Physical and Genetical Investigation of the Xp11.3 Region on the Short Arm of the Human X-Chromosome

Pia Ethena Wittwer

A thesis approved for the degree of
Doctor of Philosophy in Biochemistry
at the University of the Western Cape

2004

Department of Biotechnology
University of the Western Cape
Cape Town
South Africa

Abstract

Examination of the UBE1 Region on the Short Arm of the Human X-Chromosome at Xp11.3

Pia Ethena Wittwer

The University of the Western Cape

A thesis approved for the degree of Doctor of Philosophy in Biochemistsry

2004

The pattern of inactivation in the DXS8237E-UBE1-PCTK1 region is of particular interest, since the mechanisms of X chromosome inactivation and the escape from inactivation are, as yet, not fully understood. The inactivation status of the *DXS8237E* and *PCTK1* gene differ: the first undergoes normal inactivation and the second escapes this process. The status of the *UBE1* gene has been controversial, although it is widely excepted that it does escape X chromosome inactivation. Physical mapping of the region employing YACs and subsequently PACs has been undertaken, but was restricted in scope by the high frequency of rearrangements occurring. DNA sequences between *DXS8237E, UBE1, PCTK1* and the distal gene, *UHX1,* have been investigated with regard to LINE1 elements, which are thought to play a role in X-inactivation. The results obtained strongly suggest a link between LINE1 elements and X chromosome inactivation. Sequence analysis results also contributed to the understanding of difficulties with restriction mapping of the region.

Further, this work includes the first reported establishment of the *UBE1* exon-intron boundaries. Additionally, genomic sequence analysis showed that only 46kb separate *DXS8237E* from *UHX1,* which confirms that this region is extremely gene rich.

A new approach for detecting the inactivation status of *DXS8237E, UBE1* and *PCTK1* has been investigated. This has been carried out by employing the theory that active and inactive genes have different replication times. The results, although not 100% conclusive, indicate that the pseudoautosomal region (PAR) is distinct from the remainder of the X chromosome in that the genes in the PAR may replicate at earlier stages of the cell cycle and the latter region may replicate relatively late irrespective of their inactivation status.

Analysis of the association of the *UBE1* locus with quantitative trait loci (QTL) for cognition has been undertaken. Evidence present suggests that *UBE1* can be excluded as a QTL.

Acknowledgements

I would like to thank:

- Professor Don Cowan for giving me the opportunity to work in the Biotechnology Department at UWC

- Professor Chris Gehring for acting as my supervisor at UWC

- Professor John Edwards for making the facilities of the Genetics Laboratory in Oxford available to me

- Professor Ian Craig for his supervision at the University of Oxford and for the very enjoyable bonfire nights and dinners at his home. In this context, a big thank you to Sally Craig, too

- Joseph McClay for spending time to show me how to use various computer programs and Linzy Hill for her help with the work carried out in London

- Past and present members of the laboratory for helpful conversations and for creating an enjoyable working environment

- My friends, in England and abroad, who supported me in difficult times

- My parents for their unconditional love.

Work published

• Wittwer P.E., Hill L. Ninomiya T., Coleman M., Plomin R. and Craig I.W. Investigation of X chromosome-linked loci escaping inactivation as potential contributions to behavioural phenotypes. Abstract presented at the VIIth CEPH Annual Conference of Human Genetics (Paris, 1999).

• Wittwer P.E. and Craig I.W. Physical mapping of the UBE1 region. Abstract presented at the HGM'98 Human Genome Meeting (Turin).

To be noted

The work described in this thesis was carried out by myself, unless specifically stated.

Throughout the text of this thesis the names of genes are indicated in italic, whereas for e.g. the proteins they encode for are not. For a clearer presentation in some diagrams names of genes are not put into italics.

A few of the most commonly used abbreviations:

A= adenine

C= cytosine

G= guanine

T= thymidine

bp= base pair (s)

kb= kilobase (s)

nt= nucleotide (s)

ng= nanogram (s)

mg= milligram (s)

μg= microgram (s)

M&M = Material and Methods (of that particular chapter)

Keywords: X-chromosome, X-inactivation, UBE1, cognitive ablility, physical mapping, replication studies, ubiquitination, BrdU, HeLa cells

List of figures and tables

* If the figure is on an unnumbered page, the page on which the figure is mentioned, is listed

* If the figure is on an unnumbered page, the page on which the figure is mentioned, is listed

* If the table is on an unnumbered page, the page on which the table is mentioned, is listed

Table of contents

CHAPTER 1 - Introduction

As a consequence of the continual developments in molecular biology, the field of human genetics is progressing rapidly. This has not only helped to understand certain aspects of evolution but also, more importantly, the characterisation of genes responsible for single-gene defects and genes involved in complex diseases. This has lead to an improved understanding of the causes of multifactorial disorders. Since, on a very conservative basis, a relative high percentage of the population is affected by genetic disorders, advances in molecular genetics are likely to be of great importance with respect to gene therapy.

Some genetic diseases are caused by chromosomal aberrations. Normally there are 46 chromosomes in humans: 22 pairs of autosomes and a pair of sex chromosomes, either two X-chromosomes, or one X- and one Y-chromosome. There are two types of chromosomal aberrations. The first involves changes in the entire set of chromosomes. Since the development and functioning of the adult organism depends on the correct gene dosage, in general, the presence of fewer or extra sets of chromosomes results in a drastic perturbation of gene activities and development is aberrant. Polyploid individuals are therefore found only among aborted foetuses. The second affects individual chromosomes (examples are trisomy-21, also known as Down's syndrome; Turners syndrome, X0; or Klinefelter's syndrome, most commonly XXY), or part of single chromosomes (examples are inversions, translocations, duplications and deletions) (McConkey, 1993).

At a higher level of resolution, genes responsible for hereditary disorders were initially identified by understanding the biochemistry of a disease, a technique called functional cloning. This primarily depends on the availability of information about the protein product and/or function of the responsible gene. This enables the isolation of gene specific probes and the subsequent cloning and mapping of the locus. However, the functions of the loci responsible for the majority of single gene disorders are unknown. An alternative technique, known as positional cloning, became available about 20 years ago, where the localisation of a disease gene is essential in enabling its isolation and hence

determining its function (Collins, 1991). This approach uses genetic and physical mapping techniques to localise the position of genes and molecular cloning of appropriate regions to characterise and determine their involvement in normal development and in disease states. A number of genes have been cloned by this method, one of the first being the chronic granulomatous disease locus on the X chromosome (Royer-Pokora et al., 1986)

The work described in this thesis is concerned with a region of the X chromosome and is an analysis of the *DXS8237E-UBE1-PCTK1* region on the proximal short arm, at Xp11.3. It includes a physical map of the region, the isolation and examination of a novel polymorphism and studies of X-chromosome inactivation. An introduction to the X-chromosome, inactivation, the region under investigation and mapping techniques are outlined below

1.1 The X Chromosome

The X chromosome contains about 164Mb of DNA, equivalent to 5% of the human haploid genome and is present in two copies in females and as a single copy in males. The mechanism that leads to equal expression of genes on the X chromosome in females and males is called dosage compensation In placental mammals, this is achieved by random inactivation of one X chromosome n each cell during female embryogenesis (Lyon, 1961 Inactivation occurs randomly, affecting either the paternal or maternal X chromosome thereafter is clonally inherited (Davidson et al 1963). The inactivation process is accompanied by condensation; the tightly coiled, hypermethylated inactive X chromosome is visible as a body of heterochromatin in the nucleus, known as a 'Barr body' (Barr and Bertram, 1949) X chromosome inactivation is explained n more detail in section 1.2.

Genes found on the X chromosome are no more likely to be connected with sexual differentiation than are autosomal genes. Nevertheless, the inheritance pattern of diseases on the X chromosome differs from that of autosomal chromosomes Most X-linked disease genes that follow Mendelian inheritance are expressed in a recessive manner. This implies that the incidence of the

disease in males, the heterogametic sex, occurs at the same frequency as the allele. Females usually have a normal phenotype except when non-random X inactivation of the normal allele in the affected tissue results in a manifesting female carrier (Plenge et al., 1995). This pattern of carrier females and affected males who do not transmit the disease gene to their male offspring immediately identifies potential X-linked disease loci. The first gene to be given chromosomal assignment was the gene responsible for colour blindness, located on the X chromosome in 1911 (Wilson, 1911). It was not until 1968 that the first autosomal gene, the Duffy blood group, was assigned to chromosome 1 (Donahue et al., 1968). In addition, there are some X-linked dominant disorders such as spinal muscular atrophy and pyruvate dehydrogenase deficiency, which are usually lethal in males and are manifested in females (McConkey, 1993).

1.1.1 Proximal Xp disorders

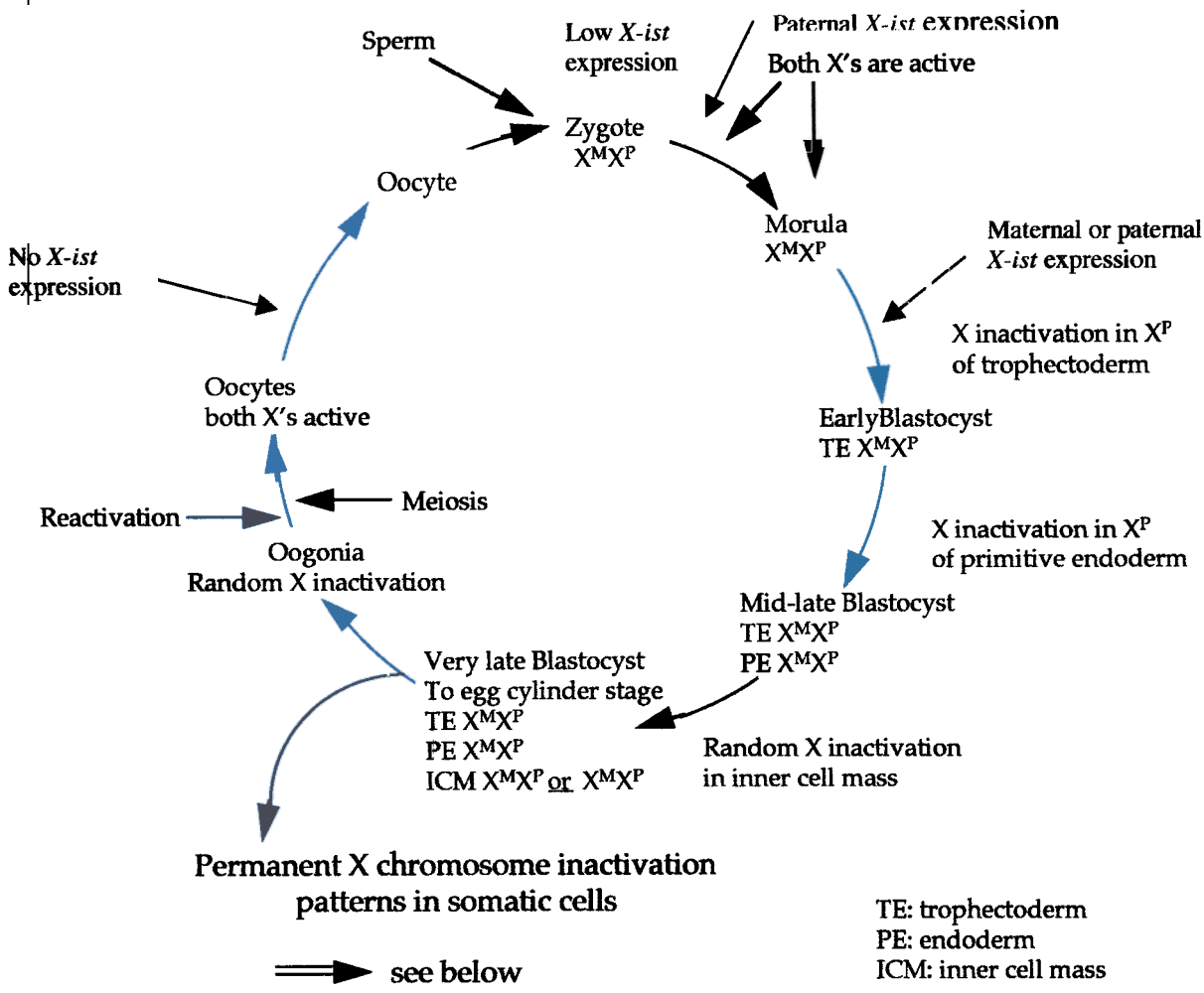
A variable number tandem repeat or VNTR (see mapping techniques) probe, M27β, which recognises the locus DXS255 at Xp11.22 has been used to link a number of disease genes to the proximal Xp including Norrie's disease (Chen et al., 1992), Åland Island eye disease (Alitalo et al., 1991, Schwartz and Rosenberg, 1991) and congenital stationary night blindness (Gal et al., 1989). Other disorders mapping to this interval include Retinitis pigmentosa (RP), a pigmentary retinal dystrophy that causes progressive visual disability and can result in blindness. A number of genes responsible for this disorder have been identified and some map to the X chromosome. One of the X-linked loci, namely RP3, lies at Xp21.1 (Schwahn et al., 1998). Another one, RP2, lies at Xp11.3 (Coleman et al., 1990; Coleman et al., 1993; Meitinger et al., 1989; Schwahn et al., 1998) and shows linkage to M27β. Both, RP2 and RP3 have been positionally cloned (Schwahn et al., 1998). Another disorder is Dent's disease, a renal tubular disorder. Analysis of deletion patients with Dent's disease and transcripts identified from a kidney cDNA library that map to this region has revealed a candidate gene for this disease (Fisher et al., 1994). In addition, X chromosome autosome translocations with clinical phenotypes which map to proximal Xp include synovial sarcoma (Gilgenkrantz et al., 1990), a number of breakpoints associated with incontinentia pigmenti 1 (Gilgenkrantz et al., 1985), Åarskog syndrome (Bawle et al., 1984) and mental retardation.

1.2. X Chromosome Inactivation

1.2.1 The cycle of X chromosome inactivation

Early investigations laid the foundations for our understanding of this topic. Prior to X chromosome inactivation in the female embryo, both X chromosomes are euchromatic and there is no dosage compensation (Adler et al., 1991; Epstein et al., 1978; Kratzer and Gartler, 1978). In mice the process of X chromosome inactivation firstly takes place in the trophectoderm in a non random fashion (Takagi, 1978; Takagi and Sasaki, 1975; West et al., 1977; West et al., 1978). It then continues in the primitive endoderm. Random X chromosome inactivation takes place later in the cells of the inner cell mass that will form the adult somatic cells of the female (Epstein, Travis et al., 1978; Kratzer and Gartler, 1978/2; Monk, 1978; Monk and Harper, 1979). X chromosomes in the germ line are inactivated in early embryogenesis and then go through a reactivation step approximately at the time of entry to meiosis (Gartler et al., 1975; Johnston and Cattanach, 1981; Kratzer and Chapman, 1981). Takagi (1974; 1978; Takagi et al., 1975) first noted that inactivation in non-embryonic tissues was non-random, with the paternal X chromosome being preferentially inactivated; this work was confirmed by others (West et al., 1977; West et al., 1978). They also noted that the replication asynchrony of the X chromosome in the trophectoderm was unusual in that the inactive X chromosome initiated replication earlier than the active X chromosome (Takagi et al., 1982). This pattern is the opposite of what occurs in regular somatic cells. These two observations indicate that the inactivation pattern in the trophectoderm and primitive endoderm is different to that in the inner cell mass. It has been shown that X chromosome inactivation takes place when cells differentiate (Martin et al., 1978; McBurney and Strutt, 1980). Since the oocyte carries an undifferentiated X chromosome, either the germ line must never undergo inactivation, or if it does, the X chromosome must undergo a reactivation event sometime during ontogeny. Various studies (Andina, 1978; Gartler et al., 1980; Gartler et al., 1975; Gartler et al., 1975; Hartung and Stahl., 1975; Jagiello et al., 1982; Kratzer and Chapman, 1981; Luciani et al., 1979; McMahon et al., 1981; Monk and McLaren, 1981; Ohno et al., 1962; Semenova-Tien-Shanskaja and Patkina, 1978; Witschi, 1957) indicate that the germ cells go through an inactivation-reactivation cycle (Gartler and Riggs, 1983). A

diagrammatic inactivation and reactivation cycle of the X chromosome modified from Gartler and Riggs (1983) is shown below:



A review of X chromosome inactivation during development (and other inactivation issues) can be found in the *Annu. Rev. Genet.*, 1997 **31**: 571-610 by E. Heard, P. Clerc and P. Avner.

1.2.2 Dosage compensation and the evolution of the sex chromosomes

In many organisms, the evolution of the X/Y sex chromosomes necessitated the co-evolution of mechanisms for equalising X-linked gene dosages between females with two X chromosomes and males with only one (Kelly and Kuroda, 1995). Dosage compensation in *Drosophila* is achieved in males by two-fold increased transcription of their single X. In nematodes dosage compensation is accomplished in females by down-regulation of X-linked gene transcription from both Xs. In mammals, dosage equivalence occurs by global transcriptional silencing of one X in females in a process along a 164Mb sequence (reviewed by

Lee and Jaenisch, 1997 (2)).

Our understanding of the organisation and behaviour of X and Y chromosomes has been guided by two influential ideas originally brought forward by Susumo Ohno (Ohno et al., 1962). The first idea was that the mammalian X chromosome has been conserved entirely because of selection against disruption X chromosome inactivation The other idea was that heteromorphic sex chromosomes evolved originally from an autosomal pair, driven initially by differences at a sex-determining locus. Questioning Ohno's hypotheses has led to a deeper understanding of the organisation, function and evolution of mammalian sex chromosomes.

The Y chromosome is a genetic and evolutionary puzzle In groups like mammals and *Drosophila*, which possess advanced systems of chromosomal sex determination, the Y chromosome lacks most of the thousands of genes carried on its pairing partner (the X chromosome) at meiosis. Comparative studies suggest strongly that the X and Y chromosomes were originally largely homologous, probably differing by only a few genes concerned with sex determination Degradation of most of the genes on the Y chromosome happened repeatedly, in dozens of independent lineages and now can easily observed by size comparison of the two chromosomes. But why has the chromosome become subjected to such attrition? A likely explanation is that it must have been an evolutionary response to the suppressed recombination between most of the X and Y chromosome, which allows them to evolve independently. The X chromosome can recombine freely in the homogametic sex, but the Y chromosome, being present in a permanently heterozygous state, cannot. The lack of genetic recombination means that natural selection is less effective in preventing the accumulation of deleterious mutations or promoting the spread of advantageous ones (Charlesworth, 1998).

At this point, it should be noted that on each of the human sex chromosomes are segments of 2.6 million base pairs on the distal tips of the short arms called the pseudoautosomal region 1 (PAR1), which have a high degree of sequence homology and where, contrary to the usual prohibition, crossing-over during male meiosis is not only allowed, but required for successful segregation of the X and Y chromosomes. PAR1 genes are present in two doses in both males and

females and they escape inactivation. Similarly, a second and smaller P (PAR2) exists on the distal tips of the long arms (Ellis, 1998; Graves et al., 1998).

Y attrition and X inactivation

Genes with copies only on the X chromosome are subject to inactivation, whereas most genes with active alleles on the Y chromosome are not (Disteche, 1997; see figure 1.2.1a). The correlation between gene dosage and inactivation suggests that progressive degradation of the Y chromosome has been offset by the spread of inactivation along the X chromosome. Thus inactivation has more-or-less kept pace with Y chromosome attrition. However, there are some exceptions to the correlation between loss of an active Y copy and inactivation of the X copy, and these reveal the process by which Y chromosome degradation and X chromosome inactivation co-evolved.

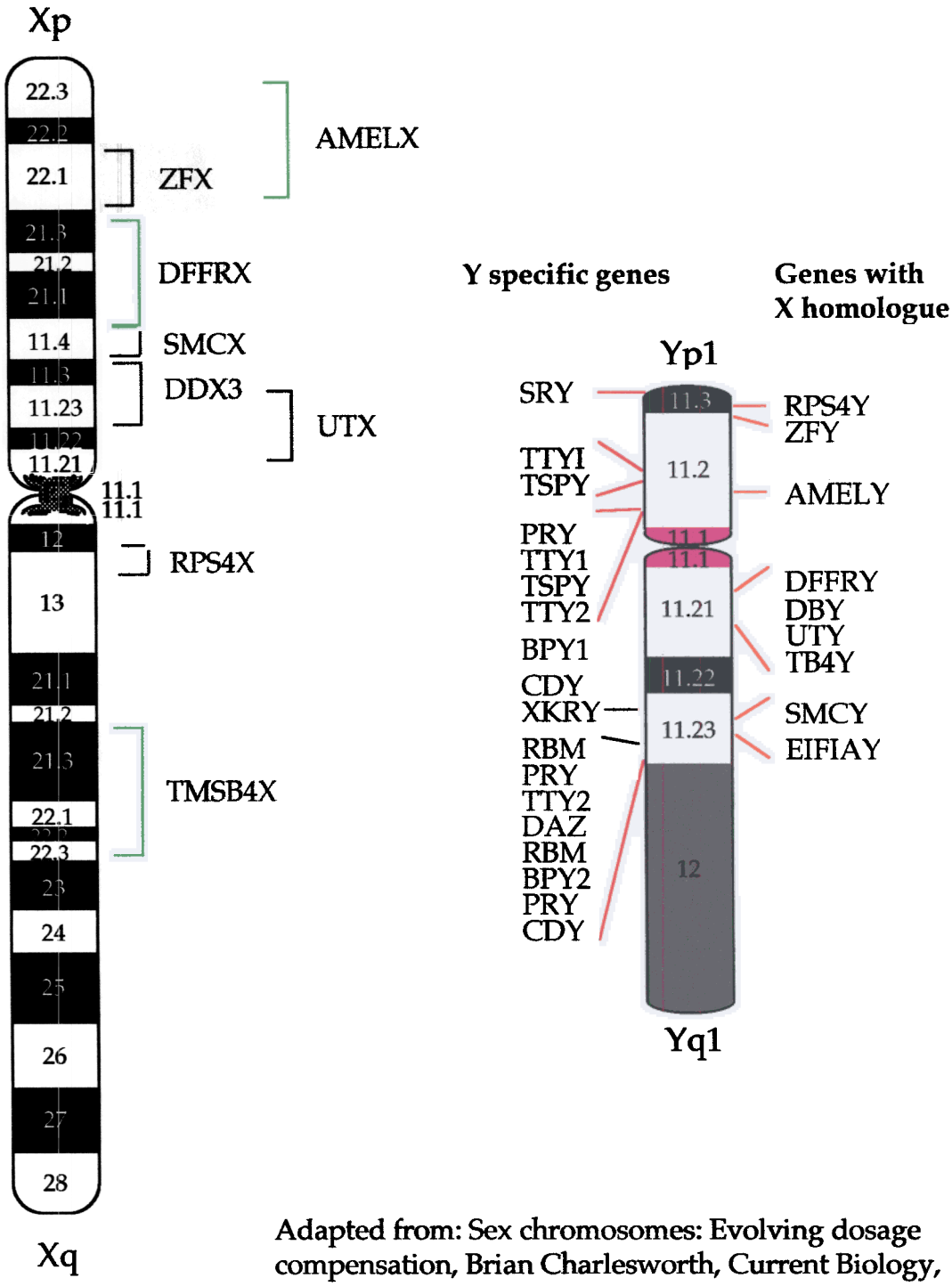
It is almost universally supposed that Y attrition was the primary event, which forced selection for recruitment of the X-linked homologues into the inactivation system. Evidence can be found for several human X chromosome linked genes which escape inactivation although they have no active allele on the Y chromosome. One example is the human *UBE1X* gene (see Chapter 1.3), which escapes inactivation although there is no *UBE1Y* on the human Y chromosome (Mitchell et al., 1998). The presence of an active Y homologue of this gene in other species, such as mice (Mitchell et al., 1991; Kay et al., 1991), suggests that loss of *UBE1Y* from the human Y chromosome was a recent event and there has not been sufficient time for X chromosome inactivation to "catch up' with the Y chromosome attrition.

The alternative possibility that X chromosome inactivation could have been the primary event is suggested by the finding of several genes on the mouse X chromosome which are subject to inactivation although they have active homologues on the Y chromosome. One example is the *Zfx* gene (Adler et al., 1991). However, the Y alleles of at least some of these may have acquired a male-specific function A diagram of the human Y chromosome, its specific genes and those with a X chromosome homologue is shown in figure .2.1b

Fig. 1.2.1 The organisation of the Y chromosome

a. showing genes that escape X inactivation with homologues on the Y chromosome

b. showing Y specific genes and genes with a homologue on the X chromosome



Adapted from: Sex chromosomes: Evolving dosage compensation, Brian Charlesworth, Current Biology, 1998, 8:R931-R933

In this context another observation is of interest: in mammals X chromosome inactivation is random in the embryo proper (see the diagrammatic presentation of the inactivation and reactivation cycle of the X chromosome above). However, in marsupials and the extra-embryonic region of the mouse it is imprinted: the paternal X chromosome is preferentially inactivated whereas the maternal one is always active (reviewed by Wang et al., 2001). Therefore random X-inactivation might have evolved from parental X-inactivation or alternatively, expression of the imprinted autosomal loci was originally random and became coordinated and parent specific with the evolution of controlling sequences analogous to the XIST gene (Ohlsson et al., 2001).

1.2.3 The mechanism of X chromosome inactivation

The major genetic locus proposed to control the X chromosome inactivation process is the X inactivation centre (XIC). XIC is defined as a region of the chromosome from which a currently ill-defined signal exerts its effect in *cis* along the chromosome; derivative X chromosomes lacking this XIC are unable to become inactivated (Lyon, 1996). The human XIC has been localised to a < 1Mb region within band Xq13.2 (Brown et al., 1991).

A gene designated *XIST* (in human) and *Xist* (in mouse) has been mapped to the XIC region (Brown et al., 1991; Borsani et al., 1991; Brockdorff 1991) and has been shown to be both necessary and sufficient for X inactivation to occur (Penny, Kay et al., 1996; Lee et al., 1996; Lee et al., 1997(1); Herzing et al., 1997; Marahrens). Its name originates from the fact that it is expressed exclusively from the inactive X chromosome: X-inactive specific transcript (Brown et al., 1991). Notably, the product of the *XIST* (and the *Xist*) gene is a noncoding RNA that remain associated with the inactive X chromosome in female interphase ucle In this RNA is 15-17kb in length and its expression coincides with the inactivation (Brown et al., 1992; Brockdorff et al., 1992; Clemson et al 1996) humans the *XIST* cDNA is said to be 19.3kb of length (Hong-Young et al., 2000). The onset of *Xist* expression precedes X inactivation and is therefore a likely cause rather than a consequence of X inactivation (Brockhoff et al 1992; Kay et al., 1994). Further evidence that *Xist* is necessary for X inactivation to occur came from mutation analysis in female embryonic stem cells carried out by Penny co-workers (1996)

Recently Migeon et al. (1999) have transfected male mouse embryonic stem (ES) cells with a YAC transgene containing 480kb of the putative XIC. Their results show that DNA included in the XIC transgene is sufficient to initiate random X inactivation, even in cells of a different species.

X chromosome inactivation occurs in 4 stages: 1. counting of chromosomes; the initiation of X chromosome inactivation; 3. the propagation of heterochromatin throughout the X chromosome and 4. the maintenance of the inactive state through subsequent mitotic divisions.

1. 'Counting'

In all mammals examined to date, the 'n-1 rule dictates that a single X remains active per diploid set of autosomes, such that XY males have no inactive X, XX females have one inactive X and XXX individuals have two inactive chromosomes etc. A favoured explanation is the presence of a mechanism in which a diploid cell produces a single 'blocking factor' that inhibits Xic action (Lyon, 1996). All other X chromosomes containing unblocked Xics are inactivated in order to maintain one active X chromosome per diploid genome. Studies of *Xist* knockouts (Penny, et al. 1996; Marahrens, et al., 1997) suggest that Xist RNA does not participate in the process of counting, although molecular details of this process are unknown. It may be possible though, that the distinct functions are encoded by an Xist transcriptional regulator. A 2.4kb CpG island lying 15kb downstream of Xist (Courtier et al., 1995) may act as an enhancer to which the hypothesised blocking factor binds, precluding the Xist induction requirement for X inactivation. Consistent with this model, the 2.4kb CpG island is hypermethylated on the active X chromosome (Lee and Jaenisch, 1997(2)).

2. Initiation

Xist induction represents the first measurable step of X chromosome inactivation. Two, not necessarily mutually exclusive models have been proposed to explain how *Xist* expression initiates heterochromatin formation at the Xic (Brockdorff et al., 1992; Brown et al., 1992). The first proposes that *Xist* produces a functional RNA which acts in *cis* to recruit silencing factors to the X

chromosome. The second model states that the critical role of Xist lies not in its RNA, but in its chromatin structure. The act of *Xist* transcription sult n conformational changes at the Xic which enable binding of heterochromatisi factors.

Three observations support the first model (Bead et al. 1995; Clemson et al. 1996 Marahrens et al., 1997; Tai et al., 1994). Firstly, in undifferentiated ES cells *Xist* is transcribed without X chromosome inactivation, arguing that transcription might not be sufficient for chromosome silencing (Bead et al 1995; 1994). Secondly, in differentiated cells, *Xist* RNA is expre abun anti associates with the inactive X chromosome and contributes to the nuclear matrix, implying that the critical factor for inactivation may be the gene product (Clemson et al., 1996). Thirdly, *Xist* transcription is not disrupted by an internal *Xist* RNA deletion which abolishes X chromosome inactivation (Marahrens et al., 1997). However, none of these arguments are conclusive. More importantly as the knockout necessarily deletes *Xist* DNA, chromatin structure may play role in the initiation process (Lee and Jaenisch, 1997(2))

Recently a gene antisense to *Xist* has been reported by Lee et al. (1999). *Tsix* is defined as a 40kb RNA originating 15kb downstream of *Xist* and being transcribed across the *Xist* locus. The *Tsix* RNA has no conserved open reading frames- it is seen exclusively in the nucleus. Before the onset of X chromosome inactivation *Tsix* is expressed from both X chromosomes At the onset of X inactivation *Tsix* is expressed from the active chromosome and persists until *Xist* is turned off and X inactivation is established. It is not found on the inactive chromosome once cells become inactivated. The authors conclude that the *Tsix* possesses features suggesting a role in the early regulation of X inactivation, such as counting of chromosomes or initiation of X chromosome inactivation.

3. Propagation and establishment

Molecular details of how silencing is communicated throughout the X chromosome are virtually unknown. The idea that X chromosome inactivation 'spreads' bi-directionally from the Xic originated from two sources. Firstly, observations in mammalian X-autosome translocations showed that inactivation can extend into the autosome in a mosaic distance-limited fashion

Secondly, analogous studies of *Drosophila* position effect variegation and yeast telomeric silencing. No definitive silencing factors have been identified for X chromosome inactivation; however, the *Xist* RNA may emerge as such a factor since it is a *cis*-acting product of the Xic with the ability to coat the entire X chromosome (Lee and Jaenisch, 1997(2)). The RNA appears to bind the human inactive X chromosome in interphase (Clemson et al., 1996), as well as to the mouse inactive X chromosome during mitosis, suggesting that the *Xist* gene product contributes to the chromosome structure.

In various studies DNA replication and histone acetylation have been directly linked to gene expression. As X chromosome inactivation is accompanied by delayed replication timing and histone deacetylation, the relationship of *Xist* to these events was addressed by Keohane et al. (1996, 1999). The authors found that in differentiating female ES cells, the appearance of a late-replicating X chromosome coincided with the onset of *Xist* induction. This was followed temporally by the appearance of X chromosomes with hypoacetylated histones and by a two-fold reduction of X-linked gene expression which indicates X chromosome inactivation (Keohane et al., 1996, 1999). The fact that a late-replicating, hypoacetylated X chromosome did not occur prior to *Xist* induction is consistent with *Xist*s role as an initiator of X-heterochromatin. Whether changes in DNA replication timing and histone acetylation are prerequisites or consequences of gene silencing, however, remains to be addressed.

Recently Lyon (1998) and Bailey et al. (2000) suggest that a non random organisation of long interspersed element (LINE) repetitive sequences on the X chromosome might be connected to X heterochromatin formation/ X chromosome inactivation. In 1990 Riggs put forward a concept that "booster" elements which were concentrated at the X inactivation centre and other positions throughout the chromosome served to amplify and spread the X inactivation signal along the X chromosome. Lyon (1998) proposed LINE-1 (L1) as a candidate for those "booster" elements. Bailey et al. (2000) detected an almost 2 fold enrichment of L1 elements on the X chromosome compared with other human autosomal chromosomes. The most significant observations the authors made were a big increase in L1 elements for sequences in Xq13, which contains XIC, and a reduction of L1 elements at genomic loci escaping X inactivation

compared to loci subject to X inactivation. Therefore it was concluded that L1 elements may serve as DNA signals to propagate X inactivation along the X chromosome

4. Maintenance

Mechanisms must exist to maintain memory of the inactive state, since X chromosome inactivation is both stable and clonally heritable through thousands of mitotic divisions. Maintenance almost certainly requires self-propagating chromatin states, none of which are unique to X chromosome inactivation. DNA methylation is stable and clonally heritable by the action of a maintenance methylase which acts preferentially on hemimethylated substrates of DNA replication. As the CpG islands of the inactive X chromosome are hypermethylated, methylation has long been proposed as a memory mechanism. Maintenance could also be achieved by the propagation of stable transcription complexes on the active X chromosome and of silencing factors on the inactive X chromosome (Lee and Jaenisch, 1997(2))

Do *Xist* and Xic play a role in the maintenance of X chromosome inactivation? One fact is that the inactive X chromosome appears to be stable in the absence of the Xic and *Xist* (Brown and Willard, 1994; Rack et al. 1994). But, if *Xist* and Xic were not required beyond establishment of X heterochromatination, why is the 15-17kb RNA continuously expressed through adult life? Maintenance could be solved by somatically imprinting the *Xist* gene. If the initial choice of which X chromosome to inactivate were accompanied by a physical mark on *Xist*, t daughter cells could faithfully retain the X inactivation pattern by allelic expression DNA methylation could act as such a mark. In mice, where X chromosome inactivation in extra-embryonic tissues is paternally imprinted, *Xist* is uniquely hypomethylated on and expressed from the paternal allele (Norris et al., 1994). Furthermore, parental imprinting correlates with *Xist* demethylation in the male germline and *Xist* is hypermethylated in the female germline (Ariel et al., 1995; Zuccotti and Monk, 1995). Other data (Bead et al., 1995; Panning and Jaenisch, 1996) suggest that *Xist* is regulated by DNA methylation and that X chromosome inactivation can occur with *Xist* induction in the absence of DNA methylation. One possible explanation for this could be that heterochromatin can be established without methylation, but that

methylation is subsequently required to 'lock in' a heterochromatic structure.

Maintenance of X inactivation through subsequent cell divisions is thought not only to be associated with methylation of CpG islands, but also the absence of acetylated histone H4, chromatin condensation and allocyclic replicatio 1992; Gartler et al., 1992)

In summary, the XIC is required for both initiation and spread of X chromosome inactivation (Rastan and Brown, 1990). The manner in which inactivity spreads along the selected X chromosome remains unclear *Xist* induction and X silencing were observed in only 5% of differentiating ES and embryor (Panning and Jaenisch, 1996). Demethylation neither induces *Xist*, nor h ffect on X chromosome inactivation in the vast majority of cells Jaenisch, 1997(2)). Thus, the issue of maintaining the inactive X chromosome and whether it involves somatically imprinting *Xist in vivo* needs further investigation. Further discussion on these points has been presented by Panning and Jaenisch (1999).

1.2.4 Escape from X chromosome inactivation

Genes that escape X inactivation were first discovered in humans Those first found were located at or near the pseudoautosomal region within band Xp22.3. The locus that determines the XG blood group was the first gene to be shown to escape X inactivation, on the basis of the expression of XG polymorphism (Fialkow, 1978). The lack of inactivation of pseudoautosomal genes is expected since genes have functional X and Y homologues that combine in male meiosis. Other pseudoautosomal genes shown to escape inactivation include which encodes the cell surface antigen CD99 (Goodfellow, 1984), *XE7*, a gene of unknown function and the gene encoding ANT3 ADP/ATP (adenine nucleotide translator 3) (Slim, 1993; Schiebel, 1993). For genes located near the X chromosome pseudoautosomal region, the lack of inactivation was attributed at first to a 'position effect" One early example of a gene located near the pseudoautosomal region that shows partial escape from X chromosome inactivation was the human steroid sulfatase (*STS*) gene (Shapiro et al. 1979). Its Y homologue (*STSY*) is a non-functional pseudogene. However, other human genes soon were found that escaped X chromosome inactivation despite being

located at considerable distances from the pseudoautosomal region. One example is *ZFX*, which encodes zinc finger protein, X-linked, at Xp22.1 (Schneider-Gädicke et al., 1989). A second example is *RPS4X* (ribosomal protein S4, X-linked) at Xq13 (Fisher et al. 1990). Like pseudoautosomal genes, both *ZFX* and *RPS4X* have functional Y homologues, *ZFY* and *RPS4Y*, respectively.

A comparison of mouse and human X chromosomes has shown a difference in the X chromosome inactivation status. For example the *Zfx*, *Rps4* and *Ube1* genes are subject to inactivation in mice, but not in humans (Adler et al., 1991; Ashworth, 1991; Kay et al., 1991; Zinn et al., 1991). Nevertheless, there are genes in mice that escape the inactivation process. For example *Smcx* (or *Xe169x*) escapes X chromosome inactivation as does its human homologue and has a widely expressed Y chromosome homologue (Aguinik et al., 1994; Wu et al., 1994; Wu, Salido et al., 1994). The existence of genes that escape X inactivation has long been suggested on the basis of the abnormal phenotype of individuals lacking one X chromosome who have Turner syndrome (Lyon, 1992). Differences in the inactivation status of human and mouse genes may explain the phenotypic differences between humans and mice lacking one X chromosome (Ashworth, 1991; Zinn et al., 1991). XO mice have a near-normal phenotype while the 45,X human female has an abnormal one including short stature, lymphoedema, ovarian failure and severe foetal loss (Turner, 1938). Now we know that haploinsufficiency of genes that escape X inactivation in humans explain at least part of the abnormal phenotype of individuals with Turner syndrome (Zinn et al., 1993). The *PHOG* gene was isolated by Ellison et al. (1997) from the, in Turner syndrome individuals deleted, PAR1 region and is proposed to be a candidate gene for involvement in the short stature of Turner syndrome. Its inactivation status is not known, but, because of its location, thought to be likely to escape X chromosome inactivation and thereby have an altered dose in Turner patients. The authors have also found that the mouse homologue is autosomal, which may explain the lack of growth abnormalities in mice.

It should be noted that patients with more than three X chromosomes (pol) are phenotypically female and show normal secondary sexual characteristics, although they often suffer from mental retardation (de la Chapelle

Genes that escape X inactivation in humans, but have no Y homologue may be important in as yet unidentified female-specific and dosage-sensitive functions involved in processes such as ovarian development that are compromised in Turner syndrome. In males, the Y chromosome protects against features of Turner syndrome (Ellison et al 1997) and many of the genes involved are therefore likely to be those with functional Y homologues. However, certain genes that escape X chromosome inactivation have no functional chromosome homologues and are therefore expressed at a higher level females than in males, suggesting that gene dosage may not be a critical factor for all X-linked genes or that they may contribute to male female dichotomy. Examples are the *UBE1* gene (see Chapter 1.3), the *SB1.8* (*DXS423*) gene and the more recently discovered *INE1* and *INE2* genes (Esposito et al., 1997).

Genes that escape X chromosome inactivation and are not pseudoautosomal tend to be clustered in certain regions of the X chromosome; for example, the proximal short arm of the human X chromosome. Although this may simply reflect the fact that relatively few genes have been studied, it may also indicate that certain regions of the X chromosome are more susceptible to escaping inactivation, perhaps as a result of the evolutionary relationships between the sex chromosomes. The majority of such transcripts mapping to the short arm of the X chromosome, Xp, implies that a genetic imbalance of Xp may be more severe clinically than imbalance of the long arm of the X chromosome, Xq. A complete X chromosome inactivation profile will provide important information for medical genetics and an insight into the role that the genomic organisation of the X chromosome plays in the regulation of epigenetic silencing on the chromosome. So far, *XIST* is the only gene escaping inactivation which expressed from the inactive and silent on the active X chromosome.

Despite the clustering of genes that escape X chromosome inactivation, they are interspersed with genes that are subject to inactivation. Two explanations can be put forward. First, the spreading of X inactivation may skip genes that lack appropriate inactivation signals. Secondly, some genes may lack appropriate signals for maintaining their initial X chromosome inactivation and consequently become reactivated during embryogenesis. The correct explanation of the two remains to be clarified. The inactivation status of genes on the X

chromosome is shown in figure 1.2.2.

As yet the boundaries between genes that are subject to X chromosome inactivation and those that escape this have not been completely characterised. By examining X inactivation boundaries essential clues to the mechanism of spreading and to the question of whether regulation is at the level of individual genes could be obtained.

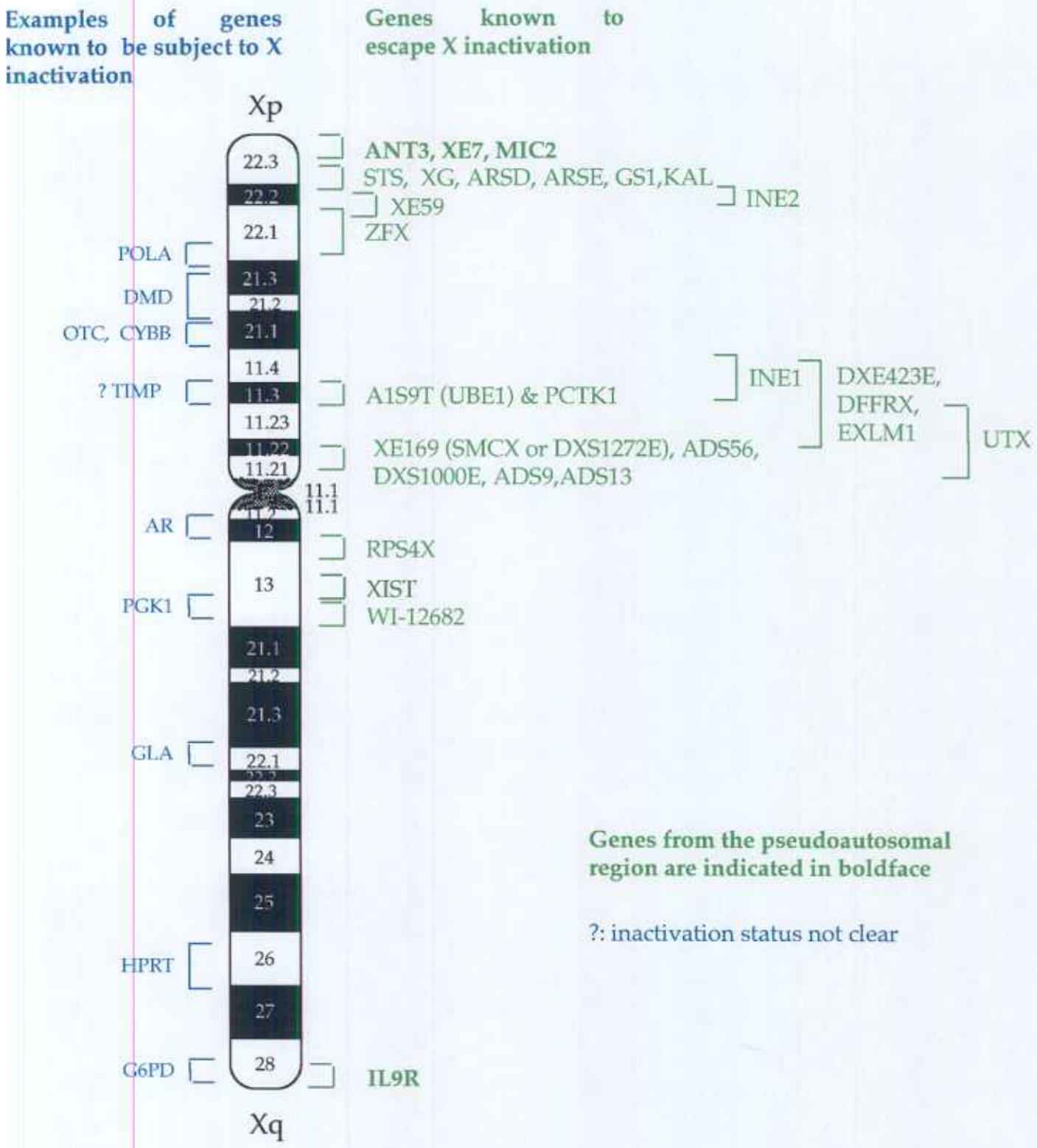
1.3 The region under investigation

1.3.1 Localisation of the *UBE1* (*A1S9T*), *PCTK1* and *DXS8237E* genes

UBE1 is an X-linked gene with a distinct Y-linked homologue in many eutherian (placental) and metatherian (marsupial) mammals which indicates that the gene has been conserved on the mammalian Y chromosome since the divergence of the eutherian and metatherian linages, over 130 million years ago (Mitchell et al., 1991; Mitchell et al., 1992). In mice a homologue of *UBE1* (85% nucleic acid identity) was cloned from the Y chromosome (Mitchell et al., 1991; Kay et al., 1991). Equally, in squirrels and new world monkeys a *UBE1* gene on the Y chromosome was identified. Nevertheless, no *UBE1* homologue is detectable on the human, chimpanzee, old world monkeys or marmoset Y chromosome (Mitchell et al., 1998). The authors conclude from their findings that *UBE1* has been lost during evolution of the primate lineage and that it therefore illustrates the key steps of 'autosomal to X-specific' evolution of genes on the sex chromosomes.

As with the mapping of the X chromosome in general, somatic cell hybrid genetics also played an important role in localising the *UBE1* gene and in assigning markers to the physical map of this region. Using somatic cell hybrids, Brown et al. (1989) localised the *A1S9T* gene (shown to be the same locus as *UBE1*) on the short arm of the X chromosome, Xpter-Xp11.1. Using Southern blot and *in situ* hybridisation, Zacksenhaus et al. (1990) located it to Xp 11.2-p11.4 and finally it was assigned to Xp11.3-p11.23 using high-resolution fluorescence *in situ* hybridisation (FISH) by Takahashi (1992). Kwan et al. (1991) localised the Wiskott-Aldrich syndrome (*WAS*) gene between OATL1 and GATA in the

Fig. 1.2.2 Ideogram of Inactivation status of genes on the X chromosome



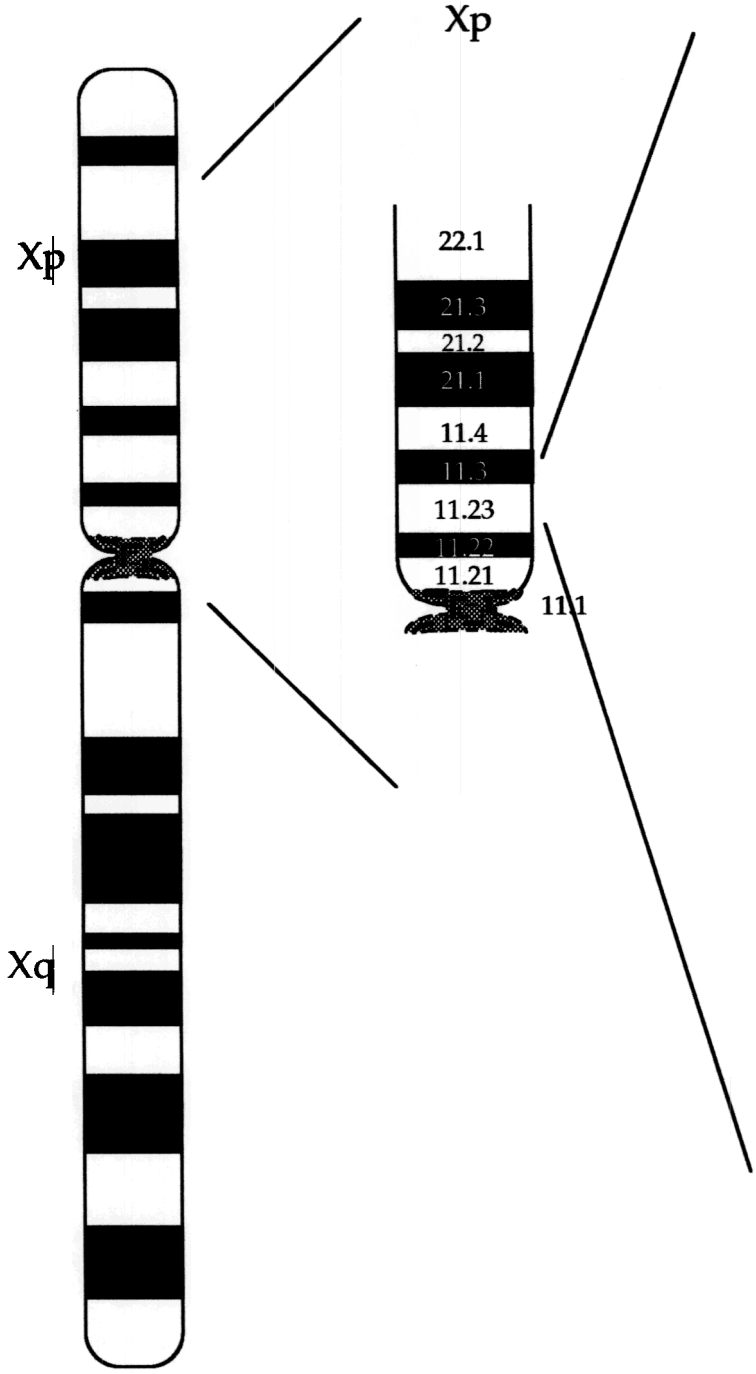
region of Xp11.23. This enabled the up-date of the physical map from *UBE1* to the GATA locus. Knight et al. in 1995 positioned the *PCTK1* gene, coding for a cdc2-related protein kinase (PCTAIRE-1) of unknown function, to the Xp11.3-p11.23 region. *PCTK1* is located within 420kb of the *UBE1* gene and both genes are highly conserved through evolution (from yeast to man). More recent studies have shown that *PCTK1* maps within 5kb of *UBE1*. Comparative mapping studies of the homologous loci in mouse established that *Ube1-x* and *Pctk1* are also within close physical proximity on the murine X chromosome (Carrel et al., 1996). Distal to *PCTK1* lies the *UHX1* gene at Xp21.1-p11.2. Coleman et al. (1996) mapped the 3' end of a novel gene, *DXS8237E*, isolated from human foetal brain cDNA, within 20kb upstream of *UBE1* in Xp 11.23. The corresponding 3kb mRNA is widely expressed in human cell lines and in mouse tissues. A detailed map showing some of the markers in this region is illustrated in figure .3.

1.3.2 Ubiquitin-activating enzyme 1 (UBE1), formerly A1S9T

Ubiquitination, the post-translational, covalent modification of various cellular proteins, is involved in different cellular activities, including DNA repair, cell-cycle control (Grenfell et al., 1994), peroxisome biogenesis and stress response (Swanson et al., 1996). One of the best understood functions of ubiquitin is its role in targeting proteins, for example oncoproteins (Ciechanover et al., 1991), for degradation. Additionally, reversible ubiquitination may play a role in regulating protein function, for example, IgE receptors are poly-ubiquitinated directly after antigen induced binding and are de-ubiquitinated once the receptors have dissociated (Paolini and Kinet., 1993).

The conjugation event occurs in the ubiquitin-dependent pathway. The first reaction is the activation of ubiquitin, a 76 amino acid polypeptide (Haas and Siepmann, 1997) to a high energy intermediate. It is catalysed by UBE1 and takes place in two steps, during which ubiquitin first forms an adenylate intermediate and then is transferred to a thiol site. In the second reaction it is transferred to one of the ubiquitin carrier proteins (E2), where it forms a similar thioester linkage. Now the ubiquitin can either be linked to a target protein or conjugat to proteins destined for degradation by ubiquitin-protein ligases (E3s). Hence, UBE1 provides the initial activated form which is necessary for any conjugation

Fig. 1.3
Markers on the short arm of the X chromosome in the region Xp11.3-p11.23
adapted from the sixth international X chromosome workshop 1995 (Banff)



DXS742
UHX1
PCTK1
UBE1
DXS8237E
ZNF157
DXS1264
DXS1055
DXS1003
DXS1146
DXS337
CXS1266
ARAF1
SYN1
TIMP1 — ZNF41
PFC — **(DXS1004E)**
DXS426
ELK1
DXS1367
ZNF81
ZNF21(DXS6849)
DXS267
DXS6616
OATL1
DXS7465E
DXS6941
DXS1011E
WASP (WAS/THC),DXS6940
DXS7466E — **DXS722**
DXS7467E DXS1358
GATA1
DXS226
HRASP
DXS1126 DXS1240
DXS7469E
DXS1470
TEF3
SYP**DXS1007E**
DXS7468E
DXS573
DXS1039

genes of interest to this work

disease and other expressed genes

event to take place.

The activation process is catalysed in the following steps (Handley et al., 1991):

$E1 + Ub + ATP \quad <=> \quad E1_{AMP\sim Ub} + PP_i$

$E1_{AMP\sim Ub} \quad <=> \quad E1^{S\sim Ub} + AMP$

$E1^{S\sim Ub} + Ub + ATP \quad <=> \quad E1^{S\sim Ub}{}_{AMP\sim Ub} + PP_i,$

where Ub = ubiquitin, S = thioester and PP_i = pyrophosphate

The cloning and sequencing of the cDNA for the human *UBE1* has been reported by Handley et al., (1991). The cDNA recognised a single 3.5kb E1 message that was ubiquitous among tissues and cell lines studied. The clone predicted a 110,450 dalton (D) protein and *in vitro* translation of the mRNA yielded a major product of approximately 110D.

1.3.3 Inactivation status of the genes in the UBE1 cluster

The inactivation status of *UBE1/ A1S9T* has proved controversial. Employing somatic cell hybrids containing an inactive X chromosome Brown and Willard (1990) determined that the *A1S9T* gene escapes X inactivation. In contrast, Zacksenhaus and Sheinin (1990) showed that the expression of the *A1S9T* gene in human diploid cells containing 3 or 5 inactive chromosomes was the same as in cells carrying no inactive chromosome and therefore concluded that it did not escape inactivation. One possible resolution of this controversy is that the *A1S9T* gene undergoes partial reactivation, because of the instability of X inactivation in somatic cell hybrids. Another explanation could be that *UBE1* behaves similar to the *REP1* gene, which in some cell lines is inactivated and in others escapes X chromosome inactivation (Carrel and Willard, 1999). Anderson and Brown (1999) have shown a gene thought to be inactivated, *TIMP1*, was expressed in some, but not all inactive X-containing somatic-cell hybrids. Their finding suggests that *TIMP1* is either prone to reactivation or variable in its inactivation. However, this is not the case for experiments *in vivo* (Bressler et al., 1993) and it was confirmed that the *UBE1* gene did escape X inactivation in humans, but that in mice it was subject to X inactivation (Bressler et al., 1993, Carrel et al., 1996). In earlier studies, Mitchell et al. (1991) and Kay et al. (1991)

had already demonstrated homology of a candidate spermatogenesis gene, *Sby*, on the mouse Y chromosome to the *Ube1* gene on the X chromosome. Kay et al. (1991) isolated part of the mouse *A1s9* gene, mapped it to the proximal portion of the X chromosome and also detected 2 copies of the gene on the short arm of the mouse Y chromosome. *A1s9X* undergoes normal X-inactivation, whereas *A1s9Y1* is expressed in testis. The two copies of this Y-liked gene may form part of a co-regulated group of genes which function during spermatogenesis. A second example of a gene which in mouse has a X and Y homologue and has a X but no Y homologue in humans is the eIF-2γ structural gene (Ehrmann et al 1998)

It should be noted at this point that the proximal gene, *DXS8237E*, does not escape inactivation (Coleman et al., 1996), but the distal *PCTK1* does (Carrel et al., 1996). Expression studies of the *Pctk1* gene (homologous locus in mouse) indicated that, similar to *Ube1-x*, it is subject to X inactivation in mouse. Methylation of CpG residues at restriction sites at the 5' end of both genes on the murine inactive X chromosome is consistent with both genes being subject to inactivation, in contrast to their expression status in humans (Carrel et al. 1996). The inactivation status of another gene on the X chromosome encoding a ubiquitin C-terminal hydrolase, *UHX1*, located distal, but in the close vicinity to *PCTK1* (Brandau et al., 1998), is not known. The investigation of the inactivation status of *UBE1* and closely linked loci would therefore be of considerable interest.

1.4 Other Members of the Ubiquitin Family and Ubiquitin-like proteins

As stated before, UBE1 generates a high-energy thiolester intermediate with ubiquitin in the ubiquitin-conjugating machinery. This process involves an internal Cys residue.

1.4.1 UBC (ubiquitin-conjugation enzymes) or E2 enzymes

The activated ubiquitin is transferred from E1 to a Cys residue of an E2 enzyme, thereby generating yet another thiolester intermediate. The genome of *Saccharomyces cerevisiae* encodes for 13 E2s and E2-like proteins and many more have been described in mammals. Some E2s are involved in specific cellular

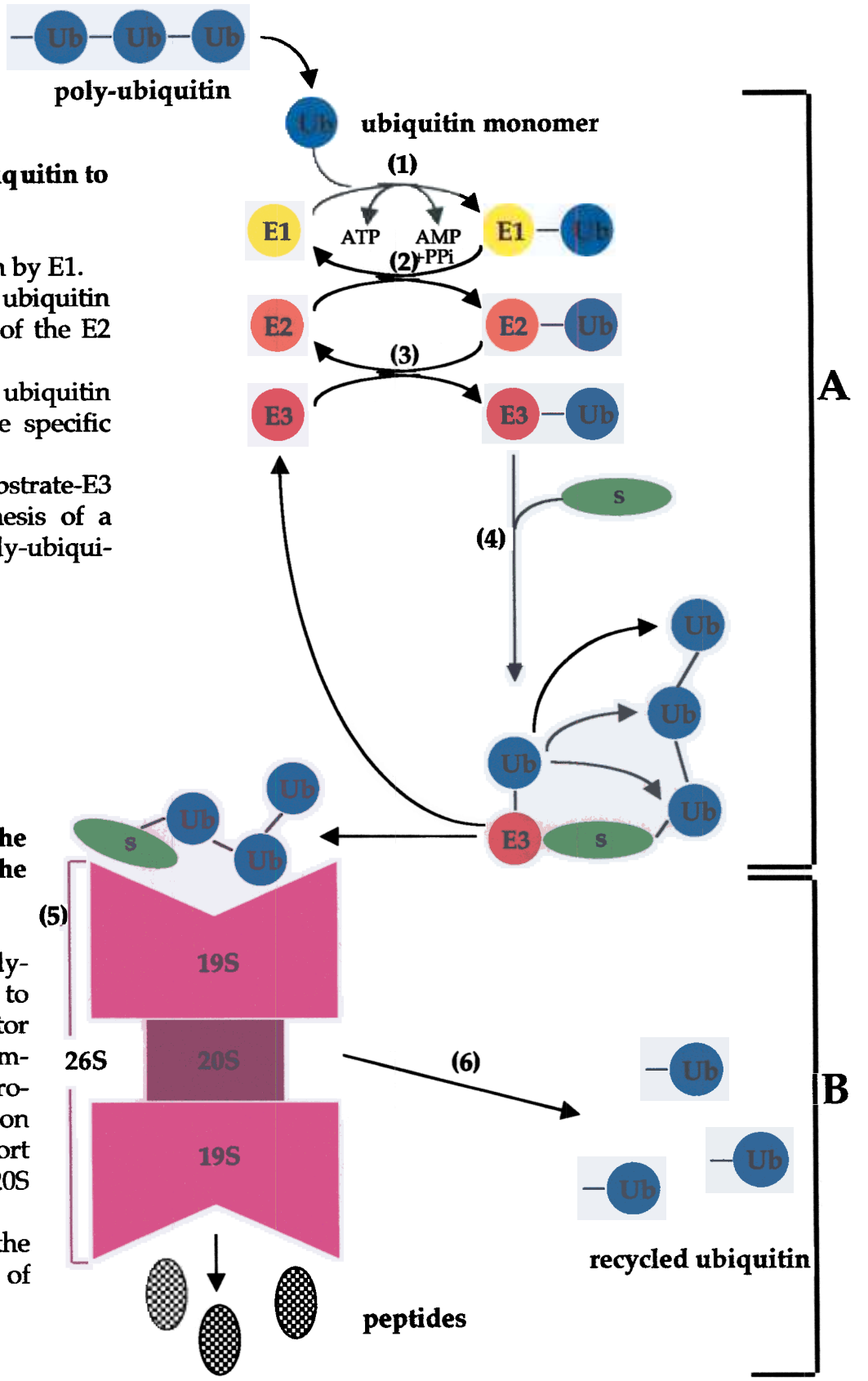
processes while the role of others is still obscure However, they all a he function as UBCs: all are inactivated by mutatio of h (Ciechanover, 998) Due to specific effects of certain 2s o define has been proposed that they can interact directly with the substrate protein While such interactions have been described in using protein-protein interaction screening methods, their physiological significance is not clear. Most probably the specific functions of E2s are due to their association with distinct E3s (Hershko and Ciechanover, 1998).

1.4.2 Ubiquitin-protein ligases or E3 enzymes

An E3 enzyme is defined as a protein that binds the target substrate, either directly or indirectly, via ancillary proteins, and catalyses transfer of ubiquitin from a thiolester intermediate on E2 or E3 to an amide linkage with the substrate or with a poly-ubiquitin chain already anchored to it. Since the target proteins bind to the ligases prior to conjugation, E3s are key players in determining the high specificity of the system. Despite their importance, the number of known E3s is few and the information concerning their mode of action is scant. Lack of sequence homology among different E3s and the frequent association of these enzymes with multisubunit complexes in which the identity of the ligase subunit is not known, make their study rather difficult (Ciechanover, 1998).

Different ligases recognise the numerous substrates of the pathway via specific motifs. These can be either primary, or secondary, post-translational modifications. Primary motifs do not necessarily lead to constitutive degradation of the proteins that contain them. They can be hidden and exposed only following misfolding or dissociation of subunits. Certain substrates will not be recognised by their ligases unless they associate with an ancillary protein or molecular chaperone that act as *trans* recognition elements. The 26S proteasome is composed of the 20S core catalytic complex, which is flanked on both sides by the 19S regulatory complex. It recognises specifically ubiquitin-tagged proteins and degrades them to small peptides. An important step in the ubiquitin pathway involves the release of ubiquitin from its various adducts. Release of ubiquitin plays an essential role in two processes, the first of which is protein degradation and the second is ubiquitin biosynthesis (Ciechanover, 1998). The ubiquitin-proteasome pathway is illustrated in figure .4

Fig. 1.4 The ubiquitin-proteasome pathway
modified from A. Ciechanover (1998)



(A) Conjugation of ubiquitin to the target molecule:

1 activation of ubiquitin by E1.
2 transfer of activated ubiquitin from E1 to a member of the E2 family.
3 transfer of activated ubiquitin from E2 to a substrate specific E3.
4 formation of a substrate-E3 complex and biosynthesis of a substrate-anchored poly-ubiquitin chain.

(B) Degradation of the tagged substrate by the 26S proteasome:

5 binding of the poly-ubiquitinated substrate to the ubiquitin receptor subunit in the 19S complex of the 26S proteasome and degradation of the substrate to short peptides by the 20S complex.
6 recycling of the ubiquitin via the action of isopeptidases.

The high evolutionary conservation of ubiquitin (only three amino acid differences between the human and yeast homologues; Haas and Siepmann, 1997) enabled the discovery of many ubiquitin-related- or ubiquitin-like proteins, which all resemble ubiquitin in their amino acid sequences. Ubiquitin-related proteins often have similar means of attachment to their target proteins, but, on the other hand, also posses unique functions. Some are involved in post-translational, single or multiple modification of target proteins that serves non-proteolytic purposes (Hochstrasser, 2000). Some of these proteins are listed below.

1.4.3 SUMO (small ubiquitin-related modifier)/ sentrin

In at least some instances the addition of SUMO to a target protein (sumoylation) changes the localisation of this protein or its interactions with other proteins. The first identified target of SUMO is the trafficking protein RanGAP1, which helps to shuttle proteins from the cytoplasm to the nucleus and can only localise to the nuclear pores in the presence of SUMO (Kretz-Remy and Tanguay, 1999). A diverse array of targets for SUMO are known to exist.

1.4.4 RUB1 (yeast ubiquitin like protein, related to ubiquitin 1)

Only one type of RUB1-modified protein is known to exist: the cullins, which are common subunits of a large group of multisubunit E3 ligases that add ubiquitin to its target proteins (Read et al., 2000). One of these E3 ligases directs the ubiquitination of I-KB (a transcription factor): the cullin subunit (Cull-1) of this complex must be conjugated to RUB1 to ensure the ubiquitination and degradation of I-KB. A curious feature of this system is that addition of RUB1 to the cullin subunit appears to depend on RUB1 ligase activity within the ubiquitin ligase itself (Kamura et al., 1998). In mammalian cells RUB1 is called NEDD8 and is a 81-amino acid polypeptide that is 60% identical and 80% homologous to ubiquitin (Kumar et al., 1993). This protein will be discussed in some greater detail in Chapter 4.

The current understanding of SUMO and NEDD8 are reviewed by Yeh et al. (2000).

1.4.5 Apg12 (yeast ubiquitin like protein) and Urm1 (ubiquitin-related modifier)

Apg12 and Urm1 are two polypeptides that have only weak sequence similarity to ubiquitin (Mizushima et al., 1998; Furukawa et al., 2000). Although these ubiquitin like proteins are not closely related in sequence to ubiquitin activation depends on E1-like activation enzymes All of these enzymes share a similar nucleotide-binding motif and other sequence similarities, including a putative cysteine active site downstream of the nucleotide-binding element.

The broad range of substrates and processes in which the ubiquitin pathway is involved means that there are aberrations in the system that have been implicated in the pathogenesis of several diseases, both inherited and acquired Ciechanover (1998) divided the pathological states into two groups: 1. those that result from loss of function, a mutation in an enzyme or substrate that leads to stabilisation of certain proteins; and 2. those that result from a gain of function resulting in accelerated degradation. Diseases include Huntington disease (Davies et al. 1997), spinocerebellar ataxias (Cummings et al., 1998), Liddle syndrome (Staub et al., 1997), cystic fibrosis (Ward et al., 1995) and syndrome (Kishino et al. 1997).

1.5 Mapping techniques

1.5.1 Genetical mapping

Genetic maps represent genetic distances separating gene loci in a linkage structure, i.e. the genes of one linkage group or chromosome. The genetic distances in turn reflect the probability of crossing over to occur in the region being mapped. The total genetic map distance of the human genome is about 3000cM and there are about one million base pairs per cM. It should be noted that there are differences between males and females and that crossing over is not quite random e.g. near the centromere it is infrequent. This means that genes at the centromere would appear to be closely linked genetically, but could be physically far apart, i.e. genetic distance is not an accurate reflection of the physical distance.

Genetical mapping makes use of polymorphisms/ polymorphic markers. A polymorphism is the presence of two or more allelic variants in a population at appreciable frequency. A polymorphic marker is a gene or DNA segment whose location in the genome is known and ideally must be highly polymorphic (i.e. a locus for which multiple alleles exist). This often makes it possible to distinguish parental alleles and therefore the identification of a marker that is co-inherited with a disease gene, thereby establishing a physical link between the two. Various types of markers are available. These include, in addition to the traditional biochemical variants and blood groups, restriction fragment length polymorphisms (RFLPs) (Botstein et al., 1980), variable number tandem repeats (VNTRs) (Nakamura et al., 1987) and microsatellite polymorphisms (e.g. (AC)n repeats).

RFLPs are inherited differences in sites for restriction enzymes that result in differences in lengths of fragments produced by cleavage with the relevant restriction enzyme and occur frequently enough in the human genome to be useful as markers. If allelic sequences are compared between any two individual chromosomes, differences in individual base pairs occur at a frequency of > 1 per 1000bp (Lewin, 2000). As well as base changes, deletions and insertions of a sequence, can be detected as RFLPs. RFLP maps of the human genome have been reported by Keller et al. (1987) and Dib et al. (1996)- the latter based on over 5000 microsatellites.

VNTRs were first detected by Jeffreys et al (1985). Also known as hypervariable minisatellites, they are of varying sequence and length and are relatively frequent within the genome, although concentrated towards the ends of the chromosomes.

Microsatellite markers are the most abundant source of highly informative polymorphisms and consist of di-, tri- or tetra-nucleotide repeats. The first and most widely used ones to be involved in genotyping are the simple (CA)n-(GT)n repeats. These markers are distributed ubiquitously throughout the genome, are variable in length and can be highly polymorphic. On average, a variable (AC)n repeat occurs approximately every 40 to 50 kb (Weber and May, 1989) with highly variable repeats occurring every 300 to 500 kb. They may occur adjacent to coding regions of genes, in introns within genes or within untranslated regions

(Hamada et al. 1984; Weber and May 1989) and hence act as useful markers for gene loci. Examination of over 100 human (AC)n repeat sequences revealed that the sequences differed from each other both in numbers of repeats and in repeat sequence type. Using a set of precise classification rules, Weber (1990) divided the sequences into three categories: perfect repeat sequences without interruptions in the runs of CA or GT dinucleotides (64% of total), imperfect repeat sequences with one or more interruptions in the run of repeats (25%) and compound repeat sequences with adjacent tandem simple repeats of a different sequence (11%). It was found that the marker informativeness increases as the average number of repeats increases.

It should be noted that the distinction between microsatellites and VNTRs is essentially an arbitrary one, with microsatellites possessing repeat units varying from 1 to 5 nucleotides and minisatellites possessing repeat units varying roughly from 9 to 64 nucleotides. The success of genetic mapping depends on the informativeness of the markers and for some diseases not many individuals may be available for analysis and therefore, even with highly informative markers, the resolution of genetic mapping may be limited. In these cases, further characterisation is carried out with physical mapping techniques.

More recently single nucleotide polymorphisms (SNPs) have been the focus in human genetics, although they are generally less informative than other types of genetic markers, such as microsatellites. However, SNPs are essentially distributed randomly throughout the genome, although less frequent in coding regions (Venter et al., 2001). They are also extremely abundant and well-suited for automated large-scale genotyping (Lindblad-Tho et al., 2000) More than 1 million SNPs have been identified in the human genome Lander e al 2001 They can be assayed via chip-based microarrays and one of their advantage their ability to locate loci that may be responsible for complex traits via linkage/ linkage-disequilibrium analysis (Zhao-Lue-Ping et al., 1998). The National Cancer for Biotechnology Information has established the dbSNP database and the contents is available to the public at website http://www.ncbi.nlm.nih.gov/SNP. Submitted SNPs can also be dow loaded via anonymous FTP at ftp://ncbi.nlm.nih.gov/snp/ (Smigielski e 200(

1.5.2 Physical mapping

The assignment of genes to chromosomes, and further, to specific regions on a chromosome, using physical mapping techniques, progressed rapidly mainly due to the development and nature of early somatic cell hybrid genetics. Somatic cell hybrids are typically produced by fusing rodent and human cells (Harris and Watkins, 1965) which preferentially maintain the rodent chromosomes and lose their human counterparts if the rodent cell line is well established for propagation in culture (Weiss and Green, 1967). Once a stable pattern of chromosome retention has been established these fused cells provide a very valuable resource for physical mapping as they contain a subset of the human genome. The human component in the cells can be determined by cytogenetic means and/or by the retention pattern of established markers. Genes and anonymous DNA sequences can then be assigned to various chromosomes on the basis of their absence or presence in such a panel of somatic cell hybrids (Creagan and Ruddle, 1975). Selection of the human material can be achieved by arranging for a gene on a human chromosome to be the sole provider of a cellular growth requirement. The classic selection method is the growth of cells in HAT (hypoxanthine-aminopterin-thymidine) medium (Littlefield, 1964). Aminopterin blocks *de novo* synthesis of purines and pyrimides and therefore the major DNA synthesis pathway. The addition of the substrates hypoxanthine and thymidine allows the cells to use an alternative pathway in the presence of the hypoxanthine phosphoribosyl transferase (HPRT) and thymidine kinase (TK) genes respectively. This enables the selection of these two genes and therefore the chromosomal segments on which they are localised, namely Xq for the HPRT gene (Riciutti and Ruddle, 1973) and 17q for the TK gene (Miller et al., 1971). Hence, the two human chromosomes on which these genes are localised were the focus for early somatic cell hybrid mapping studies. Therefore, along with the unique inheritance pattern observed for X-linked disorders, the resources available for physical mapping also favoured preferential characterisation of the X chromosome.

Developments of somatic cell hybrids brought about a technique that combines physical and genetical mapping, called radiation hybrid mapping. It is based on a strategy first devised by Goss and Harris (1975) where chromosomes are broken

by X-irradiation of human cells. The non-viable irradiated cells are fused with untreated rodent cells which are HPRT and hybrids selected in HAT medium as describe above. The distance between DNA markers is determined by the frequency with which pairs of markers remain on the same fragment; generall the further apart two markers are on a chromosome, the more likely they are tc be separated by X-irradiation. And by analysing a series of markers known tc on a single chromosome the order of the markers can be established However, this method was limited by the lack of suitable selection methods for different chromosomes until 1990, when Cox et al. assayed about 100 radiation hybrid clones for 14 markers known to occur within the proximal 20Mb on he long arm of the human chromosome 21 Jsing pulsed field gel electrop or the order of all 14 markers was determined with an average resolution of 500kb, which is better than that of *in situ* hybridisation on metaphase chromosomes (1-2Mb), but worse than that of fluorescent *in situ* hybridisation on interphase cells, which can determine the order of markers separated by no more than 50 -100kb (Trask et al., 1991). A review of the whole field has been provided by Abbott and Povey (1995)

The goal of gene mapping and the Human Genome Project is to produce a sequence of DNA representing the functional blueprint and evolutionary history of the human species. However, at present, there is no way to sequence DNA in an intact chromosome, the average size being 120 Mb. Physical mapping includes the analysis of DNA using naturally occurring or artificially induced breakpoints or deletions in the chromosome. It makes use of the cutting of DNA into smaller fragments using restriction enzymes, the *in vitro* joining of the DNA to be cloned (or insert) to a vector, a small extrachromosomal genome (like those of plasmids, phage and viruses) that can replicate in an appropriate host cell. The resulting recombinant molecule can then be transformed into host cells and clonally propagated Different kinds of vectors are available, varying in origin and the amount of endogenous DNA they can hold (see table 1.1). Two of these, namely YACs and PACs, will be discussed in further detail in Chapter 3 Physical mapping also includes the separation of DNA fragments using gel electrophoresis, the transfer of DNA onto a solid surface, usually a nitro-cellulose membrane, hybridisation techniques and the production of overlapping genomic clones, their analysis using rare cutting restriction enzymes

Table 1.1: Summary of Cloning Vectors available

vector	insert size	features
HACs	difficult to determine	- constructed from a YAC format (see below) using alpha satellite DNA - contains a selectable marker, human telomers and a putative origin of replication - stable structure of insert (even after 100 generations) - size range from 3.5-12.9 Mb
YACs	250kb - 1Mb	- propagated in *S.cerevisiae* - artificially constructed - contains sequences necessary for propagation and selection in *E.coli* - functions as a yeast chromosome 10^3 clones / μg vector
BACs	300kb	- based on the F plasmid in *E.coli* - maintained at a low copy number (1 or 2 per cell) - stable structure of insert (even after 100 generations) - transformation in bacteria is more efficient than spheroplast transformation - ease of isolation and manipulation of insert DNA - no positive selection for recombinants
PACs	100 - 300kb	combination of F-plasmid from BACs and P1 clones transformation in bacteria possible, but also higher copy number/ DNA yields and positive selection
P1s	80 - 100kb	- insert and vector DNA are ligated and packaged into P1 phage heads - virions inject DNA into bacteria where it replicates as a plasmid - single copy - copy number can be increased to 20 per cell by additon of IPTG which relieves repression leading to a good yield - cloning efficiency is that between a YAC and cosmid 50,000-member human DNA library exists
Cosmids	40 - 50kb and smaller	- replicate like a plasmid, but can be packaged in vitro into λ phage heads - high efficiency if in vitro packaging 10^6-10^7 / μg vector - contain the cohesive ends (cos sites) from lambda DNA - problem associated with instability of clones
Fosmids	35 – 45kb	- contain cohesive end sites (*cos*) for bacteriophage l packaging - derived from genetic elements of the single copy fertility (F) factor of *E. coli*, which enables fosmids to maintain genomic DNA inserts with much higher stability than cosmids
phage	20 - 30kb	
plasmids	20kb	

References: Henning et al, 1999, *Proc. Natl. Acad. Sci.* **96**: 592-597. Burke et al., 1987. Shizuya et al., 1992. Sternberg et al., 1990. See references.

and the generation of sequence data.

1.5.3 Computer analysis

This has become a very important aspect of mapping. Sequencing centres such as the Sanger Centre (Hinxton, Cambridge), the St Louis Genomic Sequencing Center (Washington University), the Whitehead Institute or the Human Genome Mapping Research Centre (funded by the Medical Research Council, UK) are producing an ever increasing amount of genomic data and sequences are available from their web sites (see table 1.2). In order to obtain an insight into the biological significance and possible function, or to do statistical analyses of newly obtained sequences, computer programs are being developed. These include programs that find splice sites (Staden, 1984; Nakata et al., 1985; Kudo et al., 1987), coding regions (Fickett, 1982; Fichant and Gautier, 1987; Konopka and Smythers, 1987), repeats and identify gene signals such as promoters (Bucher, 1990; Prestridge, 1995), predict possible exons (Lopez et al., 1994; Solovyev et al., 1994) and look for homologies to sequences in existing databases (Altschul and Lipman, 1990; Zhang and Madden, 1997).

A large number of databases are available. Different kind of sequence databases exist: nucleic acid and peptide sequence databases (for example EMBL and SwissProt, respectively) and restriction enzyme, protein motif, transcription factor databases. Each sequence has a unique accession number permanently associated with it; although it might differ if the sequence is merged with another. Each sequence also has an ID name, but this is not guaranteed to be unique between databases. Genomic databases include the GDB (Genomic Data Base), which was devised as the ultimate repository of human mapping and genomic data. Their web site address is: http://www.hgmp.mrc.ac.uk/gdb/. Although it is very easy to search for information on genes, clones etc., it does not hold any information on gene products. Clinical and Mutation databases include OMIM, which is a database of phenotypes of human diseases having a substantial genetic component and HGMD, a database of sequences and phenotypes of human disease-causing mutations. Integrated databases, such as Entrez, are a set of tightly linked databases for example including nucleic acid and protein sequences as well as MEDLINE, a bibliographic database.

Table 1.2: Sources of publicly available sequence data and other genomic information

(modified from Lander et al., 2001)

web site address	Institute	service provided
http://genome.ucsc.edu/	University of California at Santa Cruz	contains assembly of the draft genome sequence and updates
http://genome.wustl.edu/gsc/human/Mapping/	Washington University	contains links to cloneand accession maps of the human genome
http://www.ensembl.org	EBI/ Sanger Center	allows access to DNA and protein sequences with automatic baseline annotation
http://www.ncbi.nlm.nih.gov/genome/guide/	NCBI	views of chromosome and maps and loci with links to other NCBI resources
http://www.ncbi.nlm.nih.gov/genemap99/	NCBI	contains data and viewers for radiation hubrid maps of EST-based STSs
http://compbio.ornl.go/channel/index.html	Oak Ridge National Laboratory	Java viewers fro human genome data
http://hgrep.ims.u-tokyo.ac.jp/	RIKEN and University of Tokyo	gives an overview of the entire human genome structure
http://snp.cshl.org	The SNP Consortium	includes a variety of ways to query for SNPs in the human genome
http://www.ncbi.nlm.nih.gov/Omim/	Online *Mendelian Inheritance in Man*	contains information about human genes and disease
http://www.nhgri.nih.gov/ELSI/ http://www.ornl.gov/hgmis/elsi/elsi/html	NHGRI and DOE (respectively)	contain information, links and articles on a wide range of social, ethical and legal issues

Two major sequence analysis packages exist: EMBOSS (European Molecular Biology Open Software Suit), which is a collaboration of European biological software developers, who aim to develop and integrate a range of currently available packages and tools for sequence analysis into a general, publicly available, suit of programs and libraries. Their web site address is: http://www.sanger.ac.uk/Software/ EMBOSS/. The second one being GCG, an extensive package of applications for molecular biologists, containing programs for sequence editing, fragment assembly, sequence analysis, multiple sequence alignments, database searching etc. Their web site address is: http://www.hgmp.mrc.ac.uk/Registered/Option/gcg.html. These two packages enable the user to do sequence database searches, gene identifications, retrieve sequences, do graphical sequence comparisons, sequence alignments (pairwise and multiple), produce restriction maps, translate nucleotide sequences and design primers. Other programs are designed to deal with various other types of protein analysis including finding motifs and domains in protein sequences and prediction of protein secondary and tertiary structures, phylogeny linkage analysis, radiation hybrid mapping and sequence assembly. The availability of those bioinformatics services have made a crucial difference to the analysis of information obtained by scientists working in a variety of disciplines

1.6. Outline of this Study

The primary aims of the work described in this thesis include an analysis of the region around Xp11.3 with particular emphasis on an investigation of X chromosome inactivation. The study is divided into three main sections:

As a basis for the analysis, it was deemed to be important to produce a physical map and detailed characterisation of a YAC/ PAC contig around the DXS8237E UBE1-PCTK1 region.

Chapter 2 describes the materials and the techniques used for the physical and genetic mapping undertaken during the course of this project. Chapter 3 describes the production of clones from the region around Xp11.3 and a physical

achieved with rare cutting restriction enzymes and pulsed field gel electrophoresis. The clones have been converted to sequence-tagged sites (STSs),

detected by primer pairs yielding unique PCR products and a map has been constructed using these markers. Probes from genes mapping to the region have been used to define overlaps between the YACs and PACs spanning the region.

2. Examination of a unique polymorphism in the UBE1 region that may be associated with cognitive ability.

Chapter 4 describes the analysis of a (AC)n repeat, obtained from cosmid turn was subcloned from a YAC. It was selected on the basis of hybridisation to a clone near the *UBE1* gene and its unique size. The isolation, mapping and sequencing was followed up by primer design and tested by individua genotyping, as well as genome scanning for quantitative trait loci that may be important in cognition (g)- given the significance of protein sequestration/ turnover in the brain and neurodegeneration.

3. X chromosome inactivation studies of the DXS8237E-UBE1-PCTK region

The pattern of inactivation in this region is of particular interest, since the mechanisms of X chromosome inactivation and the escape from inactivation are, as yet, not fully understood. The inactivation status of the *DXS8237E* and *PCTK1* gene differ: the first undergoes normal inactivation and the second escapes this process. The inactivation status of the *UBE1* gene has been controversial, although it is widely excepted that it does escape X chromosome inactivation.

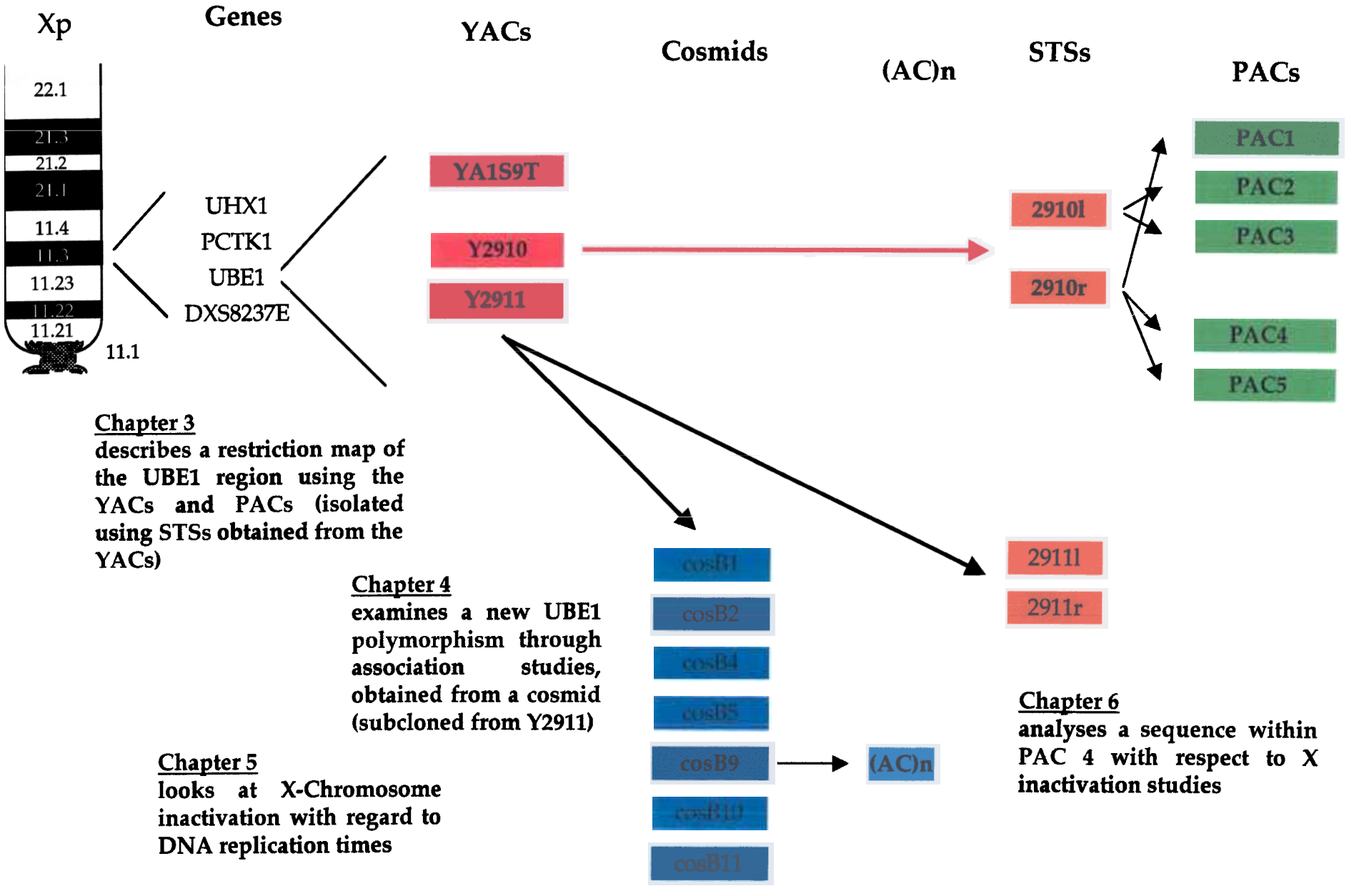
Chapter 5 describes the determination of the inactivation status of all three genes by the means of replication time studies. Late and early replicating DNA isolated from HeLa cells and separated by caesium chloride centrifugation. By means of restriction enzyme digests, hybridisation and comparison hybridisation patterns inactivation status were determined with references to standard hybridisation patterns in which probes generated from genes of known inactivation status were used.

Chapter 6 describes the investigation of the DNA sequence between *DXS8237E*, *UBE1*, *PCTK1* and the distal gene, *UHX1*, especially with regard to LINE1 elements. This may help to obtain a better understanding of the mechanism of escaping X chromosome inactivation. Further, the exon-intron boundaries of

the *UBE1* gene were established. Computer analysis was used for these studies.

A summary of the outline of study is illustrated in figure 5

Fig. 1.5 Outline of this study



Chapter 3
describes a restriction map of the UBE1 region using the YACs and PACs (isolated using STSs obtained from the YACs)

Chapter 4
examines a new UBE1 polymorphism through association studies, obtained from a cosmid (subcloned from Y2911)

Chapter 5
looks at X-Chromosome inactivation with regard to DNA replication times

Chapter 6
analyses a sequence within PAC 4 with respect to X inactivation studies

CHAPTER 2 - Materials and methods

2.1 Buffers

Church hybridisation buffer: 7%(w/v) SDS (Sigma #L4509, 99% pure), 0.5M di-sodium hydrogen orthophosphate (pH 7.2), 1% (w/v) BSA (Sigma#A4503; made as a 10% solution, prefiltered through a Sartorius 0.45µm minifilter), 1mM Na_2EDTA (pH 8.0)

Denaturing solution 1.5M NaCl, 0.5M NaOH

Enzyme dilution buffer: 50mM KCl, 10mM Tris-Cl (pH 7.4), 0.1mM Na_2EDTA, 1mM DTT, 0.2% (w/v) BSA, 50% (v/v) glycerol

Frozen storage buffer (FSB): 100mM KCl, 45mM $MnCl_2$, 10mM $CaCl_2$, hexamine (III) cobalt trichloride, 10mM potassium acetate, 10% (v/v glycerol

GTE: 50mM glucose, 25mM Tris-Cl (pH 8.0), 10mM Na_2EDTA

LiDS: 100mM Na_2EDTA, 10mM Tris-Cl (pH 8.0), 1% (w/v) lithium dodecyl sulphate (Sigma)

NDS: 0.5M Na_2EDTA, 10mM Tris-base, 1% (w/v) sodium lauryl sarcosine (Sigma) (pH 9.5)

Neutralising solution 1.5M NaCl, 1M Tris-Cl (pH 8.0)

PBS: 2.6mM KH_2PO_4, 26mM Na_2PO_4, 145mM NaCl (pH 7.4)

SE: 75mM NaCl, 25mM Na_2EDTA (pH 8.0)

SEB: 1M sorbitol, 20mM Na_2EDTA, 4.4mM ß-mercaptoethanol

SCE: 1M D-sorbitol, 100mM sodium citrate, 50 mM EDTA (pH 8.0)

20 x SET: 3M NaCl, 0.4M Tris-Cl, 20mM Na_2EDTA (pH 7.2)

SM 0.1M NaCl, 50mM Tris-Cl (pH 7.5), 0.2% (w/v) $MgSO_4.7H_2O$, 0 (w v gelatin (pH 7.5)

20 x SSC: 3M NaCl, 0.3M Na_3 citrate (pH 7.0)

STEB 1M sorbitol, 10mM Tris-Cl (pH 7.5), 20mM Na_2EDTA, 14mM ß-mercaptoethanol

5% (w/v) sucrose, 50mM Na_2EDTA, 5% (v/v) Triton x-100 (Sigma), 50mM Tris-Cl (pH 8.0)

STOP: 80% (v/v) formamide, 10mM EDTA (pH 8.0), 0.01% (w/v bromophenol blue, 0.01% (w/v) xylene cynol FF

50 x TAE: 2M Tris-acetate, 50mM Na_2EDTA

10 x TBE: 0.9M Tris-borate, 20mM Na_2EDTA (pH 8.0)

TE 10mM Tris-Cl (pH 7.4), 1mM Na_2EDTA (pH 8.0)

150mM NaCl, 10mM Tris-Cl, 10mM Na_2EDTA (pH 8.0)

2.2 Media and antibiotics

CY 0.2% (w/v) KCl, 0.2% (w/v) $MgSO_4.7H_2O$, 0.3% (w/v) NaC 0.5% (w v) Bacto yeast extract (Difco), 1% (w/v) casamino acids (pH 7.5)

LB broth: 1% (w/v) Bactotryptone (Difco), 0.5% (w/v) Bacto yeast extrac (Difco) 1% (w/v) NaCl (pH 7.5)

SD + CAT: 0.67% (w/v) Bacto yeast nitrogen base without amino acids (Difco) 2% (w/v) glucose, 1.4% (w/v) Bacto casamino acids (Difco), 55mg/ adenine hemisulphate, 55mg/l tyrosine

SOB: 2% (w/v) Bactotryptone (Difco), 0.5% (w/v) Bacto yeast extract (Difco), 10mM NaCl, 2.5mM KCl, 10mM $MgCl_2$, 10mM $MgSO_4.7H_2O$

SOC: SOB + 20mM glucose

2 x TY 1.6% (w/v) Bactotryptone (Difco), 1% (w/v) Bacto yeast extract (Difco), 0.5% NaCl (pH 7.0)

Ampicillin (Sigma) Stock solution 25mg/ml, used in media at 50µg/ml

Kanamycin (Sigma)	Stock solution 10mg/ml, used in media at 10µg/ml
IPTG (Sigma)	Stock solution 0.5M, used in media at 0.125M
X-gal (Gibco-BRL)	Stock solution 2%(w/v) in dimethylformamide, used in media at 0.004%

2.3 Bacterial strains

JM83: F- *ara* Δ(*lac* -*pro*AB) *rps*L (Str^r) (ø80 dΔ (*lac* Z) M15) Yanisch-Perron et al 1985)

JM109: *end*A1, *rec*A1, *gyr*A96, *thi*, *hsd*R17 (r_K^-,m_K^+), *rel*A1, *sup*E44, Δ(*lac*-*pro*AB), [F', *tra*D36, *pro*AB, *laq*^{lq}ZΔM15] (Yanisch-Perron et al., 1985)

Top Ten (Invitrogen): mcrA, Δ(*mrr-hsd*RMS-*mcr*BC), ø80 Δ(lacΔ) M15, Δ*lac*X74 *deo*R, *rec*A1, *ara* D139, Δ(ara,leu) 7697, *gal*U, *gal*K, l-, *rps*L, *end*A1, *nup*G

OneShot (Invitrogen): F', *end* A1, *rec*A1, *hsd*R17(r_K^-,m_K^+), l-, *sup*E44, *thi*-1, *gyr*A96, *rel*A1, ø80D*lac*DM15, D(lacZYA-argF)U169, *deo*R+

2.4 Somatic cell hybrids

Somatic cell hybrids, produced by fusing human and rodent cell lines, are a valuable resource for physical mapping of clones and anonymous sequences. Those used in this study were selected for and grown in HAT medium by A Chand, E. Hatchwell and G. Black.

The fibroblast and hybrid DNA samples used for the localisation of clones from the short arm of the X chromosome and contained in mapping panels used for this investigation include the following:

i. THYB-X: A hybrid containing an intact X chromosome as the only human material on a mouse (RAG) background (Lund et al., 1983).

ii. MOCH Hybrids: A translocation between chromosome X and 9, at Xp11.21,

associated with incontinentia pigmenti was used to make these hybrids (Gorski et al., 1992). The derived X-chromosome is termed MOCH1A and that with the reciprocal derived chromosome 9 is called MOCH1B. The derived X chromosome, in MOCH1A, is selected for by culturing the cells in the presence of HAT. The human Xp material in hybrid Moch1B cannot be selected for and is present at a low level.

iii. X;17: A translocation hybrid t(X;17)(p11;p11) containing the derived X chromosome on a mouse (RAG) background (personal communication Hatchwell).

iv. Hamster A23: A CHO, TK-, control cell line for the hybrids with a hamster background listed above (Benham et al., 1989).

v. RAG: A mouse control cell line for MCP6, MOG and ThyBX (Cockburn et al., 1992).

2.5 Frozen storage of cells

Bacterial Culture: Add 150µl of sterile glycerol to 850µl of cells in culture. Shake well to ensure complete and even dispersal of glycerol. Freeze the culture in an ethanol dry ice bath or in liquid nitrogen. Store at -70°C.

Yeast Culture: Follow the protocol outlined for bacterial cells, but increase the glycerol concentration to 50%.

Mammalian Culture: Freeze 3-5 x 10^6 cells per ml. Resuspend pellet in freeze medium (60% medium, 30% FCS, 10% DMSO). Immediately wrap in bubble wrap and leave at -70°C over night. Transfer to liquid nitrogen to store.

2.6 Preparation of plasmid DNA

2.6.1 Boiling miniprep

This was the method of choice used to analyse large numbers of clones after cloning procedures.

1 Resuspend $^1/_2$ streak of cells in 105µl STET.

2. Add 7.5µl lysozyme solution (10mg/ml; Sigma) in TEN. Leave at room temperature for 5 minutes.

3. Add 1µl diethylpyrocarbonate (DEPC, Sigma). Place in a boiling water bath for minute.

4. Spin cells in an Eppendorf centrifuge at room temperature for 15 minutes.

5. Remove pellet with toothpick

6. Add 11µl 3M sodium acetate (pH 5.2) and 115µl ice-cold isopropanol and place at -20°C for 30 minutes.

7. Spin in Eppendorf centrifuge at 4°C for 15 minutes.

8. Wash pellet in 70% ethanol, dry briefly and dissolve in TE containing 2.5µg/ml of DNAse-free RNase (Boehringer Mannheim). Incubate at 65°C for 10 minutes.

9. Store at -20°C

2.6.2 Small scale alkaline lysis

1. Inoculate 10ml of 2 x TY broth, containing the appropriate antibiotic, with a clonal bacterial colony (or 100µl of frozen stock) and incubate, shaking, at 37°C overnight

2. Centrifuge in a bench centrifuge at top speed for 10 minutes.

3. Remove supernatant and resuspend in 300µl of lysozyme (5mg/ml) in GTE. Leave at room temperature for 5 minutes.

4. Add 600µl of fresh, ice-cold 0.2M NaOH/1%SDS, mix gently by inversion, and stand on ice for 5 minutes

5 Add 500µl of 5M KAc (3ml 5M KAc, 575µl glacial acetic acid 425µl water Vortex and place on ice for 20 minutes

6. Pellet the cell debris and bacterial DNA in an Eppendorf centrifuge at 4°C for 15 minutes

7. Transfer the supernatant to a fresh tube. Add $^1/_2$ volume phenol equilibrated with TE, vortex, then add $^1/_2$ volume of chloroform (24:1 chlorofor alcohol) and vortex again Spin in Eppendor centrifuge for minute separate phases

8. Transfer the supernatant to a fresh tube and add 2 volumes of cold (4°C) absolute ethanol or 1 volume of cold (4°C isopropanol Stand at room temperature for 5 minutes

9. Spin in Eppendorf centrifuge for 5 minutes.

10. Wash DNA with 70% ethanol, spin for 2 minutes, and remove supernatant

11 Dry the DNA in vacuum desiccator, and dissolve in TE containing 2.5µg/ml DNAse-free RNase (Boehringer Mannheim). Incubate at 37°C for 30 minutes.

12. Store at -20°C

2.6.3 Promega plasmid minipreps and QIAGEN minipreps

These are commercially available kits, which are faster procedures than alkaline lysis and give good yields of DNA (10µg per 3ml cell culture). No further purification of the DNA is necessary before use. See manufacturer's manuals for protocols.

2.7 Restriction enzyme digestion of DNA in solution

Restriction digests were performed in buffers supplied by the manufac rer: except for certain double digests. For genomic and YAC plug digests sper was used at the following concentrations

[Salt]	[Spermidine]
>50mM	5mM
30mm	3mM
20mm	2mM
0mM	0mM

Spermidine was not used, in general, for other digests. Precipitation of digests using sodium acetate was avoided with digests containing spermidine as this forms a precipitate that is difficult to dissolve. Genomic digests were carried out overnight, using a three-fold excess of enzyme; other digests were performed for 1-16 hours using between 2 and 10-fold excess of enzyme.

Genomic digests were precipitated before electrophoresis in order to remove salt from the solution and thus eliminate discrepancies in the speed of electrophoresis between samples. The following protocol was used:

Add NaC to a final concentration of 200mM, followed by 2 volumes of ice-old ethanol or 1 volume of ice-cold isopropanol. Place at -70°C for 15 minutes.

2. Spin in an Eppendorf centrifuge at 4°C for 15 minutes.

3. Remove supernatant, add 0.5-1.0ml of 70% ethanol and centrifuge.

4 Dry the DNA pellet and redissolve in appropriate amount of TE

Partial digests were achieved by keeping the reaction time consta and aryin the enzyme concentration. 15-20 units of enzyme were used for complete digests and 0.015 to 5 units for partials depending upon the efficiency of the enzyme.

2.8 Conventional agarose gel electrophoresis

The Pharmacia GNA100 and GNA200 along with the Scieplas electrophoresis tanks were used for the separation of DNA fragments. The routine analysis of sizing fragments, estimating DNA concentrations and visualising PCR products were run on 0.5-2% (w/v) Type I (Sigma, low EEO) or Type II-a (Sigma, medium EEO) gels in a 1 x TBE buffer. For separation and visualisation of fragments under 400bp - 1% (w/v) Type I (Sigma, low EEO) + 2% (w/v Nuseive GTG (FMC bioproducts) 1 x TBE gels were used. Genomic digests were usually run on a 0.8% (w/v) Type I (Sigma, low EEO) gel in 1 x TBE buffer overnight at a voltage gradient of approximately 1.5V/cm. DNA fragments to be eluted from a gel matrix were electrophoresed on a 0.7-1.5% (w/v) Ultrapure (Gibco BRL) or Seaplaque (FMC Bioproducts) agarose gel in a 1 x TAE buffer.

Loading buffers contained a 5 or 10 times concentration of the appropriate running buffer, with bromophenol blue and cyanol and either glycerol (5% (v v) final concentration) or Ficoll-400 (3% (w/v) final concentration, (Pharmacia) Ficoll is superior to glycerol, in particular for genomic digests, as it gives a less marked meniscus and sharper bands on hybridisation after Southern blotting.

Gels for Southern blotting were stained in water or BE buffer containi ethidium bromide (EthBr, Sigma) at 0.5mg/ml for 5-30 minutes, wh gels contained EthBr at approximately 0.5mg/ml All were viewed on a V transilluminator (Ultra Violet Products Inc.) and then photographed with a Polaroid Land camera using Polaroid 665 film or an image taken using a IS-1000 Digital Imaging System (Alpha Innotech Corporaton).

Size markers that were used for agarose gels were lambc ph predigested with *Hind* III, 1kb ladder and 100bp ladder (Gibco BRL The zes of the markers are as follows:

λ/ Hind III: 0.56kb, 2.03kb, 2.32kb, 4.36kb, 6.56kb, 9.42kb and 23.13kb

1kb ladder: 75bp, 142bp, 154bp, 200bp, 220bp, 298bp, 344bp, 394bp, 506bp, 516bp, 1018bp, 1635bp, 2036bp, 3054bp continuing with increases of approximately 1kb.

100bp ladder: 100bp, 200bp, 300bp, continuing with increases of 100bp to 500 and a last band at 2072bp.

2.9 Transfer of DNA to a solid support

2.9.1 Southern blotting of agarose gels

Transfer of DNA from agarose gels to membranes was carried out using the basic technique described by Southern (Southern, 1975). DNA from the gels were transferred onto Hybond N+ membranes (Amersham) using alkaline blotting. After photography and 4 minute exposure to UV light, gels were placed on a blotting platform containing 0.4M NaOH transfer solution, using 3MM Whatman as a wick. The filter was marked with a laundry marker (WH Smith Ltd) and placed dry on the gels, followed by two pieces of Whatman 3MM paper, soaked in 0.4M NaOH. Paper towels were added and transfer allowed to proceed for 14-24 hours. Filters were then neutralised in 0.2M Tris-Cl (pH7.5), 2 x SSC and stored damp at -20°C.

2.9.2 Bacterial colony transfer

1. Grow bacterial colonies in duplicate on plates with 2 x TY agarose and the appropriate antibiotic in a 37°C incubator for 24 hours.

2. Place a piece of Hybond N+ on the surface of one of the plates for 30 seconds and then lay DNA side up on dry 3MM Whatman paper.

3. Submerge filter in denaturing solution for 4 minutes at room temperature.

4. Place filter in an 85°C steam bath for 4 minutes.

5. Submerge filter in neutralising solution at room temperature for 4 minutes.

6. Submerge filter in 50mM Tris-Cl (pH 8.5), 50mM Na_2EDTA, 100mM NaCl, 1%(w/v) Na-lauryl sarcosine containing 250µg/ml pronase (Boehringer Mannheim) at 37°C for 20 minutes.

7. Transfer filter to 3 layers of 3MM Whatman paper until dry.

8. UV cross link the filter for 4 minutes.

2.10 Radio labelling of DNA

2.10.1 Multiprime labelling

DNA was labelled following the protocol described by Feinberg & Vogelstein (Feinberg and Vogelstein, 1983, Feinberg and Vogelstein, 1984). This method is suitable for labelling linear DNA molecules; circular molecules were linearised before labelling. Probe specific activities were in the region of 10^8 cpm/µg DNA. Ideally, 5ng of insert DNA was labelled for every ml of Church buffer, for hybridisations to genomic DNA. If the hybridisation was to cloned DNA, 2ng/ml hybridisation mix was usually sufficient. When total genomic DNA was labelled as a probe, 50ng was labelled with 20µCi of [α ^{32}P]-dCTP.

The final reaction mixture was as follows:

Bovine Serum Albumin (BSA) - 10mg/ml in water

Oligomer Labelling Buffer (OLB)

[α ^{32}P]-dCTP (Amersham; 3000Ci/mmol, 10µCi/µl)

Klenow polymerase (Gibco BRL; 1U/µl)

add DNA and water to 50µl total reaction volume

OLB is made by mixing the following solutions in the ratio (A:B:C) 2:5:3:-

Solution A: 1.25M Tris-Cl (pH 8.0), 0.125M $MgCl_2$, 0.5mM dATP, 0.5M dGTP, 0.5M dTTP (Pharmacia), 250mM ß-mercaptoethanol

Solution B: 2M Hepes (pH 6.6) (Sigma)

Solution C: 90 OD units/ml Hexadeoxyribonucleotides (Pharmacia)

Mix the DNA and water and boil for 10 minutes to denature the DNA. Cool the tube on ice for 1 minute.

2. Add BSA, OLB, label, then enzyme and mix.

3. Incubate at 37°C for 1-4 hours (or overnight at room temperature).

4. Remove unincorporated nucleotides as described in section 2.10.3

2.10.2 End labelling

This method is used to radioactively lable primers.

1. To 30-50pmol of oligonucleotide, add 2µl 5 x T4 polynucleotide kinase buffer (Boehringer Mannheim), 2-4µl (20-40µCi) of [γ^{32}P]-dATP (Amersham 3000Ci/mmol) and 1µl of polynucleotide kinase (10u/µl) (Boehringer Mannheim), make up to 10µl with water.

2. Incubate at 37°C for 30 minutes. (Unincorporated label was not removed.)

2.10.3 Spin dialysis

In order to remove unincorporated nucleotides from labelled DNA, spin dialysis was undertaken

1 After multiprime labelling, add 50µl of TE.

2. Make a column in an Eppendorf tube as follows: pierce a half-hole in the bottom using a 21-gauge syringe needle; add about 25µl of small glass beads (Sigma; 211-300 microns, acid washed and stored in TE) to the tube; fill the column with Sepharose CL6B-200 (Sigma), stored as a slurry in TE. Close the tube and pierce the lid.

3. Part-dry the column by spinning in a bench centrifuge at 1800rpm at room temperature for 2.5 minutes.

4. Place a fresh tube under the column to collect the eluate. Add the labelled DNA mixture and spin as in step 3.

5. Add 100µl of TE to the column, and spin again as in step Th unincorporated nucleotides are retained in the column A rough estimate of incorporation can be obtained by monitoring the column and the eluate.

2.11 Preassociation of labelled DNA

Probes whoe sequence contained repeats were preassociated with sonicated human placental DNA, prior to hybridisation. The placental DNA was in large excess to drive the reaction whereby the repeat elements in the two sources of DNA would hybridise to one another and subsequently not be available for hybridisation to the filter.

1 To the labelled DNA (assumed to be 200µl) add 67µl TE and 10µl of 12mg/ml sonicated placental DNA.

2. Boil the mixture for 10 minutes to denature the DNA Place on ice for minute

3. Add 37.5µl of 1M phosphate buffer (pH 7.2) and leave on ice for 1 hour.

4. Add the prereassociated probe to the hybridisation mix, as described below

2.12 Hybridisation, washing, autography and stripping of filters

Filters were hybridised in bottles in a Hybaid hybridisation oven, using Church hybridisation buffer (Church and Gilbert, 1984).

Hybridisation was performed as follows:

1 Interleave the filters with nylon meshes (Hybaid) and roll into a tube in a tray of 2 x SSC.

2. Place the filters and meshes in a hybridisation bottle, add about 15ml of 2 x SSC, and unroll them onto the inner surface of the bottle by slowly rotating it.

3. Pour off the 2 x SSC and replace with the preheated (65°C) Church buffer. 10ml of buffer was used for 190ml bottles and 15ml for the 280ml bottles.

4. Place the bottles in a hybridisation oven with rotation, at the appropriate temperature (usually 65°C for multiprimed and 42°C for end labelled DNA) and prehybridise the filters for a minimum of 15 minutes.

5. Denature the labelled/purified DNA, and labelled marker if desired, by heating in a boiling water bath for 10 minutes, then cool on ice for 1 minute.

6. After prehybridisation, pour off some of the Church buffer into a universal tube, add the denatured probe to it, swirl and return to the hybridisation bottle.

Hybridise for 14-20 hours.

8. After hybridisation, remove the filters and meshes and rinse briefly at room temperature in 500ml of 400mM Na_2HPO_4, pH 7.2. Discard this solution and add 500ml of preheated (65°C) 400mM Na_2HPO_4, 0.1% (w/v) SDS. Wash the filters in a shaking water bath at 65°C/42°C for 15 minutes.

9. Remove the meshes and soak in warm water. Wash the filters to the required stringency by reducing the concentration of Na_2HPO_4, in the presence of 0.1% SDS, each wash being for 10-15 minutes. The stringency with which each filter was washed was probe dependant. In most cases, filters were washed in 400mM, 100mM and 40mM sodium phosphate/ 0.1% (w/v SDS solutions. For non repetitive probes it was sufficient to stop after the 100mM phosphate/0.1% (w v SDS wash.

10. Surface-dry the filters on Whatman 3MM paper, then seal in cling film. Do not allow them to become fully dry.

11 Place filters in Kodak cassettes with intensifying screens with either R-X film (Fuji), or for quicker but slightly less clear results, X-O-matic film (Kodak). In general, the filter was exposed to 2 pieces of film concurrently. The filter was taped to 1 piece of film to prevent it from being disturbed when the first was removed and developed. This allowed for flexibility and maximum exposure.

12. After autoradiography, strip the residual signal from the filters by washing in 2mM Tris-Cl (pH 7.5), 1% (w/v) SDS, 1mM EDTA at 65°C until the signal appears to have been sufficiently removed. If the signal is not removed totally, place the filters in boiling 0.5% (w/v) SDS. A few seconds may be efficient to remove the residual signal from the filter. If necessary, the efficiency of stripping may be checked by autoradiography. Filters can be stored at -20°C and may be reused.

2.13 Polymerase chain reaction (PCR)

PCR technique (Saiki et al., 1985) was described for the exponent amplification of DNA between two regions of known sequence. The process relies on repeated cycles of heat denaturation, annealing of oligonucleotides to their complementary sequences and extension of the annealed primers using DNA polymerase. The isolation of DNA polymerase from *Thermus aquaticus* (Taq), which retains its activity despite the high denaturation temperatures, has greatly improved this technique.

2.13.1 Oligonucleotide design

Oligonucleotides in the range of 17-24bp were chosen using the Primer programme (Version 0.5: copyright 1991, Whitehead Institute for Biomedical Research) available through the computer facilities at Harrow. This programme facilitates the selection of primers of a specific annealing temperature, which amplify a product within a specific size range. In addition, the programme selects against self-annealing primers or those that contain repetitive elements, such as Alu and LINE elements.

2.13.2 Deprotection and purification of oligonucleotides

Oligonucleotides were synthesised by Dr Val Cooper at the Dyson Perrins Laboratory using an Applied Biosystems Synthesiser. The yield was up to 1mg of detritylated oligonucleotides in about 3ml of concentrated ammonia. Treatment was necessary to remove base protecting groups and care had to be taken to ensure that the primers were not contaminated during the process. The protocol for deprotection and purification of oligonucleotides is as follows:

1 Incubate oligonucleotides at 55°C overnight.

2. Cool to -70°C, dry in a vacuum desiccator at 4°C.

3. Resuspend in 500µl water and extract three times with butan-1-ol

4. Add sodium acetate (pH 5.2) to a concentration of 0.3M and precipitate with ethanol at room temperature.

5. Centrifuge in an Eppendorf centrifuge for 10 minutes.

6. Wash with 70% ethanol, spin again and dry

Resuspend in water. Determine the OD_{260} of a 200 dilution, and use 1OD = 20 mg/ml to calculate the concentration of the purified oligonucleotide

2.13.3 PCR amplification conditions

A Techne Programmable Dri-Block PHC-1 was used for the heating and cooling cycles necessary for the reaction Contamination of PCR was minimised by obeying the following guidelines:

the use of pipettes, which were cleaned regularly with absolute ethanol

UV irradiation for 15-20 minutes, of the buffer, dNTPs and water. Primers without consecutive thymidine residues were also irradiated The production of thymidine dimers in consecutive thymidine residues during UV irradiation reduces the efficiency of PCR amplification, and is therefore avoided)

iii minimising the number of cycles of PCR to reduce the amplification of any contaminating DNA

PCR was generally carried out in 20, 50 or 100µl reaction volu paraffin oil The reaction mix consisted of 20-200ng of tem ate)NA 5mM mM of each primer, 1 x concentration reaction buffer, 0.2mM each of dGTP, dATP, dCTP, and dTTP, and 1U/200µl of Taq polymerase (Boehringer Mannheim, Cetus and Promega)

Samples were processed through an initial denaturation step at 94°C for 5-10 minutes Then, 30-40 cycles of denaturation, annealing and elongation followec The annealing temperature can be determined on the base composition of the primer with the following formula: annealing temperature in °C = 4 x (no. of G or C residues) + 2 x (no. of A or T residues) - 5 For certain primer pairs, non-specific amplification can be a significant problem This was overcome by the following procedures:

"hot-start" PCR:- dNTPs and Taq polymerase were added after a initial denaturation step

ii increasing the annealing temperature of the reaction above that calculated

varying the magnesium concentration of the PCR buffer from 0 to 6mM.

2.14 Purification of DNA fragments

2.14.1 Phenol chloroform purification

If very pure DNA was required, for example for cloning, the DNA was purified using a phenol extraction, followed by a phenol/chloroform extraction and then a final 24:1 chloroform/ isoamylalcohol extraction step. Phenol and 24:1 chloroform/isoamylalcohol were kept at 37°C. Salt was added to a concentration of 0.3M sodium acetate followed by 1 volume of ice-cold isopropanol he Eppendorf tube was placed at -20°C for 30 minutes and then centrifuged for minutes at 4°C. The DNA was washed in 70% ethanol, dried and dissolve water orTE buffer.

2.14.2 Gene clean

For fragments cut out from agarose gels and greater than 500bp, Gene clean (BIO 101) was the method of choice, giving recovery efficiencies of 50-75%.

1. Electrophorese DNA fragment to be purified on an Ultrapure (Gibco BRL) or Seaplaque (FMC Bioproducts) agarose, 1 x TAE gel using as narrow a well as is possible without overloading until there is a good separation between the desired fragment and the surrounding ones.

2. Cut out the band with as little extraneous agarose as possible.

3. Transfer the gel slice to an Eppendorf tube and add 2-3 volumes of 6M NaI

4. Place at 45-55°C, mixing occasionally until the agarose is dissolved. This step should be performed fairly quickly (i.e. under 10 minutes) to prevent oxidation of the NaI solution.

5. Resuspend the Glassmilk (BIO 101) by vortexing and add 5µl Glassmilk for up to 5µg of DNA and an additional 1µl per extra 0.5µg of DNA. Mix well and place on ice for 5 minutes.

6. Centrifuge at room temperature for 5 seconds. Discard the supernatant.

7. Wash the pellet three times with 50-250 x pellet volume of NEW wash (supplied with the kit, containing NaCl, EDTA, ethanol and water). Spin, to pellet, for 5 seconds between washes and vortex to resuspend in NEW wash.

8. After the final wash, spin for 30 seconds and carefully remove any residual liquid so as not to disturb the pellet.

9. Resuspend the pellet in an appropriate volume of water or TE and place at 45-55°C for 2-3 minutes. Centrifuge for 30 seconds and transfer supernatant to a fresh tube. Repeat this step.

10. The DNA concentration and purity can be tested on an agarose gel.

2.14.3 QIAEX II Agarose Gel Extraction

Alternatively, for fragments between 40bp and 50kb this kit was used. See manifacturer's manual for protocol.

2.14.4 Promega PCR minipreps

For fragments amplified by PCR (see 2.13) it was necessary to remove contaminating primers and nucleotides. This method provided a simple and quick way to purify double-stranded PCR amplified DNA. No further treatment was required before using the DNA in other experiments. See manufactures manual for protocol.

2.15 General cloning procedures

The pUC vector series were used for general cloning experiments; a particular one was used on the basis of the presence of the desired restriction enzyme sites in the multiple cloning site.

2.15.1 De-phosphorylation of the pUC vector

A procedure, which cleaves the phosphate groups of the 5' end of the cut vector to prevent it from reanealing to itself. This step is only necessary if the vector is digested with a single enzyme.

1. Digest vector with appropriate restriction enzyme (*Eco*R1) at 37°C overnight.

2. Add sodium acetate (pH 5.2) to 0.3M and mix well.

3. Add 2 volumes of ice-cold ethanol or 1 volume of ice-cold isopropanol and mix gently. Place at -20°C for 30 minutes to precipitate the DNA.

4. Centrifuge the DNA at -4°C and wash the resulting pellet with 70% ethanol.

5. Pellet the DNA; transfer to an Eppendorf tube and air dry.

6. Dissolve the DNA in 10µl TE.

7. Add 10µl of 10x calf intestinal alkaline phosphatase (CIAP) buffer (500mM Tris-HCl, pH9.0, 10mM $MgCl_2$, 1mM $ZnCl_2$, 10mM spermidine).

8. Add 0.01U/pmol of CIAP, make up reaction mix to a final volume of 100µl with dH_2O and incubate at 37°C for 1hour.

9. Stop the reaction by adding 2µl of 0.5M EDTA.

10. Purify DNA as in 2.14.1 to prevent interference of the CIAP with the ligation and transformation process.

2.15.2 Preparation and ligation of insert DNA

1. Digest and run the DNA sample on a Seaplaque or Ultrapure gel if it needs to be separated from other DNA fragments and purify as described in 2.14.2 or 2.14.3.

2. Add 50 - 100ng of the appropriately digested vector DNA to the gel slice and Gene clean the 2 DNA sources together, if possible.

3. Resuspend in 11µl of TE.

4. Add 3µl of 5 X T4 DNA ligase buffer (250mM Tris-Cl (pH 7.6), 10mM $MgCl_2$, 1mM ATP, 1mM DTT, 5%(w/v) polyethelene glycol-8000) and 1µl of T4 DNA ligase (Gibco BRL; 1u/µl).

5. Ligate the mixture at 14°C overnight.

6. Place at 65°C to heat inactivate the ligase.

The ligation mix is ready for transformation at this stage. If transformation is going to be done using electroporation the DNA needs to be reprecipitated and dissolved in 2µl dH_2O as a salt free solution is required for the procedure.

2.15.3 Preparation of bacterial cells for electroporation

1 Grow overnight culture of JM83 using a single colony to inoculate 10ml of TY media.

2. Use the overnight culture to further inoculate 500ml of TY

3. Grow cells until they give an OD reading between 0.5 and 1.0 at 550nm. This usually takes about 3 hours.

4. Transfer the culture into 250ml centrifuge tubes, spin in a Sorvall RC-5B centrifuge using the GSA rotor, at 5000rpm, 4°C, for 15 minutes

5. Resuspend cells in 500ml sterile distilled water kept at 4°C

6. Spin cells as in step 4.

7. Resuspend cells in 400µl sterile dH_2O (4°C).

8. Spin cells as in step 4.

9. Resuspend cells in 10ml of 10% (v/v) glycerol (4°C)

10. Transfer cells to falcon tube and spin at top speed in bench centrifuge at 4°C.

11 Resuspend cells in 1ml of 10% (v/v) glycerol (4°C).

12. Pipette aliquots of 50µl into nunc tubes prechilled to -70°C.

13. Flash freeze cells in liquid nitrogen and store at -70°C or in liquid nitrogen.

2.15.4 Transformation by electroporation

The cells used for electroporation were prepared as described above or obtained from Invitrogen (TopTen). Electroporation was the method of choice for cloning large plasmids.

Remove cells from the liquid nitrogen store and thaw on ice.

2. Add DNA to cells, at a volume not greater than 10%, mix anc ansf chilled cuvette (Bio-Rad) The DNA added to the cells must free alt prevent voltage arcing. The cells must be at the bottom in contac with all four sides of the cuvette, and care must be taken to prevent air bubbles from being present.

3. Place the cuvette in the holder (Gene Pulsar; Bio-Rad) and apply pulse. following settings were used: 25μF, 200Ω and 2.5kV.

4. Immediately add 1ml of SOC, at room temperature.

5. Incubate in a shaking 37°C incubator for 1 hour and then plate onto selective medium (depending on antibiotic resistant marker of cells).

2.15.5 Preparation of bacterial cells for transformation by heat shock

The *E.coli* strain JM83 was made competent by the following frozen storage protocol and used for all general transformation procedures:

1. Pick several 2-3mm colonies from a freshly streaked SOB agar plate and disperse in 1ml SOB medium by vortexing. Use one colony per 10ml of culture medium.

2. Inoculate SOB medium with the dispersed cells in an Erlenmeyer flask (100ml in a 1 litre flask) and incubate at 37°C with moderate agitation until the cell density is 6-9 x 10^7 viable cells/ml (OD_{550} of 0.2-0.4).

3. Collect the culture in 50ml Falcon tubes and chill on ice for 10-15 minutes.

4 Pellet the cells at 2000-3000 rpm at 4°C for 12-15 minutes, then drain the pellets thoroughly

Resuspend the cells in frozen storage buffer (FSB) in 3 of the culture volume by moderate vortexing and store on ice for 10-15 minutes.

6. Pellet as before and drain thoroughly.

7 Resuspend in FSB in $^1/_{12.5}$ of the original volume. Add DMSO to 3.5% (v/v

Squirt the DMSO into the centre of the cell suspension and immediately swirl the tube for 5-10 seconds. Incubate on ice for 5 minutes.

8. Add a second volume of DMSO, as in step 7 to bring the final concentration to 7%. Incubate on ice for 10-15 minutes.

9. Pipette 100 or 200µl aliquots into chilled screw cap polypropylene tubes, then flash freeze by placing the tubes in a dry ice/alcohol bath or into liquid nitrogen for several minutes. Store aliquots at -70°C or in liquid nitrogen.

The transformation efficiency of the cells was tested using 0.1ng pUC9.

2.15.6 Bacterial transformation by heat shock

Remove an aliquot of competent cells from the freezer, thaw at room temperature until just liquid and then store on ice.

2. Add DNA (in a volume of 20µl or less) and mix by swirling.

3. Incubate on ice for 10-60 minutes.

4. Heat shock the cells at 42°C for 90 seconds. Quench the heat shock on ice.

5. Add 800µl SOC and incubate at 37°C for 30-60 minutes, shaking occasionally.

Bacteria were then plated out on medium containing ampicillin for selection and X-gal for the identification of recombinant clones. As an insert interrupts the β-galactosidase gene in the vector, the chromogenic substrate X-gal cannot be converted to a blue product and white colonies will indicate the presence of a recombinant. IPTG was sometimes used in the plates as it intensifies the blue colour of non-recombinant clones.

2.15.7 Transformation using TA cloning kit

This kit, obtained from Clonetech, provides a quick, one step cloning strategy for direct insertion of a PCR product into a plasmid vector. See manufacturer's manual for protocol

2.16 Sequencing techniques

2.16.1 Manual sequencing

Sequencing reactions were carried out using the Sequenase version 2.0 enzyme (Tabor and Richardson, 1990) and the accompanying kit (USB). The method is based on the chain-termination technique (Sanger et al 1977). All sequencing of restriction fragments cloned into pUC9 used the 17-mer M13 reverse primer (Pharmacia) and the 17-mer M13 universal primer -40 (USB) to sequence from both ends. The sequencing of restriction fragments using the pGEM-11Zf vector used 22-mers. At the 5' end the T7 forward primer and at the 3' end the SP6 reverse primer to sequence from both ends. Reactions were prepared according to the protocol accompanying the Sequenase version 2.0 ki

DNA was prepared using the alkaline lysis protocol with an additional poly-ethylene glycol (PEG) purification step. After washing the DNA in 70% ethanol, dissolve the pellet in 200µl TE/RNase (10mg/ml) and incubate at 37°C for 30 minutes. Add 120µl PEG [20% PEG-6000, 2.5M NaCl], mix and leave on ice for 1 hour. Centrifuge at 4°C for 10 minutes. Remove the supernata ith c wash in 70% ethanol. Dry the pellet and resuspend in TE

A. Annealing reaction

1. Dissolve 3-5µg of DNA in an equal volume of 0.4M NaOH/ 4mM EDTA Incubate at 37°C for 30 minutes to denature the DNA.

2. Add sodium acetate (pH 5.2) to a final concentration of 0.3M followed by volumes of ice-cold ethanol or 1 volume of ice-cold isopropanol Place at -70°C for 15 minutes.

3. Spin in an Eppendorf centrifuge at 4°C for 15 minutes.

4. Pour off supernatant and wash with 70% ethanol Spin in an Eppendorf centrifuge at room temperature for 5 minutes. Carefully pour off supernatant

5. eave to air dry or lyophilise for as short a time as possible, then dissolve ir 7µl water.

6. Add 2µl of 5 x Sequenase reaction buffer (0.2M Tris-C (pH 7.5), 0.1M

$MgCl_2$,0.25M NaCl) and 1µl of primer (1pmole/µl). Incubate at 37°C for 15-30 minutes to anneal the primer.

This mix can be kept on ice for up to 4 hours.

B. Labelling reaction

1. Dilute 5 x labelling mix (dGTP) (7.5mM each of dGTP, dCTP and dTTP) 5-fold with water.

2. Dilute the Sequenase 2.0 enzyme 1:8 in ice-cold enzyme dilution buffer (10mM Tris-Cl (pH 7.5), 5mM DTT, 0.5 mg/ml BSA) - this solution should be kept on ice and used within 1 hour.

3. To the annealed template/primer add 1µl 0.1M DTT, 0.5µl [α^{35}S]-dATP (10µCi/µl) (Amersham), 2µl of the diluted labelling mix and lastly 2µl of the diluted Sequenase enzyme. Mix and incubate at room temperature for 2-5 minutes.

C. Termination reaction

1. Place 2.5µl of each dideoxynucleotide - ddGTP, ddATP, ddTTP and ddCTP (80mM of each dNTP, 8mM of ddNTP, 50mM NaCl) in a suitably labelled Eppendorf tube. Prewarm the tube, with the lid on to prevent evaporation, at 37°C for 1 minute.

2. When the labelling reaction is complete, add 3.5µl of the labelled mix to each of the tubes containing the dideoxynucleotides. Mix, centrifuge briefly and incubate at 37°C for 3-5 minutes. This step can be extended for up to 30 minutes without any detrimental effect.

3. Add 4µl of STOP solution , mix and store on ice for loading on polyacrylamide gel. If the samples are to be stored, they should be kept at -20°C.

2.16.2 Polyacrylamide gel electrophoresis (PAGE)

Sequencing reactions were run on 6% denaturing polyacrylamide gels. A stock acrylamide solution was prepared as follows:

A 40% (w/v) acrylamide solution was prepared containing acrylamide: N,N'-methylenebisacrylamide (BDH) at a ratio of 19:1. The solution was filtered through Whatman no.1 paper and stored in dark bottles at 4°C. The final polyacrylamide solution contains 6% (w/v) acrylamide, 4.2% (w/v) urea, 1 x TBE. Two 55cm glass plates were treated individually with Bind-Silane (LKB A-174) or 'Repelcote' (BDH) respectively to ensure that the gel attached to one plate only, when they were separated after running. The acrylamide solution was polymerised by the addition of 0.1% (v/v) TEMED (Sigma) and 1% (v/v) of a 10% (w/v) ammonium persulphate solution and poured between the two glass plates separated by 0.2mm spacers. The gels were allowed to set for 30 - 60 minutes. An LKB 2010 Macrophor Sequencing System was used for running the gels.

An aluminium backing plate was clamped onto the back plate of glass to reduce temperature differences within the gel. Once in place, the wells were carefully flushed out with TBE to remove urea and the gel was pre-run for 45 - 60 minutes at 27-36V/cm in 1 x TBE buffer. Prior to loading, the wells were flushed out again and the samples were denatured by heating to 94°C for 3 minutes, then placing immediately on ice. Approximately 2µl of sample was loaded into each lane using a macrophor sample syringe with glass fibre needle (LKB 2010-150) and electrophoresis allowed to proceed at 27-36V/cm.

After electrophoresis, the gel plates were removed from the apparatus and carefully separated, leaving the gel attached to the glass plate which had been treated with 'Bind-silane'. The gel was placed in 10% acetic acid/10% methanol (v/v) for 30 minutes to allow the urea to diffuse out of the gel, and was then allowed to dry for 2 hours or more at 60°C. RX X-ray film (Fuji) or hyperfilm bmax (Amersham) was exposed to the gel at room temperature overnight.

2.16.3 Dye terminator cycle sequencing

The Thermo Sequenase Dye Terminatior sequencing pre-mix kit (Amersham) was used for this method. See manufactures manual for protocol. The ready samples were given to the Sequencing Facility in the Department of Biochemisty, University of Oxford, to be run on an ABI Prism™ 377 fluorescent sequencing instrument.

CHAPTER 3 - Physical mapping of the DXS8237E - UBE1 - PCTK1 region, which is located on the short arm of the X chromosome at Xp11.3

3.1 Introduction

Physical mapping frequently depends on the assembly of 'contigs", which represent the structure of contiguous regions of the genome by specifying the overlap relationships among a set of clones. They are dependent on the continuing existence of a particular underlying clone collection and the detailed analysis of individual clones, which includes restriction enzyme mapping. By using a detailed physical map which includes restriction sites, specific stretches of DNA of a certain region can be examined in more detail. Markers may be established and subsequently investigated The DXS8237E - UBE1 PCTK region is of particular interest, since the gene *DXS8237E* is inactivate orm *UBE* and *PCTK1* gene appear to escape this process. he nactivation statu another gene on the X chromosome encoding a ubiquitin C-terminal hydrolase, *UHX1*, located distal, but in the close vicinity to *PCTK1*, is not known. A physical map will aid in the analysis of the DNA sequence containing the inactivation boundary. It will also provide a valuable comparison for the only other map existing of this region. This is of importance, particularly because the four genes lie very close to each other in an area of high guanine cytosine (GC) ontent potential instability

At the start of this investigation a preliminary physical map of this region was constructed using overlapping sets of yeast artificial chromosomes (YACs), according to their sequence tagged site (STS) content and some internal markers. These YACs, although providing starting material for further characterisation, had limitations in constructing a comprehensive map because two out of three were either chimæric or rearranging. To overcome the problems inherent to YAC technology, P1 derived artificial chromosomes (PACs) were chosen to attempt the detailed mapping of the region. These were intended to provide the basis for the isolation of a single fragment containing all three genes, *DXS8237E*, *UBE1* and *PCTK1*, which after purification, cloning and sequencing could be analysed with regard to X chromosome inactivation boundaries.

3.1.1 Yeast Artificial Chromosomes (YACs) and YAC libraries

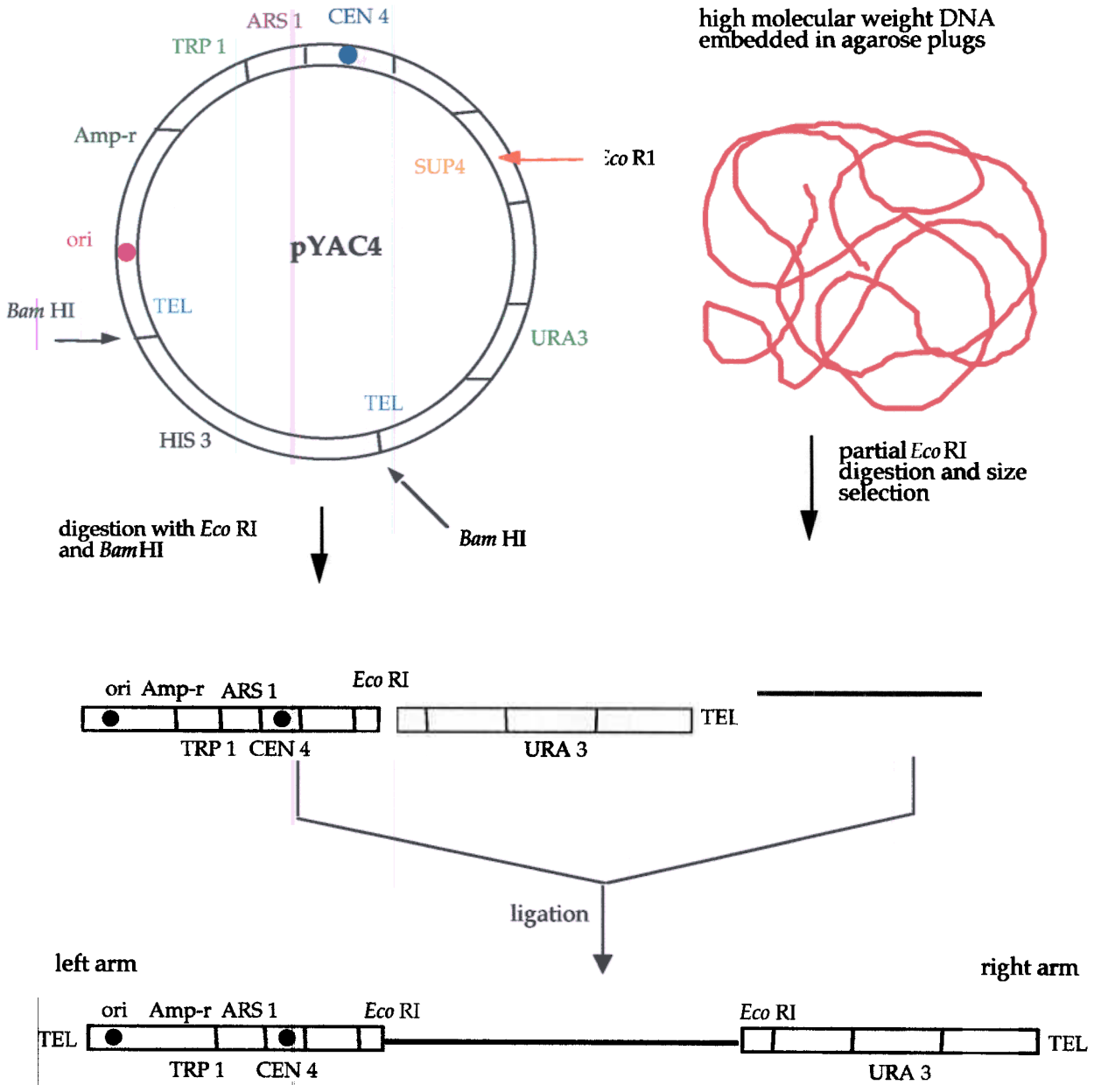
YACs are cloning systems based on the *in vitro* construction of linear DNA molecules of sizes up to and exceeding 1 megabase pairs (Mb) that can be transformed into yeast, where they are maintained as artificial chromosomes (Burke et al. 1987).

The vector system used in this study was pYAC4. This YAC is circular and contains the following pBR322 derived sequences: a replication origin (ori) and gene for ampicillin resistance (Amp). These sequences allow the YAC to be replicated as a single plasmid in the bacterial host, *Escherichia coli*. In order to replicate and be stable in the yeast *Saccharomyces cerevisiae*, three other elements are essential: a segment which can act as a fully functional centromere in mitotic and meiotic cell divisions (CEN4), an autonomous replicating sequence, which allows replication of colinear DNA (ARS1) and two sequences, which function as yeast telomeres *in vivo* (TEL) and are separated by a HIS3 gene, which is lost during the cloning procedure. CEN4 and ARS1 are cloned from yeast chromosome IV and the latter has shown to replicate once per cell division at a time coincident with the corresponding chromosomal sequence (Hieter et al., 1985). Additionally, the vector contains two selection markers, TRP1 and URA3, on either of the *Eco*R1 cloning site, created within the 14bp intron of SUP4, whose interruption is phenotypically visible by the production of red colonies. The cloning procedure is shown in figure. 3.1.1

YAC libraries have been constructed by preparing and size-fractionating high molecular weight DNA in solution by sucrose gradient density separation (Burke et al., 1987; Coulson et al., 1989; Garza et al., 1989) or in agarose by pulsed field gel electrophoresis (Albertsen et al., 1990; Anand et al., 1989; McCormick et al., 1990). Pulsed field gel electrophoresis produces more effective fractionation than sucrose density gradients, but results in DNA which is embedded in agarose and therefore is more difficult to handle than DNA in solution and which ligates and transforms yeast with reduced efficiency. This problem is overcome by a combination of high resolution separation of PFGE with subsequent removal of the agarose (Anand et al., 1989). Size fractionation before and after ligation has also been used to increase the size of clones (Albertsen et al., 1990). Addition of polyamines proved effective to protect the DNA from shearing forces prior to the

Figure 3.1.1: Cloning into pYAC4 - construction of a YAC

(adapted from Burke et al., 1987)



The insert DNA is partially digested with *Eco* RI and size fractionated. The vector is digested to completion with *Eco* RI and *Bam* HI and the sequence between the *Bam* HI sites is lost. The vector is phosphatased to prevent religation into the plasmid form. The vector and insert DNA are ligated and then transformed into spheroplasted *S. cerevisiae* using selective media for both the left and right arms. Pink transformed colonies that lack SUP 4 activity are chosen.

ARS 1: an autonomous replication sequence which allows replication in the host.
CEN 4: provides centromeric activitiy
TEL: telomeres derived from *Tetrahymena* ribosomal DNA macronuclear molecules.
SUP 4: an ochre supressing allele of tyrosine tRNA which is inactivated in recombinant clones. prevents suppression of a mutation in adenine biosyntheses resulting in the build-up of a red co metabolite.
TRP 1 and URA 3 allow for selection of both vector arms.
ori: origin of replication

melting of the agarose (Larin et al., 1991

Another major problem of YAC libraries is the co-cloning of non-contiguous fragments in a single insert (chimærism). Some chimæric YACs arise from ligation of two or more restriction fragments, but it is likely that most derive from a recombination event between two YACs or YAC fragments in the same cell via two of the many common mammalian repeats. Chimæric YACs can be identified by fluorescence *in situ* hybridisation (FISH), by mapping sequences isolated from the YAC against hybrid panels, or by chromosome walking restriction mapping. An alternative approach is to make the library from somatic cell hybrid. Because the hybrid contains only 2% of human DNA, Lee et al. (1992) were able to calculate that the rate of chimærism was 11% by determining the number of YAC inserts containing both hamster and human sequences amongst human positive YACs.

Despite the efforts put into the production of YAC libraries, the major disadvantage of YACs is still the frequency of chimærism. One more problem occurring with YACs is that of deletions. Although the cause of the instability of certain genomic regions in YACs is uncertain, the available evidence suggests that deletions occur between repetitive elements, common to the genome. The deletions may occur with a high enough frequency that the original size of the YAC clone is never seen.

3.1.2 Restriction mapping of YACs

While cloning the ends of YACs and demonstrating their presence in more than one YAC is useful in establishing an overview of YAC position within a contig, the technique does not show the exact overlap between YACs, nor does it highlight regions of other interest in the YACs. Restriction maps of YACs and localising internal and end clones within them, provide a more detailed picture. In addition, discrepancies between restriction maps of overlapping YACs may be an indication of chimæric YACs and/or regions of instability.

A common approach is to do a range of partial digests using a series of enzymes (Smith and Birnsteil, 1981). Rare cutting enzymes, those that have either an 8bp recognition sequence (e.g. *Not*I, *Sfi*I) or that have CpG dinucleotides as their recognition sequence (e.g. *Bss*HII, *Eag*I, *Nru*I, *Sac*II), generate large fragments of

DNA that can be separated and sized using pulsed field gel electrophoresis (Lindsay and Bird, 1987).

Sequence-tagged sites (STSs) derived from end fragments of YACs can facilitate the assembly of an overlapping YAC/ STS map. Improvements to the maps usefulness can be made rapidly by incorporating supplementary STSs from genes and genetic linkage probes with known locations.

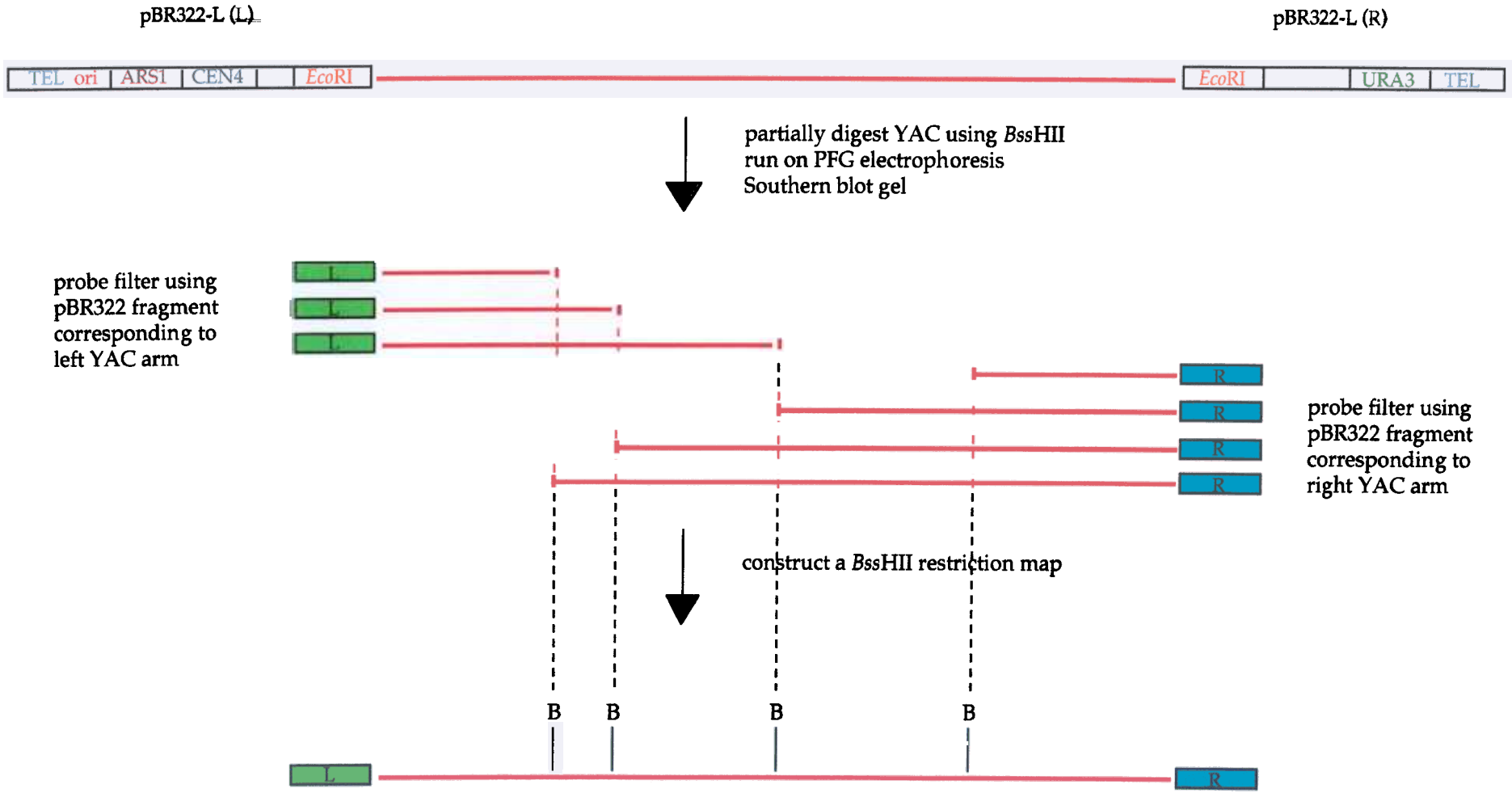
As the pYAC4 vector includes defined sequences of the plasmid pBR322 in its arms (as described above), it is possible to map both arms of the YAC separately by using them as probes. Hybridisation of filters obtained from pulsed field gels with these probes detects two series of overlapping fragments which end in the left or right end of the YAC. Hence, it is possible to construct a restriction map. One problem with this technique is that the secondary structure of DNA and the base composition of neighbouring sequences can prevent or slow down the activity of a restriction enzyme, which may result in the restriction site not being recognised. A diagram of restriction mapping of YACs using the pBR322 vector is shown in figure 3.1.2.

3.1.3 P1-derived Artificial Chromosomes (PACs) and PAC libraries

In the last few years, with the human genome project progressing rapidly, it has become increasingly important to have clones available that represent the genome in a contiguous and non-rearranged fashion, possess more accuracy and stability and produce fewer cloning artefacts than YAC clones. With recent developments in host/vector systems it is now possible to create large insert libraries using low copy number bacterial plasmid vectors. Suitable vector systems have been developed using *E. coli* F factor and P1 phage replicons. Both systems are capable of producing large clones by bacterial transformation. The latter one extensively referred to in this thesis is discussed in some detail below.

The development of P1 clones by Sternberg opened up new possibilities for the scientific world. They have a genome size of approximately 100kb with identical 10kb terminal sequences; they can accept 70-90kb of insert DNA, which compares favourable with cosmid vectors (which can accommodate only 45-48kb), but is more restricted when compared to YAC vectors (carrying 200 to >1000kb of DNA fragments). Nevertheless, they do obviate some of the difficulties encountered

Fig.3.1.2 Restriction mapping of YACs
(adjusted from Silverman et al., 1989)



with YACs : firstly, the low efficiency of transformation with vector DNA containing large inserts (about 10^5 transformants are produced per µg of vector) secondly, the decreasing transformation efficiency with increasing insert size; thirdly, the need to process transformants individually prior to screening; and fourthly, the difficulty in obtaining large amounts of insert DNA from transformed cells (Sternberg, 1990).

When the viral DNA is injected into the host cell, homologous recombination between the terminally redundant sequences produces a circular molecule Once circularised, the phage can establish either a proviral or lytic cycle. In the former, the lytic functions are repressed and the phage DNA is maintained sing copy per cell (Pierce and Sternberg, 1992). As a result of this, termole u recombination does not occur, thereby possibly eliminating one caus of instability (Sternberg et al. 1990).

Sternberg et al. (1990) constructed a 50,000-member human DNA library with DNA inserts ranging in size from 75 to 100kb. Further an arrayed human genomic library, generated with the P1 cloning system (Shepherd, 1994), was constructed by producing DNA inserts by size fractionation of a *Sau*3AI partial digest of high molecular weight genomic DNA isolated from primary cells of human foreskin fibroblasts. The inserts were cloned into the pAd10sacBII vector and packaged *in vitro* into P1 phage. These were used to generate recombinant bacterial clones. The resulting library, DMPC-HFF//1 series A, consists of 130,000 - 140,000 recombinant clones that were stored in 1500 microtiter dishes. The cloned inserts have an average size of 80kb, which in theory provides a genome coverage of 3- to 4-fold

PACs are P1-derived artificial chromosomes. Firstly designed by Ioannou et al (1994), who established an initial 15,000 clone library with and average insert size of 130 - 150kb. They combine the features of the P1 vector and the *E.coli* fertility plasmid or F-factor (from BACs) systems, to construct a new P1-derived vector, pCYPAC1. This is capable of cloning inserts in the range of 100 to 300kb in bacterial cells. Further, the authors observed stable propagation of large inserts and a lower level of cloning artifacts when creating a library. These details are important for detailed physical mapping and functional genome characterisation.

A schematic representation of the pCYPAC2 vector, on which the PACs employed in this investigation are based, is shown in figure 3.1.3. It has been constructed for the cloning of large DNA fragments following electroporation

has been achieved by the removal of the stuffer ragment form the pAd10SacBII vector (Pierce et al .992), followed by the insertion pUC plasmi nto the *Bam*HI cloning site in an non permanent fashion During the cloning process, pUC sequences are removed through a double-digestion scheme, which employs *Bam*HI and *Sca*I. The pUC plasmid cannot reinsert back into the *Bam*H loning site, which would otherwise result in the creation of non-recombinant background clones in the library.

Re-ligation into the *Bam*HI site is prevented at three levels:

1. pUC is cleaved into *Sca*I fragments (ends incompatible with *BamH*I).
2. The *Bam*HI/ *Sca*I oligo linkers are physically separated from the pUC and PAC-vector fragments.
3. All vector-derived fragments are treated with alkaline phosphatase to inhibit ligation of vector-to-vector ends.

features include that the pCYPAC2 clones have kanamycin antibiotic resistance and the *Sci*I site has been replaced with a second *Not*I site to allow th release of most inserts as a single large fragment (http://128.205.166.214/cgi-bin/vector-query?Vetor=pPCYPAC2&Choice=I). This facilitates size determination of inserts.

For the current investigation, the RPCI Human PAC libraries 3 and 5 were employed These were constructed from anonymous male DNA samples extracted from blood The cloned inserts have an average size of 100 to 30kb The libraries were arrayed into 384-well microtiter dishes and gridded onto 22x22 cm nylon high density hybridisation filters for screening purposes (http://bacpac.med.buffalo.edu/libchar.htm)

Restriction maps of PACs are produced in the same way as they are for YACs and the mapping procedure using the pCYPAC2 vector is very similar to using the pYAC4 vector. Here also, the two known end-sequences, T7 and SP6, are used as probes in order to map the two arms individually One big advantage using PACs is that they have proven to be more stable than YACs (Stoddart e al 1999

3.1.4 Restriction enzymes used for mapping

Rare cutting enzymes are a important tool in restriction mapping. In mammals, sites for rare cutting enzymes are concentrated in CpG islands. CpG islands are sequences of 0.2-3.0kb characterised by high GC levels and unmethylated CpG doublets. In addition to making the construction of long range maps possible, they are important in identifying new genes (reviewed by Larsen et al., 1992). All houskeeping genes and 40% of tissue specific or limited expressed genes are associated with CpG islands (Lewin, 2000). Larsen et al. (1992) classified 26 rare cutting restriction enzymes into 4 classes based on length and composition of their recognition sequences, the abundance of sites and percentage of sites within islands. Cleavage was known to be blocked by C_5 methylation of CpG in all, but one of those enzymes. The authors results suggest that class enzymes, which includes *Not*I, are second best for long range mapping. Class II enzymes have a high frequency of sites within CpG islands and data suggest that a cluster of three enzymes, even in cloned DNA, is a very good indicator of a CpG island and an associated gene. Further the authors conclude that the class II enzymes *Bss*HII, *Eag*I and *Sac*II are more suitable for detailed mapping than class I- and III type enzymes. Class III enzymes, which include *Nru*I, cleave genomic DNA most infrequently, resulting in fragments on average >700kb. Class IV enzymes, including *Xho*I, have one-third or fewer of their sites within CpG islands be used as additional mapping enzymes.

In this study, the class II enzymes BssHII, *Eag*I and *Sac*II were used for restriction mapping of the YACs and *Bam*HI, *Bss*HII, *Eag*I, *Eco*RI, *Hin*dIII, *Not*I, *Nru*I, *Sac*II, *Sfi*I, *Sma*I and *Xho*I were chosen for the PAC restriction map. These enzyme cover all classes established by Larsen et al. (1992) and further analysis was completed with *Bam*HI and *Eco*RI which are more frequent restriction enzyme used on its own to determine overlaps of fragments and *Sf* a 13 recognition cutter, were also used. This choice of enzymes was the same as that of Carrel et al. (1996), which made a comparison between the map generated in this study to the one published by Carrel et al (1996) feasible

3.1.5 Pulsed field gel electrophoresis (PFGE)

The principle of PFGE is best understood if DNA molecules are considered as worm-like coils. In agarose gel electrophoresis, where the dimensions of the coils are compatible to the size of the pores in the gel, the separation of different size DNA fragments results from different rates of migration through the gel matrix due to the different mobility of different sized molecules. However, extremely long DNA fragments form coils whose dimensions are significantl larger than the average pore size of the gel. When an electric field is applied, the molecules may enter two or more pores, become trapped and therefore not separate (Schwartz and Cantor, 1984). Separation by PFGE is achieved by subjecting the DNA molecules to alternatively pulsed, perpendicular orientated electric

Several related systems have been developed; field inversion gel le ·ophor (FIGE), which involves periodically reversing polarity in a standard horizontal apparatus (Anand, 1986). A second method uses a uniform circular gel in a uniform electric field. The gel is rotated through an angle of 110 degrees at each switching interval, which can be varied according to the size of the to separated fragments (Southern et al., 1987). The CHEF system is based on contour-clamped homogeneous fields which are 120 degrees apart (Chu, 1989) The wo systems have been used in the investigations reported in this thesis.

PFGE has proven essential for generating long-range restriction maps of genomic DNA. Such maps typically cover regions over 500kb and can extend well over 5Mb. Therefore the development of this technique, together with the isolation of restriction enzymes whose cleavage sites are rare in mammalian genomes was a critical breakthrough that permitted long-range mapping to define the genome structure.

This Chapter describes the production of clones from the region around Xp11.3 and a physical map achieved using rare cutting restriction enzymes and pulsed field electrophoresis. The clones have been converted to sequence-tagged sites (STSs) and a map has been constructed using these markers. Probes from genes mapping to the region have been used to define overlaps between the YACs and PACs spanning the region.

Initially, one of the aims was to extend the existing contig distally to *PCTK1*. This was going to be done by using the generated STSs to obtain further YACs, which would have subsequently been examined for CpG islands to help in identifying possible genes in the distal region to *PCTK1*.

3.2 Materials and Methods

YACs

3.2.1 Preparation of Yeast Artificial Chromosome (YAC) DNA in solution

Protocol

1. Inoculate 100ml SD+CAT medium in a 1 litre flask with a single YAC colony. Grow for 36 hours at 30°C with vigorous shaking (450rpm).

2. Harvest cells by spinning at 2000rpm in a bench centrifuge at room temperature for 20 minutes.

3. Decant and resuspend cells in 10ml of 40mM EDTA/ ß-mercaptoethanol and spin for 5 minutes.

4. Resuspend pellet in 1ml of SCE with 10µl of 100mg/ ml Novozyme (Novo Nordisk). Incubate at 37°C for 2 hours.

5. Spin at 5000 rpm for 1 minute and discard supernatant.

6. Resuspend pellet in 200µl of Y-spheroblast solution and incubate at 68°C for 15 minutes.

7. Add $^1/_2$ volume phenol equilibrated with TE, mix by inverting, then add $^1/_2$ volume of chloroform (24:1 chloroform : isoamyl alcohol) and mix as before. Spin for 5 minutes to separate phases.

8. Repeat step 7.

9. Transfer the supernatant to a fresh tube and add 1 volume of cold (4°C) isopropanol. Stand at room temperature for 15 minutes.

10. Spin at RT, 1200rpm for 5 minutes and let pellet air dry.

1 Dissolve pellet in 300µl of STE and 3µl of 10mg/ ml RNAse (Boerhinger Manheim)

12. Incubate at 37°C overnight.

12. Repeat steps 9 and 10.

13. Dissolve pellet in 200µl of TE.

For PCR reactions 2µl of the above were added to 35µl reactions.

3.2.2 Plasmid rescue for left end cloning

The left end of the YAC vector, pYAC4, can be liberated from the rest of the YAC and following recircularisation, can function as a plasmid because it contains the pBR322 origin of replication (ori) and an ampicillin resistance gene which provides a marker for the selection of the plasmid. The plasmid also contains the end fragment of the insert DNA which can be used as a hybridisation probe and/or sequenced to provide a sequence-tagged site (STS) (Olson et al., 1989).

3.2.3 Cloning left and right end fragments by inverse PCR

Inverse PCR (IPCR) is an adaptation of the general PCR technique for exponentially amplifying DNA sequences between two known primer sequences. This technique allows the amplification of DNA upstream or downstream of the region of known sequence (Ochman et al., 1988). Cloning the left and right ends (see figure 3.2.1) of YACs can be achieved using this technique. The amplified fragment can be used directly as a probe, or can be cloned into an appropriately cut vector.

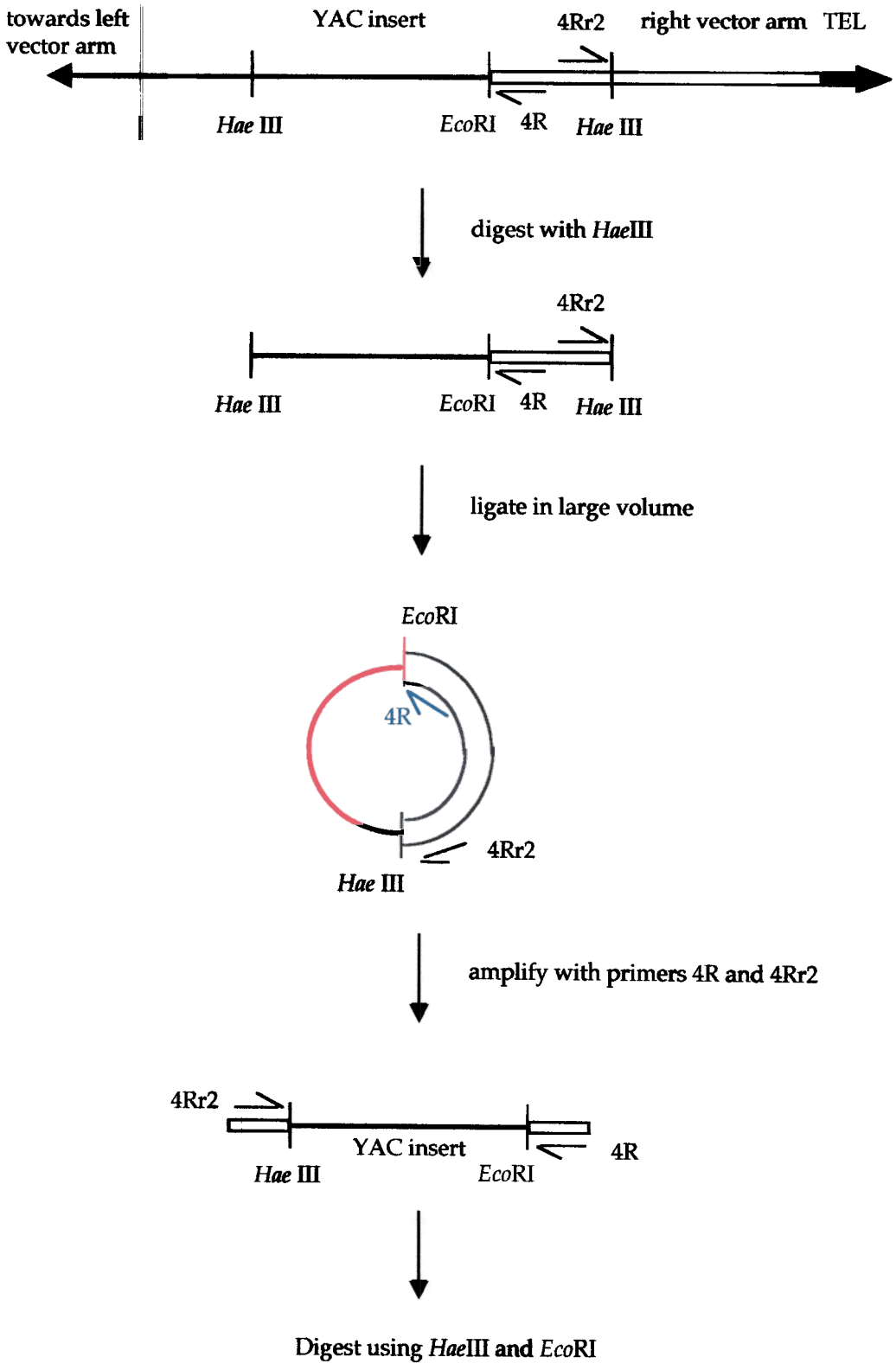
Protocol

1 Digest 100ng of DNA with *Hinc* II and/or *Hae* III in separate Eppendorf tubes

2. Phenol/chloroform extract the DNA and dissolve in 5µl dH_2O.

3. Ligate the DNA in a total volume of 100µl using 20µl T4 DNA ligase buffer and 7µl T4 ligase.

Fig. 3.2.1 Schematic diagram showing the technique of inverse PCR used to isolate the right end of a YAC insert



4. Use 1, 5 and 10µl of the ligation mix as the template in separate 100µl PCR reactions.

The primers used for amplification depend on the enzyme used for digestion. The primer 4R can be used with both *Hinc* II and *Hae* III and 4R rev1 should be used in conjunction with *Hinc* II and 4R rev2 with *Hae* III.

4L	5'	AAG TAC TCT CGG TAG CCA AG	3'
4L rev	5	TCA GAG TGA AAT TTA TAC TAA	3
4R	5	AGT CGA ACG CCC GAT CTC AA	3
4R rev1	5'	GGA GTC GCA TAA GGG AGA GC	3'
4R rev2	5'	TTC AAG CTC TAC GCC GGA	3

5. The PCR program is 94°C for 1min, 55°C for 2min and 74°C for 2min for 35 cycles with a final extension of 74°C for 5min.

6. Run 10µl of the PCR reaction on a gel. If necessary, use the PCR amplified product as template for a further round of PCR.

3.2.4 Preparation of Yeast Artificial Chromosome YAC DNA in plugs

Protocol

1 Inoculate 100ml SD+CAT medium in a 1litre flask with a single YAC colony Grow for 36 hours at 30°C with vigorous shaking (450rpm).

2. Harvest cells by spinning at 2000rpm in a bench centrifuge at room temperature.

3. Decant and resuspend cells in 40ml 50mM EDTA (pH 8.0). Centrifuge as before.

4. Resuspend in 4ml SEB containing 20U/ml lyticase (Sigma no. L5263, dissolved in 50mM Tris-Cl (pH 7.5); 1mM Na_2EDTA; 10% (v/v) glycerol).

Melt 1% Seaplaque agarose (FMC Bioproducts) in SEB. Cool to 37°C.

6. Mix 5ml yeast suspension with 5ml agarose and hold at 37°C.

7. Pipette the mixture into chilled plugmoulds.

8. Eject the plugs into 40ml of STEB containing 20U ml lyticase. Incubate at 37°C with gentle shaking for 2 hours.

9. Replace with fresh LiDS. Incubate at 37°C for 1-16 hours.

10. Replace with fresh LiDS.

11 Store at room temperature.

3.2.5 Restriction enzyme digestion of DNA in plugs

Plugs were not allowed to come into contact with ferrous metals as iron forms complexes with Na_2EDTA, which can be nucleolytic. They were cut using glass cover-slips and handled with glass loops. All instruments were washed in 20% (v/v) ethanol, 0.1% (w/v) SDS before using.

Protocol

1. Wash each YAC plug 3 x 30 minutes in 1.5ml TE (made from AnalaR Tris) at 50°C to reduce the Na_2EDTA concentration, from 100mM to 1mM, as it inhibits enzyme activity. Cut each plug into thirds for digests.

2 Equilibrate each section of the plug with 400µl of 1 x restriction buffer at room temperature for 30 minutes.

3. Replace this with 60µl of digestion buffer (0.1mg/ml Gelatin, 1 x restriction buffer, 1mM DTT, spermidine (as in section 2.7)) and leave at room temperature for 10 minutes.

4. Add the restriction enzyme and place at the reaction temperature. For partial digests leave on ice for 30 minutes to allow diffusion of enzyme before placing at the reaction temperature.

Partial digests were achieved by keeping the reaction time constant and varying the enzyme concentration. 15-20 units of enzyme were used for complete digests and 0.1 to 15 units for partials depending upon the efficiency of the enzyme.

For storage of digested plugs, 400µl of NDS was added and they were kept at 4°C. Plugs were equilibrated in 100µl of STOP solution for 15 minutes before loading onto a pulsed field gel.

3.2.6 Pulsed field gel electrophoresis (PFGE) using YACs

Electrophoresis was carried out in a rotating plate, or 'Waltzer' device (Southern et al., 1987), using 0.5 x TAE as running buffer and Type I, low EEO agarose (Sigma). The apparatus has a capacity of 6 litres, circulated by means of an Eheim pump, which is cooled by means of an LKB multi-temperature thermostatic cooler.

For the majority of pulsed field gels the following conditions were used: 1.5% agarose gels, 28-32 seconds pulse time, 28-32 hour running time, 16°C running temperature and a field strength of 6V/cm. Gels used for band elution were made of Seaplaque agarose with a Type I agarose frame around them.

Lambda oligomers (Promega) were used as markers for high molecular weight DNA. They give a ladder with steps 48.5kb apart with a size range of 50 to 1000kb. Gels and filters were treated as in section 2.8 ensuring a 4 minute UV exposure prior to blotting to nick the DNA. This aids in the efficient transfer of large fragments. Blotting was carried out for 18-24 hours.

PACs

3.2.7 Preparation of P1-derived Artificial Chromosome (PAC) DNA

After growing a single colony in 10ml of 2xTY medium containing 0.2% glycerol and 30µg/µl kanamycin at 37°C over night following procedure was followed:

Harvest cells by centrifuging in a bench centrifuge at 3000 rpm and room temperature for 10 minutes.

2. Remove supernatant, re-suspend pellet in 400µl of 50mM Tris-HCl (pH 8) containing 10mM EDTA, giving a final volume of 600µl. Divide into two 2ml Eppendorf tubes.

3. Add 400µl of 0.2M NaOH containing 1% SDS, invert to mix and leave on ice for 5 minutes.

4. Add 600µl of 1% SDS, 0.2M NaOH solution into each tube. Invert tube to mix and immediately add 450µl of 7.5M ammonium acetate, invert to mix. It is important to work with one tube at the time in order to avoid DNA damage due to alkaline pH. Incubate on ice for 5 minutes.

When the DNA solution is pipetted it is important to use cut tips from this point onwards, in order to avoid sheering of the DNA.

5. Spin at 13000rpm and room temperature for 25 minutes and supernatants into two new 2ml Eppendorf tubes.

6. Add 800µl of isopropanol to each tube, mix by inverting and centrifuge for 15 minutes

7. Resuspend resulting pellet in 100µl of dH_2O, add 50µl of 7.5M ammonium acetate to each tube and incubate on ice for 10 minutes.

8. Spin at room temperature for 15 minutes and transfer the two supernatants to a single tube.

9 Add 180µl of isopropanol and incubate at -20°C for 30 minutes (plus)

10. Spin at room temperature for 10 minutes

11 Remove supernatant and wash pellet with 70% ethanol

12. Spin briefly, remove the supernatant and leave the tube open for a few minutes, without letting the pellet to dry.

13. Dissolve pellets in 50µl dH_2O, add 50µg/ml DNA free RNase and incubate at 37°C over night.

3.2.8 Pulsed field gel electrophoresis (PFGE) using PACs

Electrophoresis was carried out in a Gene Navigator™ System (Pharmacia Biotech) using 0.5 x TAE as running buffer and Type I, low EEO agarose (Sigma). The apparatus has a capacity of 3 litres, circulated by means of an Eheim pump, which is cooled by means of an LKB multi-temperature thermostatic cooler.

For the majority of pulsed field gels the following conditions were used:

agarose gels, 11-32 seconds pulse time, 10-14 hour running time, 8°C running temperature and a field strength of 6V/cm. Gels used for band elution were made of Seaplaque agarose with a Type I frame around them.

Lambda oligomers (Promega) were used as markers for high molecular weight DNA. They give a ladder with steps 48.5kb apart with a size range of 50 to 1000kb. A "Low Range Marker" (Promega) was used to provide low molecular weight markers. The sizes are as follows: 0.13kb, 0.56kb, 2.03kb, 2.32kb, 4.36kb, 6.55kb, 9.42kb, 23.1kb, 48.5kb, 97.0kb, 15.5kb and 194.0kb.

Gels and filters were treated as in section 2.8 ensuring a 4 minute UV exposure prior to blotting to nick the DNA. This aids in the efficient transfer of large fragments. Blotting was carried out for 18-24 hours.

At the outset of the project, I was provided with three Yeast artificial chromosome (YAC) stocks which were thought to be from the UBE1 region: A1S9T (UBE1), 117.2910 and 117.2911, called YA1S9T, Y2910 and Y2911, respectively.

YAC A1S9T was obtained from M. Coleman and YACs 2910 and 2911 were isolated from the Monaco YAC library by G. Black employing the probe 117.29 which was obtained from a radiation hybrid by Alu specific PCR (see Black et al., 1992 and for primer sequence see appendix). Primer sequences to construct gene probes, which were used in this investigation are listed in table 3.2.

Table 3.2: Details of primers used to construct probes

probe	primer 1	primer 2	product size	reference
(AC)n	CAT AGC TCA CTG CAA CCT TCC	AAC AAT GAT CCC TAT CCC AGG	(267bp)	Chapter 4
DXS8237E	CAG CTC CCT GCG AGA TGA CG	CTC GAT AGG CGT TAC AAT GC	650bp (1200bp)	Carrel et al., 1996
UBE1	GAG CGG GGA CTT TGT CTC CT	CTTTGA CCT GAC TGA CGA T	150bp (750bp)	Carrel et al., 1996
PCTK1	CAC GCC AAC ATC GTT ACG CT	TGG GAT TGA CTT GGC TCG G	286bp (800bp)	Carrel et al., 1996

Sizes for probes (excluding the (AC)n repeat) were obtained using cDNA. The bp sizes in brackets were obtained by using the primer sets on genomic DNA, obtained from B. Suarez-Merino (mixture of human cell lines); resulting probes were used in experiments.

Southern blotting of gels and hybridisations was carried out as described in Chapter 2.9-2.13.

3.3 Results

3.3.1 Characterisation of the YACs

After growing and preparing plugs from YAC A1S9T, 2910 and 2911 stocks, they were sized using pulsed-field gel electrophoresis followed by Southern blotting and hybridisation to total human DNA in order to see if any of them were rearranging. The sizes were 460kb, 160kb and 364kb respectively (see figure 3.3.1) with only limited evidence of multiple forms (esp. for A1S9T) suggesting that the YACs were reasonably stable.

The set of sequence tagged site (STS) primers for probe 117.29 (see M&M) was tested to establish the optimum PCR conditions. The PCR product amplified from human DNA was 250bp and the primers were used to check YACs A1S9T, 2910 and 2911 to confirm the presence of the equivalent STS; all three gave positive signals (see figure 3.3.2)

In order to obtain additional STSs to use as internal probes when constructing the maps and, originally, to pull out novel YACs, thereby expanding the YAC contig, YAC end-cloning was carried out.

3.3.2 Cloning, examination and sequencing of the YAC ends

Plasmid rescue for left end cloning of YAC 2910 was attempted To obtain the ideal size for cloning the end into plasmid vectors (under 20kb) and for sequencing it (up to 1kb) YAC DNA was singly digested with three different frequent DNA cutting enzymes, namely *Xho*I, *Sal*I and *Nde*I. The digests were run on a gel, which was Southern blotted and hybridised with the left end probe to detect cognate sequences. The fragments obtained were either too large or too

small to be useful (results not shown). Simulaneously another method was exploited and gave pliable results, so plasmid rescue was abandoned as a method to obtain end clones.

As an alternative approach to obtain left and right ends of the YACs A1S9T, 2910 and 2911, inverse PCR was carried out after digesting the YAC DNA samples with the rare cutting restriction enzyme *Hae*III. Clear results were obtained for YACs 2910 and 2911, the inverse PCR products resulting in a single fragment for the left and right end for each of these YACs. The product sizes of 2910l, 2910r, 2911l and 2911r are approximately 800, 360, 250 and 1000bp, respectively. Results for YAC A1S9T showed two amplified fragments for each primer set used (see figure 3.3.3a and b). Closer examination revealed that both of the fragments amplified by each primer set hybridises either to the right or the left end vector sequence (results not shown), which suggests that the YAC is possibly rearranging or that two distinct forms coexist even though the YAC clones were colony purified before hand. Therefore, to avoid potential confusion from this ambiguity, the main focus of the preliminary characterisation concentrated on the left and right ends from YACs 2910 and 2911.

To sequence the obtained YAC ends, they were cloned into the plasmid pCR™II using the commercial TA cloning kit (Invitrogen). Transformants were streaked out onto LB agar plates containing 50µg/ml ampicillin and 0.02% X-gal and grown over night at 37°C. White colonies, containing the insert, were re-streaked and after overnight incubation, grown in LB containing 50µg/ml ampicillin. After the plasmid DNA was purified, the restriction enzyme *Eco*RI was used to release the insert and its authenticity was checked by running the digest on a 2% agarose gel.

All four ends were first sequenced using the universal forward and reverse primer and subsequently new sets of primers were designed for 2910r, 2910l and 2911r (see M&M and appendix for primer sequences) to obtain the complete insert sequence. The insert sequences are shown below.

2910 left:

TTCCCCCAACTTGATAAAGAGCATCTACAAAAATGCTGCAGTTAGCACCATATTTAATGGTGA

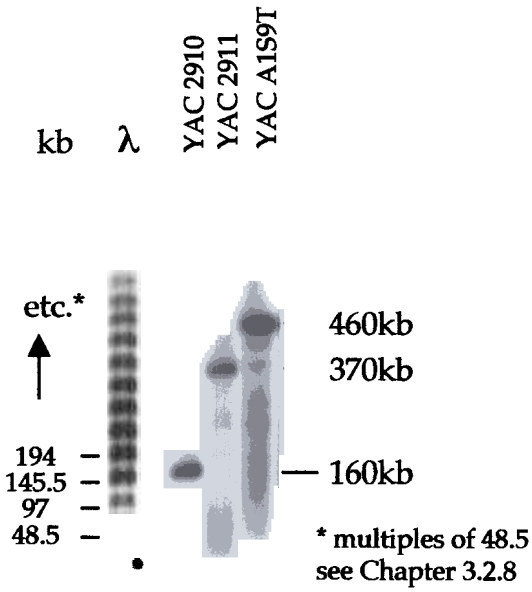


Fig. 3.3.1
Autoradiograph showing undigested YAC A1S9T, 2910 and 2911 DNA hybridised with genomic DNA, thereby showing the YAC insert sizes

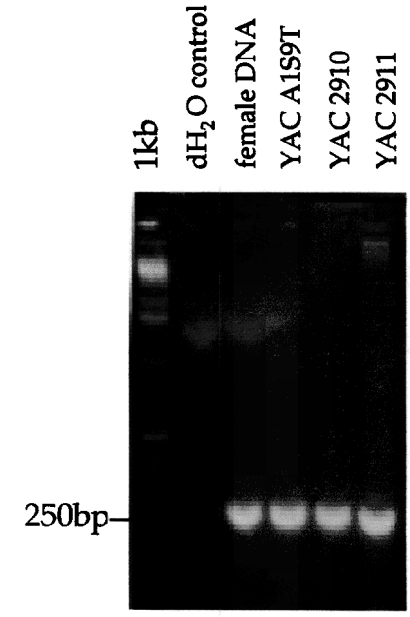


Fig. 3.3.2
Amplification of YAC A1S9T, YAC 2910 and 2911 DNA using the 117.29 primer set

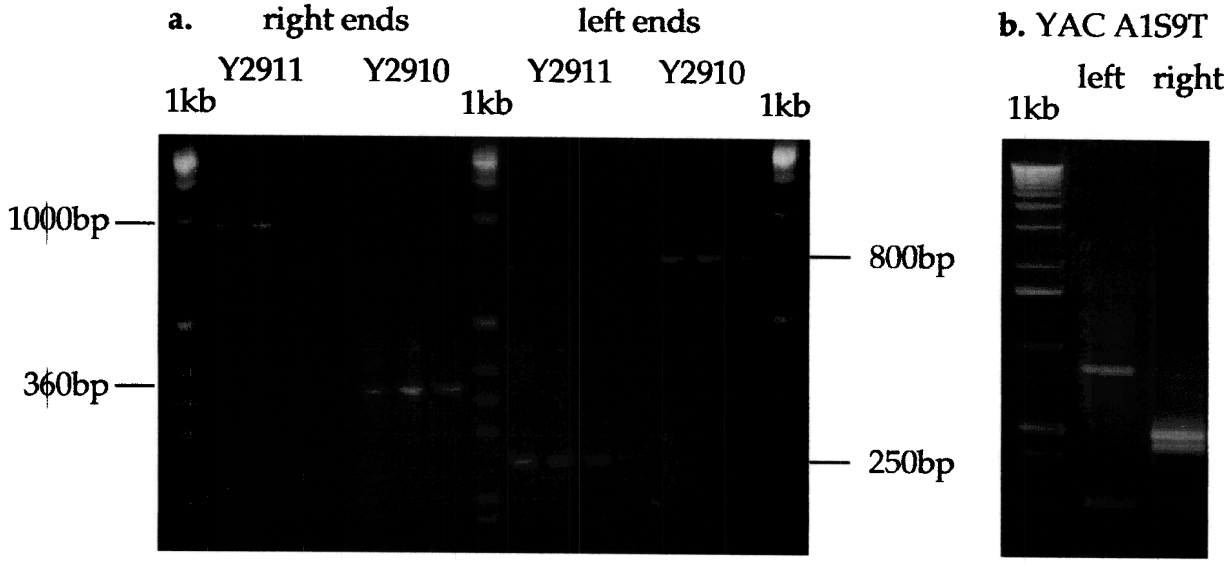


Fig. 3.3.3
Inverse PCR products from YAC 2910 and 2911 (a) and A1S9T (b), the figure 3.3.3b showing 2 products for each reaction. See text for further interpretation.

AAGACTGTATACTCTCCCCCTGAGAATGGGGAAATAAGGTAAGAATGTCTGCTCTTACCAGTCT
CATTCAACGTAATACTGGAAGTCCTAGCTAGAGCAACTAGTTGAGAAAAAAATAAAAGACA
ATGAATTGGAAAGGAAGAATAACATCTTATTTGCAGATGACGTGATTCTCCAGTAGAACATCC
CAAGAATTTACCAAGTAACTCCTAGAATAAGTGAGTCGAGTAAGGTCACAGGATATAAGATC
AACACACAATTTAATATTTCTAGATACTAGCAATGAACATATGGAAACTGAAATTAAAAACA
CAATGCCATGTACAGTCTCTCCCAAAAGAATTGAAATACAGTTATGCATTGCTGAACAACAGG
GATATGTTCTCAGAAATGCATTGTAAGGTGATTTTGTTGTTGTGTGAACATCATAGACTAGTATA
CTTATACAAACGTAGGTGATATAGCCTACTACACACCTAGCCATATTGTAT

2910 right:

CCCTCACCAGAAGCTAAGCAGATGCCAGCACCATGCTGTTAACCTTTCCAGCCACCAGAGTTG
TGAGCCAAATAAATCTCTTTCTTCGTAAGTTACCCAGACCCAGCAGGTATTCTATGATAGCAAT
ACAAAACAGATTCAGCAGCAGAAGGAAATTACACACTTTTAAAGTGCAAAAGGCTGAAAGA
AAAGAACTATTGACCCTGAATTGCATATCCAGTGAAAATGGCCTTCATGTGAAATAAAGACAT
TCTCATATGAAAGAGTGAAATAAAGACATTCTCATATGAAAGAAAAGAAAAC AAGC
TCCGTAATCTTGAGATCGGGCGTTCGACT

2911 left:

ATTCTACCAAATATTCAAGGATGAAAAAAAGGTCAATACTACACAAACTCTTACAGAGAATA
GAAGAGACTACTTCTCAATTCACTTTATGAGG

2911 right:

AGCAACTTAACACCAACACCCACTTTGAAAACTGAAAAAAGTTATGAATATTAATTAGTTTC
CCGGTAGAGACTGTGTACTTCATTATGAATAAATGCAACAGCATTAACTGTGGCAATAGCAGA
TCAAAACAAATGGGTAAGCAGCTGTTATATTATGTGTATGTTTATTTTAATTTCCAGGAATCTCT
GGGTGCCCCTTTTTCCTCTGGGGACTTCTAGCATTGTTGGGCTTGGCTTTGGTTATATCACTGATCT
TCAATATTTCCCACTATGTGAAAAGCAACGACAAGGTAAGACATTTTGACAAATTTCACATGG
TACCTGCTTGAATGGAGACTATGGAGTCACTGGAAAACCTATGGAAGCAGACAATTTTATGAA
ATGTCATTCAACAGGATATTTTTNNNNNNNNNNNNNNNNNNNGTTCTCTTATTTAAGA
ATGCAATCATATTAGAAATGTGAAATAGTTATCCAATTTAGGAACTCTTCTTTAGGGTGTAGAG
TGGGAAGGGATTATCCTGGTAAATACGTTTTTCCTCCTCCAGTTGAAAAACCTATGAAATTGTA
GTCTATGCTGGGGCAGCCATCAAGGTTCTCTCTGATACATAATAAACAAAGGGAAGTCTAGCT
CTTTAAGATTCCATTCAAGCCAACTTAATCTTCTGAGGATCAGACAGGCATTTCTCACATGCAA
CACAAAAGAGCGGGGTCAGAAGACGACTGGAAAAGGAAGAGATTGAACCATGGGCTTGCTGT
CAAACACTCCCAGAGTCTAGGAGAGGTCAAGAGTATCTGCCACTGTCAGCTTTGCTTTTTCAAC

ATGCTGGGCATAAGTTTTAACCATGTGGCTTTCAATCTGATATGCCACAACTGAAATGC

Database analysis of the sequences revealed that the insert of the YAC nd 29 has 97% homology to an exon of an unknown gene expressed in fœtal lung tissue (**bold** sequence; bases in *italic* identify exon 3' and 5' intron boundaries Primers amplifying the exon were constructed from the underlined sequence). Three mismatches were found in the 118 base sequence, two of which are unknown bases in the sequence determined Further investigation employing a monochromosomal hybrid panel showed that this exon is located or chromosome 3 (see figure 3.3.4), indicating the YAC to be chimæric. A more recent database search revealed that the sequence matches part of a T-cell antigen receptor-interacting molecule or TRIM on chromosome 3 at 3q13. Two reports are available: Bruyns et al., 1998 and Hubener et al., 2000.

3.3.3 Generation and examination of STSs

Primers amplifying the inserts were constructed (using the sequences above) and used to generate STS probes: S2910l, S2910r, S2911l and S2911r (see table 3.3.1). This was done by amplifying genomic DNA, the products of which were cut out from the gel and cleaned as described in Chapter 2.14.4. The 2911r primer set was constructed to amplify the exon sequence found in 2911r (to be used as a control) The other three primer sets were tested on THYBX DNA and all of them were X specific (see figures 3.3.5a-c).

Table 3.3.1: PCR primers designed from YAC 2910 and 2911 end sequences to generate STSs

primers	sequence 5'-3'	specific conditions	product size
2910lf 2910lr	AAT TCC CCC CAA CTT GAT AAA GAG CAC ACA ACA ACA AAA TCA CCT	T_a = 53°C, 37 cycles	250bp
2910rf 2910rr	TTA ACC TTT CCA GCC ACC AG TTC TTT TCT TTC AGC CTT TTG C	T_a = 55°C, 37 cycles	156bp
2911lf 2911lr	ATT CTA CCA AAT ATT CAA GG TCA TAA AGT GAA TTC AGA AG	T_a = 47°C, 35 cycles	90bp
2911rf 2911rr	TCC CGG TAG AGA CTG TGT ACT TCA CAA AGC TGA CAG TGG CAG ATA CTC	T_a = 54°C, 34 cycles	196bp

The denaturing and extension temperature for all the primer sets are 94°C and

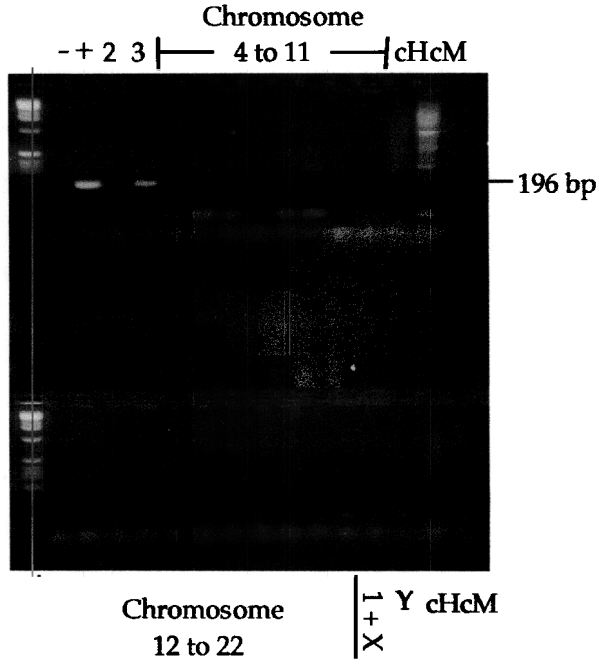


Fig. 3.3.4
PCR using 2911r primer set on the monochromosomal hybrid panel, localising the exon to chromosome 3

- : dH_2O control
+ : female DNA
The numbers indicate the chromosome number
cH: control Hamster
cM: control Mouse
X: X chromosome
Y: Y chromosome

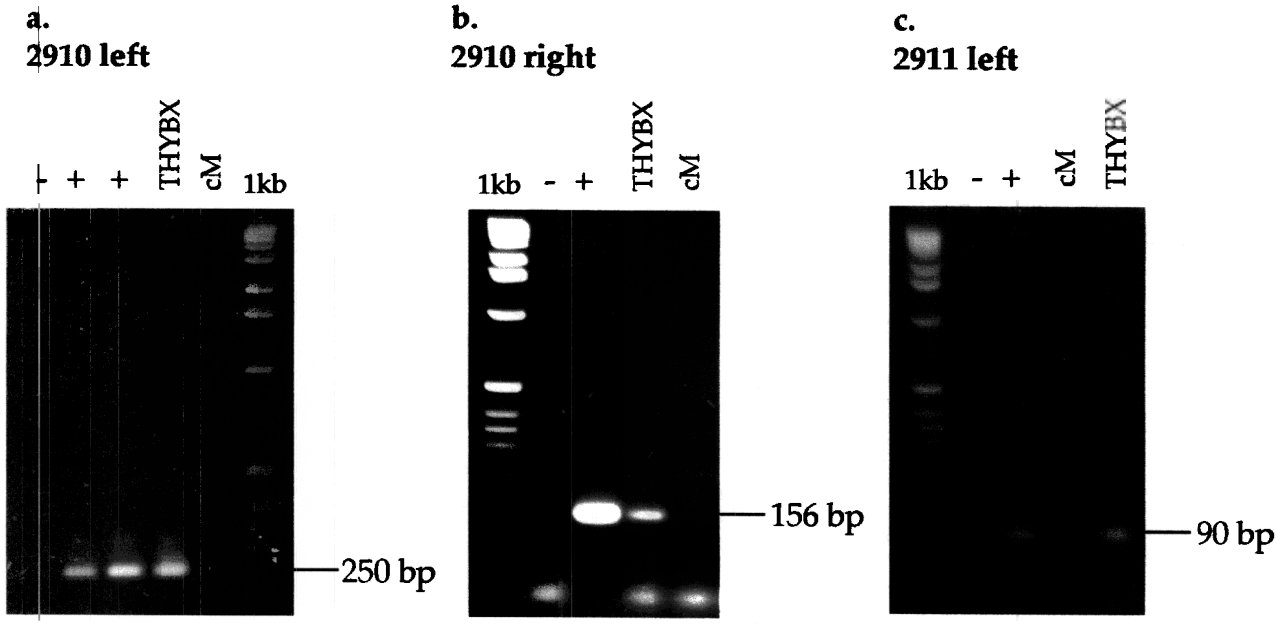


Fig. 3.3.5
Testing primer sets 2910l (a), 2910r (b) and 2911l for X-chromosome specificity
Track labelling as for figure 3.3.4

74°C respectively. The time for each step is 0.5 minute for primer sets 2910l and 2910r and 1 minute for both 2911 primer sets.

new STS markers were then employed to analyse YACs A1S9T, 2910 and 2911 in order to establish the orientation, potential irregularities and overlap of the YACs. Firstly, the primers of the STSs were used to amplify the three different YAC DNAs by PCR, but as results were not conclusive (results not shown), the actual STS markers were radioactively labelled and used as probes in hybridisation experiments on YACs and YAC digests and partial digests (see below).

3.3.4 Restriction mapping of YACs

For convenience, fragment sizes have been rounded to the nearest kb. For example, the illustration of 37.2kb is drawn as 37kb, 6.7kb as 7kb and 30.5kb as 31kb.

In hybridisation data for YACs, the enzyme concentration for a particular digest is decreasing from left to right (i.e. in the extreme left lane in the diagrams, the DNA is completely digested and in the extreme right it is hardly digested), for PACs the reverse is true (complete digestion on extreme right). If a hybridisation with a gene probe resulted in a fragment size which could fit in to more than one place on the restriction map all the possibilities are shown In general, only one fits the criteria.

3.3.5 Hybridisations

Hybridisation of partially digested YACs using the left and right vector as probes

In order to map the restriction sites of a particular enzyme to the YACs and thereby show their orientation to each other, YACs A1S9T, 2910 and 2911 were partially digested (see M&M) using the restriction enzymes *Not*I, *Bssh*II, *Eag*I and *Sst*II/*Sac*II. The digests were run on a pulsed-field gel, which was then Southern blotted. The resulting filters were hybridised at least twice with left and right end probes. At least two filters were made of the same digest; the first being hybridised with the left vector probe first, then stripped and hybridised with the right vector probe and the second hybridised with the right vector probe first,

then stripped and hybridised using the left vector probe. There were no *Not*I restriction sites detected by this approach in any of the YACs. An example of results obtained is illustrated using YAC 2911 and shown in figure 3.3.6. Results for YACs 2190 and A1S9T are given in figures 3.3.10 and 3.3.11, respectively.

Hybridisation of partially digested YACs using the end clone STSs

To investigate further the genomic organisation of the YACs, filters containing partial digested DNA of the three different YACs were hybridised with STS markers from the end clones and other markers for the region Filters containing partial digested YAC A1S9T DNA, YAC 2910 DNA and YAC 2911 DNA, were employed in different hybridisation reactions using S2910l, S2910r, S2911l and S2911r. Note: some of these experiments were in progress before the apparent chimærism of YAC A1S9T became obvious.

The results show that all STS markers hybridise to filters containing the YAC DNA they were derived from. Further, fragments obtained using, for example, the STS 2910r corresponded, as expected, to sites present on the right hand side of YAC 2910. Similar patterns of results were observed for the other STSs. These together with results for "between-YAC" hybridisations are shown in table 3.3

Table 3.3.2: Endclone STS hybridisations to partial YAC digests

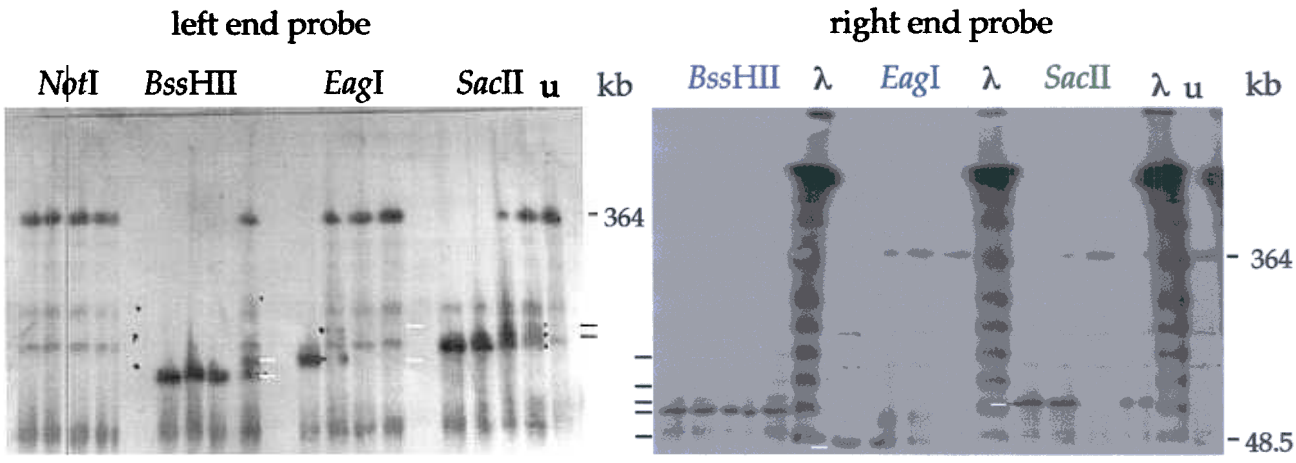
	YAC A1S9T		YAC 2910			YAC 2911		
STS	*Bss*HII (kb)	*Sst*I (kb)	*Bss*HII (kb)	*Eag*I (kb)	*Sst*I (kb)	*Bss*HII (kb)	*Eag*I (kb)	*Sst*I (kb)
S2910l	no	no	~100	49	~90	no	no	no
S2910r	no	no	49	49	~90	no	no	90kb
S2911l	<49	49, 97	no	no	no	~140	~150	~200
S2911r	no	no	no	no	no	97	<49	97

No *Eag*I site was observed in YAC A1S9T with any of the STSs (data not shown).

Hybridisation of partially digested YACs using gene probes

The position of gene probes has been based on band sizes obtained for complete digests. The possible location is indicated on the figures by a rectangle which

Fig. 3.3.6
YAC 2911 (364kb) partial digest using left and right end vector probes



λ marker did not hybridise to filter, .: after stripping and re-hybridisation of the filter the autorads were imposed onto each other and sizes determined. Faint bands present in all lanes are ignored.

u: undigested YAC

bands identified in table are demonstrated by –

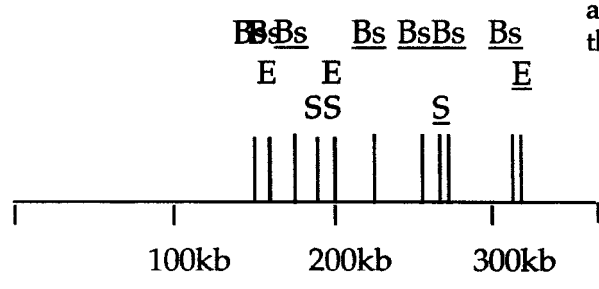
YAC 2911, left end probe			
enzyme	*Bss*HII	*Eag*I	*Sac*II
size (kb)	150	160	190
	160	200	200

YAC 2911, right end probe			
enzyme	*Bss*HII	*Eag*I	*Sac*II
size (kb)	50	47	100
	95		
	110		
	140		
	190		

restriction map

Bs = *Bss*HII
E = *Eag*I
S = *Sac*II

the underlined restriction sites are derived by mapping from the right end



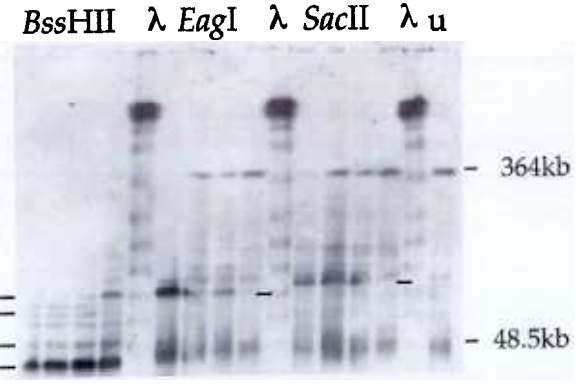
includes as many fragments as possible and feasible.

Filters containing YAC DNA treated as before were used in hybridisations using 117.29 (G. Black), DXS8237E, UBE1 and PCTK1 as probes. The marker probe 117.29 was retrieved by growing the recombinant plasmid clone, purifying the DNA and cutting out the 1kb insert by double digestion, using the res ic *Hind*III and *Bam*HI. The latter three probes were made using the relevant primer sets (see M&M) on cDNA as a template. A typical example of results obtained is illustrated using YAC 2911 and shown in figures 3.3.7 - 3.3.9. Results for YACs 2190 are presented in figure 3.3.10. Although the map for YAC A1S9T is dubious it is compared with those of the other two YACs in an attempt to detect where rearrangements have occurred. The almost symmetrical pattern obtained with left and right end probes is probably more than coincidental and suggests complex rearrangement in which segments embracing pBR322 left and right end have been exchanged. For completeness, the results for YAC A1S9T are given in figure 3.3.11 and combined restriction maps are shown in figures 3.3.12 - 3.3.15.

Further evidence for the instability of YAC A1S9T came later on in the study, when it was regrown from the original stock and showed a size decrease of almost 200kb (results not shown). In summary, YAC A1S9T is most likely rearranging, YAC 2911 is probably chimæric and the only stable YAC, the smallest of the three, is YAC 2910.

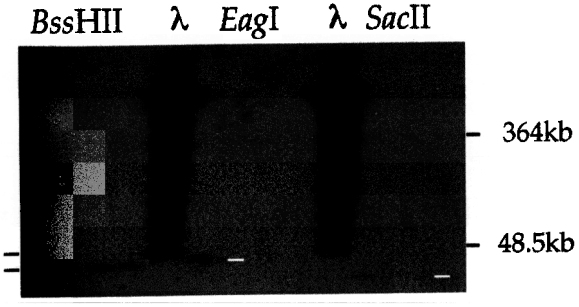
Overall, the studies on these YACs suggested that there are a multiplicity of rare enzymes cutting sites created by e.g. *Bss*HII, *Eag*I, *Sac*II pointing to a very high GC content, which may be responsible for the instability and rearrangement observed. These have impeded the simple production of an overall coherent physical map for the region. Hence this preliminary physical map had to be confirmed by other means. As YAC 2910 was the only stable one, the STS primer sequences of each end of this YAC were sent to and used by the Sanger Centre (Hinxton) to obtain PACs in the region. Altogether 21 PACs were provided and to select the most relevant ones for mapping, PCR reactions using primer sets for STSs 2910l, 2910r, 2911l and for *DXS8237E*, *UBE1*, *PCTK1* and (AC)n (see M&M) were carried out. Although the results were not uniform the following five PACs were chosen on the basis of hybridisation studies for further investigation

Fig. 3.3.7
YAC 2911 partial digest using probes for 117.29 and DXS8238E



YAC 2911, 117.29 probe			
enzyme	*Bss*HII	*Eag*I	*Sac*II
size (kb)	<49	146	163
	49		
	113		
	146		

The small Bs fragment is the same size as the Bs fragment in the hybridisations using UBE1 and DXS8237E as a probe.



YAC 2911, DXS8237E probe			
enzyme	*Bss*HII	*Eag*I	*Sac*II
size (kb)	small	47	small
	50		

The small Bs fragment is the same size as the one in the UBE1 hybridisation and bigger than the one using the PCTK1 probe.

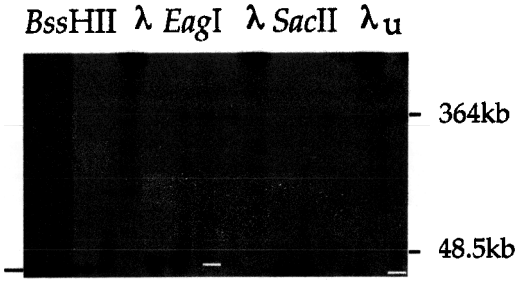
The small S fragment is bigger than the one in the PCTK1, but smaller than the one in the UBE1 hybridisation. It is the smallest considering the DXS8237E hybridisation.

u: undigested YAC

bands identified in table are demonstrated by –

Faint bands present in all lanes are ignored.

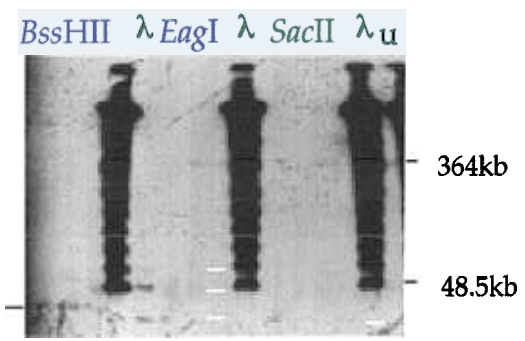
Fig. 3.3.8
YAC 2911 partial digest using probes for UBE1 and PCTK1



YAC 2911, UBE1 probe			
enzyme	*Bss*HII	*Eag*I	*Sac*II
size (kb)	small	47	small

The small Bs fragment is the same size as the one in the DXS8237E hybridisation and bigger than the one using the PCTK1 probe.

The small S fragment is the biggest out of all hybridisations, but smaller than the Bs fragment within the UBE1 hybridisation.



YAC 2911, PCTK1 probe			
enzyme	*Bss*HII	*Eag*I	*Sac*II
size (kb)	small	small	small
		49	
		97	

The small Bs fragment is the smallest out of all hybridisations. However, it is the biggest within the PCTK1 hybridisation.

The small E fragment is smaller than the Bs fragment and the same size as the S fragment.

The small S fragment is the smallest out of all hybridisations and smaller than the Bs fragment and the same size as the E fragment using the PCTK1 probe.

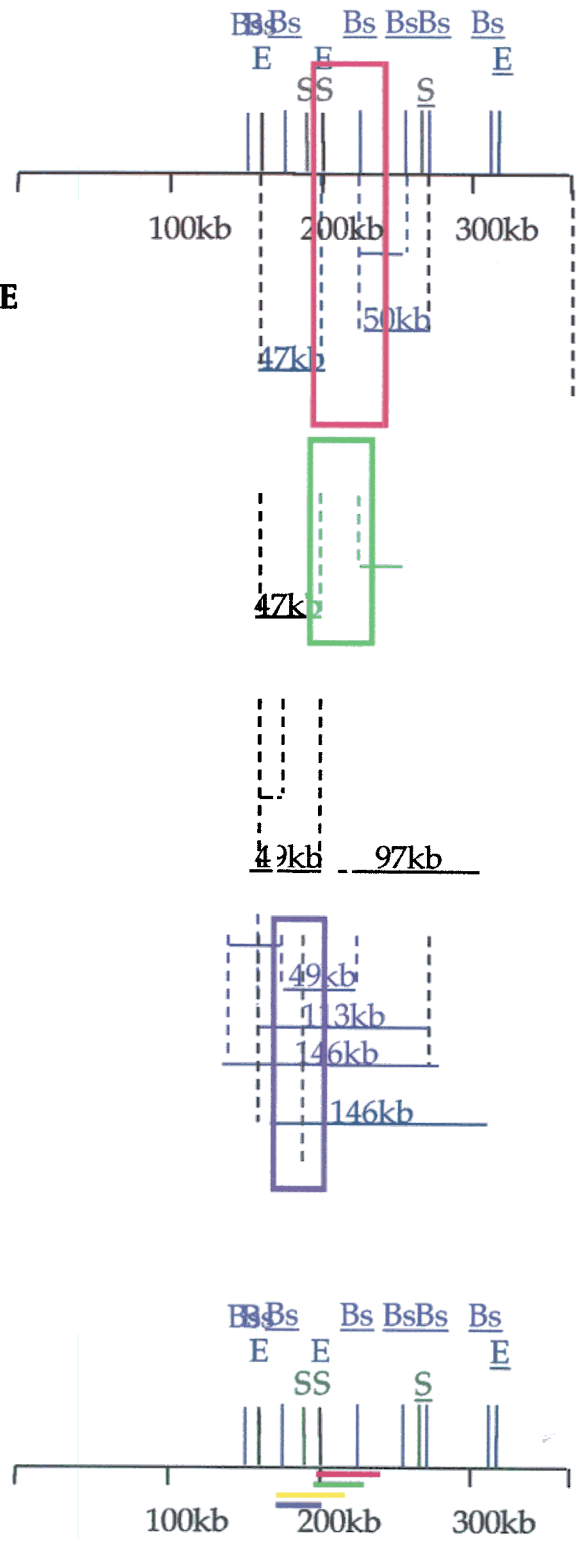
u: undigested YAC

bands identified in table are demonstrated by –

Fig. 3.3.9: Putative restriction map of YAC 2911 (364kb)

Because the order of the gene loci tested is known (see Carrel et al., 1996), the Bs fragments of unknown size could be assigned to the hybridisations for the individual probes.

Bs = *Bss*HII
E = *Eag*I
S = *Sac*II



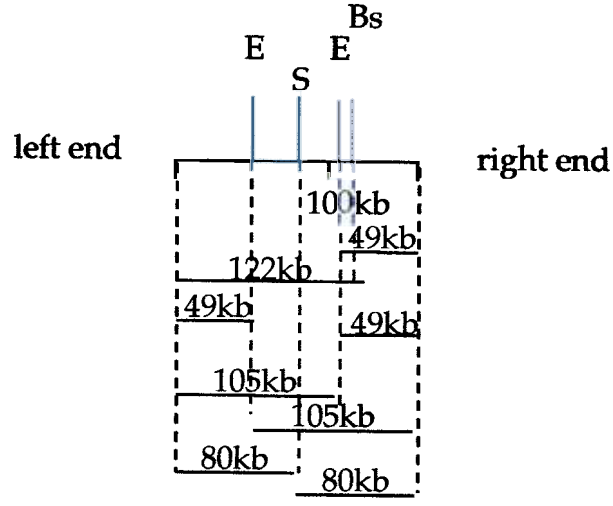
the underlined restriction sites are derived by mapping from the right end

Fig. 3.3.10 Fragment sizes obtained YAC 2910 (160kb)

YAC 2910, left end probe			
enzyme	*Bss*HII	*Eag*I	*Sac*II
size (kb)	122	49	80
		105	

YAC 2910, right end probe			
enzyme	*Bss*HII	*Eag*I	*Sac*II
size (kb)	49	49	80
		105	

Bs = *Bss*HII
E = *Eag*I
S = *Sac*II



YAC 2910, 117.29 probe			
enzyme	*Bss*HII	*Eag*I	*Sac*II
size (kb)	49	49	80
	122	105	

Since the hybridisation with probe 117.29 for *Bss*HII shows two bands: a 49 and 122kb band, it has to overlap those two bands, which are otherwise detected by left and right end probes, respectively. See diagram at right.

Slight inaccuracies might be due to the nature of restriction mapping.

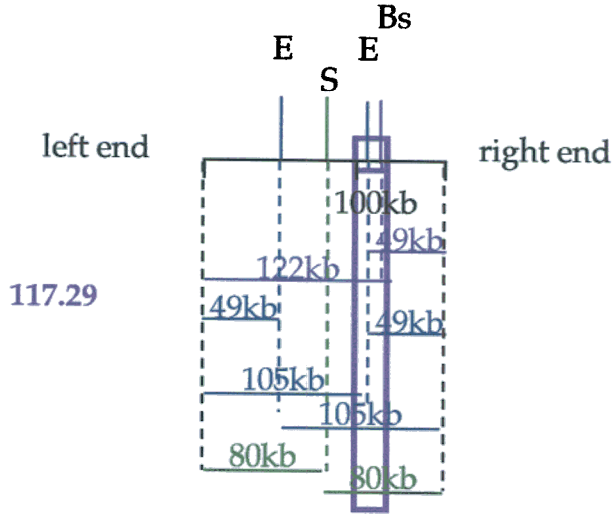


Fig. 3.3.11: Fragment sizes obtained for YAC A1S9T (460kb)

YAC A1S9T, left end probe			
enzyme	*Bss*HII	*Eag*I	*Sac*II
size (kb)	55	—	50
	122		90
	145		100?
	170		140?
	210		

YAC A1S9T, right end probe			
enzyme	*Bss*HII	*Eag*I	*Sac*II
size (kb)	55	—	50
	122		90
	155		100?
	180		140?
	250		

YAC A1S9T, DXS8237E probe			
enzyme	*Bss*HII	*Eag*I	*Sac*II
size (kb)	47	—	*
			47
			140?

* A SacII fragment much smaller than in the hybridisations using UBE1 or PCTK1 as a probe was also present.

For the hybridisation using **UBE1** as a probe there is one BsshII fragment of 47kb and one SacII fragment which is slightly smaller than the 47kb BsshII fragment.

For the hybridisation using **PCTK1** as a probe there is one BsshII fragment of 47kb and one SacII fragment, which is much smaller than the one detected in the hybridisation with UBE1.

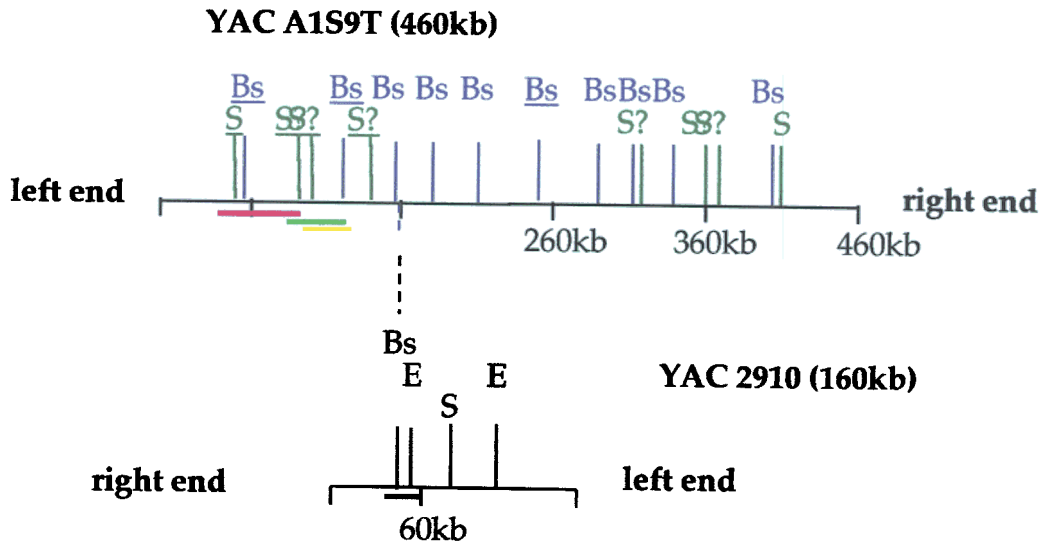
YAC A1S9T, 117.29 probe			
enzyme	*Bss*HII	*Eag*I	*Sac*II
size (kb)	47	—	47?
	73		
	122		

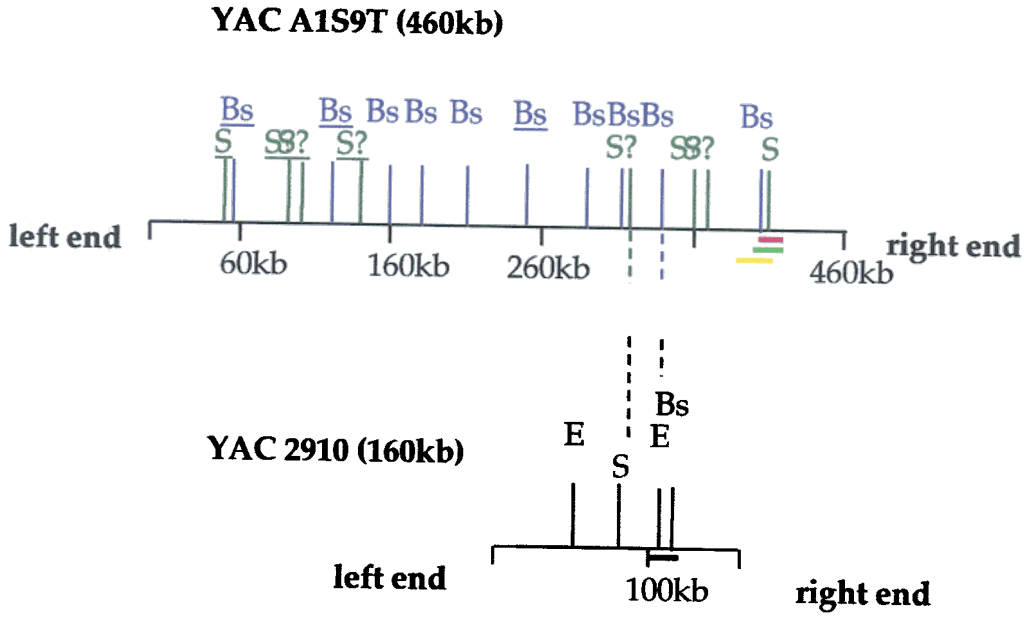
These data were used to compare this YAC to the other two in attempt to identify potential areas of rearrangements (see figures 3.3.12-3.3.14).

? : site of lower level of confidence

Fig. 3.3.12: Putative restriction maps of YAC A1S9T and YAC 2910

YAC A1S9T and YAC 2910 overlap in one of the following ways:





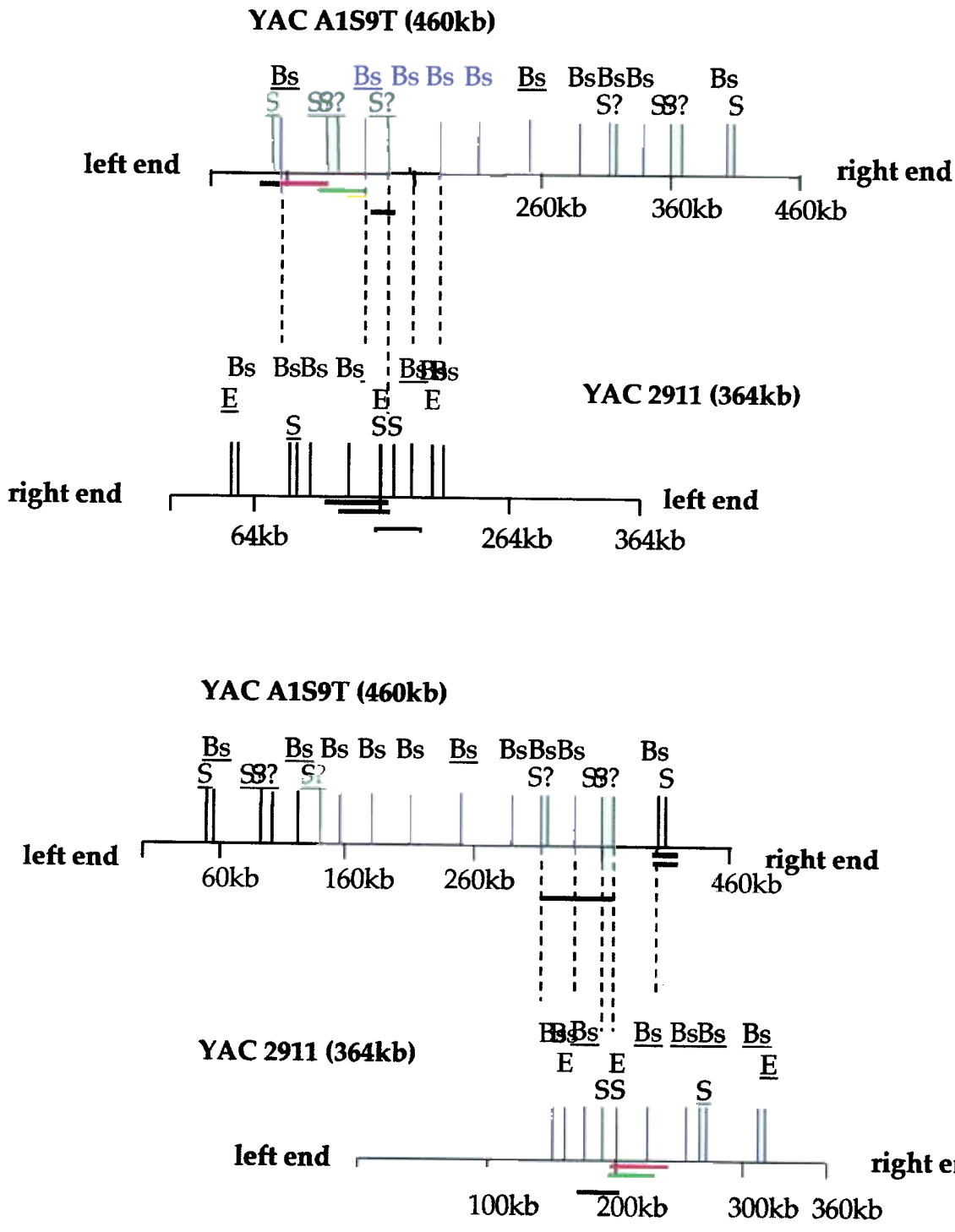
Bs = *Bss*HII
E = *Eag*I
S = *Sac*II

117.29
DXS8237E
UBE1

the underlined restriction sites are derived by mapping from the left end

Fig. 3.3.13: Putative restriction maps of YAC A1S9T and YAC 2911

Note: YAC A1S9T and YAC 2911 overlap in one of the following ways:

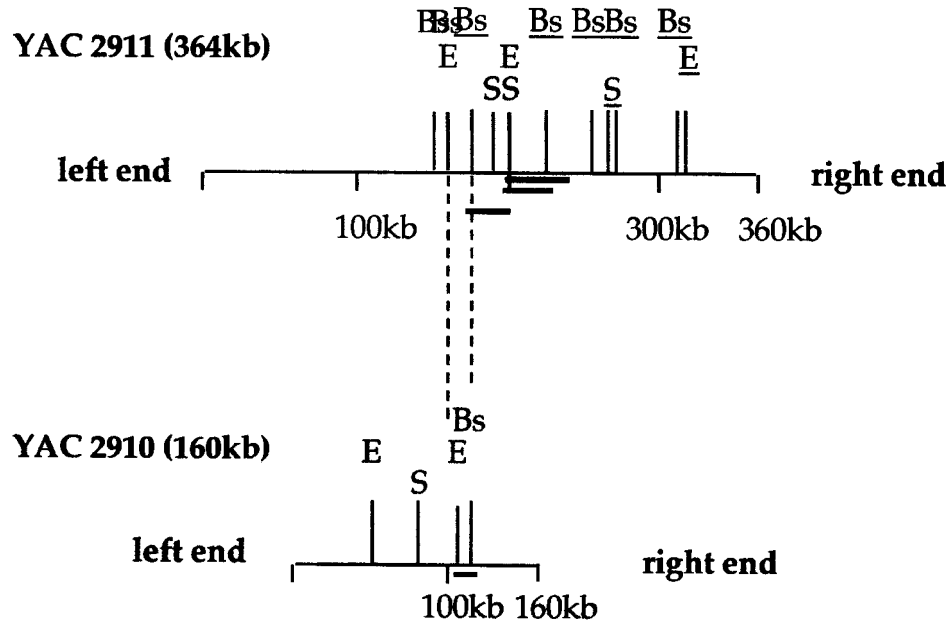


Bs = *Bss*HII
E = *Eag*I
S = *Sac*II

117.29
DXS8237E
UBE1

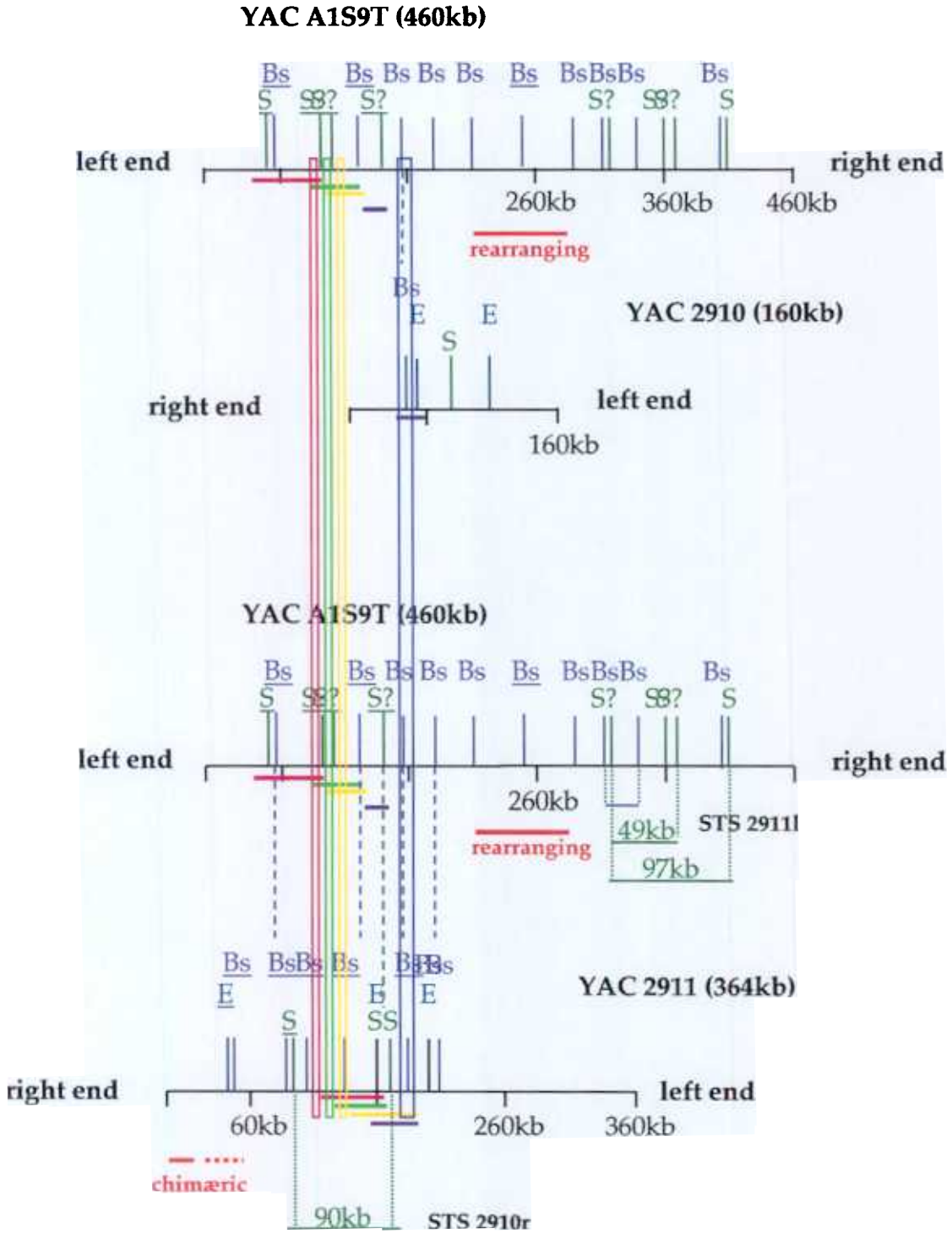
For YA1S9T the underlined restriction sites are derived by mapping from the left end and for Y2911 from the right end.

Fig. 3.3.14: Putative restriction maps of YAC 2910 and YAC 2911



Bs = *Bss*HII	**117.29**	the underlined restriction sites are derived by mapping from the right end
E = *Eag*I	**DXS8237E**	
S = *Sac*II	**UBE1**	

Fig.3.3.15: Summary of the possible correlations between restriction maps of YACs A1S9T, 2910 and 2911



← centromere telomere →

Slight mapping inaccuracies may be due to the difficulties in the assignment of sizes to fragments observed on autorads.

Bs = *Bss*HII
E = *Eag*I
S = *Sac*II

117.29
DXS8237E
UBE1

For YA1S9T the underlined restriction sites are derived by mapping from the left end and for Y2911 from the right end.

dJ0333P10, dJ0524O12, dJ0524O13, dJ0961F01 and dJ1031J13. In this study they are called PACs 1, 2, 3, 4 and 5, respectively. Results of the PCR reactions and subsequent hybridisations using the various probes are summarised in table 3.3.3

Table 3.3.3: Initial PCR reactions using the PACs selected with the endclones of YAC 2910 (Positive results confirmed by other means are shown in different colours)

√: ~~primers amplified PAC DNA~~

√: primers amplified PAC DNA

-: primers did not amplify PAC DNA

√: confirmed by hybridisation

~~√√:~~ ~~strong amplification~~

√√: very strong amplification

x: PACs grown

results confirmed by information from the Sanger Center

Confirmation of results was only obtained for the five grown PACs

Primers ->		DXS	UBE1	PCTK1	(AC)n	YACs			PACs
PACs		8237E				2910 L	2910 R	2911 L	grown
1	dA 0036 H21	-	-	-	-	√	-	-	
2	dA 0036 I21	-	-	-	-	√	-	-	
3	dA 0036 J22	-	-	-	-	√	-	-	
4	dA 0049 E14	-	-	-	-	√	-	-	
5	dA 0208 N01	-	-	-	-	√	-	-	
6	dJ 0333 P10	√√√	√√√	-	√	√	√√	-	1
7	dJ 0524 O12	√	√√	-	√√	√√	-	-	2
8	dJ 0524 O13	-	-	√	√√	√√	-	-	3
9	dJ 0563 A06	-	-	-	-	√	-	-	
10	dJ 0563 C04	-	-	-	-	√	-	-	
11	dJ 0618 M16	-	-	-	-	√	-	-	
12	dJ 0704 H13	-	-	-	-	√	-	-	
13	dJ 7010 N10	-	-	-	-	√	-	-	
14	dJ 0806 C14	-	-	-	-	√	-	-	
15	dJ 0843 O12	-	-	-	-	√	-	-	
16	dJ 0935 H12	-	-	-	-	√	√	-	
17	dJ 0961 F01	√√√	√√√	√√√	√	√	√√	-	4
18	dJ 1021 P02	-	√	√	-	√	√	-	
19	dJ 1031 J13	-	√	√	√	√	√√	-	5
20	dJ 1037 B01	-	√	√	-	√	-	-	
21	dJ 1196 E11	-	√	√	√	-	-	-	

3.3.6 Restriction mapping of PACs

PACs 1-5 were grown and their DNA extracted as described in the M&M. The human insert was cut out by complete restriction enzyme digest using *Not*I and subsequently they were partially digested using the enzymes *Bam*HI, *Bss*HII, *Eag*I, *Eco*RI, *Hind*III, *Nru*I, *Sac*II, *Sfi*I, *Sma*I and *Xho*I. The digests were run on a pulsed-field gel, which was then Southern blotted. The resulting filters were hybridised using the end sequences of the PAC vector arms, T7 and SP6 (for sequences see appendix) after end labeling them with ^{32}P (see Chapter 2.10.2), and secondly with the following probes: (AC)n, DXS8237E, UBE1 and PCTK1, which were multiprime labelled (see Chapter 2.10.1). PACs 2, 3, 5 hybridised only to the (AC)n probe, whereas PAC1 hybridised to DXS8237E and UBE1 probes and PAC4 hybridised to all three gene probes, as well as to the (AC)n probe. Results showing the sizes obtained for each PAC using the T7 and SP6 hybridisations shown in figure 3.3.16. The EthBr stained PFGE gels confirm the number and sizes of the *Not*1 fragments (results not shown). Some restriction fragments obtained using the gene probes can not be mapped, since not all enzymes have provided usable data for the T7 and SP6 hybridisations. As with the YACs, if a hybridisation with a gene probe resulted in a fragment size, which could fit in to more than one place on the restriction map, all the possibilities are shown. In general, only one fits the criteria. Only hybridisation fragments obtained f complete digests of PAC DNA were used for mapping the gene probes. The map positions of the probes are indicated by a rectangle and all possibilities consistent with the mapping data are presented for completeness . Similarly, with gene markers: all possible orientations on the map are shown. Combined data allowed some of these initial maps to be refined (for example see figures 3.3.22-26). Examples of hybridisations using gene probes are shown in figure 3.3.17 other results are shown in tables 3.3.4 -3.3.8 and in figures 3.3.18 - 3.3.27. maps produced as a result of the investigation, please see 3.4 Interpretation and Discussion

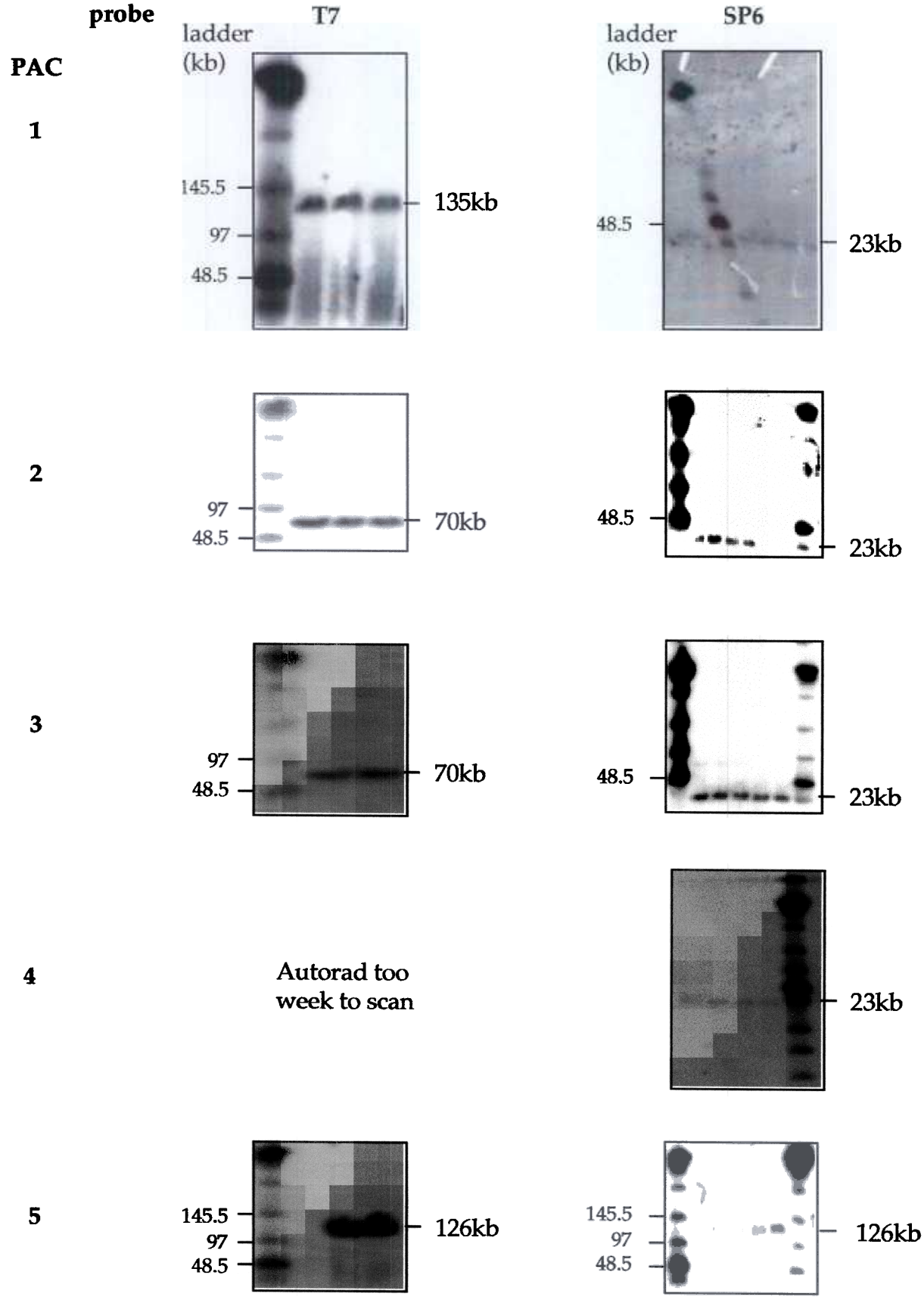
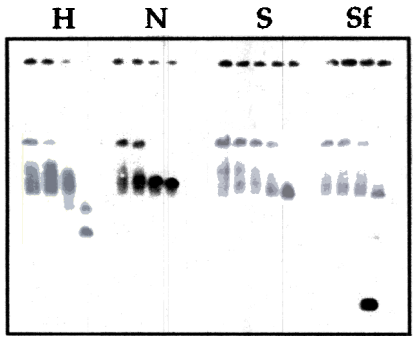


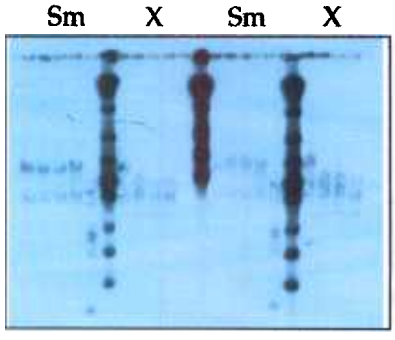
Fig. 3.3.16
Hybridisation of PACs after *Not*I digestion using the T7 and SP6 probes

PAC1 UBE1

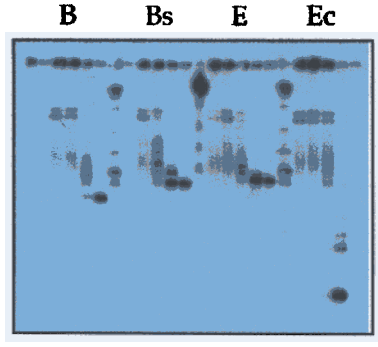


Note: band sizes are not indicated because the size reduction of the scanned autographs makes . For size data please see corresponding tables.

PAC2&3 (AC)n



PAC4 (AC)n



B = *Bam*HI

Bs = *Bss*HII

E = *Eag*I

Ec = *Eco*RI

H = *Hind*III

N = *Nru*I

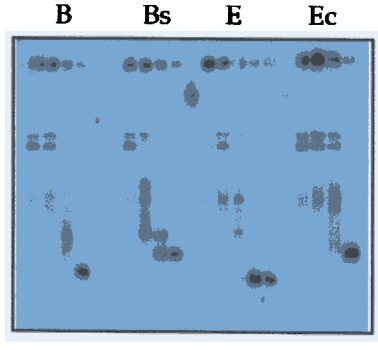
S = *Sac*II

Sf = *Sfi*I

Sm = *Sma*I

X = *Xho*I

PAC4 DSXS8237E



PAC4 PCTK1

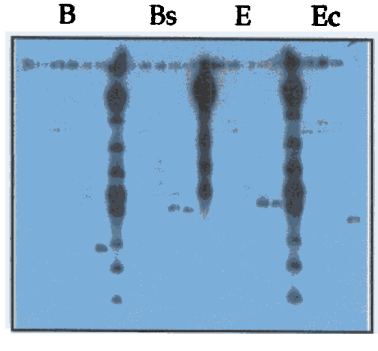


Fig. 3.3.17
Examples of partial digested PAC DNA hybridised using various gene probes

Table 3.3.4: PAC 1 partial digest using probes T7, SP6, DXS8237E, UBE1 and (AC)n

probe: T7 ***Not* I digest: 135kb**

enzyme	*Bam* HI	*Bss*HII	*Eag*I	*Eco*R I	*Hin*dIII	*Nru*I	*Sac*II	*Sfi*I	*Sma*I	*Xho*I
size (kb)	**6**	**6**	**23**	**13**	**9**	7	**5**	23		
	7	23		21	20	49	**6**			
	21	66		37	66		23			
	23				87					
	37									
	62									

probe:SP6 ***Not* I digest: 23kb**

enzyme	*Bam* HI	*Bss*HII	*Eag*I	*Eco*R I	*Hin*dIII	*Nru*I	*Sac*II	*Sfi*I	*Sma*I	*Xho*I
size (kb)	**7**	**16**								

probe: DXS8237E ***Not* I digest: 135kb**

enzyme	*Bam* HI	*Bss*HII	*Eag*I	*Eco*R I	*Hin*dIII	*Nru*I	*Sac*II	*Sfi*I	*Sma*I	*Xho*I
size (kb)	**17**	**23**	**23**	?	8	no	13	no	**9**	**49?**
		42			23	site		site	36	
					49	?		?		

probe: UBE1 ***Not* I digest: 135kb**

enzyme	*Bam* HI	*Bss*HII	*Eag*I	*Eco*R I	*Hin*dIII	*Nru*I	*Sac*II	*Sfi*I	*Sma*I	*Xho*I
size (kb)	**8**	**22**	**22**	**22**	**8**	**51**	**24**	**4**	**23**	**19**
	13	67		30?	**21**		73?	**8**	**31**	**23**
	22				**33**			**24**		**42**
	30				66					
	67				84					
	106				124					

Bands apparent in complete digests are in **bold**.
The probe PCTK1 did not hybridise to filters containing this PAC.
At least one fragment was obtained using the (AC)n probe.
It is an *Xho*I fragment of about 11kb.
? indicates a band of lower reliability for mapping purposes.

Fig. 3.3.18: PAC 1 (135 + 23kb)- DXS8237E -probe

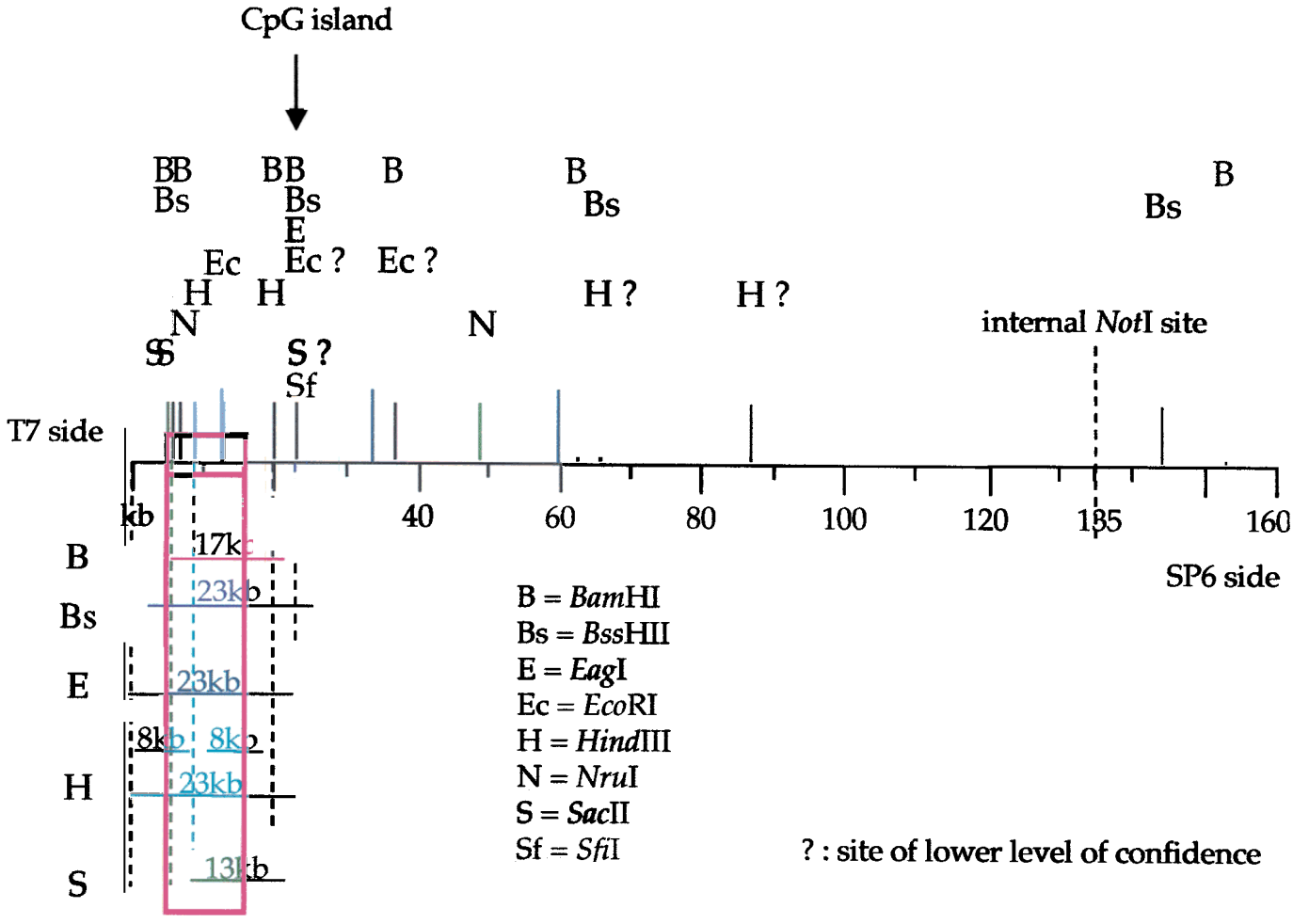


Fig. 3.3.19: PAC 1 (135 + 23kb) - UBE1 -probe

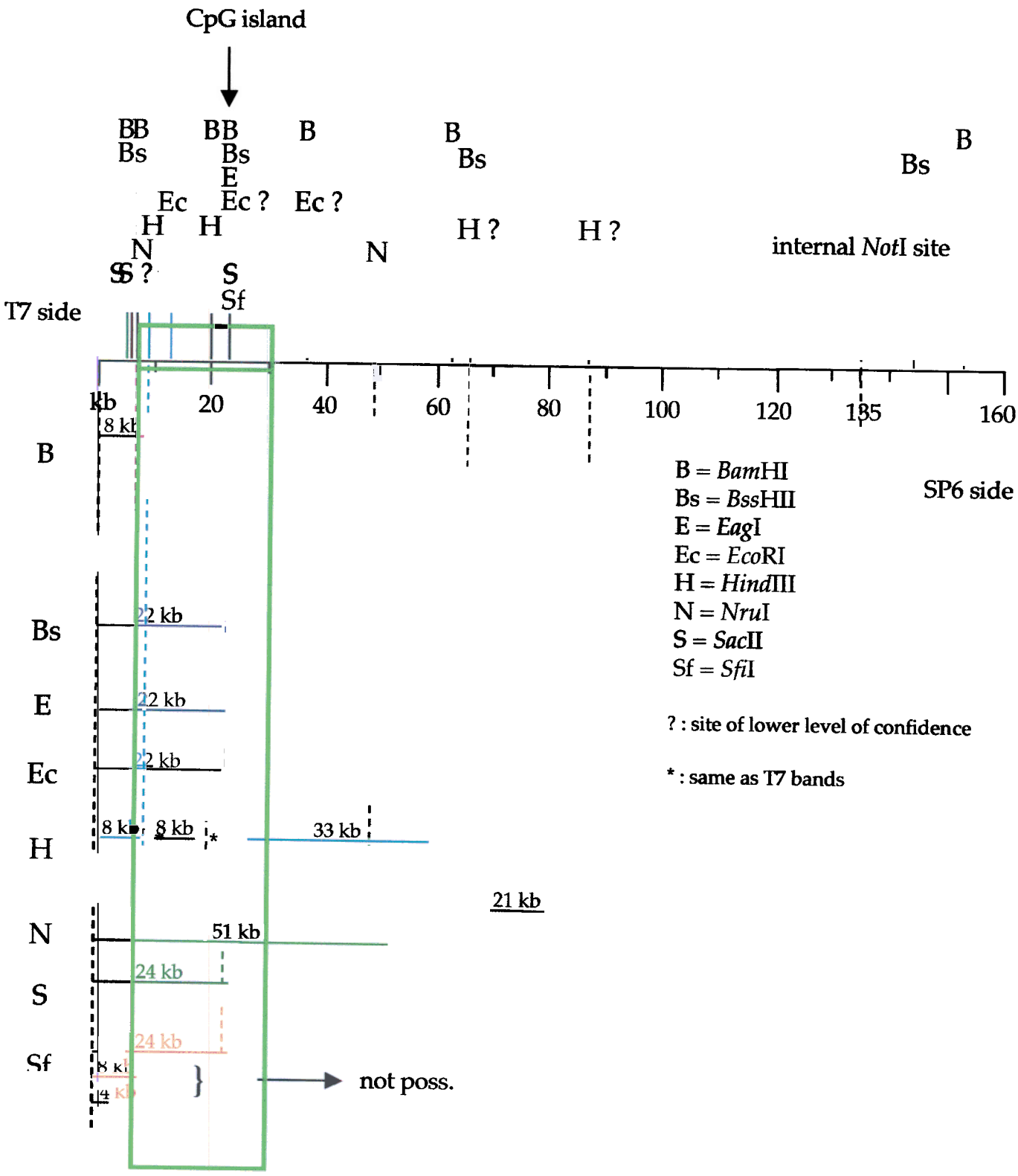


Table 3.3.5: PAC 2 partial digest using probes T7, SP6 and (AC)n

probe: T7					*Not* I only digest: 70kb					
enzyme	*Bam* HI	*Bss*HII	*Eag*I	*Eco*R I	*Hin*dIII	*Nru*I	*Sac*II	*Sfi*I	*Sma*I	*Xho*I
size (kb)	**21**	none	**41**	**7**	**4**	none	**4**	none	**20**	**49**
	23			**20**	**6**					
	32			**37**	7					
	50			49	9?					
					10?					
					22?					
					37?					
					55					
					67					

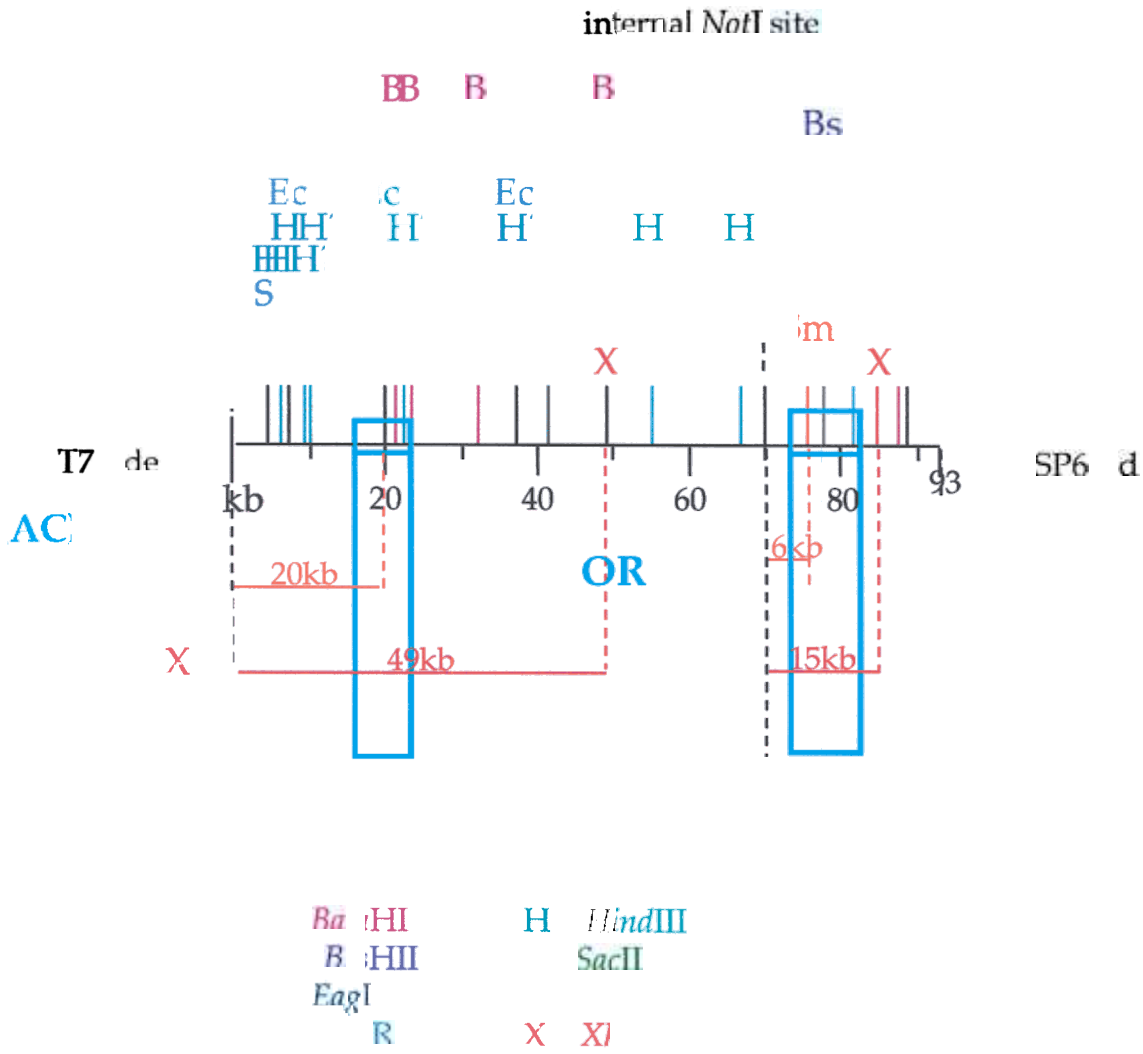
probe: SP6					*Not* I only digest: 23kb					
enzyme	*Bam* HI	*Bss*HII	*Eag*I	*Eco*R I	*Hin*dIII	*Nru*I	*Sac*II	*Sfi*I	*Sma*I	*Xho*I
size (kb)	7	15	none	4	4	none	none	none	17	8
				11						

probe: (AC)n					*Not* I only digest: 70 + 23kb					
enzyme	*Bam* HI	*Bss*HII	*Eag*I	*Eco*R I	*Hin*dIII	*Nru*I	*Sac*II	*Sfi*I	*Sma*I	*Xho*I
size (kb)									**4**	**15**
									6	**49**
			NO DATA OBTAINED						**8**	
									9	
									20	
									31	

Bands apparent in complete digests are in **bold**.

? indicates a band of lower reliability for mapping purposes.

Fig. 3.3.20: PAC2 70 + 23 kb) AC probe



ıer('eral ;m(.kb :h ann t

site of level of confidence

Table 3.3.6: PAC 3 partial digest using probes T7, SP6 and (AC)n

probe: T7				*Not* I only digest: 70kb						
enzyme	*Bam* HI	*Bss*HII	*Eag*I	*Eco*R I	*Hin*dIII	*Nru*I	*Sac*II	*Sfi*I	*Sma*I	*Xho*I
size (kb)	**23**	none	**36**	**7**	very	none	none	none	**22**	**50**
	31			**20**	small					
	49			**33**	??					

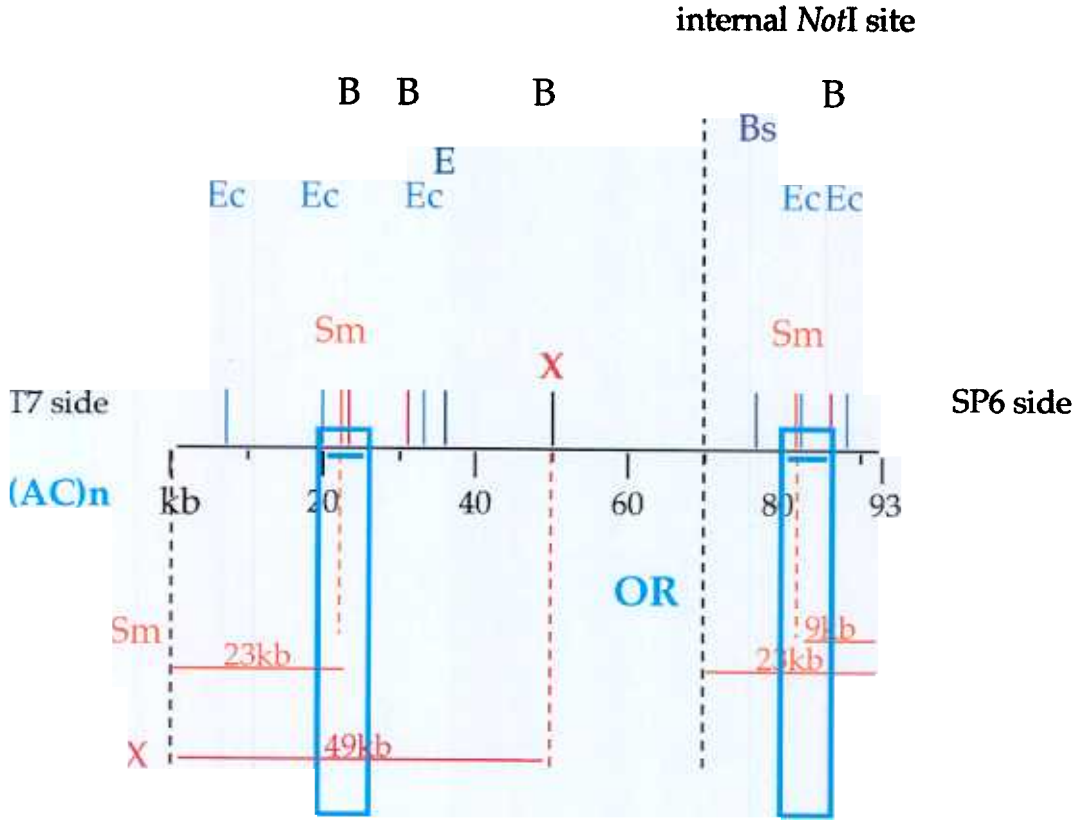
probe: SP6				*Not* I only digest: 23kb						
enzyme	*Bam* HI	*Bss*HII	*Eag*I	*Eco*R I	*Hin*dIII	*Nru*I	*Sac*II	*Sfi*I	*Sma*I	*Xho*I
size (kb)	**7**	**16**	none	**4**	none	none	none	none	**11**	none?
				10						

probe: (AC)n				*Not* I only digest: 70 + 23.1kb						
enzyme	*Bam* HI	*Bss*HII	*Eag*I	*Eco*R I	*Hin*dIII	*Nru*I	*Sac*II	*Sfi*I	*Sma*I	*Xho*I
size (kb)									**4**	**15**
									6	**49**
			NO DATA OBTAINED						**8**	
									9	
									23	
									31	

Bands apparent in complete digests are in **bold.**

? indicates a band of lower reliability for mapping purposes.

Fig. 3.3.21: PAC3 (70 + 23 kb) (AC)n-probe



B = *Bam*HI Ec = *Eco*RI
Bs = *Bss*HII Sm = *Sma*I
E = *Eag*I X = *Xho*I

There are several *Sma*I fragments (4, 6, 8 and 31kb) and one *Xho*I fragment of 15kb which cannot be accounted for.

Table 3.3.7a: PAC 4 partial digest using probes T7, (AC)n and DXS8237E

probe: T7					*Not* I only digest: 170kb					
enzyme	*Bam* HI	*Bss*HII	*Eag*I	*Eco*R I	*Hin*dIII	*Nru*I	*Sac*II	*Sfi*I	*Sma*I	*Xho*I
size (kb)	45	**23**	89	23						
	89	45		34						
	121	89		105						
		99								
		121								

no results for SP6 probe *Not* I only digest: 18kb

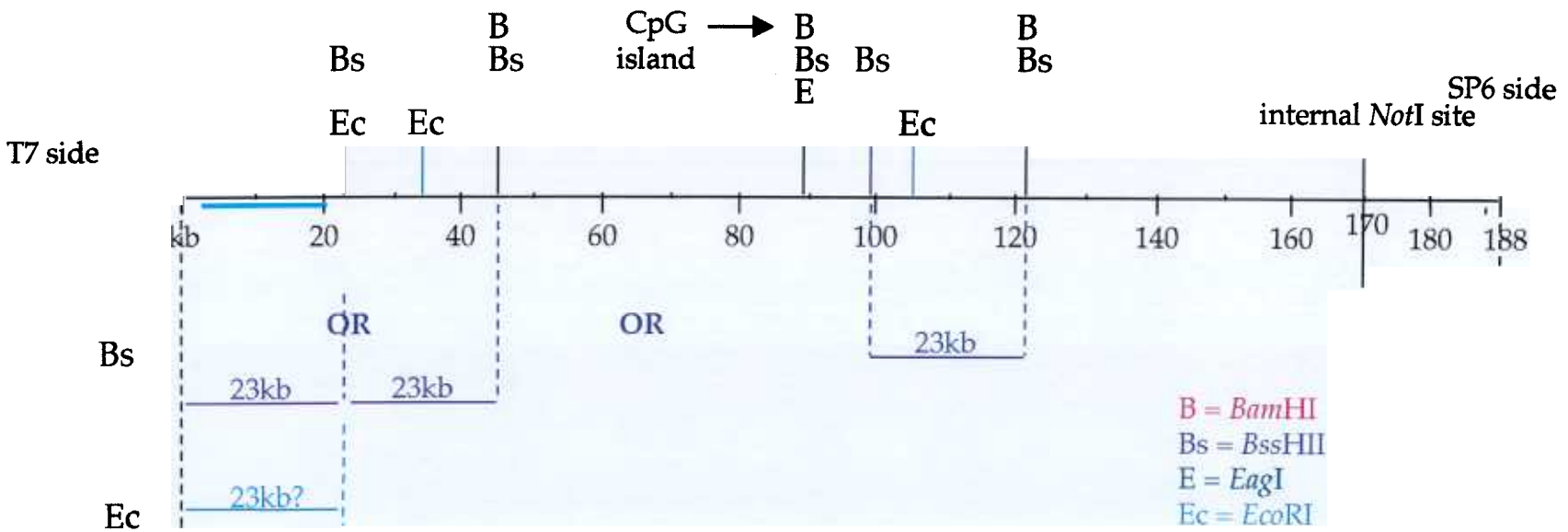
probe: (AC)n					*Not* I only digest: 170 + 188kb					
enzyme	*Bam* HI	*Bss*HII	*Eag*I	*Eco*R I	*Hin*dIII	*Nru*I	*Sac*II	*Sfi*I	*Sma*I	*Xho*I
size (kb)	**4**	**23**	**17**	**<2**	**9**	**47**	**20**	**30?**	**19**	**7**
	8?	49	**23**	**3**	14?	**93?**	23	**49**	**35?**	**8**
	16	71	36	**5**	30	75?	56?	**63?**		9
	23	97?	56	**6**	56	129?	75?	**86?**		**35?**
	36	116?	75	**9**	71?	167?	129	**136?**		
	50		90?	**23?**	86?			**167?**		
	52		104?	49?	104?					
			126?	71?	123?					
			142?	86?						
			152?	116?						

probe: DXS8237E					*Not* I only digest: 170 + 188kb					
enzyme	*Bam* HI	*Bss*HII	*Eag*I	*Eco*R I	*Hin*dIII	*Nru*I	*Sac*II	*Sfi*I	*Sma*I	*Xho*I
size (kb)	**16**	**23**	**12**	**22**	**29**	**44**	**12**	**6**	**4, 5, 6**	**22**
	32?	47	47?	77?	44	76	**16**	16	**7, 8, 10**	
	47?	54?	59?	94?	61?	90	29	**18**	**13**	
	94?	72?	82?	111?	76?	122	90	72	**17**	
		100?	94?		90?	153?	122	90	**20**	
		119?	106?		100?				**37**	
			129?						64	
			141?						84	

Bands apparent in complete digests are in **bold**.

? indicates a band of lower reliability for mapping purposes.

Fig. 3.3.22: PAC 4 (170 +18kb) (AC)n-probe

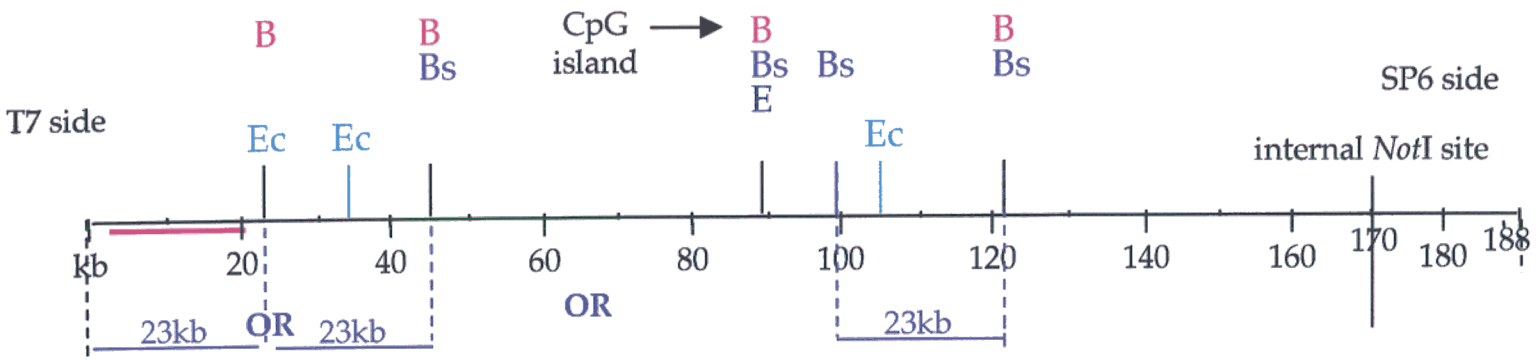


Because of the resolution at this scale of analysis the following *Bam*HI fragments cannot be accounted for: 4, 8, 16 and 23kb.

Similarly, the following *Eco*RI fragments cannot be accounted for: <2, 3, 5 and 6kb.

Similarly, the following *Eag*I fragments cannot be accounted for: 17 and 23kb.

Fig. 3.3.23: PAC 4 (170 + 18kb)-DXS8237E-probe



Bs

Ec 22kb

B = *Bam*HI
Bs = *Bss*HII
E = *Eag*I
Ec = *Eco*RI

Because of the resolution at this scale of analysis a *Bam*HI 16kb fragment cannot be accounted for.

Similarly, an *Eag*I 12kb fragment cannot be accounted for.

Table 3.3.7b: PAC 4 partial digest using probes UBE1 and PCTK1

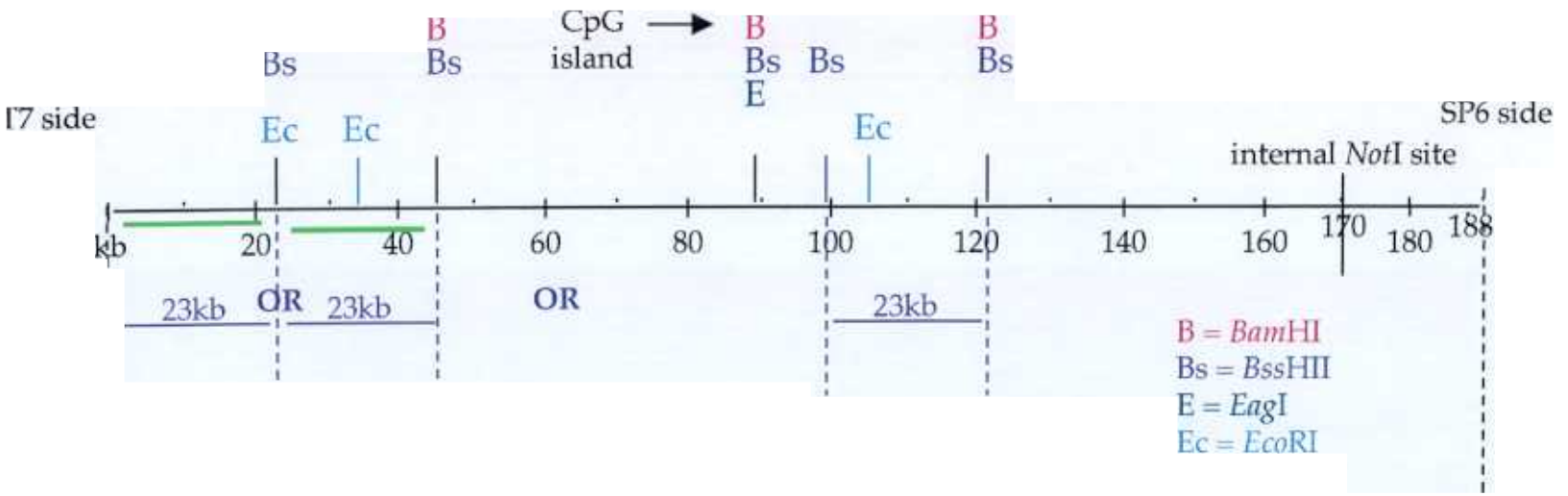
probe: UBE1	*Not* I only digest: 170 + 188kb									
enzyme	*Bam* HI	*Bss*HII	*Eag*I	*Eco*R I	*Hin*dIII	*Nru*I	*Sac*II	*Sfi*I	*Sma*I	*Xho*I
size (kb)	**8**	**23**	**23**	**20**	**9**	**47**	**20**	**3**	**19**	**35**
		48	36	86?	56?		129?	**8**	35?	
		71	56					**49**		
		97	75					63		
		116?								

probe: PCTK1	*Not* I only digest: 170 + 188kb									
enzyme	*Bam* HI	*Bss*HII	*Eag*I	*Eco*R I	*Hin*dIII	*Nru*I	*Sac*II	*Sfi*I	*Sma*I	*Xho*I
size (kb)	**9**	**23**	**26**	**17**	**21**	**34**	**15**	26?		
	26	48	37		26	45	34	**49**		
	46	70	59		53		45	**69?**		
	60	102?	74					**95?**		
	(86)	121?	92?							

Bands apparent in complete digests are in **bold.**

? indicates a band of lower reliability for mapping purposes.

Fig. 3.3.24: PAC 4 (170kb + 18kb) UBE1-probe

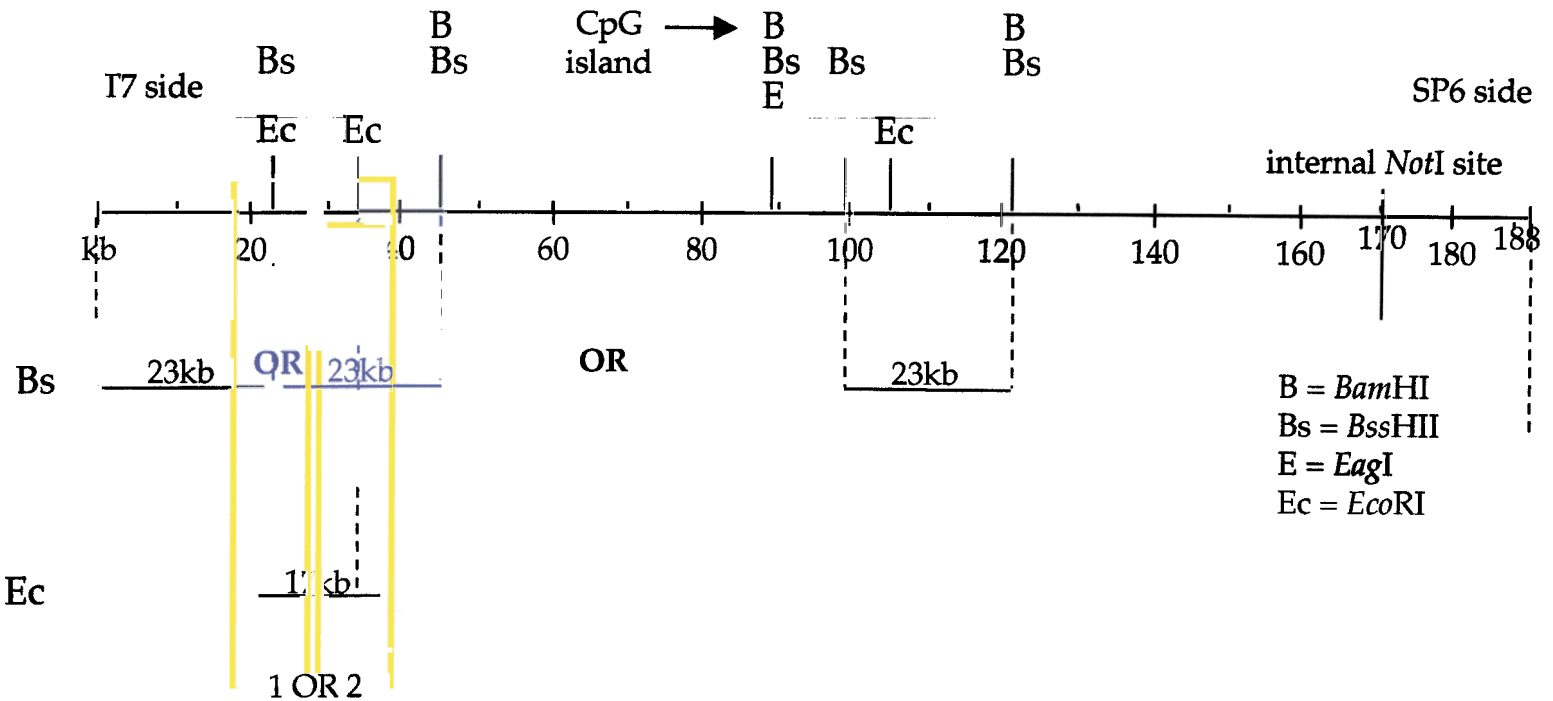


Because of the resolution at this scale of analysis a *Bam*H1 8kb fragment cannot be accounted for.
Similarly, the following *Eag*I fragment is not accounted for: 23kb.

Similarly, the following *Eco*RI fragment is not accounted for: 20kb.

Knowing the order of the genes (Carrel et al., 1996) UBE1 must lie in one of the two indicated regions

Fig. 3.3.25: PAC 4 (170 + 18kb) -probe

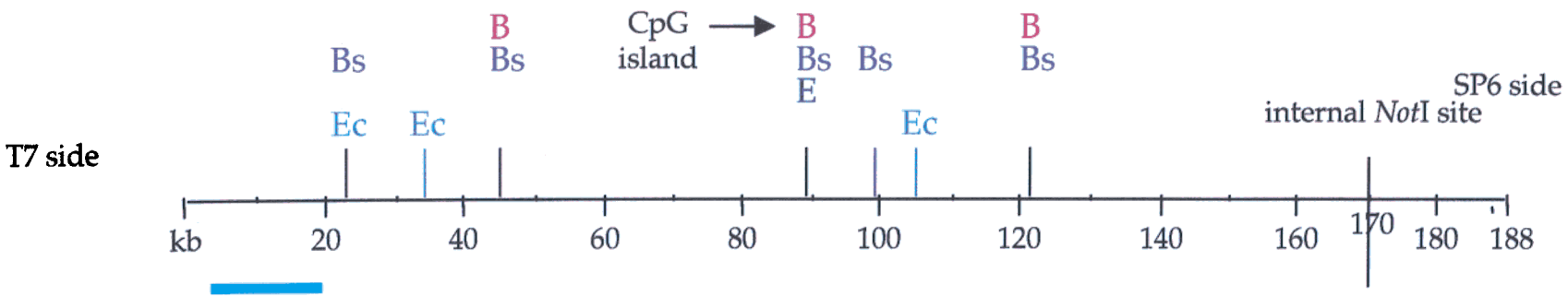


Because of the resolution at this scale of analysis a 9kb *Bam*HI fragment cannot be accounted for.

Similarly, a 26kb *Eag*I fragment cannot be accounted for.

Knowing the gene order (Carrel et al., 1996) the most likely position for is the one indicated by "2"

Fig. 3.3.26: Putative map of PAC 4 (170 +18kb)

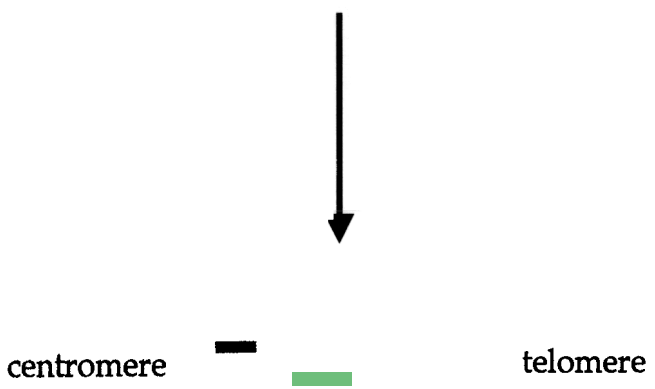


(AC)n

DXS8237E

UBE1

B = *Bam*HI
Bs = *Bss*HII
E = *Eag*I
Ec = *Eco*RI



Knowing the gene sequence (Carrel et al. 1996) and the position of the (AC)n repeat (see Chapter 6) the the most likely arrangement for the gene probes (indicated by colour coding) is indicated to the left.

Table 3.3.8: PAC 5 partial digest using probes T7, SP6 and (AC)n

probe: T7 ***Not* I only digest: 126kb**

enzyme	*Bam* HI	*Bss*HII	*Eag*I	*Eco*R I	*Hin*dIII	*Nru*I	*Sac*II	*Sfi*I	*Sma*I	*Xho*I
size (kb)	**19**	**20**	**102**	**2**	**7**	34?	**66**	**39**	**<2**	**73**
	24			**4**	41?	**66**			3	91
	38			**9**	91?				6	
	51			**20**					8	
				22					23	
				38					31	
				51?						

probe: SP6 ***Not* I only digest: 126kb**

enzyme	*Bam* HI	*Bss*HII	*Eag*I	*Eco*R I	*Hin*dIII	*Nru*I	*Sac*II	*Sfi*I	*Sma*I	*Xho*I
size (kb)	**5**	none	**19**	2	**3**	**5**	23	none	**5**	**5**
	120			**3**	**5**	**67**	**73**		6	**7**
				5	**8**				**7**	**40**
				6	**9**				**15**	
					13				**23**	
					33					
					55					

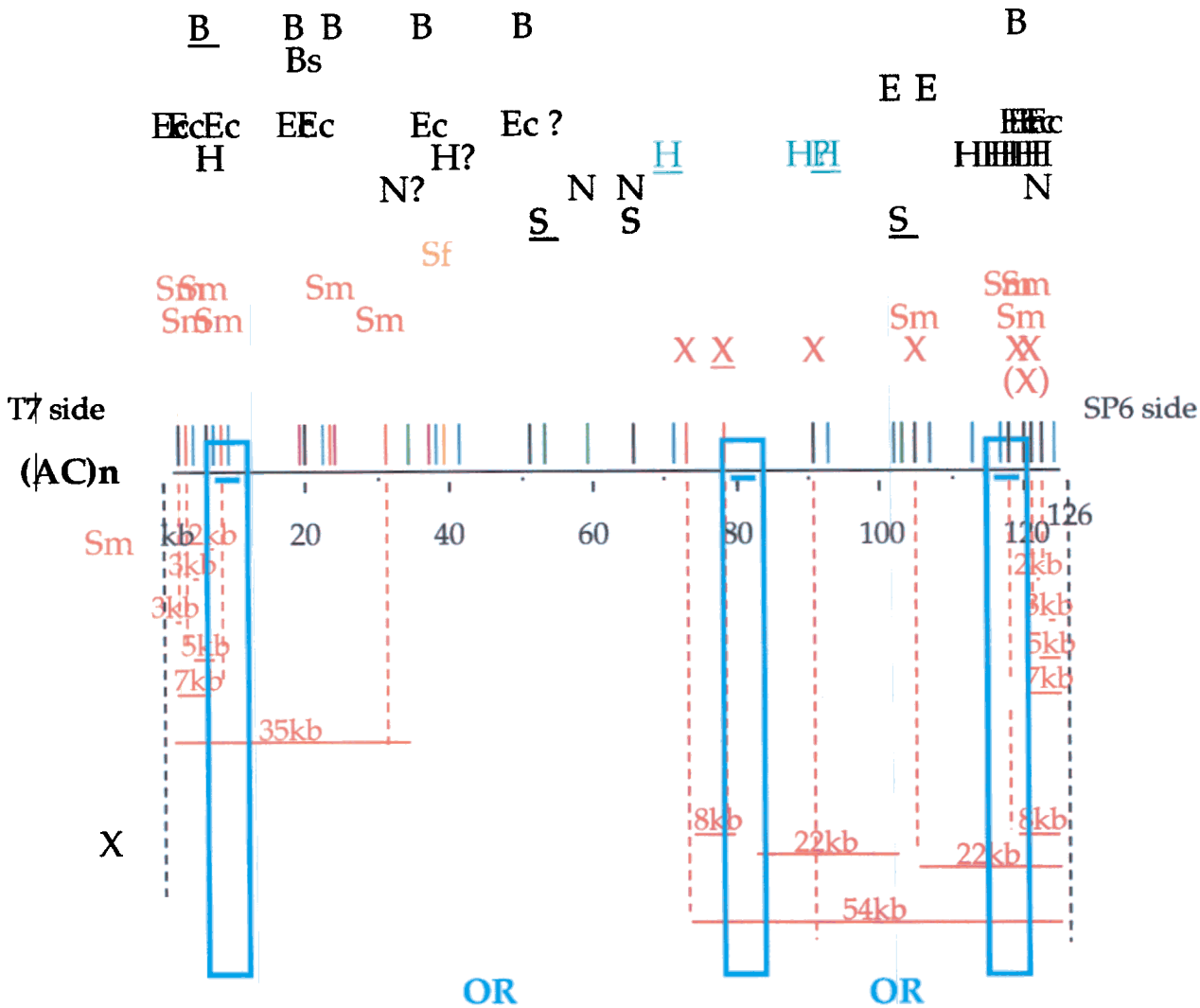
probe: (AC)n ***Not* I only digest: 126kb**

enzyme	*Bam* HI	*Bss*HII	*Eag*I	*Eco*R I	*Hin*dIII	*Nru*I	*Sac*II	*Sfi*I	*Sma*I	*Xho*I
size (kb)									**2**	**8**
									3	**22**
			NO DATA OBTAINED						**5**	**47**
									7	**54**
									11	89
									35	
									66	

Bands apparent in complete digests are in **bold**.

? indicates a band of lower reliability for mapping purposes.

Fig. 3.3.27: PAC 5 (126kb) (AC)n-probe



One *Sma*I 11kb fragment cannot be accounted for.
One *Xho*I 47kb fragment cannot be accounted for.

Bands underlined are bands on the T7 side, which were derived from the SP6 side

?: site of lower level of confidence

B = *Bam*HI
Bs = *Bss*HII
E = *Eag*I
Ec = *Eco*RI
H = *Hind*III
N = *Nru*I
S = *Sac*II
Sf = *Sfi*I
Sm = *Sma*I
X = *Xho*I

3.4 Interpretation and Discussion

3.4.1 Restriction mapping using YACs

At first, there seemed no problems with the YACs provided. Certainly YACs 2910 and 2911 were stable in size (figure 3.3.1). The marker primers 117.29, which were used by G. Black to obtain YAC 2910 and YAC 2911, were still present n both YACs (figure 3.3.2). This result was confirmed by hybridisation using the 117.29 probe, which also gave a positive result in YAC A1S9T, again giving no signs of instability. First indications that YAC A1S9T might be rearranging were observed whilst trying to clone the YAC ends when two left and two right end products were obtained by inverse PCR (figure 3.3.3b). At a later point, after the YAC had been regrown from the original stock, a considerable size difference became apparent (results not shown), underlining the instability of YAC A1S9T. The apparently symmetrical pattern of restriction enzyme sites when analysed from either end also suggest the existence of complex rearrangements.

Database analysis of the sequence obtained for the end clones from YACs 2910 and 2911 revealed that the right end of YAC 2911 contained an exon from a gene expressed in fœtal lung tissue. PCR on a monochromosomal hybrid panel revealed that this exon is located on chromosome 3 (figure. 3.3.4), proving that YAC 2911 is chimæric. A more recent database search revealed that the sequence matches part of a T-cell antigen receptor-interacting molecule or TRIM on chromosome 3 at 3q13.

A contig map covering a 2Mb region distal to DX6849 at the human Xp11.3 locus was recently published by Stoddart et al., 1999. This map is based on YACs obtained from the St. Louis, ICI and CEPH libraries and PACs from the de Jong pCYPAC library (Ioannou et al., 1994). It was derived by using STS-content mapping and hybridisation data. Mapping was carried out with a set of 21 markers localised within the Xp11.3 region and 16 region specific STSs/ STRs and ESTs. The map produced is in agreement with the consensus map published during the 6th X Chromosome workshop. However, a restriction map was not presented by Stoddart et al. (1999).

More than half of the YACs used in the physical mapping study carried out by Stoddart et al. (1999) were shown to be chimæric and frequently contain deletions. This is also true. It is possible that the observed instability is not only due to the instability inherited in YAC technology of YACs themselves, but also that this is a region-specific phenomenon in the genome. For example, in previous reports using the same YAC libraries, a 2Mb contig in neighbouring Xp11.23-p11.22 observed a 30% rate of chimærism and a 40-80% rate of internal deletions (Boycott et al., 1996), while a 3.6Mb contig constructed in Xp22.1 contained only 11% chimærism and 6-22% internal deletions (Trump et al., 1996).

Although two of the three YACs in this study were obviously not representative of the correct genomic configuration, four novel STSs could be constructed, which proved useful with the initial mapping. Combining all the hybridisation results obtained using the left and right YAC vector arms, STSs and gene markers as probes allowed a preliminary restriction map of all three YACs to be constructed (see figure 3.3.15). Difficulties in reconciling the data between the YACs are presumed to reflect rearrangements. Furthermore, two of the STSs (S2910l and S2910r), obtained from the only stable YAC (Y2910), proved to be very valuable since they were used to obtain PACs from the UBE1 region.

Interpretation of restriction mapping data

The hybridisation results using the STS 2910 left end as a probe predicted its absence from YAC A1S9T or YAC 2911. This contradicts the restriction map generated in this study. The most likely explanation is that the STS 2910l overlaps with a region of YAC A1S9T that is rearranging. Another possible explanation for STS 2910l not hybridising to either YAC filter is that both filters had been stripped several times, a process, which can result in loss of material. The hybridisation was not repeated with new filters, since at that point the problems with YAC instability had been recognised and the mapping was concentrated on the PACs obtained. The same explanation holds true obtaining a negative result when hybridising S2910r to the filter containing YAC A1S9T partially digested DNA. All other data are consistent with the restriction map generated for YACs A1S9T, 2910 and 2911 (figure. 3.3.15); however, it is recognised that the map represents a compromised interpretation generated in

an attempt to overcome known incongruities in the structures of YACs A1S9T and 2911

The gene probes DXS8237E, UBE1 and PCTK1 hybridise to the filters containing YA1S9T and Y2911 DNA, but not to the filter containing Y2910 DNA. The genomic order is known to be cen - *DXS8237E - UBE1 - PCTK1* – tel (Carrel et al., 1996). According to this YAC map the three genes under investigation seem to lie within 100kb.

To be able to size fragments smaller than 48.5kb more accurately on the map, lower molecular weight PFG markers could be used. Small fragments obtained in hybridisations using gene probes, which do not fit the restriction map might be due to CG rich regions, known to occur near transcribed sequences and providing a multiplicity of rare cutting enzymes sites (Antequera and Bird, 1993; Gardiner-Garden and Frommer, 1987). In addition, CG rich regions can make it difficult for enzymes to recognise some restriction sites because of their secondary structure formation. This can result in sites not showing up on restriction maps and also compromises the reproducibility of them This featur also holds true for restriction mapping the PACs.

3.4.2 Restriction mapping using PACs

The initial PCR results employed in an attempt to select the most promising PACs for investigation were chosen have not been included in the maps presented. Each PCR reaction was carried out five times in total; however, the results obtained were inconsistent This might have been due to slight contamination of the PAC DNA sample with other DNA, most likely from another PAC. Since PCR reactions are very sensitive, even the slightest contamination would have led to an apparently positive result. Therefore, restriction maps are based on the hybridisation results, which, although sensitive, are less liable to misinterpretation resulting from low level contamination or rearrangements

Apart from PAC5, all the PACs examined in detail have an internal *Not*I site

PAC1 contains sequence detected by the DXS8237E and UBE1 gene probes, but not those for the PCTK1 gene or the (AC)n repeat probe. PACs are generally thought

to be more stable than YACs (Sternberg et al. 1990). However, given the close clustering of the loci under study, the absence of the PCTK1 gene and the (AC)n repeat from PAC1 suggests a rearrangement event. Recent communications from M. Ross at the Sanger Centre confirmed that this PAC is unstable. The (AC)n repeat lies within *UBE1* and distal to the location of the UBE1 probe see Chapter 6) and*PCTK1* lies distal to *UBE1* Neither of the two probes being contained in the PAC must mean that a deletion has taken place within the *UBE1* gene, distal to the UBE1 probe.

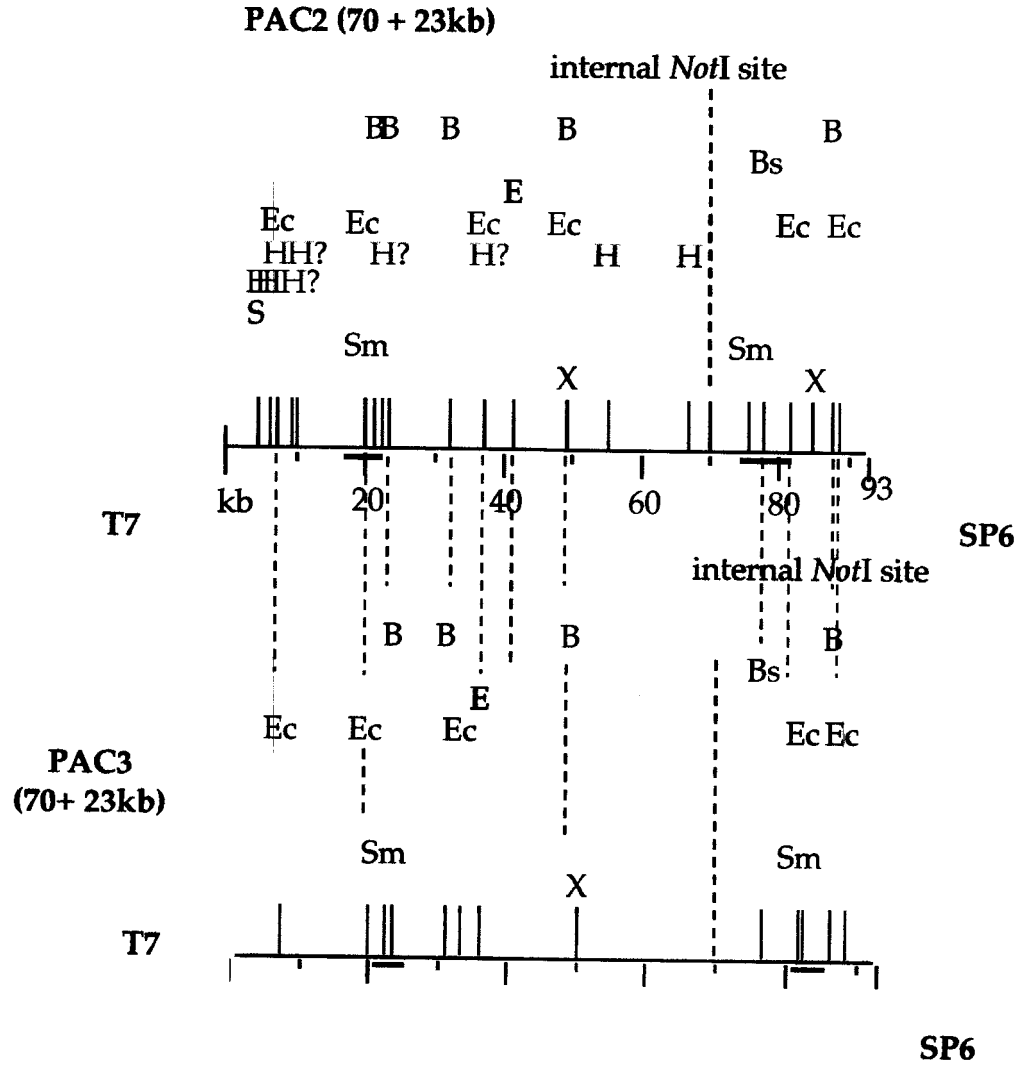
PAC2 and 3 seem very similar, if not identical, to each other (see figure 3.4.1) The results allowing this interpretation suggest that the mapping approach adapted has been fairly accurate.

PAC4 being the largest and the only PAC containing all the gene probes, is the most informative and should prove very useful in further studies. Values of 170 and 188kb for the single *Not*I digest using the (AC)n and DXS8237E probes indicate that the digest may have not gone to completion. The reason for there not being any evidence for larger fragments containing part of the vector as well is that *Not*I might preferentially cleave at the insert vector border, rather than in the midst of the human sequence because of its high GC content. The large number of *Eag*I sites observed are most likely to be internal sites, which are not detected when complete digests were hybridised with end fragments. Analysis of this PAC show the *DXS8237E, UBE1* and *PCTK1* genes fall within about 30kb of each other.

PAC5, although larger than PAC2, has great similarity to it. Two possibilities of overlap exist and both are included in figure 3.4.2 in order to illustrate the problems of aligning maps in this region. Other possibilities include that either PAC5 was derived from PAC2 as a consequence of an insertion, or that PAC2 resulted from PAC5 as a result of a deletion event.

In summary, PAC1 seems to be rearranging and PACs 2, 3 and possibly PAC5 may have been derived from the same original clone (see figures 3.4.1 and 3.4.2). Only PAC4 contains all three gene probes (see figure 3.4.26) and comparing the PAC maps to each other was complicated by the fact that not all restriction sites were obtained for all PACs. Some PACs share as little as three restriction sites that can

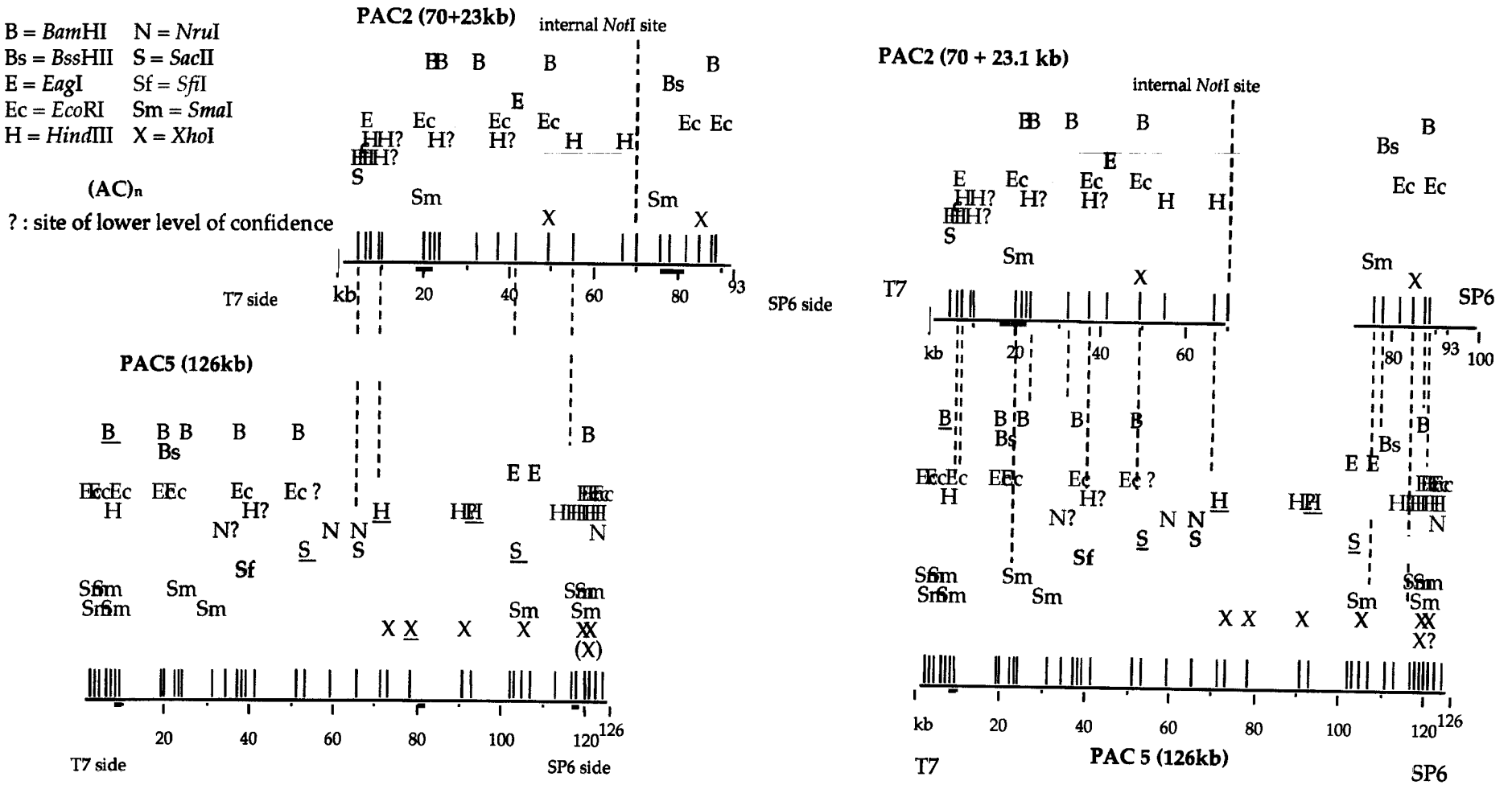
Fig. 3.4.1 Comparison of PAC 2 and PAC 3



B = *Bam*HI	N = *Nru*I	
Bs = *Bss*HII	S = *Sac*II	
E = *Eag*I	Sf = *Sfi*I	$(AC)_n$
Ec = *Eco*RI	Sm = *Sma*I	
H = *Hind*III	X = *Xho*I	

? : site of lower level of confidence

Fig. 3.4.2 Comparison of PAC2 and PAC5



Bands underlined are bands on the T7 side,
which were derived from the SP6 side

be used for comparison. The (AC)n repeat data may be confounded by the existence of related sequences which could cross hybridise (see also Chapter 6). Further, the high GC content of the region may explain its instability as evidenced by the difficulty in obtaining a totally coherent contig map for both, YACs and PACs. The instability of the region is further confirmed by the sequencing undertaken at the Sanger Centre (see Chapter 6 for detail)

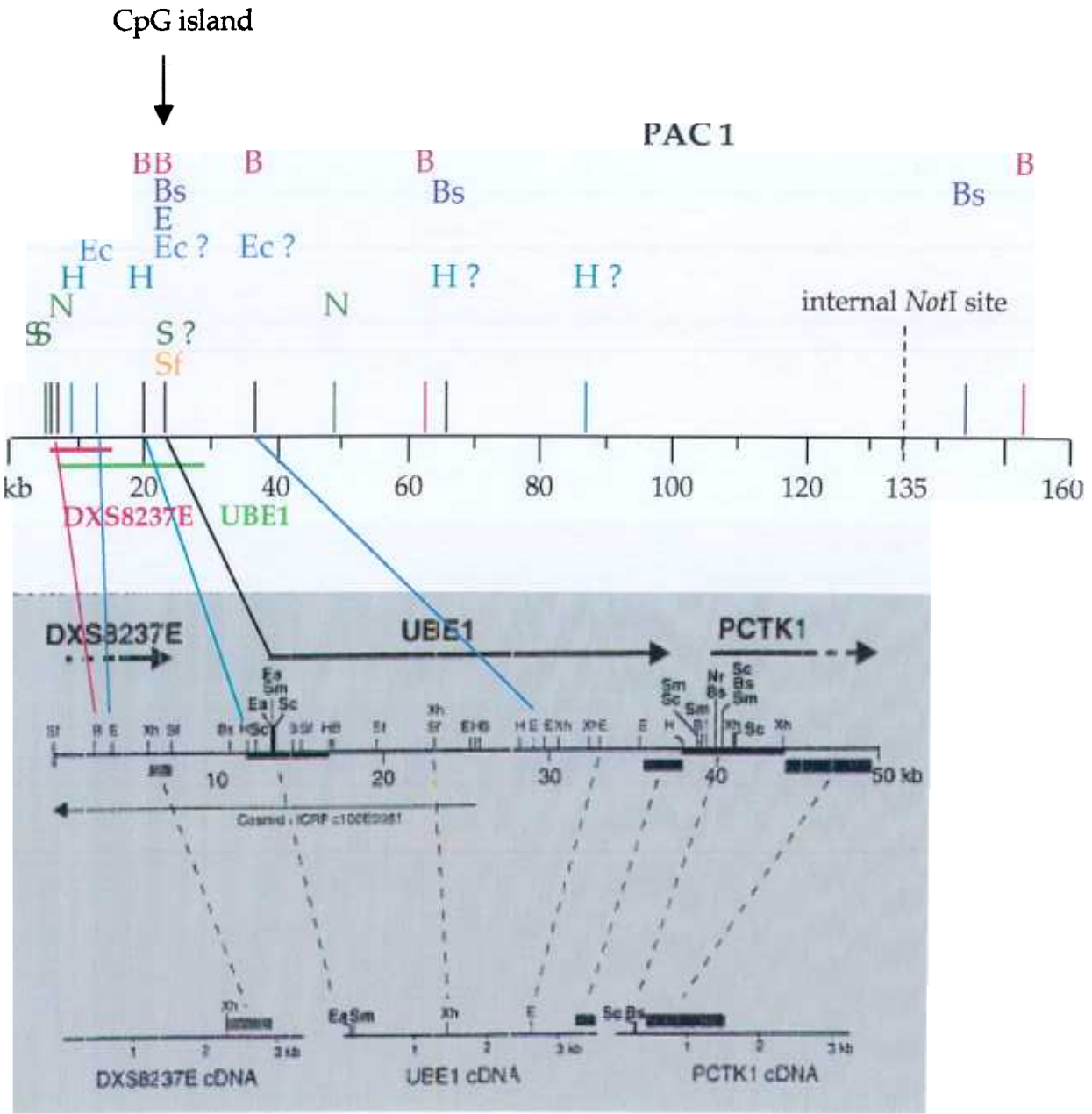
3.4.3 Comparing maps

The approaches adopted to investigate this region were modified progressively as information was generated. In the first phase, YACs were employed to gain a gross picture of the region. Then PACs were obtained with the end clones from one presumptively stable YAC, when it became obvious that isolating a PAC contig and sequencing would be the most efficient and productive strategy

The size difference of YAC and PAC maps, the varying patterns of restriction enzymes and the observed high frequency of rearrangement make it difficult to compare both maps directly to each other.

One restriction map of the DXS8237E-UBE1-PCTK1 area was published by Carrel et al. (1996), which is compared to the restriction map of PAC1, since they have most of the restriction enzymes used in common (see figure 3.4.3) Within the *DXS8237E* gene the authors established a *Bam*HI and *Eco*RI site, which are also present in the PAC1 map obtained in this study. However, a *Xho*I and *Sfi*I site were not located in this study. The authors found a CpG island at the 5' end of UBE1 containing restriction enzymes *Eag*I, *Sac*II and *Sma*I. Proximal to it there is a *Bss*HII and *Hind*III site, whereas distal a *Bam*HI, *Hind*III and *Sfi*I site were found. In my study a CpG island containing the methylation-sensitive enzymes *Bss*HII, *Eag*I, *Sac*II was detected in PAC1. Further, in close proximity to it, the sites recognised by *Bam*HI, *Eco*RI, *Hind*III and *Sfi*I were found Further distally the only restriction site that Carrel et al.'s map and the one obtained in this study have in common is the *Eco*RI site, which might be of low significance. As PAC1 is known to be rearranging, it is not surprising that the restriction sites distal to the CpG island do not coincide. The presence of a CpG island might be expected in the vicinity of a gene and the *Bam*HI, *Hind*III, *Sfi*I and *Xho*I sites the authors obtained coincide with an earlier restriction map of the PC AIRE region (Okuda

Fig. 3.4.3 Restriction map comparison of PAC1 and the one published by Carrel et al. (1996)



B = *Bam*HI
Bs = *Bss*HII
E = *Eag*I
Ec = *Eco*RI
H = *Hind*III
N = *Nru*I
S = *Sac*II
Sf = *Sfi*I

? : site of lower level of confidence

Abreviations regarding the map by Carrel et al. are as follows:
B = *Bam*HI
Bs = *Bss*HII
Ea = *Eag*I
E = *Eco*RI
H = *Hind*III
Nr = *Nru*I
Sa = *Sac*II
Sf = *Sfi*I
Sm = *Sma*I
Xh = *Xho*I

et al., 1994). Since many of the restriction enzymes used by Carrel et al. were not used in this study to digest the only PAC containing the PCTK1 gene probe (the two studies were carried out simultaneously), PAC4, it is possible that these sites are physically present, but were not detected. Allowing for the difference in scale between the two maps and the likely errors in sizing fragments they seem to be in rough overall agreement.

Initially, since large gaps existed in the contig, it was intended to extend this distally to the *UBE1* gene. This was going to be carried out after producing a restriction map of the region and using the STSs obtained from the YACs. However, although much more could be done in an attempt to obtain a detailed contig and restriction map for this region, the obvious instability and the complications arising from the multiplicity of sites for rare cutting enzymes, it was decided to concentrate on deriving sequence information on the relative small region of specific relevance to the problem of X chromosome inactivation.

3.4.4 Isolation of the interval spanning DXS8237E and UBE1

During the final investigations of the PACs, two of them, PACs 1 and 4, were found to contain the *DXS8237E* and *UBE1* genes within one 9kb *Xho*I/ *Eag*I fragment. This fragment was purified from agarose gels, but the cloning of it into a vector proved difficult, which again most likely is due to the unusually high GC content in this region. It was decided not to proceed with cloning and sequencing of this fragment, since PAC1 was already being sequenced by the Sanger Centre. Unfortunately, PAC1 turned out to be chimæric, so work was initiated on PAC4, which was subcloned into BACs, the sequences of which were combined into the genomic assembly AL513366. Sequence analysis of the DXS8237E-UBE1-PCTK1-UHX1 region was carried out as described in Chapter 6.

The argument of instability of the interval spanning DXS8237E and UBE1 is supported by the fact that a region on chromosome 3, which is thought to have arisen from an inverted duplication of the above-mentioned interval on the X chromosome, shows the same instability when cloned into YACs. On chromosome 3, the genes similar to *DXS8237E* and *UBE1* are called *RBM5* and *UBE1L*, respectively (Timmer et al., 2002).

CHAPTER 4 - Applications of a UBE1 associated SSR polymorphism through association studies

4.1 Introduction

To make use of *UBE1* or of a closely linked microsatellite marker for it as part of a genome screen for quantitative trait loci (QTLs) implicated in cognitive ability (g) is of particular interest for several reasons.

irstly brain differences between males and females hav hown impac on language and transsexuality (Shaywitz et al 1995; Zhou et al 995 The *UBE1* gene which plays an important role in protein degradation lies on one of the sex chromosomes, the X chromosome, and escapes inactivation In humans it does not have a Y chromosome homologue (see Chapter 1). Because of its sex difference in dosage it may play a role, directly or indirectly differences between male and female development

Secondly, NEDD8 is an ubiquitin like molecule, required for conjugation to target proteins. Its activation requires the human UBA3, which has homology to the C-terminal region of UBE1, and APP-BP1 (beta-amyloid precursor protein-binding protein-1), a protein with sequence homology to the N-terminal region of UBE1 (Gong and Yeh, 1999). Thereafter NEDD8 is linked to the E2-like enzyme, UBC12. The mouse homologue Nedd8 is one of a group of genes showing developmentally regulated expression in the embryonic brain (Kumar et al., 1992). Interactions between ubiquitin and NEDD8 may be important and underlining the potential role of *UBE1* in brain development

Thirdly familial frontotemporal dementia (FTD) or Pick's disease is a disease in which initial clinical patterns include personality and behavioural changes, apathy disorganisation and inability to concentrate. In ater stages anguage disorders, overeating and grabbing of food are observed biquitin immunostaining demonstrated differently shaped inclusions in various areas of the brain. These ubiquitin-positive, synuclein- and tau-negative inclusions, previously associated with a motor neuron type" of dementia, are considered indicative of a relatively common variety of Pick's disease (Kertesz et al 2000) Ubiquitin being directly linked to a disease causing mental disorders is possibly

another indication for a role for *UBE1* in cognitive ability

Fourthly the production of mutant proteins containing expanded polyglutamine repeats are due to expansions of CAG repeats in the genome polyglutamine expansions are known to cause at least eight diseases including Huntington disease), of which in six the mutant protein forms aggregates termed neuronal inclusions that also contain ubiquitin. Neurons containing these aggregates eventually undergo apoptosis, with corresponding decreases in brain mass in affected individuals (reviewed by Perrigno and Silver 200(Further the authors suggest that the nuclear inclusions containing ubiquitin are likely to have a protective function, rather than being "harmful". The exact role of ubiquitin still has to be determined Therefore, ubiquitin may play a role in these neurodegenerative diseases

Finally, allelic association studies have become easier and more feasible to carry out by a method involving DNA pooling and which can be used as a rapid preliminary screen for allelic association with simple sequence repeats (SSRs) (Daniels et al 1998) Some of these issues are elaborated on below

4.1.1 Brain differences in males and females

Remarkable differences in the functional organisation of the brain for a specific component of language, phonological processing, between normal males and females have been demonstrated (e.g. see Shaywitz et al., 1995).

Transsexuals have the strong feeling, often from childhood onwards, of having been born the wrong sex. The possible psychogenic or biological aetiology of transsexuality has been the subject of debate for many years. In 1995, Zhou et al. showed that the volume of a brain area that is essential for sexual behaviour central subdivision of the bed nucleus of the stria teminalis (BSTc (Kawakami and Kimura, 1974; Emery and Sachs, 1976.), is larger in men than in women. A female sized BSTc was found in male-to-female transsexuals. The size showed not to be influenced by sex hormones in adulthood and was independent o sexual orientation. This study supports the hypothesis that gender identity develops as a result of an interaction between the developing brain and sex hormones (Editorial Lancet, 1991; Swaab and Hofman, 1995)

4.1.2 NEDD8 (Neural precursor cell-Expressed Developmentally Down-regulated)

NEDD8 is an 81-amino acid polypeptide that is 60% identical and 80% similar to ubiquitin (Kumar et al., 1993). It is highly conserved in mammalian species with only one amino acid difference between the human NEDD8 and the mouse Nedd8 (Kamitani et al., 1997). The C terminus of NEDD8 is efficiently processed to expose Gly-76, which is required for conjugation to target proteins. The modification of Hs-cullin-4A (Cul-4A) by NEDD8 has an important role for regulation of the cell cycle. Human Cul-4A is a member of the family of human cullin/ Cdc53 proteins and functions as an essential component of a multifunctional ubiquitin-protein ligase E3 complex that has a critical role in ubiquitin-mediated proteolysis (see Chapter 1; Osaka et al., 1998). NEDD8 activation requires the human UBA3 and APP-BP1, a protein with sequence homology to the N-terminal of UBE1. A role for UBE1 in brain development is therefore a formal possibility.

UBA3 has a conserved ATP-binding site and an active site, a homologue to the C-terminal half of UBE1 It can form a beta-mercaptoethanol-sensitive conjugate with NEDD8 in the presence of APP-BP1 (Gong and Yeh, 1999). The human UBA3 is 38% identical to the yeast homologue, 22% identical to human UBA2, the gene for which has been cloned and positioned on chromosome 19q12 and 19% identical to the C-terminal of UBE1. The *UBA3* gene is located on chromosome 3p13 and gives rise to a 2.2kb pair transcript that can be detected in all tissues (Gong and Yeh, 1999).

4.1.3. The *UBE3A* gene

UBE3A, the gene encoding the ubiquitin-protein ligase E3A, has also been identified as a gene implicated in Angleman syndrome (AS). AS is characterised by moderated to severe mental retardation, absence of speech, tremor, ataxia, abnormal gait, inappropriate laughter, sleep disturbance, and seizures. The incidence of AS is estimated to be 1 in 20,000, with most cases being sporadic, although familial occurrence is not rare (Fang et al., 1999)

UBE3A is located at chromosome 15q11-q13 and 7 polymorphisms have been identified (van der Weele et al., 1999). It is expressed in fibroblasts and lymphoblasts, but imprinted in the brain (Vu and Hoffman, 1997; Rougeulle et

al., 1997). It is normally expressed from the maternal chromosome and silent on the paternal allele. Imprinted genes not only play an important role in brain function and behaviour as a result of genetic defects as in AS, Prader-Willi syndrome and Albright hereditary osteodystrophy, but also may be involved in psychiatric disorders and behavioural phenotypes (Nicholls, 2000). It important to note that many imprinted genes expressed in the central nervous system control the actions of other genes by acting as oncogenes, tumour suppressers, transcription and growth factors or by regulating the products of other genes through alternative RNA splicing and protein degradation. Hence, their effects are complicated and indirect. Again, interaction of UBE1 with these proteins may lead to its indirect importance in brain development and cognitior

4.1.4 Definition of terms used in association analysis

Allelic Association

This term describes the greater than random coinheritance of alleles at closely linked loci resulting from linkage disequilibrium. It provides the basis for case control studies designed to uncover QTLs implicated in multifactorial disorders. These are revealed by the observation of a higher frequency in the cases of an allele at an adjacent locus to that at which the mutation defining the QTL originated.

Power studies

Prior to generation of genotype data and prior to collection of DNA for case/ control studies, it is necessary to evaluate the power to detect association. Power analysis can be used to predict the sample size needed to reliably detect association. It allows the determination of -if enough data are available- a successful outcome. Power analysis can give information regarding the probability of generating a false positive result and the average exclusion region possible in a given data set.

Chi square test

This evaluates differences between observed and predicted/ expected results by hypothesis or comparison to another set of results. Chi square is calculated as follows:

$$X^2 = \sum \left[\frac{(X \text{ obtained} - X \text{ expected})^2}{X \text{ expected}} \right], \text{ where}$$

Σ is the sum of

X obtained is the experimentally obtained number

X expected is the under optimal conditions obtained or expected number.

Significance/ p-values

p-values show how likely it is that the result obtained is obtained by chance, ie. how significant the data are. The maximum value is 1 and the lower the value, the more likely it is that the result is not due to chance. If the value is lower than 0.05 the result is said to be highly significant, i.e. only 1 in 20 time would the difference be observed by chance. p-values depend on the X^2 obtained.

4.1.5 Association analysis

Allelic association studies provide the most powerful method for locating genes of small effect contributing to complex diseases and traits (Owen and McGuffin, 1993; Risch and Merikangas, 1996). Association of an allele with a particular trait is seen as a difference in allelic frequencies between groups. Comparison of allele frequencies between patients and controls from the same population is a simple method to localise susceptibility genes and to examine candidate genes.

Usually, a Pearson chi-squared test is performed on a contingency table in which the two columns contain the counts of various alleles in the patients and controls. The scope of allelic association studies has been widened considerably

by the large number of highly polymorphic markers made available through advances in molecular genetic technology (Sobell et al., 1993).

A potential problem in an association analysis comparing cases and controls with respect to allele frequencies at a highly polymorphic locus is that the conventional chi-squared test may not be valid for a large, sparse contingency table. However, statistics with known asymptotic distribution are not required anymore as Monte Carlo simulations can be performed to estimate the significance level of any statistic. Sham and Curtis (1995) have employed a Monte Carlo method for four 'chi-squared' test statistics, which they implemented in a computer program called CLUMP. This again generates a X^2 value from which a p-value can be obtained.

4.1.6 Association analysis and DNA pooling

Up to 1998 large-scale systematic searches for allelic association had not been carried out, because of the belief that they were not technically feasible, since dense marker maps and large numbers of subjects are required To reduce the amount of genotyping required Daniels et al. in 1998 suggested sample pooling DNA pooling has been used successfully in quantitative trait locus (QTL) association studies in animals, where it was combined with "selective genotyping", in which subjects were only genotyped if their scores on the trait under study lied at either extreme of the continuum. DNA from two different groups were then pooled thereby reducing the amount of genotyping (Collin et al., 1996; Darvasi and Soller, 1994). DNA pooling also has been used to study recessive diseases in inbred human populations (Carmi et al., 1995). In both of these cases, a shift towards homozygosity of particular markers was sought in pooled DNAs. However, until recently, DNA pooling had not been successfully applied to case-control association studies in outbred populations because of several problems with the application of pooling to the most common form of human genetic marker, simple-sequence repeats (SSRs). Two of these result from PCR artefacts commonly associated with SSRs: stutter banding and differential amplification, both of which make allele frequencies estimates n pooled samples impossible. In 1998 Daniels et al. reported a method involving

DNA pooling that can be used as a rapid preliminary screen for allelic association with SSRs. They pooled patients and control samples separately and markers were typed in the two pools. By using primers with fluorescent 5' ends, PCR products could be analysed on an automated sequencing apparatus. Allele image patterns (AIPs) produced for the two groups were overlaid and differences in pattern area between pools computed From this, a delta AIPs statistic was calculated from the difference in areas between the two AIPs expressed as a fraction of the total shared and non-shared area. AIPs of a range of different sized pools were generated by computer simulation for markers with a range of allele sizes and frequencies. Delta AIPs from pools and X^2 values for individual genotypings were compared with both simulated and real data from microsatellite markers and showed a high correlation. Delta AIP analysis of real microsatellite data indicated the feasibility of using this method in searches for allelic association, generating a small number of false positives, but few false negatives.

This method has allowed to detect QTLs of small effect (Risch and Merikangas, 1996; Owen et al., 1997) and is said to be simplest when functional polymorphisms in candidate genes are used. It is also possible to detect QTL associations with non-functional DNA markers that are close enough to a QTL to remain in linkage disequilibrium with the QTL for many generations (Hill et al 1999).

4.1.7 DNA pooling and cognitive ability

General cognitive ability (g), which is related to many aspects of brain functioning, is one of the most heritable traits in neuroscience. Similarly to other heritable quantitatively distributed traits, genetic influence on g is likely to be due to the combined action of many genes of small effect, perhaps several on each chromosome (Fisher et al. 1999). So far, two chromosomes have been screened for QTLs associated with cognition. Fisher et al. (1999) used DNA pooling for the first time to search a chromosome systematically with a dense map of DNA markers for allelic associations with general cognitive ability. The authors used 147 markers on chromosome 4, where 11 significant QTL associations emerged. Hill et al. in the same year performed similar studies, screening 66 markers on chromosome 22 in original and replication samples of

children of high general cognitive ability and controls of average g and excluded associations of major effect size on this chromosome.

Given the homology of the ubiquitin and NEDD8 pathways, the significance of the latter in brain differentiation, the direct or indirect association of *UBE1* with neurodegenerate diseases (eg. PICKS disease) and the "improvement" of allelic association studies in pools, the potential contribution of *UBE1*, together with the flanking and similarly non-inactivated *PCTK1* locus as QTLs determining cognition in humans, appears to be worthy of investigation his hapter describes the isolation of a novel (AC)n repeat from the UBE1 region and its examination for allelic associations with general cognitive ability using the method first described by Daniels et al. in 1998. This was done in collaboration with the Social, Genetic and Development Psychiatry Research Centre, Institute of Psychiatry in London

4.2 Materials and Methods

4.2.1 Detection, isolation and sequencing of a novel (AC)n repeat adjacent to *UBE1*

A series of X-specific cosmids derived from YAC 2911 and found on preliminary screening to contain a (AC)n repeat (M. Coleman personal communication) were re-examined by *Eco*RI digestion, Southern blotting followed by hybridisation with poly (dA-dC) x poly (dG-dT) (Pharmacia). This demonstrated that cosB9 contained an (AC)n repeat within an unique 1kb fragment, which was suitable for rapid sequencing.

The 1kb (AC)n repeat sequence of cosmid B9 was excised from a 1% low melting agarose gel and purified using the gene clean method (see Chapter 2.14.2). After confirmation that the isolated fragment was the one containing the (AC)n repeat by running 1µl on a 2% agarose gel, Southern blotting and hybridisation with a poly (AC)n repeat sequence (Pharmacia), it was ligated into a de-phosphorylated pUC9 vector. The ligation mixture was precipitated and transformed into *E.coli* by electroporation. See Chapter 2.15 for methods. Transformants were streaked out onto LB agar plates containing 50µg\ml ampicillin and 0.02% x-gal and

grown overnight at 37°C. White colonies, containing the insert, were re-streaked and after overnight incubation, grown in LB containing 50µg\ml ampicillin at 37°C. After the plasmid DNA was purified and cut with *Eco*RI to release the insert, the digest was run on a 1.5% low melting agarose gel, the 1kb fragment excised and gene-cleaned.

Sequencing of the 1kb fragment containing the (AC)n repeat was carried out using dye terminator cycle sequencing (see Chapter 2.16.3). Subsequently, a primer set capable of amplifying the repeat was constructed to use in PCR analysis and to generate a hybridisation probe.

4.2.2 Examination of the (AC)n repeat locus

The (AC)n repeat primer set was used on YAC A1S9T, YAC 2910 and YAC 2911 and on all of the PAC DNA samples in order to determine which YACs/ PACs contained the repeat.

The PCR conditions are as follows: the denaturing and extension temperature for the primer set is 94°C and 74°C respectively. The time for each step is 0.5 minute and the annealing temperature is 55°C.

The following work was carried out in collaboration with the Institute of Psychiatry in London, which I visited and worked at under the supervision of L. Hill for a total period of 2 weeks.

4.2.3 Testing the (AC)n repeat for degree of polymorphism and association studies

In the first instance one hundred children aged between 6 and 15 years were selected according to their cognitive ability (IQ): fifty of high IQ (130 to over 150, mean IQ 136) and 50 of average or mid IQ (around 100, mean IQ 105), to be used as controls. A replica study was performed with 46 subjects of high IQ (mean IQ 160) and 49 control subjects of mid IQ (around 100, mean IQ 105). The original and replica samples were combined to give an "original" sample. Further, additional samples from subjects of high and mid IQ (sample no. of 56 and 73, respectively) were obtained and designated as "replication" sample . The IQ was assessed using the Wechsler Intelligence Scale for Children, Revised (Wechsler,

1974). Only white Caucasians were chosen in order to attenuate possible problems of ethic stratification. All subjects were from the greater Cleveland metropolitan area, Ohio.

DNA samples were purified by L. Hill and derived from lymphoblastoid cell lines. DNA samples were genotyped individually and as pooled DNA (mid males, high males, mid females and high females). The sample collection was based on the hypothesis that high cognitive ability (g) would develop only if an individual had most of the positive and few of the negative QTL alleles for high g. By selecting the highest g extremes it was hoped that the power to detect QTLs that account for genetic variation throughout the distribution was increased. In contrast, at the lowest end of the distribution, it was hypothesised that idiosyncratic genetic and environmental factors could disrupt normal developmental processes, making it more difficult to find QTLs that accounted for normal variation, hence the control group were based on mid IQ values.

The (AC)n repeat forward primer was fluorescently labeled at the 5' end and PCR reactions carried out on the DNA samples with conditions as described above. One micolitre of each of the reactions was transferred to 96 well microtiter plates. To each well 0.45µl of an appropriate ladder (Genescan 500, Applied Biosystems) and 13µl of a denaturing agent (Hi-Di Formamide, Applied Biosystems) were added. The plates were sealed and after denaturation at 95°C for 5 minutes and cooling on ice, they were placed in an ABI3100 sequencer and results processed through 'Genotyper^{TM}' software. The sequencer separates a mix of DNA fragments according to their length. Bigger fragments move through the capillaries more slowly than smaller ones. In the detection cell of the system the fragments move through a laser beam which makes the fluorescent dye fluoresce. This is captured by a charge-coupled device camera, which converts the fluorescence into electric information, which is transferred to the computer workstation. Data are processed and displayed as an electropherogram: a single peak represents a single fragment. The software creates a plate record and specifies the sample type. Processed data is automatically extracted from the database, analysed and stored as sample files, which can be used exploiting the GeneScan analysis software. Comparisons of the genotypic or allele frequencies were performed using the Chi^2 test employing CLUMP and/or the 2x2 Chi^2 test

providing probability values (p).

The POWER of the investigation is such that an observed association (allele frequency difference) of about 0.15 should provide a p value of < 0.05 in a 2x2 for control allele frequencies between 0 and 0.2. For stronger association of 0.2 p-values should be < 0.02 for control frequencies between 0 and 0.5.

4.3 Results

4.3.1 Detection and isolation of a novel (AC)n repeat

A series of X-specific cosmids derived from YAC 2911 and found on preliminary screening to contain a (AC)n repeat (M. Coleman personal communication) were re-examined by *Eco*RI digestion (see figure 4.3.1a), Southern blotting followed by hybridisation with poly (dA-dC) x poly (dG-dT) (Pharmacia). This demonstrated that cosB9 contained an unique 1kb fragment possessing an (AC)n repeat and which was suitable for rapid sequencing.

The result of the hybridisation screen is shown in figure 4.3.1b.

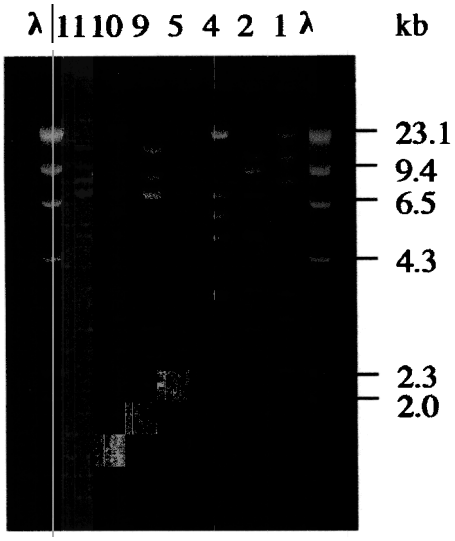


Figure 4.3.1a
*Eco*RI digest of cosmids

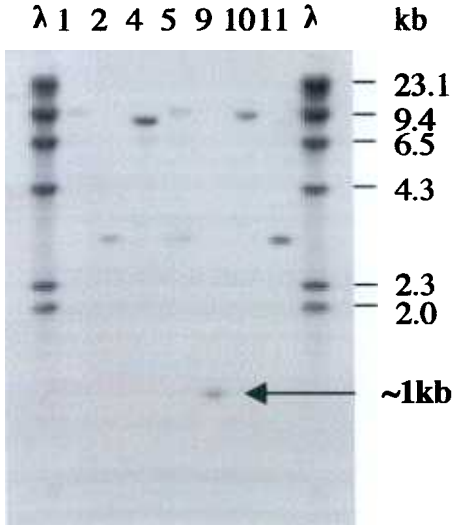


Figure 4.3.1b
Hybridisation of Southern blot obtained from *Eco*RI digest of cosmids, using (AC)n as a probe

The DNA fragment containing the (AC)n repeat was obtained and cloned as described in Chapter 4.2.1

4.3.2 Sequencing of a novel (AC)n repeat

Sequencing of the 1kb fragment containing the (AC)n repeat was carried out using dye terminator cycle sequencing (see Chapter 2.16.3 and figure 4.3.2a). The sequence revealed an imperfect (AC)n repeat (according to Weber, 1990; see Chapter 1.5.1) of 29 27 repeats with 7 5 interruptions, whic are wo bases long The two longest uninte upted AC)n repe ith he repeat sequence are 10 and 9 repeats long. Subsequently primer sets were constructed to use in PCR analysis and to generate a hybridisation probe. See figure 4.3.2b.

4.3.3 Examination of the (AC)n repeat locus

Using the (AC)n repeat primer set on DNA from YAC A1S9T, YAC 2910 and YAC 2911 and all the PACs showed that the repeat was contained in YACs A1S9T and 2911, but not in YAC 2910, as shown in figure 4.3.3a. PACs 2, 3 and 4 also include the (AC)n repeat (see figure 4.3.3b). Initially the YAC and PAC maps (see Chapter 3) were used to estimate a location for the (AC)n repeat. Towards the end of this study the (AC)n repeat was localised within the *UBE1* gene (see Chapter 6).

4.3.4 Testing the (AC)n repeat for degree of polymorphism and association studies

The genotyping results show that there are 5 different alleles observed for the (AC)n repeat; which are 282, 284, 286, 288 and 290 base pairs in size. Examples of genotypes obtained are shown in figure 4.3.4. The allele sized 288bp is the most frequent and the allele sized 286 is the second most frequent, in contrast, alleles of the size of 282, 284 and 290bp occur very rarely. There appears to be a trend in allele patterns between males and females. In males the number of the alleles 268 and 288 is higher in high IQ subjects than in the control subjects, whereas for females the number of the alleles is higher in the controls than in the high IQ subjects. The results from screening mid and high cognitive samples for males, females both, separately and combined are shown in table 4.1 and a chart corresponding to table 4.1 is illustrated in figure 4.3.5.

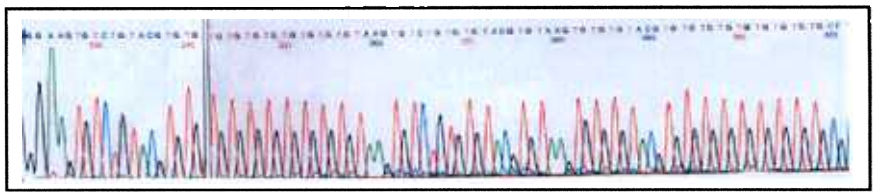


Figure 4.3.2a
(AC) repeat sequence obtained using the dye terminator cycle sequencing method

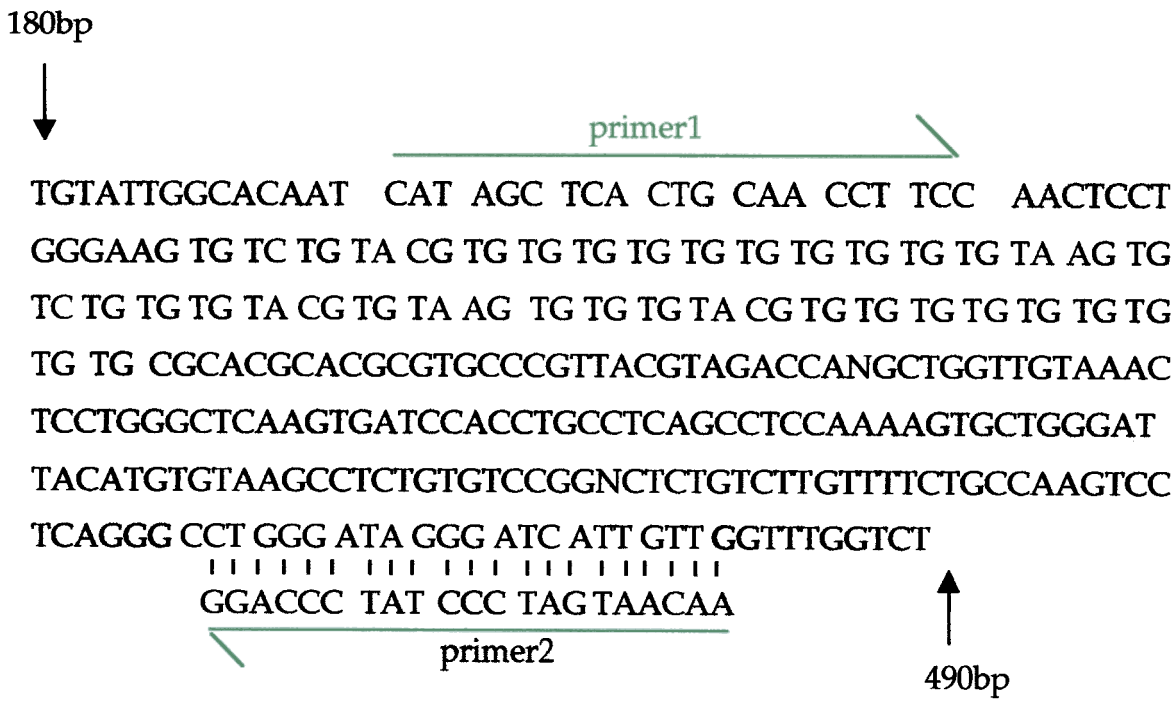


Figure 4.3.2b
The sequence containing the (AC) repeat and primer set designed from it. Letters in red highlight the (AC) repeat and the ones in blue show interruptions within the repeat sequence.

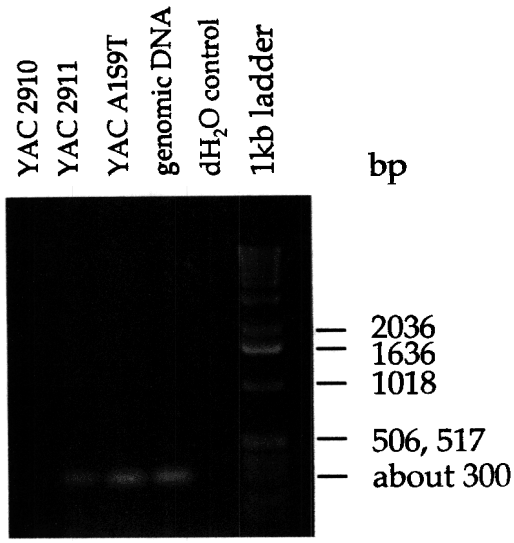


Figure 4.3.3a
(AC) repeat primers used on YAC DNA

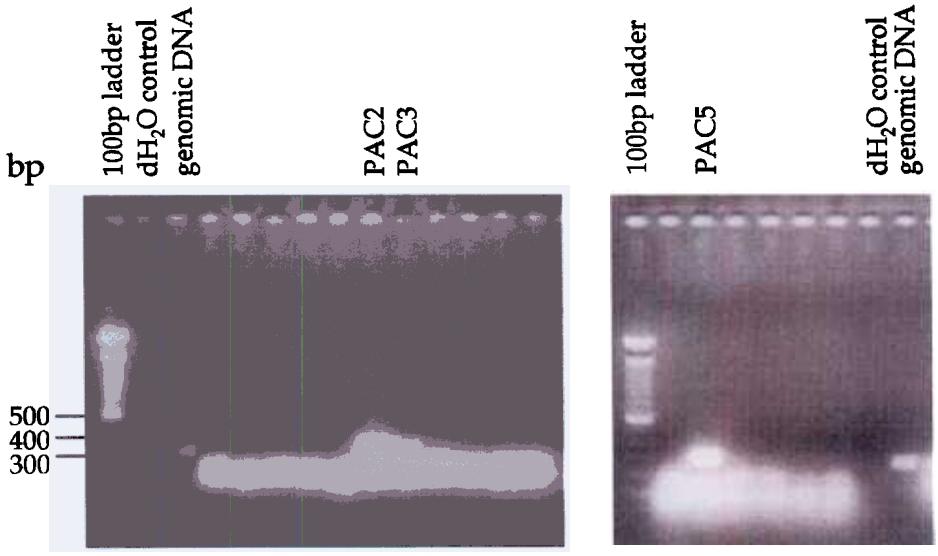


Figure 4.3.3b
(AC) repeat primers used on PAC DNA.
Only positives are identified.

Fig. 4.3.4 Examples of genotypes obtained using the CA repeat

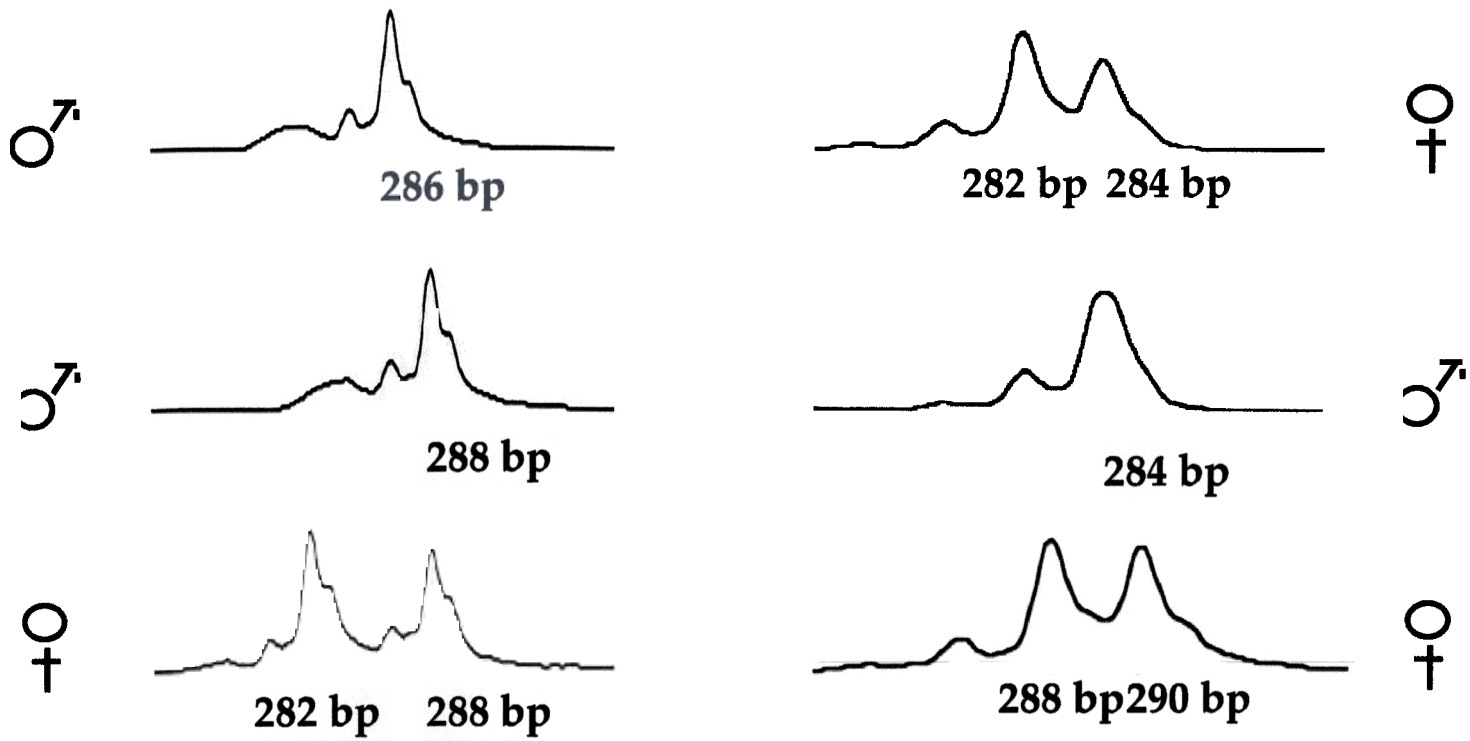


Table 4.1 Genotypes of males and females with mid and high cognitive ability

	Subjects		UBE1 AC Alleles (bp) 282	284	286	288	290	Totals	2x2 Chi square	p-value
Males	original	mid	1	0	14	33	0	48		
		high	0	1	27	40	0	68	1.33	0.24
	replica	mid	0	0	12	14	0	26		
		high	0	0	15	18	0	33	0.00	0.96
	Total	**mid**	**1**	**0**	**26**	**47**	**0**	**74**		
		high	**0**	**1**	**42**	**58**	**0**	**101**	**0.72**	**0.40**
Females	original	mid	1	0	32	69	0	102		
		high	0	0	14	41	1	56	0.66	0.41
	replica	mid	0	0	34	60	0	94		
		high	0	0	14	32	0	46	0.45	0.50
	Total	**mid**	**1**	**0**	**66**	**129**	**0**	**196**		
		high	**0**	**0**	**28**	**73**	**1**	**102**	**1.15**	**0.28**
Totals	original	mid	2	0	46	102	0	150		
		high	0	1	41	81	1	124	0.20	0.66
	replica	mid	0	0	46	74	0	120		
		high	0	0	29	50	0	79	0.05	0.82
	Total	**mid**	**2**	**0**	**92**	**176**	**0**	**270**		
		high	**0**	**1**	**70**	**131**	**1**	**203**	**0.01**	**0.90**

Results for Total Males v Total Females: $X^2 = 2.75$ and the p-value = 0.097

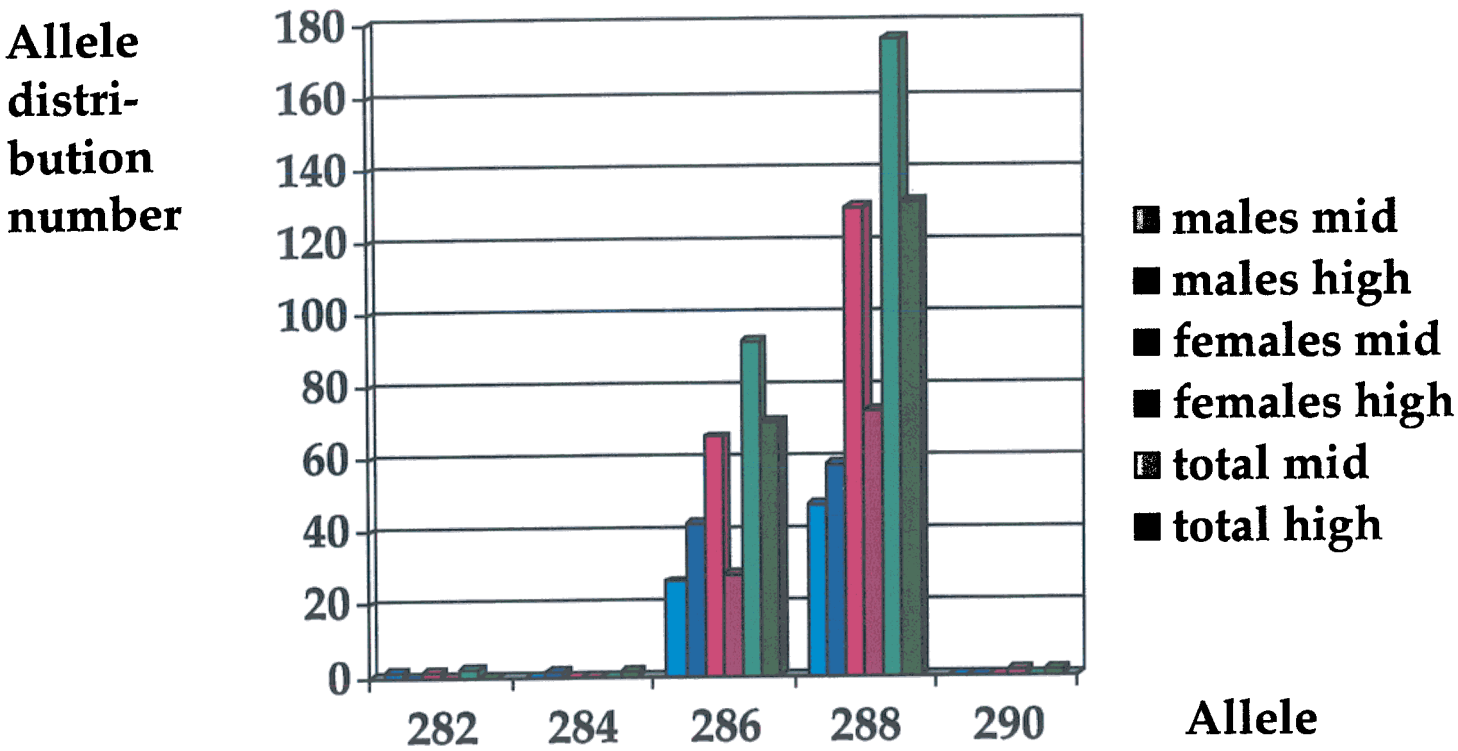


Figure 4.3.5 Chart corresponding to table 4.1

4.4 Discussion

A simple sequence repeat containing a consensus contiguous motif of at least $(AC)_{10}$ (the single base interruptions could be due to sequencing errors), variability in which confers a polymorphism comprising two alleles at high frequency and three rare alleles, has been isolated. According to sequence analysis carried out as described in Chapter 6, this marker is located within the *IBE1* gene, which lies on the short arm of the human X chromosome and is known to escape inactivation in humans. Therefore this marker is of interest in the analysis of potentially significant loci contributing to the sex differences in brain and behavioural development. Association studies for the genomic region embracing the *UBE1* locus in populations of high- and mid-range cognitive ability (g) have been carried out in collaboration with the Social, Genetic and Development Psychiatry Research Centre, Institute of Psychiatry in London.

No significant differences in frequencies were detected in a preliminary examination through allele image analysis of the available pools. Individual genotyping for each, again, indicated no highly significant differences in allele distributions for males or females, or for both combined. However, a tendency can be seen towards allele 288. The POWER of the investigation is such that an observed association (allele frequency difference) of about 0.15 should provide a p value of < 0.05 in a 2x2 for "control" allele frequencies between 0 and 0.5.

It has been suggested (Plomin, personal communication) that the results of the genomic scan for QTLs associated with 'g' suggest no locus is likely to contribute more than 2% to the total variance and it will be extremely worthwhile to extend the genotyping on the additional samples now available to improve the power of detecting QTLs of small effect.

CHAPTER 5 X-Inactivation Studies by means of Replication Timing

5.1 Introduction

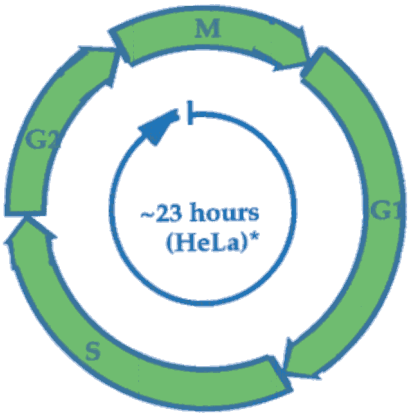
5.1.1 The Cell Cycle

The cell cycle is divided into four phases: G1, S, G2 and M phase. From the beginning of the G1 phase, in which cells are in their diploid state, to the end of the G2 phase, by which time chromosome doubling has occurred, is also called the interphase and comprises about 90% of the total cell cycle time. The M phase (division) consists of mitosis and cytokinesis (the splitting of the cell into two daughter cells). Mitosis reduces the chromosome complement to the diploid number in each daughter cell.

DNA replication in eukaryotic cells is confined to one part of the cell cycle, the S phase. The start of S phase is signalled by the activation of the first replicons. During S phase initiation events occur at other replicons. The control of S phase therefore involves two processes: release of the cell from G1 and initiation of replication at individual replicons in an ordered manner. G phases represent gaps in DNA synthesis (Alberts et al., 1989). A replicon is the basic unit of replication with 50-330kb of DNA replicating from a single origin in mammals (Edenberg and Huberman, 1975). Contiguous replicons have co-ordinated replication patterns and may be organised in temporal-spatial domains megabase or more, suggesting the existence of master control elements for replication timing (reviewed in Gartler et al., 1999).

A schematic presentation of the cell cycle is shown in figure 5.1 (see next page)

Fig. 5.1 The Cell Cycle



* : e.g. Wheatley, 1990;
Kozaki et al., 1993

5.1.2 Replication timing of active and inactive genes

Early observations (e.g. by Lima-de-Faria, 1959; Kajiwara and Mueller, 1964) suggested a relationship between chromosome condensation, late DNA replication and gene silencing. More recently a general hypothesis has emerged from molecular studies stating that expressed genes replicate early, silent genes replicate late and changes in expression status are accompanied by corresponding changes in the time of gene replication. One example are housekeeping genes, of which almost all replicate within the first half of S phase (Goldman et al., 1984; Hatton et al., 1988). Another, very prominent example is given by the X chromosomes which replicate highly asynchronously in eutherian female cells: the inactive X chromosome replicates toward the end of S-phase, whereas its active counterpart replicates much earlier (reviewed in Yeshaya et al., 1999).

This hypothesis holds true for most genes, but there are exceptions such as the *XIST* gene (Hansen et al., 1995), allele-specific replication in imprinted domains, for example different genes within the Prada-Willi/ Angelman syndrome region at 15q11-q13 (Knoll et al., 1994; Kawame et al., 1995; Sinnett 1996; White et al., 1996) and the FMR1 locus (Hansen et al., 1997; Yeshaya et al., 1999), which, when silenced, leads to fragile X syndrome, the second most frequent cause of inherited mental retardation in man (reviewed in Yeshaya et al., 1999). Mutated FMR1 alleles replicate later than wild-type alleles on both, the active and inactive X chromosome. In cases where allelic differences exist (for example in genomic imprinting or X chromosome inactivation) the expressed allele replicates before

the repressed allele (Hansen et al., 1995; Kawame et al., 1995; Hansen et al., 1996). A study by Hansen and co-workers (1996) supports a model for regulation of X-inactivated genes. It involves two control factors: one is that large chromosomal domains are placed into a transcriptionally non permissive state by late replication and the second is that transcription is blocked at the local level by promoter methylation

5.1.3 Changes in expression status of genes

The question arises whether the individual gene, or the entire multi-replicon domain is involved when there is a shift of an expressed, early replicating to an silent, late replicating form, or vice versa. There is some evidence that suggests that the multi-replicon domain is dominant. For example Hatton et al. (1988) showed that in the case of immunoglobulin κ light chain sequences non-transcribed genes replicate early because they are in proximity to transcribed ones. But since reports are controversial (Hansen et al., 1995; Torchia et al., 1995; Xiong et al., 1998) Gartler et al. (1999) investigated this issue using the *XIST* gene. The authors chose this gene because of two main reasons: firstly, it is silent on the active, early replicating X chromosome and expressed on the inactive, late replicating homologue (Brown et al., 1991) and secondly, because of its relatively close proximity to genes that are subject to X inactivation. The authors found that on the active X chromosome the silent *XIST* gene is replicated underlining the multi-replicon domain dominant hypothesis; however, the unique status of the *XIST* gene raises the question whether, or not, similar situations may be obtained at other loci.

5.1.4 Experimental approaches available to investigate replication domains in the UBE1 region

Different cell lines

Fibroblasts and lymphocytes are normal, diploid cells and therefore when used in tissue culture experiments represent an as close simulation of the correct *in vivo* situation as possible. The only drawback is that they only replicate a few times before dying and therefore are difficult to culture and use.

HeLa cells on the other hand, being cancerous (cervical carcinoma) cells replicate

continuously and are therefore easier to use. They grow attached to the flask surface and dead cells detach and float in the medium. Most HeLa sublines are of common origin, but differ in the sense that they possess variable numbers of chromosomes or parts of chromosomes (Ghosh and Ghosh, 1975). For example Mamaeva et al. (1986) detected that each of the 3 sublines they examined contained an extra copy of chromosome 1 and 5 and some cells had an additional copy of chromosomes 7, 9, 12, 14, 16 and 17. This fact also has to be taken into account when working with HeLa cells, which originated in 1952 from a woman called Henrietta Lacks and have been used ever since. Nevertheless, its robust growing properties and the fact that additional copies of X chromosomes will dilute rather than obscure any replication differentials made it the cell line of choice for the initial experiments. If promising results were obtained it was planned to concentrate subsequent work on diploid cells.

Synchronisation of cells

To study replication timing it is necessary to synchronise cells; this is usually accomplished by blocking and then releasing DNA synthesis or cell division. There are two main ways of blocking DNA synthesis: one is to apply a thymidine or 5-bromo-2-deoxyuridine (BrdU, which is an artificial analogue of thymidine) block, i.e. provide the growing cells with high concentrations of thymidine or BrdU and the other is to treat cells with colcemid. In this study BrdU could not be used as a blocking agent, as it was used in the actual replication timing experiment and using it for synchronisation could have lead to false results. Instead colcemid, a derivative of colchicine which is an anticancer chemotherapeutic agent, was used to trap cells in metaphase. Colcemid has been widely used in cytogenetics. The drug inhibits polymerisation of microtubules by binding to tubulin dimers. Since depolymerisation is unaffected, the mitotic spindle rapidly dissociates or is not formed and cycling cells accumulate in a pro-metaphase-like state (reviewed by Jha et al., 1994) from which they can later be released.

Labelling and separation of early and late replicating DNA

Growing cells in medium containing BrdU is a commonly used method to assess replication timing of DNA. BrdU as a thymidine analogue incorporates into the

replicating DNA. It may be added to the growth medium before cell-synchronisation to detect, for example, late replicating chromosome domains. Late replicating regions can be detected by binding BrdU antibodies to it, fluorescently labeling them and finally using flow cytometry to separate the BrdU incorporated DNA from the non-incorporated DNA (method from Carayon and Bord, 1992).

If cells are first synchronised, BrdU can be added following release at a specific time in the cell cycle, so that either early or late replicating DNA is BrdU labelled. Since BrdU is heavier than its thymidine analogue, BrdU incorporated DNA can be separated from DNA not containing BrdU by CsCl density gradient centrifugation. This approach using HeLa cells was used in this study.

It should be mentioned that the determination of replication timing can be carried out by fluorescence *in situ* hybridisation where no pre-labelling or synchronisation of cells is necessary and a relatively small number of cells are required. This method was first described by Selig et al. (1992), but requires access to specialist equipment and expertise.

Replication time experiments can show if DNA replication times of specifi genes are consistent with their inactivation status.

This Chapter describes the isolation of early and late replicating DNA from HeLa cells by BrdU incorporation at various times of the cell cycle followed by separation of BrdU incorporated from non-incorporated DNA by CsCl centrifugation described in outline below.

Early and late replicating DNA were digested by restriction enzymes, run on an agarose gel and Southern blotted. Hybridisation experiments using primers amplifying DNA segments of known inactivation status were carried out to confirm that replication times coincided with activation status'. Further, probes generated from primer sets DXS8237E, UBE1 and PCTK1 were used to examine their activation status.

Before carrying out the above described study, experimental conditions were established as described below.

5.1.5 Practical approach

The following preliminary investigations were undertaken:

i. Growth of HeLa cells in medium containing BrdU to estimate the time course (and extent) of any toxic effects.

ii. Determination of conditions for the preparation of karyotypes used to test synchronisation efficiency of the cells

iii. Synchronisation of HeLa cells using colcemid to establish suitable concentrations (avoiding excessive toxicity) and the percentage of synchronised cells obtained

iv. Growth curve of HeLa cells to establish the cell cycle parameters and therefore to determine times to add BrdU.

v. CsCl centrifugation to confirm conditions to separate light and heavy into two distinctive bands.

vi. Pre-hybridisation experiments to optimise conditions with regard to loading quantities of digested HeLa DNA and to determine which restriction endonuclease gives optimal fragment patterns when hybridised using NDE1, DXS8237E, UBE1, PCTK1 and MIC2 probes.

5.2 Materials and Methods

5.2.1. Growing mammalian cells

HeLa cells were grown in RPMI-1640 medium (Gibco BRL), with non essential amino acids, 0.2% (v/v) sodium bicarbonate (Imperial Laboratories), 100µg/ml penicillin and 100µg/ml streptomycin (Sigma). This was supplemented with 10% (v/v) foetal calf serum (Biological Industries) and 2mM glutamine (Gibco BRL). Cells were cultured under 5% CO_2 at 37°C. HeLa cells which grow on the surface of the flask were removed from the wall by treatment with trypsin, 0.25% (v/v) and 0.02% (w/v) versene [Na_2EDTA] in PBS, using the minimum necessary time and amount (0.5ml to a 25cm^3 flask; 2ml to a 175cm^3 flask)

were detached by gentle tapping and an equal volume of medium was added to eliminate further trypsin activity. Cells were pelleted in a bench centrifuge at 900rpm for 7 minutes, then washed in PBS before further treatment.

5.2.2 Preparation of mammalian genomic DNA

1 Resuspend cell pellet in 2.5ml STE containing 0.5% SDS and 0.2mg/ml proteinase K (Promega) and incubate at 65°C for 2 hours.

2. Add 300µl of phenol/chlorophorm mix (containing 50% ethanol), mix for 1 minute and centrifuge at 3500 rpm for 40 minutes.

3. Decant supernatant into a clean Eppendorf tube and repeat step 2.

4. Add 0.75 volumes of isopropanol, mix for 1 minute and spin as before.

5. Wash the pellet in 70% ice cold ethanol at room temperature for 10 minutes and spin as before.

6. Dry pellet at room temperature for 30 minutes and resuspend in 300µl of dH_2O.

DNA concentrations were estimated by running a defined volume of DNA on a agarose gel and comparing its brightness to that of a band of similar size of λ phage DNA/ *Hind* III marker using the documentation and analysis system Alpha Imager™ 2200. The bands relate as follows if 100ng of total marker are run

band size (bp)	weight (ng)
23 130	43
9 416	19.5
6 557	13.5
4 361	9
2 322	4.8
2 027	4.2

5.2.3 Synchronisation of mammalian cells

1. To almost confluent cells growing in a 175cm^3 flask containing 25ml of medium add 130µl colcemid (10µg/ml) (Sigma), ie. final concentration of 52ng/ml

2. Grow at 37°C for 4 hours under 5% CO_2.

3. Remove supernatant containing the metaphase arrested cells and spin down in a bench centrifuge at 900 rpm for 7 minutes.

4. Wash resulting pellet with PBS and spin as before.

5. Resuspend in appropriate solution for further experiment.

5.2.4. Chromosome preparations

1. Resuspend synchronised cells in 1ml of 37°C warm 0.56% KCl and incubate at 37°C for 50 minutes.

2. Add 1ml of fixative (3:1 Analar methanol:glacial acidic acid) drop by drop, to prevent the cells sticking together.

3. Spin for 7 minutes at 900 rpm, resuspend in 100µl fixative as before and incubate at 4°C for 16 hours.

4. From about 50 cm height drop one to two drops on a glasslide, spread by blowing in one direction over the slide and air dry.

5.2.5 Staining of chromosome preparations

1. Immerse slides in Hoechst 33258 (2mg/ml of 2xSSC) for 30 minutes.

2. Rise slides and then cover them in 2 x SSC.

3. Place slides under a long-wave UV light for 60 minutes.

4. Stain the slides using Giemsa dye (diluted to 10% in phosphate buffer pH6.8) for 3 minutes. If necessary repeat staining.

5. Rinse the slides in phosphate buffer and air dry.

6. View under microscope (Leitz Dialux 20EB) using a Plan-Neofluor 63/1.25 OIL lens

5.2.6 Base analogue incorporation into mammalian cells

1. Release synchronised cells in $75cm^3$ flask containing 15ml medium at 37°C for 6 hours (when labelling early replicating DNA) or 14 hours (when labelling late replicating DNA) under 5% CO_2.

2. Add 300µl of Bromo-deoxy-Uridine (10mg/ml) to give a final concentration of 0.2 mg/ml and continue to grow for 8 hours as in step 1

3. Harvest cells as described in 5.2.1 and extract DNA as described in 5.2.2

From the time of addition of BrdU cells, slides and DNA should be protected from direct light to avoid nicking of the BrdU-incorporated DNA

5.2.7 Cesium Cloride (CsCl) centrifugation

1. After total DNA extraction make up CsCl gradient to 4.4M using a 5ml Beckman centrifuge tube as follows: 2.2g CsCl, 250µl ethidium bromide (10mg/ml), 1500µl DNA (from 5 BrdU experiments) and 450µl of water to make up a volume of 3ml.

2. Fill tube with paraffin, heat-seal and place in Optima™ Ultracentrifuge (Beckman) using a TLA 100.4 rotor.

3. Spin at 80,000rpm, 15°C for 48 hours.

4. Purify DNA according to Sambrook, Fritsch and Maniatis

5.2.8 Probes used in this investigation

Apart from the UBE1, PCTK1 and DSX8237E probes mentioned in earlier Chapters, probes for *MIC2*, a gene located in the pseudoautosomal region of the X chromosome and escaping inactivation and NDE1 (Norrie's disease exon 1), the first exon of a gene located on the short arm of the X chromosome, which is inactivated normally, were used as controls to establish patterns for genes escaping inactivation and ones being subject to inactivation, respectively.

Primer sequences 5' to 3' end are as follows:

MIC2 primer 1: ACC CAG TGC TGG GGA TGA CT

MIC2 primer 2: TCT CCA TGT CCA CCT CCC CT

Ta = 55°C, product size: 360bp

NDE1 PCR products (350bp) were provided by B. Suarez-Merino. They were run on a 1% agarose gel and purified as described in Chapter 2.14.4

5.3 Results

5.3.1 Toxicity of BrdU to HeLa cells

To two 175cm^3 flasks containing ~ 50% confluent cells, 0.5ml BrdU (10mg/ml) was added. One flask was left to grow for 24, the other for 48 hours. Microscopy showed most cells had died at 48 hours. A preparation of total DNA showed that the yield from the 24hr BrdU incorporation was 10x higher than that from the 48hr BrdU incorporation indicating that severe toxicity developed between 24 and 48 hours of growth. It was concluded that although the base analogue is toxic to HeLa cells, growth for 8 hours following release from colcemid block would not result in unacceptable levels of toxicity. This was confirmed in later experiments.

5.3.2 Preparation of chromosome slides

Experiments were carried out at 37°C and conditions varied as illustrated in table 5.1.

Table 5.1 Experiments establishing the optimum KCl concentration and incubation time to prepare chromosome slides

experiment number	KCl concentration	incubation time
1a	0.50%	20mins
1b	0.53%	"
1c	0.50%	30mins
1d	0.53%	"
1e	0.50%	40mins
1f	0.53%	"
2a	0.53%	50mins
2b	0.56%	"

The optimal method for bursting cells and obtaining the best chromosome spreads was established to be as follows:

- incubate cells for 50 minutes in 0.53% KCl at 37°C

- cover the slide with fix, before dropping cells onto it from a height of ~ 50cm

blow over the slide after dropping the cells onto it in order to spread out chromosomes

5.3.3 Synchronisation of HeLa cells using colcemid

Various experiments were carried out to estimate the optimum condition to synchronise HeLa cells. The length of incubation time was varied, whilst the amount of colcemid was kept constant at a final concentration of 52ng/ml. Colcemid was added to almost confluent flasks of cells and incubated for 1, 2, 4, 6 and 8 hours. Microscopy showed that using an incubation time of 6 and 8 hours resulted in cell death. Cells were synchronised after either treating a flask of almost confluent cells with colcemid for 4 hours or treating a flask of almost confluent cells with colcemid for 2 hours, taking off the supernatant containing the in metaphase arrested cells, adding fresh medium plus colcemid (52ng/ml) and leaving to grow for a further two hours. This procedure was repeated once more so that in total 3 supernatants containing metaphase arrested cells were

collected over a period of 6 hours. Once the method was optimised 51 - 76% of HeLa cells were arrested in metaphase. The DNA yield per 175cm^2 flask for both cases was 80µgDNA\ flask.

5.3.4 Growth curves of synchronised HeLa cells

In this experiment the latter method of synchronisation was used. Two growth curves were obtained following colcemid treatment: one reflecting HeLa cell growth in medium without BrdU and the other describing HeLa cell growth in the presence of BrdU. The experiments were carried out in 25cm^3 flasks. Synchronised cells took six hours to attach to the bottom of the flask Therefore in order to allow most cells to attach, BrdU was added after 4 hours of inoculating the medium

Two experiments for each curve were carried out. In the first cells were counted every four hours up to 32 hours and in the second cells were counted at 12, 24, 30 and then every six hours up to 60 hours. Tables of the counts are shown in the appendix. "Raw data" growth curves are shown in figure 5.3.1a and growth curves in which the two experimental growth curves have been scaled to coincide with numbers with numbers observed in experiment 1 are illustrated in figure 5.3.1b.
Microscopical observations at 12, 24, 48 and 54 hours regarding the two growth curves were such that there was no visible difference between the cells growing in medium containing BrdU to the ones growing without BrdU up to 24 hours. At 48 hours most of the cells growing in BrdU appeared dead by microscopy and at 54 hours almost all of them were, thus confirming the results of the experiment described in 5.3.1. Again this suggests that short exposures, 4-8 hours, in BrdU are not excessively toxic.

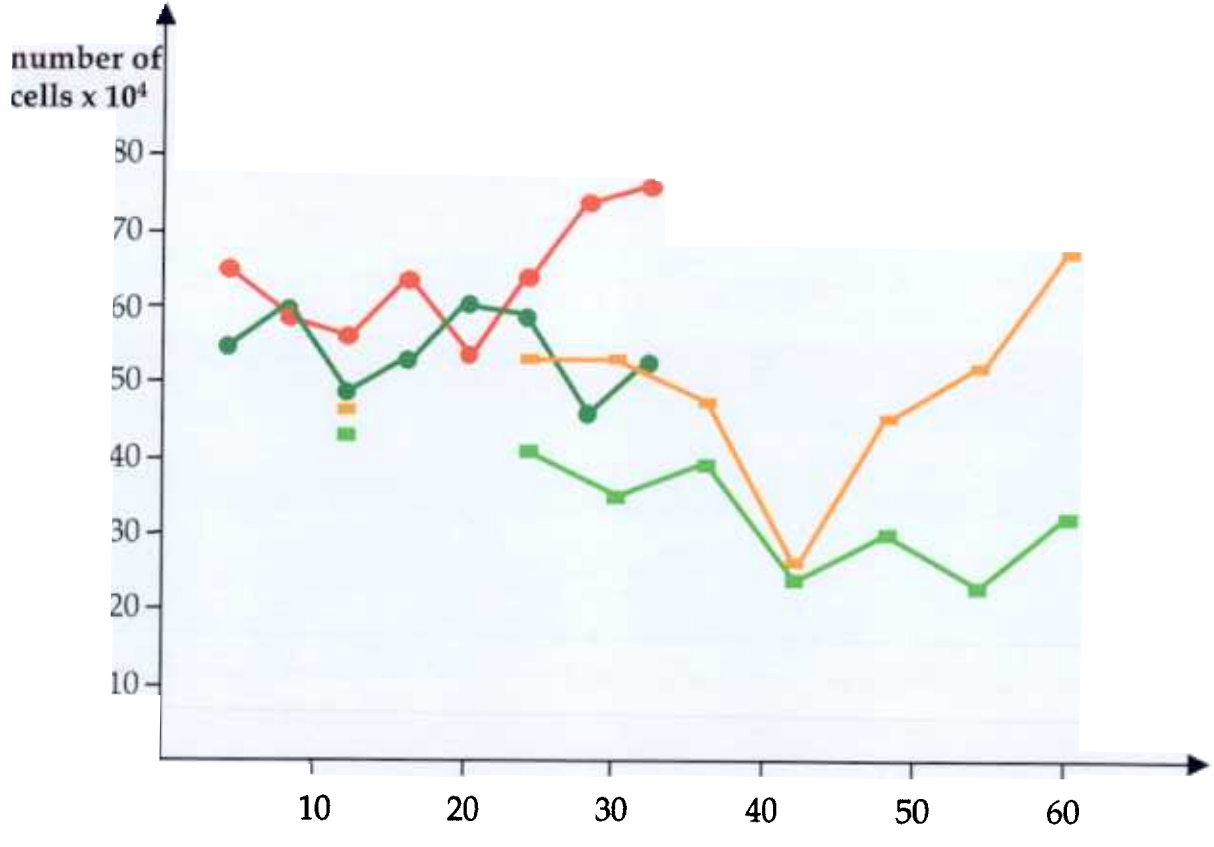
<u>Cells grown in medium without BrdU</u>

It appears that after 4 hours of growth following synchronisation about 66% of the cells survived. This value can not be taken as absolute, since the supernatant is likely to contain dead cells, as well as synchronised ones. The initial loss of viability observed from the growth curve might be explained by a period of recovery from the colcemid block and the dilution of stimulatory growth factors as a result of the manipulations undertaken. Further, some of the cells being

Figure 5.3.1a Raw data growth curves of HeLa cells without and with BrdU

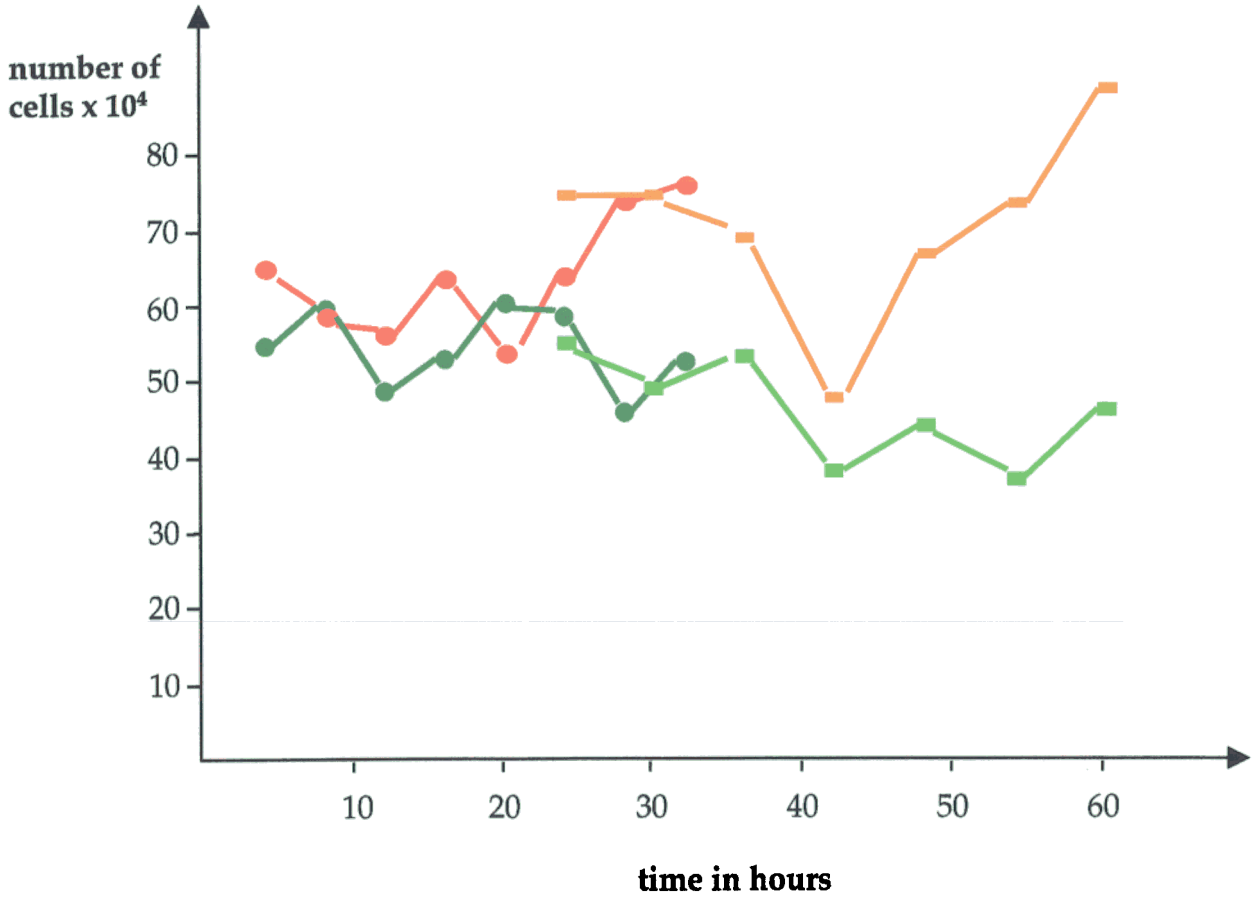
time (hrs)	without BrdU		with BrdU	
	exp. 1 cells x10⁴	exp. 2 cells x10⁴	exp. 1 cells x10⁴	exp. 2 cells x10⁴
0	**98**	51	**98**	51
4	**65**		**54**	
8	**59**		**59**	
12	**57**	46	**48**	43
16	**64**		**52**	
20	**54**		**60**	
24	**65**	53	**58**	41
28	**74**		**45**	
30		53		35
32	**76**		**52**	
36		47		39
42		26		24
48		45		30
54		52		23
60		67		32

Note: the average of cells counted has been calculated and used to determine growth curves



The graphs are colour coded to allow comparison with table data

Figure 5.3.1b Combined growth curves of HeLa cells without and with BrdU



The colour coding is consistent with that in figure 5.3.1a

kept at 4°C for over 6 hours means that some cells start to grow and others are dead when released from synchronisation. From the estimated beginning of cell division (see figure 5.3.1b) the first replication of the cells starts around 22 hours and finishes at about 30 hours. Therefore, it was estimated that DNA of active genes will replicate between 6 and 14 hours and DNA from inactive genes will replicate between 14 and 22 hours.

5.3.5 Establishing conditions for CsCl centrifugation

Experiments were first set up using total HeLa DNA and as illustrated in table 5.2. In all experiments centrifugation was carried out at 80,000 rpm and 15°C.

Table 5.2 Experiments using total HeLa DNA establishing the optimum conditions for CsCl centrifugation

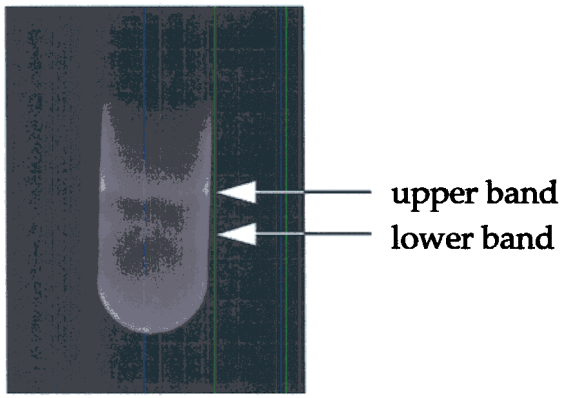
experiment number	CsCl concentration	amount of DNA	centrifugation time	result obtained
1a	6.5M	300µg	68hrs	-
1b	5.9M	"	"	-
1c	4.9M	"	"	band in upper third of tube
2a	4.9M	300µg	48hrs	+ (as above)
2b	"	150µg	"	+ (as above)
2c	4.1M	150µg	"	band in lower third of tube
3a	4.4M	75µg	36hrs	band in
3b	"	"	24hrs	middle
3c	"	10µg	24hrs	of tube

Table 5.2 shows that the optimal conditions are: 4.4M CsCl, 80,000rpm @ 15°C for 24 hours. As little as 10 µg of DNA can be visualised and about 50% of the DNA was recovered from the CsCl gradient.

For BrdU incorporated DNA, 300µg of DNA was loaded onto the gradient, since this would allow the detection of BrdU incorporated DNA representing as little as 10% total, resulting in a band containing about 30µg, which would be visible on a CsCl gradient (see above). Conditions were set at 4.4M CsCl 80,000rpm @

15°C for 39 hours. After centrifugation two bands were visible (see figure 5.3.2), the ratio of the top to the lower one (DNA containing BrdU) being about 3/2 to 3/1. The centrifugation time was increased to 48 hours, since the lower band appeared diffuse.

Fig. 5.3.2 Two distinguishable bands obtained from CsCl centrifugation of BrdU incorporated DNA



After DNA extraction from the CsCl gradient the ratio of the amounts of DNA was confirmed and about 100μg and 65μg for the top and lower band, respectively. This suggests the total DNA recovery to be over 50%.

5.3.6 Pre-hybridisation experiments

To establish the amount of DNA required for the hybridisation experiments, a conventional 1.0% agarose gel containing 1, 2 and 4μg of total HeLa cell DNA, digested to completion using the restriction enzyme *Eco*RI, was run. This was followed by Southern blotting and hybridisation of the resulting filters with the UBE1, PCTK1 and DSX8237E probes. The results show that 4μg of DNA are required to give analysable patterns (see figure 5.3.3).

In order to establish which restriction enzyme would provide the most suitable pattern, total HeLa cell DNA was digested using the following enzymes: *Bgl*I, *Bst*XI, *Eco*RI, *Hind*III, *Nde*I, *Pst*I, *Sfi*I, *Xho*I and the subsequently obtained filter hybridised with the UBE1, DSX8237E, NDE1, PCTK1 and MIC2 probes. The results showed that *Bgl*I and *Bst*XI were the best enzymes to use. An example of such a hybridisation experiment is given in figure 5.3.4.

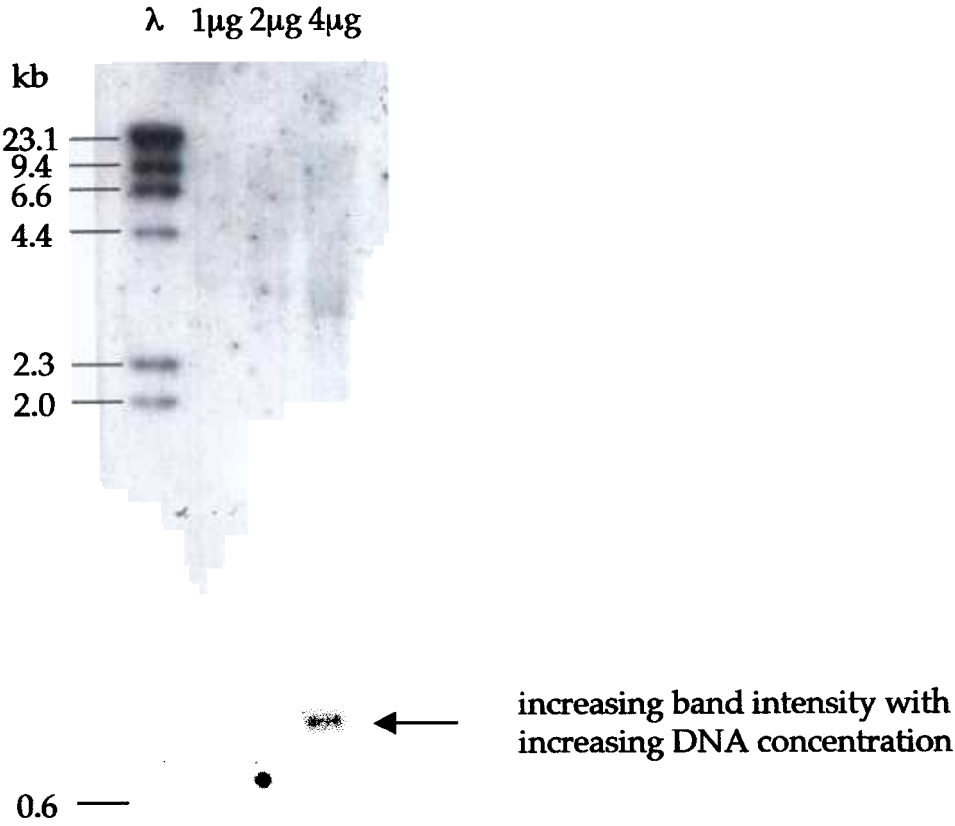


Fig. 5.3.3
Different concentrations of total HeLa DNA digested using *Eco*RI and hybridising the Southern blot with the MIC2 probe. 4μg gives a clear band.

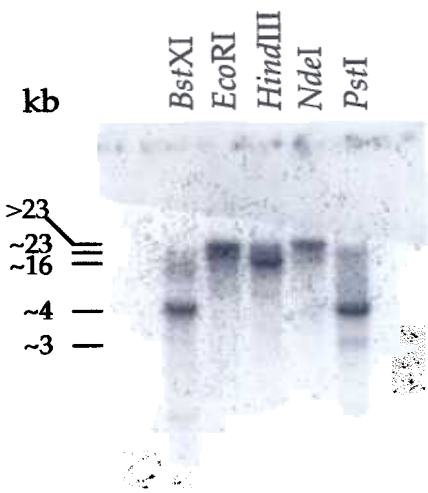


Fig. 5.3.4
Total HeLa DNA (4μg) digested using various enzymes and hybridising the Southern blot with the NDE1 probe

5.3.7 Full scale experiments

Sequential experiments were carried out as illustrated in figure 5.3.5 until sufficient material for subsequent work was available.

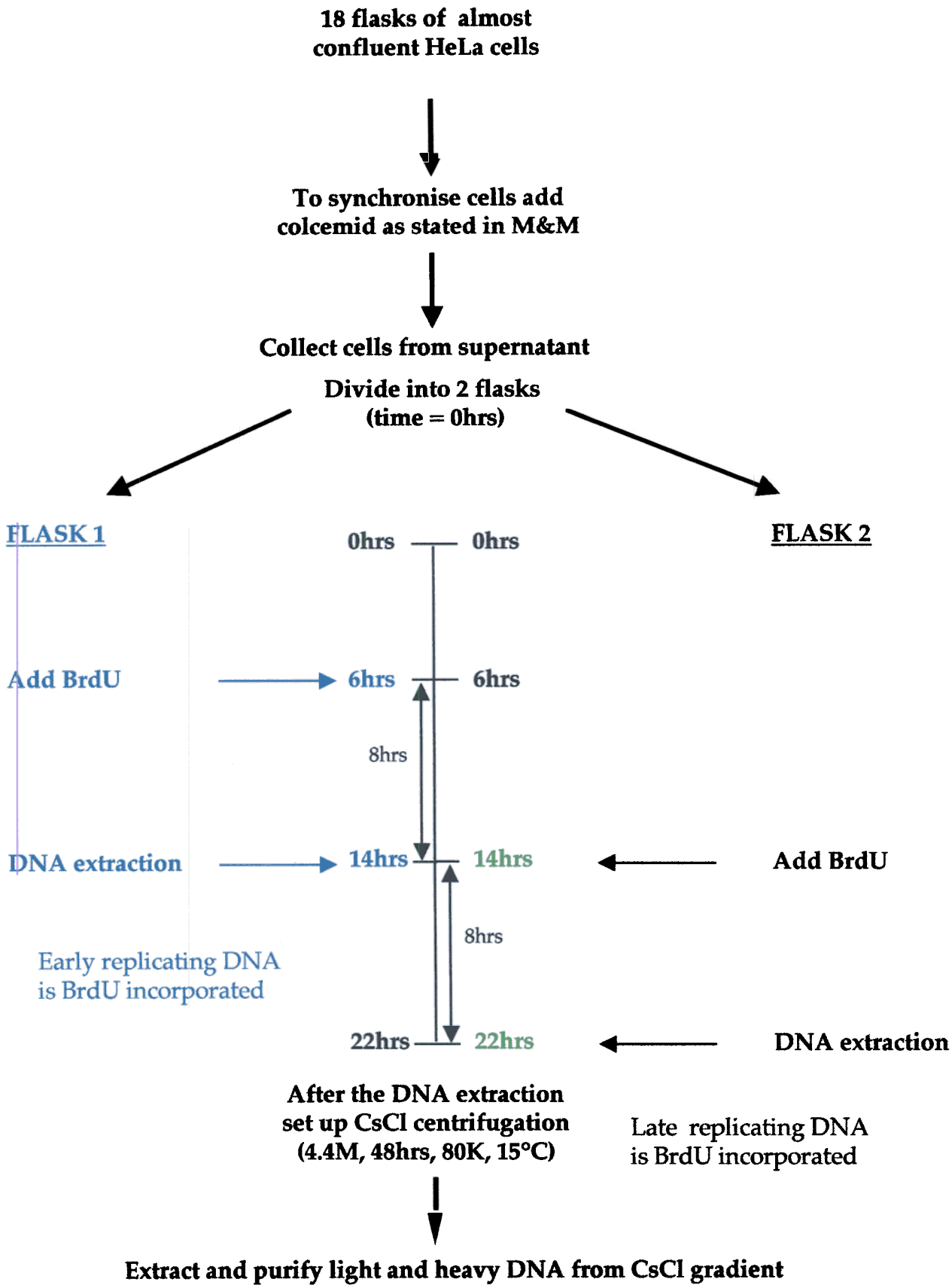
After synchronisation of the HeLa cells employing colcemid as describe in M&M, the cells were counted and chromosome slides prepared in order to determine the percentage of them in metaphase. Examples of HeLa cells in metaphase are shown in figure 5.3.6.

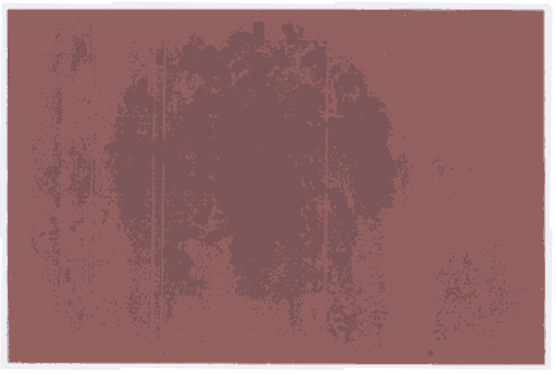
Following DNA extraction, the DNA concentration was determined and when a sufficient amount was accumulated it was used for CsCl centrifugation. Typical photographs illustrating separation of BrdU incorporated (replicated) DNA from the remainder are shown in figure 5.3.7. Yields of extracted DNA varied; but improvements to the protocol and experience in the manipulations enabled sufficient material to be obtained for subsequent analysis. Complete results are shown in table A1 of the appendix. Details of the experiments providing material for the inactivation analysis are given in table 5.3.

Table 5.3 Data obtained for DNA used in CsCl centrifugation

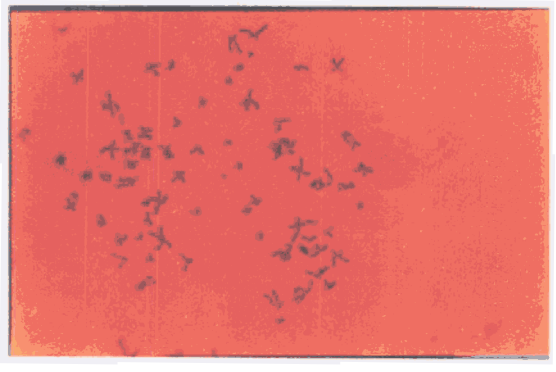
experiment number (see tab. A2)	Percentage of cells in metaphase	DNA from 6-14hr BrdU incorporation	DNA from 14-22hr BrdU incorporation
16	65	49μg (i)	55μg (ii)
18	57	30μg (i)	39μg (ii)
19	76	47μg (i)	39μg (ii)
20	62	31μg (i)	37μg (ii)
21	69	37μg (i)	35μg (ii)
DNA obtained after CsCl centri.		**U: 64μg L: 6μg**	**U: 85μg L: 10μg**
22	55	16μg (iii)	26μg (iv)
23	63	21μg (iii)	23μg (iv)
24	65	21μg (iii)	37μg (iv)
25	78	29μg (iii)	25μg (iv)
DNA obtained after CsCl centri.		**U: 34μg L: 6μg**	**U: 52μg L: 8μg**

Figure 5.3.5 BrdU incorporation experiment

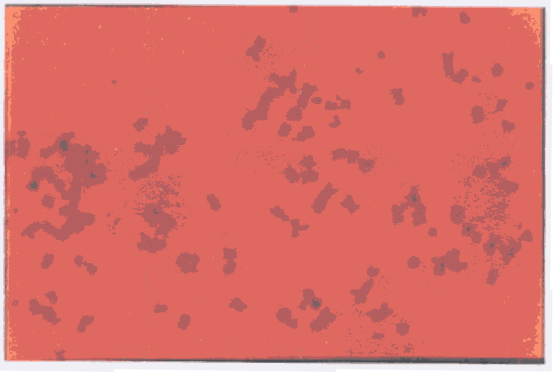




Overall magnification 1250x



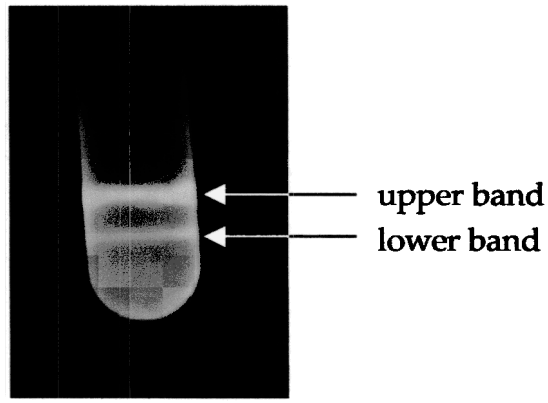
Overall magnification 1250x



Overall magnification 2500x, under oil

Fig. 5.3.6
Examples of chromosomes in metaphase
Cells with distinguishable metaphase chromosomes were counted as a proportion of total nuclei to indicate efficiency of metaphase entrapment.

CsCl 6-14hr



CsCl 14-22hr

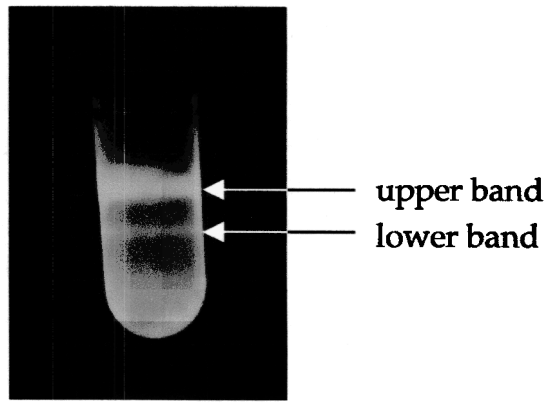
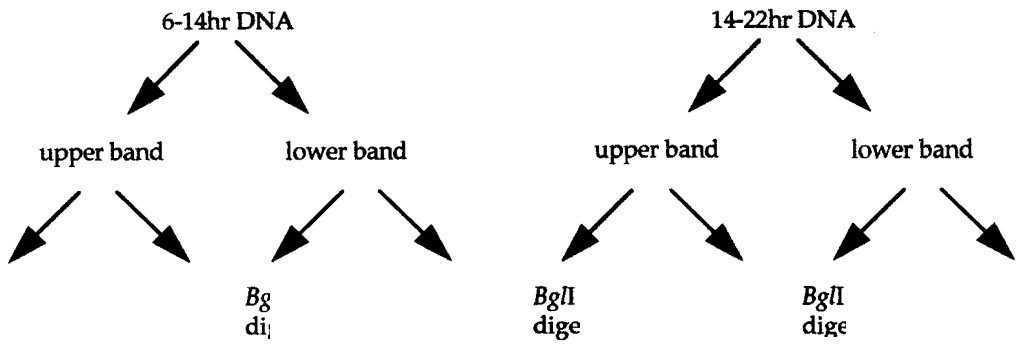


Fig. 5.3.7
Heavy and light DNA separation of DNA obtained from BrdU incorporation at 6-14hrs and 14-22hrs

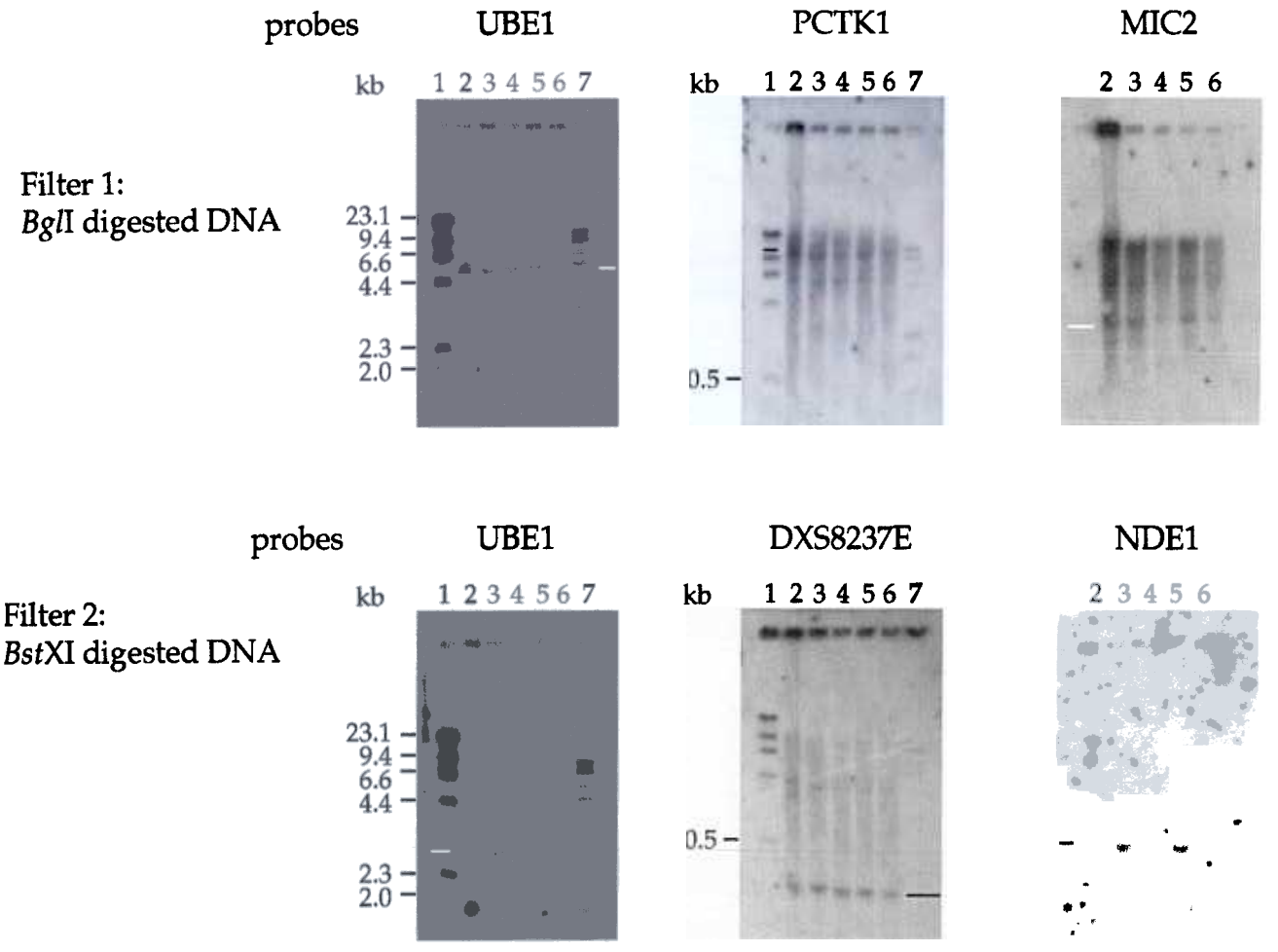
All of the DNA marked with the same number (indicated in brackets) was combined and analysed by CsCl centrifugation
U: upper band of CsCl centrifugation, i.e. non BrdU incorporated DNA
L: lower band of CsCl centrifugation, i.e. BrdU incorporated DNA.

Both DNA amounts obtained from the upper band of the 6-14hr experiment were combined, as were both DNA amounts obtained from the lower band of the 6-14hr experiment. The upper and lower bands of the 14-22hr experiment were treated in the same way. Each of the 4 different, combined DNAs were divided into 2 different tubes and digested with *Bgl*I and *Bst*XI, respectively, as illustrated below:



The digests were run on a 1% agarose gel. The amount of DNA present in the lanes was estimated by comparing the intensity of DNA staining in each of the tracks using an Alpha Imager™ 2200.

The gels were Southern blotted and hybridised with the NDE1, DXS8237E, MIC2, PCTK1 and UBE1 probes. Both filters were first hybridised with the UBE1 probe. The filter containing the *Bgl*I digested DNA was next (6 weeks later) hybridised using the PCTK1 probe and then (1 month later) using the MIC2 probe. The filter containing the *Bst*XI digested DNA was next hybridised using the probe DXS8237E and then using the NDE1 probe. For results see figure 5.3.8. The bands were analysed by comparing the autoradiographic spot densities to each other using an Alpha Imager™ 2200. An example of an image obtained is shown in figure 5.3.9 (see next page).



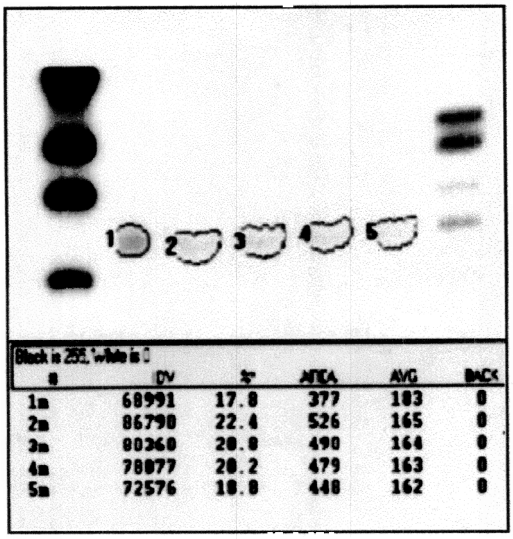
In all autorads, lanes correspond as follows:
1: λ ladder
2: total HeLa DNA
3: 6-14hr upper band
4: 6-14hr lower band
5: 14-22hr upper band
6: 14-22hr lower band
7: 1kb ladder

Note:
In the 6-14hr experiment, early replicating DNA contains BrdU
In the14-22hr experiment, late replicating DNA contains BrdU

Bands are indicated by lines (-)

Fig. 5.3.8
***Bgl*I and *Bst*XI digested HeLa DNA obtained from experiments of 6-14 and 14-22 hr BrdU incorporation, hybridised using UBE1, NDE1, DXS8237E, MIC2 and PCTK1 probes**

Fig. 5.3.9 Example of an image showing how band intensities were estimated (hybridisation of *BglI* cut HeLa DNA using the UBE1 probe)



#	DV	%	AREA	AVG	BACK
1m	68991	17.8	377	183	0
2m	86790	22.4	526	165	0
3m	80360	20.8	490	164	0
4m	78077	20.2	479	163	0
5m	72576	18.8	448	162	0

The numbered circles indicate region of the autoradiograph selected by the imager from each lane for comparison

The figures for the densities of the bands are shown in table 5.4.

5.4 Discussion

5.4.1 Synchronisation of HeLa cells using colcemid

In contrast to Gasser and Laemmli (1987) who used a colcemid concentration of 60ng/ml for 12 to 16 hours to arrest HeLa cells in metaphase, my findings, using a similar concentration of colcemid (52ng/ml), show that this length of incubation time was detrimental to the survival of HeLa cells. This might be due to using a different strain of HeLa cells, which may be less tolerant to colcemid. Nevertheless, the proportion of cells arrested in the mitotic phase were similar. Gasser and Laemmli (1987) found 50-80% of cells to be in metaphase and my results show that 51-75% of cells were arrested in metaphase, once the conditions were optimised.

5.4.2 Growth curves of synchronised HeLa cells

Six hours after releasing HeLa cells from synchronisation they attached to the bottom of the flask, this was also observed by Kozaki et al. (1993), who used a double-thymidine block to synchronise HeLa cells. Using this method, the cells

Table 5.4 Results obtained by band density measurements using an alpha imager™ 2200

General information	6-14hr upper band	6-14hr lower band early DNA	14-22hr upper band	14-22hr lower band late DNA	replication index
***Bst*XI digest**					
NDE1 IDV	6.59 *	6.56	5.54	5.51	1.00
DXS8237E IDV	5.83	5.58	5.42	6.68	1.13
UBE1 IDV	10.26	6.22	7.90	6.12	1.15
***Bgl* I digest**					
MIC2 IDV	10.57	9.26	9.85	7.58	0.93
PCTK1 IDV	4.53	4.33	4.40	4.25	1.00
UBE1 IDV	8.68	8.04	7.81	7.26	1.00

at **6 to 14 hours BrdU incorporation** the upper band without BrdU DNA contains late replicating DNA and the **lower band** with BrdU DNA contains **early replicating DNA.**

at **14 to 22 hours BrdU incorporation** the upper band without BrdU DNA contains early replicating DNA and the **lower band** with BrdU DNA contains **late replicating DNA.**

* Integrated density values (IDVs) are given in arbitrary units by the imager. They are shown $x10^{-4}$ and rounded to the closest value, although for calculations the original figures were used.

To calculate the replication index, the lower bands, i.e. the ones containing BrdU, are considered, since they contain only replicated DNA in contrast to the upper bands which additionally contains DNA from cells dying before replication for e.g. The index is calculated for each hybridisation as shown below:

$$\left(\frac{\text{IDV of the late replicating DNA}}{\text{IDV of the total amount of DNA}}\right) / \left(\frac{\text{IDV of the early replicating DNA}}{\text{IDV of the total amount of DNA}}\right)$$

arrested at the G1/S boundary. Although the authors found that the initiation time of the first DNA synthesis fluctuated between 1 and 2 hours after removal of the thymidine arrest, the cell cycle time of 23 hours was strictly maintained each cell cycle.

The growth curve obtained without BrdU added to the medium shows first signs of cell division at 22 hours rising to a clear peak at 30 hours, which indicates that the first replication cycle of the cells finishes at about 30 hours. Therefore it was estimated that DNA of active genes will replicate between 6 and 14 hours and DNA from inactive genes will replicate between 14 and 22 hours.

5.4.3 Full scale experiments

Conclusions drawn from table 5.4

Comparisons between patterns of replication for genes whose inactivation status is under investigation and those of known behaviour have been made. The Norrie gene (detected by the NDE1 probe) is inactivated in a conventional manner; in contrast, the *MIC2* gene, although X linked, is pseudoautosomal and presumed to escape X chromosome inactivation (Goodfellow, 1984).

The replication index provides an indication of the relative inactivation of the various loci by comparing the position of the replicating DNA in "late" as opposed to "early" stages of the cell cycle (**assuming late replication and inactivation are coincidental**). As, for a given filter and experiment, the total amount of DNA in upper and lower bands remains the same through the series of hybridisations, a comparison of the probe bound between lower (replicated DNA) and upper (non-replicated DNA) band provides an indication of replication activity for each locus examined during the time interval being monitored. Moreover, calculating the ratio for this figure between late stages and early stages of the cell cycle will provide an "Inactivation index". For these calculations the proportion of replicating DNA has been established by the ratio of band intensity in lower band to the total (upper + lower) (see table 5.4). From examination of the *Bgl*I digest data provided in table 5.4 it is immediately obvious that the inactivation index for MIC2 is lower than that for either PCTK1 or UBE1 indicating a preferential replication in the early stages of the cell cycle. This may reflect the status of the pseudoautosomal region as distinct from the

remainder of the X chromosome.

Overall, the figures for NDE1, DXS8237E, PCTK1 and UBE1 are higher than those for MIC2 and similar to each other with, perhaps surprisingly, the value for NDE1 (1.00) being the lowest on that filter. This may be due to the result of the relatively poor quality of the autoradiograph for NDE1 hybridisation with consequently inaccurate values for band intensities. The results indicate that all genes in the non-pseudoautosomal region are replicated relatively late irrespective of their inactivation status, indicating that the replication time of genes is domain dependent, rather than on the active/inactive state of the gene. Several limitations should be noted:

1. It would be advantageous to employ cells representing an as close simulation of the correct *in vivo* situation as possible such as normal, diploid fibroblasts and lymphocytes. Unfortunately they are difficult to culture and use since they only replicate a few times before dying. For this experiment a large number of cells were required which would have been difficult to obtain using fibroblasts or lymphocytes. HeLa cells on the other hand, being cancerous (cervical carcinoma) cells replicate continuously and are therefore easier to use, hence, they were the cell line of choice for the initial experiments. Although they are heteroploid, additional copies of X chromosomes will enhance, rather than obscure any replication differentials.

2. The length of time involved in the experimental procedures preluded the replication of the experiment. Therefore, it could not be established if the results were reproducible.

3. The proportion of non synchronised cells plus these which are dead or dying contribute to the upper layer of the CsCl centrifugation and obscure replication events by adding to the denominator in the estimates of proportions of replicating DNA.

4. Different probes lable at different intensities, which may obscure the hybridisation results.

There are several theoretical advantages to this approach to replication studies.

It does not require access to sophisticated microscopes and the expertise to undertake sensitive *in situ* hybridisation experiments. Any section of DNA from the region under investigation could be examined as long as it is a unique copy. Indeed it was the intention to extend the observations reported here to include sections within the presumed boundary area of the inactivation spread. In practice, the difficulties in obtaining sufficient DNA amounts for accurate and extensive hybridisation experiments proved the chief limitation.

CHAPTER 6 - Sequence analysis of the DXS8237E-UBE1-PCTK1 region, including exon-intron analysis of the *UBE1* gene

6.1 Introduction

The DXS8237E-UBE1-PCTK1 region on the short arm of the hromosome interest because it contains three genes within about 40kb for which there i evidence that one is inactivated (*DXS8237E*), one escapes inactivation (*PCTK1* and the inactivation status of the third gene (*UBE1*) has been controversial, although it is generally stated that it does escape X chromosome inactivation. Furthermore, a gene encoding a ubiquitin C-terminal hydrolase, *UHX1*, lies in very close proximity to *PCTK1* It was first characterised and it's cDNA cloned by Swanson et al. (1996). Although *UHX1* has been investigated by Brandau et al 1998), who performed fine mapping on *UHX1*, determined its genomic structure and undertook mutation screening to examine the possible causative role of *UHX1* in the eye disorders CSNB1 or RP2 (see Chapter 1.1.1), its inactivation status as yet is unknown. The 3' end of *DXS8237E* has been characterised by Coleman et al. (1996) and the genomic region of *PCTK1* has been described by Okuda et al. (1994) and, in more detail, by Brandau et al. (1998). Examining the sequence embracing these four genes might bring about a better understanding of the mechanism underlying the propagation of X chromosome inactivation.

During the final investigations of the PACs (see Chapter 3) it was found that two of them, PACs 1 and 4, contained the *DXS8237E*, and *UBE1* loci within one 9kb *Xho*I/ *Eag*I fragment. After preliminary attempts it was decided not to proceed with cloning and sequencing this fragment, since PAC1 was already being sequenced by the Sanger Centre (personal communication) Unfortunately PAC1 turned out to be chimæric, so work was initiated at the Centre on PAC4 which was subcloned into BACs and the sequences of which were assembled into a genomic array, AL513366. I have analysed the sequence using a computer program called NIX (described in 6.1.2), with a view to identify motifs potentially implicated in X chromosome inactivation.

Additionally, two other genes on the X chromosome, namely the Norrie disease gene (NDP) and the X-linked Zinc Finger protein (ZFX) have been analysed for

comparison with regard to Line 1 elements. Norrie disease is a X-linked recessive neurodevelopmental disease characterised by congenital blindness, which is often associated with hearing loss, mental retardation and psychiatric problems. The gene, which has been cloned lies on the short arm of the X chromosome at Xp11.4 and is known to be regularly inactivated (Meitinger et al., 1993). On the other hand, the Zinc Finger Protein at Xp22.2 - p21.3, which has a Y-homologue, escapes the inactivation process in humans, but not in mice (Luoh et al 1995)

This chapter also contains a comparison of the restriction map of PAC1 obtained in chapter 3 to a computer simulated one using the relevent region contained in clone AL513366.

6.1.1 Long Interspersed repeat elements (LINEs)

Eukaryotic gene transcription is largely controlled by sequence specific DNA-binding proteins, that may act proximally or distally to transcription start sites. Although most proteins of this type have been reported to enhance or increase the rate of transcription, a significant number of transcriptional inhibitors or repressors have now also been described. Typical gene promotors contain numerous binding sites for such positive and negative regulators but, because the binding sites are short, it is unusual to find extended regions of sequence similarity between different promotors.

Among the best-characterised conserved non-coding sequences are Alu sequences and long interspersed repeat elements (LINEs) retrotransposons, both of which occur in abundance in mammalian genomes. In humans, Alu and mammalian-wide repeat (MIR) sequences together with LINEs, which are more closely identified with having a functional role, such as retrotransposition, comprise > 25% of the entire genome (reviewed by Donnelly et al., 1999) There are believed to be between 50,000 and 100,000 LINE (L1) elements. Although greater than 90% of them are truncated, the remaining full length L1s are approximately 6kb in length and contain a 900 nucleotide GC rich, 5'-untranslated region (UTR) which contains internal transcription regulatory elements that are sufficient for efficient cell specific transcription (reviewed by McMillan and Singer, 1993). Following the UTR are two long open reading frames and a 3'-untranslated region that terminates in a poly (A) tail (reviewed

by Yang et al., 1998).

In context of chromatin inactivation it should be noted that Donnelly et al. (1999) reported the existence of a related LINE-2 fragment (ALF), which has the potential to function as a T-cell-type-specific silencer, rather than an enhancer, promotor or inhibitor as is the case with LINE-1 and Alu elements.

The Lyon repeat hypothesis

The repeat hypothesis put forward by Lyon (1998) states that X chromosome inactivation originally arose in evolution by making use of an existing cell defence mechanism to silence regions of one X chromosome which had lost homologues on the Y chromosome as those chromosomes diverged. She further presumed that this region was rich in LINE elements as these represent a potential target for selective silencing and that selection for their accumulation during evolution favoured the spread of the silencing mechanism along the X chromosome

This hypothesis is based on a number of studies. Firstly, the X chromosome in mouse and human is exceptionally rich in LINE elements (Boyle et al., 1990; Korenberg and Rykowski, 1988), which are potential candidates for "booster elements", which may serve as DNA signals to amplify and spread the X inactivation signal along the X chromosome (Riggs 1990; Lyon 1998). Secondly, examination of mouse X:autosome translocation data showed that failure of the X chromosome inactivation signals to spread often correlated with cytogenetic bands that were deficient in LINE-1 elements (reviewed by Lyon, 1998). Further, in many organisms including plants (Matzke and Matzke, 1995), *Drosophila* (Dorer and Henikoff, 1994) and fungi (Selker, 1997) reiterated transgenes are expressed poorly. The transcription rate of individual copies decreases noticeably as the copy number of inserted transgenes increases. It is speculated that there is a cell defence mechanism, present in most eukaryotes, which senses the presence of repeated sequences, which are likely to be due to invading intrachromosomal parasites, such as transposable elements, and effectively silence them The mechanism by which the repeats are recognised is unknown; however, silencing may be produced by methylation of DNA (Matzke and Matzke, 1995; Yorder et al., 1997) or by converting the repeats to a heterochromatic state (Dorer and Henikoff,

1994 The repeated elements do not need to be immediately contiguous, the could be brought into contact by chromatin being thrown nto folds Dore Henikoff, 1994). The acquisition of the heterochromatic propertie by inactive X chromosome is thought to be brought about by its coating with *XIST* RNA. Bailey et al. (2000) suggest that the L1 elements serve as binding sites for the *XIST* RNA, perhaps acting through an RNA-protein complex In this context it is important to note that in *Drosophila* although the mechanism of compensation is different, RNA-protein complexes are nvo in spre n of dosage compensation

Additionally, it is interesting that tandemly repeated LINE-1 elements are capable of forming heterochromatin-like structures in other species. For example, amongst whales and dolphins the core repeat of the α-heterochromatin satellite DNA consists of a 450bp DNA fragment with 63% similarity to L1 retrotransposons. Similarly in short-tailed field vole and Syrian hamste heterochromatin structures are highly enriched for L1 elements Kapitono e al Neitzel et al. 1998).

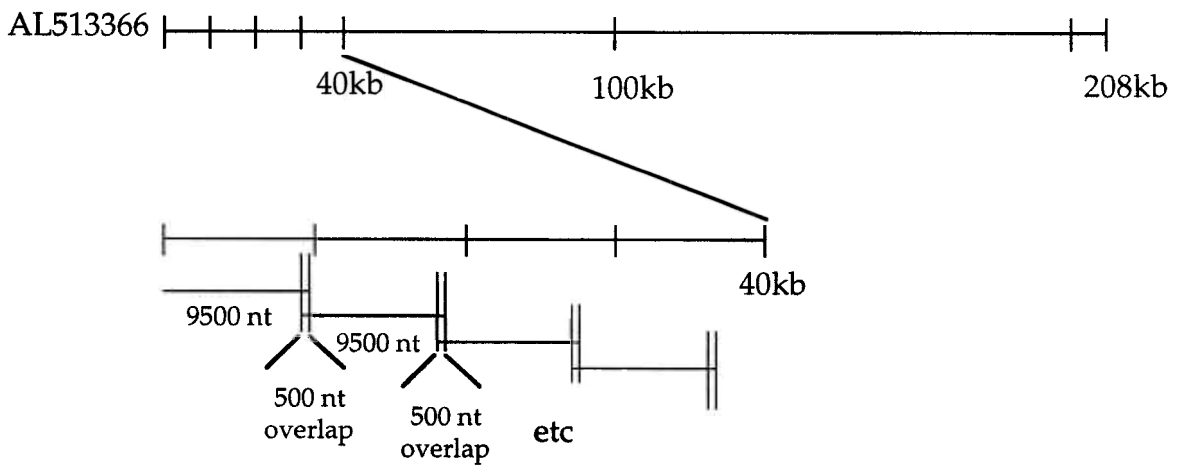
6.1.2 The Nucleotide Identify X (NIX) computer program

NIX is a tool to aid the identification of interesting genomic or transcribed nucleic acid sequences. The results of running about a dozen DNA analysis programs of a particular DNA sequence are viewed on one page and therefore the results can be compared and assessed. The NIX system selects reasonable' defaults for the programs based on whether the sequence is genomic or transcribed, its species origin and its size (up to 150kb). NIX masks the target sequence for repeat regions using Washington University's repeatmasker program (http://ftp.genome.washington.edu/RM/webhelp.html). It starts BLAST (basic local alignment search tool) searches by using the masked sequence against various databases (e.g. ecoli, EMBL, SwissProt) and outputs up to about one million alignments. The program also uses the masked sequence to predict exons and CpG islands (e.g. using GRAIL). Results are printed in colour programs with similar purposes have the same colour each of which comes in three different shades to indicate the quality of the prediction: the more intense the shade the better the prediction (http://www.hgmp.mrc.ac.uk/Registered/Webapp/nix/).

6.2 Materials and Methods

The DNA insert from PAC4 (dJ0961F01) provided the basis for the isolation of a series of BACs, the sequences of which were combined into a genomic assembly, AL513366, by the Sanger Centre. The 208,100 nucleotide (nt) long assembly had to be split into smaller fragments for analysis since the GAP4 program of the Staden package (details of which can be found at http://www.mrc-lmb.cam.ac.uk/pubseq/overview.html), which was used in this study, can only assemble chunks of DNA comprising a length of up to 10,000 nucleotides. For the ease of manipulation, the first 199,000nt of the AL513366 assembly were split into two 99,500nt fragments made up of 9500nt sub-fragments with 500nt overlaps as demonstrated in figure 6.2.1

Fig. 6.2.1 Diagrammatic presentation of splitting the AL513366 assembly into smaller fragments



The genome data base (GDB) was used to obtain the nucleic acid sequence for *UBE1*, *PCTK1* and *UHX1* using EMBL (see Chapter 1.5.3) and carrying out a chromosome specific blast search at the Sanger Centre site. The cDNA sequence of those three genes were contained in clones with the accession numbers M58028, X66363 and HSU44839, respectively. It was found that another clone, AL096791, overlapped with the HSU44839 clone. The 3' mRNA sequence of *DXS8237E* was retrieved using SRS (sequence retrieval system for further reference on SRS see http://srs.hgmp.mrc.uk/) and was contained in a clone with the accession number HSU35373. This clone only contains partial coding sequence, but, at the time, is the only one available for *DXS8237E*

The exon-intron boundaries of *UBE1* were established by aligning the cDNA to the genomic AL513366 assembly using the gap sequence alignment program and identifying the boundary sequences (exon-[GT-intron-AG]-exon). Further, the translation initiation site and polyadenylation signal (poly (A) tail) were located

Alignments of the other retrieved sequences with the AL513366 assembly were carried out to determine the end and start positions of the *DXS8237E UBE1 PCTK1* and *UHX1* genes to enable the analysis of the intervening sequences- especially for the distribution of LINE1 elements, using the NIX computer program

6.3 Results

6.3.1 Exon-Intron Boundaries of the *UBE1* gene

The genomic AL513366 sequence and retrieved cDNA UBE1 clone (M58028) were aligned to establish the exon-intron boundaries for *UBE1* (see Chapter 6.2 Results show that *UBE1* is 21211 nucleotides (genomic DNA length) long an contains 26 exons, the shortest one being made up of 60nt and the longest and last exon containing 313nt. The start codon of *UBE1* (ATG) is located at the very beginning of exon 2. Exon 26 was followed by the polyadenylation signal AATAAA. For results see table 6.1

Note: The start codon of the cDNA sequence for *A1S9T* as shown in lone X5289 is positioned within exon 8 of the *UBE1* clone M58028. However, the translat protein sequences end at the same position (result not shown)

6.3.2 Establishment of the DXS8237E, PCTK1 and UHX1 sequence boundaries

In order to be able to investigate potentially interesting motifs in the sequence between the genes in the UBE1 cluster, apart from the information obtained for *UBE1*, it was necessary to establish the end of *DXS8237E*, the beginning and end of *PCTK1* and the beginning of *UHX1*. Although the genes in the UBE1 region have been studied to a certain extent and the genomic location to each other had been determined by physical mapping (e.g. Handley et al., 1991; Okuda et al., 1994; Knight et al., 1995; Carrel et al., 1996; Coleman et al., 1996; Swanson et al., 1996;

Table 6.1 Exon-Inton analysis of the UBE1 gene

Exon	Position	Exon size (bp)	Intron size (bp)	Splice acceptor	Splice donor
1	1-129	129	4779	GAGAAGGCGGC	ACCGGCATTGgtaaaaaaacgt
2	130-246	117	128	tttttcctccagATGTCCAGCT	GCCAACCAACgtgagtgtcctc
3	247-305	59	102	ctctattcctagGGAATGGCCA	CCCGGCAGCTgtaagtggggcc
4	306-474	169	102	ctaacctggcagGTATGTGTTG	TTCCTCCCAGgtacctcttcct
5	475-609	135	1279	acacccttacagTTCTACCTGC	TGGTTTCCAGgtatcttggggg
6	610-716	107	275	atctctccacagGTGGTGGTGC	GCCTGTTTGGgtgagtggcagc
7	717-807	91	111	ttccctctacagGCAGCTCTTC	GGTTACCAAGgtaaggagacca
8	808-940	133	545	tcctttaaccagGACAACCCCG	AAAGTCCTGGgtgagctgcgac
9	941-1038	98	104	tgccctctgtagGTCCTTATAC	GATTAGCTTTgtgagtgtgtcg
10	1039-1185	147	133	cccccgccacagAAATCCTTGG	CCGCAATGAGgtgggtgagtgg
11	1186-1362	177	128	ctggcccaccagGAGGATGCAG	AGTCATGAAGgtcagcacgggt
12	1363-1467	105	86	tcccctctccagGCCTGCTCCG	GTGCCTCCAGgtatgtgggtgg
13	1468-1548	81	326	tgcattccttagCGCCAGAACC	GTATTTCCTGgtaagtggtccc
14	1549-1704	156	2251	gttgcttcttagGTGGGTGCGG	GGATGTCACGgtgagtagggta
15	1705-1870	166	134	ttaacccttagAAGTTAAAGT	GTGGATGCCCgtgagtttggag
16	1871-2067	197	3178	tgctccccacagGCATGTACAT	CACCCTGCAGgtgataagctgt
17	2068-2132	65	240	ctctgtctgcagTGGGCTCGGG	AGTACCTCACgtgagtaactcg
18	2133-2328	196	718	gattcctgatagAGACCCCAAG	TCCTGACCAGgtaatgcccagt
19	2329-2403	75	119	tccatcccttagCTCACAAGCT	TGTCAACAATgtaagtctcctt
20	2404-2593	190	1198	gccttctcctagCCCCTGCATC	GCCTCTGTTGgtgagggtgttt
21	2594-2682	89	259	ttcttcatgcagATGACAGTCG	CTTTGAGAAGgtatggggtggg
22	2683-2775	93	126	cactcataacagGATGATGACA	CCGGCACAAGgtgaggggaatc
23	2776-2967	192	1145	tctgtctctcagAGCAAGCTGA	ACGTCACCAGgtgggggcctgc
24	2968-3069	102	108	tttgtcttgcagTACTATAACC	CTATTTTAAGgtaaggcccctc
25	3070-3170	101	156	ctccctctccagACAGAGCACA	TGGATCAGCCgtgagttggaca
26	3171-3504 (end)	333		ccctatccccagGATGACAGAG	AATAAA GAATTAATAA

- intronic sequence is indicated by small letters
- EXONIC SEQUENCE IS WRITTEN IN CAPITALS
- The end/start of introns are pointed out in blue
- The start signal is shown in green
- The polyadylation signal is highlighted in red

Brandau et al., 1998; Stoddart et al. 1999), this had not been done previously by sequence alignment (see Chapter 6.1

The start of the polyadenylation signal of *DXS8237E* is located 12012 nucleotides downstream of the *UBE1* start codon. There are 8446nt between the start of the polyadenylation signal of *UBE1* and start codon of *PCTK1*. Further, the start codon of *UHX1* lies 4285nt upstream from the stop signal of *PCTK1* (the polyadenylation signal was not detectable). The total distance between the poly (A) tail of *DXS8237E* and the beginning of *UHX1* is 46256 nuclelotides. It should be noted that, according to the alignment of the HSU44839 clone to the genomic AL513366 assembly, there may be 5' sequence beyond the first identifiable exon of *UHX1*, since at the position where the sequence diverged a splice acceptor was noted.

6.3.3 Sequence analysis of the DXS8237E-UBE1-PCTK1-UHX1 region

NIX analysis was carried out on the second 100,000 nucleotides of the AL513366 assembly (see M&M), spanning the region in which the four genes lie in to give an overview. Additionally, the program was run covering the regions between *DXS8237E* and *UBE1*, *UBE1* and *PCTK1* and *PCTK1* and *UHX1*. The results (regarding the + strand only) show that the *UBE1* and *PCTK1* genes are flanked by tandemly repeated LINE1 elements. The first one is located 1494 nucleotides upstream of the beginning of the poly (A) tail of *DXS8237E* or 10518nt from the start codon of *UBE1*. It comprises a L1ME2 element, which is 595nt long and a L1M4 repeat of 380nt. When tested for the ability to form hairpin-like structures by aligning the reverse complementary sequence of one to the other, the two sequences were shown to overall match only ~42% (results not shown).
second tandem repeat is located within intron 2 (or a higher number intron depending on if there are more exons located beyond the 5' sequence of HSU44839 clone) of *UHX1*, starting 1075nt upstream from the start of *UHX1*. It comprises a L1MC4, L1MA8 and a L1MC4 repeat, which are 482, 280 and 137nt in length, respectively. Again, alignment showed an average of only ~35% overall match (results not shown). Furthermore, one more L1, a L1MC/D, is located between *DXS8237E* and *UBE1* and lies upstream from the tandemly repeated L1
For positions of the LINE1 elements throughout the DXS8237E-UBE1-PCTK1 UHX1 region see table 6.2.

Table 6.2 LINE1 elements throughout the DXS8237E-UBE1-PCTK1-UHX1 region

position of L1	+ strand		complementary strand	
	type of L1	length	type of L1	length
-27279bp from polyA of *DXS8237E*	L1P4	343bp	Between - 33142bp and -18993bp from the polyA tail of DXS8237E lie 13 line elements which are one of the following L1M4, L1MC4, L1PA16, L1MC/D, L1	
-26635bp from polyA of *DXS8237E*	L1P4	539bp		
-23906bp from polyA of *DXS8237E*	L1MD3	566bp		
-22621bp from polyA of *DXS8237E*	L1MB5	157bp		
-22250bp from polyA of *DXS8237E*	L1MB5	692bp		
-20754 bp from polyA of *DXS8237E*	L1MB5	41bp		
-17756bp from polyA of *DXS8237E*	L1MD3	198bp		
-3243bp from polyA of *DXS8237E*	L1PB3	135bp		
-2792bp from polyA of *DXS8237E*	L1PB3	33bp		
tandem repeat 1 (starting 1494bp downstream of polyA of *DXS8237E*)	L1ME2 L1M4	595bp 380bp		
+2887bp from polyA of *DXS8237E* -9125bp from start codon of *UBE1*	L1MC/D	111bp		
within intron16 of UBE1			L1MC4 L1ME3A L1ME3A	196bp 35bp 75bp
tandem repeat 2 within an intron of the *UHX1* gene and starting 1075bp downstream of the start of *UHX1*	L1MC4 L1MA8 L1MC4	482bp 280bp 137bp		
+4140bp from start of *PCTK1* -1068bp from stop of *PCTK1*			L1ME3A	296bp
-765bp from stop of *PCTK1*			L1MC5	242bp
+2259bp downstream of start of*UHX1*			L1MC4a	135bp
+2975bp downstream of start of *UHX1*	L1MB5	148bp		
+3647bp downstream of start of *UHX1*	L1M4	152bp		
+4977bp downstream of start of *UHX1*	L1MB3	60bp		
Between +17102 and +20109bp downstream of the start of UHX1			5 L1 elements which are one of the following: L1M4, L1PA3, L1MEc	

Note: the exact positions of LINE1 elements on the complementary strand are only given in the area in which the four genes lie.

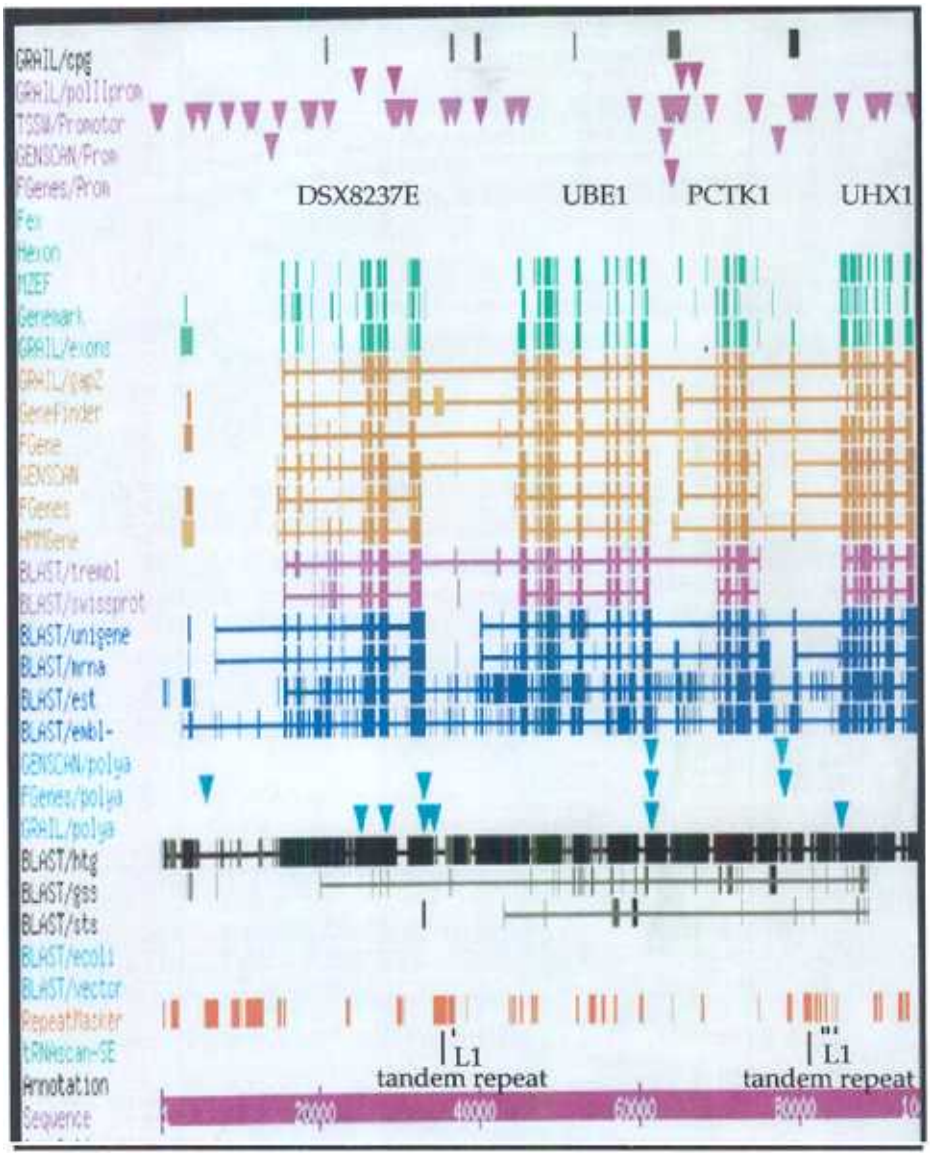
In dramatic contrast, no L1s were detected in the sequence between *UBE1* and *PCTK1* The other repetitive elements indicated by the NIX printout and shown in the appendix include SINEs (short interspersed repeat elements), such as Alu repeats, and LINE2s. See figure 6.3.1 for NIX results.

It is of interest that sequence alignment analysis showed that the (AC)n repeat (see Chapters 3 and 4) was located to intron 16 of the *UBE1* gene, 6834 nucleotides downstream of the beginning of the poly (A) tail of *UBE1*. Another (AC)n repeat sequence was detected within the same intron of *UHX1* that contains the tandem repeats described above. It is located 4727nt upstream from the start of *UHX1* and lies 2751nt upstream from the finish of the tandem repeat. Both repeats will be usefull as intragenic markers.

In addition, using sequence alignment analysis it was possible to establish the exact positions of the primers, which were used to construct the gene probes. The primer set for *DXS8237E* is located 406-1660 nucleotides downstream from its polyadenylation signal; the primer set for *UBE1* is located 2726-3431 nucleotides upstream from its start signal, with the 1st primer being located in exon 8 and the second one in exon 9 of *UBE1*. The primer set for *PCTK1* seems to be located in the centre of the gene: 2238-3051 nucleotides upstream from the start and 2972-2159 nucleotides downstream from the stop signal of *PCTK1*. There are 15142 nucleotides between the end of DXS8237E primer 2 and the start of the UBE1 primer 1 and 23558 nucleotides between the end of UBE1 primer 2 and the start of the PCTK1 primer 1 From the first nucleotide of the first DXS8237E primer to the last of the second PCTK1 primer there are 41472 nucleotides.

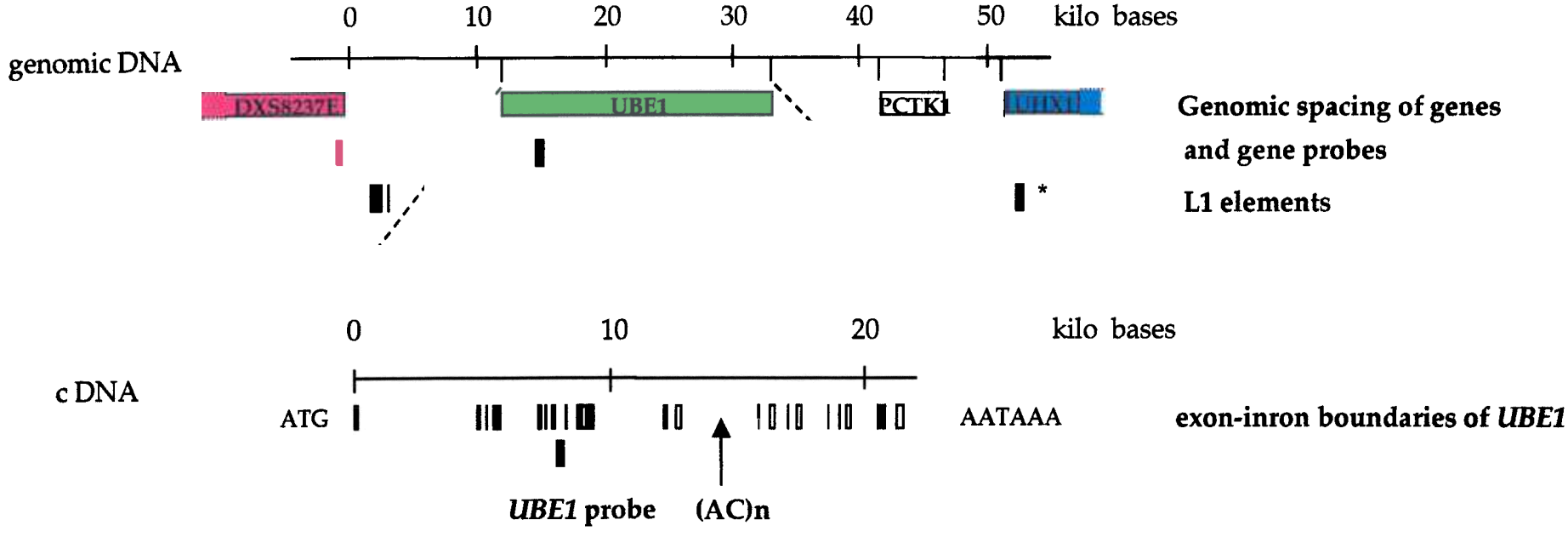
A diagrammatic presentation of the area is shown in figure 6.3.2.

Fig. 6.3.1 Modified NIX printout for the region spanning DXS8237E-UBE1-PCTK1



For results showing the sequence between 2 genes only, see appendix.

Figure 6.3.2 Gene organisation and exon-intron boundaries of *UBE1*



Note:

- positions are approximately to scale within limitations of drawing program
- the last tandom repeat is shown as one unit
- genes open to one side have not been mapped fully
- * : all three L1s have been drawn in one element
- L1s on the complementary strand are not shown

exons < 100 nts
exons 101-175 nts
exons 176-250 nts
exons 251-325 nts

6.3.4 Analysis of the Norrie disease and X-linked Zinc Finger protein gene with regard to L1 elements

The *NDP* and *ZFX* genes are contained in clones with accession numbers AL034370 and AC002404, respectively. Both clones have been analysed and details can be found at http://www.ncbi.nlm.nih.gov/entrez/viewer.cgi?val=3947696 for the *NDP* gene and http://www.ncbi.nlm.nih.gov/entrez/viewer.cgi?val=2329922 for the *ZFX* gene. Since the latter clone does not show the postition of the *ZFX* gene, this was established for all three known isoforms, the mRNA sequences of which were obtained from the GDB (accession numbers: X59738, X59739 and X59740) and alignments carried out using the GAP4 program. The results are shown in table 6.3.

Table 6.3: Positions of the *ZFX* gene-isoforms within clones AC002404 and AC002404

Isoform of *ZFX* (accession number)	start position in clone AC002404	end position in clone AC002404
1 (X59738)	92783	30887
2 (X59739)	92683	28023
3 (X59740)	92683	30180

Note: the start and end positions refer to the cDNA clones start and end, not the genes themselves.

For analysis purposes the start of the gene was taken as 92783bp and the end as 28023bp. To either side of both genes 10000 kb were examined for L1 elements. It should be noted that the *ZFX* and NPD genes lie on the complementary strands of the AC02404 and AL034370 clones, respectively

Only one L1 element can be found within 10000bp proximal to the *ZFX* gene; this is at position -6675 to -6406 (269bp). There are two L1s on the complementary strand at positions -4177 to -4070 (107bp) and -3965 to -3792 (173bp). Twelve L1s lie within the 64760bp *ZFX* gene of which half are located on the complementary strand. Distal to the gene only three L1s can be found within 10000bp, all of which are on the complementary strand at positions 3152 to 3294 (142bp), 3766 to

4122 (356bp) and 8824 to 8992 (168bp).

All L1 elements regarding the *NDP* gene lie on the complementary strand to the one containing the gene. No L1 elements lie within 10000bp proximal to *NDP* gene. The 24728bp gene contains a L1P4 and a L1M4b element and within 10000bp distal to it one L1MC1 at position 5206 to 6458 (252bp) and two L1PA16 repeats at positions 8848 to 9228 (380bp) and 9532 to 10195 (663bp), respectively.

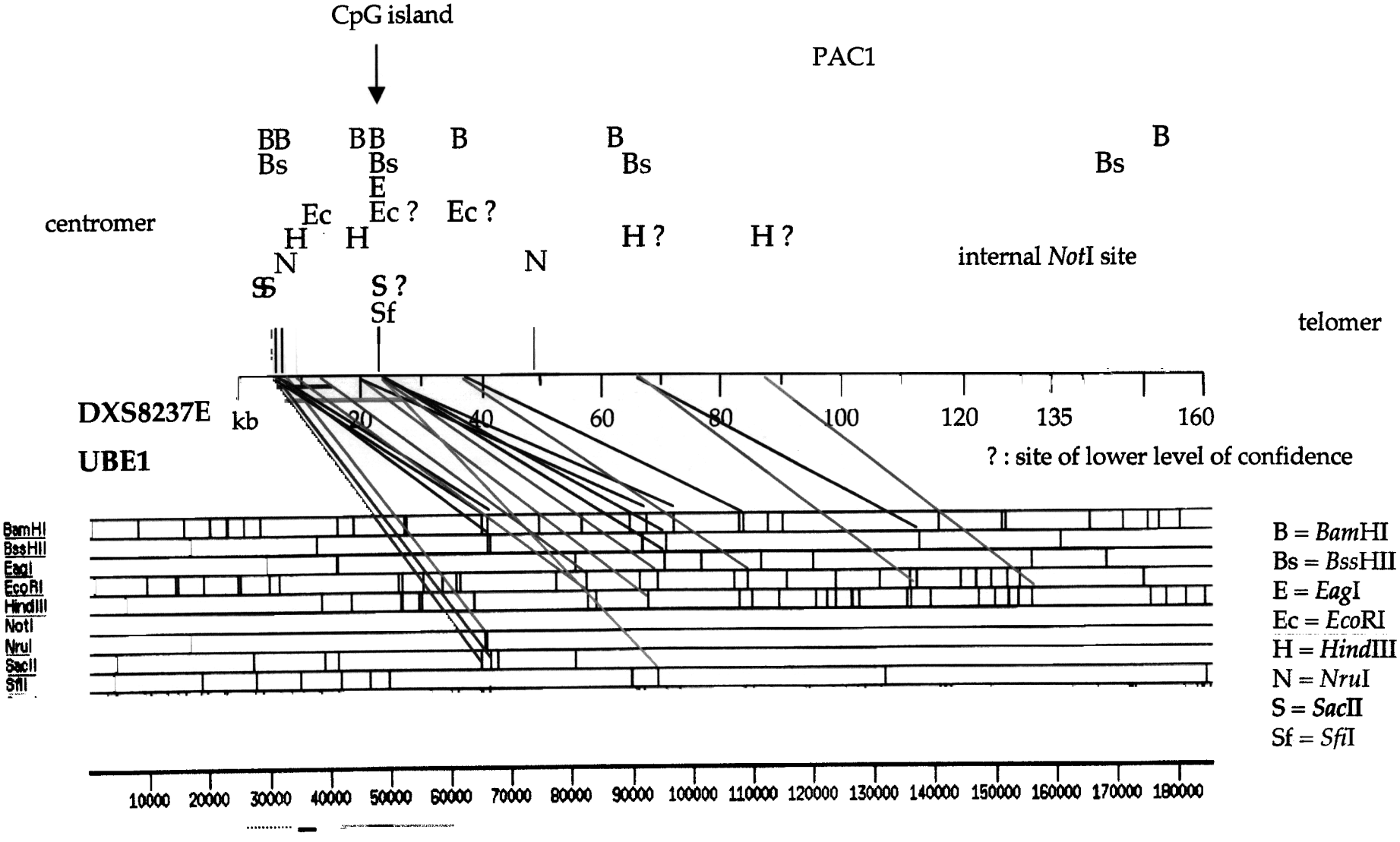
Table 6.4 Summary of the presence of L1 elements obtained for each investigated

Location of L1 elements relative to the gene	Clone AC02404 (ZFX 64760bp)		Clone AL034370 (NDP 24726bp)	
	L1s	L1s on complement. strand to gene	L1s	L1s on complement. strand to gene
LINE/L1 prox. to gene	1	2	0	0
LINE/L1 within gene	6	6	0	2
LINE/L1 distal. to gene	0	3	0	3

6.3.5 Comparison of the physical map of PAC1 to a computer simulated restriction map of the corresponding region within clone AL513366

The hgmp computer program tacg has been used to simulate an enzyme digest on the part of clone AL513366, which covers most of PAC1. PAC1 was chosen because it contains the most enzymes used for restriction mapping (*Bam*HI, *Bss*HII, *Eag*I, *Eco*RI, *Hind*III, *Not*I, *Nru*I, *Sac*II and *Sfi*I). It was also hoped that the comparison would throw light on the nature of the presumptive rearrangement noted earlier. Fewer restriction enzyme data were making comparisons difficult. Results are shown in figure 6.3.3.

Fig. 6.3.3 **Comparison of PAC1 to a computer simultated restriction enzyme digest of the relevant part of clone AL513366**
Note: the scale of both maps is the same. However, because maps for different enzymes are vertically stacked the lines connection the sites do not always appear parallel.



6.4 Discussion

6.4.1 Exon-Intron Boundaries of the *UBE1* gene

Although a lot of work has been done on this region, the information presented here is the first reported establishment of the *UBE1* exon-intron boundaries. It has brought us one step further in the analysis of this important and interesting gene. In 1990 Zacksenhaus and Sheinin reported the molecular clon primary structure of the human *A1S9T* gene, now called *UBE* he autho submitted the nucleotide and predicted amino acid sequence to EMBL (accession number X52897). By establishing the exon-intron boundaries of *UBE1*, I have been able to find that the start codon of the cDNA sequence for *A1S9T* as reported for clone X52897 was located within exon 8 of the *UBE1* clone M58028. However, the translated protein sequences end at the same position (result not shown) Just one year later, Handley et al. (1991) reported the cloning and sequencing of the cDNA-encoding human *UBE1* gene, but no exon-intron boundaries were established. In the following years physical mapping of the region was carried out (e.g. Lafrenière et al., 1991; Kwan et al., 1991; Hagemann et al., 1994; Carrel et al., 1996; Stoddart et al., 1999) and the inactivation status (e.g. Bressler et al., 1993; Salido et al., 1993) and function (e.g. Pickart et al., 1994) of the gene examined, but again, no molecular genomic structure of *UBE1* was determined.

6.4.2 Establishment of the DXS8237E, PCTK1 and UHX1 sequence boundaries

The distances between the *DXS8237E, UBE1, PCTK1 UHX1* genes have previously only been estimated by physical mapping, but exact information on the sequence of the genes and adjacent motifs has been lacking. Coleman et al. (1996) had mapped the 3' end of *DXS8237E* within 20kb upstream of *UBE1* in Xp11.23 and Knight et al. (1995) had located *PCTK1*, positioned in Xp11.3-p11.23 within 420kb of the *UBE1* gene. More recent studies had shown that *PCTK1* mapped within 5kb of *UBE1* (Carrel et al., 1996). Exon-intron boundaries were established for *PCTK1* and *UHX1* by Brandau et al. (1998) and the *DXS8237E, UBE1, PCTK1, UHX1* genes were said to lie within 70kb of each other.

In this study, it has been established that the start of the polyadenylation signal of *DXS8237E* is located 12012 nucleotides downstream of the *UBE1* start codon. There are 8446 nucleotides between the start of the polyade ylation signal

UBE1 and start codon of *PCTK1*. Further, the start codon of *UHX1* lies 4285 nucleotides upstream from the stop codon of *PCTK1* (the polyadenylation sign was not detectable). Only 46kb separate *DXS8237E* from *UHX1*. This confirms this region to be extremely gene rich.

Since the AL513366 assembly contains unfinished sequence, it is possible that the exon-intron structure and the precise distances between the genes to each other might change slightly.

6.4.3 LINE-1 elements in the DXS8237E-UBE1-PCTK1-UHX1 region

INE1 elements are thought to have an important role in X chromosome inactivation. They may act as DNA signals to amplify and spread the X inactivation signal along the X chromosome (Riggs, 1990; Lyon, 1998). Further, Bailey et al. (2000) suggest that LINE1 elements serve as binding sites for the *XIST* RNA, perhaps acting through an RNA-protein complex. Such complexes are known to be involved in spreading of dosage compensation in *Drosophila* .

The distribution of LINE1 elements in the DXS8237E-UBE1-PCTK1-UHX1 region was investigated to see if they may play a role in X chromosome inactivation. The *UBE1* and *PCTK1* genes were found to be flanked by tandemly repeated L1 elements. In addition the region between the *DXS8237E* and *UBE1* genes contained one LINE1 element. Most interestingly, no L1 elements were found in the sequence between the *UBE1* and *PCTK1* genes. The results obtained suggest a potential link between L1 elements and X chromosome inactivation. LINE1 elements found on the complementary strand in this region could possibly aid forming secondary structures. It may not be of much significance that the tandemly repeated L1s are unlikely to form hairpin-like secondary structures, with under 50% overall match. As Dorer and Henikoff (1994) point out, repeated elements forming secondary structures could be brought into contact by chromatin folding. The second tandemly repeated LINE element lying within one of the first introns of *UHX1* might indicate that this gene undergoes chromosome inactivation. Further experiments would be necessary to definitely conclude about the inactivation status of *UHX1*

6.4.4 LINE-1 elements regarding the *ZFX* and *NDE* gene

a gene escaping inactivation, contains substantially more L1 elements than *NDE*, which is inactivated. This could mean that one (or more) secondary structure(s) is (are) formed by L1 element interaction, possibly looping out the *ZFX* gene and thereby allowing the spread of inactivation to pass it unnoticed. The comparatively few L1 elements on the complementary strand of the *NDE* gene, on the other hand, may not be sufficient to form such secondary structures and therefore allowing normal inactivation.

6.4.5 Restriction map comparison of a computer simulated one to the PAC1 one constructed in this study

Figure 6.3.3. shows that the two maps agree overall. Small fragments might not have been detected because they either were too small to be picked up by pulsed field gel electrophoresis or because they did not separate from bigger bands. Other disagreements, such as the absence of restriction sites from the map generated in this study, could be due to the PAC rearranging, since only one change will alter the recognition site for any given enzyme. This comparison confirms that the restriction mapping of the PACs was essentially reliable.

6.4.6 Further investigations

In addition it will be interesting to examine further the sequences between the *DXS8237E*, *UBE1*, *PCTK1* and *UHX1* genes for other motives, such as binding for zinc finger proteins. Bell et al. (1999) identified a 42bp fragment of the chicken β-globin insulator that is both necessary and sufficient for enhancer blocking activity. This sequence is the binding site for CTCF, an 11 zinc finger protein highly conserved and ubiquitous in vertebrates. CTCF has been implicated in both transcriptional silencing and activation. CTCF sites were also found in other vertebrate enhancer-blocking elements. Interestingly, chromatin boundary/insulator elements are functionally conserved among species such chicken elements work in human and *Drosophila* cells and *Drosophila* elements work in *Xenopus* (reviewed by Bi and Broach, 2001). Therefore zinc motives may play a role in X chromosome inactivation in humans and their identification may provide a clearer picture of it.

CHAPTER 7 - Discussion

The X chromosome has traditionally been the most extensively studied of ll human chromosomes. This is a direct result of the large number of X-linked disease genes identified and due to the phenomenon of X chromosome inactivation, both of which are related to different gene numbers/ dosages of X chromosomes in males and females.

In this study the DXS8237E-UBE1-PCTK1-UHX1 region on the short arm of the X chromosome, at Xp11.3, has been investigated. This area is of particular interest since it contains at least one gene that is inactivated (*DXS8237E*), at least one that escapes this process (*PCTK1*) and genes whose the inactivation states are not quite clear (*UBE1* and *UHX1*). Two approaches to gain an insight into the X-inactivation status of the genes in the region and the nature of the X-inactivation mechanism have been carried out. One was to employ gene probes to compare the ratios of the gene copy numbers in early and late replicating DNA for the loci in question.

A relationship between chromosome condensation, late DNA replication and gene silencing was put forward as long as 40 years ago (e.g. by Lima-de-Faria, 1959; Kajiwara and Mueller, 1964). About 20 years later it was hypothesised that expressed genes replicate early, silent genes replicate late and changes in expression status are accompanied by corresponding changes in the replication time of a gene (Goldman et al., 1984; Hatton et al., 1988). However, exceptions to this general pattern exist (Keohane et al., 1996; Keohane et al., 1999). A novel experiment was set up (see Chapter 5) to determine the inactivation status of the *DXS8237E*, *UBE1* and *PCTK1* genes according to their replication times.

From the results obtained, a replication/ inactivation index was calculated, which proved to be lower for *MIC2* than for *NDE1*, *DXS8237E*, *PCTK1* and *UBE1*. This indicates that *MIC2* preferentially replicates in the early stages of the cell cycle, which might reflect the status of the pseudoautosomal region as being different to the remainder of the X chromosome, in which genes may replicate relatively late irrespective of their inactivation status.

The results obtained could be tested by cytological studies using fluorescence *in*

situ hybridisation (FISH). Probes derived from the various genes could be labeled, chromosome preparations made at different times of the cell cycle and hybridised to the probes. Conclusions would be drawn from the number of the signal(s) at each site.

The other approach to investigate the mechanism of X chromosome inactivation was to search for significant chromosomal motifs including LINE1 elements in the DXS8237E-UBE1-PCTK1 area (see Chapter 6). LINE1 elements (a total of 516,000 copies in the human genome, being abundant in G-dark regions of the chromosome), unlike Alu elements, are not associated with gene density (Venter et al., 2001), but are thought to play an important role in X-inactivation. They may amplify and spread the X inactivation signal along the X chromosome by acting as DNA signals for condensation (Riggs, 1990; Lyon, 1998). Further, Bailey et al. (2000) suggest that, perhaps acting through an RNA-protein complex, LINE1 elements serve as binding sites for the *XIST* RNA, which is encoded by the *XIST* gene, which, in turn, is necessary for X inactivation to occur (Penny, Kay et al., 1996; Lee et al., 1996; Lee et al., 1997(1); Herzing et al., 1997; Marahrens et al., 1997).

The results obtained as illustrated in Chapter 6 show that the *UBE1* and *PCTK1* genes are flanked by tandemly repeated LINE1 elements with sense sequences on the same strand. Additionally, another L1 element was found between *DXS8237E* and *UBE1* upstream from the tandem repeat. In dramatic contrast, no L1s were detected in the sequence between *UBE1* and *PCTK1*. The region overall is flanked by dense distribution of LINE1 elements. Given the prevailing balance of evidence that *PCTK1* and *UBE1* escape X chromosome inactivation and that *DXS8237E* does not, strongly suggests a link between L1 elements and chromosome inactivation. The existence of a second tandemly repeated LINE element lying within the putative first intron of *UHX1* might indicate that gene undergoes X chromosome inactivation. Nevertheless, in order to confirm the inactivation status of *UHX1* further experiments would be necessary.

Looking at two other genes on the X chromosome, one of which escapes inactivation, *ZFX*, and one that does not, *NDE* although having superfic different patterns of L1 elements to these described above nevertheless could implicate LINE-1 elements being involved in X-inactivation. The ZFX gene is associated with more L1 elements on average than the NDE gene, which could

mean that a secondary structure is formed by L1 element interaction, possibly looping out the *ZFX* gene of the linear DNA strand and thereby allowing the spread of inactivation to pass it unnoticed. Alternatively, the L1 elements could aid in heterochromatin formation, making 'access' to the gene impossible. does not necessarily mean that the Lyon hypothesis or Bailey's suggestion that L1 elements serve as binding sites for the *XIST* RNA is not correct, these ideas can coexist and do not rule each other out. The comparatively few L1 elements on the complementary strand of the *NDE* gene, on the other hand, may not be sufficient enough to form such secondary structures and it therefore does not escape X chromosome inactivation. The looping out/ heterochromatin formation theory could also hold true for the DXS8237E-UBE1-PCTK1-UHX1 region

Concerning the organisation of the region, the genomic location of *DXS8237E*, *UBE1*, *PCTK1* and *UHX1* had been previously determined by physical mapping (e.g. Handley et al., 1991; Okuda et al., 1994; Knight et al., 1995; Carrel et al., 1996; Coleman et al., 1996; Swanson et al., 1996; Brandau et al., 1998; Stoddart et al., 1999), but there was an overall lack of information concerning the gene structure organisation and distribution of significant flanking sequences

During this study (see Chapter 6), have been able to establish the *UBE* exon-intron boundaries, which will aid in further investigations of this highly interesting gene. Additionally, it has been concluded that only 46kb separate *DXS8237E* from *UHX1*, which confirms, bearing in mind the fact that the X-chromosome has, along with chromosomes 4, 18, 13 and Y, the lowest overall gene density (Venter et al., 2001), that this region is extremely gene containing four genes in this interval.

For the overall strategy adopted in this study, physical mapping was undertaken to detect an easily isolated fragment containing the genes of interest (see Chapter 3). This was then to be sucloned, sequenced and used to investigate the distribution of significant DNA motifs between the genes. Restriction mapping was initiated using three YACs (YA1S9T, Y2910 and Y2911), two of which turned out to be either rearranging or chimæric The chimærism of Y2911 was confirmed when sequence analysis of the end clone showed it contain an exon from a gene expressed in fœtal lung tissue and localised to chromosome 3. More

recently, database searches revealed that this sequence matches part of a T-cell antigen receptor-interacting molecule on chromosome 3 at 3q13

In a mapping study by Stoddart et al (1999) more than half of the YACs from this region were shown to be chimæric and frequently contain deletions It is highly probable that the observed instability is not only due to the properties of he YACs themselves, but also that this is a region-specific phenomenon Boycott al 996; Trump et al 1996)

Iowever, in spite of these difficulties, four novel STSs from the region were constructed, which proved useful for the initial mapping. Both STSs from the only stable YAC, Y2910, were sent to the Sanger Centre in orde to obtain PACs for the region he 21 PACs provided by the Sanger Centre were pre-examine and five were employed for restriction mapping

our out of the five PACs contain *Not*I sites, which are not apparent in the YACs. his is most probably due to the fact that YACs are too big to detect any *Not*I fragments by hybridisation with the vector end probes or that the enzyme did not cut. Internal *Not*I sites with PACs can be established by looking at the EthBr stained gel, but with YACs only a smear was visible and therefore fragments released by *Not*I digestion were difficult to resolve.

PAC1 is likely to be rearranging and PACs 2, 3 and possibly PAC5 may have derived from the same original clone. Only PAC4 contains all three gene probes and a comparison of the PAC maps was complicated by the fact that not all restriction sites were obtained for all PACs. Comparing restriction bands detected by the different gene probe hybridisations was not very conclusive. The (AC)n repeat data may be confounded by the existence of related sequences which could cross hybridise. The availability of detailed sequence information (Chapter 6) which led to the establishment of the exon-intron boundary of *UBE1*, locating the (AC)n repeat within intron 16 and the discovery of another (AC)n repeat within less than 30kb has contributed to understanding the difficulties encountered throughout the work described in Chapter 3 of this thesis.

However, comparing the restriction map obtained for PAC1 with a computer mulated one of the same region, indicates that the mapping carried out was accurate within the limitation of difficulties encountered

Further, the instability of the region suggested by the findings of Boycott et al. (1996) and Trump et al. (1996) and by the difficulty in obtaining a totally coherent contig map for YACs and PACs may be explained by its high GC content. This feature is clearly evident from the sequencing undertaken at the Sanger Centre (see Chapter 6 for detail).

After this work had been completed, Thiselton et al. (2002) published 2.7Mb functional gene map of the Xp11.3 – Xp11.23 region. The group used similar techniques as described in this work: PCR and hybridisation of ESTs and STSs, derived from YACs, to identify PAC clones. Additionally they used ESTs derived from the database. The workers encountered alike difficulties due to the clonal instability, GC and gene richness between DXS1146 and ZNF157, a region encompassing the UBE1 cluster. Computer analysis was used for fine mapping and map assembly. The map overlaps and expands upon the one published by Stoddart et al. (1999).

At the same time as mapping the YACs, a novel polymorphism, lying within the *UBE1* gene, was isolated and examined with regard to its possible association with cognitive ability (see Chapter 4). This is of particular interest since *UBE1* may play a role, directly or indirectly, in differences between male and female development due to its sex difference in dosage. Further, the mouse homologue of NEDD8, a ubiquitin like molecule, required for conjugation to target proteins, is one of a group of genes showing developmentally regulated expression in the embryonic brain (Kumar et al., 1992). Ubiquitin has been directly linked to Pick's disease, a disease causing mental disorders (Kertesz et al., 2000), is possibly another indication of the importance of the activating enzyme in the context of cognitive ability

Sequencing of the (AC)n repeat revealed that the two longest uninterrupted (AC)n repeats, within the 29 repeat sequence, are 10 and 9 repeats long. Single base interruptions may be due to sequencing errors. Genotyping, carried out in collaboration with the Social, Genetic and Development Psychiatry Research Centre, Institute of Psychiatry in London, shows that there are 5 different alleles for the obtained (AC)n repeat, which are 282, 284, 286, 288 and 290 base pairs n size, of which the allele sized 288bp is the most frequent the allele sized 286 the second most frequent one and alleles of the size of 282, 28· and 290 bp occur

very rarely. No differences in allele frequency patterns between males and females were observed

Association studies on pooled DNA samples for the genomic region embracing the *UBE1* locus in populations of high- and mid-range cognitive ability (g), again carried out in collaboration with the Social, Genetic and Development Psychiatry Research Centre, Institute of Psychiatry in London, showed no significant differences in allele frequencies. Individual genotyping for the two populations, again, indicated no highly significant differences in allele distributions for high versus mid cognition groups either for males or females, or for both combined.
It has been suggested (Plomin, personal communication) that the results of an initial genomic scan for QTLs associated with 'g' suggest that no locus is likely to contribute more than 2% to the total variance.
Evidence present here suggests that UBE1 can be excluded as a potential QLT, however, given the presumed sex differences in expression, further investigation of human behaviours with appropiate sex bias will include this marker.

References

Abbott C. and Povey S. (1995). Somatic Cell Hybrids, the basics. IRL Press, Oxford.

Adler D.A., Bressler S.L., Chapan V.M., Page D.C., Disteche C.M. Inactivation of the Zfx gene on the mouse X chromosome. *Proc. natl. Acad. USA* **88**: 4592-4595.

Aguinik A.I., Mithchell M.J., Mattei M-G., Borsani G., Avener P.A., Lerner J. and Bishop C.E., 1994. A novel X gene with a widely transcribed Y-linked homologue escapes X-inactivation in mouse and human. *Hum. Molec. Genet.* **3:** 879-884.

Alberts B., Bray D., Lewis J., Raff M., Roberts K. and Watson J.D., 1989. Molecular Biology of the Cell. Second edition. Garland Publishing, Inc.

Albertsen H.M., Abderrahim H., Cann H.M., Dausset, J., Le Paslier D. and Cohen D.,1990. Construction and characterisation of a yeast artificial chromosome library containing seven haploid genome equivalents. *Proc. Natl. Acad. Sci. U.S.A.* **87**: 4256-4260.

Alitalo T., Kruse T.A., Forsius H., Eriksson A.W. and de la Chapelle A., 1991. Localization of the Åland Island eye disease locus to the pericentromeric region of the X chromosome by linkage analysis. *Am. J. Hum. Genet.* **48**: 31-38.

Altschul S.F. and Lipman D.J., 1990. Protein database searches for multiple alignments. *Proc. Natl. Acad. Sci. USA* **87**: 5503-5513.

Anand R., 1986. Pulse field gel electrophoresis: a technique for fractionating large DNA molecules. *Trends in Genetics* **2**: 278-283.

Anand R., Villasante A. and Tyler-Smith C.,1989. Construction of yeast artificial chromosome libraries with large inserts using fractionation by pulsed-field gel electrophoresis. *Nucl. Acids Res.* **17**: 3425-3433.

Anderson C.L. and Brown C.J., 1999. Polymorphic X-chromosome inactivation of human TIMP1 gene. *Am. J. Hum. Genet.* **65**: 699-708.

Andina R.J., 1978. A study of X chromosome regulation during oogenesis in the mouse. *Exp. Cell Res.* **111**: 838-839.

Antequera F. and Bird A., 1993. Number of CpG islands and genes in human and mouse. *Proc. Natl. Acad. Sci. U.S.A.* **90**: 11995-11999.

M., Robinson E., McCarrey .R. and Cedar H., 1995. Gamete-specific methylation correlates with imprinting of the murine *Xist* gene. *Nat. Genet.* **9**: 312-315.

Ashworth A., Rastan S., Lovell-Badge R. and Kay G., 1991 X-chromosome inactivation may explain the difference in viability of XO humans and mice. *Nature* **351**: 406-408.

Bailey J.A., Carrel L., Chakravarti A. and Eichler E.E., 2000. Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothisis. *Proc. Natl. Acad. Sci.* **98**: 6634-6639.

M.L. and Bertram E.G., 1949. A morphological distinction between neurones of male and female and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis. *Nature* **163**: 676-677.

Bawle E., Tyrkus M., Lipman S. and Bozimawski D., 1984. Åarskog syndrome: full male and female expression associated with an X-autosome translocation. *Am. J. Med. Genet.* **17**: 595-602.

Bead C., Li E. and Jaenisch R., 1995. Loss of methylation extivates Xist in somatic but not in embryonic cells. *Genes Dev.* **9**: 2325-2334.

Bell A.C., West A.G. and Felsenfeld G., 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* **98**: 387-396.

Bi X. and Broach J.R., 2001. Chromosomal boundaries in *S. cerevisiae*. *Curr. Opinion in Genet. & Dev.* **11**: 199-204.

Black G.C.M., Hatchwell E. and Craig I.W., 1992. Cloning of x-specific sequences from irradiation hybrids using a novel nested Alu PCR approach. *Am. J. Hum. Genet.* **51**. Supp., 932.

Borsani G., Tonlorenzi R., Simmler M.C., Dandalo L., Arnaud D., Capra V., Grompe M., Pizzuti A., Muzny D., Lawrence C., Willard H.G., Avner P. and

Ballabio A., 1991. Characterization of the murine gene expressed from the inactive X chromosome. *Nature* **351**: 329-331.

Botstein D., White R.L., Skolnick, M. and Davies R.W., 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **32**: 314-331.

Boycott K.M., Halley G., Schlessinger D. and Bech-Hansen N.T., 1996. A 2-Megabase physical contig incorporating 43 DNA markers in the human X chromosome at p11.23-p11.22 from ZNF1 to DXS255. *Genomics* **33**: 488-497.

Boyle A.L., Ballard S.G. and Ward D.C., 1990. Differential distribution of long and short interspersed element sequences in the mouse genome: chromosome karyotyping by fluorescence *in situ* hybridization. *Proc. Natl. Acad. Sci.* **87**: 7757-7761.

Brandau O., Nakatura G., Jedele K.B., Platzer M., Achatz H., Ross M., Murken J., Rosenthal A. and Meindl A., 1998. *UHX1* and *PCTK1*: precise characterisation and localisation within a gene-rich region in Xp11.23 and evaluation as candidate genes for retinal diseases mapped to Xp21.1-p11.2. *Eur. J. Hum. Genet.* **6**: 459-466.

Bressler S.L., Kwang-Ho L., Adler D.A., Chapman V.M. and Disteche C.M., 1993. Maintenance of X Inactivation of the *Rps4*, *Zfx* and *UBE1* Gene In A Mouse in Vitro System. *Somatic Cell and Molecular Gentics* **19**: 29-37.

Brockdoff N., Ashworth A., Kay G.F., Cooper P., Smith S, McCabe V.M., Norris D.P., Penny G.D., Patel D. and Rastan S., 1991. Conservation of positon and exclusive expression of mouse Xist from the inactive X chromosome. *Nature* **351**: 515-526.

Brockdoff N., Ashworth A., Kay G.F., McCabe V.M., Norris D.P., Cooper P.J., Swift S. and Rastan S., 1992. The product of the mouse Xist gene is a 15kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* **71**: 515-526.

Brown C.J., Ballabio A., Rupert J.L., Lafreniere R.G., Grompe M., Tonlorenzi R. and Willard H.F., 1991. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* **349**: 38-

Brown C.J., Hendrich B.D., Rupert J.L. and Lafreniere R.G., 1992. The human XIST gene: analysis of a 17kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* **71**: 527-542.

Brown C.J., Lafreniere R.G., Powers V.E., Sebastio G., Ballabio A., Pettigrew A.L., Ledbetter D.H., Levy E., Craig I.W. and Willard H.F., 1991. Localization of the human X inactivation centre on the human X chromosome in Xq13. *Nature* **349**: 82-84.

Brown C.J., Powers V.E,. Munroe D.L., Sheinin R. and Willard H.F., 1989. Gene on Short Arm of Human X Chromosome Complements Purine tsA1S9 DNA Synthesis Mutation. *Cell and Molecular Genetics* **15**: 173-178.

Brown C.J. and Willard H.F., 1990. Localisation of a gene that escapes inactivation to the X chromosome proximal short arm: Implications for X inactivation. *Am. J. Hum. Genet.* **46**: 273-279.

Brown C.J. and Willard H.F., 1994. The human X-inactivation centre is not required for maintenance of X chromosome inactivation. *Nature* **368**: 154-156.

Bruyns E., Marie-Cardine A., Kirchgessner H., Sagolla K., Shevchenko A., Mann M., Autschbach F., Bensussan A., Meuer S. and Schraven B., 1998. T cell receptor (TCR) interacting molecule (TRIM), a novel disulfide-linked dimer associated with the TCR-CD3-zeta complex, recruits intracellular signaling proteins to the plasma membrane. *J. Exp. Med.* **188**: 561-575.

Bucher P., 1990. Weight matrix description of four eukaryotic RNA polymeras II promoter elements derived from 502 unrelated promoter sequences. *J. Mol. Biol.* **212**: 563-578.

Burke D.T., Carle G.F. and Olson M.V., 1987. Cloning of large segments of exogenous DNA into yeast by means of artificial chromosomes vectors. *Science* **236**: 806-812.

Carmi R., Rokhlina T., Kwitek-Black A.E., Elbedour K., Nishimura D., Stone E.M. and Sheffield V.C., 1995. Use of a DNA pooling strategy to identify a human obesity syndrome locus on chromosome 15. *Hum Mol Genet* **4**:9–13.

Carayon P. and Bord A., 1992. Identification of DNA-replicating lymphocyte subsets using a new method to label the bromo-deoxyuridine incorporated into the DNA. *J. Immunological Methods* **147**: 225-230.

Carrel, L., Clemson, C.M., Dunn, J.M., Miller, A.P., Hunt, P.A., Lawrence, J.B. and Willard, H.F., 1996. X inactivation analysis and DNA methylation studies of the ubiquitin activating enzyme E1 and PCTAIRE-1 genes in human and mouse. *Hum. Mol. Genet.*. **5**: 391-401.

Carrel L. and Willard H.F., 1999. Heterogeneous gene expression from the inactive X chromosome: An X-linked gene that escapes X inactivation in some human cell lines but is inactivated in others. *Proc. Natl. Acad. Sci. USA* **96**: 7364-7369.

Charlesworth B., 1998. Sex chromosomes: Evolving dosage compensation. *Current Biology* **8**:R931-R933.

Chen Z.-Y., Hendriks R.W., Jobling M.A., Powell J.F., Breakefield X.O., Sims K.B. and Craig I.W., 1992. Isolation and characterisation of a candidate gene for Norrie disease. *Nat. Genet.* **1**: 204-208.

Chu G., 1989. Pulsed field electrophoresis in contour-clamped homogeneous electric fields for the resolution of DNA by size or topology. *Electrophoresis* **10**: 291-295.

Chumakov I., Rigault P., Guillou S., Ougen P., Billant A., Gausconi G., Gerry P., LeGall I., Soularue P., Grinas L., et al., 1992. Continuum of overlapping clones spanning the entire human chromosome 21. *Nature.* **359**: 380-386.

Church G.M. and Gilbert W., 1984; *Proc. Natl. Acad. USA* 81, 1990-1995.

Ciechanover A., 1998. The ubiquitin-proteasome pathway: on protein death and

cell life. *EMBO* **17**: 7151-7160.

Ciechanover A., DiGuiseppe J.A., Bercovich B., orian A., Richter J.J., Schwartz A.L. and Brodeur G.M., 1991. Degradation of nuclear oncoproteins by the ubiquitin system in vitro. *Proc. Natl. Acad. Sci. USA* **88**: 139-143.

Clemson C.M., McNeil J.A., Willard H.F. and Lawrence J.B., 1996. XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. *J. Cell. Biol.* **132**: 1-17.

Cockburn D.J., Munro E.A., Craig I.W. and Boyd Y., 1992. Mapping of X chromosome translocation breakpoints in females with Duchenne muscular dystrophy with respect to exons of the dystrophin gene. *Human Genetics* **90**: 407-412.

Collin G.B., Asada Y. and Varnum D.S., 1996. DNA pooling as a quick method for finding candidate linkages in multigenic trait analysis: an example involving susceptibility to germ cell tumors. *Mam. Genome* **7**: 68 – 70.

Collins F.S., 1991. Of needles and haystacks: Finding human disease genes by positional cloning. *Clin. Genet.* **39**: 615-623.

Coleman M.P., Ambrose H.J., Carrel L., Németh A.H., Willard H.F. and K.E. Davies, 1996. A novel gene, DXS 8237E, lies within 20kb upstream of UBE1 in Xp11.23 and has a different X inactivation status. *Genomics* **31**: 135-138.

Coleman M., Bhattacharya S., Lindsay S., Wright A., Jay M., Litt M., Craig I. and Davies K., 1990. Localization of the microsatellite probe DXS426 between DXS7 and DXS255 on Xp and linkage to X-linked retinitis pigmentosa. *Am J. Hum Genet.* **47**: 935-940.

Coleman M.P., Nemeth A.H., Weissenbach J., Hunter N. and Davies K. E., 1993. A YAC contig spanning the OATL, DXS426 and DXS1003 loci. *Cytogenet. Cell Genet.* **64**: 148-187.

Coulson A., Waterson R., Kiff J., Sulston J. and Kohara Y., 1989. Genome linked with yeast artificial chromosomes. *Nature.* **335**: 184-186.

Courtier B., Heard E. and Avner P., 1995. Xce haplotypes show modified methylation in a region of the active X chromosome lying 3' to Xist. *Proc. Natl. Acad. Sci. USA* **92**: 3531-3535.

Cox D.R., Burgmeister M., Price E.R., Kim S. and Myers R.M., 1990. Radiation hybrid mapping: somatic cell genetic method for constructing high resolution maps of mammalian chromosomes. *Science* **250**: 245-250.

Creagan R.P. and Ruddle F.H., 1975. The clone panel; a systematic approach to gene mapping using interspecific somatic cell hybrids. *Cytogenet. Cell Genet.* **14**: 282-286.

Cummings C.J., Mancini M.A., Antalffy B., DeFranco D.B., Orr H.T. and Xoghbi H.Y., 1998. Chaperone suppression of aggregation and altered subcellular proteasome localization imply protein misfolding in SCA1. *Nature Genet.* **19**: 148-154.

Daniels J., Holmans P., Williams N., Turic D., McGuffin P., Plomin R. and Owen M.J., 1998. A simple method for analyzing microsatellite allele image patterns generated from DNA pools and its application to allelic association studies. *Am. J. Hum. Genet.* **62**: 1189-1197.

Davies S.W., Turmaine M., Cozens B.A., DiFiglia M., Sharp A.H., Ross C.A., Scherzinger E., Wanker E.E., Mangiarini L. and Bates G.P., 1997. Formation of neuronal intranuclear inclusions underlies the neurological dysfunction in mice transgenic for HD mutation. *Cell* **90**: 537-548.

Davidson R.G., Nitowsky H.M. and Childs B., 1963. Demonstration of two populations of cells in the human female heterozygous for glucose-6-phosphate dehydrogenase variants. *Proc Natl Acad Sci.* **50**: 481–485.

Dervasi, A. and Soller, M., 1994. Selective DNA pooling for determination of linkage between a molecular marker and a quantitative trait locus. *Genetics* **138**: 1365-1373.

de la Chapelle, 1983. In "Principles and Practice of Medical Genetics" (Emery

A.E.H. and Rimon D.L. Eds.), pp. 208-213, Churchill Livingstone, New York.

Dib C., Fauré S., Fizames C., Samson D., Drouot N., Vignal A., Millasseau P. Marc S., Hazan J., Seboun E., Lathrop M., Gyapay G., Morissette J. and Weissenbach J., 1996. A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380**: 152-154.

Disteche C.M., 1997. The great escape. *Am. J. hum. Genet.* **60**: 1312-1315.

Donahue R.P., Bias W.B., Renwick J. and McKusick V.A., 1968. Probable assignment of the Duffy blood group locus to chromosome 1 in man. *Proc. Natl. Acad. Sci. U.S.A.* **61**: 949-955.

Donnelly S.R., Hawkins T.E. and Moss S.E., 1999. A conserved nuclear element with a role in mammalian gene regulation. *Hum. Molec. Genet.* **9**: 1723-1728.

Dorer D.R. and Henikoff S., 1994. Expansions of transgene repeats cause heterochromatin formation and gene silencing in Drosophila. *Cell* **77**: 993-1002.

Edenberg H. J. and Huberman J. A., 1975. Eukaryotic chromosome replication *Annu. Rev. Genet* **9**: 245-284.

Editorials *Lancet*,1991 **338**: 603-604.

Ehrmann I.E., Ellis P.S., Mazeyrat S., Duthie S., Brockdorff N., Mattei M.G., Gavin M.A., Affara N.A., Brown G.M., Simpson E., Mitchell M.J. and Scott D.M., 1998. Characterisation of genes encoding translation initiation factor eIF-2γ in mouse and human: sex chromosome localization, escape from X-inactivation and evolution. *Hum. Molec. Genet.* **7**: 1725-1737.

Ellis N.A., 1998. The war of the sex chromosomes. *Nature Genetics* **20**: 9-10.

Ellison J.W., Wardak Z., Young M.F., Gehron Robey P., Laig-Webster M. and Chiong W., 1997. PHOG, a candidate gene for involvement in the short stature of Turner syndrome. *Hum. Molec. Gen.*, **6**: 1341-1347.

Emery D.E. and Sachs B.D., 1976. Copulatory behavior in male rats with lesions in the bed nucleus of the stria teminalis. *Physiol. Behav.* **17**: 803-806.

Epstein C.J., Smith S., Travis B. and Tucker G., 1978. Both X chromosomes function before visible X chromosome inactivation in female mouse embryos.

Nature **274**: 500-503.

Epstein C.J., Travis B., Tucker G. and Smith S., 1978. The direct demonstration of an X cfhroosome dosage effect prior to inactivation. *Genetic Mosaics and Chimeras in Mammals*, ed. L.B. Russel, p. 261-268. New York: Plenum.

Esposito T., Gianfrancesco F., Ciccodicola A., D'Esposito M., Nagaraja R., Mazzarella R., D'Urso M. and Forabosco A., 1997. Escape from X inactivation of two new genes associated with DXS6974E and DXS7020E. *Genomics* **43**: 183-190.

Fang P., Lev-Lehman E., Tsai T.-F., Matsuura T., Benton C.S., Sutcliffe J.S., Christian S.L., Kubota T., Halley D.J., Meijers-Heijboer H., Langlois S., Graham J.M. Jr, Beuten J., Willems P.J., Ledbetter D.H. and Beaudet A.L., 1999. The spectrum of mutations in UBE3A causing Angelman syndrome. *Hum. Mol. Genet.* **8**: 129-135.

Feinberg, A.P. and Vogelstein, B., 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132:** 6-13.

Feinberg, A.P. and Vogelstein, B., 1984. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **137:** 266-267.

Ferrigno P. and Silver P.A., 2000. Polyglutamine expansions: proteolysis, chaperones, and the dangers of promiscuity. *Neuron* **26**: 9-12.

Fichant G. and Gautier C., 1987. Statistical method for prediction protein coding regions in nucleic acid sequences. *Comput. Appl. Biosci.* **3**: 287-295.

Fickett J.W., 1982. Recognition of protein coding regions in DNA sequences. *Nucelic Acid Res.* **10**: 5303-5318.

Fialkow P.J., 1978. X-chromosome inactivation and the Xg locus. *Am. J. Hum. Genet.* **22**: 460-463.

Fisher E.M.C, Beer-Romero P., Brown L.G., Ridley A., McNeil J.A., Bentley Lawrence J., Willard H.F., Bieber F.R. and Page D.C., 1990. Homologous ribosomal protein genes on the human X and Y chromosomes: escape from X inactivation and possible implications for Turner syndrome. *Cell* **63**: 1205-1218.

Gasser S.M. and Laemmli U.K., 1987. Improved methods for the isolation of individual and clustered mitotic chromosomes. *Experimental Cell Research* **173**: 85-98.

Ghosh S. and Ghosh I., 1975. Variation of stemline karyotype in a HeLa cell line. *Z. Krebsforsch. Klin. Onkol. Cancer. Res. Clin. Oncol.* **84**: 129-133.

Gilgenkrantz S., Chery M., Tebou, M., Mujica P., Leotard B., Gregoire M.J., Boman F., Duprez A. and Hanauer A., 1990. Sublocalisation of the X breakpoint in the translocation (X;18)(p11.2; q11.2) primary change in synovial sarcomas. *Oncogene.* **5**: 1063-1066.

Gilgenkrantz S., Tridon P., Pinel-Briquel N., Beurey J. and Weber M., 1985. Translocation (X;9)(p11;q34) in a girl with incontinentia pigmenti: implications for the regional assignment of the IP locus to Xp11? *Ann. Genet.* **28**: 90-92.

Goldman M.A., Holmquist G.P., Gray M.C., Caston L.A. and Nag A., 1984. Replication timing of genes and middle repetitive sequences. *Science* **224**: 686-692.

Gong L., Li B., Millas S. and Yeh E.T.H., 1999. Molecular cloning and characterisation of human AOS1 and UBA2, components of the sentrin-activating enzyme complex. *FEBS Letters* **448**: 185-189.

Gong L. and Yeh E.T.H., 1999. Identification of the activating and conjgating enzymes of the NEDD8 conjugation pathway. *J. Biol. Chem.* **274**: 12036-12042.

Goodfellow P.N., Pym, B., Mohandas T. and Shapiro L.J., 1984. The MIC2 locus escapes X-inactivation. *Am. J. Hum. Genet.* **36**: 777-782.

Gooren L.J.G., 1990. The endocrinology of transsexualism: a view and commentary. *Psychoneuroendocrinology* **15**: 3-14.

Gorski, Boehnke, Reyner and Burright, 1992. A Radiation Hybrid Map of the Proximal Short Arm of the Human X Chromosome Spanning Incontinentia Pigmenti 1 (IP1) Translocation Breakpoints. *Genomics* **14**: 9-13

Goss S.J. and Harris H., 1975. New method for mapping genes on human chromosomes. *Nature* **225**: 680-684.

Graves J.A.M., Wakefield M.J. and Toder R., 1998. The origin and evolution of the pseudoautosomal regions of human sex chromosomes. *Hum. Molec. Genet* **7**: 1991-1996.

Grenfell S.J., Trausch-Azar J.S., Handley-Gerhart P.M., Ciechanover A., 1994. Nuclear localization of the ubiquitin-activating enzyme, E1, is cell-cycle dependent. *Biochem. J.* **300**: 701-708.

Guzman P. and Ecker J.R., 1988. Development of large DNA methods for plants: molecular cloning of large segments of *Arabidopsis* and carrot DNA into yeast. *Nucl. Acids Res.* **16**: 11091-11105.

Haas A.L. and Siepmann T.J., 1997. Pathways of ubiquitin conjugaton. *FASEB J.* **11**: 1257-1268.

Hagemann T., Surosky R., Monaco A.P., Lehrach H., Rosen F.S. and Kwan S-P., 1994. Physical mapping in a YAC contig of 11 markers on the human X chromosome in Xp11.23. *Genomics* **21**: 262-265.

Hamada H., Petrino M.G., Kakunaga T., Seidman M., Stollar B.D., 1984. Characterization of genomic poly(dT-dG)·poly(dC-dA) sequences: structure, organization, and conformation. *Mol. Cell Biol.* **4**: 2610-2621.

Handley P.M., Mueckler M., Siegel N.R., Ciechanover A. and Schwartz A.L., 1991. Molecular cloning, sequence, and tissue distribution of the human ubiquitin-activating enzyme E1. *Proc. Natl. Acad. Sci. USA* **88**: 258-262.

Hansen R.S., Canfield T.K., Fjeld A.D. and Gartler S.M., 1996. Role of late replication timiming in the silencing of X-linked genes. *Hum. Mol. Genet.* **5**: 1345-1353.

Hansen R.S., Canfield T.K., Fjeld A.D., Mumm S., Laird C.D. and Gartler S.M., 1997. A variable domain of delayed replication in FRAXA fragile X chromosomes: X inactivation-like spread of late replication. *Proc. Natl. Acad. Sci. USA* **94**: 4587-4592.

Hansen R.S., Canfield T.K. and Gartler S.M., 1995. Reverse replication timing for the *XIST* gene in human fibroblasts. *Hum. Molec. Genet.* **4**: 813-820.

Harris H. and Watkins J.F., 1965. Hybrid cells derived from mouse and man:

Artificial heterokaryons of mammalian cells from different species. *Nature* 640-646.

Hartung M. and Stahl A., 1975. Preleptotene chromosome condensation in mouse oogenesis. *Cytogenet. Cell Genet.* **18**: 309-319.

Hatton K.S., Dhar V., Brown E.H., Iqbal M.A., Stuart S., Didamo V.T. and Schildkraut C.L., 1988. Replication program of active and inactive multigene families in mammalian cells. *Mol. Cell. Biol.* **8**: 2149-2158.

Hershko A. and Ciehanover A., 1998. The ubiquitin system. *Annu. Rev. Biochem.* **67**: 425-479.

Herzing L.K.B., Romer J.T., Horn J.M. and Ashworth A., 1997. Xist has properties of the X chromosome inactivation centre. *Nature* **386**, 272-289.

Hieter P., Mann C., Snyder M. and Davis R.W., 1985. Mitotic Stability of Yeast Chromosomes: A Colony Color Assay that Measures Nondisjuncion and Chromosome Loss. *Cell* **40**: 381-392.

Hill L., Craig I.W., Asherson P., Ball D., Eley T., Ninomiya T., Fisher P.J., Turic D., McGuffin P., Owen M.J., Benbow C.P., Lubinski D., Thompson L.A., and Plomin R., 1999. DNA pooling and dense marker maps: a systematic search for genes for cognitive ablility. *NeuroReport* **10**: 843-848.Hochstrasser M., 2000. All in the ubiquitin family. *Science* **289**: 563-564.

Hochstrasser M., 2000. All in the ubiquitin family. *Science* **289**: 563-564.

Hong-Young K., Ontiveros S.D. and Strauss W.M., 2000. A revision of the human XIST gene organisation and structural comparison with mouse Xist. *Mam. Genome* **11**: 220-224.

Hubener C., Mincheva A., Lichter P., Schraven B. and Bruyns E., 2000. Genomic organization and chromosomal localization of the human gene encoding the T-cell receptor-interacting molecule (TRIM). *Immunogenetics* **51**: 154-158.

Ioannou P.A., Amemiya C.T., Garnes J., Kroisel P.M., Shizuya H., Chen C., Batzer M.A. and de Jong P.J., 1994. A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nat. Genet.* **6**: 84-89.

Jagiello G., Downey S. and Ducayen M., 1982. X-chromosome heteromorphism and appearance of meiosis in the rat fetal ovary. *Devel. Biol.* **90**: 223-225.

Jeffreys A.J., Wilson V., and Thein S.L., 1985. Hypervariable 'minisatallite" regions in human DNA. *Nature* **314**: 67-73.

Jha M.N., Bamburg J.R. and Bedford J.S., 1994. Cell cycle arrest by colcemid differs in human normal and tumor cells. *Cancer Res.* **54**: 5011-5015.

Johnston P.G. and Cattanach B.M., 1981. X chromosome activity in female germ cells of mice heterozygous for Searle's translocation T(X;16)16H. *Genet. Res.* **37**: 317-322.

Kajiwara K. and Mueller G.C., 1964. Molecular events in the reproduction of animal cells. III Fractional synthesis of deoxyribonucleic acid with 5-bromodeoxyuridine and its effect on cloning efficiency. *Biochem. Biophys. Acta* **91**: 486-493.

Kamitani T., Kito K., Nguyen H.P. and Yeh E.T.H., 1997. Characterisation of NEDD8, a developmentally down-regulated ubiquitin-like protein. *J. Biol. Chem.* **272**: 8557-28562.

Kamura T., Conrad M.N., Yan Q., Conaway R.C. and Conaway J.W., 1998. The Rbx1 subunit of SCF and VHL E3 ubiquitin ligase activates Rub1 modification of cullins Cdc53 and Cul2. *Genes Dev.* **13:** 2928-2933.

Kapitonov V.V., Holmquist G.P. and Jurka J., 1998. L1 repeat is a basic unit of heterochromatin satellites in cetaceans. *Mol. Biol Evol.* **15**: 611-612.

Kawakami M. and Kimura F., 1974. *Endocrinol. Jap.* **21**: 125-130.

Kawame H., Gartler S.M. and Hansen R.S., 1995. Allele-specific replication timing in imprinted domains: absence of asynchrony at several loci. *Hum. Molec. Genet.* **4**: 2287-2293.

Kay G.F., Ashworth A., Penny G.D., Dunlop M., Swift S., Brockdorff N. and Rastan S., 1991. A candidate spermatogenesis gene on the mouse Y chromosome is homologous to ubiquitin-activation enzyme E1. *Nature* **354**: 486-489.

Kay G.F., Barton S.C., Surani M.A. and Rastan. S., 1994. Imprinting and X chromosome counting mechanisms determine Xist expression in early mouse development. *Cell* **77**: 639-650.

Keller H.D., Green P., Helms C., Cartinhour S., Weiffenbach B., Stephens K., Keith T.P., Bowden D.W., Smith D.R., Lander E.S., Botstern D., Akots G., Rediker K.S., Gravius T., Brown V.A., Rising M.B., Parker C., Powers J.A., Watt D.E., Kauffman E.R., et al., 1987. A Genetic Linkage Map of the Human Genome. *Cell* **51**: 319-337.

Kelly R.L. and Kuroda M.I., 1995. Equality of X chromosomes. *Science* **270**: 1607-1610.

Keohane A.M., Barlow A.L., Waters J., Bourn D. and Turner B.M., 1999. H4 acetylation, *XIST* RNA and replication timing are coincident and define X;autosome boundaries in two abnormal X chromosomes. *Hum. Molec. Genet.* **8**: 377-383.

Keohane A.M., O'Neill L.P., Belyaev N.D., Lavender J.S. and Turner B.M., 1996. X inactivation and histone H4 acetylation in embryonic stem cells. *Dev. Biol.* **180**: 618-630.

Kertesz A., Kawarai M.D., Rogaeva E., George-Hyslop P.S., Poorkaj P., Bird T.D. and Munoz D.G., 2000. Familial frontotemporal dementia with ubiquitin-positive, tau-negative inclusions. *Neurology* **54**: 818-827.

Kishino T., Lalande M. and Wagstaff J., 1997. UBE3A/E6-AP mutations cause Angelman syndrome. *Nature Genet.* **15**: 70-73.

KitadaT., Asakawa S., Hattori N., Matsumine H., Yamamura Y., Minoshima S., Yokochi M., Mizuno Y. and Shimizu N., 1998. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature* **392**: 605-608.

Knight J C., Renwick P.J., Downing J.R. and Okuda T., 1995. Physical linkage of the cdc2-related gene (PCTK1) and the ubiquitin-activating enzyme E1 gene (UBE1) on human Xp11.3. *Cyto. Cell Genet* . **71**: 155-157.

Knoll J.H.M., Cheng S.-D. and Lalande M., 1994. Allele specificity of DNA replication timing in the Angelman/Prader-Willi syndrome imprinted chromosomal region. *Nat. Genet.* **6**: 41-46.

Kok K., Hofstra R., Pilz A., van den Berg A., Terpstra P., Buys C.H.C.M. and Carritt B., 1993. A gene in the chromosomal region 3p21 with greatly reduced expression in lung cancer is similar to the gene for ubiquitin-activating enzyme. *Proc.Nat. Acad. Sci.* **90**: 6071-6075.

Konopka A.K. and Smythers G.W., 1987. DISTAN- a program which detects significant distances between sort oligonucleotides. *Comput. Appl. Biosci.* **3:** 193-201.

Korenberg J.R. and Rykowski M.C., 1988. Human genome organisation: Alu, Lines, and the molecular structure of metaphase chromosome bands. *Cell* **53**: 391-400.

Kozaki Y., Kubo M., Fukuda Y., Onishi T. and Muramatu M., 1993. Amidinopiperidine-4-carboxylic acid 4-*tert*-butylphenyl ester, a trypsin inhibitor, suppresses the onset of DNA synthesis in HeLa cells synchronized by a double-thymidine block. *Biol. Pharm. Bull.* **16**: 558-564.

Kratzer P.G. and Chapman V.M., 1981. X chromosome reactivation in oocytes of Mus caroli. *Proc. Natl. Acad. Sci. USA* **78**: 3093-3097.

Kratzer P.G. and Gartler S.M., 1978. HGPRT activity changes in preimplantation mouse embryos. *Nature* **274**: 503-504.

Kratzer P.G. and Gartler S.M., 1978/2. Hypoxanthine guanine phosphoribosyl transferase expression in early mouse development. *Genetic Mosaics and Chimeras in Mammals*, ed. L.B. Russel, p. 247-260. New York: Plenum.

Kretz-Remy C. and Tanguay R.M., 1999. SUMO/sentrin: protein modifiers regulating important cellular functions. *Biochem. Cell Biol.* **77**: 299-309.

Kudo M., Iida Y. and Shimbo M., 1987. Syntactic pattern analysis of 5'-splice site of mRNA precursors in higher eukaryote genes. *Comput. Appl. Biosci.* **3**: 319-324.

Kumar S., Tomooka Y. and Noda M., 1992. Identification of a set of genes with developmentally down-regulated expression in the mouse brain. *Biochem Biophys. Res. Comm.* **185**: 1155-1161.

Kumar, S. Yoshida Y. and Noda M., 1993. Cloning of a c DNA which encodes a novel ubiquitin-like protein. *Biochem. Biophys. Res. Comm.* **195**: 393-399.

Kuspa A., Vollrath D., Cheng Y. and Kaiser D., 1989. Physical mapping of the *Myxococcus xanthus* genome by random cloning in yeast artificial chromosomes. *Proc. Natl. Acad. Sci. U.S.A.* **86**: 9817-8921.

Kwan S.-P., Lehner T., Hagemann T., Lu B., Blaese M., Ochs H., Wedgwood R., Ott J., Craig I.W. and Rosen F.S., 1991. Localisation of the gene for Wiskott-Aldrich syndrome between two flanking markers, TIMP and DXS255, on Xp11.22-Xp11.3. *Genomics* **10**: 29-33.

Lafrenière R.G., Brown C.J., Powers V.E., Carrel L., Davies K.E., Barker D.F. and Willard H.F., 1991. Physical mapping of 60 DNA markers in the p21.1 -> q21.3 region of the human X chromosome. *Genomics* **11**: 352-363.

Lander E.S. et al., 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860-921.

Larin Z., Monaco A.P. and Lehrach H., 1991. Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. *Proc. Natl. Acad. Sci. U.S.A.* **88**: 4123-4127.

Larsen F., Gundersen G. and Prydz H., 1992. Choice of enzymes for mapping based on CpG islands in the human genome. *GATA* **9**: 80-85.

Lee J.T. and Jaenisch R., 1997(1). Long range effects of ectopic X inactivation centres on a mouse autosome. *Nature* **386**: 275-279.

Lee J.T. and Jaenisch R., 1997(2). The (epi)genetic control of mammalian X-chromosome inactivation. *Curr. Opinion in Genet. & Dev.* **7**: 274-280.

Lee J.T., Davidow L.S. and Warshawsky D., 1999. Tsix, a gene antisense to Xist at the X-inactivation centre. *Nat. Genet.* **21**: 400-404.

Lee J.T., Murgia A., Sosnoski D.M., Olivos I.M. and Nussbaum R.L., 1992.

Construction and characterisation of a yeast artificial chromosome library for Xpter-Xq27.3: a systematic determination of co-cloning rate and X-chromosome representation. *Genomics.* **12**: 526-33.

Lee J.T., Strauss W.M., Dausman J.A. and Jaenisch R., 1996. A 450kb transgene displays properties of the mammalian X inactivation center. *Cell* **86**: 83-94.

Lewin Benjamin, 2000. Genes VII. Oxford University Press.

Lima-de-Faria, A., 1959. Differential uptake of tritiated thymidine into hetero- and euchromatin in *Melanoplus* and *Secale*. *J. Biophys. Biochem. Cytol.* **6**: 457-466.

Lindblad-Tho K., Winchester E., Daly M.J., Wang D.G., Hirschhorn J.N., Laviolette J.P., Ardlie K., Reich D.E., Robinson E., Sklar P., Shah N., Thomas D., Fan J.B., Gingeras T., Warrington J., Patil N., Hudson T.J. and Lander E.S., 2000. Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nat. Genet.* **24**: 381-386.

Lindsay S. and Bird A.P., 1987. Use of restriction enzymes to detect potential gene sequences in mammalian DNA. *Nature.* **327**: 336-338.

Littlefield J.W., 1964. Selection of hybrids from matings of fibroblasts in vitro and their presumed recombinants. *Science* **145**: 709-710.

Lopez R., Larsen F. and Prydz H., 1994. Evaluation of the exon predictions of the GRAIL software. *Genomics* **24**: 133-136.

Luciani J.M., Bezard J., Devictor-Vuillet M. and Mauleon P., 1979. ^{3}H-thymidine labelling pattern of preleptotene chromosome condensation stages in the foetal sheep ovary. *Ann. Biol. Anim. Biochem. Biophys.* **19(4B)**: 1241-1250.

Luoh, S.-W.; Jegalian, K.; Lee, A.; Chen, E. Y.; Ridley, A.; Page, D. C., 1995. CpG islands in human ZFX and ZFY and mouse Zfx genes: sequence similarities and methylation differences. *Genomics* **29**: 353-363.

Lyon M.F., 1961. Gene action in the X-chromosome of the mouse (*Mus musculus L.*). *Nature* **190**: 372.

Lyon, M.F., 1992. Some milestones in the history of X-chromosome inactivation.

Annu. Rev. Genet. **26**: 16-28.

Lyon, M.F., 1996. Pinpointing the centre. *Nature* **379**, 116-117.

Lyon, M.F., 1998. X-chromosome inactivation: a repeat hypothesis. *Cytogenet. Cell Genet.* **80**: 133-137.

Mamaeva S.E., Litvinchuk L.F. and Pinaev G.P., 1986. Karyotype characteristics of constant cell lines: II. Variability and chromosome content balance of M-HeLa cells. *Tsitologiya* **28**: 193-203.

Martin G.R., Epstein C.J., Travis B., Tucker G., Yatziv S., Martin D.W., Clift S. and Cohen S., 1978. X chromosome inactivation during differentiation of female teratocarcinoma stem cells *in vitro*. *Nature* **271**: 329-333.

Marahrens Y., Panning B., Dausman J., Strauss W. and Jaenisch R., 1997. Xist deficient mice are defective in dosage compensation but not spermatogenesis. *Genes Dev.* **11**: 156-166.

Matzke M.A. and Matzke A.J.M., 1995. Homology-dependent gene silencing in transgenic plants: what does it really tell us? *Trends Genet.* **11**: 1-3.

McBurney M.W. and Strutt B.J., 1980. Genetic activity of X chromosomes in pluripotent female teratocarcinoma cells and their differentiated progeny. *Cell* **21**: 357-364.

McConnkey E.H., Human Genetics, The Molecular Revolution. Jones and Bartlett Publishers, 1993.

McCormick M.K., Antonarakis S.E. and Hieter P., 1990. YAC cloning of DNA embedded in an agarose matrix. *Genet. Anal. Tech. Appl.* **7**: 114-118.

McGrath J.P., Jentsch S., and Varshavsky A, 1991. UBA1: an essential yeast gene encoding ubiquitin-activating enzyme. *EMBO J.* **10**: 227-236.

McMahon A., Fosten M. and Monk M., 1981. Random X-chromosome inactivation in female primordial germ cells in the mouse. *J. Embryol. Exp. Morphol.* **64**: 251-258.

Sternberg N., Ruether J. and DeRiel K., 1990. Generation of a 50000-member human DNA library with an average DNA insert size of 75-100 kbp in a bacteriophage P1 cloning vector. *The New Biol.* **2**: 151-162.

Stoddart K.L., Jermak C., Nagaraja R., Schlessinger D. and Bech-Hansen N.T., 1999. Physical map covering a 2Mb region in human Xp11.3 distal to DX6849. *Gene* **277:** 111-116.

Solovyev V., Salamov A. and Lawrence C.B., 1994. The prediction of human exons by oligonucleotide composition and discriminant analysis of spliceable open reading frames. *Ismb.* **2**: 354-3662.

Svaren J. and Calkley R., 1990. The stucture and assembly of active chromatin. *TIG* **6**: 52-56.

Swaab D.F. and Hofman M.A., 1995. Sexual differentiation of the human hypothalamus in relation to gender and sexual orientation. *Trends Neurosci.* **18**: 264-270.

Swanson D.A., Freund C.L., Ploder L., McInnes R.R. and Valle D., 1996. A ubiquitin C-terminal hydrolase gene on the proximal short arm of the X chromosome: Implications for X-linked retinal disorders. *Hum. Mol. Genet.* **5**: 533-538.

Tabor S. and Richardson C.C., 1990. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Effect of pyrophosphorolysis and metal ions. *J. Biol. Chem.* **265**: 8322-8328.

Tai H.H., Gordon J. and McBurney M.W., 1994. Xist is expressed in female embryonal carcinoma cells with two active X chromosomes. *Somat. Cell. Mol Genet.* **20**: 171-182.

Takagi N., 1974. Differentiation of X chromosomes in early female mouse embryos. *Exp. Cell Res.* **86**: 127-135.

Takagi N., 1978. Preferential inactivation of the paternally derived X-chromosome in mice. *Genetic Mosaics and Chimeras in Mammals*, ed. L.B. Russel, p. 341-360. New York: Plenum.

Takagi N. and Sasaki M., 1975. Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. *Nature* **256**: 640-642.

Takagi N., Suguwara O. and Sasaki M., 1982. Regional and temporal changes in the pattern of X-chromosome replication during the early postimplantation development of the female mouse. *Chromosome* **85**: 275-286.

Takahashi E.-I., Ayusawa D., Kaneda S., Itoh Y., Seno T., Hori T.-A., 1992. The human ubiquitin-activating enzyme E1 gene (UBE1) mapped to band Xp11.3 -> Xp11.23 by FISH. *Cytogenet. Cell Genet.* **59**: 268-269.

Thiselton D.L., McDowall J., Brandau O., Ramser J., d'Esposito F., Bhattacharya S.S., Ross M.T., Hardcastle A.J. and Meindl A., 2002. An Integrated, Functionally Annotated Gene Map of the *DXS8026-ELK1* Interval on Human Xp11.3-Xp11.23: Potential Hotspot for Neurogenetic Disorders. *Genomics* **79**: 560-572.

Timmer T., Terpstra P., van den Berg A., Veldhuis P.M.J.F., Elst A.T., ven der Veen A.Y., Kok K., Naylor S.L. and Buys C.H.C.M., 1999. An Evolutionary Rearrangement of the Xp11.3-11.23 Region in 3p21.3, a Region Frequently Deleted in a Variety of Cancers. *Genomics* **60**: 238-240.

Torchia B.S. and Migeon B.R., 1995. The XIST locus replicates late on the active X and earlier on the inactive X based on FISH DNA replication analysis of somatic cell hybrids. *Somat. Cell Mol. Genet.* **21**: 327-333.

Trask B.J., Massa H., Kenwrick S. and Gitschier J., 1991. Mapping of human chromosome Xq28 by two-color fluorescence *in situ* hybridisation of DNA sequences in interphase cell nuclei. *Am. J. Hum. Genet.* **48**: 1-15.

Trump D., Pilia G., Dixon P.H., Wooding C., Thakrar R., Leigh S.E.A., Nagaraja R., Whyte M.P., Schlessinger D. and Thakker R.V., 1996. Construction of a YAC contig and an STS map spanning 3.6Mb in Xp22.1. *Hum. Genet.* **97**: 60-68.

Turner, HH, 1938. A syndrome of infantilism, congenital webbed neck, and cubitus valgus. *Endocrinology* **23** :566-574.

van der Weele J.V., Gonen D., Leventhal B.L. and Cook E.H. Jr, 1999. Mutantion screening of the UBE3A/E6-AP gene in autistic disorder. *Mol Psychiartry* **4**: 64-67.

Venter J.C., Adams M.D., Myers E.W., Li P.W., Mural R.J., Sutton G.G., Smith H.O., Yandell M., Evans C.A., Holt R.A., Gocayne J.D., Amanatides P., Ballew R.M., Huson D.H., Wortman J.R., Zhang Q., Kodira C.D., Zheng X.H., Chen L., Skupski M. et al., 2001. The sequence of the human genome. *Science* **291**: 1304-1351.

Vu T.H. and Hoffman A.R., 1997. Imprinting of the Angelman syndrome (AS) gene, UBE3A, is restricted to brain. *Nat. Genet.* **17**: 12-13.

Wang J., Mager J., Chen Y., Schneider E., Cross J.C., Naby A. and Magnuson T., 2001. Imprinted X inactivation maintained by a mouse *Polycomb* group gene. *Nat. Genet.* **28**: 371-375.

Ward C.L., Omura S. and Kopito R.R., 1995. Degradation of CFTR by the ubiquitin-proteasome pathway. *Cell* **83**: 121-127.

Weber J.L., 1990. Informativeness of human (dC-dA)n·(dG-dT)n polymorphisms. *Genomics* **7**: 524-530.

Weber J.L. and May, P.E., 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* **44**: 388-396.

Wechsler D. Wechsler Intelligence Scale for Children - Revised. New York: The Psychological Corporation, 1974.

Weiss M.C. and Green H., 1967. Human-mouse hybrid cell lines containing partial complements of human chromosomes and functioning human genes. *Proc. Natl. Acad. Sci. U.S.A.* **58**: 1104-1111.

West J.D., Frels W.I., Chapman V.M. and Papaioannou V.E., 1977. Preferential expression of the maternally derived X-chromosome in the mouse yolk sac. *Cell* **12**: 873-882.

West J.D., Papaioannou V.E., Frels W.I. and Chapman V.M., 1978. Genetic Mosaics and Chimeras in Mammals, ed. L.B. Russel, p. 361-377. New York: Plenum

Wheatley D.N., 1990. Mitosis and protein synthesis 3 organelle relocation during normal and colcemid-arrested -phase in HeLa S-3 cells. *Cytobios* **63**: 109-130.

White L.M., Rogan P.K., Nicholls R.D., Wu B.L., Korf B. and Knoll J.H.M., Allele specific replication of 15q11-q13 loci: a diagnostic test for uniparental disomy. *Am. J. Hum. Genet.* **59**: 423-430.

Wilson E.B., 1911. The sex chromosomes. *Arch. Mikrosk. Anat Entwicklungsmech.* **77**: 249-271.

Witschi E., 1957. Sex chromatin and sex differentiation in human embryos. *Science* **126**: 1288-1290.

Woon, P.Y., Osoegawa K., Kaisaki P.J., Zhao B., Catanese J.J., Gauguier D., Cox R., Levy E.R., Lathrop G.M., Monaco A.P. and DeJong P.J., 1998. Construction and characterization of a 10-fold genome equivalent rat P1-derived artificial chromosome library. *Genomics* **50**, 306-316.

Wu J., Ellison J., Salido E., Yen P., Mohandas T. and Shapiro L.J., 1994. Isolation and characterisation of XE169, a novel human gene that escapes X-inactivation. *Hum. Molec. Genet.* **3** :153-160.

Wu J., Salido E.C., Yen P.H., Mohandas T.K., Heng H.H.Q., Tsui L-C., Park J., Chapman V.M. and Shapiro L.J., 1994. The murine Xe169 gene escapes X-inactivation like its human homologue. *Nature Genet.* **7**: 491-496.

Xiong Z., Tsark W., Singer S.J. and Riggs A.D., 1998. Differential replication timing of X-linked genes measured by a novel method using single-nucleotide primer extension. *Nucleic Acid Res.* **26**: 684-686.

Yang Z., Boffelli D., Boonmark N., Schwartz K. and Lawn R., Apolipoprotein (a) gene enhancer resides within a LINE element. *J. Biol. Chem.* **273**: 891-897.

Yeh E.T.H., Gong L. and Kamitani T., 2000. Ubiquitin-like proteins: new wines in new bottles. *Gene* **248**: 1-14.

Yeshaya J., Shalgi R., Shohat M. and Avivi L., 1999. FISH-detected delay in replication timing of mutated FMR1 alleles on both active and inactive X-chromosomes. *Hum. Genet.* **105**: 86-97.

Yorder J.A., Walsh C.P. and Bestor T.H., 1997. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* **13**: 335-340.

Zacksenhaus E. and Sheinin R., 1990. Molecular cloning, primary structure and expression of the human X linked A1S9 gene cDNA which complements the ts A1S9 mouse L cell defect in DNA replication. *EMBO J.* **9**: 2923-2929.

Zacksenhaus E., Sheinin R. and Wang H.S., 1990. Localization of the human A1S9 gene complementing the ts A1S9 mouse L-cell defect in DNA replication and cell cycle progression to Xp11.2 -> p11.4. *Cyto. Cell Genet* . **53**: 20-22.

Zhang J. and Madden T.L., 1997. PowerBLAST: a new network BLAST application for interactive or automated sequence analysis and annotation. *Genome Res.* **7**: 649-656.

Zhao L.P., Aragaki C., Hsu L. and Quiaoit F., 1998. Mapping of complex traits by single-nucleotide polymorphisms. *Am. J. Hum. Genet.* **63**: 225-240.

Zhou J.-N., Hofman M.A., Gooren L.J.G. and Swaab D.F., 1995. A sex difference in the human brain and its relation to transsexuality. *Nature,* **378**: 68-70.

Zinn A.R., Bressler., Beer-Romero P., Adler D.A., Chapman V.M., Page D.C. and Disteche C.M., 1991. Inactivation of the Rps4 gene on the mouse X chromosome. *Genomics* **11**: 1097-1101.

Zinn A.R., Page D. and Fisher E., 1993. Turner syndrome: the case of the missing sex chromosome. *Trends Genet.* **9:** 90-93.

Zuccotti M. and Monk M., 1995. Methylation of the mouse Xist gene in sperm and eggs correlates with imprinted Xist expression and paternal X-inactivation. *Nat. Genet.* **9**: 316-320.

Appendix

Primer sets and their conditions for PCR:

i. Primer sets obtained from G. Black

117.29 (1) 5' AAT TGG CTC CTT AAG ATT TAC 3'

117.29 (2) 5' TTA GTG GTT GCC TTA TAT AGT 3'

Standard conditions were used for PCR using T_a = 50°C.

II. Primer sets for sequencing:

i. Universal primers

universal primers (1) 5' GTT TTC CCA GTC ACG AC 3'

universal primers (2) 5' CAG GAA ACA GCT ATG AC 3'

ii. TA cloning kit primer set

M13 (-20) forward primer 5' GAG CGG GAG CAA AAT G 3'

M13 reverse primer 5' CAG GAA ACA GCT ATG AC 3'

iii. Primer sets to sequence left and right ends of YACs 2910 and 2911

The letter after the YAC name indicates the end (l = left, r = right) and the number indicates if it is the first or second primer of the set.

2910l1 5' CTC CAG TAG AAC ATC CCA AGA A 3'

2910l2 5' TCA CAC AAC AAC AAA ATC ACC T 3'

2910r1 5' TTA ACC TTT CCA GCC ACC AG 3'

2910r2 5' TTC TTT TCT TTC AGC CTT TTG C 3'

2911l1 5' ATT CTA CCA AAT ATT CAA GG 3'

2911l2 5' TCA TAA AGT GAA TTG AGA AG 3'

2911r1 5' TGG GTA AGC AGC TGT TAT 3'

2911r25' TCA AGC AGG TAC CAT GTG 3'

III. PAC end sequence primers

T7 5' TAA TAC GAC TCA CTA TAG GG 3'

SP6 5' ATT TAG GTG ACA CTA TA 3'

Growth curve of HeLa cells (1)

a series containing BrdU was set up at the same time as the cells in medium without BrdU

count time (hrs)	resusp. in (µl)	1 cells x10^4/ml	2 cells x10^4/ml	3 cells x10^4/ml	4 cells x10^4/ml	5 cells x10^4/ml	6 cells x10^4/ml	7 cells x10^4/ml	8 cells x10^4/ml	total cells x10^4/8ml	Ø cells x10^4/ml (total/8) *
0	2100	88	94	114	97					393	**98**
4	100	80	75	66	65	68	50	58	58	520	**65**
4 BrdU	100	58	57	51	49	54	55	55	54	433	**54**
8	100	60	54	70	59	54	58	48	68	471	**59**
8 BrdU	100	65	61	61	63	52	49	60	56	468	**59**
12	100	56	64	79	45	47	58	55	49	453	**57**
12 BrdU	100	48	52	44	41	54	41	60	40	380	**48**
16	100	61	95	49	50	82	60	63	55	515	**65**
16 BrdU	100	60	39	56	55	45	56	57	47	415	**52**
20	100	63	49	42	54	57	51	67	52	435	**54**
20 BrdU	100	61	63	61	69	52	56	64	50	476	**60**
24	100	59	60	52	63	69	50	53	58	522	**65**
24 BrdU	100	59	60	52	63	69	50	53	58	464	**58**
28	100	80	74	89	76	53	87	69	57	589	**74**
28 BrdU	100	64	31	52	51	41	50	32	36	359	**45**
32	100	67	75	80	74	83	71	74	80	610	**76**
32 BrdU	100	48	40	55	45	59	47	67	51	412	**52**

added about 10^5 cells to each flask

*: numbers have been rounded to nearest whole number

Growth curve of HeLa cells (2)

a series containing BrdU was set up at the same time as the cells in medium without BrdU

count time (hrs)	resusp. in (µl)	1 cells x10^4/ml	2 cells x10^4/ml	3 cells x10^4/ml	4 cells x10^4/ml	5 cells x10^4/ml	6 cells x10^4/ml	7 cells x10^4/ml	8 cells x10^4/ml	total cells x10^4/8ml	Ø cells x10^4/ml (total/8) *
0	2100	40	55	60	44	63	57	36	48	404*	**51**
12	100	60	39	52	40	52	38	48	37	367	**46**
12 BrdU	100	43	41	40	44	47	43	49	36	345	**43**
24	100	61	53	59	64	39	45	59	45	425	**53**
24 BrdU	100	50	41	56	43	31	36	35	33	325	**41**
30	100	56	64	41	56	49	49	56	54	427	**53**
30 BrdU	100	31	32	40	43	30	32	44	27	279	**35**
36	100	41	66	40	32	46	49	59	45	378	**47**
36BrdU	100	41	49	48	52	31	26	28	38	315	**39**
42	100	23	25	28	21	28	28	27	24	206	**26**
42 BrdU	100	27	36	18	20	25	22	21	23	192	**24**
48	100	47	46	47	37	43	50	52	34	356	**45**
48 BrdU	100	35	32	34	26	29	28	29	22	236	**30**
54	100	53	46	60	44	44	60	48	54	412	**52**
54 BrdU	100	23	24	26	24	16	23	20	25	182	**23**
60	100	69	73	75	51	72	67	60	66	534	**67**
60 BrdU	100	33	33	27	29	24	32	45	29	252	**32**

added 5 x10^4 cells to each flask
*: numbers have been rounded to nearest whole number

Table A1: Amount of cells, cells in metaphase and DNA concentrations obtained

exp. no.	total cell amount after tryp.	time = 0 total *	% of cells in metaphase	into each flask	time = 8 (supernat.) *	% of cells in metaphase	DNA obtained	DNA obtained (µg)	time = 16 (supernat.) *	% of cells in metaphase	DNA obtained	DNA obtained (µg)
1	not done	$2.1x10^6$	60.5	10^6	$7.8x10^5$	0	100ng		$6.0x10^5$	0	25ng	
2	$3.1x10^8$	$5.4x10^6$	0	$2.4x10^6$	$2.1x10^6$	0	500ng		$2.0x10^6$	0	0	
3	$2.4x10^8$	$7.8x10^6$	38.7	$3.5x10^6$	$3.0x10^6$	0.3	83µg		$3.4x10^6$	0	125µg	
4	$5.3x10^8$	$1.3x10^7$	22.0	$6.0x10^6$	$2.1x10^6$	1.3	5µg		$1.8x10^6$	0	75µg	
5	$3.0x10^8$	$1.3x10^7$	55.0	$6.1x10^6$	$4.6x10^6$	0	300µg**	U: 15, L: 1.5	$3.9x10^6$	0	300µg**	U: 15, L: 1.5
6	$3.4x10^8$	$1.1x10^7$	37	$5.2x10^6$	$1.6x10^6$	1.0	20µg		$3.3x10^6$	0	~ 2µg	
7	not done	$4.8x10^7$	24	$2.1x10^7$	$1.2x10^7$	14	**	U: 50, L: 0.5	$1.5x10^7$	11.5	**	U: 39, L: 0.4
8	not done	$4.2x10^6$	36	$1.9x10^5$	$2.2x10^6$	1	10µg **		$1.6x10^6$	0	10µg **	
9	$1.6x10^8$	$1.5x10^6$	43	$6.6x10^5$	$8.4x10^5$	0	19µg **		$8.3x10^5$	0	28µg **	
10	$2.1x10^8$	$4.5x10^6$	44	~ $2x10^6$	$1.4x10^6$	1	14.4µg **	U: 1.3, L: 0	$1.3x10^6$	0	28µg **	U: 2.9, L: 0
11	$1.2x10^8$	$5.3x10^6$	31	$2.6x10^6$	$1.8x10^6$	10	14µg **	U: 0.5, L: 0	$2.6x10^6$	0	49µg **	U: 22, L: 0
12	$2.9x10^7$	$4.4x10^6$	36	~$2.2x10^6$	$2.7x10^6$	1	38µg **	U: 13, L: 0.1	$2.7x10^6$	1	47µg **	U: 12.5, L: 0.5
13	$2.9x10^7$	~$5x10^6$	33	~$2.5x10^6$	$3x10^6$	1	44µg **	U: 10, L: 0.3	$3.2x10^6$	0	47µg **	U: 14, L: 0.1
14	$3x10^8$	~$9x10^6$	did not work	~$4x10^6$	$2x10^6$	did not work	26µg **	U: 16, L: 1+1	$2.4x10^6$	did not work	38µg **	U: 25.5 L: 1.5+>1
15	$4x10^8$	$7x10^6$	62	$3.3x10^6$	$1.2x10^6$	28	14µg **	U: 3.9, L: 0	$1.5x10^6$	20	18.5µg **	U: 9, L: 0.5
16		$7.6x10^6$	65	$3.8x10^6$	$1.5x10^6$		48.8µg (a)		$1.8x10^6$		54.6µg (b)	
17		$6.7x10^6$	51	$3.4x10^6$	$1.2x10^6$		21µg		$1.5x10^6$		40µg	
18		$3.7x10^6$	57	$1.9x10^6$	$7.1x10^5$		30µg (a)		$8.6x10^5$		39µg (b)	
19		$5.6x10^6$	76	$2.8x10^6$	$1.5x10^6$		47µg (a)		$1.6x10^6$		39µg (b)	
20		$8x10^6$	62	~$4x10^6$	$1.5x10^6$		31µg (a)		$1.5x10^6$		37µg (b)	
21		$6.1x10^6$	69	$3x10^6$	$1.2x10^6$		37µg (a)		$6.1x10^5$		35µg (b)	

*: 1/10 of the cells were taken for chromosome preps.

**: used DNA for CsCl centrifugation.

U: upper band of CsCl centrifugation, i.e. non BrdU incooporated DNA. Amounts in µg.

L: lower band of CsCl centrifugation, i.e. BrdU incooporated DNA. Amounts in µg.

NOTE: If do DNA extraction not the day after finishing BrdU exp., than recover less DNA

If cells "over"confluent, than get less synchronised cells (as for experiments 21-23)

Where the fields are not filled in no results were obtained

UK HGMP RESOURCE CENTRE
Funded by the MRC

Nix Results for:

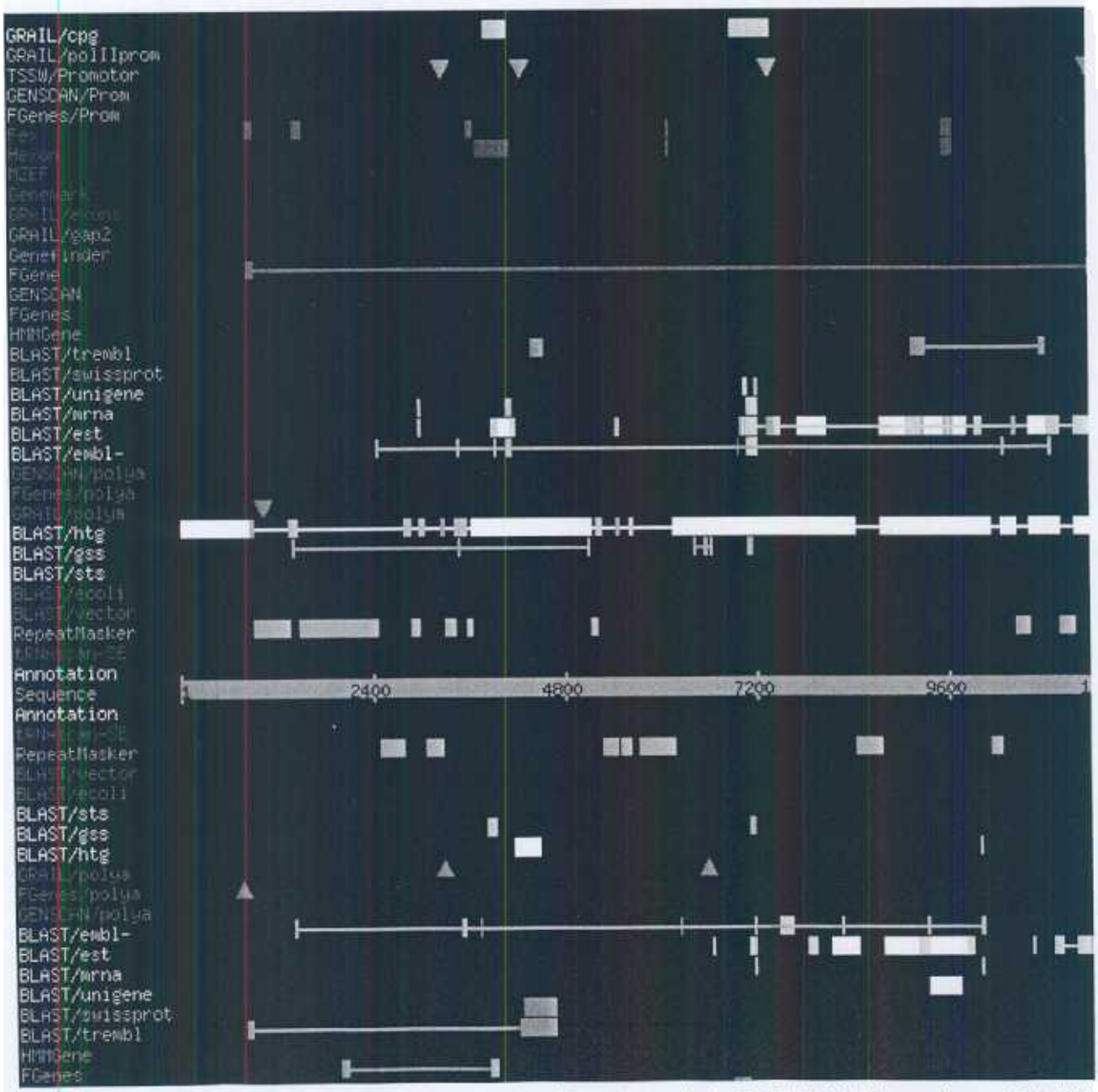
DXS_UBE1interval.seq

Information on the results viewer

Description: interval between DXS8237E and UBE1

Sequence range: 1 to 11988 (Full sequence)

If your browser has JavaScript enabled,
move your mouse over features in the image to see details



UK HGMP RESOURCE CENTRE
Funded by the MRC

Nix Results for: cutpaste-27611

Information on the results viewer

Description: end UBE1 to start codon PCTK1

Sequence range: 1 to 8418 (Full sequence)

[2470-2509] LINE/L2 (L2)

Description: End PCTK1 clone (X66363) in genomic sequence to start UHX1 exon ? (at least 3?)

Sequence range: 1 to 10170 (Full sequence)

If your browser has JavaScript enabled,
move your mouse over features in the image to see details