EFFECT OF ADJACENT SATELLITE DNA ON THE ELECTROPORATION EFFICIENCY AND ON THE STABILITY OF THE TK^+ PHENOTYPE, OF *neo* AND HSV-1 *tk* CONTAINING PLASMIDS, AND DETECTION OF SATELLITE DNA-BINDING PROTEINS

BY

CLAIRE FOUQUET

DEPARTMENT OF MICROBIOLOGY AND IMMUNOLOGY

McGILL UNIVERSITY

MONTREAL, CANADA

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

To all the ones I love,

Some people are trying to reproduce life *in vitro*
While others are trying to kill themselves
Some people are trying to find meaning to things
While others cannot find any meaning to their lives

Some days the scientists say "I found it"
Some days the others say "I found myself"
And there is a great feeling that moves you
Should we never forget it.

Claire, 1991.

ABSTRACT

A 1797 bp human EcoRI satellite II DNA sequence was cloned in vectors containing the thymidine kinase gene (HSV-1 tk) and the neomycin resistance gene, and introduced in a cell line deficient for these genes. We have observed that the electroporation efficiency of these plasmids depends on the location and/or the orientation of the satellite sequences within the transfected plasmid. Only one plasmid, pCFD1 containing one satellite fragment close to the neo gene, inhibited the formation of TK^+/NEO^+ transfectants. We have also shown that the instability of the TK^+ phenotype which was observed did not correlate with the presence of adjacent satellite DNA. In contrast, satellite DNA sequences within the transfected plasmid somehow interfered with the generation of stable TK^+ transfectants.

Moreover, we have detected (both in nuclear and partially purified HeLa whole cell extracts) the presence of proteins that specifically bind the human 1797 bp satellite II DNA sequence. Four proteins with molecular weights of 100, 93, 77 and 34 kDa were identified and named Satellite DNA-binding protein, Sbp-1, -2, -3 and -4, respectively. The function of these proteins is, as yet, unknown.

RESUME

L'ADN satellite II humain EcoRI de 1797 pb a été cloné dans des vecteurs contenant les gènes de la thymidine kinase (HSV-1 tk) et de la résistance à la néomycine qui ont été introduits dans une lignée cellulaire déficiente pour ces deux gènes. Nous avons observé que l'efficacité d'électroporation de ces plasmides dépend de la position et/ou de l'orientation des séquences satellites à l'intérieur du plasmide électroporé. Un seul plasmide, pCFD1 contenant un fragment satellite près du gène neo, inhibe la formation de transfectants TK^+/NEO^+. De plus, la présence d'ADN satellite, adjacent au gène tk, n'est pas responsable de l'instabilité du phénotype TK^+ que nous avons observé. Par contre, en utilisant différentes procédures de sélection, nous avons démontre que les séquences satellites influencent de certaine façon la formation de transfectants TK^+.

D'autre part, nous avons détecté la présence, dans des extraits protéiques nucléaires et totaux cellulaires partiellement purifiés de cellules HeLa, de protéines se liant spécifiquement au fragment satellite humain de 1797 pb. Quatre protéines de poids moléculaire de 100, 93, 77 et 34 kDa ont été identifiées et appelées Sbp-1, -2, -3 et -4 ("Satellite DNA-binding protein") respectivement. La fonction des ces protéines demeure inconnue pour l'instant.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr. Michael S. DuBow for his guidance during the course of my research and for his sometimes hard friendship.

I am much obliged to Dr. Gabriel Drapeau and Dr. Jean-Pierre Simon from Université de Montréal for their faith in me and for helping me to start graduate studies.

I would like to express my gratitude to the students of my laboratory: Chantal Autexier, Robin Cameron, George "Yogi" Kukolj, and Lina Guzzo for their advice and friendship. I also thank Peter Ulycznyj and Josée Brisebois for their precious help in the final bench work sprint.

Thanks to the girls of the hockey team and to the members of the softball team. I will always have memorable souvenirs of these games.

My most gracious thanks to Josée, my great bench neighbor, for the philosophical discussions, the late night coffees, the movies and for her understanding and friendship, and to Caro for the talks in the stairs.

I would especially like to thank Katia Sol and Louis Tremblay for their long lasting frienship and moral support, no matter what.

Special thanks goes to Robert Giasson for his special but sincere help and support.

Finally, I would like to express my gratitude to my parents, Louisette and Marcel, for their support but most importantly for transmitting their intellectual curiosity to me and for giving me self confidence.

This work has been supported by a scholarship provided by the Natural Sciences and Engineering Research Council of Canada.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

CHAPTER 1

CHAPTER 2

LIST OF ABBREVIATIONS

A	adenine
A	amper
bp	base-pair
BrdUrd	bromodeoxyuridine
C	cytosine
CENP	centromere-binding protein
Ci	currie
cpm	counts per minute
DNA	deoxyribonucleic acid
G	guanine
g	gram
HAT	hypoxanthine, aminopterin, thymidine
HMG	high molecular group
kb	kilobase-pair
kDa	kilodalton
LINEs	long interspersed repetitive sequences
mg	milligram
ml	millilitre
mM	millimolar
Msbp	minisatellite-binding protein
μg	microgram
μl	microlitre
μM	micromolar
neo	neomycin resistance gene
P.E.V.	position effect variegation

pmole ..... picomole

Sbp ..... satellite DNA-binding protein

SDS ..... sodium dodecyl sulfate

SINEs ..... short interspered repetitive sequences

stDNA ..... satellite DNA

T ..... thymidine

TFT ..... trifluorothymidine

tk ..... thymidine kinase gene

v/v ..... volume per volume

VNTR ..... variable number of tandem repeats

w/v ..... weight per volume

CHAPTER I

LITERATURE REVIEW

Eukaryotic genomes are known to contain far more DNA than that which is assumed to be required for its function. The DNA content of eukaryotic organisms can be classified as 1- unique, 2- moderately repetitive (10^4-10^5 copies per haploid genome), and 3- highly repetitive (10^6 copies per haploid genome) (Britten and Khone, 1968). The repetitive nature of these sequences was initially shown by two different methods. When total DNA is sheared to fragments of approximately 1 kb in length, then denatured and renatured under appropriate conditions, the speed with which the mixture of single-stranded fragments will renature ($Cot_{1/2}$) depends on how many complementary strands each fragment finds. For most fragments the reaction rate will be slow because the DNA strands must find their complementary strand from the millions generated by cleavage. When eukaryotic DNA was subjected to this process, 70% of the DNA renatured very slowly as expected, but 30% reannealed very quickly, presumably because these sequences were highly repeated in the genome (Britten and Kohne, 1968).

The repetitive nature of this DNA can also be observed, in agarose gels, when total DNA is cleaved with a restriction endonuclease. Somewhat darker bands can be seen over the smeary background of DNA, suggesting that some specific DNA fragments are present in a very large number of copies. Since the discovery of these repetitive DNAs, their structure and organization have been extensively studied. Repetitive DNAs are divided into moderately and highly repetitive classes based on their respective copy number within the genome.

The moderately repetitive sequences (10^4-10^5 copies per haploid

genome) can be divided into two categories, long and short. The short interspersed repeated sequences (SINEs) are typically less then 500 bp in length and are present in approximately 10^5 copies per genome. Based on their sequence, different classes composing the SINEs (e.g.: Alu in primates, B1 in mice) have been defined. They are similar in length and their sequences are related. They are found between genes, within introns, and within satellite DNA (see below for satellite DNA definition). They have a 3' end rich in A residues, are often flanked by direct repeats, are transcribed (Jelinek *et al.*, 1980; Slagel and Deininger, 1989) and are thought to have been mobile (Singer, 1982a).

The long interspersed repetitive sequences (LINEs) are over 5 kb in length and are repeated approximately 10^4 times in the genome. LINEs have been reported to exist in different genera (e.g.:KpnI or L1 in primates, MIF-1 in mice). Subfamilies can be distinguished by their restriction pattern polymorphisms. The abundance of different subfamilies varies from one species to another. LINEs have been reported to be transcribed in humans (Shafit-Zagardo *et al.*, 1983) and in lower primates (Lerman *et al.*, 1983).

Both categories of moderately repetitive sequences reside in the parts of the chromosomes that do not decondense during interphase, namely the heterochromatin. However, the Alu family is 50 fold under-represented in the centromeric heterochromatin (Moyzis *et al.*, 1989). High resolution *in situ* hybridization and quantitative solid state imaging have shown that, in metaphase chromosomes stained with fluorescent dyes, SINEs and LINEs are inversely distributed suggesting an inverse functional relationship for these sequences and that they may also be involved in chromosome structure (Korenberg and Rykowski, 1988). Many functions can be envisaged for these moderately repeated sequences. They could serve as origins of replication, signals for RNA processing, be regulatory sequences for transcription, be structural elements of chromo-

somes or even be a molecular parasite of the genome (Orgel and Crick, 1980; Doolittle and Sapienza, 1980).

Highly repetitive DNA was first called satellite DNA (stDNA) by Kit (1961). At that time stDNA was defined as a minor fraction revealed as a separate band of total cellular DNA in an equilibrium centrifugation cesium chloride density gradient. Satellite DNA has a base composition distinct from the average of the genome which allows it to form a separate fraction by virtue of its different buoyant density.

In the past thirty years, partial sequence data, renaturation kinetics, hybridization experiments, and restriction endonuclease cleavage and Southern blot analyses have allowed extensive investigation of the structure and organization of stDNAs. It became clear that stDNAs were observed as distinct components of nuclear DNA due to their differential organization and base composition (Prosser *et al*, 1981). This highly repetitive DNA is thought not to be heavily interspersed with other sequence types, but are instead tandem repeats of very short nucleotide sequences, or of longer and more complex repeat units (Gall and Atherton, 1974; Southern, 1970, 1975a; Botchan, 1974; Cooke, 1975; Manuelidis, 1976; Skinner, 1977; Maio *et al.*, 1977; Carlson and Brutlag, 1979; Hsiech and Brutlag, 1979a; Frommer *et al.*, 1982). StDNA was found to be present in all higher eukaryotic organisms. It has been found in protozoans (e.g.:trypanosoma [Sloof, 1983]), arthropods (e.g.:crab [Beattie and Skinner, 1972], *Drosophila* [Brutlag, 1980]), echinoderms (e.g.:sea cucumber [Sainz and Cornudella, 1990]), amphibians (e.g.:xenopus [Lam and Caroll, 1983], newt [Epstein *et al.*, 1986]), fish (e.g.:pollock [Denovan and Wright, 1990]), avians (e.g.:crane [Chen *et al.*, 1989], chicken [Kodama *et al.*, 1987], pheasant [Saitoh *et al.*, 1989]), mammals (e.g.:rodent [Kit, 1961; Hatch and Mazrimas, 1974], bovine [Taparowski and Gerbi, 1982], primate [monkey and human, Singer, 1982]). This short

enumeration gives an idea of the broad range distribution of these sequences.

1. HIGHLY REPEATED DNA.

1.1 *Drosophila* Satellite DNAs

As separation techniques were refined, the number of stDNAs detectable in individual species of the *Drosophila* genus increased. In some species one stDNA is revealed (*D. montana*), whereas others contain up to four stDNAs comprising 40-60% of the genome (*D. virilis*) and sometimes an even greater percentage (>60%) of stDNA was present (*D. melanogaster*) (Beridze, 1986 b).

The *Drosophila* stDNAs are named after their buoyant densities. It should be noted that the same buoyant density of stDNAs of different species does not mean that they are identical. According to Lohe and Brutlag (1986) and to John and Miklos (1979), the *D. melanogaster* stDNAs fall into four buoyant density classes: 1.672, 1.686, 1.688 and 1.705 g/ml. However, Beridze (1986) reported a different classification. The major components are 1.698, 1.700, 1.701, 1.703 and 1.715 g/ml, which are divided, based on the buoyant density of the class, into light and moderate stDNA subclasses. Most of the *D. melanogaster* stDNAs are simple satellite DNAs of 5-10 bp organized in extremely homogeneous tandem arrays that can be as long as 10^5 bp (Fry and Brutlag, 1979).

The 1.672 stDNA (AATAT, AATATAT), 1.705 stDNA (AAGAG, AAGAGAG) and the 1.686 stDNA (AATAACATAG and variants) can be represented by the $(AAN)_m(AN)_n$ consensus sequence (Endow *et al.*, 1975).

On the other hand, the 1.688 stDNA is a complex satellite, and considerable variation is found within it. The 1.688 stDNA is mainly composed of tandem repeats of 359 bp (Hsieh and Brutlag, 1979a). Most of the 359 bp units

contain a single cleavage site for HaeIII and HinfI. However some units lack one or both sites due to sequence variation. In addition, some 254 bp units are found within this satellite. The shorter repeat is 80% identical in sequence to the 359 bp monomer unit, but has a 98 bp internal deletion (Carlson and Brutlag, 1979). These characteristics make it quite distinct from the other major satellite DNAs present in *D. melanogaster*, and also distinct from the satellites found in other *Drosophila* species.

It is interesting to note that the amount of each satellite DNA within the *Drosophila* genome can vary from 0.2% to 5.6% (Lohe and Brutlag, 1986). *In situ* hybridization techniques have revealed that the locations of stDNAs are limited to the pericentric heterochromatin of the three autosomes, the proximal heterochromatin of the X chromosome and to the entirely heterochromatic Y chromosome. Each stDNA is therefore localized in specific regions of the chromosome, namely the heterochromatic regions. Individual stDNAs are not restricted to any particular chromosome, but each chromosome is characterized by a distinct amount and localization of each stDNA. For example, 98% of the 1.688 stDNA is found in the X and Y chromosomes, whereas only 5% of the 1.705 stDNA is found in the X chromosome (Peacock *et al.*, 1978).

1.2. Mouse Satellite DNAs

Mouse satellite DNAs can be qualified as both less and more complex than the *Drosophila* stDNAs. They are less complex because there are only two stDNAs in the mouse genome, but are more complex due to their organization.

The mouse major stDNA was first characterized by Southern (1975a). The structure and sequence of the mouse major stDNA was studied in great detail by Horz and Altenburger (1981). The basic repeat, obtained by

cleavage with Sau96I, is 234 bp in length. In addition, small amounts of fragments of 2, 3, 4; 1/2, 1 1/2, 2 1/2, etc.; and 1/4, 1 1/4, 2 1/4 etc. times the length of the unit size fragment were detected suggesting that the 234 bp repeat consists of four internal subrepeats of 58 and 60 bp in length. Sequencing data has shown that each subrepeat could also be divided into two homologous fragments. Therefore, the mouse major stDNA encompasses 8 shorter tandem repeats. Horz and Altenburger (1981) have proposed that these 8 repeats have originated from three similar nonanucleotide sequences (GA_5TGA, GA_6TGA, GA_5CGT) by mutation and amplification. Some variations are found within the major stDNA. For example, only 5-10% of the 234 bp repeat contain a unique TaqI cleavage site; variation in the EcoRI and Sau96I restriction pattern was also reported (Horz and Altenburger, 1981).

Pietras *et al.* (1983) reported the presence of a minor stDNA in mouse. The minor stDNA sequence has a repeat length of 130 bp and is 10-20 times less abundant than the major stDNA in *M. domesticus*, but is 5-10 fold more abundant in the genome of *M. spretus*. The total abundance of stDNA in *M. spetrus* is only 2% of the genome, compared to 10% in *M. domesticus*. Only one third of the minor stDNA sequence shares homology with the major stDNA sequence.

The mouse stDNA is extensively methylated at the internal cytosine of CpG dinucleotides. Comparisons were made between the methylation pattern of mouse major satellite sequences in DNA isolated from germinal and somatic tissues. It was shown that stDNA sequences are more methylated than non-satellite DNA sequences (Gama-Soza *et al.*, 1983). However, satellite DNA isolated from testis or mature sperm are markedly undermethylated when compared to somatic DNA (Ponzetto-Zimmerman and Wolgermuth, 1984). Enzymatic hydrolysis and base analysis by high performance liquid chromatography showed that

mouse stDNA had 30-50% less 5-methylcytosine in sperm than in somatic tissues (Feinstein *et al.*, 1985). It is not yet known if hypomethylation of satellite DNA sequences in the germ line is functionally relevant. Under-methylation of mouse stDNA is a clear case of an opposite condition to the germ line genes, since tissue-specific genes are almost fully methylated in sperm and probably in the female germ line and remain methylated throughout early development and in all non-expressing somatic tissues (Cedar, 1988).

In situ hybridization experiments (Jones, 1970; Pardue and Gall, 1970) have shown that the major stDNA is localized in the centromeric heterochromatin of the 20 mouse chromosomes. However, no stDNA was detected in the Y chromosome. The mouse minor stDNA is specific for the kinetochore region of all chromosomes (Wong and Rattner, 1988).

1.3 Primate Satellite DNAs

1.3.1 African Green Monkey Satellite DNAs

The African green monkey genome contains three stDNAs. The main satellite, the α-stDNA, overwhelms the two other satellites: β and γ. The hidden αstDNA (it cannot be isolated on classical CsCl density gradients) represents 20-25% of the monkey genome, whereas the other satellites comprise less then 3% altogether (Kurnit and Maio, 1974). For this reason, the "minor" satellites were ignored in the characterization studies.

Typically, digestion of total DNA with the restriction endonuclease HindIII produces a 172 bp fragment and its multiple oligomers. A unique nucleotide sequence, representing the most abundant residue at each of the 172 positions, has been determined and was reported in: "Highly Reiterated Sequences of SIMIANSIMIANSIMIANSIMIANSIMIAN" (Rosenberg *et al.*,1978), a title which well depicts the tandem organization of this DNA. The sequence is not internally repetitive, thus making it a complex satellite. There is some level of divergence among the population of fragments, although a single nucleotide residue predominates at each position. Other restriction endonucleases (EcoRI, HaeIII) also cleave the α-satellite DNA, yielding typical ladder patterns indicating clustering of units in defined stDNA segments. Sequences similar to the African green monkey α-stDNA occur in other monkeys (Old and New World) (Singer, 1982b; Beridze, 1986). The α-stDNA is localized at the centromere of all the 60 monkey chromosomes.

The extraordinary fluidity of these DNA sequences led Maresca *et al.* (1983, 1984) to the following question: does genomic DNA flanking the stDNA play a role in the generation of tandem arrays. When studying the junction between the non-satellite and α-stDNA, they found a new satellite DNA which

they called deca-satellite (Maresca *et al.*, 1987). The deca-satellite is 10 bp in length and forms arrays of 1-20 kb long. Even if this satellite DNA is highly polymorphic within an individual, a consensus sequence, 5' AAACCGGNTC 3', was definable. The abundance of different deca-satellite fragments varies between individuals and represents less then 1% of the monkey genome. The deca-satellite is present near the centromere of only 11 to 31 of the 60 monkey chromosomes.

1.3.2 Human Satellite DNAs

The human genome also contains a hidden (cannot be isolated as a separate band on CsCl density gradients) satellite DNA termed the alphoid (or alpha) stDNA owing to its similarities with the African green monkey α-stDNA. The complex alpha-satellite has a monomeric repeat unit of 171 bp (Waye and Willard, 1987). These monomeric units are organized into different tandem arrays that constitute higher-order repeating structures. Alphoid stDNAs make up to 3-5% of the human genome and are distributed at the centromeric regions of each human chromosome (Manuelidis, 1978b; Willard, 1985). To date, at least 33 different alphoid subfamilies have been identified. A new consensus sequence for the human alpha stDNA was derived from 293 individually cloned 171 bp monomers, representing 28 distinct alphoid subfamilies (Choo *et al.*, 1991). From this new consensus sequence it can be seen that a specific nucleotide residue predominates at 156 of 171 bp positions. Choo *et al.* (1991) have surveyed the 33 subfamilies for their chromosomal specificities and ubiquity. Some alpha stDNAs are specific for one chromosome, whereas others are present on more then one but never on more than three chromosomes. For example, the alpha-st pBS4D subfamily is present only on chromosome 2, which also does not contain any other subfamily. However, chromosome 21 has 5 different subfami-

lies that are not specific for chromosome 21, but are also present on chromosomes 13 and 14.

The classical stDNAs isolated from the human genome have been classfied as satellite I, II, III, and IV. Although the picture is complex, some authors have made an effort to clarify the confusion in the classification of these satellites.

Satellite I has been little investigated. It is an A+T rich satellite (>68%), which when cleaved by HinfI yields three fragments of 775, 820, and 875 bp. The 775 bp fragment contains restriction sites for AvaI, Sau96I, EcoRII, Fnu4HI, HinfI, and MboI, whereas the two other fragments are resistant to restriction endonuclease cleavage (Frommer *et al.*, 1982). However, RsaI generates a ladder of tandemly repeated arrays of ABABABA etc., the A and B units being 17 and 25 bp long respectively. The arrangement of the repeats may be extremely complex, so specific domains of this satellite DNA could not be elucidated and the length of the higher-order structure repeats was not determined (Prosser *et al.*, 1986).

Satellite DNAs II and III are both made of simple repeated components, related in sequence but organized differently. HinfI digestion of these satellites generates a ladder of fragments in an agarose gel formed from the 5 bp repeat 5' TTCCA 3'. The 5 bp simple repeat is degenerate in satellite II and is organized into larger repeats irregularly spaced by frequent HinfI and TaqI restriction sites. Satellite III consists of repeating units with the consensus sequence 5' $(TTCCA)_nATTCGGGTTG$ 3', is also cleaved by HinfI, but is resistant to TaqI. The canonical sequence 5' TTCCA 3' is more conserved in satellite III than in satellite II, suggesting that satellite III might have

evolved from satellite II, which have accumulated more point mutations and therefore would be older (Frommer *et al.*, 1982; Prosser *et al.*, 1986). We could depict the relationship between satellite III and II DNAs as follows: from pure satellite III DNA containing no TaqI sites, to those with a high proportion of pure pentamer plus some TaqI sites, to satellite II DNA which contains a low proportion of the 5' TTCCA 3' pentamer and many TaqI sites. The satellite II-III polymorphism seen in individual genomes is therefore attributable to the generation of TaqI sites caused by point mutational events (Fowler *et al.*, 1988). Both satellite DNAs are cleaved with EcoRI and HaeIII to yield fragments of 170 bp and its multiple oligomers, but there is variation in the concentration of each fragment. A 1770 bp satellite II EcoRI fragment has been reported to hybridize to satellite II and III DNAs (Cooke and Hindley, 1979), proving the close relationship of these two satellites. The 1.8 kb fragment seen in EcoRI digests of satellite II DNA (Frommer *et al.*, 1982) and the 1.8 kb fragment generated by KpnI digests of satellite III DNA (Higgins *et al.*, 1985) testify to the higher-order structure organization of the satellite DNAs and to the presence of distinct domains within the satellites. Furthermore, these fragments may be segments of a still longer repeating unit.

The methylation level of at least one family of satellite DNA II-III (EcoRI) has been studied. It is 66% less methylated in placental than in brain tissues (Gama-Sosa *et al.*, 1983).

Satellite IV can be isolated from the shoulder of the main band in a satellite III DNA preparative gradient. Autoradiograph of gels containing labeled fragments from digests of stDNAs III and IV suggested that these two stDNAs could be identical (Frommer *et al.*, 1982).

Altogether, these four satellite DNAs comprise 5% of the human

genome. Attempts have been made to determine the chromosomal specificity of these DNAs by *in situ* hybridization of stDNAs to metaphase chromosome spreads. Unfortunately, low chromosome specificity in the distribution of stDNAs could be detected except for chromosomes 9 and Y (Jones *et al.*, 1973; Godsen *et al.*, 1975). The lack of specificity could be ascribed to cross hybridization between these satellite DNAs and/or lack of hybridization specificity. To overcome this problem, Beauchamp *et al.* (1979) and Choo *et al.* (1990) used rodent-human somatic cell hybrids to investigate the chromosomal location of the stDNAs. Moyzis *et al.* (1987) used fluorescent probes for *in situ* hybridization to metaphase chromosomes. Frommer *et al.* (1988) used bromodeoxyuridine-labeled single-stranded satellite DNA probes, and monoclonal antibodies to BrdUrd, followed by an immunoperoxidase assay to examine the chromosomal location of the satellite DNAs. A probe containing satellite III DNA sequences hybridized to chromosomes 1, 7, 11, 15, 22 and X (Beauchamp *et al.*,1979). A cloned satellite III DNA (pHuR98) specifically hybridized to chromosome 9, and a cloned satellite II DNA (pHuR195) specifically hybridized to chromosome 16 (Moyzis *et al.*, 1987). Choo *et al.* (1990) reported the hybridization of a satellite III DNA variant to chromosomes 14 and 22. Finally, the 100 bp HinfI fragment of satellite III DNA hybridizes to a large number of chromosomes: subcentromeric heterochromatin of chromosome 9, the centromere and short arm of acrocentric chromosomes, the centromere and long arm of the Y chromosome, and the centromeric heterochromatin of chromosomes 1, 16, 17, 20 (Frommer *et al.*, 1988). From these studies, it appears that the different families of satellite DNAs are somewhat chromosome-specific, but widespread among the chromosome complement. Some chromosomes contain very little stDNA or have not yet been reported to contain any. These studies do not conclude anything about their relative abundance. The long-range organization of these sequences has not

been extensively studied. In one case (Choo *et al.*, 1990), using pulse field gel electrophoresis, an EcoRI satellite III DNA was found in large arrays of 150 kb on chromosome 14 and in 20-150 kb arrays on chromosome 22. The long-range organization and the relative abundance of each satellite family may (or may not) be relevant to their function.

A new class of human repetitive DNA, called β-satellite, was recently isolated and characterized (Waye and Willard, 1989). It is unrelated in structure or sequence to previously described satellite DNAs. It is comprised of divergent 68 bp monomer repeat units organized in tandem arrays of 50-300 kb and is localized on chromosomes 9, 13, 14, 15, 21, and 22.

The minisatellite DNAs described by Jeffreys *et al.* (1985) are 9-64 bp in length, usually G+C rich, and organized in 0.1-20 kb arrays that tend to be near the ends of chromosomes. They are also known as VNTRs (variable number tandem repeats) owing to their allelic variation which can be detected by Southern blotting techniques. The very high heterozygosities shown by VNTRs make them valuable genetic markers (Dover, 1990). Jeffreys *et al.* (1985) proposed a common core sequence that is similar to the recombination signal, CHI, of E. coli. According to Jeffreys *et al.*(1985) and Wahls *et al.* (1990), the minisatellites possibly serve as recombination signals and promote their formation. However, a number of other examples allow this hypothesis to be reconsidered. It is not clear if the minisatellites are "hot spots" of recombination or merely subjected to recombination and/or misreplication (Jarman and Wells, 1989; Dover, 1989).

Finally, GATA-GACA repetitive sequences, first isolated in a female snake (BKm sequences), are also present in man (Singh *et al.*, 1986). The accumulation of these repeats has been postulated to be an early step in sex-

chromosome differentiation. This hypothesis was tested in trout and no evidence of a sex-associated pattern could be detected (Lloyds *et al.*, 1989). This finding does not allow one to reject this hypothesis and more studies are needed to speculate on their function in other species. This type of sequence has been shown to mediate a conformational change from right-handed B-DNA to left-handed Z-DNA that could then be recognized by specific proteins, thereby having a biological role (Thomas *et al.*, 1985).

2. SATELLITE DNAs: SIMILARITIES AND DIFFERENCES.

From this brief review, it is clear that the highly repeated DNA sequences share some common characteristics but, at the same time, have peculiar features that differentiate each one.

No specific conformation has been shown for stDNAs, though A-, B-, and Z-DNA conformations have been reported. They are chromosome-specific, or widely distributed on every chromosome. The monomer unit sequences vary a great deal and can be A-T or G-C rich. StDNAs can be simple or complex, the length of each repeat fluctuating from a few base pairs to hundreds of base pairs. In addition, the length of the arrays is also heterogeneous. Furthermore, variation in the amount and kind of stDNA between individuals, species and genera are observed.

However, stDNAs always coincide with the C-bands, predominantly at centromeres, corresponding to constitutive heterochromatin which remains condensed during interphase. Satellite DNAs are more methylated than non-satellite DNAs, but the methylation level is lower in germ line stDNA than in somatic stDNA. In addition, stDNA is replicated late in the cell cycle and is usually not transcribed (see below). Their structure, always repetitive, appears to evolve rapidly and continuously.

These differences and similarities are the landmarks of satellite DNAs. The highly repeated DNAs have been studied for 30 years, yet no function can be assigned to them. The complexity of these sequences may reflect their intricate functions or evolutionary history. Nonetheless, a number of putative functions for these sequences have been put forward to explain their ubiquitous presence in eukaryotic genomes.

3. PROPOSED FUNCTIONS FOR SATELLITE DNA.

3.1 The Selfish DNA Hypothesis.

According to Orgel and Crick (1980) and to Doolittle and Sapienza (1980), repetitive DNA is excess DNA or "junk" DNA that does not confer any selective advantage to the organism that contains it. It was called selfish DNA because its only function would be its survival within the genome. The spread of selfish DNA was compared to the spread of a parasite within its host.

3.2 Satellite DNA Serves an Evolutionary Purpose.

The predominant feature of satellite DNA is its rapid turn-over. This variation could provide rapid perturbation at the cellular, individual and population level, thus making it a molecular drive for evolution that would otherwise be very slow or impossible to achieve through selection of appropriate genes (Mazrimas and Hatch, 1972).

Because some kinds of heterochromatin are known to contain stDNA, and because these sequences are found at the centromere, satellite DNA function could be related to heterochromatin function and/or centromere function.

3.3 Satellite DNA and Higher-Order Chromatin Structure.

StDNAs have been reported to be involved in nucleosome phasing (Musich *et al.*, 1977). Zhang and Horz (1984) studied nucleosomal phasing in mouse. Sixteen nucleosomal frames were detected, but no strict correlation between the stDNA repeats and the nucleosome periodicity could be made. In African green monkey, the specific arrangement of nucleosomal core particles

was reported to be related to the nucleotide sequence of an alpha-satellite DNA subfamily (Wu *et al.*, 1983). Strauss and Varshavsky (1984) have identified a protein (α-protein or HMG-1) that interacts with alphoid-stDNA at specific sites. They propose that α-protein is a phasing protein helping histone octamers to bind. It was also shown that this protein preferentially binds runs of A-T base pairs that are conserved among mammalian species. This suggests that, rather than binding to a few specific DNA sequences, α-protein recognizes a configuration of the minor groove characteristic of short runs of A-T base pairs (Solomom *et al.*, 1986). *In vitro* reconstitution experiments have shown that histones from mouse liver or chicken erythrocytes bind to the DNA in multiple precisely-defined frames in phase with a divergent 9 bp subrepeat of the satellite DNA (Linxweiler and Horz, 1985). However, this sequence dependance could not account for the different situation seen *in vivo*, where two of the frames found *in vitro* either do not exist *in vivo*, or are much less frequent. Therefore, other factors must play a role in nucleosome phasing. Moreover, in the reconstitution experiment, a pBR322 DNA fragment was used, instead of the mouse stDNA fragment, and similar results were obtained. This finding indicated that the association of histone octamers with DNA in specific frames is not unique to mouse stDNA. Finally, this group argued that the histone octamer does not need a phasing protein to know where to bind, as proposed by Strauss and Varsharvky (1984), since only histone proteins were used in the reconstitution experiment (Linxweiler and Horz, 1985).

One of the most thoroughly characterized satellite DNA-associated proteins is the chromosomal protein D1 of *D. melanogaster*, as it has properties resembling those of the HMG-1 proteins. Protein D1 is preferentially associated with the A-T rich 1.672 and 1.688 g/ml satellite DNAs. In contrast, it is not associated with the 1.705 g/ml satellite DNA which is also A-T rich

(Levinger and Varshavsky, 1982; Ashley *et al.*, 1989). The biological role of D1 remains to be clarified.

An additional report (Radic *et al.*, 1987) shows that the sequence of mouse satellite DNA exhibits a stable curvature. When the bend is alleviated by drug treatment, the condensation of centromeric heterochromatin is reduced. Other satellite DNAs have also been shown to be bent (e.g. meal worm stDNA [Plohl *et al.*,1990]). This bending may be helpful for specific positioning of DNA in nucleosomes.

These findings strongly suggest that the structure of the satellite DNA is more important than its actual sequence to achieve compaction (and transcriptional inactivity) of these heterochromatic sequences and may explain their presence with concomitant large sequence divergence.

3.4 Satellite DNA and Anchorage to the Nuclear Matrix.

Eukaryotic DNA is organized into supercoiled loops of 30-100 kb in interphase nuclei. It has been shown that these loops are attached to the nuclear matrix via protein-DNA interactions mediated by stDNA in mouse. Nuclear matrices were isolated from Friend cells, treated with lysis buffer and the DNA deproteinated by phenol leaving on the DNA only the proteins that were tightly bound to it. The constituents of the complexes were analyzed. Seven proteins (5 non-histone and 2 core histone proteins) were found to be associated to mouse stDNA and are mainly localized at the base of the loops (Patriotis and Djondjurov, 1989).

3.5 Satellite DNA and Centromere Structure.

The centromere, composed of three structural domains: the three laminar kinetochores, the central domain and the pairing domain, interacts with spindle microtubules and chromatin fibers (Rattner, 1991). It has been difficult to identify specific DNA sequences that map to the kinetochore domain. In one case, the mouse minor stDNA was shown to be localized to the outer lateral surface of the centromere (Wang and Rattner, 1988) and might be involved in centromere-kinetochore interactions.

Several specific centromere proteins (CENP-A [17 kDa], -B [80 kDa], -C [140 kDa], -D [50 kDa]) have been identified by autoimmune sera from patients with the CREST syndrome (Earnshaw and Cooke, 1989). Of these proteins, CENP-B was shown to bind specifically to the monomer unit of α-stDNA in human cells (Masumoto *et al.*, 1989). Moreover, a correlation between the amount of α-stDNA and CENP-B antigens, present at the centromere central domain, could be made using a deleted chromosome that was shown to contain only 20-30% of the normal α-stDNA sequences (Werick *et al.*, 1990). This mini-chromosome behaved normally in both mitosis and meiosis. Thus, it may be premature to conclude that satellite DNAs-CENP-B interactions have a role to play in centromere structure and/or function.

3.6 Satellite DNA and Centromere Function.

D. melanogaster has specific arrangements of different stDNAs in the centromeric heterochromatin of each chromosome. Yamamoto and Miklos (1977, 1978) proposed that stDNAs are involved in chromosome (homologs) pairing. This hypothesis proved to be wrong. Most of the heterochromatin can be deleted without affecting pairing and segregation

of the two X homologs. This hypothesis is also incorrect in males which have two different sex chromosomes. Moreover, heterochromatin was exchanged between different autosomes, by illegitimate recombination, yet pairing and segregation of these autosomes was still normal (Yamamoto and Miklos, 1977).

When mouse metaphase chromosomes were digested with restriction endonucleases that degrade the bulk of stDNA, contact between sister chromatids appeared to be disrupted and the inter-kinetochore distances were increased. Therefore, it was suggested that satellite DNA, in the centromeric region, was necessary to maintain sister chromatids together during mitosis (Lica *et al.*, 1986). However, the function of the centromere was not assessed. Furthermore, this treatment can alter other factors such as centromere-binding proteins.

A human minichromosome, comprised of the centromere of chromosome 1 that had lost centromeric heterochromatin (st-III DNA, α-st DNA, EcoR1 dimer), when present in human-hamster hybrid cells was found to segregate normally in mitosis (Carine *et al.*, 1989). Furthermore, some human chromosomes apparently lack stDNA sequences (besides α-st), yet these chromosomes pair and segregate normally. On the other hand, the tendency of acrocentric chromosomes to associate with other acrocentrics in meiosis was correlated with the amount of satellite DNA carried by these chromosomes (Guichaoua *et al.*,1986; Dumont *et al.*, 1989). Thus, it cannot be concluded that chromosome pairing and segregation depend only on satellite sequences.

3.7 Satellite DNA Transcription.

The stDNAs were at first thought not to be transcribed due to their localization in heterochromatin. In support of this concept,

several studies have failed to detect cellular RNA complementary to satellite sequences (Hsu, 1962; Flamm *et al.*, 1969; Reeder, 1973). However, there is controversial evidence reporting repetitive DNA transcription in mouse (Gaubatz and Cutler, 1990; Cohen *et al.*, 1973, Harel *et al.*, 1968), in primates (Lerman *et al.*, 1983; Shafit-Zagardo *et al.*, 1983), and in non-mammalian species (Stephenson *et al.*, 1981; Diaz *et al.*, 1981; Epstein *et al.*, 1986,1987; Bonaccorsi *et al.*, 1990). It is not clear if satellite DNA transcription is merely an accidental phenomenon due to failure of transcription termination of upstream genes, or whether it has some biological function (Epstein *et al.*, 1986). In one case, stDNA transcription was reported to be tissue and age-specific. The mechanisms that allow transcription to occur are not known, but it appears that these transcripts represent a gene regulatory dysfunction possibly leading to the aging process in mouse cardiac muscle (Gaubatz and Cutler, 1990). As changes in the methylation pattern of cytosine residues in specific DNA sequences have been correlated with transcriptional activity during cellular differentiation and development, it is possible that the extent of stDNA methylation affects interaction with specific proteins which in turn affect transcription of these sequences. In *Drosophila*, satellite DNA is transcribed on Y chromosome loops during spermatogenesis. The transcripts could serve a protein-binding function rather than a coding function (Bonaccorsi *et al.*, 1990). The gamete-associated hypomethylation of specific stDNA sequences (discussed earlier) suggests a differential role for stDNA sequences in germ line as compared to somatic tissues. These roles, whatever they are, remain to be demonstrated.

3.8 Satellite DNA and Recombination.

Owing to its homology and multiple sites of location, stDNA can increase recombination in adjacent DNA sequences through unequal

sister chromatid exhange, DNA amplification, and slip-strand mispairing, to name but a few examples.

In *Drosophila*, it has been observed that recombination of meiotic X chromosomes is reduced when the heterochromatin, and therefore satellite DNA content, is systematically deleted. Thus, heterochromatin length, and not its composition, would affect recombination (Yamamoto and Miklos, 1978).

Chromosome rearrangements and breakage appears to be due to adjacent satellite sequences. An increase in sister chromatid exchange and chromosome rearrangement was seen in hamster cells after amplification, via the use of the dhfr gene, of human α-stDNA transfected into these cells. Even if the increase in chromosome length and in the population doubling time could account (at least in part) for these effects, and if the methotrexate treatment is by itself perturbing the DNA, the authors conclude that the alphoid DNA is implicated in an active process of chromosome breakage and is a "hot spot" for DNA rearrangements (Heartlein *et al.*, 1988).

Along the same line, Butner and Lo (1986b) report the rearrangement of a transfected plasmid integrated in mouse centromeric heterochromatin. Upon further investigation (Chaterjee and Lo, 1989), the DNA rearrangements were shown to be intrachromosomal recombination and chromosome breakage. The authors conclude that stDNA was responsible for these effects. It should be pointed out here that people tend to confound satellite DNA and heterochromatic DNA. StDNA is found in heterochromatin but heterochromatic DNA is not comprised solely of satellite DNA. Therefore it is not clear if the effect on recombination is merely due to satellite sequences, or if it is a more general effect of heterochromatic DNA.

The X chromosome inactivation in female mammals might be an

example of a more general effect of heterochromatin. In females, one of the X chromosomes remains condensed throughout interphase. The inactivation of one X chromosome ensures equal dosage of gene products encoded by the X chromosome in male and female. The mechanism by which thousands of genes on only one pair of identical chromosomes are turned off is poorly understood. Both in man and mouse, genes exclusively expressed from the inactive X chromosome (XIST and Xist) have been identified. In man this gene contains an open reading frame encoding a putative protein that could be membrane bound. A 24 bp motif repeated 8 times is present at the 5' end of the gene. However, screening of sequence data banks did not reveal homology to any other sequence (Borsani *et al.*, 1991).

As mentioned before, minisatellite (VNTRs) core DNA sequences can possibly serve as a recombination signal (Jeffreys *et al.*, 1985). Wahls *et al.*, (1990) showed that hypervariable minisatellite DNAs stimulate intermolecular recombination up to 13 fold in human cells. A minisatellite DNA-binding protein (Msbp-1) was detected in mouse tissue. This protein was shown to interact tightly and specifically with a repeat sequence related to the minisatellite core sequence. Recently, 2 novel minisatellite DNA-binding proteins, Msbp-2 and -3, were purified from HeLa cell nuclear extracts and shown to specifically bind tandem repeats of the minisatellite core sequence. It is possible that these proteins are involved in recombination near minisatellite DNAs (Wahls *et al.*, 1991). However, the biological role of minisatellite DNA-binding proteins are not known (Collick and Jeffreys, 1990; Wahls *et al.*, 1991). If the recombination signal hypothesis is true, it could have implications in the generation of genetic disorders. In steroid sulfatase deficiency, it has been shown that the deletion involves recombination

between low copy repetitive elements located 1.9 megabase pairs apart on the X chromosome. It is not known if the recombination between these elements is only due to the homology of these sequences, or if it is specifically promoted by VNTR sequences (Yen *et al.*, 1990). Similarly, recombination between Alu sequences is responsible for the duplication observed in the low density lipoprotein (LDL) receptor gene (Lerhman *et al.*, 1987).

It is clear that repetitive DNA sequences can be involved in the generation of genetic disorders. What is not clear is if it is a consequence of the fluidity of the repetitive DNAs, or if these sequences actually promote these effects.

3.9 Satellite DNA and Position Effect Variegation.

Position effect variegation (P.E.V.) results from chromosome rearrangements where normal genes, having been rearranged next to heterochromatin, are inactivated in some cells but not in others, thereby producing a variegated phenotype. P.E.V. demonstrate that unusual structural and functional properties can be propagated over considerable distances along the chromosome by the formation of new heterochromatic-euchromatic junctions (Henikoff, 1990). This phenomenon was explained by propagation of the compact heterochromatin structure into the neighboring euchromatic sequences, thereby inactivating some genetic loci (Tartof *et al.*, 1989). Genetic elements can modify, by enhancing or suppressing, P.E.V.. For example, the Suvar (3)7 gene encodes a protein with 5 zinc fingers that would serve for the packaging of the chromatin fiber into heterochromatin, therefore controlling the expression from DNA domains (Reuter *et al.*, 1990). However, the interaction of this protein with DNA has not been demonstrated. The question

that arises here is the involvement of stDNA in P.E.V. since it is located in the heterochromatin. For example, for the *white* locus in *Drosophila*, the euchromatic breakpoints have been shown not to involve stDNA sequences, but sequences (called 1.688h) that share 70% homology with the 1.688 g/ml stDNA of *Drosophila*. The authors speculate that the X chromosome is folded in such a way that the 1.688 (located in heterochromatin) and 1.688h (located in euchromatin) sequences are in close proximity to each other (Tartof *et al.*, 1984). Interaction between the two related sequences may be at the nucleotide level or mediated by a protein shown to bind to the 1.688 stDNA (Hsieh and Brutlag, 1979b). StDNA would serve to bring the *white* locus into the heterochromatin since these sequences are not part of the new euchromatic-heterochromatic junction (Tartof *et al.*, 1989). However, stDNA may be involved in the formation of heterochromatic domains that would control variegation.

P.E.V.-like effects have been studied in mammalian systems. Reversion from a thymidine kinase$^+$ (TK^+) to TK^- phenotype was associated with changes in DNA methylation (Ostrander *et al.*, 1982; Christy and Scangos, 1984). A mouse transformant cell line, containing a tk gene integrated in the pericentromeric heterochromatin, showed unstable expression of the tk gene. The derepression of the tk gene was accompanied by rearrangements in the flanking mouse stDNA suggesting that the flanking stDNA may exert *cis* effects on tk gene expression (Butner and Lo, 1986 a, b). The authors suggest that the instability of the cell line may reflect the instability of the flanking stDNA.

This question was more directly addressed by Talarico *et al.* (1988) by cotransfecting mouse stDNA and the tk gene. The modulation of the tk gene, cointegrated with mouse stDNA in heterochromatin, was shown not to be due to DNA rearrangements or a change in DNA methyla-

tion. They proposed that phenotypic switching is the result of two effects: 1- the tendency of the tk promoter to maintain a transcriptionally active chromatin state and 2- the satellite DNA favoring a condensed chromatin structure, possibly by binding some specific proteins. The combination of these two effects results in an unstable chromatin conformation giving rise to phenotypic switching.

Despite all the studies on satellite DNA function, our understanding of the molecular function of satellite DNA is still vague. This is mainly due to the lack of experimental systems that would allow direct examination of the properties of satellite DNAs *in situ*. The repetitive nature of this DNA hinders the tracking of one subset of satellite repeats among all similar sequences. To circumvent this pro-blem, one could introduce satellite DNA sequences, along with a target gene, in a mammalian cell system. The ability to transfer genes provides an opportunity to study the function and physical state of exogenous DNA in the transformed host.

To elucidate some of the stDNA functions, we have transfected a shuttle vector, containing the tk gene (target gene), the neo gene (selectable marker gene) and a cloned human satellite II DNA, into a human thymidine deficient cell line and studied the effect of the adjacent stDNA on plasmid integration and on the phenotype stability. In addition, we searched for the presence of satellite DNA-binding proteins, in human cell extracts, that could mediate satellite DNA function.

4. The Mammalian Cell System.

4.1. The Shuttle Vector.

Shuttle vectors were developed to introduce genes in mammalian systems (Mulligan and Berg, 1980; Southern and Berg,1982). These vectors are capable of replicating and expressing genes in both bacterial and mammalian cells. Shuttle vectors, containing an SV40 origin of replication and transformed into mammalian cells that do not express large T antigen, do not replicate autonomously. Hence, cells that have integrated the foreign DNA into their chromosomes can be selected for (Robins *et al.*, 1981; Pelicer *et al.*,1978). The shuttle vector used in this study (pSV_2neoKT, Goring *et al.*, 1985) contains the thymidine kinase gene (tk), the neomycin resistance gene (neo), and human satellite II DNA sequences.

4.2 The Viral Thymidine Kinase Gene.

The viral thymidine kinase gene from herpes simplex virus type 1 (HSV-1) was chosen as a target gene for a variety of reasons. This gene is well characterized and has been completely sequenced (McKnight, 1980). A number of thymidine kinase deficient cell lines are available as a recipient cell line. Either the TK^+ or TK^- phenotype can be selected under appropriate growth conditions, making it a good target gene to study the stability of gene expression. The TK^+ phenotype is selected by using HAT (hypoxanthine, aminopterin, thymidine) medium in which the pyrimidine salvage pathway enzyme, thymidine kinase, is necessary for survival. The TK^- phenotype is selected with the nucleoside analogue trifluorothymidine (TFT) that is phosphorylated by the viral TK enzyme and leads to TK^+ cell death by blocking chain elongation after

incorporation into the cellular DNA.

4.3 The Neomycin Resistance Gene.

The neomycin resistance gene, encoding resistance for the aminoglycosides G418 and kanamycin, was isolated from the Tn5 transposable element (Berg *et al.*,1978; Jorgensen *et al.*,1979). The neo gene, adjacent to the tk gene, under the control of the SV40 early promoter, is used to screen out gross chromosomal rearrangements by maintaining selective pressure with G418 for a functional neo gene, while the selection for the tk gene is removed.

4.4 The Human EcoRI Satellite DNA.

A human 1797 bp EcoR1 satellite II DNA fragment was previously isolated from the genome of HeLa cells, cloned and sequenced. It contains 35% pure pentamer 5' TTCCA 3' repeats and closely related sequences. The 1.8 kb EcoR1 DNA fragment also contains a 49 bp segment devoid of satellite-like sequences. This satellite DNA fragment represents less than 1% of the HeLa genome and is mainly orgarized as 1.65, 1.95, and 3.6 kb EcoR1 fragments. In HeLa cells, this satellite DNA is clustered in 150 to 500 kb arrays residing mainly on chromosomes 7, 12, 14, 15, 16, and 22 (Sol *et al.*, 1986; Sol, 1991).

5. Gene Expression in Mammalian Cells.

The expression of foreign genes introduced into mammalian cells can be influenced by the site of DNA integration and DNA methylation.

5.1 Integration Site of Foreign DNA.

As mentioned before, satellite DNA could be involved in the regulation of gene expression. Therefore, the integration of foreign DNA into host repetitive DNA sequences could have an impact on gene expression. The isolation of the junction site between the host and foreign DNA is a direct method to characterize these sequences. Few of these junctions have been characterized, and the significance of these results have to be interpreted with circumspection. Nevertheless, analysis of integrated DNA containing viral sequences, by polymerase chain reaction, showed that it is not a site-specific process and is not chromosome- specific (Murnane *et al.*, 1990). Integration of non-viral DNA sequences (no SV40 sequences involved) was reported to occur preferentially at repetitive DNA sequences of the host (Kato *et al.*, 1986). Junctions between host DNA and foreign DNA containing rat repetitive DNA were shown not to be repetitive DNA or to involve specific host sequences (Wallenburg *et al.*, 1987).

In situ hybridization experiments reporting integration in centromeric heterochromatin (Butner and Lo, 1986a; Talarico *et al.*,1988) are only circumstantial evidence of integration into repetitive DNA. On the other hand, alterations in the chromatin structure have been correlated with changes in gene expression. Unstable tk expression was correlated with differential DNAseI sensitivity reflecting changes in the chromatin structure (Sweet *et al.*,1982; Davies *et al.*, 1982). It is possible that the foreign DNA acquires properties of the host sequences surrounding them.

5.2 DNA Methylation.

There is a tendency to believe that there is an inverse correlation between the transcriptional activity of a DNA sequence and the presence of CpG dinucleotide methylation. Housekeeping genes (e.g.: aprt and dhfr) are undermethylated in hamster and mouse tissue (Stein *et al.*, 1983). On the other hand, the vitellogenin gene in *Xenopus* (Gerber-Huber *et al.*, 1983), the α2(1)-collagen gene in chicken (McKeon *et al.*,1982) and the chicken lysozyme gene (Wolfl *et al.*, 1990) were found to be actively transcribed when methylated.

Several studies of unstable TK^+ cell lines have shown that phenotypic switching is correlated with methylation (Ostrander *et al.*, 1982; Christy and Scangos, 1982). In contrast, it has been shown that methylated tk promoters can be active and inactivation of the gene occur via methylation of a specific DNA sequence (at the EcoRI site) in the promoter (Tasseron-de Jong *et al.*, 1989). Therefore, one has to keep in mind that gene inactivation requires methylation of specific DNA sequences.

OUTLINE OF THE THESIS

Though different functions have been ascribed to stDNAs, clear evidence of their function within the genome was never obtained. This is mainly due to the fact that stDNA is found in the heterochromatin. Therefore, when studying stDNA function, one could in fact be looking at the heterochromatin function. It is possible that these two entities are the same but, they could also be different and have distinct functions.

In order to investigate stDNA function outside its native location, we have devised an experimental system in which some of the putative stDNA functions can be assessed. We have also sought to determine whether a cloned stDNA contain specific recognition signals for DNA-binding proteins that could mediate their function.

This thesis reports the results obtained while conducting this research.

Chapter II describes the effect of an adjacent stDNA on the modulation of gene expression and its effect on stDNA-containing plasmid electroporation efficiency.

Chapter III reports the detection of specific stDNA-binding proteins in HeLa cell extracts.

The experimental procedures utilized throughout these two analyses are reported in their respective chapters.

Chapter IV summarizes the results obtained. Concluding remarks and directions for future research are also included in this chapter.

CHAPTER II

EFFECT OF ADJACENT SATELLITE DNA ON THE ELECTROPORATION EFFICIENCY AND ON THE STABILITY OF THE TK$^+$ PHENOTYPE, OF *neo* and HSV-1 TK CONTAINING PLASMID.

A modified version of this chapiter has been submitted for publication to Molecular and Cellular Biology

(Claire Fouquet and Michael S. DuBow)

ABSTRACT

Satellite DNAs are tandem repeats of simple sequences (5 bp) organized in long arrays which constitute 5% of the human genome and have not, yet been associated with any phenotypic function. To study the effect of satellite DNA on the expression of adjacent genes, plasmids containing a 1797 bp EcoRI satellite DNA, plus the herpesvirus type 1 thymidine kinase gene (HSV-1 tk) and the neomycin resistance gene (neo), were electroporated into TK^- /NEO^- cell line. The presence of the satellite DNA within the plasmid was found to affect the electroporation efficiency in a location and/or orientation-dependent manner. Selection for either stable tk or neo integration and expression showed that the satellite DNA-containing plasmids integrated in a manner that led to inactivation of the tk gene sequences or interfered with its expression more frequently than the neo gene. The TK^+ phenotype was found to be highly unstable and could not be correlated with the electroporated satellite DNA sequences. Additional studies with this system may provide insights into satellite DNA function within the human genome.

INTRODUCTION

Eukaryotic genomes contain a large amount of repetitive DNA (Singer, 1982; Miklos, 1985) which, in humans, accounts for 20-30% of the genomic sequences (Marx *et al.*, 1976). Among these repetitive DNAs is a class of short, simple repeat units called satellite DNAs (stDNAs) which are organized in long tandem arrays and are generally found in the constitutive heterochromatin, predominantly at the centromeres (Kurnit, 1979; Brutlag, 1980). Satellite DNA sequences are replicated late in the cell cycle (Selig *et al.*, 1988) and are usually not transcribed (Redder *et al.*, 1973). StDNAs have been studied for many years and, while various functions have been proposed for these enigmatic sequences

(Mazrimas and Hatch, 1972; Yamamoto and Miklos, 1978; Orgel and Crick, 1980; Radic *et al.*, 1987; Heartlein *et al.*, 1988; Chaterjee and Lo, 1989; Werick *et al.*, 1990; Gaubatz and Cutler, 1990; Wahls *et al.*, 1990), their precise role in the genome remains unknown. Sequence divergence among the three human simple stDNAs (II, III and IV) can be quite extensive (Fowler *et al.*, 1988; Prosser *et al.*, 1986) and might reflect their evolution and/or function.

Modulation of gene expression was reported to occur in a variety of ways (Pelicer *et al.*, 1978; Sweet *et al.*, 1981; Christy and Scangos, 1982, 1984; Davies *et al.*, 1982; Ostrander *et al.*, 1982; Hardies *et al.*, 1983; Gebara *et al.*, 1987; Gelbert *et al.*, 1990). For example, in *Drosophila*, position effect variegation results from chromosome rearrangements where cellular genes, having been moved next to heterochromatin, are inactivated in some cells but not in others. This produces a variegated phenotype (Tartof *et al.*, 1989) and implies a *cis*-acting effect of heterochromatin and possibly stDNA sequences. Since transfected DNA integrates almost randomly intc the host genome (Wallenburg *et al.*, 1987; Murnane *et al.*, 1990), integration of foreign DNA can occur in the host heterochromatin, hence in stDNA sequences, where expression of the transfected genes may be affected (Butner and Lo, 1986a; Kato *et al.*, 1986; Talarico *et al.*, 1988).

The ability to transfect exogenous genes into mammalian cells has greatly aided the study of the stability of gene expression in mammalian cell systems (Wigler *et al.*, 1977; Pelicer *et al.*, 1978; Robins *et al.*, 1981; Scangos *et al.*, 1981). To study the *cis*-effect of stDNA sequences on plasmid DNA electroporation efficiency and stability of gene expression, we constructed a series of human fibroblast cell lines by electroporating plasmids containing a cloned and sequenced 1.8 kb human satellite II DNA (Sol *et al.*, 1986) adjacent to the neo and HSV-1 tk genes on the plasmid pSV_2neoKT (Goring and DuBow, 1985).

Following electroporation, the tk and/or neo genes were selected for and the number of transfectants was counted. We report here that stDNA sequences can affect the electroporation efficiency depending on the location and/or orientation of the stDNA relative to the two plasmid genes (tk and neo) selected for. However, the high instability of the TK phenotype could not be ascribed to the electroporated stDNA sequences.

MATERIALS AND METHODS

Cell Lines:

All cell lines used in this study are derived from the thymidine kinase deficient human transformed fibroblastic cell line 143B (ATCC CRL 8303). The KT cell line, used as a positive control, was derived from the 143B cell line and contains a single copy integrated pSV_2neoKT plasmid as described elsewhere (Goring and DuBow, 1985). All cell lines were grown in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% fetal bovine serum (FBS, Gibco), 2.5 ug/ml amphotericin B (Fungizone, Squibb), 10 U/ml penicillin and 100 μg/ml streptomycin (Pen-Strep, Gibco) at 37° C with a 5% CO_2 atmosphere. Electroporated cell lines were maintained in DMEM plus HAT (0.1 mM hypoxanthine, 0.4 μM aminopterin, 16 μM thymidine [Sigma]) and/or G418 (400 μg/ml, Sigma). The TK^- cell lines were selected and maintained in DMEM supplemented with trifluorothymidine (TFT) (0.75 μg/ml, Sigma) plus G418. Cell lines were found to be mycoplasma-free using the MycoTect detection kit (Gibco) and the Mycoplasma detection kit (Boerhinger-Mannheim).

DNA Manipulations and Cloning:

All restriction endonucleases were purchased from Bethesda Research Laboratories Inc.(BRL), New England Biolabs (NEB), Pharmacia

Canada, or Boehringer-Mannheim Canada and DNA was hydrolyzed according to the manufacturer's instructions. The 5' phosphate groups were removed with calf intestinal alkaline phosphatase (CIAP, DuPont Canada) and the recessed 3' termini were filled in with Klenow enzyme (BRL) (Sambrook *et al.*, 1989). DNA fragments were purified from agarose gels using the Gene Clean kit (Bio101) according to the manufacturer's instructions. Ligations were performed in LRB buffer with T4 DNA ligase (Pharmacia) (King and Blakesly, 1986).

The pSVL plasmid was obtained by disruption of the BamHI site 3' to the tk gene in the pSV_2neoKT plasmid (Goring and DuBow, 1985) via partial BamHI cleavage, backfilling of the 3' recessed ends and religation of the plasmid. Plasmids pCFU1 and pCFU2 were obtained by cloning a blunt 1797 bp EcoRI satellite II DNA fragment (Sol *et al.*, 1986) into the unique, backfilled, BamHI site of pSVL. Plasmid pCFD1 was obtained by cloning the blunt EcoRI satellite DNA fragment into pSVL previously linearized at the unique blunt Stu1 site. Plasmids pCF1.8 and pCF3.8 were constructed by cloning a 1.8 or 3.8 kb bacteriophage λ NdeI fragments into the NdeI restriction site of pSVL. The ligation mixtures were transformed into *Escherichia coli* strain DH1 [(F^-, gyrA1, endA1, thi, hsdR17, supE44, recA1) (Hanahan, 1983)] via the calcium chloride method of Mandel and Higa (1970).

Large scale preparations of plasmid DNA were obtained by the alkaline lysis method (Sambrook *et al.*, 1989) and purified by centrifugation on cesium chloride/ethidium bromide density gradients (Maniatis *et al.*, 1982). High molecular weight cellular DNA was prepared according to Blin and Stafford (1976).

Electroporation of Plasmid DNA into the 143B Cell Line:

Transfections were performed by electroporation according to Chu *et al.* (1987) using a Bio-Rad Gene Pulser apparatus set at 250 volts and 960 μFarads, except that the 10 minute incubations were performed at room temperature. Cells (3×10^6) were electroporated with 1.5 picomoles (pmole) of circular DNA, or plasmid DNA linearized at the NdeI site (pSVL, pCFU1, pCFU2, pCFD1) or at the HindIII site (pCF1.8, pCF3.8), then plated in DMEM. Two days after electroporation, selection for the neo and/or tk genes was applied by changing the medium for DMEM plus HAT and/or G418. After 14 days, the colonies were stained with Giemsa (Merchant *et al.*, 1964) and counted. Clones were randomly chosen and expanded. The new cell lines were named according to the form of the DNA (circular or linear), the plasmid electroporated, and the clone number (e.g.: CSV-6, LD1-5).

Reversion Analysis:

The TK^+/NEO^+ cell lines were grown in DMEM medium for 3 or 7 days. Cells (10^3) were then plated in (1) DMEM/HAT/G418, (2) DMEM/TFT/G418 or (3) DMEM medium. After 2 weeks, the colonies obtained from each selection were stained (Merchant *et al.*, 1964) and counted. The plating efficiencies were obtained by calculating the ratio of colony numbers in TFT/G418 or HAT/G418 to that in DMEM.

RESULTS

Plasmid Characterization:

The satellite DNA-containing plasmids (pCFU1, pCFU2, and pCFD1) used in this study were derived from the plasmid pSVL (Fig. 1A). Plasmid pCFU1 (11.1 kb) contains a single 1.8 kb stDNA insert. The 5' end (base pair 1)

of the stDNA (Sol *et al.*, 1986) is linked to the 5' end (base pair 1) of the tk gene (McKnight, 1980) (Fig. 1B). Plasmid pCFU2 (12.9 kb) contains two of these stDNA segments both of which are in the same relative orientation, with respect to the tk gene, as in pCFU1 (Fig. 1C). Plasmid pCFD1 (11.1 kb) contains one stDNA fragment. Compared to pCFU1 and pCFU2, the stDNA segment in pCFD1 is in the reverse orientation (Fig. 1D). Control plasmids of approximately the same sizes, but which do not contain stDNA sequences (pCF1.8 and pCF3.8), were obtained by cloning bacteriophage λ DNA fragments into pSVL as described in Materials and Methods. Plasmid pCF1.8 is the same size as plasmids pCFU1 and pCFD1, while pCF3.8 is 0.2 kb larger than pCFU2 (Fig. 1E and F).

Effect of Satellite DNA on Electroporation Efficiencies:

Human TK^- cells were electroporated with plasmid DNAs containing or lacking stDNA sequences, and were selected for the presence of the tk and neo genes. The averaged transfection efficiencies (from 3 transfections) obtained with pSVL, pCFU1, pCFU2, and pCFD1 are reported in Table 1. As expected (Toneguzzo *et al.*, 1986), the efficiencies were 2.6 to 13.4 fold higher for linear plasmids as compared to their circular forms. The presence of stDNA sequences within the transfected plasmids pCFU1 and pCFU2 decreased the electroporation efficiencies to 56% and 11% (circular forms), and to 70% and 54% (linear forms) to that observed for the parent plasmid (Table 1).

To determine if the decrease in the number of transfectants obtained was specific for stDNA or if it was due to a plasmid size effect, plasmids pCF1.8 and pCF3.8, which are the same size as the stDNA-containing plasmids but do not contain stDNA, were also electroporated (Table 1, A). These plasmids show a decrease in electroporation efficiency of 2.7 fold (pCF1.8) and 11 fold (pCF3.8) which is comparable to that observed with the stDNA-containing

plasmids. Therefore, the decrease in electroporating ability of pCFU1 and pCFU2 can be correlated with an increase in plasmid size, even when the same molar ratios were used.

Plasmid pCFD1, which contains one stDNA close to the neo gene, behaved differently from the other stDNA-containing plasmids. The circular form of pCFD1 always failed to give stable transfectants when both genes were selected for (Table 1). The electroporation efficiency obtained with linear pCFD1 was markedly decreased compared to that observed for the stDNA-deficient plasmid pSVL.

Effect of Satellite DNA on Plasmid Gene Expression:

To determine if the satellite DNA within the electroporated plasmid could affect the stable expression of plasmid genes, both single and stepwise selections were performed. Plasmid DNA was electroporated, and the cells were selected for a functional tk (HAT^R) or neo ($G418^R$) gene, or for the presence of a functional neo gene followed by selection for both a functional tk and neo gene (HAT^R and $G418^R$). The number of transfectants obtained in the stepwise (G418, then G418 plus HAT) selection represents transfectants in which at least one tk and one neo gene is functional. The electroporation efficiencies obtained in HAT only (single selection) represent the number of transfectants in which at least one tk gene is functional, whereas the electroporation efficiencies in G418 (single selection) correspond to cells in which at least one neo gene is functional (Table 2). The electroporation efficiencies obtained, in both single selections, after electroporation of circular and linear pSVL plasmids were similar. However, the presence of stDNA, in *cis*, in the electroporated circular plasmids interfered with the generation of TK^+ cells, irrespective of the stDNA copy number or location, since the efficiencies in HAT are much lower than in G418 medium. When plas-

mids were linearized at the NdeI site close to the neo gene in all cases and close to the stDNA in the case of pCFD1, similar electroporation efficiencies, in either HAT or G418 selections, were obtained (Table 2). The effect of stDNA on electroporation efficiencies seen in the single and the stepwise selections (selecting for neo, then for neo and tk) (Table 2) were different from those observed in the double selection (selecting for neo plus tk at the same time)(Table 1). First, circular and linear pCFU1 plasmids had increased or equal efficiencies, compared to those observed for pSVL, in 5 out of 6 selection experiments. Second, electroporation of circular pCFD1 and subsequent G418 single selection gave rise to a large number of transfectants. Finally, the stepwise selection applied after the electroporation of circular pCFD1 yielded NEO^+/TK^+ transfectants that could never be obtained when the selection for both genes was applied at the same time.

Stability of TK^+ phenotype:

To evaluate the effect of adjacent satellite DNA on the stability of the TK^+ phenotype, reversion to the TK^- phenotype was studied in randomly chosen transfectants. TK^+ cells were grown in non-selective medium (DMEM) for 3 or 7 days, allowing reversion to occur, and subsequently plated in DMEM, DMEM with HAT, and DMEM with TFT. The cells were continuously selected for the presence of the neo gene to maintain selection for an integrated plasmid. The plating efficiencies in each selective medium (HAT or TFT) were obtained by calculating the ratios of colony numbers in selective medium to that in non-selective medium (Table 3). The control KT cell line and cell lines obtained from the electroporation of circular plasmids always displayed lower plating efficiencies in TFT medium after 3 days of non-selection. However, after 7 days of non-selection, the plating efficiencies of these cell lines in TFT medium were higher compared to the plating

efficiencies in HAT medium. All the cell lines generated from the electroporation with linear plasmids, displayed very high plating efficiencies in both HAT and TFT media and no specific pattern of instability could be observed.

DISCUSSION

We have constructed a series of plasmids containing a 1797 bp EcoRI satellite II DNA previously isolated from the genome of HeLa cells (Sol *et al.*, 1986). These plasmids also contain two genes, tk and neo, that can be selected for after electroporation into a TK^-/NEO^- human cell line. All the plasmids used in this study are similar except that the length, the location and the orientation of the stDNA vary. Plasmids pCFU1 and pCFD1 are comprised of 16% satellite sequences located upstream and dowstream of the tk gene, respectively. Twenty eight percent of pCFU2 is composed of satellite DNA sequences, which are located upstream of the tk gene. The plasmids were introduced into a thymidine kinase deficient and neomycin sensitive cell line by the electroporation method to avoid DNA damage and multiple copy insertions reported for the calcium phosphate method (Lebkowski *et al.*, 1984; Kucherlapati and Skoultchi, 1984). The plasmid DNAs were electroporated either in their circular or linear form. The linear form of the plasmids gave rise to more transfectants than their respective supercoiled form (Table 1) presumably due to facilitation of integration of exogenous DNA *via* the double-stranded ends generated by cleavage (Toneguzzo *et al.*, 1986; Chu *et al.*, 1987; Folger *et al.*,1982). The ratio of efficiency obtained with linear DNA to that of circular DNA cannot be compared to other reported results since it seems to be cell type specific (Potter *et al.*, 1984; Gusew *et al.*, 1987; Chu *et al.*, 1987).

The presence of one or two satellite DNA, next to the tk gene in the plasmid constructs pCFU1 and pCFU2, did not affect the electroporation effi-

ciency of these plasmids compared to that of the stDNA-lacking plasmid (pSVL) as the same level of decrease in electroporation efficiency wa observed with the size control plasmids pCF1.8 and pCF3.8. This result is in contrast with the results reported by Talarico *et al.* (1988). They showed that mouse stDNA reduced the transfection efficiency, by 4-8 fold, when cotransfected with a plasmid containing the tk gene. However, the stDNA was initially located *in trans* to the tk gene. On the other hand, when the stDNA was moved close to the neo gene and its orientation reversed, the resulting plasmid (pCFD1) almost completely lost its electroporation efficiency. It is unlikely that the integration site within the host genome is solely responsible for this effect since electroporation with circular pCFD1 was repeated 11 times and always failed to give $HAT^R/G418^R$ colonies. The possibility that a component in the DNA preparation, DNA degradation, or that stDNA sequences are inhibiting plasmid integration could be rejected, since the pCFD1 electroporation efficiency was very high when selected for the neo gene only, and clones could be isolated from the stepwise selection using G418 selection followed by G418 plus HAT selection. The observed decrease in efficiency is not size dependent since the similarily sized plasmid pCF3.8 showed a higher efficiency. In addition, it is interesting to note that the pCFD1 stDNA is not in the same orientation as in pCFU1 and pCFU2 with respect to the tk gene. Thus it appears that the stDNA located near the neo gene is somehow interfering with the generation of TK^+/NEO^+ transfectants in a location and/or orientation-dependent manner. To determine if the observed decrease in electroporation efficiency is orientation-dependent, the stDNA would have to be cloned at the same position but in the opposite orientation (stDNA bp 1797 adjacent to the 5' end of the tk gene). If the inhibitory effect is orientation-independent, the same results should be obtained.

By using different selection procedures, we were able to determine

if the stDNA-containing plasmids were preferentially interfering with the generation of TK^+ or NEO^+ stable transfectants (Table 2). Plasmid pSVL, lacking stDNA, had similar electroporation efficiencies in both HAT and G418 single selections implying that the tk and neo genes are equally functional in these transfectants. The presence of stDNA sequences within the electroporated plasmids decreased the efficiencies of the circular plasmids by one order of magnitude when the tk gene was selected for alone (HAT^R), compared to those observed when the neo gene was selected for alone ($G418^R$). From these results, it can be seen that the adjacent stDNA, in circular plasmids, is either promoting plasmid integration via the tk gene or is interfering with its expression. This effect is dramatically illustrated in the case of pCFD1. However, this effect was not seen with the linear plasmids suggesting that linearized plasmids were not subjected to the same effects as the circular plasmids. In another set of stepwise selection experiments, where the presence of the tk gene was first selected for (HAT) and the transfectants then subjected to selection for the tk and neo genes (HAT then, HAT plus G418), similar results were obtained (data not shown). Highly repeated DNA sequences were reported to be involved in recombination (Wahls *et al.*, 1990; Yen *et al.*, 1990), DNA rearrangement (Heartlein *et al.*, 1988; Butner and Lo, 1986b) and chromosome breakage (Chaterjee and Lo, 1989). However, it is not clear if the repetitive sequences actually promote these effects. Since stDNA sequences within the plasmids might be involved in such processes, they could potentially serve as a "hot spot" for plasmid integration (Wallenberg *et al.*, 1984) by illegitimate recombination between the plasmid stDNA sequences and unique (Wallenberg *et al.*, 1987) or repetitive (Kato *et al.*, 1986) host sequences. It is possible that the different behaviors displayed by the cell lines obtained from the electroporation of circular plasmids, and those displayed by the cell lines obtained from electroporation of linear plasmids represent

the manner in which the electroporated plasmids integrated into the host genome. Further analyses at the molecular level will be required to rule out this issue.

Integration of plasmid sequences within the heterochromatic regions of the chromosomes, which contain stDNA, has been reported (Butner and Lo, 1986a; Talarico *et al.*, 1988). These two groups correlated unstable expression of the transfected tk gene with the presence and/or rearrangement of the flanking stDNA. To directly assess the involvement of adjacent stDNA on tk gene expression, cell lines were chosen and the stability of the TK^+ phenotype was studied (Table 3). Besides the lower plating efficiencies in TFT medium, of cell lines isolated from electroporation of circular plasmids, after 3 days of non-selection, no other significant differences were detected. It seems that the instability of the TK^+ phenotype, in these cases, is time dependent. The cell lines, electroporated with plasmids containing either 0, 1 or 2 stDNA inserts, were highly unstable. The plating efficiency mean was calculated for the cell lines obtained with each plasmid and Student t tests were performed. In no case could significant differences be detected between the KT cell line, cell lines electroporated with the control plasmid (pSVL) and those electroporated with plasmids containing stDNA adjacent to the tk gene, nor could differences be detected between the stDNA-containing cell lines suggesting that these cell lines are highly unstable, and that the observed phenotypic instabilities could not be ascribed to the transfected stDNA. The stability of the TK^- phenotype was studied, using 21 randomly chosen clones, and also appeared to be unstable (data not shown). Sweet *et al.* (1981) reported highly unstable tk gene expression resulting from heritable but unstable alterations in the tk gene. The reversion frequencies that we obtained exceeded the observed rate of mutation (10^{-6}) for a hemizygous gene (DeMars, 1974). Moreover, the reversion rates in TFT medium are much higher (10-100

fold) than those reported for the tk (Talarico *et al.*, 1988; Ostrander *et al.*, 1982; Sweet *et al.*, 1981) or the gpt gene (Gebara *et al.*, 1987). It is unlikely that reversion to the TK^- phenotype is due to of the loss of the tk gene because very few colonies would have been expected in HAT medium. Several TK^- clones were isolated and their DNA analyzed for the presence of the tk gene by the polymerase chain reaction technique. The primers used encompassed the entire tk gene coding region (base-pair 561 to 1958 [McKnight, 1980]) and yielded a 1.4 kb amplified product. The TK^- clones were found to be amplifiable and no changes in the amplified product size were detected (data not shown). Ostrander *et al.* (1982) observed that TK^+ cells could continue to synthesize tk mRNA when maintained in medium that selects for TK^- cells, leading to high TK^+ phenotype instability. DNA methylation was also shown to be involved in repression-reexpression of the tk gene (Christy and Scangos, 1984; Ostrander *et al.*, 1982; Butner and Lo, 1986a). As well, unstable expression of transfected genes was reported to be due to amplification of the exogenous genes (Gelbert *et al.*, 1990). Modulation of tk gene expression can also be correlated with differential DNAseI sensitivity, reflecting changes in the chromatin structure (Sweet *et al.*, 1982; Davies *et al.*, 1982).

The experiments described here have shown that the presence of a human EcoRI satellite II DNA within an electroporated plasmid can affect the plasmid electroporation efficiency in a location and/or orientation-dependent manner. Moreover, this satellite DNA interfered with the generation of TK^+ transfectants. The instability of the TK^+ phenotype does not appear to be related to the exogenously added satellite DNA. Characterization of the cell lines, at the molecular level, will allow the elucidation of the mechanisms by which satellite DNA mediates its effects and will provide insights into satellite DNA function within the human genome.

ACKNOWLEDGEMENTS

We would like to thank Josée Brisebois, Angelina Guzzo and Peter Ulycznyj for helpful discussions and criticism of this manuscript. C.F. was supported by a scholarship from the Natural Sciences and Engineering Research Council of Canada (NSERC). M.S.D. is a research scholar (exceptional merit) of the Fonds de la Recherche en Santé du Québec. This work was supported by a grant (OGP0003222) from NSERC.

FIGURE 1: Maps of Plasmids pSVL, pCFU1, PCFU2, pCFD1, pCF1.8, and pCF3.8. Sequences derived from pBR322 and SV40: ——; bla: ampicillin resistance gene; neo^r: Tn5 neomycin resistance gene; TK: HSV-1 thymidine kinase gene; SAT: 1.8 kb EcoR1 satellite II DNA; 1.8kb: 1.8 kb NdeI Lambda DNA fragment; 3.8kb: 3.8 kb Lambda DNA fragment. B= BamHI, H= HindIII, N= NdeI, S =StuI. The arrows indicate the orientation of the DNA segments from bp 1 to the end of the segments. Sizes are indicated in kilobase pairs (kb). Plasmids are not drawn to scale.

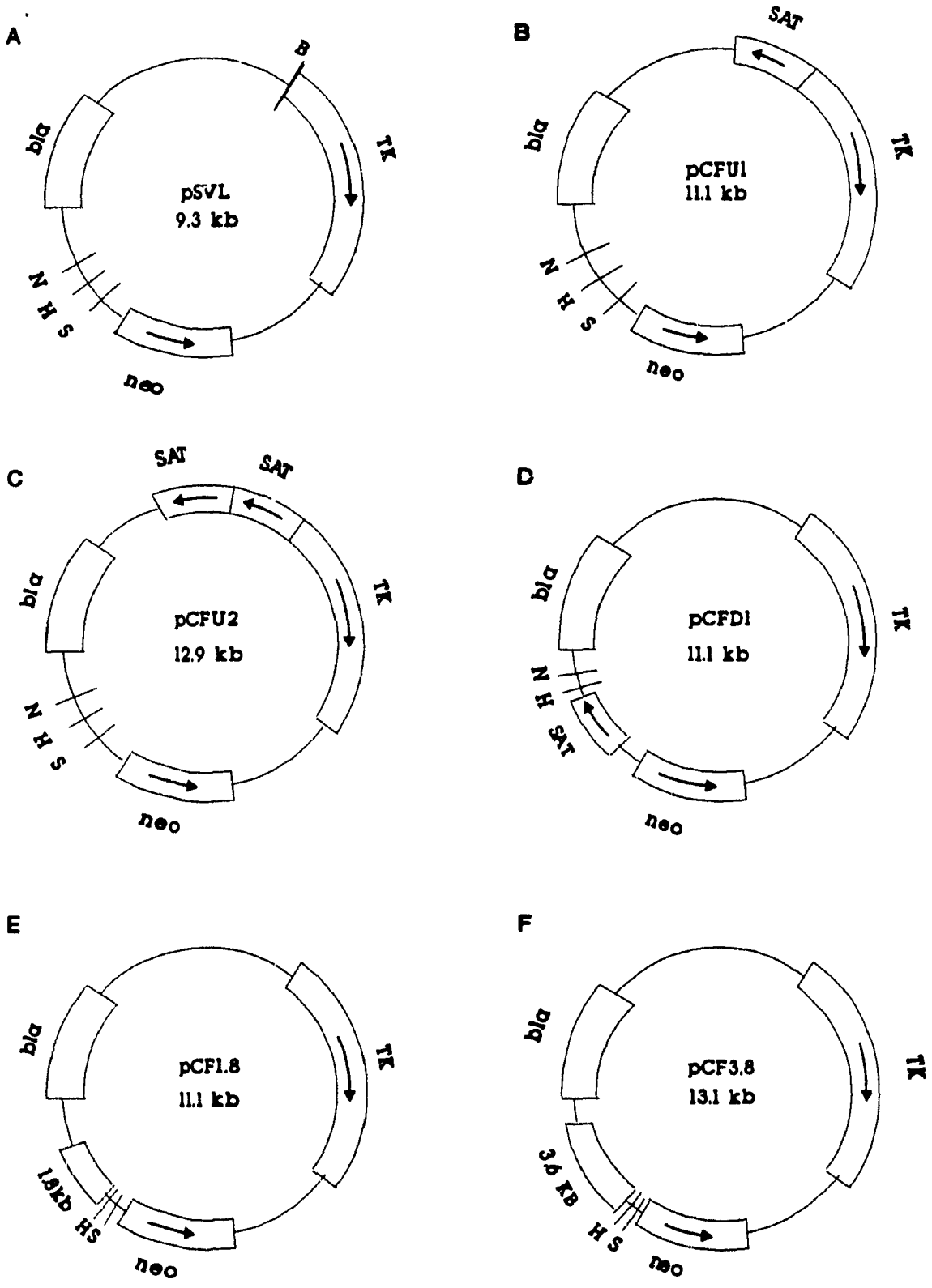

A
B
bla
TK
pSVL
9.3 kb
N H S
neo
B
SAT
pCFU1
11.1 kb
C
SAT
SAT
pCFU2
12.9 kb
D
pCFD1
11.1 kb
SAT
E
pCF1.8
11.1 kb
1.8kb
F
pCF3.8
13.1 kb
3.6 KB

TABLE 1: Electroporation Efficiencies of Plasmid Constructs

Plasmid Form	Plasmid	Efficiency	
		# of Transfectants [a]	Percentage [b]
A. Circular	pSVL	1.2×10^5	100 %
	pCFU1	6.8×10^4	56 %
	pCFU2	1.3×10^4	11 %
	pCFD1	0	0 %
	pCF1.8	4.5×10^4	37%
	pCF3.8	1.1×10^4	9%
B. Linear	pSVL	3.2×10^5	100 %
	pCFU1	2.2×10^5	70 %
	pCFU2	1.7×10^5	54 %
	pCFD1	3.7×10^3	1 %

a: number of transfectants obtained in HAT plus G418 medium per nmole of DNA electroporated

b: The # of transfectants obtained with pSVL was equaled to 100%. The % that represents the # of transfectants obtained with the other plasmids were calculated.

(# transfectants /# of pSVL transfectants) X 100

TABLE 2: Effect of Satellite DNA on Electroporation Efficiency[a]

		Selection					
PLASMID		HAT[b]	%[e]	G418[c]	%[e]	G418/HAT[d]	%[e]
A. Circular	pSVL	2.6×10^5	100%	6.6×10^5	100%	3.3×10^4	100%
	pCFU1	1.6×10^5	61%	1.1×10^6	166%	4.1×10^4	124%
	pCFU2	6.9×10^4	26%	4.3×10^5	65%	1.6×10^4	48%
	pCFD1	0	0%	5.1×10^5	77%	1.2×10^3	3.6%
B. Linear	pSVL	6.1×10^5	100%	7.4×10^5	100%	8.1×10^4	100%
	pCFU1	6.0×10^5	98%	1.3×10^6	175%	8.0×10^4	98%
	pCFU2	4.6×10^5	75%	2.7×10^5	36%	3.9×10^4	48%
	pCFD1	1.6×10^5	26%	1.8×10^5	24%	2.0×10^3	2.4%

a: number of transfectants per nmole of DNA electroporated

b: represents the electroporation efficiency in HAT

c: represents the electroporation efficiency in G418

d: represents the electroporation efficiency in the stepwise selection using G418 first, then G418 plus HAT

e: % was calculated as in Table 1

TABLE 3: Reversion Analysis on TK$^+$ Cells [a]
Relative plating efficiencies [b] after 3 and 7 days
of culture in non-selective medium

		3 DAYS		7 DAYS	
	CELL LINE [c]	HAT	TFT	HAT	TFT
A. Control	KT	1.00	0.59	0.52	0.92
B. Circular	SV-6	0.73	0.17	0.50	0.93
	SV-10	0.78	0.47	0.52	0.54
	SV-12	0.42	0.33	0.50	0.98
	U1-2	0.55	0.27	0.47	0.89
	U1-12	0.40	0.00	0.68	0.82
	U2-6	0.59	0.20	0.98	0.72
	U2-7	0.92	0.00	0.57	0.77
	U2-13	0.70	0.13	0.97	0.86
C. Linear	SV-6	0.81	0.89	0.68	0.75
	SV-21	0.60	0.12	0.83	0.51
	SV-23	0.81	1.00	0.97	0.32
	U1-2	1.00	0.43	0.69	1.10
	U1-4	0.64	0.52	0.89	0.79
	U1-6	0.77	0.99	0.61	0.55
	U2-1	0.50	0.88	0.75	0.85
	U2-2	0.86	0.39	0.67	0.99
	U2-4	0.41	0.33	0.54	0.94
	D1-4	0.25	0.94	0.32	0.45
	D1-5	0.51	0.46	0.55	0.78
	D1-8	0.15	0.04	0.20	0.20

a: TK$^+$ clones were obtained in HAT plus G418 medium

b: ratio of colony numbers in HAT plus G418 and TFT plus G418 to that in DMEM

c: see Materials and Methods for explanations

CHAPTER III

Some of the proposed satellite DNA functions are thought to be mediated through protein-DNA interactions. In addition, highly repetitive DNA-binding proteins have been isolated and characterized. However, these proteins are binding to alpha or complex satellite DNAs. In order to further elucidate the function of simple satellite DNAs we sought to determine whether proteins exist that can specifically bind to the cloned 1979 bp EcoRI satellite II DNA.

A student in our laboratory, Jerry Zaharatos, first attempted to screen a HeLa λgt11 cDNA expression library with a double-stranded satellite DNA probe, under conditions that should allow detection of protein possesing satellite DNA-binding activity. However, after screening a very large number of plaques, he was never able to detect such activity.

To circumvert this problem, I used the southwestern blotting technique utilizing the double-stranded EcoRI satellite II DNA to detect satellite DNA-binding proteins in HeLa cells. This chapter reports the results obtained while conducting these experiments.

DETECTION OF SPECIFIC STDNA BINDING PROTEINS

IN HeLa NUCLEAR AND WHOLE CELL EXTRACTS

This chapiter has been submitted for publication to

Biochemical and Biophysical Research Communications

(Claire Fouquet and Michael S. DuBow)

ABSTRACT

Using the southwestern blotting technique, we have detected specific satellite DNA-binding proteins in protein extracts (both total cellular and nuclear) prepared from HeLa cells. These polypeptides bind to a human 1797 bp EcoRI satellite II DNA, which consists of tandemly repeated pentamer (5' TTCCA 3') units, but not to a control probe from plasmid pBR322. These proteins have apparent molecular weights of 100, 93, 77 and 34 kDa and were termed satellite binding protein (Sbp)-1, -2, -3, and -4, respectively. The identification of polypeptides that bind to these tandemly arranged, simple-sequence satellite DNAs may provide clues to the function and evolution of this abundant fraction of the human genome.

INTRODUCTION

Eukaryotic genomes contain a large amount of repetitive DNA (Singer, 1982b; Miklos,1985) which, in humans, accounts for 20-30% of the genomic sequences (Marx *et al.*, 1976). Some of these repetitive sequences are known as satellite DNAs due to their characteristic banding pattern in cesium salt density gradients. StDNAs are found in constitutive heterochromatin, predominantly at the centromeres (Brutlag, 1980; Lica *et al.*,1986; Chen and Hodgetts,1989), a location that is replicated late in the cell cycle (Selig *et al.*,1988) and usually not transcribed (Reeder, 1973; Gaubatz and Cutler, 1990). Two types of stDNAs have been defined, based on whether or not they are internally repetitive. Simple stDNAs are tandem repeats of very short nucleotide sequences (5-10 bp), whereas complex stDNAs are comprised of longer and

more complex repeat units (>100 bp). The tandem organization of stDNAs, and their association with constitutive heterochromatin, may be important to obtain an understanding of any functional or evolutionary role that they may play.

Towards this goal, a number of proteins that specifically interact with different classes of complex stDNA have been identified. A nuclear protein from African green monkey cells was found to bind to a 172 bp repeat of alpha stDNA, and it was suggested that functions for this protein are mediated through its nucleosome positioning activity (Strauss and Varshavsky, 1984). StDNA has been shown to modulate transcriptional activity in mouse cells, possibly through changes in chromatin structure induced by stDNA-binding proteins (Talarico *et al.*, 1988). CENP-B, a centromere-specific protein (80 kDa), binds to alpha stDNA monomers in human cells (Masumoto *et al.*, 1989; Wewrick *et al.*, 1990), implying that these sequences may play a role in centromere structure and/or function. Recently, minisatellite DNA binding-proteins, Msbp-1 (40 kDa), -2 (77 kDa) and -3 (115 kDa), have been identified (Collick and Jeffreys, 1990; Wahls *et al.*, 1991). To our knowledge, only one group has previously reported binding of a protein (10 kDa), from phytohemagglutinin stimulated lymphocytes (Tomilin *et al.*, 1988), to a cloned human simple satellite (III).

In an attempt to gain further insights into the function and heterochromatic organization of simple stDNA sequences, we sought to determine whether proteins exist that can specifically bind to a cloned and sequenced 1797 bp human EcoRI fragment of stDNA II from HeLa cells (Sol *et al.*, 1986). This sequence contains 35% of the pure pentamer stDNA II consensus sequence (5' TTCCA 3') plus variations of this simple repeat, as well as a 49 bp region devoid of satellite-like sequences. We report here the use of this cloned DNA to detect stDNA-binding proteins

in HeLa cell extracts.

MATERIALS AND METHODS

DNA Manipulations:

All restriction endonucleases were purchased from Bethesda Research Laboratories Inc. (BRL) or Pharmacia. The DNA was hydrolysed, extracted and precipitated according to Maniatis *et al.* (1982).

The isolation, cloning (pKS36), and sequencing of the 1797 bp EcoRI satellite II DNA is reported elsewhere (Sol *et al.*, 1986). The pKS36 plasmid was cleaved with EcoRI and the 1.8 kb fragment was purified using the Gene Clean kit (BIO101) according to the manufacturer's instructions. Plasmid pBR322 was cleaved with RsaI, and a 1.5 kb fragment was isolated as a non-satellite DNA control (Fig. 1).

Probes and Southwestern Blot Analysis:

HeLa whole cell extracts, purified on heparin-agarose (Imagawa *et al.*, 1987), were purchased from Stratagene. HeLa cell nuclear extracts, prepared according to Dignam *et al.* (1983), were kindly provided by Dr. N. Sonenberg. The proteins were separated by electrophoresis through a 12.5% SDS-polyacrylamide denaturing gel (Laemmli, 1970) in 25 mM Tris-base, 192 mM glycine, 0.1% (w/v) SDS (pH 8.3). Following electrophoresis, the gel was soaked for 30 minutes in 25 mM Tris-base, 192 mM glycine, 20% (v/v) methanol (pH 8.3) and then the proteins were electroblotted (Towbin *et al.*, 1979) onto polyvinylidenedifluoride membranes (Immobilon, Millipore) for 2 hours at 0.5 A. The membrane was incubated for 1 hour at 37° C in 1X Denhardt (1966), 150 mM NaCl, 10 mM Tris-HCl (pH 7.6), 0.1 mM EDTA, 20 mg/ml yeast RNA

(Boehringer Mannheim Canada) and 50 μg/ml sonicated, double stranded *E. coli* genomic DNA. 10^7cpm of [^{32}P] labeled DNA (see below) were added and incubation was allowed to occur overnight at 37° C. The filters were washed 3 times for 30 minutes in 1X TNE (Tris-HCl (pH 7.6), 150 mM NaCl, 2 mM EDTA [Collick and Jeffreys, 1990]) at 37° C. The membranes were air-dried and exposed to XAR-5 film (Kodak) with a Dupont Cronex intensifying screen at -70° C. The protein gels were stained with Coomassie blue dye according to de Moreno *et al.* (1985).

The probes were labeled to specific activities of 10^8-10^9 cpm/μg of DNA by the random priming extension method (Feinberg and Vogelstein, 1983) with the following modifications. DNA (0.1-0.2 μg) was mixed with 4.8 μg of hexanucleotide primer (Regional Synthesis Laboratory, University of Calgary). The DNA was denatured by boiling for 5 minutes and then quickly cooled on ice for 5 minutes. The reaction was carried out in a total volume of 50 μl in 40 mM KPO_4 (pH 7.0), 4 mM $MgCl_2$, 8 mM DTT, 40 μM dCTP and dGTP, 30 μCi α-[^{32}P]dATP and -dTTP (ICN, 3000 Ci/mmole), and 5 units of Klenow enzyme (BRL). The reaction was allowed to proceed at room temperature for 3 hours. The reaction was stopped by adding EDTA to a final concentration of 20 mM and the unincorporated nucleotides were removed by passing the mixture through a Sephadex G-50 spin column (Ausubel *et al*, 1989).

RESULTS

Detection of Satellite II DNA-Binding Proteins in HeLa Cell Extracts.

The pKS36 plasmid was cleaved with EcoRI and the satellite DNA fragment was purified (Fig. 1A) and used to detect satellite DNA-binding proteins in HeLa cell nuclear extracts. The proteins were separated by SDS-polyacrylamide gel electrophoresis (Fig. 2) at four different concentrations and then transferred to an Immobilon membrane. The proteins were then incubated with radiolabeled satellite II DNA, and four nuclear polypeptides, with apparent molecular weights of 100, 93, 77 and 34 kDa, were detected (Fig. 2B) at the highest protein concentration. These proteins were called Satellite binding protein (Sbp) -1, -2, -3, and -4, respectively. Protein Sbp-3 was detectable at the lowest protein concentration, while higher concentrations were necessary to detect Sbp-1 and then Sbp-2 and Sbp-4. In contrast, the pBR322 non-specific fragment only bound to a 97 kDa protein, but the signal was very weak even at the highest protein concentration tested (Fig. 2C). In addition, non-specific binding to a polypeptide of 14 kDa was detected with both the satellite DNA and pBR322 probes (arrows in Fig. 2). In order to confirm these results, partially purified total HeLa cellular extracts (Fig. 3A) were also separated on denaturing SDS polyacrylamide gels and assayed for satellite DNA-binding proteins. Polypeptides of 77 and 34 kDa were found to bind to the 1.8 kb stDNA fragment (Fig. 3B). No polypeptides were detected that could bind to the control pBR322 DNA fragment under these assay conditions (Fig. 3C).

DISCUSSION

Using the southwestern blotting technique, we have detected satellite DNA-binding proteins in HeLa cells. We have showed that the nuclear satellite DNA-binding proteins-1, -2, -3, and -4 (Sbp-1, -2, -3, -4) specifically bind to the cloned 1797 bp EcoRI satellite II DNA. These proteins have observed molecular weights of 100, 93, 77 and 34 kDa, respectively (Fig 2B, 3B). Only Sbp-3 and Sbp-4 were present in a partially purified whole cell extract (Fig. 3B). Nuclear proteins are expected to be present in a lower concentration in a whole cell extract when the same amount of proteins is analyzed. This could, in part, explain the non-detection of Sbp-1 and Sbp-2 in the whole cell extract. On the other hand, chromatography of the whole cell extract on a heparin-agarose column selected for basic proteins, while acidic proteins did not bind to this type of column. If Sbp-1 and Sbp-2 are acidic proteins, it would be impossible to detect them when the extracts are purified by this method. Nonetheless, Sbp-3 had the highest apparent affinity for satellite II DNA (Fig. 2B), and this may explain its detection in both assays.

It is interesting to note that the Spb-3 protein detected by us has the same molecular weight as the Msbp-2 protein reported by Wahls *et al.* (1991). This group used tandem repeats (32 repeats [512 bp]) of the minisatellite consensus sequence (5' GGAGGTGGGCAGGARC 3') as a probe to detect minisatellite DNA-binding proteins in HeLa nuclear extracts. The satellite II DNA pentamer repeat (5' TTCCA 3') is not present in the minisatellite core sequence. Furthermore, the minisatellite core sequence contains 62.5% G residues, whereas the 1797 bp satellite II fragment is comprised of only 3.8% G nucleotides. We do not know

at the present time if these two proteins are identical.

The tandem repeat organization was shown to be a prerequisite of minisatellite DNAs for protein binding (Collick and Jeffreys, 1990; Wahls *et al.*, 1991). Tomilin *et al.* (1988) reported that a stDNA II-binding protein of 10 kDa does not bind to stDNA fragments of 80 bp, but could bind to fragments of 2.2 kb. This contrasts with most other DNA-binding proteins which generally require less than 20 bp for sequence-specific binding (Ptashne, 1988). It appears that the stDNA tandem organization may be a general requirement for protein binding and heterochromatization. We found that smaller fragments of the satellite II DNA yielded greatly reduced binding in our assay (data not shown).

The functions of Sbp-1, -2, -3 and -4 are not currently known. However, the satellite DNA-binding proteins reported here can now be purified, using southwestern blotting as an assay, for the study of the protein-DNA interactions on satellite II DNA and ultimate cloning and characterization of the genes which encode them.

ACKNOWLEDGEMENTS

C.F. was supported by a scholarship from the Natural Sciences and Engineering Research Council of Canada (NSERC). M.S.D. is a Research Scholar (Exceptional Merit) of the Fonds de la Recherche en Santé du Québec. This work was supported by a grant (OGP0003222) from NSERC.

Figure 1.

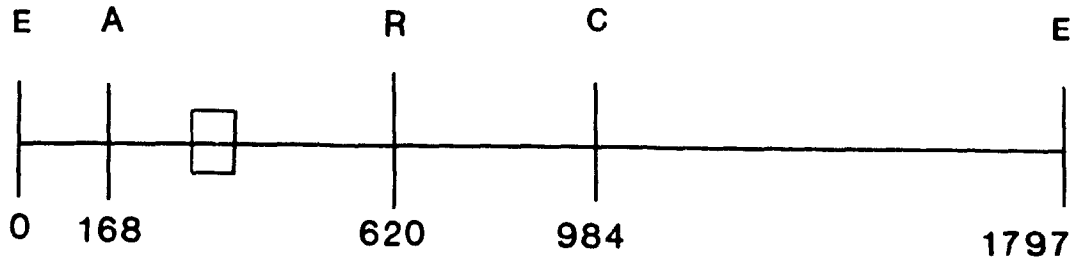
A. Partial restriction map of the 1.8 kb EcoRI satellite II DNA. The box represents the 49 bp segment devoid of satellite-like sequences (Sol *et al.*, 1986).

B. Partial restriction map of pBR322 (Maniatis *et al.*, 1982). The bracket below the map represents the 1.5 kb RsaI fragment used as a control probe.

E = EcoRI, A = AvaII, R = RsaI, C = ClaI.

The location of each restriction site is indicated in base pairs.

A



B

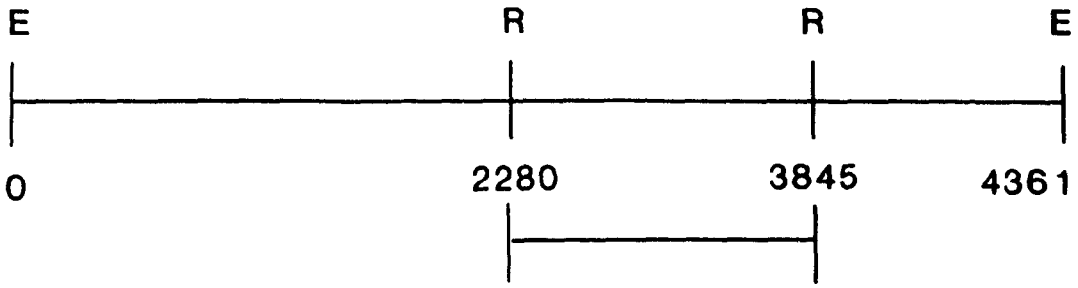


Figure 2. Detection of satellite II DNA-binding proteins by southwestern analysis (see Materials and Methods) of HeLa nuclear extracts.

Panel A: Coomassie stain of the protein gel (lanes 1-4 represent decreasing protein concentrations).

Panel B: autoradiogram of the proteins bound by the 1797 bp satellite II DNA probe.

Panel C: autoradiogram of the proteins bound by the non-specific 1.5 kb pBR322 probe.

M = molecular weight markers (97, 68, 42, 25, 18, and 15) in kDa.

Arrows indicate presumed non-specific DNA-binding proteins.

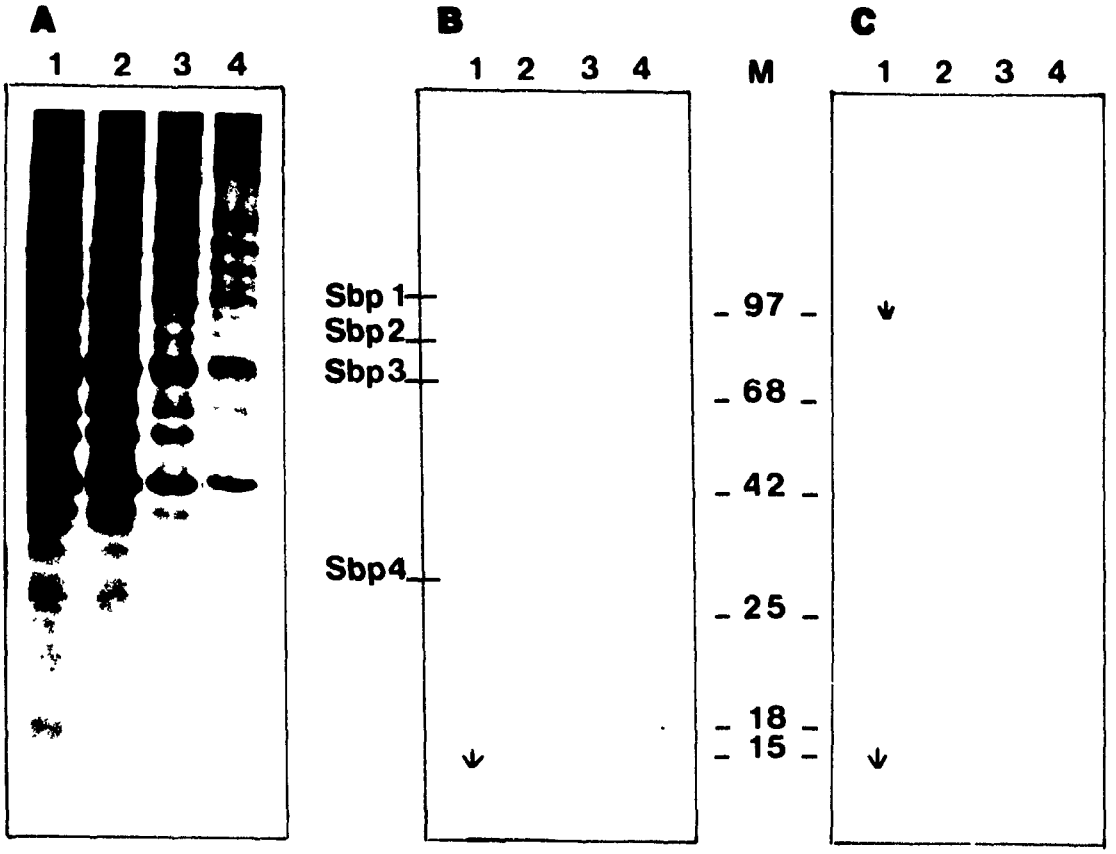

A
1 2 3 4
B
1 2 3 4
M
C
1 2 3 4
Sbp 1
Sbp2
Sbp3
Sbp4
97
68
42
25
18
15

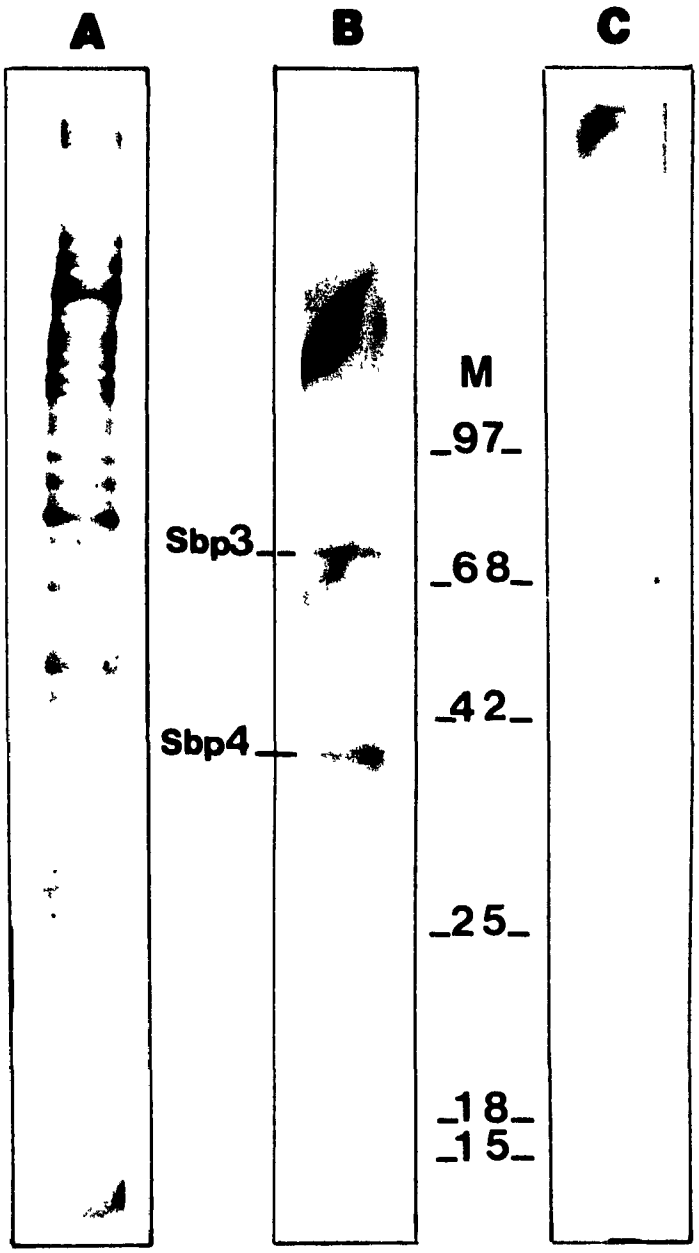
Figure 3. Detection of satellite II DNA-binding proteins by southwestern analysis (see Materials and Methods) of a partially purified HeLa whole cell extract.

Panel A: Coomassie stain of the protein gel.

Panel B: autoradiogram of the proteins bound by the 1797 bp satellite II DNA probe.

Panel C: autoradiogram of the proteins bound by the 1.5 kb pBR322 probe.

M = Molecular weight markers in kDa (as in Fig. 2).


A
B
C
M
97
68
42
25
18
15
Sbp3
Sbp4

CHAPTER IV

CONCLUSIONS AND DIRECTIONS FOR FUTURE RESEARCH

The amount of DNA in eukaryotic genomes by far exceeds what is necessary for its function. This excess DNA is highly repeated and is comprised, among other repetitive sequences, of simple satellite DNA sequences. They are characterized by their tandemly repeated patterns organized in long arrays, their sequence divergence and their persistence through evolution within the human genome. Satellite DNA sequences are found mostly within the highly condensed parts of the chromosomes, the heterochromatin. Thirty years after their discovery, their function within the genome remains to be identified. Their possible role, among other proposed functions, in heterochromatization, to specify higher-order chromosome structures and thereby the regulation of gene expression, seems extremely attractive. If this was the case, entire blocks of genes would be expressed or repressed depending on the state of the neighboring chromosomal region. Unstable expression of euchromatic genes has been shown to be related to DNA rearrangements and/or the presence of adjacent satellite DNA sequences. These effects could possibly be mediated by satellite DNA-binding proteins.

The major aim of this research was to determine the role that satellite DNA sequences may play, either directly or indirectly, on gene stability and expression, and identify proteins that may mediate their function.

In order to study the effect of satellite DNA sequences on the expression of genes adjacent to these sequences, a series of plasmids, containing a cloned and sequenced 1797 bp human EcoRI satellite II DNA, and selectable marker genes (tk and neo), were constructed Following electroporation of these plasmids (satellite DNA-containing plasmids [pCFU1, pCFU2, pCFD1] and the satellite DNA-lacking plasmids [pSVL, pCF1.8, pCF3.8]) into human

TK^-/NEO^- cells (143B), the electroporation efficiencies obtained were recorded. We observed that the decrease in electroporation efficiency obtained with plasmid pCFU1, containing one 1.8 kb stDNA sequences next to the tk gene, and with pCFU2, comprising two of these satellite DNA sequences also located next to the tk gene, was due to an increase in plasmid size. However, plasmid pCFD1, which contains only one stDNA sequence next to the neo gene, had a decreased electroporation efficiency when compared to that of the size control plasmid pCF1.8. From these results, we suggest that the presence of the satellite DNA affects the electroporation efficiency in a location-dependent manner. In addition, this effect could also be stDNA orientation-dependent since, in pCFD1 the satellite sequences are oriented differently with respect to their orientation in pCFU1 and pCFU2. However, this hypothesis remains to be confirmed.

We next investigated the stability of the TK^+ phenotype in TK^+ cells generated by the electroporation of pSVL, pCFU1, pCFU2 and pCFD1 plasmids in TK^- cells. Stable TK^+ transfectants were isolated, the selection pressure was removed for several days, the cells were plated in media selecting for the TK^+ or TK^- phenotype and the relative plating efficiencies were calculated. We found that satellite DNA sequences within the transfected plasmids, were not involved in the instability of the TK^+ phenotype. The cell lines isolated from electroporations with the control plasmid (pSVL) and the satellite DNA-containing plasmids (pCFU1, pCFU2, pCFD1) were highly unstable. Preliminary studies on TK^- cells, isolated from the parental TK^+ cell lines, showed that the TK^- phenotype is also unstable. The reasons for this instability are, as yet, unknown.

By using different selection procedures we examined the *cis* effect of satellite DNA on plasmid gene expression. We have shown that electroporation with satellite DNA-containing plasmids generates more NEO^+ than TK^+ transfectants.

To verify that satellite DNA sequences decrease electroporation efficiency in a location and/or orientation dependent manner, satellite sequences should be cloned at different places within the pSVL plasmid and in both orientations, between the tk and neo genes for example, and the electroporation efficiency of these plasmids assayed. Control plasmids containing non-satellite DNA (λ fragments cloned at the same positions as the stDNAs) should also be tested.

The high instability of the TK^+ phenotype could be ascribed to the exogenous DNA chromosomal location. To test this hypothesis, plasmid sequences could be rescued from the cell lines, or genomic libraries could be constructed, and the isolated DNA electroporated back into a thymidine kinase deficient cell line. If the tk gene, contained in the rescued DNA is also highly unstable, it could suggest that the instability is caused by intrinsic properties of that DNA rather than a chromosomal location effect. At the same token, junction regions between the plasmid and host sequences could be studied. It would be interesting to determine if the non-satellite and satellite-containing plasmids have integrated into host repetitive sequences or if the integration occurred randomly in both cases. Correlation (or lack of correlation) between phenotypic instability and chromosomal location could be made. Moreover, analysis of the rescued DNA, by restriction endonuclease cleavage and Southern blotting, should allow the determination of the integration site on the plasmid. These analyses will discriminate between the more frequent integration of satellite DNA-containing plasmid via tk sequences and inactivation of this gene, as the possible cause of the generation of less TK^+ than NEO^+ transfectants.

In the second part of this study we wanted to determine whether or not proteins exist that can specifically bind to the cloned 1797 bp EcoRI satellite II DNA. These poplypeptides may well prove to play a crucial role in the structure and function of these repetitive sequences.

Proteins from HeLa nuclear and partially purified whole cell extracts were separated by electrophoresis on SDS-polyacrylamide gels, transferred to membranes and assayed for DNA-binding activity using the labeled 1.8 kb stDNA as a probe. Utilizing this assay, proteins that were specifically bound to the satellite DNA were detected and termed satellite DNA-binding proteins (Sbp). Four Sbp's, Sbp-1, -2, -3 and -4, were detected in the nuclear extract, whereas only two of these proteins, Sbp-3 and -4, were detected in the partially purified whole cell extract. The absence of Sbp-1 and Sbp-2 in the whole cell extract could be explained by the purification procedure used, in which only the basic proteins were isolated and in which nuclear proteins were present in lower concentrations. The Sbp-1, -2, -3, and -4 have respective molecular weights of 100, 93, 77 and 34 kDa. At the present time, the function of these proteins is not known.

The southwestern blotting technique could be used as an assay to purify these proteins. Once purified, the interactions of these proteins with the satellite DNA can be directly assessed by using the 1.8 kb satellite DNA, or internal fragments, in band retardation assays and DNAseI footprinting experiments. These studies would allow the precise nucleotide sequences to which these proteins bind to be identified.

Antibodies to the satellite DNA-binding proteins could be raised and used to screen HeLa cell cDNA expression libraries to identify clones which con tain expressed mRNAs. These clones could then be used to clone the genes wich encode these polypeptides and ultimately the role of the satellite DNA binding proteins and their effect on the function of satellite DNA could be determined.

The results reported here open up several avenues of investigation that will lead to the elucidation of the role of these enigmatic satellite DNA sequences.

References

Ashley,C.T., Pendleton,C.G., Jennings,W.W., Saxena,A., and Glover,C.V. (1989) Isolation and sequencing of cDNA clones encoding *Drosophila* chromosomal protein D1. J. Biol. Chem. 264: 8394-8401.

Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Deidman,J.G., Smith,J.A., and Struhl,K. (eds) (1989) Current Protocols in Molecular Biology. Green Publishing Associates and Wiley-Interscience, New-York.

Beattie,W.G., and Skinner,D.M. (1972) The diversity of satellite DNAs of *Crustacea*. Biochim. Biophys. Acta 281: 169-178.

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

Berg,D., Jorgensen,R., and Davies,J. (1978) Transposable kanamycin-neomycin resistance determinants. In: Microbiology. D. Schessinger (ed.), pp.13-15. Washington, D.C.: ASM Publications.

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

Blin,N., and Stafford,D.W. (1976) A general method for isolation of high molecular weight DNA form eukaryotes. Nucleic Acids Res. 3: 2303-2308.

Bonnacorsi,S., Gatti,M., Pisano,C., and Allan,L. (1990) Transcription of a satellite DNA on two Y chromosome loops of *Drosophila melanogaster*. Chromosoma 99: 260-266.

Borsani,G., Tonlorenzi,R., Simmler,M.C., Dandolo,L., Ardaud,D., Capra,V., Grompe,M., Pizzuti,A., Muzny,D., Lawrence,C., WillardH.F., Avner,P., and Ballbio,A. (1991) Characterization of a murine gene expressed from the inactive X chromosome. Nature 351: 325-331.

Botchan,M.R. (1974) Bovine satellite I DNA consists of repetitive units 1400 bp in length. Nature 251: 288-292.

Britten,R.J., 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.

Brutlag,D.L. (1980) Molecular arrangements and evolution of heterochromatic DNA. Ann. Rev. Genet. 14: 121-144.

Butner,K., and Lo,C.W. (1986a) Modulation of *tk* expression in mouse pericentric heterochromatin. Mol. Cell. Biol. 6: 4440-4449.

Butner,K.A., and Lo,C.W. (1986b) High frequency of DNA rearrangement associated with mouse centromeric satellite DNA. J. Mol. Biol. 187: 547-556.

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

Carlson,M., and Brutlag,D. (1979) Different region of a complex satellite DNA vary in size and sequence repeating unit. J. Mol. Biol. 135: 483-500.

Cedar,H. (1988) Dans methylation and gene activity. Cell 53: 3-4

Chatterjee,B., and Lo,W. (1989) Chromosomal breakage associated with instability in mouse centromeric satellite DNA J. Mol. Biol. 210: 303-312.

Chen,Z.-Q., Lin,C.C., and Hodgetts,R.B. (1989) Cloning and characterization of tandemly repeated DNA sequence in the crane family (Gruidae) Genome 32: 646-654.

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

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

Christy,B., and Scangos,G. (1982) Expression of transferred thymidine kinase genes is controlled by methylation. Proc. Natl. Acad. Sci. USA 79: 6299-6303.

Christy,B.A., and Scangos,G.A. (1984) Changes in structure and methylation pattern in a cluster of thymidine kinase genes. Mol. Cell. Biol. 4: 611-617.

Chu,G., Hiroshi,H., and Berg,P. (1987) Electroporation for efficient transfection of mammalian cells with DNA. Nucleic Acids Res. 15: 1311-1325.

Church,K. (1964) Replication of chromatin in mouse mammary epithelial cells grown *in vitro*. Genetics 52: 843-849.

Collick,A., and Jeffreys,A.C. (1990) Detection of a novel mini-satellite-specific DNA-binding protein. Nucleic Acids Res. 18: 625-629.

Cohen,A.K., Huh,T.Y., and Helleiner,C.W. (1973) Transcription of satellite DNA in mouse L-cells. Cancer J. Biochem. 51: 529-532.

Cooke,H.J. (1975) Evolution of the long range structure of satellite DNAs in the genus Apodemus. J. Mol. Biol. 94: 87-94.

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

Davies,R.L., Fuhrer-Krusi,S., and Kucherlapati,R.S. (1982) Modulation of transfected gene expression mediated by changes in chromation structure. Cell 31: 521-529.

deMars,R. (1974) Resistance of cultured human fibroblasts and other cells to purine and pyrimidine analogs in relation to mutagenesis detection. Mut. Res. 24: 335-364.

de Moreno,M.R., Smith,J.F., and Smith,R.V. (1985) Silver staining of proteins in polyacrylamide gels: increased sensitivity through a combined Coomassie Blue-silver stain procedure. Anal. Biochem. 151: 466-470.

Denhardt,D.T. (1966) A membrane-filter technique for the detection of complementary DNA. Biochem. Biophys. Res. Commun. 23: 641-652.

Denovan,E.M., and Wright,J.M. (1990) A satellite DNA family from (*Pollachius virens*). Gene 87: 279-283.

Diaz,M.O., Barsacchi-Pilone,G., Mahon,K.A., and Gall,J.G. (1981) Transcripts from both strands of a satellite DNA occur in lampbrush chromosome loops of the newt Notophthalmus. Cell 24: 649-659.

Dignam,J.D., Lebovitz,R.M., and Roeder,R.G. (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. 11: 1475-1489.

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

Dover,G.A., (1989) DNA fingerprints: victims or perpetuators of DNA turnover. Nature 342: 347-348.

Dover,G.A. (1990). Mapping "frozen" accidents. Nature 344: 812-813.

Dumont,M.C., Bello,M.J., Guichaoua,M.R., and Luciani,J. (1989) Differential associative behavior of mitotic and meiotic acrocentric chromosomes. Hum. Genet. 82: 35-39.

Earnshaw,W.C., and Cooke,C.A. (1989) Proteins of the inner and outer centromere of mitotic chromosomes. Genome 31: 541-552.

Endow,S.A., Polan,M.L., anf Gali,J.G. (1975) Satellite DNA sequences of *Drosophila melanogaster.* J. Mol. Biol. 96: 665-692.

Epstein,L.M., Mahon,K.A., and Gall,J.G. (1986) Transcription of satellite DNA in the newt. J. Cell Biol. 103: 1137-1144.

Epstein,L.M., and Gall,J.G. (1987) Self cleaving transcripts of satellite DNA from the newt. Cell 48: 535-543.

Feinberg,A.P., and Vogelstein,B. (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132: 6-13.

Feinstein,S.I., Racaniello,V.R., Ehrlich,M., Gehrke,C.W., Miller,D.A., and Miller,O.J. (1985) Pattern of undermethylation of the major satellite DNA of mouse sperm. Nucleic Acids Res. 13: 3969-3978.

Flamm,W.G., Walker,P.M.B., and McCallum,M. (1969) Some properties of single strands isolated from the DNA of the nuclear satellite of the mouse (*Mus musculus*). J. Mol. Biol. 40: 423-443.

Folger,K.R., Wong,E.A., Wahl,G., and Cappechi,M.R. (1982) Pattern of integration of DNA microinjected into cultured mammalian cells: evidence for homologous recombination between injected plasmids DNA molecules. Mol. Cell. Biol. 2: 1372-1387.

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

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

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

Fry,K., and Brutlag,D. (1979) Detection and resolution of closely related satellite DNA sequences by molecular cloning. J. Mol. Biol. 135: 581-593.

Gall,J.G., and Atherton,D.D. (1974) Satellite DNA sequences in *Drosophila virilis*. J. Mol. Biol. 85: 633-664.

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. Nucleic Acids Res. 11: 3087-3095.

Gaubatz,J.M., and Cutler,R.G. (1990) Mouse satellite DNA is transcribded in senescent cardiac muscle. J. Mol. Biol. 265: 17753-17758.

Gebara,M.M., Drevon,C., Harcourt,S.A., Steingrimsdottir,H., James,M.R., Burke,J.F., Arlett,C.F., and Lehmann,A.R. (1987) Inactivation of a transfected gene in human fibroblasts can occur by deletion, amplification, phenotypic switching, or methylation. Mol. Cell. Biol. 7: 1459-1464.

Gelbert,L.M., Wilson,M.M., and Davidson,R.L. (1990) Analysis of GPT activity in mammalian cells with a chromosomally integrated shuttle vector containing altered *gpt* genes. Som. Cell Mol. Genet. 16: 173-184.

Gerber-Huber,S., May,F.E.B., Westley,B.R., Felber,B.K., Hosbach,H.A., Andres,A.-C., and Ryffel,G.U. (1983) In contrast to other *Xenopus* genes the estrogen-inducible vitellogenin genes are expressed when totally methylated. Cell 33: 43-51.

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

Goring,D.R., and DuBow,M.S. (1985) A cytotoxic effect associated with 9-(1,3-dihydroxy-2-propoxymethyl)-guanine is observed during the selection for drug resistant human cells containing a single herpesvirus thymidine kinase gene. Biochem. Biophys. Res. Comm. 133: 195-201.

Guichaoua,M.R., Devictor,M., Hartung,M., Luciani,J.M., and Stahl,A. (1986) Random acrocentric bivalent associations in human pachytene spermatocytes. Cytogenet. Cell. Genet. 42: 191-197.

Gusew,N., Nepveu,A., and Chartrand,P. (1987) Linear DNA must have free ends to transform rat cells efficiently. Mol. Gen. Genet. 206: 121-125.

Hanahan,D. (1983) Studies of transformation of *Escherichia coli* with plasmids. J. Mol. Biol. 166: 557-580.

Hardies,S.C., Axelrod,D.E., Edgell,M.H., and Hutchison,C.A. (1983) Phenotypic variation associated with molecular alterations at a cluster of thymidine kinase genes. Mol. Cell. Biol. 3: 1163-1171.

Harel,J., Hanania,N., Tapeiro,H., and Harel,L. (1968) DNA replication by nuclear satellite DNA in different mouse cells. Biochem. Biophys. Res. Commun. 33: 696-701.

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.

Heartlein,M.W., Knoll,J.H.M., and Latt,S.A. (1988) Chromosome instability associated with human alphoid DNA transfected into chinese hamster genome. Mol. Cell. Biol. 8: 3611-3618.

Henikoff,S. (1990) Position-effect variegation after 60 years. Trends. Genet. 6: 422-426.

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 KpnI sequence localized in the heterochromatin of chromosome 15. Chromosoma 93: 77-86.

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

Hsieh,T.-S., and Brutlag,D. (1979a) Sequence and sequence variation within the 1,688 g/cm^3 satellite of *Drosophila melanogaster*. J. Mol. Biol. 135: 465-481.

Hsieh,T.-S., and Brutlag,D.L. (1979b) A protein that preferentially binds *Drosophila* satellite DNA. Proc. Natl. Acad. Sci. USA 76: 726-730.

Hsu,T.C. (1962) Differential rate in RNA synthesis between euchromatin and heterochromatin. Exp. Cell Res. 27: 332-334.

Imagawa,M., Chiu,R., and Karin,M. (1987) Transcription factor AP-2 mediates induction by two different signal transduction pathways: protein kinase C and cAMP. Cell 51: 251-260.

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 "minisatellite" regions in human DNA Nature 314: 67-73.

Jelinek,W.R., Toomey,T.P., Leinwand,L., Duncan,C.H., Biro,P.A., Choudary,P.V., Weissman,S.M., Rubin,C.M., Houck,C.M., Deininger,P.L., and Schmid,C.W. (1980) Proc. Natl. Acad. Sci. USA 77: 1398-1402.

John,B., and Miklos G.L.G. (1979) Functional aspects of satellite DNA and heterochromatin. Int. Rev. Cytol. 58: 1-114.

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

Jones,K.W., Prosser,J., Corneo,G., and Ginelli,E. (1973) The chromosomal location of human satellite DNA III. Chromosoma 42: 445-451.

Jogersen,R.A., Rothstein,S.J., and Reznikoff,W.S. (1979) A restriction enzyme cleavage map of Tn5 and location of a region encoding neomycin resistance. Mol. Gen. Genet. 177: 65-72.

Kato,S., Anderson,R.A., and Camerini-Otero,D. (1986) Foreign DNA introduced by calcium phosphate is integrated into repetitive DNA elements of the mouse L cell genome. Mol. Cell. Biol. 6: 1787-1795.

King,P.V., and Blakesley,R.W. (1986) Optimizing DNA ligations for transformations. Focus 8: 1-3.

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

Kodama,H., Saitoh,H., Tone,M., Kuhara,S., Sakaki,Y., and Mizuno,S. (1987) Nucleotide sequences and unusual electrophoretic behavior of the W chromosome-specific repeating DNA units of the domesticus fowl, *Gallus gallus domesticus*. Chromosoma 96: 18-25.

Korenberg,J.R., and Rykowski,M.C. (1988) Human genome organisation: Alu, Lines, and the molecular structure of metaphase chromosome bands. Cell 53: 391-400.

Kucherlapati,R., and Skoultchi,A.J. (1984) Introduction of purified genes into mammalian cells. CRC Crit. Rev. Biochem. 16: 349-379.

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

Kurnit,D.M. (1979) Satellite DNA and heterochromatin variants: the case of unequal mitotic crossing over. Human Genet. 47: 169-186.

Laemmli,U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680-685.

Lam,B.S., and Caroll,D. (1983) Tandemly repeated DNA sequences from *Xenopus leavis*. II Disperded clusters of a 388 base-pairs repeating unit. J. Mol. Biol. 165: 587-597.

Lebkowski,J.S., Dubridge,R.M., Antel,E.A., Greisen,K.S., and Calos,M.P. (1984) Transfected DNA is mutated in monkey, mouse and human cells. Mol. Cell. Biol. 4: 1951-1560.

Lehrman,M.A., Goldstein,J.L., Russel,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.

Lerman,M.I., Thayer,R.E., and Singer,M.F.(1983) KpnI family of long interspersed repeated DNA sequences in primates: Polymorphism of family members and evidence for transcription. Proc. Natl. Acad. Sci. USA 80: 3966-3970.

Levinger,L., and Varshavsky,A. (1982) A selective arrangement of ubiquitinated and D1 protein-containing nucleosomes within the *Drosophila* genome. Cell 28: 375-385.

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

Linxweiler,W., and Horz,W. (1985) Reconstitution experiments show that sequence-specific histone-DNA interactions are the basis for nucleosome phasing on mouse satellite DNA. Cell 42: 281-290.

Lloyds,M.A., Fields,M.J., and Thoegaard,G.H. (1989) BKm minisatellite sequences are not sex-associated but reveal fingerprint polymorphism in rainbow trout. Genome 32: 865-868.

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

Maio,J.J., Brown,F.L., and Musich,P.R. (1977) Subunit structure of chromatin and the organization of eukaryotic highly repetitive DNA: recurrent periodicities and models for the evolutionary origins of the repetitive DNA. J. Mol. Biol. 117: 637-655.

Mandel,M., and Higa,A. (1970) Calcium dependant bacteriophage DNA infection. J. Mol. Biol. 53: 159-162.

Maniatis,T., Fritsch,E.F., and Sambrook,J. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratories, Cold Spring Harbor, New-York.

Manuelidis,L. (1976) Repeating restriction fragments of human DNA. Nucleic Acids Res. 3: 3063-3073.
Manuelidis,L. (1978) Chromosomal localization of complex and simple repeated humans DNAs. Chromosoma 66: 23-32.

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

Maresca,A., Singer,M.F., and Lee,T.H.N. (1984) Continuous reorganization leads to extensive polymorphism in a monkey centromeric satellite. J. Mol. Biol. 179: 629-649.

Maresca,A., Thayer,R.E., Gosselin,P., and Singer,M.F. (1987)Genomic organization of low copy number sequences that are associated with deca-satellite DNA in the monkey genome. Nucleic Acids Res. 15: 8799-8813.

Marx,K.A., Allan,J.R., and Hearst,J.E. (1976) Characterization of the repetitious human DNA families. Biochem. Biophys. Acta 425: 129-147.

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

Mazrimas,J.A., and Hatch,S.T. (1972) A possible relationship between satellite DNA and the evolution of kangaroo rat species (genus *Dipodimys ordii*). Nature 240: 102-105.

McKeon,C., Ohkubo,H., Pastan,I., and de Crombrugghe,B. (1982) Unusual methylation pattern of the a2(I) collagen gene. Cell 29: 203-210.

McNight,S.L. (1980) The nucleotide sequence and transcript map of the herpes simplex virus thymidine kinase gene. Nucleic Acids Res. 8: 5949-5964.

Miklos,G.L.C. (1985) Localized highly repetitive DNA sequences in vertebrate and invertebrate genomes. In Molecular Evolutionary Genetics, R.J. MacIntyre eds, Plenum, New-York, pp.241-321.

Moyzis,R.K., Albreight,K.L., Bartholdi,M.F., Cram,L.S., Deaven,L.L., Hildebrand,C.E., Joste,N.E., Lonhmire,J.F., Meyne,J., and Schwarzacher-Robinson,T. (1987) Human chromosome-specific repetitive DNA sequences: Novel markers for genetic analysis. Chromosoma 95: 375-386.

Moyzis,R.K., Torney,D.C., Meyne,J., Buckingham,J.M., Wu,J.-R., Burks,C., Sirotkin,K.M., and Goad,W.B. (1989) The distribution of interspersed repetitive DNA sequences in the human genome. Genomics 4: 273-289.

Mulligan,R.C., and Berg,P. (1980) Expression of a bacterial gene in mammalian cells. Science 209: 1422-1427.

Murnane,J.P., Yezzi,M.J., and Young,B.R. (1990) Recombination events during integration of transfected DNA into normal human cells. Nucleic Acids Res. 18: 2733-2738.

Musich,P.R., Maio,J.J., and Brown,F.L. (1977) Subunit structure of chromatin and the organization of eukaryotic highly repetitive DNA: indications of a phase relation between restriction sites and chromatin subunits in African green monkey and calf nuclei. J. Mol. Biol. 117: 657-677.

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

Ostrander,M., Vogel,S., and Silverstein,S. (1982) Phenotypic switching in cells transformed with the herpes simplex virus thymidine kinase gene. Mol. Cell. Biol. 2: 708-714.

Pardue,M.L., and Gall,J.G. (1970) Chromosomal localization of mouse satellite DNA. Science 168: 1356-1358.

Patriotis,C., and Djondjurov,L. (1989) Tightly bound DNA-protein complexes representing stable attachment sites of large DNA loops to components of the matix. Eur. J. Biochem. 184: 157-164.

Peacock,W.J., Lohe,A.R., GerlachW.L., Dunsmuir,P., Dennis,E.S., and AppelsR. (1978) Fine structure and evolution of DNA heterochromatin. Cold Spring Harbor Symp. Quant. Biol. 42: 1121-1135.

Pellicer,A., Wigler,M., and Axel,R. (1978) The transfer and stable integration of the HSV thymidine kinase gene into mouse cells. Cell 14: 133-141.

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

Plohl,M., Borstnik,B., Ugarkovic,C., and Gamulin,V. (1990) Sequence-induced curvature of *Tenebrio molitor* satellite DNA. Biochimie 72: 665-670.

Ponzetto-Zimmerman,C., and Wolgemuth,D.J. (1984) Methylation of satellite sequences in mouse spermatogenetic and somatic DNAs. Nucleic Acids Res. 12: 2807-2822.

Potter,H., Weir,L., and Leder,P. (1984) Enhancer-dependent expression of human K immunoglobulin genes introduced into mouse pre-B lymphocytes by electroporation. Proc. Natl. Acad. Sci. USA 81: 7176-7165.

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

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

Ptashne,M. (1988) How eukaryotic transcriptional activators work. Nature 335: 683-689.

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

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

Reeder,R.H. (1973) Transcription of chromatin by bacterial RNA polymerase. J. Mol. Biol. 80: 229-241.

Reuter,G., Giarre,M., Farah,J., Gausz,J., Spierer,A., and Spierer,P. (1990) Dependence of position-effect variegation in *Drosophila* on dose of a gene encoding an unusual zinc-finger protein, Nature 344: 219-223.

Robins,D.M., Ripley,S., Henderson,A.S., and Axel,R. (1981) Transforming DNA integrates into the host genome. Cell 23: 29-39.

Rosenberg,H., Singer,M.F., and Rosenberg,M. (1978) Highly reiterated se quences of SIMIAMSIMIANSIMIANSIMIANSIMIAN. Science 200: 394-402.

Sainz,J., and Cornudella,L. (1990) Preservation of a complex satellite DNA in two species of Echinoderms. Nucleic Acids Res. 18: 885-890.

Saitoh,H., Harata,M., and Mizumo,S. (1989) Presence of female-specific bent-repetitive DNA sequences in the genomes of turkey and pheasant and their interactions with W-protein of chicken. Chromosoma 98: 250-258.

Sambrook,J., Fritsch,E.F., and Maniatis,T. (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratories, Cold Spring Harbor, New-York.

Scangos,G.A., Huttner,K.M., Juricek,D.K., and Ruddle,F.H. (1981) Deoxyribonucleic acid-mediated gene transfer in mammalian cells: molecular analysis of unstable transformants and their progression to stability. Mol. Cell. Biol. 1: 111-120.

Selig,S., Ariel,M., Goitein,R., Marcus,M., and Cedar,H. (1988) Regulation of mouse satellite DNA replication time. Embo J. 7: 419-426.

Shafit-Zagardo,B., Broe,F.L., Zavodny,P.J., and Maio,J.J. (1983) Transcription of the KpnI families of long interspersed DNAs in human cells. Nature 304: 277-280.

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: 67-112.

Singh,L., anf Jones,K.W. (1986) Bkm sequences are polymorphic in humans and are clustered in pericentric regions of various acrocentric chromosomes including the Y. Hum. Genet. 73: 304-308.

Skinner,D.M. (1977) Satellite DNAs. Bioscience 27: 790-796.

Slagel,V.K., and Deininger,P.L. (1989) *In vitro* transcription of a cloned prosimian primate SINE sequences. Nucleic Acids Res. 17: 8699-8682.

Sloff,P., Bos,J.L., Konings,F.J.M., Borst,H.H., Borst,P., Gutteridge,W.E., and Leon,W. (1983) Characterization of satellite DNA in *Trypanosoma cruzi*. J.Mol. Biol. 167: 1-21.

Sol,K., 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 human KpnI fragments. Biochem. Biophys. Acta 868: 128-135.

Sol,K. (1991) Isolation and characterization of human highly repeated satellite II and III DNAs. Ph.D. Thesis, McGill University, Montréal.

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

Southern,E.M. (1970) Base sequence and evolution of guinea-pig a-satellite DNA. Nature 227: 794-798.

Southern,E.M. (1975a) Long range periodicities in mouse satellite DNA. J. Mol. Biol. 94: 51-69.

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

Southern,P.J., and Berg,P. (1982) Transformation of mammalian cells to antibiotic resistance with bacterial gene under control of the SV40 early region promoter. J. Mol. Appl. Genet. 1: 327-341.

Stein,R., Sciaky-Gallili,N., Razin,A., and Cedar,H. (1983) Pattern of methylation of two genes coding for house kepping functions. Proc. Natl. Acad. Sci. USA 80: 2422-2426.

Stephenson,E.C., Erba,H.P., and Gall,J.C. (1981) Histone gene clusters of the newt notophthalmus are separeted by long tracts of satellite DNA. Cell 24: 639-647.

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

Sweet,R.W., Chao,M.V., and Axel,R. (1982) The structure of the thymidine kinase promoter: Nuclease hypersensitivity correlates with expression. Cell 31: 347-353.

Sweet,R., Jackson,J., Lowy,L., Ostrander,M., Pelicer,A., Roberts,J., Robins,D., Sim,G.-K., Wold,B., Axel,R., and Silverstein,S. (1981) The expression, arrangement and rearrangement of genes in DNA-transformed cells. In Genes, Chromosomes and Neoplasia, Arrighi,F.F., Rao,P.N., and Stubblefield (eds) Raven Press, New-York.

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.

Taparowsky,E.J., and Gerbi,S.A. (1982) Structure of 1.711g/cm^3 bovine satellite DNA: evolutionary relationship to satellite I. Nucleic Acids Res. 10: 5503-5515.

Tartof,K.D., Hobbs,C., and Jones,M. (1984) A structural basis for variegating position effects. Cell 37: 869-878.

Tartof,K.D., Bishop,C., Jones,M., Hobbs,C., and Locke,J. (1989) Towards an understanding of position effect variegation. Dev. Genet. 10: 162-176.

Tasseron-de Jong,J., Aker,J., den Dulk,H., van de Putte,P., Giphart-Gassler,M. (1989) Cytosine methylation in the EcoRI site of active and inactive herpesvirus thymidine kinase promoters. Biochim. Biophys. Acta 1008: 62-70.

Thomas,R.J., Bolla,R.I., Rumbyrt,J.S., and Schlessinger,P. (1985) DNAseI-persistant non transcribed spacer segments of the mouse ribosomal DNA contain poly(dG-dT).poly(dA-dC). Proc. Natl. Acad. Sci. USA 82: 7595-7598.

Tomilin,N.V., Fedortseva,R.F., Svetlova,M.P., and Podgornaya,O.I. (1988) The presence in human lymphocytes of a protein that specifically binds a cloned fragment of human satellite DNA III. Mol. Genet. Microbiol. Virol. 4: 15-18.

Toneguzzo,F., Hayday,A.C., and Keating,A. (1986) Electric field-mediated DNA transfer: transient and stable gene expression in human and mouse lymphoid cells. Mol. Cell. Biol. 6: 703-706.

Towbin,H., Staehelin,T., and Gordon,J. (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedures and some applications. Proc. Natl. Acad. Sci. USA 76: 4350-4354.

Wahls,W.P., Wallace,L.J., and Moore,P.D. (1990) Hypervariable minisatellite DNA is a hot spot for homologous recombination in human cells. Cell 60: 95-103.

Wahls,W.P., Swenson,G., and Mooe,P.D. (1991) Two hypervariable minisatellite DNA binding proteins. Nucleic Acids Res. 19: 3269-3274.

Wallenburg,J.C., Nepveu,A., and Chartrand,P. (1987) Integration of a vector containing rodent repetitive elements in the rat genome. Nucleic Acids Res. 15: 7849-7863.

Wallenburg,J.C., Nepveu,A., and Chartrand,P. (1984) Random and non random integration of a polyomavirus DNA molecule containing highly repetitive cellular sequences. J. Virol. 50: 678-683.

Waye,J.S., and Willard,H.F. (1987) Nucleotide sequence heterogeneity of alpha satellite repetitive DNA: a survey of alphoid sequences from different human chromosomes. Nucleic Acids Res. 15: 7549-7569.

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

Wevrick,R., Earnshaw,W.C., Howard-Peebles, H.N., and Willard,H F. (1990) Partial deletion of alpha satellite DNA associated with reduced amounts of the centromere protein CENP-B in a mitotically stable human chromosme rearrangement. Mol. Cell. Biol. 10: 6347-6380.

Wigler,M., Silverstein,S., Lee,L.-S., Pelicer,A., Cheng,Y.-C., and Axel,R. (1977) Transfer of purified herpes simplex virus thymidine kinase gene to cultured mouse cells. Cell 11: 223-232.

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

Wolff,S., Schrader,M., and Wittig,B. (1991) Lack of correlation between DNA methylation and transcriptional inactivation: the chicken lysozyme gene. Proc. Natl. Acad. Sci. USA 88: 271-275.

Wong,A.K.C., and Rattner,J.B. (1988) Sequence organization and cytological localization of the minor satellite of mouse. Nucleic Acids Res. 16: 11645-11661.

Wu,K.C., Strauss,F., and Varshavsky,A. (1983) Nucleosome arrangement in green monkey a-satellite chromatin. Superimposition of non-random and apparently random patterns. J. Mol. Biol. 170: 93-117.

Yamamoto,M., and Miklos,G.L.G. (1977) Genetic dissection of heterochromatin in *Drosophila*: the role of basal X heterochromatin in meiotic sex chromosome behavior. Chromosoma 66: 283-296.

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

Yen,P.H., Li,X.-M., Tsai,S.-P., Johnson,C., Mohandas,T., and Shapiro,L.J. (1990) Frequent deletions of the human X chromosome distal short arm result from recombination between low copy repetitive elements. Cell 61: 603-610.

Zhang,X.-Y., and Horz,W. (1984) Nucleosomes are positioned on mouse satellite DNA in multiple highly specific frames are correlated with a diverged subrepeat of nine base-pairs. J. Mol. Biol. 176: 105-129.