for that defect. In some cases, the localisation of a defect has been made possible by the presence of chromosome deletions in affected individuals. 'Deletion enrichment' cloning procedures have been successfully employed to clone sequences mapping to the deletions present in a patient diagnosed as having *DMD*, chronic granulomatous disease and retinitis pigmentosa (Kunkel *et al.*, 1985; Smith *et al.*, 1986) and in a patient with choroideremia, mental retardation and deafness (Nussbaum *et al.*, 1987). In the first case, these sequences have been useful in the characterisation of the *DMD* locus. There are special circumstances where chromosomal rearrangements of known genes into regions of unknown structure or function can help to identify a region implicated in the gene defect. Worton *et al.* (1984) had shown that an X;21 translocation had juxtaposed the *DMD* region on the X chromosome next to a cluster of rRNA genes on chromosome 21. Ray *et al.* (1985) were able to clone the

breakpoint by screening a library, made from a somatic cell hybrid containing the translocation chromosome, with a human specific probe derived from a rRNA gene. In addition to *DMD*, the manifestation of X-linked defects in females carrying balanced translocations has helped in the sub-regional localisation of several other disorders, including incontinentia pigmenti (Xp11) and Hunter's syndrome (Xq26-q27) (reviewed in Goodfellow *et al.*, 1985). However, there are few cases in which these approaches will be feasible, and in general other strategies will have to be devised.

1.1.4 The uses of RFLPs in the study of the human X chromosome

Changes in DNA sequences that introduce or destroy restriction-enzyme cleavage sites and cause restriction fragment length polymorphisms (RFLPs) can provide the key to the sub-regional localisation of X-linked disorders. Many random DNA sequences have now been isolated from the X chromosome, by a variety of strategies (reviewed in Chapter 2). These sequences can be screened for RFLPs. Although it has been argued by Hofker *et al.* (1986) that the X chromosome shows less variation at restriction sites than the autosomes, of the 222 DNA probes for defined sequences localised on the X chromosome, 74 recognise sequence polymorphisms (Craig & Goodfellow, 1986), and recent compilations indicate that this number is probably much higher (Human Gene Mapping Conference 9, 1987). RFLPs have been used in the analysis of several X chromosome disorders. Retinitis pigmentosa has been shown to be closely linked to L1.28, a probe assigned to Xp11.3 that detects a *TaqI* polymorphism (Bhattacharya *et al.*, 1984). Norries disease has

also been localised in this region by linkage analysis of L1.28 in affected families (Gal *et al.*, 1985). Other examples of the linkage of anonymous sequences to disease loci have been reviewed by Goodfellow *et al.* (1985). In the absence of the cloned gene itself, the establishment of close linkage to a polymorphic marker will increase the frequency and reliability with which, by family studies, genetic disorders can be diagnosed. Markers closely linked to *haemophilia A* have been reported and are already being used for antenatal diagnosis and carrier detection (Davies, 1985). Linkage analysis is most efficient when highly informative polymorphic probes can be used. In the past few years a number of such probes have been isolated by chance (e.g Wyman & White, 1980) or more recently through the use of a screening strategy involving previously characterised highly polymorphic probes (Jeffreys *et al.*, 1985b; Nakamura *et al.*, 1987a), many of these probes have been shown to be associated with minisatellite regions of DNA (reviewed in Chapter 7). Although there have been no previous reports of such regions being present on the X chromosome, the isolation of such sequences would obviously be beneficial for linkage analysis.

Linked RFLPs have been suggested as the starting point to isolating the defective gene itself (Little, 1986). If, by recombination analysis, two RFLPs can be shown to flank the defective gene, it may in some cases be possible to clone the intervening material and so isolate the defective gene. The distance between the flanking markers must be small if these elements are to be helpful starting points from which to isolate the gene. The relationship between recombination and

physical distance along the X chromosome is not linear. For example Goss and Harris (1977), using radiation induced gene segregation, showed that the *HPRT* locus is physically close to the *G6PD* locus, whereas Francke *et al.* (1974) had previously shown that these loci are not closely linked genetically. This was the first evidence of the increase of recombination towards the end of the X chromosome long arm. Recombination is thought to be lower in pericentric regions (Drayna *et al.*, 1984), again leading to a discrepancy in the genetic and physical maps. An idea of the physical distance between flanking RFLPs can be obtained by a combination of pulsed field gel electrophoresis (reviewed in Anand, 1987) and Southern blotting. To isolate the intervening material it is necessary to chromosome 'walk' from one marker to the other (reviewed in Little, 1986). To span large distances the construction of 'jumping' and 'linking' libraries has been suggested (reviewed in Poustka & Lehrach, 1987). Once this region has been spanned the defective gene must be identified, how this is done will depend very much on the gene and the defect in question, but the identification of characteristic motifs may provide the first clue.

1.2 THE RELATIONSHIP BETWEEN THE SEX CHROMOSOMES IN MAN

1.2.1 Evolution

The evolution of the heteromorphic sex chromosomes of mammals is a matter of some debate. The human X and Y chromosomes are thought to derive from an ancient homologous pair of chromosomes differing at a single sex determining locus. It has been suggested that with developing complexity in

the sex-determining mechanisms, it became essential that recombination between the previously identical X and Y chromosomes should be suppressed; this could lead to the gradual loss of genetic information from the genetically isolated Y chromosome and the evolution of the dosage compensation mechanism (Ohno, 1967; Charlesworth, 1978). Alternatively, it has been suggested that the evolution of heteromorphic sex chromosomes is the direct consequence of the evolution of dosage compensation, and not the reverse (Lucchesi, 1978; Jones, 1984). A multitude of evidence points to the presence of a testis determining factor on the Y chromosome. The analysis of individuals with structural abnormalities of the Y chromosome had assigned such a gene (*TDF*) to the proximal region of the Y short arm (Davis, 1981). This localisation was consistent with the karyotype of several XX males who appear to have acquired Yp material (Madan, 1976; Evans *et al.*, 1979) on one X chromosome short arm or an autosome (Dosik *et al.*, 1976)*. The maintenance of this gene specific to the Y chromosome is achieved by suppressing recombination between the sex chromosomes. The existence of such suppression is exemplified by the number of X-linked loci with no counterpart on the Y chromosome.

* However, subsequent analysis of region of the Y transferred to XX males has localised *TDF* to distal Yp (see section 1.2.3).

1.2.2 Regions of X-Y homology

Despite the widely differing morphologies, the human sex chromosomes do pair at male meiosis (Pearson & Bobrow, 1970). The Y short arm pairs with the distal region of the X short arm (e.g. see Moses *et al.*, 1975). The extent of this pairing varies considerably, and can involve 11-56% of the Y

chromosome and 3.3-17.3% of the X chromosome (Chandley *et al.*, 1984). The observation of pairing between the X and Y chromosomes implies that there may be sequence homology within this region. Burgoyne (1982) proposed that recombination between the X and Y is required for the normal resolution of the sex chromosome pairing. If this is true, then the X and Y chromosomes will share sequences by exchange. These sequences will show only partial sex linkage and may behave in a 'pseudoautosomal' fashion, depending on the frequency of recombination. The existence of a pseudoautosomal region has been confirmed by cloning X and Y shared sequences. The locus *MIC2*, that controls the expression of the 12E7 antigen, had been shown to be present on both the X and Y chromosomes (Goodfellow *et al.*, 1983). Genetic mapping experiments have localised *MIC2* within the pairing region (Buckle *et al.*, 1985) and, using cloned sequences derived from *MIC2*, this gene has recently been shown to exchange between the X and Y chromosomes (Goodfellow *et al.*, 1986). Cooke *et al.*, (1985) have isolated a probe, 29C1, which detects an extremely polymorphic DNA locus, *DXYS14*, located close to the telomere of the sex chromosomes. Simmler *et al.* (1985a) have isolated a cosmid sub-clone (113D) and a cDNA clone (601) which detect the loci *DXYS15* and *DXYS17*, localised in the X-Y pairing region. Family studies combined with RFLPs have been used to demonstrate that these sequences exchange between the X and Y chromosomes in males and between the X chromosomes in females. Rouyer *et al.* (1986) have found that the loci *DXYS14*, *DXYS15* and *DXYS17* are linked to sex determination at frequencies which define a gradient of linkage. This study has been extended to include

MIC2 (Goodfellow *et al.*, 1986; Weissenbach *et al.*, 1987). Sequences at the telomere recombine with the male determining segment of the Y chromosome with a frequency of $\approx 50\%$ and *MIC2*, the most proximal marker, recombines with the male determining segment at a frequency of a few per cent. The segregation patterns of these loci indicate that X-Y recombination results from a single obligatory meiotic crossing-over in the pseudoautosomal region. Recombination at male meiosis in the terminal regions of the short arms of the X and Y chromosomes is 10-20 fold higher than between the same regions of the X chromosomes at female meiosis.* If one chiasma at least takes place in each bivalent to ensure proper segregation during the first meiotic division, then this increase in male recombination frequency is a direct consequence of a chiasma having to be formed in a chromosomal segment as limited as the pseudoautosomal region, which has been estimated not to exceed 2 to 3 Mb (Weissenbach *et al.*, 1987). Simmler *et al.* (1985a) have isolated a probe (113F) from the same cosmid as 113D, which contains a 600 bp repeat sequence present in 20 copies per haploid genome. All of the copies are dispersed within the pseudoautosomal regions. These sequences define a locus, *DXYZ2*, which Simmler *et al.* (1985) speculate may have a function in pairing or recombination. It is interesting to note that minisatellite sequences have been found to be associated with the high variability of the pseudoautosomal loci *DXYS14*, *DXYS15*, *DXYS17* and *DXYS20* (Simmler *et al.*, 1987). In line with the function proposed for this type of sequence in recombination, it is tempting to relate the high recombinational activity of the pseudoautosomal region to the

* (Rouyer *et al.*, 1986)

presence of minisatellites in this region. This point is discussed in more detail in the introduction to Chapter 7.

In addition to the region of X-Y homology present in the pairing region, several DNA fragments have been isolated from one sex chromosome which detect sequences with variable degrees of homology on the other chromosome. The first sequence formally assigned to the Y chromosome short arm was isolated by Page *et al.* (1982). This sequence, which detects the locus *DXYS1*, recognised a male specific *TaqI* restriction fragment in addition to fragments shared by males and females. *In situ* hybridisation localised the Y sequence to Yp and the X sequence to Xq13-Xq21 (Page *et al.*, 1984). The X- and Y-linked sequences show extensive homology and restriction mapping suggests the sequences match at better than 99% of the residues. *DXYS1* is also X-linked in the great apes, but these species lack the Y-linked sequences, suggesting that a specific transposition between the long arm of the X and the short arm of the Y chromosome occurred during the evolution of the human lineage. There are many other reports of X-Y homologous sequences showing a similar distribution and degree of homology (97-99%) (e.g. Wolfe *et al.*, 1984b; Geldwerth *et al.*, 1985; Koenig *et al.*, 1985; Affara *et al.*, 1986a; Buckle *et al.*, 1987). The number of clones originating from this region, isolated from a Y chromosome specific library (Bishop *et al.*, 1983) suggests that it may constitute an important part (up to one quarter) of Y chromatin. Vergnaud *et al.* (1986) and Affara *et al.* (1986a) have found that the X-Y homologous sequences on Yp fall into two groups separated by a series of Y-specific sequences. The possibilities are that two independent transpositions have

occurred at the same point in evolution, or that inversions in Yp subsequent to the original transposition event have led to the separation of the blocks of X-Y homologous sequences, the latter hypothesis is favoured. There may even be population variation in the order and organisation of non-coding regions of the Y chromosome.

Cooke *et al.* (1984) have described the isolation of a sequence (2:13) which is present on Yq and Xq. Preliminary sequence data had revealed no base change in 1kb. This study has been updated (Bickmore & Cooke, 1987a), the X-Y localisation refined to Xq26-q27 and Yqcen-q11.1, and the region of homology is now thought to extend over at least 50 kb. The sequence has been shown to be X-linked in higher primates and they postulate that a transposition of material from the X to the Y chromosome has occurred during the last one million years. Although they speculate that this sequence can be categorised along with the X-Y homologous sequences typified by *DXYS1*, its long arm localisation on the Y chromosome (and probable more distal localisation on Xq) and higher degree of homology might suggest that at least two transposition events have occurred during recent evolutionary history.

Another category of X-Y homologous sequences is represented by e.g. *DXS31* (Koenig *et al.*, 1984), *DXS69* (Kunkel *et al.*, 1983) and *GMGXY3* (Affara *et al.*, 1986a). These detect homology between the sex chromosomes at only low stringency. They are generally localised to Xp21-pter and the euchromatic part of Yq. Koenig *et al.* (1985) have shown that for the locus *DXS31*, sequences originating from the Xp22.3-pter region show only 80% homology to sequences present on Yq11. X-Y homology is

also detected in the chimpanzee, but an X-only localisation is found in the macaque. In the lemur their probe showed dosage consistent with an autosomal/pseudoautosomal localisation. This region of X-Y homology is believed to be the result of transposition event that occurred >20 million years ago.

1.2.3 Abnormal interchanges between the X and Y chromosomes

It appears that the high level of recombination between the terminal short arms of the sex chromosomes at male meiosis sometimes results in unequal crossing over which may generate various sex reversal syndromes. An abnormal X-Y interchange has been proposed by Ferguson-Smith (1966) to account for XX maleness. One in 20,000 European males have a 46,XX genotype and a male phenotype similar to Klinefelter's syndrome (47,XXY). A single, but unequal, cross-over in a region proximal to the TDF locus on the Y chromosome would result in its translocation onto the X chromosome during male meiosis. Studies have confirmed this by demonstrating that DNA from XX males contains Y-specific sequences (e.g. Guellaen *et al.*, 1984; De la Chapelle *et al.*, 1984; Koenig *et al.*, 1985; Page *et al.*, 1985; Vergnaud *et al.*, 1986; Muller *et al.*, 1986; Affara *et al.*, 1986; Andersson *et al.*, 1986; Buckle *et al.* 1987). In some of these cases, *in situ* hybridisation has revealed the existence of Y sequences on the distal region of Xp. Southern analysis on DNA prepared from flow sorted X chromosomes of XX males has also indicated the presence of Y sequences on X chromosomes (Affara *et al.*, 1986b). By following the inheritance of pseudoautosomal RFLPs it has been demonstrated that at least in some cases this transfer involves interchange

of the terminal part of both paternal sex chromosomes (Page *et al.*, 1987; Weissenbach *et al.*, 1987; Petit *et al.*, 1987). However, Petit *et al.* (1987) were also able to demonstrate that in the Y(-) XX males analysed, no abnormal X-Y interchange was observed, suggesting that maleness can occur in the absence of any Y-specific DNA. It is possible that maleness in Y(-) XX males is the result of an autosomal mutation.

The amount of Y-specific DNA transferred to Y(+) XX males varies. Vergnaud *et al.* (1986) have constructed a deletion map of the Y chromosome, comprising seven intervals, based on the presence/absence of Y-specific DNA in individuals with sex chromosome anomalies. Affara *et al.* (1986b & 1987) and Müller *et al.* (1986) have carried out similar surveys. The overall conclusions are that a testis determinant is located in distal Yp proximal to the pseudoautosomal region. However, Affara *et al.* (1986b) have found anomalies in the relative position of their probes based on a Y deletion panel (Affara *et al.*, 1986a) and their frequency of transfer to XX males. This they have explained by the unlikely event of several crossover events occurring in a relatively small area, or by the existence in the population of polymorphic Y chromosomes which contain inverted segments of Yp. However, as they also point out, the genesis of an XX male is an unusual event which may be accompanied by complex rearrangements and interstitial deletions resulting in anomalous patterns of transfer.

1.3 OUTLINE OF RESEARCH

This thesis is concerned with the isolation and characterisation of sequences specific to the human X and Y chromosomes, to investigate their evolutionary relationship and consequences in context of expression of independent and shared functions.

In Chapter 3, I describe the isolation and characterisation of sequences from a cosmid library, made from a somatic cell hybrid which has the human X chromosome as its only human component. In Chapter 4, I give an evaluation of the technique of 'deletion enrichment' cloning as a more direct approach of cloning sequences from a specific chromosome region.

The availability of cDNA and genomic sequences for the gene *STS*, shown previously to map to terminal Xp, through a collaborative study with the International Institute of Genetics and Biophysics, Naples, has enabled a preliminary investigation of sequences mapping to this region. This has included the discovery of *STS* homologous sequences on the long arm of the Y chromosome (Chapter 5).

Another region of X-Y homology has been investigated using a probe that I had previously isolated from a cosmid library, made from a somatic cell hybrid which has the human Y chromosome as its only human component (Chapter 6).

One of the sequences isolated from the cosmid library described in Chapter 2, detects a hypervariable locus on the short arm of the X chromosome. This sequence has been analysed in some detail (Chapter 7), and an attempt to clone the

hypervariable repeat associated with this locus is described in Chapter 8.

CHAPTER 2

GENERAL MATERIALS AND METHODS

2.1 BUFFERS AND MEDIA

20xSSC: 3M NaCl, 0.3M Na₃citrate (pH 7.0).

20xSSPE: 3M NaCl, 0.2M NaH₂PO₄.H₂O, 20mM Na₂EDTA.

1xSET: 0.15M NaCl, 50mM Tris.Cl, 5mM Na₂EDTA (pH 7.8).

PBS: 2.6mM KH₂PO₄, 26mM Na₂HPO₄, 145mM NaCl (pH 7.2).

10xTBE: 108g/l Tris base, 55g/l boric acid, 0.95g/l Na₂EDTA.2H₂O (pH 8.3).

1xTE: 10mM Tris.Cl, 1mM Na₂ EDTA (pH 8.0).

GTE: 50mM Glucose, 25mM Tris.Cl, 10mM Na₂EDTA (pH 8.0).

50x Denhardt's: 1% bovine serum albumin, 1% polyvinylpyrrolidone, 1% ficoll.

10x M9 salts: Per 100ml - 6g Na₂HPO₄, 3g KH₂PO₄, 1g NH₄Cl, 0.5g NaCl (after autoclaving store at 4°C).

Restriction Enzyme Buffers

Reaction buffers supplied by BRL were routinely used, otherwise the following buffers were used as specified by Maniatis *et al.* (1982).

HIGH SALT : 50mM Tris.Cl (pH 7.4), 10mM MgSO₄, 100mM NaCl.

MEDIUM SALT: 10mM Tris.Cl (pH 7.4), 10mM MgSO₄, 50mM NaCl.

LOW SALT : 10mM Tris.Cl (pH 7.4), 10mM MgSO₄.

Media

LB Broth: 10g/l Bacto-tryptone, 5g/l Bacto yeast extract, 10g/l NaCl.

2xTY Broth: 16g/l Bacto-tryptone, 10g/l Bacto yeast extract, 5g/l NaCl.

Plates were made with either LB or 2xTY broth to which had been added 1.5% w/v Bacto-agar.

Ampicillin was used at a concentration of 75-100µg/ml and tetracycline at 15µg/ml. X-gal was used at 2ml/l (from a 2% stock in dimethylformamide) and IPTG (where used) at 0.5mM. Plates were stored at 4°C for up to a week before use.

Glucose/minimal medium plates: The following reagents should be autoclaved and allowed to cool before mixing aseptically. Per 100ml - 1.5g agar in 88.5ml dH₂O, 10ml 10x M9 salts, 0.1ml 1M MgSO₄, 0.1ml 0.1M CaCl₂, 0.2ml 0.5M thiamine HCl and 1ml 20% glucose.

H plates: 10g/l Bacto-tryptone, 8g/l NaCl, 12g/l Bacto-agar.

H top agar: Per 100ml - 1g Bacto tryptone, 0.8g NaCl, 0.8g Bacto agar.

2.2 BACTERIAL STRAINS

HB101: F⁻, *hsdS20* (r_B⁻, m_B⁻), *recA13*, *ara-14*, *proA2*, *lacY1*, *galK2*, *rpsL20* (Str) *xyl-5*, *mtl-1*, *supE44*, λ⁻

JM83: *ara*, Δ(*lac proA,B*), *rpsL*, φ80 *lacZ* ΔM15 (r_K⁺, m_K⁺)

DH5α: F⁻, *endA1*, *hsdR17* (r_K⁻, m_K⁺), *supE44*, *thi-1*, λ⁻, *recA1*, *gyrA96*, *relA1*, Δ(*lacZYA-argF*)U169, φ80d *lacZ* ΔM15

JM101: Δ*lac pro*, *supE*, *thi*, (r_K⁺, m_K⁺)/F' *traD36*, *proAB*, *lacI^q*, ZΔM15

SF8: *recB*⁻, *recC*⁻, *thr1*, *leuB6*, *thi1*, *supE44*, *hsdR*, *hsdM*, *lop11*, *gal96*, *Sm^R*

DL282: *leu6*, *ara14*, *his4*, *thr1*, *lacY1*, *mtl1*, *xyl5*, *galK2*, *proA2*, *argE3*, *str31*, *tsx33*, *sup37-amber*, *recB21*, *recC22*, *sbcB15*, *hsdR*⁻, *hsdM*⁺, *thr::Tn10*

2.3 CELL CULTURE

Hybrid cell lines were grown in Minimal Essential Medium with Earle's Salts (*Flow Labs*) supplemented with glutamine, non-essential amino acids and sodium bicarbonate (all *Imperial Laboratories*); the final pH of the medium was adjusted to 7.4 with NaOH. Foetal calf serum (*Gibco*) was added to 10% v/v. Penicillin (100 units/ml) and streptomycin (100µg/ml) were added to all media. HMT medium for growing hybrids of parental cell lines bearing hypoxanthine-guanine phosphoribosyl transferase deficiencies contained hypoxanthine at 1×10^{-4} M, methotrexate at 1×10^{-5} M and thymidine at 1.6×10^{-5} M.

Lymphoblastoid cell lines were grown in RPMI 1640 medium (*Imperial Laboratories*) with penicillin, streptomycin and sodium bicarbonate, supplemented with glutamine and 10% foetal calf serum.

Cells were cultured in plastic flasks (*Nunc*) which were kept in a humidified incubator at 37°C under 5% CO₂.

Hybrid cell lines were harvested by treatment with 10ml of trypsin solution (0.25% v/v in PBS) for a few minutes. An equal volume of serum-containing medium was added and the cells detached from the plastic by firmly slapping the flask.

All cell suspensions were pelleted in an MSE minor centrifuge at 1,000 rpm for 5 minutes and were washed in PBS. Cell densities were determined using a model FN Coulter Counter or haemocytometer.

Cells were frozen at a density of 10^6 cells/ml. Cells were suspended in ice cold 5% DMSO plus 95% foetal calf serum, cooled to -70°C overnight before being transferred to liquid N₂.

2.4 MAMMALIAN CELL DNA PREPARATION

2.4.1 Preparation of DNA from cell pellets and tissue

This technique may be used to prepare high molecular weight DNA from tissues (liver, kidney etc.) or cell pellets. When using tissue, a small piece (0.5-1cm³) should be frozen in liquid N₂ and ground to a powder using a pre-chilled pestle and mortar. Cell pellets (10⁷-10⁸ cells) should be washed with 1xSET.

1. Resuspend in approx. 20ml 1xSET. Add SDS to 0.5% and proteinase K to 100µg/ml.
2. Incubate at 55°C for 4 hours - overnight until clear.
3. Add an equal volume of phenol (equilibrated with 1xSET), shake well and spin at top speed on bench centrifuge for 10 minutes. Collect upper aqueous layer.
4. Re-extract aqueous layer with an equal volume of phenol/chloroform. Shake well and spin at top speed on bench centrifuge for 10 minutes. Collect upper aqueous layer.
5. Re-extract aqueous layer with an equal volume of chloroform. Shake well and spin at top speed on bench centrifuge for 10 minutes. Collect upper aqueous layer.
6. Add pancreatic RNAase to give 75µg/ml and incubate at 37°C for 2 hours.
7. Repeat steps 3-5.
8. At room temperature add 2 volumes of absolute ethanol to aqueous DNA solution. Gently invert tube several times until precipitate forms.
9. Collect DNA by spooling on a pasteur pipette that has had the end heat-sealed. Wash in 70% ethanol.
10. Transfer DNA to a sterile *Eppendorf* tube, allow to dry and add approx. 500µl sterile 1xTE:

2.4.2 Preparation of DNA from blood (K.H. Choo, personal communication)

This technique may also be used to prepare DNA from small (10^6 cells) pellets, in which case steps 1-5 should be omitted.

1. Collect 10ml of blood in a heparinised tube (or a tube containing 0.2ml of a 3% solution of K_2EDTA). Mix well, but gently.
2. Add 5ml 5% Dextran T-500/0.9% NaCl. Mix well.
3. Leave for 30 minutes at room temperature. Red cells will precipitate, white cells should remain in suspension.
4. Collect top layer and spin at 2000 rpm (bench centrifuge) for 10 minutes.
5. Carefully remove supernatant. Pellet can now be stored at $-20^{\circ}C$ or used immediately.
6. Resuspend in 10ml of lysis buffer: 0.32M Sucrose, 10mM Tris.Cl pH7.5, 5mM $MgCl_2$, 1% TRITON X100 v/v.
7. Leave for 5-10 minutes on ice.
8. Centrifuge for 15 minutes at top speed on bench centrifuge at $4^{\circ}C$.
9. Remove supernatant and resuspend nuclear pellet in 0.5ml 1xSET.
10. Transfer to *Eppendorf* tube. Add 25 μ l 10% SDS and 5 μ l proteinase K (10mg/ml). Mix well.
11. Incubate at $55^{\circ}C$ for 4 hours - overnight until clear.
12. Extract with 0.5ml phenol (equilibrated with 1xTE), by gently inverting for 5-10 minutes.
13. Centrifuge for 2-3 minutes.
14. Transfer aqueous layer to a clean tube.
15. Re-extract phenol with 200 μ l 1xTE. Gently invert for 2-3

minutes.

16. Centrifuge for 2-3 minutes and pool aqueous layers.
17. Extract with 600 μ l chloroform. Gently invert for 5 minutes.
18. Centrifuge for 3 minutes.
19. Transfer aqueous layer to a Universal tube.
20. Add 1ml 1xTE and 60 μ l 3M NaAcetate (pH5) and mix well.
21. Precipitate DNA by adding 4.5ml absolute ethanol and gently inverting the tube several times.
22. Collect DNA by spooling on a pasteur pipette that has had the end heat-sealed. Wash in 70% ethanol.
23. Transfer DNA to a sterile *Eppendorf* tube, allow to dry and add $\approx$ 100 μ l 1xTE.

2.5 ISOLATION OF PLASMID AND COSMID DNA

2.5.1 Alkaline lysis mini prep. method (Maniatis *et al.*, 1982)

1. Inoculate 5ml of LB/2xTY medium containing the appropriate antibiotic with a single bacterial colony. Incubate at 37°C overnight with vigorous shaking.
2. Transfer 1.5ml of culture to an *Eppendorf* tube. Centrifuge for 1 min in an *Eppendorf* microfuge.
3. Remove as much medium as possible by aspiration.
4. Resuspend pellet in 100 μ l ice-cold GTE containing 5mg/ml lysozyme and leave at room temp. for 5 minutes.
5. Add 200 μ l of a freshly prepared, ice-cold solution of 0.2M NaOH/1% SDS. Mix by inverting the tube rapidly two or three times. Store on ice for 5 minutes.
6. Add 150 μ l of an ice-cold solution of potassium acetate ($\approx$ pH 4.8) made up as follows : To 60ml of 5M potassium acetate, add 11.5ml of glacial acetic acid and 28.5ml of dH₂O. The resulting

solution is 3M with respect to potassium and 5M with respect to acetate.

Vortex gently for 10 seconds then store on ice for 10 minutes.

7. Centrifuge for 10 minutes then transfer the supernatant to a fresh tube.

8. Add an equal volume of phenol/chloroform. Mix by vortexing. After centrifuging for 2 minutes, transfer aqueous layer to a fresh tube.

9. Add two volumes of ethanol at room temperature. Mix by vortexing. Stand at room temp. for 5 minutes.

10. Centrifuge for 10 minutes at room temp.

11. Remove the supernatant by aspiration, then wash the pellet with 1ml of 70% ethanol. Vortex briefly then recentrifuge.

12. Again remove all of the supernatant. Dry the pellet briefly in a vacuum desiccator.

13. Add 50 μ l of TE (pH 8.0) containing DNase-free pancreatic RNase (20 μ g/ml). Vortex briefly.

2.5.2 Large scale preparation of plasmid and cosmid DNA (Maniatis *et al.*, 1982)

1. Resuspend the bacterial pellet from a 200ml culture in 5ml of GTE containing 5mg/ml lysozyme.

2. Transfer to a 30ml polypropylene tube. Let stand at room temperature for 5 minutes.

3. Add 10ml of a freshly made solution of 0.2M NaOH/1% SDS. Cover the tube with parafilm and mix the contents by gently inverting the tube several times. Let stand on ice for 10 minutes.

4. Add 7.5 ml of an ice-cold solution of potassium acetate ($\approx$ pH 4.8) made up as follows: To 60ml of 5M potassium acetate, add 11.5ml of glacial acetic acid and 28.5ml of dH_2O . The resulting solution is 3M with respect to potassium and 5M with respect to acetate.

Cover the top of the tube with parafilm and mix the contents by inverting the tube sharply several times. Let stand on ice for 10 minutes.

5. Centrifuge at 20,000 rpm for 20 minutes at 4°C. Bacterial DNA and debris form a tight pellet on the bottom of the tube.

6. Transfer equal quantities ($\approx$ 10ml) of the supernatant into each of two, siliconised, 30ml Corex tubes.

7. Add 0.6 volumes of isopropanol to each tube. Mix well and let stand at room temperature for 15 minutes.

8. Recover the DNA by centrifugation in a 'swing-out' rotor at 10,000 rpm for 30 minutes at room temperature.

9. Discard the supernatant. Wash the pellet with 70% ethanol at room temperature. Discard as much ethanol as possible, then dry briefly in a vacuum desiccator.

10. Dissolve the pellets in a total volume of exactly 4ml of TE (pH 8.0).

11. Add 4.41g of solid CsCl and 0.8ml of a solution of ethidium bromide (10mg/ml in dH_2O). Mix well. The final volume of the solution is $\approx$ 5ml and the density should be 1.55g/ml.

12. Transfer the CsCl /DNA solution (together with the protein aggregates that form) to a 5ml polyallomer tube. Overlay the solution with a layer of light parafin oil.

13. Centrifuge at 55,000 rpm in the TV65 rotor of a *Sorvall* OTD-65B ultracentrifuge for 16-24 hours at 20°C.

14. Two bands of DNA can be seen with a long-wave U.V. lamp. The lower band consists of closed circular plasmid/cosmid DNA. Collect the lower band into a 2ml syringe through a #21 hypodermic needle inserted into the side of the tube just below the band.
15. Transfer to a glass tube and extract several times with an equal volume isopropanol saturated with 20xSSC to remove the ethidium bromide. Continue extractions until no pink colour remains in either layer.
16. Dilute with an equal volume of dH₂O, transfer to a siliconised 15ml Corex tube and then add two volumes of absolute ethanol pre-chilled to -20°C.
17. Leave at -20°C for 2-3 hours and then centrifuge at 10,000 rpm in a 'swing out' rotor for 30 minutes at 4°C.
18. Wash the pellet several times with 10ml of 70% ethanol to remove all traces of CsCl.
19. Briefly dry the pellet in a vacuum desiccator and dissolve in 200µl of TE.

2.6 TRANSFER OF BACTERIAL COLONIES TO NITROCELLULOSE (Schleicher & Schuell) AND HYBOND-N (Amersham)

1. Grow bacterial colonies on suitable selective plates overnight at 37°C. Ideally colonies should be ≈1mm in diameter.
2. Cut 3 pieces of *Whatmann 3MM* paper so that they fit neatly into a plastic tray (≈22cm x 22cm).
3. Carefully place membrane onto agar surface. Mark the membrane and agar asymmetrically by stabbing with a syringe needle filled with waterproof ink, or other means, to ensure correct orientation of colonies.

4. Leave for 1 minute then using blunt-ended, *Millipore*, forceps remove membrane and place, colony side up, on one of the sheets of 3MM paper soaked in 10% SDS. Leave for 3 minutes.
 5. Transfer to a second sheet of 3MM paper that has been saturated with denaturing solution (0.5M NaOH, 1.5M NaCl). Leave the filter for 7 minutes.
- When transferring membranes from one tray to another, use the edge of the first tray as a scraper to remove as much fluid as possible from the underside of the membrane.
6. Transfer the membrane to the third sheet of 3MM paper that has been saturated with neutralising solution (1.5M NaCl, 1M Tris.Cl [pH 8.0]). Leave for 10 minutes.
 7. Wash the membrane in 2xSSC.
 8. Transfer to dry filter paper and air dry, colony side up.
 9. Nitrocellulose: bake for 2 hours at 80°C in a vacuum oven.
- HYBOND-N: place colony side down on U.V. transilluminator for 2-5 minutes.

2.7 ANALYSIS AND FRACTIONATION OF DNA USING AGAROSE GELS

2.7.1 Agarose Gel Electrophoresis

Agarose (*Sigma*, Type II) was dissolved in 1xTBE by autoclaving or in a microwave oven, and cooled to 60°C before pouring into a gel mould. Ethidium bromide was added to the molten agarose to a final concentration of 0.5µg/ml. Sometimes gels were stained after electrophoresis in buffer containing 0.5µg/ml ethidium bromide. Large gels were run using the *Pharmacia* GNA-200 apparatus and 'mini-gels' using the *Pharmacia* GNA-100 appartatus or the *BRL* Model H6 apparatus. Before loading, 1/10 volume of loading dye (≈50% glycerol, ≈50% 10xTBE,

0.025% bromophenol blue) was added to DNA samples. The voltage and time of electrophoresis varied.

Gels were photographed on a 254nm transilluminator, using a red filter and Polaroid 665 film.

Fragment sizes were estimated by running standard restriction digests containing fragments of known size e.g. phage λ digested with *Hind*III. Using these standards, fragment sizes of unknown length could be calculated by the BASIC computer programme shown in the appendix.

The molecular weights of the size markers used routinely were (in base pairs):

λ /*Hind*III: 23130, 9416, 6557, 4361, 2322, 2027, 564, 125

λ /*Hind*III/*Eco*RI: 21226, 5148, 4973, 4263, 3530, 2027, 1904,
1584, 1375, 947, 831, 564, 125

pUC9/*Hae*III: 587, 458, 434, 298, 267, 236, 174, 102, 80, 18, 11

pUC9/*Msp*I: 501, 489, 404, 320, 242, 190, 147, 110, 101, 67, 34,
34, 26

pUC9/*Sau*3A: 945, 585, 341, 258, 130, 105, 78, 75, 46, 36, 18,
17, 12, 11, 8

(*New England Biolabs. Inc. Catalogue 1986/7*)

2.7.2 Southern blotting of agarose gels

DNA was routinely capillary blotted from agarose gels onto HYBOND-N (*Amersham*), using a modified method of Southern (1975).

1. After electrophoresis incubate gel in 0.25M HCl for 15 minutes at room temp..

This increases the efficiency of transfer of large DNA fragments (>10kb) by breaking them into smaller fragments, due to depurination.

2. Place gel in enough denaturing solution (1.5M NaCl,

- 0.5M NaOH) so as to completely cover the gel. Gently agitate on an orbital shaker for 45 minutes.
3. Replace denaturing solution with a similar amount of neutralising solution (1.5M NaCl, 1M Tris.Cl pH8.0). Gently agitate for 60 minutes.
 4. Remove neutralising solution and briefly rinse gel with de-ionised water.
 5. Cut HYBOND-N membrane to exact size of gel to be blotted.
 6. Half fill a large plastic box with 25mM sodium phosphate buffer (pH6.5). Place a glass plate onto a support in the middle of the reservoir and form a wick over the glass plate using two crossing pieces of *Whatman 3MM* filter paper.
 7. Place gel onto filter paper ensuring no bubbles are trapped. Place membrane onto gel, roll a glass pipette over membrane to remove any trapped bubbles. Place two pieces of *Whatman 3MM* (pre-soaked in phosphate buffer) on top of membrane without trapping any bubbles.
 8. Flank gel with strips of 'cling film' to ensure that the blotting buffer goes *through* the gel, not around it.
 9. Place a stack of absorbent paper towels (10cm high) on top and compress using a glass plate and 100-500g weight.
 10. Allow transfer to continue for 4-16 hours.
 11. After blotting, carefully remove membrane and rinse in phosphate buffer to remove any adhering agarose. Allow surface to dry.
 12. Wrap in 'cling film' and place, DNA side down, onto U.V. transilluminator for 5 minutes.

2.7.3 Recovery of DNA from agarose gels (Dretzen *et al.*, 1981)

1. Run DNA in a horizontal agarose gel containing 0.5µg/ml ethidium bromide. Use *Pharmacia NA* or a similar high purity agarose.
2. Cut a piece of *Schleicher & Schuell* NA45 (DEAE) membrane that will be just large enough to reach in front of entire band to be eluted.
3. Soak membrane in 2.5M NaCl for 2-4 hours (while gel is running). Rinse in several changes of dH₂O and then soak in gel running buffer for at least 5 minutes.
4. Visualise desired band in gel using long wave UV light. If short wave UV transilluminator has to be used, expose DNA for as short a time as possible to minimise 'nicking'.
5. Cut with scalpel in front of desired band.
6. Using tweezers to hold the piece of NA45, and a second pair of (*Millipore*) tweezers to hold the slit in the gel open, place the membrane into the slit. Allow the slit to close, ensuring that no bubbles have been trapped. Check that the membrane is correctly positioned and adjust if necessary.
7. Run the gel at high current (100mA) until the band as visualised with the UV light has completely stacked up on the NA45 (5-10 minutes).
8. Remove the membrane and briefly rinse in 1xTE. Place it into an *Eppendorf* tube containing 0.4ml of 1M NaCl. Do not let the membrane become dry at any stage.
9. Incubate at 70°C for 20 minutes to elute DNA from the membrane.
10. Remove the membrane from the tube and extract the DNA solution with 0.4ml phenol/chloroform.

11. Centrifuge for 3 minutes. Remove and re-extract the aqueous layer with 0.4ml chloroform.
12. Centrifuge for 3 minutes. Transfer the aqueous layer to a clean tube and add absolute ethanol, pre-chilled to -20°C , to fill the tube. Precipitate at -20°C overnight.
13. Centrifuge for 10 minutes. Wash the DNA pellet in 70% ethanol.
14. Allow the pellet to dry and dissolve in 10-20 μl 1xTE.

2.8 LABELLING DNA WITH ^{32}P

2.8.1 Nick translation of DNA

DNA was labelled with ^{32}P - α -dCTP (>3000 Ci/mmol, 1mCi/ml, Amersham) using an Amersham Nick-Translation Kit. Volumes of 30 μl were routinely used and the reaction was allowed to proceed for 2 hours at 14°C . The reaction was stopped by adding 10 μl of 25mM EDTA/0.5% SDS/0.025% bromophenol blue and the volume increased to 100 μl with 1xTE. Labelled DNA was separated from unincorporated nucleotides by spin-dialysis through a 1ml column of Sepharose CL-6B equilibrated with 1xTE. Incorporation was typically 10^8 cpm/ μg DNA.

2.8.2 End labelling of λ /*Hind*III marker fragments with ^{32}P

A.

1. Mix 2 μl λ /*Hind*III (approx. 1 μg), 13 μl dH₂O, 1 μl 2mM dATP, dTTP, dGTP and 2 μl 10x T4 DNA polymerase buffer (0.33M Tris-acetate (pH 7.9), 0.66M potassium acetate, 0.1M magnesium acetate, 5mM dithiothreitol).
2. Transfer to tube containing approx. 4 μCi ^{32}P - α -dCTP.

3. Add 2.5 μ l (2.5 units) T4 DNA polymerase, mix well and incubate at 37°C for 5 minutes.
4. Add 1 μ l 1mM dCTP and continue incubation for 10 minutes.
5. Stop reaction by adding 10 μ l 25mM EDTA/0.5%SDS/0.025% bromophenol blue and increase volume to 100 μ l with 1xTE.

B.

1. Mix 2 μ l λ /*Hind*III (approx. 1 μ g), 15 μ l dH₂O and 5 μ l Nucleotide buffer solution from *Amersham* Nick-Translation Kit.
2. Transfer to tube containing approx. 4 μ Ci ³²P- α -dCTP.
3. Add 1 μ l (1 unit) Klenow fragment, mix well and stand at room temp. for 10 minutes.
4. Stop reaction by adding 10 μ l 25mM EDTA/0.5% SDS/0.025% bromophenol blue and increase volume to 100 μ l 1xTE.

In both cases separate labelled DNA from unincorporated nucleotides by spin dialysis through a 1ml column of Sepharose CL-6B. Check activity of labelled DNA by scintillation counting of a 1 μ l sample. Aim to load 5000 cpm onto gel for an overnight signal.

2.9 REMOVAL OF UNINCORPORATED NUCLEOTIDES BY SPIN DIALYSIS

1. Pierce the bottom of a 1.5ml *Eppendorf* tube with a fine needle. Add $\approx$ 25 μ l of a suspension of *Sigma* Type IV glass balls in 1xTE. The glass balls should have been previously acid washed and autoclaved in 1xTE.
2. Fill the tube with a slurry of *Sigma* Sepharose CL-6B, equilibrated with 1xTE. Pierce several holes in the cap of the tube.

3. Place two 1.8ml *Nunc* freezing vials (or similar) in a 15ml *Nunc* culture tube. Place the tube containing the Sepharose CL-6B into the top of this and centrifuge for 3 minutes at 1500 rpm in a bench centrifuge. The volume of Sepharose in the column should reduce to about 1ml.
4. Remove the top vial and replace with a clean one. Load labelled DNA, in a volume of $\approx 100\mu\text{l}$, onto the column and centrifuge as above. Wash the column through with $100\mu\text{l}$ of 1xTE by re-centrifuging.
5. Remove the vial containing the labelled DNA and discard the column which should contain the unincorporated nucleotides.

2.10 HYBRIDISATION TO DNA ON FILTERS

Pre-hybridisation

1. Place membrane (typically HYBOND-N) into a plastic bag and add at least 5ml of hybridisation solution per 100cm^2 of membrane.

Hybridisation solution contains (final concentration):

5x Denhardt's solution, 5xSSPE, 0.3% SDS, (10% dextran sulphate).

2. Non-homologous DNA (herring/salmon sperm) is denatured by heating in a boiling water bath for 5 minutes (or in a microwave oven on full power for 2 minutes), chilled on ice and added to the hybridisation mixture at a final concentration of $100\mu\text{g/ml}$.
3. The bag is heat-sealed and placed in a water bath at 64°C for 1-2 hours for pre-hybridisation.

Hybridisation

1. Denature probe by heating in a boiling water bath for 5-10 minutes and chill on ice.
2. Remove approx. half of the pre-hybridisation solution and add to it the denatured probe. Mix well and re-introduce into bag. N.B. The probe concentration should not exceed 50-100ng/ml of hybridisation buffer and this must be reduced to 10-20ng/ml if dextran sulphate is present.
3. Hybridise for at least 12 hours at 64°C.

Washing of filters

Following removal from bag, filters were briefly rinsed in 2xSSC/0.1% SDS at room temp. Remaining washes were typically carried out at 64°C, the stringency of the wash being dependent on the nature of the experiment.

On completion of washing the membranes were sealed in plastic bags while still damp.

Autoradiography was carried out at -70°C with an intensifying screen.

Removal of probes from membranes

Hybridisation probe was removed by washing the membrane at 64°C in several changes of 1mM Tris pH 7.5/2mM EDTA/0.1% SDS for periods of 30 minutes, until little or no activity could be detected by a G.M. hand-counter.

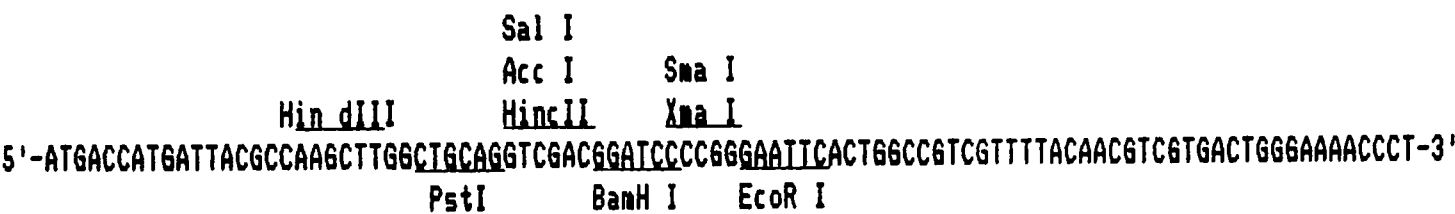
2.11 SUBCLONING OF RESTRICTION FRAGMENTS INTO PLASMID VECTORS

2.11.1 Preparation of vector

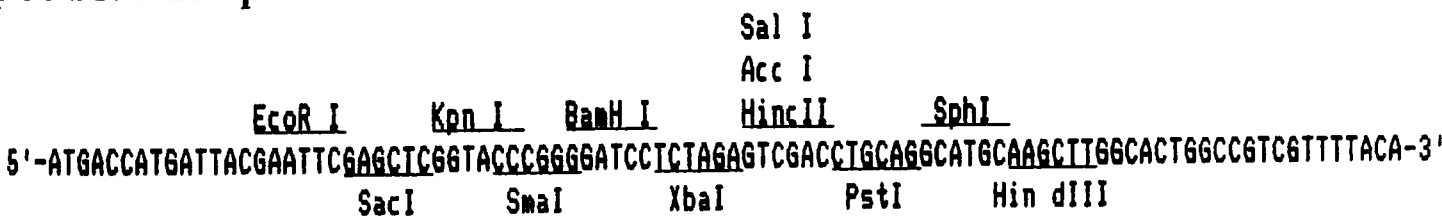
The vector routinely used was pUC9. On occasions pBR322, pUC18 and the replicative form of the single-stranded phage vector M13mp18 were also used. The advantages of using pUC9 and pUC18 over other plasmid vectors (and this is also true for M13mp18) is the presence of multiple cloning sites in the *LacZ* selectable marker and the ease of screening for recombinants (recombinants appear white as opposed to blue when grown on plates containing X-gal). Transformants can be selected through the presence of AMP^R genes in pUC9 and pUC18.

The restriction enzyme sites present in these vectors are:

pUC9



pUC18/M13mp18



Vector DNA was digested with a 5 fold excess of the required restriction enzyme(s) for 2-5 hours. Residual enzyme activity was destroyed by heating for 5 minutes at 70°C. The volume was increased to 200µl and the vector was purified by extraction with phenol/chloroform, chloroform and ethanol precipitation. The vector was dissolved in 1xTE at a concentration of ≈10ng/µl. An aliquot was examined on a mini-gel to check for complete digestion.

2.11.2 Ligation of DNA fragments to vector DNA

Fragments were ligated to vector by either 'shot gun' cloning complete restriction digests of e.g. cosmid DNA, or by using fragments eluted from agarose gels (see 2.7.3). Restriction digests were purified in the same way as the vector. Typically 10-20ng of vector was ligated to 50-100ng insert in a reaction volume of 10-20 μ l using 1 unit of T4 DNA ligase. Ligations were carried out overnight at 14°C in buffer containing 50mM Tris-HCl (pH 7.6), 10mM MgCl₂, 1mM ATP and 1mM DTT. This was made as a 10x stock and stored at -20°C. Reactions were heated at 65°C for 10 minutes and then stored on ice prior to being used for transformation.

2.12 PREPARATION OF COMPETENT CELLS AND TRANSFORMATION OF *E. coli* HB101/JM83/JM101

2.12.1 Preparation of competent cells (CaCl₂ Technique)

1. Inoculate 10ml 2xTY medium with a single colony from a plate and grow overnight in a 37°C shaking incubator.

If growing *E. coli* JM101 for transformation with an M13 derivative a colony must be picked from a glucose/minimal medium plate to ensure that the cells carry the plasmid for F-pilus synthesis, as this structure is necessary for M13 infection. Since this plasmid also carries a gene for proline synthesis (while the host cell carries a deletion of this gene), selection on glucose/minimal medium plates allows the growth only of those cells bearing the plasmid.

2. Inoculate 100ml 2xTY medium with 1ml of overnight culture and grow to O.D.₆₀₀ 0.3-0.5 (approx. 2 hours).

3. Chill on ice. Pellet 20ml of culture by centrifugation at 2000 r.p.m. for 10 minutes in a bench centrifuge at 4°C.

For all subsequent stages ensure that pipettes and solutions are pre-chilled to 4°C for maximum transformation efficiency.

4. Discard supernatant then gently resuspend in 10ml 10mM MgSO_4 by slowly sucking up and down in pipette.

5. Keep on ice for 5 minutes then pellet by centrifugation for 10 minutes.

6. Discard supernatant and gently resuspend in 10ml 50mM CaCl_2 .

7. Keep on ice for 30 minutes then pellet by centrifugation for 10 minutes.

8. Discard supernatant and gently resuspend in 2ml 50mM CaCl_2 .

9. Leave on ice for 2-3 hours.

The protocols for transformation of *E. coli* HB101, JM83 and JM101 differ in the selection for recombinants.

2.12.2 Transformation of *E. coli* HB101/JM83 and selection for recombinants

1. Transfer 0.1ml aliquots of competent cells to sterile polypropylene culture tubes on ice.

2. Add ligation mix (use HB101 for pBR322 and JM83 for pUC9 etc.) and leave on ice for 1-2 hours.

3. Heat shock cells at 42°C for 2 minutes. Place tubes back on ice.

4. Add 1ml of 2xTY and incubate at 37°C for 1 hour with moderate shaking.

5. Plate out aliquots on suitable selective plates and incubate at 37°C overnight.

Plates were made from 2xTY containing 1.5% agar. Selection for pBR322 was made with 75µg/ml ampicillin or 15µg/ml tetracycline. Selection for pUC9 was made with 75µg/ml ampicillin and 2µl/ml 2% X-gal (in dimethylformamide).

2.12.3. Transformation of *E.coli* JM101 and selection for recombinant phage

1. Transfer 0.3ml aliquots of competent cells to sterile polypropylene culture tubes on ice.

At this stage add one drop of overnight culture to 20ml 2xTY to provide fresh cells which will be ready for use at the plating out stage.

2. Add ligation mix (M13mp18 + insert) to tubes and keep on ice for 1-2 hours.

3. Heat shock cells at 42°C for 2 minutes. Place tubes back on ice.

4. While cells are being heat shocked, prepare the following mixture for each tube: 40µl IPTG (100mM), 40µl X-gal (2% in dimethylformamide) and 200µl fresh *E. coli* cells (culture should be 3-4 hours old). This may be prepared in batch.

5. Add 270µl of fresh cells/X-gal/IPTG mix to each tube of heat shocked cells.

6. To each tube add 3ml molten H top agar, kept at 42°C. Mix by gently rolling and pour immediately onto pre-warmed (37°C) H plate. Leave to set at room temp.. Invert the plates and incubate overnight at 37°C.

2.13 PREPARATION OF SINGLE-STRANDED DNA TEMPLATES FROM M13mp18 CLONES

1. Inoculate 100ml 2xTY with 1ml of an overnight culture of *E.coli* JM101 (prepared from a single colony taken from a glucose/minimal medium plate).
2. Dispense 1.5ml aliquots into sterile culture tubes.
3. Inoculate each tube with a colourless (recombinant) plaque, removed from a plate as a plug using a sterile Pasteur pipette.
4. Shake the tubes for 5 hours at 37°C. Transfer to *Eppendorf* tubes.
5. Centrifuge for 5 minutes. Pour the supernatants into fresh tubes - being very careful not to pick up any cells.
6. Re-centrifuge and transfer 1ml of the supernatant into a fresh tube.
7. Add 200µl of a 20% solution of PEG 6000/2.5M NaCl. Mix well and leave for 15 minutes.
8. Centrifuge for 5 minutes. Discard the supernatant.
9. Centrifuge for 2 minutes and carefully remove all traces of PEG using a drawn out Pasteur pipette.
10. Dissolve the viral pellet in 100µl 1xTE. Add 50µl phenol saturated with 1xTE. Vortex for 15 seconds.
11. Stand at room temp. for 15-30 minutes.
12. Vortex for 15 seconds then centrifuge for 3 minutes.
13. Transfer the aqueous layer to a fresh tube.
14. Add 10µl 3M NaAcetate. Add 250µl ethanol pre-chilled to -20°C.
15. Leave overnight at -20°C to precipitate DNA.
16. Centrifuge for 10 minutes. Wash pellet with 1ml cold (-20°C) ethanol. Pour off the ethanol and briefly dry pellet in

a vacuum desiccator.

17. Dissolve in 50 μ l 1xTE and store at -20°C.

18. The yield of DNA can be estimated by running a small sample (2-5 μ l) on a mini-gel, using single-stranded mp18 as a size marker. Recombinant DNA should have a reduced mobility, but this is only observable with inserts greater than 300bp.

2.14 GENERAL DNA METHODS

2.14.1 DNA storage

Double-stranded DNA was stored in 1xTE at 4°C or at -20°C for longer-term storage. Single-stranded DNA was stored in 1xTE at -20°C.

2.14.2 Estimation of DNA concentration

DNA concentrations were estimated spectrophotometrically ($OD_{260} \approx 1.0$ for 50 μ g/ml ds DNA, or for 40 μ g/ml ss DNA). The concentration of small amounts of DNA were estimated by running on an agarose gel and comparing to standards. Alternatively, for concentrations in the range 1-25ng/ μ l, $\approx 5\mu$ l of the sample was mixed with an equal volume of 1xTE containing 2 μ g/ml ethidium bromide. This was spotted onto a piece of 'cling film' placed on the transilluminator, along with a range of known concentrations of DNA treated in the same way. The spots were photographed and a comparison of the intensity of the unknown sample with that of the standards allowed the concentration to be estimated.

2.14.3 Miscellaneous

Phenol for DNA extraction was not routinely re-distilled, unless the crystals had a pink tinge. It was first equilibrated with 1M Tris. pH 8.0 followed by several changes of 0.1M Tris.

pH 8.0, until the pH of the phenol was ≈ 7.0 . It was finally equilibrated with either 1xSTE or 1xTE depending on the application. 8-Hydroxyquinoline was added to a final concentration of 0.1%

'Chloroform' is a 14:1 mixture of chloroform and isoamylalcohol.

RNase A (10mg/ml in 1xTE) for use in DNA preparations was boiled for 10 minutes to destroy any DNase activity.

CHAPTER 3

ISOLATION OF SEQUENCES SPECIFIC TO THE HUMAN X CHROMOSOME

3.1 INTRODUCTION

Sequences specific to the human X or Y chromosomes were first isolated using an approach that utilises nucleic acid reassociation to purify DNA of a specific 'probe' chromosome from radioactive genomic DNA. This purification depends on the reassociation between radiolabelled DNA from cells containing probe chromosome and an excess of unlabelled 'driver' DNA from cells devoid of probe chromosome, the chromosome content of the 'probe' and 'driver' cells being otherwise identical. This approach was employed to obtain DNA specific for the Y chromosome by using female DNA as the driver DNA and male DNA as the source of the probe chromosome (Kunkel *et al.*, 1976) and later to isolate X-specific DNA, by reassociating labelled DNA from human-mouse hybrid cells, containing a single human X chromosome, with an excess of mouse DNA (Schmeckpeper *et al.*, 1979). This approach gave some early insights into the organisation of human sex chromosome DNA, but unique sequence probes were not isolated until several years later.

Several methods have been used for the isolation of unique sequence probes for the X and Y chromosomes. The most obvious approach is to screen unique sequence DNA, isolated at random from human genomic or cDNA libraries, on male and female DNA or somatic cell hybrids containing just X or Y chromosomes. However, this technique is laborious and not suited to isolating a large number of probes. Nevertheless, the first

probe for a unique sequence shared by the X and Y chromosomes was isolated in this way (Page *et al.*, 1982). A more efficient method is to construct genomic libraries with DNA from X or Y chromosomes, which have been physically separated from the autosomes by the fluorescence activated cell sorter. This method has proved effective for the X chromosome (Davies *et al.*, 1981; Gillard *et al.*, 1987) and also several autosomes (Krumlauf *et al.*, 1982), but has been applied with less success to the Y (Müller *et al.*, 1983; Cooke *et al.*, 1983), since the early banks constructed suffered from contamination with sequences derived from other chromosomes. Many well characterised probes have been isolated by the combined approaches of somatic cell genetics and molecular biology. Human-rodent somatic cell hybrids which retain only a limited human chromosomal constitution have been used by several groups to construct partial genomic DNA libraries of particular human chromosomes (e.g. Gusella *et al.*, 1980). For Y chromosome libraries, human-mouse hybrids or human-hamster hybrids containing the human Y chromosome, as the only detectable human chromosome contribution, have been used to construct total genomic libraries in phage λ (Bishop *et al.*, 1983; Burk *et al.*, 1985) and cosmid vectors (Wolf *et al.*, 1984a). The two vector systems have different advantages. A cosmid vector may be chosen as clones usually have large (≈ 45 kb.) inserts which facilitates chromosome 'walking' and often has sequences necessary for expression of insert DNA in eukaryotic cells. The phage λ vector Charon 21A has been used to construct a library from the euchromatic portion of the Y chromosome (Müller *et al.*, 1986). This vector has a maximum acceptable insert size of 9kb. By

cleaving DNA from flow sorted Y chromosomes with *HindIII*, the library constructed was enriched for DNA from the Y chromosome euchromatic portion since the heterochromatic portion of distal Yq consists of highly repeated satellite DNA virtually lacking *HindIII* sites and so most of this DNA was not clonable in this system. However, these sequences are probably under-represented in cosmid libraries as well, since they may also be selected against in the cloning procedure due to the inherent instability of tandemly repeated sequences even in *RecA*⁻ hosts (Brutlag *et al.*, 1977).

There are several approaches for isolating unique sequence probes from these types of banks. Banks constructed from flow sorted chromosomes should contain only X or Y material (although in practice there is often contaminating autosomal material) and so, theoretically, every clone could be used as a probe for mapping purposes. However, due to the presence of repeat sequences in many clones, a direct approach to mapping is not usually feasible, since the smear of repeat hybridisation may obscure bands produced by low-copy components. Clones containing essentially single-copy inserts can be identified by screening with total genomic DNA labelled with ³²P under conditions where only repetitive sequences will be detected (Gusella *et al.*, 1980) and by picking those clones which fail to hybridise with the probe. However, since the phage and cosmid vectors used in the construction of libraries have a lower size limit that is often in excess of the repeated sequence interspersion distance, most clones will contain at least one representative of a family of highly repeated sequences (Maniatis *et al.*, 1978) and consequently only a few

single-copy probes can be isolated directly. For banks made from rodent-human hybrid cell-lines, it is first necessary to differentiate those clones containing human DNA inserts (presumed to be X or Y derived if this was the only detectable human chromosome contribution) from those containing rodent material. Because there is very little cross-hybridisation between human and rodent repetitive sequences, radiolabelled human genomic DNA can again be used as a probe. For the reasons given above, the majority of clones containing human inserts will hybridise with the probe and so can easily be identified. Single-copy DNA fragments within insert sequences can be identified by Southern analysis of restriction digested clones, again by using radiolabelled total human DNA to detect repeat containing fragments. Alternatively, restriction enzyme digested cosmid DNA can be subcloned into another vector. Subclones containing inserts which are not highly repeated can be selected by the method described above. There have been some attempts to use cosmid clones directly as hybridisation probes, by overcoming the problem of repeat sequences by other methods. One approach has been to add a large excess of competitor DNA (e.g. sheared total human DNA) to the probe (e.g. Weiss *et al.*, 1980) to the hybridisation mix to increase the ratio of low to high copy signal. Although this approach has met with some success, results appear not to be consistently reproducible. A more successful approach, originally described by Brison *et al.* (1982) and more recently simplified by Sealey *et al.* (1985), relies on the pre-reassociation of the probe with a large excess of total human DNA to a C_{0t} value that essentially removes repeat sequences from the probe.

The aims represented in this chapter were to isolate sequences from the X chromosome, specifically those that mapped to the region that pairs with the Y chromosome at meiosis (terminal Xp). Not only is this region of interest because it undergoes recombination with the Y chromosome, but also because sequences within, or close to this region escape from X-inactivation. Any other sequences isolated that mapped to other regions of the X chromosome of interest e.g. telomeres, centromere, regions of X/Y homology or close to any of the known disease loci, were also to be analysed further.

3.2 THE SCREENING OF A COSMID BANK FOR THE ISOLATION OF X-SPECIFIC SEQUENCES

3.2.1 Materials and Methods

Cell-lines and hybrids were grown as described in 2.3. DNA was isolated using the method described in 2.4.1.

DNA from the hybrids PIP, HORL911R8B & MCP-6 was a gift from Dr. Y. Boyd.

All genomic DNA samples were digested overnight with at least a two-fold excess of enzyme, under the conditions recommended by the supplier. Samples were routinely run on a 0.8% agarose gel overnight at 40mA/35V. Filters were prepared by blotting gels onto HYBOND-N (2.7.2). Conditions for hybridisation and washing of filters were as in 2.10.

For the screening of the bank, cosmids were grown on LB plates containing 75µg/ml AMP. Plates were incubated at 37°C for 12-18hrs until the colonies were 1-2mm in diameter. Colonies were screened after being transferred to nitrocellulose filters (2.6). Filters were hybridised with genomic DNA probes (1-10µg) that had been labelled to a specific activity of $>10^7$ cpm/µg. The volume of hybridisation fluid and the amount of probe used varied according to the number of filters being screened. Hybridisation was at 64°C

overnight in the absence of dextran sulphate. Conditions of washing are described in the text.

Bacterial cultures were grown overnight in medium containing 75µg/ml AMP. Cosmid DNA was isolated using the alkaline lysis mini-prep. (2.5.1).

Preparation of insert DNA by elution onto DEAE paper is described in 2.7.3.

All probes were labelled with ^{32}P - α -dCTP by nick translation (2.8.1).

3.2.2 Identification of Cosmids Containing Human Sequences

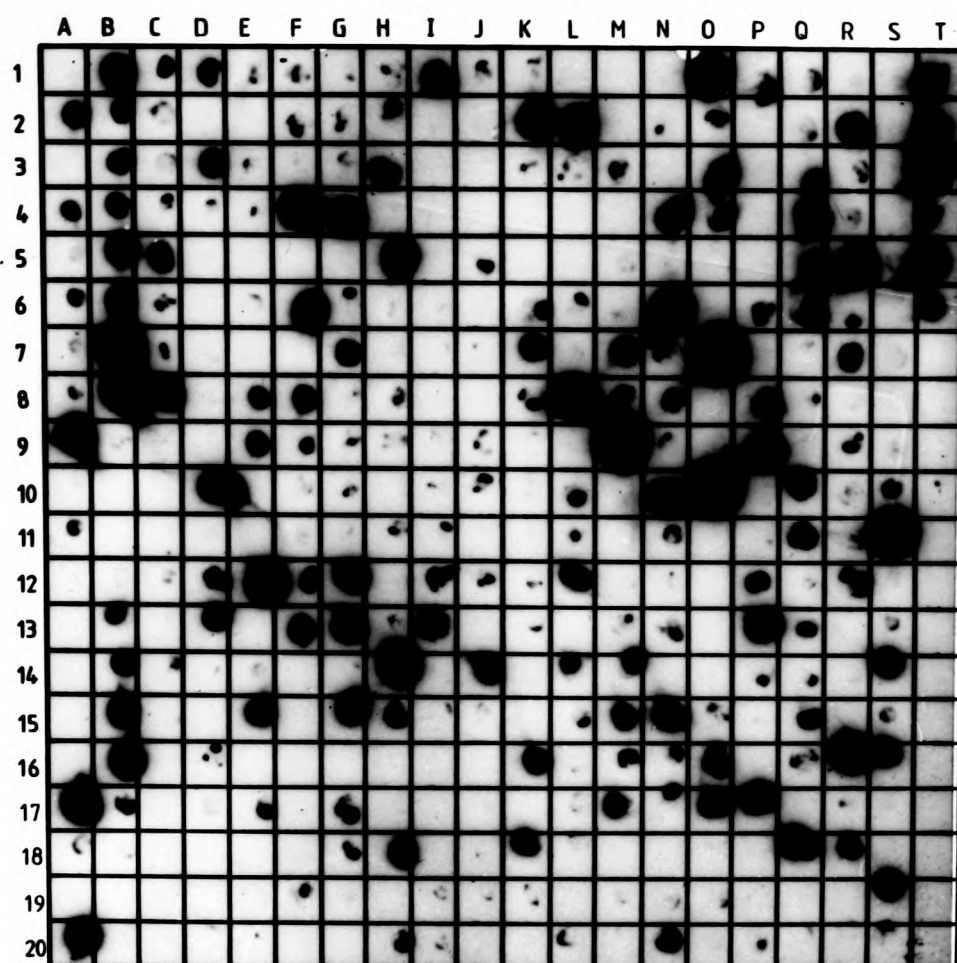
The source of X-chromosomal material used in this study was a cosmid bank recently constructed from the hybrid MOG-T (D.J.G Rees & G.G. Brownlee, Sir William Dunn School of Pathology, Oxford; unpublished information). MOG-T is derived from the hybrid MOG which contains an intact human X chromosome. However, this hybrid has two fragments of the human X chromosome translocated onto mouse chromosomes as its only human component (Boyd, 1987). Together, these fragments are thought to span the whole of the X chromosome. Briefly, size fractionated (35-45kb.) fragments from a *Sau3A* partial digest of the hybrid DNA were cloned into the cosmid vector pTM1, using methods previously described (Grosveld et al., 1982). The complexity of the bank was estimated to be $\approx 10^6$ with a background of non-recombinants of <0.3%. The majority of clones would be expected to contain mouse DNA, since the single human X chromosome present in the hybrid represents only a small fraction of the total DNA content of the cells. However, colonies containing human (and thus X chromosomal) DNA can be identified by probing colonies with radioactively labelled total human genomic DNA, as has been previously mentioned in

the introduction to this chapter.

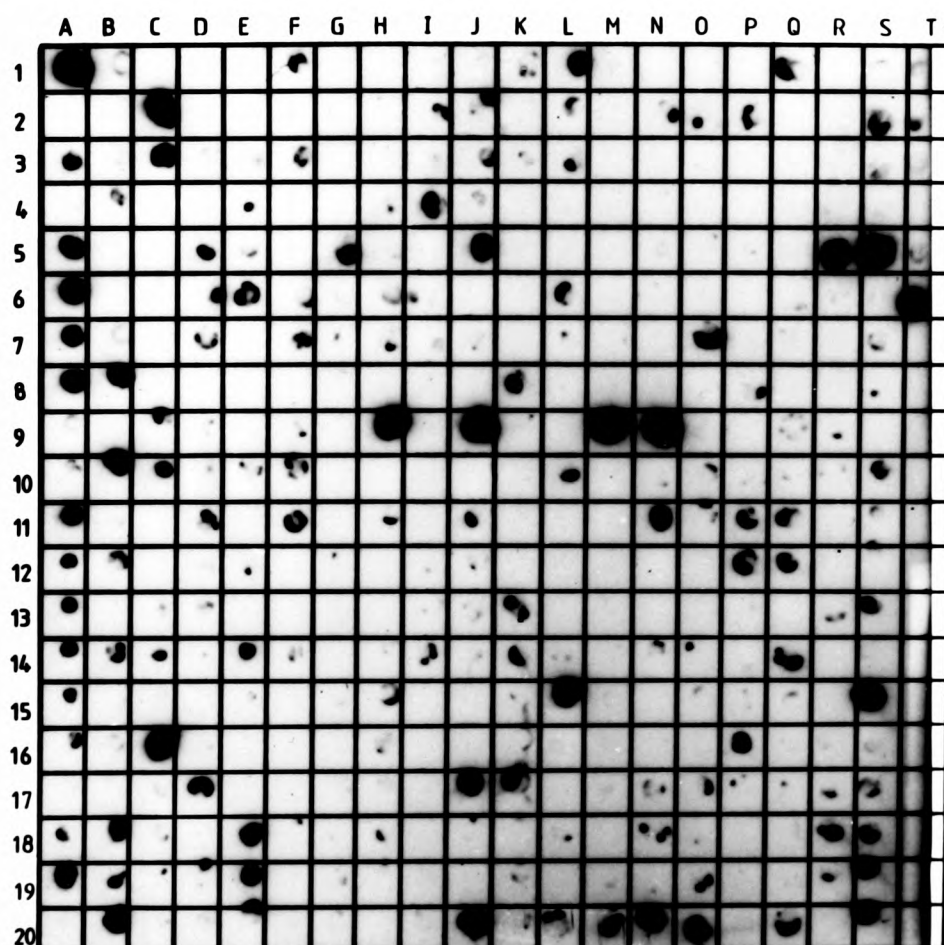
The titre of the bank was determined by plating out a series of dilutions onto LB plates and was found to be $\approx 10^5/\text{ml}$. Ten large LB/AMP plates (22x22cm) were spread to give a cell density of 10-20 colonies/cm². After overnight growth colonies were transferred to nitrocellulose filters and denatured. The plates were returned to the incubator for 3-6 hours to allow regrowth of the bacteria and were then stored at 4°C. Plates and filters were carefully marked to allow orientation of the colonies.

The bank was screened using ³²P labelled total human ϕ genomic DNA as the probe. After an overnight hybridisation, filters were washed stringently (0.1xSSC, 64°C) and autoradiographs left to expose for 1-7 days to allow identification of positive colonies.

Each plate gave ≈ 100 strong signals (1-2% of the total number of colonies). Orientation of some of these with respect to the original plates was difficult, especially if they fell in an area that was very crowded with colonies. Initially 386 putative positive colonies were picked and transferred in a grid pattern to duplicate 22x22cm LB/AMP plates. Colonies were grown overnight, transferred nitrocellulose and denatured. Plates were incubated at 37°C to allow regrowth of colonies and then stored at 4°C. One of the duplicate filters was probed with total human ϕ DNA and the other with total mouse ϕ DNA. Conditions of hybridisation and washing were as above, although volumes and amount of probe used in the hybridisation were reduced. The results of the probings are shown in Figures 3.1a & 3.1b and summarised in 3.1c.



3.1a

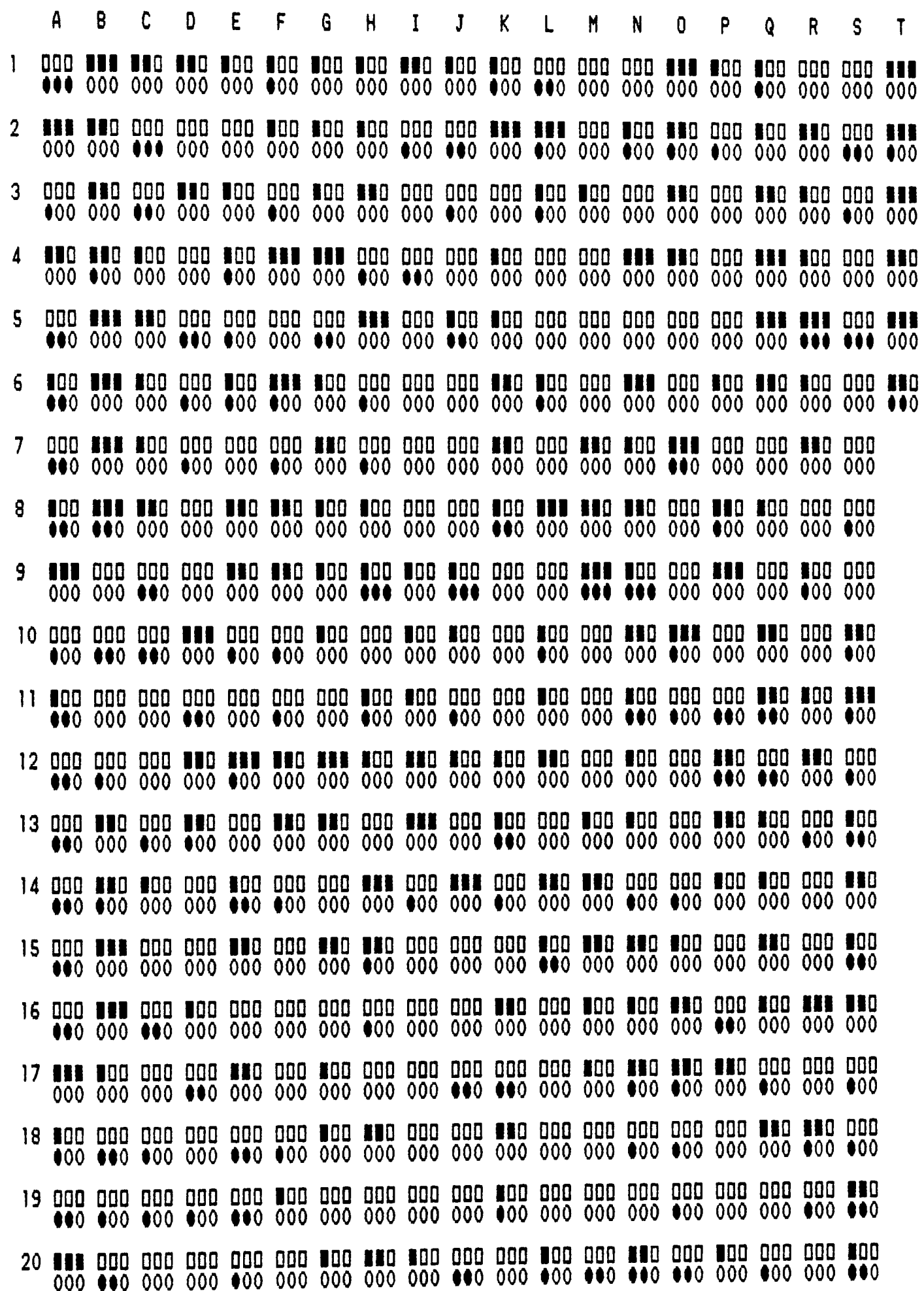


3.1b

FIGURE 3.1

Identification of cosmids containing human material.

Duplicate filters of bacterial colonies from the MOG cosmid library probed with total human DNA (3.1a) and total mouse DNA (3.1b).



KEY :

DEGREE OF CROSS-HYBRIDISATION

PROBE:

HUMAN

MOUSE

000

000

000

000

HIGH

MEDIUM

LOW

NONE

FIGURE 3.1c
 Summary of the data presented in Figures 3.1a & 3.1b.

The 386 colonies probed with both human and mouse DNA can be categorised as follows with respect to their degree of cross-hybridisation with the two probes:

 2	 2	 6	 30	 0	 5	 10	 55
 3	 12	 13	 57	 3	 34	 43	 111

For explanation of symbols see key to Figure 3.1c.

Colonies that hybridised solely or predominantly to human DNA were selected for further analysis. Overnight cultures of these colonies were grown in LB broth containing 75µg/ml AMP. Aliquots were stored frozen at -20°C and in liquid N₂, after the addition of sterile glycerol to a final concentration of 15%. This gave a total of 167 cultures which were redesignated M1-M167 (see Table 3.1).

An additional 71 clones (M168-M238) were added to this total by rescreening further colonies from the original plates in a similar manner. For the isolation of M198-M238 a total of 170 colonies were picked and screened with human and mouse probes. The results were as follows:

Human only - varying signal strength with human probe: 28
 Human signal >> mouse signal: 12
 Human signal & mouse signal (varying strength): 8
 Mouse signal >> human signal: 20
 Mouse only - varying signal strength with mouse probe: 49
 No signal with either probe: 53

The 40 colonies hybridising solely or predominantly to the human probe were designated M198-M238.

1-A2	31-E1	61-H2	91-L12	121-O10	151-R6
2-A4	32-E3	62-H3	92-L14	122-O15	152-R7
3-A9	33-E8	63-H5	93-A20	123-O16	153-R11
4-A17	34-E9	64-H8	94-G17	124-O17	154-R12
5-B1	35-E12	65-H12	95-K7	125-P1	155-R16
6-B2	36-E15	66-H14	96-M3	126-P6	156-R18
7-B3	37-E17	67-H15	97-M7	127-P8	157-S10
8-B4	38-F2	68-H18	98-M8	128-P9	158-S11
9-B5	39-F4	69-H20	99-M9	129-Q14	159-S14
10-B6	40-F6	70-I1	100-M13	130-P12	160-S16
11-B7	41-F8	71-I9	101-M14	131-P13	161-S19
12-B8	42-F9	72-I10	102-M15	132-P14	162-T1
13-B13	43-F12	73-I11	103-M16	133-P17	163-T2
14-B14	44-F13	74-I12	104-M17	134-P20	164-T3
15-B15	45-F19	75-I13	105-N4	135-Q2	165-T4
16-B16	46-G1	76-I20	106-N6	136-Q3	166-T5
17-B17	47-G2	77-J1	107-N7	137-Q4	167-T6
18-C1	48-G3	78-J10	108-N8	138-Q5	
19-C4	49-G4	79-J12	109-N10	139-Q6	
20-C5	50-G6	80-J14	110-N12	140-Q8	
21-C6	51-G7	81-K2	111-N13	141-Q10	
22-C7	52-G8	82-K4	112-N15	142-Q11	
23-C8	53-G9	83-K5	113-N16	143-Q13	
24-C14	54-G10	84-K6	114-N17	144-Q15	
25-D1	55-G12	85-K12	115-N20	145-Q16	
26-D3	56-G13	86-K16	116-O1	146-Q18	
27-D10	57-G15	87-K18	117-O2	147-R2	
28-D12	58-G18	88-L2	118-O3	148-R3	
29-D13	59-G20	89-L8	119-O4	149-R4	
30-D16	60-H1	90-L11	120-O7	150-R5	

TABLE 3.1

Re-designation of the MOG cosmid containing human inserts.
A2-T6 refer to the colonies depicted in Figures 3.1a-c.

The X chromosome represents $\approx 5\%$ of the total haploid content of the human genome i.e. $\approx 1.5 \times 10^8$ bp, the maximum size of a cosmid insert is ≈ 40 kb, this means that this partial X library of 238 clones (assuming that none are duplicated) represents a maximum of 1/16 of the total DNA content of the X chromosome.

3.2.3 Isolation of cosmid DNA and identification of single-copy restriction fragments

Colonies were grown overnight in LB broth and DNA was isolated using the alkaline lysis mini-prep.. The amount of culture grown varied from 10-30ml and the amounts of reagents used in the DNA preparation adjusted accordingly. The yield and quality of DNA obtained varied extensively and seemed to be dependent on the particular clones being examined. For a 10ml culture the total amount of DNA recovered ranged from 1-25 μ g, some clones gave consistently better yields than others. The purity of DNA recovered also seemed to be clone dependent. Some were much easier to digest with restriction enzymes and some suffered badly from degradation, this did not affect all of the preparations in a particular batch to the same degree and clones that were difficult to isolate were consistently difficult. Attempts were made to improve the methods for growing the bacteria and isolating DNA. These included using richer broth (2xTY) and attempting to amplify the cosmids by the addition of chloramphenicol once the culture had reached log phase. None of these had much effect. If DNA of sufficient quantity and/or quality could not be isolated from a particular clone it was not pursued. To allow full analysis of a

particular clone at least 10µg of DNA was required.

Before attempts could be made to sub-regionally assign any of the cosmids, it was necessary to identify and isolate any single copy sequences present. 2µg of cosmid DNA was digested with at least a two fold excess of *EcoRI* (i.e. >4 units). The times of digestion varied from 2-6 hours. With clones that suffered badly from degradation problems, the time allowed for digestion was as short as possible and a greater excess of enzyme was used. Duplicate 0.8% agarose gels of digested cosmids were run overnight to allow good separation of the bands. Gels were then blotted onto HYBOND-N overnight. One of the filters was probed with total human DNA and the other with total mouse DNA. For each hybridisation 1µg of genomic DNA was labelled to a specific activity of $\approx 10^7$ cpm/µg. Hybridisation was carried out overnight at 64°C. Filters were then washed under conditions of low stringency (2xSSC, 0.1% SDS, 64°C) any fragments not hybridising to the probe could then be assumed to be essentially single copy. Autoradiography was for 16-24 hours in the first instance, then a second film was put down for up to a week.

Typical results are shown in Figures 3.2 & 3.3. Single-copy sequences could be identified by comparing the resultant autoradiographs with photographs of the original gel and were designated α , β , γ , etc.. The molecular weights of these fragments are shown in Table 3.2. Of an additional 34 cosmids that were analysed in this way, 27 were found to have human inserts, 5 had mouse inserts and 2 contained inserts that hybridised to both the mouse and human probes. Most cosmids were found to contain at least one single copy sequence. Based

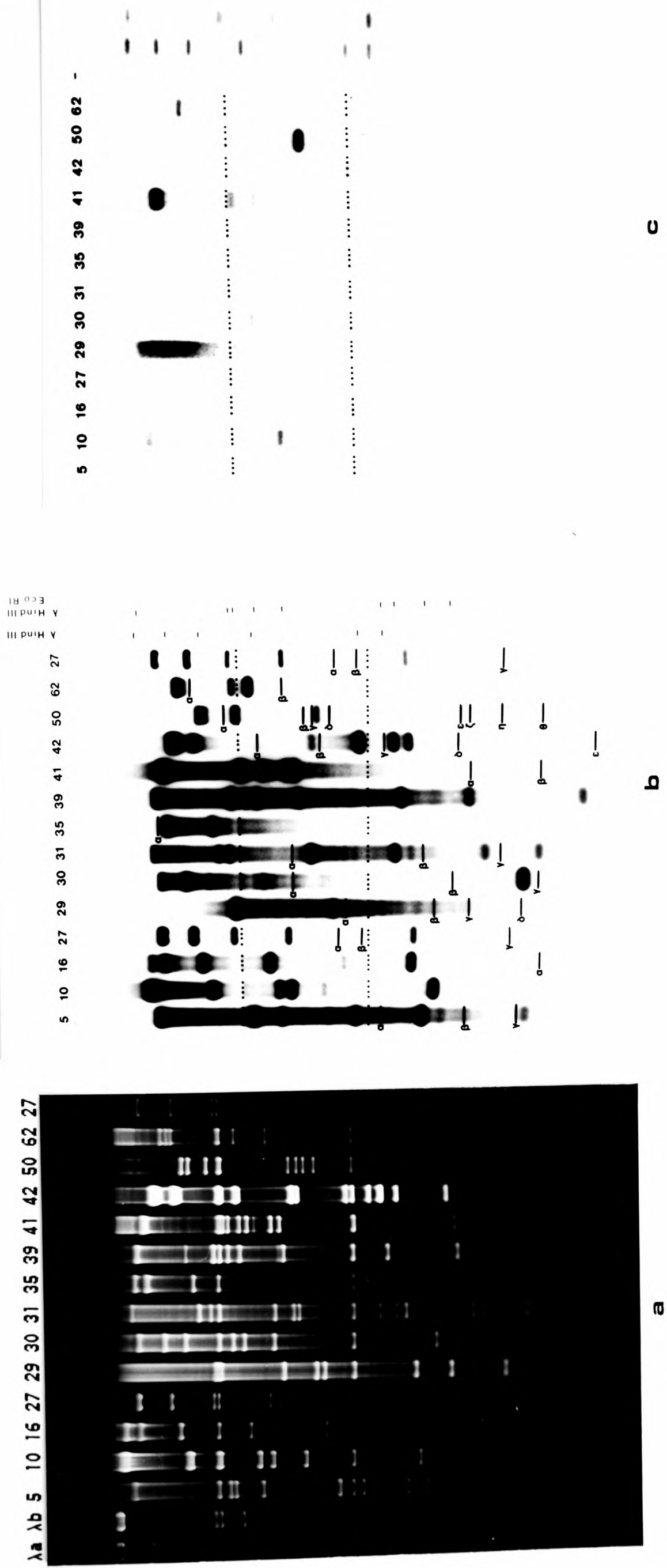


FIGURE 3.2

3.2a Ethidium bromide stained gel of *Eco*RI digests of various MOG cosmids. Duplicate filters were probed with total human DNA (3.2b) and total mouse DNA (3.2c). Cosmids with fragments that hybridise predominantly to human DNA, but not mouse, can be assumed to contain human material. Those fragments which do not hybridise with the probe represent single-copy sequences, these have been designated α , β , γ etc.. Positions of the vector bands have been marked Molecular weights of single-copy fragments were calculated from the migration of the λ /*Hind*III (λa) and λ /*Hind*III/*Eco*RI (λb) standards.

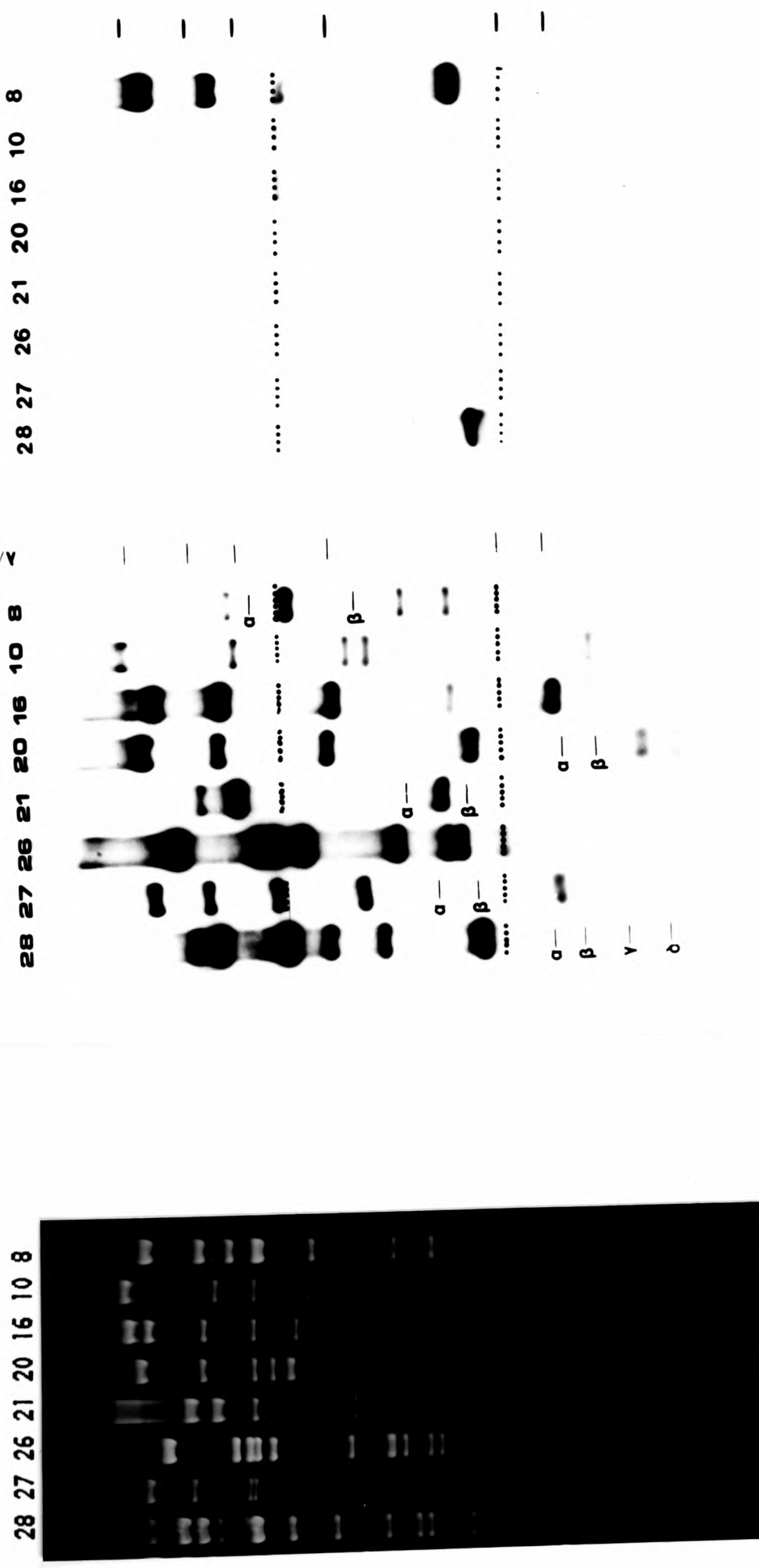


FIGURE 3.3

Identification of single-copy sequences within cosmids

3.3a Ethidium bromide stained gel of *EcoRI* digests of various MOG cosmids. Duplicate filters were probed with total human DNA (3.3b) and total mouse DNA (3.3c). Cosmids with fragments that hybridise predominantly to human DNA, but not mouse, can be assumed to contain human material. Those fragments which do not hybridise with the probe represent single-copy sequences, these have been designated α , β , γ etc.. 8α has been shown to contain a mouse single-copy sequence. Positions of the vector bands have been marked Molecular weights of the single-copy fragments were calculated from the migration of the $\lambda/HindIII$ standard (λ).

Cosmid №	Fragment	Size (kb)	Total Amount of Single Copy DNA
5	α	2.17	5.29
	β	1.66	
	γ	1.46	
8	α	6.10	9.66
	β	3.56	
10	-	-	-
16	α	1.38	1.38
20	α	1.61	3.07
	β	1.46	
21	α	2.90	5.21
	β	2.31	
26	-	-	-
27	α	2.54	6.36
	β	2.34	
	γ	1.48	
28	α	1.88	6.06
	β	1.58	
	γ	1.39	
	δ	1.21	
29	α	2.50	7.36
	β	1.80	
	γ	1.63	
	δ	1.43	
30	α	3.17	6.26
	β	1.71	
	γ	1.38	
31	α	3.17	6.54
	β	1.86	
	γ	1.51	
35	α	13.09	13.09
39	-	-	-
41	α	1.62	2.99
	β	1.37	
42	α	3.85	11.6
	β	2.75	
	γ	2.09	
	δ	1.68	
	ϵ	1.23	
50	α	5.07	19.66
	β	3.00	
	γ	2.84	
	δ	2.62	
	ϵ	1.66	
	ζ	1.62	
	η	1.49	
	θ	1.36	
62	α	7.12	10.49
	β	3.37	

TABLE 3.2

Molecular weights of the single-copy fragments identified in a selection of MOG cosmids by probing *Eco*RI restriction digests with total human genomic DNA (Figures 3.2b and 3.3b).

on the data shown in Table 3.2, the average size of a single copy *EcoRI* restriction fragment was found to be ≈ 2.6 kb, but some were as large as 13kb (e.g. M35 α). The total amount of single copy sequence present in any one cosmid was found to be ≈ 6.4 kb, i.e. $\approx 20\%$ of the insert DNA. It is possible that the amount of single-copy DNA present in any one cosmid is higher than that estimated, since regions of single-copy DNA juxtaposed to moderately or highly repetitive DNA on the same *EcoRI* restriction fragment would not have been identified as such by the screening procedure used. Where a cosmid contained more than one single-copy fragment, it is possible that they represent contiguous stretches of DNA. To ascertain whether this was so it would have been necessary to construct restriction maps of the cosmids, this was not done.

Most cosmids were found to contain some sequences that cross-hybridised with the mouse probe. Even those that had not hybridised at all to the mouse probe during the rescreening of the colonies. In nearly all cases it was fragments that had been already been identified as containing repeat sequences that hybridised, suggesting that there is some inter-specific cross-hybridisation of this type of sequence. The signal obtained with the mouse probe was usually weaker than that obtained with the human probe, even though they were of the same specific activity. Notable exceptions are some of the bands present in cosmid M8. Since there are fragments that hybridise solely to the human or mouse probe present, it is possible that this cosmid is chimeric i.e. containing both human and mouse sequences, an artefact that may have arisen during the original cloning procedure through the ligation of

two different sequences into the same vector. This is one of the problems that may arise when cloning into cosmid vectors, but is usually prevented by size fractionation, or phosphatase treatment, of DNA to be cloned. Fragments M8 α and M8 β did not hybridise to either probe so it was not possible to tell at this stage whether they were mouse or human specific.

3.2.4 Sub-regional localisation of single-copy sequences

The method chosen to sub-regionally assign the cosmids was to use eluted single-copy fragments as hybridisation probes against a somatic cell hybrid panel containing different representations of the human X chromosome. Enough cosmid DNA was digested with *EcoRI* to allow the recovery of at least 100ng of the fragments containing single-copy sequences. Fragments containing single-copy sequences were eluted from the gel onto NA45 DEAE paper. Recovery was usually 60-80%; smaller fragments (<7kb) being easier to recover. These fragments were not routinely subcloned, but were used directly as probes against genomic digests.

The panel used is shown in Figure 3.4. In summary the hybrids and regions of the X chromosome represented were:

HORL911R8B - Xpter-Xq2(2-4) (Boyd, 1987)
PIP - Xp11.4-Xp22.1 & Xq26-Xqter (Boyd, 1987)
MCP-6 - Xq13-Xqter (Goodfellow *et al.*, 1982)
WAG 8 - Xp21-Xqter (Boyd *et al.*, 1987)
UCLA_{B2} - Xp22.3-Xqter (Curry *et al.*, 1984)

The use of the hybrids PIP, HORL911R8B and WAG 8 as a mapping panel has been previously described (Boyd, 1987). Since one of the original aims of this investigation had been to isolate sequences that mapped to the region of the X chromosome that

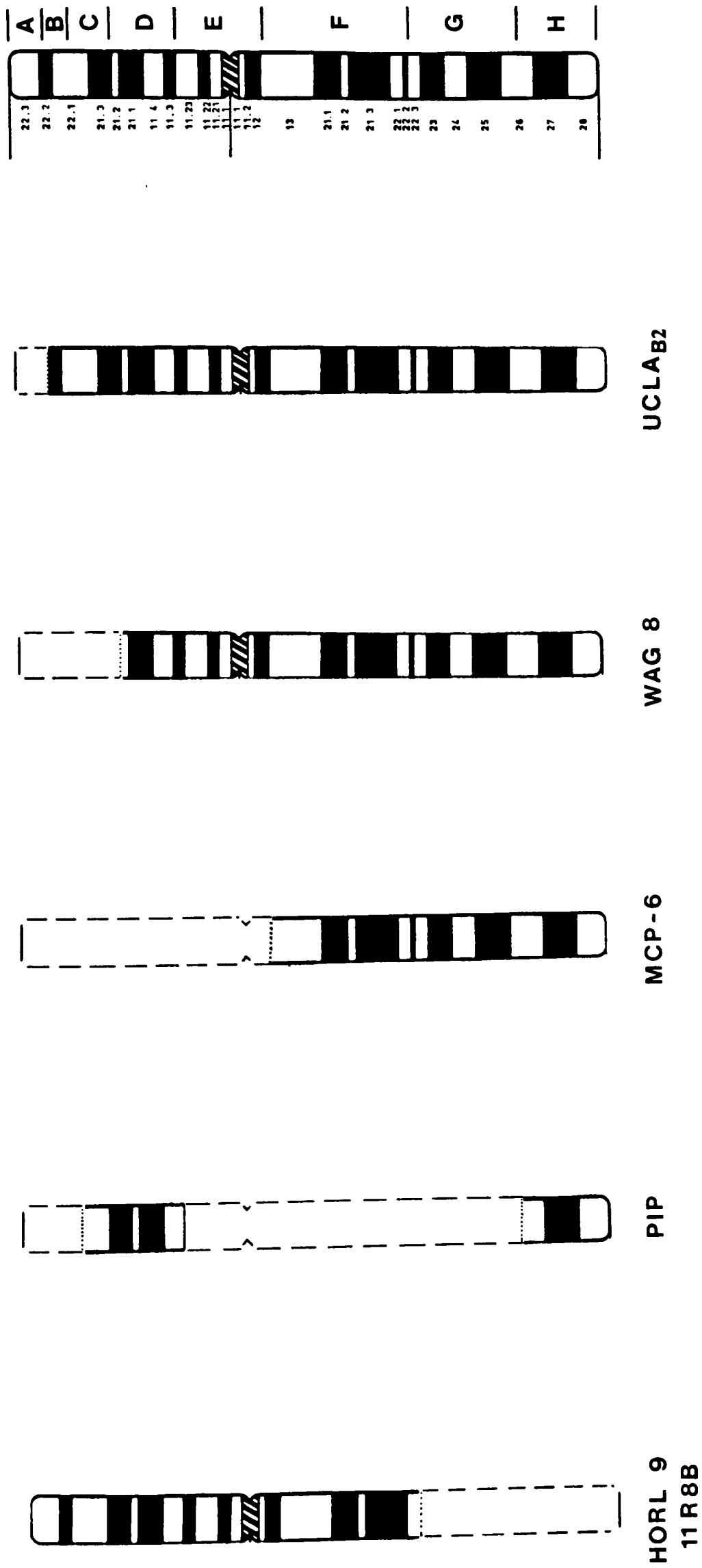


FIGURE 3.4

Diagrammatic representation of the regions of the human X chromosome retained in the hybrids used in the hybrid mapping panel described in the text. The regions to which sequences can be mapped using this panel (A-H) are shown on the right.

pairs with the Y at meiosis (terminal Xp), a panel consisting of just WAG 8 and UCLA_{B2} was routinely used. In addition to hybrid DNA normal human ♂ and ♀ DNA was used as controls. DNA from the cell-lines GM1416 (48,XXXX), OX (49,YYYYY) and 3E7 (Y-only hybrid) (Darling *et al.*, 1986) was often included to test for X and Y localisation and RAG (mouse cell-line (Klebe *et al.*, 1970)) as a test for mouse specific sequences.

The decision as to which fragments to use as hybridisation probes was based on a) ease at which the fragment could be isolated, well resolved bands could be recovered without contamination of other sequences b) size of fragment, as mentioned previously fragments >7kb were difficult to recover in high yield. Usually a fragment of ≈2-3kb was selected. Routinely only one fragment from any particular cosmid was used for mapping purposes. Eluted fragments were nick translated using 10-20μCi of ³²P-α-dCTP. 40-100ng of DNA was labelled and incorporation varied from 20-70%. Southern blots of *EcoRI* digest hybrid DNA were hybridised with the probes overnight at 64°C. Filters were routinely washed under stringent conditions (0.1xSSC, 0.1% SDS, 64°C). Single-copy fragments from cosmids 1, 3, 4, 5, 8, 10, 16, 20, 21, 27, 28, 29, 30, 31, 167, 200, 201 and 224 have been assigned to at least above or below Xp21 and sometimes more precisely. Results are summarised in Table 3.3 and typical autoradiographs shown in Figure 3.5.

As can be seen from the autoradiographs shown, most of the sequences were not entirely free from repetitive sequences and a background smear was apparent in many cases; nevertheless, the cognate band of the same molecular weight as the probe was usually easily discernable and the presence or absence of this

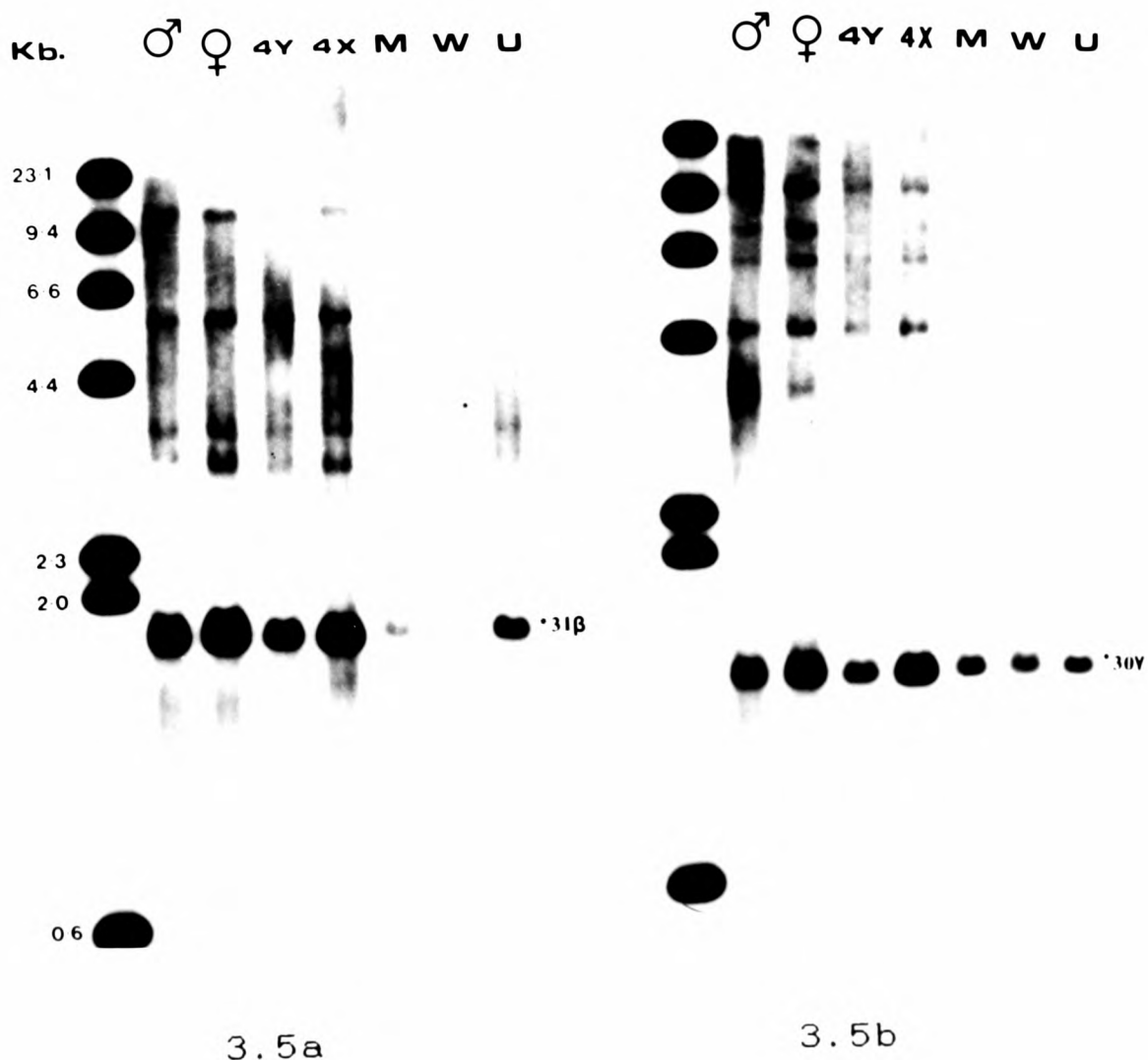
FRAGMENT	SUB-REGIONAL ASSIGNMENT	COMMENTS
M1 α	Xp21→Xqter	
M3 α	Xp21→Xqter	
M4 α	Xp21→Xqter	
M5 α	Xp21→Xqter	
M8 α	Mouse	Cosmid M8 contained repetitive sequences that hybridised to the mouse and human probes. See text.
M10 α	Xp21→Xqter	
M16 α	Xp21→Xqter	
M20 α	Xp21→Xqter	Additional 4 bands of equivalent intensity to cognate band (1.6kb) at 12kb, 7kb, 4.5kb & 4.3kb. See text.
M21 β	Xp21→Xqter (?)	Repetitive smear, cognate band not easily discernable.
M27 β	Xp21→Xqter	Multi-allelic, hypervariable. Subsequently mapped to Xp11.3→Xcen. See text.
M28 δ	Xp21→Xqter	
M29 α	Xp21→Xqter	Repetitive smear in addition to cognate band.
M30 γ	Xp21→Xqter	
M31 β	Xp22.1→Xp22.3	
M200	Xq24→Xq26	Cosmid contains multiple <i>Sfi</i> I sites (J. Boyce, Biochem. Pt. II Thesis, Oxford (1987)) See text.
M201 γ	Xp11.4→Xp21.1	Detects the locus <i>DXS231</i> . See Boyd, 1987
M224 α	Xp11.4→Xq24	

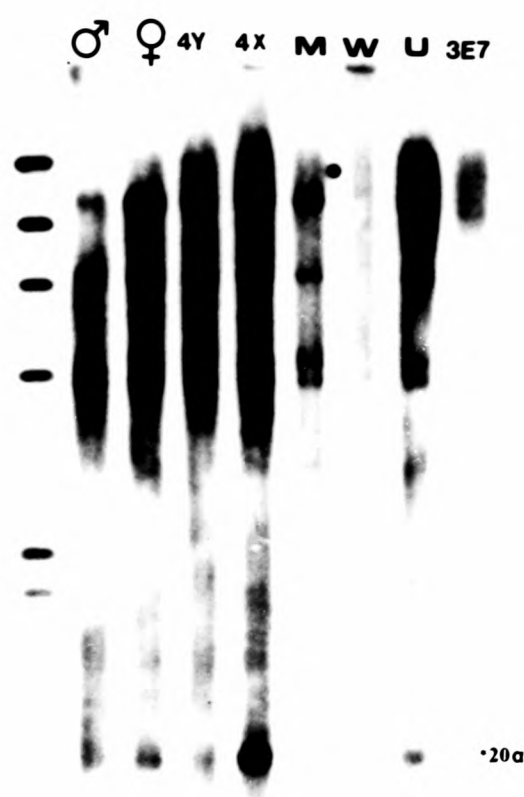
TABLE 3.3

Summary of the sub-regional localisation of single copy fragments from the MOG cosmids. Typical autoradiographs showing the sub-regional localisation of M20 α , M27 β , M29 α , M28 δ , M30 γ and M31 β are shown in Figure 3.5.

FIGURE 3.5
Sub-regional localisation of MOG cosmids

Sub-regional localisation of *Eco*RI restriction fragments from some of the MOG cosmids. In each case λ /*Hind*III molecular weight markers are shown in the far left track, the sizes of which are shown in 3.5a. The tracks contain *Eco*RI digests of DNA from normal male (σ), normal female (ϕ) and the cell-lines/hybrids GM1416 (4X), OX (4Y), MOG (M), WAG 8 (W), UCLA_{B2} (U) and in two cases, 3E7. These contain different representations of the human sex chromosomes (described in the text). In the examples shown M30 γ (3.5b), M20 α (3.5c), M28 δ (3.5d), M27 β (3.5e) and M29 α (3.5f) map proximal to Xp21 and M31 β (3.5a) maps distal to Xp21. The sub-regional localisation of M31 β has been refined to Xp22.1-Xp22.3. A cognate band the same size as the probe is present in each case except for M27 β . M27 β detects a hypervariable locus described in detail in Chapter 7. M20 α hybridises to five fragments with equal intensity, suggesting that it may be repeated several times, all of these fragments map proximal to Xp21.

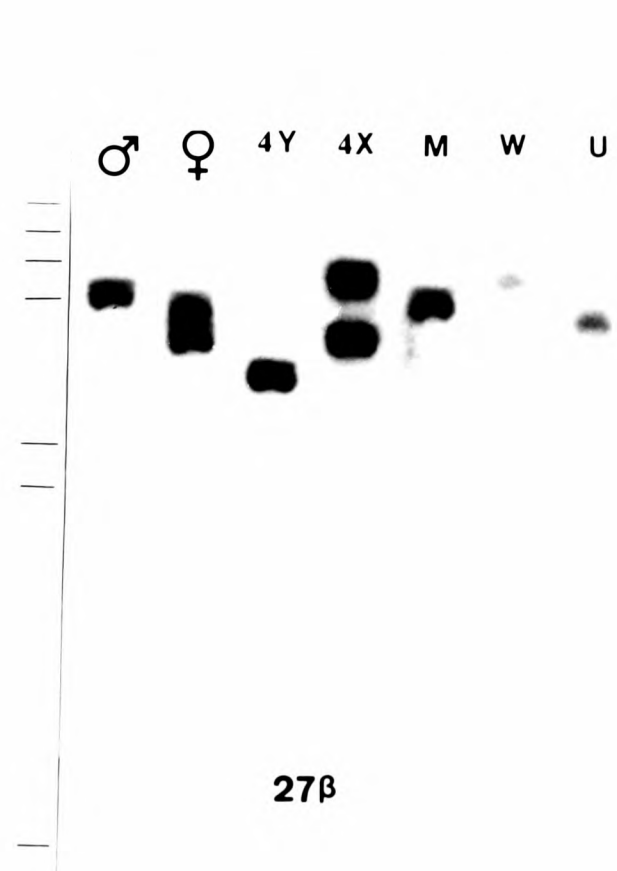




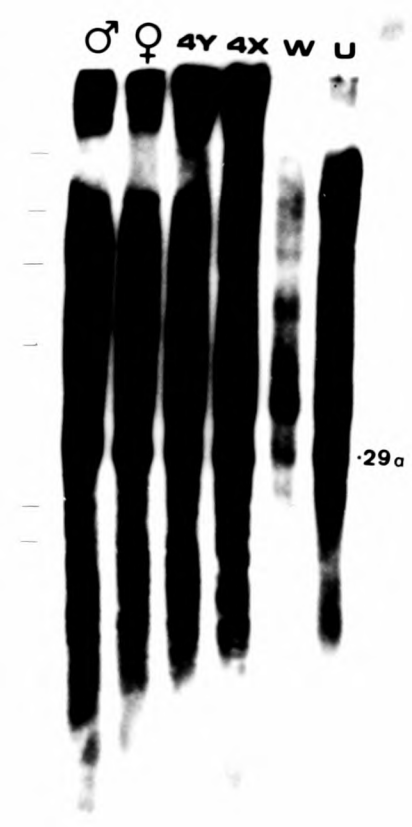
3.5c



3.5d



3.5e



3.5f

band in the different hybrids was used as the basis for sub-regional localisation. M20 α (M.W. 1.6kb) hybridised to 4 additional bands with equal intensity, suggesting that it may be repeated several times. All of these fragments were X-specific and mapped to the rather broad region Xp21→Xqter, their precise distribution was not analysed further.

The sequence M27 β which detects multi-allelic variation is discussed fully in a later chapter.

M31 β was the only sequence analysed that mapped distal to Xp21. Refinement of this original localisation to Xp21→Xp22.3, based on its pattern of hybridisation to the hybrids WAG 8 and UCLA_{B2}, was achieved through its lack of hybridisation to PIP.

3.3 DISCUSSION OF RESULTS

The original screening of the bank was successful in identifying many colonies cross-hybridising with human DNA and thus presumably containing X-chromosomal DNA. However, on subsequent rescreening with both human and mouse probes, only 43% of those originally picked appeared to contain human inserts. Nevertheless, this partial library has the probability of representation of 1/16 of the total DNA content of the X chromosome. Single-copy sequences that could be used as hybridisation probes were identified in the majority of cosmids analysed. Typically 20% of the cosmid insert was identified as being single-copy, although this is probably an under-estimate.

Some of these were successfully mapped using a somatic cell hybrid panel.

Although the investigation was by no means exhaustive, no sequences were found to map in the X-Y pairing region or were found to be X/Y homologous. The technical difficulties in isolating sufficient DNA of good quality from many of the cultures meant that this method of sub-regional localisation was very laborious and often frustrating. The adopted technique of eluting putative single-copy sequences from agarose gels and using these as hybridisation probes was time consuming. This was especially so during the early days of this investigation when recovered yields were often low and subsequent labelling by nick translation resulted in probes of low specific activity. Consequently, autoradiographs had to be left to expose for at least a week before results could be interpreted, which meant that large numbers of clones could not be screened rapidly. As has been mentioned previously, many cosmids were not analysed further because sufficiently large amounts of DNA could not be isolated easily. This need for microgram quantities of DNA was the main drawback of this approach. All of these problems precluded rapid mapping of a large number of clones. In retrospect, this was probably not the best approach to have adopted. Having identified cosmids containing putative single copy sequences it may have been more sensible to subclone complete *EcoRI* digestions of the cosmids directly into a plasmid vector (e.g. pUC9 which has been used routinely for subcloning). Clones containing inserts which are not highly repeated could have been selected again by screening with radioactively labelled total genomic DNA as described

previously. This time, clones not reacting with the probe would have been selected for further analysis. The inserts of such clones could be checked by comparing their sizes with the putative single copy fragments identified by the screening of the Southern blot of cosmid digests with the total human probe, this comparison would also have shown whether all of the putative single copy sequences had been isolated. This was the approach successfully adopted by Bishop *et al.* (1983) and also Wolfe *et al.* (1984a) for the isolation of single-copy probes from cosmid libraries of Y only hybrids. The need for several micrograms of cosmid DNA would have been obviated, since it would have been possible to subclone all of the *EcoRI* fragments from a typical cosmid with a maximum of 100ng. This technique of 'shotgun' cloning of cosmid restriction digests was subsequently employed with success for the subcloning of M31 β , M31 γ and M27 β , amongst others. Fragments subcloned into plasmid vectors are much easier to handle and yields of DNA from small cultures are substantially higher than those of cosmids.

As previously stated, only one particular single-copy sequence was routinely used from each cosmid for mapping purposes. Although this was perfectly adequate in most cases, the analysis of only one sequence means that obviously interesting sequences like M27 β could have been easily overlooked. It was very fortuitous that both M27 α and M27 β were chosen for analysis since M27 α does not display the same hypervariability as M27 β . To have analysed every single-copy sequence from each cosmid would have been very time consuming, although, in some cases it may have been possible and to label several sequence together to probe the same filter. An

alternative method that would have allowed analysis of all of the low copy sequences at the same time is the pre-association technique described by Sealey *et al.* (1985). This technique has been successfully employed for the analysis of several of the cosmids from this partial X library (Y. Boyd, Z.-Y. Cheng; unpublished results).

As well as using the partial library of X specific cosmids for the isolation of sequences mapping to particular regions of the X chromosome, it has been put to different uses as well. X-specific cosmids have been analysed for the presence of restriction sites for rare cutting enzymes. These restriction enzymes (e.g. *SfiI* - GGCCNNNN^{*}NGGCC and *NotI* GC^{*}GGCCGC) recognise very GC rich regions of DNA which it has been suggested are often associated with the 5' and 3' ends of genes (Bird, 1986). Of 25 cosmids analysed, 6 were found to contain at least one *SfiI* site and one was found to contain a *NotI* site (J. Boyce, Pt.II Biochemistry Thesis, Oxford (1987)). One of these, M200, which has been mapped to Xq24→Xq26, was found to contain multiple *SfiI* sites; this cosmid has yet to be investigated further. This approach may result in the isolation of previously unknown coding sequences or may give some insight into, the distribution and possible roles of GC rich regions of DNA.

The unscreened bank itself has been used for the isolation of the X homologue of an X/Y homologous sequence, p59 γ , (discussed in detail in Chapter 6) and may prove useful in the isolation of other specific regions for which probes are available.

Although the bank has not been fruitful in yielding sequences mapping to X-Y pairing region, other sequences of interest have been isolated from it. One of these, M27 β , is discussed in detail in Chapter 7.

An alternative method for the isolation of sequences mapping to distal Xp was attempted. The technique of 'deletion enrichment' is discussed in Chapter 4.

CHAPTER 4

AN APPRAISAL OF THE TECHNIQUE OF 'DELETION ENRICHMENT' FOR THE CLONING OF SEQUENCES MAPPING TO DISTAL Xp

4.1 INTRODUCTION

Several techniques have been devised based on subtraction hybridisation that allow the isolation of sequences that have been deleted from a source of DNA. An excess of randomly sheared driver DNA bearing the deletion is used in a reannealing reaction to compete out corresponding sequences in 'normal' tracer DNA, which has been cut with a restriction enzyme. During this process three types of DNA hybrid are predicted to form. In the first type of hybrid both strands would be derived from the sheared driver DNA. The second type would contain one sheared strand from the driver and one restriction enzyme cut strand from the tracer; tracer DNA homologous to sequences present in the driver DNA should enter this type of hybrid. In the third type of hybrid both strands are derived from the restriction enzyme cut tracer DNA. Single-stranded restriction fragments of tracer DNA that share no homology with the driver DNA should self-hybridise to form this type of hybrid. If the driver and tracer DNAs are otherwise identical, these should represent the sequences absent from the deletion in the driver (see Figure 4.1).

This approach was first used for analysing murine Y chromosome sequences by Lamar & Palmer (1984) (for the sake of this argument, Y-encoded DNA sequences can be considered 'deleted' from the female genome). It was also the technique

used by Kunkel *et al.* (1985) for isolating the *pERT87* locus, Smith *et al.* (1987) for the isolation of *HIP25*, both present in the Duchenne muscular dystrophy region of the X chromosome, and by Nussbaum *et al.* (1987) for the isolation of sequences from within a submicroscopic X chromosomal deletion around Xq21. Cotinot *et al.* (1987) have used the technique with some success to isolate bovine Y derived sequences and Bickmore & Cooke (1987b) have successfully isolated human Y chromosome specific sequences.

All of the techniques used rely on an enhancement of the rate of reassociation of tracer and driver DNA, such that almost 100% renaturation occurs in a relatively short time. The use of a phenol emulsion in the presence of chaotropic ions to enhance the reassociation rate, the PERT method, was first described by Kohne *et al.* (1977). Modifications of this technique were used by Kunkel *et al.* (1986), Nussbaum *et al.* (1987) and Bickmore & Cooke (1987b). Kohne has also devised conditions that utilise high concentrations of inorganic phosphate (2.4M) (used by Avery *et al.*, 1980 and Smith *et al.*, 1987), or ammonium sulphate (2M) (used by Lamar & Palmer, 1984 and Cotinot *et al.*, 1987) to enhance reassociation. The success of this approach has varied quite considerably and from the various reports the success rate has been 3/24 (Lamar & Palmer, 1984), 4/125 (Kunkel *et al.*, 1985), 1/6 (Smith *et al.*, 1987), 2/83 (Nussbaum *et al.*, 1987), 2/200 (Cotinot *et al.*, 1987) and 3/20 (Bickmore & Cooke, 1987b) clones analysed were specific for the region of interest.

The final yield of DNA using the PERT method has been reported to be only 5% of input (Kunkel *et al.*, 1985). This

seemed rather low and so, also for technical reasons and simplicity of approach, it was decided to try to reproduce the technique described by Lamar & Palmer (1984), which had not reported substantial losses of input material. The aim was to clone sequences from the X chromosome mapping to Xp21→pter and Xp22.3→pter, since these represent quite large deletions. The cloning scheme is shown in Figure 4.1. For the cloning of sequences mapping to Xp21→pter, WAG 8 DNA was to be used as the driver and for sequences mapping to Xp22.3→pter, UCLA_{B2}, in both cases DNA from the X only hybrid MOG was to serve as the tracer.

4.2 MATERIALS AND METHODS

4.2.1 General methods

Cell-lines and hybrids were grown as described in 2.3 and DNA was prepared using the method described in 2.4.1. Driver DNA was prepared from a human placenta.

Genomic digests were performed overnight using at least a two fold excess of enzyme under the conditions recommended by the supplier. Samples were run on 0.8% agarose gels overnight at 30mA/35V and blotted onto HYBOND-N (2.7.2).

Plasmid DNA was isolated from clones using the mini-prep. method described in 2.5.1. Insert DNA was isolated by elution onto DEAE paper (2.7.3).

The sub-cloning of *EcoRI/SalI* fragments into mp18 and isolation of recombinant phage was using methods similar to those described in 2.11, 2.12.3 and 2.13.

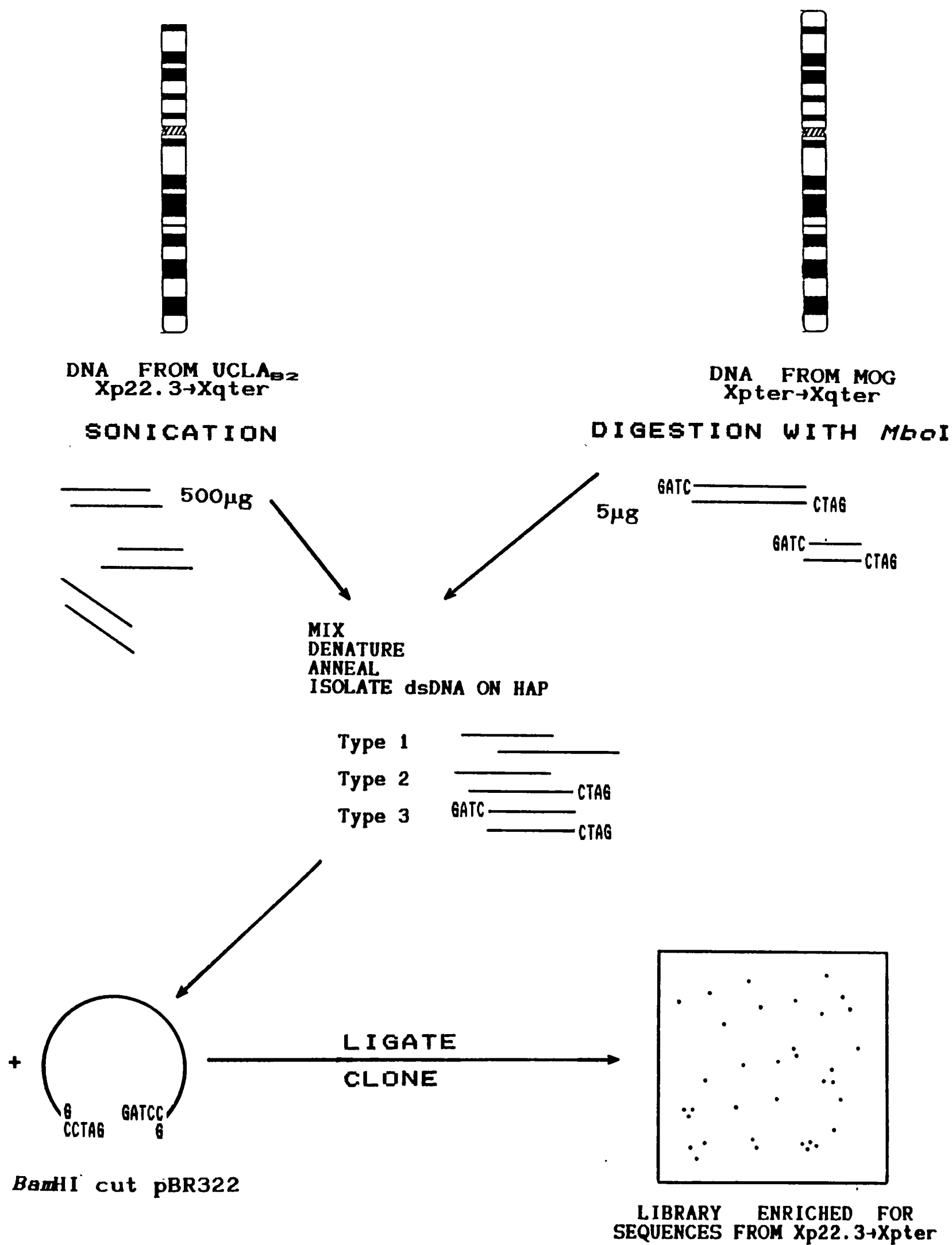


FIGURE 4.1. Deletion Enrichment Scheme for Cloning Sequences Mapping to Terminal Xp. (See text for details)

4.2.2 Deletion enrichment cloning (adapted from Lamar & Palmer, 1984)

1. Shear >500µg of driver DNA containing deleted region by sonication to fragments ≈500 bases long. The extent of shearing can be evaluated by analysing samples on a mini-gel, and comparing with standards of known molecular weight.
2. Purify by extraction with phenol/chloroform saturated with 1xTE.
3. Precipitate DNA by addition of ethanol pre-chilled to -20°C and recover by centrifugation in an *Eppendorf* microfuge. Wash pellet with 70% ethanol and dry briefly in a vacuum desiccator.
4. Dissolve in dH₂O and adjust concentration to 5µg/µl.
5. Digest >5µg of DNA without the deletion to completion with *Mbo*I.
6. Purify by extraction with phenol/chloroform saturated with 1xTE. Adjust salt concentration of DNA solution to 0.1M NaCl and precipitate DNA by addition of ethanol pre-chilled to -20°C, recover by centrifugation in an *Eppendorf* microfuge. Wash pellet with 70% ethanol and dry briefly in a vacuum desiccator.
7. Dissolve in dH₂O and adjust concentration to 0.5µg/µl.
8. Mix 100µl sonicated DNA (500µg) with 10µl *Mbo*I digested DNA (5µg) and add 40µl dH₂O. Denature by heating at 100°C for 4 minutes then immediately chill on ice.
9. Add 250µl 4M ammonium sulphate and 100µl 250mM sodium phosphate (pH 6.8)/25mM EDTA and gently mix. Overlay solution with sterile paraffin oil to prevent evaporation.
10. Transfer to a heating block and maintain at 68°C until desired C₀t is achieved (1-2 weeks).
11. De-salt the DNA solution by passing down a column of Sephadex G-50-150 (*Sigma*), formed in a Pasteur pipette. Load 200µl onto column and collect 100µl fractions. Localise the DNA peak by running 10µl samples of each fraction on a 1% agarose gel.
12. Pool the peak fractions and isolate double-stranded DNA by hydroxylapatite chromatography.
13. *Bio-Rad* DNA grade Bio-gel™ hydroxylapatite was re-hydrated in 0.01M sodium phosphate (pH 6.9), as recommended by the manufacturer. 0.5ml was placed into a *Bio-Rad* disposable Poly-

Prep™ chromatography column and was washed with several ml of 0.01M phosphate buffer (pH 6.9).

14. Load 200µl of de-salted DNA solution (containing ≈100µg DNA) and mix well with the hydroxylapatite to ensure all of the DNA is adsorbed.

15. Elute the single-stranded fraction with 0.12M sodium phosphate (pH 6.9). Collect 10, 300µl, fractions.

16. Elute the double-stranded fraction with 0.27M sodium phosphate (pH 6.9). Collect 10, 300µl, fractions.

17. Localise the DNA peaks by running 10µl samples of each fraction on a 1% agarose gel. Precise amounts of DNA present in each fraction can be measured by U.V. spectroscopy at 260nm.

18. Pool the double-stranded DNA fractions and de-salt as in 11. above.

19. Purify by extraction with phenol/chloroform saturated with 1xTE. Adjust salt concentration of DNA solution to 0.1M NaCl and precipitate DNA by addition of ethanol pre-chilled to -20°C, recover by centrifugation in an *Eppendorf* microfuge. Wash pellet with 70% ethanol and dry briefly in a vacuum desiccator.

20. Dissolve in dH₂O and adjust concentration to 1µg/µl.

21. Mix 1µg of the purified double-stranded fraction with 100ng of *Bam*HI cut, dephosphorylated pBR322 (*New England Biolabs*) in the presence of T4 ligase at 200U/ml in ligase buffer and leave at 14°C overnight.

22. Use the ligated DNA to transform 1.5ml of competent *E.coli* HB101 (2.12).

4.2.3 Single-Stranded Hybridisation Probe Generation (Hu & Messing, 1982)

High specific activity probes can be generated using the M13 hybridisation probe primer. The 13-base sequence of the primer (5'-GAAATTGTTATCC-3') is complementary to the 5' side of the multiple cloning site present in M13 mp vectors, and is used to initiate synthesis of the (-) strand from the (+) strand template by the Klenow fragment of *E.coli* DNA polymerase I. The synthesis of the complementary strand, labelled by the incorporation of ³²P-α-dCTP, does not go to completion and so the inserted probe sequence remains single-stranded. Probes prepared in this way are more sensitive than those prepared by nick translation, since there are no sequences

complementary to the probe in the hybridisation mixture competing with those immobilised on the filter.

Labelling:-

1. Mix together 50-100ng ssmp18 clone, 1.5 μ l 100mM MgCl₂/100mM Tris pH8.0, 2 μ l 'Hybridisation Probe Primer' (BRL) and dH₂O to a final volume of 8 μ l in a 0.4ml microcentrifuge tube. Place the tube in a water-filled ($\approx$ 5 ml) test tube (10x50mm) which is in a 90°C-100°C water bath. Incubate for 5 minutes to denature.
2. Remove the test tube holding the annealing mixture from the water bath and allow the tube to equilibrate to room temperature (30-45 minutes). The slow cooling step allows the primer to anneal to the template.
3. Add 1 μ l 0.5mM dATP, dTTP, dGTP, 1 μ l Klenow ($\approx$ 4 units), 2 μ l ³²P- α -dCTP (20 μ Ci) and 5 μ l dH₂O.
4. Incubate for 1 hour at 37°C.
5. Separate unincorporated nucleotides by spin dialysis through a column of Sepharose CL-6B.

Hybridisation :-

Prehybridisation and hybridisation is carried out using the same solution, consisting of: 1M NaCl, 10% w/v dextran sulphate, 1% w/v SDS , 50% v/v formamide, 250 μ g/ml denatured herring sperm DNA.

Prehybridise overnight at 43°C.

Remove the solution from the bag, add the probe (N.B. DO NOT denature !) and 10 μ g of ssmp18 DNA. Mix well and return to the bag.

Hybridise overnight at 43°C.

Washing of filters:-

Filters were routinely washed at 60°C. The stringency of washing ranged from 2xSSC/0.1% SDS to 0.1xSSC/0.1% SDS, depending on how strong a signal could be detected by a G.M. hand-counter.

4.3 AN EVALUATION OF THE TECHNIQUE BY ATTEMPTING TO CLONE Y-SPECIFIC DNA

To ascertain how successful this approach was likely to be, it was decided to set up a reassociation experiment using female DNA as the driver and DNA from the cell-line OX (49,XYYYY) as the tracer. In this way, it was hoped that Y-specific sequences would be cloned. Since the Y chromosomes present in OX represent a bigger 'deletion' than either of those present in WAG 8 or UCLA_{B2}, the number of Y-specific clones obtained would give an upper limit of those to be expected.

The reassociation was set up as described and annealed to a C_{ot} value of 2100. The rate of reannealing in the ammonium sulphate buffer is reported to be accelerated 50-fold relative to the rate in 0.14M sodium phosphate buffer (pH6.8). A reassociation using phosphate buffer was set up as well so that the degree of renaturation in the two could be compared.

Following de-salting of the DNA, aliquots from the two reassociations were passed over hydroxylapatite to ascertain the relative amounts of single-stranded and double-stranded DNA present. The amount of DNA present in each fraction was determined by measuring the absorbance of the sample at 260nm using a *Varian* spectrophotometer. The results are shown in Figure 4.2. A rough estimate of the relative amounts of ssDNA and dsDNA can be calculated by summation of the area under the peaks. Since the absorbance of a given concentration of dsDNA at this wavelength is only 80% of that of the same concentration of ssDNA, these figures have to be adjusted

FIGURE 4.2

Evaluation of the degree of renaturation attained in 'standard' phosphate buffer (4.2a) and ammonium sulphate buffer (4.2b). Details of the buffers are given in the text. Samples were denatured by boiling, and left to anneal at 68°C to a C_{ot} value of 2100. Samples were desalted using a column of Sephadex G-50-150 and then loaded onto a hydroxylapatite column. Figures 4.2a and 4.2b show the elution profile of single-stranded (fractions 1-10) and double-stranded (fractions 11-20) DNA from the hydroxylapatite columns. Single-stranded DNA was eluted using 0.12M sodium phosphate (pH6.9) and double-stranded DNA using 0.27M sodium phosphate (pH6.9). The amount of DNA present in each fraction was estimated by measuring the absorbance of a diluted aliquot at 260nm. Under these conditions the degree of renaturation was estimated to be $\approx 50\%$ in 'standard' phosphate buffer and $\approx 90\%$ in ammonium sulphate buffer.

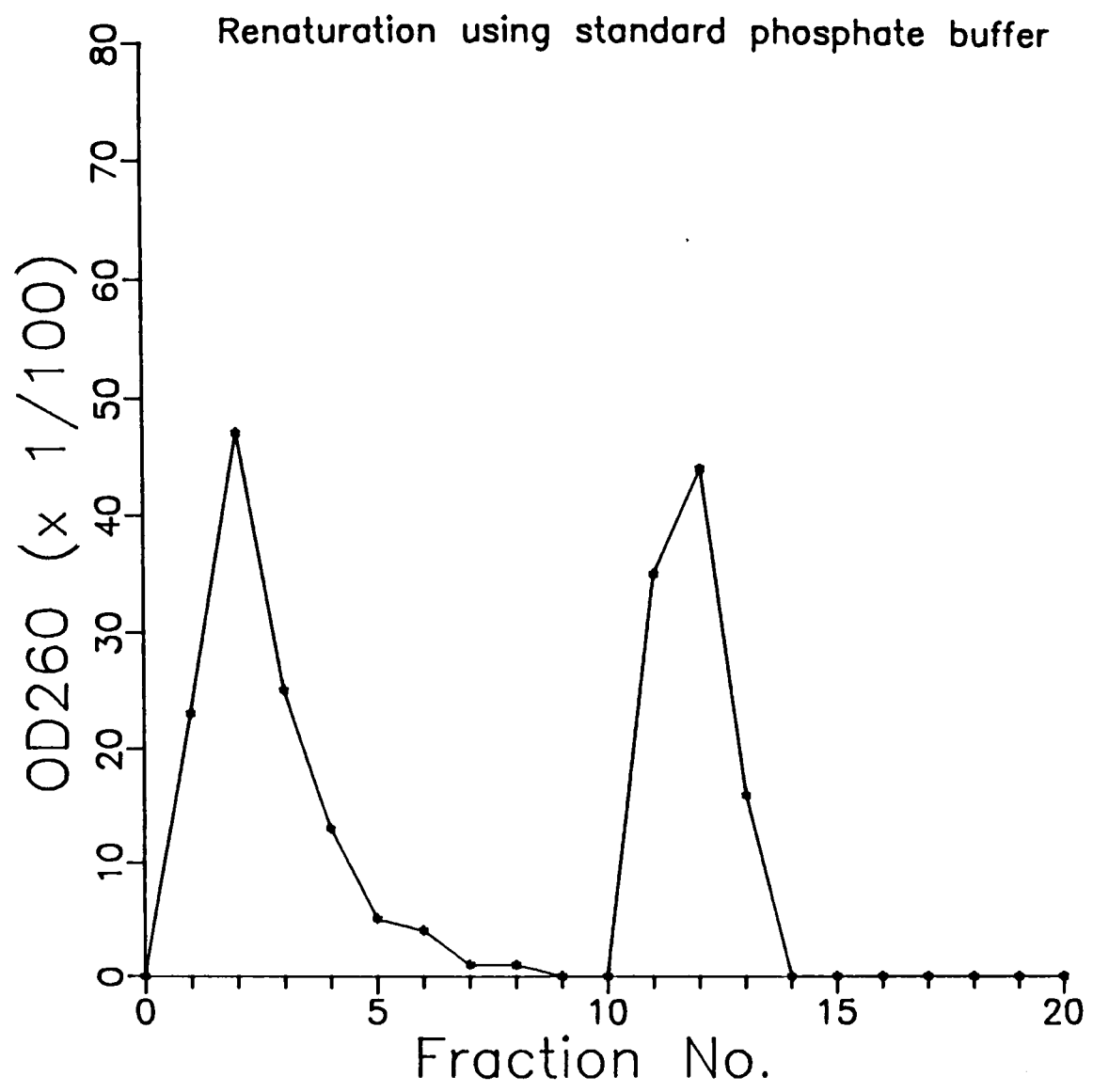
OD₂₆₀ readings used in 4.2a.

1. 0.023	11. 0.035
2. 0.047	12. 0.044
3. 0.025	13. 0.016
4. 0.013	14. 0.000
5. 0.005	15. 0.000
6. 0.004	16. 0.000
7. 0.001	17. 0.000
8. 0.001	18. 0.000
9. 0.000	19. 0.000
10. 0.000	20. 0.000

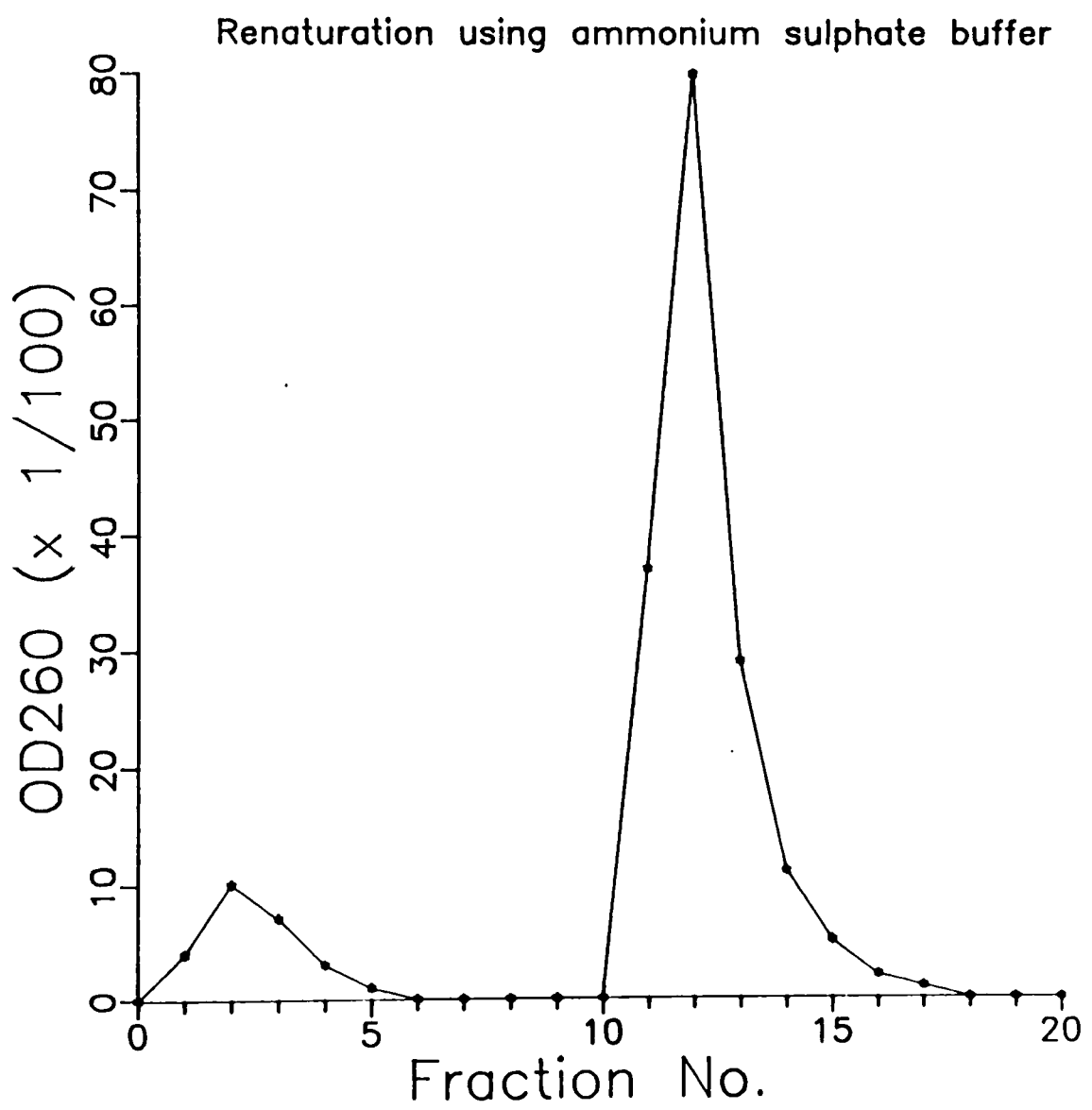
OD₂₆₀ readings used in 4.2b.

1. 0.004	11. 0.037
2. 0.010	12. 0.080
3. 0.007	13. 0.029
4. 0.003	14. 0.011
5. 0.001	15. 0.005
6. 0.000	16. 0.002
7. 0.000	17. 0.001
8. 0.000	18. 0.000
9. 0.000	19. 0.000
10. 0.000	20. 0.000

1-10 ssDNA (read using a 0.12M PO₄⁻ blank)
 11-20 dsDNA (read using a 0.27M PO₄⁻ blank)



4.2a



4.2b

accordingly. For 'standard' phosphate buffer, the relative amounts of ssDNA and dsDNA were found to be $\approx 50\%$ ssDNA, 50% dsDNA, whereas for ammonium sulphate buffer they were $\approx 10\%$ ssDNA and $\approx 90\%$ dsDNA. In subsequent attempts the level of renaturation has been as high as 96% . This double-stranded fraction was purified and attempts were made to obtain clones from it.

Ligations were set up using $1\mu\text{g}$ of dsDNA/ 100ng dephosphorylated vector as described (2.11.2) and used to transform competent *E.coli* HB101. The transformation efficiency using supercoiled pBR322 was $\approx 2 \times 10^5$ transformants/ μg . The number of AMP^R colonies obtained with the equivalent of 100ng religated vector alone was ≈ 30 . The experimental ligations yielded a total of 190 AMP^R colonies ($\approx 65/100\text{ng}$ vector). These colonies were transferred to TET plates to distinguish between recombinants (AMP^R /TET^S) and recircularised vector (AMP^R /TET^R). 57 (30%) were found to be recombinants. Recombinants were picked from the plates and their respective plasmids isolated. Since *Mbo*I fragments had been cloned into the *Bam*HI site of pBR322, it was not possible to release the insert by digestion of the plasmids with *Bam*HI because this site is no longer retained. To estimate insert size, plasmids were cut with *Eco*RI and *Sal*I which flank the insert. This fragment is normally 652 bp and so the size of the insert can be estimated from the increase in size of this fragment. Typical results are shown in Figure 4.3. Although this method for determination of the sizes of the inserts is not very accurate, the average size of the insert was found to be ≈ 140 bp, the largest one identified being ≈ 610 bp. The average size of these fragments

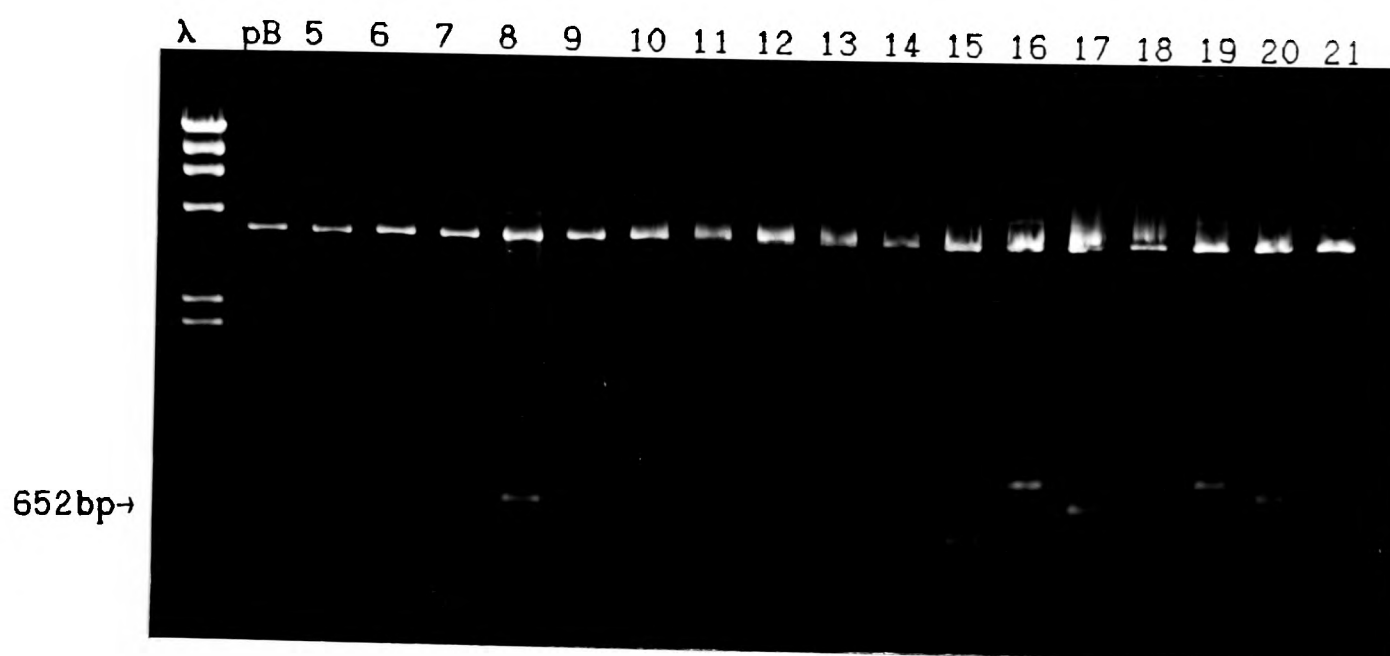


FIGURE 4.3

Determination of the sizes of deletion enrichment clones.

λ: λ *Hind*III marker fragments; pB: pBR322 digested with *Eco*RI and *Sal*I; 5-21: 'deletion enrichment' clones digested with *Eco*RI and *Sal*I. The sizes of the *Mbo*I inserts cloned into the *Bam*HI site of pBR322 can be estimated from the increase in the size of the *Eco*RI/*Sal*I fragment present in the vector. In the absence of an insert this is 652bp (see track pB). In most of the clones the mobility of this fragment has been retarded, due to the presence of an insert. In the cases of clone № 13 and № 15, this fragment has an increased mobility, either due to the presence of an *Eco*RI or *Sal*I site in the insert, or a deletion that has occurred during the cloning procedure. Molecular weights of the inserts :

Clone №	M.W. /bp.	Clone №	M.W. /bp.
5	65	14	135
6	125	15	?
7	115	16	175
8	95	17	85
9	75	18	50
10	90	19	170
11	70	20	120
12	70	21	95
13	?		

The average size of the inserts shown here is <100bp. In general the average was ≈140bp, the largest insert identified was 610bp.

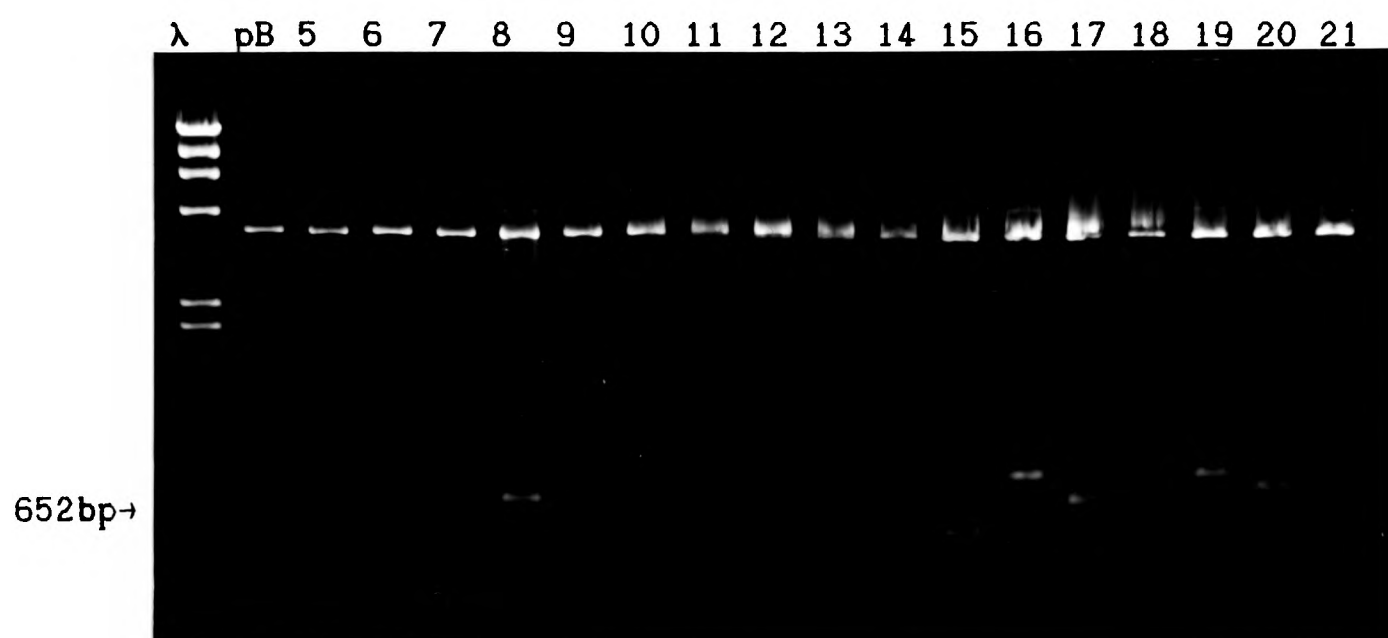


FIGURE 4.3
Determination of the sizes of deletion enrichment clones.

λ: λ *Hind*III marker fragments; pB: pBR322 digested with *Eco*RI and *Sal*I; 5-21: 'deletion enrichment' clones digested with *Eco*RI and *Sal*I. The sizes of the *Mbo*I inserts cloned into the *Bam*HI site of pBR322 can be estimated from the increase in the size of the *Eco*RI/*Sal*I fragment present in the vector. In the absence of an insert this is 652bp (see track pB). In most of the clones the mobility of this fragment has been retarded, due to the presence of an insert. In the cases of clone № 13 and № 15, this fragment has an increased mobility, either due to the presence of an *Eco*RI or *Sal*I site in the insert, or a deletion that has occurred during the cloning procedure.
Molecular weights of the inserts :

Clone №	M. W. /bp.	Clone №	M. W. /bp.
5	65	14	135
6	125	15	?
7	115	16	175
8	95	17	85
9	75	18	50
10	90	19	170
11	70	20	120
12	70	21	95
13	?		

The average size of the inserts shown here is <100bp. In general the average was ≈140bp, the largest insert identified was 610bp.

is somewhat less than would be predicted. The AT:GC ratio found in human DNA is 1.52 (60.3%) (reviewed in Lehninger, 1975). $p(A)$, the probability of finding an A at a given point, is:

$$p(A)=p(T)=0.302 \text{ and}$$

$$p(G)=p(C)=0.198$$

MboI recognises the sequence GATC, the probability of finding this particular sequence is :

$p(GATC)=(0.302)^2 \times (0.198)^2 = 3.599 \times 10^{-3}$. The average size of an *MboI* fragment is therefore $1/p(GATC)=278\text{bp}$. This suggests that there is some selection for smaller fragments during the enrichment procedure. Because of the small sizes of these inserts, the clones were difficult to analyse directly through use as probes, labelled by nick translation. Signals were very weak, if present at all, and autoradiographs had to be exposed for several weeks. Since this was not conducive to the rapid characterisation of the clones, it was decided to subclone some of the *EcoRI/SalI* fragments containing inserts into the vector mp18, so that more sensitive single-stranded probes could be prepared. Although some clones failed to give any discernable signals, using this technique it was usually possible to get a reasonable signal from clones as small as 70 bp within a few days. Over 20 clones were analysed in this way, some of the autographs are shown in Figure 4.4 and results summarised in Table 4.1. No Y-specific clones were identified.

Although a sufficiently large number clones were not analysed to say that there had been no enrichment for Y-specific material, any enrichment that had occurred could not be said to be dramatic. Since the deletions present in the hybrids WAG 8 and UCLA_{B2} are both smaller than the 'deletion'

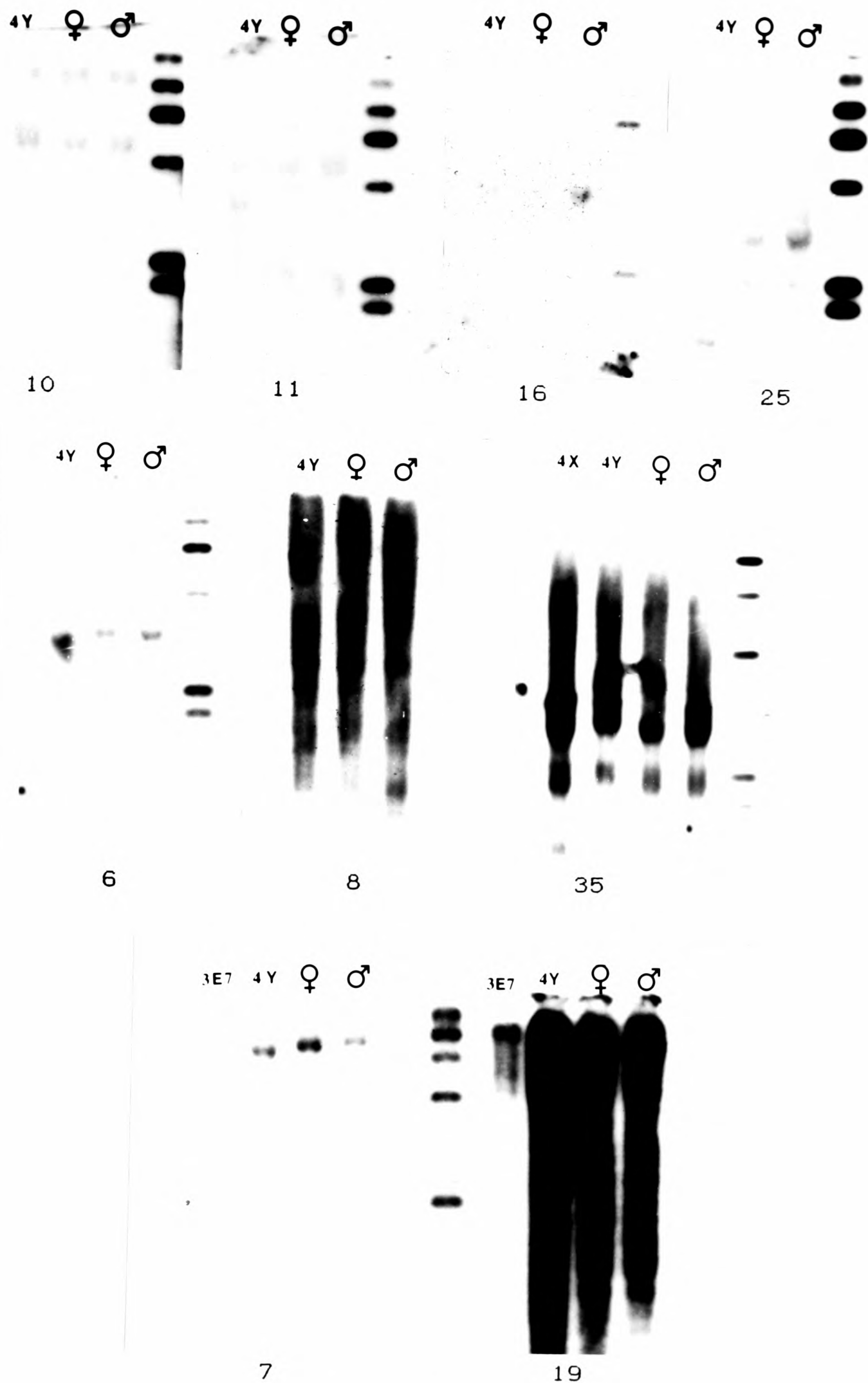


FIGURE 4.4
Analysis of 'deletion enrichment' clones.

♂: normal male; ♀: normal female; 4X: GM1416 (Four X chromosomes present); 4Y: OX (Four Y chromosomes present); 3E7: Y-only hybrid. All samples were digested with *EcoRI*. In many cases the track on the right contains λ /*HindIII* marker fragments. The filters were probed with single-stranded, 'deletion enrichment clones' under the conditions described in the text. All the clones detect autosomal sequences.

N^o of Clone	Size of Insert (bp)	Result of Hybridisation
4	610	Repetitive smear - No bands discernable
6	125	Single-copy band $\approx$ 3.5 kb
7	115	Single-copy band $\approx$ 8 kb
8	95	Repetitive smear - Many bands
10	90	Moderate repeat - Bands 4-15 kb
11	70	Moderate repeat - Bands 2-15 kb
12	70	Repetitive smear - No bands discernable
14	135	Repetitive smear - No bands discernable
16	175	Single-copy band $\approx$ 4 kb
17	85	No discernable signal
19	170	Repetitive smear - Many bands
20	120	Repetitive smear - No bands discernable
21	95	Faint single-copy band $\approx$ 15 kb
22	190	Repetitive smear - No bands discernable
23	105	Repetitive smear - No bands discernable
24	190	Repetitive smear - No bands discernable
25	155	Moderate repeat - Bands 1.5-3 kb
26	90	No discernable signal
34	60	No discernable signal
35	115	Moderate repeat - Bands 0.7-3.5 kb
38	190	Faint single-copy band $\approx$ 2 kb
40	225	Faint moderate repeat - Bands 4-10 kb

TABLE 4.1

Summary of results obtained with the hybridisation of ss 'deletion enrichment' clones. All clones had an apparent autosomal localisation.

represented by the Y chromosomes in the cell-line OX, these results suggested that it would be necessary to screen many more clones in order to isolate sequences mapping to terminal Xp, even then, because of the small size of the clones, if further analysis of this region was to be performed it would have been necessary to probe a genomic library (e.g. the MOG cosmid library) with these clones to isolate longer stretches of DNA. Reassociation experiments were nevertheless performed with both WAG 8 and UCLA₂ and clones were isolated. Again the average size of the inserts was <150bp; however, these were not analysed any further because of direct access to the region of interest, through the availability of cDNA and genomic clones for steroid sulphatase (see Chapter 5).

4.4 GENERAL DISCUSSION OF RESULTS

The two major problems with the deletion enrichment technique were a) the degree of enrichment for desired sequences was much less than apparently predicted by published reports and b) the size of clones isolated was small, which led to difficulties in their subsequent analysis. It is possible that the amount of driver DNA used (100 fold excess over tracer DNA) was insufficient, and that a larger excess of DNA would provide more efficient competition. In the cases where MOG DNA was used as the tracer and source of X chromosomal material, a lot of the 'background' clones isolated would be expected to be murine sequences. Since X chromosomal sequences represent only a small fraction of the total amount of DNA present. Smith *et al.* (1987) used inserts isolated from

a bulk preparation of an X-only library as their tracer for the isolation of sequences present in a deletion of Xp21.2. The only readily available source of purified X chromosomal material was the partial X library of 238 clones from the MOG cosmid library (discussed in Chapter 3). If these clones are representative of sequences dispersed throughout the X chromosome, then some of them would be expected to contain sequences mapping to Xp21→Xpter and Xp22.3→Xpter. In retrospect, it may have been worth using DNA from a bulk preparation of these clones as the source of tracer DNA, the inclusion of an excess of randomly sheared vector DNA in the reassociation mix would have avoided the need to isolate inserts. Although some sequences would have been over represented, due to the different growth characteristics of individual clones, this technique may have stood a better chance of success. A great advantage of this approach would have been that once a clone mapping to the desired region had been identified, it would have been necessary to screen only 238 clones in order to isolate more extensive regions of DNA in this area.

It is debatable as to how successful the other reported uses of deletion enrichment have actually been. Even where success has been achieved, this does not seem to have been reproducible. Kunkel *et al.* have tried to construct recombinant libraries using other, some larger, deletions with no success (W. Middlesworth, personal communication). The isolation of the *HIP25* locus was a unique event after many failed attempts (T. Smith, personal communication). However, in spite of the large size of the deletion used by both Smith *et al.* (1987) and

Kunkel *et al.* (1985), estimated to be $\approx 3000\text{kb}$, and the fact that the former used purified X-chromosomal material as the driver, it does seem that in these cases significant enrichment did occur. The 1% success rate reported by Cotinot *et al.* (1987), is less convincing.

A putative cDNA clone for steroid sulphatase became available at this stage. Since the steroid sulphatase locus had previously been shown to map to terminal Xp (e.g. Mohandas *et al.*, 1979; Tiepolo *et al.*, 1980; Curry *et al.*, 1984) it was decided to analyse this clone in some detail, rather than persevere with my own strategies for isolation sequences from this region.

CHAPTER 5

AN ANALYSIS OF THE HUMAN STEROID SULPHATASE LOCUS USING PUTATIVE cDNA AND GENOMIC CLONES

5.1 INTRODUCTION

The gene for the enzyme steroid sulphatase (STS) is of particular interest in the context of sex chromosome organisation and evolution as it is the only well characterised X-linked locus in humans which has no functional, Y-linked equivalent and yet escapes from inactivation. In the last few years, STS has been the target of many investigations in relation to steroid metabolism, to the genetics of the X chromosome and to clinical aspects of STS deficiency. In humans STS deficiency is the primary defect of X-linked ichthyosis (XLI) (Shapiro *et al.*, 1978), a skin disorder often associated with late parturition. The enzyme defect associated with XLI is not confined to placental tissue, in an investigation of more than 100 fibroblast clones from a female heterozygous for XLI, all of them were found to have substantial STS activity (Shapiro *et al.*, 1979); an observation in contrast to the expected pattern for X-inactivation, in which approximately equal numbers of positive and negative clones should result. This provided evidence that the *STS* locus escapes from inactivation. Non-inactivation in humans has been verified by showing expression of STS from isolated inactive X chromosomes in inter-specific cell hybrids (Mohandas *et al.*, 1980) and by demonstrating gene dosage with the number of X chromosomes (Shapiro, 1985). The female to male ratio for STS levels in

fibroblasts and placentae is an average value of 1.6 (Craig & Tolley, 1986), substantially less than that expected for a direct dosage relationship. An investigation by Migeon *et al.*, (1982) has indicated that STS expression from the inactive X is about half that from the active X, suggesting that partial inactivation occurs. This partial escape from inactivation is probably an intrinsic feature of the *STS* gene itself. Autosomal sequences juxtaposed to an inactive X can be inactivated (Mohandas *et al.*, 1982), while STS sequences, even when flanked on both sides by inactive X DNA, maintain their expression (Mohandas *et al.*, 1987). Since DNA methylation has been implicated in the maintenance of X inactivation (Mohandas *et al.*, 1981; Yen *et al.*, 1984; Wolf *et al.*, 1984; Toniolo *et al.*, 1984; Riggs *et al.*, 1985) the methylation status of the *STS* gene is of interest. Yen *et al.* (1987) using their clones isolated from the *STS* locus have reported no significant differences in methylation for the *STS* genes on active and inactive X chromosomes.

The *STS* locus in man has been mapped to the distal portion of the X chromosome short arm. This has been accomplished by studies in somatic cell hybrids bearing X/autosome translocations (Mohandas *et al.*, 1979) and through deletion mapping of individuals nullisomic for distal Xp as a result of X chromosome deletions (Curry *et al.*, 1984) or X/Y translocations (Tiepolo *et al.*, 1980). There is no functional gene on the human Y chromosome and a study of a series of X/Y translocations supports the mapping of *STS* to a position proximal to the *MIC2* locus and the human pseudoautosomal region (Geller *et al.*, 1986; Mondello *et al.*, 1987).

The behaviour of the *STS* locus in the context of other mammals is also of interest, since the rationale explaining the conservation of the X chromosome (Ohno, 1967) does not extend to genes which escape from inactivation. Although there is preliminary evidence that there is an absence of an X-linked allele for *STS* in Australian marsupials (reviewed in Graves, 1987), most studies have concentrated on wood-lemmings and mice. In the case of wood-lemmings, as in humans, the locus appears X-linked, but not subject to inactivation (Ropers & Wiberg, 1982). Mice also have an X-linked *STS* locus (Gartler & Rivest, 1983), but do not show sex differences in dosage and there is no evidence for loci which escape from inactivation on the mouse X chromosome (Crocker & Craig, 1983). This can be explained by there being in the mouse a Y-linked allele which brings the level of the enzyme in the male into a range similar to that in the female. Keitges *et al.* (1985) have been able to demonstrate, in different types of crosses, that *STS* deficiency could behave either as an X-linked or as an autosomal recessive trait and have explained these observations by postulating that both the X and Y chromosomes possess functional *STS* alleles which undergo obligatory recombination during male meiosis. Further support for a pseudoautosomal localisation is suggested by its linkage to the most distal X chromosome marker, *Crm*, and by its cosegregation with *Sxr* in the appropriate backcrosses involving sex reversed mice (reviewed in Craig *et al.*, 1987). Steroid sulphatase levels have been measured in the root vole (*Microtus oeconomus*) (Wiberg *et al.*, 1987) and the results have shown that levels are significantly higher in males than females (1.6:1). It is suggested that in the root vole, as in

the mouse, the *STS* gene(s) is X- and Y- linked, and that it is subject to partial X-inactivation on one of the X chromosomes in females.

Placental STS has been purified from microsomal fractions by treatment with detergents (Shapiro *et al.*, 1985). The apparent M.W. of the major polypeptide observed is ≈ 60 kd. Antibodies raised against the purified protein have been used to demonstrate the lack of immunologically cross-reacting material in X-linked ichthyotics (Epstein & Bonifas, 1985). cDNA sequences have been isolated by screening a λ gt11 expression library prepared from human placental mRNA with poly-clonal anti-STS antibodies (van der Loos *et al.*, 1984). One of these clones, p422 (M.W. 560 bp), has been used to isolate a longer cDNA clone from the same library. This clone, P2A7 (M.W. 2.7 kb), has in turn been used to isolate genomic sequences and together these have allowed the beginning of a detailed molecular analysis of this region.

In this chapter, I shall describe the use of some of the clones isolated by Ballabio *et al.* to map sequences corresponding to the *STS* locus to the distal short arm of the human X chromosome and provide evidence of a non-functional allele of *STS* on the long arm of the Y chromosome. The presence of STS homologous sequences in other mammals is also investigated.

Some of the work presented in this chapter has already been published (Ballabio *et al.*, 1987; Fraser *et al.* 1987a,b). Since this work was undertaken Yen *et al.* (1987a) have also published results describing the analysis of STS clones.

5.2 MATERIALS AND METHODS

Cell-lines and hybrids were grown as described in 2.3. DNA was isolated using the methods described in 2.4. Primate blood samples were a kind gift from R. Lovell-Badge and Prof. John Hearn. Many of these samples had clotted, or the cells lysed, so it was not possible to separate just the white cells for DNA preparation. The whole sample was treated with lysis buffer and nuclei were isolated as described in 2.4.2.

DNA from the hybrids 817/175 and 445/393 was a gift from Dr. P.N. Goodfellow. DNA from the hybrid 853 and the cell-line TAP was a gift from M. Ross.

All genomic DNA samples were digested overnight with at least a two-fold excess of enzyme, under the conditions recommended by the supplier. For 10 μ g of DNA, electrophoresis was carried out overnight at 35mA/40V. For 2.5 μ g of DNA electrophoresis was for 6hrs at 100mA/80V. Filters were prepared by blotting gels onto HYBOND-N (2.7.2). Conditions for hybridisation and washing of filters were as in 2.10. Modifications are described in the text.

The clones p422, P2A7, B2RI2 and B12RI5 were all gifts from Dr. A. Ballabio and Dr. M.G Persico.

Preparation of insert DNA by elution onto DEAE paper is described in 2.7.3. All inserts and plasmids were labelled with ^{32} P- α -dCTP by nick translation (2.8.1). Routinely, specific activities of $\approx 10^8$ cpm/ μ g were achieved.

5.3 THE SUB-REGIONAL ASSIGNMENT OF cDNA CLONES FOR STS

5.3.1 An analysis of the clone p422

Anti-STS polyclonal antibodies have been used for the immunoscreening of a λ gt11 expression library prepared from placental mRNA. Among 10⁵ plaque-forming units screened, 6 positive cDNA clones were isolated and purified (Ballabio et al., 1987). Preliminary analysis indicated that only one of

these, p422, was localised to the short arm of the X chromosome (A. Ballabio *et al.*, unpublished observations). This clone contains a 560bp cDNA insert (Figure 5.1a), which was subsequently subcloned into the vector pUC18. Only ≈ 200 bp of this insert actually appears to contain cDNA sequences for STS, the other sequences are of unknown origin (A. Ballabio, personal communication).

The localisation of the genomic sequences recognised by p422 was established by its hybridisation to restriction digests of DNAs from somatic cell hybrids with a variable representation of human sex chromosomes and from human cell lines with 1, 2 and 4 X chromosomes (Table 5.1). 10 μ g of DNA from the different cell lines and hybrids was digested with *Bam*HI and a mapping filter prepared. Insert was isolated following *Eco*RI digestion of the plasmid and was labelled by nick translation to a specific activity of $\approx 10^8$ cpm/ μ g. Hybridisation was carried out overnight at 64°C in the presence of dextran sulphate. In the first instance the filter was washed to high stringency (0.2xSSC/0.1% SDS, 64°C). The result is shown in Figure 5.2. In the dosage experiments, the decreasing signal, observed in lanes loaded with the same amount of DNA from 48,XXXX, 46,XX and 46,XY cell lines indicates a clear localisation to the X chromosome. There is no evidence for a Y-localised sequence, even when hybridisation and washing was repeated at a lower stringency (hybridisation: 60°C, washing: 2xSSC/0.1% SDS room temp.), no other bands appeared in any of the tracks (results not shown). The most precise information is provided by comparison of the hybridisation to the DNA of two hybrids that retain human X chromosomes with

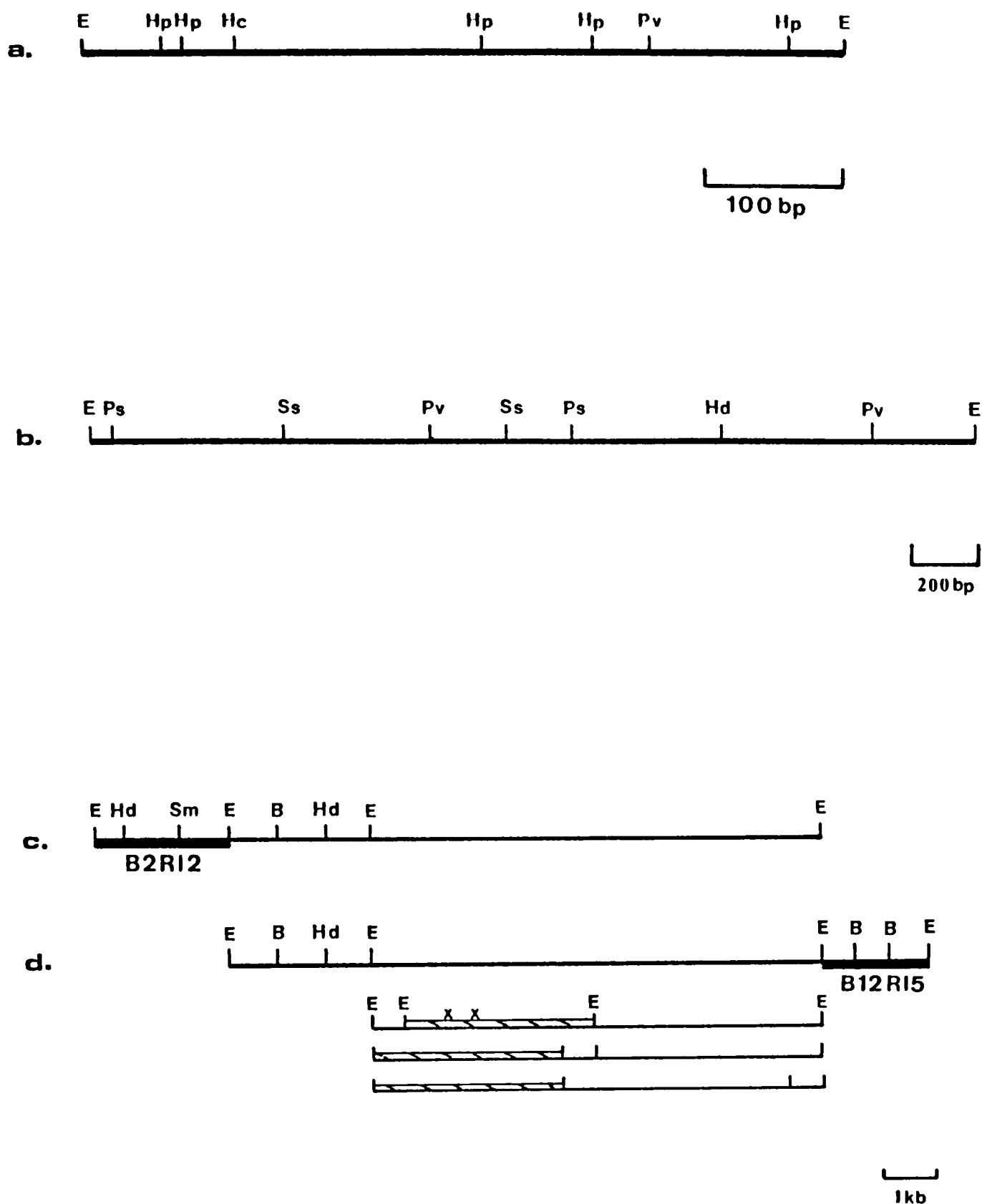


FIGURE 5.1
Restriction maps of the clones used
in the analysis of the *STS* locus.

B: *Bam*HI; E: *Eco*RI; Hc: *Hinc*II; Hd: *Hind*III; Hp: *Hpa*II;
 Ps: *Pst*I; Pv: *Pvu*II; Sm: *Sma*I; Ss: *Sst*I.

a.) p422 (cDNA)

b.) P2A7 (cDNA)

c.) λ B2 (genomic) - The subclone B2RI2 was used in this investigation.

d.) λ B12 (genomic) - The subclone B12RI5 was used in this investigation.

λ B2 and λ B12 overlap. The orientation of the *Eco*RI sites within the 7.2kb *Eco*RI fragment of these clones is unsure. The hatched area represents the fragment to which P2A7 hybridises. Two possible sites for p422 are marked (x).

(These restriction maps were provided by A. Ballabio.)

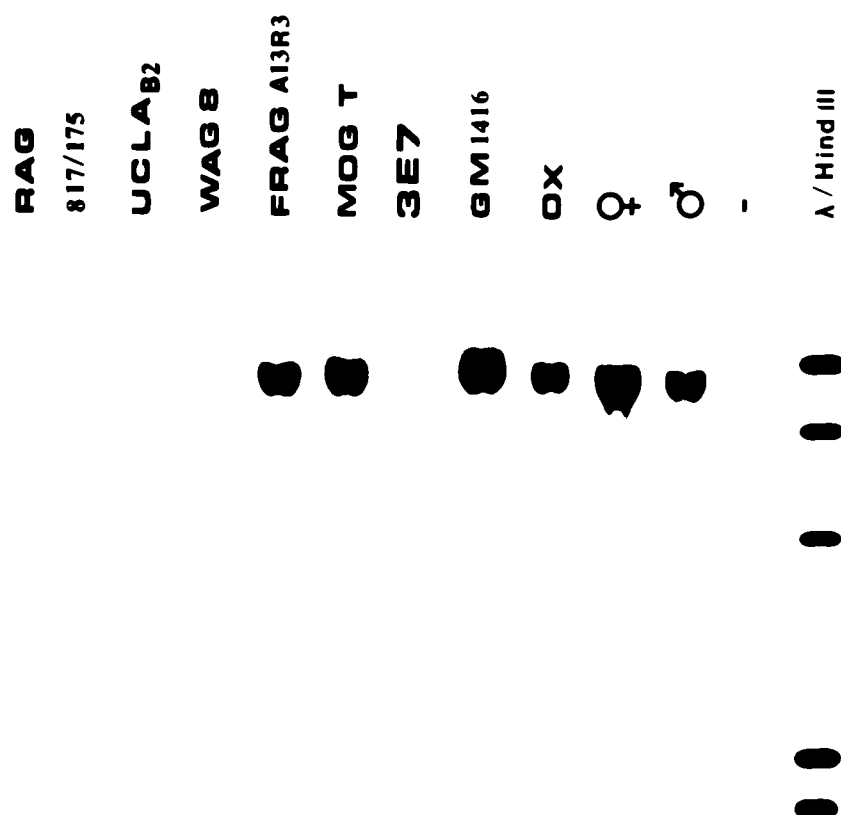


FIGURE 5.2
Sub-regional localisation of p422.

Filter prepared from *Bam*HI digested DNA samples of various hybrids and cell-lines and probed with p422. The type of cell-line and content of human sex chromosomes are indicated in Table 5.1. The pattern of hybridisation is consistent with X-linkage and a sub-regional localisation to Xp22.3→Xpter.

CELL-LINE	DESCRIPTION	HUMAN AUTOSOMES	HUMAN SEX CHROMOSOMES
RAG	MOUSE	None	None
817/175	HYBRID	Several	Xp22.3-Xqter Yq11-Yqter
UCLA _{B2}	HYBRID	Several	Xp22.3-Xqter
WAG 8	HYBRID	Several	Xp21.1-Xqter
FRAG A13R3	HYBRID	Several	Xpter-Xp21.2
MOG T	HYBRID	None	X
3E7	HYBRID	None	Y
GM1416	HUMAN	All	48,XXXX
OX	HUMAN	All	49,XYYYYY
NORMAL FEMALE	HUMAN	All	XX
NORMAL MALE	HUMAN	All	XY

TABLE 5.1

Hybrid 817/175 is derived from a human cell-line 46,X,t(X;Y)(Xqter-Xp22.3::Yq11-Yqter) (Ropers *et al.*, 1985). UCLA_{B2} retains the chromosome associated with chondrodysplasia punctata (Curry *et al.*, 1984). See Boyd *et al.* (1987) and Darling *et al.* (1986) for details of other cell-lines and hybrids.

breakpoints in Xp22.3. One hybrid (817/175) expresses human levels of STS activity but lacks the *MIC2* locus, which has been assigned to Xp22.3 (Mondello *et al.*, 1987), and the other (UCLA_{B2}) lacks both STS activity and *MIC2*. (Curry *et al.*, 1984). The observation of a positive signal in the former hybrid and not the latter provides a clear localisation of the sequences recognised by p422 to Xp22.3.

Independent confirmation of this localisation has been provided by the results of *in situ* hybridisation of this probe to replication-banded metaphase chromosome spreads derived from lymphocytes of a normal male (Ballabio *et al.*, 1987).

p422 has been used by Ballabio *et al.* to isolate other cross-hybridising sequences (unpublished results). Two of these, T17 a clone isolated from the same library and COS1A isolated from a cosmid library, were also analysed by hybridisation to the same mapping filter as used above. Even under conditions of reduced stringency (hybridisation: 60°C, washing: 2xSSC/0.1% SDS room temp.), these clones did not hybridise to any of the tracks (results not shown), suggesting that these clones contained inserts not of human origin and had presumably been detected by the sequences of unknown origin in p422. These clones were not analysed any further.

5.3.2 An analysis of the clone P2A7

p422 has been used successfully to isolate a cDNA clone containing more extensive regions representing the *STS* locus. This clone P2A7 contained a 2.7kb insert, the restriction map for which is shown in Figure 5.1b. As with p422 this insert has been subcloned into pUC18 for ease of handling and was used in

subsequent investigations.

Plasmid containing P2A7 was labelled by nick translation to a specific activity of $\approx 0.5 \times 10^8$ cpm/ μ g and was used to probe the same mapping filter as was used for the sub-regional localisation of p422. Conditions of hybridisation and washing were the same as for p422. The result is shown in Figure 5.3. P2A7 recognises four *Bam*HI restriction fragments in female DNA and an additional two fragments in male DNA. The pattern of hybridisation of the four fragments present in female DNA is consistent with them all being localised to the region Xp22.3→Xpter. The two fragments present in male DNA are also present in the Y-only hybrid 3E7. These fragments show increased dosage in the cell-line OX (49, XYYYY). On prolonged exposure of the autoradiograph, these bands are also present in the hybrid 817/175. 817/175 is derived from a human cell-line 46,X,t(X;Y)(Xqter-Xp22.3::Yq11-Yqter) (also see Figure 5.4). Together, these findings suggest that the two additional fragments detected in these tracks are present on the long arm of the Y chromosome. Under these conditions, the probe does not detect any homologous sequences present on the autosomes. An *Eco*RI digest of DNA from the hybrid 445/393 was also probed with P2A7. This hybrid also contains an X/Y translocation chromosome (46,X,t(X;Y)(Xqter-Xp22.3::Yq11-Yqter), but does not express STS (Mondello *et al.*, 1987). In this case only the Y-specific fragments can be seen, confirming that Y-specific *STS* sequences are present on Yq (Figure 5.4).

The assignment of the Y-specific bands to Yq was refined through analysis of the cell-lines TRAN and TAP. Both of these cell-lines contain Y chromosomes that have extensive terminal

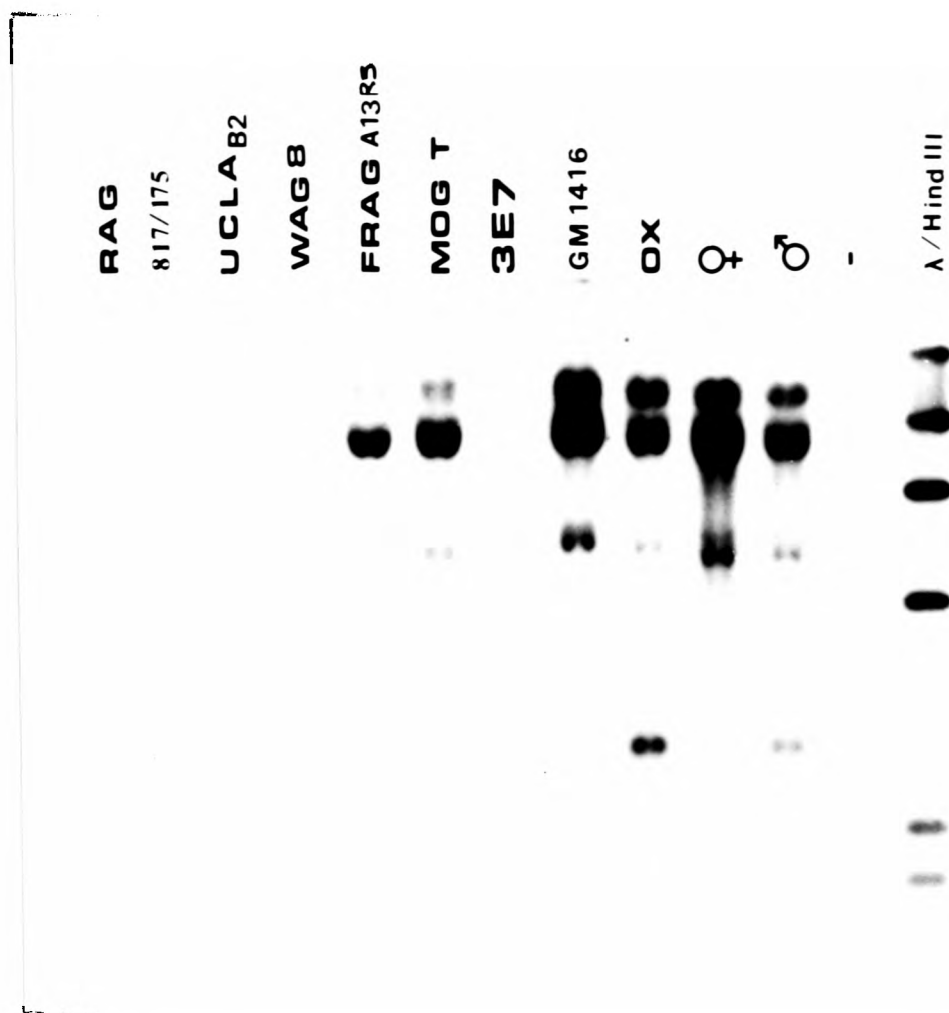


FIGURE 5.3
Sub-regional localisation of P2A7

The same mapping filter shown in Figure 5.2, this time probed with P2A7. In addition to several X-specific fragments, all of which map to Xp22.3→Xpter, Y-specific fragments are visible at $\approx 15\text{kb}$ and $\approx 3\text{kb}$.

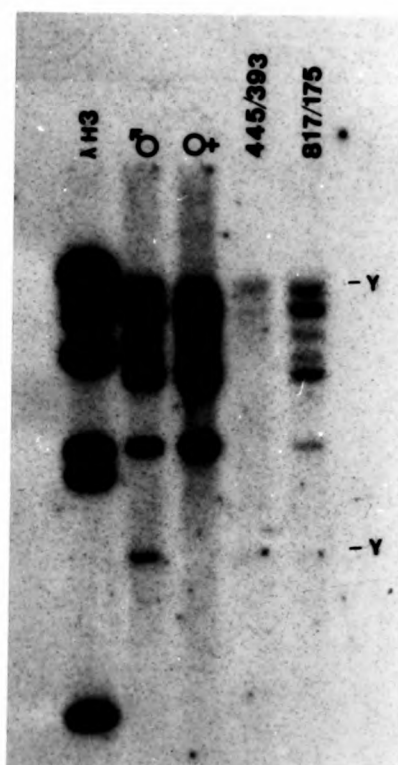


FIGURE 5.4
Localisation of STS homologous sequences to Yq.

EcoRI digests of DNA from the hybrids 817/175 and 445/393 probed with P2A7. Both of these contain X-Y translocation chromosomes (46,X,t(X;Y)(Xqter-Xp22.3::Yq11-Yqter). 817/175 expresses STS whereas 445/393 does not. X-specific fragments are present only in 817/175. The photograph is necessarily overexposed to show the Y-specific fragments ($\approx 21\text{kb}$ and 1.5kb) present in both hybrids. More convincing evidence of the localisation of STS homologous sequences to Yq has been obtained by *in situ* hybridisation (Fraser et al., 1987b).

deletions of Yq (unpublished information). Both TAP and TRAN have been shown to lack the locus *DYS20* (detected by the probe 69/6, Figure 5.5b), which has been mapped to Yq11.2 by *in situ* hybridisation (Buckle *et al.*, 1987), suggesting that the deletions of Yq in both cell lines must occur proximal to this band, unless complex rearrangements have taken place. The presence of Y-specific bands in TRAN and their absence in TAP (Figure 5.5a), maps *STS* homologous sequences proximal to the locus *DYS20*, but distal to the breakpoint in TAP.

The result of *in situ* hybridisation of this probe to replication-banded metaphase chromosome spreads derived from lymphocytes of a normal male (performed by V. Buckle) confirms its localisation to Xp22.3 and Yq11.2 (Fraser *et al.*, 1987b).

P2A7 can be cleaved by *PvuII*, to give three fragments of M.W. 0.4, 1.0 and 1.3kb. The orientation of the clone has been determined by comparing the restriction map of P2A7, with that of the *STS* cDNA isolated by Yen *et al.* (1987a) (see Figure 5.6). Thus 5' sequences are detected by the 1kb *EcoRI/PvuII* fragment. The 1.0kb and 1.3kb fragments were isolated and used as 5' and 3' probes. *BamHI* and *EcoRI* digests of male, female, 4X, 4Y, 3E7, and MOG DNA were probed with the two fragments. The results are shown in Figures 5.7a and 5.7b. The 5' probe recognises all of the fragments, including both Y-specific fragments, detected by the intact clone. The 3' probe recognises only one, this fragment is the same size as the fragment detected by p422. This suggests that the majority of the exons recognised by P2A7 are present in the 5' end of the cDNA and that p422 is part of a large 3' exon. The probing of the *EcoRI* digests allows a comparison to be made between the

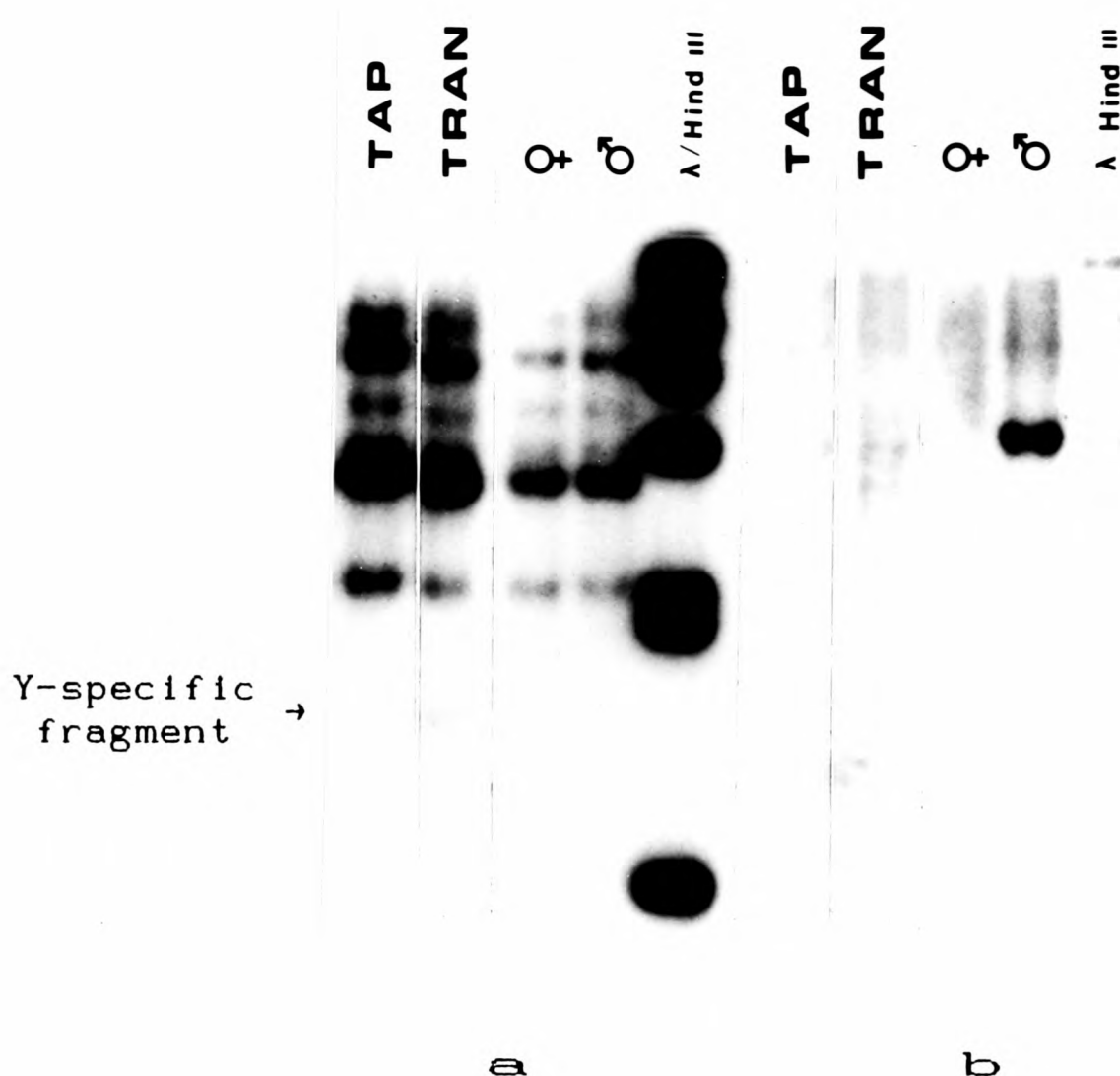


FIGURE 5.5
Localisation of STS homologous sequences to Yq11.2.

*Eco*RI digest of DNA from the cell-lines TAP and TRAN probed with P2A7 (a) and 69/6 (b). Both cell-lines contain Y chromosomes with extensive terminal deletions of Yq. Both cell-lines lack the locus detected by 69/6 (mapped to Yq11.2 by *in situ* hybridisation - Buckle et al., 1987), suggesting that the deletions of Yq in both cell-lines must occur proximal to this band. The presence of Y-specific bands detected by P2A7 (a) in TRAN and their absence in TAP, maps STS homologous sequences proximal to 69/6, but distal to the breakpoint in TAP.

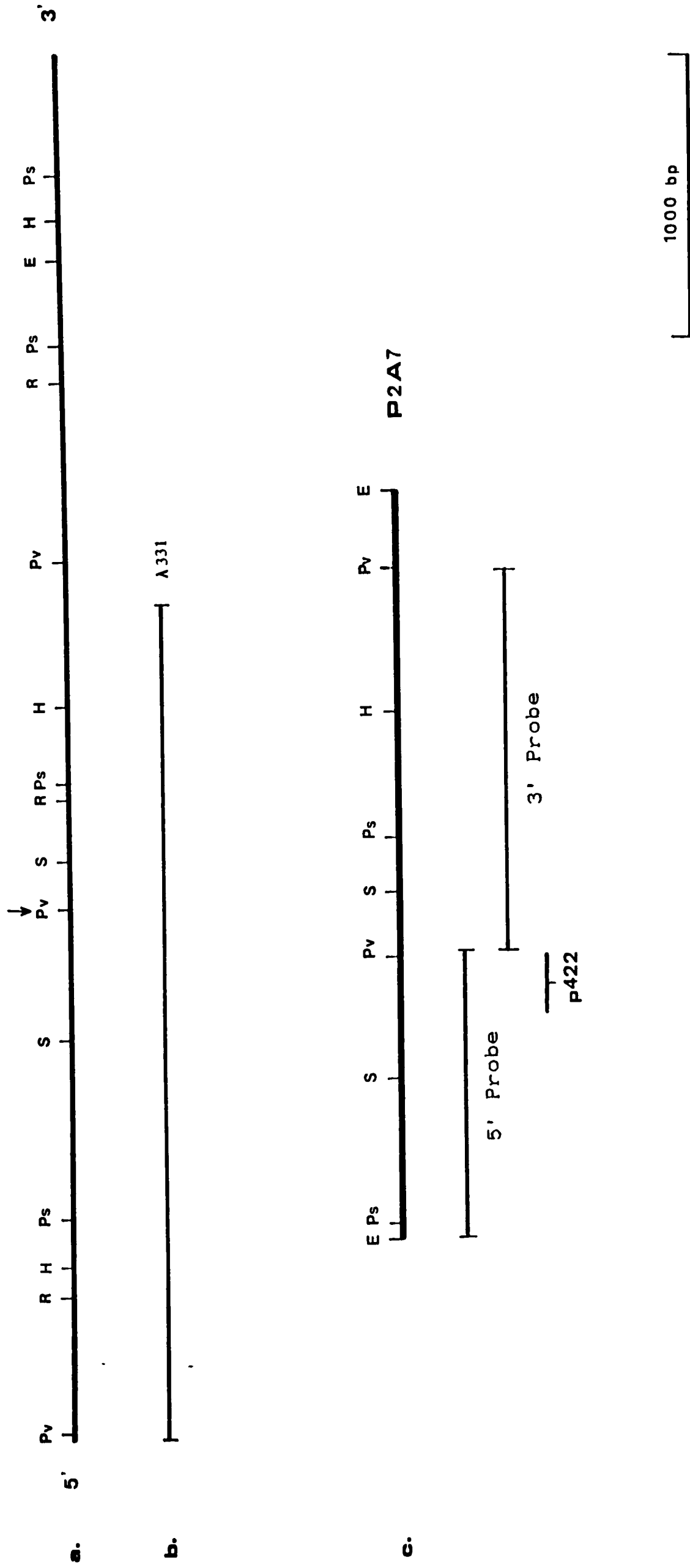
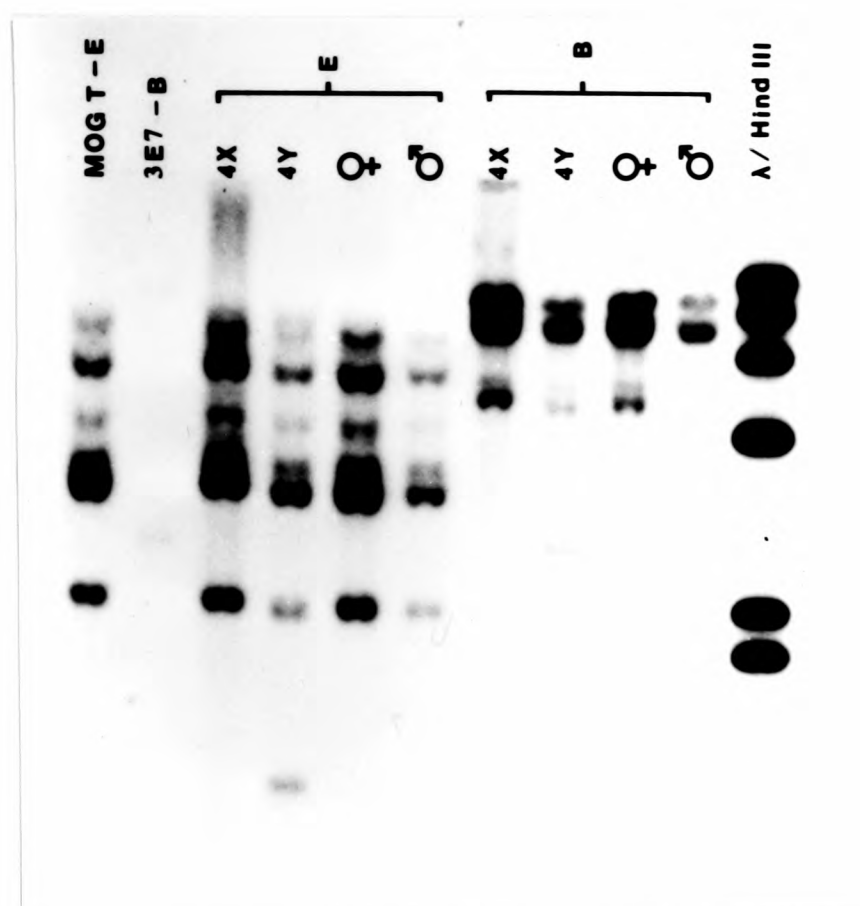


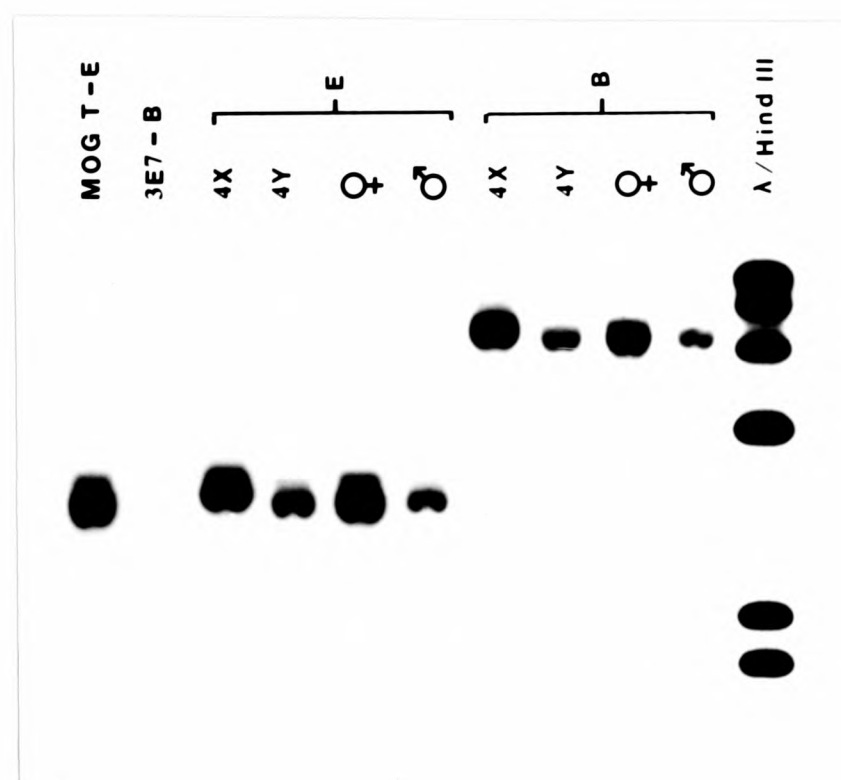
FIGURE 5.6
Restriction maps of STS cDNA clones.

E: *EcoRI*; H: *HindIII*; Ps: *PstI*; Pv: *PvuII*; R: *RsaI*; S: *SstI*.

- Presumptive complete map of the STS cDNA (from Yen et al., 1987a). Sequences 3' to the *PvuII* site in the middle of the map (↓) are colinear with genomic sequences.
- cDNA clone λ331 used in the analysis of the STS locus by Yen et al..
- cDNA clone P2A7 used in this investigation. The fragments referred to as the 5' and 3' probes are shown. The region corresponding to the cDNA clone p422 is also shown.



a



b

FIGURE 5.7

Analysis of the 5' & 3' probes derived from P2A7

4X: GM1416; 4Y: OX; 3E7: Y-only hybrid; MOG T: X-only hybrid. *Bam*HI (B) and *Eco*RI (E) digests of DNA from various hybrids and cell-lines probe with the P2A7 5' probe (a) and 3' probe (b). The 5' probe detects all of the fragments, including both Y-specific fragments. The 3' probe detects only one; this fragment is the same size as the fragment detected by p422.

fragments detected by P2A7 and those detected by the cDNA of Yen *et al.* (1987a). Their probe detects *EcoRI* fragments of M.W. 15, 10, 9.0, 6.1, 4.6, 4.2, 3.1 and 2.6kb in female DNA and additional fragments of 21 & 1.5kb in male DNA. P2A7 detects all of these fragments, with the possible exception of the X specific fragment of M.W. 10kb, although the precise size of the fragments is arguably different. The 3.1kb fragment is faint, but can be seen better in Figure 5.12a. Comparison of the restriction map of P2A7 with that postulated by Yen *et al.* shows that their clone used for probing genomic DNA (λ 331) is more extensive in the 5' direction, but less so in the 3' direction than P2A7 (Figure 5.6). It is possible that the clone λ 331 detects an additional exon. They suggest that sequences 3' to the *PvuII* site in the middle of the map ($\downarrow$) are colinear with genomic sequences, in agreement with the hypothesis that the majority of the exons are present in the 5' end of the cDNA. They also find that only sequences present in the 5' region of the cDNA detect homologous sequences on the Y chromosome.

5.3.3 An analysis of the genomic clones B2RI12 and B12RI5

p422 has been used to isolate sequences from a genomic library (Ballabio *et al.*, unpublished information). pUC18 subclones of two of these, B2RI2 (from λ B2) and B12RI5 (from λ B12), have been used in the further analysis of this region. Restriction maps of the phage and the regions corresponding to B2RI2 and B12RI5 are shown in Figures 5.1c and 5.1d. The orientation of the two clones has been determined by probing restriction digests of the phage inserts with both p422 and the

3' *HindIII/EcoRI* fragment of P2A7 (A. Ballabio, personal communication). Thus, B2RI2 detects sequences in the 5' region of the gene and B12RI5 detects sequences 3' to the cDNA sequences. Inserts were isolated from these subclones and were used to probe *BamHI* and *EcoRI* digests of DNA from normal ♂, normal ♀, 4Y, 4X, Y-only and X-only cell-lines. Labelling of the inserts by nick translation, preparation of filters and conditions of hybridisation were as described above. Washing was carried out to a stringency of 1xSSC/0.1% SDS at 64°C. The results are shown in Figures 5.8a and 5.8b.

B2RI2 hybridises to single *BamHI* (≈11kb) and *EcoRI* (≈2.4kb) fragments. Since these fragments are not detected by P2A7 and because B2RI2 maps 5' to a region of the phage insert cross-hybridising with the cDNA, it would seem that they are either present in one of the 5' introns, or that they represent sequences in an additional exon not detected by P2A7. The possibility that B2RI2 maps outside the region of the gene is unlikely. Sequences 3' to the *PvuII* site adjacent to p422 have been shown by Yen *et al.* (1987) to be colinear with genomic sequences. B12RI5 maps 3' to this site and also ≈12kb 3' to B2RI2. The size of the *STS* gene has been predicted to be >120kb (Yen *et al.*, 1987b), B2RI2 can not map 5' to the gene and therefore presumably lies within an intron or in an additional exon not detected by P2A7. Since the presumed, complete cDNA of Yen *et al.* (1987a) does not detect an additional fragment of this size, the former suggestion seems the more likely. It is interesting to note that B2RI2 does not detect any Y-specific fragments, whereas sequences in the 5' region of P2A7 do. There are three possible explanations for this: a) the sequences

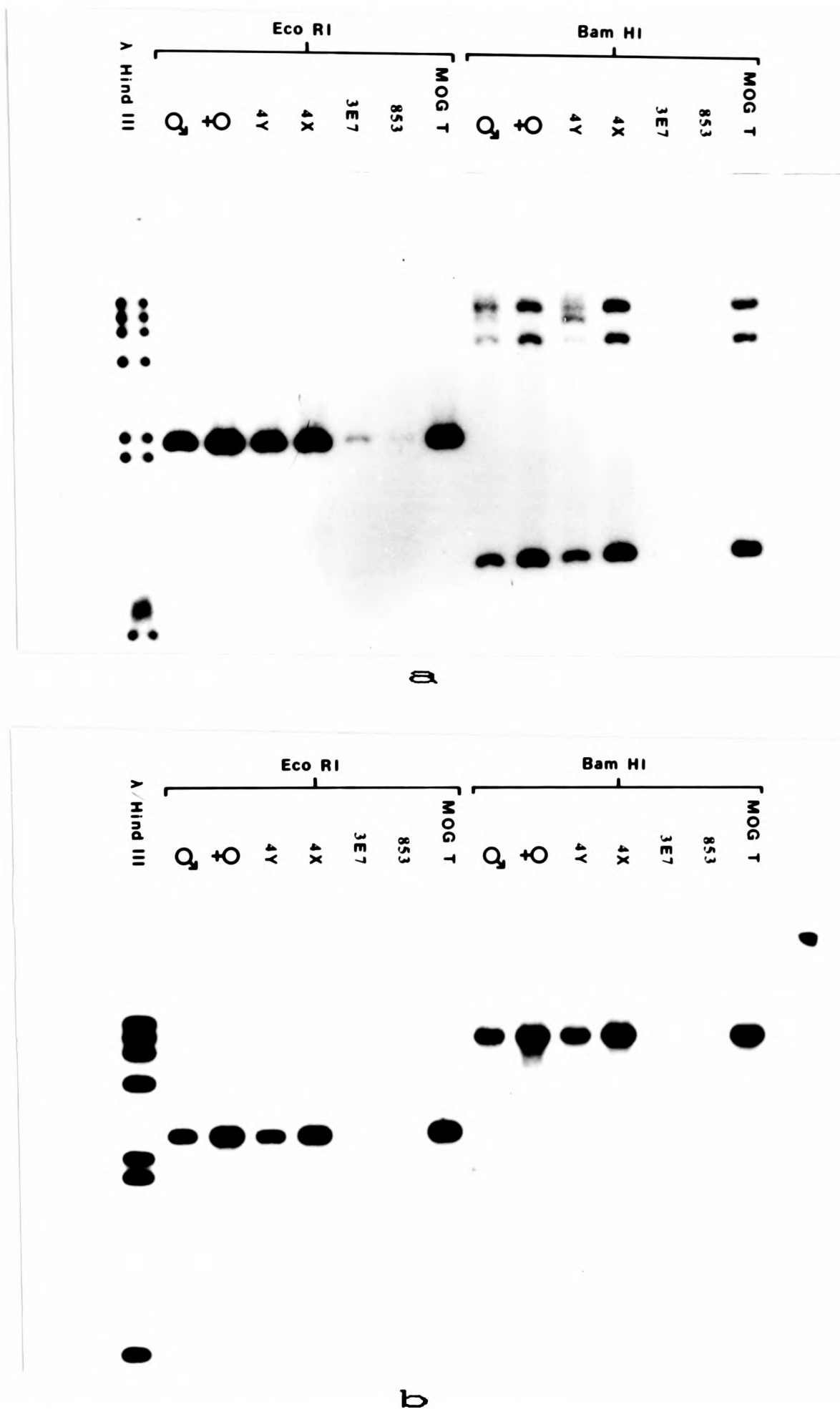


FIGURE 5.8
Analysis of the *STS* genomic clones B2RI2 & B12RI5

4X: GM1416; 4Y: OX; 3E7: Y-only hybrid; 853: Y-only hybrid, MOG T: X-only hybrid.

*Eco*RI and *Bam*HI digests of cell-lines and hybrids probed with the genomic clones B2RI2 (a) and B12RI5 (b). B2RI2 detects sequences in the 5' region of the gene and B12RI5 detects sequences 3' to the cDNA clones. No Y-homology is detected by B2RI2, however B12RI5 does detect Y-homology, which must map 3' to the coding sequences. Y-specific bands are not visible in the *Bam*HI digest of 853, due to there being much less DNA in this track.

present in P2A7 that detect Y homology map further 5' than B2RI2, b) the Y homology does not extend as far 5' as this intron, c) the regions of Y homology detected by P2A7 contain deletions. Yen *et al.* (1987b) have data that support the third hypothesis. They have isolated a group of overlapping clones that represent *STS* homologous sequences located on the long arm of the Y chromosome. It consists of two regions separated by 15kb which are homologous to the 5' end of the gene. Homology between the X and Y copies was found both within the sequences hybridising with the cDNA and within the introns. The conservation between the X and Y copies has been estimated to be as high as 92%, the majority of X/Y differences being the result of small deletions in the Y copy.

B12RI5 detects three X-specific *Bam*HI fragments (M.W. $\approx$ 12, 6.7 and 1kb) and one X-specific *Eco*RI fragment (M.W. $\approx$ 2 kb). The $\approx$ 12kb *Bam*HI fragment is the same fragment detected by the P2A7 3' probe and p422, as would be expected from the restriction maps, given the colinearity of cDNA and genomic sequences from around p422. The position on the map of λ B12 of the fragment hybridising to p422 relative to 5' restriction sites, suggests that there must be a splice site quite close to the 5' end of p422, since it is not possible for much of the sequence 5' to p422 to remain colinear with the genomic sequences, and sites 5' to p422 in the genomic clones are not found 5' to p422 in the cDNA. The $\approx$ 1kb *Bam*HI fragment is the same size as that predicted from the map, and the $\approx$ 7kb *Bam*HI fragment detected is presumably the fragment that extends further in the 3' direction. In addition to these X-specific fragments, B12RI5 detects sequences present on the Y

chromosome. These are shown as the $\approx 2\text{kb}$ *EcoRI* fragment visible in the tracks containing DNA from the Y-only hybrids 3E7 and 853 and the $\approx 30\text{kb}$ & $\approx 9\text{kb}$ *BamHI* fragments present in the σ , 4Y and Y-only tracks. The observation of the Y homology detected by this clone is intriguing, since previously only sequences present in the 5' region of the cDNA had been shown to detect sequences present on the Y chromosome. Yen *et al.* (1987b) have reported the isolation of 60kb of sequence corresponding to *STS* homologous sequences from the Y chromosome. The homology has been shown to correspond to exons 2-6 of the *STS* gene (the functional X-linked copy of the gene is thought to be composed of 10 exons), but they have not found Y homologous sequences mapping to the 3' end of the gene. B12RI5 maps 7-8kb 3' to p422 (the ambiguity being due to the uncertainty in the orientation of the *EcoRI* sites shown in Figures 5.1c and 5.1d), whereas sequences present in the presumptive, almost complete, map of the *STS* cDNA continue for only $\approx 3\text{kb}$ 3' to p422 (see Figure 5.6), suggesting that B12RI5 is detecting genomic sequences 3' to the *STS* 'gene' itself, which are also present on the Y chromosome. This observation, along with the identification by Yen *et al.* of sequences corresponding to the introns on the Y chromosome, implies that it is unlikely that the Y homology has resulted from the viral retroposition of *STS* mRNA sequences.

5.4 SEQUENCE CONSERVATION OF *STS* HOMOLOGOUS SEQUENCES DURING EVOLUTION

5.4.1 A search for *STS* homologous sequences in the mouse and rabbit

An investigation was undertaken to see whether *STS* homologous sequences could be identified in other species. In the first instance, *EcoRI* digests of DNA from mouse and rabbit livers was analysed. Filters were prepared as described above, using $\approx 10\mu\text{g}$ of DNA from a σ and ϕ mouse, rabbit and human controls. These were probed with insert from p422. Hybridisation was carried out at 60°C in the presence of dextran sulphate and washing was carried out at 60°C at a low stringency of $2\times\text{SSC}/0.1\%\text{SDS}$. The result is shown in Figure 5.9. No signal was detected in either the mouse or rabbit tracks even after prolonged exposure (not shown). Although conditions of stringency used with p422 could have been lowered, the availability of the longer P2A7 clone meant that subsequent investigations were performed using this clone. A low stringency hybridisation (55°C) was performed using P2A7 as the probe. This filter was then washed at room temperature with $2\times\text{SSC}/0.1\%\text{SDS}$. The result is shown in Figure 5.10a. Although the result is not clear (the stringency was probably too low) the bands normally detected in human DNA are just visible superimposed on the background smear, along with others possibly corresponding to cross-hybridisation with other sulphatases. It is difficult to ascertain whether bands are visible in the mouse and rabbit tracks. Apart from the background smear, the degree of cross-hybridisation is not



FIGURE 5.9
Attempt to identify sequences homologous to p422
in mouse and rabbit.

EcoRI digests of DNA from ♂ and ♀ mouse, rabbit and human controls probed with p422 and washed under conditions of low stringency (60°C, 0.2xSSC/1% SDS). There is no cross-hybridisation with sequences present in the mouse or rabbit.

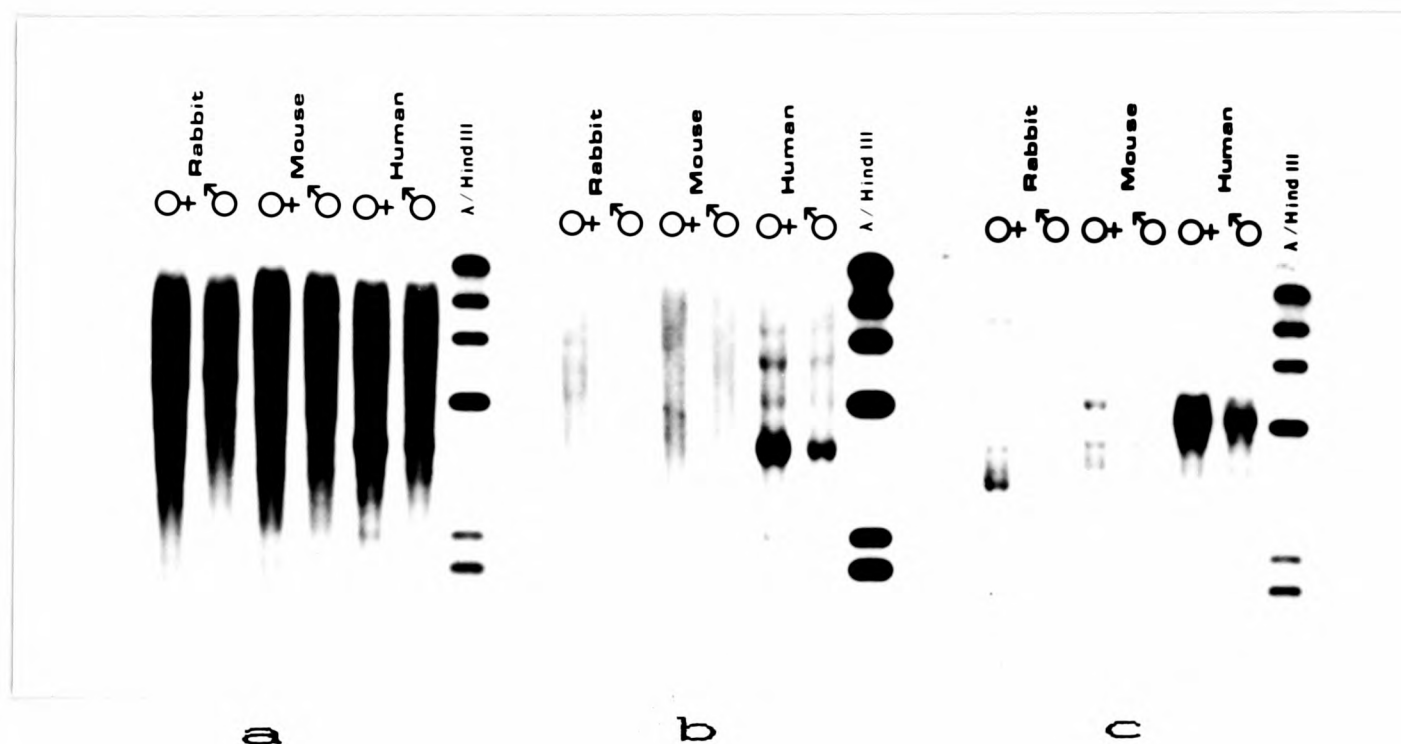


FIGURE 5.10
Attempt to identify sequences homologous to P2A7
in mouse and rabbit.

EcoRI digests of DNA from ♂ and ♀ mouse, rabbit and human controls probed with P2A7 under conditions of low stringency (55°C) and washed under conditions of low stringency (2xSSC/0.1% SDS, room temp) (a), or high stringency (0.1xSSC/0.1% SDS, 64°C) (b). The same filter was re-probed with cVII (factor IX probe) under conditions of high stringency (c). Sequences homologous to P2A7 can not be detected in the mouse or rabbit under conditions in which significant homology can be detected with human factor IX coding sequences.

very convincing. Any bands that are visible, appear to be more significant in the rabbit tracks. After overnight exposure the filter was washed stringently at 64°C using 0.1xSSC/0.1% SDS. the result is shown in Figure 5.10b. Although bands remain in the human tracks, no bands are clearly discernable in the rabbit or mouse tracks. Probe was removed from the filter and it was reprobed with the Factor IX probe cVII (Anson *et al.*, 1984). Conditions of hybridisation and washing were stringent (64°C 0.1xSSC/0.1%SDS). The result shows that under these conditions the mouse and rabbit homologues of factor IX can easily be identified (Figure 5.10c). These preliminary results suggested the lack of any close sequence homology between murine and human *STS* genes.

A range of different stringencies has been used by E. Levy (personal communication) to try and detect homology with mouse sequences. Hybridisation at 55°C has been followed by washes in 2xSSC/0.1%SDS at 25, 37, 43 and 55°C. In each case, although bands were discernable in human control tracks, only background smears were observed in the mouse tracks.

5.4.2 Conservation of *STS* homologous sequences during primate evolution

The availability of blood samples from a range of primates from which it was possible to isolate varying amounts of DNA, meant that it was possible to analyse the conservation of *STS* homologous sequences during primate evolution. DNA from ♂ and ♀ chimpanzee, orang-utan, gorilla, saki, tamarin, brown lemur and ♂ macaque was isolated of sufficient quantity and quality to allow further analysis. Filters were prepared using 2.5µg of

the above DNAs digested with either *Pst*I or *Eco*RI. In addition to these, samples of ♂ and ♀ rabbit, mouse and human controls were also included. Filters were hybridised at 60°C with P2A7 as the probe. Washing was performed at 60°C using 2xSSC/0.1% SDS. The results are shown in Figures 5.11 and 5.12a. The most striking observation is that *STS* homologous sequences have been strongly conserved throughout primate evolution and are clearly present in species as far removed as the lemur, which is thought to have diverged from the line leading to modern man ≈70 Myr ago. Significant hybridisation is seen in the tracks containing rabbit DNA, but again there are no signs of significant hybridisation in the mouse tracks. The pattern obtained with the *Eco*RI digest of chimpanzee DNA is exactly the same as the pattern for human DNA. The Y-specific fragment of 1.5 kb is clearly visible, although the Y-specific fragment of >20kb is not easily discernable in either the human or chimpanzee tracks. With *Pst*I the pattern obtained for ♀ chimpanzee and human is identical; however, the additional fragments present in the ♂ tracks of each species are different. In the ♂ human track the additional bands are of M.W. ≈6 kb and ≈2.5 kb. A single additional band of ≈2.1 kb is visible in the ♂ chimpanzee track. Male specific bands are not visible in any of the other ape tracks. Differences in the *Pst*I pattern of the ♂ and ♀ orang-utan appear to be due to a RFLP, since an additional fragment present in the male is accompanied by the loss of a fragment that is present in the female. Several differences are present in *Pst*I digests of ♂ and ♀ lemur DNA. The data are not sufficient to say whether some of these differences are sex limited, but it would be necessary to

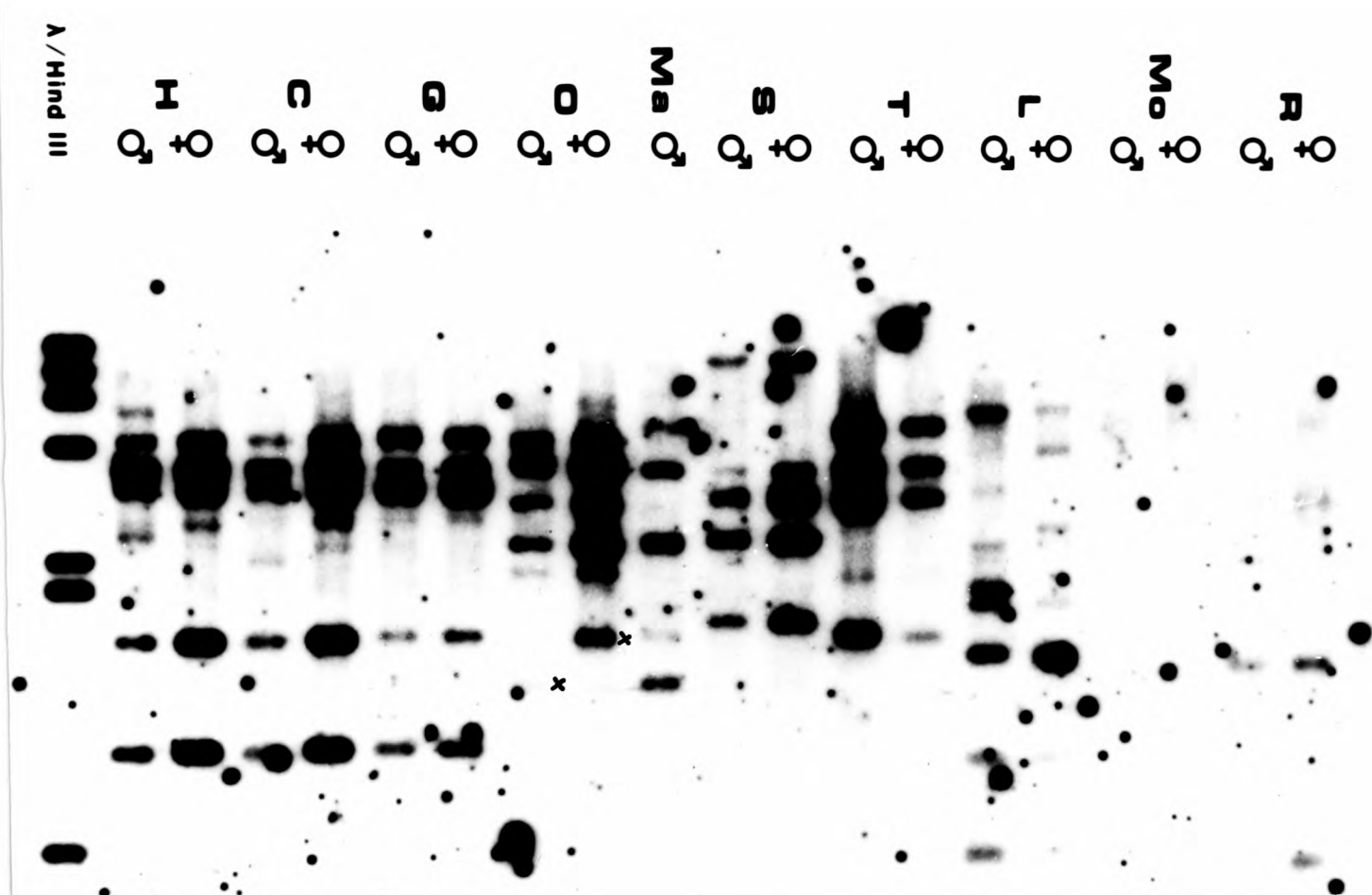


FIGURE 5.11
Evolutionary conservation of *STS* homologous sequences

H: Human; C: Chimpanzee; G: Gorilla; O: Orang-utan;
Ma: Macaque; S: Saki; T: Tamarin; L: Lemur; M: Mouse; R: Rabbit

DNA from different species digested with *Pst*I and probed with P2A7. *STS* homologous sequences have been strongly conserved throughout primate evolution. Significant hybridisation is seen in the rabbit, but not mouse, tracks. Y-specific fragments are present in the ♂ human (5.5kb & 2.5kb) and the ♂ chimpanzee (2.2kb) tracks. Differences in the ♂ and ♀ lemur tracks are more easily explained by the presence of RFLPs. A RFLP is also detected in the orang-utan, since the fragment marked (x) is a different size in the ♂ and ♀ tracks.

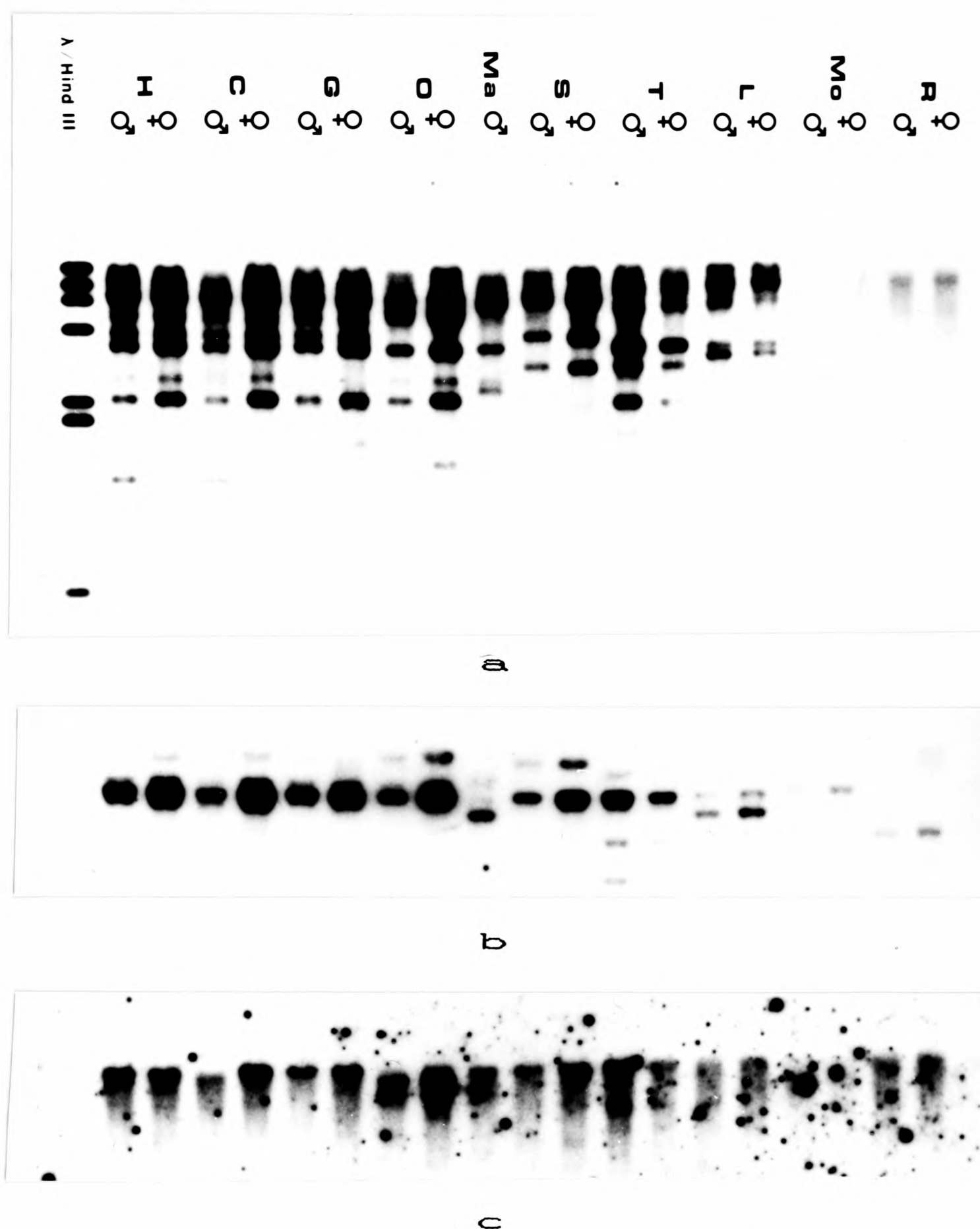


FIGURE 5.12
Evolutionary conservation of *STS* homologous sequences

H: Human; C: Chimpanzee; G: Gorilla; O: Orang-utan;
 Ma: Macaque; S: Saki; T: Tamarin; L: Lemur; M: Mouse; R: Rabbit

DNA from different species digested with *EcoRI* and probed with a) P2A7, b) cVII (factor IX probe, X-linked) & c) α -globin (autosomal). The 1.5kb Y-specific band present in the δ human track, is clearly visible in the δ chimpanzee track. The dosage observed with the factor IX and α -globin probes suggests that the amounts of DNA in each track are not exactly equivalent. However, X-, but not Y-, linkage of *STS* homologous sequences is likely in the higher primates. In the lemur & rabbit the δ : q ratio of signals is suggestive of (pseudo)autosomal linkage.

evoke polymorphism on both the X and Y chromosomes of lemurs if this is the case. Although a comparison of the dosage in signal between the δ and ϕ tracks suggests X-linkage (at least in the case of the higher primates) and hence lack of extensive representation on the Y chromosome, the amounts of DNA in each track is not strictly equivalent. The filter with the *EcoRI* digested DNAs was re-probed with cVII (an X-linked probe for factor IX) and an α -globin derived probe (autosomal), to compare the δ : ϕ ratio of intensity of hybridisation obtained with probes known to be X- and autosomally linked in all these species (Figures 5.12b and 5.12c). The factor IX probe shows an increased dosage in all cases except the tamarin where there was clearly more DNA loaded in the δ track than the female. The result from probing with α -globin is disappointing, since there is a lot of background hybridisation obscuring the tracks. However, it does suggest that in some cases there is a bias in loading. Nevertheless, X-, but not Y-, linkage of *STS* homologous sequences in higher primates does seem likely, since there is only evidence for Y-specific fragments in the chimpanzee and humans. The picture in the lower primates is less clear, X-only linkage is likely in the saki, as there is a clear δ / ϕ dosage effect which does not seem to be due to a bias in loading. The signal obtained from the δ and ϕ lemur tracks however, is suggestive of an autosomal localisation. This also appears to be true for the rabbit tracks. In both cases the dosage observed with the factor IX probe is not observed with the *STS* probe.

Obviously it is not sensible to draw any firm conclusions from these limited data, since only two individuals from each

species were examined and the X/Y dosage experiments need to be repeated and quantitated more accurately; the limited amount of material and time available has precluded this being done at this stage.

5.5 GENERAL DISCUSSION AND CONCLUSIONS

The isolation of a cDNA clone, p422, by immunological screening with anti-STS antibodies and its assignment to Xp22.3 is suggestive that it represents part of the *STS* structural gene. The observation that corresponding sequences are deleted in patients with classical ichthyosis helps to confirm this hypothesis (Ballabio *et al.*, 1987). The isolation of the longer cDNA, P2A7, has allowed a more thorough investigation of this locus. Further confirmation of the provenance of the STS clones was later provided by comparison with results obtained by Yen *et al.* (1987a). Their well characterised STS cDNA clones have been shown to correspond to the amino acid sequence of the protein. Southern blot analysis has not provided evidence for other X-encoded, or autosomal, homologous loci. However, another cDNA clone, 8A11, isolated during the immunological screening of the library has been provisionally mapped to Xq13 (E. Levy, personal communication). The nature of the locus detected by this sequence has not been characterised any further. There appear to be *STS*-related sequences present on the long arm of the Y chromosome; these sequences are not expressed. The data indicate that sequences corresponding to the 3' end of the cDNA are not represented in the Y-linked locus. However, genomic sequences mapping 7-8kb to the 3' end

of the cDNA, detected by B12RI5, do appear to be present on the Y chromosome. This observation, along with the identification of intron sequences on the Y chromosome by Yen *et al.* (1987a), suggest that it is highly unlikely that these sequences result from viral retroposition of STS mRNA sequences.

The availability of DNA samples from a number of primates has allowed the analysis of the evolutionary history of STS homologous sequences. The basic taxonomic classification of the species analysed is: Great Apes - chimpanzee, gorilla and orang-utan. Old World Monkey - macaque. New World Monkeys - saki and cotton top tamarin. Prosimian - brown lemur. The evolutionary history of the primates is a matter of some debate and has been reviewed in Passingham (1982). Placental mammals are thought to have evolved around 100 Myr ago, towards the end of the Cretaceous period. The primates diverged from the ancestral mammalian stock roughly 75 Myr ago. Man and the lemurs are thought to have been separated for ≈ 70 Myr. The molecular evolution of the hominoid primates has been studied by various techniques including DNA hybridisation, DNA sequencing, protein sequencing and a variety of immunological techniques. In spite of certain contradictions the derived phylogenies are strikingly concordant. The general conclusions are that the New World monkeys diverged 40-50 Myr ago. The great and lesser apes split from the Old World monkeys 30-40 Myr ago. The next split, which led to the lesser apes, was followed by the divergence of the ancestors of the Asian great ape, the orang-utan. The hominidae split from the African great apes as recently as 4.5 Myr ago. The human-chimpanzee-gorilla trichotomy has been difficult to resolve (reviewed in Goldman

et al., 1987). They have attempted to resolve their relationship and have constructed a molecular phylogeny for the hominoid primates, by using genetic distances from a survey of 383 radiolabelled fibroblast polypeptides resolved by two-dimensional electrophoresis. They have concluded that the Old World monkeys diverged 37 Myr ago, the gibbons 20-25 Myr ago, the divergence of the orang-utan was just prior to that of the gorilla (≈ 13 Myr) and the chimpanzee-human split was ≈ 8 Myr ago. All of these results can be summarised to give the approximate times at which different primate groups are thought to have diverged from the line leading to modern man, shown in Figure 5.13.

The observation of an identical pattern in the *EcoRI* digest of chimpanzee DNA to that in humans, when probed with P2A7, suggests that *STS* homologous sequences are present on the X and Y chromosomes of chimpanzees as they are in humans, consistent with their, presumptive, close evolutionary origin. If one accepts the evidence for *STS* location in the pseudoautosomal region of mice, then two features require explanation: a) lack of sequence conservation, b) evolutionary rearrangements of the X and Y chromosomes. The lack of sequence conservation between human and murine *STS* at the DNA level may not be too surprising. Anti-human *STS* polyclonal and monoclonal antibodies do not bind mouse *STS* (Yen *et al.*, 1987a); similarly, anti-mouse *STS* antisera do not react with the human enzyme (C.Bishop, personal communication), suggesting that the proteins have diverged markedly. Since *STS* deficient mice are apparently normal and the deficiency in man is nondebilitating, the selection for sequence conservation may be quite low. The

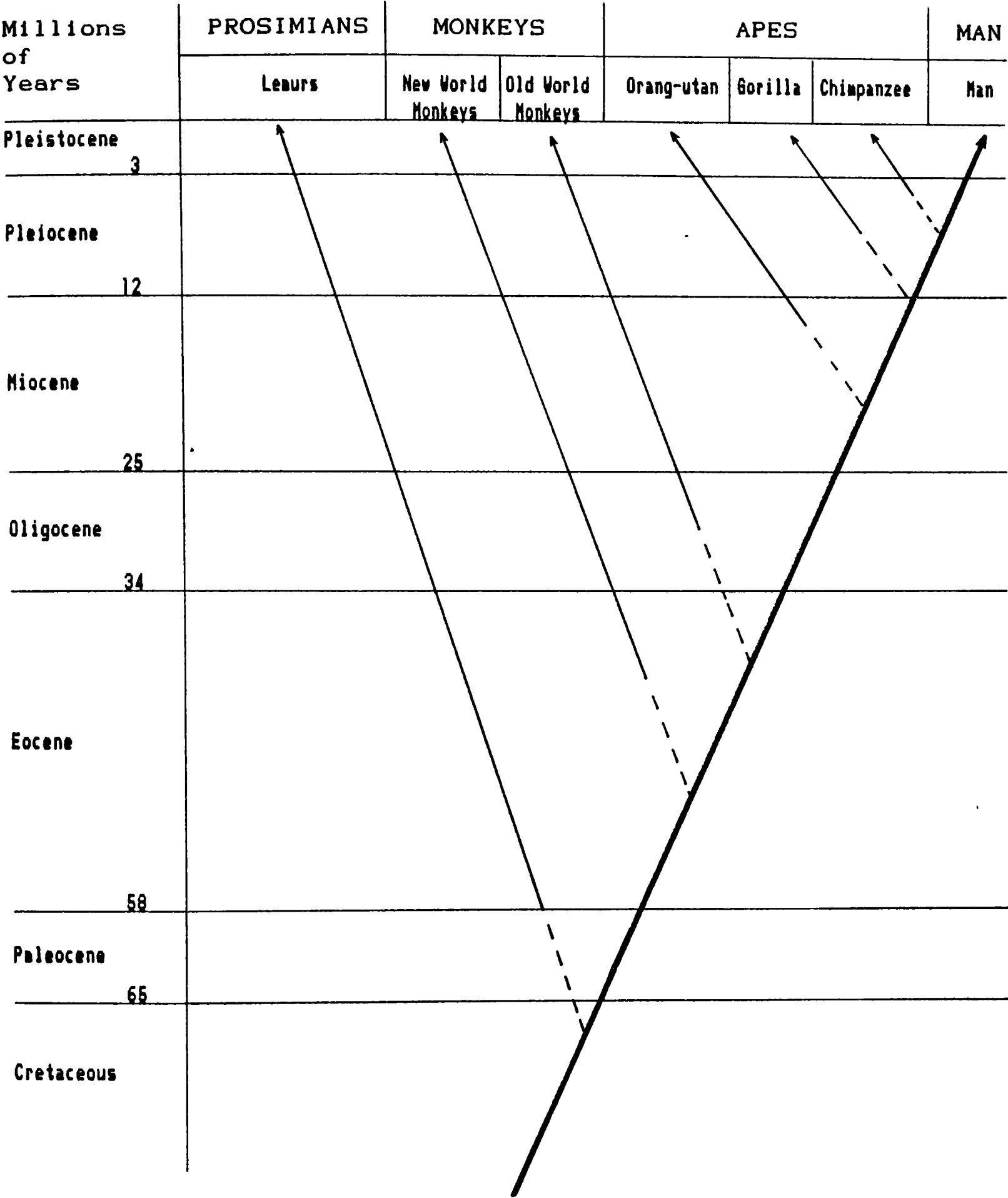


FIGURE 5.13
Primate radiation

The dates of divergence of the primate species used in the evolutionary studies of STS homologous and other DNA sequences. The geological timescale is given on the left. The dates of divergence from the line leading to modern man are rough estimates based on data from various sources (see text for details).

precise function of the *STS* locus in the mouse is not known. It may be of interest that the only other pseudoautosomal gene, *MIC2*, has not been assigned a function in man. It has been suggested that several inversions and transpositions could relate the present organisation of Y chromosomes in mouse and man (Craig *et al.*, 1987). It would be unwise to assume that the minimum path length is the correct explanation for this relationship and it appears that the mammalian Y chromosome organisation is remarkably fluid. Other sequences have been shown to have a similar Xp:Yq distribution (typified by *DXS31* (Koenig *et al.*, 1985) and *GMGXY3* (Affara *et al.*, 1986). *DXS31* has been shown to be X- and Y-linked in the chimpanzee, similar to the result described here for *STS*. In contrast to the $\approx 92\%$ homology thought to be shared by the X and Y copies of *STS* (Yen *et al.*, 1987b), the Y copies of these loci have been estimated to be only 80% homologous to sequences present on Xp. It is possible that these sequences represent a different class of X-Y homology. If mutations have accumulated as the result of neutral evolution, then, based on the lowest rate of changes (4×10^{-9} per site per year), these regions have been separated for ≈ 50 Myr. Alternatively, if many of the differences have been due to the accumulation of small deletions (as has been suggested for the differences between the X and Y copies of *STS*) and this accumulation has varied throughout the region, the transposition of material from the X to the Y may have occurred more recently and all of these sequences could have arisen from the same event. Koenig *et al.* (1985) claim to have demonstrated unequivocally that sequences related to *DXS31* are X-, but not Y-, linked in the macaque. The

results here tentatively suggest that *STS* homologous sequences are X-, but not Y-, linked in the gorilla; if this is the case, the transposition of material from Xp to Yq must have occurred during the past ≈ 13 Myr. However, based on this model, it is difficult to envisage the transition from pseudoautosomal behaviour of *STS* in the mouse (and possibly the lemur, as the results described here and for *DXS31* suggest) to X-only linkage in other primates and Xp:Yq homology in the chimpanzee and human, without postulating several rearrangements. Further studies of this block of homology in man and other species may shed some light on this conundrum.

It is of interest to note the unusually high incidence of deletions in patients with *STS* deficiency (Ballabio *et al.*, 1987; Yen *et al.*, 1987a). Gillard *et al.* (1987) have also found a high frequency of deletions in male ichthyotics, using the probe GMGX9 which has shown tight linkage with *STS* deficiency. It is tempting to speculate that this phenomenon may be related to the proximity of the *STS* locus to the X-Y pseudoautosomal region and that deletion of the *STS* locus is the result of an aberrant X-Y interchange. However, in the cases examined by Gillard *et al.* (1987), the pattern of inheritance of the *Xg* locus, thought to map distal to *STS*, suggests that this is unlikely unless accompanied by a double cross-over. In some cases, the region of homology on Yq may predispose this region of the X chromosome to abnormal pairing and recombination. Alternatively, unequal cross-over events may be particularly frequent at this site in female meiosis.

CHAPTER 6

THE ANALYSIS OF AN X-Y HOMOLOGOUS SEQUENCE PRESENT ON Yp AND Xq

6.1 INTRODUCTION

The first single-copy sequence formally assigned to the Y chromosome was isolated by Page *et al.* (1982). This sequence, which detects the locus *DXYS1*, recognised a male specific *TaqI* restriction fragment in addition to fragments shared by males and females. *In situ* hybridisation localised the Y sequence to Yp and the X sequence to Xq13-Xq21 (Page *et al.*, 1984). *DXYS1* may be regarded as the 'prototype' of a number of other sequences found to map to these regions of the X and Y chromosomes (reviewed in 1.2.2).

Cosmid Y59 is one of >100 cosmids reacting with human genomic DNA isolated from a bank made by Wolfe *et al.* (1984a). As part of a project for the Final Honours School of Biochemistry Pt.II, Oxford (1984), I had identified three putative single-copy sequences within this cosmid, by screening an *EcoRI* digest with ³²P labelled human genomic DNA, in the manner described in 3.2.3. One of these, 59γ - a 1.4kb fragment, was subsequently subcloned into the plasmid pUC9. Somatic cell hybrid analysis and *in situ* hybridisation has localised this sequence to Yp and Xq21 (Buckle *et al.*, 1987). Figure 6.1a shows an *EcoRI* digest of cosmid Y59 probed with total human genomic DNA, allowing the identification of the three single-copy sequences 59α, 59β and 59γ. Figure 6.1b shows the sub-regional localisation of 59γ to Xq13.3→q21.3, using the hybrid TAG and the cell-line CAT. Both of these contain

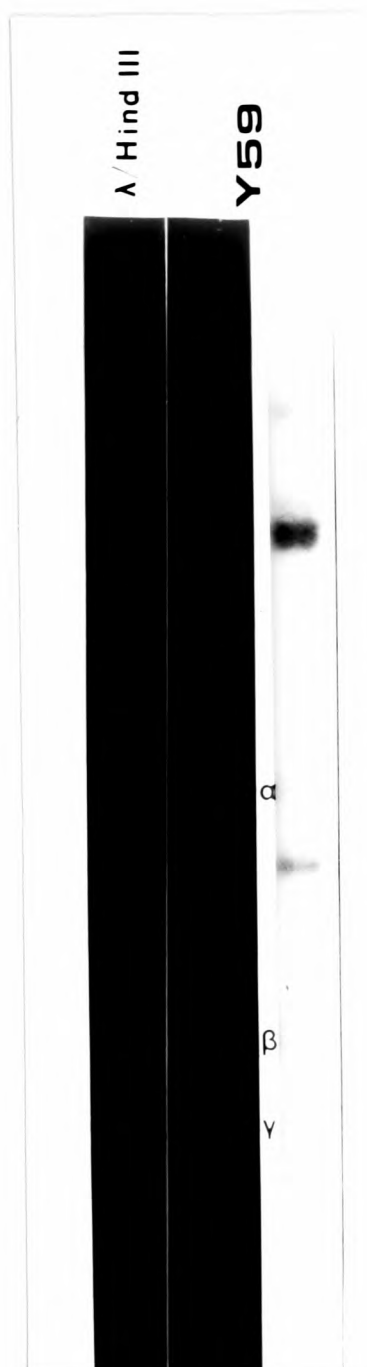


FIGURE 6.1a
Identification of single-copy
sequences in cosmid Y59

*Eco*RI digest of Y59 probed with
total human DNA.

Fragments containing single-copy
sequences have not hybridised
with the probe (α, β, γ). The
other two fragments not
hybridising, belong to the
cosmid vector.

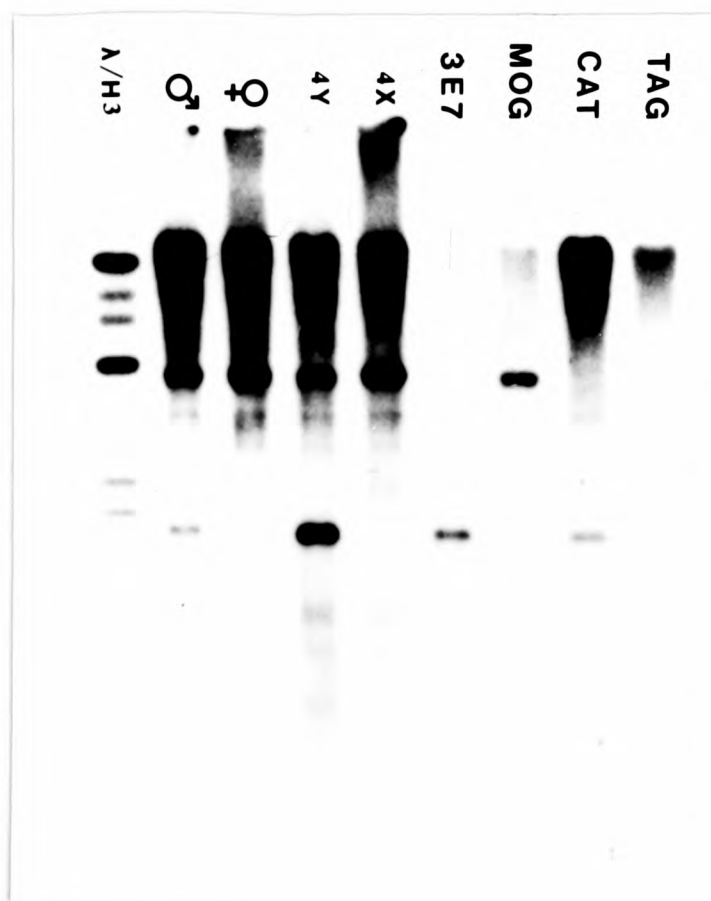


FIGURE 6.1b
Sub-regional localisation of 59 γ

*Bam*HI digests of various DNA
samples probed with 59 γ . This
probe detects a Y-specific
fragment of 1.8kb (present in
the Y-only hybrid 3E7 and shows
increased dosage in the 4Y cell-
line) and an X-specific fragment
of 4kb (present in the X only
hybrid MOG and shows increased
dosage in the 4X cell-line. TAG
is a hybrid containing an X
chromosome with an interstitial
deletion of Xq13.3→q21.3, no
fragment is visible. CAT is a
cell-line with a similar
deletion in Xq, but also
contains a Y chromosome. In this
case, only the Y-specific
fragment is visible. The X-
specific fragment can be
localised to Xq13.3→q21.3.

interstitial deletions of the X chromosome long arm encompassing this region. TAG is a hybrid derived from the cell-line TEL26 (Tabor *et al.*, 1983; Y.Boyd & E. Munro, unpublished information). The cytological characterisation of CAT, 46,Y,del(X)(q13.3→q21.3), is also unpublished (R. Lindenbaum, personal communication). The locus recognised by 59γ has been denominated *DXYS27* in the catalogue established at the 8th Human Gene Mapping Workshop and has the characteristic X-Y distribution of many other X-Y homologous sequences. The X and Y copies of these sequences are 97-99% homologous and are thought to be the result of a transposition of material from the X chromosome long arm to the Y chromosome short arm during recent evolutionary history (Page *et al.*, 1984). Sequences from this region of the Y chromosome have been shown to be transferred to the X chromosome in males with a 46,XX karyotype, as the result of an abnormal X-Y interchange during male meiosis (e.g. see Buckle *et al.*, 1987 and section 1.2.3).

The aim of this chapter was to characterise the locus *DXYS27* in more detail by a) comparing the homologies of the X and Y copies b) looking for the presence of Y material in 46,XX males c) analysing DNA from a number of primates in the hope of gaining some information on the evolutionary history of this locus. A comparison can then be made with other X-Y homologous sequences which may have been in evolutionarily recent rearrangements between the sex chromosomes.

6.2 MATERIALS AND METHODS

6.2.1 General

Hybrids and cell lines were grown as in 2.3 and DNA was isolated from cell pellets using the method described in 2.4.1. DNA was isolated from blood samples using the method in 2.4.2.

Genomic digests were performed overnight using an excess of enzyme, under the conditions recommended by the manufacturer. Samples were subject to electrophoresis overnight at 40V/35mA or for $\approx$ 6hrs at 80V/100mA. All Southern blots were performed using HYBOND-N (2.7.2). Conditions of hybridisation are described in 2.10. Hybridisations were carried out at 64°C overnight and filters were washed to 0.2xSSC/0.1% SDS at 64°C, unless otherwise stated in the text. Typically hybridisation probes were made by the nick translation (2.8.1) of gel eluted fragments.

The elution of restriction fragments is described in 2.7.3 and the subcloning of these into plasmid and sequencing vectors in 2.11. The isolation of cosmid and plasmid DNA is described in 2.5. The screening of bacterial colonies is described in 2.6.

Some of the DNAs used in the polymorphism study were a gift from Dr. A. Miciack. DNA from the hybrid TAG and the cell line CAT was a gift from M. Ross. The probes 69/6, 116/21 and 75/79 were gifts from Dr. J. Wolfe. 47c came from Dr. P.N. Goodfellow.

The XX male cell-lines RCH-C103, RCH-C131, RCH-IS-3a, the XY female cell-line RCH-C335, the two male Australian aborigine cell-lines RCH-AMB and RCH-AMD, and DNA from the XX male EJ were all acquired while at the Royal Children's Hospital, Melbourne.

Primate blood sample from which DNA was made was a gift from R. Lovell-Badge and Prof. John Hearn.

Preliminary sequence analysis of Y59y was carried out by Simon Donowho as a project for the Final Honours School of Biochemistry Pt.II, Oxford (1986).

6.2.2 DNA SEQUENCING

Strain of *E. coli* and vectors used

E. coli JM101 ($\Delta lac\ pro$, *supE*, *thi*, (r_{κ}^{+} , m_{κ}^{+})/F' *traD36*, *proAB*, *lacI^q*, ZAM15) was used for propagation of recombinant phage M13. In order to prevent curing of the F', which is essential for allowing phage M13 infection, the host strain was repeatedly colony purified and maintained on minimal plates.

The vectors mp18 and mp9 were used in this study.

Dideoxy sequencing using single-stranded DNA templates from M13 clones.

Single-stranded templates were made using the method described in 2.13.

Formamide dyes were made by stirring 100ml of formamide with 5g of Amberlite MB1 resin for 30 minutes. The resin was removed by filtration through *Whatmann 1MM* paper. 0.1g xylene cyanol and 0.1g bromophenol blue were added and the solution was made 20mM with respect to Na₂EDTA.

dNTP° mixes contained:

		<u>dNTP° mix</u>			
		A°	C°	G°	T°
0.5mM	dATP	1μl	20μl	20μl	20μl
0.5mM	dGTP	20μl	20μl	1μl	20μl
0.5mM	dTTP	20μl	20μl	20μl	1μl
1xTE	buffer	20μl	20μl	20μl	20μl

To A° was added an equal volume of 150μM dideoxy (dd)ATP; similarly, equal volumes of 20μM ddCTP, 50μM ddGTP and 500μM ddTTP were added to C°, G° and T° respectively. These mixes of dNTP/ddNTPs were stored at -20°C and used routinely in sequencing reactions. However, on some occasions the chain termination reaction was 'fine tuned' by altering the ddNTP concentrations. Increasing the dideoxynucleotide concentrations results in shorter fragments, while lower concentrations result in less frequent chain termination and hence longer fragments.

The 'Universal primer was used: 5'-TAAAACGACGGCCAGT-3'

SEQUENCING PROTOCOL

Annealing Reaction

1. To a 500 μ l microfuge tube add $\approx$ 200ng single-stranded template DNA, 2ng primer, 1 μ l of 10x Klenow reaction/annealing buffer (100mM Tris pH8.0, 50mM MgCl₂) and x μ l dH₂O to give a final volume of 10 μ l. Mix the components well and centrifuge briefly to recover contents at the bottom of the tube.
2. Place the tube in a water-filled test tube which is in a 90°C-100°C water bath. Incubate for 5 minutes to denature.
3. Remove the test tube holding the annealing mixture from the water bath and allow the tube to equilibrate to room temperature (30-45 minutes). The slow cooling step allows the primer to anneal to the template. The primer/template preparation should be stored on ice until the sequencing reactions are ready to be started, or frozen at -20°C to be used later.

Sequencing reaction

1. During the primer annealing reaction, add to the sides of four reaction microcentrifuge tubes labelled A, C, G and T 2 μ l of the relevant dNTP/ddNTP mix.
2. After temperature equilibration, add directly to the tube containing the hybridised primer/template: 1 μ l ³²P- α -dCTP (aqueous solution; 3000Ci/mmol; 10mCi/ml), 1 μ l 0.1M DTT, 1 μ l 22 μ M dCTP and 1 μ l (1 unit) Klenow.
3. Mix by gently pipetting and dispense a 3 μ l aliquot to the four reaction tubes, A, C, G and T which contain the nucleotides.
4. Mix briefly and collect by centrifugation to form a single drop and incubate at 30°C for 15 minutes.
5. Add 2 μ l of 'chase' containing 0.5mM of each dNTP and continue incubation for an additional 15 minutes.
6. Stop the reaction by adding an equal volume of formamide dye mix.
7. Remove an aliquot for gel analysis. Denature by heating at 90°C-100°C for 4 minutes then quickly transfer to ice. Immediately apply 3 μ l samples of each reaction to a gel.

Sequencing gel

7% acrylamide/6M urea gel was made up as follows:

21g urea, 7.5ml acrylamide stock solution, 5.0ml 10x TBE, dH₂O

to 50ml. This was warmed to 37°C to dissolve the urea, but stored at 4°C prior to use.

Immediately before use to this was added:

300µl fresh 10% ammonium persulphate solution and 50µl TEMED (NNN'-tetramethyl ethylenediamine).

The gel was poured immediately using a 50ml measuring cylinder. Acrylamide stock solution was made up as follows:

38g acrylamide, 2g NN'-methylenebisacrylamide (BIS), dH₂O to 100ml. This was stirred for 30 minutes with 5g of Amberlite MB1 resin and then filtered through *Whatman 1MM* paper. The stock acrylamide solution was stored in a dark bottle at 4°C.

Electrophoresis

20x40 cm glass electrophoresis plates were used, and these were thoroughly cleaned before use by washing with detergent, then ethanol, and rinsing with deionised water. The spacers and combs were cut from 0.3mm *Plastikard* and were also washed carefully. The notched plate was siliconised to ease separation of the plates for removal of the gel. The plates and spacers were clamped together using clips and the edges sealed with 2% agarose to prevent leakage. The gel was left to set for one hour after pouring before use.

The wells were flushed out with running buffer (1xTBE) just before loading 2-3µl samples from drawn out capillary tubes. Electrophoresis took place for 1-4 hours. The recommended power setting was 40 watts, but in practice this was often much less (see text). A temperature sensitive strip was placed on one of the glass plates, so that a check could be made on the running temperature; 50-55°C is the ideal temperature for running the gels. Following electrophoresis the notched, (siliconised), plate was removed and the gel, still attached to the other plate, fixed in 10% methanol/10% acetic acid for 20 minutes. The fixed gel was then transferred to a sheet of 3MM paper and dried down, with heat, using a vacuum drier. Gels were exposed to Fuji film for periods of 16hrs to 4 days.

6.3 AN ANALYSIS OF THE DEGREE OF HOMOLGY BETWEEN THE X AND Y COPIES OF DXYS27

6.3.1 X-Y differences detected by restriction enzymes

In order to ascertain whether any differences between the X and Y copies of *DXYS27* could be identified, DNA isolated from a normal ♂ and a normal ♀ was digested with a range of restriction enzymes, blotted onto HYBOND-N and probed with 59γ. The enzymes used were *AvaII*, *BamHI*, *BglII*, *EcoRI*, *EcoRV*, *HindII*, *HindIII*, *PstI*, *PvuII* and *TaqI*. ♂/♀ differences were noticed with *BamHI*, *BglII*, *PstI* and *PvuII* and these are shown in Figure 6.2. This could be accounted for by only 4 single base pair changes out of the 114 nucleotides which constitute the restriction sites detected by the ten enzymes. These preliminary results suggested that the X and Y copies could be ≈96% homologous at this locus. The high degree of homology is also suggested by the fact that both X and Y specific fragments are detected with comparable ease when filters are washed under conditions of high stringency (0.1xSSC/0.1% SDS, 64°C). The X-Y differences detected with *BamHI*, *BglII*, *PstI* and *PvuII* seem to be present in a great majority of the population. Although the investigation has not been exhaustive, only one individual has been found who does not have the characteristic X-Y pattern for these four enzymes. In this case, the Y-specific *BamHI* fragment of 1.8kb is not present, instead there are two fragments of ≈1.5kb and ≈0.3kb, which can be accounted for by the creation of an additional *BamHI* site within the normal fragment (P.J. Goodfellow, personal communication). DNA from 18 unrelated males digested with the four enzymes detecting X-Y

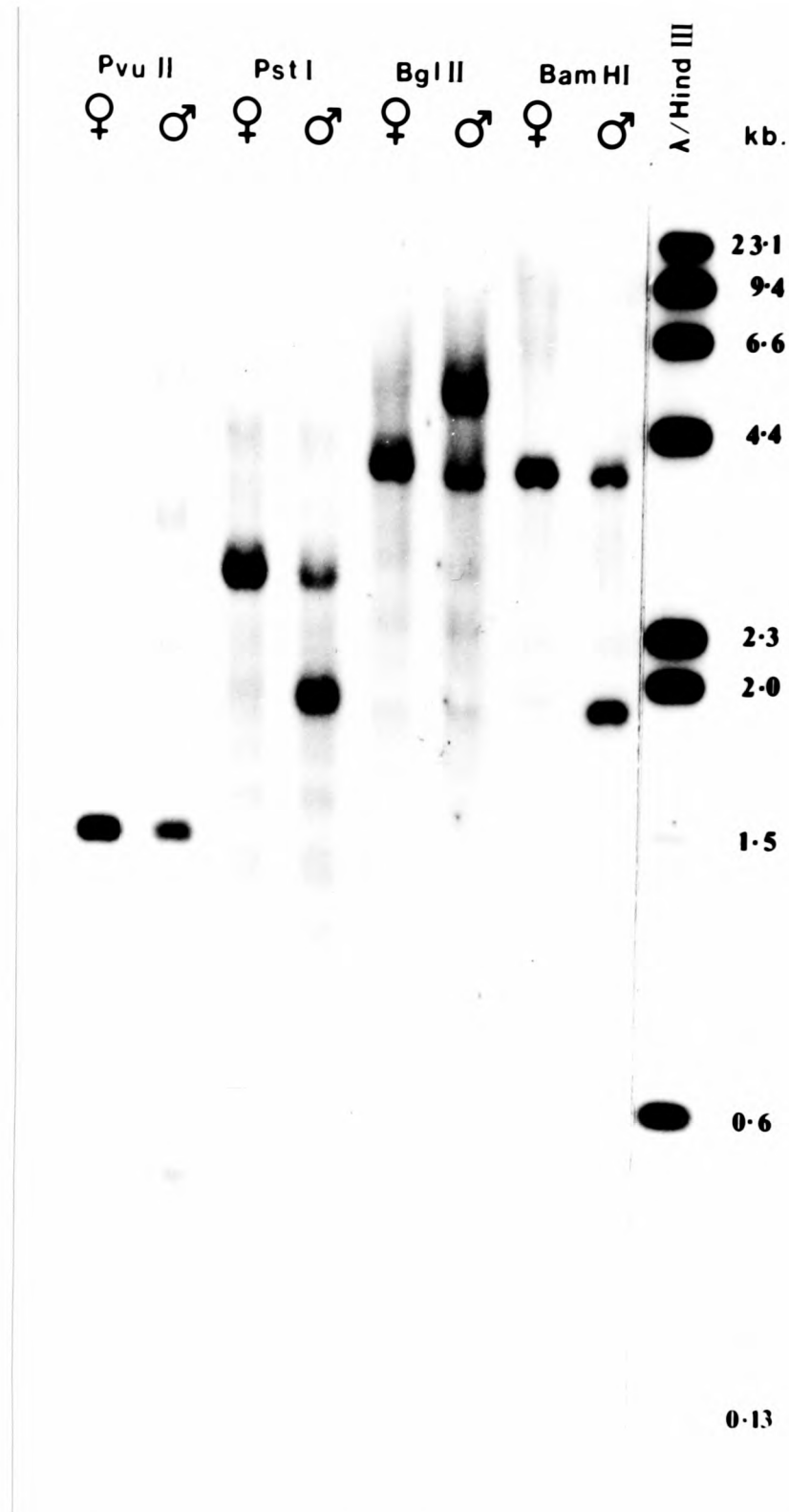


FIGURE 6.2
Identification of X/Y polymorphisms detected by 59γ.

DNA from a normal ♂ and a normal ♀ digested with *Pvu*II, *Pst*I, *Bgl*II and *Bam*HI and probed with 59γ. X-specific fragments are present at 1.3kb (*Pvu*II), 2.7kb (*Pst*I), 3.3kb (*Bgl*II) and 4kb (*Bam*HI). Y-specific fragments are present at 0.9kb & 0.4kb (*Pvu*II), 1.8kb (*Pst*I), 6.6kb (*Bgl*II) and 1.8kb (*Bam*HI).

polymorphisms was probed and in no case was an abnormal pattern observed (Figure 6.3). Although most of these individuals were of Northern European descent, two of these, AMB and AMD, are full blood Australian Aborigines and one, 5J1, is of Asiatic origin. The observation of the same X-Y polymorphisms in all of these individuals suggests that they were present at the time of racial separation. It is not possible to tell whether these polymorphisms are due to mutation of sites present on the X, Y or both chromosomes. Their constancy within the population suggests that they may have occurred at, or near, the time of the transposition event that is supposed to account for this block of X-Y homology or that a subsequent 'founder effect' took place before racial divergence, resulting in the conserved polymorphism seen today. Even in more extensive studies, very few Y polymorphisms have been detected (e.g. Geldwerth *et al.*, 1984; Koenig *et al.*, 1985; Mitchell *et al.*, 1987; J. Wolfe, personal communication), however several X polymorphisms have been found and some of these have shown strong linkage disequilibrium in a variety of ethnic groups (Mitchell *et al.*, 1987).

6.3.2 The isolation of cosmids containing the X homologue of *DXYS27*

In order to be able to compare the X-Y homologies at this locus in more detail, it was decided to try and isolate sequences corresponding to the X copy of the locus *DXYS27*. Although restriction enzyme analysis had suggested that the sequences were likely to be $\approx 96\%$ homologous, direct sequence analysis of the X and Y copies would give a more accurate

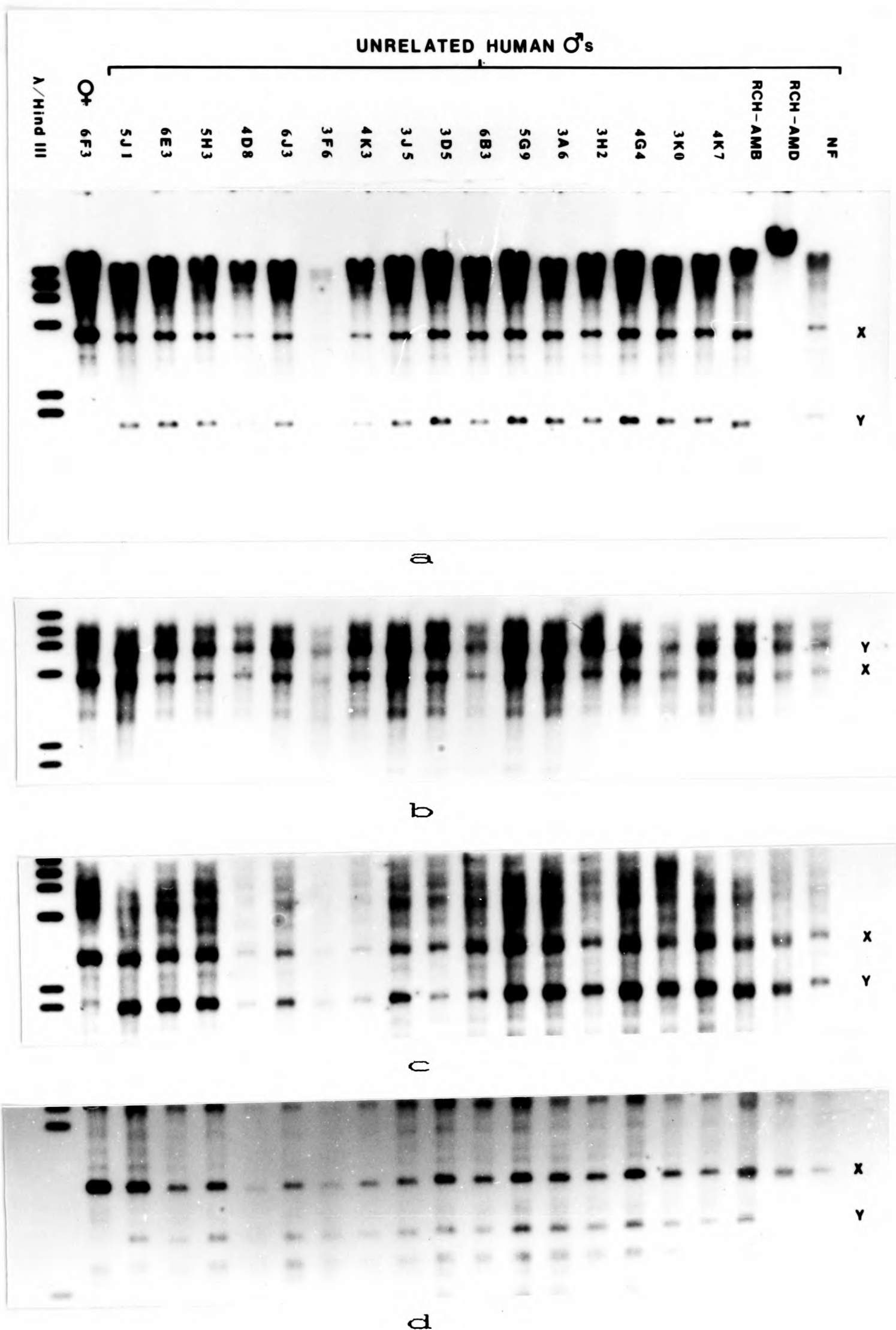


FIGURE 6.3

Population survey of the X/Y polymorphisms detected by 59γ.

DNA from 19 unrelated ♂s and a ♀ digested with a: *Bam*HI, b: *Bgl*III, c: *Pst*I and d: *Pvu*II probed with 59γ. In each case, the characteristic X- and Y-specific fragments are present.

estimation of homology. The MOG cosmid bank (described in Chapter 2) seemed to be an ideal source of material from which to isolate these sequences.

The probability of finding a single-copy DNA sequence in a genomic library is given by the formula:

$$N = \frac{\ln(1-P)}{\ln(1-f)}$$

Where N is the number of recombinants which must be screened to have probability P of finding a single-copy sequence, if a single recombinant contains the fractional proportion f of the total genome. The genome size of MOG is $\approx 10^{10}$ bp and the average insert size of a cosmid 4×10^4 bp. Thus, to have a 90% chance of isolating the sequence, it was necessary to screen $\approx 575,000$ recombinants.

Colonies were spread onto six 22x22cm plates, to give $\approx 100,000$ colonies per plate, based on the titre of the bank calculated in 3.2.2. The actual titre of stock used was less than expected, and the number of colonies on each plate was estimated to be $\approx 50,000$. Nevertheless, there was still a 70% chance of isolating the sequence by screening these colonies. The colonies were transferred to HYBOND-N and screened with insert from 59γ. Filters were washed to a final stringency of 0.2xSSC/0.1%SDS at 64°C. After an overnight exposure of the autoradiograph, two filters (4 and 5) had single signals. Since these were in crowded areas of the plates, plugs were taken using the blunt end of sterile Pasteur pipettes. These were mixed with LB broth to disperse the colonies and serial dilutions plated out. After growth overnight, plates containing ≈ 100 colonies were rescreened with 59γ insert. In both cases 5-10 positive signals were obtained. Positive colonies were

selected and grown overnight. Cosmid DNA was isolated and digested with *EcoRI*. The digested samples were compared with an *EcoRI* digest of cosmid Y59 by running on a gel. Both X-derived cosmids contained a fragment the same size as 59 γ . The two X-derived cosmids had clearly different *EcoRI* patterns, neither of these was identical to the pattern of Y59. These cosmids were designated X594 and X595. The band pattern of X595 seemed to have more fragments in common with Y59 than X594. To check that both X copies did contain sequences homologous to 59 γ , *EcoRI* digests of the cosmids were blotted and probed with insert from 59 γ . In addition to this replica filters were probed with 59 α , and 59 β . The results are shown in Figure 6.4. They confirm that both X594 and X595 contain sequences homologous to 59 γ . X595 also contains sequences homologous to 59 α and 59 β , the X and Y copies are comparable in size. The absence of these sequences in X594 suggests that X594 and X595 may overlap by only a few kilobases, and explains the widely differing *EcoRI* restriction patterns. In addition to *EcoRI*, the cosmids were digested with *BamHI*, *BglII*, *PstI* and *PvuII*, the enzymes shown to detect X-Y polymorphisms with 59 γ . A filter was prepared and hybridised with 59 γ . The result is shown in Figure 6.5. Only in some of the cases is the fragment detected the same size as that in genomic digests. This discrepancy can be accounted for by 59 γ mapping towards the ends of the cosmid inserts and consequently, incomplete restriction fragments may be present in the cosmids; in cases where the fragment detected is larger than that present in the genome, this fragment may be juxtaposed to vector sequences. However, in a number of cases the pattern of hybridisation obtained is entirely consistent

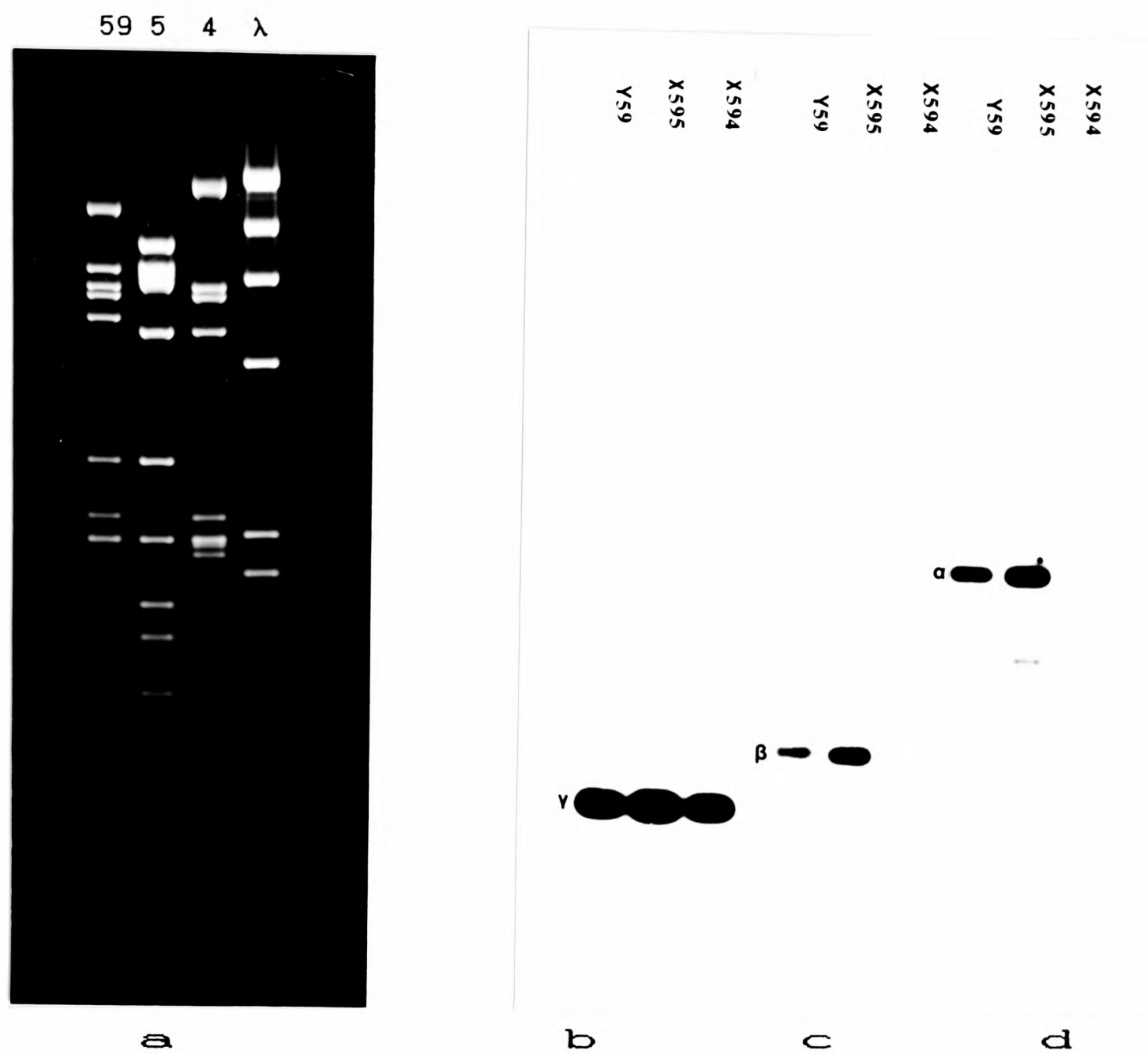
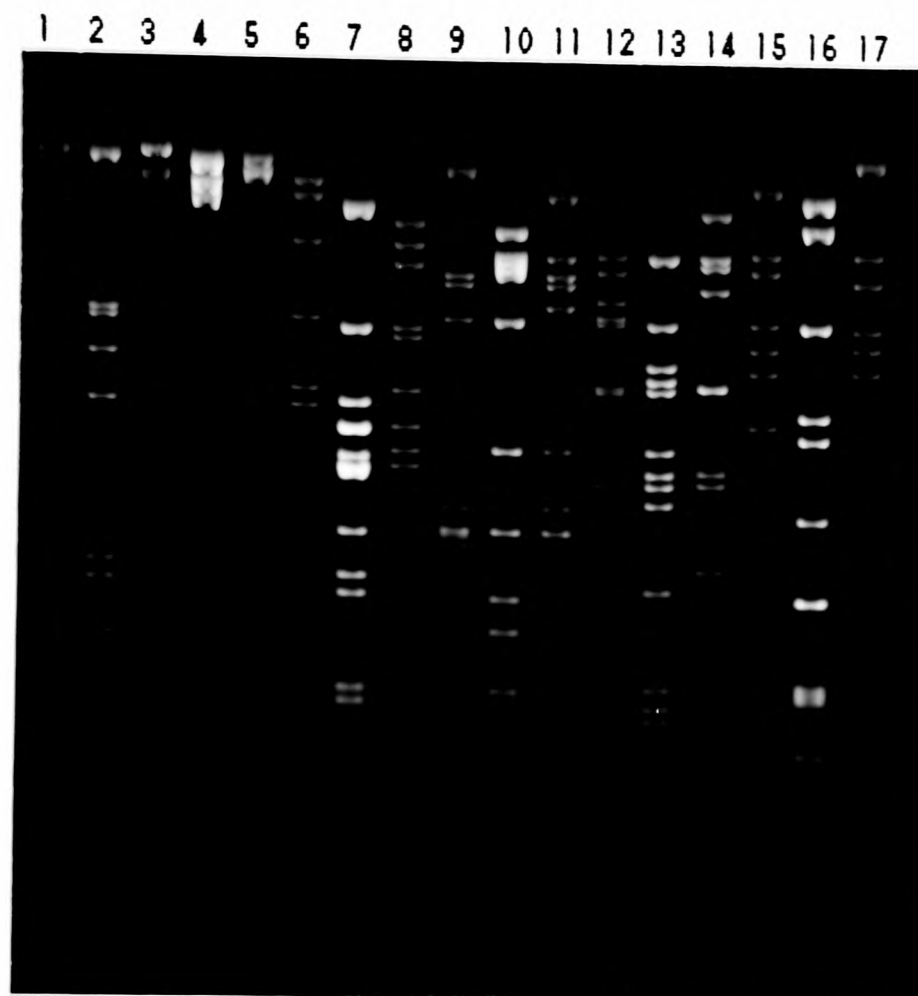


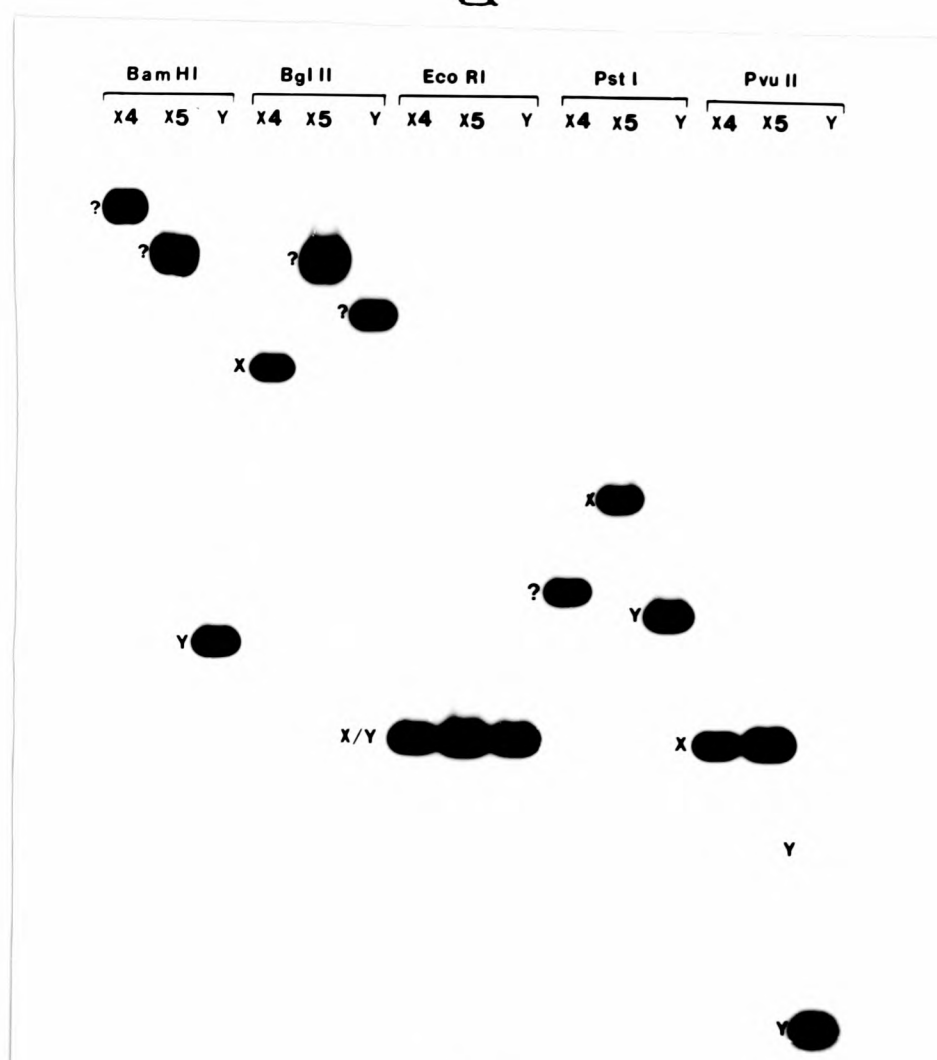
FIGURE 6.4

Detection of homology between the cosmids Y59, X594 and X595

59-Y59, 5-X595, 4-X594, λ-λ/*Hind*III marker fragments.
*Eco*RI digests of cosmids Y59, X595 and X594 (a) probed with 59γ (b), 59β (c) and 59α (d). Both X595 and X594 contain sequences homologous to 59γ. X595 also contains sequences homologous to 59α and 59β.



a



b

FIGURE 6.5

Identification of restriction fragments in cosmids X594, X595 and Y59 that hybridise with 59γ.

1:λ/*Hind*III; 2:λ/*Hind*III/*Eco*RI; 3:X494/B; 4:X495/B; 5:59/B; 6:X494/Bg; 7:X595/Bg; 8:Y59/Bg; 9:X594/E; 10:X595/E; 11:Y59/E; 12:X594/Ps; 13:X595/Ps; 14:Y59/Ps; 15:X594/Pv; 16:X595/Pv; 17:Y59/Pv. B: *Bam*HI; Bg: *Bgl*II; E: *Eco*RI; Ps: *Pst*I; Pv: *Pvu*II. Fragments detected that are the same size as those present in genomic digests are marked X and Y, those of different sizes ?. See text for details.

with X594 and X595 containing the X homologue of Y59 γ . This fragment was isolated from both X derived cosmids and subcloned into both pUC9 and the sequencing vector mp18. Routinely the subclone X595 γ was used for comparison with Y59 γ .

6.3.3 Sequence analysis of X59 γ and Y59 γ

Y59 γ had been subcloned into the sequencing vector mp9 by Simon Donowho, a project student working in the laboratory, and some preliminary sequence data had already been obtained. This was to be verified and compared with that of the X homologue. X59 γ was subcloned into mp18 rather than mp9, since this was the vector available at the time. *EcoRI* fragments are cloned in one of two orientations, thus it is possible to obtain sequence data from both ends of the insert depending on which way round it has been cloned. The opposite orientations are complementary to each other, and so in effect both strands can be sequenced. The orientation of a particular clone can be determined by hybridisation analysis with other clones. 50ng of a reference clone (a), was mixed with a similar amount of clones of unknown orientation in a final volume of 15 μ l of annealing buffer (see 6.2.2). This was boiled for 5 minutes and then allowed to cool slowly to room temperature. Clones with complementary inserts hybridise and can be distinguished from clones with inserts in the same orientation by analysis on a mini-gel. Typical results are shown in Figure 6.6. Where complementary clones hybridise a complex pattern results, as opposed to the discrete band normally observed. The two orientations of the clones were designated a and b.

One of the problems with comparing the two homologues was that in the case of X59 γ it was necessary to sequence $\approx$ 50 bases of the polylinker before sequence corresponding to the insert was reached, however it did mean that it was possible to determine the sequence right from the start of the insert. In the case of Y59 γ , the *EcoRI* cloning site in mp9 adjoins the sequence to which the sequencing primer hybridises, which meant that in practice it was not possible to determine the precise start point of the insert. Routinely, X59 γ clones were loaded onto the sequencing gel 30-45 minutes before loading Y59 γ clones.

It was not possible to compare as much of the sequences as was hoped, due to technical difficulties with the sequencing apparatus (which have still not been resolved satisfactorily). It was not possible to run the gels at the recommended voltage of 1500V without the gel overheating, leading to either hydrolysis of the gel matrix or cracking of the plates. The use of an aluminium heat sink clamped to the front of the plate alleviated the problem slightly, but to maintain a gel running temperature of $\approx$ 55°C it was necessary to run the gels at no more than 1000V. Consequently the bands were often fuzzy and difficult to resolve easily, especially in regions >250 residues from the start. Gels were run for a maximum of 4 hours to minimise overheating problems. The use of different reagents of higher purity did not make an appreciable difference.

It was possible to compare 230 residues from X/Y59 γ (a) and 218 from X/Y59 γ (b). Typical results are shown in Figure 6.7. 'Good runs' were obtained infrequently and the data presented in Figure 6.8, representing the sequences compared, is the

FIGURE 6.6
Hybridisation screening for ssmp18 clones of X59 γ and Y59 γ
in opposite orientations

1: Y59 γ (a) + X59 γ (b); 2: Y59 γ (a) + X59 γ (a);
 3: Y59 γ (a) + X59 γ (a); 3: Y59 γ (a) + X59 γ (b);
 5: Y59 γ (a) + X59 γ (b).
 A: Y59 γ (a); B: Y59 γ (b).

In each of the cases (1-5), Y59 γ (a) was used as the reference clone and was mixed under annealing conditions with an X59 γ clone of unknown orientation. X59 γ clones in the same orientation result in a single band pattern (2 & 3); clones in the opposite orientation give a more complex pattern (1, 4 & 5) due to hybridisation of the inserts.

FIGURE 6.7
Sequence analysis of Y59 γ and X59 γ

a: Comparison of X59 γ (α) and Y59 γ (a) between residues 55-189 (see Figure 6.8). No differences are apparent in this region.

b: Comparison of X59 γ (b) and Y59 γ (b) between residues 55-218 (see Figure 6.8). Three differences are apparent: at positions 98 (Y:T, X:A), 152 (Y:T, X:?) and 212 (Y:A, X:G). These differences are seen more clearly in :-

c: Position 98
 d: Position 152
 e: Position 212

1 2 3 4 5 A B

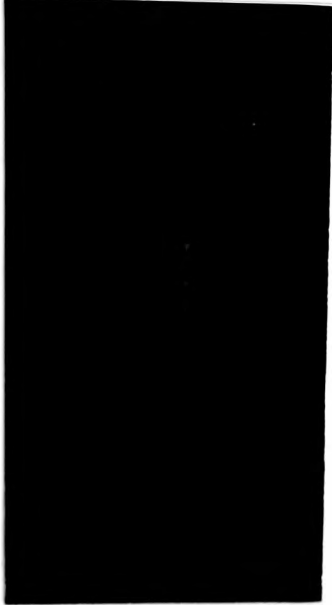


FIGURE 6.6
Hybridisation screening
for ssmp18 clones of
X59γ and Y59γ in
opposite orientations

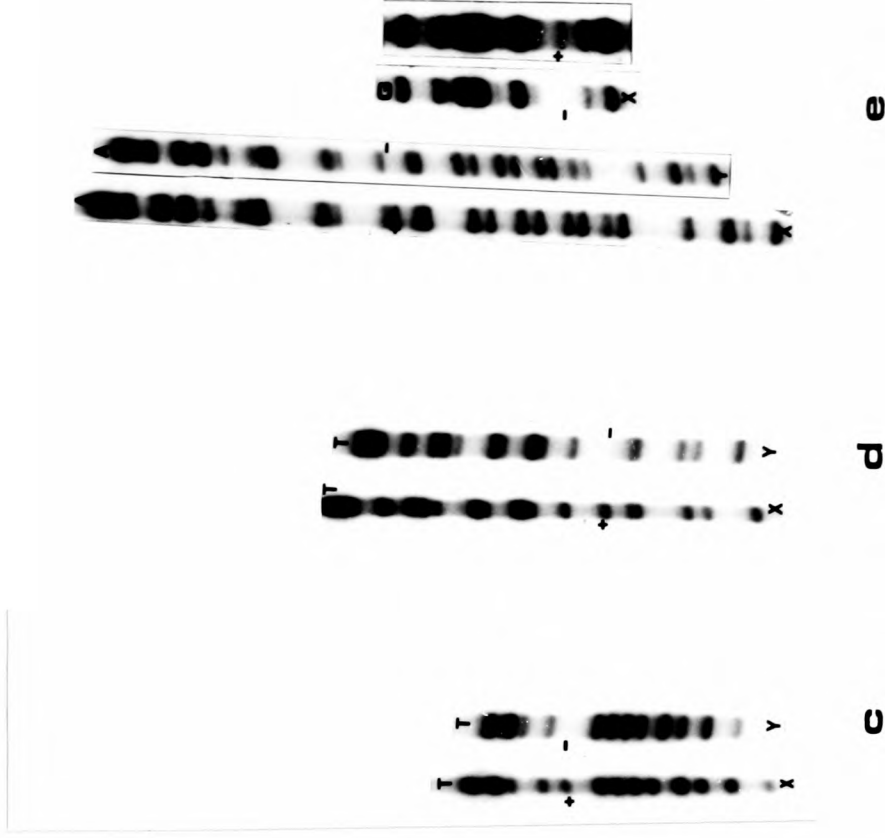
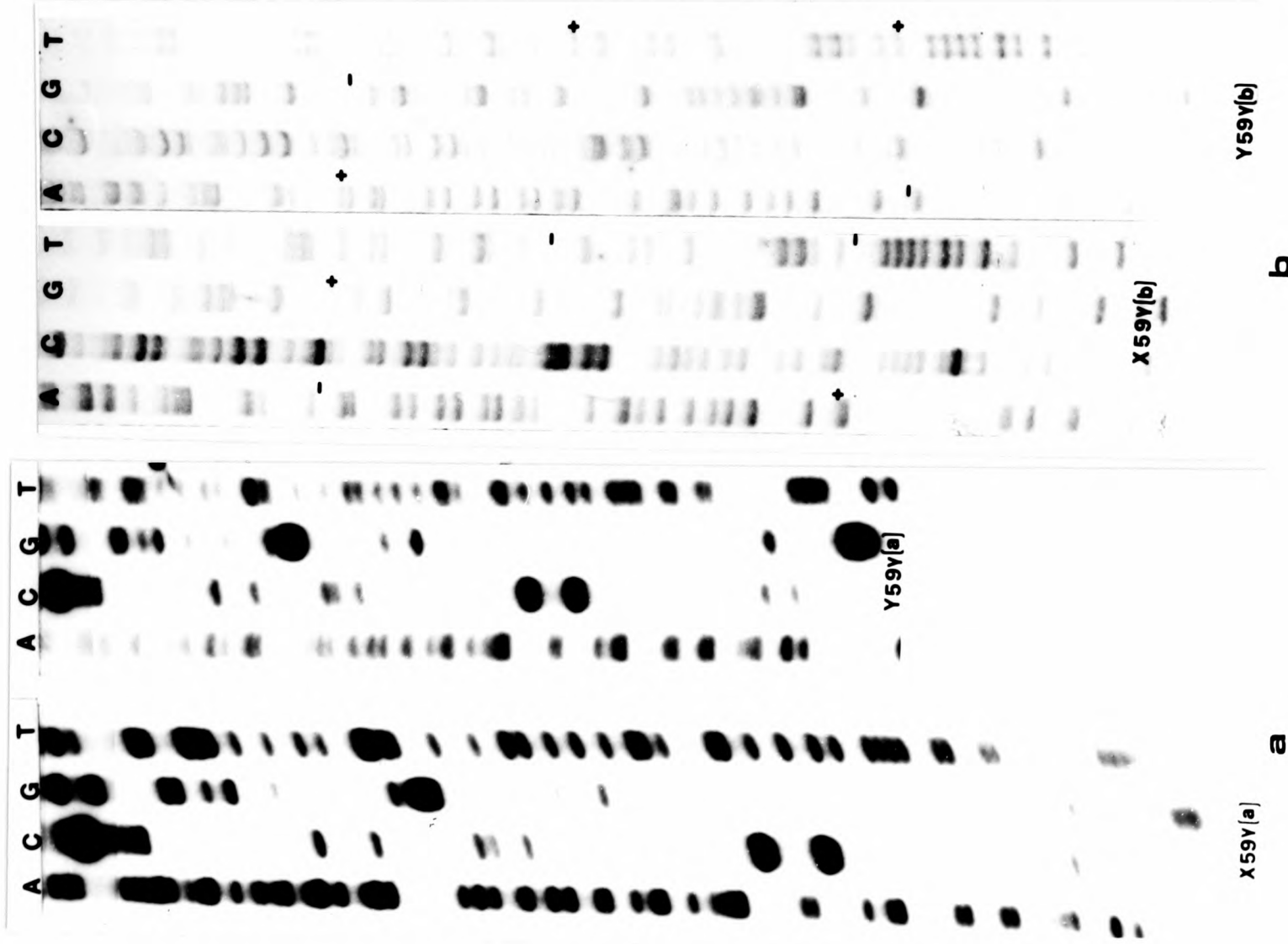


FIGURE 6.7
Sequence analysis of Y59γ and X59γ

```

[ GAATTCAACT] AGCCAAGTTGGGCACATT
                    5 0 ↓
AGAAAGCCACCACAAAGACTTATAAGCTA
TGGGTTTTTACAGCAAATTAATTAATTTTA
                    1 0 0 ↓
ATATCCTAATCCTTAAAATATTAATAGTA
ATGATATCAATCACATAGGGGTTGTTACAA
                    1 5 0 ↓
TGATCAAATGAATAAGTATGTAAAGAGTA
                                   2 0 0 ↓
TAACACACACCGGTCCTAGAAATATTGTTTA
TACTACATACATATTAACATAGAGGGGTCA
                                   2 5 0 ↓
C[ AAATTTACTATTACCTAAAA]

```

FIGURE 6.8a
Nucleotide sequence of X/Y59γ(a).

The first 10 nucleotides have been determined for X59γ(a) only. X59γ(a) and Y59γ(a) have been compared up to, and including, residue 230. The remainder of the sequence (230-250) refers to Y59γ(a) only. No differences were found in the regions compared.

```

[ GAATT] CTC TGGGTCCAAACTGTTCCAA
                    5 0 ↓
TCTCCTGGGCACCAGCAGGAGACTAGTGGC
AGACTGGAAC TGCACAATGTGGCTGCCTC
                    1 0 0 ↓
TGCTCTCTCTGGTACCTCAGTCTTCTTAG
                    1 5 0 ↓
GCAGCAGGCAGCTG CAGCGATAATGGCCA
                    2 0 0 ↓
CCCTTCCC TCAGGAACTCAGCAGTCTTAG
GCAGTCTCCAGCCAGCCAAGTGGCTGCTG
AGCTAGACTCACC ACTC

```

FIGURE 6.8b
Nucleotide sequence of X/Y59γ(b)

The first five nucleotides have been determined for X59γ(b) only. Y59γ(b) and X59γ(b) have been compared up to, and including, residue 218. Three differences have been found (*): at positions 98 (Y:T, X:A), 152 (Y:T, X:?) and 212 (Y:A, X:G).

compilation of data from many gels. Although the sequence may be inaccurate in places, it was relatively easy to see where differences between the X and Y homologues occurred. Of the 448 residues compared, only 3 differences were found. This suggests an overall homology of >99%, obviously the more sequence data obtained, the more accurate this estimate is likely to be.

6.4 THE ANALYSIS OF XX MALES AND AN XY FEMALE FOR THE PRESENCE OF THE Y HOMOLOGUE OF DXYS27

In section 1.2.3 abnormal X-Y interchange resulting in XX maleness was discussed. Since 59 γ had been mapped to the short arm of the Y chromosome, it is of interest to see whether this region of the Y chromosome is transferred to the X chromosome of XX males, in agreement with the results of several other groups (reviewed in 1.2.3). The presence of the X or Y homologues of 59 γ can easily be distinguished using the restriction enzymes revealing X-Y polymorphisms. The enzymes *Pst*I and *Bam*HI have been used previously to demonstrate the presence of Y-specific bands in the XX male cell-lines GIM and WER, when probed with 59 γ (Buckle *et al.*, 1987). These cell-lines were also shown to contain the Y-specific bands of the locus DXYS25, which shows a similar X-Y distribution to DXYS27 (Wolfe *et al.*, 1984b). The probe 75/78 (which detects this locus along with 75/79) was shown, by *in situ* hybridisation, to hybridise to the tip of the X chromosome in the XX male GIM, in addition to a site of hybridisation at Xq21 (the only site of X chromosome hybridisation in normal males). These results suggested that in some cases the Y-homologue of DXYS27 may be

transferred to the X chromosome of XX males. The availability of cell-lines from three Australian XX males and blood from another has allowed this study to be carried further.

DNA isolated from the XX male cell-lines C103, C131, IS-3a and from lymphocytes from EJ, was digested with *Bam*HI, blotted and probed with insert from 59γ. The result is shown in Figure 6.9b. A Y-specific bands is present in each of the three cell-lines, but is absent from EJ. The probing of a *Pst*I digest of the different DNAs gives a similar result (not shown). Filters were also prepared from DNAs digested with *Eco*RI and *Taq*I and probed with 75/79 and 47c respectively. 47c comes from the same block of X-Y homology as 75/79 and 59γ, but is thought to map distally and closer to TDF. Y-specific fragments detected by this probe have been shown to be present in other XX males (e.g. Guellaen *et al.*, 1984). In both cases, Y-specific fragments were detected, except in the case of EJ (Figures 6.9a and 6.9d). The filter with the *Eco*RI digests was also probed with the Y long arm probes 69/6 and 116/21 (Wolfe *et al.*, 1984b) (Figure 6.9c). In none of the cases were sequences detected, confirming that these were true XX males and not XX/XY mosaics. In addition to these XX males, Y-specific fragments have been detected in *Bam*HI digested DNA from 4/6 XX males probed with 59γ (P.J. Goodfellow, personal communication). One of the XX males negative for the Y-specific fragment of 59γ, GM1889, has also been examined by Müller *et al.* (1986). None of their eleven Y-specific probes, shown to be present in one or more XX males, hybridised to DNA from this individual. The absence of detectable Y material from this individual and also EJ classifies them as Y(-) XX males.

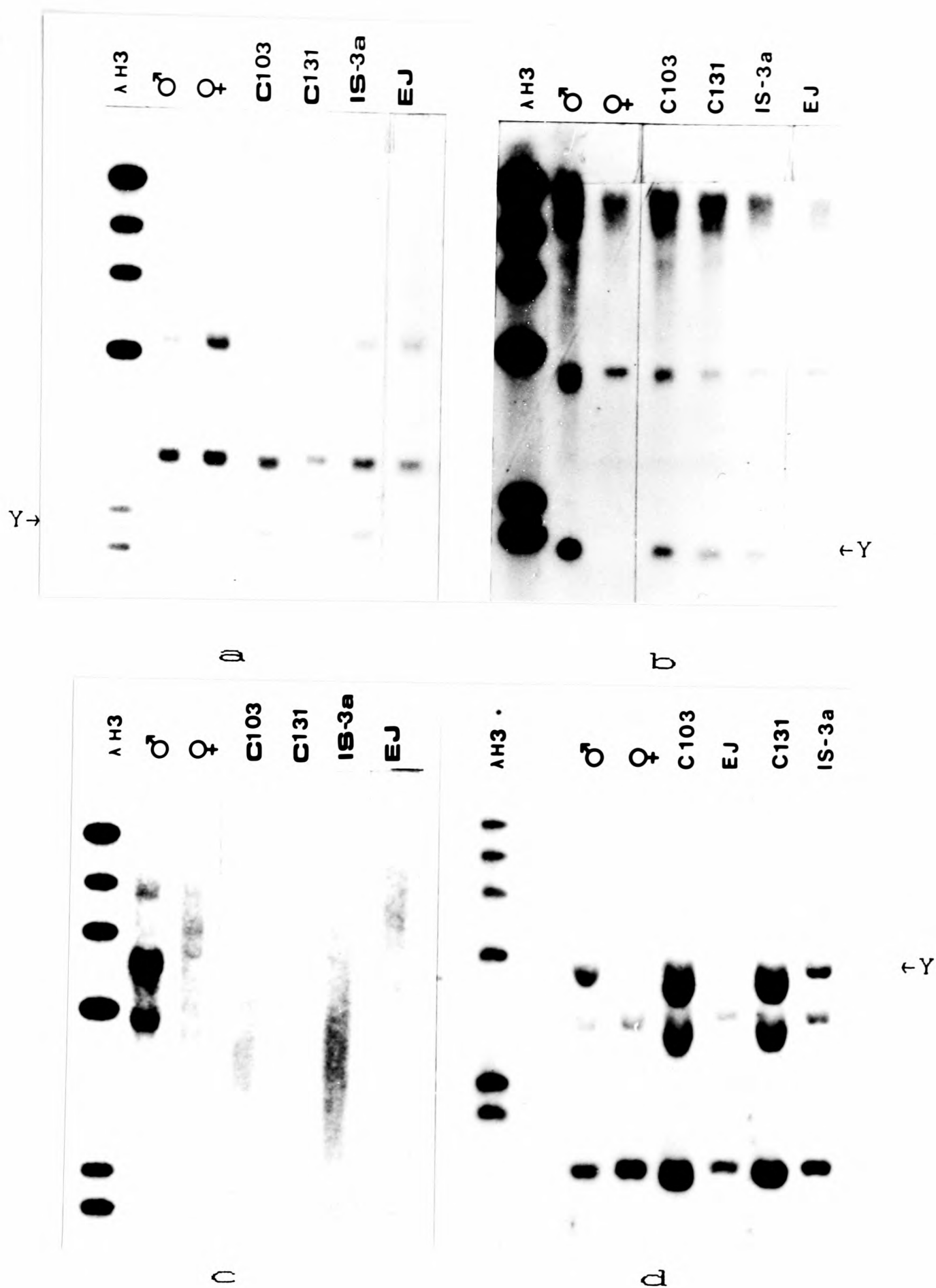


FIGURE 6.9
The analysis of XX males with Y-specific
and X/Y homologous probes.

λ /H3: λ /HindIII marker fragments; ♂ & ♀: normal controls;
C103, C131, IS-3a & EJ: XX males.
a: EcoRI digests probed with 75/79 (Probe maps to Yp and Xq.
Y-specific band: 2kb).
b: BamHI digests probed with 59 γ (Probe maps to Yp and Xq.
Y-specific band: 1.8kb).
c: EcoRI digests probed with 69/6 and 116/21 (Probes map to Yq.
Y-specific bands 3.9kb (69/6), 8.7kb & 5.1kb (116/21)).
d: TaqI digests probed with 47c (Probe maps to Yp and Yq.
Y-specific band 4.1kb)
Yp material is present in C103, C131 and IS-3a. None of the XX
males contains Yq material.

although further studies using more and well defined Yp probes would be necessary to confirm this.

The availability of DNA from a, phenotypically normal, 46,XY female has enabled this study to include the analysis of, what in some cases may be, the reciprocal of the XX male condition. The aetiology of this patient had excluded the possibility of testicular feminisation (Dr. D. Healey, personal communication) and so the female phenotype was presumably due to a disruption of the male determination locus on the Y chromosome. Cytogenetic analysis of this individual, C335, at the Royal Children's Hospital, Melbourne, indicated that the Y chromosome in lymphocytes and fibroblasts was small, although within the normal size range. The small size is most probably due to a reduction in the heterochromatic region of the long arm. No major rearrangements could be detected. The integrity of the Y chromosome present in this individual was examined by blot analysis using the probes described above. The results are shown in Figure 6.10. All of the probes were found to be normally represented. Testis determination is thought to be located quite close to the most proximally located pseudoautosomal locus *MIC2*. If some cases of XX maleness result from simple, contiguous translocations of Y material to the X chromosome, then this would result in the transfer of the *MIC2* locus from the Y chromosome. A non-reciprocal exchange of this kind in an XY female, would result in the loss of the *MIC2* locus from the Y chromosome. The presence of the *MIC2* locus on the Y chromosome of this individual has been confirmed by *in situ* hybridisation of a cDNA for *MIC2* to replication banded metaphase chromosome preparations from fibroblasts. In an

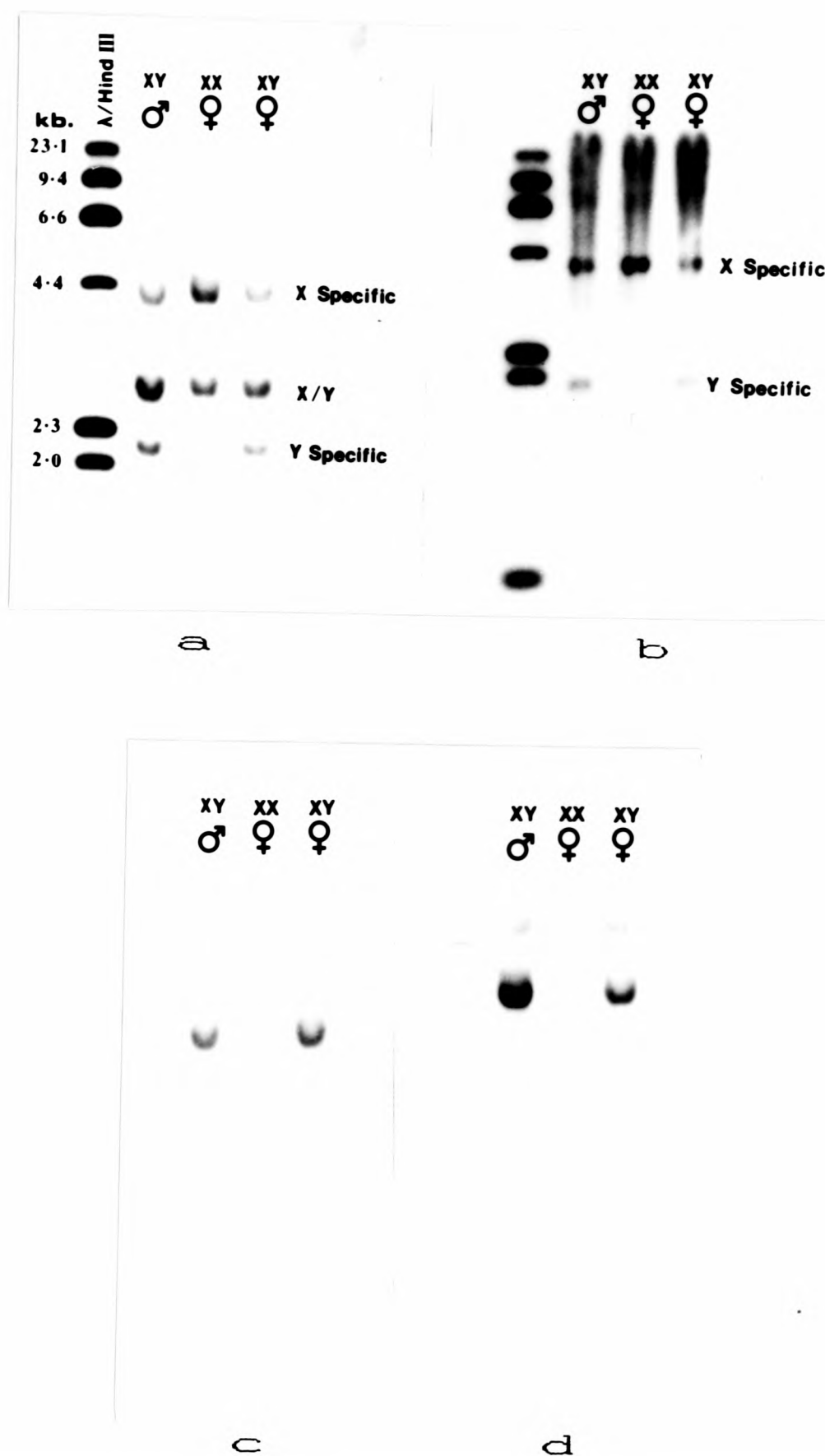


FIGURE 6.10
The analysis of an XY female with Y-specific and X/Y homologous probes.

- a: *EcoRI* digests probed with 75/79 (Probe maps to Yp and Xq. Y-specific band: 2kb).
- b: *BamHI* digests probed with 59y (Probe maps to Yp and Xq. Y-specific band: 1.8kb).
- c: *EcoRI* digests probed with 69/6 (Probe maps to Yq. Y-specific band 3.9kb).
- d: *EcoRI* digests probed with 116/21 (Probe maps to Yq. Y-specific bands 8.7kb & 5.1kb).
- TaqI* digested DNA from the XY female was also probed with 47c and a Y-specific bands was observed (result not shown).

examination of 88 cells, 23 (26%) were observed to have grains at Ypter and 33 (38%) at Xpter (I. & S. Craig, personal communication). No major deletions, or rearrangements, affecting the region of the Y short arm thought to be involved in male determination can therefore be detected in this individual.

6.5 A SEARCH FOR SEQUENCES HOMOLOGOUS TO DXYS27 IN PRIMATES

The analysis of the presence, in the sex chromosomes of primates, of sequences related to DXYS27, might be able to confirm the hypothesis that a block of material has been transposed from the long arm of the X chromosome to the short arm of the Y during recent human evolution. DXYS27 has many features analogous to the 'prototype' of this class of sequence, DXYS1. Sequences homologous to DXYS1 have been found to be X- but not Y-linked in the great apes.

DNA from a range of primates was digested with *Pst*I and probed with 59 γ . Hybridisation was carried out at 60°C and the filter was washed in several changes of 2xSSC/0.1% SDS at 60°C. The result is shown in Figure 6.11. Sequences cross-hybridising with the probe are clearly present in species as far removed as the tamarin and saki (thought to have diverged from the line leading to modern man 40 Myr ago). Cross-hybridising sequences are not present in the lemur, rabbit or mouse, whose probable time of divergence ranges from $\approx$ 70 Myr-100 Myr. Although, the amounts of DNA in each track is not exactly equivalent, e.g. the amount of DNA in the ϕ saki track is much less than in the δ saki track, the dosage effect where amounts are approximately

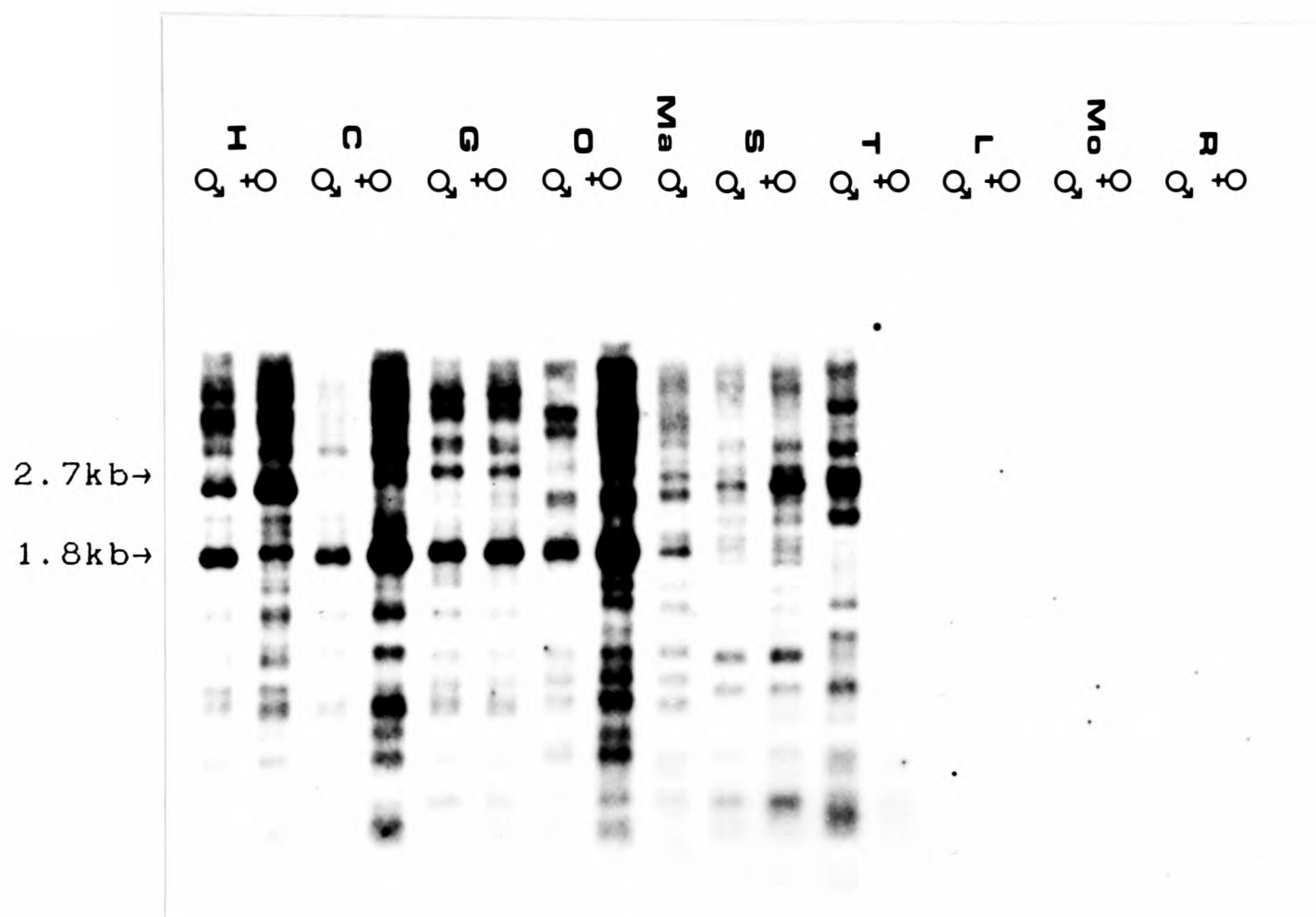


FIGURE 6.11
Conservation of sequences homologous to 59γ
during primate evolution

H: Human; C: Chimpanzee; G: Gorilla; O: Orang-utan;
 Ma: Macaque; S: Saki; T: Tamarin; L: Lemur; M: Mouse; R: Rabbit

DNA from different species digested with *Pst*I and probed with 59γ. In the human tracks the X-specific fragment is 2.7kb and the Y-specific fragment 1.8kb. In the chimpanzee, gorilla and orang-utan a band at 1.8kb is present, however this appears to be X-, but not Y-, linked in these species.

equal is indicative of X-only linkage, except in the human tracks. This is the same filter that was probed with the cDNA for STS (Figure 5.11). The issue is confused somewhat by the presence of multiple bands, due to 59γ containing elements repeated elsewhere in the genome detected at low stringency, but a significant observation is that the Y-specific fragment of 59γ is apparently the same molecular weight as the X-specific fragment in the chimpanzee, gorilla and orang-utan. The equivalent fragment in the other primates is difficult to discern, due to the presence of multiple band of similar intensity. One interpretation of this, is that the X chromosome of the great apes carries the 'prototype' *PstI* restriction fragment, these sites have been maintained on the human Y chromosome after the proposed transposition event, but subsequent mutation has altered the X-specific restriction fragment. It is therefore of great interest to determine the sizes of the restriction fragment detected by 59γ in, for example, chimpanzee DNA digested with the other enzymes showing X-Y polymorphism in humans. A filter was prepared using DNA from ♂ & ♀ human controls and ♂ & ♀ chimpanzee digested with *BamHI*, *BglII*, *EcoRI*, *PstI* and *PvuII*. The filter was probed with 59γ under more stringent conditions than were used above, hybridisation was at 64°C, and was washed to 0.2xSSC/0.1% SDS at 64°C. The result is shown in Figure 6.12. In the cases of *BamHI* and *PvuII*, the X-specific chimpanzee band is the same size as the X-specific human band; for *BglII* and *PstI*, the chimpanzee band is the same size as the human Y-specific band; the band detected in the chimpanzee *EcoRI* digest is the same size as the band common to the X and Y chromosomes in humans.

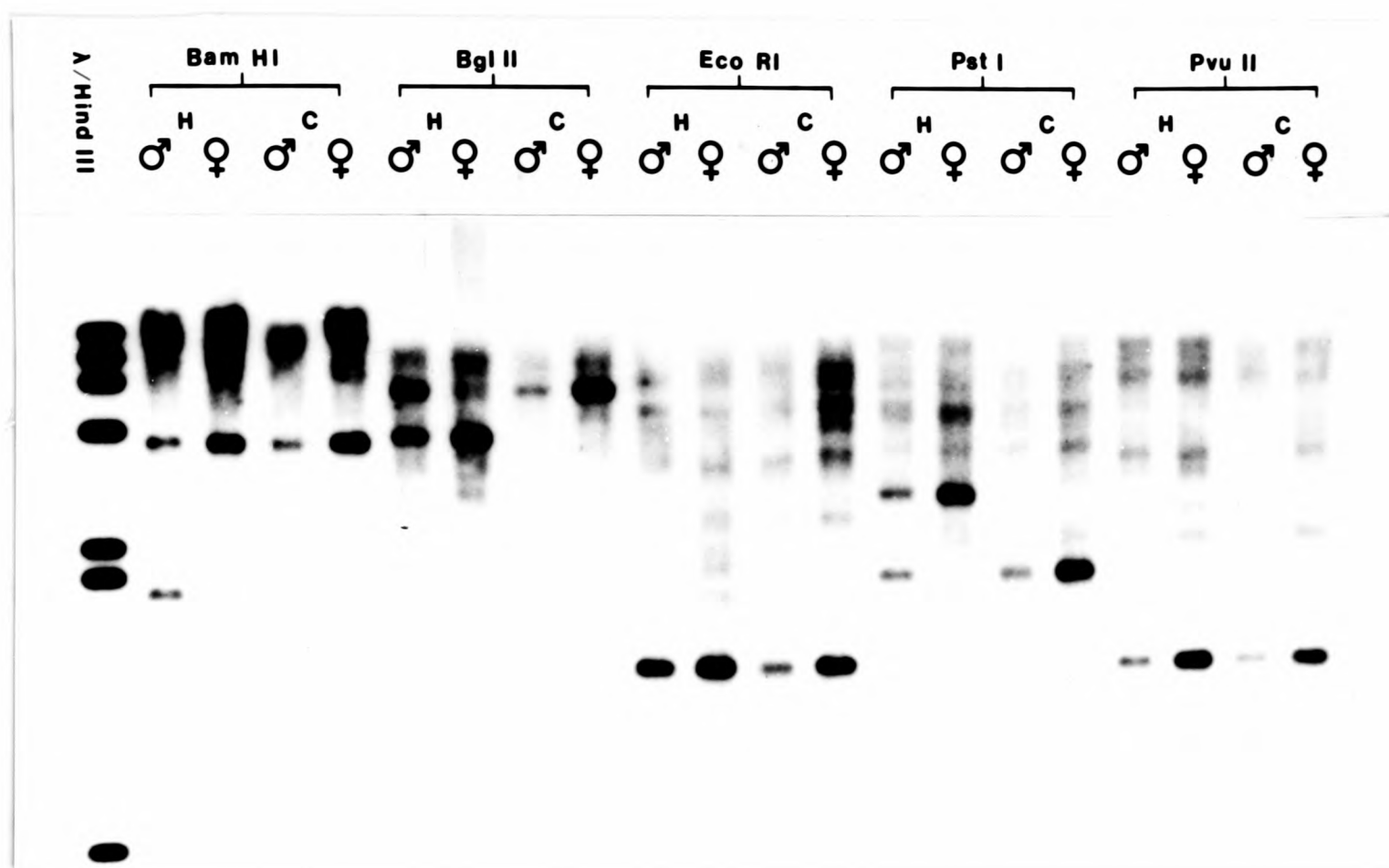


FIGURE 6.12

The sizes of restriction fragments detected in different digests of human and chimpanzee DNA when probed with 59γ.

H: Human; C: Chimpanzee

In the case of *Bam*HI and *Pvu*II, the X-specific chimpanzee band is the same size as the X-specific human band; for *Bgl*II and *Pst*I, the chimpanzee band is the same as the human Y-specific band; the band detected in the chimpanzee *Eco*RI digest is the same size as the band common to the X and Y chromosomes in humans.

Although these results may be coincidental, and if the material was available a more exhaustive study involving more chimpanzee samples digested with a range of enzymes should be carried out, the most economical interpretation of these findings is that the X-Y polymorphism seen in humans is due to mutations that have occurred and been fixed on both the X and Y chromosomes after the human line diverged.

6.6 GENERAL DISCUSSION AND CONCLUSIONS

The locus *DXYS27* is present on the short arm of the Y chromosome and maps to Xq21 on the long arm of the X chromosome. Both the X and Y copies of this locus have been isolated. The degree of homology between the X and Y copies has been estimated by restriction site analysis and direct sequence comparison. Of the 557 residues compared using the two methods 7 differences have been found. This suggests that the sequences are highly homologous ($\approx 99\%$), a result consistent with the observations of other groups. The presence of two other single copy sequences in one of the X-derived cosmids (X595), homologous to sequences present in the Y-derived cosmid (Y59), means that direct sequence comparison of over 6kb in this region is possible.

The presence of X-Y polymorphisms has facilitated the analysis of DNA from XX males and an XY female, for evidence of aberrant X-Y interchange. It appears that along with many other sequences present on Yp, the Y homologue of *DXYS27* is present in many, but not all XX males, presumably as the result of an aberrant X-Y interchange which has also mobilised *TDF*. Since it

is not present in all XX males, it is unlikely that it maps close to *TDF*. The results of *in situ* hybridisation of this sequence to chromosome spreads from a normal male, suggest that it maps proximal to *DXYS25* and so some distance from the proposed locus for testis determination (Buckle *et al.*, 1987). In the XY female examined, no major deletions, or rearrangements, affecting the region of the Y short arm thought to be involved in male determination could be detected. In other studies on XY females, small deletions of Yp have been detected. These have been shown to include the loci detected by 47b, which is adjacent to 47c (Disteche *et al.*, 1986) and *DXYS1*, which is thought to map proximal to *DXYS25* and *DXYS27*. Affara *et al.* (1987) have tested 9 XY females with a group of 19 Yp probes to examine the possibility of X-Y interchange as the causation of this condition. Only one clearly showed evidence of a Yp deletion, and in this case a deletion was detectable cytogenetically. In the case analysed, the lack of male determination could be attributed to a mutation affecting *TDF*, but which is not detectable by the approaches employed, or by mutation at another, as yet unidentified, site in the genome. A further possibility is that the small Y chromosome predisposes non-disjunction and that the primordial germ cells were of a predominantly XO, rather than XY, composition. The general observation of a lack of Yp deletions in XY females is intriguing, since on the basis of the abnormal X-Y interchange model for the genesis of XX males and XY females, one might expect the same proportion of XY females to have deletions of Yp as there are XX males with additional Yp material on Xp. One explanation for this could be that a gamete bearing gross

deletions of Yp has reduced viability, and hence is less likely to fertilise an egg and allow survival of the embryo. Alternatively, the presence of an abnormally small Y chromosome may disrupt spermeogenesis due to an increase in the frequency of non-disjunction events at meiosis. Other hypotheses have been postulated in Affara *et al.* (1987). Whatever the reasons, it seems that in XY females micro deletions and point mutations are the predominant class of lesions on the Y.

The analysis of *DXYS27* homologous loci in primates, although not conclusive, supports the view that this block of X-Y homology is the result of a transposition of material from Xq to Yp during recent evolutionary history. Since the *DXYSn* loci constitute a large block of material (estimated to constitute as much as 1/4 of the eucromatic region of Yp), it is possible that active genes are present at least in the X homologous locus if not on Yp. If the inactivity of this block of material on the Y chromosome is essential to maintain $\delta:\phi$ gene dosage, mutations in this region might have been selected for and are not the result of genetic drift. However, one might therefore expect sequence conservation of the X-homologue with X-Y differences being attributable to mutations accumulating on the Y chromosome. The results described above suggest that mutations have occurred both on the X and Y chromosomes. The presence of a much lower frequency of Y-polymorphism compared to X-polymorphism at these loci also argues against selection for mutations on the Y chromosome.

Although X polymorphisms have not been detected for *DXYS27*, other groups have found such polymorphisms. A number of disease loci have been shown to map to this region of the X

chromosome and through the use of X polymorphisms linkage has been established between *DXYS1* and Charcot-Marie-Tooth disease, Kennedy spinal muscular atrophy, Choroideremia and (tentatively) X-linked dysplasia-gigantism syndrome (McKusick, 1987). More recently linkage has been established between *DXYS1* and X-linked cleft palate (Moore *et al.*, 1987) and one form of X-linked mental retardation (Sutherland *et al.*, 1987). The identification of X polymorphisms for *DXYS27* may be useful in the further analysis of these conditions, and may be useful in the construction of a genetic map of this region.

CHAPTER 7

THE ANALYSIS OF A SEQUENCE (M27B) DETECTING A HYPERVARIABLE LOCUS (DXS255) ON THE X CHROMOSOME

7.1 INTRODUCTION

Linkage analysis based on DNA polymorphisms has localised the genes responsible for several major genetic diseases including Huntington's Chorea (Gusella *et al.*, 1983), polycystic kidney disease (Reeders *et al.*, 1985), and cystic fibrosis (Tsui *et al.*, 1985; White *et al.*, 1985; Williamson *et al.*, 1985) and, in a number of cases, the approach has been found useful in antenatal diagnosis (Weatherall, 1982). In order for this type of analysis to be of value there must be a positive correlation in inheritance of a polymorphic variant with the inheritance of the disease gene concerned. The single-copy human DNA probes normally used to detect restriction fragment length polymorphisms (RFLPs), detect changes that create or destroy a cleavage site for a specific restriction enzyme, causing a change in the length of a DNA fragment (Jeffreys *et al.*, 1979; Cooper & Schmidtke, 1984). The mean heterozygosity of human DNA is low (≈ 0.001 per base pair) (Jeffreys *et al.*, 1979; Cooper & Schmidtke, 1984; Ewans *et al.*, 1981) and so few, if any, restriction enzymes will detect a RFLP at a given locus. Even when detected, a marker locus based on a single such variant will only have two alleles (presence or absence of a cleavage site) and the chance that a parent is heterozygous at this locus is almost always less than 50%. However, it is sometimes possible to detect several such RFLPs

with a given probe very close to one another which may help if the analysis of one site proves to be uninformative. Although nearly 400 polymorphic DNA markers for human chromosomes have been described, the majority have only two alleles and are thus contribute little to analysis of genetic linkage in many families.

A few DNA marker systems have been reported that reveal a DNA restriction fragment whose length is highly variable within the population. The first highly polymorphic locus (*D14S1*) in human DNA was discovered using an anonymous DNA probe selected from a human genomic library (Wyman & White, 1980). At least eight alleles were identified and homozygotes accounted for less than 25% of the small number of unrelated individuals examined. The structural basis for multi-allelic variation at this locus appears to be the presence of a hypervariable region, although the cloning and characterisation of this has been difficult because of its inherent instability in various bacterial, host systems (Wyman *et al.*, 1985). Subsequently, and again by chance, several other highly variable regions have been discovered near the human insulin gene (Bell *et al.*, 1982), the *c-Ha-ras-1* oncogene (Capon *et al.*, 1983), the ζ -globin pseudogene (Proudfoot *et al.*, 1982), and the myoglobin gene (Jeffreys *et al.*, 1985b). In each case, it has been shown that the variable region contains a set of tandem repeats of a short (11-60 bp) oligonucleotide sequence (minisatellites). The polymorphism at these loci results from allelic differences in the number of repeats, arising presumably by mitotic or meiotic unequal exchanges or by DNA slippage during replication and can be detected using any restriction enzyme which does not

cleave the repeat unit. It has further been suggested that minisatellite sequences may represent hotspots for recombinational activity (Jeffreys *et al.*, 1985b) and may be involved in human chromosomal rearrangements (Donlon *et al.*, 1986). The tandem repeat units in a subset of human minisatellites, including the myoglobin locus, share a common 10-15 bp 'core' sequence related to the bacterial *Chi* sequence implicated as a hot spot for *recA*-mediated recombination (Smith *et al.*, 1981). Minisatellite sequences have been found to be associated with the high variability of the pseudoautosomal loci *DXYS14*, *DXYS15*, *DXYS16* and *DXYS20* (Simmier *et al.*, 1987). In line with the function proposed for myoglobin like minisatellite sequences in recombination, it is tempting to relate the high recombinational activity of the pseudoautosomal region to the presence of minisatellites in this region; however, these hypervariable regions are not related to the myoglobin core. Neither do they share any feature with the bacterial *Chi* sequence, nor with any other reported minisatellites. Nevertheless, such sequences are not confined to the pseudoautosomal region, since there appear to be numerous autosomal loci related to the *DXYS15* minisatellite. It has been shown that some limited autosomal regions recombine more frequently during male meiosis (White *et al.*, 1985), and therefore the location of *DXYS15* like minisatellites with respect to these regions is of interest.

Perhaps suprisingly, some hypervariable regions present in several species, including man, can be detected using a probe derived from a short repeat region of the M13 bacteriophage vector (Vassart *et al.*, 1986). The hybridisation pattern

obtained using this probe is clearly different to that found by Jeffreys (1985b), suggesting that a different family of minisatellite sequences is being detected. In some cases, minisatellite sequences appear to be associated with the *Mst II* short interspersed repeat family (Mermer *et al.*, 1987). It has been suggested that the various classes of minisatellites may have arisen by different means and that, in some cases, the interspersed repeat may play an important role in their genesis.

Jeffreys *et al.* (1985b) have demonstrated that a DNA probe based on a set of tandem repeats associated with the myoglobin gene locus detects several loci in human DNA. The restriction fragment pattern revealed by the sum of the loci containing such related sequences, which are apparently scattered throughout the genome, can constitute a genetic 'fingerprint' unique to an individual. These probes, designated 33.6 and 33.15, are becoming important in forensic and parentage studies and immigration disputes in man (Gill *et al.*, 1985; Jeffreys *et al.*, 1985a). They have recently been shown to reveal similar hypervariability in mice (Jeffreys *et al.*, 1987b), dogs, cats (Jeffreys *et al.* 1987a) and birds (Wetton *et al.*, 1987; Burke & Bruford, 1987). The minisatellite probes provide a powerful technique that is likely to be widely used in family, demographic and linkage analyses of many wild and domestic species. However, the large number of restriction fragments that are detected at a number of loci by partial homology to the tandem sequence probe make interpretation of the resultant allelic series very difficult. To make this task easier, minisatellite probes can be used to screen DNA libraries and to

identify clones representing unique loci. Wong et al. (1986) have used the tandem repeat probe to identify and isolate a locus that had been previously shown to cosegregate apparently with the heterocellular form of hereditary persistence of foetal haemoglobin (HPFH) (Jeffreys et al., 1986). Nakamura et al. (1987a) have extended this general approach by using synthetic oligonucleotide sequences from several of the known minisatellites, as well as other candidate sequences, as probes against human genomic libraries and have identified 77 new loci. A minimum estimate of the average heterozygosity shown by 67 of these markers that have three or more alleles is over 70%. Providing the loci are dispersed over the genome, and linkage analysis of 17 loci suggests that they are well distributed over at least ten chromosomes, they will prove very useful in linkage studies and may replace conventional RFLPs with low heterozygosity. This report has now been updated (Nakamura et al., 1987b) and 55 of their polymorphic probes examined for linkage to localised markers. 35 of these showed linkage with markers which had already been localised to chromosomes 1, 2, 3, 4, 6, 7, 9, 12, 13, 14, 15, 17, 18 or 19.

Four cloned locus specific minisatellites, detected by the minisatellite probes 33.6 and 33.15 used in DNA fingerprint analysis, have been regionally localised by *in situ* hybridisation (Royle et al., 1987). The loci detected by three of the probes have been shown to reside as follows, *D7S22*: 7q36→qter, *D5S43*: 5q35→qter, *D12S11*: 12q24.3→qter. The other probe, pms51, hybridises to a single locus which has been localised to 11q13. Another cloned minisatellite, λ ms29, recognises two loci which have been localised to 6pter→p25 and

16p11 by *in situ* hybridisation. It has been suggested that there may be a non-random distribution of highly variable minisatellite loci over the telomeric region of human autosomes, and therefore their general usefulness in linkage analysis may be restricted.

The only evidence of X-linkage of a minisatellite sequence has come from a study by Jeffreys *et al.* (1987b). They investigated the inbred murine lines C57BL/6J (B6) and DBA/2J (D2) and the B x D recombinant inbred strains, established from the F₂ hybrid of the two-way cross about 60 generations previously. An invariant region was always found to give a lower hybridisation signal in males, suggesting its X-linkage.

The most polymorphic X-linked probe to date is described by Oberlé *et al.* (1985). The probe St14, which recognises the locus *DXS52*, hybridises to a series of bands amounting to 65kb, located in the q26-q28 region of the X chromosome. It is suggested that the complex *TaqI* RFLP revealed by the probe and which is defined by a series of at least 8 allelic fragments, could be generated by sequence rearrangement. However, the same polymorphism was not detected with the few other restriction enzymes tested. Alternatively, the polymorphism might be created by independent point mutations in a region dense with *TaqI* sites (as in the case of the complex *SstI* RFLP found in the switch region of the immunoglobulin heavy chain genes (Migone *et al.*, 1983). Three independent *MspI* RFLPs are also detected which are more likely to be due to separate point mutations. To date, this was the best genetic marker on the X chromosome, with about 80% of all females being heterozygous for both the *TaqI* and *MspI* alleles. Assuming no linkage

disequilibrium, the *TaqI* and *MspI* RFLPs would be expected to generate 64 haplotypes and heterozygosity has been estimated to be $\approx 90\%$.

Dietz Band *et al.* (1987) have attempted to identify X chromosome specific highly polymorphic minisatellite markers by hybridising oligonucleotides representative of this type of sequence to an X chromosome specific library. Although many positively hybridising clones have been identified, none of these appears to have been fully characterised and definitive X localisation has not been shown.

The polymorphism shown by the probe M27 β isolated from the MOG cosmid bank, appears to be typical of that displayed by minisatellite sequences. Its behaviour suggests that the single copy probe either resides very close to a region of hypervariability or that such a repeat has been deleted from this sequence. This chapter describes the further characterisation of M27 β , including a refinement of its sub-regional localisation, an analysis of the heterozygosity detected in individuals and an investigation of its possible linkage to the locus for X-linked retinitis pigmentosa.

The locus recognised by M27 β has been designated DXS255 at the Human Gene Mapping Conference 9, held recently in Paris.

7.2 MATERIALS AND METHODS

Cell-lines and hybrids were grown as described in 2.3 and DNA was prepared using the method described in 2.4.1.

DNA from the hybrid 7-8-14-13 was a gift from Ruth Brown. Some of the DNAs used in the polymorphism and family studies were provided by Prof. J.H. Edwards, Dr. A. Miciak & Dr. T. Meitinger. Access to samples from RP families was through collaboration with the University Eye Hospital, Munich.

DNA was prepared from blood samples using the method described in 2.4.2.

All genomic DNA samples were digested overnight with at least a two-fold excess of enzyme, under the conditions recommended by the manufacturer. 2.5-5 μ g samples were routinely run on 0.8% agarose gels for $\approx$ 6 hours at 100mA/80V. Filters were prepared by blotting gels onto HYBOND-N (2.7.2). Conditions for hybridisation and washing of filters were as in 2.10. The temperature of hybridisation and washing was routinely 64°C. All genomic blots were hybridised in the presence of dextran sulphate, this was omitted when probing plasmid/cosmid digests.

Probes were made by nick translation (2.8.1) of restriction fragments eluted from gels (2.7.3)

The subcloning of restriction fragments into plasmid vectors is described in 2.11 and the transformation of *E.coli* in 2.12.

The transfer of bacterial colonies to nitrocellulose is described in 2.6.

Plasmid/cosmid DNA was prepared using the methods described in 2.5.

7.3 OBSERVATION OF MULTI-ALLELIC VARIATION DETECTED BY M27 β

The initial identification of M27 β as an X-specific single copy sequence has been described in Chapter 3. It was during the preliminary studies to sub-regionally localise the sequence

that its hypervariable nature was first noticed. When probing genomic DNA digested with *EcoRI*, instead of obtaining the pattern of hybridisation characteristic of other single-copy sequences i.e. a single cognate band of the same molecular weight as the probe present in controls and hybrids containing the locus recognised by the probe, a more confusing result was obtained. The probe could be localised to the region Xp2.1→Xqter by virtue of the presence of cross-hybridising sequences in the hybrids WAG 8 and UCLA_{B2}. However, these fragments were not the same size as M27β. In fact, none of the fragments detected in the ♂, ♀, 4X, 4Y or even MOG-T control tracks were the same sizes M27β (see Figure 7.1b). Two distinct bands were visible in the ♀ track, which could be accounted for by the presence of two different X chromosomes polymorphic at this locus, whilst three bands were clearly visible in the 4X cell line GM1416, suggesting the presence of at least three different X chromosomes. The other tracks show single bands, but each of a different size, accounted for by the single X chromosomes present in the cell-lines or hybrids. A significant observation was that the *EcoRI* restriction fragment recognised by M27β in the hybrid MOG-T was much larger than the probe itself, ≈5kb as opposed to the expected size of 2.3kb. Although the size differences observed in the other tracks could be accounted for by substantial polymorphism at this locus, the size of the fragment in MOG-T would be expected to be the same as the probe, since this is the hybrid from which the original cosmid bank was derived. This suggests that either M27β contains a deletion of ≈3kb or that it is at one of the extremes of the cosmid insert, and so represents an incomplete

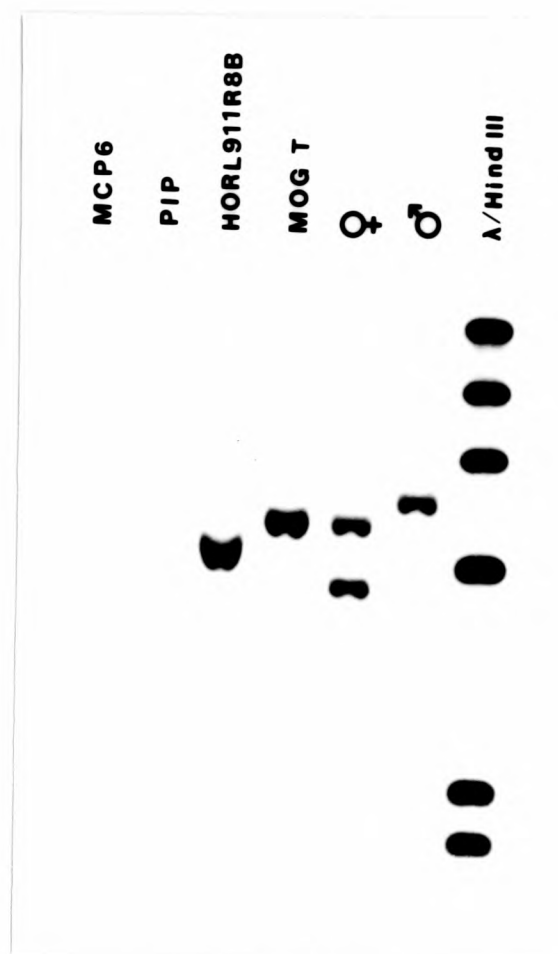
*Eco*RI restriction fragment, possibly containing vector sequences. Results discussed later suggest that M27 β is not juxtaposed to vector sequences, but that a region has been deleted from it during the initial cloning procedure.

7.4 REFINEMENT OF THE SUB-REGIONAL LOCALISATION OF THE LOCUS DETECTED BY M27 β

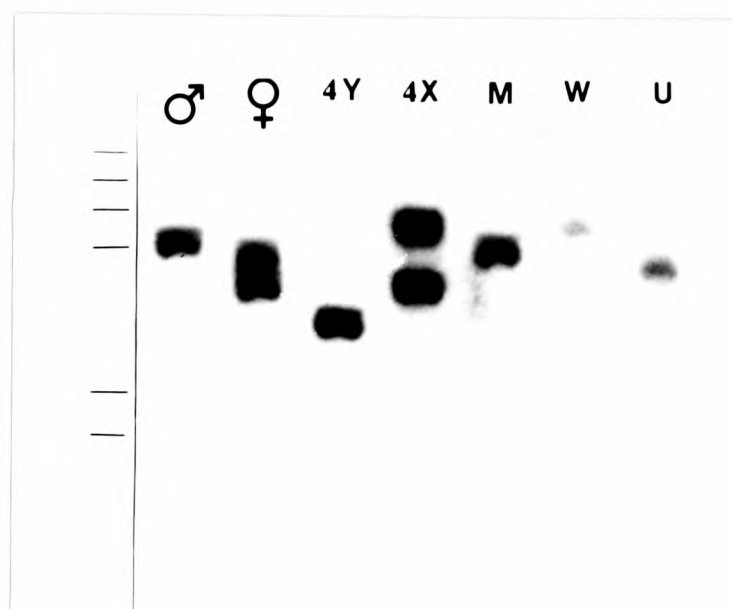
M27 β was sub-regionally assigned to the region Xp11.3-Xq13 using the somatic cell hybrid mapping panel described in 3.2.4 (Figure 7.1a), suggesting that the locus may have a pericentric localisation. This localisation was refined to Xp11.3-Xcen by probing DNA from the hybrid 7-8-14-13. This hybrid is derived from a cell-line with an iso(Xq) chromosome (R. Brown, personal communication). The absence of a signal in this track when probed with M27 β precludes a localisation of this sequence to Xq.

7.5 SUBCLONING OF M27 β INTO THE PLASMID VECTOR pUC9

Since continued elution of M27 β from the cosmid was a laborious process and yields were low, the fragment was subcloned into the plasmid vector pUC9. Rather than using the eluted band as the substrate for the ligation reaction, it was found to be easier to 'shotgun' clone an *Eco*RI digest of the whole cosmid and to identify the desired colonies by probing with a small amount of ³²P labelled M27 β insert. Recombinants were selected by plating out the bacteria on 2xTY/AMP/Xgal plates. 100 white colonies were selected and transferred in a



7.1a



7.1b

FIGURE 7.1
Sub-regional localisation of M27 β .

4X: GM1416; 4Y: OX; M: MOG T; W: WAG 8; U: UCLA_{B2}.. Details of the chromosomal representations of these, and the other hybrids are given in Chapter 3 (e.g. Figure 3.4). All samples were digested with *EcoRI* and the filters probed with M27 β . Using this hybrid mapping panel, M27 β can be mapped to Xp11.3→Xq13. The observation that different fragment sizes are detected in each sample, exemplifies the multi-allelic nature of this sequence.

grid pattern onto two 2xTY/AMP plates. Following growth overnight the colonies were transferred to nitrocellulose and screened with M27 β insert. Positive colonies were identified and grown overnight in liquid culture. *EcoRI* digestion of the isolated plasmids and analysis on a mini-gel showed that they all contained M27 β , but some were found to contain additional fragments as well. A clone that apparently contained just M27 β was used to prepare larger amounts of plasmid. Subsequent restriction analysis of this clone has shown that it contains two copies of M27 β (see Figure 7.12).

7.6 DETERMINATION OF THE CAUSE OF THE MULTI-ALLELIC VARIATION

RFLPs can result from the creation or elimination of a restriction site by base-pair substitution, or modification, or from the movement of sites by rearrangements of DNA segments (deletions, insertions or inversions). Since the preliminary analysis of M27 β had shown that it recognised *EcoRI* restriction fragments of several different sizes present on different X chromosomes, it seemed unlikely that this polymorphism was caused by point mutations, causing the loss or gain of restriction sites, unless this sequence was present in a region of the X chromosome prone to mutation. A more likely explanation is that the polymorphism is caused by an insertion/deletion process similar to that causing the hypervariability of minisatellite sequences discussed in the introduction. If this hypothesis is correct, then the same degree of hypervariability should be observed if the DNA to be probed is digested with any of a number of restriction enzymes,

so long as the enzyme does not cleave within the repeat sequence responsible.

Samples of DNA from two different females and two cell lines containing four X chromosomes were digested with three different enzymes (*EcoRI*, *HindIII* and *PstI*). Fragments were separated on a 0.8% agarose gel run overnight and blotted onto HYBOND-N. The filter was probed with M27 β insert and was washed to a final stringency of 0.1xSSC/0.1% SDS. The result is shown in Figure 7.2. The same degree of polymorphism is found with all three enzymes. In each case when the target DNA was digested with *HindIII*, the probe hybridised to a fragment ≈ 1.5 kb larger than the corresponding *EcoRI* fragment. In the case of *PstI* the difference in size was always ≈ 1.3 kb (see Table 7.1). In addition to these, multi-allelic variation has been demonstrated with the enzymes *AvaII*, *BglII*, *HindII*, *PvuII* and also with *BamHI*, *EcoRV*, *Sall* and *TaqI*, but the alleles detected are too large to be easily resolved (e.g. see Figure 7.6). These results suggest that the polymorphism is indeed due to an insertion/deletion phenomenon.

7.7 ANALYSIS OF THE DEGREE OF HETEROZYGOSITY AT THE LOCUS *DXS255*

The degree of heterozygosity at this locus was analysed by probing *EcoRI* digests of DNA from 18 unrelated females. A filter was prepared from $\approx 2.5\mu\text{g}$ samples of each digest and was probed with M27 β insert. The filter was washed to a final stringency of 0.1xSSC/0.1% SDS. Results are shown in Figure 7.3; the differences in intensity of bands in the different

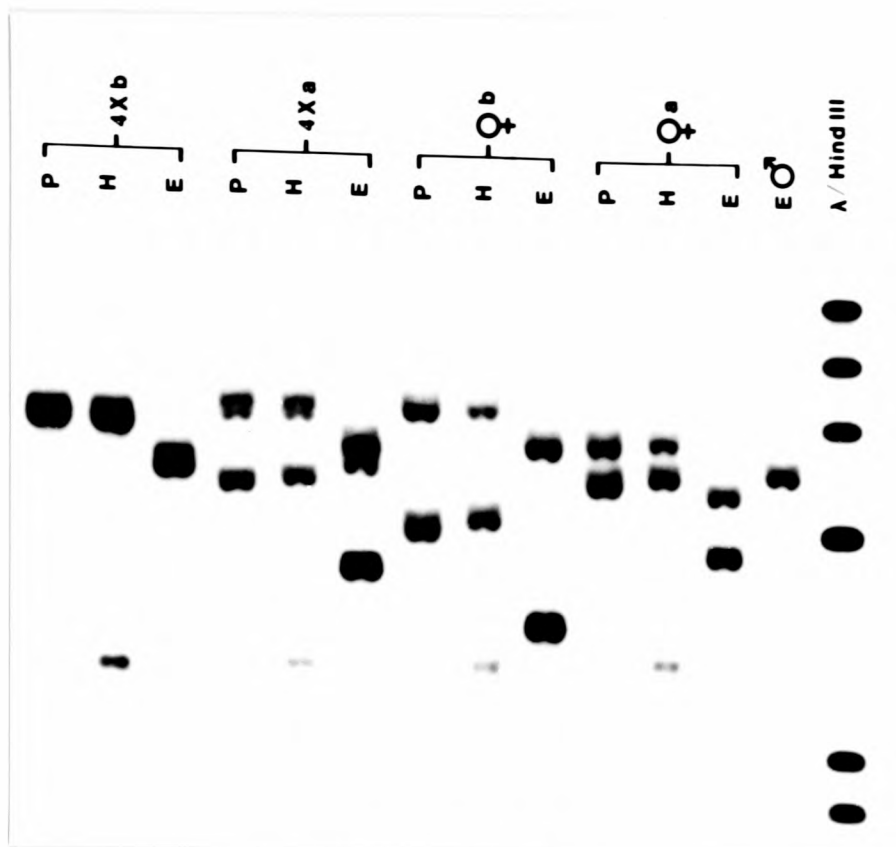


FIGURE 7.2
Determination of the nature of the multi-allelic
variation at *DXS255*.

Genomic DNA samples from two unrelated females and two 4X cell-lines digested with three different enzymes (E: *EcoRI*; H: *HindIII*; P: *PstI*) and probed with M27 β . Multi-allelic variation is seen with each of the enzymes. In the case of *HindIII*, a constant fragment of 2.7kb is detected in each sample in addition to variable fragments. The size differential of the variable fragments is shown in Table 7.1. Molecular weights were calculated from the migration of the λ /*HindIII* standard fragments.

Sample	<i>EcoRI</i> Fragments	<i>HindIII</i> Fragments	<i>PstI</i> Fragments	Difference in Size
qa	5.0	6.4	6.2	$\approx 1.1\text{kb}$
	3.8	5.3	5.1	
qb	6.2	7.7	7.4	$\approx 3.2\text{kb}$
	3.0	4.4	4.2	
4Xa	6.3	7.8	7.7	$\approx 0.8\text{kb}$
	5.5	7.0	6.9	$\approx 2.0\text{kb}$
	3.7	5.3	5.2	
4Xb	6.0	(7.6)	(7.6)	$\approx 0.7\text{kb}$
	5.4	(6.9)	(6.9)	

TABLE 7.1

In each case, the size differential of the variable fragments detected remains constant for whichever enzyme is used to digest the sample. In each case, the fragments detected in a *HindIII* digest are $\approx 1.5\text{kb}$ larger than the corresponding *EcoRI* fragments; in a *PstI* digest the fragments are $\approx 1.3\text{kb}$ larger. Together, these results suggest that the polymorphism detected is the result of an insertion/deletion phenomenon altering the size of the restriction fragments.

tracks is reflects the slightly differing amounts of DNA present. The allelic variation shown is very marked. Approximate molecular weights of the different alleles were calculated (Table 7.2), the resolution of the gel is not sufficient for this to be an accurate measurement. Nevertheless, it seems that by comparing bands in the same or adjacent lanes, there are between 11 and 17 different fragment lengths present out of the 34 bands visible on the autoradiograph. These range in size from 3.5-7kb. The interpretation of these data is that individuals with two bands are heterozygous for two fragment lengths (alleles) and individuals with only one band are homozygous for a single allele or contain two alleles that have not been resolved. Males only ever have one allele and may be termed hemizygous. 16/18 of the females shown in Figure 7.3 are heterozygous at this locus, suggesting a heterozygosity of $\approx 90\%$. Subsequent analysis of other DNA samples has suggested that this estimate might be low rather than high. This range in size variation of the alleles seems to be typical, with few being found greater than 7kb and the smallest allele found to date being $\approx 3\text{kb}$, no *EcoRI* fragments the same size as M27 β itself have been found during the analysis of over 100 individuals. The same filter was probed with the two other single copy sequences from cosmid M27 (M27 α and M27 β) under conditions similar to those described above. Both probes gave single bands of approximately the same molecular weight calculated for the probe (Figure 7.4).

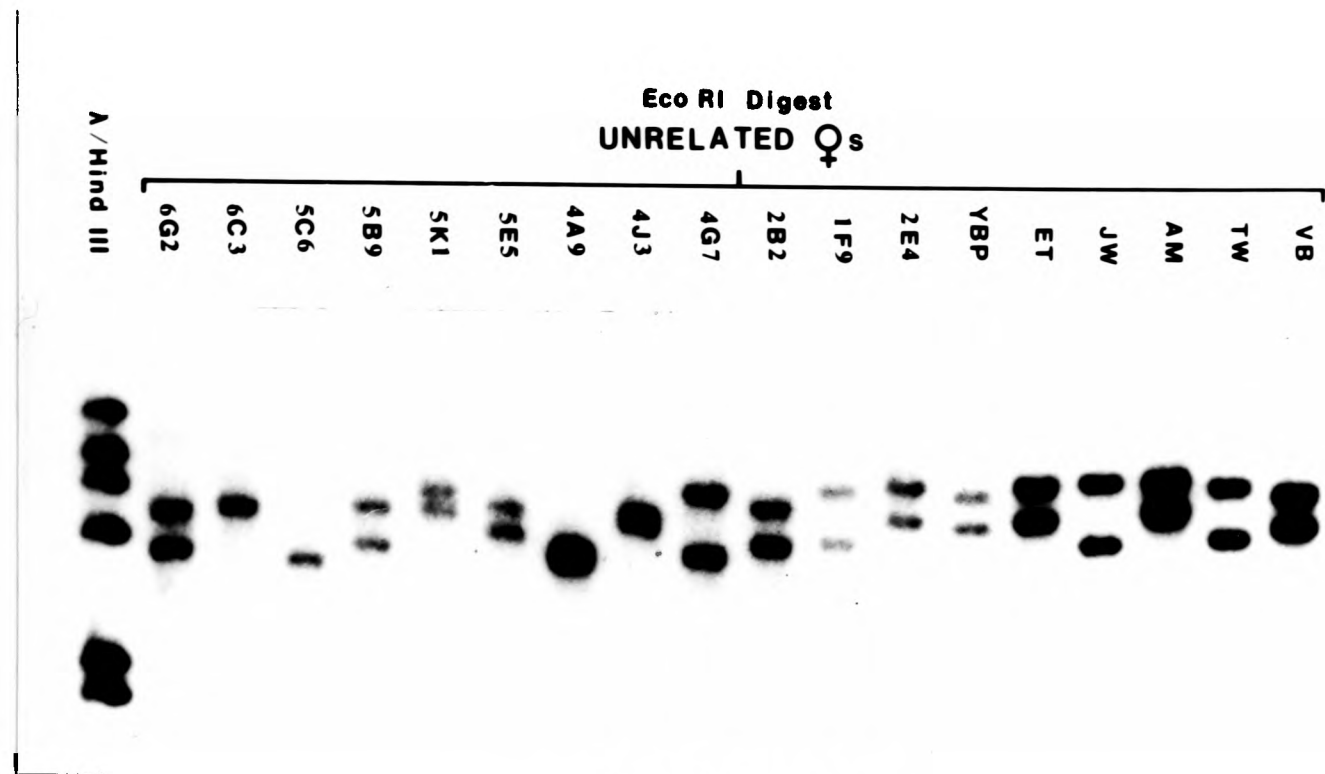


FIGURE 7.3
Determination of the heterozygosity at *DXS255*.

Genomic DNA samples from 18 unrelated qs digested with *EcoRI* and probed with M27 β . 16/18 appear to be heterozygous (exceptions being 6C3 and 5C6). The approximate molecular weights of the different alleles, calculated from the migration of the λ /*HindIII* standard fragments, are shown in Table 7.2.

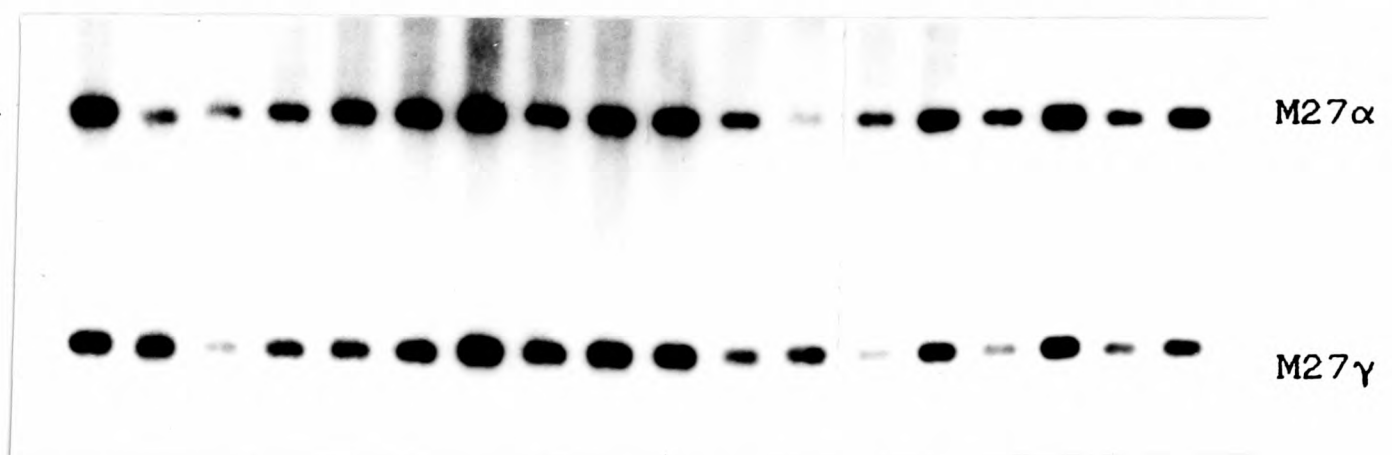


FIGURE 7.4

The same DNA samples probed with M27 α and M27 γ . These probes, from the same cosmid, do not detect multi-allelic variation.

FEMALE	SIZE OF ALLELES	FEMALE	SIZE OF ALLELES
6G2	3.7, 4.9	2B2	4.0, 5.6
6C3	5.2	1F9	4.2, 6.6
5C6	3.5	2E4	4.9, 6.9
5B9	3.9, 5.4	YBP	4.7, 6.4
5K1	5.2, 6.4	ET	4.9, 6.9
5E5	4.4, 5.2	JW	4.2, 6.9
4A9	3.4, 3.6	AM	4.9, 6.9
4J3	4.7, 5.2	TW	5.1, 6.6
4G7	3.5, 6.4	VB	4.7, 6.4

TABLE 7.2

Molecular weights (in kb) of the different alleles detected by M27 β in the 18 samples shown in Figure 7.3. Although these estimations are probably not very accurate, there appear to be 17 different fragments detected in the size range 3.5–7kb.

7.8 INVESTIGATION OF THE INHERITANCE OF RFLP

In order to examine the heritability of restriction fragment length, the DNAs from seven members of a family spanning three generations were examined. $\approx 3\mu\text{g}$ of DNA from each member was digested with *EcoRI*. A filter was prepared and probed with M27 β . A pedigree showing the familial relationships and the DNA restriction patterns is shown in Figure 7.5. Although some of the fragment lengths are very similar (if not identical) in size, they are consistent with their inheritance as X-linked Mendelian alleles through three generations. Specifically, the fragment pattern of individual 5G0 (♀) can be accounted for by the inheritance of the higher molecular size fragment from 2H6 (♀) and the single paternal fragment from 2G6 (♂). The fragment pattern of 5A8 (♂) is consistent with the inheritance of the higher molecular weight maternal fragment from 5G0, whilst 5K7 (♂) has inherited the lower molecular weight fragment from 5G0. 5F0 (♀) appears to be an incompletely resolved heterozygote for fragments of similar lengths, one of these is the higher molecular weight maternal fragment from 5G0 and the other is the single paternal fragment from 5H8 (♂). Its Mendelian mode of inheritance suggests that the polymorphism detected is based on sufficiently stable allelic variation for it to be of use in linkage analysis.

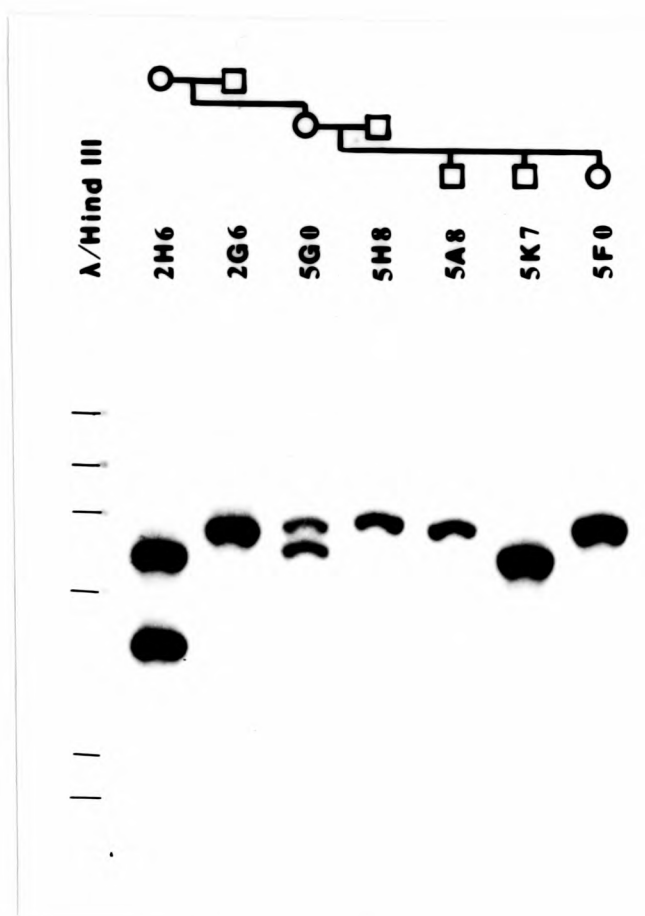


FIGURE 7.5

Mendelian inheritance of fragments detected by M27β.

Pedigree showing the familial relationship of seven individuals and *Eco*RI digests of genomic DNA from these individuals probed with M27β. Although some of the fragments detected are very similar in size (if not identical), they are consistent with their inheritance as X-linked Mendelian alleles through three generations.

7.9 THE POTENTIAL USE OF THE PROBE M27 β IN LINKAGE ANALYSIS

7.9.1 Introduction

The sub-regional localisation of M27 β to the pericentric region of the X chromosome (Xp11.3→Xcen), means that the probe may be of use for linkage analysis of disease loci thought to reside in this region. Because the level of recombination between sequences near to the centromere is thought to be lower than that between sequences located more distally (Drayna et al., 1985), it may be applicable to loci on proximal Xq as well as those on proximal Xp. Loci in this region include those for retinitis pigmentosa (Xp11.3), incontinentia pigmenti (Xp11), Menkes syndrome (Xp11-q11), testicular feminisation (Xp11-q11) and Wiskott-Aldrich syndrome (Xp11-q12).

7.9.2 Investigation of the linkage between *DXS255* and the locus for retinitis pigmentosa

Blood samples from four large German and Yugoslavian families in which retinitis pigmentosa was known to be segregating had become available, therefore it was decided to analyse DNA isolated from these samples with M27 β , to see if any of the alleles detected by the probe could be linked to the disease locus.

Blood samples were kindly collected by Dr. T. Meitinger, who also helped in the preparation of DNA using techniques previously described.

Retinitis pigmentosa (RP) is a group of retinal degenerations characterised by progressive visual field loss, night blindness and pigmentary retinopathy. Its prevalence is

in the region 1-2 in 5,000 of the general population. 36-38% of RP patients are isolated cases, the remainder show autosomal dominant, autosomal recessive or X-linked modes of inheritance. The X-linked variety (XLRP) is found in 14-22% of RP families in the U.K (these details are reviewed in Bhattacharya *et al.*, 1985). Two possible loci for XLRP have been proposed, one at Xp11.3 and another at Xp21. Linkage analysis using RFLP probes spanning Xp21 to Xcen have been used by several groups to try and refine the position of these loci (reported at the Human Gene Mapping Conference 9, 1987). Linkage of XLRP to a RFLP identified by the DNA probe L1.28 was first established by Bhattacharya *et al.* (1984). In this analysis they examined 5 kindreds and found that 9 individuals out of a total of 28 obligate heterozygotes screened were informative for the L1.28 polymorphism. In four of the kindreds there were no definite recombinants. In the fifth kindred one definite recombinant was detected. When the data from all the kindreds were combined the genetic distance between the XLRP locus and L1.28 was estimated at 3cM with a 95% confidence interval of 0-15cM. Somatic cell hybrid mapping had assigned L1.28 to close to Xp11.3, suggesting that the locus for RP1 is on the proximal part of Xp. It was suggested that the probe would possibly be of value in carrier detection and prenatal diagnosis in about 40% of cases. One of the limitations being that not all carriers would be informative (heterozygosity of the two *TaqI* alleles detected by L1.28 had been estimated to be $\approx 41\%$). Subsequent studies have confirmed this linkage, though there is significant evidence of genetic heterogeneity from both genetic and clinical observations. A second locus at Xp21 has been

proposed, through the observed tight linkage of *XLRP* with the locus for *OTC* in several families (reported at HGM9).

Since M27 β detects such a high degree of polymorphism, with heterozygosity estimated at over 90%, nearly all the individuals in any particular kindred could be informative in this type of analysis and if linkage is established within a family, then detection of suspect carriers and prenatal diagnosis of affected individuals should be possible.

Pedigrees of the individuals analysed in this study are shown in Figures 7.6-7.10, along with resultant autoradiographs from *EcoRI* digests of DNA from these individuals probed with M27 β . All the families are informative and results are consistent with simple X-linked Mendelian inheritance of the different alleles detected. Only obligate carriers and one additional carrier that had been diagnosed by clinical testing are shown in the pedigrees. For the purposes of linkage analysis, the meioses of obligate heterozygotes resulting in affected or unaffected sons are the only ones scored, unless a daughter has been clinically diagnosed as such, or has been shown to be an obligate carrier, due to there being affected males in subsequent generations.

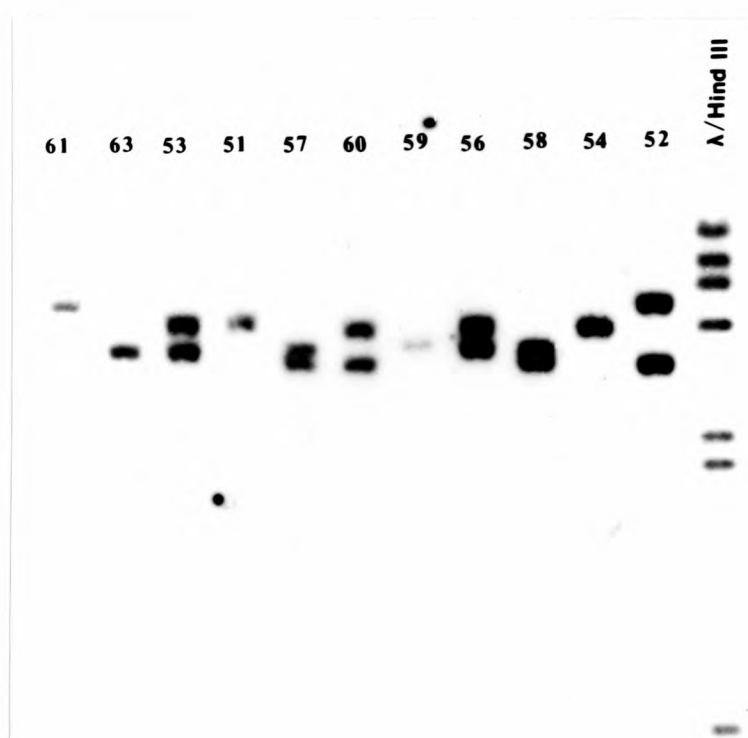
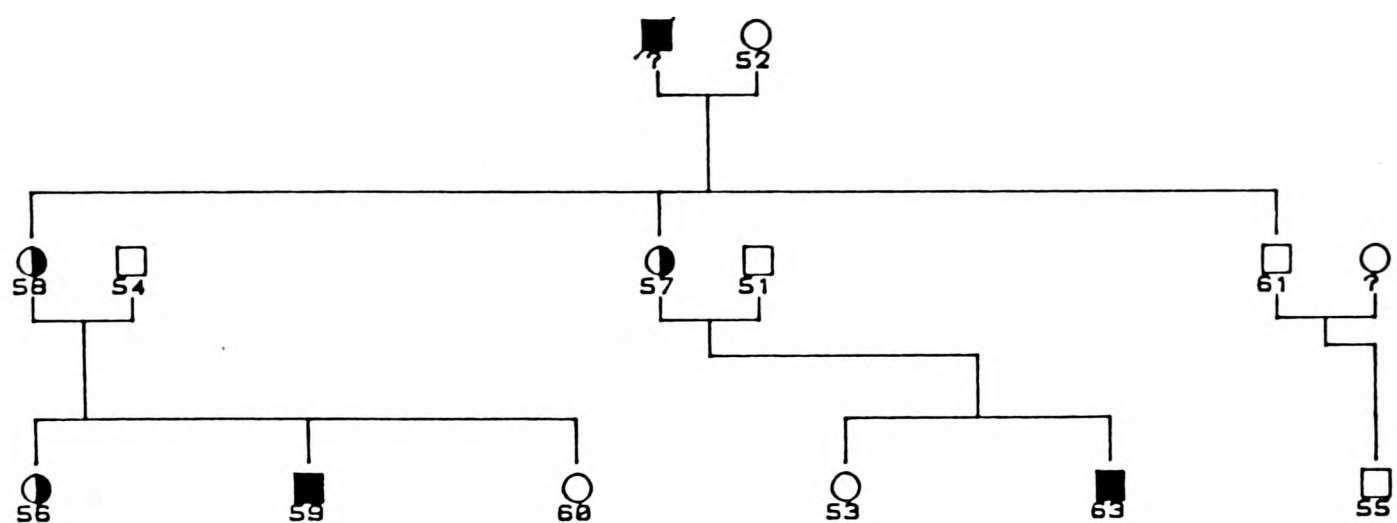


FIGURE 7.6

KEY:

- - Normal ♀ ◐ - Carrier ♀
- - Normal ♂ ■ - Affected ♂
- ∅ ▣ - Deceased
- ? - Sample not available

PEDIGREE A

From the pedigree it can be seen that ♀s 57 and 58 are obligate carriers since they are daughters of an affected ♂. Both have alleles of M.W. 3.4kb and 3.0kb. In both cases the affected sons of these individuals (63 and 59) have inherited the 3.4kb allele, suggesting that in this family the disorder is co-segregating with the 3.4kb allele. ♀ 56 has been clinically diagnosed as a carrier and has also inherited the 3.4kb allele. Thus, for the purpose of linkage analysis there are 3 informative meioses with no cross-overs detected. Additionally, ♀ 53 also appears to be a carrier.

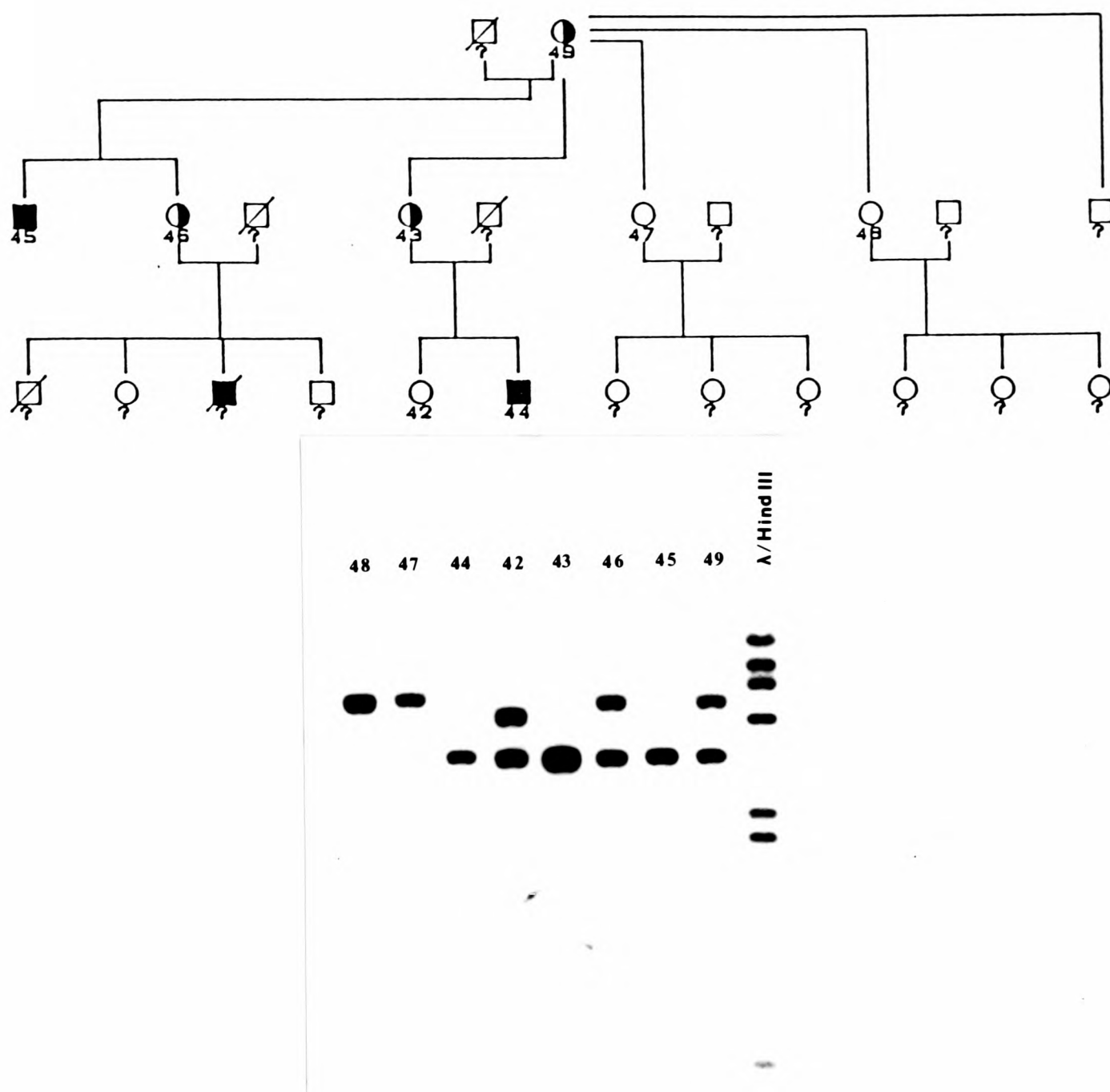


FIGURE 7.7

KEY:

- - Normal ♀ ◐ - Carrier ♀
- - Normal ♂ ■ - Affected ♂
- ⊘ ⊠ - Deceased
- ? - Sample not available

PEDIGREE B

q 49 is an obligate carrier, because she has had an affected son and daughters that have had affected sons in subsequent generations. She has alleles of M.W. 4.8kb and 2.8kb. The affected son, 45, and both the carrier daughters, 43 and 46, have the 2.8kb allele, suggesting that in this family the disorder is co-segregating with the 2.8kb allele. The affected ♂ 44 has also inherited this allele. In this family 4 informative meioses have been scored with no cross-overs detected. In addition, q 42 appears to be a carrier whereas qs 47 and 48 do not.

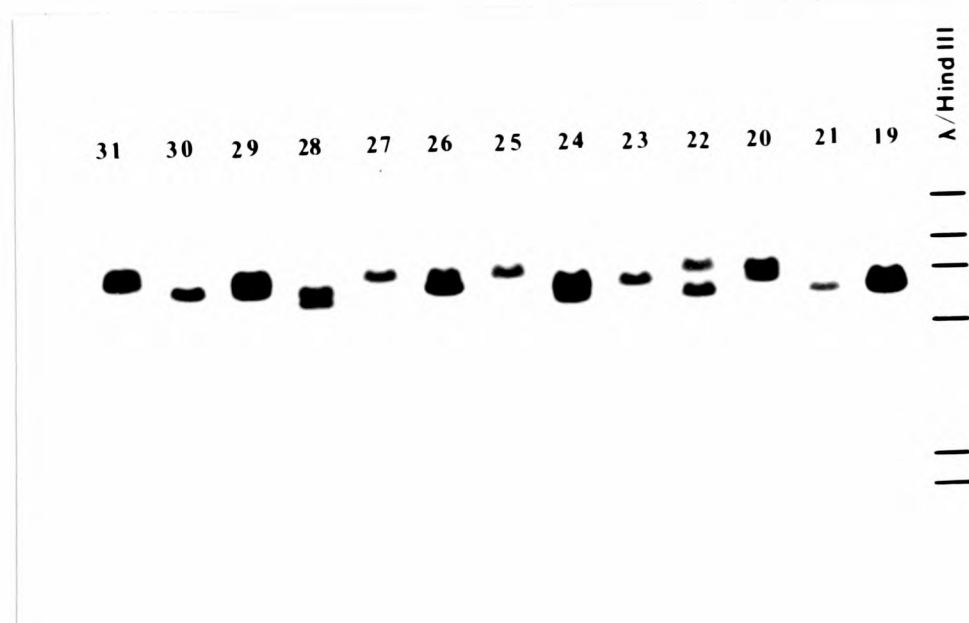
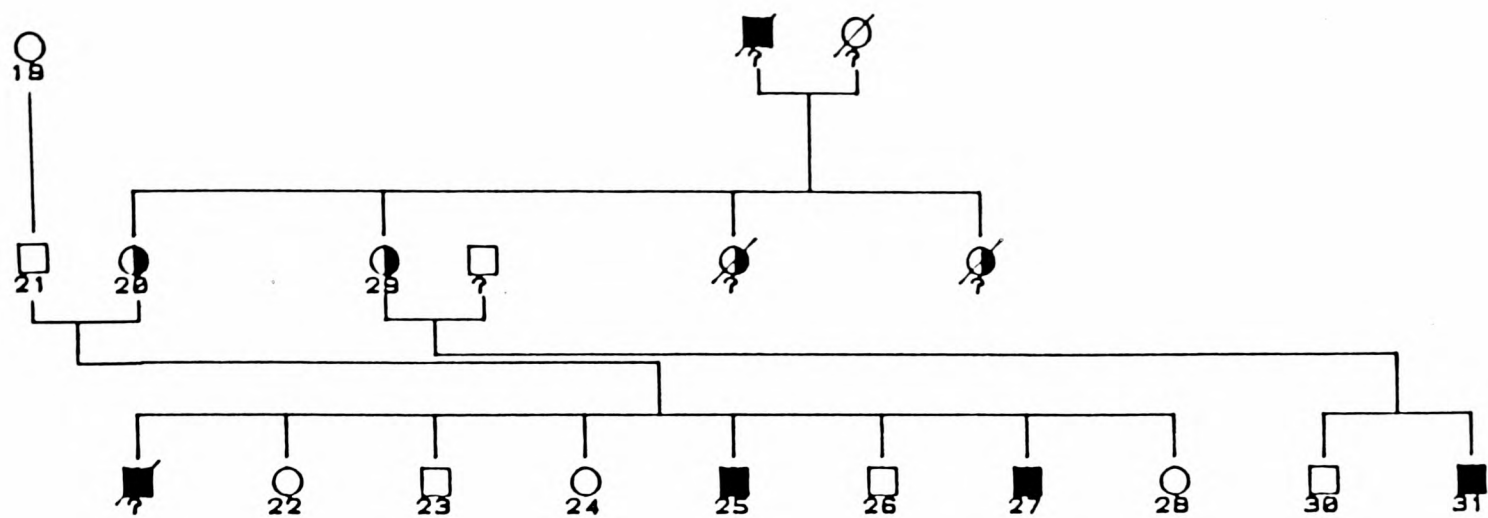


FIGURE 7.8

KEY:

- - Normal ♀ ◐ - Carrier ♀
- - Normal ♂ ■ - Affected ♂
- ⊘ ⊠ - Deceased
- ? - Sample not available

PEDIGREE C

qs 20 and 29 are obligate carriers since they are daughters of an affected ♂. Both have alleles of M.W. 5.7kb and 6.5kb. The two affected sons of ♀ 20, 25 and 27, and the affected son of ♀ 29, 31, have inherited the 6.5kb allele, suggesting that in this family the disorder is co-segregating with the 6.5kb allele. The unaffected sons, 23, 26 and 30, have inherited the 5.7kb allele. In this family it is possible to score 6 informative meioses with no cross-overs detected. In addition the data suggest that ♀ 22 is a carrier, whereas qs 24 and 28 are not.

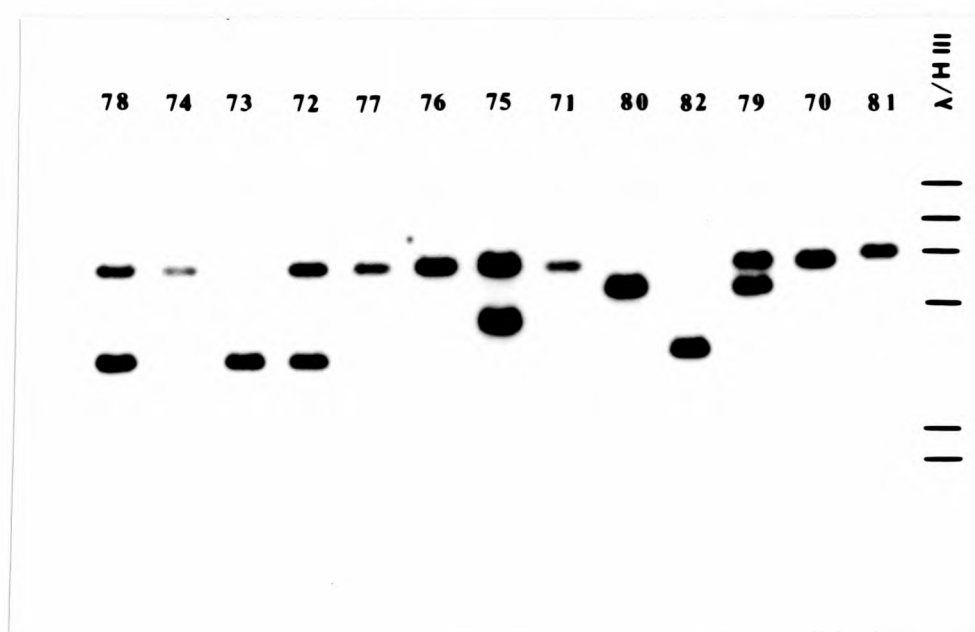
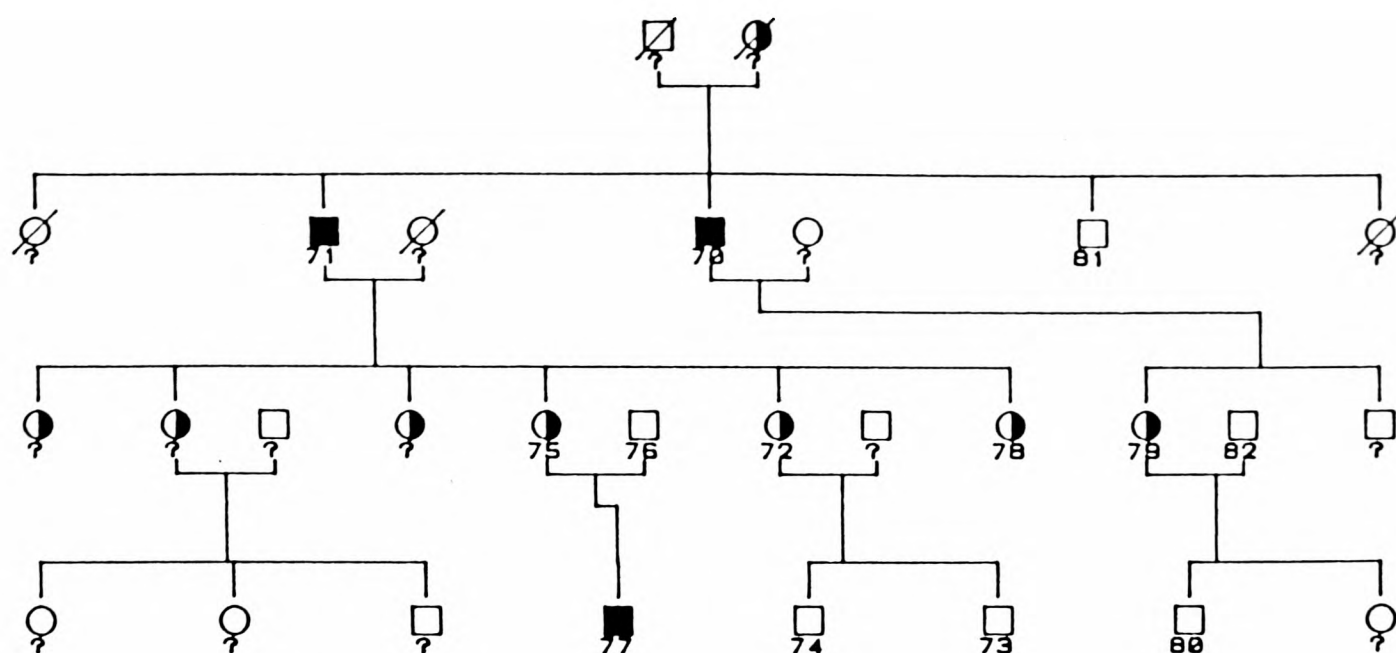


FIGURE 7.9

KEY:

- - Normal ♀ ◐ - Carrier ♀
- - Normal ♂ ◼ - Affected ♂
- ⊘ ⊠ - Deceased
- ? - Sample not available

PEDIGREE D

The affected ♂s 71 and 70 have both inherited a 5.5kb allele from their carrier mother, whereas their unaffected brother, 81, has not, suggesting that in this family the disorder is co-segregating with the 5.5kb allele. All daughters of the affected brothers are obligate carriers and consequently have the 5.5kb allele. ♂ 77, the affected son of ♀ 75 has this allele. ♂ 73 has not inherited this allele from ♀ 72, however his brother, 74, has. The diagnosis of ♂ 74 as unaffected is not definitive, since he is quite young and a more detailed medical examination is necessary to confirm his status. The result of this is not known and this individual has not been scored in the linkage analysis. ♂ 80, is unaffected and has not inherited the 5.5kb allele. If the males in the second generation are scored, even though it is not possible to confirm their mother's alleles, 6 informative meioses can be scored with no definite cross-overs.

FIGURE 7.10

PEDIGREE E

qs 90, 1 and 33 are obligate carriers, being the daughters of an affected ♂. q 90 has alleles of M.W. 4.9kb & 3.3kb, q 1: 4.9kb & 3.3kb and q 33: 4.9kb & 6.1kb. The only allele in common is 4.9kb, suggesting that in this family the disorder is co-segregating with this allele. This being the case, one can assume that the deceased mother of ♂s 99 and 7 also carried this allele. ♂ 99 is affected and has inherited this allele, whereas ♂ 7 is unaffected and has not. q 93 is an obligate carrier since she has two sons that are affected, all of these individuals have inherited the 4.9kb allele. However, ♂ 100 has also inherited this allele from q 93, but is unaffected. This individual is over 21 years old and so it seems very unlikely that he will develop the disorder. Consequently, this can be taken as evidence for a cross-over. q 33 has two affected sons, 131 and 35, that have inherited the 4.9kb allele and an unaffected son, 40, that has not. The carrier status of q 8 is unknown, however she does have the 4.9kb allele and this has been passed on to both of her sons, 10 and 11. These boys may be too young for the disorder to be expressed yet and are currently undergoing medical examination to try and definitively determine their status. These have not been included as informative meioses. Nevertheless, it has been possible to score 9 informative meioses with definite evidence for one cross-over. In addition, qs 3, 5, 6, 13, 15, 18, 37, 38 and 92 are possible carriers. Two anomalous results are present: 1) the pattern obtained for ♂ 14, shown as (?), suggests the loading of the wrong DNA sample, since the two allele pattern is not consistent with the individual being male and neither of the alleles are present in the daughters 2) although q 93 has inherited her maternal allele from q 90, the paternal allele is not consistent with it coming from her 'father', ♂ 91. The patterns obtained with her sisters, 93, 8 and 13 are consistent with the inheritance of a maternal allele from q 90 and a paternal allele from ♂ 91, suggesting that ♂ 91 is not the father of q 93.

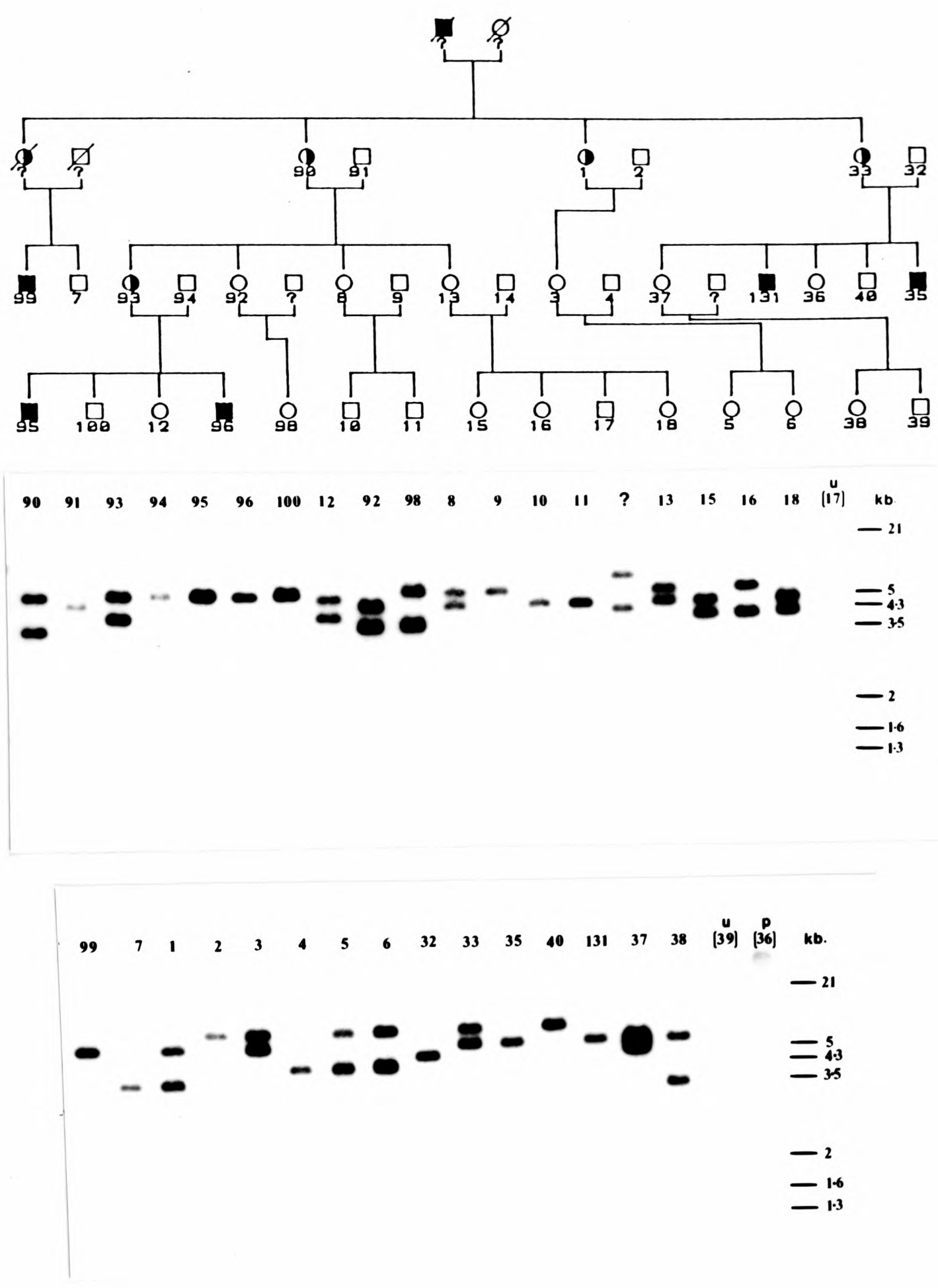


FIGURE 7.10

KEY:

- - Normal ♀
- ◐ - Carrier ♀
- - Normal ♂
- - Affected ♂
- ∅ □ - Deceased
- ? - Sample not available

u - undigested DNA sample
 p - partially digested DNA sample

Pooling all the data together from the different families, it is possible to score 28 informative meioses with evidence for only one definite cross-over. Using these data, the genetic distance between *DXS255* and the *XLRP* locus can be estimated using an 'exact' confidence limits table. The 95% confidence limits are 0.09-18.35 centimorgans.

The number of informative meioses will be increased by a detailed medical examination of putative carriers and sons under 10 years of age, some of these are currently being undertaken at the University Eye Hospital, Munich. The families are also being analysed using the probe L1.28 and probes from the *DMD* region to try and establish how closely linked these markers are to the *XLRP* locus in these families, (Dr. T. Meitinger, personal communication). It may not be wise to pool all of the data to come up with this figure for linkage, because it is not yet known whether the same locus for retinitis pigmentosa is involved in all cases. Even if the same locus is involved, the observation of a different allele co-segregating with the disorder in the five families, suggests that different mutations are responsible in each case. Alternatively, if a single mutation of the *XLRP* locus is responsible, then it no longer co-segregates with the same allele due to cross-overs in previous generations.

Nevertheless, these preliminary data suggest that probe is worthy of further investigations with respect to the analysis of *XLRP* families. They highlight that the probe may be a powerful tool if linkage is established between this, or any other disorder, due to the great number of informative meioses and simplicity of analysis.

7.10 COMPARISON OF GENOMIC SEQUENCES IN THE REGION OF M27 β WITH THOSE PRESENT IN THE COSMID M27

A comparison of restriction sites present in the cosmid M27 with those in the genome, might reveal whether the reason for M27 β being $\approx$ 3kb smaller than the *EcoRI* restriction fragment detected in MOG-T is due to a deletion that occurred during the cloning procedure, or the juxtaposition of M27 β to vector sequences.

By probing samples of σ and ϕ DNA digested with different restriction enzymes, it was possible to deduce that in some cases in addition to variable fragments, M27 β hybridised to fragments that were of invariant size in different individuals (see Figure 7.11). Three enzymes that showed this phenomenon were *HindIII* (2.7kb), *BglIII* (1.9kb & 0.8kb) and *PvuII* (2.4kb). It was decided to use these enzymes for the construction of the map.

A map of the plasmid was constructed using different combinations of these three enzymes (Figure 7.12). As is shown, the plasmid actually contains two copies of M27 β . The *BglIII*-*BglIII* fragment of $\approx$ 0.8kb corresponds to one of the invariant bands found. This suggests that sequences to the left of *BglIII*¹ or to the right of *BglIII*² are probably invariant in different individuals and that the hypervariable sequences are found in the opposite direction. To test this hypothesis, the 0.9kb *BglIII*²-*EcoRI*² was eluted from an agarose gel and used as a probe against restriction digests of the cosmid; M27 β was also used to probe the same digests. The results, shown in Figures 7.13b & 7.13c, show that M27 β hybridises to fragments

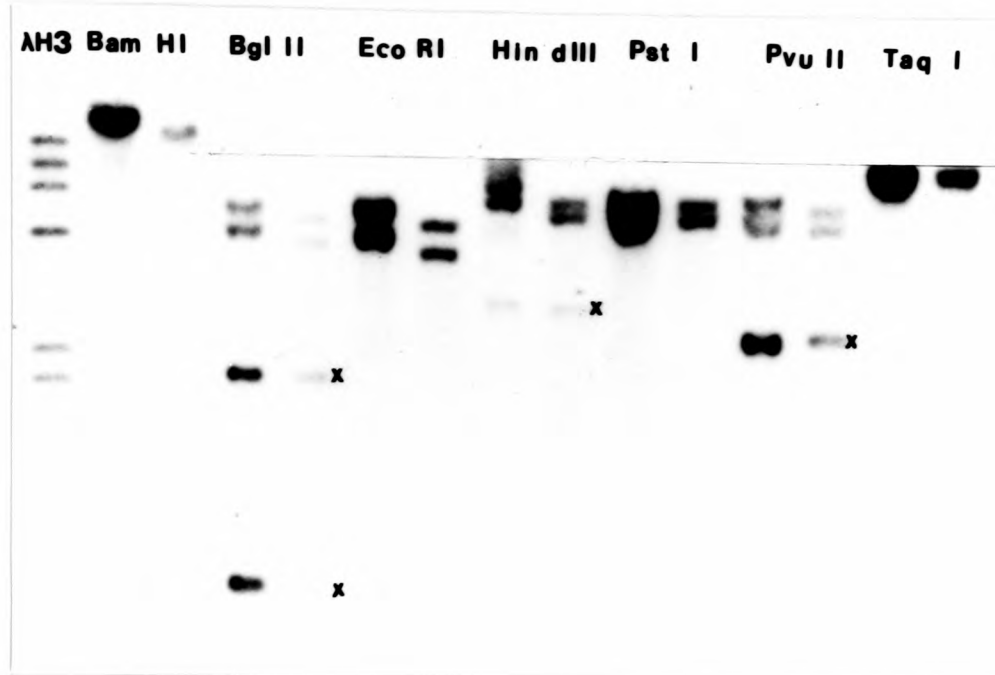


FIGURE 7.11

Polymorphism at *DXS255* detected by different enzymes.

DNA from two different females digested with *Bam*HI, *Bgl*II, *Eco*RI, *Hin*dIII, *Pst*I, *Pvu*II & *Taq*I and probed with M27 β . Polymorphism is detected in each case, with the exceptions of *Bam*HI and *Taq*I where the fragments detected are large and if polymorphism is detected with these enzymes, the fragments have not been resolved on this gel. In the cases of DNA digested with *Bgl*II, *Hin*dIII and *Pvu*II, as well as variable fragments being detected, there are fragments that are of invariant length in different individuals (marked x).

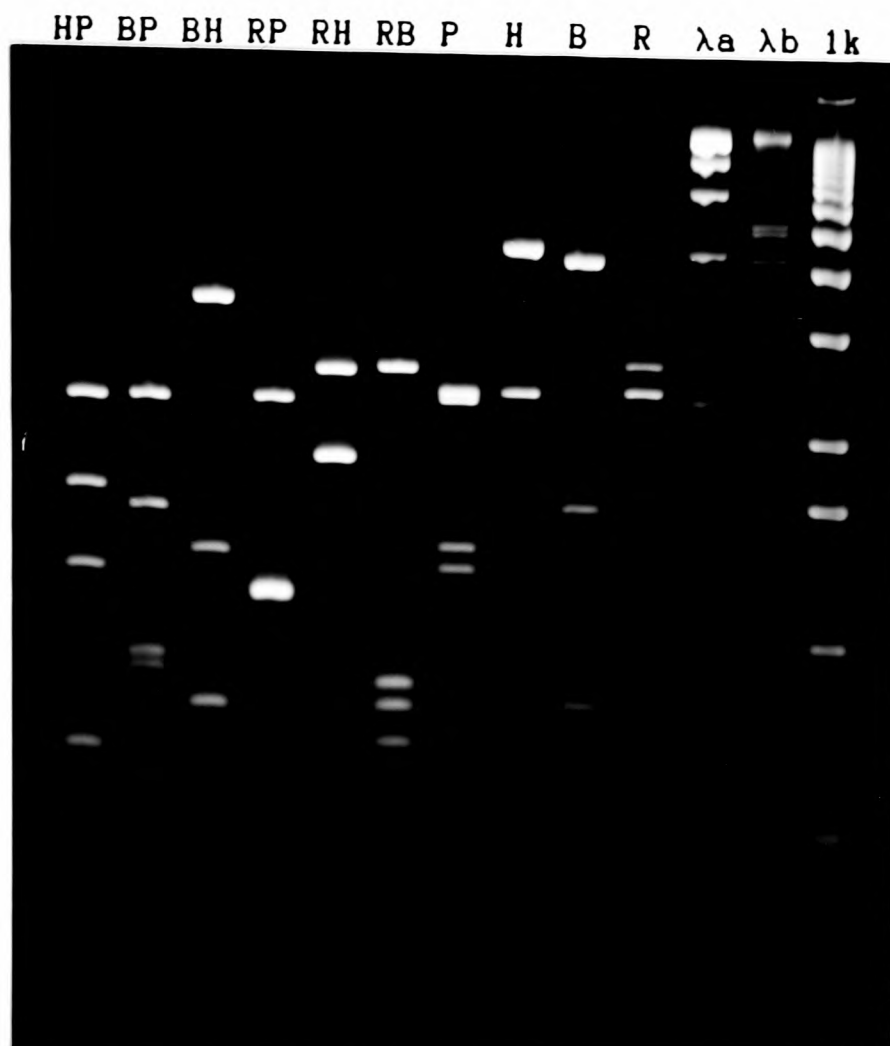
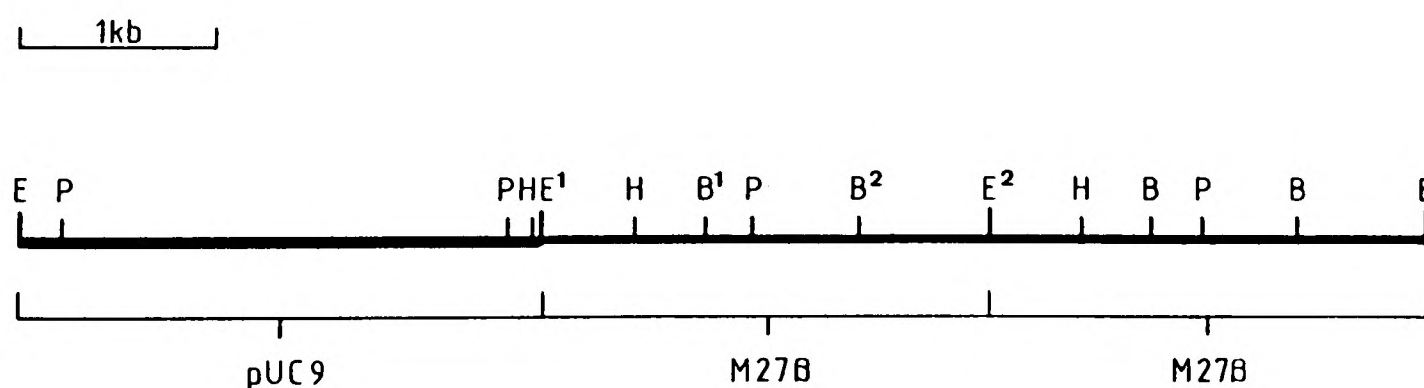


FIGURE 7.12
Restriction map of M27β.

B: *Bgl*III; E: *Eco*RI; H: *Hind*III; P: *Pvu*II; λa: λ/*Hind*III marker fragments; λb: λ/*Hind*III/*Eco*RI marker fragments, 1k: 1kb ladder (in addition to fragments differing by 1kb, this ladder contains fragments of 1636, 510, 396, 344, 298 and 200bp). The plasmid M27β was digested using the aforementioned enzymes, both singly and in paired combinations, and the sizes of the resultant fragments calculated, using the three sets of markers as standards. The restriction map shown below was constructed from these data.



B-*Bgl*III, E-*Eco*RI, H-*Hind*III, P-*Pvu*II. Restriction map of M27β, showing that the plasmid actually contains two copies of the insert. The sites E¹, B¹ etc. are referred to in the text.

that are the same size as the invariant fragments found in genomic digests, in addition to a variable fragment, whereas the *Bgl*III²-*Eco*RI² fragment hybridises to the fragments that are not of invariant size in different individuals. The combination of the data from the restriction map of M27β with that obtained by probing the cosmid digests with the two fragments as described above, allows the map shown in Figure 7.14 to be constructed. If the restriction site E² is in fact part of the vector, as was hypothesised earlier, then the sites B⁺, P⁺ and H⁺ should also be present in the vector. The published restriction map for the vector, pTM1 (Grosveld *et al.*, 1982), does not contain this arrangement of *Eco*RI, *Bgl*III, *Pvu*II and *Hind*III sites. To verify this, ³²P labelled purified vector was used to probe the same digests of the cosmid as used above. It did not hybridise to any of the fragments detected by M27β (Figure 7.13a). These results therefore suggest that the B²-E² fragment present in the cosmid contains a deletion of ≈3kb, presumably the repeat sequences responsible for the observed hypervariability are those that have been deleted. If this map is correct, it is possible to predict the sizes of the *Bgl*III, *Hind*III and *Pvu*II fragments that will be recognised by M27β in MOG-T, since they will be ≈3kb larger than those shown in the map i.e. 5.6, 6.8 and 6.3kb respectively. *Bgl*III and *Hind*III digests of MOG-T were probed with M27β and the fragments recognised were the same sizes as those predicted (Figure 7.15).

M27β has since been designated HXoxS1 and the 0.9kb *Bgl*III²-*Eco*RI² fragment subsequently cloned into the *Eco*RI/*Bam*HI sites of pUC9, HXoxS2.

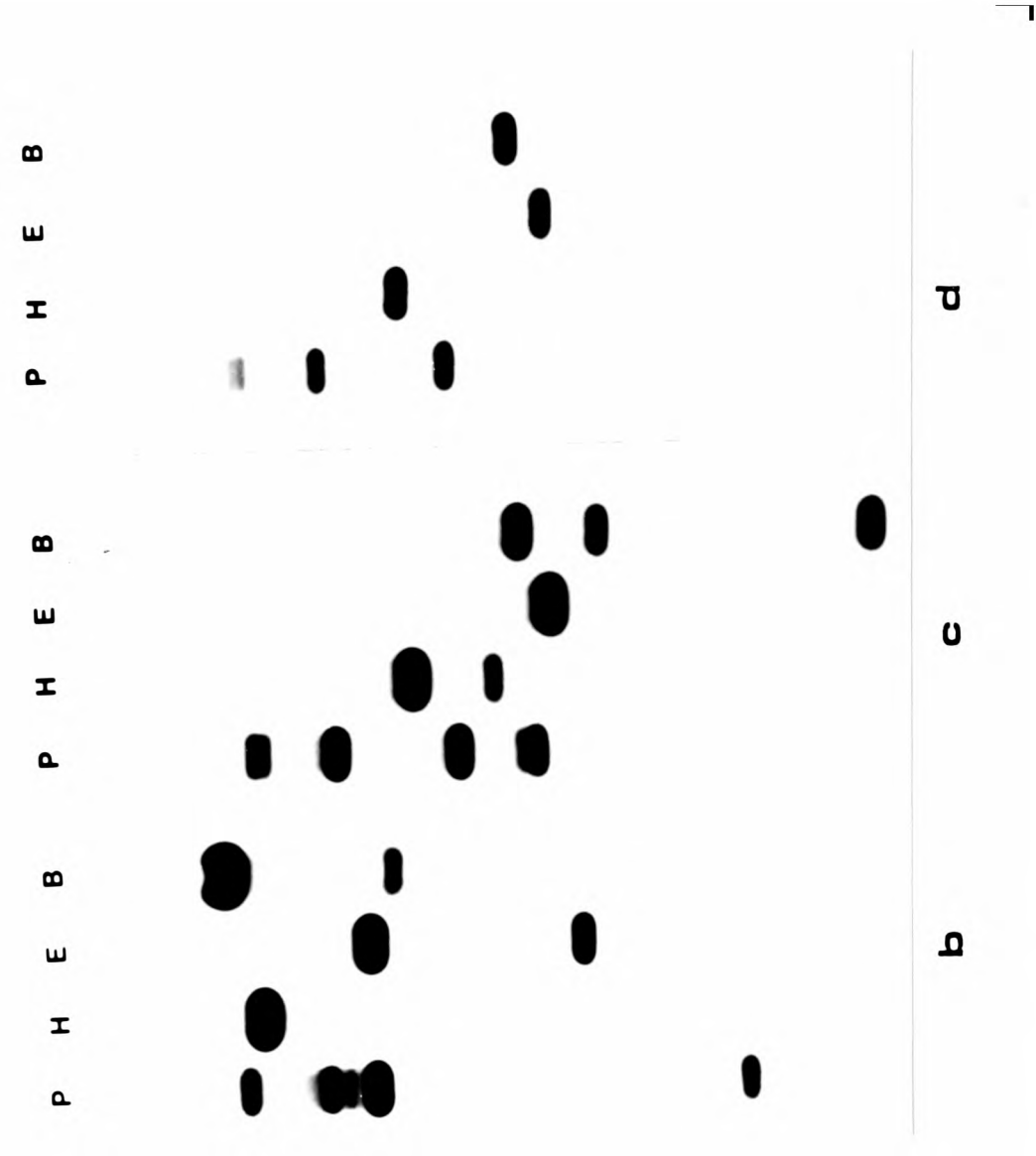
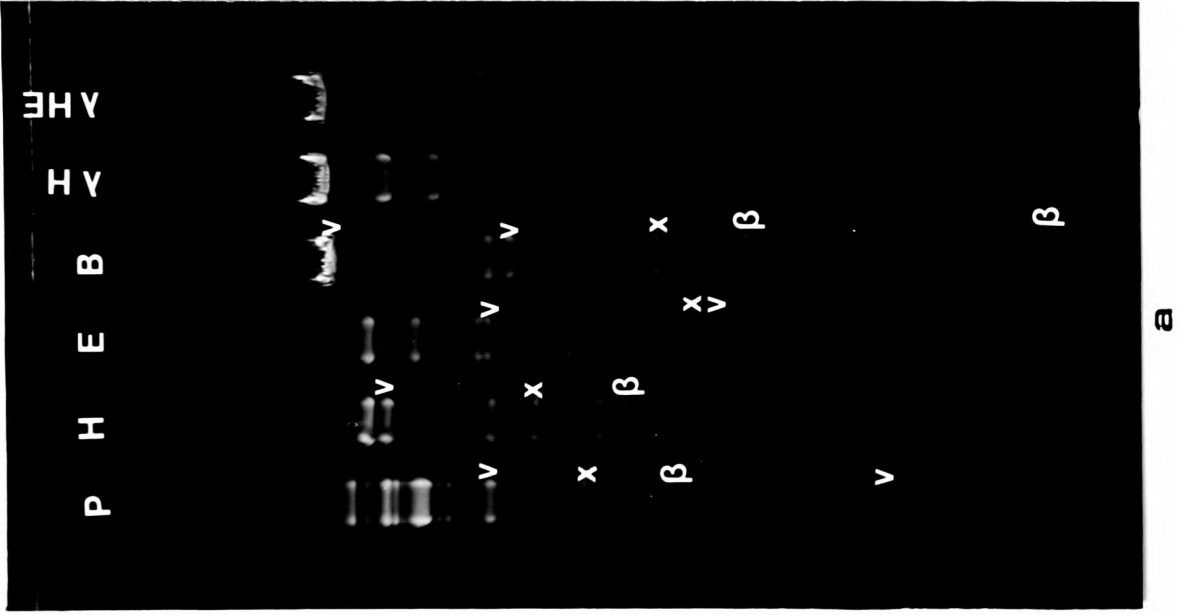


FIGURE 7.13
Characterisation of sequences in cosmid M27

FIGURE 7.13
Characterisation of sequences in cosmid M27
hybridising with M27 β .

a) Ethidium bromide stained gel of cosmid M27 digested with *Pvu*II (P), *Hind*III (H), *Eco*RI (E) and *Bgl*II (B). Identical filters were prepared from similar gels and probed with b) pTM1 (vector) c) M27 β and d) the 0.9kb *Bgl*II²-*Eco*RI² fragment of M27 β (see Figure 7.12). Fragments hybridising with the vector are marked (v), those hybridising just with M27 β (β) and those hybridising with the 0.9kb *Bgl*II²-*Eco*RI² fragment (x). Additional fragments are detected in the *Pvu*II digested cosmid, as a result of this digest being incomplete. M27 β hybridises with fragments that are the same size as the invariant fragments detected in genomic digests (B-1.9 & 0.8kb, H-2.7kb, P-2.4kb) in addition to a variable fragment (B-2.6kb, E-2.3kb, H-3.8kb, P-3.3kb). The 0.9kb *Bgl*II²-*Eco*RI² fragment detects the variable fragments. The vector does not hybridise with any of the fragments detected by M27 β . Together, these results suggest that the hypervariable region can be detected by the 0.9kb *Bgl*II²-*Eco*RI² fragment and that this fragment is not juxtaposed to vector sequences.

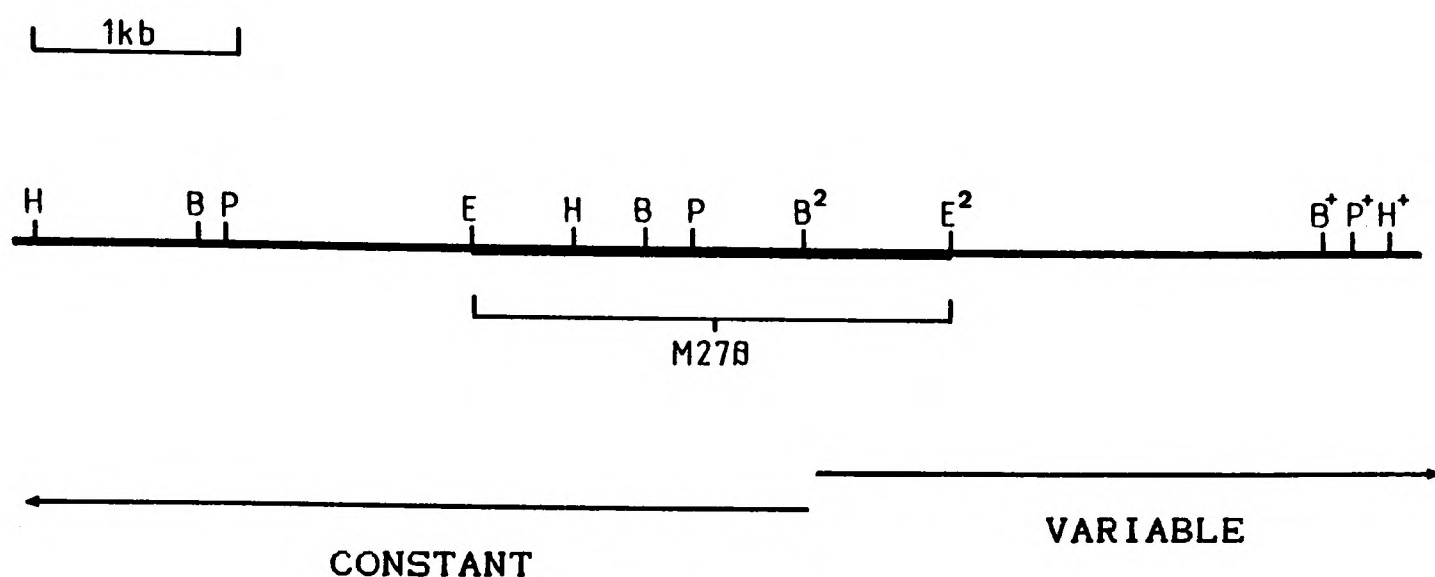


FIGURE 7.14
Restriction map of cosmid M27 in the region of M27β.

B: *Bgl*III; E: *Eco*RI; H: *Hind*III; P: *Pvu*II. The combination of the data from the restriction map of M27β (Figure 7.12) with that obtained by probing the cosmid digests shown in Figure 7.13, allows the restriction map of cosmid M27 in the region of M27β to be constructed. Fragments to the right of B² are variable in length in different individuals, those to the left of B² are of constant length. Results suggest that the B²-E² fragment contains a deletion of ≈3kb, this deletion is presumed to encompass the hypervariable repeat. Sites marked B⁺, P⁺ and H⁺ are those referred to in the text.

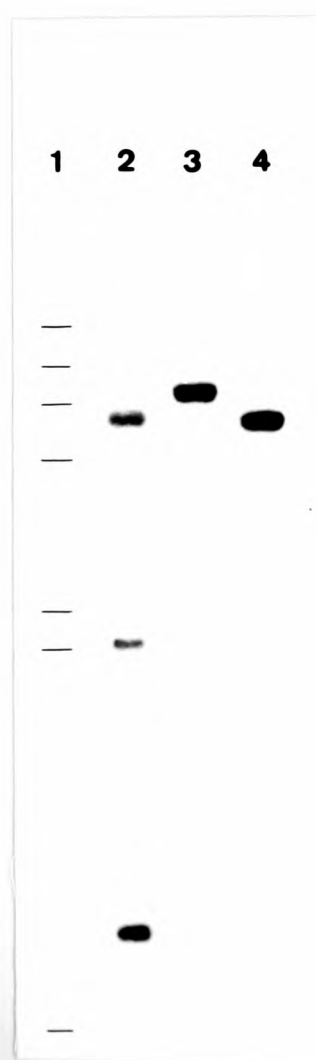


FIGURE 7.15
Restriction fragments in MOG T hybridising with M27β.

- 1: λ/*Hind*III marker fragments
- 2: MOG T/*Bgl*III
- 3: MOG T/*Hind*III
- 4: MOG T/*Eco*RI

Genomic DNA from MOG T digested with different enzymes and probed with M27β. The variable fragments detected are all ≈3kb bigger than would be predicted from the restriction map shown in Figure 7.14. These results confirm that ≈3kb has been deleted from M27β, during the propagation of cosmid M27.

7.11 IDENTIFICATION OF A 330bp RESTRICTION FRAGMENT OF HXoxS2 FROM WHICH THE HYPERVARIABLE REPEAT (HVR) HAS BEEN DELETED

Although results suggest that the HVR associated with *DXS255* had been deleted during the original generation of the cosmid bank, it is still possible that HXoxS2 contains one copy of the repeat. Direct sequence analysis of this fragment might reveal a motif characteristic of one of the HVRs already isolated (examples of these are given in Nakamura *et al.*, 1987a). However, since 900bp is not a trivial amount of DNA to sequence, it was decided to see if a smaller region containing the site of deletion of the HVR could be delineated. A restriction map of the clone HXoxS2 was constructed by digestion of the plasmid and the *EcoRI/SalI* insert with enzymes that have 4bp recognition sequences. The enzymes used were *MspI*, *Sau3A* and *HaeIII*. Both *Sau3A* and *HaeIII* sites were present in the insert. The map was constructed by running samples digested with these enzymes on a 1.4% agarose gel (Figure 7.16a). The information gained from the *Sau3A* partial digest of the insert and a knowledge of the positions of the *Sau3A* and *HaeIII* sites in the plasmid pUC9 (obtained from the *New England Biolabs* catalogue, 1986/7) allows the position of these sites and their orientation with respect to the *EcoRI* and *SalI* sites to be determined (Figure 7.16b). The fragment can thus be divided into four regions *EcoRI-Sau3A* 270bp, *Sau3A-Sau3A* 130bp, *Sau3A-HaeIII* 320bp and *HaeIII-Sau3A/BglII* 180bp. All of these fragments are more amenable to sequence analysis than the original clone of 900bp. To identify which fragment contained the HVR deletion, *EcoRI/HaeIII* and *EcoRI/Sau3A* double

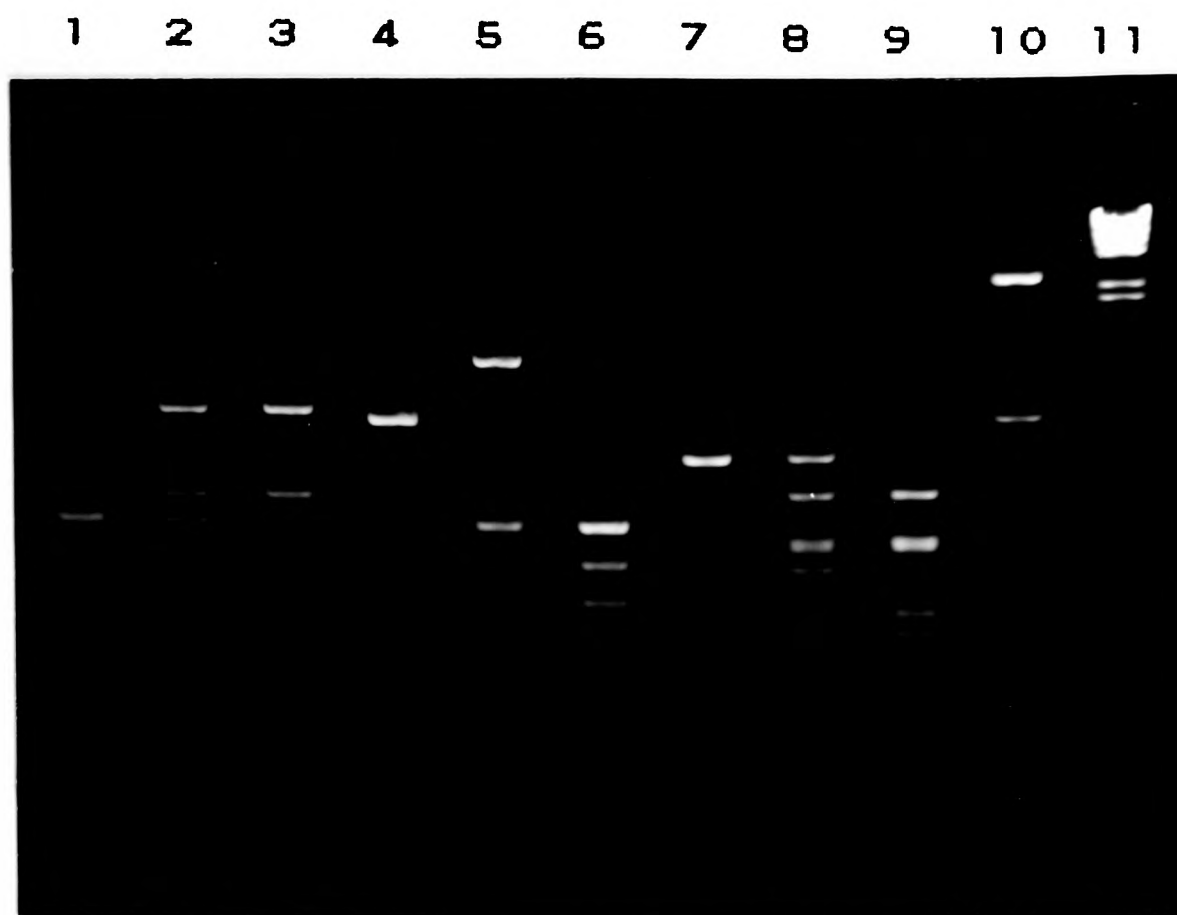


FIGURE 7.16a
Restriction map of HXoxS2

Track	Contents	Track	Contents	Track	Contents
1	I/ <i>Sau</i> 3A	5	P/ <i>Msp</i> I	9	V/ <i>Hae</i> III
2	P/ <i>Sau</i> 3A	6	V/ <i>Msp</i> I	10	P/ <i>Eco</i> RI
3	V/ <i>Sau</i> 3A	7	I/ <i>Hae</i> III	11	λ / <i>Hind</i> III
4	I// <i>Msp</i> I	8	P// <i>Hae</i> III		

I: HXoxS2 *Eco*RI/*Sa*II insert; P: HXoxS2 plasmid; V: Vector (pUC9)

Ethidium bromide stained gel of the digests used in the construction of the restriction map of HXoxS2. Molecular weights of unknown fragments were calculated using the sizes of the *Msp*I, *Hae*III & *Sau*3A fragments of pUC9 as standards. The orientation of the sites in the insert could be deduced from the altered mobility of some of the vector fragments in the digests of the HXoxS2 plasmid. The partial products of the *Sau*3A digest of the insert helped in the deduction of the positions of these sites.

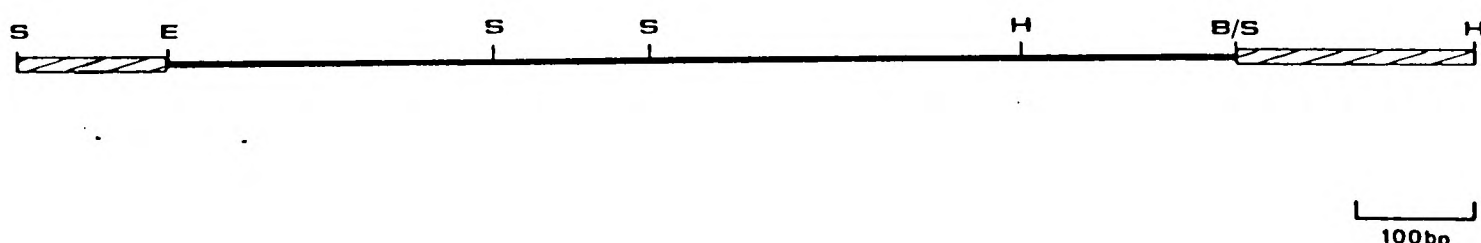


FIGURE 7.16b
B-*Bgl*III E-*Eco*RI H-*Hae*III S-*Sau*3A
Restriction map of HXoxS2. The hatched area represents sequences present in the vector.

digests of three different male DNAs were performed. These were run on a 1.4% agarose gel, blotted and the filter probed with HXoxS2. The sizes of the bands detected can be predicted from the map given in Figure 7.16b (except for the fragment extending to the right of the *Hae*III site since the position of the next *Hae*III/*Eco*RI site is not known), any band not accounted for presumably contains the HVR. The result is shown in Figure 7.17. The constant band detected in the *Eco*RI/*Hae*III digest is $\approx$ 1800bp and the 720bp fragment predicted is not present, suggesting that the HVR lies between the *Eco*RI and *Hae*III site. A constant band of 270bp is easily visible in the *Eco*RI/*Sau*3A digests, the 130bp fragment is not easily discernable and the 500bp fragment is clearly absent. The combined information from the two sets of digests suggests that the *Sau*3A/*Hae*III fragment of HXoxS2 (320bp) contains the site from which the HVR has been deleted. There are plans to sequence this fragment in the near future. An attempt to isolate, from a plasmid library, a fragment containing multiple copies of the HVR is described in Chapter 8.

7.12 CONSERVATION OF SEQUENCES HOMOLOGOUS TO DXS255 IN OTHER SPECIES

To see whether this locus is conserved in other species, 1 μ g samples of DNA from δ and ϕ rabbit, lemur, tamarin, saki, macaque, orang-utan, gorilla, chimpanzee and human controls were digested with *Eco*RI, run on a mini-gel, blotted onto HYBOND-N and probed with M27 β . Hybridisation was carried out at 60°C and washing was to a stringency of 2xSSC/0.1% SDS at 60°C.

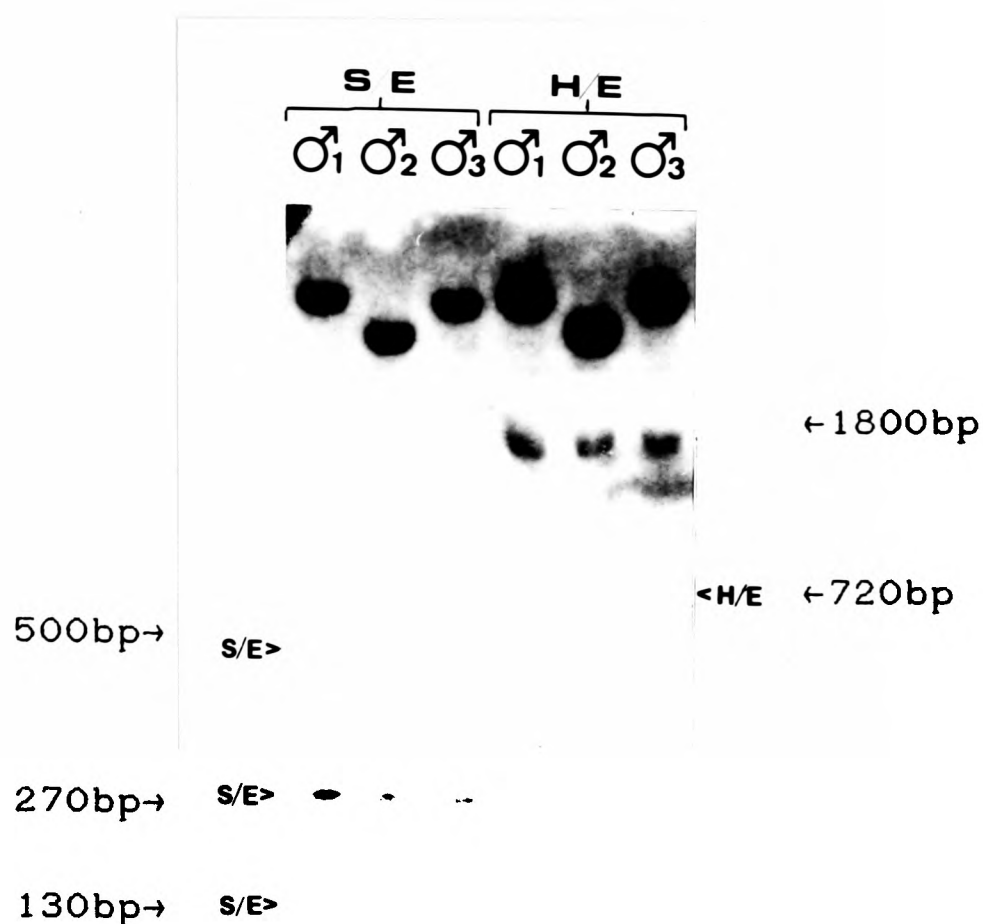


FIGURE 7.17
Determination of the *Sau3A*/*HaeIII* sites flanking
the HVR associated with *DXS255*

DNA from three unrelated males digested with *HaeIII*/EcoRI (H/E) or *Sau3A*/EcoRI (S/E) and probed with HXoxS2. The sizes of the fragments detected can be predicted from the restriction map of the clone (Figure 7.16b) and are marked (S/E) & (H/E). These positions were determined by comparing the autoradiograph with a photograph of the gel containing size markers. Fragments not present can be assumed to contain the HVR. The 500bp *Sau3A*/EcoRI fragment and the 720bp *HaeIII*/EcoRI are clearly absent. The 130bp *Sau3A*/EcoRI fragment is too faint to be seen and the 1800bp *HaeIII*/EcoRI constant fragment could not be predicted from the map, since the position of the next *HaeIII* or *EcoRI* site to the right of the *HaeIII* site shown in Figure 7.16b is not known. These results suggest that the HVR has been deleted from the 320bp *Sau3A*/*HaeIII* fragment of HXoxS2.

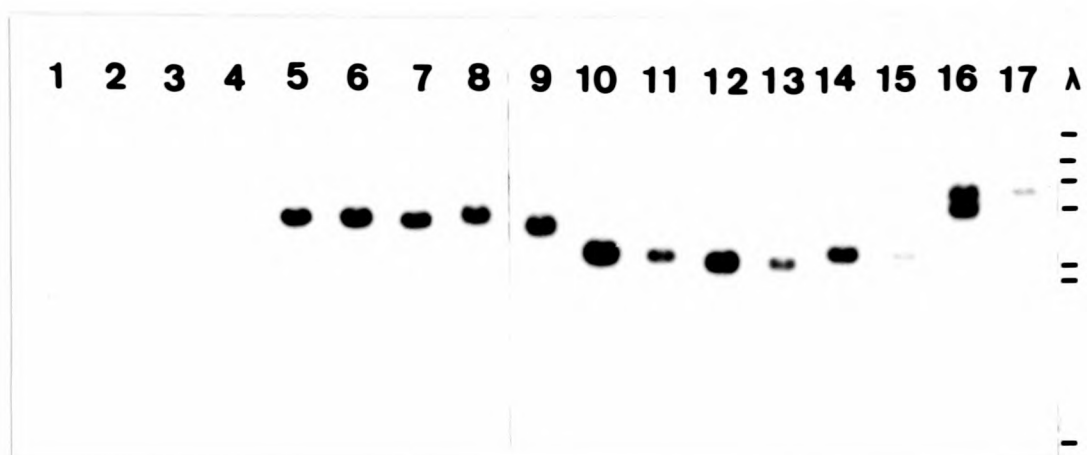


FIGURE 7.18
Phylogenetic study of sequences homologous to M27 β .

1: ♀ Rabbit; 2: ♂ Rabbit; 3: ♀ Lemur; 4: ♂ Lemur; 5: ♀ Tamarin; 6: ♂ Tamarin; 7: ♀ Saki; 8: ♂ Saki; 9: ♂ Macaque; 10: ♀ Orang-utan; 11: ♂ Orang-utan; 12: ♀ Gorilla; 13: ♂ Gorilla; 14: ♀ Chimpanzee; 15: ♂ Chimpanzee; 16: ♀ Human; 17: ♂ Human; λ : λ /*HindIII* markers.

All genomic DNA samples were digested with *EcoRI* and probed with M27 β . Homologous sequences are detected in primate species as far removed as the saki, and tamarin, thought to have diverged from the line leading to modern man ≈ 40 Myr ago. A faint signal is present in the ♀ rabbit track, this is also visible in the ♂ rabbit track on prolonged exposure. Homologous sequences are not detected in the lemur. The locus detected by M27 β appears to be hypervariable only in humans. The size of the fragment detected in the chimpanzee is similar to that present in the cloned, 'null', allele (≈ 2.3 kb).

The result is shown in Figure 7.18. Although it has been possible to analyse only three X chromosomes from each species, the most striking observation is that the locus does not appear to be hypervariable in any species other than man. The $\delta:\phi$ dosage observed in some cases suggests an X localisation; however, the amount of DNA in each track is not exactly equivalent. Sequences detected by the probe are present in primate species as far removed as the tamarin and saki, thought to have diverged from the line leading to modern man ≈ 40 Myr ago. Very weak signals are present in the rabbit tracks, but there appears to be no significant cross-hybridisation with lemur DNA, even after a prolonged exposure. There is little, or no, cross-hybridisation with sequences present in the mouse (results not shown). A perhaps significant observation is that the fragment detected in the chimpanzee (≈ 2.3 kb) is the same size as the putative 'null' allele present in the cloned sequence. Although this may be coincidental, it is tempting to speculate that the acquisition of an HVR at this locus in humans has occurred during recent evolutionary time i.e. since the divergence of the chimpanzee and man. Whether this HVR is present at other chromosomal loci in man and other species, will not be known until the sequence defining this repeat has been isolated.

7.13 GENERAL DISCUSSION OF RESULTS

This chapter has described the characterisation of a hypervariable locus (*DXS255*) on the short arm of the X chromosome. The locus has been identified using a single-copy

probe isolated from the MOG cosmid bank and sub-regionally localised to Xp11.3→Xcen using the somatic cell hybrid mapping panel described in Chapter 3. The high degree of polymorphism at *DXS255* appears to be due to an insertion/deletion phenomenon similar to that causing the hypervariability of minisatellite sequences. Heterozygosity in females has been estimated to be >90%, with alleles ranging in size from 3.5–7kb. It has not been possible to determine the basic repeat size, although the probing of well resolved blots may give some indication of this. The sequence responsible for the hypervariability appears to have been deleted during the original creation of the cosmid bank, hence the probe M27β (HXoxS1) may represent a 'null' allele. A 330bp *Sau3A/HaeIII* restriction fragment has been identified as containing the site from which the HVR has been deleted. If one copy of the repeat has remained on this fragment, it may be possible to identify it by looking for a motif characteristic of one of the HVR's already isolated. There are plans to do this in the near future. The deletion of this sequence during the cloning procedure is most probably due to the inherent instability of HVR's in many bacterial host systems. Nevertheless, family studies have shown that the polymorphism detected is based on sufficiently stable allelic variation for it to be of use in linkage analysis.

DNA has been analysed from members of five large families with X-linked retinitis pigmentosa (XLRP). One locus for this disorder has previously been localised to Xp11.3. Results were encouraging, with only one definite cross-over being identified in 28 informative meioses. This corresponds to a genetic distance of 0.09–18.35 centimorgans at 95% confidence limits.

Some members of these families are currently undergoing detailed medical examinations and so the number of informative meioses is likely to be increased when the carrier status of females at risk is known. The analysis of these families with other, well characterised, probes from this region may help to localise *DXS255* more specifically with respect to XLRP. There are a number of other disorders that have been mapped to this region of the X chromosome that will be amenable to study using this probe.

As well the possible uses of this probe in linkage analysis, it may prove to be useful in the study of the origins of the X chromosomes in individuals with disorders such as Klinefelter's (47, XXY) and Turner's (45, XO) syndromes. A preliminary investigation of a ♂ 47, XXY individual has suggested that, in the one case examined, both X's are the same and are maternally derived (results not shown). In the 48, XXXX cell line GM1416 used in these studies, it has been possible to identify three different X chromosomes, the dosage of the smallest fragment is approximately twice that of the other two suggesting that there are two copies of the X chromosome carrying this allele. This arrangement can be explained by nondisjunction at the first and second meiotic divisions of oogenesis giving rise to a $X^1X^2X^2$ gamete, subsequently fertilised by a ♂ X^3 gamete.

Many other uses for the probe have been suggested. Methylation has been proposed to have a role in X-inactivation (see Chapter 1); if M27β is able to detect a RFLP using an enzyme sensitive to the methylation of its recognition site (e.g. *HpaII* and *HhaI*) and there is a consistent difference in

methylation of these sites on inactive and active X chromosomes, it may be possible to distinguish between clonal and polyclonal populations of cells. Cells in a clonal population will all have the same X chromosome inactivated, and the simplest pattern obtained by probing DNA cut with a methylation sensitive enzyme with M27 β would be two bands, one representing the unmethylated allele on the active X and the other the methylated allele on the inactive X. In the case of a polyclonal population of cells, where X-inactivation, is random, if the probe detects a RFLP then the simplest pattern would consist of four bands, representing both the methylated and unmethylated copies of the two alleles. In practice, the pattern may be more complex, due to the differential methylation of some residues on active and inactive X chromosomes (Wolf & Migeon, 1982). A similar approach, using a probe for the *HPRT* locus, has met with some success in the analysis of the clonal origin of tumours. In this case, it was necessary to use a combination of two enzymes, one detected a RFLP and the other determined the methylation status of the gene on the active and inactive X chromosomes (Vogelstein *et al.*, 1985). Only 27% of females were estimated to be heterozygous for this RFLP, and so its use is limited. The high degree of polymorphism detected by M27 β suggests that if it can easily discriminate between active and inactive X chromosomes, then many types of female tumour may be amenable to this sort of analysis. Knowledge of the clonal origin of a tumour has important conceptual and practical implications. An example of which is that clonality can detect a preleukaemic phase during the remission of acute myelocytic leukaemia (reviewed in

Vogelstein *et al.*, 1985).

The observation of a lack of hypervariability in species other than man is intriguing. Although the data are not conclusive, they suggest that the HVR has been acquired by the *DXS255* locus during the past few million years. It is possible that the basic repeat unit may have features that allowed its transposition from another locus to Xp11.3 and that the observed hypervariability has been caused by subsequent rounds of saltatory replication, resulting in an expansion of this repeat. The molecular characterisation of the HVR itself may give some insight into its possible origins. An attempt to clone this sequence is described in Chapter 8.

CHAPTER 8

AN ATTEMPT TO CLONE THE SEQUENCES RESPONSIBLE FOR THE OBSERVED HYPERVARIABILITY OBSERVED AT *DXS255*

8.1 INTRODUCTION

A simple and rapid technique for the cloning of specific genomic DNA fragments in small, but representative, plasmid libraries has been described (Nicholls *et al.*, 1985). The technique is based on digestion of genomic DNA with multiple restriction enzymes, identification of the desired fragment by Southern blot analysis, and then size selection on agarose gels. It is especially applicable for the repetitive isolation of mutant or polymorphic alleles, the cloning of breakpoints from chromosome rearrangements, and for the cloning multi-allelic DNA sequences that are unstable or 'unclonable' in other vector/host systems. Since the hypervariable repeat associated with the multi-allelic variation of M27 β appeared to have been deleted from the cosmid M27, it suggested that this sequence is likely to be unstable in the *recA*⁺ hosts normally chosen to propagate phage and cosmid libraries, but may be amenable to this approach. They had shown that purification of genomic DNA fragments generated by a single enzyme with size selection on an agarose gel followed by electroelution results in the exclusion of >80% of genomic sequences. However, such fragments prepared from a single enrichment step when cloned into a plasmid vector results in only a partially representative library. Nevertheless, a similar approach has been successfully employed by Wong *et al.* (1986) to construct a

library in phage λ that contained an HVR tending to cosegregate with HPFH (hereditary persistence of foetal haemoglobin). Predictions from a simple model proposed by Nicholls et al. (1985) suggested that the use of several enzymes for preparation of DNA fragments would give the desired purification necessary to construct a genomic library representative of all fragments of a given size.

The model predicts the number of fragments of a given size that will be present in the haploid human genome following multi-enzymic digestion. This fraction will be given by $a=b(c/d)$, where:

b = genome fraction remaining after digestion

c = haploid size of the genome

d = size of a particular restriction fragment

Further enrichment is obtained due to the heterogenous ends in the cloning generated by the use of different enzymes.

8.2 MATERIALS AND METHODS

DNA was prepared from cell-pellets and blood as described in 2.4.

Genomic digests were performed overnight using an excess of enzyme under the conditions recommended by the manufacturers.

For the large scale digestion of DNA from GM1416 (4X cell-line), 100 μ g of DNA was digested with 100 units of both *Bgl*III and *Eco*RI overnight. The enzymes were inactivated by extraction once with phenol/chloroform once with chloroform. DNA was recovered by ethanol precipitation for 30 minutes at -70°C, resuspended in 1x React 3 (*BRL*) buffer and digested overnight with 100 units each of *Bam*HI, *Pvu*II, *Pst*I, *Sal*I & *Hind*III and a further 50 units each of *Bgl*III and *Eco*RI (to ensure that digestion with these two enzymes was complete). Again the digest was purified and recovered by ethanol

precipitation, but was finally dissolved in 1xTBE electrophoresis buffer.

The elution of DNA onto DEAE paper is described in 2.7.3.

The preparation of vector DNA and subcloning of restriction fragments is described in 2.11.

The screening of bacterial colonies is described in 2.6.

The blotting of agarose gels onto HYBOND-N is described in 2.7.2. Conditions for the hybridisation and washing of filters is described in 2.10.

All probes were labelled by nick translation of insert DNA (2.8.1).

Cells were made competent using the method of Hanahan (1983):-

1. Pick six 2-3mm diameter colonies off a freshly streaked SOB agar plate and disperse in 1ml SOB medium by vortexing.
 2. Inoculate the cells into an Erlenmeyer flask containing 60ml of SOB medium.
 3. Incubate at 37°C with moderate agitation until the culture reaches O.D.₅₅₀ 0.45-0.55.
 4. Chill on ice for 10 minutes then transfer 24ml aliquots to 50ml polypropylene centrifuge tubes.
- For all subsequent stages ensure that pipettes and solutions are pre-chilled to 4°C for maximum transformation efficiency.
5. Pellet by centrifugation at 2000 r.p.m. in a bench centrifuge at 4°C for 10 minutes. Drain the pelleted cells thoroughly and remove all traces of liquid.
 6. Resuspend in 8ml (1/3 of culture volume) TFB by sucking up and down in a pipette. Incubate on ice for 10-15 minutes.
 7. Pellet the cells and drain thoroughly as in step 5.
 8. Resuspend the cells in 1.9ml (1/12.5 of the original volume) TFB. Add 67µl of DnD solution (3.5% v/v) by squirting into the centre of the cell suspension and swirling the tube immediately for several seconds. Incubate on ice for 10 minutes.
 9. Add a second equal aliquot of DnD as in step 8. Incubate on ice for 15 minutes.
 10. Pipette 210µl aliquots into chilled polypropylene tubes.
 11. Add the DNA solution in a volume of <20µl, swirling to mix. Incubate the tubes on ice for 30-60 minutes.
 12. Heat shock the cells by placing the tubes in a 42°C water

bath for 2 minutes. Return the tubes to ice for 2 minutes.

13. Add 800 μ l of SOC medium to each tube. Incubate at 37°C with moderate agitation for 1 hour.

14. Plate out aliquots on suitable selective plates and incubate at 37°C overnight.

TFB is: 100mM KCl, 45mM MnCl $\cdot$ 4H $_2$ O, 10mM CaCl $_2$ $\cdot$ 2H $_2$ O, 3mM H $\text{A} \text{CoCl}_3$ (hexamine cobalt (III) + trichloride), 10mM K-MES, (Final pH 6.20 $\pm$ 0.10)

Preparation: Equilibrate a 0.5M solution of MES (2[N-morpholinoethane sulphonic acid]) to pH 6.3 using concentrated (5M) KOH, then sterilise by filtration through a 0.22 μ membrane, and store in aliquots at -20°C. Make a solution of 10mM K-MES, using the 0.5M MES stock and dH $_2$ O. Add the salts as solids, then filter the solution through a 0.22 μ pre-rinsed membrane. Aliquot into sterile tubes and store at 4°C. (Stable for >1 year).

DnD is: 1M DTT, 90% (v/v) DMSO, 10mM Potassium acetate.

SOB is: 2% Bacto tryptone, 0.5% Bacto yeast extract, 10mM NaCl, 2.5mM KCl, 10mM MgCl $_2$, 10mM MgSO $_4$.

SOC medium is identical to SOB medium but contains 20mM glucose in addition.

Preparation of SOB and SOC: Combine tryptone, yeast extract, NaCl and KCl in dH $_2$ O and autoclave. Use within 2-3 weeks. Make a 2M stock of Mg $^{2+}$, comprised of 1M MgCl $_2$ plus 1M MgSO $_4$. Sterilise by filtration through a 0.22 μ membrane. Prepare a 1M stock of glucose similarly, and store at -20°C. Just prior to use, combine the medium with Mg $^{2+}$ (and glucose for SOC). The final pH should be 6.8-7.0.

8.3 PREPARATION OF GENOMIC INSERTS

In order to apply this method to the cloning of the HVR associated with HXoxS1, it was first necessary to determine which restriction enzymes would cleave outside repeat, but that would also result in fragments containing the repeat

migrating above the main front of DNA. This is approximately 4.5kb after digestion with five enzymes. 3 μ g samples of DNA from the female 6G2 were digested with the enzymes *EcoRI*, *HindIII*, *PstI*, *PvuII* and *BglII*, using the enzymes both singly and in every paired combination. In addition, one 3 μ g sample was digested with a combination of all five enzymes. Data from the restriction map of the cosmid M27 in the region around M27 β suggested that the sites flanking the hypervariable region were *BglII* and *EcoRI* (Figure 7.14). The probing of a Southern blot of these multiple digests with M27 β would confirm this, since the digest giving the smallest alleles would be that containing the enzyme(s) cutting at sites flanking the HVR. This pattern would also be reflected in the digest containing all of the enzymes. The result is shown in Figure 8.1. As predicted the digest producing the smallest alleles is a *BglII/EcoRI* double digest, showing that of the enzymes chosen, these two must flank the HVR. The sizes of these restriction fragments are $\approx$ 3.4 and $\approx$ 2.3kb and so migrate below the main front of DNA in the multi-enzyme digest. Previous experiments had shown that the 4X cell line GM1416 contained more extensive HVR regions, since two of the fragments detected by M27 β in an *EcoRI* digest of DNA prepared from this cell line were >1.5kb larger than those detected in the same digest of 6G2 (GM1416 is 4X_a in Figure 7.2). Probing of a *BglII/EcoRI* double digest of GM1416 with M27 β showed that two of the alleles were present on restriction fragments of $d_1 = 5.3$ kb and $d_2 = 4.4$ kb (see Figure 8.2). These were of a more suitable size to attempt fragment enrichment, and the addition of *BamHI* and *SaII* (which had also been shown to cleave well outside the region of

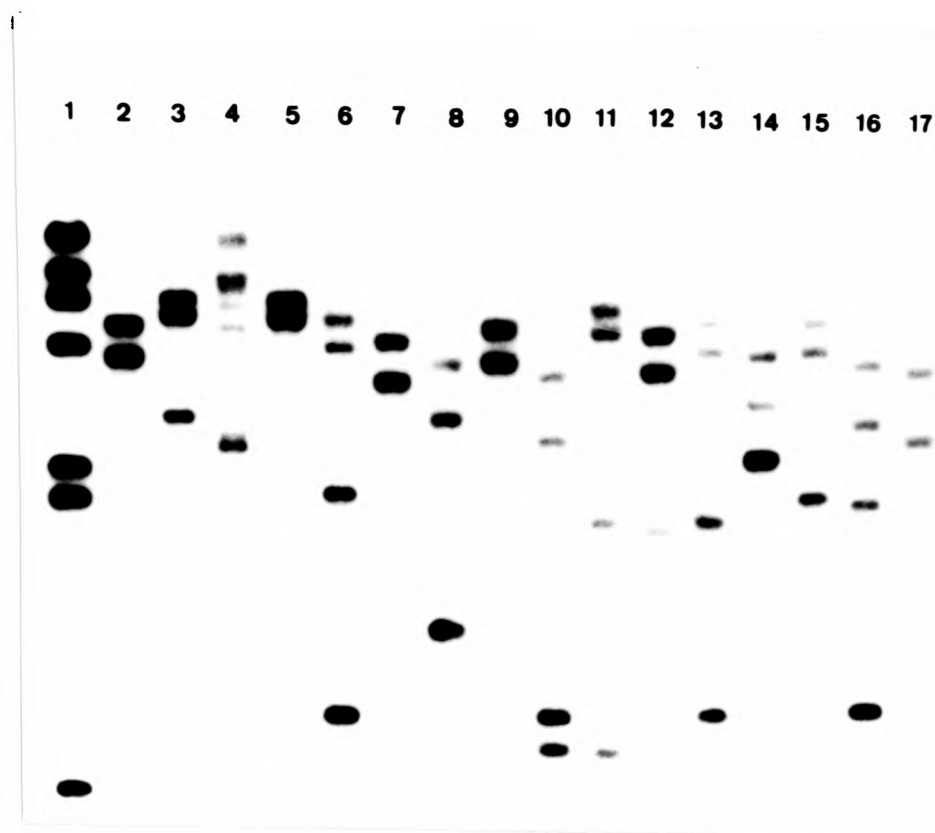


FIGURE 8.1
Identification of the restriction sites
flanking the HVR associated with *DXS255*

1: λ /*Hind*III marker fragments.

The remaining tracks contain genomic DNA from the q 6G2 digested with:-

2: *Eco*RI; 3: *Hind*III; 4: *Pvu*II (incomplete); 5: *Pst*I; 6: *Bgl*II;
 7: *Eco*RI/*Hind*III; 8: *Eco*RI/*Pvu*II; 9: *Eco*RI/*Pst*I;
 10: *Eco*RI/*Bgl*II; 11: *Hind*III/*Pvu*II; 12: *Hind*III/*Pst*I;
 13: *Hind*III/*Bgl*II; 14: *Pvu*II/*Pst*I; 15: *Pvu*II/*Bgl*II;
 16: *Pst*I/*Bgl*II; 17: A combination of all of these enzymes.

The filter was probed with M27 β .

In general variable fragments are those >2.3kb. Fragments smaller than this are invariant length in different individuals (a 2.7kb constant fragment is present in the *Hind*III digest). The smallest variable fragments are present in track 10 (3.4kb and 2.3kb), these fragments are also present in track 17, suggesting that the sites flanking the HVR are *Eco*RI and *Bgl*II.

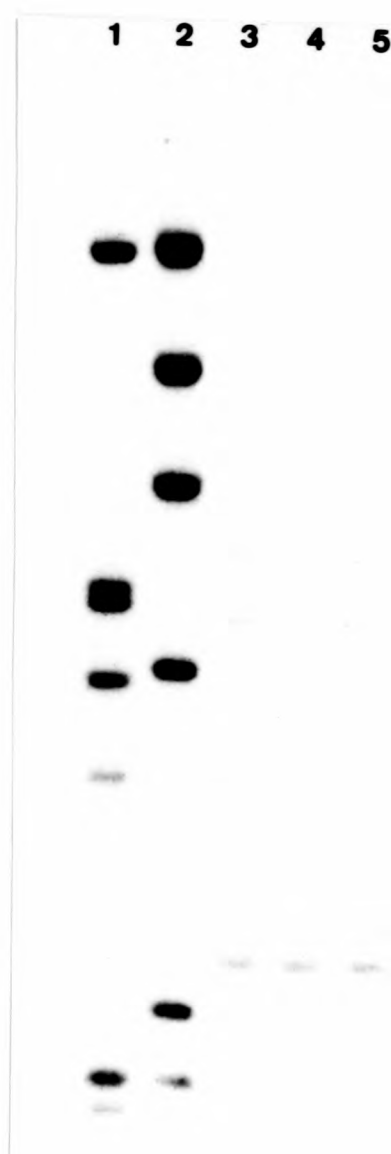


FIGURE 8.2
Determination of the size of the *EcoRI/BglII* fragments
in GM1416 containing the HVR associated with *DXS255*.

- 1: λ /*HindIII/EcoRI*; 2: λ /*HindIII*.
 3: GM1416/*BglII/EcoRI*.
 4: GM1416/*BglII/EcoRI/HindIII/PstI/PvuII/BamHI*.
 5: GM1416/*BglII/EcoRI/HindIII/PstI/PvuII/BamHI/SalI*.

The filter was probed with M27 β to determine the size of the fragments detected in the 4X cell-line GM1416 after a multi-enzyme digest. The results confirm that *BglII* and *EcoRI* sites flank the HVR and that the fragments detected in this cell-line (5.3kb and 4.4kb) migrate above the main front of DNA.

hypervariability) to a multi-enzyme digest would ensure that these fragments migrated above the main front of DNA. A *Bgl*III/*Eco*RI restriction fragment can be cloned into pUC9 that has cut with *Eco*RI and *Bam*HI since DNA cleaved with *Bgl*III (A⁺GATCT) has compatible ends with the vector cleaved with *Bam*HI (G⁺GATCC). The number of fragments (a) of these sizes present in the multi-enzyme digest can be calculated using the formula given above. In the case of digestion with seven enzymes $b=1.28 \times 10^{-5}$ (Nicholls *et al.*, 1985). For the two fragments:

$$a_1 = 1.28 \times 10^{-5} (3 \times 10^9 / 5.3 \times 10^3)$$

$$a_2 = 1.28 \times 10^{-5} (3 \times 10^9 / 4.4 \times 10^3). \text{ Thus, } a_1 = 7 \text{ and } a_2 = 9.$$

An analytical gel of 0.8% NA agarose (*Pharmacia*) was prepared. 4 μ g of the multi-enzyme digest and 4 μ g of a *Bgl*III/*Eco*RI digest together with λ /*Eco*RI/*Hind*III and λ /*Hind*III marker fragments were run at 40V (45mA) for 24 hours. The position of the fragments to be purified from a preparative gel was determined by Southern blotting of the gel and hybridisation with M27 β (Figure 8.2).

The preparative gel was also of 0.8% NA agarose with $\approx 90\mu$ g of the multi-enzyme digest loaded into a 30x1 mm well, flanked by the same size markers. The gel was run under the same conditions as above. Ethidium bromide staining of the size markers enabled size selection of the 5.3 and 4.4kb fractions. Incisions were made in the gel at 2mm intervals in a region between ≈ 4 kb and ≈ 6 kb and the fractions were eluted onto pieces of NA45 DEAE paper. The gel was then stained to ensure that all of this fraction had been eluted. The DNA recovered from the paper was dissolved in 25 μ l of 1xTE. 2.5 μ l aliquots of each

fraction were examined by gel electrophoresis to estimate the amount of DNA recovered, by comparison to size markers of known concentration (Figure 8.3). The gel included 1 μ g of 4X DNA digested with the seven enzymes and also $\approx$ 2 μ g of DNA digested with just *Bgl*III and *Eco*RI. The total amount of DNA recovered in each fraction was estimated to be 50–100ng. The gel was blotted onto HYBOND-N and the filter probed with HXoxS2. The result is shown in Figure 8.4. Fraction 5 apparently contains the largest amount of the 5.3kb allele and fraction 9 the 4.4kb allele. The degree of enrichment can be estimated by comparing the signal in these tracks with those present in the tracks containing the genomic digests. The signals present in tracks 5 and 9 are approximately 4x the intensity of the corresponding signals in the multi-enzyme digest. The amount of DNA present in these tracks is $\approx$ 5ng (equal in intensity to the 564bp fragment of the λ /*Hind*III marker fragment which is $\approx$ 5ng), thus the enrichment factor is $(1 \times 10^{-8} / 5 \times 10^{-9}) \times 4 = 800$. Therefore these fractions each contain $\approx$ 0.13% of the genome, or 100 times that expected for fragments 5.3kb or 4.4kb alone. The total number of fragments present in each fraction, calculated using the formula given above, is $\approx$ 80. The total number of recombinants that would be needed to be screened is less than this due to heterogenous ends obtained by the use of 5 additional enzymes over the two generating the desired fragment (*Bam*HI also creates ends compatible for cloning in the particular system used here). A library of >100 recombinants should therefore stand a good chance of containing the desired clone(s).

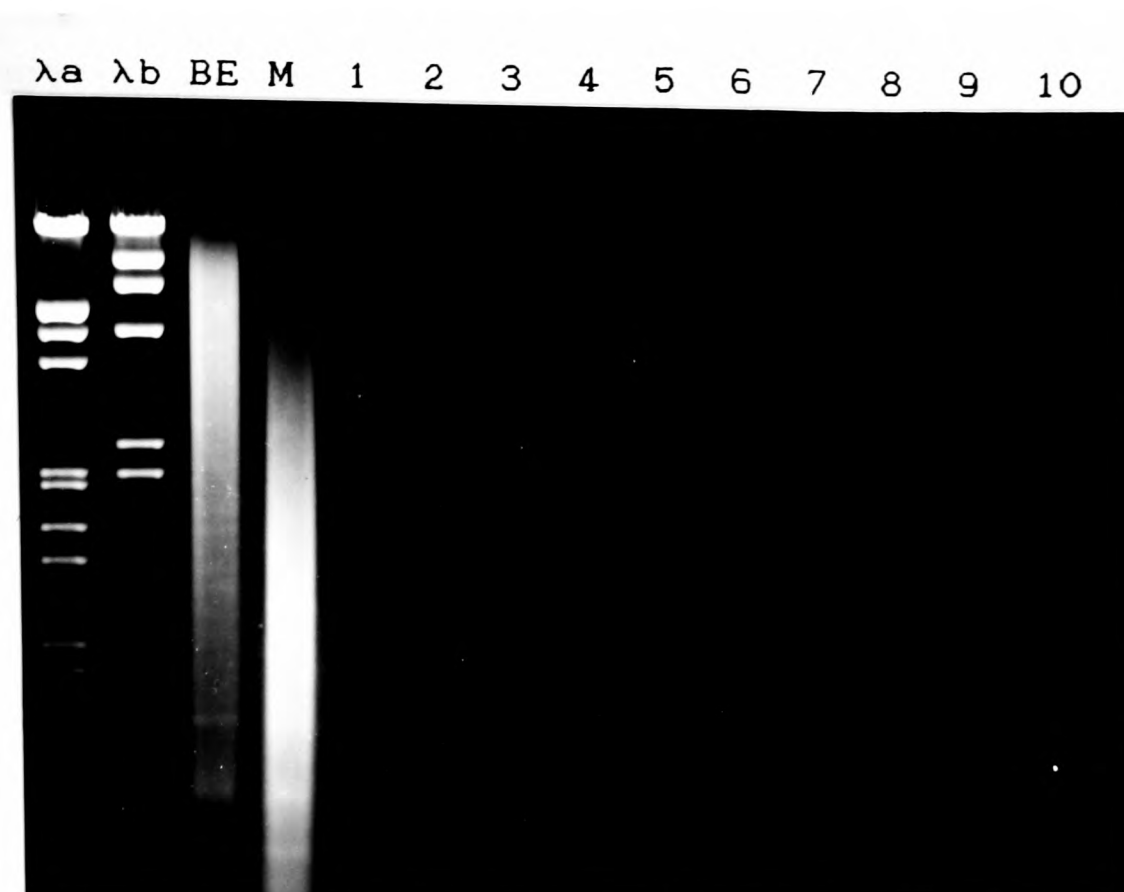


FIGURE 8.3
Recovery of 4-6kb fractions
from a multi-enzyme digest of GM1416

λa: λ/*Hind*III/*Eco*RI; λb: λ/*Hind*III marker fragments.
BE: GM1416/*Bgl*III/*Eco*RI; M: GM1416 multi-enzyme digest.
Tracks 1-10 contain samples of the 4-6kb fractions eluted from the preparative gel of the GM1416 multi-enzyme digest. The λ/*Hind*III track contains ≈500ng of DNA. The maximum amount of DNA present in any of the fractions can be estimated to be ≈5-10ng.

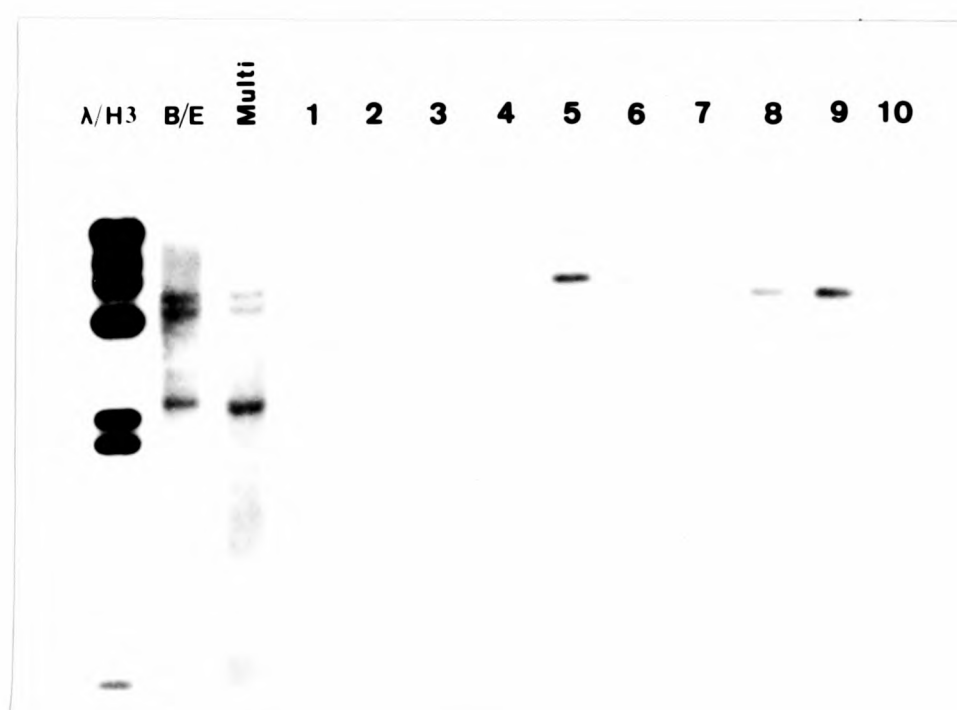


FIGURE 8.4
Identification of fractions containing sequences
homologous to HXoxS2

Filter prepared from the gel shown in Figure 8.3 probed with HXoxS2. Fraction 5 contains the greatest amount of the 5.3kb allele and fraction 9 the 4.4kb allele.

8.4 CONSTRUCTION AND SCREENING OF THE PLASMID LIBRARY

The plasmid vector chosen was pUC9 for both a small size to aid transformation efficiency and the availability of suitable cloning sites in the polylinker. *Bam*HI/*Eco*RI cut vector was prepared as described (2.11). Since a representative library was required it was decided not to use competent cells prepared by the standard CaCl_2 method, but to use the high frequency technique devised by Hanahan (1983). This technique is reported to give transformation efficiencies of $>10^8/\mu\text{g}$ pBR322.

To test that the double digest of the vector had been successful, test ligations were performed. 10ng of vector was ligated overnight to 15ng of genomic DNA that had been digested with a) *Eco*RI b) *Bgl*III c) *Eco*RI and *Bgl*III. A ligation was also set up with just the vector present. *E.coli* JM83 cells were made competent as described (8.2). The different ligations were used to transform 200 μl aliquots of cells. As a control, and to ascertain the overall transformation efficiency of the competent cells, 10ng of supercoiled pUC9 was used to transform one aliquot.

Serial dilutions were made and duplicate 100 μl aliquots spread onto 2xTY/X-gal/AMP plates. Plates were incubated at 37°C overnight. The results are shown below. The percentage of recombinants was calculated from the number of white (recombinant) and blue (non-recombinant) colonies.

Transformation efficiency: 7×10^7 transformants/ μ g pUC9

Ligation Reaction	Total Colonies/ng vector	% Recombinants
Vector alone	25	--
a	120	77
b	120	73
c	20,000	95

The results suggest that >99% of the vector molecules have *EcoRI/BamHI* ends.

The first strain of *E.coli* used in the attempt to clone the HVR associated with M27 β was HB101, since this *recA*⁻ strain had previously been shown to stabilise this type of sequence (Nicholls *et al.*, 1985). 25ng of DNA from fraction 5 was ligated to 10ng of vector as described above and used to transform the competent cells. Supercoiled pUC9, and vector ligated in the absence of insert DNA were used as controls. Since HB101 is not a strain on which the 'blue/white' test for recombinants can be performed, the transformants were plated onto 2xTY/AMP plates not containing X-gal.

The overall transformation efficiency was calculated to be 6×10^7 transformants/ μ g pUC9. The re-ligated vector gave a total of 120 colonies, and the experimental plates a total of 2000 colonies.

Colonies from the experimental plates were transferred to HYBOND-N and denatured. These were screened with HXoxS2 insert. From this initial screening, 2 colonies were found to hybridise with the probe; this was significantly less than the number expected from the calculation given earlier. The colonies on the plate were sufficiently well spread out for the positive colonies to be easily identified. These were picked and grown for 6 hours in 5ml of 2xTY/AMP and plasmid DNA was

extracted. To release the inserts it was necessary to digest the plasmids with *EcoRI* and another enzyme 3' to the *BamHI* site, since this site is lost during the cloning of fragments with a *BglIII* end; *Sall* was used routinely. The products of the digest were analysed on a mini gel. The insert was not of the expected size (5.3kb), but was the same size as HXoxS2. This suggested that the HVR may not be stable in HB101 and that it had been deleted from the clone, in the same way that it had been deleted from the original cosmid. An alternative possibility was that one of the constituents of the ligation reaction was contaminated with this fragment. Suspect solutions were either replaced or tested for the contaminant by gel analysis and subsequent blot hybridisation. In no case could a fragment corresponding to putative contaminant be detected. The experiment was repeated using fresh solutions with the similar results. The single cross-hybridising clone found in this case also contained a small insert of the same size.

Since the HVR was apparently unstable in HB101, it was decided to repeat the experiment, this time using another strain of *E.coli* - DH5 α . This strain has three advantages over HB101. i) it is capable of α -complementation (ϕ 80d1acZAM15) of β -galactosidase and so the blue/white screening for recombinants can be used ii) it contains the *recA1* allele which prevents recombination of inserts more effectively than the *recA13* allele of HB101 iii) it contains a mutation which inactivates the endonuclease-1 gene, improving the quality of plasmid DNA from mini-preps. Competent cells of this strain were purchased from BRL. In addition to these, cells were also grown up and made competent using the previously described

method. Both lots of cells were transformed in a manner similar to that described above.

The overall transformation efficiency was calculated to be 3×10^7 transformants/ μ g pUC9. The re-ligated vector gave a total of 25 colonies, and the experimental plates a total of 5000 colonies. $\approx 90\%$ of these were judged to be recombinants from the ratio of white:blue colonies.

Colonies from the experimental plates were transferred to HYBOND-N and screened as above. 3 positive colonies were identified. Subsequent analysis showed that the inserts in these clones were also the same size as HXoxS2. These results suggested that the HVR was unstable even in *recA*⁻ strains.

This observation is not unique. The locus *D14S1* mentioned in the introduction has proved to be refractory to cloning by several techniques including the use of *recA*⁻ hosts (Wyman et al., 1985). However, they were successful in propagating phage λ clones containing *D14S1* on *E.coli* hosts containing mutant *recB*, *recC*, and *sbcB* genes. Their results had shown that the DNA of *D14S1* had secondary structure consistent with the presence of inverted repetitions of a few hundred base pairs. Leach and Stahl (1983) had previously found that deliberately constructed perfect inverted repetitions could be propagated only in *recB recC sbcB* hosts. These results, along with those obtained by Wertman et al. (1986), Wyman et al. (1986) and Donlon et al. (1986), suggest that a host deficient in exonucleases V and I (*recB recC sbcB*) is essential for the propagation of this type of sequence. The RecBC enzyme demonstrates several activities including cutting at *chi* sequences (Ponticelli et al., 1985 and Taylor et al., 1985),

if the HVR responsible for the observed hypervariability at *DXS255* contains sequences homologous to this structure or regions of inverted repeats, then this may explain the observed instability on hosts not deficient in this enzyme.

Two suitable hosts were available, SF8 (*recBC*⁻) and DL282 (*recBC*⁻ *sbcB*⁻). Competent cells were prepared from the two strains, which were then transformed with fraction 5 ligated into pUC9 as above. Half of the ligation mix was used to transform each strain. The results were as follows:

For SF8 - Transformation efficiency was 2×10^7 transformants/ μ g supercoiled pUC9.

Total number of colonies on experimental plates ≈ 1500 .

For DL282 - Transformation efficiency was 3×10^6 transformants/ μ g supercoiled pUC9.

Total number of colonies on experimental plates - 120.

Colonies from the experimental plates were transferred to HYBOND-N and screened as above. Each transformation yielded one positive colony. The positive colony from the DL282 transformation showed poor growth characteristics, both on plates and in liquid culture. However, subsequent analysis showed that the inserts in both of these clones were the same size as HXoxS2, again suggesting that the HVR had been deleted.

As a 'last ditch' attempt the vector pUC18 was used to try and clone the fragment in the opposite orientation. The rationale behind this was that when Jarman et al. (1986) were characterising a HVR located downstream of the human α -globin gene complex, they found that clones were inviable when the repeat was in one orientation. If this is true for the HVR associated with *DXS255* then the attempt was worth trying.

Fraction 5 was ligated into *EcoRI/BamHI* cut pUC18 as above, and half of the ligation mix was used to transform competent DH5 α or DL282 cells. Transformation efficiencies were similar to those obtained with pUC9, in the case of DL282 a little higher. $\approx$ 2000 colonies were screened from the DH5 α experimental plates and $\approx$ 400 from the DL282 plates. There were no positive colonies on the DL282 plates, but there were 4 positive colonies on the DH5 α plates. Subsequent analysis of these clones has shown them also to contain inserts of the same size as HXoxS2.

8.5 GENERAL DISCUSSION OF RESULTS

In spite of repeated attempts the HVR associated with DXS255 has remained refractory to cloning using the technique described. A plasmid library was successfully constructed using a DNA fraction enriched for sequences of the same size as the desired fragment. The number of recombinants obtained was in excess of the number calculated to achieve a representative library. The number of colonies hybridising with HXoxS2 was much less than expected, and in every case the clone contained an insert the same size as the probe and not one containing the putative HVR. Several strains of *E. coli* with different recombination deficiencies were used with similar results. Although it is not possible to definitely rule out the presence of contaminating HXoxS2 insert that was repeatedly cloned, solutions were either replaced or screened for the presence of such a contaminant and in no case could it be detected. If present, the concentration of this contaminant is substantially less than that of the desired fragment, which would

nevertheless suggest that the fragment containing extensive HVR regions is refractory to cloning.

One remaining strategy might be to screen a phage λ library that has been propagated on one of the recombination-deficient strains. As mentioned above, λ libraries propagated on strains similar to DL282 have successfully yielded other hypervariable loci. To reduce the number of recombinants that it would be necessary to screen, the enriched fraction could still be used in conjunction with a suitable vector. One possible approach would be to use the vector λ L47 (Loenen & Bramer, 1980). When digested with *EcoRI* this vector will accept fragments in the size range ≈ 9 –20kb. If a ≈ 6 kb *EcoRI/BglIII* 'stuffer' fragment is first ligated to one of the arms to generate a *BglIII* site, it may be possible to clone the either the 4.4kb or 5.3kb *EcoRI/BglIII* fragments from the enriched fractions. There would also be a selection against the cloning of the 0.9kb *EcoRI/BglIII* fragment, because an insert of this size would be too small to result in viable phages. Although recombinant phages made with λ L47 cannot be grown on a *recA*⁻ host, they can be grown under non-recombinogenic conditions on a *recB*, *recC* host.

It is intriguing that the clone M278, and clones isolated subsequently, appear to have the same sized deletion. Based on the data from the probing of chimpanzee DNA (see previous Chapter), it is tempting to speculate that these represent 'null' alleles. The mechanism of removal of the repeat is unknown, but if it contains regions of perfect inverted repeats, these may form a secondary structure that is acted upon by host enzymes. Even in a *recB recC sbcB* mutant host, the

presence of an alternative pathway for generalised recombination may be responsible for the deletion of such structures.

Data on the nature of the repeat may be obtainable from sequence analysis of the 330bp *Sau3A/HaeIII* fragment described in the previous chapter. Even if this sequence does not contain a copy of the repeat, if the region from which it has been deleted can be delineated further, it may be possible to synthesise oligonucleotides representing sequences flanking the site of the deletion and to use these as primers to obtain sequence data directly from genomic DNA. The technique of direct genomic sequencing has recently been described in Wong *et al.* (1987) and Wrischnik *et al.* (1987).

CHAPTER 9

SUMMARY AND CONCLUSIONS

This thesis has been concerned with the isolation and characterisation of sequences specific to the human sex chromosomes. Results obtained have given some insight into exchanges that have occurred between the X and Y chromosomes during recent evolutionary history and abnormal X-Y interchanges that result in males with an apparent 46, XX karyotype. One other recent evolutionary event appears to have been the acquisition of a hypervariable sequence on the human X chromosome short arm, since the divergence of man and chimpanzee.

An attempt was made to isolate sequences from terminal Xp, both by screening a cosmid library made from a somatic cell hybrid with the X chromosome as its only human component and by the technique of 'deletion enrichment'; neither of these techniques proved successful. The technique of 'deletion enrichment' was evaluated by attempting to isolate Y-specific sequences. The results obtained suggested that the degree of enrichment obtained was far less than hoped for, and that the technique adopted would be unlikely to yield sequences mapping to terminal Xp. Although the theoretical basis of this technique suggests that it should be useful in the isolation of sequences from a specific chromosomal location, the results described here, and those reported in the literature, imply that success relies on some degree of serendipity.

The availability of cDNA and genomic sequences for human STS has allowed a preliminary investigation of this locus in man and other species. The localisation of these sequences to Xp22.3, provides confirmation of the sub-regional assignment of the structural gene for *STS*. It is also consistent with the existence of a region escaping from inactivation at the distal terminus of the X chromosome short arm. It has been presumed that the region escaping from inactivation would correspond to that portion of the X chromosome sharing homology with the Y chromosome and which pairs and recombines with it at meiosis. Results obtained make it clear that *STS* sequences cannot be exchanged regularly between the sex chromosomes. It is possible, however, that the close proximity of the *STS* gene to the pseudoautosomal region on the X chromosome may predispose rearrangements resulting in deletions which may include this locus. It is therefore of considerable interest to note the high incidence of deletions found in patients with *STS* deficiency (Ballabio *et al.*, 1987; Yen *et al.*, 1987a; Gillard *et al.*, 1987).

The observation of *STS* gene sequences located on the long arm of the Y chromosome is intriguing. It is unlikely that they result from viral retroposition as presumptive intron sequences (Yen *et al.*, 1987a), and sequences thought to map 3' to the coding regions have been detected on the Y chromosome. Furthermore, it appears that there are additional sequences unrelated to *STS* which show a similar distribution on the X and Y chromosomes i.e. Xp22.3 and Yq11 (e.g. *DXS31*, Koenig *et al.*, 1985; *DXS69*, Kunkel *et al.*, 1983; *GMGXY3*, Affara *et al.*, 1986), although their relationship to the event leading to the

localisation of *STS* homologous sequences on Yq remains unclear. Phylogenetic studies have shown that sequences homologous to *DXS31* and *STS* are X- and Y-linked in the chimpanzee, but appear to be only X-linked in more distantly related species. There is provisional evidence to suggest that these loci are (pseudo)autosomally located in the lemur and in the case of *STS*, the rabbit. Evidence has been presented for X- and Y-localisation of *STS* in the mouse with regular exchange between the alleles on the two chromosomes (Keitges *et al.*, 1985). It has not been able to confirm these data using the *STS* probes described here, since there is little or no cross-hybridisation between the human and murine sequences. The simplest explanation for the apparent differences in the organisation of the *STS* gene in mice and humans, is that the human Y chromosome has rearranged since their divergence and reflects a rearrangement from an ancestral sex chromosome organisation in which both X and Y alleles were functional and whose identities were maintained by crossing-over (see Craig *et al.*, 1987). However, it may be necessary to propose a more complex pathway when the precise distribution of *STS* sequences, and other sequences in this block of homology, is known. It is possible that more than one rearrangement has resulted in the pattern homology seen in man.

Further studies on the organisation of the *STS* gene itself, may give some insight into the mechanisms of X-inactivation. DNA methylation of the gene has been studied by Yen *et al.* (1987b) using probes from the 5' and 3' ends of the gene. In contrast to several genes studied which undergo X-inactivation and which were found to be more methylated on

an inactive X chromosome than on an active X chromosome, no significant differences in methylation were observed between the *STS* genes on active and inactive X's.

It has been possible to analyse another region of X-Y homology using the probe 59γ, which detects the locus *DXYS27* located at Xq21 and on Yp. Restriction enzyme analysis and direct sequence comparison of the X and Y copies of *DXYS27* has shown them to be ≈99% homologous, in keeping with the proposed homology of other X-Y homologous sequences that show this Yp:Xq distribution. Phylogenetic studies performed by Page *et al.* (1984) using *DXYS1*, the 'prototype' of this block of X-Y homology, had suggested that this block was X-, but not Y-, linked in the chimpanzee and that an evolutionarily recent X-Y transposition was the cause of this homology. Results presented here are in agreement with this hypothesis. Sequences homologous *DXYS27* have been identified in primate species thought to have diverged from the line leading to modern man ≈40 Myr ago. Phylogenetic studies have also suggested that the mutations resulting in the X-Y differences have occurred on both the X and Y chromosomes. Although X polymorphisms have not been detected for *DXYS27*, other groups have found such differences for similar loci. A number of disease loci have been shown to map to this region of the X chromosome and the identification of such polymorphisms will be useful in linkage analysis of *DXYS27* to these. The *DXYSn* loci constitute a large block of material (estimated to constitute as much as 1/4 Of the euchromatic region of Yp), it will be of interest to compare the organisation of these sequences on both the X and Y chromosomes. The presence of X-Y polymorphisms detected at

DXYS27 has facilitated the analysis of DNA from XX males and an XY female, for evidence of aberrant X-Y interchange. It appears that along with many other sequences present on Yp, the Y homologue of *DXYS27* is present in many, but not all XX males, presumably as the result of an aberrant X-Y interchange which has also mobilised *TDF*. Studies of XX males by other groups have enabled the identification of probes detecting low-copy Y-specific sequences that map to the interval embracing the testis determining gene and which are frequently deleted in XY females. These are being investigated by a number of groups as possible starting points for walking, long-range restriction mapping and for the identification of sequence motifs characteristic of transcribed sequence and before long the *TDF* locus itself will be cloned.

A hypervariable locus, *DXS255*, mapping to Xp11.3→Xcen has been identified using the probe M27 β , isolated from the MOG cosmid library. This locus has many features that suggest that the observed hypervariability is due to the presence of a minisatellite sequence of variable length in different individuals. However, the cloned sequence M27 β , appears to have had this region deleted and repeated attempts to isolate clones containing the repeat have failed. The inherent instability of this type of sequence has been noted previously. Other strategies for gaining information on the nature of this sequence are currently being devised. The heritability of the restriction fragment length polymorphism detected by M27 β has been examined. The Mendelian mode of inheritance suggests that the polymorphism detected is based on sufficiently stable allelic variation for it to be of use in linkage analysis. A

preliminary investigation of the linkage between *DXS255* and the locus for retinitis pigmentosa has been undertaken. This study has demonstrated the ease at which different X chromosomes can be followed through several generations, with nearly all individuals being informative. Results were encouraging, and further studies are currently in progress. Many other uses for the probe have been suggested, some of these are currently being investigated. Undoubtedly, it has great value as a '2nd probe' to verify the paternity and sex of samples used in a range of analyses.

Phylogenetic studies have suggested that *DXS255* is hypervariable only in humans. Although the study has not been exhaustive, the size of the fragment detected in ♂ and ♀ chimpanzee appears to be same size as the cloned, presumptively, 'null' allele and it is tempting to speculate that the hypervariable repeat sequence has been an evolutionarily recent acquisition by the human X chromosome. The origins of this sequence are unknown, however a concerted attempt is being made to isolate and characterise the repeat in the hope of gaining more information on its organisation and possible function.

It can be concluded from these and the results of others, that the human Y chromosome is a recent evolutionary patchwork of sequences sharing various degrees of homology with the X chromosome, and it is likely that sequences have been acquired by both the X and Y chromosomes from other sources (see Figure 9.1). Whether the rearrangements resulting in these blocks of homology were catalysed by preexisting homologies is not known, but several seem to have taken place. A region located at the

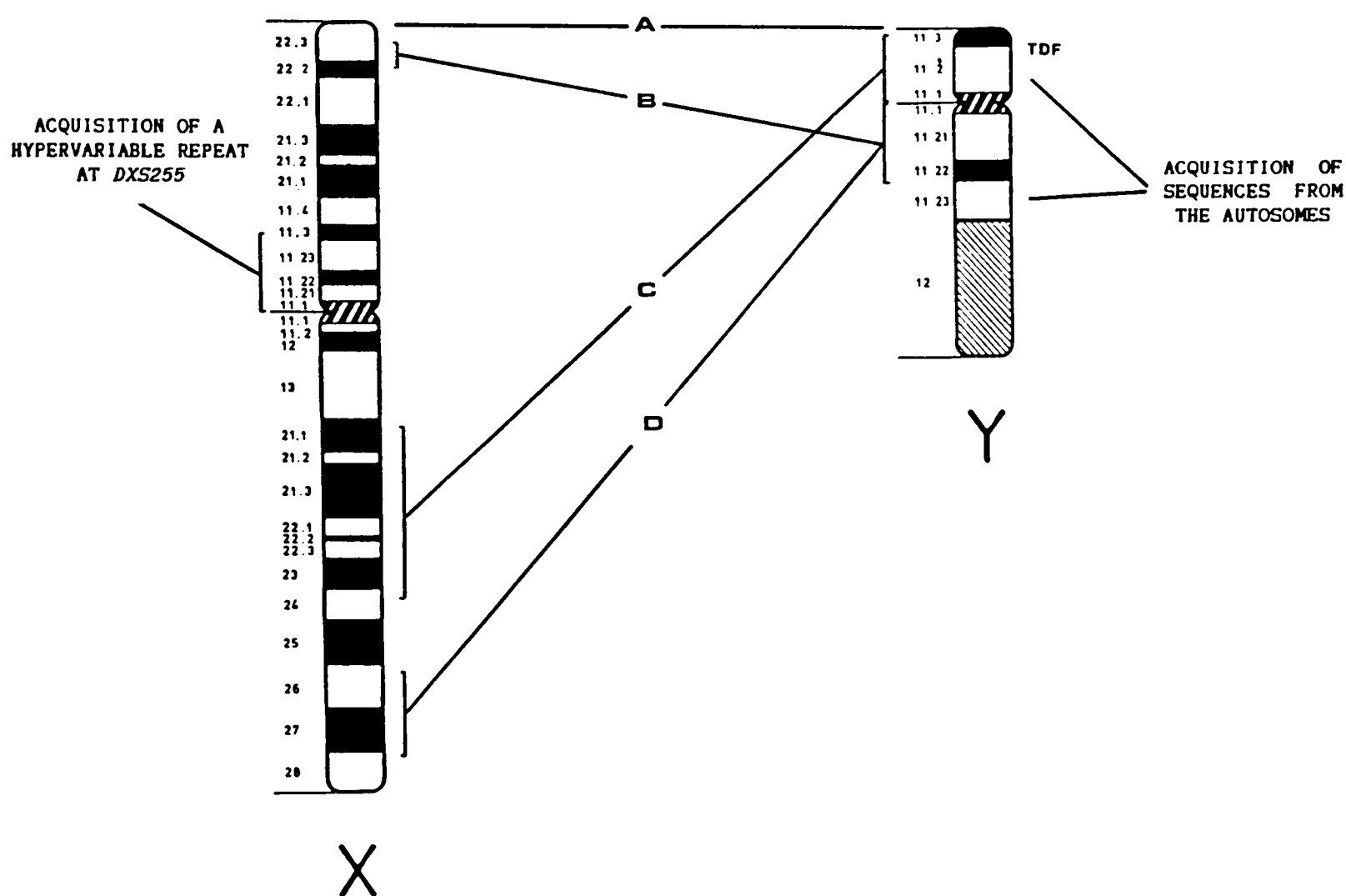


FIGURE 9.1

Evolutionarily recent exchanges involving the sex chromosomes

A: Pseudoautosomal region of regular X-Y interchange. A gradient of recombination exists in this region with sequences near the telomeres exchanging most frequently. Abnormal X-Y interchange resulting in the transfer of *TDF* to the X chromosome occurs at frequency of about 10^{-4} , this exchange often includes sequences from region C.

B: Sequences in this region include *STS*, *DXS31*, *DXS69* and *GMGXY3*. The X and Y copies are 80-90% homologous. Sequences on Yq may be the result of a transposition of material from the X chromosome or a pericentric inversion moving sequences from a previously pseudoautosomal location ≈ 15 Myr ago.

C: Sequences in this region include *DXYS1* and *DXYS27*. The X and Y copies are 97-99% homologous. The sequences on Yp appear to be the result of a transposition of material from the X chromosome during the past ≈ 5 Myr.

D: The X and Y copies of the sequence 2:13 are thought to be 100% homologous and may represent a very recent exchange between the X and Y chromosomes (< 1 Myr).

The hypervariable repeat associated with *DXS255* appears to have been acquired by the human X chromosome during the past ≈ 5 Myr. Sequences appear to have been acquired by the Y chromosome from the autosomes at various stages of evolution.

tips of the short arms of the sex chromosomes is still regularly exchanged between these chromosomes, with a gradient of recombination from the telomere down towards the centromere (Rouyer *et al.*, 1986). Normally this exchange does not involve *TDF*. However, it seems that occasionally an aberrant X-Y interchange takes place, mobilising *TDF* and often more proximal loci (including *DXYS27*), resulting in males with a 46, XX karyotype. Apart from this terminal region of 'true' X-Y homology, which does not, however, appear to be conserved in many other species (P.N. Goodfellow, personal communication), there do not appear to be regions representing the proposed 'ancient homology' of X and Y chromosomes; existing regions of homology are the result of much more recent events.

APPENDIX

A PROGRAM TO CALCULATE THE SIZES OF DNA FRAGMENTS IN AGAROSE GELS FROM THE MOBILITY OF STANDARD FRAGMENTS

The method commonly used to estimate the size of DNA fragments in agarose gels is the construction of a semi-logarithmic standard curve. $\text{Log}(\text{size})$ is plotted against migration for a set of known DNA fragments and the size of unknown fragments may then be read from the line of best fit. This method is notoriously inaccurate at both high and low molecular weights and is less accurate than other methods based on a reciprocal migration/size relationship. The program listed here is based on a program written by George. C. Russell of the Institute of Genetics, University of Glasgow, for the BBC micro. This program uses a least squares method to calculate the sizes of unknown DNA fragments from known size standards. The version shown here has been written in Mallard BASIC for the Amstrad *PCW8256*.

```

10 REM Program to calculate the sizes of DNA fragments in
20 REM agarose gels from the mobility of standard fragments.
30 REM A modified version of the program written for the BBC micro.
40 REM by George C. Russell, Institute of Genetics, University of Glasgow.
50 REM Written in Mallard Basic for the Amstrad PCW8256
60 REM Neil Fraser - March 1987
70 CLEAR
80 LPRINT CHR$(15)
90 LPRINT CHR$(27)+"3"+CHR$(15)
100 OPTION BASE 1:IX=1
110 REM Sets minimum value of array subscript to 1
120 STAN=0
130 DIM WT(50),DI(50),PR(50),DW(50),DD(50),DP(50),C(50),D(50)
140 CLS$=CHR$(27)+"E"+CHR$(27)+"H"
150 REM 'PRINT CLS$' will now clear the screen
160 PRINT CLS$
170 PRINT:PRINT:PRINT:PRINT:PRINT:PRINT
180 PRINT TAB(10) "1. END"
190 PRINT TAB(10) "2. ENTER DATA FOR M.W. STANDARDS"
200 IF STAN<1 THEN 240
210 PRINT TAB(10) "3. DISPLAY STANDARD FIT"
220 PRINT TAB(10) "4. CALCULATE M.W. OF UNKNOWN FRAGMENTS"
230 PRINT TAB(10) "5. CLEAR ALL PREVIOUS DATA"
240 PRINT:PRINT:PRINT TAB(20) "ENTER CHOICE"
250 GOSUB 1690
260 REM Waits for a key to be pressed
270 IF VAL(A$)<1 OR VAL(A$)>5 THEN 250
280 IF STAN<1 AND VAL(A$)>2 THEN 250
290 ON VAL(A$) GOTO 1670,300,1220,1370,70
300 PRINT CLS$
310 REM Enter data of MW standards
320 STAN=1
330 PRINT:PRINT "Have you used any of these standards ? (Y/N) "
340 PRINT:PRINT "Lambda/HindIII, Lambda/EcoRI, Lambda/HindIII/EcoRI"
350 PRINT:PRINT "pBR322/HaeIII, pBR322/Sau3A1"
360 PRINT:PRINT:PRINT
370 GOSUB 1690
380 IF A$="Y" OR A$="y" THEN GOTO 510
390 IF A$<>"N" AND A$<>"n" THEN 370
400 PRINT CLS$
410 INPUT "Enter number of standards ",N%
420 IF N%<4 THEN PRINT "You must enter at least 4 standards" ELSE GOTO 460
430 PRINT:PRINT "Press any key to continue"
440 GOSUB 1690
450 GOTO 400
460 PRINT
470 PRINT "Enter your data (Format: FRAGMENT SIZE, MIGRATION <RETURN>)"
480 PRINT:PRINT "FRAGMENT "
490 FOR IX=1 TO N%:PRINT: PRINT "      "IX": ";:INPUT "", WT(IX),DI(IX):NEXT
500 GOTO 890
510 PRINT CLS$
520 PRINT "Which of these standards have you used ?"
530 PRINT "ENTER NUMBER"
540 PRINT:PRINT "      1. Lambda/HindIII"
550 PRINT "      2. Lambda/EcoRI"
560 PRINT "      3. Lambda/HindIII/EcoRI"
570 PRINT "      4. pBR322/HaeIII"
580 PRINT "      5. pBR322/Sau3A1"
590 GOSUB 1690
600 IF VAL(A$)<1 OR VAL(A$)>5 THEN 590
610 ON VAL(A$) GOTO 620,630,640,650,660

```

```

620 V%=8;RESTORE 1620;GOTO 670
630 V%=6;RESTORE 1630;GOTO 670
640 V%=13;RESTORE 1640;GOTO 670
650 V%=18;RESTORE 1650;GOTO 670
660 V%=12;RESTORE 1660;GOTO 670
670 PRINT CLS$
680 PRINT;PRINT;PRINT "Enter migration of Standard Fragments"
690 PRINT;PRINT "To omit a fragment press <RETURN>"
700 PRINT;PRINT TAB(2)"Fragment";TAB(15)"Size";TAB(28)"Migration"
710 FOR J%=1 TO V%;READ X%
720 PRINT;PRINT TAB(4)J%;; PRINT TAB(14)X%;; PRINT TAB(32);;INPUT Y$
730 IF Y$="" THEN 750
740 WT(I%)=X%;DI(I%)=VAL(Y$);I%=I%+1
750 NEXT
760 N%=I%-1;PRINT "Do you want to add any more standards ? (Y/N)"
770 GOSUB 1690
780 IF A$="N" OR A$="n" THEN 890
790 IF A$<>"Y" AND A$<>"y" THEN 770
800 PRINT "1. Use standard sets";PRINT "2. Input individually"
810 GOSUB 1690
820 IF VAL(A$)=1 THEN 510
830 IF VAL(A$)<>2 THEN 810
840 INPUT "How many more ";K%
850 PRINT "Enter your data (Format: FRAGMENT SIZE, MIGRATION <RETURN>)"
860 PRINT;PRINT "FRAGMENT "
870 PRINT;FOR LX=1 TO K%; PRINT "      "I%"; ";;INPUT "", WT(I%),DI(I%)
880 I%=I%+1;NEXT;N%=I%-1
890 REM check data
900 PRINT;PRINT
910 PRINT "Check data ? (Y/N)"
920 GOSUB 1690
930 IF A$="Y" OR A$="y" THEN PRINT CLS$;PRINT TAB(5);"Fragment";TAB(16);"Size";TAB(25)"Migration";PRINT;
FOR J%=1 TO N%;PRINT TAB(8);
J%;TAB(15);WT(J%);TAB(28);DI(J%);PRINT;GOSUB 1720;NEXT;GOTO 960
940 IF A$="n" OR A$="N" THEN 1090
950 IF A$<>"N" AND A$<>"n" THEN 920
960 REM Edit data
970 PRINT "Edit data ? (Y/N)"
980 GOSUB 1690
990 IF A$="N" OR A$="n" THEN 1090
1000 IF A$<>"Y" AND A$<>"y" THEN 980
1010 PRINT;PRINT
1020 INPUT"Fragment no. ";J%
1030 INPUT"Fragment Size ";WT(J%)
1040 INPUT"Migration ";DI(J%)
1050 PRINT;PRINT;PRINT "Do you wish to edit any more (Y/N) ?"
1060 GOSUB 1690
1070 IF A$="Y" OR A$="y" THEN 1010
1080 GOTO 890
1090 REM Standard Calculations
1100 SW=0;SI=0;SP=0;CS=0;CM=0;CC=0;CP=0;CL=0;S=0;SS=0
1110 FOR I%=1 TO N%;SW=SW+WT(I%);SI=SI+DI(I%);PR(I%)=WT(I%)*DI(I%);SP=SP+PR(I%);NEXT
1120 MW=SW/N%;MD=SI/N%;MP=SP/N%
1130 FOR I%=1 TO N%;DW(I%)=WT(I%)-MW;DD(I%)=DI(I%)-MD;DP(I%)=PR(I%)-MP;NEXT
1140 FOR I%=1 TO N%;CS=CS+DW(I%)^2;CM=CM+DD(I%)^2;CC=CC+DW(I%)*DD(I%);CP=CP+DP(I%)*DW(I%);CL=CL+DP(I%)*D(I%);NEXT
1150 DT=CS*CM-CC^2;M=(CM*CP-CC*CL)/DT;LD=(-CC*CP+CS*CL)/DT
1160 FOR I%=1 TO N%;C(I%)=(WT(I%)-LD)*(DI(I%)-M);S=S+(C(I%)-C(1));SS=SS+(C(I%)-C(1))^2;NEXT
1170 CB=S/N%+C(1);SC=SQR((SS-S^2/N%)/(N%-1))
1180 DEF FNA(Z)=ABS(INT(Z+0.5))

```

```

1190 DEF FNB(Z)=ABS(INT(Z*10+0.5)/10)
1200 REM Return to main menu
1210 PRINT CLS$:GOTO 170
1220 REM Standard Fit
1230 PRINT cls$
1240 PRINT:PRINT"STD LEN  DIST  PRED LEN  DEV  %DEV":PRINT
1250 FOR J%=1 TO N%:PW=CB/(DI(J%)-M)+LQ:WD=WT(J%)-PW:PC=100*WD/WT(J%):SD=SD+WD:D=D+(WD^2)
1260 PRINT:PRINT TAB(0)WT(J%);TAB(10)DI(J%);TAB(17)FNA(PW);TAB(27)FNA(WD);TAB(34)FNB(PC):GOSUB 1720:NEXT
1270 PRINT:PRINT "Do you want a hard copy ? (Y/N)"
1280 GOSUB 1690
1290 IF A$="Y" OR A$="y" THEN 1300 ELSE 1340
1300 LPRINT:LPRINT"STD LEN  DIST  PRED LEN  DEV  %DEV":LPRINT
1310 FOR J%=1 TO N%:PW=CB/(DI(J%)-M)+LQ:WD=WT(J%)-PW:PC=100*WD/WT(J%):SD=SD+WD:D=D+(WD^2)
1320 LPRINT:LPRINT TAB(0)WT(J%);TAB(10)DI(J%);TAB(17)FNA(PW);TAB(27)FNA(WD);TAB(34)FNB(PC):NEXT
1330 LPRINT:LPRINT:LPRINT
1340 PRINT:PRINT:PRINT "Press any key to continue"
1350 GOSUB 1690
1360 GOTO 160
1370 REM Calculate M,W, of unknown fragments
1380 PRINT CLS$:INPUT "Enter number of unknown fragments ",U%
1390 PRINT "Enter migration of fragment:"
1400 FOR J%=1 TO U%
1410 PRINT TAB(28)J%;;INPUT ": ",D(J%)
1420 NEXT
1430 PRINT cls$
1440 PRINT "For migration:      Predicted size is:"
1450 FOR J%=1 TO U%:PW=CB/(D(J%)-M)+LQ
1460 PRINT
1470 PRINT TAB(8)D(J%);TAB(29)FNA(PW)
1480 GOSUB 1720
1490 NEXT
1500 PRINT:PRINT "Do you want a hard copy ?"
1510 GOSUB 1690
1520 IF A$="Y" OR A$="y" THEN 1530 ELSE 1590
1530 LPRINT "For migration:      Predicted size is:"
1540 FOR J%=1 TO U%:PW=CB/(D(J%)-M)+LQ
1550 LPRINT
1560 LPRINT TAB(8)D(J%);TAB(29)FNA(PW)
1570 NEXT
1580 LPRINT:LPRINT:LPRINT
1590 PRINT:PRINT:PRINT "Press any key"
1600 GOSUB 1690
1610 GOTO 160
1620 DATA 23130,9416,6557,4361,2322,2027,564,125
1630 DATA 21226,7421,5804,5643,4878,3530
1640 DATA 21226,5148,4973,4268,3530,2027,1904,1584,1375,947,831,564,125
1650 DATA 587,540,504,458,434,267,234,213,192,184,124,123,104,89,80,64,57,51
1660 DATA 1374,665,358,341,317,272,258,207,105,91,78,75
1670 PRINT cls$:CLEAR:END
1680 REM GOSUB routines
1690 A$="":WHILE A$="":A%=INKEY$:WEND
1700 REM Waits for a key to be pressed
1710 RETURN
1720 SC=J%/13
1730 IF INT(SC)=SC THEN PRINT:PRINT "Press any key to continue": A$="":WHILE A$="":A%=INKEY$:WEND:PRINT
cls$
1740 REM Checks to see if screen is full of data and pauses if it is
1750 RETURN

```

REFERENCES

- Affara, N.A., Florentin, L., Morrison, N., Kwok, K., Mitchell, M., Cook, A., Jamieson, D., Glasgow, L., Meredith, L., Boyd, E. & Ferguson-Smith, M.A. (1986a). Regional assignment of Y-linked DNA probes by deletion mapping and their homology with X-chromosome and autosomal sequences. *Nucleic Acids Res.* 14, 5353-5373.
- Affara, N.A., Ferguson-Smith, M.A., Tolmie, J., Kwok, K., Mitchell, M., Jamieson, D., Cooke, A. & Florentin, L. (1986b). Variable transfer of Y-specific sequences in XX males. *Nucleic Acids Res.* 14, 5375-5387.
- Affara, N.A., Ferguson-Smith, M.A., Magenis, R.E., Tolmie, J.L., Boyd, E., Cooke, A., Jamieson, D., Kwok, K., Mitchell, M. & Snadden, L. (1987). Mapping the testis determinants by an analysis of Y-specific sequences in males with apparent XX and XO karyotypes and females with XY karyotypes. *Nucleic Acids Res.* 15, 7325-7342.
- Anand, R. (1986). Pulsed field gel electrophoresis: a technique for fractionating large DNA molecules. *Trends Genet.* 2, 278-283.
- Andersson, M., Page, D.C. & de la Chapelle, A. (1986). Chromosome Y-specific DNA is transferred to the short arm of the X chromosome in human XX males. *Science* 233, 786-788.
- Anson, D.S., Choo, K.H., Rees, D.J.G., Gianelli, F., Gould, K., Huddleston, J.A. and Brownlee, G.G. (1984). The gene structure of human antihaemophilic factor IX. *EMBO J.* 3, 1053-1060.
- Avery, R.J., Norton, J.D., Jones, J.S., Burke, D.C. & Morris, A.G. (1980). Interferon inhibits transformation by murine sarcoma viruses before integration of provirus. *Nature* 288, 93-95.
- Ballabio, A., Parenti, G., Carrozzo, R., Sebastio, G., Andria, G., Buckle, V., Fraser, N., Craig, I., Rocchi, M., Romeo, G., Jobsis, A.C. & Persico, M.G. (1987). Isolation and characterization of a steroid sulfatase cDNA clone: Genomic deletions in patients with X-chromosome-linked ichthyosis *Proc. Natl. Acad. Sci. USA* 84, 4519-4523.
- Bell, G.I., Selby, M.J. & Rutter, W.J. (1982). The highly polymorphic region near the human insulin gene is composed of simple tandem repeating sequences. *Nature* 295, 31-35.
- Bhattacharya, S.S., Wright, A.F., Clayton, J.F., Price, W.H., Phillips, C.I., McKeown, C.M.E., Jay, M., Bird, A.C., Pearson, P.L., Southern, E.M. & Evans, H.J. (1984). Close genetic linkage between X-linked retinitis pigmentosa and a restriction fragment length polymorphism identified by recombinant DNA probe L1.28. *Nature* 309, 253-255.

- Bickmore, W.A. & Cooke, H.J. (1987a). Evolution of homologous sequences on the human X and Y chromosomes, outside the meiotic pairing segment. *Nucleic Acids Res.* 15, 6261-6271.
- Bickmore, W.A. & Cooke, H.J. (1987b). Characterisation of a sequence from the human Y chromosome showing a high degree of homology to the X: enrichment cloning of DNA unique to the Y chromosome. *Cytogenet. Cell Genet.* In Press.
- Bird, A.P. (1986). CpG-rich islands and the function of DNA methylation. *Nature* 321, 209-213.
- Bishop C.E., Guellaen G., Geldwerth D., Voss R., Fellous, M and Weissenbach, J. (1983). Single copy DNA sequences specific for the human Y chromosome. *Nature* 303, 831-832.
- Boyd, Y. (1987). Characterisation and use of somatic cell hybrids with interspecific translocations involving the human X chromosome. *Ann. Hum. Genet.* 51, 13-26.
- Boyd, Y, Munro, E., Ray, P., Monaco, A., Kunkel, L. and Craig, I.W. (1987). Molecular heterogeneity of translocations associated with muscular dystrophy. *Clin. Genet.* 31, 265-272.
- Brison, O., Ardeshir, F. & Stark, G.R. (1982). General method for cloning amplified DNA by differential screening with genomic probes. *Mol. Cell. Biol.* 2, 578-587.
- Brutlag, D., Fry, K., Nelson, T. & Hung, P. (1977). Synthesis of hybrid bacterial plasmids containing highly repeated satellite DNA. *Cell* 10, 509-519.
- Buckle, V., Mondello, C., Darling, S., Craig, I.W. & Goodfellow, P.N. (1985). Homologous expressed genes in the human sex chromosome pairing region. *Nature* 317, 739-741.
- Buckle, V.J. Boyd, Y., Fraser, N., Goodfellow, P.N. Goodfellow, P.J., Wolfe, J. & Craig, I.W. (1987). Localisation of Y chromosome sequences in normal and 'XX' males. *J. Med. Genet.* 24, 197-203.
- Burgoyne, P.S. (1982). Genetic homology and crossing over in the X and Y chromosomes of mammals. *Hum. Genet.* 61, 85-90.
- Burk, R.D., Ma, P. & Smith, K.D. (1985). Characterisation and evolution of a single copy sequence from the human Y chromosome. *Mol. Cell Biol.* 5, 576-581.
- Burke, T & Bruford, M.W (1987). DNA fingerprinting in birds. *Nature* 327, 149-152.
- Capon, D.J., Chen, E.Y., Levinson, A.D., Seeburg, P.H. & Goeddel, D.V. (1983). Complete nucleotide sequence of the T24 human bladder carcinoma oncogene and its normal homologue. *Nature* 302, 33-37.

- Chandley, A.C., Goetz, P., Hargreave, T.B. Joseph, A.M. & Speed, R.M. (1984). On the nature and extent of XY pairing at meiotic prophase in man. *Cytogenet. Cell genet.* 38, 241-247.
- de la Chapelle, A., Tippet, P.A., Wetterstrand, G. & Page, D. (1984). Genetic evidence of X-Y interchange in human XX males. *Nature* 307, 170-171.
- Charlesworth, B. (1978). Model for the evolution of Y chromosomes and dosage compensation. *Proc. Natl. Acad. Sci. USA* 75, 5618-5622.
- Cooke H.J., Fantes J. & Green, D. (1983). Structure and evolution of human Y chromosome DNA. *Differt.* 23(s), 48-55.
- Cooke H.J., Brown, W.A.R. & Rappold, G.A. (1984). Closely related sequences on the human X and Y chromosomes outside the pairing region. *Nature* 311, 259-261.
- Cooke, H.J., Brown, W.R.A. & Rappold, G.A. (1985). Hypervariable telomeric sequences from the human sex chromosomes are pseudoautosomal. *Nature* 317, 687-692.
- Cooper, D.N. & Schmidtke, J. (1984). DNA restriction fragment length polymorphisms and heterozygosity in the human genome. *Hum. Genet.* 66, 1-16
- Cotinot, C., Kirszenbaum, M., Bishop, C., Vaiman, M. & Fellous, M. (1987). Cloning and characterization of bovine Y derived sequences. *Development* 101, (suppl.). In Press.
- Craig, I.W. & Goodfellow, P.N. (1986). Biology of disease: Molecular genetics of the X chromosome and X-linked diseases. *Laboratory Investigation* 54, 241-253.
- Craig, I.W. & Tolley, E. (1986) Steroid sulphatase and the conservation of mammalian X chromosomes. *Trends Genet.* 2, 201-204.
- Craig, I.W., Levy, E. & Fraser, N. (1987). The mammalian Y chromosome: molecular search for the sex-determining gene - summary and perspectives. *Development* 101 (suppl.). In Press.
- Crocker, M. & Craig, I.W. (1983). Variation in the regulation of steroid sulphatase locus in mammals. *Nature* 303, 721-722.
- Curry, C.R., Magenis, R.E., Brown, M., Lanman, J.T., Tsai, J., O'Lague, P., Goodfellow, P., Mohandas, T., Bergner, E.A., & Shapiro L.J. (1984). Inherited chondrodysplasia punctata due to a deletion of the terminal short arm of an X chromosome. *N. Eng. J. Med.* 311, 1010-1015.
- Darling, S.M., Banting, G.S., Pym, B., Wolfe, J. & Goodfellow, P.N. (1986). Cloning an expressed gene shared by the human sex chromosomes. *Proc. Natl. Acad. Sci. USA* 83, 135-139.
- Davies K.E., Young B.D., Elles R.G., Hill M.E. & Williamson R. (1981). Cloning of a representative genomic library of the

human X chromosome after sorting by flow cytometry. *Nature* 293, 374-376.

Davies, K.E. (1985). Molecular genetics of the human X chromosome. *J. Med. Genet.*, 22, 243-249.

Davis, R.M. (1981). Localisation of male determining factors in man: a thorough review of structural anomalies of the Y chromosome. *J. Med. Genet.* 18, 161-195.

Dietz Band, J., Skolnick, M. & Barker, D. (1987). Isolation and characterization of highly polymorphic minisatellite markers (VNTRs) for the human X chromosome. *Cytogenet. Cell Genet.* In Press.

Dosik, H., Wachtel, S.S., Khan, F., Spergel, G.E. & Koo, G.C. (1976). Y chromosomal genes in a phenotypic male with a 46,XX karyotype. *J.A.M.A.* 236, 2505-2508.

Drayna, D., Davies, K., Hartley, D., Mandel, J.-L., Camerino, G., Williamson, R. & White, R. (1984). Genetic mapping of the human X chromosome by using restriction fragment length polymorphisms. *Proc. Natl. Acad. Sci. USA* 81, 2836-2839.

Dretzen, G., Bellard, M., Sassone-Corsi, P. & Chambon, P. (1981). A reliable method for the recovery of DNA fragments from agarose and acrylamide gels. *Anal. Biochem.* 112, 295-299.

Epstein, E.H. & Bonifas, J.M. (1985). Recessive X-linked ichthyosis: lack of immunologically detectable steroid sulphatase enzyme protein. *Hum. Genet.* 71, 201-210.

Evans, H.J., Buckton, K.E., Spowart, G. & Carothers, A.D. (1979). Heteromorphic X chromosomes in 46,XX males: evidence for the involvement of X-Y interchange. *Hum. Genet.* 49, 11-31.

Ewens, W.J., Spielman, R.S. & Harris, H. (1981). Estimation of genetic variation at the DNA level from restriction endonuclease data. *Proc. Natl. Acad. Sci. USA* 78, 3748-3750.

Ferguson-Smith, M.A. (1966). X-Y chromosomal interchange in the aetiology of true hermaphroditism and of XX Klinefelter's syndrome. *Lancet* ii, 475-476.

Francke, U, Bakay, B., Connor, J.D., Coldwell, J.G. & Nyhan, W.L. (1974). Linkage relationships of X-linked enzymes glucose-6-phosphate dehydrogenase and hypoxanthine guanine phosphoribosyltransferase: recombination in female offspring of compound heterozygotes. *Am. J. Hum. Genet.* 26, 512-522.

Fraser, N., Ballabio, A., Zollo, M., Persico, M.G. & Craig, I.W. (1987a) Identification of incomplete coding sequences for steroid sulphatase on the human Y chromosome: Evidence for an ancestral pseudoautosomal gene ? *Development* 101 (suppl.). In Press.

Fraser, N., Buckle, V., Levy, E., Ballabio, A., Zollo, M., Persico, G. and Craig, I. (1987b) Identification of incomplete

coding sequences for steroid sulphatase on the human Y chromosome: Evidence for an ancestral pseudoautosomal gene? *Cytogenet. Cell Genet.* In press.

Gal, A., Stolzenberger, C., Weinker, T., Weiacker, P., Ropers, H.-H., Friedrich, U., Bleeker-Wagemakers, L., Pearson, P. & Warburg, M. (1985) Norrie's disease: close linkage with genetic markers from the proximal short arm of the human X chromosome. *Clin. Genet.* 27, 282-283.

Gartler, S.M. & Rivest, M. (1983). Evidence for X-linkage of steroid sulfatase in the mouse: steroid sulfatase levels in oocytes of XX and XO mice. *Genetics* 103, 137-141.

Geldwerth, D., Bishop, C., Guellaen, G., Koenig, M., Vergnaud, G., Mandel, J.-L. & Weissenbach, J. (1985). Extensive DNA sequence homologies between the human Y and the long arm of the X chromosome. *EMBO J.* 4, 1739-1743.

Geller, R.L., Shapiro, L.J. & Mohandas, T. (1986). Fine mapping of the distal short arm of the human X chromosome using X/Y translocations. *Am. J. Hum. Genet.* 38, 884-890.

Gill, P., Jeffreys, A.J. & Werret, D.J. (1985). Forensic applications of DNA 'fingerprints'. *Nature* 318, 577-579.

Gillard, E.F., Affara, N.A., Yates, J.R.W., Goudie, D.R., Lambert, J., Aitken, D.A. & Ferguson-Smith, M.A. (1987). Deletion of a DNA sequence in eight of nine families with X-linked ichthyosis (steroid sulphatase deficiency). *Nucleic Acids Res.* 15, 3977-3985.

Goldman, D. Giri, P.R. & O'Brien, S.J. (1987). A molecular phylogeny of the hominoid primates as indicated by two-dimensional protein electrophoresis. *Proc. Natl. Acad. Sci. USA* 84, 3307-3311.

Goodfellow, P.J., Darling, S.M., Thomas, N.S. & Goodfellow, P.N. (1986). A pseudoautosomal gene in man. *Science* 234, 740-743.

Goodfellow, P.N., Banting, G., Trowsdale, S., Chambers, S., & Solomon, E. (1982). Introduction of a human X-6 translocation chromosome into a mouse teratocarcinoma: Investigation of control of *HLA-A,B,C* expression. *Proc. Natl. Acad. Sci. USA* 79, 1190-1194.

Goodfellow, P., Banting, G., Sheer, D., Ropers, H.-H., Caine, M.A., Ferguson-Smith, M.A., Povey, S. & Voss, R. (1983). Genetic evidence that a Y-linked gene in man is homologous to a gene on the X chromosome. *Nature* 302, 346-349.

Goodfellow, P., Pym, B., Mohandas, T. & Shapiro, L.J. (1984). The cell surface antigen locus *MIC2X* escapes X-inactivation. *Am. J. Hum. Genet.* 36, 777-782.

- Goodfellow, P., Davies, K. & Ropers H.-H. (1985). Report of the committee on the genetic constitution of the X and Y chromosomes. *Cytogenet. Cell Genet.* 40, 296-352.
- Goss, S. & Harris, H. (1977). Gene transfer by means of cell fusion: radiation-induced gene segregation. *J. Cell. Sci.* 25, 17-37
- Graves, J.A.M. (1987). The evolution of mammalian sex chromosomes and dosage compensation: clues from marsupials and monotremes. *Trends Genet.* 3, 252-256.
- Grosveld, F.G., Lund, T., Murray, E.J., Mellor, A.L., Dahl, H.H.M. & Flavell, R.A. (1982). The construction of cosmid libraries which can be used to transform eukaryotic cells. *Nucleic Acids Res.* 10, 6715-6732.
- Guellaen, G., Cassanova, M., Bishop, C.E., Geldwerth, D., Andre, G., Fellous, M. & Weissenbach, J. (1984). Human XX males with Y single-copy fragments. *Nature* 307, 172-173.
- Gusella, J.F., Andersen, M.A., Tanzi, R.E., Watkins, P.C., Ottina, K., Wallace, M.R., Naylor, S.L., Sakaguchi, A.Y., Young, A.B., Shoulson, I., Bonilla, E. and Martin, J.B. (1983). A polymorphic DNA marker genetically linked to Huntington's disease *Nature* 306, 234
- Gusella J.F., Keys C., Vassanyi-Breiner A., Kao F-T, Jones C., Puck T.T. & Housman, D. (1980). Isolation and localisation of segments from different chromosomes. *Proc. Natl. Acad. Sci. USA* 77, 2829-2833.
- Hanahan, D. (1983). Studies on transformation of *Eschericia coli* with plasmids. *J. Mol. Biol.* 166, 557-580.
- Hofker, M.H., Skraastad, M.I., Bergen, A.A.B., Wapenaar, M.C., Bakker, E., Millington-Ward, A., Van Ommen, G.J.B. & Pearson, P.L. (1986). The X chromosome shows less genetic variation at restriction sites than the autosomes. *Am. J. Hum. Genet.* 39, 438-451.
- Hu, N. & Messing, J. (1982). The making of strand-specific M13 probes. *Gene* 17, 271-277.
- Jacobs, P.A. & Strong, J.A. (1959). A case of human intersexuality having a possible XXY sex determining mechanism. *Nature* 183, 302-303.
- Jarman, A.P., Nicholls, R.D., Weatherall, D.J., Clegg, J.B. & Higgs, D.R. (1986). Molecular characterisation of a hypervariable region downstream of the human α -globin gene cluster. *EMBO J.* 5, 1857-1863.
- Jeffreys, A.J. (1979). DNA sequence variants in the $\alpha\gamma$ -, $\alpha\gamma$ -, δ - and β -globin genes of man. *Cell* 18, 1-10.

- Jeffreys, A.J., Brookfield, J.F.Y. & Semeonoff, R. (1985a). Positive identification of an immigration test-case using human DNA fingerprints. *Nature* 317, 818-819.
- Jeffreys, A.J., Wilson, V., & Thein, S.L. (1985b). Hypervariable 'minisatellite' regions in human DNA. *Nature* 314, 67-73.
- Jeffreys, A.J., Wilson, V., Thein, S.L., Weatherall, D.J. & Ponder, B.A.J. (1986). DNA 'fingerprints' and segregation analysis of multiple markers in human pedigrees. *Am. J. Hum. Genet.* 39, 11-24.
- Jeffreys, A.J. & Morton, D.B. (1987a). DNA fingerprints of dogs and cats. *Anim. Genet.* 18, 1-15.
- Jeffreys, A.J., Wilson, V., Kelly, R., Taylor, B.A. & Bulfield, G. (1987b). Mouse DNA 'fingerprints': analysis of chromosome localisation and germ-line stability of hypervariable loci in recombinant inbred strains. *Nucleic Acids Res.* 15, 2823-2836.
- Jones, K.W. (1984). The evolution of sex chromosomes and their consequences for the evolutionary process. *Chromosomes Today* 8, 241-255.
- Kaslow, D.C & Migeon, B.R. (1987). DNA methylation stabilises X chromosome inactivation in eutherians but not in marsupials: Evidence for multistep maintenance of mammalian X dosage compensation. *Proc. Natl. Acad. Sci. USA* 84, 6210-6214.
- Keitges, E., Rivest, M., Siniscalco, M. & Gartler, S.M. (1985). X-linkage of steroid sulphatase in the mouse is evidence for a functional Y-linked allele. *Nature* 315, 226-227.
- Klebe, R.J, Chen, T. & Ruddle, F.H. (1970). Controlled production of proliferating somatic cell hybrids. *J. Cell Biol.* 45, 74-82.
- Koenig, M., Camerino, G., Heilig, R. & Mandel, J.-L. (1984). A DNA fragment from the X chromosome short arm which detects a partially homologous sequence on the Y chromosomes long arm. *Nucleic Acids Res.* 12, 4097-4109.
- Koenig, M., Moisan, J.P, Heilig, R. & Mandel J.-L. (1985). Homologies between the X and Y Chromosomes detected by DNA probes: localisation and evolution. *Nucleic Acids Res.* 13, 5485-5501.
- Krumlauf R., Jeanpierre M. and Young B.D. (1982). Construction and characterisation of genomic libraries from specific chromosomes. *Proc. Natl. Acad. Sci. USA* 79, 2971-2974.
- Kunkel L.M., Smith K.D. & Boyer, S.H. (1976). Human Y chromosome specific reiterated DNA. *Science* 191, 1189-1190.
- Kunkel L.M., Smith K.D. & Boyer, S.H. (1977). Analysis of human Y chromosome specific reiterated DNA in chromosome variants. *Proc. Natl. Acad. Sci. USA* 74, 1245-1249.

- Kunkel, L.M., Tantravahi, U, Kurnit, D.M., Eisenhard, M., Bruns, G.P. & Latt, S.A. (1983). Identification and isolation of transcribed human X chromosome DNA sequences, *Nucleic Acids Res.* 11, 7961-7979.
- Kunkel, L.M., Monaco, A.P., Middlesworth, W., Ochs, H.D. & Latt, S.A. (1985). Specific cloning of DNA fragments absent from the DNA of a male patient with an X chromosome deletion. *Proc. Natl. Acad. Sci. USA* 82, 4778-4782.
- Lamar, F.F. and Palmer, K (1984). Y encoded species specific DNA in mice : Evidence that the Y chromosome exists in two polymorphic forms in inbred strains. *Cell* 37, 171-177.
- Leach, D.R.F. & Stahl, F.W. (1983). Inviability of palindrome-bearing λ phage. *Nature* 305, 448-451.
- Lehninger, A.L. (1975) *Biochemistry* 2nd edn. Worth Publishers, Inc., New York. p861.
- Loenen, W.A.M. & Brammer, W.J. (1980). A bacteriophage lambda vector for cloning large DNA fragments made with several restriction enzymes. *Gene* 20, 249-259.
- Little, P. (1986). Restriction fragment length polymorphisms: Finding the defective gene. *Nature* 321, 558-559.
- van der Loos, C.M., van Breda, A.J., van der Berg, F.M., Walboomers, J.M.M. & Jobsis, A.C. (1984) *J. Inherited Metab. Dis.* 7, 97-103.
- Lucchesi, J.C. (1978). Gene dosage compensation and the evolution of sex chromosomes. *Science* 202, 711-716
- Lyon, M.F. (1972). X-chromosome inactivation and developmental patterns in mammals . *Biol. Rev.* 47, 1-35.
- Madan, K. (1976). Chromosomal measurements on a XXp⁺ male. *Hum. Genet.* 32, 141-142.
- Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G.K. & Efstatidis, A. (1978). The isolation of structural genes from libraires of eucaryotic DNA. *Cell* 15, 687-701.
- Maniatis, M., Fritsch, E.F. & Sambrook, J. (1982). *Molecular cloning: A laboratory manual*. Cold Spring Harbour Laboratory.
- Martin, G.R. (1982). X-chromosome inactivation in mammals. *Cell* 29, 721-724.
- Mattei, M.G., Mattei, J.F., Ayme, S. & Giraud, F. (1982). X-autosome translocations; cytogenetic characteristics and their consequence. *Hum. Genet.* 61, 295-309.
- McKusick, V.A. (1987). The human gene map (August 15, 1987)

- Mermer, B., Colb, M. & Krontiris, T.G. (1987). A family of short, interspersed repeats is associated with tandemly repetitive DNA in the human genome. *Proc. Natl. Acad. Sci. USA* 84, 3320-3324.
- Migeon, B.R., Shapiro, L.J., Noreum, R.A. Mohandas, T., Axelman, J., & Dabora, R.L. (1982). Differential expression of steroid sulphatase locus on active and inactive human X chromosomes. *Nature* 229, 838-840.
- Migone, N., Fedre, J., Cann, H., van West, B., Hwang, J., Takahashi, N., Honjo, T., Piazza, A. & Cavalli-Sforza, L.L. (1983). Multiple DNA fragment length polymorphisms associated with the immunoglobulin μ chain switch-like regions in man. *Proc. Natn. Acad. Sci.* 80, 467
- Mitchell, M.J., Florentin, L., Affara, N.A. (1987). Characterisation of the regions of close homology on Yp and Xq13-24. *Cytogenet. Cell Genet.* In Press.
- Mohandas, T., Shapiro, L.J., Sparkes, R.S., Sparkes, M.C. (1979). Regional assignment of the steroid sulfatase-X-linked ichthyosis locus: implications for a non-inactivated region on the short arm of the human X chromosome. *Proc. Natl. Acad. Sci. USA* 76, 5779-5783.
- Mohandas, T., Sparkes, R.S., Hellkuhl, B., Grzeschik, K.H. & Shapiro, L.J. (1980). Expression of an X-linked gene from an inactive human X chromosome in mouse-human hybrid cells: further evidence for the non-inactivation of the steroid sulphatase locus in man. *Proc. Natl. Acad. Sci. USA* 77, 6759-6773.
- Mohandas, T., Sparkes, R.S. & Shapiro, L.J. (1981). Reactivation of an inactive X chromosome: evidence for X-inactivation by DNA methylation. *Science* 211, 393-396.
- Mohandas, T., Sparkes, R.S. & Shapiro, L.J. (1982). Genetic evidence for the inactivation of a human autosomal locus attached to an X chromosome. *Am. J. Hum. Genet.* 34, 811-817.
- Mohandas, T., Geller, R.L., Yen, P.H., Rosendorff, J., Bernstein, R., Yoshida, A., & Shapiro, L.J. (1987). Cytogenetic and molecular studies on a recombinant human X chromosome: Implications for the spreading of X inactivation. *Proc. Natl. Acad. Sci. USA* 84, 4954-4958.
- Monaco, A.P. & Kunkel, L.M. (1987). A giant locus for the Duchenne and Becker muscular dystrophy gene. *Trends Genet.* 3, 33-37.
- Mondello, C., Ropers, H.-H., Craig, I.W., Tolley, E. & Goodfellow, P.N. (1987). Physical mapping of genes and sequences at the end of the human X chromosome short arm. *Ann. Hum. Genet.* 51, 137-143.
- Moore, G.E., Ivens, A., Chambers, J., Farrall, M., Williamson, R., Page, D.C., Bjornsson, A., Arnason, A. & Jensson, O.

(1987). Linkage of a X-chromosome cleft palate gene. *Nature* 326, 91-92.

Moses, M.J., Counce, S.J. & Paulson, D.F. (1975). Synoptonemal complex complement of man in spreads of spermatocytes, with details of the sex chromosome pair. *Science* 187, 363-365.

Müller C.R, Davies K.E. Cremer C., Rappold G., and Ropers H-H (1983). Cloning of genomic sequences from the human Y chromosome after purification by dual beam flow sorting. *Hum. Genet.* 64, 110-115.

Müller, U., Lalande, M., Donlon, T. & Latt, S.A. (1986a). Moderately repeated DNA sequences specific for the short arm of the human Y chromosome are present in XX males and reduced in number in an XY female. *Nucleic Acids Res.* 14, 1325-1341.

Müller U., Lalande M., Distèche C.M. & Latt S.A (1986b). Construction, analysis and application to 46,XY gonadal dysgenesis of a recombinant phage DNA library from flow-sorted human Y chromosomes. *Cytometry* 7, 418-424.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E. & White, R. (1987). Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* 235, 1616-1622.

Nakamura, Y., Leppert, M., O'Connell, P., Lalouel, J.M. & White, R. (1987b). Variable number of tandem repeat (VNTR) markers for human gene mapping. *Cytogenet. Cell Genet.* In Press.

Nicholls, R.D., Hill, A.V.S., Clegg, J.B. & Higgs, D.R. (1985). Direct cloning of specific DNA sequences in plasmid libraries following fragment enrichment. *Nucleic Acids Res.* 13, 7569-7578.

Nüssbaum, R.L., Lesko, J.G., Lewis, R.A., Ledbetter, S.A. & Ledbetter, D.H. (1987). Isolation of anonymous DNA sequences from within a submicroscopic X chromosomal deletion in a patient with choroideremia, deafness, and mental retardation. *Proc. Natl. Acad. Sci. USA* 84, 6521-6525.

Oberlé, I., Drayna, D., Camerino, G., White, R. & Mandel, J.-L. (1985). The telomeric region of the X-chromosome long arm: Presence of a highly polymorphic DNA marker and analysis of recombination frequency. *Proc. Natl. Acad. Sci. USA* 82, 2824

Ohno, S. (1967). Sex Chromosomes and Sex-linked Genes. *Springer Verlag, New York.*

Ohno, S. (1976). Major regulatory genes in sexual development. *Cell* 7, 315-321.

Page, D., de Martinville, B., Barker, D., Wyman, A., White, R., Francke, U. & Botstein, D. (1982). Single copy sequence hybridises to polymorphic and homologous loci on the human X and Y chromosomes. *Proc. Natl. Acad. Sci. USA* 74, 5352-5356.

- Page, D.C., Harper, M.E., Love, J. & Botstein, D. (1984). Occurrence of a transposition from the X chromosome long arm to the Y chromosome short arm during human evolution. *Nature* 311, 119-123.
- Page, D.C., de la Chapelle, A. & Weissenbach J. (1985). Chromosome Y-specific DNA in related XX males. *Nature* 315, 224-227.
- Page, D.C., Brown, L.G. & de la Chapelle, A. (1987). Exchange of terminal portions of X- and Y-chromosomal short arms in human XX males. *Nature* 328, 437-440.
- Passingham, R.E. (1982). *The Human Primate*. W.H. Freeman and Co., Oxford and San Francisco.
- Pearson, P.L. & Bobrow, M. (1970). Definitive evidence for the short arm of the Y chromosome associating with the X during meiosis in the human male. *Nature* 226, 959-961.
- Petit, C., de la Chapelle, A., Levilliers, J., Castillo, S., Noël, B. & Weissenbach, J. (1987). An abnormal terminal X-Y interchange accounts for most but not all cases of human XX maleness. *Cell* 49, 595-602.
- Polani, P.E. (1982). Pairing of X and Y chromosomes, non-inactivation of X-linked genes, and the maleness factor. *Hum. Genet.* 60, 207-211.
- Ponticelli, A.S., Schultz, D.W., Taylor, A.F. & Smith G.R. (1985). Chi-dependent strand cleavage by RecBC enzyme. *Cell* 41, 145-151.
- Poustka, A. & Lehrach, H. (1987). Jumping libraries and linking libraries: the next generation of molecular tools in mammalian genetics. *Trends Genet.* 2, 174-179.
- Proudfoot, N.J., Gil, A. & Maniatis, T. (1982). The structure of the human zeta-globin gene and a closely linked, nearly identical pseudogene. *Cell* 31, 553
- Race, R.R. & Sanger, R. (1975). *Blood groups in man*. 6th ed. Oxford, London: Blackwell.
- Ray, P.N., Belfall, B., Duff, C., Logan, C., Kean, V., Thompson, M.W., Sylvester, J.E., Gorski, J.L., Schmickel, R.D. & Worton, R.G. (1985). Cloning of the breakpoint of an X;21 translocation associated with Duchenne muscular dystrophy. *Nature* 318, 672-675.
- Reeders, S., Breuning, M.H., Davies, K.E., Nicholls, R.D., Jarman, A.P., Higgs, D.R., Pearson, P.L. & Weatherall, D.J. (1985) A highly polymorphic DNA marker linked to adult polycystic kidney disease on chromosome 16. *Nature* 317, 542-544.
- Riggs, A.D., Singer-Sam, J. & Keith, D.H. (1985). Methylation of the PGK promoter region and an enhancer way station model

for X-chromosome inactivation. *Prog. Clin. and Biochem. Res.* 198, 211-222.

Ropers, H.H & Wiberg, U. (1982). Evidence for X-linkage and non-inactivation of steroid sulphatase locus in wood lemming. *Nature* 296, 766-768.

Ropers, H.-H., Zimmer, J., Strobl, G. & Goodfellow P.N. (1985). The *MIC2X* (12E7) locus maps distally from *STS* on Xp. *Cytogenet. Cell Genet.* 40, 736.

Rouyer, F., Simmler, M.-C., Johnsson, C., Vergnaud, G., Cooke, H.J. & Weissenbach, J. (1986). A gradient of sex linkage in the pseudoautosomal region of the human sex chromosomes. *Nature* 319, 291-295.

Royle N.J., Clarkson, R., Wong, Z. & Jeffreys, A.J. (1987). Preferential localisation of hypervariable minisatellites near human telomeres. *Cytogenet. Cell Genet.* In press.

Schmeckpeper B.J., Smith K.D., Dorman B.P., Ruddle F.H. and Talbot C.C. (1979). Partial purification and characterisation of DNA from the human X chromosome. *Proc. Natl. Acad. Sci. USA* 76, 6525-6528.

Sealey, P.G., Whittaker, P.A. & Southern E.M. (1985). Removal of repeated sequences from hybridisation probes. *Nucleic Acid Res.* 13, 1905-1922.

Shapiro, L.J., Weiss, R., Webster, D., & France, J.T. (1978). X-linked ichthyosis due to steroid-sulphatase deficiency. *Lancet* i, 70-72.

Shapiro, L.J., Mohandas, T., Weiss, R. & Romeo, A. (1979). Non-inactivation of an X chromosome locus in man. *Science* 204, 1224-1226.

Shapiro, L.J. (1985). Steroid sulfatase deficiency and the genetics of the short arm of the human X chromosome. *Advances in Hum. Genet.* 14, 331-381.

Simmler, M.-C., Rouyer, F., Vergnaud, G., Nyström-Lahti, Ngo, K.Y., de la Chapelle, A. & Weissenbach J. (1985). Pseudoautosomal DNA sequences in the pairing region of the human sex chromosomes. *Nature* 317, 692-697.

Simmler, M.-C., Johnsson, C., Petit, C., Rouyer, F., Vergnaud, G. & Weissenbach, J. (1987). Two highly polymorphic minisatellites from the pseudoautosomal region of the human sex chromosomes. *EMBO J.* 6, 963-969.

Smith, T.J., Wilson, L., Kenwrick, S.J., Forrest, S.M., Speer, A., Coutelle, C. & Davies, K.E. (1987). Isoaltion of a conserved sequence deleted in Duchenne muscular dystrophy patients. *Nucleic Acids Res.* 15, 2167-2174.

Smith, G.R., Kunes, S.M., Schultz, D.W., Taylor, A. & Triman, K.L. (1981). Structure of Chi hotspots of generalized recombination. *Cell* 24, 429

Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98, 503-517.

Sutherland, G.R., Gedeon, A.K., Haan, E.A, Mulley, J.C. (1987). The gene for one form of X-linked mental retardation (*MRX2*) is closely linked to *DXYS1*. *Cytogenet. Cell Genet.* In Press.

Tabor, A., Andersen, O., Lundsteen, C., Niebuhr, E. & Sardemann, H. (1983). Interstitial deletion in the 'critical region' of the long arm of the X chromosome in a mentally retarded boy and his normal mother. *Hum. Genet.* 64, 196-199.

Taylor, A.F., Schults, D.W., Pontecelli, A.S. & Smith, G.R. (1985). RecBC enzyme nicking at *chi* sites during DNA unwinding: location and orientation dependence of the cutting. *Cell* 41, 153-163.

Tiepolo, L., Zuffardi, O., Fraccaro, M., di Natale, D., Gargantini, L., Müller, C.R., & Ropers, H.H. (1980). Assignment by deletion mapping of the steroid sulphatase X-linked ichthyosis locus to Xp22.3. *Hum. Genet.* 54, 205-206.

Toniolo, D., D'Urso, M., Martini, G. Persico, M., Tufano, V., Battistuzzi, G. & Luzzatto, L. (1984). Specific methylation pattern at the 3' end of the human housekeeping gene for glucose-6-phosphate dehydrogenase. *EMBO J.* 3, 1987-1995.

Tsui, L.C., Buckwald, M., Barker, D., Braman, J.C., Knowlton, R., Schumm, J.W., Eiberg, H., Mohr, J., Kennedy, D., Plavsic, N., Zsiga, M., Markiewicz, D., Akots, G., Brown, V., Helms, C., Gravius, T., Parker, C., Rediker, K., & Donis-Keller, H. (1985). Cystic fibrosis locus defined by a genetically linked polymorphic DNA marker. *Science* 230, 1054-1057.

Vergnaud, G., Page, D.C., Simmler, M.-C., Brown, L., Rouyer, F., Noël, B., Botstein, D., de la Chapelle, A. & Weissenbach, J. (1986). A deletion map of the human Y chromosome based on DNA hybridisation. *Am. J. Hum. Genet.* 38, 109-124.

Wainwright, B.J., Scambler, P.J., Schmitke, J., Watson, E.A., Law, H.-Y., Farrell, M., Cooke, H.J., Eiberg, H. & Williamson, R. (1985). Localization of cystic fibrosis locus to human chromosome 7cen-q22. *Nature* 318, 384-385.

Weatherall, D.J. (1982). *The New Genetics and Clinical Practice*. Nuffield Provincial Hospitals Trust, London.

Weiss, E.H., Golden, L., Fahner, K., Mellor, A.L., Devlin, J.J., Bullman, H., Tiddens, H., Bud, H., & Flavell, R.A. (1980). Organization and evolution of the class I gene family in the major histocompatibility complex of C57BL/10 mouse *Nature* 310, 650-655.

- Weissenbach, J., Levilliers, J., Petit, C., Rouyer, F. & Simmler, M.-C. (1987). Normal and abnormal interchanges between the human X and Y chromosomes. *Development* 101 (suppl.). In Press
- Wertman, K.F., Wyman, A.R. & Botstein, D. (1986). Host/vector interactions which affect the viability of recombinant phage lambda clones. *Gene* 49, 253-262.
- Wetton, J.H., Carter, R.E., Parkin, D.T. and Walters, D. (1987). Demographic study of a wild house sparrow population by DNA fingerprinting. *Nature* 327, 147-149
- White, R., Leppert, M., Bishop, D.T., Barker, D., Berkowitz, J., Brown, C., Callahan, P., Holm, T. & Jerominski, L. (1985). Construction of linkage maps with DNA markers for human chromosomes. *Nature* 313, 101-105.
- Wiberg, U.H. & Fredga, K. (1987) Steroid sulphatase levels in the root vole (*Microtus oeconomus*): Yet another rodent with an active Y-linked allele. *Cytogenet. Cell Genet.* In press
- Wolf, S.F. & Migeon, B.R. (1982). Studies of X chromosome DNA methylation in normal human cells. *Nature* 295, 667-671.
- Wolf, S.F., Jolly, D.J., Lunnen, K.D., Friedmann, T., & Migeon, B.R. (1984) Methylation of the hypoxanthine phosphoribosyl transferase locus on the human X chromosome: implications for X chromosome inactivation. *Proc. Natl. Acad. Sci. USA* 81, 2806-2810.
- Wolfe J., Erickson R.P., Rigby P.W.J. and Goodfellow P.N. (1984a). Cosmid clones derived from both the euchromatic and heterochromatic regions of the Y chromosome. *EMBO J.* 3, 1997-2003.
- Wolfe J., Erickson R.P., Rigby P.W.J. and Goodfellow P.N. (1984b). Regional localisation of 3 Y-derived sequences on the human X and Y chromosomes. *Ann. Hum. Genet.* 48 253-259.
- Wong, C., Dowling, C.E., Saiki, R.K., Higuchi, R.G., Erlich, H.A. & Kazazian, H.H. (1987). Characterization of β -thalassaemia mutations using direct genomic sequencing of amplified single copy DNA. *Nature* 330, 384-386.
- Wong, Z., Wilson, V., Jeffreys, A.J. & Thein, S.L. (1986). Cloning a selected fragment from a human DNA 'fingerprint': isolation of an extremely polymorphic minisatellite. *Nucleic Acids Res.* 14, 4605-4616.
- Worton, R.G., Duff, C., Sylvester, J.E., Schmickel, R.D. & Willard, H.F. (1984). Duchenne muscular dystrophy Involving translocation of the *dmd* gene next to ribosomal RNA genes. *Science* 224, 1447-1448.
- Wrischnik, L.A., Higuchi, R.G., Stoneking, M., Erlich, H.A., Arnheim, N. & Wilson, A.C. (1987). Length mutations in human

mitochondrial DNA: direct sequencing of enzymatically amplified DNA *Nucleic Acids Res.* 15, 529-542.

Wyman, A. & White, R. (1980). A highly polymorphic locus in human DNA. *Proc. Natl. Acad. Sci. USA* 77, 6754-6758.

Wyman, A.R., Wolfe, L. & Botstein, D. (1984). Isolation of the highly polymorphic locus D14S1 and other lethal/unstable human sequences in a new cloning host. *Am. J. Hum. Genet.* 36 (suppl.), 159s (abstr.).

Vassart, G., Georges, M., Monsieur, R., Brocas, H., Lequarre, A.S. & Christophe, D. (1986). A sequence in M13 phage detects hypervariable minisatellites in human and animal DNA. *Science* 235, 683-684.

Vogelstein, B., Fearon, E.R., Hamilton, S.R. & Feinberg, A.P. (1985). Use of restriction fragment length polymorphisms to determine the clonal origin of human tumours. *Science* 277, 642-644.

Yen, P.H., Patel, P.I., Chinault, A.C., Mohandas, T., & Shapiro, L.J. (1984). Differential methylation of HPRT genes on active and inactive X chromosomes. *Proc. Natl. Acad. Sci. USA* 81, 1759-1763.

Yen, P.H., Allen, E., Marsh, B., Mohandas, T., Wang, N., Taggart, R.T. & Shapiro, L.J. (1987a) Cloning and expression of steroid sulfatase cDNA and the frequent occurrence of deletions in STS deficiency: Implications for X-Y Interchange. *Cell* 49, 443-445.

Yen, P.H., Marsh, B., Allen, E., Mohandas, T. & Shapiro, L.J. (1987b) Organization of the human steroid sulfatase gene and sequence homology between the X and Y chromosomes. (Details not present in this abstract were reported at the 1987 annual meeting of the American Society of Human Genetics). *Am. J. Hum. Genet.*, 41 (suppl.), A248 (abstr.).