ISOLATION AND CHARACTERIZATION OF HUMAN HIGHLY REPEATED SATELLITE II AND III DNAs

by

Katia Sol

Department of Microbiology and Immunology
McGill university
Montreal, Quebec, Canada

A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

To Jacky Sol-
a mother in all the best ways

To Jack and Ki-
and the rest of my family

ABSTRACT

We have isolated, cloned, and sequenced, from the genome of human (HeLa) cells, a 1.797 Kb EcoR1 satellite II DNA fragment that displayed partial identity to the Drosophila melanogaster transposable P-element.

The sequence analysis of this clone (pKS36) indicated that it has originated from the tandem amplification of pentameric repeats that were derived from satellite II and III canonical consensus sequence 5' TTCCA 3'. A two-stage decay mechanism, based on the methylation and subsequent deamination of cytosine residues within the pentamers, was proposed to explain the non-random base substitutions that were observed in pKS36, as well as in other sequenced members of the satellite II and III DNA families. In addition, this two-stage decay model correlated with the Taq1 and Hinf1 polymorphisms that were observed between related satellite DNA members.

Clone pKS36 also contains a region of 49 bp devoid of satellite sequences that was found, by southern hybridization, to be present in pKS36 closely related (pKS36-like) but absent from more divergent (pKS36-related) cloned satellite II and III DNAs. pKS36-related satellite DNAs represent up to 2% of the genomes of HeLa and MeWo cells, and are organized mostly in tandem arrays of 1.8 Kb EcoR1, Kpn1, and Sau3A DNA fragments. pKS36-like satellite DNAs represent less than 1 % of the genome of HeLa cells, and are found mainly organized as 1.65 Kb, 1.95 Kb and 3.6 Kb EcoR1 elements, though their Kpn1 and Sau3A distributions resemble that of pKS36-related satellite DNAs. Cell specific organization of satellite DNAs, that may be the result of chromosomal translocations inherent to cultured cells, was observed in the two human (HeLa and MeWo) cell lines.

The analysis, by southern hybridization, of satellite DNAs using field inversion gel electrophoresis revealed the presence, in HeLa cells, of satellite DNA clusters ranging from 150 Kb to 500 Kb in length.

Using rodent-human hybrid cell DNAs,the members of the pKS36 satellite II DNA family were found to reside mainly on human chromosomes 7, 12, 14, 15, 16, and 22.

RÉSUMÉ

Nous avons isolé, clôné et séquencé un fragment EcoR1, long de 1.797 Kb, possédant des régions d'homologie partielle avec l'élément transposable P de la mouche à vin, **Drosophila melanogaster.**

Ce clône (pKS36) est composé de séquences pentamériques, organisées en tandem, qui sont dérivées de la séquence consensus des ADN satellites II et III: 5' TTCCA 3'. Le polymorphisme Taq1 et Hinf1 observé chez les membres des ADN satellites, ainsi que la présence de bases souvent substituées au sein des pentamères (telles les deux cytosines), peuvent être expliqués par un phénomène de dégénérescence basé sur une forte méthylation des résidus cytidiques et de leur subséquente déamination.

Le clône pKS36 contient une région de séquence non-satellite de 49 paires de base, également identifiée, par hybridation "Southern", dans des clônes d'ADN satellite qui sont homologues à pKS36 (satellites **homologues**). Cette région n'est pas présente chez les ADN satellites qui sont partiellement homologues à pKS36 (clônes **semi-homologues**). Ces derniers représentent 2 % des génômes des cellules **HeLa** et **MeWo** et sont organisés en tandem de 1.8 Kb sur des fragments EcoR1, Kpn1 et Sau3A. Les ADN satellites homologues à pKS36 représentent moins d'un pour cent du génôme des cellules **HeLa**, et sont organisés, pour la plupart, sur des fragments EcoR1 de diverses tailles: 1.65 Kb, 1.95 Kb et 3.6 Kb. Leurs organisations, en tant que fragments Kpn1 et Sau3A, se rapprochent de celles des clônes semi-homologues à pKS36. Le polymorphisme, observé dans l'organisation des ADN satellites des deux lignées cellulaires humaines analysées (**HeLa** et **MeWo**), pourrait être le résultat de translocations de chromosomes, phénomènes de prédilection de cellules maintenues en culture.

La technique d'électrophorèse par inversion de champs a permis l'identification, au sein des cellules **HeLa**, de larges segments de chromosomes (contenant des ADN satellites homologues à pKS36), qui varient en taille de 150 Kb à 500 Kb.

L'analyse, par hybridation, des ADN provenant de cellules hybrides semble indiquer que la majorité des ADN satellites II se trouvent sur les chromosomes humain 7, 12, 14, 15, 16 et 22.

ACKNOWLEDGEMENTS

I am thankful to the staff at the University Paul Sabatier in Toulouse (France), for their encouragement and training. A special thank to Dr. Jean Pierre Zalta for introducing me to the fascinating field of molecular biology, and for his three years of excellent teachings.

I will always long for the fine environment that was provided during my tenure in the Department of Microbiology and Immunology at McGill University.

I would like to express my appreciation to my thesis advisor, Dr. Michael S. DuBow, for his friendship and guidance during the five years I have spent in his laboratory. His wisdom will always be integrated in my scientific thinking, and in many other aspects of my life as well. I know at times he would have liked to throw me out the window and I am glad he did not. His patience with my bursts of energy was exemplary.

I owe much gratitude to the staff members of my department. I am thankful to Drs. Teresa Keng, Michael Ratcliffe, and Nick Acheson for their help in solving particular experimental problems that I encountered during my research. I have greatly appreciated the many "philosophical" talks shared with Drs. Robert Murgita and Dal Briedis. I will cherish forever the friendship of Ann Nudo, and I thank her for smoothing out all my administrative paper work.

I could not have done it without all the love and care of my fellow lab-mates. From the "old gang" were Josée Harel, David Levin, Peter Tolias, Annie Kupelian, Alan Jarjour, Daphne Goring, Martine Lapointe, Mike MacLeod, Manon Lalumière, Linda Dupléssis, Paul Zanga, Phillip Marshall, and Kalpana Gupta. From the new crop of the DuBowian family are Robin Cameron, Chantal Autexier, George "Yogi" Kukolj, Peter "Ardvark" Ulycznyj, Lina Guzzo, Josée Brisebois, Marina Pascali, Joy Marmor, Farnas Forghani, and Alain Biguet. Science would not be fun without friends to share with.

I would especially like to thank George "Hap" Szatmari, Caroline "Math" Diorio and Claire Fouquet. Their long lasting friendships, help and constructive criticism have been instrumental in my successful research project and in my development both as a scientist and as a person.

Many other people deserve recognition for the help and/or friendship they have shown me: Louise Blois, Robert Giasson, Drs. Luc Dégroseiller, Alassane Séré, David Garfinkel, and Malcolm Casadaban (and his family).

The writing of this thesis would not have been possible without the counseling and encouragement of my colleagues at the DuPont de Nemours experimental research station: Drs. Brian Sauer, Karen Skorupski, and Shinichi Fukushigi. A very special thank to my post-doctoral advisor, Dr. Nat L. Sternberg, for his constant support and patience.

This work has been supported by a Scholarship provided by World University Service of Canada. The editing of this thesis

was supported, in part, by a post-doctoral fellowship from the National Institutes of Health.

CONTRIBUTIONS TO ORIGINAL KNOWLEDGE

1- We have identified and isolated from the genome of the human cell line HeLa, a 1.8 Kb EcoR1 DNA fragment (in plasmid pKS36), that displayed partial homology to a known **Drosophila melanogaster** transposon: the P-element.

2- We sequenced the entire 1.8 Kb insert and determined that we have presented the first complete nucleotide sequence of a long member of the highly repeated human family of satellite DNA. This EcoR1 DNA fragment contains the characteristic pentameric element, 5' TTCCA 3', of the human satellite II and III DNAs, and it is also interrupted in its pentameric repeats by a 49 bp unique, non-satellite DNA region, called the 49 mer.

3- Using this 49 mer, we were able to distinguish between this class and other members of the closely related families of human satellite II and III DNAs.

4- We cloned other members of the human satellite families and sequenced their extremities. We determined, via the analysis of their characteristic Taq1 and Hinf1 digestion patterns, their memberships to either satellite II or III DNAs.

5- DNA sequence analysis revealed the presence, within satellite DNA, of hotspots for mutations. A model that explains

how satellite DNA may have evolved from the consensus 5' TTCCA 3' repeats, was presented.

6- We determined the genomic organization of the pKS36 satellite II family, and discovered that many of these sequences are apparently dispersed within the human genome, as well as organized as tandem arrays of 1.8 Kb EcoR1 and Kpn1 families. We also determined the presence, in the genome of HeLa cells, of large blocs of satellite DNAs, ranging from 150 Kb to 500 Ko, bounded by HindIII sites.

7- We have determined that our cloned satellite DNA represents less than 1 % of the genome of HeLa cells, and that other related 5' TTCCA 3' satellite DNA make up 2 % of the genome.

8- Using a human-rodent hybrid cell panel, we have found that these satellite sequences are mainly located on human chromosomes 7, 12, 14, 15, 16, and 22.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

TABLES.....PAGE #

FIGURES.....PAGE #

Chapter I

INTRODUCTION

Rêver un impossible rêve
Porter le chagrin des départs
Brûler d'une possible fièvre
Partir ou personne ne part

Jacques Brel (1968)

I- DNA-HEREDITY THEORY AND CHROMOSOME STRUCTURE

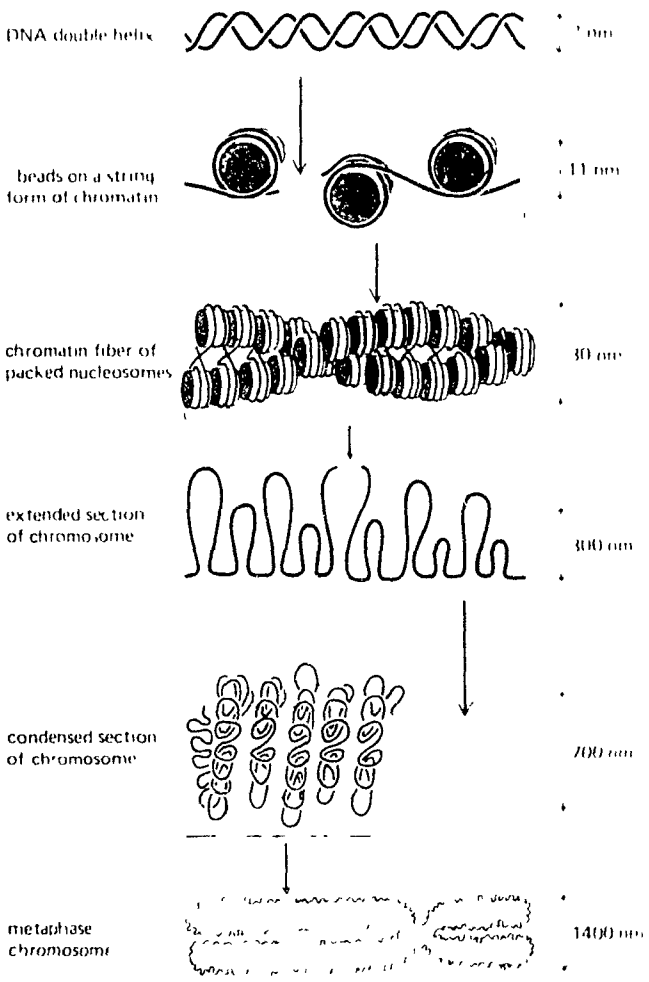
Deoxyribonucleic acid (DNA) is the molecule for the genetic information of most life forms that exist on earth today. The organization of the information contained in a DNA molecule is different for every species and often differs in particular sequences even between organisms of the same species. According to the DNA-heredity theory, it could be reasonable to presume that the more complex an organism is, the more genes it contains. Therefore, it is not surprising that, as an alternative to the simple chromosome organization of prokaryotes, eukaryotes have developed a different way for handling tremendous amount of genetic information. One of the distinguishing feature of eukaryotes is the structure of their chromosomes. In every eukaryotic cell, the complete genetic blueprint of an organism is tightly compacted in chromatin structures formed by the close and periodic association of DNA and proteins. Histones and scaffolding proteins provide a highly efficient way to wrap the DNA content of a cell into a length one ten-thousandth of its natural length. In a single human cell, 6 billion base pairs of DNA wind around histone octamers into structural subunits called nucleosomes (Alberts et al., 1983). The chromatin fibres are in turn folded and coiled into a higher level of organization, the chromosomes (figure 1).

This highly sophisticated packaging mode enables the accurate replication, segregation into daughter cells, expression and regulation of the genetic material contained in euchromatic regions. Heterochromatic regions, the most highly condensed region of chromosomes, are located predominantly at the tips (telomeres) and around the centromeres. The replication of both telomeric and centromeric heterochromatin is programmed in a sequence and species specific manner during the S-phase of cell division (McCarroll and Fangman, 1988; Ten Hagen et al., 1990).

The chromosomes of eukaryotes display an extreme flexibility. During meiosis, the chromosomes can be broken apart and recombined, thus creating constant dynamic changes. Errors during DNA replication (though infrequent), sequence rearrangements mediated by mobile genetic elements or direct repeats, and meiotic gene shuffling, provide the necessary variations from which evolutionary forces can act and select from.

Figure 1: Levels of DNA organization in chromosomes. (from figure 8-24, page 399 in "The molecular biology of the cell" Alberts et al., 1983). The association of histone proteins with DNA leads to the formation of the first level of DNA packing: the 10 nm (nanometer) nucleosome fibers. Condensation of nucleosomes into 30 nm fibers is mediated by histones H1 and H5. The organization of chromatin into looped domains (euchromatin) regulates transcriptional activity of genes. Highly condensed domains, that appear as dark bands on mitotic chromosomes, constitute the constitutive or facultative heterochromatic regions of chromosomes.

Figure 1


DNA double helix
beads on a string form of chromatin
chromatin fiber of packed nucleosomes
extended section of chromosome
condensed section of chromosome
metaphase chromosome

II- THE C-VALUE PARADOX

In eukaryotes, the flow of genetic information proceeds from a DNA template, to a RNA transcript, to a protein product. However, in the last 30 years, molecular exploration of the "building blocks of life", both within and between eukaryotic species, has revealed some discrepancies to this DNA-heredity theory. These are generally referred to, by a puzzled scientific community, as the C-value paradox (Gall, 1981).

One of the first observations made was that the total amount of DNA of a haploid eukaryotic cell (the C-value) is not always proportional to the apparent phenotypic complexity of a given organism. For example, certain amphibians contain twenty times as much DNA as man does.

Moreover, the C-value has shown a great deal of variation between closely related species. Methods of chromosome analysis, such as the C-banding technique (Arrighi and Hsu, 1971), have demonstrated enormous variability in the amount of heterochromatin (mostly centromeric) in individuals of the same (or phenotypically related), species (Pardue and Gall, 1972). In amphibians, the genome size between two individuals of the same species may range from 10^9 bp to almost 10^{11} bp (Alberts et al., 1983).

The last aspect of the C-value paradox, and probably the most striking, comes from the realization that the amount of DNA in a single cell is greater than would be predicted from the number of its genes. In humans, it is believed that less than 2 %

of the DNA encodes proteins. In any particular eukaryotic cell, the bulk of the DNA contains information that is apparently never used. It stays, for the most part, locked into closed untranscribed heterochromatic structures (constitutive heterochromatin). Yet, these non-coding sequences are generally preserved and passed on from generation to generation.

III- NATURE OF THE EXCESS DNA

We now know that some of the excess DNA found in eukaryotes is accounted for because genes are much larger than the sequences needed to code for their proteins. The flow of information from DNA to protein must be regulated in some way. Eukaryotic genes include, within and proximal to their coding sequences, elements that regulate their expression (promoters, enhancers) and their transcription (non-translated regions, introns). A fraction of the excess DNA can also be accounted for by the presence of pseudogenes (truncated nonfunctional versions of genes). However, the beginning of an answer to the C-value paradox started to emerge in the 1960s and 1970s with the development and application of technologies such as equilibrium density gradient centrifugations (Vinograd and Hearst, 1962), reassociation kinetics (Britten and Kohne, 1968; Southern, 1971) and restriction endonuclease analysis (Southern, 1975a, 1975b; Pech et al., 1979), which revealed that most of the

non-genic component of an eukaryotic genome is made of simple and complex repetitive DNA sequences.

III-1 Equilibrium density gradients via ultracentrifugation.

This method is based on the physical properties of DNA molecules, subjected to centrifugation at high speed through a cesium chloride (CsCl) density gradient, to form a band at a position corresponding to its own density (Meselson et al., 1957). The position of the band reflects the average G-C base composition of the DNA, (Schildkraut et al., 1962). In 1961, while studying the distribution in CsCl density gradients of mouse (**Mus musculus**) DNA, Kit observed a minor DNA fraction separating from the main band. Owing to its original detection as a distinct component of nuclear DNA, the fraction was termed satellite DNA (Kit, 1961).

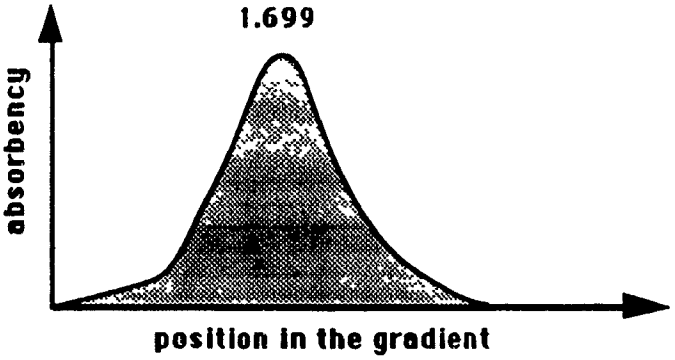
The subsequent use, in Cs_2SO_4 gradients, of heavy metals, dyes, and antibiotics that bind preferentially to either A-T or G-C rich sequences, allowed a better resolution of satellite DNA fractions in almost all eukaryotes (Skinner and Beattie, 1973; Peacock et al., 1974; Manuelidis, 1977.

Using silver ions (Ag^{+}) in Cs_2SO_4 gradients, Corneo and his coworkers (1971) demonstrated the presence, in humans, of three satellite fractions, each characterized by its own buoyant density (figure 2).

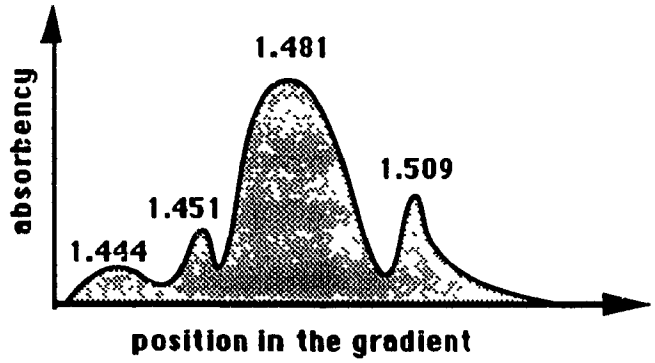
Figure 2: Distribution of human DNA in $CsCl_2$ (a) and Ag^+-Cs_2SO_4 (b) gradients. (a) In Neutral Cesium gradient the bulk of human DNA has a density of 1.699. (b) In Ag^+-Cs_2SO_4 the main band density is 1.481, and the satellite DNA fractions appear with densities of 1.444, 1.451, 1.509. Redrawn from Hearst et al., 1974 and Corneo et al., 1971, (a) and (b), respectively

Figure 2

a) Neutral CsCl2 gradient



b) Ag+-Cs2SO4 gradient



In neutral CsCl gradients (figure 2, diagram a), the density of the cryptic satellite DNAs coincide with that of the major component (Corneo et al., 1971; Hearst et al., 1974). In Ag^{+}-Cs_2SO_4 gradient, three distinct satellite DNA fractions, characterized by buoyant densities heavier or lighter (GC- or AT-rich, respectively) than that of the main band, were resolved (figure 2, diagram b).

While satellite peaks were found to be ubiquitous in eukaryotes, the nature of the DNA they contained remained unclear (except for their overall AT- or GC-rich nature). It was the application of reassociation kinetics and restriction site analysis that revealed that these fractions were made of repetitive DNA sequences.

III-2 DNA reassociation kinetics

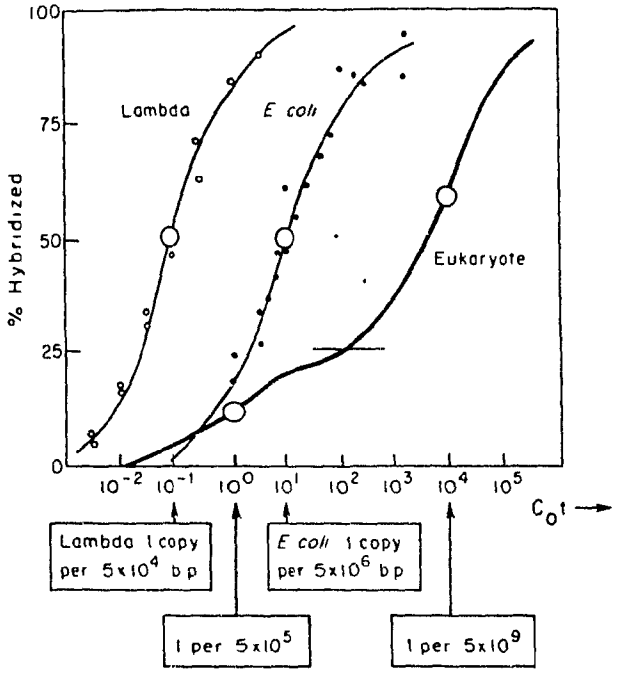
Reassociation of denatured DNA strands depends on random collision of complementary molecules. Thus, the rate of duplex formation is not only controlled by the length and complexity of the DNA, but also by its concentration (number of identical molecules), and by the time the annealing reaction is allowed to proceed (Britten and Kohne, 1968). Within a heterogeneous population of sequences, DNA molecules present in many copies have greater chances, via random collision, of rapidly finding and annealing to their complementary counterpart, than sequences that are present only once. Thus, the repetition frequency of DNA sequences can be monitored by the rate of their reassociation kinetics. Britten and his

coworkers showed that the reassociation rate of DNA chains can be characterized by the so called Cot value: the product of the DNA concentration [Co], and the time of incubation [t] (Waring and Britten, 1966; Britten and Kohne, 1968). The larger the Cot value, the slower the rate of duplex formation.

Figure 3 illustrates reassociation kinetics of genomes of diverse sequence complexity: Lambda phage, E. coli and a hypothetical eukaryote. For lambda 50 % renaturation (hybridization) of the molecules is observed at a Cot value ($Cot_{1/2}$) of 10^{-1}. E. coli, on the other hand, achieves 50 % renaturation at a $Cot_{1/2}$ value of 10, or 100 times more slowly than Lambda. This reflect directly the difference in bp content of the two DNAs: E. coli has 100 times more DNA as Lambda (50 Kb). In the case of the eukaryote, the Cot curve reflects a more complex pattern in the organization of the DNA. 25 % of the genomic DNA hybridizes at a $Cot_{1/2}$ value of 1, or 10 times faster than E. coli, whereas 75 % hybridizes at a $Cot_{1/2}$ value of 10^4. Thus, it appears that this hypothetical eukaryote contains 25 % of DNA sequences that are repeated a thousand times,on average, and that the rest of the DNA is made of single copy DNA sequences. Typically, in lower eukaryotes, 10-20 % of the DNA renatures at a fast rate. In animal cells, up to 50 % of the DNA is repetitive, whereas in plants and amphibians up to 80 % of the total hereditary material is composed of repeated DNA sequences.

Figure 3: Examples of Cot curves. (from Figure 2, page 277 in "Chromosome structure and level of organization" Ris and Korenberg, 1979). Reassociation kinetics of λ phage, E. coli and a hypothetical eukaryote. Circles positioned on the curves correspond to the Cot value at 50 % hybridization ($Cot_{1/2}$). See text for explanations.

Figure 3


100
75
50
25
0
% Hybridized
Lambda
E coli
Eukaryote
10^{-2} 10^{-1} 10^{0} 10^{1} 10^{2} 10^{3} 10^{4} 10^{5}
C_0t →
Lambda 1 copy per 5×10^4 bp
E coli 1 copy per 5×10^6 bp
1 per 5×10^5
1 per 5×10^9

Britten and Kohne, (1968) showed that mouse satellite DNA renatured more rapidly than the rest of the nuclear DNA, thus demonstrating the repetitive nature of the DNA sequences present in satellite fractions.

III-3 Restriction endonuclease analysis

If a sequence is sufficiently abundant, fragments generated by digestion of the DNA with restriction endonucleases can be observed, upon staining with ethidium bromide, over the background of genomic DNA fragments. This property is illustrated in figure 4. When human genomic DNA is hydrolyzed with KpnI, EcoRI, HindIII or HaeIII (figure 4, lanes A, B, C and D, respectively), discrete bands (indicated by arrows), characteristic of the restriction endonuclease utilized, stand out over the background stain of diverse-sized DNA fragments.

At partial digestion of genomic DNA, it was observed that repeated DNA sequences fall into two types of organizations:

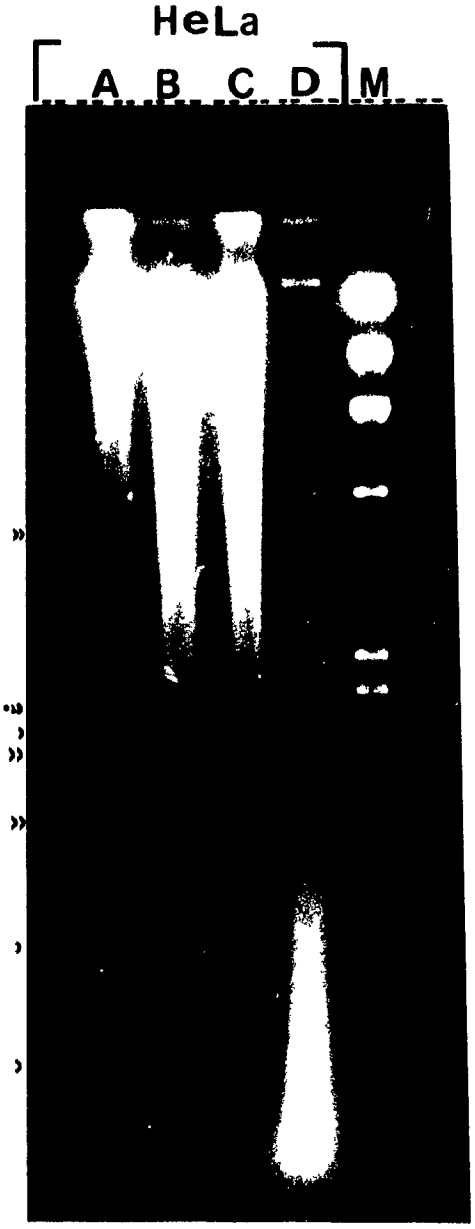
-sequences that are organized in a head to tail manner, show a characteristic ladder of fragments that are arithmetically related. The length of the tandem arrays may be characteristic of the restriction endonuclease utilized, as well as specific for the genome under study. In figure 4, this type of organization is observed after digestion (to completion) of human DNA with EcoRI (lane B). The digestion produces a heavily stained band 340 bp in length, as well as its dimeric form (680 bp) The dimer results from point mutations occuring at the EcoRI

site between two adjacent units. This EcoRI family of tandemly arrayed DNA fragments is characteristic of primate genomes, and is referred to as the EcoRI alphoid family (section IV-1.2, and Maio et al., 1981a). The term "family" defines a population of evolutionary divergent, though related, DNA sequences. Thus, members of a family that have arisen from the duplication of a common ancestor may show high degree of polymorphism at the sequence and/or organization levels.

-Sequences that are dispersed or linked to unrelated DNA appear as discrete diverse-sized fragments. In humans,the most abundant interspersed repetitive elements belong to the Alu and Kpn (or L1) families (Singer, 1982a). Both families have been named after the specific restriction endonuclease with which they were originally detected. Figure 4, displays the dispersed organization of a series of human KpnI repetitive DNA sequences (lane A) The four major fragment sizes yielded (1.2 Kb, 1.5 Kb, 1.8 Kb and 1.9 Kb), are characteristic of all primate genomes (except prosimians). Owing to this long period interspersion mode, the human KpnI elements are refered to as the human LINES (long Interspersed sequences) family. Using the restriction endonuclease AluI, a second group of repetitive DNA, characterized by shorter units (300 bp) of dispersed sequences was observed. The Alu elements are also refered to as the human SINES (short interspersed sequences) family.

Figure 4: Restriction endonuclease analysis of HeLa genomic DNA. 10 μg of HeLa DNA were hydrolyzed with restriction endonuclease Kpn1 (lane A), EcoR1 (lane B), HindIII (lane C), and HaeIII (lane D). As molecular weight marker, 2 μg of λ phage DNA were hydrolyzed with HindIII (lane M). The DNA fragments were then separated via gel electrophoresis, as indicated in Chapter II (materials and methods). The agarose gel was subsequently soaked in a solution of ethidium bromide, and the DNA fragments were visualized under a high intensity-UV source. Arrowheads point to the characteristic discrete bands formed by repeated DNA sequences (see text).

Figure 4



IV- ANALYSIS OF REPETITIVE DNA FAMILIES IN EUKARYOTES

Repetitive DNA may be classified in terms of copy number (middle Vs highly repetitive) and organization (tandem Vs interspersed), as discussed earlier, but they also may be categorized in terms of sequence complexity. The sequence complexity is accounted for by the length of the unit of reiteration. Thus, sequences whose basic unit of repetition range from 2 bp to 60 bp are called "simple", and those with longer repeat units are called "complex".

IV-1 Complex repetitive sequences

IV-1.1 Mobile elements

In the early 1950s, while studying the changes in pattern of pigmentation of leaves and kernels of maize, Barbara McClintock came upon the first observation of the presence of "jumping genes" at work in eukaryotes. The transposable elements she described (Ds-Ac family) consist of two related elements: an autonomous activator (Ac) that is able to trans-activate a related, though non-autonomous, dissociation element (Ds) (McClintock, 1951, 1956).

Since this initial discovery, transposable elements have been found to be abundant and unstable residents of the genome of most eukaryotes. DNA transposition appears to be a

ubiquitous pathway of illegitimate recombination among all life forms. The molecular mechanisms by which these elements move involves either replicative transposition via reverse transcription and integration (mammalian retroviruses and Alu sequences, **Drosophila** copia and **Saccharomyces cerevisiae** Ty elements), or conservative integration (maize Ds-Ac family, **Drosophila** P-elements). Upon insertion into a gene, the mobile element may cause mutations by insertional inactivation, but can also alter the genetic material both functionally, by affecting the expression or regulation of adjacent genes, and structurally, by promoting deletions, inversions or translocations of neighboring DNA sequences. Transposable elements may exist extra-chromosomally or be resident within the host genome. It is the extremities of the elements that contain the primary cis-acting determinants for DNA transposition.

IV-1.1;1 Drosophila P-elements

In eukaryotes, transposable elements have been implicated as causative agents in the generation of genetic defects and tumor induction.

In **Drosophila melanogaster**, P-elements have been found to be the causative agent of hybrid dysgenesis (Kidwell and Kidwell, 1976; Bingham et al., 1982). This phenomenon occurs during mating of a male from a strain carrying P-elements (P-strain) in its genome with a female lacking such elements (M-strain). After fertilization, the P-elements carried by the male genetic material are induced to transpose in the

zygote due to the lack of any regulatory functions in the fertilized egg. The subsequent hyper transposition induces chromosomal aberrations and distorted segregation in the progeny and often results in offspring sterility.

The P-elements range from 500 to 2900 bp and are present on average in approximately 50 copies per cell. They contain a 31 bp inverted terminal repeat sequence and induce an 8 bp direct duplication of the host target DNA at the site of insertion. The full-length 2.9 Kb element may be compared to the Ac element found in maize as it can **trans**-activate, in a tissue specific way, truncated non-autonomous forms of the element.

IV-1.1;2 SINES elements

Alu elements are short interspersed repetitive DNA sequences, half of which are cleaved by the restriction endonuclease AluI. The human Alu family consists of roughly 500,000 members per genome, each of which is 300 bp long, and most of which are transcriptionally inactive (Deininger et al., 1981; Deininger and Daniels, 1986). Alu elements can be organized into sub-classes of related, though evolutionary divergent, sequences (Britten et al., 1988, Quentin, 1988). In humans, the average homology between Alu members is 87 %, and they all share a 14 bp core sequence that was found to be almost identical to the SV40 origin of replication. Apparently, Alu elements were derived, by deletion of a central segment, from the 7SL RNA, a class III gene product that was found to be an essential component of a small ribonucleoprotein particle that

mediates protein secretion from cells (Ullu and Tschudi, 1984). When compared to the total human genome, SINES elements appear enriched in both unmethylated GpC dinucleotides and 5-methylcytosines (Gama-Sosa et al., 1983). Members of the Alu family are flanked by short direct repeats 10 to 19 bp long (Pan et al., 1981), and contain an internal promoter for RNA polymerase III and a 3' poly (A) tail. Thus, they appear to have been amplified and dispersed by an RNA intermediate and as such, are often refered to as retroposons (Weiner et al., 1986).

IU-1.1;3 LINES elements

Ubiquitous in mammals, LINES-elements are long and complex interspersed repetitive DNA sequences. In primates, they were first observed by virtue of their characteristic series of KpnI DNA fragments (Maio et al., 1981b). LINES elements are very polymorphic in length and sequence: in primate genomes, KpnI-like sequences may vary in size from 70 bp to 6400 bp. The short variants of the KpnI family are due to truncation at the 5' end of the element (Shafit-Zagardo et al., 1982). The elements are characterized by the presence of large ORFs, showing homology to the main components of a retroviral proviral genome (LTR, gag and pol genes), a 3' poly A track, and 14 bp direct repeats. Singer and her co-workers found (in African green monkey [AGM]) a truncated KpnI element, 829 bp long, inserted in one of the units of tandemly organized alpha satellite DNA sequences. This element, KpnI-RET is thought to have transposed to this location after the amplification of the

tandem arrays (Thayer and Singer, 1983). Other members have been found at the junction of AGM alpha satellite DNA (Grimaldi and Singer, 1983). Though the KpnI organization of the LINES elements is a characteristic of primate genomes, related sequences are found in other genera. In filamentous fungi, a transposon similar to the I-elements of **Drosophila** and to the LINES elements of mammals was found (Kinsey and Helber, 1989).

IV-1.2 Alphoid DNA

The centromeric DNA of higher eukaryotes contains long tandem arrays of highly repetitive sequences, the best characterized being the family of alphoid satellite of primates (Maio,1971; Kurnit and Maio, 1974; Singer, 1979, 1982 b; Choo *et al.*, 1990a; Mahtani and Willard, 1990). The African green monkey alpha satellite DNA comprises 20-25 % of the genome (Kurnit and Maio, 1974), and is organized in a chromosome specific manner as long tandem arrays of 172 bp repeat units. Sequences that are related to the AGM alpha satellite were detected in other primates and called, by analogy, alphoid satellite DNAs.

In human genomes, alphoid satellite DNAs account for 5 % of the genome, and are found as 340 bp EcoRI fragments organized in tandem at the centromeres of all chromosomes (Willard *et al.*, 1989). The repeat unit is composed of two related regions of 169 bp and 171 bp. There is 23 % sequence divergence between the two sub-repeats. In contrast, there is only 1 % sequence divergence between adjacent units of 340 bp. It thus

appears that the amplification, in tandem, of the 340 bp repeat units has occured after the divergence of the shorter sub-repeats.

The human EcoRI alphoid family displays tissue-specific levels of methylcytosine. In brain tissue, it was observed that the alphoid DNA sequence was 66 % more methylated, (at CpG dinucleotides) than its placental counterpart (Gama-Sosa et al., 1983). Furthermore, the authors observed that these repeated sequences were ten times less methylated in sperm cells than in somatic cells.

The striking feature about the evolution of alphoid DNA is the presence of conserved regions alternating with non-conserved regions (Rosenberg et al.,1978; Donehower et al., 1980). Furthermore, the junctions between conserved and divergent regions are marked by the presence of 5' TTCC 3' (or its invert AAGG). These simple sequences are believed by the authors to be recognition site for nucleases and / or recombination proteins.

IV-2 Simple repetitive sequences

IV-2.1 Mouse satellite DNA

The major mouse satellite DNA discovered by Kit (1961), was shown to represent approximately 9 % of the genome and appeared to contain 2 to 3 time more methylated cytosines (mC) than the DNA found in the main band (Bond et al.,1967;

Schildkraut and Maio, 1968; Salomon et al., 1969). Though satellite DNA is concentrated in heterochromatin, trace amounts are found in euchromatin (Mattoccia and Comings, 1971). Originally localized within the centromeric heterochromatin of all chromosomes, except the Y (Jones 1970), it was later found on the long arms of some chromosomes (White et al., 1975). Southern (1975b) was the first to report the tandem organization of mouse satellite DNA as 234 bp long EcoRII segments. Nucleotide sequence analysis revealed that while the basic repeat was 234 bp, it was composed of internal sub-repeats, 58 bp or 60 bp in length (Manuelidis, 1981). In turn, shorter sub-repeats of 28 bp (called alpha) alternating with 30 bp (called beta) were found. It is believed that mouse satellite DNA has evolved, from the tandem duplication of an ancestral unit, 5' TGAAAAA 3', by a process of mutation and amplification, (Horz and Altenburger, 1981).

IV-2.2 Mini-satellites

The term mini-satellites includes polymorphic repetitive DNA sequences, varying in length from 9 to 60 bp, that are organized in tandem arrays within dispersed repetitive regions (Bell et al., 1982). A subset of these, preferentially located near the ends of human chromosomes, share a common core sequence 5'GGAGGTGGGCAGGARG 3', that is similar to the CHI recombination signal of E. coli. (Jeffreys et al., 1985). One of these minisatellite DNA consists of four tandem repeats of a 33 bp sequence and is found flanked by 9 bp direct repeat, a

feature characteristic of the target site duplication generated by transposable elements.

IV-2.3 Decameric satellites DNAs

Deca-satellite DNA was found fortuitously in the early 1980s, in the genome of African green monkey (Maresca and Singer, 1983). While looking at the junction of alpha satellite DNA, decameric repeat units, organized in long tandem arrays, were detected (Maresca and Singer, 1983). Though the basic repeats are derived from the consensus 5'AAACCGGNTC 3', it is the internal CCGG core that is found most conserved.

In Kangaroo rat, decameric satellite DNAs share a different though related, decameric unit 5'ACACAGCGGG 3' (Fry et al., 1973). This family of related sequences, called HS-Beta-ST-DNA, contain DNA sequences that contain high amount (6.7 %) of methylated Cytosines (Hatch and Mazrimas 1974).

IV-2.4 GT/CA satellite DNAs

The genome of many eukaryotes contains simple sequences, composed of alternating purines and pyrimidines of the following types: (GT/CA)n, (GA/CT)n , (AA/TT)n... etc.

IV-2.5 Human classical satellite DNA

Molecular analysis of the human genome has revealed two types of tandemly organized repeated DNA sequences that are characteristic of heterochromatin: the alphoid satellite DNA and the so-called classical satellite DNAs.

The term "classical" satellite DNAs refer loosely to the collection of DNA sequences detected by buoyant density sedimentation gradients as distinct component of nuclear DNA (Corneo et al., 1971, and figure 2). Satellite peaks consist of a mixture of sequences, some of which are related as in the case of satellite II and III DNAs (Mitchell et al., 1979).

Satellite II and III peaks consist of a heterogeneous collection of repeated DNA sequences that apparently have evolved from a common pentameric ancestor 5' TTCCA 3' (Frommer et al., 1982; Prosser et al., 1986). Members of these two families of related sequences have been localized, by in situ hybridization to centromeric heterochromatin (Gosden et al., 1975).

The satellite II fraction consists of tandemly arrayed DNA sequences in which the 5 bp motif is widely degenerate (Frommer et al., 1982). This degeneracy has led to the formation of numerous Taq1 (TCGA) and Hinf1 (GANTC) restriction sites which are characteristic of satellite II DNA family members.

Figure 5: Taq1 and Hinf1 restriction site generation in simple satellite DNA. The generation of Taq1 (TCGA) and/or Hinf1 (GANTC) restriction sites may appear, within the consensus sequence (5' TTCCA 3') of human satellite II and III DNA, as a result of single point mutations. Multiple transversions affecting the two cytosine residues within the same repeat, generate Hinf1 sites (arrow a). Overlapping Taq1 and Hinf1 restriction sites may be achieved via C to G transversion in only one of the repeats (arrow b). Whenever transversions of the second and first cytosines occur, respectively, between adjacent repeats, only Taq1 sites are generated (arrow c). Bases substituted are underlined within the 5'TTCCA 3' tandem array and outlined within the mutated pentamers. Hinf1 and Taq1 sites are indicated, respectively, between brackets and parenthesis.

Figure 5

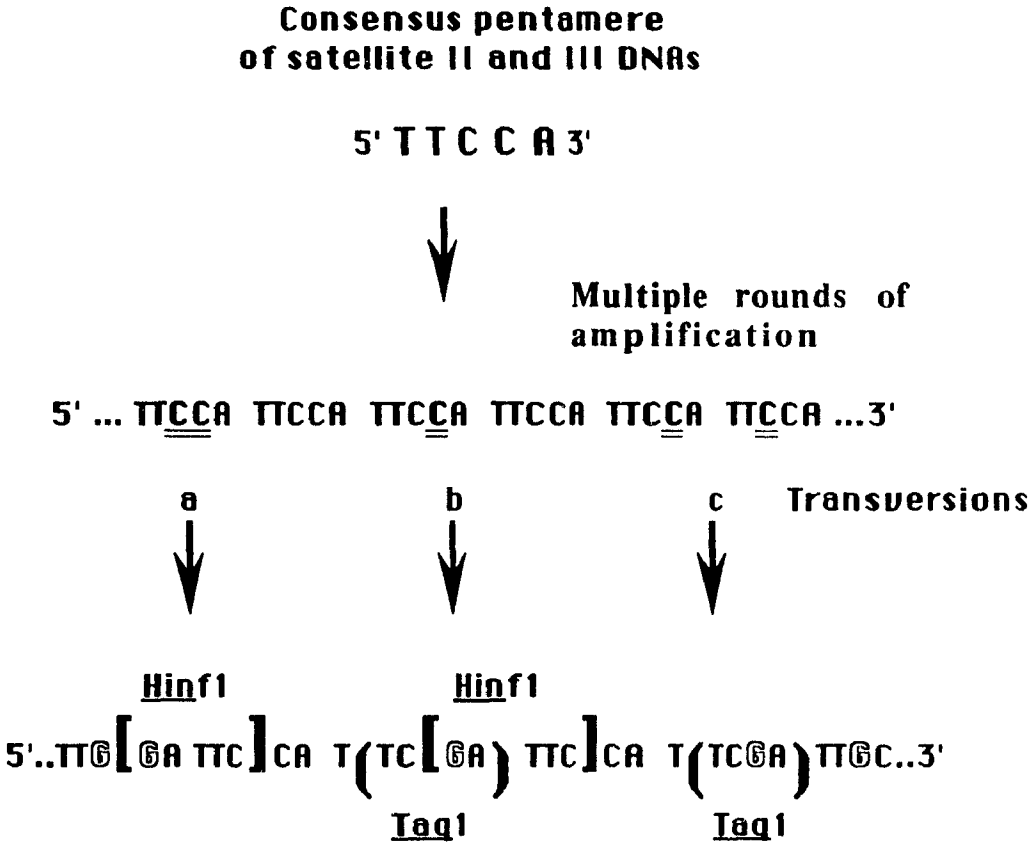

Consensus pentamere
of satellite II and III DNAs
5' T T C C A 3'
Multiple rounds of
amplification
5' ... TTCCA TTCCA TTCCA TTCCA TTCCA TTCCA ...3'
a
b
c
Transversions
Hinf1
Hinf1
5'..TTG[GA TTC]CA T(TC[GA) TTC]CA T(TCGA)TTGC..3'
Taq1
Taq1

Figure 5 shows how these restriction sites may be generated, from tandem pentamers, via a single C to G transversion.

One main distinguishing characteristic of satellite II DNA members is that the internal nucleotide [N] within the Hinf1 sites (GANTC) is more frequently a G/C than an A/T. Satellite II appears to be the only one group of simple sequences to hybridize significantly to chromosome 16 (Gosden et al., 1975).

Though the "classical" satellite III fraction contains a number of unrelated repeated DNA such as alphoid satellite DNA (Mitchell et al., 1979), it is the major family of human simple repeated sequences. Characterized by 5' TTCCA 3' pentamers, satellite III DNAs are interspersed with Hinf1 sites, and contain few to no Taq1 restriction sites (Prosser et al., 1986, Nakahori et al., 1986). In contrast with satellite II DNA, satellite III DNA family members have fewer Hinf1 sites and the internal nucleotide [N] of the restriction site (GANTC) is more often an A/T than it is a G/C.

Typical domains of satellite II and III have been identified on human autosomal and sex chromosomes. The K-domain described by Burck et al. (1985), and observed by other (Holden et al., 1985, Sol et al., 1986), refer to the tandem organization of 1.8 Kb and 3.6 Kb Kpn1 satellite DNA units. A member of this Kpn1 family (D15Z1) has been found associated with the nucleolar organizer and centromeric heterochromatin, in homogeneously staining regions (HSR) of chromosome 15 (Higgins et al., 1985; Holden et al., 1985). D- and R-domains specific to

chromosomes 16 and 1, respectively, were also identified (Burck et al., 1985). These three domains represent the autosomal homologues of the male specific 3.4 Kb HaeIII (or EcoR1) satellite DNA that was previously identified by Cooke (1976). Recently, a 1.4 Kb EcoR1 satellite III DNA, located on chromosome 14, was reported (Choo et al., 1990b). This satellite was shown, by in situ hybridization, to not only be located in the heterochromatic regions of chromosome 14, but also on chromosomes 1, 9, 22 and Y. Using pulse field gel electrophoresis, the authors showed that there was a remarkable polymorphism in the organization of this satellite DNA on different chromosomes. On chromosome 14, the 1.4 Kb EcoR1 satellite DNA is found clustered in a large tandem array of 150 Kb, whereas on chromosome 22, shorter arrays of the satellite, ranging from 20 to 150 Kb, are detected.

V- ORIGIN OF REPEATED DNA SEQUENCES

A series of diverse mechanisms have been proposed to explain the generation and maintenance of tandem and interspersed repeated DNA sequences (reviewed in Beridze, 1986).

Some of the repeated DNA sequences are thought to originate from replicative transposition, as seen for the SINES-like AluI family that appear to have arisen via self-primed reverse transcription (Schmid and Shen, 1985). The process by which mobile elements (e. g. SINES and LINES) have been amplified, homogenized and distributed among homologous and non-homologous chromosomes (in short, their concerted evolution) has been explained both by retrotransposition and gene conversion (Jackson and Fink, 1981; Dover , 1982). It is the subsequent drift that is responsible for the degeneration of the original unit (thus, creating diverged subfamilies), and for the generation of immobile copies of transposable elements (e. g. Alu pseudogenes).

Most of the data concerning the generation of simple and complex tandem DNAs can be explained by mechanismes such as slippage-replication, sister chromatid exchange, rolling circle amplification of extrachromosomal DNA, onion skin replication, and unequal crossing over (Britten and Kohne, 1969; Southern, 1974, 1975b; Smith, 1976, 1978; Schimke, 1984).

One of the first step of sequence duplication is believed to arise via aberrant replication. Though slippage of DNA

polymerases during replication or repair of DNA molecules have been shown to occur in vitro (Kornberg et al., 1964), it has not been proved, as yet, to occur in mammalian cells. Jarman and Wells (1989) proposed a model by which minisatellite DNA (containing CHI-like core sequences) may be duplicated via a DNA nicking/repair mechanism. Illegitimate recombination, at a region of microhomology (e. g. direct repeats), between two chromosomes generates two products, one of which is a deletion of one direct repeat (plus intervening DNA), and the reciprocal product is a duplication of the corresponding deleted product. This mechanism, sometimes termed unequal crossing over (Smith, 1976), has been proposed not only for deletion formation but also as the initial reaction during gene amplification (Ford and Fried, 1986). As the tandem array acts as a hotspot for recombination both within a chromosome or between sister chromatids, large regions may be generated in the germline by the continual dynamic process of unequal cross-over and gene conversion (Baltimore, 1981).

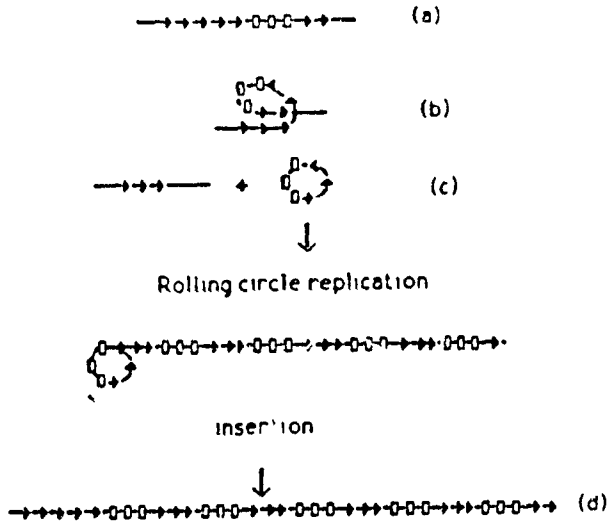
In contrast with the slow process of unequal crossing-over, it is believed that large regions of tandemly arrayed simple sequences are created suddenly via aberrant replication (Schimke,1984). Models of aberrant replication such as onion skin replication (Schimke, 1984) and rolling circle replication (Hourcade et al., 1973) are presented in figure 6. Onion skin replication, also called the replicative loop hypothesis, is due to replicon misfirings and leads to the absence of separation of the newly synthesized daughter strands. Recombination of the

endoreplicated strands may either lead to their insertion in the chromosome as tandem arrays, or to their excision as circles that may be amplified further via rolling circle replication. Rolling circle replication is based on the excision (of a region containing an origin of replication), circularization, and reintegration, probably by some recombination mechanism of the amplified DNA sequences. Variants within family members occur through mutations taking place either prior or after the amplification of the tandem array. The processes of homogenization (or randomization), by which mutations spread through a family, and fixation (genetic drift) through a population of related repeated DNA sequences, are usually referred to as molecular drive (Dover, 1986).

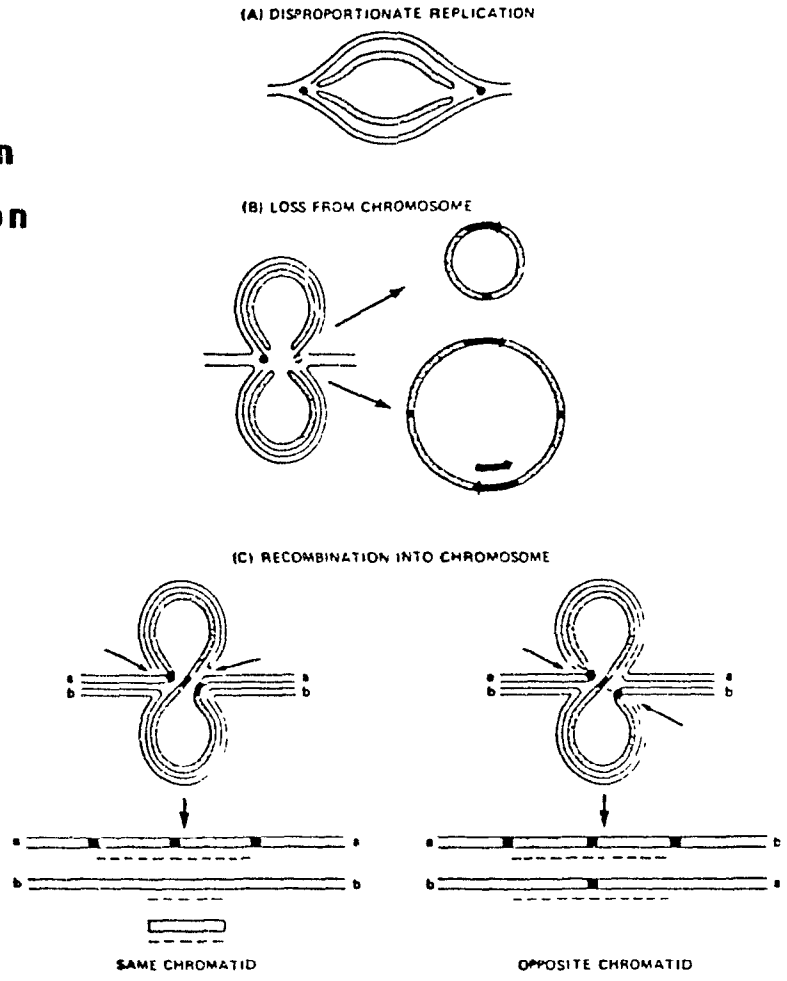
Figure 6: Models of amplification of repeated DNA sequences. Rolling circle replication (from Walsh, 1987): (a) Tandem arrays undergoes a sister-strand exchange, resulting (b) in deletion of some array members and formation of a circular plasmid. (c) Some of these plasmids may undergo rolling circle amplification before (d) reinsertion into the chromosome. Onion skin replication (from Schimke, 1984) results from the endoreplication (stippled strands) of a region proximal to the origin of replication (A). The free strands can ligate to generate free circles (B), which in turn may undergo rolling circle replication. To generate chromosomally localized tandem sequences recombination to the backbone is required (C). Recombination may occur onto the same chromatid (left panel) or onto opposite sister chromatids (right panel). Recombination joints are depicted as black sites.

Figure 6

Rolling circle replication



Onion skin replication



VI- DNAs IN SEARCH OF A FUNCTION

If the forces responsible for molecular evolution of repeated DNAs remain a mystery, so is their role in eukaryotes. There are different schools of thought about the relevance of carrying this load of apparently unimportant genetic baggage around. Despite many efforts, no function has yet been established for these enigmatic DNA sequences, although a variety of hypotheses have been put forward.

VI-1 The concept of selfish DNA

Repetitive, non-genic sequences may have accumulated over the eons and prospered as selfish genomic parasites (Orgel and Crick, 1980; Doolittle and Sapienza, 1980). According to these authors, simple sequences and moderately repeated DNAs (as well as introns), would exploit the cell machinery for their own reproduction, bringing to the host cell no benefits in return. Lacking any function but their own preservation, such elements would evolve rapidly via duplication and deletions of large blocks during genetic recombination by a process of unequal crossing over, and thus might be by-products of genome rearrangement or aberrant DNA replication.

VI-2 Alternative views

VI-2.1 Genetic evolution

As much as Nature may be objective, Evolution is surely opportunistic. Thus, the bulk of non-genic simple and complex repetitive sequences may give a greater genetic flexibility to the host by providing "the raw material" for the making of new genes, and serving as pools of DNA for use in further evolution (Mazrimas and Hatch, 1972). Molecular mechanisms by which repeated sequences propagate and amplify, such as unequal exchange, gene conversion and transposition, may cause mutations, thus providing a pool of genetic variation upon which Darwinian selection acts.

Some Alu/Alu rearrangements lead to genetic disease (Lehrman et al , 1987), whereas others lead to new genes (Gilbert, 1987), as in the proposed origin of the glycophorin B gene (Kudo and Fukuda, 1989). Alternatively, other roles have been proposed for Alu elements. Because of their homology to 7SL RNA, some Alu sequences may function at the level of a ribonucleoprotein (RNP) particles related to the signal recognition particle (SRP) containing 7SL RNA. Jang and Lachman (1989), after observing increased transcription of Alu sequences upon HSV infection, proposed that these repeated sequences may be involved as cellular anti-viral agents. Moreover, due to the presence in Alu sequences of a core sequence similar to SV40 origin of replication, some members of the Alu family may have a role to play in chromosome replication.

Simple repeated sequences favour illegitimate recombination between two separate chromosomes at a region of microheterogeneity. Moreover, the tandem array itself is a fertile ground for recombination within a chromosome. Thus, it is believed that some repeated DNAs may play an important role in gene shuffling, a function that may well be linked to the tandem structure itself rather than to a particular DNA sequence. In particular, the transition to the Z-DNA form of regions containing arrays of the (GT/CA)n may be related to the efficiency of recombination of the human embryonic gamma-globulin genes, and may participate in gene regulation (Thomas et al., 1985).

Some of the highly repeated DNA sequences that are enriched in methylcytosines may participate in negative regulation of gene expression (Gama-Sosa et al., 1983).

VI-2.2 Structural-epigenetic information

In view of the heterochromatic localization of repeated sequences, it is possible that their function may be linked to that of the heterochromatin (John, 1988). The close and periodical association of proteins and repetitive DNA sequences may stabilize the condensed chromatin structure. Roles such as involvement in meiotic pairing, recognition of homologous centromeres, and folding of chromosome fibres have also been proposed (Manuelidis, 1982).

Individual centromeric proteins (CENP A-B-C), that are necessary for proper centromere function during mitosis, have

been identified by using autoimmune sera of patients with the CREST group of scleroderma pigmentarum disorder (Brenner et al., 1981). One of these, CENP-B an 80 Kdalton protein, is localized beneath the kinetochore in the centromeric heterochromatin (Cooke et al.,1990). In 1989, Masumoto et al., presented the first in vitro evidence of specific interactions between a subclass of alphoid DNA and CENP-B. Using binding assays, the authors showed that CENP-B binds specifically to a 17 bp core sequence 5'CTTCGTTGGAAACGGGA 3' that is shared by all members of this subclass of alphoid DNA (Masumoto et al., 1989). Recently, Wevrick et al., (1990) reported the in vivo evidence of these specific interactions. The authors observed a correlation between the amount of centromeric alphoid satellite and the quantity of CENP-B antigen, a deletion of satellite DNA is associated with reduced amount of CENP-B protein. The structure formed by the association of CENP-B to the 17 bp core region may be recognized by other proteins involved in assembly of the kinetochore.

The AT-rich mouse major satellite DNA has the ability to produce a bent structure, and hence was thought to play a role in centromere condensation (Radic et al., 1987). However, this hypothesis does not seem to survive the recent findings that this satellite is not found in other species of the genus Mus (Wong et al., 1990).

VII- OUTLINE OF THE THESIS

Though highly repeated DNA sequences are believed to play a role in chromosome behavior and evolution, to date it is not clear whether they do so accidently, or have been selected to do so.

In order to unravel some of the mysteries surrounding the presence of excess DNA in the human genome, we have chosen to look for sequence conservation in the repetitive DNA between the genome of the fruit fly (**Drosophila melanogaster**) and human (**HeLa**) cells. We were interested in studying repeated mobile elements, and therefore used the P-element (a fruit fly transposon), as a probe to screen, by hybridization, for P-element-related human DNA sequences.

Serendipity led us to the isolation and characterization of a human highly repeated satellite II DNA, that appeared, however, not to be mobile.

The relevant experimental data gathered has been compiled into four chapters (introduction and concluding remarks making Chapters I and VI respectively, of this thesis).

Chapter II concerns all the materials and procedures utilized throughout this analysis.

Chapter III reports the original findings concerning the search for sequence conservation between human repeated DNA sequences and **Drosophila's** P-element transposon.

Chapter IV describes the micro-heterogeneity observed between closely related members of the satellite II and III

families of human highly repeated simple DNA sequences, as well as the attempts to clarify their chromosomal origins.

Chapter V depicts the macroscopic organization of satellite DNAs, and the extent of structural polymorphisms observed in different cell lines.

Chapter II

MATERIALS AND METHODS

How to paint a perfect painting-
Make yourself perfect and then paint naturally...

Robert M. Pirsig

Abbreviations used in this thesis are listed in Appendix I.

I- BACTERIAL STRAINS, PLASMIDS, AND GROWTH CONDITIONS.

The bacterial strains and plasmids used in this thesis are listed in table 1.

The Escherichia coli (E. coli) strains DH1 and NM522 were routinely grown at 32°C in LB media (Miller 1972), supplemented with 40 μg/ml of ampicillin (Bristol), when required.

E. coli GM33 was grown in M9 glucose media (Maniatis et al., 1982) supplemented with thiamine (10 μg/ml).

M13K07 phage (Viera and Messing 1982), were a kind gift of Dr. J. Geisselsoder of BioRad (Richmond, CA).

The Drosophila P-element containing plasmid pπ25.1 (Rubin et al., 1982) was a gift of Drs. G. Rubin and A. Spradling. Plasmid pBK1.8[20] (Shafit-Zagardo et al., 1982) was kindly provided by Dr. J. Maio (Albert Einstein College, NY).

Plating of transformed cells was performed on LB amp plates supplemented, when needed, with X-Gal (40μg/ml made in N,N dimethyl formamide) and IPTG (1 mM). All bacteriological culture media components were purchased from Difco (Detroit, Michigan) or Maknur (Ottawa, ON).

TABLE 1: Strains and plasmids.

Names:	Characteristics:	References:
E. coli strains		
DH1	F^-, recA1, endA1, gyrA96, thi-1, hsdR17, supE44	Low (1968)
NM522	Δpro-lac, rspL, thi, hsd(r^-m^+), supE44, / F'traD36, pro, lac I^q, Z M15.	Gough & Murray (1983)
Plasmids		
pBR322	ColE1 derivative, Tc^r, Ap^r, 4.3 Kb.	Sutcliffe (1979)
pUC119	derivatiuce of M13mp19, and pUC19, M13 packaging site, Ap^r, 3.2 Kb.	Viera & Messing (1982)
pπ25.1	Contains the 2.9 Kb P-element of *Drosophila melanogaster*.	Spradling & Rubin (1982)
pBK1.8[20]	Contains a human 1.8 Kb Kpn1 satellite DNA	Shafit-Zagardo *et al.*, (1982)
pKS36	Contains a human 1.797 Kb EcoR1 satellite DNA	Sol *et al.* (1986)

II- EUKARYOTIC STRAINS AND CULTURE CONDITIONS.

All mammalian cell culture media were purchased from Flow (Mississauga, ON) or Gibco (Burlington, ON). Yeast cells, *Saccharomyces cerevisiae* strain BWG1-7a (MATa, ade 3-52, Leu 2-3,112, his4-519, GAL$^+$, Guarente et al., 1982) were kindly prepared by Dr. T. Keng of our department. The yeast cells were grown at 30°C in YPD media (1 % yeast extract, 2 % peptone, 2 % dextrose). HeLa and AK143 cells were propagated in a humidified 5% CO_2 atmosphere at 37°C in Dulbecco's modified Eagle medium (DMEM) supplemented with 50 µg/ml gentamycin (Schering Canada Inc.), 10% fetal bovine serum (FBS), and 2.5 µg/ml fungizone (Squibb Canada Inc.). AK143 cells were derived from a murine sarcoma virus-transformed line R970-5 (Rhim et al., 1975). MeWo cells, purchased from the Sloan-Kettering Institute (Rye, NY) were grown in RPMI 1640 medium supplemented as indicated above with FBS, gentamycin and fungizone. U937 cells, a generous gift of Dr. M. Ratcliffe of our department, were maintained in Iscove's Modified Dulbecco's Medium (IMDM), supplemented with 5% heat-inactivated FBS, 100 µg/ml penicillin, 100 µg/ml streptomycin, 5 µg/ml transferrin (Boehringer Mannheim Canada Ltd., BMC), and 0.05 mM 2-mercaptoethanol.

DNA's extracted from human/hamster hybrid cell lines were kindly provided by Dr. M. Hansen of the Ludwig Cancer Institute (Montreal, Que.).

III- ENZYMES AND CONDITIONS.

All restriction endonucleases were purchased from BMC, Bethesda Research Labs. (BRL), or Pharmacia Canada (Montreal, Que.). Routinely, DNA was hydrolyzed in 6 mM Tris-HCl [pH 7.5], 6 mM $MgCl_2$, 75 mM NaCl, 6 mM 2-mercaptoethanol and 125 μg/ml bovine serum albumin (BSA, Pentex fraction V, Miles, Elkhart, IN). DNA hydrolysis was conducted for 4 hours at 37°C (unless otherwise indicated), using 1 unit of restriction endonuclease per microgram of plasmid or phage DNA, or 5 units of enzyme per microgram of genomic DNA.

After hydrolysis, the samples were placed at 65°C for 10 minutes to stop the reaction, and completely separate the cleaved fragments before electrophoresis.

T4 DNA ligase, E. coli DNA polymerase1, and T4 polynucleotide kinase were purchased from Pharmacia Canada Inc. Calf intestinal alkaline phosphatase (CIAP) was obtained from Dupont-New England Nuclear (NEN, Boston, MA.). RNase A, DNase 1, Pronase were purchased from BMC, and Zymolyase from Serkagaky Kogyo Co Ltd. (Tokyo, Japan). DNA pol 1 (Klenow fragment) was obtained through BRL. Table 2 presents the composition of the buffers employed with these enzymes.

TABLE 2: Enzymes and buffers

Enzymes (stock solution) :	Buffers:
RNase A (1 mg/ml)	10 mM Tris-HCl (pH 7.6) boiled 10 min to inactivate DNase activity and stored at -20° C.
Pronase (20 mg/ml)	10 mM Tris-HCl, 1 mM EDTA (pH 7.5) = 1 x TE, autodigested 2 hours at 37° C, stored at -20° C.
Lysozyme (10 mg/ml)	10 mM Tris-HCl (pH 8)
Zymolyase (2 mg/ml)	10 mM Sodium phosphate buffer (pH7.5)
DNase 1 (1 mg/ml)	10 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$, 10 mM 2-mercaptoethanol = TM
*E. coli DNA pol 1	50 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$, 10 mM 2-mercaptoethanol, 25 µg/ml BSA = Nick Translation Buffer, NTB.
*T4 DNA ligase	15 mM DTT, 1 mM spermidine, 0.75 mM ATP, 50 µg/ml autoclaved gelatin, 60 mM Tris-HCl (pH 7.4), 10 mM $MgCl_2$ = Linker Ligation Buffer
*Calf intestinal alkaline phosphatase	20 mM Tris-HCl (pH 10).
*T4 polynucleotide kinase	50 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, 5 mM DTT, 0.1 mM spermidine, 0.1 mM EDTA = Linker Kinase Buffer, LKB.
*DNA pol 1 (Klenow fragment)	7 mM Tris-HCl (pH 7.5), 7 mM $MgCl_2$, 50 mM NaCl, 1 mM DTT = KB.

* indicates enzymes already in solution

IV- DNA PURIFICATION AND CONCENTRATION.

When required, the reaction volume was increased to 200 µl with 1 x TE (10 mM Tris-HCl [pH 7.5], 1 mM EDTA) to reduce loss of material. For extraction of DNA, an equal volume of phenol saturated with 1 x TE was added, and the fluids vortexed thoroughly. The aqueous phase, containing the deproteinated DNA, was separated by centrifugation at 15,000 x g for 10 minutes at 15°C in an Eppendorf microcentrifuge. The supernatant fluid was removed and mixed with an equal volume of ether to remove any traces of phenol. After phase separation, the ether (top phase) was removed from the DNA preparation. This procedure was, when indicated, performed with phenol/chloroform/isoamyl-alcohol (PCI:45/45/10 by volume) in place of the phenol/ether procedure. For extraction of total genomic DNA, the phases were mixed by gentle inversion of the tubes to reduce shearing of the large DNA molecules.

The DNA was then precipitated in 10% salt upon addition of 2.5 volumes of ice cold absolute ethanol. Depending on the subsequent use of the DNA, the salt was either sodium acetate or ammonium acetate. The precipitations were either conducted at -20°C overnight, or for 15 minutes in a dry ice-ethanol bath. The DNA was then collected by centrifugation at 15°C for 30 minutes, at 15,000 x g, dried 5 minutes under vacuum, and resuspended in the appropriate buffer.

U- DNA ISOLATION.

U-1 Total E. coli DNA.

Logarithmically growing cells were harvested by centrifugation at 10,000 x g for 10 minutes at 4°C, and the cell pellet resuspended in 5 ml of a solution composed of 10 mM Tris-HCl [pH 8], 1 mM EDTA and 0.5 % sodium dodecyl sulfate (SDS). This suspension was incubated at 37°C for 4 hours in the presence of 0.5 ml of 20 mg/ml pronase. The solution was then extracted twice with phenol and ether and dialysed for 24 hours against 100 volumes of 1 x TE at 15°C. After treatment for 2 hours at 37°C with 0.5 ml of 1 mg/ml RNase A and a phenol/ether extraction, the DNA was further dialysed for 48 hours as indicated above. The purity and concentration of the DNA were assessed by gel electrophoresis and spectrophotometry at 260 and 280 nm (1 OD_{260}unit = 50 μg/ml).

U-2 Calf thymus DNA.

500 mg of high molecular weight calf thymus DNA fibers (Sigma, St. Louis, MO) were allowed to dissolve slowly in 1 x TE to a concentration of approximately 5 mg/ml. After 5 days at 4°C (with occasional gentle agitation), the DNA was treated for 2 hours with 0.5 mg RNase A and 3 hours with 100 mg pronase at 37°C. The DNA was extracted twice using PCI and dialysed for 48 hours against 1 x TE at 4°C. Its purity and concentration were determined as previously indicated.

U-3 Mammalian DNA extractions.

The following protocol was used to extract genomic DNA from 10 plates of confluent HeLa or MeWo cells or 3 flasks (250 ml) of U937 cells. Cells were washed twice with PBS and subsequently incubated 5 minutes with 1 mg of Trypsin. The trypsinized cells were collected and harvested by centrifugation at 2,000 x g for 10 minutes at room temperature, and washed with 2 volumes of phosphate buffered saline (PBS: 3 mM KCl, 8 mM Na_2HPO_4 [pH 6.8], 1.5 mM KH_2PO_4, 0.14 M NaCl). The cells were resuspended in 10 ml of 1 x TE, supplemented with 0.5 ml of 10 % (w/v) SDS, 0.5 ml of pronase and incubated for 5 to 8 hours at 37°C. The DNA was extracted twice with phenol and twice with chloroform/isoamyl-alcohol (24:1 by vol), and dialysed against 1 x TE for 4 days at 4°C, with two changes of buffer per day.

U-4 Oligonucleotide purification.

A crude oligonucleotide mix, consisting of the complete oligonucleotide plus premature terminations of a synthesized 49 nucleotide long non-satellite sequence present in pKS36 (Chapter III), was generously provided by Dr. D. Garfinkel (National Cancer Institute, Frederick, MD.). 100 ng of "crude" oligomer (0.2 µg/µl) was kinased (section XII-2) to be used as a radioactive marker. A 12 % polyacrylamide sequencing gel (Maniatis et al., 1982), containing 8 M urea, was constructed and pre-run at 1500 volts for 1 hour. 60 µl of crude oligomer (0.2 µg/µl) was denatured in 140 µl of freshly deionized formamide,

and placed at 50°C for 3 minutes. Once denatured, the oligomer mix was then loaded onto the gel (15 µl/slot). In one slot, 200,000 cpm of kinased oligomer was placed as a marker, and the electrophoresis was resumed at 1500 volts for 2 hours. The gel was then wrapped in saran-wrap and exposed on Kodak X-Omat film. The 49 nucleotide long oligomer band was localized on the gel by comparison with the position of the labelled bands on the autoradiograph, and purified using (as described in section U-6.1) the "crush and soak" procedure of Maxam and Gilbert (1980). The resulting oligomer pellet was resuspended in 2 ml of "crush & soak" buffer (section U-6.1) and passed through a C18 Sep-pac column as recommended by the manufacturers (Waters, Mississauga, ON). The final concentration of the pure oligomer was determined by spectrophotometry at 260 nm (OD_{260}=40 µg/ml).

U-5 Plasmid DNAs.

Large scale preparation of plasmid DNA was performed according to Maniatis et al., (1982). Briefly, late logarithmically growing broth cultures of E. coli were supplemented with chloramphenicol (50 µg/ml) to selectively increase the copy number of the relaxed plasmid DNA due to host protein synthesis arrest (Clewell, 1972). After 18 hours of amplification, the cells were harvested by centrifugation at 10,000 x g, at 4°C and resuspended in 20 ml (per liter of cells) of resuspension buffer (25 % [w/v] sucrose, 50 mM Tris-HCl [pH 8], 40 mM EDTA). The plasmolysed cells were then treated on ice with 2 ml of RNase A

for 10 minutes, and subsequently with 8 ml of 0.5 M EDTA [pH 8] plus 2 ml of pronase for 10 minutes. Ten milliliters of lysis buffer (50 mM Tris-HCl [pH 7], 50 mM EDTA, 0.3 % [v/v] Triton X100) was added and the viscous mixture was subjected to centrifugation at 4°C for 1 hour at 30,000 x g. The cleared lysate was decanted, and every 8 ml was mixed with 7.5 g of cesium chloride (CsCl) and 0.3 ml of ethidium bromide (EtBr, 5 mg/ml). The solution was transferred to 12 ml Beckman polyallomer heat-sealable tubes, and filled to the top with mineral oil. An equilibrium gradient was obtained by centrifugation for 48 hours at 39,000 x g in a Beckman type 40 rotor (L8-70 Beckman ultracentrifuge) at 20°C. The DNA bands (host and plasmid) were visualized with 254 nm UV light, and the lower plasmid band collected through a 21-gauge needle. The EtBr was extracted with isopropanol saturated with 40 x SSC (6 M NaCl, 0.6 M sodium citrate [pH 7.6]), and the DNA was dialysed exhaustively versus 1 x TE buffer.

Rapid plasmid isolation (mini-prep) was performed according to the procedure of Holmes and Quigley (1981).

V-6 Recovery of DNA fragments.

V-6.1 Crush and soak.

This technique was used to isolate small DNA fragments after electrophoresis on polyacrylamide gels. The bands of interest were visualized, cut out of the gel, and the acrylamide

piece was forced through an 18-gauge needle. The crushed acrylamide was incubated overnight at 37°C in a buffer consisting of 0.5 M ammonium acetate, 10 mM magnesium acetate, 1 mM EDTA and 0.1 % (w/v) SDS (crush and soak buffer). The eluted DNA was then separated from the acrylamide by centrifugation at 15,000 x g for 10 minutes at room temperature. The resulting aqueous phase was gently removed and the DNA was purified by phenol/ether extraction and precipitated twice in two volumes of ethanol in the presence of sodium acetate. The last pellet was washed free of salt by a resuspension in water and ethanol precipitation.

U-6.2 Electroelution.

Electroelution was used to purify large DNA fragments after electrophoresis on agarose gels and was performed according to Maniatis et al., (1982). The band of interest was cut out under a low intensity UV lamp and placed in a dialysis bag filled with 50 μg/ml of BSA prepared in 1 x E buffer (40 mM Tris-base, 20 mM acetic acid, 2 mM EDTA [pH 8]). The DNA was electroeluted in a gel box filled with 1 x E buffer for 4 hours at 75 volts. The electric field was reversed for 3 minutes to release the DNA from the side of the bag, and the eluted DNA was collected, extracted using PCI and ethanol precipitated in the presence of ammonium acetate.

VI- CLONING PROCEDURES.

In this study, two different cloning vectors were used to clone the series of HeLa 1.8 Kb EcoR1 fragments eluted from agarose gels, and the transformations were performed using two different hosts. The 1.8 Kb Kpn1 human satellite DNA found in plasmid pBK1.8[20] was recloned in the Kpn1 site of plasmid pUC119.

VI-1 Ligation.

The cloning vehicles, pBR322 and pUC119, were linearized with the restriction endonuclease EcoR1, and the DNA was purified using PCI and ethanol precipitated with ammonium acetate. The Eco RI linearized DNA pellet was resuspended in 20 mM Tris-HCl [pH 10], and incubated for 30 minutes at 37°C, with 0.5 units of calf intestine alkaline phosphatase to remove the 5' phosphate from both extremities, hence minimizing recircularization of the linear molecules (Ulrich et al., 1977). The reaction was then heated for 10 minutes at 65°C, and the dephosphorylated DNA molecules were extracted, precipitated as before, and resuspended in 1 x TE. After linearization of pUC119 with Kpn1, the DNA was purified with PCI, ethanol precipitated and resuspended in 1 X TE.

The ligation reactions, containing 0.1 mg/ml of total DNA, were performed at 15°C overnight in linker ligation buffer (LLB, table 2) plus 1 unit of T4 DNA ligase. The molecular ratio of plasmid to linear DNA to be inserted was generally 1:3.

VI-2 Transformations.

VI-2.1 E. coli DH1 cells.

E. coli DH1 cells were made competent by the Magnesium-Calcium-Hepes method. 20 ml of cells were grown to an OD_{550} of 0.4 at 32°C, and collected by centrifugation for 7.5 minutes at 3,000 x g, 4°C. The supernatant fluid was removed and the cell pellet resuspended in 12 ml of magnesium-Hepes buffer (10 mM $MgSO_4$, 2mM Hepes-NaOH [pH 7.5]). After 20 minutes on ice, the cells were collected as before, resuspended in 12 ml of calcium-Hepes buffer (50 mM $CaCl_2$, 2 mM Hepes-NaOH [pH 7.5]), and the incubation on ice resumed for another 20 minutes. The competent cells were then pelleted as above, resuspended in 2 ml of the calcium-Hepes buffer and kept on ice.

Typically, 1.5 μg of DNA was incubated with 200 μl of competent cells plus 200 μl of of the $CaCl_2$-Hepes buffer for 20 minutes on ice. The transformation reaction was then subjected to heat-shock at 37°C for 2 minutes, followed by the addition of 1.5 ml of LB broth and incubation was continued at 32°C for 1.5 hours. Transformed cells were then plated on LB amp plates and grown at 32°C overnight to allow colony formation.

VI-2.2 NM522 cells.

A 25 ml culture of strain NM522 was grown to an OD_{550} of 0.5, and harvested as for DH1 cells. The cells were resuspended in 6 ml of 0.1 M $MgCl_2$ and pelleted by centrifugation at 4°C for 7 minutes at 3,000 x g. The supernatant fluid was removed and

the cell pellet resuspended in 6 ml of Tris-calcium buffer (10 mM Tris-HCl [pH 8], 75 mM $CaCl_2$), and incubated on ice for 15 minutes. After centrifugation, the competent cells were resuspended in 1.2 ml of the Tris-calcium buffer and kept on ice.

Typically, 1.5 µg of DNA was incubated for 30 minutes on ice with 200 µl of competent cells and 200 µl of of the Tris-calcium buffer. After a two minutes incubation at 37°C, the heat-shocked cells were gently agitated on ice for 1 hour. Then, 1.5 ml of LB broth was added, and the cells were further incubated for one hour at 32°C prior to spreading on LB amp plates and overnight incubation at 32°C to allow colony formation.

VII- CRACKING CELLS FOR PLASMID ANALYSIS.

This rapid technique for plasmid isolation is a modification of that of Barnes (1977). To analyze clones for plasmid content single colonies were picked and grown overnight on duplicate gridded master plates (Miller, 1972). Each patched clone was then scraped off one of the master plate with a sterile toothpick and dispersed thoroughly in 100 µl of "cracking buffer" (50 mM Tris-HCl [pH 8], 1 % [w/v] SDS, 2 mM EDTA, 0.4 M sucrose and bromophenol blue 0.01 % [w/v]). After a 30 minute incubation at room temperature to allow cell lysis, the viscous lysates were centrifuged for 15 minutes at 15,000 x g at 4°C and the aqueous supernatant fluid removed, loaded onto agarose gels, and subjected to electrophoresis. The size of the plasmid was

estimated by comparing the migration of the covalently closed circular form with either pBR322 or pUC119.

VIII- LAMBDA PHAGE PREPARATION.

E. coli GM33 cells (table 1), were grown to mid-log phase at 32°C in M9 minimal medium, glucose and thiamine (Miller 1972). The lambda c1857 lytic cycle was induced by shifting the culture to 42°C for 20 minutes, and phage production was allowed to proceed for an additional 4 hours by incubation of the culture at 37°C. The cells were then collected by centrifugation at 3,000 x g for 30 minutes at 4°C and, after removal of the supernatant fluid, resuspended in 5 ml of 1 x Mu buffer (Ljungquist and Bukhari, 1977). Two milliliters of chloroform were added, followed by gentle agitation for 30 minutes at 37°C until cell lysis (as denoted by a marked increase in the viscosity of the solution). The host genomic DNA was hydrolyzed on ice by the addition of 1 mg of DNase 1 for 15 minutes. The non-viscous solution was then subjected to centrifugation at 6,000 rpm, at 4°C for 10 minutes to remove unlysed cells and cell debris, and the blue, opalescent supernatant fluid was removed and gently layered on top of a CsCl block gradient. The block gradient was constructed by the slow addition of CsCl solutions of decreasing densities (ρ) : 0.5 ml of (ρ=1.6) CsCl, 1 ml of (ρ=1.5) CsCl, 0.5 ml of (ρ=0.4) CsCl and 0.5 ml 20 % [w/v] sucrose. All CsCl and sucrose solutions were prepared in 1 x Mu buffer 24 hours prior to use.

The block gradient was subjected to centrifugation for 2 hours at 40,000 rpm in a Beckman SW50.1 rotor at 4°C. The phage band (approximately 1.0 ml) was collected through a 21-gauge needle and mixed with to 3.5 ml of a 1.5 g/ml CsCl solution in Mu buffer. After 48 hours of centrifugation at 40,000 rpm (Beckman SW50.1, 4°C), the phage band was recovered as before and dialysed for 24 hours against 100 vol of 1 x Mu buffer. The phage particles were stored at +4°C. If needed, DNA was isolated by extraction with phenol and ether and subsequently dialysed extensively against 1 x TE.

IX- UNIDIRECTIONAL GEL ELECTROPHORESIS.

IX-1 Agarose gels.

Horizontal agarose gel electrophoresis of DNA was carried out on slab gels in 1 x E buffer at room temperature. The DNA samples were loaded along with a tracking dye consisting of 25 % [w/v] sucrose, 1 x E buffer, 0.05 % [w/v] bromophenol blue and 0.05 % [w/v] xylene-cyanol. Large DNA fragments were routinely separated on a 0.75 % [w/v] agarose gel whereas smaller fragments were separated in 1 % [w/v] agarose gels. Electrophoresis was performed either at 20 volts overnight or at 50 volts for 1 to 5 hours, depending on the size of the gel.

IX-2 Acrylamide gels.

A solution of 30 % [w/v] acrylamide-bis acrylamide (29/1) was diluted to the required concentration in 1 x TBE buffer and polymerization was obtained in the presence of 0.04 % [w/v] ammonium persulfate and 0.01 % [v/v] TEMED (final concentrations). The gel was pre-run for 30 minutes at 40 volts prior to loading the DNA samples.

Denaturing sequencing gels (ranging from 10 to 20 % in acrylamide concentration) were constructed with acrylamide containing 8 M urea. The DNA was subjected to electrophoresis at 1500 volts in 1 x TBE according to Maxam and Gilbert (1980).

IX-3 Staining of gels and DNA detections.

To locate the appropriate DNA fragments, agarose and neutral polyacrylamide gels were routinely stained in a 1 μg/ml ethidium bromide solution (EtBr) for 5 to 15 minutes at room temperature and subsequently destained in deionized water for 30 minutes When required, the gels were photographed, under a high intensity-UV source, using a Polaroid camera and type 57 Polaroid film.

Detection of small DNA fragments (down to 10 bp) was performed using a combination of Coomassie brilliant blue (G250) and silver staining. The polyacrylamide gels were soaked 3 times for 15 minutes in a solution composed of 40 % [v/v] methanol and 10 % [v/v] acetic acid. They were then placed in a glass tray filled with a stain solution (0.25 % [w/v] Coomassie brilliant blue, 50 % [v/v] methanol and 6 % [w/v] trichloroacetic

acid [TCA]) for 30 minutes at room temperature under constant agitation. The gels were then destained extensively in 5 % [w/v] TCA for 10 minutes, and soaked twice in 40 % [v/v] methanol/10 % [v/v] acetic acid, and twice in a solution made of 10 % [v/v] ethanol, 5 % [v/v] acetic acid. The gels were subsequently stained with silver nitrate (0.012 M, made in water)) for 30 minutes. After several washes with water, the DNA bands were visualized in 100 ml of 0.28 M sodium carbonate and 0.5 ml/l of formalin. This developing process was stopped by placing the gels in 5 % [v/v] acetic acid. The gels were kept in water overnight at room temperature, and photographed.

X- FIELD INVERSION GEL ELECTROPHORESIS

X-1 Mammalian cell plug preparation.

HeLa, MelWo and U937 cells were grown to confluency (as previously described, section II), trypsinized, harvested by centrifugation and resuspended in PBS to 1×10^7 cells/ml (final concentration) at 37°C. An equal volume of cells and 1 % low gelling temperature agarose (LGT Sigma type VII made in PBS) were mixed at 37°C and aspirated into silicon tubing (3/32 inner diameter, Cole-Parmer). The tubing was then placed at +4°C for 10 minutes to allow the agarose to harden. The polymerized agarose containing the cells was extruded onto a sheet of saran-wrap (Fisher Scientific, Montreal, Que.), and cut into 1 cm "plugs". Each plug contained approximately 4×10^5 cells.

The plugs (50 on average), containing the embedded cells (HeLa, MelWo and U937), were placed in three separate 50 ml conical tubes and incubated at 50°C for 24 hours in 25 ml of a lysis buffer consisting of 0.5 M EDTA, 1 % [w/v] sarkosyl and 2 mg/ml proteinase K. The plugs were subsequently transferred to new conical tubes and washed with ice-cold 1 x TE buffer in order to stop the lysis process. After three washes with 1 x TE at room temperature, the plugs were incubated for 2 to 4 hours at 50°C in 0.04 mg/ml PMSF (phenyl methyl sulfonylfluoride) prepared in 1 x TE. The washes were thereafter resumed as described above and the plugs were stored in 0.5 M EDTA at 15°C.

Just before use, the required number of plugs were washed extensively in 1 x TE, and preincubated overnight (one plug per 1.5 ml eppendorf tube) at 15°C in 0.5 ml of a buffer consisting of autoclaved gelatin (0.2 mg/ml), spermidine (5 mM), and the appropriate digestion buffer as recommended by the manufacturers. The next day, the plugs were transferred to new tubes containing 0.2 ml of freshly made buffer. Fifty units of the appropriate restriction endonuclease were added, and the reaction was allowed to incubate for 5 hours at the temperature recommended by the manufacturers. The reactions were then stopped by placing each plug in a new tube containing 0.5 ml of 25 mM EDTA, and stored at 4°C prior to electrophoresis.

H-2 Molecular weight markers.

1.5 ml of yeast cells, **Saccharomyces cerevisiae** strain BWG1-7a (1 x 10^6 cells/ml), 2.5 ml of 1 % [w/v] LGT agarose (prepared in 0.125 M EDTA) and 75 µl of zymolyase (prepared at 2 mg/ml in 0.01 M sodium phosphate buffer, [pH 7.5], and 50 % [v/v] glycerol) were gently mixed at 38°C, and aspirated in siliconized tubing as previously described. The 1 cm long plugs were transferred to a conical tube filled with LET buffer (0.5 M EDTA, 0.01 M Tris-HCl [pH 7.5], and 7.5 % [v/v] 2-mercaptoethanol), and incubated at 38°C for 24 hours. The plugs, containing yeast spheroplasts, were transferred to NDS buffer (0.5 M EDTA, 0.01 M Tris-HCl [pH 7.5], 1 % [w/v] Lauroyl sarcosine and 2 mg/ml proteinase K). The incubation was performed at 50°C for 48 hours and followed by three successive washes in ice-cold 1 x TE to stop cell lysis as for the mammalian cell plug preparation. The plugs were stored in 0.5 M EDTA at 15°C.

100 µl of lambda phage particles, purified as described in section VIII, were mixed with an equal volume of PBS and placed at 37°C. The phage solution was mixed with 200 µl of 1 % [w/v] LGT agarose, and aspirated in tubing as before. The plugs were incubated in lysis buffer (0.5 M EDTA, 1 % [w/v] sarkosyl, 1 mg/ml proteinase K, in 10 mM Tris-HCl [pH 9]), for 4 hours at 55°C with gentle agitation. The plugs were then placed in a 50 ml conical tube filled with ice-cold 1 x TE, and agitated for 2 minutes in order to stop the lysis. This washing step was

repeated once and the plugs were then transferred to a new tube and stored in 0.5 M EDTA.

H-3 Electrophoresis.

The plugs to be analyzed were soaked for 30 minutes in a solution made of 0.5 x TBE plus 0.05 % [w/v] bromophenol blue. The stained plugs were cut in half and placed, one per well, in the slots of a 0.75 % [w/v] agarose gel (GTG agarose, Sigma) made in 0.5 x TBE. The gel was submerged in 0.5 x TBE and field inversion gel electrophoresis was conducted at 4°C for 96 hours at 90 volts using a programmable device. The field inversion was carried out automatically by a PP1-100 device (MJ Research Inc., Cambridge, MA). The program used in this study was number 9 and had the following characteristics:

A: reverse time at beginning of ramp = 2 sec.
B: amount added to reverse time at each step = 2 sec.
C: forward time at beginning of ramp = 6 sec.
D: amount added to forward time at each step = 6 sec,
E: number of complete reverse and forward runs before starting over with initial values = 22
F: added to reverse increment at each step = -0.1
G: added to forward increment at each step = -0.6

Before starting the reverse/forward field flow, the gel was run for 10 minutes in the forward direction at 80 volts to allow migration of the DNA out of the wells and into the gel.

XI- DNA TRANSFER.

In this study, three different methods were used to transfer DNA to nylon membranes.

XI-1 Southern transfer.

Transfer of DNA fragments from agarose gels to nylon membranes were carried out using the technique of Southern (1975).

After electrophoresis was completed, the gel was stained in EtBr and photographed. The DNA was then denatured by soaking the gel in 0.4 M NaOH, 0.1 M NaCl for 45 minutes at room temperature with gentle agitation, rinsed with water and subsequently neutralized in 2 M NaCl, 0.7 M Tris-HCl [pH 7.5] for another 45 minutes. The gel was then soaked in 20 x SSC (3 M NaCl, 0.3 M Na Citrate pH 7.6), along with 3 pieces of 3 MM Whatman paper and one piece of nylon membrane (Genescreen, NEN-Dupont). Subsequently, the gel was placed in a glass tray filled with 20 x SSC on an inverted gel casting tray that was wrapped in a piece of 3MM Whatman paper saturated with 20 x SSC. The paper had extended wicks dipping into the solvent. The nylon membrane was placed on top of the gel, and this was in turn covered with the three pieces of Whatman paper and a stack (six inches) of paper towels. The next day, the set-up was disassembled and the membrane was rinsed for 5 minutes at room temperature in 20 x SSC and air dried. The DNA was fixed

to the membrane by placing the dry membrane in a vacuum oven at 80°C for 4 hours

XI-2 Bi-directional transfer.

This procedure, derived from Smith and Summers (1980), involves the acid depurination of large DNA fragments which improves their transfer to nylon membranes.

The depurination of DNA was performed for 15 minutes in 0.25 M HCl at room temperature with agitation. After three washes with water, the gel was transferred to a denaturation solution consisting of 0.5 M NaOH plus 1.5 M NaCl, and agitated for another 15 minutes. Neutralization of the gel was performed for 30 minutes in 20 mM NaOH, 1 M ammonium acetate. The denatured DNA was transferred simultaneously onto two membranes positioned on either side of the gel for 2 hours at room temperature. The membranes were baked at 80°C for 4 hours under vacuum and stored until needed.

XI-3 Dot blots.

Human-rodent DNA (obtained from hybrid cells), total genomic DNA and plasmid DNA were transferred using a BioRad dot blot filtration unit. Serial dilutions of the DNA was made in 200 µl of 1 x TE. To each tube, 40 µl of 1 M NaOH was added to denature the DNA. After 10 minutes of incubation at room temperature, 40 µl of 1 M Tris-HCl [pH 7.5] and 40 µl of 1 M HCl were added to the reactions and the tubes were, thereafter, kept on ice until needed. A Genescreen nylon membrane was cut

to the size of the filtration unit, and soaked for 10 minutes in water, and then in 6 x SSC. It was assembled on the dot blot filtration unit as specified by the manufacturers. The membrane was washed under vacuum with 6 x SSC, and the samples were applied, vacuum off, in the slots made by the apparatus. The samples were filtered by vacuum and the slots were washed twice with 6 x SSC. After filtration was complete, the membrane was released from the dot blot apparatus and washed in 2 x Denhardt's solution (100 x Denhart's solution is 2 % [w/v] polyvinyl pyrollidone, 2 % [w/v] BSA, 2 % [w/v] ficoll 400,). It was then baked at 80°C under vacuum, and stored at room temperature for future use.

XII- LABELING OF DNAS.

XII-1 Nick translation of DNAs.

Routinely, 0.5 to 1 μg of DNA was incubated for 10 minutes at 42°C in a solution containing 1 x NTB (table 2), 3 μl of dATP (250 μM), 3 μl of dGTP (250 μM), 4 μl of (α-^{32}P) dTTP (3,000 Ci/mM, ICN), 4 μl of (α-^{32}P) dCTP (3,000 Ci/mM, ICN) and 1 μl of 10 mM $MgCl_2$. DNase 1 (1 μg/ml) was made fresh in TM buffer (table 2). One unit of DNA polymerase 1 and 3 μl of the DNase 1 solution were added and the reaction was incubated at 15°C for 1 hour. The reaction was then heat-inactivated at 65°C and the DNA was extracted with PCI, and ethanol precipitated 4 times using ammonium acetate. The last pellet was thoroughly resuspended

in 200 µl of 1 x TE and the radioactivity counted using an LKB scintillation counter.

XII-2 Kinasing of DNAs.

200 ng of pure oligonucleotide (49 nucleotides) were labelled with 100 µCi of γ-^{32}P ATP (5,000 Ci/mmole, ICN) in the presence of 10 units of T4 polynucleotide kinase in 1 x LKB (table 2). After 1 hour at 37°C, the reaction was stopped by PCI extraction, and the DNA was purified through a series of ethanol precipitations as for nick-translation except that 1 µg of carrier DNA (E. coli genomic DNA) was added in the first precipitation in order to reduce oligonucleotide loss.

XII-3 Backfilling of 5' protruding termini.

Routinely, 40 µl of purified DNA restriction fragments were incubated, for 1 hour at room temperature, in 1 x Klenow buffer (table 2), and the protruding ends were made flush with the appropriate nucleotides using 5 units of the large Klenow fragment of E. coli DNA polymerase 1. The incubation was then placed for 10 minutes at 65°C and the DNA extracted with phenol/ether, ethanol precipitated using sodium acetate, and resuspended in deionized water.

XIII- HYBRIDIZATIONS.

The three procedures employed in this study for the detection of DNA sequence homology are presented below. The prehybridization, hybridization, and washes were carried out in heat-sealable bags. All membranes were dried, wrapped in saran-wrap and exposed in X-ray film holders with Kodak XAR5 film under intensifying screens (Dupont Cronex), at -70°C.

XIII-1 P-element related sequences.

This procedure is described in Sol et al., (1986). Briefly, nylon membranes, containing fixed DNA fragments, were incubated overnight at 42°C in 0.4 x buffer A (1 x buffer A is 0.15 M NaCl, and 0.011 M sodium citrate [pH 8.3]), supplemented with 1 x Denhardt's solution and 100 µg/ml E. coli DNA.

The hybridizations were performed, for 48 hours at 42°C in prehybridization solution supplemented with 100 µM ATP, 0.2 % [w/v] SDS and $3x10^6$ cpm of nick translated probe.

The washing solution was 0.25 x buffer A and 0.2 % [w/v] SDS. The blots were washed twice under constant agitation at room temperature for 30 minutes and then air-dried.

XIII-2 Oligonucleotide.

To minimize signal loss, DNA fragments that were to be hybridized with the kinased oligonucleotide probe were not depurinated prior to transfer to nylon membranes.

The membranes were soaked for 10 minutes in 3 x SSC and prehybridized for 2 hours at 42°C in 6 x SSC, 1 x Denhardt's solution, 0.5 % [w/v] SDS, 0.05 % [w/v] sodium pyrophosphate, and 50 µg/ml of E. coli DNA (Woods, 1984).

The hybridization was performed for 24 hours at 42°C in 6 x SSC, 1 x Denhardt's solution, 25 µg/ml E. coli DNA and $4x10^6$ cpm of probe, followed by two washes of 15 minutes at room temperature and one for 10 minutes at 42°C in 6 x SSC and 0.05 % [w/v] sodium pyrophosphate.

XIII-3 Satellite DNAs.

The nylon membranes were presoaked in 5 x SSC and prehybridized in 50 % [v/v] deionized formamide, 5 x Denhardt's solution, 5 x SSC, and 50 µg/ml E. coli DNA, for 24 hours at 42°C.

The hybridization was achieved in 50 % [v/v] formamide 5 x SSC, 1 x Denhardt's solution, 25 µg/ml E. coli DNA, 0.1 % [w/v] SDS and 0.05 % [w/v] sodium pyrophosphate for 24 hours at 42°C. Washes were performed in 0.2 % [w/v] SDS and 0.5 x SSC twice for 30 minutes at 37°C, and once at room temperature.

XIV- SEQUENCING STRATEGIES.

XIV-1 Chemical sequencing.

The complete nucleotide sequence of the 1.797 Kb EcoR1 satellite DNA cloned in pKS36 was performed using the chemical sequencing method of Maxam & Gilbert (1980). Restriction

fragments of this satellite DNA were recovered from agarose or acrylamide gels, labelled at their 5' end and strand separation was done according to Maniatis et al., (1982). The sole exception was that the "crush & soak" buffer used contained 1 mM EDTA (instead of 0.1 mM) and the denaturing buffer consisted of 0.05 % [w/v] bromophenol blue, 0.05 % xylene-cyanol, 1 mM EDTA, 50 mM NaOH and 7 % [w/v] sucrose. Both strands were sequenced in an overlapping manner, and the sequence was verified for accuracy by identification of predicted restriction endonuclease sites and fragment lengths.

M13-2 Dideoxy sequencing.

The partial sequences of the Kpn1 and EcoR1 satellite DNA cloned in pUC119 were performed by the dideoxy chain terminator sequencing procedure (Sanger et al., 1977).

Briefly, strains containing pUC119 derivatives were grown at 37°C in superbroth (Ljungquist and Bukhari, 1977), to an OD_{600} of 0.1 (1×10^8 cfu [colony forming units]/ml). The cells were infected with 0.1 ml of M13K07 phage (5×10^{10} plaque forming units [pfu]/ml) at a multiplicity of infection of 50, and grown overnight in superbroth media containing 70 µg/ml kanamycin and 150 µg/ml ampicillin. 1.5 ml of the overnight culture was transferred to an eppendorf tube and the cells pelleted by a 5 minutes centrifugation at 15,000 x g, 4°C. The cell-free supernatant fluid was mixed thoroughly with 200 µl of 20 % [w/v] PEG 6000 (polyethylene glycol, made in 2.5 M NaCl) and incubated for 15 minutes at room temperature. The phage

pellet was then harvested by centrifugation at 15,000 x g for 5 minutes at 4°C and resuspended in 100 µl of 1 x TE. After extraction with 50 µl of phenol, the resulting single stranded (SS) DNA was precipitated in 0.3 M sodium acetate and 70 % ethanol. This DNA was used directly for dideoxy sequencing using the BioRad M13 Klenow sequencing kit. Deoxyadenosine 5'[athio]-triphosphate (^{35}S, 1500 Ci/mmol.) was obtained from Dupont. The electrophoresis of samples (G, A, T, C reactions) was carried out at 1500 volts in an 8 % polyacrylamide gel (acrylamide/bis-acrylamide = 19/1), containing 8 M urea and 1 x TBE buffer. The gels were dried and exposed on Kodak XAR 5 film for 2 to 10 hours under intensifying screens at -70°C.

Chapter III

CHARACTERIZATION OF P-ELEMENT RELATED HUMAN REPETITIVE DNA SEQUENCES

" I am not willing to accept position effect variegation as some extraordinarily subtle effect of chromosome structure as a whole"

F. Crick (1979)

INTRODUCTION

Since the discovery and characterization of the maize mobile DS-Ac family by B. McClintock in the early 1950s (McClintock, 1951), the importance of mobile genetic elements has become increasingly apparent, as they are likely to be involved in mutagenic processes (such as DNA inversion, translocation, insertion etc) as well as in evolution. For instance, eukaryotic mobile genetic elements and DNA transposition have been found to play a key role as etiological agents in tumor induction (e.g., feline leukemia virus) and in a wide range of somatic and genetic variegation (e.g., Ac elements of maize, Ty elements of yeast [Roeder et al., 1980]). More recently, in **Drosophila melanogaster**, a repetitive and mobile element, called the P-element, has been found to play a role in birth defects and infertility in crosses between certain strains of the fruit fly (Rubin et al., 1982). This syndrome, called hybrid dysgenesis, was observed when males from a P-strain were mated with females from an M-strain, but not in other crosses (M-males x M-females, P-males x P-females, or M-males x P-females). The molecular basis of this phenomenon was extensively analyzed (Bingham et al., 1982, O'Hare and Rubin 1983, Spradling and Rubin 1983; Engels et al., 1990). Discrete DNA segments, the P-elements, were found in the genome of P-strains ranging in size from 500 base pairs (bp) to 2.9 Kb. The

shorter version of the P-elements appeared to have arisen from internal deletions of the 2.9 Kb active transposable element (Laski et al., 1986). The P-elements have 31 bp inverted terminal repeats and generate an 8 bp duplication of host sequences at the site of insertion. Furthermore, the functional 2.9 Kb long element contains four large open reading frames (ORFs) and encodes for its own transposase. This type of mobile genetic element has not, as yet, been detected as components of animal cell genomes, although remnants of ancient transposition events have been found in the form of direct repeats flanking dispersed repetitive DNA segments.

In an attempt to screen the human genome for the presence of repetitive and potentially mobile DNA segments, we have examined the human genome for P-element related sequences. The rationale behind our approach was based on the recent findings indicating a high degree of sequence conservation among certain eukaryotic genes such as the human and **Drosophila** epidermal growth factor genes (Livneh et al., 1985). Furthermore, the P-family of transposable elements induces, in **D. melanogaster**, genetic traits which are equivalent to certain birth defects in humans (Kidwell and Kidwell, 1976). The P-element that was used in this study had originally transposed in the locus 17C, and was cloned with its flanking **D. melanogaster** sequences in the plasmid pπ25.1 (Spradling and Rubin, 1982).

Reported in this section is the cloning and characterization of a human repetitive 1,797 bp EcoR1 fragment that hybridizes, under low stringency, to a probe consisting of internal fragments of the P-element (Sol et al., 1986). This HeLa DNA fragment was cloned into the EcoR1 site of pBR322 and the resulting plasmid, pKS36, was mapped by restriction endonucleases and completely sequenced. Our analysis indicates that this DNA segment contains many tandem copies of the human satellite II and III consensus sequence 5' TTCCA 3', and variations of this pentameric repeat, but the pure sequence comprises only 35% of the total DNA. Though there are several stretches of sequence similarities to the P-element, no structural feature indicative of a mobile genetic element were discernible. Thus, this cloned segment of DNA appears to be a widely divergent member of the highly repetitive satellite II and III family (Corneo et al., 1971; Cooke and Hindley, 1979; Deininger et al., 1981).

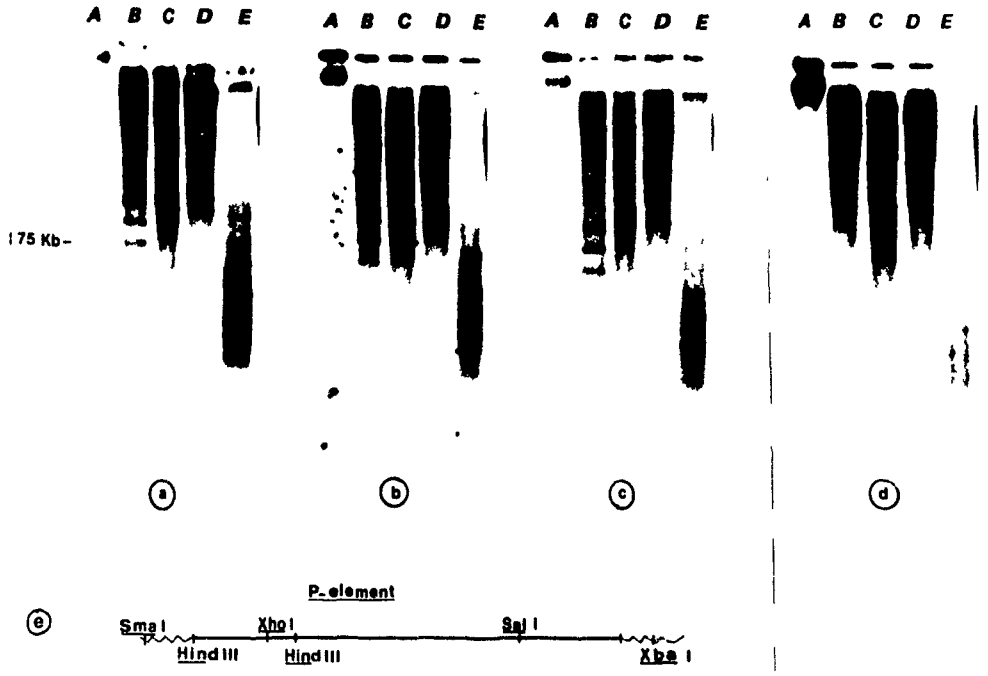
RESULTS

1- DETECTION OF P-ELEMENT RELATED SEQUENCES IN THE HUMAN GENOME.

The P-element, a transposable element found in **D. melanogaster**, was cloned along with flanking **Drosophila** sequences in the plasmid pπ25.1 (Spradling and Rubin, 1982). Depicted in figure 7 -e is a simplified map of the cloned element along with the restriction sites used in this experiment. To determine which part of the P-element or its flanking **Drosophila** sequences would be conserved in the human genome, three restriction fragments, each containing discrete regions of the cloned DNA, were generated. Fragment a), containing the entire P-element plus the flanking **Drosophila** sequences, was obtained via Sma1-Xba1 double digestion of the plasmid. Fragment b) generated with HindIII, consisted of the small P-element segment from its left extremity. Fragment c), a SalI-XhoI restriction fragment, represented internal sequences of the P-element. Each of these three fragments were nick-translated and used to probe Southern blots of **HeLa** DNA digested with EcoR1, Pst1, HindIII and HaeIII, lanes B, C, D, and E, respectively (figure 7 panels a to c).

Figure 7 : Hybridization of the fragments of the 2.9 Kb long P-element and flanking Drosophila sequences to Southern blots of HeLa DNA. The genomic DNA was hydrolyzed with EcoR1 (lanes B), Pst1 (lanes C), Hind III (lanes D), and Hae III (lanes E). Lanes A of each blot contain uncut DNA. The genomic DNA was isolated, hydrolyzed with restriction endonucleases, subjected to agarose gel electrophoresis, blotted and hybridized in 0.4 x buffer A as per Materials and Methods with (α-^{32}P)-labelled fragments of the P-element: a) Sma1-Xba1 probe (containing the genomic flanking sequences); b) Hind III P-element probe; c) Sal1-Xho1 probe (containing the internal sequences of the P-element). In panel d) Adenovirus 2 DNA was used as a probe. Panel e) displays a simplified map of the pπ25.1 insert; the restriction sites, within the P-element and the Drosophila genomic sequences, used in this experiment, are indicated.

Figure 7


A B C D E
A B C D E
A B C D E
A B C D E
175 Kb-
a
b
c
d
e
P-element
Sma I
Hind III
Xho I
Hind III
Sal I
Xba I

All hybridizations were performed in 0.4 x buffer A as indicated in Materials and Methods. Under these very permissive conditions, very small regions of partial homology to repetitive sequences could be detected (Beltz et al., 1983), with ATP and calf thymus DNA included to reduce the background of non-specific hybridization expected under these low stringency conditions. The hybridization profiles generated by the three P-element probes appeared to be similar, thus indicating that the flanking Drosophila genomic sequences did not substantially contribute to these results. The most striking feature observed in the hybridizations was the series of EcoR1 bands (lanes B) that shared sequence homology to the various probes. The major band appearing at approximately 1.75 Kb, over the background of non-specific hybridization expected under these permissive conditions, is indicated on the left side of the figure. No other distinct bands were observed over the background when HeLa DNA was hydrolyzed with Pst1, HindIII or HaeIII (lanes C, D, and E). In order to study the specificity of the hybridizations, Adenovirus 2 DNA was labelled by nick-translation and used to probe, under the same conditions, an identical Southern blot of HeLa-cleaved DNA (panel d). Although the A+T content of the viral DNA (62 %) is very similar to that of the P-element (O'Hare and Rubin, 1983), no bands could be distinguished over the background of non-specific hybridization. This suggested that the hybridization patterns observed

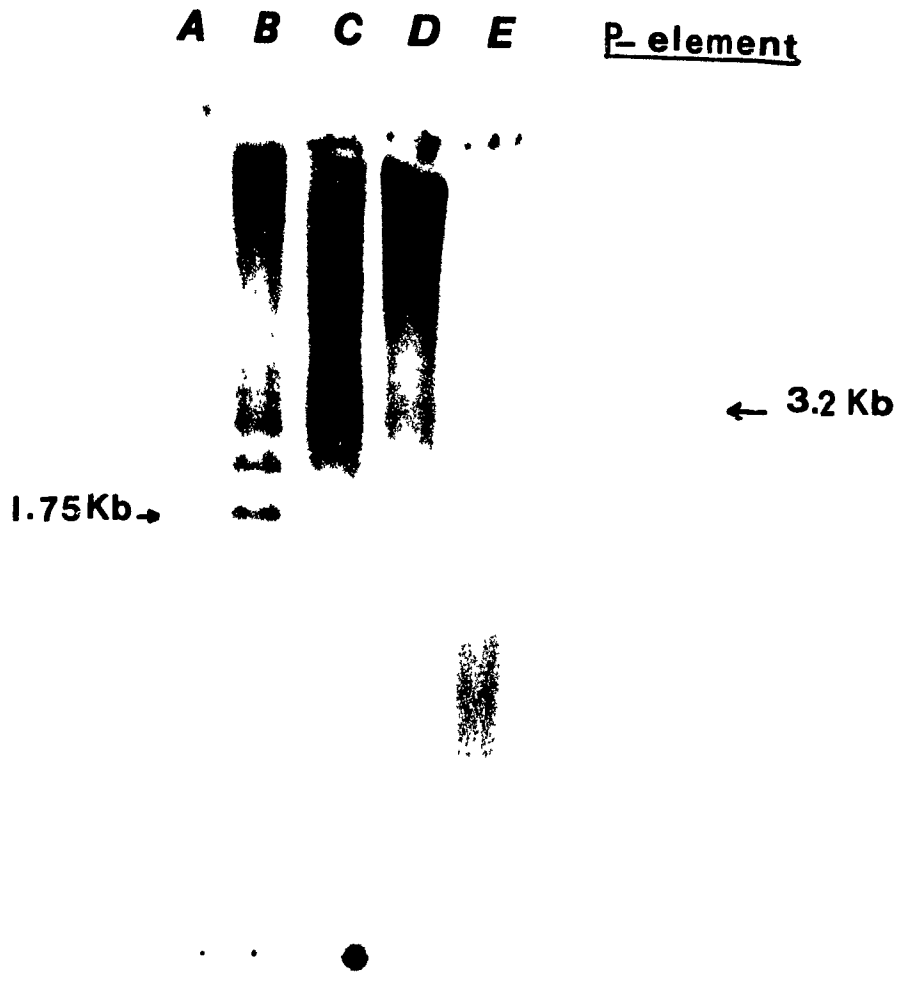
between the P-element and the EcoR1-cleaved HeLa DNA were not due to the A-T richness of the sequences alone.

II- THE P-ELEMENT RELATED SEQUENCES ARE REPETITIVE.

The human EcoR1 fragments of 1.7-1.8 Kb were electroeluted from a 1 % agarose gel, and nick-translated as per Materials and Methods. The heterogeneous probe, as indicated in figure 8, was hybridized to Southern blots of restriction enzyme cleaved-HeLa and pπ25.1 DNAs. The hybridization conditions were the same as before. A similar pattern of hybrid bands was observed with EcoR1-cleaved genomic DNA (lane B), and again no distinctive bands were observed when HeLa DNA was hydrolyzed with Pst1, HindIII, or HaeIII (lanes C, D and E, respectively). The observable series of arithmetical EcoR1 restriction fragments suggested that the isolated DNAs were rich in repetitive DNA sequences. Plasmid pπ25.1, when hydrolyzed with HindIII, generated three fragments: a large fragment consisting of the cloning vector, a 3.2 Kb fragment containing most of the P-element and some flanking genomic sequences, and a small 500 bp fragment containing the extreme left-end of the cloned element (refer to figure 7 e).

Figure 8 : Southern blots of hydrolyzed HeLa and pπ25.1 DNAs hybridized to the human 1.7-1.8 Kb EcoR1 DNA fragments. The genomic DNA was hydrolyzed as in figure 5 with EcoR1 (lane B), Pst1 (lane C), Hind III (lane D), and Hae III (lane E). Lane A is uncut DNA. Plasmid pπ25.1 was hydrolyzed with the restriction endonuclease Hind III. The 3.2 Kb fragment indicated here corresponds to the P-element and **Drosophila** sequences located to the right of the HindIII site (refer to the simplified map in figure 7). The cleaved DNAs were subject to electrophoresis, transferred to nylon membranes and hybridized as per Materials and Methods. The probes consisted of human EcoR1 fragments ranging from 1.7 to 1.8 Kb that were eluted from an agarose gel and labelled by nick-translation.

Figure 8



Human 1.7- 1.8 Kb EcoR I fragments

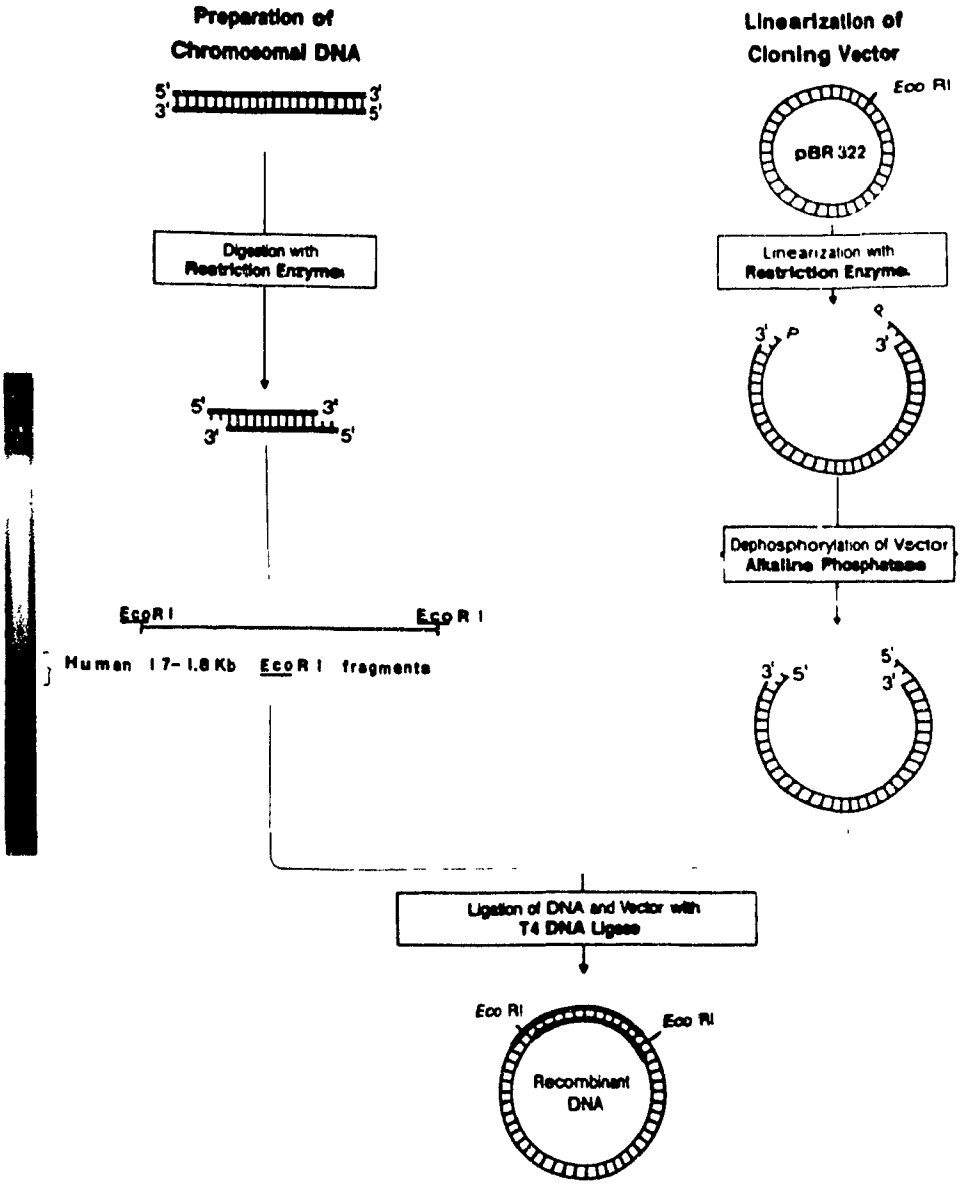
Upon hybridization of the hydrolyzed plasmid with the human 1.7-1.8 Kb EcoR1 fragments as a probe, a single hybrid band was detected, as seen on the right panel of figure 8, that corresponded to the large P-element 3.2 Kb fragment. No hybridization was detected with the cloning vehicle, or with the smallest of the HindIII fragments. It is entirely possible that, under these conditions, this fragment was too short for quantitative transfer to the nylon membrane or to be able to form stable hybrids.

III- CLONING OF THE P-ELEMENT RELATED SEQUENCES.

To isolate and characterize the human DNAs that shared apparent homologies with the Drosophila P-element, the EcoR1 restriction fragments ranging from 1.7 Kb to 1.8 Kb were gel purified and cloned into pBR322 (figure 9). The cloning vector was linearized at its unique EcoR1 site and its 5' ends were dephosphorylated via CIAP treatment as described in Materials and Methods. The ligation mixture was transformed into E. coli DH1 competent cells and plated on LB plates supplemented with ampicillin. The transformants were subsequently grown on gridded master plates and the size of their plasmid was assessed by gel electrophoresis. The gels were then blotted onto Genescreen nylon membranes.

Figure 9 : Cloning strategy of the 1.7-1.8 Kb EcoRI DNA fragments. Left panel: HeLa genomic DNA was isolated, as per material and methods, and hydrolyzed with the restriction endonuclease EcoRI. After electrophoresis on a 1% agarose gel, the series of EcoRI DNA fragments ranging from 1.7 to 1.8 Kb were electroeluted as per material and methods. Right panel: Plasmid pBR322 was linearized at its unique EcoRI site, and the ends of the molecule were dephosphorylated via alkaline phosphatase, as described in material and methods.

Figure 9

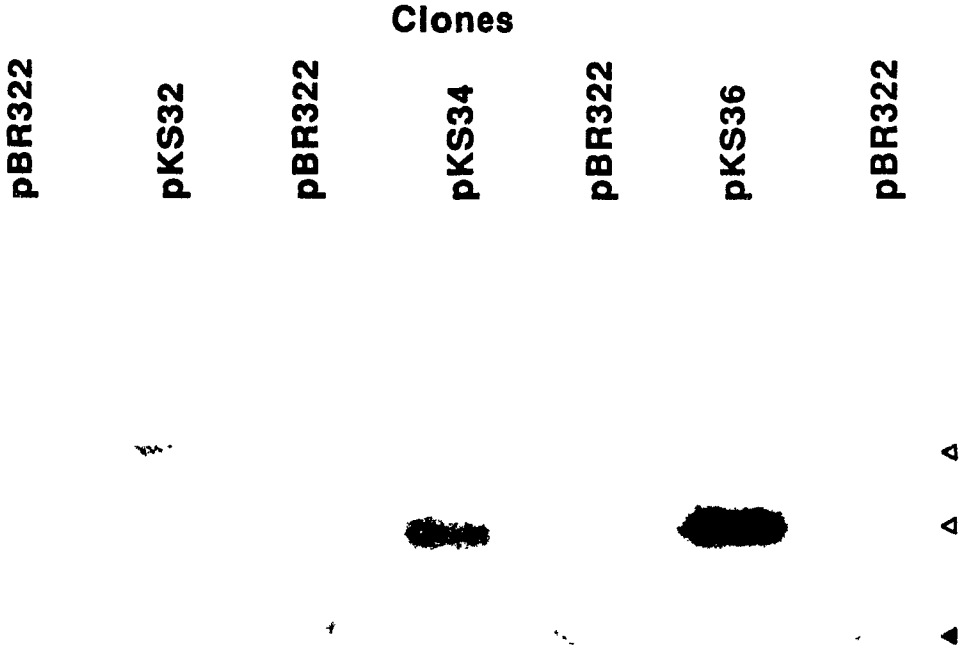

Preparation of Chromosomal DNA
Linearization of Cloning Vector
Eco RI
pBR 322
Digestion with Restriction Enzymes
Linearization with Restriction Enzyme
Dephosphorylation of Vector Alkaline Phosphatase
EcoR I
EcoR I
Human 17-18 Kb EcoR I fragments
Ligation of DNA and Vector with T4 DNA Ligase
Eco RI
Eco RI
Recombinant DNA

A total of 725 clones were screened, under low stringency, for P-element related sequences using the Sal1-Xho1 internal restriction fragment as a probe. Among the clones tested, 20 displayed varying degrees of homology to the probe as determined by an increased hybridization signal over the background of non-specific hybridization to pBR322.

One of these clones, showing the strongest signal, is represented in figure 10. The signal of hybridization between pKS36 and the P-element was retained, unlike that of the other remaining clones, when the salt concentration was increased (from 0.4 x buffer A to 2.5 x buffer A). The hybridization was completely lost with 5 x buffer A (data not shown). These results suggested that the sequence homology between the two heterologous DNAs (the human EcoR1 insert of pKS36 and the Sal1-Xho1 fragment of the Drosophila P-element) was due to either short stretches of sequence or long sequences of weak homology (Beltz et al., 1983).

Figure 10 : Hybridization to the P-element of the recombinant plasmids. Transformed cells were cracked open, their DNA content subjected to electrophoresis, transferred to nylon membranes and hybridized with the nick-translated internal restriction fragment of the P-element as per Materials and Methods. Three colonies containing recombinant plasmids are shown: pKS32, pKS34, and pKS36. Note that pKS32 contains a double insert whereas pKS34 and pKS36 contain a single insert.

Figure 10

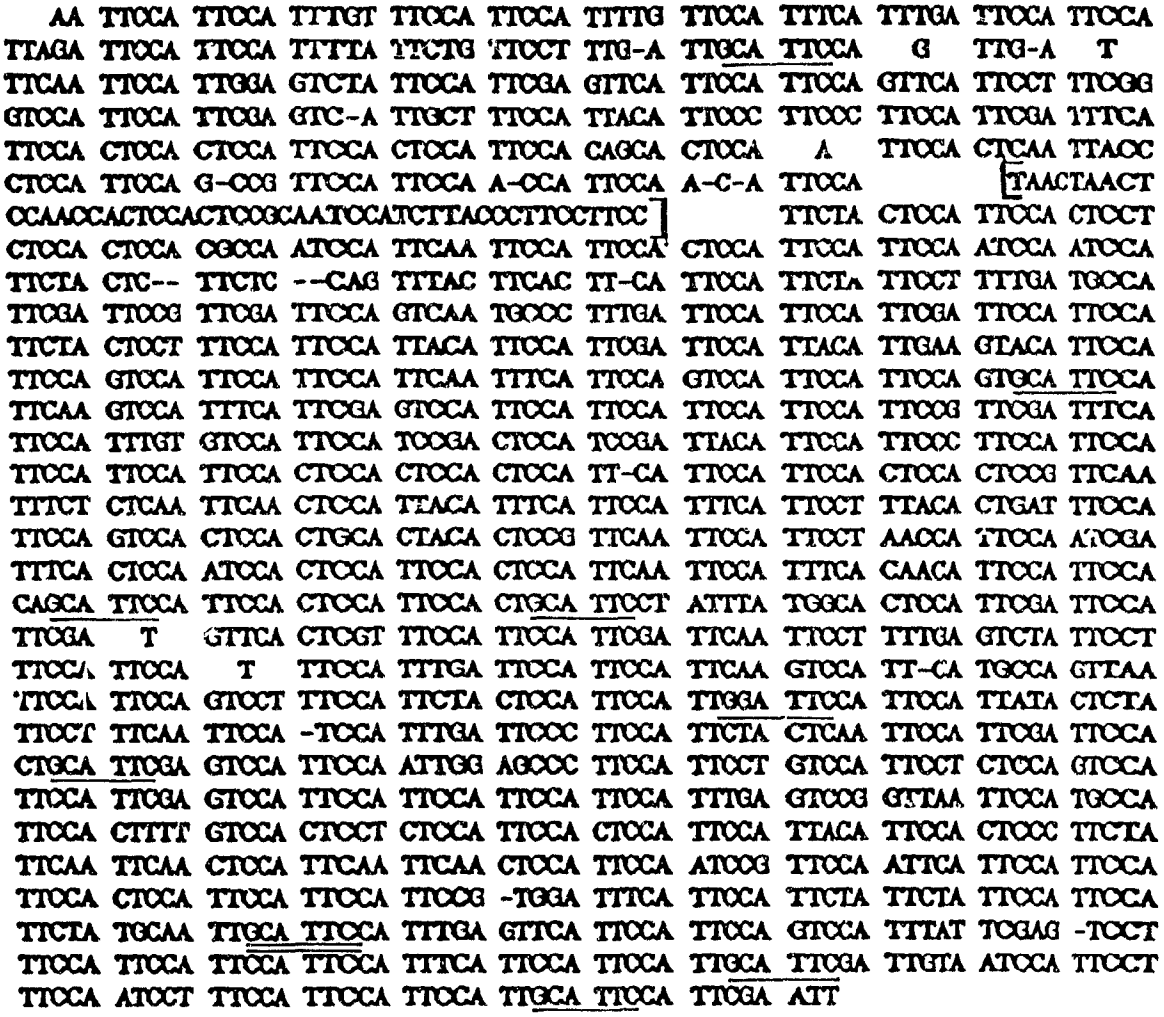


IV- SEQUENCE OF PKS36.

The EcoR1 insert contained in pKS36 was mapped extensively with a battery of restriction endonucleases. Few unique restriction sites were detected in this analysis (Cla1, Rsa1, Ava1 and HpaII), whereas numerous Taq1 and Hinf1 sites were present throughout the insert. The availability of these different sites allowed us to prepare a strategy for the sequencing of the entire insert via the method of Maxam and Gilbert (1980). Both strands were sequenced in an overlapping manner. The entire EcoR1 fragment of pKS36 is 1.797 Kb in length (figure 11). This human DNA fragment did not contain any of the characteristic traits of a eukaryotic mobile element, such as terminal inverted repeats (like the P-element) or a poly(A) tail (like the human Alu element). No large open reading frames (ORF) could be found, although numerous short potential ORFs were present. The striking feature, as depicted by the sequencing of this DNA, was the presence of numerous pentameric repeats seemingly derived from the consensus 5' TTCCA 3' (figure 11). The sole exception in the whole sequence was the presence of a region of 49 bp, located between bp 324 and bp 372, seemingly devoid of pentameric repeats (region within brackets in figure 11).

Figure 11 : Pentameric organization of pKS36. Illustration of the repetitive nature of clone pKS36. The sequence is presented as blocks of 5 bp organized in a head to tail manner. The recurrent pentamer 5' TTCCA 3' or its derivative pentameric blocks are present throughout the insert with the exception of the region from bp 327 to bp 372, indicated within brackets, called the "49 mer". Sequences underlined and double underlined are reference points for Chapter V, discussion.

Figure 11


AA TTCCA TTCCA TTTGT TTCCA TTCCA TTTTG TTCCA TTTCA TTTGA TTCCA TTCCA
TTAGA TTCCA TTCCA TTTTA TTCTG TTCCT TTG-A TTGCA TTCCA G TTG-A T
TTCAA TTCCA TTGGA GTCTA TTCCA TTCGA GTTCA TTCCA TTCCA GTTCA TTCCT TTCGG
GTCCA TTCCA TTCGA GTC-A TTGCT TTCCA TTACA TTCCC TTCCC TTCCA TTCGA TTTCA
TTCCA CTCCA CTCCA TTCCA CTCCA TTCCA CAGCA CTCCA A TTCCA CTCAA TTACC
CTCCA TTCCA G-CCG TTCCA TTCCA A-CCA TTCCA A-C-A TTCCA TAACTAACT
CCAACCACTCCACTCCGCAATCCATCTTACCCTTCCTTCC TTCTA CTCCA TTCCA CTCCT
CTCCA CTCCA CGCCA ATCCA TTCAA TTCCA TTCCA CTCCA TTCCA TTCCA ATCCA ATCCA
TTCTA CTC-- TTCTC --CAG TTTAC TTCAC TT-CA TTCCA TTCTA TTCCT TTTGA TGCCA
TTCGA TTCCG TTCGA TTCCA GTCAA TGCCC TTTGA TTCCA TTCCA TTCGA TTCCA TTCCA
TTCTA CTCCT TTCCA TTCCA TTACA TTCCA TTCGA TTCCA TTACA TTGAA GTACA TTCCA
TTCCA GTCCA TTCCA TTCCA TTCAA TTTCA TTCCA GTCCA TTCCA TTCCA GTGCA TTCCA
TTCAA GTCCA TTTCA TTCGA GTCCA TTCCA TTCCA TTCCA TTCCA TTCCG TTCGA TTTCA
TTCCA TTTGT GTCCA TTCCA TCCGA CTCCA TCCGA TTACA TTCCA TTCCC TTCCA TTCCA
TTCCA TTCCA TTCCA CTCCA CTCCA CTCCA TT-CA TTCCA TTCCA CTCCA CTCCG TTCAA
TTTCT CTCAA TTCAA CTCCA TTACA TTTCA TTCCA TTTCA TTCCT TTACA CTGAT TTCCA
TTCCA GTCCA CTCCA CTGCA CTACA CTCCG TTCAA TTCCA TTCCT AACCA TTCCA ATCGA
TTTCA CTCCA ATCCA CTCCA TTCCA CTCCA TTCAA TTCCA TTTCA CAACA TTCCA TTCCA
CAGCA TTCCA TTCCA CTCCA TTCCA CTGCA TTCCT ATTTA TGGCA CTCCA TTCGA TTCCA
TTCGA T GTTCA CTCGT TTCCA TTCCA TTCGA TTCAA TTCCT TTTGA GTCTA TTCCT
TTCCA TTCCA T TTCCA TTTGA TTCCA TTCCA TTCAA GTCCA TT-CA TGCCA GTTAA
TTCCA TTCCA GTCCT TTCCA TTCTA CTCCA TTCCA TTGGA TTCCA TTCCA TTATA CTCTA
TTCCT TTCAA TTCCA -TCCA TTTGA TTCCC TTCCA TTCTA CTCAA TTCCA TTCGA TTCCA
CTGCA TTCGA GTCCA TTCCA ATTGG AGCCC TTCCA TTCCT GTCCA TTCCT CTCCA GTCCA
TTCCA TTCGA GTCCA TTCCA TTCCA TTCCA TTCCA TTTGA GTCCG GTTAA TTCCA TGCCA
TTCCA CTTTT GTCCA CTCCT CTCCA TTCCA CTCCA TTCCA TTACA TTCCA CTCCC TTCTA
TTCAA TTCAA CTCCA TTCAA TTCAA CTCCA TTCCA ATCCG TTCCA ATTCA TTCCA TTCCA
TTCCA CTCCA TTCCA TTCCA TTCCG -TGGA TTTCA TTCCA TTCTA TTCTA TTCCA TTCCA
TTCTA TGCAA TTGCA TTCCA TTTGA GTTCA TTCCA TTCCA GTCCA TTTAT TCGAG -TCCT
TTCCA TTCCA TTCCA TTCCA TTTCA TTCCA TTCCA TTGCA TTCGA TTGTA ATCCA TTCCT
TTCCA ATCCT TTCCA TTCCA TTCCA TTGCA TTCCA TTCGA ATT

U- SEQUENCE HOMOLOGY SEARCH.

In order to clarify the relationship between the P-element and pKS36, sequence comparisons were carried out using the Sequence Analysis Facilities provided by the University of Montreal through the Clinical Research Institute of Montreal (programs Bestfit and Seqfit). Several short stretches of partial homology were found between the two sequences. Three of these, displayed in figure 12, show homology between internal sequences of the P-element and non-clustered segments of pKS36. The percent of sequence similarities, ranging from 72 % and 65 %, are consistent with the hybridization detected, under permissive hybridization condition, between the two DNAs.

Figure 12 : Regions of sequence similarity between the pKS36 EcoR1 insert (top strand) and the P-element (bottom strand). The matches between the nucleotide sequences are indicated by stars (*). The nucleotide positions of each sequence are indicated at the 5' and 3' ends of each strand. At the right of each sequence is indicated the percent similarity, calculated by considering pKS36 as the reference strand. Dots and supplementary nucleotides correspond to misalignment and folding of one of the strands.

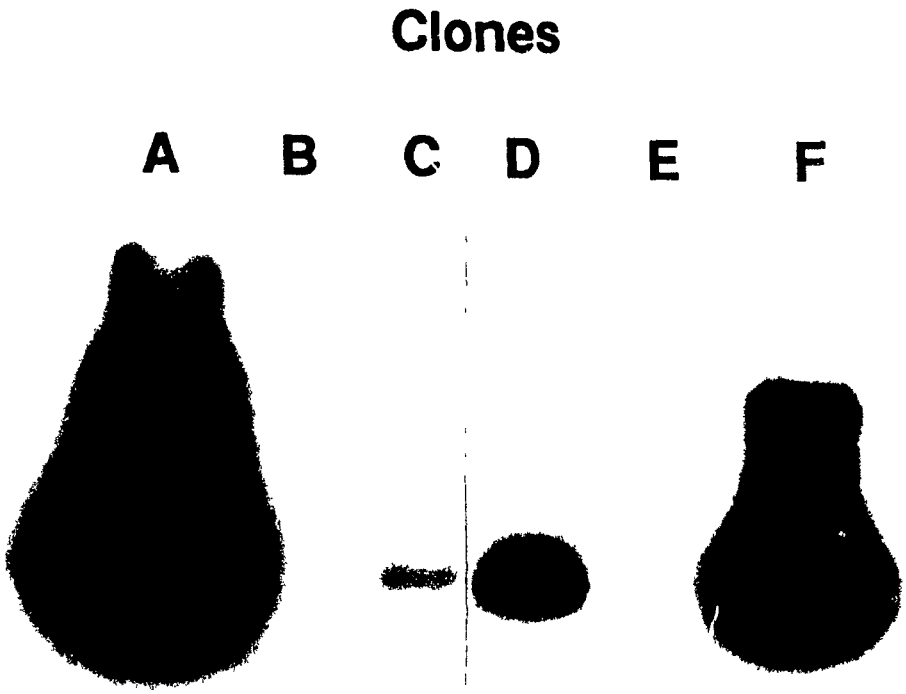
Figure 12

VI- RELATIONSHIP BETWEEN THE P-ELEMENT "POSITIVE" CLONES.

Of the remaining clones that showed reduced levels of hybridization to the P-element under low salt concentration (0.4 x buffer A), seven were randomly picked and tested for their homology to clone pKS36. The hybridization conditions were increased to a salt concentration of 5 x SSC, which corresponds to low stringency when comparing DNA sequences within the same genus. With the exception of clone pKS544, all clones displayed some degrees of homology to pKS36. Three of these are presented in figure 13. The background of non-specific hybridization appeared very low, as seen by the signal displayed by the negative control plasmid pBR322 (lanes B). The positive control (plasmid pKS36, lane A) displayed, as expected, an intense signal of hybridization. Plasmids pKS40, pKS30 and pKS28 (lanes C, D and F respectively) displayed increasing amounts of hybridization to pKS36. From this analysis, it appeared that the clones screened for P-element related sequence were more related to one another than to the P-element.

Figure 13 : Relationship between the clones cross-hybridizing to the P-element. Southern blot of recombinant plasmids (lanes A to F) hybridized under low stringency (5 x SSC) to the (α-^{32}P)-labelled EcoR1 insert of pKS36. Lane A contains pKS36 as a positive control for 100 % homology to the probe. Lanes B is pBR322, and lane E contain, as negative control, a recombinant plasmid that do not cross-hybridize to the P-element. Plasmids pKS40 (lane C), pKS30 (lane D), and pKS28 (lane F) cross-hybridized under permissive conditions (0.4 x buffer A) to the P-element and display here increasing homology to pKS36, respectively.

Figure 13



DISCUSSION

In order to screen the human genome for the presence of repetitive (and potentially mobile) elements, HeLa DNA was cleaved with various endonucleases and examined, via Southern blotting, for homology to a Drosophila transposon, the P-element (Rubin et al., 1982). The Southern hybridizations were performed under low stringency, but with a vast excess of unlabeled competitor DNA and ATP, to reduce the background of non-specific hybridization that one might expect under such permissive conditions (Beltz et al., 1983). After detection, human EcoR1 fragments containing sequences that cross-hybridized with the P-element were cloned into pBR322, and the entire 1.797 Kb nucleotide of one of these clones was determined (Sol et al., 1986).

Considering the very permissive hybridization conditions used to probe HeLa DNA with the P-element, it was entirely possible that no large region of homology would be detected (Beltz et al., 1983). This supposition was reinforced when it was found that the hybrid formation between the P-element and the EcoR1 cloned fragments was destabilized by an increase in the salt concentration from 0.4 x buffer A to 2.5 x buffer A.

A computer analysis performed between the P-element and the sequenced insert of clone pKS36 revealed the presence, on the P-element, of several short regions ranging from 30 to 50

bp, and non-uniformly distributed along the sequence, with 65 to 72 % similarity to segments of the pKS36 cloned sequence. The three largest regions displayed in figure 12 occur within the ORF1 and ORF2 of the P-element encoded transposase (O'Hare and Rubin, 1983), however, no homologies were detectable at the protein level between potential polypeptides encoded by the two DNAs. The biological significance of the homologous DNA sequences is not apparent at this time, as the roles of these homologous segments in pKS36 are not currently known. The degree of relationship between different clones analyzed in this study revealed the existence of two classes of related sequences: the pKS36 closely related sequences (e.g., pKS28) and the pKS36 divergent sequences (e.g.: pKS30). The inter-relationships that were displayed by the clones obtained in this study indicated that the human DNA segments that showed cross-hybridization to the P-element did not appear to be evolutionary related to the D. melanogaster element.

Nonetheless, our serendipitous cloning of the pKS36 EcoR1 insert allowed us to prepare the first complete nucleotide sequence of a human satellite DNA (Sol et al., 1986). Previous studies have demonstrated that human satellite II and III DNA's were characterized by the presence of multiple copies of the pentameric consensus sequence 5' TTCCA 3' (Frommer et al., 1982, Deininger et al., 1981, Cooke and Hindley, 1979). None of the other human satellite DNAs sequenced to date have been found to be interrupted by a non-satellite segment in the way

pKS36 appears to be. This unique feature, characteristic of pKS36-like satellite DNA, has been termed "the 49 mer".

In the next chapters of this thesis, the relationship between different members of this human satellite DNA family will be studied, and the organization of these sequences at the genomic and chromosomal levels will be analyzed. Finally, mechanisms which may play a role in the maintenance and amplification of such non-genic repetitive sequences will be discussed.

Chapter IV

ANALYSIS OF CLONED SATELLITE FAMILY MEMBERS.

The history of life is decidedly non-random
Simpson, (1964)

INTRODUCTION

The highly repeated human satellite DNAs are generally defined as compacted and late-replicating sequences that are organized in tandem arrays within the centromeric and telomeric heterochromatic regions of mammalian chromosomes (Corneo et al., 1971; Beauchamp et al., 1979). A function for these enigmatic DNA segments has yet to be discovered, but it is currently believed that they play a role in the structural integrity of chromosomes (Yamamoto and Miklos 1978; Manuelidis, 1982; Holden et al., 1985). Alternatively, such simple repeated sequences may have no indispensable function and may be by-products of genome rearrangement and/or aberrant DNA replication, as suggested by the selfish DNA model (Doolittle and Sapienza, 1980; Orgel and Crick, 1980). These sequences thus would be free to amplify and mutate in the absence of selection pressure.

It is possible that development of ideas for alternative roles for satellite DNA will be achieved by the detailed analysis of their structure and evolutionary behavior as non-coding elements.

Originally isolated as distinct peaks by virtue of their buoyant density in cesium salt density gradients, the classical satellite II and III DNAs soon appeared to represent a

heterogeneous population of divergent (though related) tandemly organized, simple, repeated DNA sequences.

The study presented in this chapter was undertaken to gain insight into the manner in which members of a family of related satellite II and III DNAs have evolved from a potential parental consensus sequence, namely the pentameric array 5' TTCCA 3'. We first screened a genomic library of human 1.8 Kb EcoR1 fragments for the presence of repeated satellite sequences that share sequence homology to a previously cloned and characterized EcoR1 satellite II DNA, clone pKS36 (Sol et al., 1986) and/or to a Kpn1 satellite II DNA, clone pBK1.8[20] (Shafit-Zagardo et al., 1982). Clone pKS36 contains a 1,797 bp EcoR1 satellite DNA that was shown to be interrupted in its tandem pentameric array by a region of 49 bp (called the 49 mer) devoid of the satellite II and III consensus sequence 5' TTCCA 3' (Chapter III). Clone pBK1.8[20], containing a 1.8 Kb Kpn1 satellite DNA, was kindly provided by Dr. J. Maio. In this study, pBK1.8[20] is shown by southern hybridization analysis not to contain the 49 mer non-satellite region of pKS36, even though it is highly homologous to the rest of the pKS36 satellite sequence. This property was taken into account in order to differentiate between satellite DNA directly related to pKS36, and therefore containing the "49 mer", from divergent clones (such as pBK1.8[20]).

The microheterogeneity of pKS36-like satellite DNA clones isolated in our laboratory was determined by virtue of their specific Taq1 and Hinf1 restriction patterns, as any of these sites

may be generated by a single point mutation within the consensus pentamers (Chapter I, figure 5). This series of clones, as well as pBK1.8[20], were partially sequenced, and compared to published satellite DNA sequences. These analysis revealed the hypervariability of the cytosine residues present in the pentameric consensus repeat 5' TTCCA 3'. Mechanisms that may cause such mutational hotspots within the consensus sequence are then discussed.

Using the DNA of a panel of rodent-human hybrid cell lines, the chromosomal location of two families of related satellite DNA, represented respectively by pKS36 and pBK1.8[20], was determined. The family of sequences related to pKS36 and pBK1.8[20] was found to comprise 2 to 3 % of the total genomic DNA of HeLa cells, and to reside mainly on chromosomes 7, 12, 14, 15, 16 and 22.

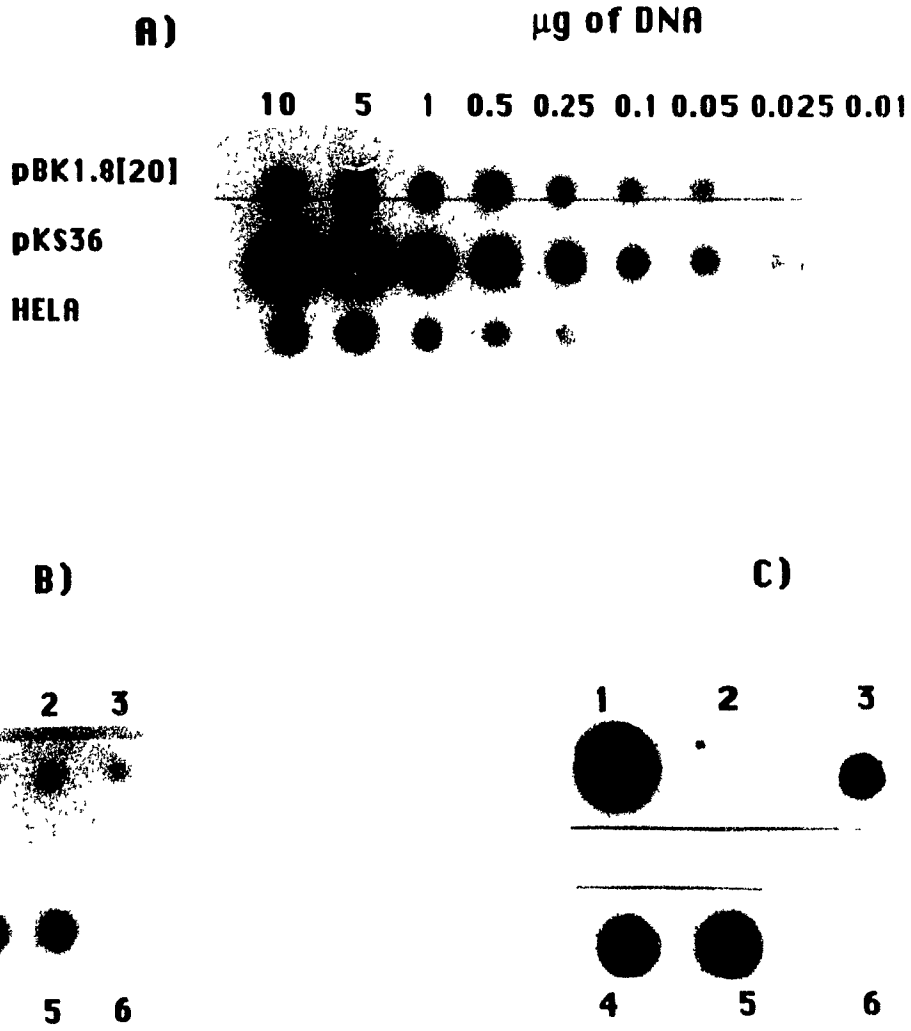
RESULTS

I- THE 49 MER REGION OF PKS36 DEFINES SPECIFIC CLASSES OF REPEATED DNA SEQUENCES.

A series of dot blot hybridizations were performed on DNA from a variety of species in an effort to determine the genetic distribution, relatedness and copy number of two cloned satellite DNAs: clone pKS36 (Sol et al., 1986) and pBK1.8[20] (Shafit-Zagardo et al., 1982). The later clone contains a 1.8 Kb Kpn1 satellite DNA, as the prototype for the human Kpn1 tandemly arrayed family of satellite DNAs. The probes used in the present experiments were either the gel-purified and nick-translated EcoR1 or Kpn1 inserts of plasmids pKS36 or pBK1.8[20], respectively. In addition, a 49 nucleotide long oligomer, of the unique region from base pair 324 to base pair 372 of clone pKS36 (called the 49 mer), was graciously synthesized by Dr. D. Garfinkel (NCI, Frederick, MD), end-labeled, as indicated in Chapter 2, and used as a probe. Figure 14 display the autoradiograms of hybridization experiments performed on serial dilutions of pBK1.8[20], pKS36, HeLa, AK143, Rat and E. coli DNAs, using either the 1.797 Kb EcoR1 insert (panel A), the 1.8 Kb Kpn1 insert (panel B) or the 49 mer oligonucleotide (panel C) as probes. Clone pBK1.8[20] displayed a high degree of homology to pKS36 as seen by hybrid formation between the two DNAs under high stringency (figure 14-A, lane pBK1.8[20] and 14-B, lanes 1 to 3).

Figure 14: Dot blot hybridizations. The probes used in this experiments were; Panel A: pKS36 nick translated EcoR1 satellite DNA; Panel B: pBK1.8[20] nick-translated Kpn1 satellite DNA. Panel C: pKS36 kinased 49 mer. The DNAs were transferred to nylon membranes as per Chapter 2. Panel A: serial dilution of plasmid pBK1.8[20], pKS36, and genomic HeLa DNA. Panel B: Lane 1, 2 and 3 contain 10 μg, 0.1 μg and 0.01 μg of pKS36 plasmid DNA respectively. Lanes 4, 5, and 6 contain 10 μg, 5 μg, and 0.05 μg of genomic HeLa DNA. Panel C: Lanes 1, 2, 3, and 6 contain 10 μg of pKS36, E. Coli, Rat2, and pBK1.8[20] DNAs, respectively. Lanes 4 and 5 contain 20 μg of HeLa and AK143 genomic DNAs, respectively.

Figure 14


A)
µg of DNA
10 5 1 0.5 0.25 0.1 0.05 0.025 0.01
pBK1.8[20]
pKS36
HELA
B)
1 2 3
4 5 6
C)
1 2 3
4 5 6

However, the Kpn1 element did not contain sequences homologous to the 49-mer found in pKS36, as it failed to cross-hybridize to this unique region of pKS36 (figure 14-C, lane 6). The oligonucleotide probe (the 49 mer) failed to hybridize to E. coli DNA (figure 14-C, lane 2), but a strong hybridization signal was detected, under these low stringency conditions, with Rat DNA (figure 14-C, lane 3). The significance of this observation remains, at the moment, unclear. The intensities of the hybridization signals of the EcoR1 probe to known amounts of its homologous DNA, pKS36 (figure 14-A), were recorded by laser densitometry, and the ratio of copy number/signal intensity determined. This signal ratio was then used to determine, by recording the hybridization signal of the EcoR1 probe to known amount of HeLa genomic DNA, the genomic copy number of this satellite DNA element. An example of the calculations involved in determining the copy number of pKS36 is presented in Appendix II. Assuming the content of DNA per nucleus to be 6 billion base pairs (Alberts et al., 1983), this 1.797 Kb satellite DNA (and other closely related sequences) was estimated to represent approximately 2 % of the genome of HeLa cell. Though similar results were obtained with the 1.8 Kb Kpn1 satellite probe, less than 1 % of the genomes of HeLa and AK143 cells could hybridize to the oligonucleotide probe (figure 14-C, lanes 4 and 5). These results may suggest that the actual copy number of the pKS36 family of satellite DNA (containing one copy of the 49 mer) is less than 1 % of the human genome.

II- ANALYSIS OF A RODENT-HUMAN HYBRID CELL PANEL.

The human chromosome make-up of the hamster/human cell-line panel used in this study is presented in table 3. Equal quantities of the 10 hybrid DNA's (wc1 to wc10), along with the control hamster DNA sample, were denatured, applied onto a nylon membrane in a BioRad dot blot apparatus at different concentrations, and hybridized under high stringency to the 1.797 Kb EcoR1 insert contained in clone pKS36. After autoradiography, the probe was removed from the membrane by treatment with a basic solution (0.5 M KOH). Subsequently, the stripped membrane was hybridized, as per Materials and Methods, to the 49-mer specific to clone pKS36. The process of probe removal was repeated, and the membrane was finally hybridized, under high stringency, to the nick-translated 1.8 Kb Kpn1 satellite DNA of clone pBK1.8[20]. The complete removal of the probes by KOH treatment was assessed by autoradiography of the stripped membrane. Figure 15 presents the autoradiograms of these hybridization experiments, along with the corresponding laser densitometer tracings. In all cases, no hybridization signal to wc1 and wc2 was detected (data not shown). Therefore no significant amount of the satellite DNA (or 49 mer region) appears to be present on human chromosomes 6, 8, 11, and X. Neither of the EcoR1 or Kpn1 probes hybridized to the control hamster DNA (data not shown). Because of the kinship relationship displayed by the EcoR1 and Kpn1 satellite DNA, their hybridization profiles to the different subsets of human chromosomes appeared virtually identical.

TABLE 3

Rodent-Human Hybrids	Human Chromosomes content
WC1	6, 8, 11, X
WC2	X
WC3	1, 3, 4, 5, 8, 12. 13. 14. 16, 20, 21, Y
WC4	1, 2, 5, 7, 8, 12, 13, 14, 15, 17, 18, 19, 21, 22, X
WC5	1, 3, 4, 5, 6, 7, 8, 12, 14, 15, 16, 17, 19, 21, 22, X
WC6	3, 4, 5, 6, 8, 10, 11, 12, 13, 14, 15, 19, 20, 21, 22. Y
WC7	3, 4, 8, 9, 10, 15, 17, 19, 20, 22, X, Y
WC8	6, 12, 13
WC9	3, 4, 5, 6, 9, 11, 14, 17, 22
WC10	2, 3, 6, 7, 8, 11, 12, 13, 14, 15, 17, 20, 21, X, Y

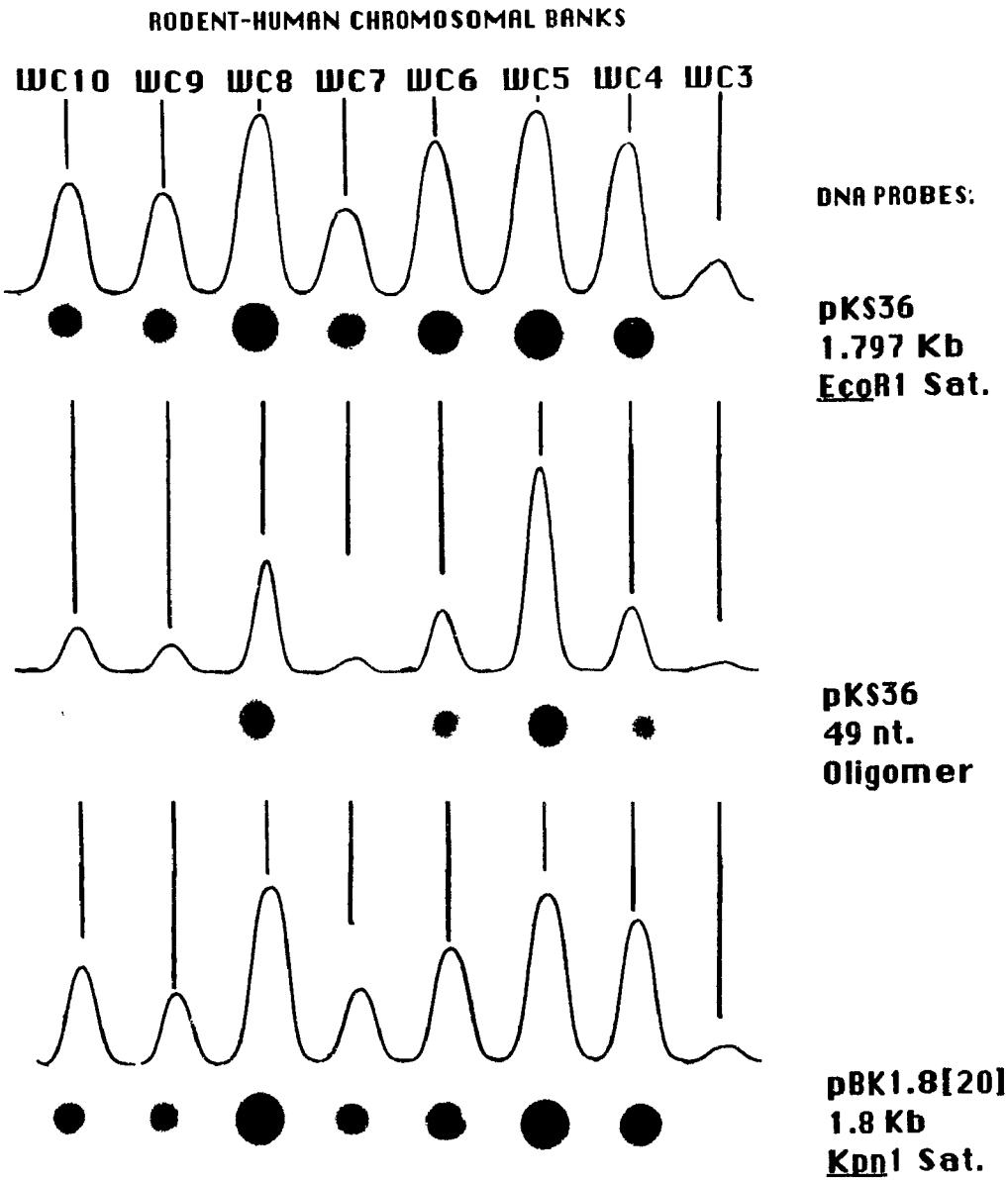
Figure 15: Analysis of a rodent-human hybrid cell panel. Dot blot autoradiogram and densitometer profiles of hamster-human chromosomal banks (lanes wc10 to wc3) hybridized, as per Chapter 2, with: pKS36 1,797 bp EcoR1 DNA, pKS36 49-mer non-satellite sequence, and pBK1.8[20] 1.8 Kb Kpn1 satellite DNAs. The peaks correspond to the hybridization intensity recorded for each dot blot using an LKB laser densitometer scanner. The distribution of satellite DNAs was analysed according to Waye et al. (1988).

For each of the human chromosomes the degrees of discordance (D) and concordance (C) were determined, and the percent discordance (D / D+C) calculated.

D = numbers of positive hybrids in which the chromosome is absent + numbers of negative hybrids in which the chromosome is present.

C = numbers of positive hybrids in which the chromosome is present + numbers of negative hybrids in which the chromosome is absent.

Figure 15


RODENT-HUMAN CHROMOSOMAL BANKS
WC10 WC9 WC8 WC7 WC6 WC5 WC4 WC3
DNA PROBES:
pKS36
1.797 Kb
EcoR1 Sat.
pKS36
49 nt.
Oligomer
pBK1.8[20]
1.8 Kb
Kpn1 Sat.

The data obtained from these experiments should be considered as preliminary as the degree of discordance is found to be elevated. Thus, from these preliminary results, sequences related to both pKS36 and pBK1.8[20] were found to reside mainly on chromosomes 7, 12, 14, 15, 16, and 22. The absence of significant hybridization of the 49 mer to wc1 and wc2 hybrid DNA also suggest that the 49 mer sequences are not present in significant quantities in hamster DNA.

III- CLONING OF RELATED HUMAN SATELLITE DNAs

HeLa EcoR1 DNA fragments, 1.8 Kb in length, were excised and electroeluted from an agarose gel, and cloned into the EcoR1 site of plasmid pUC119. The ligation mixtures were transformed into E. coli strain NM522, and the cells were subsequently plated on LBamp X-Gal/IPTG plates as per Materials and Methods. Colonies containing recombinant plasmids were regrown on master grids (Miller, 1972) and the plasmids were isolated via the method of Barnes (1977). After agarose gel electrophoresis, the recombinant plasmids were screened for the presence of satellite DNA on duplicate Southern blots, under high stringency (5 x SSC, 50 % formamide), using the nick-translated 1.797 Kb EcoR1 fragment of pKS36, or the 1.8 Kb Kpn1 fragment of pBK1.8[20]. These results are summarized in table 4, columns 1 and 2 (column 3 compiles the results from figure 16, see below). Colonies containing satellite DNA sequences were then selected and their plasmid DNA hydrolyzed with EcoR1.

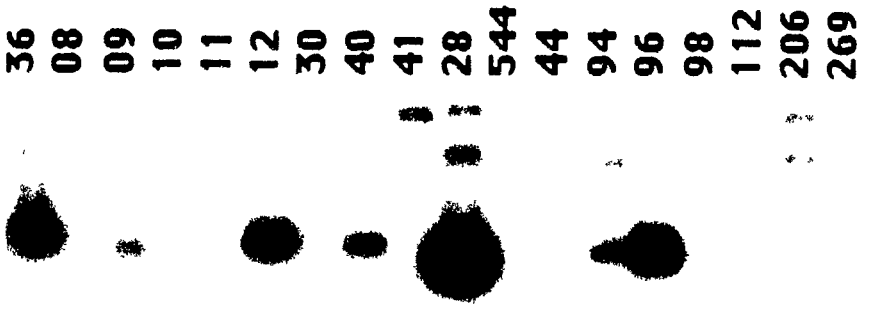
TABLE 4

Clones	Hybridizations results* pKS36	pBK1.8[20]	49 mer
pKS36	+++	+ +	+++
pKS08	+	+	–
pKS09	+	+	+
pKS10	+	+	–
pKS11	+	+	–
pKS12	+++	+	+ +
pKS30	+	+	–
pKS40	+ +	+	+ +
pKS41	+	+	+
pKS28	+++	+	+++
pKS544	–	–	–
pKS44	+	+	–
pKS94	+ +	+	+
pKS96	+++	+	+++
pKS98	+	+ +	–
pKS112	+	+	–
pKS206	+	+	+
pKS269	+/–	+	–

* The numbers of + (or –) signs reflect the intensity of the hybridization signals observed on the autoradiograms.

Figure 16: Southern blot/hybridization analysis of 1.8 Kb cloned EcoR1 satellite DNA fragments. Autoradiogram of a Southern blot containing EcoR1-cleaved satellite DNA clones hybridized with the kinased 49-mer specific for pKS36. pKS41, 44, and 206 are uncut DNA.

Figure 16



With the exception of pKS36 (diluted 1:10 prior electrophoresis), equal quantities of the EcoR1-cleaved DNAs were blotted and hybridized to the kinased 49-mer specific to pKS36 (figure 16). The hybridization results are presented in table 4 column 3. Of the 18 clones presented, clone pKS544 is a negative control containing a human 1.8 Kb EcoR1 DNA fragment that does not appear to contain any satellite DNA consensus sequences, and clones pKS41, 44, and 206 are uncut DNAs. The three clones displaying the strongest signal of hybridization to the 49-mer (pKS12, 28 and 96), gave intense hybridization signals to pKS36, as indicated in the table by plus signs (+++). Among the remaining satellite DNA clones, 5 clones showed partial homology (pKS09, 40, 41, 94, and 206) to the 49-mer, whereas clones pKS08, 10, 11, 30, 44, 98, 112 and 269, did not show any significant hybridization signal to the 49-mer. Note that pKS269 appears more closely related to pBK1.8[20] than to pKS36, and that it does not display cross-hybridization to the 49-mer. Thus, it appears that there is a good correlation between the presence or absence of cross-hybridization to the 49-mer and the degree of relatedness of the EcoR1 satellite DNA fragments of the diverse satellite DNA clones to pKS36.

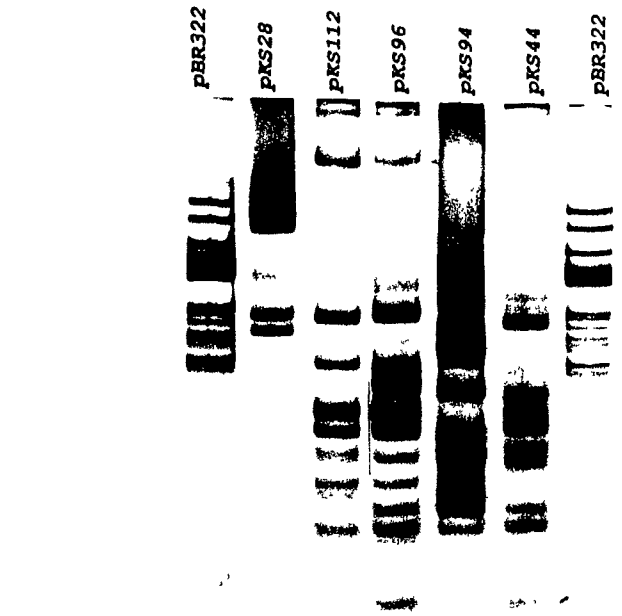
IV- MICROHETEROGENEITY WITHIN FAMILY MEMBERS

Nine of the satellite DNA clones studied in the previous section were randomly picked, and their DNA's purified on CsCl-EtBr equilibrium density gradients. These plasmids were hydrolyzed with EcoR1 (or Kpn1 in the case of pBK1.8[20]), and

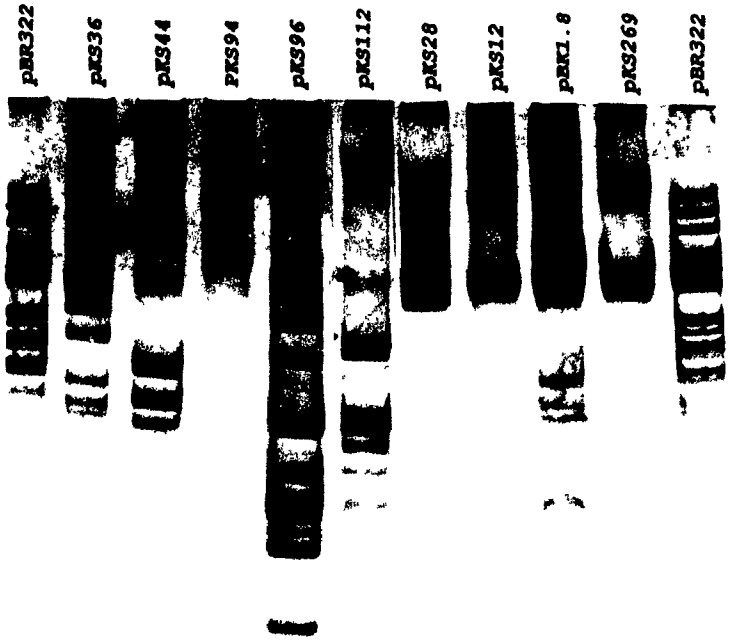
the inserts isolated by electroelution from a 1 % agarose gel. The purified DNA fragments were then hydrolyzed, under optimal conditions, with Hinf1 (at 37°C) and Taq1 (at 65°C), and the resulting fragments separated on 7 % vertical polyacrylamide slab gels. The hydrolyzed DNA fragments were visualized as indicated in Materials and Methods with a combination of Coomassie brilliant blue and silver staining. On the top half of figure 17 are displayed the Hinf1 restriction patterns of the EcoR1 satellite DNA clones, and the bottom part of the figure contains the Taq1 patterns of all the clones studied. Although some of the lanes appear to contain partially digested DNA fragments, the diverse restriction patterns demonstrated the existence of a high degree of polymorphism among the cloned fragments. This type of analysis allowed the classification of the related clones into groups as characterized by their restriction site distribution. Clones containing numerous Taq1 (and Hinf1, when tested) included pKS36, pBK1.8[20], pKS44 , pKS96, and pKS112. Clones containing few (less than five) to no sites for Taq1 included pKS12, pKS28, pKS94 and pKS269. Of the cloned tested with both enzymes, only pKS28 displayed rare to no restriction sites for either one of the enzymes. Although the hybridization experiments presented in figure 16 indicated that clones pKS12, pKS28 and pKS96 were very homologous to pKS36 and displayed strong cross-hybridization to the 49-mer, the restriction fragment analysis indicates that these clones, having distinct Taq1 restriction patterns, are divergent members of the sub-family of satellite DNAs containing the 49-mer region.

Figure 17: *Taq*1-*Hin*f1 polymorphism of satellite DNA clones. Coomassie blue and silver stain coloration of 7 % polyacrylamide gels containing *Eco*R1 cloned satellite DNA hydrolyzed, as indicated, with *Hin*f1 (top panel) and *Taq*1 (bottom panel). Molecular marker: pBR322 hydrolyzed with *Hpa*II.

Figure 17



Hinf1



Taq1

Table 5 presents some of the characteristics of the satellite DNA's analyzed in the next section.

U- FAMILY STUDIES.

U-1 Partial DNA sequencing of the satellite clones.

The Kpn1 satellite DNA insert of clone pBK1.8[20] was cloned into the Kpn1 site of pUC119, and the resulting clone was named pKpn1. The sequences at the extremities of the Kpn1 element were determined using the dideoxy chain terminator sequencing method (Sanger et al., 1977). The partial nucleotide sequences of three EcoR1 satellite DNA's isolated by our group (pKS44, 94, and 96) were also determined and a listing of these is shown in figure 18. A mismatch analysis was conducted, for each of the sequences generated, in which the number of base substitutions from the original consensus unit 5' TTCCA 3' were recorded (table 6). The number of base pair changes did not appear to occur randomly, as depicted by the presence of numerous C residues that were substituted. Clones pKS44 and pKpn1 are characterized by a high degree of substitution of the second deoxycytidine, whereas the other bases were changed at the same apparent frequency.

TABLE 5

Clone	Length in bp	Characteristics	Ref
pKS36	1797	EcoR1 monomer	1
pKS44	170	49-mer negative	2
pKS94	165	49-mer negative	2
pKS96	185	49-mer positive	2
pKpn1	263	Kpn1 monomer	2, 7
pHY10	3564	Y-specific sat. DNA	4
HumpPD9	168	Satellite III DNA	5
HumpPD17	244	Satellite II-III DNA	5
HumpPD18	166	Satellite II-III DNA	5
Humsat3R	332	Satellite III	6
HumsatII	250	Satellite II	3

1. Sol et al. (1986)
2. This work
3. Hollis and Hindley (1988)
4. Nakahori et al. (1986)
5. Deininger et al. (1981)
6. Fowler et al. (1987)
7. Shafit-Zagardo et al. (1982)

Figure 18: Partial sequences of four satellite DNAs. These sequences were obtained by the chain termination sequencing technique as indicated in Chapter 2.

Figure 18

pKS44

A TTTCA TTCGA TTCCA
TT AA TTCGA GTCAA TTCCA
TTCCA TTCTA TTCAA TGCCA
TTCCA TTCAA TTCTA TTCCA
TTCGA CACCA TTCCT TTCCA
TTCCG TTCCA TCCGA TTCTA
TTCCA TTGTA TTCCA TTCAA
TTCCA TTCCA TTGCA TTCCA
TTCGT TTCCA TTCCA

pKpn1

A TTCCA TGCCA TTCCA
TCT TGCCA TTCCA TTCCA
GTCCA TTCCA TTGCA T CGA
TTCCA TTCCA CTCCA TTCCA
ATTTA TTCCA TTCAA CTCCA
TTCCA ATCTA TTAC TTTCA
TTCCA TTCAA TTCCA TTCCA
TTTGA TTCCA TTCCA CTCTG
T CA TTCCA TTTGA ATCCA
CTCCA TTCCA TTCCA TTCCA
TTCCA TTCAG TTAAA TTCCA
TTCTG TTCCA TTCCA TTCCG
TTCCA TTCAA TTCCA TTCCA
TTCGG TACGA TGC

PKS94

GA TTCAA TCCA TTCAA
TTCCA CTGG TT A ATCCT
TTCCA TTCCA TTCCA ATCGA
TTCCC TTCCA TTCAA TTCCA
CTCGA TTCAA T CAA TTCCA
TTCTA TTCCA TTCTG TTCG
TTCCA CTCCA TTCCA TTGCA
TTCCA TACCA TTCCA TTCCA
CTCGG GTTGA TTCC

PKS96

TTTGA TACCA TCAG TTCGA
TT CA TTTG TTCCA GTCTA
TTCCT TTCCA GTTCA TTCCA
TTCCA TTTCA CTCCA TTCGA
TTTCA TTCCA CTCGA TT CA
CTCCG TTCCA TTTCA CTGCA
TTCCA TTCTA TTCAA TTCTA
TTGCA TTCCA TTCCA TTCCA
TTTGA TTACA TTCCA TGCT
T A TTCCA TTC

TABLE 6

Clones	Substitution[a]				
	T	T	C	C	A
pKS44	2	3	4	14	3
pKS94	9	1	4	15	6
pKS96	6	4	11	15	5
pKPN1	7	8	7	16	7
pKS36	88	15	67	83	46
HumpPD9	9	0	4	5	4
HumpPD17	7	1	10	20	2
HumpPD18	4	1	7	11	4
Humsat3R	2	11	16	18	14
pHY10	124	164	165	231	129

[a] Substitution: only complete pentamers were analyzed. Deletions (or insertions) were not included.

U-2 Sequence analysis

The satellite DNA sequences, reported in table 5, were compared using programs NUCALN and OVRLAP provided by the University of Montreal through the Clinical Research Institute of Montreal (Montreal, Que.).

None of the published sequences contained a sequence related to the 49-mer present in pKS36 (Sol et al., 1986). Although no complete identity was found between the satellite DNA members that were analyzed, Humsat05 (Prosser et al., 1986) appeared 86 % homologous to clone pHY10 (Nakahori et al., 1986) whereas HumpPD9 (pPD9, Deininger et al., 1981) displayed as much as 83 % homology to Humsat3R (Fowler et al., 1987).

If mutations occur at random within the five bases of the consensus sequence 5' TTCCA 3', the degree of base substitution should be equivalent at the five position. However, the mismatch analysis performed on the published sequences (table 6) revealed the presence, in the pentameric consensus repeats (5' TTCCA 3'), of distinct hotspots for base pair substitution. All clones tested showed hypervariability in the second C of the consensus pentamer. The sole exception was HumpPD9, in which it was the first T that was the most frequently substituted. An additional base was found frequently substituted in some clones: the first T in clones pKS36 and 94, and the first C in clones pKS96 and HumpPD17.

Along with mutational hotspots, the mismatch analysis of these diverse satellite DNA sequences revealed the presence of bases that were more conserved within the pentameric repeats.

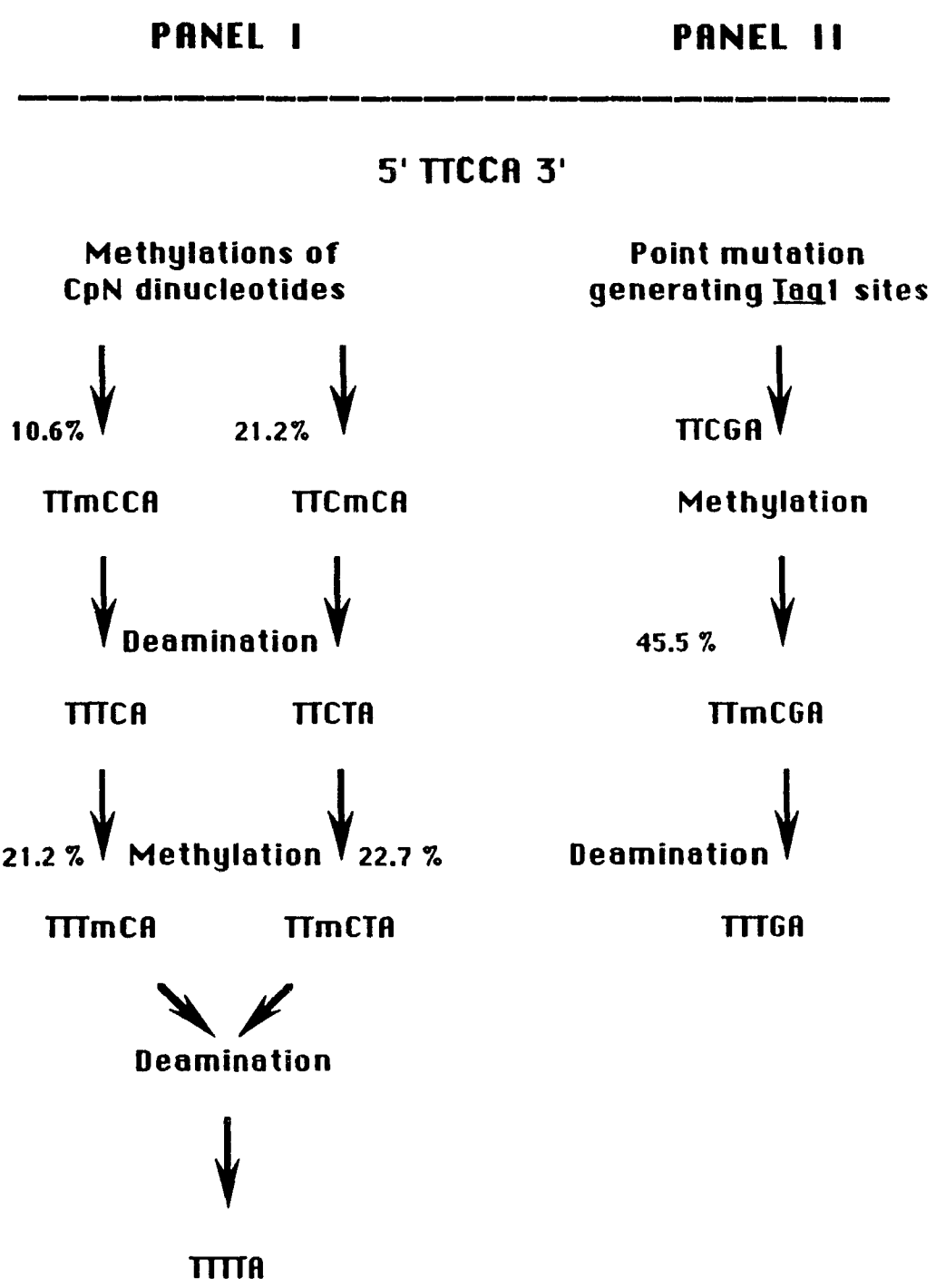
Humsat3R displayed fewer substitutions of the first thymidine, whereas pKS36, 94, HumpPD9, 17 and 18 showed an increase in the conservation of the second thymidine when compare to the other bases. Clone HumpPD17, and to a lesser extent pKS36, displayed fewer substitution of the adenine residues.

Since methylated cytosines have been shown to be one source of point mutations in mammalian cells (Bird, 1980), we examined the satellite DNA sequences for the presence of pentamers that could have arisen from the original consensus unit by spontaneous deaminations at methyl-deoxycytidine sites (mC->T) within CpC and CpA dinucleotides. This potential decay mechanism is presented in figure 19, panel I. Moreover, one of the differential characteristics between satellite II and III DNA is the presence of many Taq1 (TCGA) sites in the former and few to none in the latter (Sol et al., 1986; Nakahori et al., 1986; Hollis and Hindley, 1988). The Taq1 restriction sites, generated by C to G transversions within the 5' TTCCA 3' tandem units, are potential methylation sites (mCpG) and thus may degenerate by subsequent deamination of the methylated base (figure 19, panel II).

The frequencies of the pentamers, contained in the analyzed sequences, that may have arisen from one of the two decay mechanisms are presented in table 7. As deletions and insertions of bases that occurred during the propagation of the satellite DNA's appeared to be random, only intact pentamers were considered in this analysis.

Figure 19: Proposed two-stage decay mechanism for base substitution in satellite DNA. Mechanisms for the degeneration of the consensus pentamer 5' TTCCA 3' by methylation of cytosine residues and point mutations are proposed. The underlined pentamers represent the outcomes of the decay mechanisms that were searched in the sequences analyzed in table 8. mC represents methylated cytosine. N is any nucleotide (G, A, T or C).

Figure 19



N.B.: The potential methylation levels of mCpN dinucleotides are expressed in accordance to Woodcock et al. (1987) as percent of total genomic mC.

TABLE 7

	PENTAMER					
Clones	TTCCA	TTCGA	TTTGA	TTTCA	TTCTA	TTTTA
pKS36	129	18	9	13	10	1
	36%	5%	2.5%	3.6%	2.8%	0.3%
pHY10	229	58	20	22	27	6
	32%	8%	2.8%	3.1%	3.8%	0.8%
HumpPD9	22	0	0	0	0	0
	65%	(0)	(0)	(0)	(0)	(0)
HumpPD17	12	8	3	2	1	0
	24.6%	16.4%	6%	4%	2%	(0)
HumpPD18	14	2	1	0	2	0
	42%	6%	3%	(0)	6%	(0)
Humsat3R	36	2	0	1	0	0
	54%	3%	(0)	1.5%	(0)	(0)
HumsatII	12	10	3	0	0	0
	24%	20%	10%	(0)	(0)	(0)
pKS44	14	3	0	1	3	0
	41%	8.8%	(0)	3%	8.8%	(0)
pKS94	13	0	0	0	1	0
	39%	(0)	(0)	(0)	3%	(0)
pKS96	12	2	2	3	2	0
	32%	5.3%	5.3%	8%	5.3%	(0)
pKpn1	26	0	2	1	0	0
	49.5%	(0)	3.8%	1.9%	(0)	(0)

NB: The frequencies of a given pentamer are calculated in reference to the sequenced region of the satellite DNA.

DISCUSSION

The group of simple sequences found in the family of satellite II and III DNA consists of heterogeneous sequences which are likely to have originated from the tandem duplication of the pentameric repeat 5' TTCCA 3' (Beauchamp et al., 1979). During the process of evolution, this non-coding DNA has undergone mutations that may have resulted in the degeneration of the original unit of repetition, and in the segregation (perhaps to different chromosomal environments) of large units of repetition in related, though distinct, sub-families. The heterogeneity, yet overall sequence relatedness, of the members of a satellite DNA family has made it unusually difficult to study the precise location and organization of specific subsets of satellite DNA's within the human genome (Corneo et al., 1971; Cooke and Hindley, 1979).

During this study, a 49 nucleotide sequence, specific to pKS36, has made it possible to distinguish members of two related sub-families of satellite II DNAs: the sub-family of satellite II DNA containing the 49 mer (e. g., pKS36), and the sub-family of satellite II DNA devoid of it (e. g., pBK1.8[20]). Furthermore, this unique region enabled us to differentiate closely related members of the satellite III DNA family. The satellite II DNA family analyzed represents 2% of the genome of HeLa cells and its members are found mainly clustered on

chromosomes 7, 12, 14, 15, 16 and 22. While we did not see any significant hybridization to chromosomes 9 and Y, the major region of satellite II DNA has been previously mapped, by in situ hybridization and TaqI restriction endonuclease banding analysis, to chromosomes 1, 9, 15, 16 and Y (Gosden et al., 1975; Tagarro et al., 1991). The EcoRI sub-family of satellite II DNA, represented by clone pKS36, appeared to represent no more than 1 % of the human genome.

By probing a genomic library of 1.8 Kb EcoRI satellite DNA clones, produced during this work, with either pBK1.8 [20], pKS36, or the 49-mer, the degree of relatedness of the different members of the satellite II and III families was determined. This analysis was extended by examining the TaqI and HinfI restriction site polymorphisms of selected family members. The distribution of TaqI (TCGA) and HinfI (GANTC) restriction sites in the repeated elements is a diagnostic of satellite sequences, as these sites are generated by a single C to G transversion within the 5' TTCCA 3' consensus motif. These studies allowed the discrimination between apparently identical sequences (e. g. pKS36, pKS12, pKS28 and pKS96) and the classification of the cloned sequences into two main classes: a) satellite DNA containing very few TaqI (or/and HinfI) sites (e. g. pKS12, pKS28 pKS94 and pKS269); b) the class of sequences sensitive to TaqI (e. g. pKS36, pBK1.8[20], pKS44, pKS96 and pKS112). Satellite III DNA has been previously defined as the group of simple sequences containing very few to no TaqI sites (Frommer et al., 1982), whereas satellite II DNA was defined as the group of

sequences extensively digested with TaqI (Frommer et al., 1982; Prosser et al., 1986; Hollis and Hindley, 1988). Using this nomenclature, clones pKS12, pKS28, pKS269, and pKS94 belong to the family of satellite III DNA, whereas pKS36, pKS44, pKS96, pKS112 , and pBK1.8[20] belong to the family of satellite II DNA. The TaqI and HinfI polymorphisms, observed between the cloned satellite DNAs, correlate with the TaqI genomic polymorphisms of satellite-like sequences that were reported by Fowler et al., (1988).

It is interesting to note that some members of the satellite II family (such as pKS36 and 96) and members of the satellite III family (such as pKS12 and 28) have in common the 49 mer region (as depicted in figure 16). It is most likely that these satellite DNA members have evolved from a common ancestor containing the unique region, and have thereafter diverged as members of two distinct families, probably by segregating to a different genomic environment.

We have examined, at the nucleotide level, several satellite DNA's in an attempt to clarify the source of sequence polymorphisms observed within related members of the family of highly repeated human satellite II and III DNA. With the exception of pKS36 (Sol et al., 1986), clone pHY10 is the only other long satellite DNA completely sequenced (Nakahori et al., 1986). Clone pHY10 contains a 3.6 Kb EcoR1 dimeric satellite DNA that is located on the human Y chromosome. This clone has been reported as a satellite III DNA though, it has been shown to be enriched in TaqI restriction sites (Fowler et al., 1988). To date,

most of the analyzed sequences reveal hypermutability in the dinucleotide CpC present in the consensus repeat 5' TTCCA 3' (these are the two residues least conserved within the pentameric repeats). Such mutational hotspots in non-genic DNA should be rare, as non phenotypic DNA is hypothesized to degenerate via random mutations.

However, it is possible that these hypervariable bases may result from DNA methylation. CpN dinucleotides (N being either A, T, C, or G) have been found to be sites for methylation in most mammalian genomes (Grafstrom et al., 1985, Nyce et al., 1986). Woodcock et al. (1987) reported that, in human spleen tissue samples, only 55.5 % of the total mC was present in mCpG dinucleotides. The methylation levels of the other 3 CpN dinucleotides were 22.7 % (CpT), 21.2 % (CpA) and 10.6 % (CpC). Though it is still unclear whether methylation levels of satellite DNA are higher or lower than the rest of the genome, it is possible that the high rate of spontaneous deamination (mC->T) observed at the mCpN sites in the mammalian genomes (Bird, 1980) may thus be a cause for the accumulation of point mutations within these satellite CpN-rich DNA. Methylation of either of the two cytosine residues within TTCCA pentamers could lead to the generation of TTTCA and TTCTA pentamers (figure 19). Subsequent methylation and deamination of the TpmC and mCpT dinucleotides would generate TTTTA pentamers. Methylation within TaqI sites T(TCGA), and subsequent deamination of the methyl deoxycytidines, would convert the sequence to TTTGA. Upon examination of published satellite II

and III DNA sequences for the presence of pentamers resulting from a methylation/deamination decay mechanism (figure 19), it was observed that in clones pKS44, 36, 96, pHY10, HumpPD17 and HumpPD18, such pentamers (TTTCA and TTCTA) represent from 6 to 13 % of the total sequence, more than would be predicted via random mutations.

Furthermore, the hypermutability of the mCpN dinucleotide may be largely responsible for the TaqI/HinfI polymorphisms observed between closely related satellite DNA's in figure 17. As observed in table 7, clone pKS36, 96, pHY10, HumsatII and HumpPD17 contain a considerable fraction of TTCGA (from 5 to 20 %) pentamers, which, followed by deamination of the methyl deoxycytidines, would convert the sequence to TTTGA and result in the loss of the TaqI site (this scenario holds true for HinfI sites). Since CpG dinucleotides have been shown to contain a high proportion of methyl deoxycytidines (Nyce et al., 1986; Gruenbaum et al., 1982), it can be postulated that an appreciable number of TaqI sites resident on satellite DNA contains mCpG. In agreement with the report made by Fowler and his colleagues (Fowler et al., 1988), we observed that, in these clones, an increase in the frequency of TTCGA pentamers was generally accompanied by an increase in TTTGA pentamers (table 7, clones pKS36, pHY10, humpPD17, and humpPD18). All other pentamers that may have arisen from random mutations were found underrepresented in most of the analyzed sequences.

This analysis thus suggests a multi-step decay mechanism in which, as illustrated in figure 19, both C to G mutations and

CpN dinucleotide methylation (followed by subsequent deoxycytidine spontaneous deaminations) have a role to play in the generation of sequence micro-heterogeneity within members of the families of satellite II and III DNAs.

Chromosomes are the stage of both stability (to insure accurate distribution in daughter cells) and fluidity (to insure evolutive flexibility) of the heriditary material. The apparent "stability" of the families (or groups) of satellite DNAs may result from low level of exchange between non-homologous chromosomes. The satellite DNA segments on each chromosome pairs would be isolated and thus free to diverge from one another and create new collection (families) of size fragments. The relative rate of the sequence homogenization (or diversity) of members of a satellite DNA family may thus depend on their location within defined regions of the human genome. The particular molecular environment of the highly repeated sequences may also dictate, via processes such as gene conversion, the observed microheterogeneity.

The variations observed between the different satellite DNAs that were analyzed in this study may thus result from differences in the micro- and macro-organization of satellite sequences, as well as their unique chromosomal environment, within the human genome.

The next chapter of this thesis is devoted to the study of the genomic organization of the members of the satellite II and III families, that are closely related to pKS36.

Chapter V

GENOMIC ORGANIZATION OF THE SATELLITE DNA FAMILY

There is a theory which states that if ever anyone discovers exactly what satellite DNA is for and why it is here, it will instantly disappear and be replaced by something even more bizarre and inexplicable.

There is another which states that this has already happened...

Katia Sol (1991)
Inspired from Doug Adams, (1989)

INTRODUCTION

Satellite II and III DNA consist of families of simple repeated sequences whose members, though evolutionary related, appear to be highly polymorphic in sequence. As this polymorphism may be linked to specific genomic environments, we have chosen to investigate, by southern hybridization, the genomic distribution of these simple repeated DNA sequences both in the short (using conventional gel electrophoresis) and long (using field inversion gel electrophoresis) range.

The ability of related, but non-identical, sequences to recognize each other during hybridization experiments can be controlled by the stringency of the conditions imposed for hybrid formation. By varying the concentration of formamide in the hybridization reaction, we were able to distinguish between the genomic organization of the family of related satellite II and III DNA sequences, and the specific organization of the sub-group of pKS36-like satellite II DNA sequences. HeLa genomic DNA was extensively analyzed, via southern hybridization, using a battery of restriction endonucleases, and the organization of pKS36-like (or closely related) satellite II sequences was investigated under high stringencies. Satellite II and III DNAs were found mainly organized in tandem arrays of 1.8 Kb long EcoR1 and Kpn1 domains.

In addition to the original cell line HeLa from which our cloned satellite was isolated, the human melanoma derived-cell line MeWo was analyzed. In MeWo, Holden et al., (1986) have identified a related satellite DNA that is amplified within homogeneously staining regions (HSRs) of chromosome 15. The analysis of these human cell lines, by southern hybridization using the pKS36 cloned satellite DNA as a probe, revealed distinct patterns in the organization of satellite DNA.

Using the technique of field inversion gel electrophoresis (FIGE), the long range organization of satellite DNA was explored. This technique makes possible the separation and resolution of DNA molecules (up to millions of bp in length) that would otherwise move, in conventional unidirectional electrophoresis, at limiting mobility (Schwartz and Cantor, 1984). Molecules that are smaller than 50 kilobase (Kb) migrate at a rate determined roughly by their size. Those greater than 50 Kb move at the same rate, probably because they assume a conformation which allows them to migrate in a wormlike motion through the gel's pores (Lumpkin and Zimm, 1982). During field inversion gel electrophoresis (FIGE), the molecules are forced to change direction when the electric field is reversed, and to assume new configurations before they can move again. The time it takes to change configuration depends on the size of the molecules and therefore allows separation of large DNA molecules. Although it is not clear to date if large blocs of satellite DNA are interspersed with unrelated sequences, we were able to identify, via FIGE, discrete blocks

(ranging from 150 Kb to 500 Kb in size) containing satellite II DNA sequences closely related to pKS36.

RESULTS

I-SHORT-RANGE GENOMIC ORGANIZATION OF PKS36-RELATED FAMILY MEMBERS.

The sizes of the DNA fragments, reported in this section, were estimated using the migration rates of known molecular markers (phage Lambda hydrolyzed with either HindIII or EcoR1).

I- 1 Short range organization in HeLa cells

In order to define the genomic organization of the satellite DNA related to the cloned 1.8 Kb EcoR1 satellite DNA of pKS36, a series of Southern blot analysis of restriction enzyme hydrolyzed HeLa DNA was probed, under increasing stringency conditions, with the insert of pKS36 (figure 20, panels 1, 2 and 4). The stringency was controlled, from low to high stringency, by the percent of formamide added to the hybridization mix. Low stringency was defined by the presence of 15 % formamide (figure 20, panels 1 and 2), whereas high stringency was obtained in the presence of 50 % formamide (figure 20, panel 4). Panel 2, a shorter exposure of panel 1, is presented to facilitate the identification of the EcoR1 satellite DNA species positioned around 1.8 Kb.

Under low and high stringency (lanes K, panels 1 and 4), two intense bands located at approximately 1.8 Kb (monomers) and 3.6 Kb (dimers) were detected, along with faint hybrid bands

(indicated by arrowheads) migrating as trimers (5.4 Kb), tetramers (7.2 Kb) and pentamers (9 Kb), thus suggesting a regular tandem organization for Kpn1 satellite DNAs. No other Kpn1 size fragments appeared after over-exposure of the blot (figure 20, panel 1, lane K).

In contrast, the pattern of hybridization to the EcoR1-cleaved HeLa DNAs appeared more complex. Under low stringency (Panel 2, lane E) prominent EcoR1 hybrid bands are formed around 1.8 Kb and 2.35 Kb (indicated by a dot). The other detectable bands (observed in panel 1, lane E), formed under these permissive conditions were approximately 1.5 Kb, 1.65 Kb, 1.95 Kb, 2.75 Kb, 2.95 Kb, 3.05 Kb, 3.6 kb, 5.4 Kb, 7.2 Kb, and 9 Kb in length, along with many diverse sized fragments. Although the presence of multimeric forms (3.6 Kb, 5.4 Kb, 7.2 Kb and 9 Kb), indicative of a tandem organization, could have been generated by point mutations at the EcoR1 site between adjacent 1.8 Kb EcoR1 repeat units, the presence of the non-multimeric sized hybrid bands found in the EcoR1-cleaved DNA lanes suggest an interspersed organization for some of the genomic sequences related to the cloned satellite II DNA. Under high stringency conditions (panel 4, lane E), hybridization was found to occur predominantly, not with the expected 1.8 Kb EcoR1 DNA fragment but with a 1.95 Kb EcoR1 DNA fragment. In addition, major bands were observed at 1.65 Kb and 3.6 Kb. It thus appear that pKS36-like satellite II sequences define a sub-class of the satellite II family whose members are likely to be clustered on 1.65 Kb, 1.95 Kb and 3.6 Kb EcoR1 DNA fragments.

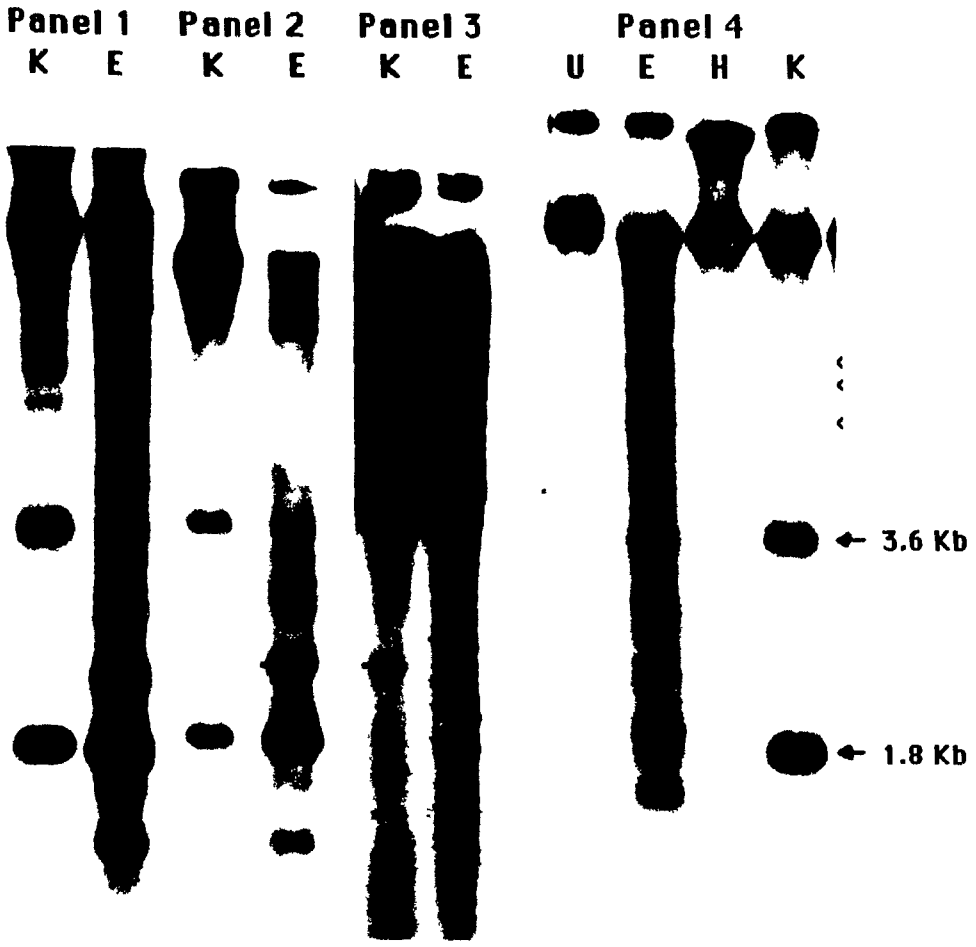
Hybridization occured, to a much lesser degree, with bands previously detected under low stringency conditions.

In addition, the distribution of satellite DNAs containing the 49 mer region specific to pKS36 was investigated (figure 20, panel 3). The 49 mer, hybridized under mild condition as indicated in Chapter II, annealed with EcoR1-cleaved HeLa DNA fragments in a pattern similar to that observed under low stringency for the complete EcoR1 element, but with a very high background (panel 3, lane E). Though no hybrids with the 49 mer could be detected around the positions of 1.8 Kb or 3.6 Kb characterizing the Kpn1 satellite DNAs, a band was detectable at approximately 2.35 Kb (panel 3, lane K).

Under the electrophoresis conditions used in these experiments to separate the hydrolyzed DNAs, no bands were detected with restriction endonuclease HindIII (figure 20, panel 4, lane H).

Figure 20: Southern hybridization under increasing stringency of HeLa genomic DNA. Genomic DNA was isolated as per Chapter II and hydrolyzed with restriction endonucleases Kpn1 (K), EcoR1 (E), or HindIII (H). U is uncut DNA. Panels 1 and 2 present the autoradiograms of southern hybridizations performed under low stringency (15 % formamide), whereas panel 4 presents the autoradiogram of southern hybridization performed under high stringency (50% formamide), using the 1.797 Kb EcoR1 DNA fragment of pKS36 as a probe. Note that panel 2 is a lighter exposure of the panel 1 autoradiogram. Panel 3 is the autoradiogram of the southern blot hybridization performed using the 49 mer as a probe. The positions of the 1.8 Kb and 3.6 Kb DNA bands are indicated. Arrowheads indicate the positions of the 1.8 Kb multimeric DNA fragments. Dots indicate the positions of the 2.35 Kb DNA fragments.

Figure 20



I- 2 Polymorphisms between cell lines

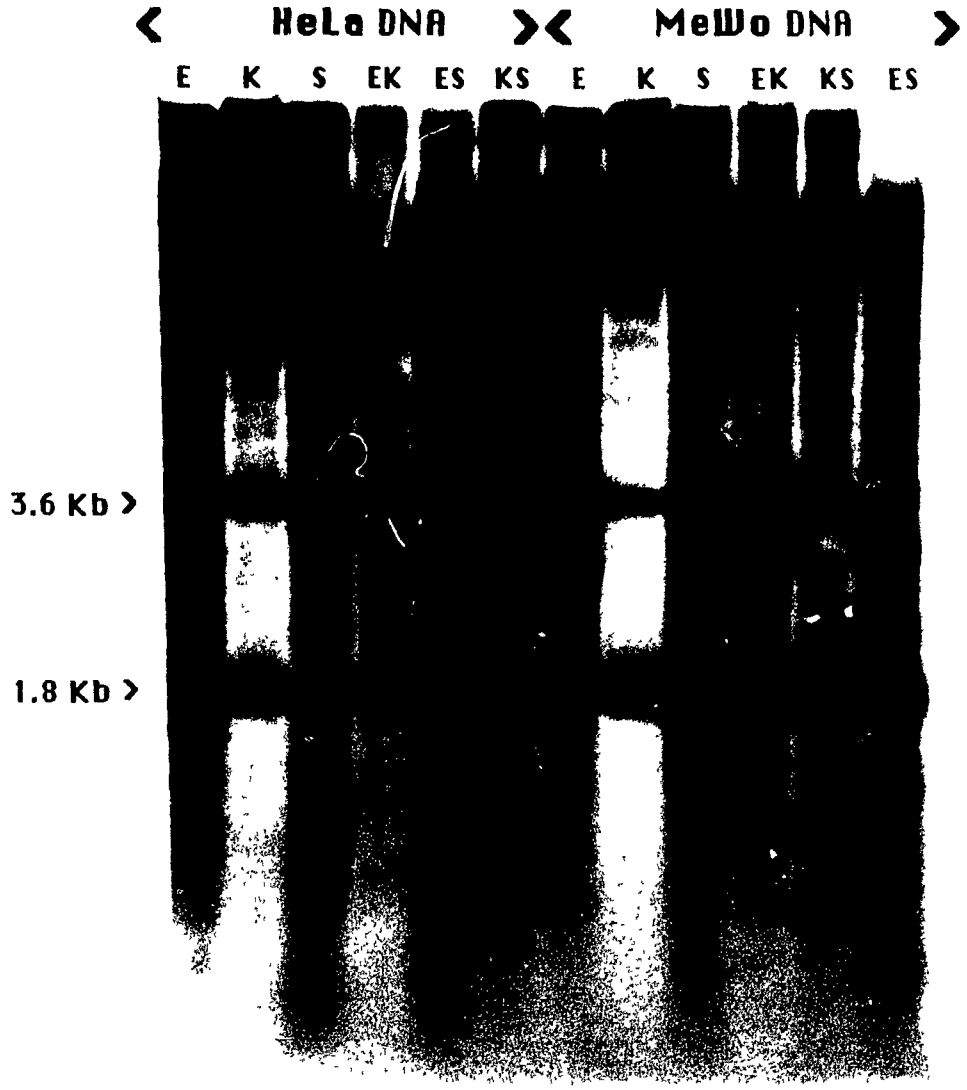
Figure 21 presents the autoradiogram of a southern blot of HeLa and MeWo DNAs hydrolyzed with EcoR1 (E), Kpn1 (K), Sau3A (S), EcoR1 plus Kpn1 (EK), EcoR1 plus Sau3A (ES), Kpn1 plus Sau3A (KS).

As seen in the above section, HeLa pKS36-like satellite II DNAs are exclusively organized in tandem arrays as Kpn1 fragments (1.8 Kb and 3.6 Kb repeat units), and their EcoR1 pattern reveals a complex distribution (figure 21, lanes K and E, respectively). In addition to tandem array distribution of 1.8 Kb and 3.6 Kb Sau3A DNA fragments, pKS36-like sequences are present in abundance on 3.25 Kb long Sau3A fragments (figure 21, lane S).

In MeWo cells, this class of satellite II DNA was found mainly in tandem arrays of 1.8 Kb long Kpn1 and Sau3A DNA fragments (figure 21, lane K and S, respectively). When MeWo genomic DNA was hydrolyzed with EcoR1 (figure 21, MeWo, lane E), the majority of the satellite II DNAs was found as 2.75 Kb, 2.95 Kb and 3.6 Kb long DNA fragments. Minor bands were detected around positions of 1.5 Kb, 1.65 Kb, 1.8 Kb, and 2.35 Kb. No EcoR1 hybrid band was detected around position 1.95 Kb, as seen in HeLa cells, but when MeWo DNA was hydrolyzed for longer incubation time (6 hours) with an excess of EcoR1, approximately 25 % of the 3.6 Kb EcoR1 DNA appeared as 1.8 Kb EcoR1 monomers (data not shown). Thus a fourth of the 3.6 Kb EcoR1 DNA fragments observed in figure 21 (MeWo, lane E) are the results of partial digestion of the DNA.

Figure 21: Southern blot hybridization of HeLa- and MeWo-hydrolyzed genomic DNAs. Total genomic DNAs were isolated from the genomes of two human cell lines HeLa and MeWo, and digested as indicated with EcoR1 (E), Kpn1 (K), Sau3A (S) or with a combination of two restriction endonucleases (EK, ES, and KS). The DNAs hydrolyzed with either EcoR1 or Kpn1 were phenol/chloroform extracted, ethanol precipitated and resuspended in the desired digestion buffer prior to the second restriction endonuclease hydrolysis. Southern hybridization were conducted under high stringency, with the pKS36 EcoR1 DNA insert as a probe. The positions of the 1.8 Kb and 3.6 Kb DNA fragments are indicated.

Figure 21



Although, the majority of the 1.8 Kb Kpn1 and 2.75 Kb EcoR1 DNA fragments do not seem affected by the process of double hydrolysis of HeLa DNA with EcoR1 and Kpn1 (lane EK), the major EcoR1 DNA fragments (1.65 Kb, 1.95 Kb and 3.6 Kb) virtually disappear, indicating that recognition sites for the enzyme Kpn1 are present on these EcoR1 fragments. In contrast, the hybridization profile of MeWo DNA hydrolyzed with EcoR1 and Kpn1 (lane EK) was virtually a combination of the simple digestion profiles (lanes E and K), indicating the absence of EcoR1 sites on Kpn1 DNA units..

The Kpn1-Sau3A double digestion of HeLa DNA suggested that these restriction sites are interspersed within the satellite II DNA sequences. The reduced hybridization of the satellite probe to the previously observed 1.8 Kb and 3.6 Kb Kpn1 and Sau3A DNA fragments, as well as the disappearance of the major 3.25 Kb Sau3A DNA fragment and presence of new junction fragments (1 Kb, 1.3 Kb 1.65 Kb and 2.55 Kb, approximately), indicated that the majority of the Kpn1 fragments contained at least one Sau3A restriction site and vice versa.

The same type of observations were made with the Kpn1-Sau3A (figure 21, lane KS) digestion pattern of MeWo DNA, though it seemed that either the 1.8 Kb Kpn1 (or Sau3A) satellite DNA was devoid of Sau3A (or Kpn1) sites, as seen by the predominant 1.8 Kb DNA fragment species resistant to the double digestion.

The DNA banding pattern of HeLa cells corresponding to the EcoR1 and Sau3A double hydrolysis virtually resemble that of the

superimposed simple digestions(figure 21, lanes ES, and E, S, respectively), thus suggesting that the EcoR1 and Sau3A units of amplification are not neighboring sequences but rather that they define two divergent regions of repetitive DNAs. The sole exception was the virtual disappearance of signal of pKS36 to the HeLa genomic 1.95 Kb EcoR1 DNA fragment after Sau3A hydrolysis (lane ES).

In contrast, when MeWo DNA was hydrolyzed with EcoR1 and Sau3A, the patterns of the simple digestions did not superimpose. Hybridization of the 1.797 Kb EcoR1 satellite probe occured mainly with fragments 1.45 Kb, 1.8 Kb, 2.35 Kb, 2.95 Kb, and 3.6 Kb in length, and, to reduced degrees, with DNA fragments 900 bp, 1.1 Kb, and 2.75 Kb in length (figure 21, lane KS).

Hydrolysis of HeLa genomic DNA with restriction endonuclease Cla1 (ATCGAT), which is present once in pKS36 and contains an internal Taq1 recognition site, showed an unresolved bulk of bands greater than 20 Kb. Endonucleases BamH1 and Hpa11 also gave unresolved smear of hybridization. However, 3.6 Kb long Msp1 satellite DNA and Rsa 1 (internal Kpn1 recognition site) discrete hybrid bands (approximately 850 bp, 1000 bp and 1.8 Kb in size) were detected (data not shown).

Figure 23, presented in the discussion section, illustrates the possible modes of organization of satellite DNA in human cells.

II LONG RANGE ORGANIZATION OF SATELLITE DNA

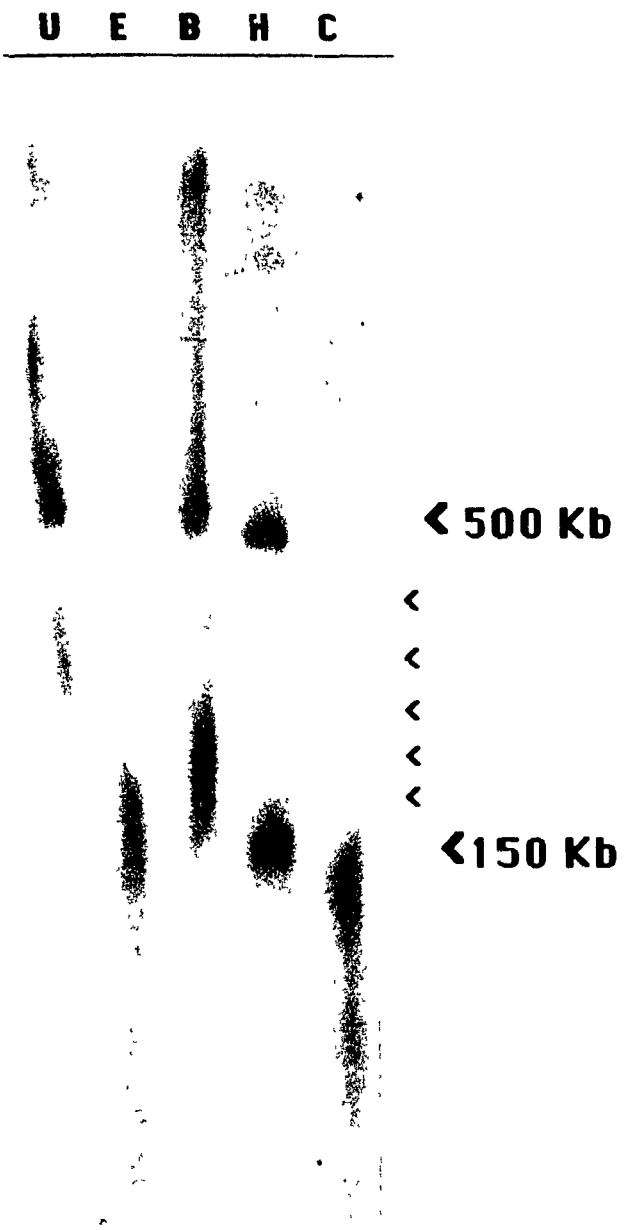
In order to study the higher order organization of satellite DNA, and determine whether or not a tandem organization of large blocks could be detected, field inversion gel electrophoresis (FIGE) of hydrolyzed HeLa DNA was conducted as per Chapter 2. Because the technique requires the manipulation of very high molecular weight DNA, embedding eukaryotic cells in agarose plugs has been used to prevent shearing of the DNA in solution during isolation and restriction enzyme digestion (Gardiner et al., 1986).

Agarose plugs containing lysed HeLa cells were hydrolyzed with EcoR1, BamH1, HindIII, and Cla1 (Figure 22, lanes E, B, H and C, respectively). After FIGE, the DNA was denatured, blotted to nylon membranes and hybridized under high stringency to the 1.797 Kb satellite DNA of clone pKS36. No clear banding pattern was resolved by EcoR1 BamH1 or Cla1 restriction analysis. Previous results showed that no banding pattern could be detected, in the 50 Kb and lower range, when HeLa DNA cleaved with HindIII was hybridized to pKS36 EcoR1 satellite DNA (figure 20, and Sol et al., 1986), suggesting that the satellite DNA family under study was in a genomic environment devoid of HindIII restriction sites. However, using FIGE, a discernable "ladder" of satellite DNA fragments (indicated by arrowheads in figure 22), "flanked" by two distinct blocs 150 to 500 Kb in size, was identified (lane 4) when human DNA was hydrolyzed with HindIII. The difference in intensity between these diverse HindIII

segments may reflect a polymorphic distribution in the number of satellite repeats that are present (see discussion).

Figure 22: Long range organization of satellite DNAs. HeLa. Cells embedded in agarose plugs were hydrolyzed, and the DNA fragments separated by field inversion gel electrophoresis, as per Chapter II. The autoradiogram displays the hybridization patterns, obtained under high stringency (50 % formamide), of HeLa cells hydrolyzed with EcoR1, BamH1, HindIII and Cla1 (lanes E, B, H, and C, respectively), using the nick-translated 1.797 Kb EcoR1 satellite II DNA as a probe. Lane U contains lysed HeLa cells whose DNAs were not subjected to restriction endonuclease hydrolysis, as a control for nuclease activity triggered by the lysis process. The molecular markers in the experiments were, as described in length in Chapter II, Lambda concatamers and yeast chromosomes.

Figure 22



DISCUSSION

Repetitive satellite II and III DNAs consist of families of sequences that are, though evolutionary related, highly polymorphic in sequence, as seen in Chapter IV. The polymorphism of satellite DNAs extend to the level of their organization into diverse domains, as depicted in the series of southern blot analysis presented in this chapter. The organizations of satellite II DNA in HeLa cells are proposed in figure 23, and discussed below. In addition to the cell line HeLa, we examined a second cell line (MelWo) derived from a human melanoma. MelWo was chosen because it exhibits on chromosome 15 homogeneously staining regions containing amplified copies of D15Z1, a satellite II DNA (Simmons et al., 1984; Holden et al., 1986).

Southern blot analysis, performed under low stringency, has revealed that, though the members of the satellite DNA II and III are organized as 1.8 Kb and 3.6 Kb EcoRI and KpnI tandem repeat units, a large fraction of its members are found interspersed as diverse sized EcoRI fragments (figure 23, panels A and B). Alternatively, these diverse sized fragments may indicate that there is considerable heterogeneity in EcoRI site distribution in tandemly repeated satellite elements (figure 23, panel A). In HeLa cells clone pKS36 appeared to be a minor member of the newly observed sub-family characterized by clusters of 1.65 Kb, 1.95 Kb and 3.6 Kb Eco RI satellite DNAs. If

the 1.65 Kb and 1.95 Kb units are organized in consecutive tandems (···1.65-1.95-[1.65-1.95]-1.65-···), a mutation at the EcoR1 site between two consecutive units (indicated within brackets) would result in the formation of the observed 3.6 Kb EcoR1 composite dimer (···1.65-[3.6]-1.65-···). Unfortunately, upon limited EcoR1 hydrolysis of HeLa DNA, the numerous diverse size partial EcoR1 fragments of other inter-spersed satellite II DNA family members, prevented verification of the tandem organization of 1.65 Kb and 1.95 Kb EcoR1 fragments.

Examination of the sequence of the cloned 1.8 Kb satellite DNA revealed the presence of a "near" EcoR1 site 5' GCATTC 3' located 150 bp from the 3' end of the element (double underlined in figure 11, Chapter III). As seen in Chapter IV, satellite DNA repeats are characterized by the hypervariability of cytosine residues. A single C->A point mutation within the 5' GCATTC 3' sequence would generate an EcoR1 site (G|AATTC) in the original 1.8 Kb monomeric unit and a new 1.65 Kb EcoR1 fragment would appear upon hydrolyzes. Moreover, the same mutation affecting one of the sub-repeat of a dimerized 1.8 Kb repeat would generate two new EcoR1 fragments 1.65 Kb and 1.95 Kb in size. Further amplification in tandem of these two fragments could generate the type of EcoR1 organization that is observed in HeLa cells. Furthermore, point mutations in other "near" EcoR1 sites that are scattered within the cloned element (underlined in figure 22, Chapter IV) may result in the "apparent" interspersed organization of EcoR1 satellite DNAs in human cells.

Using the "49 mer" as a probe it was observed that, under the permissive hybridization conditions utilized, this element hybridized in much the same way as the 1.797 Kb EcoR1 satellite element did under low stringency. Though the 49 mer did not anneal to any of the Kpn1 satellite DNA fragments, it was found to associate with a major 2.35 Kb Kpn1 non-satellite DNA. The identity of this series of Kpn1 DNA fragments remains unknown, but the intensity of the hybridization signal indicate that they may be repetitive elements.

The sub-family of pKS36-like satellite II DNA appear to be characteristic of the genome of HeLa cells as it is absent (as 1.65 Kb and 1.95 Kb EcoR1 DNA fragments) from the genome of the human cell line MeWo. Restriction analysis of these two human cell lines for satellite II DNAs closely related to pKS36 revealed the presence of diverse domains, polymorphic in their organization.

In HeLa cells, the 1.65 Kb and 1.95 Kb EcoR1 DNA fragments, characterizing members of the sub-family of pKS36-like satellite II DNAs, are interrupted with Kpn1 sites. though a fraction of the 1.95 Kb EcoR1 contain Sau3A sites, the bulk of the members of the pKS36-like satellite family are devoid of such sites.

In contrast, the members of the MeWo pKS36-like satellite II DNA family are characterized by 2.95 Kb and 3.6 Kb EcoR1 DNA fragments. Furthermore, the units of repetition of these satellite DNAs as Kpn1 and Sau3A was largely found to be 1.8 Kb

in length, (as opposed to the equal distribution of 1.8 Kb and 3.6 Kb long units in HeLa cells).

In MeWo cells, the characteristic EcoR1 DNA fragments appear to be interrupted by Sau3A sites, whereas only a fraction contain Kpn1 sites.

Higgins et al. (1986) determined the genomic distribution of a pKS36-related satellite III DNA (D15Z1), by hybridization to MeWo DNA and to male and female placental DNA. The D15Z1 probe hybridized to placental male DNA and MeWo (male cell line) DNA in a pattern similar to that observed with pKS36 in MeWo. The author detected low level of EcoR1 restriction fragments with a 1.8 Kb periodicity in male DNA, but detectable hybridization was displayed with 1.8 Kb and 3.6 Kb EcoR1 fragments in female placental DNA. The Kpn1, Msp1, Sau3A and Rsa1 hybridization patterns of D15Z1 were similar to the one obtained with pKS36. Is is not known, to date, if D15Z1 contain a region homologous to the 49 mer of pKS36. However, from the similar hybridization patterns of the two cloned satellite and the presence of numerous Taq1 sites in D15Z1, it is likely that this satellite DNA belong to the satellite II DNA rather than to the satellite III DNA family.

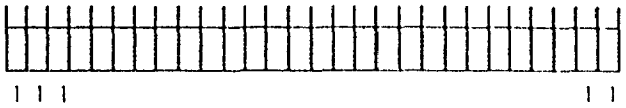
Figure 23-C presents the organization of pKS36-like satellite DNA as it would occur in HeLa cells. The hybridization results of pKS36 suggest that related DNA sequences may be organized on domains which differ by their restriction endonuclease site distribution. Domain I, consists of the few EcoR1 units that are devoid of Kpn1 or Sau3A recognition sites.

Figure 23: Organization of satellite DNA. The proposed organization was drawn from the observed patterns of hybridization of pKS36 to **HeLa**-hydrolyzed DNA (figures 20 and 21). A) satellite II and III DNA are organized in tandem arrays of 1.8 Kb EcoR1 monomeric units (unit 1). Mutations (*) between adjacent unit 1 repeats generate 3.6 Kb dimers (2), 5.4 Kb trimers (3), 7.2 Kb tetramers (4) etc... EcoR1 sites generated within tandem arrays (#) results in the new spacing of these sites, as it is the case for the HeLa characteristic 1.65 Kb (a) and 1.95 Kb (b) units. B) Interspersed organization of the diverse sized EcoR1 satellite DNA units. Non-satellite DNAs are represented by different shades. Vertical bars represent EcoR1 sites. C) organization of satellite DNA in domains. Domain I, II, III, and IV are as discussed in text. E: EcoR1, K: Kpn1, S: Sau3A. The two composites illustrate the chromosomic organization of the four domains; (a) adjacent or (b) interspesed with non-satellite sequences. The broken line (—S—) indicates that these domains may also be present on different chromosomes.

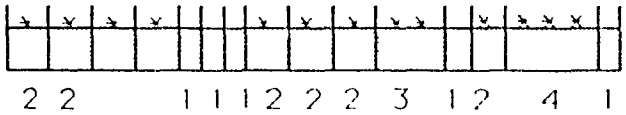
Figure 23

A)

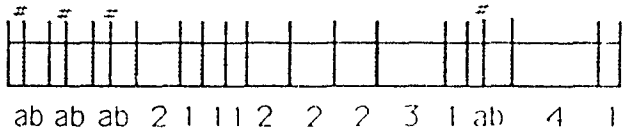
Amplification in tandem of 1.8 Kb long repeat units



Mutation at EcoR1 sites *



Mutation generating new EcoR1 sites #



B)

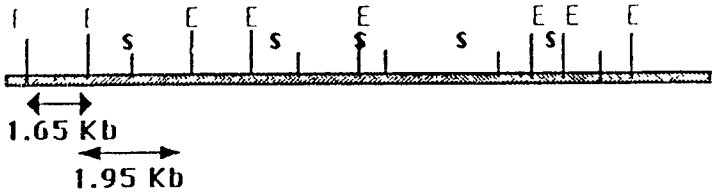
Interspersed organization

Figure 23 (Continues)

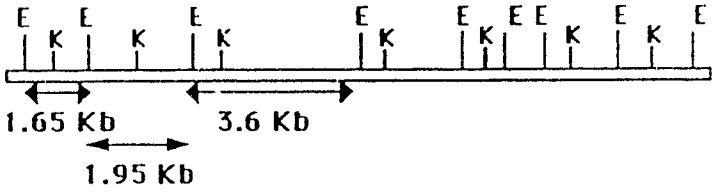
C)

Domain I

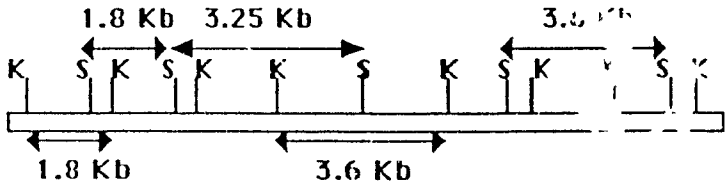
Domain II



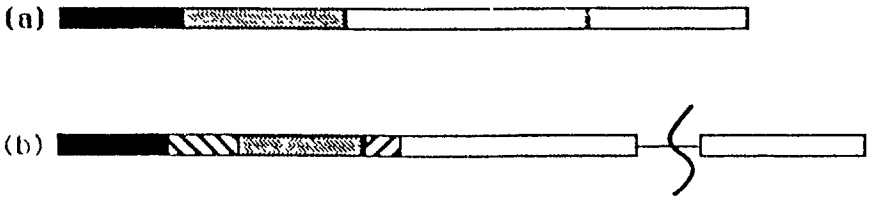
Domain III



Domain IV



Composite



As discussed earlier, these diverse EcoR1 units may either be derived from tandem arrays or be locally interspersed with other non-related DNA fragments (figure 23, panels A and B). Domain II is composed by the fraction of EcoR1 repeats containing Sau3A sites. Domain III, consists of the numerous EcoR1 elements interrupted by Kpn1 sites. Domain IV contains Sau3A units that are interrupted by Kpn1 sites. It is not known, as yet, if each of these domains are present of different chromosomes or if they are found, simultaneously, on the same chromosome.

Using FIGE to separate very large restriction fragments, we analyzed, by southern blot hybridization, the macro-organization of satellite DNAs in HeLa cells. The restriction endonucleases EcoR1, Cla1, and BamH1 do not appear to define large units of amplified satellite DNAs as we did not observe discrete sized hybridization bands, under the electrophoretic conditions utilized. However, we were able to identify HindIII fragments, ranging from 150 Kb to 500 Kb, that contain satellite II DNAs. To date, it is unclear if these large blocks contain tandem or discontinuous tandem arrays of satellite DNAs. The intensity of the 150 Kb and 500 Kb HindIII fragments suggest that they may be repetitive. Alternatively, they may contain numerous copies of satellite DNAs and be present in single copy in HeLa cells. The faint but discernable series of fragments, ranging approximately from 200 Kb to 400 Kb, would represent single copy DNA fragments containing low concentration of satellite DNAs, or highly divergent copies of pKS36-hybridizing

sequences. The diverse distribution of satellite DNAs on HindIII fragments may reflect their heteromorphic distribution on human chromosomes. The two major HindIII blocks might represent the satellite DNA organization common to a subset of human chromosomes. In contrast, each of the minor blocks might represent the satellite DNA organization that is specific for a particular chromosome.

Chapter VI

CONCLUDING REMARKS

"All Things considered,
I'd rather be in Philadelphia.."

W. C. Field
Epitaph on tombstone

SYNOPSIS

I-The search for repetitive mobile elements

When the amount of DNA in eukaryotes was found not to be correlated with organism complexity, it was proposed that eukaryotic cells contain "superfluous" DNAs that are either remains of extinct genes or the result of accidental aberrant DNA replication and recombination. It soon became apparent that the bulk of a eukaryotic genome was composed of distinct families of repeated sequences: the complex sequence Kpn1 and Alu1 DNA, and the simple sequence satellite DNA families. In analyzing the latter, still stigmatized as "junk" as no known function has yet been attached to it, it was shown that its presence might not be accidental as members of this family were found localized to specific regions of the chromosomes (centromeres, telomeres, heterochromatin).

In an attempt to shed some light on the nature and potential function of human satellite DNA, we primarily chose to focus our interests on certain reiterated elements that have the capability of moving (or transposing) their genomic information. Specifically, our interest was to look for evolutionary conserved mobile elements in two divergent species: Man and **Drosophila melanogaster** (a fruit-fly). Even though such findings would not have constituted an absolute proof for a functional

relationship between the related sequences, they might have presented suggestive evidence of an evolutionary structure/function link that would not have been detected otherwise.

In **Drosophila melanogaster**, mobile P-elements appear to be "recent" invaders, as they are found absent from the genome of fruit flies captured more than 40 years ago. These elements are the cause of the hybrid-dysgenesis syndrome, and may play an essential role in speciation, as their presence or absence in the **Drosophila** genome (in P-strains or M-strains, respectively) determines the viability of the offsprings from inter-strain mating.

As millions of years of divergent evolution separate fruit-flies from men, the search for conserved elements was conducted under permissive (or low) stringency conditions.

Under these conditions, serendipity led us to the isolation of a human EcoR1 DNA fragment that, though displaying several regions of partial homology to the P-element, turned out to be a member of the human satellite II DNA family.

II-Human satellite DNA

Originally separated and identified as distinct DNA fractions via ultracentrifugation in density gradients (Corneo et al., 1971), satellite DNAs appeared to fall into diverse families of simple DNA sequences, organized in tandem arrays in heterochromatic regions of human chromosomes (Gosden et al., 1975; Mitchell et al., 1979; Prosser et al., 1986). Satellite II and III families

consist of members that have evolved from the duplication (and amplification) of a basic pentameric unit 5' TTCCA 3'. When compared to satellite III DNA, satellite II DNA family members show a great deal of divergence from this basic pentameric unit.

The analysis of human satellite II and III DNAs has not been an easy task for several reasons. First, as these two families of simple DNA consist of "like-sequences", which are virtually undistinguishable by hybridization techniques, it has been difficult to assess their specific chromosomal location and organization. Moreover, as repeated sequences are prone to recombination, cloning tandemly arrayed simple satellite DNA has proved to be difficult. Brutlag and his co-workers (1977) have reported cases of instability in recombinant plasmids carrying repeated DNA sequences. In their experiments, the cloning of satellite DNA fragments ranging from 5 to 10 Kb in length, generated unstable recombinant plasmids containing satellite inserts 1.6 Kb in size. In a recent article (Neil et al., 1990), the stability in YACs (yeast artificial chromosomes) of chromosome Y-derived tandemly repeated DNA sequences was investigated. Most of the YAC clones, containing DNA insert of approximately 200 Kb, showed evidence of instability and additional rearrangements were also observed during transformation of the clones to a new background. During the course of this study, our group has encountered many difficulties in cloning some of the human satellite II and III DNA sequences. In our hands the 3.6 Kb dimeric version of the 1.8 Kb EcoR1 satellite DNA family (Sol et al., 1986) has totally eluded

cloning attempts (Chapter IV, this work). Although the reason for the instability of these EcoR1 clones has yet to be determined, it can be speculated that base modification (i. e., methylation) or the presence of SS-gaps along the DNA molecules (making them highly recombinogenic) impeded plasmid maintenance in the RecA$^+$ genetic background of NM522 cells. It is quite possible that a fraction of the satellite DNA population is refractory to cloning, therefore, the amount of data gathered to date on the structure and organization of satellite II and III may well be biased and reflect only a part of what is occuring in the genome. Indeed, the experiments conducted by Neil and his colleagues clearly demonstrated that "there is a bias against cloning DYZ1 (the 3.5 Kb HaeIII male-specific satellite DNA observed by Cooke et al., 1979, 1982) alone in YACs" (Neil et al., 1990).

III-Mutation hotspots in satellite II and III DNAs

Due to limited satellite DNA sequence information available at the time this work was started, the degenerate nature of satellite II (and III) DNAs was believed to be the results of mutations that occured randomly at any of the five positions of the ancestral pentameric unit.

However, in determining the first complete nucleotide sequence of a human 1.797 Kb EcoR1 satellite II DNA (clone pKS36), we were able to look at the basic structure of the tandem arrays. This analysis revealed that what looked like randomness, within the degenerate 5' TTCCA 3' repeats, had in

fact an exquisite non-chaotic fine structure. Mutational hotspots, as well as specific base conservations, were discovered within the pentameric repeats of pKS36. These features did not appear to be specific to our cloned satellite, as the analysis of other satellite DNA sequences (published by other groups or generated during this work) revealed a common theme in the non-random occurrence of pentamers. All sequences tested displayed hypervariability of the second cytosine and, to a lesser extent, hypervariability of the first cytosine. In eukaryotes, it is known that mutations within CpN dinucleotides may be the result of a two-stage methylation-deamination decay mechanism (Bird, 1980; Gruenbaun et al., 1982; Grafstrom et al., 1985; Nyce et al., 1986; Woodcock et al., 1987). Methylated deoxycytidines, which have been shown to be abundant in CpN dinucleotides (Nyce et al., 1986) can be deaminated in mammals, thus creating C->T transition mutations (Bird, 1980). Satellite II and III DNA members, having evolved from the amplification of a common CpN-rich ancestral pentameric unit, 5' TpTpCpCpA 3', may have therefore decayed via the methylation-deamination mechanism described above. Indeed, sequence analysis of the satellite DNAs revealed a high proportion of 5' TTTCA 3' and 5' TTCTA 3' pentamers that may be the result of such a mechanism. Furthermore, as TaqI (TCGA) sites are known landmarks of satellite II (and III, to a lesser extent) DNAs (Frommer et al., 1982), we have searched within the available satellite DNA sequences for pentamers that may be derived from a two-stage decay occuring at these CpG-rich

sites. This analysis showed that an increase in the occurrence of 5' TTCGA 3' Taq1-containing pentamers was correlated with an increase in 5' TTTGA 3' pentamers that may be generated by methylation/deamination at the cytosine residue. Our findings correlate with the report of Fowler et al. (1988) in strongly suggesting that satellite II and III DNAs do not evolve uniquely by random chance, as is expected for non-coding DNA elements. The presence of conserved bases within some members of the satellite DNA families, such as the first T of the pentameric unit, though not explainable at present, seems to reiterate the non-random evolution of these sequences. In addition to our and the Fowler et al. (1988) analysis, Bernardi and Bernardi (1986) found that both coding and non-coding DNA (they did not look at satellite DNA, at the time) appear to be, at the sequence level, under the same base compositional constraints, and thus under the same selection pressure. These authors postulated that "non-coding sequences may play a physiological role, which may have to do with the modulation of basic genome functions".

IV-The 49-mer non-satellite region

The close analysis of the satellite II DNA member isolated by our group, pKS36, revealed that the cloned element was a peculiar member of the satellite II DNA family, as a unique region (called the 49 mer) of non-satellite sequence interrupted the tandem arrays. This region, 49 bp long, was used as a probe to test diverse EcoR1 satellite DNA clones that were isolated during the course of this work, as well as a Kpn1 generated

satellite DNA, clone pBK1.8[20] (Shafit-Zagardo et al., 1982). This analysis revealed that the 49 mer was not an artifact of pKS36, as it was also detected in pKS36-closely related satellite clones, and absent in pBK1.8[20]-closely related satellite DNAs.

Moreover the genomic distribution of the 49 mer resembled the EcoR1-cleaved HeLa genomic hybridization pattern of the entire 1.787 Kb EcoR1 element of pKS36, under low stringency.

U-Sequence polymorphism of satellite DNA

In studying the Taq1 and Hinf1 site distribution patterns of cloned satellite DNAs, it was possible to classify them as members of either the Taq1-rich satellite II or Taq1-rare satellite III DNA families (Frommer et al., 1982; Prosser et al., 1986; Hollis and Hindley, 1988). The distribution of these restriction sites appeared to be highly polymorphic within the analyzed cloned satellite DNAs. In addition, it was observed that some members from the two different satellite DNA families (as defined by the frequency of Taq1 site occurrence) were more closely related (by southern blotting) than members belonging to the same family. Such was the case for clones pKS12 and pKS28, members of the satellite III DNA family, and clones pKS36 and pKS96,members of the satellite II DNA family, which appeared after hybridization experiment analysis to be closely related and to contain the 49 mer unique region. It is possible that these

members were originally derived from a common element containing the 49 mer landmark. Thus, the appearance of the 49 mer region in satellite DNA may have occured prior to the divergence of the satellite II and III families.

VI-Genomic distribution of satellite DNAs

In exploring the genomic distribution of clone pKS36 under different hybridization stringency conditions, it was possible to study the genomic organization of pKS36-like (closely related) sequences, as well as the genomic organization of the pKS36-related satellite II and III DNAs. As was previously reported (Burk et al., 1985; Sol et al., 1986), pKS36-related satellite II and III DNAs were found organized in tandemly arrayed 1.8 Kb and 3.6 Kb Kpn1 elements. The satellite II and III DNAs were also found in tandem of 1.8 Kb EcoR1 DNA fragments, though a large fraction could be seen as diverse sized-EcoR1 elements. Under high stringency, pKS36-like sequences still displayed the distinct Kpn1 organization. In HeLa cells, pKS36-like sequences are found mainly organized on 1.65 Kb and 1.95 Kb EcoR1 long units, though it is not clear at this time if these two abundant EcoR1 elements are adjacent or interspersed sequences. In the human melanoma-derived cell line MeWo, pKS36-like sequences appeared to be tandemly organized as 1.8 Kb and 3.6 Kb EcoR1 DNA fragments. The atypical EcoR1 organization of pKS36-like satellite DNAs, in HeLa cells, may thus have resulted from abnormal chromosome replication and rearrangement inherent to cells maintained in culture. Alternatively, the HeLa 1.65 Kb

and 1.95 Kb EcoR1 elements may have resulted from point mutations within a 3.6 Kb unit of amplification generating new spacing of the EcoR1 sites. Further heterogeneity in EcoR1 site distribution within tandemly arrayed satellite DNA may also generate diverse sized restriction fragments that would appear as if interspersed in human cells. HeLa pKS36-like family members appeared distributed in four polymorphic domains characterized by an heteromorphic restriction site distribution. It is possible that these domains define specific satellite organization on distinct chromosomes or, alternatively, they may be characteristic of sub-regions common to a number of human chromosomes.

Using the technique of field inversion gel electrophoresis, we were able to identify large HindIII blocks (or segments), ranging from approximately 150 Kb to 500 Kb in length, containing different density of satellite DNA sequences homologous to pKS36. Satellite DNA appear to be present in high copy number in two of the HindIII segments (150 Kb and 500 Kb long), whereas they are found in low copy number in segments ranging from 200 Kb to 400 Kb in length. Alternatively, the difference in the distribution of satellite DNA may reflect a difference in the copy number of large HindIII blocks (repetitive versus single copy blocks). In a recent article published by Neil et al. (1990), satellite DNAs, homologous to the 3.5 Kb HaeIII satellite DNA located in the heterochromatic region of the long arm of chromosome Y (Cooke et al., 1979), were found organized

on a series of repetitive Sst1 fragments ranging from 90 Kb to over 300 Kb in length, in the human cell line OHEN (49, XYYYY).

VII-Chromosomal distribution of satellite DNAs

Using Human/rodent hybrid-cell line DNAs, we attempted to determine the specific chromosomal location of two related satellite II DNAs, the 1.8 Kb EcoR1 insert of pKS36 and the 1.8 Kb Kpn1 insert of clone pBK1.8[20]. Though these two satellite DNAs were shown, by restriction analysis as well as by southern hybridization, not to be identical, their hybridization pattern to a hybrid cell DNA panel appeared identical. Human satellite II DNA related to both the EcoR1 and the Kpn1 satellite were found mainly to reside on chromosomes 7, 12, 14, 15, 16 and 22. In order to narrow down the chromosomal distribution of pKS36-like satellite sequences, the 49 mer (unique region found in pKS-like sequences) was used to probe, under permissive stringency conditions, the same hybrid DNA panel. Unfortunately, due to cross-reactivity of this probe to rat DNA, we were not able to give a more precise distribution of pKS36-like sequences. Satellite II DNAs related to pKS36 and pBK1.8[20] were found by dot blot hybridization to represent 2 to 3 % of the genome of HeLa cells. However, pKS36-like sequences, containing the 49 mer element, appeared to make up less than 1 % of HeLa DNA.

DIRECTIONS FOR FUTURE RESEARCH

A clearer picture of the chromosomal organization of pKS36-like satellite DNAs, could be readily achieved by the use of human cell sorted chromosomes, a technology that is not currently available (unfortunately) in our laboratory. Analysis of sorted chromosomes would clarify wether or not satellite II DNA display a chromosome-specific organization, as is the case for the alphoid centromeric DNAs (Chapter I).

Recently, DNA binding proteins that specifically recognize and bind to mini-satellite DNA (see Chapter I) were discovered in several species (Collick and Jeffreys, 1990). In these experiments, diverse nuclear protein extracts were size separated by SDS-polyacrylamide gel electrophoresis and subsequently transferred to nitrocellulose filters. As substrates, a series of mini-satellite DNAs, composed of diverse numbers of the repeat unit 5'GTGGGCAGGAAG 3' in tandem (200 to 500 bp long), were labeled by random priming and allowed to bind to the immobilized proteins on the membrane. The results show that minisatellite-binding proteins are found in mouse brain and rabbit (40 Kdalton protein), **Xenopus** (35 and 39 Kdalton proteins), and in **Drosophila** (37 Kdalton protein). Moreover, their experiments show that not only the sequence, but the numbers of tandem repeats, are important factors for stable DNA-protein interactions, as increasing the number of repeats

from 4 to 5 induces a ten-fold increase in the binding affinity of the protein to its substrate. Analysis of satellite DNA-binding proteins and the genes that encode them may provide clues to the function or behavior of satellite DNAs, as well as their evolution.

During the course of this study, attempts were repeatedly made to identify DNA binding proteins that would specifically recognize and bind to our cloned EcoR1 satellite II DNA member. Using whole cell, and subsequently nuclear, protein extracts in gel retardation assays (Strauss and Varshavski, 1984), I was unable to identify any specific binding activity over the background of non-specific protein-DNA interactions. Recently, however, a student from our laboratory (Claire Fouquet) attempted to approach the same problem using the technique of Collick and Jeffreys (1990) as described above. In these experiments, HeLa nuclear protein extracts were size separated by SDS-polyacrylamide gel electrophoresis, transferred (via electroblotting) to membranes, and binding assays performed using random-primed fragments of the EcoR1 satellite II DNA. Proteins that may be specific to the cloned satellite, when compared to a non-satellite control DNA (pBR322), were detectable. They are now pending analysis.

With regard to possible satellite-specific DNA binding proteins, it is possible that interactions between satellite DNA and nuclear proteins, if they do occur, may be tissue-specific or else may depend on the developmental stage of a given cell. Gaubatz and Cutler (1990) analyzed a mouse satellite DNA that

was assumed not be transcribed in mouse cells, as no satellite transcripts were detected in mouse liver, kidney and brain tissues. However, in analyzing total RNA preparations from heart tissues at different ages, they found that, though no satellite DNA-related transcripts were detectable in young animals (up to 6 months old), the level of satellite-derived transcripts began to appear at 12 months and increased with the aging of the animal. With this in mind, it may be worthwhile to extend the type of gel retardation and binding assays performed on the cloned EcoR1 satellite DNA using nuclear extracts from diverse tissue origins and/or ages.

Unstable expression of euchromatic genes has been shown to be connected with rearrangements of DNA sequences which result, via non-homologous recombination or the action of transposable elements, in positioning heterochromatic regions close to the gene under study. This phenomenon, called "position effect variegation", has been identified in **Drosophila melanogaster** and in mouse cells (Spofford, 1976; Talarico et al., 1988). Mouse satellite DNA was shown to influence the expression of adjacent TK (thymidine kinase) genes in mouse cells (Talarico et al., 1988). These authors suggested that the phenotypic instabilities they observed resulted from satellite DNA induced- heterochromatization of the inserted exogenous TK gene.

In order to study the effect of human satellite DNA sequences on adjacent gene expression, Claire Fouquet has

cloned, upstream or downstream from the HSV-TK gene of plasmid pSV2neoKT (Goring and DuBow, 1985), one or multiple copies of the pKS36-EcoR1 satellite II DNA. Following electroporation of these constructs into a TK-deficient human cell line (AK143B), the frequencies of stable TK^+ transfectants were recorded. Plasmid pSV2neoKT, devoid of satellite sequences, was used as a control. Preliminary results suggest an inhibitory effect of these satellite sequences on the expression of the TK gene in mammalian cells. Furthermore, it seems that the inhibition increases (though not proportionally) with an increase in the number of 1.8 Kb adjacent satellite DNA. It is not clear, however, if the inhibition is sequence specific. To address this point, plasmids containing 1.8 Kb (or 3.6 Kb) long non-satellite DNA in place of the satellite element will be included as controls in further transfection analysis.

Comprehension of satellite DNAs biology is still a long way off. One way to fully understand their significance may only be achieved by the manipulation of model chromosomes. In this approach the need is great for the development of new technologies that would allow the placement of these repeated DNAs into context on human artificial mini-chromosomes (HAC). Though HAC technology is not yet at hand, the ideal eukaryotic vehicle should contain an origin of replication (capable of replicating long linear DNA molecules), a functional centromere (to ensure proper segregation into daughter cells), terminal telomeric sequences, and one or several reporter genes. As

previously reported by Neil et al., (1990), yeast artificial chromosome (YAC) cloning vectors may not be suitable for these types of analysis as satellite DNAs appeared unstable in this system. A likely alternative to YACs may be the use of the newly developed P1 cloning technology (Sternberg, 1990). P1 vector has been shown to support the stable maintenance, under the tightly regulated P1-plasmid stringent origin of replication, of 90 Kb to 95 Kb human DNA inserts in **E. coli**. In addition, the presence on the vector of a loxP site (recognition sequence of the P1 phage cre site specific recombinase), allows its linearization **in vitro** (Sol and Sternberg, in preparation). After the cloning of diverse lengths (and sequences) of satellite DNA into the P1 vehicle, linearization via cre-mediated site-specific recombination and transfer by microinjection into mammalian cells, one might be able to assess the effects (or absence thereof) of these sequences on the behavior of the mini-chromosomes during cell growth and division.

It is difficult, to say the least, to decipher the role that the highly repeated simple satellite sequences may play in human cells. Walker (1971) discussed at length the selective advantages (or disadvantages) of organisms having amplified DNA sequences. He argued that "amplified mouse satellite DNAs confer an advantage, not to the mouse, but to its chromosomes". It seems that tandemly arrayed satellite DNAs became molecular steps in the construction of chromosomes, since they are common (in structure if not in sequence) to most

eukaryotes. The tandem organization of non-coding DNA has been preserved for millions of years of evolution, and thus can hardly be expected to be non-functional. If indeed they contribute to cell survival, I believe that satellite DNAs may do so as epigenetic (or structural) information.

The bulk of the work presented in this thesis has pointed out the microscopic and macroscopic molecular dynamic structure of human satellite II and III DNAs, as a means to convey a better "feel" for satellite DNAs as entities that seem to be haunting eukaryotic chromosomes (and their investigators) for, as yet, no apparent reason.

LIST

OF

REFERENCES

Adams D., (1989) Quote from "The restaurant at the end of the universe" In: The more than complete Hitchiker's guide to the galaxy.

Alberts B., Bray D., Lewis J., Raff M., Roberts K. and Watson J. D., (1983) "Molecular biology of the cell" Garland Publishing Inc. NY NY)

Arrighi F. E., and Hsu T. C., (1971) "Localization of heterochromatin in human chromosomes" Cytogenetics 10, 81-86.

Baltimore D., (1981) "Gene conversion: some implications for Immunoglobulin genes" Cell 24, 592-594.

Barnes W. M.,(1977) "Plasmid detection and sizing in single colony lysates" Science 195, 393-394.

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.

Bell G. I., Selby M. J., and Rutter W. J., (1982) "The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences" Nature 295, 31-35.

Beltz G. A., Jacobs K. A., Eickbush T. H., Cherbas P. T., and Kafatos F. C., (1983) "Isolation of multigene families and determination of homologies by filter hybridization methods" Methods Enzymol. 100, 266-285.

Beridze T., (1986) "Satellite DNA" Springer-Verlag, Berlin Heidelberg New york Tokyo.

Bernardi G. and Bernardi G., (1986) "The human genome and its evolutionary context" C. S. H. Symp. Quant. Biol. Vol LI, CSH labs.

Bingham P. M., Kidwell M. G., and Rubin G. M., (1982) "The molecular basis of P-M dysgenesis: the role of the P-element, a P-strain specific transposon family" Cell 29, 995-1004.

Bird A. P., (1980) "DNA methylation and the frequency of CpG in animal DNA" Nucleic Acids Res. 8, 1499-1504.

Bond H. E., Flamm W. G., Burr H. E., and Bond S. B., (1967) "Mouse satellite DNA. Further studies on its biological and physical characteristics and its intracellular localization" J. Mol. Biol. 27, 289-302.

Brel J., (1968) Quote from "La Quête" In the opera: L'homme de la Mancha (The man from Mancha).

Brenner S., Pepper D., Berns M. W., Tan E., and Brinkley B. R., (1981) "Kinetochore structure duplication and distribution in mammalian cells: analysis by human autoantibodies from Scleroderma patients" J. Cell Biol. 91, 95-102.

Britten R. J., and Kohne D. E., (1968) "Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms" Science 161, 529-540.

Britten R. J., and Kohne D. E., (1969) "Repetition of nucleotide sequence in chromosomal DNA" In: Handbook of Molecular cytology, A. Lima de Faria (Ed.), North Holland Publ. Co, London, 21-51.

Britten R. J., Baron W. F., Stout D. B., and Davidson E. H., (1988) "Sources and evolution of human Alu repeated sequences" Proc. Natl. Acad. Sci. USA 85, 4770-4774.

Brutlag D. L., Carlson M., Fry K., and Hsieh T. S., (1977) "Synthesis of hybrid bacterial plasmids containing highly repeated satellite DNA" Cell 10, 509-519.

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

Choo K. H., Earle E., Vissel B., and Filby R. G., (1990a) "The centromere of the acrocentric chromosomes carry different subfamilies of alpha satellite DNA" Genomics 7, 143-151.

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

Clewell D. B., (1972) "Nature of Col E 1 plasmid replication in **Escherichia coli** in the presence of chloramphenicol" J. Bacteriol. 110, 667-676.

Collick A., and Jeffreys A. J., (1990) "Detection of a novel minisatellite-specific DNA-binding protein" Nucleic Acids Res. 18, 625-629.

Cooke H. J., (1976) "Repeated sequence 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 chromosomes" Nucleic Acids Res. 6, 3177-3197.

Cooke H. J. Schmidtke J., and Gosden J. R., (1982) "Characterization o. a human Y chromosome repeated sequence and related sequences in higher primates" Chromosoma 87, 491-502.

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

Corneo G., Ginelli E., and Polli E., (1971) "Renaturation properties and localization in heterochromatin of human satellite DNA's" Biochim. Biophys. Acta 247, 528-534.

Crick F., (1979) Quote in discussion of a paper by Lima de Faria "A prediction of gene location and classification of genes according to the chromosome field" pp25-38, In: A. Benzon Symposium 13, J. Engberg, H. Klenow, and V. Leick (eds) Munksgaard-Copenhagen.

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

Deininger P. L., and Daniels G. R., (1986) "The recent evolution of mammalian repetitive DNA elements" Trends Genet. 2, 76-80.

Donehower L., Furlong C., Gillespie D., and Kurnit D., (1980) "DNA sequence of baboon highly repeated DNA: evidence for evolution by nonrandom unequal crossovers" Proc. Natl. Acad. Sci. USA 77, 2129-2133.

Doolittle W. F., and Sapienza C., (1980) "Selfish genes, the phenotype paradigm and genome evolution" Nature 284, 601-603.

Dover G. A., (1982) "Molecular drive: a cohesive mode of species evolution" Nature 299, 111-117.

Dover G. A., (1986) "Molecular drive in multigene families: How biological novelties arise, spread and are assimilated. " Trends Genet. 2, 159-165.

Engels W. R., johnson-Schlitz D. M., Eggleston W. B., and Sved J., (1990) "High frequency P-element loss in **Drosophila** is homology independent" Cell 62, 515-525.

Ford M., and Fried M., (1986) "Large inverted duplications are associated with gene amplification" Cell 45, 425-430

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

Fowler C., Drinkwater R., Skinner J., and Burgoyne L., (1988) "Human satellite-III DNA: an example of "macrosatellite" polymorphism" Hum. Genet. 79, 265-272.

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

Fry K., Poon R., Whitcome P., Idriss J., Salser W., Mazrimas J., and Hatch F., (1973) "Nucleotide sequence of HS-beta satellite DNA from Kangaroo rat, **Dipodomys ordii**" Proc. Natl. Acad. Sci. USA 70, 2642-2646.

Gall J. G., (1981) "Chromosome structure and the C-value paradox" J. Cell. Biol. 91 3s-14s.

Gama-Sosa M. A., Wang R. Y-H., Kuo K. C., Gehrke C. W., and Ehrlich M., (1983) "The 5-methylcytosine content of highly repeated sequences in human DNA" Nucleic Acids Res. 11, 3087-3095.

Gardiner K., Laas W., and Patterson D., (1986) "Fractionation of large mammalian DNA restriction fragments using vertical pulse-field gradient gel electrophoresis" Somatic cell Mol. Genet 12, 185-195.

Gaubatz J. W., and Cutler R. G., (1990) "Mouse satellite DNA is transcribed in senescent cardiac muscle" J. Biol. Chem. 265, 17753-17758.

Gilbert W., (1978) "Why genes in pieces?" Nature 271, 501.

Goring D. R., and DuBow M. S., (1985) "A cytotoxic effect associated with 9-(dihydroxy-2-propoxymethyl)-guanine is observed during the selection for drug resistant human cells containing a single Herpesvirus Thymidine kinase gene" Biochem. Biophys. Res. Com. 133, 195-201.

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

Gough J. A., and Murray N. E.,(1983) "Sequence diversity among related genes for recognition of specific target in DNA molecules" J. Mol. Biol. 166, 1-19

Grafstrom R. H., Yuan R., and Hamilton D. L., (1985) "The characteristics of DNA methylation in an in vivo DNA synthesizing system from mouse fibroblasts" Nucleic Acids Res. 13, 2827-2842.

Grimaldi G., and Singer M. F., (1983) "Members of the KpnI family of long interspersed repeated sequences join and interrupt alpha-satellite in the monkey genome" Nucleic Acids Res. 11, 321-338.

Gruenbaum Y., Cedar H., and Razin A., (1982) "Substrate and sequence specificity of a eukaryotic DNA methylase" Nature 295, 620-622.

Guarente L., Yocum R. R., and Gifford P., (1982) "A Gal10-CYC1 hybrid yeast promotor identifies the GAL4 regulatory region as an upstream site" Proc. Natl. Acad. Sci. USA 79, 7410-7414.

Hatch F. T., and Mazrimas J. A., (1974) "Fractionation and characterization of satellite DNAs of the kangaroo rat (Dipodomys ordii)" Nucleic Acids Res. 1, 559-575.

Hearst J. E., Cech T. R., Marx K. A., Rosenfeld A., and Allen J. R., (1974) "Characterization of the rapidly renaturing sequences in the main CsCl density bands of Drosophila, mouse and man" Cold Spring Harbor Symp. Quant. Biol. 38, 329-339

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 93, 77-86.

Holden J. J. A., Reimer D. L., Higgins M. J., Roder J. C., and White B. N., (1985) "Amplified sequences from chromosome 15, including centromeres, nucleolar organizing regions, and centromeric heterochromatin, in HSR in the human melanoma cell line MelWo" Cancer Genet. Cytogenet. 14, 131-146.

Holden J. J. A., Reimer D. L., Higgins M. J., Roder J. C., and White B. N., (1986) "Rearrangements of chromosomal regions containing ribosomal genes and centromeric heterochromatin in the melanoma cell line MelWo" Cancer Genet. Cytogenet. 21, 221-237.

Holmes D. S., and Quigley M., (1981) "A rapid boiling method for the preparation of bacterial plasmid" Analytical Biochem. 114, 193-197.

Horz W., and Altenburger W., (1981) "Nucleotide sequence of mouse satellite DNA" Nucleic Acids Res. 9, 683-696.

Hourcade D., Dressler D., and Wolfson J., (1973) "The amplification of ribosomal RNA genes involves a rolling circle intermediate" Proc. Natl. Acad. Sci. USA 70, 2926-2930.

Jackson J. A., and Fink G. R., (1981) "Gene conversion between duplicated genetic elements in yeast" Nature 292, 306-311.

Jang K. L., and Latchman D. S., (1989) "HSV infection induces increased transcription of Alu repeated sequences by DNA polymerase III" FEBS Lett. 258, 255-258.

Jarman A. P., and Wells R. A., (1989) "Hypervariable minisatellites: recombinators or innocent bystanders" Trends Genet. 5, 367-371.

Jeffreys A. J., Wilson V., and Thein S. L., (1985) "Hypervariable mini-satellite regions in human DNA" Nature 314, 67-73.

John B., (1988) "The biology of heterochromatin" In: Verma R. S., (Ed) Heterochromatin: Molecular and structural Aspects. Cambridge Press New York.

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

Kidwell M. G., and Kidwell J. F., (1976) "Selection for male recombination in **Drosophila melanogaster**" Genetics 84, 333-351.

Kinsey J. A., and Helber J., (1989) "Isolation of a transposable element from **Neurospora crassa**" Proc. Natl. Acad. Sci. USA 86, 1929-1933.

Kit S., (1961) "Equilibrium sedimentation in density gradient of DNA preparations from animal tissues" J. Mol. Biol. 3, 711-716.

Kornberg A., Bertsch L. L., Jackson J. F., and Khorana H. G., (1964) "Enzymatic synthesis of DNA. XVI. Oligonucleotides as templates and mechanism of their replication" Proc. Natl. Acad. Sci. USA 51, 315-318.

Kudo S., and Fukuda M., (1989) "Structural organization of glycophorin A and B genes: Glycophorin B gene evolved by homologous recombination at Alu repeated sequences" Proc. Natl. Acad. Sci. USA 86, 4619-4623.

Kurnit D. M., and Maio J. J., (1974) "Variable satellite DNA's in the African green monkey **Cercopithecus aethiops**" Chromosoma 45, 387-400.

Laski F. A., Rio D. C., and Rubin G. M., (1986) "Tissue specificity of **Drosophila** P-element transposition regulated at the level of mRNA splicing" Cell 44, 7-19.

Lehrman M. A., Goldstein J. L., Russell D. W., and Brown M. S., (1987) "Duplication of seven exons in LDL receptor gene caused by Alu-Alu recombination in a subject with familial hypercholesterolemia" Cell 48, 827-835.

Livneh E., Glazer L., Segal D., Schlessinger J., and Shilo B. Z., (1985) "The Drosophila EGF receptor gene homolog: conservation of both hormone binding and kinase domains" Cell 40, 599-607.

Ljungquist E., and Bukhari A. I., (1977) "State of prophage Mu DNA upon induction" Proc. Natl. Acad. Sci. USA 74, 3143-3147.

Low B., (1968) "Formation of merodiploids in matings with a class of Rec-recipient strains of Escherichia coli K12" Proc. Natl. Acad. Sci. USA 60, 160-167.

Lumpkin O. J., and Zimm B. H., (1982) "Mobility of DNA in gel electrophoresis" Biopolymers 21, 2315-2316.

Mahtani M. M., and Willard H. F., (1990) "Pulse field gel analysis of alpha-satellite DNA at human X chromosome centromere: High frequency polymorphisms and array size estimate" Genomics 7, 607-613.

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

Maio J. J., Brown F. L., and Musich P. R., (1981a) "Toward a molecular paleontology of primate genomes. I. The HindIII and EcoR1 dimer families of alphoid DNAs" Chromosoma 83, 103-125.

Maio J. J., Brown F. L., and Musich P. R., (1981b) "Toward a molecular paleontology of primate genomes. II The KpnI family of alphoid DNAs" Chromosoma 83, 127-144.

Maniatis T., Fritsh E. F., and Sambrook J., (1982) "Molecular cloning, A laboratory manual" Cold Spring Harbor Laboratory, N. Y.

Manuelidis L., (1977) "A simplified method for preparation of mouse satellite DNA" Anal. Biochem. 78, 561-568.

Manuelidis L., (1981) "Consensus sequence of mouse satellite DNA indicates it is derived from tandem 116 bp repeats" FEBS Lett. 129, 25-28.

Manuelidis L., (1982) in Dover G. A., and Flavell R. B., (Eds), Genome Evolution, Academic Press N. Y., 263-285.

Maresca A., and Singer M. F., (1983) "Deca-satellite: a highly polymorphic satellite that join alpha satellite in the African green monkey genome" J. Mol. Biol. 164, 493-511.

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

Mattoccia E., and Comings D. E., (1971) "Buoyant density and satellite composition of DNA of mouse heterochromatin" Nature New Biol. 229, 175-176.

Maxam A. M., and Gilbert W., (1980) "Sequencing end-labeled DNA with base-specific chemical cleavages" Methods Enzymol 65, 499-560.

Mazrimas J. A., and Hatch S. T., (1972) "A possible relationship between satellite DNA and the evolution of Kangaroo rat species (genus **Dipodomys ordii**)" Nature New Biol. 240, 102-105.

McCarroll R. M., and Fangman (1988) "Time of replication of yeast centromeres and telomeres" Cell 54, 505-513.

McClintock B., (1951) "Chromosome organization and genic expression" CSH Symp. Quant Biol XVI, 17-47.

McClintock B., (1956) "Controlling elements and the gene" CSH Symp. Quant Biol XXI, 197-216.

Meselson M., Stahl F. W., and Vinograd G., (1957) "Equilibrium sedimentation of macromolecules in density gradients" Proc. Natl. Acad. Sci. USA 43, 581-588.

Miller J. H., (1972) "Experiments in molecular genetics" Cold Spring Harbor, N. Y.

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.

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

Neil D. L., Villasante A., Fisher R. B., Vetrie D., Cox B., and Tyler-Smith C., (1990) "structural instability of human tandemly repeated DNA sequences cloned in yeast artificial chromosome vectors" Nucleic Acids Res. 18, 1421-1428.

Nyce J., Liu L., and Jones P. A., (1986) "Variable effects of DNA-synthesis inhibitors upon DNA methylation in mammalian cells" Nucleic Acid Res. 14, 4353-4367.

O'Hare K., and Rubin G. M., (1983) "Structure of P transposable elements and their sites of insertion and excision in the Drosophila melanogaster genome" Cell 34, 25-35.

Orgel L. E., and Crick F. H. C., (1980) "Selfish DNA: the ultimate parasite" Nature 284, 604-607.

Pardue M. L., and Gall J. G., (1969) "Formation and detection of RNA-DNA hybrid molecules in cytological preparations" Proc. Natl. Acad. Sci. USA 63, 378-383.

Peacock W. J., Brutlag D., Goldring E., Appels R., Hinton C. W., and Lindley D. L., (1974) "The organization of highly repeated DNA sequences in **Drosophila melanogaster** chromosomes" Cold Spring Harbor Symp. Quant. Biol. 38, 405-416.

Pech M., Igo-Kemenes T., and Zachau H. G., (1979) "Nucleotide sequence of a highly repetitive component of rat DNA" Nucleic Acids Res. 7, 417-432.

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

Quentin Y., (1988) "The Alu family developped through successive waves of fixation closely connected with primate lineage history" J. Mol. Evol. 27, 194-202.

Radic M. Z., Lundgren K., and Hamkato B. A., (1987) "Curvature of mouse satellite DNA and condensation of heterochromatin" Cell 50, 1101-1108.

Rhim J. S., Cho H. Y., and Heubner R. J., (1975) Int. J. Cancer 15, 23-29.

Ris P. W. J., and Korenberg J., (1979) "Chromosomes structure and levels of chromosomes organization" pp267-361, In: Prescott and Goldstein (Eds) Cell Biology vol. 2, Academic Pres New York.

Roeder G. S., Farabaugh P. J., Chaleff D. T., and Fink G. R., (1980) "The origine of yeast instability in yeast" Science 209, 1375-1380.

Rosenberg H., Singer M., and Rosenberg M., (1978) "Highly reiterated sequences of SimianSimianSimianSimianSimians" Science 200, 394-402.

Rubin G. M., Kidwell M. G., and Bingham P. M., (1982) "The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations" Cell 29, 987-994.

Saedler H., and Nevers P., (1977) "Transposable genetic elements as agents of gene instabilities and chromosomal rearrengements" Nature 268, 109-115.

Salomon R., Kaye A. M., and Herzberg M., (1969) "Mouse nuclear satellite DNA: 5-methylcytosine content, pyrimidine isoplith distribution and electron microscopic appearance" J. Mol. Biol. 43, 581-592.

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

Schildkraut C. L., Marmur J., and Doty P., (1962) "Determination of the base composition of DNA from its buoyant density in CsCl" J. Mol. Biol. 4, 430-443.

Schildkraut C. L., and Maio J. J., (1968) "Studies on the intranuclear distribution and properties of mouse satellite DNA" Biochim. Biophys. Acta 161, 76-93.

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

Schmid C. W., and Shen C. K. J., (1985) "The evolution of interspersed repetitive DNA sequences in mammals and other vertebrate" pp323-358. In: R. J. MacIntyre (Ed), Molecular evolutionary genetics. Plenum Publishing Corp, New York.

Schwartz D. C., and Cantor C. R., (1984) "Separation of yeast chromosome-sized DNAs by pulse gradient gel electrophoresis" Cell 37, 67-75.

Shafit-Zagardo B., Maio J. J., and Brown F. L., (1982) "Kpn1 family of long, interspersed repetitive DNAs in human and other primate genomes" Nucleic Acids Res. 10, 3175-3193.

Simmons M. C., Maxwell J., Haliotis T., Higgins M. J., Roder J. C., White B. N., and Holden J. A., (1984) "Amplified Kpn1 repetitive DNA sequences in homogeneously staining regions of a human Melanoma cell line" JNCI 72, 801-808.

Singer D. S., (1979) "Arrangement of highly repeated DNA sequence in the genome and chromatin of African green monkey" J. Biol. Chem. 254, 5506-5514.

Singer M. F., (1982a) "SINES and LINES: Highly repeated short and long interspersed sequences in mammalian genomes" Cell 28, 433-434.

Singer M. F., (1982b) "Highly repeated sequences in mammalian genomes" Int. Rev. Cytol. 76, 67-117.

Skinner D. M., and Beattie W. G., (1973) "Cs2So4 gradients, containing both Hg++ and Ag++, effect the complete separation of satellite DNA having identical densities in neutral CsCl gradients" Proc. Natl. Acad. Sci. USA 70, 3108-3110.

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

Smith G. P., (1978) "What is the origin and evolution of repetitive DNAs?" Trends Biochem. Sci. 3, 34-36.

Smith E. G., and Summers M. D., (1980) "The bidirectional transfer of DNA and RNA to nitrocellulose or diazobenzylomethyl paper" Anal. Biochem. 109, 123-129.

Sol K., Lapointe M., MacLeod M., Nadeau C., and DuBow M. S., (1986) "A cloned fragment of **HeLa** DNA containing consensus sequences of satellite II and III DNA hybridizes with the **Drosophila** P-element and with the 1.8 Kb family of Kpn1 fragments" Biochim. Biophys. Acta 868, 128-135.

Southern E. M., (1971) "Effects of sequence divergence on the reassociation properties of repetitive DNAs" Nature 232, 82-83.

Southern E. M., (1974) "Eukaryotic DNA" In: Burton (Ed) Int. Rev. Sci. Biochem. Nucleic Acids 6, 101-139.

Southern E. M., (1975a) "Detection of specific sequences among DNA fragments separated by electrophoresis" J. Mol. Biol. 98, 503-507.

Southern E. M., (1975b) "Long range periodicities of mouse satellite DNA" J. Mol. Biol. 94, 51-69.

Spofford J., (1976) "Position effect variegation in **Drosophila**" In M. Ashburner and E. Novitski (ed.). The genetics and biology of **Drosophila**. Academic Press. Inc. New York.

Spradling A. C., and Rubin G. M., (1982) "Transposition of cloned P-element into **Drosophila** germline chromosomes" Science 218, 341-347.

Spradling A. C., and Rubin G. M., (1983) "The effect of chromosomal position on the expression of the **Drosophila** Xanthine deshydrogenase gene" Cell 34, 47-57.

Sternberg N. L., (1990) "Bacteriophage P1 cloning system for the isolation, amplification, and recovery of DNA fragments as

large as 100 kilobase pairs" Proc. Natl. Acad. Sci. USA 87, 103-107.

Strauss F., and Varshavski A., (1984) "A protein binds to a satellite DNA repeat at three specific sites that would be brought into mutual proximity by DNA folding in the nucleosome" Cell 37, 889-901.

Sutcliffe (1979) "Complete nucleotide sequence of the Escherichia coli plasmid pBR322" Cold Spring Harbor Symp. Quant. Biol. 43, 77-90.

Talarico D., Peverali A. F., Ginelli E., Meneveri R., Mondello C., and Della Valle G., (1988) "Satellite DNA induces unstable expression of the adjacent Herpes Simplex Virus TK gene cotransfected in mouse cells" Mol. Cell. Biol. 8, 1336-1344.

Ten Hagen K. G., Gilbert D. M., Willard H. F., and Cohen S. N., (1990) "Replication timing of DNA sequences associated with human centromeres and telomeres" Mol. Cell. Biol. 10, 6348-6355.

Thayer R. E., and Singer M. F., (1983) "Interruption of an alpha satellite array by a short member of the Kpn1 family of interspersed highly repeated monkey DNA sequences" Mol. Cell. Biol. 3, 967-973.

Tagarro I., Gonzalez-Aguilera J. J., and Fernandez-Peralta A. M., (1991) "Taq1 digestion reveals fractions of satellite DNAs on human chromosomes" genome 34, 251-254.

Ullu E., and Tschudi C., (1984) "Alu sequences are processed 7SL RNA genes" Nature 312, 171-172.

Ulrich A., Shine J., Chirgwin J., Pictet R., Tischer E., and Rutter W. J., (1977) "Rat insulin genes: construction of plasmid containing the coding sequences" Science 196, 1313-1319.

Viera J., and Messing J., (1982) "The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers" Gene 19, 259-268.

Vinograd J., and Hearst J. E., (1962) "Equilibrium sedimentation of macromolecules and viruses in density gradients" Fortschr Chemie Org. Naturst 20, 372-422.

Walker P. M. B., (1971) "Repetitive DNA in higher organisms" Prog. Biophys. Mol. Biol. 23, 145-190.

Walsh J. B., (1987) "Persistence of tandem arrays: Implications for satellite and simple sequence DNAs" Genetics 115, 553-567.

Waring M., and Britten R. G., (1966) "Nucleotide sequence repetition: A rapidly reassociating fraction of mouse DNA" Science 154, 791-794.

Waye J. S., Mitchell A. R., and Willard H. F., (1988) "Organization and genomic distribution of 82H alpha satellite DNA; Evidence for a low-copy or single-copy alphoid domain located on human chromosome 14" Hum. Genet. 78, 27-32.

Weiner A. M., Deininger P. L., and Efstratiadis A., (1986) "Non viral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information" Annu. Rev. Biochem. 55, 631-661.

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

White R., Pasztor L. M., and Hu F., (1975) Mouse satellite DNA in noncentromeric heterochromatin of cultured cells" Chromosoma 50, 275-282.

Willard H. F., Wevrick H., and Warkurton P. E., (1989) "Human centromere organization and potential role of alpha satellite DNA" Prog. Clin. Biol. 318, 9-18.

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

Woodcock D. M., Crowther P. J., and Diver W. P., (1987) "The majority of methylated deoxycytidines in human DNA are not in CpG dinucleotide" Biochem. Biophys. Res. Commun. 145, 888-894.

Woods D., (1984) "Oligonucleotide screening of cDNA libraries" Focus 31, 1-3.

Yamamoto M., and Miklos G. L. G., (1978) "Genetic studies on **heterochromatin** in **Drosophila melanogaster** and their implication for the function of satellite DNA" Chromosoma 66, 71-98.

APPENDIX I

LIST OF WORD ABBREVIATIONS

Word	Abbreviation
5-bromo-4-chloro-3-indolyl-b-D-galacto-pyranoside	X-Gal
Ampicillin	Ap, Amp.
Base pair	bp
Bovine serum albumin	BSA
Calf intestinal alkaline phosphatase	CIAP
Cesium chloride	CsCl
Colony forming unit	Cfu
Counts per minute	cpm
Curie	Ci
Density	ρ
Deoxyribonuclease I	DNaseI
Dithiothreitol	DTT
Ethidium bromide	EtBr
Ethylenediamine-tetraacetic acid	EDTA
Fetal bovine serum	FBS
Gram	g
Isopropyl b-D-thiogalactopyranoside	IPTG
Kilo base pair	Kb
liter	l
Luria broth	LB
Megabase	Mb
Molecular weight	MW
Nanometer	nm
Nucleotide	N
Open reading frame	ORF
Phosphate buffered saline	PBS
Plaque-forming units	PFU
DNA polymerase I	pol I
Polynucleotide kinase	PNK
Revolutions per minute	rpm
Ribonuclease A	RNase A
Single stranded DNA	SS-DNA
Sodium dodecyl sulfate	SDS
Tetramethyl-ethylenediamine	TEMED
Ultraviolet	UV
Volt	V
Volume	vol, or v

APPENDIX II

pKS36 COPY NUMBER

---

1 Kb of DS-DNA = 6.6 10^5 daltons

1 dalton = 1.66 10^{-24} g

---

pKS36 = 6.1 Kb => 1 copy of the 1.8 Kb DNA insert = 4 10^6 daltons (1)

0.1µg of pKS36 = 6.02 10^{16} daltons (2)

(1) and (2) => 0.1 µg of pKS36 = 1.505 10^{10} copies of the 1.8 Kb DNA insert

Hybridization signal of 0.1 µg of pKS36 = 3.31 10^7 peak area (relative number recorded by densitometry).

=> 3.31 10^7 peak area = 1.505 10^{10} copies of the 1.8 Kb DNA insert (3)

Hybridization signal of 0.1 µg of **HeLa** = 20.454 10^5 peak area (4)

(3) and (4) => 0.1 µg of **HeLa** = 9.3 10^8 copies of 1.8 Kb satellite DNA (5)

One **HeLa** cell (6 10^6 Kb) contains 3.3 10^6 copies of 1.8 Kb long DNA segments (6)

1 **HeLa** genome = 6.57 10^{-6}µg => 0.1 µg of **HeLa** = 1.52 10^4 genomes (7)

(6) and (7) => 0.1 µg of **HeLa** = 5.016 10^{10} copies of 1.8 Kb long DNA segments = 100 % (8)

(5) and (8) => satellite DNA represent 1.8 % of the **HeLa** genome

---