INFORMATION TO USERS

The most advanced technology has been used to photograph and reproduce this manuscript from the microfilm master. UMI films the original text directly from the copy submitted. Thus, some dissertation copies are in typewriter face, while others may be from a computer printer.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyrighted material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each oversize page is available as one exposure on a standard 35 mm slide or as a 17″ × 23″ black and white photographic print for an additional charge.

Photographs included in the original manuscript have been reproduced xerographically in this copy. 35 mm slides or 6″ × 9″ black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Accessing the World's Information since 1938

300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA

Order Number 8821103

Isolation and characterization of a minor domain of human Satellite II DNA

Macoska, Jill Anne, Ph.D.

City University of New York, 1988

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark __✓__.

1. Glossy photographs or pages __✓__
2. Colored illustrations, paper or print _____
3. Photographs with dark background __✓__
4. Illustrations are poor copy _____
5. Pages with black marks, not original copy _____
6. Print shows through as there is text on both sides of page _____
7. Indistinct, broken or small print on several pages __✓__
8. Print exceeds margin requirements _____
9. Tightly bound copy with print lost in spine _____
10. Computer printout pages with indistinct print _____
11. Page(s) _____ lacking when material received, and not available from school or author.
12. Page(s) _____ seem to be missing in numbering only as text follows.
13. Two pages numbered _____. Text follows.
14. Curling and wrinkled pages _____
15. Dissertation contains pages with print at a slant, filmed as received _____
16. Other _____

_____

_____

ISOLATION AND CHARACTERIZATION OF A MINOR DOMAIN OF HUMAN SATELLITE II DNA

by

JILL A. MACOSKA

A dissertation submitted to the Graduate Faculty in Biochemistry in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

1988

Thia manuscript has been read and accepted for the Graduate Faculty in Biochemistry in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

4/25/88
Date

Chair of Examining Committee

4/25/88
Date

Executive Officer

Rivka Rudner

Harvey Ozer

Diane Robin

Reba Goodman

Thomas Schmidt-Glenwinkel
Supervisory Committee

The City University of New York

ABSTRACT

ISOLATION AND CHARACTERIZATION OF A
MINOR DOMAIN OF HUMAN SATELLITE II DNA

by

Jill A. Macoska

Advisor: Professor Ann Henderson

Specific sequence heterogeneities and a novel minor chromosome 7-specific domain were identified within human Satellite II DNA. Satellite II DNA was purified using cesium gradient ultracentrifugation and propagated as recombinant DNAs. Twelve of 106 cloned DNA fragments were randomly chosen for study. Three contained unique or not detectably repetitive sequences. One contained alphoid DNA. Four contained sequences homologous to L1 KpnI LINES element DNA and one of these, 43 DNA, localized specifically to the centromeres of chromosome 4 and the X chromosome. The identification of unique, alphoid and L1 sequences defined nonsatellite-associated sequence heterogeneity within Satellite II DNA. The remaining four cloned DNAs studied contained

Satellite II sequences associated with Alu SINES element DNA. DNA base sequencing showed that the 5' terminal 200 basepair (bp) region of one of these, 33 DNA, contained a tandem array of six degenerate 26-nucleotide Satellite II repeat units that conserved an average 56% of the consensus sequence residues. This region was subcloned as XL3 DNA and comprised basepairs 1-750 of 33 DNA. Southern blot analysis showed that a 1400 basepair sequence homologous to XL3 DNA occurred 10-100X in genomic DNA. Moreover, XL3 DNA revealed a 3.4 kilobasepair (kb) HaeIII fragment in human male placental DNA and hybridized in situ to the pericentromeric region of human chromosome 7. This indicated that XL3 DNA was homologous to the non Y-specific (NYS) 3.4 kb HaeIII family of repeated sequences. Together, these experiments showed that XL3 DNA defined a novel minor domain of human Satellite II that occurred in 10-100 copies in genomic DNA, was associated with NYS 3.4 kb HaeIII repeated sequences and localized to the pericentromeric region of human chromosome 7.

ACKNOWLEDGEMENTS

Many people at Hunter College provided invaluable assistance towards the preparation of this dissertation. Drs. Harvey Ozer and Rivka Rudner were particularly helpful in this regard and have generally guided my progress through graduate work. I would like to thank members of the Ozer and Rudner laboratories, especially Jim Dermody and Russel Widom, for scientific and technical advice, and Brian Wojcik, for moral support. I would also like to thank Dr. Reba Goodman, for her unfailing good humor and encouragement throughout the years and Ms. Sally Ripley, for the quality of the cytogenetic data included in this study.

I could not have completed graduate work without help from my friend, Cindy Hemenway, and my fiancee, David Sigmon, who saw me through difficult periods in the course of these studies. They deserve my unqualified gratitude for their support.

I would like to thank my parents, for their patience, and my undergraduate advisor, Dr. Owen Lovejoy, for inadvertently steering me towards science. I would like to thank, posthumously, Dr. Sherman Beychok for inspiring me to pursue graduate studies and for giving me the final push towards a career in Biochemistry.

Finally, I would like to extent heartfelt thanks to the person who was most instrumental to the completion of these studies, my advisor, Dr. Ann Henderson. Her unfailing guidance and sponsorship throughout my graduate career will be gratefully remembered.

TABLE OF CONTENTS

SECTION		PAGE
INTRODUCTION		1.
MATERIALS AND METHODS		9.
I.	Isolation of High Molecular Weight DNA	9.
II.	Cesium Gradient Ultracentrifugation	10.
III.	Restriction Enzyme Digestion	13.
IV.	Agarose Gel Electrophoresis	14.
V.	Molecular Cloning	17.
VI.	Southern Blot Analysis	26.
VII.	Hybridization *In Situ* to Human Metaphase Chromosomes	31.
VIII.	Nucleotide Sequencing	35.
RESULTS		
PART ONE: MOLECULAR CLONING OF SATELLITE II FRACTION DNA		41.
I.	Isolation of Satellite II-Related Sequences	41.
II.	Molecular Cloning of Satellite II-Related Sequences	42.
III.	Determination of Insert Sizes of Satellite II-Related Cloned Sequences	47.
IV.	Restriction Enzyme Mapping of Cloned Sequences	51.
V.	Southern Blot Analysis of Cloned DNA	55.
PART TWO: CHARACTERIZATION OF SUBCLASS 1 AND 2 CLONED FRAGMENTS		71.
VI.	Genomic Organization of Cloned 43 and 33 DNA	71.
PART THREE: GENOMIC ORGANIZATION OF A DEGENERATE SATELLITE II DNA		94.
VII.	Partial Nucleotide Sequence of the Cloned 33 Fragment	94.
VIII.	Analysis of the Satellite II Component of 33 DNA	112.
IX.	Summary of Results	134.
DISCUSSION		137.
I.	Heterogeneity of Human Satellite II Fraction DNA	137.
II.	The XL3 Cloned Fragment, a Variant Satellite II Sequence	139.
III.	XL3 DNA Defines a Minor Domain of NYS 3.4 kb HaeIII Sequences	145.

IV. Association of Satellite II and Alu Sequences in Subclass 1 DNAs.....150.
V. Origin of Satellite II DNA and Specific Origin of the XL3 Cloned Fragment.....153.

APPENDIX I.....158.

APPENDIX II.....160.

REFERENCES.....170.

AUTOBIOGRAPHICAL STATEMENT.....177.

LIST OF TABLES

TABLE		PAGE
TABLE I	Sizes (in bp) of Satellite II-related fragments	48.
TABLE II	Summary of initial selection of cloned DNA for study	52.
TABLE III	Homologies among Satellite II-related clones: Southern blot data	57.
TABLE IV	Homologies among Satellite II-related clones: analysis of Southern blot data	70.
TABLE V	Genomic organization of fragment 43 DNA: Southern blot analysis	77.
TABLE VI	Cloned 33 fragment sequences S1, S2, S3 and S4	98.
TABLE VII	Base composition of cloned 33 fragment sequences	100.
TABLE VIII	Satellite III pentameric repeat units of the S1 sequence	101.
TABLE IX	Alignment of the partial nucleotide sequence of the cloned 33 fragment with the Satellite II consensus sequence	104.
TABLE X	Alignment of the pPD17, HS3 and pHuR195 sequences with the Satellite II consensus sequence	105.
TABLE XI	Genomic organization of XL3 DNA: Southern blot analysis	127.

LIST OF FIGURES

FIGURE PAGE

FIGURE 1. Cesium gradient isolation of Satellite DNA.....43.

FIGURE 2. Distribution of insert fragments sizes among the Satellite II-related cloned DNAs.....49.

FIGURE 3. Restriction enzyme maps of the cloned fragments.....53.

FIGURE 4. Southern blot analysis of cloned 33 DNA.....58.

FIGURE 5. Southern blot analysis of cloned 2 DNA.....61.

FIGURE 6. Southern blot analysis of Satellite II-related cloned fragments: Hybridization to BLUR 8.....63.

FIGURE 7. Southern blot analysis of cloned 43 DNA.....65.

FIGURE 8. Southern blot analysis of Satellite II-related cloned fragments: Hybridization to pE1.....68.

FIGURE 9. Southern blot analysis of plasmid 43: Hybridization to human genomic DNA, 24 hour exposure.....73.

FIGURE 10. Southern blot analysis of plasmid 43: Hybridization to human genomic DNA, 96 hour exposure.....75.

FIGURE 11. Relationship of the cloned 43 fragment to the L1 KpnI family of repeated sequences.....79.

FIGURE 12. Southern blot analysis of plasmid 33: Hybridization to human genomic DNA.....87.

FIGURE 13. Southern blot analysis of cloned XL3 DNA: Relationship of the XL3 fragment ot Alu element and cloned 33 fragment DNAs.....92.

FIGURE 14. Plasmid constructions used in DNA sequencing.....95.

FIGURE 15. Panel of Satellite II-related and other cloned DNAs..114.

FIGURE 16. Southern blot analysis: Homology of XL3 and BLUR 8 to restriction fragments of Satellite II-related DNAs.....116.

FIGURE 17. Southern blot analysis of XL3 DNA: Hybridization to genomic DNA using relaxed conditions.....120.

FIGURE 18. Southern blot analysis of XL3 DNA: Hybridization to genomic DNA using intermediate conditions.....124.

FIGURE 19. Southern blot analysis of XL3 DNA: Hybridization to genomic DNA using stringent conditions.....128.

FIGURE 20. Genomic RE maps of XL3-homologous DNA.....131.

ABBREVIATIONS

The following abbreviations are used in the text: weight/volume, w/v; volume/volume, v/v; gram, g; milligram, mg; microgram, ug; nanogram, ng; liters, l; milliliter, ml; microliter, ul; sodium dodecyl sulfate, SDS; nanometer, nm; disodium ethylenediamine tetraacetate, EDTA; revolutions per minutes, rpm; units, U; hour, hr; minute, min; centimeter, cm; millimeter, mm; volt, v; ultraviolet, UV; basepair, bp; kilobasepair, kb; molecular weight, MW; gravity, g; ampicillin, amp; dimethylsulfoxide, DMSO; molar, M; millimolar, Mm; micromolar, uM; nanomolar, Nm; picomolar, Pm; deoxyribonuceic acid, DNA; ribonucleic acid, RNA; optical density, O.D.; curie, Ci; millicurie, mCi; microcurie, uCi; counts per minute, cpm; bovine serum albumin, BSA; phytohaemagglutinin, PHA.

INTRODUCTION

Eukaryotic DNA is comprised of an extremely complex array of DNA sequences. At least 55% of the human genome consists of unique sequence DNA (Schmid and Deininger, 1975). The remaining DNA is repeated sequences at various levels of complexity characterized on the basis of length, number of repeats and the type of surrounding or interspersed DNA. Repetitive DNA is organized either as copies of a repeated sequence interspersed with other types of sequences or as arrays of tandemly repeated sequences.

Repeated sequences with interspersion are characterized as short interspersed repeated sequences (SINES) less than 500 basepairs (bp) in length or as long interspersed repeated sequences (LINES) several hundred or thousand bp in length (Singer, 1982). The prototype SINES sequences are Alu elements, consisting of two similar units of approximately 130 bp sequences each. One of the 130 bp units contains a conserved 30 bp region not found in the first 130 bp sequence. The entire element is cleaved roughly in half by the restriction endonuclease AluI and is flanked by nonconserved short (7-12 bp) direct repeats. Alu elements are abundant and occur in approximately 500,000 copies in the human genome (Houck et al., 1979; Rubin et al., 1980; Schmid and Jelinek,

1982). Other less abundant SINES families include the Sau3A and MstII elements, named for the restriction endonucleases that release them from genomic DNA (Menevere et al., 1985; Mermer et al., 1987).

The prototype LINES sequences are the L1 KpnI elements. The L1 family consists of 6.4 kb sequences digested by KpnI into four fragments arranged, 5'->3', as 1.8, 1.5, 1.2 and 1.8 kb sequences. These elements occur in 30,000-50,000 copies in the human genome (Schmeckpeper et al., 1981; Shafit-Zagardo et al., 1982a,b). Both Alu and L1 elements have been described as retrotransposons (reviewed by Hardman, 1986), though direct evidence of transpositional activity by these sequences is limited.

Organization of highly repeated sequences into tandem arrays is characteristic of alphoid (alpha satellite) and Satellite I, II and III DNA. Alphoid sequences are defined by selective restriction endonuclease digestion of human and other primate DNA which reveals highly repetitive sequences approximately 170 bp in length. Alphoid DNA comprises 1-2% of total human genomic DNA and is localized to the centromeric regions of all human chromosomes (Wu and Manuelides, 1980; Mitchell et al., 1985).

Satellite I, II and III DNA consists of simple repeated sequences comprising approximately 5% of human genomic DNA initially isolated using buoyant density cesium gradient ultracentrifugation (Corneo et al., 1970; 1971). The Satellite DNAs comprise different types of repeated simple sequences. Satellite I

DNA occurs as a complex arrangement of two A+T-rich units, A (17 bp) and B (25 bp). Satellite III DNA occurs in tandem arrays of a pentameric 5'ATTCC 3' sequence whereas Satellite II sequences occur in tandem arrays of a 26 nucleotide 5' ATTCC ATTCG G/AGTA/CC ATTCG ATGATG 3' repeat unit (Prosser et al., 1986). Satellite I DNA shares no homology with Satellite II or III DNA, but 40% of Satellite III sequences share homology with 10% of Satellite II sequences. C$_o$t analysis and renaturation kinetics shows that all three Satellites comprise internally heterogeneous collections of different but related sequences (Mitchell et al., 1979). The Satellite DNAs as a group localize predominantly to the centromeres of chromosomes 9, 15, 21, 22, 1, 5, 7, 13, 14, 16, 17, and 20 and to the long arm of the Y chromosome. As expected, chromosomal specificities differ somewhat among specific Satellite DNA fractions, e.g., the major sites of Satellite III homology are the centromeres of human chromosomes 9 and 15, whereas the major sites of Satellite II localization are to the centromeres of human chromosomes 1 and 16 (Gosden et al., 1975; Jones and Corneo, 1971).

Sequences homologous to Satellite I, II and III DNA are found in higher primates. Human Satellite III DNA hybridizes in situ to nonhuman higher primate chromosomes, whereas Satellite I and II sequences do not (reviewed by Jones, 1977). Since 40% of Satellite III sequences share homology with 10% of Satellite II sequences, this suggests that Satellite II arose from Satellite III sometime after the divergence of Hominidae and Pongidae. The degeneracy of

the Satellite III 5'ATTCC 3' pentameric repeat in Satellite II DNA, however, has been interpreted as evidence that Satellite II sequences are evolutionarily "older" than Satellite III sequences (Prosser et al., 1986). Whether Satellite III sequences arose from or preceeded Satellite II sequences in primate evolution, or vice versa, remains to be determined.

Early studies noted that sequence heterogeneities within cesium gradient-isolated human Satellite I, II and III DNA fractions hindered molecular characterization (Mitchell et al., 1979; Prosser et al., 1981; Frommer et al., 1982). Recently, characteristic DNA sequences within Satellite I (Frommer et al., 1984), Satellite II and Satellite III (Cooke and Hindley, 1979; Moyzis et al., 1985; Burke et al., 1985; Jeanpierre et al., 1985; Higgins et al.,1985) DNA have been cloned. The molecular organization of specific cloned sequences has been determined, resulting in the elucidation of some aspects of the heterogeneity within Satellite DNA families. For example, specific members of a 3.4 kb HaeIII family of repeated sequences contain Satellite III (Cooke and McKay, 1978; Bostock et al., 1978; Fowler et al., 1987) or Satellite II (Bostock et al., 1978; Cooke and Hindley, 1979) DNA. Cloned Satellite III-containing 3.4 kb HaeIII repeat sequences localize predominantly to chromosome 9 or 15, and cloned Satellite II-containing sequences localize predominantly to chromosome 1 or 16 (Burk et al., 1985; Jeanpierre et al., 1985; Cooke and Hindley, 1979). The internal sequence heterogeneity of Satellite II and III may be due to the association

of different DNA sequences with components of the 3.4 kb HaeIII repeated elements, conferring chromosome specificity.

Previous studies have provided some information on the molecular organization and internal heterogeneity of Satellite DNA. Many questions, however, remained, and formed the basis for the present study. Specifically, what is the molecular organization of Satellite II sequences in the human genome? This is of importance in light of the paucity of published information describing this complex class of DNAs in a cloned DNA population. Corneo et al. (1970) estimated that Satellite II sequences accounted for 1-2% of total genomic DNA, yet the major localizations of Satellite II DNA are to only two chromosomes, chromosomes 1 and 16 (Jones and Corneo, 1971). Thus, other sites of homology are expected to be present within the human genome. Cloned Satellite II sequences localized primarily to chromosomes 1 or 16, but never to both chromosomes (Cooke and Hindley, 1979; Burk et al., 1985; Jeanpierre et al., 1985; Moyzis et al., 1987). The remainder of Satellite II genomic sequences probably comprised other, related sequence domains that localized to sites traditionally considered "minor" Satellite II chromosomal localizations, i.e., that did not involve chromosomes 1 or 16.

This study defined other sequence components or domains within Satellite II DNA. The investigation was divided into two parts. First, definition of nonSatellite II-derived sequences within Satellite II DNA, and secondly, definition of previously

undescribed Satellite II sequence domains.

The first part of this investigation sought to determine if the internal sequence heterogeneity of Satellite DNA molecular populations could be succinctly defined. If so, could this definition be based on the identification of discrete, relatively homogeneous sequence subclasses within Satellite DNA? Satellite II DNA was initially isolated using cesium-gradient isolated Satellite II fraction DNA to facilitate comparison to previous studies. This fraction, however, could vary in composition in different preparative isolations, and each isolation comprised only microgram quantities of purified DNA (Manuelides, 1978; Singer, 1982). To ensure that the Satellite II DNA used in this study consisted of a constant population of sequences in a quantity sufficient for analysis, Satellite II fraction DNA was cloned. The identification of Satellite- and nonSatellite-related sequence subclasses within all or part of the resulting "library" of Satellite II-related cloned fragments was accomplished by comparison of restriction endonuclease sites and Southern blot analysis involving the cross-hybridization of specific sequences. Nonsatellite II-derived sequence subclasses were further characterized for the presence or absence of unique and other repetitive sequences through detailed Southern blot analysis of representative cloned fragments from different identified subclasses. These experiments defined the nonsatellite II-derived sequence heterogeneity within Satellite II DNA.

The second part of this investigation sought to determine if domains of Satellite II-derived sequences existed in addition to those described in the literature. If so, how did the divergence of sequence components or domains within a Satellite DNA molecular population relate to the molecular organization of Satellite DNA in the human genome? A major aim was to compare the molecular organization of one Satellite component or domain to other components or domains, and relate this to the chromosomal localization of the component or domain sequences.

The Satellite II subclass cloned fragments identified in these experiments were analyzed in detail to determine if they comprised the same or different Satellite II sequence domains. The analysis involved determination of the nucleotide base sequence and molecular organization of Satellite II-derived cloned fragments. The nucleotide base sequences of Satellite II-derived cloned fragments was compared with those of previously described cloned Satellite II-derived DNAs. The degree of homology between these sequences was then used as a measure of the homogeneity or heterogeneity within Satellite II DNA. This was correlated with the domains to which the sequences belonged and served to delineate sequence divergence between members of different Satellite II domains.

The genomic organization of the Satellite II cloned fragments was determined by hybridization to specific restriction endonuclease-produced fragments in genomic DNA, and chromosomal

localization. Human genomic DNA digested with the restriction endonucleases HinfI, TaqI, EcoRI and HaeIII reveals characteristic fragments when hybridized to Satellite II-derived DNA (Cooke and Hindley, 1979; Mitchell et al., 1979; Frommer et al., 1982; Frommer et al., 1984; Burk et al., 1985; Prosser et al., 1986). Hybridization in situ to human metaphase chromosomes was used to determine if specific Satellite II cloned fragments localized to the major Satellite II sites (the centromeric regions of chromosomes 1 and 16), or to minor Satellite II sites on other chromosomes (Jones and Corneo, 1971; Gosden et al.,1975). Together, these experiments determined whether components or domains in addition to those described in the literature existed within Satellite II fraction DNA. In this manner, the internal sequence heterogeneities known to exist within Satellite II could be more completely defined. Moreover, the type of study proposed here would allow a basis for the analysis of the origin, amplification and divergence of these sequences in the human genome. In this way, the evolutionary history of Satellite II DNA could begin to be elucidated.

MATERIALS AND METHODS

The laboratory techniques used in this study include the isolation of high molecular weight human DNA from placental tissue; cesium gradient ultracentrifugation; restriction endonuclease digestion and mapping of restriction sites within human cloned DNA fragments; agarose gel electrophoresis; molecular cloning; purification of plasmid DNA; Southern blot analysis; radiolabelling of DNA; hybridization *in situ* to human metaphase chromosomes and DNA nucleotide sequencing. The methodology of these techniques is presented here with attention to their specific application to the present study.

I. Isolation of High Molecular Weight DNA

High molecular weight human DNA was prepared by a modification of the method of Gall (1968). Twenty to thirty grams (g) of frozen human placental tissue were minced on ice and incubated at 37°C for 12-16 hours in one half volume (w/v) 1X Galls buffer (10mM Tris-HCl, pH 8.0; 10mM NaCl; 10mM EDTA and 1-5% SDS) containing 1 milligram (mg)/milliliter (ml) Pronase (Sigma) previously self-digested at 37°C for 30 minutes (Gall, 1968). The digested tissue was extracted sequentially with equal volumes of phenol, 1:1 phenol:Sevags (Sevags is 24:1 chloroform:isoamyl alcohol), and Sevags and the aqueous phase recovered by centrifugation at 6000xg (Maniatis *et al.*, 1982). The nucleic acids were precipitated by the

addition of two volumes (v/v) of ice cold 95% ethanol and spooled onto a glass rod (Marmur, 1961). The pellet was dissolved in 3-5 ml 0.1X SSC (1X SSC is 0.15M NaCl, 0.015M sodium citrate) and digested with 50 ug/ml boiled pancreatic RNase A (Boehringer Mannheim Biochemicals, BMB) for 1-2 hours at 37°C. The sample was adjusted to 10mM EDTA, pH 8.0 and 0.5% SDS and digested with Pronase at 100 micrograms (ug)/ml for 1-2 hours at 37°C. Following sequential extractions and ethanol precipitation as described above, the purified DNA was dissolved in 1 ml sterile triple-distilled (STD) water. Spectrophotometric readings at 260 nanometers (nm) and 280 nm were read to determine DNA concentration and protein contamination, respectively. The DNA was adjusted with water to a final concentration of 0.5-1.0 mg/ml and stored at 4°C.

II. Cesium Gradient Ultracentrifugation

The initial isolation procedure used $Ag^{++}Cs_2SO_4$ density gradient centrifugation, where the differential binding of silver ions to A+T-rich DNA alters the buoyant density of these sequences and allows their separation from main band DNA as satellites. Satellite II sequences were purified from cesium gradient ultracentrifugation of high molecular weight human DNA essentially as described by Corneo et al. (1970; 1971). The preparation of Cs_2SO_4 gradients, from which Satellite II sequences were cloned, was as follows: per 5.0 ml gradient tube, 250 ug human DNA

purified as described above was mixed with $AgClO_4$ to a final $AgClO_4$:DNA ratio equal to 0.27. The $AgClO_4$/DNA mixture was brought to 0.1M Na_2SO_4 and 0.025M borate buffer (pH 9.2) in a volume of 0.52 ml. To this was added 0.48 ml of a solution of Cs_2SO_4 (Calbiochem) to 1.25 g/ml in 0.1M Na_2SO_4. These components were mixed and added to 4.0 ml of a solution containing 0.6 g/ml Cs_2SO_4. The final concentrations of the various gradient constituents were: DNA at 50 ug/ml, for a total of 250 ug DNA (400 uM); $AgClO_4$ at 95 uM for an $AgClO_4$:DNA ratio equal to 0.27; 0.1M Na_2SO_4; 0.005 M borate, and 0.6 g/ml Cs_2SO_4 for a final density of 1.485 g/ml. The density of the solution was determined by measuring the refractive index on a refractometer (Bausch and Lomb) using the equation

$$p = 13.6986n - 17.3233$$

where p is the density of the solution and n is the refractive index of the solution at 25°C (Vinograd and Hearst, 1962).

Gradients were centrifuged in a Beckman SW50 rotor at 28,000 rpm (100,000 x g) at 25°C for periods of 72-96 hours. Equilibrium was usually achieved by 72 hours. Fractions were collected from the top of the gradients with an ISCO fraction collector. Twenty-five 0.2 ml fractions were collected from each tube and the refractive index of each fraction was determined. On the basis of refractive index and A_{260} readings, fractions 10-12 were pooled as Satellite I and III DNA; fractions 13-18 were discarded as heterogeneous main band DNA, and fractions 19-21 were pooled as Satellite II DNA. An

$AgClO_4$:DNA ratio of 0.27 favors separation of Satellite II rather than Satellites I and III from main band DNA (Corneo et al., 1970). Under these conditions, a peak was observed for Satellite II, whereas Satellites I and III separated as small shoulders.

Isolated satellite fractions were pooled separately and dialyzed exhaustively against 4M NaCl, 0.1M Tris-HCl, pH 8.0, and 0.1M EDTA, pH 8.0, to remove the heavy salt (as modified from Gosden et al., 1975). After ethanol precipitation, the dialyzed DNAs were dissolved in TE (10mM Tris-HCl, pH 8.0, 1mM EDTA, pH 8.0) and the concentrations determined spectrophotometrically at 260 nm.

CsCl density gradient ultracentrifugation was used to achieve further purification of the Satellite DNA fractions. Satellite II and III fractions, pooled separately and dissolved in TE, were brought to 0.95 g/ml CsCl to achieve a final DNA concentration of 10-50 ug/ml and a final density of 1.700 g/ml. The density of the solution was determined by measuring its refractive index with the refractometer using the equation:

$$p = 10.8601n - 13.4974$$

where p is the density of the solution and n is the refractive index of the solution at 25°C (Vinograd and Hearst, 1962).

The Satellite II or III fractions were each centrifuged in CsCl gradients in a Beckman SW50 rotor at 28,000 rpm (100,000 x g) at 25°C for 48-72 hours. Fractions were collected and the buoyant densities determined. Satellite II fractions were identified at p = 1.694 g/ml, and Satellite III fractions at p = 1.697 g/ml (Corneo

et al., 1970; 1971). Heavier fractions of densities equal to or greater than 1.700 g/ml were discarded as heterogeneous main band DNA.

Satellite II or III fractions were readjusted to 5.0 ml with solid CsCl and TE to ρ = 1.700 g/ml and recentrifuged as described to homogeneity, that is, until the measurable DNA in the gradient equilibrated to a density equal to 1.694 (Satellite II) or 1.697 (Satellite III) g/ml, and a single peak at A_{260} was apparent. The Satellite II and III fractions used in molecular cloning were purified through one Cs_2SO_4 and two CsCl ultracentrifugations.

III. Restriction Enzyme Digestion

The common restriction endonucleases AluI, BamHI, BclI, BglII, EcoRI, HaeIII, HindIII, HinfI, HpaI, KpnI, PstI, PvuII, SalI, SstI, TaqI, XbaI and XhoI were used for analysis of Satellite II and III DNAs. Reaction conditions were: 1X low salt buffer (10mM Tris-HCl, pH 7.5; 10mM $MgCl_2$; 1mM DTT) for BglII, HpaI, KpnI, SstI and TaqI; 1X medium salt buffer (10mM Tris-HCl, pH 7.5; 10mM $MgCl_2$, 1mM DTT, 50mM NaCl) for AluI, BamHI, BclI, HaeIII, HindIII, HinfI, PstI and PvuII, and 1X high salt buffer (50mM Tris-HCl, pH 7.5; 10mM $MgCl_2$, 1mM DTT, 100mM NaCl) for EcoRI, SalI, XbaI and XhoI (Maniatis et al., 1982). All reactions contained 1-10 units (U) restriction enzyme per 1-10ug DNA in the appropriate reaction buffer. Digestions were incubated at 37°C, except those involving BclI or TaqI, which were carried out at 60-65°C.

Reactions involving digestion with more than one enzyme that utilized the same reaction conditions were incubated 1-2 hours at the appropriate temperature. Reactions utilizing two enzymes requiring different reaction conditions were performed as follows: 1) 1X low salt/1X medium salt buffer conditions were met by incubating the reaction for one hour with the first enzyme in 1X low salt buffer, adjusting the reaction mixture to 50mM NaCl by the addition of 0.5 microliters (ul) 1M NaCl per 10 ul reaction, adding the second enzyme and continuing digestion for an additional hour; 2) 1X medium salt/1X high salt buffer conditions were met as described in 1), except that the reaction mixture was adjusted to 100mM NaCl, 50mM Tris-HCl, pH 7.5, by the addition of 0.5 ul 1M NaCl and 0.4 ul 1M Tris-HCl, pH 7.5, to the reaction mixture; 3) 1X low salt/1X high salt buffer conditions were met as in 1) and 2), except that the reaction was adjusted to 100mM NaCl, 50mM Tris-HCl, pH 7.5, by the addition of 1 ul 1M NaCl and 0.4 ul 1M Tris-HCl, pH 7.5, to the reaction mixture.

IV. Agarose Gel Electrophoresis

Agarose gels (McDonell et al., 1977) were prepared by dissolving solid agarose (Bethesda Research Laboratories, BRL) in 1X E Buffer (40mM Tris-HCl, 20mM NaOAc, 1mM EDTA adjusted to pH 7.7 with glacial acetic acid) with heat to a final concentration of (w/v) 0.7%, 1.0%, 1.2% or 1.5%. Agarose was poured to 0.5 cm thickness in horizontal gel trays measuring 6 cm X 8 cm (minigel)

or 12 cm X 20 cm (medium gel). Well-forming combs were positioned in the tray slots and the agarose allowed to polymerize for 30 minutes (min). The combs were removed and the gels immersed in 1 X E Buffer in the horizontal electrophoresis tanks for 5-15 min. prior to use.

Loading buffer (0.25% bromphenol blue dissolved in 1:1 1 X E Buffer:glycerol) was added to the DNA samples prior to electrophoresis to a final glycerol concentration of 12-15%. The samples were heated to 68°C for 10 min. to dissociate aggregated fragments and loaded immediately into the gel wells. Minigel electrophoresis was carried out at 12-15 V/cm for 1-1.5 hours or until the bromphenol blue migrated 3/4 the way down the gel (~6 cm from the wells). Medium gel electrophoresis was carried out at 5V/cm for 3-4 hours or at 1V/cm for 12-16 hours until the bromphenol blue migrated 3/4 the way down the gel (~15 cm from the wells) (Maniatis et al., 1982).

Upon completion of electrophoresis, the gels were stained in 0.5 ug/ml ethidium bromide (Sharp et al., 1973) (from a 10 mg/ml stock) for 15 min. (minigel) or 30 min. (medium gel), rinsed 3X with distilled water and the DNA visualized on a long wave ultraviolet (UV) transilluminator (Spectroline). The gels were photographed on the transilluminator using a Konica instant camera equipped with a Kodak 22A red Wratten filter using Polaroid type 667 Land film (Maniatis et al., 1982).

After photography, the molecular weights of the electrophoresed

DNA fragments were determined by comparison with HindIII-digested lambda DNA, which was electrophoresed in the last well of each gel as a molecular weight marker. The log molecular weights of the 23.1, 9.64, 6.64, 4.34, 2.26, 1.98 and 0.56 kb lambda/HindIII fragments were plotted on the log (y) axis of 3 cycle semilog paper versus the distances migrated from the gel wells (in mm) plotted on the linear (x) axis, creating a $\log_{MW}$ vs. d(mm) graph. A straight line was drawn through the points, providing a direct correlation between $\log_{MW}$ and d(mm). The distances migrated from the gel wells of other fragments on the same gel were measured, plotted on the x axis of the graph and the corresponding log molecular weights read from the y axis. The molecular weight values were divided by 660 daltons/basepair to determine the basepair lengths of the gel fragments (Helling et al., 1974). This procedure was repeated for each gel to ensure accuracy and was used to determine the sizes, in basepairs (bp), of Satellite II and III fragments inserted as recombinant DNAs in pBR322 and to determine the location of restriction enzyme (RE) sites in the recombinant plasmid sequences. Twenty four, 6cm x 8cm, 0.7% agarose gels were required to determine the insert sizes of Satellite II and III cloned fragments, and twenty, 12cm x 20cm, 1.0% agarose gels were required to construct RE maps of the twelve cloned fragments chosen for study.

V. Molecular Cloning

Molecular cloning techniques were used to: 1) propagate DNA fragments produced by HindIII digestion of Satellite II fraction DNA (recombinant plasmids 1-170); 2) propagate DNA fragments produced by HindIII digestion of Satellite III fraction DNA (recombinant plasmids B, C and D); 3) prepare a particular Satellite II cloned fragment, 33 DNA, for DNA sequencing (recombinant plasmids 4A, 5N, HX3 and XL3), and 4) isolate Satellite II from Alu sequences within 33 DNA by separately subcloning the large 750 bp HindIII/XbaI fragment (recombinant plasmid XL3).

A. Preparation of Vector DNA

Two ug restriction enzyme-digested pBR322, pSP64 or pSP65 plasmid DNA was extracted once with 1:1 phenol:Sevags, ethanol precipitated, and dephosphorylated in a 50 ul reaction containing excess (10-20U) calf intestinal phosphatase (CIP) (BMB) and 1X CIP buffer (50mM Tris-HCl, pH 9.0; 1mM $MgCl_2$; 0.1mM $ZnCl_2$; 1mM spermidine) incubated at 37°C for 2 hours (Maniatis et al., 1982). The reaction mix was extracted once with 1:1 phenol:Sevags, ethanol precipitated and resuspended in TE to a final concentration of 100 nanograms (ng)/ul, determined by co-electrophoresis with known quantities of linearized pBR322 DNA.

B. Preparation of Insert DNA

One ug Satellite II fraction DNA was digested with HindIII at 37°C for 2 hours, extracted with an equal volume of 1:1

phenol:Sevags, brought to 50 ug/ml with bacterial tRNA as carrier, ethanol precipitated and resuspended in 5 ul TE to a final concentration of 100-200 ng/ul prior to the ligation reaction. 0.5 ug Satellite III fraction DNA was similarly treated prior to the ligation reaction. All DNA for sequencing was isolated for ligation into pSP64/65 plasmid vectors by digestion of 5 ug DNA with HindIII, electrophoresis of the digestion products through 0.7% agarose gels, excision of fragments from the gels and elution from gel slices using electroelution (IBI), as per manufacturer's instructions. 50 ug/ml bacterial tRNA was added as carrier and recovered DNA fragments were precipitated with ethanol. The pellet was resuspended in 10 ul water and the DNA concentration calculated at 50-100 ng/ul by comparison in co-electrophoresis of aliquots with 50, 100 and 200 ng quantities of HindIII-digested pBR322 DNA. HindIII/XbaI 33 DNA fragments were obtained for ligation into the pSP65 plasmid vector by digestion of 5 ug plasmid 33 with HindIII+XbaI followed by gel electrophoresis and electroelution of the HindIII/XbaI fragments as described above.

C. Ligation of Vector and Insert DNA

Vector and insert DNAs prepared as described above were coprecipitated with ethanol prior to the ligation reactions. The Satellite II cloned DNAs (recombinant plasmids 1-170) were prepared using 0.8 ug HindIII-digested Satellite II DNA coprecipitated with 200 ng HindIII-digested, dephosphorylated pBR322 DNA. The Satellite III cloned DNAs (recombinant plasmids B, C and D) were prepared

using 0.4 ug HindIII-digested Satellite III DNA coprecipitated with 200 ng HindIII-digested, dephosphorylated pBR322 DNA. The DNA subcloned for use in DNA sequencing (recombinant plasmids 4A and 5N) was obtained by coprecipitation of 0.4 ug HindIII-digested, electroeluted cloned 33 DNA fragment with 100-150 ng HindIII-digested, dephosphorylated pSP64 and pSP65 plasmid DNA in two separate reactions. Regions of 33 DNA were subcloned for use in DNA sequencing (recombinant plasmid HX3) by coprecipitation of 0.2-0.25 ug XbaI-digested 33 DNA with 0.1-0.15 ug HindIII/XbaI-digested, dephosphorylated pSP65 DNA. The large 750 bp HindIII/XbaI region of 33 DNA was subcloned for use in DNA sequencing (recombinant plasmid XL3) by coprecipitation of 0.3 ug electroeluted 750 bp HindIII/XbaI 33 fragment with 0.1-0.2 ug HindIII/XbaI-digested, dephosphorylated pSP65 DNA.

Coprecipitated insert and vector DNAs were centrifuged at 13,000xg, the pellet was resuspended in a 10ul volume consisting of 0.5U T4 DNA Ligase (BMB), 1X Ligation buffer (50mM Tris-HCl, pH 7.4; 10mM $MgCl_2$; 10mM DTT; 1mM spermidine; 1mM ATP, pH 7.0; 0.1 mg/ml bovine serum albumin [BSA]) and water (Maniatis et al., 1982). The vector-insert mixture was incubated 12-16 hours at 14-16°C. 2-3ul of the ligation mixture was electrophoresed on a 1.0% minigel and the products of the ligation reaction were visualized with UV light after staining the gel with ethidium bromide, as previously described. The ligation reaction was judged successful if the band corresponding to unligated vector DNA was largely absent and laddered or smeared bands suggestive of ligated

DNAs were present.

D. Preparation of Competent Cells

DH1 recA- strain E. coli cells were made competent to accept plasmid DNA by modification of the method of Hanahan (1983) (as developed by Small, unpublished). Fifty ul DH1 cells (equal to 1 x 10^8 cells) from a 2 x 10^9 cells/ml frozen stock were innoculated into 5.0 ml TYM broth (2.0% Bacto-Tryptone [Difco]; 0.5% Bacto Yeast Extract [Difco]; 0.1M NaCl and 0.01M $MgCl_2$ or $MgSO_4$). The cells were grown with aeration at 250 rpm for 18 hours at 37°C. 0.1-0.4 ml of the culture was used to innoculate 25.0 ml TYM broth to an initial cell density of 1 x 10^7 cells/ml, equivalent to a Klett colorimeter value of 10 units. The cells were grown with aeration at 250 rpm and 37°C for 3-4 hours until they achieved an approximate density of 5 x 10^7 cells/ml (equal to a Klett colorimeter value of 150-170 units). The cells were divided into two sterile screw-capped polypropylene tubes (Corning) and pelleted at 1000xg for 5 min. at 4°C. The supernatants were discarded and the pellets resuspended in a total volume of 8.0 ml Transformation Buffer I (30mM KOAc; 50mM $MnCl_2$; 100mM KCl; 10mM $CaCl_2$; 15%[w/v] glycerol, to pH 5.8 with 0.2N glacial acetic acid). After a 10 min. incubation on ice, the cells were pelleted at 1000xg, the supernatants discarded and the cells resuspended in a total of 1 ml Transformation Buffer II (10mM Na+-MOPS, pH 7.0; 75mM $CaCl_2$, 10mM KCl and 15% [w/v] glycerol). The cells were divided into 0.2 ml aliquots and pipetted into 1.5

ml round-bottomed sterile polypropylene freeze vials (Corning), quick-frozen for 5 minutes in a dry ice/ethanol bath and stored at -70°C.

E. Transformation of Cells

Cells were transformed essentially as described by Hanahan (1983). 10-100 ng supercoiled pBR322 DNA were mixed with 0.2 ml competent DH1 cells, placed on ice for 30 min., then heat-shocked at 42°C for 90 seconds. 0.8 ml TYM broth was added and the cells incubated with aeration at 225 rpm and 37°C for 60 min. 100 ul cells were plated on TYM agar plates (1.5% [w/v] agar [Difco]) that were Amp- and Amp+ (75 ug/ml ampicillin from a 25mg/ml stock [BMB]), respectively. The plates were inverted and incubated 18 hours at 37°C. The cells were judged competent if confluent or near-confluent colony densities (>2000 colonies) were visible on the Amp+ plate. Cells shown to be competent were transformed as described using ligated recombinant DNA. In some experiments, frozen competent DH5 E. coli cells (BRL) were used in the transformation procedures. Transformations were carried out as described above, except that 0.1 rather than 0.2 ml competent cells were used per transformation.

The Satellite II cloned fragments (recombinant plasmids 1-170) were obtained by transformation of 100 ul competent DH5 cells with 100 ng ligated Satellite II/pBR322 DNA. The Satellite III cloned fragments (recombinant plasmids B, C and D) were obtained using 50 ng ligated Satellite III/pBR322 DNA. The 33 DNA subcloned

fragments for use in DNA sequencing were obtained using 200 ng ligated 33 DNA/pSP64 DNA (recombinant plasmid 4A) and 200 ng ligated 33 DNA/pSP65 DNA (recombinant plasmid 5N). The HindIII-XbaI DNA fragments subcloned for use in DNA sequencing were obtained using 100 ng ligated HindIII-XbaI 33/pSP65 DNA (recombinant plasmid HX3) and 100 ng ligated 750 bp HindIII-XbaI 33 fragment/pSP65 DNA (recombinant plasmid XL3). 100 ul of 1X, 1/10X and 1/100X dilutions were plated in duplicate as plates A and B on ampicillin-containing TYM-agar plates. 100 ul of 1X transformed cells were also plated on one Amp- plate. The plates were inverted and incubated 18 hours at 37°C.

F. Purification of Plasmid DNA

1. Minipreparation of Plasmid DNA

Cloned DNAs were initially analyzed using small-scale recombinant plasmid isolations, termed " minipreparations". Two ml TYM supplemented with 50-75 ug/ml ampicillin were innoculated with a single colony and the transformed bacteria were grown 18 hours with aeration at 250 rpm and 37°C. 0.5 ml of the 2.0 ml culture was transferred to a freeze vial (Corning), adjusted to 7.0% dimethylsulfoxide (DMSO) and stored at -20°C as a stock culture. The remaining 1.5 ml of a 2.0 ml culture was subjected to alkaline lysis to extract plasmid DNA as described by Birnboim and Doly (1979). The culture was transferred to a 1.5 ml microcentrifuge tube and centrifuged at 13,000xg for 1 min. The pelleted cells were resuspended and lysed in 100 ul cold 50mM glucose, 10mM EDTA, 25mM

Tris-HCl, pH 8.0, 250 ug/ml RNase (Sigma) and 5 mg/ml egg white lysozyme (Worthington). After incubation at room temperature for 5 min., 200 ul of an ice cold solution of 0.2N NaOH, 1% SDS was added to each tube to denature DNA and proteins. Following incubation on ice for 5 min, the mixtures were neutralized by the addition of 150 ul of ice cold 3M KOAc (to pH 5.2 with glacial acetic acid for a final acetate concentration of 5M). The tubes were inverted to mix and incubated on ice for 5 min, then centrifuged at 13,000xg for 5 min. to pellet proteins and chromosomal DNA. The supernatants containing plasmid DNA were transferred to fresh tubes and extracted with an equal volume 1:1 phenol:Sevags. The aqueous phase was recovered after centrifugation at 13,000xg for 1 min. and the plasmid DNA precipitated with 1 ml 95% ethanol per tube at room temperature. The pellets were washed once with 70% ethanol to remove debris, resuspended in 10 ul water and stored at 4°C. This method typically recovered 2-5 ug plasmid/2ml minipreparation.

2. Large Scale Isolation of Plasmid DNA

Restriction enzyme mapping and Southern blot analysis of cloned sequences required large amounts of DNA, which were prepared using large scale isolation procedures. Two ml TYM supplemented with 75 ug/ml ampicillin were innoculated with 20 ul frozen stock culture and grown with aeration at 250 rpm and 37°C for 18 hours. 0.1-0.4 ml of the overnight culture was transferred to 25 ml TYM supplemented with 75 ug/ml ampicillin to an increase in the Klett colorimeter value of 10 units (equal to 1 x 10^7 cells/ml final cell density). The bacteria were grown with aeration at 250 rpm and

37°C for 3-4 hours until they reached mid-late logarithmic growth and a density equivalent to an $O.D._{660}$ = 0.6, measured as an increase in Klett colorimeter values equal to 120-140 units. The 25 ml culture was transferred to 500 ml TYM supplemented with 75 ug/ml ampicillin and grown with aeration at 250 rpm and 37°C for 2.5 hours. Chloramphenicol (BMB) dissolved in 95% ethanol at 34 mg/ml was added to the media to a concentration of 170 ug/ml and growth was resumed with aeration at 250 rpm and 37°C for 18 hours.

Each 500 ml culture was divided into two 250 ml bottles (Nalgene) and the bacteria pelleted by centrifugation at 4000xg for 10 minutes at 4°C. The supernatants were discarded and the pellets washed with 50 ml each ice cold STE (0.1M NaCl, 10mM Tris-HCl, pH 8.0, 1mM EDTA) (Maniatis et al., 1982).

Plasmid DNA was purified from cell pellets by alkaline lysis (Birnboim and Doly, 1979). Pellet were resuspended (by vortexing) in 2.0 ml ice cold lysis buffer (50mM glucose, 25mM Tris-HCl, pH 8.0, 10mM EDTA, 150 ug/ml RNase and 5 mg/ml egg-white lysozyme). The lysates were transferred to 10 ml screw-capped polypropylene ultracentrifuge tubes (Nalgene) and incubated 5 minutes at 25°C. 4.5 ml of an ice cold 0.2N NaOH, 1.0% SDS solution was added. The tube contents were mixed by inversion and incubated 10 minutes on ice. 3-3.5 cold 5M KOAc (to pH 5.2 with glacial acetic acid) was added to fill the tube, the contents of the tube mixed by inversion and incubated on ice for 10 minutes. After centrifugation in the Ti50 rotor (Beckman) at 29,000xg for 20 minutes at 4°C, the

supernatants were collected and transferred to 15 ml Corex (Corning) tubes. The supernatants were each mixed with 0.6 volumes (5-6 ml) isopropanol and incubated at room temperature for 15 minutes. DNA was recovered by centrifugation at 13,000xg for 30 minutes at room temperature. Pellets were resuspended in 8 ml TE. One gram of CsCl was added per ml solution and ethidium bromide was added to 80 ug/ml. The preparations were transferred to high speed screw capped ultracentrifuge tubes (Beckman) and centrifuged at 42,000 rpm in the Ti50 rotor for 36-48 hours at 25°C. The plasmid DNA was visualized as a distinct fluorescent band in the lower half of the tube. A 1 ml syringe fitted with a long (1 1/2 inch) 18 gauge needle (Beckton-Dickson) was used to transfer the plasmid DNA into separate 15 ml polypropylene screw capped centrifuge tubes (Corning). The samples were extracted 6-8 times with water-saturated n-butanol followed by centrifugation at 1000xg at room temperature for 5 minutes to remove ethidium bromide. The extracted DNA was transferred to boiled dialysis tubing and dialyzed at room temperature against several 1 liter (l) changes of TE buffer for 3-4 hours until the refractometer reading indicated that the heavy salt had been removed. The samples were transferred to 15 ml Corex tubes and combined with 1 ml TE washes of the tubing, adjusted to 150mM NaCl and precipitated with 2 volumes of ice cold ethanol. The pellets were resuspended in 0.5 ml TE and the DNA concentration determined by spectrophotometric readings at 260 nm and comparison of electrophoresis patterns on a 1.0% minigel

with known amounts of pBR322 (typically, 100, 200 and 400 ng aliquots). The plasmid concentration was adjusted to 1-5 mg/ml with TE and the DNA stored at 4°C.

VI. Southern Blot Analysis

A. Southern Transfer of DNA

DNA was transferred from agarose gels to GeneScreen Plus (NEN) essentially as described by Southern (1975). Restriction site analysis of cloned DNA used digestion of 2 ug each recombinant plasmid DNAs, 8 ug placental DNA and 1 ug lambda DNA with HindIII, followed by electrophoresis on 1.0% agarose gels. Genomic DNA was analyzed with digestion of 8 ug each placental DNA using various restriction endonucleases. The pattern was compared with that obtained from digestion of 1 ug pBR322 DNA, 0.5 and 1.0 ug cloned DNA and 1 ug lambda DNA with HindIII followed by electrophoresis through 1.0% agarose gels. The gels were stained with ethidium bromide following electrophoresis and photographed. The DNA in the gel was then denatured by soaking the gel in 0.5 - 1.0 liter 1.5M NaCl, 0.5N NaOH for 30 minutes at room temperature. The gel was rinsed with 1 liter distilled water and neutralized by immersion in 0.5 - 1.0 liter 3.0M NaCl, 0.5M Tris-HCL, pH 7.0, for 30 minutes at room temperature. After rinsing in 1 liter distilled water the gel was inverted and laid onto 3mM Whatman paper on a platform elevated in a tank containing 2.5 liters 2 X SSC (transfer of plasmid DNA) or 6 X SSC (transfer of genomic DNA). A gel-sized piece of

GeneScreen Plus saturated with the appropriate (2X or 6X) concentration of SSC was placed on the gel. Air bubbles between the gel and filters were removed, paper towels and weights applied and the transfer allowed to proceed for 18 hours. The filter with bound DNA was then air dried or dried *in vacuo* at 60°C for one hour.

B. Radiolabelling of Probe DNA

DNA was labelled with ^{32}PdATP (>3000 Ci/mmole; NEN) using nick-translation or random primer extension techniques.

Nick-translation of DNA was accomplished essentially as described by Maniatis *et al.* (1975). Between 100-500 ng DNA was mixed with 20uM each dTTP, dCTP and dGTP and 50-100 uCi (100-150 pmoles) ^{32}PdATP in 50mM Tris-HCl, pH 7.9; 5mM $MgCl_2$; 10mM 2-mercaptoethanol, 50 ug/ml BSA and distilled water to a final 30 ul volume. DNase (BMB), in 10mM Tris-HCl, pH 7.9; 5mM $MgCl_2$ and 50 ug/ml BSA, was added to a final concentration of 1 ng/ml and the reaction incubated 5 minutes at room temperature. 5U DNA Polymerase I (Klenow fragment) (BMB) was added and the reaction incubated at 15°C for 2 hours. Unincorporated nucleotides were removed by passing the mixture through a pD10 Sephadex G-25 column (Pharmacia) followed by the addition of 50 ug/ml bacterial tRNA as carrier and ethanol precipitation. Alternately, the reaction mixture was passed through a NENSORB (NEN) column and the eluate evaporated to dryness under N_2. In either case, the pelleted DNA was resuspended in 100 ul STD water. Two or three 1 ul aliquots were counted in 0.5 ml water and 10 ml scintillation cocktail (Ecoscint) using a preset

^{32}P beta radiation channel of the liquid scintillation counter (Packard).

Random primer extension labelling was accomplished using a kit and accompanying instructions supplied by the manufacturer (BMB) (Feinberg and Vogelstein, 1984).

This method requires prior linearization of probe DNA. DNA was linearized using BamHI, which digests the vector pBR322 DNA at one site, but did not digest any of the inserted Satellite II DNA fragments used in these experiments. The pE1 plasmid, which contains four tandemly repeated alphoid monomer units, was used in these studies to detect the presence of alphoid sequences in cloned Satellite fragments and was linearized with BamHI prior to radiolabelling (Gray et al., 1985). The BLUR 8 plasmid, which contains an Alu SINES element, was used to identify Alu sequences within Satellite II cloned fragments (Rubin et al., 1980). BLUR 8 DNA was linearized with HindIII, since BamHI removes the entire inserted Alu element from the plasmid. One hundred ng of digested plasmid DNA was denatured by boiling for 10 minutes, then placed immediately on ice to prevent renaturation. The reaction components dTTP, dCTP and dGTP (25 uM each), 2 ul hexanucleotide/reaction buffer mix, 50uCi ^{32}PdATP (>3000 Ci/mmole) (NEN) and 5U (1 ul) DNA Polymerase I (Klenow fragment) and distilled water were added to the denatured DNA to a 20 ul volume. The mixture was incubated at 37°C for 30-60 minutes. Unincorporated nucleotides were removed by passing the reaction volume through a NENSORB (NEN)

column and incorporated radioactivity was assayed as described above. Either procedure labelled DNA to specific activities between 5 x 10^6-5 x 10^8 cpm/ug.

C. Hybridization of Southern Transfers

The radiolabelled probe DNA was denatured with alkali prior to hybridization by adjusting the sample to 0.1N NaOH and incubating at room temperature for 15 minutes. The denatured DNA was placed on ice and neutralized by the gradual addition of 1N HCl to between pH 7 and 8, determined with pH paper. The probe was denatured less than 30 minutes prior to use and stored on ice to prevent renaturation.

The transfer filter was pre-hybridized between 3-4 hours in a solution containing 6 X SSC, 0.5% SDS, 5 X Denhardt's (50 X Denhardt's is, w/v, 1.0% Ficoll, 1% polyvinylpyrrolidone and 1% BSA) and 100 ug/ml sheared, denatured salmon sperm DNA in a volume equal to 0.2 ml solution/cm^2 filter in a heat-sealed plastic bag (Dazey) (Southern, 1975; Maniatis et al., 1982). Prehybridization temperatures of 55-60°C were used in the analysis of cloned fragments and genomic DNA with the exception of the cloned XL3 fragment, which was prehybridized under relaxed (50°C), intermediate (60°C) and stringent (72°C) conditions. After prehybridization, the volume in the bag was reduced to an amount equal to 50 ul/cm^2 filter and adjusted to 10mM EDTA. The probe was added, the bag re-sealed and hybridization was conducted at prehybridization temperatures for 12-18 hours.

Following hybridization, the filter was removed from the bag and washed 3X for 5 minutes each in 500 ml 2 X SSC, 0.5% SDS at room temperature, followed by a 15 minute wash in 500 ml 2 X SSC, 0.1% SDS at room temperature. The filter was then washed at a temperature equal to 5°C below the prehybridization temperature in 1 liter of 0.1 X SSC, 0.5% SDS with three changes for 2.5 hours (Maniatis et al., 1982). The transfers were dried in vacuo at 60°C for 30 minutes or until dry, wrapped in plastic wrap and autoradiographed using XAR-5 film (Kodak) with an intensifying screen (Cronex Lightning Plus, DuPont) at -70°C. The filters were exposed to X-ray film for appropriate time intervals (between 1 and 96 hours) to visualize hybridization signals of different intensities. The film was developed for 1-5 minutes in 1 liter of Kodak GBX Developer, or until the hybridization signals were visualized, immersed 1 minute in 1 liter 3% acetic acid to arrest development, then hardened in 1 liter rapid fixer (Kodak) for 10 minutes. The film was rinsed under cold running water for 15 minutes and air dried.

Southern blots were used to determine molecular weights of cloned and genomic DNAs. This was accomplished by constructing a $\log_{MW}$ vs. d(mm) graph for the agarose gel from which the DNA was transferred, then correlating the d(mm)s of fragments of known molecular weight on the gel with the same fragments on the autoradiograph. The fragments of known molecular weight used on the gel and apparent on the autoradiograph were pBR322 DNA and the

probe DNA. The correlation between gel and autoradiograph, expressed as a multiplication coefficient equal to N where $d(mm)_{gel} = N \times d(mm)_{autoradiograph}$, was used to convert the distances migrated by fragments on the gel with the distances migrated by fragments on the autoradiograph.

VII. Hybridization In Situ to Human Metaphase Chromosomes

A. Preparation of Slides

The following procedures, with the exception of DNA radiolabelling and statistical calculations, were conducted by Ms. Sally Ripley, a cytogeneticist in the laboratory of Dr. Ann Henderson. Human metaphase chromosomes were prepared from phytohaemagglutinin (PHA)-stimulated normal peripheral blood lymphocytes. Lymphocytes were isolated from whole blood by centrifugation through LeukoPREP Tubes (Becton-Dickinson). The lymphocyte layer was recovered and the cells cultured in RPMI 1640 media (Gibco) supplemented to 10% with fetal calf serum (Gibco). The T-cell population was stimulated to divide by the addition of PHA to 90 ug/ml (Gibco) and grown at 37°C for 72 hours. Colcemid was added to 0.6 ug/ml and the cells incubated 60 minutes at 37°C to arrest division at mitotic metaphase. The cells were harvested by centrifugation at 1000xg for 5 minutes at room temperature and resuspended in 5 ml hypotonic 0.075M KCl for 15 minutes to rupture the cell membranes. Cell nuclei were recovered by centrifugation at 1000xg for 5 minutes. Membranes and cellular debris were removed by repeated resuspensions of nuclear pellets in 5 ml fixative (3:1

v/v glacial acetic acid:methanol) and centrifugation for 5 minutes at room temperature. The nuclei were stored in fixative until use. 3-4 drops resuspended nuclei were dropped onto each slide and the slides passed through a flame to burst the nuclear membranes and fix the released chromosomes to the slide. The chromosomes were stained by exposing the slides to a solution of (v/v) 10% Giemsa, 1X E Buffer for 5 minutes. The slides were then rinsed, blotted dry and the chromosomes visualized under light microscopy. Slides that displayed an acceptable metaphase index (at least 20 out of 100 nuclei) were used for hybridization *in situ* experiments.

Chromosomes were G-banded by a modification of the method of Bignone *et al.* (1983) and identified using conventions established by the International System for Human Cytogenetic Nomenclature (ISCN 1981). Slides were stored *in vacuo* at 42°C and exposed to a solution of 0.5M monobasic phosphate, pH 8.5 (not adjusted), 20% (v/v) methanol, 8% (v/v) trypsin (Gibco) and 2% (v/v) Giemsa for 5-7 minutes at room temperature or until banded. The slides were rinsed, blotted dry, and examined using light microscopy. This procedure was repeated until the chromosomes were G-banded. 20-30 metaphase plates of well-banded chromosomes per slide were photographed using an 80X objective under the light microscope with Pan-X film (Kodak) and the coordinates recorded. The chromosomes on "prephotographs" were specifically identified for later analysis after hybridization *in situ* to probe DNA.

Prior to hybridization, the slides were treated with 50 ug/ml

RNase in 2 X SSC at room temperature for 1 hour to remove endogenous RNA, rinsed once in 2 X SSC, twice in 70% ethanol and three times in 95% ethanol, then air dried.

B. Preparation of Probe

Probe DNA was radiolabelled as previously described, except that ^{125}IdCTP (>2000 Ci/mmole) (NEN) was used rather than ^{32}PdATP. 100-200 uCi (100-200 pmoles) ^{125}IdCTP (100-200 ul), in ethanol, was evaporated to dryness under N_2, resuspended in 5 ul distilled water and added to the reaction mixture. The specific activity of the probe was determined as described except that the ^{125}I gamma radiation channel of the liquid scintillation counter was used. Carrier tRNA was added to 50 ug/ml, the probe was ethanol precipitated, resuspended in 100 ul STD water and the final counts determined immediately prior to use. The specific activities of probes labelled in this manner were usually in the range of 5 x 10^7 to 5 x 10^8 cpm/ug.

The DNA probe was alkali denatured by incubation at room temperature for 15 minutes in 0.1N NaOH, then placed on ice and neutralized by the gradual addition of 1N HCl to pH 7-8, as assayed with pH paper. The DNA was adjusted to 3 X SSC:50% formamide by the addition of 20 X SSC, formamide (adjusted to pH 7 with HCl) and STD water to a final volume equal to 30-50 ul solution per slide. The denatured probe was kept at -20°C no longer than 30 minutes prior to use.

C. Hybridization of Probe and Chromosomal DNA

The chromosomal DNA was denatured by incubating the slides in 95% formamide:1 X SSC (to pH 7.1 with HCl) for 1.5 hours at 68°C, as described by Henderson (1982). The slides were immersed in cold 70% ethanol for 10 minutes at -20°C to stop denaturation, washed 3X each in ice cold 70% and 95% ethanol and blotted dry. 30-50 ul of the probe solution was applied to each slide and held in place with a coverslip. The slides were positioned horizontally atop slide mounts in separate petri plate hybridization chambers containing filter paper soaked in 3 X SSC:50% formamide. Metaphase plates were incubated 18 hours under reasonably stringent conditions at 42°C (cloned 43 probe DNA) or under less stringent conditions at 38°C (cloned XL3 probe DNA). The slides were placed in a dish containing 3 X SSC:50% formamide to float off the coverslips, then incubated for 10 minutes at 45°C in 2 X SSC to remove unhybridized probe DNA. They were then washed for 3 hours at room temperature in 2 liters of 2 X SSC with several changes and rinsed successively in several volumes of 70% and 95% ethanol. After air-drying, the slides were dipped in 1:1 Kodak NTB-2 emulsion:distilled water, boxed and exposed at 4°C for 24-72 hours. Test slides were removed at 24 hour intervals to check grain densities. All slides were developed for 2.5 minutes in 50 ml Kodak D-19 developer. Development was stopped by immersing the slides in 50 ml 1% acetic acid for 45 seconds and fixed by immersing the slides in 50 ml Kodak Rapid Fix for 2.5 minutes. The slides were

rinsed under cold running water, air dried and stained (up to 20 minutes for control slides or up to 2 hours for prephotographed slides), as described.

D. Statistical Analysis

The known coordinates of prephotographed metaphase plates allowed the relocation of these chromosomes on the hybridized slides. The hybridized chromosomes were re-photographed and compared with their pre-hybridization photographs. Positive identifications were made and grain density was determined by dividing the chromosome complement into 96 approximately equal segments. The grains over all metaphase plates were totalled for each segment and submitted to statistical analysis using a X^2 test to determine deviation from a random grain distribution. Segments that scored a X^2 value > 10 were arbitrarily considered sites of significant labelling.

VIII. Nucleotide Sequencing

A. Sequencing Reaction

Nucleotide sequencing of plasmid DNA was accomplished as described by Sanger et al. (1977) and modified by Chen and Seeburg (1985). 2 ug plasmid template DNA was adjusted to 0.1N NaOH, 2mM EDTA in a 20 ul volume and incubated 5 minutes at room temperature to denature the DNA. The sample was neutralized by the addition of 3M NaOAc, pH 5, to a 1M concentration and precipitated with ethanol. The pellet was washed with 70% ethanol and resuspended in

6 ul distilled water. 30 ng (3 ul) primer/2 ug template DNA and 1 ul of 10 X Annealing Buffer (1X Annealing Buffer is 7mM Tris-HCl, pH 7.5; 7mM $MgCl_2$; 30mM NaCl; 10 mM DTT and 0.1mM EDTA) was added to a final 10 ul volume. The mixture was incubated at 37°C for at least 30 min. to allow the primer to anneal to the template DNA. Primed template was used immediately or was stored at -20°C and thawed gently at room temperature prior to use. 5U (1 ul) DNA Polymerase I (Klenow fragment) was added to the annealing mixture along with 4 ul (~40 uCi) ^{32}PdATP (>800 Ci/mmole) (NEN). 3 ul of the mixture was added separately to each of four reaction tubes containing, respectively, A mix, G mix, T mix or C mix. A mix contained 100 um each dGTP, dTTP and dCTP and 100 um ddATP. T mix contained 1.4 um dTTP, 100 uM each dCTP and dGTP and 100 um ddTTP. G mix contained 2.5 um dGTP, 100 uM each dTTP and dCTP and 100 um ddGTP. C mix contained 2.5 um dCTP, 100 uM each dGTP and dTTP and 100 uM ddCTP. The four nucleotide reaction tubes were incubated at 37°C for 15 minutes and the reactions chased by the addition of 1 ul chase solution (125 uM each dATP, dTTP, dGTP and dCTP) to each tube with incubation at 37°C for an additional 15 minutes. The reactions were stopped by the addition of 5 ul stop solution (98% [w/v] deionized formamide; 0.2% [w/v] bromphenol blue; 0.2% [w/v] xylene cyanol and 10 mM EDTA, pH 8.0) to each tube and stored on ice until use.

B. Sequencing Gel Electrophoresis

Two gel plates measuring 33 x 38 cm and 33 x 40 cm,

respectively, were cleaned with a 0.1% SDS solution, rinsed with distilled water and rubbed with 95% ethanol. The ethanol was allowed to evaporate and the gel sandwich was assembled using 0.4 mm thick plastic side and bottom spacers. Dabs of vacuum grease were applied to the junctures of the side and bottom spacers to prevent leakage. The sides and bottom of the assembled gel plates were sealed with electrical tape and large clips.

The 6% polyacrylamide/7M urea gels were made by mixing 14.5 ml 40% (w/v) acrylamide (19:1 acrylamide : N,N' methylenebisacrylamide), 10 ml 10X Tris-Borate-EDTA buffer (TBE) (1X TBE is 100mM Tris, 83 mM Boric acid, 1 mM EDTA at pH 8.3, not adjusted) and 40.5 ml distilled water. 42 g solid urea (BRL) was added and dissolved into the mixture with stirring over low heat. The gel mixture was degassed by stirring under vacuum 15-30 minutes. 2.0 ml of the gel mixture was removed, adjusted with 16 ul 10% (w/v) ammonium persulfate (BioRad) and 1 ul TEMED (NNNN'-tetramethylethylenediamine) (BioRad), vortexed and pipetted into the gel sandwich. The gel sandwich was laid at a 30° angle to the horizontal and the 2.0 ml acrylamide plug allowed to polymerize 15-30 min. The rest of the gel mixture was then adjusted with 0.8 ml 10% (w/v) ammonium persulfate and 60 ul TEMED, mixed, and quickly injected between the gel plates using a 30cc syringe. The 0.4 mm comb spacer (flat side of sharkstooth comb) (BRL) was positioned at the top of the gel. The gel sandwich was laid at a 30° angle to the horizontal, excess gel mixture was pipetted

across the comb spacer to seal the gel and the acrylamide was allowed to polymerize 30-60 minutes.

After polymerization was complete, the comb and bottom spacer were removed. The top of the gel was flushed with 1X TBE and the gel sandwich clipped into the vertical gel electrophoresis apparatus (BRL). The top and bottom electrophoresis tank reservoirs were filled with 1X TBE and air bubbles trapped at the bottom of the gel were removed using a syringe filled with 1X TBE with a bent 20 gauge needle. The sharkstooth comb was inserted and the gel was preelectrophoresed at 15-20 mA (1500-2000 V) for 15-30 minutes.

Upon completion of preelectrophoresis, the sequencing reaction samples were heated at 70°C for 3 minutes. 3-4 ul of each sample were loaded quickly into separate adjacent wells and the gel electrophoresed at 20-25 mA (2000-2500 V) until the bromphenol blue dye reached the bottom of the gel (1.5-2.0 hours). The remainder of the samples, which had been kept on ice during this time, were reheated at 70°C for 3 minutes and 3-4 ul of each sample were loaded into separate adjacent wells. Electrophoresis was resumed at 20-25 mA (2000-2500 V) until the bromphenol blue again reached the bottom of the gel (1.5-2.0 hours). The "double" loadings and electrophoresis were conducted to allow the reading of more nucleotides from the top and bottom of the gel autoradiograph. After electrophoresis, the gel sandwich was removed from the tank and laid flat. The combs, spacers, and top (short) glass were carefully removed. A large sheet of 3mm Whatman

chromatography paper was laid onto the gel surface and smoothed flat with tissues. The gel adhered to the paper and was lifted off the glass plate. The orientation of the gel (first lane) was marked with a notch and the gel was wrapped in plastic wrap and exposed directly or after drying under vacuum to Kodak X-Omat AR film at room temperature for 16-48 hours. The X-ray film was processed as previously described.

Sequences were read from the bottom to the top of the gel in both "runs" so that the nucleotide base sequences were obtained in the 5' (bottom of the gel) to the 3' (top of the gel) direction. Overlaps in the base sequence at the bottom of the first "run" and top of the second "run" provided sequence continuity.

C. Interpretation of Nucleotide Base Sequences

1. Determination of Base Composition

The base compositions of the nucleotide sequences were determined by dividing the number of dATP + dTTP and dGTP + dCTP residues by the total number of nucleotides in the sequence. The resulting values were multiplied by 100 to express the base composition of the sequence as %AT and %GC content.

2. Determination of Intersequence Homologies

Sequences were aligned and compared to detect regions of homology with computer assistance using the NUCALN program developed by D.J. Lipman and W.J. Wilbur at the Mathematical Research Branch of the NIADDK, NIH. The parameters of the program

define "k-tuple" values of 2-7 which designate the number of consecutive bases required to assign a match; "window" size values (default at 20) which define the number of bases permitted on either side of a match and "gap penalty" values (1-7) which do (gap penalty = 1) or do not (gap penalty = 7) permit the introduction of gaps in one sequence to produce best alignment with another sequence. Ranges of homology between two sequences may be obtained by varying the parameter values. For the purposes of this study, parameter values were set at k = 2, window = 20 and gap = 1 or 7 to determine maximal (at gap = 1) and minimal (at gap = 7) homologies between sequences. This allowed calculation of maximal and minimal % homologies between two sequences based upon the number of matched bases found between the two sequences (divided by the nucleotide length of the shorter of the two sequences) and the alignment of the sequences with each other.

3. Identification and Translation of Open Reading Frames

Open reading frames (orfs) in the nucleotide base sequences were identified and translated with computer assistance using the ORF and TRANSLATION programs developed at the NIH. The ORF program examines each of three possible orfs in each DNA strand (six possible frames) for the presence of translation initiation ATG (AUG) and termination TAA, TAG and TGA (UAA, UAG and UGA) codons. TRANSLATION then reads the open reading frame codons and assigns the appropriate amino acid residue to each codon.

RESULTS

PART ONE: MOLECULAR CLONING OF SATELLITE II FRACTION DNA

The following study was undertaken to reevaluate previous investigations characterizing and localizing satellite DNAs in the human genome. Prior studies of the molecular organization of human satellite DNAs noted that Satellite I,II and III molecular populations were heterogeneous. They contained somewhat complex fractions of A+T-rich DNAs that copurified within gradient-isolated Satellite DNA preparations (Mitchell et al., 1979; Prosser et al., 1981; Frommer et al., 1982). The heterogeneity of the satellite fractions themselves rendered their detailed molecular analysis difficult. The studies presented here better define the degree of sequence heterogeneity within Satellite II DNA and provide new information concerning the the molecular and genomic organization of a discrete, cloned Satellite II fragment.

I. Isolation of Satellite II-Related Sequences

Significant purification of Satellite DNA was achieved by density gradient ultracentrifugation (Corneo et al., 1970; 1971) and was used to avoid isolation of repeated sequences with procedures involving complex C_ot analysis (Deininger et al.,

1981) or genomic library surveys for repeated sequences. This method also allowed a comparison of purified satellite DNAs with those previously isolated by identical methods. Satellites I, II and III DNA, which are A+T-rich, purify as easily identifiable peaks or shoulders separated from nonsatellite "main band" genomic DNA (Corneo et al., 1970; 1971).

Initial isolation of Satellites II and III fractions used isopycnic Ag^{++}/Cs_2SO_4 buoyant density gradients. Visible Satellite II peaks and Satellite III shoulders were obtained (Figure 1A), One hundred eighteen ug DNA was isolated as Satellite II, and 53 ug as Satellite III sequences from 1.5 mg total DNA. Centrifugation of the crude Satellite II and III preparations through two successive rounds of CsCl gradients isolated the Satellite fractions as single peaks with buoyant densities equal to 1.694 g/ml and 1.697 g/ml, respectively (Figure 1B) (Corneo et al.,1970; 1971). 23 ug Satellite II and 3 ug Satellite III were obtained, corresponding to 1.5% and 0.2%, respectively, of the initial 1.5 mg DNA sample. Less Satellite III than Satellite II DNA was obtained because the Ag^{++}/Cs_2SO_4 gradient conditions used (Ag^{++}:DNA ratio >0.2) optimized purification of Satellite II rather than Satellite III sequences (Corneo et al., 1971), in keeping with the objective to study Satellite II DNA.

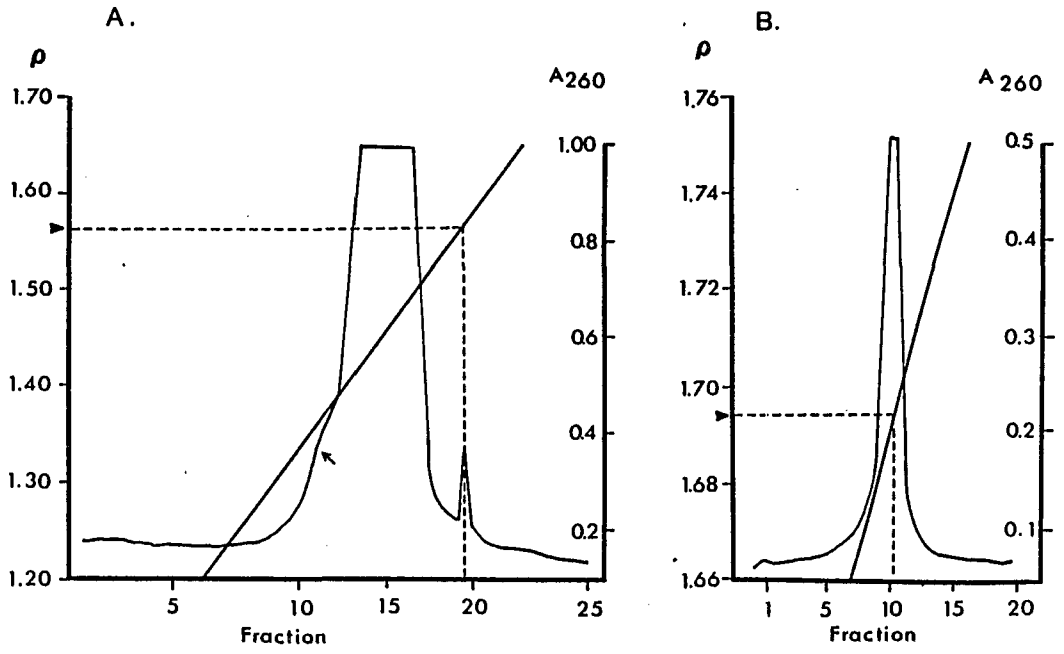
II. Molecular Cloning of Satellite II-Related Sequences

Satellite II and III fractions isolated by buoyant density

FIGURE 1. Cesium Gradient Isolation of Satellite DNA.

A. Ag^{++}/Cs_2SO_4 Ultracentrifugation. 50 ug/ml high molecular weight human male placental DNA was centrifuged in an isopycnic Cs_2SO_4 gradient at an intial density equal to 1.485 g/ml and an Ag^{++}/DNA ratio > 0.2. Twenty five fractions, at 0.2 ml each, were collected. The buoyant density and absorbance at A_{260} is shown for each fractions. Satellites I and III DNA appear as a "shoulder" at a mean density = 1.32 g/ml; "mainband" DNA spans the gradient between densities 1.39-1.50 g/ml, and Satellite II DNA appears as a unimodal peak at density = 1.56 g/ml.

B. CsCl Ultracentrifugation. Satellite II and III fractions collected from the Cs_2SO_4 gradients were dialyzed, precipitated, resuspended TE and centrifuged in isopycnic CsCl gradients at intial densities equal to 1.700 g/ml. Twenty-five fractions, at 0.2 ml each, were collected. The buoyant densities and absorbance at A_{260} is shown for each fraction. After successive ultracentrifugations under these conditions, the Satellite II and III fractions appeared as unimodal peaks at a densities equal to 1.694 g/ml and 1.697 g/ml, respectively.


A.
ρ
A260
1.70
1.60
1.50
1.40
1.30
1.20
1.00
0.8
0.6
0.4
0.2
5
10
15
20
25
Fraction
B.
ρ
A 260
1.76
1.74
1.72
1.70
1.68
1.66
0.5
0.4
0.3
0.2
0.1
1
5
10
15
20
Fraction

ultracentrifugation are known to consist of heterogeneous populations of A+T-rich sequences. The composition of these fractions can further vary between different preparative gradient isolations (Singer, 1982; Manuelides, 1978). Precise characterization of Satellite DNA fractions necessitated obtaining a constant population of DNAs in each preparative isolation and large enough quantities of DNA to permit detailed molecular analysis. These conditions were met by propagating Satellite II and III fractions as recombinant DNAs in the bacterium E. coli. Molecular cloning of the Satellite fractions allowed characterization and mapping of component DNAs, as well as comparison with other cloned Satellite fragments.

Accurate characterization of complex Satellite fractions depended upon the retention of native Satellite and associated flanking sequences in the cloned DNA. Determining sequence structure of DNA flanking Satellite repeats allowed a specific determination of the genomic organization and localization of discrete cloned Satellite elements. To ensure that flanking DNA sequences would be cloned along with Satellite sequences, Satellite DNA predicted to be large enough to include both types of sequences were cloned. Some restriction enzymes digest Satellite sequences into small fragments (HinfI, TaqI, EcoRI, HaeIII) whereas others do not appreciably digest Satellite DNA (BamHI, HindIII, XbaI) (Manuelides, 1978; Mitchell et al., 1979; Frommer et al., 1982). Preliminary Southern blot analysis using radiolabelled Satellite II

fraction DNA hybridized to HindIII-digested human placental DNA (not shown) revealed a series of fragments ranging in size from approximately 0.5 kb to 9.0 kb. Thus, genomic DNA digested with HindIII was likely to include a range of large fragments homologous to Satellite II DNA, including those large enough to possess both flanking and Satellite sequences. Further, ligation of fragments into the HindIII site of the vector plasmid pBR322 disrupts the coding region of the tetracycline gene, rendered the bacteria tetracycline sensitive, which can be used as a screening device to select bacterial colonies harboring recombinant plasmids.

Transformation of competent DH1 E. coli cells with supercoiled or cut, religated pBR322 was overall successful, resulting in $>10^6$ transformants/ug plasmid. Similar experiments utilizing Satellite II-pBR322 recombinant plasmids were not initially successful, even though agarose gel electrophoresis showed that the DNAs were ligated. One published report had cited similar difficulties in cloning Satellite III sequences, e.g., of 117 transformants obtained from ligated Satellite III-lambda gt wes lambda B DNA, only two contained inserted sequences, and one of those had undergone a putative recombination event during the cloning process (Cooke and Hindley, 1979). This suggested that Satellite DNA repeated sequences may be difficult to maintain as part of recombinant molecules in bacterial hosts. In the present study, this problem was partially overcome by increasing transformation efficiencies through the use of "supercompetent" DH5

cells (BRL). Transformation of DH5 cells with Satellite II-pBR322 yielded 103 colonies (1X plate A) and 113 colonies (1X plate B). No colonies were observed on any of the dilution plates. The transformation efficiency equalled 10^2 colonies/0.1 ug DNA/0.1 ml cells, or 1 x 10^4 colonies/ug DNA/ml cells. In a similar experiment, Satellite III-pBR322 was transformed into DH5 competent cells to an efficiency of 1 x 10^3 colonies/ug DNA/ml cells. A comparable transformation with supercoiled pBR322 gave $>10^7$ colonies/ug DNA/ml cells.

Of 216 Satellite II and 18 Satellite III colonies, 170 and 3, respectively, were tetracycline sensitive. The remaining colonies were resistant to tetracycline, and probably harbored non-recombinant plasmids.

III. Determination of Insert Sizes of Satellite II-Related Clones

Initial characterization of recombinant plasmids involved determination of basepair lengths of the cloned DNA fragments. The data obtained from electrophoretic analysis is summarized in Table I and shown graphically in Figure 2A. Of the 170 tetracycline sensitive colonies, 69 either did not harbor recombinant plasmids or possessed inserted DNAs too small (<400 bp) to be easily detected. The remaining 101 colonies possessed recombinant plasmids containing inserted fragments ranging in size from ~500-4000 bp. Ninety-seven plasmids contained single inserts, three plasmids (numbers 11, 20 and 62) possessed two inserted fragments and one

TABLE I

Sizes (in bp) of Satellite II-Related Fragments

PLASMID	INSERT SIZE(bp)	PLASMID	INSERT SIZE(bp)	PLASMID	INSERT SIZE(bp)
1	3600	59	----	116	----
2	4000	60	----	117	----
3	----	61	1000	118	----
4	3300	62	1750, 750	119	----
5	1970	63	1000	120	----
6	3000	64	1000	121	----
7	2500	65	2200	122	----
8	2200	66	----	123	----
10	3000	67	----	124	----
11	2000,1200	68	2000	125	----
12	2200	69	750	126	3300
13	2500	70	2000	127	2200
14	750	71	----	128	----
15	3600	72	2700	129	750
16	1700	73	1970,910; 550	130	3300
17	3300	74	----	131	----
18	750	75	----	132	----
19	----	76	----	133	3300
20	1970,2050	77	----	134	4000
21	----	78	1000	135	1000
22	1000	79	1400	136	----
23	----	80	2200	137	3300
24	----	81	----	138	3300
25	*500	82	----	139	4000
26	4000	83	----	140	1800
27	2200	84	----	141	----
28	----	85	1100	142	----
29	*500	86	----	143	3300
30	----	87	----	144	1800
31	*500	88	2500	145	1000
32	1400	89	----	146	750
33	1400	90	1000	147	2200
34	3000	91	----	148	----
35	2800	92	----	149	----
36	2300	93	----	150	----
37	----	94	----	151	----
38	1300	95	2200	152	2200
39	2200	96	750	153	----
40	2200	97	1000	154	2200
41	*500	98	750	155	----
42	----	99	1200	156	----
43	2200	100	1200	157	3300
44	----	101	750	158	3300
45	2800	102	3300	159	1000
46	----	103	3300	160	1000
47	750	104	----	161	750
48	----	105	2200	162	1000
49	----	106	1000	163	1000
50	----	107	1000	164	----
51	----	108	1000	165	----
52	----	109	750	166	2200
53	3000	110	1000	167	3300
54	----	111	----	168	----
55	----	112	1750	169	1000
56	----	113	1400	170	1000
57	----	114	2200		
58	----	115	3300		

----indicates that the plasmid was not recombinant or possessed inserted fragments too small (<400 bp in length) to be detected on the gels.

FIGURE 2. Distribution of Insert Fragments Sizes among the Satellite II-Related Cloned DNAs.

A. Graph of the data presented in Table I showing the number of cloned fragments per 200 bp insert size group.

B. Graph of the data presented in Table II showing the number of cloned fragments per larger insert size category. Fragment size was determined as described using electrophoresis of 0.7% agarose gels.

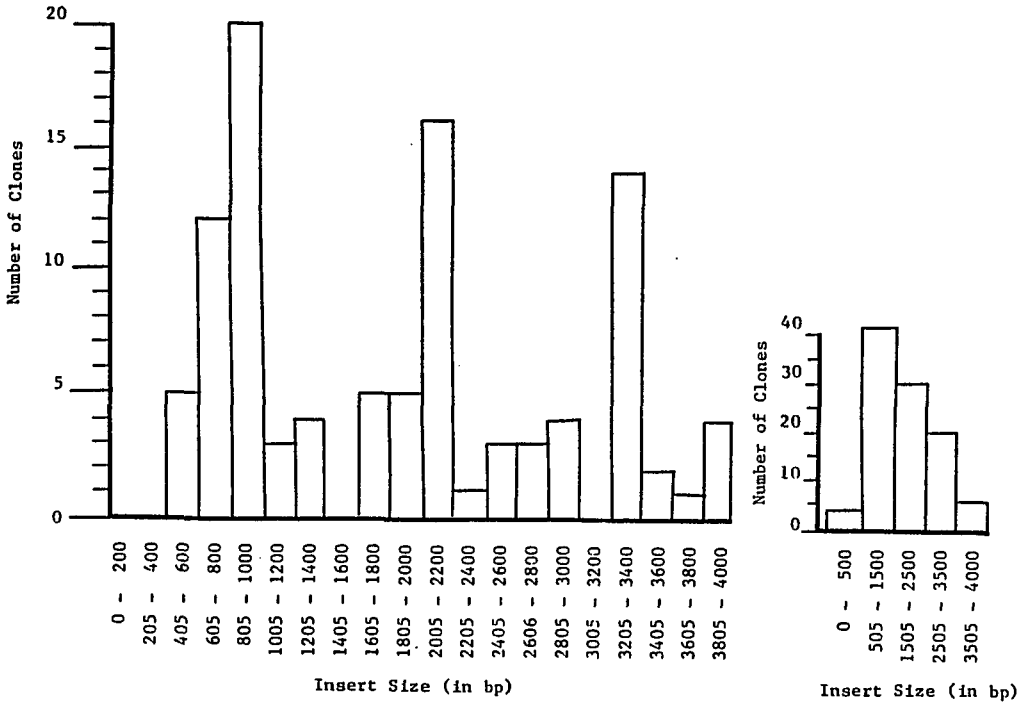


FIGURE 2. Distribution of size of inserted fragments among the Satellite II-related Clones. A. Graph of the data presented in Table I showing the number of clones per 200 bp insert size group. B. Graph of the data presented in Table II showing the number of clones per larger insert size category.

plasmid (number 73) had three inserted fragments. Thus, the actual number of HindIII inserts represented by the 101 recombinants was 106.

Cloned DNA was placed into four broad categories based on insert size (Table II and Figure 2 B). The cloned DNA fragments were organized into classifications based on size ranges of 550-1500 bp, 1550-2500 bp, 2550-3500 bp and 3550-4000 bp, and DNA for further study was selected randomly from the molecular weight categories (Table II).

The Satellite III recombinant plasmids possessed inserts of 4000 bp (B and D) and 1600 and 500 bp (C). Of these, plasmid D was chosen for further study on the basis of its large size to provide comparison of a Satellite III-derived DNA and Satellite II-derived DNA.

IV. Restriction Enzyme Mapping of Cloned Sequences

Twelve Satellite II cloned fragments and the Satellite III plasmid D were further characterized by determination of internal restriction endonuclease (RE) recognition sites. Restriction enzymes BamHI, EcoRI, HindIII, HpaI, PstI, PvuII, SalI and XbaI were used singly and in combination. Cloned fragments 33, 43 and 5 were also characterized by digestion with AluI, BclI, BglII, HaeIII, HinfI, KpnI, SstI, TaqI and XhoI for further analysis. The restriction enzyme maps determined in this manner are given in Figure 3.

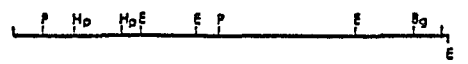
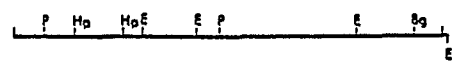
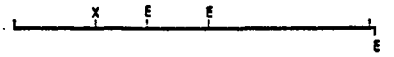
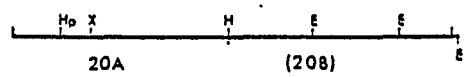
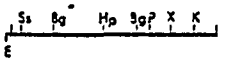
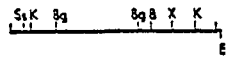
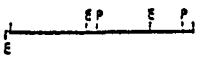
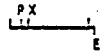
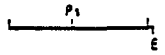
TABLE II

Summary of Initial Selection of Cloned DNA for Study

ALL CLONED DNAs Fragment Length (in bp)	Number of Fragments	Percent Occurence	Selected Plasmids	Fragment Length (in bp)
1 - 500	4	3.8%	--	----
501 - 1500	42	40.0%	#14 #85 #38 #33	750 950 1300 1400
1501 - 2500	32	30.5%	#16 #43 #5 #20A #20B #36	1700 1900 1900 1900 2050 2300
2501 - 3500	21	20.0%	#103	3300
3501 - 4000	6	5.7%	# 2	4000
TOTAL	105	100.0%	11	----

FIGURE 3. Restriction Enzyme Maps of the Cloned Fragments.

The restriction maps, determined as described in Materials and Methods, are shown for the eleven Satellite II fraction-derived (plasmids 33, 2, 103, 20 A and B, 43, 5, 16, 85, 36, 14 and 38) and one Satellite III fraction-derived (plasmid D) DNAs studied. The length, in bp, of each cloned fragment is indicated. The recognition sites of restriction endonucleases that digest the DNAs are shown. Restriction endonucleases that do not digest the cloned fragments are indicated. All fragments are defined by 5' and 3'terminal HindIII sites. The EcoRI site of pBR322 is shown to provide orientation of the fragment in the cloning vector, except for fragment 85, which could not be oriented. The position and orientation of the Alu element within the cloned 33 fragment is shown with an arrow. The DNAs are organized into four subclasses, as explained in the text. Due to covalent linkage with 20A, 20B is shown in the Subclass 1 category though it is a member of the Subclass 4 group. H, HindIII; B, BamHI; E, EcoRI; P, PvuII; Ps, PstI, Hp, HpaI; S, SalI; X, XbaI; K, KpnI; Bg, BglII; Ss, SstI, and A, AluI.

RESTRICTION ENZYME MAP	PLASMID	INSERT SIZE (bp)	LACKS SITES FOR:
SUBCLASS 1			
	33	1.400	B E Ps Hp S X K Bg Ss
	2	4,000	B Ps S X K Ss
	D	4,000	B Ps S X K Ss
	103	3,300	B P Ps Hp S - - -
	20	2,000 (A) 2,100 (B)	B P Ps S K - -
SUBCLASS 2			
	43	1,900	B E Ps S
	5	1,900	E P Ps Hp S
	16	1,700	B Ps Hp S X - - -
	85	950	B E P Ps Hp S X - - -
SUBCLASS 3			
	36	2,300	P Hp S - -
SUBCLASS 4			
	14	750	B E Ps Hp S - - -
	38	1,300	B E P Hp S X - - -

Restriction digests showed that some of the recombinant plasmids shared specific RE sites. Digestion of plasmids 2, D and 103 with HindIII + EcoRI produced a 1.2 kb fragment, and digestion with EcoRI alone generated a 1.5 kb fragment. Plasmids 2, D and 103 also possessed small (300-500 bp) EcoRI fragments, as did plasmid 16. HindIII + XbaI digestion of the DNA of plasmids 33, 20a and 103 produced a 750 bp fragment from each of the three inserted fragments. Digestion of plasmids 2, D and 43 with HindIII + PvuII produced a 1.4 kb fragment. Identical RE maps were generated for the inserted fragments of plasmids 2 and D with similar maps for 43 and 5. In contrast, the inserted fragments of plasmids 36, 38, 14 and 85 bore no obvious RE map relationship to each other or to any of the other DNA fragments.

The possession of common restriction enzyme sites among some of the cloned fragments suggested that they might also possess common base sequences. Since the cloned fragments were derived from purified Satellite fractions, it was expected that some would share base sequence homologies.

V. Southern Blot Analysis of Cloned DNA

Homologous sequences within the cloned Satellite II-related fragments were detected using Southern blot analysis. Plasmids 33, 2, D, 43, 16, 38, BLUR 8 and pE1 were each radioactively labelled with ^{32}PdATP and used as probes against transfers of gels containing RE-digested 33, 2, D, 38, 43, 16, 103, 20A and B, 5, 14, 36, 85, 4, BLUR 8 and pE1 DNAs. The BLUR 8 plasmid, which contains

an Alu SINES element (Rubin et al., 1980), was used in these studies to identify Satellite cloned fragments containing Alu elements. The pE1 plasmid, which contains four tandemly repeated alphoid monomer units (Gray et al., 1985), was included to detect the presence of alphoid sequences. The results of these experiments are summarized in Table III.

Southern blot analysis of the cloned fragments showed that they could be resolved into four subclasses, based upon sequence homologies. Subclass 1 contained homologous cloned fragments 33, 2, D, 103 and 20A and was characterized by the presence of Alu SINES element DNA. Subclass 2 comprised homologous cloned fragments 43, 5, 16 and 85 amd was characterized by the presence of L1 KpnI LINES element DNA. Subclass 3 consisted of the cloned 36 fragment containing sequences homologous to alphoid DNA. Subclass 4 consisted of nonhomologous cloned 38, 20B, and 14 sequences which were not common to any of the other cloned DNAs.

Subclass 1 cloned elements. Cloned 33 DNA was hybridized to HindIII+EcoRI-digested 33, 2, 16, 20 and 73 DNA (Figure 4 and Table III). 33 DNA displayed homology to the 1.5, 1.2 and 0.3 kb HindIII/EcoRI fragments of 2 DNA and the 2.0 kb HindIII fragment of 20A DNA. 33 DNA was also hybridized to HindIII-digested 33, 2, D, 38, 43, 16, 20A and B, 5, 14, 36 and 85 DNAs; a positive signal was obtained with the 2, D, 103, 20A and 43 cloned fragments.

DNA of plasmids 2 and D were separately hybridized to HindIII-digested 33, 2, D, 38, 43, 16, 20A and B, 5, 14, 36 and 85

TABLE III

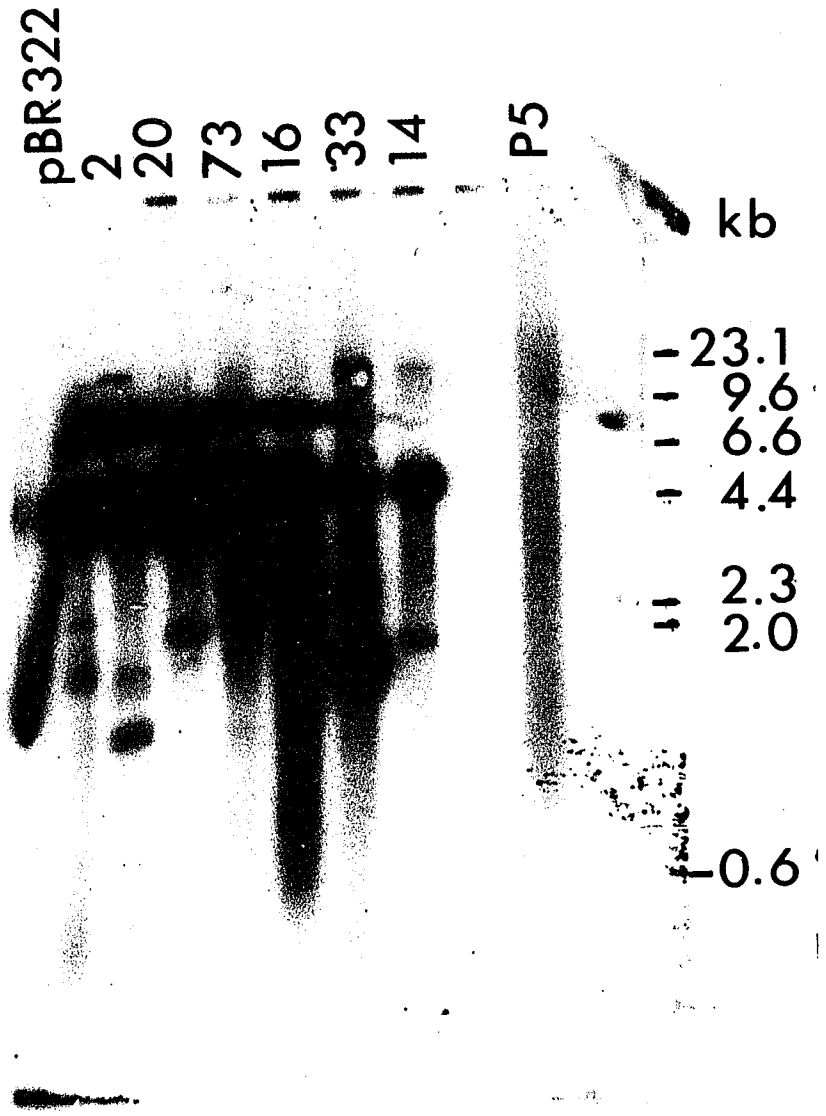
Homologies Among Satellite II-Related Clones: Southern Blot Data

	PROBES/EXPERIMENT										
CLONES	33A	33B	2	D	43	16A	16B	38	Alu	αpE1	XL3
33	+	+	+	+	+	-	-	-	+	-	+
2	+	+	+	+	-	-	-	-	+	-	+
D	+	x	+	+	-	-	-	-	+	-	+
38	x	-	-	-	-	-	-	+	-	-	-
43	x	+	-	-	+	+	o	-	-	-	-
16	-	-	-	-	+	+	+	-	-	-	-
103	x	+	+	+	-	-	-	-	+	-	+
20A	+	+	+	+	-	-	-	-	+	-	+
20B	-	-	-	-	-	-	-	-	-	-	-
5	x	-	-	-	+	+	+	+	-	-	-
14	x	-	-	-	-	-	-	-	-	-	x
36	x	-	-	-	-	-	-	-	-	+	-
85	x	-	-	-	-	+	+	-	-	-	x
4	x	x	x	x	x	x	-	x	x	x	x
Alu	x	x	x	x	x	x	x	x	+	x	-
αpE1	x	x	x	x	x	x	x	x	x	+	x
XL3	x	x	x	x	x	x	x	x	x	x	+
73A	+	x	x	x	x	x	x	x	x	x	-

LEGEND: + = positive hybridization signal; - = no hybridization signal; x = not included on gel of this experiment; o = not included on gel of this experiment due to error. Specific activities of the probes were: 33A, 1.5×10^8 cpm/ug; 33B, 1.4×10^7 cpm/ug; 2, 1.1×10^7 cpm/ug; D, 1.1×10^7 cpm/ug; 43, 8.9×10^6 cpm/ug; 16A, 2.2×10^7 cpm/ug; 16B, 2.1×10^7 cpm/ug; 38, 1.1×10^7 cpm/ug; Alu (BLUR 8), 3.7×10^7 cpm/ug; αpE1, 1.1×10^8 cpm/ug; XL3, 5.7×10^7 cpm/ug. Hybridization conditions were as described in Materials and Methods.

FIGURE 4. Southern blot analysis of cloned 33 DNA; Relationship of 33 to other Satellite II-related cloned fragments.

100 ng plasmid 33 DNA was nick-translated using α^{32}PdATP to a specific activity of 1.5 x 10^8 cpm/ug and hybridized to the Southern transfer of a 1.0% agarose gel. The gel contained 500 ng pBR322 DNA digested with HindIII (Lane 1); 500 ng each plasmid 2, 20, 73, 16, 33 and 14 DNAs digested with HindIII + EcoRI (Lanes 2-7); 2 ug human placental (P) DNA digested with HindIII (Lane 9, designated P5), and 1 ug lambda DNA digested with HindIII (Lane 10). The lengths, in kb, of the HindIII-digested lambda DNA fragments are indicated for reference as molecular weight markers. Hybridization and washing were conducted at 55-60°C as described in Materials and Methods. The autoradiogram was produced by exposure of the blot filter for 1 hour at -70°C with an intensifying screen.



CLONE 33

DNAs. Cloned 2 DNA hybridized to itself and to D, 33, 103 and 20A DNA (Figure 5 and Table III). Cloned D DNA hybridized to itself and to 2, 33, 103 and 20A DNA (Table III). 43 DNA did not hybridize to cloned 2 or D DNA. On this basis, cloned fragments 33, 2, D, 103 and 20A were placed in Subclass 1. Subclass 1 consisted of highly repetitive DNAs, since one of its members, 33 DNA, hybridized to a wide range of genomic sequences present in HindIII-digested placental DNA (see P5 in Figure 4).

The BLUR 8 probe also hybridized to 33, 2, D, 103, and 20A cloned fragments (Figure 6 and Table III). This showed that Subclass 1 cloned fragments contained Alu elements, whereas the other Satellite cloned fragments did not. The hybridization of 33 DNA to a wide range of genomic fragments was a product of homology between the Alu element in the 33 sequence and thousands of other Alu elements dispersed throughout the human genome (Figure 4) (Manuelides and Biro, 1982).

Subclass 2. Members of Subclass 2 were identified on the basis of Southern blot experiments (Table III). Cloned 43 DNA was hybridized to HindIII-digested 33, 2, D, 103, 20A and B, 43, 5, 16, 85, 38, 36, 14 (Figure 7 and Table III). DNA from the cloned 43 fragment showed homology to cloned fragments 5 and 16. The relationship between cloned 43, 5 and 16 DNAs was further investigated where plasmid 16 was used as a probe. Here, plasmid DNAs were digested with HindIII + EcoRI rather than HindIII alone. The 16 DNA fragment hybridized to 43, 5 and 85 DNA. This result was

FIGURE 5. Southern Blot Analysis of Cloned 2 DNA: Relationship of 2 DNA to other Satellite II-Related Cloned Fragments.

2 ug of plasmids pBR322, 14, 85, 38, 16, 33, 5, 20, 43, 36, 103, 2 and D; 8 ug placental DNA (Lane P5), and 2 ug lambda DNA (Lane L) digested with HindIII were electrophoresed on a 1.0% agarose gel, transferred to GeneScreen Plus (NEN) and hybridized to 400 ng α^{32}PdATP-labelled plasmid 2 DNA (specific activity = 1 x 10^7 cpm/ug). The autoradiograph indicates homology between 2, D, 103, 20A and 33 DNA. These cloned DNAs define the Subclass 1 Satellite II-related elements. Other apparent hybridization signals are artefacts, since they do not correspond to the positions of the cloned fragments on the gel. Plasmid vector homologies to pBR322 DNA appear aligned with linearized pBR322 DNA (Lane I). Plasmid 16 was not digested with HindIII and appears as two bands with homology to pBR322 DNA representing relaxed circular (upper band) and supercoiled circular (lower band) forms. Homology to a wide range of genomic fragments (Lane P5) are also apparent due to the presence of an Alu SINES element in the 2 DNA sequence. The autoradiograph was produced by exposure of the blot filter for 48 hors at -70°C with an intensifying screen.

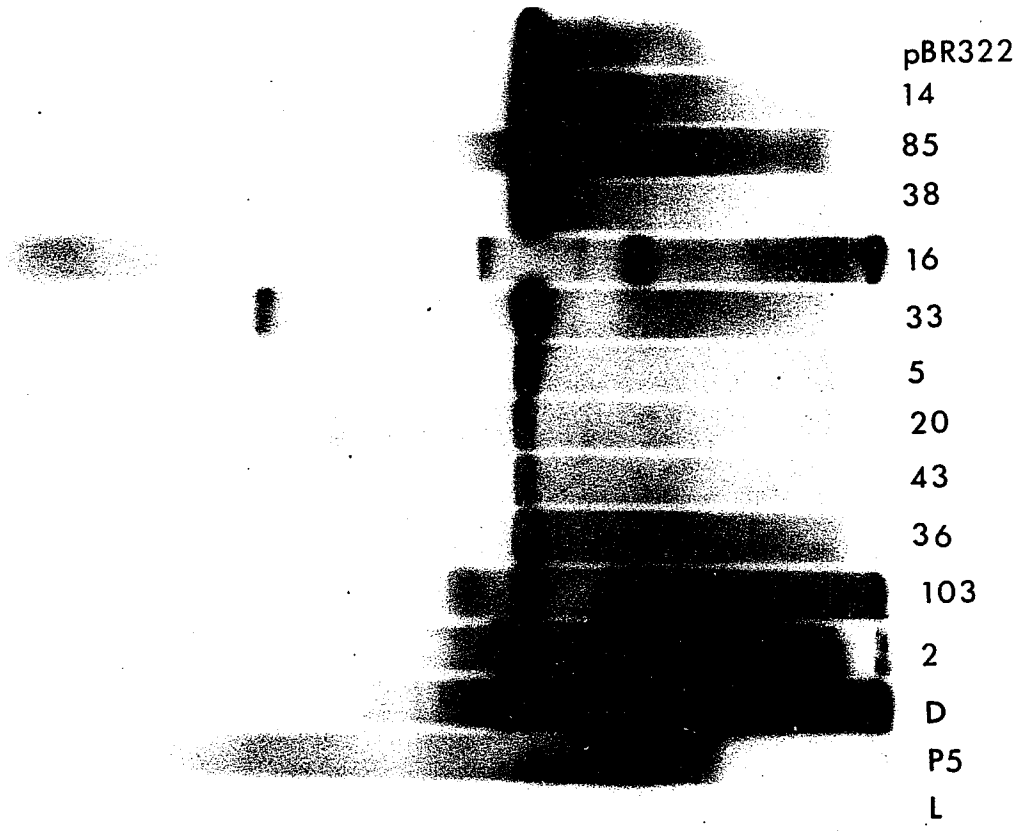

pBR322
14
85
38
16
33
5
20
43
36
103
2
D
P5
L

FIGURE 6. Southern blot analysis of Satellite II-related cloned fragments: Hybridization to BLUR 8.

200 ng BLUR 8 DNA was radiolabelled with α^{32}PdATP using random primer extension to a specific activity of 3.7 x 10^7 cpm/ug and hybridized to a 1.0 % agarose gel. The gel contained 2 ug plasmid BLUR 8 digested with BamHI (Lane 1); 2 ug each plasmids 14, 85, 38, 16, 33, 5, 20, 43, 36, 103, 2 and D digested with HindIII (Lanes 2-13); 8 ug human placental DNA digested with AluI (Lane 14), and 2 ug lambda DNA digested with HindIII (Lane 15). The lengths, in kb, of the HindIII-digested lambda DNA fragments are indicated for reference as molecular weight markers. Hybridization and washing were conducted at 55-60°C as described in Materials and Methods. The autoradiogram was produced by exposure of the blot filter for 18 hours at -70°C with an intensifying screen.

Hybridization of Blur 8 to Selected Satellite II–Related Clones

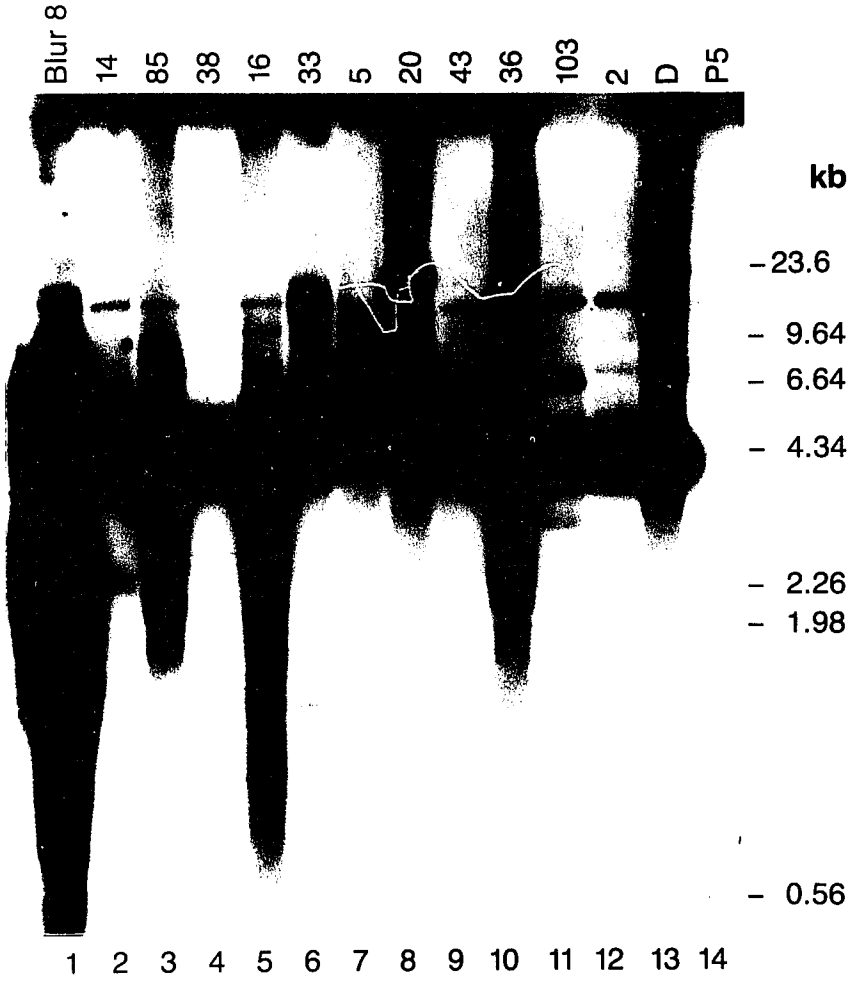
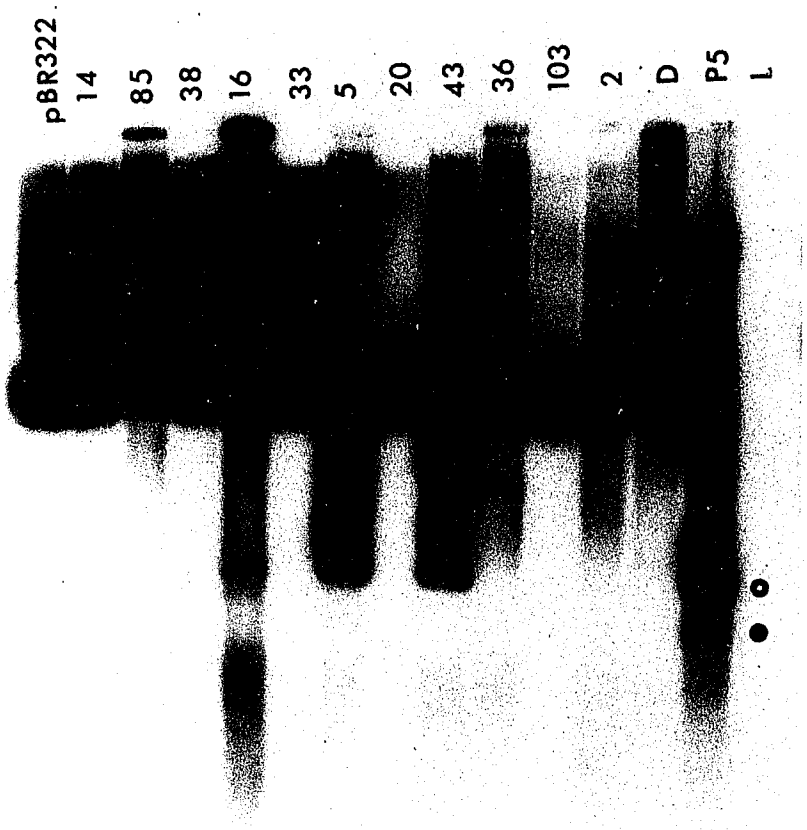


FIGURE 7. Southern Blot Analysis of Cloned 43 DNA: Relationship of 43 to other Satellite II-Related Cloned Fragments.

2 ug each plasmids pBR322, 14, 85, 38, 16, 33, 5, 20, 43, 36, 103, 2 and D; 8 ug placental DNA (Lane P5), and 2 ug lambda DNA (Lane L) digested with HindIII were electrophoresed through a 1.0% agarose gel, transferred to GeneScreen Plus (NEN) and hybridized to 400 ng α^{32}PdATP-labelled plasmid 43 DNA (specific activity = 9 x 10^8 cpm/ug), as described in Materials and Methods. The autoradiograph of the hybridized filter, shown here, indicates strong homology between the cloned 43 fragments and the cloned 5 and 16 sequences and weak homology between the cloned 43 fragment and the cloned 85 sequence. These cloned DNAs define the Subclass 2 Satellite II-related elements. Plasmid vector homologies to pBR322 DNA are apparent, as are homologies to genomic 1.9 (°) and 1.5 (•) kb fragments (Lane P5). The autoradiograph was produced by exposure of the blot filter for 24 hours at -70°C with an intensifying screen.


pBR322
14
85
38
16
33
5
20
43
36
103
2
D
P5
L

confirmed in another experiment where plasmid 16 was hybridized to a Southern transfer of the HindIII-digested cloned DNAs. It was apparent that cloned fragments 43, 5, 16 and 85 describe a second subclass among the cloned DNAs. Subclass 2 cloned fragments were homologous to each other but not to Subclass 1 cloned fragments and did not contain Alu sequences (Figure 5).

Subclass 3. The alphoid sequence probe, plasmid pE1, hybridized only to plasmid 36, defining Subclass 3 (Figure 8 and Table III). Plasmid 36 was not homologous to other Satellite II-related cloned fragments.

Subclass 4. The Subclass 4 cloned fragments - 38, 14 and 20 B DNA - were not homologous to the other cloned DNAs, and together formed a group of unrelated sequences.

The Southern blot experiments show that the Satellite II-related cloned sequences could be divided into four subclasses. Subclass 1 cloned DNAs - 33, 2, D, 103 and 20A - were homologous to each other and to the BLUR 8 probe, indicating that they possessed Alu elements, but not alphoid sequences. Subclass 2 cloned DNAs - 43, 5, 16 and 85 - were homologous to each other but not to Alu or alphoid DNA sequences. Subclass 3 cloned DNA - the 36 fragment - possessed alphoid but not Alu sequences and was not homologous to any other satellite II-related cloned DNA. Last, the Subclass 4 cloned sequences - 38, 20B and 14 - showed no homology to each other or to any of the other recombinant plasmids (Table IV).

FIGURE 8. Southern blot analysis of Satellite II-related cloned fragments: Hybridization to pE1.

400 ng plasmid pE1 was nick-translated with α^{32}PdATP to a specific activity of 1.1 x 10^8 cpm/ug and hybridized to the Southern transfer of a 1.0% agarose gel. The gel contained 2 ug plasmid pE1 digested with XbaI (Lane 1); 2 ug each plasmids 4, 85, 38, 16, 33, 5, 20, 43, 36, 103, 2, and D digested with HindIII (Lanes 2-13) and 2 ug lambda DNA digested with HindIII (Lane 14). The lengths, in kb, of the HindIII-digested lambda fragments are shown for reference as molecular weight markers. Hybridization and washing were conducted at 55-60°C as described in Materials and Methods. The autoradiogram was produced by exposure of the blot filter for 2 hours at -70°C with an intensifying screen.

Hybridization of αpEI to Selected Satellite II Related Clones

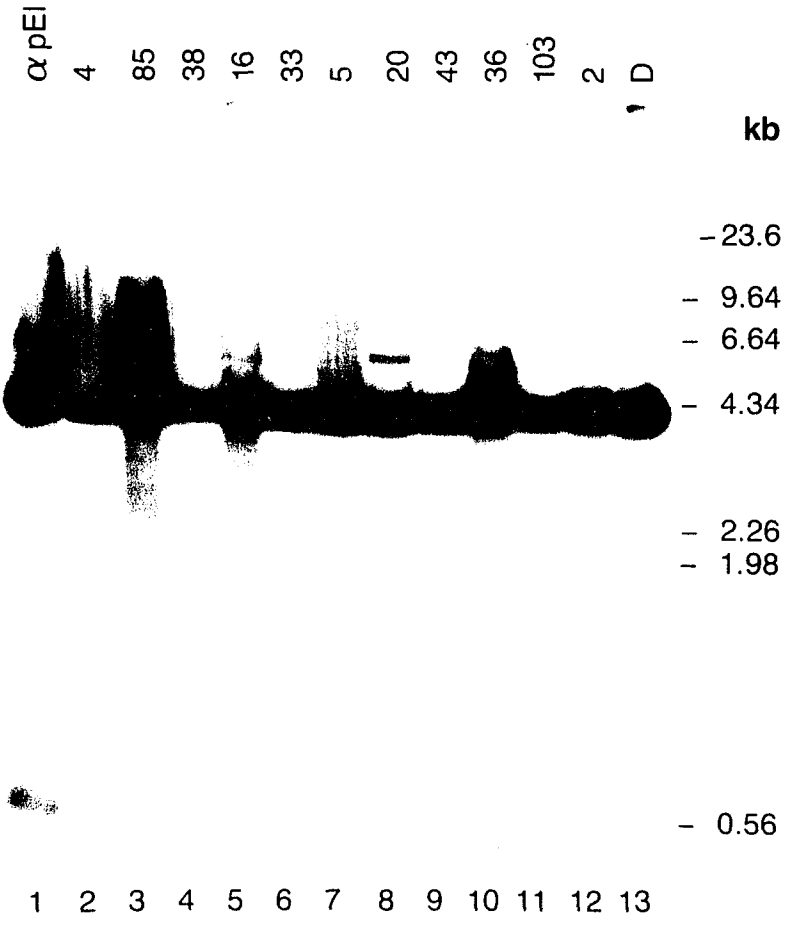


TABLE IV

Homologies Among Satellite II-Related Clones: Analysis of Southern Blot Data

	Alu	33	2	D	103	20A	XL3	20B	16	43	5	85	38	14	36	73A	apE1
Alu	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-
33	+	+	+	+	+	+	+	-	-	*	-	-	-	-	-	+	-
2	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-		-
D	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-		-
103	+	+	+	+	+		+		-	-	-		-				-
20A	+	+	+	+		+	+		-	-			-				-
XL3	-	+	+	+	+	+	+	-	-	-	-		-		-	-	-
20B	-	-	-	-			-	+	-	-			-				-
16	-	-	-	-	-	-	-	-	+	+	+	+	-	-	-		-
43	-	-	-	-	-	-	-	-	+	+	+	-	-	-	-		-
5	-	-	-	-			-		+	+	+		-				-
85	-	-	-	-					+	-		+	-				-
38	-	-	-	-	-	-	-	-	-	-	+	-	+	-	-		-
14	-	-	-	-					-	-			-	+			-
36	-	-	-	-			-		-	-			-		+		+
73A		+					-									+	
apE1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+		+

LEGEND: + = homology; - = no homology; * = inconsistent demonstration of homology. Blank areas indicate that direct demonstration of homology or lack of homology has not been experimentally shown. Hybridization conditions were as described in Materials and Methods. Data upon which Table IV is based is presented in Table III.

PART TWO: CHARACTERIZATION OF SUBCLASS 1 AND 2 CLONED FRAGMENTS

The above studies did not allow a basis for the identification of Satellite II sequences among the cloned fragments. The presence of Alu, alphoid or other sequences by no means precluded the presence of Satellite II sequences. In order to determine which recombinant plasmids, if any, contained Satellite II sequences, the genomic organization of Subclass 1 member 33 DNA and Subclass 2 member 43 DNA was determined and compared. Comparison of this information with that published for various satellite sequences would allow the provisional assignment of Satellite II sequences to cloned 33 and/or 43 DNA.

VI. Genomic Organization of the Cloned 43 and 33 DNA

Subclass 2 cloned 43 DNA and Subclass 1 cloned 33 DNA were used as probes in Southern transfers from gels containing male human placental DNA digested with various restriction enzymes. Hybridization patterns were compared with those of known repeated DNAs. Cloned 43 DNA was also hybridized *in situ* to human metaphase chromosomes. These experiments showed that the 43 cloned fragment was a variant Kpn1 L1 LINES element and did not contain Satellite sequences. It localized to the centromeric regions of human chromosome 4 and the X chromosome. The 33 cloned fragment was resolved into regions containing Satellite and Alu SINES sequences.

A. Genomic Distribution of Subclass 2 Cloned 43 DNA

1. Southern Blot Analysis of 43 DNA

Hybridization of 43 DNA to total human genomic DNA showed that it was a variant KpnI L1 LINES element. Human placental DNA was digested with the HindIII, EcoRI, BamHI, PvuII, PstI, XbaI, TaqI, HaeIII, KpnI and HpaI, which generated many fragments homologous to 43 DNA (Figures 9 and 10). The molecular weights of these fragments are summarized in Table V.

Two aspects of the genomic organization of 43 DNA were immediately apparent. The first is that the 43 sequence is a repetitive fragment in the genome, since multiple bands were generated by digestion of genomic DNA with BamHI, PvuII, XbaI and KpnI. The second is that smears of high molecular weight DNA undigested by the restriction endonucleases (with the exception of HaeIII) hybridized to 43 DNA. This type of hybridization to high molecular weight genomic DNA indicates that the probe sequence occurs genomically as an interspersed repeated sequence (Manuelides and Biro, 1982). On this basis, the cloned 43 fragment was provisionally identified as a LINES element.

Comparison of the hybridization patterns of 43 DNA to human genomic DNA with those of known LINES elements - the 1.9 kb HindIII repeat and its superfamily, the KpnI L1 LINES elements - revealed similarities (Table V) (Schmeckpeper et al., 1981; Shafit-Zagardo et al., 1982; Manuelides, 1982; Manuelides and Biro, 1982; Schindler and Rush, 1985; Hwu et al, 1986). The analysis showed

FIGURE 9. Southern blot analysis of Plasmid 43: Hybridization to Human Genomic DNA.

500 ng of plasmid 43 DNA was nick-translated using α^{32}PdATP to a specific activity of 5 x 10^7cpm/ug and hybridized to the Southern transfer of a 1.0% agarose gel. The gel contained 2 ug plasmid pBR322 digested with HindIII (Lane 1); 8 ug each human placental DNA digested with HindIII (Lane 2), EcoRI (Lane 3), BamHI (Lane 4), PvuII (Lane 5), PstI (Lane 6), XbaI (Lane 7), TaqI (Lane 8), HaeIII (Lane 9), KpnI (Lane 10) and HpaI (Lane 11), and 2 ug lambda DNA digested with HindIII (Lane 12, not shown). The lenghts, in kb, of the HindIII-digested lambda DNA fragments are indicated for reference as molecular weight markers. Hybridization and washing were conducted at 60°C as described in Materials and Methods. The autoradiogram was produced by exposure of the blot filter for 24 hours at -70°C with an intensifying screen.

Hybridization of pBsIIHIII 43 to Restricted Human Genomic DNA

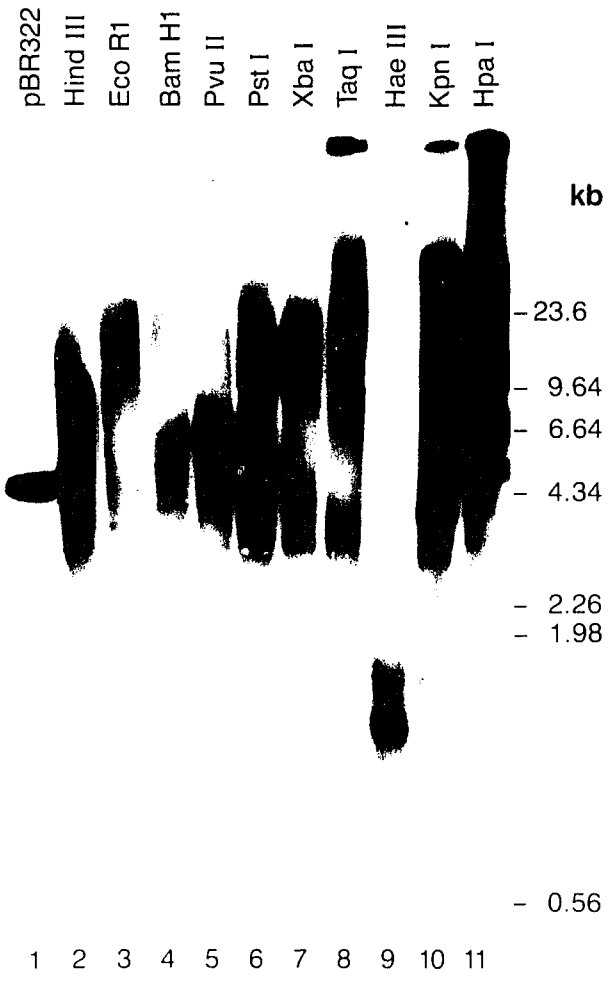


FIGURE 10. Southern Blot Analysis of Plasmid 43: Hybridization to Human Genomic DNA.

Autoradiograph of the same blot filter shown in Figure 9 except that exposure was for 96 hours rather than 72 hours at -70°C with an intensifying screen. Minor bands at 1100, 1000 and 900 bp (>) produced by digestion of genomic DNA with PvuII are evident. Bands at 1200 bp (*) produced by digestion of genomic DNA with BamHI and at 1000 bp (°) produced by digestion of genomic DNA with EcoRI are apparent.

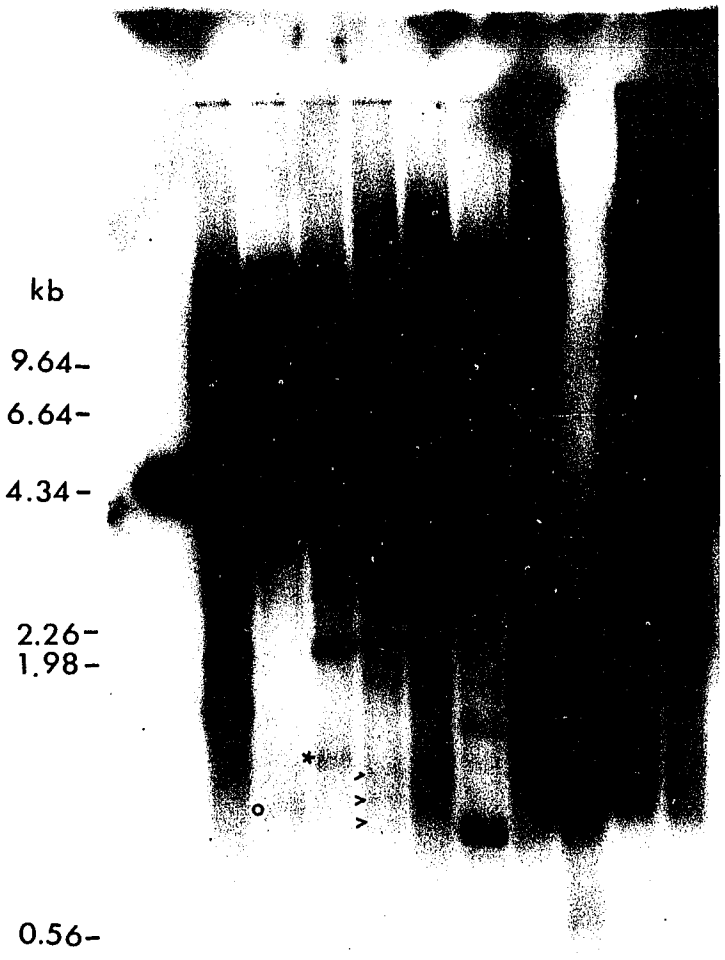

kb
9.64-
6.64-
4.34-
2.26-
1.98-
0.56-
pBR322
H E B Pv Ps X T Ha K Hp

TABLE V

Genomic Organization of Fragment 43 DNA: Southern Blot Analysis

RESTRICTION ENZYME	FRAGMENT SIZE (in bp)		CHARACTERISTIC OF: 1.9 kb HindIII family*	Kpn1 L1 LINES elements+
	MAJOR	MINOR		
HindIII	1900		+	+
	1500		-	+
EcoRI	3200		+	+
		1000	-	-
BamHI	4000		-	+
	3300		-	-
	2700		-	-
	2000		-	-
		1200	-	-
PvuII	3200		x	x
	2200		x	x
		1200	x	x
		1000	x	x
		900	x	x
PstI	--	--	--	x
XbaI	3200		+	x
	2300		+	+
	1500		-	x
	800		+	x
TaqI	--		x	x
HaeIII	1100		+	+
		500	+	+
		400	-	+
KpnI	3300		+	+
	2800		+	+
	1900		-	+
	1800		-	+
	1500		+	+
	1200		+	+
	1000		-	+
HpaI	3600		x	x
	3000		x	x
		2100	x	x
		1700	x	x

LEGEND: Major bands were visible on a 24 hour exposure of the blot; minor bands were visible on 96 but not 24 hour exposure of the blot; + = restriction fragments common to cloned 43 DNA, the 1.9 kb HindIII repeat and the KpnI LINES elements; - = restriction fragments homologus to 43 but not common to the 1.9 kb Hind III repeat and/or the Kpn1 L1 LINES elements; x = not described in the literature. Fragment sizes, in kb, were determined from the Southern blot shown in Figure 8 using $\log_{MW}$ plots, as described in Materials and Methods. *From Manuelides (1982); Manuelides and Biro (1982). +From Schindler and Rush (1985); Schmeckpeper et al., 1981; Shafit-Zagardo et al., 1982a,b.

that 43 DNA was more representative of the KpnI L1 LINES superfamily rather than the 1.9 kb HindIII repeat per se.

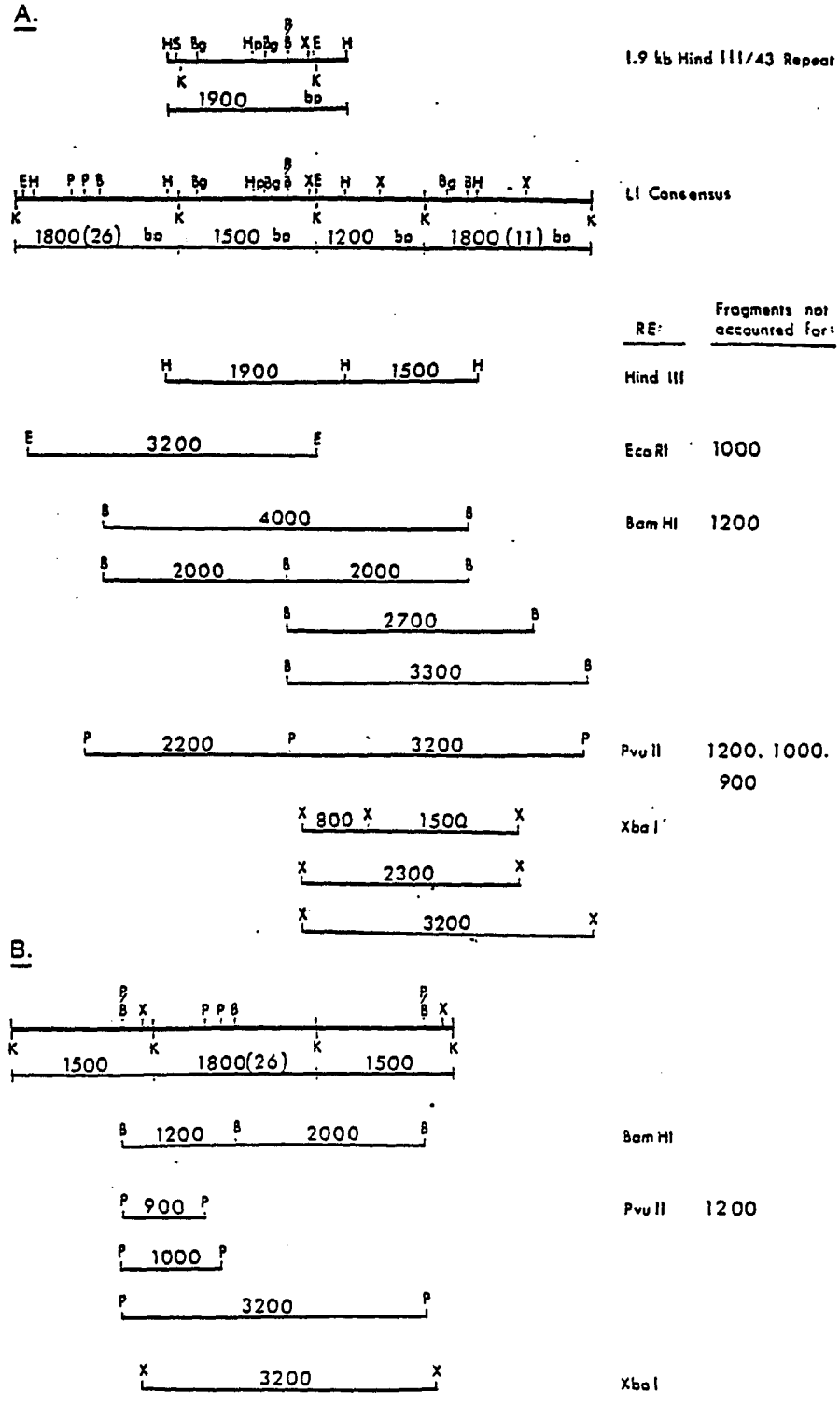
Cloned 43 DNA differed from both 1.9 kb HindIII and KpnI L1 LINES elements in that it was homologous to several BamHI-generated fragments in human DNA. Neither the 1.9 kb HindIII nor the KpnI L1 LINES repeats have been shown to hybridize to other than high molecular weight (>4 kb) BamHI fragments in the human genome (Schmeckpeper et al., 1981; Manuelides and Biro, 1982). The 43 DNA fragment, although related to the 1.9 kb Hind III family and perhaps more closely related to the KpnI L1 LINES elements, is not identical to either group of interspersed repeated sequences.

Composite RE maps of 43 DNA, the 1.9 kb HindIII repeat and the L1 KpnI LINES 6.4 kb repeat is shown in Figure 11A (Schindler and Rush, 1985). The 1.9 kb HindIII repeat overlaps 5' with the 1.8(26) kb KpnI element, encompasses the 1.5 kb KpnI element, and overlaps 3' with the 1.2 kb KpnI element (Schindler and Rush, 1985). The 43 fragment shares two BglII, one XbaI, one SstI and the 3' KpnI sites with the 1.9 kb HindIII repeat. The 5' KpnI site is present in the 1.9 kb HindIII sequence but absent in 43, as is an EcoRI site (Manuelides, 1982). The 43 fragment possesses a HpaI site ~900 bp downstream and a PvuII site ~1300 bp downstream from the 5' HindIII site (Figure 11A). The 1.9 kb HindIII repeat lacks both these sites, though single base changes at positions 867 (A>T) and 1306 (G>C) would generate HpaI and PvuII sites (Manuelides, 1982). Plasmid 5 DNA, whose RE map is almost identical to that of 43,

FIGURE 11. Relationship of the Cloned 43 Fragment to the L1 KpnI Family of Repeated Sequences.

A. Comparison of restriction endonuclease maps of 43/5 DNA with those of the 1.9 kb HindIII repeat and the 6.4 kb L1 KpnI family of repeated sequences. The first RE map shows restriction enzymes sites shared by the 43/5 cloned fragment and the 1.9 kb HindIII repeat (_____), present in the 43 fragment but not the 1.9 kb HindIII repeat (---) or present in the 1.9 kb HindIII repeat but not in the 43 fragment (···). The second RE map shows the consensus restriction enzyme sites within the 6.4 kb L1 KpnI family of repeated sequences. The 5' 1.8 kb element is designated "1.8(26)" and the 3' 1.8 kb element is designated "1.8(11)" after cloned homologues to distinguish them. The sizes, in basepair, of restriction fragments produced by digestion with the indicated enzymes are given. Fragments produced by digestion with these enzymes that cannot be derived from the 6.4 kb L1 KpnI consensus RE map are indicated. The 6.4 kb KpnI L1 consensus RE map is based on information given in Manuelides (1982); Manuelides and Biro (1982); Schmeckpeper et al., 1981; Shafit-Zagardo et al., 1982a,b, and Schindler and Rush, 1985.

B. Proposed rearrangement of the 6.4 kb L1 KpnI 5' 1.8 kb (1.8[26]) and 1.5 kb elements to produce the indicated restriction enzyme fragments. Fragments that cannot be derived from the RE maps in A. or B. are indicated and their lengths given in basepairs. B, BamHI; Bg, BglII; E, EcoRI; H, HindIII; Hp, HpaI; K, KpnI; P, PvuII; Ss, SstI, and X, XbaI.


A.
HS Bg Hpa Bg B XE H
K K
1900 bp
1.9 kb Hind III/43 Repeat
EH P P B H Bg Hpa Bg B XE H X Bg BH X
K K K K K
1800(26) bp 1500 bp 1200 bp 1800(11) bp
L1 Consensus
RE:
Fragments not accounted for:
H 1900 H 1500 H
Hind III
E 3200 E
EcoRI 1000
B 4000 B
Bam HI 1200
B 2000 B 2000 B
B 2700 B
B 3300 B
P 2200 P 3200 P
Pvu II 1200. 1000. 900
X 800 X 1500 X
Xba I
X 2300 X
X 3200 X
B.
B X P P B B X
K K K K
1500 1800(26) 1500
B 1200 B 2000 B
Bam HI
P 900 P
Pvu II 1200
P 1000 P
P 3200 P
X 3200 X
Xba I

lacks the HpaI site and possesses a BamHI site in place of the PvuII site. A BamHI site would be generated by a single base change (A>C) at position 1328 in the nucleotide sequence of the 1.9 kb HindIII repeat.

Other RE sites are consistent with the L1 map restriction enzyme site map shown in Figure 11A (Figures 9 and 10; Table V). The 1900 bp HindIII fragments originate from HindIII digestion of the 5' 1.8(26) kb and 1.5 kb KpnI elements. The large 3300 EcoRI fragment arises from digestion at the EcoRI sites of the 5' 1.8 kb and 1.5 kb KpnI repeats. The 3300, 2800, 1800, 1500 and 1200 bp KpnI fragments are generated by digestion of the entire L1 unit at the KpnI recognition sites. The 4000, 3200 and ~2000 bp BamHI fragments arise from digestion of the L1 unit at internal BamHI sites, including the BamHI site observed in the cloned 5/43 sequence. Digestion at the 5/43 BamHI site and at a proposed extreme 3' BamHI site would yield the observed 2700 bp BamHI fragment. Similarly, placement of a PvuII site at the extreme 3' end of the L1 unit, in addition to those present within the 5' 1.8(26) kb KpnI and 1.5 kb KpnI elements, would generate the observed 2200 and 3200 PvuII digestion products. The 2300, 1500 and 800 bp XbaI fragments are generated by digestion at the L1 unit XbaI sites shown in Figure 11A. Elimination of the extreme 3' XbaI site in some copies of the L1 element would result in production of the 3200 and 2300 XbaI site fragments if digestion occurred instead at XbaI sites located at the extreme 3' end of the 6.4 kb L1 unit.

Some of the RE fragments observed cannot be directly correlated with the L1 map. The 1.5 kb HindIII genomic fragment (Figures 9 and 10) shows homology to the 1.2 kb KpnI element, but not the 1.5 kb KpnI element internal to the 1.9 kb HindIII fragment (Shafit-Zagardo et al., 1982a). Homology of 43 DNA to the 1.5 kb HindIII genomic fragment arises from possession of 5' 1.2 kb KpnI sequences (Shafit-Zagardo, personal communication). The positions of the HindIII sites defining the 1.5 kb HindIII genomic fragment within the L1 consensus map are unclear.

Other 43 DNA-homologous genomic fragments observed were not expected from the L1 consensus RE map and include small 1100, 1000 and 900 bp PvuII, 1200 bp BamHI and 1000 bp EcoRI fragments (visible as minor bands on the Figure 10 blot after 96 hours of exposure). Additional BamHI, PvuII and EcoRI sites would be required within either the 1.5 or 1.2 kb KpnI elements to produce small fragments. There is no previous evidence that such sites exist, even as minor polymorphisms within cloned or genomic KpnI element DNAs (Schmeckpeper et al.,, 1981; Shafit-Zagardo et al.,, 1982; Manuelides, 1982; Manuelides and Biro, 1982; Schindler and Rush, 1985; Hwu et al., 1986). The observation that two PvuII sites and one BamHI site occur within the 5' half of the 1.8(26) KpnI element and that the cloned 43/5 DNA possesses closely spaced BamHI and PvuII sites suggests that the two sequences may be involved in the generation of the small 43 DNA-homologous BamHI and PvuII fragments. For example, juxtaposition of the 1.8(26) and 1.5 kb

KpnI element PvuII and BamHI sites would require a simple rearrangement event whereby a 1.8 (26) KpnI element is inserted at the 3' end of a 1.5 kb 43/5-like KpnI element. If the point of insertion occurred at the 5' end of the 1.8 (26) kb and 3' end of the 1.5 kb KpnI elements, the resulting alternative arrangement would resemble the one depicted in Figure 11B. Digestion of the altered arrangement with BamHI would yield the observed 1200 bp BamHI, and digestion with PvuII would generate the observed 900 and 1000 bp PvuII genomic fragments. The alternative arrangement also offers additional derivations for the 3200 BamHI, PvuII and XbaI genomic fragments homologous to 43 DNA. Thus, rearrangement of some copies of the L1 unit accounts for production of three small BamHI and PvuII fragments, as well as the 3200 bp BamHI, PvuII and XbaI fragments. Sequence "scrambling" or rearrangement events have been observed in some KpnI family members that change the relationships of these elements to each other (Shafit-Zagardo et al., 1982a, 1982b; Manuelides and Biro, 1982). Assignment of the small BamHI and PvuII fragments as minor bands indicates that the proposed rearrangement event occurred within relatively few copies of the 6.4 kb L1 unit, and does not constitute a major genomic event. It should also be noted that neither the L1 consensus map (Figure 11A) nor the proposed rearranged sequence (Figure 11B) account for the origins of the 1100 bp PvuII or 1000 bp EcoRI minor fragments.

Calculations based on the relative intensities of the 1.9 kb HindIII repeat band and titrations of cloned 43 fragment DNA probed

simultaneously with radiolabelled 43 DNA estimated the number of genomic sequences with homology to 43 DNA at 30,000-40,000 copies. This is expected based on published results of the copy number of the 1.9 kb HindIII repeat and L1 KpnI LINES elements (Schmeckpeper et al., 1981; Shafit-Zagardo et al., 1982b).

In summary, the Southern blot data shows that the cloned 43 fragment is a member of the 1.9 kb HindIII repeat family and, more generally, of the L1 KpnI LINES sequences. Therefore, the Subclass 2 members - 43, 5, 16 and 85 - comprise L1-homologous sequences that do not contain Satellite sequences. Their isolation from the cesium gradient Satellite II fraction suggests, but does not prove, that the Subclass 2 cloned fragments may be associated genomically with Satellite DNA.

2. Hybridization In Situ of the Cloned 43 Fragment to Human Metaphase Chromosomes

Hybridization in situ of 43 DNA to human metaphase chromosomes showed that it localizes primarily to centromeric regions of chromosome 4 and the X chromosome. The hybridization of 43 DNA to 60 metaphase plates taken from three individuals was analyzed. A representative hybridized metaphase plate from this experiment is shown in Appendix I. Although labelling occurred throughout the chromosome complement, (reflecting the interspersion of KpnI LINES sequences homologous to 43 DNA), high grain densities were observed at centromeric regions of chromosome 4 and the X chromosome. Statistical analysis using the X test showed that chromosome 4 and

the X chromosome were significantly labelled, with overall X^2 values of 88.5 and 21.0, respectively. The centromeric regions of chromosome 4 and the X chromosome had X^2 values of 57.3 and 59.8. More precise localization of grains to these regions was accomplished by recounting grains within smaller, identifiably banded pericentromeric areas. Of 206 grains within bands 4p11-14, 4q11,12 and 4q13, 88 (43%) localized to bands 4q11,12, demonstrating that centromeric bands 4q11,12 were the major regions of homology to the 43 DNA. Similarly, of 133 grains counted within bands Xp11, Xq11,12 and Xq13, 55 (42%) localized to bands Xq11,12, indicating that centromeric bands Xq11,12 were the major regions of homology to to 43 DNA on the X chromosome. Calculations based on grain density (Henderson 1982) showed that approximately 500 copies of 43 DNA were at the centromere of chromosome 4, accounting for 500 copies/30,000-40,000 total copies, or 1.0-2.0% of the total number of genomic sequences homologous to 43 DNA. Three hundred copies, or 0.5-1.0%, of the total number of genomic sequences homologous to 43 DNA were localized to the centromere of the X chromosome.

The localization of 43 DNA specifically to centromeric regions of chromosome 4 and the X chromosome contradicts a published report that the 1.9 kb HindIII repeat, which is homologous to 43 DNA, localized to Giemsa dark-banded regions, and not to specific centromeric or telomeric loci (Manuelides and Biro, 1982). The previous research suggests, however, that the build-up of biotin

labelling at the Giemsa dark-banded regions was not random, suggesting that L1 sequences might "cluster" at specific loci, and "scrambling" or other rearrangement events might permutate L1 sequences in such a way that chromosomal specificity was conferred to those sequences. The "clustering" of sequences homologous to 43 DNA at centromeric regions of chromosome 4 and the X chromosome is interpreted here as a specific instance of nondispersed L1 sequences. Alternatively, the clustering of L1 repeats with homology to 43 DNA at these loci could reflect specificity conferred by the permutation of some of those sequences, as suggested by the results of the Southern blot analysis of genomic DNA hybridized to the cloned 43 fragment. Although it cannot be conclusively demonstrated, the link between scrambling of some L1 sequences homologous to 43 DNA and specific localization of 43 DNA to the centromeric regions of chromosome 4 and the X chromosome is certainly reasonable.

B. Genomic Distribution of Cloned 33 DNA, a Subclass 1 Fragment

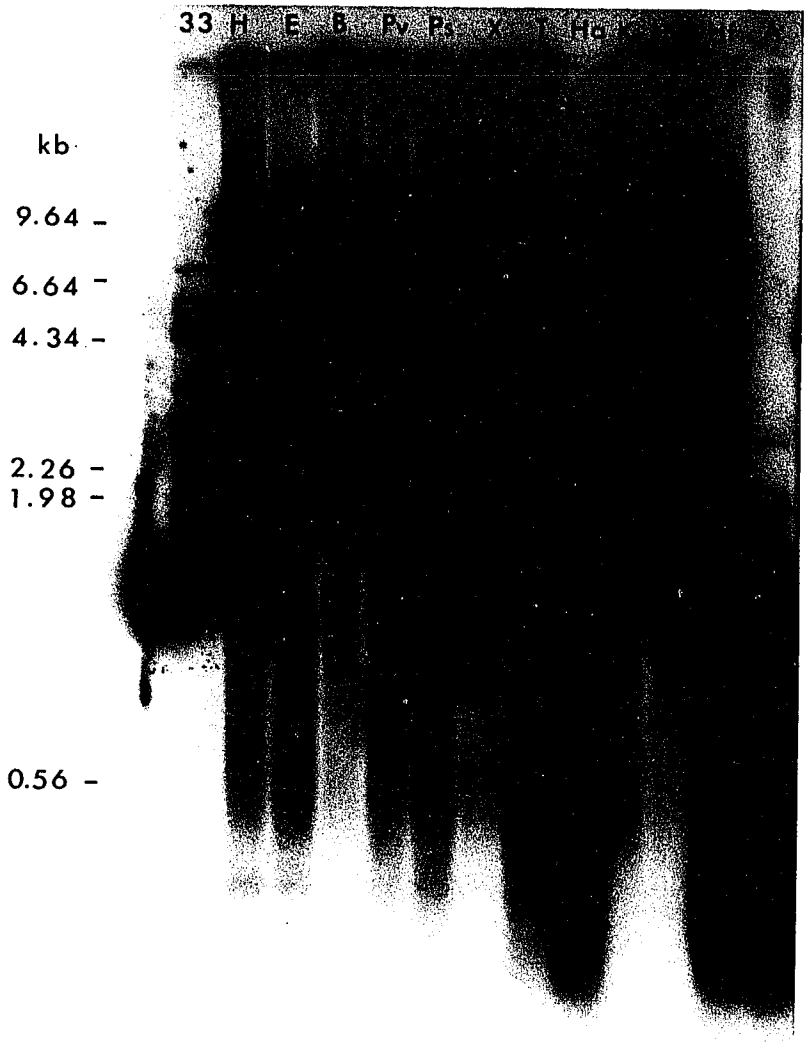
The hybridization of 33 DNA to total human genomic DNA showed that it contained Satellite sequences and redemonstrated the presence of Alu element sequences.

Radiolabelled 33 DNA was hybridized to human male placental DNA digested with HindIII, EcoRI, BamHI, PvuII, PstI, XbaI, TaqI, HaeIII, KpnI, HpaI, HinfI or AluI (Figure 12).

The vast majority of signal is due to hybridization of the Alu element with Alu SINES elements dispersed throughout the human

FIGURE 12. Southern blot analysis of Plasmid 33: Hybridization to Human Genomic DNA.

300 ng of plasmid 33 DNA was radiolabelled with α^{32}PdATP using random primer extension to a specific activity of 4.0×10^7 cpm/ug and hybridized to the Southern transfer of a 1.0% agarose gel. The gel contained 2 ug plasmid 33 digested with HindIII (Lane 1); 8 ug each human placental DNA digested with HindIII (Lane 2), EcoRI (Lane 3), BamHI (Lane 4), PvuII (Lane 5), PstI (Lane 6), XbaI (Lane 7), TaqI (Lane 8), HaeIII (Lane 9), KpnI (Lane 10), HpaI (Lane 11), HinfI (Lane 12) and AluI (Lane 13), and 2 ug lambda DNA digested with HindIII (Lane 14, not shown). The lengths, in kb, of the HindIII-digested lambda fragments are indicated for reference as molecular weight markers. Hybridization and washing were conducted at 60°C as described in Materials and Methods. The autoradiogram was produced by exposure of the blot filter for 60 hours at -70°C with an intensifying screen.



CLONE 33 GENOMIC

genome. A similar experiment utilizing labelled cloned 2 DNA as probe (not shown), which also contains Alu element sequences, generated the same intense hybridization patterns. Not all genomic fragments homologous to 33 DNA are obscured by hybridization to Alu elements. As seen in Figure 12, 340 bp fragments with homology to 33 DNA are clearly visible in both the HindIII and EcoRI-digested human genomic DNA. BamHI-generated 1.4 and 1.0 kb bands are apparent, as are 1.4 and 0.90 HpaI fragments. AluI digestion revealed clear 3.2 and 2.2 and less intense 2.1 and 1.9 kb fragments homologous to clone 33.

The homologous restriction fragments generated by digestion with HaeIII and HinfI are particularly interesting. HaeIII digestion of human genomic DNA released large 9.5 and 6.8 kb fragments. Closer inspection of the HaeIII digestion products revealed a "ladder" of bands beginning with a 9.5 kb fragment and ending at about 4 kb, where it merges with smaller heavily labelled (Alu-specific) fragments. The "laddering" effect is more apparent in the HinfI digest of genomic DNA. The HinfI ladder begins at about 9 kb and ends at about 3 kb, where it merges into heavily labelled Alu-specific hybridization signal. HaeIII and HinfI "laddering" has been previously observed in genomic DNA digested with those enzymes and hybridized to Satellite II or III DNA (Bostock et al., 1978; Mitchell et al., 1979; Frommer et al., 1982; Prosser et al., 1986). The appearance, albeit obscured, of HaeIII and HinfI laddering suggested that 33 DNA might possess Satellite

II or III sequences in association with Alu SINES elements.

C. Summary: Genomic Organization of Satellite II-Related Clones

These experiments clarified the relationships among the cloned DNAs and characterized the DNA of two cloned fragments, 43 and 33, with respect to genomic DNA.

Cloned 43 DNA was shown to be a derivative 1.9 kb HindIII repeat and KpnI L1 LINES element. The data did not support the assignment of Satellite sequences to 43 DNA. The analysis indicated that 43 DNA was homologous to both conserved and rearranged genomic L1 sequences. Hybridization *in situ* of 43 DNA to human metaphase chromosome showed that sequences homologous to 43 DNA were present throughout the chromosome complement but were concentrated at centromeric regions of chromosome 4 and the X chromosome. It was proposed that homology between the 43 fragment and permuted, rearranged genomic L1 sequences was responsible for this specific localization.

The cloned 33 DNA possessed an Alu SINES element that hybridized to thousands of similar sequences dispersed throughout the human genome, as well as non-Alu DNA. The observation of HaeIII and HinfI "ladderring" in genomic DNA homologous to 33 DNA suggested that 33 DNA possessed Satellite II sequences.

C. Localization of Alu Element Sequences in 33 DNA

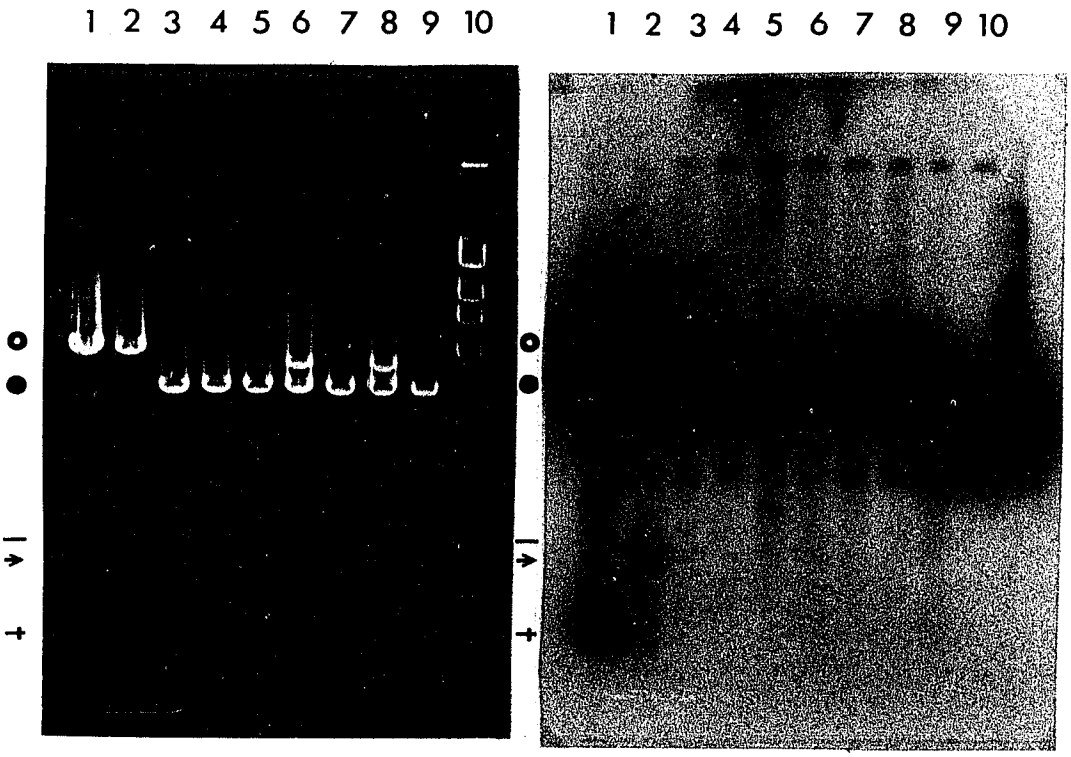
The preliminary assignment of Satellite DNA sequences to the 33 DNA warranted further investigation. Further characterization of 33 DNA first required the exact localization of internal Alu

sequences. The Alu element mapped to the 3' region of the 33 fragment between nucleotides 751-1400 (Figure 13). An AluI restriction site had also been mapped to this region, between nucleotides 1000-1100 (Figure 3). Since the 300 bp Alu element sequences are cleaved in half by digestion with AluI (Rubin et al., 1980), the location of an AluI site at bp 1000-1100 in 33 DNA suggested that the middle of the Alu element occurred there. Thus, the 300 bp Alu element spanned nucleotides 900-1200 of 33 DNA. On the basis of these experiments, it could not be determined if the purported Satellite sequences flanked the Alu element or were further upstream in the 5' region of 33 DNA. In order to specifically identify and locate the Satellite and Alu sequences within 33 DNA, the nucleotide base sequence of 33 DNA was partially determined and compared with published Satellite II, III and Alu element consensus sequences.

FIGURE 13. Southern Blot Analysis of Cloned XL3 DNA: Relationship of the XL3 Fragment to Alu Element and Cloned 33 Fragment DNAs.

A. 1 ug each plasmid BLUR 8 digested with BamHI (Lane 1); plasmid 33 digested with HindIII+XbaI (Lane 2); plasmid XL3 digested with HindIII+BamHI (Lane 3) and HindIII+XbaI (Lane 4); plasmid XL4 digested with HindIII+BamHI (Lane 5) and HindIII+XbaI (Lane 6); plasmid XL7 digested with HindIII+BamHI (Lane 7) and HindIII+XbaI (Lane 8); pSP65 digested with HindIII (Lane 9) and 1 ug lambda DNA digested with HindIII producing fragments 23.1, 9.64, 4.64, 2.26, 1.98 and 0.56 kb in length (Lane 10) were electrophoresed through a 1.0% gel, stained with ethidium bromide and photographed under UV light, as described in Materials and Methods. Plasmids XL3, XL4 and XL7 contain the 750 bp HindIII/XbaI fragment (—) of 33 DNA which spans nucleotides 1-750 of 33 DNA and is released by digestion with HindIII+BamHI or HindIII+XbaI from the multiple cloning site of the plasmid vector pSP65 (●). pBR322 DNA is indicated by (⊙), the 650 bp HindIII/XbaI fragment of 33 DNA that spans nucleotides 751-1400 of 33 DNA and contains Alu element DNA is indicated by (→) and the 300 bp Alu SINES element is indicated by (+).

B. Autoradiograph of the Southern transfer of the gel in A. after hybridization to 50 ng ^{32}PdATP-labelled BLUR 8 DNA (specific activity = 2.8 x 10^8 cpm/ug), as described in Materials and Methods. The plasmid vectors pBR322 (⊙) and pSP65 (●) and cloned Alu element (—) and 650 bp HindIII/XbaI 33 DNA fragment (→) are labelled. The 750 bp HindIII/XbaI 33 fragment, HindIII/XbaI and HindIII/BamHI 750 bp XL3, XL4 and XL7 cloned fragments are not labelled (+) and are not homologous to Alu element sequences. The autoradiograph was produced by exposure of the blot filter for 3 hours at -70°C with an intensifying screen.



A. B.

PART THREE: GENOMIC ORGANIZATION OF A DEGENERATE SATELLITE II DNA

VII. Partial Nucleotide Sequence of the Cloned 33 Fragment

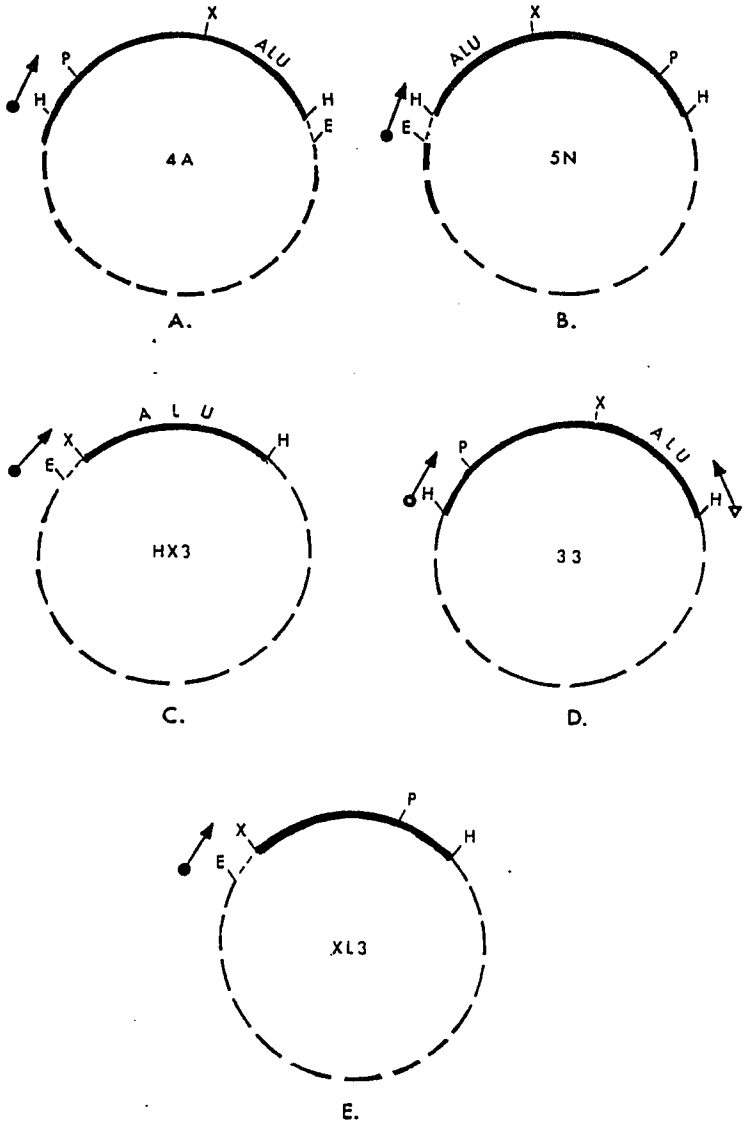
A. Preparation of the 33 DNA for Sequencing

The only portion of 33 DNA known with certainty not to contain Satellite sequences were bp 900-1200, or the location of the Alu element. Therefore, Satellite sequences could theoretically map between bp 1-900 and bp 1200-1400. Initial nucleotide sequencing determined the base sequences of portions of these areas at bp 1-201 (S1), bp 521-749 (S4), bp 750-906 (S3) and bp 1285-1400 (S2). These areas comprised the 5' end of the 33 fragment (bp 1-201), a region downstream from the 5' end (bp 521-749) and the areas flanking the Alu element (bp 750-906 and bp 1285-1400) which included the 3' end of the 33 fragment (bp 1285-1400).

Competent DH5 cells were transformed with clone 33 HindIII insert/pSP64 and clone 33 HindIII insert/pSP65 ligation mixtures to efficiencies of 2 x 10^3 colonies/ug DNA/ml cells and 2 x 10^4 colonies/ug DNA/ml cells, respectively. Several recombinants were obtained from each transformation and two of these, 4A and 5N, were chosen for sequencing. Restriction enzyme digestion of plasmid 4A showed that the inserted fragment was oriented in the pSP64 vector so that sequencing would begin at the 5'end of 33 DNA (Figure 14). Similarly, restriction enzyme digestion of plasmid 5N showed that the inserted DNA was oriented such that sequencing would begin at

FIGURE 14. Plasmid constructions used in DNA sequencing.

Each plasmid shows the approximate M13 universal (●), EcoRI cw (⊙) or HindIII ccw (▽) primer annealing/DNA synthesis initiation site, the direction of DNA synthesis (➡), the 33 fragment (heavy solid line), pBR322/pSP64/pSP65 (heavy dashed line), and the polylinker multiple cloning site (light dashed line). Recognition sites for the restriction enzymes EcoRI (E), HindIII (H), PvuII (P) and XbaI (X) and the position of the Alu element (ALU) are shown to indicate the orientation ofthe inserted fragments in the vector plasmids. A. Plasmid 4A, used to sequence the 5′ end (bp 1-201) of the 33 fragment. B. Plasmid 5N, used to sequence the 3′ end (bp 1285-1400) of the 33 fragment. C. Plasmid HX3, used to sequence the internal region (bp 750-906) of the 33 fragment. D. Plasmid 33, the original construct, used in the sequencing procedures to verify the 5′ and 3′ terminal sequences obtained used plasmids 4A and 5N. E. Plasmid XL3, used to sequence the internal region (bp 521-749) of the 33 fragment.


X
ALU
P
H
H
E
4A
A.
X
ALU
P
H
H
E
5N
B.
A L U
X
E
H
HX3
C.
X
P
ALU
H
H
33
D.
P
X
H
E
XL3
E.

the 3' end of 33 DNA (Figure 14). In addition, competent DH5 cells were transformed with a ligation mixture comprised of HindIII + XbaI-digested 33 DNA and pSP65 vector to an efficiency of 1.6 x 10^4 colonies/ug DNA/ml cells. Several recombinants were obtained and one of them, plasmid HX3, was chosen for sequencing because it possessed a small 650 bp HindIII/XbaI fragment extending from nucleotides 750-1400 of 33 DNA oriented so that sequencing would begin at the XbaI site (Figure 14). Competent DH5 cells were also transformed with a ligation mixture comprised of gel purified 750 bp HindIII/XbaI 33 DNA restriction fragment and HindIII/XbaI-digested pSP65 vector DNA to an efficiency of 1 x 10^4 colonies/ug DNA/ml cells. Many recombinants were obtained from this transformation and one of them, XL3, was chosen for sequencing because it possessed a large 750 bp HindIII/XbaI fragment extending from nucleotides 1-750 of the 33 fragment oriented so that sequencing would begin at the XbaI site (Figure 14).

B. Sequence of bp 1-201 of the 33 Fragment: Sequence S1

Nucleotide base sequencing of plasmid 4A allowed sequencing to begin at the 5' HindIII site of 33 DNA. The sequence (S1) is given in Table VI. The S1 sequence was confirmed by sequencing the 33 plasmid itself using the EcoRI primer (clockwise) specific to pBR322 (Promega). Sequence S1 comprised bp 1-201 of 33 DNA.

Inspection of S1 in Table VI shows that the entire 201 bp region has an overall A+T content of 58.7% and G+C content of

TABLE VI

CLONED 33 FRAGMENT SEQUENCES S1, S2, S3 AND S4

Subclone 4A Sequence S1: Cloned 33 Fragment bp 1-201 Sequence

1(1)	11	21	31	41	51
TTTTTCATTC	ATTTTCTACT	CTTATTGCTT	GTTGCTTAAT	ATTGGTATTC	TCTAGCATTT
61	71	81	91	101	111
ATTCCCTTCT	CATTTTCCTT	TACCTATCTC	ACCCACTCAC	CTCGCTTCAC	TACACTGCCT
121	131	141	151	161	171
GCACTCTCAA	TCAGTGTGTG	CGCTGATGTC	TTCTGTCACG	CGATGTGTAT	CTAGACGAGT
181	191	201(201)			
ACAGCTGTGC	ACTGCATGTA	T			

PvuII

Subclone XL3 Sequence S4: Cloned 33 Fragment bp 521-749 Sequence

1(521)	11	21	31	41	51
ACAATAAGAG	TAGTTCTCGT	GTTCAGGATG	TGTACGTATC	GATTACAGAG	AGATCAATGG
61	71	81	91	101	111
GTTTGAAAGG	TGGACGAAGT	ACTTGATAAA	CATACTATTG	GGACTCTCAA	GACGAAATAG
121	131	141	151	161	171
AGAAAGATAG	AGTACTTATA	AATCAGGATA	GAGATATGAC	ATATATTTAT	AAAGAGTGAA
181	191	201	211	221(749)	
GAGAAGATGA	GAGAGGTATT	ATTGTGATCG	ATTCTAGGGG	CCGTGAGTT	

Subclone HX3 Sequence S3: Cloned 33 Fragment bp 750-906 Sequence

1(750)	11	21	31	41	51
TAATTGGAAT	TCGAGCTCGC	CCGGGGATCT	AGA]polylinker		
AATGGCCATA	AAGAATGCTT	CTTCTGAGTT	ACGGACATCC	AATAGAGGAA	ACATATTATA
61	71	81	91	101	111
AACTATTACT	ATATACTAAC	GAAAATGTTC	TTATTTTGTT	AGGACGAACT	ATTAACGTGA
121	131	141	151(906)		
CGTGTACTGT	TCACTGTAGA	GCTACACTCT	CAAAAA		

Subclone 5N Sequence S2: Cloned 33 Fragment bp 1285-1400 Sequence

1(1285)	11	21	31	41	51
ATGCTTCATA	TGTCAGTGAC	GCGCCATTTC	ATCGGCTATA	TTTCGTGATA	TGATGCAGCA
61	71	81	91	101	111(1400)
				polylinker CTC	GAACCCGACG
GTATGTTGAT	GTGXAGCAGC	TCTAGTGATC	GCATGTTXCA	GTTGGTTGGA	GATCT

Overlined nucleotides represent polylinker sequence nucleotides undetectable on sequencing gel. Sequences derived from the 4A and XL3 plasmids lacked polylinker nucleotide information. The numbers in parentheses correspond to nucleotide positions within the 33 fragment. The PvuII site is underlined in the S1 sequence.

41.3%. This is in agreement with the base composition of the Satellite II consensus sequence (Prosser et al., 1986), which is 56% A+T and 44% G+C, near that of human placental DNA (Table VII).

The 5' bp 1-201 region of 33 DNA contains 17 variations on the 5'ATTCC 3' Satellite III pentameric repeat, summarized in Table VIII. Satellite II sequences have been characterized as degenerate Satellite III sequences with random base changes in the 5'ATTCC 3' pentameric repeat unit (Frommer et al., 1982; Prosser et al., 1986). Repeat 1 of the S1 sequence is a perfect 5'ATTCC 3' Satellite III pentameric sequence. Repeats 2-4 share 4 of the 5 pentameric repeat nucleotides with the Satellite III consensus sequence. Repeats 5-17 are more imperfect, sharing only 3 of the 5 pentameric nucleotides with the Satellite III sequence. Expressed in percentage form, 42% (repeats 1-17, or 85bp/201 total bp) of the S1 sequence conserves at least 3 out of 5 nucleotides of the Satellite III repeat unit; 10% (repeats 1-4, or 20 bp/201 total bp) conserves at least 4 out of 5 nucleotides, and 2% (repeat 1, or 5 bp/201 total bp) conserves all 5 nucleotides of the Satellite III repeat unit. Superimposed on the degenerate Satellite II repeats is a larger repeat, 5'TTCATTCACTT 3'. This type of repeat has not been previously described.

The degenerate 5'ATTCC 3' pentameric repeat exhibited by the S1 sequence confirmed that it was Satellite II-derived. The relationship of the S1 sequence to other Satellite II DNAs was determined by base sequence comparisons. The sequences chosen for

Table VII

Base Composition of Cloned 33 Fragment Sequences

Portion of 33 DNA	Sequenced Fragment	Content dA	dT	dG	dC	%AT	%GC
S1 bp 1-201	201 bp	36	82	29	54	58.7	41.3
S4 bp 521-749	228 bp	81	62	62	23	62.7	37.3
S3 bp 750-906	156 bp	56	48	25	27	66.7	33.3
S2 bp 1285-1400	115 bp	24	39	30	20	55.8	44.2
TOTAL	700 bp					61.0	39.0
Satellite II+						56.0	44.0
Placental DNA*						59.2	40.8

+Calculated from the base composition of the Satellite II consensus sequence (Prosser et al., 1986). *As reported in Mitchell et al., 1979. %AT for the cloned sequences was calculated as total number of A+T nucleotides divided by the total number of nucleotides in the sequenced portion of the fragment. %GC was similarly calculated using the total number of G+C nucleotides.

TABLE VIII

Satellite III Pentameric Repeat Units of the S1 Sequence

Conservation	DNA	Nucleotides	Sequence		Repeat #
5/5	Satellite III		5'ATTCC	3'	
5/5	S1	61-65	5'ATTCC	3'	1
4/5	S1	47-50	5'ATTC*	3'	2
4/5	S1	7-10	5'ATTC*	3'	3
4/5	S1	72-77	5'ATTT^{c}C	3'	4
3/5	S1	41-45	5'ATTGG	3'	5
3/5	S1	57-60	5'ATTT*	3'	6
3/5	S1	11-15	5'ATTTT	3'	7
3/5	S1	123-127	5'ACTCT	3'	8
3/5	S1	114-118	5'ACTGC	3'	9
3/5	S1	109-113	5'ACTAC	3'	10
3/5	S1	178-182	5'AGTAC	3'	11
3/5	S1	146-150	5'ATGTC	3'	12
3/5	S1	86- 90	5'ATCTC	3'	13
3/5	S1	151-154	5'*TTCT	3'	14
3/5	S1	169-172	5'AT*CT	3'	15
3/5	S1	31-35	5'GTTGC	3'	16
3/5	S1	66-70	5'CTTCT	3'	17

Conservation indicates number of nucleotides (out of five) present in pentameric repeat. *=missing nucleotide. Lower case superscript letters indicate nucleotides inserted into pentameric repeat.

comparison were: 1) three chromosome 1-specific sequences derived from a 1.77 kb cloned fragment, HS3, originally thought to be a Satellite III sequence but later found to be Satellite II-derived (Cooke and Hindley, 1979; Prosser et al., 1986); 2) a cloned Satellite sequence, pPD17 (autosome specific), (Deininger et al., 1981) shown here to be a Satellite II-derived sequence, 3) two chromosome 16-specific Satellite II sequences derived from a cloned fragment, pHuR 195 (Moyzis et al., 1987), and 4) a 26 nucleotide Satellite II consensus sequence (termed Satellite 2 in Prosser et al., 1986). The S1 sequence was also compared with other internal 33 DNA sequences, S2, S3 and S4, to determine if common homologies existed between them. The S1, S2, S3 and S4 sequences were also compared with the Alu consensus sequence (Deininger et al., 1981) in order to precisely locate and orient the Alu element within 33 DNA.

C. Comparison of the Satellite II-Derived S1, pPD17, HS3 and pHUR 195 Sequences with the Satellite II Consensus Sequence.

Prosser et al. (1986) defined the Satellite II consensus sequence as 5'ATTCC ATTCG G/AGT/ACC ATTCG ATGAT 3'. It has since been shown (Moyzis et al., 1987; this study) that a dGMP residue is the final nucleotide of the Satellite II consensus sequence. The sequence is more correctly written as 5'ATTCC ATTCG G/AGT/ACC ATTCG ATGATG 3'. Each of the Satellite II cloned sequences previously described - HS3, pPD17, pHuR 195 and S1 - in addition to the other 33 internal sequences, S2, S3 and S4, were aligned with the

26-nucleotide Satellite II consensus sequence to evaluate homologies with Satellite II DNA. The results of these alignments are shown in Tables IX (sequences S1, S2 and S2 from the cloned 33 fragment) and X (sequences pPD17, HS3 and pHuR 195).

As seen in Table IX, five complete and one partial Satellite II 26-nucleotide repeats are present in the S1 sequence, comprising 77% of the S1 sequence. A portion of the S1 sequence - nucleotides 81-107 - bear little homology to the Satellite II consensus sequence (30%), and represents a 26 basepair region of non-Satellite II sequences. The S1 sequence aligns with the Satellite II consensus sequence so that an average of 56% of the consensus residues are present per S1 Satellite II repeat block. This shows that the S1 sequence is a highly degenerate Satellite II-derived DNA. For comparison, a similar alignment was made for the HS3, pHuR195 and pPD17 DNAs (Table X). The sequences can be arranged in descending order of percent conservation of the Satellite II consensus sequence so that pPD17 (82%) > internal HS3 (77%) = 3' terminal pHuR 195 (77%) > 5' terminal pHuR 195 (76%) > 3' terminal HS3 (75%) > 5' terminal HS3 (69%) > S1 (56%). Based upon these comparisons, the S1 sequence may be classified as a highly degenerate Satellite II-derived repetitive sequence related to the pPD17, HS3 and pHuR 195 sequences.

Comparison with the Alu element consensus sequence showed that the S1 sequence is not homologous to Alu sequences, as expected from previous experiments (Figure 13). The ORF and TRANSLATE

TABLE IX

Alignment of the Partial Nucleotide Sequence of the Cloned 33 Fragment with the Satellite II Consensus Sequence

Sat.II Consensus		ATTCC	ATTCG	G/AGT/ACC	ATTCG	ATGATG
Cloned 33 Fragment						
position	%conserv.					
bp 7- 30;	61%	ATTC-	ATTTT	CTACT	CTTA-	TTGCTT
bp 31- 56;	48%	GTTGC	TTAAT	ATTGG^{t}	ATTC-	TCTAGC
bp 57- 80;	50%	ATTT-	ATTCC	CTTCT^{c}	ATTT^{t}C	CTT---
bp 81-107;	30%	tacct	atctc	accca	ctcac	ct^{c}ctt)
bp 109-139;	50%	ACTAC	ACTGC^{gtgc}	ACTCT	CAATC	A^{g}TG-TG
bp 140-168;	50%	GCGCT^{g}	ATGTC	-TTCT^{gtc}	ACGCG	ATGTGT
bp 169-198;	78%	AT-CT	AGACG	AGTAC	AGCTG^{tgc}	A^{c}TGA^{c}TG

Average Percent Conservation of Sat.II Consensus Sequences = 56%
Percent of Sequence Containing Sat.II Repeat Elements = 77%

bp 536-562;	23%	tctcg	tgttc^{agg}	atgtg	acgta	tcg---
bp 563-586;	35%	attac	agag^{a}g	atcaa	tgggt	ttg---
bp 589-617;	35%	aa^{ag}tgg	acga^{a}g^{t}	a^{c}tttg	ataaa	catact
bp 618-647;	42%	attgg^{g}	actcg^{c}	aag^{a}cg	aa^{a}tag	------
bp 648-670;	38%	atag-	agtac	ttata	aatca	-ggat-
bp 649-695;	35%	agaga^{t}	atgac	atata	tttat	aa^{a}g---
bp 696-718;	27%	agtga	agaga	agatg	agaga	ggt---
bp 719-742;	42%	attgt	tgtg-	atcg-	attct	aggggc

Average Percent Conservation of Sat.II Consensus Sequences = 35%
Percent of Sequence Containing Sat.II Repeat Elements = 35%

bp 750-769;	42%		aat	ggcca^{t}	aaaga	atgctt
bp 770-797;	58%	AAACA^{t}	AGTTA	CGGAC	ATCCA	AT^{a}GAGG
bp 798-812;	46%	atac-	taacg	aaact	attac	tat---
bp 813-838;	54%	ATAC-	TAACG	AAAAT	GTTCT	ATTTTG
bp 839-864;	23%	ttagg	acgaa	ctatt	aacgt	gacgtg
bp 865-890;	31%	actgt	tcact	gtaga	gctac	actctc

Average Percent Conservation of Sat.II Consensus Sequence = 42%
Percent of Sequence Containing Sat. II Repeat Elements = 34%

bp1288-1309;	27%	cttca	tatgt	cagtg	acgcg	cc----
bp1310-1339	69%	ATTTC	ATCGG	CTATA	TTTCG^{tgat}	ATGATG
bp1340-1358	31%	-cagc	agta-	tgttg	-----	atgtga
bp1359-1379	31%	a--gc	agctc	tagtg	atcgc	atg---
bp1380-1400	-ttnc	agttg	gttgg	agatc	ta	

Average Percent Conservation of Sat. II Consensus Sequence = 40%
Percent of Sequence Containing Sat. II Repeat Elements = 23%

Alignment of the partial nucleotide sequence of the 33 cloned fragment with the Satellite II consensus sequence (Prosser et al., 1986). The positions of the aligned sequences within the cloned 33 fragment are indicated in basepairs (bp). The percent conservation of each sequence aligned with the Sat II. consensus sequence is indicated. The average percent conservation of the aligned sequences is calculated, as is the percent of the sequence segment (S1, bp 1-201; S4, bp 521-749; S3, bp 750-906, and S2, bp 1285-1400) organized into Sat. II sequences. Nucleotides within a Sat. II consensus sequence block are shown in upper case letters; nucleotides that are not within a Sat. II consensus sequence block are shown in lower case letters. Nucleotides inserted within a Sat. II block that do not match the consensus sequence are shown in lower case letters as superscripts.

TABLE X

Alignment of the pPD17, HS3 and pHUR195 Sequences with the Satellite II Consensus Sequence

Sat.II Consensus		ATTCC	ATTCG	G/AGT/ACC	ATTCG	ATGATG
pPD17 Sequence position ;	%conserv.					
bp 3- 22;	92%	ATTCC	ATTAG	ATTCC	ATTCG	ATGATG
bp 29- 51;	81%	ATTCC	ATTCG	ATTCC	ATTTG	ATG---
bp 52- 77;	81%	ATTGC	ATTCT	ATTTC	ATTTG	ATGATG
bp 78-100;	85%	ATTCC	ATTCG	AGTCC	ACTCG	ATG---
bp 101-123;	81%	ATTCC	ATTCG	AGTCC	ATTCA	TTG---
bp 124-149;	85%	ATTCC	ATTCG	ATTTC	ATTGG	ATGATG
bp 150-173;	85%	ACTCC	ATTCG	AGTCC	ATTCG	TTG---
bp 174-198;	81%	ATTCC	ATTGG	AGTCC	ATTTG	ATTGTT
bp 201-221;	71%	ATTCG	ATTCC	ATTCG	ATTCC	T

Average Percent Conservation of Sat. II Consensus Sequence = 82%
Percent of Sequence Containing Sat. II Repeat Elements = 98%

HS3 Sequence 5' Terminal Sequence position ;	%conserv.					
bp 17- 42;	85%	ATTCC	ATTCA	ATACC	AATTG	ATGATG
bp 43- 68;	73%	GTTAT	TTTTG	ATTCC	ATTTG	ATGATG
bp 69- 94;	54%	ATTAC	ATTCC	ATTTC	ATCAT	AATTCC
bp 95-116;	58%	ATTCG	ATTCC	ACTCG[a-g]	ATTCC	------
bp 117-132;	75%	ATTCG	ATTCC	ATTCA	A	

Average Percent Conservation of Sat. II Consensus Sequence = 69%
Percent of Sequence Containing Sat. II Repeat Elements = 87%

HS3 Sequence Internal Sequence position ;	% conserv.					
bp 11- 35;	69%	AGTCC	AT-CC	ATTTC	ATTTC	ATGATA
bp 37- 62;	85%	ATTCC	ATTCG	TTTCA	ATTCG	ATGGTG
bp 63- 75;	85%	TTTCC	ATTCG	ATT		

Average Percent Conservation of Sat. II Consensus Sequence = 77%
Percent of Sequence Containing Sat. II Repeat Elements = 84%

HS3 Sequence 3' Terminal Sequence position ;	%conserv.					
bp 1- 23	88%	TTC	ATTCG	ATT-C	ATTTG	ATGATG
bp 24- 49	73%	ATTCA[t]	GCGCG	ATT-C	ATTAG	ATGATG
bp 50- 74	65%	ACCCC	-TTTC	ATTCC	ATTCA	ATGGAG
bp 75- 92	89%	[g]ATTCC	ATTCG	GTTCC	AT	

Average Percent Conservation of Sat.II Consensus Sequence = 75%
Percent of Sequence Containing Sat. II Repeat Elements = 100%

pHuR195 Sequence
5' Terminal Sequence
position ;% conserv.

bp 1- 18;	61%		TA	ATTCC	CTTTA	ATGATG
bp 19- 42;	58%	AATCC	ATTTC	ATTCA	TTTCG	GTA---
bp 43- 67;	62%	TTTCA	TTTA-	ATTCC	ATTCA	ATAC-G
bp 68- 93;	88%	TTTCC	ATTCC	ATTCC	ATTCG	ATGATG
bp 94-122;	92%	ATTCC	ATTCG^{g}	ATTCC	ATTCG^{g}	ATTAT^{a}G
bp 123-149;	81%	ATTCC	ATTTA	ATTCC	ATTCG^{g}	ATAATG
bp 150-172;	50%	ATTCC	TTAAG	TGCTT	ACCTG	ATG---
bp 173-198;	85%	AATAC	ATTGG	ATTCC	ATTCG	TTGATG
bp 199-225;	92%	ATTCC	ATTCG	GGTTC^{c}	ATTCA	ATGATG
bp 226-248;	69%	ATTCC	ATTCA	ATTCC	TTTCA	TTG---
bp 249-278;	81%	TTTCC	ATTCA	ATTCC	ATTAC^{gg}	AT^{t}GAT^{t}G
bp 279-304;	73%	ATTCC	ATTTG	ATTTG	ATTCG	ACAGTG
bp 305-329;	81%	A-GCC	ATTCG	ATTCA	ATTCC	ATGATG
bp 330-345;	38%	ATTCA	ATTGG	TTGAC	A-----	------
bp 346-371;	92%	TTTCC	ATTCG	ATTCC	ATTCG	ATGATG
bp 372-376;		ATTCC				

pHuR 195 Sequence
3' Terminal Sequence
position ;% conserv.

bp 1- 21;	67%	-----	CTTTG	AATCC	GTTCT	CAGATG
bp 22- 44;	73%	ATTCC	ATTCT	AGCCC	ATTTA	ATG---
bp 45- 70;	85%	AATCC	ATTCG	ATTCT	ATTCA	ATGATG
bp 71- 93;	65%	ATTCC	ATTCA	TGTCG	ATTTG	GTT---
bp 94-119;	88%	ATTCC	ATTTG	ACTTC	ATTTG	ATGATG
bp 120-142;	73%	ATTCC	TTTCA	AGTTC	ATTAG	ATG---
bp 143-167;	85%	ATTCC	ATTCG	ATTCC	ATTTG	ATGAT

Average Percent Conservation of Sat. II Consensus Sequence = 75%
Percent of Sequence Containing Sat. II Repeat Elements = 99%

Alignment of Sequences pPD17 (Deininger et al., 1981), HS3 (Cooke and Hindley, 1979) and pHUR 195 (Moyzis et al, 1987) with the Sat. II consensus sequence (Prosser et al., 1986). The percent conservation of each sequence aligned with the Sat. II consensus sequence is indicated. The average percent conservation of the aligned sequences is calculated, as is the percent of the sequence segment organized into Sat II. sequences. Nucleotides within a Sat. II consensus sequence block are shown in upper case letters; nucleotides inserted within a Sat. II consensus sequence block are shown in lower case letters as superscripts. Dashes indicate missing nucleotides.

programs did not reveal the presence of open reading frames within the S1 sequence. Based upon this information, the 5' terminus of the 33 fragment is comprised of degenerate Satellite II-related sequences unrelated to other repetitive or coding sequences.

D. Sequence of bp 521-749 of the 33 Fragment: Sequence S4

Plasmid XL3, possessing the large 750 bp HindIII/XbaI segment of the 33 fragment, is shown in Figure 14. Two hundred twenty eight nucleotides of subclone XL3 DNA originating from the XbaI site were sequenced by Ms. Kathi McDowell and are shown as sequence S4 in Table VI. The S4 sequence comprised bp 521-749 of the 33 fragment. The base composition of this fragment, calculated in Table VIII, was 62.7% A+T and 37.3% G+C, showed that it is somewhat A+T-rich.

Alignment of the S4 sequence with the Satellite II consensus sequence (Table XI) showed that it did not contain any Satellite II consensus sequence repeats and shared an average of only 35% sequence homology with the Satellite II consensus sequence. This indicated that bp 521-749 of 33 DNA was not Satellite II-derived. Some 5'TTCATTCACTT 3' repeats are superimposed upon the S4 sequence. Alignment of the S4 sequence with the Alu consensus sequence (Deininger et al., 1981) using the NUCALN computer program did not reveal any areas of extensive homology. No open reading frames were found within the S4 sequence using the ORF and TRANSLATION computer programs. This analysis showed that bp 521-749 of the 33 fragment was a noncoding sequence that was not derived from either Satellite II or Alu element DNA.

E. Sequence of bp 750-906 of the 33 Fragment: Sequence S3

Plasmid HX3 contained a small 650 bp HindIII/XbaI segment of 33 DNA (Figure 14). One hundred fifty six nucleotides (bp 750-906 of 33 DNA) of subclone HX3 originating from the XbaI site were sequenced (S3 in Table VI). The 156 bp S3 sequence is somewhat A+T-rich in that its base composition is 66.7% A+T and 33.3 % G+C (Table VII).

Alignment of the S3 sequence with the Satellite II consensus sequence (Table IX) showed that it contained only two complete 26-nucleotide repeats that conserved 58% and 54%, respectively, of the consensus sequence residues. The S3 sequence aligned overall with the Satellite II consensus sequence with an average 42% homology, and could not be classified as a Satellite II-derived sequence. Alignment with the Alu consensus sequence suggested that the terminal 3'AAAAA region of the S3 sequence probably represent part of the dAMP-rich region at the 3' end of the Alu consensus sequence (Deininger et al., 1981; Schmid and Jelinek, 1982). If so, the Alu element within 33 DNA is oriented in an "inverted" position, that is, the Alu element is reversed (3'-->5') in orientation with respect to the orientation of the rest of the 33 DNA sequence (Figure 3). Alu sequence inversion is not unusual in the human genome (Schmid and Jelinek, 1982; Tsukada et al., 1982; Lehrman et al., 1985). The 3' end of the Alu element occurs at approximately bp 900 of 33 DNA. This position is in agreement with Southern blot (Figure 13) and restriction map (Figure 3) data which

mapped the Alu element between approximately bp 900-1200. The 3' end of the Alu element is at ~bp 900 and the 5' end is at ~bp 1200.

The ORF and TRANSLATION computer programs identified a short open reading frame between bp 2-61 of the S3 sequence (bp 752-811 of the 33 fragment sequence). The 19-residue peptide sequence was translated as:

MET-Ala-Ile-Lys-Asn-Ala-Ser-Ser-Glu-Leu-Arg-Thr-Ser-Asn-Arg-Gly-Asn-Ile-Leu-TER,

where MET = ATG (AUG) and TER = TAA (UAA). A second termination codon, TAA (UAA), occurs further downstream in the S3 sequence between bp 79-81 (bp 829-831 of the 33 fragment sequence).

Inspection of sequences upstream (within the S4 sequence comprising bp 521-749 of the 33 fragment) of the S3 ORF showed that a transcription initiation "TATA" element existed between bp 718-723 of the 33 fragment, 30 bp upstream of the translation start site (Corden et al., 1980). No variants of the "CAATT"-box were observed between 40 and 110 bp upstream of the translation start site (McKnight and Kingsbury, 1982). The poly-A+ addition sequence recognition element, "AATAAA", was not evident (Proudfoot and Brownlee, 1976). This suggests that transcription and translation of the putative S3 ORF does not occur in vivo.

F. Sequence of bp 1285-1400 of the 33 Fragment: Sequence S2

The initial 115 nucleotides of the 3' end of the 33 fragment

were sequenced within plasmid 5N. The sequence, S2, is presented in Table VI and its base composition is calculated in Table VII. The base composition of the sequence S2 is 55.8% A+T and 44.2% G+C. These values are near those for human placental DNA, indicating that the 3' terminus is not particularly A+T- or G+C-rich.

The S2 sequence was established over five sequencing trials using subclone 5N and one trial using the cloned 33 fragment itself in conjunction with a HindIII (counterclockwise) pBR322 primer (Promega). Together, the six experiments produced one short clear sequence comprising nucleotides 1285-1400 of 33 DNA. It is not known why the 3' end of the 33 fragment was difficult to sequence, since inspection of the sequence itself does not reveal the presence of unusual Satellite, inverted repeat, or other repeated sequences that may have interfered with sequencing (Chen and Seeburg, 1985).

Alignment of the S2 sequence with the Satellite II consensus sequence showed that it contained one Satellite II 26-nucleotide repeat that conserved 69% of the consensus sequence residues (Table IX). The rest of the sequence did not contain Satellite II-type repeats, and the alignment of S2 with the Satellite II consensus sequence showed that only an average 40% of the consensus residues were present. Alignment of the S2 sequence with the Alu consensus sequence (Deininger et al., 1981) using the NUCALN computer program did not reveal any areas of extensive homology. No open reading frames were found within the S2 sequence using the ORF and

TRANSLATION computer programs.

G. Correlation of 33 Fragment RE Sites with Sequencing Data

Other RE sites were present in addition to those observed in restriction digests of 33 DNA. Based on the sequence of 33 DNA, AluI sites occurred at bp 183-186 (within the mapped PvuII site) and bp 889-893, in addition to the site previously mapped at approximately bp 1100. A HaeIII site occurred at bp 740-743 in addition to the one previously mapped at approximately bp 450. A HinfI site occurred at bp 623-627 in addition to the site previously mapped at approximately bp 550. RE sites not previously mapped in 33 DNA but evident from the sequence data included a TaqI site at bp 560-563, a BglII site at bp 1394-1400 and a XbaI site at bp 171-176. HinfI, TaqI, HaeIII, HincI and HpaI sites that occur within the Alu element consensus sequence may also be present in the 33 DNA Alu element.

H. Summary: Partial Nucleotide Base Sequence of 33 DNA

The nucleotide base sequences for four portions of the 33 fragment were determined. The first sequence, S1, comprised bp 1-201 of 33 DNA. The S1 sequence imperfectly conserves the 26-nucleotide Satellite II consensus sequence, permitting classification of the S1 sequence as a highly degenerate Satellite II-derived repeat element. No homology to the Alu element consensus sequence or evidence of coding regions were found within the S1 sequence. The second sequence, S4, comprised bp 521-749 of the 33 fragment. The S4 sequence was a noncoding region which was not

derived from Satellite II or Alu element DNA. The third sequence, S3, comprised bp 750-905 of the 33 fragment. The S3 sequence tentatively localized the 3' end of the Alu element to the S3 5' end, suggesting that the Alu element exists in an inverted orientation within 33 DNA. S3 possessed a short orf coding for a 19-residue peptide. Actual transcription and translation of the S3 orf, however, cannot be demonstrated due to the lack of transcriptional control elements in the upstream and downstream sequences. The fourth sequence, S2, comprised bp 1285-1400 of the 33 fragment. S2 showed little homology to the Satellite II or Alu consensus sequences and did not possess any open reading frames. It is a non-coding, non-repetitive sequence.

VIII. Genomic Organization of the Satellite II Component of 33 DNA.

The results of Southern blot experiments tentatively assigned Satellite sequences to the 33 fragment. This assignment was confirmed by identification of a degenerate Satellite II element within bp 1-201 of the 33 sequence. The degenerate Satellite II element was present in the other Subclass 1 cloned fragments and its genomic organization was determined using plasmid XL3, containing bp 1-750 of 33 DNA.

A. Relationship of XL3 to Satellite II-Related Cloned Fragments

Southern blot analysis showed that subclass 1 cloned 2, D, 103 and 20A fragments shared homology with the 33 Satellite II subcloned XL3 fragment. This suggested that the Subclass 1 cloned

DNAs also possessed Satellite sequences in association with an Alu element.

Radiolabelled XL3 DNA was hybridized to BLUR 8 and other DNAs cloned from Satellite II and Satellite II fractions (Figure 15). The subcloned XL3 fragment demonstrated homology with Subclass i cloned 33, 2, 103 and 20A DNAs (Figure 16, A and B). Specifically, XL3 DNA hybridized to the 1500 bp EcoRI and 1200 bp HindIII/EcoRI fragments of plasmid 2; the 1500 and 1200 bp HindIII/EcoRI fragments of plasmid 103 and the 1200 bp HindIII/XbaI fragment of plasmid 20A. These sequences are indicated by the arrows in Figure 16, A and B. Homology to plasmid D is probable, but obscured due to excessive nonspecific signal in that area of the blot. The relationship of the XL3 fragment ot the Satellite II-related clones is summarized in Tables III and IV. In a similar experiment, radiolabelled Alu element-containing BLUR 8 DNA was hybridized to an identical gel (Figures 15 and 16). It is evident from that BLUR 8 DNA hybridized to the 1500 bp EcoRI and 1200 bp HindIII/EcoRI of plasmid 2, the 1500 bp and 1200 bp HindIII/EcoRI fragments of plasmid 103 and the 1200 bp HindIII/XbaI fragment of 20A DNA (indicated by the arrows in Figure 16 C). Hybridization of BLUR 8 DNA to 33 DNA, but not to XL3 DNA, was evident, but is not shown due to excessive signal on this half of the blot. The BLUR 8 homologies are summarized in Tables III and IV.

Comparison of the hybridization of XL3 and BLUR 8 DNA to panels of Satellite II-related cloned fragments showed that the

FIGURE 15. Panel of Satellite II-Related and Other Cloned DNAs.

The gel shown is the one used in the Southern analysis shown in Figure 16 A. and B. The DNA samples were digested with restriction enzyme and electrophoresed through a 1.0% agarose gel. The gel was stained, visualized and photographed as described in Materials and Methods. The gel contained 2 ug each of plasmid XL3 digested with HindIII (Lane 1), BLUR 8 digested with BamHI (Lane 2), 33 digested with HindIII (Lane 3), 73 digested with HindIII + XbaI (Lane 4), 33 digested with HindIII + XbaI (Lane 5), 5 digested with HindIII + XbaI (Lane 6), 20 digested with HindIII + XbaI (Lane 7), 43 digested with HindIII + XbaI (Lane 8), 36 digested with HindIII + EcoRI (Lane 9), 103 digested with HindIII + EcoRI (Lane 10), 2 digested with HindIII + EcoRI (Lane 11), D digested with HindIII + EcoRI (Lane 12), C digested with HindIII (Lane 13), 16 digested with HindIII + EcoRI (Lane 14) and lambda DNA digested with HindIII (Lane 15). The lengths of the fragments produced by digestion of lambda DNA with HindIII are 23.0, 9.64, 6.64, 4.34, 2.26, 1.98 and 0.56 (barely visible) kb and were used for reference as molecular weight markers.

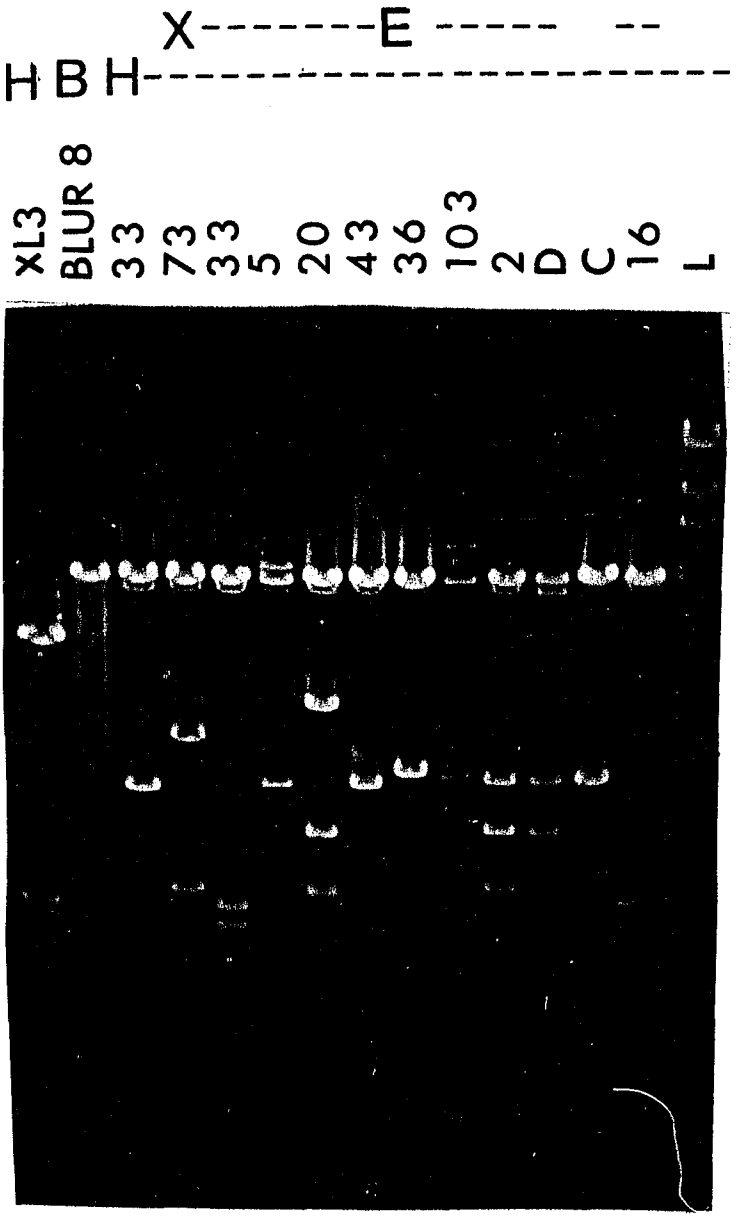
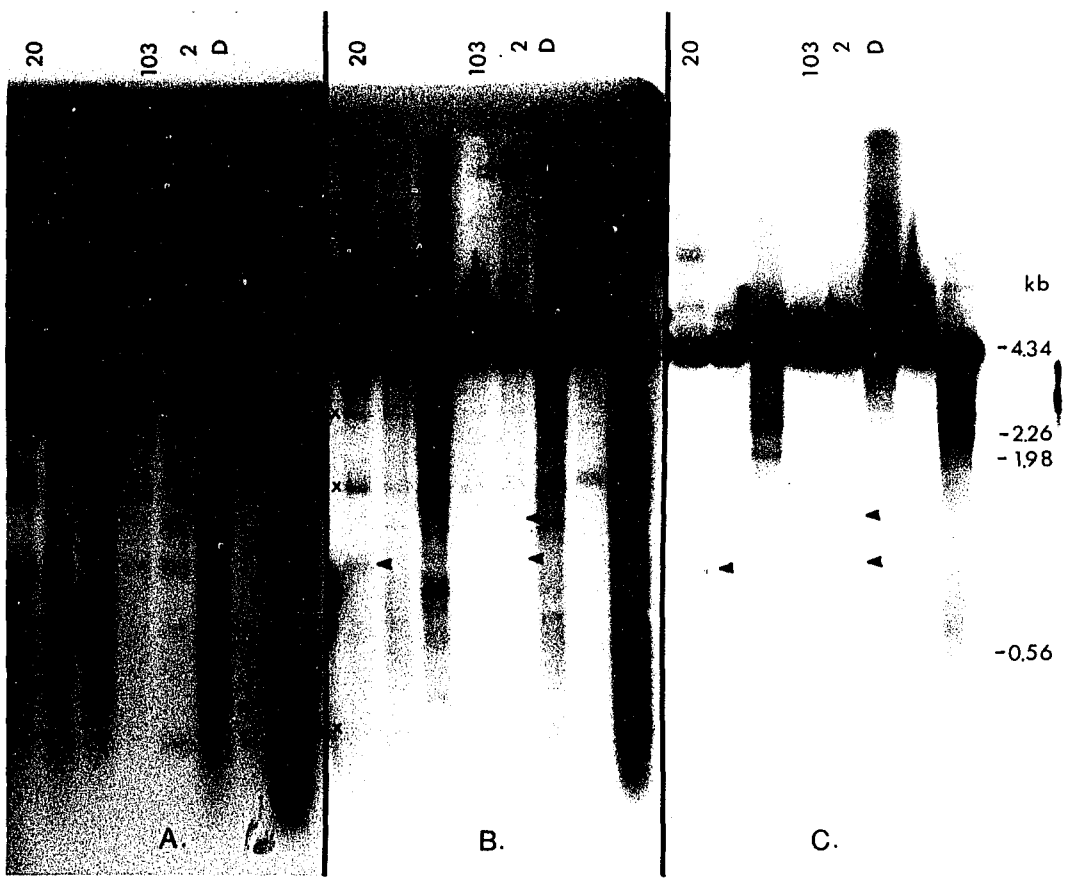

X-------E ----- --
H B H----------
XL3
BLUR 8
33
73
33
5
20
43
36
103
2
D
C
16
L

FIGURE 16. Southern Analysis: Homology of XL3 and BLUR 8 to Restriction Fragments of Satellite II-Related DNAs.

100 ng each XL3 and BLUR 8 plasmids were linearized with HindIII and radiolabelled with α^{32}PdATP using random primer extension to specific activities of 6.0×10^7 and 4×10^8 cpm/ug, respectively. The XL3 probe was hybridized to the Southern transfer of the 1.0% agarose gel shown in Figure 15.

A. 72 hour and B. 36 hour exposures, respectively, of the XL3-hybridized blot. The BLUR 8 probe was hybridized to the Southern transfer of a 1.0% agarose gel identical to that shown in Figure 15.

C. Four hour exposure of the BLUR 8-hybridized blot. Arrows indicate positions of the 1200 bp HindIII/XbaI fragment of 20A; the 1200 and 1500 bp HindIII/EcoRI fragments of 103, and the 1200 bp HindIII/EcoRI and 1500 bp EcoRI fragments of 2. The x's indicate artefacts and do not correspond to any DNA fragments shown on the gel in Figure 14. Hybridization and washing were at 55-60°C as described in Materials and Methods. The autoradiograms were produced by exposure of the blot filters for the indicated times at -70°C with an intensifying screen.


20
103
2
D
20
103
2
D
20
103
2
D
kb
-4.34
-2.26
-1.98
-0.56
A.
B.
C.

degree of homology between XL3 DNA and the Subclass 1 cloned DNAs is not high. BLUR 8-Subclass I homologies were evident after a 4 hour exposure of the Southern blot, but XL3-Subclass 1 homologies were established with 36 and 72 hour exposures. Differences in probe specific activities or probe size did not account for differences in exposure times required to visualize hybridization to Subclass 1 members. These experiments indicate that some homology exists between XL3 sequences and specific restriction fragments of Subclass 1 DNAs 2, 103 and 20A, but the homology is greater to BLUR 8 Alu element DNA. XL3 DNA homology to 2, 103 and 20A DNA restriction fragments suggests that both the Satellite and Alu elements coexist within Subclass 1 DNAs, and that both repetitive elements specifically occur within the 1200 bp restriction fragments of plasmids 2 and 20A and the 1500 bp restriction fragments of plasmids 2 and 103. This suggests that a Satellite II-Alu association occurs twice within the 2 and 103 fragments. Alu elements have been found associated with A+T-rich regions (Hyrien et al., 1987) and with Satellite I DNA (Frommer et al., 1984), but not with Satellite II DNA. The specific Satellite II-Alu association in the current study represents a previously undescribed association.

B. Genomic Distribution of XL3 DNA: Southern Blot Analysis of Genomic DNA

Southern blot analysis hybridizing the XL3 fragment to total human genomic DNA digested with various restriction endonucleases

showed that XL3 DNA was a member of the large family of Satellite II-related sequences, specifically, the non Y-specific Satellite-associated 3.4 kb HaeIII family of repeats. Since XL3 DNA possessed repetitive Satellite II sequences, it was possibile that the XL3 sequence would hybridize to many genomic fragments, and exhibit degrees of homology to genomic DNA ranging from very high (identical sequences) to very low (mismatched sequences) (Mitchell et al., 1979). To test these possibilities, each genomic DNA Southern blot analysis was conducted using conditions defined by different temperatures of hybridization: relaxed (50°C), intermediate (60°) and stringent (72°). Altering the hybridization conditions differentiated homologies of XL3 DNA to a variety of abundant Satellite-containing genomic fragments and more XL3-specific genomic fragments.

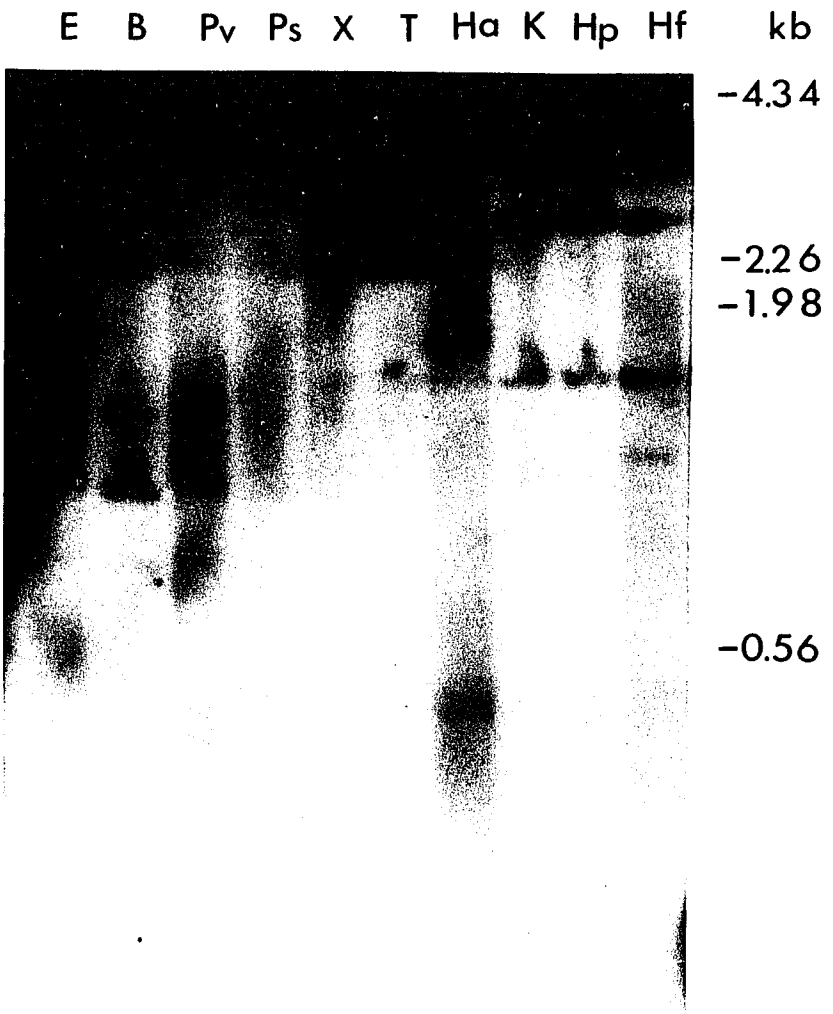
1. Hybridization of Subclone XL3 to Genomic DNA Using Relaxed Conditions

Hybridization of the XL3 fragment to genomic DNA using relaxed conditions showed that XL3 DNA was homologous to many, presumably Satellite II-associated, genomic fragments. Radiolabelled XL3 DNA was hybridized to male human placental DNA digested with EcoRI, BamHI, PvuII, PstI, XbaI, TaqI, HaeIII, KpnI, HpaI or HinfI (Figure 17).

Nonspecific probe-filter binding was a problem using low temperature hybridization and washing conditions. The top portion of the blot, comprising genomic fragments >4000 bp in length, was

FIGURE 17. Southern Blot Analysis of the XL3 Fragment: Hybridization to Genomic DNA using Relaxed Conditions.

100 ng plasmid XL3 was radiolabelled with α^{32}PdATP using random primer extension to a specific activity of 2.5 x 10^8 cpm/ug and hybridized to the Southern transfer of a 1.0% agarose gel. The gel contained 8 ug each human placental DNA digested with EcoRI (Lane 2), BamHI (Lane 3), PvuII (Lane 4), PstI (Lane 5), XbaI (Lane 6), TaqI (Lane 7), KpnI (Lane 8), HpaI (Lane 9), HinfI (Lane 10); 2 each ug plasmid 33 digested with HindIII (Lane 1, not shown) and HindIII + PvuII (Lane 11, not shown), and 2 ug lambda DNA digested with HindIII (Lane 12, not shown). The lengths, in kb, of the HindIII-digested lambda DNA fragments are shown for reference as molecular weight markers. Hybridization and washing were at 50°C as described in Materials and Methods. The autoradiogram was produced by exposure of the blot for 36 hours at -70°C with an intensifying screen. Only the lower portion of the autoradiogram is shown to eliminate very high background signal in the upper portion.


E
B
Pv
Ps
X
T
Ha
K
Hp
Hf
kb
-4.34
-2.26
-1.98
-0.56

generally obscured and is omitted from Figure 17. The 33 DNA hybridization signal was very intense in all but the shortest (16 hour) exposure autoradiograph and is omitted from the 36 hour exposure shown in Figure 17 to enhance the clarity of the remaining portion of the blot.

XL3 DNA specifically hybridized to genomic 6400 (not shown) and 1000 bp EcoRI fragments; 2300, 1300 and 1000 bp BamHI fragments; a 1000 bp PvuII fragment; 4500 (not shown), 1900, 400 bp and small, <400 bp, HaeIII fragments; 2800 and 1400 bp KpnI, HpaI and HinfI fragments, and an additional 1100 HinfI fragment. Hybridization of the 33 fragment to 1000 and 1400 bp genomic fragments was previously shown (see Figure 12). The observation that these same sequences were evident in hybridization of the XL3 fragment to genomic DNA shows that the homology to 33 DNA is associated with XL3, rather than Alu, sequences. An intense hybridization signal was also evident in genomic DNA digested with HaeIII, KpnI, HpaI and HinfI, at a position corresponding to a DNA fragment 3.4 kb in length.

The intense hybridization signal at 1400 bp produced by digestion of genomic DNA with KpnI, HpaI and HinfI suggests that the basic 1400 bp unit defined by these enzymes was homologous to XL3 DNA. The 2800 bp signal may represent dimers of the 1400 bp unit that are incompletely digested or have lost single sites for these enzymes. Similarly, production of a common 1000 bp fragment upon digestion of genomic DNA with EcoRI, BamHI and PvuII indicated

that a basic 1000 bp unit is homologous to XL3 DNA.

In summary, hybridization of XL3 to genomic DNA using relaxed conditions showed that XL3 DNA was homologous to many, probably Satellite II-associated genomic fragments. Hybridization of XL3 DNA to 1000 and 1400 bp fragments was identified.

2. Hybridization of Subclone XL3 to Genomic DNA Using Intermediate and Stringent Conditions

In order to better define the genomic organization of XL3, more complex restriction endonuclease digestions of placental DNA were included in experiments using intermediate and stringent hybridization conditions. Enzymes known to digest XL3 DNA - HindIII, PvuII, XbaI, HinfI and HaeIII - were employed, since they would produce the genomic equivalents of XL3 DNA. Enzymes that do not digest XL3 - EcoRI and HpaI - were used to better define other sites with homology to XL3 DNA.

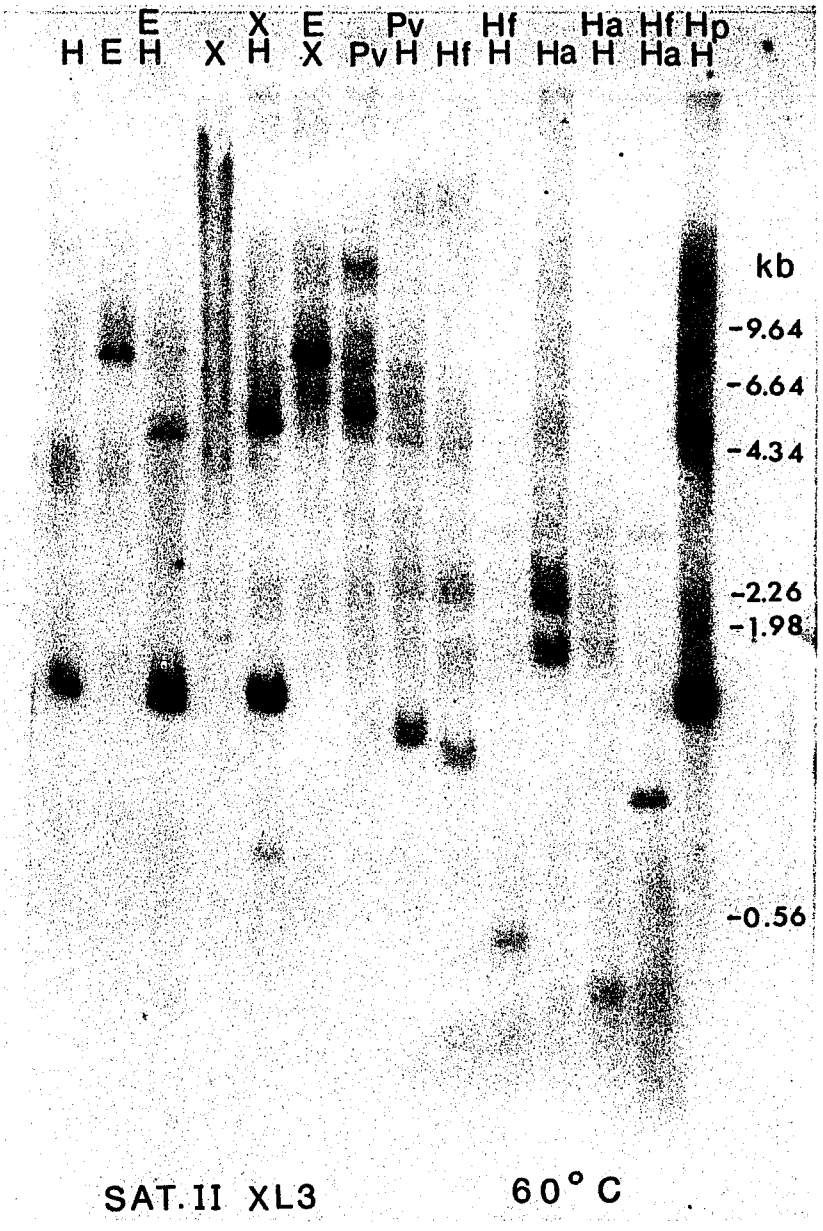
Hybridization of the XL3 fragment to genomic DNA using intermediate and stringent conditions showed that XL3-homologous sequences occurred between 10-100X in the human genome and were consistent in genomic RE sites. Based on RE maps, XL3 DNA is related to non Y-specific 3.4 kb HaeIII sequence.

a. Hybridization of Subclone XL3 DNA to Genomic DNA Using Intermediate Conditions

Hybridization of the XL3 fragment to genomic DNA using intermediate conditions is shown in Figure 18. Many bands of different intensities are evident. These are summarized as major

FIGURE 18. Southern Blot Analysis of the XL3 Fragment: Hybridization to Genomic DNA Using Intermediate Conditions.

100 ng plasmid XL3 was radiolabelled with α^{32}PdATP using random primer extension to a specific activity of 6 x 10^7 cpm/ug and hybridized to the Southern transfer of a 1.0% agarose gel. The gel contained 8 ug each human placental DNA digested with HindIII (Lane 1), EcoRI (Lane 2), HindIII+EcoRI (Lane 3), XbaI (Lane 4), HindIII+XbaI (Lane 5), EcoRI+XbaI (Lane 6), PvuII (Lane 7), HindIII+PvuII (Lane 8), HinfI (Lane 9), HindIII+HinfI (Lane 10), HaeIII (Lane 11), HindIII+HaeIII (Lane 12), HaeIII+HinfI (Lane 13), HindIII+HpaI (Lane 14) and 2 ug lambda DNA digested with HindIII (Lane 15). The lengths, in kb, of the HindIII-digested lambda DNA fragments are shown for reference as molecular weight markers. Hybridization and washing were conducted at 60°C as described in Materials and Methods. The autoradiogram was produced by exposure of the blot filter for 72 hours at -70°C with an intensifying screen.


H E E H X X H X E Pv Pv H Hf Hf H Ha Ha H Hf Ha Hp H
kb
-9.64
-6.64
-4.34
-2.26
-1.98
-0.56
SAT. II XL3
60° C

and minor genomic fragments in Table XI. The 33 and XL3 DNA RE maps predicted the observed bands: 1) a 1.4 kb HindIII fragment; 2) a 750 bp HindIII + XbaI fragment; 3) a 1200 and 250 bp HindIII + PvuII fragments; 4) a 550 bp HindIII + HinfI fragment, and 5) a 450 bp HindIII + HaeIII fragment. Complete homology to XL3 DNA (750 bp HindIII/XbaI fragment) was present in low copy number in the human genome. Other highly homologous genomic DNA fragments of higher molecular weight (1400 bp HindIII+XbaI/EcoRI/HpaI fragments) were more abundant in the genome than the native XL3 locus.

Some of the genomic fragments homologous to XL3 using relaxed hybridization conditions are not apparent in the blot hybridized using intermediate hybridization conditions. The 1000 bp EcoRI and PvuII fragments and 1400 bp HinfI fragment apparent using relaxed hybridization conditions are not present in Figure 18, indicating weak homology to XL3 DNA (Mitchell et al., 1979).

b. Hybridization of Subclone XL3 to Genomic DNA Using Stringent Conditions

Hybridization of the XL3 fragment to genomic DNA using stringent conditions is shown in Figure 19. The fragments lengths corresponding to these bands are summarized in Table XI.

Hybridization under intermediate and stringent conditions gave similar results. Notable differences included the disappearance in the stringent hybridization of 250 bp HindIII + PvuII, 2.5 kb HaeIII and <400 bp HinfI, HindIII + HinfI, HaeIII, HindIII + HaeIII, HaeIII/HinfI DNA fragments. Since the 250 bp HindIII/PvuII

TABLE XI

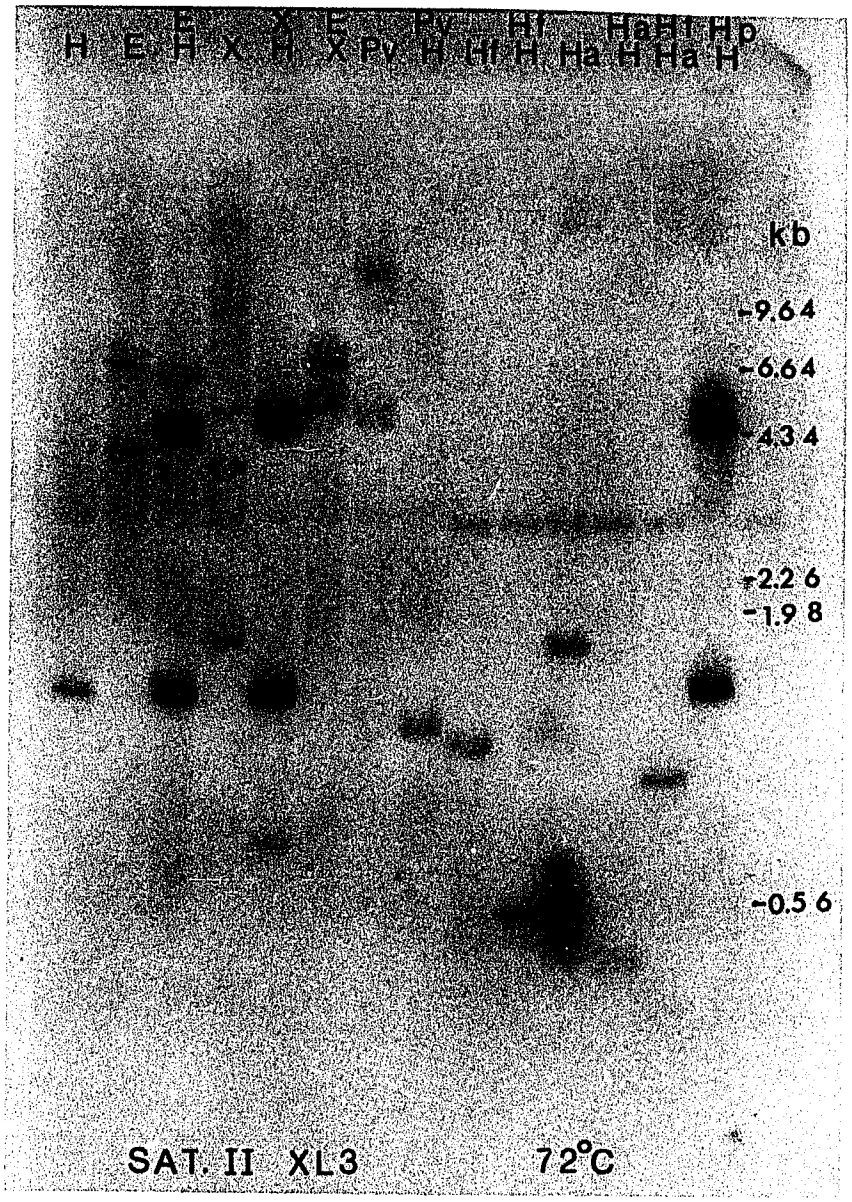
GENOMIC ORGANIZATION OF XL3 DNA

RESTRICTION ENZYMES(S)	GENOMIC FRAGMENTS (BP)			
	MAJOR	MINOR	S	I
HindIII	1400		+	+
EcoRI	6400		-	+
		4800	-	+
		3400	-	+
HindIII + EcoRI	6400		+	-
	5000		+	+
	1400		+	+
XbaI	5500		+	/
		4200	+	/
		3400	+	/
		2000	+	/
HindIII + XbaI	5000		+	+
	1400		+	+
		750	+	+
EcoRI + XbaI	6400		+	+
	5500		+	+
PvuII	10400		+	+
	5200		+	+
HindIII + PvuII	1200		+	+
	250		-	+
HindIII + HpaI	5000		+	+
	1400		+	+
HinfI	3400		+	+
	1150		+	+
		<400	-	+
HindIII + HinfI	3400		+	+
	550		+	+
		<400	-	+
HaeIII	3400		+	+
	2500		-	+
	1900		+	+
		<400	-	+
HindIII + HaeIII	3400		+	+
	450		+	+
		<400	-	+
HaeIII + HinfI	3400		+	+
	1050		+	+
		<400	-	+

LEGEND: + = present; - = absent; / = presence or absence unclear on blot; S = hybridization under stringent (72°C) conditions; I = hybridization under relaxed (60°C) conditions; Major bands were visible after 36 hours of exposure; Minor bands were visible after 72 hours of exposure.

FIGURE 19. Southern Blot Analysis of the XL3 Fragment: Hybridization to Genomic DNA Using Stringent Conditions.

100 ng plasmid XL3 was radiolabelled with α^{32}PdATP using random primer extension to a specific activity of 6 x 10^7 cpm/ug and hybridized to the Southern transfer of a 1.0% agarose gel. The gel contained 8 ug each human placental DNA digested with HindIII (Lane 1), EcoRI (Lane 2), HindIII+EcoRI (Lane 3), XbaI (Lane 4), HindIII+XbaI (Lane 5), EcoRI+XbaI (Lane 6), PvuII (Lane 7), HindIII+PvuII (Lane 8), HinfI (Lane 9), HindIII+HinfI (Lane 10), HaeIII (Lane 11), HindIII+HaeIII (Lane 12), HaeIII+HinfI (Lane 13), HindIII+HpaI (Lane 14) and 2 ug lambda DNA digested with HindIII (Lane 15). The lengths, in kb, of the HindIII-digested lambda DNA fragments are shown for reference as molecular weight markers. Hybridization and washing were conducted at 72°C as described in Materials and Methods. The autoradiogram was produced by exposure of the blot filter for 72 hours at -70°C with an intensifying screen.


H E H X H X Pv H Hl H Ha H Ha H
kb
~9.6 4
~6.6 4
~4.3 4
~2.2 6
~1.9 8
~0.5 6
SAT. II XL3
72°C

fragment of XL3 DNA contained the 201 bp highly degenerate Satellite II sequence (Table VI), the higher hybridization temperature employed under stringent conditions prevented annealing of the XL3 Satellite II DNA sequence to its genomic equivalent The 2.5 kb Hae III and <400 bp smeared fragments may have failed to hybridize to XL3 DNA under stringent conditions for similar reasons.

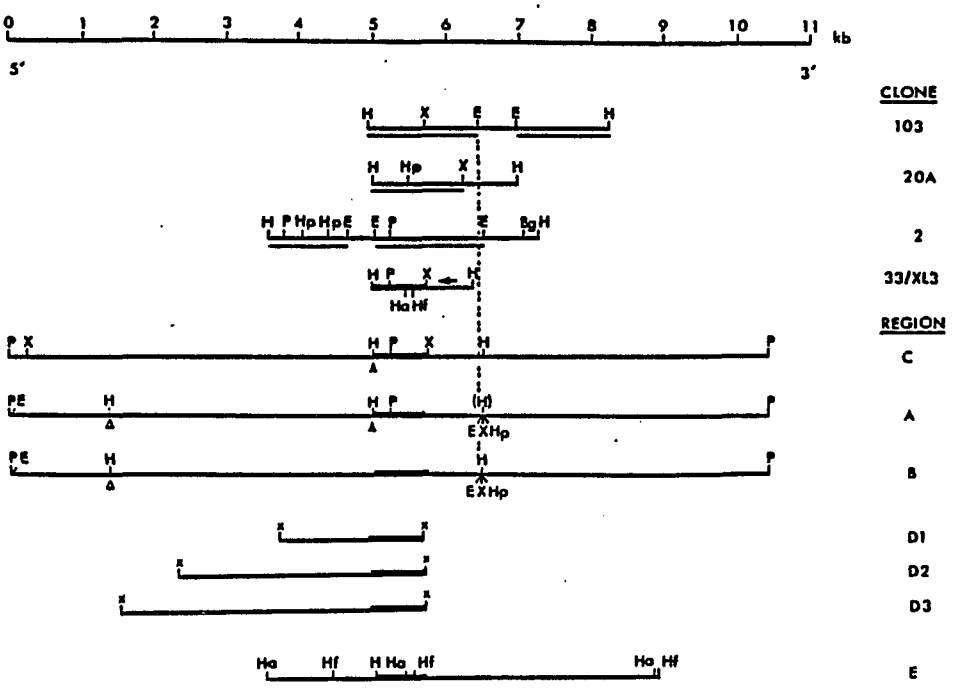
c. Construction of Genomic XL3 Restriction Enzyme Maps

The genomic restriction enzyme maps are shown in Figure 20. The detailed derivation of the maps is provided in Appendix II to this study.

The major genomic fragments homologous to XL3 DNA were 1.4 and 5.0 kb HindIII + EcoRI/XbaI/HpaI sequences (Region A in Figure 20). Also derived from Region A were the large 6.4 kb EcoRI and 5.2 kb PvuII fragments (Figure 20). The persistence of a 5.0 kb fragment in the HindIII + EcoRI/XbaI/HpaI digests shows that it occurs less frequently and is generated independently from the 1.4 kb fragments. The genomic region that produced the 5.0 kb, but not the 1.4 kb HindIII + EcoRI/XbaI/HpaI fragments, is shown as Region B in Figure 20. Region B also contains the large 6.4 kb EcoRI and 10.2 kb PvuII genomic fragments. A reconstruction of the native 33/XL3 DNA locus is shown as Region C (Figure 20). The native 750 bp HindIII/XbaI fragment, the 250 and 1200 bp HindIII/PvuII fragments and the 1.4 kb HindIII fragments are present, as well as the 5.5 kb XbaI and 5.2 kb PvuII sequences. Regions D1, D2 and D3 contain 2.0,

FIGURE 20. Genomic RE Maps of XL3-Homologous DNA.

Representation of the genomic regions identified in male placental DNA by the XL3 fragment, as determined from the Southern blot analyses shown in Figures 18 and 19 and summarized in Table XI. A detailed derivation of the maps is provided in Appendix II. The size of the entire XL3-homologous region is indicated in kilobasepairs (kb) and the genomic regions are drawn to this scale. The Subclass 1 cloned fragments homologous to XL3 DNA are shown in partial alignment with XL3-homologous genomic regions. Underlined portions of the cloned 103, 20A and 2 fragments delineate sequences homologous to both XL3 and BLUR 8 DNA. The 33/XL3 cloned fragment is aligned with the Subclass 1 cloned fragments and genomic Regions A, B, C, D1-3 and E; the XL3 sequence is indicated by the heavy black line and the position and orientation of the fragment 33 Alu element is indicated by the arrow. Region A shows the derivation of the major 1.4 and 5.0 kb genomic fragments homologous to XL3 DNA. The 5' HindIII site of the 1.4 kb genomic fragment is indicated by (▲); the 5' HindIII site of the 5.0 kb genomic fragment is indicated by (△). The 3' HindIII, EcoRI, XbaI and HpaI sites are indicated and aligned in the cloned fragments and genomic regions with the vertical dashed line. Region B is identical to Region A except that the proximal 5' HindIII site is omitted to show independent derivation of the 5.0 kb fragment. Region C shows that native 33/XL3 sequence locus. Regions D1, D2, and D3 define minor XbaI fragments homologous to XL3 DNA. Region E shows the derivation of the HinfI and HaeIII fragments, including the 3.4 kb HaeIII repeated sequence.


0 1 2 3 4 5 6 7 8 9 10 11 kb
5'
3'
CLONE
103
20A
2
33/XL3
REGION
C
A
B
D1
D2
D3
E

3.4 and 4.2 kb genomic fragments (relationship to Regions A, B and C are undetermined) (Figure 20). Region E contained the native 33/XL3 DNA 450 bp HindIII/HaeIII and 550 bp HindIII/HinfI genomic fragments, as well as the 1.9 and 2.5 kb HaeIII, 1.15 kb HinfI, and 1.05 kb HaeIII/HinfI fragments. Region E also contained the prominent 3.4 kb fragment produced by digestion with the HindIII, HaeIII and HinfI enzymes singly and in combination. Region E may be coincident with Region A, since their restriction enzyme maps are compatible, but further mapping would be necessary to show this conclusively.

Comparison of the hybridization signals produced by digestion of genomic DNA with HindIII + XbaI showed that the 750 bp 33/XL3 DNA HindIII/XbaI fragment occurred in only one copy, whereas the 1.4 kb HindIII/XbaI fragment occurred in 10-100 copies.

d. XL3 Is a Variant Non Y-Specific (NYS) 3.4 kb Hae III Repeated Sequence.

Hybridization of the XL3 fragment to a 3.4 kb HaeIII fragment and other 3.4 kb genomic restriction fragments in male placental DNA (Figures 18 and 19; Table XI) shows that XL3 DNA contains sequences common to members of the 3.4 kb HaeIII family (Cooke, 1976). The 3.4 kb HaeIII repeat molecules contain two types of sequences: Y chromosome-specific (it-Y), and non Y chromosome-specific (NYS) (Kunkel et al., 1976, 1977; Cooke and McKay, 1978). Some of the NYS sequences contain Satellite III (Cooke and McKay, 1978; Bostock et al., 1978; Fowler et al., 1987)

and Satellite II (Bostock et al., 1978; Cooke and Hindley, 1979) DNA. In order to explore the possibility that the degenerate Satellite II XL3 fragment was a variant 3.4 kb HaeIII repeat element, its chromosomal localization was determined via hybridization in situ to human metaphase chromosomes.

C. Chromosomal Localization of the XL3 Fragment

Hybridization in situ localized XL3 DNA to the pericentromeric region of chromosome 7, indicating that it was a NYS rather than it-Y Satellite II-associated 3.4 kb HaeIII sequences. A representative metaphase plate localizing XL3 DNA is shown in Appendix I. Significant grain densities were observed at a region distal to the centromere on the q arm of chromosome 7 (7q21) (X^2 = 19) and at the centromere of the same chromosome (X^2 = 19). In a second experiment, significant grain density occurred at a region distal to the centromere on the p arm of chromosome 7 (7p13) (X^2 = 29). The combined data suggested that XL3 localized to the pericentromeric region of human chromosome 7. Satellite II DNA (Gosden et al., 1975) and of NYS 3.4 kb HaeIII repeat sequences (Cooke and McKay, 1978) have both been previously localized to chromosome 7. The specific assignment of XL3 DNA to the pericentromeric region of chromosome 7 supports the assignment of Satellite II DNA to XL3 DNA and NYS 3.4 kb HaeIII sequences.

IX. Summary of Results

The aim of this study was to investigate the heterogeneity of

Satellite II DNA. To accomplish this goal, Satellite II and III fraction sequences were propagated as recombinant DNAs in E. coli. One hundred and one Satellite II and three Satellite III recombinant plasmids were recovered. Eleven plasmids derived from Satellite II DNA and one derived from Satellite III were extensively characterized and grouped into one of four subclasses based upon the presence or absence of repeated DNAs in their sequences.

Subclass 4 cloned fragments - 14, 38 and 20B - were characterized as nonaffiliated in that they possessed nonrepetitive or unique sequences which were not homologous to sequences present in other cloned fragments.

The subclass 3 cloned fragment - 36 - possessed alphoid but not Alu, L1 or Satellite elements and was not homologous to any other cloned fragments.

Subclass 2 cloned fragments - 43, 5, 16 and 85 - showed homology to the L1 KpnI LINES elements. The 43 fragment was identified as a variant 1.9 kb HindIII repeat that hybridized to characteristic and noncharacteristic L1-derived human genomic restriction enzyme fragments. The noncharacteristic fragments could be generated by rearrangement of some L1 sequences. The 43 sequence hybridized in situ to the centromeres of chromosome 4 and the X chromosome. Although 43 DNA is only slightly different from 1.9 kb HindIII repeat, the differences are enough to account for the homology of 43 DNA to rearranged L1 sequences and the marked

specificity of the 43 sequence to the centromeres of chromosome 4 and the X chromosome.

Subclass 1 cloned fragments - 33, 20A, 103, 2 and D - possessed Alu, but not alphoid or L1 elements. The cloned 33 DNA also contained a highly degenerate Satellite II sequence which was subcloned as XL3 DNA and studied, accomplishing the second aim of this work, the characterization of a discrete, cloned Satellite II sequence. The XL3 derivative Satellite II-type sequences were homologous to sequences within the Subclass 1 cloned fragments, indicating that these DNAs contained both Alu element and Satellite II sequences. The XL3 fragment shared some characteristics of the NYS 3.4 kb HaeIII family of Satellite-related sequences and localized specifically to the centromere of human chromosome 7. XL3 DNA is the first Satellite II-type NYS 3.4 kb HaeIII-associated sequence found to localize to an identified Satellite II DNA chromosomal site other than the centromeres of chromosomes 1 and 16 (Gosden et al., 1981; Burk et al., 1985; Jeanpierre et al., 1985; Moyzis et al., 1987). This suggests that the chromosomal specificity of discrete Satellite DNA sequences is a consequence of the simple nucleotide base variations known to exist within heterogeous Satellite DNA populations.

DISCUSSION

This research defined some of the levels of heterogeneity within human Satellite II DNA and characterized, in depth, a variant Satellite II DNA member.

I. Heterogeneity of Human Satellite II Fraction DNA

Total human Satellite II DNA has been previously described in terms of its buoyant density, renaturation kinetics, restriction endonuclease digestion products, chromosomal localization and nucleotide base sequence. These characteristics have been described differently in various studies, even though the sequences under discussion in all cases were purportedly Satellite II DNA.

Problems in identification begin in initial isolation procedures. Ag^{++}/Cs_2SO_4 buoyant densities for Satellite II fraction DNA have been reported as 1.549 g/ml (Corneo et al., 1970), 1.509 g/ml (Corneo et al., 1971), 1.560 g/ml (this study) or have been reported generally as "heavy" (Manuelides et al., 1978; Mitchell et al., 1979) using apparently similar gradient compositions and centrifugation conditions. These discrepancies make it difficult to assess the purity or uniformity of putative Satellite II DNAs isolated in different preparations. Any of the "Satellite II DNAs" obtained at different buoyant densities in Cs_2SO_4 have a single peak at A_{260} with a buoyant density equal to 1.693-1.694 g/ml.

Though the criteria defining purity are more uniform in CsCl gradient isolation procedures than in the primary Cs_2SO_4 gradient centrifugation, it is still unclear if different studies initially isolated the same or different DNA components.

Regardless of technical variations, it is apparent that Satellite II sequences are intrinsically heterogeneous. This is in marked contrast to other types of Satellite DNAs, such as those isolated from mouse and other mammals, which are internally homogeneous (review in Singer, 1982). C_ot analysis has shown that Satellite II DNA renatures as a single kinetic component. The melting curve of the reassociated Satellite II component, however, is biphasic, indicating that Satellite II DNA consists of at least two different, but related sequences. Some of these sequences (10%) are homologous to Satellite III DNA (Mitchell et al., 1979). The remaining 90% are presumably Satellite II-specific and other, possibly related, sequences.

The current study investigated the heterogeneous nature of Satellite II DNA and identified five types of DNA sequence: the simple sequence Satellite II consensus sequence (termed Satellite 2 in Prosser et al., 1986); Alu SINES elements; KpnI L1 LINES elements; alphoid DNA, and unique DNA. The most abundant types were DNA having a degenerate Satellite II sequence and Alu element DNAs. These occurred together within Subclass 1 cloned sequences and accounted for 38% of the cloned DNAs studied. The next most abundant sequences were the KpnI L1 LINES elements (Subclass 2),

which accounted for one third of the cloned fragments studied. Subclass 4, primarily unique sequences, accounted for 21% and Subclass 3 was a single cloned DNA containing alphoid sequences. The majority of the sequences chosen randomly for study - 79% - were composed of repetitive DNAs, and half of these consisted of Satellite II-related simple sequences and Alu element sequences.

Since no comparable study exists, it is not known if the identification of these varied repeated DNAs within the Satellite II fraction accurately or completely defines the heterogeneity of Satellite II DNA. This study, however, illustrates the kind of heterogeneity that could characterize gradient-isolated Satellite II fraction sequences and may account for the origin of some of the discrepancies apparent in published descriptions of Satellite II DNA. A subcloned region, XL3 DNA, of the Subclass 1 sequence, 33 DNA, was used to characterize a subset of Satellite II-derived sequences that will be discussed in detail. These sequences 1) occur in 10-100 copies in the human genome; 2) are degenerate Satellite II simple sequences; 3) are related to the 3.4 kb HaeIII repeat family of Satellite II sequences, and 4) localize predominantly to the pericentromeric region of human chromosome 7.

II. The XL3 Cloned Fragment, a Variant Satellite II Sequence

A. The Partial Nucleotide Base Sequence of the XL3 Fragment is a Degenerate Satellite II Repeat

The partial nucleotide sequence of the Subclass 1 member 33

fragment was characterized by the presence of six tandemly repeated degenerate 26-nucleotide Satellite II-type sequences in the 5' bp 1-201 region. The six repeat units aligned with one suggested Satellite II consensus sequence (the Satellite 2 simple sequence defined by Prosser, 1986) so that an average of 56% of the consensus residues were present per repeat unit. Random base changes, random insertions of 1-4 nucleotides and the insertion of a large 26 nucleotide non-Satellite region between bp 81-107 contributed to the degeneracy of the Satellite 2-type repeats within the 5' bp 1-201 33 DNA sequence. This was compared with the only three other known sequences purported to be Satellite II in order to redefine or establish a consensus sequence. One of these sequences, the pHuR 195 sequence, was originally characterized as a variant Satellite II DNA possessing 8 (5' terminal) and 4 (3' terminal) complete Satellite 2-type repeats with the sequence 5'ATTCC ATTG/AC ATTGG/T ATTG/AC/T ATGATG 3'; these were interspersed with other internally repetitive and highly diverged sequences (Moyzis et al., 1987). Direct alignment of the pHuR 195 sequence with a previously suggested consensus sequence, 5'ATTCC ATTCG G/AGT/ACC ATTCG ATGATG 3'(Prosser et al., 1986, modified to show a final dGMP residue), shows that the 5' terminal region possess 10 complete and 4 partial Satellite 2-type repeats arranged tandemly with no significant nucleotide insertions or deletions. Similarly, the 3' terminal region possess 4 complete and 3 partial Satellite 2-type repeats arranged tandemly with no significant

nucleotide insertions or deletions. This alignment incorporates the 9-nucleotide 5' ATTCC ATTC 3' repeats and highly diverged sequences identified by Moyzis et al. into Satellite II consensus sequence-homologous repeat units. The tandemly repeated nature of Satellite II DNA sequences (Frommer et al., 1982; Prosser et al., 1986) suggests that direct alignment of the pHuR 195 sequences with the Satellite II consensus sequence, as proposed here, is a more accurate definition of these sequences than the indirect alignment proposed by Moyzis et al. (1987). The directly aligned sequences range in homology to the Satellite II consensus sequence from 50% to 92%, with averages of 76% (5' terminal region) and 77% (3' terminal region). These values quite adequately reflect the Satellite II derivation of the pHuR 195 sequences.

Similarly, direct alignment of the pPD17 cloned fragment with the Satellite II consensus sequence reveals 4 complete and 5 partial Satellite 2-type repeat units arranged tandemly with an average 82% homology to the Satellite II consensus sequence. The pPD17 sequence has not previously been compared extensively with the Satellite II consensus sequence in the manner shown here. Direct alignment of the pPD17 and Satellite II consensus sequences showed that, of the cloned Satellite II sequences described in the literature to date, it most closely fits the suggested consensus sequence.

Direct alignment of the 5' terminal 132 bp, internal 75 bp and 3'terminal 92 bp HS3 sequence nucleotides with the Satellite 2 sequence showed that they share average sequence homologies of 69%

(5' terminal 132 bp), 77% (internal 75 bp) and 75% (3' terminal 92 bp). The HS3 sequences are organized in tandem arrays of the 26 nucleotide Satellite II repeat unit comprising a total of 8 complete and 4 partial consensus sequence repeats.

The current study shows that the cloned Satellite II sequences characterized to date are organized into tandemly repeated 26-nucleotide Satellite II consensus sequence units. This is confirmed as the basic organization of Satellite II DNA in the human genome. Deviations from the consensus sequence occur as random base changes within the 26-nucleotide repeat unit and, less frequently, as random nucleotide insertions of varying lengths within the repeated sequences units.

The S1 sequence, comprising the 5' terminal 201 bp of the cloned 33 and subcloned XL3 fragments, is a particularly degenerate member of the Satellite II-type sequences characterized by tandem arrays of the basic 26-nucleotide Satellite II simple sequence.

B. Genomic Restriction Enzyme Sites Identified by the XL3 Fragment are Associated with Satellite II DNA

Recognition sites for the restriction enzymes HinfI, TaqI, EcoRI and HaeIII have been characterized in Satellite II fraction DNA as well as specific cloned DNAs (Cooke and Hindley, 1979; Moyzis et al., 1987). These sites also occurred in digests of human placental DNA hybridized to the cloned XL3 fragment.

Digestion of Satellite II DNA with HinfI generally produces a ladder of very small fragments due to the occurrence of the 5'

GAG/CTC 3' HinfI site at positions 10-14 of the 26-nucleotide Satellite 2 simple sequence (Frommer et al., 1982; Prosser et al., 1986). Large, discrete fragments of approximately 1500 bp and 1100 bp in length are also prominent in Southern blot analyses of HinfI-digested genomic DNA hybridized to Satellite II fraction DNA (Frommer et al., 1984).

Human genomic DNA digested with HinfI and hybridized to the XL3 fragment produced fragments 1100 bp in length using both intermediate and stringent annealing conditions and produced very small (<400 bp) fragments using intermediate conditions, indicating that the degenerate Satellite II sequence in XL3 DNA is homologous to characterized Satellite II-derived HinfI restriction fragments in genomic DNA. Although there is a paucity of HinfI sites within the XL3 fragment itself resulting from the degeneracy of the XL3 Satellite 2 simple sequence, homology of the XL3 fragment to HinfI fragments identified the XL3 sequence as homologous to genomic Satellite II DNA.

Similarly, TaqI digestion at 5' TCGA 3' recognition sites at positions 18-21 of the 26-nucleotide Satellite II consensus sequence produces multiple small fragments in Satellite II DNA (Frommer et al., 1982; Prosser et al., 1986). TaqI sites are not found in the 5' 201 bp XL3 sequence. Further, genomic DNA digested with TaqI did not hybridize well to the XL3 fragment (Figure 17), though the poor quality of the Southern blot makes this information difficult to assess. Characterization of the XL3 fragment as a

Satellite II-derived DNA is consequently inconclusive using the criterion of TaqI digestion. Digestion of Satellite II DNA with EcoRI is expected to produce two fractions: high molecular weight DNA undigested by the enzyme, and several fragments 4.2, 3.4, 2.2, 1.8 and 1.4 kb in length, the most prominent being the 1.8 kb fragment (Mitchell et al., 1979; Frommer et al., 1982). The XL3 fragment was homologous to a prominent 6.4 kb fragment and a smear of higher and lower molecular weight material when hybridized to EcoRI-digested genomic DNA using intermediate conditions. Under stringent conditions, XL3 DNA identified discrete 6.4, 4.2 and 3.4 kb EcoRI genomic fragments. This further placed XL3 DNA into the category of Satellite II DNA, specifically the NYS 3.4 kb HaeIII repeated EcoRI-defined sequences.

Hybridization of total Satellite II fraction sequences to male-derived genomic DNA digested with HaeIII is expected to produce a large, variable fraction of high molecular weight DNA; a prominent 3.4 kb fragment, and a degenerate ladder of bands comprising multiples of 170 bp superimposed upon a range of other fragments of apparently random sizes. The three hybridization fractions do not appreciably cross-react, indicating that they are not truly homologous (Mitchell et al., 1979). When hybridized to HaeIII-digested, male-derived human genomic DNA, XL3 DNA did not appreciably identify high molecular weight, undigested fragments but did identify a prominent 3.4 kb fragment, and a smaller 1.9 kb genomic fragment. No "ladder" of fragments was apparent, although a

2500 bp fragment and small, <400 bp fragments did appear using intermediate hybridization conditons. This indicated that XL3 DNA was homologous to some Satellite II sequences present in HaeIII-digested DNA, particularly the 3.4 and 1.9 kb fragments.

III.XL3 DNA Defines a Minor Domain of NYS 3.4 kb HaeIII Sequences

The 3.4 kb HaeIII fragment of human male DNA was first described by Cooke (1976). Kunkel et al. (1976) showed that these sequences comprised both Y-chromosome specific (it-Y) and non Y-chromosome specific (NYS) DNAs. The non Y-chromosome specific sequences hybridized to female DNA, indicating that they were autosomal in origin (Kunkel et al., 1977; Cooke and McKay, 1978).

C$_o$t analysis and renaturation kinetics showed that ~7500 copies of the 3.4 kb HaeIII molecule occurred per male diploid genome and that both it-Y and NYS sequences occurred within each 3.4 kb HaeIII molecule. Three ~800 bp it-Y sequences were interspersed with and covalently linked to four ~200 bp NYS sequences per molecule. Together, as components of the 3.4 kb HaeIII repeat molecules, it-Y and NYS sequences accounted for 40% of Y chromosome DNA. NYS DNA was rapidly reassociating, suggesting that it contained simple repetitive sequences. Further, NYS DNA was shown to consist of at least two sequence "families" with overall homologies of 0.2% and 0.34%, respectively, to total female and male genomic DNA. Based on this information, the 3.4 kb HaeIII repeat could be generally characterized as a heterogeneous collection of it-Y sequences

covalently linked to more highly repetitive, less complex NYS sequences within the same molecules (Kunkel et al., 1979). A later study showed that approximately 50% of it-Y sequences were truly Y chromosome specific, whereas the other 50% cross-reacted with female DNA using relaxed reassociation criteria. This study also showed that ape DNA did not hybridize to the Y-specific it-Y sequences but did hybridize to the cross-reacting it-Y and NYS sequences. The study concluded that the cross-reacting it-Y and NYS sequences were more divergent, hence "older", than the Y specific it-Y sequences, suggesting that the 3.4 kb HaeIII repeat molecules were primarily autosomal, not Y-chromosomal, in origin (Kunkel and Smith, 1982).

Previous studies showed that Satellite III (Cooke and McKay, 1978; Bostock et al., 1978) and Satellite II (Bostock et al., 1978; Cooke and Hindley, 1979) DNA contained 3.4 kb HaeIII repeat sequences. Both male and female DNA hybridized to the Satellite III 3.4 kb HaeIII element, suggesting that Satellite III DNA contained both it-Y and NYS sequences (Cooke and McKay, 1978). Since Satellite II and III DNA consists of highly reiterated, simple sequences, it could compose a major part of NYS 3.4 kb HaeIII DNA. Recent studies have shown that this is the case. Burk et al. (1985) have defined three "domains" of sequences within the 3.4 kb HaeIII-type DNAs, all of which are related to Satellite DNA. Domain K, defined by hybridization of genomic 3.4 kb HaeIII sequences and a cloned, pY-3.4A, probe to 1.8 and 3.6 kb KpnI-generated genomic

fragments, occurs in many copies of a tandemly repeated 1.8 kb sequence specific to chromosome 15. Domain D, defined by hybridization of genomic 3.4 kb HaeIII sequences and a cloned probe (6D) to a large 6.2 kb fragment and three smaller 2.6, 2.0 and 1.6 kb (which sum to 6.2 kb) DdeI-generated genomic fragments, occurs as many copies of a tandemly repeated 6.2 kb sequence specific to chromosome 16. The cloned pY-3.4A and 6D fragments do not cross-react, indicating that they represent different types of 3.4 kb HaeIII repeat sequences. A third domain, Domain R, was defined by hybridization of genomic 3.4 kb HaeIII sequences and the cloned 6D probe to a distinct 3.4 kb EcoRI-generated genomic fragment and several other minor EcoRI-generated bands. These bands were similar in appearance to those homologous to the Satellite II-derived HS3 cloned fragment. This suggested that Domain R represented a chromosome 1-specific NYS 3.4 kb HaeIII sequence based on its identity with the chromosome 1 centromere-specific HS3 fragment (Gosden et al., 1981). In another study, a cloned chromosome 22 library-derived fragment, QP23, identified the 3.4 kb HaeIII repeat fragment in male genomic DNA. The QP23 probe defined a Domain K-like 1.8 kb tandemly repeated sequence specific to chromosomes 9 and 15 and a Domain D-like 6.1 kb tandemly repeated sequence specific to chromosome 16 (Jeanpierre et al., 1985). A third study showed that a Domain K-like cloned repetitive 1.8 kb KpnI sequence comprised tandem arrays of the Satellite III 5'ATTCC 3' pentameric repeat and localized to the pericentromeric region of chromosome 15

(Higgins et. al., 1985). Taken together, these studies indicate that the NYS 3.4 kb HaeIII repeat sequences may be divided into Satellite III-derived chromosome 9 and 15-specific domains and Satellite II-derived chromosomes 1 and 16-specific domains. Burk et al. (1985) proposed that other domains undoubtedly exist, but have yet to be identified.

The relationship between inter-and intra-domain sequences is not consistent. The 1.8 kb tandemly repeated Domain K-like sequence has been described as predominantly localized to chromosome 15 with possible minor sites on other chromosomes (Burk et al., 1985), or as predominantly localized to chromosome 9 with a minor localization to chromosome 15 (Jeanpierre et al., 1985). These discrepancies may be attributed to the use of two cloned probes that differ somewhat in sequence but were both homologous to the 1.8 kb tandem repeats and NYS 3.4 kb HaeIII sequences. This suggests that Domain K-like DNAs comprise a family of similar Satellite III-derived sequences that have diverged somewhat from each other. The divergence may be due to the segregation of Domain K-like sequences onto different chromosomes sometime during the evolutionary history of these sequences, as has been proposed for Satellite III DNA (Beauchamp et al., 1979).

The 6.1-6.2 kb tandemly repeated Domain D-like sequences have been described as predominantly localized to chromosome 16 (Burk et al., 1985; Jeanpierre et al., 1985)) whereas the Domain R-like sequences localized to chromosome 1 (Burk et al., 1985). Since the

3.4 kb HaeIII repeated sequences contain Satellite II DNA (Bostock et al., 1978; Cooke and Hindley, 1979), and the major sites of Satellite II localization in the human genome are the centromeric regions of chromosomes 16 and 1 (Jones and Corneo, 1971), it is reasonable to propose that Domain D- and R- like sequences represent two families of related Satellite II-derived DNAs. The cloned HS3 sequence (Cooke and Hindley, 1979; Gosden et al., 1981) may be a Domain R-like DNA, since it contains Satellite II sequences and is specific to the centromere of chromosome 1. The precise relationship between Domain D and R sequences is unclear, however, in that a cloned probe, 6D, identifies both Domains in genomic DNA but is specific only to chromosome 16, not chromosome 1 (Burk et al., 1985). Another Domain D-like sequence is also specific to chromosome 16, but not to chromosome 1 (Jeanpierre et al., 1985). This indicates that Domain D and R DNAs represent two distinct Satellite II-related sequence families within the NYS 3.4 kb HaeIII repeat sequences. The divergence may be due to the segregation of Domain D-like sequences predominantly to chromosome 16 and of Domain R-like sequences to chromosome 1 sometime during the evolutionary history of Satellite II DNA.

The present study proposes that the XL3 cloned fragment represents a third Satellite II-related domain of sequences associated with the NYS 3.4 kb HaeIII repeats. This domain, Domain H, is defined by DNA homologous to XL3 sequences produced by digestion of human male genomic DNA with the restriction

endonucleases HaeIII and HinfI. It is a minor component of the Satellite II-derived NYS 3.4 kb HaeIII sequences, is specific to the pericentromeric region of chromosome 7, and consists of 10-100 repeated sequences 1400 bp in length. Though it is unclear how these sequences are tandemly repeated, the segregation to one chromosomal site suggests that homologous sequences are clustered together along one continuous stretch of DNA. XL3 does not localize to either chromosome 1 or 16, and does not identify a 1.8 kb EcoRI genomic fragment often associated with chromosome 1 (Cooke and Hindley, 1979;) or chromosome 16 (Burk et al., 1985; Jeanpierre et al., 1985). This indicates that Domain H exists independently of Domains D and R. The identification of homology to minor Domain H DNA suggests that other minor domains will be identified in future studies. Together, this information may be used to define the heterogeneity of Satellite II DNA and the NYS 3.4 kb HaeIII sequences in particular.

IV. Association of Satellite II and Alu Sequences in Subclass 1 DNA

The degenerate Satellite II-type sequences between bp 7-198 were associated with an inverted Alu SINES element spanning approximately bp 900-1200 in the 33 fragment. The XL3 subcloned fragment, consisting of bp 1-750 of 33 DNA and containing the degenerate bp 7-198 Satellite II sequences, some slight homology to restriction fragments of cloned 2 DNA, cloned 103 DNA, cloned 20A DNA. These restriction fragments also contained sequences

homologous to BLUR 8 Alu SINES element DNA. Thus, the XL3 DNA and Alu element association in the 33 fragment also occurred in the 2, 103 and 20A cloned fragments. Moreover, the XL3 and Alu DNA association was repeated twice within cloned 2 and 103 DNAs.

As seen in Figures 3 and 20, the 1500 and 1200 bp sequences of the cloned 2 and 103 fragments are not contiguous but are separated by sequences 300 bp apart in 2 DNA and 500 bp apart in 103 DNA. This suggests that the Alu elements within the cloned 2 fragment are minimally 300 bp and maximally 3000 bp (1200 bp + 300 bp + 1500 bp) apart (Figure 20). Similarly, the Alu elements within the cloned 103 fragment are minimally 500 bp and maximally 3200 bp (1500 bp + 500 bp + 1200 bp) apart. Random dispersion throughout the human genome places Alu elements approximately 4200 bp apart from each other. Kinetic data, however, suggests that Alu elements are not randomly dispersed but may be "bunched up" in 60% of the genome, occurring approximately every 2500 bp. The remaining 40% of the genome possesses Alu elements that are more "spread out", or widely dispersed, among unique and other repetitive sequences (Rinehart et al., 1981). Maximal separation of the Alu elements in 2 and 103 DNA places them closer to 2500 bp apart than 4200 bp apart, suggesting that the 2 and 103 fragments were purified from the 60% of genomic sequences containing closely spaced Alu elements.

Within the 60% of human genomic DNA associated with frequently occurring Alu sequences are regions where Alu elements are found in

"clusters" (Calabretta et al., 1982; Tsukada et al., 1982). Alu elements have also been found in "paired" arrangements in the human genome. Some paired Alu elements have been implicated in specific genomic rearrangements, including intrastrand recombination (Lehrman et al., 1985). Inter-Alu sequences have also been shown to occur in 50 polymorphic copies dispersed in normal lymphocytic DNA. Some copies of the inter-Alu sequence existed as small polydisperse circular DNAs (Calabretta et al., 1982). An increase in the number of the polydisperse circular inter-Alu sequences was associated with in vitro and in vivo ageing (Shmookler Reis et al., 1983). These studies proposed that the inter-Alu sequences had undergone replication, excision and reintegration events which amplified and dispersed them throughout the genome as extrachromosomal circular DNAs. This transposon-type mechanism may have involved the flanking Alu sequences, which could have functioned in a manner similar to prokaryotic transposon insertion sequences (Shmookler Reis et al., 1983).

As previously stated, 2 and 103 DNA possess two associated Alu element-XL3 DNA sequence complexes. If these sequences are oriented as they are in 33 DNA, then the 2 and 103 sequences are arranged in an XL3-Alu-XL3-Alu array. This configuration would place an XL3 degenerate Satellite II sequence between two Alu elements in both cloned fragments. This suggests that the occurrence of 10-100 copies of degenerate Satellite II-containing XL3 sequences in the genome could theoretically represent an Alu element-mediated

transposon-like amplification of these specific sequences.

V. Origin of Satellite II DNA and Specific Origin of the XL3 Cloned Fragment

The identification of XL3 DNA as a minor component of Satellite II DNA raised questions concerning the specific origin of XL3 sequences in relation to the general origin and genomic distribution of Satellite II DNA.

Previous studies have shown that Satellite II DNA was composed of populations of distinct, but related, sequences comprising approximately 2.0% of genomic DNA (Mitchell et al., 1979; Corneo et al., 1970). Total Satellite II DNA localized primarily to the centromeric region of chromosome 1 with secondary localization to the centromeric region of chromosome 16 and minor localizations to the centromeric regions of many other chromosomes (Jones and Corneo, 1971; Gosden et al., 1975). Satellite II sequences were associated with NYS 3.4 kb HaeIII repeated DNAs (Bostock et al., 1978; Cooke and Hindley, 1979). Renaturation kinetics showed that NYS 3.4 kb HaeIII sequences were probably autosomal in origin (Kunkel and Smith, 1982). Taken together, this information is consistent with a centromeric origin of a major portion of Satellite II DNA on human chromosome 1. The secondary concentration of Satellite II sequences at the centromeric region of chromosome 16 and minor concentrations on the centromeres of other chromosomes could have arisen subsequently through amplification and dispersion

of Satellite II sequences to these sites sometime during human evolution.

Two mechanisms that account for the amplification and dispersion of Satellite II sequences in the human genome are unequal crossing over and disproportionate replication. Unequal crossing over was recently elucidated as a major mechanism contributing to the amplification of alphoid sequences in the human genome (Willard and Waye, 1987). In unequal crossing over, homologous recombination between two nonallelic repeated sequences organized in a tandem array results in reciprocal amplification and deletion of the repeated sequence. These events also act to maintain sequence homogeneity so that the DNA eventually consists of repeats of a single sequence through a process called crossover fixation. This mechanism also allows that similar repeated sequences on different chromosomes that do not participate in unequal crossing over are free to diverge (Smith, 1976; Lewin, 1980). Unequal crossing over adequately accounts for the initial amplification of Satellite II sequences at the centromere of chromosome 1, the relative homogeneity of these sequences, and the divergence of these sequences at other sites. It does not explain the dispersion of these sequences to other sites or the absence of reciprocal deletion products from unequal sequence exchanges.

The dispersion of Satellite II sequences to many sites in the genome is accounted for by a second mechanism, disproportionate replication. In this process, multiple initiations of DNA

replication result in the production of some free strands of newly synthesized DNA which may undergo end-to-end ligation. The resulting closed circular DNAs may then recombine into chromosomal DNA at or near the initial site of DNA synthesis or into a different site on the same or on a different chromosome (Schimke, 1984). Disproportionate replication accounts for the initial amplification of Satellite II DNA at the centromere of chromosome 1 and its subsequent dispersion to chromosome 16 and other chromosomes. Rounds of amplification and dispersion of Satellite II sequences from these to the same and other sites would then account for the presence of different amounts of Satellite II sequences on different chromosomes. This mechanism also allows the independent divergence of Satellite II sequences from each other on separate chromosomes.

Amplification-independent dispersion of Satellite II sequences could also have occurred through their transposition to other sites in the genome as "inter-Alu" sequences, a process alluded to earlier. In this type of event, Satellite II sequences sandwiched between two Alu elements could have been transposed to other genomic sites through Alu-Alu recombination events. This particular mechanism is of special interest here because of the possible XL3(Satellite II)-Alu to XL3(Satellite II)-Alu sequence arrangement in the cloned 2 and 103 fragments.

Some combination of these mechanisms is consistent with the general amplification and dispersion of Satellite II sequences in

the human genome. None of the mechanisms, however, explain the preferential localization of Satellite sequences to heterochromatic, centromeric regions. It has been proposed that the almost exclusive placement of Satellite sequences at centromeres is a "defense" mechanism designed to "protect" the genome from the intrusion of masses of repeated sequences into loci critical to cell function and propagation (Singer, 1981). This presupposes a necessary function of Satellite DNA so that Satellite sequences are maintained in the genome. A lack of selective pressure on the existence of Satellite sequences, balanced with the exertion of selective pressure to maintain the heterochromatic localization of Satellite DNA, may have resulted in the observed abundance of Satellite II sequences at selected centromeric loci in the human genome.

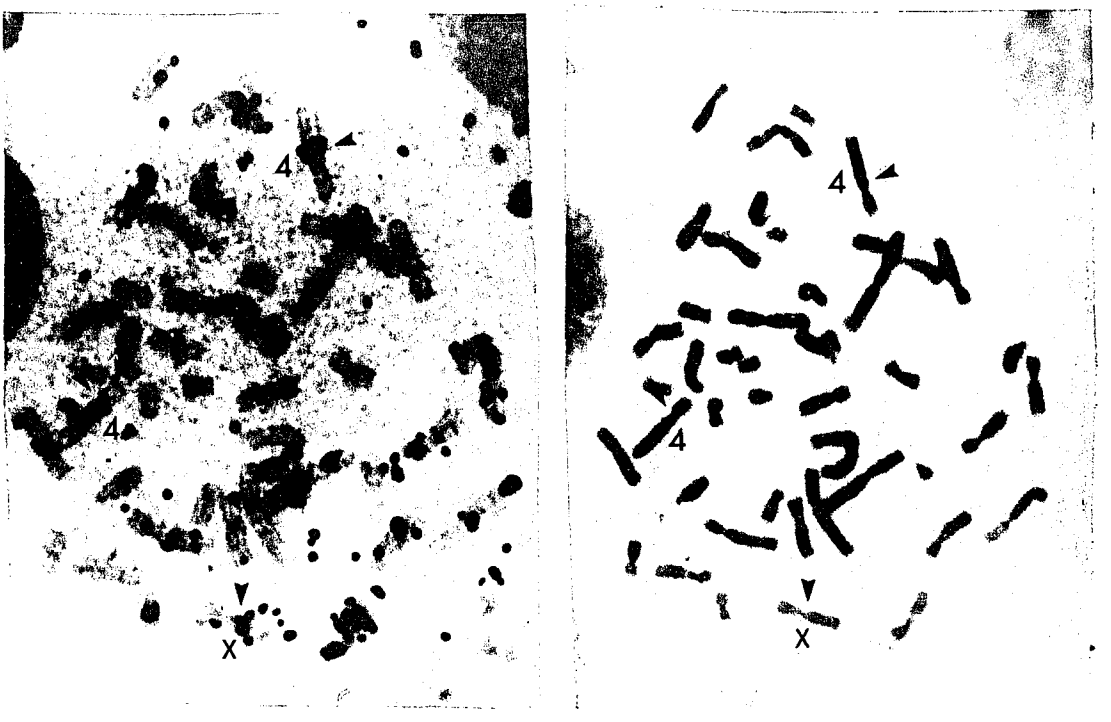
At some point in evolutionary history, one or a few copies of Satellite II DNA was dispersed to the centromere of chromosome 7. These sequences, segregated from other Satellite II sequences, underwent random changes and diverged considerably from the original Satellite II consensus sequence. Limited amplification occurred so that, eventually, 10-100 copies of this highly variant Satellite II sequence existed at that chromosomal location. This proposed mode of origin of the XL3 sequence accounts for its localization to a minor Satellite II chromosomal site, its divergence from the Satellite II consensus sequence and its low copy number. This mechanism may account for the existence of other

variant Satellite II sequences in the human genome which, together, define the heterogeneity of Satellite II DNA.
loci in the human genome.

The mechanisms involved the origin of XL3 DNA evolve directly from those just described for the origin of Satellite II DNA. At some point during evolutionary history, one or a few copies of Satellite II DNA was dispersed to the centromere of chromosome 7. These sequences, segregated from other Satellite II sequences, underwent random mutations and diverged considerably from the original Satellite II consensus sequence. They underwent limited amplification so that eventually 10-100 copies of this highly variant Satellite II sequence existed at that chromosomal location. This proposed mode of origin of the XL3 sequence accounts for its localization to a minor Satellite II chromosomal site, its divergence from the Satellite II consensus sequence and its low copy number. This mechanism may account for the existence of other variant Satellite II sequences in the human genome which, together, define the heterogeneity of Satellite II DNA.

APPENDIX I

Hybridization In Situ of Human Metaphase Chromosomes



1. 2.

A. Chromosomal Localization of 43 DNA

A representative metaphase plate is shown prior to hybridization (1.) and as it appeared after hybridization (2). The centromeric regions of chromosome 4 and the X chromosome, the major sites of 43 DNA localization, are indicated. The slide was exposed 48 hours at -4°C.

1. 2.

B. Chromosomal Localization of XL3 DNA

A representative metaphase plate is shown prior to hybridization (1.) and as it appeared after hybridization (2). The centromeric region of chromosome 7, the major site of XL3 DNA localization, is indicated. The slide was exposed 72 hours at 4°C.

APPENDIX II.

Derivation of the Genomic RE Maps of XL3 DNA

The information obtained from the Southern blot analyses using relaxed (Figure 17), intermediate (Figure 18) and stringent (Figure 19) hybridization conditions, as summarized in Table XI, was used to construct the genomic RE maps shown in Figure 20.

A. Characterization of the Major 1.4 and 5.0 kb XL3-Homologous Genomic Fragments

Mapping of the restriction enzyme sites defining genomic fragments homologous to XL3 DNA was derived from the analysis of blots shown in Figures 18 and 19 and summarized in Table XI. The most intense regions of genomic homology to the XL3 fragment correspond to fragments 1.4 and 5.0 kb in length produced by digestion of genomic DNA with HindIII + EcoRI, HindIII + XbaI and HindIII + HpaI. These fragments are not evident upon digestion of genomic DNA with either EcoRI or XbaI alone (Figures 17,18,19). The 1.4 kb but not the 5.0 kb fragments are observed upon digestion of genomic DNA with HindIII (Figure 18 and 19) or HpaI (Figure 17) alone. The production of similar 1.4 and 5.0 kb fragments upon digestion with HindIII + EcoRI, + XbaI and + HpaI suggests that

these fragments arise from similar genomic loci. EcoRI, XbaI and HpaI sites clustered 1.4 kb 3' or 5' from a HindIII site would produce the observed identical fragments upon digestion with HindIII + EcoRI/XbaI/HpaI. Movement of the HindIII site to a position 5.0 kb away from the clustered EcoRI, XbaI and HpaI sites would produce the observed identical 5.0 kb fragments which occur following digestion with HindIII + EcoRI/XbaI/HpaI.

The 1.4 and 5.0 kb HindIII + EcoRI/XbaI/HpaI fragments appear with equal intensity in Figure 19 but the 5.0 kb fragments appear with much less intensity than the 1.4 kb fragments in Figure 18. This suggests that some XL3-homologous loci possess two 5' HindIII sites: one located 1.4 kb and one located 5.0 kb from the clustered EcoRI/XbaI/HpaI sites. Complete restriction enzyme digestion of these sites cleaves the larger 5.0 kb fragment producing the smaller, more intense 1.4 kb fragment. The intensity of the 1.4 kb signal suggests that it comprises most of the genomic fragments homologous to XL3 DNA. These fragments share the expected restriction enzyme sites, which are mapped as shown in Region A (Figure 20). Though not shown, an additional HpaI site may occur near the 5' HindIII site, since a 1.4 kb genomic fragment homologous to XL3 is produced upon digestion with HpaI alone (Figure 17).

The persistence of the 5.0 kb HindIII + EcoRI/XbaI/HpaI fragment indicates that this fragment is not always digested to the smaller 1.4 kb fragment by HindIII (Figures 18 and 19). Region B

represents the place of origin of the 5.0 kb fragments. The 1.4 kb HindIII fragment is probably generated independently at a third locus, since the intensity of this fragment is much less than that of the HindIII + EcoRI/XbaI/HpaI fragments (Figures 18 and 19). This locus is depicted as Region C in Figure 20, and may represent the native 33/XL3 locus since the 33 sequence is a 1.4 kb HindIII fragment.

Analysis of the data shown in Figures 17, 18 and 19 and the derived genomic map shown in Figure 20 indicate that full homology to XL3 DNA exists at only a few genomic loci. The faint signal produced by the 750 bp HindIII/XbaI fragment (Figures 18 and 19) suggests that this fragment is the genomic counterpart of the XL3 fragment and is present in only one copy. Comparison of signal intensities shows that the number of copies of homologous 1.4 kb HindIII + EcoRI, + XbaI and + HpaI genomic sequences is between 10-100, i.e., the XL3-homologous Region A occurs 10-100 times in the human genome. Similarly, comparison of signal intensities suggests that Region B also occurs many times in the human genome, though perhaps less frequently than Region A. This discrepancy prevents a precise determination of the number of genomic copies of Region B, but allows a conservative estimation of 1-10 total copies in the human genome. The signal intensity of the 1.4 kb HindIII fragment is consistent and comparison with that of the 750 bp HindIII/XbaI fragment shows that the 1.4 kb HindIII fragment, designated Region C in Figure 20, occurs approximately 10 times in

the human genome. Since there are more genomic copies of the Region C 1.4 kb HindIII fragment than there are of the 33-derived XL3 HindIII/Xba genomic fragment, it is likely that the XbaI site occurs in only one of these fragments, specifically, the 33 fragment itself. The combined number of copies of XL3-homologous genomic DNAs, then, - Regions A + B + C - may be approximated at 10-100. It should be noted that this data does not show whether the Region A, B and C fragments occur as contiguous arrays or as non-contiguous sequences.

The placement of HindIII, EcoRI, XbaI and HpaI sites within genomic Regions A, B and C is supported by alignment of the RE maps of Subclass 1 cloned fragments homologous to XL3 DNA (Figure 20). Fragments 103 and 2 possess EcoRI sites about 1.4 kb downstream from the 5' terminal HindIII site (indicated with arrows in Figure 20) and about 700 bp downstream from an internal XbaI site. Superposition of the 103 and 2 fragment RE maps on the XL3/33 fragment RE map shows that: 1) the 5' HindIII sites of all three fragments coincide; 2) the XbaI sites of 103 and XL3 DNA coincide and 3) th3 3' (dashed line in Figure 20) HindIII site of fragment 33, 3' EcoRI site of fragment 103 and XbaI site of fragment 20A nearly coincide. Also, alignment of the homologous region of the RE map of fragment 2 with that of fragment XL3/33 demonstrates that internal PvuII sites coincide, and that an EcoRI site about 1.2 kb downstream from the PvuII site in 2 DNA coincides with the 3' terminal HindIII site of XL3/33 DNA. Taken together, these data

strongly suggest the placement of EcoRI and XbaI sites at the 3' end of the 1.4 kb XL3-homologous genomic fragment. Demonstration that a HpaI site also occurs here is not provided by the RE maps of any of the highly repetitive cloned fragments but is suggested by the Southern blot data shown in Figure 17.

The intensity of the 1.4 kb HindIII + EcoRI/XbaI/HpaI fragments suggests that these sequences comprise the major regions of homology between XL3 and genomic DNA (Figures 18 and 19).

B. Mapping of Genomic EcoRI Sites in Fragments Homologous to XL3 DNA

Construction of the genomic maps of Regions A and B allows the placement of other RE (Figures 18 and 19). Digestion of human genomic DNA with EcoRI produces a large 6.4 kb fragment homologous to XL3 upon digestion with HindIII + EcoRI, and is replaced by the 1.4 and 5.0 kb HindIII + EcoRI fragments. The 6.4 kb EcoRI fragment would be generated if an EcoRI site is placed 6.4 kb upstream from the 3' EcoRI site of Regions A and B. Digestion with HindIII + EcoRI produces the 1.4 kb HindIII + EcoRI fragment from Region A, and the 5.0 kb HindIII + EcoRI fragment from Region B, as observed (Figures 18 and 19).

It should be noted that alignment of the RE map of the homologous region of the 2 DNA with that of XL3/33 DNA suggests the existence of genomic EcoRI (and HpaI) sites in addition to those already described. However, digestion of genomic DNA with EcoRI, HindIII + EcoRI and HindIII + HpaI (Figures 17, 18 and 19) does not

reveal the fragments expected from digestion at these sites. Therefore, these fragments are assumed to be much less homologous to XL3 DNA, and genomic EcoRI and HpaI corresponding to the same sites in 2 DNA are not considered further.

C. Mapping of Genomic XbaI Sites

The major 1.4 and 5.0 kb HindIII/XbaI fragments were mapped as previously described in Regions A and B (Figure 20). A less intense band corresponding to a 750 bp HindIII + XbaI sequence represents the 750 bp HindIII/XbaI XL3 fragment itself which is mapped as shown in Region C of Figure 20 (Figures 18 and 19). Placement of external XbaI sites indicated by the 2.0, 3.4, 4.2 and 5.5 kb fragments is determined from information obtained by the HindIII + XbaI and EcoRI + XbaI digestions (Figures 18 and 19). External XbaI sites are mapped 2.0 kb, 3.4 kb and 4.2 kb upstream of the internal XL3 XbaI site of Region C and are depicted as Regions D1, D2 and D3 (Figure 20). Assignment of these fragments to independent regions D1, D2 and D3 is based on the observation that none of them occur in the HindIII + XbaI or EcoRI + XbaI digestions and are, therefore, not generated by digestion at Regions A, B or C.

One XbaI band, the large 5.5 kb fragment, is also seen in a subsequent XbaI + EcoRI digestion, and may represent a more major sequence than those of Regions D1, D2 and D3 (Figures 18 and 19). The XbaI site generating the 5.5 kb XbaI fragment is assigned accordingly to Region C at a location 5.5 kb upstream from the

internal XbaI site. This assignment supports the observation that EcoRI + XbaI digestion of human genomic DNA produces two fragments of 6.4 and 5.5 kb, respectively. The 6.4 kb band is probably the large 6.4 kb fragment is generated by digestion of Regions A and B with EcoRI, and the 5.5 kb band is the large 5.5 kb sequence produced by digestion of Region C with XbaI. These fragments are not the same, however, and are therefore assigned to different genomic map regions.

D. Mapping of Genomic PvuII Sites in Fragments Homologous to XL3 DNA

The PvuII sites are easily placed. Two bands, 5.2 and 10.4 kb in length, are visible upon digestion of human genomic DNA with PvuII (Figures 18 and 19). Placement of PvuII sites 5.2 kb 5′ and 3′ of the XL3 internal PvuII site accounts for the generation of the observed 5.2 kb band(s). Evidence that the internal PvuII site occurs genomically is seen in HindIII + PvuII digestion, which yield 1200 and 250 bp fragments (the latter is barely visible in Figure 18 and is not visible in Figure 19). The low signal intensities of these bands and the presence of the 10.4 kb PvuII band indicates that the internal PvuII site is not always present. It is certainly absent in in Region B, since digestion with HindIII + PvuII does not produce the 3.9 kb fragment that would be generated if the internal PvuII site were retained. Region B is consequently the likely origin of the large 10.4 kb PvuII fragment. As previously noted, the 250 bp HindIII + PvuII fragment hybridized

under intermediate conditions but not under stringent conditions. The absence of the 250 bp HindIII + PvuII fragment from the "stringent" blot and its presence on the "intermediate" blot supports the assignment of degenerate Satellite II sequences to this fragment, since such A+T-rich Satellite sequences may "melt off" at the higher temperature used under stringent conditions (Mitchell et al., 1979).

E. Mapping of Genomic HinfI and HaeIII Sites in Fragments Homologous to XL3 DNA

The XL3 fragment also hybridized to several fragments of HaeIII and HinfI-digested human genomic DNA. Internal HaeIII and HinfI sites are located 450 and 550 bp, respectively, downstream to the 5' HindIII site within the XL3 sequence. Digestion of genomic DNA with HindIII + HaeIII and HindIII + HinfI produced the expected 450 and 550 bp fragments (Figures 18 and 19). Other fragments observed after digestion of genomic DNA with HinfI or HaeIII may be generated by the placement of a HinfI site 1150 bp upstream of the internal HinfI site and a HaeIII site 1900 bp upstream of the internal HaeIII site. Digestion with HinfI would then generate a 1145 bp fragment; HaeIII, a 1900 bp fragment, and HaeIII + HinfI, a 1050 bp fragment (Figure 20, Region E). All of these fragments were observed (Figures 18 and 19); the 1900 bp HaeIII and 1150 bp HinfI fragments were also observed (Figure 17).

3.4 kb fragments homologous to XL3 recurred in restrictions of genomic DNA with HinfI, HaeIII and HinfI + HaeIII (Figures 18 and

19. The appearance of the 3.4 kb sequence in each case suggested that the closely-spaced internal HaeIII and HinfI sites occurred again in close proximity to each other 3.4 kb 3' or 5' to the internal sites, and these external sites were not separated from internal sites by a HindIII site, since HindIII did not digest the 3.4 kb fragment into smaller sequences. Based on these observations, the HaeIII and HinfI sites producing the 3.4 kb fragment may be aligned as shown in Region E of Figure 20. Placement of external HaeIII and HinfI sites 3.4 kb 3' to the internal sites produces the 3.4 kb fragment observed upon digestion of genomic DNA with HaeIII, HinfI, HindIII + HaeIII, HindIII + HinfI and HaeIII + HinfI. The HindIII site in the upstream portions of Regions A and C must be assumed missing in Region E, since no appropriate digestion products involving it are observed (Figures 18 and 19).

The assignment of the HinfI and HaeIII sites to Region E does not preclude the identity of Region E with any of the other major or minor Regions already discussed. Superposition of Region E on Regions A, B or C shows that the RE map of Region E is compatible with that of Region A if the 3' HindIII site of the 1.4 kb genomic fragment is not present (Figure 20). The positioning of a HindIII site at the 3' terminus of the 1.4 kb genomic fragment of Region C disqualifies this minor Region from identity with Region E, whereas the absence of that same HindIII site plus the removal of the 5' HindIII site to a position about 3.6 kb further upstream similarly

disallows identity of Region B with Region C. It seems, then, that congruence between Region E and the major Region A is the most likely of those just considered. The relative intensities of the HinfI and HaeIII-generated bands agrees with the assignment to a more major rather than minor Region, and supports the possible identity of Region E with Region A.

REFERENCES

Beauchamp, R.S., Mitchell, A.R., Buckland, R.A. and Bostock, C.J. (1979) Specific arrangements of human satellite III DNA sequences in human chromosomes. Chromosoma 71, 153-166.

Bignone, F., Panarello, C. and Ginelli, G. (1983) G-bands without pretreatment of slides, in chemically defined conditions. Hum. Genet. 63, 63-66.

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

Bostock, C.J., Gosden, J.R. and Mitchell, A.R. (1978) Localisation of a male-specific DNA fragment to a sub-region of the human Y chromosome. Nature 272, 324-328.

Burk, R.D., Szabo, P., O'Brien, S., Nash, W.G., Yu, L. and Smith, R.D. (1985) Organization and chromosomal specificity of autosomal homologs of Y chromosome repeated DNA. Chromosoma 92, 225-233.

Calabretta,B., Robberson, D.L., Barrera-Saldana, H.A., Lambrou, T.P. and Saunders, G.F. (1982) Genome instability in a region of human DNA enriched in AluI repeated sequences. Nature 296, 219-225.

Chen, E.Y. and Seeburg, P.H. (1985) Supercoil sequencing: A fast and simple method for sequencing plasmid DNA. DNA 4, 165-170.

Cooke, H.J. (1976) Repeated sequences specific to human males. Nature 262, 182-186.

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

Cooke, H.J. and McKay, R.D.G. (1978) Evolutin of a human Y chromosome-specific repeated sequence. Cell 13, 453-460.

Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corse, D., Kedinger, D. and Chambon, P. (1980) Promoter sequences of eukaryotic protein coding genes. Science 209, 1406-1413.

Corneo, G., Ginelli, E. and Poli, E. (1970) Repeated sequences in human DNA. J. Mol. Biol. 48, 319-327.

Corneo, G., Ginelli, E. and Poli, E. (1971) Renaturation properties and localization in heterochromatin of human satellite DNAs. Biochim. et Biophys. Acta 247, 528-534.

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

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

Fowler, C., Drinkwater, R. Burgoyne, L. and Skinner, J. (1987) Hypervariable lengths of human DNA associated with a human satellite III sequence found in the 3.4 kb Y-specific fragment. Nucl. Acids Res. 15, 3929.

Frommer, M., Prosser, J., Tkachuk, D., Reisner, A.H. and Vincent, P. (1982) Simple repeated sequences in human satellite DNA. Nucl. Acids Res. 10, 547-563.

Gall, J.G. (1968) Differential synthesis of of the genes for rRNA during amphibian oogenesis. Proc. Natl. Acad. Sci. USA 60, 553-560.

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

Gosden, J.R., Lawrie, S.S. and Cooke, H.J. (1981) A cloned repeated DNA sequence in human chromosome heteromorphisms. Cytogenet. Cell Genet. 29, 32-39.

Gray, K.M., White, J.W., Costanzi, C., Gillespie, D., Schroeder, W.T., Calabretta, B. and Saunders, G.F. (1985) Recent amplification of an alpha satellite DNA in humans. Nucl. Acids Res. 13, 521-535.

Hanahan, D. (1983) Studies on transformation of Escherichia coli with plasmids. J. Mol. Biol. 155, 557-580.

Hardman, N. (1986) Structure and function of repetitive DNA in eukaryotes. Biochem. J. 234, 1-11.

Helling, R.B., Goodman, H.M. and Boyer, H.W. (1974) Analysis of R. EcoRI fragments of DNA from lambdoid bacteriophages and other viruses by agarose gel electrophoresis. J. Virol. 14, 1235-1244.

Henderson, A.S. (1982) Cytological hybridization to mammalian chromosomes. Int. Rev. Cytol. 76, 1-46.

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

Houck, C.M., Rinehart, F.P. and Schmid, C.W. (1979) A ubiquitous family of repeated DNA sequences in the human genome. J. Mol. Biol. 132, 289-306.

Hwu, H.R., Roberts, J.W., Davidson, E.H. and Britten, R.J. (1986) Insertion and/or deletion of many repeated DNA sequences in human and higher ape evolution. Proc. Natl. Acad. Sci. USA 83, 3875-3879.

Hyrien, O., Debatisse, M., Buttin, G. and de Saint Vincent, B.R. (1987) A hotspot for novel amplification joints in a mosaic of Alu-like repeats and palindromic A+T-rich DNA. The EMBO J. 6, 2401-2408.

ISCN 1981. An International System for Human Cytogenetic Nomenclature - High Resolution Banding (1981). Birth Defects: Original Article Series, Vol. XVII, No. 5 (March of Dimes Birth Defects Foundation, New York, 1981).

Jeanpierre, M., Weil, D., Gallano, P., Creau-Goldberg, N. and Junien, C. (1985) The organization of two related subfamilies of a human tandemly repeated DNA is chromosome specific. Hum. Genet. 70, 302-310.

Jones, K.W. (1977) in Molecular Structure of Human Chromosomes, edited by G. Yunis. Academic Press, New York., pp 295-323.

Jones, K.W. and Corneo, G. (1971) Location of satellite and homogenous DNA sequences on human chromosomes. Nature 233, 268-271.

Kunkel, L.M., Smith, K.D. and Boyer, S.H. (1976) Human Y chromosome-specific reiterated DNA. Science 191, 1189-1190.

Kunkel, L.M., Smith, K.D. and Boyer, S.H. (1977) Molecular organisation of the human Y chromosome. Am. J. Hum. Genet. 29, 65a.

Kunkel, L.M., Smith, K.D. and Boyer, S.H. (1979) Organization and heterogeneity of sequences within a repeating unit of human Y chromosome deoxyribonucleic acid. Biochemistry 18, 3343-3353.

Kunkel, L.M. and Smith, K.D. (1982) Evolution of human Y-chromosome DNA. Chromosoma 86, 209-228.

Lehrman, M.A., Schneider, W.J., Sudhof, T.C., Brown, M.S., Goldstein, J.L. and Russel, D.W. (1985) Mutation in LDL receptor: Alu-Alu recombination deletes exons encoding transmembrane and cytoplasmic domains. Science 227, 140-146.

Lewin, B. in Gene Expression 2: Eukaryotic Chromosomes, 2nd Edition. John Wiley and Sons, 1982.

Maniatis, T., Jeffrey, A. and Kleid, D.G. (1975) Nucleotide synthesis of the rightward operator of phage lambda. Proc. Natl. Acad. Sci. USA 72, 1184.

Maniatis, T., Fritsch, E.F. and Sambrook, J., in Molecular Cloning: A Laboratory Approach. Cold Spring Harbor Laboratory Press, 1982.

Manuelides, L. (1978) Complex and simple sequences in human repeated DNAs. Chromosoma 66, 1-21.

Manuelides, L. (1982) Nucleotide sequence definition of a major human repeated DNA, the HindIII 1.9 kb family. Nucl. Acids Res. 10, 3211-3219.

Manuelides, L. and Biro, P.A. (1982) Genomic representation of the HindIII 1.9 kb repeated DNA. Nucl. Acids Res. 10, 3221-3239.

Manuelides, L. and Ward, D.C. (1984) Chromosomal and nuclear distribution of the HindIII 1.9 kb human DNA repeat segment. Chromosoma 91, 28-38.

Marmur, J. (1961) A procedure for the isolation of deoxyribonucleic acids from microorganisms. J. Mol. Biol. 3, 208-218.

McDonell, M.W., Simon, M.N. and Studier, F.W. (1977) Analysis of the restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels. J. Mol. Biol. 110, 119-146.

McKnight, S.L. and Kingsbury, R. (1982) Transcriptional control signals of a eukaryotic protein-coding gene. Science 217, 316-324.

Meneveri, R., Agresti, A., Della Valle, G., Talarico, D., Siccardi, A.G. and Ginelli, E. (1985) Identification of a human clustered G+C-rich DNA family of repeats (Sau3A family). J. Mol. Biol. 186, 483-489.

Mermer, B., Colb, M. and Koontiris, T.D. (1987) A family of short, interspersed repeats is associated with tandemly repetitive DNA in the human genome. Proc. Natl. Acad. Sci. USA 84, 3320-3324.

Mitchell, A.R., Beauchamp, R.S. and Bostock, C.J. (1979) A study of sequence homologies in four satellite DNAs of Man. J. Mol. Biol. 135, 127-149.

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

Mottez, E., Rogan, P.K. and Manuelides, L. (1986) Conservation in the 5' region of the long interspersed mouse L1 repeat: implications of comparative sequence analysis. Nucl. Acids Res. 14, 3119-3136.

Moyzis, R.K., Albright, K.L., Bartholde, M.F., Cram, L.S., Deaver, L.L., Hildebrand, C.E., Joste, N.E., Longmire, J.L., Meyne, J. and Schwartzacher-Robinson, T. (1987) Human chromosome-spedific repetitive DNA sequences: novel markers for genetic analysis. Chromosoma 95, 375-386.

Prosser, J., Reisner, A.H., Bradley, M.L., Ho, K. and Vincent, P.C. (1981) Buoyant density and hybridization analysis of human DNA sequences including three satellite DNAs. Biochim. Biophys. Acta 656, 93-102.

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

Proudfoot, N.J. and Brownlee, G.G. (1976) 3′ noncoding region sequences in eukaryotic mRNA. Nature 263, 211-214.

Rinehart, F.P., Ritch, T.G., Deininiger, P.L. and Schmid, C.W. (1981) Renaturation rate studies of a single family of interspersed repeated sequences in human deoxyribonucleic acid. Biochemistry 20, 3003-3010.

Rubin, C.M., Houck, C.M., Deininger, P.L., Friedmann, T.and Schmid, C.W. (1980) Partial nucleotide sequence of the 300-nucleotide interspersed repeated human DNA sequences. Nature 284, 372-374.

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

Schimke, R.T. (1984) Gene amplification in cultured animal cells. Cell 37, 705-713.

Schindler, C.W. and Rush, M.G. (1985) The KpnI family of long interspersed nucleotide sequences is present on discrete sizes of circular DNA in monkey (BSC-1) cells. J. Mol. Biol. 181, 161-173.

Schmeckpeper, B.J., Willard, H.F. and Smith, K.D. (1981) Isolation and characterization of cloned human DNA fragments carrying reiterated sequences common to both autosomes and the X chromosome. Nucl. Acids Res. 9, 1853-1873.

Schmid, C.W. and Deininger, P.L. (1975) Sequence organization of the human genome. Cell 6, 345-358.

Schmid, C.W. and Jelinek, W.R. (1982) The Alu family of dispersed repetitive sequences. Science 216, 1065-1070.

Shafit-Zagardo, B., Brown, F.L., Maio, J.J. and Adams, J.W. (1982a) KpnI families of long, interspersed repetitive DNAs associated with the human B-globin gene cluster. Gene 20, 397-407.

Shafit-Zagardo, B., Maio, J.J. and Brown, F.L. (1982b) KpnI families of long, interspersed repetitive DNAs in human and other primate genomes. Nucl. Acids Res. 10, 3175-3193.

Sharp, P.A., Sugden, B. and Sambrooke, J. (1973) Detection of two restriction endonuclease activities in Haemophilus parainfluenzae using analytical agarose. Biochemistry 12, 3055-3063.

Shmookler Reis, R.J., Lumpki, C.K.,Jr., McGill, J.R., Riabowol, K.T. and Goldstein, S. (1983) Extrachromosomal circular copies of an "inter-Alu" unstable sequence in human DNA are amplified during in vitro and in vivo ageing. Nature 301, 394-398.

Singer, M. (1982) Highly repeated sequences in mammalian genomes. Int. Rev. Cytol. 76, 67-113.

Smith, G.P. (1976) Evolution of repeated DNA sequences by unequal crossovers. Science 191, 528-535.

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

Tsukada, T., Watanabe, Y., Nakai, Y., Imura, H., Nakanishi, S. and Numa, S. (1982) Repetitive DNA sequences in the human corticotropin-B-lipoprotein precursor gene region: Alu family members. Nucl. Acids Res. 10, 1471-1479.

Vinograd, J. and Hearst, J.E. (1962) Progr. Chem. Org. Nat. Prod. 20, 395.

Willard, H.F. and Waye, J.S. (1987) Hierarchical order in chromosome-specific human alpha satellite DNA. Trends in Genetics 3, 192-198.

Wu., J.C. and Manuelides, L. (1980) Sequence definition and organization of a human repeated DNA. J. Mol. Biol. 142, 363-386.

AUTOBIOGRAPHICAL STATEMENT

Ms. Macoska is originally from Cleveland, Ohio. She graduated Magna Cum Laude and Phi Beta Kappa from Kent State University with a B.A. in Physical Anthropology in 1978. She was awarded the the Beatrice Konheim Graduate Scholarship in the Life Sciences from Hunter College in 1985. Ms. Macoska received a Ph.D. in Biochemistry from the City University of New York in 1988 under the sponsorship of Dr. Ann Henderson at Hunter College. She will pursue postdoctoral work in the laboratory of Dr. Matthew Meselson in the Department of Biochemistry and Molecular Biology at Harvard University.