Cloning of human satellite III DNA: different components are on different chromosomes

Howard J.Cooke

MRC Mammalian Genome Unit, King's Buildings, West Mains Road, Edinburgh EH9 3JT, and

John Hindley

Department of Biochemistry, University of Bristol, Medical School, University Walk, Bristol
BS8 1TD, UK

Received 30 May 1979

ABSTRACT

Two fragments cloned from purified human satellite III DNA do not cross-react with each other. One fragment, for which a partial sequence is reported, hybridises to satellite II as well as III and is shown to originate on chromosome 1. The other cloned fragment originates from the Y chromosome. This fragment has undergone considerable changes in size when cloned in λ gt WES λ B. Human satellite III is shown to consist of a number of non-cross-reacting sequences which nevertheless are related by the presence of closely spaced *Hin* F1 sites.

INTRODUCTION

Highly repeated DNAs which can be separated from the bulk DNAs by density gradient centrifugation have been known for many years. Their structure was regarded as simple and many theories have been proposed to explain their function, but as yet no satisfactory answer to this problem has been provided. Restriction enzyme analysis of these satellite DNAs has complicated the simple picture of a large number of copies of closely related sequences (1, 2) distributed relatively evenly throughout the karyotypes (3, 4). For instance, in mouse satellite there have been shown to be subsets of the population of repeats which can be recognised by digestion with restriction endonucleases (5).

The human genome has been reported to contain a number of repetitive DNA components which can be isolated by density gradient centrifugation. Four major components (satellites I, II, III and IV) have been described (6 - 8) which constitute in total about 2 - 5% of the genome and are AT rich compared to main band DNA. *In situ* hybridisation has been used to identify those chromosomes which contain DNA homologous to these sequences (9). In general, all satellites hybridise to the same regions of the same chromosome suggesting a uniform distribution of sequences. This is in contrast to the

finding that two particular highly repeated sequences, identified by comparison of restriction enzyme digests of total male and female DNA, were present only on the Y chromosome (10). By using mouse-human somatic cell hybrids containing one or a few chromosomes it was possible to show that the related sequences found on each autosome gave rise to distinct patterns of fragments when digested with restriction enzymes (11). This approach has been extended using human satellite DNAs, to probe the human sequences in somatic cell hybrids (12), and has shown that different restriction repeat length classes of human satellite III DNA are present on different autosomes.

These observations raise the question of the structure of human satellite III. In the case of mouse satellite, restriction enzyme analysis can be used to demonstrate that virtually all of the DNA is composed of repeats of 240 base pairs (1) but human satellite III contains sequences with a number of different repeat lengths revealed by digestion with different restriction enzymes (12). Are these fragments related in DNA sequence although differing in repeat length, or do the different repeat lengths represent unrelated sequences? We have attempted to answer this question and that of the chromosomal distribution of different sequences by the use of recombinant DNA techniques to create clones of bacteriophage λ containing sequences from human satellite III.

These cloned fragments of repeated sequence human DNA can then be compared directly by restriction site mapping, hybridisation or sequencing. The components of total human DNA related to the cloned fragments can be analysed and the chromosomal origin and organisation determined. The advantages of this approach are that the cloned DNA sequences should be easily available, completely homogeneous and, compared to a satellite DNA preparation, of relatively low complexity.

MATERIALS AND METHODS

DNA Preparation

Human satellite III DNA (a gift from A. Mitchell) was prepared by centrifugation in $\text{CsSO}_4/\text{AgClO}_4$ density gradients as described by Beauchamp *et al.* (12). The criteria of purity were formation of a single homogeneous component with a density of 1.697 g.cm^{-3} in neutral CsCl analytical density gradients and assessment of the fragment pattern given on digestion with *Eco* RI and *Hae* III restriction endonucleases. Bacteriophage DNAs were prepared

from CsCl density gradient purified phage according to Leder (λ gt WES "Owners Manual"). DNA from somatic cell hybrids was a gift from C. Bostock.

Restriction Endonuclease Digestion

DNA's were dissolved in buffer at a concentration of 250 μ g/ml. Eco RI digestions were carried out in 0.1 M Tris HCl, pH 7.5, 0.05 M NaCl, 0.01 M MgCl₂; Hae III, Alu I and Hind I digestions in 6.6 mM NaCl, 6.6 mM MgCl₂, 6.6 mM Tris HCl, pH 7.4, 6.6 mM β mercaptoethanol, 100 μ g/ml gelatin. Completeness of digestion was assessed by sampling reaction mixes at increasing times and analysing digests by gel electrophoresis. Agarose gel electrophoresis was carried out in slab gels 0.3 cm thick and 20 cm long, of concentration between 0.5 and 3% depending on the molecular weight range of the DNA fragments to be analysed.

Acrylamide slab gels were as described by Sanger and Coulson (13) but with urea omitted and DNA loaded without formamide for analysis of double-stranded DNA. DNA was recovered from agarose gels by freezing and thawing the gel slice (14), passing it through a 21 g. syringe needle and shaking the macerated gel with dH₂O overnight at 37°C. The agarose was pelleted by centrifugation at 75,000 g for 20 mins. and the supernatant concentrated by butanol extraction. The DNA was then desalted on a 1 ml Sephadex G50 column before further use.

Nick Translation, Hybridisation and Autoradiography

DNA was nick translated according to Maniatis *et al.* (15) using α -³²P dTTP and dGTP (Radiochemical Centre, Amersham) at a specific activity of 300 Ci/mmol to give DNA with a specific activity of 0.5 to 5 x 10⁷ dpm/ μ g. DNA was transferred from agarose gels to nitrocellulose filters according to Southern (16). Filters were hybridised in Denhardt's solution (17) containing 3 x SSC for 18 hours at 65°C. 0.1 to 0.5 μ g of labelled DNA at a concentration of 10 to 50 ng/ml was used as a probe after heat denaturation. After hybridisation, filters were washed for 8 hours in 2 x SSC, 0.5% SDS at 65°C followed by drying at room temperature. Autoradiography was carried out according to Laskey and Mills (18) using Fuji RX film and Hanimex Mach 2 intensifying screens at -70°C.

Cloning of DNA

Construction and growth of recombinants between λ gt WES λ .B (19) and human satellite III was carried out in C3 conditions as approved by the Genetic Manipulation Advisory Group. T₄ ligase was purchased from Miles

Laboratories Ltd., Slough, England. Ligation, transfection and screening were carried out essentially according to Tilghman et al. (20).

DNA Sequencing

DNA sequencing was carried out according to Maxam and Gilbert (21) using 5' ³²P labelled DNA. Some gels of Maxam and Gilbert fragments were run according to Sanger and Coulson (13).

Hybrid Cell Lines

Most of the hybrid cell lines have been described by Beauchamp et al. (12) and the DNA used here was from the same preparations as used by the above workers. Lines not described by Beauchamp are PG19Me5, A23 x IIIF and CHO x 9X. PG19Me5 is derived from a mouse PG19 x human fibroblast fusion, approximately 10% of the cells contain a human chromosome 9. A23 x IIIF is a chinese hamster A23 x human fibroblast cross in which enzyme markers for human chromosomes 1, 21 and 22 can be detected. Both of these cell lines were a gift of Dr. P. Pearson. CHO x 9X is a fusion between a CHO cell and fibroblasts from a patient carrying a 9/X translocation. Frozen cell lysates of these cells were a gift from Dr. J. Mohandes.

RESULTS AND DISCUSSION

Cloning of Satellite III Fragments

DNA fragments for cloning were isolated from an Eco R1 digest of human satellite III such as that shown in Fig. 1b. The two commonest molecular species are a 3.4 kb fragment, a repeat length present on the Y chromosome (10) and a 1.77 kb fragment. A gel slice was taken which contained the 3.4 kb fragment and DNA from this slice was recovered, ligated into λ gt WES λ B DNA and transfected into E. coli LE392.

The yield of plaques from this ligation was low, 0.1 μ g of λ gt WES DNA gave 117 plaques. These were picked onto a lawn of E. coli Dp50supF (19) to form a grid of plaques. After overnight growth these plaques were screened by hybridising nick translated satellite III DNA to a nitrocellulose replica of the grid (22). Two positive plaques were detected; one plaque of each recombinant was grown in 10 ml culture. The DNA from each phage preparation was digested with Eco R1 restriction endonuclease. Fig. 1, lane c, shows one recombinant phage termed λ gt WES λ B.HS3 and lane f shows the other recombinant termed λ gt WES HS5; both recombinant phage DNAs have surprising

Figure 1

DNA digests after electrophoresis and ethidium staining, and corresponding Southern transfers after hybridisation with recombinant DNAs λ HS5 and λ HS3. (a) λ C1857 Hind III, Eco RI double digest, fragment sizes are given in kb. (b) Human satellite III, Eco RI. (c) λ HS3, Eco RI. (d) total human female DNA, Eco RI. (e) total human male DNA, Eco RI. (f) λ HS5, Eco RI. (g) total human male, Hae III. (h) total human female, Hae III. (i) human satellite III, Hae III. Gel electrophoresis was on a 1% agarose gel at 2 V/cm.

features. Since a 3.4 kb fragment of satellite DNA was purified by gel electrophoresis one would not expect the DNA fragment of 1.77 kb found in λ HS3. In an RI digest of λ HS5 there is a complex pattern of bands larger than 3.4 kb. All these DNA molecules presumably originated from a single transfected cell and it is likely that recombination was responsible for the heterogeneous nature of the inserted DNA fragments in this bacteriophage.

In order to find out more about the relationship between the DNA fragments carried by λ HS3 and λ HS5 and their relationship to sequences in total human DNA, two gels like the one shown in Fig. 1 were transferred to a nitrocellulose filter and hybridised, one with nick translated λ HS3 DNA and the other with nick translated λ HS5 DNA. Negatives of autoradiographs of these filters are shown in Fig. 1. Three major points are demonstrated by this experiment. Firstly, the inserts in the two phage do not hybridise to each

other, the 1.77 kb fragment in track c is not visible when the probe is λ HS5 and the 3.24 kb fragment in track f is not visible with λ HS3 as the probe. Only vector fragments are hybridised in tracks λ HS5 c and λ HS3 f. Secondly, λ HS3 hybridises equally to male and female DNA (Fig. 1, tracks d and e) and the principal size class of fragments which hybridise this phage is 1.77 kb - the same size as the human DNA fragment carried by this phage. The third point is that λ HS5 shows hybridisation to a male DNA repeated sequence of 3.4 kb which is not present in female DNA and so behaves in a way similar to the sequences of this size already described (10, 11). However, it is clear that the major insert size class in this phage is smaller than 3.4 kb by about 250 bases.

Structure of Recombinant DNAs

(a) λ gt WES.HS3

This recombinant retains the λ B fragment of λ gt WES. λ B which can be replaced by a new DNA fragment. Partial digestion with Eco R1 gives fragments with a molecular weight consistent with the 1.77 kb fragment of human DNA being inserted into the Eco R1 site between the left arm of λ and the λ B fragment (data not shown). Sac I digestion of this phage DNA gave three fragments, the smallest of which corresponded to the fragment from the λ B fragment of λ gt WES. λ B. Neither of the large fragments correspond in size to the fragments produced by Sac I digestion of λ gt WES. λ B, but the size of the fragments corresponds to that expected if the orientation of the λ B fragment has returned to that found in wild type λ and opposite to the orientation in λ gt WES. λ B (Fig. 2a). This was confirmed by electron microscopy of heteroduplexes formed between λ gt WES. λ B and λ gt WES. λ B HS3 which showed an inversion bubble rather than the deletion loop expected if the orientation of the λ B fragment had been maintained.

Restriction mapping of the 1.77 kb R1 fragment was carried out on sucrose gradient purified fragments. Initial experiments (data not shown) indicated that restriction endonucleases AccI, Bam H1, Bgl II, Hae III, Hind III, Hpa I, Sac I and Sal I did not cleave this sequence. Restriction endonuclease Mbo I and Alu I did cleave the 1.77 kb fragment into discrete fragments, whilst Taq I and Hin F1 produced fragments too small to be detected on agarose gels. Fig. 2b shows the fragment pattern given by digestion with Mbo I and Alu I. Mbo I digestion is never complete and gives fragments of 1.4, 1.02, 0.77 and 0.42 kb - a total of 3.51 kb, two times the length of the Eco R1 fragment. Since the sums of sizes of the

Figure 2

A) Comparison of Sac I digests of λ HS3 and λ gt WES λ .B. 0.2 μ g of Sac I digested DNA was run on a 0.5% agarose gel at 0.5 V/cm after heating at 70 $^{\circ}$ for 10 mins. to melt the cohesive ends of λ . Fragment sizes are given in kb. B) Alu I and Mbo I digestion of the cloned 1.77 kb fragment of human DNA. (a) ϕ 17A DNA after Taq I digestion, fragment sizes are 2.9, 1.2, 0.40, 0.32, 0.23 kb. (b) 1.77 kb fragment after Alu I digestion. (c) Mbo I digest of 1.77 kb fragment. Sizes given are in kb. C) Eco R1 and Alu I digests of λ HS3. The 1.77 kb Eco R1 fragment is marked.

largest and smallest bands and of the two intermediate bands are both close to 1.77 kb, and since Mbo I sites can be protected from restriction enzyme digestion by methylation of the A in the restriction site (23), the most likely explanation of this observation is that there are two Mbo I sites but that the level of methylation is such that few molecules have one site unmethylated and very few have both sites unmethylated. Densitometry of gel photographs suggests that the probability of one site being unmethylated is between 0.1 and 0.05 and the probability of two sites being unmethylated would then be 0.01 to 0.0025 if methylation occurred randomly. Less than 1% of the DNA would be found in fragments from molecules in which both restriction sites had been cleaved. By transfer of such a digest and hybridisation with human satellite III DNA as a probe, only one new band can be seen, this fragment has a size of 370 bp and suggests the restriction map shown in Fig. 3.

Figure 3

Restriction endonuclease site map of 1.77 kb fragment. Sizes marked are in kb.

Digestion of the 1.77 kb Eco R1 fragment with Alu I produces two fragments (Fig. 2b) indicating that there is a single site for this enzyme. The position of the Alu I site in relation to the Mbo sites has not been determined. Hin F1 digestion gives fragments which are too small to be resolved on a 3% agarose gel. Electrophoresis of kinase labelled Hin F1 fragments on an 8% acrylamide gel gives a pattern of at least 20 bands, all of which are smaller than 80 base pairs (Fig. 4a).

The obvious way in which a single Eco R1 fragment of 1.77 kb could be found in a recombinant phage which should contain fragments of 3.4 kb is that a dimer of the 1.77 kb fragment has been cloned. The recombinant would thus contain 3.5 kb of human DNA and would have four Eco R1 sites. The 3.5 kb human fragment could arise by partial digestion of tandem repeats of 1.77 kb. This possibility was not supported by the sizes determined for the Sac I fragments of the recombinant, nor by size determination of fragments produced by partial Eco R1 digests of this recombinant phage. To confirm the presence of a single copy of the 1.77 kb Eco R1 fragment per phage genome the DNA was digested with Eco R1 and with Alu I and the digests run on adjacent slots of a gel. Tandem repeats of the Eco R1 fragment would give a 1.77 kb fragment in both digests since all Eco R1 fragments contain a single Alu I site. Fig. 2c shows that there are no fragments of 1.77 kb in the Alu I digest of λ HS3 and that this recombinant does not contain tandemly repeated 1.77 kb Eco R1 fragments. Inverted repeats cannot be ruled out by this method but no hairpin structures were seen in electron microscope spreads of the heteroduplexes described above. This recombinant phage must contain a single Eco R1 fragment of human DNA.

From this information it is possible to draw some conclusions on the origin of λ HS3. It seems improbable that this phage could have arisen via recombination from HS5 as there is no apparent sequence homology between the human sequences present in the two phage. It is possible that the 3.4 kb

Figure 4

A) 5' ^{32}P labelled fragments produced by digestion of the 1.77 kb Eco R1 fragment from λHS3 with Hin F1. Electrophoresis on 8% acrylamide 7M urea gel. The positions of 5' ^{32}P labelled ligated synthetic oligonucleotide fragments 8 bases long are marked. DNA was loaded in 99% formamide. B) Taq I digest of human satellite III. Bands which are multiples of 340 base pairs are marked. In addition, some minor fragments are visible below 340 base pairs. B2 - Taq I digest of ϕX174 . C1 - Hin F digest of human satellite III. C2 - Taq I digest of ϕX174 DNA.

size class excised from the gel contained some dimers of the 1.77 kb fragment; these would be 3.54 kb in length and not easily resolvable from the 3.4 kb fragment. Such fragments can be detected in a partial R1 digest of human satellite III DNA (data not shown). Recombination might then generate phage with only a single copy of the 1.77 kb fragment. Lastly, the 3.4 kb size class might have been contaminated with 1.77 kb DNA for trivial reasons - this sequence is the most abundant single fragment in an Eco R1 digest of human satellite III (See Fig. 1a). This explanation is probably the most likely.

DNA Sequencing

Since λ HS5 contained a population of human DNA fragments of differing size, no attempt was made to determine any of the sequences. In the case of λ HS3 the strategy was to isolate the 1.77 kb Eco R1 fragment on sucrose gradients, label the ends of the native fragment with ^{32}P using T4 polynucleotide kinase, then to separate the ends by cleavage of the fragment with Alu I followed by gel electrophoresis to separate the two fragments which were then sequenced. In addition, an attempt was made to determine the sequence around the Alu I site by digesting the Eco R1 fragment with Alu I, kinase labelling the two fragments and separating them on an agarose gel. The separated fragments were then strand separated by neutral agarose gel electrophoresis of alkali denatured fragments and the strands isolated. A sequence at the Alu I end of the 0.71 kb fragment was determined in this way. Figure 5 shows two sequencing gels and Figure 6 the total sequence determined. Three blocks of sequence have been determined, one from each end of the fragment and one from the centre. The total sequence length is 300 bases. There are numerous repetitive elements in the sequence, the most striking of which is the sequence ATTCGATTC which occurs four times as such with the longer related sequences ATCCATTCGATTC and ATCCATTCGATTCATTC occurring once each. The sequence ATTCATT occurs twice and ATTC six times. The tetranucleotide ATTC occurs a total of twenty-four times and there are a further nine tetranucleotides ATTN of which six end in T. Nearly half of the sequence can be derived from ATTC by changes in the last base, but most of the remaining half of the sequence could not be derived by single base changes in the other positions of this tetranucleotide. Such an element is the sequence GATGATGAT which occurs twice.

The sequence ATTCGATTC contains overlapping Hin F1 and Taq I sites and so the sequence confirms the observation that digestion of the 1.77 kb fragment with these enzymes produces many small fragments. The spacing of this sequence and the related sequences is not regular in the regions for which the overall sequence has been determined and this is reflected in the fragment pattern shown in Fig. 4a, in which a large range of fragment sizes is present. In fact, this lack of regularity must apply throughout the length of the 1.77 kb fragment. The overall pattern is of regions which can be derived from the tetranucleotide ATTC interspersed with other regions of varying length which contain a higher proportion of GC base pairs but which have limited similarity with each other.

Figure 5

DNA sequencing gel. 0.04 x 20 x 40 cm 8% acrylamide 7M urea gel. Cleavage was according to (21). The first four lanes are from the 1.06 kb Alu I, Eco R1 fragment, the second four lanes are from the Eco R1 end of 0.71 kb Alu I, Eco R1 fragment and the third from the Alu I end of this fragment.

This sequence is different from that found in a human and African green monkey Hind III restriction fragment (24) which showed no repeats in a 170 bp fragment. The presence of a conserved segment interspersed with less conserved segments of different lengths means that this repeated sequence cannot be classed either as a very simple or as a complex repeated sequence, but has elements which fall into both categories.

```

5' AATTCATTTGAAGACAATTCATTCAATACCAATTGATGATGTTATTTTCATTC
   CATTGATGATGATTACATTCCATTTTCATCATAATTCATT*GATTC*CACTCGAGA
   TTCCATT*CGATTCATTCAA.....
   *CGAATGAATGAGTCCATCCATTTCAATTCATGATAATTCATT*CGTTTCAATT*CG
   ATGGTGTTCATT*CGATT.....
   TTCATT*CGATTCATTGATGATGATTCATGCGCGATTCAATAGATGATGACCCCTT
   TCATTTCATTCAATGGAGGATTCATT*CGGTTCCAT..... 3'

```

Figure 6

DNA sequences determined for the 1.77 kb fragment of λ HS3 *Hin* F1 and *Taq* I sites are underlined, CG dinucleotides are marked by an asterisk. The 5' end shown is that nearest the *Alu* I site.

(b) λ gt WES HS5

This recombinant phage has inserts which are heterogeneous in molecular weight, the most common insert size is 3.2 kb, smaller than the original fragment inserted by 200 bases; the next commonest fragment is 4.1 kb and there are fragment of 1.7 and 1.3 kb. Since this phage is clearly unstable, it is necessary to determine the origin of the DNA enclosed by *Eco* R1 sites. Fig. 8a1 shows an *Eco* R1 digest of λ HS5 DNA which was transferred to a nitrocellulose filter and hybridised separately with nick translated human satellite III DNA λ C1857 and *E. coli* K12 DNA. The only probe which hybridised to the fragments below the size of the known λ fragments was human satellite III which gave the pattern shown in Fig. 7a3 and 5. A range of fragments are present which contain sequences homologous to human satellite III.

Repeated sequence DNAs have been reported to be unstable when cloned in plasmids (25) although in the case of the *Drosophila* satellites the sequences cloned had a very short repeat and contained few substitutions, a situation which might well be more favorable for a recombinational mechanism leading to DNA deletion. However, in this case not only are deletions occurring but some phage particles must also contain larger DNA fragments than those originally inserted *in vitro* into the bacteriophage. The related 3.4 kb *Hae* III fragment found on the Y chromosome showed a clear 0.6 kb repeat length when digested with *Mbo* I (11). Restriction enzymes *Mbo* I and *Hpa* II digest the inserts in λ HS5 into many fragments of different size (data not shown). Digestion

Figure 7

Analysis of λ HS5. A1 and 2, Ethidium bromide stained gel of Eco R1 digests of λ HS5 and C1857 respectively. A3 and 4, Southern transfer of A1 and 2, hybridised with human satellite III. A5 is a shorter exposure of A3. Sizes of human fragments are given in kb. B1 and 2, Ethidium bromide stained gel of total human DNA digested with Hin F1 and ϕ X174 DNA digested with Taq I. B3 and 4, Southern transfer of B1 and 2 after hybridisation with nick translated λ HS5 and ϕ X174 DNAs. B5 and 6, Ethidium bromide stained gel of λ HS5 DNA digested with Hin F1, ϕ 174 DNA digested with Taq I. B7 and 8, Southern transfer of 5 and 6 hybridised with nick translated human satellite III and ϕ X174 DNA.

of λ HS5 with Hin F1 followed by transfer and hybridisation with satellite III shows only one fragment containing human sequences of 220 bp (Fig. 7b7). When a Hin F1 digest of total DNA is transferred to millipore and hybridised with λ HS5, two bands are visible, one of 220 bp and one of 170 bp. This experiment shows that the cloned human sequence in λ HS5 is internally repetitive and this provides an obvious explanation for the heterogeneity of the insert. Recombination between DNA molecules which were out of register in terms of the Eco R1 sites, but in which the Hin F sites were aligned, would generate inserts of varying size (26). In this case the size differences would be multiples of 220 base pairs. The initial fragment taken for cloning was 3.41 kb = 15.5 x 220. The major size classes of human sequence after cloning are 1.7, 3.2, 4.1, 4.8 and 5.6 kb; the first

three sizes can be fitted into such a scheme, e.g. $3.4/2 = 1.7$, $3.4 - .22 = 3.2$, $3.4 + 3(.22) = 4.1$, but this may be no more than suggestive.

Taken together, the experiments shown in Fig. 7b demonstrate that the human sequences present in HS5 have maintained at least part of the restriction repeat length found in human DNA, even though some process has resulted in a change in size of the sequence once it is inserted into the bacteriophage. Fragments of this size are not found in a detectable amount in Eco RI digests of human DNA; the events here must occur because of a property of the recombinant phage/bacteria system rather than simply a property of the human DNA alone.

Restriction Analysis of Human Satellite III

Digests of total human satellite III are shown in Fig. 4b. Taq I gives a prominent band of 340 bases with multiples present at 680 and 1020 bases, Hin F1 digestion also gives a band of 340 base pairs but in this case the major digestion product is a band between 75 and 85 bases which constitutes a very large part of the DNA. The range of small fragments produced by Hin F1 or Taq I digestion of the 1.77 kb in λ HS3 fragment would not be visible on this gel. They must be present in human DNA since when a Hin F1 or Taq I digest of human DNA is transferred to millipore and hybridised with HS3 DNA no detectable hybrids are formed (data not shown). Since the majority of these DNA fragments are below 50 bp they do not bind efficiently to nitro-cellulose and hence cannot be hybridised.

The significance of these observations is that they show that this satellite DNA has a relatively regular distribution of some restriction sites although others may only be present on a proportion of these sequences. For instance, the digests of satellite III in Fig. 1 show large amounts of high molecular weight DNA and only a small proportion of the total digested into fragments resolvable on a 1% gel. In this way human satellite III appears more like mouse satellite in having some overall similarities between many different sets of sequences (5).

Arrangement of Sequences Homologous to λ HS3 and λ HS5 in Human DNA

When restriction endonuclease digests of total human DNAs or satellites are transferred to millipore and hybridised with one of the cloned human DNA fragments described above, each cloned fragment hybridised to a different set of bands. By comparing these band patterns and the ethidium bromide stained DNA a number of conclusions can be made about the relationships between the

different repeated sequences. Firstly, the prominent ladder of bands of length $n(170 \text{ bp})$ found in a Hae III digest of human DNA (Fig. 1g and h) and in human satellite III does not hybridise either cloned sequence, nor do the Eco RI fragments (of 370 and 720 bp) and which are also found in human satellite III. Secondly, each recombinant phage hybridises not only to fragments of the same size as the human DNA fragment it contains (or slightly larger in the case of the 3.2 kb fragment in λ HS5) but also to a number of other fragments which do not fall into a simple pattern of multiples of a given repeat length as in the case of mouse and many other satellites (1). Some fragment sizes hybridise both of the recombinant phage probes, e.g. the 3.4 kb band is present in Eco RI and Hae III digests of satellite III and male DNA when probed with λ HS5 and is present in both male and female Eco RI digests but only in the male Hae III digest when probed with λ HS3. This demonstrates that in some cases a band of apparently uniform size may contain two non-cross reacting DNA sequences. Further, there must be a minimum of three different sequence classes in human satellite III, that reassociate with λ HS3, those that reassociate with λ HS5, and third and largest, those that do not reassociate with either of the cloned sequences.

Hybridisation of Recombinants to Human Satellite Restriction Digests

The human DNA fragments cloned in these bacteriophage were produced from an Eco RI digest of human satellite III; it is of interest to know if these sequences are present in other human satellite DNAs. To test this possibility human satellites I, II and III were digested with Eco RI and with Hae III and the fragments transferred to a nitrocellulose filter after gel electrophoresis. The filter was hybridised with nick translated λ HS5 DNA and autoradiographed to give the pattern shown in the left section of Fig. 8. The filter was then washed in 5mM EDTA, pH 8, to remove hybrids and, in addition, allowed to decay for 30 days. Autoradiographs at this stage showed that only traces of the λ DNA could be detected with an overnight exposure. The filter was then rehybridised with λ HS3, washed and autoradiographed overnight to give the pattern shown in the right section of Fig. 8.

As expected, in Fig. 8 most hybridisation of λ HS5 is to satellite III although both satellite II and I show some hybridisation to a 3.4 kb size class and to DNA which has not been cleaved by Eco RI or Hae III. When this filter is hybridised with λ HS3 a different pattern emerges; satellites II and III now give a very similar pattern both qualitatively and quantitatively. Satellite I shows almost no hybridisation. If this was due simply to cross-

Figure 8

Hybridisation of cloned sequences to human satellites. Digests of satellites I, II and III with Eco R1 and with Hae III were transferred to millipore and hybridised with λ HS3 and λ HS5. Numbers above the gel tracks refer to the satellite DNA; the unnumbered track is a λ C1857 Hind III/Eco R1 digest.

contamination of the two satellite DNAs, one would expect equal hybridisation with the λ HS5 probe. Since this is not the case there must be a component common to both satellites of which λ HS3 is representative. Since these satellite DNAs have different buoyant densities and thermal stabilities (8) there must also be components which are not common to both. The sequence found in λ HS5 is an example of a component found only in satellite III. Buoyant density differences between satellites which contain common unique components suggests that these two types of sequence may be interspersed.

Chromosomal Origin of DNA Contained by Recombinants

A 3.4 kb repeated sequence found on the Y chromosome has already been described (10, 11) and this was the fragment inserted in vitro into the bacteriophage λ DNA. Furthermore, when this probe is hybridised to male and

female DNA digested with Hae III or Eco RI, a 3.4 kb band is found in males but not in females indicating that this repeated sequence is found on the Y chromosome. However, the pattern of Mbo sites differs from that described (11) due either to the effect of sequence changes in the cloned DNA or to the original repeat cloned being atypical of those previously described.

The sequence present in λ HS3 does not behave in this way; there are at most only minor differences between the pattern of hybridisation found with male and female DNA. To identify the chromosomal origin of the cloned sequence, use has been made of somatic cell hybrids carrying known human chromosomes. This approach has been used previously to determine which chromosomes carried sequences related to the 3.4 kb repeat found on the human Y chromosome (11) and to look at the chromosomal distribution of restriction fragments of human satellite III DNA (12). In both studies it was found that the sequences on different chromosomes have a different pattern of restriction sites and it therefore seemed possible that the sequence cloned in λ HS3 originated on a single chromosome. DNA prepared from hybrid cell lines was digested with Eco RI and transferred to a nitrocellulose filter which was then hybridised with nick translated human satellite III, λ HS5 or λ HS3 DNAs.

The autoradiograms shown in Fig. 9 demonstrate that of the chromosomes tested in these hybrids only chromosome 1 carries a 1.77 kb repeat which is homologous to that found in λ HS3. The pattern of fragments on chromosome 1 is very much simpler than that found in the complete genome, so other chromosomes must also carry sequences related to this but with different restriction site spacings. None of the chromosomes tested here carries a significant amount of the related sequences. Miller *et al.* (27) have shown that a sequence which contains 5 methyl-cytosine is present at the c-band region of chromosome 1. This is consistent with the observation that the 1.77 kb sequence is present on this chromosome and contains a relatively high level of the dinucleotide CG. As expected from its Y chromosomal origin described earlier, λ HS5 does not hybridise significantly to a 3.4 kb band in these hybrids. The pattern of hybridisation is similar to that reported previously on autosomes (11) when a 3.4 kb Y DNA probe is used, chromosomes 7 and 15 show most hybridisation. The patterns given with total satellite III are identical to those reported previously (12). When compared to the patterns given by the cloned fragments there are more bands for a given chromosome and more chromosomes contain sequences which are present in satellite III. This confirms the conclusion above that there are sequences in satellite III which do not hybridise to either cloned fragment and shows

Figure 9

Hybridisation to somatic cell hybrids. Three identical gels were transferred to nitrocellulose filters and hybridised with the nick translated λ HS3, λ HS5 and total human satellite III DNAs. The cell lines in each track and the principal human chromosomes are: (a) A23, 1/17 translocation including the centromere of chromosome 1. (b) C121, chromosome 7. (c) Horl 9 - 11, chromosome 11. (d) Horl 9 - 15, chromosome 15. (e) PG19 - 22, chromosome 22. (f) PG19 Me5, chromosome 9. (g) A3 x I1IF, chromosomes 1, 22 and 21 and others. (h) CHO x 9X, a chromosome 9-X translocation. (i) Horl 9X, chromosome X. (j) Total human male. (k) PG19, mouse parent line for (b), (c), (d), (e), (f) and (i). (l) A23, chinese hamster line for comparison with (a), (g) and (h). Three microgrammes of each DNA was loaded on each sample slot after Eco R1 digestion.

that at least two of these components have a very limited chromosomal distribution.

DISCUSSION

Horst and Zachau (5) have shown that the set of mouse satellite DNA sequences contain a number of subsets defined as DNA sequences which include the recognition site for a particular restriction endonuclease. Human satellite III is shown here to contain not only this type of diversity but a more surprising form in which fragments generated by the same restriction enzyme represent not only different restriction repeat length classes but also sequences which are sufficiently distantly related not to cross-hybridise

with each other. At least three and probably many more of these subsets of sequences exist in human satellite III. In this case is there any basis for considering satellite III to be a distinct component rather than simply a collection of sequences which happen to co-isolate in density gradients? The feature which all of these sequences share is a high frequency of *Hin* F1 sites, suggesting that the components are indeed related in sequence. The nature of the sequence determined for the 1.77 kb fragment suggests a possible explanation. This DNA consists of a number of very similar sequences related to the nine base pair element $\text{ATTC}\overset{\text{C}}{\underset{\text{G}}{\text{A}}}\text{TTC}$ interspersed with other sequences which differ from each other considerably. If this basic nine base pair element is retained in all components of satellite III but the rest of the sequences show as much variation as is seen in the 1.77 kb fragment, then the *Hin* F1 sites might be retained in a set of non-cross-reassociating but related sequences.

Manuelidis (28) makes the distinction between simple and complex human satellite sequences and concludes that the sequences represented in restriction bands are 'complex' and DNA which is isolated as a satellite peak is 'simple' in sequence terms. The work reported here shows that a high molecular weight satellite can appear to have different levels of complexity revealed by different methods of analysis. Human satellite III appears to consist of a set of sequences with frequent *Hin* F1 sites in common but in which other restriction sites may exist only on a few of the total number of molecules. Two cloned representatives of fragments having *Eco* R1 sites do not cross-hybridise and appear to be largely confined in origin to a single chromosome although related sequences are present elsewhere in the genome.

A number of mechanisms of satellite DNA evolution have been proposed (see 29 for review) most of which would predict sequence homology throughout a repeat unit of constant length. Within the 1.77 kb fragment there are clearly repeated elements but these are spaced at irregular intervals and separated by sequences which show few similarities with each other. The replicative loop hypothesis as modified by Musich *et al.* (30) has a requirement for sequence homology occurring at a spacing such that the repeated sequence appears in the interstices of the nucleosome; it places no constraint on other sequences. This mechanism might be more consistent with the occurrence in the 1.77 kb sequence of short repeated sequences separated by non-repeated sequences, but it would seem to require a constant spacing of the conserved sequence $\text{ATTC}\overset{\text{C}}{\underset{\text{G}}{\text{A}}}\text{TTC}$ which is not found at the level

of the sequence length determined here. A mechanism such as unequal crossing over (26) might be expected to make a sequence such as the one described here homogeneous if the register of crossing over were of the same size as this conserved DNA unit. Such a sequence would contain infrequent R1 sites; the observation of regularly spaced R1 sites suggests that if unequal cross-over occurs then the register must be larger than the short repeats seen here, or else the recombinational events must be infrequent. Neither of these mechanisms seem compatible with the type of sequence found in the 1.77 kb fragment. An analysis of the sequence variation between copies of the 1.77 kb fragment might be necessary to understand the evolution of this DNA.

The data presented here show that within a satellite DNA there can be many different components which, as well as differing in sequence, have distinct chromosomal localisations, a contrast to the previously described concept of widespread distribution of related sequences throughout the karyotype (3, 4, 9). These satellite DNA sequences seem to be capable of identifying a particular chromosome in a way similar to that found for Drosophila satellites (31). The hypothesis that these sequences may thus identify chromosomes for some cellular process such as meiotic pairing becomes more tenable in the light of these new observations.

If individual chromosomes have repeated sequences which have little or no homology to repeated sequences on other chromosomes and/or restriction patterns which are unique to individual chromosomes, then it is feasible to use these sequences as molecular probes for individual chromosomes. The cloned sequence in λ HS5 has been used to investigate DNA from a number of individuals having suspected Y-autosome translocations (32). This is the simplest case as the normal homologue of the chromosome to be detected is not present in females. However, as more repeated sequence clones are characterised some will be useful as hybridisation probes for chromosomes or fragments of chromosomes.

ACKNOWLEDGEMENTS

This work was supported by the Medical Research Council. We should like to thank C.J. Bostock and E.M. Southern for their advice and critical readings of the manuscript.

REFERENCES

1. Southern, E.M. (1975) *J.Mol.Biol.* 94, 51.
2. Gall, J.G. and Atherton, D.D (1974) *J.Mol.Biol.* 85, 633.
3. Pardue, M.L. and Gall, J.G. (1970) *Science* 168, 1356-1368.

4. Jones, K.W. (1970) *Nature* 225, 912-915.
5. Horz, W. and Zachau, H.G. (1977) *Europ.J.Biochem.* 73, 383-392.
6. Corneo, G., Ginelli, E. and Polli, E. (1967) *J.Mol.Biol.* 23, 619-622.
7. Corneo, G., Ginelli, E. and Polli, E. (1970) *J.Mol.Biol.* 48, 319-327.
8. Corneo, G., Ginelli, E. and Polli, E. (1971) *Biochem.Biophys.Acta* 247, 528-534.
9. Gosden, J.R., Mitchell, A.R., Buckland, R.A., Clayton, R.P. and Evans, H.J. (1975) *Exp. Cell Res.* 92, 148-158.
10. Cooke, H.J. (1976) *Nature* 262, 182-186.
11. Cooke, H.J. and McKay, R.D.G. (1978) *Cell* 13, 453-460.
12. Beauchamp, R.S., Mitchell, A.R., Buckland, R.A. and Bostock, C.J. (1979) *Chromosoma*, in press.
13. Sanger, F. and Coulson, A.R. (1978) *F.E.B.S. Letters* 87, 107-110.
14. Subramanian, K.N., Pan, J., Zain, S. and Weissman, S.M. (1974) *Nucleic Acid Res.* 1, 727-733.
15. Maniatis, T., Jeffrey, A. and Kleid, D. (1975) *Proc.Nat.Acad.Sci., U.S.A.* 72, 1184-1188.
16. Southern, E.M. (1975) *J.Mol.Biol.* 98, 503-517.
17. Denhardt, D.T. (1966) *Biochem.Biophys.Res.Comm.* 23, 641-646.
18. Laskey, R.A. and Mills, A.D. (1977) *F.E.B.S. Letters* 83, 314-318.
19. Leder, P., Tiemeier, D.C. and Enquist, L.N. (1977) *Science* 196, 175-177.
20. Tilghman, S.M., Tiemeier, D.C., Polsky, F., Edgell, M.H., Seidman, J.G., Leder, A., Enquist, L.N., Norman, A. and Leder P. (1977) *Proc.Natl.Acad. Sci., U.S.A.* 74, 4406-4410.
21. Maxam, A.M. and Gilbert, W. (1977) *Proc.Natl.Acad.Sci., U.S.A.* 74, 560-564.
22. Benton, W.D. and Davis, R.W. (1977) *Science* 196, 180-182.
23. Roberts, R.J. (1976) *C.R.C. Crit. Rev. Biochem.* 123-164.
24. Rosenberg, H., Singer, M. and Rosenberg, M. (1978) *Science* 200, 394-402.
25. Brutlag, D.L., Fry, K., Nelson, T. and Huey, P. (1977) *Cell* 10, 509.
26. Smith, G.P. (1976) *Science* 191, 528-535.
27. Miller, O.J., Schnedl, W., Allen, J. and Erlanger, B.F. (1974) *Nature* 251, 636-637.
28. Manuelidis, L. (1978) *Chromosoma* 66, 1-21.
29. Southern, E.M. (1974) in *MTP International Review of Science, Biochemistry Series One, Vol. 6*, Burton, K. ed. pp. 101-141, Butterworths, London.
30. Musich, P.R., Brown, F.L. and Maio, J.J. (1978) *Cold Spring Harbor Symp. Quant. Biol.* 42, 1147-1160.
31. Peacock, W.J., Lohe, A.R., Gerlach, W.L., Dunsmuir, P., Dennis, E.S. and Appels, R. (1977) *Cold Spring Harbor Symp. Quant. Biol.* 42, 1121-1135.
32. Cooke, H.J. and Noel, B. (1979) In preparation.