THE FRA16B LOCUS:

LONG RANGE RESTRICTION

MAPPING OF

16q13 - 16q22.1

BY

NARAS MYKOLAS LAPSYS. B.Sc. (Hons)

A thesis submitted for the Degree of Doctor of Philosophy

Department of Paediatrics, Adelaide Childrens Hospital,
The University of Adelaide

November 1993

Awarded 1994

TABLE OF CONTENTS

Page

STATEMENT

This work contains no material which has been accepted for the award of any other degree or diploma in any University or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying.

Naras Lapsys

ACKNOWLEDGEMENTS

PUBLICATIONS

Some of the work in this thesis has been published in the following paper.

Callen, D.F., Doggett, N.A., Stallings, R.L., Chen, L.Z., Whitmore, S.A., Lane, S.A., Nancarrow, J.K., Apostolou, S., Thompson, A.D., Lapsys, N.M., Eyre, H.J., Baker, E.G., Shen, Y., Holman, K., Phillips, H., Richards, R.I., and Sutherland, G.R. (1992).

High resolution cytogenetic - based physical map of human chromosome 16.

Genomics 13 : 1178 - 1185.

SUMMARY

Fragile sites are seen as abnormal features of chromosomes when viewed at the metaphase stage of mitotic cell division. They are classified into five specific types according to the tissue culture conditions under which they are induced. Fragile sites can be further classified into two classes, rare or common, depending on their frequency in the population.

At the commencement of this project no fragile site had been characterized at the molecular level. Since there was no information on the gene product to allow the direct cloning of a fragile site, alternative strategies were being applied. One approach was to attempt to identify fragile sites on the basis of their known chromosomal map positions by positional gene cloning.

The region of DNA encompassing 16q13-16q22.1 contained the rare distamycin A inducible fragile site FRA16B localized to the interface between 16q21 and 16q22.1. In addition, 16q22.1 contained an unusually large number of translocation breakpoints. A number of genes and anonymous DNA markers had been localized with respect to the translocation breakpoints and/or to FRA16B. The primary objective of this project was to construct a pulsed field gel electrophoresis (PFGE) derived long range restriction map of this region by physically linking adjacent DNA probes to common high molecular weight genomic DNA fragments. The construction of a contiguous map would physically order the probes along the chromosome and may physically define the probes flanking translocation breakpoints and more importantly the probes flanking FRA16B. Identification of the flanking markers and the distance separating them would direct the cloning of these DNA regions.

16q13-16q22.1 was estimated to be approximately 10 megabases in size. Upon completion of hybridization of all the suitable DNA probes to PFGE Southern blot filters, a non-contiguous long range restriction map totalling approximately 2.1 megabases of DNA was constructed for this region. PFGE mapping of genomic DNA did not physically link markers flanking FRA16B. However, DNA markers flanking the distal translocation breakpoint of an M4Eo leukaemia were found to be physically linked, at a maximum physical distance of 450 kb.

In attempts to overcome the limitations of long range restriction mapping using uncloned genomic DNA, two loci which mapped proximal and distal to FRA16B were chosen as initiation points for chromosome walking towards this fragile site using a newly developed yeast artificial chromosome (YAC) cloning system. A probe for each locus was used to screen a human genomic YAC library and a YAC clone was recovered from each marker. Neither YAC contained both proximal and distal markers and a comparison of restriction maps constructed from both YACs established that the YACs did not overlap and form a contig spanning FRA16B. *In situ* hybridization of the YAC clones to chromosomes expressing FRA16B indicated that they were unlikely to encompass the fragile site. Endclones were isolated from the human DNA inserts in each of the YACs, to be used as probes for the identification of overlapping YAC clones to continue the chromosome walk towards FRA16B.

The YAC walk was halted at this stage due to the arrival of a cosmid contig that had been mapped in close proximity to FRA16B. Preliminary physical mapping of the contig localized the cosmids to either side of the fragile site. However, detailed physical mapping identified a false overlap establishing that the cosmids did not form an uninterrupted series and that FRA16B was unlikely to have been cloned.

Endprobes were isolated from cosmids localized close to FRA16B. *In situ* hybridization of the endprobes to chromosomes expressing FRA16B determined that these cosmids did not span the fragile site and thus FRA16B had not been cloned.

CHAPTER 1

LITERATURE REVIEW

1.1   INTRODUCTION

Since their first discovery in 1965, the heritable sites of gaps and breaks in the human genome known as fragile sites have been the subject of clinical concern and laboratory investigation.   Fragile sites have been identified on nearly all chromosomes and a fragile site on the long arm of the X chromosome was found to be associated with a common form of X-linked mental retardation.   A combination of genetic and physical mapping studies initiated in the mid 1980's recently led to the first identification of a fragile site at a molecular level.   The molecular identification of additional fragile sites would greatly increase the understanding of these chromosomal mutations.

The literature review in this chapter can be divided into two main sections.   The first section defines, classifies and discusses the features of fragile sites on human chromosomes.   The second section discusses a strategy for the cloning of a fragile site and describes the advent and mechanisms of newly developed technologies which ultimately brought about the first molecular characterisation of a fragile site.   The second section also describes the molecular identification of two fragile sites as well as the main aims of this thesis.

1.2   THE DISCOVERY AND DEFINITION OF FRAGILE SITES

The first chromosomal lesion that was a fragile site was described in 1965 by Dekaban. Lejeune *et al* (1968) were the first to show that fragile sites were inherited chromosomal features by describing a constriction on the long arm of chromosome two in a woman and her daughter.

The association between mental retardation and fragile sites was first reported by Lubs (1969). Family studies demonstrated a pattern of mental retardation consistent with X-linked recessive inheritance associated with an unusual X-chromosome termed a "marker X". The morphology of this marker X-chromosome was described as a secondary constriction near the end of the long arm giving the appearance of large satellites (Lubs 1969). It took seven years for the association of the marker X and X-linked mental retardation to be confirmed by independent reports of Giraud *et al* (1976) and Harvey *et al* (1977). The delay in confirming Lub's original observation was due to the use of tissue culture conditions that suppressed the expression of the fragile site. Sutherland (1977) demonstrated that it was necessary to culture lymphocytes in medium 199 rather than other commercial culture media such as RPM1 1640, Ham's F10, Eagle's (basal) and CMRL 1969. In the 1970's many laboratories had ceased to use the medium 199 in favour of the newly developed culture media. It was soon discovered that the effectiveness of medium 199 for the induction of fragile site expression was its deficiency in folic acid and thymidine (Sutherland 1979a).

Based on fragile site distribution, cytogenetics and expression in lymphocytes (Sutherland 1979a, Sutherland 1979b), the definition of a fragile site was proposed. Fragile sites were described as specific non-staining points in the genome that displayed the following features:-

1.   In all affected cells of an individual, the fragile site always occurs in the same homologue and in the same position on the chromosome.

2.   The breaks or constrictions are of variable width usually involving both chromatids.

3.   The fragile site is inherited in a Mendelian co-dominant fashion.

4.   As a result of the fragile sites accentric fragments, deleted chromosomes and triradial figures develop (under appropriate tissue culture conditions).

1.2.1   General Features of Fragile Sites.

Fragile sites are specific points on chromosomes that are prone to breakage (Sutherland 1979a). All known chromosomal fragile sites can be induced *in vitro* to express themselves cytologically. Their expression is dependent on specific chemical agents or modifications of normal tissue culture conditions (Sutherland and Hecht, 1985).

Fragile sites exhibit a broad range of frequencies in the general population ranging from rare to common (see Table 1.1). The rare fragile sites are generally found in less than 1% of the general population and expression in individuals is most often heterozygous (Nussbaum and Ledbetter, 1986). Common fragile sites may be found in almost all individuals, present on both homologues (de la Chapelle and Berger, 1983).

The chromatin structure of one fragile site at Xq 27.3 has been analysed by scanning electron microscopy, SEM (Harrison *et al* 1983). The studies revealed that the most common morphology observed was that of a prominent isochromatid gap with the distal chromosome fragments well separated from the proximal long arm. The fragments often remained attached to the main part of the chromosome by individual fibres. Nussbaum and Ledbetter (1986) suggested that the SEM appearance of the

Table 1.1   Fragile Site Classification

Type	Class	No sites[a]	Carrier Frequency	Reference
Folate sensitive	Rare	18	-1/500[b]	Sutherland (1983)
Distamycin A - inducible	Rare	5	-1/20[c]	Schmid *et al* (1986)
BrdU inducible	Rare	2	-1/40	Sutherland *et al* (1984)
5-azaC inducible	Common	4	d	Sutherland *et al* 1985(b)
Aphidicolin - inducible	Common	75	d	Glover *et al* (1984)
BrdU inducible	Common	7	d	Sutherland *et al* (1985b)
Unclassified	Rare	1	-	-
Unclassified	Common	1	-	-

[a]   The number of fragile sites for each class. Values obtained from O'Brien (1989).

[b]   Approximate frequency for the group.

[c]   Frequency for FRA16B, in the German population.

[d]   May be present in all individuals.

fragile site is similar to the incompletely replicated regions of S-phase prematurely condensed chromosomes.

Aside from the fragile X which causes a common form of mental retardation, no pathogenic role for any other fragile site, rare or common, has been shown (Sutherland and Hecht, 1985). However, fragile sites have been implicated as predisposing factors in the pathogenesis of human cancer, the hypothesis based primarily on statistical evidence supporting an association between the locations of fragile sites and non random rearrangements associated with human cancer breakpoints (Yunis, 1983, Yunis and Soreng, 1984, Shabtai et al 1985, Sutherland and Hecht, 1985). Revised statistical analyses (Sutherland and Simmers, 1988) and gene mapping studies of fragile sites and proposed associated cancer breakpoints (Simmers et al 1987, Simmers and Sutherland, 1988, Puspurs et al 1988) indicated that the fragile sites under investigation were not directly related to cancer breakpoints and that the association was probably only circumstantial.

It has been shown that fragile sites predispose to intrachromosomal recombination as measured by sister chromatid exchanges (SCE), further supporting that fragile site expression is often, if not always, accompanied by DNA strand breakage (Glover and Stein, 1987).

Fragile sites have been used as markers for genetic linkage analysis studies in families with fragile sites (Magenis et al 1970, Mulley et al 1983, Sutherland et al 1982, reviewed Sutherland and Hecht, 1985). Mulley et al (1983) observed that fragile sites do not appear to disrupt recombination in chromosomal segments near the loci of expression, hence making them valid markers for linkage analysis giving unbiased recombination values.

Fragile sites are precisely mapped by chromosomal banding therefore human

metaphase chromosomes expressing fragile sites have been used for the regional localization of genes (Fratini *et al* 1986, Simmers *et al* 1986, Sutherland *et al* 1988, Lapsys *et al* 1992) and anonymous DNA markers (Callen *et al* 1988, Callen *et al* 1989) with respect to the site.

Recently, the cloning of FRAXA, the fragile site associated with fragile X syndrome, has been reported (Yu *et al* 1991). In addition, the cloning and sequencing of a second fragile site FRAXE has been achieved (Knight *et al* 1993). In both circumstances the site of fragility corresponds to an expanded unstable trinucleotide repeat sequence, p(CCG)n (Kremer *et al* 1991b, Knight *et al* 1993).

1.2.2   Classification of Fragile Sites.

Fragile sites are classified into five specific types - folate sensitive, distamycin A-inducible, 5-bromodeoxyuridine (BrdU) requiring, 5-azacytidine (5-azaC) inducible and aphidicolin inducible, according to the conditions of tissue culture under which they are expressed. They can be further classified into two major groups, rare or common depending on their frequency in the population (Table 1.1).

At the commencement of this study, 113 fragile sites had been identified on human chromosomes (Table 1.1, O'Brien, 1989). At its completion an additional folate sensitive fragile site at 1p 21.3 (Baker and Sutherland, 1991) and two folate sensitive fragile sites in Xq27 - q 28 (Flynn *et al* 1993, Hirst *et al* 1993) had been identified, bringing the total to 116.

1.2.3   The Folate Sensitive Fragile Sites.

Currently, 21 folate sensitive fragile sites have been located on 13 autosomes and the X-chromosome (O'Brien, 1989, Baker and Sutherland,1991, Flynn *et al* 1993, Hirst *et al* 1993). Two folate sensitive fragile sites FRAXA and FRAXE have been characterized at the molecular level (Yu *et al* 1991, Knight *et al* 1993, see

section 1.3.4).

The folate sensitive fragile sites were first shown to be expressed in lymphocytes cultured in the absence of folic acid and thymidine, or by the presence of inhibitors of folate metabolism, such as methotrexate (Sutherland, 1979a). The frequency of expression of fragile sites could be altered by changes in the composition of the culture medium in the final hours prior to harvest indicating that expression of the sites was determined at the time of spiralization of the chromatids in late S or early G2 phase (Sutherland, 1979a). The use of a thymidylate synthetase inhibitor fluorodeoxyuridine (FUdR) to induce folate sensitive fragile sites was independently demonstrated by Tommerup *et al* (1981) and Glover (1981). Glover (1981) also showed that the inhibiting effect of folic acid to fragile site expression but not that of thymidine could be negated by the addition of FUdR to the culture medium. A less cytotoxic thymidylate synthase inhibitor fluorodeoxycytidine (FCdR) was also found to induce fragile sites (Jacky and Sutherland, 1983). Nussbaum *et al* (1985) demonstrated that somatic cell hybrids made with a fragile X chromosome in a rodent cell deficient in thymidylate synthase also exhibited the fragile site when deprived of exogenous thymidine. These findings implied that a thymidylate depletion in the cell was essential for fragile site expression. It was proposed that the limiting of the thymidine monophosphate (dTMP) pool and thus the deoxythymidine triphosphate (dTTP) pool in the cell arrested DNA synthesis causing fragile site expression (Sutherland 1979a, Glover 1981).

Sutherland (1979a) further suggested that the expression of a fragile site was due to a section of thymidine rich DNA which could not complete DNA synthesis when the thymidine supply was restricted, thereby leaving gaps.

Thymidylate stress during DNA synthesis is also known to increase uracil misincorporation in the DNA (Graftstrom *et al* 1978). Although mechanisms exist for the removal of uracil residues in DNA, Krumdieck and Howard-Peebles (1983) proposed that inadequate removal of uracil bases in the fragile site DNA may disrupt chromatin conformation inducing the expression of a fragile site.

Experiments conducted by Sutherland *et al* (1985a) found that high concentrations of thymidine but not BrdU (an analogue of thymidine) also induced fragile site expression. This required a modification of the theory that dTMP depletion was the necessary condition for fragile site expression. It was proposed that the high concentrations of thymidine resulted in a feedback inhibition or thymidine block which led to a deficiency in deoxycytidine triphosphate (dCTP) for DNA synthesis. Sutherland *et al* (1985a) had found that the addition of deoxycytidine after high concentration of thymidine stopped fragile site expression. Therefore, not only a deficiency in dTTP but deficiency in dCTP for DNA synthesis could express a fragile site. From these data, Sutherland *et al* (1985a) proposed the folate sensitive fragile site to be a section of repetitive polypurine/polypyrimidine rich DNA such as poly d(AG)/polyd(TC).

DNA methylation has been considered a factor in FRAXA expression as a result of experiments involving the effects of methionine, a major methyl group donor. Methionine was shown to be required for the expression of FRAXA in peripheral lymphocyte cultures (Howard-Peebles and Pryor, 1981), in lymphoblastoid cell lines (Mixon and Dev, 1982, Mixon and Dev, 1983) and fibroblasts (Gardner *et al* 1982). The addition of 5-azaC and 5-adenosylhomocysteine (SAH) both inhibitors of DNA methylation, were shown to reverse the effects of methionine as well as that of FUdR (Mixon and Dev, 1983, Dev and Mixon, 1984). This was considered as supporting evidence for the involvement of DNA methylation in the expression of the FRAXA.

Mixon and Dev (1983) reported that methionine alone did not increase FRAXA expression but required FUdR and methionine together, indicating that methionine was not solely involved in DNA methylation. In contrast, Abruzzo *et al* (1985) found that different doses of methionine did not alter the frequency of FRAXA expression in the presence of FUdR. In addition, it was indicated that 5-azaC appeared to inhibit FRAXA expression only in high doses and that the inhibition might be a non-specific cytotoxic effect rather than a methylation effect (Abruzzo *et al* 1985). It has since been regarded that the methylation of DNA is unlikely to be the mechanism that disrupts chromatin conformation and produces a fragile site (Nussbaum and Ledbetter, 1986).

Caffeine, a known inhibitor of DNA repair mechanisms (Roberts, 1984) has been reported to enhance the expression of FRAXA in lymphocyte cultures (Yunis and Soreng, 1984) and somatic cell hybrids (Ledbetter *et al* 1986a,b), however attempts to replicate the caffeine enhancement in lymphocytes have not been confirmed (Glover *et al* 1984, Ledbetter *et al* 1986a, Abruzzo *et al* 1986). The somatic cell hybrids exhibiting caffeine enhanced FRAXA expression under thymidylate stress were characterized by a very short cell cycle (including G_2). Ledbetter *et al* (1986a) suggested that caffeine increased FRAXA expression by inhibiting the mitotic delay usually associated with DNA damage and thus interfered with the DNA repair process.

The relationship of cell density and fragile site expression has been examined (Brookwell and Turner, 1983, Krawczun *et al* 1986). It was shown that using FUdR to induce FRAXA, a higher frequency of expression was obtained by low density cell cultures. At lower densities FUdR remained in the medium to suppress thymidylate synthase and allowed the enhanced expression of the fragile site (Krawczun *et al* 1986).

The duration of culture time for FRAXA expression has also been reported. Sutherland (1983) found that 48 hour cultures resulted in lower frequencies of expression whereas

96 hour cultures provided higher frequencies of FRAXA expression. 72-96 hour cultures for the studies of FRAXA were also preferred by other investigators (Jennings *et al* 1980, Jacobs *et al* 1980, Gustavson *et al* 1981). Cultures of 120 hours or longer resulted in a decline in the frequency of expression along with the quality of the preparation (Sutherland 1983).

The association of the rare folate sensitive FRAXA with an X-linked mental retardation led many investigators to examine the behaviour of this fragile site in a foreign background. Human/rodent somatic cell hybrids were established with chromosomes containing FRAXA (termed fragile-X chromosomes) derived from affected males (Nussbaum *et al* 1983, Warren and Davidson 1984, Ledbetter *et al* 1986b, Lin *et al* 1987), transmitting males (Ledbetter *et al* 1986a,b,c) and carrier females (Ledbetter *et al* 1986b,c). Somatic cell hybrids were also established from X chromosomes from normal males (Warren and Davidson, 1984, Ledbetter *et al* 1986b). Ledbetter *et al* (1986a) observed that following caffeine enhancement of fragile site expression, the hybrids from normal males expressed the fragile site at low frequencies. It was concluded that non-fragile X males had a common fragile site at Xq 27.3 which fully mutates to a rare heritable fragile X. Recent studies by Sutherland and Baker (1990) indicated that the "fragile X" sites observed by Ledbetter *et al* (1986b) were in fact a common fragile site FRAXB located at Xq 27.2.

A reduced somatic cell hybrid containing the Xq24 - qter segment from a human fragile X chromosome has been constructed and was found to express FRAXA (Nussbaum *et al* 1986). FRAXA expression in the somatic cell hybrids containing only a single fragile X chromosome indicated that the presence of human autosomes was not necessary for expression of FRAXA (Nussbaum *et al* 1983). Furthermore, the expression of FRAXA in the reduced somatic cell hybrid suggested that the short arm and proximal long arm

of the X chromosome were not necessary for fragile site expression (Nussbaum *et al* 1986). Although the data did not eliminate the possibility of a distinct locus closely linked to FRAXA being responsible for fragile site expression, it did provide support that expression was an intrinsic property of the fragile site itself (Nussbaum *et al* 1986). In experiments designed to specifically induce chromosome rearrangements in hybrid cells containing fragile X chromosomes, Warren *et al* 1987 observed, by the use of RFLP markers flanking FRAXA, that these rearrangements frequently occurred at or near the fragile site. Somatic cell hybrids were identified containing either Xpter-Xq27 or Xq27-qter of a human fragile X chromosome (Warren *et al* 1987, Warren *et al* 1990). Expression of FRAXA was observed at the translocation junction, but in two of the hybrids significantly lower frequencies of expression were observed than in the intact fragile X chromosome of the parental hybrid (Warren *et al* 1987). These findings indicated that under certain culture conditions the fragile site was indeed fragile, demonstrating actual chromosomal breakage. In addition, the lower frequency of FRAXA expression at the translocation junction suggested that the fragile site may be composed of a repeated DNA sequence (Warren *et al* 1987). These translocation hybrids were used in mapping DNA markers very close to FRAXA and provided a reagent for the molecular cloning of FRAXA (See 1.3.4). Increased sister chromatid exchanges observed at the fragile X sites in affected males provided further support that breakage occurred at the fragile site (Wenger *et al* 1987).

Linkage analysis studies of FRAXA with RFLP's detected by anonymous DNAs and cloned genes played an integral role in diagnostic purposes such as carrier detection and prenatal diagnosis of the fragile site and were of major importance towards its molecular cloning. The first identified RFLP linked to FRAXA was reported by Camerino *et al* (1983). A polymorphic probe to the coagulation factor IX gene was found to be closely

linked to FRAXA. However subsequent studies showed that the results of Camerino *et al* were fortuitous and that the probe, F9, was)20 cM from the fragile site. As a result of these findings, other investigators attempted to isolate informative markers closer to FRAXA for linkage analysis. Polymorphic DNA probes exhibiting genetic linkage with FRAXA were isolated and ordered with respect to the fragile site (reviewed by Davies, 1986, Patterson *et al* 1987). The closest flanking markers to FRAXA in 1988 were mapped 7 cM proximal (Brown *et al* 1988) and 10-12 cM distal (Oberle *et al* 1985a). In 1989(a) Hyland *et al* isolated a series of anonymous DNA probes from a human/mouse hybrid containing only the region Xq26-qter. By genetic linkage analysis a marker (DXS296) was mapped 1-2 cM distal to FRAXA (Suthers *et al* 1989).

Before the cloning of the FRAXA mutation, highly polymorphic probes located very close to FRAXA were used in linkage analysis for the identification of transmitting males and carrier females and prenatal diagnosis (Oberle *et al* 1985b, Tommerup *et al* 1985, Oberle *et al* 1986, Mulley *et al* 1987, Veenema *et al* 1987). Importantly, in prenatal diagnosis genetic linkage analysis could only establish whether the fetus carried the FRAXA genotype. It could not distinguish between carrier or affected status.

Homozygous expression for the folate sensitive autosomal fragile sites have not been identified, therefore it is unknown whether such homozygosity would be deleterious (Sutherland and Hecht,1985).

1.2.4   The BrdU Induced Fragile Sites.

Bromodeoxyuridine (BrdU) is a halogenated thymidine analogue that has a number of effects on mammalian cells, including toxicity, mutagenesis, suppression of differentiation and induction of sister chromatid exchanges (reviewed by Goz 1978).

Only two BrdU induced fragile sites have been recorded. Localized to 10q25, FRA10B was independently reported by Scheres and Hustinx (1980) and Sutherland *et al* (1980).

A second BrdU requiring fragile site FRA12C was identified and localized to 12q24.2 (Voiculescu *et al* 1988).

FRA10B and FRA12C have been induced by fluorodeoxyuridine, FUdR (Gollin *et al* 1985a, Voiculescu *et al* 1988). FRA10B has also been induced by another halogenated pyrimidine iododeoxyuridine, IdU (Gollin *et al* 1985a) and bromodeoxycytosine, BrdC (Sutherland *et al* 1984). Being less toxic BrdC was found to be a preferred alternative to BrdU, inducing expression of FRA10B at slightly higher frequencies (Sutherland *et al* 1984). Spontaneous expression of FRA10B has been reported (Taylor and Bundey, 1983, Gollin *et al* 1985b) but at a very low frequency. Frequency of expression was significantly enhanced by the addition of BrdU. FRA10B was shown to be common in the Australian population with a heterozygote incidence in approximately 1 in 40 persons and was considered to constitute the first true chromosomal polymorphism to be described in humans (Sutherland 1982). Phenotypically normal homozygote individuals have been reported for FRA10B (Sutherland 1981) and FRA12C (Voiculescu *et al* 1991) indicating that there was no evidence to suggest that a phenotypic abnormality was associated with the expression of these fragile sites.

Cytogenetic studies of metaphase chromosomes induced with BrdU identified chromosome breakage at or very close to FRA10B, observed by sister chromatid exchanges in the region of the fragile site (Sutherland *et al* 1980).

BrdU enhancement of fragile site expression was found to be dependent on the amount incorporated into the DNA rather than the concentration in the cell (Gollin *et al* 1985a). The critical time of incorporation was between eight and nine hours prior to mitosis (Sutherland *et al* 1980, Gollin *et al* 1985a). Based on FRA10B expression by FUdR enhancement, Gollin *et al* (1985a) suggested that FRA10B is a region of DNA that has failed to complete DNA synthesis. When present in sufficient quantities, a

"thymineless" state is induced in the cell by FUdR blocking thymidylate synthase (Gollin *et al* 1985a). Thymidine stress is also required for the expression of folate sensitive fragile sites (Sutherland 1979a). However the BrdU induced fragile site FRA10B was not inhibited by folic acid nor induced in the absence of BrdU by methotrexate (Sutherland *et al* 1980).

1.2.5   The Distamycin A Inducible Fragile Sites.

A total of five rare fragile sites have been reported which are inducible by distamycin A (O'Brien, 1989, Table 1.1). They include FRA16B (Schmid *et al* 1980), FRA17A (Sutherland *et al* 1984) and three sites FRA8E, FRA11I and FRA16E that have only been described in the Japanese (Takahashi *et al* 1988a,b, Hori *et al* 1988a,b).

FRA16B, the first fragile site of this type was reported in 1970 by Magenis *et al*. A large kindred was observed in which a chromosome with a fragile site in the long arm of chromosome 16 was segregating without apparent phenotypic effect (Magenis *et al* 1970). It was shown that the fragile site segregated in a simple Mendelian fashion and was fully penetrant. Its location was measured and found to be 76% of the distance from the centromere on 16q (Hecht *et al* 1971). A detailed analysis with giemsa positive (R) and giemsa negative (G) banding showed this fragile site to be finely localised at 16q22.1, at the interface of q21 and q22 (Callen *et al*, 1988)

Expression of FRA16B was initially considered to be unaffected by conditions of tissue culture until in 1980 Schmid *et al* reported enhancement of expression of the fragile site at 16q22 with the oligopeptide antibiotic distamycin A. FRA16B and FRA17A have also been induced with the thymidine analogue BrdU (Croci 1983, Sutherland *et al* 1984). However, the three fragile sites in the Japanese population are not BrdU inducible (Takahashi *et al* 1988a,b, Hori *et al* 1988a,b). All five of the distamycin A inducible fragile sites have been induced with the DNA binding compounds Hoechst

33258, netropsin, berenil and diaminophenylindol, DAPI (Schmid *et al* 1984, 1986, 1987a, Sutherland *et al* 1984, Takahashi *et al* 1988a,b, Hori *et al* 1988a,b). FRA16B has also been induced with the antiviral agent interferon (Thestrup-Pedersen *et al* 1980, Shabtai *et al* 1983), a DAPI derivative D287/170 and methyl-green (Schmid *et al* 1986). Berenil was found to be the most effective inducing agent for FRA16B (Schmid *et al* 1986).

A population study of berenil induced FRA16B lymphocytes demonstrated a population frequency of 5.1%, making it the most frequent of all rare autosomal fragile sites (Schmid *et al* 1986, Table 1.1). Spontaneous induction of FRA16B has been repeatedly reported but occurs rarely (Magenis *et al* 1970, Sutherland 1982, Croci 1983, Schmid *et al* 1984).

FRA17A has been induced in fibroblast culture by BrdU (Sutherland *et al* 1984). Lymphoblastoid cell lines derived from heterozygous carriers for FRA8E, FRA16B and FRA16E have been established that retain their original phenotype (Tsuji *et al* 1991).

With the exclusion of interferon and BrdU, the treatment of lymphocytes with the various classes of DNA binding compounds showed that fragile site expression is only induced with A-T specific DNA ligands that bind externally to DNA without intercalation (Schmid *et al* 1986). A-T specific compounds known to interact with DNA by intercalation such as quinacrine mustard and daunomycin did not enhance FRA16B expression (Schmid *et al* 1986). G-C specific DNA ligands methramycin and chromomycin A_3 also failed to enhance FRA16B expression (Schmid *et al* 1986).

It was proposed that the ability of the A-T specific compounds to induce FRA16B and FRA17A indicate that a region of repetitive A-T rich DNA sequences are located at these fragile sites (Schmid *et al* 1986, 1987a). The attachment of the A-T specific non-intercalating ligand to the DNA sequence during S and/or G_2 phase may inhibit

chromosomal compaction for mitosis and result in expression of a fragile site (Sutherland *et al* 1984, Schmid *et al* 1986). The observation that maximum FRA16B expression was achieved when BrdU was added to the cultures close to the time of harvest further indicated a late replication of the A-T rich sequences (Sutherland *et al* 1984).

It has also been proposed that FRA16B expression is caused by integrated viral DNA (Sutherland 1979a, Sorensen *et al* 1979). Shabtai *et al* (1983) reported that both interferon and distamycin A have antiviral activity and that distamycin A triggers the excision of an integrated viral genome or allows expression of viral function by changing the hosts cellular activity.

Schmid *et al* (1987b) observed that the rate of sister chromatid exchanges (SCE) at FRA16B were increased following treatment with berenil and BrdU or with BrdU alone. These findings were supported by Lukusa *et al* (1991) who reported increased SCE in spontaneously expressed FRA16B. Schmid *et al* (1987b) proposed that SCE's preferentially occur in chromosomal regions where segments with highly different DNA base pair compositions are joined, supporting the assumption that FRA16B is a very short A-T rich DNA sequence surrounded by euchromatin. Lukusa *et al* (1991) concluded that at some point during the expression of fragile sites FRA8E, FRA16B and FRA16E a DNA break occurs because DNA fragility and SCE are variable expressions of the same mechanisms.

The first established linkage of a gene locus to a fragile site was observed for FRA16B (Magenis *et al* 1970). Since the location of FRA16B could be determined with precision, this chromosomal marker permitted the localisation of the gene for the alpha-polypeptide chain of haptoglobin (Magenis *et al* 1970). Polymorphic markers mapping to bands 16q21 and 16q22 have been examined for genetic linkage to FRA16B (Mulley

et al 1989). *In situ* hybridization using cells expressing the fragile site <u>FRA16B</u> have been used to directly localize both polymorphic and non-polymorphic anonymous DNA markers with respect to the fragile site (Callen *et al* 1988, Hyland *et al* 1989a,b). Homozygous expression for <u>FRA16B</u> (per comm. Sutherland) and <u>FRA17A</u> (Izakovic 1984) have been described. All individuals were phenotypically normal indicating that homozygosity for these fragile sites is harmless.

1.2.6 The Common Fragile Sites.

Many studies have shown that "spontaneous" chromosome gaps or breaks initially termed "hot spots" or "hot points" have been distributed non-randomly along human chromosomes (Aula and von Koskull, 1976, Ayme *et al* 1976, Mattei *et al* 1979). These breaks were often observed under conditions of thymidylate stress created by folic acid deprivation (Sutherland 1983). In 1984, Glover *et al* found that these regions could be induced by the addition of aphidicolin, and termed them aphidicolin induced "common fragile sites". Expression of these sites have also been induced by MTX (Barbi *et al* 1984) caffeine (Yunis and Soreng, 1984), uridine (Li and Zhou, 1985), 1-B-D-arabinofuranosyl-cytosine, araC (Li *et al* 1986a), the addition of cytidine or guanosine to medium free of folic acid (Li *et al* 1986b) and FUdR (Smeets *et al* 1986). At the commencement of this study, a total of 75 aphidicolin induced common fragile sites had been reported (O'Brien, 1989, Table 1.1).

The observations that common fragile sites were induced under conditions of thymidylate stress and fragile site expression was enhanced by aphidicolin led Glover *et al* (1984) to propose a mechanism by which these fragile sites may be expressed. Due to aphidicolin being a specific inhibitor of DNA polymerase-alpha which is primarily associated with chromosomal DNA repair, it was suggested that common fragile sites are induced by inhibition of replication fork progression or joining of DNA

intermediates preferentially at these sites (Glover *et al* 1984). In addition thymidylate stress condition may also partially inhibit polymerase alpha (Glover *et al* 1984).

The induction of common fragile site FRA3B in folate deficient medium by araC and caffeine, both inhibitors of DNA polymerase alpha (Yunis and Soreng, 1984, Li *et al* 1986a), provided further support that the inefficiency of the DNA repair process plays an important role in the expression of common fragile sites.

Li and Zhou (1985) suggested that the expression of common fragile sites by uridine induction was the result of uridine unbalancing the nucleotide pools, hence enhancing the misincorporation of uracil into DNA resulting in DNA breakage. It was shown that the effect of uridine on chromosome fragility could be reversed by the addition of low concentrations of thymidine prior to two to three hours before harvest (Lin *et al* 1987).

Yunis and Soreng (1984) and Daniel (1986) reported a total of six common fragile sites mapping to the same chromosomal positions as rare heritable fragile sites. Daniel (1986) proposed that mutational events at the common fragile sites may generate the rare fragile sites.

Aphidicolin-induced expressed fragile sites were found to be associated with SCE 's (Glover and Stein, 1987, Hirsch, 1991). Hirsch (1991) extended these observations to non-expressed fragile sites supporting the hypothesis that aphidicolin induced common fragile sites are preferred sites for the induction of chromatid breaks and SCE.

The common fragile site FRA3B was found to be the most readily induced common fragile site of this type (Glover *et al* 1984, Smeets *et al* 1986). Smeets *et al* (1986) observed that in a study of 70 normal healthy subjects, every individual examined was shown to express the fragile site. It was observed that in a given person homozygous expression for FRA3B was approximately 50%, suggesting that the variation in expression is less caused by inherited molecular variation than by environmental factors

(Smeets *et al* 1986). Sutherland and Hecht (1985) suggested that the common fragile sites are highly prevalent, being present in all individuals in a homozygous form.

The two other types of common fragile sites recorded are four sites induced by 5-azaC and two sites induced by BrdU (Sutherland *et al* 1985b, Table 1.1).

1.2.7 Fragile Sites in Animals.

Relatively few reports have been published on the occurrence of fragile sites in species other than man, therefore the evolutionary course of this type of human genomic instability remains largely unknown.

Common fragile sites have been identified in gorilla, chimpanzee and orangutan (Yunis and Soreng, 1984, Schmid *et al* 1985, Smeets and van de Klundert, 1990) demonstrating that fragile sites in primates are conserved among closely related species. Conservation of common fragile sites have also been reported in mammalian karyotypes. Common fragile sites have been reported in the Persian vole Ellobius lutescens (Djalali *et al* 1985), the laboratory mouse (Djalali *et al* 1987, Hunt and Burgoyne, 1987, Djalali *et al* 1990), the laboratory rat (Robinson and Elder, 1987), the domestic dog (Stone *et al* 1991a,b), racoon dogs (Wurster-Hill *et al* 1988), the pig (Riggs and Chrisman, 1989) and the rabbit (Poulsen and Ronne, 1991). Putative common fragile sites have also been reported in the horse karyotype (Ronne 1992).

BrdU induced fragile sites have been identified in the genome of the chinese hamster (Hsu and Somers, 1961), the cactus mouse (Schneider *et al* 1980) and recently in the dog (Stone and Stevens, 1993).

A fragile site has been induced by distamycin A in the chromosome 17 of the gorilla (Schmid *et al* 1981). Hoechst 33258 has been used to induce specific fragile site - like lesions in the chromosomes of marsupials of the family Macropodidae (Hayman and Sharp, 1981).

The evidence of chromosome fragility observed over the large range of mammalian genomes demonstrate that fragile sites are conserved among chromosomally distant related species.

1.3   POSITIONAL GENE CLONING

Positional gene cloning or "reverse genetics" is a strategy designed for the isolation of genes without known gene products. Since no information on gene product is available to allow for its direct cloning, the gene is identified on the basis of its chromosomal map position without direct knowledge of its function.

The initial step in the isolation of a gene by positional cloning is to map the gene to a particular chromosome by linkage analysis. Its regional localization can be further refined by the construction of a detailed linkage map around the gene. However, upon the completion of the linkage map markers flanking the gene may be separated by distances of up to several million base pairs. The resolution of linkage maps is in recombination units (cM) which correspond, on the average, to 1×10^6 bp of DNA (Barlow and Lehrach, 1987). These distances are greater than the amounts of DNA that can be cloned into conventional vectors such as cosmids for restriction mapping. Ultimately though, detailed restriction maps are required to estimate the distance between closest flanking markers to guide cloning of the interval between them and to identify candidate sequences for the gene.

The gap between the linkage map and the cloning capacity of conventional vectors can be closed by the use of a technique called pulsed field gel electrophoresis (PFGE), which allows the analysis of DNA fragments of up to millions of base pairs in size (Schwartz *et al* 1982).

1.3.1   Pulsed Field Gel Electrophoresis.

Agarose gel electrophoresis is one of the most widely used separation techniques with an enormous range of biochemical applications.   However, conventional electrophoresis separations of native DNA molecules has a practical resolution limit of 40-50 kb (Fangman, 1978).   In 1982, Schwartz *et al* introduced the concept of pulsed field gel electrophoresis PFGE, whereby molecules greater than 50 kb in size could be separated by using electrode configurations that generated electric fields in alternating orientations. Since then, many types of instruments have been developed based on these principles and DNA separations of up to 12 Mb have been achieved (Orbach *et al* 1988).

1.3.2   Development of PFGE.

The inability of conventional agarose gel electrophoresis to resolve large DNA molecules is believed to be due to a phenomenon called biased reptation.   Long DNA strands must snake their way through gel pores with one end leading and with the rest of the molecule following the same path.   Once DNA molecules become larger than the pores in the agarose, they have to distort their shape to enter the gel, and then the sieving power of the matrix becomes irrelevant resulting in size independent mobilities (Lerman and Frisch, 1982, Lumpkin *et al* 1985).   Fangman (1978) found that reducing the agarose concentration and reduction in the electrophoretic voltage during conventional electrophoresis could extend the range of separation.   However run times of up to 92 hours were necessary and a maximum resolution of 500 kb was obtained.

Klotz and Zimm (1972) demonstrated that the conformation of a DNA molecule was perturbed by a voltage gradient and that after removal of the electric field the DNA molecules relaxed back to their unperturbed state.   The measured relaxation time was found to be very sensitive to the molecular weight of the DNA (Klotz and Zimm, 1972). Schwartz *et al* (1982) attempted to regain high molecular weight DNA resolution by

perturbing the orientation and extension in gel matrices. To create a gel electrophoretic analogue exploiting the relaxation properties of large DNA molecules, Schwartz subjected DNA in agarose gels to alternatively pulsed, perpendicularly oriented electric fields. The mechanism by which this method resolves high molecular weight DNA is not understood with certainty but Schwartz *et al* (1982) proposed the following theory: when large DNA enters a gel in response to an electric field the DNA must elongate parallel to the field. The field is shut off and a new field is applied perpendicular to the long axis of the DNA. The coiled molecule now may lie across the openings of several pores in the gel. It will have to reorient and enter one of them to move efficiently in response to the new field. Schwartz *et al* (1982) found that the time required for reorientation was very sensitive to molecular weight. Larger DNA molecules took a longer time to realign after the fields were switched than smaller molecules.

The techniques of PFGE were first applied to the separation of yeast chromosomes in 1984 by Schwartz and Cantor. By employing alternatively pulsed, perpendicularly oriented electric fields, one homogeneous and the other non homogeneous, DNA separations from 30-2000 kb were achieved with fractionated intact <u>Saccharomyces cerevisiae</u> chromosomes. Also in 1984, Carle and Olson developed a similar apparatus which utilized two non-homogeneous electric fields, called orthogonal-field-alternation gel electrophoresis (OFAGE). Both of these apparatuses produced skewed DNA migration patterns because the electric fields were not uniform. DNA molecules migrated at different rates depending on their location in the gel, hindering lane to lane comparisons and limiting the number of available lanes in each gel. McPeek *et al* (1986) found that these disadvantages could be reduced by modifying the electric field geometry of OFAGE together with an altered switch pattern. Resolution was improved and DNA molecules followed a migration path that was straighter down the gel.

Gardiner *et al* (1986) modified the PFGE system by orienting the electric field transversely to the gel and since the electric fields produced were homogeneous across the gel, the characteristically skewed lanes were straightened. However, because of the varied angle from the top to the bottom of the gel, the molecules did not move at a constant velocity.

In attempts to make further improvements in large scale electrophoresis, Carle *et al* (1986) discovered that the electrophoresis of large DNA molecules could be made strongly size dependent by periodically inverting a uniform electric field in one direction (i.e. 180°). Termed field-inversion gel electrophoresis (FIGE), forward migration was achieved when a longer portion of each switching cycle was set to forward than reverse. Similar results were obtained with a constant switching interval in combination with different forward and reverse voltages. Increasing the net forward switching cycle or voltage increased the molecular weight range of size separation (Carle *et al* 1986). Carle *et al* (1986) also found that molecules of different molecular weights could unexpectedly comigrate under FIGE conditions. The problem was overcome by progressively changing the switching interval during electrophoresis, a technique known as switch time ramping. Ellis *et al* (1987) found that switch time ramping did not fully overcome this phenomenon. Using a range of lambda oligomers Ellis found that ramped runs extended the range of oligomers separated in sequential order of size, but with the largest oligomers out of sequence. However, FIGE was found to yield better resolution for molecules smaller than 750 kb than all the previous PFGE apparatuses (Lai *et al* 1989).

By applying the principles of electrostatics to gel electrophoresis, Chu *et al* (1986) developed an electrophoretic technique that generated a homogeneous electric field. The electric field was generated by a method in which multiple electrodes (24 in all) were

arranged along a polygonal contour and clamped to predetermined electric potentials. Alternation in the orientation was achieved by electronic switching. A square array of multiple electrodes generated a reorientation in the electric field of 90°. A hexagonal array generated reorientation angles of either 120° or 60° depending on the placement of the gel with respect to the hexagon and the assignment of polarity to the electrodes. DNA was separated in straight lanes using this contour-clamped homogeneous electric field, CHEF (Chu *et al* 1986). Southern *et al* 1987 also developed a method of electrophoretic separation using homogeneous electric fields. Called crossed field gel electrophoresis, the orientation of the electric field in relation to the gel was changed by discontinuously and periodically rotating the gel. Due to the inability to switch the field quickly, DNA molecules less than 50 kb in size could not be resolved. Another system based on homogeneous electric fields called homogeneous orthogonal field gel electrophoresis, PHOGE has been developed (Bancroft and Wolk, 1988). A 90° field orientation was applied with the DNA molecules undergoing four reorientations per cycle instead of two. The separation was described as a similar effect of combining pulsed field and field inversion geometry. With similarities to the CHEF system, a programmable autonomously controlled electrode (PACE) system has been constructed (Clark *et al* 1988). Twenty-four electrodes were arranged in a closed contour with the homogeneous electric field being controlled by independent regulation of the voltages of each electrode. An unlimited number of electric fields of controlled homogeneity could be generated, including OFAGE, FIGE, PHOGE and CHEF.

Using CHEF based PFGE systems very high molecular weight DNA separations have been achieved. For example Saccharomyces pombe chromosomal DNA have been resolved (Vollrath and Davis, 1987). The largest of these molecules was estimated to be 7 Mb in size. Orbach *et al* (1988) used similar technology to resolve the chromosomes

of the filamentous ascomycete <u>Neurospora crassa</u>, which range in size from four to 12.6 Mb.

In addition to the biased reptation model for DNA reorientation in PFGE, two other classes of models have been proposed to provide a quantitative explanation of PFGE. These are the switchback model (Southern *et al* 1987) and a hairpin extension model (Deutsch, 1988). These two classes of models assume linear configurations of DNA in the gel once reorientation is complete (Southern *et al* 1987, Deutsch, 1988). Deutsch (1988) proposed that when the field is rotated, a nucleation step occurs with hairpin kinks budding out at various points on the DNA chain. The kinks compete with each other, some growing longer at the expense of others, ending up with a single loop. This in turn may unfold to yield a linear molecule. The switchback model postulates that moving a kink through a gel would require more energy than moving an end (Southern *et al* 1987). It suggests that for each alternate pulse, the leading and trailing ends of the chain must reverse if the electric fields are at obtuse angles. Reorientation occurs as the new leading end pulls the rest of the molecule out of its previous configuration and into a new direction. Further understanding of the mobility of large DNA molecules through agarose gel matrices was obtained by the observations of individual DNA molecules undergoing conventional electrophoresis and PFGE, viewed with the aid of a fluorescent microscope (Smith *et al* 1989). DNA molecules were found to advance lengthwise through the gel in an extended configuration and the DNA molecules fluctuated between compact and extended states during electrophoresis. The molecules were often observed becoming hooked around obstacles in a U-shape for extended periods and the molecules displayed elasticity as they extended from both ends at once (Smith *et al* 1989).

Due to the large sizes of DNA molecules that can be separated by PFGE, careful preparation of the DNA samples has become a critical procedure. The average size of

DNA molecules prepared in solution by conventional techniques is usually less than 500 kb because the DNA is vulnerable to hydrodynamic shearing (Fangman, 1978). To preserve the integrity of large DNA molecules, Schwartz and Cantor (1984) developed a method of isolating DNA from intact cells after they had been embedded in agarose blocks. The agarose matrix stabilized the DNA molecules after the removal of cell membranes and proteins. This method has been applied to the preparation of DNA from the cells of yeast (Carle and Olson, 1985, Schwartz and Cantor, 1984) and mammalian DNA (Yanamandra and Lee, 1989). Whilst still in agarose the high molecular weight DNA could be run in gels as intact molecules or be digested by restriction enzymes. As a result of the large size and limited surface area of agarose blocks, large amounts of proteinase K and restriction enzymes were needed for complete digestion of cellular material and the DNA. Overhauser and Radic (1989) developed a method for encapsulating tissue culture or yeast cells in agarose beads. Agarose beads increased the surface area available and allowed the rapid diffusion of enzymes resulting in shorter digestion times and reduced enzyme concentrations.

The need to subdivide chromosome sized DNA molecules into large defined fragments requires restriction endonucleases with infrequently occurring recognition sites. Prior to the introduction of PFGE technology only a few restriction enzymes cut mammalian DNA at low frequency. These included the only two known enzymes with eight-base recognition sites (*Not* I and *Sfi* I) and a few six-base cutters with CpG in their recognition sites (Lai, 1991). Restriction endonuclease recognition sequences that contain CpG are very rare in mammalian genomes (Brown and Bird, 1986, Lindsay and Bird, 1987). With the expanded separation capability of PFGE a number of additional infrequent-cutting restriction enzymes have been isolated since 1985.

DNA markers of known sizes are vital to all pulsed field gel applications. By preparing

concatemers of lambda phage DNA, molecules of a known size which span a size range of 50 to over 1000 kb, in 50 kb increments have been developed (Waterbury and Lane, 1987). Chromosomes from <u>Saccharomyces cerevisiae</u> spanning a size range from approximately 200 kb - 3 Mb (Carle and Olson, 1985) have also commonly been used as size markers. <u>Saccharomyces pombe</u>, with chromosomes ranging from 3 to 7 Mb may be used as size markers for the separation of megabased size DNA molecules (Vollrath and Davis, 1987).

1.3.3   Factors Affecting PFGE.

The resolution of large DNA molecules by PFGE produces three zones of separation (Vollrath and Davis, 1987, Southern *et al* 1987). The first two zones exhibit a linear decrease in mobility of DNA with increasing molecular weight. Very large DNA molecules do not resolve at all and bunch together in the compression region or zone three. Matthew *et al* (1988b) showed that DNA size separation in zone two is better than in zone one by a factor between 1.5 and 2. It is necessary to adjust PFGE parameters to enhance DNA separation of a desired size range within this zone.

The most important determinant of mobility in PFGE is the switching interval of the electric field. Schwartz and Cantor (1984) demonstrated that increasingly long switch intervals were required to resolve DNA fragments of increasing size. They suggested that the time required for the DNA to orient when the field was switched is size dependent. Larger molecules take longer to change direction and have less time to move during each switch interval so they migrate slower than smaller molecules. The choice of an appropriate switching interval for pulsed field gels must reflect the size range of the fragments to be resolved.

As with switch time, the choice of electric field strength used in PFGE must be varied with the size of the DNA to be separated. In general lower fields are required to

achieve separation of megabase sized DNA in well resolved bands (Smith *et al* 1987, Vollrath and Davis, 1987). At fields of more than $10V.cm^{-1}$, poor recovery and smeared bands were observed (Matthew *et al* 1988b). It was not clear whether this was due to failure of the large molecules to enter the gel pores at high field or to shear damage. Smith *et al* (1987) demonstrated that there is a relationship between the voltage at which a separation is carried out and the duration of the switch interval which must be used. By increasing the voltage gradient shorter switch times were needed to separate similar sized DNAs with similar resolution. Reducing the voltage to separate large DNAs required lengthening the switching intervals.

The effects of agarose concentration and the type of agarose effect the separation obtained with PFGE. Faster DNA migrations were observed in gels of lower agarose concentration (Mathew *et al* 1988a, Lai *et al* 1989). Mathew *et al* (1988a) also observed that DNA resolution during PFGE was slightly improved in agarose gels with small pores sizes. This was in contrast to conventional electrophoresis where the opposite was observed (Fangman, 1978). Buffers of lower ionic strength such as 0.5 X TAE have been reported to produce higher DNA velocities (Carle and Olson, 1984). Mathew *et al* (1988a) demonstrated that PFGE mobility was very sensitive to changes in temperature and that gel runs at high temperature exhibited diminished DNA resolution. When gel conditions such as temperature or agarose concentration are changed, a compensating change in the switch interval will restore the relative mobility characteristics of the original conditions because pulsed field gel parameters are interdependent (Lai *et al* 1989).

PFGE has played a major role in the expanded size capacity for cloning DNA. New cloning technologies have increased the size scale of cloning by at least 10-fold. The preparation and analysis of large DNA fragments required for generating jumping

libraries (Collins and Weissman, 1984) or for the cloning in yeast artificial chromosomes, YACs (Burke *et al* 1987) have relied on the use of pulsed field gels.

1.3.4   The Molecular Identification of FRAXA and FRAXE.

Until 1991, nothing was known about the molecular nature of any fragile site. In order to clone a fragile site, very close probes were needed as starting points for chromosome walking or jumping strategies.

Using a hybrid cell panel with cell lines specifically generated in conditions favouring the break at or near the Xq 27.3 fragile site (Warren *et al* 1990) probes very close to FRAXA were physically localized (Rosseau *et al* 1991).   Recombinant clones derived from microdissection of the FRAXA region were also shown to be close to the FRAXA site (Hirst *et al* 1991a).

Somatic cell hybrid DNA carrying the human fragile X chromosome (Poustka *et al* 1991) and genomic DNA from both normal and fragile X males (Vincent *et al* 1991, Bell *et al* 1991) were used to construct long range restriction maps around the FRAXA locus.   These maps determined the order and position of loci around FRAXA and physically identified the distance between probes flanking the fragile site.

Using probes that mapped very close to the locus of FRAXA, YACs containing human DNA from an X chromosome which expressed FRAXA (Kremer *et al* 1991a., Dietrich *et al* 1991) were isolated and by chromosome *in situ* hybridization were shown to span FRAXA. Kremer *et al* (1991a) and Heitz *et al* (1991) identified individual YACs that hybridized the probes which flanked FRAXA.

A single CpG island was identified in the cloned DNA of the FRAXA region that was methylated in a large number of fragile X affected individuals (Hirst *et al* 1991b., Dietrich *et al* 1991, Heitz *et al* 1991). The CpG rich region identified a logical position at which to commence the search for a gene that may be involved in the clinical

phenotype of the syndrome.

The isolation of DNA fragments in the immediate proximity of this CpG island detected highly localized DNA rearrangements that constituted the FRAXA mutation (Yu *et al* 1991., Oberle *et al* 1991). Kremer *et al* (1991b) localized the region of instability to a trinucleotide repeat $p(CCG)_n$. Normal X chromosomes were found to have between six and approximately 50 copies of the repeat. Carrier males (or transmitting males) had an increased copy number in the range of approximately 50 to approximately 200. Affected individuals had more than approximately 200 copies (rev. Richards and Sutherland, 1992).

The $p(CCG)_n$ repeat was found to be located adjacent to a CpG island at the 5' end of a gene designated FMR-1. (Verkerk *et al* 1991). The identification of this gene at the FRAXA locus suggested that it was involved in the fragile X syndrome. Pieretti *et al* (1991) reported an absence of expression of the FMR-1 gene in a majority of male fragile X individuals, suggesting a close involvement of this gene in the development of the syndrome. Methylation of the CpG island was found to repress FMR-1 transcription (Sutcliffe *et al* 1992., Hansen *et al* 1992).

Knight *et al* (1993) cloned FRAXE and demonstrated that individuals with this fragile site also possessed amplifications of a CGG repeat adjacent to a CpG island in Xq28. Normal individuals were found to have 6-25 copies of the repeat whereas FRAXE positive individuals contained greater than 200 copies. In affected individuals the CpG island was also found to be methylated (Knight *et al* 1993). Flynn *et al* (1993) observed that in some families expression of FRAXE was seen in mildly mentally retarded individuals suggesting that there may be another gene involved in X-linked mental retardation in Xq28. However, Sutherland and Baker (1992) first reported FRAXE expression in individuals who were not mentally retarded indicating that the mental

retardation may have been a chance association.

1.3.5   Project Strategy and Thesis Aims.

Several genes and anonymous DNA markers were known to be physically and/or genetically close to the rare distamycin A inducible fragile site FRA16B, located at 16q 22.100 (Callen *et al*, 1988). DNA markers had been localized and their locations refined within the region 16q13-16q22.1 by physical mapping using somatic cell hybrid panels which were derived from translocation breakpoints located in the same region. An unusually large number of breakpoints were clustered in close proximity to FRA16B. Markers had been further refined by *in situ* hybridization to metaphase chromosomes expressing FRA16B. The use of fixed structural alterations on chromosomes such as fragile sites allowed increased resolution of the *in situ* hybridization technique (Callen *et al*, 1988). The DNA markers mapping to the region 16q13-16q22.1 included the gene for cholesterol ester transfer protein, CETP (Drayna *et al* 1987) and the metallothionein gene cluster, MT (Simmers *et al* 1987). Lecithin-cholesterol acyltransferase was finely mapped at 16q22.1 (Callen *et al* 1988). Uvomorulin (UVO) was localized to 16q22.1 (Mansouri *et al* 1988) as was haptoglobin, HP (Callen *et al* 1988). Aldolase-A (ALDOA) had been regionally assigned to 16q22-16q24 (Mukai *et al* 1991). A number of anonymous DNA markers had also been assigned to and finely mapped within the region 16q13-16q22.1 (Sutherland *et al* 1987., Callen *et al* 1988., Hyland *et al* 1988., Hyland *et al* 1989a,b). In particular, an anonymous DNA marker D16S4 had been finely localized to an interval estimated to be as small as 1-2 Mb distal to FRA16B (Callen *et al* 1988). Genetic linkage analysis of polymorphic markers mapping to 16q21-16q22 established a tight linkage group surrounding FRA16B (Mulley *et al* 1989). The order of markers determined by this study were: centromere - MT, D16S10 - FRA16B - D16S4, HP with no recombination observed between D16S10 and

D16S4, the markers which flanked FRA16B.

With chromosome 16 constituting approximately 3% of the human genome, the size of this chromosome could be estimated to be 100 Mb. 16q13-16q22.1 accounted for approximately 10% of chromosome 16, therefore this region was roughly estimated to span approximately 10Mb of DNA. Using the available genes and anonymous DNA markers that had been finely mapped to 16q13-16q22.1 with respect to translocation breakpoints and/or FRA16B, the initial aims of this project were to:

1.　　　Construct a PFGE derived long range restriction map of this region by physically linking adjacent DNA probes to common high molecular weight genomic DNA fragments. The construction of a contiguous map of this region would physically order the probes along the chromosome and may physically define the probes flanking translocation breakpoints and more importantly the probes flanking FRA16B. Identification of these flanking markers and the physical distance separating them would pinpoint the region for subsequent high resolution restriction mapping and identification of FRA16B at the molecular level.

During the course of this study, significant new technologies associated with PFGE which improved long range restriction mapping had been developed. They included: the cloning of large DNA fragments in yeast artificial chromosomes, YACs (Burke et al 1987) and the cloning of large sections of DNA into a contiguous series of overlapping cosmid clones called cosmid contigs (Stallings et al 1992a). Subsequently these technologies were included into the project and as a result additional aims of the project were to:

2.　　　Initiate a chromosome walk with YACs to cover the region surrounding

　　　　FRA16B.

3.　　　Analyse a cosmid contig that mapped in close proximity to FRA16B.

CHAPTER 2

MATERIALS AND METHODS

2.1   INTRODUCTION

Most of the methods described in this Chapter were well established and used routinely in the Department of Cytogenetics and Molecular Genetics at the Adelaide Children's Hospital (Adelaide, Australia).   Only materials and methods used in more than one chapter will be presented here, otherwise, they will be described in their corresponding chapters.   The technologies and strategies involved in long range restriction mapping of DNA will be described in greater detail in this chapter as they provided the framework of this project.   These methods include Pulsed Field Gel Electrophoresis (PFGE), Yeast Artificial Chromosome (YAC) cloning systems and cosmid cloning systems.

All enzymes were obtained from commercial sources and were used in accordance with the manufacturers specifications.   All chemicals and solvents were of analytical grade.

2.2 MATERIALS

2.2.1 Enzymes
The following enzymes were obtained from:

Calf Intestinal Phosphatase	Boehringer Mannheim, Mannheim, Germany
E.Coli DNA Polymerase I (Klenow Fragment)	Amersham, Australia Pty Ltd
Lysozyme	Boehringer Mannheim
Not I Methylase	New England Biolabs Inc., MA, USA
Proteinase K	Sigma Chemical Co. St. Louis Missouri, USA
RNase A	Boehringer, Mannheim
Sequenase	USB (United States Biochemical Corp.)
T4 DNA Ligase	Boehringer Mannheim
T4 Polynucleotide Kinase	Pharmacia, Uppsala, Sweden
Taq Polymerase	Boehringer Mannheim

All restriction endonucleases were obtained from New England Biolabs (Beverly, Massachusetts, USA)

2.2.2 Electrophoresis.
The reagents were obtained from the following companies:

Acrylamide	Ajax, Auburn NSW, Australia

Agarose - Type N.A
      - Low Gelling Temperature
      - Pulsed Field Grade
	Pharmacia
	Seaplaque (FMC)
	Biorad

Ammonium Persulphate	Ajax
Bromophenol Blue	B.D.H. Chemicals LTD, Poole, Dorset England
Ethidium Bromide	Boehringer Mannheim

Molecular weight markers
- SppI   Bresatec
- DRIgest ™ III   Pharmacia
- Lambda-PFGE   Pharmacia
- Yeast-PFGE   Pharmacia

N,N,N1,N1-tetramethylethylene diamine   Biorad, Richmond
(TEMED)   California USA

Urea   Ajax

Xylene Cyanol   Tokyo Kasei, Tokyo, Japan

2.2.3 Radiochemicals.

alpha -^{32}P-dCTP, 3000 Ci/m.mole   Radiochemical Centre Amersham

gamma-^{32}P-ATP, 5000 Ci/m.mole   Radiochemical Centre, Amersham

2.2.4 Buffers and Solutions.

Buffers and solutions routinely used in this study were as follows:

Formamide Loading Buffer   92.5% (v/v) formamide
                 20mM EDTA
                 0.1% (w/v) xylene cyanol
                 0.1% (w/v) bromophenol blue

10x Loading Buffer   50% (v/v) glycerol
                 1% (w/v)SDS
                 100mM EDTA
                 0.1% xylene cyanol
                 0.1% (w/v) bromophenol blue

10x Ligation Buffer   0.5M Tris-HCl
                 (pH7.4)
                 0.1M $MgCl_2$
                 0.1M dithiothreitol
                 10mM spermidine
                 10mM ATP
                 1mg/mL bovine serum albumin

M9 salts   1.05% K_2HPO_4
                 0.45% (w/v) KH_2PO_4
                 0.1% (w/v) $(NH_4)_2SO_4$
                 0.05% (w/v) sodium citrate

Phosphate Buffered Saline (PBS)

130mM NaCl
10mM NaHPO$_4$
10mM NaH$_2$PO$_4$, pH7.2

2xPCR mix

33mM (NH$_4$)$_2$SO$_4$
133mM Tris-HCl
2% (v/v)ß-mercaptoethanol
13mM EDTA
0.34mg/ml BSA
20% (v/v) DMSO
3mM dATP, dGTP, dTTP, dCTP.

20xSSC

3 M NaCl
0.3 M tri-sodium citrate
2H$_2$0, pH7.0

TAE

40mM Tris-acetate
2mM EDTA, pH8.5

TBE

89mM Tris-base
89mM boric acid
2.5mM EDTA, pH 8.3

TE

10mM Tris-HCl (pH 7.5)
0.1mM EDTA

TES

25mM Tris-HCl (pH 8.0)
10mM EDTA
15% (w/v) sucrose

TSB

10% (w/v) polyethelyene glycol
(PEG) mw 3600
5% (v/v) DMSO
100mM MgCl$_z$
100mM MgS0)$_4$
L-Broth to appropriate volume

2.2.5 Bacterial Media.

2.2.5.1 Liquid Media.

All liquid media were prepared using millipore water and were sterilised by autoclaving. The compositions of the various media were as follows:

AHC medium

0.67% (w/v) yeast nitrogen base
w/o amino acids
1% (w/v) casein hydrolysate-acid
0.006% adenine
2% glucose

| L-Broth | 1% (w/v) Bacto-tryptone
0.5% (w/v) Bacto yeast extract
1% (w/v) NaCl, pH to 7.5 with NaOH. |

| Minimal Medium | 1mM $MgSO_4$
0.1mM $CaCl_2$
1mM thiamine-HCl
0.2% (w/v) glucose
1 x M9 salts |

| 2 x YT | 1.6% (w/v) Bactotryptone
1% (w/v) Bacto yeast extract
0.5% (w/v) NaCl |

2.2.5.2 Solid Media.

| L-Agar | L-broth
1% (w/v) Bacto agar |

| L-Amp | L-broth
1% (w/v) Bacto agar
ampicillin (100 μg/ml) |

| L-Kanamycin | L-broth
1% (w/v) Bacto agar
kanamycin (50μg/ml) |

2.2.6 Antibiotics.

Ampicillin	Sigma
Kanamycin	Boehringer Mannheim
Tetracycline	Sigma

2.2.7 Bacterial Strains.

The *E. coli* strains used in this project are listed below:

| XLI-Blue | genotype: rec A1, end A1, gyr A96, thi-1, hsd R17, sup E 44, rel A1, lac [F^{1}pro AB, lac 1^{q}ZAMI5, tn 10 (tet^r)]. |

MV1190

2.2.8 Vectors.

The vectors used in this study are the following:

Filamentous Phage:
M13mp18   Boehringer Mannheim
M13mp19   ref. Yanish-Perron *et al.*, (1985)

Phagemid Vectors:

pBLUESCRIPT SKII   Stratagene, La Jolla, California
                                                 USA

Plasmid Vectors:
pUC19   Bresatec, Yanish-Perron *et al.*,
                                                 (1985)

2.2.9   Miscellaneous materials.

Hybond N^{+TM} Nylon Membrane   Amersham

Sephadex G-50   Pharmacia P-LBiochemicals

Streptavidin Coated Magnetic Beads   Promega

T.A. Cloning Kit   Invitrogen

X-ray film   Kodak

2.2.10   Miscellaneous Fine Chemicals.

5-bromo-4-chloro-3-indolyl-ß-D
galactoside   Boehringer Mannheim

Chemicals for Oligonucleotide Synthesis   Applied Biosystems

Cesium chloride   Boehringer Mannheim

Deoxynucleotides   Boehringer Mannheim

Dideoxynucleotides   Boehringer Mannheim

Dideoxysequencing Kits   USB

Dimethylsulphoxide (DMSO)   Sigma Chemical Co.

Isopropyl thio-ß-D-galactoside (IPTG)   Boehringer Mannheim

Random Priming Oligolabelling Kits   Amersham

Phenol   Wako

Salmon-sperm DNA   Calbiochem

Sarkosyl   Ciba Geigy, Basle, Switzerland

SDS   Sigma Chemical Co

Spermidine   Sigma Chemical Co

Taq Dye Deoxy™ Terminator   Applied Biosystems
Cycle Sequencing Kit

2.3 METHODS

2.3.1 DNA Isolation.

2.3.1.1 Large Scale Isolation of Plasmid DNA and Cosmid DNA.
(modification of Maniatis *et al.*, 1982)

10ml Luria Beroni (LB) medium containing ampicillin ($50\mu g/ml$) was inoculated with

a single fresh bacterial colony. The culture was incubated at 37°C for 5-7 hours with

vigorous shaking, and then transferred to 100ml LB containing ampicillin ($50\mu g/ml$).

After overnight incubation at 37°C with vigorous shaking, the culture was transferred

to two 50ml falconer tubes. The tubes were left on ice for 15 minutes and then spun

at 3000 rpm for 15 minutes in a Jouan CR3000 centrifuge at 4°C. The supernatant

was discarded and the cell pellet was gently resuspended in $300\mu l$ TE and glucose

containing $60\mu l$ of 80mg/ml lysozyme. The cell suspension was left at room

temperature for four minutes and on ice for one minute. 1.2mls of 0.2M NaOH/1%

SDS was added to the cell suspension, gently mixed and incubated on ice for a

further five minutes. $900\mu l$ of ice cold 3M potassium acetate (pH4.3) was then

added, and mixed by inversion. After resting on ice for 10 minutes, the lysed cell

debris was cleared by centrifugation in a Beckman J2-21M/E centrifuge with a JA20

rotor at 15,000rpm for 15 minutes. The supernatant was mixed with 5.5 mls of

ethanol and after five minutes at room temperature, precipitated nucleic acids were

pelleted by centrifugation at 15,000rpm for 15 minutes. The DNA pellet was washed

twice in 2ml of 70% ethanol, air-dried and resuspended in $200\mu l$ TE. To eliminate

RNA in the DNA preparation, $10\mu l$ of 1mg/ml RNase was added to the DNA

solution and incubated at 37°C for one hour. To eliminate proteins in the DNA

preparation, $100\mu l$ of 3x proteinase K buffer, $10\mu l$ of 10% SDS and $2\mu l$ of 10mg/ml

proteinase K were added to the DNA solution and incubated for one hour at 37°C.

Following incubation, the DNA was phenol extracted, ethanol precipitated and dissolved in 200μl of TE. If required the DNA was purified by banding on a CsCl gradient.

2.3.1.2 Small Scale Isolation of Plasmid DNA.
(modification of Birnboim and Doly, 1979)

A single bacterial colony was inoculated to 1.5ml of LB medium containing ampicillin (50μg/ml) in a 10ml tube. The culture was incubated at 37°C overnight with vigorous shaking. The culture was transferred to an eppendorf tube and spun in an eppendorf centrifuge at 14,000rpm for two minutes. After discarding the supernatant, the cell pellet was well resuspended in 100μl of cold fresh TES medium and 0.25ml of 100mg/ml lysozyme. The cell suspension was left at room temperature for five minutes before 200μl of 0.2N NaOH/1% sodium dodecyl sulphate (SDS) was added and mixed well. The mixture was incubated on ice for five minutes and then 150μl of cold 3M sodium acetate (pH4.6) was added. After five minutes on ice the mixture was spun in an eppendorf centrifuge for 10 minutes. The supernatant was carefully drawn off and the nucleic acids precipitated with two volumes of ethanol. The DNA pellet was finally washed twice with 70% ethanol, air-dried and resuspended in 50μl TE.

2.3.1.3 Isolation of Peripheral Lymphocyte DNA.
(modification of Wyman and White, 1980)

Blood samples were collected in 10ml tubes containing EDTA and were allowed to cool to room temperature before being stored at -70°C. For DNA lymphocyte isolation, the frozen blood sample was thawed and transferred to a 50ml falconer tube. Cell lysis buffer was added to the tube until the 30ml mark was reached. After mixing, the tube was left on ice for 30 minutes. The cell suspension was spun in the Jouan centrifuge at 3500rpm for 15 minutes at 4°C. The supernatant was

aspirated down to 5 mls, then cell lysis buffer was added again to the 30ml mark.

Centrifugation was repeated. The supernatant was carefully aspirated and 3.25ml

Proteinase K buffer, 0.5ml of 10% SDS and 0.2ml of Proteinase K (10mg/ml) were

added and well mixed with the cell pellet. The tube containing the cell suspension

was sealed with parafilm, secured on a rotating wheel (10rpm) and incubated

overnight at 37°C. DNA extraction was performed twice with phenol and twice with

phenol/chloroform. Following ethanol precipitation the DNA pellet was dissolved in

0.1ml of TE.

2.3.1.4 Preparation of Single Stranded M13 DNA.

M13 were picked by touching the surface of the plaque with a sterile toothpick and

inoculating 1.5ml 2 x TY broth in a 10ml tube. This was shaken at 37°C for 5-8

hours. The cells were transferred to an eppendorf tube and centrifuged for five

minutes at 14,000rpm. The supernatant was carefully removed using a Pasteur

pipette until the bacterial pellet started to move from the sides of the tube.

Centrifugation was repeated and the remaining supernatant carefully removed. To

the supernatant 200μl of 2.5M NaCl, 20% polyethylene glycol (PEG) 6000 was

added. After mixing, the tube was incubated at 4°C overnight. The precipitated

phage was pelleted by a 15 minute centrifugation in an eppendorf centrifuge. The

supernatant was removed and discarded and any residual PEG was carefully wiped

away from the inner walls of the centrifuge tube using a tissue. The phage pellet was

dissolved in 180μl H$_2$O and 20μl of 10 x lysis buffer was added. The suspension was

incubated at 80°C for ten minutes then allowed to cool to room temperature. Lysed

phage were centrifuged at room temperature for fifteen minutes at 12,000rpm,

washed once in 70% ethanol, air-dried and redissolved in 30μl H$_2$0. Phage DNA

was stored at -20°C.

2.3.1.5   Purification of DNA.

2.3.1.5.1   Phenol, Phenol/Chloroform Extraction of DNA.

Solutions of DNA were extracted with phenol/chloroform to remove proteins and other contaminants. Equal volume of phenol (TE saturated, 10mM Tris HCl pH 8.0, 1mM EDTA) were added to the DNA solution and vortexed vigorously for one minute. After vortexing, the mixture was centrifuged for five minutes at full speed in an eppendorf centrifuge. The upper aqueous phase which contained the DNA was removed leaving a white interface of denatured protein and the lower organic phase. When small quantities of DNA were being handled, the organic-phase was re-extracted with TE and the aqueous phases pooled. For better phase separation and optimal purification of DNA by this method, phenol extraction was sometimes followed by a phenol/chloroform extraction. Here, 0.5 volume of phenol and 0.5 volume of chloroform were added to the DNA solution, vigorously mixed and centrifuged for five minutes. The aqueous phase was removed and added to an equal volume of chloroform : isoamyl alcohol (24:1). The vortex, centrifugation and aqueous phase removal steps were repeated once again. Following either extraction procedure, the DNA was then ethanol precipitated.

2.3.1.5.2   Ethanol Precipitation of DNA.

The DNA sample was made 300mM with respect to sodium acetate using a 3M stock solution at pH5.2. 2.5 volumes of cold ethanol were added and the tube mixed well. The mixture was incubated at -20°C for one hour or longer. Precipitated DNA was pelleted at 14,000rpm for ten minutes and washed once in 70% ethanol. After drying in air or under vacuum, the DNA was redissolved in H_2O or TE.

2.3.1.5.3 Purification of Plasmid/Cosmid DNA by Cesium Chloride (CsCl) Centrifugation.

The DNA to be purified was made up to 720μl with TE and in this solution nucleic acid extracts were adjusted to a final density of 1.8gm/ml by the addition of 1.26gms of CsCl and 120μl of ethidium bromide (10mg/ml). 1.6mls of 65% w/v CsCl solution (density = 1.470g/ml) was made up and placed into a 2.2ml Quick Seal (Beckman) tube. Approximately 600μl of the dense nucleic acid-containing solution was layered beneath the less dense CsCl solution ensuring that the solution interface was not disturbed. The tube was then filled to capacity with the less dense CsCl solution. The tube was sealed and placed in the TLA 100.2 rotor in the Beckman TL-100 Tabletop ultracentrifuge and run at 100,000rpm for 2.5 hours. Slow acceleration and slow deceleration were set at "5" on the TL-100. During centrifugation two bands of DNA were resolved on the gradient and were subsequently visible under long wavelength (304nm) ultra-violet light. The upper band comprised of nicked circular DNA whilst the lower contained the supercoiled plasmid or cosmid DNA. The lower band was collected directly through the tube wall by means of a needle and syringe. The DNA solution was extracted 3-5 times with N-butanol (equilibrated against TES) to remove ethidium bromide and then was dialyzed against three changes of TES to remove CsC1. After phenol/chloroform extraction and ethanol precipitation, the DNA was redissolved in TE or H$_2$0.

2.3.1.6 Preparation of Agarose Beads Containing Mammalian DNA/Cell Line DNA (-refer to 2.6.1.1).

2.3.1.7 Preparation of Agarose Beads Containing Yeast Cells (- refer to 2.6.1.2).

2.3.1.8 Recovery of DNA from Agarose Gels.

2.3.1.8.1 Prep-a-Gene

The following protocol was obtained from the Biorad Prep-a-gene handbook, the reagents used in this protocol were provided in the form of a Prep-a-gene kit.

Agarose containing the DNA band was excised from the ethidium bromide (EtBr) stained agarose gel. The wet weight of the agarose gel slice was determined to estimate its volume (1 mg = 1μl). Three volumes (to the volume of the gel slice) of binding buffer was added to the gel slice in an eppendorf centrifuge tube and the tube was placed in a 50°C water bath until the agarose had completely melted. Frequent inversion of the tube was required to enable the agarose to melt. 5μl of well mixed silica matrix suspension was added to the solution (5μl for up to 5μg DNA in the gel, an additional 1μl of matrix was added to every additional 1μg of DNA contained in the gel slice) and the solution was left to stand for five minutes at room temperature to allow the DNA to bind to the silica matrix. The bound DNA was pelleted by spinning in an eppendorf centrifuge at 14,000rpm for twenty seconds. The bound DNA was washed 2x with 10-50 volumes (of the silica matrix volume) of binding buffer and 3x with 10-50 volumes of washing buffer. The DNA was finally eluted from the silica matrix with TE at 50°C for five minutes.

2.4 SUBCLONING OF HUMAN DNA SEQUENCES
This protocol is a modification of Maniatis *et al.*, (1982)

2.4.1 Preparation of Plasmid Vector DNA and Human DNA Inserts.

500ng of vector DNA (M13mp18/mp19, pUC19, Bluescript phagemid vector SK II) was digested with an enzyme(s) which cleave the polylinker, in a total volume of 20μl at the required temperature for one hour. Digestion was tested by running 1μl of digested and undigested vector DNA samples side by side on a minigel which was stained with EtBr and visualised under UV light. Human DNA (cloned in YAC or cosmid, or a PCR product) was digested with the same restriction enzyme that

cleaved the vector. The digested DNA sample was checked on a minigel for complete digestion then was extracted once with an equal volume of phenol/chloroform followed by ethanol precipitation. Final DNA concentration of insert DNA was adjusted to 20ng/μl with TE.

2.4.2 Dephosphorylation of Vector DNA.

To prevent self-ligation of vector digested with a single restriction enzyme, the 5' terminal phosphate group was removed with alkaline phosphatase. The vector DNA was digested to completion with an appropriate restriction enzyme in a total volume of 20μl. 2μl($^1/_{10}$ volume) of 10mM EDTA (pH8.0) and 1μl of 2.8u/μl calf intestinal alkaline phosphatase (CIAP) were added to the digests. The reaction was carried out at 37°C for 30 minutes, then an additional 1μl of CIAP (2.8u/μl) was added and incubation was continued at 37°C for a further 30 minutes. After incubation, 5μl of 5% SDS was added and the mixture was heated to 65°C for ten minutes. The solution was phenol/chloroform extracted and the DNA precipitated with 2.5 volumes of ethanol. The DNA pellet was rinsed with two changes of 70% ethanol at room temperature to remove all traces of SDS, desiccated and dissolved in 50μl of TE.

To test the efficiency of dephosphorylation, 1μl of dephosphorylated vector DNA was ligated and transformed into *E. coli* strain MV1190. If the 5' terminal phosphate group was removed, the vector could not recircularize, therefore only a few colonies/plaques were seen on the plate because linear form DNA is very inefficient in transformation. The vector was then ready for use.

2.4.3 Ligation Reactions.

Ligation reactions were carried out with a vector:insert molar ratio of approximately 1:3 to maximise intermolecular ligation rather than intramolecular ligation. Usually, for 100ng of linearized and phosphatased vector, 2μl of 10x ligation buffer, 1-2 units

of T4 DNA ligase and insert DNA (200ng) were added and the reaction mixture (in a total volume of 20μl) was incubated at 12-16°C overnight.

The efficiency of the ligation reaction was normally checked by religation of the *Hind* III digested lambda DNA under the same conditions as the sample DNA. The religated and unreligated lambda DNA samples were separated on an agarose minigel. The disappearance of low molecular weight bands and increasing intensity of the large molecular weight bands indicated efficiency of the ligation reaction.

2.4.4 Competent Cells and Transformation.
(modification of Chung *et al.*, 1989)

E. coli strain MV1190 cells for M13 were made competent with a method modified from Chung *et al.*, (1989). Stationary phase MV1190 cells from an overnight culture were diluted 1:100 (v/v) into 20ml LB (TY for M13 transformations). The cells were grown at 37°C with constant shaking for 2-2.5 hours. The cells were pelleted by centrifugation in the Jouan centrifuge at 3,000rpm for ten minutes and then the cell pellet was resuspended in 2mls ($^1/_{10}$ of the original volume) of ice cold fresh TSB. The competent cells were ready for use after leaving on ice for ten minutes.

M13 transformations were plated out at this stage. An aliquot of the transformation was added to 100μl of log phase cells, 30μl of 0.1M isopropylthio-β-D-galactoside (IPTG) 30μl of 2% 5-bromo-4-chloro-3-indolyl-β-D-galactoside (x-gal) and 3 mls of ➡

molten H-top agar (45°C). This was mixed and poured onto an 86mm 2 x TY plate. After the top-agar had set, the plate was incubated overnight at 37°C. Recombinant M13 phage were detected as white plaques amongst blue plaques comprising wild type phage. If transformed cells containing plasmids/cosmids were to be selected by appropriate antibiotic resistance, the transformation reaction was diluted with 1ml of LB and then incubated at 37°C for thirty minutes. The cells were then pelleted by gentle centrifugation and resuspended in 100μl of LB. The resuspended cells, 30μl of 2% X-gal and 30μl 0.1M IPTG were spread onto appropriate antibiotic plates and incubated overnight at 37°C. Recombinant plasmids/cosmids were detected as white colonies.

2.5 ENZYME DIGESTION, GEL ELECTROPHORESIS AND SOUTHERN BLOT ANALYSIS

2.5.1.1 Restriction Endonuclease Digestion of DNA.

Restriction endonuclease digestion of DNA was carried out using the buffer systems provided by New England Biolabs (see product handbook of Biolabs). Generally, four units of enzyme was added for each microgram of DNA to be digested and the reaction mix was incubated for at least 12 hours for genomic DNAs (plasmid, cosmid and phage DNA digests were incubated for 2-4 hours) to ensure complete digestion. To ensure that the enzymic activity was not affected by glycerol, the volume of restriction enzyme(s) did not exceed $^1/_{10}$ of the final volume of reaction mix, especially when two or more different enzymes were used simultaneously. Reactions were terminated by the addition of 0.1 volume of 10 x agarose gel loading buffer.

2.5.1.2 Restriction Endonuclease Digestion of Agarose Beads (See 2.6.2).

2.5.2 Agarose Gel Electrophoresis of DNA.

Electrophoresis of DNA to be used for Southern blot analysis was carried out using

agarose (0.8% - 1.2%) dissolved in 0.5 or 1 x TBE and cast on 14cm x 11cm x 0.3cm

perspex horizontal casts. Electrophoresis was performed in BRL horizontal tanks

containing 0.5-1 x TBE buffer at 15-100mA, until the bromophenol blue had migrated

an appropriate distance to ensure that adequate separation of the DNA fragments had

taken place.

Analytical agarose minigels (for checking digestions etc) were electrophoresed for one

hour at 100 volts in a Biorad Mini-sub™ DNA cell. DNA was visualised under UV light

after staining the gel in 0.02% ethidium bromide solution for 10-30 minutes.

2.5.3 Molecular Weight Markers.

Eco RI digested SppI phage or DRIgest were used as molecular weight markers in
Southern blot analysis.

2.5.4 ^{32}P Radio-Isotope Labelling of DNA.

2.5.4.1 5' End-labelling of Oligonucleotides.

Synthetic oligonucleotides were 5' end labelled using T4 DNA polynucleotide kinase and

gamma-^{32}P-dATP as described by Chaconas and Van de Sande (1980) with the addition

of spermidine to a final concentration of 0.1mM.

2.5.4.2 Primer Extension.

Labelling of double stranded DNA was performed by primer extension of random

oligonucleotides (Feinberg and Vogelstein, 1983) using the Amersham Multiprime DNA

labelling systems kit. In brief, a small quantity of DNA insert (25-50ng) was denatured

at 100°C for two minutes and added to a solution containing random hexamers, dATP,

dGTP, dTTP, alpha ^{32}P-dCTP, Klenow enzyme and buffer. The mixture was incubated

at 37°C for 1-1.5 hours or at room temperature overnight.

2.5.4.3 Probe Purification.

Unincorporated radionucleotides were removed from labelled probe by running the

sample through a Sephadex G-50 column. Two drop fractions were collected, the first peak detected by the mini-monitor ß counter being saved since it contained the incorporated labelled probe.

2.5.4.4 Prereassociation of Repetitive DNA.
(Sealy *et al.*, 1985)

^{32}P-labelled DNA thought to contain repetitive DNA sequences was pre-reassociated prior to DNA hybridization. The labelled DNA was mixed with a 2000 fold excess of sonicated human placental DNA (Sigma) and made 5 x SSC. The sample was denatured in 100°C water bath for 10 minutes, cooled on ice for one minute and then incubated at 65°C for 1-2 hours. The mixture was then added to prewarmed hybridization mix, and applied to the Southern blot filters.

2.5.5 Transfer of DNA to Nylon Membranes.

2.5.5.1 Plaque/Colony Lifting.
(Grunstein and Hogness 1975; Benton and Davis, 1977)

Bacteria harbouring a recombinant vector were plated out as described. After growth overnight at 37°C a replica was made by gently laying a nylon membrane disc onto the plate surface. Plates containing top-agar were first cooled to 4°C. The filter was keyed to the plate and then transferred to 0.5M NaCl/0.5M NaOH for 10 minutes to lyse host cells and denature DNA. After neutralisation for 10 minutes in 2M NaCl/0.5M Tris HCl pH 8.0, the filter was rinsed in 2xSSC and baked in the microwave at high heat for 30 seconds.

2.5.5.2 Southern blotting.
(Reed and Mann, 1985)

Restriction endonuclease digested DNA was separated on agarose gels and transferred to Hybond N^{+}™ (Amersham) nylon membrane using the alkaline transfer method (Reed and Mann, 1985). If the DNA to be transferred was over 1kb in size, an acid nicking

step was included by soaking the gel twice in 0.25M HCl for 15 minutes with gentle shaking. The gel was then immersed in 0.4M NaOH twice for 15 minutes with gentle shaking. The nylon membrane was cut to the size of the gel and placed in the transfer solution (0.6M NaCl, 0.4M NaOH) briefly before transfer. DNA in the agarose gel was transferred to the prepared filter (using the transfer solution) by capillary action for 1-16 hours. The DNA was then washed in 2 x SSC and allowed to dry at room temperature.

2.5.6 Prehybridization, Hybridization and Washing.

2.5.6.1 With Oligonucleotide Probes.

Prehybridization was carried out in a solution consisting of 20% (v/v) 5 x P (1% (w/v) bovine serum albumin, 1% (w/v) polyvinylpyrrolidone M_r 40,000, 1% (w/v) ficoll M_r 40,000, 250mM Tris HCl pH7.4, 0.5% (w/v) pyrophosphate), 1M NaCl, 1% SDS, 10% (w/v) dextran sulphate and 100μg/ml denatured salmon sperm DNA for at least one hour with shaking. Hybridizations were performed in the same solution and incubated overnight at 42°C with between 1-10ng/ml of 5' end-labelled probe. Filters were washed under conditions determined by the melting temperature of the primer and target sequence. Usually, two washes in 6 x SSC, 0.1% SDS for 5-10 minutes would suffice. If necessary the wash solution could be gradually increased to 65°C. Autoradiography was carried out at room temperature or, for detection of low levels of radioactivity, at -80°C in the presence of tungsten intensifying screens.

2.5.6.2 With Oligolabelled Probes.

Prior to hybridization, nylon filters were prehybridized at 42°C for 1 hour in a solution consisting of 50% (v/v) deionized formamide, 5 x SSPE, 2% SDS, 5 x denharts and 100μg/ml salmon sperm DNA. Filters were then further prehybridized in the same solution with the addition of 10% (w/v) dextran sulphate for a further one hour.

Hybridizations were performed in the same solution and incubated overnight at 42°C with between 1-10ng/ml of oligolabelled probes.

The filters were washed twice for 10 minutes in Wash A (2 x SSPE, 1% SDS) at 42°C. They were then washed twice for 10 minutes in wash A at 65°C. Finally the filters were washed twice for 15 minutes in wash B (0.1 x SSPE, 0.1% SDS) at 65°. Autoradiography was carried out at -80°C in the presence of tungsten intensifying screens.

2.6 PULSED FIELD GEL ELECTROPHORESIS

For the construction of a long range physical map encompassing 16q13-16q22.1 pulsed field gel electrophoresis (PFGE) provided the frame work for the separation and analysis of high molecular weight human DNA fragments and cloned human DNA (yeast artificial chromosomes, cosmids). In this project the Pulsaphor Plus 2015 (LKB) with Chef insert and Chef Mapper™ (Biorad) electrophoretic systems were used.

In brief, the gel was prepared by casting 150ml of 1% agarose in 0.5 x TBE directly into the gel support tray (15cm x 15cm). For each sample, 50-100μl of agarose beads containing high molecular weight DNA was loaded into the well, using cut-off tips. PFGE was performed in 0.5 x TBE at 14°C at the selected switching interval for the required time. The operation and application of the PFGE systems were according to the manufacturer's instructions.

2.6.1 Encapsulation of Cells in Agarose Beads for Use with PFGE.
(Overhauser and Radic, 1989)

Due to the large sizes of DNA molecules that can be separated by PFGE, careful preparation of the DNA samples was a critical procedure. To minimise shearing and to preserve the integrity of the large DNA molecules it was necessary to isolate the DNA

from intact cells after they had been embedded in agarose. The agarose matrix stabilises and immobilises the DNA molecules after the removal of cell membranes and proteins.

2.6.1.1 Preparation of Agarose Beads Containing Genomic/Cell Line DNA.

Hybrid cell-line cells were scraped from tissue culture plates (peripheral lymphocytes were collected from blood samples) and gently centrifuged. The cell pellet was adjusted to 2×10^7 cells with cold PBS. The cells were washed twice in PBS and finally resuspended in a total volume of 5mls. The cell suspension, 1% low melting point agarose in 1 x PBS and a bottle containing parafin oil were all equilibrated to 45°C. A beaker containing 100ml ice-cold PBS and a magnetic stir bar was placed in an ice bucket on a stir plate set at medium speed. 5ml of 1% low melting point agarose was added to the warmed cell suspension and mixed. 20ml of prewarmed parafin oil was then added to the suspension of cells in agarose and the mixture was swirled vigorously for 30 seconds to form a uniform emulsion. The emulsion was quickly poured into the cold PBS and stirred for 2-3 minutes. The mixture was transferred to several 50ml falconer tubes and centrifuged in the Jouan centrifuge at 3500rpm for 10 minutes. The parafin oil layer at the top of the tube was removed leaving a layer of beads just below and also at the bottom of the tube. The beads were dispersed by repeated pipetting with a large bore pipette and the centrifugation repeated. After centrifugation the excess PBS and any unpelleted beads were removed and the remaining pellets were combined in a single tube. Centrifugation was repeated, the supernatant removed and the inside of the tube was carefully wiped with a tissue to remove all excess parafin oil. 20ml of 1% SDS, 25mM Na$_2$EDTA pH8.0 was added to the beads and pipetted repeatedly to break up any clumps. The beads were centrifuged and suspended in 20mls 1% (w/v) sarcosyl, 25mM Na$_2$EDTA pH8.0, 50μg/ml proteinase K, well mixed and incubated overnight at

50°C. After proteinase K digestion the beads were pelleted, the supernatant discarded and 20ml TE containing 0.1M phenylmethylsulphonyl fluoride (PMSF) was added. The beads were mixed, centrifuged and washed twice in TE before storing in TE. Up to 10ml of beads was recovered.

2.6.1.2 Preparation of Agarose Beads Containing Yeast DNA.

200mls of AHC medium (1 litre of AHC medium consists of 6.7mg of yeast nitrogen base, 10mg of casein hydrolysate, 20mg of adenine and 20mg of glucose) in a one litre flask was inoculated with a single yeast clone and grown at 30°C with constant shaking until stationary phase was reached (2-3 days). The yeast cells were pelleted by centrifugation at 3500rpm in the Jouan centrifuge. The supernatant was removed and the yeast cells were resuspended in 10ml SE (75mM NaCl, 25mM Na_2EDTA pH8.0). After two washes in SE the cells were suspended in 4ml SE. Beads were made using the same procedure as already described except for the use of SE in place of PBS. For digestion of the cells, 0.5 ml 2-mercaptoethanol, 5mg zymolase (Sigma) was added and the final volume was adjusted to 10ml with SE. The beads were mixed and incubated at 37°C for two hours. The beads were then pelleted, resuspended in 20ml 1% (w/v) sarkosyl, 25mM Na_2EDTA pH8.0, 50μg/ml proteinase K, and incubated overnight at 50°C to lyse the prepared spheroblasts. The beads were rinsed using the same procedure described for beads containing cell line/genomic DNA.

2.6.2 Restriction Digestion of Agarose Beads.

For restriction endonuclease digestion of beads, approximately 200μl of bead suspension was placed in an eppendorf centrifuge tube. The tube was filled with 1 x restriction enzyme buffer, the beads thoroughly resuspended and allowed to equilibrate for five minutes before pelleting in the eppendorf centrifuge for two minutes. The buffer was

removed and the beads were washed twice more with 1 x buffer. Restriction endonuclease was added at a concentration of approximately 2-5 units/μg DNA. Incubation was performed at the appropriate temperature for at least four hours (usually overnight). The beads were centrifuged and the supernatant removed before addition of $^1/_{10}$ volume gel loading buffer. Pelleted beads were reduced to approximately 50-100μg volume by the centrifugation and represented up to 3μg DNA/50-100μl (for genomic DNA).

2.6.3 Loading Agarose Beads into the Wells.

Agarose beads were dry loaded prior to the gel being submerged into the buffer. The DNA sample was loaded into the well using a cut-off tip, the top of each well was sealed with 2-3 drops of 1% low melting point agarose before being lowered into the buffer.

2.6.4 Switching Intervals.

The most critical parameters of resolving large DNA fragments by PFGE is determining the molecular size range of DNA separation required and the switching interval that will achieve the desired result. When using the Pulsaphor 2015 with Chef insert, a variety of switching intervals were tested to separate standard PFGE molecular weight markers (see section 2.6.5). The switch time which gave good resolution at the size range of analysis was then used to separate the DNA digests.

The Chef Mapper™ (Biorad) contains an integrated software system whereby the size parameters of DNA separation are logged in and a computer algorithm automatically selects the appropriate switching ratios to maximise resolution at that scale.

The type of equipment used, the size range of DNA separation and switching intervals determined for analysis of DNA in this project is described in the relevant chapters.

2.6.5   DNA Size Markers.

A combination of commercially prepared lambda DNA and whole yeast chromosomes were used as molecular weight markers. For lower weight DNA separations Lambda *Hind* III (Pharmacia) provided DNA markers spanning from 2kb - 23 kb. For higher molecular weight separations, lambda DNA-PFGE (Pharmacia), a preparation of bacteriophage concatemers provided markers over a range of sizes from 50kb-1000kb in 50kb intervals. In addition, yeast DNA-PFGE (Pharmacia) comprising of whole *Saccharomyces cerevisiae* chromosomes were also used for high level separations as the chromosomes range in size from 200kb to 2000kb.

2.7   POLYMERASE CHAIN REACTION (PCR)

All PCR's were performed in a Perkin Elmer-Cetus thermal cycler according to the manufacturers instructions. Incubations were performed in a $20\mu l$ final volume comprising of $10\mu l$ 2 x PCR mix, $6\mu l$ $MgCl_2$ (concentration optimised for each pair of primers), 150ng of each primer template DNA, 1 unit of Taq DNA polymerase and sterile water to $20\mu l$. The solution was mixed well and overlayed with one drop of parafin oil. PCR's were performed with denaturation, annealing and elongation performed at the designated temperature and time depending on the primer pairs being used and the length of PCR product expected. An appropriate number of cycles were performed and positive and negative controls were always included in each set of reactions. $100\mu l$ of chloroform was added to the tube after the PCR run was completed. The mixture was vortexed, briefly centrifuged and the upper aqueous phase removed for further manipulation or analysis.

2.7.1 PCR Primers.

Oligonucleotides for PCR were designed to contain a similar proportion of purine and pyrimidine bases, no long runs of any one base and no repeat sequence DNA. In addition, primer pairs were carefully checked at their 3' ends to avoid the possibility of primer-dimer formation. The standard length of oligonucleotides used in this study was 25-nt. All oligonucleotides were synthesized using an Applied Biosystems 391 DNA synthesizer in the Department of Chemical Pathology at the Adelaide Children's Hospital (Kathy Holman).

2.7.2 Oligonucleotide Deprotection and Cleavage.

A manual deprotection apparatus which included the synthesis column containing the sythetised oligonucleotide was assembled. 3 mls of ammonium hydroxide was pipetted into a 5 ml tube. The ammonium hydroxide was drawn into the column/syringe, minimising the voluming in the syringe. The needle was inserted into the rubber stopper and the apparatus was allowed to stand for 30 minutes at room temperature. The needle was removed and the ammonia solution was collected in an eppendorf centrifuge tube. The ammonia treatment was repeated two or three times. The ammonia/oligonucleotide solution was diluted to 3 mls by combining all the collected ammonia solution/oligonucleotide and the unused ammonium hydroxide in the 5 ml tube. The solutions were mixed and pipetted into two screw topped eppendorf centrifuge tubes and incubated at 55°C overnight.

2.7.2.1 Oligonucleotide Purification-n-butanol method (from Sawadogo and Van Dyke, 1991).

The cleaved and deprotected oligonucleotide in ammonium hydroxide solution was cooled to room temperature. 700 μl of oligonucleotide was then added to 7 mls of n-butanol in a centrifuge tube (three or four tubes per oligonucleotide were required).

The solution was vortexed for 15 seconds, then centrifuged at 10,000 rpm for five minutes. The supernatant was discarded and the pellet was resuspended in 700 μl of water. An additional 7 mls of n-butanol was added, then vortexed, centrifuged and supernatant discarded as before. The pellet was dried under vacuum and resuspended in 100 μl of water. The DNA was quantitated and the concentration of oligonucleotide was adjusted to 1 mg/ml.

2.7.3   Subcloning PCR Products.

PCR products were purified by the "Prep-a-gene" method, as already described (2.3.1.8.1) and subcloned into a 3 kb plasmid vector pCR™ 1000 (Invitrogen) by the methods described in the TA Cloning™ Instruction manual (Invitrogen). The TA Cloning™ system takes advantage of the thermostable polymerases used in PCR that add single deoxyadenosines to the 3'-end of all duplex molecules. These A-overhangs are used to insert the PCR product into the specifically designed vector pCR™ 1000 which provides single 3'-T-overhangs at the insertion site.

2.8   DNA SEQUENCING

2.8.1   Sequence Gels.

Sequence gels were formed between 52cm x 21cm glass plates separated by 0.25-0.4mm wedge spacers. Gel mixes were comprised of 5% polyacrylamide (20 acrylamide:1 bisacrylamide) in 1 x TBE with 7M urea. The mixture was filtered through Millipore filter paper discs and degassed in a vacuum before the addition of 0.1% fresh ammonium persulphate and 0.05% TEMED (N, N, N^1, N^{1}tetramethylethylenediamine). The gel was poured and wells were formed between the teeth of a 0.25mm sharks tooth comb inserted into the top surface of the gel. The gel was pre-electrophoresed in 1 x TBE for 30 minutes prior to loading so as to warm the apparatus. 3μl of denatured sample (see

2.8.2) was loaded per slot and the gel run at 2000 volts for the appropriate time. Up to 500 bp of sequence could be read from wedge gels.

After electrophoresis the gel was transferred to a piece of pre-cut filter paper, covered with plastic film and dried at 80°C for 3-5 hours under vacuum on a 583 Gel Drier (Biorad). The dried gel was exposed to X-ray film overnight.

2.8.2   DNA Sequencing by Chain termination.
(Sanger *et al.*, 1977)

DNA sequencing by the chain termination method is possible for both single stranded (M13) DNA and double stranded (plasmid) DNA. For single stranded M13 sequencing, the DNA to be sequenced was subcloned into bacteriophage M13 and template DNA was prepared as described in section 2.3.1.4. When sequencing plasmid DNA subclones it was first necessary to denature the double stranded template.

2.8.2.1   Denaturation of Double Stranded DNA Template.
(Current Protocols in Molecular Biology)

To an eppendorf centrifuge tube $10\mu l$ of template (1-2μg of plasmid DNA containing insert) was denatured by the addition of $2\mu l$ 2M NaOH/2mM EDTA. The solution was briefly vortexed and incubated at room temperature for 10 minutes. The solution was neutralised by the addition of $3\mu l$ sodium acetate (pH4.5) and $7\mu l$ H_2O. The DNA was precipitated by the addition of $60\mu l$ ethanol and incubated on ice for 30 minutes. The pellet was washed with 70% ethanol, vacuum dried and resuspended in $10\mu l$ of H_2O.

2.8.2.2   Annealing.

$10\mu l$ of template DNA (M13 template or denatured plasmid DNA) was added to $2\mu l$ of 5 x reaction buffer (200mM Tris-HCl, pH7.5, 100mM $MgCl_2$, 250mM NaCl) and $1\mu l$ of primer (M13 used -21 sequencing primer - Applied Biosystems, for pUC sequencing PUC-F and PUC-R primers were used). The mixture was heated to 70°C for two

minutes and then allowed to cool to $<30°C$ by placing the tube into $70°C$ water and allowing the water to gradually cool.

2.8.2.3 Labelling and Chain Termination.

The primer annealed DNA template ($13\mu l$ total volume) was added to $2\mu l$ of labelling mix ($1.5\mu M$ dGTP, $1.5\mu M$ dATP, $1.5\mu M$ dTTP), $0.5\mu l$ alpha ^{32}P-dCTP and $2\mu l$ Sequenase (USB). The solution was mixed and polymerisation allowed to occur at $45°C$ for five minutes in a water bath. Termination was initiated by the addition of $4\mu l$ of the polymerisation reaction to four separate tubes, each containing $4\mu l$ of either C, G, T, or A termination mix (each mix contained either $80\mu M$ dGTP, dCTP, dTTP or dATP and 50mM NaCl. In addition each appropriate mix contained either $8\mu M$ of ddGTP, ddCTP, ddTTP or ddATP). The samples were well mixed, incubated at $70°C$ and cooled to room temperature. $4\mu l$ of formamide loading buffer was added to each tube, heat denatured at $95°C$ for five minutes and loaded onto a 5% sequencing gel.

2.9   YEAST ARTIFICIAL CHROMOSOME (YAC) CLONING SYSTEM

The YAC vector system is suitable for cloning of very large DNA fragments up to several hundred kilobase pairs in size (Burke et al, 1987). This system is particularly appropriate in physical mapping studies and accelerates the process of chromosome walking and gene cloning.

2.9.1   YAC Vectors.

A simplified map of the yeast artificial chromosome vector (pYAC) is illustrated in Figure 2.1 (adapted from Burke et al, 1987). The vector was constructed from both yeast DNA sequence (thatched box) and plasmid pBR322 derived DNA (thin line), and incorporates all necessary functions into a single plasmid that can replicate in Escherichia coli. All the sequences necessary for an artificial chromosome are carried
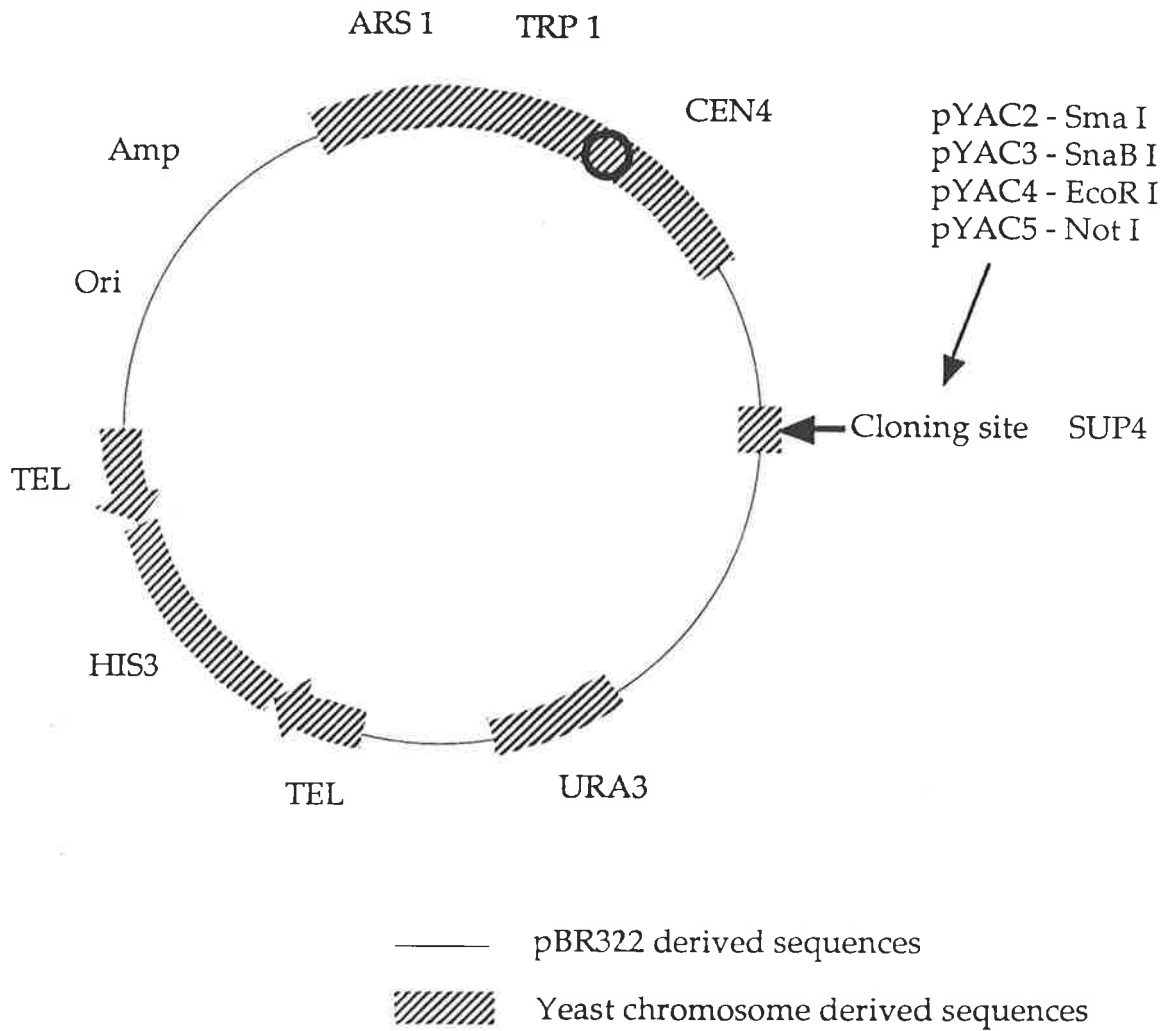
Fig. 2.1 Simplified map of pYAC vector (adapted from Burke *et al*, 1987). The vector, approximately 11.3 kb in size, is derived from yeast DNA sequences (hatched box) and the entire pBR322 DNA sequences (thin line).

by the vector. The centromere (CEN 4) autonomous replicating sequence (ARS 1), telomeres (TEL), selectable markers (TRP 1, URA 3) and SUP 4 gene are derived from yeast DNA, whilst ampicillin resistance gene (Amp) and origin of DNA replication (ORI) are from plasmid pBR322. The ARS I, CEN 4 and TEL sequences confer replication and mitotic/meiotic stability on the YAC during propagation in yeast. Therefore, a YAC can replicate in the same manner as its host's chromosomes. The selectable marker URA 3 is for positive selection of transformants in ura 3 hosts. The Amp gene and ORI are essential for growth and amplification of the YAC vector in *E. coli*. In each pYAC vector the cloning site is different, such as *Sma* I in pYAC 2, *Sna* BI in pYAC 3, *Eco* R1 in pYAC 4 and *Not* I in pYAC 5, although all of the cloning sites are in SUP 4 gene sequences. When exogenous DNA is cloned into the cloning site, SUP 4, an ochre-suppressing allele of a tyrosine transfer RNA gene, is interrupted and produces red colonies in place of white ones. Since pYAC contains the Amp gene, *E. coli* cells containing any of the pYAC vectors can grow in LB medium plus Amp.

2.9.2   YAC Library.

The YAC clones analysed in this project were isolated from a YAC library constructed by Dr D LePaslier and colleagues at the Centre d'Etude du Polymorphisme Humain (CEPH), Paris, France.   A complete human YAC library was produced as it was intended to serve as a source for high density physical mapping for the entire human genome. The library was constructed by ligation of partial *Eco* RI digested total human DNA with the pYAC 4 vector. The vector had been digested with *Eco* RI (cloning site) and *Bam* HI. The details of library construction are shown diagrammatically in Figure 2.2, are detailed below in 2.9.3 and have been published elsewhere (Burke *et al* 1987).
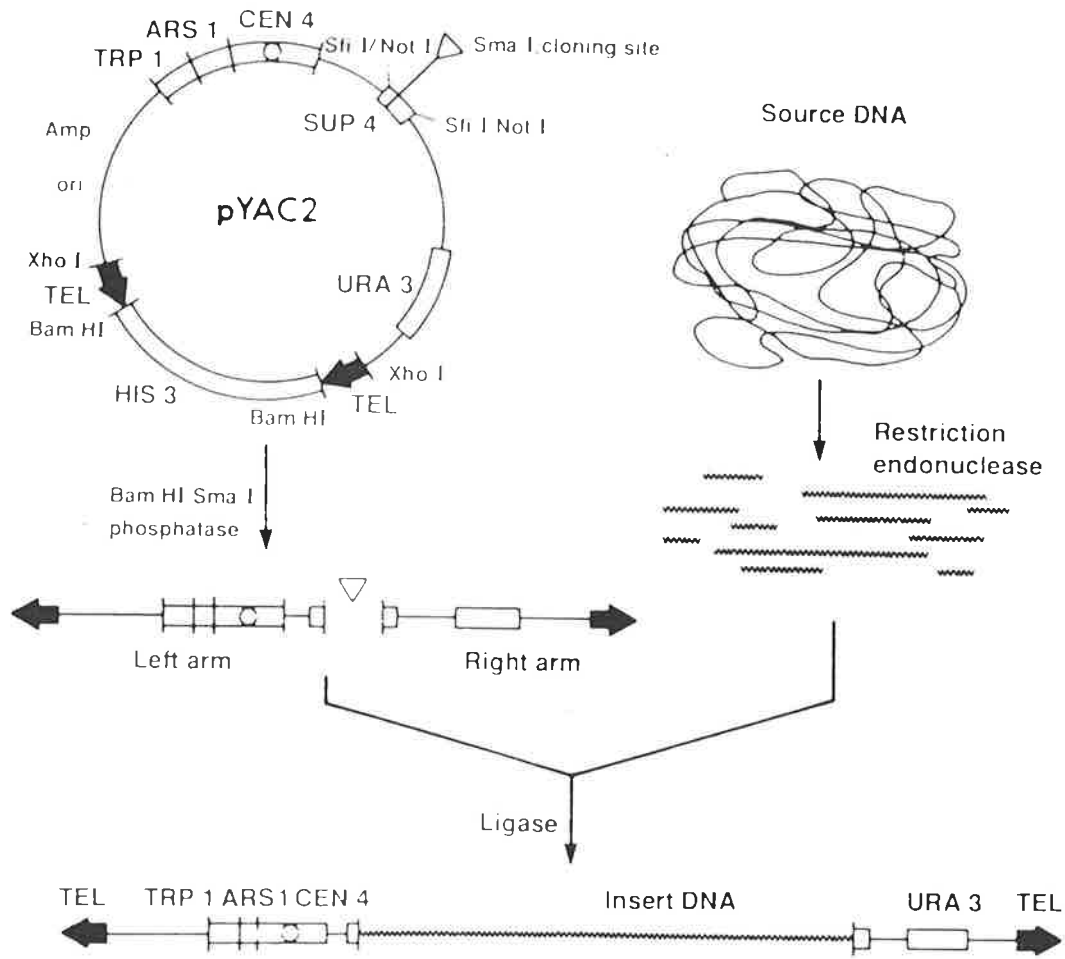
Fig. 2.2 The yeast artificial chromosome (YAC) cloning system (from Burke *et al*, 1987). pYAC 2 is chosen to demonstrate the YAC cloning procedure. The vector is digested with the restriction enzymes *Bam* HI and *Sma* I (cloning site). Three fragments are generated; the left arm (containing the centromere), the right arm and the discard *Bam* HI fragment containing H1S 3. The source DNA is digested with a restriction enzyme to generate large fragments with blunt ends (i.e. compatible with *Sma* I site). The left and right vector arms are ligated to either end of the DNA inserts to form the YAC.

2.9.3 YAC Cloning Strategy.

The strategies for cloning large human DNA fragments have been reported (Abidi *et al* 1990, Burke *et al* 1987). The YAC cloning procedure for pYAC 2 is presented in Figure 2.2.

In general, the pYAC 2 vector is prepared by double digestion with *Bam* HI and *Sma* I (cloning site). Three fragments are generated: the left arm (including the centromere), the right arm, and a discard region that separates the two TEL sequences in the circular plasmid. The two arms are then treated with alkaline phosphatase to prevent religation. Large human DNA inserts, obtained by partial digestion of high molecular weight human DNA with a restriction enzyme that leaves *Sma* I compatible ends (i.e. blunt ends) are ligated to the YAC vector arms. The ligation products are then transformed into yeast spheroblasts. The transformants containing an extra linear form YAC are selected for complementation of a *trp 1* marker, which ensures that the YAC contains both arms of the vector. Finally, they are tested for loss of expression of SUP 4 gene, which is interrupted by insertion of exogenous DNA at the *Sma* I cloning site (Burke *et al* 1987). The final structure of a YAC, shown at the bottom of figure 2.2, is that of a human DNA insert located between right and left YAC vector arms.

2.9.4 PCR Based Screening of the CEPH YAC Library.

The systematic screening of YAC libraries by use of the PCR have been described (Green and Olsen,1990a). The general approach for screening the human YAC library is illustrated in Figure 2.3.

Individual clones were grown in arrays of 384 colonies per nylon filter (in 96 well microtitre plates). The yeast cells from each filter were pooled and the DNA was purified, yielding single filter pools of DNA. Equal aliquots from single filter pools were
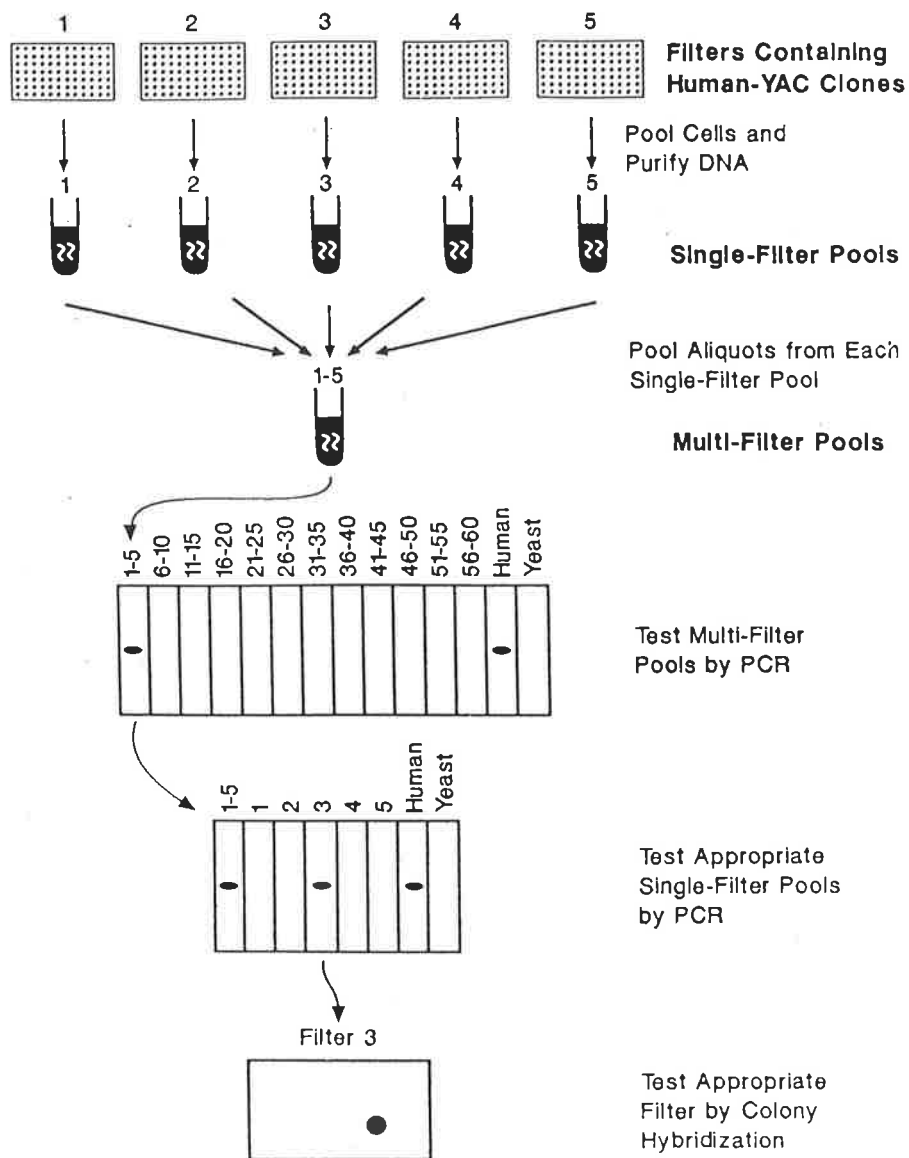
Fig. 2.3   Schematic representation of the PCR-based strategy for screening human YAC libraries (from Green and Olsen, 1990a).

mixed together in groups of four to yield multi-filter pools. A total of 113 multi-filter pools were provided to our laboratory, with the single filter pools and the filters containing the human YAC clones remaining at CEPH, France.

The multi-filter pools of DNA were then analysed individually for the presence of a specific human DNA segment using PCR. Approximately 7 ng of each multi-filter pooled DNA was added to a 20 μl final volume PCR combining 150 ng/μl of each oligonucleotide primer designed for a specific human DNA segment, optimized Mg^{2+}, and 0.05 units/μl Amplitaq. The PCR cycling conditions were: 94°C denaturation for 1 minute, 55°C annealing for 1.5 minutes and 72°C extension for 1.5 minutes, over 25 cycles. A final 72°C extension for 10 minutes was performed when the 25 cycles were completed. The PCR's were then separated on an agarose gel, ethidium bromide stained and visualized under UV light. A Southern blot filter of the PCR's was prepared and a radiolabelled PCR product derived from the PCR primers used in the screening was used as a probe to confirm any positive multifilter pool PCR signals.

If DNA from a multi-filter pool was found to generate the appropriate PCR product, then each constituent single filter pool was analysed individually by the same PCR assay. Upon generation of the appropriate PCR product with a single filter pool the location of the positive clone within the 96 well array was established by colony hybridization using the radiolabelled PCR product as a probe. The single filter pool PCR assay and colony hybridization was performed at CEPH, Paris, and positive YAC clones in stabs were returned back to our laboratory.

2.9.5   End Probes for YAC Mapping (YAC-L, YAC-R, pUC 19).

The YAC vectors were designed so that both end probes were easily derived from the plasmid pBR 322. The end probes had no cross hybridization with either yeast or human

genomic sequences.

When constructing the YAC vector (Figure 2.1), yeast DNA SUP 4 (containing the cloning site) was ligated into the *Bam* HI site of pBR 322 DNA (Figure 2.4, position 375). The pBR 322 DNA sequences from the *Bam* HI site (Position 375) to the *Pvu* II site (position 2066) containing Amp and ORI form part of the centromere arm (left arm) of the YAC vector, the rest of the pBR 322 sequences form part of the no centromere arm (right arm). Therefore, appropriate pBR 322 DNA sequences in the left arm or the right arm of the YAC vector can be used as an end probe to detect the corresponding end of a human DNA insert in a YAC. To construct YAC - left (YAC-L) and YAC - right (YAC-R) probes, pBR 322 DNA was digested with three restriction enzymes (*Bam* HI, *Pst* I and *Nru* I) to generate three DNA fragments. As shown in Figure 2.4, YAC-L, a 1 kb *Bam* HI/*Pst* I fragment detected the left arm of the YAC; while YAC-R, a 600 bp *Bam* H1/*Nru* I fragment detected the right arm. Plasmid vector pUC 19 contains a *Pvu* II/*Eco* RI fragment of pBR 322 (from position 2066 to 4361 containing Amp and ORI, Figure 2.4), so that it can be used as an end probe instead of YAC-L to identify the centromere arm of the YAC.

2.9.6   YAC Mapping Strategy.

In this project the restriction mapping of human DNA fragments in YAC's was performed primarily to determine whether the YACs' 275H10 and 377C4 overlapped one another and thus spanned the FRA16B locus. Since the human DNA fragment in a YAC is usually several hundred kilobases in size, restriction enzymes that can generate large DNA fragments are chosen for long range mapping.

In general, agarose embedded high molecular weight YAC DNA was digested to completion with various restriction enzymes. The restriction fragments were separated
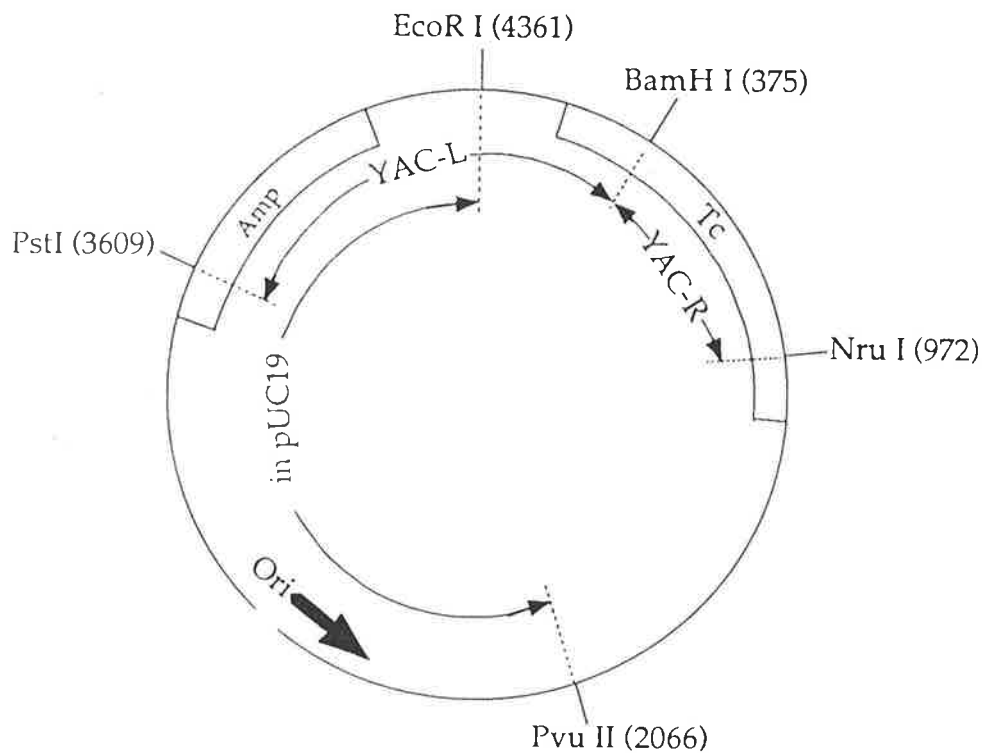
Fig. 2.4   Restriction map of pBR322 indicating the YAC end probes.   The entire

pBR322 sequence is contained in pYAC vector, with the *Bam* HI/*Pvu* II fragment

containing Amp and Ori in the left arm of the YAC vector and the remaining pBR322

sequence in the right arm.   For YAC restriction mapping, the *Bam* HI/*Pst* I fragment

(YAC-L) or pUC19 containing the *Eco* RI/*Pvu* II fragment was used as the endprobe

to identify the left arm of the YAC.   To identify the right arm of the YAC, the *Bam*

HI/*Nru* I fragment (YAC-R) was used.

by PFGE, Southern blotted onto nylon membranes, and successively hybridized with end probes (YAC-R, YAC-L or pUC 19), as well as human DNA probes, that had been labelled by the primer extension method (2.5.4.2). Standard prehybridization, hybridization and washing methods of the filters were performed (2.5.6). The size of a restriction fragment detected by one end probe indicated the distance from the restriction site to the corresponding telomere, so the restriction map could be built up from both ends. Complete digestion with the restriction enzymes can generate a complete map only for the enzymes which have one cleavage site in a human DNA insert. To generate a complete map of all restriction sites, partial digestion and double digestion of the restriction enzymes would have to be performed. However, for the construction of a restriction map to determine YAC overlap, complete restriction mapping of all restriction enzyme sites was not required.

2.9.7   End Cloning Human DNA Fragments in YACs.

The isolation and subcloning of both ends of the human insert in YAC's was essential to commence chromosome walking towards FRA16B. The isolation of unique YAC end fragments as new marker loci can be aided by the presence of YAC vector sequences adjacent to the cloned termini. Two methods which exploit this fact; vector/*alu* PCR and capture PCR, were used to recover YAC end clones.

2.9.7.1 Vector/*alu* PCR.

The use of *alu* PCR and vector-*alu* PCR to isolate YAC end clones have been described (Breukel *et al* 1990, Ledbetter *et al* 1990, Orita *et al* 1990, Tagle and Collins, 1992). Figure 2.5 illustrates the basic procedure for the vector-*alu* PCR assay.

In this procedure, 100 ng of YAC DNA was added to a 20μl final PCR containing 1μM of either internal biotinylated YAC left primer (BYAC-L) or biotinylated internal YAC

right primer (BYAC-R), and 1μM of *alu* primer. A range of five different *alu* primers were used (see Table 2.1). The Mg^{2+} concentration was optimized for the reaction, and 0.05 units/μl of Amplitaq was added. The cycling parameters were: 94°C for two minutes, 55°C for two minutes and 72°C for three minutes. Thirty cycles were performed.

The integrity and location of the amplified PCR products were confirmed by repeating the PCR assay but replacing the BYAC-L and BYAC-R primers with the YAC4-L and YAC4-R primers located 5' downstream (see Figure 2.5). For positive confirmation, the size of the new PCR product should be smaller than the first product by a size difference that reflects the distance from the internal primer to the biotinylated primer.

The YAC derived PCR primers used in this procedure are listed on Table 2.2.

2.9.7.2 Capture PCR.

The recovery of end clones by capture PCR and related methods have been reported (Riley *et al* 1990, Lagerstrom *et al* 1991, Rosenthal *et al* 1991). The following method is a modification adapted from the procedure described by Lagerstrom *et al* (1991). The general procedure for capture PCR is illustrated in Figure 2.6.

In this procedure, 200 ng of YAC DNA containing human insert was digested with a restriction enzyme that produced either a 3'-GC overhang or blunt ends (Figure 2.6A). The resulting DNA fragments were ligated to a synthetic oligonucleotide duplex containing an appropriate GC overhang or blunt ends which would anneal to the fragment ends of the digested YAC (Figure 2.6B). For this ligation reaction 100 ng of digested YAC DNA was incubated with a pair of oligonucleotide linkers (linker - 1 plus "sticky end" oligo, for GC overhang ligation, or linker - 1 plus "blunt end" oligo for blunt end ligation) at 2μM each. Five units of restriction enzyme and four units of T4
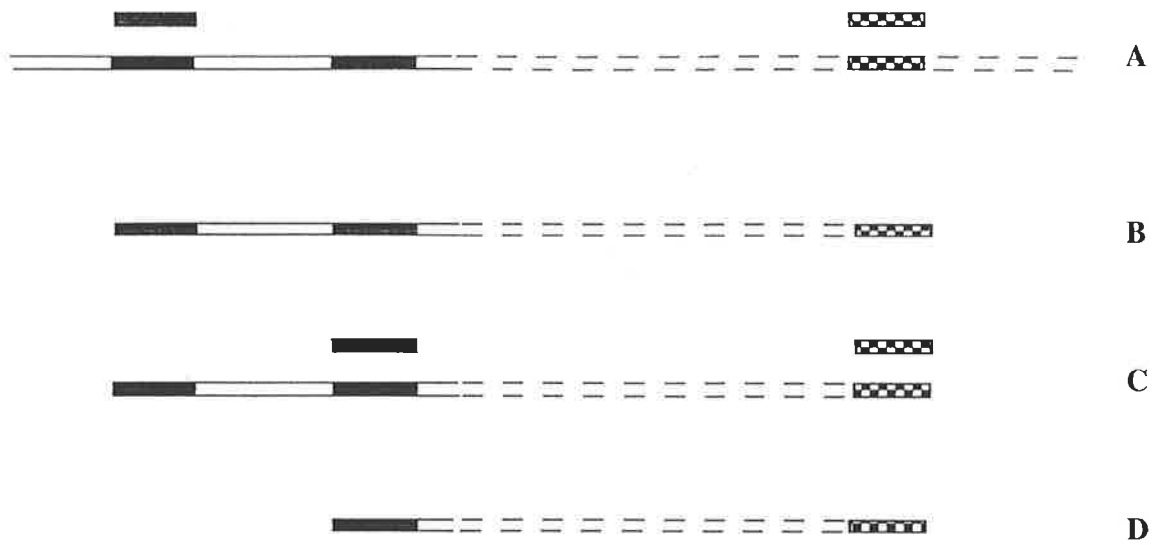
Fig. 2.5 A schematic representation of vector/*alu* PCR. The solid lines indicate yeast DNA and the dashed lines human insert DNA. PCR extensions are performed (A) using YAC DNA as template together with a specific yeast primer (solid box) and a primer specific for *alu* repeat DNA (hatched box). The vector/*alu* PCR product (B) may then be used as a template for a PCR with a second yeast specific primer located downstream from the first and the same *alu* primer, to confirm the integrity of the PCR product. A PCR product of the expected size should result (D).

Table 2.1 *Alu* PCR Primer Sequences.

Primer	Sequence
alu 1 *	TGAGCCGAGATCGCGCCACTGCACTCCAGCC AGCCTGG
alu 2 * (559)	AAGTCGCGGCCGCTTGCAGTGAGCCGAGAT
alu 3 * (517)	CGACCTCGAGATCTYRGCTCACTGCAA Y - pyrimidine R - purine
alu 4**	GGATTACAGGCGTGAGCCAC
alu 5**	GATCGCGCCACTGCACTCC

* *alu* primer sequences 1, 2, 3 were published in Nelson *et al* (1989).

** *alu* primer sequences 4 and 5 were published in Tagle and Collins (1992).

Table 2.2   Synthetic Oligonucleotide and YAC Derived Oligonucleotide Sequences.

Primer Name	Sequence 5' to 3'
YAC 4-L	CGG AAT TCG CCA AGT TGG TTT AAG GCG CAA GAC
YAC 4-R	GGA AGC TTG GCG AGT CGA ACG CCC GAT CTC AAG
Linker 4-R	GCG GTG ACC CGG GAG ATC TGA ATT C
Sticky end	CGG AAT TCA GAT
Blunt end	GAA TTC AGA TC
BYAC-R	ATG CCG GCC ACG ATG CGT CCG GCG
BYAC-L	TCG GAG CAC TGT CCG ACC GCT TTG

Note: The primer sequences for YAC 4-L and YAC 4-R were kindly provided by Dr D Nelson, Institute for Molecular Genetics Baylor College of Medicine, Houston, Texas. The remaining primer sequences were published (Lagerstrom *et al* 1991).

DNA ligase were added to a final volume of $10\mu l$ containing 0.5mM ATP, 10mM Tris - HCl pH 7.5, 10mM $MgCl_2$ and 50mM potassium acetate. The reaction was incubated at 37°C for three hours.

Multiple extension rounds were then performed using the biotinylated oligonucleotide primer derived from the YAC right arm or YAC left arm vector sequence (Figure 2.6C). 1 μl of the ligation mix was added to a final volume of $20\mu l$ PCR buffer containing $0.1\mu M$ biotinylated YAC vector (B YAC-L or BYAC-R), optimized Mg^{2+} and 0.05 units/μl of Amplitaq. The sample was overlaid with mineral oil and the temperature was varied between 94°C for one minute, 55°C for one minute and 72°C for 2.5 minutes, in 35 cycles followed by a final incubation at 72°C for seven minutes.
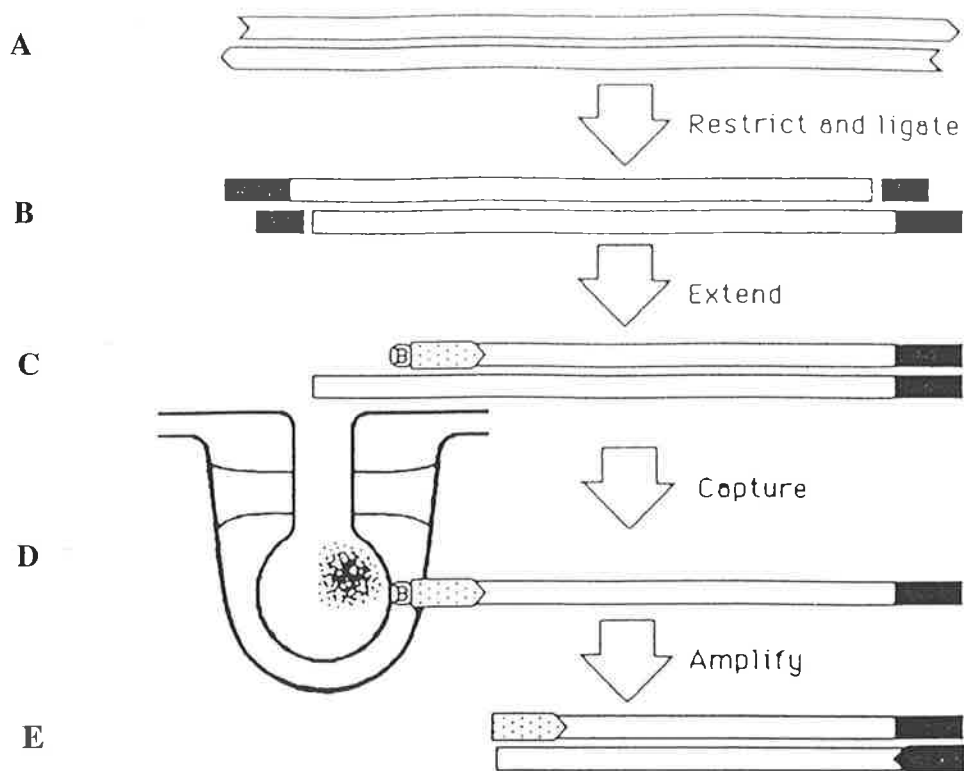
Fig. 2.6 A schematic representation of capture PCR from Lagerstrom *et al* 1991. Linker oligonucleotides are shown in black and primers specific for a given genomic region are dotted. Genomic DNA is digested with a restriction enzyme (A) and in the same reaction a linker is ligated to the cleavage site (B). The sample is subjected to PCR extension using a 5' biotinylated primer (C). Biotin labelled extension products are captured (D). PCR is performed using the isolated extension products as templates, with a second specific primer hybridizing downstream of the biotinylated oligonucleotide and the linker oligonucleotide (E).

The biotin-labelled extension products were then added to $40\mu l$ of pre-washed streptavidin-coated magnetic particles in an eppendorf centrifuge tube and incubated at room temperature for 30 minutes. The biotinylated PCR products bound to the streptavidin beads. Using a magnet, the beads containing the biotinylated PCR products were immobilized to the side of the centrifuge tube (Figure 2.6D). The beads were washed once in a washing solution of 0.1M Tris-HCl, pH7.5, 1M NaCl, and 0.1% Triton X-100, once with a denaturing solution of 0.1M NaOH, 1MNaCl, 0.1% Triton X-100, and again with washing solution. Finally, the immobilized PCR products were washed and resuspended in $10\mu l$ PCR buffer.

The extension products were then used as templates in a nested PCR reaction (Figure 2.6E). The bound streptavidin-coated magnetic beads were present during the entire PCR. $2.5\mu l$ of the bound PCR products were used as a template in a $50\mu l$ PCR reaction containing the linker-1 and a specific primer (YAC 4-R or YAC 4-L), representing a sequence located downstream to the biotinylated primer, at 1 μM each. The same PCR cycling conditions were used as in the first reaction. The amplified DNA sequences were separated on an agarose gel and the PCR products were visualized by standard means. The PCR primers generated for the capture PCR method are listed in Table 2.2.

2.10   COSMID VECTORS

Cosmid vectors were originally designed to clone and propagate large segments of genomic DNA. In their simplest form, cosmid vectors are modified plasmids that carry the DNA sequences (*cos* sequences) required for packaging DNA into bacteriophage lambda particles. Because cosmids carry an origin of replication and a drug resistance marker, cosmid vectors can be introduced into *E. coli* by standard transformation procedures and propagated as plasmids.

Cosmid vectors have been constructed to contain a large variety of structural elements designed to improve sequence representation, to simplify or expedite the structural or functional analysis of cloned DNA, or to simplify the construction of high quality representative libraries.

2.10.1 Scos Vectors.

Scos vectors are a family of vectors that have been designed specifically for the mapping and functional analysis of human chromosomes. The construction of these vectors, described elsewhere (Evans *et al* 1989), included the presence of two *cos* sites so that packaging could be carried out with high efficiency and without requiring size selection of the insert DNA. The vectors also include the presence of T7 and T3 bacteriophage promotors for the synthesis of "walking" probes. Unique restriction sites were incorporated for the removal of the insert from vector, for example; *Not* I sites for Scos 1, *Not* I and *Sac* II sites for Scos 2 and *Sfi* I sites for Scos 4. To aid in restriction mapping, the vectors contained selectable genes for gene transfer in eukaryotic cells (i.e. Amp and SV2 Neo). A plasmid origin of replication (ori) was also included for giving high yields of cosmid DNA when preparing templates. In this project Scos 1 cosmid vector was used. A simplified map of the Scos vector including the nucleotide sequence of the Scos 1 cloning site, is presented in Figure 2.7.

2.10.2 Cosmid Cloning Strategy.

The cosmid cloning strategy has been described in detail elsewhere (Hohn and Collins, 1980, Chia *et al*, 1982). In brief, segments of foreign DNA approximately 35-45 kb in length are isolated and ligated to linearized cosmid vector DNA to form a structure as shown in Figure 2.8. Two *cos* sites, arranged in the same orientation flank the foreign DNA. Also located within this region of DNA is an entire complement of plasmid genes.
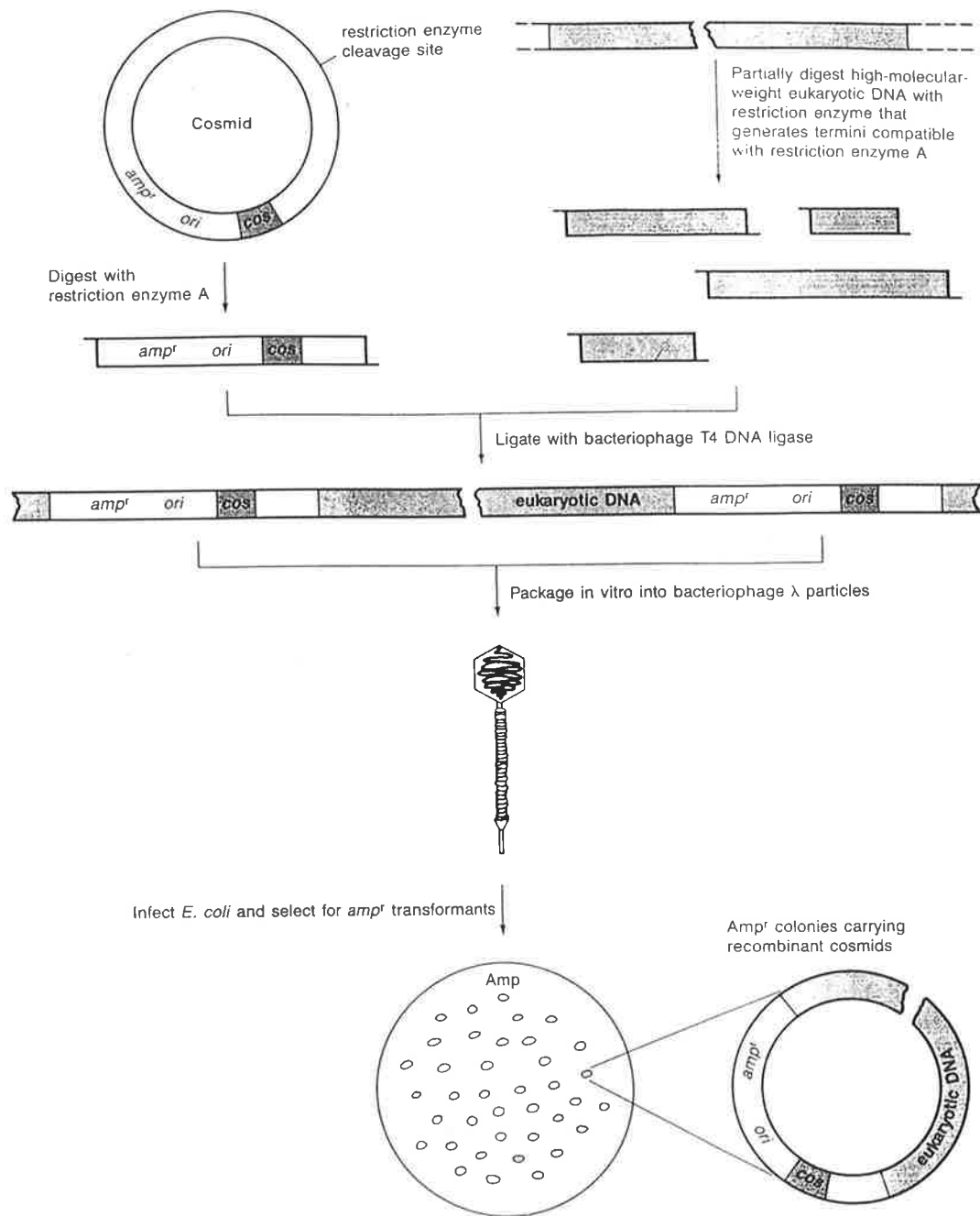
A.

B.

Fig. 2.7 A simplified map of the Scos vector (A), including the nucleotide sequence of the Scos 1 cloning site (B).

Fig. 2.8 The cosmid cloning strategy (from Current Protocols in Molecular Biology). Segments of foreign DNA are isolated and ligated to linearized cosmid vector DNA. These complexes are then used as substrates in an *in vitro* packaging reaction into bacteriophage lambda particles. *E. Coli* is infected by the bacteriophage particles and bacteria carrying recombinant cosmids can be selected using media containing the appropriate antibiotic.

These complexes are then used as substrates in an *in vitro* packaging reaction whereby the *cos* sites are cleaved by the *ter* function of the bacteriophage lambda gene. A protein and the DNA between the two *cos* sites is packaged into mature bacteriophage lambda particles. Bacteriophage lambda heads will accommodate up to 52 kb of DNA (Williams and Blattner,1979) therefore with conventional cosmid vectors accounting for up to 5 kb of DNA, recombinant cosmids containing up to 47 kb of foreign DNA can be constructed. The smallest piece of DNA that can be packaged into bacteriophage lambda particles is approximately 38 kb (Williams and Blattner,1979) so the minimum size of DNA that can be cloned is approximately 33 kb. During infection of *E. coli* by the bacteriophage particles, the linear recombinant DNA is injected into the cell, and via the cohesive ends of the *cos* sites the DNA is circularized. The resulting circular molecule contains a complete copy of the cosmid vector and replicates as a plasmid conferring drug resistance upon its bacterial host. Therefore bacteria carrying recombinant cosmids can be selected using media containing the appropriate antibiotic.

2.10.3   Construction of a Chromosome 16 Specific Ordered Cosmid Library.

The chromosome 16 specific cosmid library was constructed at the Los Alamos National Laboratory, New Mexico and has been described (Stallings *et al* 1990). In general, human chromosomes 16 were isolated from a somatic cell hybrid CY 18. A single chromosome 16 was the only human chromosome present in this hybrid. After partial digestion with *Sau* 3a and dephosphorylation with calf intestinal alkaline phosphatase, the chromosomal DNA was ligated to the cloning arms from the cosmid vector Scos 1. *In vitro* packaging and infection of *E. coli* yielded 1.75×10^5 independent recombinants, giving a 67 fold statistical representation.

2.10.3.1 Construction and Screening of High Density Cosmid Grids (performed at Los Alamos National Laboratory, New Mexico).

A Beckman Biomek 1000 was used to stamp bacterial colonies onto Biodyne nylon hybridization membranes (1536 clones/membrane) as already described (Longmire *et al* 1991). Membranes were hybridized overnight in 6xSSC, 10mM EDTA pH 8.0, 10X Denhardts, 1% SDS, 0.1 mg/ml denatured sonicated salmon sperm DNA, at 65°C. Following hybridization membranes were washed in 2 x SSC, 0.1% SDS at room temperature (quick rinse); once in 2x SSC, 0.1% SDS at room temperature for 15 minutes; and twice in 0.1% SDS at 50°C for 30 minutes. Probes were labelled with ^{32}P to a specific activity of $\rangle 10^8$ cpm/μg by primer extension labelling (2.5.4.2) and prereassociated (2.5.4.4) prior to hybridization of the probe to the filters.

2.10.3.2 Repetitive Sequence Fingerprinting and Assembly of Contigs (performed at Los Alamos National Laboratories).

An approach has been developed for the identification of overlapping cosmid clones by exploiting the high density of repetitive sequences in complex genomes as described by Stallings *et al* (1990).

By coupling restriction digestion mapping with oligomer probes targeting abundant interspersed repetitive sequences such as *alu* (Jelinek and Schmid, 1982), L1 (Scott *et al* 1987) and (GT)n (Rich *et al* 1984, Weber and May, 1989), a "fingerprint" is obtained. The initial analysis of fingerprint data are the pairwise comparison of the fingerprint between cosmid clones. Clones that overlap will share restriction fragments of similar size with similar repetitive sequences. A probability of overlap is assigned to each pair of clones based on the number of shared restriction fragments with the same repetitive sequences. The analysis of the treatment of fingerprint data and the algorithm used to detect pairwise overlap has been described (Balding and Torney,1991). Once overlapping

pairs have been identified, the clones are ordered and assembled into contigs based upon the information in overlapping pairs. For contig construction a computer programme was used based on a genetic algorithm - genetic contig assembly algorithm, GCAA (Fickett and Cinkostsky, 1992). GCAA represents possible maps as strings of numbers encoding the lengths and positions of the clones. The quality of any possible map is measured by a fitness function that takes into account the overlap likelihoods, overlap extents and clone lengths that the map is intended to fill. By the repetitive sequence fingerprinting of approximately 4000 cosmid clones obtained from the chromosome 16 specific library, a cosmid contig map was developed (Stallings *et al* 1992a). The clones were organised into 576 contigs and 1171 cosmid clones not contained within a contig.

2.10.4   Isolation and Hybridization of Cosmid Endprobes.

Cosmid DNA was digested with a range of restriction enzymes including *Pst* I, *Pvu* II and *Xmn* I according to the manufacturer's instructions.   The digests were electrophoresed and nylon filters were prepared by Southern blot transfer (2.5.5.2). DNA fragments representing the terminal sequences of the genomic insert were identified by the successive hybridization of T7 and T3 bacteriophage promotor sequences to the filters. These sequences flank the Scos 1 *Not* I cloning site. T7 and T3 oligomers were labelled by 5'-end labelling (2.5.4.1). Prehybridization, hybridization and washing of the filters have been described (2.5.6).

Preparative gels were made and the DNA fragments that hybridized to the T7 and T3 oligomers were excised and purified by the Prep-a-Gene method (2.3.1.8.1). The length of genomic DNA represented in either T7 or T3 generated endprobes was determined by subtracting the length of vector DNA present from the total length of the isolated DNA fragment. For each restriction enzyme used to generate an endprobe, the amount

of vector DNA present was calculated by establishing the distance from the cloning site to the first restriction enzyme site in either the T7 or T3 end of the vector. The T7 and T3 vector lengths have been calculated for *Pst* I, *Pvu* II and *Xmn* I restriction enzymes and are presented in Table 2.3.

Table 2.3 **The Distance (in kb) from the Insert Cloning Site to the First Restriction Enzyme Site for T7 and T3 vector ends of Scos 1.**

Restriction Enzyme	T7 end	T3 end
Pst I	2.0 *	0.7
Pvu II	2.6 *	3.2
Xmn I	0.27	0.4

*   The distance calculated for these restriction enzyme sites account for the approximately 1 kb of DNA that is cleaved between the two *cos* sites (from position 1673 to 1733 at T7 end of vector) during bacteriophage packaging.

CHAPTER 3

LONG RANGE RESTRICTION MAPPING

OF GENOMIC DNA AT

16q13 - 16q22.1

3.1   SUMMARY

Using single copy probes for five genes <u>ALDOA</u>, <u>CETP</u>, <u>LCAT</u>, <u>MT1B</u>, <u>UVO</u> and three anonymous DNA markers D16S4, D16S91 and D16S124, a non contiguous long range restriction map totalling approximately 2.1 Mb of DNA was constructed within the chromosome region 16q13 - 16q22.1. A long range restriction map encompassing 990 kb of DNA established physical linkage between D16S4 and D16S91. The two anonymous DNA markers are separated by a maximum physical distance of approximately 450 kb and may span the M4Eo leukaemia long arm breakpoint. A long range restriction map of 150 kb was constructed around the genes <u>LCAT</u> and <u>ALDOA</u> separating the two genes at a maximum physical distance of 70 kb and finely mapping <u>ALDOA</u> to 16q22.1. The genes <u>MT1B</u> and <u>CETP</u> were found to be physically linked at a maximum distance of 770 kb and a long range restriction map of 1000 kb was established around them. The PFGE analysis of gene probes and anonymous DNA markers from 16q13-16q22.1 did not establish a long range restriction map spanning the <u>FRA16B</u> locus.

3.2 INTRODUCTION

The chromosomal location of the distamycin A inducible fragile site <u>FRA16B</u> has been known for almost 15 years but cloning of the mutation itself has not been achieved. The main reason has been the lack of closely linked DNA markers in this region. By 1989, a number of genes and anonymous DNA markers had been identified in proximity to this fragile site. By genetic linkage analysis the closest polymorphic markers known to flank <u>FRA16B</u> were D16S10 proximal and D16S4 distal (Mulley *et al*, 1989). The markers were in close proximity on both the physical and genetic maps and delineated a chromosomal segment of up to 7 Mb. As a first step towards the cloning of <u>FRA16B</u> long range restriction mapping provided an ideal technique to cover such a large distance because DNA fragments up to several Mb in size could be analysed.

In this section of the project the aim was to generate a genomic DNA based long range restriction map of the chromosome region spanning 16q13-16q22.1, thereby including <u>FRA16B</u> located at the 16q21-16q22.1 interface. Analysis of the region was divided into two separate areas - proximal to <u>FRA16B</u> (16q13-16q21) and distal (16q21-16q22.1). In each region DNA markers had been finely mapped into sections with respect to translocation breakpoints and/or <u>FRA16B</u>. Using a panel of restriction enzymes that cleaved genomic DNA infrequently, high molecular weight genomic DNA fragments were resolved using PFGE. A PFGE derived map could be constructed by physically linking adjacent DNA probes to identical high molecular weight genomic DNA fragments. Using this approach, the order of markers within sections and the physical distances between them could be established. The long range restriction maps constructed from the two regions were to be linked together to form a contiguous map encompassing <u>FRA16B</u>. A contiguous map would physically

identify the markers flanking this fragile site and the distance separating them. Localizing the fragile site within a defined boundary would greatly facilitate its eventual molecular characterization.

3.3   MATERIALS AND METHODS

3.3.1   DNA Probes.

Details of the probes used in this study are given in Table 3.1. Probes for D16S4, D16S91, D16S124, ALDOA, CETP, LCAT, UVO and MT1B were identified as being single copy and probes for D16S163 and D16S174 were identified as containing repetitive DNA. The localizations of the probes with respect to hybrid cell line breakpoints are shown on a physical map of the long arm of chromosome 16 (Figure 3.1). DNA inserts were purified from plasmid vectors according to the methods already described (2.3) and were labelled as hybridization probes by the primer extension method (2.5.4.2). The probes containing repetitive DNA were subjected to prereassociation prior to hybridization (2.5.4.4).

3.3.2   Hybrid Cell Line Panel.

The DNA from a number of somatic cell hybrids containing either a cytogenetically normal chromosome 16 or portion thereof, and human parental cell line DNA (i.e. the cell line from which the somatic cell hybrid was derived) were used in this study. The portion of chromosome 16 present in each cell line are detailed in Table 3.2. In addition, the portion of chromosome 16 present in the hybrid cell lines displayed on the physical map of the long arm of chromosome 16 (Figure 3.1) are included in Table 3.2. The construction of somatic cell hybrids and details of the human parental cell lines have been previously described (Callen, 1986).

3.3.3   DNA Preparation and Restriction Digests for PFGE.

DNAs for digestion with the restriction enzymes *Bss* HII, *Mlu* I, *Nae* I, *Not* I, *Nru* I, *Sac* II and *Sal* I were isolated from lymphoblastoid cell lines from normal individuals and hybrid cell lines within agarose beads, as already described (2.6.1.1).

Restriction digestion of agarose embedded DNA was performed using the

Table 3.1.   **DNA Probes Mapping to the Long Arm of Chromosome 16 Used in this Study.**

LOCUS	PROBE	INSERT SIZE (kb)	INSERT SITE	REFERENCE
D16S4	ACH207	4.7	*Hind*III	Hyland *et al* (1988)
D16S91	LE 12	1.2	*Eco*R1	Callen *et al* 1988
D16S124	1.99	0.88	*Pst*I	Hyland *et al* (1988)
D16S163	16-10	2.6	*Hind*III	Chen *et al* (1991)
D16S174	16-53	2.8	*Hind*III	Chen *et al* (1991)
ALDOA	pHAAL 116-3	0.7	*Pvu*II/*Pst*I	Mukai *et al* (1991)
CETP	pCETP.11	1.5	*Eco*R1	Drayna *et al* (1987)
LCAT	pLCAT.2	0.8	*Sst*I	Callen *et al* (1988)
MT1B	hMT-11ASC	1.3	*Bam*HI/*Hind*III	West *et al* (1990)
UVO	pV961	1.3	*Eco*R1	Mansouri *et al(1988)*
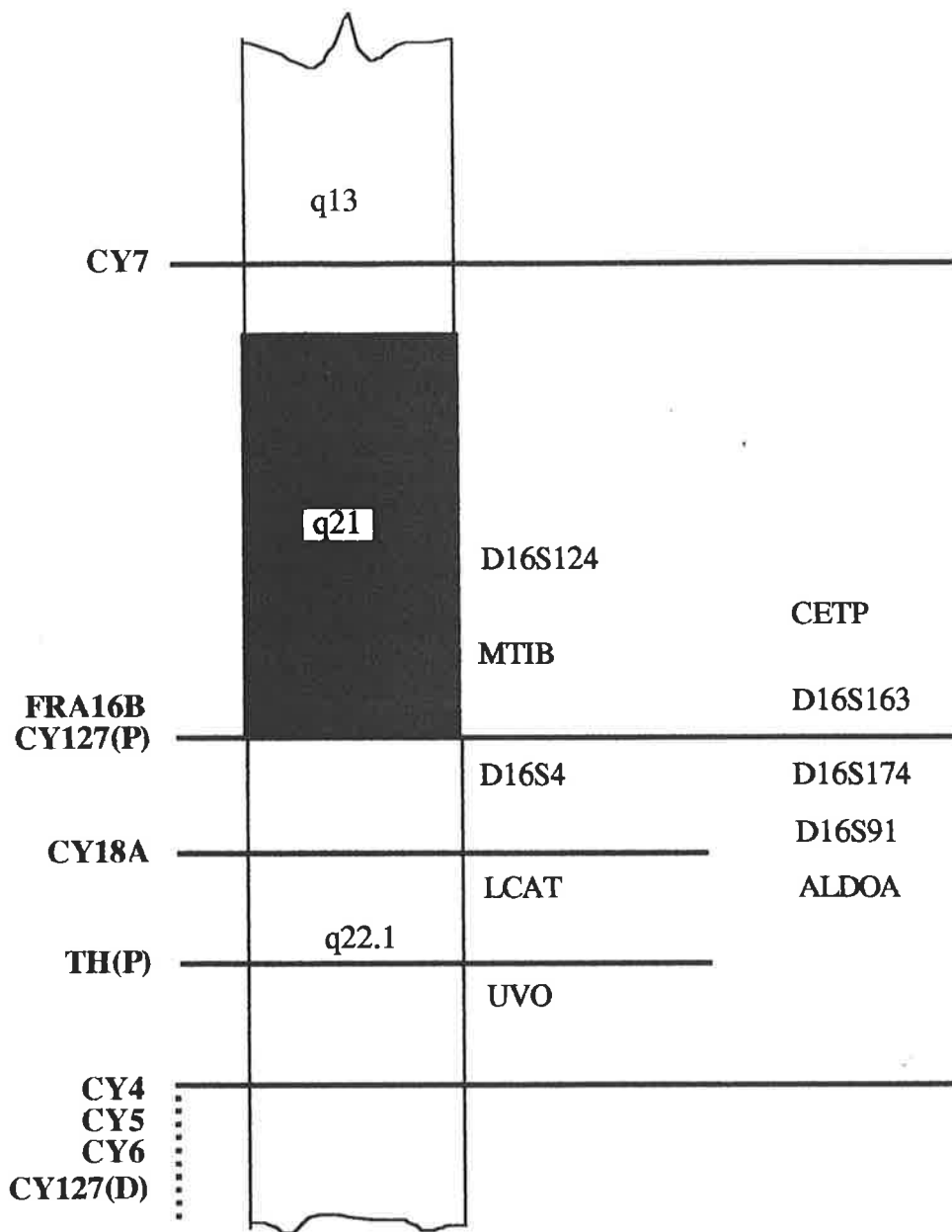
Fig. 3.1. Physical map of the chromosome region 16q13-16q22.1. The horizontal lines indicate the breakpoints of hybrid cell lines (CY) and the fragile site (FRA16B). CY127 and TH are interstitial deletions. Their proximal breakpoints are located by "P" and their distal breakpoints by "D".

Table 3.2   Hybrid Cell Lines and Parental Cell Lines Containing DNA from the Long Arm of Chromosome 16 used in this Study.

CELL LINE	PORTION OF 16 PRESENT	REFERENCE
CY 18	Complete 16	Callen, (1986)
CY 5	q22.3 -)qter	Callen, (1986)
KS (del)	pter - q22.1, q23.2 -qter	-
TH (del)	pter - q22.1, q22.3 -qter	-
CY 127	pter -)q21, q22.1 -)qter	-
CY 4 (parental)	pter - q 22.1, q22.1 -qter	Callen *et al*, (1990a)
CY 5 (parental)	pter - q22.1, q22.1 -qter	Callen, (1986)
CY 6 (parental)	pter - q22.1, q22.1 -qter	Callen *et al*, (1988)
CY 120 (parental)	pter - q22.1, q22.1 -qter	Callen *et al* (1990a)
CY 4	q22.1 -)qter	Callen *et al*, (1990a)
CY 5	q22.1 -)qter	Callen, (1986)
CY 6	q22.1 -)qter	Callen *et al*, (1988)
CY 7	q13 -)qter	Callen, (1986)
CY 18A	q22.1 -)qter	-

manufacturers recommended buffers by the methods already detailed (2.6.2).

3.3.4   PFGE.

PFGE was performed using the LKB Pulsaphor Plus 2015 apparatus with CHEF inserts by the methods described (2.6). High molecular weight DNA fragments were separated in two ranges of resolution.

(i)   high resolution PFGE. DNA fragments in the size range of 50kb-900kb were resolved. The switching interval was a 70 second pulse for 24 hours at 150 mA.

(ii)   low resolution PFGE. DNA fragments in the size range 200kb-1600kb were resolved. Low resolution PFGE was performed using ramped switching intervals of 200 sec for 12 hours, 150 sec for 12 hours and 100 sec for 12 hours, at 150 mA.

Lambda and yeast PFGE size markers (2.6.5) were used. Any modifications to the switching intervals are detailed where appropriate.

3.3.5   Southern Blotting and Hybridization.

After electrophoresis Southern blot filters were made from pulsed field gels by the method described (2.5.5.2). Standard prehybridization, hybridization and washing of filters was performed (2.5.6). Probes were stripped from filters before hybridization with the next probe.

3.3.6   Construction of Unique Sequence PCR Products from D16S163 and D16S174.

The 2.6 kb insert for D16S163 and the 2.8 kb insert for D16S174 were assymetrically ligated and subcloned into *Eco* R1/*Hind* III cloning sites of M13mp18 and M13mp19 vectors. DNA sequencing of human DNA inserts were performed using the chain termination method as described (2.8), with the M13-21 universal

primer (5'-AAC AGC TAT GAC CAT G-3', Messing *et al* 1977). The consensus sequence data generated from either end of the human DNA inserts were compared to known mammalian repetitive DNA sequences such as *alu* and L1 (Jelenik and Schmid, 1982., Scott *et al*, 1987., Wiener *et al* 1986., Moyzis *et al* 1987) using the genome data base Genbank. Using DNA sequences that did not exhibit any homology with human repetitive DNA, PCR primers were constructed for a unique sequence PCR product for each of the anonymous DNA probes.

Using genomic DNA as template and the appropriate forward and reverse PCR primers the PCR product was amplified under the following cycling conditions. A step cycle with a 94°C denaturation for one minute, 60°C annealing for 1.5 minutes and 72°C extension for 1.5 minutes was performed for 10 cycles followed by a 94°C denaturation for one minute, 55°C annealing for 1.5 minutes and 72°C extension for 1.5 minutes for 25 cycles. A final 10 minute extension of 72°C was performed. The PCR product was separated on a low melting point agarose gel in 1 X TAE (2.5), the gel slice extracted and purified using the Prep-a-gene method (2.3.1.8.1). The PCR product was radiolabelled using the primer extension method (2.5.4.2).

3.4   RESULTS

3.4.1   Analysis of High Molecular Weight Restriction Fragments with Unique Probes from the Region Distal to <u>FRA16B</u> (16q21-16q22.1).

The hybridization fragments observed for D16S4, D16S91, <u>LCAT</u>, <u>ALDOA</u> and <u>UVO</u> analysed by high and low resolution PFGE using the restriction enzymes *Sal* I, *Not* I, *Nru* I and *Mlu* I, are listed in Table 3.3. Lymphoblastoid cell line DNA from three unrelated individuals were used in the analysis and undigested lymphoblastoid cell line DNA was included as a control. The approximated and averaged sizes in kb are shown for each marker and for four enzymes.

3.4.1.1   Linkage between D16S4 and D16S91.

The successive hybridization of probes for D16S4 and D16S91 to high and low resolution PFGE filters indicated physical linkage between the two markers. The high resolution filter produced an identical 450 kb *Mlu* I fragment of high intensity for the two probes (Figure 3.2). D16S4 also hybridized to a less intense 230 kb *Mlu* I fragment and D16S91 hybridized to an additional 200 kb *Mlu* I fragment (Figure 3.2). A 90 kb D16S91 derived *Mlu* I fragment was observed in one lymphoblastoid cell line DNA sample (Figure 3.2, lane 2). The 230 kb and 200 kb fragments did not align on the same filter and approximately totalled the size of the common 450 kb *Mlu* I fragment. It is proposed that the two markers are physically linked to a common 450 kb *Mlu* I fragment and that an *Mlu* I restriction site is located between the two markers but is poorly cleaved. The 90 kb D16S91 derived *Mlu* I fragment may indicate an additional *Mlu* I restriction site between the two markers that is also rarely cleaved. Varying degrees of methylation at these restriction sites may influence cleavage.

Table 3.3   Fragment Sizes (kb) of Markers Distal to <u>FRA16B</u> Analysed by PFGE.

ENZYMES				
LOCUS	*Sal* I	*Not* I	*Nru* I	*Mlu* I
D16S4	70 380 650 990	1150 1250	170 370	230 450 990
D16S91	180 700	260 430 630 730	790 1280	90 200 450 990
<u>LCAT</u>	-	〈 50	180	70
<u>ALDOA</u>	-	〈 50 90	〈 50 130 230 350	70 270
<u>UVO</u>	320 480 570	650	220 260 440 500 750 930 1070	1420 〉1600

Note: " - " indicates that individual hybridization fragments could not be resolved.
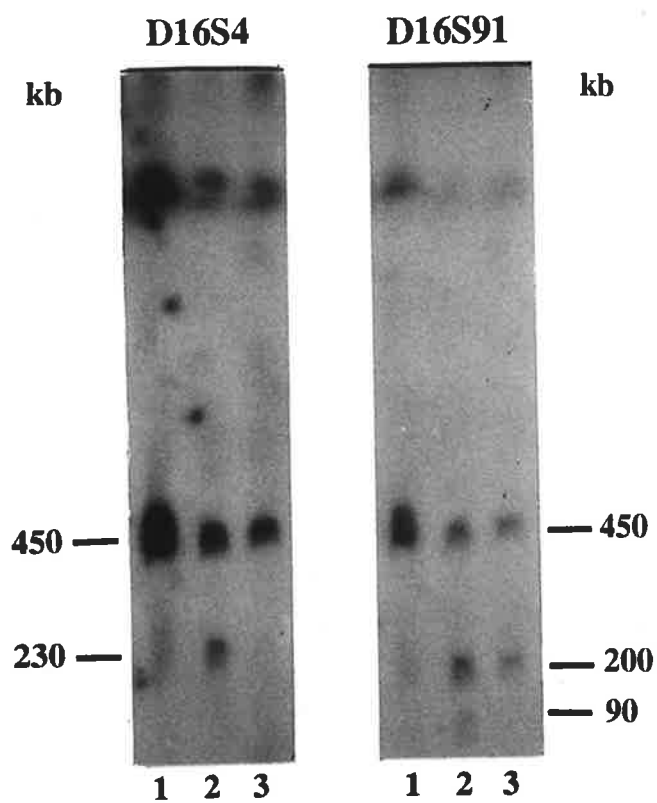
Fig. 3.2. Identification of a common 450 kb *Mlu* I fragment between D16S4 and D16S91. The two probes were hybridized to the same high resolution PFGE filter containing *Mlu* I digested DNA samples from three unrelated individuals (1, 2, 3).

The low resolution PFGE filter identified the same common 450 kb *Mlu* I fragment for D16S4 and D16S91 (Figure 3.3). An identical 990 kb *Mlu* I fragment was also shared by both probes (Figure 3.3) further supporting a physical link. No additional bands were detected by either probe. Partial cleavage of one of the *Mlu* I sites flanking D16S4 and D16S91 would give rise to the common 990 kb band. *Not* I, *Nru* I and *Sal* I digests of lymphoblastoid cell line DNA did not produce any common fragments between D16S4 and D16S91 (see Table 3.3) indicating that at least one of each of these restriction sites lie between the two markers.

To further verify the physical link between D16S4 and D16S91, *Mlu* I digested hybrid cell line panel filters comprised of hybrid cell lines and parental cell lines that contain breakpoints within the region 16q21-16q22.1 were successively hybridized to with the two probes. Lymphoblastoid cell line DNAs from two normal individuals were included as positive controls and the cell lines CY5, KS (del) and TH (del) were included as negative controls. The DNA from the hybrid cell line panel produced the same *Mlu* I fragments for D16S4 and D16S91 with the only differences being the variation in cleavage of *Mlu* I restriction sites from one hybrid cell line to another (Table 3.4). The additional analysis of D16S4 and D16S91 to the hybrid cell line panel DNA further reduced the likelihood that the common 450 kb and 990 kb *Mlu* I fragments were a result of coincidental migration. The physical linkage of D16S91 to D16S4 implies the fine localization of D16S91 to the hybrid cell line interval between FRA16B and CY18A. The proposed long range restriction map linking D16S4 and D16S91 at a maximum physical distance of 450 kb is presented in Figure 3.4.
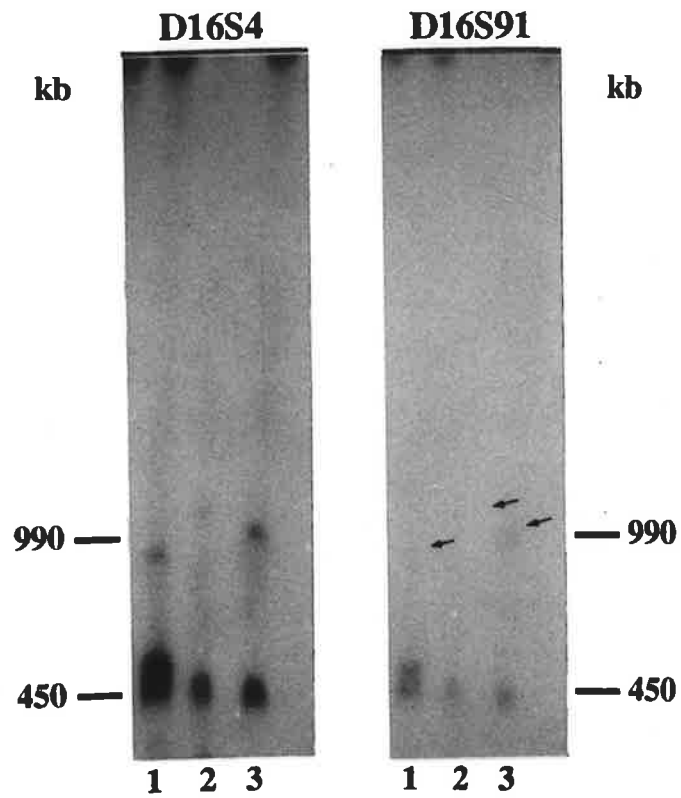
Fig. 3.3. Identification of common 450 kb and 990 kb *Mlu* I fragments between D16S4 and D16S91. The two probes were hybridized to the same low resolution PFGE filter containing *Mlu* I digested DNA samples from three unrelated individuals (1, 2, 3). The faint bands are indicated with arrows.

Table 3.4   Fragment Sizes (kb) of an *Mlu* I Digested Hybrid Cell Line Panel Analysed with D16S4 and D16S91.

DNA SAMPLE	D16S4	D16S91
Normal lymphoblastoid	450	450
Normal lymphoblastoid	450, 990	450, 990
CY 4 (par)	450	450
CY 6 (par)	230, 990	200
CY 18	230	200,990
CY 5	-	-
KS (del.)	-	-
TH (del.)	-	-
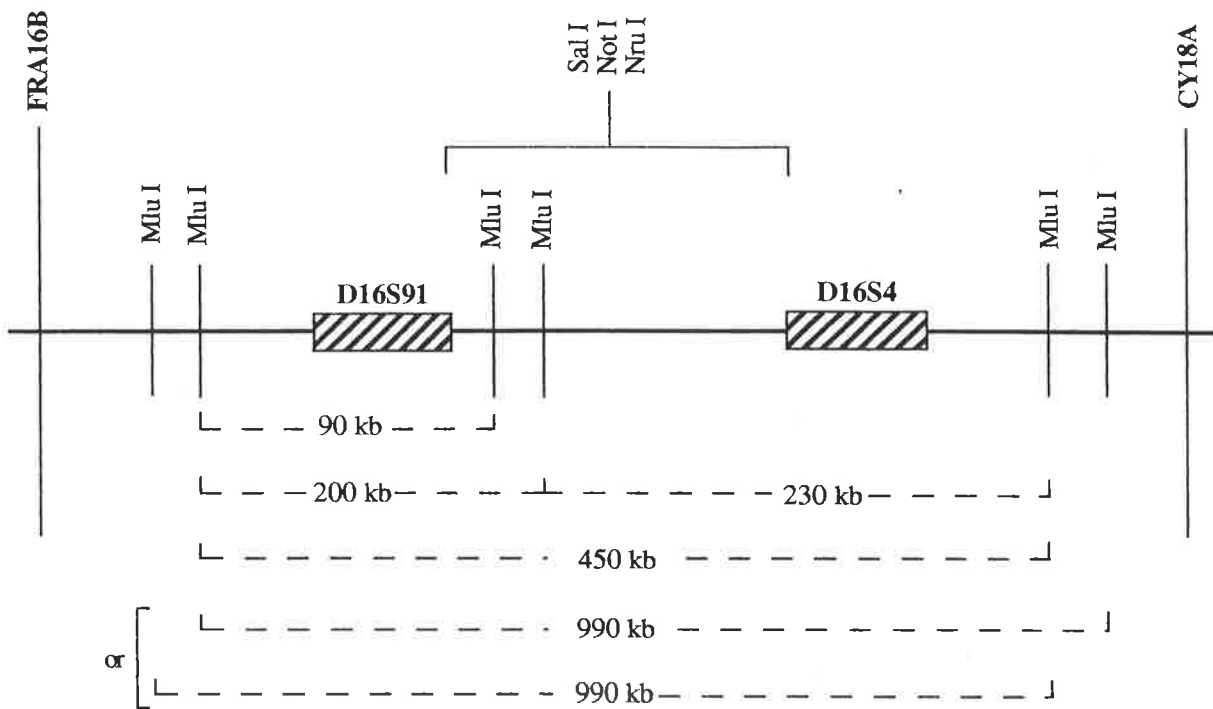
Fig. 3.4. Long range restriction map linking D16S4 and D16S91 at a maximum physical distance of 450 kb. At least one *Sal* I, *Not* I and *Nru* I restriction enzyme site is located between D16S91 and D16S4. The order of loci relative to FRA16B are based on genetic linkage data (Kozman *et al* 1991). The map is not to scale.

3.4.1.2 Linkage between <u>LCAT</u> and <u>ALDOA</u>.

The hybridization of probes for <u>LCAT</u> and <u>ALDOA</u> to the same high resolution filters located both gene probes to a common 70 kb *Mlu* I fragment (Figure 3.5). In addition to the 70 kb *Mlu* I fragment a 270 kb *Mlu* I band was observed for <u>ALDOA</u>. This band was of equal intensity to the 70 kb signal and <u>LCAT</u> did not hybridize to it (Figure 3.5). No hybridization fragments were observed for the hybridizations of either gene probe to low resolution filters. Hybridization of <u>LCAT</u> and <u>ALDOA</u> to *Sal* I, *Not* I and *Nru* I digested DNA samples produced small hybridization signals, none of which were common to both probes (Table 3.3). Therefore, if the two genes are physically linked, at least one restriction enzyme site for each of these enzymes is located between them.

In an attempt to confirm physical linkage between <u>LCAT</u> and <u>ALDOA</u> the restriction enzyme panel was expanded to include *Bss* HII, *Nae* I and *Sac* II. High resolution PFGE filters were prepared using lymphoblastoid cell line DNAs from three unrelated individuals, and undigested cell line DNA was included as a control. The approximated and averaged hybridization fragments (in kb) observed for <u>LCAT</u> and <u>ALDOA</u> probings are listed on Table 3.5.

The hybridization of <u>LCAT</u> and <u>ALDOA</u> to *Sac* II digested DNA samples (Figure 3.6) resulted in common 110 kb and 150 kb fragments between them providing further support for physical linkage. <u>ALDOA</u> hybridized to additional 50 kb, 200 kb and 590 kb *Sac* II fragments that did not cohybridize with <u>LCAT</u>. An additional 340 kb *Sac* II band of hybridization was identified to the <u>LCAT</u> probe which was not in common with <u>ALDOA</u>.

No common hybridization signals were observed between <u>LCAT</u> and <u>ALDOA</u> for the restriction enzymes *Bss H* II and *Nae* I (Table 3.5), suggesting that at

Fig. 3.5. Identification of a common 70 kb *Mlu* I fragment between <u>LCAT</u> and <u>ALDOA</u>. The two probes were hybridized to the same high resolution PFGE filter containing *Mlu* I digested DNA samples from two unrelated individuals (1, 2).

Table 3.5. Fragment Sizes (kb) of <u>LCAT</u> and <u>ALDOA</u> Analyzed by the Extended Restriction Enzyme Panel.

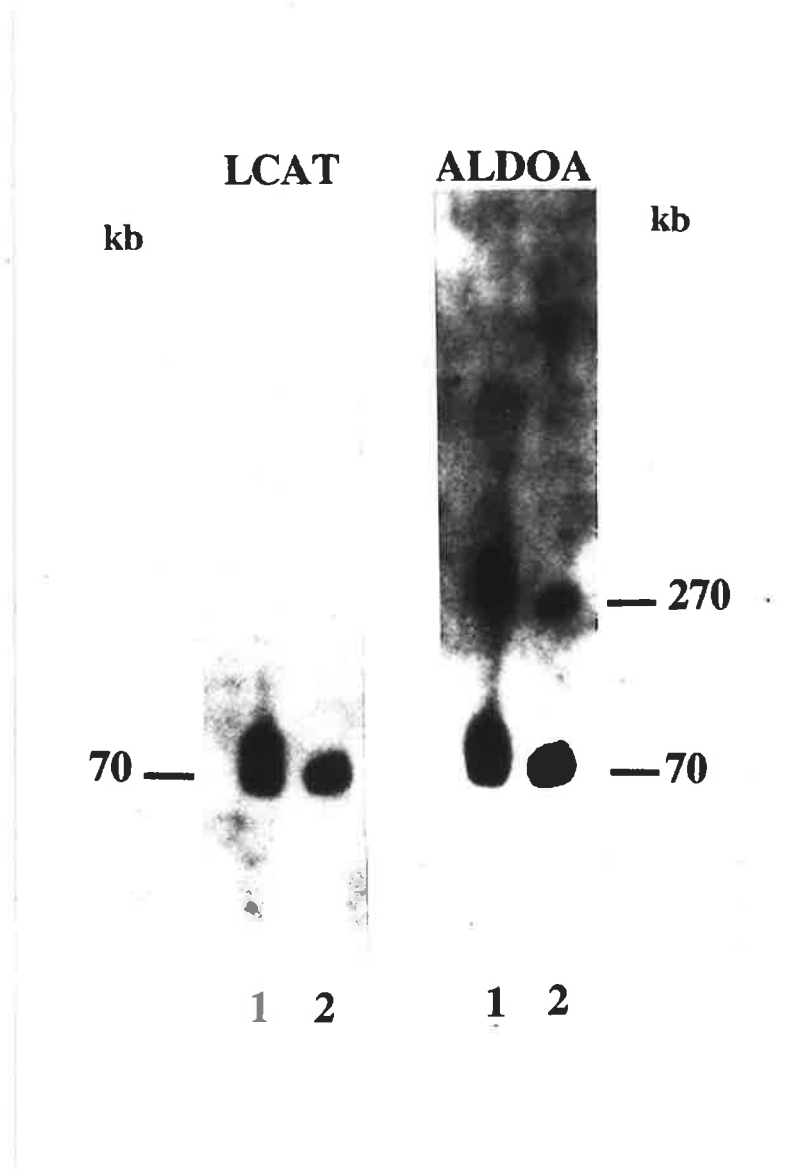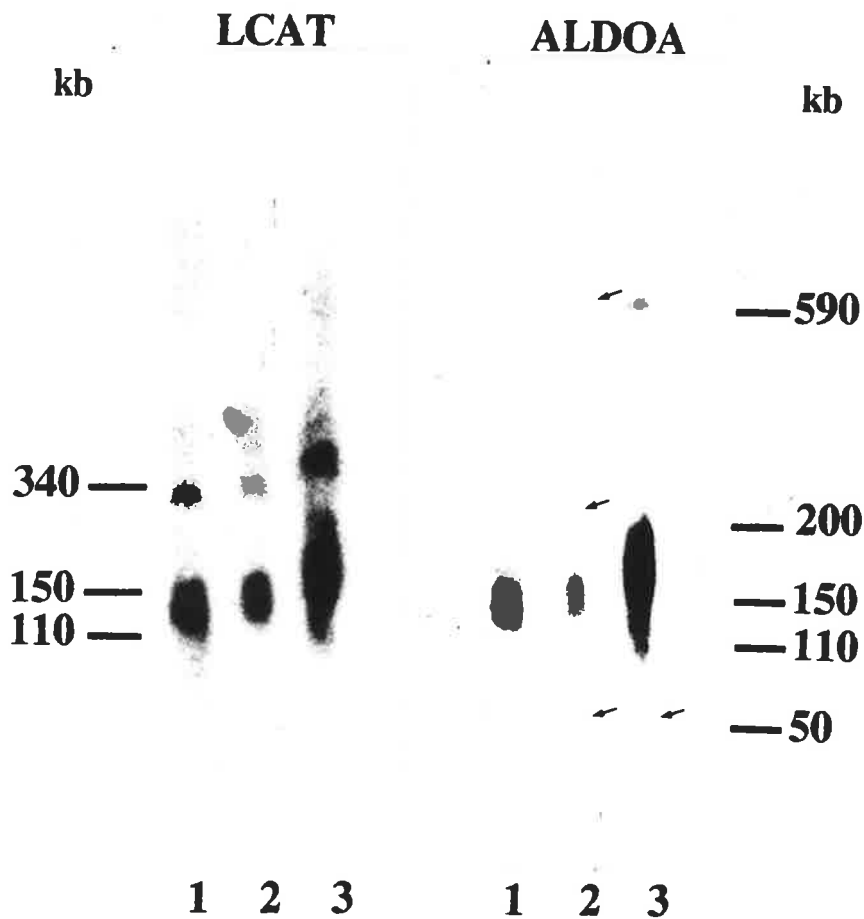| LOCUS	ENZYMES		
	BssH II	*Nae* I	*Sac* II
<u>LCAT</u>	50 120	⟨ 50 150	110 150 340
<u>ALDOA</u>	70 420 530 740	70 180	50 110 150 200 590

Fig. 3.6. Identification of common 110 kb and 150 kb *Sac* II fragments between LCAT and ALDOA. The two probes were hybridized to the same high resolution PFGE filter containing *Sac* II digested DNA samples from three unrelated individuals (1, 2, 3). Faint bands are indicated with arrows.

least one restriction site for these two enzymes is located between the two genes.

Although evidence for physical linkage between LCAT and ALDOA was supported by their hybridization to three common restriction fragments from two restriction enzymes, the presence of additional hybridization fragments presented a number of inconsistencies and ambiquities. If LCAT and ALDOA are physically linked and do not contain *Sac* II or *Mlu* I restriction sites between the two loci, the additional 270 kb *Mlu* I fragment, 200 kb and 590 kb *Sac* II fragments observed for ALDOA could not be a result of partial digestion because the LCAT probe did not hybridize to them. Similarly, the origin of the additional band of hybridization to the LCAT probe of 340 kb in *Sac* II digests could not be a result of partial digestion. Although the *Sac* II digests were clearly incomplete, the ALDOA probe did not hybridize to them. The possibility of the 340 kb *Sac* II fragment resulting from the LCAT probe hybridizing to an LCAT pseudogene could be eliminated because the probe for LCAT hybridized to a unique sequence in the *Mlu* I digest. The fragment could not be the result of cleavage of the LCAT gene by an internal *Sac* II recognition site because a scan of the complete LCAT gene sequence yielded no internal *Sac* II restriction sites.

The restriction map linking the two loci of LCAT and ALDOA to a maximum physical distance of 70 kb is presented in Figure 3.7. Physical linkage between LCAT and ALDOA localizes the ALDOA gene to 16q22.1 and indicates a fine localization of the gene to the same hybrid cell line interval as LCAT, between CY 18A and TH (P).

3.4.2 Analysis of High Molecular Weight Restriction Fragments with Unique Probes from the Region Proximal to FRA16B (16q13-16q21).

The hybridization fragments observed for MT1B, CETP and D16S124 analyzed by high and low resolution PFGE using the restriction enzymes *Sal* I, *Not* I, *Nru* I
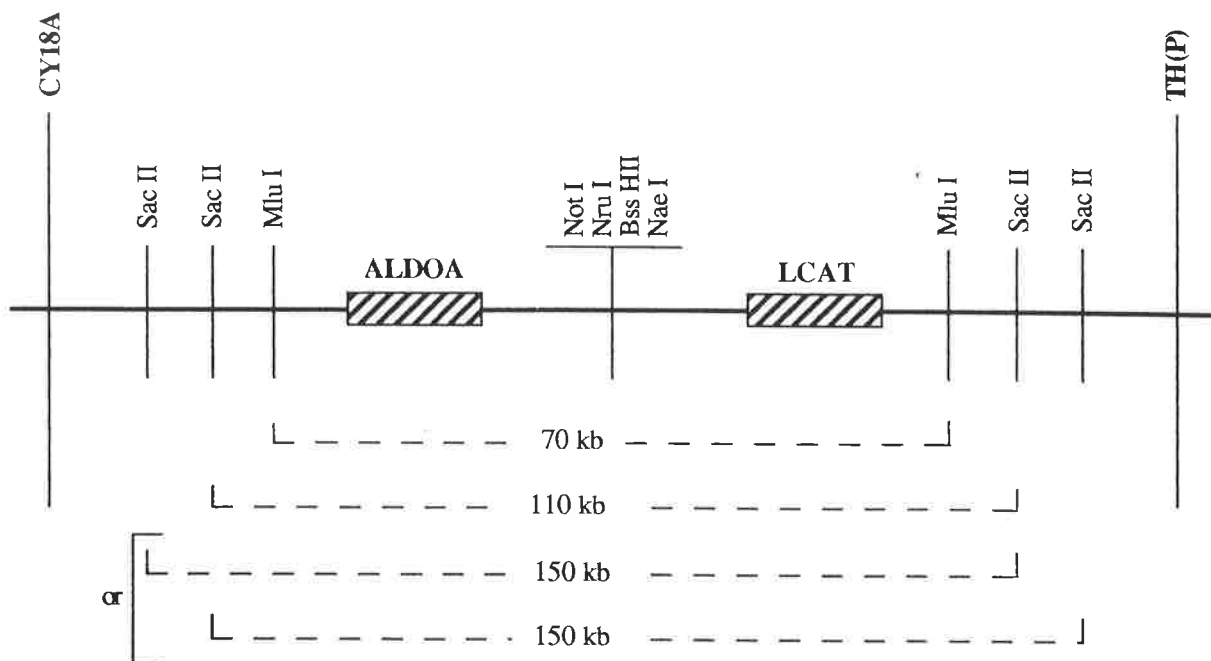
Fig. 3.7. Long range restriction map linking <u>LCAT</u> and <u>ALDOA</u> at a maximum physical distance of 70 kb. At least one *Sal* I, *Not* I, *Nru* I, *Bss* H II and *Nae* I restriction enzyme site is located between <u>LCAT</u> and <u>ALDOA</u>. The relative order of loci with respect to the somatic cell hybrid breakpoints CY 18A and TH(P) are not known. The map is not to scale.

and *Mlu* I are listed on Table 3.6. Lymphoblastoid cell line DNAs from three unrelated individuals were used in the analysis and undigested lymphoblastoid cell line DNA was included as a control. The approximated and averaged sizes in kb are shown for each marker and for four enzymes.

3.4.2.1   Linkage between CETP and MT1B.

The successive hybridization of CETP and MT1B to high and low resolution PFGE filters provided evidence of physical linkage between the two gene probes. The high resolution filter detected a common *Not* I hybridization fragment of 740 kb for the two probes in two of the three cell lines analyzed (Figure 3.8). The probe for CETP also identified a 480 kb *Not* I fragment of hybridization and the probe for MT1B hybridized to an additional 190 kb *Not* I fragment in all three cell lines. The sum of the two hybridization fragments approximately totalled the size of the 740 kb *Not* I fragment common to both probes. It is proposed that CETP and MT1B are physically linked to the 740 kb *Not* I fragment and that a *Not* I restriction enzyme site is located between the two probes.

Hybridization of probes for CETP and MT1B to the low resolution filter detected the common 740 kb *Not* I fragment and also identified an additional 1000 kb *Not* I fragment of hybridization common to both gene probes (Figure 3.9). A 1300 kb *Not* I fragment was also identified for the MT1B probe that did not cohybridize with CETP. This hybridization fragment may be the result of a partial *Not* I digestion having arisen from the cleavage of the proposed *Not* I restriction enzyme site located between the two loci. For both high and low resolution PFGE filters, the remaining restriction enzyme panel did not produce additional evidence for physical linkage between CETP and MT1B (See Table 3.6). This suggests that at least one restriction enzyme site for each of these enzymes is located between the two genes.

Table 3.6. **Fragment Sizes (kb) of Markers Proximal to FRA16B Analysed by PFGE.**

	ENZYMES			
LOCUS	*Sal* I	*Not* I	*Nru* I	*Mlu* I
CETP	210 520 830	480 740 1010	580 770	260 1020
MT1B	290 560 860	190 740 1010 1300	100 230 450 660 840	630 1040 1520
D16S124	180 300 430 530 680	170 250 420 530 680	280 350 420 590 800	80 130 260 590 1040

Fig. 3.8. Identification of a common 740 kb *Not* I fragment between <u>CETP</u> and <u>MT1B</u>. The two probes were hybridized to the same high resolution PFGE filter containing *Not* I digested DNA samples from three unrelated individuals (1, 2, 3).
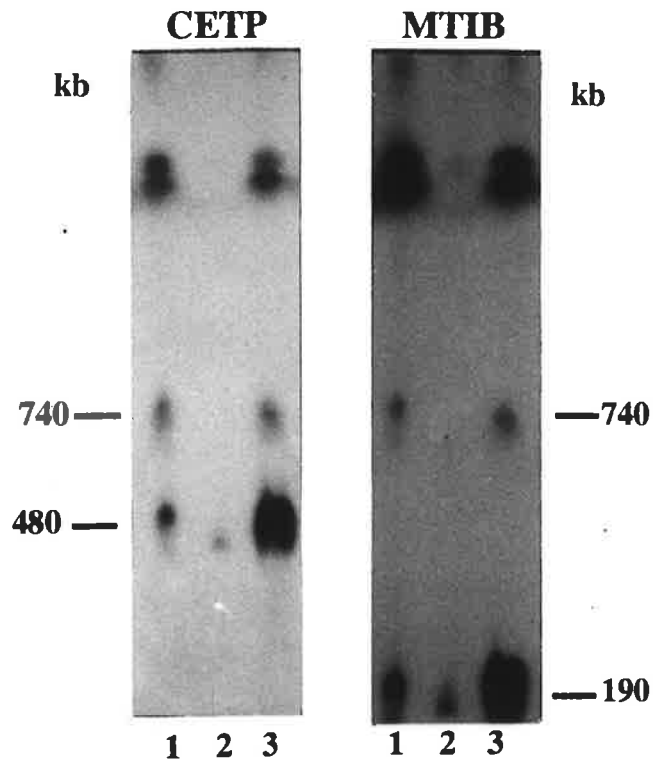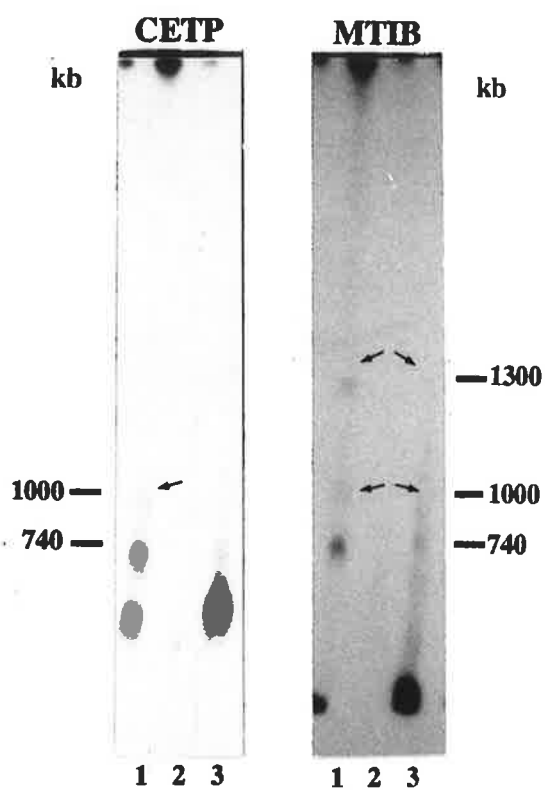
Fig. 3.9. Identification of common 740 kb and 1000 kb *Not* I fragments between CETP and MT1B. The two probes were hybridized to the same low resolution PFGE filter containing *Not* I digested DNA samples from three unrelated individuals (1, 2, 3). Faint bands are indicated with arrows.

In attempts to further verify the 740 kb and 1000 kb common *Not* I fragments of hybridization, a *Not* I digested hybrid cell line panel containing DNA within the region 16q13-16q21 was hybridized with probes for <u>CETP</u> and <u>MT1B</u>. Lymphoblastoid cell line DNA from two unrelated individuals were included as positive controls and mouse A9, CY 5, KS (del) and TH (del) were included as negative controls. The hybridization fragments identified for the two probes are presented in Table 3.7. The lymphoblastoid cell line DNA's exhibited complete digestion, producing the 190 kb *Not* I fragment for <u>MT1B</u> and the 480 kb *Not* I fragment for <u>CETP</u>. The parental cell lines, CY 127 and CY 18 were either completely digested or partially cleaved producing either the 190 kb <u>MT1B</u> or 480 kb <u>CETP</u> individual *Not* I fragments and also the common 740 kb and 1000 kb *Not* I fragments. The 1300 kb *Not* I fragment for <u>MT1B</u> was not observed. No hybridization signals were observed for the negative controls.

The results obtained from the hybrid cell panel complement the initial evidence for the physical linkage between <u>CETP</u> and <u>MT1B</u> reducing the possibility of coincidental hybridization of same sized but different DNA fragments. Based on these results the long range restriction map of the region containing <u>CETP</u> and <u>MT1B</u> is presented in Figure 3.10.

3.4.3 Comparison Between High Molecular Weight Restriction Fragments Hybridizing to Unique Probes Proximal and Distal to <u>FRA16B</u>.

High molecular weight hybridization signals of probes mapping proximal and distal to the <u>FRA16B</u> locus were compared in order to determine whether any two markers that mapped to either side of <u>FRA16B</u> were hybridizing to common restriction fragments, thereby physically spanning the <u>FRA16B</u> locus. Hybridization

Table 3.7   **Fragment Sizes (kb) of an _Mlu_ I Digested Hybrid Cell Line Panel Analyzed with CETP and MT1B.**

DNA SAMPLE	MT1B	CETP
Normal lymphoblastoid	190	480
Normal lymphoblastoid	190	480
CY 4 (par)	190	480
CY 5 (par)	190	480
CY 6 (par)	190, 740, 1000	480, 740, 1000
CY 120 (par)	190, 740, 1000	480, 740, 1000
CY 127	190, 740, 1000	480, 740, 1000
CY 18	190, 740, 1000	480, 740, 1000
Mouse A9	-	-
CY 5	-	-
KS (del)	-	-
TH (del)	-	-

Note:- A modification of the low resolution PFGE parameters were used for the hybrid cell line panel filter. Switching times were: 200 seconds for 12 hours, 150 seconds for 10 hours, 100 seconds for 7 hours.
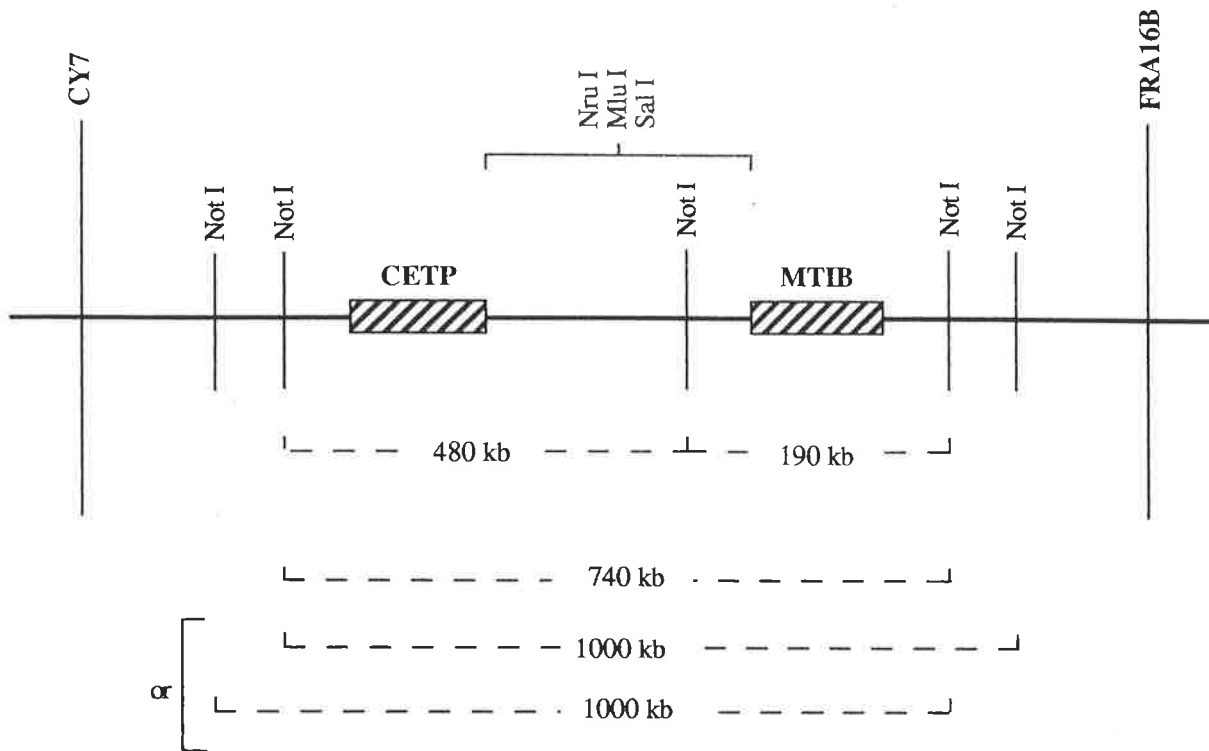
Fig. 3.10. Long range restriction map linking CETP and MT1B at a maximum physical distance of 740 kb. At least one *Nru* I, *Mlu* I and *Sal* I restriction enzyme site is located between the two genes. The relative order of loci between CY 7 and FRA16B are not known. The map is not to scale.

signals of probes for D16S4 and D16S91, the two distal markers mapping to the interval closest to FRA16B (i.e. between FRA16B and CY 18A), were compared with hybridization signals of probes for the three markers MT1B, CETP and D16S124 which were localized to the interval immediately proximal to FRA16B (i.e. between FRA16B and CY 7). A comparison of the hybridization fragments (distal fragments, Table 3.3, proximal fragments, Table 3.6) between proximal and distal probes did not detect any physical linkage. The probe for D16S91 did produce a 730 kb *Not* I fragment which was of similar size to the common 740 kb *Not* I fragment observed for both CETP and MT1B probes. The probe for D16S91 also yielded 180 kb and 680 kb *Sal* I fragments which corresponded to 180 kb and 700 kb *Sal* I fragments observed for D16S124. As all probes had been successively hybridized to the same high and low resolution PFGE filters it was possible to overlay the resulting autoradiographs to determine that these potentially common bands were only of similar size and were not identical fragments of hybridization.

Of the four restriction enzymes used and the PFGE conditions chosen in this study, no two unique probes immediately distal and immediately proximal to the FRA16B locus hybridized to identical high molecular weight restriction fragments. By use of genomic DNA mapping, no physical links were established across FRA16B.

3.4.4 Analysis of High Molecular Weight Genomic DNA with Probes Containing Repeat Sequences.

The hybridization of prereassociated anonymous DNA probes for D16S163 and D16S174 to high and low resolution PFGE filters containing genomic DNA digested with *Sal* I, *Not* I, *Nru* I, *Mlu* I and undigested DNA, did not identify any discrete resolvable DNA fragments. Over 20 hybridizations using a total of six filters were attempted. Non specific hybridization of the probes produced a high level of lane

background in all digested DNA samples indicating that the repetitive DNA present was not being adequately blocked (data not shown).

In attempts to increase the effectivity of prereassociation prior to hybridization of the probes to PFGE filters, various modifications were made to the prereassociation parameters. Probes were prereassociated with 2,000, 4,000, 6,000, 8,000 and 10,000 times excess of sonicated human placental DNA or $C_o t$ I fractionated DNA. Prereassociation incubation periods of ten minutes, thirty minutes, one hour, three hours and overnight were performed. Extending the prereassociation parameters did not significantly reduce non specific hybridization.

Filter washing after probe hybridization was also modified in attempts to reduce lane background. High stringency washes in 0.1XSSPE, 0.1% SDS were increased in 5-10 minute increments from thirty minutes up to several hours or until the probes were completely stripped from the filters. High stringency washes were also performed with the temperature increased to 70°C. Extending the stringencies of washes eventually stripped the probes from the filters without any discrete hybridization fragments being identified.

In attempts to determine that the non specific hybridization was not a result of the DNA samples having been degraded, filters were made using new batches of nylon membrane, restriction enzymes and agarose. Fresh DNA samples and running buffers were made up. Hybridization specificity was not improved suggesting that DNA degradation was probably not the cause.

Under all modifying conditions the hybridization of the probes to PFGE derived Southern blot filters did not produce discrete DNA fragments or significantly reduce lane background.

3.4.5 Construction of Unique Sequence PCR Products from D16S163 and D16S174.

A strategy had been developed whereby unique probes suitable for PFGE analysis could be generated from DNA probes containing repetitive DNA. A model system had been constructed that demonstrated that short lengthed PCR products containing unique sequence DNA produced discrete hybridization signals of the expected size when used as probes to PFGE Southern blot filters of genomic DNA (data not shown). This system was applied to the anonymous DNA markers D16S163 and D16S174.

The consensus sequences resulting from single stranded M13 DNA sequencing of D16S163 subcloned into M13 mp 19 yielded 469 bases of DNA (Figure 3.11). No sequence data was obtained for D16S163 subcloned into M13 mp 18. D16S174 insert DNA subcloned into M13 mp 18 and M13 mp 19 yielded 324 bases of sequence (Figure 3.12a) and 421 bases of DNA (Figure 3.12b) respectively. A comparison of the DNA sequences generated from D16S163 and D16S174 with known mammalian repeat sequence DNA using Genbank exhibited no homology. Internal forward and reverse PCR primers were designed for both D16S163 and D16S174. Their sequences are underlined and listed in Figures 3.11 and 3.12 respectively. PCR amplification of genomic DNA using the D16S163 derived primer pairs produced a single 333 bp product (Figure 3.13, lane 1). For the D16S174 derived primer pairs, a 324 bp PCR product was synthesized (Figure 3.13, lane 2).

3.4.5.1 Analysis of the D16S163 and D16S174 Derived PCR Products Used as Probes to Genomic DNA Southern Blot Filters.

The D16S163 and D16S174 derived PCR products were hybridized to Southern blot filters containing *Eco* RI and *Taq* I digested peripheral lymphocyte DNA from two unrelated individuals and undigested DNA samples were included as controls.

```
5' - C C A G T G T T G A C C C C A A T T A C G G G T G
      T T C A G C C T C G C C A G G C C C C T T T C A T
      T T G T C C T G C T C A C C A G C A A T G T G C T
      G A A A A G A T C C C C A C A C C A G A C C T C A
      C A C T C T G G T C C C T C T G T T C C C G T C G
      G A G T G T C T G A A T G G G T G G A G T G A T C
      A C A G T G T T A A C A C A T T T G C T T T C A G
      T T T C T T T G T C T T C C C T C A G A T G A G G
      C C T T G G T A G A A A A A C A A A C C T A A A
      T C T T T C A G T G A C T A T A C T A G A C T T T
      D A A A C T G T A A G G A T A A T A A A G A C T T
      C A G T G T G T T T T G A T A G A C T G A G A A C
      C A G T G C T A C C A G T T T T C A A A A G C A T
      A T G T G T T G C C G A T G A A T A A T G T A G C
      A G T T T T T C T T C C T C C T T G A T G G A A T
      T A G C T A G G T T A T T T G T C A T T G A A A T
      G T G A G A A T A A A G C A A A T T G C T C T A T
      G G A A G A G T C T T T A C A A T A A T G T C T T
      G G A G A C G G A T A C A C C A T G G T A C - 3'
```

D16S163 forward primer

5' - G T C C T G C T C A C C A G C A A G T G C T G A A - 3'

D16S163 reverse primer

5' - A A C C T A G C T A A T T C C A T C A A G G A G G - 3'

Fig. 3.11. Consensus DNA sequence of 469 bases for D16S163. D16S163 was subcloned into *Eco* RI/*Hind* III cloning sites of M13mp19. The forward and reverse primers were constructed from the DNA sequences underlined.

a) 5'- A C A T T T T G G A C A G G G C C T T C T A G T G
        C C A T T T C T A A A T G G A A C C T A C A A A G
        C A T T T T A T T C C A G A T A T C T G A A G C C
        A A A C T C A G C C T C T C T T C C T C A T T T C
        C C A T C T C C T A G A A A T A A T T C T G A A G
        G C A A G G T T G T G C T T T A A T G T G T A G A
        T T T T A T G T G A G A A A A T A G G T G T C T T
        G A A A A C A A A G A T A T A T A G G C A T A G C
        A T G A G A A A A T C C A G C A T T G C A G A A G
        A G T A C A T G T G A C T G G G A G A T A G G G A
        C A G A G A C T C A A C A C T C T G T T C T T T C
        T G G T G A G A G G A A A G T T A C G C A G A A A
        G A G T C T G G G T T T A G T T A T G G T T T T - 3'

D16S174 forward primer

5'- G G C C T T C T A G T G C C A T T T C T A A A T G - 3'

D16S174 reverse primer

5'- C C A G A C T C T T T C T G C G T A A C T T T C C - 3'

b) 5'- T T A A G A C A C T G C A T T T T T C A A T T C T
        A A A A T T T C A A T T T G G T T C T G T T T T T
        A G A T G G T T T T T T G A G A A C C T C T A T C
        T T T C T A T A C A C T A A T C C T T A T C T C A
        T T G A A T G T G A C T A T A A T A T C T A C G T
        T A A A G T C T T T G T C T G G T A A T T C C A A
        C A T C T A G G T C A T C T C A A G C A T G G A T
        A T T T A T T G G G A T T C T G T T T T T C T G T
        G G T G A G G A T C G G C A C A C T T T T T C T G
        T C C A G G G T C A G A T A G T A A T T G T T T T
        T A G C T G T G T G G G A C A T A T C G T C T C T
        G C G G C A C A T A T C G T C T C T G C G G C A A
        C T T C T C A A C T C T G A C A T T G C A G C A T
        G A A A G C C A T G C A G A C A A C A C A T G A A
        T G A A T G A G T G G T A G T T G T A T T C C G A
        T A A A A C A T T A G A C A T A G A T A C T G A A
        A T T T G A C T G A C A T A A A T A T T C A C G T
        T T T - 3'

Fig. 3.12. Consensus DNA sequence of 324 bases for D16S174 subcloned into *Eco* RI/*Hind* III cloning sites of M13mp18(a) and 421 bases for D16S174 subcloned into *Eco* RI/*Hind* III cloning sites of M13mp19(b). The forward and reverse primers were constructed from the DNA sequences underlined in (a).
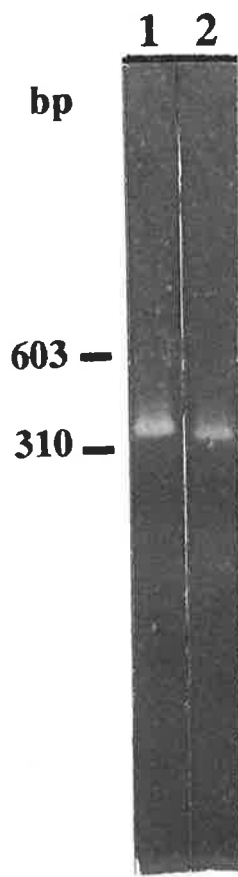
Fig. 3.13. Synthesis of a 333 bp PCR product for D16S163 derived PCR primers (lane 1) and a 324 bp PCR product for D16S174 derived PCR primers (lane 2). Genomic DNA was used as a template. The size of molecular weight markers are shown (in bp) on the left side of the gel.

Discrete hybridization signals were observed for both probes. The D16S163 derived PCR product hybridized to a 9.5 kb *Eco* R1 fragment and a 2.6 kb *Taq* I fragment (Figure 3.14a). The D16S174 derived PCR product hybridized to a 3.5 kb *Eco* RI fragment and a 1.9 kb *Taq* I fragment (Figure 3.14b).

3.4.5.2 Analysis of D16S163 and D16S174 Derived PCR Products Used as Probes to Genomic DNA PFGE Southern Blot Filters.

High resolution PFGE filters were prepared containing lymphoblastoid cell line DNA from three unrelated individuals digested with the restriction enzymes *Not* I, *Nru* I and *Mlu* I and undigested lymphoblastoid control DNA was included. Over sixteen probings of the D16S163 and D16S174 derived PCR products to at least four different PFGE filters resulted in non specific hybridization obscuring any identification of discrete fragments (data not shown).

The attempts to increase probe specificity included subjecting the labelled probes to prereassociation under a range of varying parameters. After probe hybridization the PFGE filters were washed under a range of varying stringencies in attempts to reduce lane background. The D16S163 and D16S174 derived PCR products were subjected to the same series of modified prereassociation parameters and washing stringencies as previously described in section 3.4.4. No modifications resulted in any discrete signals being observed when the PCR products were hybridized to genomic DNA PFGE filters.

3.4.6 Towards a Long Range Restriction Map of 16q13-16q22.1.

Long range restriction mapping of 16q13-16q22.1 with unique probes for five genes and three anonymous DNA markers produced a non-contiguous map encompassing approximately 2.1 Mb of DNA. Distal to FRA16B the long range restriction map linking D16S4 and D16S91 included 990 kb of DNA (Figure 3.4) and the restriction
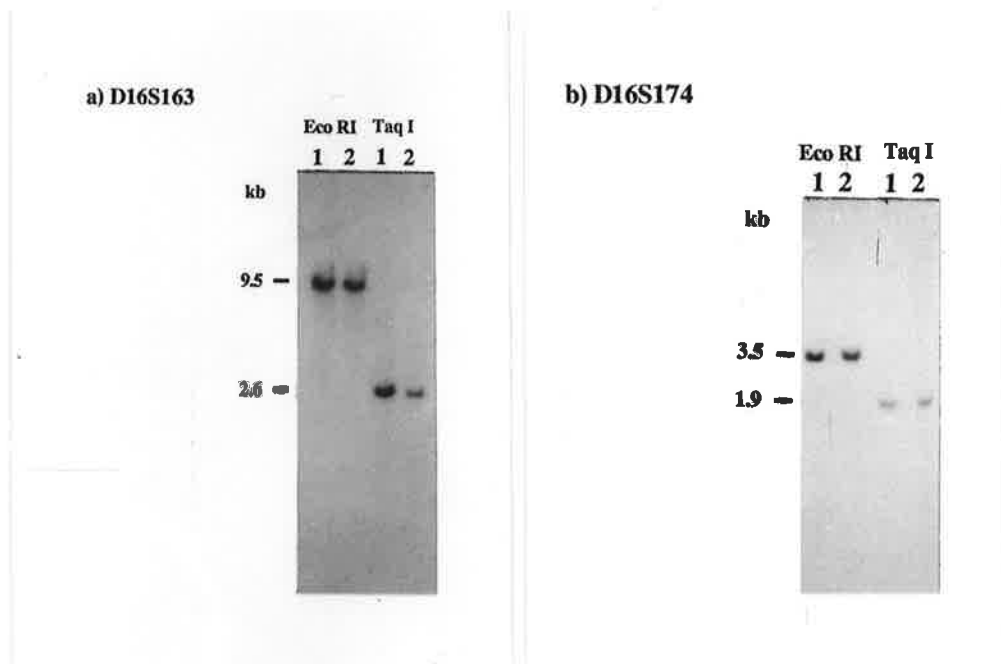
Fig. 3.14. Hybridization of a) D16S163 derived PCR product and b) D16S174 derived PCR product to genomic DNA Southern blot filters from two (1, 2) unrelated individuals digested with *Eco* RI and *Taq* I restriction enzymes.

map linking LCAT and ALDOA encompassed 150 kb of DNA (Figure 3.7).
Proximal to FRA16B the physical linkage between MT and CETP was identified
within a long range restriction map extending 1000 kb of DNA (Figure 3.9). No
physical linkage was observed between probes proximal and distal to FRA16B.

The detailed cytogenetic-based physical map of markers making up the non-
contiguous long range restriction map of 16q13-16q22.1 are summarized on the
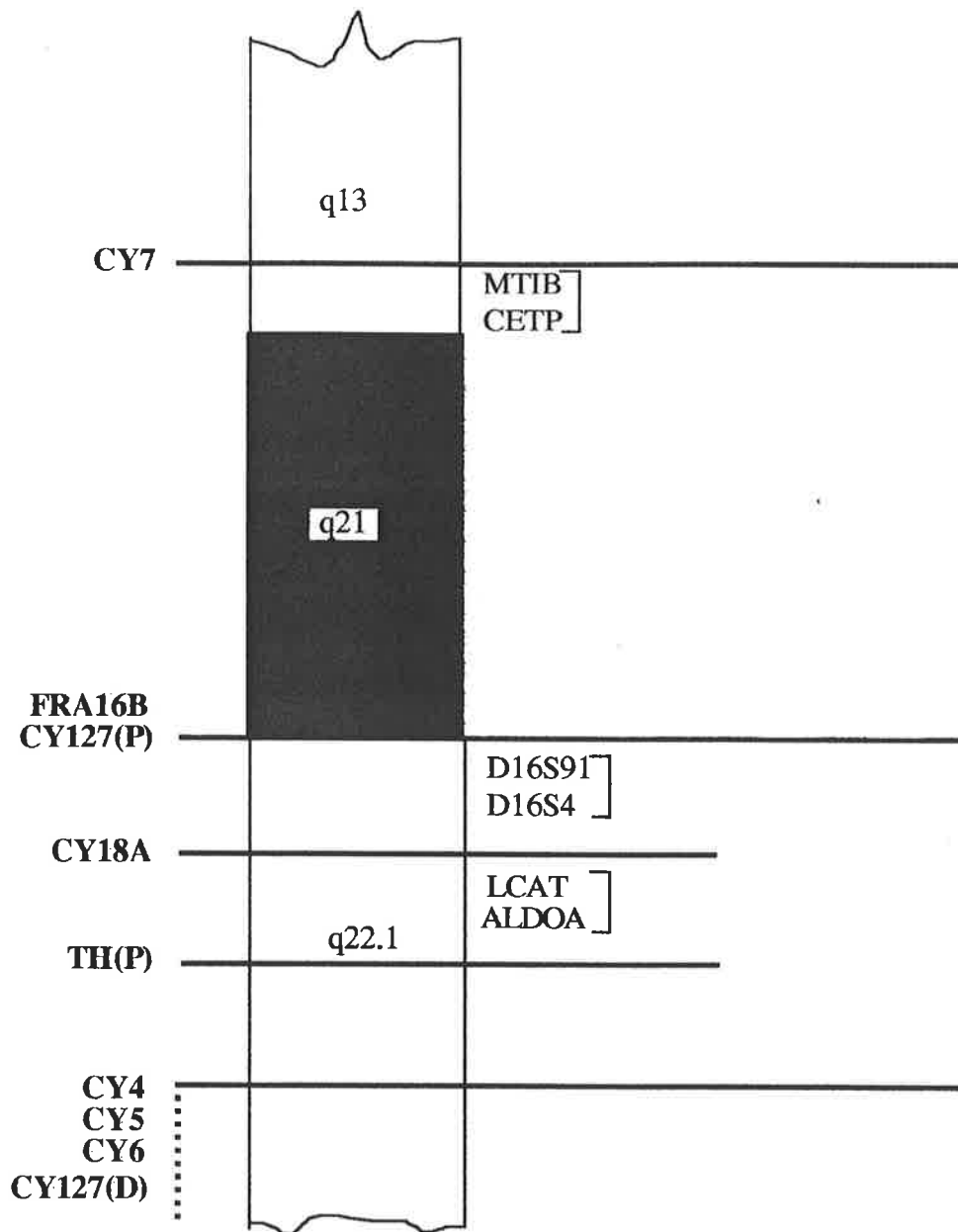physical map of the long arm of chromosome 16 (Figure 3.15).

Fig. 3.15. Physical map of the chromosome region 16q13-16q22.1 exhibiting the regions of physical linkage between markers making up the non contiguous long range restriction map. The horizontal lines indicate the breakpoints of hybrid cell lines (CY) and the fragile site (FRA16B). CY 127 and TH are interstitial deletions. Their proximal breakpoints are located by "P" and their distal breakpoints by "D".

3.5   DISCUSSION

A long range restriction map spanning the chromosome region 16q13-16q22.1 was the initial step towards the localization and characterization of translocation breakpoints and the ultimate goal of cloning the FRA16B locus. Genes and anonymous DNA markers had been localized into sections of the chromosome with respect to cytogenetic breakpoints (See Figure 3.1). Attempts were then made to physically link these markers within sections and across breakpoints as defined by comigrating high molecular weight DNA fragments.

The construction of a long range restriction map of 16q13-16q22.1 was complicated by several factors. In general, the region under investigation was very large, estimated to span approximately 10 Mb of DNA. Its magnitude, coupled with the limitation to using single copy probes in the analysis reduced the total number of probes available to establish physical linkage. More specifically, the construction of a long range map was hindered by the presence of a high frequency of unmethylated CpG dinucleotides in the region of interest. Long range restriction maps around the cystic fibrosis (CF) locus (Fulton *et al* 1989), polycystic kidney disease, PKD1 (Harris *et al* 1990). Huntington disease gene (Pritchard *et al* 1989), alphaglobin region (Fischel-Ghodsian *et al* 1987) and others have also been complicated by these factors.

3.5.1   CpG Islands.

Most of the hybridization fragments generated by the probes in this study were of sizes much smaller than expected. For example, the expected cleavage frequency for *Not* I restriction enzyme is every 800-1200 kb in the human genome (Barlow and Lehrach, 1987, Brown and Bird, 1986). In this study, complete digest hybridization fragments identified with *Not* I approached an average size of only 375 kb. The other

restriction enzymes revealed a similar trend, identifying complete restriction digest hybridization fragments of the following averaged sizes, *Sal* I - 210 kb, *Nru* I - 295 kb, *Mlu* I - 355 kb, *Nae* I - 60 kb, *Bss* H *II* - 60 kb, *Sac* II - 80 kb. These values were calculated by averaging the size of the smallest hybridization fragments generated for each restriction enzyme.

This high frequency of CpG residues in the region (the recognition sites of the restriction enzyme panel have a high CpG content) may be correlated with the apparent concentration of genes in this area. Associated with the 5'-end of genes are HTF (*Hpa* II tiny fragment) or CpG islands (Bird 1986, Brown and Bird, 1986). CpG islands are short, dispersed regions of unmethylated DNA with a high frequency of CpG dinucleotides relative to the bulk genome. In the bulk of the genome, the cytosines in the CpG dinucleotides in mammalian DNA are highly methylated (Bird 1986). It has been estimated that 89% of the *Not* I sites in the human genome are localized in CpG islands (Lindsay and Bird, 1987). At the time of this study up to seven genes had been localized within the boundaries of 16q13-16q22.1 (Reeders and Hildebrand, 1989), five of which were included in the long range restriction mapping. It is proposed that the associated CpG islands in this gene rich region of chromosome 16 complicated the construction of the long range restriction map. The higher than expected frequency of DNA cleavage by the restriction enzyme digest panel reduced the likelihood of adjacent probes hybridizing to common bands. Therefore, probes that may lie in close proximity to one another were not physically linked thereby leading to the generation of a non-contiguous restriction map.

Harris *et al* (1990) observed that restriction enzymes recognising sites consisting of two CpG dinucleotides plus A and T residues (i.e. *Mlu* I and *Nru* I) may produce larger restriction fragments around CpG islands. Such sequences are less likely to be

found in typical CpG rich islands (Lindsay and Bird, 1987) and thus the CpG dinucleotides may be partially methylated so that the site is not completely cleaved. In this study *Mlu* I and *Nru* I digested genomic DNA did not produce larger fragments of hybridization than restriction enzymes with only CpG residues in their recognition sites such as *Bss H* II, *Nae* I, *Not* I and *Sac* II.

3.5.2   Partial Digest Fragments.

Physical links between probes can often be verified by the hybridization of probes to common partial digest fragments. In this study, most probings resulted in a number of partial digest hybridization fragments that were consistently reproduced and provided evidence for physical linkage between markers. These fragments were produced even following extended digestion by the restriction enzyme panel. Termed natural partial digestion products (Rappold and Lehrach, 1986), these fragments are most likely to result from partial methylation of the corresponding restriction site. For probes in the region 16q13-16q22.1, the partial digest products were rarely greater than 1,000 kb in size, once again limiting the likelihood of physical linkage between adjacent probes. In other long range restriction mapping studies such as the mapping of the distal region of the long arm of human chromosome 21 (Burmeister *et al* 1991, Gardiner *et al* 1990), the cystic fibrosis region (Fulton *et al* 1989, Rommens *et al* 1989) and the human multidrug resistance locus, MDR (Meese *et al* 1989), a majority of the physical links between adjacent probes were only established by very high molecular weight partial digest fragments (i.e. 1,000-3,000 kb).

Apart from natural partial digests, other investigators have attempted methods to induce reproducible partial digest for improved long range restriction mapping. Albertsen *et al* (1989) attempted partial cleavage of genomic DNA by using less enzyme than required for complete digestion or limiting concentrations of Mg^{2+}.

Reproducible partial digestion with *Not* I was difficult to obtain by these methods. The enzyme cleaved the DNA faster than it diffused through the agarose matrix in which the genomic DNA was suspended. As a result, the outer molecules were completely digested and the interior DNA was left intact.

Hanish and McClelland (1990) developed a method for partial *Not* I cleavage of DNA embedded in agarose by introducing a methylase and restriction endonuclease with the same site specificity used in a competition reaction. Enough methylase and endonuclease were used to respectively methylate or cleave the entire sample. When used in a pre-determined ratio, the enzymes competed for their recognition sites resulting in a partial digest. The method was applied to the long range restriction mapping of the human <u>ABL</u> locus (Hanish *et al* 1991) and successfully identified hybridization fragments up to 1600 kb in size. It was reported that commercially prepared methylases were generally unsuccessful due to impurities present. I was unable to obtain purified methylase from a non commercial source, so this method was not attempted for the long range restriction mapping of 16q13-16q22.1.

3.5.3 Probes Containing Repetitive DNA.

Limitations were also imposed upon the construction of a long range restriction map of 16q13-16q22.1 due to the small proportion of probes mapping within this region that were suitable for PFGE analysis of genomic DNA.

Probes containing repetitive sequence DNA produced non specific hybridization signals when probed against genomic DNA PFGE filters. The repetitive DNA present in probes can normally be blocked by the method of prereassociation (Sealey *et al* 1985). The technique involves the incubation of the labelled DNA probe with an excess of unlabelled sonicated total human DNA prior to probe hybridization. The total human DNA binds to the repetitive DNA in the probe and upon

hybridization only unique sequence DNA binds to the filter, resulting in discrete hybridization signals. However, even after prereassociation anonymous DNA probes containing repetitive sequence DNA (i.e. D16S163 and D16S174) produced non specific hybridization signals when probed to genomic DNA PFGE Southern blot filters. Extensive modifications to the prereassociation parameters had no significant effect on increasing probe specificity. The small PCR products free of repeat sequence DNA that were constructed from the anonymous DNA probes D16S163 and D16S174 also produced non specific hybridization when used as probes to PFGE Southern blot filters. Further investigation would be required to determine why these probes did not yield discrete signals.

Recently, human chromosome 16-specific low abundance repetitive (CH16LAR) DNA sequences have been identified to chromosome bands p13, p12, p11 and to q22 (Stallings *et al* 1992b). These sequences were termed low abundance repeat sequences because they could not be suppressed by prereassociation with human genomic DNA or C_0t I fractionated DNA. It is possible that anonymous DNA markers containing repetitive DNA that map to 16q22 may contain CH16LAR DNA sequences. The inability of probes for D16S163 and D16S174 to suppress the repetitive DNA by prereassociation supports this proposal.

3.5.4   A Long Range Restriction Map Linking D16S4 and D16S91.

Physical linkage was established between D16S4 and D16S91 at a maximum physical distance of 450 kb and encompassed a long range restriction map spanning 990 kb of DNA (Figure 3.4).

D16S4 was the closest physical marker distal to FRA16B (Callen *et al* 1988) and also the closest genetic marker distal to the fragile site, as determined by linkage analysis (Mulley *et al* 1989). Physical linkage was established between D16S4 and D16S91 by

two common partial *Mlu* I hybridization fragments observed in lymphoblastoid cell line DNA and confirmed using hybrid cell line DNA. No additional enzymes exhibited physical linkage between the two probes. It is proposed that a number of recognition sites for the restriction enzyme panel are located between the markers, probably attributed to the location of at least one CpG island in this area. D16S91 is also known as APOEL1 (Davidson *et al* 1987), a pseudogene related to the gene for apolipoprotein located on chromosome 19. A CpG island may be associated with this gene.

Although the order of D16S4 and D16S91 along chromosome 16 could not be ordered by physical means, a multipoint genetic linkage map of the region 16q12.1-16q22.1 was constructed during the course of this project which provides the most likely order of the two loci with respect to FRA16B. No recombination was observed between D16S4 and D16S91, but the multipoint linkage analysis (for ten loci) placed D16S91 closer to FRA16B than D16S4 (Kozman *et al* 1991).

3.5.5 D16S4 and D16S91 May Span the M4Eo Breakpoint.

Myelomonocytic leukaemia with abnormal eosiniphils (M4Eo) has been characterized by a pericentric inversion of chromosome 16,#inv(16)(p13q22) or rearranged chromosome 16 homologues t(16;16)(p13q22) as described by Le Beau *et al* (1983) and Arthur and Bloomfield (1983). Using *in situ* hybridization, the long arm breakpoint was mapped to the proximal half of the band 16q22 between the locus D16S4 and the haptoglobin, HP gene (Sutherland *et al* 1990). A number of probes were then chosen that mapped between D16S4 and HP. These probes were used for *in situ* hybridization to metaphase spreads from cultured bone marrow cells containing the inversion 16. The probes D16S4 and D16S91 were found to flank the long arm breakpoint of the inv(16) at q22.1 (L Baker, H Eyre, data not published).

The PFGE data imply that the long arm M4Eo breakpoint lies within the 450 kb *Mlu* I fragment common to both probes. The proximal marker was D16S91 as it hybridized to the short arm on inversion 16 metaphase spreads. D16S4 remained in its position on the long arm of chromosome 16 when hybridized to the same metaphase spreads, deeming it the distal marker. These results support the genetic linkage data that the order of the distal loci relative to FRA16B is FRA16B - D16S91 - D16S4.

3.5.6   A Long Range Restriction Map Linking LCAT and ALDOA.

Although ALDOA had only been regionally assigned by *in situ* hybridization to 16q22-16q24 (Kukita *et al* 1987), it was included in PFGE analysis on the chance that it would be at 16q22.1 and physically link with one or more of the genes localized to the same region.

A comparison of hybridization fragments between LCAT, UVO and ALDOA provided evidence for physical linkage between LCAT and ALDOA at a maximum physical distance of 70 kb (Figure 3.7). Although, in total, three identical fragments from two different enzymes supported a physical link between LCAT and ALDOA, further analysis would be required for confirmation. It was not possible to attribute the hybridization fragments from ALDOA to only chromosome 16. Human aldolase gene mapping has shown that the structural gene for ALDOA is located on the long arm of chromosome 16 and that probes to the ALDOA gene also localized respectively on chromosomes three and ten. The related ALDOA sequences observed were attributed to pseudogenes located on these chromosomes (Maire *et al* 1987, Serero *et al* 1988). The hybridization of a probe for ALDOA to additional 200 kb and 590 kb *Sac* II fragments (Figure 3.6) and to a 270 kb *Mlu* I fragment on genomic DNA samples may be attributed to the probe hybridizing to the pseudogenes located

on chromosomes three and ten.

Further investigation using hybrid cell line DNA containing only human chromosome 16 (i.e. CY 18) as the sample DNA may verify which ALDOA hybridization fragments are chromosome 16 specific and may eliminate human pseudogenes as the origin of the additional fragments. Double digests of the enzyme panel and subsequent analysis with probes for LCAT and ALDOA may determine whether the three physical links are the result of chance alone.

The 150 kb long range restriction map linking LCAT and ALDOA at a maximum physical distance of 70 kb finely mapped the gene for ALDOA to 16q22.1. The relative order of the two genes along chromosome 16 were not established.

3.5.7 A Long Range Restriction Map Linking MT1B and CETP.

The region proximal to FRA16B (16q13-16q21) represented approximately two thirds of the total chromosomal DNA under investigation. In that region only probes for MT1B, CETP and D16S124 were suitable for PFGE analysis.

Physical linkage was observed between MT1B and CETP. Complete and partial digests supported the observations as did additional analysis with hybrid cell line DNA. A 1000 kb long range restriction map linking the two genes at a maximum physical distance of 770 kb was constructed (Figure 3.9). No evidence of physical linkage was observed between D16S124 and either of the gene probes.

MT1B was the closest physically mapped marker proximal to FRA16B (Simmers *et al* 1987, Callen *et al* 1988). Early genetic linkage studies (Mulley *et al* 1989) had shown no recombination between MT1B and the closest flanking distal markers to FRA16B.

A recent comprehensive genetic linkage map of the region (Kozman *et al* 1991) combining physical and genetic mapping has localized MT1B near the CY 7

breakpoint in 16q13. Giemsa light regions like 16q13 have previously been thought to represent CG rich regions of DNA (Comings 1978) possibly representing genes and their associated CpG islands.

Prior to this study CETP had not been finely localized. Its physical linkage with MT1B finely localizes CETP to the same giemsa light region near the CY 7 breakpoint on the long arm of chromosome 16 (Figure 3.11).

3.5.8 Long Range Restriction Mapping Across the FRA16B Locus.

A comparison of hybridization fragments from probes proximal and distal to FRA16B did not reveal any physical linkage across the fragile site. The probes for D16S91 and D16S4 have been established as the closest distal markers to FRA16B. MT1B and CETP had been localized to 16q13 near the CY 7 breakpoint so physical linkage between these sets of probes would be unlikely. The fine localization of the proximal probe for D16S124 was not established but this probe did not physically link with any distal marker.

Due to the lack of a contiguous long range restriction map around the FRA16B locus it was not possible to physically determine the closest flanking markers either side of the fragile site or the physical distance between them.

3.5.9 An Alternative Strategy.

The limited success of restriction mapping large regions of the genome using the same methods employed to construct a long range restriction map of 16q13-16q22.1, have led a number of investigators to explore alternative approaches. Some have already been described, such as partial digestions and *in vitro* methylation of restriction sites. In addition, chromosome jumping, linking clones and yeast artificial chromosome (YAC) cloning systems have also been developed.

Chromosome jumping can be used when the size of the region between flanking

markers is large and the region between the markers may be difficult to subclone. Chromosome jumping does not require cloning all the intervening DNA between the nearest DNA markers and only relies on conventional vectors. The technique involves the circularization of very large DNA fragments, followed by cloning of the junction fragments of these circles. This brings together DNA sequences that were originally located a considerable distance apart in the genome (Collins and Weissman, 1984 and independently by Poustka and Lehrach, 1986). This method could be applied to proximal and distal markers to FRA16B with aims to chromosome jump from both ends towards the fragile site. However, it was observed (Nguyen *et al* 1989, Iannuzzi *et al* 1989) that circularization of large DNA fragments is difficult due to instability and shearing of the DNA in solution. Subsequently, the chromosome jumps were obtainable only in approximately 100 kb intervals. The markers flanking FRA16B may be several Mb apart which would require too many chromosome jumps to be considered practical.

Linking clones are genomic clones that contain a rare cutting restriction site. They are useful for long range restriction mapping as they detect two adjacent fragments when genomic DNA is digested with that enzyme (Smith *et al* 1987).

A linking library must be constructed and positive clones isolated and mapped to the region of interest, before PFGE analysis. However, linking clones would still identify CpG islands, therefore in regions rich in CpG islands the same limitations would arise as in standard long range restriction mapping.

The YAC cloning system, described in 1987, by Burke *et al*, is suitable for the cloning of very large DNA fragments up to several hundred kb in size. It has been shown that very large DNA fragments foreign to Sacharomyces cerevisiae can be stably maintained in this host when cloned into a specially constructed vector that

mimics a natural yeast chromosome. Its benefits for long range restriction mapping includes the avoidance of complications associated with performing many small cloning steps and it circumvents the problem of CpG islands as all intervening sequences are cloned. It was proposed that the incorporation of YAC's to the long range restriction mapping of 16q13-16q22.1 may overcome some of the limitations encountered by standard long range restriction mapping. The introduction and development of YAC based long range restriction mapping of 16q13-16q22.1 is presented in Chapter 4.

CHAPTER 4

CHROMOSOME WALKING WITH YACs TOWARDS FRA16B

4.1   SUMMARY

Two loci which mapped proximal and distal to FRA16B were chosen as initiation points for a bidirectional chromosome walk using YACs to construct a YAC contig spanning the fragile site.   One YAC, 275H10, was identified from the proximal marker and a YAC 377C4, was identified from the distal marker.   Their sizes were shown to be 350 kb and 320 kb respectively.   The integrity of the human insert DNA in 377C4 was confirmed but it was established that the insert DNA in 275H10 had been modified.   *In situ* hybridization of total YAC DNA to FRA16B chromosomes indicated that both YACs were physically close to FRA16B.   A comparison of long range restriction maps constructed for each YAC did not identify any overlap between them, indicating that the two YACs were unlikely to form a contig and thus span FRA16B.   Three of the four endclones from these YACs were isolated, partially sequenced and PCR primers were constructed in order to rescreen YAC libraries for overlapping YACs to continue the chromosome walk towards FRA16B.

4.2 INTRODUCTION

In order for a defined region of DNA to be cloned, DNA markers flanking either side of the locus of interest must be bridged by a single cloned DNA sequence or a series of overlapping clones (i.e. contigs) that collectively span the region. Cloning in conventional vectors such as cosmids or lambda provided access to segments of DNA up to a maximum size of 40 kb, with individual contigs of only 100-200 kb often being identified (Wahl *et al* 1987, Page *et al* 1987, Spies *et al* 1989). When the distance between markers was large (i.e. megabase distances) one possible approach for bridging the markers was chromosome walking with YACs. Facilitated by the development of PFGE, the introduction of YAC vectors had allowed the cloning of large regions of DNA thus extending recombinant DNA analysis to megabase sized regions of the genome. The YAC system had a cloning capacity that was 10-20 times greater than that of cosmids or lambda vectors. Furthermore, due to the large size of the cloned DNA inserts, both the number of clones needed to represent a DNA region and the number of walking steps required to link up existing markers was considerably reduced.

In attempts to overcome the limitations of long range restriction mapping of genomic DNA within the region 16q13-16q22.1, two loci which mapped proximal and distal to FRA16B were chosen as starting points for chromosome walking with YACs. It was proposed that bidirectional walking from points proximal and distal to the fragile site would eventually clone the region spanning FRA16B. The strategy chosen to achieve these goals was to:

1.   Screen for YAC clones containing the proximal and distal markers.

2.   Size and establish the integrity of the DNA inserts in the YAC clones.

3.   Construct and compare long range restriction maps of the YAC clones in order to determine whether the YACs overlapped and thus spanned the fragile site.

4.   Isolate endclones from the human DNA inserts in the YACs for chromosome walking until FRA16B was spanned.

By physical mapping studies and *in situ* hybridization to FRA16B chromosomes, a probe for D16S163 was mapped just proximal to FRA16B and a probe for D16S174 was finely mapped just distal (Chen *et al* 1991, E. Baker per. comm.). The *in situ* hybridization analysis established that both probes were located in close proximity to FRA16B. Although the physical distance separating these flanking probes was not known, they were considered to be good starting points for the chromosome walk. In this chapter the isolation and molecular analysis of YACs identified from the initiation of the chromosome walk will be described. This includes their recovery, sizing, establishment of integrity, construction of long range restriction maps and isolation of endclones.

4.3 MATERIALS AND METHODS

4.3.1 Screening for YAC Clones.

PCR based screening of a complete human YAC library, kindly provided by Dr D LePaslier at the Centre d'Etude du Polymorphisme (CEPH), was used to recover YAC clones (see 2.9.4). The construction of PCR primers for the anonymous DNA markers D16S163 and D16S174 have been detailed in Chapter 3.

4.3.2 PCR.

PCRs were performed by methods already described (2.7). For PCR assays using whole yeast colonies as template DNA, single colonies were picked using a sterile toothpick and resuspended in $10\mu l$ of 2 x PCR buffer. $1\mu l$ of the suspension was used as a DNA template in the reaction. The PCR cycling parameters for all DNA templates were: a step cycle with a $94°C$ denaturation for 1 minute, a $55°C$ annealling for 1.5 minutes and a $72°C$ extension for 1.5 minutes, for 25 cycles. Following this, a 10 minute extension at $72°C$ was performed.

4.3.3 DNA Probes.

The probes YAC-L, YAC-R and PCR products derived from D16S163 and D16S174 primers were used to construct the restriction maps of the recovered YAC clones. The origins of YAC-L and YAC-R probes were described in detail in Chapter 2 (see 2.9.5). The PCR products were synthesized using the PCR cycling conditions described above (4.3.2). The products were resolved by gel electrophoresis and recovered by the methods already described (2.5). The probes were purified by the "Prep-a-gene" method (2.3.1.8.1).

All probes were radiolabelled using the primer extension method (2.5.4.2).

4.3.4 Restriction Enzyme Digestion.

To generate the YAC restriction maps, complete digestion of agarose beads containing yeast DNA was performed with the restriction enzymes *Bss H*II, *Eag* I, *Nae* I, *Nar* I, *Not* I, *Nru* I, *Rsr* II, *Sac* II, *Sal* I, *Sfi* I and *Sna* BI according to the manufacturers instructions. The preparation of yeast DNA embedded in agarose beads has been described (2.6.1.2).

In order to establish the integrity of YAC clones, agarose embedded DNA from YACs and lymphoblastoid cell line DNA constructed from normal individuals were digested with *Eco* R1. The restriction fragments were separated by gel electrophoresis and transferred to nylon membranes by Southern blot transfer (2.5). The existence of restriction fragments of common size in both YAC and lymphoblastoid cell line DNA was taken as a confirmation of DNA integrity in the human YAC insert.

4.3.5 PFGE.

In this section of the project PFGE was performed using the CHEF Mapper™ (Biorad) by the methods already described (2.6). For the PFGE separation of undigested YAC DNA samples, DNA in a size range from 100-800 kb was resolved. For the PFGE separation of YAC DNA digested with the restriction enzyme panel, DNA in a size range from 10-400 kb was resolved. The switching parameters used were those discussed in the CHEF Mapper™ Handbook (Biorad).

4.3.6 End Rescue and Subcloning.

To clone the ends of human inserts in YACs for chromosome walking, vector/*alu* PCR and capture PCR methods were used (see 2.9.7). PCR products were subcloned into the plasmid vector PCR™ 1000 by the methods described (2.7.3).

4.3.7 Automated DNA Sequencing and PCR Primer Construction.

All automated DNA sequencing was performed by Julie Nancarrow. An Applied Biosystems 373A DNA Sequencer was the apparatus used. Double stranded DNA templates were sequenced using the Taq Dye Deoxy™ Terminator Cycle Sequencing Kit (Applied Biosystems) using the procedure described in the Applied Biosystems Cycle Sequencing Handbook.

The design, construction and purification of PCR primers has been described (2.7.1).

4.4 RESULTS

4.4.1 PCR Based Screening of the CEPH YAC Library with Primers for D16S163 and D16S174.

Of the 113 multifilter pools representing the CEPH YAC library, three faint signals from pools 56, 69 and 101 produced the expected 333 bp PCR product from D16S163 derived primers. Three faint signals from multifilter pools 16, 46 and 95 produced the expected 294 bp PCR product from D16S174 derived primers. Hybridization of radiolabelled D16S163 and D16S174 derived PCR products to the appropriate Southern blot filters containing the positive multifilter signals confirmed that multifilter pool number 56 and 69 but not 101 were positive for D16S163 (Figure 4.1). The PCR products from all three multifilter pools 16, 46 and 95 hybridized to the D16S174 derived PCR product (Figure 4.2) confirming their integrity.

The second round of PCR screening performed at CEPH, Paris using single filter pools and subsequent colony hybridizations identified a total of three YACs. Two YACs, 275H10 and 276C11 screened positive for the D16S163 derived PCR primers and one YAC, 377C4, screened positive for D16S174 derived primers.

4.4.2 Confirmation of Positive YAC Clones by PCR Analysis.

Single YAC colonies isolated from 275H10, 276C11 and 377C4 were used directly as templates in PCR assays with the appropriate D16S163 and D16S174 derived PCR primers. A PCR product of the expected 333 bp size was produced when D16S163 derived PCR primers and a 275H10 YAC colony were combined in a PCR assay (Figure 4.3, lane 1). No PCR products were observed when a 276C11 YAC colony was used as a template (Figure 4.3, lane 2) suggesting that 276C11 was a false positive. A PCR product of the expected 294 bp size was observed when D16S174 derived PCR primers and a 377C4 YAC colony were used in a PCR assay (Figure 4.3,
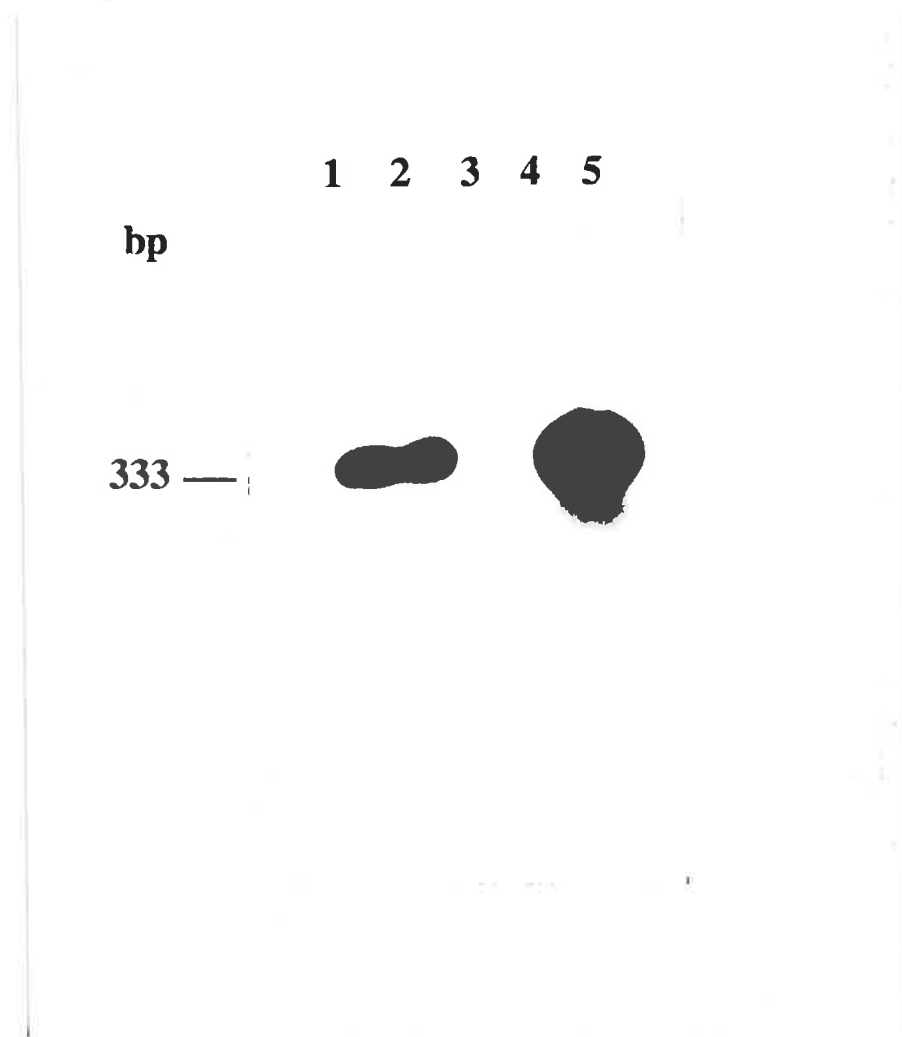
Fig. 4.1 Confirmation of positive multifilter CEPH YAC pools from D16S163 derived PCR primers. A radiolabelled D16S163 derived PCR product was hybridized to a Southern blot filter containing PCR products resulting from assays using the D16S163 derived PCR primers combined with multifilter pool numbers 56 (lane 1), 69 (lane 2), 101 (lane 3), no DNA template (lane 4) and total human DNA (lane 5).
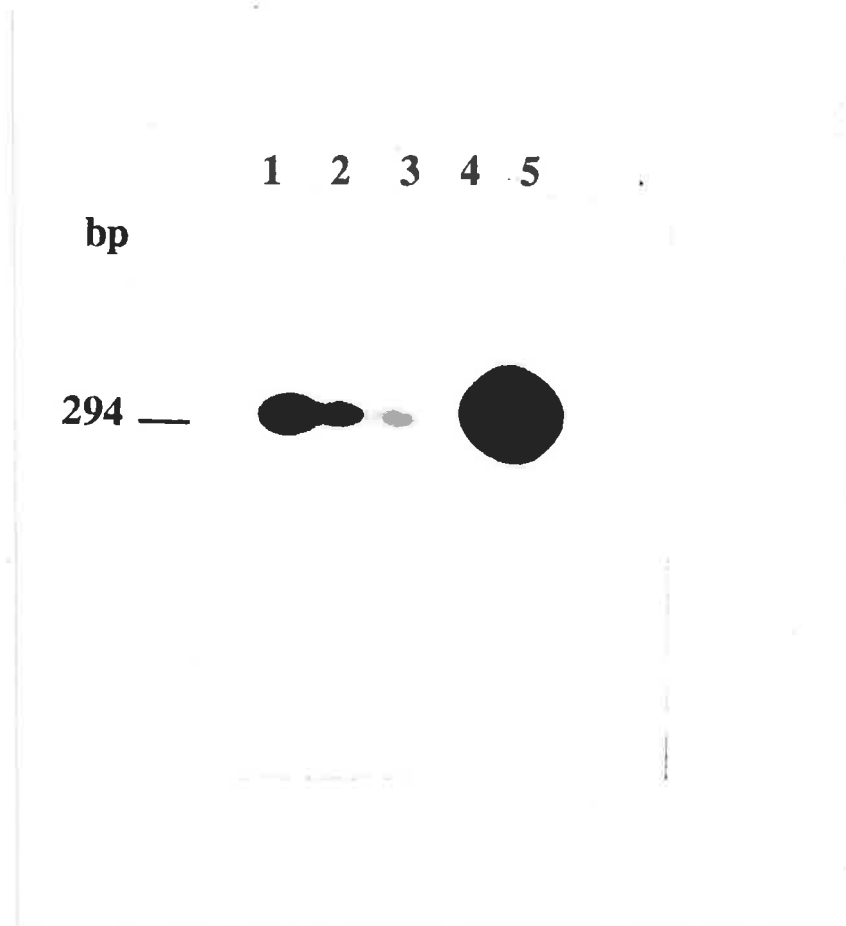
Fig. 4.2 Confirmation of positive multifilter CEPH YAC pools from D16S174 derived PCR primers. A radiolabelled D16S174 derived PCR product was hybridized to a Southern blot filter containing PCR products resulting from assays using the D16S163 derived PCR primers combined with multifilter pool numbers 16 (lane 1), 46 (lane 2), 95 (lane 3), no DNA template (lane 4) and total human DNA (lane 5).
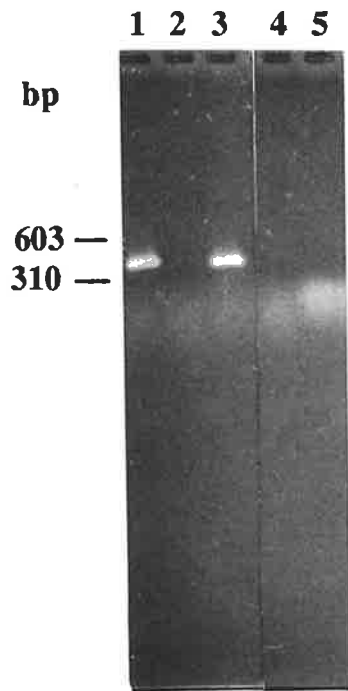
Fig. 4.3 Confirmation of positive YAC clones by PCR analysis. Single YAC colonies from 275H10 (lane 1) and 276C11 (lane 2) were used as templates for PCR assays with D16S163 derived PCR primers. A single YAC colony from 377C4 (lane 3) was used as a template for a PCR assay with D16S174 derived PCR primers. A YAC colony from 275H10 was used as a template with primers derived from D16S174 (lane 4) and a YAC colony from 377C4 was used as a template with primers derived from D16S163 (lane 5). Molecular weight markers are shown (in bp) on the left side of the gel.

lane 3).

In attempts to confirm these observations, the PCR assays were repeated using purified YAC DNA in place of single YAC colonies. Identical results were obtained (data not shown). Therefore of the PCR based CEPH YAC library screening, one positive YAC for D16S163 derived PCR primers and one positive YAC for D16S174 derived PCR primers were identified.

In addition, template DNA from 377C4 was combined with the PCR primers derived from D16S163 (Figure 4.3, lane 4), and DNA from 275H10 was combined with PCR primers derived from D16S174 (Figure 4.3, lane 5) in a PCR assay in an attempt to determine by PCR whether either of the YACs contained flanking sequence to FRA16B and thus spanned the fragile site. Positive controls using total human DNA as template and no DNA template negative controls were included. The results of this PCR assay were negative.

4.4.3 Sizing of YACs.

PFGE Southern blot filters containing preparations of undigested YAC DNA beads from 275H10, 377C4 and an unrelated YAC as a control were successively hybridized with radiolabelled PCR products derived from D16S163 and D16S174 PCR primers. 275H10 was identified as a single YAC approximately 350 kb in size (Figure 4.4A) and 377C4 was identified as being approximately 320 kb in size (Figure 4.4B).

To ensure that the yeast bead DNA preparations contained only one YAC per preparation, the PFGE filters were probed with radiolabelled total human DNA

(Figure 4.4C). Only the 350 kb and 320 kb YACs were observed for preparations of 275H10 and 377C4 respectively. No additional YACs were detected confirming that the only human insert DNA present in the yeast beads were contained in the appropriate YACs.
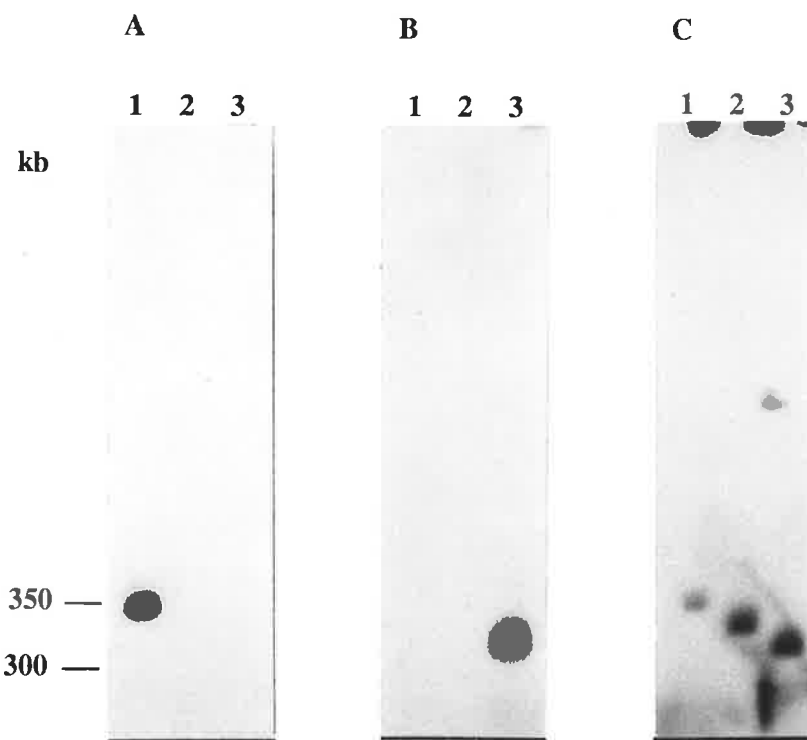
Fig. 4.4 Sizing of positive YAC clones 275H10 and 377C4. PFGE Southern blot filters containing uncut YAC DNA preparations of 275H10 (lane 1), unrelated YAC control (lane 2) and 377C4 (lane 3) were probed with radiolabelled A) a PCR product derived from D16S163, B) PCR product derived from D16S174 and C) total human DNA. Molecular weight markers (in kb) are shown on the left side of the gel.

4.4.4 YAC DNA Integrity.

A Southern blot filter containing *Eco* RI restriction enzyme digestions of YACs 275H10 and 377C4 and *Eco* RI digests of genomic DNA samples from five unrelated individuals, was sequentially probed with radiolabelled plasmids containing probes for D16S163 and D16S174. The probe for D16S174 hybridized to a single 3 kb *Eco* RI fragment in 377C4 and genomic DNA samples (data not shown), supporting the integrity of the human DNA insert in this YAC. The probe for D16S163 hybridized to a single 9 kb *Eco* RI fragment in genomic DNA samples but not in 275H10 DNA (data not shown). The probe hybridized to a single 20 kb *Eco* RI fragment in the YAC, indicating that the human DNA insert of 377C4 had been modified.

4.4.5 YAC Localizations using *in situ* Hybridization (performed by E. Baker).

12 cells were scored for the *in situ* hybridization of total 275H10 DNA to metaphase spreads of chromosomes that expressed FRA16B. 10 proximal, 2 central and no distal signals to FRA16B were recorded, which localized 275H10 close but proximal to FRA16B (ideogram, Figure 4.5A).

28 metaphase spreads were scored for *in situ* hybridization of total 377C4 DNA to FRA16B chromosomes. 18 proximal, 6 central and 4 distal signals were observed (ideogram, Figure 4.5B). The *in situ* hybridization results indicated that although 377C4 is localized distal to FRA16B, it may extend into or just across the fragile site.

4.4.6 Long Range Restriction Mapping of 275H10 and 377C4.

PFGE derived long range restriction maps were constructed for 275H10 and 377C4.

4.4.6.1 275H10.

The hybridization fragments detected with YAC-L and YAC-R endprobes are shown in Figure 4.6A and Figure 4.6B respectively. The hybridization fragments detected with the 275H10 derived PCR product are illustrated in Figure 4.7. The sizes of the
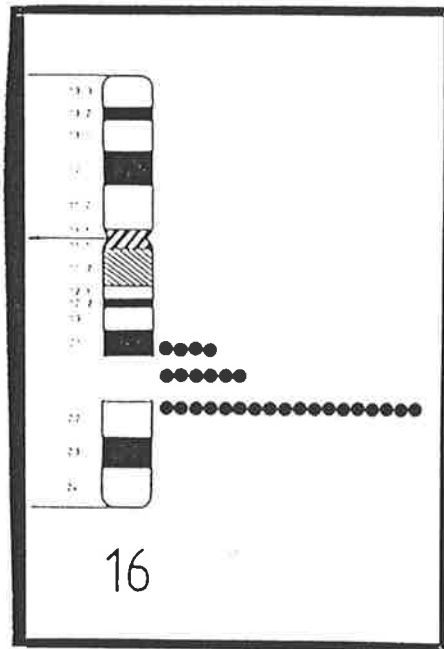
A. D16S163
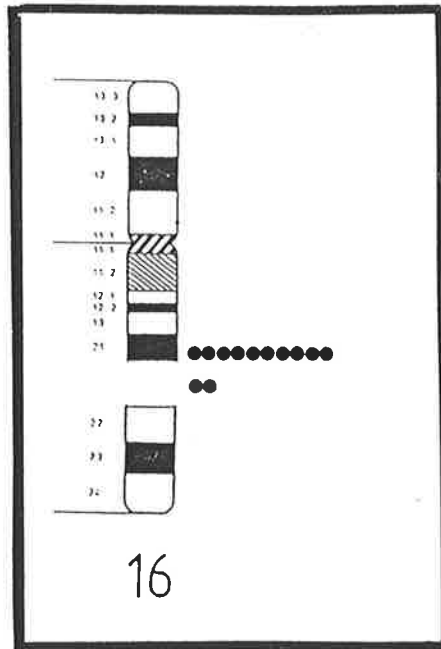
B. D16S174

Fig. 4.5 Ideogram of chromosome 16 depicting the observed distribution of signals proximal, central and distal to <u>FRA16B</u> from *in situ* hybridizations of A) D16S163 and B) D16S174 to chromosomes expressing the fragile site.
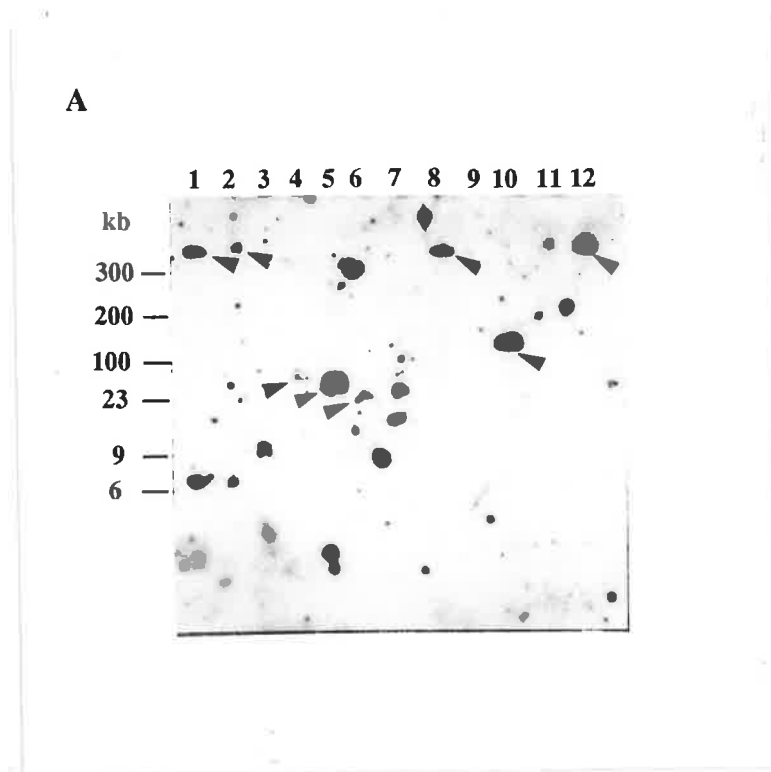
A

Fig. 4.6A) Endlabel mapping of 275H10 with different restriction enzymes. A single PFGE Southern blot filter containing 275H10 DNA digested with *Nru* I (lane 1), *Not* I (lane 2), *Nae* I (lane 3), *Nar* I (lane 4), *Sal* I (lane 5), *Sfi* I (lane 6), *Sna* BI (lane 7), *Sac* II (lane 8), *Rsr* II (lane 9) *Eag* I (lane 10), *Bss* H II (lane 11) and uncut 275H10 (lane 12) were sequentially hybridized with radiolabelled YAC-L (A, this figure) and YAC-R (B, over page). The sizes (in kb) of DNA molecular weight markers are indicated on the left side of the gel. The hybridization signals are indicated with arrows.

B

Fig. 4.6B) Endlabel mapping of 275H10 with different restriction enzymes. A single PFGE Southern blot filter containing 275H10 DNA digested with *Nru* I (lane 1), *Not* I (lane 2), *Nae* I (lane 3), *Nar* I (lane 4), *Sal* I (lane 5), *Sfi* I (lane 6), *Sna* BI (lane 7), *Sac* II (lane 8), *Rsr* II (lane 9), *Eag* I (lane 10), *Bss* HII (lane 11) and uncut 275H10 (lane 12), were sequentially hybridized with YAC-L (previous page) and YAC-R (B this figure). The sizes (in kb) of DNA molecular weight markers are indicated on the left side of the gel. The hybridization signals are indicated with arrows.
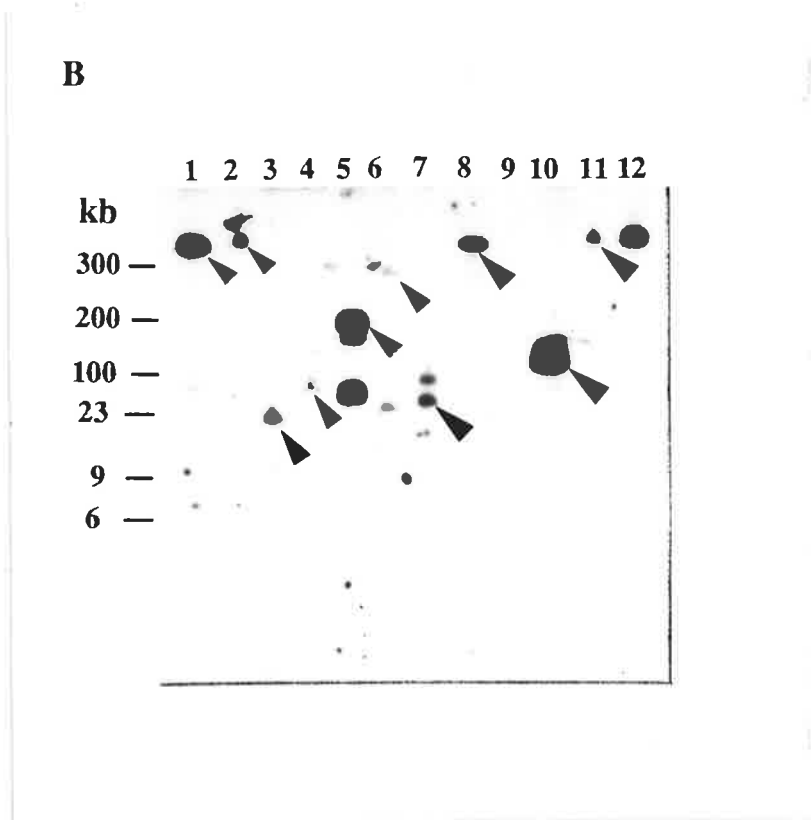
Fig. 4.7 Restriction enzyme digest fragments of 275H10 detected by a D16S163 derived PCR product. A PFGE Southern blot filter of 275H10 DNA digested with *Sfi* I (lane 1), *Not* I (lane 2), *Nae* I (lane 3), *Nar* I (lane 4), *Sal* I (lane 5), *Nru* I (lane 6), *Sna* BI (lane 7), *Sac* II (lane 8), *Rsr* II (lane 9), *Eag* I (lane 10) and *Bss* HII (lane 11) was hybridized with a radiolabelled D16S163 derived PCR product. The sizes (in kb) of DNA molecular weight markers are indicated on the left side of the gel. Faint bands are indicated with arrows.

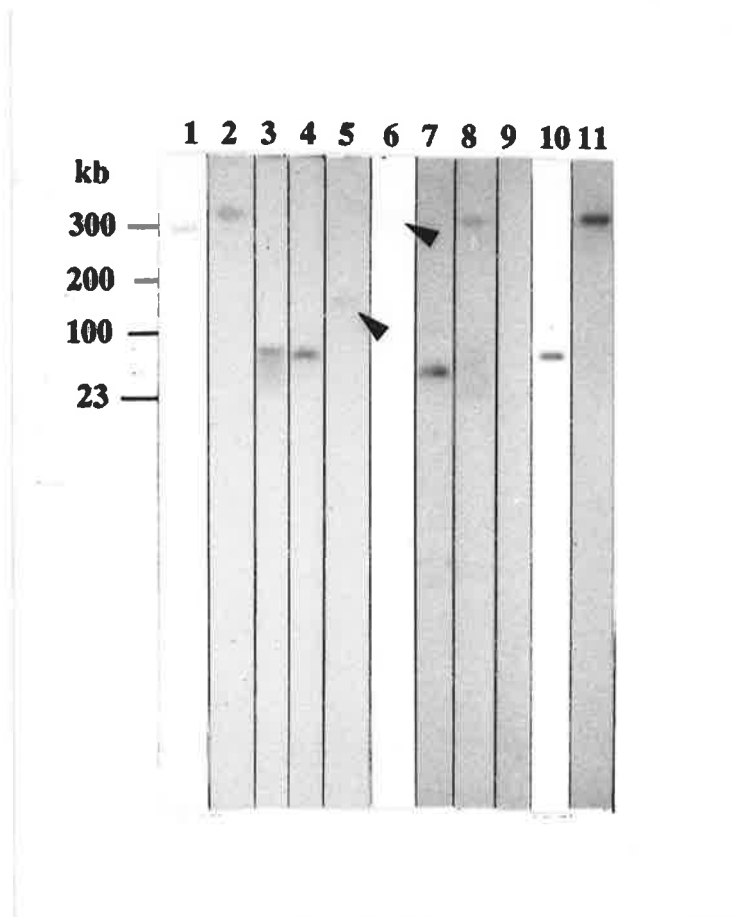hybridization fragments detected by these three probes are listed on Table 4.1.

With complete enzyme digestion, YAC-L detected a 30 kb *Sfi* I fragment (Figure 4.6A, lane 6) and YAC-R detected a 280 kb *Sfi* I fragment (Figure 4.6B, lane 6). Allowing for up to a 20% error in the determination of band sizes, these two fragments totalled the size of 275H10 indicating that the human DNA insert contained only one *Sfi* I site located 30 kb in from YAC-L. Of the restriction enzyme panel used, *Sfi* I was the only enzyme with a single recognition site in 275H10.

Hybridization with the YAC-L and YAC-R endprobes to *Eag* I, *Sal* I and *Sna* BI digested 275H10 DNA resulted in hybridization fragments that did not total the size of 275H10 (Figure 4.6A,B, Table 4.1) indicating that more than one recognition site for these enzymes exist in the YAC. *Nar* I and *Nae* I digested DNA samples yielded uninterpretable results when probed with YAC-L and YAC-R, therefore the number of recognition sites in 275H10 for those two enzymes could not be determined. *Not* I, *Nru* I, *Sac* II and *Bss* H II did not cut 275H10 (Figure 4.6A,B, Table 4.1). The only internal marker that was mapped to 275H10 was the PCR product derived from the D16S163 PCR primers. The PCR product hybridized to the same sized 175 kb *Sal* I fragment (Figure 4.7, lane 5), 280 kb *Sfi* I fragment (Figure 4.7, lane 1) as the YAC-R probe. The PCR product did not detect the 20 kb *Nae* I restriction fragment detected by YAC-R. The probe hybridized to a 75 kb *Nae* I fragment (Figure 4.7, lane 3) that was not detected by either endprobe therefore representing an internal DNA fragment. Based on these observations, the D16S163 derived PCR product was localized to the region of DNA between the *Sna* BI and *Nae* I restriction enzyme sites located near the right arm vector end of 275H10. The long range restriction map of 275H10, including the localization of the D16S163 derived PCR product is shown on Figure 4.8.

Table 4.1. Restriction Enzyme Digest Hybridization Fragments (in kb) Detected for 275H10.

ENZYME	PROBE		
	YAC-L	YAC-R	D16S163 PCR
Not I	350	350	350
Nru I	350	350	350
Sal I	40	175	175
Nae I	N.R.	20	75
Nar I	40	N.R.	75
Sfi I	30	280	280
Sac II	350	350	350
Rsr II	N.R.	N.R.	N.R.
Eag I	125	80	80
Bss HII	N.R.	350	350
Sna BI	25	60	60

Note: N.R. indicates that no hybridization was observed.

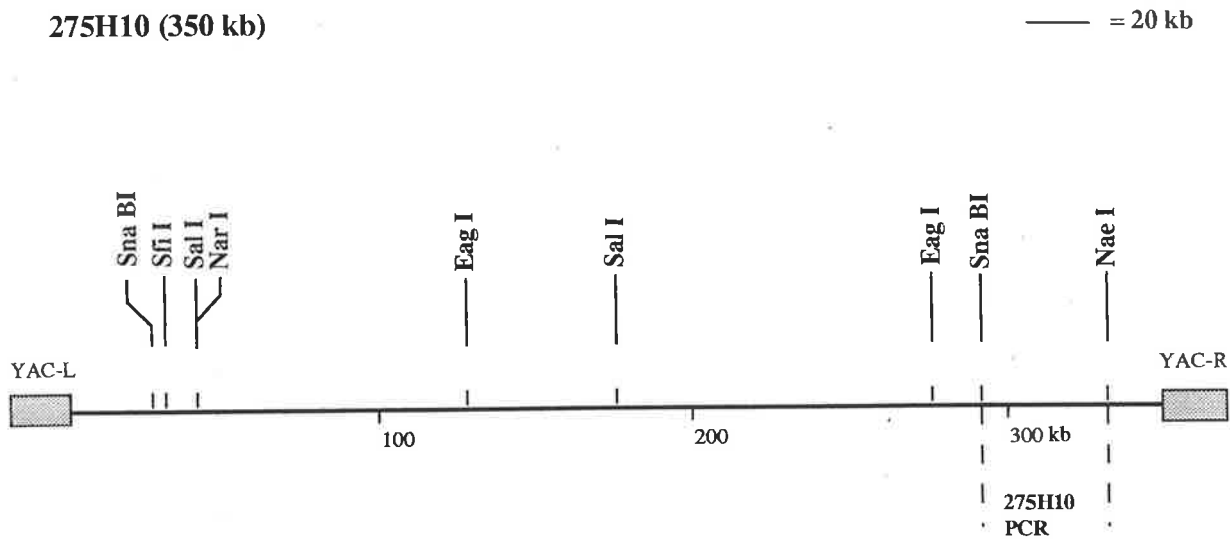**275H10 (350 kb)** ──── = 20 kb

Fig. 4.8 Long range restriction map of YAC 275H10 (A), showing the location of the D16S163 derived PCR product (B). Complete map restriction enzyme sites in human insert DNA are shown only for *Sfi* I. Other enzymes may have one or more additional restriction enzyme sites.

4.4.6.2 377C4.

The hybridization fragments detected by YAC-L and YAC-R endprobes are shown in

Figure 4.9A and Figure 4.9B respectively. The fragments of hybridization detected

with the D16S174 derived PCR product are illustrated in Figure 4.10. The sizes of the

restriction digest fragments detected by these probes are listed on Table 4.2.

With complete enzyme digestion of 377C4 DNA, YAC-L detected an 80 kb *Sal* I

fragment (Figure 4.9A, lane 5) and YAC-R detected a 220 kb *Sal* I fragment (Figure

4.9B, lane 5). The two fragments totalled the size of 377C4 indicating that only one

Sal I recognition site is present in the 377C4 human DNA insert, located 80 kb in

from the YAC-L vector arm. Hybridization of YAC-L and YAC-R endprobes to *Nru* I,

Nae I, *Nar* I, *Sfi* I, *Sac* II, *Eag* I, *Bss* HII and *Sna* BI digested 377C4 detected signals

that did not total 320 kb (Figure 4.9A,B, Table 4.2).   More than one recognition site

must be present for these enzymes in 377C4 human DNA insert. *Not* I did not cut

377C4.

The D16S174 derived PCR product was used as a probe to determine its location

within the 377C4 restriction map. The PCR product hybridized to the same sized

130 kb *Sac* II fragment (Figure 4.10, lane 8) and 90 kb *Eag* I fragment (Figure 4.10,

lane 10) as YAC-L, but did not hybridize to the 60 kb *Bss* HII fragment detected by

YAC-L. The probe also hybridized to the same sized 220 kb *Sal* I fragment detected

by YAC-R (Table 4.2).   The PCR product hybridized to internal *Nae* I, *Sna* BI, *Sfi* I

and *Bss* HII fragments (Figure 4.10).   Based on these observations the D16S174

derived PCR product was localized to the region of DNA between the *Sal* I and *Eag* I

restriction enzyme sites located near the left arm vector end of 377C4. The long range

restriction map for 377C4 including the localization of the D16S174 PCR product is

shown in Figure 4.11.

Fig. 4.9A) Endlabel mapping of 377C4 with different restriction enzymes. A single

PFGE Southern blot filter containing 377C4 DNA digested with *Nru* I (lane 1), *Not* I

(lane 2) *Nae* I (lane 3), *Nar* I (lane 4), *Sal* I (lane 5), *Sfi* I (lane 6), *Sna* BI (lane 7),

Sac II (lane 8), *Rsr* II (lane 9), *Eag* I (lane 10), *Bss* H II (lane 11) and uncut 377C4

(lane 12) was sequentially hybridized with radiolabelled YAC-L (A this figure) and

YAC-R (next page). The sizes (in kb) of DNA molecular weight markers are shown

on the left side of the gel. Hybridization signals for YAC-L may be indicated with
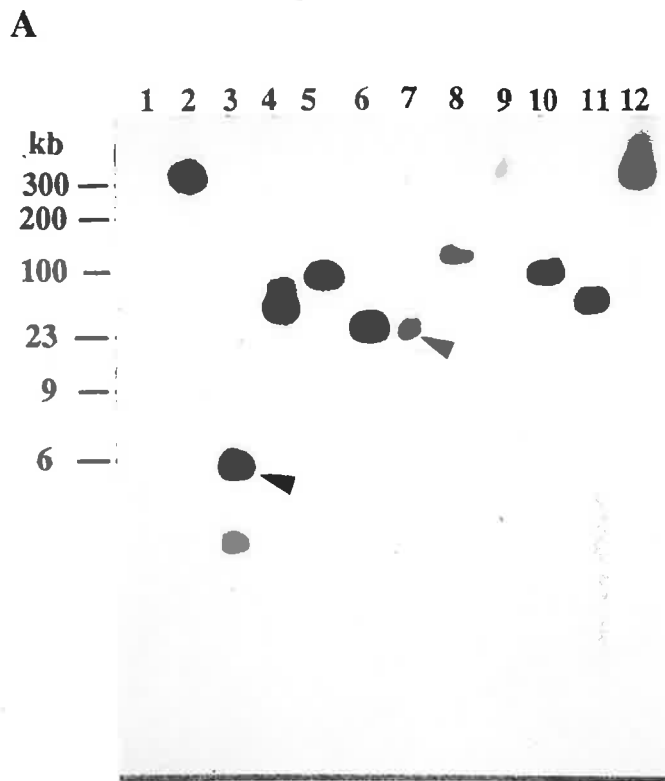
arrows.

Fig. 4.9B) Endlabel mapping of 377C4 with different restriction enzymes. A single

PFGE Southern blot filter containing 377C4 DNA digested with *Nru* I (lane 1), *Not* I

(lane 2), *Nae* I (lane 3), *Nar* I (lane 4), *Sal* I (lane 5), *Sfi* I (lane 6), *Sna* BI (lane 7),

Sac II (lane 8), *Rsr* II (lane 9), *Eag* I (lane 10), *Bss* HII (lane 11), and uncut 377C4

(lane 12) were sequentially hybridized with radiolabelled YAC-L (previous page) and

YAC-R (B, this figure). The sizes (in kb) of DNA molecular weight markers are

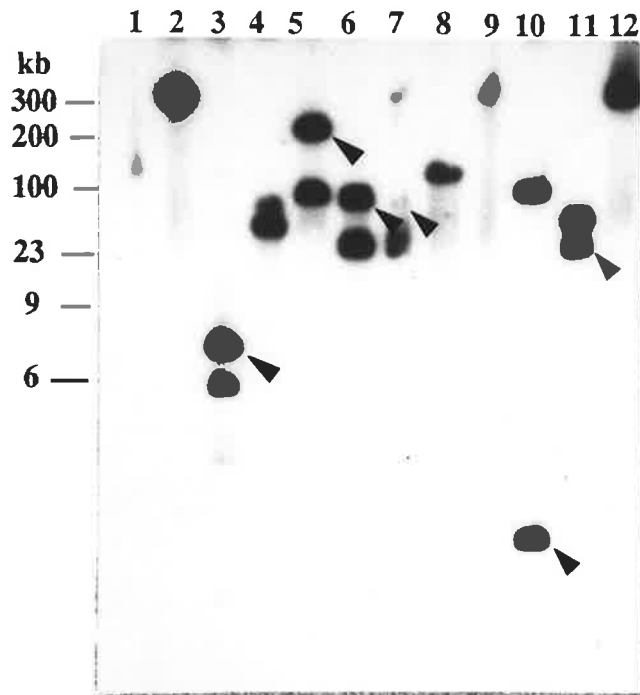shown on the left of the gel. Hybridization signals for YAC-R may be indicated with

arrows.
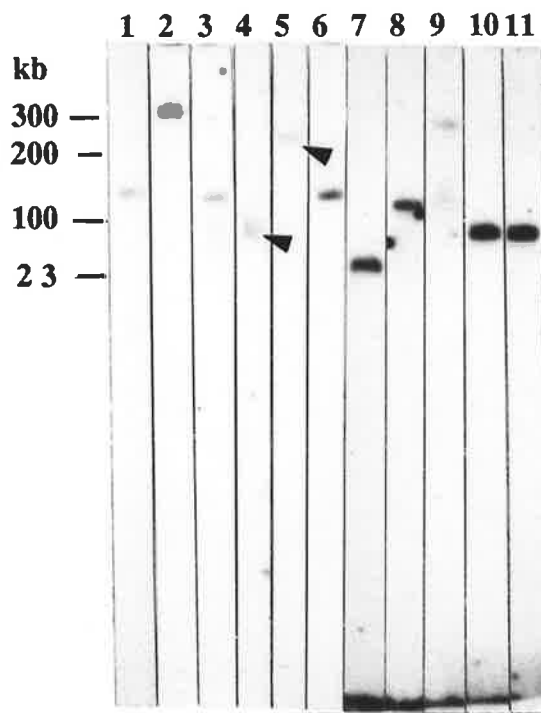
Fig. 4.10 Restriction enzyme digest fragments of 377C4 detected by a D16S174 derived PCR product. A PFGE Southern blot filter of 377C4 DNA digested with *Sfi* I (lane 1), *Not* I (lane 2), *Nae* I (lane 3), *Nar* I (lane 4), *Sal* I (lane 5), *Nru* I (lane 6), *Sna* BI (lane 7), *Sac* II (lane 8), *Rsr* II (lane 9), *Eag* I (lane 10) and *Bss* H II (lane 11), was hybridized with a radiolabelled D16S174 derived PCR product. The sizes (in kb) of DNA molecular weight markers are shown on the left of the gel.

Table 4.2 Restriction Enzyme Digest Hybridization Fragments (in kb) Detected for 377C4.

	PROBE		
ENZYME	YAC-L	YAC-R	D16S174 PCR
Not I	320	320	320
Nru I	N.R.	N.R.	130
Sal I	80	220	220
Nae I	4	5	130
Nar I	40	N.R.	70
Sfi I	20	90	130
Sac II	130	N.R.	130
Rsr II	N.R.	N.R.	N.R.
Eag I	90	2	90
Bss HII	60	20	90
Sna BI	20	70	50

Note: N.R. indicates that no hybridization was observed.

377C4 (320 kb)   ——— = 20 kb
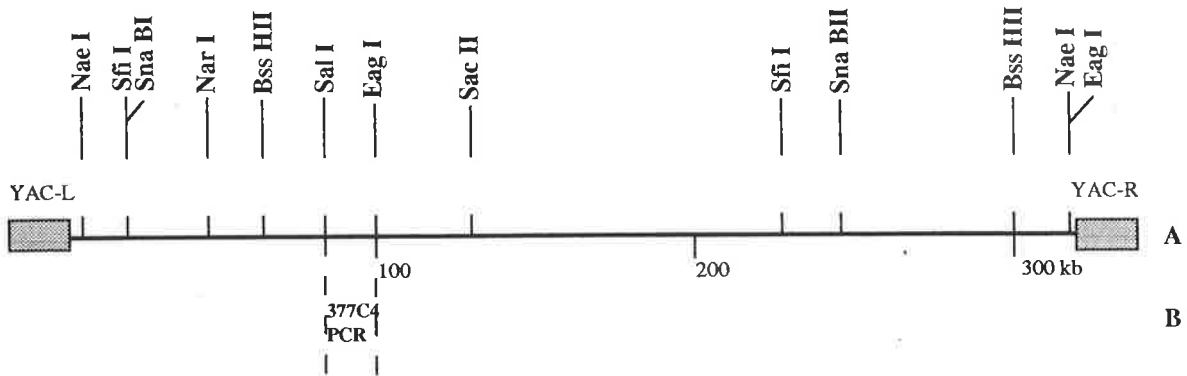

Fig. 4.11 Long range restriction map of YAC 377C4 (A) showing the location of the D16S174 derived PCR product (B). Complete map restriction enzyme sites in human DNA insert are shown only for *Sal* I. Other enzymes may have one or more additional restriction enzyme sites.

4.4.7   Recovery of YAC Endclones.

YAC endclones from 275H10 and 377C4 were recovered by the methods of vector/*alu*

PCR and capture PCR.

4.4.7.1   Vector/*Alu* PCR.

Vector/*alu* PCR assays were performed using 275H10 and 377C4 DNA as templates.

Control PCRs included no DNA template, *alu* primers alone and combinations of *alu*

primers, so as to distinguish internal *alu* products from endclones.

No potential endclones were detected for the left or right ends of 275H10 when *alu* I,

alu 2 or *alu* 3 primers were used in PCR assays with BYAC-L or BYAC-R vector

primers. With 377C4 as template DNA, BYAC-R and *alu* 3 primers produced a 600

bp PCR product (Figure 4.12, lane 16) that was not present in any of the control lanes.

This indicated that the PCR product may be the result of vector/*alu* PCR.

No unique products were generated from the left arm end of 377C4.   A second PCR

assay was performed in attempts to confirm that the 600 bp PCR product generated

from BYAC-R and *alu* 3 primers was a vector/*alu* product located at the right arm end

of 377C4. In place of the BYAC-R primer the PCR assay was repeated using YAC4-R

primer located downstream from BYAC-R.   The 600 bp PCR product from the first

assay and 377C4 DNA were both used as templates. A repeat of the first PCR assay

was included as a positive control.   The results of the second assay produced PCR

products of a size slightly smaller than the original assay (Figure 4.13). The size

difference corresponded to the distance from YAC4-R to BYAC-R. This indicated that

the 600 bp PCR product represented the stretch of DNA that contained the BYAC-R

primer and YAC4-R primer and extended into the human insert DNA sequence of
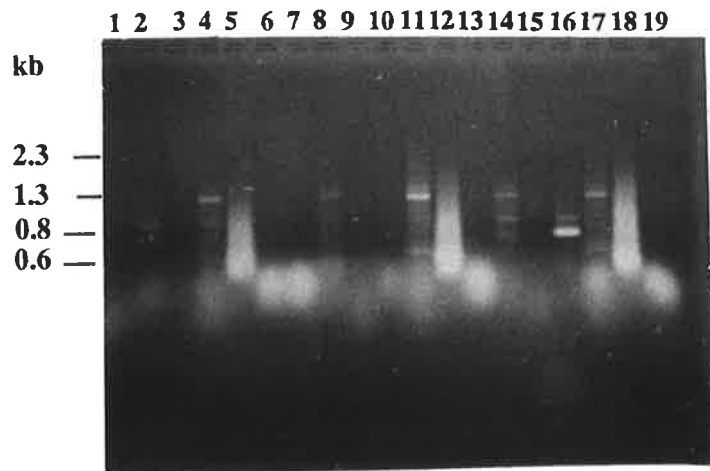
Fig. 4.12 Generation of endclones from 377C4 using vector/*alu* PCR. PCR assays were performed combining 377C4 DNA template with the PCR primers *alu* I (lane 1), *alu* 2 (lane 2), *alu* 3 (lane 3), *alu* 1 and 2 (lane 4), *alu* 1 and 3 (lane 5), *alu* 2 and 3 (lane 6), *alu* 1, 2 and 3 (lane 7), BYAC-L and *alu* 1 (lane 8), BYAC-L and *alu* 2 (lane 9), BYAC-L and *alu* 3 (lane 10), BYAC-L, *alu* 1 and *alu* 2 (lane 11), BYAC-L, *alu* 1 and *alu* 3 (lane 12), BYAC-L, *alu* 2 and *alu* 3 (lane 13), BYAC-R and *alu* 1 (lane 14), BYAC-R and *alu* 2 (lane 15), BYAC-R and *alu* 3 (lane 16), BYAC-R, *alu* 1 and *alu* 2 (lane 17), BYAC-R, *alu* 1 and *alu* 3 (lane 18), BYAC-R, *alu* 2 and *alu* 3 (lane 19). The sizes (in kb) of DNA molecular weight markers are shown on the left side of the gel.

Fig. 4.13 Identification and confirmation of a 600 bp endclone from the right arm end of 377C4. PCR assays were performed using BYAC-R/*alu* 3 PCR product as template with YAC-4R and *alu* 3 primers (lane 1 and 2), 377C4 DNA as template with YAC-4R and *alu* 3 primers (lane 3) and 377C4 DNA as template with BYAC-R and *alu* 3 primers (lane 4). The sizes (in kb) of DNA molecular weight markers are shown on the left size of the gel.
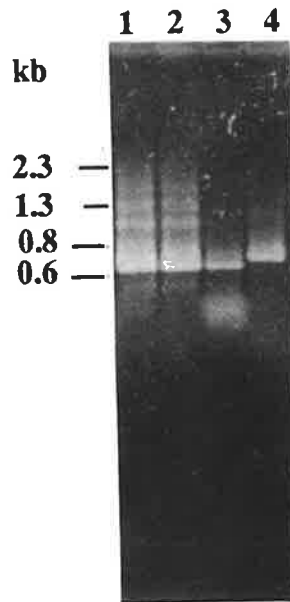
377C4 until it reached the first *alu* 3 primer recognition site.

4.4.7.2 Capture PCR.

275H10 and 377C4 DNA were digested with the restriction enzymes *Taq* I, *Hpa* II and *Hha* I to generate 3'-GC overhangs and *Alu* I, *Rsa* I, *Hpa* I, *Hinc* II and *Hae* III restriction enzymes were used to generate blunt ends, to allow for the ligation of synthetic linkers in the recovery of endclones by capture PCR. The left arm end of 275H10 produced two possible capture PCR products; a 700 bp *Hpa* I product (Figure 4.14, lane 1) and a 700 bp *Hinc* II product (Figure 4.14, lane 2). No capture PCR products were detected for the right arm end of 275H10.

An 800 bp *Rsa* I capture PCR product was detected for the left arm end of 377C4 (Figure 4.15, lane 2). The right arm end did not generate any capture PCR products.

4.4.8   *In situ* Hybridization of YAC Endclones (performed by L. Baker).

In an attempt to confirm that the vector/*alu* PCR product and the capture PCR products were not PCR artifacts, recombinant PCR™ 1000 plasmids containing the YAC endclones were used as probes for *in situ* hybridization to metaphase spreads of chromosomes that expressed FRA16B. For each probe, cells were scored for signals proximal, central and distal to FRA16B and are listed on Table 4.3. Based on the overall hybridization scores, Table 4.3 also shows the localization of each endclone with respect to FRA16B.

4.4.9   Endclone Sequencing and PCR Primer Construction.

Approximately 420 bp of sequence was generated from the 377C4 LHS endclone (Figure 4.16A), 230 bp of sequence from the 377C4 RHS endclone (Figure 4.16B) and 500 bp of sequence from the 275H10 LHS *Hpa* I endclone (Figure 4.17). DNA sequence was obtained from only one 275H10 LHS endclone. A Genbank homology search performed on the DNA sequences did not identify any significant homology

Fig. 4.14 Capture PCR products generated from 275H10. Identification of capture PCR products from the left arm end of 275H10 generated by restriction enzyme digestion with *Hpa* I (lane 1) and *Hinc* II (lane 2).
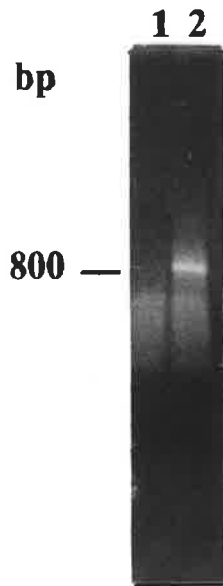
Fig. 4.15 A capture PCR product generated from 377C4. Identification of a capture

PCR product from the left arm end of 377C4 generated by restriction enzyme

digestion with *Rsa* I (lane 2). Lane 1 was a no DNA template control.

Table 4.3 *In situ* Hybridization of YAC Endclones to <u>FRA16B</u> Chromosomes.

ENDCLONE	*In situ* HYBRIDIZATION			LOCALIZ- ATION[a]
	PROXIMAL	CENTRAL	DISTAL	
377C4 LHS[b]	3	3	18	distal
377C4 RHS	-	-	-	-
275H10 LHS/*Hinc* II[d]	10	1	1	proximal
275H10 LHS/*Hpa* I[e]	11	2	5	proximal

a. - indicates the localization of the endclone with respect to <u>FRA16B</u>.
b. - refers to the 800 bp *Rsa* I derived capture PCR product from the left arm end of 377C4.
c. - refers to the 600 bp vector/*alu* PCR product derived from the right arm end of 377C4.
d. - refers to the 700 bp *Hinc* II derived capture PCR product from the left arm end of 275H10.
e. - refers to the 700 bp *Hpa* I derived capture PCR product from the left arm end of 275H10.

A.

```
5'- TTATAAAAAGAAGAAGGAATAAAG
     CTGTCCTCTTATACATCCTCATGTA
     AGGNTGCTTACATTCTTTGAGCATC
     CTTCTAAGCGCTTTATTCAGTCAAA
     AAACATTTACTAAGTATCTACTTTA
     AAAAACCCCTCTATTAAGGGCTAGA
     CAAAAATCCCTGCACTTGCGAAATT
     TGCATTCTCCTGGAGAGAAAGATGC
     ATAAAANATAAGCATAATACATTTG
     ATATTTGAAGATGACATGTGCTGTA
     ACAAAGAAAAACGAATAGATCACAG
     TAAGGAAGACTTACATGTACTACCT
     CAAAGGCCATTATTGTATAGACAGA
     TGCTTTTTTTATTCCTCGBTTTTA
     TACACGTAGANANTGANGCACAGAG
     AGGTTAAGTANCCTGCCCAACTATC-3'
```

377C4 LHS forward primer

5'-   CATTCTTTGAGCATCCTTCTAAGCG-3'

377C4 LHS reverse primer

5'-   GCATCTGTCTATACAATAATGGCCT-3'

B.

```
5'- AGAAGGGTTCGACCTTGAGATACTA
     CGAGACTCACTGCAAGCTCCACCTC
     CGGGTTCACGCCATTCTCCTGCCTC
     AGCGTCCCGAGTAGCTGGGACTACA
     GGTGCACGCCACCACACCCAGCTAA
     TTTTTTGTATTTTTTAGTAGAGACG
     GAGTTTCACCATGTTAGCCAGGATG
     GTTTCGACCTCCGACCTATGATCTG
     CCACCTGGCCTCCAAAGTCTGGATT
     ACAGG-3'
```

Fig. 4.16 Partial DNA sequence from A) 377C4 LHS endclone and B) 377C4 RHS endclone. PCR primer pairs are underlined in the DNA sequence data and are listed below each sequence. "N" denotes an unresolved base.

5' - T A C G N C T C A C T A T A G G G C G A A T T G G

G C C C T C T A G A T G C A T G C T C G A C G G G

C C G C C A G T G T G A T G G A T A T C T G C A G

A A T T C G G C T T G C G G T G A C C C G G G A G

A T C T G A A T T C A A C T T A T G A T T A G C T

T T T G A A T C C C A G T G C T G C T A G T G T T

T C C T T G A A A T A G T T C T T C C T A A T A T

C C C C A G T T A T A T G C T G C C A A T A C T T

T T N A C A T T A T T G A T T C A T A G T T A T C

T A T T A A T A T G A A T A T T T T C A A T A A T

G T G C A T T T T C T T T T C T G T A A T C C A C

C A C T G G G A A A A A A G A G A C C A C A C T T

A T T T T G C T C A G C A C T G A A T G A C T A G

G C A T T C A C C T A G G C A G G T C T T G T G G

A T G C C T T G G A T A T G A G G T G T G A A G G

G G A G A G C A A G A T T A A A A G A A A G T T A

T G G T A T T T G C C A A G T C A A G T G A T T T

T G A T C A C T C T T A A T T G G G A T G G G A G

G A G A T T T G G G G A A A A A A T A G G G T T T

T G G A A T G G G G T T G G N G G A T T A T G G G - 3'

275H10 LHS forward primer

5' - T G G A T A T C T G C A G A A T T C G G C T T G C - 3'

275H10 LHS reverse primer

5' - C A C T T G A C T T G G C A A A T A C C A T A A C - 3'

Fig. 4.17 Partial DNA sequence from 275H10 LHS *Hpa* I endclone. The PCR primer pairs are underlined in the sequence data and are listed below the sequence. "N" denotes an unresolved base.

between the sequences and known genes or repetitive DNA.

PCR primer pairs were constructed for the 377C4 LHS endclone and the 275H10 LHS *Hpa* I endclone. The primer sequences for the endclones are underlined in their corresponding figures and are listed below the sequence data. The PCR primers for the 377C4 LHS endclone produced a 267 bp PCR product and the primer pairs for the 275H10 LHS *Hpa* I endclone produced a 358 bp PCR product (data not shown).

4.5 DISCUSSION

The introduction of YACs and YAC chromosome "walking" to this project was a reflection of the technological advancements occurring with the large scale mapping of genomes. The shortfalls of genomic based long range restriction mapping already described (Chapter 3), was experienced by many investigators and the advent of cloning large DNA fragments in YAC's provided a strategy that circumvented these obstructions. In addition, it has also been shown that YAC libraries contain regions of DNA that could not be recovered from other vector libraries (Coulson *et al* 1988). One of the major aims in the long range restriction mapping of 16q13-16q22.1 was the physical spanning of the FRA16B locus. PFGE mapping with uncloned genomic DNA did not succeed in this goal. The introduction of YAC chromosome walking towards the fragile site presented a feasible strategy by which the goal may be achieved. It was by the use of YAC mapping strategies that the isolation and sequencing of the fragile site FRAXA was established (Kremer *et al* 1991a, Heitz *et al* 1991). Other regions of the human genome associated with human disease have also been cloned in YAC's, for example the entire cystic fibrosis (CF) gene (Green and Olsen,1990b).

4.5.1 YAC Clones.

The initiation of the YAC chromosome walk towards FRA16B was facilitated by the isolation of individual YAC clones from the flanking markers D16S163 and D16S174. PCR primers necessary for the PCR based screening of the YAC library had already been constructed from both markers. The ordered YAC library at CEPH represented approximately 20,000 clones with an average insert size of 250 kb. In such a library, single copy sequences were expected, on average, to be represented twice. The PCR based strategy for screening YAC libraries was chosen in favour of

other methods such as gel transfer hybridization or colony hybridization, due to its high sensitivity and high specificity. A single positive clone could be detected using only 100 ng of DNA from a pool representing thousands of YAC clones. Gel transfer hybridization of a similar sized YAC pool would require 300 times the amount of DNA. In addition, the speed by which the PCR based assay produced a result was many times faster than that of other procedures.

Two positive YAC clones were isolated from the markers flanking FRA16B. YAC 275H10 was isolated from the proximal flanking marker D16S163 and YAC 377C4 was isolated from the distal flanking marker D16S174. When the YACs were sized, both were found to be larger than the expected average of 250 kb. Only one YAC was present in each yeast preparation, however it was observed that the DNA insert in 275H10 had been modified. Problems associated with basic YAC technology may account for this loss of DNA integrity. The yeast's splicing mechanism, normally used for repairing DNA damage, is also the same mechanism that allows it to splice foreign DNA into its own genome. This mechanism is so effective that it can result in chimeric YACs by stitching together fragments of foreign DNA. Anderson (1993) suggested that this mechanism may also be responsible for the problems of deletions observed in YAC clones. Approximately 10% of the clones in YAC libraries were predicted to consist of DNA fragments from different parts of the genome that had been spliced together (Burke *et al* 1987), however, investigators have been finding the percentage of chimeric clones in practice to be much higher. For example, Foote *et al* (1992) experienced up to 59% chimeric clones in the YACs they have isolated.

However, the *in situ* hybridization of 275H10 to metaphase spreads of chromosomes expressing FRA16B not only confirmed the localization of 275H10 close to but proximal to FRA16B (Figure 4.5A) but did not localize the YAC to any other region

of the genome, indicating that chimerism may not be responsible for the modifications of 275H10. Internal rearrangements, duplications or deletions may be the cause. *In situ* hybridization of 377C4 to FRA16B chromosomes exhibited distal signal to FRA16B with a significant proportion of central and proximal signals also being recorded (Figure 4.5B). The results produced by this YAC reflected the *in situ* hybridization results obtained by XTY-26, the YAC that was found to span FRAXA (personal comm. L. Baker, Kremer *et al* 1991a).

With both flanking YAC's hybridizing in such a close proximity to FRA16B, it was necessary to determine whether the YAC's were overlapping and thus physically spanned the fragile site. Initially, a PCR based assay was performed to determine whether either of the flanking YAC's contained the PCR products constructed from both proximal and distal markers. The results for this assay were negative (Figure 4.3). If a relatively large overlap existed between 275H10 and 377C4, a comparison of the long range restriction maps of each YAC may have indicated the extent of the overlap. The long range restriction maps for 275H10 (Figure 4.8) and 377C4 (Figure 4.11) did not reveal any overlap between the two YAC's. *Bss H* II did not cleave 275H10, and 377C4 contained at least two *Bss* HII sites; one being located 60 kb in from the left end of the YAC and another 20 kb in from the right end. Therefore any overlap between the two YAC's would have to exist within either one of these regions. The distribution of restriction sites located at the extremities of 377C4 did not align with any clusters of sites in 275H10 indicating that any overlap would be unlikely. *Alu*-PCR has also been validated as a "fingerprinting" method for the assessment of YAC overlaps (Butler *et al* 1992). *Alu*-PCR allows the amplification of regions flanked by *alu* repeat elements within PCR distance of each other (Nelson *et al* 1989). In the generation of YAC endclones from 275H10 and

377C4 by vector/*alu* PCR (See Figure 4.12), the control lanes included various combinations of three *alu* primers with 275H10 and 377C4 DNA as templates. No common PCR products or "fingerprints" were observed between 275H10 and 377C4. Based on this method no YAC overlap could be established. Although not included in this study, non PCR based fingerprinting methods have also been described (Wada *et al* 1990). Overlapping YAC's have been established by identifying a sub-population of restriction fragments in each YAC using hybridization with radioactive repetitive sequences such as *alu*'s or LI sequences (which occur in the human genome, on average, every 4 kb and 50 kb respectively; Singer 1982). The hybridizations result in a "fingerprint" not unlike the *alu* PCR derived "fingerprint".

It was concluded that if any overlap existed between 275H10 and 377C4, it would be small and would require finer restriction mapping of the individual YAC's for detection. Rather than focusing on detailed restriction maps for YAC's that may not overlap or span FRA16B it was proposed that the generation of end clones from each YAC would not only be used to continue the YAC walk towards FRA16B but would be incorporated as probes back to 275H10 and 377C4 to align any overlaps that may exist between them.

4.5.2   Recovery of Endclones.

A number of methods have been established to recover endclones from YAC's, including end rescue subcloning (Burke *et al* 1987, McCormick *et al* 1990), genomic end rescue PCR (Silverman *et al* 1989) inverse PCR (Ochman *et al* 1988) and conventional subcloning (Bellanne-Chantelot *et al* 1991). In addition, vector/*alu* PCR (Nelson *et al* 1989, Breukel *et al* 1990, Ledbetter *et al* 1990, Orita *et al* 1990) and capture PCR (Riley *et al* 1990, Lagerstrom *et al* 1991, Rosenthal *et al* 1991) have also been described. Based on the rapidity of these final two techniques and of their

high success rates, these methods were used to generate endclones for 275H10 and 377C4. The identification of these unique sequences on the specified clones have also been termed sequence tagged sites or STS's.

The vector/*alu* PCR based assays utilized PCR primers developed to identify the left and right terminal sequences of the YAC vector and human-specific sequences directed at the human *alu* repeat element (Nelson *et al* 1989). Repeat elements homologous to the human *alu* repeat are also found in rodents but there is sufficient sequence divergence to reduce cross hybridization (Jelenik and Schmid,1982, Britten *et al* 1988, Jurka and Smith,1988). The *alu* PCR primers were constructed from highly conserved regions of the repeat and were located close to one of the ends of the consensus sequence (Nelson *et al* 1989). Vector/*alu* PCR using 275H10 and 377C4 as substrate DNA's only recovered one verified endclone from the right arm end of 377C4. The lack of products for the other YAC ends may indicate that no *alu* recognition sequences are located within PCR distance of these ends. Alternatively, the actual *alu* repeats may have diverged from the consensus sequence, resulting in a lower homology of primer to template. Tagle and Collins (1992) described a pair of improved *alu* PCR primers that were designed from the consensus sequences of *alu*. In general, the improved primers were designed to recognise and prime from the highly conserved regions of *alu* elements, thus amplifying the more abundantly dispersed but anciently inserted and more divergent *alu* elements (Tagle and Collins, 1992). The two primers, *alu* 4 and *alu* 5 (see Table 2.1) were included in PCR assays with 275H10 and 377C4 but without any additional endclones being produced, further supporting that no *alu* repeats were present in close proximity to the remaining vector ends.

The method of capture PCR did not rely on the presence of repeat DNA in the

human DNA inserts. This method generated endclones from the left arm end of 377C4 and from the left arm end of 275H10. Alternative strategies may be required to obtain endclones from the right arm end of 275H10.

In situ hybridization of the endclones from 377C4 and from the left arm end of 275H10 did not establish whether 377C4 or 275H10 spanned FRA16B. The left arm end of 377C4 hybridized distal to FRA16B but no specific *in situ* hybridization signals could be generated from the right arm end of 377C4. The 600 bp of endclone sequence may have been too small to use as an effective probe. The left arm end of 275H10 hybridized proximal to FRA16B. These results could only establish that the left arm end of 377C4 and left arm end of 275H10 did not extend across the fragile site.

Sequencing of the three endclones generated enough sequence to construct PCR primer pairs for the left arm ends of 377C4 and 275H10. PCR primers can be constructed for the right arm end of 377C4 but for only a small PCR product.

The chromosome walk was abandoned at this stage due to the arrival and immediate analysis of a cosmid contig that was reported by our collaborators in Los Alamos to contain FRA16B (see Chapter 5). I was unable to complete this section of the project, it has been continued by other investigators. The remaining manipulations considered necessary to complete this section will be described.

In order to continue the chromosome walk towards FRA16B, PCR primers are required from each endclone of 377C4 and 275H10. PCR primers need to be constructed from the right arm endclone of 377C4, which may first require additional sequencing. Alternative methods such as end rescue subcloning, genomic end rescue PCR and inverse PCR may be necessary to generate endclones from the right arm end of 275H10.

These endclones may be used as probes in order to determine whether any overlap exists between 377C4 and 275H10. If so, then FRA16B has been cloned. The endprobes may also be mapped against hybrid cell line panels whose breakpoints are located near FRA16B. If 377C4 and/or 275H10 are found to span any of these breakpoints it will be possible to orientate the YACs along chromosome 16. Most importantly, the PCR primers will be used to rescreen the YAC library for the isolation of overlapping YAC clones. By using *in situ* hybridization of the new YACs to FRA16B chromosomes, it may be determined which overlapping YACs extend towards the fragile site and which extend away. In addition, the introduction and availability of a CEPH constructed "mega YAC" library (Bellane-Chantelot *et al* 1992); the entire human genome cloned in extra large YACs (up to 1000 kb inserts) may allow for the isolation of very large YACs for this region of the chromosome. The subsequent restriction mapping of a number of YACs for any one region of the chromosome may identify any internal rearrangements and deletions that have been found in YACs and especially in "mega YACs" (Anderson 1993). By building up a YAC contig from both sides of FRA16B using the original YACs and "mega YACs" it is possible that only one round of rescreening would be necessary to span FRA16B.

CHAPTER 5

ANALYSIS OF A COSMID CONTIG THAT MAPPED
TO THE SAME PHYSICAL INTERVAL AS <u>FRA16B</u>

5.1 BACKGROUND TO CHAPTER 5

Whilst I was attempting to clone FRA16B by chromosome walking with YACs, a series of overlapping cosmid clones had been identified that, by preliminary analysis, appeared to span FRA16B. A representative cosmid 328F12, from a cosmid contig encompassing approximately 190 kb of DNA had been finely mapped by hybrid cell line panel analysis to the same physical interval as FRA16B (Z. Chen, unpublished data). *In situ* hybridization of the two terminal cosmids of this contig to FRA16B chromosomes localized the cosmids to either side of FRA16B, indicating that the contig spanned the fragile site.

Based on these initial findings and assuming that the cosmids did form a contig, the YAC walk towards FRA16B was suspended. Had FRA16B been cloned within a YAC the next step in the project would have been to further localize the fragile site by constructing a cosmid library from the YAC insert containing FRA16B and then identifying cosmids that spanned the fragile site. The 190 kb contig from Los Alamos provided such cosmids so further YAC analysis was considered to be redundant. It was proposed that the immediate analysis of the overlapping cosmid clones encompassing FRA16B would facilitate the rapid isolation and characterization of this fragile site.

5.2   SUMMARY

A 190 kb cosmid contig was mapped to the same physical interval as FRA16B. It was assembled by use of a genetic contig assembly algorithm, GCAA. The total contig was comprised of two separate contig series linked together. The overlapping cosmids 10c6, 60h5, 45b11 and 57b6 were included in contig 9 and 68g12, 40e6 and 14d9 were included in contig 70. The algorithm predicted a small overlap between 57b6 and 68g12 joining the two series together. *In situ* hybridization of these cosmids to FRA16B chromosomes mapped the clones to either side of FRA16B indicating that the fragile site had been spanned. In an attempt to confirm this localization, representative cosmids from contig 9 and 70 were finely mapped with respect to FRA16B using a hybrid cell line panel that contained breakpoints located near FRA16B. Contig 9 was mapped to the same physical interval as FRA16B but contig 70 was mapped distal to 16q22.3, a region of DNA many megabases distal to FRA16B. These findings established that the cosmids did not form an uninterrupted contig across FRA16B. The GCAA algorithm predicted overlap between 57b6 and 68g12 was identified as false and that contig 9 and 70 were not linked. The false overlap between these two cosmids indicated that FRA16B was unlikely to have been cloned.

The *in situ* hybridization of whole cosmids to FRA16B chromosomes had identified three cosmids from contig 9 - 57b6, 45b11 and 60h5 that were localized in close proximity to FRA16B possibly encompassing the fragile site. *In situ* hybridization of endprobes generated from these cosmids localized all three cosmids proximal to FRA16B. However, endprobes derived from the T7 end of 57b6 exhibited a high proportion of signals central and distal to FRA16B identifying these probes as the closest proximal markers to FRA16B possibly extending into the fragile site. Using

these markers as probes, a comparison of hybridization fragments of genomic DNA isolated from normal individuals and individuals expressing FRA16B did not detect any variable bands that may have been associated with FRA16B. A T7 endprobe derived from 57b6 was also used to initiate a cosmid "walk" aimed to isolate overlapping cosmids that may span FRA16B. The screening of an ordered cosmid library recovered one overlapping cosmid 328F12. *In situ* hybridization of the cosmid and its derived endprobes to FRA16B chromosomes established that this cosmid did not encompass FRA16B.

5.3   INTRODUCTION

The development of ordered clone or contig maps for human chromosomes provide an immediate source of cloned DNA from virtually any region of the genome. Recently, the detection of overlapping cosmid clones based on restriction fragment sizes and the distribution of repetitive sequences within these fragments had been described (Stallings *et al* 1990).   Using this method, called repetitive sequence fingerprinting, a chromosome 16 contig map had been constructed with an estimated chromosome coverage of approximately 84% (Stallings *et al* 1992a).   A 190 kb cosmid contig constructed by this method had been finely localized by hybrid cell line panel analysis to the same physical interval as FRA16B (Z. Chen, unpublished data). The aim of this section of the project was to determine whether the contig spanned FRA16B, and if so, to systematically localize the fragile site to a region small enough to allow for its molecular characterization.

A strategy was devised for achieving these goals based on the assumption that the cosmids formed an uninterrupted contig.   Initially, a combination of *in situ* hybridization to FRA16B chromosomes coupled with hybrid cell line panel analysis using whole cosmids as probes, would be used to determine whether the contig spanned FRA16B.   The *in situ* hybridization results may also target any individual cosmids that appear to encompass the fragile site.   Endprobes would be isolated from these candidate cosmids to provide smaller, more sensitive probes.   *In situ* hybridization of endprobes from a cosmid to either side of FRA16B would confirm that the cosmid flanked the fragile site.   Once FRA16B had been localized within a cosmid then finer localization would be implemented by direct Southern blot analysis of genomic DNA.   A probe for FRAXA was found to hybridize to different sized fragments of DNA between normal and FRAXA affected individuals (Yu *et al* 1991).

Based on these observations individual restriction digest fragments from the cosmid would be used as probes to compare hybridization signals from the DNA of normal individuals and individuals expressing <u>FRA16B</u>. Probes that produced different hybridization signals between normal and affected individuals may be associated with <u>FRA16B</u>. The sequencing of these altered fragments may lead to the molecular characterization of this fragile site.

This chapter describes the implementation of this strategy to the 190 kb cosmid contig.

5.4   MATERIALS AND METHODS

5.4.1   Cosmid Contig Clones.

Seven overlapping cosmid clones 10c6, 60h5, 45b11, 57b6, 68g12, 40e6 and 14d9 representing a total of 190 kb of DNA, were kindly provided by R. Stallings, Los Alamos National Laboratories, New Mexico. Although the total contig comprised of 10 cosmids, if more than one cosmid represented the same region of DNA only the larger cosmid was provided. The estimated sizes of the individual cosmids and the predicted overlap between them as depicted in the cosmid contig map (Figure 5.1), were determined by the genetic algorithm GCAA (See 2.10.3.2). Screening of an ordered cosmid library recovered an additional cosmid 328F12 which was also made available.

5.4.2   Hybrid Cell Line Panel Analysis.

The hybrid cell lines used to finely map cosmid probes, including their portion of chromosome 16 present, are illustrated in Table 5.1.

5.4.3   Isolation of Cosmid Endprobes.

DNA fragments representing the terminal sequences of genomic inserts in cosmids (T3 and T7 ends) were isolated by the methods already described (2.10.4). Endprobes were radiolabelled by 5' end labelling methods (2.5.4.1).

5.4.4   Restriction Enzyme Digestion.

Genomic DNA isolated from lymphoblastoid cell lines representing normal individuals and individuals expressing FRA16B were digested with the restriction enzymes *Bam* HI and *Pst* I. Hybrid cell line DNA was digested with *Taq* I and cosmid DNA was digested with *Pst* I. All restriction enzyme digestions were performed according to the manufacturer's instructions.

190 kb

(70) 14d9
38 kb

(70) 40e6
31 kb

(70) 23e10
18 kb

(70) 68g12
35 kb

(9) 57b6
38 kb

(9) 328F12
32 kb

(9) 51F3
34 kb
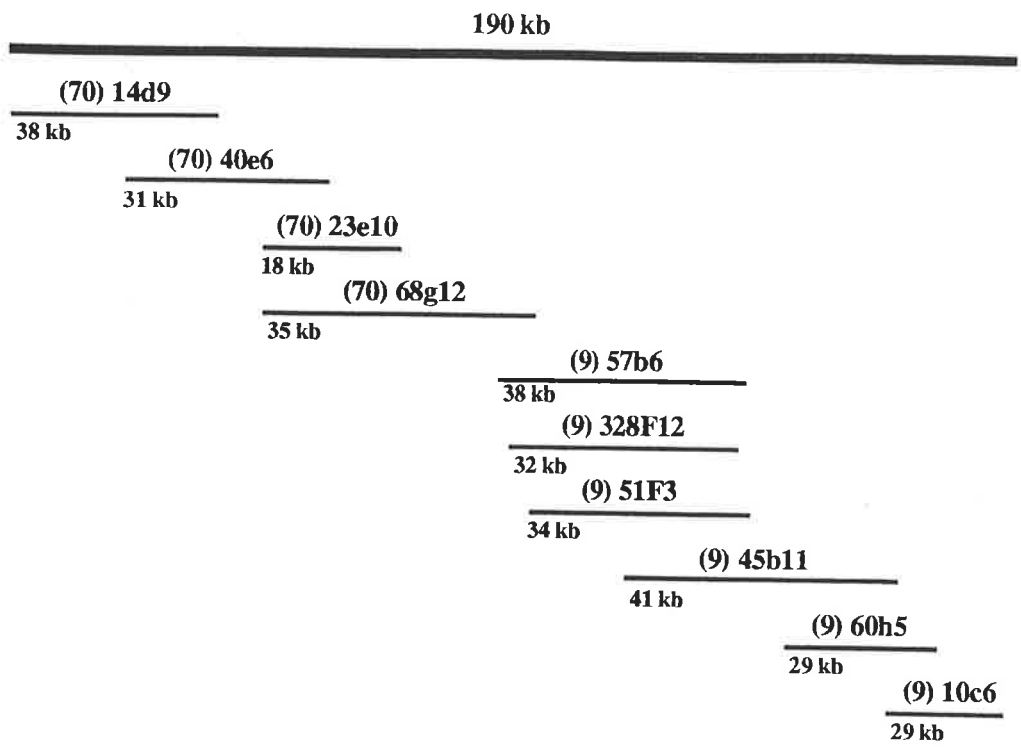
(9) 45b11
41 kb

(9) 60h5
29 kb

(9) 10c6
29 kb

Fig. 5.1 The GCAA derived 190 kb cosmid contig. The solid bars represent each cosmid and depict the predicted overlap between cosmids. The cosmid sizes are shown below each bar. The map was provided by the Los Alamos National Laboratories, New Mexico.

Table 5.1 Chromosome 16 Long Arm Hybrid Cell Line Panel Used to Finely Map Cosmid Probes.

Cell Line	Portion of 16 present	Reference
CY 18	Complete 16	Callen (1986)
CY 140	q12.1 -)qter	Callen *et al* (1990a)
CY 135	q13 - qter	Callen *et al* (1990a)
CY 138	pter -)q12.1, q13 -)qter	-
CY 143	q13 -)qter	-
CY 126	q13 -)qter	-
CY 7	q13 -)qter	Callen (1986)
CY 122	q21 -)qter	-
CY 130	pter -)q21, q21 -)qter	Callen *et al* (1990a)
CY 125	pter -)q21, q22.1 -)qter	-
CY 127	pter -)q21, q22.1 -)qter	Callen *et al* (1990a)
CY 4	q22.1 -)qter	-
CY 128	q22.1 -)qter	Callen *et al* (1988)
CY 6	q22.1 -)qter	-
CY 13A	q22.2 -)qter	Callen (1986)
CY 5	q22.3 -)qter	

5.4.5 Screening the Ordered Cosmid Library.

High density cosmid grids were screened with a T7 derived endprobe from 57b6 by the methods described (2.10.3.1). The results of the screening were sent to the Los Alamos National Laboratories, New Mexico, where the ordered chromosome 16 cosmid contig library was located. The construction of this library has been described (2.10.3). Cosmids positive for the endprobe were returned in glycerol stabs.

5.5 RESULTS

5.5.1 *In situ* Hybridization of the Cosmid Contig to FRA16B Chromosomes.
(performed by L. Baker and H. Eyre).

In situ hybridization of the seven overlapping cosmids 10c6, 60h5, 45b11, 57b6, 68g12, 40e6 and 14d9 to chromosomes expressing FRA16B, was used to localize each cosmid with respect to the fragile site. For each cosmid, the total number of cells scoring proximal, central and distal signals to FRA16B is presented in Table 5.2.

The two cosmids 10c6 and 14d9 representing either end of the contig, were shown to be localized on either side of FRA16B. Of the 20 cells scored for 10c6, only proximal signals were observed (Table 5.2a). 19 cells were scored for 14d9 with only distal signals being observed (Table 5.2g).

Three central overlapping cosmids 60h5, 45b11 and 57b6 exhibited proximal, central and distal signals to FRA16B targeting them as candidate cosmids for containing the fragile site. 30 cells were scored for 60h5 with 50% of the cells presenting proximal signals, 27% central and 23% distal (Table 5.2b). Of the 35 cells scored for 45b11, 45% of the cells expressed proximal signals 42% central and 13% distal (Table 5.2c). 31 cells were scored for 57b6 with 74% of the cells locating 57b6 proximal, 14% central and 12% distal (Table 5.2d).

The predominantly proximal signals observed for 57b6 contrasted with the almost entirely distal signals observed by its overlapping cosmid 68g12. 95% of the 20 cells scored for 68g12 expressed distal signals (Table 5.2e). A similar contrast was observed between 60h5 and 10c6. 60h5 expressed a range of signals across FRA16B but its overlapping cosmid 10c6 expressed only proximal signals. Based on these observations, the integrity of these overlaps were questioned.

Table 5.2   *In situ* Hybridization Scores of the Cosmid Contig to __FRA16B__
Chromosomes.

Cosmid	*In situ* Hybridization		
	Proximal	Central	Distal
a) 10c6	20	0	0
b) 60h5	15	8	7
c) 45b11	16	15	4
d) 57b6	23	5	3
e) 68g12	1	0	19
f) 40e6	0	0	20
g) 14d9	0	0	19

Numbers indicate the total number of cells scored for each cosmid.

5.5.2 Fine Localization of the Cosmid Contig by Hybrid Cell Line Panel Analysis.

The two cosmids 10c6 and 14d9, located at either end of the contig were probed against a hybrid cell line panel comprised of *Taq* I digested cell line DNA containing breakpoints located along the long arm of chromosome 16 (see Table 5.1 for the chromosome 16 content of each hybrid cell line). Total human DNA and mouse A9 cell line DNA were included as positive and negative controls respectively.

The proximal cosmid 10c6 hybridized to cell lines CY 140, CY 135, CY 138, CY 126, CY 7, CY 122 and to total human DNA. No hybridization was observed to the cell lines CY 130, CY 125, CY 127, CY 4 and to mouse A9 (Figure 5.2), thus localizing 10c6 to the same physical interval as FRA16B.

The distal cosmid 14d9 hybridized to all cell lines in the hybrid cell line panel except for the mouse A9 negative control (Figure 5.3). The hybridization of 14d9 to CY 130, CY 125, CY 127 and CY 5 indicated that the cosmid was located distal to CY 5 (16q22.3) representing a distance of many megabases distal to FRA16B.

The localization of 14d9 distal to CY 5 indicated that the cosmids did not form a contiguous length of DNA. Based on the GCAA information sent with the cosmid contigs from Los Alamos, 14d9, 40e6 and 68g12 formed one contig (number 70) which was linked to a second contig (number 9) that encompassed 57b6, 45b11, 60h5 and 10c6. It was proposed that the overlap between 57b6 and 68g12 did not exist and that contig 70 was located distal to 16q22.3.

The localization of 68g12 was established by its hybridization to the hybrid cell panel. Positive signals were observed for all hybrid cell lines including CY 130, CY 125, CY 127 and CY 5, (Figure 5.4) further supporting the location of contig 70 distal to the CY 5 breakpoint.
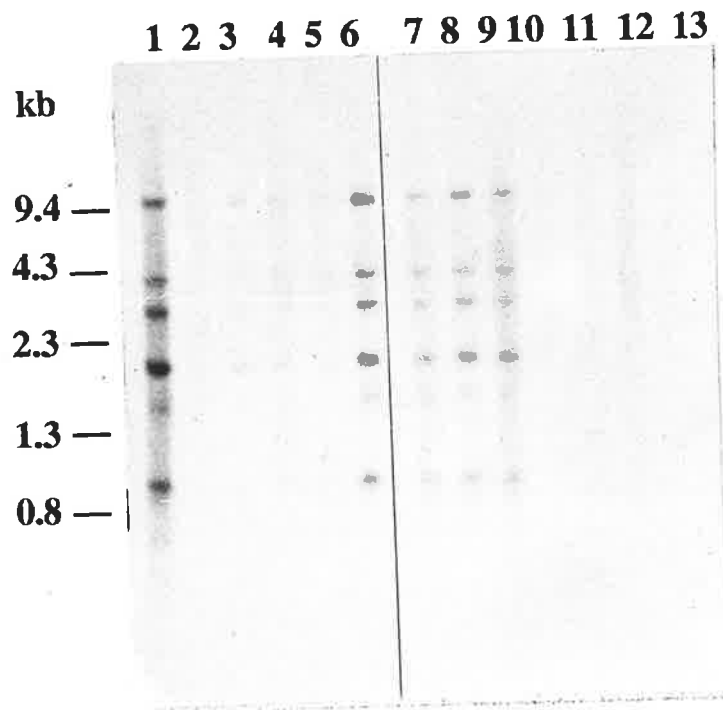
Fig. 5.2 Hybridization of cosmid 10c6 to a hybrid cell line panel. Radiolabelled total cosmid DNA from 10c6 was hybridized to a Southern blot filter of *Taq* I digested DNA samples from total human (lane 1), mouse A9 (lane 2), CY 18 (lane 3), CY 140 (lane 4), CY 135 (lane 5), CY 138 (lane 6), CY 126 (lane 7), CY 7 (lane 8), CY 122 (lane 9), CY 130 (lane 10), CY 125 (lane 11), CY 127 (lane 12) and CY 4 (lane 13). The portion of chromosome 16 present in each hybrid cell line is shown on Table 5.1. The sizes of DNA molecular weight markers are shown (in kb) on the left side of the gel.

Fig. 5.3 Hybridization of cosmid 14d9 to a hybrid cell line panel. Radiolabelled total cosmid DNA from 14d9 was hybridized to a Southern blot filter of *Taq* I digested DNA samples from total human (lane 1), mouse A9 (lane 2), CY 18 (lane 3), CY 126 (lane 4), CY 130 (lane 5), CY 128 (lane 6), CY 143 (lane 7), CY 127 (lane 8), CY 6 (lane 9), CY 125 (lane 10), CY 13A (lane 11) and CY 5 (lane 12). The portion of chromosome 16 present in each hybrid cell line is shown in Table 5.1. The sizes of DNA molecular weight markers are shown (in kb) on the left side of the gel.
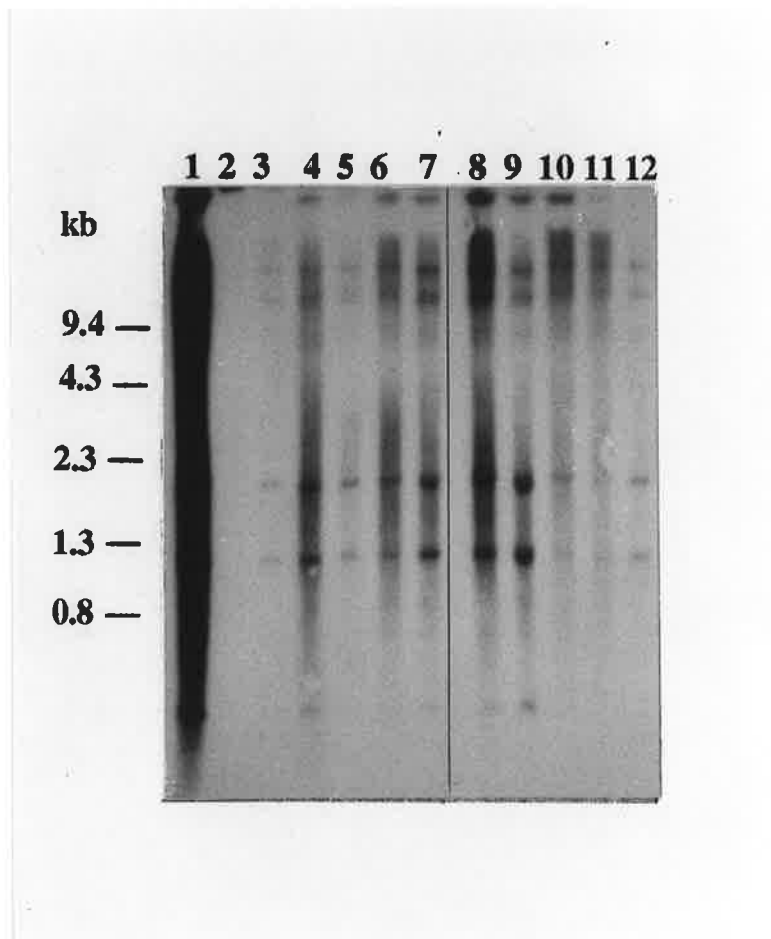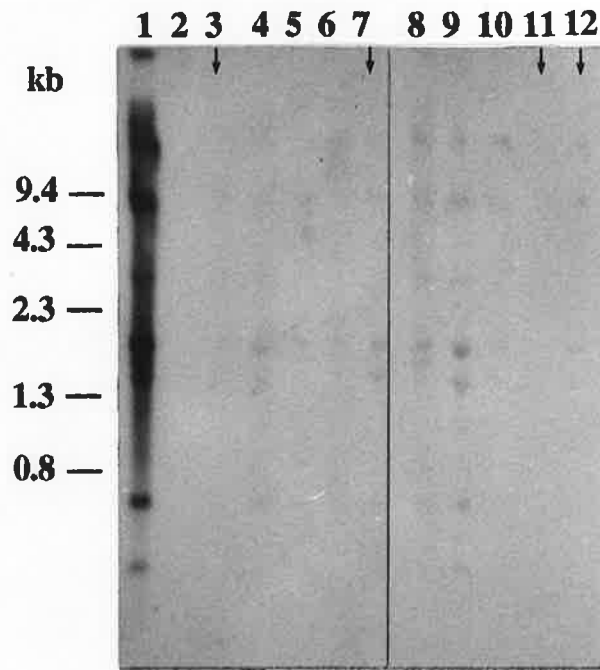
Fig. 5.4 Hybridization of cosmid 68g12 to a hybrid cell line panel. Radiolabelled total cosmid DNA from 68g12 was hybridized to a Southern blot filter of *Taq* I digested DNA samples from total human (lane 1), mouse A9 (lane 2), CY 18 (lane 3), CY 126 (lane 4), CY 130 (lane 5), CY 128 (lane 6), CY 143 (lane 7), CY 127 (lane 8), CY 6 (lane 9), CY 125 (lane 10), CY 13A (lane 11) and CY 5 (lane 12). The portion of chromosome 16 present in each hybrid cell line is shown on Table 5.1. The sizes of DNA molecular weight markers are shown (in kb) on the left side of the gel. Lanes with faint bands are marked with an arrow.

5.5.3 Isolation of Endprobes from 57b6, 45b11 and 60h5 and Analysis by *in situ* Hybridization (*in situ* hybridization performed by H. Eyre).

Endprobes were isolated from the overlapping cosmids 57b6, 45b11 and 60h5 in order to determine the finer localization of these cosmids with respect to FRA16B. The endprobes provided a series of smaller probes representing six defined regions of DNA for the overlapping cosmids. The sizes of endprobes (i.e. T7 and T3 ends) for cosmids 57b6, 45b11 and 60h5 are shown in Table 5.3. *In situ* hybridization scores of cosmid endprobes to FRA16B chromosomes are shown in Table 5.4 for 57b6 and Table 5.5 for 45b11 and 60h5. Three endprobes were generated from 57b6 representing 0.7 kb, 3.0 kb and 3.8 kb of genomic DNA from the T7 end (Figure 5.5). A total of 240 cells were scored for *in situ* hybridization of the T7 endprobes to FRA16B chromosomes (Table 5.4). Although 63% of the cells expressed proximal signals, 17% resulted in central signals and 20% were distal. Of the six endprobe regions analysed, the 57b6 T7 endprobes recorded the highest proportion of central and distal signals to FRA16B. More significantly, a total of 14% of the 88 cells with fragile sites greater than the width of a chromatid expressed distal signals when probed with the 57b6 T7 endprobes. The presence of central and distal signals across such a large physical gap in the chromosome provided greater confidence that the probe actually contained sequences that may be located on either side of the fragile site. Therefore it was proposed that the T7 end of 57b6 may be the closest proximal marker to FRA16B, possibly extending into the fragile site.

Three endprobes were isolated from the T3 end of 57b6, representing 0.4kb, 0.7kb and 2.8kb of terminal sequence genomic DNA (Figure 5.6). A total of 51 cells were scored with 73% of the cells exhibiting proximal signal, 17% central and 10% distal (Table 5.4). A lesser proportion of the total number of cells produced central and

Table 5.3   Sizes of T7 and T3 Endprobes from 57b6, 45b11 and 60h5.

Cosmid	T7 Endprobes	T3 Endprobes
57b6	0.7 kb *Pvu* II	2.8 kb *Pvu* II
	3.0 kb *Pst* I	0.7 kb *Pst* I
	3.8 kb *Xmn* I	0.4 kb *Xmn* I
45b11	1.7 kb *Pst* I	2.5 kb *Pst* I
	0.25 kb *Xmn* I	9.5 kb *Xmn* I
60h5	1.8 kb *Pst* I	2.3 kb *Pst* I

Table 5.4 *In situ* Hybridization of 57b6 Endprobes to <u>FRA16B</u> Chromosomes.

| Cosmid End	End Probe	Size of Fragile site	*In situ* Hybridization			
			Proximal (P)	Central (C)	Distal (D)	Other Signals
T7 (57b6)	*Pvu* II (0.7kb)	⟨width of chromatid	20	14	14	P + D:4
		⟩width of chromatid	21	2	4	
	Pst I (3.0kb)	⟨width of chromatid	23	16	11	P + D:7
		⟩width of chromatid	21	0	4	
	Xmn I (3.8kb)	⟨width of chromatid	24	7	12	
		⟩width of chromatid	30	2	4	
T3 (57b6)	*Pvu* II (2.8kb)		10	4	2	-
	Pst I (0.7kb)		10	2	0	-
	Xmn I (0.4kb)		17	3	3	-

Table 5.5   *In situ* Hybridization of 45b11 and 60h5 Endprobes to <u>FRA16B</u> Chromosomes.

Cosmid End	Endprobe	Size of fragile site	*In situ* Hybridization			
			Proximal(P)	Central(C)	Distal(D)	Other Signals
T7 (45b11)	*Pst* I (1.7kb)	⟨width of chromatid	24	4	6	
		⟩width of chromatid	2	1	2	
	Xmn I (0.25kb)	-	4	1	5	high background
T3 (45b11)	*Pst* I (2.5kb)	⟨width of chromatid	25	7	6	
		⟩width of chromatid	6	1	0	
	Xmn I (9.5kb)	⟨width of chromatid	16	7	3	
		⟩width of chromatid	6	1	0	
T7 (60h5)	*Pst* I (1.8kb)	-	18	4	0	-
T3 (60h5)	*Pst* I (2.3kb)	-	19	5	1	-
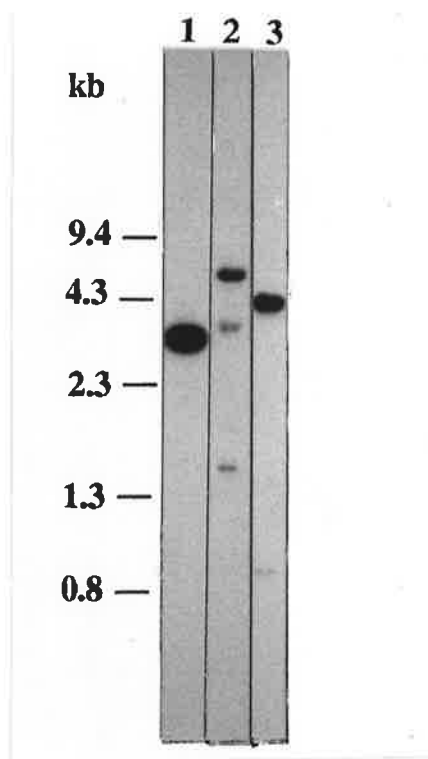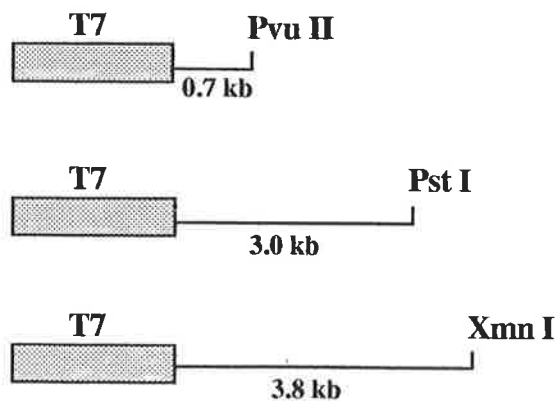
Fig. 5.5 Isolation of T7 endprobes from 57b6. (A) A radiolabelled T7 oligomer was hybridized to a Southern blot filter of cosmid 57b6 DNA digested with *Pvu* II (lane 1), *Pst* I (lane 2) and *Xmn* I (lane 3). Molecular weight DNA markers are shown (in kb) on the left side of the gel. The sizes of genomic insert DNA present in each endprobe are shown in (B) and the restriction map of the insert at the T7 end of the cosmid is shown in (C).
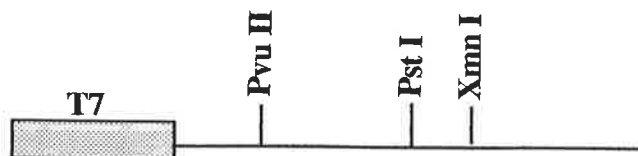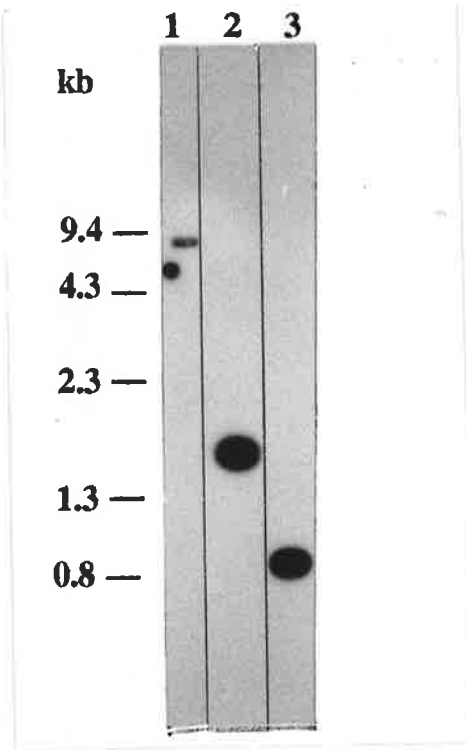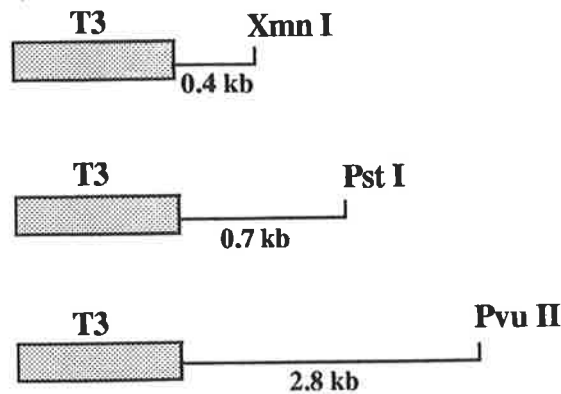
Fig. 5.6 Isolation of T3 endprobes from 57b6. (A) A radiolabelled T3 oligomer was hybridized to a Southern blot filter of cosmid 57b6 DNA digested with *Pvu* II (lane 1), *Pst* I (lane 2) and *Xmn* I (lane 3). Molecular weight DNA markers are shown (in kb) on the left side of the gel. The sizes of genomic insert DNA present in each endprobe are shown in (B) and the restriction map of the insert at the T3 end of the cosmid is shown in (C).
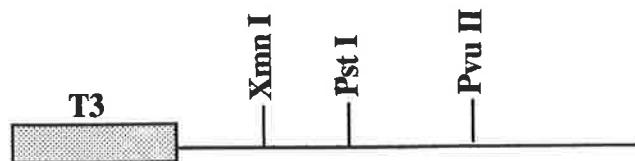
distal signals when compared to the T7 end of 57b6 suggesting that the T3 end may lie in the overlapping region of the contig under investigation.

Two endprobes were isolated from the T7 end of 45b11, representing 0.25 kb and 1.7 kb of genomic DNA, and two endprobes were isolated from the T3 end representing 2.5 kb and 9.5 kb of genomic DNA (Figure 5.7). A total of 49 cells were scored for the T7 endprobes and 78 cells were scored for the T3 ends. For the T7 end, 64% of the cells produced proximal signal, 12% were central and 24% recorded distal signals (Table 5.5). The high proportion of distal signals observed for the T7 endprobes of 45b11 may be falsely skewed by the signals recorded for the 0.25 kb *Xmn* I endprobe. The size of this probe was too small for accurate *in situ* hybridization, with a high background being observed. As a result, only nine cells were scored and five were recorded distal.

The T3 endprobes of 45b11 recorded 68% proximal signals, 20% central and 12% distal (Table 5.5). No cells recorded distal signals when the size of the fragile site was greater than the width of a chromatid.

One endprobe was isolated from either end of 60h5, the T7 endprobe representing 1.8 kb of terminal genomic sequence and the T3 endprobe representing 2.3 kb of genomic DNA (Figure 5.8). In situ hybridization of the T7 end resulted in a total of 22 cells being scored. 78% of the cells expressed proximal signal, 22% were central and no distal signals were observed (Table 5.5). *In situ* hybridization of 25 cells with the T3 end produced 76% of the cells expressing proximal signal, 20% central and 4% distal (Table 5.5). These results indicated that of the three overlapping cosmids, 60h5 may be located a greater distance away from FRA16B than 45b11 and 57b6.

Overall, the *in situ* hybridization of T7 and T3 endprobes for 57b6, 45b11 and 60h5
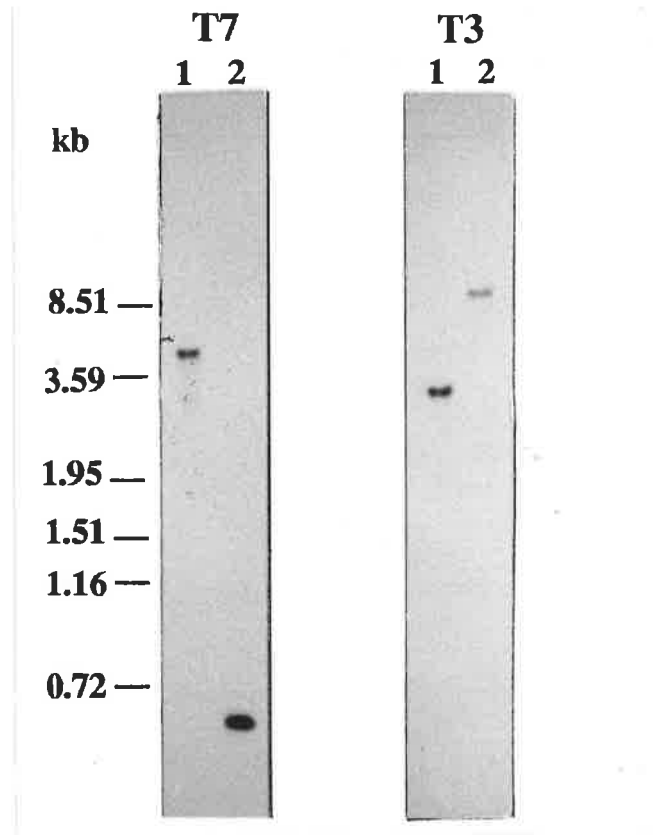
Fig. 5.7 Isolation of T7 and T3 endprobes from 45b11. (A) Radiolabelled T7 and T3 oligomers were hybridized to Southern blot filters of cosmid 45b11 DNA digested with *Pst* I (lane 1) and *Xmn* I (lane 2). Molecular weight DNA markers are shown (in kb) on the left side of the gel. The sizes of genomic insert DNA present in each endprobe are shown in (B) and the restriction map of the insert at the T7 and T3 ends of the cosmid are shown in (C).

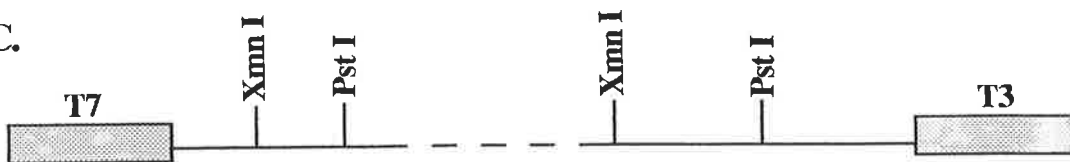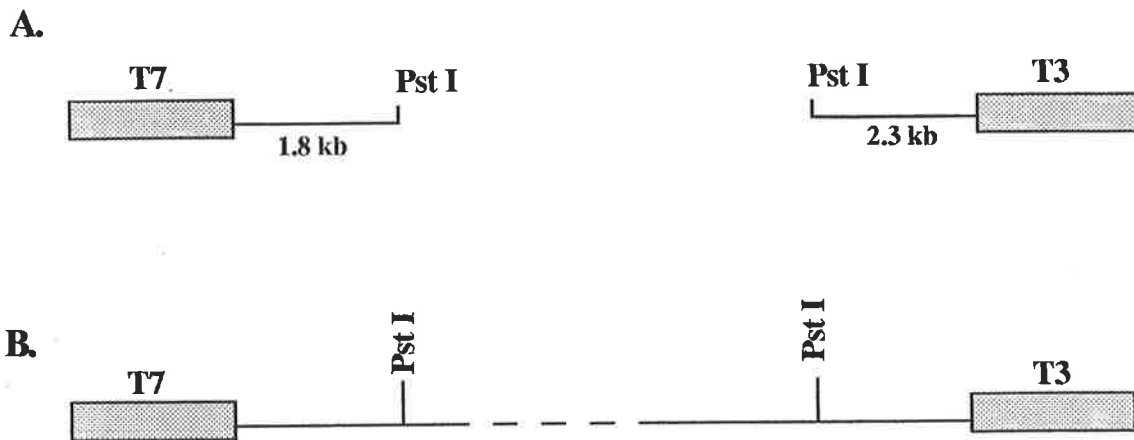**Fig. 5.8** Isolation of T7 and T3 endprobes from 60h5. Radiolabelled T7 and T3 oligomers were hybridized to Southern blot filters of cosmid 60h5 DNA digested with *Pst* I. The sizes of genomic insert DNA present in each endprobe are shown in (A) and the restriction map of the insert at the T7 and T3 ends of the cosmid are shown in (B).

yielded predominantly proximal signals to FRA16B, suggesting that these three cosmids did not span the fragile site.

5.5.4 Hybridization Fragments of Genomic DNA from Normal and FRA16B Expressing Individuals for a 57b6 - T7 Derived Endprobe.

In an attempt to determine whether the T7 endprobes derived from 57b6 actually extended into FRA16B a 3.8kb *Xmn* I derived T7 endprobe from 57b6 was used as a probe to compare *Bam* HI and *Pst* I restriction fragments in human genomic DNA from normal individuals and individuals expressing FRA16B. The DNA from 12 normal individuals and the DNA from three unrelated individuals expressing FRA16B were included in the analysis. The 57b6-T7 endprobe hybridized to a single 2.3 kb fragment in *Bam* HI digested genomic DNA samples and to 4.3 kb and 3.1 kb fragments in *Pst* I digested genomic DNA. Two bands representing adjacent *Pst* I fragments were observed for *Pst* I digested DNA samples because the endprobe contained a *Pst* I restriction enzyme site. No variation in hybridization fragments were observed between normal individuals and individuals expressing FRA16B (Figure 5.9).

5.5.5 Cosmid "Walking" : Screening a Cosmid Library with a 57b6-T7 Derived Endprobe.

As a result of the cosmids not forming a contig and thus FRA16B having not been cloned, a cosmid walk was initiated from the closest proximal endprobe to FRA16B in an attempt to isolate overlapping cosmids that may span the fragile site. The screening of high density cosmid grids with the 3.8 kb *Xmn* I derived T7 endprobe from 57b6 resulted in the hybridization of the probe to one clone, 328F12 (data not shown). A faint hybridization signal was also observed for 57b6. 328F12 had previously been established in contig 9, located between 57b6 and 45b11. A revised
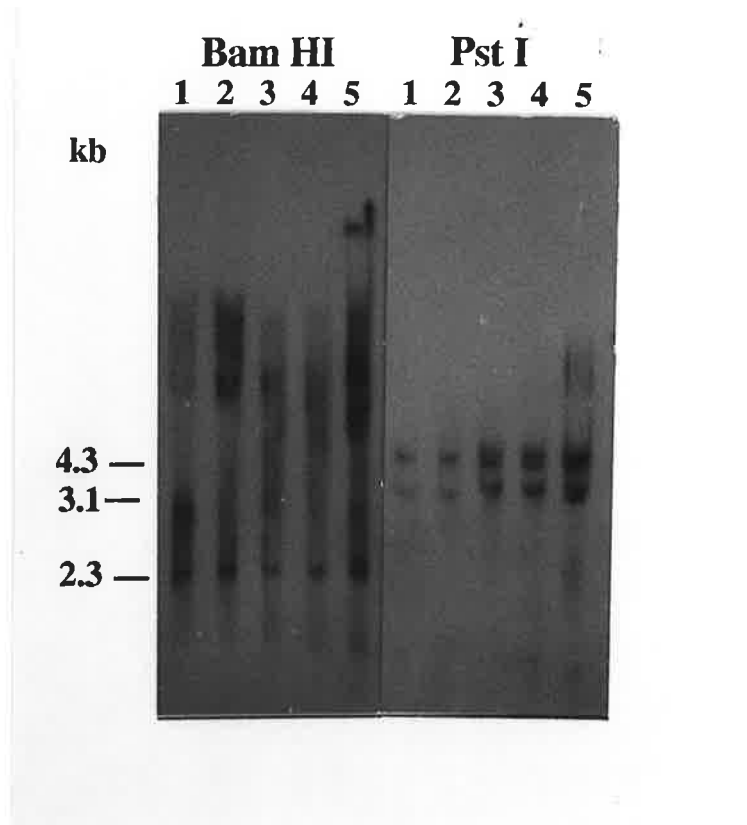
Fig. 5.9 Comparison of hybridization fragments generated by a 57b6 T7 derived endprobe, between DNA from normal individuals and individuals expressing FRA16B. A radiolabelled 3.8 kb *Xmn* I derived T7 endprobe from 57b6 was hybridized to a Southern blot filter containing *Bam* HI and *Pst* I digested lymphoblastoid cell line DNA from normal individuals (lanes 1 and 2) and individuals that expressed FRA16B (lanes 3, 4 and 5). The fragment sizes are shown (in kb) on the left side of the gel.

order for the overlapping cosmids in contig 9 generated by GCAA analysis accompanied the 328F12 clone from Los Alamos. 328F12 was placed as a terminal cosmid in the contig, followed by 51F3 and then 57b6 (Figure 5.10). Large overlaps between all three cosmids were indicated.

5.5.6 Analysis of 328F12 and T7 and T3 Derived Endprobes by *In situ* Hybridization (performed by E. Baker and H. Eyre).

Total 328F12 cosmid DNA and endprobes isolated from the cosmid were analysed by *in situ* hybridization to FRA16B chromosomes in order to determine the localization of the cosmid with respect to the fragile site, and to determine whether either end of the cosmid spanned FRA16B. The *in situ* hybridization scores for 328F12 endprobes are shown on Table 5.6.

In situ hybridization of whole 328F12 to chromosomes expressing FRA16B localized the cosmid proximal to this fragile site. A total of 70 cells were scored with 62% of the signals hybridizing proximal to FRA16B, 15% central and 23% distal (Table 5.6). Of the 17 cells that expressed fragile sites greater than the width of a chromatid, 18% exhibited distal signals when probed with 328F12.

Three endprobes were isolated from the T7 end of 328F12, representing 0.6 kb, 1.1 kb and 1.8 kb of terminal genomic sequence (Figure 5.11). 46 cells were scored for *in situ* hybridization of two of the three T7 endprobes to chromosomes expressing FRA16B. 70% of the cells exhibited proximal signal, 15% were central and 15% distal signals were scored.

Two endprobes were isolated from the T3 end of 328F12, representing 2.1 kb and 5.2 kb of terminal genomic sequence (Figure 5.11). 42 cells were scored, with 57% of the cells producing proximal signal, 17% central and 26% distal (Table 5.6). Neither endprobe scored predominantly distal signal to FRA16B indicating that this

120 kb

(9) 328F12

(9) 51F3

(9) 57b6

(9) 45b11

(9) 60h5

(9) 10c6
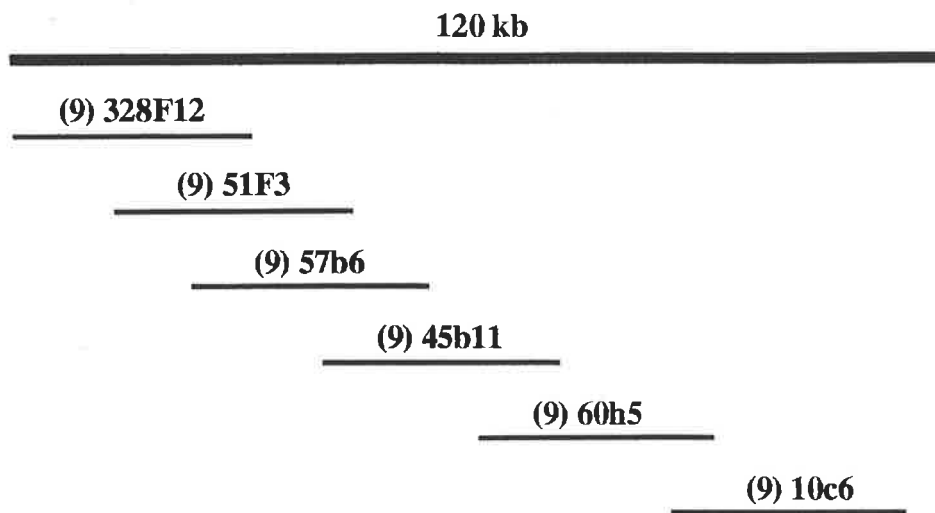
Fig. 5.10 The updated GCAA-derived map of contig. 9. The solid bars represent individual cosmids and depict their predicted overlaps.

Table 5.6 *In situ* Hybridization of 328F12 and T7 and T3 Endprobes to FRA16B Chromosomes.

Cosmid End	Endprobe	Size of fragile site	*In situ* Hybridization			
			Proximal(P)	Central(C)	Distal(D)	Other Signals
Whole Cosmid 328F12	- - -	⟨width of chromatid ⟩width of chromatid	31 12	9 2	13 3	
T7 (328F12)	*Pst* I *Xmn* I	⟨width of chromatid ⟩width of chromatid ⟨width of chromatid	10 1 21	1 1 5	2 0 5	
T3 (328F12)	*Pst* I *Xmn* I	⟩width of chromatid ⟩width of chromatid ⟨width of chromatid	12 4 8	4 1 2	5 1 5	P + D:2

Fig. 5.11 Isolation of T7 and T3 endprobes from 328F12. (A) Radiolabelled T7 and T3 oligomers were successively hybridized to a Southern blot filter of cosmid 328F12 DNA digested with *Pst* I (lane 1), *Xmn* I (lane 2) and *Pvu* II (lane 3). Molecular weight DNA markers are shown (in kb) on the left side of the gel. Hybridization fragments for the T3 oligomer are marked with arrows. The sizes of genomic insert DNA present in each endprobe are shown in (B) and the restriction map of the insert at the T7 and T3 end of the cosmid is shown in (C).

cosmid was unlikely to span FRA16B. Similar *in situ* hybridization signals observed by 328F12 to 57b6 and 45b11 supported the localization of 328F12 to contig 9.

5.5.7 The Integrity of the Overlap between 57b6 and 328F12.

In attempts to confirm the GCAA predicted overlap between 57b6 and 328F12, the endprobe that was used to recover 328F12 from the cosmid library screening (i.e. the 3.8 kb *Xmn* I derived endprobe from the T7 end of 57b6) was probed against Southern blot filters containing *Pst* I digested 328F12 and 57b6. Confusing results were obtained. Hybridization of the T7 endprobe to 57b6 produced two fragments of sizes 3.1 and 5.0 kb (Figure 5.12). These two bands represented adjacent *Pst* I fragments on 57b6 because the 57b6 *Xmn* I derived T7 endprobe contained a *Pst* I restriction site. The 5.0 kb band was the T7 terminal *Pst* I fragment of 57b6 (the T7 RNA promotor hybridized to the same sized fragment, see Figure 5.5), and the 3.1 kb *Pst* I fragment was adjacent to it. A 3.1 kb *Pst* I fragment was also observed when the same T7 endprobe was hybridized to Southern blot filters containing genomic DNA (Figure 5.9). Hybridization of the 57b6 T7 endprobe to *Pst* I digested 328F12 produced only one band of a size 3.8 kb (Figure 5.12). Two circumstances were proposed whereby this may occur (Figure 5.13).

In situation A (Figure 5.13A), only a small overlap must exist between 328F12 and 57b6 because 328F12 doesn't hybridize to the 3.1 kb *Pst* I fragment of 57b6. The 57b6 T7 endprobe hybridizes to only one internal *Pst* I fragment of 328F12 but not to the terminal *Pst* I fragment. If the terminal *Pst* I fragment were small, for example the T7 end *Pst* I fragment of 328F12 is only 0.6 kb (Figure 5.11) and contained repetitive DNA, then the prereassociated T7 probe may not hybridize to it. However, the T7 *Pst* I endprobe for 328F12 produced signals for *in situ* hybridization so it is unlikely to contain large amounts of repetitive DNA. In addition, the
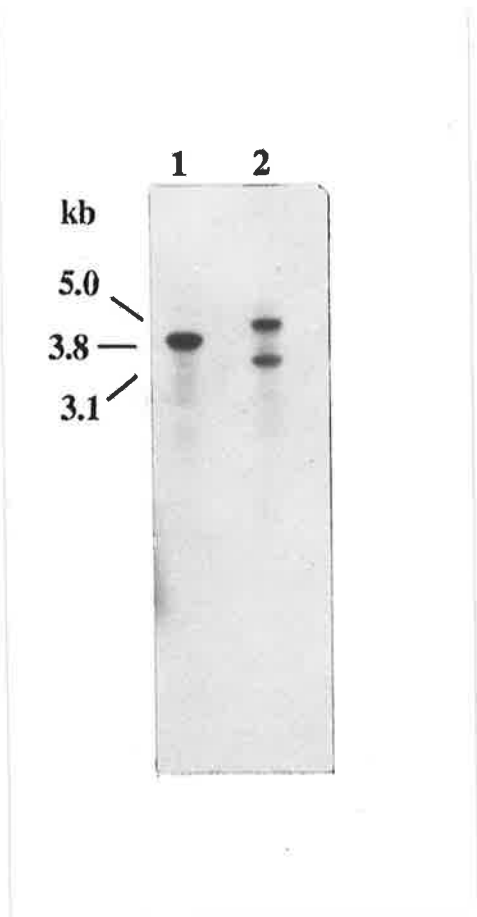
Fig. 5.12 Hybridization fragments of cosmid DNA for a 57b6 T7 derived endprobe.

The 3.8 kb *Xmn* I derived endprobe from the T7 end of 57b6 was hybridized to a

Southern blot filter containing *Pst* I digested DNA from cosmid 328F12 (lane 1) and

57b6 (lane 2). The size of hybridization fragments are shown (in kb) on the left side

of the gel.

A.
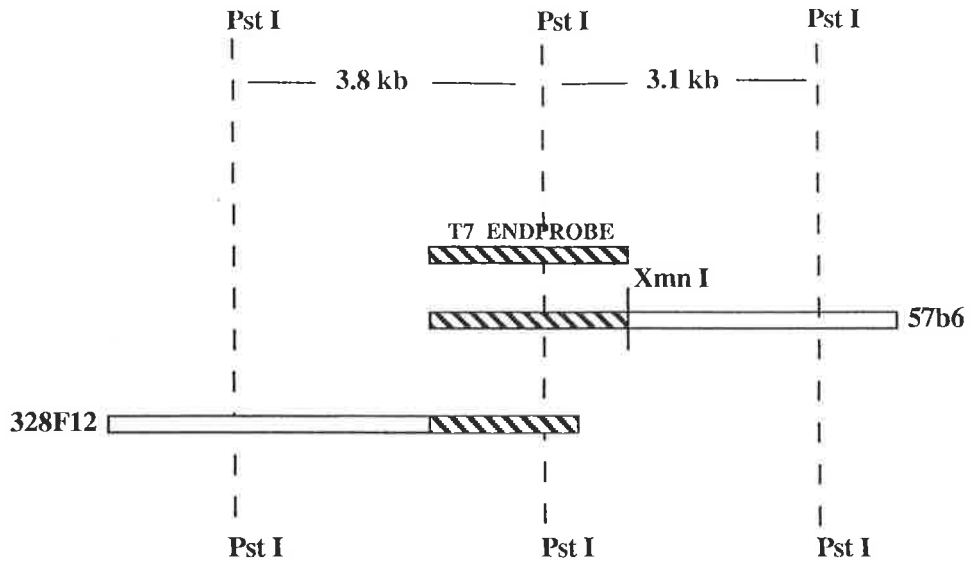
B.
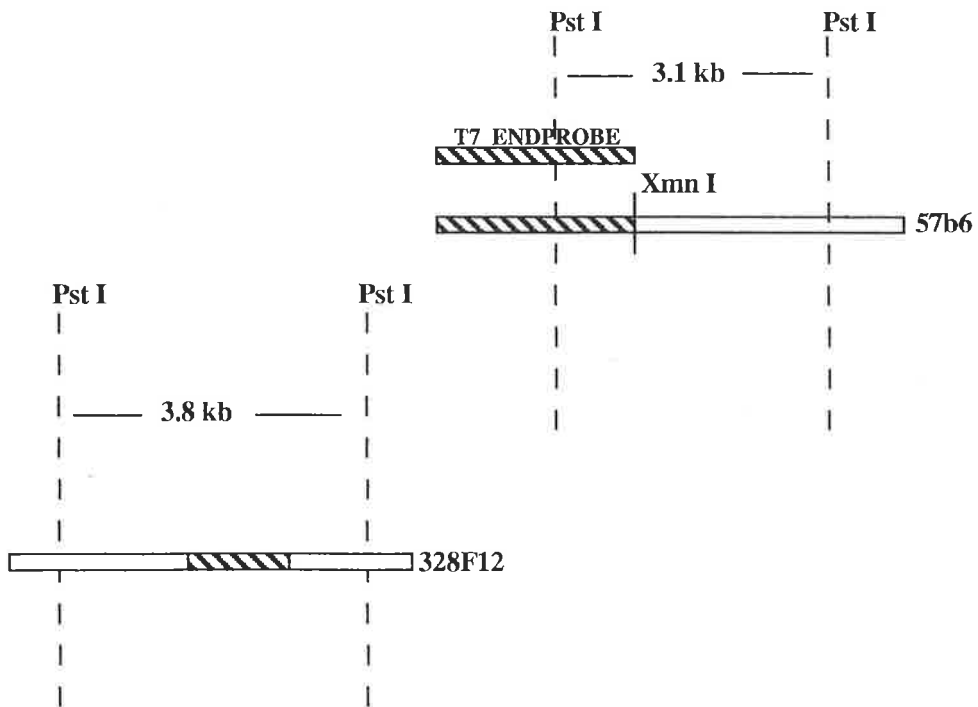
Fig. 5.13 Two situations (A) and (B) whereby hybridization of a 57b6 T7 derived endprobe to *Pst* I digested cosmid 328F12 DNA produced one band of a size 3.8 kb. The hatched regions indicate hybridization of the T7 endprobe.

hybridization of the 57b6 T7 endprobe to a 3.8 kb *Pst* I fragment on 328F12 does not correspond with the data obtained from genomic DNA in which the adjacent fragment is 4.3 kb in size (Figure 5.9).

In situation B (Figure 5.13B), the two cosmids do not overlap. A region of DNA within the 3.1 kb *Pst* I fragment of 57b6 has homology within a 3.8 kb *Pst* I fragment of 328F12. Conserved *Eco* R1 and *Hind* III restriction enzyme sites within this region of DNA account for the overlap being established by GCAA analysis. If this proposal is true, then also by chance, 328F12 is located proximal and very close to FRA16B as the *in situ* hybridization results indicated. In addition, a large overlap was predicted between 328F12 and 57b6 which is not consistent with a small region of conserved *Eco* R1 and *Hind* III sites between the two cosmids. A maximum of only 0.8 kb of DNA could be homologous between 328F12 and 57b6 (0.8 kb is the distance from the *Pst* I site to the *Xmn* I site on 57b6).

The results neither support nor dismiss the circumstances described whereby the 57b6 T7 endprobe could hybridize to only one band on 328F12. It was therefore not possible to establish the integrity of the overlap between 328F12 and 57b6.

5.6   DISCUSSION

The analysis of overlapping sets of cosmid clones or contigs on human chromosome 16 have successfully led to the isolation of a cosmid contig containing the candidate sequences for the PKD1 gene at 16p13.3 (Gillespie *et al* 1990), the isolation of the chromosome 16 microdeletion region in Rubenstein-Taybi syndrome (Breuning *et al* 1992) and the closure of the metallothionein gene family contig map at 16q13 (West *et al* 1990).   The initial *in situ* hybridization analysis of a 190 kb cosmid contig mapping to the same physical interval as FRA16B indicated that the cosmid cloning of FRA16B may have been achieved.   However, the finer localization of individual cosmids by hybrid cell line panel analysis revealed that the cosmids did not form a contiguous series and that the GCAA algorithm that linked the cosmids had established a false overlap.   The original 190 kb contig was comprised of one series of overlapping cosmids including 10c6, 60h5, 45b11 and 57b6 called contig 9 and a second series of cosmids including 68g12, 40e6 and 14d9 called contig 70. Representative cosmids from contig 9 had been finely mapped in the same physical interval as FRA16B, however hybrid cell line panel analysis established that cosmids from contig 70 were localized many megabases distal to FRA16B.   The GCAA algorithm had predicted that the two contigs (9 and 70) representing significantly overlapping cosmids were linked at 57b6 and 68g12 by a very small overlap.   The results from this section of the project indicated that this link was a false overlap. The integrity of the overlap between these two cosmids was critical to the spanning of FRA16B by this contig because by *in situ* hybridization 57b6 represented the proximal marker to FRA16B and 68g12 the distal marker.   A break at this junction indicated that FRA16B was unlikely to have been cloned within the contig.   Further confirmation of the false overlap was provided when an updated and refined version

of the GCAA-derived map of the contig indicated that no overlap existed between 57b6 and 68g12. The initial algorithm used to identify overlaps by repetitive sequence fingerprinting (Stallings *et al* 1990) may not have used stringent enough criteria which resulted in false positives. The presence of low abundance repetitive DNA sequences have also been found to lead to false overlaps in cosmid contig construction (Stallings *et al* 1992b). As described in Chapter 3, human repetitive DNA sequences have been isolated that give specific hybridization to chromosome 16 (Moyzis *et al* 1987), and in particular low abundance repeat DNA have been identified to chromosome bands p13, p12, p11 and q22 (Stallings *et al* 1992b). The low abundance repeats can interefere with contig construction when using the fingerprinting based approaches if they contain restriction sites for the enzymes employed in the fingerprinting strategy.

Although the likelihood of FRA16B having been cloned in one of the cosmids was low, the *in situ* hybridization of whole cosmids to FRA16B chromosomes identified three cosmids 60h5, 45b11 and 57b6 as potential candidates for encompassing FRA16B. The isolation and *in situ* hybridization of endprobes generated from these cosmids to FRA16B chromosomes did not identify any cosmids that spanned the fragile site. All probes localized the three cosmids proximal to FRA16B. However, the T7 endprobes from 57b6 were targeted as the closest proximal markers to this fragile site, possibly extending into it. The distribution of signals across FRA16B by the 57b6-T7 derived endprobes were similar to those observed by lambda subclones isolated from the YAC (XTY-26) that was found to span FRAXA (Kremer *et al* 1991a). The subsequent Southern blot analysis of one of these subclones, lambda subclone 5, to DNA isolated from carriers of FRAXA exhibited a variable DNA region associated with FRAXA (Yu *et al* 1991). However, in this project, Southern

blot analysis of genomic DNA from carriers expressing FRA16B with a 57b6 T7 derived endprobe did not identify any variable DNA regions.

The identification of a false overlap and the proximal localization of the cosmids that were potential candidates for containing FRA16B indicated that the fragile site had not been cloned. As a result, a modification to the original strategy was required. A cosmid "walk" was initiated from the closest proximal marker to FRA16B, in an attempt to detect overlapping cosmid clones which may extend across the fragile site. The screening of an ordered cosmid library detected one overlapping clone 328F12, already a part of contig 9. The updated GCAA derived contig map had localized 328F12 at the terminal end of contig 9, exhibiting a large overlap with 57b6. *In situ* hybridization of whole 328F12 DNA and T7 and T3 endprobes to FRA16B chromosomes supported the GCAA derived map by localizing the cosmid proximal to FRA16B. The results also indicated that this cosmid did not span the fragile site. In an attempt to verify the integrity of the overlap between 57b6 and 328F12, a 57b6 T7 derived endprobe was used to compare hybridization fragments in the cloned DNA of 57b6 and 328F12 with that of genomic DNA. The hybridization fragments observed for cloned DNA did not correlate with the genomic DNA fragments inabling the physical verification of the overlap. My involvement with the cosmid walk was suspended at this stage.

The isolation of additional overlapping clones to "walk" towards FRA16B need not be limited to cosmid libraries where the expansion of the contig is restricted to cosmid sized increments. For example, the screening of a YAC library would isolate clones containing longer segments of DNA, minimalizing the number of steps required to span FRA16B. Sequence data is being generated by other investigators from the 57b6-T7 derived endprobes and from 328F12 to construct primers for a

PCR based screening of a YAC library.

CHAPTER 6

CONCLUDING REMARKS

At the commencement of this study in 1989, no fragile site had been characterised at the molecular level. In its essence, this project involved the analysis of large DNA fragments in the vicinity of the rare distamycin A inducible fragile site FRA16B with aims that the cloning and characterisation of the fragile site may be achieved. Interestingly, this study was undertaken at a time when the technology involving the analysis of large DNA fragments was being rapidly developed. As a direct consequence of these advancements, two fragile sites FRAXA and FRAXE were characterised at the molecular level during this project's course (Yu *et al* 1991, Kremer *et al* 1991b, Knight *et al* 1993). Subsequently, the methods that led to their successful cloning were introduced to the analysis of DNA in the region near FRA16B, in attempts to facilitate the rapid cloning of this fragile site.

The initial aim of the project was the construction of a long range restriction map of the region 16q13-16q22.1 which included FRA16B localized to the 16q21-16q22.1 interface (Callen *et al*, 1988). At the time, analysis of large DNA fragments was limited to uncloned genomic DNA. PFGE mapping of this region (Chapter 3) did not produce a contiguous map, therefore the linear order of markers spanning FRA16B and the physical distances separating them were not established.

The decision to change mapping strategies in the region surrounding FRA16B was spurred by the cloning of FRAXA in YACs (Kremer *et al* 1991a, Heitz *et al* 1991, Dietrich *et al* 1991). In addition, restriction mapping with YACs circumvented many of the difficulties that were encountered with the mapping of uncloned regions of DNA. Bidirectional "walking" with YACs was initiated from markers flanking FRA16B (Chapter 4). The YACs recovered from the first round of screening of a human genomic YAC library did not overlap or span FRA16B. However, the isolation of additional overlapping YACs will eventually form a YAC contig

physically linking the flanking markers resulting in the cloning of the fragile site. The utilization of a recently introduced YAC library containing inserts up to 1000 kb in size (Bellanne - Chantelot *et al* 1992) may significantly reduce the number of YACs required to create such a contig.

The initial physical mapping of a 190 kb cosmid contig that was preliminarily localized across FRA16B relied upon fluorescent *in situ* hybridization (FISH) to identify the position of probes on metaphase chromosomes expressing FRA16B (Chapter 5). The resolution of this mapping approach was unable to establish that the cosmid probes did not form a contiguous series and that FRA16B was unlikely to have been cloned. Recently, FISH has been applied to interphase cell nuclei for higher resolution mapping. The DNA is less condensed in interphase cell nuclei than in metaphase chromosomes and resolution in the 50-100 kb range has been obtained (Lawrence *et al* 1990, Trask *et al* 1991). In addition, the recent adaption of FISH to include multicoloured labelling and detection of DNA probes (reviewed by Smith *et al* 1992) has further developed the ordering of probes along the chromosome and the measurement of their distance apart (both in metaphase and interphase). For example, Reid *et al* (1992) have demonstrated the simultaneous visualisation of six different DNA probes to map six cosmids clones in a single hybridization experiment. The difficulties experienced in mapping the distance between two probes by interphase mapping has been recently addressed by Parra and Windle (1993). They have described a procedure to map decondensed nuclear DNA by a means of forcibly stretching DNA in a linear fashion. When FISH techniques were used, a high resolution multicoloured map was obtained. Applying the methods of interphase mapping or DNA stretching with multicoloured FISH to the future analysis of cosmid contigs in the region of FRA16B may provide a more accurate method for the

establishment of contig integrity and the ordering of cosmids along the chromosome. These mapping strategies may also be extended to determine the orientation and overlap of YAC sequences. Possible deletions, insertions or rearrangements in the genome or cloned sequences may also be identified by these methods.

The molecular characterisation of the folate sensitive fragile site <u>FRAXA</u> (Yu *et al* 1991, Kremer *et al* 1991b) and more recently <u>FRAXE</u> (Knight *et al* 1993) revealed a new mechanism of genetic mutation in humans - the amplification of a trinucleotide repeat. This new class of human mutation has also been identified in Kennedy disease (La Spada *et al* 1991) and myotonic dystrophy, DM (Aslanidis *et al* 1992, Buxton *et al* 1992, Brook *et al* 1992, Fu *et al* 1992, Mahadevan *et al* 1992). The molecular basis of these two disorders have been determined as an amplification of a $p(AGC)_n$ repeat sequence (La Spada *et al* 1991, Brook *et al* 1992, Fu *et al* 1992). The triplet amplifications observed for <u>FRAXA</u>, <u>FRAXE</u> and DM exhibit a genetic phenomenon termed 'anticipation' (Kremer *et al* 1991b, Harley *et al* 1992, Buxton *et al* 1992, Knight *et al* 1993). Anticipation can be defined as the appearance of increasing disease severity or earlier onset in successive generations with a heritable disorder (Caskey *et al* 1992). The mechanisms of these amplifications however remain unknown.

The discovery of the mutational mechanism of triplet repeat amplification for the two folate sensitive fragile sites, DM and Kennedy disease may lead to the discovery of additional fragile sites and disease related genes. In fact, synthetic oligonucleotides containing triplet sequences (there are 10 possible trinucleotide repeats) were used in a scanning strategy to pinpoint the causative mutation for DM (Fu *et al* 1992). Similar strategies may also be applied in attempts to identify the <u>FRA16B</u> mutation. However triplet sequences are common in the human genome (Caskey *et al* 1992)

and amplification of trinucleotide repeats are not solely the domain of fragile sites. Furthermore, the mode of induction of FRA16B differs from that of FRAXA and FRAXE indicating that the molecular nature of distamycin A inducible fragile sites may also differ. The cloning of fragile sites other than the folate sensitive class will allow a better understanding of the molecular basis of these chromosomal mutations.

References

Abidi, F.E., Wada, M., Little, R.D., and Schlessinger, D. (1990).
Yeast artificial chromosomes containing human Xq24-Xq28
DNA: library construction and representation of probe sequences.
Genomics 7 : 363-376.

Abruzzo, M., Mayer, M., and Jacobs, P.A. (1985).
The effects of methionine and 5-azacytidine on fragile X expression.
Am. J. Hum. Genet. 37 : 193-198.

Abruzzo, M.A., Pettay, D., Mayer, M. and Jacobs, P.A. (1986).
The effect of caffeine on fragile X expression.
Hum. Genet. 73 : 20-22.

Albertsen, H.M., LePaslier, D., Abderrahim, H., Dausset, J., Cann, H., and
Cohen, D. (1989).
Improved control of partial DNA restriction enzyme digest in agarose using limited
concentrations of Mg2+.
Nucl. Acids Res. 17 : 808.

Anand, R., Ogilvie, D.J., Butler, R., Riley, J.H., Finniear, R.S., Powell, S.J.,
Smith, J.C. and Markham, A.F. (1991).
A yeast artificial chromosome contig encompassing the cystic fibrosis locus.
Genomics 9 : 124-130.

Anderson, C. (1993).
Genome shortcut leads to problems.
Science 259 : 1684-1687.

Arthur, D.C., and Bloomfield, C.D. (1983).
Partial deletion of the long arm of chromosome 16 and bone marrow eosinophilia in
acute non lymphocytic leukemia: A new association.
Blood 61 : 994-998.

Aslandis, C., Jansen, G., Amemiya, C., Shutter, G., Mahadevan, M., Tsilfidis, C.,
Chen, C., Alleman, J., Wormskamp, N., Vooijs, M., Buxton, J., Johnson, K.,
Smeets, H., Lennon, G., Carrano, A., Korneluk, R., Wieringa, B. and de Jong, P.
(1992).
Cloning of the essential myotonic dystrophy region and mapping of the putative
defect.
Nature 355 : 548-551.

Aula, P., and von Koskull, H. (1976).
Distribution of spontaneous chromosome breaks in human chromosomes.
Hum. Genet. 32 : 143-148.

Ayme, S., Mattei, J.F., Mattei, M.G., Aurran, Y., and Giraud, F. (1976).
Non random distribution of chromosome breaks in cultured lymphocytes of normal subjects.
Hum. Genet. 31 : 161-175.

Baker, E. and Sutherland, G.R. (1991).
A new folate sensitive fragile site at 1p 21.3.
J. Med. Genet. 28 : 356-357.

Balding, D.J., and Torney, D.C. (1991).
Statistical analysis of DNA fingerprint data for ordered clone physical mapping of human chromosomes.
Bull. Math. Biol. 53 : 853-879.

Bancroft, I., and Wolk, C.P. (1988).
Pulsed homogeneous orthogonal field gel electrophoressis (PHOGE).
Nucl. Acids Res. 16 : 7405-7418.

Barbi, G., Steinbach, P., and Vogel, W. (1984).
Non random distribution of methotrexate-induced aberrations on human chromosomes. Detection of further folic acid sensitive fragile sites.
Hum. Genet. 68 : 290-294.

Barlow, D.P. and Lehrach, H. (1987).
Genetics by Gel Electrophoresis: The impact of pulsed field gel electrophoresis on mammalian genetics.
Trends Genet. 3 : 167-171.

Bell, M.V., Hirst, M.C., Nakahori, Y., MacKinnon, R.N., Roche, A., Flint, T.J., Jacobs, P.A., Tommerup, N., Tranebjaerg, L., Froster-Iskenius, U., Kerr, B., Turner, G., Lindenbaum, R.H., Winter, R., Pembrey, M., Thibodeau, S. and Davies, K.E. (1991).
Physical mapping across the fragile X : Hypermethylation and clinical expression of the fragile X syndrome.
Cell 64 : 861-866.

Bellanne-Chantelot, C., Barillot, E., Lacroix, B., LePaslier, D. and Cohen, D. (1991).
A test case for physical mapping of human genome by repetitive sequence fingerprints : construction of a physical map of a 420 kb YAC subcloned into cosmids.
Nucl. Acids Res. 19 : 505-510.

Bellanne-Chantelot, C., Lacroix, B., Ougen, P., Billault, A., Beaufild, S., Bertrand, S., Georges, I., Gilbert, F., Gros, I., Lucotte, G., Susisn, L., Codani, J., Gesnovin, P., Pook, S., Vaysseix, G., Lu-Kuo, J., Reid, T., Ward, D., Chumakov, I., LePaslier, D., Barillot, E., and Cohen, D. (1992).
Mapping the whole human genome by fingerprinting yeast artificial chromosomes.
Cell 70 : 1059-1068.

Benton, W.D., and Davis, R.W. (1977).
Screening lambda gt recombinant clones by hybridization to single plaques *in situ*.
Science 196 : 180-182.

Berger, R., Bloomfield, C.D., and Sutherland, G.R. (1985).
Report of the committee on chromosome rearrangements in neoplasia and on fragile sites.
Cytogenet. Cell Genet. 40 : 490-535.

Bird, A.P. (1986).
CpG-rich islands and the function of DNA methylation.
Nature (London) 321 : 209-213.

Birnboim, H.C., and Doly, J. (1979).
A rapid alkaline extraction procedure for screening recombinant plasmid DNA.
Nucl. Acid Res. 7 : 1513-1522.

Breukel, C., Wijnen, J., Tops, C., Klift, H., Dauwerse, H. and Khan, P.M. (1990).
Vector-alu PCR : a rapid step in mapping cosmids and YACs.
Nucl. Acids Res. 18 : 3097.

Breuning, M.H., Dauwerse, H.G., Fugazza, G., Saris, J.J., Spruit, L., Wijnen, H., Tommerup, N., van der Boogaard, M.J., dePater, J.M., Mariman, E., Hamel, B., Himmelbauer, H., Frischauf, A.M., Stallings, R.L., van Ommen, G. and Hennekam, R.C.M. (1992).
Submicroscopic deletions of chromosome 16 in patients with Rubenstein-Taybi Syndrome.
Submitted for publication.

Britten, R.J., Baron, W.F., Stout, D.B. and Davidson, E.H. (1988).
Source and evaluation of human *alu* repeated sequences.
Proc. Natl. Acad. Sci. USA 85 : 4770-4774.

Brook, D.J., McCurrach, M.E., Harley, H., Buckler, A.J., Church, D., Aburatani, H., Hunter, K., Stanton, V.P., Thirion, J.P., Hudson, T., Sohn, R., Zemelman, B., Snell, R.G., Rundle, S.A., Crow, S., Davies, J., Shellbourne, Buxton, J., Jones, C., Juvonen, V., Johnson, K., Harper, P., Shaw, D., and Houseman, D. (1992).
Molecular Basis of myotonic dystrophy : Expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member.
Cell 68 : 799-808.

Brookwell, R. and Turner, G. (1983).
High resolution banding and the locus of the Xq fragile site.
Hum. Genet. 63 : 77.

Brown, W.R.A. and Bird, A.P. (1986).
Long range restriction site mapping in mammalian genomic DNA.
Nature 322 : 477-481.

Brown, W.T., Ye, W., Gross, A.C., Chan, C.B., Dobkin, C.S. and Jenkins, E.C. (1988).
Multipoint linkage of 9 anonymous probes to HPRT, Factor 9, and fragile X.
Am. J. Med. Genet. 30 : 551-566.

Burke, D.T., Carle, G.F. and Olson, M.V. (1987).
Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors.
Science 236 : 806-812.

Burmeister, M., Kim, S., Roydon Price, E., DeLange, T., Tantravahi, U., Myers, R.M. and Cox, D.R. (1991).
A map of the distal region of the long arm of human chromosome 21 constructed by radiation hybrid mapping and pulsed field electrophoresis.
Genomics 9 : 19-30.

Butler, R., Ogilvie, D.J., Elvin, P., Riley, J.H., Finniear, R.S., Slynn, G., Morten, J.E.N., Markham, A.F., and Anand, R. (1992).
Walking, cloning and mapping with yeast artificial chromosomes : A contig encompassing D21S13 and D21S16.
Genomics 12 : 42-51.

Buxton, J., Shellbourne, P., Davies, J., Jones, C., Van Tongeren, T., Aslanidis, C., deJong, P., Jansen, G., Anvret, M., Riley, B., Williamson, R., and Johnson, K. (1992).
Detection of an unstable fragment of DNA specific to individuals with myotonic dystrophy.
Nature 355 : 547-551.

Callen, D.F. (1986).
A mouse - human hybrid cell panel for mapping human chromosome 16.
Ann. Genet. 29 : 235-239.

Callen, D.F., Hyland, V.J., Baker, E.G., Fratini, A., Simmers, R.N., Mulley, J.C. and Sutherland, G.R. (1988).
Fine mapping of gene probes and anonymous DNA fragments to the long arm of chromosome 16.
Genomics 2 : 144-153.

Callen, D.F., Hyland, V.J., Baker, E.G., Fratini, A., Gedeon, A.K., Mulley, J.C., Fernandez, K.E., Breuning, M.H. and Sutherland, G.R. (1989).
Mapping the short arm of human chromosome 16.
Genomics 4 : 348-354.

Callen, D.F., Baker, E., Eyre, H.J., and Lane, S.A. (1990a).
An expanded mouse-human hybrid cell panel for mapping human chromosome 16.
Ann. Genet. 33 : 190-195.

Callen, D.F., Baker, E., Eyre, H.J., Chernos, J.E., Bell, J.A., and Sutherland, G.R. (1990b).
Reassessment of two apparent deletions of chromosome 16p to an ins (11;16) and a t(1;16) by chromosome painting.
Ann. Genet. 33 : 219-221.

Camerino, G., Mattei, M.G., Mattei, J.F., Jaye, M. and Mandel, J.L. (1983).
Close linkage of fragile X-mental retardation syndrome to haemophilia B and transmission through a normal male.
Nature 306 : 701-704.

Carle, G.F., and Olsen, M.V. (1984).
Separation of chromosomal DNA molecules from yeast by orthogonal field alteration gel electrophoresis.
Nucleic. Acids Res. 12 : 5647-5664.

Carle, G.F., and Olson, M.V. (1985).
An electrophoretic karyotype for yeast.
Proc. Natl. Acad. Sci. U.S.A. 82 : 3756-3760.

Carle, G.F., Frank, M., and Olsen, M.V. (1986).
Electrophoretic separation of large DNA molecules by periodic inversion of the electric field.
Science 232 : 65-68.

Caskey, C.T., Pizzuti, A., Fu, Y.H., Fenwick, R.G., and Nelson, D.L. (1992).
Triplet repeat mutation in human disease.
Science 256 : 784-789.

Chaconas, G., and van de Sande, J.H. (1980).
5'-^{32}P labelling of RNA and DNA restriction fragments.
Methods Enzymol 65 : 75-88.

Chen, L.Z., Harris, P.C., Apostolou, S., Baker, E., Holman, K., Lane, S.A., Nancarrow, J.K., Whitmore, S.A., Stallings, R.L., Hildebrand, C.E., Richards, R.I., Sutherland, G.R., and Callen, D.F. (1991).
A refined physical map of the long arm of human chromosome 16.
Genomics 10 : 308-312.

Chia, W., Scott, M.R.D., and Rigby, P.W.J. (1982).
The construction of cosmid libraries of eukaryotic DNA using the Homer series of vectors.
Nucl. Acids Res. 10 : 2503-2520.

Chu, G., Vollrath, D., and Davis, R.W. (1986).
Separation of large DNA molecules by contour - clamped homogeneous electric fields.
Science 234 : 1582-1585.

Chung, C.T., Niemela, S.L. and Miller, R.H. (1989).
One step preparation of competent Escherichia coli : Transformation and storage of bacterial cells in the same solution.
Proc. Natl. Acad. Sci. U.S.A. 86 : 2172-2175.

Clark, S.M., Lai, E., Birren, B.W. and Hood, L. (1988).
A novel instrument for separating large DNA molecules with pulsed homogeneous electric fields.
Science 241 : 1203-1205.

Collins, F.S. and Weissman, S.M. (1984).
Directional cloning of DNA fragments at a large distance from an initial probe : a circularization method.
Proc. Natl. Acad. Sci. U.S.A. 81 : 6812-6816.

Comings, D.E. (1978).
Mechanisms of chromosome banding and implications for chromosome structure.
Annu. Rev. Genet. 12 : 25-46.

Coulson, A., Waterston, R., Kiff, J., Sulston, J., and Kohara, Y. (1988).
Genome linking with yeast artificial chromosomes.
Nature 335 : 184-186.

Croci, G. (1983).
BrdU-sensitive fragile site on long arm of chromosome 16.
Am. J. Hum. Genet. 35 : 530-533.

Current Protocols in Molecular Biology.
Wiley Interscience, 1993, Canada.

Daniel, A. (1986).
Clinical implications and classification of the constitutive fragile sites.
Am. J. Med. Genet. 23 : 419-427.

Davidson, P.J., Jobsis, J., Davis, M., Wallis, S., and Humphries, S.E. (1987).
Identification of a gene on chromosome 16, which is related to that for apolipoprotein E (apo E) on chromosome 19.
Cytogenet. Cell Genet. 46 : 604-605.

Davies, K.E. (1986).
DNA studies of X-linked mental retardation associated with a fragile site at Xq27.
Am. J. Med. Genet. 23 : 633-642.

Dekaban, A. (1965).
Persisting clone of cells with an abnormal chromosome in a woman previously irradiated.
J. Nucl. Med. 6 : 740-746.

de la Chapelle, A., and Berger, R. (1983).
Seventh international workshop on human gene mapping. Report of the committee on chromosomal rearrangments in neoplasia and on fragile sites.
Cytogenet. Cell. Genet. 37 : 274-311.

Deutsch, J.M. (1988).
Theoretical studies of DNA during gel electrophoresis.
Science 240 : 922-924.

Dev, V.G., and Mixon, C. (1984).
5-azacytidine decreases the frequency of fragile X expression in peripheral blood lymphocyte culture.
Am. J. Med. Genet. 17 : 253-254.

Dietrich, A., Kioschis, P., Monaco, A., Gross, B., Korn, B., Williams, S., Sheer, D., Heitz, D., Oberle, I., Toniolo, D., Warren, S., Leharch, H., and Poustka, A. (1991).
Molecular cloning and analysis of the fragile X region in man.
Nucl. Acids. Res. 19 : 2567-2572.

Djalali, M., Barbi, G., and Steinbach, P. (1985).
Folic acid sensitive fragile sites are not limited to the human karyotype.
Demonstration of non random gaps and breaks in the Persian vole Ellobius lutescens Th. inducible by methotrexate, fluorodeoxyuridine and aphidicolin.
Hum. Genet. 70 : 183-185.

Djalali, M., Adolph, S., Steinbach, P., Winking, H. and Hameister, H. (1987).
A comparative mapping study of fragile sites in the human and murine genomes.
Hum. Genet. 77 : 157-162.

Djalali, M., Adolph, S., Steinbach, P., Winking, H. and Hameister, H. (1990).
Fragile sites induced by 5-azacytidine and 5-azadeoxycytidine in the murine genome.
Hereditas 112 : 77-81.

Drayna, D., Jarnagin, A.S., McLean, J., Henzel, W., Kohr, W., Fielding, C., and Lawn, R. (1987).
Cloning and sequencing human cholesteryl ester transfer protein cDNA.
Nature 327 : 632-634.

Ellis, T.H., Cleary, W.G., Burcham, K.W., and Bowen, B.A. (1987).
Ramped field inversion gel electrophoresis : a cautionary note.
Nucl. Acids. Res. 15 : 5489.

Evans, G.A., Lewis, K., and Rothenberg, B.E. (1989).
High efficiency vectors for cosmid microcloning and genomic analysis.
Gene 79 : 9-20.

Fangman, W. (1978).
Separation of very large DNA molecules by gel electrophoresis.
Nucl. Acids. Res. 5 : 653-665.

Feinberg, A.P. and Vogelstein, B. (1983).
A technique for radiolabelling DNA restriction fragments to high specific activity.
Analyt. Biochem. 132 : 6-13.

Fickett, J.W. and Cinkostsky, M.J. (1992).
A genetic algorithm for assembling chromosome physical maps.
Submitted for publication.

Fischel-Ghodsian, N., Nicholls, R.D., and Higgs, D.R. (1987).
Long range genome structure around the human X-globin complex analysed by PFGE.
Nucl. Acids Res. 15 : 6197-6201.

Flynn, G.A., Hirst, M.C., Knight, S.J.L., MacPherson, J.N., Barber, J.C.K., Flannery, A.V., Davies, K.E., and Buckle, V.J. (1993).
Identification of the FRAXE fragile site in two families ascertained for X-linked mental retardation.
J. Med. Genet. 30 : 97-103.

Foote, S., Vollrath, D., Hilton, A., and Page, D.C. (1992).
The human Y chromosome : overlapping DNA clones spanning the euchromatic region.
Science 258 : 60-66.

Fratini, A., Simmers, R.N., Callen, D.F., Hyland, V.J., Tischfield, J.A., Stambrook, P.J. and Sutherland, G.R. (1986).
A new location for the human adenine phosphoribosyl transferase gene (APRT) distal to the haptoglobin (HP) and fra(16)(q23) (FRA16D) loci.
Cytogenet. Cell. Genet. 43 : 10-13.

Fu, Y.H., Pizzuti, A., Fenwick, R.G., King, J., Rajnarayan, S., Dunne, P.W., Dubel, J., Nasser, G.A., Ashizawa, T., deJong, P., Wieringa, B., Korneluk, R., Perryman, M.B., Epstein, H.F., and Caskey, C.T. (1992).
An unstable triplet repeat in a gene related to myotonic muscular dystrophy.
Science 255 : 1256-1258.

Fulton, T.R., Bowcock, A.M., Smith, D.R., Daneshvar, L., Green, P., Cavalli-Sforza, L.L. and Donis Keller, H. (1989).
A 12 megabase restriction map at the cystic fibrosis locus.
Nucl. Acids Res. 17 : 271-284.

Gardiner, K., Laas, W., and Patterson, D. (1986).
Fractionation of large mammalian DNA restriction fragments using vertical pulsed field gradient gel electrophoresis.
Somatic Cell Molec. Genet. 12 : 185-195.

Gardiner, K., Horisberger, M., Kraus, J., Tantravahi, U., Korenberg, J., Rao, V., Reddy, S., and Patterson, D. (1990).
Analysis of human chromosome 21;   correlation of physical and cytogenetic maps; gene and CpG island distributions.
The EMBO Journal 9 : 25-34.

Gardner, A.P., Howell, R.T., and McDermott, A. (1982).
Fragile X chromosome : consistent demonstration of fragile site in fibroblast cultures.
Lancet 1 : 101.

Gillespie, G.A., Germino, G.G., Somlo, S., Weinstat-Saslow, D., Breuning, M.H., and Reeders, S.T. (1990).
Cosmid walking and chromosome jumping in the region of PKD1 reveal a locus duplication and three CpG islands.
Nucl. Acids Res. 18 : 7071-7075.

Giraud, F., Ayme, S., Mattei, J.F., and Mattei, M.G. (1976).
Constitutional chromosomal breakage.
Hum. Genet. 34 : 125-136.

Glover, T.W. (1981).
FUdR induction of the X chromosome fragile site : Evidence for the mechanism of folic acid and thymidine inhibition.
Am. J. Hum. Genet. 33 : 234-242.

Glover, T., and Stein, C.K. (1987).
Induction of sister chromatid exchanges at common fragile sites.
Am. J. Hum. Genet. 41 : 882-890.

Glover, T.W., Berger, C., Coyle, J. and Echo, B. (1984).
DNA polymerase alpha inhibition by aphidicolin induces gaps and breaks at common fragile sites in human chromosomes.
Hum. Genet. 67 : 136-142.

Gollin, S.M., Bock, H., Caskey, C.T. and Ledbetter, D.H. (1985a).
A new family with fra(10)(q25): Spontaneous expression and 100% expression with 100μM BrdU.
Am. J. Med. Genet. 21 : 643-648.

Gollin, S.M., Holmquist, G.P. and Ledbetter, D.H. (1985b).
Fra(10)(q25): The BrdU effect is substitution-dependent.
Am. J. Hum. Genet. 37 : 208-214.

Goz, B. (1978).
The effects of incorporation of 5-halogenated deoxyuridines into the DNA of eukaryotic cells.
Pharmacol. Rev. 29 : 249-272.

Graftstrom, R.H., Tseng, B.Y., and Goulian, M. (1978).
The incorporation of uracil into animal cell DNA *in vitro*.
Cell 15 : 131-140.

Green, E.D., and Olson, M.V. (1990a).
Systematic screening of yeast artificial chromosome libraries by use of the polymerase chain reaction.
Proc. Natl. Acad. Sci, USA 87 : 1213-1217.

Green, E.D. and Olson, M.V. (1990b).
Chromosomal region of the cystic fibrosis gene in yeast artificial chromosomes:
A model for human genome mapping.
Science 250 : 94-98.

Grunstein, M. and Hogness, D.S. (1975).
Colony hybridization : a method for the isolation of cloned DNAs that contain a specific gene.
Proc. Natl. Acad. Sci. USA 72 : 3961-3965.

Gustavson, K.H., Holmgren, G., Blomquist, H.G., and Mikkelsen, M. (1981).
Familial X-linked mental retardation and fragile X chromosomes in two Swedish families.
Clin. Genet. 19 : 101-110.

Hanish, J. and McClelland, M. (1990).
Methylase-linked partial *Not* I cleavage for physical mapping of genomic DNA.
Nucl. Acids. Res. 18 : 3287-3291.

Hanish, J., Rebelsky, M., McClelland, M., and Westbrook, C. (1991).
Application of methylase-limited partial *Not* I cleavage for a long range restriction map of the human ABL locus.
Genomics 10 : 681-685.

Hansen, R., Gartler, S., Scott, C., Chen, S., and Laird, C. (1992).
Methylation analysis of CGG sites in the CpG island of the human FMR1 gene.
Hum. Mol. Genet. 1 : 571-578.

Harris, P., Lalande, M., Stroh, H., Bruns, G., Flint, A. and Latt, S.A. (1987).
Construction of a chromosome 16 - enriched phage library and characterization of several DNA segments from 16p.
Hum. Genet. 77 : 95-103.

Harris, P.C., Barton, N.J., Higgs, D.R., Reeders, S.T. and Wilkie, A.O.M. (1990).
A long range restriction map between the a-globin complex and a marker closely linked to the polycystic kidney disease 1 (PKD1) locus.
Genomics 7 : 195-206.

Harrison, C.J., Jack, E.M., Allen, T.D., and Harris, R. (1983).
The fragile X : a scanning election microscope study.
J. Med. Genet. 20 : 280 - 285.

Harvey, J., Judge, C., and Wiener, S. (1977).
Familial X-linked mental retardation with an X chromosome abnormality.
J. Med. Genet. 14 : 46-50.

Hayman, D., and Sharp, P. (1981).
Hoechst 33258 induced uncondensed sites in marsupial chromosomes.
Chromosoma 83 : 249-262.

Hecht, F., Tolby, B., Magenis, R.E., Kimberling, W.J., Wyandt, H., and Lovrien, E.W. (1971).
Chromosomal localization of the heterochromatic region 16qh (.76) linked to alpha-haptoglobin in Man.
Nature 233 : 480.

Heitz, D., Rosseau, F., Devys, D., Saccone, S., Abderrahim, H., LePaslier, D., Cohen, D., Vincent, A., Toniolo, D., Della Valle, G., Johnson, S., Schlessinger, D., Oberle, I., and Mandel, L. (1991).
Isolation of sequences that span the fragile X and identification of a fragile X-related CpG island.
Science 251 : 1236-1239.

Hirsch, B. (1991).
Sister chromatid exchanges are preferentially induced at expressed and non expressed common fragile sites.
Hum. Genet. 87 : 302-306.

Hirst, M.C., Roche, A., Flint, T.D., MacKinnon, R.N., Bassett, J.H.D., Nakahori, Y., Watson, J.E.V., Bell, M.V., Patterson, M.N., Boyd, Y., Thomas, S.T., Knight, S.J.L., Warren, S.T., Hors-Cayla, M., Schmidt, M., Sutherland, G.R., and Davies, K.E. (1991).
Linear order of new and established DNA markers around the fragile site at Xq27.3.
Genomics 10 : 243-249.

Hirst, M.C., Rack, K., Nakahori, Y., Roche, A., Bell, M.V., Flynn, G., Christadoulou, Z., MacKinnon, R.N., Francis, M., Littler, A.J., Anand, R., Poustka, A., Lehrach, H., Schlessinger, D., D'Urso, M., Buckle, V.J., and Davies, K.E. (1991b).
A YAC contig across the fragile X site defines the region of fragility.
Nucl. Acids. Res. 19 : 3283-3288.

Hirst, M.C., Barnicoat, A., Flynn, G., Wang, Q., Daker, M., Buckle, V.J., Davies, K.E., and Bobrow, M. (1993).
The identification of a third fragile site, FRAXF in Xq27-28 distal to both FRAXA and FRAXE.
Hum. Molec. Genet. 2 : 197-200.

Hohn, B. and Collins, J. (1980).
A small cosmid for efficient cloning of large DNA fragments.
Genet. 11 : 291-298.

Hori, T., Takahashi, E., Ishihara, T., Minamihisumatsu, M., Kaneko, Y., and Murata, M. (1988a).
Distamycin A-inducible fragile sites and cancer proneness.
Cancer Genet. Cytogenet. 34 : 177-187.

Hori, T., Takahashi, E., and Murata, M. (1988b).
Nature of distamycin A-inducible fragile sites.
Cancer Genet. Cytogenet. 34 : 189-194.

Howard-Peebles, P.N. and Pryor, J. (1981).
Fragile sites in human chromosomes. I.
The effect of methionine on the Xq fragile site.
Clin. Genet. 19 : 228-232.

Hsu, T.C. and Somers, C.E. (1961).
Effects of 5-bromodeoxyuridine on mammalian chromosomes.
Proc. Natl. Acad. Sci. USA 47 : 396-403.

Hunt, P.A., and Burgoyne, P.S. (1987).
A tissue specific fragile site associated with the sex reversed (Sxr) mutation in the mouse.
Chromosome 96 : 67-71.

Hyland, V.J., Grist, S., Callen, D.F., and Sutherland, G.R. (1988).
Anonymous DNA probes to human chromosome 16 derived from a flow-purified library.
Am. J. Hum. Genet. 42 : 373-379.

Hyland, V.J., Fernandez, K.E., Callen, D.F., MacKinnon, R.N., Baker, E., Friend, K., and Sutherland, G.R. (1989a).
Assignment of anonymous DNA probes specific to intervals of chromosome 16 and X.
Human Genet. 83 : 61-66.

Hyland, V.J., Callen, D.F., Fernandez, K.E., MacKinnon, R.N., Friend, K., Mulley, J.C., Fratini, A.F., Baker, E., Breuning, M.H., Keith, T., and Sutherland, G.R. (1989b).
Anonymous DNA probes to specific intervals of human chromosome 16.
Cytogenet. Cell Genet. 51 : 1017.

Iannuzzi, M.C., Dean, M., Drumm, M.L., Hidaka, N., Cole, J.L., Perry, A., Stewart, C., Gerrard, B., and Collins, F.S. (1989).
Isolation of additional polymorphic clones from the cystic fibrosis region, using chromosome jumping from D7S8.
Am. J. Hum. Genet. 44 : 695-703.

Izakovic, V. (1984).
Homozygosity for fragile site at 17p12 in a 28 year old healthy man.
Hum. Genet. 68 : 340-341.

Jacky, P.B. and Sutherland, G.R. (1983).
Thymidylate synthetase inhibitors and fragile site expression in lymphocytes.
Am. J. Hum. Genet. 35 : 1276-1283.

Jacobs, P.A., Glover, P.W., Mayer, M., Fox, P., Gerrard, J.W., Dunn, H.G., and Herbst, D.J. (1980).
X-linked mental retardation : a study of seven families.
Am. J. Med. Genet. 7 : 471-489.

Jelenik, W.R., and Schmid, C.W. (1982).
Repetitive sequences in eukaryotic DNA and their expression.
Annu. Rev. Biochem. 51 : 813-844.

Jennings, M., Hall, J.D., and Hoehn, H. (1980).
Significance of phenotypic and chromosomal abnormalities in X linked mental retardation. (Martin Bell or Renpenning Syndrome).
Am. J. Med. Genet. 7 : 417-432.

Jurka, J. and Smith, T. (1988).
A fundamental division in the *Alu* family of repeated sequences.
Proc. Natl. Acad. Sci. USA 85 : 4775-4778.

Kievits, T., Dauwerse, J.G., Wiegant, J., Devilee, P., Breuning, M.H., Cornelisse, C.J., Van Ommen, G-J.B., and Pearson, P.L. (1990).
Rapid subchromosomal localization of cosmids by non-radioactive *in situ* hybridization.
Cytogenet. Cell Genet. 53 : 134-136.

Klotz, L.C. and Zimm, B.H. (1972).
Retardation times of deoxyribonucleic acid solutions. II. Improvements in apparatus and theory.
Macromolecules 5 : 471-481.

Knight, S.J., Flannery, A.V., Hirst, M.C., Campbell, L., Christodoulou, Z., Phelps, S.R., Pointon, J., Middleton-Price, H.R., Barnicoat, A., Pembrey, M.E., Holland, J., Oostra, B.A., Bobrow, M., and Davies, K.E. (1993).
Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE/mental retardation.
Submitted for Publication.

Kozman, H.M., Gedeon, A.K., Whitmore, S., Suthers, G.K., Callen, D.F., Sutherland, G.R. and Mulley, J.C. (1991).
Addition of MT, D16S10, D16S4, and D16S91 to the linkage map within 16q12.1-16q22.1.
Genomics 11 : 756-759.

Kozman, H.M., Phillips, H.A., Callen, D.F., Sutherland, G.R. and Mulley, J.C. (1993).
Integration of the cytogenetic and genetic linkage maps of human chromosome 16 using 50 physical intervals and 50 polymorphic loci.
Cytogenet. Cell Genet. (in press).

Krawczun, M.S., Lele, K.P., Jenkins, E.C., and Brown, T. (1986).
Fragile X expression increased by low cell culture density.
Am. J. Med. Genet. 23 : 467-473.

Kremer, E.J., Yu, S., Pritchard, M., Nagaraja, R., Heitz, D., Lynch, M., Baker, E., Hyland, V.J., Little, R.D., Wada, M., Toniolo, D., Vincent, A., Rousseau, F., Schlessinger, D., Sutherland, G.R., and Richards, R.I. (1991a).
Isolation of a human DNA sequence which spans the fragile X.
Am. J. Hum. Genet. 49 : 656-661.

Kremer, E.J., Pritchard, M., Lynch, M., Yu, S., Holman, K., Baker, E., Warren, S.T., Schlessigner, D., Sutherland, G.R. and Richards, R.I. (1991b).
Mapping of DNA instability at the fragile-X to a trinucleotide repeat sequence $p(CCG)_n$.
Science 252 : 1711-1714.

Krumdieck, C.L. and Howard-Peebles, P.N. (1983).
On the nature of folic acid sensitive fragile sites on human chromosomes : an hypothesis.
Am. J. Med. Genet. 16 : 23-28.

Kukita, A., Yoshida, M.C., Fukushige, S., Sakakibara, M., Joh, K., Mukai, T. and Hori, K. (1987).
Molecular gene mapping of human aldolase-A (ALDOA) gene to chromosome 16.
Hum. Genet. 76 : 20-26.

Lagerstrom, M., Parik, J., Malmgren, H., Stewart, J., Petterson, U., and Landegren, U. (1991).
Capture PCR : Efficient amplification of DNA fragments adjacent to a known sequence in human and YAC DNA.
PCR Methods and Applications (in press).

Lai, E. (1991).
Pulsed field gel electrophoresis - an overview.
The NEB Transcript 3 : 1-3.

Lai, E., Birren, B.W., Clark, S.M., Simon, M.I., and Hood, L. (1989).
Pulsed field gel electrophoresis.
Biotechniques 7 : 34-42.

Lalley, P.A. and McKusick, V.A. (1985).
Report of the committee on comparative mapping : Human gene mapping 8.
Cytogenet. Cell Genet. 40 : 536-566.

Lapsys, N.M., Layfield, R., Baker, E., Callen, D.F., Sutherland, G.R., Abedinia, M., Nixon, P.F. and Mattick, J.S. (1992).
Chromosomal location of the human transketolase gene.
Cytogenet. Cell Genet. 61 : 274-275.

LaSpada, A.R., Wilson, E.M., Lubahn, D.B., Harding, A.E. and Fischbeck, K.H. (1991).
Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy.
Nature 352 : 77-79.

Lawrence, J.B., Singer, R.H. and McNeil, J.A. (1990).
Interphase and metaphase resolution of different distances within the human dystrophin gene.
Science 249 : 928-932.

LeBeau, M.M., Larson, R.A., Bitter, M.A., Vardiman, J.W., Golomb, H.M., and Rowley, J.D. (1983).
Association of an inversion of chromosome 16 with abnormal marrow eosinophils in acute myelomonocytic leukemia.
The New England Journal of Medicine 309 : 630-636.

Ledbetter, D.H., Airhart, S.D. and Nussbaum, R.L. (1986a).
Caffeine enhances fragile (X) expression in somatic cell hybrids.
Am. J. Med. Genet. 23 : 445-456.

Ledbetter, D.H., Ledbetter, S.A., and Nussbaum, R.L. (1986b).
Implications of fragile X expression in normal males for the nature of the mutation.
Nature 324 : 161-163.

Ledbetter, D.H., Airhart, S.D. and Nussbaum, R.L. (1986c).
Somatic cell hybrid studies of fragile (X) expression in a carrier female and transmitting male.
Am. J. Med. Genet. 23 : 429-443.

Ledbetter, S.A., Nelson, D.L., Warren, S.T. and Ledbetter, D.H. (1990).
Rapid isolation of DNA probes within specific chromosome regions by interspersed repetitive sequence polymerase chain reaction.
Genomics 6 : 475-481.

Lejeune, J., Dutrillaux, B., Lafourcade, J., Berger, R., Abonyi, D., and Rethore, M.O. (1968).
Endoreduplication selective du bras long du chromosome 2 chez une femme et sa fille.
C.R. Acad. Sci. Paris 266 : 24-26.

Lerman, L.S., and Frisch, H.L. (1982).
Why does the elctrophoretic mobility of DNA in gels vary with the length of the molecule?
Biopolymers 21 : 995-997.

Li, N. and Zhou, X. (1985).
IV. Uridine-induced hot-point breaks at 3p14 and 16q23-24 and increased expression of fragile site Xq27 in folate-free medium.
Hum. Genet. 71 : 363-365.

Li, X., Yan, Z., and Zhou, X. (1986a).
The effect of 1-B-D-arabinofuranosyl-cytosine on the expression of the common fragile site at 3p14.
Hum. Genet. 74 : 444-446.

Li, N., Wu, Y., and Zhou, X. (1986b).
V. The effect of four nucleosides on chromosomes in folate-free medium.
Hum. Genet. 74 : 101-103.

Lin, M.S., Shimanuki, K., and Wilson, M.G. (1987).
Expression of fragile X in human-mouse somatic cell hybrids.
Cytogenet. Cell. Genet. 44 : 118-122.

Lin, S., Figueiredo, C., Sciorra, L., and Lee, M. (1987).
Study of human chromosome V. The effects of thymidine concentration and timing on the expression of uridine-induced constitutive fragile sites.
Human Genet. 76 : 173-175.

Lindsay, S., and Bird, A.P. (1987).
Use of restriction enzymes to detect potential gene sequences in mammalian DNA.
Nature (London) 327 : 336-338.

Litt, M., and White, R.L. (1985).
A highly polymorphic locus in human DNA revealed by cosmid derived probes.
Proc. Natl. Acad. Sci. USA 82 : 6206-6210.

Longmire, J.L., Brown, N.C., Ford, A.F., Naranjo, C., Ratliff, R.L., Hildebrand, C.E., Stallings, R.L., Costa, A.K., and Deaven, L.L. (1991).
Automated construction of high density gridded arrays of chromosome-specific cosmid libraries.
Lab. Robotics Automation 3 : 195-198.

Lubs, H.A. (1969).
A Marker X chromosome.
Am. J. Hum. Genet. 21 : 231-244.

Lukusa, T., Meulepas, E., Fryns, J.B., Van den Berghe, H., and Cassiman, J.J. (1991).
"Spontaneous" FRA16B is a hot spot for sister chromatid exchanges.
Hum. Genet. 87 : 583-586.

Lumpkin, O.J., Dejardin, P., and Zimm, B.H. (1985).
Theory of gel electrophoresis of DNA.
Biopolymers, 24 : 1573-1593.

Magenis, E. and Chamberlin, J. (1979).
Location of fragile site : 16q22.00.
Cytogenet. Cell Genet. 25 : 180.

Magenis, R.E., Hecht, F., and Lovrien, E.W. (1970).
Heritable fragile site on chromosome 16 : Probable localization of haptoglobin locus in Man.
Science 170 : 85-87.

Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, C., Jansen, G., Neville, C., Narang, M., Barcelo, J., O'Hoy, K., LeBlond, S., Macdonald, J., deJong, P., Wieringa, B., and Korneluk, R.G. (1992).
Myotonic dystrophy mutation: An unstable CTG repeat in the 3' untranslated region of the gene.
Science 255 : 1253-1255.

Maire, P., Gautron, S., Hakim, V., Gregori, C., Mennecier, F., and Kahn, A. (1987).
Characterization of 3 optional promotors in the 5'-region of human aldolase A gene.
J. Mol. Biol. 197 : 425-438.

Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982).
Molecular cloning : A laboratory manual. Cold spring Harbor Laboratory, Cold Spring Harbor, New York.

Mansouri, A., Spurr, N., Goodfellow, P.N., Kemler, R. (1988).
Characterization and chromosomal localization of the gene encoding the human cell adhesion molecule uvomorulin.
Differentiation 38 : 67-71.

Mathew, M.K., Smith, C., and Cantor, C.R. (1988a).
High resolution separation and accurate size determination in pulsed field gel electrophoresis of DNA. 1. DNA size standards and the effects of agarose and temperature.
Biochemistry 27 : 9204-9210.

Mathew, M.K., Smith, C.L., and Cantor, C.R. (1988b).
High resolution separation and accurate size determination in pulsed field gel electrophoresis of DNA. 2. Effect of pulse time and electric field strength and implications for models of the separation process.
Biochemistry 27 : 9210-9216.

Mattei, M.G., Ayme, S., Mattei, J.F., Aurran, Y., and Giraud, F. (1979).
Distribution of spontaneous chromosome breaks in man.
Cytogenet. Cell Genet. 23 : 95-102.

Meese, E., Meltzer, P., and Trent, J. (1989).
Application of natural partial digests to pulsed field gel analysis of the amplified MDR locus.
Genomics 5 : 371-374.

Messing, J., Gronenborn, B., Muller-Hill, B., and Hofschneider, P. (1977).
Filamentous coliphage M13 as a cloning vehicle. Insertion of a *Hind* II fragment of the *lac* regulatory region in M13 replicative form *in vitro*.
Proc. Natl. Acad. Sci. 74 : 3642-3646.

Mixon, C., and Dev, V.G. (1982).
Effect of FUdR and methionine on the expression of fragile X in a lymphoblastoid cell line.
Clin. Res. 30 : 891A.

Mixon, J.C., and Dev, V.G. (1983).
Fragile X expression is decreased by 5-azacytidine and 5-adenosylhomocysteine.
Am. J. Hum. Genet. 35 : 1270-1275.

Moyzis, R.K., Albright, K.L., Bartholdi, M.F., Cram, L.S., Deaven, L.L., Hildebrand, C.E., Joste, N.E., Longmire, J.L., Meyne, J. and Schwarzacher-Robinson, T. (1987).
Human chromosome-specific repetitive DNA sequences : Novel markers for genetic analysis.
Chromosoma 95 : 375-386.

Moyzis, R.K., Torney, D.C., Meyne, J., Buckingham, J.M., Wu, J.R., Burks, C., Sirotkin, K.M. and Goad, W.B. (1989).
Distribution of interspersed repetitive DNA sequences in the human genome.
Genomics 4 : 273-289.

Mukai, T., Arai, Y., Yatsuki, H., Joh, K., and Honi, K. (1991).
An additional promotor function in the human aldolase A gene, but not in rat.
Eur. J. Biochem. 195 : 781-787.

Mulley, J.C., Hay, J., Sheffield, L.J., and Sutherland, G.R. (1983).
Regional localization for HLA by recombination with a fragile site at 6p23.
Am. J. Hum. Genet. 35 : 1284-1288.

Mulley, J.C., Gedeon, A.K., Thorn, K.A., Bates, L.J., and Sutherland, G.R. (1987).
Linkage and genetic councelling for the fragile X using DNA probes 52A, F9, DX13, and ST14.
Am. J. Med. Genet. 27 : 435-448.

Mulley, J.C., Hyland, V.J., Fratini, A., Bates, L.J., Gedeon, A.K. and Sutherland, G.R. (1989).
A linkage group with FRA16B (the fragile site at 16q22.1).
Hum. Genet. 82 : 131-133.

McCormick, M.K., Shero, J.H., Connelly, C.J., Antonarakis, S.E., and Hieter, P. (1990).
Methods for cloning the large DNA segments as artificial chromosomes in S. cerevisiae.
Technique 2 : 65-71.

McGill, J.R., Yang, F., Baldwin, W.D., Brune, J.L., Barnett, D.R., Bowman, B.H. and Moore, C.M. (1984).
Localization of the haptoglobin α and B-genes (HPA and HPB) to human chromosome 16q22 by in situ hybridization.
Cytogenet. Cell Genet. 38 : 155-157.

McLean, J., Wien, J., Drayna, D., Fielding, G. and Lawn, R. (1986).
Human lecithin : cholesterol acyltransferase gene : complete gene sequence and sites of expression.
Nucl. Acids Res. 14 : 9397-9406.

McPeek, F.D., Coyle-Morris, J.F., and Gemmil, R.M. (1986).
Separation of large DNA molecules by modified pulsed field gradient gel electrophoresis.
Anal. Biochem. 156 : 274-285.

Natt, E., Westphal, E.M., Toth-Fejel, S.E., Magenis, R.E., Buist, N.R., Rettenmeier, R., and Scherer, G. (1987).
Inherited or *de novo* deletion of the tyrosine aminotransferase gene locus at 16q22.1-q22.3 in a patient with tyrosinemia type II.
Hum. Genet. 77 : 352-358.

Natt, E., Magenis, R.E., Zimmer, J., Mansouri, A., and Scherer, G. (1989).
Regional assignment of the loci for uvomorulin (UVO) and chymotrypsinogen B (CTRB) on human chromosome 16q.
Cytogenet. Cell Genet. 51 : 1050.

Nelson, D.L., Ledbetter, S.A., Corbo, L., Victoria, M.F., Ramirez-Solis, R., Webster, T.D.., Ledbetter, D.H., and Caskey, C.T. (1989).
Alu polymerase chain reaction : A method for rapid isolation of human specific sequences from complex DNA sources.
Proc. Natl. Acad. Sci. USA 86 : 6686-6690.

Nguyen, C., Poostka, A.M., Djabali, M., Roux, D., Mattei, J.F., Lehrach, H., and Jordan, B.R. (1989).
Large scale mapping and chromosome jumping in the q27 region of the human X chromosome.
Genomics 5 : 298-303.

Nussbaum, R.L. and Ledbetter, D.H. (1986).
Fragile X syndrome : A unique mutation in man.
Ann. Rev. Genet. 20 : 109-145.

Nussbaum, R.L., Airhart, S.D., and Ledbetter, D.H. (1983).
Expression of the fragile (X) chromosome in an interspecific somatic cell hybrid.
Hum. Genet. 64 : 148-150.

Nussbaum, R.L., Walmsley, R.M., Lesko, J.G., Airhart, S.D. and Ledbetter, D.H. (1985).
Thymidylate synthase-deficient Chinese hamster cells : a selection system for human chromosome 18 and experimental system for the study of thymidylate synthase regulation and fragile X expression.
Am. J. Hum. Genet. 37 : 1192-1205.

Nussbaum, R.L., Airhart, S.D. and Ledbetter, D.H. (1986).
A rodent - human hybrid containing Xq24 -qter translocated to a hamster chromosome expresses the Xq 27 folate - sensitive fragile site.
Am. J. Med. Genet. 23 : 457-466.

O'Brien, J.J. (1989).
Genetic Maps. Locus Maps of Complex Genomes.
5th Edition. Cold Syprings Harbour Laboratory Press, New York.

Oberlè, I., Drayna, D., Camerino, G., White, R., and Mandel, J.L. (1985a).
The telomeric region of the human X chromosome long arm : Presence of a highly polymorphic DNA marker and analysis of recombination frequency.
Proc. Natl. Acad. Sci. USA 82 : 2824-2828.

Oberlé, I., Mandel, J.L., Boue, J., Mattei, M.G., and Mattei, J.F. (1985b).
Polymorphic DNA markers in prenatal diagnosis of fragile X syndrome.
Lancet. i : 871.

Oberlé, I., Heilig, R., Moisan, J.P., Kloepfer, C., Mattei, M.G., Mattei, J.F., Boué, J., Froster-iskenius, U., Jacobs, P.A., Lathrop, G.M., Lalouel, J.M. and Mandel, J.L. (1986).
Genetic analysis of the fragile X mental retardation syndrome with two flanking polymorphic DNA markers.
Proc. Natl. Acad. Sci. USA 83 : 1016-1020.

Oberlé, I., Rousseau, F., Heitz, D., Kretz, C., Devys, D., Hanauer, A., Boue, J., Bertheas, M.F., and Mandel, J.L. (1991).
Instability of a 550 base pair DNA segment and abnormal methylation in fragile X syndrome.
Science 252 : 1097-1102.

Ochman, H., Gerber, A.S., and Hartl, D.C. (1988).
Genetic applications of an inverse polymerase chain reaction.
Genetics 120 : 621-623.

Oostra, B.A., Hupkes, P.E., Perdon, L.F., van Bennekom, C.A., Baker, E., Halley, D.J.J., and Schmidt, M. (1990).
New polymorphic DNA markers close to the fragile site FRAXA.
Genomics 6 : 129-132.

Orbach, M.J., Vollrath, D., Davis, R.W. and Yanofsky, C. (1988).
An electrophoretic karyotype of *Neurospora crassa.*
Mol. Cell. Biol. 8 : 1469-1473.

Orita, M., Sekiya, T. and Hayashi, K. (1990).
DNA sequence polymorphisms in *Alu* repeats.
Genomics 8 : 271-278.

Overhauser, J. and Radic, M.Z. (1989).
Encapsulation of cells in agarose beads for use with pulsed - field gel electrophoresis.
Focus 9 : 8-9.

Page, D.C., Mosher, R., Simpson, E.M., Fisher, E.M., Mardon, G., Pollack, J., McGillivray, B., de la Chapelle, A. and Brown, L. (1987).
The sex determining region of the human Y chromosome encodes a finger protein.
Cell 51 : 1091-1104.

Parra, I. and Windle, B. (1993).
High resolution visual mapping of stretched DNA by fluroescent hybridization.
Nature Genetics 5 : 17-21.

Patterson, M., Schwartz, C., Bell, M., Sauer, S., Hofker, M., Trask, B., van den Engh, C., and Davies, K.E. (1987).
Physical mapping studies on the human X chromosome in the region Xq27-Xqter.
Genomics 1 : 297-306.

Pieretti, M., Zhang, F., Fu, Y., Warren, S.T., Oostra, B., Caskey, C., and Nelson, D. (1991).
Absence of expression of the FMR-1 gene in fragile X syndrome.
Cell 66 : 817-822.

Poulsen, B.S., and Ronne, M. (1991).
High resolution R-banding and localization of fragile sites in Orycytolagus cuniculus.
Genet. Sel. Evol. 23 : 183-186.

Poustka, A., and Lehrach, H. (1986).
Jumping libraries and linking libraries : the next generation of molecular tools in mammalian genetics.
Trends Genet. 2 : 174-179.

Poustka, A., Dietrich, A., Langenstein, G., Toniolo, D., Warren, S.T., and Lehrach, H. (1991).
Physical map of human Xq27-qter : Localizing the region of the fragile X mutation.
Proc. Natl. Acad. Sci. USA 88 : 8302-8306.

Pritchard, C.A., Casher, D., Uglum, E., Cox, D.R. and Myers, R.M. (1989).
Isolation and field inversion gel electrophoresis analysis of DNA markers close to the Huntington disease gene.
Genomics 4 : 408-418.

Puspurs, A.H., Baker, E., Callen, D.F., Fratini, A., and Sutherland, G.R. (1988).
Translocation breakpoint at 11q13.3 in the t(11;14) in B-cell leukemia is not at the rare fragile site at 11q13.3.
Cancer Genet. Cytogenet. 31 : 25-30.

Rappold, G., and Lehrach, H. (1986).
A long range restriction map of the pseudo autosomal region by partial digest PFGE analysis from the telomere.
Nucl. Acids Res. 16 : 5361-5377.

Reed, K.C. and Mann, D.A. (1985).
Rapid transfer of DNA from agarose gels to nylon membranes.
Nucl. Acids Res. 13 : 7207-7221.

Reeders, S.T. and Hildebrand, C.E. (1989).
Report of the committee on the genetic constitution of chromosome 16.
Cytogenet. Cell Genet. 51 : 299-318.

Reid, T., Baldini, A., Rand, T., and Ward, A. (1992).
Simultaneous visualization of seven different DNA probes by *in situ* hybridization using combinatorial fluorescence and digital imaging microscopy.
Proc. Natl. Acad. Sci. USA 89 : 1388-1392.

Rich, A., Nordheim, A., and Wang, A. (1984).
The chemistry and biology of left-handed Z-DNA.
Ann. Rev. Biochem. 53 : 791-846.

Richards, R.I., and Sutherland, G.R. (1992).
Heritable unstable DNA sequences.
Nature Genetics 1 : 7-9.

Riggs, P.K., and Chrisman, C.L. (1989).
Preliminary analysis of aphidicolin induced fragile sites in pig chromosomes.
6th North Am. Coll. Cytogenet. Domest. Anim. Abs. 4.

Riley, J., Butler, R., Ogilvie, D., Finniear, R., Jenner, D., Powel, S., Anand, R., Smith, J.C., and Markham, A.F. (1990).
A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones.
Nucl. Acids Res. 18 : 2887-2890.

Roberts, J.J. (1984).
Mechanisms of potentiation by caffeine of genotoxic damage induced by physical and chemical agents. In : DNA repair and its inhibition.
Collins, A., Downs, C.S. and Johnson, R.T., eds. IRL Press. Oxford.

Robinson, T.J., and Elder, F.F.B. (1987).
Multiple common fragile sites are expressed in the genome of the laboratory rat.
Chromosoma 96 : 45-49.

Rommens, J.M., Zengerling-Lentes, S., Kerem, B., Melmer, M.B., and Tsui, L.C. (1989).
Physical localization of 2 DNA markers closely linked to the cystic fibrosis locus by pulsed field gel electrophoresis.
Am. J. Hum. Genet. 45 : 932-941.

Ronne, M. (1992).
Putative fragile sites in the horse karyotype.
Hereditas 117 : 127-136.

Rosenthal, A., MacKinnon, R.N., and Jones, S. (1991).
PCR walking from microdissection clone M54 identifies three exons from the human gene for the neural cell adhesion molecule L1 (CAM-L1).
Nucl. Acids Res. Vol. 19 : 5395-5401.

Rousseau, F., Vincent, A., Rivella, S., Heitz, D., Triboli, C., Maestrini, E., Warren, S.T., Suthers, G.K., Goodfellow, P., Mandel, J.L., Toniolo, D., and Oberlé, I. (1991).
Four chromosomal breakpoints and four new probes mark out a 10 cM region encompassing the fragile X locus (FRAXA).
Am. J. Hum. Genet. 48 : 108-116.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1985).
Primer directed enzymatic amplification of DNA with a thermostable DNA polymerase.
Science 230 : 1350-1354.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977).
DNA Sequencing with chain terminating inhibitors.
Proc. Natl. Acad. Sci. USA 74 : 5463-5467.

Sawadogo, M. and Van Dyke, M.W. (1991).
A rapid method for the purification of deprotected oligodeoxynucleotides.
Nucl. Acids. Res. 19 : 674.

Scheres, J.M. and Hustinx, T.W. (1980).
Heritable fragile sites and lymphocyte culture medium containing BrdU.
Am. J. Hum. Genet. 32 : 628-629.

Schmid, M., Klett, C., and Niederhofer, A. (1980).
Demonstration of a heritable fragile site in human chromosome 16 with distamycin A.
Cytogenet. Cell Genet. 28 : 87-94.

Schmid, M., Poppen, A., and Engel, W. (1981).
The effects of distamycin A on gorilla, chimpanzee, and orangutan lymphocyte cultures.
Cytogenet. Cell Genet. 30 : 211-221.

Schmid, M., Hungerford, D.A., Poppen, A., and Engel, W. (1984).
The use of distamycin A in human lymphocyte cultures.
Hum. Genet. 65 : 377-384.

Schmid, M., Ott, G., Haaf, T., and Scheres, J. (1985).
Evolutionary conservation of fragile sites induced by 5-azacytidine and 5-azadeoxycytidine in man, gorilla and chimpanzee.
Hum. Genet. 71 : 342-350.

Schmid, M., Feichtinger, W., Jessberger, A., Kohler, J., and Lange, R. (1986).
The fragile site (16)(q22). I. Induction by AT-specific DNA-ligands and population frequency.
Hum. Genet. 74 : 67-73.

Schmid, M., Feichtinger, W., Deubelbeiss, D., and Weller, E. (1987a).
The fragile site (17)(p12) : induction by AT-specific DNA-ligands and population cytogenetics.
Hum. Genet. 77 : 118-121.

Schmid, M., Feichtinger, W., and Haaf, T. (1987b).
The fragile site (16) (q22). II. Sister chromatid exchanges.
Hum. Genet. 76 : 365-368.

Schneider, N.R., Chaganti, R.S.K., and German, J. (1980).
Analysis of a BrdU-sensitive site in the cactus mouse Peromyscus eremicus : chromosomal breakage and sister chromatid exchange.
Chromosoma 77 : 379-389.

Schwartz, D.C., and Cantor, C.R. (1984).
Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis.
Cell 37 : 67-75.

Schwartz, D.C., Saffron, W., Welsh, J., Haas, R., Goldenberg, M. and Cantor, C.R. (1982).
New techniques for purifying large DNAs and studying their properties and packaging.
Cold Springs Harbor Symp. Quant. Biol. 47 : 189-195.

Scott, A.F., Schmeckpeper, B.J., Abdelrazik, M., Comey, T.C., O'Hara, B., Rossiter, J., Couley, T., Heath, P., Smith, K.D. and Margolet, L. (1987).
Origin of the human L1 elements : Proposed progenitor genes deduced from a consensus DNA sequence.
Genomics 1 : 113-125.

Sealey, P.G., Whittaker, P.A., and Southern, E.M. (1985).
Removal of repeated sequences from hybridization probes.
Nucl. Acids Res. 13 : 1905-1922.

Serero, S., Maire, D., Nguyen Von Gong, Cohen-Haguenaeur, O., Gross, M.S., Jegou-Foubert, C., deTand, M.F., Kahn, A. and Frezal, J. (1988).
Localization of the active gene of aldolase on chromosome 16 and two aldolase A pseudogenes on chromosome 3 and 10.
Hum. Genet. 78 : 167-174.

Shabtai, F., Klar, D., Bichacho, S., Hart, S., and Halbrecht, I. (1983).
Familial fragility on chromosome 16 (fra 16q22) enhanced by both interferon and distamycin A.
Hum. Genet. 63 : 341-344.

Shabtai, F., Klar, D., Hart, J., and Halbrecht, I. (1985).
On the meaning of fragile sites in cancer risk and development.
Cancer Genet. Cytogenet. 18 : 81-85.

Silverman, G.A., Ye, R.D., Pollock, K.M., Sadler, J.E. and Korsmeyer, S.J. (1989).
Use of yeast artificial chromosome clones for mapping and walking within human chromosome segment 18q21.3.
Proc. Natl. Acad. Sci. USA. 86 : 7485-7489.

Simmers, R.N., Stupans, I., and Sutherland, G.R. (1986).
Localization of the human haptoglobin genes distal to the fragile site at 16q22 using *in situ* hybridization.
Cytogenet. Cell Genet. 41 : 38-41.

Simmers, R.N., Sutherland, G.R., West, R., and Richards, R.I. (1987).
Fragile sites at 16q22 are not at the breakpoint of the chromosomal rearrangement in AMMoL.
Science 236 : 92-94.

Simmers, R.N. and Sutherland (1988).
Further localization of ETS 1 indicates that the chromosomal rearrangement in Ewing sarcoma does not occur at fra(11)(q23).
Hum. Genet. 78 : 144-147.

Singer, M.F. (1982).
Highly repeated sequences in mammalian genomes.
Int. Rev. Cytol. 76 : 67-112.

Smeets, D.F., and van de Klundert, F.A. (1990).
Common fragile sites in man and three closely related primate species.
Cytogenet. Cell Genet. 53 : 8-14.

Smeets, D., Scheres, J., and Hustinx, T. (1986).
The most common fragile site in man is 3p14.
Hum. Genet. 72 : 215-220.

Smith, C.L., Lawrance, S.K., Gillespie, G.A., Cantor, C.R., Weissman, S.M., and Collins, F.S. (1987).
Mapping and cloning macro-regions of mammalian genomes.
In "Methods in Enzymology" (M. Gottesman, Ed.). Vol. 151 p.p. 461-489, Academic Press, San Diego.

Smith, S.B., Aldridge, P.K., and Callis, J.B. (1989).
Observation of individual DNA molecules undergoing gel electrophoresis.
Science 243 : 203-206.

Smith, E., Robson, L., and Schwanitz, G. (1992).
Fluorescence *in situ* hybridization.
Life Science : 56-63.

Sorensen, K., Nielsen, J., Holm, V. and Haahr, J. (1979).
Fragile site long arm chromosome 16.
Hum. Genet. 48 : 131-134.

Southern, E.M., Anand, R., Brown, W.R., and Fletcher, D.S. (1987).
A model for the separation of large DNA molecules by crossed field gel electrophoresis.
Nucl. Acids Res. 15 : 5925-5943.

Spies, T., Breshnahan, M., and Strominger, J.L. (1989).
Human major histocompatibility complex contains a minimum of 19 genes between the complement cluster and HLA-B.
Proc. Natl. Acad. Sci 86 : 8955-8958.

Stallings, R.L., Torney, D.C., Hildebrand, C.E., Longmire, J.L., Deaven, L.L., Jett, J.H., Doggett, N.A., and Moyzis (1990).
Physical mapping of human chromosomes by repetitive sequence fingerprinting.
Proc. Natl. Acad. Sci USA 87 : 6218-6222.

Stallings, R.L., Doggett, N.A., Callen, D., Apostolous, S., Chen, L.Z., Nancarrow, J.K., Whitmore, S.A., Harris, P., Michison, H., Breuning, M., Saris, J.J., Fickett, J., Cinkosky, M., Torney, D.C., Hildebrand, C.E. and Moyzis, R.K. (1992a).
Evaluation of a cosmid contig physical map of human chromosome 16.
Genomics 13 : 1031-1039.

Stallings, R.L., Doggett, N.A., Okumura, K., and Ward, D.C. (1992b).
Chromosome 16 - specific repetitive DNA sequences that map to chromosomal regions known to undergo breakage/rearrangement in Leukemia cells.
Genomics 13 : 332-338.

Stone, D.M., Jacky, P.B., Hancock, D.D., and Prieur, D.J. (1991a).
Chromosomal fragile site expression in dogs : I. Breed specific differences.
Am. J. Med. Genet. 40 : 214-222.

Stone, D.M., Jacky, P.B., and Prieur, D.J. (1991b).
Chromosomal fragile site expression in dogs : II. Expression in Boxer dogs with mast cell tumors.
Am. J. Med. Genet. 40 : 223-229.

Stone, D.M., and Stephens, K.E. (1993).
Bromodeoxyuridine induces chromosomal fragile sites in the canine genome.
Am. J. Med. Genet. 46 : 198-202.

Sutcliffe, J., Nelson, D., Zhang, F., Pieretti, M., Caskey, C., Saxe, D., and Warren, S.T. (1992).
DNA methylation represses FMR-1 transcription in fragile X syndrome.
Hum. Mol. Genet. 1 : 397-400.

Sutherland, G.R. (1977).
Fragile sites on human chromosomes : Demonstration of their dependence on the type of tissue culture medium.
Science 197 : 265-266.

Sutherland, G.R. (1979a).
Heritable fragile sites on human chromosomes. I. Factors affecting expression in lymphocyte culture.
Am. J. Hum. Genet. 31 : 125-135.

Sutherland, G.R. (1979b).
Heritable fragile sites on human chromosomes. II. Distribution, phenotypic effects and cytogenetics.
Am. J. Hum. Genet. 31 : 136-148.

Sutherland, G.R. (1981).
Heritable fragile sites on human chromosomes. VII. Children homozygous for the BrdU - requiring fra (10)(q25) are phenotypically normal.
Am. J. Hum. Genet. 33 : 946-949.

Sutherland, G.R. (1982).
Heritable fragile sites on human chromosomes. IX. Population cytogenetics and segregation analysis of the BrdU - requiring fragile site at 10q25.
Am. J. Hum. Genet. 34 : 753-756.

Sutherland, G.R. (1983).
The fragile X chromosome.
Int. Rev. of Cytol. 81 : 107-143.

Sutherland, G.R. and Baker, E. (1990).
The common fragile site in band q27 of the human X chromosome is not coincident with the fragile X.
Clinical Genet. 37 : 167-172.

Sutherland, G.R., and Baker, E. (1992).
Characterization of a new rare fragile site easily confused with the fragile X.
Hum. Mol. Genet. 1 : 111-113.

Sutherland, G.R., and Hecht, F. (1985).
Fragile sites on human chromosomes.
Oxford University Press, New York.

Sutherland, G.R. and Ledbetter, D.H. (1989).
Report of the committee on cytogenetic markers.
Cytogenet. Cell Genet. 51 : 452-458.

Sutherland, G.R. and Simmers, R.N. (1988).
No statistical association between fragile sites and non random chromosome breakpoints in cancer cells.
Cancer Genet. Cytogenet. 31 : 9-15.

Sutherland, G.R., Baker, E., and Seshardi, R.S. (1980).
Heritable fragile sites on human chromosomes. V. A new class of fragile site requiring BrdU for expression.
Am. J. Hum. Genet. 32 : 542-548.

Sutherland, G.R., Baker, E., and Mulley, J.C. (1982).
Genetic length of a human chromosomal segment measured by recombination between two fragile sites.
Science 217 : 373-374.

Sutherland, G.R., Jacky, P.B., and Baker, E.G. (1984).
Heritable fragile sites on human chromosomes. XI. Factors affecting expression of fragile sites at 10q25, 16q22 and 17p12.
Am. J. Hum. Genet. 36 : 110-122.

Sutherland, G.R., Baker, E. and Fratini, A. (1985a).
Excess thymidine induces folate sensitive fragile sites.
Am. J. Hum. Genet. 22 : 433-443.

Sutherland, G.R., Parslow, M.I. and Baker, E. (1985b).
New classes of common fragile sites induced by 5-azacytidine and bromodeoxyuridine.
Hum. Genet. 69 : 233-237.

Sutherland, G.R., Reeders, S., Hyland, V.J., Callen, D.F., Fratini, A., and Mulley, J.C. (1987).
Molecular Genetics of human chromosome 16.
J. Med. Genet. 24 : 451-456.

Sutherland, G.R., Baker, E., Callen, D.F., Hyland, V.J., May, B.K., Bawden, M.J., Healy, H.M., and Borthwick, I.A. (1988).
5-ominoevulinate synthase is at 3p21 and thus not the primary defect in X-linked sideroblastic anemia.
Am. J. Hum. Genet. 43 : 331-335.

Sutherland, G.R., Baker, E., Callen, D.F., Margaret Garson, O., and West, A.K. (1990).
The human metallothionein gene cluster is not disrupted in myelomonocytic leukaemia.
Genomics 6 : 144-148.

Suthers, G.K., Callen, D.F., Hyland, V.J., Kozman, H.M., Baker, E., Eyre, H., Harper, P.S., Roberts, S.H., Hors Cayla, M.C., Davies, K.E., Bell, M.V. and Sutherland, G.R. (1989).
A new DNA marker tightly linked to the fragile X locus (FRAXA).
Science 246 : 1298-1300.

Tagle, D., and Collins, F.S. (1992).
An optimized *Alu*-PCR primer pair for human specific amplification of YACs and somatic cell hybrids.
Hum. Molec. Genet. 1 : 121-122.

Takahashi, E., Kaneko, Y., Ishihara, T., Minamihisamatsu, M., Murata, M., and Hari, T. (1988a).
A new rare distamycin A-inducible fragile site, fra(11)(p15.1), found in two acute nonlymphocytic leukaemia (ANLL) patients with t(7:11)(p15-p13; p15).
Hum. Genet. 80 : 124-126.

Takahashi, E., Hori, T., and Murata, M. (1988b).
Population cytogenetics of rare fragile sites in Japan.
Hum. Genet. 78 : 121-126.

Taylor, A.M.R. and Bundey, S. (1983).
Spontaneous expression of the chromosome fragile site fra(10)(q25).
Am. J. Hum. Genet. 35 : 123-125.

Thestrup-Pedersen, K., Essman, V., Jensen, J.R., Hastrup, J., Thorling, K., Saemundsen, A.K., Bisballe, S., Pallesen, G., Madsen, M., Grazia-Masucci, M., and Ernberg, I. (1980).
Epstein-barr virus induced lymphoproliferative disorder converting to fatal burkitt-like lymphoma in a boy with interferon inducible chromosomal defect.
Lancet. ii : 997-1001.

Tolan, D.R., Niclas, J., Bruce, B.D., and Lebo, R.V. (1987).
Evolutionary implications of the human aldolase-A-B-C and pseudogene chromosome locations.
Am. J. Hum. Genet. 41 : 907-924.

Tommerup, N., Nielson, K.B., Mikkelsen, M. (1981).
Marker X chromosome induction in fibroblasts by FUdR.
Am. J. Med. Genet. 9 : 263-264.

Tommerup, N., Sondergaard, F., Tonnesen, T., Kristensen, M., Arveiler, B. and Schinzel, A. (1985).
First trimester prenatal diagnosis of a male fetus with fragile X.
Lancet i : 870.

Trask, B., Massa, H., Kenwrick, S., and Gitschier, J. (1991).
Mapping of human chromosome Xq28 by two-colour fluorescence *in situ* hybridization and of DNA sequences to interphase cell nuclei.
Am. J. Hum. Genet. 48 : 1-15.

Tsuji, H., Hitomi, A., Takahashi, E., Murata, M., Ikeuchi, T., Yamamoto, K., Tsuji, S., and Hari, T. (1991).
Induction of distamycin A-inducible rare fragile sites and increased sister chromatid exchanges at the fragile site.
Hum. Genet. 87 : 254-260.

Veenema, H., Carpenter, N.J., Bakker, E., Hofker, M.H., Millington Ward, A., and Pearson, P.L. (1987).
The fragile X syndrome in a large family. III. Investigations on linkage of flanking DNA markers with the fragile site Xq27.
J. Med. Genet. 24 : 413-421.

Verkerk, A., Pieretti, M., Sutcliffe, J., Fu, Y., Kuhl, A., Pizzuti, A., Reiner, O., Richards, S., Victoria, M., Zhang, F., Eussen, B., van Ommen, G., Blanden, L., Riggins, G., Chastain, J., Kunst, C., Galjaard, H., Caskey, T., Nelson, D., Oostra, B. and Warren, S. (1991).
Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome.
Cell 65 : 905-914.

Vincent, A., Heitz, D., Petit, C., Kretz, C., Oberlé, I., and Mandel, J.L. (1991).
Abnormal pattern detected in fragile X patients by pulsed field gel electrophoresis.
Nature 349 : 624.

Voiculescu, I., Hausmann, C., Wolff, G., and Backe, E. (1988).
A BrdU - requiring fragile site on chromosome 12.
Hum. Genet. 78 : 183-185.

Voiculescu, I., Back, E., and Schempp, W. (1991).
Homozygous condition for a BrdU - requiring fragile site on chromosome 12.
Hum. Genet. 86 : 416-417.

Vollrath, D., and Davis, R.W. (1987).
Resolution of greater than 5 megabase pair DNA molecules by contour-clamped homogeneous electric fields.
Nucl. Acids Res. 15 : 7865-7876.

Wada, M., Little, R.D., Abidi, F., Porta, G., Labella, T., Cooper, T., Volle, G.D., D'Urso, M., and Schlessinger, D. (1990).
Human Xq24-Xq28 : Approaches to mapping with yeast artificial chromosomes.
Am. J. Hum. Genet. 46 : 95-106.

Wahl, G.M., Lewis, K.A., Ruiz, J.C., Rothenberg, B., Zhao, J., and Evans, G.A. (1987).
Cosmid vectors for rapid genomic walking, restriction mapping and gene transfer.
Proc. Natl. Acad. Sci. 84 : 2160-2164.

Walsh, F.S., Barton, C.H., Putt, W., Moore, S.E., Kelsell, D., Spurr, N., and Goodfellow, P.N. (1990).
N-cadherin gene maps to human chromosome 18 and is not linked to the E-cadherin gene.
S. Neurochem. 55 : 805-812.

Warren, S.T. and Davidson, R.L. (1984).
Expression of fragile X chromosome in human - rodent somatic cell hybrids.
Somatic Cell Molec. Genet. 1- : 409-413.

Warren, S.T., Zhang, F., Licameli, G.R., and Peter, J.F. (1987).
The fragile X site in somatic cell hybrids : an approach for molecular cloning of fragile sites.
Science 237 : 420-423.

Warren, S.T., Knight, S.J.L., Peters, J.F., Stayton, C.L., Consalez, G.G. and Zhang, F. (1990).
Isolation of the human chromosomal band Xq28 within somatic cell hybrids by fragile X site breakage.
Proc. Natl. Acad. Sci. USA 87 : 3856-3860.

Waterbury, P.G., and Lane, M.J. (1987).
Generation of lambda phage concatemers for use as pulsed field electrophoresis size markers.
Nucl. Acids Res. 15 : 3930.

Weber, J.L. and May, P.E. (1989).
Abundant class of human DNA polymorphisms which can be typed using Polymerase Chain Reaction.
Am. J. Hum. Gen. 44 : 388-396.

Weiner, A.M., Deininger, P.L., and Efstratiadis, A. (1986).
Non viral retroposons : Genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information.
Annu. Rev. Biochem. 55 : 631-661.

Wenger, S.L., Hennessey, J.C., and Steele, M.W. (1987).
Increased sister chromatid exchange frequency at Xq27 site in affected fragile X males.
Am. J. Med. Genet. 26 : 909-914.

West, A.K., Stallings, R.L., Hildebrand, C.E., Chiu, R., Karin, M., and Richards, R.I. (1990).
Human metallotheionein genes : structure of the functional locus 16q13.
Genomics 8 : 513-518.

Wilkes, D., Shaw, J., Anand, R., Riley, J., Winter, P., Wallis, J., Driesel, A.G., Williamson, R., and Chamberlain, S. (1991).
Identification of CpG islands in a physical map encompassing the Friedrich's ataxia locus.
Genomics 9 : 90-95.

Williams, B.G., and Blattner, F.R. (1979).
Construction and characterization of the hybrid bacteriophage lambda Charon vectors for DNA cloning.
J. Virology 29 : 555-575.

Wurster-Hill, D.H., Ward, O.G., Davis, B.H., Park, J.P., Moyzis, R.K. and Meyne, J. (1988).
Fragile sites, telomeric DNA sequences, B chromosomes, and DNA content in raccoon dogs, Nyctereutes procyonoides, with comparative notes on foxes, coyote, wolf, and raccoon.
Cytogenet. Cell Genet. 49 : 278-281.

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

Yanamandra, G., and Lee, M.L. (1989).
Isolation and characterization of human DNA in agarose block.
Genet. Anal. Tech. 6 : 71-74.

Yanisch-Perron, C., Vieira, J., and Messing, J. (1985).
Improved M13 phage cloning vectors and host strains : nucleotide sequence of the M13mp18 and pUC 19 vectors.
Genet. 33 : 103-119.

Yu, S., Pritchard, M., Kremer, E., Lynch, M., Nancarrow, J., Baker, E., Holman, K., Mulley, J.C., Warren, S.T., Schlessinger, D., Sutherland, G.R., and Richards, R.I. (1991).
Fragile X genotype characterized by an unstable region of DNA.
Science 252 : 1179-1181.

Yunis, J.J. (1983).
The chromosomal basis of human neoplasia.
Science 221 : 227-236.

Yunis, J.J., and Soreng, A.L. (1984).
Constitutive fragile sites and cancer.
Science 226 : 1199-1204.

ERRATA

1. Page 28, line 2 and page 61, section 2.6.5, line 6.
   The increment of each step in the lambda concatemer is 48.5 kb and not 50 kb as stated.

2. Page 33, line 3.
   Chromosome 16 has been reported to be 98 Mb (Morton, 1991, Proc. Natl. Acad. Sci. 88: 7474-7476). Use this number for the size of chromosome 16.

3. Page 66, line 2.
   The telomere sequences (TEL) in the pYAC vector are derived from *Tetrahymena* and not yeast.

4. Page 66, section 2.9.2, line 1.
   A first generation CEPH YAC library comprised of YACs containing human DNA inserts of sizes 100-800 kb was screened.

5. Fig 2.2, line 5.
   Change "The source DNA is digested with a restriction enzyme...." to read "The source DNA is partially digested with a restriction enzyme....".

6. Page 67, Section 2.9.4, line 4.
   Change "... 384 colonies per nylon filter (in 96 well microtiter plates) to read ".... 384 colonies per nylon filter (stamped from 96 well microtiter plates)."

7. Page 74, section 2.10.1, line 8.
   The statement "To aid in restriction mapping, the vectors contained selectable genes for gene transfer in eukaryotic cells (i.e. Amp and SV2 Neo)" is inaccurate. SV2Neo is in the vector to aid in functional analysis not restriction mapping, by providing a selectable marker for transfection into mammalian cells. Amp is a selectable marker for growth in *E. Coli*.

8. Page 76, section 2.10.3.1, line 7.
   Change the sentence "Probes were labelled with ^{32}P to a specific activity of $\rangle10^8$ cpm/μg by primer extension labelling (2.5.4.2) and prereassociated (2.5.4.4) prior to hybridization of the probes to the filters." to read "Probes were labelled with ^{32}P to a specific activity of $\rangle10^8$cpm/μg by primer extension labelling (2.5.4.2) and prereassociated with an excess of sonicated total human DNA (2.5.4.4) prior to hybridization of the probes to the filters".

9. Page 116, section 4.3.4.
   The probes that were used in the Southern blot filters of restricted YAC DNA for confirmation of the integrity of human YAC inserts for YACs 275H10 and 377C4 were radiolabelled plasmids containing probes for D16S163 and D16S174 respectively.

10. Page 140, section 5.4.1, line 3 and page 141, section 5.4.5, line 3.
    Change "Los Alamos National Laboratories" to read "Los Alamos National Laboratory".