THE FINE STRUCTURE AND EVOLUTION
OF HUMAN REPEATED DNA

A. J. CLARK

Thesis submitted for the degree of
Doctor of Philosophy
University of Edinburgh

1982
DECLARATION

The work described in this thesis was done by me and the words written in it written by me.

A. J. Clark.
The Fine Structure and Evolution of Human Repeated DNA.

A number of human highly repeated DNAs are described. These include the classic satellites I-IV (Gosden et al., 1975) and a number of repeating restriction fragments, in particular a 170 bp repeating sequence (Manuelides, 1976) and a 3.4 kb Y specific sequence (Cooke, 1976).

Using a combination of restriction endonuclease digestion, molecular hybridisation and pyrimidine fingerprinting the fine structure and inter-relationship of these sequences has been elucidated. In all, three distinct repeated DNA families are apparent, a very AT rich group of sequences (satellite I), a diverse group of sequences sharing an underlying 5 bp periodicity (the Hinfl family) and a 170 bp long repeated sequence.

The location of these sequences and their arrangement on individual chromosomes has been investigated by means of in situ hybridisation and Southern 'blotting' of DNA prepared from human/rodent hybrid cell lines. Individual members of a particular family are found to have a discrete chromosomal location and arrangement, and it is postulated that one of the main sources of variation of these sequences is their independent evolution on different chromosomes.
ACKNOWLEDGEMENTS

My thanks to Dr. John Bishop (Genetics, Edinburgh) and Dr. John Gosden (MRC, CAPCU) for supervision, advice and encouragement. Also to Professor H. J. Evans for allowing me to work at the MRC, CAPCU. I am indebted to both Arthur Mitchell and Richard Buckland for access to the somatic cell hybrid DNAs, described in Chapter 8, and also for many a stimulated discussion. Sandy Bruce did the photography for this thesis and Lucian Begg the typing. To these two I extend my gratitude and apologies for the protracted length of time it took to put together. Finally, to Helen for her support, encouragement and continual misunderstanding.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>INTRODUCTION</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Non-Coding DNA</td>
<td>1</td>
</tr>
<tr>
<td>1.2</td>
<td>The Structure and Evolution of Highly Repeated DNA</td>
<td>4</td>
</tr>
<tr>
<td>1.3</td>
<td>Human Highly Repeated DNA</td>
<td>11</td>
</tr>
<tr>
<td>1.4</td>
<td>Aims and Approaches</td>
<td>14</td>
</tr>
<tr>
<td>2.1</td>
<td>Introduction</td>
<td>16</td>
</tr>
<tr>
<td>2.2</td>
<td>Restriction Analysis</td>
<td>16</td>
</tr>
<tr>
<td>2.3</td>
<td>Pyrimidine Tract Analysis</td>
<td>16</td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>21</td>
</tr>
<tr>
<td>3.2</td>
<td>Restriction Analysis</td>
<td>21</td>
</tr>
<tr>
<td>3.3</td>
<td>Pyrimidine Tract Analysis</td>
<td>32</td>
</tr>
<tr>
<td>3.4</td>
<td>Methylation of the Basic Sequence</td>
<td>36</td>
</tr>
<tr>
<td>3.5</td>
<td>Higher Order Repeating Structures of Satellite II</td>
<td>41</td>
</tr>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>44</td>
</tr>
<tr>
<td>4.2</td>
<td>Restriction Analysis</td>
<td>44</td>
</tr>
<tr>
<td>4.3</td>
<td>Pyrimidine Tract Analysis</td>
<td>47</td>
</tr>
<tr>
<td>4.4</td>
<td>Base Sequence and Evolution of Satellite IV</td>
<td>48</td>
</tr>
<tr>
<td>5.1</td>
<td>Introduction</td>
<td>56</td>
</tr>
<tr>
<td>5.2</td>
<td>Thermal Chromatography</td>
<td>56</td>
</tr>
<tr>
<td>5.3</td>
<td>Restriction Analysis</td>
<td>58</td>
</tr>
<tr>
<td>5.4</td>
<td>The Relationship between Satellite II and Satellite III</td>
<td>60</td>
</tr>
<tr>
<td>5.5</td>
<td>The Relationship between Satellite IV and Satellite III</td>
<td>63</td>
</tr>
<tr>
<td>5.6</td>
<td>The Sequence Composition of Satellite III</td>
<td>65</td>
</tr>
<tr>
<td>6.1</td>
<td>Introduction</td>
<td>71</td>
</tr>
<tr>
<td>6.2</td>
<td>Restriction Analysis</td>
<td>71</td>
</tr>
<tr>
<td>6.3</td>
<td>Evolution of the Sequence</td>
<td>78</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Pyrimidine tract analysis of satellite I</td>
<td>17</td>
</tr>
<tr>
<td>3.1</td>
<td>HinfI restriction of satellite II</td>
<td>22</td>
</tr>
<tr>
<td>3.2</td>
<td>HinfI and TaqI restriction of nick translated satellite II</td>
<td>24</td>
</tr>
<tr>
<td>3.3</td>
<td>Analysis of HinfI and TaqI fragments of satellite II &gt;25 bp</td>
<td>28</td>
</tr>
<tr>
<td>3.4</td>
<td>Restriction of the HinfI and TaqI 22 and 25 bp fragments of satellite II</td>
<td>30</td>
</tr>
<tr>
<td>3.5</td>
<td>Restriction map of the HinfI and TaqI 22 and 25 bp fragments of satellite II</td>
<td>31</td>
</tr>
<tr>
<td>3.6</td>
<td>Pyrimidine tract analysis of the HinfI, 10 bp, 12 bp and 15 bp fragments of satellite II</td>
<td>33</td>
</tr>
<tr>
<td>3.7</td>
<td>Pyrimidine tract analysis of the fast and slow strands of the 10 bp HinfI fragment of satellite II</td>
<td>35</td>
</tr>
<tr>
<td>3.8</td>
<td>Analysis of 5MeC within the TaqI sites of satellite II</td>
<td>38</td>
</tr>
<tr>
<td>3.9</td>
<td>Higher order repeats within satellite II</td>
<td>43</td>
</tr>
<tr>
<td>4.1</td>
<td>HinfI restriction of satellite IV</td>
<td>45</td>
</tr>
<tr>
<td>4.2</td>
<td>Pyrimidine tract analysis of the HinfI, 50 bp and 100 bp fragments of satellite IV</td>
<td>49</td>
</tr>
<tr>
<td>4.3</td>
<td>Pyrimidine tract analysis of the fast and slow strands of the 45 bp HinfI fragment of satellite IV</td>
<td>50</td>
</tr>
<tr>
<td>4.4</td>
<td>Pyrimidine tract analysis of satellite IV</td>
<td>51</td>
</tr>
<tr>
<td>4.5</td>
<td>Thermal chromatography of satellite IV</td>
<td>53</td>
</tr>
<tr>
<td>4.6</td>
<td>5' base analysis of C2T2 from satellite IV</td>
<td>55</td>
</tr>
<tr>
<td>5.1</td>
<td>Thermal chromatography of satellite III</td>
<td>57</td>
</tr>
<tr>
<td>5.2</td>
<td>Restriction analysis of the four human satellite DNAs</td>
<td>59</td>
</tr>
<tr>
<td>5.3</td>
<td>Restriction analysis of the four human satellite DNAs</td>
<td>61</td>
</tr>
<tr>
<td>5.4</td>
<td>Comparison of the MboI restriction patterns of satellite II and satellite III</td>
<td>62</td>
</tr>
<tr>
<td>5.5</td>
<td>The relationship between satellite III and satellite IV</td>
<td>64</td>
</tr>
<tr>
<td>5.6</td>
<td>Comparative restriction analysis of satellites II, III and IV</td>
<td>66</td>
</tr>
<tr>
<td>5.7</td>
<td>Analytical buoyant density centrifugation of satellite III before and after HaeIII digestion</td>
<td>70</td>
</tr>
<tr>
<td>FIGURE</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>6.1</td>
<td>Visualisation of the 170 bp repeat sequence in restriction digests of satellite III and total DNA</td>
<td>72</td>
</tr>
<tr>
<td>6.2</td>
<td>Restriction analysis of the 170 bp repeat sequence</td>
<td>74</td>
</tr>
<tr>
<td>6.3</td>
<td>Restriction analysis of the 170 bp repeat sequence</td>
<td>76</td>
</tr>
<tr>
<td>6.4</td>
<td>Restriction of the DdeI 170 bp monomer</td>
<td>77</td>
</tr>
<tr>
<td>6.5</td>
<td>Plot of $\log \frac{a_n}{n}$ vs $n - 1$ for the 170 bp repeat sequence after restriction of satellite III by DdeI and HaeIII</td>
<td>80</td>
</tr>
<tr>
<td>6.6</td>
<td>Thermal chromatography of the 340 bp HaeIII fragment</td>
<td>82</td>
</tr>
<tr>
<td>7.1</td>
<td>Gel electrophoresis of the 3.4 kb Y specific fragment</td>
<td>87</td>
</tr>
<tr>
<td>7.2</td>
<td>Hybridisation of the 3.4 kb Y fragment to total male and to total female DNA</td>
<td>88</td>
</tr>
<tr>
<td>7.3</td>
<td>Restriction map of the 3.4 kb Y fragment</td>
<td>90</td>
</tr>
<tr>
<td>7.4</td>
<td>Pyrimidine tract analysis of the 3.4 kb Y fragment</td>
<td>92</td>
</tr>
<tr>
<td>7.5</td>
<td>HinfI and TaqI restriction of the 3.4 kb Y fragment</td>
<td>95</td>
</tr>
<tr>
<td>7.6</td>
<td>Reassociation of the 3.4 kb Y fragment and satellite IV</td>
<td>97</td>
</tr>
<tr>
<td>7.7</td>
<td>Reassociation of the 3.4 kb Y fragment and satellite IV with total female DNA and total male DNA</td>
<td>98</td>
</tr>
<tr>
<td>7.8</td>
<td>Reassociation of the 3.4 kb Y fragment with female DNA and satellite IV at 50°C.</td>
<td>102</td>
</tr>
<tr>
<td>7.9</td>
<td>Thermal chromatography of the 3.4 kb Y fragment</td>
<td>104</td>
</tr>
<tr>
<td>8.1</td>
<td>In situ hybridisation of satellite I</td>
<td>112</td>
</tr>
<tr>
<td>8.2</td>
<td>Hybridisation of satellite I to DNA from somatic cell hybrid lines</td>
<td>114</td>
</tr>
<tr>
<td>8.3</td>
<td>In situ hybridisation of satellite II</td>
<td>116</td>
</tr>
<tr>
<td>8.4</td>
<td>Hybridisation of satellite II to DNA from somatic cell hybrid lines</td>
<td>118</td>
</tr>
<tr>
<td>8.5</td>
<td>Hybridisation of satellite II at low stringency</td>
<td>120</td>
</tr>
<tr>
<td>8.6</td>
<td>HinfI and TaqI restriction of isolated MboI fragments from satellite II</td>
<td>122</td>
</tr>
<tr>
<td>8.7</td>
<td>In situ hybridisation of human satellite IV</td>
<td>124</td>
</tr>
<tr>
<td>FIGURE</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>8.8</td>
<td>Hybridisation of satellite IV to DNA from somatic cell hybrids</td>
<td>126</td>
</tr>
<tr>
<td>8.9</td>
<td>Hybridisation of the 3.4 kb Y fragment at low stringency to somatic cell hybrid DNA</td>
<td>129</td>
</tr>
<tr>
<td>8.10</td>
<td>Restriction map of the 3.4 kb KpnI fragment</td>
<td>131</td>
</tr>
<tr>
<td>8.11</td>
<td>Hybridisation of satellite IV at low stringency</td>
<td>133</td>
</tr>
<tr>
<td>8.12</td>
<td>Hybridisation of the 680 bp HaeIII fragment to somatic cell hybrid DNA and to total human DNA</td>
<td>137</td>
</tr>
<tr>
<td>10.1</td>
<td>First cycle Ag⁺ Cs₂SO₄ gradients</td>
<td>168</td>
</tr>
<tr>
<td>10.2</td>
<td>Analytical CsCl profiles of purified human satellite DNAs</td>
<td>170</td>
</tr>
<tr>
<td>10.3</td>
<td>Calibration of agarose gels</td>
<td>176</td>
</tr>
<tr>
<td>10.4</td>
<td>Calibration of acrylamide gels</td>
<td>178</td>
</tr>
<tr>
<td>10.5</td>
<td>The separation of pyrimidine tracts by thin layer chromatography</td>
<td>184</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Pyrimidine tract analysis of nick translated satellite I.</td>
</tr>
<tr>
<td>3.1</td>
<td>Size determination of HinfI and TaqI fragments of satellite II.</td>
</tr>
<tr>
<td>3.2</td>
<td>The Stoichiometry of HinfI and TaqI digests of satellite II.</td>
</tr>
<tr>
<td>3.3</td>
<td>Pyrimidine tract analysis of the 10, 12 and 15 bp fragments of satellite II.</td>
</tr>
<tr>
<td>4.1</td>
<td>Size determination of HinfI fragments of satellite IV.</td>
</tr>
<tr>
<td>4.2</td>
<td>Pyrimidine tract analysis of the 50 bp and 100 bp HinfI fragments and of total satellite IV.</td>
</tr>
<tr>
<td>6.1</td>
<td>Densitometry of HaeIII and DdeI digests of satellite III.</td>
</tr>
<tr>
<td>7.1</td>
<td>Comparison of the pyrimidine fingerprints of the 3.4 kb Y fragment and satellite IV.</td>
</tr>
<tr>
<td>7.2</td>
<td>Summary of the reassociation data for the 3.4 kb Y fragment and satellite IV.</td>
</tr>
<tr>
<td>8.1</td>
<td>Enzyme marker and chromosome analysis of hybrid cell lines.</td>
</tr>
<tr>
<td>10.1</td>
<td>The human satellite DNAs</td>
</tr>
<tr>
<td>10.2</td>
<td>List of restriction enzymes</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

1.1 Non-Coding DNA

One of the characteristic features of the genomes of higher organisms is the substantial portion of DNA that does not code for proteins and the variability with which it is organised when closely related species, or even individuals within the same species, are compared. For the purpose of description these sequences can be grouped into a number of classes, although these distinctions are becoming much less rigid as the detailed organisation of eukaryotic genomes is elucidated by powerful techniques, such as molecular cloning and rapid sequencing, that are now available.

Highly Repeated (Satellite) DNA

These sequences are primarily confined to the condensed heterochromatic regions of chromosomes and are usually comprised of long tandem arrays of a sequence that is repeated many thousands or even millions of times per haploid genome. As human highly repeated DNA is the subject of this study the structure and evolution of this class of sequence is discussed separately, and in some detail, in the next section.

Intermediately Repeated Sequences

These sequences were first characterised by Britten and Kohne (1968). They are generally thought to be repeated only a few hundred times per haploid genome. They are not organised in tandem arrays, but are interspersed with sequences present only once or a very few times per haploid genome and are generally found dispersed throughout the genome. Many eukaryotes share a common pattern of interspersion in which middle repeat sequences about 300 bp long are interspersed with single copy sequences about 2000 bp long (Davidson et al., 1973;
Graham et al., 1974; Chamberlin et al., 1975). There are indications that some 'middle' repeat sequences may be much more highly reiterated than was at first thought. The human genome has a family of interspersed 300 bp repeats comprising about 3% of the total genome, characterised by the presence of an AluI restriction site about 170 bp from one end (Houck et al., 1979). Thus the reiteration frequency of this sequence (approximately 300,000) is what might be expected of a highly repeated satellite DNA, whereas its pattern of interspersion is characteristic of an intermediately repeated sequence family.

In D. melanogaster a number of middle repetitive DNA sequences are transcribed forming abundant poly(A) containing RNA (Rubin et al., 1976; Potter et al., 1979). A remarkable property of these sequences, and also for many of the other middle repetitive sequences of the D. melanogaster genome (Young, 1979), is that they are transposable and so different patterns of dispersion are seen when their chromosomal arrangements are examined in genetically isolated laboratory stocks or non-interbreeding wild-type populations (Potter et al., 1979, Strobel et al., 1979). Transposable middle repetitive DNA families have also been described in yeast (Cameron et al., 1979).

Foldback DNA

These sequences were discovered by their ability to form duplexes very rapidly and independently of DNA concentration. They are distributed throughout the genome and, in some cases, are due to the inversion of adjacent or near adjacent sequences such as the AluI family (Houck et al., 1979; Shen and Maniatis, 1980). In Xenopus it has been shown that these sequences have different patterns of dispersion in different individuals, thus indicating that they are transposable (Perlman et al., 1976).
Spacer Sequences

These are sequences that separate repetitive genes such as the ribosomal or histone genes. In some cases, for example the 5S and ribosomal genes of *Xenopus laevis*, the spacer sequences are internally repetitive (Wellauer *et al.*, 1976; Botchan *et al.*, 1977; Federoff and Brown, 1978) though this is not necessarily always the case, as for example in the complex spacer that separates the histone genes in *S. purpuratus* (Sures *et al.*, 1978). Spacers from different repeats within a tandem array often show considerable length heterogeneity (Wellauer *et al.*, 1976; Botchan *et al.*, 1977; Federoff and Brown, 1978) and homologous gene clusters from closely related species as exemplified by the oocyte 5S DNA of *X. laevis* and *X. borealis* can have very different spacer sequences (Korn and Brown, 1978).

Homopolymer Clusters

These are long runs of one base on one strand of DNA and are distributed throughout the genome (Flavell *et al.*, 1977).

Introns

The finding that a great many eukaryote genes are split by non-coding sequences (introns) was one of the most important discoveries of molecular biology during the 1970's (Breathnach *et al.*, 1977; Flavell *et al.*, 1978; van den Berg *et al.*, 1978; Tilghman *et al.*, 1978; Tonegawa *et al.*, 1977). It has profound implications for understanding the functioning and evolution of the genomes of higher animals (Crick, 1979; Gilbert, 1978). The length and sequence of introns appears to be variable when homologous genes from different species are compared (van den Berg *et al.*, 1978) or when related members of a gene family within an organism are compared (Lawn *et al.*, 1978; Tiemeier *et al.*, 1978). The sequences at the non-coding/coding junctions are conserved and it has been postulated that these sequences are important in the process...
of RNA splicing that generates mature mRNA molecules (Lerner et al., 1980). Many introns appear to be single copy sequences but they may be repeated as, for example, in the 28S genes of D. melanogaster, some of which are split by a sequence formed in tandem arrays at sites distant from the nucleolar organiser (Glover and Hogness, 1977; Glover et al., 1978; Dawid and Wellauer, 1977).

1.2 The Structure and Evolution of Highly Repeated DNA

Most of the sequence arrangements described above share the property of repetition and variability. It is difficult to assign unique distinguishing features to any of them and each class of sequence can have a range of properties. It could be argued that highly repeated DNAs are in a different class because they seem to be confined to the condensed, heterochromatic regions of the chromosomes. The detection of small amounts of these sequences interspersed throughout much of the genome, as for example in the African green monkey (Singer, 1979) and the red crab (Christie and Skinner, 1979) might argue that such differences are quantitative rather than qualitative and that these sequences are not in an autonomous class but are, perhaps, best considered as extremes on a common theme of repetition.

The tandem arrangement of a repeating sequence often results in molecules of DNA sufficiently large and sufficiently different in base composition that they can be separated from the bulk of the genomic sequences by isopycnic centrifugation as 'satellite' DNAs (Kit, 1961; Mazrimas and Hatch, 1977; Gall and Atherton, 1974). Other highly repeated sequences may require the presence of metal ions (Corneo et al., 1970) or antibiotics (Barnes et al., 1978) for their visualisation and isolation. Not all highly repeated sequences can be isolated by isopycnic centrifugation and in these cases their kinetics of reassociation
or pattern of restriction (Shmookler-Reis and Goldstein, 1980; Bedbrook et al., 1980a, b) has been used for their isolation and/or characterisation. In some cases these highly repeated sequences are not in tandem arrays but are interspersed widely throughout the genome (Lagowski et al., 1973; Houck et al., 1979), whereas in other cases they conform to classic notions of satellite DNAs and are present in long tandem arrays that are concentrated in the heterochromatin (Bedbrook et al., 1980a, b).

The structure of many highly repeated DNAs is extremely simple as exemplified by the crab satellite which is >95% poly d(AT) (Skinner and Kerr, 1971) and a number of closely related satellite DNAs, poly (5'ACAAATT3'), poly (5'ACAAACT3') and poly (5'ATAAAACT3') which are present in D. virilis (Gall and Atherton, 1974). Other highly repeated DNAs are composed of a number of simple related oligonucleotides that seem to have been derived from a common ancestral sequence (Biro et al., 1975; Pech et al., 1979). Short repeats may be arranged in larger more complex repeating units as, for example, in the mouse satellite (Southern, 1975a) and calf satellites (Streeck and Zachau, 1978; Pech et al., 1979). Not all complex highly repeated DNAs show evidence of a simpler origin. The 359 bp repeat of the Drosophila 1.688 g/cm³ satellite and a 170 bp repeated sequence common to many primates both appear to lack any internal repetitiveness (Hsieh and Brutlag, 1979a; Rosenberg et al., 1978; Manuelidis and Wu, 1978; Donehower et al., 1980).

Maio et al. (1977) and Musich et al. (1977) showed that there appears to be a relationship between the subunit structure of a chromatin, as imposed by the nucleosome spacing, and the repeat periodicities of highly repeated DNA sequences. Thus, a number of these sequences
from a variety of organisms share an underlying periodicity of around 170 bp, although whether there is a true phasing of the nucleosomes with the actual repeat sequence has been questioned by a number of other workers (Fittler and Zachau, 1979; Singer, 1979). Early studies on highly repeated DNAs indicated that even closely related species possessed completely different sequences when analysed by hybridisation or by density gradient centrifugation (Hennig and Walker, 1970; Beattie and Skinner, 1972). The elucidation of the simple repeating nature of the sequences involved explained some of the enormous variation because if a single base-pair was changed in a short repetitive unit it would change the physical properties of the DNA dramatically (Blumenfeld et al., 1973; Gall and Atherton, 1974). In some groups of animals the changes in highly repeated DNA seem to be more quantitative than qualitative. Thus Salser et al. (1976) and Fry and Salser (1977) have shown great similarities between some of the highly repeated sequences found in rodents. A number of different rodents share the sequence 5'GGGTTA3' and variants of it in their α satellites. The main difference between the different species is the relative amounts of this sequence. It is probably fair to say that as techniques for investigating highly repeated sequences have improved so has their inter-relationship in different species become more apparent. For example, digestion of total DNA from a number of primates with a variety of restriction enzymes yields a characteristic pattern due to the presence of a 170 bp repeated sequence. These can comprise as much as 20% (African green monkey, Singer, 1979) or as little as 3% (man, Manuelidis, 1976, 1978a). Restriction mapping and sequence analysis of these repeats shows them to be related and derived from a common ancestral repeat present many millions of years ago (Manuelidis and Wu, 1978;
Donehower and Gillespie, 1979; Donehower et al., 1980). In fact, regions of homology between highly repeated DNA derived from such distantly related species as fruit flies and cows and humans and rats has now been reported (Brutlag, 1980); a far cry from the earlier notions that even closely related species had totally unrelated highly repeated sequences.

Theories for the evolution of highly repeated DNA must account for both the quantitative and qualitative variation of these sequences that is generally observed. One of the most successful theories was first presented by Southern (1970) and postulated that highly repeated DNAs evolve by a process of random mutation causing sequence divergence of short repeating units followed by amplification of the more complex sequences generated and was prompted by the finding that the pyrimidine fingerprint of the guinea pig α satellite was much simpler than the kinetic complexity of this sequence indicated. This model considered the basic repeat units of highly repeated DNA to be continually undergoing alteration in both amount and sequence and hence explained the high degree of variation observed.

One of the strongest pieces of evidence for this theory is the finding of more than one distinct periodicity within a highly repeated DNA, indicating that divergence within tandem arrays of simple repeats followed by amplification can create a new periodicity. Thus, the mouse satellite is comprised of a number of related simple oligonucleotides less than 20 nucleotides (Biro et al., 1975) which are organised into repeating units about 240 bp in length (Southern, 1975a) and examination of calf satellites with different restriction enzymes indicates very short and very long periodicities within the same sequence (Streeck and Zachau, 1978; Pech et al., 1979). Restriction enzymes have been used to demonstrate random divergence within tandemly repeated sequences.
(Southern, 1975a; Fittler, 1977; Altenberger et al., 1977). In these cases the restriction enzymes were shown to generate an oligomeric series of DNA fragments whose lengths were multiples of a monomer repeat and whose relative amounts fitted the distribution expected if the restriction sites had been lost by random base changes.

Although some of the bases in a repeating unit appear to diverge randomly there is considerable evidence that the frequency of sequence divergence is not necessarily constant throughout the repeat unit. Direct sequence analysis of the monomers from the African green monkey a satellite shows that some of the sites within the repeat have a much higher frequency of divergence than others (Rosenberg et al., 1978). Sequence analysis of 15 adjacent monomers was almost sufficient to account for the heterogeneity observed in a Drosophila sequence (Hsieh and Brutlag, 1979a). The limitation of the divergence to a few prominent locations may be the result of the amplification and fixation of a few repeating units that contain these particular sequences.

Non-random alteration in restriction sites has also been observed in some highly repeated DNAs by quantitating the amounts of monomer, dimer, trimer etc. that result from restriction. In these cases the amount of DNA in the oligomers does not decrease monotonically as predicted from random mutation (Christie and Skinner, 1980; Dennis et al., 1980) and this could result from selective amplification of certain regions of the tandem array. In fact, Brutlag and Peacock (1975) have postulated that much of the overall heterogeneity that is observed in highly repeated DNA sequences is not so much due to the accumulation of mutations as to the presence of a number of related, regularly arranged alternative sequences (i.e. sequence isomers). This view of the heterogeneity within highly repeated DNAs has received considerable experimental support. For example, a satellite DNA present in Apodemus
sylvaticus has been shown by Cooke (1975) to be comprised of two related sequences with the same 370 bp periodicity. These sequences reassociate to give well matched intrafamily duplexes but poorly matched interfamily duplexes. Similarly, restriction enzymes can also divide up a highly repeated sequence into relatively homogeneous families or segments that contain a particular recognition sequence or completely lack it. This clustering of restriction sites, as opposed to their random distribution throughout the whole sequence, suggests that amplifications of only a part of the sequence may occur (Altenburger et al., 1977; Fittler, 1977; Horst and Zachau, 1977).

That different segments may reside on different chromosomes has been indicated by the observation that the restriction pattern of sequences complementary to human satellite III differed between individual chromosomes examined in somatic cell hybrids (Beauchamp et al., 1979; Chapter 8).

In addition to the single base changes observed in repeat units divergence has sometimes occurred more dramatically. Carlson and Brutlag (1979) have described tandem arrays of a repeated sequence 254 bp in length that hybridises to the 359 bp repeating unit of the *D. melanogaster* 1.688 g/cm³ satellite and is derived from it by a 105 bp deletion. Divergence can also occur by the insertion of an unrelated sequence into a tandem array, as reported for certain highly repeated DNAs in rye (Bedbrook et al., 1980a, b).

A number of molecular mechanisms have been proposed to account for the structure of highly repeated DNAs and the ways in which they change during evolution. One mechanism that has received much support is unequal exchange between homologous tandem arrays (Smith, 1973, 1976). This envisages crossing over between two tandem repeating arrays that are not in perfect register and will result in recombinants shorter or
longer than the parental arrays. Such a mechanism can lead to rapid quantitative changes in the number of repeats in a tandem array. This process can both maintain the repeats within a tandem array homogeneous if the exchange rate is greater than the rate of mutation (Smith, 1973) or can result in the amplification and fixation of a newly diverged sequence which happened to duplicate. The structure of many highly repeated DNAs indicates that this process may be occurring. For example, the presence of 1-mer repeats in the mouse satellite and the fact that the monomer can reassociate out of register is consistent with the notion of unequal crossing over occurring in an ancestral subrepeat (Southern, 1975a). The structure of satellite II from the hermit crab is a series of short simple repeats with the formula 5'(CAG)$_n$CTGACT3' where $n$ is normally distributed between 3 and 12, and is consistent with unequal crossing over being responsible for the propagation and maintenance of this sequence (Chambers et al., 1978). A recent study in yeast in which a genetic marker was inserted into the tandem array of ribosomal genes showed a high level of unequal recombination between ribosomal gene clusters located on sister chromatids during meiosis (Petes, 1980).

Although unequal crossing over can occur during either meiosis or mitosis only those events that occur in the germ line have any evolutionary significance. Crossing over during meiosis within constitutive heterochromatin seems very infrequent, if not entirely absent, and hence Kurnit (1979) has concluded that it is unequal sister chromatid exchange during germ line mitoses which is the source of much of the variation apparent in highly repeated DNA.

Although it seems pretty certain that unequal crossing over between tandem repeats does occur, it is by no means clear whether this
mechanism accounts entirely for the structure and variability of highly repeated DNA. Other mechanisms have been proposed such as chromosomal excision and extrachromosomal replication as in the rolling circle mechanism (Hourcade et al., 1973). Some workers, for example consider that amplifications of tandem arrays may occur by such mechanisms but favour unequal crossing over for short range events such as spreading a restriction site throughout a segment of a tandem array (Horz and Zachau, 1977).

From the foregoing account it is apparent that many details of the structure and evolution of highly repeated DNA have been established. However, relatively little information exists as to how these sequences spread to the different chromosomes of the genome and how important such chromosomal isolation is in the continuing evolution of these sequences. I hope to show that the study of human highly repeated DNA can profitably address itself to some of these questions.

1.3 Human Highly Repeated DNA

Centrifugation of total human DNA in CsCl gradients reveals no obvious buoyant density satellites however, the use of silver and mercuric ions does resolve a number of cryptic satellites (Corneo et al., 1968; Corneo et al., 1970; Corneo et al., 1971; Corneo et al., 1972). The total amount of these satellites has been estimated to be about 4%-6% of the total genome (Maio, 1973; Jones, 1977; Manuelidis, 1978a) although this is probably an underestimate as these sequences are not necessarily completely resolved from main band DNA (Chapter 5). Originally four satellites, designated I-IV were identified by Corneo and his co-workers however, other workers have reported the presence of additional human satellite DNAs (Schildkraut and Maio, 1969; Saunders et al., 1972; Saunders, 1974). Some of these additional
satellite DNAs may have been reported because of confusions in identity due to inaccurate buoyant density measurements and also because of the vagaries of satellite DNA preparation by heavy metal binding (Chapter 5). The isolation and characterisation of the satellites originally designated I-IV has proved the most reproducible in the hands of different workers (Jones and Corneo, 1971; Jones et al., 1973; Gosden et al., 1975; Marx et al., 1976; Mitchell et al., 1979). Even so, there still remains a large amount of confusion as to the precise nature and inter-relationship of these human highly repeated DNAs. For example, human satellite II was originally identified and isolated from the heavy side of main band $^{+}$Cs$_2$SO$_4$ gradients (Corneo et al., 1970) and this has been reproduced by Jones and Corneo (1971), Gosden et al. (1975) and Mitchell et al. (1979), yet Manuelidis (1978a) and Miklos and John (1979) reported that this satellite bands on the light side of main band DNA in $^{+}$Cs$_2$SO$_4$ gradients. Similarly, the restriction properties of apparently the same satellite seem to differ widely in the hands of different workers; Mitchell et al. (1979) report that satellite II is cut by EcoRI whereas Manuelidis (1978a) reports that this satellite is resistant to digestion by this enzyme. Mitchell et al. (1979) report extensive sequence homology between the different human satellites but Manuelidis (1978a, b) regards them as distinct, unrelated sequences and that any homology between them is due to cross-contamination during their isolation. The pattern of in situ hybridisation obtained for satellite I and satellite II by Jones and Corneo (1971) and Jones et al. (1974) is very different from that reported by Gosden et al. (1975), again underlying the confusion in identity of these sequences. Even cloning and sequencing does not seem to have clarified the issue. Cooke and Hindley (1979) have reported the sequence of a cloned 1.77 kb EcoRI fragment isolated from human satellite
III. However, the properties of this sequence bear little resemblance to those of satellite III reported by other workers (Beauchamp et al., 1979; Mitchell et al., 1979; Gosden et al., 1981a).

The reason for much of the confusion in the field is most likely due to the difficulty of isolating these sequences in a pure form because they are a small proportion of the total genome and because of the somewhat erratic nature of Ag⁺ binding as a means of purifying satellite DNAs. Three of the satellites have similar buoyant densities (Chapter 10) which are also very close to that of main band DNA, giving rise to a situation where the cross-contamination of sequences during their isolation is likely.

Human highly repeated DNAs have been studied without recourse to satellite DNA purification. For example, Manuelidis and co-workers (1977, 1978a, b, 1980) have described the repeating structure, location and base sequence of a sequence that is prominent in restriction digests of total human DNA and is characterised by its digestion with EcoRI and HaeIII. It consists of tandem arrays of a 340 bp repeat and comprises about 3% of the genome. Sequence analysis (Wu and Manuelidis, 1980) has shown that the 340 bp repeat is comprised of two related 170 bp repeats, but no shorter periodicity is evident. In its arrangement of restriction sites this sequence family is very similar to some of the sequences that comprise satellite III (Bostock et al., 1978; Mitchell et al., 1979), although Wu and Manuelidis (1980) report that it is not isolated as a satellite DNA (Chapters 5 and 6).

Comparison of restriction digests of total male and total female DNA has revealed repeating restriction fragments that are specific to the Y chromosome (Cooke, 1976). One prominent 3.4 kb fragment co-purifies with human satellite III in Ag⁺ Cs₂SO₄ gradients (Bostock et al., 1978) and comprises about 40-50% of the total Y chromosome. This
fragment has been reported by Cooke (1976) to be comprised of a tandem array of short simple repeats, but Kunkel et al. (1979) have presented evidence showing that this repeating fragment contains a large number of unrelated and relatively complex sequences (Chapter 7).

1.4 Aims and Approaches

In the light of the previous discussion it is apparent that the current description of human highly repeated DNA leave much to be desired. This is unfortunate as humans offer some unique advantages for studying the structure and evolution of these sequences because of their well defined karyotype, the existence of numerous heterochromatic polymorphisms and, more recently, the construction of various human/rodent hybrid cell lines which make it possible to study the precise arrangement of highly repeated DNA sequences on individual chromosomes (Beauchamp et al., 1979). However, an overall understanding of the structure of these sequences is clearly essential before questions as to their chromosomal location and arrangement can be asked. Consequently, one of the primary aims of this work was to provide a comprehensive description of the structure and inter-relationship of the various human highly repeated DNAs.

The starting point for this was the isolation of the classical satellites I-IV in Ag\(^{+}\) Cs\(_2\)SO\(_4\) gradients. This was facilitated by the use of vertical ultracentrifuge rotors (Sorvall TV850 and TV865) which dramatically improved the resolution of these sequences from main band DNA and thus increased their yield and purity relative to preparative centrifugation in angle-head rotors. The structure of these sequences was investigated by pyrimidine tract analysis and digestion with restriction enzymes. Some of the satellites had simple fingerprints and were digested to very small fragments by certain
restriction enzymes so that their simple, basic repeating structure could be elucidated. Not all the sequences could be described in such fine detail, but nevertheless, digestion with a variety of restriction enzymes served to identify the various sequences of which the four satellite DNAs were comprised. The structure of repeating restriction fragments such as the 340 bp repeated sequence (Manuelidis, 1978) and the 3.4 kb Y-specific fragment (Cooke, 1976) was also investigated and related to the structure of the satellite DNAs.

Once the structure and inter-relationship of these various highly repeated DNAs had been established their location and arrangement on different chromosomes was investigated. This was accomplished by in situ hybridisation and by hybridisation to Southern 'blots' of DNA prepared from human/rodent hybrid cell lines. One of the aims of the project was to assess the extent that these sequences evolve independently on different chromosomes and to relate this to current ideas (section 1.2) on the evolution of highly repeated DNA.

Up until now the highly repeated DNA of man has not been well characterised. At present many other aspects of the structure of this genome are being investigated. Nevertheless, I feel that a correct description of all its sequence components, from highly repeated DNA to unique structural gene, is essential to an understanding of its overall structure, workings and evolution.
CHAPTER 2

THE STRUCTURE OF HUMAN SATELLITE I

2.1 Introduction

Human satellite I was first described by Corneo et al. (1967) and is the lightest of the four human satellites ($\rho = 1.686 \text{ g/cm}^3$) isolated from $\text{Ag}^+ \text{Cs}_2\text{SO}_4$ gradients. In first cycle gradients (section 10.2) it bands many fractions away from the light side of main band DNA and has been estimated to comprise about 0.5% of total genomic DNA (Maio, 1973). In the analytical ultracentrifuge the purified satellite bands sharply in neutral CsCl (Section 10.2) and appears virtually free of contaminating sequences.

2.2 Restriction Analysis

Purified satellite I proved highly resistant to digestion with a wide variety of restriction endonucleases although some enzymes, Alu I, Taq I and Kpn I did digest a small proportion of the satellite into discrete fragments.

2.3 Pyrimidine Tract Analysis

Pyrimidine tract analysis (section 10.7) was carried out on satellite I which had been labelled by nick translation with a single $\alpha^{32}$P deoxypurine triphosphate (A or G) or with a mixture of $\alpha^{32}$P deoxypyrimidine triphosphates (C + T) of the same specific activity. C + T labelling and subsequent depurination yields a mixture of pyrimidine oligonucleotides of general formula $5' (\hat{\text{P}}C)_n (\hat{\text{P}}T)_m 3'$ (* indicates a labelled phosphate residue). DNA labelled with A or G yields pyrimidine tracts labelled at their 3' end having the general formula $5' (\hat{\text{P}}C)_n (\hat{\text{P}}T)_m 3'$. Figure 2.1 shows the three types of fingerprint obtained with satellite I. Table 2.1 tabulates the relative
Figure 2.1
Pyrimidine tract analysis of satellite I

Pyrimidine tracts were prepared by depurinating nick translated satellite I. They were separated by thin layer chromatography (Section 10.7). The first dimension (1) was developed in 2M pyridine formate and the second dimension (2) was developed in 1M LiCl.

i  C+T labelled satellite I
ii  A labelled satellite I
iii  G labelled satellite I.

The amount of radioactivity in the spots was measured by cutting them out and counting in a scintillation counter (Table 2.1).
TABLE 2.1

Pyrimidine Tract Analysis of Nick-Translated Satellite I.

The figures in columns 2, 4 and 6 were obtained by cutting out and counting the respective spots from the TLC plates. Those in column 3 estimate the purine residues, as % total nucleotides, that are 3' to a given pyrimidine tract, and were calculated from the formula $\frac{\% P_{CT}}{21}$, where $\% P_{CT}$ is the % phosphate in a C+T labelled tract 1 nucleotides long. The figures in columns 5 and 7 estimate the A and G residues, as % total nucleotides, 3' to a given pyrimidine tract and were calculated by multiplying the % phosphate in a given tract labelled by A or G by the % A or % G of satellite I estimated from the GC composition (= 23.7%). Each figure in column 8 is obtained by adding the corresponding figures in columns 5 and 7. Ideally the figures in column 8 should equal those in column 3.
<table>
<thead>
<tr>
<th>Tract</th>
<th>C+T Label</th>
<th>A Label</th>
<th>G Label</th>
<th>A+G</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% PCT</td>
<td>% PCT</td>
<td>% PA</td>
<td>% PA x% A</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>T1</td>
<td>51.3</td>
<td>25.65</td>
<td>51.9</td>
<td>19.8</td>
</tr>
<tr>
<td>T2</td>
<td>3.8</td>
<td>0.95</td>
<td>1.6</td>
<td>0.6</td>
</tr>
<tr>
<td>T3</td>
<td>3.7</td>
<td>0.6</td>
<td>0.2</td>
<td>0.1</td>
</tr>
<tr>
<td>T4</td>
<td>5.1</td>
<td>0.65</td>
<td>2.9</td>
<td>1.1</td>
</tr>
<tr>
<td>C1</td>
<td>11.8</td>
<td>5.9</td>
<td>9.5</td>
<td>3.7</td>
</tr>
<tr>
<td>C2</td>
<td>1.3</td>
<td>0.3</td>
<td>0.7</td>
<td>0.25</td>
</tr>
<tr>
<td>C3</td>
<td>1.7</td>
<td>0.3</td>
<td>0.7</td>
<td>0.30</td>
</tr>
<tr>
<td>C4</td>
<td>0.4</td>
<td>0.05</td>
<td>&lt;0.1</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>C1T1</td>
<td>9.5</td>
<td>2.4</td>
<td>1.6</td>
<td>0.6</td>
</tr>
<tr>
<td>C1T2</td>
<td>2.3</td>
<td>0.4</td>
<td>0.2</td>
<td>0.1</td>
</tr>
<tr>
<td>C1T3</td>
<td>6.2</td>
<td>0.8</td>
<td>0.6</td>
<td>0.25</td>
</tr>
<tr>
<td>C2T1</td>
<td>0.9</td>
<td>0.15</td>
<td>0.5</td>
<td>0.20</td>
</tr>
<tr>
<td>C2T2</td>
<td>1.0</td>
<td>0.15</td>
<td>0.4</td>
<td>0.15</td>
</tr>
<tr>
<td>C2T3</td>
<td>1.0</td>
<td>0.10</td>
<td>0.6</td>
<td>0.20</td>
</tr>
<tr>
<td>P</td>
<td>-</td>
<td>-</td>
<td>28.3</td>
<td>-</td>
</tr>
</tbody>
</table>

TOTAL C = 23.7
proportions of the pyrimidine oligonucleotides estimated by cutting out the spots and counting in a scintillation counter. Table 2.1 also compares the C + T labelling with the A and G labelling and although the figures in column 3 and column 8 are not entirely in agreement, the correspondence between the two columns is reasonable considering the errors involved. From the data in Table 2.1 it can be calculated that about 55% of all T is in the form of \( Pu \ T \ n \ A_n \), where \( n \) is mainly 1; most of the remaining T and most of C is in the form of \( Pu \ T \ G_n \) or \( Pu \ TG \), where \( n \) is, again, mainly 1. Thus satellite I consists primarily of AT interspersed with C, G, CT and longer runs of T. The simplicity of the pyrimidine fingerprints of satellite I does not necessarily imply a simple periodic repeat such as is seen in the satellite DNAs of a number of species (Southern, 1970; Brutfalag and Peacock, 1979; Skinner, 1977). If such a simple periodic structure does underly the structure of human satellite I then the repeating unit must be highly diverged because, despite the simplicity of its fingerprint, this sequence appears to be the least repetitious of the four human satellites, having a kinetic complexity estimated to be in excess of 600 (Mitchell et al., 1979).

A putative satellite I isolated from \( Ag^+CsSO_4 \) gradients has been described by Manuelidis (1978a). The pyrimidine fingerprint of the ATP labelled DNA was similar to that described here except that the amount of C(A) was greater than the amount of T(A) and there was significant labelling of \( C_2T_2(A) \). The most likely explanation for this is that the satellite DNA described by Manuelidis was contaminated with other satellite sequences, most probably satellite IV which bands next to satellite I in first cycle gradients (pyrimidine tract analysis of
satellite IV yields prominent C(A) and C_2T_2(A) tracts). The cross contamination of human satellite sequences during their preparative isolation is a common occurrence and is discussed more fully in Chapters 5 and 9.

The main features of human satellite I are its simple pyrimidine fingerprint and its resistance to cleavage by restriction endonucleases. These features can be used to distinguish it from the sequences that comprise the other human satellites (Chapter 8) and which do not seem to be related to it in any way.
3.1 Introduction

Satellite II was first described by Corneo et al. (1970). It differs from the three other human satellites that are isolated from $\text{Ag}^+ \text{Cs}_{2}\text{SO}_4$ gradients in that it strongly binds silver and is present as a heavy shoulder to main band DNA, while its buoyant density in neutral CsCl is 1.694 g/cm$^3$, considerably lighter than that of main band ($\rho = 1.700$ g/cm$^3$). This satellite has been estimated to comprise about 2% of total genomic DNA (Maio, 1973).

3.2 Restriction Analysis

The fine structure of satellite II was investigated with the restriction enzymes HinfI and TaqI both of which digest the bulk of the satellite down to very small fragments. Figure 3.1.i shows satellite II digested with HinfI and electrophoresed on an agarose gel. Figure 3.1.ii shows the resolution of the small fragments generated by digestion with HinfI obtained by electrophoresing the end labelled fragments on a 7M urea, 20% acrylamide gel (Section 10.4). The bulk of the label is present in five major fragments which migrate faster than the xylene cyanol marker. These five fragments are single-stranded. They were sized with reference to the bromophenol blue and xylene cyanol markers (which, under these conditions, are equivalent to 10 and 28 nucleotides, respectively, Maxam and Gilbert (1978)) and to synthetic oligonucleotides, poly $d\text{T}_{10}$ and poly $d\text{T}_{12-18}$ (Miles Laboratories) that had been end labelled with polynucleotide kinase.

The five HinfI fragments when sized with respect to the dye markers, on the assumption of a log/linear relationship between MW and mobility were estimated to have sizes of 10.1, 12.3, 15.8, 22.4 and 25.6 nucleotides
Figure 3.1
HinfI restriction of satellite II

i 1.5% agarose gel stained with 2 μg/ml ethidium bromide.
   a, φXRF DNA restricted with HaeIII - the arrow marks the
   position of the 115 bp fragment that was just visible on
   the original negative; b, 1.5 μg of purified satellite II
   restricted with HinfI.

ii End-labelled fragments of HinfI restricted satellite II
   electrophoresed on a 20% polyacrylamide, 7M urea gel.
   BB, position of bromophenol blue marker; XC, position of
   xylene cyanol marker.
(averages determined from four gels). Sizing with respect to the synthetic oligonucleotides estimated the fragments as 9.5, 12.4, 15.2, 22.0 and 25.1 nucleotides. The sizes were concluded to be 10, 12, 15, 22 and 25 nucleotides.

Fragments migrating above the xylene cyanol marker were found to be double stranded. This was shown by isolating the first two fragments above the marker and rerunning on another 7M urea, 20% acrylamide gel. The individually isolated bands from the gel resolved into two bands on the second gel, one having the same electrophoretic mobility as the original and the other, the single stranded derivative, migrating much faster (Figure 3.4). The separated strands were not resolved under the conditions employed.

The first two fragments above the xylene cyanol marker are the duplex 22 and 25 base-pair molecules. The size of these double stranded fragments was measured independently by reference to XRF size markers and estimated to be 21.6 and 24.8 bp (Table 3.1) The remainder of the Hinf I restricted satellite II consists of a number of duplex fragments ranging in size from about 32-70 bp.

The Hinf I pattern of restriction was compared to that of Taq I by digesting nick translated satellite II (double-labelled with $\alpha^{32}$P dCTP and $\alpha^{32}$P dTTP) and separating the products on 7M urea, 20% acrylamide gels (Fig. 3.2). For Hinf I the same pattern of fragments is seen as with DNA end-labelled by means of polynucleotide kinase except that a large amount of label is excluded from the gel. Since this labelled DNA would have to be of very high MW to be excluded from the gel and because its relative amount varied between different nick-translated preparations, it was concluded to be 'network DNA' which is an artefact of the nick translation reaction caused by strand switching of the DNA polymerase I.
Figure 3.2
HinfI and TaqI restriction of nick translated satellite II.

C+T labelled satellite II was restricted by HinfI, TaqI or both enzymes and the products electrophoresed on 7M urea 20% polyacrylamide gels (Section 10.4). The DNA was not denatured prior to loading onto the gel and the 22 bp and 25 bp fragments were present both as single stranded (ss) and double stranded (ds) fragments.

a, HinfI; b, TaqI; c, TaqI followed by HinfI; d, HinfI followed by TaqI. BB, position of bromophenol blue marker; XC, position of xylene cyanol marker. Occasionally bands that were not reproduced on other gels were seen, such as the one marked by an arrow in lane d; these may arise by strand separation of the small, single stranded restriction fragments.
<table>
<thead>
<tr>
<th>EndoR Fragment</th>
<th>Av. Size $\sigma(n=3)$</th>
<th>EndoR Fragment</th>
<th>Av. Size $\sigma(n=3)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HinfI</td>
<td></td>
<td>TaqI</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>21.6 ±0.15</td>
<td>1</td>
<td>22.4 ±0.8</td>
</tr>
<tr>
<td>2</td>
<td>24.8 ±0.4</td>
<td>2</td>
<td>25.1 ±1.1</td>
</tr>
<tr>
<td>3</td>
<td>33.6 ±0.6</td>
<td>3</td>
<td>33.4 ±1.2</td>
</tr>
<tr>
<td>4</td>
<td>35.7 ±0.7</td>
<td>4</td>
<td>36.6 ±1.0</td>
</tr>
<tr>
<td>5</td>
<td>39.3 ±1.1</td>
<td>5</td>
<td>40.2 ±1.2</td>
</tr>
<tr>
<td>6</td>
<td>48.8 ±2.0</td>
<td>6</td>
<td>48.2 ±1.5</td>
</tr>
<tr>
<td>7</td>
<td>50.7 ±1.1</td>
<td>7</td>
<td>51.8 ±1.6</td>
</tr>
<tr>
<td>8</td>
<td>58.5 ±2.2</td>
<td>8</td>
<td>58.3 ±2.3</td>
</tr>
<tr>
<td>9</td>
<td>61.3 ±2.3</td>
<td>9</td>
<td>66.8 ±1.7</td>
</tr>
<tr>
<td>10</td>
<td>65.9 ±1.9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 3.1**

Size Determination of HinfI and TaqI Fragments of Satellite II.

HinfI and TaqI fragments of satellite II larger than 20 bp were sized on 20 cm or 30 cm 20% polyacrylamide gels by reference to $\Phi$XRF DNA markers. The fragments correspond to those marked in Figure 3.4, their average size was determined from three estimates.
<table>
<thead>
<tr>
<th>Fragment</th>
<th>% DNA</th>
<th>Molar Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HinfI</td>
<td>TaqI</td>
</tr>
<tr>
<td>&lt;10</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>19.6±4.8</td>
<td>14.4±2.9</td>
</tr>
<tr>
<td>12</td>
<td>10.6±0.65</td>
<td>10.4±1.8</td>
</tr>
<tr>
<td>15</td>
<td>13.0±1.4</td>
<td>7.5±1.4</td>
</tr>
<tr>
<td>18</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>22</td>
<td>13.1±5.0</td>
<td>7.3±1.2</td>
</tr>
<tr>
<td>25</td>
<td>18.2±2.1</td>
<td>10.5±1.2</td>
</tr>
<tr>
<td>&gt;25</td>
<td>25.6±3.7</td>
<td>49.6±3.9</td>
</tr>
</tbody>
</table>

TABLE 3.2

The Stoichiometry of HinfI and TaqI digests of Satellite II.

Autoradiographs of HinfI and TaqI satellite II, such as the one shown in Figure 3.2, were scanned on a Joyce-Loebl microdensitometer and the %DNA in each fragment was estimated by cutting out and weighing the respective peaks. The molar ratios were estimated relative to the 10 bp fragments.
Figure 3.2 and similar autoradiographs were scanned on a Joyce-Loebl microdensitometer and the amount of DNA in each restriction fragment estimated. Table 3.2 lists the average values obtained from a number of different gels. In this analysis the 'network' DNA excluded from the gel was ignored.

The pattern of fragments 25 bp and below obtained by digestion with Hinf I and Taq I was the same, indicating that the recognition sequences of both enzymes (GANTC and TCGA) are part of the same short repeating sequence which must comprise a large proportion of the satellite DNA. With Hinf I about 75% of the satellite was cut to 25 bp or below, whereas only about 50% of the satellite was cut to these small sizes with Taq I. The molar ratio of the 10 bp repeat to the combined 12 and 15 bp repeats approximated to 1:1 for both Hinf I (1:0.88) and Taq I (1:0.95) digestion. The stoichiometry of the pattern indicated that for every 10 bp repeat a 12 bp or 15 bp repeat was released upon digestion with either enzyme. The arrangement of these, the three smallest, repeats was indicated by the fact that the next two largest fragments released by digestion with either enzyme were 22 bp and 25 bp long. This strongly suggests that the 10 bp and 12 bp repeats are tandemly arranged and that loss of the intervening restriction site by point mutation has generated 22 bp repeats. Similarly, the 25 bp repeat would consist of a 10 bp and a 15 bp repeat tandemly arranged. Furthermore, for the fraction of the satellite cut to below 30 bp it is unlikely that the three basic repeats are arranged in any other way because fragments of 20 bp (10-10), 24 bp (12-12), 27 bp (12-15) or 30 bp (15-15) are not produced in any significant amount by restriction with either enzyme.

With Hinf I digestion and with double digestion with both enzymes the molar ratio of 12 bp:15 bp repeats is close to 1 (Table 3.2).
Figure 3.3

Analysis of Hinfl and TaqI fragments of satellite II >25 bp.

i  Hinfl restricted satellite II labelled by polynucleotide kinase and electrophoresed on a 20% polyacrylamide, 7M urea gel.

ii Satellite II labelled with C+T by nick translation, restricted by TaqI and electrophoresed on a 20% polyacrylamide gel. Small fragments were run off the bottom of the gel to increase the resolution of the larger fragments. The numbered fragments refer to those tabulated in Table 3.1. These were sized a number of times by reference to $\Phi XRF$ marker DNAs. XC, position of xylene cyanol marker; ss, single stranded 25 bp fragment.
For Taq I digestion this ratio was estimated as 1:1.47. The overall ratio of 12 bp + 22 bp fragments: 15 bp + 25 bp fragments, averaged for all four columns of Table 3.2 is 1:1.02. Thus, the 12 bp and 15 bp repeats may be present in equimolar amounts and a possible structure for a large portion of the satellite DNA is tandem arrays of a repeat such as 10-12-10-15, itself built up of the three basic 10 bp, 12 bp and 15 bp repeats each of which contains within it the recognition sequence for Hinf I and Taq I.

Fragments larger than 25 bp produced by restriction of satellite II were investigated on 20 cm and 30 cm 20% polyacrylamide gels. The separation of the fragments was increased by running the smaller fragments off the bottom of the gel. The remaining fragments were sized by reference to $\phi$XRF markers (Section 10.4). Figure 3.3 shows the pattern obtained with both Hinf I and Taq I digested satellite II and Table 3.1 lists the average molecular weights that were determined from a number of gels. The restriction patterns of both enzymes are similar, and there are probably a number of equally sized fragments, but the different mobility of Hinf I and Taq I fragments (Section 10.4) makes direct comparison difficult. Both restriction patterns are similar in that very few fragments are produced between 25 bp and 35 bp. This would be expected if the higher MW fragments were predominantly comprised of tandem arrays of 10, 12 and 15 bp repeats but in which two or more adjacent restriction sites had been lost. However, the large number of fragments predicted and the impossibility of absolute sizing makes the testing of this model by direct MW determination of restriction fragments very difficult.

Double digestion of nick translated satellite II showed that 35% of the Hinf I fragments and 66% of the Taq I fragments greater than 25 bp were digested to fragments 25 bp or smaller with the second
Figure 3.4
Restriction of the HinfI and TaqI 22 and 25 bp fragments of satellite II

Nick translated satellite II was digested with HinfI or TaqI and the double stranded 22 bp and 25 bp fragments extracted from a 20% polyacrylamide gel (Section 10.4). These were digested with the other enzyme and the products separated by electrophoresis on a 20% polyacrylamide, 7M urea gel alongside undigested controls.

a, HinfI 22 bp fragment - undigested; b, HinfI 22 bp fragment - cut by TaqI; c, HinfI 25 bp fragment - undigested; d, HinfI 25 bp fragment - cut by TaqI; e, TaqI 22 bp fragment - undigested; f, TaqI 25 bp fragment - cut by HinfI; g, TaqI 25 bp fragment - undigested; h, TaqI 25 bp fragment - cut by HinfI; ds, double stranded 22 bp fragment; ss, single stranded 22 bp fragment; XC, position of xylene cyanol; BB, position of bromophenol blue.
The size of some of the fragments is indicated on the right hand side of the gel.
Figure 3.5
Restriction map of the HinfI and TaqI 22 and 25 bp fragments of satellite II

The relative positions of the possible HinfI and TaqI sites were mapped onto the 22 bp and 25 bp repeats of satellite II by digestion of the isolated HinfI and TaqI fragments with the other enzyme (Fig. 3.4).

i The 22 bp repeats; a and b, the two possible HinfI fragments generated by loss of HinfI sites by point mutation; c and d, the two possible TaqI fragments.

ii The 25 bp repeats; a and b, the two possible HinfI fragments; c and d, the two possible TaqI fragments.

iii The recognition sequences of the two restriction enzymes were mapped onto the repeats; a, the 22 bp repeat; b, the 25 bp repeat. H, site of HinfI cut; T, site of TaqI cut. Sites marked with a closed circle are presumed lost by point mutation, thus generating 22 bp and 25 bp repeats from 10-bp + 12-bp repeats and 10 bp + 15 bp repeats. The 22-bp repeat can be generated from the 25 bp repeat by deletion of the 3 bases indicated between the closed triangles.
```plaintext
i

\[
\begin{array}{c}
\text{a} \\
\text{b} \\
\text{c} \\
\text{d}
\end{array}
\begin{array}{c}
H & H^* & H \\
T & H & T & H^* & H \\
H & T & H \\
T & T & T
\end{array}
\]

ii

\[
\begin{array}{c}
\text{a} \\
\text{b} \\
\text{c} \\
\text{d}
\end{array}
\begin{array}{c}
H & H^* & H \\
T & H & T & H^* & H \\
H & T & H \\
T & T & T
\end{array}
\]

iii

\[
\begin{array}{c}
\text{a} \\
\text{b}
\end{array}
\begin{array}{c}
\text{GIA \cdot TC \cdots TCGA \cdot TC \cdots TCGA} \\
\text{GIA \cdot TC \cdots TCGA \cdot TC \cdots TCGA}
\end{array}
\]
```
enzyme (Table 3.2). Only two major new fragments were visible after double digestion and these would be predicted from the relative positions of the Hinf I and Taq I sites (Figure 3.5). At a minimum, then, at least 83% of satellite II is composed of tandem arrays of the basic repeats.

The relative positions of the Hinf I and Taq I sites were mapped on the 22 and 25 bp fragments. This was accomplished by extracting the double-stranded 22 and 25 bp fragments, restricting with the other enzyme and electrophoresing the products on a second 7M urea, 20% acrylamide gel. Figure 3.4 is an autoradiograph of such a gel and Fig. 3.5 shows the relationship of the Hinf I and Taq I sites deduced from the fragments obtained. Both the 22 and 25 bp fragments are the same in that the 10-12 and 10-15 repeats defined with one enzyme are staggered with respect to those defined by the other enzyme. Thus, for each enzyme one restriction site overlaps that of the other enzyme and one is independent.

3.3 Pyrimidine Tract Analysis

The Hinf I 10 bp, 12 bp and 15 bp repeats were further analysed by examining their pyrimidine fingerprints. Figure 3.6 shows the pyrimidine fingerprints of these three fragments and Table 3.3 lists the proportion of each tract estimated by cutting out the spots and counting in 'Instagel' (section 10.7). All three fragments give a similar fingerprint with $C_1T_1$, $T_1$, $CT_2$ and $C_2T_2$ as the predominantly labelled tracts. The 22 and 25 bp fragments and total satellite II also gave similar pyrimidine fingerprints. All three fragments are heterogeneous as shown by the fact that the calculated repeat length for a number of the pyrimidine tracts is much greater than the size of the restriction fragment (Table 3.3).
Figure 3.6

Pyrimidine tract analysis of the HinfI 10 bp, 12 bp and 15 bp fragments of satellite II.

C+T labelled satellite II was restricted with HinfI and electrophoresed on a polyacrylamide gel. The 10 bp, 12 bp and 15 bp fragments were extracted from the gel (Section 10.4), depurinated and the pyrimidine tracts separated by thin layer chromatography (Section 10.7). The first dimension (1) was developed in 2M pyridine formate pH 3.4 and the second dimension (2) was developed in 1M LiCl.

i  10 bp HinfI fragment

ii  12 bp HinfI fragment

iii 15 bp HinfI fragment.
<table>
<thead>
<tr>
<th>Tract</th>
<th>10 bp</th>
<th>12 bp</th>
<th>15 bp</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%P</td>
<td>Repeat length</td>
<td>%P</td>
</tr>
<tr>
<td>( T_1 )</td>
<td>9.6</td>
<td>10.4</td>
<td>19.6</td>
</tr>
<tr>
<td>( T_2 )</td>
<td>2.2</td>
<td>90.9</td>
<td>7.6</td>
</tr>
<tr>
<td>( T_3 )</td>
<td>2.4</td>
<td>12.5</td>
<td>3.9</td>
</tr>
<tr>
<td>( C_1 )</td>
<td>1.7</td>
<td>58.8</td>
<td>1.3</td>
</tr>
<tr>
<td>( C_1 T_1 )</td>
<td>21.0</td>
<td>9.5</td>
<td>32.0</td>
</tr>
<tr>
<td>( C_1 T_2 )</td>
<td>22.6</td>
<td>13.2</td>
<td>13.9</td>
</tr>
<tr>
<td>( C_1 T_3 )</td>
<td>2.40</td>
<td>16.6</td>
<td>2.9</td>
</tr>
<tr>
<td>( C_2 T_1 )</td>
<td>6.70</td>
<td>44.7</td>
<td>9.2</td>
</tr>
<tr>
<td>( C_2 T_2 )</td>
<td>31.50</td>
<td>12.6</td>
<td>9.6</td>
</tr>
</tbody>
</table>

**TABLE 3.3**

Pyrimidine Tract Analysis of the 10, 12 and 15 bp Fragments of Satellite II.

The %P was estimated by cutting out each spot and counting in a scintillation counter. The repeat length estimates the length of DNA in bp in which a given pyrimidine tract repeats and is given by \( \frac{100 \times n}{\%P} \) where \( n \) is the number of nucleotides in the given pyrimidine tract.
The isolated 10 bp Hinfl fragment (C+T labelled) was separated into fast (F) and slow (S) strands by electrophoresis on a 20% polyacrylamide gel containing no urea. The two strands were separately isolated, depurinated and their pyrimidine tracts separated by thin layer chromatography.

BB, position of bromphenol blue; XC, position of xylene cyanol.

Pyrimidine fingerprint of the fast strand.

Pyrimidine fingerprint of the slow strand.

1, the first dimension was developed in 2M pyridine formate pH 3.4;
2, the second dimension was developed in 1M LiCl.
The 10 bp fragment was resolved into a lightly labelled fast strand and a heavily labelled slow strand by re-electrophoresing it, after isolation from a 7M urea, 20% acrylamide gel, on a 20% acrylamide gel containing no urea (Figure 3.7). Depurination of these two strands showed that the major pyrimidine tracts of the slow strand were \( C_2 T_2 \) and \( C_1 T_2 \) whereas \( C_1 T_1 \) predominated on the fast strand. Strand separation and subsequent depurination of the 12 bp and 15 bp fragment gave similar results in that \( C_1 T_1 \) predominated on the fast strand and \( C_2 T_2 \) and \( C_1 T_2 \) on the slow strand.

The similarity of their overall pyrimidine fingerprints and of the fingerprints of their fast and slow strands indicates that the same sequence comprises each fragment. This sequence probably accounts for most of satellite II because the pyrimidine fingerprint of total satellite II is very similar to those derived from the three fragments (see Fig. 10.6). From the pyrimidine tract data it is not possible to discern the actual sequence and judging from the heterogeneity of the three fragments, it is no doubt highly diverged. However, the structure of the 22 and 25 bp repeats (Fig. 3.5) shows that the doublet 5' TC3' repeats every 5 bp four out of five times and the doublet 5' GA3' repeats every 10 bp three out of four times. These periodicities would be expected of a sequence that had an underlying 5 bp periodicity. If this is the case then the 10 bp, 15 bp and 25 bp fragments are multiples of the basic, albeit highly diverged 5 bp repeat, whereas the 12 bp and 22 bp have been derived either by 2 bp insertion in the 10 bp repeat or a 3 bp deletion from the 15 bp repeat.

3.4 Methylation of the Basic Sequence

For the fine structure of satellite II proposed the recognition sequences Hinf I and Taq I are part of the short repeats that comprise
Figure 3.8

Analysis of $^{5}\text{Me}C$ within the TaqI sites of satellite II

The C residues within the TaqI sites of satellite II and $^{5}$XRF DNA (an unmethylated control) were labelled with $^{32}\text{P}$ by means of polynucleotide kinase. Labelled DNA was digested with snake venom phosphodiesterase or bacterial alkaline phosphatase (to release inorganic phosphate) and the products separated by high pressure liquid chromatography (Section 10.8) along with cold marker nucleosides.

i Satellite II digested with snake venom phosphodiesterase.

ii $^{5}$XRF DNA digested with snake venom phosphodiesterase.

iii $^{5}$XRF DNA digested with bacterial alkaline phosphatase.

The arrows labelled 1 and 2 mark the peaks of the cold C and $^{5}\text{Me}C$ nucleosides respectively whose elution was monitored optically. A small lag of 15-30 seconds was observed between the optical peak and the radioactive peak of elution.
this DNA and restriction fragments of larger size than the basic repeats are considered to have arisen primarily by the loss of recognition sites by point mutation. The number of sites lost by point mutation for each enzyme can be calculated on the assumption that the entire sequence consists of arrays of 10 bp, 12 bp and 15 bp repeats. The fraction of altered sites, P, is then given by the formula $P = a_n (1 - \frac{1}{n})$, where $a$ is the fraction of DNA in a particular size group, and $n$ the number of adjacent restriction sites lost to produce fragments of that size. Such an analysis indicated that for satellite II significantly more Taq I sites had been lost than Hinf I sites ($P_{\text{Taq I}} = 0.44$, $P_{\text{Hinf I}} = 0.34$).

The recognition sequences of both enzymes contain all four bases (G A N T C and T C G A ) and therefore on point mutation alone the rate of loss of Taq I sites should equal that of Hinf I sites. The Taq I recognition sequence however, contains within it the doublet 5'CpG3' and in mammalian DNA the C of this doublet can be modified by methylation to 5-methylcytosine (Bird, 1980). Many restriction enzymes are blocked by methylation of bases within their recognition sequence (Bird, 1978) and so Taq I may be cutting satellite II less than Hinf I because of this base modification.

This possibility was investigated by analysing the C residue within the Taq I sites to determine whether or not they were methylated. This was done by restricting satellite II with Taq I, end labelling the 5'C generated with polynucleotide kinase and digesting this labelled DNA with snake venom phosphodiesterase (Worthington). The resulting 5' mononucleotides were separated by high pressure liquid chromatography (section 10.8) and the ratio of 5-methylcytosine to cytosine measured. Figure 3.8 shows the results of this experiment. Both C and 5MeC were released from the C of the
cut Taq I site showing that the action of the enzyme was not blocked by methylation of this residue (see also van den Ploeg, 1980). About 40% of the total cytosine residues from the Taq I site were found to be methylated and since CpG methylation is symmetrical (Bird, 1978) this would also equal the percentage of methylated Taq I sites.

Although methylation of the internal C of the Taq I site does not block the action of the restriction enzyme it may explain the greater loss of Taq I sites by point mutation. Thus a number of studies have shown that 5 methyl cytosine is prone to point mutation because it is susceptible to deamination producing a thymidine residue. A $5\text{Me}C + T$ transition would not alter the Hinf I sites of satellite II as judged from the relative positions of the two sites (Fig. 3.5).

If the differential loss of Taq I sites is due to the greater rate of point mutation of the $5\text{Me}C$ residues then the rate of mutation of this residue compared to the other unmodified bases can be estimated. This calculation assumes that satellite II is an equimolar array of Hinf I and Taq I sites which are diverging by point mutation, that the rate of mutation in the four unmodified bases is the same and that $\text{Me}C + C$ is zero.

Considering an initial array of Taq I sites of which ($x$) are unmethylated and ($1-x$) are methylated. The fraction of unmethylated Taq I sites lost through point mutation = the fraction of Hinf I sites lost = 0.34. Therefore:

number of unmethylated Taq I sites lost = 0.34x

number of unmethylated Taq I sites remaining = 0.66x

number of methylated Taq I sites lost = ($1-x$)PM

number of methylated Taq I sites remaining = ($1-x$)(1-PM)
where $P_M$ is the fraction of methylated sites lost and total Taq I sites lost = $0.34x + (1-x)P_M = 0.44$. (1)

The ratio of unmethylated to methylated Taq I sites was estimated by means of HPLC (Fig. 3.8) as 60/40 = $0.66x/(1-x)(1-P_M)$. (2)

Solving (1) and (2) simultaneously gives a value for $P_M$ of 0.55. This contrasts with the value of 0.34 estimated for unmethylated restriction sites. The average rate change ($p$) for each of the four bases in the unmethylated restriction site is given by $P = 1-(1-p)^4$ where $P$ = fraction of restriction sites lost. For Hinf I sites $p = 0.1$. If the same rate of change is assumed for the three unmodified bases of methylated Taq I sites then the rate of change ($p^*$) in the $5^{\text{Me}}C$ is given by the formula $P_M = 1-((1-p)^3 x (1-p^*))$ and $p^* = 0.375$. Thus the rate of mutation of 5 methyl cytosine residues would be estimated to be 3.75 times greater than that of unmodified bases.

3.5 Higher Order Repeating Structures of Satellite II

The recognition sequences for the enzymes Hinf I and Taq I are part of the short repeating sequences that comprise satellite II. Other restriction enzymes cut satellite II but they yield restriction fragments much larger than these basic repeating units. Thus, as well as the short order periodicities described in the previous sections the sequences of satellite II are also arranged in larger repeating units. Figure 3.9.i shows satellite II digested with the enzyme MboI. Most of the satellite is cut to below 1700 bp with this enzyme and a complicated pattern of restriction fragments is generated. EcoRI cuts about 50% of the purified satellite to a pattern of fragments very different to that seen with MboI. Most of the fragments are greater than 1700 bp and show no obvious relationship to those
generated by MboI. Thus there are at least two distinct higher order periodicities contained within the sequences that comprise satellite II.

The pattern of restriction fragments generated by MboI or EcoRI can be used to detect and distinguish satellite II sequences in Southern transfers of, for example, total genomic DNA (see also Chapters 5 and 8). Figure 3.9.ii is a Southern 'blot' of total female DNA which has been restricted with MboI, EcoRI and AluI, and hybridised to $^{32}$P labelled satellite II. The pattern of bands is essentially the same as that seen with ethidium bromide stained gels of purified satellite II. AluI cuts satellite II yielding a pattern of fragments almost identical to that seen with EcoRI. The similarity of the AluI and EcoRI restriction patterns indicates that both enzymes are cutting the same higher order repeats of satellite II.

A human highly repeated DNA having a structure resembling satellite II has been described by Cooke and Hindley (1979). This sequence was derived from a 1.77 kb EcoRI fragment of satellite III and consisted of a number of very similar sequences related to a 9 bp element 5'ATTC G/C ATTC 3' interspersed with other sequences which differed from each other considerably. This sequence resembles satellite II in being located predominantly on chromosome 1 (Chapter 8) and in having EcoRI and AluI repeats 1.77 kb long. However, it does not have such regular elements as the 10, 12 and 15 bp repeats described for satellite II although it has closely spaced Hinf I and Taq I sites, some of which overlap. The relationship of this sequence to satellite II is consistent with the notion that both are related and members of a diverse simple sequence family to which a number of other human highly repeated DNAs also belong (Chapters 8 and 9).
Figure 3.9
Higher order repeats within satellite II

Restriction of satellite II with a number of enzymes reveals the presence of higher order repeating periodicities.

i 1.5% agarose gel of satellite II digested by \(a\), MboI and \(b\), EcoRI. Tracks \(a\) and \(b\) are from different gels and contain 2 \(\mu g\) and 1 \(\mu g\) of purified satellite II, respectively.

ii Hybridisation of restricted female DNA to \(^{32}P\) labelled satellite II. \(b-g\), 3 \(\mu g\) female DNA per track (1.5% agarose gel) digested with \(b\), MboI; \(c\), AluI; \(d\), EcoRI; \(e\), EcoRI/MboI; \(f\), AluI/EcoRI; \(g\), AluI/MboI. The open triangles in track \(b\) indicate some of the characteristic MboI fragments of satellite II whereas the closed triangles in track \(d\) indicate some of the characteristic EcoRI fragments (sizes given to the left and right of the gel respectively). A small amount of nick translated lambda DNA was included in the hybridisation to visualise the lambda EcoRI/HindIII digest in track \(a\).
CHAPTER 4
THE STRUCTURE OF SATELLITE IV

4.1 Introduction

Human satellite IV was originally purified by binding to MAK columns. It was found to band close to satellite III, on the light side of main band DNA, in Ag\textsuperscript{+}Cs\textsubscript{2}SO\textsubscript{4} gradients (Corneo et al., 1972). In the present study satellite IV was prepared directly from first cycle Ag\textsuperscript{+}Cs\textsubscript{2}SO\textsubscript{4} gradients centrifuged in vertical rotors. In these gradients it was found to band as a definite peak between satellite I and the light shoulder of main DNA that constituted satellite III (Section 10.2). The purified satellite banded as a sharp unimodal peak in neutral CsCl and had a buoyant density of 1.699-1.700 g/cm\textsuperscript{3}.

4.2 Restriction Analysis

The structure of satellite IV was investigated with the restriction enzymes Hinfl and TaqI (Section 10.3). Hinfl digests of satellite IV migrated on agarose gels as a smear of low MW DNA fragments (Fig. 4.1.i). In a limit digest about 60% of the Hinfl restricted DNA was cut to fragments below 100 bp. TaqI also cuts satellite IV, but a limit digest with this enzyme yields larger fragments than Hinfl (see Fig. 5.7).

The small Hinfl fragments of satellite IV were resolved by electrophoresis on polyacrylamide gels. Fig. 4.1.ii is an etidium bromide stained gel of purified satellite IV and Fig. 4.2.i an autoradiograph of \textsuperscript{32}P labelled satellite IV after Hinfl restriction. The fragments up to 100 bp were sized from a number of gels using \textsuperscript{4}XRF DNA markers (Table 4.1). The fragments formed a series of bands that regularly increased in size by 5 bp. The shortest restriction fragment that was readily detectable using \textsuperscript{32}P labelled satellite IV
Figure 4.1

HinfI restriction of satellite IV

i 1.5% agarose gel. a, 1.5 µg of purified human satellite IV digested by HinfI; b, *XRF DNA digested by HaeIII.

ii 15% polyacrylamide gel. a, *XRF DNA digested by HaeIII; b, 2.0 µg of purified satellite IV digested by HinfI; c, *XRF DNA digested by HinfI.
### TABLE 4.1

Size Determination of HinfI Fragments of Satellite IV.

HinfI fragments of satellite IV were sized on 20 cm or 30 cm, 15% or 20% polyacrylamide gels by reference to XRF DNA markers. The values tabulated above are the averages from a number of separate determinations.

<table>
<thead>
<tr>
<th>Fragment</th>
<th>Average Size</th>
<th>Increment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\sigma$</td>
<td>$n$</td>
</tr>
<tr>
<td>1</td>
<td>23.8 ±1.0</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>28.9 ±1.12</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>34.7 ±1.12</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>39.8 ±1.18</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>44.5 ±0.85</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>49.8 ±0.36</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>54.8 ±0.27</td>
<td>5</td>
</tr>
<tr>
<td>8</td>
<td>59.7 ±0.47</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>66.1 ±1.37</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>69.2 ±0.52</td>
<td>5</td>
</tr>
<tr>
<td>11</td>
<td>74.9 ±1.33</td>
<td>4</td>
</tr>
<tr>
<td>12</td>
<td>78.8 ±1.56</td>
<td>3</td>
</tr>
<tr>
<td>13</td>
<td>83.5 ±1.85</td>
<td>3</td>
</tr>
<tr>
<td>14</td>
<td>87.5 ±1.32</td>
<td>3</td>
</tr>
<tr>
<td>15</td>
<td>93.0</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>97.1 ±2.73</td>
<td>5</td>
</tr>
</tbody>
</table>


was 25 bp long, although at longer periods of exposure of the autoradiographs trace amounts of DNA corresponding to 10, 12, 15 and 20 bp restriction fragments were observed.

Fragments could be resolved and sized reasonably accurately up to 100 bp but thereafter they were too closely spaced to be sized. Between the 100 bp fragment and the $\alpha$XRF 141 bp fragment, 9 fragments could be counted on the negative of Fig. 4.1.ii suggesting that the restriction fragments were still increasing in size by increments of 5 bp. In the well resolved 25-100 bp range no bands, other than those which formed the 5 bp incremental series, were produced by HindI restriction.

4.3 Pyrimidine Tract Analysis

The pyrimidine fingerprints of isolated HindI restriction fragments were analysed (Section 10.4). Fig. 4.2 shows the pyrimidine fingerprints of $\alpha^{32}$P dCTP/TTP labelled 50 bp and 100 bp HindI restriction fragments. The fingerprints of both bands were essentially the same and produced $T$ and $C_2T_2$ as the major spots. This same pattern was observed for other isolated restriction fragments (10, 30 and 45 bp) that were analysed in this manner.

Restriction fragments were separated into a fast and slow strand by loading the individually isolated fragments (Section 10.4) in 0.2M NaOH and electrophoresing on 20% TBE gels. Fingerprints of these isolated strands showed that $C_2T_2$ was on the slow strand, whereas the main pyrimidine tract of the fast strand was $T$ (Fig. 4.3).

The pyrimidine fingerprints of total satellite IV were also investigated after labelling with $\alpha^{32}$P dCTP/TTP, dATP or dGTP (Fig. 4.4). The $C + T$ fingerprint of total satellite IV was essentially
the same as that for individually isolated restriction fragments (Table 4.2). dATP labelled satellite IV yielded T, C and C₂T₂ as the major spots whereas dGTP labelled DNA yielded T alone.

The heterogeneity of the sequences that comprise satellite IV was examined by comparing the Tₘ of native nick translated satellite IV with the Tₘ of these sequences after melting and reassociating them (Fig. 4.5). The difference in Tₘ (ΔTₘ) between native and reassociated satellite IV was 15°C, indicating a surprisingly high degree of mismatch (15%, on an estimate of a 1°C ΔTₘ = 1% mismatch, Bonner et al., 1973).

4.4 Base Sequence and Evolution of Satellite IV

About 60% of satellite IV was restricted to below 100 bp with HinfI and the regular pattern of bands obtained showed that the HinfI sites were spaced at intervals which were higher multiples of 5 bp. Although above 100 bp fragments could not be accurately sized, it was felt likely that the same pattern of Hinf spacing was shared by these larger fragments, rather than there being any great discontinuity in the overall structure of this satellite DNA. The similarity of the pyrimidine fingerprints of fragments below 100 bp with the fingerprint of total satellite IV is in accord with this view and suggests that this satellite DNA is comprised of only one, albeit highly diverged, basic sequence.

The recognition sequence of HinfI, GANTC, is unlikely to form a basic 5 bp repeat of this DNA because very few fragments less than 25 bp are produced by digestion with this enzyme. It is felt more probable that the HinfI sites have been generated by point mutation at some time during the amplification of the sequence, thus serving to 'tag' the underlying 5 bp periodicity.
Figure 4.2

Pyrimidine tract analysis of the Hinfl 50 bp and 100 bp fragments of satellite IV.

i  a, Satellite IV was labelled with C and T by nick translation, restricted by Hinfl and electrophoresed on a 15% polyacrylamide gel;
b, nick translated φXRF DNA digested by HaeIII.

ii Pyrimidine fingerprint of the 50 bp fragment isolated from i.

iii Pyrimidine fingerprint of the 100 bp fragment isolated from i.
   1, the first dimension was developed in 2M pyridine formate pH 3.4
   2, the second dimension was developed in 1M LiCl.
Figure 4.3

Pyrimidine tract analysis of the fast and slow strands of the 45 bp Hinfl fragment of satellite IV.

i The isolated 45 bp Hinfl fragment of satellite IV was loaded onto a 20% polyacrylamide gel in 0.2M NaOH and the fast (F) and slow strands (S) separated by electrophoresis.

ii Pyrimidine fingerprint of the fast strand.

iii Pyrimidine fingerprint of the slow strand.

1, first dimension - 2M pyridine formate pH 3.4;
2, second dimension - 1M LiCl.
Figure 4.4
Pyrimidine tract analysis of satellite IV

i  Pyrimidine fingerprint of C+T labelled satellite IV

ii  Pyrimidine fingerprint of A labelled satellite IV

iii  Pyrimidine fingerprint of G labelled satellite IV
    1, first dimension - 2 M pyridine formate pH 3.4;
    2, second dimension - 1M LiCl.
<table>
<thead>
<tr>
<th>Tract</th>
<th>50 bp %P C+T Label</th>
<th>100 bp %P C+T Label</th>
<th>Sat.IV %P C+T Label</th>
<th>Sat.IV %P A Label</th>
<th>Sat.IV %P G Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>64.8</td>
<td>36.0</td>
</tr>
<tr>
<td>T₁</td>
<td>37.4</td>
<td>42.7</td>
<td>33.0</td>
<td>6.1</td>
<td>59.0</td>
</tr>
<tr>
<td>T₂</td>
<td>6.9</td>
<td>6.9</td>
<td>7.5</td>
<td>0.8</td>
<td>0.7</td>
</tr>
<tr>
<td>T₃</td>
<td>-</td>
<td>-</td>
<td>1.37</td>
<td>0.01</td>
<td>-</td>
</tr>
<tr>
<td>T₄</td>
<td>-</td>
<td>-</td>
<td>0.97</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C₁</td>
<td>8.7</td>
<td>16.5</td>
<td>9.27</td>
<td>9.2</td>
<td>3.1</td>
</tr>
<tr>
<td>C₂</td>
<td>1.9</td>
<td>2.5</td>
<td>1.60</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C₃</td>
<td>2.3</td>
<td>2.8</td>
<td>1.90</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C₁T₁</td>
<td>9.6</td>
<td>3.4</td>
<td>6.90</td>
<td>4.03</td>
<td>1.2</td>
</tr>
<tr>
<td>C₁T₂</td>
<td>5.8</td>
<td>2.2</td>
<td>5.02</td>
<td>0.9</td>
<td>-</td>
</tr>
<tr>
<td>C₁T₃</td>
<td>-</td>
<td>-</td>
<td>1.60</td>
<td>0.7</td>
<td>-</td>
</tr>
<tr>
<td>C₂T₁</td>
<td>5.8</td>
<td>5.4</td>
<td>5.05</td>
<td>1.2</td>
<td>-</td>
</tr>
<tr>
<td>C₂T₂</td>
<td>21.6</td>
<td>18.3</td>
<td>20.4</td>
<td>12.3</td>
<td>-</td>
</tr>
<tr>
<td>C₃T₁</td>
<td>-</td>
<td>-</td>
<td>1.9</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

**TABLE 4.2**

Pyrimidine Tract Analysis of the 50 bp and 100 bp Hinf Fragments and of Total Satellite IV.

The %P was estimated by cutting out each spot and counting in a scintillation counter.
The melting profiles of native nick translated satellite IV (O) and renatured nick translated satellite IV (●) were determined on a water jacketed HAP column by the method outlined in Section 10.9. Prior to renaturation the single stranded radioactive probe was checked to ensure the absence of fold back sequences. At each temperature point single stranded $^{32}$P labelled DNA was eluted by 3 x 1.0 ml washes of 0.12M PB and Cerenkov radiation was counted directly in the scintillation counter. Complete melting of the sequences was checked by a final wash of 0.3M PB which elutes both single stranded and double stranded DNA from the column.
Pyrimidine tract analysis showed that \((C_2T_2)A\) was present on one strand only and was predominant throughout the satellite sequence. However, it is unlikely that one of the isomers of this forms a basic 5 bp repeat because 5' base analysis of the \(C_2T_2\) spot shows it to be comprised of a heterogeneous mixture of \(C_2T_2\) isomers (Fig. 4.6). It is improbable that such a mixture of \(C_2T_2\) isomers could have arisen by point mutation alone from one basic 5 bp repeat. It is felt more likely that satellite IV consists of an array of related 5 bp oligonucleotides because such a structure would account for the regular propagation of \(HinfI\) sites at 5 bp intervals by unequal crossing over during the amplification of the sequence.

\(TaqI\) cuts satellite IV to yield a smear of DNA (see Fig. 5.6). Only a small proportion of the sequence is digested down to the range of small fragments that is obtained by restriction with \(HinfI\). Nevertheless, the small fraction of \(TaqI\) fragments in the \(HinfI\) size range could be resolved into the same 5 bp incremental pattern (see Fig. 7.5) suggesting that the \(TaqI\) sites are also "tagging" the underlying 5 bp periodicity. It is postulated that the \(TaqI\) sites have been introduced at a later stage of evolution of satellite IV than the \(HinfI\) sites, but have been propagated through the sequence in the same way.

No other restriction enzymes were found to significantly digest satellite IV. This reflects the simple nature of the basic underlying sequence and also indicates that it lacks any higher order periodicity. Thus satellite IV seems to resemble satellite II (Chapter 3) in having an underlying 5 bp periodicity defined with either \(HinfI\) or \(TaqI\). However, the nature of this short underlying periodicity is very different when the two satellites are compared. In addition, the sequences that comprise satellite II are arranged in higher order periodicities, whereas those that comprise satellite IV are not. The relationship between these two sequences is discussed further in Chapters 8 and 9.
5' labelled pyrimidine tracts were prepared by depurinating satellite IV and then labelling with $^{32}$P by means of polynucleotide kinase. The pyrimidine tracts (general formula $C_nT_mP_{n+m}$) were separated on PEI cellulose plates by development in 2M pyridine formate pH 3.4 followed by development in the second dimension with 1M LiCl. The $C_2T_2$ spot was eluted onto a phosphocellulose flag (Sanger et al., 1965), eluted from the flag into distilled H$_2$O, digested with snake venom phosphodiesterase and the products analysed by high pressure liquid chromatography (Section 10.8). The T and C labelled arrows indicate the position of the cold marker nucleosides, whose elution was monitored optically. For the $C_2T_2$ pyrimidine tract of satellite IV the ratio of 5'T to 5'C was estimated to be about 2 to 1.
5.1 Introduction

Satellite III bands on the light side of main band DNA in Ag+ \( \text{Cs}_2\text{SO}_4 \) gradients (Ctomeo et al., 1971). In the first cycle gradients described in this study (Section 10.2) it was found to band as a light shoulder to main band DNA and was recycled in both \( \text{Cs}_2\text{SO}_4 \) and \( \text{CsCl} \) gradients before a reasonably unimodal band of DNA \( (\rho = 1.697 \text{ g/cm}^3) \) was observed in the ultracentrifuge.

Satellite III has been studied by a number of workers. It has been shown to consist of a mixture of sequences by thermal chromatography (Ctomeo et al., 1971, Prosser et al., 1973) and by restriction enzyme analysis (Mitchell et al., 1979).

5.2 Thermal Chromatography

Figure 5.1.i shows the melting profile of native nick translated satellite III compared to that of renatured satellite III. Two components to the melt of the latter are evident, a well matched component with a \( \Delta T_m \) of 5°C and a poorly matched component with a \( \Delta T_m \) of 15°C. Similar two step melts have been observed in other satellite DNAs. For example, the reassociated duplexes of an isolated restriction fragment prepared from a satellite of \( A. sylvaticus \) melted in two distinct steps (Cooke, 1975). This was because the satellite was composed of a number of related sequence families that shared a common pattern of restriction. Poorly matched sequences were due to interfamily reassociation and well matched sequences due to intrafamily reassociation. The two step melt observed in human satellite III has been postulated to be due to such inter and intra family reassociation.
Figure 5.1
Thermal chromatography of satellite III

i. Melting profile of native nick translated satellite III (●) and renatured nick translated satellite III (○).

ii. Renatured satellite III sequences, that were poorly matched (cut a in i) and those that were well matched (cut b in i) were independently prepared by using a heated centrifuge. These were then denatured, reassociated back to an excess of cold satellite III and subsequently remelted.

●, remelt of poorly matched satellite III sequences;
○, remelt of well matched satellite III sequences.

At each temperature point $^{32}$P labelled DNA was eluted by 3 x 1.0 ml washes of 0.12M PB and Cerenkov radiation was counted directly in the scintillation counter.
(Mitchell et al., 1979). However, the results presented in Figure 5.1.ii would indicate that this is not the case. In this experiment the well matched and poorly matched sequences were reassociated back separately to satellite III. In both cases the reassociated sequences remelted off satellite III monophasically, the poorly matched sequences with a large $\Delta T_M$ and the well matched sequences with a small $\Delta T_M$. Thus satellite III is comprised of at least two unrelated sequences (at least at the level of stringency employed for reassociation, i.e. $60^\circ C$, 0.12M PB) that can be distinguished by the thermal stability of their reassociated duplexes.

5.3 Restriction Analysis

The restriction enzymes HaeIII and EcoRI have been shown to digest satellite III to a distinctive pattern of fragments (Bostock et al., 1978; Mitchell et al., 1979). Figure 5.2.i shows the human satellites I-IV restricted by HaeIII and EcoRI. Satellite III is distinguished from the other highly repeated DNAs by its restriction with HaeIII to yield a multimeric series of fragments that has a basic repeat length of 170 bp. The most prominent fragment in this series is the dimer (340 bp). Other members of the series are clearly visible up to a size of about 1870 bp (11-mer). Between the 11-mer and the DNA that is unresolved by the gel there is a prominent 3.4 kb fragment in male satellite III which is absent from HaeIII digests of female satellite III, which has been shown to be located on the Y chromosome (Bostock et al., 1978). With HaeIII about 30-40% of male satellite III is digested to the 170 bp 'ladder', 10% to the 3.4 kb Y fragment and the rest is uncut. These figures are approximate because it was noted that the relative proportions of these three components varied with different satellite III preparations.
Figure 5.2

Restriction analysis of the four human satellite DNAs

i Satellite DNAs (1.5 μg per track) were digested by HaeIII, EcoRI and MboI and the products separated by electrophoresis on a 1.5% agarose gel. a, lambda DNA digested by EcoRI and HindIII; b, satellite I/HaeIII; c, satellite II/HaeIII; d, satellite III/HaeIII; e, satellite IV/HaeIII; f, satellite I/EcoRI; g, satellite II/EcoRI; h, satellite III/EcoRI; i, satellite IV/EcoRI; j, satellite II/MboI; k, satellite III/MboI.

ii Hybridisation of the 340 bp HaeIII fragment purified from satellite III and labelled with $^{32}$P by nick translation to a Southern 'blot' of i. Hybridisation was in 3 x SSC at 65°C and the filter was finally washed in 2 x SSC at 65°C, before exposure to X-ray film.
Figure 5.2.i is a Southern 'blot' of the gel shown in Figure 5.2.i that has been hybridised to the purified 340 bp HaeIII fragment. When HaeIII restricted satellite III was probed with this fragment hybrids were formed almost exclusively with the members of the multimeric series. The fragment did not hybridise significantly with the 3.4 kb Y fragment, or with the satellite III sequences uncut by HaeIII. Nor did it significantly hybridise to any of the other human satellite DNAs. Thus HaeIII digestion defines a sequence that makes up 30-40% of satellite III. This sequence is distinct from the rest of the sequences that comprise satellite III and is not present in the other three satellite DNAs investigated. EcoRI and MboI also cut this sequence yielding fragments based on a 170 bp repeat. However, in contrast to HaeIII, these enzymes cut only a portion of the sequence and in both cases a considerable amount of it remains uncut at the origin of the gel.

5.4 The Relationship between Satellite II and Satellite III

Other sequences present in satellite III were also investigated by Southern blotting. Figure 5.3.i again shows the four human satellites restricted with HaeIII and EcoRI, and Figure 5.3.ii is the Southern 'blot' of this gel, this time after hybridisation to satellite II. EcoRI digestion of satellite III yields a number of fragments which are the same size as the EcoRI fragments of satellite II plus a number of additional fragments. The fragments held in common between the two satellites hybridised to satellite II whereas the additional fragments present in satellite III did not. In a HaeIII digest of satellite III, however, satellite II hybridised only to the DNA sequences at the origin of the gel. This indicates that some of the satellite III
Figure 5.3

Restriction analysis of the four human satellite DNAs

i Satellite DNAs (1.5 μg per track) were digested by HaeIII or EcoRI and the products separated by electrophoresis on a 1.5% agarose gel. a, lambda DNA digested with EcoRI and HaeIII; b, satellite I/HaeIII; c, satellite II/HaeIII; d, satellite III/HaeIII; e, satellite IV/HaeIII; f, satellite I/EcoRI; g, satellite II/EcoRI; h, satellite III/EcoRI; i, satellite IV/EcoRI.

ii Hybridisation of satellite II, labelled with $^{32}$P by nick translation to a Southern 'blot' of i. Hybridisation was in 3 x SSC at 65°C and the filter was finally washed in 2 x SSC at 65°C before exposure to X-ray film.
Figure 5.4
Comparison of the MboI restriction patterns of satellite II and satellite III.

i a, satellite II and b, satellite III were digested by MboI and the products separated by electrophoresis on a 1.5% agarose gel.

ii Southern 'blot' of i hybridised to satellite II labelled with $^{32}$P by nick translation. The final wash was in 2 x SSC at 65°C.
sequences that are uncut by HaeIII are satellite II. The EcoRI fragments of satellite III that do not hybridise to satellite II under the standard conditions of hybridisation (2 x SSC, 65°C) are comprised of the 3.4 kb Y fragment and a fraction of the 170 bp repeat sequence.

A similar comparison of satellite II and satellite III was performed using the restriction enzyme MboI. This cuts satellite II yielding a distinctive pattern of fragments (Section 3.5). It also cuts satellite III yielding a pattern of fragments not obviously related to that of satellite II. Nevertheless, hybridisation of nick translated satellite II to a blot of MboI restricted satellite III yields a pattern of labelling very similar to that obtained with MboI restricted satellite II (Fig. 5.4.ii). Washing the filter shown in Figure 5.4.ii at increasingly higher temperatures showed that the thermal stability of the duplexes formed between the labelled satellite II probe and the MboI fragments of either satellite DNA was the same (data not shown).

5.5 The Relationship between Satellite IV and Satellite III

A similar analysis of satellite III was also carried out using purified satellite IV. Figure 5.5.i shows Southern 'blots' of HaeIII restricted satellites I, II and III that have been hybridised to satellite IV. Satellite IV did not hybridise to satellite I or satellite II in 2 x SSC at 65°C but did hybridise to the 3.4 kb Y fragment and to the sequences uncut by HaeIII within satellite III. Figure 5.5.ii shows that satellite IV is largely uncut by HaeIII but is reduced to a smear of low molecular weight fragments upon digestion with HinfI. Similarly a HaeIII/HinfI double digest of satellite III
Figure 5.5

The relationship between satellite III and satellite IV

i  Hybridisation of $^{32}$P labelled nick translated satellite IV to HaeIII digests of a, satellite I; b, satellite II; c, satellite III.

ii Hybridisation of satellite IV to HaeIII digests of a, satellite III; b, satellite IV and to HaeIII/HinfI double digests of c, satellite III; d, satellite IV. OR indicates the sequences remaining at the origin of the gel, Y indicates the 3.4 kb Y specific fragment (see Chapter 7).

In both i and ii the final wash was in 2 x SSC at 65°C.
<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ii</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
shows that the sequences complementary to satellite IV are also
reduced to a smear of low molecular weight fragments by HinfI.
Electrophoresis on 15% or 20% acrylamide gels of HinfI digests of
total satellite III and of HinfI digests of the HaeIII resistant
satellite III sequences showed that some of the sequences were res-
stricted to a ladder of low MW fragments that increased by 5 bp.
This pattern was very similar to that observed for HinfI digests of
purified satellite IV (Section 4.2).

A heterologous reassociation of satellite IV to satellite III
and subsequent thermal chromatography showed that satellite IV melted
off its complementary sequences in satellite III in a similar man-
ner to the melting of the homologous (satellite IV/satellite IV)
reassociated sequences (data not shown).

Thus, as for satellite II, sequences similar, if not identical,
to satellite IV comprise a portion of satellite III.
Figure 5.6 is a gel of satellites II, III and IV after restriction
with six different enzymes. The comparative patterns of restriction
are consistent with the notion that both satellite II and IV comprise
a considerable fraction of satellite III.

5.6 The Sequence Composition of Satellite III

The foregoing account shows that satellite III is not a 'pure'
satellite DNA but is comprised of a number of different highly
repeated sequences that copurify at or around a buoyant density of
1.697 g/cm³ in neutral CsCl. Thus, although satellite II is prepar-
atively isolated from the heavy side of main band in first cycle
$\text{Ag}^+\text{Cs}_2\text{SO}_4$ gradients, a significant portion of this sequence co-purifies
with a number of different sequences on the light side of main band
Comparative restriction analysis of satellites II, III and IV

The patterns of restriction of satellites II, III and IV were
compared using six different restriction endonucleases.
a and t, $\phi XRF/HaeIII; b, c$ and $d$ are II, III and IV/HaeIII;
e, f and $g$ are II, III and IV/EcoRI; $h, i$ and $j$ are II, III and
IV/XbaI; $k, l$ and $m$ are II, III and IV/MboI; $n, o$ and $p$ are
II, III and IV/HinfI; $q, r$ and $s$ are II, III and IV/TaqI.
1 μg of purified satellite was loaded per track on a 1.5%
agarose gel.

The patterns of restriction are consistent with the notion
that satellite III contains sequences of both satellites II and
IV. For example, EcoRI digestion gives rise to fragments held
in common between satellites II and IV (lanes e and f) and TaqI
digestion gives rise to a smear of DNA with some diffuse bands
visible in satellite IV which also appears to be present in
satellite III (lanes r and s).
as satellite III. It is felt unlikely that the satellite II sequences identified in satellite III are significantly different from those isolated from the heavy side of main band DNA because of the resemblance of the MboI and EcoRI digests of the two satellites when probed by satellite II in Southern 'blot' experiments. Likewise, at least some of the satellite IV sequences present in satellite III would seem to be the same as those isolated as purified satellite IV because of the similarity of their melting profiles and the similarity of their fine restriction patterns. In satellite III isolated from males a fourth component, the 3.4 kb Y fragment also co-purifies with these other sequences (Bostock et al., 1978). In 2 x SSC at 65°C the Y fragment appears to have some homology with satellite IV but not with satellite II or the 170 bp repeat sequence. The structure of the Y fragment and its relationship to other human highly repeated DNAs is discussed in more detail in Chapters 7 and 8.

The initial purification of the human satellite sequences involves binding silver ions to total genomic DNA. Satellite II is prepared by virtue of the fact that it has a high affinity for silver ions and thus bands on the heavy side of main band DNA. The factors that govern silver binding to DNA are not at all well understood. It may be that the presence of satellite II sequences in a preparation of satellite III is due to the incomplete binding of the silver ions to satellite II. Certainly the variability of the composition of satellite III from one preparation to the next would support this view. This variability can be quite high, for example, one preparation of satellite III that restricted quite normally with HaeIII contained virtually no satellite II sequences, as judged by EcoRI restriction. Similarly, the presence of satellite IV sequences in satellite III may be due to the inconsistencies of the silver binding reaction.
However, it should be noted that these two satellites band close together in first cycle gradients and thus the presence of satellite IV in satellite III may be due to an inability to resolve completely satellite IV sequences from the light shoulder to the main band which comprises satellite III.

Similar experiments to those described above were also carried out with satellite I. In one preparation of satellite I significant amounts of satellite IV were detected. In other preparations no satellite IV was detected (see Fig. 5.5.i) again emphasising the somewhat variable nature of the human satellite preparations caused either by the inconsistencies of silver binding and/or the inability to completely resolve first cycle satellite bands.

The relationship between satellite III and satellites II and IV is not so much one of homology, implying related sequences, but one of co-purification of different sequences with similar buoyant densities. The results of an experiment that substantiate this view are presented in Figure 5.7. A sample of satellite III was split into two halves; one half was incubated with HaeIII and the other half incubated with the restriction enzyme buffer but no nuclease was added. After incubation the buoyant density profiles of the restricted and control DNA were analysed on the analytical ultracentrifuge. The undigested satellite III gave a reasonably sharp unimodal peak of buoyant density $1.697 \text{ g/cm}^3$, whereas that digested with HaeIII gave a bimodal peak superimposed on a spread of heterogeneously banding DNA. The buoyant densities of the two peaks were estimated to be $1.695 \text{ g/cm}^3$ and $1.699 \text{ g/cm}^3$. It is postulated that these two peaks correspond to the satellite II ($\rho = 1.694 \text{ g/cm}^3$) and the satellite IV ($\rho = 1.699 \text{ g/cm}^3$) components present in satellite III. Thus,
after HaeIII digestion the 170 bp repeat sequence and the 3.4 kb Y fragment would not be of sufficient molecular weight to band sharply in CsCl. However, satellite II and satellite IV, which are not significantly cut by HaeIII, would still be able to band at their own buoyant density but would not be masked by the predominant 1.697 g/cm$^3$ buoyant density of the digested sequences.

In conclusion satellite III is a composite satellite consisting of at least three highly repeated DNA sequences. The sequences band independently in CsCl but a mixture of them gives rise to an apparently homogeneous satellite DNA (when analysed by analytical ultracentrifugation) with an average buoyant density of about 1.697 g/cm$^3$. As judged by restriction and thermal stability the satellite II and satellite IV sequences are no different from the sequences present in the purified satellites. The complex melting behaviour of satellite III (Fig. 5.1) is thus due to the different thermal stabilities of these distinct sequences upon reassociation, e.g. reassociated satellite IV is poorly matched with a $\Delta T_M$ of 15°C whereas satellite II is moderately well matched with a $\Delta T_M$ of 8°C.
Figure 5.7
Analytical buoyant density centrifugation of satellite III before and after HaeIII digestion.

2 μg of purified satellite III was split into 2 aliquots. One half was incubated with HaeIII and the other half was incubated with the restriction enzyme buffer alone. At the end of the incubation period the samples were briefly heated to 70°C to inactivate the restriction enzyme, mixed with Micrococcus lysodeikticus DNA and their buoyant density profiles analysed on the MSE Centriscan (Section 10.2).

i Undigested satellite III, \( \rho = 1.697 \text{ g/cm}^3 \).

ii Satellite III after digestion by HaeIII. Two peaks, \( \rho = 1.695 \text{ g/cm}^3 \) and \( \rho = 1.699 \text{ g/cm}^3 \), superimposed on a background of heterogeneously banding DNA are evident.
6.1 Introduction

20-40% of satellite III consists of a sequence that is restricted by HaeIII to yield a series of fragments based on a repeat length of 170 bp. This sequence is distinct from the other sequences which copurify with satellite III (Chapter 5). It is also apparent in HaeIII digests of total human DNA as a distinctive 170 bp 'ladder' visible above the background smear of DNA (Figure 6.1.ii) and has been estimated to comprise about 2% of the total genome (Manuelidis, 1978a).

When HaeIII digests of total DNA are probed with the HaeIII 340 bp fragment isolated from satellite III the pattern of labelling is essentially the same as that seen with a 'blot' of purified satellite III. The major bands present in purified satellite III (Fig. 5.2) correspond to those present in total DNA (Fig. 6.2) and in both cases the enzyme digests all the sequences complementary to the probe. Thus the members of this sequence family which copurify as satellite III seem to be representative of the sequence family as a whole and an initial satellite purification provides a convenient means for studying them in relative isolation from the rest of the DNA of the genome.

6.2 Restriction Analysis

A number of other restriction enzymes also cut this sequence to yield a characteristic series of fragments based on a monomeric 170 bp repeat. This can be demonstrated either by restricting purified satellite III or by 'blotting' restriction digests of total DNA and probing with a specific sequence such as the 340 bp HaeIII fragment.
Figure 6.1
Visualisation of the 170 bp repeat sequence in restriction digests of satellite III and total DNA

i  HaeIII and DdeI restriction of male satellite III. a, lambda DNA/EcoRI and HindIII; b, satellite III/DdeI; c, satellite III/HaeIII; d, satellite III/HaeIII and DdeI double digest. 1.5% agarose gel - 1.5 μg of purified satellite DNA was loaded per track.

ii HaeIII and DdeI restriction of total DNA. a, lambda DNA/EcoRI and HindIII; b, male DNA/HaeIII; c, female DNA/HaeIII; d, male DNA/DdeI; e, female DNA/DdeI; f, male DNA/HaeIII and DdeI; g, female DNA/HaeIII and DdeI. 1.5% agarose gel - 3 μg of human DNA was loaded per track.
<table>
<thead>
<tr>
<th>Fragment</th>
<th>Size (bp)</th>
<th>Fraction of DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>HaeIII</td>
</tr>
<tr>
<td>1-mer</td>
<td>170</td>
<td>.008</td>
</tr>
<tr>
<td>2-mer</td>
<td>340</td>
<td>.084</td>
</tr>
<tr>
<td>3-mer</td>
<td>510</td>
<td>.020</td>
</tr>
<tr>
<td>4-mer</td>
<td>680</td>
<td>.027</td>
</tr>
<tr>
<td>5-mer</td>
<td>800</td>
<td>.002</td>
</tr>
<tr>
<td>6-mer</td>
<td>920</td>
<td>.020</td>
</tr>
<tr>
<td>7-mer</td>
<td>1090</td>
<td>.002</td>
</tr>
<tr>
<td>8-mer</td>
<td>1360</td>
<td>.032</td>
</tr>
<tr>
<td>9-mer</td>
<td>1530</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>10-mer</td>
<td>1700</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>11-mer</td>
<td>1870</td>
<td>.028</td>
</tr>
<tr>
<td>Total</td>
<td>170 bp</td>
<td>-</td>
</tr>
<tr>
<td>Y fragment</td>
<td>3400</td>
<td>.082</td>
</tr>
<tr>
<td>Uncut</td>
<td>-</td>
<td>.690</td>
</tr>
</tbody>
</table>

TABLE 6.1
Densitometry of HaeIII and DdeI Digests of Satellite III

The original negative of Figure 6.1.i was scanned on a Joyce-Loebl microdensitometer and the fraction of DNA in each fragment was estimated by cutting out and weighing the corresponding peaks.
Figure 6.2
Restriction analysis of the 170 bp repeat sequence

Total human DNA was digested with a variety of restriction enzymes, the products electrophoresed on a 1.5% agarose gel and subsequently transferred to nitrocellulose by the method of Southern. This was hybridised to the 340 bp HaeIII fragment, purified from satellite III and labelled with $^{32}$P by nick translation. 

a, lambda DNA/EcoRI and HindIII; b, male DNA/EcoRI; c, female DNA/EcoRI; d, male DNA/XbaI; e, female DNA/XbaI; f, male DNA/HaeIII; g, female DNA/HaeIII; h, female DNA/TaqI.

The final wash of the filters was in 2 x SSC at 65°C. Tracks a–g were co-electrophoresed, track h is from another gel.
Figure 6.2 shows 'blots' of total DNA restricted with a variety of enzymes that cut this sequence. All the enzymes yield fragments based on the 170 bp repeat but each enzyme has its own distinctive pattern. In contrast to HaeIII the three other enzymes investigated, EcoRI, XbaI and TaqI cut only a portion of the total sequence and a considerable amount of the sequence remains at the origin of the gel.

The restriction enzyme DdeI also cuts this sequence. Figure 6.1 compares the HaeIII and DdeI restriction patterns of satellite III and of total genomic DNA. The digestion of satellite III with DdeI is similar to that observed with HaeIII in that about 65% of the total satellite DNA remains uncut and 10% is cut to the 3.4 kb Y fragment (Table 6.1). In contrast to HaeIII, instead of a 'ladder' of restriction fragments based on a repeat length of 170 bp, the 170 bp monomer is the major fragment present. Considerably smaller amounts of the dimer and trimer are also visible. After DdeI restriction the amount of satellite III present as the monomer, dimer and trimer was almost the same as that estimated for the entire HaeIII ladder of fragments (Table 6.1), i.e. DdeI completely digests this sequence forming predominantly the fundamental monomeric repeat length. This was confirmed by the HaeIII/DdeI double digest which showed that virtually all the HaeIII ladder was cut by DdeI to the 170 bp monomer plus faster migrating fragments expected of 170 bp repeats that contained both HaeIII and DdeI sites. Hybridisation of the 340 bp HaeIII fragment to DdeI digests of total DNA also showed that DdeI was completely digesting this sequence family because no labelled material was observed at the origin of the gel (Fig. 6.3). Digestion of the end-labelled DdeI 170 bp fragment with HaeIII showed the HaeIII site to be ≈50 bp from one of the DdeI sites (Fig. 6.4).
Figure 6.3

Restriction analysis of the 170 bp repeat sequence.

The 340 bp HaeIII fragment, purified from satellite III was hybridised to a Southern 'blot' of the gel shown in Figure 6.1.ii.

α-γ, as for Figure 6.1.ii.

The final wash of the filter was in 2 x SSC at 65°C.

The low signal from the 170 bp fragment is because small DNA fragments below about 200 bp do not bind efficiently to nitrocellulose.
Satellite III was restricted by DdeI and the products end-labelled with $^{32}$P by means of polynucleotide kinase (Section 10.6). The labelled products were electrophoresed on a 1.5% agarose gel and the 170 bp fragment extracted from the gel (Section 10.4). The isolated fragment was redigested with HaeIII or HinfI and the products analysed on a 15% polyacrylamide gel.

a, *XRF DNA restricted by HinfI and end labelled with $^{32}$P by means of polynucleotide kinase; b, DdeI 170 bp monomer/undigested; c, DdeI 170 bp monomer/HaeIII; d, DdeI 170 bp monomer/HinfI.
6.3 Evolution of the Sequence

From the above section it is apparent that the 170 bp repeat sequence is characterised by at least three different types of restriction pattern. These three types of pattern provide much information on the evolution of this sequence family.

Figure 6.5 is a 'Slack plot' (Slack, 1974; Southern, 1975a) for the amounts of monomer, dimer and trimer released by DdeI digestion of satellite III. The plot of \( \log \frac{an}{n} \) against \((n-1)\) is very nearly a straight line indicating that the loss of DdeI restriction sites is essentially random and that the sequence behaves as a tandem array of 170 bp repeats, with one DdeI site per repeat, which is diverging at random by point mutation. The level of divergence in the DdeI sites was estimated to be between 3\% and 4\%.

The HaeIII restriction site is also present throughout the sequence. However, a similar 'Slack plot' for HaeIII shows that this site is not behaving in the same random fashion as the DdeI site (Fig. 6.5). Restriction of satellite III by HaeIII yields the 340 bp dimer as the most prominent fragment (on an estimate from Table 6.1 it accounts for 42\% of the total HaeIII 'ladder') and 72\% of the sequence is present as multiples of 340 bp. Thus the mode of evolution of this sequence, when considered with respect to the HaeIII sites, is consistent with the amplification of a dimeric 340 bp repeat, containing two DdeI sites but only one HaeIII site. Recently, direct sequencing studies have shown this interpretation to be correct. Wu and Manuelidis (1980) have sequenced the 340 bp EcoRI fragment isolated from total human DNA and found this fragment to consist of two 170 bp repeats (actually one of 169 bp and the other of 171 bp). Both '170' bp repeats contained a DdeI site whereas only one contained a HaeIII site. The two 170 bp arms of the repeat, although clearly related to each other, had only 74\% of their base sequence in common.
Both restriction analysis and sequencing indicate that this particular DNA family, whilst consisting of a fundamental 170 bp repeating unit, has been amplified as a 340 bp repeating unit. However, a 'Slack plot' of the HaeIII digest considering only the dimeric molecules (Fig. 6.5) whilst less erratic than that plotted for all the fragments, is still not consistent with a simple model in which HaeIII dimers have amplified and the 4-mers, 6-mers etc. been generated by random point mutation in the restriction site. Figure 6.5 shows that the HaeIII 6-mer and 8-mer are very much over represented and, of course, such an analysis takes no account of the considerable proportion of the sequence that is restricted by HaeIII to odd numbered repeats.

Considering the predominant HaeIII 340 bp repeat as the basic unit of amplification, then the odd numbered HaeIII fragments correspond to the ½-mers of such a series. The presence of ½-mers has been shown by Southern (1975a) to indicate the presence of subrepeats that might allow pairing and unequal crossing over in the ½-mer register. This allows the restriction site under study to recombine at ½-mer intervals. The DdeI 170 bp periodicity and the sequencing data provide unequivocal evidence for two subrepeats within the HaeIII 340 bp fragment and so the odd numbered HaeIII fragments may have arisen by unequal crossing over in the 170 bp register. The measured ratio of the HaeIII 170 bp fragment and the HaeIII 510 bp fragment is 1:2.5 (Table 6.1) which is close to the ratio of 1:3 predicted if these two HaeIII fragments were being generated from the 340 bp array solely by unequal crossing over.

Unequal crossing over in the 170 bp register thus allows for the generation of odd numbered HaeIII repeats. It may also account for the generation of HaeIII fragments whose relative amounts cannot be
Figure 6.5
Plot of $\log \frac{an}{n}$ vs n-l for the 170 bp repeat sequence after restriction of satellite III by DdeI and HaeIII.

The negative of the gel shown in Figure 6.1.1 was scanned with the Joyce-Loebl microdensitometer and the amount of DNA in each fragment estimated. For the estimate of $an$ (the fraction of DNA in fragment n of a multimeric series) the DNA at the origin of the gel and the 3.4 kb Y fragment were excluded as they did not hybridise to the 340 bp HaeIII fragment.

- HaeIII restricted satellite III - the monomer was taken as 170 bp;
- HaeIII restricted satellite III - the monomer was taken as 340 bp;
- DdeI restricted satellite III - the monomer was taken as 170 bp.

For random distribution the plot of $\log \frac{an}{n}$ vs n-l is a straight line whose slope is $\log P$ and intercept $2\log (1-P)$ (Slack, 1974; Southern, 1975a) where P gives the fraction of altered sites. The divergence, p, can be calculated from P assuming random divergence in the bases which comprise the restriction sites. For a 4 base restriction site $P = 1 - (1-p)^4$ (Southern, 1975a).
explained on a simple amplification/mutation hypothesis. Successive recombination of HaeIII sites by unequal cross overs between two different 170 bp repeats would generate novel HaeIII repeats, which if subsequently amplified would give rise to a non-random pattern of HaeIII digestion. Only a few recombination/amplification events would be required to generate the overall pattern of fragments seen with HaeIII. This mode of evolution allows for the generation of a non-random HaeIII pattern of digestion whilst the DdeI pattern would be unaffected because this restriction site is present on both 170 bp repeats.

The model outlined above predicts a considerable heterogeneity between different fragments from the HaeIII ladder because of recombination between the two different 170 bp repeats. Figure 6.6 compares the thermal stability of the 340 bp fragment after reassociation to itself and after reassociation to total satellite III. This fragment melts off itself monophonically with a $\Delta T_M$ of about 6°C, whereas it melts off satellite III biphasically, having a well matched component with a $\Delta T_M$ of about 4°C and a poorly matched component with a $\Delta T_M$ of about 11°C. These contrasting melting profiles indicate that the 340 bp fragment is better matched to itself than it is to the rest of the sequences to which it hybridises. A similar conclusion was reached by washing a Southern 'blot' of satellite III, hybridised to the 340 bp fragment, at increasingly higher temperatures. It was observed that duplexes formed with the 'blotted' 340 bp fragment were more stable than duplexes formed with other members of the multimeric series. Mitchell et al. (1979) have shown by the technique of two dimensional hybridisation that members of the series are usually better matched to themselves than they are to other members of the series. Such intrafragment homogeneity and interfragment heterogeneity
The 340 bp HaeIII fragment, isolated from satellite III, was labelled with $^{32}$P by nick translation. Single stranded DNA was prepared using a heated centrifuge and reassociated with excess 340 bp HaeIII fragment and excess satellite III (the % reassociation of the probe was 86% and 84% respectively). The thermal stability of the resulting duplexes was investigated using a water jacketed HAP column (Section 10.9).

- , native 340 bp fragment;
- , 340 bp fragment reassociated to itself;
- , 340 bp fragment reassociated to satellite III.

Prior to reassociation the $^{32}$P labelled single stranded probe was checked to ensure the absence of foldback sequences.
is what would be expected if recombination were giving rise to new sequence variants composed of elements from the two basic 170 bp repeats recombined with one another.

The third class of restriction enzymes characterised with respect to the restriction pattern they generate with this sequence are similar to HaeIII in that they produce fragments that are predominantly multiples of the 340 bp periodicity. They differ from HaeIII in that they cut the sequence overall much less frequently, they do not generate such a prominent 340 bp fragment and they cut only a portion of the sequence.

One possible explanation for the type of pattern seen with these enzymes is that their recognition sequence has arisen de novo on a portion or segment (Horz & Zachau, 1977) of the total repeated sequence and this has been subsequently amplified. TaqI, for example, cuts only three major fragments of the series, the 8-mer, the 16-mer and the 11-mer. It is difficult to imagine this enzyme site being present in the ancestral array with a periodicity of 340 bp as no fragments are cut to this size by the enzyme. Rather, it is felt more likely that this site rose at a later stage on a segment of the 340 bp array whose subsequent amplification then defined a new repeating periodicity. The observation that the HaeIII 11-mer and the TaqI 11-mer are very likely the same molecules (Section 8.6) is consistent with this interpretation. On the scheme outlined above the portion of this highly repeated sequence which lacks TaqI sites is due to its isolation or segmentation from the array on which the TaqI sites arose.

In conclusion, the 170 bp repeat sequence in humans exhibits three different types of restriction pattern which underly its course of evolution from a 170 bp repeat to a segmental array of repeats developing higher order periodicities. This chapter has primarily considered
the evolution of this sequence with regard to its repeating structure, determined by restriction analysis. Section 8.6 considers the topic further by considering the distribution of these repeats on different chromosomes and Section 9.3 discusses the matter in the context of related sequence families that are present in other primate species.
THE STRUCTURE OF THE 3.4 kb Y-SPECIFIC FRAGMENT

7.1 Introduction

When male satellite III is restricted with HaeIII about 10% is cut to a 3.4 kb fragment which is absent in similar digests of female satellite III (Bostock et al., 1978). This fragment was first described by Cooke (1976) by comparing HaeIII digests of total male and female DNA. Fingerprint analysis revealed it to be comprised of a simple sequence DNA.

Analysis of polymorphic Y chromosomes and 'in situ' hybridisation (Bostock et al., 1978) located this fragment on the distal fluorescent arm of the Y chromosome. Initially it was thought that the sequences which comprise this repeating fragment were specific to the Y chromosome but subsequent work (Bostock et al., 1978; Cooke and McKay, 1978; McKay et al., 1978) revealed the presence of sequences on autosomes that had some homology to it.

Densitometric measurements have estimated that the 3.4 kb fragment comprises between 0.4% and 0.5% of total genomic DNA (Cooke, 1976; Kunkel et al., 1979). On the basis that the diploid male genome is $5.3 \times 10^9$ bp in size (Sober, 1968) of which the Y chromosome constitutes about 0.8% (Cooke, 1976) then the 3.4 kb fragment would be estimated to be present in 6,200-7,800 copies and to comprise 50-60% of the total DNA of the Y chromosome.

Other Y chromosome specific sequences have been described. A 2.1 kb fragment, unrelated to the 3.4 kb fragment, is also specific to HaeIII digests of human male DNA. This fragment has recently been located on the distal tip of the Y chromosome (Schmidtke and Schmid, 1980).
Another approach to the investigation of Y specific sequences has been described by Kunkel et al. (1976) in which male specific sequences were isolated by competition hybridisation of male DNA with vast excesses of female DNA. The male specific sequences obtained by this method were intermediately repeated and apparently interspersed along the 3.4 kb repeat (Kunkel et al., 1979).

In this study the 3.4 kb fragment was isolated from purified male satellite III restricted with HaeIII. Although the yield of satellite DNA was not particularly good, initial satellite purification did ensure a 20-25 fold enrichment of the sequence. The yield of the fragment after extraction from an agarose gel (Section 10.4) was low; about 0.4 ug of the fragment was recovered from a total of 50 ug of satellite III, representing a yield of about 8%. A small amount of the purified fragment was electrophoresed on an agarose gel, to ensure purity (Fig. 7.1).

7.2 Restriction Analysis

Four restriction enzymes were used to cut the 3.4 kb repeat to yield Y specific restriction fragments. Three of these, HaeIII, EcoRI and MboI have been described before (Cooke and McKay, 1978) and a fourth, DdeI, is described here. Figure 7.2.i shows an agarose gel, stained with ethidium bromide, of total male and female DNA restricted by these four enzymes. Figures 7.2.ii and 7.2.iii are autoradiographs of the 'blot' of this gel after hybridisation to the purified HaeIII 3.4 kb fragment labelled with $^{32}$P by nick translation. Figure 7.2.ii shows the 'blot' after hybridisation and a low stringency wash (3 x SSC, 58°C) whereas Figure 7.2.iii shows the 'blot' after a further, high stringency, wash (0.12M PB, 76°C). At low stringency duplexes between the 3.4 kb fragment and sequences present in female DNA are
The 3.4 kb Y fragment was isolated from male satellite III restricted with HaeIII by electrophoresis into a dialysis bag. A small amount of this purified product was analysed by electrophoresis on a 1.5% agarose gel. a, lambda DNA/EcoRI and HindIII; b, ϕXRF DNA/HaeIII; c, purified HaeIII 3.4 kb Y fragment; d, HaeIII 3.4 kb Y fragment after HinfI restriction.

The negative of Figure 7.1 was used to scan tracks a and b on the Joyce-Loebl microdensitometer to demonstrate the relationship between log MW and mobility of DNA restriction fragments and also to show that the amount of DNA in a restriction fragment could be reliably estimated from the trace (see Section 10.3).
Figure 7.2

Hybridisation of the 3.4 kb Y fragment to total male and to total female DNA.

i Male and female DNA (3 µg per track) were restricted by the enzymes specified and the products separated by electrophoresis on a 2% agarose gel.

a and m, lambda DNA/EcoRI and HindIII; b, female DNA/HaeIII;
c, female DNA/EcoRI; d, female DNA/MboI; e, female DNA/DdeI;
f, male DNA/HaeIII; g, male DNA/EcoRI; h, male DNA/MboI;
i, male DNA/DdeI; j, male DNA/HaeIII and DdeI double digest;
k, male DNA/EcoRI and DdeI double digest; l, male DNA/MboI and DdeI double digest. After electrophoresis and staining with ethidium bromide the DNA was transferred to a nitrocellulose filter by the method of Southern.

ii The purified 3.4 kb Y fragment, labelled with $^{32}$P by nick translation, was hybridised to the Southern 'blot' of i in 3 X SSC at 58°C. The final wash was at this level of stringency, a-m, as for i. EcoRI fragments characteristic of satellite II are marked with a small open circle (lanes c and g); MboI fragments characteristic of satellite II are marked with a dash (lanes d and h).

iii The filter shown in ii was washed at 76°C in 0.12M PB and re-exposed to X-ray film until the lambda DNA markers were about the same intensity as they were in ii.
stable. At a high level of stringency most of these duplexes formed with related sequences are no longer stable and the 3.4 kb probe hybridises reasonably specifically to the 3.4 kb repeats that are present only in the male target DNA.

From Figure 7.2.iii it is apparent that virtually all the truly male specific sequences are cut by each of the four restriction enzymes because after a stringent wash very little signal remains at the origin of the gel. Three of the restriction enzymes cut the 3.4 kb repeats once, whereas MboI cuts most of them twice yielding fragments about 1730 and 1770 bp in size. The relative positions of the restriction enzyme sites were mapped by MboI digestion of the isolated HaeIII fragment and also by digestion of the kinase labelled DdeI 3.4 kb fragment, purified from satellite III, with the other three enzymes. The relative positions of the enzyme sites mapped in both ways were mutually consistent and showed that all the sites could be mapped onto the same 3.4 kb repeat, shown in Figure 7.3. This restriction map is consistent with that obtained by Cooke and McKay (1978) for the relative positions of the EcoRI, HaeIII and MboI sites on the 3.4 kb fragment isolated from total male DNA.

With HaeIII and DdeI, which cut the 3.4 kb repeat only once, there is a small amount of Y specific material present as a band that was sized approximately as 6.8 kb. This suggests that the 3.4 kb repeats may be tandemly arranged and that the 6.8 kb fragment has been generated by the loss of a restriction site. A 6.8 kb fragment is not apparent with MboI digested male DNA or DNA that has been double digested with HaeIII and DdeI or EcoRI and DdeI. Furthermore the amount of material remaining as a 3.4 kb fragment after double digestion with HaeIII and DdeI or EcoRI and DdeI (i.e. those repeats with 1 of 3 restriction sites lost) very approximately equals the amount of
Figure 7.3
Restriction map of the 3.4 kb Y fragment.

The relative positions of the HaeIII, EcoRI, MboI and DdeI sites on the 3.4 kb repeat were determined by MboI digestion of the isolated HaeIII 3.4 kb fragment - digestion products were analysed by electrophoresis followed by Southern 'blotting' and HaeIII, EcoRI and MboI digestion of the isolated DdeI fragment that had been end-labelled by polynucleotide kinase. The sizes indicated are in bp.

•, MboI; ▼, DdeI; †, EcoRI; †, HaeIII.

See also Cooke and McKay, 1978.
3.4 kb fragment seen after MboI digestion (i.e. those repeats with 1 of 3 MboI sites lost). Thus the overall pattern of restriction would suggest that in the genome this Y specific DNA is arranged as a tandem array of at least two 3.4 kb repeats each carrying 1 HaeIII, 1 EcoRI, 1 DdeI and 2 MboI sites per repeat. There is heterogeneity in the array as evidenced by the loss of restriction sites and this appears to be about equal for the different sites. No heterogeneity in terms of the relative positions of the 5 restriction sites was observed.

7.3 Pyrimidine Tract Analysis

The pyrimidine fingerprints of the 3.4 kb fragment after CT, A or G labelling are shown in Figure 7.4. After CT and A labelling $T_1$ is the predominantly labelled pyrimidine tract; $C_2T_2$ and $C_1$ are the next most heavily labelled tracts. After G labelling $T_1$ is heavily labelled, but other pyrimidine tracts are only faintly labelled. $C_2T_2$, so prominent with CT or A labelled DNA, is not labelled by G. The presence of just a few predominantly labelled spots is characteristic of a simple sequence DNA and, indeed, the fingerprints of the 3.4 kb fragment are very similar to those observed for satellite IV (Section 4.3) suggesting that satellite IV and the 3.4 kb fragment may be related. The fingerprints of these two sequences are compared in Table 7.1.

7.4 The Relationship between the 3.4 kb Y Fragment and Other Highly Repeated Sequences

Figure 7.3 ii shows that at low stringency the 3.4 kb Y fragment hybridises with sequences that are present in both male and female DNA. One family of sequences to which it hybridises under these conditions can be identified by its pattern of restriction with MboI and
Figure 7.4
Pyrimidine tract analysis of the 3.4 kb Y fragment

The purified HaeIII 3.4 kb Y fragment was labelled by nick translation with dCTP + dTTP, dATP or dGTP, depurinated and the pyrimidine tracts separated by thin layer chromatography by the method outlined in Section 10.7.

i  dCTP + dTTP labelled Y fragment.

ii  dATP labelled Y fragment.

iii dGTP labelled Y fragment.

1, the first dimension was developed in 2M pyridine formate, pH 3.4; 2, the second dimension was developed in 1M LiCl.

The %P in the spots was estimated by cutting them out and counting in a scintillation counter. These values are compared to those outlined for satellite IV in Table 7.1.
<table>
<thead>
<tr>
<th>Tract</th>
<th>Y Frag. %P</th>
<th>Sat. IV %P</th>
<th>Y Frag. %P</th>
<th>Sat. IV %P</th>
<th>Y Frag. %P</th>
<th>Sat. IV %P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C+T Label</td>
<td>C+T Label</td>
<td>A Label</td>
<td>A Label</td>
<td>G Label</td>
<td>G Label</td>
</tr>
<tr>
<td>P</td>
<td>-</td>
<td>-</td>
<td>52.0</td>
<td>64.8</td>
<td>49.5</td>
<td>36</td>
</tr>
<tr>
<td>T₁</td>
<td>44.6</td>
<td>33.0</td>
<td>10.5</td>
<td>6.1</td>
<td>39.4</td>
<td>59</td>
</tr>
<tr>
<td>T₂</td>
<td>9.1</td>
<td>7.5</td>
<td>2.8</td>
<td>0.8</td>
<td>3.2</td>
<td>0.7</td>
</tr>
<tr>
<td>T₃</td>
<td>0.8</td>
<td>1.37</td>
<td>0.9</td>
<td>0.01</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>T₄</td>
<td>-</td>
<td>0.97</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C₁</td>
<td>14.7</td>
<td>9.27</td>
<td>17.7</td>
<td>9.2</td>
<td>3.7</td>
<td>3.1</td>
</tr>
<tr>
<td>C₂</td>
<td>0.7</td>
<td>1.6</td>
<td>0.5</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C₃</td>
<td>-</td>
<td>1.9</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C₁₁</td>
<td>6.1</td>
<td>6.9</td>
<td>2.9</td>
<td>4.0</td>
<td>4.0</td>
<td>1.2</td>
</tr>
<tr>
<td>C₁₂</td>
<td>3.9</td>
<td>5.02</td>
<td>2.2</td>
<td>0.9</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C₁₃</td>
<td>0.9</td>
<td>1.6</td>
<td>1.7</td>
<td>0.7</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C₂₁</td>
<td>2.9</td>
<td>5.1</td>
<td>-</td>
<td>1.2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C₂₂</td>
<td>16.6</td>
<td>20.4</td>
<td>8.7</td>
<td>12.3</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C₃₁</td>
<td>0.7</td>
<td>1.9</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

**TABLE 7.1**

Comparison of the Pyrimidine Fingerprints of the 3.4 kb Y Fragment and Satellite IV.

The %P was estimated by cutting out each spot and counting in a scintillation counter.
EcoRI. Thus the complicated pattern of fragments down to 360 and 330 bp seen after MboI digestion, and the 1300, 1770 and 2240 bp fragments seen after EcoRI digestion are both characteristic of satellite II.

Another prominent family of sequences evident in both male and female DNA to which the 3.4 kb Y fragment hybridises at low stringency is cut by MboI to yield fragments of about 1.8 and 3.4 kb. This family of sequences is discussed further in Section 8.5.

The 3.4 kb Y fragment also hybridises at low stringency with sequences that are not cut by any of the four restriction enzymes and remain at the origin of the gel. Some of these sequences are undoubtedly satellite IV because this sequence is also resistant to digestion by the four enzymes employed (Section 4.4), has a similar simple pyrimidine fingerprint to the 3.4 kb Y fragment and hybridises to it at normal stringency (Section 5.5). Satellite IV has an underlying 5 bp periodicity that can be observed after HinfI and, to a lesser extent, TaqI digestion. It was, therefore, of interest to see if the 3.4 kb Y fragment had a similar pattern of restriction with these enzymes.

After digestion by HinfI and electrophoresis on agarose gels the 3.4 kb Y fragment shows a smear of low MW fragments similar to that observed after HinfI digestion of satellite IV (Fig. 7.1). Figure 7.5 compares nick translated 3.4 kb Y fragment with nick translated satellite IV after digestion by HinfI and TaqI and electrophoresis on a 20% polyacrylamide gel. The HinfI patterns are similar in that virtually all the bands produced appear to be related to one another by increments of 5 bp. However, the fine restriction pattern of the 3.4 kb fragment differs from that of satellite IV in terms of the relative proportions of DNA in each band, in particular prominent...
Figure 7.5

HinfI and TaqI restriction of the 3.4 kb Y fragment and satellite IV.

Satellite IV and the 3.4 kb Y fragment were labelled with $^{32}$P by nick translation. The labelled DNA was digested by HinfI or TaqI and the products separated on a 40 cm 20% polyacrylamide gel. 

- a, Y fragment/HinfI; 
- b, satellite IV/HinfI; 
- c, Y fragment/TaqI; 
- d, satellite IV/TaqI.

The sizes indicated on the right of the gel refer to the HinfI digests of satellite IV. XC, xylene cyanol marker.

Note that a TaqI restriction fragment has a slightly greater mobility than a HinfI fragment of the same size (see Section 10.4).
10 and 15 bp fragments were observed with the 3.4 kb fragment (not shown). Furthermore, TaqI gave a similar pattern of restriction to that observed with HinfI, whereas only very faint bands were observed with TaqI digests of satellite IV. In the 3.4 kb fragment the HinfI and TaqI sites probably overlap because of the great similarity of the two restriction patterns and also because double digestion did not noticeably digest this sequence any more than did digestion with a single enzyme (data not shown).

The 3.4 kb Y fragment is related to a number of other highly repeated DNA sequences. It is distantly related to satellite II, hybridising to it at low stringency. However, at the standard level of stringency (2 x SSC, 65°C) these two sequences do not form stable duplexes (Section 5.4, see also Section 8.5). The 3.4 kb fragment and satellite IV do form stable duplexes in 2 x SSC at 65°C and are therefore more closely related. These two sequences share a similar fine structure as evidenced by their pyrimidine fingerprints and fine restriction pattern that is characterised by an underlying 5 bp periodicity. The sequences that comprise satellite IV are probably not arranged with a higher order repeat periodicity (Section 4.4) whereas the related Y chromosome sequences are, of course, arranged in 3.4 kb repeats.

7.5 Renaturation Kinetics

The nature of the 3.4 kb fragment and its relationship to other repeated sequences, in particular satellite IV, were further investigated by an analysis of renaturation kinetics. The 3.4 kb Y fragment and satellite IV both appear simple in terms of their pyrimidine fingerprints and underlying 5 bp periodicity, so they might be expected to exhibit the very fast rates of reassociation that are
DNA reassociation was carried out by the methods outlined in Section 10.9 and the % reassociation was plotted as a function of the Cot. Cot is the product of the initial concentration of single stranded DNA and the time of incubation and has units of moles litre$^{-1}$sec. In all the experiments described the % reassociation plotted for a given Cot is an average of at least two determinations.

i  Reassociation of the 3.4 kb Y fragment. Trace amounts of $^{32}$P labelled 3.4 kb Y fragment were driven by an excess of unlabelled 3.4 kb Y fragment in 0.12M PB at 60°C.

ii  Reassociation of satellite IV and the 3.4 kb Y fragment. $^{32}$P labelled Y fragment DNA and $^3$H labelled satellite IV were driven with an excess of unlabelled satellite IV in 0.12N PB at 60°C. The % reassociation was determined as a function of the Cot. Samples at a given time point were usually counted simultaneously using two channels of the Packard Tri-Carb scintillation counter, although some points were determined independantly. At the settings employed 1.2% of the $^{32}$P counts were detected in the $^3$H channel; no $^3$H counts were detected in the $^{32}$P channel. o, $^{32}$P labelled 3.4 kb Y fragment driven by satellite IV; o, $^3$H satellite IV driven by satellite IV. The self reassociation of the $^{32}$P labelled Y fragments (A) was determined in the absence of driver DNA and the % reassociation at a given time point plotted as a function of the Cot of the driver DNA.
Reassociation of the 3.4 kb Y fragment and satellite IV with total female DNA and total male DNA.

$^3$H satellite IV and $^{32}$P 3.4 kb Y fragment DNA were driven by an excess of total female or total male DNA in 0.12M PB at 60°C by the methods outlined in Section 10.9. The % reassociation of the $^{32}$P and $^3$H labelled probes was usually determined simultaneously using two channels of the Packard Tri-Carb scintillation counter, although some points were determined independently. The self reassociation of the two probes was determined in the absence of driver DNA and was plotted as a function of the Cot of the driver DNA.

i Female DNA driven reassociation.

- $^3$H satellite IV;
- $^{32}$P 3.4 kb Y fragment;
- $\Delta$, self reassociation of $^3$H satellite IV;
- $\Delta$, self reassociation of $^{32}$P 3.4 kb Y fragment.

ii Male DNA driven reassociation.

- $^3$H satellite IV;
- $^{32}$P 3.4 kb Y fragment;
- $\Delta$, self reassociation of $^3$H satellite IV;
- $\Delta$, self reassociation of 3.4 kb Y fragment.
characteristic of highly repeated DNAs of low complexity (Britten and Kohne, 1968).

Satellite IV when driven by itself (Fig. 7.6.ii) or by total genomic DNA (Fig. 7.7) reassociates over two decades of the Cot curve. The sequence has a rapid rate of reassociation with a Cot of about $7 \times 10^{-4}$. When driven by total male or female DNA the Cot was estimated to be about $6 \times 10^{-2}$. These figures, when corrected for the effect of mismatch on the rate of reassociation (Southern, 1971) would estimate satellite IV to have a complexity of about 33 bp and to be present in $1.5 - 2.0 \times 10^6$ copies per diploid genome (Table 7.2).

In contrast to satellite IV the 3.4 kb fragment, whether driven by itself (Fig. 7.6.i) or by total male DNA (Fig. 7.7.i) reassociates over at least four decades of the Cot curve and therefore appears to be comprised of at least two sequences of very different repetition frequency. At least two sequence components are also apparent in the melting behaviour of the 3.4 kb fragment (Section 7.6). On the simplest analysis the self reassociation of the isolated 3.4 kb fragment can be considered to have a fast component with a Cot of about $5 \times 10^{-3}$ and a slow component with a Cot of $1 \times 10^0$. Table 7.2 shows the calculated repetition frequency of the fast and slow components and it can be seen that the rapidly reassociating sequences are reiterated sufficiently to be present several times per 3.4 kb molecule, whereas the slowly reassociating sequences would be estimated to be repeated only once every 100 3.4 kb molecules.

When driven by total male DNA the 3.4 kb fragment also exhibits a fast and slow component to the reassociation (Fig. 7.7.ii). About 50% of the trace molecules reassociate with a Cot of $10^{-7}$ (driven by sequences repeated about $1.1 \times 10^6$ times per diploid genome - Table 7.2) and the other 50% reassociate with a Cot of $3 \times 10^1$
(driven by sequences repeated about 520 times per diploid genome - Table 7.2). At 60°C in 0.12M PB the slowly reassociating sequences are specific to male DNA, as evidenced by a similar trace/driver experiment with female DNA (Fig. 7.7.i). When the 3.4 kb Y fragment is driven by female DNA about 50% of the trace molecules reassociate with a Cot\textsuperscript{2}\textsubscript{1/2} of 1.5 x 10\textsuperscript{-1} but female DNA fails to drive the probe any further. However, when the experiment is repeated at 50°C (Fig. 7.8.i) female DNA does drive the 3.4 kb Y fragment probe to completion showing that there are DNA sequences present in the female genome which are distantly related to the slowly reassociating 'Y specific' sequences.

Assuming the 3.4 kb fragment to be present in 7000 copies per male genome (Section 7.1), then the fast component of the self reassociation can be estimated to be reiterated about 14.5 x 7000 = 100,000 copies genome and the slow component to be reiterated about 0.01 x 7000 = 70 copies genome (Table 7.2). When the 3.4 kb fragment is driven by total male DNA the slowly reassociating 'Y specific' component is driven by sequences estimated to be reiterated about 520 times per genome and the rapidly reassociating non specific sequences estimated to be driven by sequences reiterated about 10\textsuperscript{6} times per genome. It would seem unlikely that the rapidly reassociating sequences observed in the self reassociation (100,000 copies genome) correspond to the slowly reassociating 'Y specific' sequences (540 copies genome) observed in the total male driven reassociation. It would seem more likely that the slowly reassociating 'Y specific' sequences correspond to the slowly reassociating sequences seen in the self reassociation. However, there is quite a large disparity between the two estimates of the reiteration frequency (540 and 70 copies genome) of the slowly reassociating sequences. This disparity may be due to the considerable errors inherent in making these estimates, not least the factors
TABLE 7.2

Summary of the Reassociation Data for the 3.4 kb Y Fragment and Satellite IV.

(a) Corrected for mismatch according to the equation \( \log \frac{R_m}{R_0} = n \log (1-p) \) (Southern, 1971)

where \( R_m \) = the measured rate of reassociation, \( R_0 \) = the corrected rate, \( n \) = a constant taken in this calculation to equal 15, and \( p \) = the degree of mismatch. The mismatch was estimated from the \( \Delta T_M \) on the basis of a 1°C \( \Delta T_M \) = 1% mismatch (Bonner et al., 1973).

(b) Reiteration per diploid genome. Estimated relative to the Cot\( \frac{1}{2} \) of E. Coli taken to be 8 mole l\(^{-1}\).sec (from Mitchell et al., 1979; reassociation experiments were performed in a similar manner) and assuming the genome size of E. Coli to be \( 4.25 \times 10^6 \) bp (Cairns, 1963) and the diploid human genome size to be \( 5.3 \times 10^9 \) bp (Sober, 1968).

(c) Reiteration per 3.4 kb molecule; estimated from the ratio of the corrected Cot\( \frac{1}{2} \) to the calculated Cot\( \frac{1}{2} \) for a sequence with a complexity of 3.4 kb = \( \frac{8 \times 3.4 \times 10^3}{4.25 \times 10^6} = 6.4 \times 10^{-3} \) mole l\(^{-1}\).sec.

(d) Reiteration per diploid male genome calculated by multiplying the reiteration per 3.4 kb molecule by the copy number of 3.4 kb repeats, taken to be 7000 (Section 7.1).

(e) Complexity of satellite IV relative to E. Coli DNA.

(f) Fast renaturing component of the 3.4 kb fragment. (g) Slow renaturing component of the 3.4 kb fragment.
<table>
<thead>
<tr>
<th>Driver DNA</th>
<th>Trace</th>
<th>( \text{Cot}_{1/2} ) (mole ( 1^{-1} ) sec)</th>
<th>( \Delta T_M ) (°C)</th>
<th>(a) Corrected ( \text{Cot}_{1/2} ) (mole ( 1^{-1} ) sec)</th>
<th>Reiteration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male DNA</td>
<td>(^{32}\text{P HaeIII} )</td>
<td>( 1 \times 10^{-1} ) (f)</td>
<td>15</td>
<td>( \approx 9 \times 10^{-3} )</td>
<td>( \approx 1.1 \times 10^6 ) (b)</td>
</tr>
<tr>
<td>Male DNA</td>
<td>Y fragment</td>
<td>( 3 \times 10^{1} ) (g)</td>
<td>3</td>
<td>( \approx 1.9 \times 10^{1} )</td>
<td>( \approx 5.2 \times 10^{2} ) (b)</td>
</tr>
<tr>
<td>Male DNA</td>
<td>(^{3}\text{H Sat. IV} )</td>
<td>( 5 \times 10^{-2} )</td>
<td>14</td>
<td>( \approx 4.5 \times 10^{-3} )</td>
<td>( \approx 2 \times 10^{6} ) (b)</td>
</tr>
<tr>
<td>Female DNA</td>
<td>(^{32}\text{P HaeIII} )</td>
<td>( 1.5 \times 10^{-1} )</td>
<td>15</td>
<td>( \approx 1.3 \times 10^{-2} )</td>
<td>( \approx 6.8 \times 10^{5} ) (b)</td>
</tr>
<tr>
<td>Female DNA</td>
<td>Y fragment</td>
<td>( 7 \times 10^{-2} )</td>
<td>15</td>
<td>( \approx 6.3 \times 10^{-3} )</td>
<td>( \approx 1.5 \times 10^{6} ) (b)</td>
</tr>
<tr>
<td>HaeIII</td>
<td>(^{32}\text{P HaeIII} )</td>
<td>( 5 \times 10^{-3} ) (f)</td>
<td>15</td>
<td>( \approx 4.4 \times 10^{-4} )</td>
<td>( \approx 1.45 \times 10^{1} ) (c) (1.01 ( \times 10^{5} )) (d)</td>
</tr>
<tr>
<td>HaeIII</td>
<td>Y fragment</td>
<td>( 1 \times 10^{0} ) (g)</td>
<td>3</td>
<td>( \approx 6.3 \times 10^{-1} )</td>
<td>( \approx 1 \times 10^{-2} ) (c) (7.0 ( \times 10^{1} )) (d)</td>
</tr>
<tr>
<td>Sat. IV</td>
<td>(^{32}\text{P HaeIII} )</td>
<td>( 1.5 \times 10^{-3} )</td>
<td>17</td>
<td>( \approx 9 \times 10^{-4} )</td>
<td></td>
</tr>
<tr>
<td>Sat. IV</td>
<td>(^{3}\text{H Sat. IV} )</td>
<td>( 7 \times 10^{-4} )</td>
<td>15</td>
<td>( \approx 6.2 \times 10^{-5} )</td>
<td>33 bp (e)</td>
</tr>
</tbody>
</table>

101
Figure 7.8

Reassociation of the 3.4 kb Y fragment with female DNA and satellite IV at 50°C.

i. $^{32}$P labelled 3.4 kb Y fragment DNA was driven with an excess of total female DNA in 0.12M PB at 50°C. Double stranded and single stranded DNA were separated by the methods outlined in Section 10.9, except that the operations were performed at 50°C.

$\Delta$, self reassociation of $^{32}$P 3.4 kb Y fragment determined at 50°C in the absence of driver DNA.

ii. $^{32}$P labelled 3.4 kb Y fragment DNA was driven with an excess of purified satellite IV in 0.12M PB at 50°C.

$\circ$, $^{32}$P 3.4 kb Y fragment driven by satellite IV;

$\Lambda$, self reassociation of $^{32}$P 3.4 kb Y fragment.
involved in correcting the rate of reassociation for sequence divergence. Also, in the self reassociation of the 3.4 kb fragment 100% reassociation was not observed and, therefore, an accurate estimate of the Cot\textsubscript{2} of the slow component was probably not obtained. Similar results for the kinetics of reassociation of the 3.4 kb fragment have been published by Kunkel et al. (1979). These workers also showed that the fast and slow components of both types of reassociation (self and trace/driver) were equivalent and correspond to the non specific and the 'Y specific' sequences. Furthermore, they demonstrated by a comparison of HAP and nuclease S1 assayed reassociation that these two classes of sequence were interspersed on the same 3.4 kb molecule.

The cross hybridisation on filters of the 3.4 kb fragment and satellite IV (Section 5.5) and their similar fine structures suggest that it might be the latter which is responsible for the fast non-specific reassociation of the 3.4 kb fragment when it is driven by total female DNA. Figure 7.6.ii shows the reassociation of trace amounts of the 3.4 kb fragment when driven by satellite IV. Only a small proportion of the trace sequences (about 20%) are driven by satellite IV indicating that, although some of the sequences which drive the 3.4 kb Y fragment can be accounted for by satellite IV there are also other sequences present in female DNA which drive it under the conditions employed (0.12M PB, 60°C). When the temperature of reassociation is lowered to 50°C about 40% of the trace sequence is driven by satellite IV, indicating that there are also other, more distantly related, sequences present on the 3.4 kb Y fragment (Fig. 7.8.ii).

7.6 Thermal Chromatography

The heterogeneous nature of the 3.4 kb Y fragment was confirmed by thermal chromatography. Figure 7.9.i shows that the renatured
The $^{32}$P labelled 3.4 kb Y fragment was denatured and single stranded DNA isolated using a heated centrifuge. This was reassociated with an excess of various driver DNAs and the thermal stability of the renatured duplexes investigated using a water jacketed HAP column (Section 10.9).

- native 3.4 kb Y fragment;
- reassociated with unlabelled 3.4 kb Y fragment to Cot of $5 \times 10^{-1}$;
- reassociated with total male DNA to a Cot of $2 \times 10^{2}$;
- reassociated with total female DNA to a Cot of $2 \times 10^{2}$;
- reassociated with satellite IV to a Cot of $2 \times 10^{-2}$.

The $^{32}$P labelled 3.4 kb fragment was driven with excess total male DNA to a Cot of $10^0$ and to a Cot of $2 \times 10^{2}$ and the thermal stability of the duplexes investigated using a water jacketed HAP column.

- thermal stability of duplexes after reassociation to a Cot of $10^0$;
- thermal stability of duplexes after reassociation to a Cot of $2 \times 10^{2}$.

Prior to reassociation the $^{32}$P labelled single stranded probe was checked to ensure the absence of foldback sequences.
3.4 kb fragment has a two step melt, indicating the presence of both poorly matched duplexes with a $\Delta T_M$ of about 15°C and well matched sequences with a $\Delta T_M$ of about 3°C. The 3.4 kb Y fragment, when melted off total male DNA, did not exhibit such an obviously biophasic melt as the self reassociated sequences. However, the duplexes formed between the 3.4 kb fragment and total male DNA were better matched than those formed between the 3.4 kb fragment and female DNA, which exhibited a monophasic melt with a $\Delta T_M$ of about 15°C (Fig. 7.9.i).

The poorly matched component of the renatured 3.4 kb fragment and the duplexes formed between it and total female DNA have about the same large $\Delta T_M$ as renatured satellite IV. The duplexes formed between satellite IV and the 3.4 kb fragment are also poorly matched, having a large $\Delta T_M$ estimated to be about 17°C (Fig. 7.9.i).

The sequences of the 3.4 kb Y fragment that form poorly matched duplexes with total male DNA are the rapidly reassociating non specific sequences. This was shown by driving a labelled 3.4 kb Y fragment trace with total male DNA to a Cot at which only highly repeated sequences would anneal. Melting of these duplexes showed an increase in the proportion of poorly matched sequences when compared to the melting profile of more completely reassociated sequences (Fig. 7.9.ii).

When analysed by thermal chromatography or by the kinetics of reassociation the sequences that comprise the 3.4 kb Y fragment appear very heterogeneous. This repeating fragment is comprised of sequences held in common to the whole population of molecules interspersed with sequences that are present on only a portion of the total population of molecules (Kunkel et al., 1979). Sequences related to those sequences present on all the 3.4 kb repeats are also present in large copy numbers elsewhere in the genome and hybridise to this fragment forming
poorly matched duplexes. Some of these sequences are probably satellite IV, although this particular family of repeated sequences cannot account for all the sequences present in the female genome that are capable of hybridising to the 3.4 kb Y fragment under standard conditions. In contrast the sequences that are present on only a fraction of the 3.4 kb repeats are specific to the Y chromosome (in 0.12M PB at 60°C) and reassociate to form well matched duplexes.

In conclusion, the 3.4 kb Y fragment presents a number of paradoxical features. From the restriction analysis it appears to be a tandem array of repeats on the Y chromosome that exhibit a small amount of heterogeneity, as indicated by the loss of restriction sites that define the 3.4 kb periodicity. This loss seems roughly equal for the restriction sites examined and no drastic rearrangement of these sites is apparent. The fine structure of the 3.4 kb repeat as evidenced from its pyrimidine fingerprints and restriction pattern with HinfI and TaqI is consistent with the notion that it is comprised of a simple sequence DNA having an underlying 5 bp periodicity. The fine structure of this sequence is similar to that observed for satellite IV and it is postulated that both sequences have been derived from a common ancestral sequence. The restriction and pyrimidine fingerprint analysis, however, gives no indication of the large amount of heterogeneity that is observed by an analysis of reassociation kinetics and thermal chromatography.

The 3.4 kb Y fragment is absent from the great apes although sequences complementary to it have been detected at autosomal locations in both chimps and gorillas (Szabo et al., 1979) and so it seems fair to conclude that it has originated as a relatively recent amplification on the human Y chromosome. It is unlikely that a number of different 3.4 kb molecules, all carrying the same restriction sites
(i.e. HaeIII, EcoRI, DdeI and 2 x MboI) in the same relative positions, could have arisen independently. The present array of repeats most probably evolved from an ancestral 3.4 kb molecule containing the five restriction sites and comprised of a simple sequence DNA with an underlying 5 bp periodicity. Presumably the amplification of the 3.4 kb repeat on the human Y chromosome was accompanied by divergence and/or rearrangements that created a series of non hybridising sequence families interspersed between the more generally represented sequences.
8.1 Introduction

In the previous chapters a number of different highly repeated human DNA sequences have been characterised. Where possible, the fine structure of these sequences has been related to a possible course of evolution. In this chapter the location and arrangement of these sequences on the different chromosomes of the human genome is described. The main purpose of this was to assess the extent to which these repeated sequences evolve independently at different chromosomal locations, and whether such isolation is an important consideration for the fuller understanding of their structure, evolution and possible function.

Two main approaches have been taken, in situ hybridisation to metaphase spreads of human chromosomes (Section 10.10) and 'blot' hybridisation to human/mouse or human/chinese hamster cell line DNA (these cell lines contain only one or, at most a few human chromosomes in a mouse or hamster cell line).

In situ hybridisation (Pardue and Gall, 1970) allows the visualisation of sequences complementary to a radioactive probe directly on the chromosomes, which can be identified by Q-banding (Caspersen et al., 1971) either before or after the hybridisation reaction (Gosden et al., 1975; Lawrie and Gosden, 1980). This technique has a number of drawbacks in the analysis of repeated DNA sequences. These are the unknown factors inherent in hybridising to sequences present in the remnants of a chromosome structure and whether or not the target DNA on a given chromosome is exactly the same sequence as the purified probe or a related cross-reacting sequence. The second approach, hybridising labelled DNA to Southern 'blots' of hybrid cell line DNA avoids some of these problems. With this approach the precise nature of the target DNA in
terms of its pattern of restriction can be determined. When the sequence occurs on more than one chromosome and these are present in different hybrid cell lines then the distribution of restriction sites can be compared from one location to the next. A different distribution of restriction sites on different chromosomes has been demonstrated for total human satellite III by Beauchamp et al. (1979). As discussed in Chapter 5, satellite III is comprised of a number of copurifying sequences and although these workers established differences between chromosomes, the sequence families underlying these differences were not well defined. In the work described below, well characterised sequences have been used both for in situ experiments and for 'blot' hybridisations to hybrid cell line DNA. The combination of these two approaches has allowed a detailed description of the location, inter-relationship and evolution of the human highly repeated sequence families whose structures are described in the previous chapters.

The use of hybrid cell lines is a powerful tool in the study of highly repeated DNA. Nevertheless, results obtained using them must be viewed with some caution because it is often extremely difficult to determine the precise composition of such cell lines. This is especially so when the chromosomes are not maintained in the line by any selective pressure and the chromosomal constitution may vary from one clone to the next. Another consideration is the stability of highly repeated sequences on the chromosomes. A recent study has shown that highly repeated sequences can be lost during long term culture of human fibroblasts (Shmookler-Reis and Goldstein, 1980) without any evident loss of chromosomes and it would seem likely that similar losses will occur during the culture of a human chromosome in rodent cells.

In this study eight hybrid cell lines were employed (Table 8.1). They did not cover all the chromosomes present in the human complement but the autosomes that contain the bulk of human highly repeated DNA
TABLE 8.1  Enzyme marker and chromosome analysis of hybrid cell lines.

The enzyme analysis was carried out by Dr. S. Povey at the Galton Laboratory and the cytogenetic analysis was carried out by R. Buckland at the MRC, C.A.P.C.U. SK 82 was derived from a CHO A23 x human fibroblast fusion and contains a human 1, 17 (q 12; q21) translocation. The other cell lines were derived from mouse x human fusions. PG 25 is tetraploid and was the gift of Dr. P. Pearson. The Horl series of cells were the gift of Dr. V. van Heyningen. Horl 9.X (not shown in the Table) carries only the human X chromosome and is positive for G6PD. The X/9 cell line (not shown in the Table) contains an X, 9 translocation consisting of the long arm and C band of chromosome 9. It was the gift of Dr. T. Mohandis and DNA from the cell line was kindly supplied by A. Mitchell.

*Indicates the fraction of cells containing the chromosome.
| Chromosome | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | X | Y |
|------------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| Enzyme     | DN01- | IDH1- | ACY1- | HEXB- | ME1- | GOT1- | LDH4+ | PEPS- | NP- | MP1- | APRT+ | PEPA+ | GP1+ | ADA+ | SOD1+ | G6PD+ | >.9 | >.9 | >.9 | GPRM | 1.6 |
| Markers    | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- |
| NOSL 9.15  | CHR. | a | ANALYSIS | |    | 0.5 | >.9 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Enzyme     | PGM1- | IDH1- | ACY1- | HEXB- | ME1- | AK1- | GOT1- | LDHA- | LDHB- | PEPB- |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Markers    | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- |
| SK 82      | CHR. | a | ANALYSIS | |    | >.9 | <.1 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Enzyme     | PGD+ | ACP1+ |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Markers    | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- |
| PG 25      | CHR. | a | ANALYSIS | |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Enzyme     | PGM1- | MHI1- |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Markers    | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- | PEPC- |

**Note:** The table represents a genetic analysis involving various enzymes and markers. The values indicate genetic distances or frequencies associated with these genetic markers.
8.2 Satellite I

There have been a number of \textit{in situ} hybridisation studies describing the location of human satellite I. Gosden et al. (1975) located this sequence primarily on the C band of chromosome 9 with secondary sites of hybridisation on the Y and the short arms of the acrocentric chromosomes. Jones et al. (1974) and Manuelides (1978a), however, reported the main site of hybridisation of this sequence as the long arm of the Y chromosome but also reported considerable hybridisation to the short arm of the acrocentric chromosomes and to the C band of some of the metacentric chromosomes, in particular chromosome 9.

The results of an \textit{in situ} hybridisation experiment using nick-translated satellite I are presented in Fig. 8.1. These results are broadly in agreement with the findings of Jones et al. (1974) and Manuelides (1978b) although there are some differences. The sequence was found to hybridise primarily to the Y chromosome and to the short arm of four of the five acrocentric chromosomes.

The presence of sequences complementary to satellite I was investigated in the DNA of the hybrid cell lines. Unfortunately, no restriction enzyme has been found that usefully cuts satellite I (Section 2.2). Consequently this sequence's resistance to digestion by restriction endonucleases was used as its major distinguishing feature and the hybrid cell line DNAs were tested for HaeIII and HinfI resistant sequences which hybridised to a satellite I probe. The probe hybridised strongly to only one cell line, PG25, although a weak signal was seen from three others. PG25 contains chromosome 22, a chromosome 13 and a chromosome 9 (Table 8.1) and from the \textit{in situ} results this cell line might be expected to contain considerable amounts of this sequence. However, the intensity of the signal compared to that from other cell
Figure 8.1
In situ hybridisation of satellite I

Satellite I, labelled with $^3$H by nick translation was hybridised to human metaphase spreads by the methods described in Section 10.10. In i and ii the Q-banded spread was pre-selected, photographed with U.V. light and then the quinacrine washed out with running tap water prior to hybridisation. After hybridisation and autoradiography the spreads were stained in a 10% solution of Gurr's Giemsa dissolved in Gurr's buffer for 10-20 minutes, the spreads relocated and then photographed with visible light.

i  Q-banded human metaphase spread

ii  The same spread after hybridisation to $^3$H satellite I and Giemsa staining. 12 week exposure, photographed with visible light.

iii  Cumulative grain count for a total of 17 spreads some of which were analysed as described above and some of which were analysed by simultaneous observation of the Q-banding and silver-grains (see Section 10.10).
lines is surprising. SK 82 contains one chromosome 22 and W47 contains one chromosome 21 per diploid cell but the same amount of DNA (3 µg) from each gave only a very weak signal compared to that from PG 25.

2 WI contains one chromosome 13 per diploid set of chromosomes. Sequences on the centromere or short arm of this chromosome hybridised strongly to satellite I in the in situ hybridisation experiment (Fig. 8.1) yet this cell line apparently contains no sequences homologous to satellite I.

One possible explanation for the rather poor fit of the in situ and cell hybrid data for satellite I could be the almost exclusive autosomal location of this sequence on the short arms or centromeres of the acrocentric chromosomes. These regions are known to be highly variable both in the size of the C band, the length of the short arm and the presence or absence of intensely fluorescent tips or 'satellites' after quinacrine staining (Paris Conference, 1971). The chromosome 22 present in PG25 has a very intensely fluorescent 'satellite' on its short arm (pers. observation) and such a region is absent from the chromosome 13 present in cell line 2WI (R. Buckland, pers. communication). Thus the presence of satellite I sequences may be correlated with the presence of certain cytogenetic features such as the intensely fluorescing 'satellite' regions. Similarly the chromosome preparation used in the in situ hybridisation had two different chromosome 21s, one having a large short arm and a bright satellite region, the other having a short arm with no satellite region visible. In nine spreads out of the total 17 that were counted, it was estimated that the large chromosome 21 hybridised about five times as many grains as its smaller homologue, suggesting that such polymorphisms may be associated with changes in the amount of this repeated DNA (see also Gosden et al., 1981b).

* Not to be confused with 'satellite DNA'.
Figure 8.2
Hybridisation of satellite I to DNA from somatic cell hybrid lines.

DNA from the various somatic cell hybrid lines (see Table 8.1), total human DNA and purified satellite I were restricted with HaeIII or HaeIII and HinfI together, the products were separated by electrophoresis on a 2% agarose gel and then transferred to nitrocellulose by the method of Southern. The filters were hybridised to $^{32}$P labelled satellite I and finally washed to 2 X SSC at 65°C.

i  HaeIII digests. $a$, lambda DNA/EcoRI and HindIII; $b$-$i$, somatic cell hybrid DNA - 3 µg per track; $b$, W47; $c$, Mirl; $d$, 2WI; $e$, Horl - 9-15; $f$, P25; $g$, X-9; $h$, SK82; $i$, Horl - 9-X; $j$, 1.5 µg female DNA; $k$, 1.5 µg male DNA; $l$, 0.1 µg purified satellite I; $m$, 3 µg mouse DNA; $n$, 3 µg CHO DNA.

ii HaeIII and HinfI double digests. The loading is the same as that indicated in i.

Aliquots from the same batch of HaeIII and HaeIII-HinfI digested DNA used above were also probed with satellite IV (see Fig. 8.8).
Alternatively the disparity between *in situ* and cell hybrid data may be that satellite I sequences are very unstable in cell lines and have been lost from the chromosomes during their culture in the host cell line. Thus the disparity between *in situ* and cell hybrid data may not be a reflection of the polymorphic nature of satellite I in the population but a reflection of this sequence's instability when cultured in heterologous cell lines.

The data presented here for satellite I is preliminary and any conclusions drawn from it necessarily tentative. However, the apparent extreme variation in the amount of this sequence at a chromosomal location and the possibility that this may underly such cytological features as the acrocentric 'satellites' would certainly make this topic worth pursuing.

8.3 Satellite II

*In situ* hybridisation of this sequence has been carried out by Jones *et al.* (1971) who reported that the main site of hybridisation was the C band of chromosome 1 with lesser amounts hybridising to the C band of chromosome 16 and small amounts present on the C band of 9. Gosden *et al.* (1975), however, reported that the main site of hybridisation of this sequence was the C band of chromosome 9. Figure 8.3 shows the results of an *in situ* hybridisation using \(^{3}H\) labelled nick translated satellite II. In agreement with Jones *et al.* (1971) the major site of hybridisation of this sequence was found to be the C band of chromosome 1. Lesser, but still significant hybridisation was observed to the C band of chromosome 16 and to the Y chromosome.

A number of restriction enzymes cleave satellite II yielding a
Figure 8.3

In situ hybridisation of satellite II

$^{3}H$ satellite II, labelled by nick translation was hybridised to human metaphase spreads as outlined in Section 10.9. In this experiment the spreads were Q-banded after autoradiography and the chromosome banding and silver grains were observed simultaneously (see Section 10.9).

i Q-banded spread - observed with U.V. light source only.

ii Q-banded spread - observed with U.V. light source and visible light source.

iii Cumulative grain count from 16 spreads. The number of grains was counted per segment of chromosome (a chromosome segment was taken to equal $\frac{1}{2}$ the length of chromosome 22 on that spread).
characteristic pattern of fragments that can readily be identified in total genomic DNA by 'blot' hybridisation experiments (Section 3.5). Thus it should be possible to unambiguously identify satellite II sequences in the DNA extracted from the cell hybrid lines. Figure 8.4 shows Southern 'blots' of the cell line DNA after restriction with MboI and hybridisation to $^{32}\text{P}$ labelled nick translated satellite II. Only two of the cell lines, SK82 and W47, gave MboI fragments that hybridised to the probe; these two cell lines also gave the strongest signal with this probe. SK82 contains two human chromosomes per diploid cell, a 1/17 translocation that consists of most of chromosome 1 and the short arm of chromosome 17 and a chromosome 22 (Table 8.1). The cell line PG 25 also contains a chromosome 22, but does not exhibit any MboI restriction fragments and reacts only weakly to the probe when compared to SK82 and W47. It would seem likely, therefore, that the pattern of fragments seen in SK82 is derived from sequences present on the 1/17 translocation and, in the light of the in situ data, that these are located in the C band of chromosome 1. The pattern of MboI bands observed in SK82 could be clearly related to the MboI pattern of total genomic DNA and/or purified satellite II. For example, two bands at 330 bp and 360 bp are evident in all three DNA samples as are strong bands at 640 bp and 1050 bp (Figure 8.4). However, not all the fragments that are seen in digests of total DNA or purified satellite II are present in SK82 DNA. This is most apparent inbetween the 360 bp and 640 bp fragments and between the 640 bp and 1050 bp fragments. At least five intervening MboI fragments are seen in total DNA or purified satellite II digests that are absent in SK82 and not detectable even after prolonged exposure (Figure 8.4.ii). However, the restriction pattern of W47 DNA,
Figure 8.4
Hybridisation of satellite II to DNA from somatic cell hybrid lines.

DNA samples were restricted by MboI, the products separated on a 2% agarose gel and then transferred to a nitrocellulose filter by the method of Southern. The filter was hybridised to $^{32}\text{P}$ labelled satellite II. The final wash was in 2 X SSC at 65°C before exposure to X-ray film.

i 24 hour exposure. a, lambda DNA/EcoRI and HindIII; b-i, DNA from somatic cell hybrid lines - 3 μg per track; b, W47; c, Mirl; d, 2WI; e, HorI 9-15; f, PG25; g, X-9; h, SK32; i, HorI - 9-X; j, 3 μg mouse DNA; k, 3 μg CHO DNA; l, 0.1 μg purified satellite II; m, 1.5 μg total female DNA.

ii 4 day exposure. a-m as above.

The closed triangles (▲) indicate fragments present in total DNA (and purified satellite II) thought to correspond to fragments present in cell line SK32. The open triangles (◇) indicate fragments present in total DNA (and purified satellite II) thought to correspond to fragments present in cell line W47. The fragments marked with ▲ and those marked with ◇ were purified independently and their fine structure compared using the restriction enzymes HinfI and TaqI (see Fig. 8.6).
whilst lacking the 330 bp, 360 bp, 640 bp and 1050 bp fragments, does possess at least some of these 'intervening' fragments. To some extent the restriction patterns of the two cell lines are complementary in that combining them would generate a pattern more similar to that of the purified satellite than is either one alone. The fit is not perfect. In W47 a large fraction of the sequences that hybridise satellite II appear not to be cut by MboI, even though in MboI digests of total human genomic DNA and the purified satellite most of the sequence is cut. The large amount of satellite II sequences present at the origin in W47 after MboI restriction was, to some extent, due to partial digestion and subsequent experiments showed a more complete digestion of this DNA. W47 contains a number of human chromosomes in each mouse cell. It is the only cell line of the eight employed in this study to contain a human chromosome 16, the only autosome other than chromosome 1 that hybridised significant amounts of satellite II in the in situ hybridisation experiment (Figure 8.3). It would seem reasonable to postulate that the MboI patterns seen by restriction of SK82 and W47 are the result of a distinct chromosomal location of satellite II on chromosomes 1 and 16 and that the sequence differs in its arrangement of MboI sites at these two locations.

The arrangement of satellite II sequences in the two cell lines was further investigated by restriction with EcoRI and AluI (Fig. 8.5). These enzymes cleave only a portion of total satellite II and both yield virtually the same pattern of fragments indicating that the restriction sites for these two enzymes may be on the same subset of satellite II sequences (Section 3.5). Investigation of the cell hybrids by restriction and Southern transfer showed that satellite II sequences
DNA samples were electrophoresed on a 2% gel and were subsequently transferred to a nitrocellulose filter by the method of Southern. This filter was hybridised to $^{32}$P labelled satellite II in 3 X SSC at 58°C and the final wash was at this level of stringency.

- a, lambda DNA/EcoRI and HindIII
- b, 1.5 µg female DNA/MboI
- c, 3 µg SK82 DNA/MboI
- d, 3 µg W47 DNA/MboI
- e, 3 µg SK82 DNA/AluI
- f, 1.5 µg female DNA/AluI
- g, 3 µg W47 DNA/AluI
- h, 1.5 µg female DNA/EcoRI
- i, 3 µg SK82 DNA/EcoRI
- j, 3 µg W47 DNA/EcoRI
- k, 1.5 µg male DNA/KpnI
- l, 1.5 µg male DNA/HaeIII
- m, 0.05 µg satellite IV/HaeIII
- n, 0.05 µg satellite II/HaeIII

The filter shown in i was washed in 0.12M PB at 76°C and then re-exposed to X-ray film until the lambda DNA markers were about the same intensity as in i.

The open and closed triangles indicate the two preparations of MboI fragments whose pattern of HinfI and TaqI restriction is investigated in Figure 8.6 (see also Figure 8.4).
that had AluI restriction sites were present only in SK82, i.e. presumably on chromosome 1. The sequences complementary to the satellite II probe present in W47 were not restricted by AluI (complete restriction of W47 DNA was ensured by (a) inspection of the gel after electrophoresis which showed very little DNA remaining at the origin, and (b) the inclusion of φXRF DNA in the digestion mixture of each sample. As expected, digestion of SK82 with EcoRI yielded a pattern of satellite II fragments similar to that seen with AluI, whereas W47 DNA yielded no detectable fragments. However, the pattern of EcoRI fragments yielded by SK82 was very faint compared to that obtained with AluI. This was due to partial digestion of SK82 by EcoRI (inspection of the gel prior to 'blotting' showed that most of the EcoRI restricted SK82 DNA was still at the origin) even though total human DNA appeared to be completely digested. Repeated attempts to fully digest SK82 DNA with EcoRI failed. This problem was encountered with a number of DNA preparations and was attributed to this restriction enzyme's notorious sensitivity (SK82 DNA was purified on a CsCl gradient; residual traces of Cs⁺ could inhibit the activity of the enzyme).

Despite the problems encountered with EcoRI, it is evident that restriction with AluI defines a subset of satellite II sequences that are most probably located on chromosome 1 and a subset of sequences lacking these enzyme sites very likely present on chromosome 16. Are these two subsets thus distinct sequences that copurify as satellite II or do the differing chromosomal locations contain the same basic sequence arranged with different higher order periodicities? This question can be answered in the case of satellite II because this sequence consists of tandem arrays of 10, 12 and 15 base pair repeats which can be visualised by HinfI or Taq 1 digestion (Chapter 3). Accordingly MboI fragments specific to SK82 and MboI fragments specific
Purified satellite II was restricted by MboI and the products separated on a 1.5% agarose gel. Two pools of MboI fragments were extracted from the gel. Pool 1 was comprised of 4 fragments thought to be present in SK 82 DNA (marked by closed triangles in Figures 8.4 and 8.5) and pool 2 was comprised of 5 fragments thought to be present in W47 DNA (marked by open triangles in Figures 8.4 and 8.5). The two pools of MboI fragments were nick translated with dTTP, restricted with HinfI and TaqI and the products electrophoresed on a 20% polyacrylamide, 7M urea gel.

a, HinfI digested satellite II end-labelled with $^{32}$P by polynucleotide kinase; b, pool 1 MboI fragments/HinfI; c, pool 1 fragments/TaqI; d, pool 2 fragments/HinfI; e, pool 2 fragments/Taq I.

The HinfI and TaqI patterns of the nick translated MboI fragments are rather different although they are much the same when unrestricted satellite II is nick translated and digested by them (see Fig. 3.2). This difference presumably reflects the positions of the HinfI and TaqI sites relative to the MboI sites at which the DNA polymerase I will, to some extent, end-label by 'filling-in' the bases complementary to the 5' single stranded tail.
to W47 were prepared from purified satellite II. Four MboI fragments corresponding to those present in SK82 (↑) and five fragments corresponding to those present in W47 (↓) were excised from a preparative gel (Fig. 8.4). The two sets of bands were separately pooled and extracted from the agarose gel by electrophoresis into a dialysis bag (Section 10.4). The two preparations of DNA were nick translated, restricted with HinfI and TaqI and the products analysed on 20% acrylamide, 7 M urea gels (Section 10.4). Both preparations were very similar in their HinfI and TaqI patterns of restriction (Fig. 8.6) indicating that the fine structure of the sequence is the same in the two preparations of MboI fragments. Therefore it seems that the fine structure of the sequences that comprise satellite II is the same at the two chromosomal locations. However, the higher order arrangement of the short repeats that comprise this sequence are specific to one chromosomal location.

8.4 Satellite IV

The location of satellite IV, as determined by in situ hybridisation has been determined by Gosden et al. (1975). These workers reported that the major site of hybridisation of this satellite is to the C band of chromosome 9 with lesser amounts hybridising to the long arm of the Y chromosome and to the short arm of the acrocentric chromosomes, in particular to chromosome 15. Fig. 8.7 shows the results of an in situ hybridisation using satellite IV as a probe. In agreement with Gosden et al. (1975) the major site of hybridisation of this satellite was determined to be the C band of chromosome 9 with lesser amounts hybridising to the long arm of the Y chromosome and to the short arms of the acrocentric chromosomes. The distribution and arrangement of satellite IV sequences were investigated by restriction analysis and
In situ hybridisation of human satellite IV

$^{3}$H satellite IV, labelled by nick translation, was hybridised to human metaphase chromosomes as outlined in Section 10.9. In this experiment the spreads were Q-banded after autoradiography and the chromosome banding and silver grains were observed simultaneously (Section 10.9).

i Q-banded spread - observed with U.V. light source only.

ii Q-banded spread - observed with U.V. light source and visible light source.

iii Cumulative grain count from 11 spreads. The number of grains was counted per segment of chromosome (a chromosome segment was taken to equal $\frac{1}{2}$ the length of chromosome 22 on that spread).
Southern 'blots'. This approach suffers from a similar problem to that encountered with satellite I (Section 8.2) in that no enzyme is available for this satellite DNA which gives a distinctive restriction pattern suitable for Southern 'blotting'. This sequence can however be distinguished by its resistance to digestion with a number of enzymes e.g. HaeIII, MboI and by its susceptibility to others e.g. HinfI and TaqI. Fig. 8.8 shows two Southern 'blots' in which the target DNAs have been digested either with HaeIII alone or double digested with HaeIII and HinfI. The DNA samples used in these two transfers were aliquots from the same digestions used in the investigation of satellite I (Fig. 8.2). After HaeIII digestion the probe reacted with sequences that were uncut by the enzyme and in each case the sequences complementary to satellite IV remained at the origin of the gel (an exception to this was total male DNA in which hybridisation was observed between the satellite IV probe and the 3.4 kb specific male band, see Section 7.4). In contrast to the Southern blots hybridised to satellite I, digestion with HinfI reduced the high MW DNAs at the origin of the gel to a smear of low MW DNAs that did not efficiently or reproducibly bind to nitrocellulose.

The cell hybrid DNA that hybridised most strongly to the probe was that prepared from the X-9 cell line. This hybrid cell line contains the long arm, C band and part of the short arm of chromosome 9 translocated to the short arm of an X chromosome (Table 8.1). Since Horl-9-X, which contains one X chromosome per diploid cell, gave no signal with the satellite IV probe it must be concluded that it is the chromosome 9 of the translocation in which the large amount of sequences complementary to satellite IV are located. Thus the results from the cell hybrid lines are in good agreement with those derived from the in situ hybridisation experiment.
DNA from the various somatic cell hybrid lines (see Table 8.1) mouse and CHO control DNA, total human DNA and purified satellite IV were restricted with HaeIII or HaeIII and HinfI together, the products were separated by electrophoresis on a 2% agarose gel and then transferred to nitrocellulose by the method of Southern. The filters were hybridised to $^{32}$P labelled satellite IV and finally washed in 2 X SSC at 65°C.

i  HaeIII digests. a, lambda DNA/EcoRI and HindIII;  
   b-i, somatic cell hybrid DNA - 3 μg per track;  
   b, W47; c, M1rl; d, 2WI; e, Horal-9-15; f, PG25;  
   g, X-9; h, SK82; i, Horal-9-X; j, 1.5 μg female DNA;  
   k, 1.5 μg male DNA; l, 0.1 μg purified satellite IV;  
   m, 3 μg mouse DNA; n, 3 μg CHO DNA.

ii  HaeIII and HinfI double digests. The loading is the same as that indicated in i.

Aliquots from the same batch of HaeIII and HinfI digested DNA used above were also probed with satellite I (see Fig. 8.2). The 3.4 Kb Y specific fragment is indicated on left of Figure 8.8.i.
Other cell lines, W47, PG25, SK82 and Harl-X-15 gave weaker but quite significant signals with the probe indicating that sequences complementary to satellite IV are also present on other chromosomes. Digestion of the cell hybrid DNAs with other restriction enzymes, e.g. MboI and KpnI indicated that the chromosome 9 sequences complementary to satellite IV lacked any high order periodicity in the form of a distinctive repeating pattern, whereas the sequences on the other chromosomes were restricted to some type of pattern. Thus the sequences detected on the C band of chromosome 9 are similar to those isolated as purified satellite IV in that they lack any demonstrable arrangement into higher order repeating units. Given the very large amount of satellite DNA detected on the C band of chromosome 9 (Jones et al., 1974; Gosden et al., 1975; Marx et al., 1976; Fig. 8.7) and the large signal given by X-9 cell line DNA when hybridised to satellite IV, it does not seem unreasonable to suppose that the purified satellite (Section 10.2) which is about 2% of the total genome, is predominantly derived from this one chromosome. Certainly the other major sites where sequences complementary to it are detected by in situ hybridisation, the Y chromosome (Section 7.4) and chromosome 15 (Section 8.5) probably do not contain satellite IV proper in that the complementary sequences seem to be arranged into higher order repeating units, revealed by restriction enzyme analysis, which are absent in the purified satellite DNA. The sequences at these other sites are more correctly described as related cross reacting sequences; for example, in the case of the Y chromosome, it is felt likely that the hybridisation of satellite IV to it is because of the cross hybridisation of satellite IV and the 3.4 kb male specific repeat (Section 7.4) rather than the presence of satellite IV proper on this chromosome.
In conclusion, *in situ* hybridisation of human satellite IV to human chromosomes shows that there are a number of different chromosomal locations which hybridise this probe. The major site of hybridisation is the C band of chromosome 9 and it is postulated that the sequences physically purified as satellite IV in caesium gradients are predominantly derived from this one location and that hybridisation to other locations may be due to cross reaction of this sequence to sequences sharing a similar fine structure such as the 3.4 kb Y specific fragment. If this is the case, then the arrangement of 'satellite IV-like' sequences at different chromosomal locations in the genome is seen to be a similar case to that described for satellite II in which an underlying fine structure is held in common but each location has a highly specific higher order arrangement of the sequence.

8.5 The 3.4 kb 'Y Specific' Fragment

The experiments described in Chapter 7 indicate that some of the sequences interspersed along this 3.4 kb repeated fragment cross react with other sequences, in particular satellite IV, to which it has a similar fine structure. The extent of hybridisation depends on the stringency of the reaction conditions. At low stringency a number of different sequences cross hybridise with it but at high stringency the sequence hybridises only to itself (Fig. 7.2).

*In situ* hybridisation was carried out with the purified fragment at 65°C in 2 x SSC. The localisation of the \(^3\)H probe was very specific. Nearly all the grains were found over the Y chromosome; although trace amounts were detected on chromosomes 1, 9, 13, 14 and 15 (data not shown). Considering the hybridisation of the 3.4 kb fragment and satellite IV (Chapter 7) it might be expected that the former would also hybridise to autosomal locations. Bostock *et al.* (1978) did find that at longer
Figure 8.9

Hybridisation of the 3.4 kb Y fragment at low stringency to somatic cell hybrid DNA

Somatic cell hybrid DNA (3 μg per track), total human DNA (1.5 μg per track), mouse DNA (3 μg per track), and CHO DNA (3 μg per track) were restricted with KpnI or MboI, the products separated on a 2% agarose gel by electrophoresis and transferred to nitrocellulose by the method of Southern. The 3.4 kb Y fragment, labelled with $^{32}\text{P}$ by nick translation was hybridised to the filters in 3 X SSC at 58°C; the final washes were at this level of stringency.

i  KpnI digests. a, lambda DNA/EcoRI and HindIII; b, W47; c, Mirl; d, 2WI; e, Horl-9-15; f, PG 25; g, SK82; h, Horl-9-X; i, total female DNA; j, total male DNA; k, mouse DNA; l, CHO DNA; m, X-9 DNA. The KpnI fragments present in Horl-9-15 are indicated by the closed triangles.

ii  MboI digests. The order of loading and the amount of DNA loaded is the same as for i. The MboI fragments present in Horl-9-15 are indicated by the open triangles. At this low level of stringency hybridisation is also detected between the 3.4 kb Y fragment and the MboI fragments of satellite II present in total DNA (see Fig. 7.2) and in SK82 and W47 DNA (see Fig. 8.4).
exposures the 3.4 kb Y specific fragment did hybridise significantly to the C band of chromosome 9 and to the short arm of chromosome 15. Presumably in the in situ hybridisation described in the present work the slides were not exposed for long enough for grains to accumulate over these autosomal sites containing related sequences.

Because of the low levels of hybridisation of the sequence to the autosomes, encountered under the standard conditions of 2 x SSC/65°C, sequences related to the 3.4 kb fragment were studied in cell hybrid DNA under more relaxed conditions of hybridisation of 58°C/3 x SSC. Fig. 8.9.i is a Southern 'blot' of total human DNA and cell hybrid DNA restricted with the enzyme KpnI (Section 7). A number of cell hybrid DNAs gave a strong signal at the origin. However, the complementary sequences present in the Horl-9-15 were mostly restricted to two fragments sited at 1.8 kb and 3.4 kb, which were also prominent in KpnI restricted total male and female DNA.

The 3.4 kb KpnI fragment was investigated further by eluting it from a preparative agarose gel and restricting it with a number of other enzymes. MboI was found to cut this sequence to yield four major fragments sized 700 bp, 1100 bp, 1800 bp and 2800 bp. A tentative restriction map of this fragment is shown in Fig. 8.10 and indicates that within the KpnI 3.4 kb repeat there is a MboI 1.8 kb repeat. It would therefore seem quite likely that this repeated sequence is comprised of tandem arrays of 1.8 kb, cut either by MboI or KpnI. Loss of a restriction site would result in fragments of 3.6 kb in size (the difference between the measured 3.4 kb at the estimated 3.6 kb is within the limits of experimental error in 1.5-2.0% agarose gels). The arrangement of MboI sites in sequences that cross hybridised to the male band was also investigated in the cell hybrid DNA's (Fig. 8.9.ii). MboI digestion of Horl-9-15 generated fragments sized at 1.8 kb and
The 3.4 kb KpnI fragment was preparatively isolated from total female DNA (Section 10.3). The position of MboI sites on the 3.4 kb KpnI fragment were tentatively assigned by digesting it with MboI, electrophoresing the products on a 1.5% gel, transferring to nitrocellulose and hybridisation to $^{32}$P labelled Y fragment DNA.
3.4 kb as well as two other prominent fragments sized to be approximately 5.4 kb and 7.0 kb, which indicated a tandem array of the MboI 1.8 kb fragments (faint traces of such higher multimers were also visible in KpnI digests of total DNA and Horl-9-15 DNA on closer inspection).

No hybridisation of the 3.4 kb probe was observed to Horl 9X DNA (X chromosome) and therefore it was concluded that the KpnI/MboI repeats were located on the chromosome 15 of Horl-9-15. None of the other cell hybrids contained such repeating fragments and it would seem possible that they are specific to chromosome 15, although it must be borne in mind that not all 23 human chromosomes are covered by the eight cell lines used in this study. Assuming the KpnI/MboI repeats to be specific to chromosome 15 this sequence is analogous to the 3.4 kb repeated sequence that is specific to the Y chromosome. They are also very similar in terms of their higher order structure; the Y specific sequence is comprised of a 1.7 kb and 1.8 kb repeat, whereas the KpnI/MboI sequence consists of a tandem array of 1.8 kb fragments. Furthermore, the two sequences hybridise (at ordinary levels of stringency as well as under relaxed conditions, not shown) and presumably both have a similar underlying fine structure, although this has not been determined for the KpnI/MboI fragments.

Satellite IV also hybridises to the KpnI fragments (Fig. 8.11.i) to form stable hybrids that are not melted off under stringent conditions of washing (Fig. 8.11.ii). Although this satellite is not cut by either KpnI or MboI a trace of these fragments can be detected copurifying with it. This was considered to be caused by cross contamination from satellite III in which a large amount of these sequences are found. Thus it could be the presence of contaminant KpnI fragments that cause the hybridisation seen in Figure 8.11. This was thought to be unlikely for a number of reasons, (a) the KpnI fragments form a negligible fraction
Figure 8.11

Hybridisation of satellite IV at low stringency

i DNA samples were electrophoresed on a 2% gel and were subsequently transferred to a nitrocellulose filter by the method of Southern. This filter was hybridised to $^{32}$P labelled satellite IV in 3 X SSC at 58°C and the final wash was at this level of stringency.

a, lambda DNA/EcoRI and HindIII; b, 1.5 µg female DNA/KpnI; c, 1.5 µg female DNA/MboI; d, 1.5 µg female DNA/MboI and KpnI double digest; e, 0.05 µg satellite IV/KpnI; f, 0.05 µg satellite III/KpnI.

ii The filter shown in i was washed in 0.12M PB at 76°C and then re-exposed to X-ray film until the lambda DNA markers were about the same intensity as in i.

The 1.8 kb and 3.4 kb KpnI and MboI fragments are indicated by open triangles.
of satellite IV and are not detectable by ethidium bromide staining; (b) KpnI fragments are HincII sensitive and are digested to a smear of low MW DNA as expected of sequences related to satellite IV; (c) the inter-relationship of satellite IV and the 3.4 kb Y specific DNA has been studied in some detail and the basis for their cross hybridisation (i.e. an underlying 5 bp periodicity) described. KpnI fragments do not purify with the 3.4 kb Y specific fragment yet do hybridise to it and, therefore, might be expected to hybridise related sequences such as satellite IV.

The overall conclusion is that these three distinct sequences, the 3.4 kb Y DNA, the KpnI/MboI repeats and satellite IV share a similar fine structure. The 3.4 kb repeat is undoubtedly specific to the Y chromosome, it seems likely that the KpnI fragments are specific to chromosome 15 and that satellite IV may be predominantly located on chromosome 9. Thus these similar simple sequences are arranged in a different manner at different chromosomal locations. This is reflected by the in situ hybridisation results in which the secondary sites of hybridisation of satellite IV are the Y chromosome and chromosome 15, whereas the secondary sites of hybridisation for the 3.4 kb repeat (at long exposures) are chromosomes 9 and 15 (Bostock et al., 1978).

The relationship between satellite IV, the KpnI/MboI fragments and Y DNA is similar to that described for satellite II (Section 8.3) in which sequences of similar fine structure, are arranged differently on different chromosomes. Satellite II also seems to be distantly related to these three sequences. At normal stringency (2 x SSC, 65°C) it hybridises only weakly to the 3.4 kb Y DNA (Figure 5.2), satellite IV (Figure 5.4) and the KpnI fragments. However, under relaxed conditions of hybridisation it reacts strongly with all three of these sequences. Figure 8.5.i shows the hybridisation of satellite II to
these three sequences under relaxed conditions of hybridisation (3 x SSC, 58°C). Washing the filter under more stringent conditions (0.12M PB 76°C) shows that the duplexes formed between the satellite II probe and these three sequences are not as stable as the duplexes formed with the homologous satellite II sequences present in the target DNA.

In conclusion, it would appear that the four sequences, satellite II, satellite IV, the KpnI/MboI sequence and the 3.4 kb Y DNA are all members of the same family in which there are sequences of widely varying degrees of relatedness. A particular member of this family may have a specific chromosomal location and this is reflected in a specific higher order arrangement of repeating units. The relationship between the 3.4 kb Y DNA and satellite IV has been described in Chapter 7. The relationship between these sequences and the more distantly related satellite II, and the evolution of the family as a whole, are discussed in Section 9.3.

8.6 The 170 bp Repeated Sequence

In Chapter 6 the structure and evolution of this repeated sequence family was investigated by restriction with a variety of enzymes. DdeI digestion showed that the fundamental repeat unit was about 170 bp, whereas HaeIII digestion indicated that the sequence had amplified as a unit of 340 bp containing two of these repeats. It was postulated that a combination of the loss of restriction sites by mutation and unequal crossing over had generated new and larger HaeIII repeats from the basic 340 bp array and that segmental amplification (i.e. rapid independent amplification of only a part of the sequence; Horst and Zachau, 1977) resulted in some of these repeats comprising a considerable proportion of the total sequence family. Similarly, segmental amplification was deemed responsible for the pattern of restriction
obtained with other enzymes, in particular TaqI. In this case the major periodicities observed upon digestion were the 8-mer, 11-mer and 16-mer and the pattern thought to be due to the appearance of this restriction site by mutation subsequent to the initial amplification of the 340 bp array, i.e. Type B restriction pattern (Horst and Zachau, 1977). In the work described below the arrangement of the HaeIII and TaqI sites of the 170 bp repeat sequence is described on the different chromosomes present in the cell hybrids and related to the overall structure and evolution of this repeated sequence family.

Figure 8.12.i shows a Southern 'blot' of cell hybrid DNA digested with HaeIII and Fig. 8.12.ii a similar 'blot' of DNA digested with TaqI. Both filters were hybridised to a $^{32}$P probe made by nick translating the 640 bp fragment (4-mer) purified from a HaeIII digest of satellite III (Section 10.4). Unfortunately, use of this probe rather complicated matters in that it cross reacted with sequences at the origin of the HaeIII gel (Fig. 8.12). In view of the fact that the HaeIII 340 bp fragment (2-mer) does not hybridise to sequences at the origin of a gel of either HaeIII restricted satellite III (Fig. 5.1) or HaeIII restricted total human DNA (Fig. 6.1), the hybridisation observed with the 4-mer was felt to be due to impurities in the probe. These impurities were presumably due to sequences smearing down from the origin of the preparative gel and being co-extracted with the 4-mer. That these sequences are not cut by DdeI (the amount of signal at the origin of DdeI restricted DNA is about the same as that observed with HaeIII restricted DNA (Figure 8.12.iii) also suggests that they are unrelated to the 170 bp repeated sequence family. They are, however, completely digested by TaqI, probably to a smear, suggesting they may be related to the satellite II/satellite IV family of sequences
Figure 8.11

Hybridisation of the 680 bp HaeIII fragment to somatic cell hybrid DNA and to total human DNA.

DNA samples were digested by the enzymes indicated, the products separated on a 2% agarose gel and transferred to nitrocellulose by the method of Southern. The filters were hybridised to the 680 bp fragment, isolated from a HaeIII digest of satellite III (i.e. the 4-mer), in 3 x SSC at 65°C. The final wash of the filters was in 2 X SSC at 65°C.

i HaeIII digests. a, lambda DNA/EcoRI and HindIII; b–i, somatic cell hybrid DNA - 3 μg per track; b, W47; c, Mirl; d, 2WI; e, Horl-9-15; f, PG25; g, Horl-9-X; h, SK82; i, X/9; j, 1.5 μg of female DNA.

ii TaqI digests. a–j, as for i. Track h was run independently of the other tracks on another gel. The apparent small fragment present in it (indicated by the closed triangle) was background smear and did not diminish, as did the signal from the DNA fragments, when the filter was washed at a high temperature.

iii Female DNA (1.5 μg per track) digested by a variety of enzymes. a, lambda DNA/EcoRI and HindIII; b, DdeI; c, HaeIII; d, XbaI; e, TaqI; f, HaeIII and XbaI double digest; g, HaeIII and TaqI double digest; h, XbaI and TaqI double digest; i, EcoRI; j, EcoRI and HaeIII double digest; k, EcoRI and XbaI double digest; l, EcoRI and TaqI double digest.

The numbers at the right of each autoradiograph indicate multiples of 170 bp.
(Section 9.2) which are known to comprise a substantial portion of satellite III.

Despite the problems encountered concerning the purity of the probe, a reasonably clear picture emerges concerning the arrangement of the 170 bp sequence. The HaeIII 2-mer, 3-mer and 4-mer, although at varying amounts, are observed in all the cell lines. This could be due to the ubiquity of this basic array of repeats but because the X chromosome which is present in five of the eight cell lines has (albeit very faintly) these repeats, it may well reflect the ubiquity of this chromosome in the cell hybrids. Horl-9-15 and Mrl1 have similar amounts of this sequence, in a similar arrangement to that observed on Horl-9-X so that there may be little or none of this sequence family present on the other chromosomes (15 and 7) in these cell lines. In contrast, the other two X chromosome containing cell lines, X-9 and W47 gave considerably stronger signals than Horl-9-X and also contained some differently sized fragments. It was concluded that chromosomes other than the X present in these cell lines also contained some of the 170 bp repeat sequences, e.g. chromosome 9 in the X-9 cell line.

In contrast to the HaeIII 2-mer, 3-mer and 4-mer some of the larger HaeIII repeating fragments are specific to just a few cell lines, e.g. the 5-mer is present only in SK82 and X-9, the 6-mer in W47, SK32 and PG25 and the 11-mer present only in W47, 2WI and PG25. Some of the cell lines contain the same chromosome and therefore it might be expected that they have repeated fragments in common. PG25 and SK82 both contain a chromosome 22 and both have a HaeIII 6-mer and 8-mer in about the same ratio. The similar HaeIII pattern is correlated with the fact that both cell lines have a prominent TaqI 8-mer that is not present in any of the other cell lines (Fig. 8.12.ii). Thus this
arrangement of repeated sequences may be specific to chromosome 22. If so it would seem likely that the HaeIII 6-mer and 8-mer repeats are arranged in TaqI 8-mer repeats on this chromosome.

It was felt unlikely that the TaqI 8-mer corresponded solely with the HaeIII 8-mer in these cell lines because in each case the proportion of the former was much greater than the latter. Certainly the observation that TaqI extensively digested the HaeIII 8-mer and HaeIII 6-mer in HaeIII/TaqI double digests of total DNA is in agreement with this view. Likewise a TaqI/HaeIII double digest of PG25, hybridised to the 340 bp HaeIII monomer, gave a fragment just smaller than the 11-mer (see below), a prominent fragment of about 830 bp in size and less prominent fragment about 500 bp in size. No 8-mer or 6-mer was visible in this double digest, indicating that both HaeIII 6-mers and 8-mers have TaqI sites within them and that the TaqI 8-mer has HaeIII sites within it. In the context of the inter-relationship between the HaeIII 6-mer and 8-mer it is of interest to note that it has been shown by two dimensional hybridisation that sequences within these two repeating fragments are very well matched to each other (as well matched as they are to themselves, Section 6.3) indicating a very close relationship of the sequences that comprise these fragments (Mitchell et al., 1979).

Fig. 8.12.i shows that W47 also exhibits a prominent HaeIII 6-mer. However, in contrast to SK82 and PG25 this was not correlated with the presence of a HaeIII 8-mer nor was a TaqI 8-mer prominent in this cell line. Thus the HaeIII 6-mer in W47 is different to that in SK82 and PG25. W47 does not have a chromosome in common with both SK82 and PG25 and so it may be that the 6-mer in this cell line is the result of a separate and different amplification event on a different chromosome to that which produced the same sized HaeIII repeat in SK82 and PG25. It would certainly be most illuminating to compare the fine
restriction map of the HaeIII 6-mer from the three cell lines in which it is present.

PG25 and SK82 have a chromosome 22 in common but they also contain chromosomes that are not in common. This is presumably reflected in PG25 by the presence of a prominent HaeIII ll-mer that is absent in SK82. PG25 also contains a chromosome 13 and a chromosome 9. The absence of any detectable HaeIII ll-mer in X-9 DNA would seem to indicate that this fragment is derived from the chromosome 13. This repeat is prominent in two other cell lines, 2WI and W47. 2WI also contains a chromosome 13 as well as an X chromosome and two chromosome 7s. Comparison with cell lines Mirl and Horl-9-X indicates that this repeat was very likely located on either a chromosome 13 or perhaps a chromosome 19 and that the similarity between these two cell lines could thus be due to the presence of a common chromosome 13. W47 does not contain a chromosome in common with both of the other two cell lines and so the ll-mer repeat must be present on at least two chromosomes. TaqI digestion produced a prominent ll-mer in all three cell lines in about equal amounts to that observed with HaeIII. HaeIII/TaqI double digestion of total DNA produced a fragment just slightly smaller than the Hae or TaqI ll-mer indicating that in these repeats the HaeIII and TaqI sites are very close to each other (Fig. 8.13.iii). Digestion of the isolated HaeIII ll-mer with TaqI resulted in a fragment slightly smaller than the original HaeIII ll-mer. Thus the HaeIII ll-mer and TaqI ll-mer appear to be the same repeat and it would seem reasonable to conclude that the segmental amplification that created the HaeIII ll-mer repeat (Section 6.3) simultaneously created the TaqI ll-mer.

HaeIII/TaqI double digestion of the three cell lines in each case gave rise to a fragment just perceptibly smaller than the ll-mer
fragment seen by digestion with one enzyme, showing that the same repeat is shared by all three. It is felt unlikely that such a repeat, containing HaeIII and TaqI sites in the same place, could have arisen twice de novo. Yet from the analysis of the cell hybrids it seems to be located on at least two chromosomes. It seems more likely that this 11-mer repeat was created by a combination of mutation and recombination (Section 6.3) at one chromosomal location, amplified there and subsequently translocated to other sites in the genome. In contrast the HaeIII 6-mer, previously discussed, seems to be comprised of at least two 6-mer sequences that have arisen independently at different locations in the genome.

Fig. 8.12.ii shows that TaqI digestion does not generate any fragments in X-9 DNA; all the sequences complementary to this probe were at the origin of the gel. Because of the problem encountered using the 640 bp probe the experiment was repeated using the nick-translated 340 bp (2-mer) fragment as a probe. This probe does not hybridise to sequences at the origin of gels of HaeIII restricted total DNA (Fig. 6.3) or HaeIII restricted satellite III (Fig. 5.1). When hybridised to X-9 DNA that was digested with TaqI virtually all the sequences remained at the origin of the gel. These sequences were, however, digested completely by HaeIII to give a pattern of fragments very similar to that observed using the 640 bp probe. Most of the 170 bp repeat sequence in the X-9 cell line is probably present on chromosome 9, and therefore on this chromosome the sequence is characterised by the lack of TaqI sites. Thus the presence or absence of TaqI sites on different chromosomes provides another example of the segmentation of the 170 bp repeat sequence. In the discussion above it was shown that much of the sequence that contains the TaqI restriction sites is the same as the HaeIII 8-mer, 11-mer and
some of the 6-mer repeats. Thus, the lack of these HaeIII repeating fragments might be expected to correlate with the lack of TaqI fragments, as indeed it does on chromosome 9.

In conclusion, the arrangement of the 170 bp repeat sequence on different chromosomes supports the idea that the independent amplification of segments of this sequence has taken place on different chromosomes. In considering the HaeIII pattern of restriction it should be noted that only when new combinations of repeats occur (generated from the basic 340 bp array by mutation and/or recombination) can these segmental amplifications be distinguished from the primary amplification that generated the 340 bp array. However, in the case of TaqI, because the recognition sequence for this enzyme appeared subsequent to the primary amplification of the sequence, and in only a portion of the sequence, any TaqI restriction fragments are segmental variants. The evolution of this interesting family is considered further in Chapter 9 in the context of related sequences that are present in the primates.
9.1 Introduction

In the previous chapters the structure and evolution of a variety of human highly repeated DNAs has been described. All the sequences discussed can be isolated as components of cryptic satellites in Ag⁺ Cs₂SO₄ gradients. Fine structural analysis shows that a number of these sequences may be related. If this is taken into account then all the sequences can be grouped into three distinct repeated DNA families. These are satellite I, the family of sequences cut by HinfI (and TaqI) to a smear of short fragments - the HinfI family, and the 170 bp repeat sequence. In the following two sections the HinfI family and the 170 bp repeat family are discussed in more detail. Satellite I has to some extent been ignored, mainly because of the difficulty in analysing it as a result of the lack of restriction sites within the sequence. It is considerably more AT rich than the other sequences described (consisting primarily of AT interspersed with C, G and longer runs of T) and appears to be completely distinct from them.

9.2 The HinfI Family

This family of human highly repeated DNA is characterised by sequences cut to a smear of short fragments by the enzymes HinfI (and TaqI). The members of the family that have been analysed in detail have an underlying 5 bp periodicity and are characterised by relatively simple pyrimidine fingerprints. They are satellite II (Chapter 3), satellite IV (Chapter 4) and the 3.4 kb Y fragment (Chapter 7). Also included in this family are a large proportion of the sequences that comprise satellite III (assumed to be very
similar, if not identical to satellites II and IV) and the KpnI 1.8 kb and 3.4 kb fragments.

Restriction of satellites II and IV and the 3.4 kb Y fragment with Hinfl released very small fragments because of an underlying short order periodicity. Visualisation of these fragments often involved digesting nick translated DNA which, because of the presence of 'network DNA' (Section 3.2), made it impossible to say whether the entire repeated DNA had this structure. This was noted particularly in the case of the 3.4 kb Y fragment, and after Hinfl restriction of this fragment a considerable smear of high MW DNA was seen, although TaqI did appear to cut it more completely (Figure 7.5). Thus it is possible that sequences completely unrelated to the Hinfl family are present on the 3.4 kb fragment and certainly its complex reassociation and melting profile might argue for this view. However, the simplicity of its pyrimidine fingerprints (Figure 7.4) and their similarity to satellite IV would indicate that the bulk of this repeat is comprised of simple sequences of the Hinfl family type and that base changes and rearrangements within such simple repeats are responsible for the complex kinetics of reassociation and melting behaviour (Section 7.5).

Members of the Hinfl family cross-hybridise with one another to some extent, although some members only form duplexes with others when the stringency of hybridisation is relaxed. Such cross hybridisation experiments using uncloned sequences must be interpreted with caution because of the possibility of cross contamination of sequences during their purification. For example satellite II and satellite III are distinct 'classical' satellite DNAs that hybridise with one another. However, as shown in Chapter 5, the basis of this cross hybridisation is the copurification of satellite II sequences within
Satellite III. In contrast, at 60°C in 2 x SSC satellite II did not react with satellite I or the 170 bp repeat sequence, and only very weakly if at all, with satellite IV and the 3.4 kb Y fragment. However, when the stringency of hybridisation was lowered, satellite II did hybridise with the 3.4 kb Y fragment and with satellite IV, indicating a distant relationship to these sequences.

Satellite IV does not hybridise significantly to satellite II in 2 x SSC at 60°C but does appear to hybridise to the 3.4 kb Y fragment under these conditions (Figure 5.5). This positive hybridisation could be due to the copurification of the 3.4 kb Y fragment with satellite IV, as was observed with one preparation of this DNA (Figure 5.3). However, preparations of satellite IV (e.g. see Figure 5.2) were isolated virtually free of the 3.4 kb Y fragment and still hybridised to it. In addition, duplexes formed between the 3.4 kb Y fragment and satellite IV melted monophasically with a large ΔT_M (Figure 7.9), a melting profile quite different from that observed with the homologous reassociation of the 3.4 kb Y fragment.

One way of minimising the problems of cross contamination of sequences is to further purify them by cutting out restriction fragments from a gel. The 3.4 kb Y fragment was purified in this manner and although it hybridised to a variety of sequences at low stringency it was quite specific for itself at higher levels of stringency (Figure 7.2) indicating a distant relationship with other repeated sequences in the genome.

Despite the problems encountered with cross contamination it appears that some of the distinct highly repeated DNA sequences do cross hybridise. Satellite IV hybridises with some of the sequences present on the 3.4 kb Y fragment at 60°C in 2 x SSC, but under these conditions satellite II does not hybridise to either of these two repeated DNAs.
The HinfI family is a diverse simple sequence family. Within each member of the family there is considerable sequence variation. When the various members of the family are compared this variation can be seen to be discontinuous. It is this discontinuity in variation that serves to identify each member as a distinct sequence and is correlated with their distinct chromosomal location. In other words, the generation of discontinuous sequence variation in this diverse simple sequence family may have occurred because of the independent evolution of the sequences at different sites in the genome. In the case of satellite II discontinuous variation is apparent in terms of the presence of AluI (or EcoRI) restriction sites. Sequences of satellite II that are cut by AluI seem to be specific to chromosome 1 whereas there are satellite II sequences that lack AluI restriction sites present at another chromosomal location, possibly on chromosome 16 (Section 8.3).

Members of the HinfI family have been described by other workers. The sequence described by Cooke and Hindley (1979), and its similarity to satellite II has been previously mentioned (Chapter 3). Recently Deininger et al. (1981) have presented the sequence of a number of cloned human sequences prepared from renatured 300 nucleotide S1 resistant repeats. Three of the sequences isolated in this manner had a simple underlying 5 bp periodicity with 5'TTCCA' as the predominant repeating unit although they were quite distinct from each other. Each contained HinfI and two contained overlapping HinfI and TaqI sites. Two of the sequences hybridised to the 3.4 kb Y fragment, whereas the third one showed no sex specific differences. Unfortunately, with the present data it is not possible to identify these sequences with the ones described in this thesis. Their description by Deininger and co workers does, to a large extent, confirm the
notion of a prevalent simple sequence family with an underlying 5 bp periodicity.

The Hinfl family accounts for about 80% of the four satellite DNAs described in this work and therefore for a minimum of about 5% of total genomic DNA. It is felt likely that other members of this sequence family may also exist in the human genome, e.g. at low stringency the 3.4 kb fragment is driven by sequences present in female DNA that reassociate as intermediately repeated sequences and which may represent less reiterated members of this sequence family than the ones described in this work.

9.3 The 170 bp Repeat Sequence

This sequence is comprised of a fundamental repeat of about 170 bp that has been amplified as a dimer (Chapter 6). Subsequent evolution of the sequence has resulted in the generation of segmental variants that are present on different chromosomes (Section 8.6).

A 340 bp fragment from EcoRI digests of total human DNA has been sequenced by Wu and Manuelidis (1980). The sequence of this fragment shows it to be very similar to the 170 bp repeat sequence isolated as a component of satellite III. The 340 bp EcoRI fragment is comprised of two unequal halves - a 169 bp sequence and a 171 bp sequence which share 72% homology with each other. Both halves of the fragment have a DdeI site in an equivalent position. One of the halves has a HaeIII site that maps 51 bp from the DdeI site and a common variant has a Hinfl site that maps 87 bp from a DdeI site. Fragments of this size were seen by digesting the isolated DdeI monomer with HaeIII and Hinfl (Figure 6.4). The HaeIII 340 bp fragment isolated from satellite III hybridises to a multimeric series of fragments based on a repeat length of 170 bp in EcoRI digests of
total human DNA (Figure 6.2). Similarly the 340 bp EcoRI fragment isolated from total DNA hybridises to a characteristic HaeIII 'ladder' in digests of total DNA (Maio et al., 1981a).

A variety of primate DNAs have been investigated with restriction enzymes and have been shown to contain tandemly repeated sequences that are based on a 170 bp periodicity. About 20% of the genome of the African Green Monkey (Cercopithecus aethiops) is comprised of a satellite DNA called component α (Maio, 1971). Restriction of this DNA with HindIII or EcoRI shows the sequence to be tandem arrays of a 172 bp repeat. The West African baboon (Papio papio) has a 340 bp tandemly repeated DNA that is cut once by BamHI whereas the colobus monkey (Colobus badius) is characterised by a tandemly repeated DNA with a prominent monomeric length of 680 bp (Donehower and Gillespie, 1979). Both the 340 bp repeats of the baboon and the 680 bp repeats of the colobus monkey are characterised by an internal 170 bp periodicity.

Some of the primate highly repeated DNAs have been sequenced. The 340 bp fragment described for the West African baboon is in fact 343 bp in length, consisting of two unequal halves of 171 bp and 172 bp. The closely related bonnet monkey (Macaca radiata) has about 6% of its genome as a 343 bp repeat again with two unequal halves of 171 bp and 172 bp (Rubin et al., 1980a). The sequence of the African green monkey 172 bp repeat has also been published (Rosenberg et al., 1978). Comparison of the sequences of these primate highly repeated DNAs, including the human EcoRI fragment, shows that they are all related and therefore presumably derived from an ancestral sequence of 170 bp or thereabouts (Donehower et al., 1980). In the case of the human EcoRI fragment the divergence between the 169 bp and 171 bp halves of the repeat (27%) is almost
as great as the divergence between each half and the African green monkey repeat (35%) (see Section 9.6). In the case of the West African baboon the 172 bp half shares more homology with the African green monkey repeat than it does to the 171 bp half (Donehower, et al., 1980). Comparison of all the '170 bp' sequences indicate that some regions of the repeat have undergone more base changes than others. However, it is likely that the more homogeneous regions of the repeats are more diverged than the sequence data would indicate. For example, the DdeI restriction site is 'conserved' (see also Section 9.6) in the four primate repeats. From the analysis of the restriction of satellite III a level of base change of 3-4% was estimated (Section 6.3) at this site. The sequence data for the primate repeats has been obtained primarily from uncloned restriction fragments and presumably such a level of base divergence would not be detectable with the usual sequencing techniques. Nevertheless, the comparison of the four species indicates that at one level sequence divergence has been non-random. This pattern of divergence has been considered by Donehower et al. (1980) to be consistent with a model in which the more variable regions represent exchanges of stretches of DNA in an ancestral sequence with similar but non-identical sequences. Thus the divergence observed between the various primate species may be analogous to that which occurs within the 170 bp repeat sequence in humans (Section 6.3) where it was postulated that new recombinant sequences had been generated by unequal crossovers thus accounting for the interfragment heterogeneity of the HaeIII series and the non-random distribution of the HaeIII restriction sites. Similarly the amplification of the segmental variants of the 170 bp repeat sequence which seem to have occurred at distinct chromosomal locations (Section 8.6) seems analogous to the
amplification of multimeric repeats in some of the primate genomes, e.g. the predominant 6-mer in the colobus monkey. Whatever the mode of evolution of these sequences they appear to extend far back into evolutionary history and they are at least as old as the primate order. Thus the presence of 170 bp repeats homologous to the African green monkey α component has been demonstrated in the DNA of both the lemur and galago (prosimians) by Maio et al. (1981a) as well as in six other primate orders. The α component probe hybridised to sequences present in human DNA but at the level of stringency employed (1 x SSC, 65°C) it did not hybridise to the EcoRI ladder of fragments (Maio et al., 1981a). So, even though these sequences are related at this level of stringency, they could not form stable duplexes (from a comparison of the two sequences the longest stretch of uninterrupted homology is only 14 bp) even though there are other sequences present in the human genome that can. The results obtained with the α component of the African green monkey genome thus indicate that there is more than one family of 170 bp repeats present in the human genome.

In conclusion, the 170 bp repeat sequence that comprises a portion of human satellite III is only a fraction of the total family of 170 bp repeat sequences present in the human genome. The sequence is related to other tandemly arranged 170 bp repeats that are present in other primate genomes. The processes that have been operating to generate sequence variation between the various primate species (unequal crossovers, amplification of higher multimeric repeats) seem similar to those that have been operating to generate variation within this one human repeated DNA family and in many ways the variation that is observed between different species seem analogous to the interchromosomal variation seen within the human genome.
9.4 Other Human Highly Repeated DNA

The Hinfl family and the 170 bp repeat sequence conform, in general, to classic notions of satellite DNA. They are tandemly arranged, thus allowing their isolation from main band DNA as discrete buoyant density components and they seem to be predominantly located within heterochromatin (Chapter 8, Manuelidis, 1978b).

Recently a number of other human highly repeated DNA families have been described. Some of these are interspersed throughout the genome.

Maio et al. (1981b) have described a number of repeating restriction fragments that are present in primate DNA. The KpnI digestion of human DNA yields four prominent fragments visible above the background smear of DNA. These have been designated KpnIA (1.2 kb), KpnIB (1.55 kb), KpnIC (1.9 kb) and KpnID (1.8 kb). KpnIA and B fragments are present in the DNA of all anthropoid apes, including new world Capuchin monkeys (but not prosimians) whereas the KpnIC fragment was detected only in the genomes of old world monkeys.

The KpnID fragment seemed the most recently evolved repeating fragment and was found only in the DNA of higher primates (pongids and man). Probing KpnI digests of primate DNA with purified component α from the African green monkey or the human EcoRI 340 bp fragment showed that all four fragments appeared to have homology with the former, whereas only KpnIA and B hybridised with the latter. Partial KpnI digests of total DNA did not yield the pattern of multimeric fragments expected of tandemly arranged repeats and it was therefore concluded that these repeats may be interspersed.

The hybridisation of the KpnI fragments to component α and to the human EcoRI 340 bp fragment must be viewed with some caution because of the possible cross contamination of sequences (see Section 9.2). In this context it is surprising that neither the isolated
KpnIA or B fragments showed an internal 170 bp periodicity when digested with other restriction enzymes as might be expected of sequences homologous to these repeated DNAs (Section 9.3). It is also surprising that the KpnID fragment hybridised to component a because this fragment is most probably the 1.8 kb KpnI fragment described in Section 8.5 and a putative member of the Hinfl family. Thus this fragment is evident in KpnI digests of satellite III and it co-electrophoreses with the KpnID fragment in digests of total DNA, whereas the KpnIA, B and C fragments are not evident in satellite III. When KpnI digests of total human DNA are probed with satellite III the KpnD fragment hybridises strongly (along with a 3.4 kb fragment) whereas the KpnIA, B and C fragments are not detectable (data not shown). From the analysis of cell hybrids the KpnI 1.8 kb fragment may be specific to chromosome 15 (Section 8.5); it is certainly not interspersed throughout the genome. It probably represents a relatively recent segmental amplification of the Hinfl family, being restricted to pongids and man, and might be expected to occupy the equivalent chromosomal location in these species.

The AluI family of repeats comprises about 300,000 copies of a 300 bp sequence that is interspersed throughout the human genome and was first identified by the finding that most (approximately 60%) of the 300 nt sequences liberated by S1 digestion of human renatured DNA share a common site for AluI (Houck et al., 1979). The identification of the AluI family as a closely related group of sequences was subsequently confirmed by direct sequencing and cloned sequencing studies of S1 liberated 300 nt repeats (Rubin et al., 1980b; Deininger et al., 1981). AluI sequences have been found flanking a number of genes, e.g. insulin (Bell et al., 1980), γ globin (Jelinek et al., 1980) and ε globin (Baralle et al., 1980).
The sequencing studies have shown that the AluI repeats are distinct from the 170 bp repeat sequence and that they do not have a short underlying periodicity characteristic of the Hinfl family. Divergence between members of the AluI family seems to be random, nor do members of the family appear to be arranged into subfamilies showing any especially strong sequence homologies (Deininger et al., 1981). The absence of subgroups within the AluI family is in contrast to what is observed for the Hinfl family or the 170 bp repeat sequence. In these cases distinct discontinuities are observed in their sequence variation. These discontinuities may be so large that completely distinct satellite DNAs result, e.g. satellite II and satellite IV. This contrasting pattern of divergence may reflect the difference between tandemly repeated sequences and interspersed repeated sequences, the former being able to undergo rapid segmental amplifications thus generating the discontinuities of divergence within the sequence family as a whole.

AluI repeats share one feature in common with the 170 bp sequence family, namely their organisation as dimers. Sequencing studies have shown that each AluI repeat consists of a head to tail dimer of an original sequence of approximately 130 bp with an insert of 31 bp in one half of the dimer sequence compared with the other. The 130 bp Alu monomer is of similar size to the equivalent repeat found in the mouse genome (Krayev et al., 1980) and shares considerable homology with it. AluI sequences and the mouse equivalent are often flanked by short direct repeats (Bell et al., 1980; Jelinek et al., 1980; Baralle et al., 1980; Krayev et al., 1980) indicating that they may be transposable elements.

A repeated DNA 6.4 kb long has been identified 3 kb 3' to the human P globin gene (Adams et al., 1980). This sequence family has
about 3,000-4,800 copies per haploid genome (equivalent to about 1% of the genome) which are interspersed throughout the genome. This repeat does not appear to be composed of any smaller repeating units and in contrast to the AluI family of repeats does not appear to be transcribed.

9.5 Sequence Variation of Human Highly Repeated DNA

Highly repeated DNAs may be broadly classified as simple or complex. This classification is based on the length and homogeneity of the repeat unit. Thus, many crab satellite DNAs which are composed predominantly of poly dAT (Skinner, 1977) and the D. viridis satellite DNAs which are composed of 7 bp repeats (Gall and Atherton, 1974) are simple sequence DNAs, whereas the 1.688 g/cm\(^3\) satellite DNA of D melanogaster (Hsieh and Brutlag, 1979a) which is composed predominantly of a 354 bp repeat unit, is a complex highly repeated DNA. However, the distinction between simple and complex highly repeated DNA is not clear cut. For instance, when highly repeated DNAs from a variety of organisms are compared, a continuum of repeat lengths is apparent, e.g. 2 bp, crab (Skinner, 1977); 10 bp, D. melanogaster (Peacock et al., 1977); 23 bp, calf (Pech et al., 1979); 28 bp and 30 bp, mouse (Horz and Altenburger, 1981); 46 bp, calf (Poschl and Streeck, 1980); 81 bp, red crab (Christie and Skinner, 1980); 120 bp, rye (Bedbrook et al., 1980a); 170 bp, primates (Section 9.3); and 354 bp, D melanogaster (Hsieh and Brutlag, 1979a). Furthermore, some highly repeated DNAs, although being composed of short repeats exhibit such a high degree of heterogeneity between the different repeats that the sequence appears quite complex. The 1.706 g/cm\(^3\) calf satellite is comprised of a 23 bp repeat which may have been derived from an even shorter repeat unit of 11 bp or 12 bp
(Pech et al., 1979). Sequence analysis of the cloned regions shows that adjacent 23 bp repeating units differ markedly from each other and that at least 20 of the possible 23 positions have been altered by base changes or deletions. Conversely, some apparently complex sequences may reveal weak indications of a simpler substructure. For example, three related species of Hawaiian Drosophila share a 189 bp repeat which may have been derived from a repeat of about 35 bp (Miklos and Gill, 1981).

In humans the HinfI family of sequences, exhibiting an underlying 5 bp periodicity, are probably best classified as simple, whereas the 170 bp repeat sequence is probably best classified as complex. However, it should be noted that the human 170 bp repeat sequence and the related sequences that are present in different primate species (Section 9.3) suggest permutations on a simpler common prototype or 'core' sequence (Donehower et al., 1980). Over 50% of the sequence is characterised by pyrimidine runs of 4 or more nucleotides, in particular 5'CTTT 3', 5'TTTT 3', 5'CCTT 3' and 5'TTCC 3'. Pyrimidine runs, particularly C2T2 are prominent in the HinfI family and so it is possible that these two now quite distinct families of human highly repeated DNA are very distantly related.

The HinfI family is extremely heterogeneous and different members of the family may be so diverged from each other that they will only cross hybridise at reduced levels of stringency (Section 9.2). One of the main sources of this heterogeneity appears to be the independent evolution of the various members of the family at discrete chromosomal locations and this is paralleled by their arrangement into distinct higher order repeating periodicities.

Within each member of the HinfI family there is also considerable sequence heterogeneity. For example, satellite II is comprised of 3
basic repeats of 10, 12 and 15 bp. These are related to each other and have presumably arisen from an ancestral repeat by base changes and at least one deletion or insertion (Section 3.2). Even the basic repeating units (10, 12 or 15 bp) are heterogeneous as shown by their pyrimidine fingerprints and by base changes in HinfI or TaqI sites giving rise to 22 bp and 25 bp repeats. The 3.4 kb Y fragment also seems to have undergone considerable sequence variation such that some regions of the 3.4 kb repeat no longer appear to hybridise with other regions, thus generating very complex kinetics of reassociation (Section 7.6).

C band polymorphisms especially of human chromosomes 1, 9 and 16 (Craig-Holmes et al., 1973) and the Y chromosome (Unnerus et al., 1963) are frequent in the population. In a number of cases the variation in the size of the C band has been correlated with variation in the amount of human highly repeated DNA. McKay et al. (1979) have shown that C band polymorphisms of the human Y chromosome are correlated with the amount of 3.4 kb Y fragment present in that chromosome and Gosden et al. (1981a) have shown by in situ hybridisation that C band polymorphisms of chromosome I are correlated with the level of hybridisation of a cloned repeated human DNA which shows a number of similarities to satellite II (Section 3.5). Hybridisation of cRNA made from satellite IV to an individual with a very large C band on one chromosome 9 (approximately 110% larger than its homologue) showed a similar correlation with about 150% more grains hybridising to the C band region of the larger chromosome (data not shown). Thus expansion (and presumably contraction) of human highly repeated DNA, in particular the HinfI family, may underlie many of the C band polymorphisms and judging from the frequency of these polymorphisms in the population (Craig-Holmes et al., 1973) such
changes are frequent events. In the case of the polymorphisms of the Y chromosome the increases in the amount of the 3.4 kb Y fragment were consistent with a saltatory amplification of the sequence rather than a gradual increase (McKay et al., 1978).

The Hinfl family has not been well characterised in other primates. It is present as indicated by the fact that satellite IV hybridised strongly to gorilla and chimpanzee DNA in Southern 'blot' experiments (data not shown) and by the finding that the 3.4 kb Y fragment from humans, although absent in other primates does hybridise strongly to sequences present at the C band of a number of gorilla autosomes (Szabo et al., 1979). In addition, the KpnID fragment, described by Maio et al. (1981b) and present in humans and pongidae, is very probably the 1.8 kb KpnI fragment that is present on human chromosome 15 (Section 9.2).

The 170 bp repeat sequence also exhibits sequence variation in humans. This is indicated by the loss of DdeI sites creating higher multimers of the 170 bp periodicity (Section 6.4) and by the presence of variants in the 340 bp EcoRI fragment that were directly detected by sequencing (Wu and Manuelidis, 1980). One of the most significant features of this highly repeated DNA is its organisation into a dimer. It is postulated that this has allowed unequal crossing over between the two related halves thus giving rise to a non random distribution of HaeIII sites throughout the sequence and to recombinant sequence variants (Section 6.4). As for the Hinfl family, one of the main sources of sequence variation appears to have been the independent evolution of different members of the family at discrete chromosomal locations (Section 8.9).

It is not known how polymorphic the 170 bp repeat sequence is in the population. Its relationship to similar sequences present in
other primates has been discussed in Section 9.3. In general it seems that the processes that have been operating to generate sequence variation between species and within species are the same. In particular, the parallel between the sequence variation between different chromosomes within the human genome and the sequence variation between different primate species has been emphasised (Section 9.3).

In humans the HinfI family and the 170 bp repeat sequence are present on a number of different chromosomes and therefore they, or their ancestral sequences, must have been transposed at various times during their evolutionary history. In some cases when two members of the same family are compared (e.g. satellite II and satellite IV) they exhibit large differences in their fine structure, indicating that considerable sequence divergence has occurred after the transpositions and expansions from the ancestral array. In other cases different chromosomal sites seem to share a very similar basic sequence, but differ in the higher order arrangement of that sequence (e.g. satellite II, Section 8.5). The TaqI 11-mer of the 170 bp repeat sequence is present on more than one chromosome (Section 8.7). This repeat seems to be very similar at the different chromosomal sites because it is also defined by HaeIII and the HaeIII and TaqI sites seem to be at equivalent positions on different chromosomes. Thus relatively little sequence divergence has occurred in this repeating array after its transposition.

In conclusion a high level of sequence variation is observed in human highly repeated DNA. This is accounted for by a number of processes including base changes, deletions, recombinations, transpositions and expansions of tandem arrays. Sequence divergence is apparent in the fine structure of the human repeated DNA families. The high level of C band polymorphisms in the population indicates
that the contraction and expansion of arrays of highly repeated DNA may be a frequent event. Transpositions are indicated \textit{a priori} by the fact that members of the same sequence family are found on different chromosomes. This may occur by translocations. Cooke and Noel (1979) have shown that an extra fluorescent body on the short arm of chromosome 15 in 3 individuals and on the short arm of chromosome 13 in one individual contained highly repeated DNA (including the 3.4 kb Y fragment) derived from the Y chromosome.

9.6 Function of Highly Repeated DNA

It is often stated that highly repeated DNA is conserved, implying a functional role for these sequences. The guinea pig and the kangaroo rat, thought to have diverged about 50 million years ago, share a major repeat unit, \textit{GGGTTA 3'}, in their \textalpha{} satellites, as well as an array of similar variants and it has been proposed that these sequences have been conserved by functional constraints (Fry and Salser, 1977). However, use of the term conservation when highly repeated sequences are compared across different species implies that a true measure of sequence divergence has been made. This can be done for a unique sequence such as \textit{\beta{}} globin and rates of divergence have been calculated for the intronic and exonic segments of this gene (Efstratiadis et al., 1980). However, it is usually not possible to measure the true divergence of highly repeated DNA between species because of the substantial within species variation. Thus the HinfI family in humans (Section 9.2) is undoubtedly related to sequences that are present in other primate genomes, yet it would be impossible to measure the divergence of this entire family of sequences between humans and other species of primates because of the very large variation that occurs within humans - between different individuals,
between different chromosomes and even between sequences that occur in tandem on the same chromosome.

In some cases the variation within a particular highly repeated DNA may be quite small. Thus in the *D. melanogaster* 1.688 g/cm³ satellite a large fraction of the total sequence divergence may be represented by variation at a few positions only (Hsieh and Brutlag, 1979a). Such non-random sequence divergence has been thought by some workers to indicate functional constraints acting on the sequence, but such a situation could also arise by the amplification and fixation of repeat units containing these particular changes. Similarly a low level of sequence variation observed within a stretch of repeated DNA may simply be the result of a recent amplification rather than any evolutionary conservation. Such amplification events would also tend to obscure variants present in low copy number and would serve to underestimate the overall level of sequence divergence.

If the sequence *per se* of highly repeated DNAs were of significance then some form of interaction between them and proteins might be expected. Such an interaction has been described for the *D. melanogaster* 1.688 g/cm³ satellite and a protein has been partially purified with a specific affinity for this sequence (Hsieh and Brutlag, 1979b). However as yet no other specific protein/highly repeated DNA interaction has been described. When the possibility of such interactions is considered for the more heterogeneous highly repeated DNAs such interactions must either be fairly non-specific, or, if they are specific, then the number of binding proteins will approach the number of sequence variants. Miklos and Gill (1981) have calculated that for the known variants of the 1.706 g/cm³ calf satellite, if protein - highly repeated DNA interactions were specific then the number of proteins required would exceed the normal coding capacity of the genome!
It has been generally agreed for some time now that the informational content of most highly repeated DNAs precludes them from being translated into any sensible protein structure. Their transcription has been observed in the lampbrush chromosome stage in *Triturus cristatus aornifex* by Varley *et al.* (1981) but these workers consider that the most likely explanation for this phenomenon is random overrun transcription from meaningful translated sequences into regions containing highly repeated DNA.

A number of specific functions for highly repeated DNAs have been proposed. One of the most popular speculations was that these sequences mediated chromosome pairing (Walker, 1971) and in this context the quite distinct chromosomal locations of a number of the human sequences would seem to be supportive of this view. However, such hypotheses are not supported by the direct experimental evidence. Thus Yamamoto and Miklos (1977) have shown that the chromosomal segregation of the X chromosome of *D. melanogaster* is not appreciably altered by deletions of centromeric heterochromatin. Chromosome pairing seems to depend on euchromatic homologies. For example, when a small fragment which contains the centromeric region of chromosome 1 and the telomeric region of chromosome 13 is added to a normal mouse genome it preferentially pairs with the telomeric region of chromosome 13 and not the centromeric region of chromosome 1, which is rich in highly repeated DNA (de Boer and de Groan, 1974). Similarly in *D. melanogaster* a small chromosome 4 prefers to associate on the basis of euchromatic homology rather than on centric homology (Yamamoto, 1979a).

It has been proposed that highly repeated DNAs may be involved in speciation. One hypothesis holds that differences in highly repeated DNA on homologues may lead to pairing problems, resulting
in infertility and hence leading to sympatric speciation. The high level of highly repeated DNA polymorphisms encountered in natural populations and the direct experimental evidence on chromosome pairing makes this theory untenable. It has also been proposed that highly repeated DNAs may underlie speciation by being responsible for chromosomal rearrangements. In particular, Hatch et al. (1976) claim that there is a correlation between the presence of large amounts of highly repeated DNA and the frequency of chromosomal rearrangements. However, in the first place, it is not clear how important chromosome rearrangements are to speciation; from an analysis of 22 species of Hawaiian Drosophila, Carson et al. (1967) concluded that major chromosomal rearrangements might be incidental to speciation. In a comparison of *R. vili* and *R. sordidus* Miklos et al. (1980) concluded that the eleven major centric arrangements in these two species had been fixed in the absence of large amounts of highly repeated DNA.

The theories that link highly repeated DNA with speciation have considered the variation between species and to some extent have ignored the extensive variation within a species. Furthermore, other parts of the genome also change during speciation. For example, in N. American Plethodontid salamanders, Mizuno and Macgregor (1974) have shown huge differences in the C value of different genomes which affect both unique and repetitive portions.

Tandem arrays of highly repeated DNA are not entirely without effect on the genomes which contain them. Experimental data reveals that the probability of meiotic recombination decreases near to blocks of highly repeated DNA and that such effects decrease with distance from these blocks (Miklos and Nankivell, 1976; John and Miklos, 1979). When regions of highly repeated DNA are experimentally deleted not only are the recombinant characteristics of the chromosome
altered, but other chromosomes are also perturbed (Yamamoto, 1979b). Thus the amount of highly repeated DNA at a given location can exert an effect on the combination of genes during evolution although the magnitude of this effect is unknown.

The lack of any apparent functional constraints on the sequence of highly repeated DNA and the lack of any demonstrated somatic function have led a number of workers to postulate a 'selfish' role for this type of DNA (Doolittle and Sapienza, 1980; Orgel and Crick, 1980). 'Selfish' DNA sequences are considered to have no functional significance but their presence stimulates the further accumulation of sequences of the same kind. The concept of 'selfish' DNA implies a level of intragenomic selection and introduces the idea of preferred replicators, i.e. that some sequences by their very nature are better able to accumulate than others. In this context Dover (1980) has pointed out that if highly repeated DNAs were preferred replicators then one might predict a limited range of sequence types as opposed to the variety of sequence organisation that is observed. It is possible that there is some as yet unrecognised commonality between different highly repeated DNAs, or that a variety of mechanisms exist for their accumulation, each favouring a different type of sequence organisation.

In conclusion, no clear somatic function has been demonstrated for highly repeated DNA and it is felt likely that they may be of no phenotypic significance at all. In this context it is felt that the distinct chromosomal location of members of the Hinl family and members of the 170 bp repeat family merely represents the independent evolution of these sequences at different locations. These sequences however may mediate important effects on the recombination of the chromosomes on which they are present.
9.7 Concluding Remarks

This thesis describes the structure and evolution of human highly repeated DNA. The confusing state of affairs described for human satellite DNAs (Section 1.4) has been resolved by the identification of the sequence families that comprise them. Independent evolution at different chromosomal sites has been one of the main sources of sequence variation. In the case of the HinfI family this has been so marked that quite distinct sequences are now found on different chromosomes.
10.1 DNA Preparation

Placental DNA

Large quantities of human DNA were prepared from fresh placentae by a modification of the procedure of Marmur (1961). The tissue was chopped roughly with scissors in the cold and washed with physiological saline. To about 300 g of this material an equal volume of 0.3 M sodium perchlorate, 0.3 M Tris-HCl, pH 8.3 was added and the tissue homogenised in a Potter-Elvehjem Teflon homogeniser. 10% sodium dodecyl sulphate (SDS) was added to give a final concentration of 2%. After bringing to room temperature, an equal volume of water saturated phenol was added and the mixture agitated gently for ten minutes. Phenol and aqueous phases were separated by centrifugation in a Sorvall Type GSA rotor on a Sorvall RC2-B refrigerated centrifuge at 6000 rev/min for 20 minutes. The aqueous phase was re-extracted a number of times with chloroform:octan-2-ol (24:1) until the interphase was reasonably clear. 2 volumes of ethanol were added, the precipitated nucleic acid was spooled out onto a glass rod and then redissolved in 100-200 ml of 50 mM Tris-HCl, 10 mM NaCl, 5 mM EDTA pH 8.0 (Buffer A). To this solution was added boiled pancreatic RNAase A (Sigma) to 50 \( \mu \)g/ml and RNAase T1 (Worthington) to 1 \( \mu \)g/ml. The solution plus RNAase was dialysed vs. 2 litres of Buffer A at 37°C for 4 hours. At the end of this period SDS was added to 0.5% and Proteinase K (Boehringer) to 50 \( \mu \)g/ml, the mixture transferred to fresh tubing and dialysed against fresh buffer A for 12 hours at 37°C. The preparation was then extracted once with phenol and then a number of times with chloroform:octan-2-ol until the interphase was clear. The purified DNA was ethanol precipitated, spooled out and dissolved in 0.001 M EDTA, 0.01 M Tris pH 8.0, to
give a final concentration of approximately 1 mg/ml. The yield of DNA obtained by this procedure was in the order of 100 mg DNA/200 g placental tissue.

The concentration and purity of the sample were measured in a Beckman DB-GT or a Unicam SP750 spectrophotometer using either Beckman microcells (Type 97260) or Unicam 2 ml UV cells. The concentration of DNA was estimated assuming an absorbance at 260 nm (A_{260}) of 1.0 O.D unit equivalent to a DNA concentration of 50 μg/ml. DNA preparations with an A_{260}/A_{280} ratio between 1.7 and 1.8 and an A_{260}/A_{320} ratio >100 were considered pure.

DNA from Cell Lines

DNA was extracted from cell cultures by a modification of the above procedure designed to increase the yield from small amounts of material e.g. from about 10^7-10^8 cells. Cells were harvested by spinning down at 1000 rpm in a SS 34 rotor. The cells were resuspended in Dulbecco's solution and centrifuged again at 1000 rev/min. The washed cell pellet was resuspended in 5 mls of Buffer A. An equal volume of Buffer A, 1% SDS was then added while vortexing the cell suspension. Proteinase K was then added, to a concentration of 50 μg/ml and the cell lysate incubated for 4 hours at 37°C. After incubation the mixture was phenol extracted once, ether extracted twice and 23.1 g CsCl were added. The solution was made up to a final volume of 33 ml with distilled water to give a final density of 1.700 g/cm^3 and transferred to a Sorvall ultracentrifuge tube (Type 03141), covered with 1 ml of paraffin oil and then centrifuged overnight at 42,000 rev/min in a Sorvall TV 850 vertical rotor on an OTD-2 ultracentrifuge at a temperature of 18°C.

After centrifugation the ultracentrifuge tube was attached, by means of a three-way tap, to a reservoir of paraffin oil and a syringe.
The bottom of the tube was pierced with a hot needle and 1 ml aliquots removed by pumping the same volume of paraffin oil into the top of the tube from the reservoir via the syringe.

The absorbance at 260 nm was measured on a spectrophotometer, the DNA containing aliquots pooled and then dialysed for 24 hours against several changes of 0.001 M EDTA, 0.01 M Tris pH 8.0. DNA was pelleted by spinning overnight at 4°C in an MSE 10 x 10 or 8 x 35 angle-head titanium rotor on an MSE superspeed 65 ultracentrifuge at 30,000 rev/min. The pelleted DNA was dissolved in a suitable volume of 0.001 M EDTA, 0.01 M Tris pH 8.0.

10.2 Satellite DNA Preparation

Satellite DNA was prepared from total human DNA in preparative \( \text{Ag}^+ \text{Cs}_2\text{SO}_4 \) gradients (Corneo et al., 1970). Three satellites were isolated from the light side of main band DNA and one from the heavy side.

The three light side satellites were prepared in the following manner. First cycle gradients were run at a DNA concentration of 70 \( \mu \text{g/ml} \) in a total volume of 33.0 ml. DNA was dialysed into 0.1 M \( \text{Na}_2\text{SO}_4 \) and 0.1 M borate buffer (pH 9.1) added to a final concentration of 0.005 M. 0.001 M \( \text{AgClO}_4 \) was added dropwise to give a \( \text{Ag}^+/\text{DNA-P} \) ratio of 0.24. Solid \( \text{Cs}_2\text{SO}_4 \) (Merck) was added to give a final density as judged on a Bellingham and Stanley refractometer of 1.48 g/cm\(^3\).

The gradients were centrifuged to equilibrium in an OTD-2 ultracentrifuge using a TV 850 vertical rotor at 44,000 rev/min at 18°C for 12 hours. 0.3 ml fractions were collected from the top by piercing the bottom of the ultracentrifuge tube (Sorvall Type 03141) with a syringe needle and pumping in Fluorinert (ISCO Ltd) with an LKB peristaltic pump at a rate of 1 ml/minute. The fractions were diluted to 0.6 ml.
Figure 10.1
First cycle $\text{Ag}^+\text{Cs}_2\text{SO}_4$ gradients.

i  Light side gradient. Satellite DNAs were prepared from the light side of main band DNA in $\text{Ag}^+\text{Cs}_2\text{SO}_4$ gradients at a $\text{Ag}^+/\text{DNA-P}$ ratio of 0.24 by the methods outlined in Section 10.2. Three cuts were taken, a, b and c, which correspond to satellites I, IV and III respectively.

ii  Heavy side gradient. Satellite II (cut d) was prepared from the heavy side of main band DNA in $\text{Ag}^+\text{Cs}_2\text{SO}_4$ gradients at a $\text{Ag}^+/\text{DNA-P}$ ratio of 0.36.
Increasing Density:

DECREASING DENSITY:

A260

FRACTION
with 0.1 M Na₂SO₄ and the A₂₆₀ read in a Unicam UV microcell on a Unicam SP1750 spectrophotometer.

Figure 10.1.1 is a typical first cycle gradient profile, with two distinct satellite peaks and a shoulder to the main band. The arrows in Fig. 10.1.i indicate the cuts of the gradient taken. Cuts a and b, pooled from each individual gradient, were dialysed extensively against 5 M NaCl, 0.01 M Tris pH 8.0 to remove silver ions and then dialysed into 0.001 M EDTA, 0.01 M Tris pH 8.0. The DNA was pelleted in an MSE 8 x 35 Ti angle head rotor and redissolved in 2.94 ml 0.01 M Tris pH 8.0 and 3.78 g solid CsCl added to give 4.2 ml of a solution with a density of 1.70 g/cm³. These second cycle gradients were centrifuged at 50,000 rev/min at 18°C to equilibrium, fractionated into 0.1 ml aliquots and the A₂₆₀ read. Tubes containing satellite DNA were pooled, dialysed into 0.001 M EDTA, 0.01 M Tris pH 8.0, the DNA pelleted and redissolved in 0.001 M EDTA, 0.01 M Tris pH 8.0 to a concentration of 200 µg/ml; samples were stored frozen at -20°C.

Gradient cuts from the shoulder of the main band (Cut c) were subject to an extra round of Ag⁺ Cs₂SO₄ centrifugation, at a mean density of 1.44 g/cm³, before being finally centrifuged in neutral CsCl.

Isolation of satellite DNA from the heavy side of main band was carried out in a similar manner except that the Ag⁺/DNA-P ratio was 0.36 and the gradients were fractionated from the bottom (Section 10.1). Figure 10.1.ii is a typical gradient profile; d indicates the cut taken for recycling in neutral CsCl.

Analytical Ultracentrifugation

CsCl analytical ultracentrifugation was carried out on an MSE Centriscan using an MSE analytical rotor. About 0.5 µg each of sample DNA and marker DNA (Micrococcus lysodekticus) and in 0.01 M Tris pH 8.0 were added to optical grade CsCl and the density adjusted by refractometry to 1.700 g/cm³. Samples were centrifuged to equilibrium at
Figure 10.2

Analytical CsCl profiles of purified human satellite DNAs.

CsCl analytical ultracentrifugation was carried out on an MSE Centriscan using an MSE analytical rotor. The absorbance profile at 254 nm was scanned using the U.V. optical system and the buoyant density calculated from the equation

\[ \rho_s = \rho_m + 9.2 \times 10^{-3} (R_s^2 - R_m^2) \] (Flamm et al., 1969a) where

- \( \rho_s \) = buoyant density of sample DNA;
- \( \rho_m \) = buoyant density of marker DNA (Micrococcus lysodeikticus, \( \rho = 1.731 \text{ g/cm}^3 \)),
- \( R_s \) = radius of sample peak, and
- \( R_m \) = radius of marker peak.

i Satellite I, \( \rho = 1.687 \text{ g/cm}^3 \)

ii Satellite II, \( \rho = 1.694 \text{ g/cm}^3 \)

iii Satellite III, \( \rho = 1.697 \text{ g/cm}^3 \)

iv Satellite IV, \( \rho = 1.700 \text{ g/cm}^3 \)
<table>
<thead>
<tr>
<th>Satellite</th>
<th>$\rho$ g/cm$^3$</th>
<th>$\text{Ag}^+\text{Cs}_2\text{SO}_4$ Cut (Fig. 10.1)</th>
<th>%Genomic DNA</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>1.687</td>
<td>a</td>
<td>=0.5%</td>
<td>1, 2, 3, 4</td>
</tr>
<tr>
<td>II</td>
<td>1.694</td>
<td>d</td>
<td>=2.0%</td>
<td>4, 5, 6</td>
</tr>
<tr>
<td>III</td>
<td>1.697</td>
<td>c</td>
<td>=1.5%</td>
<td>4, 7, 8</td>
</tr>
<tr>
<td>IV</td>
<td>1.700</td>
<td>b</td>
<td>=2.0%</td>
<td>4, 9</td>
</tr>
</tbody>
</table>

**TABLE 10.1**  
The Human Satellite DNAs.  
The human satellite DNAs were identified by their position in 1st cycle $\text{Ag}^+\text{Cs}_2\text{SO}_4$ gradients and by their buoyant density in neutral CsCl gradients, as previously described by other workers, references 1-9.  
44,700 rev/min for 20 hours at 25°C. The absorbance profile was scanned using the UV optical system (at 254 nm) and the XY recorder of the Centriscan. Buoyant densities were calculated by the method of Flamm et al. (1969a).

The procedure used to isolate the human satellite DNAs gave high MW homogeneous DNA products as judged by their banding profiles in the analytical ultracentrifuge (Fig. 10.2). The satellites were identified according to Corneo et al. (1970) by their buoyant density and position in the 1st cycle Ag⁺Cs₂SO₄ gradients (Table 10.1).

10.3 Restriction Endonucleases

A wide variety of Type II restriction endonucleases were used to investigate the structure and arrangement of the four human satellites. These were purchased from either New England Biolabs or Bethesda Research Laboratories. The recognition sequence and digestion conditions for the restriction enzymes used in this study are tabulated in Table 10.2.

In general, a suitable amount of enzyme was added to ensure complete digestion of the DNA under study. This was checked, either by adding fresh enzyme to an aliquot of the digestion mixture and continuing incubation prior to electrophoresis, or by including an aliquot of marker DNA (lambda or ψXRF DNA) in the digestion mixture.

The two suppliers of restriction enzymes have different unit definitions for the enzyme activity. Biolabs considers 1 unit as the amount required to completely degrade 1.0 μg of lambda DNA in 50 μl at 37°C in 15 minutes, whereas the Bethesda unit is the amount of enzyme that gives complete digestion in 1 hour. The Biolabs unit does not equal four times the Bethesda unit because many enzymes are progressively inactivated during 60 minutes of incubation. In practice,
<table>
<thead>
<tr>
<th>EndoR</th>
<th>Recognition Sequence</th>
<th>Digestion Buffer</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Tris-HCl pH 7.5 (mM)</td>
<td>NaCl (mM)</td>
<td>MgCl₂ (mM)</td>
<td>β-mercaptoethanol (mM)</td>
</tr>
<tr>
<td>AluI</td>
<td>AGCT</td>
<td>6</td>
<td>50</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>DdeI</td>
<td>CINAG</td>
<td>100</td>
<td>100</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>EcoRI</td>
<td>GAATT</td>
<td>100</td>
<td>80</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>HaeIII</td>
<td>GGCC</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>HinfI</td>
<td>GANTC</td>
<td>6</td>
<td>50</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>HindIII</td>
<td>AAGCTT</td>
<td>20</td>
<td>50</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>¹HpaII</td>
<td>COCG</td>
<td>20</td>
<td>-</td>
<td>6</td>
<td>-</td>
</tr>
<tr>
<td>KpnI</td>
<td>GGTACC</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>MboI</td>
<td>GATC</td>
<td>6</td>
<td>150</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>MspI</td>
<td>CCGG</td>
<td>20</td>
<td>-</td>
<td>6</td>
<td>-</td>
</tr>
<tr>
<td>²TaqI</td>
<td>TCGA</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>XbaI</td>
<td>TCTAGA</td>
<td>6</td>
<td>100</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

**TABLE 10.2**

List of Restriction Enzymes

The digestion conditions were those recommended by the manufacturer.

¹HpaII is blocked by methylation of the internal C residue.

²All incubations were carried out at 37°C except in the case of TaqI which was used at 45°C or 65°C.
it was found that 2 Bethesda units or 1 Biolabs unit per μg of DNA usually gave complete digestion after 1 hour. Only in exceptional circumstances, e.g. when a restriction site was very highly repeated in a purified satellite DNA, was it found necessary to increase the enzyme concentration.

10.4 Gel Electrophoresis

DNA that had been digested with restriction endonucleases was analysed using both agarose and polyacrylamide gel systems.

Agarose Gel Electrophoresis

Agarose gels varying in concentration from 0.8%-2.5% w/v were used to separate restriction fragments between 100 bp and 5000 bp in size. Digested DNA samples were made 5% in glycerol and loaded onto 20 cm x 20 cm x 0.3 cm slab gels. Electrophoresis was carried out for about 12 hours at 10 mAmp in E-Buffer (40 mM Tris-HCl, 20 mM Sodium acetate, 1 mM EDTA, pH 8.2) using a Chandos PSU powerpack. Gels were stained for 15 minutes in 2 μg ethidium bromide/ml (Sigma), excited with UV light from a germicidal tube and photographed on a 5 x 4 Monorail camera through a Wratten W25 barrier filter using +PanX film. Lambda Cl857 phage DNA, digested with a mixture of EcoRI and Hind III restriction endonucleases and φX 174 RF DNA digested with Hae III, Hind I or Taq I were used as molecular weight markers (Table 10.2).

Figure 8.1 is a photograph of a 1.5% agarose gel that shows the electrophoretic separation of EcoRI/Hind III digested lambda DNA and Hae III digested φX 174 RF DNA. The negative of this photograph was scanned on a Joyce-Loebl densitometer and the distance migrated by each band plotted against the log of its molecular weight (in base-pairs). Figure 10.3 illustrates that the relationship between log MW and the distance migrated is linear over much of the gel (from 300 bp-3400 bp) and that the molecular weight of a restriction fragment can be reliably estimated by reference to this curve.
Figure 10.3
Calibration of agarose gels.

Tracks a and b of Figure 7.1 were scanned with the Joyce-Loebl microdensitometer. The trace obtained is a linear expansion of the gel (2:1 or 5:1) and thus greatly facilitates the measurement of the distance migrated by a given restriction fragment.

i The log MW of *XRF* DNA fragments after restriction by HaeIII (●) and the log MW of lambda DNA after restriction by EcoRI and HindIII (○) were plotted against the distance migrated.

ii The areas under the peaks were estimated by cutting them out and weighing. The % of DNA in a fragment, established in this fashion was plotted against its MW (○) and compared to that expected theoretically (○).
The relative amounts of DNA in restriction fragments were estimated by cutting out the corresponding peaks on the scan and weighing them. Figure 10.3 indicates that this simple method gives a reasonably accurate measurement of the relative amounts of DNA although bands with large amounts of DNA in them may be slightly underestimated.

Polyacrylamide Gels

Polyacrylamide gels were prepared in the following manner. A 30%/0.8% solution of acrylamide/bis-acrylamide (Bio-Rad) was diluted in water and 10 x E buffer (E buffer is 50 mM Tris-borate, 1 mM EDTA pH 8.3) to the required volume and acrylamide concentration. The gel was polymerised by the addition of N, N, N' N'-tetramethyl-ethylenediamine (TEMED) and 10% ammonium persulphate (for a 20% acrylamide gel, 60 µl TEMED and 200 µl 10% ammonium persulphate were added per 100 ml of gel; for gels of lower acrylamide concentration the amount of these were reduced accordingly). Gels were cast between 20 cm x 20 cm or 20 cm x 40 cm glass plates using 1.5 mm spacers. Samples were loaded in 5% glycerol with a mixture of 0.025% Orange G, 0.025% bromphenol blue, 0.025% xylene cyanol and electrophoresed at 8 mAmp.

7M urea, 20% acrylamide gels were prepared similarly except that solid acrylamide, bis and urea were weighed out together and then dissolved in E buffer.

In general, DNA analysed by acrylamide gels was radioactively labelled and so was visualised by autoradiography. The gel was backed onto 3 mm Whatman filter paper, placed into a plastic bag, clamped between two aluminium sheets with X-ray film (Kodak X-Omat H) and frozen at -70°C in a light-tight box.

Small, single-stranded restriction fragments were sized on 7M urea, 20% acrylamide gels by reference to the marker dyes bromphenol blue and xylene-cyanol, whose electrophoretic mobilities in this system...
Figure 10.4

Calibration of acrylamide gels.

$\times$RF DNA labelled with $^{32}$P by nick translation was restricted by HinfI or TaqI and the products separated on a, 20 cm or b, 40 cm 20% polyacrylamide gels. The log MW of the fragments was plotted against the distance migrated.

●, $\times$RF/TaqI; ○, $\times$RF HinfI.

Note that under these conditions TaqI fragments migrate a little faster (± 1-2 bp) than HinfI fragments.
DISTANCE MIGRATED (cm)

Log MW

2.2
2.1
2.0
1.9
1.8
1.7
1.6
1.5
1.4
1.3

2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32

a b

1.9
1.8
1.7
1.6
1.5
1.4
1.3

DISTANCE MIGRATED (cm)
are equivalent to 10 nucleotides and 26-27 nucleotides, respectively (Maxam and Gilbert, 1978). That the mobility of bromophenol blue was equivalent to 10 nucleotides was checked by comparing its mobility relative to poly (dT)$_{10}$ (Miles) end labelled by polynucleotide kinase (Section 10.6).

The size of double stranded fragments in the range of 20-100 bp was estimated by using restriction digests of φX RF DNA radioactively labelled either by nick translation or with polynucleotide kinase (Section 10.6).

Figure 10.4 is the plot of log MW vs distance migrated for restriction fragments generated with Hinf 1 and Taq 1 from φX RF DNA and electrophoresed on 20 cm and 40 cm 20% polyacrylamide gels. It is apparent in both cases that fragments cut with Hinf 1 run slightly slower than do those cut with Taq 1, presumably reflecting the fact that Hinf 1 fragments have 3 nucleotide single-strand tails, whereas Taq 1 fragments have 2 nucleotide single-strand tails. In sizing fragments generated from a satellite DNA by a particular restriction enzyme, it was therefore deemed necessary to size relative to φX RF fragments cut with the same enzyme. An accuracy in sizing of about ±1 bp was estimated for fragments up to about 40 bp.

DNA restriction fragments were recovered from gels in two ways. Larger fragments, separated on agarose gels were cut from the gel and recast with fresh 1.5% agarose in 6 cm x 0.5 cm glass tubes. A piece of dialysis tubing, tied at one end and filled with 0.5 mls of E-buffer was attached to one end of this cylindrical gel, taking care to exclude all air bubbles. DNA was electrophoresed out of the gel using a disc gel apparatus at 10 mAmp/tube into the dialysis membrane, immersed in the bottom (+ve) tank. Electrophoresis was monitored with bromophenol blue. At the end of the electrophoresis the current was reversed for
30 seconds to prevent the DNA from sticking to the dialysis membrane. If the DNA had been stained with ethidium bromide the preparation was made 2 M with sodium chloride and extracted once with n-butanol and then twice with water saturated ether. After dialysis into 0.01 M Tris, 0.001 M EDTA, pH 8.0 (Tris-EDTA) the DNA was precipitated overnight at -20°C by the addition of 0.1 volume of 20% sodium acetate, 2 volumes of absolute ethanol. The precipitated DNA was collected by spinning at 30,000 rev/min for 30 minutes in an MSE 6 x 5.5 ml swing out rotor, and dissolved in a small volume of 0.01 M Tris, 0.001 M EDTA pH 8.0.

Small restriction fragments, <100 bp, were isolated in the following manner. The band, cut from the gel, was chopped up with a scalpel blade and then passed several times through a 21G hypodermic needle, in a small volume of Tris-EDTA into a 2 ml Eppendorf centrifuge tube. An equal volume of water saturated phenol was added to the macerated gel suspension and the phases mixed by vortexing. After repeating this a number of times over a period of several hours the aqueous phase was separated by spinning for 2 minutes at full speed on an Eppendorf (Type 3200) centrifuge, ether extracted twice and ethanol precipitated.

10.5 Southern Transfer and Filter Hybridisation

DNA in agarose gels was denatured in situ, at room temperature with 1.5 M NaCl, 0.5 M NaOH for 1 hour, neutralised for 1 hour with 3 M NaCl, 1.5 M Tris, pH 7.0 and the DNA transferred by blotting onto a 20 cm x 20 cm nitrocellulose membrane filter (Schleicher and Schull BA 85, 0.45 μm) by the method of Southern (1975). Transfer was carried out overnight in 20 x SSC (SSC is 0.15 M NaCl, 0.015 M Na\textsuperscript{+} citrate), and the DNA baked onto the filters at 80°C. Subsequently, hybridisation was carried out by the method of Jeffreys and Flavell (1977).
The baked filter was pre-incubated for 3-6 hours at 65°C in a solution of 0.02% polyvinylpyrrolidone, 0.02% Ficoll (Fisons), 0.02% BSA, Denhardt (1966). The $^{32}$P DNA probe, labelled by nick translation (Section 10.6) was denatured by heating for 5 minutes at 100°C. In general, between 2-5 x $10^6$ dpm of nick translated probe, so between $10^7$ x $10^8$ dpm/μg (Section 10.6) was added to 20 ml of hybridisation mixture (3 x SSC, 0.1% SDS) and the filter hybridised overnight at 65°C by rolling in a stoppered tube or shaking in a sealed plastic bag. After hybridisation the filter was washed with several changes of 2 x SSC at 65°C, backed onto 3 mm Whatman filter paper, put into a plastic bag and autoradiographed using Kodak X-OMAT X-ray film. Exposure was either at room temperature or at -70°C using intensifying screens.

After the initial wash and exposure some filters were washed in 0.12M PB (PB is phosphate buffer, pH 6.8) at increasingly higher temperatures and autoradiographs of the filters were made at each incremental level of temperature.

10.6 In-vitro Labelling of DNA

Nick Translation

$^{32}$P nick translated DNA was prepared by a modification of the procedure of Rigby et al. (1977). For purified satellite DNAs and marker DNAs (0.5 - 1 μg of each) 30 μCi of an $α^{32}$P nucleotide triphosphate (usually TTP), specific activity 400 Ci/mmol, obtained from the Radiochemical Centre, Amersham, was dried down with an air pump.

DNA; 10 x nick translation buffer (1 x nick translation buffer is 50 mM Tris-HCl (pH 7.8), 5 mM MgCl$_2$, 10 mM β-mercaptoethanol, 50 μg/ml BSA), H$_2$O and 0.1 mM solutions of the other three non-labelled triphosphates (Sigma) added to give a final DNA concentration of
10 μg/ml and cold triphosphate concentration of 0.0018 mM. The mixture was preincubated at 13°C for 10 minutes and then 1 unit/μg DNA of DNA polymerase I (Boehringer) added and the mixture incubated for 60 minutes at 13°C. After incubation the specific activity of the nick translated product was determined by spotting an aliquot from the reaction mix onto a Whatman GF/B filter, precipitating in 5% trichloroacetic acid, washing in ethanol and counting the filter in 5 ml of a toluene based scintillation cocktail (PPO/PPOP). The rest of the mixture was extracted twice with chloroform-octanol 24:1 and separated from unincorporated nucleoside triphosphates by chromatography on Sephadex G50 (medium). Maximal specific activity was about 3 x 10^7 dpm/μg DNA.

Very small quantities of DNA, <0.1 μg e.g. restriction fragments prepared from a gel (Section 10.4.) were nick translated in a similar fashion except that deoxy nucleoside triphosphates at a specific activity of 2000 Ci/mmol (Amersham) were used and the reaction volume was reduced to 20 ulitres. The amounts of cold triphosphates and DNA polymerase I were correspondingly reduced. The specific activity of the DNA was in the region of 10^8 dpm/μg.

3H nick-translated DNA was also prepared in a similar fashion except that all four triphosphate precursors (Amersham) were labelled each at a concentration of 0.0018 mM in the reaction mix. The specific activity of the labelled product was usually about 3 x 10^6 dpm/μg.

End-labelling with Polynucleotide Kinase

Restriction fragments were labelled at the 5' termini by means of polynucleotide kinase. After restriction 2 units of bacterial alkaline phosphatase (BRL) per μg of DNA was added and the mixture incubated a further 30' at 37°C. The preparation was then phenol extracted once and ether extracted twice. The volume was checked and
an equal volume of 2 x kinase buffer added (kinase-buffer is 50 mM Tris-HCl pH 9.5, 10 mM MgCl₂, 5 mM dithiothreitol, 5% glycerol).

This solution was transferred to an Eppendorf tube in which γ³²P ATP, specific activity 2000 Ci/mmol (Amersham) had been dried down by an air pump. 20-30 pmols of γ³²P ATP were used per µg of restricted DNA. To this mixture 2 units/µg DNA of polynucleotide kinase were added and the preparation incubated for 30' at 37°C. At the end of this period the mixture was heated to 65°C for 5' to inactivate the polynucleotide kinase and the labelled DNA separated from unincorporated precursors, by chromatography on Sephadex G50. Incorporation varied from 2 x 10⁵ - 2 x 10⁶ dpm/µg.

10.7 Pyrimidine Tract Analysis

DNA was fingerprinted by depurinating the nick translated product and separating the pyrimidine tracts by thin layer chromatography (Southern and Mitchell, 1971). ³²P labelled DNA in about 20µl H₂O was mixed with 2 volumes of 3% (w/v) diphenylamine (Koch Ltd.) in 98% formic acid (Burton, 1967). After incubation at 30°C overnight the mixture of pyrimidine oligomers, of general formula Cₙ₋₅₇₇₅₄ₙ₊₁Pₙ₊₁ was extracted once with 6 volumes of water saturated ether, then five times with an equal volume of ether, lyophilised and then redissolved in 1-2 µlitres of double d H₂O. This solution was applied, using a 'Drumond' microcapillary pipette, to a 20 cm x 20 cm "Polygram" Cel 300 PEI TLC plate (Camlab). Before development a 20 cm wide piece of 3 mm filter paper was stapled to the top of the plate so that chromatography could be continued after the solvent front had reached the top of the plate. Development in the first dimension was in 2 M pyridine-formate pH 3.4 and in the second in 1 M LiCl. In between the first and second dimension runs pyridine-formate was removed by
Figure 10.5

The separation of pyrimidine tracts by thin layer chromatography.

DNA labelled by nick translation was depurinated, applied to a TLC plate and pyrimidine tracts were separated by development in the first dimension with 2M pyridine formate, pH 3.4, followed by development in the second dimension with 1M LiCl. After chromatography the plate was exposed to X-ray film.

i Pyrimidine fingerprint of nick translated satellite II (labelled with dCTP and dTTP).

ii Oligonucleotide map of i (after Southern and Mitchell, 1971).

1, first dimension; 2, second dimension; OR, origin.
leaving the plate overnight in a fume cupboard. LiCl was removed by washing the plate in anhydrous methanol (Randerath and Randerath, 1967). The plate was exposed against Kodak X-Omat H X-ray film. Figure 10.5.i is a fingerprint of a satellite DNA, nick translated with dCTP and TTP each at the same specific activity and Figure 10.5.ii is an oligonucleotide map of this fingerprint.

Pyrimidine oligonucleotides of the general formula $C_{n}T_{m}P_{(n+m)}$ were obtained by end labelling their 5' termini with polynucleotide kinase (Szelký and Sanger, 1969) and separated in the same way.

The amount of radioactivity in a spot was quantified by cutting it out with reference to the autoradiogram of the plate, scraping the layer from the plastic backing into 3 ml of $H_{2}O$ in a scintillation vial, adding 10 ml 'Instagel' (Packard) scintillation cocktail and counting on a Packard Tri-Carb scintillation counter.

Nucleotides were eluted from a spot onto phosphocellulose flags (made from Whatman P80 phosphocellulose paper) in 2 M-triethylammonium carbonate (30% triethylamine, saturated with carbon dioxide), washed in ether and eluted from the flags in distilled $H_{2}O$ (Sanger et al., 1965).

10.8 High Pressure Liquid Chromatography

The 5' terminal base composition of DNA labelled with polynucleotide kinase was determined on a DuPont 830 liquid chromatograph using an A179 ion exchange column. To the DNA in 10 μlitres of 0.01 M NaCl, 0.1 M MgCl$_2$, 0.005 M Tris pH 7.5, 2 μlitres of 2 mg DNAase I/ml (Worthington) were added and the mixture was incubated for 2 hours at 25°C after which time 2 μlitres of 1 M Tris pH 7.4, 2 μlitres CaCl$_2$ and 6 μlitres of 1 mg/ml snake venom phosphodiesterase (Boehringer) ml added and the incubation continued at 45°C for 2 hours. For small
oligonucleotide fragments, such as pyrimidine tracts eluted from TLC plates, the initial DNAase I digestion was omitted. Unlabelled marker nucleosides (1 μg each) were added at the end of the incubation and the mixture injected into the column which was equilibrated with 0.03 M phosphate-buffer pH 4.65 and the column run at a pressure of 15,000 lbs/sq in. The elution of the markers was detected on the instruments UV optical systems at 254 nm and recorded on a Hewlet-Packard 3380 A Integrator. The elution of radioactive samples was followed by collecting 30 second fractions off the column (about 0.5 ml) and counting them in 'Instagel'.

10.9 DNA Reassociation and Thermal Chromatography

DNA reassociation was performed by driving $^{32}$P and/or $^3$H labelled nick translated probes with a large excess of driver DNA. In general, the ratio of $^3$H probe to driver DNA was 1:100 and the ratio of $^{32}$P probe to driver DNA was 1:1000. Single stranded and double stranded DNA were separated on hydroxylapatite (Bernardi, 1965), using a series of small columns, each with a glass scintered at the bottom. The columns were heated together in a common water bath, designed to allow collection of the column eluates.

Prior to reassociation DNA samples were sonicated in 0.06 M PB at 0°C in an MSE sonicator fitted with a microprobe at maximum power in two ten second bursts separated by an interval of five minutes. The average size of single stranded DNA prepared in this fashion is between 600-100 nucleotides (Mitchell et al., 1979). Before sonication of $^{32}$P or $^3$H labelled nick translated DNA unlabelled Micrococcus lysodiioticus DNA was added to a concentration of 2.5 μg/ml.
DNA was denatured by heating to 100°C for five minutes in 0.12 M PB in a 1.5 ml 'Eppendorf' centrifuge tube and subsequently cooled rapidly to 0°C. Driver and probe were mixed to give a final driver concentration of between 0.1 and 4 µg/ml for purified satellite DNA and between 10 and 400 µg/ml for total human DNA. The annealing mixture was placed under paraffin oil, incubated at 60°C, and portions removed at various times which were diluted several fold into 1 ml ice cold 0.06M PB. The proportion of double stranded and single stranded DNA was determined for each time point at 60°C by adding the diluted sample to the column containing fresh (0.1 g) hydroxylapatite (Biorad) that had been washed several times at 60°C in 0.06M PB. After equilibration the column was opened by removing the clip attached to the piece of tubing leading from the bottom of the column and the PB allowed to drip through under gravity. The hydroxylapatite was subsequently washed with 2 x 2 ml of 0.06M PB. Single stranded DNA was eluted by washing the hydroxylapatite with 3 x 1 ml of 0.12M PB at 60°C. The eluate from the column was collected in a scintillation vial. Double stranded DNA was subsequently eluted by washing the column with 3 x 0.4 ml of 0.3M PB into a scintillation vial. This eluate was diluted to 0.12M PB by the addition of 1.8 mls of dH₂O. Samples were counted in the scintillation counter after the addition of 7 ml of 'Instagel'.

Determination of foldback sequences (zero-time reassociation) and the extent of self reassociation during annealing of the labelled probes were determined in the absence of driver DNA.

The % reassociation at each time point was calculated after correction for the small amount of sonicated native DNA (f) that was eluted in 0.12M PB and the fraction of foldback DNA (F). Foldback DNA was considered an artefact of the nick translation procedure. The relationship of these various quantities is illustrated in the
Corrected double stranded counts, \( D = d - (F((s+d)-(s+d)f)) \)
Corrected single stranded counts \( S = s - (s+d)f \)
% reassociation \( = \frac{D}{D + S} \times 100 \)

\( s = \text{cpm eluted in 0.12M PB} \)
\( d = \text{cpm eluted in 0.3M PB} \)
\( f = \text{fraction of native DNA eluted in 0.12M PB} \)
\( F = \text{fraction of foldback DNA} \)

DNA melting experiments were carried out on hydroxylapatite using water-jacketed columns as described by McCallum and Walker (1967). Melting profiles of renatured DNA were investigated using \( ^{32}\text{P} \) of \( ^{3}\text{H} \) nick-translated probes from which foldback sequences had been removed by an initial hydroxylapatite separation, using a heated centrifuge (Flamm et al., 1969b). These single-stranded probes were renatured with an excess of homologous or heterologous driver DNA at 60°C in 0.12 M PB. After renaturation the reassociated duplex was bound to HAP at 60°C and its melting profile determined by elution in 0.12 M PB at each increment of temperature.

10.10 Chromosome Preparation and In Situ Hybridisation

Human metaphase chromosomes were prepared from peripheral blood lymphocytes. 0.8 ml of fresh blood was added to 10 ml of blood culture medium (Hams F10 supplemented with 0.8% PHA, 100 µg/ml streptomycin sulphate) in a McCartney tube and incubated at 37°C. After 2 days colchicine was added to a concentration of \( 10^{-5} \)M and the culture
incubated for a further 2 hours. The cells were spun down at 3000 rev/min for five minutes in a bench centrifuge and swollen for 10 minutes at R.T. in 4 ml of 0.075M KCl. The KCl was removed, the preparation fixed three times in 3:1 methanol:acetic acid (Sumner et al., 1970) and the cells finally resuspended in about 0.25 mls of fixative. One drop of this cell suspension was spread onto alcohol washed microscope slides.

The chromosome spreads were hybridised to $^3$H nick translated probes in the following manner. Chromosomes were denatured in 0.2 N HCl for 30 minutes at R.T. and dehydrated through an alcohol series. Nick translated DNA (Section 10.6) was ethanol precipitated and re-dissolved in a suitable volume of dH$_2$O. The DNA was denatured by heating it at 100°C for five minutes and then rapidly cooled on ice. The final volume was measured and 20 x SSC (pH 7.4) added to a final concentration of 2 x SSC. 15 ulitres of this solution was spotted onto a 20 x 40 mm coverslip and subsequently sealed on top of the chromosome spread with 'Holdtite' rubber glue. Usually about 5 x $10^5$ dpm of $^3$H DNA (approx 0.15 µg) were used per chromosome spread. The slides were incubated at 65°C for four hours after which time the coverslip was removed and the slides washed in 200 mls of 2 x SSC pH 7.4 at 65°C for two hours and then in 4 litres of 2 x SSC at 4°C overnight. The slides were dehydrated through an alcohol series, dipped in Ilford L4 liquid emulsion and exposed in light tight boxes at 4°C for an appropriate time.

Chromosome banding and silver grains were simultaneously observed by a procedure described by Lawrie and Gosden (1980). After development in Kodak D19 the slides were stained for seven minutes in 0.25% quinacrine hydrochloride made up in double distilled deionised H$_2$O pH 4.5 (adjusted with 0.1 N HCl) at room temperature, rinsed momentarily and washed for two minutes in 2 litres of the same pH 4.5 water. The
slides were mounted in 2 M sucrose made up in DDDH₂O pH 4.5 and a 20 mm x 40 mm coverslip overlaid. A Leitz ortholux microscope with both visible and UV light sources was used for photomicrography. Spreads were photographed with the UV alone and then with the 70X objective. The visible light was then switched on and its intensity increased until the autoradiographic grains were clear, when a second photograph was taken. Kodak Panatomic X was used for both fluorescence and combined fluorescence/visible-light photomicrography.
REFERENCES


ABBREVIATIONS

DNA  deoxyribonucleic acid
RNA  ribonucleic acid
dATP deoxyadenosine triphosphate
dGTP deoxyguanosine triphosphate
dTTP deoxythymidine triphosphate
dCTP deoxycytosine triphosphate
RNAase ribonuclease
RF replicative form
5MeC 5 methyl cytosine
polyA polyriboadenylic acid
PHA phytohaemagglutinin
EDTA diaminoethanetetraacetic acid
Tris tris (hydroxymethyl) aminomethane
TEMED NNN'N' tetramethyl-1,2-diaminoethane
PB phosphate buffer, pH 6.8
SSC .15M NaCl, .015M sodium citrate
SDS sodium dodecyl sulphate
gram
milligram
microgram
litre
cubic centimetre
microlitre
molar
millimolar
micromolar
counts per minute
disintegrations per minute
Curie
molecular weight
base-pairs
kilo base-pairs
absorption at 260 nanometers
thin layer chromatography
melting temperature
rev/min revolutions per minute