Analysis of the centromeric region of the human Y chromosome

by

Katrina Cooper

Thesis submitted for the degree of Doctor of Philosophy at the University of Oxford

Trinity College
Trinity term 1992



Abstract

Katrina Cooper
Trinity College

Submitted for the
degree of Doctor
of Philosophy
Trinity Term 1992

Analysis of the centromeric region of the human Y chromosome

The centromere is an important region of the chromosome which ensures correct segregation at cell division. The DNA sequences which make up human centromeres are poorly understood. An analysis of the human Y chromosome centromere DNA has therefore been undertaken. The structures of 23 yeast artificial chromosomes (YAC) clones and 4 cosmid clones have been determined and these have contributed to a map of ~7Mb of DNA which span the centromere.

The centromeric region of the human Y chromosome contains a single major block of tandemly repeating alphoid DNA which is variable in size. The 5.7kb alphoid subunits are all orientated in one direction and become diverged at the edges of the array. Flanking the alphoid DNA are small blocks of other known tandemly repeated sequences, the 5bp, 48bp and 68bp satellites. These satellites are arranged in an asymmetric manner and are interspersed with a range of low to moderate copy number repeats. Only one putative single copy sequence has been detected. Nearby lie two regions of X-Y homology: a more proximal region which contains a gene (amelogenin) and a more distal region which has previously been shown to result from a recent X-Y transposition event.

These results show that the centromeric region of the human Y chromosome is a complex mosaic of tandem repeats and other repeats. Furthermore, they provide a detailed map of the region and thus provide a solid basis for functional studies of candidate centromere determining sequences.

ACKNOWLEDGEMENTS

I would like to thank Chris Tyler-Smith for all his help, encouragement and supervision of this thesis and guidance over the last three years. Also I thank him for all his endless time and energy in helping me to prepare talks for conferences. My thanks go to E. M. Southern for his constructive criticism and suggestions of my research and for reading this thesis. I thank the Medical Research Council for funding me during the three years and the MRC, Graduate studies, Trinity College and E. M. Southern for supplying funds for attending conferences.

I would also like to thank Leslie Taylor for growing all the cell lines I needed for this research in tissue culture, Martin Johnson for mending the numerous bits of equipment I "found" broken in the lab and for being the excellent resident bike mechanic and Kirsty Kvaløy and Alison Murray for endless discussions over long coffee breaks. Also I would like to thank Zoia Larin and Steve Taylor for many lively discussions and for making 4.11 a much more enjoyable place to work in over the last year.

My thanks also extend to Louis Quesnel, tutor and friend at Manchester University (undergraduate degree), whose faith in my abilities far exceeded my own and who enthused me to pursue science as a career. Also unlimited amounts of thanks go to Simon Cutting who first introduced me to the rigours of scientific research and for all his support, help and encouragement over the last 4 years. Finally, I would like to thank my parents, Jean and David, and all my family for their unlimitless encouragement throughout the years I've spent studying in Manchester and Oxford for both degrees.

CONTENTS

Tittle page I

Abstract II

Acknowledgements III

1 **Introduction** 1
1.1 The centromere 1
1.1.1 Structure 2
1.1.1.a Kinetochore 2
1.1.1.b Centromeric DNA 4
1.1.1.c Microtubule attachment 4
1.1.2 Function 4
1.2 Structure and function of yeast centromeres 8
1.2.1 The budding yeast, *Saccharomyces cerevisiae* 8
1.2.1.a Structure and centromeric activity 8
1.2.1.b Structural analysis of *S. cerevisiae* centromeric DNA. 8
1.2.1.c Functional analysis of *S. cerevisiae* centromeres 9
1.2.2 The fission yeast, *Schizosaccharomyces pombe* 12
1.2.2.a Structure and centromeric activity 12
1.2.2.b Structural analysis of *S. pombe* centromeric DNA 12
1.2.2.c Functional analysis of *S. pombe* centromeres 14
1.3 Higher eukaryotic centromeres 15
1.3.1 *Drosophila* 15
1.3.1.a Structural analysis of centromeric DNA 16
1.3.1.b Functional analysis of *Drosophila* centromeres 16
1.3.2 Mouse 18
1.3.2.a The mouse major satellite 18
1.3.2.b The mouse minor satellite 18
1.3.2.c The role of the minor satellite in centromere function 20
1.4 Centromeres of *Homo sapeins* 21
1.4.1 Structure of the human kinetochore 21
1.4.2 Structure of human centromeric DNA 23
1.4.2.a Centromeric tandem repeats present in the human genome 23
1.4.2.b Alphoid satellite DNA 25
1.4.3 Human centromeric proteins 26
1.4.3.a Satellite binding proteins 26
1.4.3.b Other centromere-associated proteins 28
1.4.4 Functional analysis of human centromeric DNA 29
1.4.2.a Methods 29
1.4.2.b Candidate DNA sequences 31
1.5 The human Y chromosome centromere 32

1.5.1 Structure of DNA at the human Y chromosome centromere.. 33
1.5.1a Identification of satellite sequences present..... 33
1.5.1.b Structure and organisation of Y chromosome alphoid DNA..... 33
1.5.1.c Proteins..... 35
1.5.2 Definition of the human Y chromosome centromeric region.. 37
1.5.3 Long range map and determination of the DNA structure of the centromeric region..... 37
1.6 The use of YACs to determine the structure of genomes..... 40
1.7 Project aims..... 42

2 Materials and Methods..... 44
2.1 Cell lines and tissue culture..... 44
2.2 Probes..... 44
2.3 Preparation of DNA..... 49
2.3.1 Genomic DNA from tissue culture cells..... 49
2.3.2 Yeast artificial chromosome DNA..... 49
2.3.3 Plasmid DNA..... 50
2.3.4 Cosmid DNA..... 50
2.4 Analysis of DNA..... 50
2.4.1 Restriction enzyme digests..... 50
2.4.2 Gel electrophoresis..... 51
2.4.2a Pulsed-field gels..... 51
2.4.2.b 0.5 -1% conventional agarose gels..... 51
2.4.2.c 4% agarose gels..... 52
2.4.3 DNA transfer, hybridisation and autoradiography..... 52
2.4.3a Hybond N+..... 52
2.4.3.b Nitrocellulose..... 53
2.4.3.c The Washington, St. Louis YAC library filters..... 53
2.4.3.d Competition..... 53
2.5 Polymerase chain reaction..... 53
2.5.1 Purification of oligonucleotides..... 53
2.5.2 Precautions taken to avoid contamination during PCR..... 55
2.5.3 PCR conditions..... 55
2.6 YAC library screening and analysis..... 56
2.6.1 Department of Biochemistry, Oxford, YAC library..... 56
2.6.2 Washington, St. Louis YAC library..... 56
2.6.3 CEPH library, Paris..... 56
2.7 Analysis and subcloning of YACs..... 57
2.7.1 Colony purification of YACs..... 57
2.7.2 Mapping YACs..... 57
2.7.3 Subcloning into plasmids..... 57
2.7.4 Plasmid end-rescue..... 58
2.8 Screening the cosmid library..... 58
2.9 Sequence analysis..... 59
2.7.1 Sequencing subclones from YAC 17C12C..... 59
2.7.2 STS for CEPH library PCR..... 59

3 Structural analysis of the pericentric long arm region of the human Y chromosome 60
3.1 Search for sequnces located in the centromeric region of the human Y chromosome 60
3.1.1 *Schizosaccharomyces pombe* B, K & L repeats 60
3.1.2. Human repeated sequences 61
3.1.2.a 14L family 61
3.1.2.b H7 family 61
3.1.2..c Lambdaphage CM8 subclones 62
3.1.2. Detection of the 68bp and 48bp satellites (tandem repeats) on the Y chromosome 62
3.1.3.a 68bp satellite 63
3.1.3.b 48bp satellite 65
3.1.3..c 42bp satellite 65
3.2 Localisation of the satellites on the human Y chromosome 65
3.3 Analysis of YAC clones from the region 67
3.3.1. Screening the Biochemistry, Oxford YAC library 67
3.3.1a Isolation of YAC clones 67
3.3.1.b Isolation of YAC 9C3C (telomeric Yq) 69
3.3.2. Detection of 48bp and 5bp satellites in pericentric Yq 71
3.3.2a 48bp satellite 71
3.3.2.b 5bp satellite 71
3.3.3. Restriction site map of the YACs 71
3.3.4. Analysis of plasmid subclones derived from YAC 17C12C 76
3.3.5. Sequence analysis 80
3.4 Analysis of cos 64 82
3.5 Long range mapping of genomic DNA 82
3.6 Summary of chapter 3 86

4 Isolation and selection of YACs from the pericentric region of the human Y chromosome 89
4.1 Screening the CEPH YAC library 89
4.1.1. Primary screening of the 113 pools by PCR 89
4.1.2. Secondary screening of the library 91
4.2 Screening the Washington, St. Louis YAC library 94
4.2.1. Colony hybridisation 94
4.2.2. PCR screening 96
4.3 YAC selection, CEPH library 96
4.3.1. Selection of alphoid YACs for further analysis 96
4.3.1.a Identification of Y alphoid YACs 96
4.3.2.b Alphoid block size of Y alphoid YACs 99
4.3.2. Analysis of YαII positive YACs 101
4.3.3. Analysis of 64a positive YACs 104
4.3.4. Analysis of AMGL positive YACs 105
4.4 YAC selection, Washington, St. Louis library 105
4.5 Summary of chapter 4 111

5 Structural analysis of the pericentric short arm and centromeric region of the human Y chromosome 113
5.1 Ananlysis of YACs from the alphoid contig 113
5.1.1 Colony purification and instability of alphoid YACs 113
5.1.1.a Instabilty 113
5.1.1.b Problems with colony purification of YACs BB8H10/2 and AA15G8/6 116
5.1.2 Restriction site maps of the alphoid contig YACs 117
5.1.2.a YAC281C5 117
5.1.2.b YAC 62C12 117
5.1.2.c YACBB8G/4 117
5.1.2.d YAC AA4A7/11 122
5.1.2.e YAC 13C1F 122
5.1.2.f YAC AA15G8/6 122
5.1.2.g YACs 368A6 and 417D3 128
5.1.2.h Restriction site map of the alphoid contig 130
5.1.3 Further analysis of YACs containing alphoid DNA 130
5.1.3.a Organisation of alphoid DNA in YACs 281C5, 62C12 & 417D3 132
5.1.3.b Alu-alphoid-5bp satellite PCR on YAC417D3 137
5.1.3.c Location of the short arm block of the 48bp satellite 137
5.1.3.d Divergence of alphoid DNA in YACs 139
5.1.3.e Orientation of alphoid DNA 141
5.2 Analysis of YACs from the AMGL contig 145
5.2.1 Restriction site maps of the AMGL YACs 145
5.2.2 Restriction site maps of the AMGL contig 150
5.3 Investigation of the nature of sequences in the pericentric short arm region 152
5.3.1 Plasmid end-rescue of YACs from the alphoid contig 152
5.3.1a pAA15e 153
5.3.1b pAA4e 153
5.3.2 Plasmid end-rescue of YACs from the AMGL contig 153
5.3.3 Analysis of plasmid subclones isolated from YAC 235F3 156
5.3.4 Screening the Y chromosome cosmid library 157
5.3 Summary of chapter 5 159
5.4 Summary of complete results 160

6 Discussion 163
6.1 The use of YACs to analyse the centromeric region of the human Y chromosome 163
6.1.1 Screening YAC libraries 163
6.1.1.a Colony hybridisation 163
6.1.1.b PCR based assays 164
6.1.1.c Under-representation of sequences 165
6.1.2 Analysis of YACs 165
6.1.2.a STS/probe content mapping 165
6.1.2.b Restriction site maps of the YACs 166
6.1.2.c Isolation of YAC ends 167
6.1.2.d Subcloning 167

6.1.3 Problem with YACs..... 168
6.1.4 Conclusions on the use of YACs to analyse centromeres..... 170
6.2 Sequence organisation of the Y chromosome centromere..... 170
6.2.1 Long range map of the Y chromosome centromere..... 170
6.2.2 Sequence organisation..... 173
6.2.2.a The 5bp satellite..... 174
6.2.2.b The 48bp satellite..... 175
6.2.2.c The 68bp satellite..... 175
6.2.2.d Alphoid DNA (the 170bp satellite)..... 176
6.2.2.e Overall structure..... 177
6.3 Implications for structure and function..... 177
6.3.1 Comparison of the structure of the human Y chromosome centromere with other centromeres..... 177
6.3.1a Comparison with human centromeres..... 177
6.3.1.b Comparison with the centromeres of other organisims..... 179
6.3.2 Analysis of naturally occurring rearranged Y chromosomes.... 180
6.3.3 Conclusions about the location of the centromere..... 182
6.3.4 Future work..... 184
6.4 Conclusions..... 186

References 187

CHAPTER 1

INTRODUCTION

1.1. THE CENTROMERE.

During cell division it is usually important that daughter cells contain the same genetic information as the mother cell. There are a number of ways in which this can be achieved. For example, in many prokaryotes all of the genetic information is packaged into a single (circular) chromosome allowing a simple segregation mechanism to ensure that each daughter cell receives one copy of the mother chromosome at cell division. However, all known eukaryotes have multiple chromosomes. This increases the possibilities for genetic variation (through independent assortment of the chromosomes), but requires a more complex segregation mechanism.

Eukaryotic nuclear chromosomes are usually linear. The ends are specialised structures called telomeres which ensure the integrity of a linear chromosome by replicating terminal DNA and protecting it from degradation (Szostak and Blackburn, 1982) Two other *cis* acting DNA elements are also required for the function of a linear chromosome. These are the origins of DNA replication of which there are many per chromosome (Struhl *et al.*, 1979) and centromeres which are responsible for the proper segregation of chromosomes at mitosis and meiosis.

The functions of the centromere at cell division are not understood at a molecular level. It is known that for all linear eukaryotic chromosomes centromeres capture microtubules (Mitchison and Kirschner, 1985) and hold the sister chromatids together on the metaphase plate until anaphase, ensuring the faithful segregation of sister chromatids during mitosis and meiosis II and the segregation of homologous chromosomes in meiosis I . Furthermore, it is suspected that the signal which triggers the metaphase-anaphase transition is located at the centromere, as the centromere is the last point at which the sister chromatids remain attached prior to separation. Therefore, because centromeres are involved in the important, widespread biological phenomenon of mitosis and meiosis, understanding how they function at a molecular level is very important. In addition, understanding the movement and control of chromosomes during cell division in a normal cell may ultimately lead to understanding how anueploidy occurs in the

developing embryo and in neoplastic cells. It may also lead to the development of artificial chromosomes for a wide range of organisms including mammals.

1.1.1. Structure.

The cytological structure of the centromere was first visualised by light microscopy. It was originally described as the region on the chromosome which becomes attached to the spindle at mitosis and meiosis (Schrader, 1953). In most organisms this region corresponds to a narrow constriction on the metaphase chromosome known as the primary constriction. With the advent of electron microscopy it was discovered that a distinct structure for microtubule attachment to the chromosomes existed and this was called the kinetochore. Therefore the region of primary constriction i.e. the DNA is now commonly referred to as the "centromere" and the term "kinetochore" is reserved to describe the structural components (proteins etc) that are assembled on the centromeric DNA. However, despite this classification these terms are still not strictly adhered to.

1.1.1.a. Kinetochore.

There are two principal types of kinetochores: diffuse and localised. Diffuse kinetochores occur in several plant and algae groups. They also occur in certain insects, worms such as *Caenorhrabditis* and some protozoan (Bostock and Summer, 1978; Brinkley *et al.*, 1991). These chromosomes lack a primary constriction and their kinetochore microtubules attach all along the length of the chromosome. Consequently, as the chromosomes move to the poles in anaphase they remain parallel to each other. This movement is called holokinetic and therefore these chromosomes are said to be holocentric.

Localised or discrete kinetochores are those that are confined to a single site along the chromosome. Two types have been described: the "ball and cup" kinetochores are characteristic of higher plants and some insects (Reider, 1982) and when examined by electron microscopy appear as diffuse balls of lightly staining material which are attached to a depression on the surface of the chromosome (Luykx, 1970); trilaminar kinetochores appear in a wide variety of organisms from algae to mammals (including humans). Under electron microscopy they appear as crescent shaped trilaminar plates (Brinkley and Stubblefield, 1966) and will be discussed in more detail in section 1.4.1. Figure 1.1. shows the spatial relationship between these structures in a higher eukaryotic chromosome.

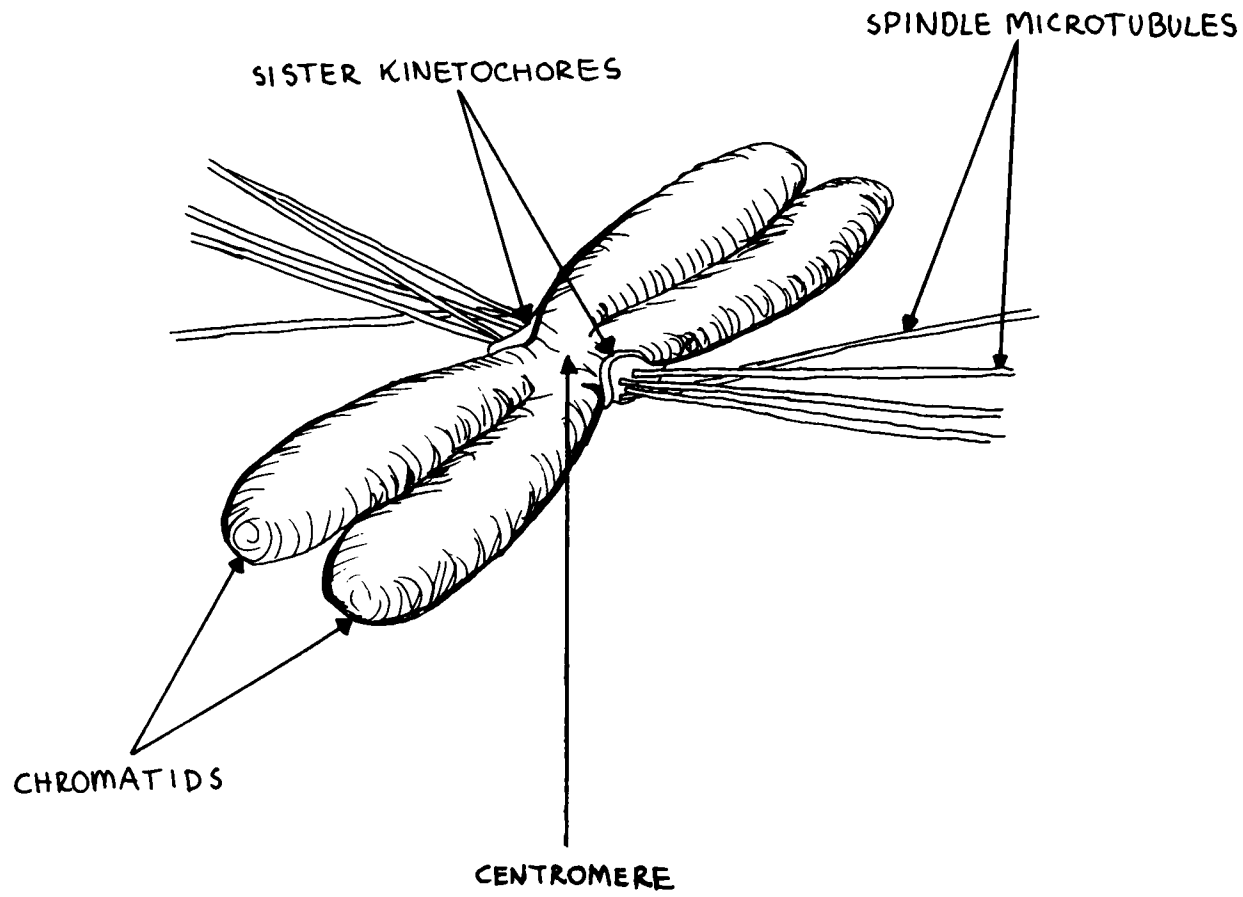


Figure 1.1. Diagramatic representation of the mammalian centromere showing the primary constriction, the trilaminar kinetochore plates and spindle microtubules.

Some organisms e.g the budding yeast *S. cerevisiae*, have no distinct kinetochore and microtubules appear to attach directly to the chromatin fibres (Peterson and Ris, 1976). However, with *S. cerevisiae* it has been suggested that the centromeric chromatin is organised in such a way as to form a "primitive kinetochore" (Bloom and Carbon, 1982). Alternatively, it has been suggested (Brinkley *et al.*, 1991) that in these chromosomes a kinetochore might not be detected because so few microtubules are involved in the attachment site.

1.1.1.b. Centromeric DNA.

The type of DNA present at the centromere also varies between different organisms. In many higher eukaryotes the primary constriction consists of heterochromatin which fails to decondense after cell division, is replicated late in S phase (Vig, 1984) and contains in some organisms (eg. mouse, human, flies) tandem arrays of highly repeated sequences (Willard, 1990; Lohe and Brutlag, 1986). In the fission yeast *S. pombe* repeated sequences are present at the centromere (Clarke, 1990 and section 1.2.2) but the centromeres of the budding yeast are devoid of repeats (Clarke, 1990 and section 1.2.2). Therefore the relationship between centromeric heterochromatin, repeated sequences and centromere structure and function remains unclear.

1.1.1.c. Microtubule attachment.

The number of microtubules per chromosome varies between species. For example in *S. cerevisiae* one microtubule attaches per chromosome (Bloom *et al.*, 1984) and in *H. sapiens* the exact number is not known but it is suggested to be around 15 per chromosome (Pluta *et al.*, 1990). The reason for this variation in number is not as yet understood but has been reported to be related to kinetochore size (Bostock and Summer, 1978).

1.1.2. Function.

The role of the centromere-complex during cell division was first postulated by Schrader (1953) who concluded that "the kinetochore is an element of fundamental importance in movement of chromosomes". Today this view is widely supported (Vallee 1990; Brinkley *et al.*, 1991; Hyman and Mitchison, 1992) and with the advent of new techniques in immunocytochemistry, biochemistry and molecular biology a more detailed understanding of the structure and function of the centromere-complex (including the kinetochore) has been gained. Therefore this section will describe the current

knowledge of the role of the centromere-complex at cell division and chromosome movement. The description given is based on knowledge of mammalian centromeres and it must be emphasised that this is just one example of the centromere-related events occurring at cell division from a huge range of diverse biology.

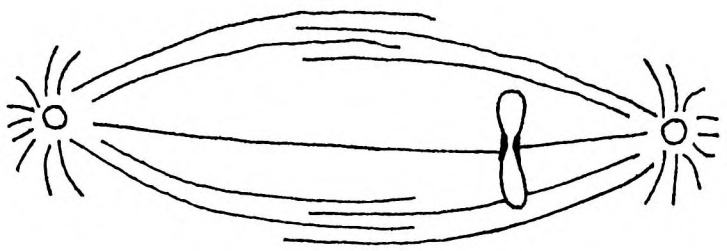
Chromosomes segregate at mitosis by interacting with the microtubules of the mitotic spindle. Originally it was suggested that chromosomes were passive and were pulled to the spindle poles by the attached microtubules. However, it has been shown that the kinetochore is the site of force production during chromosome movement (Mitchison *et al.*, 1986). Two different types of mechanisms have been considered in order to explain how this force is generated. (1) Kinetochores have the ability to both polymerise (Mitchison *et al.*, 1985) and depolymerise (Nicklas 1989; Hyman and Mitchison 1990; Reider and Alexander 1990) microtubules. During this process energy is released from GTP hydrolysis which could drive chromosome movement directly. (2) Kinetochores contain motor molecules capable of moving in both directions along the microtubule (Pfarr *et al.*, 1990; Steuer *et al.*, 1990). The motors are proteins that bind and hydrolyse ATP and generate force in the process. Furthermore, Hyman and Mitchison (1992) examined the movement of microtubules on kinetochores by video microscopy and demonstrated that two different microtubule based motors were present which have opposite directions of movement, the activities of which are controlled by phosphorylation.

Many models of chromosome movement during mitosis have been proposed (Hays and Salmon 1990; Salmon 1991). The latest model of Hyman and Mitchison (1992) takes into account all the evidence discussed above. In this model the kinetochores first attach to a microtubule at prophase and then rapidly slide towards the spindle pole (Figure 1.2.a). This movement is minus-end directed as microtubules possess minus and plus ends (Reider and Alexander, 1990). The attachment of a single microtubule soon matures into the attachment of many polar microtubules (~15 per kinetochore for mammalian cells) though only one is drawn in Figure 1.2 for clarity. From this point on all further movement of the chromosome towards the spindle pole is accompanied by microtubule depolymerisation at the kinetochores, and conversely all movement towards the metaphase plate is accompanied by microtubule polymerisation (Figure 1.2.b). By this point metaphase has been reached and the chromosomes oscillate on the metaphase plate, sometimes

Early prophase



Prometaphase congression



Anaphase A

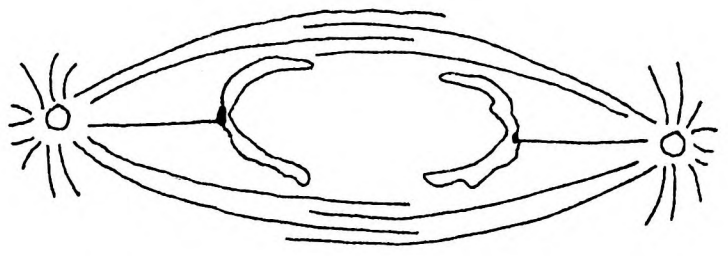


Figure 1.2. Mitotic chromosome movements. In prometaphase, the chromatid pair is first drawn to the spindle pole, then migrates towards the centre of the spindle. At the onset of anaphase, the chromatids separate and migrate slowly towards the spindle poles. (Adapted from Vallee, 1990).

for many hours, using both plus and minus end directed movement to maintain a central position on the spindle. An as yet unknown signal triggers the onset of anaphase (Figure 1.2.c) where the sister chromatids split apart and the sisters move only by minus-end directed movements towards their respective poles.

This model also proposes that two different types of force generating mechanisms play important roles in chromosome movement though the stages at which each force is employed is unknown. The discovery of plus (dynein) and minus (kinesin-like protein) motor molecules (Hyman and Mitchison, 1992) located at the kinetochore which are regulated by phosphorylation has strongly suggested that these motors are important during mitosis. Furthermore, the changes in phosphorylation levels that are known to occur during the cell cycle could not only be controlling whether a plus or minus end directed motor is active but also trigger the metaphase-anaphase transition.

The identification of the these different centromeric activities has raised many basic questions. For example, how are these activities generated at a molecular level? What is the role of centromeric DNA in organising these activities and how is the high fidelity of segregation maintained? Finally, how is the spatial position of the chromosomes on the mitotic spindle regulated? The ability to answer these question in higher eukaryotes has been hindered by the complexity of their centromeric structure and function but recently with the development of *in vitro* assays (Koshland *et al.*, 1988; Hyman and Mitchison, 1992) some progress has been made. Some answers to these questions are now starting to emerge in the budding yeast (discussed in section 1.2.1) where knowledge of centromeric DNA and centromeric proteins is the most comprehensive. However, the marked differences between the structure of centromeric DNA in this yeast and higher organisms and also the relationship between mitosis in higher eukaryotes and mitosis in yeast still remains unclear. Therefore the sections below will discuss what is known about the structure and possible function of centromeres from a range organisms (the evolutionarily distinct budding and fission yeasts, flies, mice and humans). Finally, what is known about the DNA structure of the human Y chromosome and the approaches adopted in this work to analyse it further will be discussed in detail (sections 1.5 and 1.6).

1.2. STRUCTURE AND FUNCTION OF YEAST CENTROMERES.

1.2.1. The budding yeast, *Saccharomyces cerevisiae.*

1.2.1.a. Structure and centromeric activity.

The cell cycle of the budding yeast is unlike that of higher eukaryotes in many ways. For example, it has a extended mitotic phase, the chromosomes do not condense and the spindle persists throughout much of the cell cycle (Nurse, 1985). In addition, during mitosis only one microtubule attaches per chromosome (16 in total (Carle and Olson, 1985)) and no discrete kinetochore structure between the microtubules and centromeric chromatin is seen (Peterson and Ris, 1976).

1.2.1.b. Structural analysis of S. cerevisiae centromeric DNA.

The centromeric DNA of *S. ceresviae* has been identified and the relationship between the DNA sequences and centromere function established (reviewed in Clarke, 1990) The method originally used for *CEN3* was to isolate DNA segments containing centromere-linked genes known to lie on either side of the centromere and to link the two regions by chromosome walking. The centromere must then lie within the cloned region and was identified by its ability to confer stability at mitosis and to initiate reductional segregation at meiosis on automously replicating plasmids introduced into yeast (Clarke and Carbon, 1980). Using similar assays 13 CEN-DNAs of the 16 *S. cerevisiae* chromosomes have been identified. Subsequent analysis has shown that the minimal functional centromere is contained within ~125bp (reviewed in Newlon, 1988) which can be divided into three elements, CDEI, CDEII and CDEIII. CDEI is an 8bp sequence, PuTCACPuTG (Pu, purine), which is completely conserved in all the *S. cerevisiae* centromeres analysed to date. CDEII is ~80bp in length with an unusually high A-T content (>90%) but lacks recognisable conservation of primary sequence. CDEIII contains a 25bp sequence with partial dyad symmetry.

Centromeres from the related yeast, *S. uvarum,* can stabilise minichromosomes in *S. cerevisiae* and contain the conserved sequences elements CDE1-III (Huberman *et al.,* 1986). However centromere DNA from another genus of Ascomycetes, *K. lactis,* does not function in *S. cerevisiae.* The sequences determining centromere function have been cloned using an autonomously replicating plasmid assay in *K. lactis* for 5 of this yeast's 6 chromosomes and the functional domains are 2-5 kb long (Heus *et al.,* 1990).

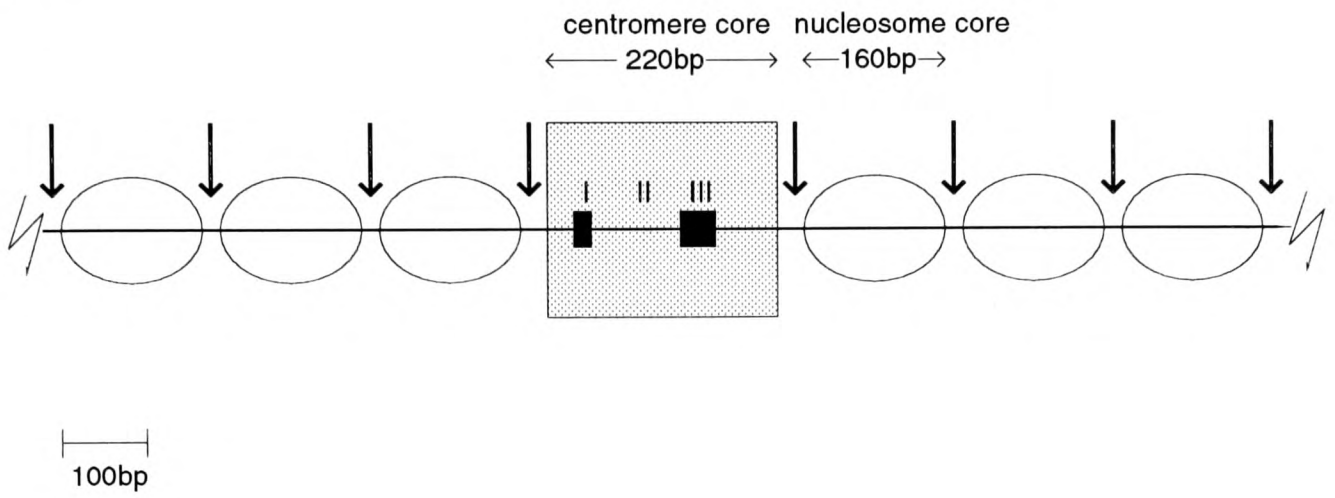
The chromatin structure of the centromeric region has been determined by mapping nuclease-sensitive sites and the consensus centromeric DNA sequences were found to be contained in a 220-250bp nuclease resistant core. This core was also were found to be flanked by hypersensitive cleavage sites and a highly ordered array of nucleosomes (Bloom and Carbon, 1982). This is represented schematically in Figure 1.3. Bloom and Carbon estimated that the core was folded into a chromatin particle 15-20 nm in diameter. Given that microtubules are ~20nm in diameter, and yeast make direct contact with the chromatin fibres (Peterson and Ris, 1976), Bloom and Carbon also proposed that this core might serve as a microtubule attachment site thereby acting as a "primitive" kinetochore. Finally, the contribution of histone proteins to the unique chromatin structure of the nuclease resistant core has recently been investigated by Saunders *et al.* (1990). This group demonstrated that nucleosome depletion by depression of either histone H2B or H4 rendered the cell incapable of chromosome segregation as well as rendering the central core sensitive to digestion.

1.2.1.c. Functional analysis of S. cerevisiae centromeres.

The function of the budding yeast centromere has been addressed by analysing the segregational properties of autonomously replicating plasmids carrying deleted portions of CEN-DNA, or by altering wild-type DNA in the chromosome using fragment-mediated homologous recombination. The results show that sequence alterations in CDEI or CDEII can reduce but not entirely inactivate centromere function (Cumberledge and Carbon, 1987; Gaudet and Fitzgerald, 1987; Hedgemann *et al.*, 1988 whereas point mutations in the central -CGG- of CDEIII completely inactivate the centromere (McGrew *et al.*, 1986; Ng and Carbon, 1987). Furthermore, inversions of CDEIII relative to CDEI and CDEII is deleterious for centromere function. Thus the requirements for centromere function must lie in CDEIII and its spatial orientation relative to the A-T rich CDEII.

Two proteins have been identified that specifically interact with *S. cerevisiae* centromeric DNA: CBF1 (centromere binding factor) and CBF3. CBFI was isolated by several groups (Bram and Kornberg, 1987; Baker *et al.*, 1989; Cai and Davis, 1989; Jiang and Philippsen, 1989) and is a relatively abundant protein (~600 copies per cell) containing a single polypeptide of ~58kd. As its name suggests the protein binds specifically to the 8bp CDEI DNA sequence and is apparently involved in both centromere function and some type of transcriptional control, since CDEI binding sequences have been localised

(a) *S. cerivisiae*
CEN3



(b) *S. pombe*
cen2

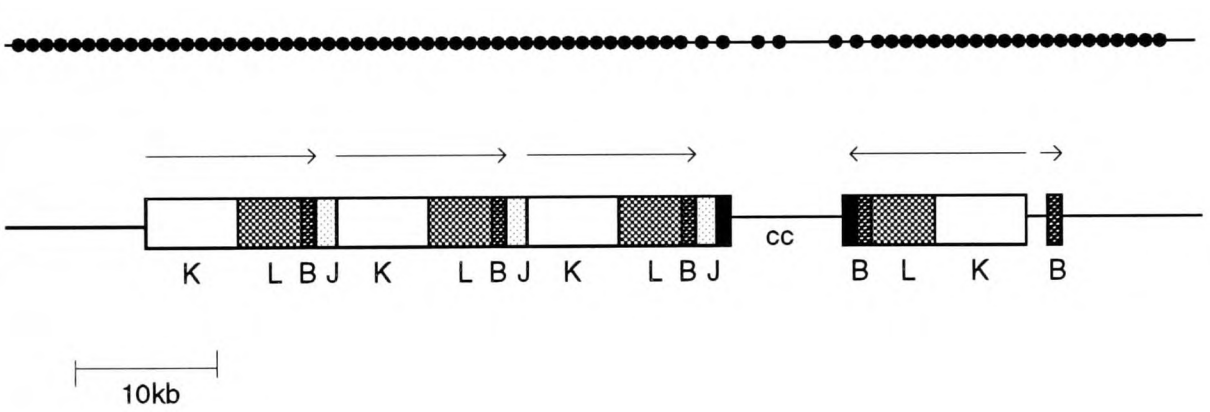


Figure 1.3. Schematic representation of the centromere regions from *S. cerevisiae CEN3* (a) and *S. pombe cen2* (b) adapted from Bloom, 1991 and Clarke, 1990. (a) The relative positions of CDEI, II and III within the protected centromere core are shown along with flanking nuclease-sensitive sites (arrows) and the surrounding ordered nucleosomal array. (b) The organisation and orientation of centromeric repeats K, L, B, and J and the location of the nonhomologous central core (cc) are shown. The solid black boxes directly flanking the central core each represent 1.5kb of inverted repeat sequences that occur only at this location in the *S. pombe* genome. The filled circles drawn above the repeats represent the approximate locations of nucleosomal subunits.

upstream of several genes, and the inactivation of the single gene specifying CBF1 leads to both increased mitotic chromosome loss and auxotrophy for methionine (Cai and Davis, 1990). CBF3 was first isolated by Ng and Carbon (1987) who showed that yeast extracts bind to CDEIII in DNA fragment mobility shift assays. The same group (Lechner and Carbon, 1991) later isolated and characterised three proteins (CBF3A, -B and -C) that form a 240kd complex (CBF3) which binds specifically to yeast wild-type centromeres but not to nonfunctional CEN DNA containing a single base substitution in the critical CDEIII consensus sequence. Furthermore, Lechner and Carbon (1991) also showed that the *in vitro* formation of the CBF3-centromere DNA complex requires the phosphorylated state of at least one of the CBF3 proteins and the presence of a chaperone-like assembly factor.

With a knowledge of CEN DNA sequences and some centromeric binding proteins in *S. cerevisiae* their function at a molecular level in important cellular processes such as binding microtubules and moving chromosomes can be analysed. However, functional analysis of the individual roles of the three centromere specific CBF3 proteins has been hampered by an inability to isolate the individual proteins separately under non-denaturing conditions. In addition, the number of CBF3 molecules in the cell appears to be as low as one copy of CBF3 per centromere. It is though possible that one of the proteins could be similar to the kinetochore associated mechanomotor proteins (dynein and kinesin-like protein) of higher eukaryotes as both types of proteins are regulated by changes in phosphorylation levels and have high ATPase activity (Hyman and Mitchison, 1992; Lechner and Carbon, 1991). Elucidation of the exact functional roles of CBF3 in kinetochore action will require isolating, sequencing and mutagenesis of the genes specifying these proteins. Experiments by Kingsbury and Koshland (1991) have started to address these functional questions. By using an *in vitro* assay measuring the binding of *S. cerivisiae* wild-type and mutated minichromosomes to bovine microtubules they determined that CDEII and CDEIII are absolutely necessary to bind microtubules.

All of these studies suggest that CDEIII sequences are crucial for centromeric function. It could be possible therefore that at a defined point in the cell cycle, in response to changes in phoshorylation levels, the CBF3 complex assembles and either one or more of the proteins attach to CDEIII. This in turn could facilitate the assembly of factors responsible for microtubule binding on CDEII. Alternatively, as yet unidentified factors could bind to CDEII and

stimulate the binding of the single microtubule possibly to the CBF3 complex. Finally, separate factors could bind to CDEII and CDEIII which together form a complex with the microtubule-binding site. However, this uncertainty and the cell cycle regulation of centromere and hence chromosome movements is not understood and further *in vitro* assays systems are needed in order to address these questions.

1.2.2. The fission yeast, *Schizosaccharomyces pombe.*

1.2.2.a. Structure and centromeric activities.

The chromosomes of the fission yeast *S. pombe* are on average 6 times larger than those of *S. cerevisiae* and during cell division behave more like higher eukaryotes (Nurse, 1985). The mitotic cell cycle includes discrete G1, S, G2 and M phases and at mitosis a microtubular spindle forms and the three chromosomes condense and become visible by light microscopy. Also, like higher eukaryotes, a discrete kinetochore can be seen under electron microscopy but the number of kinetochore microtubules is unknown and is probably one or very few (Chikashige *et al.*, 1989).

1.2.2.b. Structural analysis of S. pombe centromeric DNA.

The centromeric regions of *S. pombe* resemble those of higher eukaryotes in containing several classes of repeated sequences that span many kilobases of DNA (Clarke *et al.*, 1986; Fishel *et al.*, 1988). Because of their large size and structural complexity, attempts to isolate functional centromeres by using the automously replicating plasmid assay failed (Clarke *et al.*, 1986; Nakaseko *et al.*, 1986). However, centromeric DNA from all three chromosomes has now been analysed using three different approaches: (1) by chromosomal walking (Clarke *et al.*, 1986; Nakaseko *et al.*, 1986; Fishel *et al.*, 1988), (2) by cleaving at *Not*I sites integrated into chromosomal DNA and determining the long range maps of partially digested fragments (Chikashige *et al.*, 1989), (3) as linear minichromsomes (Hahnenberger *et al.*, 1989; Niwa *et al.*, 1989). Furthermore a plasmid integration assay, has recently been developed to facilitate walking through regions containing repetitive DNA (Clarke and Baum, 1990) and has allowed the structure of the*cen2* region to be analysed in detail.

As a result of using all these techniques the structural organisation of the centromeric DNA of all three chromosomes has been determined by two groups. Unfortunately, the centromeric repeat elements isolated were called B, K, L and J by the American group and dg, dh and yn/tm by the Japanese

group, complicating description of the region. However, the relationship between these different repeats has been determined (dh=portion of K-L, dg=K and yn/tm=B) and the American terminology which is easier to follow will be used in this thesis. The structure of the *cen2* region has been studied in the most detail and the overall arrangements of the different centromeric repeats for this chromosome is shown in Figure 1.3. The region contains four copies of a 14kb tandemly repeating unit. Each unit is composed of three repetitive elements K (6.4kb), L (4.5kb) and B (1-2kb) and related sequences occur at all three *S. pombe* centromeres (Fishel *et al.*, 1988). Another repeat element, designated J, is also present (Clarke, 1990). Finally, two of the 14kb units are arranged in an inverted repeat flanking a non-conserved single copy central core cc2 (7kb in *cen2*) (Fishel *et al.*, 1988). The arrangement of centromeric sequences in *cen1* and *cen3* is similar in that a central core region is present which is also flanked by an inverted repeat configuration of K, L, and B subunits. However, the overall size of the centromeric DNA sequences varies as functional centromere sequences are contained within 65kb and 150kb *Sal*I restrction fragments for the centromere regions of chromosome I and III (Hahnenburger *et al.*, 1989) and within 100kb for chromosome II (Clarke and Baum, 1990). Also a portion of the J repeat is only found within *cen2* and *cen3* but not *cen1* (Clarke 1990). Thus the structural organisation of *S. pombe* centromere regions is complex. Repeats occur that are common to all three centromeres but there are also repeated elements that are specific to a particular chromosome.

Finally, both the American (Kuhn *et al.*, 1991) and the Japanese groups (Takahashi *et al.*, 1991) have determined that a large number of tRNA genes are located in clusters in the B element of *S. pombe* centromeric repeats (see Figure 1.3). Several lines of evidence indicate that these tRNA genes are actively transcribed. For example four tRNAs are not found elsewhere in the genome and all of the centromeric-region tRNAs possess typical structural features known to be common to tRNAs. However, whether these tRNAs are involved in centromeric function has not yet been determined.

The structural organisation of centromeric chromatin of *S. pombe* has recently been determined by nuclease digestion studies (Polizzi and Clarke, 1991). This group demonstrated that the repeated sequences exhibit nucleosome arrays that typify the bulk of the chromatin DNA but found that ordered nuclosomes arrays were absent from the central core. However, nucleosome monomers and dimers are found in the central core. The

authors also determined that the atypical nucleosome pattern remains constant throughout the cell cycle for chromosome 2 and proposed that the unusual arrangement of chromatin at the central core forms the basis of higher order structural differentiation that distinguishes the centromere from the chromosome arms and specifies the essential structure for chromosome function.

1.2.2.c. Functional analysis of S. pombe centromeres.

The function of *S. pombe* centromeric DNA sequences has been determined by analysing the segregation of *S. pombe* minichromosomes containing wild-type or deleted centromeric DNA at mitosis and meiosis. The first minichromosomes were constructed by Niwa and Yanagida (1986) by gamma irradiating unstable strains carrying two copies of chromosome III and selecting for stable, truncated aneuploids. A minichromosome was isolated (Ch16) which was 530kb in length, stable mitotically and segregated 2+:2- through meiosis indicating the presence of a functional centromere. Later, the same group (Niwa *et al.*, 1989) isolated smaller (>125kb) minichromosomes by deleting portions of the left and right hand arms of Ch16. The resulting constructs were analysed and it was determined that deleting potions of the centromeric repeat (left hand arm) leads to loss of the mimichromosome at mitosis and precocious sister chromatid separation at meiosis. The second method of constructing minichromosomes was to clone *S. pombe* centromeric sequences into *S. cerevisiae* artificial chromosome (YAC) vectors (Burke *et al.*, 1987). The entire centromeric regions from all three chromosomes have been cloned in this manner and upon introduction into *S. pombe* host cells, the minichromosomes are faithfully segregated during mitosis and meiosis and are stably maintained without selection (Hahnenberger *et al.*, 1989; Clarke and Baum, 1990b). The deletion analysis of the functional minichromosomes has been carried out by several groups (*cen1* Hahnenberger *et al.*, 1991; *cen2* Clarke and Baum, 1990 and *cen3* Niwa *et al.*, 1989) and has shown that for all three chromosomes deletion of a portion of the inverted repeat results in increased sister chromatid separation in the first meiotic division but has no effect on mitotic stability or the second meiotic division. However, further work by Hahnenberger *et al.* (1991) also showed that all mitotic and meiotic function is abolished when the K" repeat sequence (a subclass of K repeat) is deleted from *cen1*. Furthermore this group also showed that the central core, flanked by subunits B and at least one copy of K, is required for mitotic function in chromosome 1.

Unlike *S. cerevisiae,* no proteins have been identified which interact specfically with the *S. pombe* centromeric sequences. However, the detailed analysis of the *cen1* DNA by Hahnenberger *et al.* (1991) has suggested that K' (another subclass of K) and L repeats are involved in maintaining sister chromatid attachment in meiosis and the K" sequence element is implicated in spindle attachment. To date though the central core and B sequences have yet to be assigned functions. It is possible that the central core sequences also bind microtubules especially as the nucleosome array is disrupted for several kilobases within this region. Therefore to precisely determine the roles of the *S. pombe* sequence elements, an *in vitro* assay sytems needs to be designed. For example it would be interesting to see if the *S. pombe* minichromosomes also bind bovine microtubules and if so which sequences are necessary for this binding.

1.3. HIGHER EUKARYOTIC CENTROMERES.

The centromeres of higher plants and animals are much larger and much more complex than the centromeres of the budding and fission yeasts. An insight into their structure was first gained when Pardue and Gall (1970) and Jones (1970) showed that satellite sequences (tandemly repeated DNA) in mouse chromosomes were associated with constitutive heterochromatin. Thereafter, with the development of a variety of staining procedures e.g. C banding and cd (centromere dot) staining, the presence of heterochromatin in the centromeric and pericentric regions of many chromosomes was discovered. However, the DNA that actually determines centromere function in higher eukaryotes is not known. With the advent of pulsed-field gel electrophoresis (Schwartz and Cantor, 1984) and YAC cloning (Burke *et al.*, 1987) the structure of DNA present at the centromeric regions of many chromosomes is starting to emerge. Therefore the next two sections will discuss the possible DNA structure and relationship to function of the centromeres of an invertebrate (*Drosophila*) and a mammal (mouse). Both of these species have condensing chromosomes, hemispherical kinetochores with bundles of microtubules attached.

1.3.1. *Drosophila.*

The fruit fly *Drosophila* has been studied for many years to address a number of developmental and genetic problems. It is well suited for cytogenetic analysis because the larvae contain large polytene chromosomes thus making the location and analysis of genes easier to study. However, the centromeric

heterochromatin in polytene chromosomes is under-represented and aggregates together to from a chromocenter, thus hampering the analysis of centromeric DNA in these enlarged chromosomes.

1.3.1.a. Structural analysis of centromeric DNA.

The centromeres of *Drosophila melanogaster* consist of heterochromatin which makes up a high proportion (~20%) of the chromosomes (Peacock *et al.*, 1973). As well as the functional centromere, embedded in the heterochromatin are also many important loci; for example X and Y chromosome ribosomal RNA genes and Y chromosome fertility factors (Hilliker *et al.*, 1980). The structural organisation of the heterochromatin has been analysed. Four different satellite DNAs were isolated on CsCl gradients (Peacock *et al.*, 1973) and found to be arranged in long tandem arrays restricted to heterochromatin and chromosomal regions deficient in genes (Peacock *et al.*, 1977). Further cloning of the satellites identified eleven different simple sequences which in each satellite are organised in long, tandemly repeating homologous arrays (Lohe and Brutlag, 1986).

More recently a tandemly repeated DNA sequence which is solely located in the centromeric regions of some *D. melanogaster* chromosomes has been isolated (Abad *et al.*, 1992). Sequence analysis revealed that the repeat has a 11-12bp periodicity and consequently it has been called the dodeca-satellite. Furthermore, at low stringency the satellite was shown to cross-hybridise to five other species of *Drosophila* and to *Arabidopsis thaliana*, hamster, mouse and human DNAs. The hybridisation pattern to human DNA is similar to the pattern detected by the 5bp satellite (Prosser *et al.*, 1986 and see Table 1.1). The biological significance of such satellite DNA cross-hybridisation is not clear. However, by using *in situ* hybridisation the satellite was not detected on all the chromosomes of *Drosophila* suggesting that the dodeca-satellite is not essential for centromeric function. It is possible though that the satellite is present on the other *Drosophila* chromosomes but is either sufficiently diverged from the probe used or present in too small amounts to be detected by *in situ* hybridisation. Therefore the role of the dodeca-satellite in *Drosophila* and other species remains unclear.

1.3.1.b. Functional analysis of Drosophila centromeres.

Despite *Drosophila* being amenable to genetic manipulation, genetic approaches to determining the role of the centromeric heterochromatin have been hampered by the inability to manipulate heterochromatin on a large

scale. Recently Abad *et al.* (1992) characterised a linear 1.3Mb minichromosome (Dp1187) created by Krivshenko and Cooper (1954) (cited in Lindsley and Grell, 1968). Judging from its maintenence in *Drosophila* stocks for several decades it is stably transmitted through mitosis and meiosis. It contains 1Mb of heterochromatin including the dodeca-satellite. This suggests that all the sequences necessary for centromere function are present in the minichromosome and further analysis of its structure will be very revealing.

A few other minichromosomes in *Drosophila* have been recovered. For example, Block *et al.* (1990) isolated a minichromosome (TE225) which is about 1Mb in size and is mitotically stable. This is therefore similar to the Dp1187 minichromosome and suggests that centromeric heterochromatin can be substantially deleted without loss of centromeric function. However, recently Wines et al. (1992) have isolated a mitotically unstable *Drosophila* minichromosome *(Dp(3;f)Th)*. Although the molecular structure of this chromosome has yet to be determined the cytological data suggests that the translocation breakpoint which generated this chromosome occurs deep within pericentric heterochromatin. As a result it seems that in this minichromosome essential centromeric sequences have been lost, resulting in mitotic instability. Molecular analysis of this chromosome will be interesting and also the comparison of the structure of the stable and unstable minichromosomes would be informative.

Genetic approaches also have been used to study the role of mitotic motors. For example, the claret gene product in *Drosophila* is a kinesin like-protein and although the protein has yet to be localised to the kinetochore or centromere, mutations in the *claret* locus lead to high frequency meiotic nondisjunction and chromosome loss (Endow *et al.*, 1990). Similarly mutations in the *nod* locus, which encodes another kinesin-like protein, leads to abnormal segregation of chromosomes at mitosis and meiosis (Zhang *et al.*, 1990). Two proteins D1 and HP-1 have been isolated that appear to interact with centromeric DNA (Alfageme *et al.*, 1980; James *et al.*, 1989). Mutations in HP-1 directly impinge upon heterochromatin physiology and suggest that these proteins play a role in higher levels of heterochromatin condensation and subsequent formation of the kinetochore. However, both proteins interact with A+T rich DNA and are unlikely to bind to the dodeca-satellite which is G+C rich.

1.3.2. Mouse.

Until recently, most of the analysis on the centromeres of the mouse had focussed on determining the DNA structure of the centromeric heterochromatin. However, later work has also concentrated on analysing centromere binding proteins as well as the role of centromere DNA at cell division. All the chromosomes of *Mus musculus* (except the Y chromosome) are generally described as being acrocentric with no cytogenetically detectable short arms (Evans, 1981). The centromeric heterochromatin contains two different satellite DNAs: the major and minor satellites.

1.3.2.a. The mouse major satellite.

The mouse major satellite was one of the first satellites to be isolated and hybridises (*in situ*) to the pericentric region of all chromosomes except the Y (Pardue and Gall, 1970). Later studies have shown that the satellite consists of a 234bp tandemly repeating units (Hörz and Altenburger, 1981) and is located in heterochromatin that immediately flanks the centromere on the long arm (Joseph *et al.*, 1989). However, it is now clear that this satellite lies further away from the centromere than the minor satellite and also lies outside the location of centromere antigens detected by indirect immunofluorescence with anti-centromere antibodies (Wong and Rattner, 1988).

1.3.2.b. The mouse minor satellite.

The mouse minor satellite consists of tandem repeats of a 120bp monomer unit and has been localised to the primary constriction in most species of mouse tested (Pietras *et al.*, 1983; Wong and Rattner, 1988; Wong *et al.*, 1990). Long range physical analysis of *M. musculus* centromeres indicate that the minor satellite is present predominantly as uninterrupted tandem arrays located very close to one end of each chromosome. Furthermore, the array often lies on the same restriction enzyme fragment as telomere associated sequences, thus suggesting that mouse chromosomes are truly telocentric (Kipling *et al.*, 1991). The overall arrangement of the satellites, suggested by all these results is shown schematically in Figure 1.4. If this arrangement of sequences is correct it then very much suggests that the mouse minor satellite may determine sequences which are necessary for centromeric structure except perhaps on the Y.

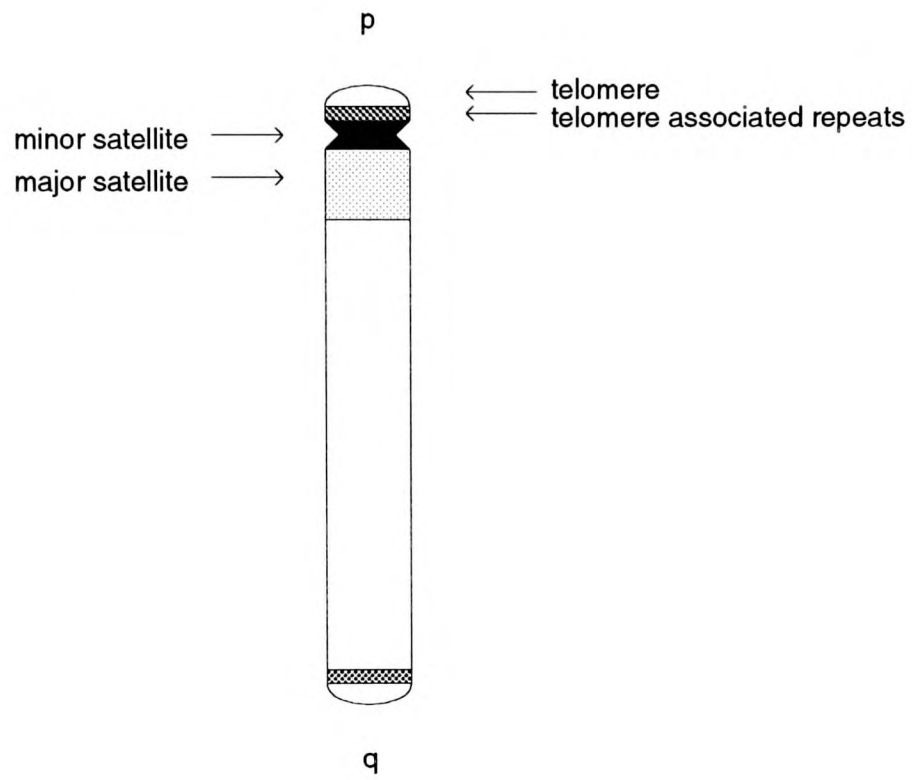


Figure 1.4. Schematic representation of the structural organisation of the mouse major and minor satellites on all the mouse chromosomes except the Y. The short arm has been drawn larger than reality for clarity. Data interpreted from literature cited in text, section 1.3.2.

1.3.2.c. The role of the minor satellite in centromere function.

The experiments of Wong and Rattner in 1988 first suggested that mouse minor satellite sequences play a role in the organization of the kinetochore. Using a two step double-labelling procedure they specifically localised the satellite to the outer lateral surface of the centromere, at or adjacent to the site of anti-centromere antigens. Importantly, the minor satellite remained in discrete patches that coincided with the site of anti-centromere staining even when the region was markedly elongated by treatment with the dye 33258 Hoechst. Further experiments by Masumoto *et al.* (1989) also support the notion that the minor satellite maybe involved in centromere function. These workers identified a 17bp sequence in human alphoid DNA which binds the protein CENP-B. (see section 1.4). The same 17bp sequence, with a single base pair change, is present in the mouse minor satellite consensus sequence. The presence of this conserved protein binding sequence in both humans and mouse repetitive sequences suggests that they have a common function.

However, the role of the mouse minor satellite is not as clear cut as the above paragraph might suggest. The major reason for suggesting that the satellite is not essential for centromere function is that it has not been located on the mouse Y chromosome. However, this could be because the amount of satellite present on the Y is too small to be detected by the methods used. Alternatively, on the Y chromosome the satellite could have become sufficiently diverged from the 120bp consensus sequence for cross-hybridisation not to occur. Whatever the outcome, it makes the Y chromosome particularly interesting.

If the satellite is not directly involved in centromeric function then other sequences must be responsible. Alternatively, the satellite could be involved in conjunction with other sequences. The junction between mouse major and mouse minor satellites has been cloned (Wong and Rattner, 1988) and no other sequences are found in this region. Similarly there seems to be very little "typical" genomic DNA between the minor satellite and the telomere as this region lacks sites for commonly occurring restriction enzymes (Kipling *et al.*, 1991). There could be functionally important sequences embedded in the minor satellite but as the structure of this region is further analysed this possibility becomes less likely. Therefore the role of the mouse minor satellite at cell division remains uncertain.

In summary, the sequences present in the centromeric region of *Drosophila* are just starting to be defined and further analysis of the 1.3Mb minichromosome should rapidly determine which of these sequences are important for chromosome segregation. In the mouse, the structure of the centromeric region has been well defined and at present the mouse minor satellite is the best candidate sequence for determining centromere function. However, in the absence of a minichromsome or an alternative functional assay, this conclusion must remain uncertain.

1.4 CENTROMERES OF *HOMO SAPIENS*.

Despite vasts amounts of literature, in comparison to what is known about the centromeres of yeast and mouse very is little known about human centromeres. Cytogenetic studies have determined that the chromosomes condense during cell division and possess a trilaminar kinetochore. Staining techniques have shown that heterochromatin is present at and near the primary constriction of all human chromosomes as well as in a few other locations. However, improvements in *in situ* hybridisation and immunocytogenetic techniques, the invention of pulse-field gel electrophoresis to obtain long range maps and the development of YACs to clone large amounts of DNA have provided the techniques with which the analysis of human centromeres can begin. Current knowledge of human kinetochores, centromeric DNA, centromeric proteins and the possible roles of these structures at cell division will be discussed below.

1.4.1. Structure of the human kinetochore.

Human centromeres are composed of three structural domains: the kinetochore domain, the central domain and the pairing domain and this is represented diagrammatically in Figure 1.5. In turn the kinetochore is composed of four layers: an electron-dense inner plate that is composed of a specialised layer of chromatin, a lightly staining electron-lucent zone, a dense outer plate that is apparently composed of tightly packed fibres (Ris and Witt, 1981; Rattner, 1991), and an ill-defined fibrous corona, seen most clearly in the absence of microtubules. When the spindle microtubules bind to the chromosome they typically do so by inserting into the outer dense plate although they have also been observed to interact with the fibrous corona and with neighbouring centromeric heterochromatin.

The role of the kinetochore at cell division has been discussed in detail in section 1.1.3. Recently, studies of the kinetochore have focussed on its role in

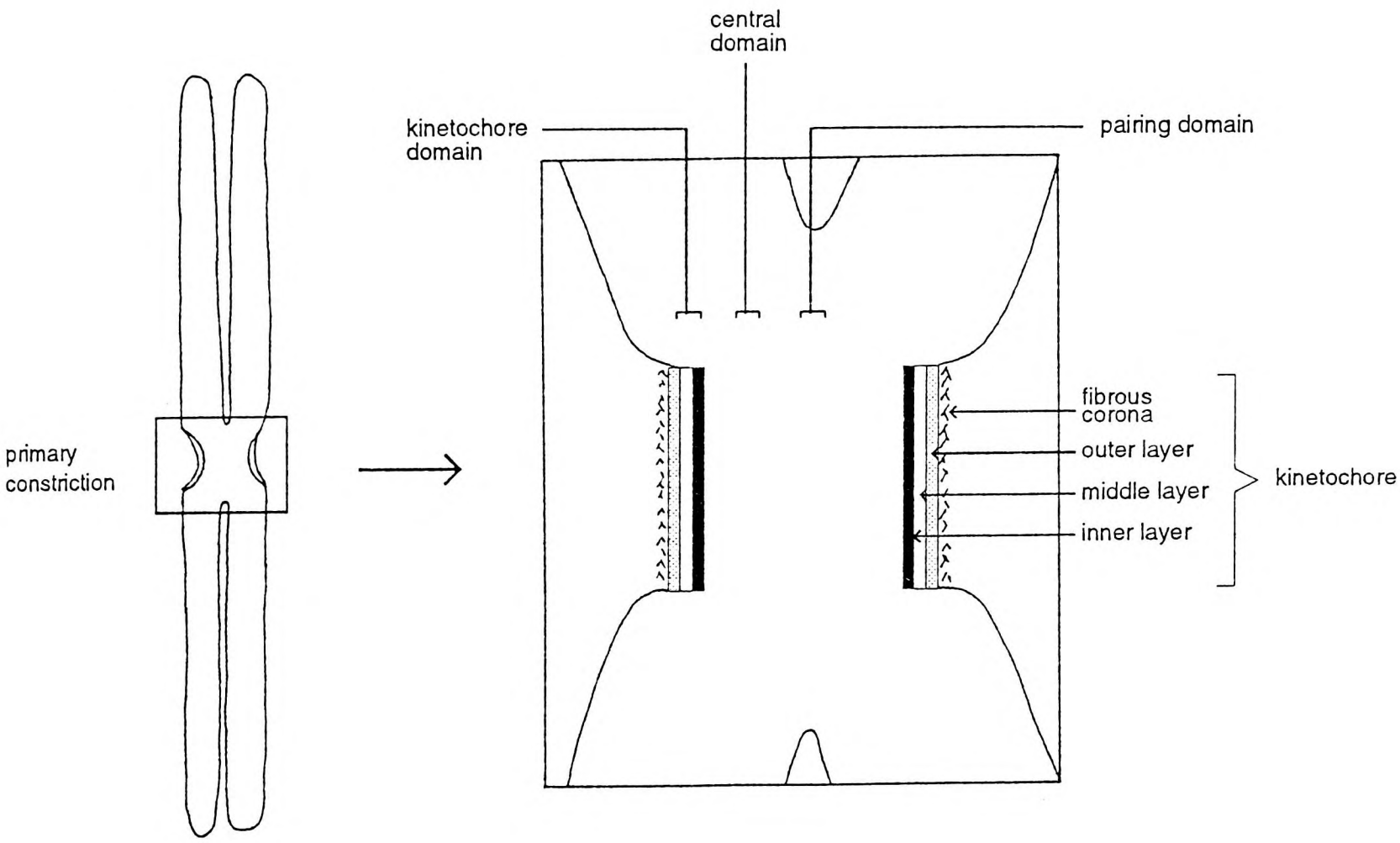


Figure 1.5. Domains of the human centromere-kinetochore complex.

mitotic chromosome movements since it now appears that the mechanomotors for chromosome movement are located in or near its structure. Although the relationship between the kinetochore, centromeric heterochromatin and associated proteins on a molecular level is still not known, experiments of Bernat *et al.* (1991) and Simerly *et al.* (1990) have started to analyse some of these relationships. These groups demonstrated that they could disrupt mitosis by injecting into human cells anticentromere antibodies (see section 1.4.3) that recognise proteins located in the heterochromatin beneath the kinetochore. Furthermore they observed that mitosis was disrupted because the structure of the kinetochore had been altered by the anticentromere antibodies. These experiments show that the proteins recognised by the antibodies (called CENP - section 1.4.3) are essential for the assembly of a functional kinetochore at mitosis.

1.4.2. Structure of human centromeric DNA.

1.4.2.a. Centromeric tandem repeats present in the human genome.

Human satellite DNAs (also called tandemly repeated DNA) were first discovered over twenty years ago as species of DNA, which due to their unusual base composition, band at densities distinct from bulk DNA upon isopycnic centrifugation (Corneo, 1967). Four different satellites (termed classical satellites I-IV) were identified and it was determined that they were localised to the centromeric regions in human chromosomes (Gosden *et al.*, 1975). Later work by various groups demonstrated that these gradient purified satellites consisted of mixtures of several components and determined the structure of many of these components (Prosser *et al.*, 1986) as well as isolating two novel centromeric tandem repeats (48bp and 68bp satellites). However, throughout the literature the nomenclature and relationships of the different satellites are confusing as different authors have used different names for the same sequence. Therefore throughout this work the satellites will be referred to by their different periodicities except for the 170bp satellite which will be referred to as alphoid DNA. Using this classification system five classes of satellites 5bp, 42bp, 48bp, 68bp and 170bp are found; their location determined by *in situ* hybridisation in the human genome is summarised in Table 1.1.

Within many of these families of tandem repeats there occur chromosome specific variants of the unit structure. For example in 1991 Choo *et al.* listed at least 33 subfamilies of alphoid DNA. Some subfamilies were specific for a

Table 1.1. Centromeric tandemly repeated sequences found in the human genome.

repeat unit	satellite	location	reference
5bp	2 + 3	major site in pericentric region of chromosome 9 and long arm of the Y chromosome minor sites near centromeres of chromosomes 1, 5, 7, 10, 12, 13, 14, 15, 16, 17, 19, 20, 21, & X	Gosden *et al.*(1975)
		centromeric heterochromatin of Y chromosome	Frommer *et al.*(1988)
42bp	1	as 5bp repeat but not on chromosome 10, 16, 17, 19 & X or centromeric heterochromatin of Y chromosome	Gosden *et al.* (1975)
48bp		pericentric region of chromosome 22	Metzdorf *et al.*(1988)
68bp	beta	1, 13, 14, 15, 21 & 22 (heterochromatin) pericentric heterochromatin of chromosome 9	Agresti *et al.*(1987) Waye and Willard (1989)
170bp	alphoid (α)	at or close to centromeres of all chromosomes	Mannuelidis (1978); Mitchell *et al.*(1985)

single chromosome while others were common to a small group of chromosomes. However, this data is ambiguous as it is not clear if every subfamily represents the major centromeric array of alphoid DNA. In addition, chromosome specific variants of the 5bp satellite (Cooke and Hindley, 1979; Higgins *et al.*, 1985; Nakahori *et al.*, 1986; Choo *et al.*, 1990) and the 68bp satellite (Meneveri *et al.*, 1985; Waye and Willard, 1989) have been found. In contrast the 48bp satellite has only been reported to be present on one chromosome (Metzdorf *et al.*, 1988; Metzdorf *et al.*, 1990) and whether the 42bp satellite forms subfamilies is unknown. The most studied of these families is alphoid DNA and it will be discussed in more detail in the next section.

1.4.2.b. Alphoid satellite DNA.

Alphoid satellite DNA has been reported to be present at the primary constriction of every human chromosome (Manuelidis, 1978; Mitchell *et al.*, 1985) and for this reason it has been considered as a candidate centromere sequence. It is a primate specific sequence, first isolated in the African green monkey (Maio, 1971) and is based on a monomer repeat unit length of ~170bp which is further organised into multimeric higher order repeat units (Willard and Waye, 1987; Choo *et al.*, 1991). The structure and arrangement of alphoid DNA in the human genome has been studied in great detail and has been reviewed elsewhere (Willard and Waye, 1987; Choo *et al.*, 1991; Willard, 1990) and will not be discussed in detail here. However, a few salient points will be mentioned. (1) In the human genome alphoid DNA is arranged into chromosome specific subsets. These subsets show between 10-40% divergence between the repeat units whereas a high degree of sequence homology exists between independent copies of a higher-order repeat on a particular chromosome. However, chromosomes 13 and 21 have very similar arrangements of their alphoid DNA (Jorgensen *et al.*, 1987) suggesting that the alphoid DNA on these chromosomes is the result of a recent exchange event. (2) Each centromere consists of many copies of individual higher order repeat units which define a chromosomal locus spanning ~250kb to several megabases on different chromosomes (Wevrick and Willard, 1989; Oakey and Tyler-Smith, 1990; Willard, 1990) and the size of the alphoid block on a single chromosome can vary between different individuals (Oakey and Tyler-Smith, 1990). (3) On most chromosomes there is a single block of alphoid DNA which appears to be uninterrupted by other sequences (Oakey and Tyler-

Smith, 1990; Warburton and Willard, 1990) but on chromosome 7 two distinct alphoid arrays exist (Wevrick and Willard, 1991).

1.4.3. Human centromeric proteins.

1.4.3.a. Satellite binding proteins.

Two chromosomal proteins, HMG-I and CENP-B, are known to be associated with alphoid DNA (Solomon *et al.*, 1986; Masumoto *et al.*, 1989). HMG-1 was isolated from monkey CV-1 cells based on its ability to bind *in vitro* to runs of six or more A-T base pairs found in alphoid DNA. However, further work by Disney *et al.* (1989) has shown that the protein is not alphoid DNA specific as it binds to other locations of satellite heterochromatin. Therefore it has been proposed that HGM-I is similar to the HP-I and DI, heterochromatin associated proteins of *Drosophila* (section 1.3.1) and this group of proteins may be involved in heterochromatin condensation and consequent kinetochore formation.

CENP-B is an acidic protein (Earnshaw, 1987) which despite being very similar to HMG-I in chemical composition (Sullivan and Glass, 1991), binds specifically to alphoid DNA. It is one of the five centromere specific proteins, called CENP-A to E (see Table 1.2), which were identified after Moroi *et al.* (1980) discovered that the sera of patients with the autoimmune disease CREST* recognise antigens localised to the centromere. The CENP-B protein was first isolated with CENP-A and CENP-C by characterisation of the antigens recognised by the sera (Earnshaw and Rothfield, 1985). Since then, the identification of antisera specific to CENP-B and the cloning of its gene has allowed the protein to be extensively characterised (Brinkley *et al.*, 1992). However, its function has not yet been unambiguously established.

The relationship between CENP-B, the kinetochore and centromeric heterochromatin has been examined by immunoelectron microscopy and DNA protein binding studies. Results from this work have shown that CENP-B is largely confined to the chromatin of the central domain and not to the trilaminar kinetochore (Cooke *et al.*, 1990) and that it appears to interact with a short 17bp DNA sequence (called the CENP-B box) found in subsets of alphoid monomers (Masumoto *et al.*, 1989). Recent work by Pluta *et al.*, (1992)

---

* Autoimmune disease of progressive systemic sclerosis charaterised by Calcinosis, Raynaud's phenomenon, Esophageal involvement, Sclerodactyly and Telangiectasis.

Table 1.2. Human centromere proteins.

Protein type	size (kDa)	Localisation	Comments	Reference
CENP-A	17-19	centromere	centromere specific histone	Palmer *et al.* (1991)
CENP-B	80	central domain	satellite DNA-binding protein	Masumoto *et al.* (1989)
CENP-C	140	inner kinetochore plate	specific to active centromeres	Saitoh *et al.* (1992)
CENP-D	50	kinetochore domain	may be involved in chromosome condensation	Bischoff *et al.* (1990)
CENP-E	250-300	changes during mitosis	may be involved in regulation of metaphse to anaphse transition	Yen *et al.*(1991)
INCENs	135, 155	inner centromere	sister chromatid association	Cooke *et al.* (1987)
CLIPs	?	pairing domain	sister chromatid linking	Rattner *et al.* (1988)
HMG-I	~10	heterochromatin	heterochromatin condensation	Solomon *et al.* (1986)
kinesin-like	?	kinetochore	mechanochemical ATPase, (plus-end directed)	Hyman and Mitchison (1992)
dynein	205	kinetochore	mechanochemical ATPase, (minus-end directed)	Pfarr *et al.* (1990); Steuer *et al.* (1990)

has also demonstrated that a single domain <158 residues from the terminus of the protein is all that is necessary to localise CENP-B to centromeres. All these results are consistent with the observations by Earnshaw *et al.* (1987) and Wevrick *et al.* (1990) that the apparent abundance of CENP-B on different chromosomes varies in direct correlation with the abundance with alphoid DNA. Furthermore, 16 of the 17 CENP-B box nucleotides are also found in the mouse minor satellite which is otherwise unrelated to alphoid DNA thereby strengthening the case for the mouse minor satellite containing sequences which determine centromere function. However, to throw a different light on the story, antibodies to CENP-B have been located on all human chromosomes except the Y chromosome. Furthermore, the CENP-B box is absent from all Y chromosome specific alphoid DNA repeats sequence to date (Masumoto *et al.*, 1989). Thus CENP-B might have a role in organising the chromatin underlying the kinetochore. However, its apparent absence from the human Y chromosome and also its location at the inactive centromere of a dicentric chromosome (Earnshaw *et al.*, 1989) suggests that it is not essential for the proper segregation of chromosomes at cell division and that the presence of CENP-B alone is not sufficient for centromeric activity.

Finally, in addition to the above properties of CENP-B it has been shown the protein can bind tubulin *in vitro* (Balczon and Brinkley, 1987) but whether this is a non-specific protein-protein interaction has not been made clear.

1.4.3.b. Other centromere-associated proteins

The location and possible functions of the other CENP proteins are summarised in Table 1.2. In addition to CENP-B, the precise location of CENP-C has been determined. Using immuno-electroscopy Earnshaw *et al.* (1992) have demonstrated that CENP-C localises to the inner kinetochore plate. Also this group (Earnshaw *et al.*, 1989) had previously shown that CENP-C is only present at the active centromere of mitotically stable dicentric chromosome. These two observations suggest that CENP-C may be intimately involved in centromeric function.

Table 1.2. also summarises the role and possible functions of two other proteins, INCENPs and CLIPs, which specifically localise to the pairing domain of the primary constriction. The centromere-associated proteins which may be responsible for providing the energy for the movement of chromosomes during cell division are also included.

1.4.4. Functional analysis of human centromeric DNA.

The sequences and some of the proteins involved in centromeric function in the yeasts *S. cerevesiae* and *S. pombe* were identified by analysing the segregational properties of various artificial minichromsomes carrying truncated amounts of centromeric DNA (sections 1.2.1. and 1.2.2) However the lack of such an assay system and the complexities of centromeric DNA have hampered the functional analysis of human centromeres. Three different approaches have been used to try and understand human centromere function and to date three different DNA sequences (discussed below) and numerous proteins (discussed above) have been proposed to be involved in the segregation of chromosomes at cell division.

1.4.4.a. Methods.

The closest comparable assay system to the ones used in yeast has been to analyse naturally occurring human mini- or microchromosomes; chromosome 17 (Anderson *et al.*, 990); chromosome 11 (Haaf *et al.*, 1992) or a artificial mini chromosome 1 (Carine *et al.*, 1989). The chromosomes 17 and 11 mini and microchromosomes were each formed by deletion/excision events from the normal chromosome and are maintained in cell lines as ring chromosomes along with the usual complement of chromosomes. The chromosome 1 minichromosome is autonomous and is the only detectable human DNA in a mouse-hybrid cell line. However, despite these chromosomes being called mini or micro they are still very large (probably many Mb: their sizes have not been accurately determined because it has not been possible to visualise them on a gel) in comparison to the minichromosomes of yeast (a few kb to ~250kb) thereby making their molecular analysis complex. At present their molecular analysis is far from complete but has shown that they all contain alphoid DNA. The articial chromosome 1 minichromsome also contains the 5bp satellite. Furthermore it has been shown that the minichromosome 17 does not contain any detectable euchromatin or telomeres. Also in this minichromosome it seems that the majority of the heterochromatin appears to have been transferred from the deleted chromosome 17 to the supernummary chromosome without interfering with each chromosome's function and survival. This suggests two points: (1) that vital centromeric sequences are retained on both chromosomes (2) that the centromere cannot be a unique sequence since it can be split into two functioning fragments. Thus, these reports support the hypothesis that alphoid DNA is the primary determinant of the human

centromere but until these supernumary chromosomes are analysed further (e.g. pulsed-field mapping data) no more conclusions can be drawn.

Further insight into identifying functional sequences has been gained by analysing dicentric chromosomes. In humans dicentric chromosomes can only be stably maintained if one of the centromeres is inactive. Dicentrics can be detected cytologically as the chromosome has two primary constrictions but the inactive centromere is often smaller than the active centromere (Zinkowski *et al.*, 1986). Interestingly the CENP-B protein is present at both centromeres, consistent with it binding to alphoid DNA whereas CENP-C seems to be restricted to the active centromere (Earnshaw *et al.*, 1989), suggesting that the inactive centromere is devoid of key binding sites for segregation function, or altered in such a way that DNA packaging and/or size precludes visualisation of antibody staining.

An experimental approach to analysing the functional domains of a human centromere has been carried out by Hadlaczky *et al.* (1991). This group isolated a putative centromeric ~14kb fragment by screening a human λ library with DNA extracted from chromatin fragments immunoprecipitated by anticentromere antibodies. The resulting λ clone, CM8, contained a sequence estimated to be present 16-32 times in the genome which was difficult to localise by *in situ* hybridisation but appeared to be located at the primary constriction of all groups of human chromosomes. Sequence comparison showed no homology of λCM8 to any known sequences. To test the possible function of these sequences the clone was cotransfected with a selectable marker into a mouse cell line and surprisingly a cell line was produced that carried the transfected DNA as a functional additional centromere either as a dicentric chromosome or as a minichromosome. Furthermore the extra centromeres bound anti-kinetochore antibodies. Unless the mouse sequences have contributed to the formation of the new centromere, these observations suggest that the sequences contained within the 14kb λ clone can determine centromere function. Further experiments by the same group (Pradnovszky *et al.*, 1991) have shown that the minichromosome, when fused with hamster cells, forms a stable full sized functional chromosome (λneo chromosome). *In situ* hybridisation showed that the λneo chromosome consists of seven tandem arrays of amplified human sequences interspersed with mouse DNA. Interestingly small amounts of the mouse sequences hybridised to the mouse major satellite but only hybridised to the minor satellite if 5 times the usual concentration of probe was used. This later result was therefore considered to

be due to cross-hybridisation. Unfortunately, long range mapping of this novel chromosome has yet to be reported but it will be very interesting to see how a mouse-human chromosome can faithfully segregate when it does not appear to contain either human alphoid or mouse minor satellite DNA, the two most favoured repeats for determining centromere function in man and mouse respectively. Finally, whatever the nature of the DNA sequences used by Hadlaczky's group, their experiments demonstrate that a simple transfection assay can be used to test centromeric function of a DNA fragment.

1.4.4.b. Candidate DNA sequences.

So far three different types of DNA have been proposed to be involved in centromeric function in humans. Alphoid DNA is the most commonly discussed and probably the most favoured candidate as it is present at or near the centromeres of all human chromosomes. In addition subsets of its ~170bp monomers bind to CENP-B, the only heterochromatin protein identified to date. Furthermore it is the only sequence found in all three minichromosomes and also has been shown to be deleted or modified at the inactive centromere of dicentrics (Maraschio *et al.*, 1990). However, as yet no CENP-B has been located on the human Y chromosome and the 17bp binding motif has not been found in the alphoid monomers sequenced to date, thereby arguing that either CENP-B and/or alphoid DNA are not absolutely crucial for a chromosome to be stably maintained.

The second candidate sequence that has been recently proposed by Grady *et al.* (1992) to be a component of the functional human centromere is the 5bp satellite. This group isolated clones from a human plasmid library enriched for repetitive DNA by screening the library at high stringency with mouse and hamster repetitive DNA and discarding clones which contained $(GT)_{25}$ and telomeric sequences. Eight resulting clones were analysed and they all contained the 5bp satellite. Thereafter a 5bp satellite oligomer was used for a number of experiments including cross-hybridsation between different species, thermal stability and DNA mobility shift assays. The results showed that the 5bp satellite is present in many species, located in the chromocenter of *Drosophila* chromosomes and is similar in base pair composition to the centromeric element CDEIII of *S. cerevisiae.* These results together with the observations that the sequence exhibits unusual hydrogen bonding and has a high affinity for nuclear proteins has led the authors to speculate that the 5bp satellite may be a component of the functional centromere. However in the

absence of long range mapping data of an entire centromere these conclusions are rather tentative.

The final candidate sequences for determining centromere function are the sequences contained in the λ clone CM8 discussed above. These sequences are not related to alphoid DNA or any previously reported sequences, apparently localise to the centromeric regions of most human chromosomes and induce the formation of additional centromeres or stable minchromosomes upon transfection into mouse cells. It is possible therefore that the sequence could be a centromere organising element but more extensive structural analysis of the λneo chromosome would be informative.

In summary, alphoid DNA, the 5bp satellite and λ CM8 sequences have been proposed as being important for centromere function. However, the only definitive way of confirming their roles at mitosis and meiosis will be to construct human artificial chromosomes (HACs) with known structures. Truncated and complete versions of each candidate sequence could be tested for centromeric function This idea is discussed in detail in chapter 6.

1.5. THE HUMAN Y CHROMOSOME CENTROMERE.

Cytogenetic analysis of the human Y chromosome has shown that it is one of the smallest human chromosomes (~60Mb) of DNA and in addition to the centromeric primary constriction a large band of heterochromatin is found on the long arm (see Figure 1.9). Most of the reported literature on the human Y chromosome has been focussed on the pseudoautosomal region and the sex-determining region, located at the distal end of the short arm and also on the location and analysis of ~15 male specific genes (HGM 11 report). In addition, a report by Page *et al.* (1984) demonstrated that the human X-Y homology is not limited to to the pairing portions of the sex chromosomes as they located extensive homology between a single-copy locus which maps to the X long arm and the Y short arm. These findings suggest an X-Y transposition during human evolution. Reports on the structural analysis and long range mapping of the human Y chromosome centromere have been very few and most of the knowledge about the structure of the centromere has been determined by our group (Tyler-Smith and Brown, 1987; Tyler-Smith 1987; Tyler-Smith *et al.*, 1988; Oakey and Tyler-Smith, 1990; Neil *et al.*, 1990; Cooper *et al.*, 1992).

The human Y chromosome centromere was chosen for a number of reasons. Firstly, the chromosome is not essential for the survival of a developing

embryo. Consequently many Y chromosome rearrangements and deletions survive and these abnormal chromosomes can be used to form the basis of a deletion map. Secondly, the chromosome is haploid and therefore it can be analysed without constructing somatic cell hybrids to separate the rearranged and normal copies of the chromosome. Finally, the amount of centromeric heterochromatin detected by C banding and the size of the kinetochore is smaller than found on other human chromosomes (Cherry and Johnston, 1987), perhaps making the analysis of the centromere easier.

1.5.1 Structure of DNA at the human Y chromosome centromere.

1.5.1.a. Identification of satellite sequences present.

It can be seen from Table 1.1. that the 5bp satellite and alphoid DNA are the only centromeric tandem repeats that have been reported (by using *in situ* hybridisation) to be present on the human Y chromosome. Furthermore, both these sequences have been reported to be present at the primary constriction. However, long range restriction mapping of Y chromosome alphoid DNA in cell lines that contain the Y chromosome as their only human component has suggested that the 5bp satellite flanks alphoid DNA on the long arm side of the block (Tyler-Smith, 1987).

1.5.1.b. Structure and organisation of Y chromosome alphoid DNA.

The organisation and structure of human Y chromosome centromeric alphoid DNA has been studied by various groups (Willard 1985; Wolfe *et al.*, 1985; Tyler-Smith and Brown, 1987). The details of its structure, summarised in Figure 1.6, were determined by Tyler-Smith and Brown (1987) who analysed nine Y-specific cosmid clones isolated from a Y cosmid library constructed from a Y-only cell hybrid line 3E7 (Marcus *et al.*, 1976) by screening the library with the alphoid clone p82H (Mitchell *et al.*, 1985). Restriction enzyme mapping, subcloning and sequencing of these cosmids revealed that Y chromosome alphoid DNA is made up of 170bp repeating units (monomer units) which are between 76-86% homologous to each other and between 87-92% homologous to the Y chromosome consensus sequence. Restriction enzyme mapping of both the Y cosmids and genomic DNA showed that the monomer units were organised into a higher order structure with a periodicity of approximately 5.7kb. In some individuals, a subset of variant 6.0kb are also found amongst the 5.7kb units.

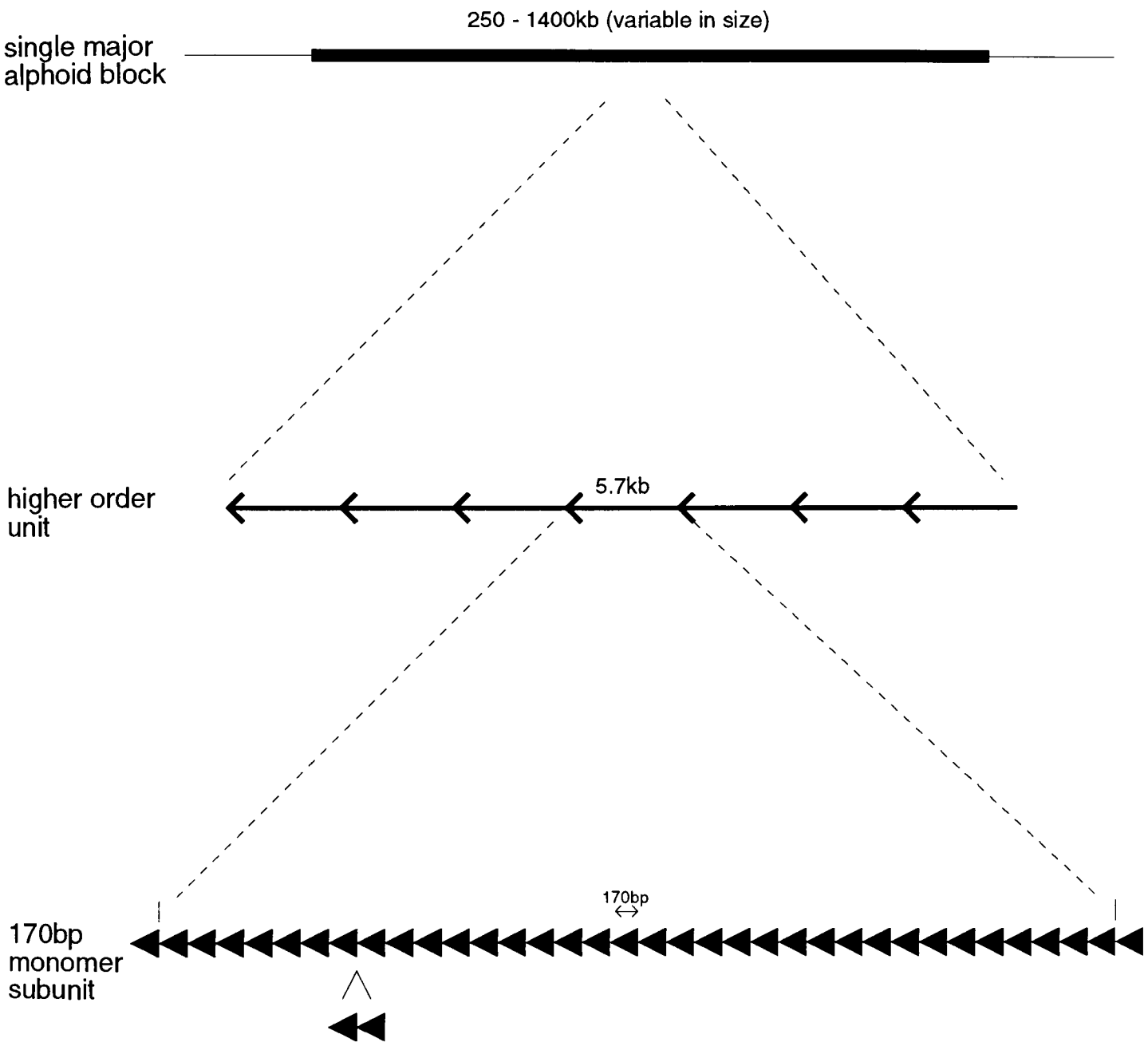


Figure 1.6. Structure of the major block of alphoid DNA on the human Y chromosome. The thick line at the top of the diagram represents the single major block of alphoid DNA found on the human Y chromosome which can vary in size bewteen different individuals. The array consists mainly of tandemly repeating 5.7kb units but some units are 6.0kb long. In turn, the 5.7kb unit consists of tandemly repeating, 170bp monomer subunits and the 6.0kb units have two additional monomer subunits.

The long range structure of alphoid DNA has also been determined (Tyler-Smith, 1987; Tyler-Smith and Brown, 1987; Oakey and Tyler-Smith, 1990). These workers used double digests with restriction enzymes that do not cut within alphoid DNA and pulsed field gel electrophoresis to demonstrate that on the human Y chromosome there is a single major block of alphoid DNA which varies in size between 245-1050kb in different individuals and is uninterrupted by other sequences. Furthermore, detailed analysis of the restriction sites present at the long arm and short arm alphoid-nonalphoid DNA boundaries in 853, one of the Y-hybrids (Burk *et al.*, 1985), showed that the long arm boundary did not have the restriction enzyme distribution expected for typical human DNA. The closest *Apa*I site was 400kb away and *Kpn*I 200kb away; these should cut every 6.4 and 8.5kb respectively. However the short arm boundary contained the expected frequency of these restriction sites for typical euchromatic DNA and therefore was suggested to be the boundary between centromeric heterochromatin and euchromatin (Tyler-Smith, 1987).

Finally, two other small, diverged blocks of alphoid DNA have been located on the human Y chromosome (C. Tyler-Smith, personal communication). These are called YαII and YαIII. At the beginning of this work it was known that YαII was located a few hundred kb from the edge of the major block of alphoid DNA on the short arm and YαIII outside the defined centromeric region (see next section and Figure 1.8) on the long arm of the chromosome. Due to its centromeric location YαII has been analysed and consists of approximately 18kb of alphoid DNA which is flanked by non-alphoid low copy number repeated sequences. The alphoid DNA consists of 170bp monomer units but does not have the 5.7kb higher order unit structure typical of the Y chromosome major block.

Figure 1.7 summarises what was known about the structure of the Y chromosome centromere at the beginning of this work.

1.5.1.c. Proteins.

Apart from HMG-I no proteins have been shown to be specifically associated with Y chromosome centromeric DNA. Even the CENP-B, which has been shown to be present at the centromeric regions of all other human chromosomes has not been shown to be present. The possible explanations for this absence though have been discussed in section 1.4.3.a. However, mixed anti-centromere antiserum does hybridise to the Y chromosome so

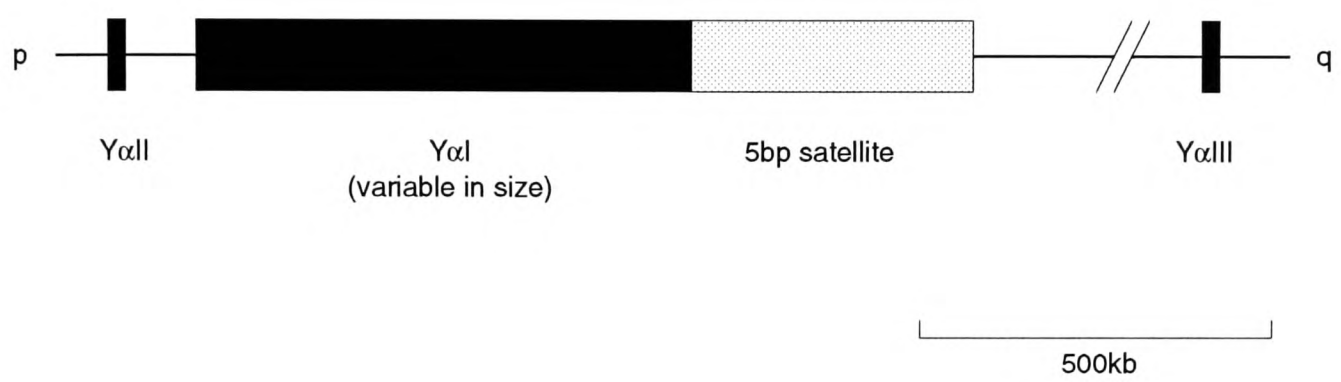


Figure 1.7. Schematic representation of the human Y chromosome centromere interpreted from Tyler-Smith (1987) and C. Tyler-Smith, personal communication. The approximate locations of Y α II and Y α III are also shown.

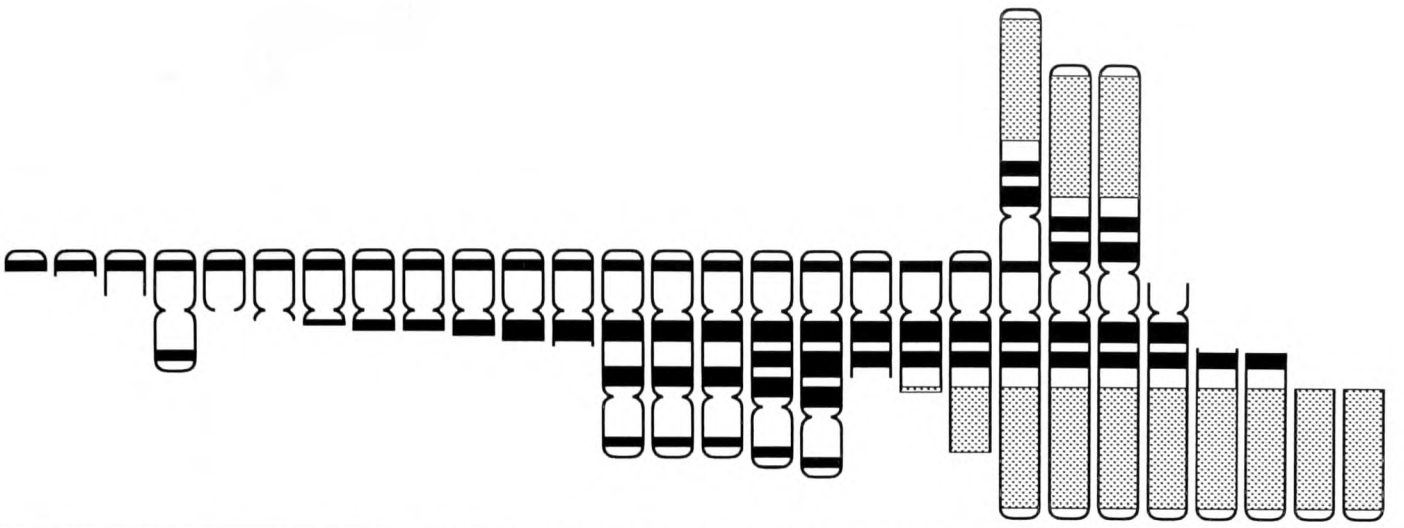
presumably one or more of the CREST proteins (CENP-A, C, D or E) must be present on the Y chromosome.

1.5.2. Definition of the human Y chromosome centromeric region.

In principle, mapping and characterising the Y chromosome centromere is similar to mapping a gene. In practice, it is more difficult because it is not clear how the centromere can be defined and it may not be a unique sequence or a unique region. In order to define a centromeric interval 28 cell lines with rearranged Y chromosomes were analysed with a range of probes known to be present on the human Y chromosome (C. Tyler-Smith, personal communication; Fisher, 1992). The resulting deletion map, shown in Figure 1.8, divides the chromosome into 22 intervals which have an average euchromatic interval size of about 1.5Mb. By comparing the cytological data available on the rearranged chromosomes with the presence or absence of the Y chromosome probes, all the chromosomes which have a Y chromosome primary constriction also contain interval 8. The simple conclusion is that interval 8, in which the major block of alphoid DNA lies, contains the centromere. However, this may be an over-simplification. In Figure 1.8 two criteria are informative for localising the centromere : (1) the presence of a functioning Y centromere in the absence of some Y DNA sequences, and (2) the absence of a functioning Y centromere in the presence of Y DNA sequences. The second criteria are less rigourous because the Y centromere can be suppressed (inactivated) without detectable deletion of DNA sequences (see Chapter 6). If the first, more rigourous, criteria are used, the centromeric interval is defined as shown in Figure 1.9. Cell line A (#651) has a functioning Y centromere but lacks many short arm sequences such as AMY8e. Similarly, cell line C (GM2668) has a functioning Y centromere but lacks many long arm sequences such as RBF5. Thus AMY8e lies outside the centromeric interval on the short arm side, and RBF5 lies outside the centromeric interval on the long arm side. The centromere must lie between AMY8e and RBF5.

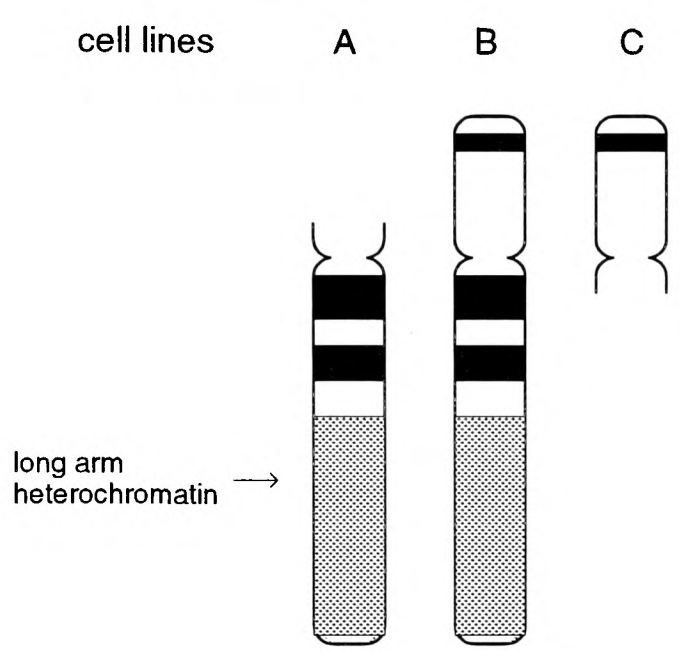
1.5.3. Long range map and determination of DNA structure of centromeric region.

The 5bp satellite and alphoid DNA were first located on the human Y chromosome by *in situ* hybridisation. Thereafter with the advent of pulsed-field gels (Schwartz and Cantor, 1984) and the isolation of rare cutting restriction enzymes long range maps around alphoid DNA have been



	4532	RFY	8903	842	CHoP	KIB	M-6	76R	GM	MSM	TAP	BIT	80R	WSM	863	863	870	DAA	JOW	BRG	GM	861	870	#	GM	GM	GM	GM
		1	0291	968			R-1		2668					184	829	712	740				6967	748	377	651	2103	2469	118	7970
sex	M	M	M	M	M	M	M	M	A	M	M	M		M	M			M	M	M	A	F	F	F	F	F	F	F
ZFY 1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-
47z 2	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-
RBF6 B	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-
LLY25h 3	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-
50f2 B 4	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-
50f2 A	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-
Y-190 maj	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-
35R9 A	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-
5.1 B	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-
RBF2 min	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-
LLY22g ABCF	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-
52d B 5	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-
52d C	-	-	+r	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+r	+r	+r	-	-	-	-
LLY56b	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
55H4	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
35R9 B	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
92R7 6	-	-	+r	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+r	+r	+r	-	-	-	-
GMG XY8	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	-	-	-	-
pDP34	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	-	-	-	-
AMGL	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	-	-	-	-
Y-190 min 7	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	-	-	-	-
50f2 D	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
2D29 A	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
αII	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
αI 8	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
64a	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
pKFC68 9	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
pHY10 min 10	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
14A3C 11	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
RBF5 12	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
GMG Y6 13	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
GMG XY3	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
STSP	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
RBF7	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
25H4 14	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
αIII	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
12H2 15	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
GMG XY3*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	-	-	-
RBF3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	-	-	-
RBF4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	-	-	-
RBF9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	-	-	-
RBF12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	-	-	-
LLY46j	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	-	-	-
52d A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	-	-	-
50f2 E	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	-	-	-
2D29 C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	-	-	-
LLY22g EG 16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	-	-	-
LLY22g D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	-	-
RBF2 maj 17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+r	-	-
Y-216a maj	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	-	-
21A1 18	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	-	-
50f2 C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	-	-
RBF8 19	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	-	-
2D29 D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	-	-
49f	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	-	-
9H4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	-	-
poxY1 A 20	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	-	-
pHY10 maj 21	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+r	+	+	+	+	+	+	+	+	+
pHY2.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+r	+	+	+	+	+	+	+	+
pY9.9 22	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+

Figure 1.8. The 22 interval deletion map of the human Y chromosome derived from all available probes in the laboratory. "r" indicates that the band intensity seen in the autoradiographs of the probe hybridised to that panel member is reduced. Deletion interval 8 contains the centromere.



AMY8e	-	+	+
RBF5	+	+	-

Figure 1.9. Definition of the region containing the Y chromosome centromere. See text, section 1.5.2 for further details.

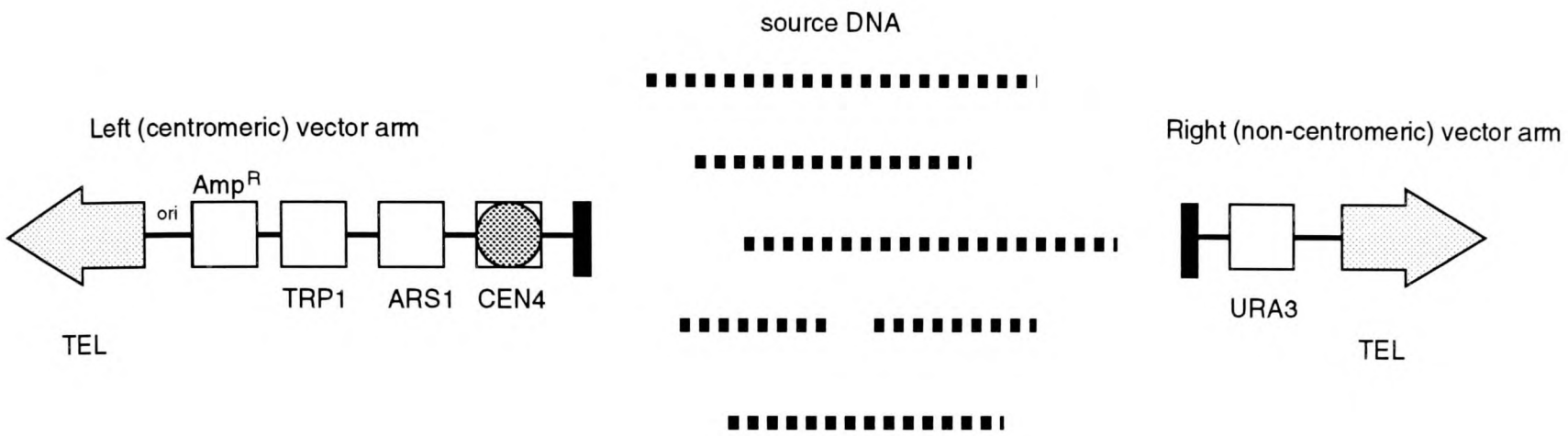
constructed (discussed above) (Fisher 1992; C. Tyler-Smith, personal communication), but none of these maps completely span centromeric region. Obtaining these maps has been complicated for a number of reasons. Firstly, the defined centromeric region of the human Y chromosome is highly repeated (Fisher, 1992; C. Tyler-Smith, personal communication; this thesis) and unique sequences are under-representation or absent. Long range maps can be constructed by using repeated sequences which have a single copy on the Y chromosome using Y-only somatic cell hybrids, but this severely limits the choice of probes and cell lines. Secondly, repeated sequence regions rarely contain the methylation-free HTF islands which allow complete cleavage by many rare-cutter restriction enzymes which are used for mapping. Therefore digestion can produce complex patterns of faint bands which are difficult to interpret. As a result, some mapping data was available at the beginning of this thesis, but there was no extensive long range map.

1.6. THE USE OF YACS TO DETERMINE THE STRUCTURE OF GENOMES.

The yeast artificial chromosome cloning system allows large fragments of cloned DNA to be maintained as linear chromosomes in a yeast host. The basic technique is outlined in Figure 1.10. Large fragments of insert DNA, generated by partial digestion with a restriction enzyme are ligated between two vector arms. Each of these arms ends in a telomere and bears other necessary chromosomal elements: a yeast origin of replication and centromere, as well as markers selectable in the yeast host. This linear construct is then transformed into spheroplasted yeast cells and colonies containing YACs selected on suitable media.

Many groups have constructed large insert YAC libraries (Brownstein *et al.*, 1989; Anand *et al.*, 1989; Albertsen *et al.*, 1990; Larin *et al.*, 1991; Monaco *et al.*, 1991) which have been used to isolate large genes and loci; for example factor IX (Brownstein *et al.*, 1989) cystic fibrosis (Green and Olson 1990b; Anand *et al.*, 1991), Duchenne muscular dystrophy (Monaco *et al.*, 1992). Furthermore, chromosome specific libraries (McCormick *et al.*, 1989), tumour cell-specific libraries to identify breakpoints between chromosomes (Feingold *et al.*, 1990) and telomere specific libraries (Brown 1989) have been constructed as well as libraries from other species: for example *C. elegans* (Coulson *et al.*, 1988), *Drosophila* (Garza *et al.*, 1989) and mouse (Larin *et al.*, 1991).

The use of YACs to analyse complex genomes is relatively recent (Anand, 1992). At the start of this project it was not known to what extent the system



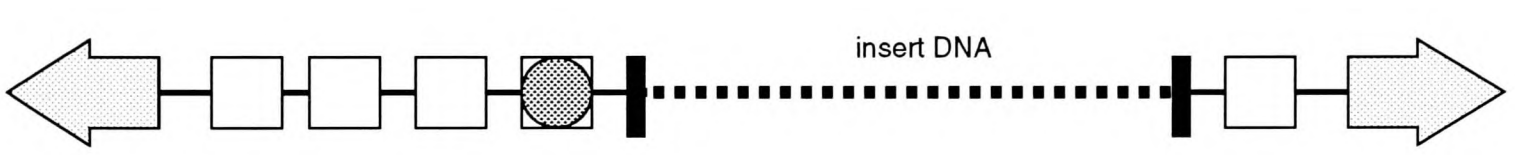


Figure 1.10. Structure of a yeast artificial chromosome (YAC). The exogenous insert DNA is ligated bewteen the vector arms to form a YAC. The arms end in telomere sequences (TEL) and contain a centromere (CEN), a replication origin (ARS) and selectable markers (TRP and URA) to stabilise the YACs in yeast. The two black boxes represent the suppressor tRNA gene SUP4 which is split by insert DNA, thus allowing pink YAC colonies to be recognised. The vector also contains pBR322 sequences (origin of replication and ampicillin resistance) which facilitate the recovery of the left end of the insert as a plasmid in *E. coli.* The diagram is not drawn to scale: the vector arms total about 10kb, the insert hundreds of kb.

would prove to be a broadly useful method for understanding the structure of genomes. For example, the stability of cloned DNA in the YACs, the extent at which the source DNA is randomly sampled, the ease of manipulation of YACs and the biological activity of cloned DNA had yet to be examined. Therefore the first part of this project was spent in helping to construct a small YAC library made from a 49,XYYYY cell line (Fisher, 1992). Thereafter much time was spent in analysing the YACs from this library to determine if YACs from this region would be a useful method of analysing centromeric sequences. By the end of this project many YAC libraries had been constructed and analysed, allowing some of these technical problems to be addressed; these will be discussed in chapter 6..

1.7. PROJECT AIMS.

The aim of this project has been to understand the DNA structure of the human Y chromosome centromere. Two method were adopted: with other workers a long range map of the Y chromosome centromeric interval was constructed (with one gap) in which it was known the functional centromere must lie; concomittently, by using YACs, the sequences present in this interval were analysed. It is hoped that understanding the nature of the sequences present within a defined centromeric interval will guide analysis and experiments which enable DNA sequences to be related to function.

This structural analysis was begun by investigating the nature of sequences in this part of the genome. The approach adopted was to determine whether candidate sequences, isolated from other human and non-human centromeres were present at or near the centromere of the human Y chromosome. The 68bp satellite was found to be located in pericentric Yq so 3 YAC clones from this region were identified and characterised. Also clones were isolated which were used to build up the long range map of the centromeric region.

This analysis (Chapter 3) and a large amount of work in other laboratories demonstrated that YACs were a generally useful method for analysing human DNA. Therefore two further libraries (Albertsen *et al.*, 1990; Burke *et al.*, 1987), which by this time were publically available, were screened using a polymerase chain reaction (PCR) based assay and colony hybridisation respectively, with all suitable probes from the Y chromosome centromeric region (Chapter 4). Characterisation of the YACs obtained (Chapter 5) forms the remaining part of this thesis. Taking into account mapping data from

various abnormal Y chromosome centromeres (C. Tyler-Smith, personal communication) this thesis has allowed the structure of the human Y chromosome centromere to be determined. It has provided the first detailed DNA analysis of a human centromere and it will be interesting to see how other centromeres compare.

CHAPTER 2

MATERIALS AND METHODS

2.1 CELL LINES AND TISSUE CULTURE.

The cell lines used in this work and their sources are listed in Table 2.1. All the human lymphoblastoid cell lines were grown in RPMI 1640 medium (Imperial Laboratories) supplemented with 10 to 20% (v/v) foetal calf serum (Flow laboratories) and antibiotics. The cell hybrid lines used contain the Y chromosome as the only cytogenetically detected chromosome of human origin. Hybrids 3E7 (mouse), 853 (Chinese-hamster) and their respective revertant cell lines were maintained by growing in alpha-modified Eagles medium (Imperial Laboratories) plus 10% foetal calf serum and antibiotics. Q998-8, another Chinese hamster cell line, was a kind gift from Dr. Carol Jones of the Eleanor Roosevelt Institute for Cancer Research, Colorado, U. S. A. and was grown in Ham's F12 medium with 5% foetal calf serum, penicillin (100units/ml) and streptomycin (100µg/ml). The Q998-8 derivative cell line was established by growing the hybrid cells for approximately six months and then testing individual clones for the spontaneous loss of the human Y chromosome long arm by colony hybridisation to the DYZ1 sequence pHY10 (Nakahori *et al.*, 1986). This derivative was found to contain Y chromosome short arm sequences in filter hybridisation experiment (C. Tyler-Smith, personal communication). All the cells lines were maintained by Lesley Taylor ecept FEI and Boleth which were maintained by Maggie Foley and Chris Tyler-Smith respectively.

2.2 . PROBES.

The probes used in this work and their sources are listed in Table 2.2. Novel probes which were created during this work have been listed in Table 2.3. In addition Table 2.3 summarises the location, source of DNA, type of vector, insert excision restriction enzyme and insert size of the novel probes.

The 68bp satellite probe (pKFC68) was constructed using the polymerase chain reaction (PCR). Oligonucleotide primers were chosen from the consensus sequence of Agresti *et al.*, (1989). Both primers contained four additional nucleotides at their 5' ends followed by a *Bgl*II site for primer 68A (CACTAGATCTGTGCAGAGATATGTC) and a *Bam*HI site for primer 68B

Table 2.1. Descriptions and sources of human and hybrid cell lines used in this work.

Cell line	Human Y chromosome content	Source	Reference
OXEN	YYYY	P. Goodfellow, ICRF, U.K.	Sirota *et al.* (1981)
MTL (male)	Y	C. Tyler-Smith, Oxford, U. K.	Tyler-Smith *et al.* (1988)
FEI (female)	-	I. Craig, Oxford, U. K.	
861748	iso(dic)Yq	M. A. Ferguson-Smith, Cambridge, U. K.	Affara *et al.* (1986a) (case 9)
CHoP	Yp	M. Münke, Philadelphia, U. S. A.	Münke *et al.*, (1988)
no. 651	Yp-	J. M. J. Scheres, Nijmegen, Netherlands	Müller *et al.* (1986)
89030291	Yp distal	ECACC	Affara *et al.* (1986b) (cannot establish which XX male it corresponds to)
Boleth	Y	D. Le Paslier, CEPH, France	Ref. 9031, 10th International Histocompatibility workshop.
3E7	Y-hybrid	P. Goodfellow, ICRF, U. K.	Marcus *et al.* (1976)
3E7 revertant	-	C. J. Bostock, Edinburgh, U. K.	
853	Y-hybrid	C. J. Bostock, Edinburgh, U. K.	Burk *et al.* (1985)
853 revertant	-	C. Tyler-Smith, Oxford, U. K.	
Q998-8	Y-hybrid	C. Jones, Colorado, U. S. A.	Emrie *et al.* (1988)
Q998-8 derivitive	Yp	C. Tyler-Smith, Oxford, U. K.	

(GCTAGGATCCCAGGTGATGTAACTC). Target DNA from 1 x 10^5 OXEN cells was amplified using a Perkin-Elmer Cetus thermal cycler in a volume of 100µl containing 10mM Tris-HCl pH 8.3, 50mM KCl, 10mM $MgCl_2$, 500µM dNTPs and 50pmoles of each primer. The DNA was denatured at 94°C for 10 minutes followed by 30 cycles of amplification. Each cycle consisted of denaturation at 94°C for 1 minute, primer annealing at 55°C for 1 minute and extension using Taq polymerase (Cetus, 5 units) at 72°C for 2 minutes. A final extension at 72°C for 10 minutes was added. After amplification, the DNA was phenol extracted, centrifuged through a Sephadex G-25 column (Maniatis *et al.*, 1982) and ethanol precipitated. In order to increase the overall length of the amplified sequence, the PCR product was digested with *Bgl*II and *Bam*HI, purified through a G-25 column and self-ligated overnight at 15°C in the presence of *Bgl*II and *Bam*HI. The products were then ligated into *Bam*HI digested, dephosphorylated pTZ19R vector and used to transform competent *E. coli* XL1 blue cells by electroporation (Dower *et al.*, 1988). The transformation mix was plated onto LB plates (LB is 10g/l tryptone, 5g/l yeast extract, 5g/l NaCl, 1g/l glucose) which contained 100µg/ml ampicillin. Recombinants were detected by colony hybridization using the ligated PCR product as a probe. pKFC68 contains an insert of approximately 0.7kb, and sequence analysis confirmed that it contained variants of the expected 68bp satellite.

The 42bp satellite probe, KFC42, was constructed by synthesizing two complementary but offset 42mer oligonucleotides corresponding to the degenerate A + B consensus sequence of Prosser *et al.* (1986). The sequences were :- SAT1A (ACATAAAATAT G/C AAAGTACCCAA A/G G/C T/C AT A/G T/C ATT A/C TATACTGT); SAT1B (ACTTT C/G ATATTTTATGTACAGTATA T/G AAT A/G T/C AT A/G C/G T/C TTGGGT).The overall size of the probe was extended by annealing the two oligonucleotides together at 37°C for 10 mins in the presence of dNTPs (500µM) and DNA polymerase I (Klenow fragment), followed by overnight ligation. The product was subsequently ethanol precipitated and used as a probe.

Table 2.2 List and source of probes used in this work.

Probe	D number	Source	Reference
Y190 (91H3.5 & 91H4.5)	DYZ4 /DYZ5	C. Tyler-Smith, Oxford, U. K.	Tyler-Smith *et al.* (1988)
GMGXY8	DXYS34Y	ECACC [1]	Affara *et al.* (1986a)
pDP34	DXYS1Y	ATCC [2]	Page *et al.* (1984)
AMGL		M.Snead, Los Angeles, U. S. A.	Lau *et al.* (1989)
2D5		R. Fisher, Oxford, U. K.	Fisher (1992)
50f2		J. Weissenbach, Paris, France.	Vergnaud *et al.* (1986)
YαII 5.1		C. Tyler-Smith, Oxford, U. K.	unpublished results
YαII 2.1		C. Tyler-Smith, Oxford, U. K.	unpublished results
YαI	DYZ3	C. Tyler-Smith, Oxford, U. K.	Tyler-Smith and Brown (1987)
64a		J. Weissenbach, Paris, France	Guellaen *et al.* (1984)
p22hom48.4		ATCC [2]	Metzdorf *et al.* (1988)
14A3C		R. Fisher, Oxford, UK	Fisher (1992)
GMGY6	DYS66	ECACC [1]	Affara *et al.* (1986a)
Blur 8		ATCC [2]	Deininger *et al.* (1981)
pCM1		G. Hadlaczky, Szeged, Hungary	Hadlaczky *et al.* (1991)
pCM6		G. Hadlaczky, Szeged, Hungary	Hadlaczky *et al.* (1991)
pCM11		G. Hadlaczky, Szeged, Hungary	Hadlaczky *et al.* (1991)
pCM16		G. Hadlaczky, Szeged, Hungary	Hadlaczky *et al.* (1991)
pCM21		G. Hadlaczky, Szeged, Hungary	Hadlaczky *et al.* (1991)
pK1620		G. Hadlaczky, Szeged, Hungary	Hadlaczky *et al.* (1991)
pH7-1		J. Doering, Chicago, U. S. A.	Graham *et al.* (1984)
p14L		D. Houseman, Boston, U. S. A.	Glaser *et al.* (1989)
pKH-K		M. Baum, California, U. S. A.	Fishel *et al.* (1988)
pKH-B		M. Baum, California, U. S. A.	Fishel *et al.* (1988)
pKH-L		M. Baum, California, U. S. A.	Fishel *et al.* (1988)

1. European Collection of Animal Cell Cultures
2. American Type Culture Collection

Table 2.3. List of novel probes created in this work.

Probe	Location	Source	Vector	Cloning site(s)	insert size (kb)
pAA3e	pericentric Yp	YAC AA3 B7/13	pBR322[1]	*Eco*RI & *Nde*I	4.0
p296e	pericentric Yp	YAC 296B12	pBR322[1]	*Eco*RI & *Nde*I	1.0
pKFC3	pericentric Yp	YAC 235F3	pTZ19R	*Hind*III	1.4
pKFC5.1	pericentric Yp	YAC 235F3	pTZ19R	*Hind*III	1.1
pAA15e	pericentric Yp	YAC AA15 G8/6	pBR322[1]	*Eco*RI & *Nde*I	1.8
64b	pericentric Yq	cos 64	pTZ19R	*Eco*RI & *Bss*HII[2]	0.6
p15X1	non-Y	YAC 15B12F	pBR322[1]	*Eco*RI & *Xho*I	3.0
pY5	pericentric Yq	YAC 17C12C	pTZ19R	*Eco*RI	2.2
pY48	pericentric Yq	YAC 17C12C	pTZ19R	*Eco*RI	2.5
pY68	pericentric Yq	YAC 17C12C	pTZ19R	*Eco*RI	2.3
pKFC52	pericentric Yq	YAC 17C12C	pTZ19R	*Eco*RI	4.4
pKFC37	pericentric Yq	YAC 17C12C	pTZ19R	*Eco*RI	3.5
pKFC11	pericentric Yq	YAC 17C12C	pTZ19R	*Eco*RI	2.0
pKFC43	pericentric Yq	YAC 17C12C	pTZ19R	*Eco*RI	2.5
pY9.9	telomeric Yq	YAC 9C3C	pTZ19R	*Eco*RI	1.3
pKFC68	many	OXEN 49,XYYYY	pTZ19R	*Eco*RI & *Hin*dIII	0.7
KFC42	many	nucleotides	-	-	-

1. Plasmid-end rescued probes - pBR322 like vector
2. *Bss*HII site in insert

2.3. PREPARATION OF DNA

2.3.1. Genomic DNA from tissue culture cells.

High molecular weight DNA from the cell lines was prepared in agarose plugs by the method of Anand (1986). Cells were harvested, washed in phosphate buffered saline (PBS) and diluted to 2 times the final concentration. Plugs were made at a final concentration of 1 x 10^7 cells/ml, except for 3E7 and 3E7 revertant plugs which were prepared at 5 x 10^6 cells/ml for standard use and at 1.5 x 10^6 cells/ml for use on pulsed field gels. The cells were then mixed with an equal volume of 1% low gelling temperature (LGT) agarose (FMC) in PBS which had been cooled to 37°C and the mixture dispensed into a plug mould. Once the agarose had set the plugs were transfered to a falcon tube containing NDS (0.5M EDTA, 10mM Tris-HCl,1% lauroyl sarcosine, pH 9.5) and 1mg/ml pronase (Boehringer-Mannheim) and incubated overnight at 50°C. The solution was then replaced with fresh NDS and pronase and incubated for a further 24 hours at 50°C. Finally the plugs were rinsed in NDS to remove the pronase and stored in NDS at 4°C.

2.3.2. Yeast artificial chromosome DNA.

Yeast chromosomal DNA minipreps were prepared in agarose plugs by the lithium dodecyl sulphate method (Southern *et al.*, 1987; Anand *et al.*, 1989). Single colonies were used to inoculate 10mls of SD medium (6.7g/l Bacto yeast nitrogen base without amino acids, 20g/l glucose, 55mg/l adenine, 55mg/l tyrosine, 14g/l casamino acids) and grown at 30°C until saturated. Cells were harvested, washed once in 50mM EDTA and resuspended in 450µl of 1M sorbitol, 20mM EDTA solution containing Zymolyase 20T (ICN Biomedicals) at 1mg/ml. An equal volume of 1% LGT in the same solution without Zymolyase was added and the mixture dispensed into a plug mold. Once set, the plugs were transferred to the sorbitol solution with Zymolyase and also 14mM ß-mercaptoethanol and incubated for two hours at 37°C. The solution was then changed for yeast lysis buffer (1% lithium dodecyl sulphate, 100mM EDTA, 10mM Tris-HCl pH 8.0) and the plugs incubated at 50°C. After 1 hour, fresh yeast lysis buffer was added and the plugs were incubated at 50°C overnight. Finally the plugs were stored at room temperature in yeast lysis buffer.

The same method, scaled-up, was used to prepare large quantities of YAC DNA (maxipreps) except that 100mls of SD was inoculated with 100µl of a 10ml overnight culture. Also yeast chromosomal markers (*S. cerevisiae*

YP148) were prepared in the same way except that this yeast was grown in YPD medium (2% glucose, 2% Bactopeptone, 1% yeast extract).

2.3.3. Plasmid DNA.

E. coli cells containing plasmids were grown at 37°C in 2mls of LB broth plus 100µg/ml ampicillin unless otherwise stated. Minipreps of plasmid DNA were prepared using the alkaline lysis method of Birnboim and Doly (1979) except that lysozyme was not used and 5M potassium acetate replaced the sodium acetate. The same method was used to prepare large quantities (500ml cultures) of DNA with one addition. In order to precipitate the RNA an equal volume of 5M lithium chloride was added to the DNA solution, just prior to the ethanol precipitation step, and left on ice for 10 minutes. The mixture was then centrifudged at 8K and the supernatent precipitated twice in 100% ethanol.

Large quantities of plasmid DNA were also occasionally prepared by the ethidium bromide/caesium chloride method as described by Maniatis *et al.* (1982).

2.3.4. Cosmid DNA.

Cos 64 DNA was grown and prepared by the method used for preparing miniplasmid DNA. Large quantities of cos 64 DNA were prepared by the ethidium bromide/caesium chloride method (Maniatis *et al.*, 1982). The LLY cosmids were grown in LB broth containing 20µg/ml kanamycin and prepared by the method of Little (1987) except that the DNA was phenol-chloroform extracted post digestion.

2.4. ANALYSIS OF DNA.

2.4.1. Restriction enzyme digests.

Both genomic and YAC plugs were digested to completion with restriction enzymes as described by Tyler-Smith *et al.* (1988). Prior to digestion the plugs were cut into fragments containing approximately 3.3µg (PFGE) or 2.5µg DNA (conventional gels) and equilibrilated (3 x 30 min) in TE followed by once in 1 x the restriction enzyme buffer recommended by the suppliers (Boehringer, NBL, New England Biolabs). The low concentration 3E7 plugs were cut into fragments containing approximately 1µg DNA. Restriction enzyme digests were then carried out in the buffers using between 10 and 40 units of enzyme

per sample for 4 hours. The digests were always carried out in the presence of 1mg/ml gelatin. After digestion the plugs were again equilibrilated in TE.

YAC plugs were treated in a similar manner for partial digests except that decreasing concentrations of enzyme were used until the conditions were found which gave good partials. The plugs were either digested for 1 or 3 hours, (Yq and Yp YACs respectively).

Plasmid and cosmid DNA was digested to completion according to the enzyme manufacturers instructions. To achieve partial digests of the cosmids decreasing amounts of enzyme were used and the digestion time was exactly 1 hour.

2.4.2. Gel electrophoresis.

2.4.2.a. Pulsed-field gels.

DNA was analysed by pulsed-field gel electrophoresis (PFGE) using a "waltzer" apparatus (Anand, 1986; Southern *et al.*, 1987). 1.5% agarose (Sigma type 1) "waltzer " gels were run in 0.5 xTAE (1 x TAE is 40mM Tris acetate, 2mM EDTA) at 150V with a temperature of 20°C. Preparative 1.5% LGT gels were prepared and run using the same conditions. The 1% agarose gels were also run using these conditions but a different voltage (44V) was used. Pulse times and run times for individual gels are given in the figure legends in the results chapters.

For all PFGE *S. cerevisiae* YP148 chromosomal DNA (Jones *et al.*, 1989) and lambda DNA concatemers (New England Biolabs) were used as size markers. Also for certain pulse times lambda DNA digested with *Hind*IIII was used as a size marker.

2.4.2.b. 0.5 - 1% conventional agarose gels.

Conventional gels were used to analyse DNA fragments smaller than 40kb in size. Digested DNA was electrophoresed in 0.75 or 1% agarose (Sigma type 1) in either a 20cm long vertical gel box at 1.25V/cm (overnight) or a 10cm vertical gel box at 5V/cm (minigel). 0.5% agarose gels were run for 24 hours at 0.5V/cm. All gels were run in 0.5 x TAE buffer with recirculation. LGT agarose conventional gels were run using similar conditions except that horizontal gels were used. Lambda DNA digested with *Hind*III and a 5kb ladder (New England Biolabs) were used as size markers.

2.4.2.c. 4% agarose gels.

4% agarose gels were used to analyse DNA fragments smaller than 400bp in size (PCR products). The gels were run in 3% Nuseive agarose (FMC) and 1% Seakem agarose (FMC) at 10V/cm in a horizontal gel box. The buffer used was 1 x TBE (0.9M Tris-borate, 0.008M EDTA). pTZ19R plasmid DNA was digested with *Hpa*II and used as size markers.

All types of gels were stained with ethidium bromide and photographed by exposing for 2 minutes to Kodak technical pan film 4415 through an interferance filter transmitting at 590nm with a half band width of 10nm. The negative was then developed in a light-tight developing tank using 1/4 strength Kodak developer for 12 minutes, ~5% acetic acid stop solution for 1 minute and 1/4 strength of concentrate of Ilford fix for 10 minutes. Finally, the negative was washed in running water for a further ten minutes and allowed to dry.

2.4.3. DNA transfer, hybridization and autoradiography.

2.4.3.a. Hybond N+.

After ethidium bromide staining and photography, DNA fractionated by conventional gels was transfered (Southern, 1975) to Hybond N+ nylon filters (Amersham) in 0.4M NaOH and hybridised to oligolabelled probes (Feinberg and Vogelstein, 1984). DNA fractionated by PFGE or 0.5-0.75% conventional agarose gels was first depurinated in 0.025M HCl for 15 or 5 minutes respectively and then denatured in 1.5M NaCl and 0.5M NaOH (2 x 25 mins) before transfer. For all Hybond N+ filters the hybridisation was in 5 x Denhardt's (1 x Denhardts is 0.02% bovine serum albumin, 0.02% ficoll, 0.02% polyvinylpyrrolidone), 1% sodium dodecyl sulphate (SDS), 0.08M sodium dihydrogen phosphate, 0.42M disodium hydrogen phosphate, 10% dextran sulphate and 200µg/ml autoclaved salmon sperm DNA in horizontal rotating tubes at 68°C unless otherwise stated. Filters were either washed in 0.1 x SSC or 2 x SSC (1 x SSC is 150mM NaCl, 15mM sodium citrate) at 65°C and exposed to preflashed Fuji Rx X-ray film at -70°C. Signal quantitation was carried out using a Molecular Dynamics phosphorimager and "imagequant" software.

After filters had been exposed to film they were "stripped" by placing into boiling 0.5% SDS and immediately turning off the heat. Once cool, the stripped filters were either exposed to film or checked with a Geiger counter

to determine the effectiveness of the stripping procedure (normally 90-100%). The filters were then either stored at -20°C or rehybridised.

2.4.3.b. Nitrocellulose.

All DNA which was transfered to nitrocellulose filters was first denatured (2 x 25 mins) and then neutralised in 1M Tris-HCl, 1.5MNaCl, pH 7.0. (1 x 25 mins). The transfer medium was 20 x SSC. The remaining hybridisation and autoradiography procedures were identical to the conditions above except that the hybridisation solution was 6 x SET (1 x SET = 150mM NaCl, 20mM Tris-HCl, 1mM EDTA, pH 7.8), 6 x Denhardts, 10% (w/v) dextran sulphate, 0.1% (w/v) sodium pyrophoshate, 0.1% sodium dodecyl sulphate and 200µg/ml autoclaved salmon sperm DNA.

2.4.3.c. The Washington, St. Louis YAC library filters.

The St. Louis, Washington YAC library (Burke *et al.*, 1987) was screened by colony hybridisation. The filters were a kind gift from Ian Dunham, Guy's Hospital, London and were hybridised at 68°C in 6 x SSC, 50mM Tris-HCl pH 7.4, 10 x Denhardts, 10% dextran sulphate, 1% sarkosyl and 200µg/ml autoclaved salmon sperm.

2.4.3.d. Competition.

Human repeated sequences located in a probe were removed by reassociating the repeats with total human DNA (Sealy *et al.*, 1985). The probe was labelled as described above and purified through a Sephadex G-25 column made with TE. 30µl of 20 x SSC was then added to the 30µl of the purified probe and 60µl of prewarmed 20mg/ml autoclaved human placental DNA (Sigma). The probe mix was then boiled for 5 minutes, cooled on ice for 1 minute and finally reassociated at 68°C for 10 minutes before adding to the hybridisation mix.

2.5. POLYMERASE CHAIN REACTION.

2.5.1. Purification of oligonucleotides.

All oligonucleotides were synthesized by Val Cooper, Department of Organic Chemistry, Oxford. They were purified by heating overnight at 55°C in a sealed vial, cooled and the ammonia allowed to evaporate in a fume hood or in a rotary evaporator. They were then resuspended in 100µl sterile double distilled water, extracted three times with butan-1-ol, and precipitated twice using 4 volumes of 100% ethanol and 1/10th volume 3M sodium acetate,

Table 2.4. List of the nucleotide sequences of the PCR primers and PCR conditions used

Name	Sequence	Conditions	Reference
AMGL.A	5' AAATAAAGTGGTTTCTCAAGTGGTC 3'	1	Nakahori *et al.* (1991)
AMGL.B	5' GAGGAAATTATTTGCCCAAAGTTAG 3'	1	Nakahori *et al.* (1991)
YαII.A	5' TGCAAAATGGGGAGTCTCCAATATA 3'	1	C. Tyler-Smith (personal communication)
YαII.B	5' TGATTTCTGCATTTCCCTGTGATGT 3'	1	C. Tyler-Smith (personal communication)
Yα1.A	5' TGAAAACTACACAGAAGCTG 3'	1 & 2	Warburton *et al.* (1991)
Ya1.B	5' ACACATCACAAAGAACTATG 3'	1	Warburton *et al.* (1991)
Yα1.C	5' TGCTTGGAAACTGCACAATG 3'	2	C. Tyler-Smith (personal communication)
64a.A	5' TTCTGGCCAGCTAGCATGAACCCTC 3'	1	
64a.B	5' GGGCCTCTTGTTCCTGCTGAACACA 3'	1	
pBR.A	5' TTGGAGCCACTATCGACTAC 3'	2	Maniatis *et al.* (1982)
pBR.B	5' ATCGGTGATGTCGGCGATAT 3'	2	Maniatis *et al.* (1982)
5SAT.A	5' GGAATGGAATGGAATGGAATGGAAT 3'	3	
5SAT.B	5' ATTCCATTCCATTCCATTCCATTCC 3'	3	
LIH 559	5' AAGTCGCGCCCGCTTGCAGTGAGCCGAGAT 3'	3	Ledbetter *et al.* (1990)
LIH 517	5' CGACCTCGAGATCT(C/T)(G/A)GCTCACTCCAA 3'	3	Ledbetter *et al.* (1990)

1. 1.5mM $MgCl_2$, 30 cycles, 94°C denaturation for 30 seconds, 55°C annealing for 30 seconds and 72°C extension for 1 minute. (CEPH PCR)

2. 2.0mM $MgCl_2$, 30 cycles, 94°C denaturation for 30 seconds, 55°C annealing for 30 seconds and 72°C extension for 2minutes. (alphoid orientation)

3. 2.0mM $MgCl_2$, 30 cycles, 94°C denaturation for 30 seconds, 55°C annealing for 30 seconds and 72°C extension for 2minutes. (Alu-5bp satellite and alphoid PCR)

Table 2.5. List of PCR amplification product sizes.

Primer pairs	Product size (bp)
YαI.A + YαI.B	1000
YαI.A +pBR.A	1300
YαI.A + pBR.B	1000
YαI.C +pBR.A	1300
YαI.C + pBR.B	1000*
YαII.A +YaII.B	222
AMGL.A +AMGL.B	Y=295, X=470
64a.A +64a.B	270

*1800bp product amplified from YAC 17.2 (section 5.1.3.e)

dried and resuspended in 50µl TE. After determining their OD at 260nm they were resuspended at a final concentration of 1mg/ml (100 x stock) by using the formula 1 OD=20µg/ml DNA. A list of the nucleotide sequence of all the oligonucleotides used (PCR primers) is given in Table 2.4 and the product sizes in Table 2.5. Each primer was used at a final concentration of 10µg/ml.

2.5.2. Precautions taken to avoid contamination during PCR.

The following precautions were taken in order to avoid PCR containmination. PCR reactions were always set up in a PCR hood. All solutions, pipette tips and eppendorfs were kept in the hood and were designated PCR items and were never taken out to the general work area. 30 minutes before the reactions were commenced the UV light in the PCR hood was switched on. Finally positive pressure pipettes (Eppendorf) were used.

2.5.3. PCR conditions.

The amplification conditions used to synthesize pKFC68 have been given in section 2.2. The conditions listed below apply for all other PCR reactions. DNA for amplification was prepared as follows. 900µl of double distilled sterile water was added to an agarose plug (~10µg DNA) and melted at 65°C

for 10 minutes. The mixture was vortexed very well and dispensed into 100µl aliquots. This DNA was then treated as 10 x stock. The reactions were performed in 10mM Tris-HCl, 50mM KCl, 0.01% gelatin, 250µM each dNTP (Boehringer-Mannheim) and 1 x primer. They were amplified by 5 units/100µl of Taq polymerase (Cetus). The concentration of $MgCl_2$, the cycle times and temperatures all varied and are listed in Table 2.4.

2.6. YAC LIBRARY SCREENING AND ANALYSIS.

2.6.1. Department of Biochemistry, Oxford, YAC Library.

This YAC library was constructed by R. Fisher (Fisher 1992) from OXEN 49,XYYYY DNA using partial *Eco*RI digestion, the pYAC4 vector and *S. cerevisiae* AB1380 host cells (Burke *et al.*, 1987) as described by Anand *et al.* (1989). Approximately 19 000 clones with an average insert size of 150kb - 200kb (about 2-fold coverage of the Y chromosome) were screened by colony hybridisation (Brownstein *et al.*, 1989) using 8 x offsetting. The hybridisation conditions used were identical to the conditions used for nitrocellulose filters except that the dextran sulphate was ommited and instead of using rotating tubes the filters were hybridised in a plastic box in a water bath. The filters used were immobilon (Millipore). YAC clones identified were colony purified and analysed as described below.

2.6.2. Washington, St Louis YAC Library.

The Washington, St Louis YAC Library YAC library contains 57 000 clones with an average insert size of well over 200kb (about 4.5 - 5 genome equivalents) (Burke *et al.*, 1987). The library was screened by colony hybridisation (section 2.4.3.c.) and the filters were a kind gift from Ian Dunham, Guy's Hospital, London. YAC clones were identified, colony purified and analysed as described below.

2.6.3. CEPH Library, Paris.

The CEPH YAC library contains 50 000 clones with an average insert size of 400kb (about 7 haploid genome) (Albertsen *et al.*, 1990). The library was screened by PCR using the method first outlined by Green and Olson (1990a). 113 primary pools of DNA pellets were obtained from D. Le Paslier, CEPH, Paris, and were resuspended in 50µl TE. 1µl of each pool was used per PCR reaction (30µl). The PCR conditions and primers used are given in Table 2.4. The sequence tagged sites (STS) (Olson *et al.*, 1990) used to screen the library were 64a, Yα1, Yα11, and AMGL. Positive pools were identified and double

checked and then the primers and conditions used were sent to CEPH for the secondary screen by PCR. YAC clones containing the locus were identified by CEPH and either 1 or a few clones sent back. The final identification of positive YACs was done by hybridisation. Finally, the YACs were colony purified and analysed as described below.

2.7. ANALYSIS AND SUBCLONING OF YACS.

2.7.1. Colony purification of YACs.

The YACs were colony purified by growing to saturation in 10mls of SD medium at 30°C and then plating out for single colonies on SD agar. After three days incubation at 30°C, pink, single colonies were isolated, grown to saturation in 10mls of SD medium and DNA prepared as described in section 2.3.2. The YACs were visualised by fractionating the DNA by PFGE and staining the gel with ethidium bromide and daughter clones which contained the same sized YAC as the parent clone were further analysed.

Yeast clones which contained more than YAC per cell were colony purified as described above except that a the YACs were first grown to saturation in YPD medium and then 10µl of the culture was used to innoculate a further 10mls of medium. This was repeated another time before selecting for single colonies on SD agar plates.

2.7.2. Mapping YACs.

The YACs were mapped using the technique of partial end-labelling as first described by Burke *et al.* (1987) which is based on the approach of restriction site mapping outlined by Smith and Birnsteil (1976). YAC DNA was either completely or partially digested with rare cutting enzymes, fractionated by PFGE, transfered to a filter and hybridised to both the L ("left", centromeric) and R ("right", non-centomeric) arm YAC vector probes. Partial digests were used to determine the order of restriction enzyme sites and complete digests to determine the sizes of the fragments. Complete digests were also hybridised to total human DNA and any loci known to be located in the YAC.

2.7.3 Subcloning into plasmids.

YAC DNA was excised from a low melting point agarose PFG and digested to completion with *Eco*RI (YAC 17C12C and 9C3C) or *Hin*dIII (YAC 235F3) in the gel slice. The DNA was purified using a "Geneclean" kit (Stratech Scientific) and ligated overnight at 15°C to *Eco*RI or *Hin*dIII digested, dephosphorylated

pTZ19R vector DNA (Maniatis *et al.*, 1982). The ligated DNA was used to transform competent XL1 blue cells as described in section 2.2. Recombinant clones were picked (768 for YAC 17C12C and 384 for both other YACs) into 96-well micro-titre dishes and stored at -70°C in 20% glycerol. The libraries were plated onto Hybond N+ filters using a 96-well prong device, and screened by colony hybridisation.

2.7.4. Plasmid end-rescue.

Insert DNA from the left hand edge of YAC clones was recovered by using the plasmid end-rescue method described by Burke *et al* (1987). YAC DNA was digested to completion with either *Xho*I or *Nde*I and fractionated by conventional gel electrophoresis. The DNA was transfered to a membrane and hybridised to the L vector arm fragment to determine whether there was a restriction enzyme site near enough the end of the insert to be plasmid end-rescued. If there was, approximately 50µg DNA in agarose plugs was digested with the chosen enzyme, fractionated on a vertical 1% LGT agarose gel and the appropriate sized gel slice excised. The DNA was then purified using a "Geneclean" kit and ligated to itself overnight at 15°C (*Xho*I) or 4°C (*Nde*I) in 200µl using 3 units of T4 ligase (Boehringer-Mannheim) in the presence of ligase buffer recommended by the enzyme manufacturer. The ligation mix was precipitated, resuspended in 10µl double distilled water and electroporated into competent *E.coli* XL1 blue cells as described above. The transformation mix was plated onto LB plates which contained ampicillin at a concentration of 30µg/ml. True recombinants plasmids were identified by digestion with *Eco*RI and either *Xho*I or *Nde*I which should release the insert. At all times during this procedure great care was taken to avoid contamination with spurious plasmid DNA.

2.8. SCREENING THE COSMID LIBRARY.

The Y chromosome-specific cosmid library LLOYNCO3 "M" was constructed at the Biomedical Sciences Division, Lawrence Livermore National Laboratory, Livermore, California, U. S. A. and was a kind gift to the lab from Pieter de Jong. The source of the Y chromosome was a human-hamster hybrid which contains human chromosomes Y, 22 and 9. Flow sorted Y chromosome DNA was partially digested with *Mbo*I, cloned into the Lawrist 16 vector and propagated in DH5αMCR (BRL) and recombinant clones arrayed into 130 microtitre dishes. The library was determined to contain 82% human clones, 13% hamster clones and 5% non-recombinants and is

calculated to represent 10 x coverage of the human Y chromosome (P. de Jong, personal communication).

The library was plated onto filters (Hybond N+) with a 4 x offsetting by M. Colomer, M. Jobling and C. Tyler-Smith using a 96-well prong device. It was screened by colony hybridisation in rotating tubes using the conditions stated in section 2.4.3. Recombinant clones were easily identified and as yet all clones analysed have been of human Y chromosome derivation (C. Tyler-Smith, personal communication).

2.9. SEQUENCE ANALYSIS.

2.9.1. Sequencing subclones from YAC 17C12C.

Single stranded DNA was prepared by infecting cells carrying pTZ recombinants with the helper phage M13K07 according to the suppliers instructions (Pharmacia) or by denaturing double stranded DNA with 0.2M NaOH and 2mM EDTA, and sequenced using the dideoxy method of Sanger *et al.*, (1977). The following subclones, containing fragments of the original *Eco*RI satellite inserts, were made in order to obtain the sequence from the second strand:- 5bp satellite, ~200bp *Hpa*II- *Hpa*II fragment; 48bp satellite,~260bp *Eco*RI- *Nco*I fragment; 68bp satellite, ~220bp *Eco*RI-*Bgl*II fragment. All sequences were read independently by two people. Sequence comparisons were carried out using the University of Wisconsin GCG package COMPARE and DOTPLOT programs (Devereux *et al.*, 1984).

2.9.2. STS for CEPH library PCR.

The nucleotide sequence of approximately 150bp of plasmid 64a was determined by sequencing single stranded DNA by the method of Sanger *et al.* (1977). From this data two 20mer oligonucleotides were chosen which were balanced in their purine and pyrimidine content for use as PCR primers. Sequence data used to design PCR primers for YαI, YαII and AMGL loci was obtained from the sources listed in Table 2.4.

CHAPTER 3

STRUCTURAL ANALYSIS OF THE PERICENTRIC LONG ARM REGION OF THE HUMAN Y CHROMOSOME

3.1. SEARCH FOR SEQUENCES LOCATED IN THE CENTROMERIC REGION HUMAN Y CHROMOSOME.

A structural analysis of the pericentric region of the human Y chromosome was begun by investigating the nature of sequences in this part of the genome. The approach adopted was to determine whether candidate sequences (sections 3.1.1, 3.1.2 and 3.1.3), isolated from other human and non-human centromeres, were present at or near the centromere of the human Y chromosome. Sections 3.1.1 and 3.1.2 describe the analysis of sequences thought not to be tandem repeats, while section 3.1.3 describes analysis of known human satellite DNAs.

3.1.1. *Schizosaccharomyces pombe* B, K & L repeats.

The first sequences tested were the three different classes of repeats found at the centromeres of the fission yeast *S. pombe* (Clarke, 1990). These repeats were chosen because the centromeric regions of this yeast resemble those of higher eukaryotes in containing several classes of repeated DNA sequences that encompass many kb on each chromosome. If centromeres, like telomeres, are conserved throughout evolution, then probes isolated from the centromeric regions of *S. pombe* will cross-hybridise to mammalian centromeric sequences. Plasmids pKH-K, pKH-B and pKH-L which contain the *S. pombe* K, B, and L centromeric repeats respectively (Fishel *et al.*, 1988) were hybridised to OXEN (49,XYYYY) DNA which had been digested with either *Hin*dIII or *Bam*HI, fractionated by conventional gel electrophoresis and transferred to a filter. As a positive control *S. pombe* DNA, digested with *Hin*dIII, was included on the filter. After hybridisation using standard conditions the filters were washed at low stringency in order to detect any cross hybridisation between the species. The results (not shown) were inconclusive as a faint smear with no distinct bands in human DNA was seen. The experiment was repeated using DNA digested with *Sfi*I and fractionated by PFGE. Despite a two week exposure to film no clear bands were detected so it was concluded that the *S. pombe* centromeric repeats show no

significant homology to the human genome and cannot be used to identify human centromeres.

3.1.2. Human repeated sequences.

3.1.2.a. 14L family.

The 14L family was first described by Glasser *et al.* (1989) as a repeated sequence that was distributed over a small pericentric region of human chromosome 11. The repeat was reported to be unrelated to alphoid DNA and on chromosome 11 was solely located in the centromeric region. It is not known whether p14L is located on other human chromosomes. In order to determine whether the 14L family was present on the human Y chromosome, plasmid p14L was hybridised to a panel of female, male and "Y-only" hybrid DNA (3E7 & 853) which had been digested with *Eco*RI, fractionated by conventional gel electrophoresis and transferred to a filter. After hybridisation the filters were washed at low stringency, (2 x SSC), in order to achieve maximum hybridisation. The results (not shown) indicated that p14L does detect sequences on the human Y chromosome. However, the hybridisation pattern was identical to the sequences detected by the Y specific alphoid probe (Yαl). Therefore it is probable that p14L contains sequences which are related to alphoid DNA and that on the Y chromosome it does not detect any novel sequences.

3.1.2..b. H7 family.

The H7 family was first described by Graham *et al.* (1984) and the probe detects two different low copy number repeated sequences in the human genome. By using *in situ* hybridisation the family was localised to the centromeric regions of chromosomes 1, 13, 14, 15, 21, and 22 (Devine *et al.*, 1985). However, I wanted to ask whether this family of repeats could be found on the Y chromosome by more sensitive filter hybridisation experiments using "Y-only" hybrid cell lines. DNA samples were digested with *Eco*RI, fractionated by conventional gel electrophoresis, transferred to a filter and probed with pH7-1 at low stringency. From the results shown in Figure 3.1 it is clear that this repeat is present on the human Y chromosome as it detects bands in all three Y hybrid cell lines (tracks 5, 7 and 9) and is absent from their revertants and derivatives (tracks 6, 8, and 10). Approximately 50kb of the H7 family is therefore located on the human Y chromosome and must be on the long arm since there is no hybridisation to the Q998-8 derivative (track 10) which

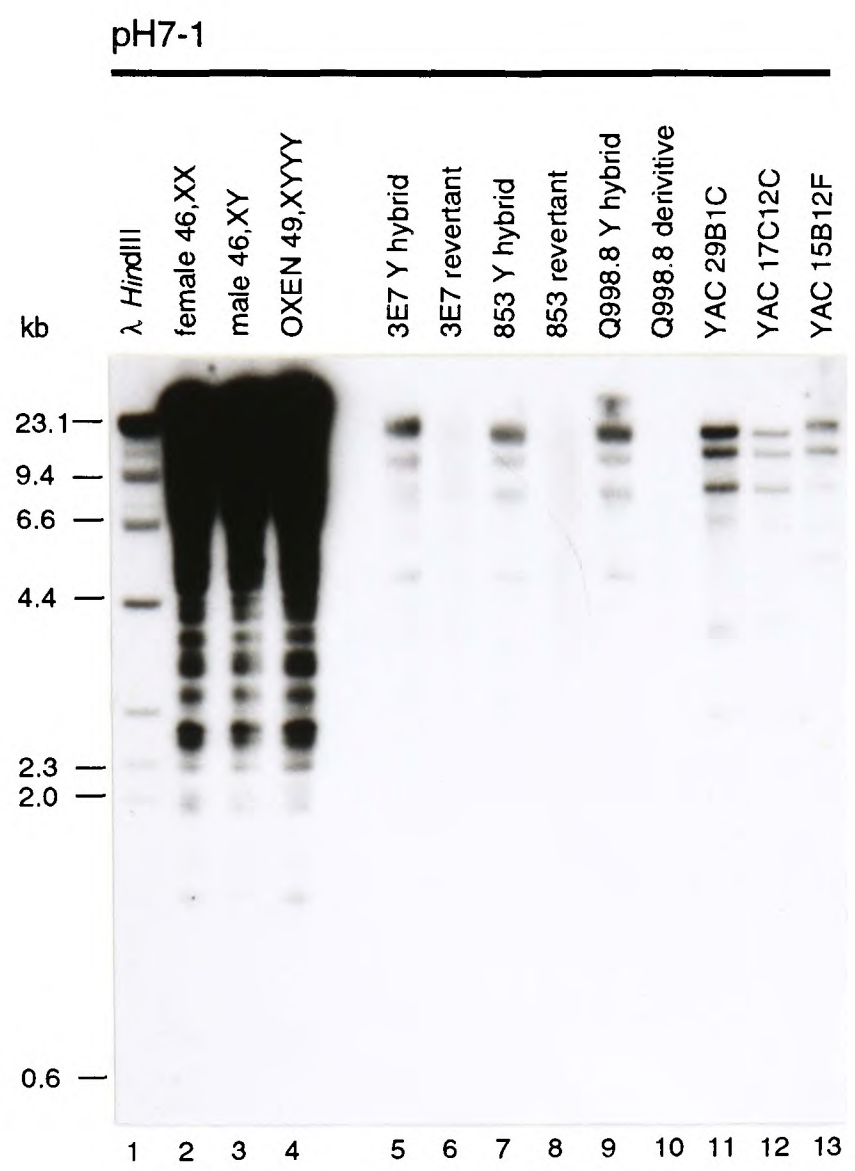


Figure 3.1. Hybridisation of pH7-1 to human, Y hybrid and YAC DNA. The DNA was digested with *Eco*RI, fractionated by gel electrophoresis through a 1% conventional agarose gel, transferred to a membrane and hybridised to pH7-1. The filter was washed in 2 x SSC. The position of λ *Hin*dIII markers are shown on the left hand side.

retains the short arm of the chromosome (C. Tyler-Smith, personal communication).

However, further analysis of the the of sequences detected by probe H7-1 revealed that the sequences it identified were identical to those detected by the 68bp satellite (discussed in the next section). The authors of the paper describing this probe were informed of this result and probe H7-1 was not used in any further investigations.

3.1.2.c. Lambdaphage CM8 subclones.

Six plasmid subclones, (pCM1, pCM6, pCM11, pCM16, pCM21 and pK1620) spanning the putative centromeric clone λCM8 (discussed in section 1.4.4) were a kind gift from Gyula Hadlaczky, Hungary. They were hybridised to *Eco*RI panels as described above to determine whether any of the sequences in λCM8 were present on the human Y chromosome. The results are not shown but none of the sequences were present in the 3 "Y-only" hybrids indicating that these sequences are not present on the human Y chromosome.

3.1.3 Detection of the 68bp and 48bp satellites (tandem repeats) on the Y chromosome.

3.1.3.a. 68bp satellite.

The presence of the known human satellite DNA families on the Y chromosome was investigated. Alphoid DNA and the 5bp satellite have previously been detected in the centromeric or pericentric region of the Y chromosome by *in situ* hybridisation (Table 1.1) but the 42bp, 48bp and 68bp satellites were not found. It was therefore asked whether these sequences could be detected on the Y chromosome by more sensitive filter hybridisation experiments using "Y-only" hybrid cell lines. DNA samples were digested with *Hae*III, fractionated by conventional gel electrophoresis, and a filter transfer was probed with the 68bp satellite probe, pKFC68, at low stringency. The method of construction of pKFC68 was expected to link 68bp satellite monomers from several different chromosomes, and extensive hybridisation to many regions of the genome would be expected under the conditions used. Figure 3.2a tracks 2-4 shows that there is indeed strong hybridisation to female, male and 49,XYYYY human samples. A band at 16.9kb can be seen in the male sample (track 3) which is absent from female DNA (track 2). The intensity of this band is increased in the 49,XYYYY DNA (track 4) and it is present in the three independent Y hybrids (tracks 5, 7 and 9) but absent from

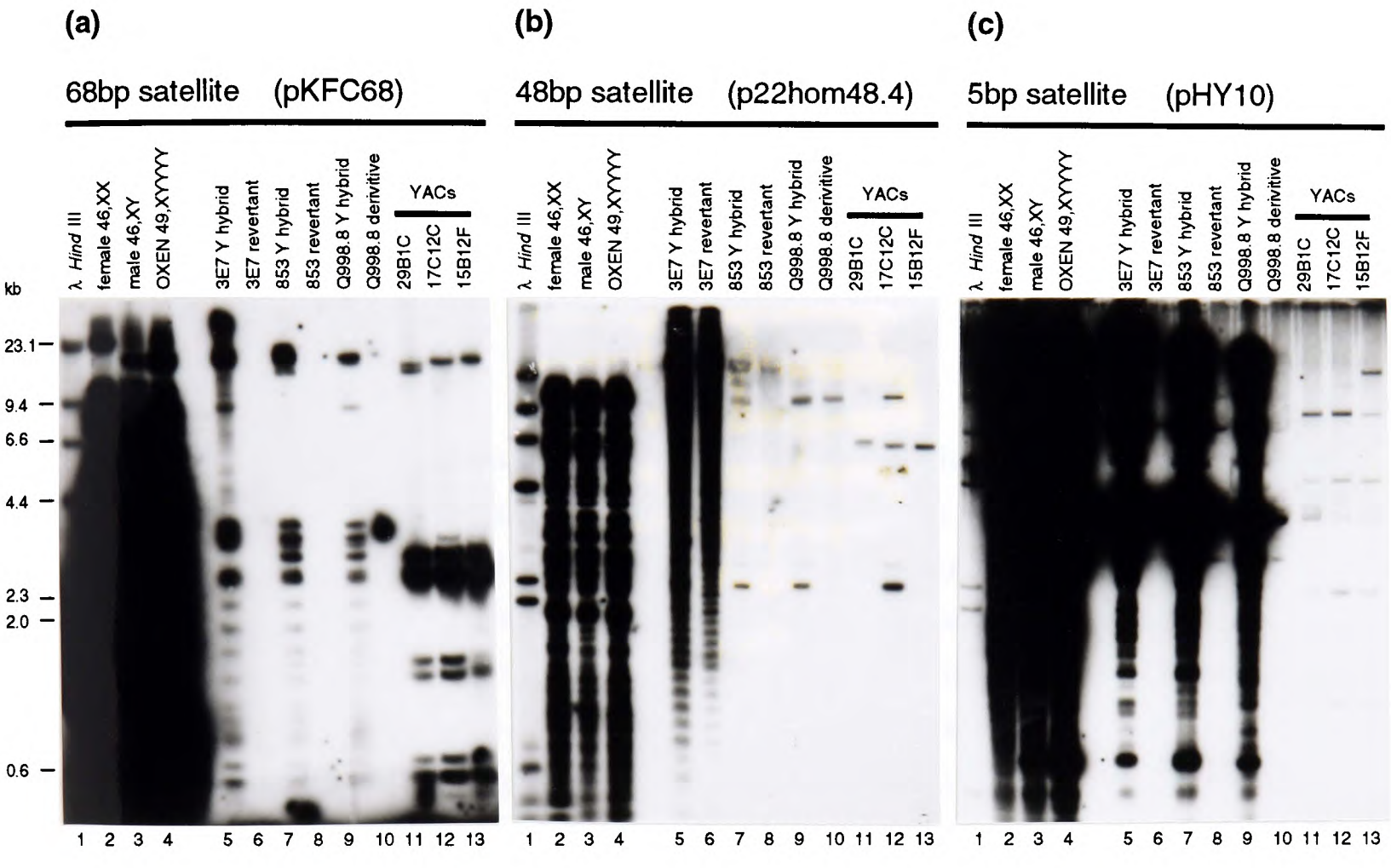


Figure 3.2. Hybridisation of the 68bp, 48bp and 5bp satellites to human, Y hybrid and YAC DNA. The DNA was digested with *Hae*III, (a), or *Eco*RI, (b) and (c), fractionated by gel electrophoresis through a 0.75% (a) or 1% (b and c) conventional agarose gel, transferred to a membrane and hybridised to pKFC68 (a), p22hom48.4 (b) or pHY10 (c). All filters were washed in 2xSSC. The position of the λ *Hin*dIII markers are shown on the left hand side.

their revertants or derivatives (tracks 6, 8 and 10). This fragment is therefore located on the Y chromosome, and like pH7-1 must be on the long arm since it is not present in the Q998-8 derivative.

The hybrids also contain additional fragments hybridising to pKFC68. The sum of the sizes of these *Hae*III fragments is ~54kb and this therefore represents an approximate estimate for the amount of 68bp satellite on the Y chromosome. Some minor differences in the patterns can be seen in the different hybrids. These could represent non-Y sequences, rearrangements which have occurred during culture, or pre-existing polymorphisms on the Y chromosome.

3.1.3.b. 48bp satellite.

Figure 3.2b shows the results of hybridisation of the 48bp satellite probe p22hom48.4 to *Eco*RI digests of the samples used in Figure 3.2a. The 48bp satellite detects an abundant repeated sequence in human DNA (tracks 2, 3 and 4) and two bands of 19.0kb and 2.0kb in the hamster hybrids (tracks 7 and 9). These bands are absent from the 853 revertant and the 2.0kb band is absent from the Q998-8 derivative. More extensive cross-hybridisation is seen to a repeated sequence in the mouse genome (tracks 5 and 6) but, on the original autoradiograph, the 19.0kb and 2.0kb fragments can also be seen in track 5. These bands therefore represent a small amount of 48bp satellite located on the Y chromosome. The 2.0kb band is located on the long arm and one copy of the 19.0kb band is located on the short arm: see section 3.3. below.

3.1.3.c. 42bp satellite.

The 42bp satellite probe, KFC42, was also used to examine the *Eco*RI panel. Even under low stringency washing conditions (2 x SSC), the repeat was not detected on the human Y chromosome (results not shown).

These results show that the 5bp, 48bp, 68bp and 170bp satellites are present on the Y chromosome, but no sequence that hybridises to the 42bp satellite is detected.

3.2. LOCALISATION OF THE SATELLITES ON THE HUMAN Y CHROMOSOME.

The 68bp satellite was localised on the human Y chromosome using the panel of 28 rearranged Y chromosomes discussed in section 1.5.2. The resulting deletion map is summarised in Figure 3.3. The major centromeric block of

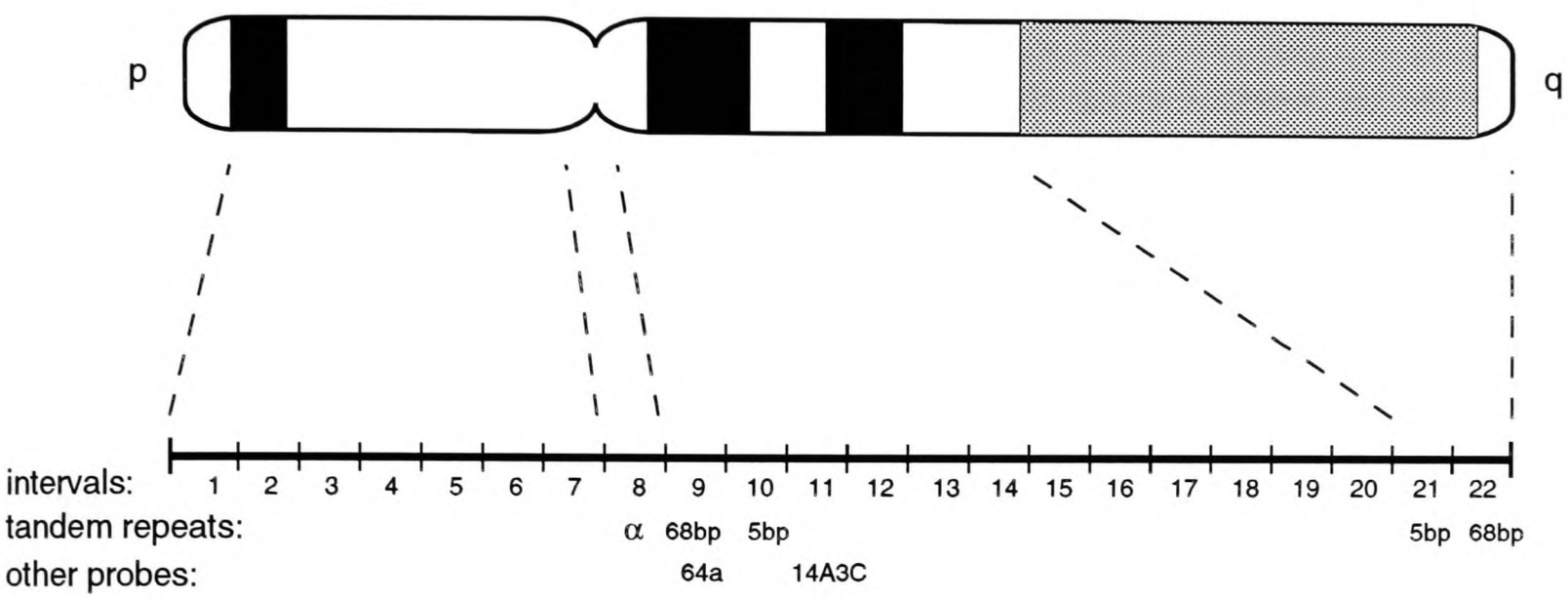


Figure 3.3. Deletion map of the human Y chromosome showing the locations of tandemly repeated DNA sequences and additional probes used in this chapter.

alphoid DNA is located in interval 8. Most of the 5bp satellite in the long arm heterochromatin lies in interval 21, but minor 5bp satellite sequences are found in interval 10. The diagnostic 16.9kb Y chromosome specific 68bp satellite band is placed in interval 9, adjacent to alphoid DNA on the long arm of the chromosome. The 48bp satellite could not be localised using the deletion panel because the Y chromosome bands were obscured by intense hybridisation from other regions of the genome. A random probe, 64a (Guellaen *et al.*, 1984) also fell into interval 9, and another random probe, 14A3C (Fisher, 1992) was located in interval 11. Localisation of all the sequences on the deletion map was done by C. Tyler-Smith and R. Fisher.

This analysis shows that some 68bp satellite sequences are located in the pericentric region of the long arm of the Y chromosome, and that probes such as 64a are nearby.

3.3. ANALYSIS OF YAC CLONES FROM THE REGION.

The existing probes were not sufficient to construct a restriction site map of the region, and provided little detail about the sequence organisation. Therefore it was decided to isolate YAC and cosmid clones. These clones could then be used both as a source of additional probes for mapping, and for detailed analysis of the region. In addition, the results would show whether analysing YACs was a suitable and accurate way of understanding the structure of genomic DNA from the centromeric region of the human Y chromosome.

3.3.1. Screening the Biochemistry, Oxford YAC library.

3.3.1.a. Isolation of YAC clones.

The Biochemistry, Oxford YAC library (described in section 2.6.1) was screened with pKFC68 by colony hybridisation at low stringency and 22 positive clones were isolated. To determine whether any of the clones originated from the Y chromosome they were digested with *Hae*III, fractionated by conventional gel electrophoresis and a filter transfer probed with pKFC68. Figure 3.4 shows the results from 11 of the clones. The YAC in track 5 contains the 16.9kb diagnostic male specific band which is also seen in the three "Y-only" hybrid DNA. YACs 3, 4, 7-11 are not derived from the Y chromosome as they do not have the 16.9kb diagnostic band. YACs 2 and 6 which do not contain the 68bp satellite were false positives. YAC 1 is derived from the Y chromosome but not from the centromeric region and is described in the next section. In total,

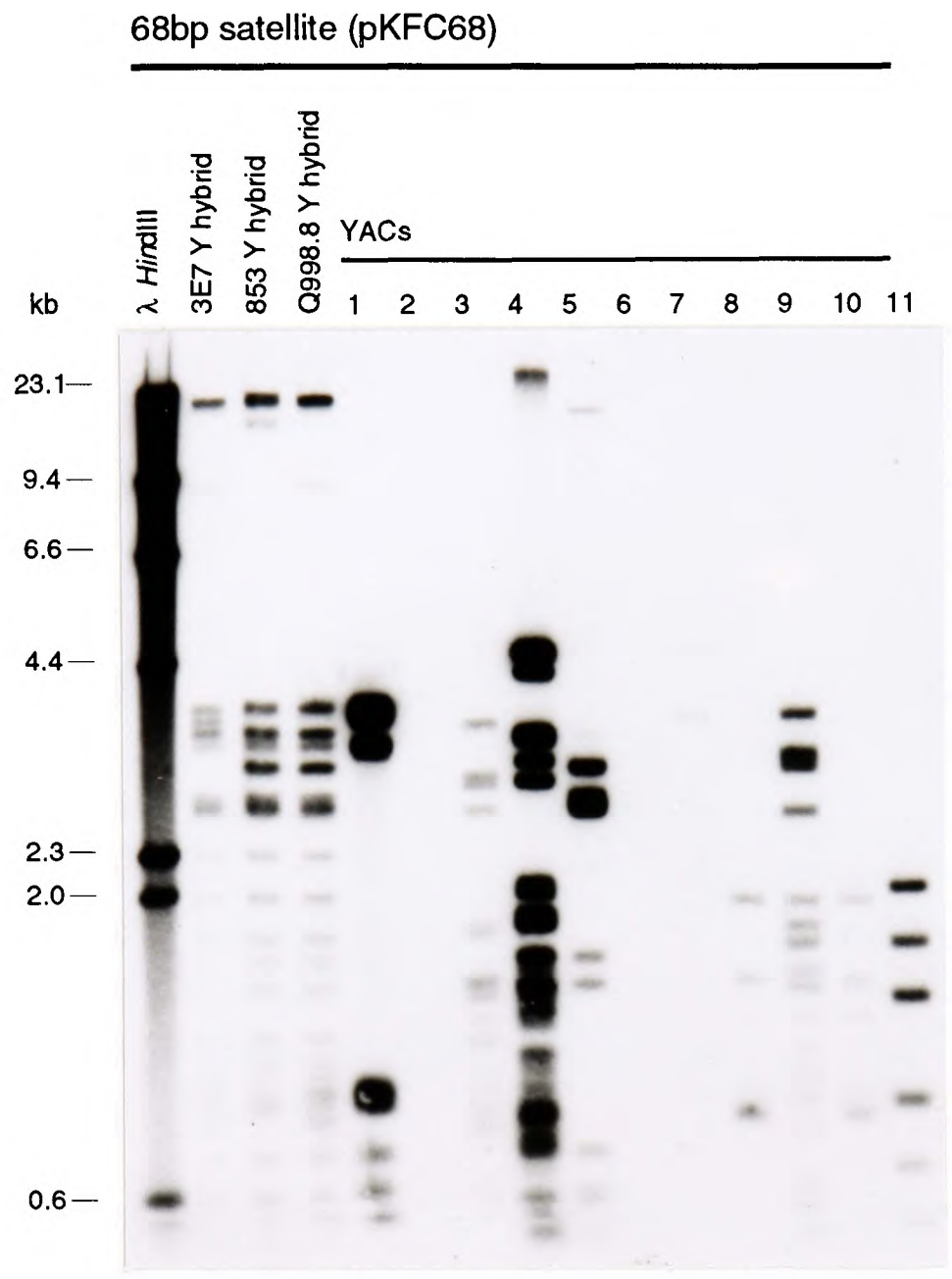


Figure 3.4. Identification of 68bp satellite containing YAC clones which are derived from the human Y chromosome. Y hybrid and YAC DNA was digested with *Hae*III, fractionated through a 0.75% agaose gel, transferred to a filter and hybridised with the 68bp satellite probe, pKFC68. The filter was washed in 2 x SSC and the position of the λ *Hin*dIII size markers is shown.

three clones, 29B1C (110kb), 17C12C (233kb) and 15B12F (255kb), contained the diagnostic 16.9kb Y chromosome fragment and a subset of the other bands seen in the Y hybrids (Figure 3.2a, tracks 11, 12 and 13). These clones must originate from the pericentric region of the long arm of the Y chromosome. In one of these clones (29B1C, track 11) the 16.9kb band appears as a doublet and is slightly reduced in size, showing that there is some instability of this sequence in the YAC.

3.3.1. Isolation of YAC 9C3C (telomeric Yq).

A fourth YAC (9C3C - 115kb) was identified which contained a further subset of 68bp satellite Y specific bands but not the diagnostic 16.9kb band (Figure 3.4, YAC 1). It was therefore suspected that this YAC was derived from another part of the Y chromosome. To confirm this, DNA from YAC 9C3C, YAC 17C12C, Y hybrids 3E7 and 853 was digested with a *Eco*RI, *Eco*RV, *Ava*II and *Pvu*II and analysed by conventional gel electrophoresis, filter transfer and hybridisation with the 68bp satellite. The results obtained from the *Eco*RI panel are shown in Figure 3.5a. It can be seen that each band detected by this satellite in YAC 9C3C (track 4) is identical in size to a band detected in both the Y hybrids (tracks 2 and 3). Furthermore, YAC 17C12C detects a different set of Y bands to 9C3C. The hybridisation patterns from the other panel were similar i.e. YAC 9C3C always detected the same sized bands as those found in the hybrid cell lines.

However, these results still did not confirm beyond doubt that YAC 9C3C was derived from the Y chromosome as it could still be possible that a "non-Y" YAC could give the results obtained above by chance. To confirm the chromosomal origin of YAC 9C3C a plasmid library was constructed from the YAC and a resulting subclone (pY9.9) was used to probe a small panel of genomic DNA. The results are given in Figure 3.5.b and it can be seen that pY9.9 detects a 1.3kb band in male, OXEN, the three Y hybrids and YAC 9C3C DNA (tracks 3, 4, 5, 6, and 7). The band is difficult to see in hybrids 3E7 and 853 though it can be seen on the original autoradiograph. The band is not seen in female DNA (track 2). These results confirm that YAC 9C3C is indeed derived from the human Y chromosome. Finally, clone pY9.9 was shown to map into the most distal interval on the long arm of the Y chromosome (interval 22, Figure 3.3). Due to its telomeric location, this YAC is not relevant to the centromere and will not be discussed any further.

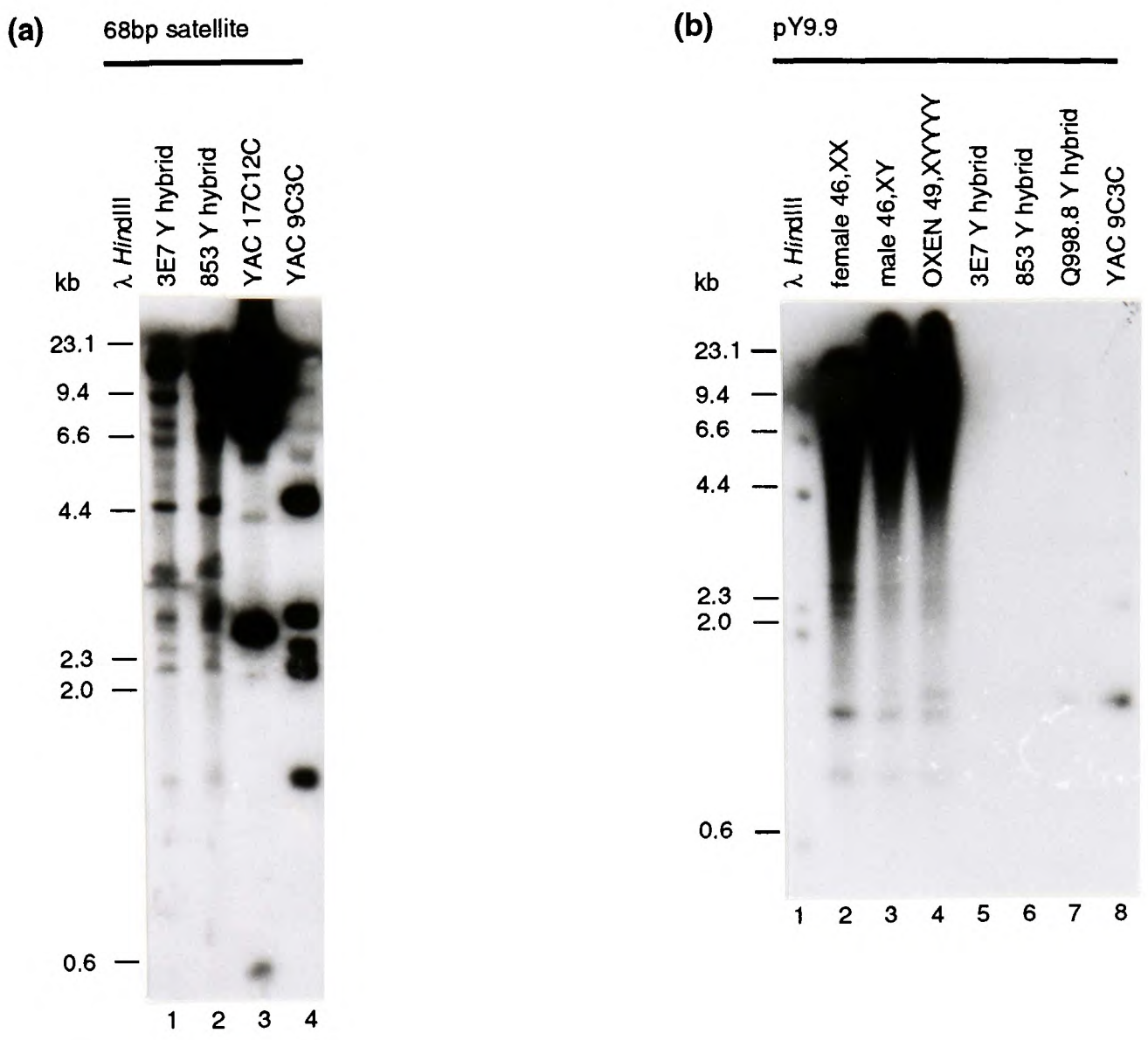


Figure 3.5. Analysis of YAC 9C3C. (a) Hybridisation of the 68bp satellite to Y hybrid and YAC DNA. (b) Hybridisation of pY9.9 to human, Y hybrid and YAC DNA. For (a) and (b) the DNA was digested with *Eco*RI, fractionated through a 1% conventional agarose gel and a filter transfer hybridised to (a) pKFC68 and (b) pY9.9. Probe pY9.9 was hybridised using competition. The filters were washed in 2 x SSC (a) or 0.1 x SSC (b). The position of λ *Hin*dIII markers are given on the left hand side.

3.3.2. Detection of 48bp and 5bp satellites in pericentric Yq.

3.3.2.a. *48bp satellite.*

Hybridisation of the three YAC clones to the 48bp satellite probe (Figure 3.2b, tracks 11, 12 and 13) showed that YAC 17C12C contained both of the Y chromosome *Eco*RI fragments and that these must be from the pericentric Yq region. The band common to all three YAC tracks represents cross-hybridisation to vector sequences. The 19.0kb band must be a doublet in genomic DNA: one copy, present in the Q998-8 derivative cell line (track 10), is located on the short arm and the other copy, present in YAC 17C12C, is located on the long arm. This conclusion is supported by phosphorimager quantitation: in the hamster hybrids the intensity of the 19.0kb band is similar to the intensity of the 2.0kb band, whereas in the YAC the 19.0kb band has only about half of the intensity of the 2.0kb band. Further analysis of other YACs show that the short arm copy is located close to the alphoid array (section 5.1.3.c and Figure 5.17).

3.3.2.b. 5bp satellite

The 5bp satellite is extremely abundant in the human genome (Figure 3.2c, tracks 2, 3 and 4). On the Y chromosome it is mainly located in the long arm heterochromatin. Small amounts of this satellite are also found in all three YACs (tracks 11, 12 and 13) and therefore represent an additional array of this tandem repeat in the pericentric region of Yq.

This further analysis shows that some of the 5bp and 48bp satellites are located in pericentric Yq near the 68bp satellite.

3.3.3. Restriction site maps of the YACs.

The three YACs were mapped by partial digestion with *Bss*HII, *Mlu*I, *Sfi*I and *Sst*II and indirect end labelling (Burke *et al.*, 1987). These rare-cutting enzymes were chosen as they had previously been shown to be informative for analysing the centromeric region of the human Y chromosome (C. Tyler-Smith, personal communication). Partial digests of the YACs were fractionated by PFGE, transferred to a filter and probed with the right and left arm YAC vector probes. The results for YACs 29B1C, 17C12C and 15B12F are given in Figures 3.6a, 3.7a and 3.8a respectively. The 68bp, 48bp and 5bp satellites were placed on the maps by hybridisation to complete digests (Figure 3.9). The resulting individual maps are given in Figures 3.6b, (29B1C), 3.7b (17C12C) and 3.8b (15B12F) respectively. For each map the results from the

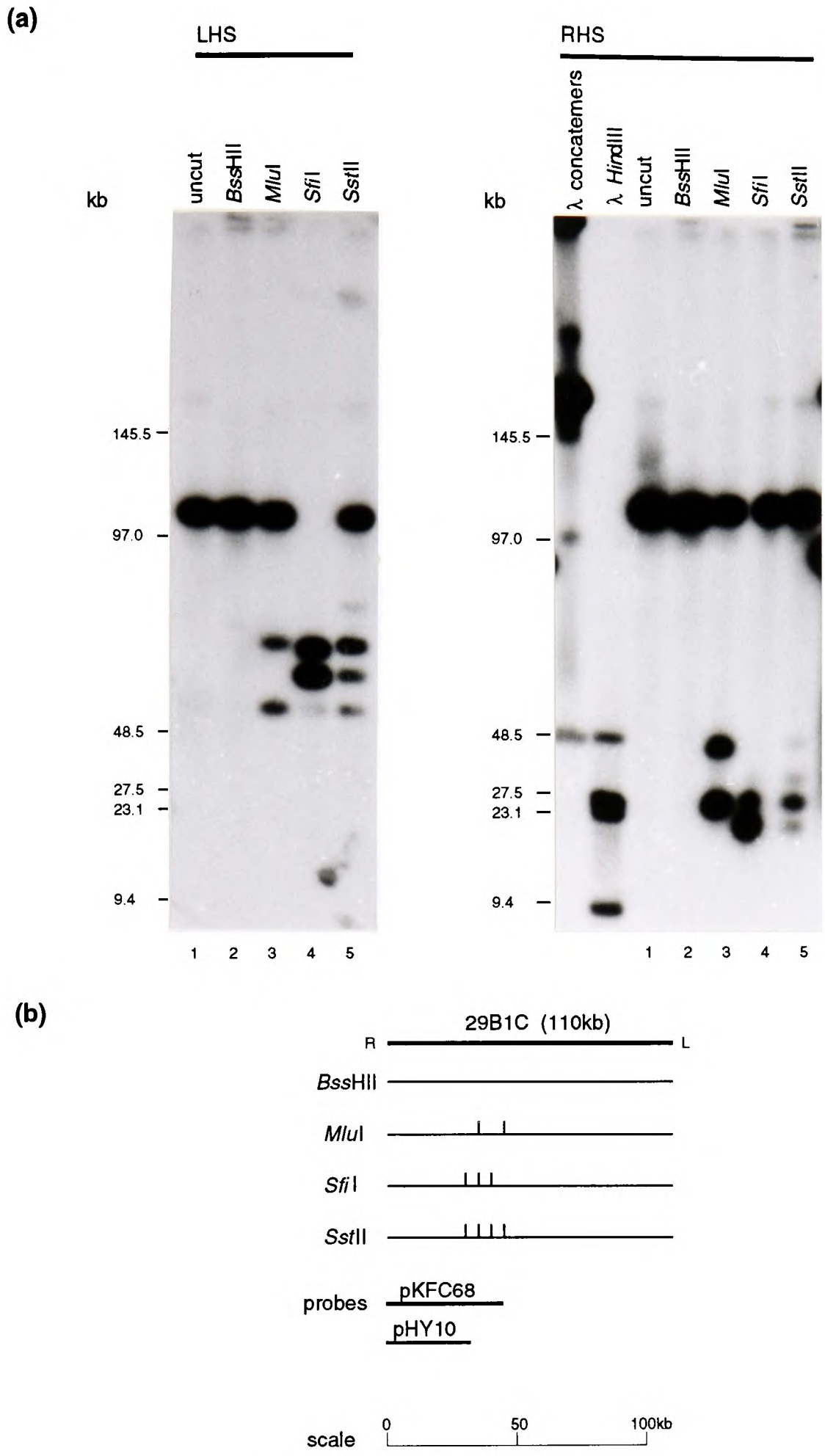


Figure 3.6. Analysis of YAC 29B1C. (a) Hybridisation of the left and right hand vector probes to partial digests of the YACs. The DNA was partially digested with the enzymes indicated, fractionated by PFGE, (effective pulse time 3 secs, run time 50hrs) and a filter transfer hybridised with the right and left hand vector probes. The filters were washed in 0.1 x SSC. The position of the λ *Hin*dIII and λ concatemers are shown at the side of each photograph. (b) Restriction site map of the YAC showing the location of the 68bp (pKFC68), and 5bp (pHY10) satellites.

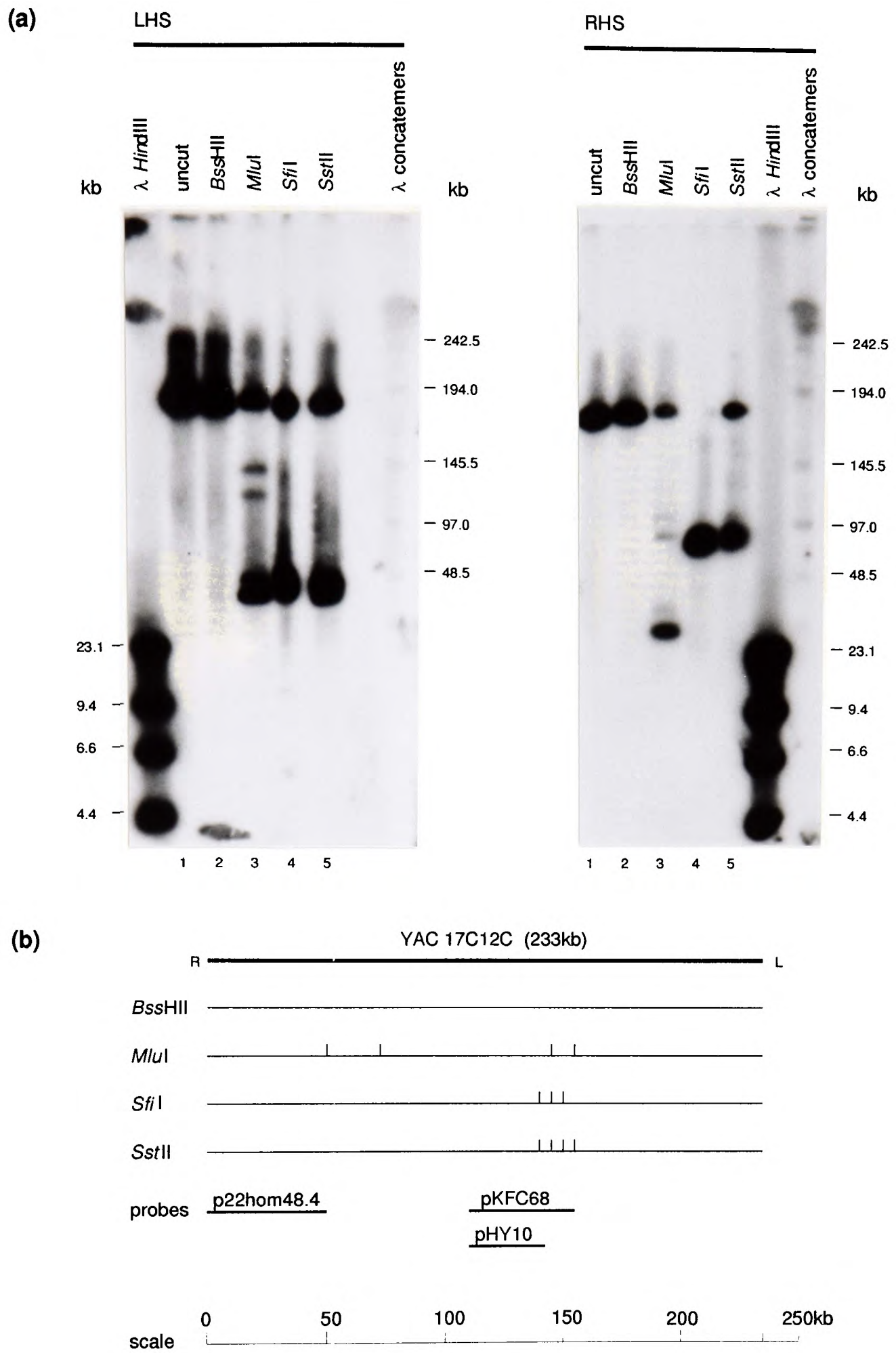


Figure 3.7. Analysis of YAC 17C12C. (a) Hybridisation of the left and right vector probes to partial digests of the YAC. The DNA was partially digested with the enzymes indicated, fractionated by PFGE, (effective pulse time 11 secs, run time 26 hrs) and a filter transfer probed with the right and left hand vector probes. The filters were washed in 0.1 xSSC. The position of the λ *Hin*dIII and λ concatemers are shown at the side of each photograph. (b) Restriction site map of the YAC showing the location of the 68bp (pKFC68), 48bp (p22hom.48.4) and 5bp (pHY10) satellites.

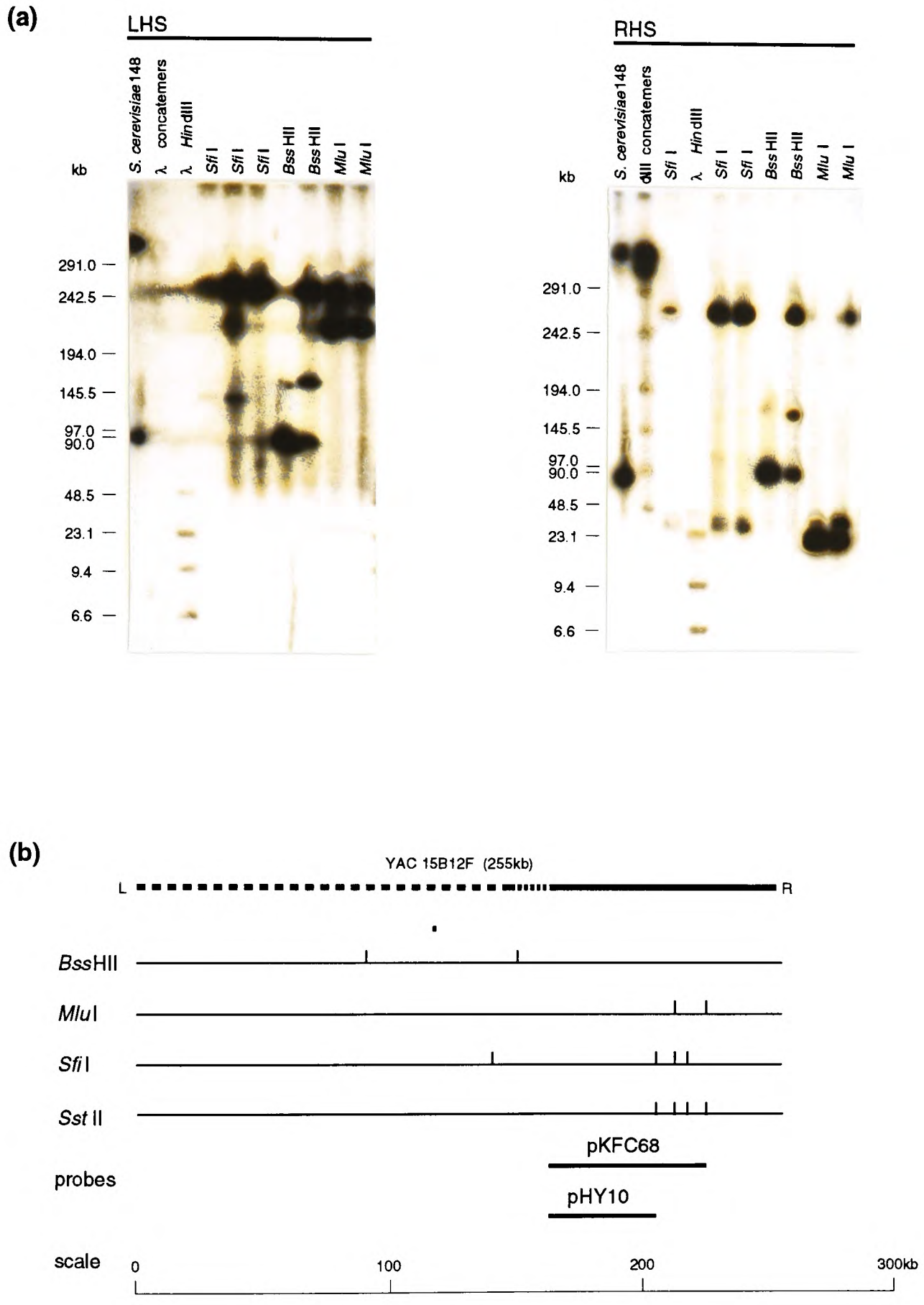


Figure 3.8. Analysis of YAC 15B12F. (a) Hybridisation to the left and right hand vector probes to partial digests of the YAC. The DNA was digested with the enzymes indicated, fractionated by PFGE and a filter transfer hybridised with the left and right hand vector probes. The pulsed-field gel was run for 26hrs with an effective pulse time of 11 secs and the *Sst*II partial were run on a separate gel (not shown). (b). Restriction site map of the YAC showing the location of the 68bp (pKFC68) and 5bp (pHY10) satellites. The stippled line represents the portion of the YAC which is not derived from the Y chromosome.

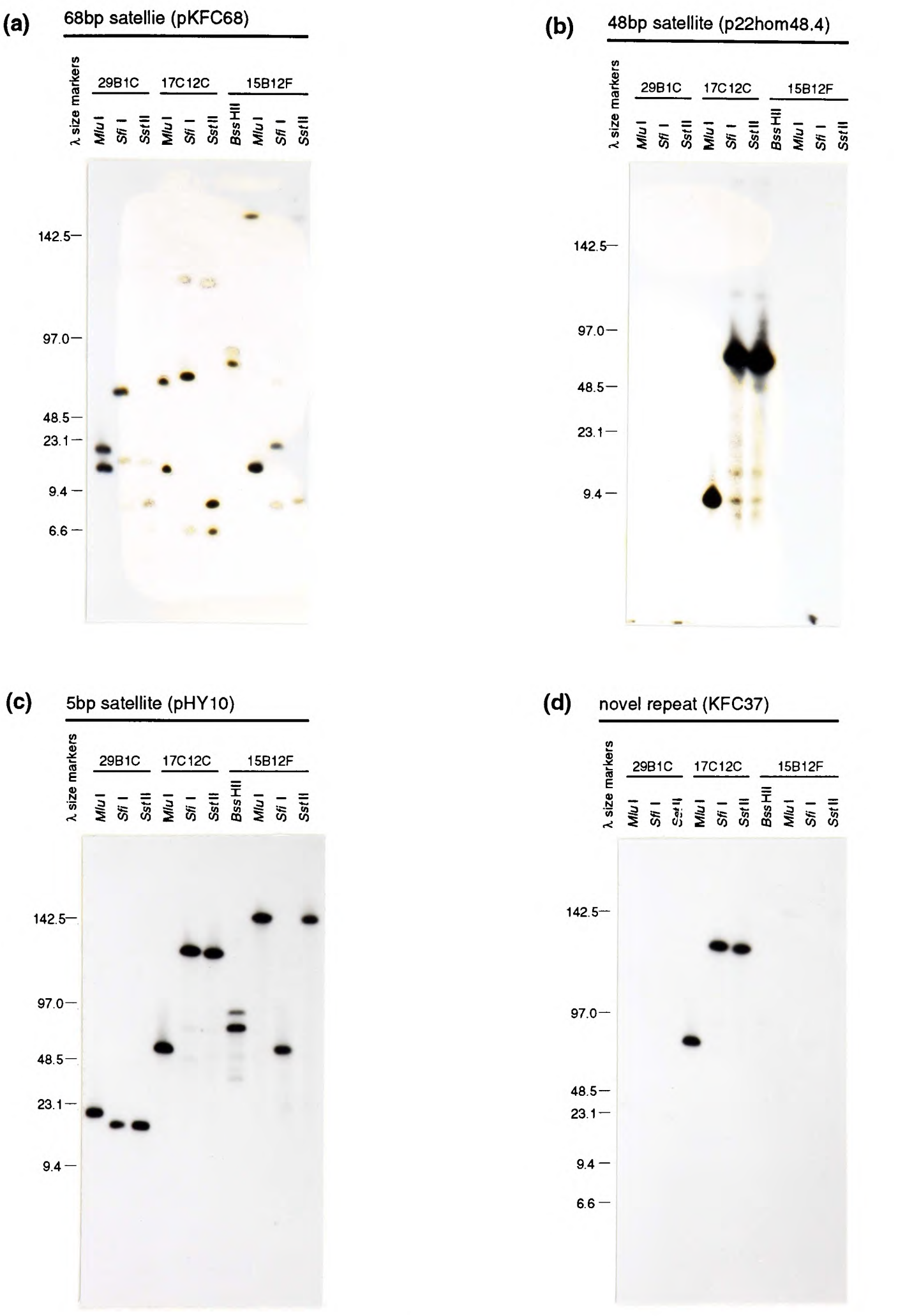


Figure 3.9 Localisation of the 5bp, 48bp and 68bp satellites on YACs derived from the long arm pericentric region of the human Y chromosome. The YAC DNA was digested to completion with the enzymes indicated, fractionated by PFGE, transferred to a membrane and hybridised with probes (a) - (d). The filters were washed in 2 x SSC and the position of the λ concatemers and λ *Hin*dIII size markers are shown at the side of each photograph. The pulsed-field gels were run for 39hrs with an effective pulse time of 5 secs.

partial digests were used to determine the order of the restriction enzyme sites and the results from the complete digests to accurately determine the sizes of the fragments.

As expected, the three maps show a region of overlap and the resulting YAC contig is given in Figure 3.10. However the maps of 15B12F and 17C12C show inconsistencies indicating that one of these YACs was likely to be co-ligated. To confirm this, the left hand end of 15B12F was recovered by plasmid rescue and the resulting probe (p15X1) was used to probe the same DNA panel as used in Figure 3.2. The results, given in Figure 3.11, show that p15X1 detects a 3.0kb band in female, male, OXEN and YAC 15B12F DNA (tracks 2, 3, 4 and 13). This band is not present in the Y hybrids, the revertants and derivative or the other YACs (tracks 5, 6, 7, 8, 9, 10, 11 and 12). Therefore p15X1 is a non-Y single copy sequence. In contrast, further analysis of 17C12C showed that it was derived from the Y chromosome (see below). YAC 15B12F is therefore co-cloned; the junction between the Y and non-Y inserts is located between the pHY10-hybridising sequences derived from the Y chromosome and a *Bss*HII site present in the non-Y DNA (Figure 3.10a).

Approximately 230kb of genomic DNA is covered by the YACs and this was divided into 10 genomic regions, defined by the restriction enzyme sites and the edges of the YAC clones (Figure 3.10c). The satellites were scored as hybridising or non-hybridising within each region and the distribution is shown in Figure 3.10d. Sizes of the arrays can be estimated very approximately using frequent-cutter enzymes; taking these into account, the most probable locations are shown in Figure 3.10e. The 68bp satellite lies close to the 5bp satellite, and the 48bp satellite array is located approximately 100kb away.

3.3.4. Analysis of plasmid subclones derived from YAC 17C12C.

In order to analyse the structure of this region in more detail, plasmid subclones were constructed from YAC 17C12C. In addition, the plasmid subclones were used to look for probes which would allow mapping in genomic DNA. Filters containing 768 plasmid subclones (384 colonies per filter) were made as described in materials and methods and screened by colony hybridisation with the 68bp, 48bp and 5bp satellites. The results from one filter are shown in Figure 3.12. In total a single subclone was detected with the 68bp satellite probe (a), seven subclones for the 48bp satellite (b) and many subclones for the 5bp satellite (c). Clones containing Y copies of each

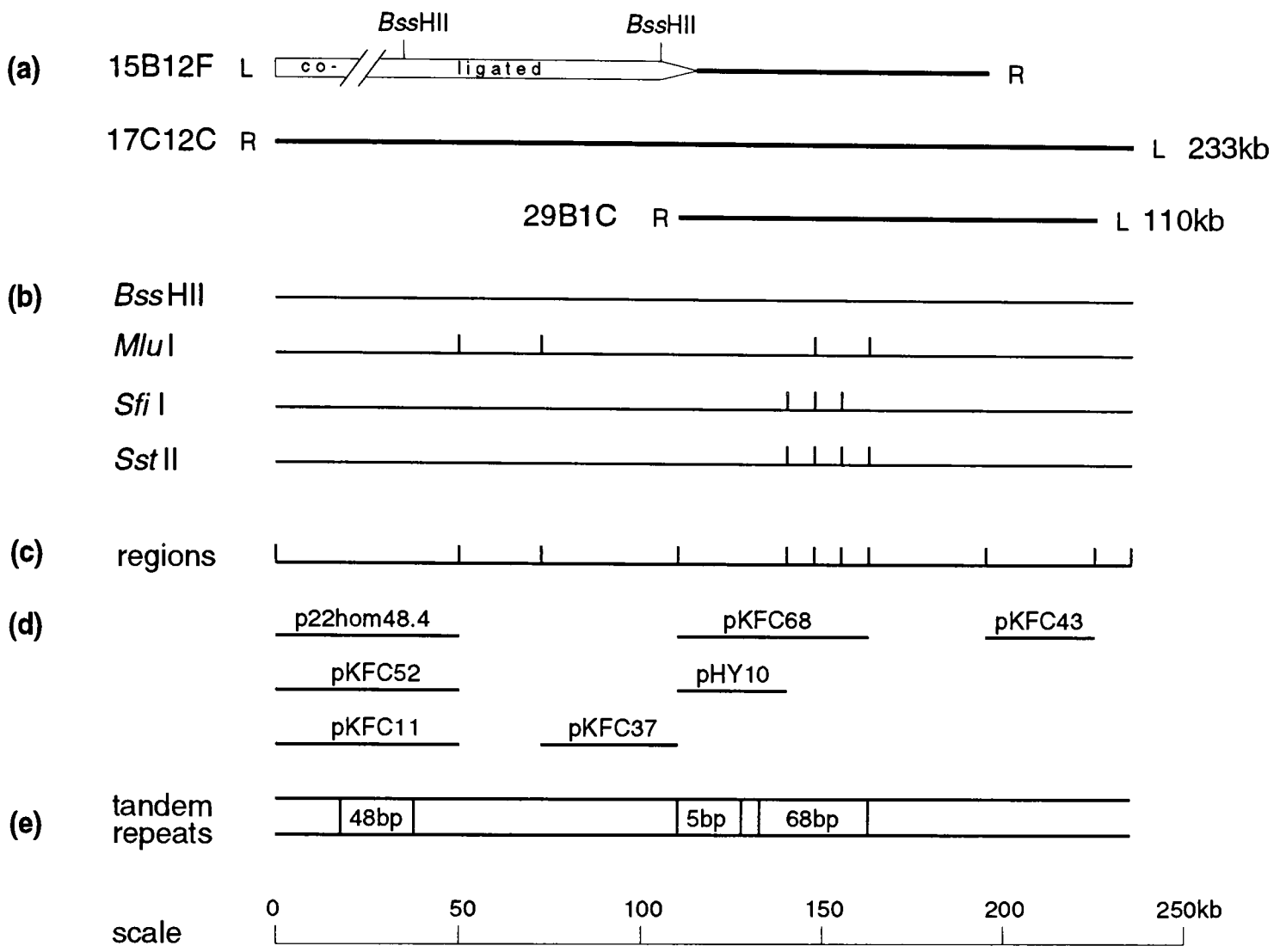


Figure 3.10. Restriction site map of the YACs showing the location of the repeated sequences. (a) YACs; (b) restriction site maps; (c) regions defined by the edges of the YACs and restriction enzyme sites; (d) location of probes; (e) position of the tandem repeats. L and R represent the left (centromeric) and right (non-centromeric) arms of the YAC vector.

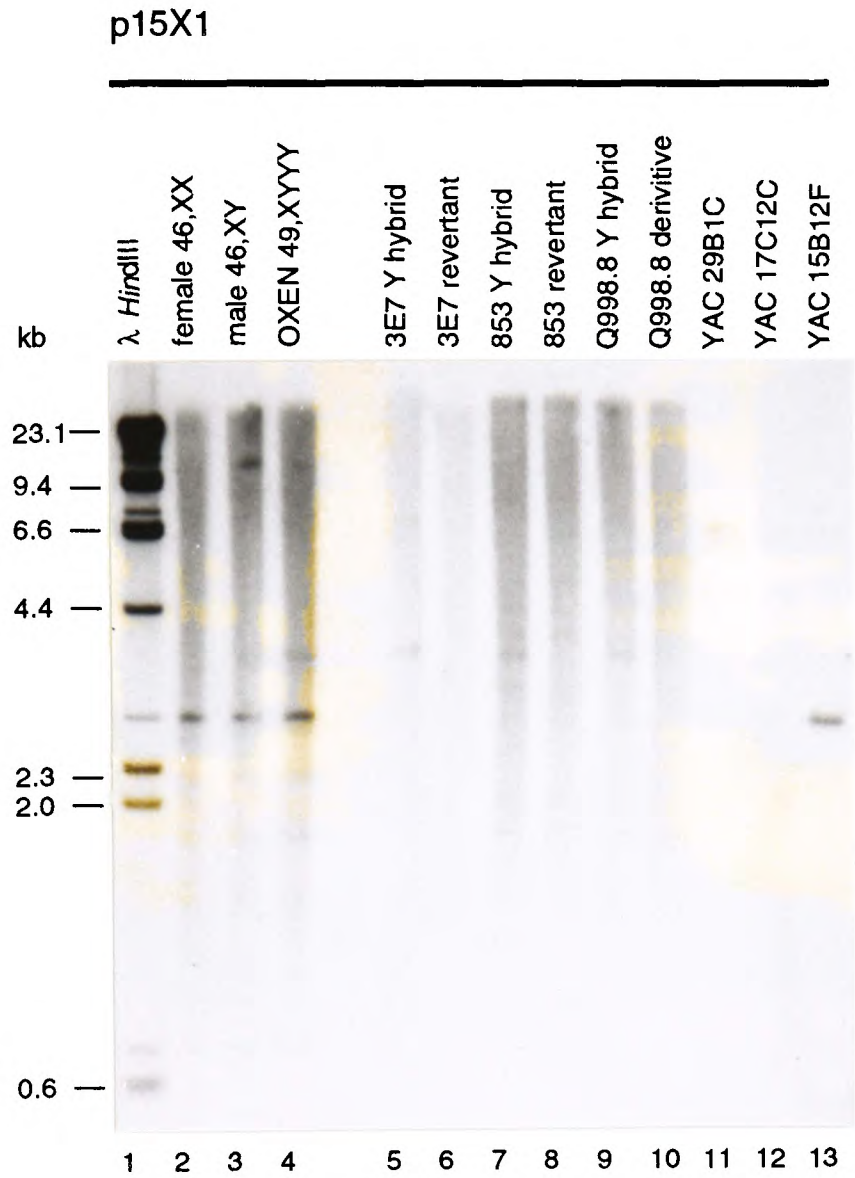


Figure 3.11. Hybridisation of p15X1 to genomic, Y hybrid and YAC DNA.The DNA was digested with *Eco*RI, fractionated by gel electrophoresis through a 1% conventional agarose gel, transferred to a membrane and hybridised to p15X1. The filter was washed 0.1 x SSC. The position of λ *Hin*dIII markers are shown on the left hand side.

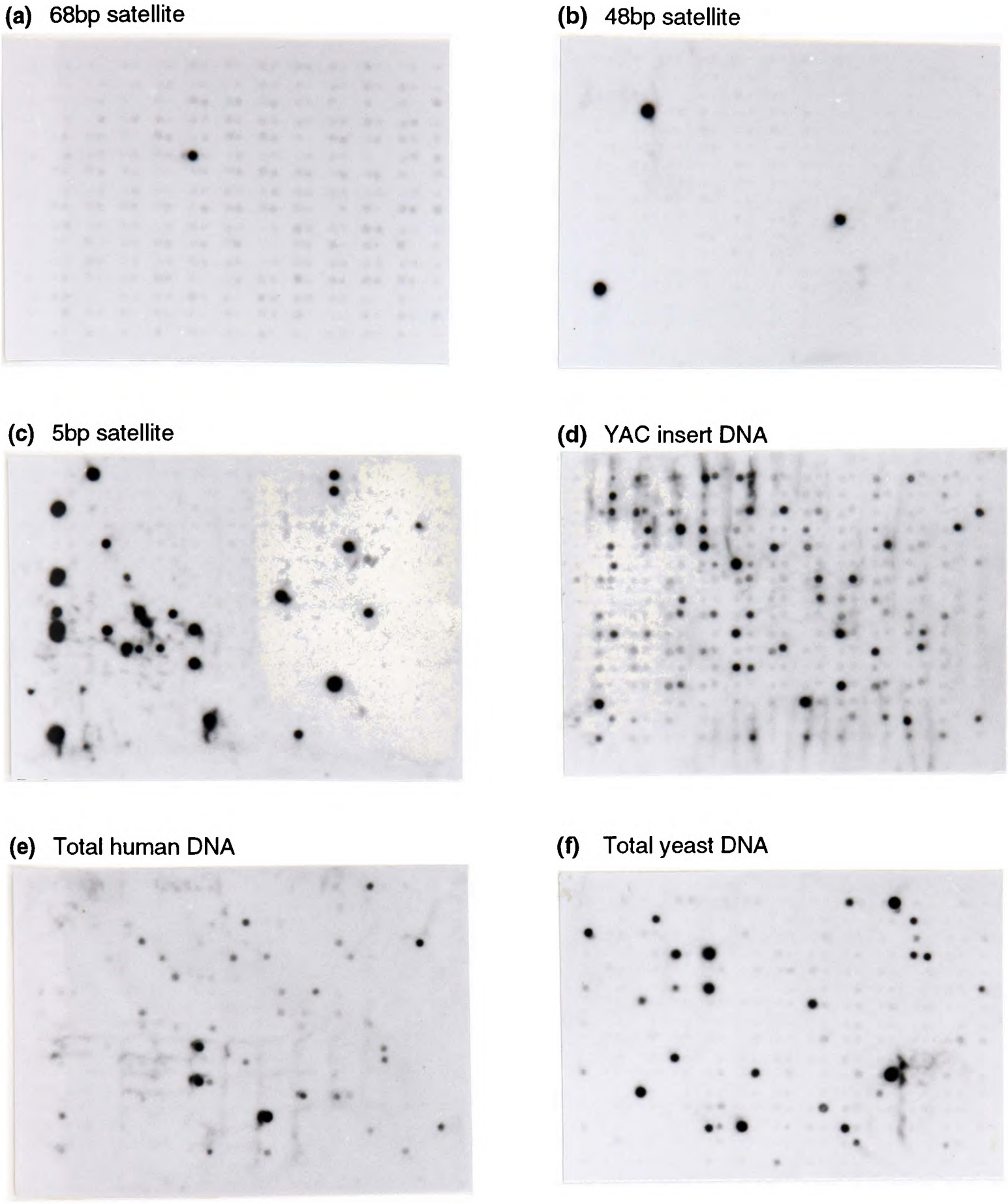


Figure 3.12. Screening of the plasmid library made from YAC 17C12C. The colonies were plated onto filters using a 96-well prong device and hybridised with (a) 68bp satellite (pKFC68), (b) 48bp satellite (p22hom.48.4), (c) 5bp satellite (pHY10), (d) YAC 17C12C insert DNA, (e) total human DNA and (f) total yeast DNA. Filters (a) - (d) were washed in 2 x SSC and filters (e) and (f) in 0.5 x SSC.

satellite were analysed in more detail (see section 3.3.5). The subclones were also screened with YAC 17C12C, total human and total yeast DNA and the results from one filter are shown in Figure 3.11 (d), (e) and (f) respectively. The filters were screened with total human DNA to detect clones which contain *Alu* (Deininger *et al.*, 1981) and *Kpn* (Singer and Skowronski 1985) repeated sequences. The library was also screened with total yeast DNA to detect clones which contain yeast DNA as the YAC insert used for subcloning was difficult to separate from a yeast chromosome. Clones were analysed which did not hybridise to total human and total yeast DNA probes but were detected by the YAC insert probe. These clones were expected to contain human sequences which were enriched in the YAC. Using this strategy clones of several novel repeated sequences were detected but no unique sequences were found in this region. The novel repeats (pKFC11, pKFC37, pKFC43 & pKFC52) were placed on the YAC maps (Figures 3.9.d and 3.10d) and used to probe *Eco*RI panels. Examples of the results are shown in Figure 3.13. In each case the novel repeat detects a single band in YAC 17C12C which is the same size as the plasmid insert. This band can also be seen in the somatic cell hybrids, 49,XYYYY (OXEN), male and female DNA. An additional band is detected by pKFC11 in two of the hybrids and many additional bands are always detected in human DNA (tracks 2, 3 and 4). These clones therefore represent Y copies of moderately repeated sequence families. KFC11 detects a polymorphism in human DNA (compare track 2 with tracks 3 and 4). Neither restriction enzyme analysis of human DNA nor partial sequencing of these clones provided evidence for a tandemly repeated organisation (results not shown).

3.3.5 Sequence analysis.

Plasmid subclones derived from YAC 17C12C and representing Y copies of the 68bp satellite (pY68), the 48bp satellite (pY48) and the 5bp satellite (pY5) were sequenced according to the strategies shown in Figure 3.14. The clones were sequenced on both strands but pY68 was unreadable on one strand. In addition, one anomaly was encountered in the sequencing of the 68bp satellite: since a *Bgl*II site was present at position 220-225, nucleotide 220 is expected to be an A. However, in three independent subclones this position was unambiguously a C. The sequence data is also presented in Figure 3.14. Comparison with the published sequence of each satellite confirms that the subclones are indeed Y variants of the known repeats. Further comparison of the Y sequence to the consensus sequences was performed for the 68bp and

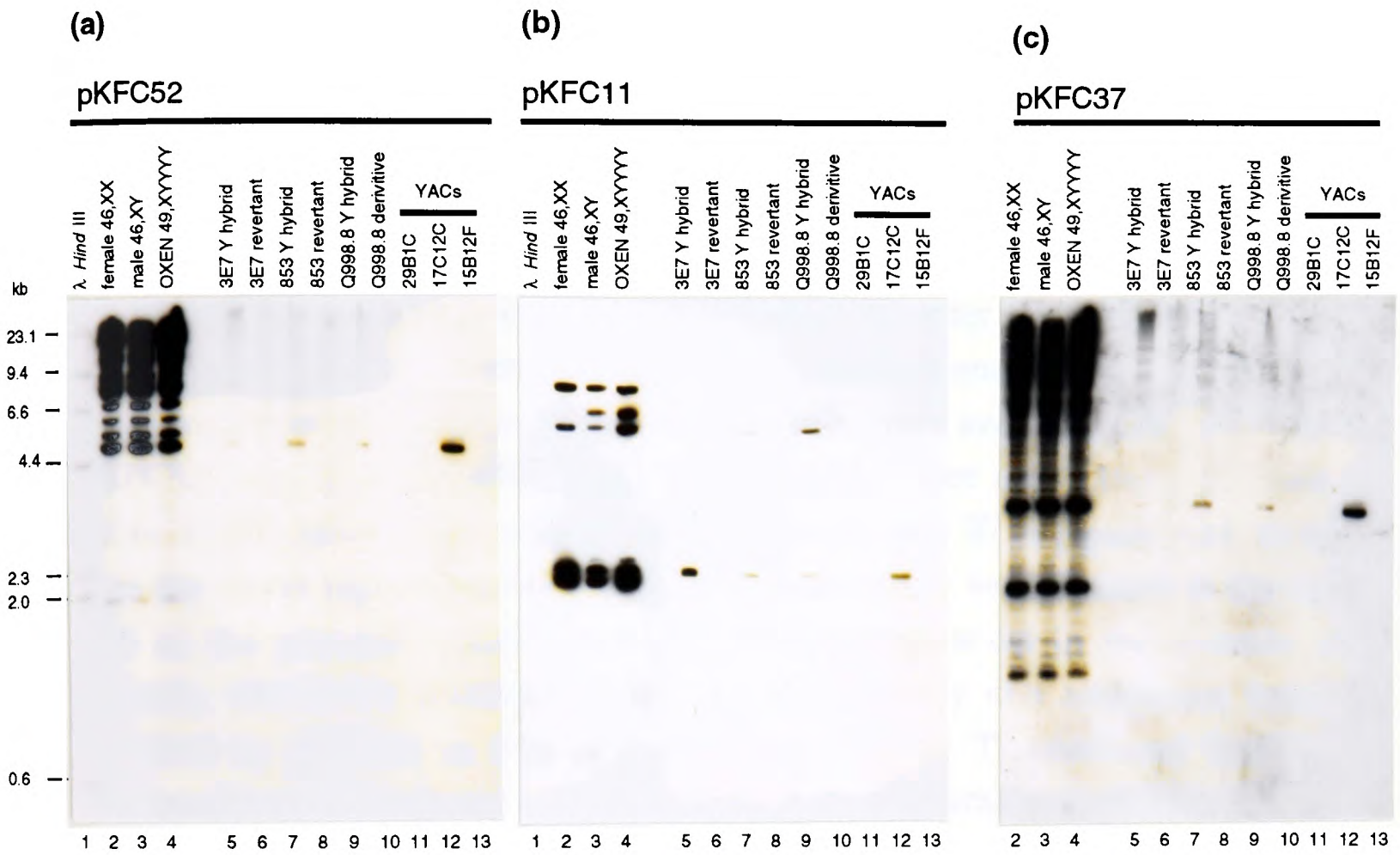


Figure 3.13. Hybridisation of the novel repeats to human, Y hybrid and YAC DNA. The DNA was digested with *Eco*RI and fractionated by gel electrophoresis through a 1% conventional agarose gel, transferred to a membrane and hybridised to pKFC52 (a), pKFC11 (b) or pKFC37 (c). All filters were washed in 0.1 x SSC. The position of the λ *Hin*dIII size markers is shown on the left hand side.

48bp satellites using DOTPLOT analysis and the results are shown in Figure 3.14a and 3.14b. The Y copy of the 48bp repeat is highly homologous to the chromosome 22 consensus sequence. The Y copy of the 68bp repeat, however, is more diverged from the genomic consensus sequence.

3.4. ANALYSIS OF COS 64.

The probe 64a detects one Y locus and one autosomal locus in human DNA (Guellaen *et al.*, 1984). The Y locus falls into the same deletion interval as the pKFC68 16.9kb fragment (Figure 3.3) and can be linked to probes from YAC 17C12C by PFGE (section 3.5, below). Three independently derived YAC libraries were screened with probe 64a (see Chapter 4) but no YACs were isolated which were derived from the Y chromosome. Therefore cos64, the original source of subclone 64a, was included in this analysis of the pericentric Yq region. Sites for *Bss*HII, *Mlu*I, *Sfi*I and *Sst*II were mapped in addition to sites for the frequently cutting enzymes *Eco*RI and *Xba*I. A map of the rare-cutter sites is shown in Figure 3.15a. None of the satellite sequences were detected in cos64. A fragment was subcloned from the distal end of cos 64 and designated 64b. In genomic DNA it detects a family of moderately repeated sequences which included one Y copy (Figure 3.15b). Thus the organisation of sequences detected by 64b resembles that of the moderately repeated sequences in YAC 17C12C.

3.5. LONG RANGE MAPPING OF GENOMIC DNA.

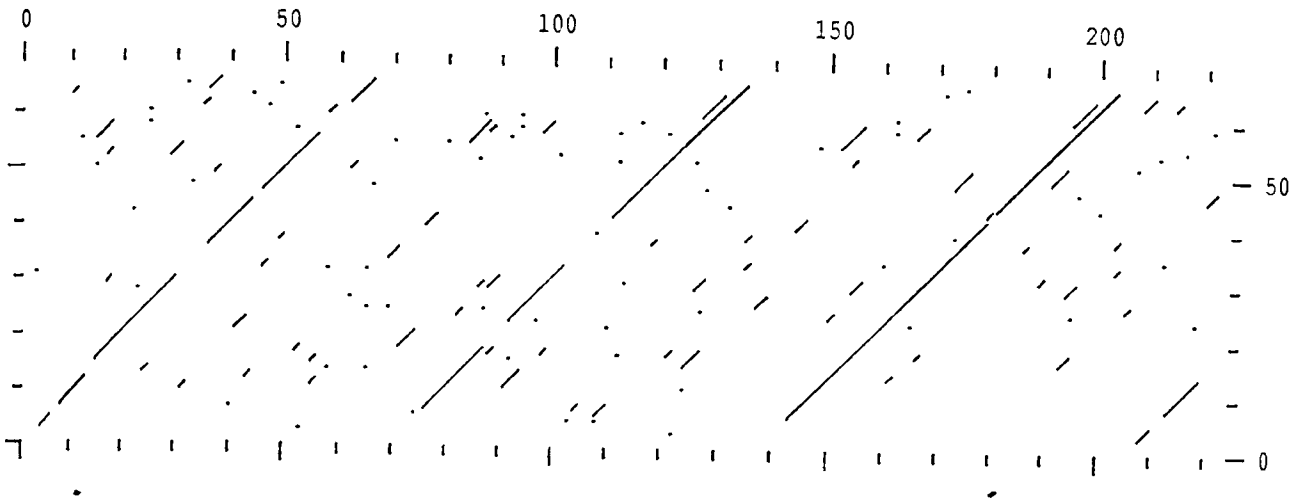
To complete this structural analysis of the pericentric Yq region it was necessary to construct a long range restriction site map. However, the repeated nature of all the available probes made this impossible in human male DNA. It was therefore investigated whether it would be possible to map in the Y hybrids using probes which are repeated in human genomic DNA but have only one copy on the Y. Initial experiments showed that methylation-sensitive rare-cutter restriction enzymes cleaved the two hamster hybrids very poorly in this part of the genome. However, sites in the mouse hybrid 3E7 were cleaved to a low but detectable extent. Figure 3.16 shows *Mlu*I, *Sfi*I and *Sst*II digests of 3E7 DNA probed with 64a and probes from each end of YAC 17C12C, pY48 and pKFC43. Hybridisation to the *Sfi*I fragments is very much stronger than to the *Mlu*I or *Sst*II fragments because the *Sfi*I sites are cleaved to completion while the *Mlu*I and *Sst*II sites are only partially cleaved. All three probes detect the same ~500kb and ~1300kb *Sst*II fragments, showing that they are located within ~500kb of each other and

(a) 68bp satellite




```
GAATTCCCTG  TAGC AGAGC  CTAGACAAGT  GTAATATCAC  CTAGGTTATC  AGTGCAGGTA  TAAGTCATA     1 - 69
CAGCCTCCTG  TAGGCAGAGA  GTAGACAAGA  GTTCCCTCCC  CAGGGTGATC  AGTGCAGAGA  TGTGTCAC     70 - 137
AAAACCCCTG  TAGGCAGAGC  TTAGACAAGA  GTTTCATCAC  TTGGTTGATC  AGTTCAGAGA  TGTGTCAC    138 - 205
AATGTCCATG  TAGGC GATC                                                           206 - 224

AATGCCCCTG  TAGGCAGAGC  CTAGACAAGA  GTTACATCAC  CTGGGTGATC  AGTGCAGAGA  TATGTCAC   genomic consensus
```

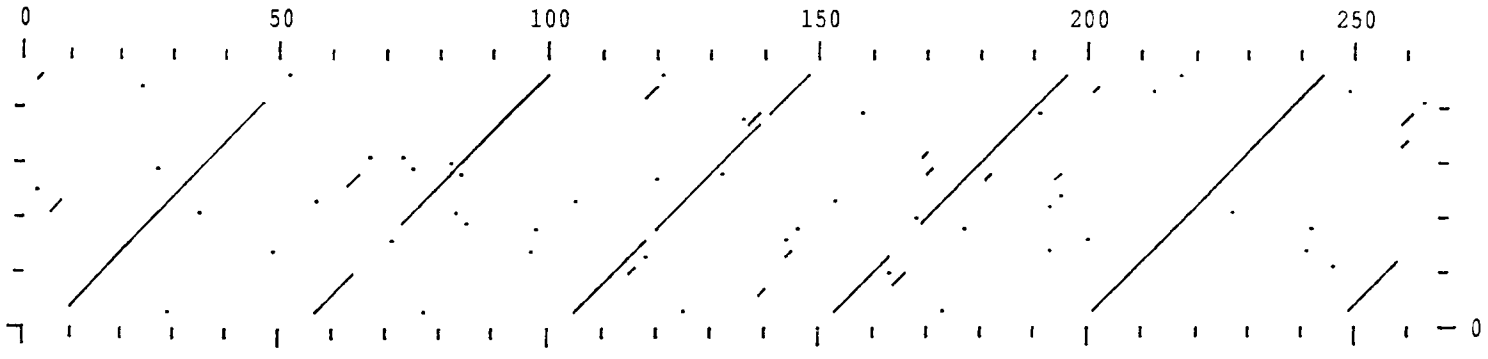


(b) 48bp satellite

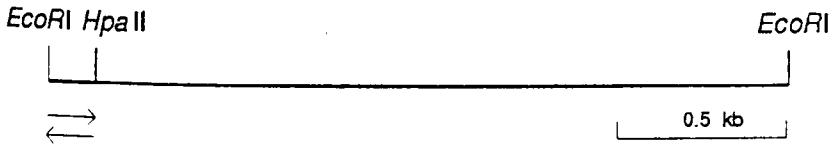



```
GAATTCTGGA  ACACTGCTAC  AAGGGTCTGA  ATGTTTGTCC  TTCACATA      1 - 48
CCATTCCAGA  ACACTGTTGC  CGTGGTCTGA  ATGTGTGTCC  CTCACATA     49 - 96
GGATTCCAGA  ACACTGCTAC  TAGGTTCTGA  ATGTTTTTCC  CACACCTA     97 - 144
GGATTCCAGA  ACACTTCTAC  TGGTGTCTGA  ATGGTTGTTC  CTCACATA    145 - 192
TGATTCCAGG  ACACTGCTAC  GAGAGTCTTA  ATGTTTGTCC  TTCACATA    193 - 240
GGATTCTAGA  ACACTGCTCC  CATGG                             241 - 265

GGATTCCAGA  ACACTGCTAC  GAGGGTCTGA  ATGTTTGTCC  CTCACATA    chromosome 22 consensus
```



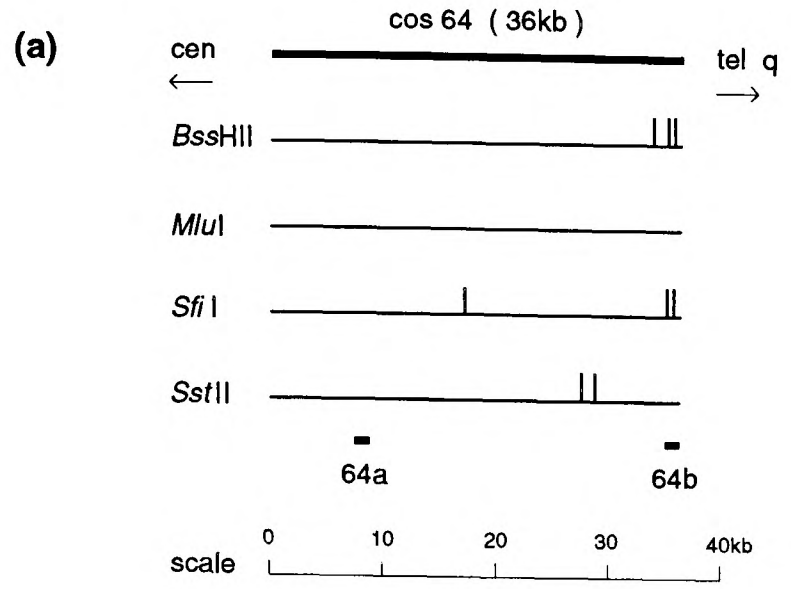
(c) 5bp satellite




```
 GAAT  TCTAT  GGAAT  GAAAT  AAACC  CGAGA  GGAAT  GGAAT  TGAAC  GGAAT  GGAAG  GGAAT  GGAAT     1 - 64
GGAAT  GGAAT  CAACA  CGATT  TGAAT  GGAAT  GGAAT  GGAAA  GGAAA  AGAAT  GGAAT  GGAAT  GGAAT    65 - 129
GTATT  GAACC  CGAGT  GGAAT  TGAAT  GGAAT  GGAAA  GGAAT  GGAAT  GGAAT  GGAAT  CAACC  GG      130 - 191

GGAAT  genomic consensus
```

Figure 3.14. The sequence and DOTPLOT of the pericentric Yq copies of the 68bp (a), 48bp (b), and 5bp (c) satellites. At the top of each section, the subclone containing the repeat is drawn and the arrows underneath represent the extent of the DNA sequenced. The published genomic consensus sequences are given below (Agresti *et al.*, 1989; Metzdorf *et al.*, 1988 and Prosser *et al.*, 1986 respectively). The DOTPLOTS drawn in sections (a) and (b) show the comparisons of the Y copies of the 68bp and 48bp repeats (horizontal axis) to their genomic consensus sequences (vertical axis). In each case the window size is 5 and the stringency 4.



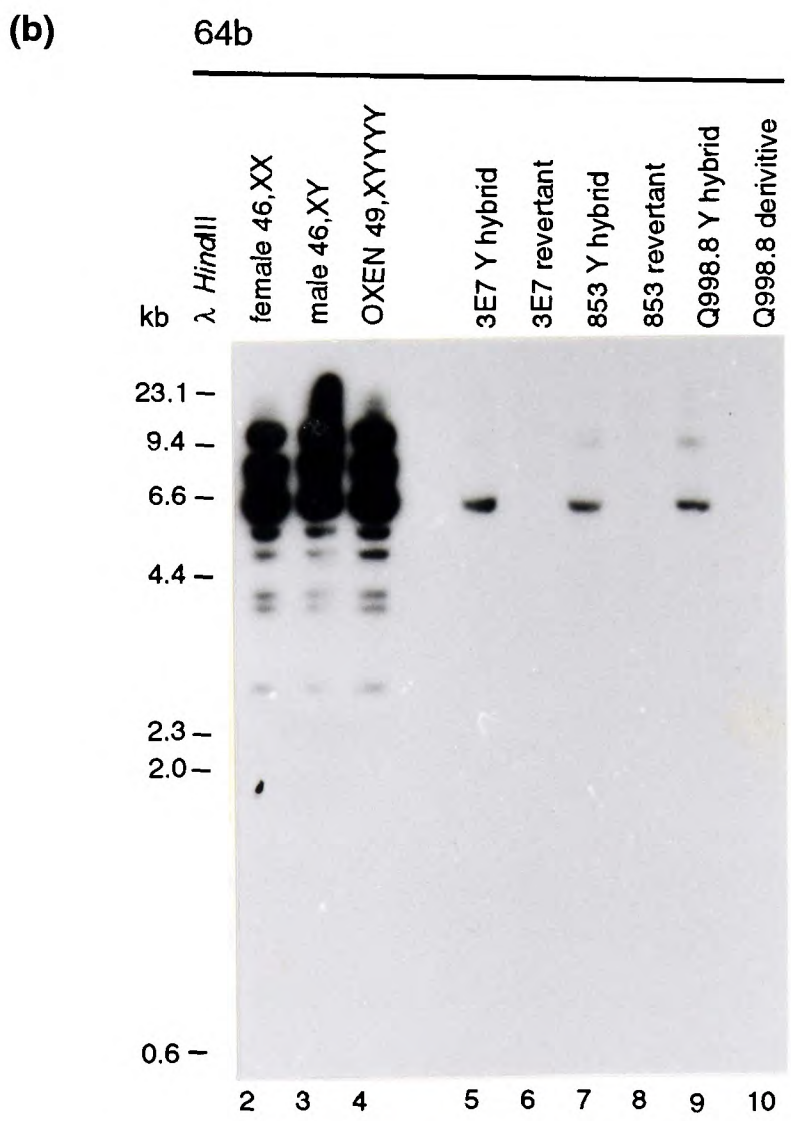


Figure 3.15. Analysis of cos 64. (a) Restriction site map of cos 64. The location of 64a and 64b are shown. (b) Hybridisation of 64b to human and Y hybrid DNA. The DNA was digested with *Eco*RI, fractionated by gel electrophoresis through a 1% conventional agarose gel, transferred to a membrane and hybridised to 64b. The filter was washed in 0.1 x SSC. The position of λ *Hind*III markers are shown on the left hand side.

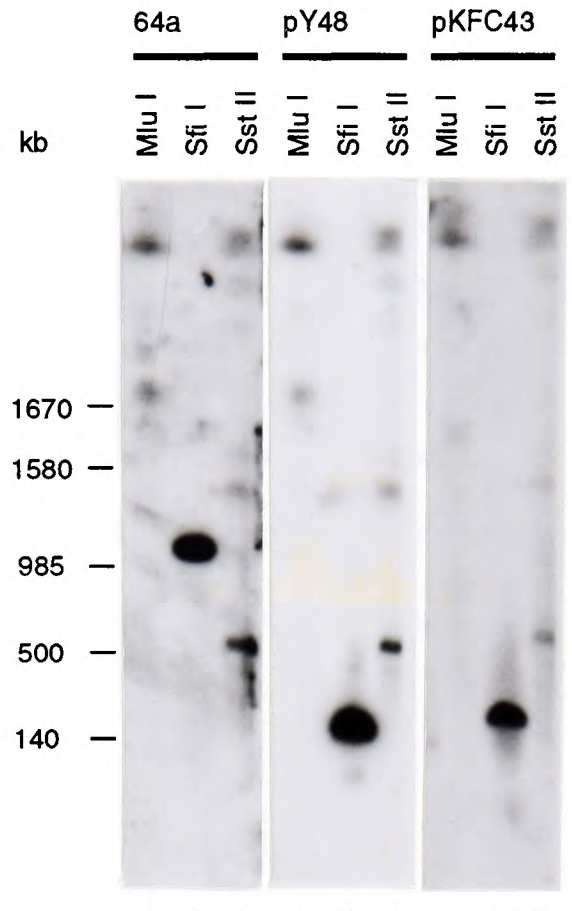


Figure 3.16. Linkage of 64a, pY48 and pKFC43 in genomic DNA. DNA from the mouse hybrid cell line 3E7 was digested with *Mlu*I, *Sfi*I or *Sst*II as shown and analysed by pulsed-field gel electrophoresis using a 1% agarose gel at 60V with a 4 minute pulse time for 50 hours. A filter transfer was hybridised sequentially with the probe indicated at the top of each track. The sizes of the hybridising fragments (in kb) are shown on the left-hand side.

Table 3.1. Fragment sizes in 3E7 genomic DNA.

Enzyme	64a	64b	pY48	pKFC43
*Bss*HII	90	1940	1940	1940
*Mlu*I	1670	1670	1670	1580
*Sfi*I	985	*	140	140
*Sst*II	500	500	500	500
Sfi I+ *Sst*II	170	*	140	110
*Mlu*I + *Sst*II	400			

All sizes are in kb. When multiple fragments were seen in a track, only the size of the smallest fragment is given.
* no bands were seen in this track because it had run off the gel.

linking cos 64 to YAC 17C12C. 64a and pY48 detect a common *Mlu*I fragment of ~1670kb while pKFC43 detects a ~1580kb *Mlu*I fragment. This shows that 64a is closer to pY48 than to pKFC43 and that at least one of the *Mlu*I sites within 17C12C is cleaved in 3E7 DNA. All the probes lie on separate *Sfi*I fragments, although the pY48 and pKFC43 *Sfi*I fragments are similar in size and are not distinguishable at the pulse time used in Figure 3.16. A series of single and double digests (Table 3.1) allowed a long range restriction site map of the region to be constructed (Figure 3.17). Additional experiments showed that the Y alphoid probe, pYαI, detected the same *Sfi*I and *Mlu*I fragments as 64a, while the more distal probe 14A3C detected the same *Bss*HII and *Mlu*I fragments as pKFC43 (results not shown). These findings show that the map in Figure 3.17 spans intervals 8 - 11 of the Y deletion map.

3.6. SUMMARY OF CHAPTER 3.

In this chapter the sequence organisation of the DNA in the pericentric region of the long arm of the human Y chromosome was analysed. This was achieved by determining the structures of one cosmid and three YAC clones. The region was found to consist of a mosaic of the known 5bp, 48bp and 68bp tandemly repeated sequences and at least five novel repeated sequence families. A long range map of ~3.5Mb of genomic DNA was constructed

which placed the clones between ~500kb and ~850kb from the long arm edge of the centromeric alphoid array.

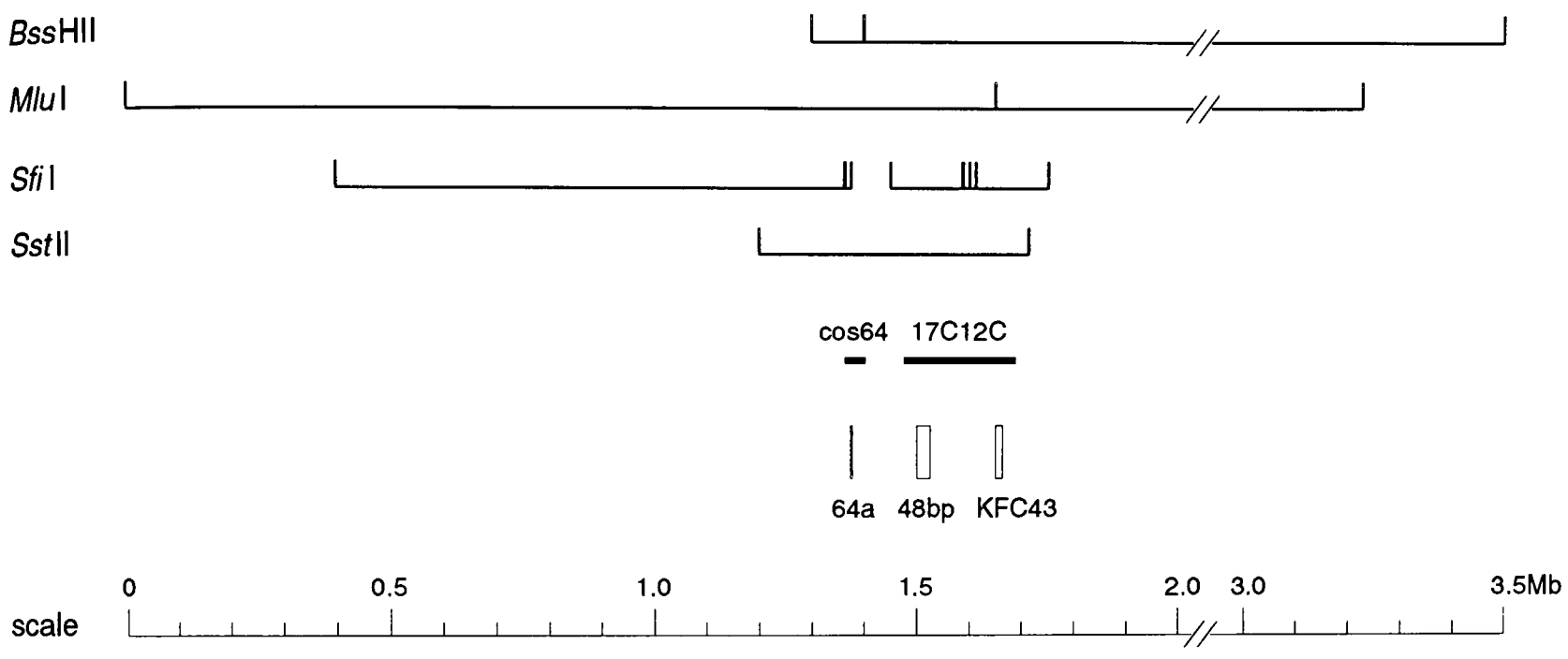


Figure 3.17. Long range map of the pericentric region of Yq. The genomic restriction enzyme sites are shown and the position of cos 64 and YAC 17C12C drawn underneath. The location of 64a, the 48bp satellite and the novel repeat probe pKFC43, (the loci used for mapping), are also shown.

CHAPTER 4

ISOLATION AND SELECTION OF YACS FROM THE PERICENTRIC REGION OF THE HUMAN Y CHROMOSOME.

The results from the previous chapter demonstrated that it was possible to use YACs to determine the structure of DNA from the pericentric long arm region of the human Y chromosome. Therefore it was decided that a major part of this project/thesis would be to analyse YACs from the entire centromeric region of the human Y chromosome. To achieve this YACs were isolated from the CEPH (Albertsen *et al.*, 1990) and Washington, St. Louis, (Burke *et al.*, 1987) YAC libraries by screening each with selected centromeric Y chromosome probes.

The probes used to screen each library were detremined by the method in which the library was screened. Filters were avaliable for the Washington, St. Louis YAC library for screening by colony hybridisation. Therefore centromeric, Y chromosome specifc probes (GMGY6 and YαII 5.1), X-Y chromosome specific probes (pDP34 and GMGXY8) and non-Y specifc, low copy number probes (64a, RBF5 and 14A3C) were used to screen the library. The CEPH library was screened by PCR by the method outlined by Green and Olson (1990a) as filters for colony hybridisation were not avaliable. The sequence tagged sites (STS) chosen to screen the library with ideally needed to amplify products that were Y chromosome specific. However only one centromeric STS used was Y specific (YαI primers). The other loci either were X-Y specific (AMGL) or low copy number repeats (64a and YαII).

4.1. SCREENING THE CEPH YAC LIBRARY.

4.1.1. Primary screening of the 113 DNA pools by PCR.

The CEPH library was screened for YAC clones containing YαI, YαII, AMGL and 64a loci by PCR. Details of the method used, the nucleotide sequence of the primer pairs, the amplication product size and the PCR conditions have been described in section 2.6.3. and summarised in Table 2.4. The loci chosen to screen the library were selected as they map into deletion intervals 7 (AMGL), 8 (YαII and YαI) and 9 (64a) of the Y chromosome deletion map (C. Tyler-Smith, personal communication and see Figure 1.8). An example of the results obtained is given in Figure 4.1 which shows the results of screening

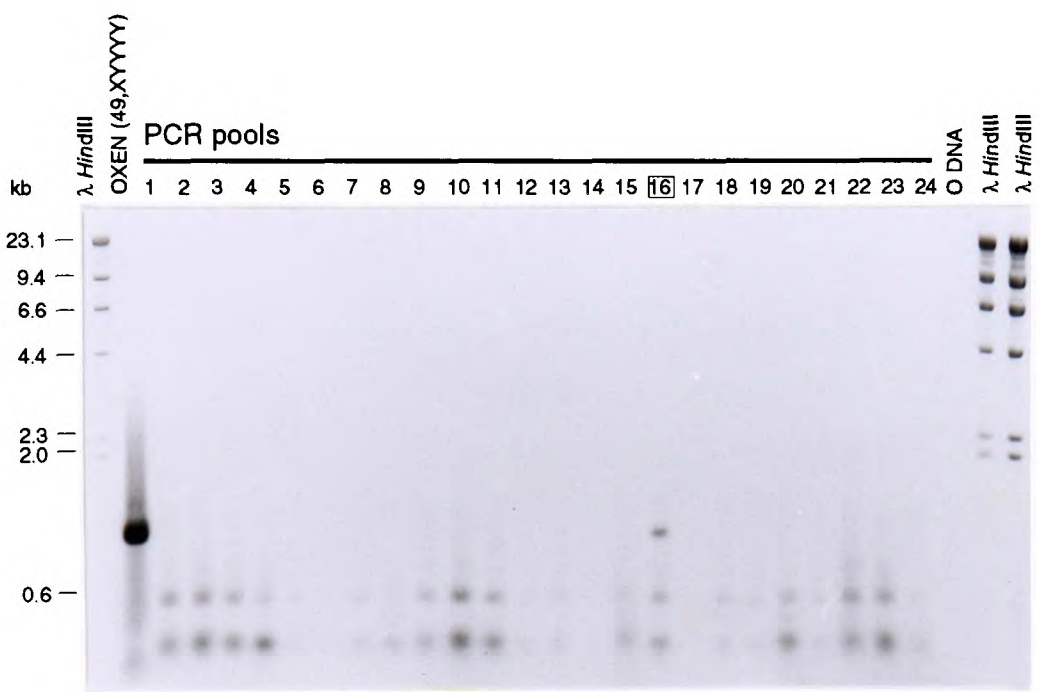


Figure 4.1. Photograph of an ethidium bromide stained agaorse gel showing the results of the primary screening of the CEPH library PCR pools 1-24 with Y αI primers. Only OXEN (49, XYYYY) DNA and pool 16 amplify the 1.0kb product. The position of the λ *Hin*dIII size markers is shown on the left hand side.

primary pools 1-24 with YαI primers. It can be clearly seen that pool 16 amplifies a 1.0 kb fragment which is the same size as the positive male control (OXEN). All the other pools and the no DNA control do not amplify. Therefore pool 16 is positive for this locus and pools 1-15, 17-24 are negative.

Five additional primary pools were found to be positive for YαI and these are listed in Table 4.1. However, before the secondary screen was implemented, the positive pools were reamplified to check that the original screening was correct. The results are shown in Figure 4.2.a. The six positive pools (tracks 5-10) and male DNA (track 2) all amplified a 1.0kb fragment which is absent from female DNA (track 3), the no DNA control (track 11) and the *S. cerevisiae*, yeast only, control (track 4). These results confirm that the six positive pools contain alphoid YACs that are derived from the Y chromosome. Finally, for positive pools 56, 71 and 105 additional bands were amplified which were not seen in male DNA. It is possible that these bands could represent sequences which are derived from the edges of the alphoid block but a fuller interpretation of this result was not possible at this stage.

The CEPH library was screened in a similar fashion with YαII, 64a and AMGL (amelogenin) primer pairs and the results are shown in Figure 4.2.b, c and d respectively. In each case, primary pools were identified which amplified a product the same size as male DNA. However, for primers YαII and 64a the size of the product was identical for male and female DNA (Figure 4.2.b tracks 4 and 5 for YαII and results not shown for 64a). Therefore YACs isolated from these pools would not necessarily be derived from the Y chromosome. The PCR results for the AMGL primers were easier to interpret as the Y locus amplifies a smaller product (Figure 4.2.c, track 2) than the X (track 3). Positive pools were therefore identified that contained YACs derived from both chromosomes (tracks 6-11 for the Y chromosome and tracks 12 (faint) and 13 for the X chromosome).

4.1.2. Secondary screening of the library.

The PCR primers, photographs of the primary screening results and conditions used were sent to D. Le Paslier at CEPH, Paris who performed the secondary screening of the library by PCR. A list of the YACs identified and the corresponding primer pairs are given in Table 4.1. and YACs which were later confirmed to be derived from the Y chromosome are marked with an asterisk. The YACs which are true positive but not derived from the human Y chromosome and YACs which are negative are also summarised in this

Table 4.1. List of the positive PCR pools and the corresponding YACs isolated after the secondary screen by CEPH.

PCR primers	Pool	YACs identified by CEPH
Yαl	16	63A2, 63A3, 64B9, 62C12*, 64G11
	47	none
	56	224H121, 224H1, 224G3*, 222F5, 221C1, 221A5, 223H4, 224A2, 224A8, 222B9, 222B10, 221G1
	71	284B4, 281B10, 284D12, 282G2, 284H1, 281C5*
	92	365A6, 366A6, 367A6, 368A6*
	105	420H1, 419F6, 417D3*, 419A2, 417B4, 418C4, 418H8, 417H10, 420C7
YαII	2	8D3
	5	20E8
	12	45H4
	16	62C3
	71	281C5*
AMGL	5	18C4*, 18C6, 18H4, 18H6, 19H4, 19H6, 19C4, 19C6
	19	73F9*
	59	235F3*
	68	271B5*
	74	296D12*
	74	297B7*
64a	19	74D2
	38	non-specific amplification
	70	non-specific amplification
	85	none
	107	427D11

*YACs derived from the Y chromosome.

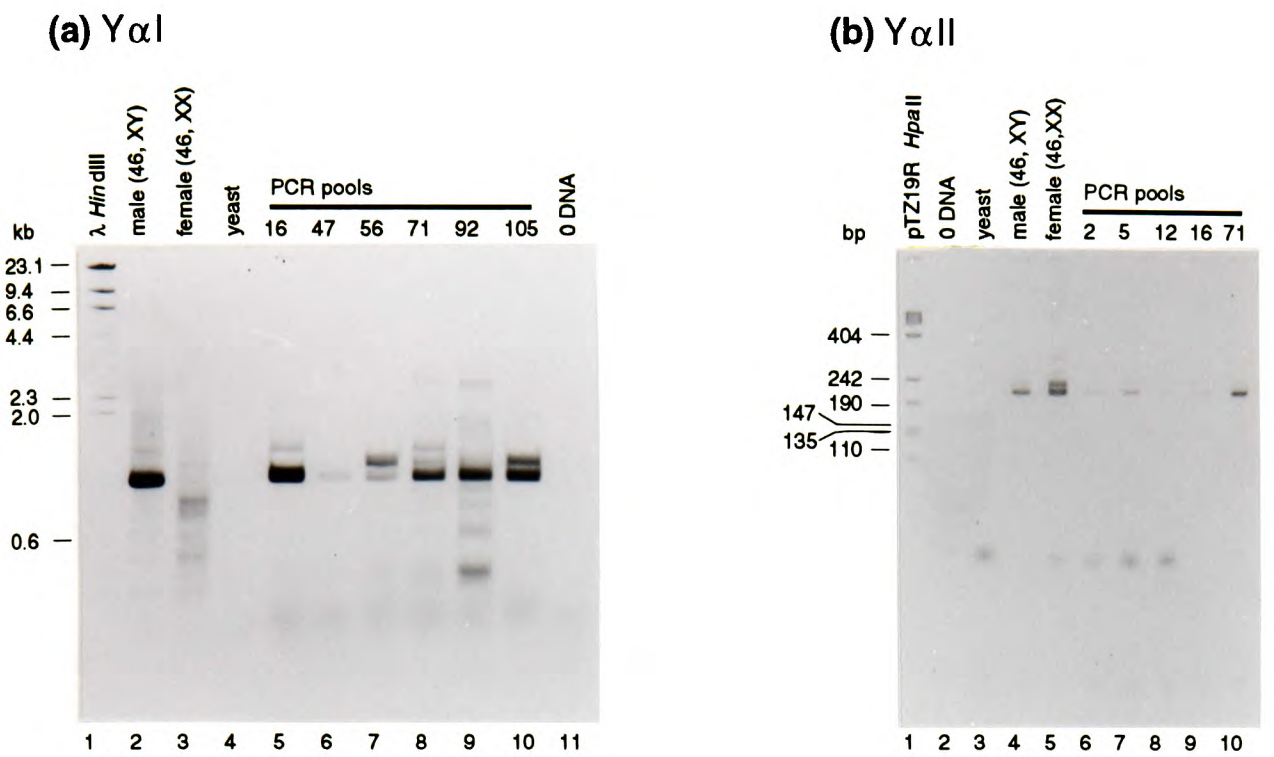


Figure 4.2. Photographs of ethidium bromide stained gels showing the results of the primary screening of the CEPH library by PCR. The primers used were (a) Yαl, (b) YαII, (c) 64a and (d) AMGL. For each primer pair a product was amplified in male DNA and the positive PCR pools. This product was not amplified in the no DNA and yeast only controls. The sizes of the λ *Hin*dIII and pTZ19R *Hpa*II size markers are shown on the left hand side of each photograph.

table. However, for several pools, (YαI pools 16, 56, 71, 92 and 105; AMGL pool 5), more than one YAC was returned from CEPH after the secondary screen. Therefore further screening of these YACs was required to determine which YACs were derived from the Y chromosome. This was achieved by hybridisation or PCR analysis of the YACs. The results for the alphoid YACs are discussed in section 4.3.1. In AMGL pool 5 YAC 18C4 was positive (result not shown).

The results obtained by D. Le Paslier at CEPH for the secondary screen of the library were not entirely satisfactory as YACs were not isolated from several pools (YαI pool 47; 64a pools 38, 70 and 85) which clearly amplified in the primary screen. Therefore these primary pools were reamplified using the appropriate primers and again for every pool the correct products were obtained, reconfirming that the pools were positive. However, despite repeating the secondary screen twice, YACs were never isolated from these pools.

4.2. SCREENING THE WASHINGTON, ST. LOUIS YAC LIBRARY.

4.2.1. Colony hybridisation.

The Washington, St. Louis YAC library was screened by colony hybridisation with selected Y chromosome centromeric probes. In total, the library was screened twice, each screening being performed with a different probe mixture (see Table 4.2). The first mixture contained non-centromeric probes (*DXYS64* and *ZFY*) as the library was screened in conjunction with another researcher who wanted to locate YACs for these loci. For each screening a number of YACs were identified and these results are given in Table 4.2. alongside the corresponding probe mixture. Specific filters from the library were screened a third time with probes 64a and *DXYS64* which had been used in the first mixture. This was because further analysis of the YACS identified with the first probe mixture revealed that there were no YACs which contained these loci. In total, 77 YACs were isolated from this library and a description of which YACs were selected for further analysis is given in section 4.4.

Table 4.2. List of YACs selected for analysis from the St. Louis, Washington YAC library after the primary screen by colony hybridisation.

Probe mix	YACs identified
64a, GMGY6, DXYS64, ZFY	AA3 C9/3, AA3C9/4, AA3B2/13, AA8F1/7, AA8F1/8, AA9E9/3, AA11E11/15, AA12D12/12, AA12D12/16, BB2D2/3, BB2G11/12, BB3G2/11, BB3G3/211, BB4E5/5, BB6A4/8, BB7B9/12, BB7C8/11, BB7C8/15, BB7H3/11, BB7H4/15, BB11F10/5, BB11F10/6, BB12A11/5, BB15E5/3, BC17A5/8
GMGXY8, pDP34, 2D5, YαII5.1, 14A3C, RBF5	AA1G10/11, AA3B7/13, AA4A7/11, AA5F2/14, AA6D2/16, AA6 A6/3, AA6 A6/4, AA7 E6/9, AA9G9/13, AA15G8/6, AA16 C3/4, AA16G2/16, AA16H4/2, AA16H12/8, AA17E7/5, AA17E7/6, AA17E8/5, AA17E8/6, BB1D1/1, BB3 F6/14, BB3G7/5, BB3G7/6,BB3F8/14, BB3A10/10, BB3A10/11, BB5F8/4, BB7F3/1, BB8G1/3, BB8G1/4, BB8H10/1, BB8H10/2, BB14F2/2, BB15D2/11, BB15D2/12, BB15D2/15, BB15D2/16, AC19E4/14, AC19E5/14, AC19F4/14, AC19F4/15, CC1C10/6, CC1F8/5,CC1F9/5
64a, DXYS64	AA12D12/8, BB2G11/12, BB2G11/8, BB2G11/6, BB11F10/5, BB11F10/6, BB11F10/9, BB11F10/10, BB17A6/5

4.2.2. PCR screening.

The Washington, St. Louis library was also screened by PCR with YαI and 64a sequence tagged sites. This screening was performed by The MRC Resourse Centre, Harrow and unfortunately at the time of writing this no YACs had been isolated despite the primers being sent to the centre in January.

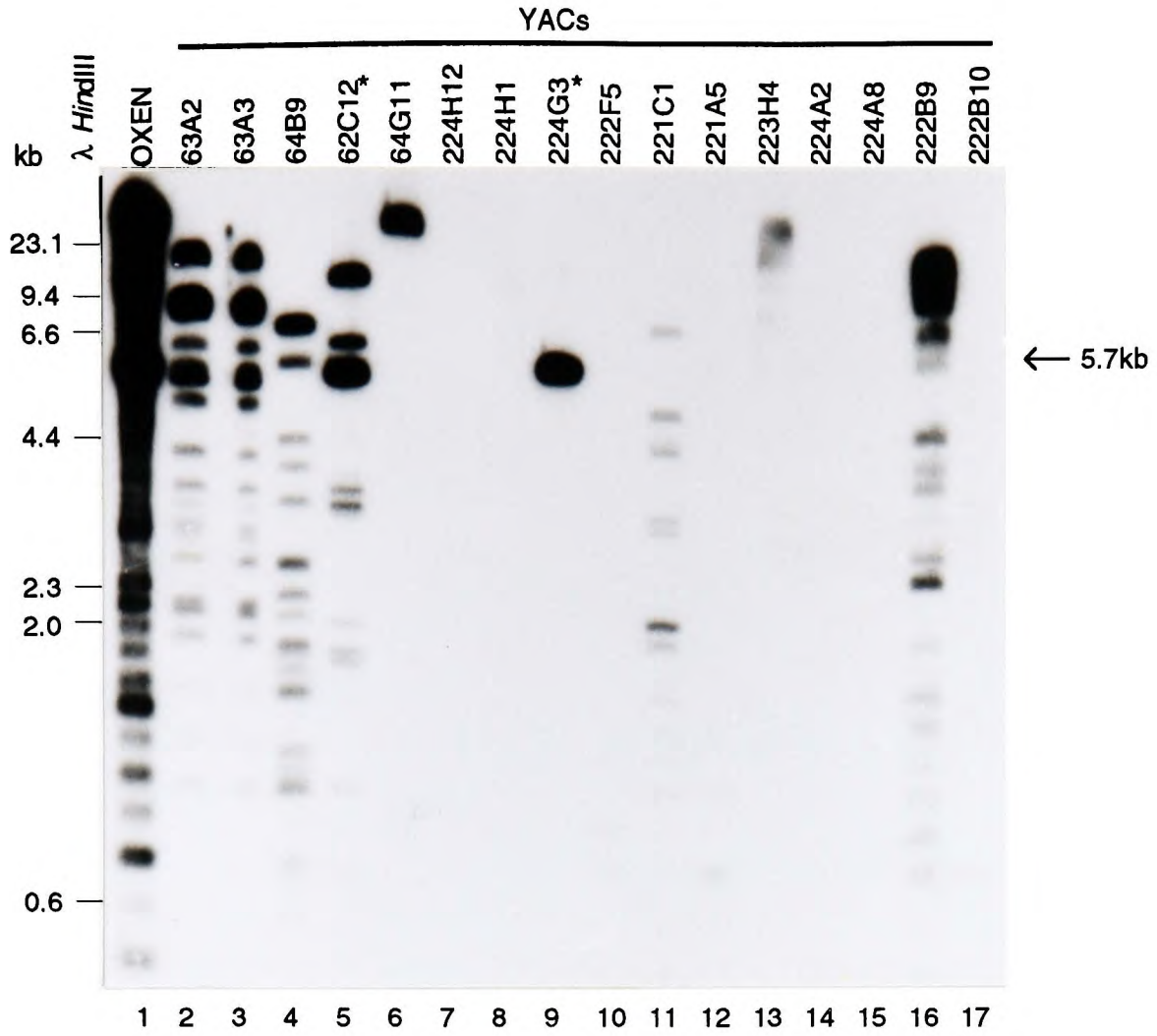
4.3. YAC SELECTION, CEPH LIBRARY.

4.3.1. Selection of alphoid YACs for further analysis.

4.3.1.a. Identification of Y alphoid YACs.

In total, 36 YACs were identified at CEPH by screening the six positive PCR pools using the YαI primers (Table 4.1). The first experiments performed were therefore designed to determine which of these YACs contained Y chromosome alphoid DNA. The structure of the human Y chromosome alphoid DNA was discussed in section 1.5.1 but it is important to reiterate here that on the Y chromosome the single major centromeric block of alphoid DNA consists of 170bp tandemly repeating subunits which are organised into a higher order unit structure with a periodicity of 5.7kb which is revealed by *Eco*RI. DNA from the YACs (isolated from pools 16, 56, 71 and 105) and the OXEN cell line was digested with *Eco*RI and analysed by conventional agarose gel electrophoresis, filter transfer and hybridisation at 68°C with YαI. These low stringency conditions were used to identify YACs which might contain alphoid DNA from the edges of the alphoid block which could be diverged. In addition, YACs which contain alphoid DNA which is not derived from the Y chromosome would be identified. The results of the filter hybridisation are shown in Figure 4.3. In tracks 5, 9, 25 and 28 much of the hybridisation was to a 5.7kb fragment suggestive of these YACs being derived from the human Y chromosome. Additional hybridising fragments were also detected in three of these Y chromosome alphoid YACs (tracks 5, 25 and 28). It was thought that these YACs were therefore derived from the edges of the alphoid block where the 5.7kb higher order unit structure becomes more diverged.

The filters were stripped and reprobed with the 5bp satellite, (pHY10), as previous work (Tyler-Smith and Brown, 1987) had demonstrated that this satellite could be physically linked to the long arm edge of Y chromosome alphoid DNA. The results are not shown but YAC 417D3 was the only "Y-alphoid YAC" which contained sequences which hybridised to the 5bp



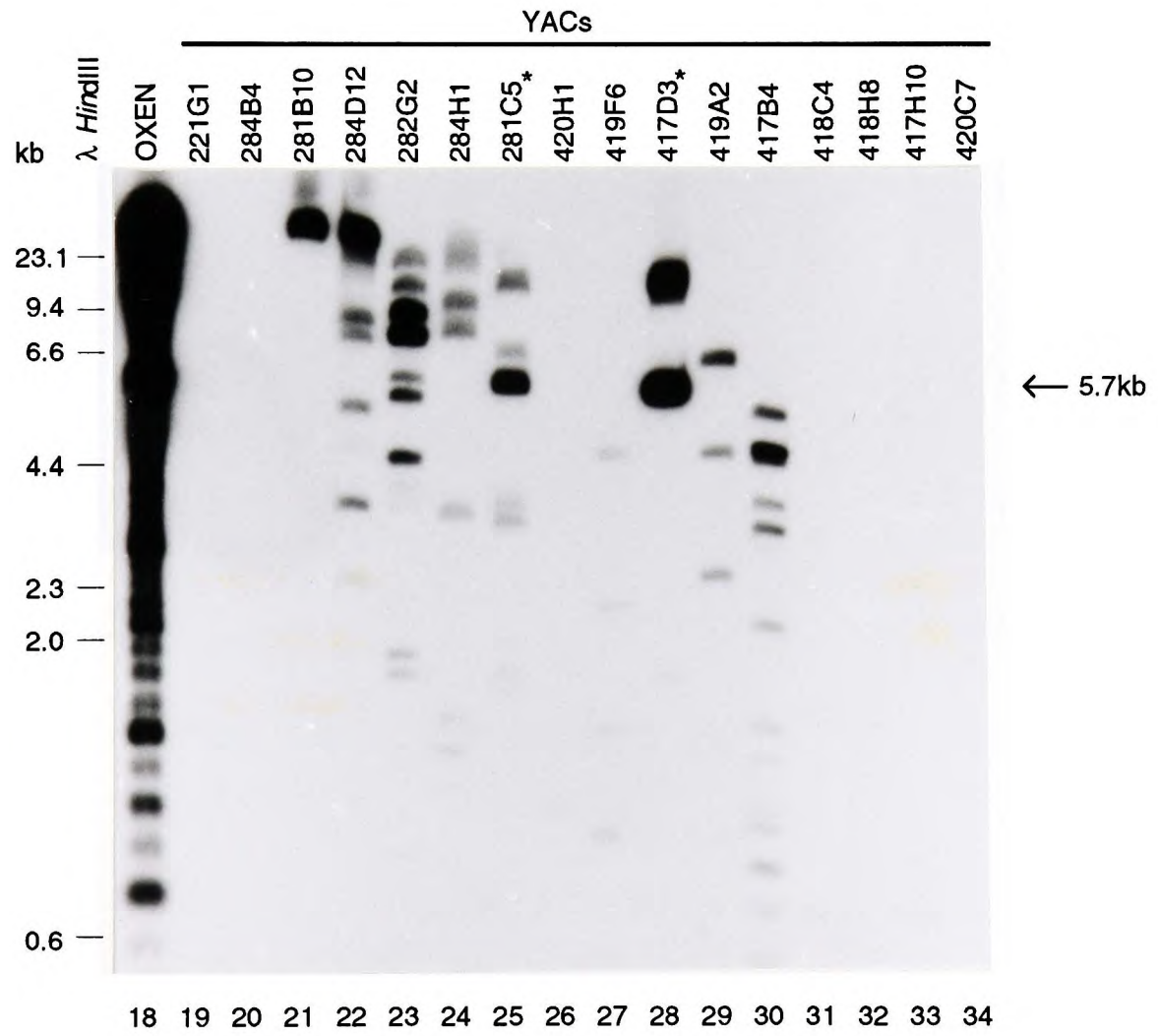


Figure 4.3. Southern blot analysis of the alphoid YACs isolated from the CEPH library showing how YACs derived from the human Y chromosome were detected. OXEN and YAC DNA was digested with *Eco* RI , analysed by conventional agarose gel electrophoresis and a filter transfer hybridised with YαI at 68°C. The filters were washed in 0.1 x SSC. The position of the 5.7kb alphoid unit is indicated as well as the position of the λ *Hin*dIII size markers. YACs derived from the Y chromosome are marked with an asterisk and 63A2 and 63A3 are duplicates of the same YAC.

Table 4.3. List of YACs identified from the CEPH library containing alphoid DNA.

YAC	comment
63A2	non-Y
63A3	non-Y
64B9	non-Y
62C12	Y, diverged edge of block
64G11	non-Y
224G3	Y, middle of the block (co-ligated)
221C1*	non-Y
223H4	non-Y
222B9*	non-Y
281B10	non-Y
284D12	non-Y
281G2	non-Y
284H1	non-Y
281C5	Y, diverged edge of block
417D3*	Y, diverged edge of block
419A2*	non-Y
419F6*	non-Y
417B4	non-Y
368A6	Y, middle of the block

*These YACs also hybridise to pHY10, (Y specific 5bp satellite probe), at low stringency.

satellite. A detailed analysis of this YAC is presented in section 5.1.3. Therefore it was concluded that YACs 62C12, (tracks 5), and 281C5 (track 25) were derived from the short arm edge of Y-alphoid DNA, YAC 224G3 (track 9) from the middle of the alphoid block and YAC 417D3 (track 28) from the long arm edge.

Hybridisation of YαI was also seen to many of the other YACs as the YαI probe is not Y-specific when it is used at 68°C. None of these YACs contained the 5.7kb higher order unit structure so it was decided that they were not derived from the human Y chromosome. However it was interesting to note that 4 of these YACs also contained the 5bp satellite as well as alphoid DNA, suggesting that alphoid DNA could be physically linked to the 5bp satellite on other human chromosomes.

The YACs isolated from pool 92 were primarily analysed by PCR to determine which YACs were derived from the human Y chromosome. The results are not shown but only YAC 368A6 amplified the 1.0kb male specific product using the YαI primers. Further analysis of this YAC demonstrated that it is derived from the middle of the Y chromosome block of alphoid DNA. Finally, Table 4.3 lists the YACs isolated from the CEPH library which contain alphoid DNA.

4.3.1.b. Alphoid block size of Y alphoid YACs.

It was necessary to determine the block size of alphoid DNA cloned in each "Y-alphoid YAC". In addition, the single YAC (BB8G1/4) isolated from the Washington, St. Louis library which contained alphoid DNA (see section 4.4) was also included in this analysis.

Initial studies on the long-range structure of Y chromosome alphoid DNA were reported by Tyler-Smith and Brown, 1987. They identified five restriction enzymes that did not cut within the major centromeric block of alphoid DNA. One of these enzymes, *Pvu*II was chosen to determine the amount of alphoid DNA cloned in the "Y alphoid YACs". To achieve this the YACs were either digested with *Pvu*II or left intact, fractionated by pulsed-field gel electrophoresis and a filter transfer was hybridised with YαI at 68°C. The results are shown in Figure 4.4. YACs 62C12, 281C5, BB8G1/4, 224G3 and 63A2 (non-Y-alphoid) all contain a small amount of alphoid DNA in proportion to the size of the YAC. This result can be explained for YACs 62C12, 281C5 and BB8G1/4 if these YACs are derived from the short arm edge

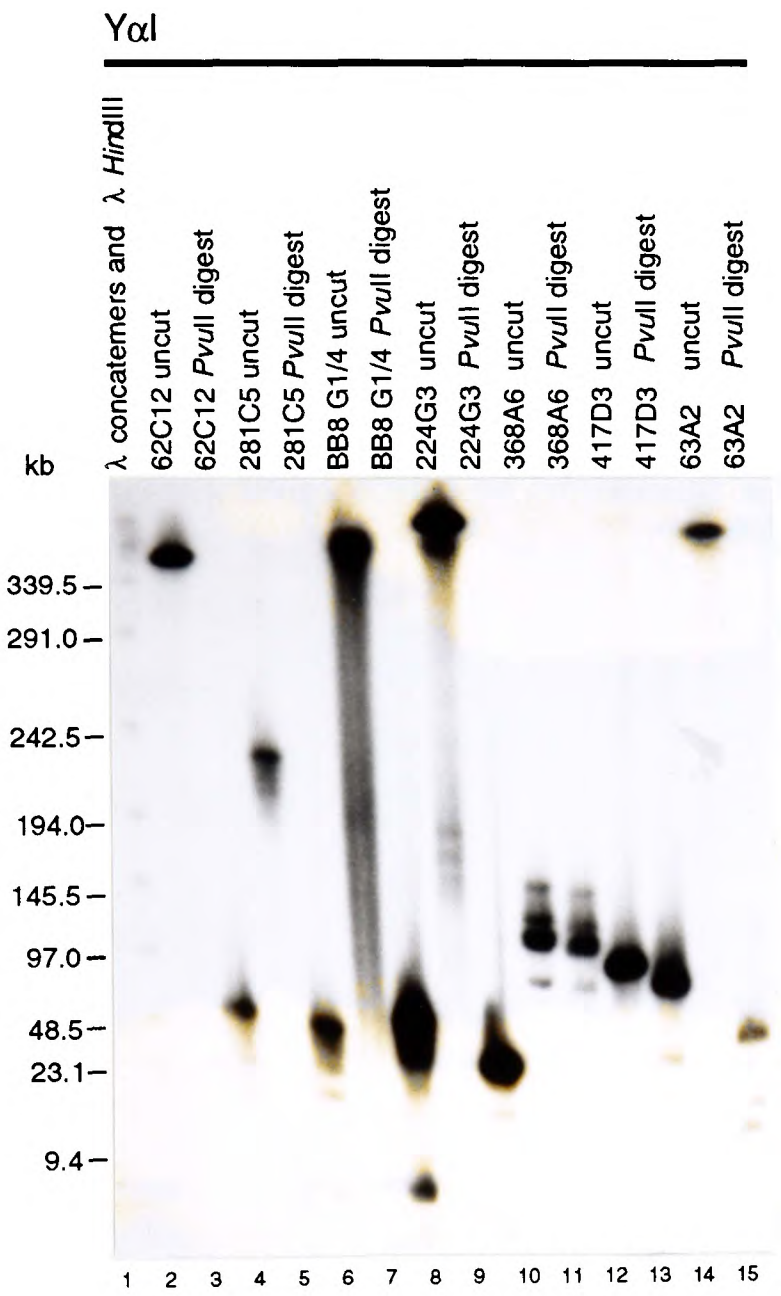


Figure 4.4. Determination of the amount of alphoid DNA located in the Y alphoid YACs. DNA for each YAC was either left intact or digested with *Pvu*II, and analysed by pulsed-field gel electrophoresis using an effective pulse time of 18 seconds for 32 hours. A filter transfer was hybridised with Yαl at 68 ° C. The filter was washed in 0.1 x SSC and the position of the λ size maerkers is shown on the left hand side.

of alphoid DNA (see section 4.3.1.a) and in addition to alphoid DNA contain other Y specific sequences (discussed in section 4.4). However, the small amount of alphoid DNA present in YAC 224G3 (~60kb) was unexpected as the results presented in section 4.3.1.a suggested that this large (>1Mb) YAC should entirely consist of alphoid DNA. It was concluded that YAC 224G3 must be co-ligated as it does not span either edge of the alphoid array which is ~300kb in the individual the YAC is derived from (C. Tyler-Smith, personal communication). Therefore this YAC was not analysed any further.

The size of the alphoid DNA found in YACs 368A6 and 417D3 was entirely consistent with the results presented in section 4.3.1.a. Both of these YACs are small, (110kb and 90kb respectively), and after digestion with *Pvu*II the size of their alphoid block either remains unchanged (368A6) or is only slightly reduced (417D3). Therefore YAC 368A6 is not co-ligated and consists entirely of middle-of-block alphoid DNA whereas YAC 417D3 consists mainly of edge-of-block alphoid DNA but in addition does contain ~10kb of the 5bp satellite sequence. Middle-of-block alphoid DNA refers to alphoid DNA which derived from the centre of the major centromeric block of Y alphoid DNA. It is entirely composed of the 170bp unit and has the higher order periodicity of 5.7kb but has no other distinguishing features. Edge-of-block alphoid DNA however retains the 170bp subunit structure but the 5.7kb unit is often disrupted as the DNA becomes more diverged.

Thus as a result of this analysis, 3 out of 36 CEPH alphoid YACs were selected for detailed analysis: 62C12 and 281C5 which possible span the short arm edge of the array, and 417D3 which possible spans the long arm edge.

4.3.2. Analysis of YαII positive YACs.

The PCR primers used for isolating YαII YACs were not Y chromosome specific. They were derived from the nucleotide sequence adjacent to the 170bp alphoid repeating subunit (see Figure 4.5). YαII itself is a small (approximately 18kb) diverged block of alphoid DNA located about 100kb distal to the main block of centromeric alphoid DNA on the short arm of the Y chromosome (C. Tyler-Smith, personal communication and this work). Flanking YαII there are two non-alphoid low copy number repeated sequences (YαII 5.1 and YαII 2.1). The first of these sequences, YαII 5.1, is Y chromosome specific and there are two copies of the locus. These are called YαII 5.1-proximal (centromeric) and YαII 5.1-distal and their location on the

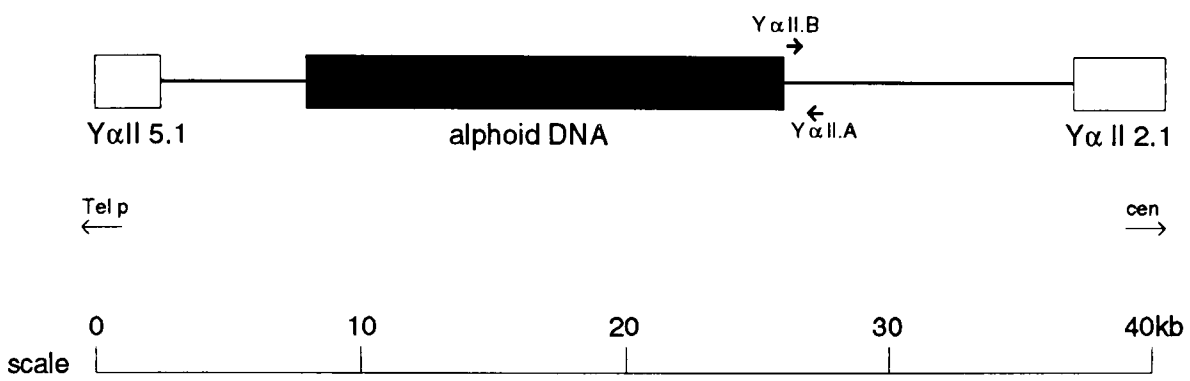


Figure 4.5. Structure of the Yα II locus showing the location of PCR primers Yα II.A and Yα II.B.

Table 4.4. Analysis of YαII YACs isolated from the CEPH YAC library.

YAC	PCR pool	PCR	YαII5.1	YαII2.1	YαI	size (kb)
8D3	2	+	-	+	+	~300
20E8	5	+	-	+	+	~100
45H4	12	+	faint	+	+	~600
62C3	16	-	-	-	-	~450
62C12	16	+	+	+	+	370
281C5	71	+	-	+	+	230

deletion map is given in Figure 1.8. YαII 2.1 however, is not Y chromosome specific and the location of other copies of this repeat are unknown.

Since the YαII primers did not specifically amplify the Y locus it was necessary to identify YACs which were derived from the Y chromosome after the PCR screening was completed. In total, CEPH identified 5 YACs containing this locus, one YAC corresponding to each positive PCR pool (Table 4.1). The YACs were colony purified and examined to determine which were derived from the human Y chromosome. Firstly the YACs were digested with *Hin*dIII, fractionated by conventional gel electrophoresis and a filter transfer sequentially probed with YαII 5.1, YαII 2.1 and YαI. Secondly, the size of each YAC was determined by separating the YAC from the host cell's chromosomes by pulsed-field gel electrophoresis.

The results, are summarised in Table 4.4. YACs 8D3, 20E8, 45H4 and 281C5 all hybridised to YαII 2.1 and YαI but not to YαII 5.1. None of the probes were located in YAC 62C3 and the lack of hybridisation to YαII 5.1 in all these YACs suggested a non-Y origin. Further insight into their chromosomal origin was gained by comparing their YαI hybridisation patterns to the pattern obtained by a YAC, (62C12), known to contain Y chromosome short arm alphoid DNA. Using this criterion YAC 281C5 was found to be derived from the Y chromosome as it detected bands in common with 62C12. Furthermore, the PCR screening of the CEPH library with YαI primers identified this YAC as

being derived from the Y chromosome (section 4.3.1). Despite YACs 8D3, 20E8 and 45H4 not being useful for this work it was interesting to discover that the locus detected by YαII 2.1 lies close to alphoid DNA on at least two other human chromosomes.

It was mentioned above that YAC 62C3, (pool 16), did not hybridise to any of the probes tested. Therefore it was checked to see if it would amplify a product using the YαII primers. The results, summarised in Table 4.4, show that 62C3 did not amplify a product. However, YAC 62C12 which was also isolated from pool 16 using the YαI PCR primers, also amplifies the YαII locus. Thus this YAC must account for pool 16 amplifying with YαII primers during the original screening of the library. Further analysis of this YAC (section 4.3.1 and chapter 5) confirmed that it was derived from the Y chromosome and also spans the YαII locus). The isolation of YAC 62C3 by CEPH however still remains a mystery as the YAC has consistently given negative results for all of these experiments.

4.3.3. Analysis of 64a positive YACs.

The primary screening of the CEPH library identified 5 pools using the 64a primers (Table 4.1). However, isolating the actual YAC which was responsible for the amplification was difficult. For pools 38 and 70 no YACs were identified at the secondary screening stage as CEPH always reported "non-specific amplification". Furthermore, no YAC was ever isolated from pool 85. YACs were isolated for pools 19 and 107 and these were further analysed to determine whether they contained the Y chromosome or the autosomal copy of the locus. YACs were digested with *Eco*R1, fractionated by conventional gel electrophoresis and a filter transfer was hybridised to 64a. No band was detected for YAC 427D11 (pool 107), but YAC 74D2, (pool 19), contains the autosomal copy of the locus as 64a detected a 2.0kb fragment as opposed to the 1.5kb Y chromosome locus.

Further experiments were therefore required to determine why YAC 427D11 was selected by CEPH. Firstly, the YAC was reamplified using the 64a primers and the correct sized product was obtained. Secondly, both 64a YACs were separated from the host cell's chromosomes by pulsed-field gel electrophoresis and a filter transfer sequentially probed with 64a and the left hand arm YAC vector probe. The vector probe was used to determine whether there was more than one YAC present in clone 427D11. The results are not shown but it was determined that 64a only hybridised to the

autosomal YAC (74D11), in agreement with the *Eco*RI digest (above), and the vector hybridised to both samples but only one YAC band was visualised for 427D11.

Two possible explanations for these results can be suggested. Firsty, there could be a third, undiscovered 64a locus whose only similarity to the known loci was that the primer binding sites are the same and in the appropriate orientation. Thus it would give a PCR product but no hybridisation. Secondly, there could be two YACs present in clone 427D11: the one containing the 64a locus being present at a very low concentration and hence giving a PCR product but no hybridisation. To test the first of these theories, the PCR product amplified from YAC 427D11 was used to probe the pulsed-field gel discussed above. Again the only hybridisation seen was to the autosomal YAC 74D1. Therefore YAC 427D11 does not contain a third copy of this locus. In conclusion, clone 427D11 is likely to contain two YACs, the 64a YAC being present at a very low concentration which then accounts for the positive PCR and negative hybridisation results. Unfortunately, lack of time prevented any further work being done on this YAC.

4.3.4. Analysis of AMGL positive YACs.

The primary screening of the library identified 6 pools using the AMGL primers (Table 4.1). One YAC from each pool was consequntly isolated at CEPH. The YACs were probe content mapped and the results are shown in Table 4.6. YAC 73F9 was not analysed any further for because it gave only weak hybridisation signal when probed with AMGL. Furthermore it did not contain pDP34 or GMGXY8 loci and as the region was well covered by other YACs it was put aside.

4.4. YAC SELECTION, WASHINGTON, ST. LOUIS LIBRARY.

The screening of this YAC library by colony hybridisation was discussed in section 4.2. In total, 77 YACs were identified using 10 different probes and firstly therefore it was necessary to sort out which loci were present in each YAC (probe content mapping). However, most of the probes used for mapping detect complicated loci and these are summarised in Table 4.5 and their exact location on the Y chromosme deletion map is shown in Figure 1.8. Therefore restriction enzyme analysis and hybridisation was used to determine which of the 77 YACs were (a) derived from the human Y chromosome, (b) derived from the human X chromosome, (c) derived from

Table 4.5. List of the different loci detected by the probes used for mapping the YAC clones.

Probe	Y specific[1]	Type of sequence detected	Location	Diagnostic restriction enzyme fragment[2]
Y190	+	repeat	-Y190 major Yp (distal to centromere) -Y190 minor Yp (proximal to centromere)	*Taq*I - 4.3kb[3]
GMGXY8	-	X-Y sequence	-Yp -Xq	*Taq*I - 14.0kb
pDP34	-	X-Y sequence	- centromeric region Yp -Xq	*Taq*I - 13.0kb
AMGL	-	gene	- centromeric region Yp -Xp	*Taq*I - 3.5kb
2D5	-	low copy number repeat	- centromeric region Yp (main locus) - several other non-Y locations	*Hin*dIII - 12.0kb
50f2	-	low copy number repeat	-1 locus centromeric region Yp (50f2d) - 4 loci non-centromeric Y (50f2a, b, c, & e) -1 non-Y locus	*Hin*dIII - 11.0kb[4]
YαII5.1	+	low copy number repeat	- centromeric region Yp (proximal to centromere) -Yp (distal to centromere)	*Hin*dIII - 7.8kb (proximal) *Hin*dIII - 3.5kb (distal)
YαII2.1	-	low copy number repeat	- centromeric region Yp - several other autosomal locations	*Hin*dIII
YαI	+ (at high stringency)	satellite	-Y chromosome centromere	*Eco*RI - 5.7kb
64a	-	low copy number repeat	- centromeric Yq - one other autosomal location	*Eco*RI - 1.5kb
GMGY6	+	Y sequence	- centromeric region Yq	*Hin*dIII - 1.6kb
14A3C	-	low copy number repeat	- centromeric region Yq - one other non-Y location	*Eco*RI - 4.5kb

1. Loci which are solely located on the Y chromosomeare scored as +.
2. Only the sizes of the Y chromosome centromeric loci are given.
3. The major and minor Y190 blocks cannot be distinguished between using this type of analysis. The Y190, 3.5kb probe was used.
4. The size of the locus detected by 50f2d is given.

Table 4.6. List of YACs which were found to be derived from the pericentric short arm region of the human Y chromosome. Details of the different loci present in each YAC is given. YACs which were isolated using these loci but which are not derived from the Y chromosome are also listed.

YAC	Y190 major	YαII 5.1 (D)[7]	GMG XY8	pDP34	AMGL	Y190 minor	2D5	50f2d	YαII 5.1 (P)[7]	YαII 2.1	YαI
AA17E7/5[1,4]	+	+	-	-	-	-	-	-	-	-	-
AA17E8/5[1,4]	+	+	-	-	-	-	-	-	-	-	-
BB3A10/11[1]		+	-	-	-	-	-	-	-	-	-
AA7E6/9[1]		+	-	-	-	-	-	-	-	-	-
CC1C10/6[1,5]	-	-	+ (Y)	-	-	-	-	-	-	-	-
CC1F8/5[1,5]	-	-	+ (Y)	-	-	-	-	-	-	-	-
CC1F9/5[1,5]	-	-	+ (Y)	-	-	-	-	-	-	-	-
AA1G10/11[1]	-	-	+ (Y)	+ (Y)	-	-	-	-	-	-	-
18C4[2]	-	-	-	+(Y)	+	-	-	-	-	-	-
AA3B7/13[1]	-	-	-	+ (Y)	+	-	-	-	-	-	-
297B7[2]	-	-	-	+(Y)	+	-	-	-	-	-	-
296D12[2]	-	-	-	-	+	-	-	-	-	-	-
271B5[2]	-	-	-	-	+	-	-	-	-	-	-
235F3[2]	-	-	-	-	+	-	-	-	-	-	-
AA4A7/11[1]	-	-	-	-	-	+	+	+	+	-	-
AA15G8/6[1]	-	-	-	-	-	+	+	-	-	-	-
BB8G1/4[1]	-	-	-	-	-	+	+	+	+	-	+
62C12[2]	-	-	-	-	-	+	+	+	+	+	+
13C1F[3]	-	-	-	-	-	-	+	+	-	-	-
BB8H10/2[1]	-	-	-	-	-	-	+	+	+	+	-
281C5[2]	-	-	-	-	-	-	-	-	-	+	+
368A6[2]	-	-	-	-	-	-	-	-	-	-	+
AA6D2/16[1]	-	-	+ (X)	+ (X)	-	-	-	-	-	-	-
AC19E4/14[1,6]		-	+ (X)	+ (X)	-	-	-	-	-	-	-
AC19E5/14[1,6]		-	+ (X)	+ (X)	-	-	-	-	-	-	-
BB15D2/15[1]		-	-	+ (? X or Y)	-	-	-	-	-	-	-
BB3G7/6[1]	-	-	+ (unknown locus)	-	-	-	-	-	-	-	-

1. YACs isolated from the St. Louis, Washington library.
2. YACs isolatedfrom the CEPH library.
3. YAC isolated from the Biochemistry, Oxford library by R. Fisher.
4., 5. & 6. Duplicates of the same YAC.
7. Distal (D) and proximal (P) loci of YαII 5.1.

Table 4.7. List of YACs derived from the pericentric region long arm of the human Y chromosome. Details of the different loci present in each YAC are given.

YAC	YαI	pHY10	64a	64b	KFC68	RBF5	GMGY6
417D3[1]	+	+	-				
74D12[1]	-	-	+[4]	-	-		
17C12C[3]	-		-	-	+	-	-
29B1C[3]	-		-	-	+	-	-
15B12F[3]	-		-		+	-	-
AA8F1/8[2]			-			+	+
BB3G2/11[2,5]			-			-	+
BB3G3/11[2,5]			-			-	+
BB7B9/12[2]			-			-	+
427D11[1]	-	-	+(non Y)	-	-		

1. YACs isolated from the CEPH library.
2. YACs isolated from the St. Louis, Washington library.
3. YACs isolated from the Biochemistry, Oxford library (discussed in detail in Chapter 3).
4. Positive by PCR but negative by hybridisation (Y or non-Y location unknown).
5. Duplicates of the same YAC.

Table 4.8. Location, size and library origin of YACs analysed in detail in this work.

YAC	Location	Loci used to isolate YAC	size (kb)	Library origin
AA3B7/13	pericentric Yp	pDP34	290	Washington, St. Louis
18C4	pericentric Yp	AMGL	340	CEPH
297B7	pericentric Yp	AMGL	350	CEPH
296D12	pericentric Yp	AMGL	290	CEPH
271B5	pericentric Yp	AMGL	450	CEPH
235F3	pericentric Yp	AMGL	500	CEPH
AA15G8/6	pericentric Yp	2D5	250	Washington, St. Louis
AA4B7/11	pericentric Yp	many e.g. 2D5	290	Washington, St. Louis
62C12	pericentric Yp	many e.g. YαI	370	CEPH
13C1F	pericentric Yp	50f2	140	Biochemistry, Oxford
281C5	pericentric Yp	YαI	230	CEPH
BB8G1/4	pericentric Yp	many e.g. 2D5	665	Washington, St. Louis
BB8H10/2	pericentric Yp	many e.g. 2D5	185	Washington, St. Louis
368A6	pericentric Y	YαI	110	CEPH
417D3	pericentric Yq	YαI	90	CEPH
15B12F	pericentric Yq	KFC68	255	Biochemistry, Oxford
17C12C	pericentric Yq	KFC68	230	Biochemistry, Oxford
29B1C	pericentric Yq	KFC68	110	Biochemistry, Oxford
9C3C	telomeric Y q	KFC68	115	Biochemistry, Oxford

Table 4.9. Location, size and library origin of YACs isolated in this work but not analysed in detail.

YAC	Location	Loci used to isolate YAC	size (kb)	Library origin
AA17E7/5	pericentricYp	Yα11, 5.1		Washington, St. Louis
AA7E6/9	pericentricYp	Yα11, 5.1		Washington, St. Louis
BB3A10/11	pericentricYp	Yα11, 5.1	180	Washington, St. Louis
CC1C10/11	pericentricYp	GMGXY8 (Y)	165	Washington, St. Louis
BB3G7/6	pericentricYp	GMGXY8 (Y)	220	Washington, St. Louis
AA1A10/11	pericentricYp	GMGXY8 (Y)	260	Washington, St. Louis
AA6D2/16	Xq	GMGXY8 (X)		Washington, St. Louis
AC19E4/14	Xq	GMGXY8 (X)	150	Washington, St. Louis
BB15D2/15	non-Y	GMGXY8 (X or Y)	200	Washington, St. Louis
73F9	pericentric Yp	AMGL	310	CEPH
8D3	non-Y	YαII	280	CEPH
20E8	non-Y	YαII	680	CEPH
45H4	non-Y	YαII	600	CEPH
62C3	non-Y	YαII	450	CEPH
224G3	pericentric Y	YαI	>1Mb	CEPH
74D2	unknown	64a	660	CEPH
427D11	non-Y	64a	340	CEPH
AA8F1/8	pericentric Yq	RBF5	195	Washington, St. Louis
BB3G2/11	pericentric Yq	GMGY6	160	Washington, St. Louis
BB7B9/12	pericentric Yq	GMGY6	360/200	Washington, St. Louis

the human Y chromosome but not located near the centromere and (d) not derived from the X or Y chromosome.

DNA was prepared from the YACs, digested separately with either *Hin*dIII, *Eco*RI or *Taq*I, fractionated through 1% conventional agarose gels and filter transfers hybridised to the probes listed in Table 4.5. The results are summarised in Table 4.6 (short arm YACs) and Table 4.7 (long arm YACs). The YACs isolated from the CEPH library and the Biochemistry, Oxford YAC libraries were also probe content mapped and the results are included in these Tables. It can been seen from Table 4.6 that the YACs from the short arm form three contigs with the gaps occuring between the YαII 5.1 distal and GMGXY8 loci and AMGL and Y190 minor loci. Two contigs were isolated for the long arm, YACs containing the 68bp satellite YACs forming the more proximal contig (see Chapter 3) and YACs containing RBF5 and GMGY6 loci forming the distal contig.

Finally, not all the YACs described in Tables 4.6 and 4.7 were analysed any further. The YACs which were analysed in detail are listed in Table 4.8 and are either discussed in Chapter 3 or Chapter 5. The YACs which were put aside are listed in Table 4.9.

Tab.4.6

4.5. SUMMARY OF CHAPTER 4.

The CEPH and the Washington, St. Louis YAC libraries were screened for YACs derived from the centromeric region of the human Y chromosome using a variety of selected probes. In total, 133 YACs were isolated from both libraries and the selection procedures were used to determine which YACs were derived from the centromeric region of the human Y chromosome. Finally, the 19 YACs which were selected for detailed analysis form 5 contigs and this result is drawn schematically in Figure 4.6. The 3 YACs which make up the 68bp satellite contig and a further 68bp satellite containing YAC, (9C3C), were described in detail in Chapter 3. Most of the remaining YACs fall into either the AMGL or alphoid contig and their analysis is described in detail in the next chapter.

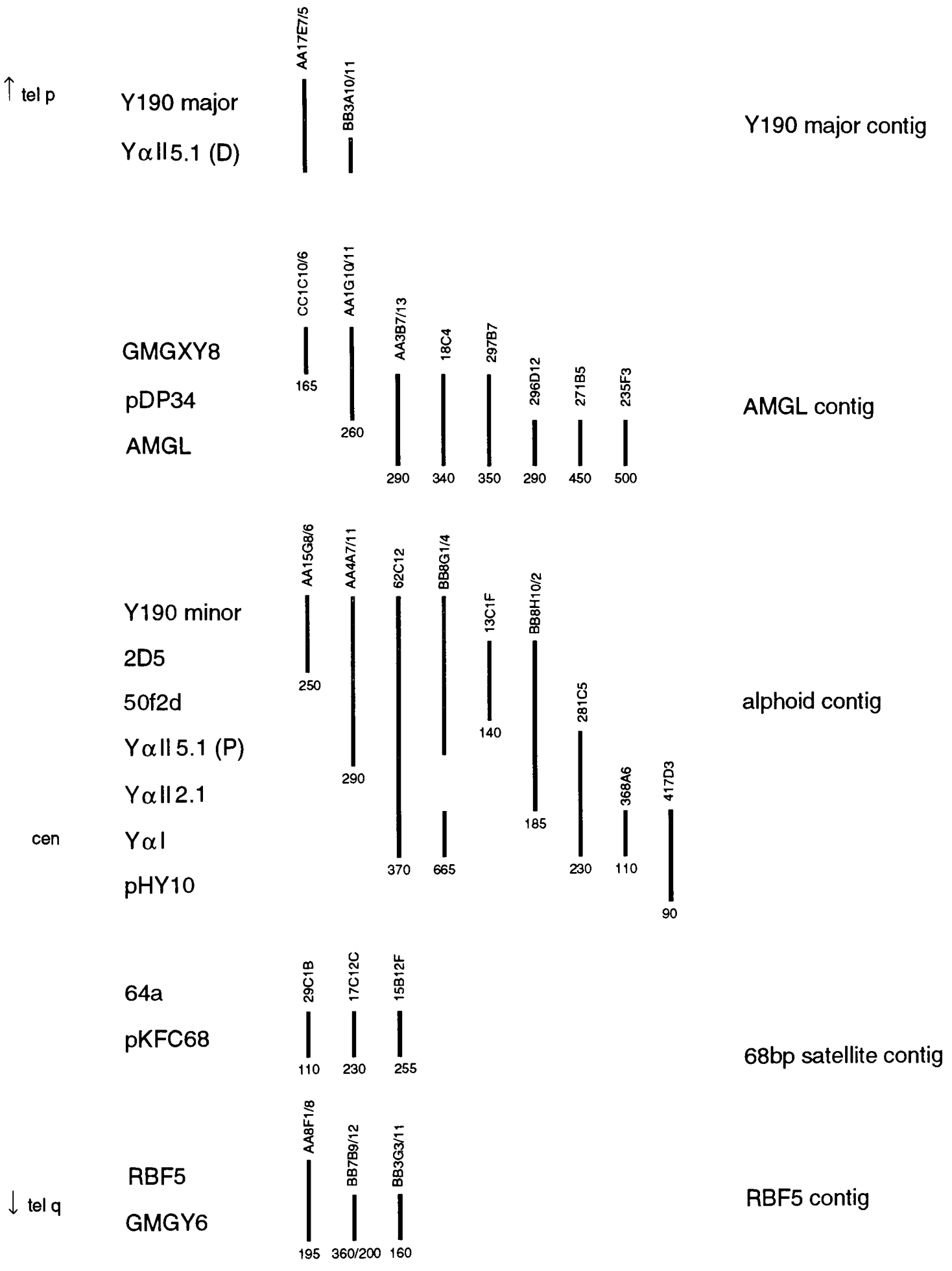


Figure 4.6. Diagramatic representation of the 5 YAC contigs isolated in Chapter 4 summarising which probes are present in each YAC. The YACs are not drawn to scale but the size of each YAC is indicated where known.

CHAPTER 5

STRUCTURAL ANALYSIS OF THE PERICENTRIC SHORT ARM AND CENTROMERIC REGION OF THE HUMAN Y CHROMOSOME.

The Y chromosome short arm YACs which were selected in Chapter 4 for further analysis form two major contigs: the alphoid and the AMGL contig. The alphoid contig contains 9 YACs which extend from the major centromeric block of alphoid DNA to the Y190 minor locus and the AMGL contig consists of 8 YACs which extend from the AMGL locus to GMGXY8. This was represented schematically in Figure 4.6 which predicts that these two contigs span an extensive centromeric region of the short arm. However, it was not known at this stage whether the distal ends of YACs from the alphoid contig overlapped with the proximal end of YACs from the AMGL contig. The gap between these two short arm contigs also corresponds to the gap in the genomic map of the whole of the centromeric region (C. Tyler-Smith, personal communication) which was discussed in section 1.5.2. Therefore one aim of analysing these contigs was to find probes which would close the physical map (section 5.3). The contigs were also analysed in detail in order to determine the underlying structure of genomic DNA from these regions and the results are presented in sections 5.1 (alphoid contig) and 5.2 (AMGL contig). Finally, YACs from the Y190 major contig were put aside because they lay outside the defined centromeric region and the RBF5 contig YACs were analysed by another researcher.

5.1. ANALYSIS OF YACS FROM THE ALPHOID CONTIG.

5.1.1. Colony purification and instability of alphoid YACs.

5.1.1.a. Instability.

Many YACs containing repeated sequences, such alphoid DNA, are unstable (Vilageliu and Tyler-Smith, 1992) and this instability can occur at a frequency of more than one event per two generations (Neil *et al.*, 1990). The instability of some sequences can be overcome in a *rad52* background (Vilageliu and Tyler-Smith, 1992) However, the YACs analysed here were not transformed into this new host because only the clones containing alphoid DNA (417D3, 368A6, 281C5, 62C12 and BB8G1/4) were found to be very unstable and it was

felt that the previous work on alphoid instability would allow the data obtained from these YACs to be interpreted. YAC 235F3 was found to be slightly unstable and this is discussed further in section 5.2.1.

The first step in the detailed analysis of these YACs was colony purification. For the unstable clones it was important to select a daughter colony representative of the parenet clone. This was done by fractionating the undigested parent and 16 colony purified daughter clones by pulsed-field gel electrophoresis and hybridising a filter transfer with YαI. The results for YAC 417D3 are shown in Figure 5.1.a. If the clones were stable a single band of the same size would be seen in each track. However, due to the structural instability of the DNA in the YAC, multiple bands are seen in both the parent and daughter clones: sequences are deleted during growth. Therefore for each alphoid YAC the largest stable daughter clone was analysed. For YAC 417D3 this was clone 5 (Figure 5.1, track 6).

YAC 417D3 contains 5bp satellite sequences as well as alphoid DNA. Therefore to ascertain whether this satellite was deleted from any of the daughter clones the membrane re-hybridisied with pHY10. The results are shown in Figure 5.1.b and it can be seen that all the daughter clones retained alphoid DNA and some 5bp satellite but the amount of 5bpsatellite in clone 7 is reduced.

The Y190 major locus (DYZ5) has also been reported by Neil *et al.* 1990 to be unstable in YAC clones. However, it is not known whether the Y190 minor locus is unstable in YAC clones. Both of these loci are located on the short arm of the human Y chromosome and the major block is hypervariable in size and has a periodicity of ~20kb. The smaller, minor block is not variable in size and the details of its structure are not known (Tyler-Smith *et al.*, 1988). Two of the YACs from the alphoid contig which contain Y190 minor sequences (62C12 and AA4A7/11) were therefore observed to see if they were unstable. A further two YACs, (BB8G1/4 and AA15G8/6), from the alphoid contig which also contain Y190 minor sequences were not included in this analysis as YAC BB8G1/4 is co-ligated and deleted and YAC AA15G8/6 was difficult to colony purify (see section 5.1.2.b) The final results are not shown but clone 62C12, which contains small amount (~10kb) of the Y190 minor locus, was found to be unstable. However, this was probably because the clone also contains alphoid DNA. The second clone, AA4A7/11 spans the Y190 minor locus (see Table 4.6) but did not show any signs of instabilty. Therefore

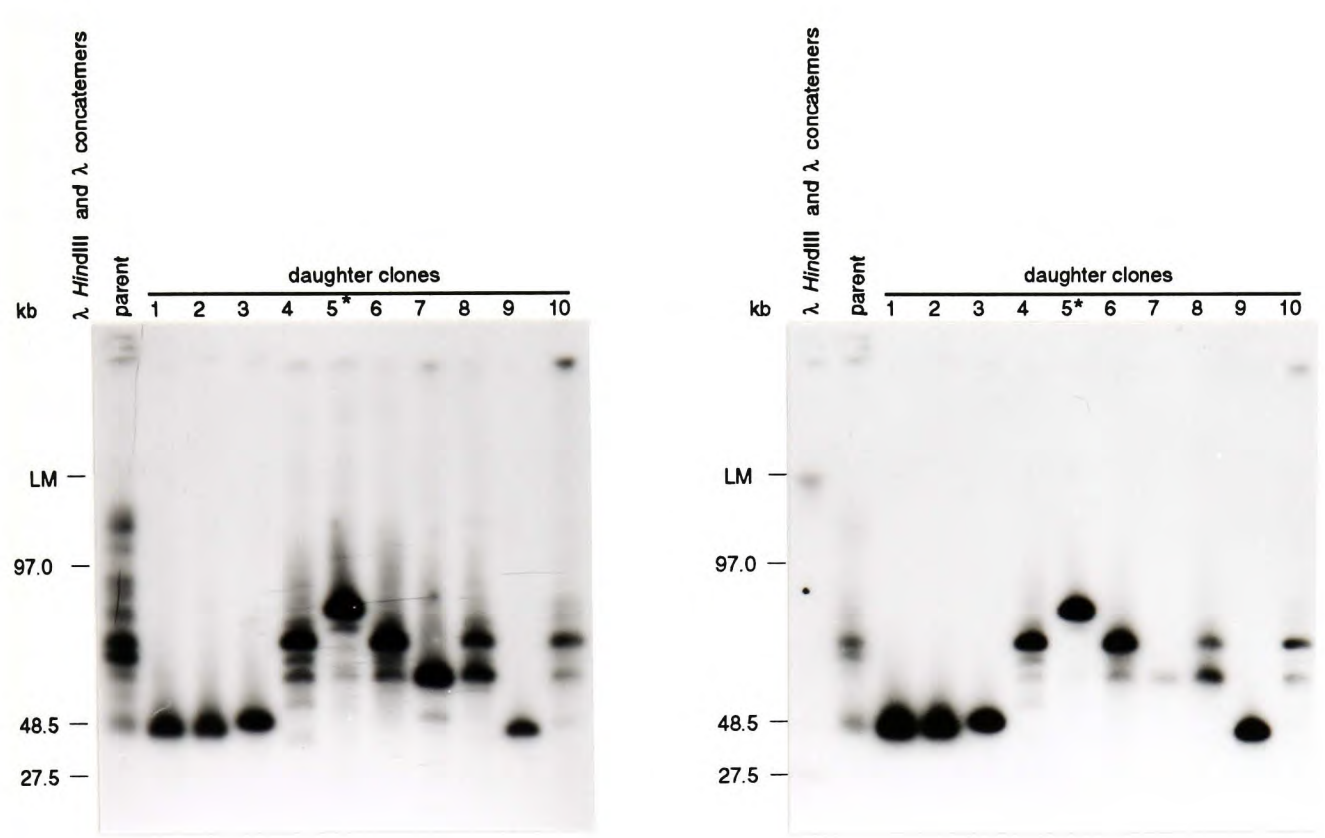


Figure 5.1. Analysis of ten undigested daughter clones from YAC 417D3 demonstrating the instability of alphoid DNA cloned in YACs. The clones were analysed by pulsed-field gel electrophoresis (50hr run time and 4 secs effective pulse time) and hybridisation to (a) Yαl and (b) pHY10. The filters were washed in 0.1 (a) and 2 x SSC (b).The daughter clone which was chosen for further analysis is marked with an asterisk and the position of the λ size markers are given on the left hand side of each photograph. LM = limit mobility.

the Y190 minor locus, unlike the Y 190 major locus is not unstable in YAC clones.

5.1.1.b. Problems with colony purification of YACs BB8H10/2 and AA15G8/6.
Analysis of the undigested YAC clones BB8H10/2 and AA15G8/6 would be expected to show a single YAC band. However the clones showed two and three bands respectively. To investigate whether the extra bands corresponded to unrelated clones or reflect instability of the sequences originally cloned, the cultures were streaked out on SD agar and single colonies were picked for analysis. After separation by PFGE and hybridisation with 2D5 only one band was seen in each sample showing that the clones contain unrelated YACs. This experiment also determined the size of the YAC which was derived from the human Y chromosome as probe 2D5 had previously been shown to be present in the YACs (see Table 4.6). For AA15G8/6 this was 250kb and for BB8H10/2 185kb. Therefore an attempt was made to obtain yeast cells containing only one YAC by growing the clones to saturation in YPD medium and then streaking out for single colonies on selective SD agar as outlined in section 2.7.1. DNA was then prepared from sixteen single colonies for each clone and analysed by fractionation through a pulsed-field gel.

The results of the pulsed-field gel fractionation experiment are not shown but several of the AA15G8/6 daughter clones contained only two YACs, each clone having lost the smallest non-Y YAC. All the daughter clones derived from BB8H10/2 still contained two YACs. Therefore one of the AA15G8/6 daughter clones containing two YACs was re-purified as described above to again select for one YAC per clone. Unfortunately, the 250kb Y chromosome YAC was never isolated on its own. However, later experiments (section 5.1.2.f) determined that this 250kb YAC does not contain the right hand vector arm which could explain why the largest, non-Y YAC was always selected for during these procedures. Despite not being able to isolate the 250kb YAC, clone AA15G8/6 was still analysed because it contains DNA from the distal end of the alphoid YAC contig which is only covered by one other YAC (AA4A7/11 and see Figure 5.12.). Clone BB8H10/2 however contains sequences from the proximal end of the alphoid contig which are well covered by other YACs and therefore was not analysed any further. Throughout the rest of this work reference to YAC AA15G8/6 will be referring to the 250kb Y chromosome YAC found in the clone.

5.1.2. Restriction site maps of the alphoid contig YACs.

The YACs from the alphoid contig were mapped by partial digestion with *Bsi*WI, *Bss*HII, *Mlu*I, *Sfi*I and *Sst*II and indirect end-labelling (Burke *et al.*, 1987 and as described in section 2.7.2). Partial digests of the YACs were fractionated by PFGE, transferred to a filter and sequentially probed with the right and left arm vector probes. The loci discovered to be present in each YAC, (see Table 4.6), were placed on the maps by hybridisation to complete digests. The resulting individual maps are shown in Figures 5.2 - 5.11 and are briefly discussed in sections (a) - (g) below.

5.1.2.a. YAC 281C5.

This YAC was difficult to map as it was very unstable. However this was expected as the results from section 4.3.b determined that approximately 50% of the YAC consisted of alphoid DNA. The results of the indirect end-labelling experiments are given in Figure 5.2, alongside the restriction enzyme map. For each partial digest either one or more principal bands were detected above a number of less intense deleted bands. This deletion was more pronounced when the right arm vector probe was used as alphoid DNA in the YAC extends from the middle of the YAC to the right arm. The principal bands were used to compile the restriction map along with the data obtained from the complete digests (Figure 5.3). The overall structure of the region covered by YAC 281C5 was confirmed by the analysis of another overlapping, more stable YAC, 62C12 (see Figure 5.12).

5.1.2.b. YAC 62C12.

Apart from slight instability of alphoid DNA no problems were encountered in analysing YAC 62C12. The resulting data is shown in Figures 5.4 (partial digests and map) and 5.5 (complete digests). However, by comparing the restriction site maps of YACs 62C12 and 281C5 there appears to be a discrepency in the number of *Sst*II sites present, YAC 62C12 clearly having an extra site. Also some differences in the distances was found. Due to the difficulties in mapping YAC 281C5 the data from YAC 62C12, which is clearer and easier to interpret, was taken to be correct.

5.1.2.c. YAC BB8G1/4.

During the process of analysing YAC BB8G1/4 it was discovered that it was co-ligated; the junction between Y and non-Y inserts was located between the Yαl-hybridising sequences derived from the Y chromosome and a *Sfi*I site

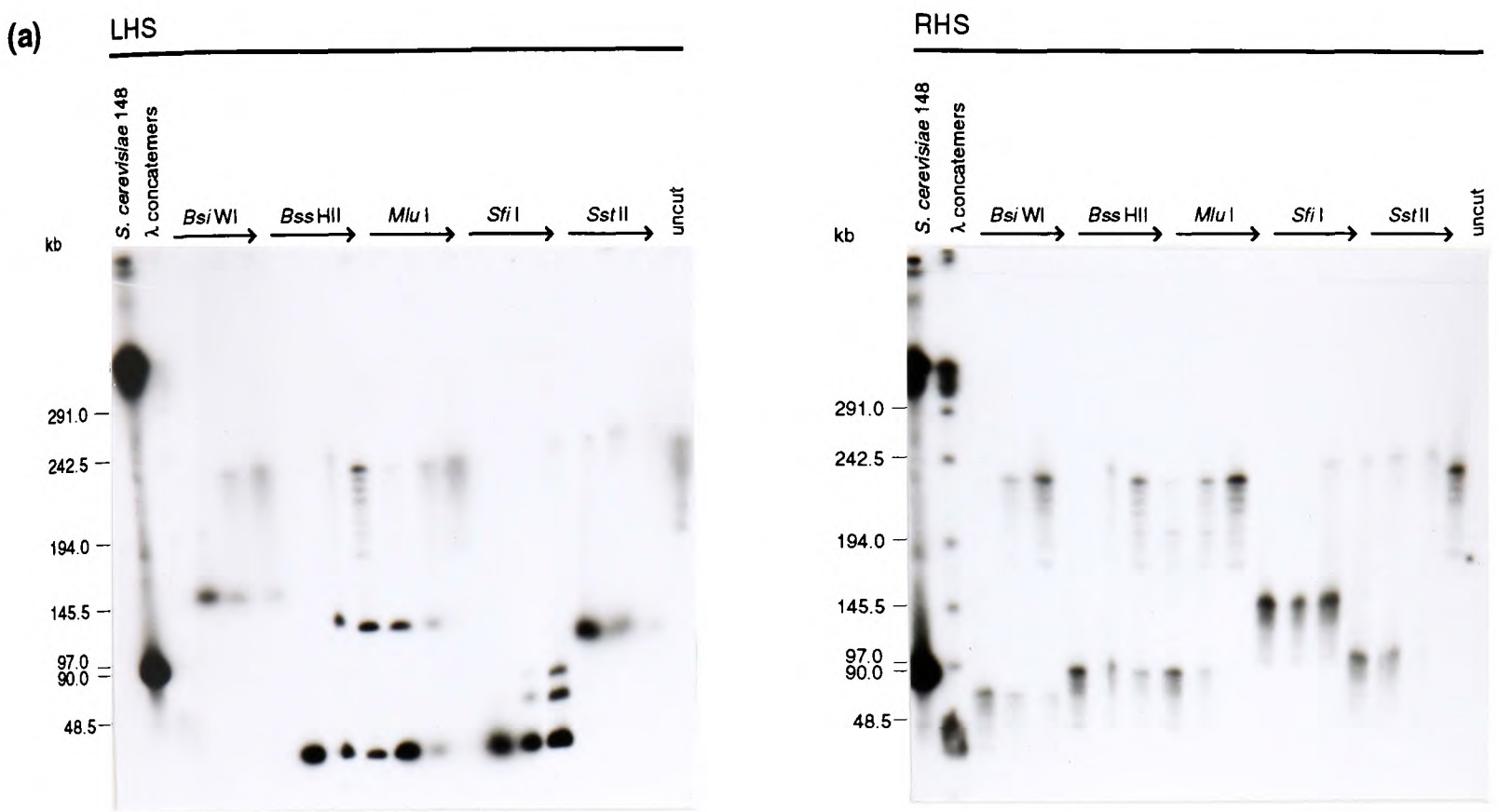


Figure 5.2. Analysis of YAC 281C5. (a) Hybridisation of the left and right hand vector probes to partial digests of the YAC. The DNA was partially digested with the enzymes indicated, fractionated by PFGE (effective pulse time 11 secs, run time 32hrs) and a filter transfer probed with the right and left hand vector probes. The filters were washed in 0.1 x SSC. The position of the λ concatemer size markers are shown on the left hand side of each photograph. (b) Restriction site map of the YAC showing the location of various repeated sequences.

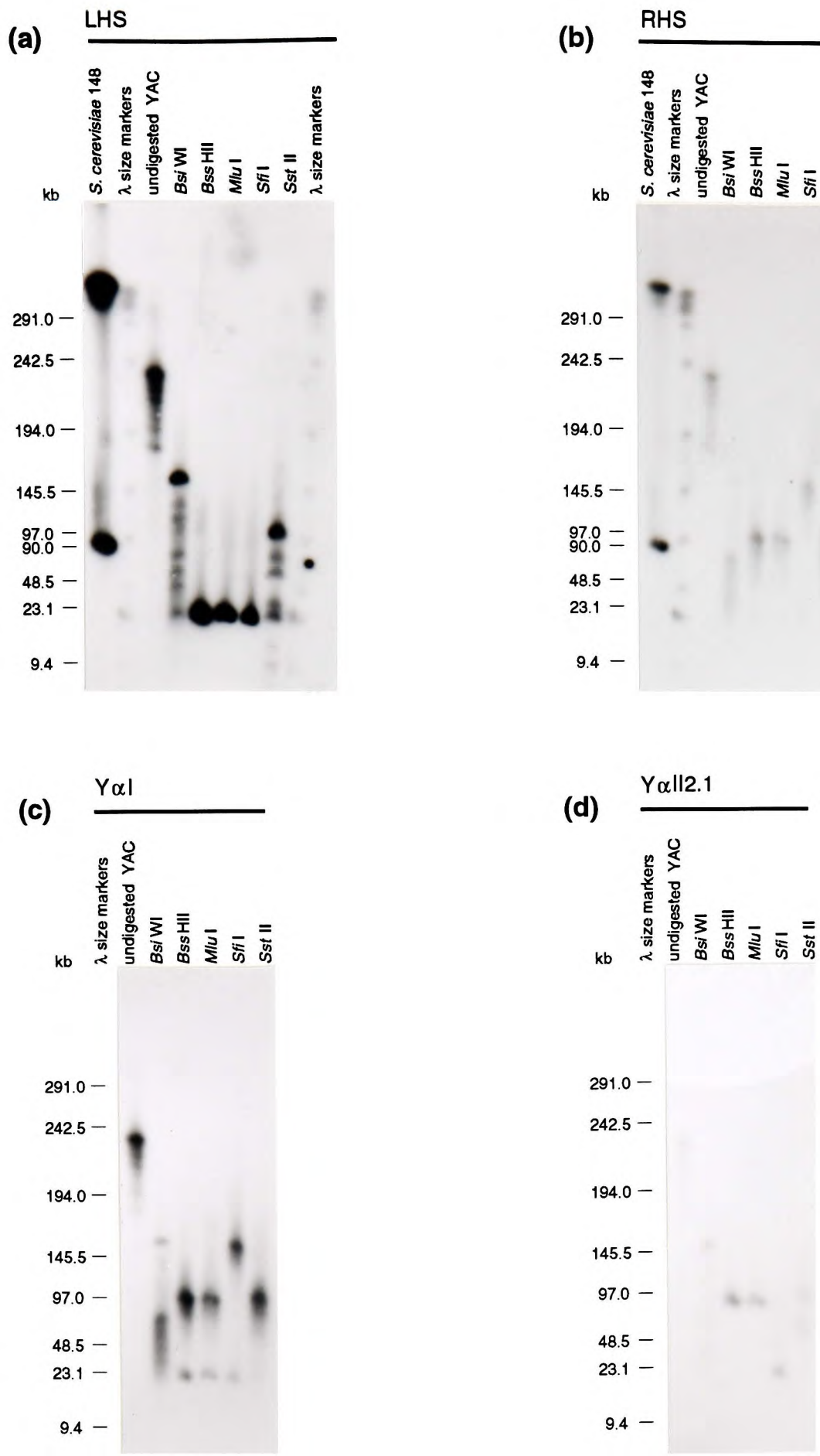
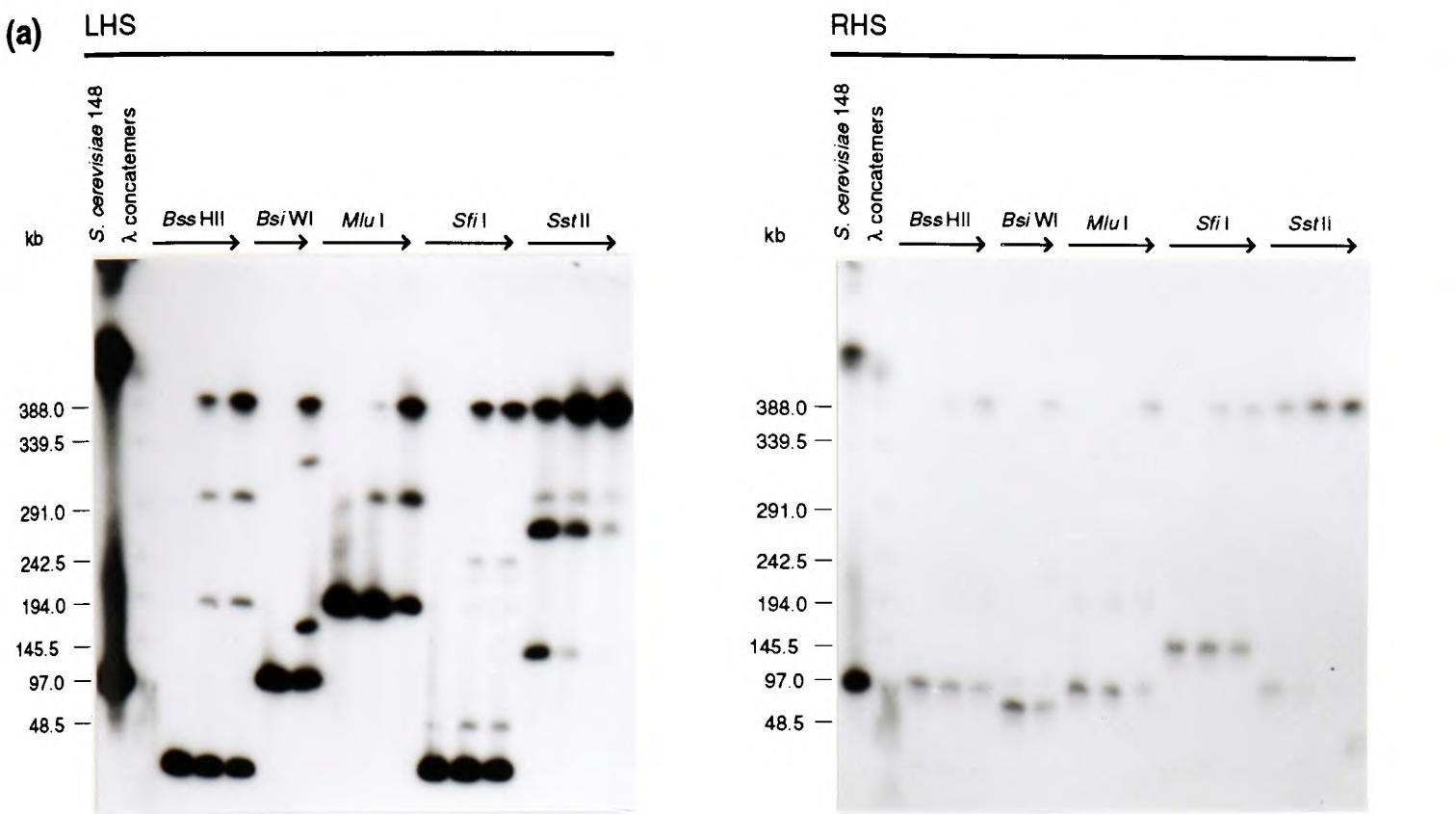


Figure 5.3. Localisation of various probes on YAC 281C5. The YAC DNA was digested to completion with the enzymes indicated, fractionated by PFGE, transferred to a filter and hybridised with probes (a) -(d). The filters were washed in 0.1 x SSC and the position of the λ *Hin*dIII, λ concatemers and *S. cerevisiae* YP148 size markers are shown at the side of each photograph. The pulsed-field gel was run for 32hrs with an effective pulse time of 10 secs.



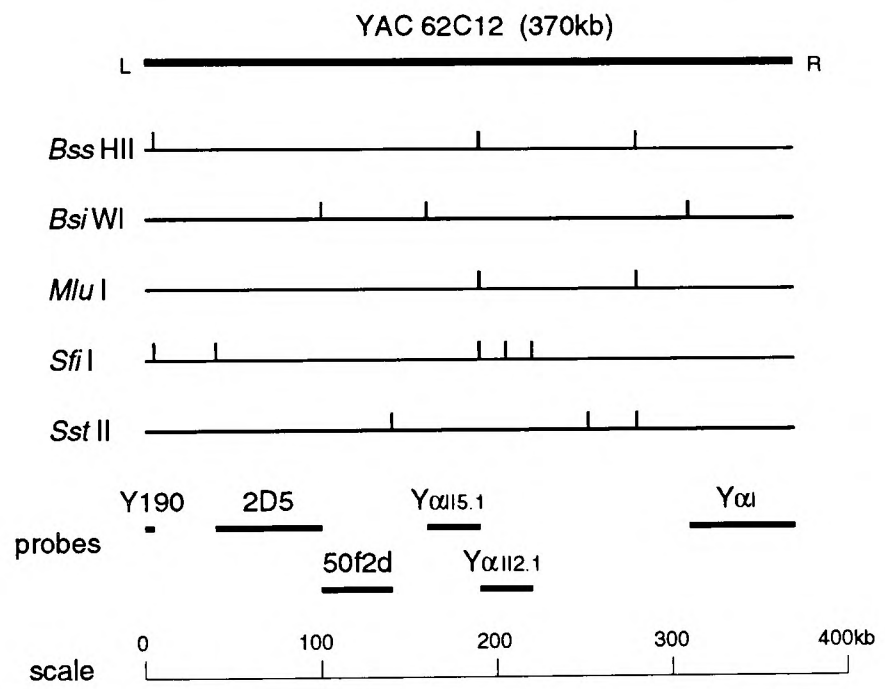


Figure 5.4. Analysis of YAC 62C12. (a) Hybridisation of the left and right hand vector probes to partial digests of the YAC. The DNA was partially digested with the enzymes indicated, fractionated by PFGE (effective pulse time 16 secs, 26hrs run time) and a filter transfer probed with the right and left hand vector probes. The filters were washed in 0.1 x SSC. The position of the λ concatemer size markers are shown on the left hand side of each photograph. The arrows below the enzymes represent decreasing number of units used. (b) Restriction site map of the YAC showing location of various sequences.

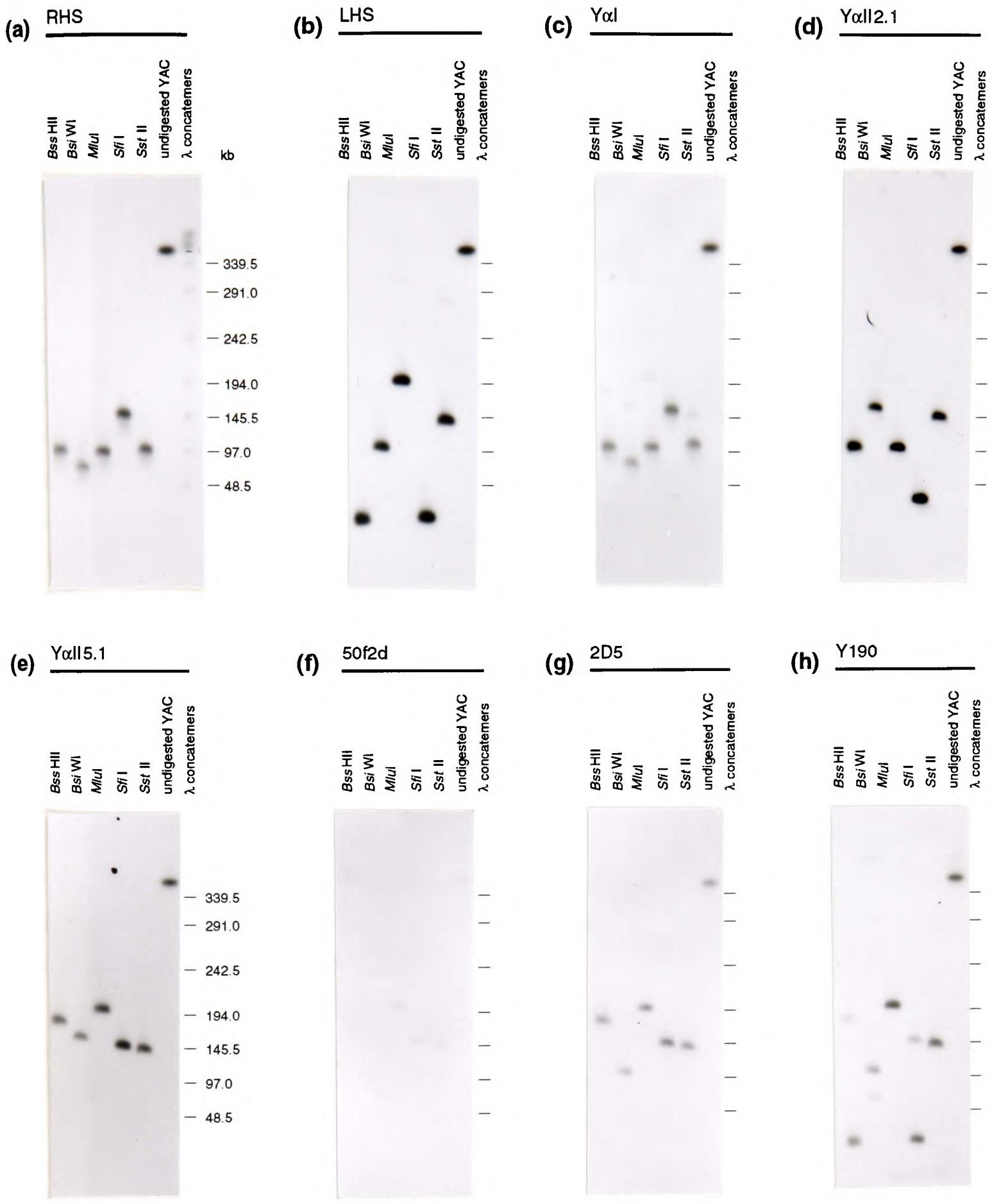


Figure 5.5. Localisation of various probes on YAC 62C12. The YAC DNA was digested to completion with the enzymes indicated, fractionated by PFGE, transferred to a filter and hybridised with probes (a) - (h). The filters were washed at 0.1 x SSC and the position of the λ concatemer size markers is given at the side of each photograph. The pulsed-field gel was run for 32hrs with an effective switch time of 15 secs.

present in the non-Y DNA. These results are shown in Figures 5.6. and 5.7. It was also determined that the Y chromosome insert in this YAC did not represent a true copy of genomic DNA because the YAC contains YαI and YαII 5.1 sequences but not YαII 2.1. Anaylsis of YACs AA4A7/11, 62C12 and 281C5 places YαII 2.1 between YαI and YαII 5.1 (see Figure 5.12) and therefore the arrangement in YAC BB8G1/4 was taken to be abnormal. It was possible that the deletion in BB8G1/4 could have arisen by recombination between the alphoid DNA at YαII and YαI.

5.1.2.d. YAC AA4A7/11.

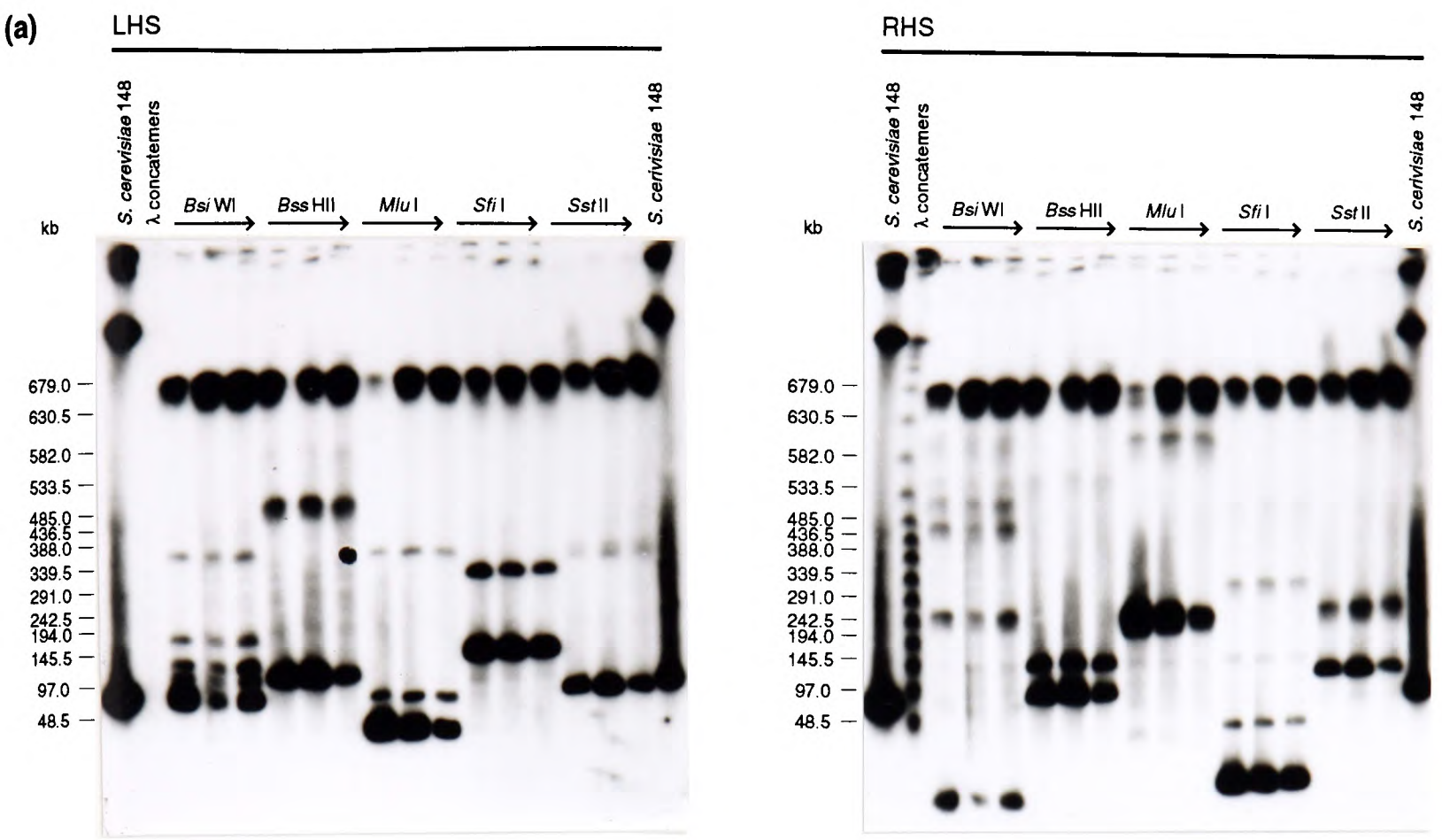
YAC AA4A7/11 does not contain any alphoid DNA and should have been easy to map. However, the results from the end-labelling experiments (not shown) were difficult to interpret because when the right arm vector was used to probe the partial digests of the YAC both the right arm and left arm sequences were detected. Further experiments revealed that the probe was not contaiminated with the left arm sequenecs and that the structure of the right vector arm in YAC AA4A7/11 was abnormal. The restriction site map of this YAC (Figure 5.8.) was therefore determined by analysing the results from the hybridisation of various probes to complete digests (Figure 5.9) and the left arm vector hybridisation to the partial digests (not shown).

5.1.2.e. YAC 13C1F.

This YAC, like 62C12, was not co-ligated and the mapping data was straight forward. However, as this YAC was so small and contained very few restriction enzyme sites the map was obtained by analysing the results of the complete digests only. The results are shown in Figure 5.10.

5.1.2.f. YAC AA15G8/6.

The problems associated with the colony purification of YAC AA15G8/6 were discussed in section 5.1.1.b. It was concluded that two separate YACs co-existed in the same clone and that the Y chromosome YAC could not be isolated from the other non-Y YAC because of its abnormal structure. Therefore the restriction site map of YAC AA15G8/6(Y) could not be determined by partial end-labelling as the non-Y YAC would also be detected. However it was possible to determine the structure of the YAC by using a variation of the partial end-labelling technique and also by hybridising Y chromosome probes to complete digests of the clone. The variation in the partial end-labelling experiment involved determining the order of restriction enzyme sites from



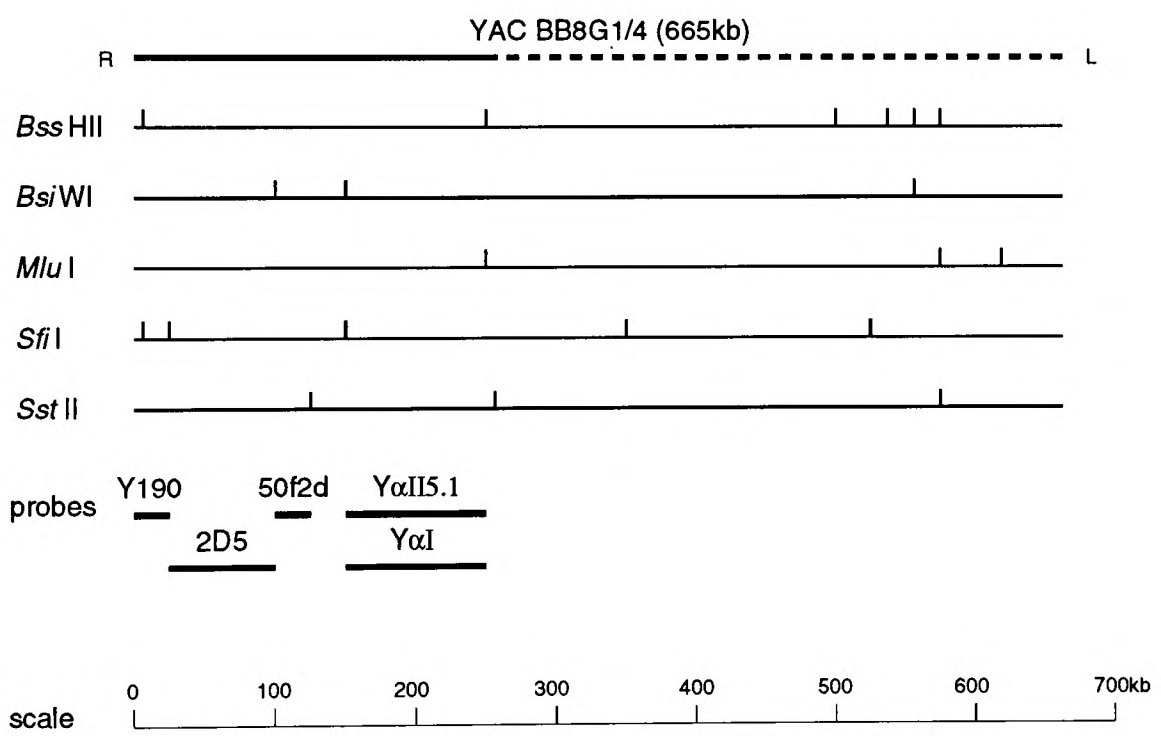


Figure 5.6. Anaylsis of YAC BB8G1/4. (a) Hybridisation of the left and right hand vector probes to partial digests of the YAC. The DNA was partially digested with the enzymes indicated, fractionated by PFGE (effective pulse time 40 secs, run time 26 hrs) and a filter transfer probed with the right and left hand vector probes. The filters were washed in 0.1 x SSC. The position of the λ concatemer size markers are shown on the left hand side of each photograph. (b) Restriction site map of the YAC showing the location of various sequences. The dashed line represents the section of the YAC which is not derived from the Y chromosome.

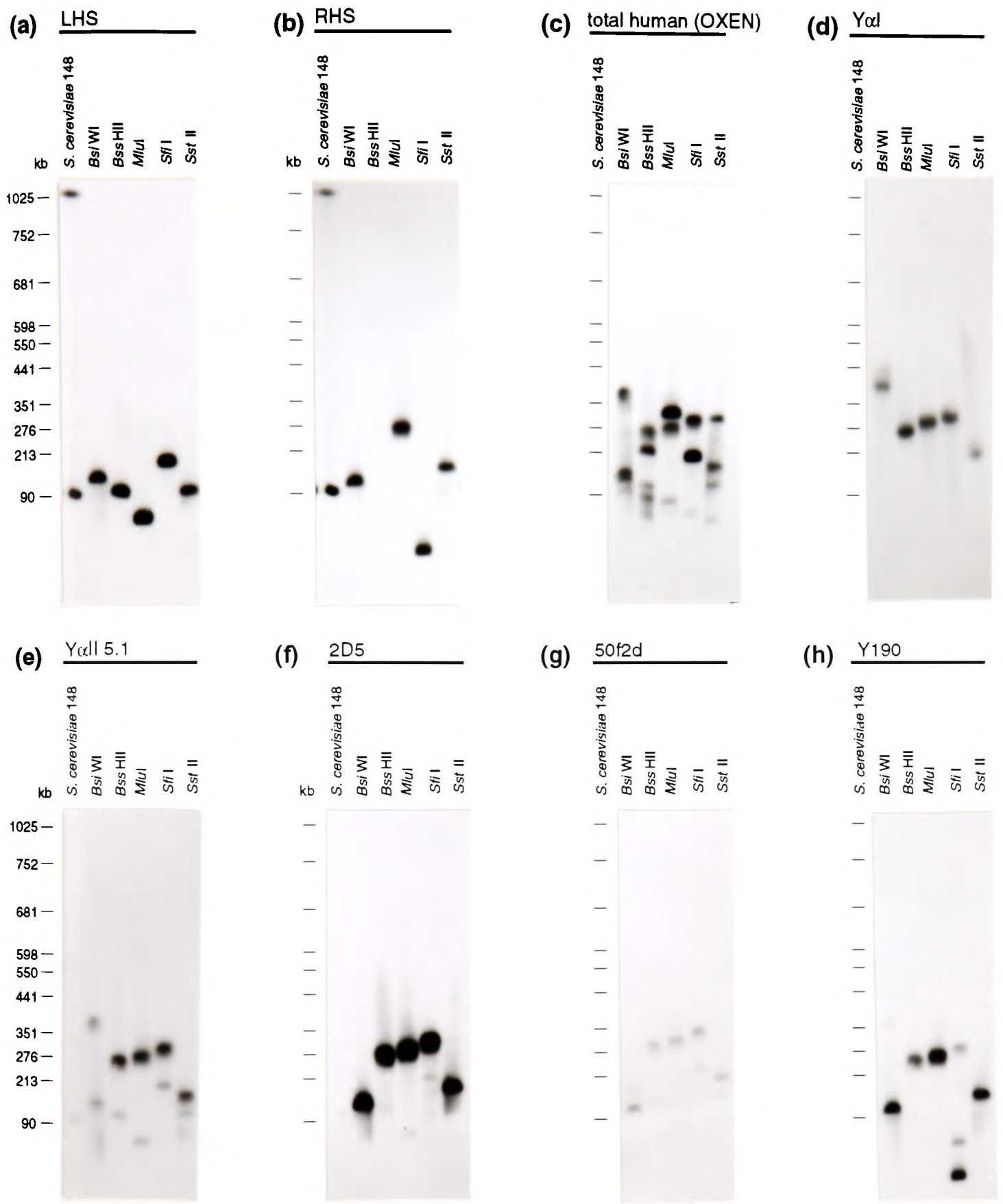


Figure 5.7. Localisation of various probes on YAC BB8G1/4. The YAC DNA was digested to completion with the enzymes indicated, fractionated by PFGE, transferred to a membrane and hybridised with probes (a) - (h). The filters were all washed in 0.1 x SSC except (c) which was washed in 0.5 x SSC. The position of the *S. cerevisiae* 148 size markers is given and the pulsed-field gel was run for 26hrs with an effective pulse time of 40 secs.

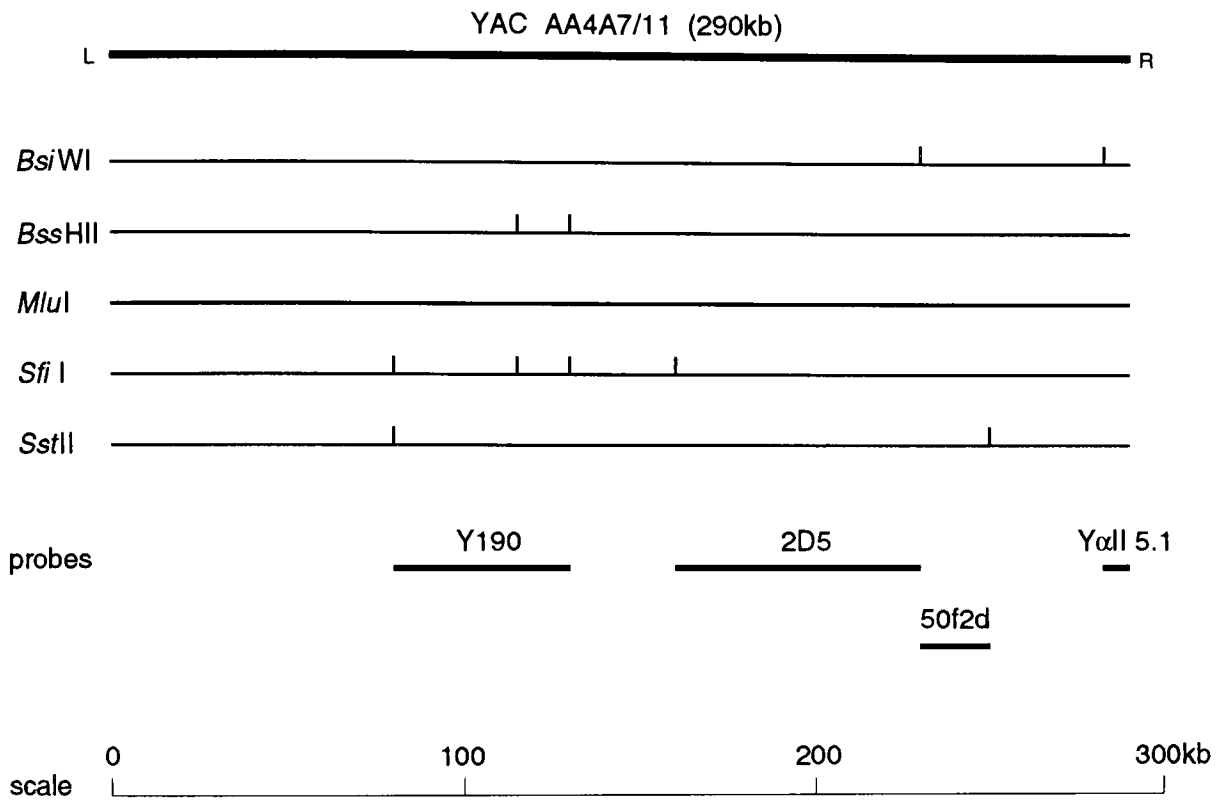


Figure 5.8. Restriction site map of YAC AA4A7/11 showing the location of various probes.

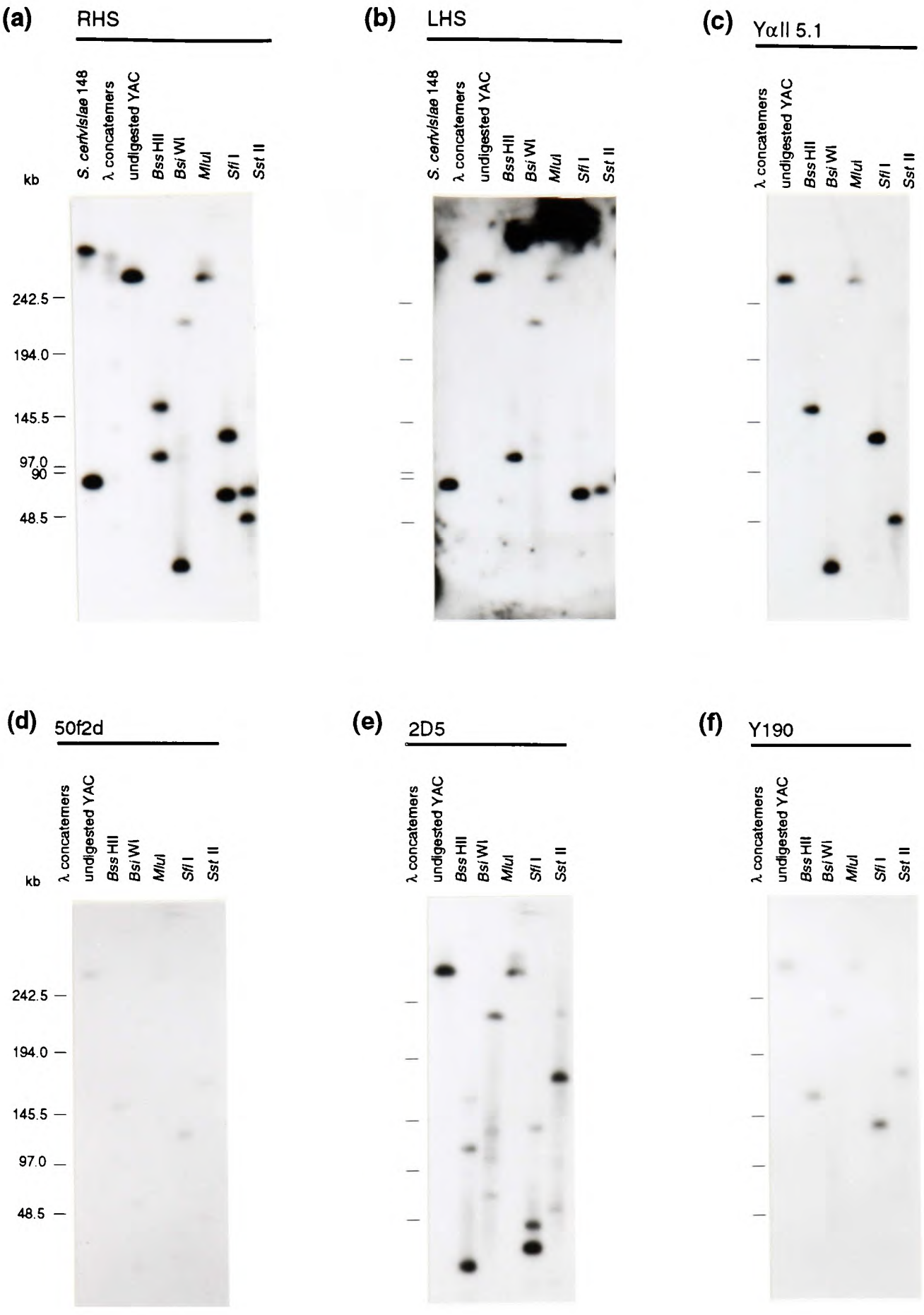
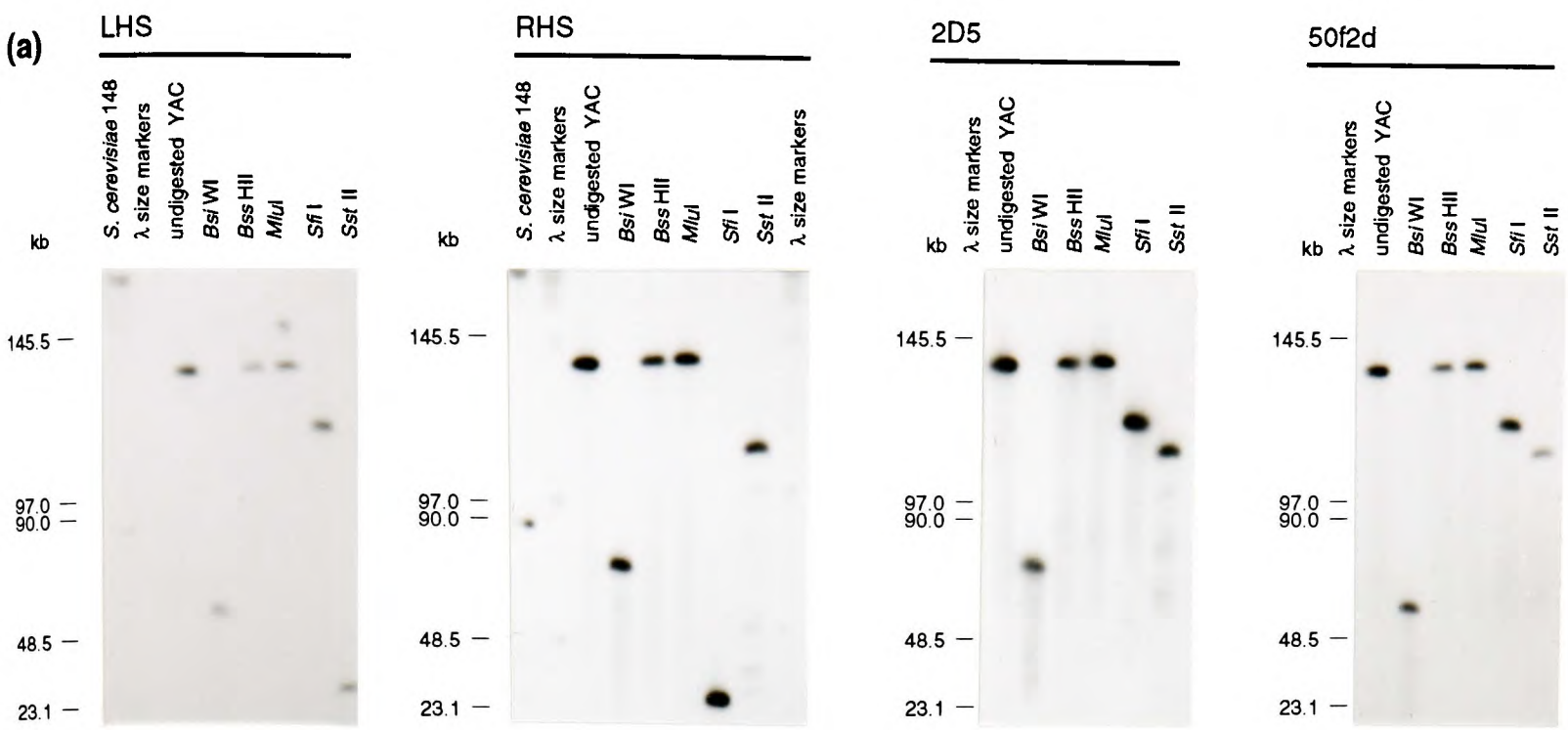


Figure 5.9. Localisation of various probes on YAC AA4A7/11.The YAC DNA was digested to completion with the enzymes indicated, fractionated by PFGE, transferred to a filter and hybridised with probes (a) - (f). The filters were washed at 0.1 x SSC and the position of the λ concatemer size markers is given at the side of each photograph. The pulsed-field gel was run for 32hrs with an effective pulse time of 10 secs. Due to the complex natue of the right hand arm vector in this YAC both the left and right arm restriction fragments are detected by the RHS vector probe (a). However, the RHS probe detects the same sized fragments as Y α ll 5.1.



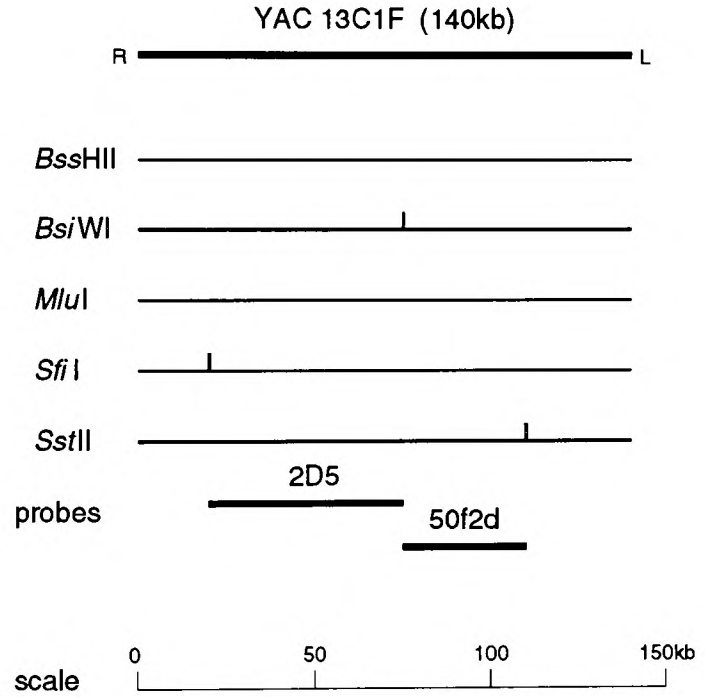


Figure 5.10. Analysis of YAC 13C1F. (a) Hybridisation of probes indicated to complete digests of the YAC. The enzymes used are shown at the top of each track and after digestion the DNA was fractionated by PFGE , transferred to a filter and sequentially hybridised with the probes shown. The filters were washed in 0.1 x SSC and the position of the λ *Hin*dIII, λ concatemer and *S. cerevisiae* size markers is shown. (b) Restriction site map of YAC 13C1F showing the location of probes 2D5 and 50f2d.

the left arm end of the YAC by using the Y specific end-rescued probe, AA15e (section 5.3.1.a), instead of the left arm vector probe. These results are not shown but the mapping data obtained is given in Figure 5.11. Here it can be seen that the left arm probe (a) detects two bands in every track as would be expected if there were two YACs in the same clone. However, the right arm probe (b) only detects one band in every track suggesting that one of the YACs had deleted the right vector arm. It was decided that the right vector arm of YAC AA15G8/6(Y) was deleted because the size of the undigested YAC detected by the right arm vector probe is larger than the size of the undigested YAC detected both by Y190 (c) and 2D5 (d). The discovery that YAC AA15G8/6(Y) did not contain right hand arm vector sequences explains why the non-Y YAC was never lost during the colony purification because only clones containing *URA3* marker, which is carried on the right vector arm, will grow in the selective SD medium. However this deletion event does not explain why YAC AA15G8/6(Y) was never lost during growth. In addition, AA15G8/6 would be expected on the basis of its size to contain the 50f2d locu. It does not. This indicates that the right-hand ~50kb of the YAC does not correspond to the genomic structure and probably represents a co-cloning event.

5.1.2.g. YACs 368A6 and 417D3.

A restriction map was not obtained for YAC 368A6 as the results of the experiments done in section 4.3.1.b predicted that this YAC was derived from "middle of block" Y chromosome alphoid DNA. Therefore determining the restriction sites of *Bsi*WI, *Bss*HII, *Mlu*I, *Sfi*I and *Sst*II in the YAC would not have been informative as these enzymes do not cut within the alphoid block (Tyler-Smith, 1987).

YAC 417D3 is derived from the long arm edge of the alphoid block of DNA and it was not known whether the five restriction enzymes chosen to map this region would be informative in determining the structure of this YAC. Therefore 417D3 was partially digested with these enzymes as described in section 2.7.2, fractionated by pulsed-field gel electrophoresis and a filter transfer hybridised with the left arm vector probe. The results are not shown but it was determined that YAC 417D3 did not contain the recognition sites of these enzymes. The structure of the YAC was determined by digesting the YAC to completion with a range of frequent cutter enzymes which are known

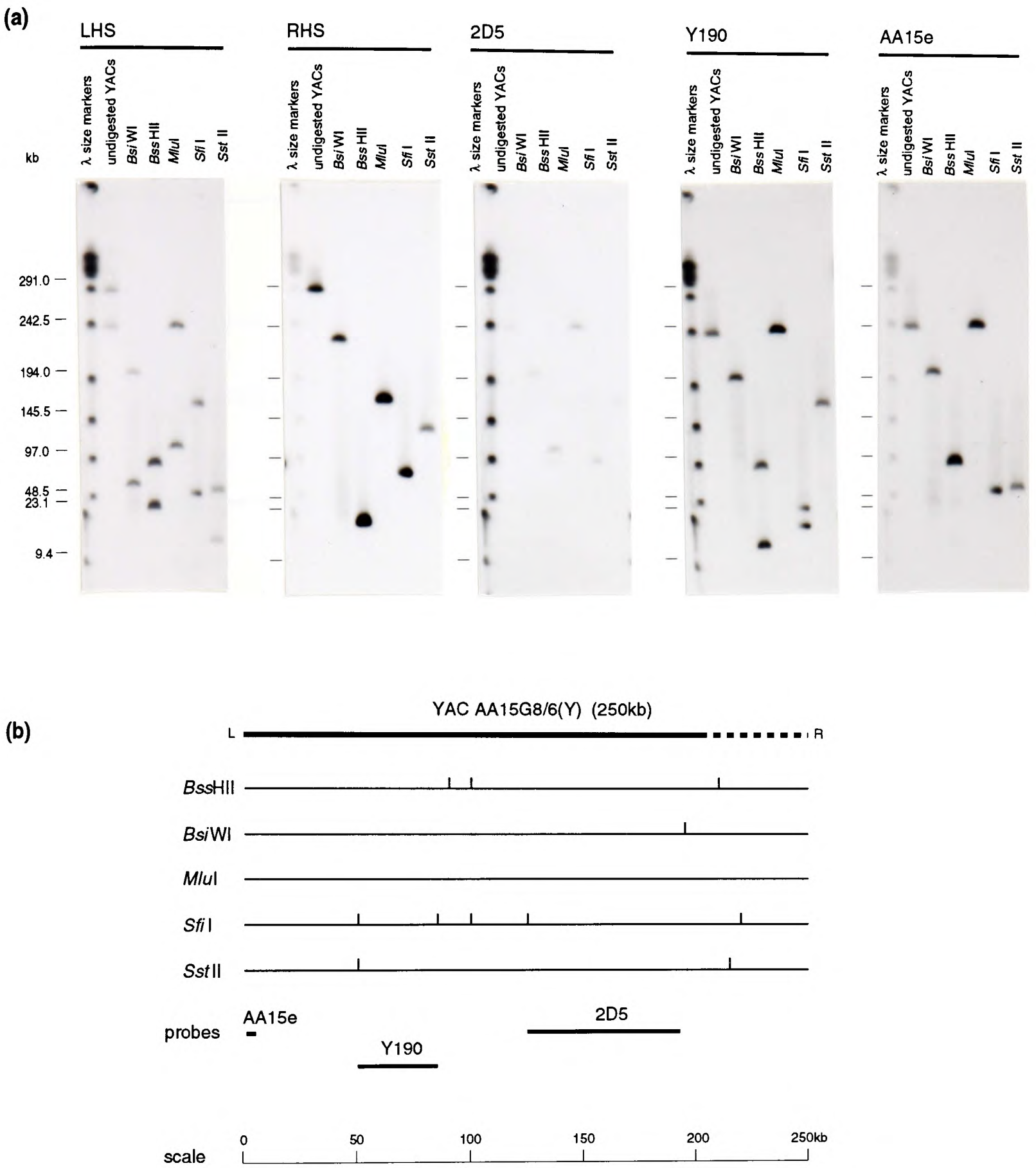


Figure 5.11. Analysis YAC AA15G8/6. (a) Localisation of various probes. The YAC DNA was digested to completion with the enzymes indicated, fractionated by PFGE , transferred to a filter and hybridised with the probes shown above each photograph.The pulsed-field gel was run for 32hrs with an effective pulse time of 16 secs. The position of the λ *Hind*III and λ concatemer size markers are shown. (b). Restriction site map of YAC AA15G8/6. The stippled represents the section of the YAC which is not derived from this region of the Y chromosome.

Table 5.1. Summary of the difficulties encountered in analysing the YACs from the alphoid contig.

YAC	Difficulties encountered
281C5	Very unstable
62C12	Slightly unstable
BB8G1/4	Co-ligated and deleted
13C1F	No problems
AA15G8/6(Y)	Cannot be colony purified (Two different YACs in the same clone)
AA4A7/11	Abnormal structure of right vector arm

to recognise sites outside the alphoid block (Tyler-Smith, 1987) and these results are discussed in section 5.1.3.

5.1.2.h. Restriction site map of the alphoid contig.

It was not anticipated that determining the restriction site map of these YACs would be so problematic and Table 5.1 summarises the difficulties occurred. Despite these problems, individual maps of the YACs were obtained. As expected the YACs all overlapped and the restriction site map of the resulting contig is given in Figure 5.12. Excluding YACs 417D3 and 368A6, approximately 500kb of genomic DNA is covered by the YACs spanning from centromeric alphoid DNA to the Y190 minor locus.

5.1.3. Further analysis of YACs containing alphoid DNA.

The YACs which contained alphoid DNA were derived from the short arm edge (281C5, 62C12 and BB8G1/4), middle of block (368A6) and long arm edge (417D3) of Y chromosome alphoid DNA. They were therefore analysed further to determine:- (1) the organisation and degree of divergence of alphoid DNA at the edges of the block; (2) the orientation of the 170bp repeating subunit within the array; (3) the type of sequences present immediately adjacent to alphoid DNA at the edges of the block. The results

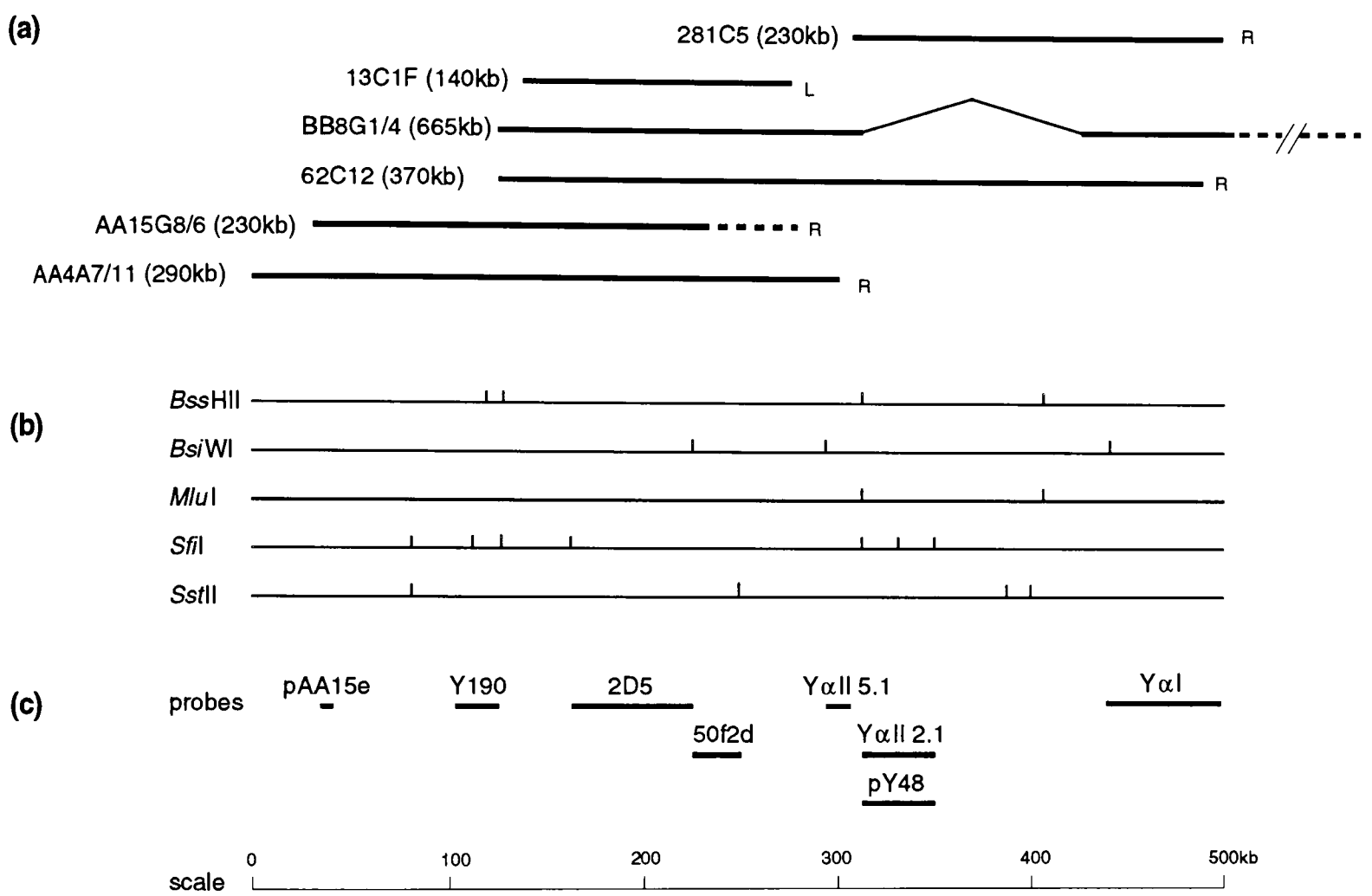


Figure 5.12. Restriction site map of the YACs which form the alphoid YAC contig. (a) YACs; (b) restriction site maps; (c) location of probes. L and R represent the left (centromeric) and right (non-centromeric) arms of the YAC vector. The arrow and dotted line drawn in YAC BB8G1/4 represent deletion and co-ligation events respectively.

are presented in sections (a)-(e) below. However, two of the alphoid YACs, BB8G1/4 and 368A6, were not included in this analysis for the following reasons. The former YAC is co-ligated and deleted and as a result it was impossible to determine whether its alphoid DNA was derived from the middle or an edge of the major array. The latter YAC, which contains solely "middle of block" alphoid DNA, would be similar in structure to many other YACs derived from this region in other individuals (Neil *et al.*, 1990; Fisher, 1992)

5.1.3.a. Organisation of alphoid DNA in YACs 281C5, 62C12 and 417D3.

The alphoid satellite DNA locus consists of tandemly repeated ~170bp subunits organised into 5.7 or 6.0kb units which are themselves tandemly repeated in a single array several hundreds kilobases long (Tyler-Smith and Brown, 1987; Tyler-Smith, 1987). These authors identified 30 restriction enzymes that do not cut within the alphoid array and determined that the location of their recognition sites divide the alphoid region into three: the alphoid block and its left and right (short and long arm) flanking regions. The YACs containing alphoid DNA were therefore analysed with 16 of these frequently cutting enzymes to confirm that they were derived from the short arm (281C5, 62C12) and long arm (417D3) edges of the major alphoid block. To achieve this the YACs were digested and fractionated by pulsed-field gel electrophoresis, transferred to a filter and hybridised with Yαl and the results are shown in Figures 5.13 (YACs 281C5, 62C12,) and Figure 5.14 (YAC 417D3). The location of the 5bp and *Alu* sequences previously shown to be present in YAC 417D3 were also determined by hybridising the 417D3 filter with these probes.

In order to interpret these results (discussed below) the restriction site map of the Y chromosome alphoid array from the normal individual (Boleth) the YACs were derived from was consulted. This map was constructed by C. Tyler-Smith (personal communication) and is shown in Figure 5.15. In a typical male the recognition sites for many of the enzymes on the short arm of the chromosome are located in a cluster, close to the the alphoid-nonalphoid boundary. In addition, sites for *Sst*I and *Hpa*II are located ~45 and 60kb away respectively. The long arm flanking sequences though are typically characterised by a smaller cluster of these sites but noticably the recognition sites for the frequent cuting enzymes *Kpn*I and *Apa*I are located ~240 and

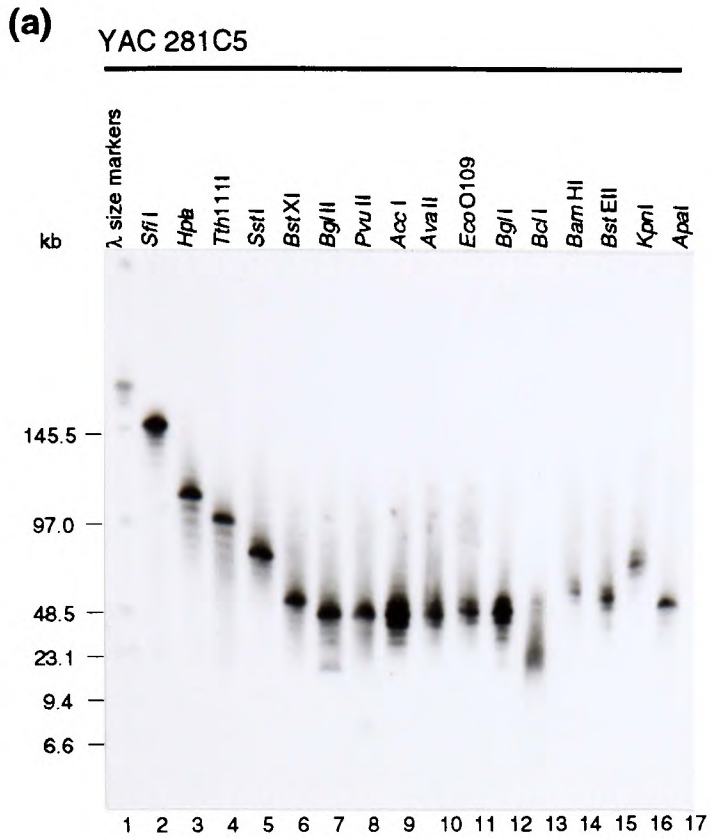


Figure 5.13. Organisation of alphoid DNA in YACs 281C5 (a) and 62C12 (b). YAC DNA was. digested to completion with the enzymes indicated and analysed by PFGE, filter transfer and hybridisation to Y α I. The filters were washed in 0.1 x SSC and the position of the λ *Hin*dIII and λ concatemer size markers is shown. Both pulsed-field gels were run for 50hrs with an effective pulse time of 5 secs.

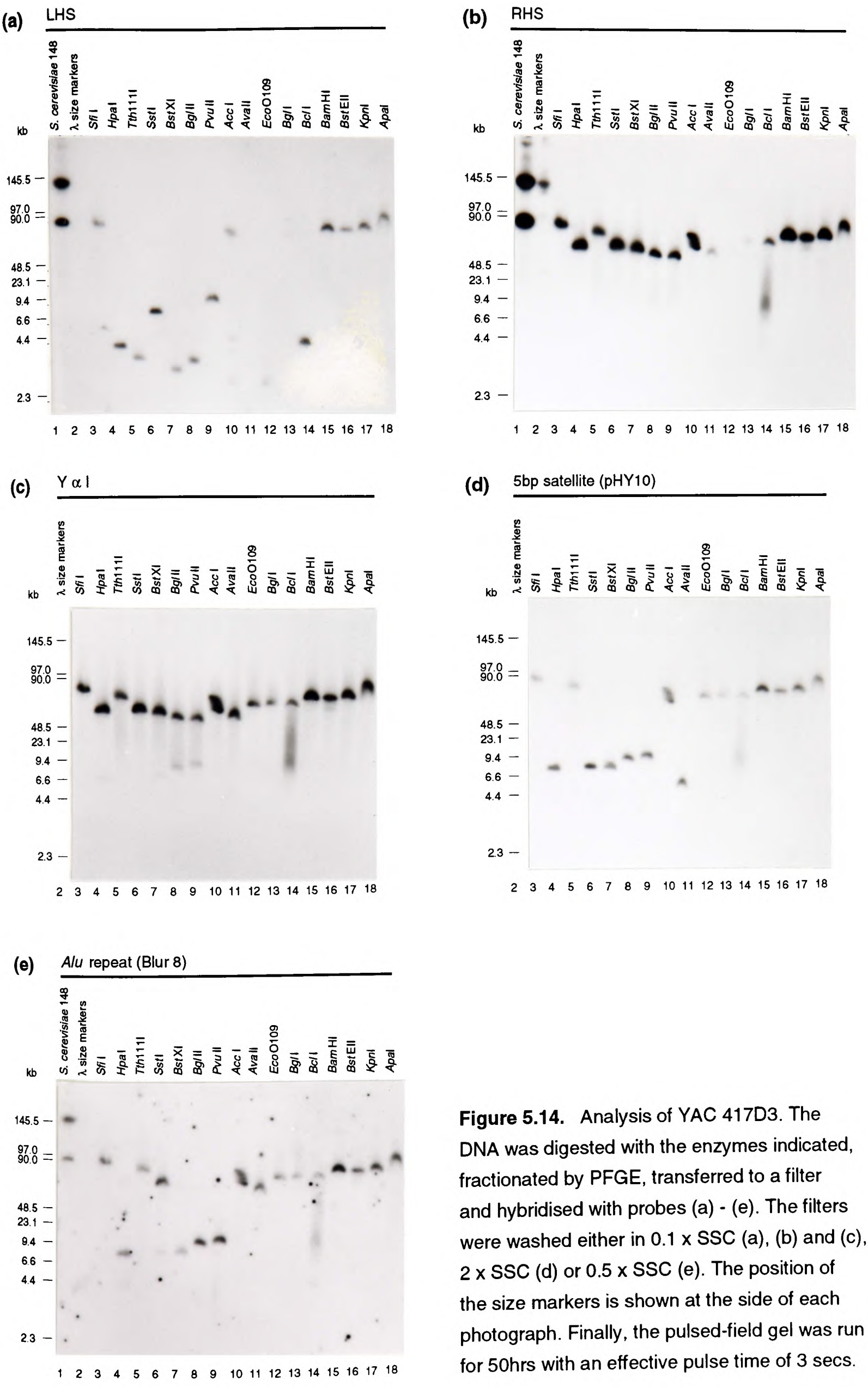


Figure 5.14. Analysis of YAC 417D3. The DNA was digested with the enzymes indicated, fractionated by PFGE, transferred to a filter and hybridised with probes (a) - (e). The filters were washed either in 0.1 x SSC (a), (b) and (c), 2 x SSC (d) or 0.5 x SSC (e). The position of the size markers is shown at the side of each photograph. Finally, the pulsed-field gel was run for 50hrs with an effective pulse time of 3 secs.

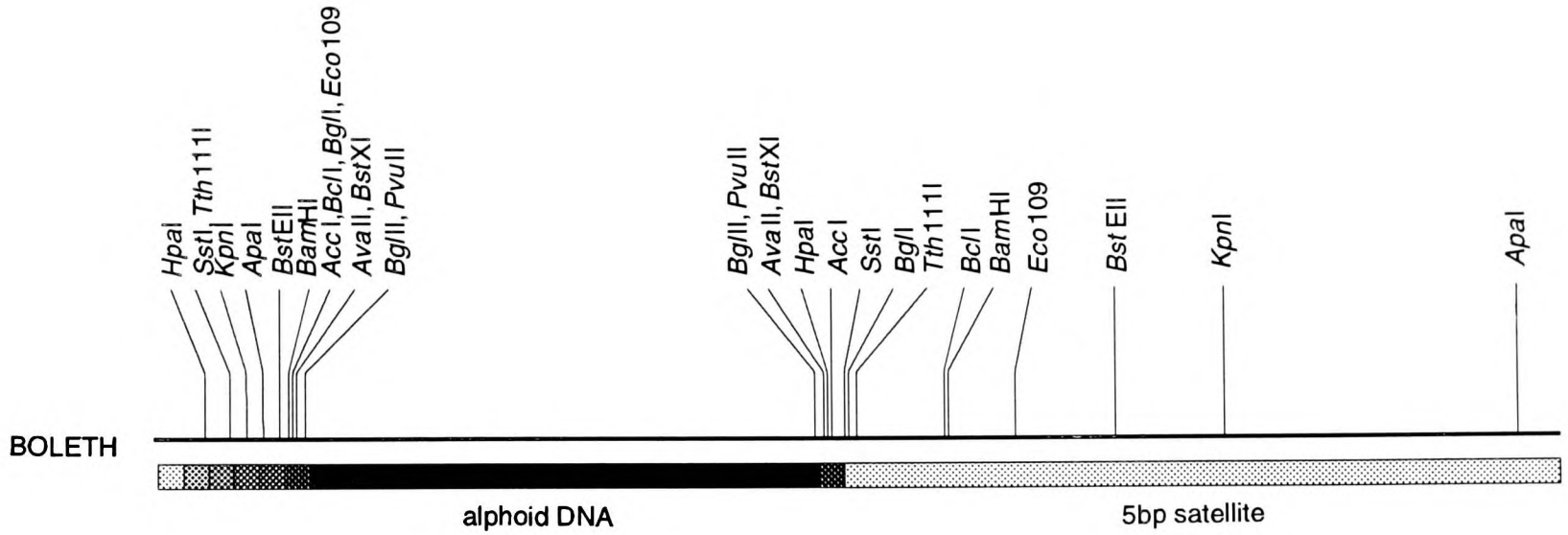
~405kb away from alphoid DNA. These landmarks were therefore used to confirm which edge of the alphoid block the YACs were derived from.

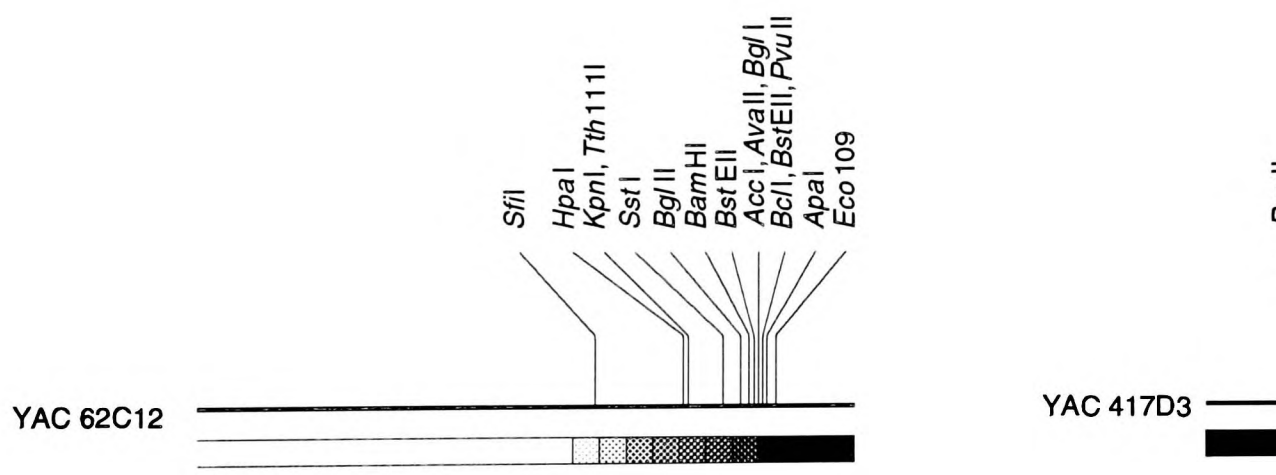
i) The short arm alphoid YACs.

The results shown in Figure 5.13. determine the location of the 16 alphoid restriction enzymes in YACs 281C5 and 62C12 and confirm that these YACs are derived form the short arm edge of the major array of alphoid DNA. This conclusion was made because for both YACs the alphoid DNA is located on larger *Sfi*, *Tth*111I, *Hpa*I and *Sst*I fragments (tracks 3, 4, 5 and 6) than the *Pvu*II fragment (track 8) but on similar sized *Kpn*I and *Apa*I fragments (tracks 16 and 17). This indicates that these YACs contain the recognition sites characteristic of the short arm alphoid non-alphoid boundary but not of the long arm boundary. The location of these 16 restriction enzymes used to characterise alphoid DNA in the YACs is shown in Figure 5.15. These results show that there appear to be slight differences between the location of restriction sites *Hpa*I and *Tth*111I, *Bst*XI and *Bgl*I. The most likely explanation is that in YAC 62C12 the *Hpa*I and *Tth*111I, *Bst*XI and *Bgl*I plugs have been loaded in reverse order.

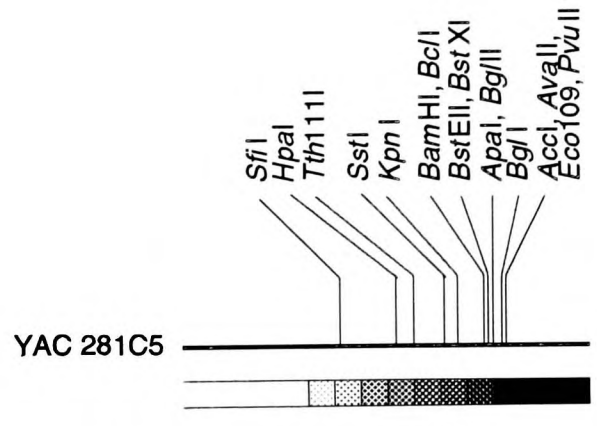
ii) The long arm alphoid YAC.

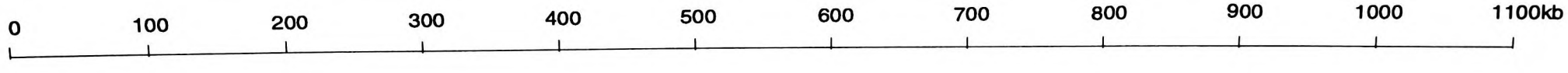
It was determined in section 5.1.2.g that the rare cutter restriction enzymes chosen to map the centromeric region of the human Y chromosome did not cut in this YAC. Therefore it was analysed as discussed above to define its structure. These results and the restriction site map of the YAC are shown in Figures 5.14 and 5.15 respectively. Most of the restriction enzymes do not cleave within the YAC insert. However, the insert contains a single cluster of sites for enzymes which also are present at the alphoid-nonalphoid long arm boundary in Boleth (*Pvu*II, *Bgl*I, *Bst*XI, *Hpa*I, *Ava*II and *Sst*I). The cluster appears to contain sites for just within alphoid DNA as they cleave alphoid DNA to ~80 fragments and the 5bp satellite to <10kb fragments. Four of the enzymes (*Pvu*II, *Bgl*I, *Bst*XI and *Hpa*I) also cleave alphoid DNA into small <10kb fragments thus placing the sites proximal to *Ava*II and *Sst*I. The YAC also contains an *Alu* repeat (see Figure 5.14.e). All the enzymes in the cluster except SstI sites (Figure 5.14.c tracks 4, 7, 8, and 9) all must lie proximal to the *Alu* sequence because in these digests *Alu* sequences hybridise to the same sized fragment as the 5bp satellite. The *Sst*I site must lie distal to the *Alu* element because in this digests the *Alu* sequence hybridises to the same sized fragment as alphoid DNA (track 6). The resulting restriction site map of YAC


BOLETH
HpaI
SstI, Tth111I
KpnI
ApaI
BstEII
BamHI
Acc I,BclI, BglII, Eco109
AvaII, BstXI
BglII, PvuII
BglII, PvuII
AvaII, BstXI
HpaI
Acc I
SstI
BglI
Tth111I
BclI
BamHI
Eco109
Bst EII
KpnI
ApaI
alphoid DNA
5bp satellite


SfiI
HpaI
KpnI, Tth111I
Sst I
BglII
BamHI
Bst EII
Acc I,AvaII, BglI
BclI, BstEII, PvuII
ApaI
Eco109
YAC 62C12

PvuII
BglII
Bst XI, HpaI
AvaII
SstI
YAC 417D3


SfiI
HpaI
Tth111I
SstI
KpnI
BamHI, BclI
BstEII, Bst XI
ApaI, BglII
BglI
AccI, AvaII,
Eco109, PvuII
YAC 281C5


0
100
200
300
400
500
600
700
800
900
1000
1100kb

417D3 is shown in Figure 5.15 and the location of the different repeats in the YAC represented schematically in Figure 5.16.

5.1.3.b. Alu-alphoid-5bp satellite PCR on YAC 417D3.

The evidence presented in the previous section showed that YAC 417D3 contains Y chromosome alphoid DNA that is derived from the long arm edge of alphoid DNA. The location of the alphoid restricition enzyme sites was in agreement with the location of the sites at the genomic long arm alphoid-nonalphoid boundary. This strongly suggests that YAC 417D3 truely does represent the structure of human centromeric DNA from this region. However, because only a single YAC from this region was available, it was desirable to confirm the structure and the following PCR experiments were performed.

PCR primers for the 5bp satellite (5A and 5B), *Alu* repeat (LIH517 and LIH559) and alphoid repeat (YαI.A and YαI.B) were used in all possible combinations (in both orientations) to amplify YAC 417D3 DNA. The PCR conditions were given in section 2.7.2 and Table 2.4. These preliminary experiments determined that primers YαI.A with 5A, YαI.B with 5A, LIH517 with 5A and LIH559 with 5B all amplified a range of products in YAC 417D3. To determine whether any of these products were Y chromosome specific, male, female and two Y-hybrid cell lines (3E7 and 853) were amplified using these primers. The results are not shown because the patterns were too complex to interpret. Unfortunately lack of time prevented further work being carried out on this YAC.

5.1.3.c. Location of the short arm block of the 48bp satellite.

On the human Y chromosome there are two small blocks of the 48bp satellite (this thesis). The results described in chapter 3 showed that one of the blocks is located on the long arm ~600kb distal to alphoid DNA. Also these results determined that the other block must be located on the short arm as the satellite hybridises to a ~19kb *Eco*RI fragment in the Q9988.8 derivitive which only contains the short arm of the chromosome. Therefore the satellite was hybridised to all the short arm YACs available and surprisingly the 19kb fragment was found to be present in YACs 281C5 and 62C12. These results are shown in Figure 5.17.b. The satellite was located on the YACs and maps close to YαII2.1 (Figure 5.12), ~100kb distal to alphoid DNA. However, the YαII

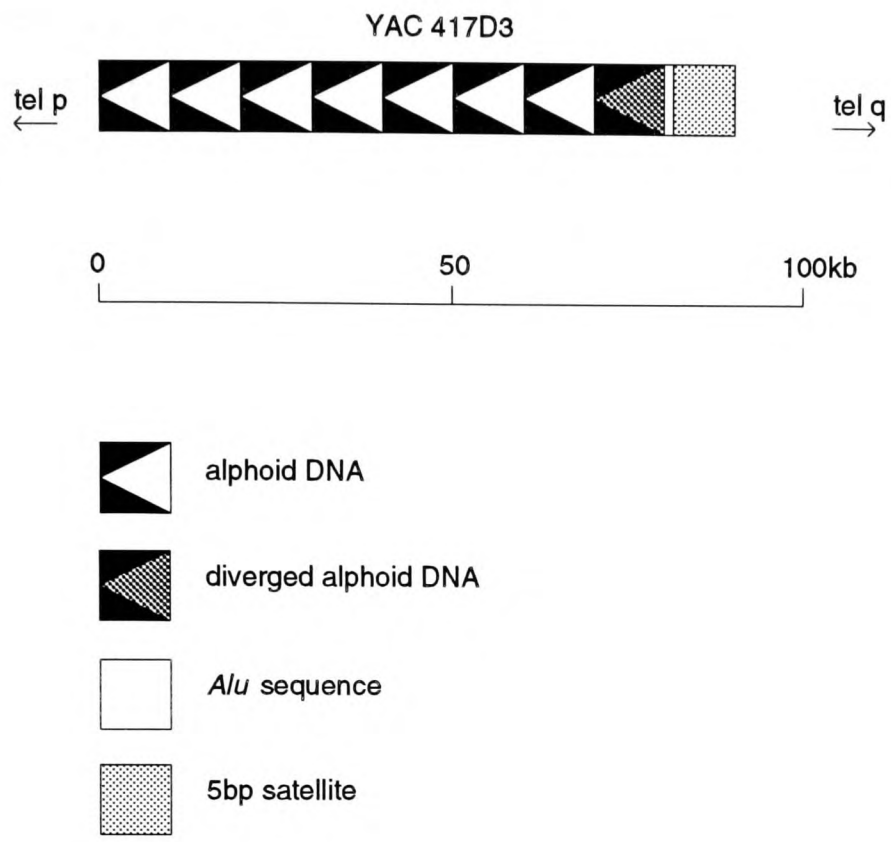


Figure 5.16. Schematic representation of the structure of YAC 417D3. The large white and stippled arrows represent the orientation of alphoid DNA within the array (discussed in section 5.1.3.e). The size of the Alu sequence is drawn larger than reality for clarity.

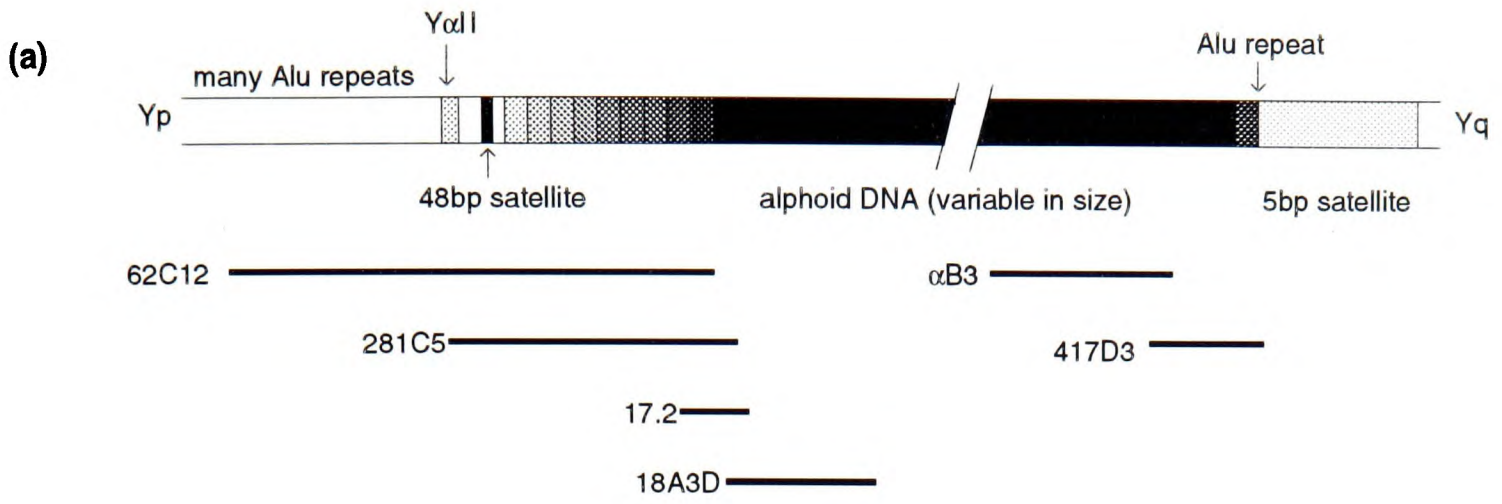
cosmids analysed by C. Tyler-Smith do not contain a 19kb *Eco*RI fragment and therefore the 48bp satellite must lie closer to alphoid DNA than YαII 2.1.

5.1.3.d. Divergence of alphoid DNA in YACs.

The results from the section above and also from previous work (Tyler-Smith and Brown, 1987) on the long range structure of alphoid DNA predict that the long arm edge of alphoid DNA is flanked by the 5bp satellite. However, the types of sequences present adjacent to the short arm edge of the alphoid-nonalphoid boundary were unknown though Tyler-Smith and Brown (1987) had predicted that this region consisted of a mixture of unique and interspersed repeated sequences. The analysis of these YAC from both the long and short arm edges of alphoid DNA therefore not only allowed the type of sequences flanking alphoid DNA to be determined but also defined the degree of divergence of alphoid DNA at the edges of the block.

DNA from YACs 281C5, 62C12 and 417D3 (CEPH library constructed from Boleth DNA) and 4 YACs derived from the Y chromosome alphoid block of OXEN DNA were analysed. Two of these latter 4 YACs (18A3D and 1B9F) were isolated from the Biochemistry, Oxford YAC library and localised on the Y chromosome by R. Fisher (Fisher, 1992). The other two, 17.2 and αB3 YACs were constructed by A. Villasante andD. Vetrie respectively whilst at Oxford and were described by Neil *et al.* (1990). The position of the YACs relative to the centromeric block of alphoid DNA is drawn schematically in Figure 5.17.a. To determine the degree of divergence of alphoid DNA in these YAC they were all digested with *Eco*RI, fractionated by conventional agarose gel electrophoresis, transferred to a membrane and sequentially hybridised with the *Alu* repeat (Blur 8), the 48bp satellite (p22hom48.4), 5bp satellite (pHY10) and YαI. These results are shown in Figure 5.17.b. and discussed below.

YAC 1B9F (track 6) contains mainly homogenous 5.7kb units of alphoid DNA and its position on the alphoid array is unknown. It is included as an example of typical "middle of block" alphoid DNA. The other 3 OXEN YACs contain restriction enzyme sites which are located in known positions near the edges of the alphoid array in OXEN DNA (the cell line the YACs were derived from). This allowed YACs 18A3D and 17.2 to be placed near the short arm edge and YAC αB3 of the alphoid array. The degree of divergence of alphoid DNA at either edge of the major block was then determined by analying these YACs and the YACs from the CEPH library as described above.



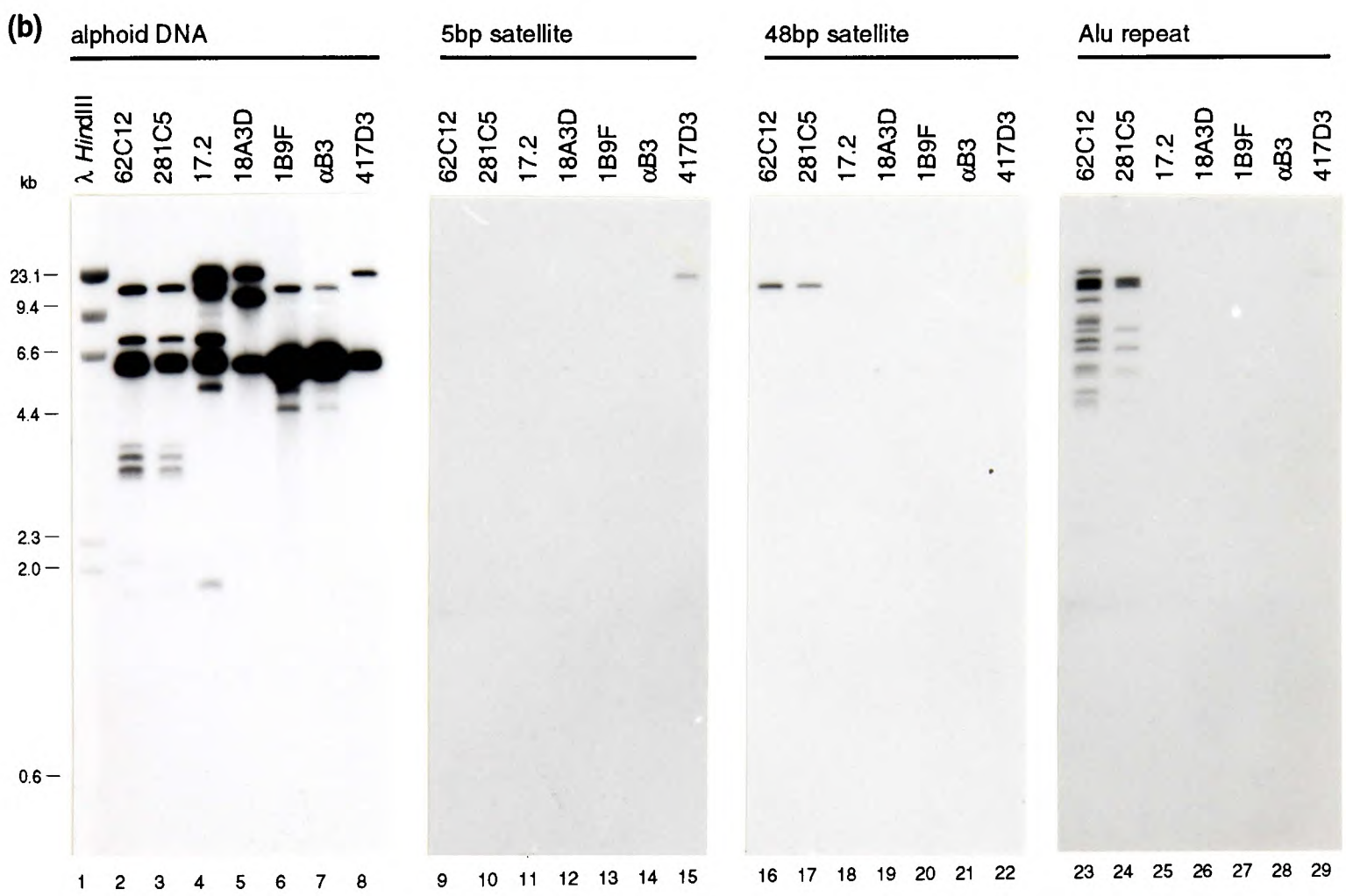


Figure 5.17. Structure of the edges of the alphoid array. (a) Location of the YACs containing alphoid DNA on the major array (black box). YAC 1B9F cannot be positioned on the array as it contains only homogenous 5.7kb units of alphoid DNA. The decreasing intensities of the stippled boxes represents increasing degrees of divergence of alphoid DNA. (b) Autoradiographs of the YACs, digested with *Eco* RI, fractionated through a 1% conventional agarose gel, transferred to a membrane and hybridised with alphoid DNA, the 5bp satellite (pHY10), the 48bp satellite (p22hom48.4), and the Alu repeat (Blur 8). The position of the λ *Hin*dIII size markers is shown.

The hybridisation of YαI to these YACs showed that YACs 1B9F and αB3 (track 6 and 7) contain mainly the 5.7kb units: αB3 also contains a few 6.0kb units which are seen more clearly in a *Hind*III digest (Neil *et al.*, 1990). YAC 18A3D (track 5) contains additional bands which represent dimers and trimers of the 5.7kb unit. These arise as *Eco*RI sites are lost as the DNA becomes more diverged towards the edge of the array. YACs 62C12, 281C5 and 17.2 (tracks 1-3) also still contain the 5.7kb unit (though the hybridisation signal is less intense than to 1B9F) but also contain many additional bands which cannot be accounted for by the simple loss of the occaissional *Eco*RI site. These band therefore represent a high degree of divergence of alphoid DNA at the edge of the array. YAC 417D3 (track 8), which spans the alphoid-nonalphoid boundary on the long arm side of the alphoid array, also contains the 5.7kb higher order unit but two additional bands are seen, representing a small degree of divergence.

The results from hybridising the YACs to the 5bp satellite repeat were as expected: only YAC 417D3 (track 15) contains the 5bp satellite. The *Alu* hybridisation results were also as expected. YACs 62C12 and 281C5 which extend ~300 and ~150kb repectively away from the major block of alphoid DNA contain many *Alu* sequences. Furthermore, the *Alu* bands in YAC 281C5, which is encompassed by 62C12, are a subset of the 62C12 bands. The location of *Alu* sequences in these YACs was therefore consistent with the normal distribution in the human genome. YAC 417D3 also contains an *Alu* sequence at the boundary between alphoid and nonalphoid DNA and this was discussed in detail in section 5.1.3.b. Finally, this Figure demonstrates that YACs 62C12 and 281C5 (tracks 16 and 17) contain the single 19kb short arm array of the 48bp satellite (as discussed in section 5.1.3.c.).

5.1.3.e. Orientation of alphoid DNA.

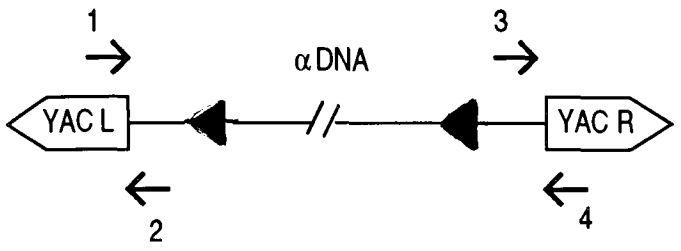
The centromeric regions of the three chromosomes of the fission yeast *S. pombe* contain repeated sequences that are arranged in a palindromic fashion (Chikashige *et al.*, 1989). This suggested a simple question about human centromeres: is alphoid satellite DNA also palindromic? This question was addressed by determining the orientation of Y chromosome alphoid DNA in the OXEN and Boleth individuals by analysing the alphoid YACs derived from these cell lines using a PCR based assay outlined in 5.18. Since the exact position and orientation of the YACs with respect to Yq and Yp on both alphoid arrays was known (see Figure 5.19.b) this assay will reveal the

orientation of alphoid DNA on the chromosome. The DNA structure of the YACs analysed was discussed in section 5.1.3.c but in addition to these YACs the orientation of alphoid DNA in YAC 14A8A (Biochemistry, Oxford YAC library; Fisher, 1992) was determined. This YAC contains DNA from the edge of the alphoid block on the the short arm of the chromososme (see Figure 5.19) However, it is co-ligated and therefore it was not used in the experiments described in section 5.1.3.c as it would have given results which would have been difficult to interpret.

The design of the PCR based assay is shown in Figure 5.18. Each YAC was amplified using the combinations of the four primers listed (one vector and one alphoid per reaction) and the possible outcomes for each primer pair in both alphoid orientations are given. YACs which contain only alphoid DNA are expected to give a product in two of the reactions, which YACs which extend into flanking sequences or are co-ligated are expected to give product only in one of the reactions. For a given vector primer, (1 or 4), determining which of the alphoid primers (2 or 3) produces the product reveal the orientation of the alphoid DNA adjacent to the vector.

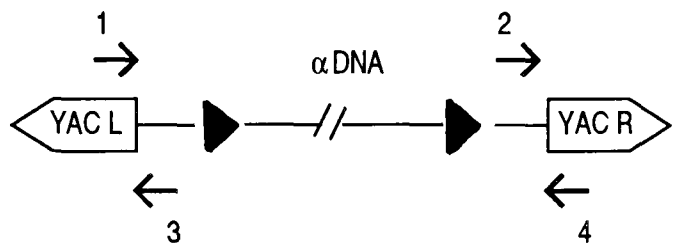
Figure 5.19.a shows the results of the PCR reactions. YACs 62C12, 281C5, 17.2, 14A8A and 417D3 give a product in only one reaction, while 18A3D, 1B9F and αB3 give products in two reactions. These results are as expected, except for 17.2 which lies entirely within the alphoid array and so is expected to amplify from both ends. There could be two reasons for its failure to amplify the outer end: the alphoid sequence could be so diverged that the primer no longer binds, or the amplification product could be rather large since the primers were chosen to be near the *Eco*RI site in the alphoid unit and 17.2 is an *Sst*I - *Sst*I clone. Experiments by Fisher (1992) indicate that both of these factors contribute: under low stringency PCR conditions a larger product is seen. The size of the products from the other YACs is as expected, and the orientations determined are shown in Figure 5.19.b. The two thick black lines represent the major blocks of Y chromosome centromeric alphoid DNA in the Boleth and OXEN individuals. The lines are different sizes because in OXEN the alphoid block size is ~1Mb (Oakey and Tyler-Smith, 1990) and in Boleth ~310kb (Figure 5.15). This variation in size is due to the polymorphic nature of Y chromosome alphoid DNA in different individuals (Oakey and Tyler-Smith, 1990). The YACs, derived from each individual, are positioned on the arrays. The arrows below the YACs represent both the ends of the YACs which amplified in the PCR assay and also the orientation of alphoid DNA.

(a) YAC clone, α orientation 1



PCR primers:	1+2	3+4	1+3	2+4
product :	+	+	-	-

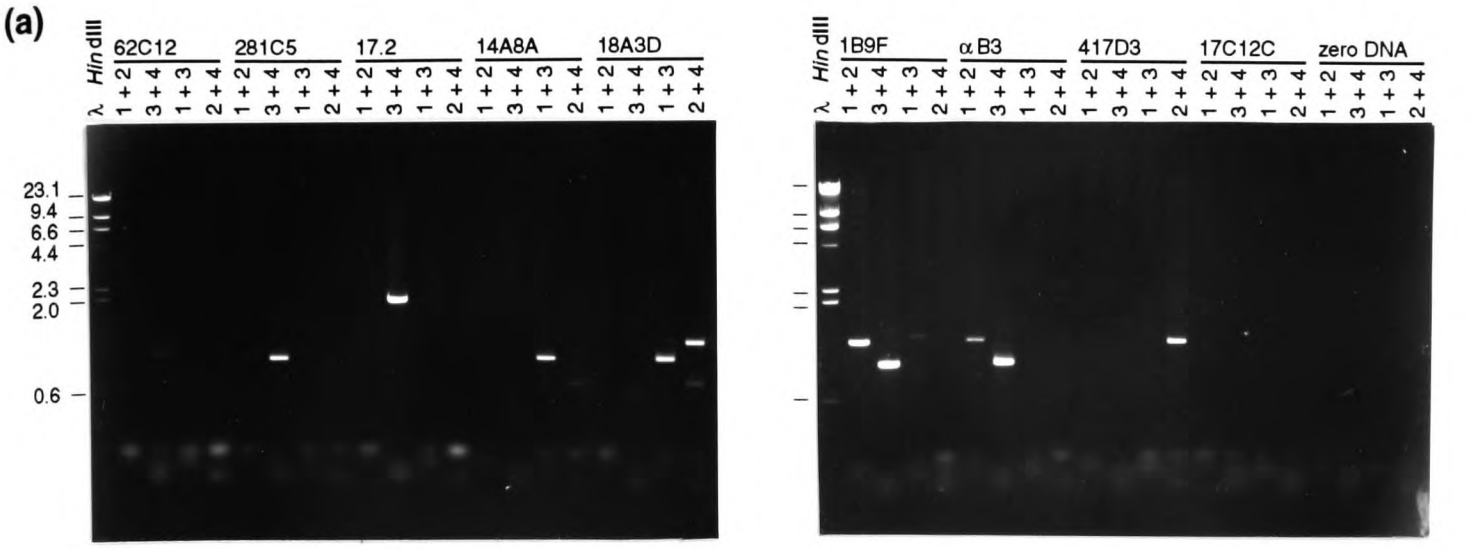
(b) YAC clone, α orientation 2



1+2	3+4	1+3	2+4
-	-	+	+

primer 1 :- pBR.A
primer 2 :- Y α I.A
primer 3 :- Y α I.C
primer 4 :- pBR.B

Figure 5.18. Determination of the Y alphoid DNA in YAC clones using PCR.



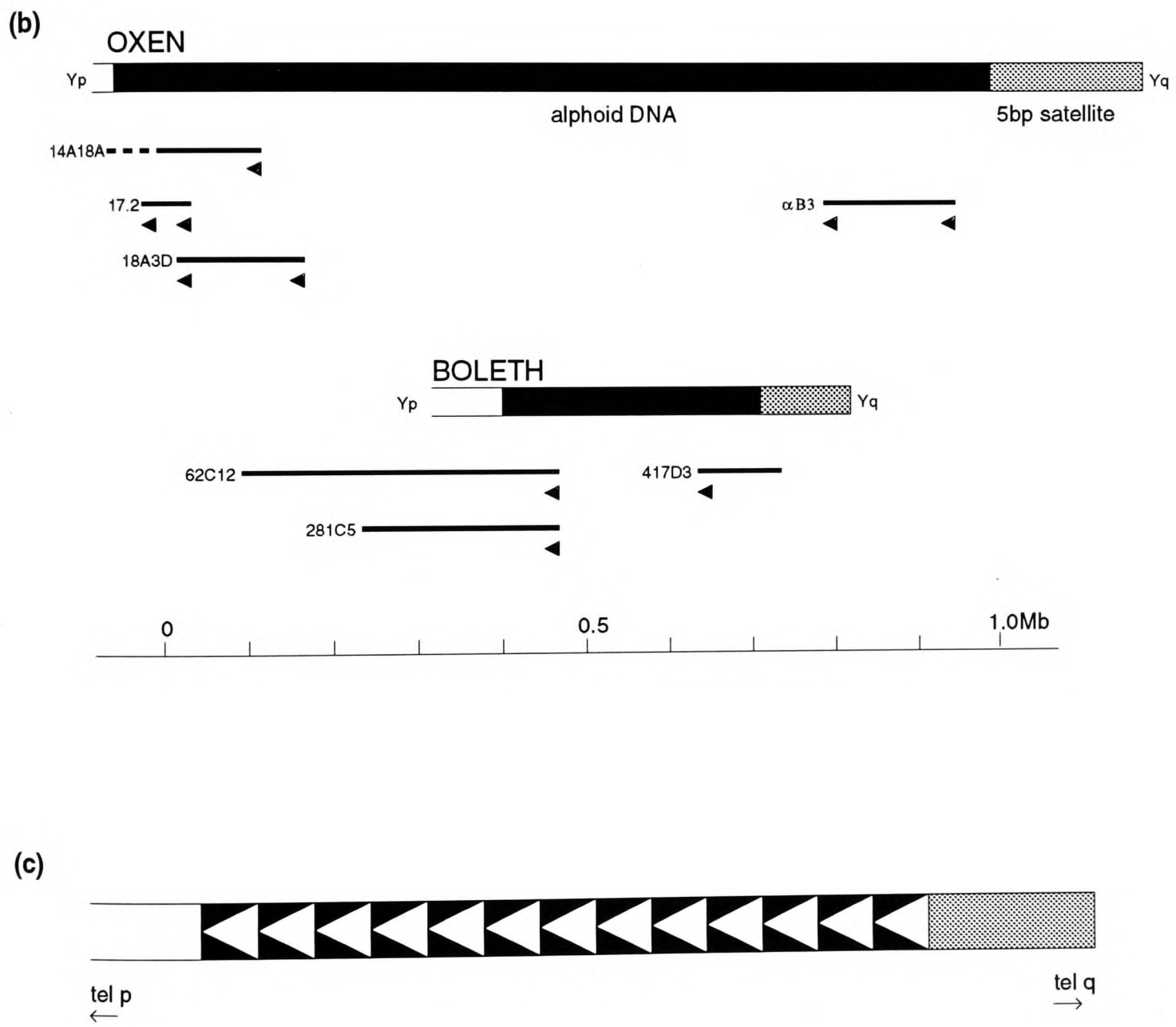


Figure 5.19. Determination of the orientation of alphoid DNA in the major array. (a) Ethidium bromide stained 1% conventional agaraose gel of the products obtained after amplification of the YACs shown using the combinations of the primers 1-4 stated (see Figure 5.18). (b) Interpretation of the PCR results. The YACs are located on the alphoid arrays from which they derived and the arrows represent the orientation of alphoid DNA. The stippled line represents a co-ligated YAC. (c) Schematic representation of the orientation of alphoid DNA on the human Y chromosome.

These results demonstrate that alphoid DNA on the human Y chromosome is orientated in one direction and this is drawn schematically in Figure 5.19.c. This result shows that, unlike the organisation of the *S. pombe* centromeric repeats, alphoid satellite DNA on the human Y chromosome is not palindromic and is probably a simple tandem repeat throughout.

5.2. ANALYSIS OF YACS FROM THE AMGL CONTIG.

The YACs from the AMGL contig were analysed for several reasons :- (1) They are located in the defined Y chromosome centromeric region and therefore analysis of the YACs would define the underlying structure of DNA from this region. (2) The YACs span a region of X-Y homology (pDP34 and GMGXY8) and also contain a homologous X-Y gene (amelogenin) which is expressed during tooth development from both sex chromosomes (Salido *et al.*, 1992). (3) Some of the YACs are quite large and it was anticipated that probes could be isolated from them which subsequently could be used to determine (a) whether the alphoid and AMGL contigs overlap and (b) whether the long range physical map could be closed.

After screening the three YAC libraries discussed in Chapter 4, nine overlapping YACs were isolated which make up the AMGL contig. However, two additional YACs, (Amy8 and Amy11-4), were isolated from a fourth library (Monaco *et al.*, 1991) by C. Tyler Smith, which also overlap with YACs from the AMGL contig. These YACs and YACs AA1A10/11 and CC1C10/6 which contain the GMGXY8 locus were analysed by A. Kay (Department of Biochemistry, University of Oxford) and the results of this work is summarised in Figure 5.28

5.2.1. Restriction site maps of the AMGL YACs.

6 YACs from the AMGL contig were mapped in the way described for the alphoid YACs (section 5.1.2). The YACs proved straight-forward to map except for YAC 235F3. Therefore only results of the mapping data obtained for the most distal YAC of the contig analysed (18C4, Figures 5.20 and 5.21) and the proximal YAC 271B5 (Figures 5.22 and 5.23) are shown. This latter YAC is not the most proximal in the contig but the results are shown because it is the most proximal YAC that does not have a complex structure.

YAC 235F3 proved difficult to map because a small, but noticeable, percentage, of the yeast cells harboured a deleted version of YAC 235F3. This results can be seen in Figures 5.23.a and 5.23.b. Here the YAC has been hybridised with

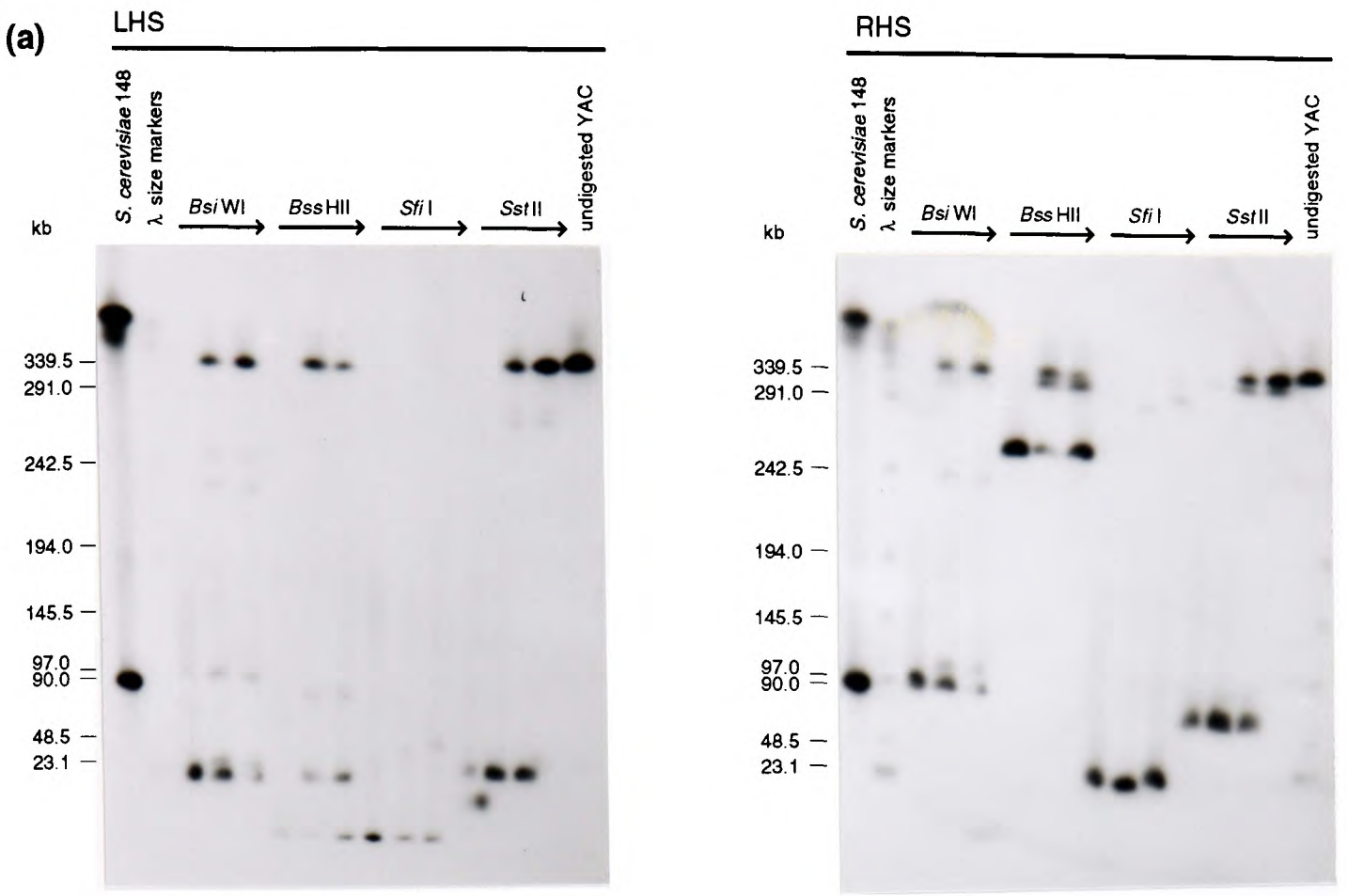


Figure 5.20. Analysis of YAC 18C4. (a) Hybridisation of the left and right hand vector probes to partial digests of the YAC. The DNA was partially digested with the enzymes indicated, fractionated by PFGE (effective pulse time 11 secs, run time 32hrs), transferred to a filter and hybridised with the right and left arm vector probes. The filters were washed in 0.1.x SSC and the position of the λ concatemer, λ *Hin*dIII and *S. cerevisiae* 148 size markers are shown at the side of each photograph. The arrows drawn below each enzyme represent decreasing numbers of units used. (b) Restriction site map of the YAC showing the location of the loci detected by pDP34 and AMGL.

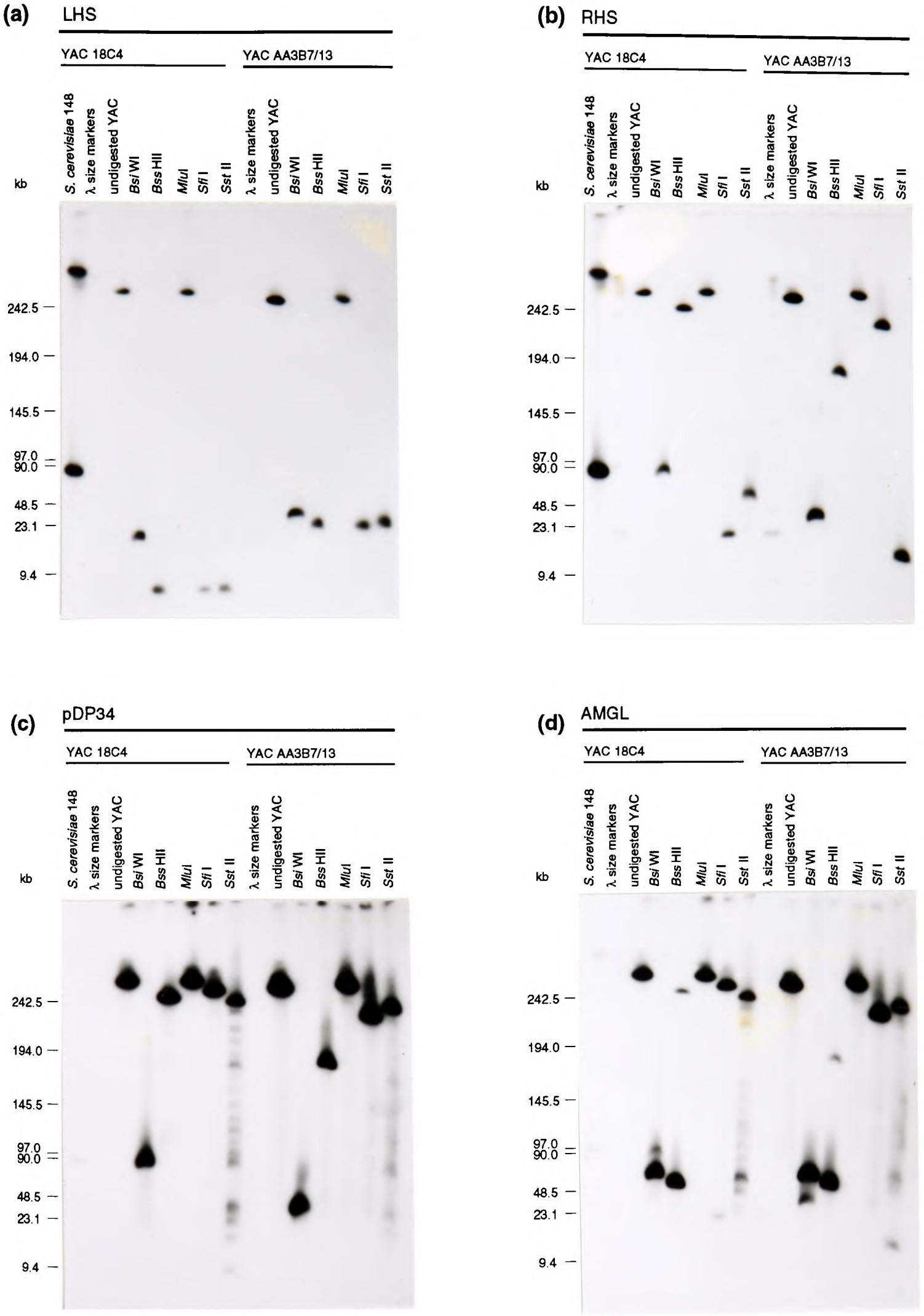
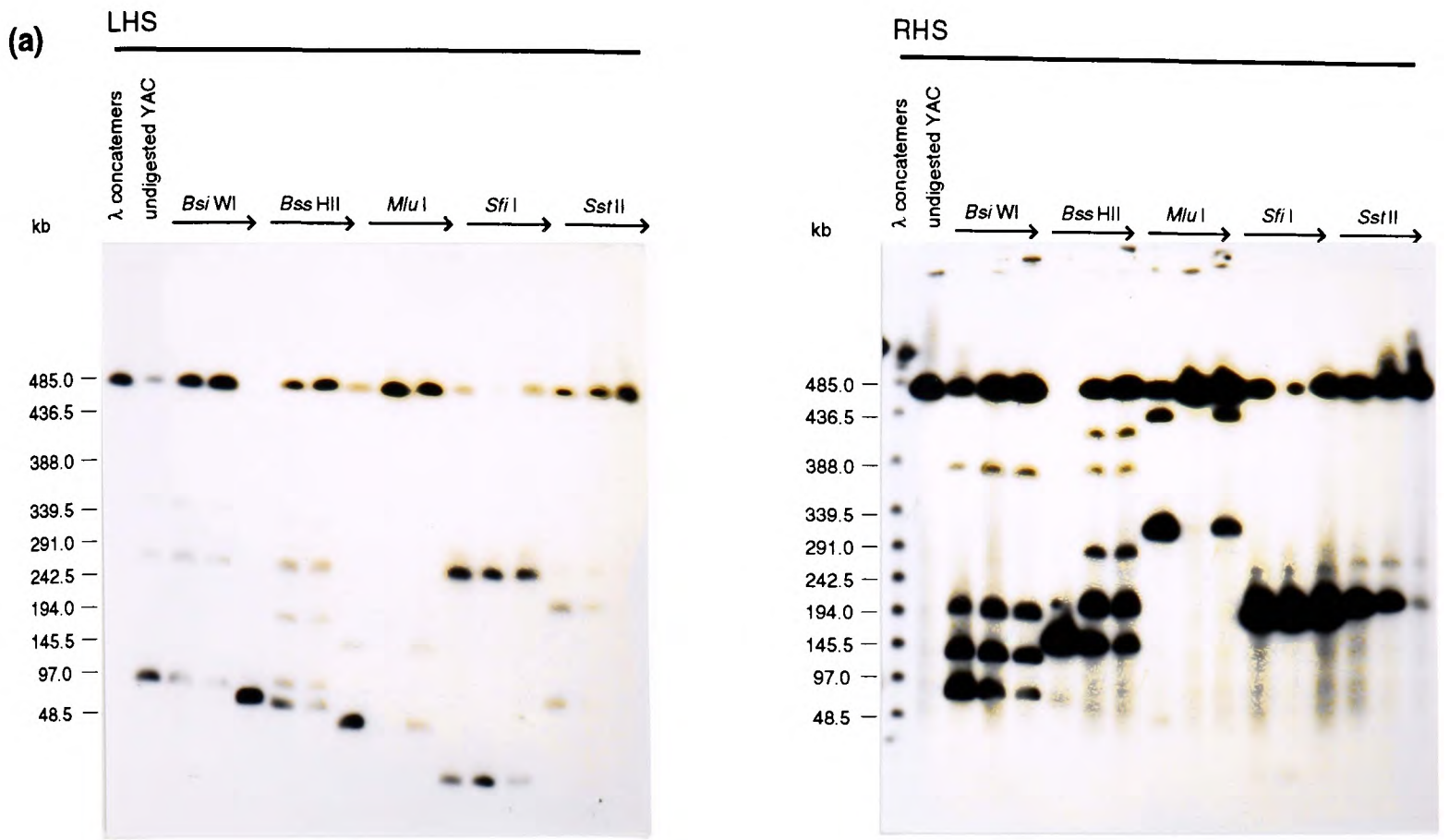


Figure 5.21. Localisation of various probes on YAC18C4 and AA3B7/13. The YAC DNA was digested to completion with the enzymes indicated, fractionated by PFGE, transferred to a membrane and hybridised with probes (a) - (d). The filters were washed in 0.1.x SSC and the position of the *S. cerevisiae* 148, λ concatemers and λ *Hin*dIII size markers are shown at the side of each photograph. The pulsed-field gel was run for 32hrs with an effective pulse time of 10 secs.



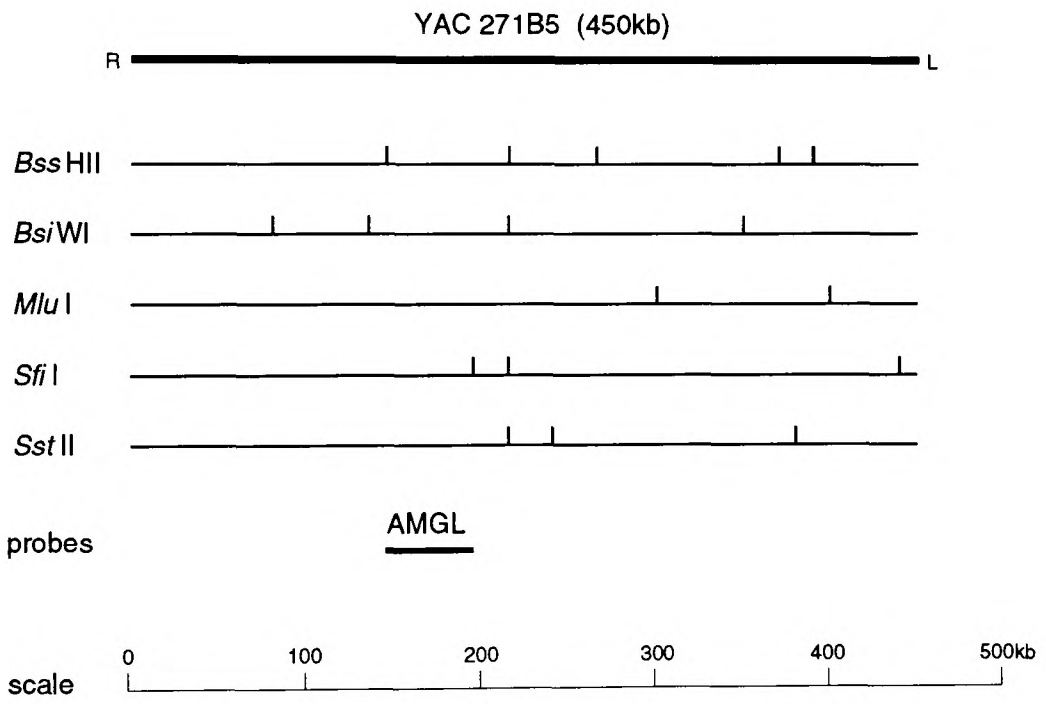


Figure 5.22. Analysis of YAC 271B5 (a) Hybridisation of the left and right hand vector probes to partial digests of the YAC. The DNA was partially digested with the enzymes indicated, fractionated by PFGE (effective pulse time 24 secs, run time 24hrs), transferred to a filter and hybridised with the right and left arm vector probes. The filters were washed in 0.1.x SSC and the position of the λ concatemer size markers are shown on the left hand side of each photograph. The arrows drawn below each enzyme represent decreasing number of units used. (b) Restriction site map of the YAC showing the location of the AMGL locus.

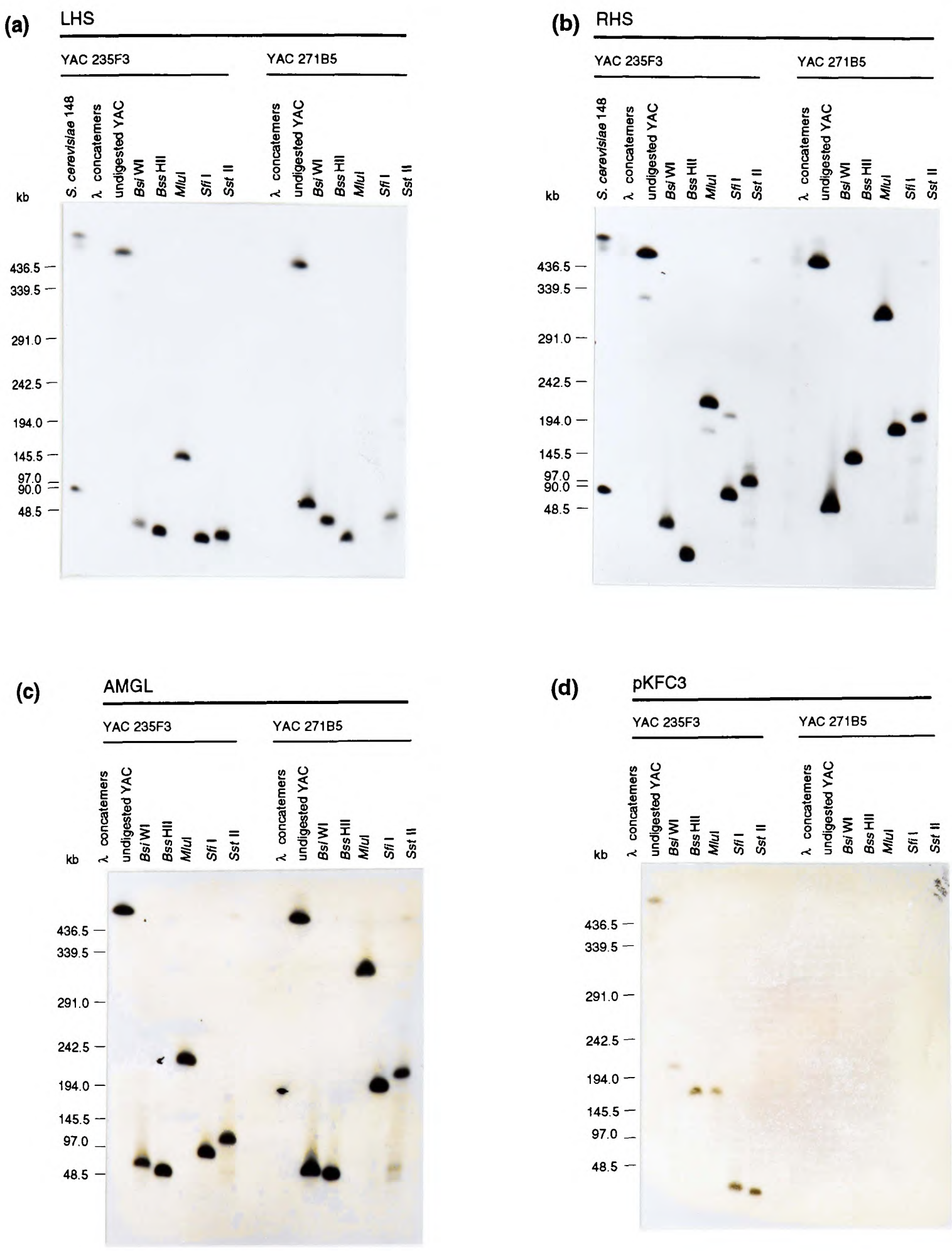


Figure 5.23. Localisation of various probes on YACs 235F3 and 271B5. The YAC DNA was digested to completion with the enzymes indicated, fractionated by PFGE, transferred to a membrane and hybridised with probes (a) - (d). The filters were washed in 0.1.x SSC and the position of the *S. cerevisiae* 148 and λ concatemer size markers is shown at the side of each photograph. The pulsed-field gel was run for 32hrs with an effective pulse time of 15 secs.

the left and right arm vector probes respectively and for both probes two bands are seen in the undigested YAC, the smaller, fainter band representing the deleted version of the YAC. This second band could not be due to contaimination with another YAC because when the digested YAC is hybridised with the left arm vector probe only one band is seen. One band is also seen when the digested YAC is hybridised to AMGL (c) and KFC3 (d) but two bands for the right arm vector probe in the *Mlu*I and *Sst*II digests but not in *Bsi*WI or *Bss*HII digests (b). Therefore this suggests that the deleted YAC has lost about 170kb of DNA starting 80-90kb from the RHS and encompassing the AMGL locus. The restriction site map of this YAC was therefore obtained by analysing the results from the complete digests, interpreting the partial digests and comparing the map to maps obtained from more stable overlapping YACs.

5.2.2. Restriction site map of the AMGL contig.

As expected the 6 AMGL YACs all overlapped and the resulting restriction site map of the contig is given in Figure 5.24. The YACs span ~600kb and the region of X-Y transposition (pDP34 and GMGXY8 loci) is located approximately 170kb away from the AMGL gene. This is interesting because this region of X-Y homology is located on the long arm of the X chromosome (Page *et al.*, 1984; Affara *et al.*, 1986a) whereas the X copy of the amelogenin gene is located on the short arm (Salido *et al.*, 1992). Therefore the junction between these two events must be located in the YACs. Another intersting feature of the map is the presence of a cluster of the CpG restriciton enzyme sites very close to the location of the AMGL locus. This is suggestive of there being a CpG island (Bird, 1986) located upstream of the gene as sites for these enzymes are rare or blocked by methylation in the rest of the genome (Brown and Bird, 1986). The total region of X-Y homology (GMGXY8 to AMGL) terminates around 100kb proximal to AMGL since probe 296e (discussed in section 5.3.2) detects a Y chromosome specific single copy sequence. Finally, some fragment sizes are not quite consistent between YACs 235F3 and 271B5. Thus, as this part of the map is not confirmed by other YACs it has to be regarded as provisional.

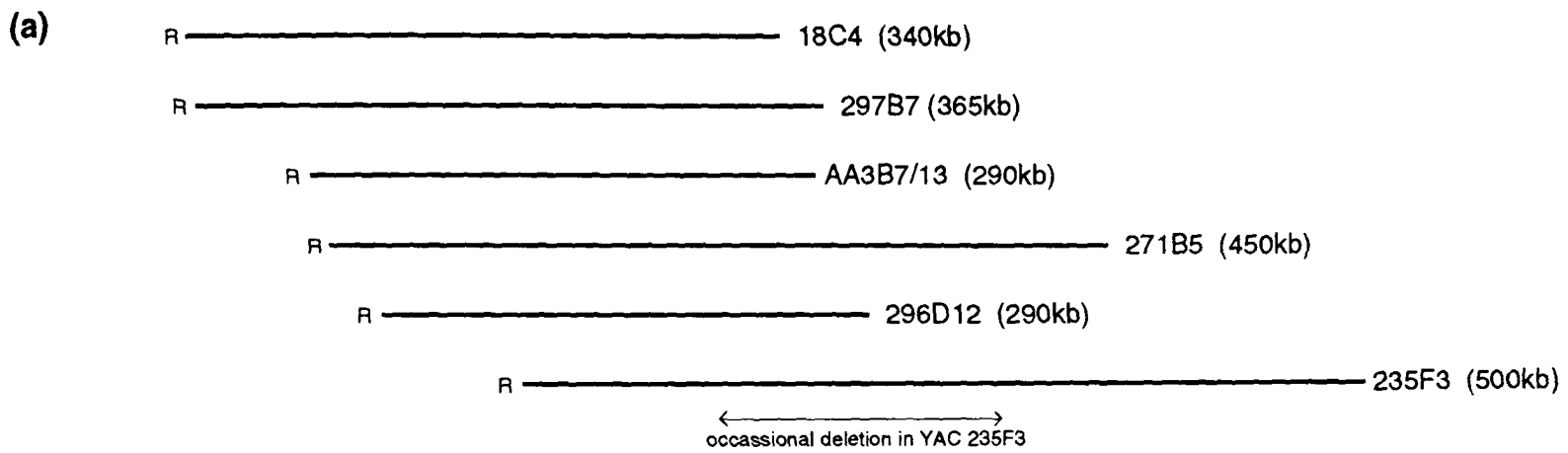


Figure 5.24. Restriction site map of the YACs which form the AMGL contig. (a) YACs; (b) restriction site maps; (c) location of probes. R represents the right (non-centromeric) arm of the YAC vector.

5.3. INVESTIGATION OF THE NATURE OF THE SEQUENCES IN THE PERICENTRIC SHORT ARM REGION.

The YACs from both the alphoid and AMGL contigs were further analysed to investigate the nature of sequences present in the pericentric short arm and also to search for probes which could be used to complete the long range physical map of the Y chromosome centromeric region. The cell lines and probes used to define this region have been discussed in section 1.5.2. and at the time the alphoid and AMGL contig YACs were isolated the long range map contained one gap. This gap coincides with the region where the two YAC contigs did not obviously meet. The simplest way therefore to obtain potential "gap-closing" probes from the YACs was to plasmid-end rescue (Burke *et al.*, 1987 and section 2.7.4) sequences from the distal end of the alphoid contig and the proximal end of the AMGL contig. Probes isolated from these positions would determine whether the YACs from both contigs overlapped and reveal the nature of the sequences in this part of the chromosome.

5.3.1. Plasmid-end rescue of YACs from the alphoid contig.

YACs 62C12, AA4A7/11 and AA15G8/6 from the alphoid contig were all found to be in the correct orientation for trying to obtain "gap-closing" probes; i.e. the left arms of the YACs were distal to alphoid DNA (see Figure 5.12). However, the mapping data presented in this Figure showed that the left arm of YAC 62C12 is located at the Y190 minor locus. Therefore, a plasmid-end rescued probe derived from this YAC would contain Y190 sequences which are already available and are not useful for long range mapping as there are two loci on the Y chromosome. (Tyler-Smith *et al.*, 1988) Consequently, probes were only isolated by plasmid-end rescue from YACs AA4A7/11 and AA15G8/6.

Before using the probes for long range mapping it was essential to determine whether they would be suitable by analysing the type of sequences they contained. In addition, it was important to determine whether these probes, isolated from the distal most location in the YACs, were present in any of the YACs from the AMGL contig. These two aims were achieved by hybridising the probes, with competition, to a panel of human, Y-hybrid, rearranged cell lines and YAC DNA from both contigs which had been digested with *Hin*dIII, fractionated through a 1% conventional agarose gel and transferred to a filter.

5.3.1.a. pAA15e.

The results for AA15e, the end-rescued probe derived from YAC AA15G8/6, are shown in Figure 5.25. The probe detect two major bands in YAC AA15G8/6 and AA4A7/11 and a complicated pattern in human DNA. Two bands are seen in the YAC DNAs because the *Nde*I-*Eco*RI fragment used as a probe spans a *Hin*dIII site and so detects two *Hin*dIII fragments. The lower band is different in size because it represents the very end of YAC AA15G8/6 but an internal fragment in AA4A7/11. The patterns in human DNA are much more complex, showing that AA15e detects a moderately repeated family of sequences, most of which are not located on the Y chromosome. The Y bands are difficult to score against this background, and so cannot be reliably localised on the Y chromosome. Hence the probe would not be suitable to use for long range mapping in genomic DNA. However, AA15e has shown that the two YAC contigs discussed in this chapter do not overlap. The probe does not hybridise to YACs 18C4 , 296D12 and 235F3 (tracks 11-13) and YAC235F3 is the most proximal YAC in the AMGL contig.

5.3.1.b. pAA4e.

The plasmid-end resuced probe derived from YAC AA4A7/11 was isolated by C. Tyler-Smith. The results, (personal communication), are not shown because the plasmid-end rescued probe derived from this YAC did not contain human DNA but additional plasmid DNA. This result agrees with the end-labelling experiments (section 5.1.2.d) which suggested that the structure of the vector arms in this YAC was abnormal.

5.3.2. Plasmid end-rescue of YACs from the AMGL contig.

The restriction site maps of the AMGL YACs, discussed in section 5.2, revealed that all of them were orientated in the correct orientation for plasmid-end rescue; i.e. with their left arms extending towards alphoid DNA (see Figure 5.24). However, probes were only successfully isolated from YACs AA3B7/13 and 296D12. Again, these probes were tested, as described in section 5.3.1, to determine if they would be suitable to use for long range mapping. The results of the hybridisation of the end-rescued probe recovered from YAC 296B12 (p296e) to the *Hin*dIII panel are shown in Figure 5.26. This probe detects a major ~23kb band in male, OXEN, Y-hybrid 3E7, ChoP and 651 cell lines and YACs 296B12, 271B5 and 235F3 (tracks 3, 4, 5, 8, 9, 13, 15 and 16) which is absent from female, 3E7 revertant, 89030291 and 861748 cell lines and all the other YACs (tracks 2, 6, 7, 10, 11, 12, 14, 17, 18 and 19). The additional

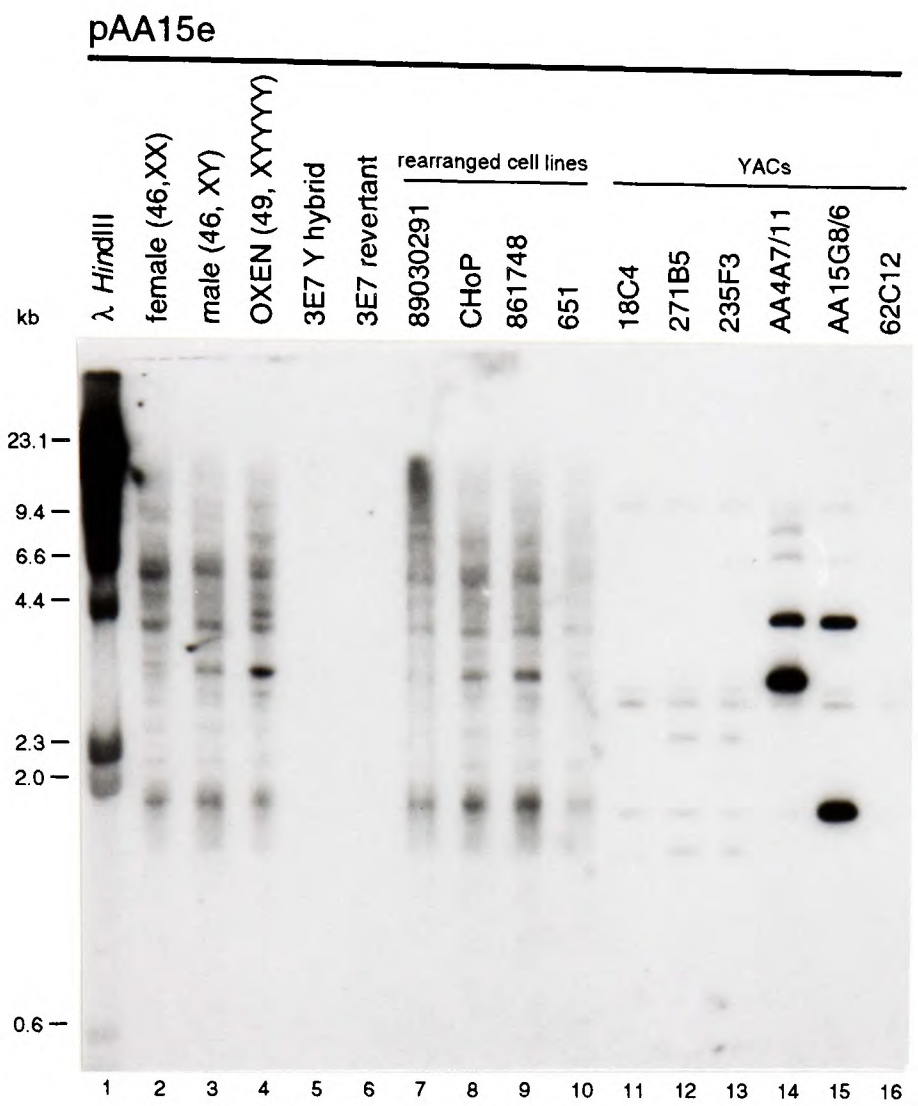


Figure 5.25. Determination of the type of sequences detected by plasmid-end recued probe AA15e. The DNA was digested with *Hin*dIII, fractionated by conventional agarose gel electrophoresis and a filter transfer hybridised to pAA15e using competition. The filter was washed in 0.1 X SSC. The postion of the λ *Hin*dIII size markers are given on the left hand side and the rearranged cell lines are described in detail in the text.

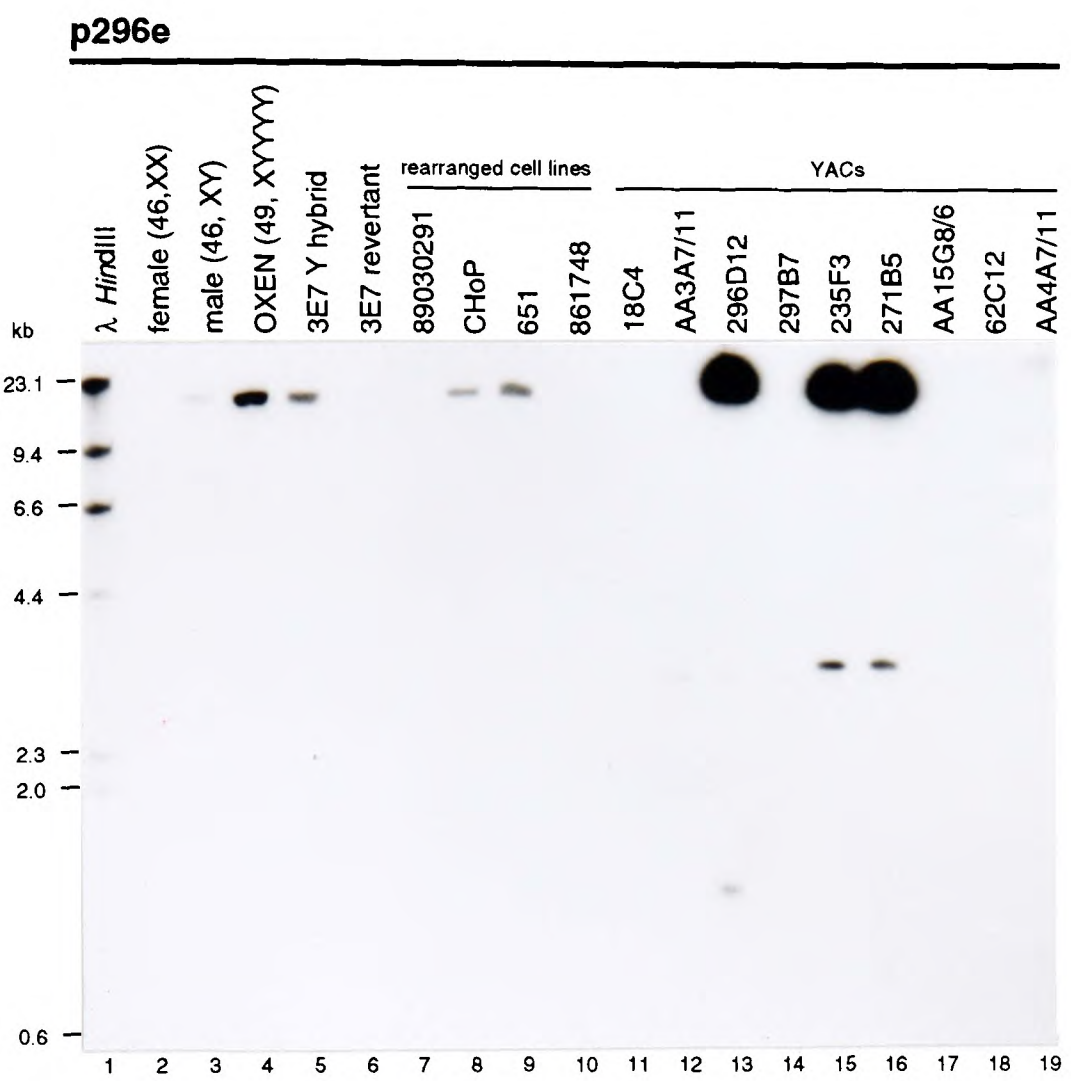


Figure 5.26. Determination of the type of sequences detected by plasmid-end rescued probe 296e. The DNA was digested with *Hin*dIII, fractionated by conventional agarose gel electrophoresis and a filter transfer hybridised to p296e using competition. The filter was washed in 0.1 x SSC. The position of the λ *Hin*dIII size markers are given on the left hand side and the rearranged cell lines are described in detail in the text.

bands seen in the other YAC represents background vector hybridisation. Therefore as probe 296e detects a single copy Y chromosome specific sequence it was an ideal candidate to use for long range mapping.

The results from testing the suitability of the end-rescued probe AA3e (recovered and analysed by C. Tyler-Smith), derived from YAC AA3B7/13, for long range mapping are not shown. However, when the probe was hybridised using competition to the *Hin*dIII panel discussed above it detected a large Y specific band in the same DNA samples as p296e but in addition it detected a smear in each of the human DNA tracks. Therefore this probe must contain the ubiquitous repetitive *Alu* and/or *Kpn* sequences and for the same reasons as pAA15e (section 5.3.1) would not be suitable for long-range mapping. Finally, because pAA3e is located very close to p296e (see Figure 5.28) which is an ideal probe for long-range mapping, it was not analysed any further.

YACs 235F3 and 271B5 from the AMGL contig are 500 and 450kb in size respectively and are orientated in the correct direction for obtaining potential "gap-closing" probes. However, the plasmid-end rescue experiments were not successful as the appropriate restriction sites were not present in the YACs. Therefore, probes were obtained from YAC 235F3 which extends the furthest proximally (Figure 5.24) by subcloning the YAC into plasmids and these results are discused below.

5.3.3. Analysis of plasmid subclones isolated from YAC 235F3.

It was very important to obtain clones from YAC 235F3 because the YAC extends approximately 300kb more proximally than p296e. (See Figures 5.24 and 5.28). As a result it was anticipated that clones isolated from the left hand arm of this YAC might be able to link up the long-range map.

A plasmid library was created from this YAC as described in section 2.7.3. The library was plated out onto filters and screened with total human, total yeast, YAC 235F3 and 271B5 DNA. Clones were isolated that hybridised to YAC 235F3 but not to the other probes. This screening strategy was employed to detect clones which were: (a) not highly repeated, (b) not yeast and (c) not also present in YAC 271B5. This last point was important because YAC 271B5 does not extend as far proximally as 235F3. In total, nine clones were isolated and the type of sequence they contained was determined by hybridising the insert DNA, using competition, to a small panel of *Hin*dIII digested human and

YAC DNA. The results for the best two probes (pKFC3 and pKFC5.1) are shown in Figure 5.27. In both cases the probes detected a major band (marked with an arrow) in YAC 235F3 (tracks 7) which was not present in YAC 271B5 (tracks 6). For both probes the band was present in male (tracks 3) and 3E7 Y-hybrid DNA (tracks 4) but absent from female and 3E7 revertant DNA (tracks 2 and 6 respectively).

These results show that the clones isolated from the proximal region of YAC 235F3 are derived from the human Y chromosome, providing more evidence that YAC 235F3 was unlikely to be co-ligated in this region. However, both the probes detected a smear in male and female DNA and therefore were not suitable candidates to use for long range mapping in male DNA for the reasons discussed above (section 5.3.1). Probe KFC3, which detected a single band in the Y hybrid 3E7, was used for long range mapping in this cell line. The results are not shown but showed that in 3E7 pKFC3 does not lie on the same restriction fragments as 2D5, the most distal useable probe from the alphoid contig (see Figure 5.28. Finally the probes were located on YAC 235F3 by hybridising them to the same filter as discussed in section 5.2.1.e, containing complete digests of the YAC. The results (an example of which is shown in Figure 5.23.d and summarised in Figure 5.28) confirm that the probes are located very near the left vector arm of the YAC.

5.3.4. Screening the Y chromosome cosmid library.

It was considered important to continue to try and obtain a clone from this region which could be used for long range mapping and therefore the Y chromosome cosmid library was screened with pKFC3 and pKFC5.1, using competition, as described in section 2.8. The results are not shown but three cosmids, (56A9, 110F4 and 125G4) were isolated. It was determined that the cosmids were different but overlapped; 56A9 containing both KFC3 and KFC5.1 and cosmids 110F4 and 125G4 containing KFC3 or KFC5.1 respectively (see Figure 5.28).

In order to obtain a probe from these cosmids that did not contain the ubiquitous *Alu* and *Kpn* repeats the cosmids were subcloned into plasmids. Unfortunately all the subclones contained repeats and therefore could not be used for long range mapping in male DNA. The largest suclones were further digested with a range of frequent cutting enzymes but still no non-repeated fragments were detected. Unfortunately, lack of time prevented the further analysis of the cosmid subclones.

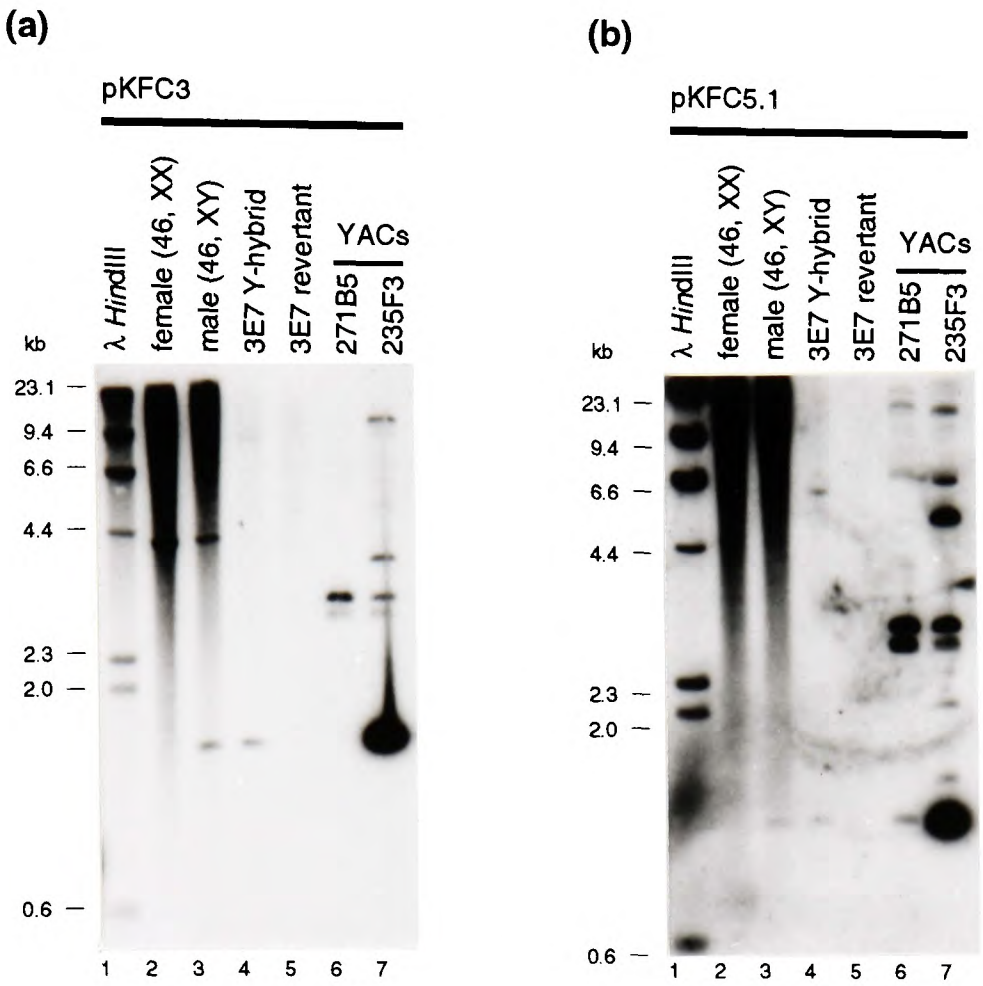


Figure 5.27. Determination of the type of sequences detected by probes KFC3 and KFC5.1 which were derived from YAC 235F3Human, Y-hybrid and YAC DNA was digested with *Hin*dIII , fractionated by convetional agarose gel electrophoresis and a filter transfer hybridised using competition with pKFC3 (a) and pKFC5.1 (b). The filters were washed in 0.1 x SSC anf the position of the λ *Hin*dIII size markers is shown at the side of each photograph.

Therefore due to the highly repetitive nature of DNA from the centromeric region of the human Y chromosome the only suitable clone isolated from either the AMGL or alphoid contigs was p296e which could be used for long range mapping. However, this probe still did not detect overlapping restriction enzyme fragments which would close the gap in the long range physical map.

5.4. SUMMARY OF CHAPTER 5.

Fourteen YACs derived from the short arm pericentric and centromeric region of the human Y chromosome were analysed and it was determined that the YACs fall into two main contigs which in total cover ~1.2Mb of genomic DNA. These results are represented schematically in Figure 5.28. In addition to the 6 AMGL YACs discussed in this work, the location of the 4 other, more distal, overlapping YACs is also shown. These YACs were analysed by A. Kay (Department of Biochemistry, Oxford). By analysing these YACs the types of sequences present in this region of human DNA were determined. The region predominantly contains low to moderate copy repeated sequences, indicated in Figure 5.28 as white boxes. Also present is a region of X-Y transposition, a gene (stippled box), a single copy sequence (striped box) and the alphoid and 5bp satellites which are tandem repeats.

One of the aims of analysing the YACs was to obtain a probe which could be used to close the single remaining gap in the long range restriction site map of the human Y chromosome centromeric region. Figure 5.28. shows the location of the novel probes isolated (p296e, pAA3e, pAA15e, pKFC3 and pKFC5.1). Unfortunately the only probe which was suitable to use for long range mapping in normal human DNA was p296e and this probe did not close the gap. The other probes could not be used because they were all repeated in nature. However the isolation of probes pAA15e and pKFC3 which are located at the distal and proximal ends of the alphoid and AMGL contigs respectively confirmed that these two contigs do not overlap.

Three YACs from the alphoid contig containing alphoid DNA, (62C12, 281C5 and 417D3), were analysed in detail to determine the organisation of alphoid DNA on the human Y chromosome.The PCR assay showed that alphoid DNA is not palindromic and is probably a simple tandem repeat throughout. Analysis of these YACs also determined that the 5.7kb alphoid higher order unit structure becomes very diverged at the short arm edge of the major block of alphoid DNA and is flanked by a small block of the 48bp satellite and low to

moderate copy sequences. The long arm edge of the alphoid block shows less divergence and is flanked by the 5bp satellite.

5.5. SUMMARY OF COMPLETE RESULTS.

Throughout this work the two fold approach of obtaining a long range restriction map and analysing YACs was adopted to provide detailed information about the sequences present at the centromeric region of the human Y chromosome. The final diagram (Figure 5.29) is therefore a schematic representation of the structure. There is a central block of alphoid satellite DNA. This is flanked on the long arm side by a large block of the 5bp satellite and more distally the 48bp and 68bp satellites and a range of other repeats. The short arm side is characterised by a small block of diverged alphoid DNA and a small block of the 48bp satellite but is predominantly made up of low to moderate copy number repeats.

To conclude therefore, the human Y chromosome centromeric region has a complex asymmetric structure, made up of a mosaic of tandem and other repeats.

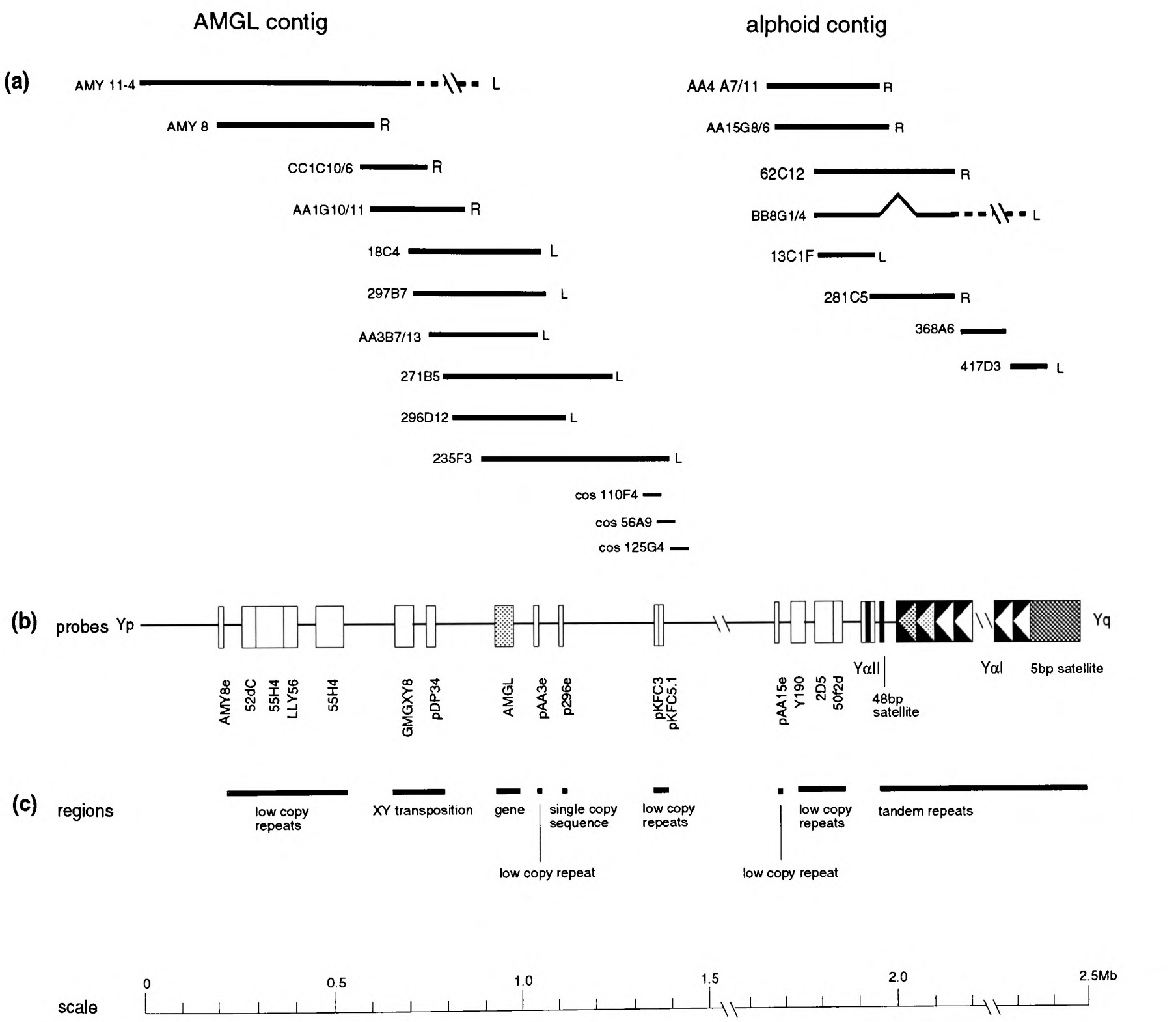


Figure 5.28. Schematic representation of the organisation of sequences found in the short arm centromeric/pericentric region of the human Y chromosome. (a) Overlapping YACs in the AMGL and alphoid contigs; (b) probes; (c) regions defined by the types of sequences present.

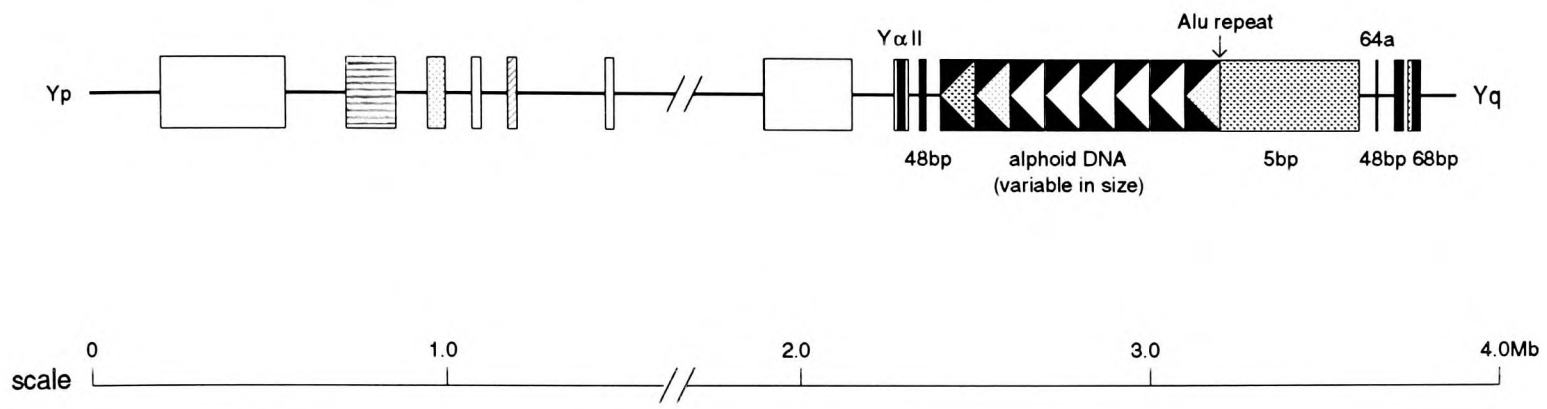


Figure 5.29. Detailed schematic representation of the centromeric region of the human Y chromosome.

CHAPTER 6

DISCUSSION

This thesis has described the molecular analysis of the centromeric region of the human Y chromosome. Five YAC contigs were assembled (3 of which were analysed in detail) and placed within the region. By combining data from the analysis of these clones with the results of deletion mapping of abnormal chromosomes and restriction site mapping of genomic DNA, it has been possible to determine in outline the sequence organisation of the human Y chromosome centromere. In this discussion the use of YACs to analyse this poorly understood part of the genome will be considered (section 6.1), followed by conclusions about the sequence organisation (6.2) and the implications for centromere structure and function (6.3).

6.1. THE USE OF YACS TO ANALYSE THE CENTROMERIC REGION OF THE HUMAN Y CHROMOSOME.

6.1.1. Screening YAC libraries.

During the course of this work three YAC libraries were screened with all suitable probes from the centromeric region of the human Y chromosome: two (Burke *et al.*, 1987; Fisher, 1992) by colony hybridisation and one (Albertsen *et al.*, 1990) by using a PCR based assay The advantages and disadvantages of using each technique will be discussed below.

6.1.1.a. Colony hybridisation.

The major advantage of using colony hybridisation to screen the YAC libraries was that only one round of screening was required to identify the clone. In addition, the libraries were gridded out onto filters which were screened numerous times with different probes. There proved to be two major disadvantages. (1) Considerable numbers of false positives were detected by cross-hybridisation especially as many of the Y centromeric probes contain repeated sequences found in a number of YACs. For example, out of 77 YACs picked from the Washington, St. Louis library only 26 were derived from the human Y chromosome. However, this statistic is rather biased because if the YACs were not easy to identify 3 or 4 colonies were picked and probe content mapped to ensure that the colony responsible for the signal was

obtained. When the library was screened with YαI, the Y chromosome alphoid probe, over 200 colonies were identified. Approximately 50 of these were analysed but none of them was derived from the Y (C. Tyler-Smith; personal communication). (2) Preparing the filters for hybridisation for the Biochemistry, Oxford YAC library was a labour intensive task. However, two groups (Nizetic *et al.*, 1991; Bentley *et al.*, 1992) have now developed robots to produce high density grids of clones, thereby making hybridisation screening a very fast and easy method to identify YACs of interest. For example a 2.4Mb contig containing 36 overlapping YACs has been assembled which encompasses the Duchenne muscular dystrophy (DMD) locus (Monaco *et al.*, 1992).

6.1.1.b. PCR based assays.

The PCR based method used to screen the CEPH library was first described by Green and Olson (1990a). The advantages of using PCR to screen a library are numerous. For example, screening results in binary information which makes scoring easier, the process is amenable to automation, no radioactivity is needed and primers pairs can define sequence tagged sites (STS, Olson *et al.*, 1990) which can subsequently be used as milestones to construct maps of the YACs. This research has also shown that PCR based assays can be more discriminating than screening by colony hybridisation. For example the X and Y chromosome loci for the amelogenin gene could be discriminated by comparing the sizes of the product (see Table 2.5) whereas the colony hybridisation using a mouse cDNA probe identifies both X and Y copies.

The major disadvantage of using this technique was that several rounds of screening were required before the YAC of interest was identified. As a result, after the screening of the primary pools, which was in itself time consuming, it took approximately 6 weeks for the YAC responsible for the signal to be identified. Furthermore, YACs were never isolated from 4 of the 22 positive primary pools. The other disadvantage of screening a library by PCR was that the initial phase of PCR primer development, which involves sequencing a DNA fragment, is more time consuming and expensive than generating a probe to use for hybridisation. However, thereafter the expense of synthesizing the primers is no more than to prepare a plasmid, digest it with restriction enzymes and purify and label the probe for hybridisation.

Since the original paper by Green and Olson (1990a), YAC libraries have been screened using STSs to isolate YACs encompassing many loci. For example, Green and Olson (1990b) described the isolation of 30 YAC clones spanning ~1.5Mb encompassing a region of the cystic fibrosis (CF) locus. Also the Huntingdon disease region (HD) has recently been cloned in YACs isolated in this way (Zuo *et al.*, 1992). Finally, Anand *et al.* (1991) have used a joint PCR/colony hybridisation approach to screening libraries and have assembled a 1.5Mb contig encompassing the complete CF locus.

6.1.1.c. Under-representation of sequences.

During this analysis of the centromeric region of the human Y chromosome three YAC libraries (~ 7 x coverage in total) were screened with 11 probes (though not each library was screened with each probe - see chapter 4). For one of these probes, 64a no YACs derived from the Y chromosome locus were isolated. However, a single YAC containing the autosomal copy of 64a was isolated. It is possible therefore that this region is under-represented. Furthermore screening approximately two Y chromosome equivalents with the 68bp satellite produced 3 clones from pericentric Yq and one clone from distal Yq in good agreement with random expectation but the number of non-Y clones was less than expected. Other sequences have also been shown to be under-represented in YAC libraries. For example Zuo *et al.* (1992) screened six to eight genome-equivalents with 12 STS's and isolated 28 clones. This number of YACs was lower than the authors expected; based on random distribution and the observed average insert size they calculated that at least 72 clones should have been obtained. As yet no explanations have been given for under-representation of certain sequences in YAC libraries. Experiments by Abidi *et al.* (1990) showed that cloning of Xq sequences is essentially random and predicted that all or nearly all sequences are equally clonable in YACs. However, for 64a this under-representation could be explained by the fact that it lies very close to the ~400kb centromeric block of the 5bp satellite (Figure 6.2) and resulting clones could be too unstable to survive.

6.1.2. Analysis of YACs.

6.1.2.a.STS/probe content mapping.

Once YACs were isolated from the libraries the different loci they contained were determined by hybridisation and PCR (STS mapping). These techniques provided a quick means to fingerprint the YACs and allowed clones for

detailed analysis to be selected. Other groups have used probe content mapping to construct overlapping contigs. For example Green and Olson (1990b) first demonstrated that STS mapping could be used to construct a 1.5 Mb contig containing over 30 YACs encompassing part of the cystic fibrosis gene; and Ragoussis *et al.* (1991) used hybridisation to isolate YACs forming a 1.5Mb contig spanning the class II major histobility complex. However, both of these techniques rely upon the STSs and probes detecting unique sequences or loci which can be distinguished by size. In the case of Y chromosome centromere this is rare. For example, 0 out of 4 of the STSs used to screen the CEPH library and 1 out of 11 hybridisation probes (GMGY6) used to screen the Washington and Oxford libraries detect a single copy sequence. Consequently, most of the YACs derived from the Y chromosome centromeric region could only be differentiated from Y YACs derived from other regions of the chromosome and non-Y YACs by restriction enzyme digestion and hybridisation (see Table 4.5).

6.1.2.b. Restriction site maps of the YACs.

In this project the YACs were characterised mainly by restriction site mapping. The YACs were accurately aligned by comparing the individual restriction maps, and a consensus map of the region was obtained. However, many differences were observed between the restriction analysis of YAC inserts and the corresponding uncloned genomic DNA. This was because the cleavage by most rare-cutting enzymes except *Sfi*I is prevented by CpG methylation and this type of methylation occurs in genomic DNA of mammalian cells (Brown and Bird, 1986) but is lost in yeast.

These differences between yeast and mammalian DNA complicate the use of restriction mapping for determining correspondence between genomic and YAC DNA and have been reported to occur in the analysis of other regions, for example the CF region analysed by Anand *et al.* (1991) and the HD region analysed by Zuo *et al.* (1992). As a result Anand *et al.* (1991) found that it was not always possible to unequivocally orient YACs on the genomic map. However, in this thesis overlapping restriction maps largely confirmed each other, but some discrepancies in the number of restriction sites between individual YACs were observed.

6.1.2.c. Isolation of YAC ends.

The sequences from the left hand arm of 7 YACs were isolated by using plasmid end rescue. The end rescued probes from the distal and proximal ends of the alphoid and 68bp contigs respectively were most informative. Their subsequent analysis quickly revealed that the most proximal YAC in the 68bp contig (15B12F, Figure 3.10) was co-ligated and that the alphoid and the AMGL contig did not overlap (Figure 5.28). Unfortunately, due to the repeated nature of the alphoid contig end-rescued probe (pAA15e, Figure 5.25), this sequence could not be used to walk in YAC libraries using colony hybridisation. The major limitation of using this technique is that the pYAC4 vector was designed to facilitate plasmid end rescue of the left hand arm only and it is dependent on generating a suitably sized *Xho*I or *Nde*I fragment for plasmid propagation. However, now new vectors (McCormik *et al.*, 1989; Traver *et al.*, 1989; Shero *et al.*, 1991) have been designed in which bacterial plasmids can be rescued from both the left and right hand vector arms.

YAC end sequences can also be isolated by subcloning (discussed below), Alu-vector PCR (Nelson *et al.*, 1989), inverse PCR (Silverman *et al.*, 1989) and adaptations of ligation mediated PCR termed the "vectorette method" (Riley *et al.*, 1990). The latter three methods have proved to be fast and quite efficient but problems still arise. For example: the Alu-Vector PCR method requires the presence of an Alu element in the insert within ~1kb of the vector and in the correct orientation; inverse PCR requires purified YAC DNA separated from the bulk of the natural yeast chromosomes and the vectorette system requires digesting each YAC with a range of enzymes to determine which sites are present near the vector-insert junction. In this thesis Alu-vector PCR was used but found not to be useful for generating end clones. However, these techniques they have been successfully used by many workers to isolate YAC ends. For example Zuo *et al.* (1992) isolated 37 YAC ends from 23 different YACs mainly using inverse PCR (32/37 ends), Palmeiri *et al.* (1992) used 8 Alu-vector PCR probes to build up a 20 YAC, 1.2Mb contig surrounding the iduronate sulfatase gene and Anand *et al.* (1991) used the vectorette method to obtain end probes from YACs encompassing the CF locus.

6.1.2.d. Subcloning.

The AMGL and 68bp YAC contigs covered regions of genomic DNA which were not very well understood. Therefore detailed analysis of these regions was achieved by subcloning the YACs into plasmids. This technique for

obtaining subclones was chosen, as opposed to using Alu-PCR (Nelson *et al.*, 1989), as the centromeric region of the human Y chromosome is very repeated. Arrays of tandemly repeated sequences rarely contain Alu sequences and thus would be missed. For example, it seems unlikely that the 5bp, 48bp and 68bp satellite would have been detected using Alu-PCR. The subclones were subsequently located on the YACs to provide new landmarks, and the type of DNA present analysed and sequenced. Furthermore, an end clone from YAC 235F3, the most proximally extending YAC in the AMGL contig, was easily isolated this way.

6.1.3. Problems with YACs.

The above sections have discussed the virtues of using YACs to understand complex genomes. However, during the analysis of YACs from the centromeric region of the human Y chromosome many problems were encountered. These are summarised in Table 6.1.and discussed below.

The major problem with analysing these YACs was that 63% of them were unstable. This instability was exhibited by all of the clones containing alphoid DNA and the 68bp satellite and was expected as Neil *et al.* (1991) had previously shown that clones containing tandemly repeated sequences can be unstable. The alphoid instability was fairly severe, especially in the YACs which contain a high proportion of this sequence (281C5 and 417D3), as the 5.7kb higher order units easily become deleted (see Figure 5.1). In comparison, the 68bp satellite instability was mild (the intensity of hybridisation to the 16.9kb band was less than expected and YAC 29B1C has obvious deletion of this fragment- Figure 3.2), possibly due to the presence of multiple classes of repeats and lack of regular tandemly repeated higher order structure. Instability has also been reported to occur in other YACs. For example, 2/467 YACs from the human Xq24-Xq28 (Abidi *et al.*, 1991); 2/148 from a chromosome 21 enriched library (Bellis *et al.*, 1991); 8/28 YACs from the 2Mb HD YAC contig (Zuo *et al.*, 1992); 1/1 YACs containing the *Drosophila* dodecca satellite (Abad *et al.*, 1992).

Additional deletion events were encountered in a further two of the YACs analysed. YAC BB8G1/4 from the alphoid contig contains an internal deletion of > 100kb (see Figure 5.12) and some clones of YAC 235F3 (AMGL contig) had deleted ~170kb (Figure 5.23). Restriction enzyme analysis of these YACs made these deletion events easy to recognise and thus did not hamper the analysis

Table 6.1. Problems associated with analysing YACs from the centromeric region of the human Y chromosome.

Problem	Number of YACs (19 analysed in total)
Two YACs per cell	2
Co-ligated YACs	4
Rearranged YACs	2
Instability	12
Discrepencies between different YAC clones	2

of the region. However, the structure of YAC BB8G1/4 would have been difficult to interpret if other YACs, (62C12 and 281C5), spanning the same region, had not been identified. The rearrangement events which took place in the vector arms of two of the alphoid contig YACs were difficult to recognise. The reasons for this rearrangement is unknown and has only been reported elsewhere by Zuo *et al.* (1992) and Sleister *et al.* (1992).

Co-ligated YACs or chimeric YACs are clones which contain more than one fragment of insert DNA derived from different parts of the genome co-cloned in the same YAC. They have been reported to occur in 10-30% of YACs in libraries (Schlessinger, 1990) though co-cloning has been measured in the range of 50% for contigs in the CF (Green and Olson, 1990b) and HLA (Bronson *et al.*, 1991) regions. Only 4/18 (21%) of YACs isolated from the centromeric region of the human Y chromosome in this thesis were found to be co-ligated. The co-cloned YACs were identified easily (comparison of restriction enzyme maps and plasmid end rescue) and therefore did not prove to be a serious problem. It has been suggested that either two inserts ligate together before transformation or a recombination event between two incipient YACs at a homologous region occurs during or after transformation.

Another complication of the YAC cloning system encountered during this work was the occurrence of two YACs per cell. 11% of the clones analysed contained more than one YAC per cell. Although it is sometimes possible to

segregate multiple YACs into separate yeast clones, the YACs of interest in these clones could not not be isolated alone, thus making their analysis complex. Similar results have been previously reported in other YAC studies (Schlessinger *et al*, 1991; Zuo *et al.*, 1992; Anand *et al.*, 1991, Abidi *et al.*, 1990) and are thought to be the consequence of the stable uptake of two different YACs during transformation.

6.1.4. Conclusions on the use of YACs to analyse centromeres.

At the outset of this thesis it was not known whether YACs would be suitable for analysing large amounts of DNA. More importantly their suitability for analysing the repeated sequences known to be present at human centromeres (see Table 1.1) was unknown. This thesis has shown that using YACs to understand the structure of the centromeric region of the human Y chromosome produced some complications but in general was highly successful. Other workers have similarly used YACs to determine the structure of other regions of the genome: for example the DMD locus (Monaco *et al.*, 1992), the HD locus (Zuo *et al.*, 1992). This thesis and previous work by Neil *et al.* (1990) and Fisher (1992) has revealed that YACs containing tandemly repeating sequences are often unstable, making their subsequent analysis difficult to interpret. However, Villagelieu and Tyler-Smith (1992) have shown that transforming unstable YACs into a recombination deficient host strain in some cases circumvents this problem. Despite these problems this thesis has shown that cloning in YACs is an efficient and accurate means of determining the structure of a human centromere.

6.2. SEQUENCE ORGANISATION OF THE Y CHROMOSOME CENTROMERE.

6.2.1. Long range map of the Y chromosome centromere.

To fully understand the structure of the human Y chromosome centromere a long range map of the region is needed as well as determining the nature of the sequences present. During the course of this work a long range map of the region was being built (C. Tyler-Smith, personal communication; R. Fisher, 1992; this thesis). The resulting map, which spans over 7Mb, and its relationship to the defined centromeric region is drawn in Figure 6.1. Only the restriction sites for *Mlu*I and *Bss*HII are shown for clarity but in addition sites for *Bsi*WI, *Sfi*I and *Sst*II have been mapped. The location of the five YAC contigs isolated during the course of this work are shown. It can be seen that they provide a fairly good coverage of the short arm centromeric sequences

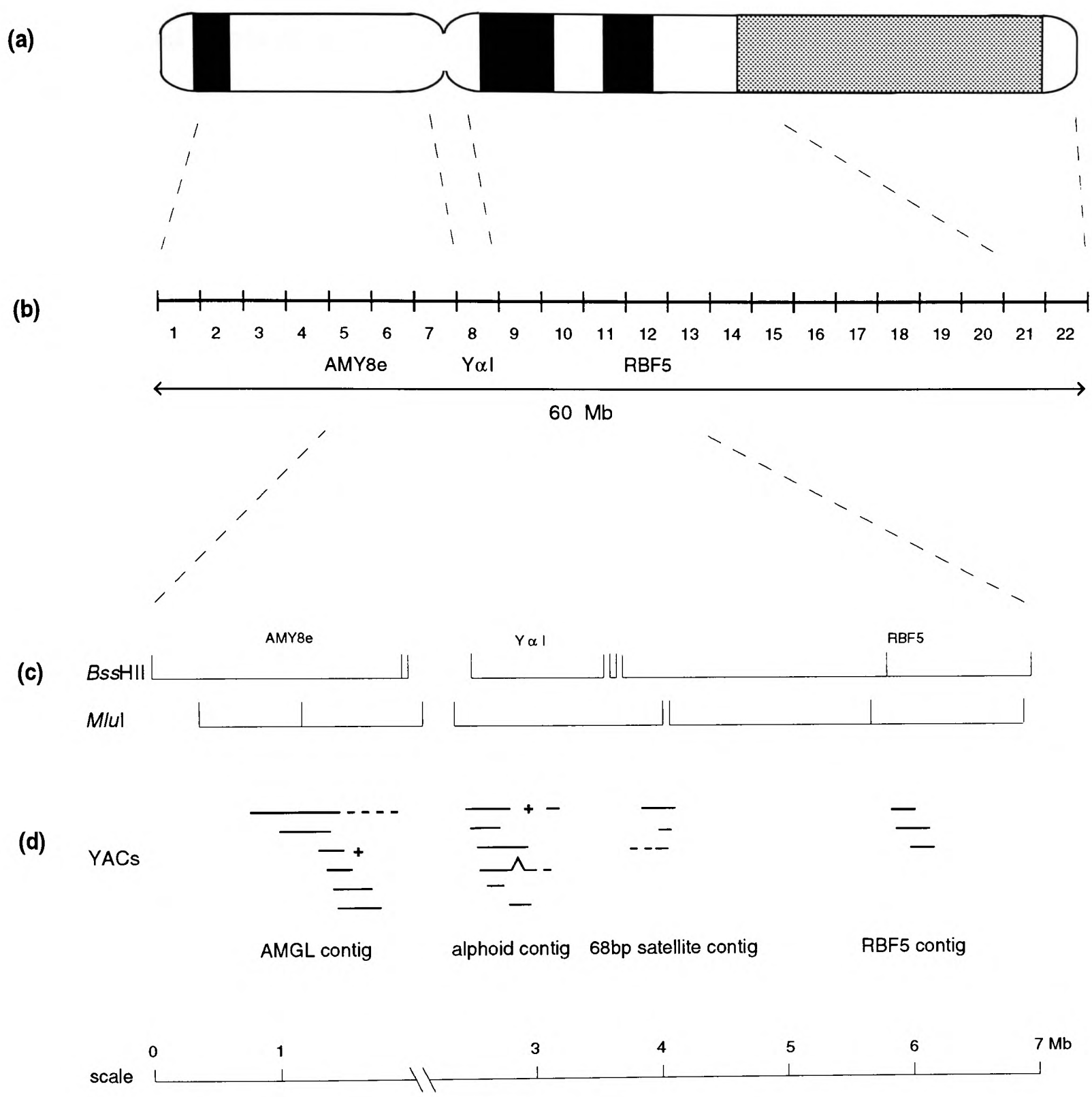


Figure 6.1. Long range restriction site map of the centromeric region of the human Y chromosome showing the location of the YAC contigs. (a) ideogram of the human Y chromosome; (b) deletion intervals showing the location of probes which define the region and Y α I; (c) restriction site map; (d) position of the contigs. The stippled lines represent chimeric YACs, the plus sign many YACs and the arrow head a deleted YAC.

but limited coverage of the long arm. Significantly, no YACs containing the 64a locus and only one YAC spanning the long arm alphoid-nonalphoid boundary have ever been isolated but the possible reasons for this have been discussed above in section 6.1.

The most useful probes isolated during the course of this work to use for building this map have been isolated from YAC 17C12C in the 68bp contig. For example the most proximally located probe in the YAC (pY48) was used to link this region to the major alphoid array and the most distally located probe (pKFC43) to a locus in the next deletion interval (14A3C). However, both these probes detect repeated non-Y chromosome specific loci in male genomic DNA which is typical of the probes located throughout of the whole of the Y centromeric region. This lack of unique sequence probes not only complicates walking in YAC libraries but has also restricted restriction site mapping experiments to Y-only hybrids. Only one out of the three available hybrids showed useful cleavage at rare-cutter restriction sites within this region and even here many bands were faint. Despite these difficulties, as shown in Figure 6.1. and discussed above it was possible to construct a map of the region.

It can also be seen from Figure 6.1 that there is one gap in the long range map of the Y chromosome centromeric region which corresponds to the region where the AMGL and alphoid YAC contigs do not overlap. Much time was spent in trying to find suitable probes from the proximal and distal end of each contig respectively but most of the probes isolated from the alphoid contig (pAA15e, Figure 5.25; many more from cosmids analysed from this region, C. Tyler-Smith, personal communication) were repeated. One probe isolated from the distal end of the alphoid contig (probably 2 Y copies) and probes isolated from the proximal end of the AMGL contig (KFC3 and KFC5.1, Figure 5.27) were less repeated and only detected one band in the Y hybrid 3E7 but when the best of them, KFC3, was used for long range mapping of the region the results were difficult to interpret. However, by analysing rearranged chromosomes several workers eg. Affara *et al.* (1986a), Vergnaud *et al.* (1986), Tyler-Smith *et al.* (1988) have suggested that in some individuals an inversion polymorphism may exist in this region. This could explain why this gap in the genomic map is difficult to close. Further work by C. Tyler-Smith and M. Colomer (personal communication) has shown that a probe isolated from this region also detects a polymorphism on pulsed-field gels in males. Whether the possible inversion event is related to the polymorphism

is difficult to determine because of the repeated nature of the sequences and remains unknown.

The analysis of these YACs has allowed the types of sequences present at the Y chromosome centromere to be determined (discussed below). However, the completion of the long range map of the region in which the centromere lies will require the exact nature of the complications to be understood in detail.

6.2.2. Sequence organisation.

The centromeric region of the human Y chromosome, as defined in section 1.5.2, contains a mixture of different types of sequences. This was summarised diagramatically in Figures 5.28 and 5.29. Starting from distal Yp (and working towards distal Yq) there is a ~400kb region of moderately repeating sequences (52dC, 55H4, LLY56 and 55H4). Approximately 150kb proximal to this lies a region of recent X-Y homology (pDP34 and GMGXY8). The X copy of these sequences have been mapped to Xq21-q22 (Page *et al.*, 1984). This transpostion is described as recent as pDP34 is only found on the long arm of ape X chromosomes and not on ape Y chromosomes (Page *et al.*, 1984). Furthermore, homology between the human X and Y sequences remains high and very few restriction enzyme sites within the region are specific either to the X or the Y chromosome. The extent of the transpostion was estimated by Page *et al.* (1984) to be at least 36kb. Later Affara *et al.* (1986a) demonstrated that GMGXY8 mapped to the same region. This thesis has shown that these two loci lie are contained within ~150kb thereby extending the previously estimated size of the region.

Approximately 150kb proximal to this region of recent X-Y homology lies a region of ancient X-Y homology (AMGL). The X copy is found on the short arm of the chromosome in the p22.1-p22.3 region and sequence comparison of the two loci show major differences suggesting that this region of homology is ancient. In contrast to the non-functional pDP34 region this region encodes a gene, amelogenin, which is expressed during embryonic development from both the X and the Y chromosome (though the Y locus expresses only 10% of that of AMGX) and is necessary for enamel development (Salido *et al.*, 1992). Again the extent of the homology is not known but this thesis has shown that the AMGL region on the Y chromosome is must be smaller than 300kb. It is known that a probe 296e, which is located ~100kb proximal to AMGL is not present on the X

chromosome. Therefore this region of ancient homology could be at least 150kb in length. However, no probes have been found which lie in between AMGL and pDP34. Consequently it is not known whether the recent and ancient regions of X-Y homology lie next to each other on the human Y chromosome or are separated by other sequences.

The region proximal to the block of X-Y homology is characterised by a single copy sequence (p296e) and moderately repeated sequences. Proximal to the gap between the contigs low to moderately repeated sequences predominate (AA15e, Y190 minor, 2D5, 50f2d, YαII5.1 and YαII2.1). Thereafter, through to Yq the region is characterised by a range of tandemly repeated sequences which are discussed in detail below. Distal and intermingled with the pericentric Yq tandem repeats (68bp, 48bp and 5bp satellites) low to moderate copy repeats also are present and this type of sequence probably predominates right to the Yq distal edge of the centromeric interval (defined by probe RBF5 - low copy repeat).

The centromeric region of the human Y chromosome therefore consists of a mosaic of low to moderate repeats and tandem repeats. Also abundant Kpn (LINE-1) and Alu (SINE-1) interspersed repeats occur. The novel repeats range in copy number from two (64a) to several dozen per genome (pKFC37). They either show clear band patterns DNA (KFC11, 37, 43, 52 and YαII 2.1 and 5.1) in genomic DNA. The novel repeats located on the long arm were studied in detail and no evidence for short (<200bp) tandemly repeating units, by limited sequencing, was found. In addition genomic hybridisation experiments did not reveal larger tandemly repeated units. Further experiments are required to elucidate their structures and other genomic locations. Finally single copy sequences are either absent or under-represented in clones from the centromeric region as throughout this analysis only one, p296e, was found (see Figure 5.28). However, even here the possibility of two copies of this locus being present in the genome with the same restriction pattern is not excluded. The tandemly repeated sequences will now be discussed in more detail.

6.2.2.a. The 5bp satellite.

The 5bp satellite was first detected in the human Y chromosome centromeric region by *in situ* hybridisation (Frommer *et al.*, 1987). Later Tyler-Smith, (1987) showed that the 5bp satellite lies on the same size restriction enzyme fragments as alphoid DNA and suggested that a ~400kb block of the satellite is

linked to alphoid DNA on the long arm of the chromosome. This thesis has not only confirmed this linkage (YAC 417D3) but has also shown that a further small (~17kb) array lies about 400kb distal to this on the long arm of the chromosome (YAC 17C12C). In addition, a third Yp pericentric 5bp satellite array has been detected in a slightly more distal position (interval 10, Figure 1.8; C. Tyler-Smith, personal communication). There may be additional arrays on the chromosome, but the 5bp satellite was not detected on the short arm. In the sequenced region of YAC 17C12C, 19/37 (51%) of the 5mers correspond to the consensus GGAAT; an additional nine 5mers (24%) differ at a single position. The sequences of other pericentric 5bp satellite arrays are not known, but the Y long arm DYZ1 sequence has been determined (Nakahori *et al.*, 1986). Here 229/713 (32%) of the 5mers were GGAAT and 297/713 (42%) differed by one base.

6.2.2.b. The 48bp satellite.

The 48bp satellite had previously only been reported to be present in the pericentric region of chromosome 22 (Metzdorf *et al.*, 1988). It probably was not detected before on the Y chromosome due to the small amount present. Analysis of YACs 17C12C (68bp contig, Figure 6.2), 281C5 and 62C12 (alphoid contig) has shown that <21kb is located ~500kb from the edge of the major alphoid block on the long arm and <19kb within ~100kb on the short arm. The Y sequence in the long arm is 88% similar to chromosome 22 consensus sequence. Individual 48bp monomers had between 40 and 43 out of 48 base pairs corresponding to the consensus sequence.

6.2.2.c. The 68bp satellite.

The 68bp satellite was also not previously detected on the Y chromosome, again probably because of the small amount present. Analysis of the three somatic cell hybrids, the Q998.8 derivative and YACs 17C12C and 9C3C has shown that the *Hae*III fragments totalling ~27kb are located in pericentric Yq, ~10kb in distal Yq and ~17kb in an unknown position in Yq. None of this satellite was detected in the Q998.8 derivative and therefore it is not present in Yp. The non-centromeric distal Yq location of the 5bp satellite and the 68bp satellite is striking. One possible explanation is that a recent long arm inversion event has placed these normally pericentric sequences in a non-centromeric position.

6.2.2.d. Alphoid DNA (the 170bp satellite).

Alphoid DNA has been known to be present at or near the centromeres of all human chromosomes for over 10 years (Mannuledis, 1978). Its molecular structure on the Y chromosome was described by Tyler-Smith and Brown in 1987 who analysed 9 cosmid clones derived from the region. Further long range mapping studies (Tyler-Smith, 1987; Oakey and Tyler-Smith, 1990) determined that on the Y chromosome there is one major, uninterrupted, centromeric block of alphoid DNA which varies in size from ~250-1025kb though later work by N. Mathias (personal communication) has shown that the upper size limit is nearer 1400kb. By analysing YACs this thesis has determined the DNA structure of both the long and short arm alphoid-nonalphoid boundaries, the divergence of alphoid DNA at the edges of the block, the orientation of alphoid DNA within the major array as well as positioning a small diverged ~18kb block of alphoid DNA (Yαll) on the long range map.

The types of sequences present at the alphoid-nonalphoid boundaries were as predicted by Tyler-Smith in 1987. Analysis of the YACs derived from the short arm boundary demonstrated that alphoid DNA in this region becomes very diverged from the rigid 5.7kb higher order subunits (see Figure 1.6) and is flanked by typical euchromatic DNA. Conversely analysis of YACs αB3 and 417D3 determined that alphoid DNA at the long arm edge of the array is only diverged to a small amount and is flanked by a Alu element followed by the 5bp satellite .

The results of determining the orientation of alphoid DNA within the array suggested that on the Y chromosome alphoid DNA is arranged in a head-to-tail fashion throughout. Unfortunately the orientation of alphoid DNA on other human chromosomes has not as yet been reported so the significance of this observation is hard to weight. However, the head to tail arrangement of Y chromosome alphoid DNA repeats is unlike the centromeric repeats of *S. pombe* which exhibit a palindromic arrangement (Clarke, 1990). This could possibly suggest that the arrangement of repeat units at centromeres is not important but as the differences between human and *S. pombe* DNA are so vast then this is a speculative conclusion.

6.2.2.e. Overall structure.

The combination of analysis of these YACs and long range mapping of the Y chromosome centromeric region (Tyler-Smith, 1988; Oakey and Tyler-Smith, 1990; this thesis) has allowed the structure of the human Y chromosome centromere to be determined and this is drawn schematically Figure 6.2. It revels a complex organisation with seven blocks of four different satellites interspersed with other repeats. The location of these repeats, which are striking in their numbers, have not been drawn in Figure 6.2 for clarity but their location in the Y chromosome centromeric region was shown in Figure 5.29.

6.3. IMPLICATIONS FOR STRUCTURE AND FUNCTION.

One indicator of functional importance is sequence conservation. Are any of the features seen in the centromeric region of the human Y chromosome (Figure 6.1) shared by other human centromeres or the centromeres of other organisms?

6.3.1. Comparison of the structure of the human Y chromosome centromere with other centromeres.

6.3.1.a. Comparison with human centromeres.

Comparable, detailed molecular maps are not available for other human chromosomes except for an outline of the arrangement of alphoid DNA on chromosome 7 (Wevrick and Willard, 1991). In this chromosome, two centromeric blocks of centromeric alphoid DNA exist which span ~4000kb and are separated by between 0kb and 1Mb of DNA of unknown sequence. Both blocks are variable in size, the larger more centromeric array spanning between 1500-3800kb, the smaller, short arm array 100-550kb. Whether other satellites are present at or near the centromere of this chromosome has not been analysed at the molecular level, although *in situ* hybridisation has suggested the presence of the 5bp and 68bp satellites.

In the absence of other comparable maps the presence or absence of these satellites at other human chromosomes, detected by using *in situ* hybridisation, is the only source of information (Table 1.1). This technique shows that only alphoid DNA is present at or near the centromeres of all human chromosomes, thus making it a prime candidate for determining

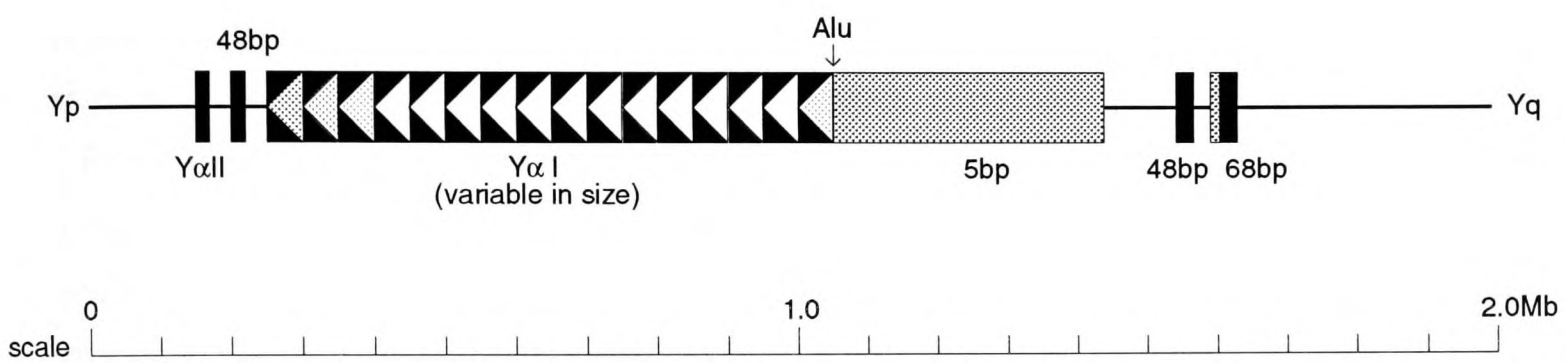


Figure 6.2. The structure of the human Y chromosome centromere.

centromere function. In addition Table 1.1. shows that there is great superficial variation in the distribution of the other satellite sequences on different chromosomes. Nevertheless this apparent variability could mask an underlying similarity. The alphoid and the 5bp satellites are present at the centromeres of most human chromosomes, and this work has demonstrated that the 48bp and the 68bp satellites are more widely distributed than shown in Table 1.1. Therefore it could be possible that in addition to alphoid DNA other satellites and repeated sequences are common to the centromeric regions of all human chromosomes (being present in too small quantities to be detected by *in situ* hybridisation) and could have a role to play in centromere function. For example 4 out of 5 YACs which contained non-Y copies of the low copy repeat YαII 2.1 also contained non-Y alphoid DNA (this thesis; Fisher, 1992) and 4 out of 14 non-Y alphoid YACs isolated from the CEPH library contained the 5bp satellite. However, until similar molecular analysis of other human centromeres is complete these questions about sequence conservation, which in turn lead to questions about the roles (if any) of these individual centromeric satellites and repeats at cell division, will remain unanswered.

As well as alphoid DNA and the 5bp satellite (discussed above) the sequences present in λCM8 (Hadlaczky *et al.*, 1991, section 1.4.4.b) have been considered as candidate sequences which determine centromere function. However, this work has shown (using Southern hybridisation) that they appear to be present at less than the 16-32 copies estimated and are not present on the human Y chromosome. One possible explanation is that λCM8 is a single copy sequence from an unknown location. Its ability to form a centromere after transformation into mouse cells is then puzzling and is discussed further in section 6.3.2.

6.3.1.b. Comparison with the centromeres of other organisms.

The most striking feature of the centromeric regions of other higher eukaryotes is that they all generally contain satellite sequences. Even the centromeric regions of the fission yeast *S. pombe* are made up of repeated sequences. Therefore it seems logical to postulate that repeated sequences, satellite DNA and centromeric function must be related, but until the precise function of different types of centromeric DNA is determined this relationship cannot be understood.

Satellites are not usually conserved between distantly related species. For example, alphoid DNA is only found in primates. However exceptions to this have been shown:- (1) this thesis has shown that the 48bp satellite cross hybridises to mouse DNA under low stringency conditions (Figure 3.2.b). The target sequence in mouse DNA is unknown: the periodicity (~0.13kb) shows that it is unlikely to be the major satellite, but it could be the mouse minor satellite which has a similar periodicity on agarose gels (Pietras *et al.*, 1983). Comparison of the consensus sequences revealed a region with 18/31bp similarity, partially overlapping the CENP-B box in the mouse minor satellite; (2) 16 out of the 17bp recognition sequence for alphoid DNA binding protein CENP-B are found in the mouse minor satellite. Interestingly CENP-B has not been detected on either mouse or human Y chromosomes but the significance of this is not as yet understood; (3) The *Drosophila* dodeca-satellite cross-hybridises to a variant of the human 5bp satellite family and to *Arabidopsis* sequences (Abad *et al.*, 1992). The 5bp satellite has also been reported to be conserved and to be similar to CDEIII sequences of *S. cerevisiae* centromeres (Grady *et al.*, 1992). Furthermore, the similarity (8 out of 9 matches) occurs in the region of CDEIII in which a single base pair change abolishes all mitotic function. However, by using Southern hybridisation the authors could not detect the 5bp satellite hybridising to *S. cerevisiae* DNA. Similarly the repeated sequences, (B, K. L) present at the centromeres of *S. pombe*, do not cross-hybridise to human DNA (section 3.1).

The biological significance of this satellite cross-hybridisation is unclear and difficult to assess. Aa an alternative to the hypothesis that these sequences have been conserved because they have a functional role, it is possible that all the cross-hybridisation events mentioned above could reflect the limited range of oligonucleotides which form satellites.

6.3.2.Analysis of naturally occurring rearranged chromosomes.

A second way of identifying the DNA sequences which are important for centromere function is to analyse mutant centromeres. A few naturally occuring rearranged Y chromosomes have been analysed (C. Tyler-Smith, unpublished obsevations). A monocentric chromosome (Affara *et al.*, 1986a) contains a functional Y chromosome centromere which segregates at mitosis but lacks all long arm sequences. The breakpoint is within the alphoid DNA and the resulting structure is a perfect palindrome - Figure 6.3.b. Twelve dicentric chromosomes have also been analysed. They segregate in a normal

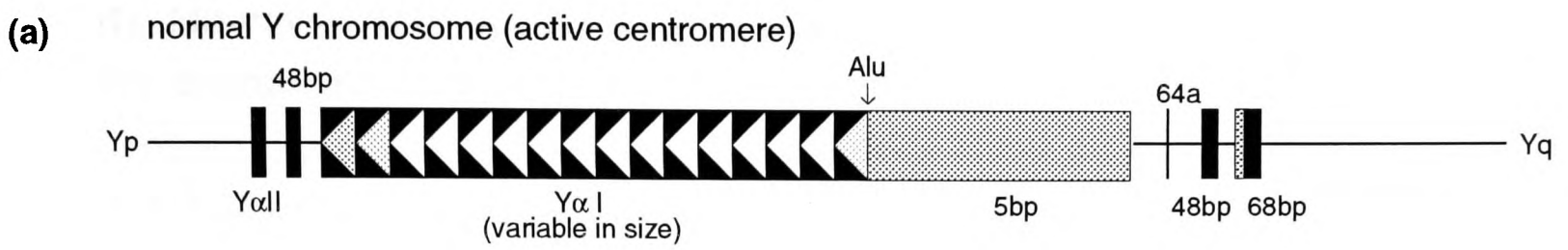


Figure 6.3. Arrangement of tandem repeats found at the centromere of the monocentric isoYp chromosome (b) and at the inactive centromeres of 12 dicentric Y chromosomes (c) - (e). The structure of a normal active Y chromosome is also shown (a). The vertical line drawn through centromeres (b), (d) and (e) represents the breakpoint in alphoid DNA.

fashion because one centromere has been inactivated. The inactive centromere of one (ZAC: Maraschio *et al.*, 1990) shows a deletion of part of the alphoid array and several megabases of adjacent long arm sequences (Figure 6.3.d). However, the inactive centromere of another (PIP: M. Jobling, 1991) shows partial deletion of alphoid DNA without alteration to the flanking sequences (Figure 6.3.e). The other 10 dicentrics show no detectable change to the DNA structure at the inactive centromere (Figure 6.3.c)

These results are quite difficult to interpret in a simple way. The monocentric isochromosome shows that sequences on the long arm side of the alphoid DNA are not required for mitotic centromere function, or can be replaced by short arm sequences. However, in the dicentric isochromosome ZAC deletion of these sequences appears to be responsible for inactivation of the centromere, and in the dicentric isochromosome PIP a deletion entirely within the alphoid DNA appears to inactivate the centromere. In each case the residual block of alphoid DNA still falls within the size range of normal Y alphoid arrays. Thus the size restriction in the alphoid block does not entirely account for centromeric inactivation. Also deletion of long arm sequences only occurs in ZAC so it seems unlikely that this event is responsible for the inactivation of a centromere. Finally, although these results are difficult to interpret in detail they do (like the comparison of the Y chromosome centromere with other centromeres) focus on the region of tandem repeats in the centre of Figure 6.2 as the region most likely to be responsible for centromere function.

6.3.3. Conclusions about the location of the centromere.

The ultimate aim of analysing the stucture of human chromosome centromeres is to determine which sequences are necessary for centromeric function. Therefore the question arises of whether a model for a functional human centromere can be proposed at this stage. Such a model would have to account for the published structural data, the data presented in this thesis and include the known centromere binding proteins.

The possibilities are that either an unknown unique sequence, an unknown repeat or a known repeat could determine centromere function. The first of these possibilities is doubtful for two reasons. Firstly, a chromosome 17 minichromosome (section 1.4.4.a) and a deleted chromosome 17 both function and survive, thereby suggesting that the centromere cannot be a unique sequence since it can be spilt into two functioning segments. Secondly,

structural analysis of the centromeric region of the mouse *Mus musculus* (section 1.3.2) has suggested that it consists of adjacent arrays of telomeric repeats (TTAGGG), telomeric associated repeats, mouse minor satellite and mouse major satellite (see Figure 1.4): it seems unlikely that there is any extensive stretch of unique sequence in this region.

Analysis of the centromeres of the rearranged Y chromosomes (section 6.3.2) and knowledge of the arrangements of the sequences present at normal Y chromosome centromeres (this thesis) strongly suggest that alphoid DNA and/or adjacent short arm region are necessary for centromeric function. However, it is also possible that adjacent long arm sequences are important for meiotic function although it is known that at miosis they can be replaced by short arm sequences (monocentric isoYp data). In the absence of a functional monocentric isoYq chromosome the reverse of this cannot be established. Furthermore, the detailed structural analysis of the human Y chromosome centromere presented in this thesis has established that it is unlikely that on the Y chromosome an unknown repeat is responsible for centromere function.

In summary, all these results can be pooled together to form a very simplistic model of the sequences necessary for a human centromere. This model proposes that alphoid DNA, which is present at the primary constriction of all human chromosomes, binds to the heterochromatin specific protein CENP-B (Cooke *et al.*, 1990; Masumoto *et al.*, 1990; section 1.4.3.a). As this protein has not been detected on the Y chromosome an equivalent unknown protein could bind to the Y alphoid DNA. An adjacent sequence(s) either on one or both sides of alphoid DNA (for example the 5bp, 48bp and 68bp satellites or Yαll 2.1) could bind an additional protein. CENP-C would be an ideal candidate as it has recently been shown to specifically localise to the inner kinetochore plate (Saitoh *et al.*, 1992) and also is only present at the centromeres of active chromosomes (Earnshaw *et al.*, 1989). Furthermore Saitoh *et al.* (1992) suggest that CENP-C is an DNA binding protein and is unlikely to bind to alphoid DNA. In turn, CENP-C alone or with CENP-B could bind the rest of the kinetochore.

This model accounts for most of the data reported to date on the structure of human centromeres. However, it does not explain in detail how two putative DNA binding proteins (CENP-B and C) can bind to other proteins which must make up the kinetochore. Also it does not take into account the sequences

contained in λCM8 (sections 1.4.4.a and 6.3.1.a) which can form a centromere after transfection into mouse cells. This thesis has proposed that λCM8 sequences are not present 16-32 copies per cell as previuosly reported (Hadlaczky *et al.*, 1991) but are probably present as a single copy. To account for this data in the above model it could be possible that λCM8 sequences provide a binding site for the second protein (proposed above to be CENP-C). When it is single copy (before transfection) it does not interfere with centromere function as little CENP-C binds but the amplification of the sequence (which occurs upon transfection into mouse cells) confers centromeric properties to the region as many molecules of the second protein and hence the kinetochore could bind and form a pseudo-centromere.

6.3.4. Future work.

Direct functional assays are need to unequivocally demonstrate which sequences in Figure 6.1 are needed for centromeric function at mitosis and meiosis. The resources of structural information described in this thesis can be used to design such experiments.

At present the only functional assay of human centromeric function in existence is the one described by Hadlaczky *et al.* (1991) in which candidate centromeric sequences can be tested for centromeric function. However, the major disadvantage of this system is that the final structure of the *de novo* centromere is difficult to control, thus making interpretation of the results difficult.

A more controlled method to test candidate sequences for centromeric function would be to introduce YACs into mammalian cells and assay for centromeric function. To do this YACs containing known sequences (for example alphoid DNA, 5bp, 48bp and 68bp satellites) would have to be modified by homologous recombination in yeast to contain mammalian selectable marker genes before being introduced into mammalian cells. As no established assay of centromeric function exists the biological consequences of the anticipated integration of the YAC into the host genome could be investigated (formation of a second primary constriction, anaphase bridge formation, disruption of mitosis). Using this technique sequences immediately adjacent to alphoid DNA on both sides of the block could be functionally assessed. This thesis has identified many YACs which could potentially be very interesting to analyse in this way and at present such experiments are being commenced.

Introducing YACs into mammalian genomes has been done by several different groups. All the groups reported that transfected YACS are not maintained as an separate chromosomes, but become integrated into the host cell lines, and Eliceiri *et al.* (1991) demonstrated that during integration they loose varying amounts of sequences near their ends In particular, Huxley *et al.* (1991) and D'Urso *et al.* (1990) have demonstrated that genes cloned in YACs are functional and capable of restoring defects present in endogenous genes of the transfected cell lines. However, the behaviour of repeated sequences in such experiments may be different. Hence comparison of the structure of YACs before and after fussion would be vital in interpreting the fates of YACs transferred to a human cells.

An extension of this type of integration experiment would be to recombine YACs *in vitro* to construct clones containing DNA spanning the human Y chromosome centromere. These clones could subsequently be introduced into mammalian cells and their potential centromeric function observed. The YAC contigs built up in this work therefore will provide a good basis for these experiments. Homologous recombination of overlapping YACs to reconstruct large regions of genomic DNA in one YAC has already been performed by several groups. For example, Den Dunnen *et al.* (1992) have reconstructed a 2.3Mb YAC containing the human Duchenne muscular dystrophy gene using meiotic recombination. Likewise Green and Olson (1990b) reconstructed the CF gene in a single YAC by using meiotic recombination. Furthermore, Ragoussis *et al.* (1992) have constructed a 650kb single clone by mitotic recombination containing 18 genes from the MHC class II region.

A more elegant method to precisely define the roles of the candidate centromeric sequences would be to construct human artificial chromosomes which upon introduction into human cells would replicate independently of the host chromosomes. To implement these experiments a vector containing human telomeres, an origin of replication and mammalian selectable markers would need to be designed. At present human telomeres (Brown, 1989) and various selectable markers have been cloned but human origins of replications are undefined. Either viral origins could be used or alternatively a ~100kb stretch of euchromatic DNA could be cloned in the vector in the hope that it would contain one of the muliple origins of replication. To define which centromeric sequences are necessary for mitotic function

fragments of centromeric DNA could be cloned into the vector and the segregational properties conferred on the construct analysed.

Finally, telomere directed chromosome breakage experiments (Farr *et al.*, 1991) would be another method to engineer human minichromosome. In this system a linear vector containing human telomeres linked to defined centromeric sequences could be targeted by homologous recombination to specific sequences adjacent to alphoid DNA. The telomeres would cause chromosome breakage at the site of recombination and a functional minichromosome should result. To proceed with this type of experiment a good molecular map of a centromere is needed. This thesis has provided such a map of the human Y chromosome and some of these experiments are now in progress.

6.4. CONCLUSION.

Centromeres are essential for the survival of most eukaryotic cells. Their structure and how this structure relates to function is therefore very important in understanding the movement of chromosomes during cell division. This thesis has described the structure of one eukaryotic centromere, the human Y chromosome centromere. The results presented not only provide a first insight into the detailed stucture of a human centromere but also establish that despite their instability YACs can be used to analyse such regions. As a consequence it is hoped that other human centromeres will be similarly analysed. Finally, this detailed knowledge of the sequences at and surrounding the Y chromosome centromere will provide a solid basis from which functional studies can be started.

REFERENCES

Abad, J. P., Carmena, M., Baars, S., Saunders, R. D., Glover, D. M., Ludena, P., Sentis, C., Tyler-Smith, C. and Villasante, A. (1992). Dodeca satellite: A conserved G+C-rich satellite from the centromeric heterochromatin of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA.* **89:** 4663-4667

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

Affara, N. A., Fergusson-Smith, M. A., Tolmie, J., Kwok, K., Mitchell, M., Jamieson, D., A.Cooke and Florentin, L. (1986b). Variable transfer of Y-specific sequences in XX males. *Nucl.Acids Res.* **14:** 5375-5387

Affara, N. A., Floentin, L., Morrison, N., Kwok, K., Mitchell, M., Cook, A., Jamieson, D., Glasgow, L., Meredith, L., Boyd, E. and Ferguson-Smith, M. A. (1986a). Regional assignment of Y-linked DNA by deletion mapping and their homology with X-chromosome and autosomal sequences. *Nucl. Acids Res.* **14:** 5353-5373

Agresti, A., Rainaldi, G., Lobbiani, A., Magnani, I., Lernia, R. D., Meneveri, R., Siccardi, A. G. and Ginelli, E. (1987). Chromosomal location by in situ hybridisation of the human Sau3A family of DNA repeats. *Hum. Genet.* **75:** 326-332

Agresti, A., Meneveri, R., Siccardi, A. G., Marozzi, A., Corneo, G., Gaudi, S. and Ginelli, E. (1989). Linkage in human heterochromatin between highly divergent *Sau* 3A repeats and a new family of repeated DNA sequences (*Hae*III family). *J. Mol. Biol.* **205:** 625-631

Albertsen, H. A., Abderrahim, H., Cann, H. M., Dausset, J., Paslier, D. L. and Cohen, D. (1990). Constuction and characterisation of a yeast artificial chromosome library containing seven haploid genome equivalents. *Proc. Natl. Acad. Sci. USA.* **87:** 4256-4260

Alfageme, C. R., Rudkin, G. T. and Cohen, L. H. (1980). Isolation, properties and cellular distribution of D1, a chromosomal protein of *Drosophila*. *Chromosoma.* **78:** 1-31

Anand, R. (1986). Pulsed field gel electrophoresis: a technique for fractionating large DNA molecules. *Trends Genet.* **2:** 278-283

Anand, R., Villasante, A. and Tyler-Smith, C. (1989). Construction of yeast artificial chromosome libraries with large inserts using fractionation by pulsed-field gel electrophoresis. *Nucl. Acids Res.* **17:** 3425-3433

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

Anand, R. (1992) Editor. " Techniques for the analysis of complex genomes". Academic Press, London

Anderson, L. B., Tommerup, N. and Koch, J. (1990). Formation of a minchromosome by excision of the proximal region of 17q in a patient with von Recklinghausen neurofibromatosis. *Cytogenet. Cell Genet.* **53:** 206-210

Baker, R. E., Fitzgerald-Hayes, M. and O'Brien, T. C. (1989). Purification of yeast centromere binding protein CPI and a mutational analysis of its binding site. *J. Biol. Chem.* **264:** 10843-10850

Balczon, R. D. and Brinkley, B. R. (1987). Tubulin interaction with kinetochore proteins: analysis by in vitro assembly and chemical cross-linking. *J. Cell Biol.* **105:** 855-862

Bellis, M., Gérard, A., Charlieu, J. P., Marçais, B., Brun, M. E., Viegas-Péquignot, E., Carter, D. A. and Roizès, G. (1991). Construction and characterization of a partial library of yeast artificial chromosomes from human chromosome 21. *DNA and Cell Biology.* **10:** 301-310

Bentley, D. R., Todd, C., Collins, J., Holland, J., Dunham, I., Hassock, S., Bankier, A. and Giannelli, F. (1992). The development and application of automated griding for efficient screening of yeast and bacterial ordered libraries. *Genomics.* **12:** 534-541

Bernat, R. L., Delannoy, M. R., Rothfield, N. F. and Earnshaw, W. C. (1991). Disruption of centromere assembly during interphase inhibits kinetochore morphogenesis and functions in mitosis. *Cell.* **66:** 1229-1238

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

Birnboim, H. C. and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucl. Acids Res.* **7:** 1513-1523

Bischoff, F. R., Maier, G., Tilz, G. and Ponstingl, H. (1990). A 47-kDa human nuclear protein recognized by antikinetochore autoimmune sera is homologous with the protein encoded by *RCC1,* a gene implicated in onset of chromosome condensation. *Proc. Natl. Acad. Sci. USA.* **87:** 8617-8621

Block, K., Ising, G. and Ståhl, F. (1990) Minichromosome in *Drosophila melanogaster* derived from the transposing element *TE1. Chromosoma.* **99:** 336-343

Bloom, K. S. and Carbon, J. (1982). Yeast centromere DNA is in a unique and highly ordered structure in chromosomes and small circular minichromosomes. *Cell.* **29:** 305-317

Bloom, K. S., Amaya, E., Carbon, S., Hill, A. and Yea, E. (1984). Chromatin configuration of yeast centromeres. *J. Cell Biol.* **99:** 1559-1568

Bostock, C. J. and Summer, A. T. (1978). The eukaryotic chromosome. Elsevier/ North Holland Biomedical Press. Amsterdam

Bram, R. J. and Kornberg, R. D. (1987). Isolation of a *Saccharomyces cerevisiae* DNA-binding protein, its human homolog, and its possible role as a transcription factor. *Mol. Cell Biol.* **7:** 403-409

Brinkley, B. R. and Stubblefield, E. (1966). The fine structure of the kinetochore of a mammalian cell *in vitro*. *Chromosoma.* **19:** 28-43

Brinkley, B. R., Valdivia, M. M., Tousson, A. and Balczon, R. D. (1991) The Kinetochore: Structure and molecular organization. In "Mitosis: Molecules and Mechanisims". (ed. J. S. Hymans and B. R. Brinkley) Academic Press, London. pp. 77-118.

Brinkley, B. R., Ouspenski, I. and Zinkowski, R. P. (1992). Structure and molecular organization of the centromere-kinetochore complex. *Trends Cell Biol.* **2:** 15-21

Bronson, S. K., Pei, J., Taillon-Miller, P., Chorney, M. J., Deraghty, D. E. and Chaplin, D. D. (1991). Isolation and characterization of yeast artificial chromosomes linking the HLA-B and HLA-C loci. *Proc. Natl. Acad. Sci. USA.* **88:** 1676-1680

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

Brown, W. R. A. (1989). Cloning of human telomeres by comlpementation in yeast. *Nature.* **338:** 774-776

Brownstein, B. H., Silverman, G. A., Little, R. D., Burke, D. T., Korsmeyer, S. J., Schlessinger, D. and Olson, M. V. (1989). Isolation of single-copy human genes from a library of yeast artificial chromosome clones. *Science.* **244:** 1348-1351

Burk, R. D., Ma, P. and Smith, K. D. (1985). Characterization and evolution of a single-copy sequence from the human Y chromosome. *Mol. Cell Biol.* **5:** 576-581

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

Cai, M. and Davis, R. W. (1989). Purification of a yeast centromere-binding protein that is able to distinguish single base-pair muations in its recognition site. *Mol. Cell Biol.* **9:** 2544-2550

Cai, M. and Davis, R. W. (1990). Yeast centromere binding protein CBF1, of the helix-loop-helix protein family, is required for chromosome stability and methionine prototrophy. *Cell.* **61:** 437-446

Carine, K., Jacquemin-Sablon, A., Waltzer, E., Mascarello, J. and Scheffler, I. E. (1989). Molecular characterisation of human minichromosomes with centromere from chromosome 1 in human-hamster hybrid cells. *Somat. Cell Mol. Genet.* **15:** 445-460

Carle, G. F. and Olsen, M. V. (1985). An electrophorectic karyotype for yeast. *Proc. Natl. Acad. Sci. USA.* **82:** 3756-3760

Cherry, L. M. and Johnston, D. A. (1987). Size variation in kinetochores of human chromosomes. *Hum. Genet.* **75:** 155-158

Chikashige, Y., Kinoshita, N., Nakaseko, Y., Masumoto, T., Murakami, S., Niwa, O. and Yanagida, M. (1989). Composite motifs and repeat symmetry in *S. pombe* centromeres: direct analysis by integration of *Not*I restriction sites. *Cell.* **57:** 739-751

Choo, K. H., Earle, E. and McQuillan, C. (1990). A homologous subfamily of satellite III DNA on human chromosomes 14 and 22. *Nucl. Acids Res.* **18:** 5641-5648

Choo, K. H., Vissel, B., Nagy, A., Earle, E. and Kalitsis, P. (1991). A survey of the genomic distribution of alpha satellite DNA on all the human chromosomes, and derivation of a new consensus sequence. *Nucl. Acids Res.* **19:** 1179-1182

Clarke, L. and Carbon, J. (1980). Isolation of a yeast centromere and construction of functional small circular chromosomes. *Nature.* 504-509

Clarke, L., Amstutz, H., Fishel, B. and Carbon, J. (1986). Analysis of centromeric DNA in the fission yeast *Schizosaccharomyces pombe. Proc. Natl. Acad. Sci. USA.* **83:** 8253-8257

Clarke, L. (1990). Centromeres of budding and fission yeasts. *Trends Genet.* **6:** 150-154

Clarke, L. and Baum, M. P. (1990). Functional analysis of a centromere from fission yeast: A role for centromere-specfic repeated DNA sequences. *Mol. Cell Biol.* **10:** 1863-1872

Cooke, C. A., Heck, M. S. and Earnshaw, W. C. (1987). The inner centromere protein (INCENP) antigens: Movement from inner centromere to midbody during mitosis. *J. Cell Biol.* **105:** 2053-2067

Cooke, C. A., Bernat, R. L. and Earnshaw, W. C. (1990). CENP-B: a major human centromere protein located beneath the kinetochore. *J. Cell Biol.* **110:** 1475-1488

Cooke, H. J. and Hindley, J. (1979). Cloning of human satellite III DNA: different components are on different chromosomes. *Nucl. Acids Res.* **6:** 3177-3197

Cooper, K. F., Fisher, R. B. and Tyler-Smith, C. (1992). Structural analysis of the pericentric long arm region of the human Y chromosome. *J. Mol. Biol.* "in press"

Corneo, G. (1967). A satellite DNA isolated from human tissues. *J. Mol. Biol.* **23:** 619-622

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

Cumberledge, S. and Carbon, J. (1987). Mutational analysis of meiotic and mitotic centromere function *Saccharomyces cerevisiae. Genetics.* **117:** 203-212

Deininger, P. L., Jolly, D. J., Rubin, C. M., Freidmann, T. and Schmid, C. W. (1981). Base sequence studies of 300 nucleotide renatured repeated human DNA clones. *J. Mol. Biol.* **151:** 17-33

Devereux, J., Haerberli, P. and Smithies, O. (1984). A comprehensive set of sequence analysis programmes for the vax. *Nucl. Acids Res.* **12:** 387-395

Devine, E. A., Nolin, S. L., Houck, G. E., Jenkins, E. C. and Brown, W. T. (1985). Chromosomal localization of several families of repetitive sequences by in situ hybridization. *Am. J. Hum. Genet.* **37:** 114-124

Disney, J., Johnson, K. R., Magnuson, N. S., Sylvester, S. R. and Reeves, R. (1989). High-mobility group protein HGM-I localizes to G/Q- and C-bands of human and nouse chromosomes. *J. Cell Biol.* **109:** 1975-1982

Dower, W. J., Miller, J. F. and Ragsdale, C. W. (1988). High efficiency transformation of *E. coli* by high voltage electroporation. *Nucl. Acids Res.* **16:** 6127-6145

D'Urso, M., Zucchi, I., Ciccodicola, A., Palmieri, G., Abidi, F. E. and Schlessinger, D. (1990). Human glugose-6-phosphate dehydrogenase gene carried on a yeast artificial chromosome encodes active enzyme in monkey cells. *Genomics.* **7:** 531-534

Earnshaw, W. C. and Rothfield, N. (1985). Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. *Chromosoma.* **91:** 313-321

Earnshaw, W. C. (1987). Anionic regions in nuclear proteins. *J. Cell Biol.* **105:** 1479-1482

Earnshaw, W. C., Halligan, N., Cooke, C. and Rothfield, N. (1987). Molecular cloning of a cDNA for CENP-B, the major centromere autoantigen. *J. Cell Biol.* **104:** 817-829

Earnshaw, W. C., Ratrie, H. and Stetten, G. (1989). Visualization of centromeric proteins CENP-B and CENP-C on a stable dicentric chromosome in cytological spreads. *Chromosoma.* **98:** 1-12

Emrie, P. A., Jones, C., Hofmann, T. and Fisher, J. H. (1988). The coding sequence for the human 18,000-dalton hydrophobic pulmonary surfactant protein is located on chromosome 2 and identifies a restriction fragment length polymorphism. *Som. Cell & Molec. Genet.* **14:** 105-110

Eliceiri, B., Labella, T., Hagino, Y., Srivastavia, A., Schlessinger, D., Pilia, G., Palmieri, G. and D'Urso, M. (1991). Stable integration and expression in mouse of yeast artificial chropmosomes harboring human genes. *Proc. Natl. Acad. Sci. USA.* **88:** 2179-2183

Endow, S. A., Heinkoff, S. and Soler-Niedziela, L. (1990). Mediation of meiotic and early mitotic chromosome segrgation in *Drosophila* by a protein related to kinesin. *Nature.* **345:** 81-83

Evans, E. P. (1981). Karyotype of the house mouse. *Symp. Zool. Soc. London.* **47:** 127-139

Farr, C., Fantes, J., Goddfellow, P and Cooke, H. (1991) Functional reintroduction of human telomeres into mammalian cells. *Proc. Natl. Aca. Sci. USA.* **88:** 7006-7010

Feinberg, A. P. and Vogelstein, B. (1984). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **137:** 266-267

Feingold, J. M., Ogden, S. D. and Denny, C. T. (1990). Streamlined approach for creating yeast artificial chromosome libraries from specialized sources. *Proc. Natl. Acad. Sci. USA.* **87:** 8637-8641

Fishel, B. H., Amstutz, H., Baum, M., Carbon, J. and Clarke, L. (1988). Structural organisation and functional analysis of centromeric DNA in the fission yeast *Schizosaccharomyces pombe. Mol. Cell Biol.* **8:** 754-763

Fisher, R. B. (1992). Molecular studies of the human Y chromosome. D. Phil Thesis. Dept. of Biochemistry. Oxford

Frommer, M., Paul, C. and Vincent, P. C. (1988). Localisation of satellite DNA sequences on human metaphase chromosomes using bromodeoxyuridine-labelled probes. *Chromosoma.* **97:** 11-18

Garza, D., Ajioka, J. W., Burke, D. T. and Hartl, D. L. (1989). Mapping the Drosophila genome with yeast artificial chromosomes. *Science.* **246:** 641-646

Gaudet, A. and Fitzgerald, M. (1987). Alterations in the adenine-plus-thymine-rich region of *CEN3* affect centromere function in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **7:** 68-75

Glasser, T., Houseman, D., Lewis, W. H., Gerhard, D. and Jones, C. (1989). A fine-structure deletion map of human chromosome 11p: analysis of J1 series hybrids. *Somat. Cell & Mol. Genet.* **15:** 477-501

Gosden, J. R., Mitchell, A. R., Buckland, R. A., Clayton, R. P. and Evans, H. J. (1975). The location of four human satellite DNAs on human chromosomes. *Exp. Cell Res.* **92:** 148-158

Grady, D. L., Ratliff, R. L., Robinson, D. L., McCanlies, E. C., Meyne, J. and Moyzis, R. K. (1992). Highly conserved repetitive DNA sequences are present at human centromeres. *Proc. Natl. Acad. Sci. USA.* **89:** 1695-1699

Graham, G. J., Hall, T. J. and Cummings, M. R. (1984). Isolation of repetitive DNA sequences from human chromosome 21. *Am. J. Hum. Genet.* **36:** 25-35

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. U. S. A.* **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

Guellaen, G., Casanova, M., Bishop, C., Geldwerth, D., Andre, G., Fellous, M. and Weissenbach, J. (1984). Human XX males with Y single-copy DNA fragments. *Nature.* **307:** 172-173

Haaf, T., Summer, A. T., Köhler, J., Willard, H. F. and Schmid, M. (1992). A microchromosome derived from chromosome 11 in a patient with the CREST syndrone of scleroderma. *Cyto. Genet. Cell Genet.* **60:** 12-17

Hadlaczky, G., Praznovszky, T., Cserpán, I., Keresö, J., Péterfy, M., Kelemen, I., Atalay, E., Szeles, A., Szelei, J., Tubak, V. and Burg, K. (1991). Centromere formation in mouse cells cotransformed with human DNA and a dominant marker gene. *Proc. Natl. Acad. Sci. USA.* **88:** 8106-8110

Hahnenberger, K., Baum, M. P., Polizzi, K., Carbon, J. and Clarke, L. (1989). Construction of functional artificial minichromosomes in the fission yeast *Schizosaccharomyces pombe*. *Proc. Natl. Acad. Sci. USA.* **86:** 577-581

Hahnenberger, K. M., Carbon, J. and Clarke, L. (1991). Identification of DNA regions required for mitotic and meiotic functions within the centromere of *Schizosacchromyces pombe* chromosome I. *Mol. Cell Biol.* **11:** 2206-2215

Hays, T. S. and Salmon, E. D. (1990). Poleward force at the kinetochore in metaphse depends on the number of kinetochore microtubules. *J. Cell Biol.* **110:** 391-404

Hedgemann, J. H., Shero, J. H., Cottarel, G., Philippsen, P. and Hieter, P. (1988). Mutational analysis of the centromere DNA from chromosome VI of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **8:** 2523-2535

Heus, J., Zonneveld, B. J. N., Steensma, H. Y. and Berg, J. A. V. d. (1990). Centromeric DNA of *Kluyveromyces lactis*. *Curr. Genet.* **18:** 517-522

Higgins, M. J., Wang, H., Shtromas, I., Haliotis, T., Roder, J. C., Holden, J. J. A. and White, B. N. (1985). Organization of a repetitive 1.8kb KpnI sequence localized in the heterochromatin of chromosome 15. *Chromosoma.* **93:** 77-86

Hilliker, A. J., Appels, R. and Schalet, A. (1980). Genetic analysis of *D. melanogaster* heterochromatin. *Cell.* **21:** 607-619

Hörz, W. and Altenburger, W. (1981). Nucleotide sequence of mouse satellite DNA. *Nucl. Acids Res.* **9:** 683-696

Huberman, J. A., Pridmore, R. D., Jager, D., Zonneveld, B. and Philippsen, P. (1986). Centromeric DNA from *Saccharomyces uvarum* is functional in *Saccharomyces cerevisiae*. *Chromosoma.* **94:** 162-168

Huxley, C., Hagino, Y., Schlessinger, D. and Olson, M. V. (1991). The human HPRT gene on a yeast artificial chromosome is functional when transferred to mouse cells by fusion. *Genomics.* **9:** 742-750

Human Gene Mapping 11. (1991). *Cytogent. Cell Genet.* **58:** 967-984

Hyams, J. S. and Brinkley, B. R. (1989) Editors. "Mitosis: Molecules and Mecahnisms". Academic Press, London.

Hyman, A. A. and Mitchison, T. J. (1990). Modulation of microtubule stability by kinetochores in vitro. *J. Cell Biol.* **110:** 1601-1616

Hyman, A. A. and Mitchison, T. J. (1992). Two different microtubule-based motor activities with opposite polarities in kinetochores. *Nature.* **351:** 206-211

James, T. C., Eissenbach, J. C., Craig, C., Dietrich, C., Hobson, A. and Elgin, S. C. R. (1989). Distribution patterns of HP1, a heterochromatin associated, non-histone chromosomal protein of *Drosophila*. *Eur. J. Cell Biol.* **50:** 170-180

Jiang, W. and Philippsen, P. (1989). Purification of a protein binding to the CDEI subregion of *Saccharomyces cerevisiae* centromere DNA. *Mol. Cell Biol.* **9:** 5585-5593

Jobling, M. (1991). Molecular studies of the human Y chromosome. D. Phil Thesis. Dept. of Genetics. Oxford.

Jones, C. P., Janson, M. and Nordenskjöld, M. (1989). Separation of yeast chromosomes in the megabase range suitable as size markers for pulsed-field gel electrophoresis. *Technique.* **1:** 90-95

Jones, K. W. (1970). Chromosomal and nuclear location of mouse satellite DNA in individual cells. *Nature.* **225:** 912-915

Jorgensen, A. L., Bostock, C. J. and Bak, A. L. (1987). Homologous subfamilies of human alphoid repetitive DNA on different nucleolus organizing chromosomes. *Proc. Natl. Acad. Sci. USA.* **84:** 1075-1079

Joseph, A., Mitchell, A. R. and Miller, O. J. (1989). The organization of mouse satellite DNA at centromeres. *Exp. Cell Res.* **183:** 494-500

Kingsbury, J. and Koshland, D. (1991). Centromere-dependent binding of yeast minichromosome to microtubules in vitro. *Cell.* **66:** 483-495

Kipling, D., Ackford, H. E., Taylor, B. A. and Cooke, H. J. (1991). Mouse minor satellite DNA genetically maps to the centromere and is physically linked to the proximal telomere. *Genomics.* **11:** 235-241

Koshland, D., Mitchison, T. J. and Kirschner, M. W. (1988). Chromosome movement driven by microtubule depolymerization *in vitro*. *Nature.* **311:** 499-504

Kuhn, R. M., Clarke, L. and Carbon, J. (1991). Clustered rRNA genes in *Schizosaccahromyces pombe* centromeric DNA sequence repeats. *Proc. Natl. Acad. Sci. USA.* **88:** 1306-1310

Larin, Z., Monaco, A. P. and Lehrach, H. (1991). Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. *Proc. Natl. Acad. Sci. USA.* **88:** 4123-4127

Lau, E. C., Mohandas, T. K., Shapiro, L. J., Slavkin, H. C. and Snead, M. L. (1989). Human and mouse amelogenin gene loci are on the sex chromosomes. *Genomics.* **4:** 162-168

Lechner, J. and Carbon, J. (1991). A 240kd multisubunit protein complex, CBF3, is a major component of the budding yeast centromere. *Cell.* **64:** 717-725

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 sequnece polymerase chain reaction. *Genomics.* **6:** 475-481

Lica, L. M., Narayanswami, S. and Hamkalo, B. A. (1986). Mouse satellite DNA, centromere structure and sister chromatid pairing. *J. Cell Biol.* **103:** 1145-1151

Lindsley, D. and Grell, E. H. (1968). In "Genetic variations of *Drosophila melanogaster*". Carneigie Inst., Washington.*Publ.* **No. 627:** p. 316

Little, P. F. R. (1987) Choice and use of cosmid vectors. In " DNA cloning, Volume III". (ed. D. M. Glover) IRL Press, Oxford. pp. 19-42

Lohe, A. R. and Brutlag, D. L. (1986). Multiplicity of satellite DNA sequences in *Drosophila melanogaster. Proc. Natl. Acad. Sci. USA.* **83:** 696-700

Luykx, P. (1970). Cellular mechanisms of chromosome distribution. *Int. Rev. Cytol. Suppl.* **2:** 1-173

Maio, J. (1971). DNA strand reassociation and polyribonucleotide binding in the Afican green monkey, *Cercopithecus aethiops. J. Mol. Biol.* **56:** 579-595

Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982). Molecular Cloning. A Laboratory Manual. Cold Spring Harbour Press, New York

Mannuelidis, L. (1978). Chromosomal location of complex and simple repeated human DNAs. *Chromosoma.* **66:** 23-32

Maraschio, P., Zuffardi, O., Caiulo, A., Daniotti, E., Piantanida, M., Rivera, H. and Tupler, R. (1990). Deletion of specific sequences or modification of centromeric chromatin are responsible for Y chromosome centromere inactivation. *Hum. Genet.* **85:** 491-494

Marcus, M., Tantravahi, R., Dev, V. G., Miller, D. A. and Miller, O. J. (1976). Human-mouse cell hybrid with human multiple Y chromosomes. *Nature.* **262:** 63-65

Masumoto, H., Masukata, H., Muro, Y., Nozaki, N. and Okazaki, T. (1989). A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. *J. Cell Biol.* **109:** 1963-1973

McCormick, M. K., Shero, J. H., Cheung, M. C., Kan, T. W., Hieter, P. A. and Antonarakis, S. E. (1989). Construction of a human chromosome-21 specific yeast artificial chromosome library. *Proc. Natl. Acad. Sci. USA.* **86:** 9991-9995

McGrew, J., Diehl, B. and Fitzgerald-Hayes, M. (1986). Single base-pair mutations in centromere elemnet III cause aberrant chromosomal segregation in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **6:** 530-538

Meneveri, R., Agresti, A., Valle, G. D., Talarico, D., Siccardi, A. G. and Ginelli, E. (1985). Identification of a human clustered G + C-rich DNA family of repeats (*Sau* 3A family). *J. Mol. Biol.* **186:** 483-489

Metzdorf, R., Göttert, E. and Blin, N. (1988). A novel centromeric repetitive DNA from chromosome 22. *Chromosoma.* **97:** 154-158

Metzdorf, P., Herzog, R., Göttert, E. and Blin, N. (1990). Isolation and characterisation of DNA sequences from flow-sorted human chromosome 22 libraries. *Hum Hered.* **40:** 145-152

Mitchell, A. R., Gosden, J. R. and Miller, D. A. (1985). A cloned sequence, p82H, of the alphoid repeated DNA family found at the centromeres of all human chromosomes. *Chromosoma.* **92:** 369-377

Mitchison, T. J. and Kirschner, M. W. (1985). Properties of the kinetochore *in vitro* I. Microtubule nucleation and tubulin binding. *J. Cell Biol.* **101:** 755-765

Mitchison, T. J., Schultze, E., Evans, I. and Kirschner, M. W. (1986). Sites of microtubule assembly and disassembly in the mitotic spindle. *Cell.* **45:** 515-527

Monaco, A. P., Müller, U., Larin, Z., Meier-Ewert, S. and Lehrach, H. (1991). Isolation of the human sex determining region from a Y-enriched yeast artificial chromosome library. *Genomics.* **11:** 1049-1053

Monaco, A. P., Walker, A. P., Millwood, I., Larin, Z. and Lehrach, H. (1992). A yeast artificial chromosome contig containing the complete Duchenne muscular dystrophy locus. *Genomics.* **12:** 465-475

Moroi, Y., Peebles, C., Friztler, M. J., Steigerwald, J. and Tan, E. M. (1980). Autoantibody to centromere (kinetochore) in sceleroderma sera. *Proc. Natl. Acad. Sci.* **77:** 1627-1631

Müller, U., Donlon, T., Schmid, M., Fitch, N., Richer, C.-L., Lalande, M. and Latt, S. A. (1986). Deltion mapping of the testis determining locus with DNA probes in 46,XX males and in 46, XY and 46,X,dic(Y) females. *Nucl. Acids Res.* **14:** 6489-6505

Münke, M., Page, D. C., Brown, L. G., Armson, B. A., Zackai, E. H., Mennuti, M. T. and Emanuel, B. S. (1988). Molecular detection of a Yp/18 translocation in a 45,X holoprosencepalic male. *Hum Genet.* **80:** 219-223

Nakahori, Y., Mitani, K., Yamada, M. and Nakagome, Y. (1986). A human Y-chromosome specfic repeated DNA family (DYZ1) consists of a tandem array of pentanucleotides. *Nucl. Acids Res.* **14:** 7569-7580

Nakahori, Y., Takenaka, O. and Nakagome, Y. (1991). A human X-Y homologous region encodes "amelogenin". *Genomics.* **9:** 264-269

Nakaseko, Y., Adachi, Y., Funahashi, S., Niwa, O. and Yanaguda, M. (1986). Chromosome walking shows a highly repetitive sequence in all the centromere regions of fission yeast. *EMBO J.* **5:** 1011-1021

Neil, D. L., Villasante, A., Fisher, R. B., Vetrie, D., Cox, B. and Tyler-Smith, C. (1990). Structural instability of human tandemly repeated DNA sequences cloned in yeast artificial chromosomes. *Nucl. Acids Res.* **18:** 1421-1428

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

Newlon, C. S. (1988). Yeast chromosome replication and segregation. *Microbiol. Rev.* **52:** 568-601

Ng, R. and Carbon, J. (1987). Mutational and in vitro protein-binding studieson centromere DNA from *Saccharomyces cerevisiae. Mol. Cell Biol.* **7:** 4522-4534

Nicklas, R. B. (1989). The motor for poleward chromosome movement in anaphase is in or near the kinetochore. *J. Cell Biol.* **109:** 2245-2255

Niwa, O. and Yanagida, M. (1986). Construction of a mini-chromosome by deletion and its mitotic and meiotic behaviour in fission yeast. *Mol. Gen. Genet.* **203:** 397-405

Niwa, O., Matsumoto, T., Chikashige, Y. and Yanagida, M. (1989). Characterisation of *Schizosaccchromyces pombe* minichromosome deletion derivitives and a functional allocation of their centromere. *EMBO J.* **8:** 3045-3052

Nizetic, D., Zehetner, G., Monaco, A. P., Gellen, L., Young, B. D. and Lehrach, H. (1991). Construction, arraying and high density display of large insert libraries of the human chromosomes X and 21: their potential use as reference libraries. *Proc. Natl. Acad. Sci. USA.* **88:** 3233-3237

Nurse, P. (1985). Cell cycle control genes in yeast. *Trends Genet.* **3:** 51-55

Oakey, R. and Tyler-Smith, C. (1990). Y chromosome DNA haplotyping suggests that most European and Asian men are descended from one of two males. *Genomics.* **7:** 325-330

Olson, M. V., Hood, L., Cantor, C. and Botstein, D. (1990). A common language for physical mapping of the human genome. *Science.* **245:** 1434-1435

Page, D. C., Harper, M. E., Love, J. and Botstein, D. (1984). Occurrence of a transposition from the X-chromosome long arm to the Y-chromosome short arm during human evolution. *Nature.* **311:** 119-123

Palmer, D. K., O'Day, K., Trong, H. L., Charbonneau, H. and Margolis, R. L. (1991). Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. *Proc. Natl. Acad. Sci. USA.* **88:**3734-3738

Palmieri, G., Capra, V., Romano, G., D'Urso, M., Johnson, S., Schlessinger, D., Morris, P., Hopwood, J., Natale, P. D., Gatti, R. and Ballabio, A. (1992). The iduronate sulfatase gene: Isolation of a 1.2Mb YAC contig spanning the entire gene and identification of heterogeneous deletions in patients with Hunter Syndone. *Genomics.* **12:** 52-57

Pardue, M. L. and Gall, J. G. (1970). Chromosomal localization of mouse satellite DNA. *Science.* **168:** 1336-1338

Pardue, M. L., Lowenhaupt, K., Rich, A. and Nordheim, A. (1987). $(dC\text{-}dA)_n$ and $(dG\text{-}dT)_n$ sequences have evolutionarily consereved chromosomal locations in Drosophila with implications for roles in structure and function. *EMBO. J.* **6:** 1781-1789

Peacock, W. J., Brutlag, D., Goldring, E., Appels, R., Hinton, C. W. and Linsley, D. L. (1973). The organization of highly repeated sequences in *Drosophila melanogaster* chromosomes. *Cold Spring Harbor Symp. Quant. Biol.* **38:** 405-416

Peacock, W. J., Lohe, A., Gerlach, W. L., Dunsmuir, P., Dennis, E. S. and Appels, R. (1977). Fine structure and evolution of DNA in heterochromatin. *Cold Spring Harbor Symp. Quant. Biol.* **42:** 1121-1135

Peterson, J. B. and Ris, H. (1976). Electron microscopic study of spindle and chromosome movement in the yeast *Saccharomyces cerevisiae. J. Cell Sci.* **22:** 219-242

Pfarr, C. M., Cove, M., Grissom, P. M., Hay, T. S., Porter, M. E. and McIntosh, J. R. (1990). Cytoplasmic dynein localises to kinetochore during mitosis. *Nature.* **345:** 263-265

Pietras, D. F., Bennett, K. L., Siracusa, L. D., Woodworth-Gutai, M., Chapman, V. M., Gross, K. W., Kane-Hass, C. and Hastie, N. D. (1983). Construction of a small *Mus musculus* repetitive DNA library: identification of a new satellite sequence in *Mus musculus. Nucl. Acids Res.* **11:** 6965-6983

Pluta, A. F., Cooke, C. A. and Earnshaw, W. C. (1990). Structure of the human centromere at metaphase. *Trends Biochem.* **15:** 181-185

Pluta, A. F., Saitoh, N., Goldberg, I. and Earnshaw, W. C. (1992). Identification of a subdomain of CENP-B that is necessary and suffient for localization to the human centromere. *J. Cell Biol.* **116:** 1081-1093

Polizzi, C. and Clarke, L. (1991). The chromatin structure of centromeres from fission yeast: differentiation of the central core that correlates with function. *J. Cell Biol.* **112:** 191-201

Pradnovszky, T., Keresö, J., Tubak, V., Cseprán, I., Fátyol, K. and Hadlaczky, G. (1991). *De novo* chromosome formation in rodent cells. *Proc. Natl. Acad. Sci. USA.* **88:** 11042-11046

Prosser, J., Frommer, M., Paul, C. and Vincent, P. C. (1986). Sequence relationship of three human satellite DNAs. *J. Mol. Biol.* **187:** 145-155

Ragoussis, J., Monaco, A. P., Mockridge, I., Kendall, E., Campbell, R. D. and Trowsdale, J. (1991). Cloning of HLA class II region in yeast artificial chromosomes. *Proc. Acad. Natl. Sci. USA.* **88:** 3753-3757

Rattner, J. B. (1991). The structure of the mammalian centromere. *BioEssays.* **13:** 51-56

Rattner, J. B., Kingwell, B. G. and Fritzler, M. J. (1988). Detection of distinct structural domains within the primary constriction using autoantibodies. *Chromosoma.* **96:** 360-367

Reider, C. and Alexander, S. P. (1990). Kinetochores are transported poleward along a single astral microtubule during chromosome attachment to the spindle in newt lung cells. *J. Cell Biol.* **110:** 81-95

Reider, C. L. (1982). The mammalian kinetochore. *Int. Rev. Cytol.* **79:** 1-58

Riley, J., Butler, R., Oligvie, D., Finniear, R., Jenner, D., Powell, S., Anand, R., Smith, J. C. and Markham, A. F. (1990). A noevl, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. *Nucl. Acids Res.* **18:** 2887-2890

Ris, H. and Witt, P. L. (1981). Stucture of the mammalian kinetochore. *Chromosoma.* **82:** 153-170

Saitoh, H., Tomkiel, J., Cooke, C. A., Ratrie III, H., Maurer, M., Rothfield, N. F. and Earnshaw, W. C. (1992) CENP-C, an autoantigen in scleroderma, is a component of the human inner kinetochore plate. *Cell.* **70:** 115-125

Salido, E. C., Yen, P. H., Koprivnikar, K., Yu, L. and Shapiro, L. H. (1992). The Human Enamel Protein Gene Amelogenin Is Expressed from Both the X and Y Chromosomes. *Am. J. Hum. Genet.* **50:** 303-316

Salmon, E. D. (1991) Microtubule dynamics and chromosome movement. In "Mitosis: Molecules and Mechanisms". (ed. J. S. Hyams and B. R. Brinkley) Academic Press. London. pp. 119-181.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977). DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA. **74:** 5463-5467

Saunders, M. J., Yeh, E., Grunstein, M. and Bloom, K. (1990). Nucleosome depletion alters the chromatin structure of *Saccharomyces cerevisiae* centromeres. *Mol. Cell Biol.* **10:** 5721-5727

Schlessinger, D. (1990). Yeast artificial chromosomes: tools for mapping and analysis of complex genomes. *Trends Genet.* **6:** 248-258

Schlessinger, D., Little, R. D., Freije, D., Abidi, F., Zucchi, I., Porta, G., Pilia, G., Nagaraja, R., Johnson, S. K., Yoon, J. Y., Srivastava, A., Kere, J., Palmieri, G., Ciccoicola, A., Montanaro, V., Romano, G., Casamassimi, A. and D'Urso, M. (1991). Yeast artificial-based genome mapping: some lessons from Xq24-28. *Genomics.* **11:** 783-793

Schrader, F. (1953). In "Mitosis. The movement of chromosomes in cell division". (ed. L. C. Dunn) Columbia Univ. Press, New York. pp. 1-170

Schwartz, D. C. and Cantor, C. R. (1984). Separation of yeast chromosome-sized DNAs by pulsed-field gel electrophoresis. *Cell.* **37:** 67-75

Sealy, P. G., Whittaker, P. A. and Southern, E. M. (1985). Removal of repeated sequences from hybridisation probes. *Nucl. Acids Res.* **13:** 1905-1922

Shero, S. H., McCormick, M. K., Antonarakis, S. E. and Hieter, P. (1991). Yeast artificial vectors for efficient clone manipulation and mapping. *Genomics.* **10:** 505-508

Siverman, G. A., Ye, R. D., Pollock, K. M., Sadler, J. E. and Korsmeyer, S. J. (1990). (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

Simerly, C., Balczon, R., Brinkley, B. R. and Schatten, G. (1990). Microinjected kinetochore antibodies interfere with chromosome movement in meiotic and mitotic oocytes. *J. Cell Biol.* **111:** 1491-1504

Singer, M. and Skowronski, J. (1985). Making sense out of LINES: long interspersed repeat sequences in mammalian genomes. *Trends Biochem. Sci.* **10:** 119-122

Sirota, L., Zlotogora, Y., Shabtai, F., Halbrecht, I. and Elian, E. (1981). 49,XYYYY. A case report. *Clinical Genetics.* **19:** 87-93

Sleister, H. M., Mills, K. A., Blackwell, S. E., Killary, A. M., Murray, J. C. and Malone, R. E. (1992). Construction of a human chromosome 4 YAC pool; and analysis of artificial chromosome stability. *Nucl. Acids Res.* **20:** 3419-3425

Smith, H. O. and Birnsteil, M. L. (1976). A simple method for DNA restriction site mapping. *Nucl. Acids Res.* **3:** 2387-2398

Solomon, M. J., Strauss, F. and Varshavsky, A. (1986). A mammalian high mobility group protein recognizes any stretch of six A-T base pairs of duplex DNA. *Proc. Natl. Acad. Sci. USA.* **83:** 1276-1280

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

Southern, E. M., Anand, R., Brown, W. R. A. 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

Steuer, E. R., Wordeman, L., Schroer, T. A. and Sheetz, M. P. (1990). Localization of cytoplasmic dynein to mitotic spindles and kinetochore. *Nature.* **345:** 226-268

Struhl, K., Stinchcomb, D. T., Scherer, S. and Davis, R. W. (1979). High-frequency transformation of yeast: autonomous replication of hybrid DNA molecules. *Proc. Natl. Acad. Sci. USA.* **76:** 1035-1039

Sullivan, K. F. and Glass, C. A. (1991). CENP-B is a highly conserved mammalian centromere protein with homology to the helix-loop-helix family of proteins. *Chromosoma.* **100:** 360-370

Szostak, J. and Blackburn, E. H. (1982). Cloning yeast telomeres on linear plasmid vectors. *Cell.* **29:** 245-255

Takahashi, K., Murakami, S., Chikashige, Y., Niwa, O. and Yanagida, M. (1991). A large number of tRNA genes are symmetrically located in fission yeast centromeres. *J. Mol. Biol.* **218:** 13-17

Traver, C. N., Klapholz, S., Hyman, R. W. and Davis, R. W. (1989). Rapid screening of a human genomic library in yeast artificial chromosomes for single-copy sequences. *Proc. Natl. Acad. Sci. USA.* **86:** 5898-5902

Tyler-Smith, C. (1987). Structure of repeated sequences in the centromeric region of the human Y chromosome. *Development (Suppl).* **101:** 93-100

Tyler-Smith, C. and Brown, W. R. A. (1987). Structure of the major block of alphoid satellite DNA on the human Y chromosome. *J. Mol. Biol.* **195:** 457-470

Tyler-Smith, C., Taylor, L. and Müller, U. (1988). Structure of a hypervariable tandemly repeated DNA sequence on the short arm of the human Y chromosome. *J. Mol. Biol.* **203:** 837-848

Vallee, R. (1990). Dynein and the kinetochore. *Nature.* **345:** 206-207

Vergnaud, G., Page, D. C., Simmler, M. C., Brown, L., Rouyer, F., Noel, B., Botstein, D., Chapelle, A. d. l. and Weissenbach, J. (1986). A deletion map of the human Y chromosome based on DNA hybridization. *Am. J. Hum. Genet.* **44:** 756-765

Verna, R. A. (1990) " The genome". VCH Publishers, New York.

Vig, B. K. (1984). Sequence of centromere separation: orderly separation of multicentric chromosomes in mouse L cells. *Chromosoma.* **90:** 39-45

Vilageliu, L. and Tyler-Smith, C. (1992) Structural Instability of YAC clones and the use of recombination deficient yeast host strains. In "Techniques for the analysis of complex genomes". (ed. R. Anand) Academic Press, London. pp. 92-112.

Warburton, P. E. and Willard, H. F. (1990). Genomic analysis of sequence variation in tandemly repeated DNA. Evidence for localized homogenous sequence domains within arrays of α-satellite. *J. Mol. Biol.* **216:** 3-16

Warburton, P. E., Greig, G. M., Haaf, T. and Willard, H. F. (1991). PCR amplification of chromosome -specific alpha satellite DNA: definition of centromeric STS markers and polymorhic analysis. *Genomics.* **11:** 324-333

Waye, J. S., England, S. B. and Willard, H. F. (1987). Genomic organisation of alpha satellite DNA on human chromosome 7: evidence for two distinct alphoid domains on a single chromosome. *Mol. Cell Biol.* **7:** 349-356

Waye, J. S. and Willard, H. F. (1989). Human ß satellite DNA: Genomic organization and sequence definition of a class of highly repetitive tandem DNA. *Proc. Natl. Acad. Sci. USA.* **86:** 6250-6254

Wevrick, R. and Willard, H. F. (1989). Long-range organization of tandem arrays of α satellite at the centromeres of human chromosomes: High frequency array-length polymorphism and meiotic stability. *Proc. Natl. Acad. Sci. USA.* **86:** 9394-9398

Wevrick, R., Earnshaw, W. C., Howard-Peebles, P. N. and Willard, H. F. (1990). Partial deletions of alpha satellite DNA associated with reduced amounts of the centromere protein CENP-B in a mitotically stable human chromosome rearrangement. *Mol. Cell. Biol.* **10:** 6374-6380

Wevrick, R. and Willard, H. F. (1991). Physical map of the centromeric region of human chromosome 7: relationship between two distinct alpha satellite arrays. *Nucl. Acids Res.* **19:** 2295-2301

Willard, H. F. (1985). Chromosome-specific organisation of human alpha satellite DNA. *Am J. Hum. Genet.* **37:** 524-532

Willard, H. F. and Waye, J. S. (1987). Hierarchical order in chromosome-specfic alpha satellite DNA. *Trends Genet.* **3:** 192-198

Willard, H. F. (1990). Centromeres of mammalian chromosomes. *Trends Genet.* **6:** 410-416

Wines, D. R. and Henikoff, S. (1992). Somatic instabilty of a Drosophila chromosome. *Genetics.* **131:** 683-691

Wolfe, J., Darling, S. M., Erickson, R. P., Craig, R. W., Buckle, V. J., Rigby, P. W. J., Willard, H. F. and Goodfellow, P. N. (1985). Isolation and characterization of an alphoid centromeric repeat family from the human Y chromosome. *J. Mol. Biol.* **182:** 477-48

Wong, A. K. C. and Rattner, J. B. (1988). Sequence organization and cytological organization of the minor satellite of the mouse. *Nucl. Acids Res.* **16:** 11,645-11,661

Wong, A. K. C., Biddle, F. G. and Rattner, J. B. (1990). The chromosomal distribution of the major and minor satellite is not conserved in the genus *Mus*. *Chromosoma.* **99:** 190-195

Yen, T. J., Compton, D. A., Wise, D., Zinkowski, R. P., Brinkley, B. R., Earnshaw, W. C. and Cleveland, D. W. (1991). CENP-E, a novel human centromere-associated protein required for progression from metaphse to anaphase. *EMBO J.* **10:** 1245-1254

Zhang, P., Knowles, B. A., Goldstein, L. S. B. and Hawley, R. S. (1990). A kinesin-like protein required for distributive chromosome segrgation in Drosophila. *Cell.* **62:** 1053-1062

Zinkowski, R. P., Vig, B. K. and Broccoli, D. (1986). Characterization of kinetochores in multicentric chromosomes. *Chromosoma.* **94:** 243-248

Zuo, J., Robbins, C., Taillon-Miller, P., Cox, D. R. and Myers, R. M. (1992). Cloning of the Huntindons disease region in yeast artificial chromosomes. *Hum. Mol. Genet.* **1:** 148-159