GC RICH ISLANDS ON THE HUMAN Y CHROMOSOME;
AN APPROACH TO CLONING Y-LINKED GENES

VIVIENNE PATRICIA SHORTLE

A thesis submitted in partial fulfilment
for the degree of Doctor of Philosophy
of the University of London
November 1989

Department of Genetics and Biometry
The Galton Laboratory
University College
University of London
Genetic analysis of the human Y chromosome has been slow despite its central role in sex determination. At present genetic analysis has identified only about ten genes on the Y chromosome, and only three genes (one coding for a cell surface antigen, a second coding for a regulatory zinc finger protein and a third coding for a testis specific transcript) have been cloned. This study has attempted to identify further genes on the Y chromosome using techniques which also facilitate cloning. This thesis therefore describes an investigation into the GC rich regions (HTF - Hpall Tiny Fragment Islands) of the Y chromosome as an approach to cloning Y-linked genes.

Using restriction analysis, 150 Y-specific cosmid clones have been screened. Four putative HTF island containing clones were selected and studied in more detail using Southern mapping, northern analysis, library screening, DNA sequencing and methylation analysis. Although data indicate that three of the clones are not associated with transcribed sequences, the fourth clone has been shown through sequence analysis, to contain a possible open reading frame.

A direct approach to cloning Hpall tiny fragments from a Y chromosome specific library, obtained from the American Type Culture Collection, is also described. Following the isolation and purification of the human inserts away from the phage vector within this library, the inserts were further restricted using the enzyme Hpall and the resulting tiny fragments re-cloned and analysed for their possible association with transcribed sequences. Although a clone was identified which hybridizes to human genomic sequences, data
indicates this to be of autosomal origin. Evidence is also discussed that the human Y chromosome appears to be relatively deficient of HTF islands when compared with the remainder of the genome, and that this deficiency could also reflect the small number of genes which are carried on the human Y chromosome.
ACKNOWLEDGEMENTS

I am indebted to my supervisor, Dr. Jonathan Wolfe, for his unfailing support, enthusiasm and inspiration. I am also grateful for the support and companionship of all past and present members of the group who have made the past four years so enjoyable. In particular, I would like to thank Kay, Zanna, Paul and Steve for being such entertaining company. I would also like to thank all the members of the Department and the MRC Human Biochemical Genetic Unit for their help and friendship throughout my stay at the Galton Laboratory.

I am also grateful to Georges for his constant support and sound advice throughout the years I have known him. I would also like to thank my parents for their encouragement, and Michael for his patience and unreserved support.
CHAPTER 1. INTRODUCTION

1.1. The Y chromosome 13

1.2. The classical model 16

1.2.1. The short arm pairing region 16

1.2.2. The sex determining region 25

   a) Genes involved in primary sex determination 25

   b) Short arm Y - long arm X homology 37

1.2.3. Centromeric region 39

1.2.4. The long arm euchromatic region 41

   a) Sequences involved in spermatogenesis 41

   b) Steroid sulphatase 45

   c) Amelogenin sequences 47
d) Other sequences
1.2.5. The long arm heterochromatic region
1.3. Background and aims of this study
1.3.1. CpG or HTF islands

CHAPTER 2. METHODS AND MATERIALS
2.1. General techniques and reagents
2.2. Bacterial competence and transformation
2.3. Repackaging cosmid DNA
  2.3.1. Testing E. coli strain ED8767 for recA− function
  2.3.2. Packaging reaction
2.4. DNA isolation
  2.4.1. Preparation of cosmid and plasmid DNA
  2.4.2. Preparation of bacteriophage λ DNA
  2.4.3. Preparation of DNA from genomic and hybrid cell lines
  2.4.4. Preparation of DNA from flow sorted chromosomes
2.5. DNA analysis
  2.5.1. Restriction enzyme digestion
  2.5.2. Restriction mapping by partial digestion
  2.5.3. Gel electrophoresis
  2.5.4. Isolation of fragments for subcloning and radioactive labelling
  2.5.5. Subcloning DNA fragments
2.6. Radioactive labelling of nucleic acid
2.7. Southern blot analysis
2.8. Construction of chromosome-specific HpaII libraries
2.8.1. Library construction using DNA from flow sorted chromosomes

2.8.2. Library construction using DNA from a chromosome specific λ library

2.9. Screening bacteriophage λ libraries

2.10. Northern blot analysis

2.11. Nucleotide sequence analysis

2.12. Materials

CHAPTER 3. RESULTS: CpG islands within Homer V cosmid clones

3.1. Identification of CpG rich cosmid clones

3.2. Analysis of CpG rich clones

   3.2.1. Restriction analysis

   3.2.2. Southern blot analysis

      a) Chromosomal localisation

      b) Quantitative analysis

   3.2.3. Sequence analysis

   3.2.4. Methylation studies

CHAPTER 4. DISCUSSION: CpG islands within Homer V cosmid clones

CHAPTER 5. RESULTS: Cloning Hpal tiny fragments from chromosome specific libraries

   5.1. Cloning fragments from flow sorted chromosomes

       5.1.1. Analysis of clones

   5.2. CloningMspI fragments from a chromosome-specific λ library

       5.2.1. Analysis of clones
CHAPTER 6. DISCUSSION: Cloning *HpaII* tiny fragments from chromosome specific libraries

CHAPTER 7. SUMMARY

REFERENCES
FIGURES

1.1. The classical model of the human Y chromosome 14
1.2. Gradient of sex linkage within the pseudoautosomal region 23
2.1. Restriction map of the vector Bluescript 87
3.1. MspI digests of 20 Y-linked Homer V cosmid clones 100
3.2. Distribution of non-vector MspI fragments within cosmid clones 101
3.3. Restriction enzyme maps of cY019, cY022, cY24 and cY113 103
3.4. 'CG'-restriction enzyme analysis of cY019, cY022, cY24 and cY113 105
3.5. 'CG'-restriction enzyme sites within p019P3 106
3.6. Southern analysis using cY113 as a probe 109
3.7. Southern analysis using cY24 as a probe 109
3.8. Southern analysis using cY022 as a probe 110
3.9. Southern analysis of mouse DNAs using cY022 as a probe 111
3.10. Southern analysis of human, hybrid and mouse DNAs to localise cY022 subclones
   a) Using p022E1 as a probe
   b) Using p022E4 as a probe
3.11. Southern analysis using cY019 as a probe 115
3.12. Southern analysis of mouse DNAs using cY019 as a probe 116
3.13. Southern analysis of TaqI digested genomic DNAs probed with p019P3 117
3.14(1) Northern analysis (1) using p019P3 as a probe 119
3.14(2) Northern analysis (2) using p019P3 as a probe 119a
3.15. Screening the λgt11 Fetal Testis library with p019P
3.16. Southern analysis of hybrid cell lines to localise p019P3
3.17. In situ hybridization analysis using p019P3 as a probe
3.18. Quantitative blot analysis to estimate p01P3 copy number
3.19. Nucleotide and derived amino acid sequences from p019P3
3.20. G+C content and CpG/GpC incidence across a 1.115kb region of p019P3
3.21. Southern analysis of Hpall and Mspl digested genomic DNAs using p019P3 as a probe
3.22. Restriction map of the sequenced region
3.23. Southern analysis of 'CG'-enzyme digested genomic DNAs using p019P3 as a probe
5.1. Flow karyotype of OXEN cell line
5.2. Pvull restriction endonuclease analysis of random library clones
5.3. In vitro transcription of clones Mspl inserts generating labelled RNA probes
5.4. Southern blot analysis of various plasmid digests using a labelled in vitro transcript as a probe
5.5. Colony screen and EcoRI restriction endonuclease analysis of clones derived from Mspl library
5.6. Ethidium bromide stained agarose gels showing HindIII cleaved Charon 21A library DNA
5.7. Pvull restriction endonuclease analysis of random library clones
5.8. Southern blot analysis of genomic and hybrid cell line DNAs using two MspI inserts from the Y chromosome library as probes.

5.9. Southern blot analysis of genomic and hybrid cell line DNAs using an anonymous MspI insert from the chromosome 9 library as a probe.

NOTE: For coherence, some figure legends have been typed on the backs of pages, facing figures on opposite pages.
ABBREVIATIONS

bp base pairs

cpm counts per minute

DNase deoxyribonuclease

EDTA diaminoethane tetraacetic acid, disodium salt

GAPDH Glyceraldehyde 3 phosphate dehydrogenase

G6PD Glucose 6 phosphate dehydrogenase

HPRT hypoxanthine phosphoribosyl transferase

kb kilobase pairs

mb megabase pairs

mins minutes

MIS Mullerian inhibiting substance

NA nucleic acid

OD optical density

PFU plaque forming units

PGK phosphoglycerate kinase

RNase ribonuclease

RNaseIn ribonuclease inhibitor

s seconds

SDS sodium dodecyl sulphate

SSC sodium saline citrate

TDF Testis determining factor

UV ultra violet

vol volume
CHAPTER 1

INTRODUCTION

1.1 THE Y CHROMOSOME

It has been shown that in many though not all vertebrate and invertebrate species, the sex of an individual is determined by the individual's chromosomal constitution. Indeed sexual dimorphism in insects was among the first traits shown to have a chromosomal basis (Wilson 1905).

In humans, the existence of X and Y chromosomes was initially demonstrated cytologically by Wilson in 1911. Karyotypic analysis shows that a normal male has a Y chromosome and only one X chromosome, whereas a normal female lacks a Y chromosome but possess two X's. Until 1959, however, it was not known whether the apparent phenotypic differences between the sexes was due to X chromosome dosage or the presence of a Y chromosome. Studies then revealed the existence of XO female mice (Welshons & Russell 1959) and a 45,X human subject who was female (Ford et al. 1959), indicating the association between the Y chromosome and maleness. Subsequently, XXY male mice were found (Russell & Chu 1961) and humans with as many as four X chromosomes and a Y were clearly shown to be males with unequivocally testicular gonadal histology (Atkins et al. 1963). It is now believed that one or more dominant genes on the Y chromosome control sex determination.
FIG. 1.1 The Classical Model of the Human Y Chromosome
by initiating an orderly cascade process, propagated through genetic and hormonal controls, inducing the bipotential foetal gonad to differentiate as a testis, even when multiple copies of the X chromosome are present (Jacobs & Strong 1959, Russell & Chu 1961).

In man, the Y is one of the smallest chromosomes comprising 1-2% of the total length of the haploid genome (Paris Conference 1971) and therefore contains approximately 3-6 x 10^7 bp of DNA. It consists of a minute short arm (Yp) with no associated satellite and a long arm (Yq). 30-50% of the chromosome is made up of heterochromatin located at the distal end of Yq, whilst the remaining portion (approximately 2 x 10^7 bp) of the Y constitutes the euchromatic region covering the entire short arm and the proximal region of Yq (Ypter-Yq11).

According to classical cytogenetic models, using karyotypic-phenotypic correlations to assign male specific genes to certain regions, the human Y chromosome can be further sub-divided into five broad areas (Fig. 1.1):

a) a short arm X-Y homologous pairing region

b) a sex determining region on Yp

c) a centromeric region

d) a long arm euchromatic region associated with spermatogenesis

e) the long arm heterochromatic region.
1.2 The Classical Model

1.2.1 The short arm pairing region

The chromosomal determination of sex requires that the X and Y chromosomes migrate to opposite poles during male meiosis and are incorporated into different gametes. Correct segregation of chromosomes at meiosis appears to be a function of pairing and chiasmata formation, and telomeric associations between the short arms of the X and Y have been documented using light and electron microscopy (Chen & Falek 1971; Chandley et al. 1984), where pairing has been shown to involve the terminal 30% of Xp and most of Yp (Solari 1980). In addition, it was predicted that such pairing would be a consequence of sequence homology and that recombination would occur between these sequences (Koller & Darlington 1934). Consequently such sequences would fail to show classical sex linkage but would simulate autosomal inheritance. The term 'pseudosoma1' was therefore introduced to describe their predicted genetic behaviour (Burgoyne 1982).

In addition to the observations of XY pairing, DNA sequence homologies between the mammalian X and Y chromosomes have been predicted on the basis that a) these chromosomes have a common ancestral origin, b) there must be genes on the Y chromosome that are homologous to a group of genes on the X chromosome, and that these genes must function in double dose in the somatic cells of XX and XY individuals. The presence of these genes in only a single dose in XO individuals contributes to the somatic defects of Turner's syndrome (Burgoyne 1982), and c) there appears to be a lack of X-inactivation of X-linked loci, such as Xg (Race & Sanger 1975) and steroid
sulphatase $STS$ (Mohandas et al. 1979), which map to the pairing region of Xp. Polani (1982) therefore suggested that the X and Y chromosomes carry a homologous segment of DNA at Xpter and Ypter and that during the zygotene stage of meiosis an obligatory recombination occurs between one chromatid of the X and Y chromosomes. More recently the results from several genetic and molecular experiments have supported the existence of a pseudoautosomal region.

Studies of the sex reversing factor in mice, $Sxr$, provided the first indication for the existence of the mammalian pseudoautosomal region. This dominant mutation, first described by Cattanach et al. in 1971, causes XX and XO mice to develop as phenotypically normal, though sterile, males. Since carrier males transmit $Sxr$ equally to XX and XY offspring, it was initially assumed that $Sxr$ was located on an autosome although extensive linkage analysis failed to map the locus. Subsequently, however, it was found that XY carriers have a high incidence of X-Y pairing failure (Winsor et al. 1978) indicating a change within the pairing region where pseudoautosomal genes are located. In 1982 Singh and Jones used the Y-specific Bkm satellite sequence as a probe to show that XX$Sxr$ male mice did carry a small proportion of the Y in their genomes and additional in situ hybridization data confirmed that Bkm hybridized to the pericentric region of the normal Y but to both the pericentric and telomeric regions of the Y in XYSxr carrier males. $Sxr$ is therefore viewed as a duplication of the Y-linked testis determining factor which has been translocated to the distal tip of the Y. During meiosis in the male $Sxr$ carriers there is an obligatory recombination event which transfers the telomeric copy of the sex determining region to the X chromosome. Thus, male carriers produce equal numbers of four types
of gametes containing different sex chromosomes: Y, YSxr, X and XSxr. The offspring which receive the YSxr chromosome will be the new male carriers; the offspring which receive the XSxr chromosome will be sex reversed females.

In man, the genetics of the MIC2 locus are indicative of a pseudoautosomal region. MIC2 encodes a cell surface antigen which is recognised by the 12E7 monoclonal antibody (Goodfellow et al. 1980). The antibody was shown to react with a cell surface antigen expressed on all human cells but not with the surface of rodent cells (summarized in Goodfellow 1983). Exploiting this species-specificity, Goodfellow et al. (1980) used somatic cell hybrids to assign the gene controlling 12E7 expression to the short arm of the X chromosome and to localize the gene to the Xp22.32-Xpter region (Goodfellow et al. 1983, Curry et al. 1984). During further studies, cell lines which retained the human Y chromosome, but not the human X chromosome, were also found to express the antigen. Subsequently this Y-linked gene was localized to the euchromatic region, Ypter-Yq11 (Goodfellow et al. 1983).

The assignment of both X and Y linked genes to the postulated sex chromosome pairing regions led to the suggestion that MIC2 is pseudoautosomal, and biochemical analysis of the MIC2 gene products, which showed that the X and Y derived forms do not differ in either size or charge (Banting et al. 1985) supported this view. Subsequently a cloned cDNA probe for MIC2, pSG1, was isolated (Darling et al. 1986) and shown by in situ hybridization to recognize two loci in the human genome, one in the region Xp22.32-Xpter and the other at Yp11.2-Ypter (Buckle et al. 1985a), whilst Southern blot analysis using pSG1 as a probe failed to detect specific X and Y chromosome differences between
the two genes. A result consistent with MIC2 being pseudoautosomal. Finally, conclusive proof for MIC2 being pseudoautosomal was obtained from family studies using cDNA and genomic probes which recognise RFLPs, demonstrating that MIC2 recombines between the X and Y chromosomes during male meiosis with a frequency of 2% (Goodfellow et al. 1986).

Two other expressed genes map to the tip of the short arm of the human X chromosome; the locus encoding steroid sulphatase STS (termed Sts in the mouse), and the locus encoding the XGa red blood cell antigen, XG (Mohandas et al. 1979, Tiepolo et al. 1980, Ferguson-Smith et al. 1982, Curry et al. 1984). MIC2, XG and STS all escape X-inactivation (Ducos et al. 1971, Race and Sanger 1975, Shapiro et al. 1979, Mohandas et al. 1980, Goodfellow et al. 1984). It might be expected therefore, that genes which escape X-inactivation would have functional Y-linked homologues, thereby maintaining male/female dosage for such loci. However, unlike MIC2, there is no evidence for the existence of functional Y-linked genes for either STS (Craig and Tolley, 1986) or XG (Race and Sanger, 1975) in man. In the mouse Sts is pseudoautosomal. Keitges et al. (1985) demonstrated that deficiency of Sts could be transmitted to XO offspring via the X chromosome. To account for the apparent lack of sex linkage, functional Sts genes were proposed to exist on the mouse X and Y chromosomes, within the pseudoautosomal region, which undergo obligatory recombination during meiosis. In humans however, the functional STS gene maps proximal to MIC2 on the X chromosome (Geller et al. 1986, Ropers et al. 1985) whilst a Y-linked STS pseudogene has been characterized and assigned to the long arm of the human Y chromosome (Fraser et al. 1987, Yen et al. 1988). Data in humans and
other primates therefore suggests that a recent pericentric inversion of the Y chromosome during primate evolution (Yen et al. 1988) has disrupted the former pseudoautosomal arrangement of these genes. Incomplete X-inactivation at STS (Migeon et al. 1982) might therefore represent an intermediate stage in acquiring full dosage compensation at this locus.

The discovery that MIC2 is located within close proximity of XG in Xp22.3 and that both genes encode cell surface antigens and escape X-inactivation (Ferguson-Smith et al. 1982, Race and Sanger 1975) raised the question of whether these loci are identical. As MIC2 is pseudoautosomal and XG is strictly X-linked, this is now known not to be the case. In addition an example of recombination between XG and MIC2 suggests that these loci are independent (Goodfellow et al. 1987). However a complex interaction between the expression of MIC2 and XG on red blood cells has been discovered (Goodfellow and Tippett 1981). Classically, XG has been described as the X-linked structural locus for the red blood cell antigen XGa. This locus is polymorphic with two alleles XGa and XG. The XGa-antigen is defined by reaction with XG antisera. XG is the reciprocal "null" allele. The XG locus is located close to MIC2 in Xp22.3 and escapes X-inactivation (Race & Sanger 1975). Extensive family and population studies have ruled out the possibility of a Y-located form of XG. Similarly 12E7 antigen expression on red blood cells is polymorphic; two phenotypes can be distinguished: low level expression and high level expression. In quantitative assays, red cells with the low level antigen phenotype bind approximately one third the amounts of 12E7 antibody as red cells from high level antigen expressors.
In females the XG antigen phenotype predicts the 12E7 antigen phenotype: XG antigen positive females (XGa/XGa; XGa/XG) are 12E7 antigen high level expressors; XG antigen negative females (XG/XG) are low level 12E7 antigen expressors. In males further complexity is found: all XG antigen positive males are 12E7 antigen high level expressors; however, XG antigen negative males can either be low level or high level 12E7 antigen expressors. To explain this sex-limited behaviour, Goodfellow and Tippett (1981) proposed the existence of a polymorphic Y-located gene, YG, which had the property of regulating 12E7 antigen expression on red blood cells. In a fashion analogous to XG, YG was proposed to be polymorphic: the YGa allele would cause high level 12E7 expression; theYG allele would cause low level expression. It was predicted that XGa-antigen-negative brothers would all share the same 12E7 phenotype and family studies strongly support the YG hypothesis (Tippett et al. 1984), however, of several hundred individuals examined, seven did not fit the expectation of a Y-located locus regulating MIC2 gene expression on red blood cells (Tippett et al. 1986). One case, in which a normal male with the apparent genotype XG/YG had failed to inherit his father's YGa allele (XG/YGa), was investigated in more detail (Goodfellow et al. 1987). DNA analysis showed that he is a recombinant at MIC2. This exchange at MIC2 could have destroyed YG activity, however Goodfellow et al. (1987) proposed a new model in which a cis-acting regulator controls both XGa-antigen expression and 12E7-antigen expression on red blood cells.

Three closely linked genetic elements are postulated to exist:

1) XGS a structural locus encoding the XGa-antigen
2) XGR a structural locus which controls in cis the expression of the MIC2 and XGS loci.
3) **MIC2** the structural gene which encodes the 12E7-antigen.

MIC2 and XGR are pseudoautosomal and XGS is strictly X-linked. XGR is polymorphic with two alleles, A and B. In cis the A allele induces XGa-antigen expression from the XGS locus and high level 12E7 antigen expression from the MIC2 locus. In cis the B locus results in failure to express XGa-antigen from the XGS locus and in low level 12E7 antigen expression. Formal proof of this model requires the ascertainment of other families exhibiting abnormal XGa/12E7 inheritance.

Conclusive evidence of the existence of the pseudoautosomal region was obtained from family studies using cloned probes for random DNA sequences. Besides MIC2, the human X and Y chromosomes are known to share five other homologous DNA loci (Cooke et al.1985; Simmler et al.1985; Affara et al.1986a; Rouyer et al.1986a; Page et al.1987a). This homology extends up to the telomere (Cooke et al.1985) but is restricted to the terminal part of the pairing region (Simmler et al.1985, Vergnaud et al.1986) observed at male meiosis between the short arm of the Y chromosome and the distal short arm of the X chromosome (Chandley et al.1984). DNA probes detecting RFLPs at these loci have been isolated and used in family studies to test for sex linkage. Most of the loci recombine with sexual phenotype (Cooke et al.1985; Simmler et al.1985; Rouyer et al.1986a,b; Goodfellow et al.1986; Page et al.1987a) giving experimental support to the concept of pseudoautosomal loci proposed by Burgoyne (1982).

The segregation of four pseudoautosomal DNA loci has been followed through family analysis of about 100 male and female meioses (Rouyer et al.1986b; Goodfellow et al.1986; Wiessenbach et al.1987).
FIG 1.2 The Pseudoautosomal Region: A map derived from the recombination frequencies of four pseudoautosomal loci with respect to sexual phenotype. This map is a summary of data in Weissenbach et al. (1987) and Goodfellow et al. (1986).
Such a linkage analysis first showed that the different loci analysed recombine with sex at different frequencies according to a gradient of sex linkage. This gradient can be represented on a map, where the four loci are ordered with respect to their recombination distances with the X and Y sex-specific chromosomal blocks (Fig.1.2.). Thus the telomeric locus DXYS14 recombines with the proposed testis-determining factor at a frequency of almost 50% and is thus not sex linked, whereas exchange at the MIC2 locus is relatively rare (<5% recombination), making MIC2 one of the most proximal markers (Goodfellow et al.1986). Three-point analysis of these different loci has also shown that loci recombining less frequently never segregate independently from loci recombining more frequently (Rouyer et al.1986a,b). This indicates that X-Y interchange of pseudoautosomal loci results from a crossing over and not from other genetic exchange events, such as gene conversion. These results also strongly suggest that a single crossover event occurs between the X and Y chromosomes since no examples of double recombination have yet been found. In addition, since different pseudoautosomal loci do not recombine at the same frequency, the exchange events must be scattered throughout the pseudoautosomal region.

Several indirect arguments, based on the number of pseudoautosomal sequences isolated from genomic libraries, suggest that the size of the pseudoautosomal region is approximately $3 \times 10^3$ Kbp (Rouyer et al.1986b) This estimation is consistent with deletion analysis (Mondello et al.1987) and with physical maps of the pseudoautosomal region obtained from pulsed field gel electrophoresis experiments (Brown 1988; Petit et al.1988; Rappold & Lehrach 1988; Pritchard et al.1987). This implies that male recombination rates in
the pseudoautosomal region are approximately ten times higher than elsewhere in the genome. In contrast female recombination rates in the pseudoautosomal region are close to expected values. The relatively small physical size of the human pseudoautosomal region may explain the absence of double recombination events within this region (Rouyer et al. 1986b; Page et al. 1987a).

1.2.2 The sex determining region

a) Genes involved in primary sex determination

There must exist on the Y chromosome one or more genes whose products are responsible for the induction of testis from the indifferent foetal gonad, with secondary sexual characteristics a direct consequence of this primary event. In man, this signal has been termed the Testis Determining Factor (TDF) (Miller et al. 1983), whilst in the mouse the equivalent gene (or genes) is termed the Testis Determining-Y or Tdy. Until recently, nothing was known about the link between the Y chromosome and the development of a testis, either at the molecular level or in terms of a developmental process. Recently, however, a candidate gene for TDF has been cloned (Page et al. 1987b), termed ZFY (Page 1988) as it encodes a zinc-finger protein. Although its mode of action is not known a possible role in transcriptional control has been suggested (see below).

The early development of the gonad proceeds identically in both males and females. In both sexes the primordial germ cells are
observed to migrate from their site of origin and colonise the gonadal ridges on the median surface of the mesonephros. The mesonephros functions as an embryonic kidney and has its own duct, the Wolffian duct. Later in development the embryo acquires the metanephros, and then the mesonephros and its duct become obsolete. However, this duct has the potential to develop into the male reproductive tract, including vas deferens, seminal vesicle and epididymis. In addition, the embryo also possesses a paramesonephric, or Müllerian duct, which has the potential to develop into the female reproductive tract, including Fallopian tube (oviduct), uterus and upper part of the vagina. The embryo therefore, has the forerunners of both the male and female reproductive tracts of which one is destined to develop and the other to regress.

The Y-chromosomal determination of gonadal sex manifests itself only when colonization of the gonadal ridges is completed during the sixth week in the human embryo and the second week in the mouse. As the indifferent gonad develops into either a testis or an ovary, the first cell lineage to show sex-specific differentiation is the supporting cell lineage that surrounds the germ cells, giving rise to Sertoli cells in the male, and follicle cells in the female. Other cell types in the testis such as the cells of the tunica albuginea, and the Leydig cells that produce testosterone, differentiate later, and may be dependent upon prior Sertoli cell differentiation. Since the Sertoli cells are the first testicular cell type to be distinguished morphologically (Jost et al. 1981) it has been suggested that differentiation of these cells is dependent upon expression of the testis determining gene.
The foetal testis performs two early functions. The first is to produce testosterone, which is excreted by the developing testicular Leydig cells and binds to an androgen receptor found in target cells stimulating the development of the male reproductive tract. The second function is to cause regression of the Müllerian duct system, which it does with the production of a Müllerian inhibiting substance (MIS) produced by the foetal Sertoli cells (Vigier et al. 1987). MIS has been purified to homogeneity and demonstrated to have a cytotoxic effect on human ovarian tumour cells in vitro and in vivo (Donahoe et al. 1979 and 1981). In addition the bovine and human MIS genes have been cloned (Cate et al. 1986; Picard et al. 1986) enabling the exact role of MIS in male embryonic development to be elucidated.

Initially the primary signal for male development, represented by TDF, was thought to be identical to the male-specific histocompatibility antigen H-Y (Watchel et al. 1975) known to be controlled by a gene or genes on the Y chromosome. The H-Y antigen was originally defined immunologically by skin grafting; Eichwald and Silmser (1955) showed that in certain inbred strains, female mice can reject skin grafts from otherwise identical male mice. However, as transplantation assays are cumbersome, two other assays have also been developed. A serological assay which uses male specific antibodies isolated from the serum of female mice grafted with male skin has been described (Goldburg et al. 1972). However the validity of this test is questionable as it now appears that the antibodies are not directed at the same determinant as that recognized by H-Y specific cloned T cells (Silvers et al. 1982; Simpson et al. 1982). H-Y antigen can also be detected by an in vitro killing assay which employs female cytotoxic T-lymphocytes generated in MLC (Gordon et al. 1975). In contrast with
the serological tests, results obtained with the T cell killing assay are more reproducible and correlate closely with the H-Y typing obtained from the transplantation experiments (Andrews 1984).

Two experiments, one in mouse and the other in man, prove conclusively that the H-Y antigen is not the testis determining factor. In mice, XXSxr males, like XY males are positive for the male-specific antigen H-Y as defined by transplantation (Bennett et al. 1977), T cell killing and serological studies (Simpson et al. 1981) indicating that the gene controlling the expression of this antigen (Hya) is contained within the same region as Tdy, a finding predicted from Wachel's hypothesis. In normal XX individuals, only one X is active and X inactivation is random (Lyon 1961). In mice carrying one X chromosome bearing Sxr, although X inactivation remains random, inactivation may not spread to the translocated Sxr gene in all cells in which the Sxr-bearing X chromosome is inactivated. Thus XXSxr mice are mosaics with more than 50% of their cells expressing H-Y. Subsequently McLaren and Monk (1982) and Cattanach et al. (1982) mated female mice carrying the balanced translocation T(16;X)16H (T16H) (Lyon et al. 1964) with Sxr carrier males to produce progeny which although possessing Tdy developed as females because the paternally derived XSxr chromosome was invariably inactivated. Following H-Y typing however, these mice were found to be H-Y positive (presumably because Hya is not inactivated with Tdy) suggesting that the presence of H-Y antigen is not sufficient condition for testis development (McLaren et al. 1984). In addition, one female mouse was found to be H-Y negative (McLaren et al. 1984). This individual was then mated with a normal male. All H-Y tested descendants of this female who inherited her mate's normal Y chromosome were found, as expected, to be positive for H-Y antigen;
however, those who received her $Sxr$ region in the absence of a $Y$ chromosome, that is $X/Sxr$ males and $T16H/XSxr$ individuals were all $H-Y$ negative. These results implied that these mice carried a variant form of the $Sxr$ region (termed $Sxr'$) in which $Tdy$ is retained but the sequences controlling the expression of $H-Y$ antigen have been lost, and that the presence of $H-Y$ antigen is not a necessary condition for testis development. The transition from $Sxr$ to $Sxr'$ is believed to involve a small deletion (Roberts et al. 1988) brought about perhaps by unequal crossing over.

Similarly, Simpson et al. (1987a) separated the genetic loci for testis determination and the $H-Y$ antigen in man. $H-Y$ typing using cytotoxic T cell assays was carried out on a series of sex-reversed humans (XX males and XY females), each shown by DNA hybridization to carry part but not all of the $Y$ chromosome. This deletion analysis maps the gene for $H-Y$ to the long arm or centromeric region of the human $Y$ chromosome, a locus clearly distinct from the $TDF$ locus which is now known to map to the distal portion of $Yp$ (Simpson et al. 1987b) (see below).

Until recently very little was known about the nature or possible mode of action of the $TDF$ gene product so that cloning strategies based on expression were limited. However, in common with several genetic disorders which were characterized by a phenotypic expression of a disease for which an aberrant genetic product had not been identified eg. Duchenne muscular dystrophy and retinoblastoma cloning could be accomplished from knowledge of the chromosomal position. Through karyotypic analysis of $Y$ chromosome structural abnormalities, $TDF$ has been localised to the pericentric region of the short arm (Davies 1981; Magenis et al. 1984). Subsequently, through molecular
studies involving sex-reversed males (46XX karyotype) and sex-reversed females (46XY karyotype) a small portion of the Y chromosome has been identified which appears to be sufficient to induce testicular development.

Human XX maleness, first described by de la Chapelle (1981), is the best known example of error in primary sex determination. The syndrome occurs with a frequency of 1:20,000 newborn boys and although individuals have a 46,XX karyotype, phenotypically they are sterile males. The hypotheses put forward to account for XX maleness fall into three groups 1) XX maleness is due to a non Y related mechanism triggered by an autosomal or X-linked mutation (de la Chapelle et al. 1977), 2) there is an undetected Y chromosome mosaicism, 3) the male phenotype is the result of the expression of some cytologically undeletable Y-chromosome material. The first hypothesis was suggested to explain rare familial cases of XX males, however the latter hypothesis is now believed to account for the occurrence of most sporadic cases of XX males, and is thought to arise due to an unequal X-Y exchange during paternal meiosis which results in the transfer of a Y chromosome fragment harbouring TDF, to the X chromosome short arm by 'accidental crossing-over', with the simultaneous loss of Xp sequences through reciprocal exchange (Ferguson-Smith 1966). This hypothesis parallels the mouse model in which the Sxr locus is transferred from the Y to the X chromosome by what appears to be an obligate chiasmata (Evans et al. 1982). Females receiving an X chromosome carrying the Sxr locus are sex-reversed phenotypic males unless the Sxr locus is inactivated (Cattenach et al. 1982; McLaren & Monk 1982). Molecular studies using Y-specific DNA probes have now shown Y-chromosomal DNA sequences in most XX males (Guellaen et al.)
1984; Koenig et al.1985; Page et al.1985; Müller et al.1986a and b; Affara et al.1986b; Vergnaud et al.1986). The percentage of Y DNA positive XX males reported so far ranges between 60% (Vergnaud et al.1986) and 80% (Müller et al.1987). Those XX males which fail to react with Y derived probes may have either smaller, undetectable, translocations of Y chromosome material or may be caused by mutations elsewhere in the genome. Other results provide even more direct support of the interchange model. In many families, the XX male does not express his father's XG allele (see de la Chapelle 1981), and in one case this loss of paternal XG expression was associated with the acquisition of the Y-linked allele for MIC2 (de la Chapelle et al.1984). In addition, in situ hybridization studies of Y-DNA positive XX males (Magenis et al.1984; Buckle et al.1985b and 1987; Andersson et al.1986) and a 47,XXX male (Müller et al.1987) have demonstrated that Y-DNA in these individuals is located at the tip of the short arm of one X chromosome, giving further support to the notion that an aberrant X-Y interchange during paternal meiosis is the most frequent underlying cause of the XX male syndrome.

These results however, do not show if the paternal X chromosome of Y(+)XX males actually arises from an interchange involving the terminal part of both parental sex chromosomes. Using pseudoautosomal probes, inheritance of the paternal pseudoautosomal region has been studied in nine Y(+)XX males by segregating the paternal X chromosome in somatic hybrids (Petit et al.1987) or by family analysis (Page et al.1987c). All these individuals have inherited the entire pseudoautosomal region from the Y chromosome and lost the pseudoautosomal region from the paternal X chromosome. The term 'X-Y interchange males' has therefore been suggested for those individuals.
arising by such abnormal terminal exchange (Ferguson-Smith et al. 1987). By contrast the occurrence of XX males lacking detectable Y-specific sequences cannot be explained by the terminal transfer model but may arise by other forms of X-Y interchange, by mutation in a non-Y linked gene or through Y mosaicism.

Since many X-Y interchange males are heterogeneous with respect to the amount of Y chromosome material present (Guellaen et al. 1984; Page et al. 1985; Page 1986) deletion maps centred around TDF on the Y chromosome short arm can be constructed. This mapping approach requires testing each individual for the presence or absence of Yp DNA sequences. A transfer gradient can then be established in which physical distance from TDF depends upon the frequency each DNA sequence is detected. Because of the underlying assumptions of the terminal exchange model, this gradient is orientated only for the proximal side of TDF; sequences located distally are transferred in all X-Y interchange males and therefore cannot be ordered with respect to each other or with respect to TDF.

Several independent researchers have made use of X-Y interchange males to construct deletion maps of the short arm of the Y chromosome (Affara et al. 1986a; Müller et al. 1986b; Vergnaud et al. 1986). Guellaen et al. (1984) used ten Y chromosome probes and four XX males to define four intervals on Yp. TDF was subsequently assigned to deletion interval 1 as all Y(+)XX males have this region in common, and the locus defined by probe 47c, found in the greatest number of XX males, maps to this interval. Vergnaud et al. (1986) using DNAs from 19 XX males, 2 XX hermaphrodites and six persons with microscopically detectable anomalies of the Y chromosome, defined the deletion map of the complete Y chromosome in terms of seven intervals, assigning each

- 32 -
of the 23 Y-specific probes to one of the intervals. In this map deletion intervals 1, 2 and 3 are within Yp (interval 1 containing \textit{TDF}), interval 4 contains the centromere, and intervals 5, 6 and 7 are on Yq. The XX males could then be divided into three classes according to the number of Y-specific fragments present.

This deletion map was extended further, dividing the short arm of the Y chromosome into 13 intervals, using 135 Y-DNA probes (Page et al. 1987a). In the course of this study, one individual, although phenotypically male, was found to carry less than 0.5% of the Y chromosome, derived from deletion intervals 1A1 and 1A2. This finding, together with deletion analysis of other males carrying part but not all of the Y chromosome, suggested that the entire \textit{TDF} locus must reside within intervals 1A1 and 1A2. Conversely, a second individual, although phenotypically female, was found to carry 99.8% of the Y chromosome, lacking only 160kb that comprise intervals 1A2 and 1B. This result, together with deletion analysis of other females carrying part but not all of the Y chromosome, strongly indicated that an essential portion of \textit{TDF} is contained within the 140kb interval 1A2.

By chromosome walking Page et al. (1987b) cloned a 230kb portion of the Y chromosome which included all of interval 1A2. Sequences within this region were shown to be highly conserved, with homologous sequences present on the Y chromosome of all mammals examined. Subsequent nucleotide sequencing of this conserved DNA revealed it encodes a protein with a tandem array of cysteine- and histidine-rich finger domains similar to those found in many transcription factors, for example \textit{Xenopus} 5S gene transcription factor \textit{Al1A} (Miller et al. 1985; Brown et al. 1985) and human transcription factor \textit{Sp1} (Kadonaga
et al. 1987). It is postulated that, in each such finger domains, a pair of cysteines and a pair of histidines are arranged about a central zinc ion in a tetrahedral coordination complex, and that such domains can interact with nucleic acids in a sequence specific manner (Miller et al. 1985). In common with other transcription factors, it is possible that the Y-encoded zinc finger protein (designated ZFY) is capable of binding to the regulatory sequences of one or more genes and positively or negatively regulating their transcription. This finding is in accord with the hypothesis that the testis determining factor is a switch which initiates a cascade of events, inducing male development. Page et al. (1987b) therefore propose that ZFY is TDF.

However, in addition to the conserved sequences on the Y chromosome, homologous sequences have also been found on the X chromosome of all mammals examined so far. In humans, this related sequence (ZFX) is located on the short arm of the X chromosome at Xp21-Xp22.3 (Page et al. 1987b) and more specifically at Xp21.3 (Müller & Schempp 1989). The high degree of evolutionary conservation suggests that neither the Y nor the X locus is a pseudogene, but that both loci encode similar proteins. These results together with the current notion of a dominantly acting sex determining factor unique to the Y chromosome, appear to argue against ZFY being TDF. To encompass these unexpected findings within a model which proposes ZFY to be TDF, Page et al. (1987b) have suggested four possible hypotheses:

1) The X-encoded protein does not function in gonadal sex determination, whilst the Y-encoded protein is sex determining simply by its presence or absence.
2) The X and Y loci act antagonistically in sex determination. For example by encoding negative and positive regulatory factors that bind to the same regulatory sequence.

3) The X and Y loci act co-operatively but are not functionally interchangeable. For example by encoding different subunits of a multimeric structure.

4) The X and Y loci are interchangeable, but the X locus is subject to X inactivation, and gonadal sex is determined by the total number of loci expressed. Thus males would have two active copies of the gene whereas females would only have one (see German 1988).

Scherer et al. (1989) have argued in favour of the 'antagonistic' model based on results obtained from a study of two 46,XY females with tandem duplications of an X short arm segment. These two individuals have cytogenetically normal Y chromosomes and in one case the presence of 18 Y-specific loci, including ZFY, was confirmed by Southern blot analysis using probes spanning the entire Y chromosome. However there was no indication of masculization in either these or two previously reported cases (Bernstein et al. 1980). In all four cases the region Xp21.2-Xp22.2 which includes the locus for ZFX, was duplicated. Although the possibility cannot be excluded that the duplication of other genes within this region or point mutations within TDF resulted in sex reversal in these individuals, it was concluded that in sporadic cases where sex reversal is accompanied by the duplication of Xp21-Xp22, ZFX was the critical locus.

Based on these findings Scherer et al. (1989) have argued that female development occurred despite the presence of ZFY thus excluding model (1). Likewise, model (3) -the 'cooperative' model- was excluded because although both ZFY and ZFX are present the individuals are
female. Finally, the single X chromosome in these individuals is not inactivated so that three active copies of the gene are assumed to be present, thus excluding model (4).

In the model proposed by Scherer et al. (1989) ZFY has a positive effect and ZFX has a negative effect with respect to the induction of male development. However ZFY is epistatic to ZFX if the ratio of active genes is 1:1, as in the normal male. In cases with a duplication of ZFX on an active X chromosome the resulting 2:1 ratio of ZFX to ZFY is sufficient to overcome the epistasis of ZFY. Consequently, testicular development is not induced and the constitutive female phenotype develops.

In contrast to these results however, it has been found that unlike other placental mammals, the mouse Y chromosome normally carries two distinct homologs of the human ZFY gene, termed Zfy-1 and Zfy-2 (Mardon & Page 1989), resulting from an intrachromosomal duplication which occurred during mouse evolution (Mardon et al. 1989). Both Zfy loci map to the sex-determining region of the mouse Y chromosome with complementary results indicating that X5Xr DNA resembles Y DNA in having two such regions whilst X5Xr' DNA shows only one (confirming that X5Xr' represents a deleted version of X5Xr) (see McLaren 1988). Northern analysis reveals that both Zfy-1 and Zfy-2 are transcribed in mouse adult testis, however, not both are required for testis determination; testis differentiation can occur in the absence of Zfy-2 (Mardon et al. 1989). The addition of Zfy-1 and Zfy-2 via transgenic manipulation, to XX mouse embryos should resolve the functional relationship of these genes to each other and their role in gonadal sex determination as well as testing the 'dosage model' of sex determination.
In marsupials the Y chromosome also appears to be testis-determining; the few XXY animals described possess testes, whereas XO animals lack testes (Sharman et al. 1970). However, unlike human XXY and XO individuals, marsupials with aneuploid sex chromosomes are not unambiguously male or female, suggesting that whereas the Y chromosome determines testicular differentiation, the dosage of X chromosomes may influence the differentiation of scrotum, pouch and mammary glands.

Using Southern blot analysis and the ZFY probe, Sinclair et al. (1988) found that ZFY homologous sequences are not present on either the X or Y chromosomes in marsupials, but map to the autosomes, implying that ZFY is not the primary sex-determining gene in marsupials. Either the genetic pathways of sex-determination in marsupials and eutherians differ, or they are identical and ZFY is not the primary signal in human sex-determination. This will be elucidated by the identification and isolation of the primary sex-determining gene borne on the Y chromosome of marsupials.

b) Short arm Y - long arm X homology

The ZFY and ZFX loci represent the first example of sequence homology on Yp and Xp outside the pseudoautosomal region (Page et al. 1987b), although short arm Y-long arm X homology has been well documented (Bishop et al. 1984; Wolfe et al. 1985; Affara et al. 1986a). Indeed the first single-copy sequence to be mapped to the Y chromosome was localised to Yp and to the region Xq13-Xq21 by in situ hybridization studies (Page et al. 1982 and 1984). This sequence, defining the locus DXYS1, produces sex-specific hybridization patterns.
on digestion of genomic DNA with *Taq* I. The homology between the X and Y-specific fragments at the *DXYS1* locus is extensive; it is estimated to be better than 99% over a region of at least 36kb (Page et al. 1984). Such limited divergence in the sequences homologous between Xq13-Xq24 and Yp and the fact that in great apes, sequences homologous to *DXYS1* are found only on the X chromosome, suggest that the human Y chromosome is, in part, derived from very recent transpositions of material from the X chromosome (Page et al. 1984; Koenig et al. 1985).

Like *DXYS1*, the loci defined by other probes recognizing Yp-Xq sequence similarities also show greater than 99% X-Y sequence homology (Geldwerth et al. 1985). However, deletion mapping studies have shown that such sequences are not clustered within a single block but are located in two distinct areas on Yp; *DXYS1* sequences (defined by probe pDP34) occur near the centromere in deletion interval 4, while a group of *DXYS1*-like sequences (defined by probes 47a, 47z, 13d and 115) are found more distally, in intervals 1 and 2 (Vergnaud et al. 1986). Y-specific or Y-autosome sequences appear to separate these two blocks. If this order is correct, then either multiple Xq-Yp transpositions occurred during recent evolution or, following a single transposition, some event resulted in the intercalation of Y-specific DNA between X-Y homologous sequences.
Tandemly repeated simple sequence DNA, often called satellite DNA, is a ubiquitous component of eukaryotic genomes. In primates, such a repeated DNA family, known as alphoid satellite DNA (reviewed by Singer 1982) exists as tandem arrays of a small unit, often found near centromeres. Sequences homologous to this alphoid family are also located at the centromeres of all human chromosomes (Manuelidis 1978; Mitchell et al. 1985), although the molecular organisation of these repeats appears to be chromosome-specific (Willard 1985a and b; Wolfe et al. 1985; Tyler-Smith & Brown 1987). As shown by Wolfe et al. (1985) and Tyler-Smith & Brown (1987), the Y chromosome has a unique set of alphoid repeats which can be distinguished from the alphoid repeats of other chromosomes by sequence and by the periodicity of particular restriction sites within the tandem array. Probes corresponding to the alphoid-like sequences (termed DYZ3) have been cloned by both groups. Wolfe et al. (1985) showed that such probes hybridize to a characteristic 5.5kb EcoRI fragment within male DNA, whilst the probes isolated by Tyler-Smith & Brown (1987) recognize a 5.7kb band (probably corresponding to the 5.5kb band of Wolfe et al. (1985)) and a minor band of 6.0kb. Restriction enzyme mapping of the cloned DNA with HaeIII, together with sequence analysis showed a repeating structure of approximately 170bp in length, with the 5.7kb unit composed of approximately 34 such subunits and the variant 6.0kb unit containing 36. Additional sequence analysis revealed that although Y-chromosome alphoid subunits are heterogeneous, being between 76% and 86% homologous to each other, they are generally more similar to one
another than they are to alphoid sequences from other chromosomes; despite showing 70% homology to X-chromosome alphoid subunits in situ hybridization studies, using an alphoid probe, revealed stronger homology to the centromeres of chromosomes 13, 14 and 15 (Wolfe et al. 1985).

It is estimated by dosage analysis that there are approximately 100 copies of the conserved EcoRI repeat unit present on the Y chromosome (Wolfe et al. 1985). This result together with the finding that alphoid DNA is not interspersed with other sequences (Tyler-Smith & Brown 1987) suggested that there is almost certainly only a single block per Y chromosome. This was confirmed by Tyler-Smith & Brown (1987) using five restriction enzymes which were found not to cut within the alphoid DNA block. Following digestion of high molecular weight DNA, the products were analysed by pulsed-field gel electrophoresis. The five enzymes each produced a single major alphoid DNA fragment, the smallest fragment size being estimated to be 475kb. In addition, they discovered a 100kb variation in the length of the alphoid block in two different Y chromosomes indicating that the tandemly repeated sequences may be very variable on a large scale in the human genome, thought to reflect the occurrence of unequal recombination and mutation events (Smith 1976).

More recently, the long-range structure of alphoid DNA has been mapped in greater detail (Tyler-Smith 1987). A total of 36 restriction enzymes were found not to cut Y chromosome DNA within the alphoid block giving rise to fragment sizes estimated to be between 540kb (Avall digests) to approximately 1200kb (Nael digests). Using these results and the results from double digests a 1.1mb map of the alphoid DNA region was constructed. This provided some indirect
evidence about the nature of the sequences flanking the alphoid block. In particular, sequences to one side of the alphoid DNA had an unexpected restriction site distribution more suggestive of a simple sequence such as that found in other satellite DNAs, whilst sequences flanking the other side of the block had a restriction site distribution similar to that expected in a stretch of typical human DNA consisting of a mixture of unique sequences and interspersed repeated sequences. Tyler-Smith (1987) suggests this flanking region could represent the boundary between euchromatic and heterochromatic DNA. Evidence supporting this would result from a more detailed map of a human centromeric region which would be expected to extend from the euchromatic DNA on one arm of the chromosome, through the heterochromatic DNA, to the euchromatic DNA on the other, and which would form the basis for a more detailed structural analysis of the region and a systematic search for functional mammalian centromeric sequences.

1.2.4 The long arm euchromatic region

a) Sequences involved in spermatogenesis

Although Y chromosome anomalies account for only 1 to 3% of male sterility cases (Koulischer & Schoysman 1974) the euchromatic portion of the long arm has been associated with a factor required for spermatogenesis (AZF) through karyotypic analysis. Azoospermia, in particular, has been related to deletions within this region (Tiepolo
& Zuffardi 1976; Neu et al. 1973; Yunis et al. 1977; Steinbach et al. 1979; Cohen et al. 1983; Schempp et al. 1985; Chandley et al. 1986; Hartung et al. 1988). In humans, data relating to this factor are very limited. Mice studies however, have provided a key to the role of the Y chromosome in male germ cell differentiation.

By comparing aspects of the male phenotype in mice which differ only in the strains from which their Y chromosome is derived, differences have been attributed to strain-specific variants at Y-chromosomal loci. Using this approach, Y-chromosomal effects have been identified on testis weight, serum testosterone levels, serologically detected male antigen levels, target organ sensitivity to androgen, the frequency of abnormal sperm and the frequency of Y-chromosome non-disjunction (Stewart 1983; Jutley & Stewart 1985; Stewart & Jutley 1987). From this it has been suggested that a minimum of four loci are involved.

It is now well established that XX germ cells in the mouse testis are able to enter the male gametogenic pathway (McLaren 1988). However, cytogenetic and histological studies of XO/XY or XO/XY/XXX mosaic mice reveals that germ cells lacking a Y chromosome rarely survive into the adult testis, whilst those which do survive to enter meiosis degenerate before the first meiotic metaphase (Levy & Burgoyne 1986). More recently an XO + XY chimeric male was discovered which appeared from tissue analysis, to be predominantly XO. Nevertheless, spermatogenesis was supported entirely by the XY cell component. It appears therefore, that XO germ cells are failing despite an XY Sertoli cell environment and that the germ cells themselves require a Y chromosome in order to take part in spermatogenesis (Burgoyne 1987). Levy & Burgoyne (1986) concluded therefore, that the mouse Y
chromosome harbours a spermatogenesis gene (or genes) required for
male gametogenesis which must be expressed in the germ line. This
gene is termed Spy.

Further insight into spermatogenesis has been gained from studies
involving sex-reversed mice. Crosses between XYSxr males and XO
females give rise to XOSxr males (Cattenach et al. 1971) which, despite
carrying out all stages of spermatogenesis, are sterile - the sperm
produced are misshapen and few in numbers (Burgoyne & Baker 1984).
Unlike XO germ cells therefore, XOSxr germ cells escape early
spermatogenic failure, implying that Spy must reside within the Sxr
region. From these results it was also concluded that other
sequences, outside the Sxr region, must be required for complete
spermatogenesis and normal sperm development. Two models have been
postulated to explain why the absence of a complete Y chromosome leads
to spermatogenic anomalies: one model, favoured by Cattenach et
al. (1971) and by Burgoyne & Baker (1984) proposes that the Y
deficiency is triggering a 'meiotic quality control mechanism' which
acts to eliminate the products of pachytene spermatocytes which have
unsynapsed or incomplete synapsed chromosomes (Miklos 1974). In the
germin line therefore, the Y chromosome acts as a pairing partner for
the X during meiotic prophase. The second model proposes that there
is a Y gene (or genes), absent from Sxr, which are involved in sperm
morphogenesis and function (Eicher & Washburn 1986). Although
further studies are required to assess these two models it is possible
that spermatogenesis in XOSxr mice is being affected by both a Y-gene
deficiency as well as by a pairing deficiency.

Additional studies have also been carried out on the Sxr variant
Sxr'. Burgoyne et al. (1986) have shown that XOSxr' mice have an early
block to spermatogenesis which is indistinguishable from that in the 'XO' tubules of XO/XY mosaics. It was concluded that when $Sxr$ became altered to $Sxr'$, genetic information required for spermatogenesis ($Spy$) had been lost along with the genetic information required for H-Y antigen expression ($Hya$). From these results two explanations could be offered: either $Spy$ is the Y-chromosomal gene which controls H-Y expression implying H-Y antigen has a role in spermatogenesis; or $Spy$ and $Hya$ are separate loci which have both been deleted in $Sxr'$.

In parallel with these investigations, studies to map the male-specific transplantation antigen, H-Y, have confirmed that in humans the H-Y antigen gene is located on Yq where the gene controlling spermatogenesis has also been placed (Tiepolo & Zuffardi 1976). By H-Y typing a series of XX males and XY females which were used to construct a Y chromosome deletion map (Vergnaud et al. 1986; Disteche et al. 1986; Page 1986; Simpson et al. (1987b) were able to map the H-Y gene between deletion intervals 4B and 7 on the Y chromosome long arm. Although additional studies are required to define the map position more precisely, the results so far do not exclude the possibility that the gene controlling spermatogenesis could be coincident with the H-Y gene (Simpson et al. 1987b).

Furthermore, Page (1987) has used a similar approach to map a putative Y-linked gene which predisposes morphologically abnormal gonads to develop gonadoblastomas. These neoplasms defined histologically by the occurrence of germ cells, immature Sertoli and granulosa cells within well circumscribed nests (Scully 1970), arise in gonadal tissue which often lacks the usual architecture of either an ovary or a testis and is therefore termed 'dysgenetic'. In addition, gonadoblastoma is not only restricted to those individuals
with dysgenetic gonads but more exclusively to those who also carry Y chromosomal material (reviewed by Verp & Simpson 1987). However, it is rarely, if ever, found in normal males (or females) but occurs mainly among individuals with dysgenetic gonads who, although phenotypically female, have a 46XY or mosaic 45X/46XY karyotype. Page (1987) has postulated that in the context of the dysgenetic gonad and the apparent necessity for Y chromosomal sequences, a structural or regulatory gene on the Y chromosome, which in normal males has a physiological function in the testis (perhaps in or prior to spermatogenesis), behaves as an oncogene, in concert with other unrecognised factors, in the presence of a markedly abnormal gonad. This postulated gene has been termed GBY. Limited deletion mapping evidence suggests that GBY is located near the centromere or on the long arm of the Y chromosome within deletion intervals 4B to 7. As this is the same region to which AZF is thought to map and to which H-Y has been located (Simpson et al. 1987) Page (1987) suggests that although these could be three distinct loci, GBY, H-Y and AZF may be synonyms for a single gene merely denoting the pleiotropic manifestations of that gene. Again, additional mapping studies, which will define these loci more precisely, should also demonstrate if they are separate factors.

b) Steroid sulphotase

Steroid sulphotase, STS, is ubiquitously expressed in mammalian tissues and appears to play an important role in the conversion of sulphated steroid precursors to oestrogens during human pregnancies.
In man, STS deficiency results in a syndrome of X-linked ichthyosis. Through somatic cell hybrid studies, deletion mapping and linkage analysis, a functional gene for STS has been assigned to the region Xpter-Xp22.3 (Mohandas et al. 1979; Müller et al. 1980; Tiepolo et al. 1980; Weiacker et al. 1984) and more precisely to Xp22.3 by in situ hybridization studies (Ballabio et al. 1987). In common with other loci in this region, STS shows X-inactivation (Shapiro et al. 1979) although additional evidence suggests that escape from inactivation may not be complete (Migeon et al. 1982). These results, together with lack of evidence for a functional Yp locus suggests STS lies just centromeric to the X-Y exchange region. This is in contrast to the mouse steroid sulphatase gene which behaves pseudoautosomally with an apparently functional homologue on the Y chromosome (Kéltès et al. 1985; Nagamine et al. 1987). Recently Yen et al. (1987), Ballabio et al. (1987) and Conary et al. (1987) independently isolated STS cDNA clones using specific anti-STS antibodies. Using these clones as probes, evidence for the presence of cross-hybriding sequences on the long arm of the Y chromosome has been obtained (Fraser et al. 1987; Yen et al. 1987). Further subregion assignments, using somatic cell hybrids and cell lines which contain deleted Yq chromosomes, map these related sequences to Yq11.2 (Fraser et al. 1987). Subsequently, twelve clones, covering 100kb of Y-STS-related sequence, have been isolated and certain regions sequenced (Yen et al. 1988). Analysis shows the similarity index between these X and Y sequences ranges from 85% to 94% although it appears that several exons as well as the region corresponding to the STS-X promoter, have been deleted. Furthermore, considerable sequence divergence from the STS-X gene has produced numerous stop codons. In addition to sizeable deletions the STS-Y
locus has at least four large insertions which consist of complete or truncated LINE sequences. Analysis shows that these alterations in the Y sequence render it incapable of producing a functional protein product; the STS-Y locus therefore appears to be a pseudogene in the process of degeneration. From results obtained from genomic blots containing primate DNA probed with a sequence from the STS-Y region, Yen et al. (1988) concluded that the STS gene was pseudoautosomal in ancestral rodents, insectivores and primates, and that a putative pericentric inversion of the Y chromosome occurred at the time after the divergence of the prosimians from the higher primate lineage. However further evaluation of the mouse and human steroid sulphatase sequences is required since the degree of homology between the human and rodent steroid sulphatase genes appears to be very limited (Fraser et al. 1987).

c) Amelogenin sequences

Recently a cDNA clone for the mouse amelogenin gene has been used to determine the chromosomal locations of the human and mouse amelogenin (AMEL) loci (Lau et al. 1989). This gene codes for the amelogenin class of protein which, together with enamelin, compose the enamel matrix covering the teeth. In humans, results obtained from Southern blot analysis, using somatic cell hybrids, have localized AMEL to the distal region of the X chromosome short arm (Xp22.1-p22.3) and to the centromeric region of the Y chromosome, possibly Yq11. Unlike the locus in humans, however, the mouse AMEL locus appears to be confined to the X chromosome. It is not known whether the AMEL-Y
locus is actively transcribed or whether it represents a pseudogene, as in the case of STS. However, it is possible that the pericentric inversion of the Y chromosome which translocated the STS-Y gene to the long arm, disrupting meiotic exchange and permitting sequence divergence, also relocated AMEL-Y. Further studies, involving DNA sequencing and phylogenetic hybridization studies are therefore required to determine the extent of homology between AMEL-X and AMEL-Y.

The chromosomal assignments of the human AMEL loci are consistent with the hypothesis that perturbation of the amelogenin gene is involved in X-linked types of amelogenesis imperfecta, a condition affecting enamel formation (see Lau et al. 1989). Previous investigations have suggested that a Y-chromosome gene located near the q11 region, known as TSY, participates in regulating the size of human teeth (Alvesalo & de la Chapelle 1981). The localization of AMEL to the Y chromosome strongly indicates that this and TSY are either the same or closely related loci.

d) Other sequences

In common with STS and AMEL, several random human genomic sequences have been isolated, which also map to short arm of the X chromosome and to the euchromatic region of the Y long arm. DXS31 (Koenig et al.1984), DXS69 (Kunkle et al.1983), GMGXY3 (Affara et al.1986a) and CRI-S232 (Knowlton et al.1989) all detect homology between the sex chromosomes and have been localized to Xp21-pter and Yq11-Ypter. Furthermore in the one case examined, DXS31, X-Y homology
is also found in the chimpanzee, but an X only location is found in the macaque (Koenig et al. 1984). In correlation with the results obtained for STS therefore, these sequences may have been involved in an X to Y transposition event occurring prior to the divergence of the human and chimpanzee lineages but after the split of the higher primates from the Old World monkeys (Bickmore & Cooke 1987).

Other sequences within Yq are known to be homologous to sequences on Xq. Cooke et al. (1984) have described a sequence (pUC9H1) which has been localized using somatic cell hybrids, hybrids with X-autosome translocations and in situ hybridization, to Xq24-Xqter and distal Yq, whilst Bickmore & Cooke (1987) have isolated a sequence, 2:13, which has been mapped to Xq12-28 and to Yqcen-q11.1. In addition two pseudogenes are known to be present on Yq. Heilig et al. (1984) found actin-like sequences, resembling cytoskeletal actin rather than muscle α actin, on both the X and Y chromosomes. Through hybridization of a human α skeletal actin cDNA probe to DNAs from a panel of human-mouse hybrid cell lines, the X-linked actin sequences have been assigned to the centromeric region Xp11-Xq11 whilst the isolation of Y-linked genomic actin sequences and flanking regions has allowed the localization of the Y-linked sequences (ACTP1) to the euchromatic part of the Y long arm (Koenig et al. 1985). Subsequent sequencing of the Y-located actin gene has confirmed its inability to be expressed (Koenig et al. 1985). Sequences cross reacting with a probe for the argininosuccinate synthetase gene are also present on the X and Y chromosomes as well as dispersed throughout the genome (Daiger et al. 1982). In common with the actin-like sequences, the two Y-linked argininosuccinate synthetase sequences, ASSP4 and ASSP5, are known to be pseudogenes (Freytag et al. 1984). Although these sequences occur
on both the X and Y chromosomes it is thought that the ASS pseudogenes on the X and Y chromosomes were probably dispersed by a mechanism common to all the ASS pseudogenes and do not reflect strict X and Y homologies (Daiger & Chakraborty 1985).

1.2.5. The long arm heterochromatic region

The heterochromatic region of the long arm constitutes approximately one third to one half of the Y chromosome and stains intensely with quinacrine mustard. It is probable that there are no transcribed sequences within this portion of the long arm as extensive variation in the length of this region is seen within the population of healthy fertile men (Bobrow et al. 1971; McKay et al. 1978); the region is present in 1 in 3000 normal females (Cooke & Noel 1979). It is primarily composed of two tandemly repeated male specific sequences, DYZ1 and DYZ2 (Cooke 1976). Digestion of male genomic DNA with the restriction enzyme Haelll yields two distinct ethidium staining bands at 3.4kb (representing DYZ1) and 2.1kb (representing DYZ2) that are not seen in similar digests of female DNA. Members of these two repeat families have since been purified, cloned and analysed by several laboratories. Independently, Kunkle et al. (1976) reported the isolation of Y-chromosome repeated DNA (termed it-Y) by hybridizing single stranded male DNA with a large excess of female DNA and recovering the remaining single stranded fraction. This fraction when used as a probe reacted with DNA from normal males but failed to react with DNA from males carrying a Y chromosome deleted for the
heterochromatic region). It-Y sequences were subsequently shown to be largely accounted for by the 3.4kb repeat family described by Cooke (Kunkle et al.1977).

In situ hybridization studies have localized the majority of the members of the DYZ1 and DYZ2 families to the heterochromatic portion of the Y (Yq12-Ypter) (Bostock et al.1978; Cooke et al.1982) with the 2.1kb family restricted to the distal tip whilst the 3.4kb family appears to be dispersed along the entire long arm (Szabo et al.1979). In addition, the amount of the 3.4kb repeat sequence present in any male seems to reflect the length of the heterochromatin present in that male (Kunkle et al.1977; Bostock et al.1978). Cooke (1976) has estimated there to be between 4000 and 6000 copies of the 3.4kb repeat and 2000 copies of the 2.1kb repeat.

While the 3.4 and 2.1kb fragments are themselves from the Y chromosome, each contains sequences which cross react with DNA from other chromosomes. Kunkle & Smith (1982) have shown that the 3.4kb HaeIII fragment is composed of Y-specific sequences interspersed with non-Y-specific ones and that the major sites of autosomal homology are chromosomes 1, 9, 15, 16, 21 and 22 (Burk et al.1985). Even under considerably reduced stringencies of hybridization, however, no sequences homologous to the X chromosome are found (Cooke & McKay 1978; Willard 1985b). In addition studies of the 3.4kb repeat, involving reassociation kinetics or direct sequencing have shown considerable heterogeneity among DYZ1 sequences. At least a portion of the 3.4kb repeat consists of simple repeats of the pentamer 5'-TTCCA-3'.

The DYZ2 family appears to be more complex than the DYZ1 family. The 2.1kb probe does not cross react with the 3.4kb HaeIII fragment,
indicating the sequence independence of the *DYZ1* and *DYZ2* repeats. Studies by Young *et al.* (1983) and Cooke *et al.* (1982) have established that the 2.1kb *Haelll* repeat is not the fundamental repeat unit on the Y. Rather, partial digests of male DNA which would be expected to reveal a ladder of fragment sizes corresponding to exact multimers of the basic tandem sequence, indicate a basic repeat unit of 2.4kb (Cooke *et al.*1982; Young *et al.*1983; Frommer *et al.*1984). Sequence analysis of members of the *DYZ2* repeat has revealed a complex structure involving both A-T and G-C rich subregions and the presence of an *Alu* repeat (Frommer *et al.*1984). However, it is not certain if all *DYZ2* repeats have this structure.

Higher primates share *DYZ1* and *DYZ2*; however, only in humans are these sequences concentrated on the Y chromosome (Cooke *et al.*1982; Szabo *et al.*1980). Thus just based on these repeats, half of the sequences on the Y chromosome are of recent evolutionary origin.

In addition to the *DYZ1* and *DYZ2* repeats, several low copy sequences have been provisionally assigned to the heterochromatic region. Cooke *et al.* (1984) reported the isolation of a probe Y21.3, which, by *in situ* hybridization recognizes a locus in the Yq satellite region as well as a locus on the long arm of the X chromosome (Xq24-Xpter). Likewise, Rappold *et al.* (1984) used *in situ* hybridization to map two Y-derived sequences to Yq11-Yqter, both of which also react with homologous loci on other human chromosomes.
1.3. Background and aims of this study

In classical genetic investigations, the direction of analysis flows from the phenotype to the gene and DNA level. The Y chromosome, however, is largely refractory to this form of investigation as a direct consequence of its defined function in sex determination and the lack of recombination. The need to retain the sex determining gene on the Y chromosome is a strong theoretical reason for the absence of recombination between the sex chromosomes, and the large number of X linked genes not found on the Y chromosome testifies to the absence of extensive sex chromosome recombination. Genetic analysis by family linkage studies is therefore impossible when dealing with sequences outside the pseudoautosomal region. Within this region the X and Y chromosomes must pair during male meiosis, a function achieved by specific sequence homology between the two chromosomes. Theoretically any gene could be found in the pseudoautosomal region as it will be present identically in male and females, assuming that X-located pseudoautosomal genes escape X-inactivation. Outside the pseudoautosomal region the Y chromosome appears to be a mosaic of sequences derived from several different sources at different evolutionary times.

A better understanding of the Y chromosome has resulted from molecular genetics in which DNA sequences are isolated and then used to study phenotypes and function. Apart from the repeat elements located at the centromere and Yq heterochromatic regions, five approaches have been used to clone sequences from the Y chromosome:
1) Initially unique Y DNA fragments were obtained by screening random probes, isolated from total human genomic libraries, for their ability to hybridize differentially to the male as opposed to the female genome (Daiger et al. 1982; Page et al. 1982). As this strategy prohibited the isolation of large numbers of probes, procedures for producing libraries highly enriched for the Y chromosome were subsequently employed.

2) Human-rodent somatic cell hybrids, containing the Y chromosome as the only detectable human contribution (e.g. Marcus et al. 1976) have been used to construct genomic libraries in cosmid (Bishop et al. 1983; Wolfe et al. 1984; Cooke et al. 1985) or bacteriophage vectors (Burk et al. 1985). By screening the resultant libraries with a human repetitive probe, recombinant clones specific for the Y chromosome were obtained. Isolation of single copy sequences was subsequently achieved by sub-cloning smaller fragments of the phage or cosmid clones.

3) The Y chromosome has been physically isolated from other human chromosomes using fluorescence activated flow cytometry. On the basis of their size, chromosomes are assigned to specific peaks in the resultant flow karyotype and peaks of interest isolated. Chromosomal DNA is then extracted, restricted and cloned, normally into high-efficiency phage lambda vectors. Using this method Y-chromosome enriched libraries have been constructed for the mouse (Bishop et al. 1985) as well as the human (Fantes et al. 1983; Müller et al. 1983; Müller et al. 1986). Such banks, however, are rarely complete or totally pure.

4) To obtain DNA sequences specifically from the mouse Sxr region, Bishop et al. (1987) microdissected the region of interest from metaphase chromosomes and microcloned the picogram quantities of DNA
obtained. This technique however, relies on the easy identification of the relevant chromosome without staining. In this case a marker Y chromosome was used; a metacentric Y formed by a pericentric inversion.

5) Several sequences derived from the X chromosome have been shown to detect homologous loci on the Y chromosome (Kunkel et al. 1983; Koenig et al. 1984). Such sequences have been useful for mapping the Y as well as the X and for examining the evolutionary relationship between these two chromosomes.

Many of the Y-linked (or putatively Y-linked) genes described to date, have been localised using deletion mapping techniques eg. ZFY (Page et al. 1987), H-Y (Simpson et al. 1987) and GBY (Page 1987). In addition, using the methods described, several Y-linked single copy sequences have been isolated which detect transcripts during Northern blot analysis. Using a Y specific random DNA probe taken from a highly enriched mouse Y chromosome library, Bishop & Hatat (1985) identified and cloned an mRNA transcript expressed solely in the adult mouse testis. The genomic sequence appears to be repeated about 250 times and in situ hybridization studies suggest that it maps along the entire length of the Y (Bishop et al. 1987). Although the function of the transcript is unknown, its failure to hybridize to DNA from XX.Sxr mice suggests it is unlikely to be involved in sex determination, and its absence from both the mouse germ cell derived teratocarcinoma PCC7 and the presumptive mouse Sertoli cell line TM4, appears to suggest it is not involved in the early stages of spermatogenesis.

Similarly, Arnemann et al. (1987) isolated a Y-specific clone from a human Y-chromosome enriched cosmid library. Deletion mapping localised this clone to the median region of Yp, neighbouring TDF,
whilst northern blot analysis showed a transcription signal in poly(A)+ RNA of human testis. Although sequence analysis revealed an open reading frame of 522bp, no significant homologies with known DNA or protein was found. Likewise, Leroy et al. (1987) used a human Y-chromosome random DNA sequence obtained from a human Y-specific cosmid library to probe a northern blot of human and mouse RNAs. The sequence, derived from the long arm of the Y chromosome, was found to detect mRNAs specifically expressed in the testis. However, Southern analysis and in situ hybridization studies using the mouse cDNA sequence as a probe under high-stringency conditions, revealed that in rodents at least, this sequence is autosomally located. From signal intensities on Southern blot analysis LeRoy et al. (1987) predict that the Y-chromosome sequence is not the active gene, which is probably autosomally located, but represents a Y-located pseudogene.

In an effort to identify transcripts specifically from the Sxrr region of the mouse Bishop et al. (1987) probed a flow-sorted Y library with a mixture of total mouse DNA and various repeated sequences. Negative plaques were then replica-plated and screened with cDNA probes synthesized from male testis, male liver and female liver. Clones which were positive with the testis cDNA but negative with the other cDNA probes were identified and shown by Southern blot analysis to detect restriction fragments in normal male, XXSxrr and XYsxxr DNAs but not female DNA.

My primary objective in this study has been to identify other transcribed sequences on the human Y chromosome, using molecular techniques which facilitate cloning. More specifically, to detect CpG islands, which are characteristic of many eukaryotic genes and use these as 'markers' in the search for Y-linked genes.
1.3.1. CpG or HTF Islands

The dinucleotide CpG is rare in vertebrate DNA, occurring at only one-fifth of the frequency expected from base composition (Russell et al. 1976). Since CpGs in bulk DNA are often methylated at position 5 of the pyrimidine ring, it has been proposed that the rarity of CpG is caused by the high frequency of deamination of 5-methylcytosine to give thymine (Bird 1980). Owing to the high level of CpG methylation and rarity of the dinucleotide, vertebrate DNA is poorly cleaved by restriction endonucleases which are blocked by 5mCpG. Cooper et al. (1983) however found that about 1% of the genome, in a wide range of vertebrates, is very frequently cut by the methylation-sensitive enzymes Hpall and Hhal. In chicken DNA, for example, Hpall sites are on average 120bp apart in this fraction compared with an average spacing of 1.8Kb in the chicken genome as a whole, so that Hpall cleavage of these regions gives rise to numerous small fragments - the Hpall tiny fragments (HTF). Such sequences therefore contain CpG at more than 10 times its density in bulk DNA (Bird et al. 1985) with the CpG dinucleotides clustered, forming discrete 'islands', usually 1-2kbp long. Although no extensive sequence homologies have been detected between different CpG islands (Bird et al. 1985), their atypical sequence composition gives rise to restriction endonuclease recognition sites which are C+G rich and contain one or more CpGs. Such 'rare cutting' or 'CG' enzymes, including BssHII (G+C GGC), Nael (GCC+GGC), SacII (CCGCGG), Eagl (C+GGCCG), NotI (GC+GGCGGC) and NarI (GG+CGCC), show a strong preference for island over inter-island DNA (Brown & Bird 1986); a preference which is most extreme in chromosomal DNA as inter-island CpGs are not only rare but are usually methylated.
and therefore immune to cutting by CG enzymes. Indeed, 75% of the
total number of sites for each of these CG enzymes has been calculated
to occur in CpG islands, and on average every island will contain one
site for each enzyme (Brown & Bird 1986).

Bird (1986) has proposed two criteria by which the HTF character
of a region of DNA can be judged. The first criterion depends upon
CpG frequency. A test of HTF character therefore, is to determine the
G+C content of a likely region of several hundred base pairs, and
count CpGs and GpCs. For an area to be considered CpG island-like,
the G+C content must be over 50% and the CpGs must roughly equal the
G+Cs. The second criterion depends upon lack of CpG methylation
within the putative HTF island region. If both these conditions are
satisfied the region is termed HTF-like.

It is now clear that in addition to their methylation properties
and high C+G content, many CpG islands are associated with genes, and
particularly with the region where transcription begins. A recent
detailed survey (Gardiner-Garden & Frommer 1987) shows that all
sequenced 'housekeeping' genes that are transcribed by RNA polymerase
II, have islands at their 5' ends, when judged by CpG frequency,
(however, lack of methylation has yet to be established at the great
majority of candidate CpG islands). In addition many tissue specific
genes also have typical islands; for example the human retinol binding
protein gene (d'Onofrio et al.1985), the Thy-1 gene of mouse (Kolsto
et al.1986) and the α-globin genes of man (Bird et al.1987). A
comparison of the approximate number of islands (about 30,000) with
the calculated number of genes (20-50,000) suggests that a high
proportion of islands will be gene-associated.
At present the role of these islands is unknown. There is evidence however, that DNA methylation provides one mechanism for stably altering the local structure of a gene thereby altering DNA-protein interactions and thus playing a role in the regulation of gene activity (Groudine et al. 1981, Keshet et al. 1986). Furthermore, there is evidence that transcription of genes with CpG islands is inhibited when the island is artificially methylated, whilst methylation of non-island regions shows no obvious correlation with activity. The most likely hypothesis states therefore, that it is the methylation status of CpGs in specific 5' upstream regions of genes that plays an important role in regulating gene activity. This was first illustrated by McGhee and Ginder in 1979, who reported that certain specific HpaII methylation sites in the 5' region of the chicken β-globin genes are unmethylated in erythrocytes and reticulocytes but methylated in oviduct tissue. Additional support for this hypothesis has come from studies by several groups who have used direct tests to establish the relationship between DNA methylation and gene activity. The protocol involves the in vitro methylation of a DNA template and an assessment of its transcription after transfection into eukaryotic cells (Stein et al. 1982; Doerfler et al. 1985). Busslinger et al. (1983), for example, established that methylation of the upstream region of the human γ-globin gene abolished its expression by (non-tissue-specific) transcription factors in L-cells. In an initial approximation an upstream region between -790 and +92 was found to be important in methylation/suppression. Similarly, transfection experiments with an artificially methylated hamster APRT gene demonstrate that methylation of the body of the gene does not interfere with transcription, but methylation of the 5' region, which
is a CpG island, markedly reduces the level of transcription (Keshet et al. 1985).

Bird (1986) has suggested that methylation-free islands distinguish regions of the genome available for interaction with nuclear factors in all cells. Their frequent association with the 5' domains of genes suggests that they may be preferred sites of interaction between DNA and DNA-binding proteins. For example, any nuclear protein that can bind to non-methylated CpG-containing sequences will tend to concentrate at islands, and hence at many (housekeeping gene) promoters. The transcription factor Spl is a candidate for such a protein, as its consensus binding site contains a CpG and is G+C rich (Kadonaga et al. 1986). Moreover the great majority of non-viral genes whose expression is thought to be facilitated by Spl have CpG islands at their 5' ends (Dynan 1986). Such regions could constitutively bind common transcription factors which might initiate transcription from constitutive promoters (such as the HPRT gene). The promoters of tissue-specific genes with islands, such as Thy-1, would however require additional tissue-specific factor(s) or chromatin alterations to activate transcription in a tissue-specific manner (Kolsto et al. 1986; Bird 1986). Alternatively, tissue-specific genes with islands could be silenced in appropriate non-expressing tissues, by efficient trans-acting repressors. The gene for retinol binding protein, for example is associated with island-like DNA sequences and is expressed only in the liver (d'Onofrio et al. 1985). The cloned gene retains tissue-specific expression upon transformation into hepatic or non-hepatic cells. It is expressed in inappropriate cells, however, when a small region of the promoter is mutated before transformation (Colantuoni et al. 1987).
The implication is that the gene is potentially expressible in many tissues, but is normally blocked by binding a repressor. Mutation of the repressor binding site leads to constitutive expression. In contrast, the promoter regions of tissue-specific genes without CpG islands would be unavailable to the ubiquitous transcription factors and would require highly tissue-specific factors to initiate transcription. The human β-globin genes, for example, require a transcriptional enhancer in cis for efficient transcription whereas the transfected α genes are expressed in several cell types, at a high rate without the presence of an enhancer (Mellon et al. 1981).

Overall however it is possible that CpG islands can lead to simplification of the large vertebrate genome by effectively disqualifying non-island DNA from interactions with ubiquitous transcription factors. DNA methylation would play an essential part in this process, by minimizing the number of non-island CpGs through mutation, and by blocking any spurious binding sites for proteins that are intended to interact with genes.

Further evidence for the inhibitory effect of methylation at islands is provided by studies of genes on the inactive X chromosome of mammals. The HPRT, PGK and G6PD genes have been studied in detail; the CpG clusters in their 5' regions were found to be methylated on the inactive X but unmethylated on the active X (Toniolo et al. 1988; Wolf et al. 1984; Yen et al. 1984). Transfection experiments which tested the ability of purified DNA to confer the HPRT* phenotype on HPRT− recipient cells showed that the HPRT gene on the somatic inactive X is unable to express. However, DNA from the active X chromosomes can transform cells to HPRT* (Liskay & Evans 1980; Chapman et al. 1983; Venolia & Gartler 1983), as can DNA from an inactive X
chromosome that has been re-activated by the inhibitor of DNA methylation, 5-azacytidine (Venolia et al.1982). Methylation studies of X-linked genes and several random sequences have shown that no global methylation differences exist between the active and inactive X chromosome DNA and the inactive state of the X-linked genes has been correlated only with methylation of CpG islands (Toniolo et al.1988; Wolf & Migeon 1985; Lindsay et al.1985). These findings have suggested that CpG islands, regularly spaced along the X chromosome, could be control elements for the maintenance of X chromosome inactivation and that the methylation state of the CpG islands could directly affect the chromatin structure of neighbouring regions (Riggs et al.1985). Evidence supporting this has been obtained by examining the chromatin structure of genes introduced into cells by DNA-mediated gene transfer. While unmethylated DNA adopts a DNase 1 sensitive structure, fully methylated sequences are resistant to DNase 1 and are structurally similar to the inactive genes of the cell (Keshet et al.1986). The role of CpG clusters could therefore involve the stabilizing and maintenance of a chromatin state which renders the genes inaccessible to the transcriptional machinery of the cell. Indeed Lock et al. (1987) reported that during embryonic development in mouse, methylation of the CpG islands of the HPRT gene occurs after inactivation of the X chromosome, suggesting that methylation is important in the maintenance of X inactivation rather than in the initial events of the process.

The two Y-linked genes which have been studied in most detail, are also known to be associated with CpG islands. Primary sequence analysis of the MIC2 cDNA revealed a high C+G level and an abundance of CpG pairs (Darling et al.1986). Subsequent cloning of genomic
sequences identified a CpG island associated with the 5' end of MIC2, which methylation studies revealed to be unmethylated on the active and inactive X chromosomes as well as the Y chromosome. The hypomethylated state of the MIC2 CpG island on the inactive X chromosome contrasts with similar islands identified for other X-linked genes and correlates with expression of MIC2 from the inactive X (Goodfellow et al. 1988). Similarly Page et al. (1987) found five BssHII sites, four EagI sites, and two SacII sites grouped in two clusters very close to ZFY. However, it is unknown whether either of these two CpG islands correspond to the 5' end of the gene or whether the CpG dinucleotides in these islands are unmethylated in human genomic DNA. The existence of one of these islands was inferred from pulsed-field gel electrophoresis (PFGE) studies by Pritchard et al. (1987). Since 'rare cutting' enzymes cleave DNA almost exclusively within CpG islands, very large DNA fragments result from digestion. Such fragments can then be analysed using PFGE (Schwartz & Cantor 1984; Carle & Olsen 1984) and long-range restriction maps constructed. From such maps, clusters of rare restriction enzyme sites can be identified and used to diagnose the presence of a CpG island (Brown & Bird 1986). Using this technique to map the boundary of the pseudoautosomal region of the human Y chromosome, Pritchard et al. (1987) found a group of rare restriction sites corresponding to the CpG island at the 5' end of the MIC2 gene. A second putative island, centromerically proximal to the first, was also discovered less than 80kb away. Because MIC2 was known to map close to the pseudoautosomal boundary and TDF was inferred from XX male studies also to map close to the pseudoautosomal boundary, but on the other side, it was speculated that this second island could lie 5' to TDF.
Similar studies to map the Y chromosome pseudoautosomal region have identified other putative CpG islands (Petit et al. 1988; Brown 1988; Rappold & Lehrach 1988). Analysis reveals a telomeric region of high CpG density which extends 500kb into the pseudoautosomal region. Though many of the 'rare' restriction enzyme recognition sites are clustered within this area, the extreme density of the sites is not reminiscent of classical CpG islands. Indeed, studies of DNA from a patient with Turner's syndrome indicate that these sites are methylated and therefore cannot be classified as CpG islands (Brown 1988). However, this CpG rich region is similar to that found on the tip of chromosome 4 (Bucan et al. 1988) and might reflect some structural feature possibly related to telomeric function (Petit et al. 1988). Two, possibly three, further CpG rich clusters were found within the rest of the pseudoautosomal region. These appear more typical of classical CpG islands and are therefore more likely to label gene locations (Petit et al. 1988; Brown 1988).

In this investigation I have approached the study of CpG rich regions of the human Y chromosome in two ways:

1) By screening Y-specific cosmid clones (Wolfe et al. 1984) using \textit{Hpall}/(or \textit{MspI}) restriction analysis and selecting those which contain numerous \textit{Hpall}-tiny fragments. The character of selected clones has been studied in more detail using Southern mapping, northern analysis, library screening, DNA sequencing and methylation studies.

2a) By directly cloning \textit{Hpall}/fragments from fluorescence activated flow sorted human Y chromosomes provided by the MRC Population Genetics Unit, Edinburgh, and analysing a selected number of random clones for their association with transcribed sequences.
2b) By directly cloning MspI tiny fragments from a Y-chromosome specific λ library obtained from the American Type Culture Collection and analysing a selected number of random clones. In parallel to this study I have also cloned MspI tiny fragments from a chromosome 9 specific λ library also obtained from the American Type Culture Collection.

Using these methods it was proposed that an insight into the frequency of CpG islands, and thus genes, on the human Y chromosome would be gained.
METHODS AND MATERIALS

2.1 General Techniques and Reagents

DNA concentrations were determined spectrophotometrically at 260nm (Maniatis, Fritsch & Sambrook 1982).

DNA precipitations were routinely performed by the addition of sodium acetate, pH 4.5, to a final concentration of 0.3M and adding 2 volumes of ethanol, before chilling at -20°C overnight or at -70°C for 1 hour. Precipitated DNA was recovered by centrifugation at 10-15K rpm and washed in 70% ethanol before briefly drying in vacuo and resuspending in H2O or TE.

G50 spun column chromatography was used to remove low molecular weight contaminants and unincorporated nucleotides from DNA (Maniatis et al. 1982). Sephadex G50 resin was expanded in sterile TNE containing 0.1% SDS and columns were spun at 1500rpm for 3 mins before applying the sample and repeating the spin.

Buffers and media were prepared using Analar grade reagents and where appropriate sterilized by autoclaving (15psi, 121°C for 25 mins) or by filtration (0.22μm pore size).

Biological containment. Manipulations of recombinants were carried out in accordance with the stipulations for GMP, as recommended by the Genetic Manipulations Advisory Group (GMAG) and the UCL genetic manipulations safety subcommittee.
Standard buffers and media

Denhardt's solution (1x) 0.02% Ficoll, 0.02% polyvinylpyrrolidone, 0.02% BSA (fraction V)

SSC (1x) 150mM NaCl, 15mM sodium citrate pH7.0

SSPE (1x) 180mM NaCl, 1mM EDTA, 10mM NaH₂PO₄

TAE (1x) 40mM Tris, 20mM sodium acetate, 2mM EDTA pH7.0

TBE (1x) 89mM Tris-borate, 89mM boric acid, 2mM EDTA pH8.3

TE (1x) 10mM Tris, 1mM EDTA pH8.0

TNE (1x) 100mM NaCl, 10mM Tris, 1mM EDTA pH8.0

PSM (1x) 100mM NaCl, 1mM MgSO₄·7H₂O, 50mM Tris-HCl pH7.5, 0.01% gelatin

L- Broth (1 litre) 10g tryptone, 5g yeast extract, 5g NaCl, 1g glucose (for L-agar add 15g nutrient agar)

LM Broth (1 litre) 10g tryptone, 5g yeast extract, 5g NaCl, 2g MgSO₄·7H₂O (for LM agar add 15g nutrient agar, for LM agarose add 7g agarose)

2xYT (1 litre) 16g tryptone, 10g yeast extract, 5g NaCl

Gel loading buffer (10x) 0.25% bromophenol blue, 0.25% xylene cyanol, 30% glycerol
2.2 Bacterial competence and transformation

_E. coli_ strains HB101 and XL-1 Blue were rendered competent for transformation as described by Hanahan (1983). A single colony was picked into 5ml of L-broth and grown to late log phase (approximately 3 hrs). A 1ml aliquot of this was then used to inoculate 100ml of L-broth pre-warmed to 37°C and allowed to grow to an O.D._550=0.2 (approximately 5x10^7 cells/ml). Cells were harvested by spinning for 5 mins at 1000g (3000rpm), 4°C and washed twice in 50ml of ice cold 0.1M NaCl, 10mM Tris-HCl pH8 before being resuspended in 50ml of ice cold 80mM CaCl2, 10mM Tris-HCl pH7.6. The suspension was allowed to stand on ice for 20 mins and cells were then pelleted by centrifugation and resuspended in 2.5ml of the CaCl2-Tris solution. The final suspension was used immediately or made to 25% glycerol. 200μl aliquots of these competent cells were stored in eppendorf tubes at -70°C. Cells prepared and stored in this way routinely gave transformation efficiencies of greater than 10^8 transformants per μg of supercoiled plasmid for up to 4 months after preparation.

Competent XL1-Blue cells required to prepare the HTF libraries were obtained from Stratagene. The transformation efficiencies of these cells were guaranteed to be greater than 10^6 transformants per μg of supercoiled plasmid.

**Transformation:** Frozen aliquots of competent cells were thawed immediately before transformation and placed on ice. 50ng of plasmid DNA in 10μl of TE was added to each 200μl aliquot and the mixture allowed to stand on ice for 30 mins. The cells were then 'heat-shocked' for 45 seconds at 42°C before the addition of 800μl of L-broth. The cells were then incubated at 37°C for 30mins and 250μl
aliquots of these cells were plated onto L-plates containing 100μg/ml ampicillin. Bacterial colonies were allowed to grow overnight at 37°C.

2.3 Repackaging cosmid DNA

Y-specific Homer V cosmid clones were repackaged before analysis to select against those which may have undergone recombination in vivo. Because of packaging restraints only clones which have full-length inserts are selected.

2.3.1. Testing E.coli strain ED8767 for recA- function

The recA mutation inactivates the generalized recombination systems of the host cell, thereby minimizing recombination between the endogenous DNA and the exogenous added recombinant genomes, as well as minimizing the possibility of recombination between repeated sequences within the inserts which might lead to deletion of intervening sequences.

ED8767 cells were streaked onto an L-agar plate from a glycerol stock and grown overnight at 37°C. Four small colonies were picked and restreaked onto a fresh L-agar plate. At the same time they were restreaked as continuous, parallel streaks across a second L-agar plate. Using an UV opaque material, this second plate was covered. Portions of the plate, and thus regions of the bacterial streaks were then exposed, for known time periods (0”, 10”, 20”, 30”, 45”), to UV light provided by a transilluminator positioned 750cm above the plate.
Following the overnight incubation at 37°C, of both streaked plates, the UV exposed plate was examined for non-growth; mutations in recA prevent repair of UV-induced damage so that irradiated recA- cells do not grow. The colony exhibiting the greatest susceptibility to UV irradiation was selected and a corresponding colony from the unexposed L-agar plate was picked. A 10ml culture of this colony was then grown overnight at 37°C in L-broth containing 2% maltose but no glucose. The following day the cells were harvested by centrifugation at 2500 rpm for 5mins and then resuspended in 5ml of 10mM MgSO4. This cell suspension was then used immediately as a recipient for the packaged cosmid.

2.3.2 Packaging Reaction

Sonicated Extract (SE) and Freeze Thaw Lysate (FTL), prepared by a slightly modified method of that described by Maniatis et al. (1982) were provided by Dr. J. Wolfe.

SE and FTL were removed from -70°C storage and placed in liquid nitrogen to ensure they did not thaw. The reaction components were added in the following order: 7μl of buffer A (20mM Tris.HCl pH8, 1mM EDTA, 5mM MgCl2, 0.05% v/v β-mercaptoethanol) prewarmed to 30°C, 1μl (10ng) of cosmid DNA, 1μl buffer Q (6mM Tris.HCl pH7.5, 18mM MgCl2, 60mM spermidine pH7.5, 15mM ATP pH7.6, 0.2% v/v β-mercaptoethanol). The reaction mixture was then incubated at 30°C for 5mins. 3.5μl of SE and 5.0μl of FTL were then added as quickly as possible, as each thawed, and the incubation continued at 30°C for a further 60mins. Following the addition of 180μl of PSM, 20μl of this diluted packaged material was removed and added to 20μl of resuspended recA- ED8767.
This was then incubated at 30°C for 30mins before 160μl of L-broth was added. Incubation was continued for a further 30mins. The cells were then plated onto L-agar plates containing 100μg/ml ampicillin and allowed to grow overnight at 37°C. Resulting colonies were picked for DNA isolation and analysis.

2.4. DNA Isolation

2.4.1. Preparation of cosmid and plasmid DNA

Small and large scale preparations of plasmid and cosmid DNA were performed by modifications of the procedures described by Maniatis et al. (1982).

a) Maxipreparations: Large scale (250ml) cultures were inoculated from 10ml saturated cultures at a density of 1/100 and were allowed to grow overnight at 37°C with vigorous shaking. Cells were harvested by centrifugation (6000rpm, 10mins, 4°C) and resuspended in 3ml of glucose lysing solution (solution 1; 50mM glucose, 10mM EDTA, 25mM Tris-HCl pH8). 1ml of Lysozyme solution (8mg/ml of solution 1) was added and the mixture left at room temperature for 5mins. 8ml of solution 11 (0.2M NaOH, 1% SDS) was then added, carefully mixed and left on ice for 5mins. This was neutralized by the addition of 6ml of solution 111 (potassium acetate solution pH4.8; 3M with respect to potassium and 5M with respect to acetate) and left on ice for a further 5mins. The mixture was centrifuged at 10,000rpm for 10mins and the resultant supernatant precipitated with 0.6 volumes of isopropanol at room temperature for 30mins. After centrifugation at
10,000rpm for 20mins, the pelleted nucleic acids were resuspended in 10ml of 50mM Tris.HCl pH8, 50mM EDTA. Solid caesium chloride (1g/ml of solution) was then added and allowed to dissolve. The volume was measured and ethidium bromide (0.8ml of a 10mg/ml stock per 10ml of NA solution) was added. The solution was pipetted into 'quick seal' tubes and centrifuged overnight in a vertical rotor at 45K rpm for 17hrs. The banded plasmid or cosmid DMA was visualised under long wave UV light and collected in a 1ml volume. Ethidium bromide was removed by repeated extraction with caesium chloride-saturated isopropanol and the DNA dialysed against several changes of TE before ethanol precipitation.

b) Minipreparations: Plasmid and cosmid DNA was released from 1.5ml of an overnight bacterial culture by the alkaline lysis method described above (with volumes adjusted accordingly). Following isopropanol precipitation the nucleic acid pellet was washed with 70% ethanol, respun, dried under vacuum and redissoved in an appropriate volume of H2O for restriction endonuclease digestion. DNase-free RNase (final concentration 20μg/ml) was added for the final 20mins of these digests, before the reaction was terminated by the addition of 1/10 vol. gel loading buffer.

2.4.2. Preparation of Bacteriophage λ DNA

200ml of prewarmed LM broth was inoculated with 3ml of an overnight culture of LE392 bacteria (grown in LM broth containing 0.2% maltose) and incubated at 37°C until the culture had reached a density of 3-4 x 10⁸ cells/ml (OD₆₀₀=0.3-0.4). 5 x 10⁶ pfu of phage stock were then added and the phage allowed to adsorb onto the cells at 37°C.
for 20mins without shaking. After standing the culture was grown with vigorous shaking for 5-6 hours by which time lysis had occurred and the culture cleared. After continued shaking with 1% chloroform for 20mins, the culture was incubated with 1M NaCl, 1μg/ml DNase and 1μg/ml RNase, at room temperature for 1 hour. Cell debris was removed by centrifugation at 10K rpm for 10mins at 4°C and PEG 6000 added to the supernatant to 10%w/v, to precipitate the phage particles at 4°C overnight. Phage were pelleted by spinning at 10K rpm for 10mins at 4°C and resuspended in 10ml of PSM. The phage were purified by banding in a CsCl₂ gradient at 55,000rpm overnight.

The banded phage particles were visualized and collected in a volume of approximately 2ml. After dialysis at room temperature overnight against 1 litre of 10mM NaCl, 50mM Tris.HCl pH8, 10mM MgCl₂, the phage solution was treated with 20mM EDTA pH8, 50μg/ml proteinase K and 0.5% SDS at 65°C for 1 hour to remove phage protein. Phage DNA was recovered by two phenol/chloroform extractions one chloroform extraction and ethanol precipitation. Finally the DNA was dissolved in TE and the concentration determined.

2.4.3 Preparation of DNA from Genomic and Hybrid Cell Lines

Cell Culture: IRE3, HORL9X, 853, 7/2 and 3E7 cells were grown in tissue culture flasks (75cm²) in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% v/v foetal calf serum (FCS), 3μg/ml L-glutamine, 0.3% w/v NaHCO₃ and 100units of both Penicillin and Streptomycin. HORL9X cells were also grown under HAT selection so 100μM hypoxanthine, 10μM methotrexate and 10μM thymidine were also added to the culture medium. To passage the cells, the medium was
removed and replaced with a solution of 0.25% trypsin in PBS, until the cells were sufficiently free to aspirate off by pipette. They were then re-seeded at the appropriate density in complete medium. 

Oxen and GM1416B cells were cultured in flasks in RPMI 1640 supplemented with 10% v/v FCS, 3μg/ml glutamine, 0.2% w/v NaHCO₃ and 100 units of Penicillin and Streptomycin. For passaging, the cells were pelleted by centrifugation and reseeded by dilution to the appropriate density. All cultures were grown in 37°C humidifying incubators. The cells grown in DMEM were grown in closed systems. Those grown in RPMI were grown under 5% CO₂. All manipulation were performed in sterile flow hoods and the cells routinely monitored under a phase contrast microscope. Cells were stored in 1ml aliquots of 5% v/v dimethyl sulfoxide (DMSO), 95% v/v FCS in freezing vials in liquid nitrogen.

DNA preparation: DNA for Southern analysis was prepared by the isopropanol precipitation procedure described by Hofker et al. (1985) as follows: Approximately 2 x 10⁶ cells were harvested, washed three times in PBSa and digested overnight at 37°C in 20ml of 75 mM NaCl, 25 mM EDTA, pH 8, 1% SDS and 100 μg/ml Proteinase K. DNA was purified away from proteins by two extractions with an equal volume of phenol:CHCl₃:isoamylalcohol (25:24:1). After centrifugation (2500rpm, 10mins), the lower organic phase was discarded. The upper aqueous phase was collected and precipitated with 0.1 vol 3M Na acetate and 1 vol isopropanol. Precipitated nucleic acid was recovered by spooling. It was then washed briefly in 70% ethanol, air dried and dissolved in 10ml of TE for 1 hour at 50°C. After treatment with 50μg/ml RNase (37°C, 1 hour) and 100μg/ml Proteinase K, 1% SDS (37°C, 3 hours), the DNA was extracted once with an equal volume of phenol/chloroform and
once with an equal volume of chloroform alone. The final aqueous phase was precipitated with 0.1 vol 3M Na acetate and 1 vol isopropanol and the DNA was recovered again by spooling. Following a 70% ethanol wash the sample was air dried and dissolved in TE to give an approximate concentration of 500µg/ml. DNA concentration was confirmed by spectrophotometry.

2.4.4 Preparation of DNA from Flow Sorted Chromosomes

Flow sorted human Y chromosomes were provided by Dr. J. Fantes (Fantes et al. 1983). Essentially metaphase chromosomes were isolated from the lymphoblastoid lineOX (49XYYYY) using hypotonic shock. Following a clearing spin to remove cell debris, the chromosomes were resuspended in polyamine buffer (15mM tris, 0.2mM spermine, 0.2mM spermidine, 0.5mM EGTA, 2mM EDTA, 80mM KCl, 20mM NaCl, 14mM β-mercaptoethanol) containing 1mg/ml digitonin and stained with 0.5µg/ml Hoechst 33258 and 2µg/ml chromomycin A3 in 1mM MgCl₂. Samples were analysed on a dual beam fluorescence activated cell sorter (FACS IV) at a flow rate of 100-800 chromosomes per second. Following sorting the sample containing the Y chromosome was dialysed against 10mM Tris.HCl pH8, 1mM EDTA, 1mM EGTA, 100mM NaCl to remove the stain and polyamines. The chromosomes were then concentrated by spinning at 3000 rpm for 20mins at 4°C and resuspended in 400µl of buffer (150mM NaCl, 10mM EDTA, 10mM Tris.HCl pH8). The sample was stored at -20°C.

DNA was isolated from the Y chromosomes as described by Fuscoe et al. (1986). The chromosomes were resuspended in lysis buffer (150mM NaCl, 10mM Tris.HCl pH8, 10mM EDTA, 0.5% SDS, 100µg/ml Proteinase K) and incubated for 6 hours at 37°C. The proteins were then extracted
twice with phenol/chloroform and once with chloroform. Because of the small amount of DNA, each organic phase was re-extracted with a small amount of TE. Aqueous layers were combined and the DNA was precipitated at -20°C by the addition of 0.1 vol of 3M Na acetate and 2 vol ethanol, using 20μg/ml tRNA as carrier. Following centrifugation at 8K rpm at 4°C for 20mins the DNA pellet was washed using 70% ethanol and dried in vacuo. The pellet was resuspended in 25μl of TE and stored at -20°C.

2.5 DNA Analysis

2.5.1 Restriction Enzyme Digestion

Restriction endonuclease digest were carried out according to the suppliers recommendations using buffers supplied with the enzymes. Usually, plasmid and cosmid digests were carried out in a final volume of 20μl using 1-3 units of enzyme per μg of DNA. for 2-3 hours whilst genomic DNA was digested overnight with a 10-fold excess of enzyme.

2.5.2 Restriction Mapping by Partial Digestion

Restriction mapping of cosmids was carried out using a modified method of Smith & Birnstiel (1976). Essentially the cosmid clone was cleaved at a unique site within the vector leaving a 'left hand' and a 'right hand' end. A partial digestion reaction was then carried out and the products separated on a low percentage agarose gel to resolve the larger fragments. Following Southern blotting, the filter was
probed using a DNA fragment which hybridized to either the 'right hand' end or the 'left hand' end of the cosmid clone. The resulting pattern of bands corresponded to a simple overlapping series of fragments with a common labelled terminus; the ascending order of fragments in the gel corresponding directly to the order of restriction sites along the clone. In addition, the products of partial digestion by several different enzymes could be analysed simultaneously on the same gel, so that the relative positions of restriction sites could be obtained from a single autoradiograph.

20μg of Homer V cosmid clone DNA was linearized using Sall and an aliquot removed to confirm digestion by agarose gel electrophoresis. The remainder was subjected to phenol/chloroform extraction and ethanol precipitation. Following microfugation, washing and drying the DNA was resuspended in H₂O to give a concentration of 1μg/μl. The enzymes, whose sites were to be mapped, were diluted to 0.5 units/μl using the recommended buffer and stored on ice. The restriction enzyme reactions were set up using 5μl (5μg) of linearized cosmid DNA, 4μl of 10x recommended buffer, and 30μl of H₂O. 1μl of the appropriate enzyme was then added at reaction time 0. 8μl aliquots of each reaction were then removed at intervals of 2, 5, 10, 15 and 20mins, and each immediately added to the same 100μl aliquot of phenol/chloroform, to quench the reaction. Following the final time point the aqueous phase containing the partial products, was separated from the organic phase by microfugation and the phenol/chloroform layer discarded. Gel loading buffer was then added and the product mixture divided into two 20μl aliquots to run in duplicate on a 0.4% agarose gel. Electrophoresis was performed at 1.5V/cm for 24 hours. The duplicate halves of the gel were then separated and subjected to
Southern blotting. One blot was then probed with the Homer V SalI-HindIII labelled fragment (corresponding to the 'right hand' vector fragment) whilst the second blot was probed with the SalI-PstI labelled fragment (corresponding to the 'left hand' vector fragment). Following autoradiography the series of resulting bands provided information about both ends of the clone.

2.5.3 Gel Electrophoresis

Agarose gels: DNA was separated on agarose gels prepared and run in either 1xTBE or 1xTAE buffer using a flat-bed apparatus. DNA samples were loaded with the addition of 0.1 vol of 10x loading buffer. Electrophoresis was generally performed at 2.5V/cm at room temperature overnight. The gel was then stained using 2µg/ml ethidium bromide for 15mins before the DNAs were visualised by UV transillumination and photographed against a metric ruler. HindIII-cut or BstEII-cut λ DNA was used for markers.

Non-Denaturing Polyacrylamide Gels: These gels were used to analyse the small restriction fragments resulting from the NspI digests of cosmids. 5% polyacrylamide gels were prepared and run as described by Maniatis et al. (1982). These 1mm thick gels were pre-electrophoresed for 30mins prior to use. Radioactively labelled samples were loaded with the addition of 0.1 vol of loading buffer and the gels run at a constant power of 10W (550-700V, 14mA) for approximately 3 hours. Following electrophoresis the gels were dried onto Whatman 3MM paper and autoradiographed overnight.

Denaturing Polyacrylamide Gels: These gels were used to analyse the products from the in vitro transcription reaction used in the
production of radiolabelled RNA probes, and also to analyse the methylation patterns of one of the cosmids clones. 5% or 8% gels were prepared and run as described by Maniatis et al. (198). The gels were pre-electrophoresed for 30 mins prior to use, to warm them. Before loading the samples, 0.1 vol of loading buffer were added and the radioactively labelled samples placed in a boiling water bath for 5 mins to denature the nucleic acids. Electrophoresis was carried out for 1½-2 hours at 11W (for the methylation studies) or 35W (for the RNA studies). Gels required for methylation studies were then prepared for electroblotting (section 2.7). Gels used to analyse radiolabelled RNA were dried onto Whatman 3MM paper and autoradiographed overnight.

Sequencing reactions were run on 6% denaturing polyacrylamide gels using a buffer gradient system of 2.5-0.5xTBE (Biggin et al. 1983). Gels were pre-electrophoresed at 30W to warm the system to approximately 45°C before samples were loaded. Samples were run at approximately 40W to maintain the temperature before the gel was dried onto Whatman 3MM paper ready to be autoradiographed overnight.

2.5.4 Isolation of fragments for Subcloning and Radioactive Labelling

Following electrophoresis, restriction fragments required for subcloning or radioactive labelling were purified from agarose gels by three methods:

Electroelution: The desired band was excised from the gel and placed into prepared dialysis tubing with 0.5ml of sterile 0.2xTBE. Air was excluded from the bag and the bag secured. This was then submerged in an electrophoresis tank also containing 0.2xTBE and a
current passed through the bag (100V for 2-3 hours) to electrophorese the DNA out of the gel into the bag. The polarity of the current was then reversed for 2mins to release the DNA from the wall of the dialysis bag and the DNA solution from within the bag removed. DNA was recovered by direct ethanol precipitiation or by Elutip-d purification.

**Elutip-d purification:** Using the procedure exactly as described by the manufacturer, Schleicher & Schuell Inc., DNA fragments were purified from either low-melting agarose gels or from DNA solutions, following electroelution, using Elutip-d columns. In the final stage of this procedure, DNA was eluted from the column using a high salt (1M NaCl) buffer and DNA recovered by precipitation at -20°C by the addition of 2 vol of ethanol.

**Low-melting agarose gels:** DNA fragments required for multiprime labelling reactions were excised from low-melting agarose gels, following electrophoresis, and transferred to pre-weighed Eppendorf tubes. Sterile H₂O was then added at a ratio of 3ml of H₂O per gm of gel and the tube was placed in a boiling water bath for 7mins to melt the gel and denature the DNA. If the DNA was not required immediately it was stored at -20°C, in aliquots, each containing approximately 50ng of DNA, ready for subsequent multiprime reactions.

### 2.5.5 Subcloning DNA fragments

DNA restriction fragments used for subcloning were purified by electroelution or by Elutip-d purification as described in section 2.5.4. and subcloned into linearized vector DNA, usually pS1EML or
pBluescript (fig. 2.1), with compatible, phosphatase treated cohesive ends.

**Phosphatase reaction:** Linearized vector DNA was treated with calf alkaline phosphatase (CAP) to prevent self-ligation. The phosphatase reaction was carried out in a final volume of 50μl containing 50mM Tris.HCl pH9, 1mM MgCl₂, 0.1mM ZnCl₂ and 1mM spermidine. A total of 2 units of enzyme were added in two aliquots, with 30min incubations at 37°C after each addition. The reaction was stopped by the addition of EDTA to 1mM, SDS to 0.5%, NaCl to 50mM and Tris.HCl pH8 to 10mM and the enzyme heat inactivated by incubation at 68°C for 15mins. Prepared vector DNA was then phenol/chloroform extracted and recovered by ethanol precipitation.

**Ligation reaction:** All ligations were carried out in a final volume of 10μl containing 66mM Tris.HCl pH 7.5, 6.6mM MgCl₂, 0.5mM spermidine, 5mM DTT and 2mM ATP. Ligations generally contained 200ng of vector and a 3-fold molar excess of insert. T4 DNA ligase (1μl, 10units) was added and the reaction mix incubated overnight at 4°C. Ligated recombinant DNA was transformed into *E.coli* HB101 or XL1-Blue as described in section 2.2.

### 2.6 Radioactive labelling of Nucleic Acid

**Labelling of 3'-ends with 5'-overhangs:** This method was used to label restriction enzyme sites where cleavage generated a single-stranded 5' overhang of DNA, and where a dCTP is amongst the residues required to "fill-in" and generate the blunt end. *MspI* sites are an example of this type.
0.5µg of cosmid clone DNA was cleaved using *MspI* in a low salt restriction buffer, in a total volume of 10µl. 0.04µCi [α-3²P] dCTP and 1 unit of DNA polymerase 1 'Klenow' fragment was then added to the digest and incubated at room temperature for 15mins. The reaction was terminated by the addition of 0.1 vol of gel loading buffer.

**Multiprime labelling of DNA probes:** A commercial random-prime labelling kit (Amersham International plc) was used to generate labelled DNA probes. This is essentially a modified version of Feinberg & Vogelstein (1983). Approximately 50ng of DNA in 29µl of H₂O was placed in a boiling water bath for 2mins to denature the DNA. This was then chilled briefly on ice. The reaction components were then added in the following order: 10µl of Buffer 1 (containing unlabeled dATP, dGTP,dTTP and the reaction buffer), 5µl of Buffer 2 (containing the random hexanucleotide primers), 4µl [α-3²P] dCTP (10µCi/µl) and 1µl (1 unit) DNA polymerase 1 'Klenow' fragment. The reaction was incubated at room temperature overnight and then terminated by the addition of 50µl TNE containing 1% SDS. Unincorporated nucleotides were removed by passage through a Sephadex G50 column as described in section 2.1. Before addition to the hybridizing solution the probe was boiled for 5mins to denature the DNA and chilled quickly on ice.

**Preparation of high-specific activity labelled RNA probes:** T7 RNA polymerase was used to generate high specific activity radiolabeled RNA probes from the *Hpall* tiny fragments cloned into pBluescript. All solutions involved in this reaction were freshly prepared and treated with 0.1%v/v diethylpyrocarbonate to inhibit RNase activity.

The reaction components were added in the following order: 5.2µl H₂O, 5µl 5x transcription buffer (200mM Tris.HCl pH8,40mM MgCl₂,10mM
spermidine, 250mm NaCl), 4.8μl ATP/CTP/GTP mix (10mM each), 1μl (1μg) linearized DNA template, 2μl (250mM) DTT, 1μl (1unit) RNaseIn, 5μl [α-32P] UTP (40mCi/ml; 800Ci/mmol) and 1μl (10units) T7 RNA polymerase. Following incubation at 37°C for 30mins the nucleic acids were phenol extracted and ethanol precipitated at -70°C for 2 hours. They were recovered by microfugation, washed in 70% ethanol and dried in vacuo. The pellets were then redissolved in 10μl of buffer (10mM Tris.HCl pH7.5, 0.1mM EDTA). A 1μl aliquot of this was diluted in 100μl of buffer. 1μl of the diluted solution was then loaded onto a 5% denaturing polyacrylamide gel to confirm the size and integrity of the transcript, whilst a second 1μl aliquot was used to determine 32P UTP incorporation by Cherenkov counting.

2.7 Southern Blot Analysis of DNA

For Southern analysis (Southern 1975), DNA was transferred onto either GeneScreen Plus or Hybond-N membranes by capillary blotting (from agarose gels) or by electro-blotting (from denaturing polyacrylamide gels).

Capillary blotting: Following depurination in 0.25M HCl, denaturation in 0.5M NaOH, 1.5M NaCl and neutralization in 1.5M NaCl, 0.5M Tris.HCl pH 7.5, 1mM EDTA, DNA was transferred in 10x SSC for GeneScreen Plus or 20x SSC for Hybond-N, overnight. After transfer GeneScreen filters were treated as recommended by the suppliers, with 0.4M NaOH for 45secs and then with 2x SSC, 0.2M Tris.HCl pH7.5 for 5mins, before air drying. Hybond-N filters were air dried and then UV irradiated for 5mins on a transilluminator.
**Electro-blotting:** Following denaturing polyacrylamide gel electrophoresis the gel was blotted onto Whatman 3MM paper and trimmed to fit the electroblot apparatus. The gel was then overlaid with a Hybond-N filter and this arrangement was fitted into the apparatus cassette and held in place by thin sponges and layers of Whatman 3MM, all soaked in 1xTBE. This complete cassette was then lowered into the tank of the electroblotter which contained 1xTBE. Transfer was accomplished by passing a current of 0.25A across the gel for 1 hour and then increasing the current to 0.35A for the second hour. The Hybond-N filter was then removed and floated in 0.5M NaOH, 3M NaCl for 5mins, and then neutralised for 10mins in 3M NaCl, 0.5M Tris.HCl pH7. Finally the filter was air dried and UV irradiated for 5mins.

**Hybridization of DNA probes:** GeneScreen filters were prehybridized in 1M NaCl, 1% SDS, 10% dextran sulphate whilst Hybond-N filters were prehybridized in 6x SSC, 0.5% SDS, 5x Denhardt's solution and 100µg/ml sonicated herring sperm DNA. In both cases prehybridization occurred overnight at 65°C. Denatured ³²P labelled probe (final concentration 10ng/ml) was added directly to both types of filter. However, when it was known that the probe contained highly repeated sequences, these were 'removed' by pre-associating the probe with a large excess of sonicated human and mouse DNA thus allowing signals from single or low copy number components to be detected (Sealey et al.1985). Following the passage of the labelled DNA through the G50 Sephadex column, 50µl (10mg/ml) sonicated human DNA, 50µl (10mg/ml) sonicated mouse DNA and 60µl of 20x SSC were added. This mixture was then placed in a boiling water bath for 10mins to denature all the DNAs, before being incubated at 65°C for 1 hour.
This 'competed' probe was then added directly to the filter. In all cases hybridization was allowed to proceed at 65°C for 16-24 hours.

**Hybridization of RNA probes:** Following UV treatment of Hybond-N filters, the blots were prehybridized at 42°C for 17 hours in 50% deionized formamide, 5x SSC, 5X Denhardt's solution, 50mM sodium phosphate buffer pH6.5, 0.1% SDS, 10µg/ml poly C, 10µg/ml poly A, 250mg/ml sonicated herring sperm DNA, 10mM Ribonucleoside-Vanadyl complex (Cross & Little 1987). The RNA probe was added directly to this hybridization solution at a concentration of 10⁶cpm/ml and hybridization allowed to proceed at 42°C overnight.

Excess, non-specifically bound probe was removed by washing the filters in 2x SSC for 5mins at room temperature. This wash was repeated and followed by a series of washes: 2x SSC, 1% SDS at room temperature for 20 mins, 0.1x SSC, 0.1% SDS at room temperature for 20 mins and finally 0.1X SSC, 0.1% SDS at 65°C for 15 mins. After washing, filters were blotted and exposed to X-ray film at -70°C with an intensifying screen.

---

2.8 Construction of chromosome-specific *Hpall* Libraries

2.8.1 Library construction using DNA from flow-sorted chromosomes

DNA was isolated as described in section 2.4.4. Following recovery of the DNA by ethanol precipitation, the DNA was redissolved in 17µl of H₂O and cleaved with *Mspl* (an isoschizomer of *Hpall*) in a 20µl reaction containing 10 units of enzyme. The reaction was incubated at 37°C for 6 hours before 0.1 vol of gel loading buffer
were added and the DNA separated by electrophoresis through a 1xTBE/1.5% agarose gel run overnight at 30V. *Mspl* cleaved pBR322 was also loaded onto this gel to provide DNA size markers. After ethidium bromide staining the DNA was visualized under UV light. As expected, due to the very small quantity of DNA present, no DNA was visible in the sample track. However, a gel slice was excised corresponding to the position of the *Mspl* tiny fragments (50bp-250bp), as judged by the pBR322 size markers, and the DNA isolated by electroelution and Elutip purification (section 2.5.4). Following ethanol precipitation, the DNA was redissolved in 4µl of H₂O and stored at -20°C.

The pBluescript (pBS; fig. 2.1) vector DNA was linearized with the enzyme *Clal* to generate ends which are compatible to those produced by *Mspl*, and phosphatased as described in section 2.5.5. 1µg of this vector DNA was then ligated to the *Mspl* insert DNA in a 10µl reaction (see section 2.5.5), containing the complete 4µl aliquot of insert DNA. The following day a 1µl aliquot of this reaction was used to transform a 100µl aliquot of competent XL1-Blue bacteria (section 2.2) and the transformed cells were plated onto L-agar plates containing 100µg/ml ampicillin, 0.5mM IPTG and 0.01% w/v X-Gal. After overnight incubation at 37°C, blue and white colonies were counted and a random selection of each were picked and grown as small L-broth cultures. These were then 'minipreped' (section 2.4.1b) and the DNA cleaved to release the insert. Following agarose gel electrophoresis, six clones which appeared to have small inserts were selected for further analysis. Glycerol stocks were prepared from the remaining clones, and stored at -70°C.
FIG. 2.1 Restriction map of the Bluescript vector showing the multiple cloning site in detail.
2.8.2 Library construction using DNA from a chromosome-specific \( \lambda \) Library

A Y-specific and a chromosome 9-specific Charon 21A library were obtained from the American Type Culture Collection. These libraries were constructed using DNA, derived from flow sorted human chromosomes, cloned into the \textit{HindIII} site of the \( \lambda \) vector Charon 21A. The average insert size of each library was reported to be 4kb and the yield of independent recombinants for the Y-specific library was given as \( 2.5 \times 10^8 \), and that for the chromosome 9 library as \( 3 \times 10^8 \).

Titration of recombinant phage: The titre of the library was verified using a LE392 plaque assay. A culture of LE392 bacteria was grown overnight in 10ml L-broth + 0.2% maltose. This was then spun down at 1.5K rpm for 5mins and the pelleted cells resuspended in 0.4 vol. of PSM. 200µl of these resuspended cells were then added to approximately 3ml warm LM-agarose and the mixture was poured onto a small LM-agar plate and allowed to set. 10-fold serial dilutions of the phage stock were then prepared in PSM. 10µl aliquots of each dilution were then spotted onto the prepared LE392/LM-agar plate. This was incubated overnight at 37°C and the number of plaque forming units counted and the overall titre of the library calculated.

Amplification of the library: A culture of LE392 bacteria was grown overnight in L-broth + 0.2% maltose. 2ml of this was then spun down at 1.5K rpm for 5mins and the pelleted cells resuspended in 100µl of PSM. 10µl (\( 3 \times 10^8 \) pfu) of the phage stock were added and the suspension left at 37°C for 20mins to allow the phage to adsorb onto the bacteria. 15ml of warm LM-agarose was then added and the mixture poured onto a large (20cm x 20cm) LM-agar plate. This was then
incubated overnight at 37°C by which time the majority of the bacteria had lysed. 15ml of PSM was then poured onto the plate and this was left at 4°C, overnight with gentle agitation. The PSM was removed and spun at 4K rpm for 10mins to remove cell and agar debris and the supernatant stored in a glass bijou with a drop of chloroform, to prevent bacterial growth. The library was then re-titred as described above.

**DNA isolation and cloning:** DNA was isolated from the library as described in section 2.4.2. The DNA concentration was determined spectrophotometrically and using various enzymes, it was verified that the DNA could be cleaved. 50μg of the library was then cleaved with HindIII to release the inserts from the vector, in a 100μl reaction containing 100units of enzyme. 10μl of gel loading buffer was then added and the DNA fragments separated by electrophoresis on a 1xTBE/0.7% agarose gel run overnight at 30V. Following ethidium bromide staining, the DNA was visualised under UV light and the insert DNA, seen as a smear within the track, was excised, leaving the vector arms at the top of the gel. The insert DNA was then isolated by electroelution and Elutip purification (section 2.5.4). Following ethanol precipitation, the insert DNA was redissolved in 20μl of H2O. A 3μl aliquot was removed to confirm the presence of the DNA by agarose gel electrophoresis and the remainder was cleaved by HpalI in a 20μl reaction containing 10 units of enzyme, and incubated at 37°C for 6 hours. Following phenol/chloroform extraction and ethanol precipitation, the DNA was redissolved in 4μl of H2O. The procedures used in generating the pBS vector and the ligation and transformation reactions used, are similar to those described above (section 2.8.1).
Likewise, six clones which, following miniprep analysis, appeared to have inserts, were selected and analysed further.

2.9 Screening Bacteriophage \( \lambda \) Libraries.

A \( \lambda \text{gt}11 \) foetal testis cDNA library, constructed by Drs. Y. Edwards and J. Wolfe, and a \( \lambda \text{gt}11 \) gut cDNA library, constructed by Dr. Y. Edwards, were screened, using putative CpG island containing clones in order to identify expressed sequences.

Using methods described in section 2.8, each library was titred to confirm the numbers of phage present. A representative portion of each library (2.5 \( \times \) \( 10^8 \) pfu/plate) was then plated onto large (20cm x 20cm) LM-agar plates (as described in section 2.8) and the plates incubated at 37°C overnight. Filter lifts were prepared by the method of Benton & Davies (1977). Two successive lifts were taken from each plate; the Hybond-\( \text{N} \) filter (20cm x 20cm) was laid onto the cold library plate and the phage allowed to adsorb for 1min (first lift) or 2mins (second lift). Orientation marks were made in identical positions on both the filters, corresponding to reference marks on the plate. The filters were transfered to denaturing buffer (0.5M NaOH, 1.5M NaCl) for 2mins, neutralization buffer (1.5M NaCl, 0.5M Tris.HCl pH8) for 5-10mins, rinsed in 4x SSC and allowed to air dry. Following a 5min UV treatment, the filters were prehybridized, hybridized, washed and autoradiographed as described in section 2.7.
**2.10a Northern Blot Analysis**

Northern blots were prepared by a method described by Maniatis et al (1982). A 1.5% agarose gel was prepared by melting the agarose in water and allowing it to cool to 60°C. Gel running buffer (5x running buffer: 0.2M morpholinopropanesulphonic acid (MOPS) pH7, 50mM sodium acetate, 5mM EDTA) and formaldehyde were then added to give final concentrations of 1x and 2.2M respectively. The gel was then allowed to set in the fume cupboard.

The RNA samples were prepared by adding 10µg (4.5µl) of each RNA sample to 2.0µl of 5x running buffer, 3.5µl formaldehyde and 10.0µl formamide. The samples were then incubated at 55°C for 15 mins. Finally 2.0µl of loading buffer (50% glycerol, 1mM EDTA, 0.4% Bromophenol Blue, 0.4% Xylene Cyanol and 5µg/ml ethidium bromide) were added and the samples loaded. The gels were run in 1x running buffer at 5V/cm for approximately 8 hours.

Following electrophoresis the gels were rinsed in DEPC-treated water and 20xSSC for 15 mins each. The RNA was then capillary blotted onto Hybond-N in 20xSSC overnight. The filters were finally rinsed briefly in 50mM sodium phosphate (pH7) before being baked at 80°C for 1 hour.

The filters were prehybridized in 7% SDS, 0.5M sodium phosphate pH7, 1mM EDTA at 65°C overnight. The following day this was replaced with fresh buffer containing the radiolabelled DNA probe. Hybridization was allowed to proceed at 65°C overnight. Excess, unbound probe was removed by a series of washes: 2X SSC, 0.1% SDS at room temperature for 5 mins repeated twice; 0.1x SSC, 0.1% SDS at 50°C for 10 mins. The filters were finally blotted and autoradiographed using an intensifying screen.
2.10 Northern Blot Analysis

Northern blots of adult testis, heart, liver and muscle RNA were provided by Dr. Y Edwards.

The filters were prehybridized overnight at 42°C in 10ml of prehybridization buffer (5x SSPE, 0.3% SDS, 5x Denhardt's solution, 50% v/v formamide, 250μg/ml sonicated herring sperm DNA). The following day this was replaced by 10ml of fresh buffer containing the radiolabelled DNA probe. Hybridization was allowed to proceed for 24 hours at 42°C with gentle agitation. Excess, unbound probe was removed by a series of washes: 2x SSC, 0.1% SDS at room temperature for 5mins repeated twice; 0.1x SSC, 0.1% SDS at 50°C for 10mins. The filter was finally blotted and autoradiographed using an intensifying screen.

2.11 Nucleotide Sequence Analysis

A commercial kit, supplied by USB Corporation, was used for DNA sequencing. The method employed involves a modified version of the chain-termination sequencing reactions of Sanger et al. (1977), using ssDNA as the template and a modified bacteriophage T7 DNA polymerase known as 'Sequenase' (Tabor & Richardson 1987).

Once the ssDNA had been primed, an initial labelling reaction elongated the DNA from this primer, using limiting concentrations of dGTP, dTTP and dCTP and radioactively labelled dATP. In the second step, the concentrations of all the dNTPs was increased and a specific ddNTP was added, one for each of the four nucleotides sequenced. The reactions were terminated by the addition of EDTA and formamide,
denatured by heating and run on denaturing polyacrylamide gels. The advantage of using this enzyme is its high processivity and low 3' to 5' exonuclease activity.

**Preparation of ssDNA:** This method uses ssDNA as the template for the elongation reactions. DNA fragments to be sequenced were therefore subcloned into the vector M13mp19 to generate ssDNA templates (Messing & Vieira 1983). M13mp19 vector DNA was digested with the blunt end restriction enzyme *SmaI*, and phosphatased as described (section 2.5.5). Insert DNA was isolated by gel electrophoresis and purified by electroelution (section 2.5.4). In each ligation 20ng of prepared vector DNA was ligated to 100ng of purified insert. The resulting mixture was then transformed into the host strain *E.coli* JM101, made competent by the CaCl$_2$ procedure (section 2.2) and plated with 0.8mM IPTG and 0.02% w/v X-Gal. Under these conditions recombinant and non-recombinant phage can be distinguished due to insertional inactivation of the β-gal gene of M13mp19; recombinants appear as white plaques on the chromogenic substrate X-Gal, following overnight incubation at 37°C.

Single-stranded template was prepared from 1.5ml cultures of single white plaques picked into 1:100 dilutions of *E.coli* JM101 and grown for 4-5 hours. The cells were removed by centrifugation and the phage precipitated in 4% polyethylene glycol (PEG-800) and 0.5M sodium acetate (pH7) for 15mins at 4°C. Following centrifugation the protein coats of the particles were removed by phenol extraction and the ssDNA recovered by ethanol precipitation.

**Priming reaction:** 1-2μg ssDNA were annealed to 2ng of M13 universal primer in 10μl of 'Sequenase' buffer (40mM Tris.HCl pH7.5, 20mM MgCl$_2$, 50mM NaCl) at 65°C for 2mins. The temperature of the water
bath was then allowed to fall to 35°C when annealing was considered to be complete.

**Labelling reaction:** The labelling mix (7.5μM dGTP, 7.5μM dCTP, 7.5μM dTTP) was diluted 5-fold with H₂O, and the enzyme diluted 1:8 with ice-cold TE buffer. The following were then added to the 10μl annealing reaction: 1μl DTT (0.1M), 2μl diluted label mix, 0.5μl [α-³²S]dATP (10μci/μl, 1000Ci/mmol) and 2μl diluted 'Sequenase'. This was incubated at room temperature for 5-10mins before being divided between four Eppendorfs.

**Termination reaction:** 2.5μl of each termination mix (80μM dGTP, 80μM dATP, 80μM dCTP, 80μM dTTP, 50mM NaCl and 8μM of the appropriate ddNTP) were placed into the appropriate tubes and prewarmed at 37°C for at least 1min. 3.5μl of the labelling reaction was transferred to each tube and this was then returned to the 37°C water bath for a further 5mins. The reaction was terminated by the addition of 4μl of stop solution (95% formamide, 20mM EDTA, 0.05% Bromophenol Blue, 0.05% Xylene Cyanol) and the samples placed in a boiling water bath for 2-5mins before 2μl of each were immediately loaded onto an 6% denaturing polyacrylamide gel (section 2.5.3).
Cell Lines

Cell lines were provided by Dr. J. Wolfe:

**3E7**  
Mouse/Human hybrid containing human Y chromosome as the only cytologically detectable human component.  
*Marcus et al.* (1976)

**853**  
Hamster/Human hybrid containing human Y chromosome as the only cytologically detectable human component

**OXEN**  
49, XYYYY  

**HORL9X**  
Mouse/Human hybrid containing human X as the only cytologically detectable human component

**GM1416B**  
48, XXXX

**1RE3**  
A mycoplasma-free mouse cell line derived by Alan Tonnacliffe from 1R (*Nabholz et al.* 1969)

**a23**  
Hamster cell line

**7/2**  
Mouse revertant of 3E7; human Y chromosome lost

Hybrid cell lines provided by Dr. C. Pritchard:

**E1P4.10, E2P3.5, IP2.1, IP2.2, IP2.6, JP5.3, K1P4.2, K2P2.4,**  
**K3P2.2**
Bacterial strains

JM101  \( \Delta \text{lacpro}, \text{thi}, \text{supE}, F', \text{traD36}, \text{lacI}^\Delta \text{M15} \)
(Messing 1983)

HB101  \( F^-, \text{hsdS}20(\text{r}^-, \text{m}_{\text{S}_{\text{S}}}^-), \text{recA}13, \text{ara}14, \text{proA}2, \text{lac}Y1, \text{galK}2, \text{rpsL}20(\text{Sm}^r), \text{xyl}5, \text{mt}11, \text{sup}E44, \lambda^- \)

XL-1  \( \text{endA}1, \text{hsdR}17, (\text{r}^-, \text{m}_{\text{S}_{\text{S}}}^+), \text{sup}E44, \text{thi}^- 1, \lambda^-, \text{recA}1, \text{gyr}A96, \text{relA}1, \Delta\text{m} \text{lac}, [F', \text{proAB}, \text{lacI}^\Delta \text{M15}, \text{Tn}10(\text{tet}^r)] \)

LE392  \( F^-, \text{hsd}(\text{r}^-, \text{m}_{\text{S}_{\text{S}}}^-), \text{sup}E44, \text{sup}F58, \text{lac}Y1, \text{galK}2, \text{galT}22, \text{met}B1, \text{trp}R55, \lambda^- \)

Y1088  \( \Delta \text{lacU}169, \text{supF}, \text{supF}, \text{hsd}R^-, \text{hsd}M^-, \text{met}B, \text{trp}R \), \( \text{ton}A21, \text{proC}: \text{Tn}5, \text{pMC9} \)  (Young & Davis 1983)

Chemicals, Enzymes and DNAs

Supplier: BDH Chemicals Ltd., Poole, UK.

Boric acid
Caesium chloride
Chloroform
Diaminoethanetetra-acetic acid, disodium salt (EDTA)
Dimethyl sulphoxide (DMSO)
Disodium hydrogen orthophosphate
Formamide
Glucose
Glycerol
Hydrochloric acid (HCl)
Isoamyl alcohol
Isopropanol
Magnesium chloride (MgCl₂)
Magnesium sulphate (MgSO₄)
Maltose
Sodium acetate
Sodium chloride (NaCl)
Sodium citrate
Sodium hydroxide (NaOH)
Sucrose
Phenol
Polyethylene glycol 6000

**Supplier:** Sigma Chemicals, Poole, UK.

Ampicillin
Deoxyribonuclease (DNase)
Dimethylformamide
Ethidium bromide
Herring Sperm DNA
Lysozyme
Sodium dodecyl sulphate (SDS)
Ribonuclease
Tris. Hydrochloride

**Supplier:** Gibco Europe, Glasgow, Scotland, UK

Foetal Calf Serum
Supplier: Flow Laboratories Ltd., Rickmansworth, UK
Dulbecco's modified Eagle's medium (DMEM)
RPMI
Glutamine
Penicillin/Streptomycin

Supplier: Pharmacia P-L Biochemicals, Milton Keynes, UK.
Dextran sulphate
Ribonucleotides
Sephadex G50

Supplier: James Burrough, Milton Keynes, UK.
Absolute ethanol

Supplier: Boehringer Mannheim, West Germany.
Bovine Serum Albumin
Isopropylthiogalactoside (IPTG)
Nucleotide triphosphates
Proteinase K
Restriction endonucleases
T4 Ligase
5-bromo-4-chloro-3-indoly-β-D-galactoside (X-gal)

Supplier: Amersham International plc, Amersham, UK.
Cytidine 5'-[α-32P] triphosphate, triethanolamine salt (3000Ci/mmol)
Uridine 5'-[α-32P] triphosphate, triethylammonium salt (800Ci/mmol)
[α-35S] dATPαS in stabilized aqueous solution (1000Ci/mmol)
Random labelling kit
Supplier: Anglian Biotechnology Ltd., Colchester, Essex, UK.

Restriction endonucleases

Supplier: New England Biolabs Inc., MA 01915, USA.

Restriction endonucleases
Ribonucleoside-vanadyl complex
T7 RNA Polymerase

Supplier: Bethesda Research Laboratories, USA.

Restriction endonucleases

Supplier: Difco Laboratories, Detroit, Michigan, USA.

Bactoagar
Tryptone
Yeast Extract

Supplier: Seakem, Miles Laboratories, Slough, UK.

Agarose
Low gelling temperature agarose
3.1 Identification of CpG rich cosmid clones

The initial study to find CpG islands on the human Y chromosome involved characterizing Y-specific cosmid clones by restriction analysis. The Y-specific cosmid clones were originally isolated from a library constructed in the vector Homer V, using the human-mouse hybrid 3E7 as the source of human Y chromosomal material (Wolfe et al. 1984).

The restriction endonuclease *Mspl*, an isoschizomer of *Hpall*, cleaves the sequence C+CGG regardless of the methylation state of the internal cytosine and therefore gives a more defined estimate of the number of CCGG sites, than *Hpall* which does not cleave if the internal C residue is methylated. By end-labelling *Mspl* digests of each cosmid clone and separating the fragments on an 8% polyacrylamide gel, it was possible to identify non-vector fragments of between 100 and 200bp (HTF), following autoradiography (Fig. 3.1).

Initially, small scale DNA samples were prepared from each cosmid clone and this DNA repackaged to select against those which may have undergone recombination *in vivo*. Because of packaging restraints only clones which have full-length inserts were selected. DNA from each of these repackaged clones was then prepared by the maxipreparation method (see Section 2.4.1) and caesium chloride gradient centrifugation. From the 220 Y-specific clones available, DNA from 150 was prepared in this way. These were then screened using the *Mspl/Hpall* restriction analysis indicated above.

Fig 3.2 illustrates the distribution of non-vector *Mspl*
FIG. 3.1 *Mspl* digests of 20 Y-linked Homer V Cosmid Clones

*Mspl* digests of cY019, cY022, cY24 and cY113 are shown following end-labeling and resolution of the fragments on 5% non-denaturing polyacrylamide gel electrophoresis.

V - Homer V vector
FIG. 3.2 Distribution of non-vector EcoRI fragments within cosmid clones

A. Fragments larger than 300bp
B. Fragments smaller than 300bp
fragments greater than 300bp (A) and less than 300bp (B), throughout the 150 cosmid clones. As shown, the distribution falls into two main groups; the vast majority of clones containing either a small (0-9) number of large fragments (greater than 300bp), or a small number of small fragments (less than 300bp) fragments. A minor group of clones contains a large number (8-24) of \textit{Mspl} fragments; one subset of this group containing a large number of large \textit{Mspl} fragments whilst a second subset contains a large number of small \textit{Mspl} fragments. From this analysis, clones found in the latter subset were chosen for further analysis on the basis that numerous fragments (10-11), smaller than 300bp, were generated by \textit{Mspl} cleavage. These clones were cY019, cY022, cY24 and cY113.

3.2 Analysis of CpG rich clones

3.2.1 Restriction analysis

Restriction maps of the four clones were compiled from data generated by the partial digest method described in Section 2.5. Southern analysis of restriction digests of each clone were also probed with either total human or total mouse DNA, to show regions containing sequences of highly repeated human and mouse DNA. Restriction fragments which did not contain repeated sequences were then isolated in low-melting point agarose (see Section 2.5.4.) and used in subsequent Southern blot analyses, as probes against single and double digests of the entire cosmid clone, to define each map further. Fragments 019P1, 019P3, 019P10, 113P4, 24E7 022E1, and 022E4 were also subcloned into the vector pBluescript (see Section 2.5.5).
As the restriction maps illustrate (fig. 3.3), Mspl sites are clustered in only one (cY019) out of the four chosen clones. However, additional information regarding the CpG character of the four cosmid clones, was obtained from a series of restriction enzyme digests using enzymes whose recognition sites are CpG rich. Fig 3.4 shows this series of single (PstI) and double digests (PstI plus a 'CG' enzyme).

The 5.6Kb PstI fragment of cY022 contains sites for NunII (GGiCGCC), NotI (GCiCGCCGC), Nael (GCCiGGC) and EagI (CiGGCCG). The 6.7Kb PstI fragment of cY24 contains sites for NunII and Nael (GCCiGGC) and the 8.2Kb fragment of cY113 has sites for NunII, Nael, Nrul (TCGIAGA) and MluI (AICGCCT). The 5.5kb PstI fragment of cY019 contains sites for NunII and whilst the 5.7kb PstI fragment of this cosmid contains sites for NunII and BssHII (GICGCCT).

As revealed by restriction mapping, the 5.7kb PstI fragment of cY019 contains ten Mspl sites (8 Mspl/Hpall tiny fragments) and is therefore a region of high CpG density. To investigate this further, a subclone of this PstI fragment (p019P3) was used to study the occurrence of CG enzyme sites in this region in more detail. Fig. 3.5 illustrates that, in addition to BssHII and NunII (from fig. 3.4), NarI (GGiCGCC) and Smal (CCCIAGG) cleave this region of cY019. Only two fragments (3.7kb and 1.85kb) result from Smal cleavage indicating the existence of only one Smal recognition site within this region. The NarI and EagI digests shown in fig. 3.5 appear to be partial digests (the 019P3 band is still visible in both tracks) although the results indicate that both enzymes have at least two recognition sites within this region. Partial digestion with EagI gives rise to three fragments of sizes 5.2kb, 3.5kb and 1.7kb. It is probable that cleavage at one of the sites gives rise to the 5.2kb fragment and a
FIG. 3.4 Single and double restriction enzyme digests of cY019, cY022, cY24 and cY113 using Fstl and various 'CG' enzymes resolved on 0.8% agarose gels.

Lanes: 1) Pstl  2) NruI  3) NotI
4) Nael  5) EagI  6) BssHII
7) Nrul  8) MluI  9) SaccI

-105-
The cY019 subclone p019P3 was subjected to double restriction enzyme cleavage using Fstl and various CG enzymes. The fragments were separated on a 0.8% agarose gel.

Lanes:
1) p019P3 DNA cleaved with Fstl and the 'CG' enzyme shown
2) PBS vector DNA cleaved with Fstl and the 'CG' enzyme shown
0.5kb fragment which is not shown on this gel, whilst cleavage at an 
*Eagl* site within the 5.2kb fragment would give rise to the 3.5kb and 
1.7kb fragments. Similarly partial digestion with *Narl* gives rise to 
four fragments of sizes 3.7kb, 3.35kb, 2.0kb and 1.7kb. Cleavage at 
one site would give rise to the 3.7kb and 2.0kb fragments, whilst 
cleavage at a *Narl* site within the 3.7kb fragment would give rise to 
the 3.35kb fragment and a 0.35kb fragment (not shown on this gel) and 
cleavage at a *Narl* site within the 2.0kb fragment would give rise to 
the 1.7kb fragment and a 0.3kb fragment (also not shown on this gel). 
Although this accounts for the bands seen in these digests it does not 
explain why a more complex series of bands is not seen due to the 
presence of additional partial products. In both cases however, there 
may be preferential cleavage at certain sites and that once these have 
been cut, cleavage at secondary sites is then possible.

From these results it became apparent that although cY113, cY022 
and cY24 contain greater numbers of *Mspl* sites and thus CpGs, relative 
to other cosmid clones, these CpG pairs are not clustered, but 
dispersed throughout the clones. They therefore do not contain 
island-like regions. In contrast, cY019 contains a 5.7kb region in 
which *Mspl* sites are clustered and in which several CG enzymes cleave 
the DNA. Therefore, although further analysis of cY113, cY022 and 
cY24 was undertaken, studies of cY019 were carried out in greater 
detail.
3.2.2. Southern analysis

a) Chromosomal localisation

The four cosmid clones have been shown to be Y-linked through Southern analysis. Figs. 3.6, 3.7, 3.8 and 3.11 show a panel of genomic and hybrid DNAs probed with either cY019, cY022, cY113 or cY24.

cY113 sequences show homology to Y chromosome, X chromosome, autosomal and mouse sequences. Fig. 3.6 reveals cY113 homology to a major 2.3kb autosomal *EcoRl* fragment (band A, tracks 1, 2, 4 and 6), a major 9.8kb X-linked fragment (band X, tracks 1, 2, 4, 6 and 7) and a 4.5kb mouse fragment (band M, tracks 3, 5, 7 and 8). In addition, cY113 hybridization to a variable fragment (band X/Ya) in all tracks except 7/2 and IRE3, implies the presence of an homologous *EcoRl* fragment present on both the human X and Y chromosomes (it is possible that such positional differences are due to loading differences). cY113 also hybridizes to single 5.1kb fragment in GM1416B DNA. (The additional fragments revealed in 7/2 DNA are thought to be plasmid contamination, as are the 11.2kb and 9.2kb fragments in the 3E7 track, also shown in fig.3.7 Arrowed)

Fig. 3.7 shows the hybridization pattern of cY24 to genomic and hybrid cell DNAs following Southern analysis. cY24 shows homology to three Y-linked *EcoRl* fragments (6.4Kb, 5.9Kb and 5.1Kb; bands a,b and c tracks 1, 2 and 3). Although a hybridization signal is not evident in the OXEN track of Fig. 3.7, the original autoradiograph shows a very weak signal indicating homology to a 5.9Kb *EcoRl* fragment. The faint 3.0Kb band (A) present in tracks 1 and 2 appears as a very faint band in track 3 again implying homology to
FIG. 3.8 Southern Analysis using cY022 as a probe

cY022 was used as a probe against a panel of EcoRI digested genomic and hybrid DNAs following prehybridization of the labelled cosmid clone with total human and mouse DNAs to 'eliminate' highly repeated sequences.

Exposure time - 10 days.

Lanes: 1) 3E7  2) 853  3) Male Placenta
        4) OXEN  5) 7/2  6) a23
        7) HORL 9X  8) GM1416B  9) 1RE3
FIG. 3.9 Southern Analysis of Mouse DNA using cY022 as a probe

cY022 was used as a probe (following prehybridization with total human and mouse DNAs to 'eliminate' highly repeated sequences), against EcoRI or PstI digests of mouse DNAs.

Exposure time - 10 days.

Lanes: 1) Male Mouse 2) Female Mouse 3) XXSxr
Y-linked sequences. However as the signal is so weak the precise origin of cY24 is difficult to prove.

The hybridization of cY022 to genomic and hybrid DNAs also reveals a complicated pattern. This cosmid clone is also derived from Y-linked sequences. Fig. 3.8 shows cY022 hybridizes to at least four EcoRI fragments (10.1Kb, 9.3Kb, 6.4Kb and 5.1Kb) within human Y containing DNAs (tracks 1-4), although fragment size differences are revealed (eg. an addition 11.0Kb fragment is present in 3E7 DNA). In addition, cY022 also shows cross hybridization to autosomal sequences (a 4.5Kb fragment-band A in GM1416B, placenta and OX tracks) and to mouse sequences (a 11.8kb fragment-band M in 3E7, 7/2, HORL 9X and IRE3 tracks). In fact, further Southern analysis shows that cY022 hybridizes to 2 EcoRI fragments (11.8kb and 1.8kb) or 4 PstI fragments (2.0kb, 1.3kb and 2 fragments less than 1kb) within mouse DNA digests, although no differences between male, female or XXSxr DNAs are revealed (Fig. 3.9). This raised the possibility that cY022 contained sequences which showed cross species homology and were therefore conserved. However as the cosmid library was prepared from a mouse/human hybrid (3E7) which was known to contain a re-arranged human Y chromosome, it was possible that either in the cell line or during the construction of the library, unrelated mouse and human sequences had become juxtaposed. To investigate this further cY022 subclones were used as probes against Southern blots containing mouse and human DNA, to establish if the subclones would hybridize to both human and mouse DNA thereby implying sequence homology. Fig. 3.10 shows the results from two such hybridization experiments. As shown, the EcoRI subclone p022E1 hybridizes to mouse sequences only (3E7, 7/2 and 1RE3 tracks) whilst the subclone p022E4 hybridizes to human
FIG. 3.10 Southern Analysis using two cY022 subclones as probes

Two cY022 subclones were used as probes against a panel of EcoRI digested genomic and hybrid DNAs. The blots were washed at 65°C (0.1 X SSC; 0.1% SDS).

Fig. A) p022E4 (4.8kb) probe
Lanes: 1) 1RE3 2) GM1416B 3) OXEN 4) H0RL9X
       Male     Male
      5) 7/2  6) 3E7  7) Blood  8) Placenta

Fig. B) p022E1 (12.1kb) probe
Lanes: 1) Placenta 2) 853 3) 3E7 4) Blood
       Male
      5) OXEN 6) GM1416B 7) 7/2 8) 1RE7
sequences. In all cases, each subclone exhibited strict homology to either human DNA or mouse DNA but not to both, a result consistent with the cosmid clone containing contiguous sequences of human and mouse DNA.

cY019 is derived from human Y-linked DNA. Fig. 3.11 shows cY019 hybridizes specifically to DNAs which contain a human Y chromosome (3E7, 853, male placenta and OX); there is no cross hybridization to either mouse sequences (IRE3), autosomal sequences (GM1416B) or to X chromosome sequences (HORL 9X). 8 Y-linked EcoRI fragments are revealed by this hybridization (major bands a-d 15Kb,13kb, 11.5Kb and 10.2Kb; minor bands e-h 8.9Kb, 6.2Kb, 5.0Kb and 4.1Kb tracks 1-4). However, between the different DNAs there are some fragment size differences (eg. in tracks 2 and 3, bands c and d are 11.0Kb and 9.4Kb respectively, and band b is missing). An additional band (i) is also present in tracks 3 and 4. Such differences, reveal two types of band pattern resulting from the hybridization of cY019 to the different Y-containing DNAs; 3E7 and OX DNAs share one such pattern whilst 853 and placenta DNAs share a second.

The lack of homology between mouse sequences and cY019 sequences was verified by Southern analysis; a 5.7kb PstI subclone from cY019 (termed p019P3, see fig.3.3) did not hybridize to male, female or XXSxr mice DNAs (fig. 3.12). However, the strict human Y homology was confirmed by probing a Southern blot containing a series of male and female DNAs digested with TaqI. Fig. 3.13 shows p019P3 hybridizes to a major 2.7kb fragment and to minor 4.6kb, 4.9kb, 5.9kb,7.2kb and 12.1kb TaqI fragments within male DNA but does not hybridize to any sequences within female DNA. A limited study, in which 57 TaqI-
FIG. 3.11 Southern Analysis using cY019 as a probe

cY019 was used as a probe against a panel of EcoRI digested genomic and hybrid DNAs following prehybridization of the labelled cosmid clone with total human and mouse DNAs to 'eliminate' highly repeated sequences.

Exposure time - 10 days.

Lanes:  1) 3E7     2) 853     3) Male Placenta
        4) OXEN    5) 7/2     6) a23
        7) HORL 9X 8) GM1416B 9) 1RE3
FIG. 3.12 Southern Analysis of Mouse DNAs using cY019 as a probe

cY019 was used as a probe (following prehybridization with total human and mouse DNAs to 'eliminate' highly repeated sequences) against a panel of EcoRI or PstI digested mouse DNAs.
Exposure time - 10 days.

Lanes: 1) Male Mouse 2) Female Mouse 3) XXSxr
DNA samples were digested with TaqI. The blot was probed with p019P3 and washed at 65°C (0.1 x SSC; 0.1% SDS).
Exposure time - 10 days.
The northern analysis was subsequently repeated using a newly prepared blot containing the panel of RNAs indicated in Fig. 3.14(2). Again the 019P3 fragment from cY019 was used as a probe, however after a prolonged exposure of 10 days faint hybridization signals were revealed in the tracks containing adult testis RNAs (arrowed). These signals are very weak however when compared with those revealed when a *MIC2* cDNA clone (p1A) was used as a probe (exposure time = 4 days) or when a GAPDH subclone (exposure time = 6 hours) was used to probe the same northern blot.

No hybridization signals were revealed in the adult or foetal muscle tracks or the adult and foetal liver tracks.
digested DNAs derived from 5 different families, were similarly probed, revealed no TaqI restriction fragment length polymorphisms.

These preliminary results indicated that cY019 was a good candidate clone for a Y-linked CpG island. As previous studies had shown an association between genes and CpG rich sequences (Bird 1986; Gardiner-Garden & Frommer 1987), preliminary northern analysis was carried out to investigate whether the 5.7kb region of cY019 contained expressed sequences. A northern blot of adult muscle, liver, heart and testis RNA was therefore screened with p019P3. However even after prolonged exposure, no hybridization signals were revealed (fig. 3.14(1)). A similar result was obtained when a foetal testis cDNA library was also screened with p019P3 (fig. 3.15). In parallel to this study, single copy subclones of cY022, cY113 and cY24 were also used as probes to screen the northern blot and the foetal testis library (plated at a density of 2.5 x 10⁶ pfu/20cmx20cm plate and plated onto 2 plates). Again no hybridization signals were revealed (data not shown), although a control DNA probe (GAPDH subclone) did reveal a signal on the northern blot (Fig.3.14(1)) and a skeletal muscle α actin control probe revealed 19 positive signals when used to screen the foetal testis cDNA library (Fig.3.15B).

Further chromosomal mapping data for cY019 has been obtained by probing a series of hybrid DNAs, containing various regions of the Y chromosome (Pritchard & Goodfellow 1986) with the 5.7kb PstI subclone p019P3. Fig. 3.16 shows that this subclone appears to preferentially hybridize to DNAs which contain a centromeric region. Indeed, these data are consistent with cY019 localization being between the centromere and DYS8, within a subdivision of DYS1-like sequences. However, as the construction of these hybrids has led to many
FIG. 3.14 Northern analysis of poly A⁺ RNA from different human tissues probed with (A) ³²P labelled p019P3 and (B) ³²P labelled GAPDH clone (arrowed)

Exposure time - (A) 20 days  (B) 7 days

Lanes:  1) Testis    2) Heart    3) Liver    4) Muscle

(Signal shown in Fig. 3.14A is the result of a previous exposure using GAPDH as the probe)
(8) Postnatal muscle poly A
(7) Postnatal muscle poly A (etiquetted)
(6) Adult muscle poly A
(5) Adult liver poly A
(4) Adult liver (debrayed)
(3) Total adult liver
(2) Adult testis poly A

Lanes: 1) Total adult testis

Exposure time - (A) 10 days (B) 4 days (C) 6 hours

MICS cDNA clone p1A (C) etp labelled GAPDH clone.

Human tissues probed with (A) etp Labelled p0193, (B) etp labelled

Fig. 3.14(2) Northern analysis of total and poly A RNA from different

8 7 6 5 4 3 2 1

8 7 6 5 4 3 2 1

8 7 6 5 4 3 2 1
FIG. 3.15 λgt11 Foetal Testis cDNA Library (plated at 2.5 X 10⁶ pfu/plate) screened with (A) ³²P labelled p019P3 and (B) ³²P labelled skeletal muscle α actin clone.
FIG. 3.16 Southern analysis to localize p019P3 on the Y chromosome

A blot of PstI digests of genomic and hybrid cell lines (containing various regions of the human Y chromosome as indicated by the absence or presence of known genetic markers) was probed with $^{32}$P labelled p019P3 (5.7kb). The blot was washed at 65°C (0.1 X SSC; 0.1% SDS). Exposure time - 10 days.

Hybrid cell line data supplied by C. Pritchard.
rearrangements of the Y sequences a precise localization was not possible. Therefore an in situ hybridization study, again using the probe p019P3, against metaphase spreads, was carried out to resolve cY019s chromosomal position. Fig. 3.17 shows the results from this study. As illustrated, p109P3 hybridizes almost exclusively to sequences on the Y chromosome; there is no significant hybridization to sequences elsewhere in the genome. In addition however, it is apparent that sequences homologous to the 5.7kb PstI fragment are present throughout the human Y chromosome and are not localized to a specific region; a result indicative of a repeated sequence. This was further substantiated by the fact that, in common with other repeated sequences when used as probes, the signals resulting from the hybridization were very strong after only a few hours (L. West pers. comm.).

b) Quantitative analysis

To estimate the number of copies of p019P3 on the human Y chromosome, a quantitative blot was prepared. The number of pg of the 019P3 fragment, corresponding to 30-100 copies of this sequence were calculated and adjusted to account for the complete p019P3 subclone. Each quantity of DNA was then digested with PstI and the fragments resolved on a 0.8% agarose gel alongside PstI-digested DNAs from hybrids 853 and 3E7 and male placenta. Following hybridization using only the 019P3 fragment as a probe, comparisons of signal strengths were used to estimate the numbers of copies of the 019P3 sequence in each of the hybrid or genomic DNAs. Fig. 3.18 illustrates that the strength of signals generated in the 3E7 and 853 tracks correspond to
FIG. 3.17  *In situ* hybridization

(A) *In situ* hybridization of p019P3 to a 46,XY derived metaphase spread. All metaphases examined showed at least two grains over the Y chromosome, as indicated.

(B) Grain distribution on the complete male karyotype.

(C) Grain distribution on the Y chromosome of 20 metaphase spreads after hybridization *in situ* with p019P3. The grains scored on the autosomes are at a diploid level, whereas the signal of the X and of the Y is haploid.

*In situ* hybridization carried out by L. West.
B.

1 2 3 4

5 6 7 8 9 10

11 12 13 14 15 16 17 18

19 20 21 22 X Y

C.

-123-
50-60 copies of 019P3. Unfortunately only a very weak signal was generated in the placental track due to the inaccurate estimation of the DNA concentration.

From karyotypic analysis it is known that 853 cells may be tetraploid with respect to hamster but carry 4 human Y chromosomes, or are diploid and carry 2 Y chromosomes and therefore contain 1 Y chromosome per haploid genome. 3E7 cells however are diploid with respect to mouse but carry, on average, 4 human Y chromosomes and therefore contain 2Y chromosomes per haploid genome. Therefore whilst the signal strength shown in the 853 track represents 50-60 copies of 019P3, a signal strength corresponding to 50-60 copies of 019P3 in the 3E7 track only represents 25-30 copies of 019P3 per single Y chromosome. Although further quantitative analysis should include a human male DNA track to give a more defined 019P3 copy number, these results indicate that 019P3 occurs as a relatively low copy number repeat on the human Y chromosome and that different Y chromosomes may carry varying numbers of this repeat element.

3.2.3. Sequence analysis

To obtain additional data on the CpG content of this region of cY019 and to determine whether it contained any open reading frames or was homologous to known sequences, DNA sequencing was carried out. The 3.2kb PvuII fragment which lies within the 5.7kb PstI region (see fig. 3.3) was purified. This was then cleaved with either Rsal or Sau3A and the fragments cloned into either Smal-cut M13 or BamHI-cut M13, respectively. Overlapping clones were therefore generated.
FIG. 3.18 Quantitative blot to estimate the number of copies of p019P3 on the Y chromosome.

(A) Ethidium bromide stained gel prior to Southern transfer.

(B) Autoradiograph of blot following hybridization using only the 019P3 fragment as a probe.

If a 5.7kb sequence is present once per haploid genome then $5 \mu g$ of genomic DNA will contain:

$$5 \times 5.7 \times 10^3$$

$$\frac{3 \times 10^3}{3 \times 10^3}$$

$$= 9.5 \times 10^{-6} \mu g$$

of a 5.7kb fragment.

Using this value to represent 1 copy of the sequence, multiple amounts of p019P3, representing 30 to 100 copies were loaded onto a 0.8% agarose gel alongside $5 \mu g$ of genomic DNA. Following Southern transfer, the blotted DNA was hybridized with the $^{32}P$ labelled 019P3 fragment only. The blot was washed at 65°C (0.1xSSC:0.1%SDS). Exposure time - 4 days.
Fig. 3.19. Partial nucleotide sequence of p019P3 and derived amino acid sequence. Underlined regions denote sequence differences from those described by Arnemann et al. (1987). Letters above the sequence denote restriction endonuclease recognition sites. H HpaI E EngI B BssHII N NarI

* indicates possible left splicing (splice donor) sites.
Random clones were then chosen and sequenced according to the method described in section 2.11.

Fig. 3.19 shows a 1.15kb DNA sequence from this region. A computer aided search of the nucleotide sequence revealed an open reading frame of 522 basepairs. Although no start or polyadenylation signals were found, two potential left splicing (splice donor) sites were found at positions 703 and 724 (relative to the start of the open reading frame). In addition, a computer search of the GenBank Data library revealed no significant homologies to DNA sequences described so far. However to complete this search the amino acid sequence was used in a computer aided search of the HBRF protein sequence data bank. This revealed a sequence showing 98% homology, termed a hypothetical Y chromosome-specific protein, described by Arnemann et al. (1987). DNA sequence comparisons showed that the two sequences were almost identical although base changes give rise to 4 changes in the amino acid composition as indicated (fig. 3.19).

The G+C content for this region was calculated to be 58% which is above the 40% average for the genome as a whole (Bird 1986). However the G+C content increases to 65-70% within the body of the open reading frame. This change in G+C content is roughly paralleled by the CpG density which increases three-fold in this area. Similarly GpC density also increases three-fold within this region although overall the GpC density is higher than CpG density (fig. 3.20).
FIG. 3.20 G+C content and the incidence of CpG (●—●) and GpC (●—●) doublets (per 50bp) across a 1.15kb region of p019P3. The positions of several 'CG' enzyme sites are shown and the 40% G+C content of bulk genomic DNA is indicated for reference in the upper graph (dashed line).

H HpaII  B BssHII  E EagI  N NotI
% G+C

CpG and GpC
incidence 50bp
3.2.4. Methylation studies

To further define the CpG island-like character of the 019P3 fragment, the methylation status of this sequence, within genomic DNA, was investigated. By comparing the fragment pattern generated by \textit{Mspl} (which cleaves DNA whether the central cytosine is methylated or not) with that generated by \textit{Hpall} (which does not cleave DNA when the central cytosine is methylated) it is possible to predict if a site within the DNA is methylated. Male and female placental DNAs were therefore cleaved with \textit{Mspl} or \textit{Hpall} and the fragments separated on an 8% denaturing polyacrylamide gel. Following electroblotting, the filter was probed with the entire 019P3 fragment. Fig. 3.21 illustrates that 019P3 does not show any homology to female DNA, however, three identical bands corresponding to fragments of size 435bp, 285bp and 80bp are revealed in both the male placental DNA tracks. An additional band (fragment size 260bp) in the \textit{Mspl} track reveals that at least one site within the 019P3 sequence is methylated whilst at least four out of the remaining ten 5'CCGG3' sites clustered within 019P3 are not methylated in placental DNA.

Using the sequence data (fig. 3.19) it is possible to correlate some of the methylation data with the known \textit{Hpall} sites. Fig. 3.22 illustrates that three, possibly four of the \textit{Hpall} sites within this region are not methylated, so that cleavage with \textit{Mspl} or \textit{Hpall} gives rise to a 288bp fragment (sites 4 and 5) and an 80bp fragment (sites 5 and 6). The methylation status of sites 2 and 3 within the open reading frame is not known however if these are not cleaved and site 1 is not methylated, cleavage at sites 1 and 3 would give rise to a
FIG. 3.21 30µg of Male (M) or Female (F) Placental DNA were digested with either \textit{HpaII} or \textit{MspI} and loaded onto an 8% denaturing polyacrylamide gel. Following electroblotting, the DNA was hybridized with $^{32}$P labelled p019P3 as a probe. The blot was washed at 65°C (0.2xSSC;0.2%SDS). Initial exposure (shown) - 7 days. After long exposure (14 days) of the autoradiograph, the 80bp fragment became visible but the high background made it difficult to reproduce photographically.
FIG. 3.22 Restriction map of sequenced region showing positions of HpaII sites and distances between them (in bp).

- Unmethylated HpaII sites
- Methylation status of HpaII site unknown
- Shaded region represents the open reading frame
- X Positions of potential splicing signals
- B BssHII
- E EagI
- N NotI
430bp fragment. The location of the 260bp fragment is not revealed by the sequence data.

Since all CpGs are not accessible to restriction endonuclease cleavage it is possible that some methylated cytosines within the sequence have escaped analysis, however this preliminary study provides evidence supporting the notion that this sequence is to a degree hypomethylated and therefore CpG island-like, in this tissue at least.

Additional information regarding the methylation state of 019P3 in a second tissue has been gained by preparing a Southern blot of male placental and blood DNA cut with *PstI* and a 'CG' enzyme (known from previous data, fig. 3.4, to cleave the 019P3 fragment) in a series of double digests. In common with *HpaII*, these 'CpG' enzymes are unable to cleave DNA when one or more of the cytosine residues are methylated. Following hybridization, using p019P3 as the probe, and autoradiography, a series of bands are revealed in all tracks (fig. 3.22). A major 5.2 kb fragment and minor 9.5 kb, 7.9 kb, 7.0 kb, 6.2 kb, 3.8 kb, 3.3 kb, 2.6 kb and 1.6 kb fragments are shown. In addition, a 22 kb fragment and 10.2 kb fragment are revealed in all placental DNA tracks. Variations between the tracks are also shown. Male placental DNA cut with *Smal* and *PstI* gives rise to a 4.2 kb fragment which is not present in the similar digest of male blood DNA; 4.3 kb, 2.8 kb and 1.8 kb fragments are present in male placental DNA cleaved with *Eagl* and *PstI* but again not present in a similar digest of blood DNA, and additional 4.05 kb, 3.55 kb and 1.5 kb fragments are revealed when placental DNA is cleaved with *BssHII* and *PstI*, but are not shown in a similar digest of male blood DNA.
FIG. 3.23 Southern blot analysis of Male Blood (1) and Male Placenta (2) DNA digested with PstI and a 'CG' enzyme and hybridized to $^{32}$P labelled p019P3. The blot was washed at 65°C (0.1xSSC; 0.1%SDS). Exposure time - 5 days.
CpG ISLANDS WITHIN HOMER V COSMID CLONES

In the previous chapter I have described a strategy for isolating CpG islands from a Y chromosome specific library by cleaving Y derived clones with the Hpall ischorschizomer Mspl, and selecting clones which contained numerous CCGG sites. Using this approach four candidate clones were chosen. Of these only one, cY019, complied with many of the criteria related to CpG islands.

Although cY022, cY113 and cY24 gave rise to many small fragments when cleaved with Mspl, restriction mapping analysis revealed that these fragments were not clustered. In addition, CG-enzyme cleavage of these clones was limited, failing to reflect the large numbers of CG-enzyme recognition sites usually associated with CpG islands (Lindsay & Bird 1987).

These three clones were shown to be Y-linked through Southern blot analysis. Studies on cY022 revealed homology to human Y chromosome and mouse sequences. Further analysis indicated that such homology was due to the presence of contiguous mouse and human sequences within cY022 as a probable consequence of cloning artefacts.
during the preparation of the Homer V library. Limited data on cY24 showed sequence homology to Y-linked and possibly autosomal sequences, whilst cY113 sequences have been shown to be homologous to human Y chromosome, X chromosome, autosomal and mouse sequences.

Preliminary northern analysis revealed that sequences within these clones are not expressed in human heart, liver, muscle or testis. Similarly, such sequences are not present in a foetal testis cDNA library (data not shown).

Therefore although these clones are CpG rich relative to other clones within the Homer V library, as judged by the frequency of MspI cleavage, they do not represent regions of CpG islands. As these clones were originally selected from the Homer V library because of their ability to hybridize to total human DNA, it is more probable that these clones contain human repeat elements and originate from bulk, inter-gene regions of Y chromosomal DNA.

In contrast to cY022, cY113 and cY24, analysis of Homer V clone cY019 revealed a more CpG island like sequence. In addition to the numerous small MspI fragments which the cloned DNA contained, restriction mapping analysis showed that the majority of the MspI recognition sites were clustered within a 5.7kb PstI fragment of the clone. Furthermore, restriction analysis of this PstI fragment revealed the presence of 4 CG-enzyme recognition sites. A result consistent with the clone being CpG island-like.

Southern blot analysis with probe p019P3 demonstrated Y-specificity not only with different enzymes, but also under non-stringent hybridization and washing conditions. However, through such analyses no restriction fragment length polymorphisms were revealed with either PstI or TaqI although an RFLP was revealed when the
** Subsequent northern analysis has shown that sequences within cY019 (019P3) are not expressed in foetal testis, adult or foetal muscle or adult or foetal liver but are expressed in adult testis although at a very low level when compared with MIC2 and GAPDH expression (Fig. 3.14(2)). In addition recent screening of an adult testis cDNA library with the probe 019P3 has revealed 3 positive clones (K. Taylor per. comm.). It therefore appears that sequences homologous to 019P3 are expressed in the human adult testis as indicated by Arnemann et al (1987).
complete cosmid was used as a probe against a panel of EcoRI digested hybrid and genomic cell lines (fig. 3.11). In addition p019P3 sequences showed no homology to sequences within a foetal testis cDNA library or a muscle cDNA library (data not shown), nor did p019P3 contain sequences which are expressed in adult liver, heart, muscle or testis as judged by northern blot analysis, although such analyses would need to be repeated to confirm these results.**

Sequence analysis of a small region within p019P3 did however reveal an open frame of 522bp, and although computer aided searches were undertaken, no homology to DNA sequences present in the GenBank data library was found. To complete this analysis the amino acid sequence of the open reading frame was used to search the NBRF protein sequence data bank. In this instance a sequence was revealed which showed 98% homology to the p019P3 sequence. Reports termed this sequence 'a hypothetical Y chromosome specific protein' (Arnemann et al.1987). Comparisons of the DNA sequences revealed almost complete homology over the 1115bp region, with the exception of 8 base changes which result in four amino acid changes. Only one of these changes is conserved:

<table>
<thead>
<tr>
<th>Position relative to start of open reading</th>
<th>Codon &amp; amino acid from Arnemann et al (1987)</th>
<th>Codon &amp; amino acid from p019P3 frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>+232</td>
<td>Glycine</td>
<td>Serine</td>
</tr>
<tr>
<td></td>
<td>GGA</td>
<td>TCG</td>
</tr>
<tr>
<td>+250</td>
<td>Cysteine</td>
<td>Valine</td>
</tr>
<tr>
<td></td>
<td>TGC</td>
<td>GTC</td>
</tr>
<tr>
<td>+494</td>
<td>Glycine</td>
<td>Glutamic acid</td>
</tr>
<tr>
<td></td>
<td>GGA</td>
<td>GAG</td>
</tr>
</tbody>
</table>

-136-
In common with the protocol initially used to select cY019 (Wolfe et al. 1984), Arnemann et al. (1987) isolated a cosmid clone (cos 36) initially selected from a human Y-chromosome enriched cosmid library (Cooke et al. 1985), constructed from DNA of the mouse-human cell hybrid 3E7. Subclone pJA36B (corresponding to p019P3) was found to be a low copy sequence within the cos36 cosmid clone.

In common with p019P3, pJA36B (termed DYS14, Goodfellow et al. 1985) was shown to be male specific and to hybridize to a 2.8kb TaqI fragment within human male DNA correlating with the 2.7kb TaqI fragment shown in this study. Furthermore, in agreement with this study, Arnemann et al. (1987) estimated that at least 10 copies of this sequence are present on the human Y chromosome, and that such elements appeared not to be contiguous. In addition, using DNA samples from different patients with aberrant sex chromosomes of known deletions Arnemann et al. (1987) localized pJA36B to the median region of Yp a result consistent with the hybrid cell line mapping data shown in this study and the limited in situ hybridization analysis undertaken. Such a localization would therefore place p019P3 sequences within a region of the Y chromosome to which Bkm related sequences also map. Bkm (banded krait minor) related DNA is preferentially present on the heterogametic sex chromosome of snakes and mice (Schäfer et al. 1986; Epplen et al. 1982; Singh et al. 1980) and is transcribed in different tissues (Epplen et al. 1983). It consists of a repeat sequence composed of two tetranucleotides, GATA and GACA, repeated several hundred times within each repetitive unit (Epplen et al. 1982). In man such sequences have been mapped to the median part of Yp and the
proximal part of Yq (Arneemann et al. 1986). Although analysis of the partial sequence from p019P3 shows no precise sequence homology to this repeat motif, the G and A richness of the sequence is similar. In contrast to this study however, Arneemann et al. (1987) showed that pJA36B was able to detect a poly (A)+ RNA transcript of human testis although total RNA from human thymus, lung and liver failed to hybridize with pJA36B. The inability to detect hybridization of p019P3 to RNA from the tissues examined however could indicate that either the transcription in these tissues is below the limit of detection or that transcription is limited to specific tissues or developmental stages. In addition, the data imply that p019P3 contains only part of the transcribed sequence, as initiation or polyadenylation signals are missing. However, two potential left splicing (splice donor) sites have been found in the non-coding region of the sequence following the open reading frame:

Consensus (the subscripts indicate the percentage of junctions at which the base is found)  

\[
\begin{align*}
\text{Arneemann et al. sequence:} & \quad A_{73} G_{100} T_{100} A_{2} A_{64} G_{64} T_{2}\text{G} \\
\text{p019P3 sequence:} & \quad G \ T \ G \ T \ A \ A \ G \ T 
\end{align*}
\]

Although these do not conform precisely to the splicing consensus sequence (Breathnach & Chambon 1981; Mount 1982) they may function in splicing any transcripts originating from this sequence to generate the mature message.

As this clone appears to be only one of a family of such sequences on the Y chromosome, it is possible that another member is the active gene whilst other elements such as p019P3, are diverging sequences and therefore unable to be transcribed efficiently. The base changes seen from the sequence data would be consistent with this notion. Although differences between the sequence published by
Arnemann et al. (1987) and the sequence reported here may be due to sequencing error, it is also possible that the two sequences are derived from different cloned loci and are therefore not identical.

Methylation studies on male placental DNA reveals information about only five out of the ten CCGG sites known to be present within the 5.7kb PstI fragment of cY019, the remaining sites being undetected, although the remaining fragments may well be outside the range of detection by Southern blotting, indeed, from sequence data, one HpaII fragment is known to be only 30bp long. Despite these limitations analysis shows that at least three similar fragments are generated by both Mspl and HpaII cleavage. Together with the sequence data this indicates that in male placental DNA at least three, possibly four CCGG sites are not methylated. However methyl moieties at these specific sites (CCGG) represent only a fraction of the total methylations of a region. If 5mC is generally found at CpG sequences, methylation at the CCGG site will represent only about 6% of the total methylated sequences. Thus the methylation pattern with these enzymes provides only partial methylation information so that it is probable that many of the other CpGs within p019P3 will be unmethylated consistent with subclone p019P3 containing CpG island-like sequences.

Further methylation data is revealed following hybridization of labelled p019P3 to PstI and PstI/CG-enzyme digests of both blood and placental DNAs. Analysis reveals that all four CG-enzymes which cleave cloned 019P3 sequences, fail to cleave sequences within male blood DNA, as similar hybridization patterns are shown in the presence or absence of these enzymes. It therefore appears that the recognition sites for these enzymes within the 019P3 sequences of male blood DNA are methylated. However, it cannot be overlooked that lack of
cleavage at these sites could also result from the absence of these sites within this sequence in this tissue. Since a series of isoschizomers which will cleave the DNA regardless of the methylation status are unavailable, it is possible that such digestion patterns could also be due to such primary sequence differences.

In contrast, similar digests using male placental DNA do give rise to additional fragments indicating the ability of these enzymes to cleave the DNA and therefore the lack of methylation at these recognition sites. These results agree with other data produced during methylation studies on other repeat sequences of the human Y chromosome. The DYZ2 repeat family, for example, is differentially methylated in different tissues, being methylated in DNA from a variety of somatic tissues (fibroblast, liver and leukocyte) whilst strikingly unmethylated in DNA from sperm (Young et al. 1981). Results with DNA from male placenta indicated an intermediate level of methylation (Cooke et al. 1982; Young et al. 1981). However, although such studies give an indication of the level of methylation within these sequences, it is difficult to ascertain whether such an intermediate level of methylation is due to partial methylation of a single locus or whether it represents an overall methylation status for the complete sequence family. Furthermore, since these tissues contain many cell types such an intermediate level of methylation could represent the 'average' methylation status of all the cell types. However, by isolating additional cosmid clones which hybridize to sequences within cY019, other members of this repeat family could be analysed in order to define methylation and transcriptional differences between them. In addition, pulsed field gel electrophoresis would allow long range mapping of these sequences on
the Y chromosome, defining their numbers and positions more accurately.

The cloning and structural analysis of such sequences should also reveal whether they have arisen through independent events of duplication or viral retroposition. Although p019P3 shows no obvious retroviral origin, Phillips et al. (1982) have shown that in the mouse, DNA containing retroviral elements are present in the male but not in females DNAs, and that male specific fragments varied in copy number from one to as many as one hundred. Such concentrations of dispersed repeated sequences could therefore reflect the poor 'housekeeping' ability of the monosomic Y chromosome, or indicate that the Y is a 'target' chromosome for retrovirus integration or amplification. Additional Southern analysis would also reveal the degree to which this family of repeats is evolutionary conserved.

From these results it is apparent that although cY019 contains sequences which appear to be CpG island-like, features of these sequences do not strictly obey the defined character of CpG islands based on CpG density and G+C numbers. Despite this, the results do indicate the suitability of this approach to finding CpG islands and thus transcribed sequences, particularly in light of results from other studies which demonstrate that one of the selected clones appears to contain expressed sequences.
CLONING *Hpall* TINY FRAGMENTS FROM CHROMOSOME SPECIFIC LIBRARIES

5.1 Cloning fragments from flow sorted chromosomes

Flow sorted human Y chromosomes were provided by Dr. J. Fantes from the lymphoblastoid cell line OXEN (49, XYYYY). Following hypotonic shock to release the metaphase chromosomes, the isolated chromosomes are equilibrated with two DNA specific fluorescent dyes for flow cytometry and sorting; Hoechst 33258 (Ho - binds preferentially to DNA rich in adenine and thymine, Latt & Wohlleb 1975) and chromomycin A3 (CA3 - binds preferentially to DNA rich in guanine and cyosine, Behr et al. 1969). The use of these two dyes enables the chromosomes to be distinguished on their total DNA content and their base composition. (Gray et al. 1979) The Ho and CA3 contents of individual chromosomes are determined by measuring the intensities of fluorescence that are emitted as they pass the UV 458nm and laser beams of the FACS. These measurements are then accumulated to form fluorescent intensity distributions termed flow karyotypes. The flow
FIG. 5.1 Flow Karyotype of OXEN Cell Line
karyotype resulting from the Y chromosome sort from OXEN cells is shown in fig 5.1. Following sorting the sample was dialysed to remove the stains and the chromosomes collected by centrifugation. It was estimated that approximately $1.5 \times 10^6$ Y chromosomes were collected, representing approximately 150ng DNA (B.Young pers. comm). 85% of the sample was thought to be Y chromosome with contamination by chromosomes 18 and 21.

DNA from flow sorted Y chromosomes was cleaved with MspI, and fragments less than 200bp were isolated following agarose gel purification and cloned into Clal-cut pBS as described in section 2.8.1. Following transformation of competent E.coli XL1-Blue, the percentage of recombinants was assessed by plating the bacteria on L-agar plates (with 100µg/ml ampicillin) in the presence of the lac operon inducer, IPTG, and a chromogenic substrate of ß-galactosidase (ß gal), X-gal. Non-recombinant XL1-Blue bacteria produce blue colonies under these conditions, in contrast to recombinants where the interruption of the ß gal protein coding sequence leads to loss of ß gal activity and white colonies. Using this technique, approximately 7% of bacteria were shown to be non-recombinants.

5.1.1. Analysis of clones

To screen this library, random clones were selected and DNA prepared by the miniprep method (section 2.4.1). This was then cleaved with Pvull (see fig. 2.1) to assess the insert size (Fig 5.2). From this it was estimated that approximately 78% of the white colonies were recombinants and that the insert sizes ranged from approximately 110bp to 20bp.
FIG. 5.2 *PvuII* Restriction endonuclease digests of miniprep DNA from random clones selected from the *Mspl* library resolved on a 0.8\% agarose gel

V  pBS vector
6 clones were chosen for further analysis and DNA was prepared by the maxiprep method (section 2.4.1). In order to use these clones as efficient probes, despite the small insert size, \textit{in vitro} transcription reactions were used to generate labelled RNA probes specifically from the insert. By cleaving the DNA downstream of both the T7 promoter and the insert, high specific activity 'run off' transcripts were produced (fig. 5.3). 1\mu g of the maxiprep DNA was therefore cleaved with \textit{AcoI} and \textit{in vitro} transcription reactions carried out as described in section 2.6.

The integrity of these transcripts was verified by loading 1% of the reaction onto denaturing polyacrylamide gels. Fig.5.3 illustrates the labelled transcripts resulting from one of these reactions giving rise to a major RNA species of 240b. In all cases the percentage incorporation of label into the transcripts was estimated to be >95% and the specific activity calculated to be \(1-1.5 \times 10^6 \text{ dpm/\mu g RNA.}\)

To localise the clones to the Y chromosome, Southern blot analysis was carried out, using the labelled RNA probes against a panel of genomic and hybrid cell line DNAs. However even after prolonged exposure, no hybridization signals were revealed. Therefore to investigate the nature of the cloned inserts, the RNA probes were used in Southern blot analysis against a series of pBR322, Homer V and pBS digests. Fig.5.4B shows the results of one of these analyses. Hybridization signals were revealed after only 1 hour exposure, demonstrating the efficiency of the hybridization technique, however the probe hybridizes to several pBR322 fragments (tracks 1-6), to the 5.2kb Homer V \textit{EcoRI} fragment (track 7) and the 2.5kb pBS \textit{PvuII} fragment (track 8). Further evidence about the nature of these inserts was gained from a colony screen. 70 recombinant clones were
FIG. 5.3 *In vitro* transcription of cloned *Mspl* inserts generating labelled RNA probes.

(A) Diagrammatic representation of clone and *in vitro* transcription

(B) Autoradiograph following separation of the labelled RNA probes on a 5% denaturing polyacryamide gel.

Lanes:  
(1) RNA probe (92b) generated from PBS vector without insert DNA. (Varying amounts loaded onto 3 adjacent tracks)
(2) RNA probe (240b) generated from a cloned *Mspl* insert (Varying amounts loaded onto 3 adjacent tracks)
T7 Promoter +1(638)

Insert cloned into Clal site

AccI (730)

Ampf

ori

AccI

in vitro transcription

RNA

B bp 1 2 bp

622
404
309
242
190
160
110
76
240
92

-147-
FIG. 5.4 Southern blot analysis to investigate nature of cloned inserts

Lanes: (1-6) pBR322 cleaved with: (1) AccI, (2) PvuII, (3) BglII, (4) HaeIII, (5) Aval, (6) Hpal
(7) Homer V cleaved with EcoRI
(8) pBS cleaved with PvuII

(A) Ethidium bromide stained agarose gel of pBR322, Homer V and pBS fragments generated by cleavage with various enzymes.

(B) Southern blot analysis of gel (A) hybridized with an RNA probe generated from an anonymous cloned Nspl insert.

(C) Southern blot analysis of gel (A) hybridized with a $^{32}$P labelled 20 pBR322 fragment.
FIG. 5.5  (A) Colony screen of clones derived from the MspI library probed with a $^{32}$P labelled 2.033kb pBR322 fragment.  
(B) EcoRI restriction endonuclease digests of clones from (A) which showed no homology to pBR322.  
Digests resolved on 0.8% agarose gel.  
(V) pBS vector
dotted onto a gridded Hybond filter. Following overnight growth the colonies were lysed and the DNA denatured, neutralised and bonded to the filter. A 2033bp pBR322 fragment (which was shown to have no homology to pBS, fig.5.4C) was then used as a probe to screen these DNAs. 43% of the clones were shown to be homologous to pBR322 sequences (fig 5.5A). DNA was subsequently prepared from the clones which showed no pBR322 homology, and linearised with EcoRI. Unfortunately agarose gel electrophoresis of these digests revealed the absence of inserts in any of these clones (fig.5.5B). From these results it became apparent that pBR322 Mspl fragments had been preferentially cloned during this study. This was probably a consequence of using such fragments as size markers on the initial size selection gel following cleavage of the Y chromosome DNA with Mspl, even though they were on non-adjacent tracks to the one from which the target DNA was purified.

5.2 Cloning Mspl fragments from a chromosome-specific λ Library

Since limited quantities of DNA could be obtained from the flow sorted chromosomes, additional Mspl libraries were constructed using DNA derived from Charon 21A chromosome specific libraries obtained from the American Type Culture Collection. As a control, to represent a more 'typical' human chromosome, a chromosome 9 Mspl library was also constructed in parallel with the Y Mspl chromosome library.

DNA was isolated from the ATCC Charon 21A libraries as described in section 2.4.2. Following HindIII cleavage of 50μg of library DNA, to release the human DNA inserts, the digests were run on a 0.7%
FIG. 5.6 Ethidium stained agarose gels showing *Hind*III cleaved Charon 21A library DNA. Right (R) and Left (L) vector arms.

(A) DNA from ATCC Charon 21A Y Chromosome Library
(B) DNA from ATCC Charon 21A Chromosome 9 Library
agarose gel (fig. 5.6). The inserts were then excised from the gel leaving the vector arms and the DNA purified by electroelution and EluTip procedures. It was estimated that approximately 1µg of DNA was recovered and this complete aliquot was digested with MspI. However from lessons learnt from the previous cloning experiment, these MspI fragments were not size selected. In this instance, the complete range of fragment sizes was cloned into ClaI-cut and phosphatased pBS. Following transformation of competent *E. coli* XL1-Blue, the bacteria were plated onto L-agar plates containing 100µg/ml ampicillin, supplemented with IPTG and X-gal in order to differentiate between recombinant and non-recombinant clones. Using this technique it was estimated that 86% of the clones forming the Y-specific library were recombinants whilst 94% of the clones forming the chromosome 9 library were recombinants.

5.2.1. Analysis of clones

White colonies were selected at random and DNA prepared by the miniprep method. This DNA was then cleaved with *PvuII* to release the inserts enabling an estimation of the insert size to be made. Fig. 5.7 shows an example of these gels. The inserts from the Y library range in size from 160bp to 50bp however not all the Y library clones contain inserts (fig. 5.7A). Therefore on this basis it was estimated that approximately 55% of the Y clones were recombinants. In contrast, all clones from the chromosome 9 library contain inserts, with the insert size ranging from approximately 4kb to 100bp. In addition it is evident that many of the inserts contain several *PvuII* sites as shown by the numerous fragments within each track (fig. 5.7B).
FIG. 5.7 *Pvu*II restriction endonuclease digests of random *Hspl* cloned inserts derived from the Charon 21A libraries, resolved on a 0.8% gel.

V PBS vector

(A) DNA from Y chromosome library

(B) DNA from chromosome 9 library

(Line in fig. A. drawn as a guide to illustrate which clones contained inserts)
Six clones from each library were selected at random and DNA prepared by the miniprep method. The DNA was then cleaved with Fvull and the fragments separated on a low-melting agarose gel. These small insert fragments were then excised from the gel and the DNA labelled, without prior extraction from the gel, as described in section 2.5.4. These labelled insert DNAs were then used as probes against a panel of genomic and hybrid DNAs in Southern blot analysis to chromosomally localise each clone (fig. 5.8). Insert DNA from 5 of the clones chosen from the Y library did not hybridize to mouse or human DNA (fig. 5.8A). These clones did however show homology to λ DNA as indicated by the intense signal generated in the λ BstEI1 marker track. However, one clone, termed pBSY2, did show homology to human DNA. The insert DNA from this clone hybridizes to a 4.2kb EcoRI fragment in male placental, 3E7, OX, female blood and GM1416B DNAs (fig. 5.8B). There is no homology to sequences in either HORL9X DNA or mouse (IRE3) DNA. This result is consistent with pBSY2 showing homology to Y and autosomal sequences.

Fig. 5.9 shows the result of Southern blot analysis of one of the chromosome 9 clones. In common with the Y chromosome clones, all chromosome 9 clones analysed showed homology to the λ markers and no homology to either genomic or hybrid cell line DNA.
FIG. 5.8 Southern blot analysis of genomic and hybrid cell lines hybridized with:

(A) an anonymous $^{32}$P labelled cloned Mspl insert from the Y chromosome library;

(B) $^{32}$P labelled Mspl Y chromosome library clone pBSY2

Lanes:  
1) Male Placenta DNA  
2) 3E7  
3) 7/2  
4) HORL 9X  
5) Female Placenta DNA  
6) OX  
7) GM1416B  
8) 1RE3
Fig. 5.9 Southern analysis of genomic and hybrid cell lines hybridized with an anonymous $^{32}$P labelled clone from the chromosome 9 'HTF' library.

Lanes: 1) Male Placenta DNA  2) Male Blood  3) 3E7  4) OX  5) GM1416B  6) 7/2  7) 1RE3
DISCUSSION

CLONING \textit{HpaII} TINY FRAGMENTS FROM 
CHROMOSOME SPECIFIC LIBRARIES

Recent progress in flow cytometry has greatly facilitated the search for chromosome specific cloned DNA sequences. By sorting and subsequent cloning of individual chromosomes from a complex karyotype, chromosome-specific libraries have been established for the human X chromosome (Davies et al. 1981; Kunkel et al. 1982), for several autosomes (Krumlauf et al. 1982) and for the human Y chromosome (Müller et al. 1983; Fuscoe et al. 1986). Flow sorted chromosomes therefore provide an excellent source of chromosome specific DNA. In this study approximately 100ng of DNA was recovered from the flow sorted Y chromosomes. As CpG island sequences are estimated to represent 1% of bulk genomic DNA it was hoped that it would be possible to clone 1ng of material following restriction enzyme cleavage and gel purification. Subsequent results have proved the difficulty in handling such small quantities of DNA. In addition, results have shown the care which must be exercised when choosing DNA markers. In this instance pBR322 \textit{MspI} markers, having compatible ends with the
phosphatase vector, and being present in much greater quantities than the sample DNA to be cloned (despite being loaded away from the sample at the edge of the gel), proved to clone at a greater efficiency than the required sequences.

Due to the difficulty in handling the limited quantities of DNA which result from flow sorted chromosomes, a second approach to cloning $MspI$ fragments from chromosome specific libraries was employed. Y chromosomal DNA was therefore isolated from a Charon 21A library which had been prepared from flow sorted Y chromosomes (Fuscoe et al. 1986). In this case it was hoped that larger quantities of material would be more easily prepared, indeed following purification approximately 1μg of chromosome specific DNA was recovered from each library. Following $MspI$ digestion however it was thought more prudent, in this case, not to size select the small $MspI$ fragments, but to clone the entire aliquot and subsequently select clones with small inserts.

Analysis of a small sample of chromosome 9 clones resulting from this study, showed that many large fragments had been cloned, however further analysis demonstrated that all of these were of $\lambda$ origin and no chromosome 9 derived clones were found. In contrast, analysis of 19 Y chromosome $MspI$ clones showed that in this case, many small fragments had been cloned, although the majority of these also showed homology to $\lambda$ sequences. It therefore appears that although $HindIII$ cleavage of the Charon 21A library released the insert DNA, it may also have caused a degree of partial digestion of the $\lambda$ vector. $\lambda$ sequences together with human chromosome specific sequences would therefore have been co-purified, $MspI$ cleaved and subsequently recloned. As $\lambda$ DNA contains a greater density of $MspI$ sites than human DNA it is probable
that λ *Mspl* fragments may constitute a large proportion of the *Mspl* digest and are therefore preferentially cloned.

From calculations based on CpG density within the bulk genome, it is estimated that there are of the order of 30 000 islands in the haploid genome, and the average spacing between them is ≈ 100kb (although variation about this mean is large) (Brown & Bird 1986). If it is assumed that the heterochromatic region of the human Y chromosome accounts for 50% of the chromosome, it would be expected that the Y chromosome would harbour at least 150 genes. Indeed, theoretically any gene could be found in the pseudoautosomal region as it will be present identically in males and females, assuming that the X-located gene escapes X-inactivation. To date however, only three expressed sequences have been found on the Y chromosome. Apart from TDF and perhaps a few genes which contribute to the male phenotype it is possible that other expressed genes will not be found on the Y specific part of the Y chromosome. This is a direct consequence of the chromosomal basis of sex determination and the evolution of the X-inactivation mechanism. It seems consistent, therefore, that in this study from the small sample of clones analysed so far, only one clone, pBSY2, has shown homology to Y chromosomal sequence and/or autosomal sequences.
Although it is to be expected that the human Y chromosome will soon be thoroughly mapped on the molecular level, and comparisons of karyotype/phenotype relations has led to an assignment of functions to various subregions of the chromosome, very little is known about Y-linked expressed genes. Apart from MIC2, ZFY and the testis specific transcript described by Arnemann et al (1987) no other expressed sequences have been found despite evidence for the presence of genes controlling H-Y antigen expression, genes involved in spermatogenesis and 'anti-Turner' genes which rescue males from the Turners phenotype exhibited by XO females.

Evidence suggests that a high proportion of CpG islands are associated with expressed sequences (Bird 1986; Gardiner-Garden & Frommer 1987) and as a consequence it has been proposed that CpG islands can be used to identify new genes. This study has therefore examined the occurrence of CpG islands on the human Y chromosome as an approach to cloning Y-linked genes.

An initial analysis used MspI-cutting frequency to select Y-linked cosmid clones which had a high CCGG density and thus high CpG density. Although many of the selected clones did not have clustered MspI sites, one candidate clone, cY019, did contain island-like
sequences and evidence from other studies confirmed that this region was associated with expressed sequences. This assay therefore proved to be a rapid technique for assessing a large number of clones for island-like sequences and expressed sequences.

The second approach to find CpG islands has made use of flow sorted chromosomes as a source of pure chromosomal DNA. By cleaving this DNA with Mspl it was hoped that Mspl fragments could be cloned directly and these clones used to probe larger chromosome specific cosmid or λ libraries in an attempt to localize the Mspl clones within CpG islands. Unfortunately this approach proved more difficult and although one Y-linked clone was found, the low numbers of chromosome specific clones has demonstrated the inefficiency of this technique. Such low numbers of Y-specific Mspl clones is however surprising with respect to the fact that the Y-linked 2.1kb Haelll repeat contains a single Mspl site near the middle of the repeat (Cooke et al. 1982; Young et al. 1981), and this alone should therefore contribute to Mspl clones.

Using additional techniques to diagnose the presence of CpG islands it may be possible to define other HTF island-like sequences on the human Y chromosome. As CG-enzyme sites are clustered in island-like sequences, the cloning of DNA sequences spanning the recognition sites of these rare cutting restriction enzymes facilitates immediate access to the CpG rich DNA regions. Estivill et al (1987), for example constructed a CG-enzyme cosmid library, and used this Xmal (Bgl)/Hindll library in a search for a candidate for the cystic fibrosis gene. Following screening, a single copy sequence was found associated with a CpG island sequence which further analysis showed was conserved throughout mammalian evolution and was
expressed in lung, placenta, intestine and kidney. Similarly Buiting et al. (1988) constructed a SacI library from chromosome 15 DNA derived from flow sorted chromosomes. Following cleavage of the chromosomal DNA with HindIII, the fragments were circularized by self-ligation and then subjected to SacI digestion. Fragments which contained a SacI site were therefore linearized and available for subsequent recloning into an appropriately cut vector, whilst fragments which lacked a SacI site remained circularized and were therefore excluded from the ligation reaction. Using this technique Buiting et al. (1988) identified a large number of single copy clones which contained evolutionary conserved sequences.

Furthermore, such cloned sequences can provide the necessary landmarks for physical mapping. Using these clones to probe pulsed field gradient gel blots (Schwartz & Cantor 1984) can provide information about neighbouring CpG islands, linking genomic restriction fragments generated by these enzymes (Poustka & Lehrach 1986; Smith et al. 1987). Thus construction of a highly detailed molecular map would give further insight into the molecular organisation of the Y chromosome providing a framework within which genes could be placed.

Evidence suggests that the Y chromosome is strikingly deficient in known expressed genetic material. This deficiency could reflect an ascertainment bias, however, a more likely explanation is the absence of genes from the Y chromosome. Although the two Y-linked genes which have been described in most detail so far appear to be associated with CpG islands it is possible that other Y chromosome located genes lack such 'marker' sequences and will therefore remain undetected using these techniques. Despite many tissue-specific genes showing
associations with CpG island-like sequences, many others, for example β-globin gene family, human nerve growth factor gene and the mouse insulin gene, show no evidence of HTF-like sequences (see reviews by Bird 1986; Gardiner-Garden & Frommer 1987). As other Y-linked genes may be expressed in a tissue specific manner they may also show no association with CpG islands. In addition, having detected a putative Y-linked CpG island it may be difficult to ascertain its expression, as candidate Y-linked genes, such as those involved in spermatogenesis or sex determination, may be expressed for only short periods of time or at certain stages in development.
REFERENCES


-165-


CHANDLEY, A.C., AMBROS, P., McBEATH, S., HARGREAVE, T.B.,
KILANOWSKI, F. & SPOWART, G. (1986) Short arm dicentric Y
chromosome with associated statural defects in a sterile man.
Hum. Genet. 72 350-353.

CHAPMAN, V.M. KRATZER, P.G., SIRACUSA, L.D., QUARANTILLO, B.M.,
EVANS, R. & LUSKAY, R.M. (1982) Evidence of DNA modification in
the maintenance of X-chromosome inactivation of adult mouse

CHEN, A. & FALEK, A. (1971) Cytological evidence for the association
of the short arms of the X and Y in the human male. Nature 232
555-556.

deletion of heterochromatin only of the Y chromosome in an
azoospermic male. Hum. Genet. 64 297-300.

control of liver specific gene expression: cloned human
retinol-binding protein gene is expressed in HeLa cells.
EMBO J. 6 631-636

CONARY, J., LORAKOWSKI, G., SCHMIDT, B., POHLMANN, R., NAGEL, G.,
MEYER, H.E., KRENTLER, C., CULLY, J., HASILIK, A. & von FIGURA,
Res. Comm. 144 1010-1017.

262 182-186.

Hum. Genet. 50 39-44.

Chromosoma 87 491-502.


Trends in Genetics 2 201-211.


-171-
chromosome. In: The Y chromosome, Part A: Basic characteristics
Alan R. Liss. N.Y.

DARLING, S.M., BANTING, G.S., PYM, B., WOLFE, J. & GOODFELLOW, P.N.
(1986) Cloning an expressed gene shared by the human sex

DAVIES, K.E., YOUNG, B.D., ELLES, R.G., HILL, M.E. & WILLIAMSON, R.

thorough review of structural anomalies of the Y chromosome.


Two XX males in one family and additional observations bearing

Genetic evidence of X-Y interchange in a human XX male. Nature
307 170-171.

DISTEACHE, C.M., CASANOVA, M., SAAL, H., FRIEDMAN, C., SYBERT, V.,
deletions of the short arm of the Y chromosome in 46,XY females.

DOERFLER, W., LANGER, K.D., KREBEL, D., WEYER, U., DORBRZANSKI, P. &
KNUSTKRON, B. (1985) Progress in Clinical and Biological Research


ESTIVILL, X., FARRALL, M., SCAMBLER, P.J., BELL, G.M. HAVLEY, K.M.F.,
LENCH, N.J., BATES, G.P. KRUYER, H.C., FREDERICK, P.A., STANIER,
candidate for the cystic fibrosis locus isolated by selection for

evidence for meiotic crossing over between X and Y chromosomes
of male mice carrying the sex reversing (Sxr) factor.

chromosomes by flow cytometry and sorting.
Cytometry 4 88-91

DNA restriction endonuclease fragments to high specific activity.
Anal. Biochem. 137 266-267

FERGUSON-SMITH, M.A. (1966) X-Y chromosomal interchange in the
aetiology of true hermaphroditism and XX Klinefelter's syndrome.
Lancet 11 475-476.

FERGUSON-SMITH, M.A., SANGER, R., TIPPETT, P., AITKEN, D.A. & BOYD, E.
(1982) A familial translocation which assigns the Xg blood group
locus to the region Xp22.5.

Y-specific sequences by deletion mapping and analysis of X-Y
interchange males and females. Development 101 (Supplement) 41-50

FORD, C.E., JONES, K.W., POLANI, P.E., de ALMEIDA, J.C. & BRIGGS, J.H.
(1959) A sex chromosome anomaly in a case of gonadal dysgenesis
(Turner's syndrome). Lancet i 711-713.


Production of HY antibody by female mice that fail to reject male skin. Nature (New Biol.) 238 55-57

GOODFELLOW, P.N., BANTING, G., LEVY, R., POVEY, S. & McMICHAEL, A.  


GOODFELLOW, P.N. (1983) Expression of the 12E7 antigen is controlled independently by genes on the human X and Y chromosomes. Different. 23(s) 35-39.


-176-
GOODFELLOW, P.J., MONDELL, C., DARLING, S.M., PYM, B., LITTLE, P.
region at the 5' end of the MIC2 gene on the active X, the
85 5605-5609.

GORDON, R.D., SIMPSON, E. & SAMELSON, L. (1975) In vitro cell mediated
immune responses to the male specific (H-Y) antigen in mice.
J. Exp. Med. 142 1108.

GRAY, J.W., PETERS, D., MERRILL, J.T., MARTIN, R. & VAN DILLA, M.A.
(1979) Slit-scan flow cytometry of mammalian chromosomes.
J. Histochem. Cytochem. 27 441-444.

GUELLAEN, G., CASANOUVA, M., BISHOP, C., GELDWERTH, D., ANDRE, G.,


Yq deletion and failure of spermatogenesis.

HEILIG, R., HANAUER, A., GRZESCHIK, K.H., HORS-CAYLA, M.C. & MANDEL,
J.L. (1984) Actin like sequences are present on the X and Y
chromosomes. EMBO J. 3 1803-1807.

HOFKER, M.H., WAPENAAR, M.C., GOOR, N., BAKKER, E. VAN OMMEN G.&
PEARSON, P.L. (1985) Isolation of probes detecting restriction
fragment length polymorphisms from X-chromosome specific
libraries: potential use for diagnosis of Duchenne muscular


human X chromosome DNA sequences.

developments in the detection of deoxyribonucleic acid 
synthesis by 33258 Hoechst fluorescence.
J. Histochem. Cytochem. 23 493-505

LAU, B.C., MOHANDAS, T.K., SHAPIRO, L.J., SLAVKIN, H.C. & SNEAD, M.L. 
(1989) Human and mouse amelogenin gene loci are on the sex 
chromosomes. Genomics 4 162-168

Testis-specific transcripts detected by a human Y-DNA-derived 
probe. Development 101 (supplement) 177-183.

of spermatogenic impairment in XO/Sxr and T31H/+ male mice.

LINDSAY, S., MONK, M., HOLLIDAY, R., HUSCHTSCHA, L., DAVIES, K.E.
RIGGS, A.D. & FLAVELL, R.A. (1985) Differences in methylation on 
the active and inactive human X chromosomes.


MOHANDAS, T., SHAPOILO, L. J., SPARKES, R. S. & SPARKES M. C. (1979)


Nucleic Acids Res. 16 6813-6824.


Hum. Genet. 64 110-115.

Hum. Genet. 82 82-84.


Nucleic Acids Res. 14 6489-6505.

Am. J. med. Genet. 28 393-401.


Nucleic Acids Res. 15 9227-9238


PAGE, D.C., MOSHER, R., SIMPSON, E.M., FISHER, E.M.C., MARDON, G.,
POLLACK, J., McGILLIVRAY, B., de la CHAPELLE, A. & BROWN, L. (1987b)
The sex-determining region of the human Y chromosome encodes a zinc

PAGE, D.C., BROWN, L.G. & de la CHAPELLE, A. (1987c) Exchange of
terminal portions of X- and Y-chromosomal short arms in human XX

National Foundation.

PERLMAN, J. & FUSCOE, J.C. (1986) Molecular characterization of the
purity of seven human chromosome-specific DNA libraries.

PETIT, C., de la CHAPELLE, A., LEVILLIERS, J., CASTILLO, S.,
accounts for most but not all cases a human XX maleness.
Cell 49 595-602.

of the pseudoautosomal region; comparison with genetic linkage
map. EMBO J. 7 2369-2376.

Male and female mouse DNAs can be discriminated using retroviral

Cloning and expression of cDNA for anti-Mullerian hormone.
POHL, T.M., ZIMMER, M., MacDONALD, M.E., SMITH, B., BUCAN, M.,
POUSTKA, A., VIOLINI, S., SEARLE, S., ZEHETNER, G.,
WASMUTH, J.J. et al. (1988) Construction of a NotI linking
library and isolation of new markers close to the Huntington’s
disease gene. Nucleic Acids Res. 16 9185-9198

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.

libraries: The next generation of molecular tools in mammalian
genetics. Trends Genet. 2 174-179.

human gene mapping: selection for fragments of the human Y
chromosome after chromosome mediated gene transfer.
EMBO J. 5 979-985

the limits of the human pseudoautosomal region and a candidate


RAPPOLD G.A., CREMER, T., CREMER, C., BLACK, W., BOREMBERGER, J. &
probes hybridizing predominantly to human sex chromosomes. Hum.
Genet. 64 110.

pseudoautosomal region by partial digest PFGE analysis from the
telomere. Nucleic Acids Res. 16 5361-5377.


A 45,X male with evidence of a translocation of Y euchromatin 
onto chromosome 15.  

SCHEMPP, W., MULLER, G., SCHERER, G., BOHLANDER, S.K., ROMMERSKIRCH,  
sequences and X chromosome replication studies in XX males.  

SCHERER, G., SCHEMPP, W., BACCICHETTI, C., LENZINI, E., BRICARELLI,  
segment that includes the ZFX locus causes sex inversion in man.  

SCHMIDTKE, J., ARNEMANN, J., SCHIM, H., BAUN, F., MAYEROVA, A.,  
isochromosome and presence of a Yp-specific DNA sequence.  
Hum. Genet. 69 135-137.

Cancer 25 1340-1356.

repeated sequences from hybridization probes.  
Nucl. Acids. Res. 3 1905-1922

SHAPIRO, L.J., MOHANDAS, T., WIESS, R. & ROMEO, G. (1979) Non- 
inactivation of an X-chromososome locus in man.  
Science 204 1224-1226

Sex chromosomes and reproductive anatomy of some intersex 
marsupials. J. Reprod. Fert. 21 57-68.


SINCLAIR, A.H., FOSTER, J.W. SPENCER, J.A., PAGE, D.C., PALMER, M.,
homologous to ZFY, a candidate human sex-determining gene, are

Int. rev. Cytol. 76 67-112.

is caused by the recurrent non-reciprocal cross over involving
the X and an aberrant Y chromosome. Cell 28 205-216.

associated nucleotide sequences in eukaryotes. Cold Spring Harb.


SMITH, C.L., MATSUMOTO, T., MIWA, O., KLEO, S., FLAN, J.-B., YANAGIDA,
Schizosaccharomyces pombe by pulsed field gel
electrophoresis. Nucleic Acids Res. 15 4481-4489.


restriction site mapping. Nucl. Acids. Res. 3 2387.

structures in microspread human spermatocytes.
Chromosoma 81 315-37.

SOUTHERN, E. (1975) Detection of specific sequences among DNA
fragments separated by gel electrophoresis. J. Mol. Biol. 98 503.


-193-

WAALWIJK, C. & FLAVELL, R. (1978) DNA methylation at a CCGG sequence in the large intron of the rabbit & globin gene: tissue specific variations. Nucleic Acids Res. 5 4631-4641


