INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

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

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

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

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

Studies on the inhibition of establishment of stable transfection by a cloned human satellite DNA

By

Djenann Saint-Dic

Department of Microbiology and Immunology
McGill University, Montreal
August, 1999

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

ABSTRACT

Eukaryotic genomes possess significant amounts of highly tandemly-arrayed repetitive DNA sequences in their constitutive heterochromatin. Heterochromatin has been involved in processes, such as position-effect variegation (PEV), that disrupt normal expression of euchromatic genes. The role of highly repetitive DNA sequences, also termed satellite DNAs, in gene silencing has therefore aroused a great deal of interest. Previous plasmid transfection studies into tk^-/neo^- human cells had shown that the location (in *cis*) of a 1.797 kb human satellite II DNA sequence next to the 5' end of the *neo* gene, exerted a severe negative effect on the recovery of stable tk^-/neo^- transfectants, in the presence of HAT and G418 drugs. As shown in this thesis, positioning of the same satellite II DNA sequence at another location, between the 3' ends of the *tk* and *neo* genes, led to the loss of this negative effect. To identify a possible size limit to the silencing effect exerted by the satellite II DNA, fragments of 813 and 983 bp of the 1.797 kb DNA sequence were also inserted near the 5' end of the *neo* gene. These sequences were still able to exert a drastic negative effect on the recovery of stable tk^-/neo^- transfectants. The effect was found to be dependent on the orientation of the satellite II DNA fragments relative to the marker genes. Further reduction in the size of the satellite II DNA sequence, to fragments of 354 to 620 bp, revealed a significant reduction in the negative effect. Selection of the transfected cells in media containing single drugs (i.e. HAT or G418) showed that the *neo* gene in the plasmids was more affected than the *tk* gene. Analysis of the integration pattern of the plasmid vector and a plasmid containing the 1.797 kb satellite II DNA indicated that the satellite II DNA did not influence the plasmid site of integration.

RÉSUMÉ

Le génome des eukaryotes contient dans l'hétérochromatine constitutive une importante quantité de séquences d'ADN hautement répétitives organisées en tandem. L'hétérochromatine a été impliquée dans des processus, tel que le PEV ("position-effect variegation"), capables de perturber l'expression normale de gènes euchromatiques. Le rôle de séquences d'ADN hautement répétitives, aussi appelées ADN satellite, dans la défectuosité de l'expression de gènes a donc suscité un grand intérêt. Des études de transfection de plasmides, dans des cellules humaines *tk*$^-$/*neo*$^-$, ont précedemment démontré que l'emplacement (en *cis*) d'une séquence d'ADN satellite II humain de 1.797 kb près du gène *neo* a eu un effet néfaste sur le rendement de transfectants stables, en présence des composés HAT et G418. Il est demontré dans cette thèse que l'insertion de cette même séquence d'ADN satellite II à un emplacement différent, entre les extrémités 3' des gènes *neo* et *tk*, a conduit à l'élimination de l'effet néfaste. Pour déterminer s'il y a une limite de grandeur quant à l'effet créé par l'ADN satellite II, des fragments de 813 et 983 pb de la séquence d'ADN de 1.797 kb ont aussi été insérés près de l'extrémité 5' du gène *neo*. Ces séquences ont encore pû exercer un effet néfaste sur le rendement de transfectants stables *tk*$^-$/*neo*$^-$. Cet effet s'est avéré dépendre de l'orientation des fragments d'ADN satellite II par rapport aux gènes marqueurs. Une réduction additionnelle de la taille de la séquence d'ADN satellite II en des fragments de 354 à 620 pb a démontré une grande atténuation de l'effet négatif. La sélection des cellules transfectées en présence de milieux contenant chacun des composés, HAT ou G418 individuellement, a montré que le gène *neo* des plasmides était plus affecté que le gène *tk*. L'analyse du mode d'intégration du plasmide vecteur et d'un plasmide contenant l'ADN satellite II de 1.797 kb a indiqué que l'ADN satellite II n'a pas influencé le site d'intégration du plasmide.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisor, Dr. DuBow, for allowing me to carry my studies as a member of his laboratory. His continuous inspiration has been very enriching and has helped me further my scientific knowledge. I also want to thank all members of the DuBow laboratory past and present: David Alexander, Dr. Josée Brisebois, Nicha Chaorensri, Dr. Angelina Guzzo, Julie Guzzo, Dr. Gina McIntyre, Manuelle Rongy, Ian Siboo, Felix Sieder, Madani Thiam, Dr. Peter Ulycznyj. I would also like to thank Drs Nicolas Acheson and Michael Ratcliffe for helpful discussions and comments.

My most sincere thanks go to Caroline Diorio for her technical assistance and for helping me with analysis of the data. Her numerous suggestions and discussions have been precious to me. I would also like to thank her for her constant support that has proved invaluable. I am forever grateful to Kirsty Salmon for her friendship and her assistance in so many aspects during the course of my studies. Her words of encouragement and her immense help have been priceless. I would like to thank her for final suggestions and corrections to the thesis. I also want to thank Dr. Hans W. Ackermann for his continuous help and for his helpful suggestions and advice.

I want to express my very special thanks to my parents, whose presence and support are precious. I want to thank them for their sacrifices, for guiding me through life and teaching me perseverance and determination. These lessons will accompany me forever. My thanks also go to my sister, Fabienne, for the complicity we share, and for helping me relieve the stress in so many occasions. Finally my very sincere thanks go to my husband, Carl-Félix, for his unconditional support, his invaluable patience, for knowing how to listen and for helping me to believe in my abilities.

I would also like to thank the "Fonds pour la formation de chercheurs et l'aide à la

recherche" (FCAR), the Department of Microbiology and Immunology of McGill University (F.C. Harrison Award), as well as my supervisor for providing funds for my research project.

CONTRIBUTIONS TO ORIGINAL KNOWLEDGE

1. By transfection of a pSVL-based plasmid, containing a 1.797 kb *Eco*RI human satellite II DNA sequence inserted between the 3' ends of the *neo* and *tk* genes, I have shown that a severely negative effect on transfection efficiency is not observed, as opposed to the effect obtained when this sequence is located at the 5' end of the *neo* gene.

2. By transfection of various plasmid constructs, I have shown that satellite II DNA fragments of 813 bp and 983 bp, inserted near the 5' end of the *neo* gene in the plasmids, are capable of exerting the same negative effect on the recovery of stable transfectants observed with the 1.797 kb satellite II DNA sequence.

3. By transfection of the various plasmid constructs containing satellite II DNA sequences of 813 and 983 bp, I have shown that the severely negative effect exerted by satellite II DNA is only observed with particular orientations of the fragments relative to the marker genes in the plasmids.

4. By transfection of several plasmid constructs containing satellite II DNA fragments of 354, 364, 458 and 620 bp, I have shown that the very severe effect produced by the satellite II DNA on the recovery of stable transfectants starts to disappear with these smaller satellite II DNA sequences.

5. I have shown that plasmids containing 354, 458, and 620 bp satellite II DNA fragments display an orientation-dependent effect on the recovery of stable transfectants from transfection experiments.

6. By subjecting transfected cells to the presence of single drugs (HAT or G418), I have shown that the *neo* gene, in plasmids containing the full size or fragments of the 1.797

kb satellite II DNA sequence, is more affected than the *tk* gene by the effect produced by the satellite II DNA, as indicated by transfection efficiencies in media containing G418.

7. I have examined, by Southern blotting, the integration pattern of the transfected plasmid vector pSVL and the plasmid containing the 1.797 kb satellite II DNA sequence in various cell lines. I have shown that there was not a specific site in the plasmids used for integration, and that integration occurred at different chromosomal locations.

PREFACE

The work presented in this thesis represents the original contributions of the author, Djenann Saint-Dic. Some information was derived from the following publication: Saint-Dic D. and DuBow M.S. 1995. Effect of a *cis*-located human satellite DNA on electroporation efficiency. *Methods in Molecular Biology* 48: 199-210. All plasmids mentioned in this 1995 article, as well as transfection efficiency results, were the work of Claire Fouquet. The manuscript was prepared by the first author, Djenann Saint-Dic. Plasmids pSVL, pCFD1 and pCF1.8 used in the present thesis, and in the article "Saint-Dic D. and DuBow M.S. Analysis of the inhibition of stable plasmid transfection by a cloned human satellite DNA" that will be submitted for publication, were constructed by Claire Fouquet. A review article is also in preparation to be submitted for publication: Saint-Dic D. and DuBow M.S. The presence and significance of satellite DNAs in mammalian cells. Both articles in preparation are based on material used in this thesis. I would like to thank Caroline Diorio for her help in the statistical analyses presented in this thesis.

TABLE OF CONTENTS

ABSTRACT.....i

RÉSUMÉ.....ii

ACKNOWLEDGEMENTS.....iii

CONTRIBUTIONS TO ORIGINAL KNOWLEDGE.....v

PREFACE.....vii

TABLE OF CONTENTS.....viii

LIST OF FIGURES AND TABLES.....xi

LIST OF ABBREVIATIONS.....xiii

CHAPTER 1.....1

INTRODUCTION.....1

1.1. ORGANIZATION OF CHROMOSOMAL DNA.....1
1.1.1. MODERATELY REPETITIVE DNA SEQUENCES.....4
1.1.2. HIGHLY REPETITIVE DNA SEQUENCES.....6

1.2. DISCOVERY AND CHARACTERIZATION OF SATELLITE DNA.....8

1.3. PARTICULAR FEATURES OF SATELLITE DNA.....12
1.3.1. QUANTITY OF SATELLITE DNA IN EUKARYOTIC GENOMES.....12
1.3.2. GENETIC AND EPIGENETIC ASPECTS OF SATELLITE DNA.....14

1.4. SEQUENCE AND EVOLUTION OF SATELLITE DNA.....19
1.4.1. SALTATORY REPLICATION.....20
1.4.2. UNEQUAL CROSSING-OVER.....23
1.4.3. CONCERTED EVOLUTION.....23

1.5. ROLES ATTRIBUTED TO SATELLITE DNA.....25
1.5.1. CENTROMERE STRUCTURE.....26
1.5.2. ANCHORAGE TO THE NUCLEAR MATRIX.....28
1.5.3. SATELLITE DNA AS A SPACER.....29
1.5.4. BINDING TO SPECIFIC PROTEINS.....30
1.5.5. CENTROMERE FUNCTION.....33
1.5.6. CHROMOSOME SEGREGATION.....36
1.5.6.1. THE USE OF MAMMALIAN ARTIFICIAL CHROMOSOMES TO DEFINE CENTROMETRIC ELEMENTS.....36
1.5.6.2. SATELLITE DNA AND THE DEFINITION OF CENTROMERE IN MULTICELLULAR EUKARYOTES.....37
1.5.7. SATELLITE DNA TRANSCRIPTION.....44

1.6. EFFECTS OBSERVED WITH REPETITIVE DNA SEQUENCES....45
1.6.1. POSITION-EFFECT VARIEGATION....45
1.6.2. POSITION EFFECTS OBSERVED WITH ISOLATED GENES AND/OR SATELLITE DNA SEQUENCES....53

1.7. OUTLINE OF THE THESIS....56

CHAPTER 2....57

MATERIALS AND METHODS....57

2.1. PLASMIDS....57
2.1.1. RESTRICTION ENZYME HYDROLYSIS AND DNA MODIFICATIONS....57
2.1.2. ISOLATION OF DNA FRAGMENTS....58
2.1.3. DNA LIGATION....59
2.1.4. SCREENING OF THE DESIRED PLASMID CONSTRUCTS....59

2.2. ORIGIN OF PLASMID VECTOR AND DNA INSERTS....60
2.2.1. PLASMID VECTORS....60
2.2.2. DNA INSERTS....63

2.3. ISOLATION OF PLASMID CONSTRUCTS....66
2.3.1. SMALL-SCALE RAPID ISOLATION OF PLASMID DNA....66
2.3.2. LARGE-SCALE ISOLATION OF PLASMID DNA....67

2.4. CELL CULTURE....67
2.4.1. THE AK143B CELL LINE....67
2.4.2. MEDIA USED FOR PROPAGATION OF CELL LINES....67
2.4.3. CULTURE ENVIRONMENT....68
2.4.4. PRESERVATION OF CELL LINES....68
2.4.5. SUBCULTURING OF CELLS....69

2.5. ELECTROPORATION OF THE AK143B CELL LINE....69
2.5.1. PLASMID DNA PURIFICATION....69
2.5.2. ELECTROPORATION PROCEDURE....70
2.5.3. DRUG SELECTION AND COLONY STAINING....73

2.6. PROPAGATION OF CELL LINES FROM INDIVIDUAL COLONIES....73

2.7. SOUTHERN BLOTTING....74
2.7.1. ISOLATION OF TOTAL GENOMIC DNA FROM HUMAN CELLS....74
2.7.2. SOUTHERN BLOTTING TECHNIQUE....74
2.7.3. REMOVAL OF RADIOACTIVELY-LABELLED DNA PROBES FROM SOUTHERN BLOTS....75

2.8. NORTHERN BLOTTING....75
2.8.1. ISOLATION OF TOTAL RNA....75
2.8.2. RNA GEL ELECTROPHORESIS....76
2.8.3. NORTHERN BLOTTING TECHNIQUE....76

2.8.4. PREPARATION OF POLY(A)$^+$ RNA.....77
2.8.5. REMOVAL OF RADIOACTIVELY-LABELLED DNA PROBES FROM NORTHERN BLOTS.....77

2.9. CONSTRUCTION OF RADIOACTIVELY-LABELLED DNA PROBES.....78

CHAPTER 3.....80

RESULTS.....80

3.1. FREQUENCY OF TRANSFECTION OF PLASMID CONSTRUCTS CONTAINING SATELLITE II DNA SEQUENCES.....80
3.1.1. PLASMIDS CONTAINING THE 1.797 KB HUMAN SATELLITE II DNA SEQUENCE AT DIFFERENT LOCATIONS.....80
3.1.2. TRANSFECTION EFFICIENCIES WITH PLASMIDS CONTAINING AN 813 BP OR 983 BP FRAGMENT OF THE 1.797 KB SATELLITE II DNA SEQUENCE.....84
3.1.3. TRANSFECTION EFFICIENCY WITH PLASMIDS CONTAINING EVEN SMALLER FRAGMENTS OF THE 1.797 KB SATELLITE II DNA SEQUENCE.....87
3.1.4. EFFECT ON THE REPORTER GENES.....95

3.2. ANALYSIS OF PLASMID INSERTION PATTERN INTO CHROMOSOMAL DNA.....101
3.2.1. SOUTHERN BLOTTING WITH CELLS TRANSFECTED WITH pSVL.....101
3.2.2. CELLS CONTAINING THE pCFD1 PLASMID.....119

3.3. ANALYSIS OF SHORT-TERM EXPRESSION OF THE TRANSFECTED TK GENE.....131

CHAPTER 4.....137

DISCUSSION.....137

REFERENCES.....156

APPENDIX.....174

LIST OF FIGURES AND TABLES

FIGURES

1.1. Simple representation of an eukaryotic chromosome showing its basic chromatin organization....2
1.2. Presence of satellite DNA fractions around the main DNA peak following CsCl density gradient ultracentrifugation....9
1.3. Models representing A) underacetylation or B) hypermethylation of heterochromatic regions of eukaryotic genomes associated with genetic inactivation....16
1.4. Models for the occurrence and evolution of satellite DNA sequences....21
1.5. Compaction and association of chromosomes through joining of satellite DNA regions mediated by binding of CENP-B proteins to CENP-B boxes....31
1.6. Association of centromeric proteins with the primary constriction....34
1.7. Participation of flanking satellite DNA sequences in the assembly and stabilization of sister chromatids and of a functional kinetochore....39
1.8. Model showing disturbance of gene expression induced by DNA rearrangements and spreading of heterochromatin 47
1.9. Model for *trans*-inactivation of a gene following DNA rearrangements....50

2.1. Plasmid vector pSVL used for construction of the various plasmids 61
2.2. Nucleotide sequence of the 1.797 kb *Eco*RI human satellite II DNA....64
2.3. Outline of the transfection and drug selection protocol used for plasmid introduction and maintenance into AK143B human cells....71

3.1. Plasmid pDH1, containing a 1.797 kb *Eco*RI satellite II DNA sequence....81
3.2. Constructed plasmids containing either the 813 bp (A) or 983 bp (B) fragments of the 1.797 kb *Eco*RI satellite II DNA sequence....85
3.3. Size and position of restriction fragments from the 1.797 kb satellite II DNA sequence and plasmids obtained following insertion into the *Stu*I site of pSVL 90
3.4. Plasmid constructs containing small restriction fragments from the 1.797 kb satellite II DNA sequence inserted into the *Stu*I site of pSVL....92
3.5. Construction of plasmid pDSL1 used as size control, containing a 604 bp fragment from bacteriophage λ DNA inserted at the *Stu*I site of pSVL 96
3.6. Southern blot analysis of total cellular DNA from the SVLG-1 cell line 102
3.7. Southern blot analysis of total cellular DNA from the SVLG-6 cell line 106
3.8. Southern blot analysis of total cellular DNA from the SVLH-6 cell line 109
3.9. Southern blot analysis of total cellular DNA from the SVLHG-2 cell line....111
3.10. Southern blot analysis of total cellular DNA from the SVLHG-4 cell line....114
3.11. Southern blot analysis of total cellular DNA from the SVLHG-5 cell line....117
3.12. Southern blot analysis of total cellular DNA from the DIG-2 cell line....120
3.13. Southern blot analysis of total cellular DNA from the DIG-3 cell line....123
3.14. Southern blot analysis of total cellular DNA from the DIH-2 cell line....126
3.15. Southern blot analysis of total cellular DNA from the DIH-3 cell line....129
3.16. Northern blot analysis of total RNA from cells transfected with plasmids pSVL, pCFD1 or pDSD2....132
3.17. Northern blot analysis of poly(A)$^+$ RNA from cells transfected with plasmids pSVL, pCFD1 or pDSD2....134

4.1. Proposed direction of heterochromatinization in the 983 and 813 bp fragments of the 1.797 kb satellite DNA sequence.....141
4.2. Proposed direction of heterochromatinization in the various fragments of the 1.797 kb satellite DNA sequence.....144
4.3. Representation of the position of curved regions in the various fragments of the 1.797 kb satellite DNA sequence.....147

TABLES

1.1. Classification of diverse repetitive DNA sequences present in mammalian genomes.....5
1.2. Possible events implicated in the acquisition of centromere function by acentric chromosomes.....43

3.1. Frequency of transfection (± standard deviation) with plasmids containing the 1.797 kb *Eco*RI satellite II DNA sequence at the *Stu*I and *Hpa*I sites into media containing HAT+G418.....83
3.2. Frequency of transfection (± standard deviation) by plasmids containing either the 813 bp or 983 bp satellite DNA fragments, into media containing HAT+G418.....88
3.3. Frequency of transfection (± standard deviation) by plasmids containing small fragments of the 1.797 kb *Eco*RI satellite II DNA, into media containing HAT+G418.....94
3.4. Frequency of transfection (± standard deviation) by pCFD1, which contains the 1.797 kb satellite II DNA sequence, and by plasmids containing 813 bp and 983 bp satellite DNA fragments, in media containing HAT+G418, HAT or G418 98
3.5. Frequency of transfection by plasmids containing various small fragments of the 1.797 kb *Eco*RI satellite II DNA in media containing HAT+G418, HAT or G418 100

4.1. List of the copy number and sites of integration of integrated plasmids in various cell lines.....150

LIST OF ABBREVIATIONS

ATCC:	American Type Culture Collection
ATP:	adenosine triphosphate
bla:	ampicillin resistance gene
bp:	base pair(s)
BSA:	bovine serum albumin
CENP:	centromere protein
CIAP:	calf intestinal alkaline phosphatase
CREST:	Calcinosis, Raynaud's phenomenon, Esophageal dysmotility, Sclerodactyly telangiectasiae
CsCl:	cesium chloride
dCTP:	deoxycytidine triphosphate
DEPC:	diethyl pyrocarbonate
DMEM:	Dulbecco's modified Eagle medium
DMSO:	dimethyl sulfoxide
DNA:	deoxyribonucleic acid
dNTP:	deoxyribonucleotide triphosphate
DTT:	dithiothreitol
EDTA:	ethylene diaminetetraacetic acid
FBS:	fetal bovine serum
HAT:	hypoxanthine, aminopterin, thymidine
HBS:	HEPES-buffered saline
HSV-1:	herpes simplex virus type 1
ICF:	Immunodeficiency, Centromere instability, Facial abnormalities
kb:	kilobase pair(s)
LB:	Luria-Bertani medium
LINES:	long interspersed sequences
LTAg:	Large T antigen
MAC:	mammalian artificial chromosome
min:	minute(s)
MOPS:	3-[N-Morpholino]propanesulphonic acid
neo:	neomycin resistance gene
ORF:	open reading frame
PAGE:	polyacrylamide gel electrophoresis
PBS:	phosphate-buffered saline
Pen-Strep:	penicillin-streptomycin
PEV:	position-effect variegation
RNA:	ribonucleic acid
rpm:	revolutions per minute
Sat:	satellite DNA
SDS:	sodium dodecyl sulfate
SINES:	short interspersed sequences
tk:	thymidine kinase gene
tRNA:	transfer RNA
μF:	microFarads

CHAPTER 1

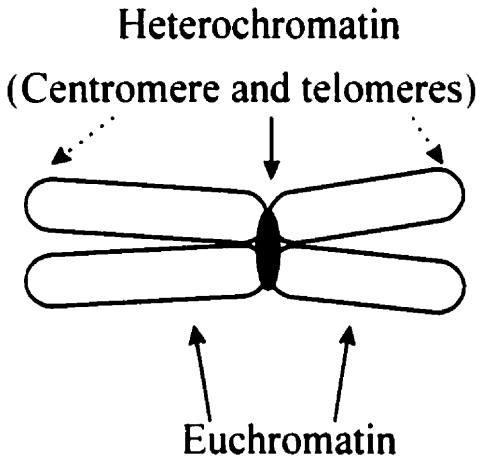
INTRODUCTION

1.1. Organization of Chromosomal DNA

Eukaryotic cells possess a well developed membrane system. They have evolved a nucleus that contains most of the cell's DNA, keeping it separate from the other contents of the cell present in the cytoplasm. The chromosomal DNA is divided into separate molecules, each packaged in an individual chromosome. A eukaryotic cell's genetic material is very extensive. The human haploid genome contains about $3x10^9$ nucleotide pairs carried on 23 chromosomes (Felsenfeld, 1985). The human karyotype therefore consists of a total of 46 chromosomes. All chromosomes are linear and possess at each end a telomere, and located more centrally, a single centromere, which appears as a constriction in metaphase. The chromosomes are packed and organized in the nucleus by a complex of proteins that together constitute the chromatin. A specially condensed form of chromatin, known as heterochromatin, has been identified at the centromere and telomeres (Brown, 1966). Euchromatin, which is less condensed, constitutes the rest of the chromatin (Figure 1.1).

The euchromatin can be separated into its principal constituents that comprise both unique and moderately repetitive DNA sequences. Unique DNA sequences contain the genetic coding material, along with a variety of diverse regulatory elements. These sequences are interspersed with the moderately repetitive DNA sequences, which represent the tandem arrangement of unit DNA sequences over long uninterrupted stretches. Based on the length of these stretches and their proportion in the genome, repetitive DNA sequences have been separated into two main classes: moderately repetitive sequences,

Figure 1.1. Simple representation of an eukaryotic chromosome showing its basic chromatin organization. Heterochromatin is located at the centromere (center) and telomeres (ends). Euchromatin is found at the chromosome arms.


Heterochromatin
(Centromere and telomeres)
Euchromatin

located between genes in euchromatin and dispersed in heterochromatin, and highly repetitive sequences mainly present in heterochromatin (Miklos, 1985; Smit, 1996).

1.1.1. Moderately Repetitive DNA Sequences

Transcribed genes exhibiting dosage repetition such as tRNA, ribosomal RNA (rRNA) genes (Long and Dawid, 1980) and histone genes (Kedes, 1979) are represented among moderately repetitive DNAs. However, in mammalian genomes, the majority of these sequences, which comprise 30% of the human genome (Smit, 1996), are composed of retroposons. They consist in two main families: the LINEs (long interspersed elements) and the SINEs (short interspersed elements) (Singer, 1982). LINEs in mammalian genomes are represented by the LINE-1 (L1) family, which is thought to have originated as a mobile gene, and presents features of a processed pseudogene (Fanning and Singer, 1987). The unit sequence of the L1 elements measures around 6,000 base pairs (Adams *et al.*, 1980) (Table 1.1). It is present at a level of 10^5 copies per haploid genome. Other members of LINEs, more than 10^4 family members in total, contain deletions or rearrangements in their unit sequence which also encompasses 6,000 bp. LINEs encode a reverse transcriptase (Mathias *et al.*, 1991) as well as an endonuclease (Feng *et al.*, 1996; Jurka, 1997). Active L1 elements were identified in human (Dombroski *et al.*, 1991) and mice (Martin, 1991) by detection of de novo insertions of truncated L1 on a human X chromosome and the presence of ribonucleoprotein particles, containing LINE-1 RNA, in a mouse embryonal carcinoma cell line. Expression of these sequences is part of the retrotransposition process that has produced the high copy number of the L1 element in mammalian genomes.

The most prominent SINEs of the human genome are the *Alu* elements consisting of repeat units of 80 to 400 base pairs (Singer, 1982). There are nearly 10^6 copies of *Alu* elements per haploid genome (Schmid, 1996). The *Alu* family, along with the rodent B1 families of SINEs, is derived from a component of the signal recognition particle (SRP), the

Table 1.1. Classification of diverse repetitive DNA sequences present in mammalian genomes.

	Moderately repetitive DNA sequences		Highly repetitive DNA sequences		Simple repeated DNA sequences	
	LINEs (L1) (Singer, 1982)	SINEs (eg. *Alu*, B1) (Singer, 1982)	Satellite DNA (Beridze, 1986)		Minisatellite DNA (Tautz, 1993)	Microsatellite DNA (Tautz, 1993)
			Complex satellite DNA	Simple satellite DNA		
Size of repeat unit (bp)	6,000	80-400	100-1400	2-10	9-100	1-6
Copy number per haploid genome	10^4	10^6	10^3-10^7	10^3-10^7	2-hundreds	5-100
Location in genome	dispersed	dispersed	centromere	centromere	dispersed	dispersed

7SL RNA (Ullu and Tschudi, 1984; Weiner, 1980). Members of the *Alu* family are flanked by short direct repeat sequences. These flanking direct repeats are also found in LINEs (Jurka, 1997). Other families of SINEs are derived from tRNAs, and possess a 5' region homologous to tRNA, a middle tRNA-unrelated region and a 3' A-rich tract (Ohshima *et al.*, 1993; Okada, 1991). Mammalian-wide interspersed repeats (MIRs) constitute prominent tRNA-derived SINEs in humans, and are found in all mammalian genomes (Jurka *et al.*, 1995). Small cytoplasmic transcripts of *Alu* and B1 elements have been observed (Maraia and Sarrowa, 1995). However, unlike LINEs, SINEs do not encode a reverse transcriptase or an endonuclease, factors that could be provided to them by LINEs (Okada *et al.*, 1997). In fact, several pairs of SINEs and LINEs were shown to have the same tail sequence at their 3' ends (Okada *et al.*, 1997). The 3' ends of several families of SINEs are derived from the 3' ends of corresponding LINEs, present in the same organisms (Ohshima *et al.*, 1996).

1.1.2. Highly Repetitive DNA Sequences

Two main types of heterochromatin are known in mammals: the constitutive heterochromatin present on homologous chromosomes (Heitz, 1928) and the facultative heterochromatin resulting from the inactivation of one of the two X chromosomes in females (Barr and Bertram, 1949; Grant and Chapman, 1988). Constitutive heterochromatin is found around the centromere of chromosomes, at telomeres, and on the Y chromosome (Miklos, 1985; Yunis and Yasmineh, 1971). The peculiar property of the constitutive heterochromatic portion of the genome is that it remains condensed throughout the whole cell cycle, including interphase, replicates late in S-phase, and is usually not transcribed. The constitutive heterochromatin is composed to a large extent of highly repeated DNA sequences without any protein-coding capacity (Miklos, 1985). However, in *Drosophila*, heterochromatin contains a number of genes located within regions of high or middle

repetitive DNAs (Gatti and Pimpinelli, 1992), some of which have vital functions (Hilliker *et al.*, 1980). The highly repetitive fraction of the genome is designated, around the centromere, as satellite DNA (Yunis and Yasmineh, 1971). Satellite DNAs are currently defined as tandemly repeated DNA sequences present in 10^3 to 10^7 consecutive repetitions at each locus (Tautz, 1993).

Based on the length of the basic repeat unit, satellite DNAs can be separated into complex or simple sequence satellites. Complex satellite DNAs are composed of long unit sequences (Table 1.1). The alphoid satellite DNA is one example of a complex satellite DNA. In the African Green Monkey, alpha satellite DNA has a monomeric unit length of 172 bp. In *Drosophila melanogaster*, the unit is 359 bp. Sometimes the repeat units of complex satellite DNAs may be even longer, for instance in the amphibian, *Xenopus laevis*, where it is around 740 bp (Beridze, 1986). These long monomeric unit sequences are usually not internally repetitive. Simple sequence satellite DNAs are defined by their shorter unit sequences, varying between 2 and 10 bp (Table 1.1). Examples of simple satellite DNAs include the 5 bp 5'-TTCCA-3' unit of human satellite DNA II (Beridze, 1986; Sol *et al.*, 1986), or the 4 bp 5'-TAGG-3' unit of the hermit crab satellite DNA I (Beridze, 1986). Within each category, complex or simple satellite DNAs usually consist of repeat units of different base composition, which may or may not be represented on all the chromosomes.

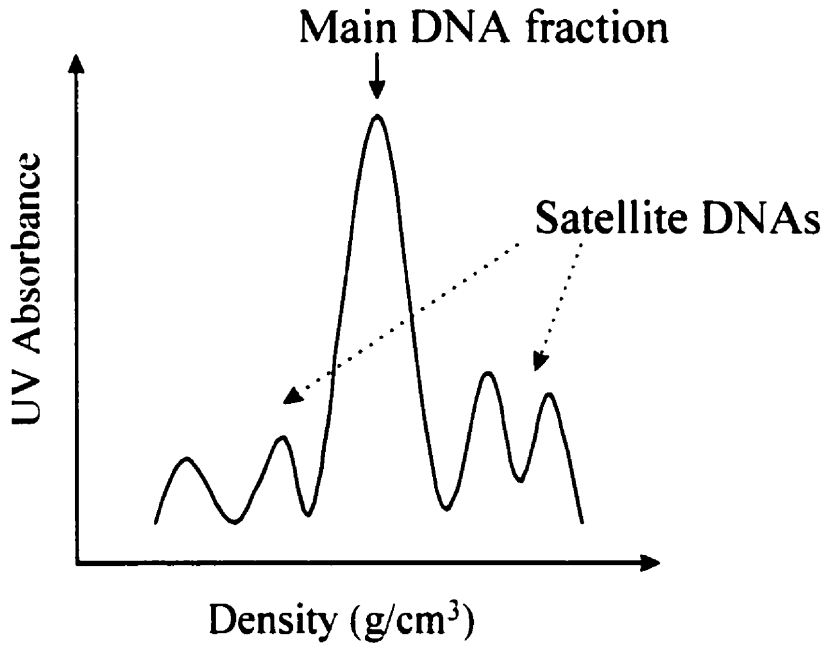
There is a distinction to be made between simple sequence satellite DNAs and two other classes of simple repeated DNA sequences: the minisatellite and the microsatellite DNAs (Table 1.1). Minisatellite DNA sequences consist of longer repeat units of the range of 9 to 100 bp (Tautz, 1993). These units have a moderate degree of repetition, forming stretches of about two to several hundred units at each locus. Minisatellites are more dispersed throughout the genome and are found interspersed in the euchromatic regions of the genome of vertebrates, fungi and plants, with a tendency to cluster near telomeric

regions in humans (Royle *et al.*, 1988). In addition, minisatellite sequences are highly variable in their array size, tending to be associated with length polymorphisms, contrary to simple sequence satellite DNAs that are much less variable. Microsatellite sequences are also composed of arrays of short unit sequences 1 to 6 bp in length. The microsatellite arrays fluctuate between 5 to about 100 repetitions at each locus (Tautz, 1993). They are found scattered throughout the vertebrate, insect and plant genomes. In humans, there are about 30,000 microsatellite loci and they are located in the euchromatin (Charlesworth *et al.*, 1994).

1.2. Discovery and Characterization of Satellite DNA

Density gradient ultracentrifugation was the first tool that allowed scientists to observe the dual state, comprised of unique sequences and repetitive fractions, of higher eukaryotic DNAs. The first eukaryotic DNA to be analyzed in CsCl density gradients, calf thymus DNA, showed an asymmetrical distribution that was an indication of heterogeneity (Meselson *et al.*, 1957). This asymmetry was shown to be due to the difference in density between the main DNA fraction and a satellite component. Using the same type of gradients, later observations from a number of animals revealed minor fractions separate from the main band of DNA (Figure 1.2). These separate fractions were termed "satellite" DNA by Kit (1961). The basis of the observed discrepancy was demonstrated to lie in the different GC contents of the DNA fractions, the density of DNA in CsCl being directly proportional to this parameter. DNA preparations containing fractions of sheared DNA fragments with a different base composition will, in such gradients, separate into several bands. Therefore, depending on whether the different satellite DNAs are poorer or richer in GC base pairs than the main DNA band, they would appear as separate peaks around it.

Figure 1.2. Presence of satellite DNA fractions around the main DNA peak following CsCl density gradient ultracentrifugation.


Main DNA fraction
Satellite DNAs
UV Absorbance
Density (g/cm3)

There also exist so-called "hidden" satellite DNAs that migrate with the major DNA fraction in CsCl gradients (Beridze, 1986). These satellite DNAs can be separated using heavy metal ions (e.g. Ag^+ or Hg^+) in Cs_2SO_4 gradients because Cl ions inhibit the binding of the heavy metal ions to the DNA (Corneo *et al.*, 1971). Reassociation kinetics studies have shown that there was a very rapid reannealing of a particular region of the separate DNA chains of higher organisms and that this was due to the high repetition of the DNA (Britten and Khone, 1966; Waring and Britten, 1966). This region was found to be made of satellite DNA. Following these observations, satellite DNA was defined not only as a minor fraction of chromosomal DNA giving a separate band in CsCl gradients, but also as a fraction with tandemly arranged highly repetitive short sequences. It was also found to be specific to eukaryotes, as it was not observed in prokaryotic organisms.

An additional property of satellite DNAs is the ladder of fragments they generate, often over a smear of total DNA, upon digestion of total nuclear DNA by particular restriction endonucleases (Pech *et al.*, 1979a). These fragments, which are observable after gel electrophoresis, are multiples of the basic repeat unit which is found in enough repetitions to form several apparent bands of various sizes. One fragment will be obtained for every repeating unit in which the restriction site occurs. As will be seen later, mutation events cause the elimination of certain restriction sites along the sequence, leading to the occurrence of larger fragments which are multimers (dimers, trimers, etc.) of the basic repeat unit. Because of the high degree of repetition of satellite DNA sequences, a large number of fragments of identical or almost identical size are created, leading to the appearance of sharp bands.

The restriction enzyme pattern of satellite DNAs constitutes one basis for their classification. Satellite DNAs I to IV are the major classes of satellite sequences in humans, along with the alphoid satellite DNA (Beridze, 1986). One of the features that distinguishes satellite I, II and III DNAs is their digestion pattern with *Hinf*I. The smallest fragment

produced by digestion of satellite I DNA with *Hinf*I is 770 bp, whereas satellite II and III DNAs are almost completely digested with the same enzyme, yielding 10 to 80 bp and 15 to 250 bp fragments respectively (Frommer *et al.*, 1982). In addition, what distinguishes satellite II DNA from satellite III DNA is the presence of *Taq*I sites in the former and their absence in the latter (Frommer *et al.*, 1982).

1.3. Particular Features of Satellite DNA

1.3.1. Quantity of Satellite DNA in Eukaryotic Genomes

Although the proportion of satellite DNA contained within eukaryotic genomes varies between species and between organisms, it is generally represented in fairly high quantities from plants to higher eukaryotes. Plant satellite DNAs were discovered in 1967, slightly after those of most animals (Beridze *et al.*, 1967). The amount of satellite DNA was found to vary between closely related higher plant species, like those of the bean genus *Phaseolus* which range from 0 to 30% of the genome (Beridze, 1972). Among the invertebrates, the Nematoda class of worms has been more frequently analyzed in terms of satellite DNA. In this group, the percentage of satellite DNA varies between less than 1% for *Caenorhabditis elegans* (La Volpe *et al.*, 1988) to 80% for *Parascaris equorum* (Moritz and Roth, 1976). In two other nematode species, *Meloidogyne hapla* and *Meloidogyne incognita*, two satellite DNA families were shown together to make up 5% and 2.5% of the genome, respectively (Piotte *et al.*, 1994). In total, the amount of repeated DNA in *Meloidogyne* represents 20% of the genome (Pableo and Triantaphyllou, 1989), with other classes of satellite DNAs, plus middle repetitive elements, contributing to this number. In another example, the satellite I and II DNAs of the mealworm *Tenebrio obscurus* constitute 36% of the genome (Plohl and Ugarkovic, 1994).

Satellite DNAs of insects have also been well studied. The satellite DNA sequences of tenebrionids (beetles) represent 30 to 50% of the total genome (Davis and Wyatt, 1989).

In the darkling beetle, *Misolampus goudoti*, two different satellite DNA families were demonstrated to be present in about 70,000 and 120,000 copies per haploid genome (Pons *et al.*, 1993), and a satellite DNA from the flour beetle *Tribolium confusum* was shown to make up about 40% of the genome, representing 577,000 copies (Plohl and Ugarkovic, 1994). In the cave cricket, *Dolichopoda schiavazzi*, one species-specific satellite DNA family was shown to contribute about 30% to the entire genome (Bachmann *et al.*, 1996). Among the arthropods, the *Drosophila* genus has been extensively studied in terms of satellite DNA, although these studies began later because of underreplication of the satellite DNA during polytenization. These studies have revealed that about 25% of the *D. melanogaster* genome are made of satellite DNA (Peacock *et al.*, 1978). In *D. virilis*, satellite DNA constitutes 41% of the total DNA in diploid tissues (Gall *et al.*, 1971).

Satellite DNAs of mammals have also been much studied. The proportion of satellite DNA in the mammalian genome varies between 10 and 20% (Beridze, 1986). Mouse satellite DNA is one of the satellite DNAs studied in most detail. It was discovered by Kit (1961), during density ultracentrifugation analysis. The mouse genome was found to contain about 9% satellite DNA (Beridze, 1986). Among primates, 20 to 25% of the African Green Monkey genome is made of a major class of satellite DNA, the alphoid satellite DNA (Kurnit and Maio, 1974). Although this satellite DNA is not considered as a "classical" satellite DNA because it does not form a separate band in ultracentrifugation gradients, it is still classified as a satellite DNA because of its highly repetitive organization. Finally, it is known that 20% of the human genome is heterochromatic (Beridze, 1986). Although the heterochromatic portion of the genome is not entirely made of satellite DNAs, these sequences represent about 10% of the genome. The human satellite DNAs I to IV together constitute 6%, and the alpha satellite DNA 0.75% of the genome (Beridze, 1986). There exist other classes of satellite DNA in humans, such as satellite DNAs A, B and C that are less well characterized than the other human satellite DNAs.

1.3.2. Genetic and Epigenetic Aspects of Satellite DNA

One feature that distinguishes satellite DNA from the other DNA sequences in the eukaryotic genomes is its different base composition. Schildkraut *et al.* (1962) have derived a formula to determine the GC content of DNA molecules on the basis of their buoyant density in CsCl [ρ=1.660 + 0.098 (GC)]. Based on the densities determined by Kit (1961) for the mouse main DNA component and its satellite DNA, the proportion of GC amounts to 41.8% and 30.6%, respectively. However, calf thymus satellite DNA is comprised of 54% GC, whereas the main DNA is composed of only 39% GC base pairs (Schildkraut *et al.*, 1962). On the other hand, some classes of satellite DNAs such as the human satellite I, II, III and IV DNAs, are A+T-rich compared to main band DNA (Corneo *et al.*, 1972). The base composition pattern therefore varies according to the species and the class of satellite DNA under consideration.

At the level of chromatin structure, there is a noticeable difference between centromeric regions and other parts of the chromosomes. This difference is displayed in the condensed appearance of heterochromatic regions. The DNA molecules of chromosomes are packed by proteins known as histones, which help organize the DNA into nucleosomes. The nucleosome consists of 146 bp of DNA wound around a histone octamer composed of two each of histones H2A, H2B, H3, and H4 (McGhee and Felsenfeld, 1980). A fifth histone, H1, mediates the folding of the linker DNA between adjacent nucleosomes (Wolffe, 1992). Chromatin in transcriptionally-active genes is deficient in H1, and addition of this histone to active chromatin greatly reduces transcription (O'Neill *et al.*, 1993; Zlatanova and Van Holde, 1992).

Upon closer examination, it has been found that histone H4 acetylation is nonuniform along human metaphase chromosomes (Jeppesen *et al.*, 1992). Histones in heterochromatic regions of diverse species are underacetylated. The centric and telomeric

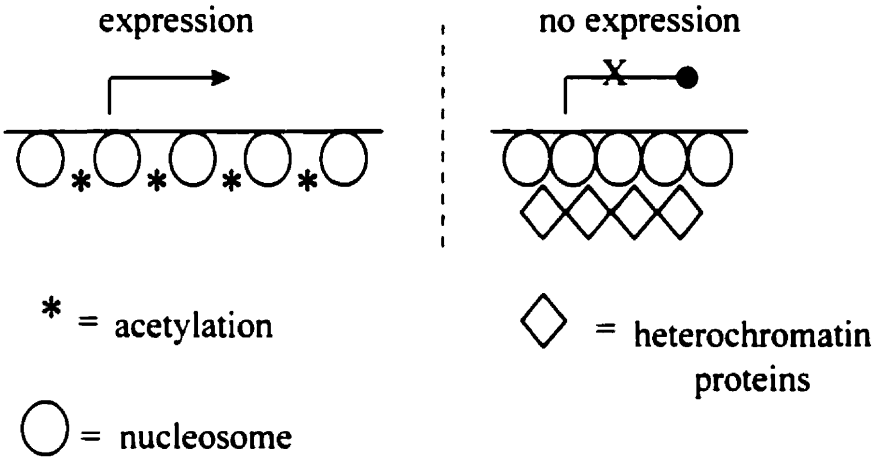
Upon closer examination, it has been found that histone H4 acetylation is nonuniform along human metaphase chromosomes (Jeppesen *et al.*, 1992). Histones in heterochromatic regions of diverse species are underacetylated. The centric and telomeric heterochromatin in human and mouse show underacetylation of the N-terminal lysine residues of histone H4 (O'Neill and Turner, 1995). The same pattern is observed for the H3 and H4 histones on the heterochromatin of the inactive X chromosome (Belyaev *et al.*, 1996; Jeppesen and Turner, 1993). Conversely, increased levels of histones H3 and H4 acetylation are found in CpG islands (Tazi and Bird, 1990), a concentration of dinucleotides (1000 to 2000 bp long) found in selected regions of the genome (Bird, 1987). These CpG islands are found at the promoters of "housekeeping genes", that encode proteins essential for cell viability and are expressed in most cells, and at promoters of some tissue-specific genes. Increased levels of histones H3 and H4 acetylation are found in regions of the genome enriched in expressed DNA sequences (Holmquist, 1987; O'Neill and Turner, 1995). It is postulated that acetylation is a landmark for genetic transcriptional activity, whereas underacetylation defines regions of the genome that are transcriptionally silent (Jeppesen and Turner, 1993). Underacetylated regions would be targetted for heterochromatin formation (Figure 1.3A). Histone underacetylation has also been observed in silent cassettes in yeast *Saccharomyces cerevisiae*, which indicates the link of heterochromatin hypoacetylation to silencing (Braunstein *et al.*, 1996). In addition, trichostatin A (TSA), a specific inhibitor of histone deacetylase interferes with the repression of marker genes in centromeric heterochromatin in fission yeast (Ekwall *et al.*, 1997).

The frequency of methylated cytosine residues (m^5C) in eukaryotic DNA has been reported to be greater in highly repetitive DNA sequences than in other fractions of the genome. High methylation of satellite DNA has been shown in a number of organisms, including humans (Miller *et al.*, 1974) and higher primates (Schnedl *et al.*, 1975). In the

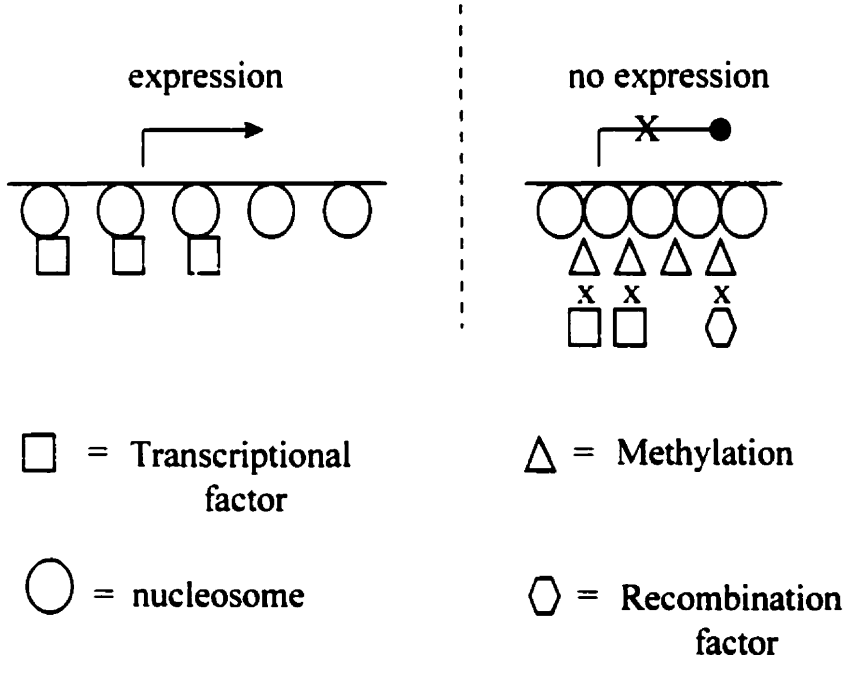
Figure 1.3. Models representing A) underacetylation (Inspired by Braunstein *et al.*, 1996) or B) hypermethylation of heterochromatic regions of eukaryotic genomes associated with genetic inactivation (second panels).

Genetically active regions Satellite DNA

A)



B)



somatic cells of the bovine genome, up to 20% of cytosines, usually of CpG dinucleotides,of satellite DNA are methylated, as opposed to less than 3% in the rest of the genome (Pech *et al.*, 1979b). The high level of cytosine methylation has been found to be associated with somatic tissues as opposed to germ cells in bovine and mice (Adams *et al.*, 1983; Feinstein *et al.*, 1985), humans (Miller *et al.*, 1974), and higher primates (Schnedl *et al.*, 1975).

The significance of higher methylation patterns in somatic cells has been proposed to reside in the prevention of recombination (Hsieh and Brutlag, 1979a). Methylation would prevent the binding of proteins involved in recombination (Figure 1.3B). It has been shown that hypomethylation of satellite DNA in ICF syndrome (Immunodeficiency, Centromere instability, Facial abnormalities) favors recombination (Miniou *et al.*, 1994). In these cases, highly demethylated satellite DNAs are preferential targets for breakpoints of rearrangements such as deletions or duplications. In addition, decondensation of heterochromatin was also observed, although hypomethylation of satellite DNA was not thought to be the sole culprit. Hypomethylated satellite DNA may therefore contribute to the formation of abnormal chromatin structure and genome instability. In the case of ICF patients, hypomethylation of satellite 2 and alpha satellite DNAs revealed a correlation between the extent of hypomethylation and the degree of malignancy (Qu *et al.*, 1999a). In addition, these same hypomethylated satellite DNAs in Wilms tumors were linked to chromosome aberrations (Qu *et al.*, 1999b), indicating that DNA hypomethylation in pericentromeric and centromeric regions is related to rearrangements. Additional studies have shown that DNase I-hypersensitive sites of chromosomes, which are considered to be genetically active, are less methylated than inactive regions (McGhee *et al.*, 1981). Methylation of satellite DNA could play a role in the regulation of gene expression by controlling or preventing the binding of certain proteins (Ehrlich and Wang, 1981) (Figure 1.3B). However, there are cases where the cytosine in satellite DNAs is not methylated, for

example in *Drosophila melanogaster*, where m^5C is not detected at all in the genome (Hsieh and Brutlag, 1979b).

1.4. Sequence and Evolution of Satellite DNA

In mammals, the different families of satellite DNAs show a certain divergence between their tandem repeats (Southern, 1970). However, one feature of satellite DNA sequences that is often observed is the internal repetition in long repeat units, usually accompanied by variations in the sequence of the units (Pech, 1979a and 1979b; Southern, 1975a). The variations are believed to be the result of mutational events. Several mammalian satellite DNAs therefore seem to be constructed from repeats that have arisen from shorter sequences. For instance, the major satellite DNA of the mouse (*Mus musculus*) consists of a predominant monomeric fragment of 234 bp (Southern, 1975a). Analysis of the DNA sequence of this 234 bp monomer shows that it must have been generated by duplication of a 120 bp unit, within which two further subunits can be identified. A more thorough analysis reveals that a 9 bp sequence would have constituted the basis for the formation of the 120 bp satellite DNA sequence. However, the reiteration of a short basic sequence within long repeating DNA units is not always observed, and constitutes the main characteristic of complex satellite DNAs. The mouse minor satellite DNA, for instance, is made of a repeating DNA unit of 130 bp that does not show any internal repetition (Pietras *et al.*, 1983). The occurrence of a sequence of mutational events over a long period of time might be responsible for masking or eliminating shorter DNA repetitions. There may also exist other events responsible for the formation of complex satellite DNAs from longer DNA repeating units.

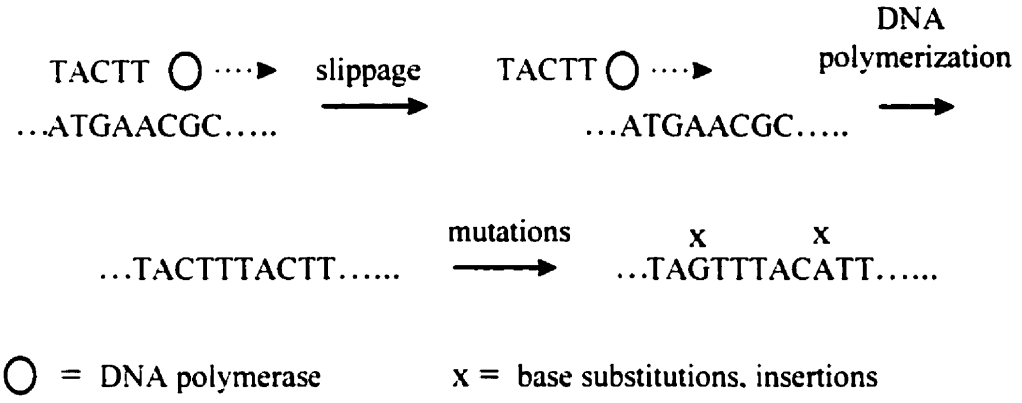
1.4.1. Saltatory Replication

The basic principle of formation of satellite DNAs seems to consist in the amplification of a repeating unit or a group of units to generate a large number of identical tandem copies. Such a process is designated as "saltatory replication" (Beridze, 1986) and is thought to be the initiating step. This event is thought to occur through two mechanisms: a strand "slippage" event during the DNA polymerization reaction, or/and a "rolling circle" event (Beridze, 1986; Charlesworth *et al.*, 1994). The slippage event may be important in the initial formation of satellite repeats, resulting in the amplification of short units (Charlesworth *et al.*, 1994) (Figure 1.4A). As repeating units are amplified by this process, mutations, base substitutions including insertions of one or a few bases, take place in such a way that smaller amplified repeats give rise to longer repeating units. The final repeating units are the results of a succession of such events.

The rolling circle mechanism is thought to occur only after a short sequence has been preliminarily amplified, since a large enough piece of DNA is needed to form a circle (Beridze, 1986). Extra-chromosomal circular sequences of an unstable human alphoid-like DNA have been observed (Kiyama *et al.*, 1987), suggesting that extra-chromosomal rolling circle replication followed by reinsertion into the genome may occur with long satellite DNA repeats (Figure 1.4B). Long satellite DNA repeats present on chromosomes could also be subject to homologous recombination between direct repeats in their sequence, followed by excision and reintegration into another repeat unit containing identical or similar DNA sequences (Figure 1.4C). This event would explain the variations observed within and between classes of satellite DNAs. Simple satellite DNAs are thought to have originated through a two-step amplification mechanism (Walker, 1971). The process would have started by the formation of a relatively long segment from a short basic sequence, followed by the appearance of a highly repeating unit by amplification of the long segment.

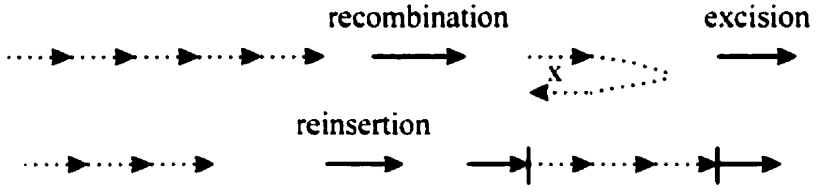
Figure 1.4. Models for the occurrence and evolution of satellite DNA sequences. A) Slippage event during the DNA polymerase. B) Rolling circle replication of a long satellite DNA sequence followed by integration into the genome. C) Homologous recombination between repeats of a satellite DNA sequence followed by excision and integration within another satellite DNA sequence. D) Unequal crossing-over between repeats on homologous chromosomes.

A) DNA polymerase slippage

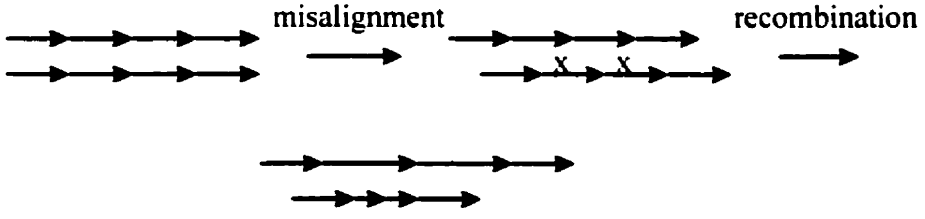


B) Rolling circle replication

C) Homologous recombination



D) Unequal crossing-over



1.4.2. Unequal Crossing-Over

In satellite DNAs containing higher-order (longer) repeating units, the process of unequal crossing-over appears to be responsible for the amplification event (Smith, 1976; Warburton *et al.*, 1993). This process would occur because the repetitive nature of satellite DNA could lead to frequent misalignments between tandem clusters of identical or nearly identical repeats on homologous chromosomes (Southern, 1975a) (Figure 1.4D). Therefore, pairing prior to recombination would involve repeating units that do not have exactly corresponding locations in their clusters. A recombination event occurring within the unevenly paired region would result in one recombinant carrying a longer cluster of repeats, and the other having a shorter cluster. The accumulation of several such deletions and tandem duplications during a specific period of time results in the formation of repetitive DNAs. In the case of complex satellite DNAs, Southern (1975a) also proposed a two-step mechanism of formation similar to that inferred for simple satellite DNAs. In this case, the first stage involves saltatory replication, while the second stage occurs through unequal crossing-over.

1.4.3. Concerted Evolution

It is thought that satellite DNAs may have evolved in eukaryotes in response to an increase in genome size (Yunis and Yasmineh, 1971) since it is not found in prokaryotes. The earliest organisms in evolution to contain satellite DNA appear to be the higher protists (algae, protozoa, fungi, slime molds). It is not known in which order the development of the nucleolus (where packaging of ribosomes takes place), centromere, heterochromatin, and satellite DNA has occurred. Nonetheless, the study of the nuclei of primitive unicellular eukaryotes (such as slime molds for instance) seems to indicate that some repetitive DNAs could have evolved prior to the emergence of the centromere and heterochromatin. Such organisms do not possess kinetochores, which represent the structure through which

replicated chromosomes bind to the mitotic spindle microtubules; they also do not contain heterochromatin, but have a nucleolus and repeated DNA sequences (Britten and Khone, 1968; Dupraw, 1970).

The amounts of repetitive DNA vary in quantity between species, but show little variation in copy number within populations (Jabs *et al.*, 1989). It has been observed that the selection for amounts of repetitive DNA varies with life history (Charlesworth *et al.*, 1994). It is tempting to believe that slowly developing species would accumulate larger quantities of repetitive DNA sequences, since a larger genome size would be required by large cells to lead to low rates of cell division. However, it can also be argued that the occurrence of excess DNA served to prevent the slow development that would result from the appearance of larger cells (Cavalier-Smith, 1980).

Not only is there a difference in the size of the repeating DNA arrays, but there is also a difference, between species, in the sequence of repeating units. Repetitive DNA sequences seem to show patterns of "concerted evolution" (Zimmer *et al.*, 1980) which is defined as "the nonindependent evolution of repetitive DNA sequences resulting in a sequence similarity of repeating units that is greater within than among species" (Dover *et al.*, 1982). This concept reflects the fact that all or most members of a family evolve together. However, conservation of satellite DNA sequences within a genus can also be observed, as in the case of four species of *Coleoptera* (Mestrovic *et al.*, 1998). Four A+T-rich satellite DNAs were found to be present in all four species. These satellite DNA sequences were highly conserved in terms of the sequence, monomer length and tandem repeat organization. The species were said to share a "library" of conserved satellite DNA sequences.

The intraspecific homogenization described by Dover *et al.* (1982) is believed to take place through mechanisms of molecular drive, achieved by inducing concerted phenotypic changes in a population. In other words, the effect of mutations on the

phenotype does not necessarily affect the organisms fitness, and thus goes unrecognized by natural selection because it is appearing in concert in a whole population. Molecular drive imparts cohesiveness to a population. It englobes those processes that induce a particular sequence to perpetuate itself and spread through the genome (Dover *et al.*, 1982). These processes include unequal crossing-over, replicative transposition followed by amplification of a particular sequence, and gene (or sequence) conversion (Elder and Turner, 1995; Roizes and Pages, 1982). Gene conversion refers to a process in which normal cross-over and repair mechanisms result in the replacement of one DNA sequence by another (Elder and Turner, 1995).

Repeated DNA sequences are thought to use a combination of unequal crossing-over, slippage replication and gene conversion to evolve, since unequal crossing-over alone eventually leads to the loss of sequences. In addition, random genetic drift and selection can also have a strong effect on the accumulation of tandem repetitive DNA sequences in the genome (Charlesworth *et al.*, 1994). Selective pressure is assumed to be very weak or nonexistent in satellite DNAs. Since they do not seem to code for any proteins, these sequences can tolerate more mutations than euchromatin. Mutations occurring in satellite DNA do not display any obvious phenotypic effects, and numerous mutations would be needed to impair the roles that have been proposed for these sequences (Yunis and Yasmineh, 1971). This property of satellite DNA renders it more apt to degenerate and consequently require replacement.

1.5. Roles Attributed to Satellite DNA

Because highly repetitive DNA does not possess any protein-coding capacity and in general does not generate any mRNA transcript, it has long been considered as "junk" DNA. This term reflects the idea that this portion of the genome does not have any real function. In this line of thought, repetitive DNA has been referred to as "selfish" DNA

(Cavalier-Smith, 1980; Orgel and Crick, 1980), which does not make any contribution to the fitness or the phenotype of the organism. In addition, its survival in the genome has been compared to the spread of a parasite in its host (Orgel and Crick, 1980), in the sense that it exploits its presence in the genome as a way to perpetuate and propagate itself without conferring any advantage or causing any harm to the cell or the organism. Because of its ability not to cause any phenotypic disturbance, satellite DNA would evade elimination and persist in the genome. Satellite DNA, which is by nature highly repetitive, inevitably falls within the junk or selfish DNA category. However, only about 3% of the human genome has protein-coding capacity (Nowalk, 1994) which would, if the junk DNA theory holds true, leave a major portion of the genome without function. Considering that only a portion of the higher eukaryotic genome has been functionally defined, it cannot be ruled out that this major part of the genome, the highly repetitive DNA, may in fact be of some importance to the organism. The host organism could have used some selfish DNA sequences for its own purpose, especially if these sequences give the organism a selective advantage by preventing it from growing too slowly (Cavalier-Smith, 1978). Middle repetitive DNA sequences, being interspersed with genes, are thought to contain the regulatory regions which reside outside the protein-coding sequences (Nowalk, 1994). The evolution of complex higher organisms must have required the development of new control sequences for different sets of genes (Wilson, 1976). Satellite DNAs, on the other hand, because of their pericentromeric location, are thought to participate in the structure of the centromeric region, a role that could be related to the function of the centromere.

1.5.1. Centromere Structure

The main idea that emerges upon consideration of the constitutive heterochromatin is that satellite DNA, which is mainly located in the regions of the centromere and telomeres, participates actively in the formation of this special type of chromatin. Satellite

DNA has been shown to bind to several centromeric proteins, as will be discussed later. The sequence of human satellite II and III DNA has been recognized with high specificity by HeLa nuclear proteins (Fouquet and DuBow, 1992a; Grady *et al.*, 1992). Two members of the high mobility group (HMG) family of nonhistone chromosomal proteins, HMG-I in mammalian cells and D1 in *Drosophila melanogaster* (Alfageme *et al.*, 1980; Lund *et al.*, 1983), were found to interact specifically with highly repetitive A+T-rich satellite DNA. These proteins were localized to heterochromatin and centromeres (Alfageme *et al.*, 1980; Disney *et al.*, 1989). Drugs binding A+T-rich DNA, such as distamycin or Hoechst 33258, displace HMG-I from DNA and disrupt the structure of centromeres (Lica *et al.*, 1986), indicating that HMG proteins localized to heterochromatin may play an important role in heterochromatin condensation. It still remains to be determined whether it is the satellite DNA sequence itself or a higher order structure that mediate the protein binding.

It has been documented in several studies that various satellite DNAs adopt a sequence-induced curved conformation (Barceló *et al.*, 1997; Martinez-Balbas *et al.*, 1990; Radic *et al.*, 1987). Long tracts of adenine residues, which have been described as inducers of curvature (Koo and Crothers, 1988), seem to be the structural elements responsible for the curvature of satellite DNAs (Barceló *et al.*, 1997). The A+T-rich sequences commonly observed in satellite DNA are very prone to bending. However, satellite DNA sequences lacking repeated A-tracts were shown to be curved in Tenebrionids (beetles) (Barceló *et al.*, 1998). Other sequence elements must therefore participate in the DNA bending. It is possible that the condensation of heterochromatin originates from a higher-order structure. Satellite DNA curvature could provide the element necessary to attain this structure. It has also been suggested that DNA curvature could account for the positioning or phasing of nucleosomes by the inability of the curved DNA to be packed into nucleosomes, or by providing a conformation prone to interaction with nuclear proteins (Drew and Travers, 1985).

Mouse cells treated with distamycin A, a drug with a binding preference for AT-rich DNA, which results in apparent straightening of the satellite DNA curvature, showed impaired chromosome condensation and elongated centromere regions (Radic *et al.*, 1987). The same effect was seen with mouse cells grown in the presence of Hoechst 33258, also an AT-specific drug, which inhibited centromeric heterochromatin condensation (Lica *et al.*, 1986; Vig and Willcourt, 1998). In these cases, satellite DNA seemed to resist compaction because of a reduced capacity to bend that would, as was suggested, affect its interaction with specific proteins. In addition, aberrations in the pattern of chromosome separation were observed (Vig and Willcourt, 1998), indicating that the structure of pericentromeric satellite DNA must play an important role in this process. These experiments support the hypothesis that satellite DNA may represent an important, or even essential, structural element involved in centromeric heterochromatin condensation. It has also been shown that some satellite DNAs, despite their high content in adenine and thymine, display conformations other than curvature (Plohl and Ugarkovic, 1994). It is therefore not excluded that other satellite DNA higher-order structures, apart from curvature, could participate in heterochromatic and centromeric chromatin condensation.

1.5.2. Anchorage to the Nuclear Matrix

Due to their base composition and resulting curved conformation, satellite DNAs could also serve as sites of attachment of DNA to the nuclear matrix. The nuclear matrix structure appears to be the structural framework responsible for higher order chromatin organization (Nickerson *et al.*, 1989). Sites of attachment to the nuclear matrix have been defined as matrix-associated or scaffold-associated regions (MAR or SAR) (Gasser *et al.*, 1989). These sites generally contain 70 to 75% (A+T). Short repetitive simple sequences containing blocks of As and Ts and displaying a curved conformation (Homberger, 1989) are the main part of nuclear matrix-associated DNA regions (Vogt, 1992). The involvement

of satellite DNAs, which are AT-rich and also adopt a folding structure, is suggestive of the development of a specific chromatin folding structure participating in the nuclear matrix attachment sites. A human nuclear matrix protein complex binding to alpha satellite DNA was identified (Enukashvily *et al.*, 1999). The association, observed *in vitro*, was seen to implicate two proteins binding directly to alpha satellite DNA, and a third component enhancing the binding of the two others.

1.5.3. Satellite DNA as a Spacer

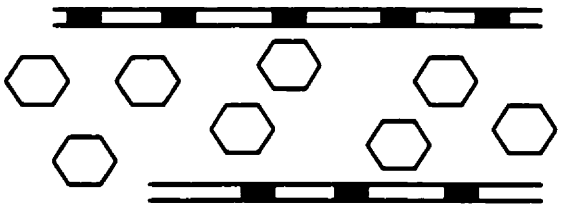
It has been suggested that satellite DNA may serve as a spacer around the nucleolar organizer where the cluster of ribosomal RNA (rRNA) genes is found (Yunis and Yasmineh, 1971). Such a spacer would protect the ribosomal RNA cistrons from mutation and cross-over to preserve their evolutionary conservation, by either spacing individual cistrons (Birnstiel *et al.*, 1968; Lima-de-Faria, 1969) or by forming a heterochromatic shield around them (Lima-de-Faria, 1969). This protective role of satellite DNA would come into play because of the reduced frequency of genetic recombination observed in the region of centromeres (Charlesworth *et al.*, 1986). In addition, sequence similarities are found between human satellite DNA I sequences and the centromere DNA element II (CDE II) of yeast centromeres that is enriched in A+T sequences (Grady *et al.*, 1992). This region is responsible for the spacing of the CDE I and CDE III domains. It has been suggested that human satellite DNA I could perform a similar spacing function in human centromeres (Vogt, 1992). Overall, satellite DNA could fulfill important structural functions such as maintaining a specific nuclear organization as well as protecting the integrity of important regions.

1.5.4. Binding to Specific Proteins

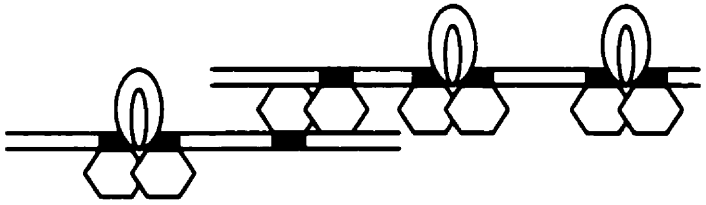
Satellite DNA seems to offer the sequences necessary for the binding and accumulation of specific proteins. Several centromere proteins (CENPs) have been identified from autoimmune sera of patients with CREST syndrome (Calcinosis, Raynaud's phenomenon. Esophageal dysmotility, Sclerodactyly Telangiectasiae) (Moroi *et al.*, 1980). Autoantibodies reacted with these proteins, which appeared to be tightly bound to DNA and located at the kinetochore. Alpha satellite DNA, which is the major DNA detected at the primary constriction of all human chromosomes (Willard, 1990), has been shown to be tightly associated with centromere proteins throughout the cell cycle (Masumoto *et al.*, 1989b). More specifically, a direct interaction has been demonstrated between CENP-B, the most abundant human centromere protein (Earnshaw *et al.*, 1987), and a 17 bp sequence of the human α satellite repeat unit (Masumoto *et al.*, 1989a). A CENP-B box was also found in the mouse minor satellite DNA (Willard, 1990).

Although the DNA binding domain of CENP-B is able to bind the CENP-B box as a monomer (Kitagawa *et al.*, 1995), the protein appears to form a stable dimer at its COOH terminus in mammalian cells. This dimer was found *in vitro* to bind two molecules of α satellite DNA: a binding which could occur intra- or inter-chromosomally (Yoda *et al.*, 1992). An internal domain of CENP-B responsible for the protein's self-association was identified (Sugimoto *et al.*, 1994). The mutual association of the CENP-B proteins at the centromere may help bring the CENP-B boxes in close proximity. The interaction at the dimerization domains of two CENP-B proteins will further stabilize this overall structure (Figure 1.5). The association may mediate the assembly or packing of the centromeric heterochromatin and help organize a higher order chromatin structure (Muro *et al.*, 1992). The possibility of CENP-B playing an important role in the assembly of higher order structures at the human centromere was suggested by Yoda *et al.* (1998). The CENP-B dimer was seen to assemble two alpha satellite DNA fragments each containing a CENP-B

Figure 1.5. Compaction and association of chromosomes through joining of satellite DNA regions mediated by binding of CENP-B proteins to CENP-B boxes. (Adapted from Sugimoto *et al.*, 1994)



Binding of CENP-B
and looping of interstitial DNA



■ = CENP-B box

⬡ = CENP-B

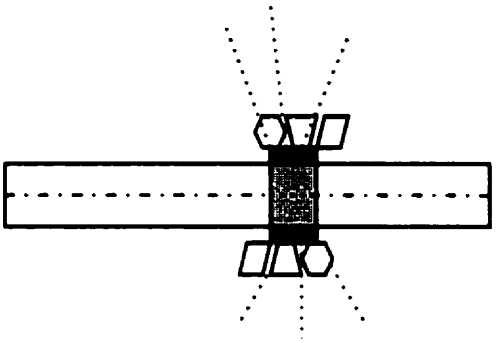
box. Intramolecular interactions of the CENP-B dimer were also observed, with the formation of DNA loop structures. CENP-B was shown to induce nucleosome positioning between pairs of CENP-B boxes implying that it may establish a particular centromeric pattern of nucleosome positioning. It has been noticed that the dimerization and DNA-binding domains of CENP-B are perfectly conserved between human and mouse (Sullivan and Glass, 1991). The central domain is also conserved perfectly between these two species (Kitagawa *et al.*, 1995). Although these observations suggest that CENP-B might perform an important function, this protein was recently found not to be essential in mice (Kapoor *et al.*, 1998). Mice null for the *cenpB* gene were produced. These mice were viable and did not present any apparent defect in growth or morphology. The centomere and kinetochore were found to be intact, and mitosis was not affected. These observations suggested that CENP-B is not required for the assembly of heterochromatin or the kinetochore, and that it may not be essential for mitosis.

1.5.5. Centromere Function

One of the important properties that centromeres are known to possess in eukaryotic cells is the responsibility of the accurate segregation of chromosomes (Miyazaki and Orr-Weaver, 1994). The centromeric protein CENP-B was found to have an acidic serine-rich region similar to the one found in the yeast CTF13 protein (Doheny *et al.*, 1993). In yeast, this protein has been found to be responsible for chromosome segregation (Goh and Kilmartin, 1993). Among the other immunologically defined centromere proteins, CENP-C and CENP-D were found to be associated with the kinetochore domain (Cooke *et al.*, 1990; Willard, 1990) (Figure 1.6A). CENP-C is the centromere protein most closely associated with the kinetochore (Willard, 1990). This protein has been shown to be essential for mitosis after anti-CENP-C antibodies inhibited mitotic progression (Bernat *et al.*, 1990; Tomkiel *et al.*, 1994). In addition, CENP-C was detected in the functional, but not the

Figure 1.6. Association of centromeric proteins with the primary constriction. A) Binding of CENP-C, CENP-D and CENP-E at the kinetochore region. B) Binding of INCENPs and CLIPs at the inner centromeric region and of a protein forming a ring around the primary constriction. (Adapted from Willard, 1990)

A)



= centromere

= Kinetochore

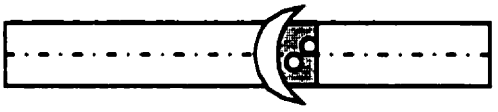
= CENP-C

= CENP-D

= CENP-E

= spindle microtubules

B)



= INCENPs; CLIPs

= ring protein

inactive, centromere of dicentric chromosomes (Earnshaw *et al.*, 1989; Page *et al.*, 1995; Sullivan and Schwartz, 1995). CENP-C may have DNA binding properties, as was inferred from analysis of its cDNA (Saitoh *et al.*, 1992), but the exact centromeric DNA sequences to which it binds are currently unknown. Simple sequence satellite DNAs could be a target for CENP-C binding, but this still remains to be shown. CENP-D has been shown to be a regulator of the onset of chromosome condensation (Bischoff *et al.*, 1990), and CENP-E has been characterized as a kinesin-like protein (Thrower *et al.*, 1995; Yen *et al.*, 1992) important for chromosome movement (Lombillo *et al.*, 1995).

In addition to the CENP antigens, other proteins have been localized at the primary constriction of chromosomes (Willard, 1990). These proteins, called INCENPs (inner centromere proteins) or CLIPs (chromatid linking proteins), are presumably involved with sister chromatid association (Figure 1.6B), but their DNA binding sites are unknown. An antigen forming a ring around the centromere was also recently identified (Holland *et al.*, 1995) with a possible function of holding the sister chromatids together, imparting physical stability to the centromere (Figure 1.6B). The identification of the DNA sequences to which these proteins bind may reveal satellite DNAs as active players in these centromeric processes.

1.5.6. Chromosome Segregation

1.5.6.1. The Use of Mammalian Artificial Chromosomes to Define Centromeric Elements

In aiming at developing vectors that can be used in important applications such as gene therapy, much effort has been focused on the construction of human or mammalian artificial chromosomes (HACs or MACs) (Harrington *et al.*, 1997; Schlessinger and Nagaraja, 1998; Willard, 1998b). In order to achieve this goal, a detailed understanding of structural and functional elements of the chromosomes of complex eukaryotes is required.

Extensive studies in yeast have established that at least three chromosomal elements are needed to produce a linear mitotically stable chromosome: a centromere, telomeres at each end, and origins of replication (Clarke and Carbon, 1980; Murray and Szostak, 1983). The molecular and genetic requirements for each of these chromosomal elements have been well defined in yeast (Bloom and Carbon, 1982; Marahrens and Stillman, 1992; Shampay *et al.*, 1984). In mammalian chromosomes, the principal missing link appears to be the centromere. This part of the chromosome is considered responsible for directing the formation of the kinetochore and mediating attachments to and movements of the chromosome along the spindle apparatus (Willard, 1998a). However, the full nature of the DNA sequences competent to provide complete centromere function in multicellular eukaryotes is still not fully understood.

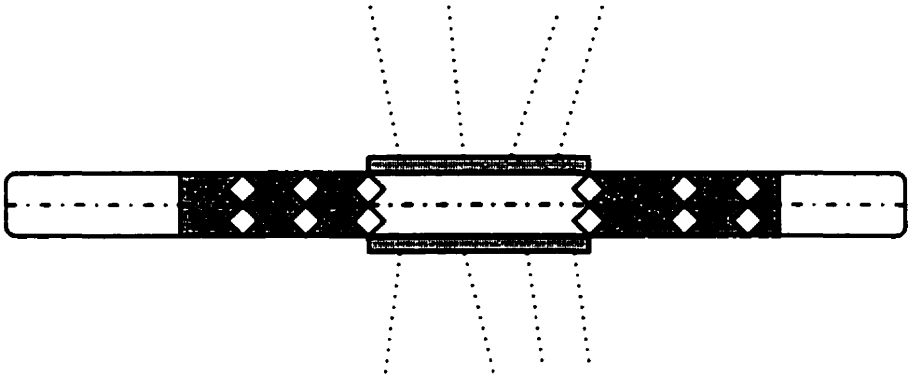
1.5.6.2. Satellite DNA and the Definition of Centromere in Multicellular Eukaryotes

As described earlier, the centromeres of many eukaryotic organisms are known to consist of megabase-sized arrays of tandemly repeated satellite DNA. Alpha satellite DNA, being the best characterized of the centromeric satellite DNAs and the only one to be found to the centromeric regions of all normal human chromosomes (Murphy and Karpen, 1998), has been used in studies with MACs. Arrays of alpha satellite DNA are found in association with centromere function, and it is known that as little as about 200 kb of alpha satellite DNA from the Y chromosome can sustain normal mitotic segregation (Heller *et al.*, 1996). Most recently, the construction and functional reintroduction into human cells of the first MAC was reported (Harrington *et al.*, 1997). In this study, microchromosomes were formed using synthetic alpha satellite DNA arrays produced through the multimerization of a single alpha satellite repeat unit. These microchromosomes were shown to be mitotically stable and to bind the centromeric proteins CENP-C and CENP-E.

Ikeno *et al.* (1998) have constructed MACs using yeast artificial chromosomes (YACs) which they combined with about 100 kb of human alpha satellite DNA in addition to telomeres. These constructs were introduced into human cells and the resulting minichromosomes were seen to behave as natural chromosomes by replicating and segregating with high fidelity. All the chromosome rearrangements examined to date retain at least 140 kb of alpha satellite DNA, which strongly suggests that this sequence plays a key role in centromere function (Heller *et al.*, 1996). Human artificial chromosomes constructed from 1 Mb yeast artificial chromosomes, containing alpha satellite DNA along with human telomeric DNA and origins of replication, were observed to bind CENP-E and found to be mitotically stable for close to 100 generations (Hennig *et al.*, 1999). Although there is evidence suggesting that alpha satellite DNA is implicated in chromosome pairing and movement, the question remains as to whether it is necessary and sufficient for centromere function.

Deletion studies using a *Drosophila* minichromosome (*Dp1187*) have shown that 200 kb or more of heterochromatic DNA, which contains AATAT satellite sequences, are required for full transmission (Murphy and Karpen, 1995). These satellite DNA sequences by themselves are not sufficient for centromere activity, but their requirement led to the suggestion that they are responsible for sister chromatid cohesion and the assembly of a fully functional kinetochore, whereas the central core mediates the attachment to spindle microtubules (Figure 1.7). The flanking heterochromatin could affect kinetochore function at a distance and play an accessory role in optimizing kinetochore formation through chromatin modification, or it could stabilize or regulate interactions with the spindle. Recently, murine pericentric satellite DNA was shown to participate in the formation of a stable murine chromosome (Telenius *et al.*, 1999). The constructed MAC was seen to support stable expression of a beta-galactosidase reporter gene over a period of months in

Figure 1.7. Participation of flanking satellite DNA sequences in the assembly and stabilization of sister chromatids and of a functional kinetochore. These flanking satellite DNA sequences may contribute additional factors helping in the organization and stability of the kinetochore. (Adapted from Murphy and Karpen, 1995)



= flanking satellite DNA

= centromeric DNA

= kinetochore

= spindle microtubules

◇ = satellite DNA binding proteins

murine, bovine and human cell lines. It therefore appears that functional MACs can be produced using pericentromeric satellite DNA, at least from mice.

It has been observed that noncentromeric DNA can acquire centromere function. Rearranged human marker chromosomes have been identified that do not contain detectable alpha satellite DNA (Blennow *et al.*, 1994; Vance *et al.*, 1997; Voullaire *et al.*, 1993). Although some of these chromosomes lack a primary constriction, others possess one. In addition, a positive reaction with CREST antiserum is observed with some of these chromosomes. Mitotic dysfunction of one marker was observed, but the regular transmission of another marker indicated the presence of a functional centromere. These observations seem to indicate that sequences that are normally euchromatic can assemble a kinetochore. Neocentromeric activity is also seen with acentric derivatives of the *Drosophila Dp1187* minichromosome (Murphy and Karpen, 1995). Although these acentrics lack elements associated with the centromere, they are transmitted well in males, indicating that the DNA sequences required for chromosome inheritance differ between sexes.

"Centromere activation" is the process describing the appearance of a centromere at a new position on a chromosome (Brown and Tyler-Smith, 1995). The mechanism of appearance of a centromere at a new position remains unknown. It has been proposed to occur through a change in the DNA sequence by transposition of a functional sequence element from a pre-existing centromere (Brown and Tyler-Smith, 1995). Other explanations involve mutational processes leading to the increase in the length of a tandemly repeated DNA sequence to form an array that functions as a centromere (Brown and Tyler-Smith, 1995). Another possibility is that centromere function is determined by an epigenetic system that activates a non-centromeric sequence. This epigenetic mechanism would involve an activation step that "imprints" the DNA, subjecting it to a self-replicating modification (Karpen and Allshire, 1997). Cytosine methylation may be such a

modification; underacetylation of histones could be another. In fact, aberrant centromeric acetylation patterns in fission yeast are correlated with abnormal centromere function involving chromosome loss (Ekwall *et al.*, 1997). The centromere activation mechanism may involve an initial activation event biased towards a particular characteristic of alpha satellite DNA. A recurring motif like the A+T-richness, repetitiveness of the DNA, or a shared higher order structure like curvature could play a role in specifying an epigenetic mechanism for centromere activity (Murphy and Karpen, 1998) (Table 1.2). DNA sequences responding to these criteria could be able to fulfill the role of a centromere. It has been shown that the presence of centromeric DNA does not necessarily support kinetochore formation and function. The major satellite DNA of the *Drosophila Dp1187* centromere is found in regions that do not fulfill centromere activity (Murphy and Karpen, 1995). Therefore, primary sequences of centromere-associated satellite DNA may not be important for centromere function.

In neocentromere formation, the disappearance of the normal centromere is alleviated by the activation of other sequences, though perhaps at a lower frequency. It should be noted that neocentromere formation is a rare event, occurring in particular situations such as chromosome rearrangements. It has been suggested that the acquisition of centromere function by acentric chromosomes might require proximity to a normal centromere (Karpen and Allshire, 1997; Williams *et al.*, 1998), involving the spreading of centromere function or proteins from the normal centromere into adjacent regions, or their action *in trans*. Although the functional elements required for providing a centromere in mammals have not been fully defined, alpha satellite DNA appears to be an important component of centromere function. Other satellite DNA sequences, flanking and interrupting alpha satellite DNA arrays, may also play a role in or contribute to centromere activity.

Table 1.2. Possible events implicated in the acquisition of centromere function by acentric chromosomes.

Genetic events	Epigenetic events
Transposition of functional sequence from another centromere (Brown and Tyler-Smith, 1995)	Increased cytosine methylation (Brown and Tyler-Smith, 1995)
Increase in length of tandemly repeated sequence (Brown and Tyler-Smith, 1995)	Histone underacetylation (Brown and Tyler-Smith, 1995)
Repetitiveness of a sequence (Murphy and Karpen, 1998)	Secondary structure (e.g. curvature) (Murphy and Karpen, 1998)
DNA composition (e.g. A+T rich) (Murphy and Karpen, 1998)	Spreading of centromeric factors from normal centromere (Karpen and Allshire, 1997)

1.5.7. Satellite DNA Transcription

Most satellite DNAs are transcriptionally inactive. However, satellite DNA transcription has been demonstrated in certain organisms such as the newt, *Notophthalmus viridescens*, (Diaz *et al.*, 1981) and amphibian, *Xenopus laevis*, (Jamrich *et al.*, 1983). In addition, the rare transcription of a 33 bp centromeric satellite DNA on lampbrush chromosomes of the newt *Triturus cristatus cristatus* (Baldwin and Macgregor, 1985) was shown. Lampbrush chromosomes represent meiotically paired chromosomes, in growing oocytes, that are covered with newly transcribed RNA packed into RNA-protein complexes (Callan, 1982). The rare transcription of satellite DNA was suggested to occur by read-through from adjacent normally transcribed sequences. This phenomenon is presumed to occur sometimes with satellite DNA sequences that are located in close proximity to transcribed regions (Diaz *et al.*, 1981; Varley *et al.*, 1980).

Studies done with the satellite 2 DNA of the newt *Triturus*, which consists of 300 to 350 bp repetitive DNA dispersed throughout the genome, have shown that this satellite DNA sequence does generate transcripts (Coats *et al.*, 1994). The satellite 2 DNA has the particularity of possessing a promoter which contains two functional domains, the proximal sequence element (PSE) and the distal sequence element (DSE), with similarities to the promoters of vertebrate small nuclear RNA (snRNA) genes. The entire promoter of the *Xenopus* U1b2 gene can be replaced by the *Triturus* satellite 2 promoter, without loss of function. Although it is not clear if this satellite 2 DNA has a cellular function, it contains a functional RNA domain that is a self-cleaving hammerhead, an extended secondary structure, which appears to be involved in the production of satellite 2 monomer RNAs in the liver and testes (Epstein and Coats, 1991). In addition, a newt ribozyme was assembled using the satellite 2 RNA (Luzi *et al.*, 1997), and it was shown to be capable of cleaving a target RNA, which suggests that it could retain the function of processing other cellular RNAs. It remains to be determined whether other satellite DNAs have the ability to carry

similar or related functions. Most recently, satellite DNA transcripts were observed in hymenopterans (Renault *et al.*, 1999). The transcripts were not polyadenylated, and one major transcript of 1.9 kb was accompanied by several smaller transcripts. The function of these transcripts, if any, remains to be determined.

1.6. Effects Observed with Repetitive DNA Sequences

Although much remains to be done in terms of defining the function and importance of satellite DNA sequences, observations have been made that provide important information on the behaviour of these sequences. As many studies have revealed, there are harmful consequences of having euchromatic genes positioned in the vicinity of heterochromatic sequences.

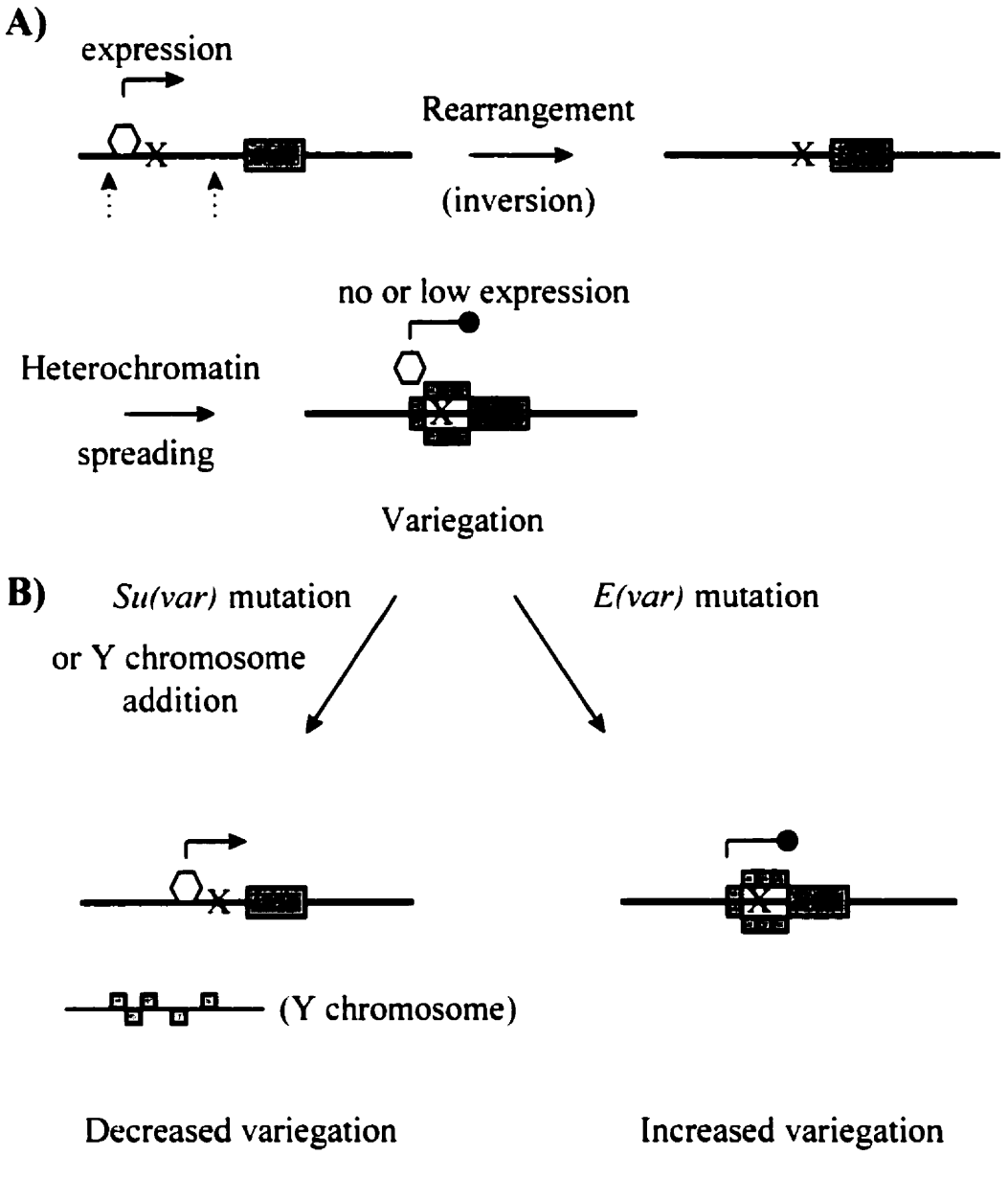
1.6.1. Position-Effect Variegation

Position-effect variegation (PEV) was first described in *Drosophila melanogaster* by Muller (1930). Muller observed cell-to-cell variations in expression of the *white*$^+$ (*w*$^+$) gene, responsible for red eye color, caused by a chromosomal rearrangement. This euchromatic gene was displaced from its normal euchromatic location by an inversion, and placed in the vicinity of heterochromatin. These rearrangements resulted in red and white patches in the adult eye. The variable inhibition was heritable, being maintained through multiple cell divisions, and provided the characteristics of silencing by epigenetic mechanisms. In another example, an inversion of the *Drosophila* X chromosome placed the *white* gene in the vicinity of heterochromatin, giving rise to a mosaic appearance of red and white eye color (Henikoff, 1990). In mammals, the first reports of PEV were describing X chromosome inactivation. Translocations have been identified in the mouse between the X chromosome and autosomes (Russell and Cacheiro, 1978; Russell and Montgomery, 1970). For example, a translocation involving the X chromosome and chromosome 8 containing

the *brown* (*br*) locus gave rise to female mice with mottled black and brown coat color, the functional *br* gene producing black coat color and a mutant or an affected gene producing brown coat color.

One of the most popular models to explain the underlying basis of PEV involves the packaging of DNA being extended, in rearranged regions, from heterochromatin into adjacent euchromatic genes (Kellum and Schedl, 1991) (Figure 1.8A). In other words, the effect would reside in the *cis*-spreading of a condensed, heterochromatic chromatin state past the rearrangement breakpoint (Elgin, 1997; Wakimoto, 1998). The altered chromatin conformation would prevent access of the transcriptional machinery to the genes and result in transcriptional repression. Variegation could be accounted for by variations among cells in the extent of linear spreading. A number of modifiers of PEV have been identified (Grigliatti, 1991). Suppressor [SU(VAR)] proteins repress euchromatic gene expression whereas enhancers [E(VAR)s] are viewed as transcriptional activators (Wakimoto, 1998). Mutations in the *E(var)7* locus have been observed to enhance variegation (Lindsley *et al.*, 1960). Deletions of the *Su(var)205* or *Su(var)3-7* loci suppress PEV of a classical variegating rearrangement involving inversion of a region of the X chromosome, $white^{mottled4}$ (Reuter *et al.*, 1987; Spofford, 1976) (Figure 1.8B), while overexpression of those genes enhances the variegation phenotype (Eissenberg *et al.*, 1992; Reuter *et al.*, 1990). Additionally, insertion of an Y chromosome, which is entirely heterochromatic in *Drosophila*, is seen to decrease PEV of euchromatic genes, presumably because of competition for heterochromatic factors (Gowen and Gay, 1933). The products of the *Su(var)205* gene, HP-1, and of *Su(var)3-7* are good candidates for structural components of heterochromatin (Elgin, 1997). HP-1 localizes predominantly to the chromocenter, which corresponds to the aggregation of centromeric regions of chromosomes, in salivary gland nuclei of *Drosophila* (James *et al.*, 1989). While having numerous binding sites throughout euchromatin, HP-1 is enriched in heterochromatin (Kellum *et al.*, 1995). SU(VAR)3-7 is

Figure 1.8. Model showing disturbance of gene expression induced by DNA rearrangements and spreading of heterochromatin (A). B) Mutations in suppressor of variegation [*Su(var)*] genes or addition of an Y chromosome can decrease the negative effect created by the rearrangements by decreasing the amount of heterochromatin proteins available; whereas mutations in enhancers of variegation [*E(var)*] increase the variegation phenotype.


A)
expression
Rearrangement
(inversion)
no or low expression
Heterochromatin
spreading
Variegation
B)
Su(var) mutation
or Y chromosome
addition
E(var) mutation
(Y chromosome)
Decreased variegation
Increased variegation
= centromere
= Transcription factor
= heterochromatin proteins

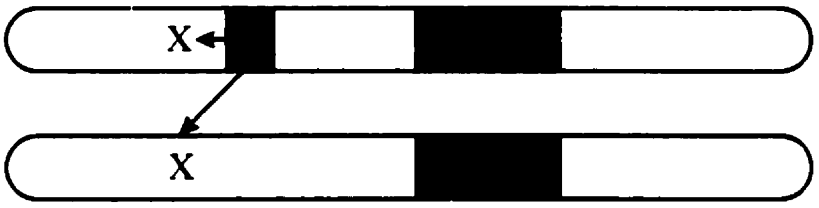
cytologically associated with centromeric heterochromatin and co-localizes there with HP-1 (Cléard *et al.*, 1997).

More observations, in favor of changes in chromatin structure in PEV, come from cytogenetic studies showing that the euchromatin adjacent to the rearrangement breakpoint undergoes visible changes in its banding pattern in *Drosophila* polytene chromosomes (Wakimoto, 1998). The variegation can also be modified by changing histone gene dosage or levels of histone acetylation. Gene inactivation, associated with PEV in mammals, is also thought to correlate with a closed chromatin configuration. This is illustrated by elimination of DNase I sensitivity in a variegated transgene located within pericentromeric heterochromatin in mice (Festenstein *et al.*, 1996).

The *cis*-spreading of heterochromatin does not explain all the features of PEV observed in *Drosophila*. Some rearrangements creating heterochromatin-euchromatin breakpoints induce the variegation of genes located several megabases away from the breakpoint (Weiler and Wakimoto, 1995). The severity of the variegated phenotype can be influenced by the position of the breakpoint, and the distance of the euchromatic genes relative to the heterochromatic block. Interchromosomal interactions (Henikoff and Dreesen, 1989) also influence the variegation. The *brown* eye pigment locus undergoes *cis*-silencing when it is juxtaposed with heterochromatin. However, silencing is also observed with the homologous copy, a case of dominant variegation (Henikoff and Comai, 1998) (Figure 1.9). This example indicates that *trans*-inactivation can occur. A rearrangement causing variegation of the *white* gene is also seen to produce *trans*-inactivation of a *white* transgene on the homologous chromosome (Martin-Morris *et al.*, 1997).

Long-distance effects could be explained by associations bringing the rearranged gene region *in trans* close to pericentromeric heterochromatin (Henikoff and Comai, 1998). This is the case for the *Drosophila brown*^{Dominant} (*bw*^{D}) heterochromatic insertion. Insertion of a heterochromatic block near the *bw* gene effectively silences this gene when

Figure 1.9. Model for *trans*-inactivation of a gene following DNA rearrangements. Inactivation of the homologous copy of a gene (X) from a rearranged chromosome presenting a block of heterochromatin adjacent to this gene. (Inspired from Henikoff and Comai, 1998)



= centromeric heterochromatin

= heterochromatic block

the homologous chromosome carries a rearrangement that brings the gene *in trans* close to pericentromeric heterochromatin (Henikoff and Comai, 1998). A nuclear compartment model has also been proposed to explain the variable and long-distance features of PEV (Wakimoto, 1998). This idea originated from the fact that *Drosophila* chromosomes at interphase have their centromeres and the bulk of heterochromatin at one end of the nucleus and the telomeres at the opposite end. This feature could explain the cell-to-cell variable phenotype observed in PEV, where the presence of the displaced gene within a particular nuclear compartment would vary. The same effect could occur with genes located megabases away from the breakpoint. Cytological studies have shown that a displaced region of heterochromatin variably associates with other regions of heterochromatin in certain cell types (Csink and Henikoff, 1996).

The effects observed in PEV are in contradiction with the behavior of genes normally located within heterochromatin (Zuckerland and Hennig, 1995). Heterochromatic genes appear to require the proximity of heterochromatin for their expression (Hilliker, 1976). Position effects on heterochromatic genes placed in an euchromatic environment have been observed. This type of effect is exemplified with the *rolled* (*rl*) gene of *Drosophila*, which is located deep in centromeric alpha heterochromatin (Eberl *et al.*, 1993). Position effects on the *rl* gene depend upon the size of the block of heterochromatin associated with the displaced gene, as well as the distance from other heterochromatic regions. The above examples illustrate that the general context in which a gene resides is very important for the proper behaviour of this gene. Changes in the normal environment of a sequence can bring about perturbations that are difficult to dissect.

1.6.2. Position Effects Observed with Isolated Genes and/or Satellite DNA Sequences

The effect of centromeric heterochromatin on euchromatic genes has also been analyzed in systems where exogenous genes are introduced into eukaryotic cells. Following introduction of foreign DNA molecules into cells, a series of events occurs. There is a period of time during which the introduced DNA sequences remain unintegrated. Ligation of DNA molecules with each other may then occur, leading to the formation of concatameric structures (Perucho *et al.*, 1980). Expression of genes present on the DNA sequences can usually occur before integration of these sequences into the genome (Milman and Herzberg, 1981). A few DNA sequences become stably integrated (Pellicer *et al.*, 1980). Butner and Lo (1986) have shown that *tk*$^-$ mouse cell line transformants, obtained by microinjection of *tk*$^-$ cells with a plasmid containing the herpes viral thymidine kinase gene (*tk*) (Lo, 1983), presented a high frequency of DNA rearrangement at the site of integration of the exogenous gene. They related this instability to the host environment flanking the integrated gene, which was shown to consist of pericentromeric satellite DNA. In light of the property of satellite DNA to be highly unstable, the effect observed was presumed to be propagated by the presence of these large blocks of repetitive DNA. Foreign DNA has not yet been demonstrated, in mammalian cells, to integrate into homologous sequences at a high frequency. Kato *et al.* (1986) conducted a study in which mice Ltk$^-$ cells were transformed with the herpes simplex virus type 1 (HSV-1) thymidine kinase gene and pBR322 DNA. The cellular DNA at the integration sites was found to consist mostly of repetitive DNA elements, indicating that foreign DNA integrates preferentially into these sequences. It was suggested that there may be cellular mechanisms protecting unique protein coding sequences by preventing integration into them.

In a similar study, a transfected HSV-1 *tk* gene was shown to display phenotypic instability when co-transfected with mouse satellite DNA (Talarico *et al.*, 1988). Co-

transfection with satellite DNA was seen to be less efficient than co-transfection with total or main band DNA. In addition, the *tk* gene and satellite DNA were physically linked in the transformants, and were preferentially located in centromeric or telomeric regions. The instability of the TK phenotype was displayed in the fact that tk^- cell lines could be readily recovered from tk^- cells, and rerevertants to the tk^- phenotype were also seen to occur at a high frequency. The phenotypic switching was assumed to be the result of the influence of the adjacent highly repetitive DNA sequences. Due to the analogy with PEV events, satellite DNA was suggested to influence the expression of the adjacent genes by promoting alterations in their chromatin structure.

Analysis of the site of integration in exogenous DNA has previously been investigated for a polyoma-derived vector containing rodent repetitive DNA sequences (Wallenburg *et al.*, 1987). It was found that integration occurred preferentially within the repetitive sequences present in the transfected DNA. The effect of repetitive DNA sequences on adjacent DNA was also studied in a heterologous system (Maiorano *et al.*, 1997). Hexamers of a telomeric AT-rich satellite DNA from the brine shrimp *Artemia franciscana* were juxtaposed to a *lacZ* reporter gene, and analysis of *lacZ* expression was performed in *Saccaromyces cerevisiae*. The functionality of the reporter gene was seen to be affected, decreasing progressively with the increase in repetitive DNA sequences present in its proximity. In this case, the orientation of the repeats also seemed to play a role.

A similar situation, where a satellite DNA sequence is positioned *in cis* with marker genes, has been studied in human cells (Fouquet and DuBow, 1992b). The satellite DNA used was a human 1.797 kb *Eco*RI satellite II DNA isolated from HeLa cells (Sol *et al.*, 1986). This satellite II DNA was estimated to represent about 2% of the genome, and to reside mainly on chromosomes 7, 12, 14, 15, 16, and 22 (Sol and DuBow, 1992). Transfections with plasmids containing the 1.797 kb satellite II DNA sequence were performed into tk^-/neo^- human cells (Fouquet and DuBow, 1992b). It was observed that the

satellite DNA, placed at a particular position in the plasmid, next to the 5' end of a *neo* gene (plasmid pCFD1), was exerting a dramatic effect on the recovery of stable *tk*$^+$/*neo*$^-$ transfectants. The transfection efficiency was consistently found to be lower or equal to one. On the other hand, the plasmid containing the satellite DNA sequence next to the 5' end of the *tk* gene did not produce such a negative effect on transfection efficiency. The satellite DNA sequence was in different orientations, based on the position of base pair 1 of the sequence, relative to the *neo* or *tk* genes, in these different plasmids. It was proposed that the satellite II DNA was either interfering with the transfected genes expression, or that it was affecting the site of plasmid integration. In addition, it was suggested that the satellite DNA was exerting its negative effect in a location and/or orientation-dependent manner.

As inferred from the above studies, it appears that satellite DNA sequences tend to exert negative effects on euchromatic genes. In certain cases, satellite DNA-directed heterochromatinization seems to be responsible for the effects observed. In other situations, the cause of the negative effects is not fully defined. Additional investigations into the mode of action of satellite DNA sequences will help in gaining a better understanding of this part of the genome.

1.7. Outline of the Thesis

In an effort to understand how the presence of satellite II DNA sequences affect transfection efficiency or/and gene expression, the present study was undertaken. As outlined above, previous observations made with the human 1.797 kb human satellite II DNA sequence indicated that its presence caused a severe negative effect on transfection efficiency (Fouquet and DuBow, 1992b). The purpose of this thesis was to further these results and to begin to understand how the satellite II DNA sequence exerted its effect. In Chapter 2, the Materials and Methods used are described. A presentation of the results is made in Chapter 3. Outlines of the transfection studies with the satellite II DNA sequence, studies designed to determine the mode of integration of the transfected plasmids, and studies into the short-term expression of the *tk* gene present in the plasmids, are presented. A discussion of the significance of the results and suggestions for future studies are presented in Chapter 4.

CHAPTER 2

MATERIALS AND METHODS

2.1. Plasmids

2.1.1. Restriction Enzyme Hydrolysis and DNA Modifications

Restriction enzyme hydrolysis of most plasmids was performed, using three units of enzyme per microgram of DNA, at 37°C in digestion buffer composed of 10 mM Tris-acetate (pH 7.5), 10 mM magnesium acetate, 50 mM potassium acetate and 100 mg/mL bovine serum albumin (BSA, Pentex Fraction V) (Sigma, Oakville, ON). All restriction enzymes were obtained from Amersham Pharmacia Biotech (Baie d'Urfé, QC), Bethesda Research Laboratories (Burlington, ON), Boehringer Mannheim (Laval, QC) or New England Biolabs (Mississauga, ON).

Hydrolyzed DNA molecules were purified prior to modification of any 5' or 3' extensions. After restriction enzyme hydrolysis, the total DNA volume was brought up to a volume of 200 µL with 1xTE [10 mM Tris-HCl (pH 7.5), 1.25 mM EDTA]. Phenol extraction of the DNA was performed using an equal volume of phenol, followed by an equal volume of chloroform ($CHCl_3$): Isoamyl alcohol (24:1, v/v). The DNA was then precipitated by the addition of 0.1 volume of 2.5 M ammonium acetate and 2.5 volumes of absolute ethanol. The precipitation was allowed to proceed for 1 hour at –70°C. The DNA was collected by centrifugation for 30 min at 4°C, and the DNA pellet was resuspended in deionized water.

To produce blunt ends for DNA ligation reactions, recessed 3' termini created by hydrolysis with several restriction enzymes were backfilled as follows: the large fragment of *Escherichia coli* DNA polymerase I (Klenow fragment) from GIBCO-BRL (Burlington, ON) was used. The reaction was performed at 25°C for 60 min in 50 mM KH_2P0_4 (pH 7.5), 3 mM $MgCl_2$, and 1 mM 2-mercaptoethanol, in the presence of 40 µM of each of the

required deoxynucleotides (dNTPs). To suppress self-ligation of cloning vectors, the 5'-phosphate groups from the hydrolyzed vector DNA were removed using 1 unit of calf intestinal alkaline phosphatase (CIAP) (GIBCO-BRL, Burlington, ON) per milligram of DNA. The reaction was performed at 37°C for 30 min in 1 M diethanolamine buffer, 10 mM 4-nitrophenyl phosphate, and 0.25 mM $MgCl_2$ (pH 9.8). The CIAP was removed using a phenol extraction procedure as described above.

2.1.2. Isolation of DNA Fragments

Separation of the DNA molecules of a given restriction endonuclease hydrolysis reaction was performed by electrophoresis through agarose (0.75%) or polyacrylamide (PAGE) (5-7%), and staining of the gels in an ethidium bromide solution. The buffer used for electrophoresis was either TAE (40 mM Tris-acetate, 20 mM glacial acetic acid, 1 mM EDTA, pH 7.5) for agarose gels or TBE (90 mM Tris-borate, 1 mM EDTA, pH 7.5) for PAGE. Gel slices containing the desired DNA fragments were purified either by the Geneclean procedure (Bio101, Mississauga, ON) for agarose gels, or by the crush-and-soak procedure for polyacrylamide gels (Sambrook *et al.*, 1989). The Geneclean process was performed according to the manufacturer's specifications. The crush-and-soak protocol was performed as follows: PAGE gel slices were incubated at 37°C overnight in crush-and-soak buffer [500 mM ammonium acetate, 10 mM magnesium acetate, 1 mM EDTA, and 0.1% (w/v) SDS]. The incubation was followed by centrifugation at 5,000 x g for 10 min at room temperature. The DNA in the supernatant fluid was precipitated in absolute ethanol by incubation at -70°C for 20 min, and centrifugation at 14,000 x g for 30 min at 4°C. The DNA pellets were resuspended in 2.5 M ammonium acetate and absolute ethanol, followed by freezing, centrifugation and resuspension of the pellets in deionized water. After a third purification, the final pellets were resuspended in 10-20 µL of 1xTE.

2.1.3. DNA Ligation

DNA ligation reactions were performed using a ratio of 3:1 (DNA fragment insert: plasmid vector), with the total amount of DNA not exceeding 1 μg in a volume of 10 μL. The reactions were allowed to proceed at 15°C for 20 hours in the presence of 2 units of T4 DNA ligase (GIBCO-BRL, Burlington, ON), in ligation buffer [50 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, 1 mM ATP, 1mM DTT, 5% (w/v) polyethylene glycol 8,000].

2.1.4. Screening of the Desired Plasmid Constructs

Ligated DNA molecules were introduced by electro-transformation (Dower *et al.*, 1988) into competent cells of *E. coli* DH1 (*supE*44, *hsdR*17, *recA*1, *end*A1, *gyrA*96, *thi*–1, *relA*1). Competent cells were prepared by inoculating 1 L of Luria-Bertani (LB) broth [1% (w/v) Bacto tryptone, 0.5% (w/v) Bacto yeast extract, 0.17 M NaCl, 3 mM NaOH] (Sambrook *et al.*, 1989) with 1/100 volume of a fresh overnight culture. The cells were grown at 37°C to an optical density (A_{600}) of 0.5 to 1, chilled for 15 to 30 min and centrifuged at 4,000 x g for 15 min at 4°C. The cell pellets were resuspended in 1 L of cold water, followed by centrifugation as above and resuspension in 0.5 L of cold water. Pellets were centrifuged again and resuspended in 20 mL of 10% (v/v) glycerol. This was followed by a final centrifugation and resuspension in 2.5 mL of 10% (v/v) glycerol. Aliquots of 80 μL were frozen and stored at –70°C.

For electro-transformation, cells were thawed at room temperature and placed on ice. A total of 2 μL of DNA (0.15-0.2 μg) was mixed with the cell suspension and placed on ice for 1 min. The mixture was transferred to a cold electroporation cuvette with a 0.2 cm electrode gap (Bio-Rad Laboratories, Mississauga, ON) and pulsed at 25 μF, 2.5 kV and 200 Ω in a Gene Pulser (Richmond, CA). The electroporation mixture was then removed and added to 1 mL of superbroth (2% Bacto tryptone, 0.5% Bacto yeast extract, 10 mM

NaCl. 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgS0_4$, 20 mM glucose). The cell suspension was incubated at 37°C for 1 hour and plated on selective medium (containing 40 μg/mL of ampicillin) to select for the presence of the *bla* (ampicillin resistance) gene.

To identify the desired plasmid constructs, bacterial transformants were streaked as patches on plates containing ampicillin. The next day, half the patch was lysed in "cracking" buffer [50 mM Tris-HCl (pH 7.5), 2 mM EDTA, 1% (w/v) SDS, 0.4 M sucrose, 0.01% (w/v) bromophenol blue] (Sambrook *et al.*, 1989) at room temperature for 15 min. Cellular debris were separated from the DNA by centrifugation at room temperature for 15 min at 14,000 x g, and the supernatant fluid was electrophoresed through a 0.75% (w/v) agarose gel.

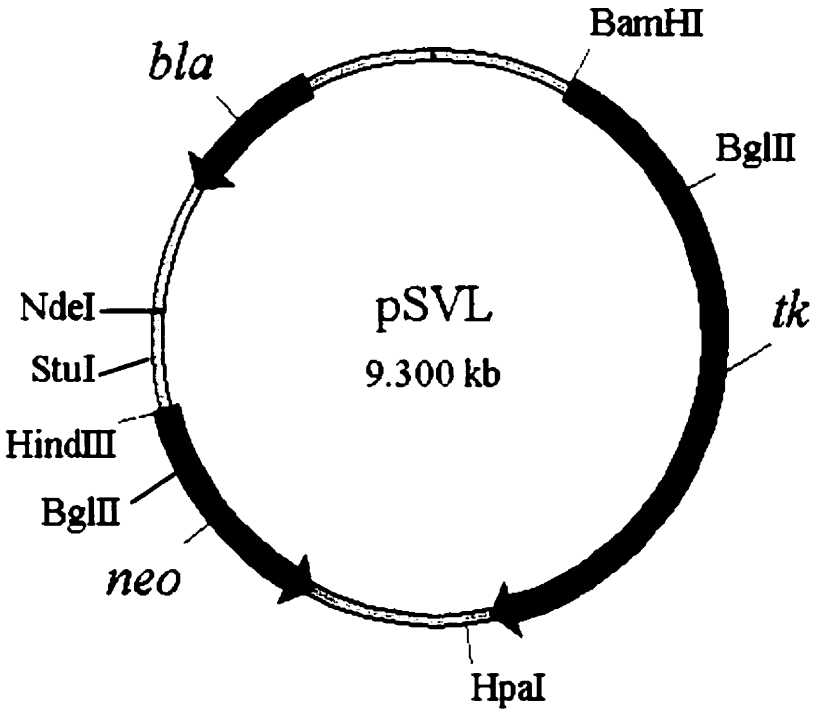
2.2. Origin of Plasmid Vectors and DNA Inserts

2.2.1. Plasmid Vectors

The vector used for construction of all plasmids, pSVL (Fouquet and DuBow, 1992b) (Figure 2.1), is a derivative of pSV2*neo*KT (Goring and DuBow, 1985) and was previously constructed by Claire Fouquet. It contains an ampicillin resistance gene (*bla*) to allow selection in bacterial cells, plus two selectable eukaryotic markers: the prokaryotic *neo* (neomycin resistance) gene, controlled by the SV40 early promoter (Southern and Berg, 1982), which confers resistance in mammalian cells to the aminoglycoside antibiotic G418, and a 3.6 kb *Bam*HI fragment containing the HSV-1 *tk* (thymidine kinase) gene which provides resistance against HAT (hypoxanthine, aminopterin, thymidine) (Littlefield, 1964). The pSVL vector can be propagated in bacterial cells because it contains a ColE1 origin of replication, but can only be propagated in eukaryotic cells that express the SV40 LTAg, such as simian COS cells (Gluzman, 1981).

Figure 2.1. Plasmid vector pSVL used for construction of the various plasmids.

The arrows indicate the direction of transcription in the *bla*, *neo*, and *tk* genes. The location of *Bam*HI, *Bgl*II, *Hin*dIII, *Hpa*I, *Nde*I, and *Stu*I restriction sites in the plasmid is indicated. (Adapted from Fouquet and DuBow, 1992b)


BamHI
bla
BglII
pSVL
9.300 kb
NdeI
tk
StuI
HindIII
BglII
neo
HpaI

2.2.2. DNA Inserts

The 1.797 kb satellite II DNA fragment used as an insert was obtained from the plasmid pKS36 (Sol *et al.*, 1986) (Figure 2.2). This satellite DNA fragment was isolated following hybridization of the *Drosophila* repetitive and mobile P-element to *Eco*RI cleaved HeLa DNA. The satellite DNA fragment contains tandem copies of the pentameric sequence 5'-TTCCA-3' including variations of this motif. A region encompassing 49 bp (nucleotides 324-372), devoid of the pentameric DNA motif, is also present. The satellite II DNA fragment was isolated from pKS36 by digestion with *Eco*RI. The recessed 3' termini of the DNA fragments were backfilled using the Klenow fragment of DNA polymerase I as indicated in section 2.1.1. The DNA fragments were then purified through agarose gels following the Geneclean procedure as previously described. Following digestion of pKS36 with *Eco*RI and purification of the 1.797 kb satellite II DNA sequence, two smaller DNA fragments of 813 bp and 983 bp were obtained from digestion of the purified satellite DNA sequence with *Cla*I. The 813 bp and 983 bp DNA fragments were isolated by electrophoresis through 5% polyacrylamide gel (PAGE), and purification by the crush-and-soak procedure (Sambrook *et al.*, 1989).

Satellite II DNA fragments of 364 bp and 620 bp were obtained from digestion of the purified 983 bp sequence with *Rsa*I. Fragments of 354 bp and 458 bp were obtained from digestion of the purified 813 bp sequence with *Hpa*II. These DNA fragments were isolated by electophoresis through 8% polyacrylamide gel, and purification by the crush-and-soak procedure.

Figure 2.2. Nucleotide sequence of the 1.797 kb *Eco*RI human satellite II DNA. The sequence is composed of tandem repetitions of the pentameric repeat unit 5'-TTCCA-3', including variations of this sequence. The boxed section indicates a region of 49 nucleotides (49-mer) devoid of the pentameric repeat unit sequence. (Adapted from Sol *et al.*, 1986) (GenBank accession number M25636)

1	AATTCCATTC	CATTTGTTTC	CATTCCATTT	TGTTCCATTT	CATTTGATTC	CATTCCATTA	GATTCCATTC	CATTTTATTC
81	TGTTCCTTTG	ATTGCATTCC	AGTTGATTTC	AATTCCATTG	GAGTCTATTC	CATTCGAGTT	CATTCCATTC	CAGTTCATTC
161	CTTTCGGGTC	CATTCCATTC	GAGTCATTGC	TTTCCATTAC	ATTCCCTTCC	CTTCCATTCG	ATTTCATTCC	ACTCCACTCC
241	ATTCCACTCC	ATTCCACAGC	ACTCCAATTC	CACTCAATTA	CCCTCCATTC	CAGCCGTTCC	ATTCCAACCA	TTCCAACATT
321	CCATAACTAA	CTCCAACCAC	TCCACTCCGC	AATCCATCTT	ACCCTTCCTT	CCTTCTACTC	CATTCCACTC	CTCTCCACTC
401	CACGCCAATC	CATTCAATTC	CATTCCACTC	CATTCCATTC	CAATCCAATC	CATTCTACTC	TTCTCCAGTT	TACTTCACTT
481	CATTCCATTC	TATTCCTTTT	GATGCCATTC	GATTCCGTTC	GATTCCAGTC	AATGCCCTTT	GATTCCATTC	CATTCGATTC
561	CATTCCATTC	TACTCCTTTC	CATTCCATTA	CATTCCATTC	GATTCCATTA	CATTGAAGTA	CATTCCATTC	CAGTCCATTC
641	CATTCCATTC	AATTTCATTC	CAGTCCATTC	CATTCCAGTG	CATTCCATTC	AAGTCCATTT	CATTCGAGTC	CATTCCATTC
721	CATTCCATTC	CATTCCGTTC	GATTTCATTC	CATTTGTGTC	CATTCCATCC	GACTCCATCC	GATTACATTC	CATTCCCTTC
801	CATTCCATTC	CATTCCATTC	CACTCCACTC	CACTCCATTC	ATTCCATTCC	ACTCCACTCC	GTTCAATTTC	TCTCAATTCA
881	ACTCCATTAC	ATTTCATTCC	ATTTCATTCC	TTTACACTGA	TTTCCATTCC	AGTCCACTCC	ACTGCACTAC	ACTCCGTTCA
961	ATTCCATTCC	TAACCATTCC	AATCGATTTC	ACTCCAATCC	ACTCCATTCC	ACTCCATTCA	ATTCCATTTC	ACAACATTCC
1041	ATTCCACAGC	ATTCCATTCC	ACTCCATTCC	ACTGCATTCC	TATTTATGGC	ACTCCATTCG	ATTCCATTCG	ATGTTCACTC
1121	GTTTCCATTC	CATTCGATTC	AATTCCTTTT	GAGTCTATTC	CTTTCCATTC	CATTTCCATT	TGATTCCATT	CCATTCAAGT
1201	CCATTCATGC	CAGTTAATTC	CATTCCAGTC	CTTTCCATTC	TACTCCATTC	CATTGGATTC	CATTCCATTA	TACTCTATTC
1281	CTTTCAATTC	CATCCATTTG	ATTCCCTTCC	ATTCTACTCA	ATTCCATTCG	ATTCCACTGC	ATTCGAGTCC	ATTCCAATTG
1361	GAGCCCTTCC	ATTCCTGTCC	ATTCCTCTCC	AGTCCATTCC	ATTCGAGTCC	ATTCCATTCC	ATTCCATTCC	ATTTGAGTCC
1441	GGTTAATTCC	ATGCCATTCC	ACTTTTGTCC	ACTCCTCTCC	ATTCCACTCC	ATTCCATTAC	ATTCCACTCC	CTTCTATTCA
1521	ATTCAACTCC	ATTCAATTCA	ACTCCATTCC	AATCCGTTCC	AATTCATTCC	ATTCCATTCC	ACTCCATTCC	ATTCCATTCC
1601	GTGGATTTCA	TTCCATTCTA	TTCTATTCCA	TTCCATTCTA	TGCAATTGCA	TTCCATTTGA	GTTCATTCCA	TTCCAGTCCA
1681	TTTATTCGAG	TCCTTTCCAT	TCCATTCCAT	TCCATTTCAT	TCCATTCCAT	TGCATTCGAT	TGTAATCCAT	TCCTTTCCAA
1761	TCCTTTCCAT	TCCATTCCAT	TGCATTCCAT	TCGAATT				

All DNA inserts were introduced into the *Stu*I site of the pSVL vector, following backfilling of the recessed 3' termini using the Klenow fragment of DNA polymerase I. The 5' phosphate groups of the plasmid vector were removed using CIAP prior to ligation to the fragments. The 1.797 kb satellite II DNA fragment was also introduced into the *Hpa*I site of pSVL, between the 3' ends of the *tk* and *neo* genes, yielding the plasmid pDH1. The pDSL1 control contained a 604 bp *Stu*I bacteriophage λ DNA fragment inserted into the *Stu*I site of pSVL. The pCFDI plasmid construct, which contains the 1.797 kb satellite DNA sequence at the *Stu*I site of pSVL next to the 5' end of the *neo* gene, was constructed previously (Fouquet and DuBow, 1992b). The pCF1.8 size control containing a 1.8 kb *Nde*I bacteriophage λ DNA fragment at the *Nde*I site of pSVL was also obtained from a previous construction done by Claire Fouquet (Fouquet and DuBow, 1992b).

2.3. Isolation of Plasmid Constructs

2.3.1. Small-Scale Rapid Isolation of Plasmid DNA

Transformed bacterial cultures containing the desired plasmids were grown overnight and subjected to centrifugation at 4,000 x g for 15 min at room temperature. The cells were lysed by resuspension in lysis buffer [50 mM glucose, 25 mM Tris-HCl (pH 8.0), 10 mM EDTA, 4 mg/mL lysozyme] and incubation for 30 min on ice. An alkaline solution (0.2 N NaOH, 1% SDS) was added, and incubation was allowed to proceed for a further 15 min at room temperature. Proteins, high molecular weight RNA and chromosomal DNA were precipitated by the addition of 7.5 mM sodium acetate (pH 4.8) for 30 min on ice. Following centrifugation at 14,000 x g for 5 min at 4°C, the plasmid DNA in the supernatant fluid was precipitated with absolute ethanol. The DNA pellet was resuspended in deionized water following centrifugation at 14,000 x g for 5 min at 4°C. After 5 min on ice, absolute ethanol was added again and the mixture centrifuged for 5 min at 4°C. The pellet was dried and resuspended in 1xTE.

2.3.2. Large-Scale Isolation of Plasmid DNA

Bacterial cells, containing the desired plasmid, were grown in 1 L of LB broth at 37°C until the culture reached an optical density (A_{550}) of 0.5 to 1. To selectively amplify the plasmids, 75 μg/mL of chloramphenicol was added, and the culture was allowed to incubate further at 37°C for 16 hours. Plasmid DNA extraction was then performed according to the alkali lysis procedure of Sambrook *et al.* (1989). The DNA was purified by equilibrium centrifugation at 39,000 rpm in a cesium chloride-ethidium bromide gradient for 40 hours at 18°C, using a Beckman Ti-75 rotor in a Beckman L8-70 ultracentrifuge (Palo Alto, CA). The resulting plasmid DNA band was removed from the gradient using a 21 gauge needle. Ethidium bromide was extracted from the DNA solution by three extractions with an equal volume of isopropanol saturated with 40xSSC (6 M NaCl, 0.6 M sodium citrate, pH 7.0). The DNA solution was dialized against 1xTE for two days at 4°C.

2.4. Cell Culture

2.4.1. The AK143B Cell Line

AK143B (ATCC CRL 8303) is a transformed human fibroblastic cell line. This cell line, which is thymidine kinase-deficient, is a *tk*$^-$ derivative of line R970-5 transformed by a murine sarcoma virus (Rhim *et al.*, 1975).

2.4.2. Media Used for Propagation of Cell Lines

Cells were propagated in Dulbecco's modified Eagle medium (DMEM) (high glucose, L-glutamine, pyridoxine hydrochloride, sodium bicarbonate) (GIBCO-BRL, Burlington, ON), supplemented with 10% (v/v) fetal bovine serum (FBS, GIBCO-BRL, Burlington, ON), 100 units/mL penicillin, 100 μg/mL streptomycin (Pen-Strep) (GIBCO-BRL, Burlington, ON) and 2.5 μg/mL amphotericin B (Fungizone) (Squibb, Montreal, QC). For cells containing the HSV-

l *tk* gene or/and the *neo* gene, HAT mixture (hypoxanthine, aminopterin, thymidine) (Sigma, Oakville, ON) was added to a final concentration of 18 μg/mL, and G418 (Geneticin) (GIBCO-BRL, Burlington, ON) was used at a final concentration of 400 μg/mL.

2.4.3. Culture Environment

Cells were grown in tissue culture flasks (260 ml; 80 cm^2) (GIBCO-BRL, Burlington, ON) or tissue culture dishes (100x20 mm or 150x20 mm) (GIBCO-BRL, Burlington, ON). They were incubated at a temperature of 37°C in a humidified 5% CO_2 atmosphere.

2.4.4. Preservation of Cell Lines

For preservation, cell lines were grown to 75% confluency, washed and detached with 0.06% trypsin (Difco Laboratories, Detroit, MI) [625 μg/mL in versene (0.2 mg/mL EDTA in phosphate buffered saline (PBS))]. Trypsin was diluted by the addition of DMEM containing FBS, Pen-Strep and Fungizone at the concentrations indicated above. The cells were collected by centrifugation at 5,000 x g for 5 min at room temperature and were resuspended in ice-cold FBS. Buffered DMSO [50% dimethyl sulfoxide (v/v), 500 mM Trizma base, 73 mM dextrose, 174.9 mM citric acid, pH 6.7] was added to a final concentration of 16.6% (v/v). The cell suspension was transferred to –70°C and stored after 1 hour in liquid nitrogen.

2.4.5. Subculturing of Cells

For propagation, cells were briefly washed with PBS (137 mM NaCl, 3 mM KCl, 8 mM Na_2HPO_4, 1 mM KH_2PO_4, pH 7.3) and detached with trypsin as described above. Cells were resuspended in DMEM, containing FBS, Pen-Strep, Fungizone and the appropriate selective drug(s) (HAT and/or G418) when required. An aliquot of the resuspended cells was diluted into fresh DMEM medium.

2.5. Electroporation of the AK143B Cell Line

2.5.1. Plasmid DNA Purification

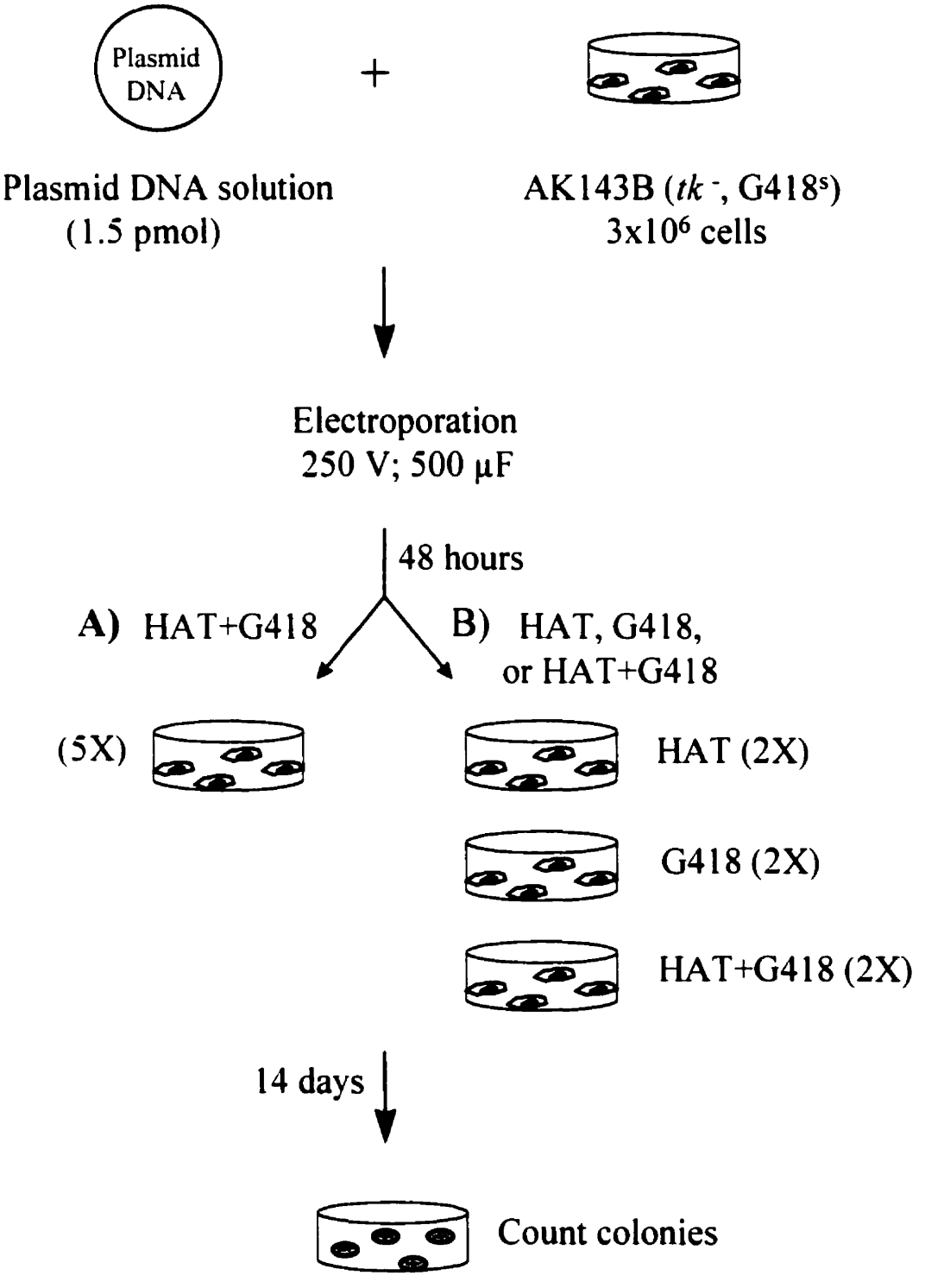
CsCl-purified plasmid DNA (0.5-1 mg) was precipitated in 0.1 volume of 2.5 M ammonium acetate and 2.5 volumes of absolute ethanol. The precipitate was allowed to form for 1 hour at -70°C and the DNA was recovered by centrifugation at 14,000 x g for 30 min at 4°C. The pelleted DNA was resuspended in 250 µL of 1xTE, and then mixed with 125 µL of 50% glycerol. The DNA/glycerol mixture was placed on a 10 mL Sephacryl® S-1000 column (Pharmacia Fine Chemicals, Uppsala, Sweden) to separate plasmid DNA from chromosomal DNA. The column was prepared by filling a 10 mL pipette with a solution of Sephacryl S-1000 beads, allowing the beads to settle and washing with 1xTE. This procedure was repeated until the level of the settled beads reached 10 mL. The flow-through was collected as 250 µL fractions. The collected DNA fractions were subjected to 0.75% agarose gel electrophoresis to identify fractions containing pure plasmid DNA. The selected fractions were used for further experiments. The plasmid DNA concentration was evaluated UV spectrophotometrically at a wavelength of 260 nm with 1 OD_{260} equalling 50 µg/mL of double stranded DNA.

2.5.2. Electroporation Procedure

Plasmids were introduced into the AK143B cell line by electroporation (Saint-Dic and DuBow, 1995) according to the following procedure: one day prior to transfection, an average of $1.0x10^6$ AK143B cells per flask (260 mL; 80 cm^2) was seeded in order to attain 50 to 75% confluency the next day. The cells were incubated overnight in DMEM supplemented with FBS, Pen-Strep and Fungizone at the concentrations specified in section 2.4.2. The next day, the cells were briefly washed with cold PBS and detached with trypsin as described in section 2.4.4. The trypsin was diluted with DMEM containing FBS, Pen-Strep and Fungizone and the cells were resuspended and collected by centrifugation for 5 min at 5 000 x g at room temperature. Pelleted cells were then washed twice with ice-cold HEPES-buffered saline (HBS) [140 mM NaCl, 5 mM KCl, 0.75 mM $Na_2HP0_4.2H_20$, 6 mM dextrose, 25 mM HEPES (ICN Biochemicals, Cleveland, OH), pH 7.05] by resuspension and centrifugation. Following a second wash, cell numbers were determined using a hemocytometer.

A total of $3x10^6$ cells were centrifuged at 5,000 x g for 5 min at room temperature, and resuspended in 0.5 mL of ice–cold HBS. The cell suspension was transferred into an ice-cold electroporation cuvette with a 0.4 cm electrode gap (Bio-Rad, Mississauga, ON). A total of 1.5 pmol of plasmid DNA was added to the cuvette and the mixture was allowed to stand for 10 min at room temperature. The DNA/cell suspension was electroporated at 250 V and 500 μF in a Bio-Rad Gene Pulser (Richmond, CA). Immediately following the electric pulse, a volume of 1.4 mL of DMEM containing FBS, Pen-Strep and Fungizone, at the concentrations specified in section 2.4.2, was added to the cuvette. Resuspended cells were seeded into non-selective medium in five 100x20 mm petri dishes, to a concentration of 10^5 cells/mL in each dish, for subsequent selection for resistance in HAT and G418. When doing resistance selection against single drugs (HAT or G418), cells were seeded into six dishes, and the different drugs, HAT and G418, HAT or G418 alone, were subsequently added separately to two dishes (Figure 2.3).

Figure 2.3. Outline of the transfection and drug selection protocol used for plasmid introduction and maintenance into AK143B human cells. A) Following electroporation, cells were separated into five petri dishes and subjected to the presence of both HAT and G418. B) To grow cells in the presence of individual drugs, they were separated into six dishes, in media containing HAT alone (two dishes), G418 alone (two dishes) or HAT and G418 (two dishes).


Plasmid DNA
+
Plasmid DNA solution
(1.5 pmol)
AK143B (tk -, G418s)
3x10^6 cells
Electroporation
250 V; 500 µF
48 hours
A) HAT+G418
B) HAT, G418,
or HAT+G418
(5X)
HAT (2X)
G418 (2X)
HAT+G418 (2X)
14 days
Count colonies

2.5.3. Drug Selection and Colony Staining

Twenty-four hours post-transfection, the non-selective medium was replaced by fresh non-selective medium. Forty-eight hours post-transfection, the non-selective medium was replaced by a medium containing the selective drugs (HAT and G418) to final concentrations specified in section 2.4.2. To select for resistance against single drugs, a medium was added containing HAT alone at a concentration of 18 μg/mL or G418 alone at a concentration of 400 μg/mL. The selective medium was replaced by fresh selective medium every three days. Fourteen days following electroporation, the plates were washed with PBS, and 10 mL of 10% buffered formalin phosphate (pH 6.9-7.1) [4% (v/v) formaldehyde, 0.4% (w/v) methanol] was added to each cell culture dish. Petri dishes were incubated for 1 to 12 hours at room temperature. The fixative was removed and replaced by 5 mL of Giemsa stain [50% (v/v) glycerol, 50% (v/v) methanol, 1% (w/v) Giemsa] per dish diluted just before use [1:50 (v/v) with deionized water]. The stain was left for 1 to 12 hours at room temperature. Petri dishes were rinsed with deionized water and the colonies were counted.

2.6. Propagation of Cell Lines from Individual Colonies

To propagate cell lines derived from individual colonies, the selected colonies were gently scraped (Goring *et al.*, 1987). Detached cells were then aspirated with a Pasteur pipette containing a few drops of the appropriate selective medium, and transferred to a 3.5x1.0 cm multi-well plate containing DMEM medium with the corresponding drug (HAT and/or G418). The plate was incubated at 37°C until cells became confluent. The cells were transferred to successively larger plates as they multiplied.

2.7. Southern Blotting

2.7.1. Isolation of Total Genomic DNA from Human Cells

Cell monolayers were detached by trypsinization as described in section 2.4.4. Cells were washed with PBS, transferred to 50 mL disposable conical tubes (Fisher Scientific, Whitby, ON) and subjected to centrifugation at 2,000 x g for 10 min at room temperature. The cell pellet was washed with PBS and resuspended in 10 mL of 1xTE. Autodigested pronase (Boehringer Mannheim, Laval, QC) (20 mg/mL in 10 mM Tris-HCl, pH 7.6) and 10% (w/v) sodium dodecyl sulfate (SDS) were added to final concentrations of 1 mg/mL and 0.5%, respectively. The mixture was incubated with gentle agitation for 5 to 8 hours at 37°C. The genomic DNA was then extracted by adding one volume of phenol and one volume of chloroform/isoamyl alcohol (24:1, v/v). The phenol extraction was repeated using one half volume of phenol, followed by two (one volume) chloroform/isoamyl alcohol (24:1, v/v) extractions. The genomic DNA suspension was dialyzed against 1xTE for 2 to 3 days.

2.7.2. Southern Blotting Technique

Genomic DNA (10 μg) was hydrolyzed as described in section 2.1.1 with ten units of the following enzymes per microgram of DNA: *Bgl*II, *Hin*dIII, *Hpa*I, or *Sac*I. The DNA used as a marker was the plasmid pSV2*neotk* hydrolyzed with *Bam*HI, *Bgl*II, and *Nde*I (Goring *et al.*, 1987). The hydrolyzed DNA was subjected to electrophoresis through a 0.75% (w/v) agarose gel. The DNA was transferred to a Hybond-N nylon membrane (Amersham Pharmacia Biotech, Baie d'Urfé, QC) following the method described by Southern (1975b). The DNA was fixed to the membrane by UV crosslinking in a Stratalinker UV Crosslinker (1800 mJoules, 30 sec) (Stratagene, Aurora, ON). The prehybridization reaction was performed using prehybridization solution [0.5 M NaH_2PO_4 (pH 7.2), 1 mM EDTA, 1% (w/v) bovine serum albumin (BSA, Pentex Fraction V) (Sigma, Oakville, ON), 7% (w/v) SDS] at 60°C for 4 hours. DNA hybridization was performed at the same temperature for 20 hours in hybridization buffer

[30% (v/v) deionized formamide, 0.2 M NaH_2PO_4 (pH 7.2), 1 mM EDTA (pH 8.0), 1% (w/v) BSA, 7% (w/v) SDS]. Approximately $1x10^9$ cpm/μg of denatured radioactively-labelled DNA (*tk*, *neo* or *bla* genes) (see section 2.9) was added to the hybridization buffer prior to incubation. The membranes were washed in 40 mM NaH_2P0_4 (pH 7.2), 1 mM EDTA (pH 8.0), 1% (w/v) SDS, four times at 60°C, and exposed to Kodak XAR-5 autoradiographic film (Rochester, NY) at –70°C for 8-10 days. Alternatively, membranes were exposed to a PhosphorImager Intensifying Screen (Sunnyvale, CA) for 24-48 hours.

2.7.3. Removal of Radioactively-Labelled DNA Probes from Southern Blots

To remove *tk* or *neo* DNA probes from the nylon membranes used in the Southern blotting technique, a boiling solution of 0.1% SDS was poured onto the membranes. The solution was left to cool to room temperature for 15-20 min. These two steps were repeated twice. The membranes were rinsed briefly in 2xSSC and, after lightly drying between two sheets of Whatman 3MM paper, they were stored in plastic bags at 4°C until further use. The membranes were re-hybridized with the *bla* (ampicillin resistance) gene following the procedure described in section 2.7.2.

2.8. Northern Blotting

2.8.1. Isolation of Total RNA

Following electroporation, cells were trypsinized as described in section 2.4.4, at 0, 24, 48 and 72 hours. Total RNA was extracted from pelleted cells using TRIzol Reagent (GIBCO-BRL, Burlington, ON) according to the manufacturer's instructions. The RNA concentration was determined by spectrophotometry at a wavelength of 260 nm with 1 A_{260} unit equalling 40 μg/mL of RNA. The RNA was stored at –70°C.

2.8.2. RNA Gel Electrophoresis

To render them RNase-free, all solutions used were treated with 0.1% (v/v) diethyl pyrocarbonate (DEPC) (Sigma, Oakville, ON) by mixing for several hours and autoclaving 20 min. Samples containing 10 μg of RNA were mixed with 20 μL of RNA loading buffer containing 0.5 μL of 10 mg/mL ethidium bromide. The RNA loading buffer was composed of 50% deionized formamide, 1xMOPS [40 mM MOPS (pH 7.0), 10 mM sodium acetate, 1 mM EDTA], 6.7% formaldehyde, 0.03% (w/v) bromophenol blue, 0.03% (w/v) xylene cyanol, and 6% (v/v) glycerol. Samples were heated at 65°C for 10 min and cooled on ice for 5 min prior to loading. Electrophoresis was performed using 1.2% denaturing agarose gels containing 1xMOPS and 2% formaldehyde, in a running buffer consisting of 1xMOPS and 3% formaldehyde. An RNA ladder (0.24 to 9.5 kb) (GIBCO-BRL, Burlington, ON) was used as a molecular size standard. Following electrophoresis, the gel was photographed under UV trans-illumination.

2.8.3. Northern Blotting Technique

The RNA was transferred onto a nitrocellulose membrane (Protran) (Xymotech, Ville Mont-Royal, QC) by laying the gel onto three sheets of Whatman 3MM paper previously soaked with 20xSSC. The gel was covered with a nitrocellulose membrane, that had been pre-soaked for 5 min in deionized water and 7 min in 20xSSC. The membrane was followed by 3 sheets of pre-soaked Whatman 3MM paper, layers of paper towel (2 inches) and a weight. After a transfer time of 15 to 24 hours, the RNA was fixed to the membrane by UV crosslinking, as previously described.

The prehybridization and hybridization procedures were performed in 50% deionized formamide, 5xSSC, 2x Denhardt's reagent [0.04% (w/v) Ficoll, 0.04% (w/v) polyvinylpyrrolidone, 0.04% (w/v) BSA], 0.1% (w/v) SDS and 100 μg/mL of denatured

salmon sperm DNA (GIBCO-BRL, Burlington, ON). The prehybridization reaction was performed by incubating the membrane at 42°C for no more than 4 hours, after which the solution was replaced and the radioactively-labelled DNA probe (*tk* gene) (section 2.7.2) was added. The hybridization reaction was performed overnight at 42°C. The membranes were washed in 2xSSC, 0.5% (w/v) SDS for 20 min at room temperature, followed by two incubations for 10 min in 1xSSC, 0.1% (w/v) SDS at room temperature, with a final wash in 0.2xSSC, 0.1% SDS for 5 min at 65°C.

2.8.4. Preparation of Poly(A)$^+$ RNA

Poly(A)$^+$ RNA was isolated from total RNA using oligo$(dT)_{25}$ Dynabeads (Dynal Inc., Lake Success, NY) according to the following procedure: aliquots containing 45 μg of total RNA were mixed with deionized water in a total volume of 500 μL. The samples were placed at 65° C for 2 min, and mixed with washed Dynabeads. Dynabeads were washed by resuspension once in 500 μL of 2x binding buffer (20 mM Tris-HCl, pH 7.6, 1 M LiCl, 2 mM EDTA). The RNA/Dynabeads mixture was allowed to stand at room temperature for 3 to 5 min. Beads were magnetically precipitated, and washed twice with 500 μL of washing buffer (10 mM Tris-HCl, pH 7.6, 0.15 M LiCl, 1mM EDTA). The poly(A)$^+$ RNA was eluted by resuspension in 10 μL of deionized water, and heating at 65° C for 2 min. The RNA was magnetically separated from the beads.

2.8.5. Removal of Radioactively-Labelled DNA Probes from Northern Blots

To remove the *tk* DNA probe from the nitrocellulose membranes, a boiling solution of 1xSSC and 0.1% SDS was poured onto the membranes. The solution was left to cool to room temperature for 15-20 min. These two steps were repeated twice. The membranes were lightly dried between two sheets of Whatman 3MM paper and stored at 4°C in plastic bags. The

membranes were re-hybridized with a metallothionein-II gene (Karin and Richards, 1982) (see section 2.9) following the procedure described in section 2.8.3.

2.9. Construction of Radioactively-Labelled DNA Probes

The *tk* gene, used as a probe in Southern and Northern blotting reactions, was isolated through restriction endonuclease hydrolysis of the pSV2*neo*KT plasmid (Goring and DuBow, 1985) with *Bam*HI. The resulting 3.6 kb *tk* fragment was separated from the rest of the DNA through a 0.75% agarose gel, and was recovered by the Geneclean procedure. The 762 bp fragment of the *neo* gene, used as a probe in Southern blotting reactions, was purified as above after restriction endonuclease hydrolysis of pSV2*neo*KT with *Pvu*II. The 2.0 kb fragment containing the *bla* gene which was used as a probe in Southern blots was isolated after restriction endonuclease hydrolysis of the pSVL vector (described in section 2.2.1) with *Eco*RI and *Nde*I. This DNA fragment was recovered as described above for the *tk* DNA fragment. The metallothionein-II fragment used as a probe in Northern blotting reactions measures approximately 350 bp. It was isolated by digestion of plasmid pAG62 (Guzzo *et al.*, 1994) with *Eco*RI and *Hind*III. The desired DNA fragment was isolated by electrophoresis through 5% polyacrylamide gel, and purification by the crush-and-soak procedure.

The probes were labelled via DNA random priming using Ready-To-Go DNA labelling beads (-dCTP) (Amersham Pharmacia Biotech, Baie d'Urfé, QC). The reaction was performed according to the manufacturer's instructions using 50 ng of DNA fragments and [α-^{32}P] dCTP (3000 Ci/mmole) (Amersham Pharmacia Biotech, Baie d'Urfé, QC) in a total volume of 50 µL. The probes were cleaned on Sephadex® G-50 columns (Amersham Pharmacia Biotech, Baie d'Urfé, QC) constructed in 1 mL syringes (Becton Dickinson & Co., Franklin Lakes, NJ). The syringes were prepared with a solution of Sephadex G-50 beads. The beads were allowed to settle by centrifugation at 5,000 x g for 2 min at room temperature. The process was repeated until the level of beads reached 0.8-1 mL. The labelled DNA fragments were then mixed with

50 μL of 1xTE and 100 μL of 50% (v/v) glycerol to bring the total volume of the sample to 200 μL. The mixture was added to the prepared columns. Purified radioactively-labelled DNA molecules, free of unincorporated dNTPs, were collected by centrifugation at 5,000 x g for 2 min at room temperature.

CHAPTER 3

RESULTS

3.1. Frequency of Transfection of Plasmid Constructs Containing Satellite II DNA Sequences

3.1.1. Plasmids Containing the 1.797 kb Human Satellite II DNA Sequence at Different Locations

Previous studies concerning the effect of the 1.797 kb satellite II DNA sequence on plasmid transfections had been performed with two plasmids containing this sequence in different locations (Fouquet and DuBow, 1992b). Plasmid pDH1 was therefore constructed with the satellite II DNA sequence in a third plasmid location. This plasmid was constructed by inserting the 1.797 kb human satellite II DNA sequence into the *Hpa*I site of pSVL, between the 3' ends of the *neo* and *tk* genes (Figure 3.1). In the original plasmid pCFD1, previously constructed by Claire Fouquet, the satellite DNA sequence had been inserted at a *Stu*I site, near the 5' end of the *neo* gene, with base pair 1 of the sequence (Figure 2.2) close to the 5' end of this gene. In pDH1, this satellite DNA sequence is inserted in the same orientation, as compared to its orientation in pCFD1, with base pair 1 close to the 3' end of the *tk* gene.

The dramatic negative effect that was observed on the recovery of stable tk^-/neo^+ transfectants from AK143B cells (Fouquet and DuBow, 1992b) with the pCFD1 plasmid was confirmed when the cells were grown in media containing HAT plus G418. The pCFD1 plasmid yielded a transfection efficiency lower than 1% (Table 3.1). The transfection efficiencies correspond to an average of the percentage of colonies obtained. A

Figure 3.1. Plasmid pDH1, containing a 1.797 kb *Eco*RI satellite II DNA sequence. The satellite DNA sequence has been inserted between the 3' ends of the *tk* and *neo* genes. The arrows indicate the direction of transcription in the *bla*, *neo*, and *tk* genes, and the direction of insertion of satellite DNA (bp 1 close to the 3' end of the *tk* gene). The location of *Nde*I and *Stu*I sites within the plasmid map is indicated.

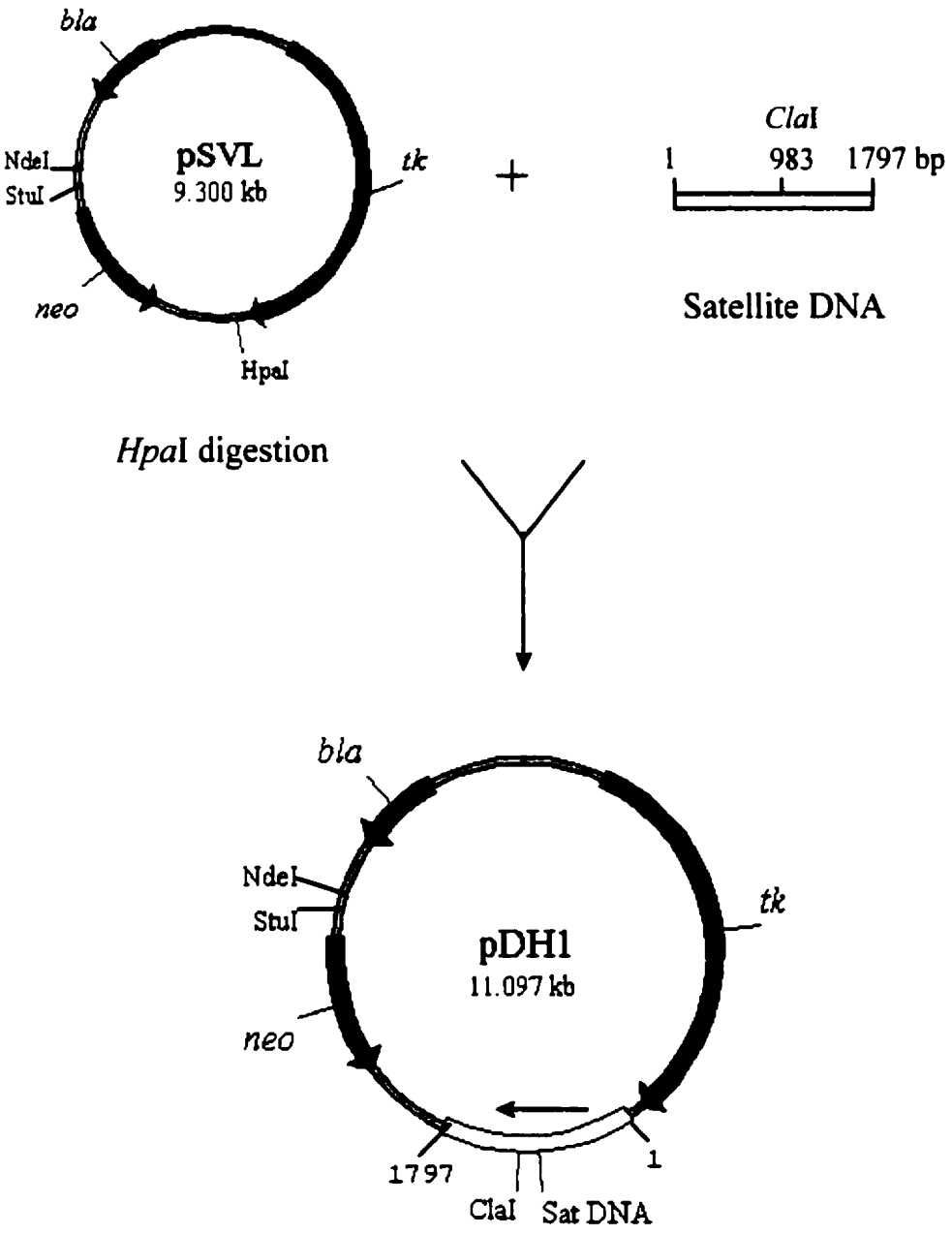

bla
NdeI
StuI
pSVL
9.300 kb
tk
neo
HpaI
ClaI
1
983
1797 bp
Satellite DNA
HpaI digestion
bla
NdeI
StuI
pDH1
11.097 kb
tk
neo
1797
1
ClaI
Sat DNA

Table 3.1. Frequency of transfection (± standard deviation) with plasmids containing the 1.797 kb *Eco*RI satellite II DNA sequence at the *Stu*I and *Hpa*I sites into media containing HAT+G418. Plasmid control pCF1.8, which contains a 1800 bp λ DNA fragment, is also represented. The orientation of the satellite DNA sequence is indicated (with respect to bp 1 at the beginning of the arrow), as is the number of experiments performed (N).

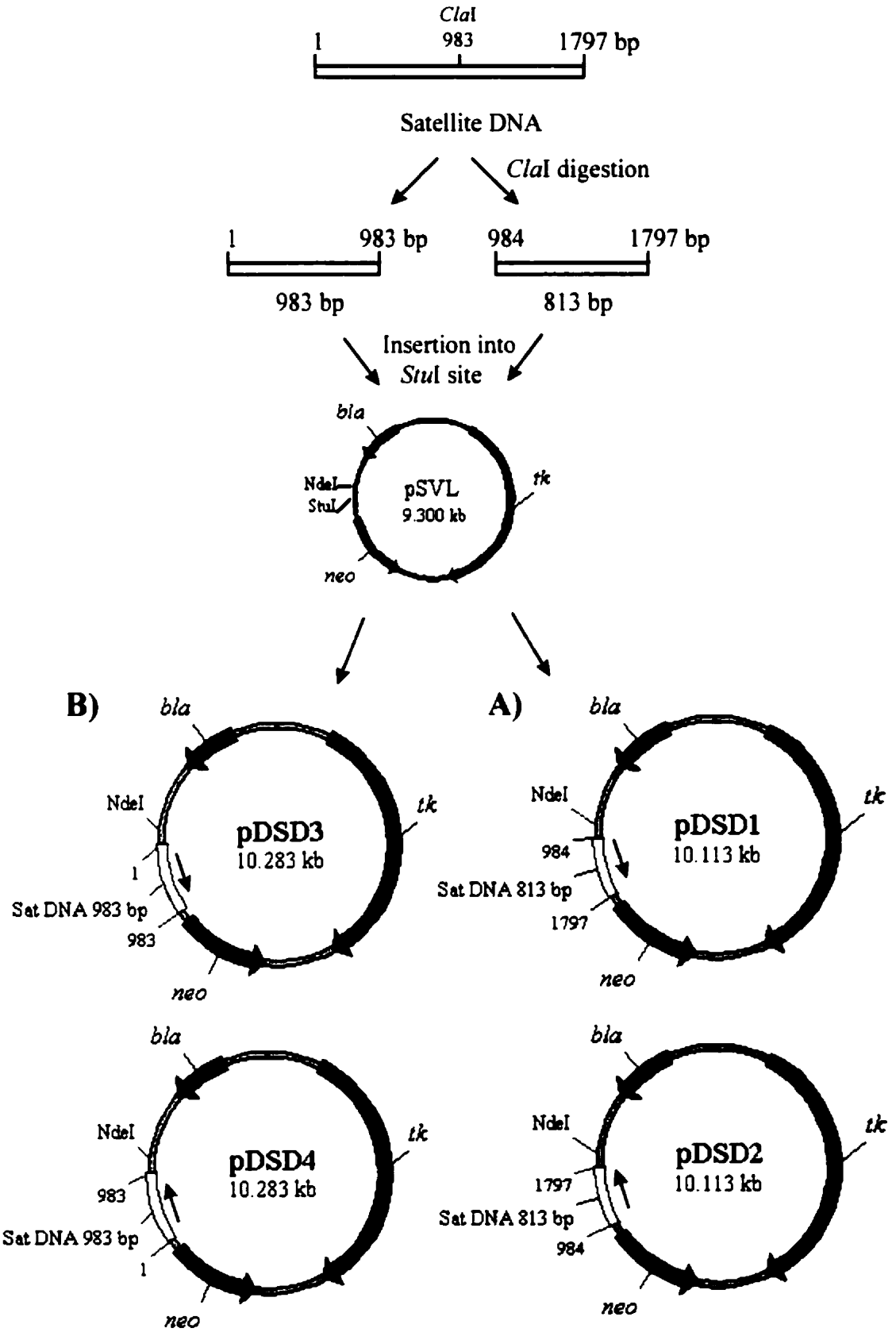
Plasmid	Satellite DNA insert location	Satellite DNA orientation	Average transfection efficiency (%)	N
pSVL	--	--	100 ± 27.8	14
pCFD1	*Stu*I	←	<1 ± 0.6	7
pDH1	*Hpa*I	←	28 ± 5.5	5
pCF1.8	--	--	66 ± 35.7	7

presentation of the different values obtained is shown in the appendix. The plasmid vector does not contain any satellite II DNA sequence. Transfection of AK143B cells with plasmid pDH1 yielded a 28% transfection efficiency, which is higher than the value observed with pCFD1 (Table 3.1). Plasmid pCF1.8, previously constructed by Claire Fouquet, which contains a 1.8 kb fragment of bacteriophage λ DNA (Fouquet and DuBow, 1992b), was used as a size control. A transfection efficiency of 66% was observed when pCF1.8 was used to transfect the AK143B cells, which corresponds to only a 1.5 fold reduction as compared to pSVL (Table 3.1).

3.1.2. Transfection Efficiencies with Plasmids Containing an 813 bp or 983 bp Fragment of the 1.797 kb Satellite II DNA Sequence

To identify the minimal size of the satellite DNA sequence which can produce a negative effect on the recovery of stable transfectants, smaller fragments of the 1.797 kb satellite II DNA sequence were inserted in pSVL at the same location (*Stu*I) for which the negative effect had been initially observed (Fouquet and DuBow, 1992b). These fragments were generated by digestion of the 1.797 kb *Eco*RI satellite II DNA sequence with *Cla*I (see Materials and Methods). Plasmids pDSD1 and pDSD2 contain, near the 5' end of the *neo* gene, an 813 bp fragment from the 1.797 kb satellite II DNA sequence in the two possible orientations (bp 984 away from or close to the 5' end of the *neo* gene, respectively) (Figure 3.2A). In addition, plasmids pDSD3 and pDSD4 contain a 983 bp fragment from the first part of the satellite DNA sequence, also in the two possible orientations (bp 1 away from or close to the 5' end of the *neo* gene, respectively) (Figure 3.2B). Transfection of pDSD2 or

Figure 3.2. Constructed plasmids containing either the 813 bp (A) or 983 bp (B) fragments of the 1.797 kb *Eco*RI satellite II DNA sequence. At the top of the figure, the position of the fragments within the 1.797 kb satellite DNA sequence is indicated. Each of the smaller satellite DNA fragments has been inserted at the *Stu*I site of pSVL, in the two possible orientations, yielding plasmids pDSD1 and pDSD2 for the 813 bp fragment (A), and pDSD3 and pDSD4 for the 983 bp fragment (B). The arrows indicate the direction of transcription of the *bla*, *neo*, and *tk* genes, as well as the direction of insertion of the satellite DNA fragments (bp 1 or bp 984 at the beginning of the arrow for the 983 bp fragment and the 813 bp fragment, respectively). The location of the *Nde*I and *Stu*I sites within pSVL is indicated.


ClaI
1
983
1797 bp
Satellite DNA
ClaI digestion
1
983 bp
984
1797 bp
983 bp
813 bp
Insertion into
StuI site
bla
NdeI
StuI
pSVL
9.300 kb
tk
neo
B)
A)
pDSD3
10.283 kb
1
Sat DNA 983 bp
983
pDSD1
10.113 kb
984
Sat DNA 813 bp
1797
pDSD4
10.283 kb
pDSD2
10.113 kb

pDSD3, in their circular conformation, and growth in media containing HAT plus G418, yielded the same results as was observed with pCFD1 (Table 3.2). Transfection efficiencies of less than 1% were obtained. However, transfection with either pDSD1 or pDSD4 showed a much higher frequency of stable *tk*$^-$/*neo*$^-$ transfectants, producing 25% and 22% transfection efficiency, respectively. Since linear plasmids have been shown to produce higher stable transfection efficiencies (Chu *et al.*, 1987; Toneguzzo *et al.*, 1986), plasmids pDSD1, pDSD2, pDSD3 and pDSD4 were linearized and used in the transfection experiments. An increase in transfection efficiency of 1.9 and 1.5 fold was seen with plasmids pDSD1 (47% transfection efficiency) and pDSD4 (32% transfection efficiency), respectively, as compared to results obtained with the circular plasmids (Table 3.2). However, only a very slight increase (from less than 1% to 2% transfection efficiency) was observed when pDSD2 was linearized, and no increase was seen with pDSD3. A similar pattern of transfection efficiency was therefore observed for circular and linear plasmids, with a very negative effect obtained using pDSD2 and pDSD3. In addition, the transfection efficiency in both cases also appears to be dependent on the orientation of the satellite DNA fragments within pSVL.

3.1.3. Transfection Efficiency with Plasmids Containing even Smaller Fragments of the 1.797 kb Satellite II DNA Sequence

Since two of the plasmids containing 805 bp and 983 bp fragments of the 1.797 kb satellite II DNA sequence were still producing a very low number of stable *tk*$^-$/*neo*$^-$ transfectants (pDSD2 and pDSD3), more plasmids were constructed with smaller DNA fragments to determine if there is a size limit to the negative effect observed with the

Table 3.2. Frequency of transfection (± standard deviation) by plasmids containing either the 813 bp or 983 bp satellite DNA fragments, into media containing HAT+G418. Circular plasmid control pCF1.8, which contains a 1800 bp λ DNA fragment, is also represented. The orientation of the satellite DNA fragments is indicated (with bp 1 for pDSD3 and pDSD4, and bp 984 for pDSD1 and pDSD2, at the beginning of the arrow). The number of experiments performed (N) is also indicated. All plasmids were linearized at the *Nde*I site.

Plasmid conformation	Plasmid	Satellite DNA size (bp)	Satellite DNA orientation	Average transfection efficiency (%)	N
Circular	pSVL	--	--	100 ± 22.6	10
	pDSD1	813	→	25 ± 9.3	5
	pDSD2	813	←	<1 ± 0.3	5
	pDSD3	983	→	<1 ± 0.3	5
	pDSD4	983	←	22 ± 16.9	5
	pCF1.8	--	--	66 ± 35.7	4
Linear	pSVL	--	--	100 ± 18	5
	pDSD1	813	→	47 ± 37.9	5
	pDSD2	813	←	2 ± 2.9	4
	pDSD3	983	→	<1 ± 0.5	5
	pDSD4	983	←	32 ± 30.9	5

satellite II DNA sequence. Four different fragments of 620, 364, 458 and 354 bp were obtained by restriction endonuclease hydrolysis of the larger 1.797 kb satellite II DNA sequence (Figure 3.3). Insertion of each of these fragments, in both possible orientations, into the *Stu*I site of pSVL, yielded a series of different plasmids (Figures 3.3 and 3.4). Plasmids pSA1 and pSA2 contain the 620 bp *Eco*RI-*Rsa*I fragment (bp 1–bp 620) of the 1.797 kb satellite II DNA sequence, inserted in both orientations (bp 1 away from or close to the 5' end of the *neo* gene, respectively) (Figure 3.4A). Plasmids pSB1 and pSB2 contain the 364 bp *Rsa*I-*Cla*I satellite II DNA fragment (bp 621-bp 983) in the two different orientations (bp 621 away from or close to the 5' end of the *neo* gene, respectively) (Figure 3.4B). Plasmids pSC1 and pSC2 contain the 458 bp *Cla*I-*Hpa*II DNA fragment (bp 984-bp 1442) in both orientations (bp 984 away from or close to the 5' end of the *neo* gene,respectively) (Figure 3.4C). Plasmids pSD1 and pSD2 contain the 354 bp *Hpa*II-*Eco*RI satellite II DNA fragment (bp 1443-bp 1797) in the two different orientations (bp 1443 away from or close to the 5' end of the *neo* gene, respectively) (Figure 3.4D).

Transfection efficiencies obtained for each of these circular constructs in AK143B cells, and growth in media containing both HAT and G418, are presented in Table 3.3. In general, the transfection efficiencies obtained with these plasmids increased more than two fold as compared to the results obtained with the previous plasmids containing larger satellite DNA inserts (Table 3.2). Transfections using plasmids pSA1, pSB1, pSB2, pSC1, and pSD1 yielded transfection efficiencies ranging from 56 to 97% (Table 3.3). Atypical values that were omitted (e.g. for pSB1 and pSC2) correspond to single values isolated from the range of frequencies obtained for a given plasmid. Some plasmids, such as pSB1 and pSD1, produced values (79% and 97%, respectively) that were very close to the control,

Figure 3.3. Size and position of restriction fragments from the 1.797 kb satellite II DNA sequence and plasmids obtained following insertion into the *Stu*I site of pSVL. Arrows indicate the direction of insertion of the fragments with respect to the 5' end of the *neo* gene in the plasmids [the end of the arrow points towards (→) or away (←) from the 5' end of this gene].

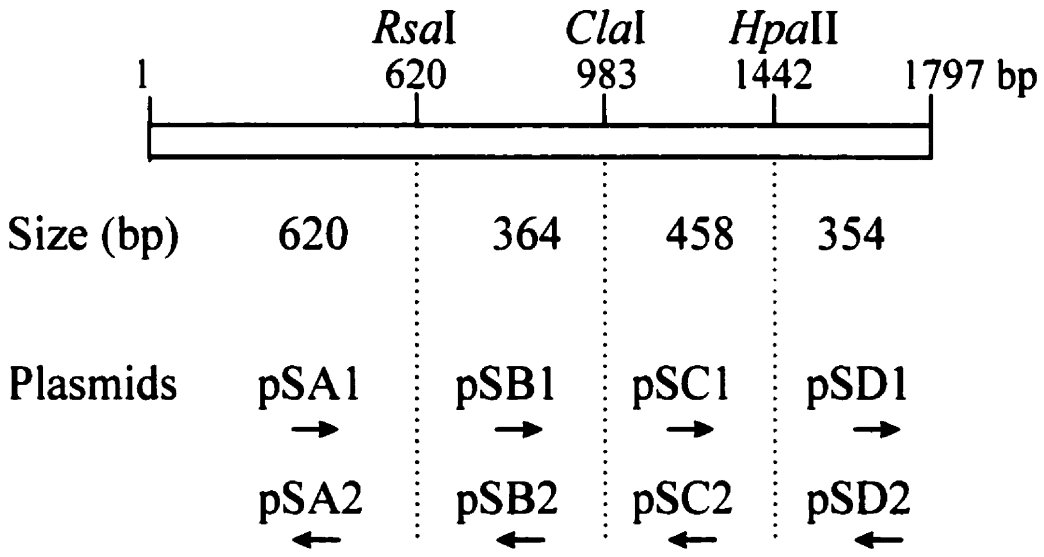

RsaI
ClaI
HpaII
1
620
983
1442
1797 bp
Size (bp)
620
364
458
354
Plasmids
pSA1
pSB1
pSC1
pSD1
pSA2
pSB2
pSC2
pSD2

Figure 3.4. Plasmid constructs containing small restriction fragments from the 1.797 kb satellite II DNA sequence inserted into the *Stu*I site of pSVL. A) Plasmids containing the 620 bp satellite II DNA fragment with bp 1 away from (pSA1) or close to (pSA2) the 5' end of the *neo* gene. B) Plasmids containing the 364 bp DNA fragment with bp 621 away from (pSB1) or close to (pSB2) the 5' end of the *neo* gene. C) Plasmids containing the 458 bp DNA fragment with bp 984 away from (pSC1) or close to (pSC2) the 5' end of the *neo* gene. D) Plasmids containing the 354 bp DNA fragment with bp 1443 away from (pSD1) or close to (pSD2) the 5' end of the *neo* gene. Arrows indicate the direction of transcription of the *bla*, *neo*, and *tk* genes, in addition to the direction of insertion of satellite DNA fragments.

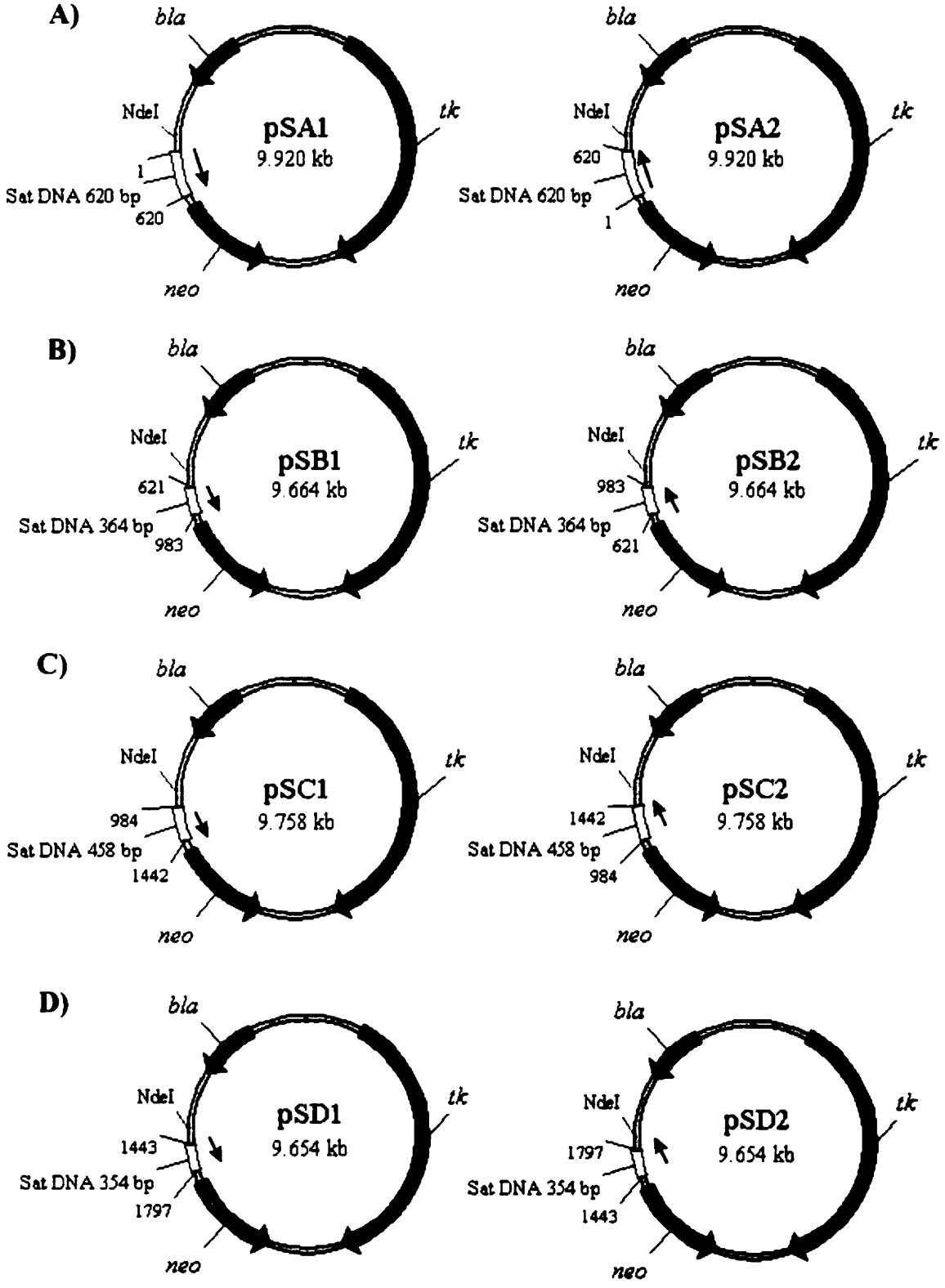

A)
bla
NdeI
1
Sat DNA 620 bp
620
pSA1
9.920 kb
tk
neo
bla
NdeI
620
Sat DNA 620 bp
1
pSA2
9.920 kb
tk
neo
B)
bla
NdeI
621
Sat DNA 364 bp
983
pSB1
9.664 kb
tk
neo
bla
NdeI
983
Sat DNA 364 bp
621
pSB2
9.664 kb
tk
neo
C)
bla
NdeI
984
Sat DNA 458 bp
1442
pSC1
9.758 kb
tk
neo
bla
NdeI
1442
Sat DNA 458 bp
984
pSC2
9.758 kb
tk
neo
D)
bla
NdeI
1443
Sat DNA 354 bp
1797
pSD1
9.654 kb
tk
neo
bla
NdeI
1797
Sat DNA 354 bp
1443
pSD2
9.654 kb
tk
neo

Table 3.3. Frequency of transfection (± standard deviation) by plasmids containing small fragments of the 1.797 kb *Eco*RI satellite II DNA, into media containing HAT+G418. Plasmid control pDSL1, containing a 604 bp λ DNA fragment, is also represented. The orientation of the satellite II DNA fragments is indicated with respect to the 5' end of the *neo* gene, as is the number of experiments performed (N).

Plasmid	Satellite DNA size (bp)	Satellite DNA orientation	Average transfection efficiency (%)	N
pSVL	--	--	100 ± 61.3	20
pSA1	620	→	56 ± 33	5
pSA2	620	←	4 ± 4.3	8
pSB1	364	→	79 ± 35.3	5*
pSB2	364	←	77 ± 29	9
pSC1	458	→	76 ± 27.1	6
pSC2	458	←	17 ± 16.3	8*
pSD1	354	→	97 ± 47.3	5
pSD2	354	←	27 ± 17.5	9
pDSL1	--	--	75 ± 26.1	9

*, One experiment was excluded as atypical (eg. 6 – 1 = 5)

pSVL. However, three plasmids, pSA2, pSC2 and pSD2, still yielded lower transfection frequencies (4%, 17%, and 27%, respectively). In these cases, the effect on transfection efficiency appeared to be satellite II DNA orientation-dependent. To confirm that this effect on transfection efficiency was not due to plasmid size, plasmid pDSL1 was constructed, that contains a 604 bp λ DNA fragment in pSVL at the *Stu*I site (Figure 3.5). A transfection efficiency of 75% was obtained with pDSL1 (Table 3.3).

3.1.4. Effect on the Reporter Genes

In order to establish which of the two reporter genes contained in the pSVL-based plasmids (*tk* or *neo*) was the most affected by the effect exerted by the satellite DNA sequences, the cellular growth following transfection was evaluated in media containing only one of the two drugs, HAT or G418. Transfections performed with pCFD1, which contains the complete 1.797 kb satellite II DNA sequence (Fouquet and DuBow, 1992b), demonstrated only a 5% transfection efficiency when G418 was used for the selection of the *neo* gene (Table 3.4). However, in the presence of only HAT, which selects for expression of the *tk* gene, a significant increase in the transfection efficiency (69%) was observed (Table 3.4). In both cases, the frequency of transfectants in media containing only one of the selection drugs is higher than that seen when both HAT and G418 are used.

As seen with pCFD1, transfections with plasmids containing either the 813 bp or the 983 bp satellite DNA fragment, yielded a lower transfection efficiency in G418 selection (2% to 43%) as compared to the results seen with selection by only HAT (76% to 113%) (Table 3.4). Plasmids pDSD2 and pDSD3, which produced a very low frequency of transfectants (<1%) in media containing both HAT and G418, also exhibited a very low transfection efficiency (2%) in the presence of G418 alone. However, media containing

Figure 3.5. Construction of plasmid pDSL1 used as size control, containing a 604 bp fragment from bacteriophage λ DNA inserted at the *Stu*I site of pSVL. Arrows indicate the direction of transcription of the *bla*, *neo*, and *tk* genes. The location of the *Nde*I and *Stu*I sites in pSVL is indicated.

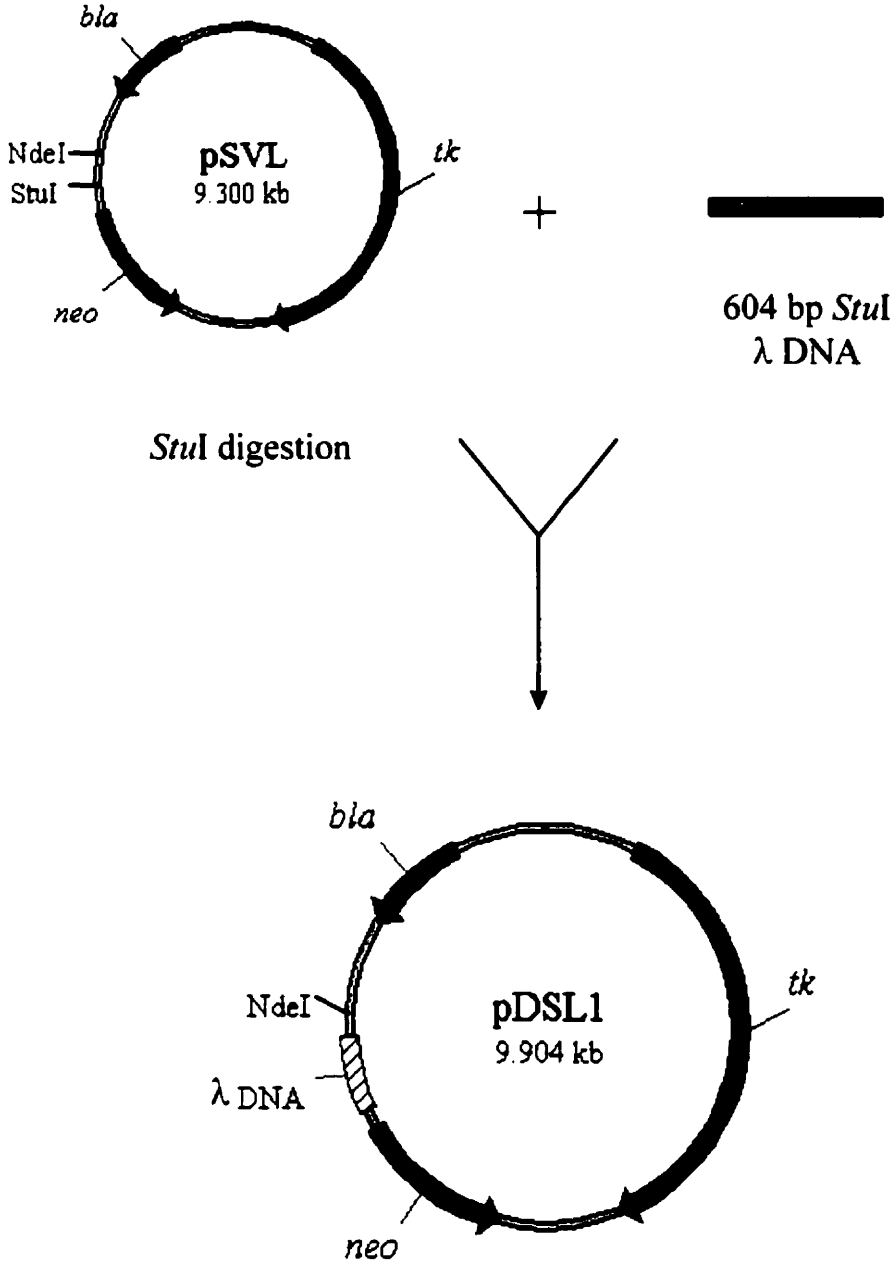

bla
NdeI
StuI
pSVL
9.300 kb
tk
neo
604 bp StuI
λ DNA
StuI digestion
bla
NdeI
λ DNA
pDSL1
9.904 kb
tk
neo

Table 3.4. Frequency of transfection (± standard deviation) by pCFD1, which contains the 1.797 kb satellite II DNA sequence, and by plasmids containing 813 bp and 983 bp satellite DNA fragments, in media containing HAT+G418, HAT or G418. Plasmid control pCF1.8, which contains a 1.8 kb λ DNA fragment is also represented. The orientation of the satellite DNA sequences is indicated [with the arrow pointing towards (→) or away (←) from the 5' end of the *neo* gene in the plasmids]. The number of experiments performed (N), in each type of media, is also indicated.

Plasmid	Satellite DNA size (bp)	Satellite DNA orientation	Average transfection efficiency (%)			$N_{HAT+G418}$, N_{HAT}, N_{G418}
			HAT+G418	HAT	G418	
pSVL	--	--	100 ± 33.8	100 ± 32.1	100 ± 43.1	5,5,5
pCFD1	1797	←	<1 ± 0.6	69 ± 23	5 ± 5.3	7,6*,7
pDSD1	813	→	33 ± 4.2	113 ± 27.8	43 ± 37.6	4,4,4
pDSD2	813	←	<1 ± 0.3	102 ± 34.9	2 ± 1.6	4,4,4
pDSD3	983	→	<1 ± 0.3	76 ± 38.3	2 ±1.8	4,4,4
pDSD4	983	←	24 ± 20.5	107 ± 17.4	21 ± 9.5	3,3,3
pCF1.8	--	--	66 ± 35.7	58 ± 14.7	101 ± 17.2	7,4,4

*, One experiment was excluded as atypical.

HAT alone did not appear to have a significant effect on the recovery of stable *tk*$^-$ transfectants, yielding transfection frequencies of 102% and 76% for pDSD2 and pDSD3, respectively. Plasmids pDSD1 and pDSD4, which were able to produce stable transfectants in media containing both HAT and G418 (33% and 24%, respectively), yielded transfection efficiencies of 43% and 21%, respectively, in the presence of G418 alone. However, no negative effect was observed with HAT alone (113% and 107% for pDSD1 and pDSD4, respectively). The plasmid control pCF1.8 shows transfection efficiencies of 58% and 101% in the presence of HAT or G418 alone, respectively.

Plasmids containing the smaller satellite II DNA inserts also demonstrated a lower frequency of *neo*^{r} transfectants in selection media containing G418 alone as opposed to HAT alone (Table 3.5). For example, plasmids pSA2 (620 bp satellite II DNA fragment) and pSC2 (458 bp satellite II DNA fragment) yielded low transfection efficiencies of 8% and 17%, respectively, in the presence of both HAT and G418. In the presence of G418 alone, the transfection efficiency again remained low at 13% and 19%, respectively. However, in the presence of HAT alone, the transfection efficiency rose to 126% and 113% for pSA2 and pSC2, respectively. A noticeable difference in transfection efficiencies with different media was also seen when pSA1 (620 bp satellite II DNA fragment) was used (Table 3.5). Although a transfection frequency of 64% was observed in media containing both HAT and G418, the value dropped to 20% in the presence of G418 alone. Conversely, the transfection efficiency in media containing HAT alone reached 162%.

With the remaining plasmids, pSB1 and pSB2 (364 bp satellite II DNA fragment), pSC1 (458 bp satellite DNA fragment), and pSD1 and pSD2 (354 bp satellite II DNA fragments), transfection efficiencies ranging from 50% to 90% were obtained in the

Table 3.5. Frequency of transfection by plasmids containing various small fragments of the 1.797 kb *Eco*RI satellite II DNA in media containing HAT+G418, HAT or G418. Plasmid control pDSL1, containing a 604 bp λ DNA fragment, is also represented. The orientation of the satellite DNA sequences is indicated [with the arrow pointing towards (→) or away (←) from the 5' end of the *neo* gene in the plasmids]. The number of experiments performed (N), in each type of media, is also indicated.

Plasmid	Satellite DNA size (bp)	Satellite DNA orientation	Average transfection efficiency (%)			$N_{HAT-G418}$, N_{HAT}, N_{G418}
			HAT+G418	HAT	G418	
pSVL	--	--	100 ± 44.2	100 ± 38	100 ± 39	10,10,10
pSA1	620	→	64 ± 32.6	162 ± 34.6	20 ± 6.7	5,5,5
pSA2	620	←	8 ± 3.4	126 ± 29.8	13 ± 5.7	4,4,4
pSB1	364	→	78 ± 13	114 ± 24.5	60 ± 22.5	5*, 6,6
pSB2	364	←	84 ± 21	115 ± 28.9	78 ± 3.8	5, 4*,4*
pSC1	458	→	105 ± 46.5	96 ± 16.4	72 ± 22.5	5, 4*,5
pSC2	458	←	17 ± 4.9	113 ± 32.1	19 ± 9.3	4,4,4
pSD1	354	→	136 ± 29.7	115 ± 39.4	90 ± 16.7	4*, 4*,5
pSD2	354	←	30 ± 23.9	96 ± 33.4	50 ± 14.5	4,4,4
pDSL1	--	--	75 ± 26.1	113 ± 31.3	97 ± 24.9	9,9,9

*, One experiment was excluded as atypical.

presence of G418. However, with these same plasmids, higher values (from 96% to 115%) were obtained in the presence of HAT alone (Table 3.5). With the plasmid control pDSL1, the difference in the frequency of transfectants, in the presence of HAT or G418 alone, is not as pronounced (Table 3.5). Values of 113% and 97% are produced with media containing only HAT or G418, respectively.

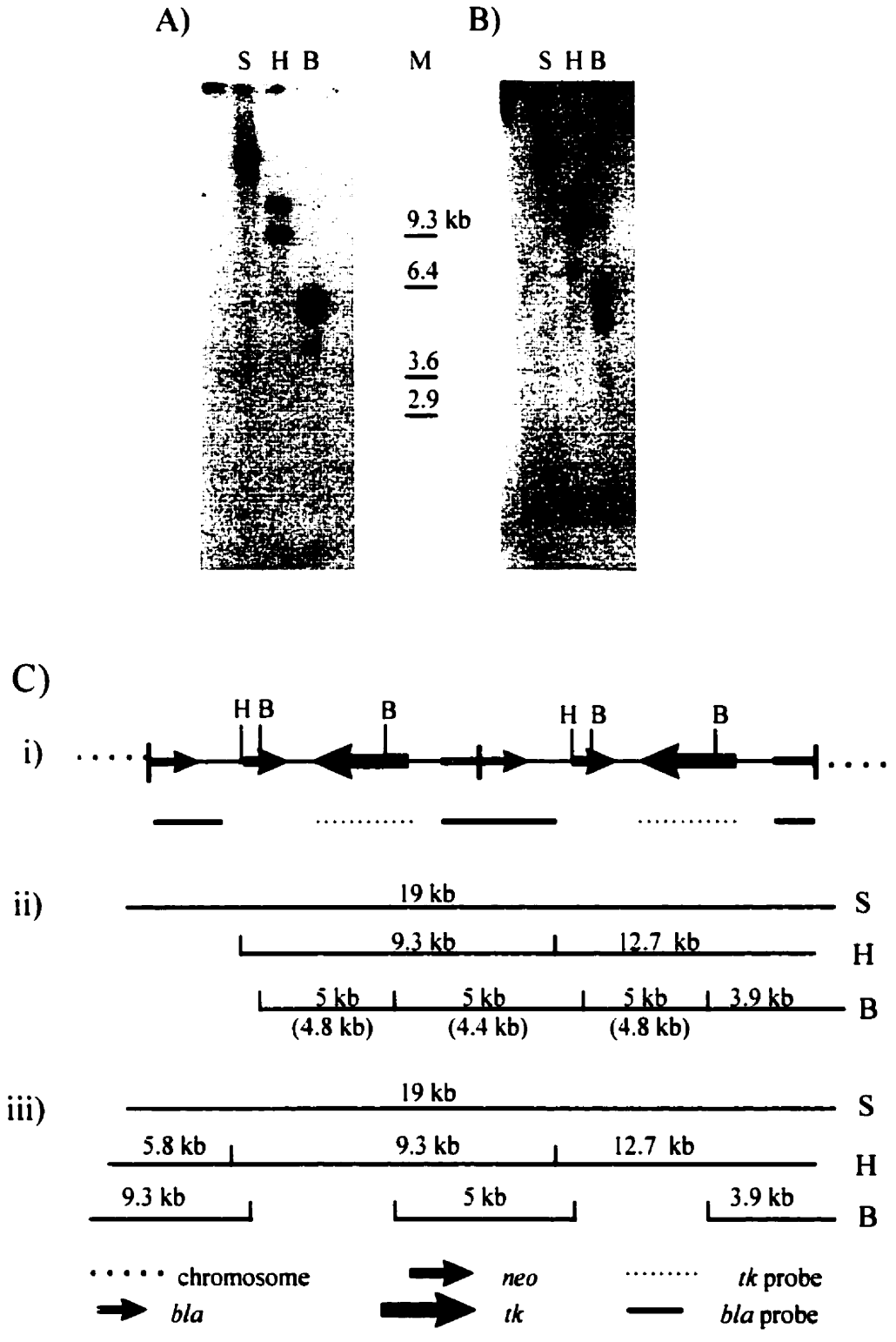
3.2. Analysis of Plasmid Insertion Pattern into Chromosomal DNA

In order to determine if the effects on stable transfection efficiency caused by the satellite DNA were due to a particular site(s) of plasmid insertion, Southern blot analyses of selected stable transfectants were performed. Total cellular DNA from each of the stable transfectant cell lines was digested with the restriction endonucleases *Bgl*II, *Hin*dIII, or *Hpa*I present on the plasmid DNA. The *Hin*dIII and *Hpa*I sites are unique within the plasmid DNA, whereas there are two *Bgl*II sites present each into the *neo* and *tk* genes. In addition, the total cellular DNA was also digested with *Sac*I, for which no restriction endonuclease site was present within the plasmid DNA. Cells transfected with either pSVL or pCFD1, which contains the 1.797 kb satellite II DNA sequence, were subjected to Southern blot analysis.

3.2.1. Southern Blotting with Cells Transfected with pSVL

Cell line SVLG-1 cell line, consisting of AK143B cells stably transfected with pSVL, was grown in a medium containing G418. Total cellular DNA was extracted (see Materials and Methods), transferred to a nylon membrane and probed with either a *tk* probe (Figure 3.6A) or a *bla* probe (Figure 3.6B). The DNA used as a marker (M) in both panels

Figure 3.6. Southern blot analysis of total cellular DNA from the SVLG-1 cell line. (A) Hybridization of the Southern blot with the *tk* probe. (B) Hybridization of the Southern blot with the *bla* probe. (C) Putative integration pattern of the pSVL plasmid within the SVLG-1 cellular DNA: (i) tandem insertion of pSVL; (ii) corresponding restriction endonuclease pattern of total cellular DNA from SVLG-1 hybridized with the *tk* probe; (iii) corresponding restriction endonuclease pattern of total cellular DNA from SVLG-1 hybridized with the *bla* probe. The sizes (in kb) indicate the bands obtained in the blots. M, marker DNA: pSV2*neotk* plasmid digested with *Bam*HI, *Bgl*II, and *Nde*I and hybridized with the *tk* probe; B, *Bgl*II; H, *Hind*III; S, *Sac*I.


A)
S H B
B)
M
S H B
9.3 kb
6.4
3.6
2.9
C)
i)
H B
B
H B
B
ii)
19 kb
S
9.3 kb
12.7 kb
H
5 kb
(4.8 kb)
5 kb
(4.4 kb)
5 kb
(4.8 kb)
3.9 kb
B
iii)
19 kb
S
5.8 kb
9.3 kb
12.7 kb
H
9.3 kb
5 kb
3.9 kb
B
chromosome
neo
tk probe
bla
tk
bla probe

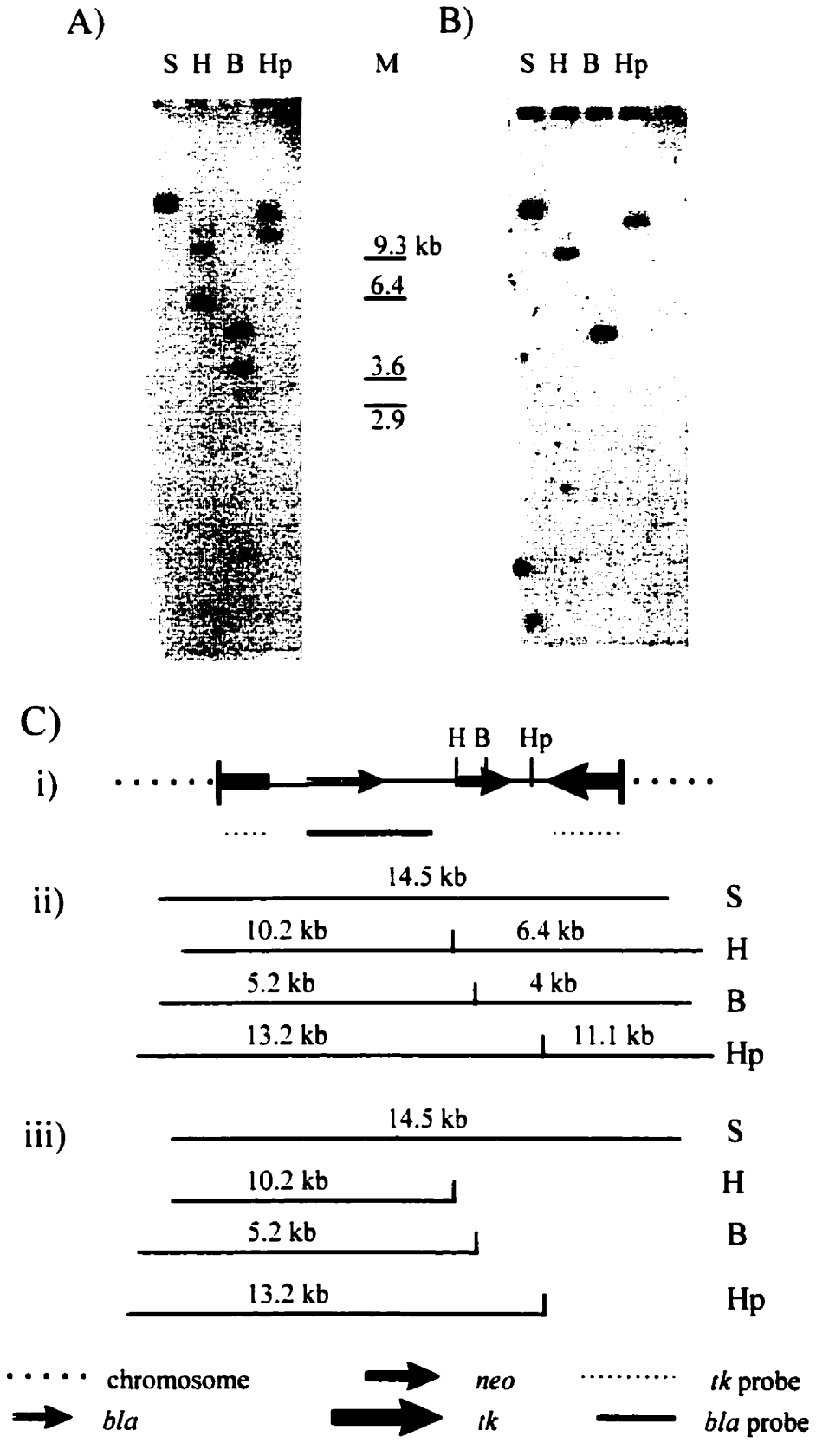
is a mixture of fragments resulting from digestion of the pSV2*neotk* plasmid with *Bam*HI, *Bgl*II and *Nde*I (Goring *et al.*, 1987), and hybridized with the *tk* gene (lane M). Plasmid pSV2*neotk* differs from pSVL in that it possesses *Bam*HI sites at both ends of the *tk* gene, whereas the *Bam*HI site at the 3' end of the *tk* gene in pSVL was previously eliminated by a backfill procedure (Fouquet and DuBow, 1992b). Of the three restriction endonucleases used to digest the total cellular DNA, there is no *Sac*I restriction site present within the integrated plasmid, and therefore this enzyme should digest within chromosomal DNA. Digestion with *Hpa*I is not represented because it was unsuccessful. The hybridization pattern obtained with the *Sac*I digestion helped to determine the number of copies of the inserted plasmid. In the case of SVLG-1, only one band was obtained (Figure 3.6A, lane S). Assuming that digestion of the total cellular DNA was complete, the intensity of the band and its large size (19 kb) suggest that two plasmids may have integrated in tandem, since the plasmid pSVL alone measures only 9.3 kb. Figure 3.6C (i) illustrates a diagram of a potential pattern of integration of the plasmid into the chromosomal DNA. The two bands produced by *Hin*dIII with the *tk* probe (9.3 and 12.7 kb) (Figure 3.6A, lane H), also support the hypothesis that two plasmids are present within the cellular DNA in tandem [Figure 3.6C (ii)]. In addition, the hybridization pattern observed from the DNA digested with *Bgl*II aided in the determination of the site of insertion of the plasmid. Digestion by this enzyme produced three fragments with the *bla* probe (3.9, 5 and 9.3 kb) (Figure 3.6B, lane B), which suggests that the plasmid has integrated by a recombination event either within its *bla* gene, or at a site very close to this gene [Figure 3.6C (i) (iii)].

The hypothesis for the method of plasmid insertion would expect three bands to be generated upon digestion of the DNA with *Bgl*II and after probing with the *tk* gene.

However, only two bands were observed (Figure 3.6A, lane B). This result may be explained by the fact that the three expected fragments would be very close in size (4.4, 4.8, and 4.8 kb) [Figure 3.6C (ii)]. Three bands were also expected after *Hind*III and *Bgl*II digestions and probing with the *bla* probe [Figure 3.6C (iii)]. Of the three bands obtained, one was very faint (Figure 3.6A, lanes H, B). These faint bands might be the result of weak hybridization.

The SVLG-6 cell line also consists of AK143B cells stably transfected with pSVL, and grown in the presence of G418. The hybridization pattern of SVLG-6 is shown in Figures 3.7A and B. In this case, it appeared that only one copy of the plasmid has been integrated, since only one band was observed following digestion with *Sac*I (14.5 kb) and probing with either the *tk* or *bla* probes (lane S, Figures 3.7A and B, respectively). The two bands resulting from digestion with *Hind*III and hybridization with the *tk* probe (6.4 and 10.2 kb) (Figure 3.7A, lane H) suggest that in this cell line, the plasmid has integrated through a site near or within the *tk* gene [Figure 3.7C (i)]. The number of bands generated by hybridization with the *tk* probe in the *Bgl*II lane (4 and 5.2 kb) and *Hpa*I lane (11.1 and 13.2 kb) (Figure 3.7A, lanes B and Hp, respectively) were also suggestive of insertion within the *tk* gene [Figure 3.7C (ii)]. The pattern of hybridization obtained with the *bla* probe (Figure 3.7B) generated single bands whose sizes (10.2, 5.2 and 13.2 kb for *Hind*III, *Bgl*II and *Hpa*I, respectively) coincided with bands observed following hybridization with the *tk* probe [Figure 3.7C (iii)]. This data supported the postulated pattern of single plasmid insertion [Figure 3.7C (i)].

Figure 3.7. Southern blot analysis of total cellular DNA from the SVLG-6 cell line. (A) Hybridization of the Southern blot with the *tk* probe. (B) Hybridization of the Southern blot with the *bla* probe. (C) Putative integration pattern of the pSVL plasmid within the SVLG-6 cellular DNA: (i) single insertion of pSVL; (ii) corresponding restriction endonuclease pattern of total cellular DNA from SVLG-6 hybridized with the *tk* probe; (iii) corresponding restriction endonuclease pattern of total cellular DNA from SVLG-6 hybridized with the *bla* probe. The sizes (in kb) indicate the bands obtained in the blots. M, marker DNA: pSV2*neotk* plasmid digested with *Bam*HI, *Bgl*II, and *Nde*I and hybridized with the *tk* probe; B, *Bgl*II; H, *Hind*III; Hp, *Hpa*I; S, *Sac*I.


A)
S H B Hp
M
9.3 kb
6.4
3.6
2.9
B)
S H B Hp
C)
i)
H B Hp
ii)
14.5 kb
S
10.2 kb
6.4 kb
H
5.2 kb
4 kb
B
13.2 kb
11.1 kb
Hp
iii)
14.5 kb
S
10.2 kb
H
5.2 kb
B
13.2 kb
Hp
chromosome
neo
tk probe
bla
tk
bla probe

The SVLH-6 cell line also consists of AK143B cells transfected with the pSVL plasmid, but was selected in medium containing HAT. Since hybridization with the *tk* probe was not successful (data not shown), a probe consisting of a fragment ofthe *neo* gene was used (Figure 3.8A), in addition to the *bla* probe (Figure 3.8B). The unique band obtained from hybridization following digestion with *Sac*I (12.1 kb) (lane S, Figures 3.8A and 3.8B) suggested that a single copy of the plasmid was integrated. The single band observed (7.7 kb) following digestion with *Hin*dIII and hybridization with either the *neo* or *bla* probes (lane H, Figures 3.8A and 3.8B) supported this hypothesis. The presence of single bands following *Bgl*II digestion, and hybridization with either the *neo* (5.6 kb) or the *bla* probes (4.7 kb) (lane B, Figures 3.8A and 3.8B, respectively), suggested that the point of integration in the plasmid has occurred within the region between the 3' ends of the *neo* and *tk* genes [Figure 3.8C (i)]. The size of the band obtained following hybridization with the *bla* probe (4.7 kb), was similar to the band containing the *bla* gene within the plasmid (4.4 kb) [Figure 3.8C (iii)], and also supported the pattern of integration indicated. The identical size of the bands following *Hpa*I digestion and hybridization with either probe (13.8 kb) [Figure 3.8C (ii) (iii)] also supported the proposed pattern of integration.

Growth of AK143B cells transfected with the pSVL plasmid, in a medium containing both HAT and G418, has produced the stable SVLHG cell lines. The pattern of hybridization of SVLHG-2 total cellular DNA with either the *tk* or the *bla* probe is shown in Figures 3.9A and 3.9B, respectively. The single band resulting from hybridization of *Sac*I (13.4 kb) and *Hin*dIII (6.2 kb) digests (lanes S and H, respectively), suggested that a single copy of the pSVL plasmid had been integrated. Hybridization following *Hpa*I digestion,

Figure 3.8. Southern blot analysis of total cellular DNA from the SVLH-6 cell line. (A) Hybridization of the Southern blot with the *neo* probe. (B) Hybridization of the Southern blot with the *bla* probe. (C) Putative integration pattern of the pSVL plasmid within the SVLH-6 cellular DNA: (i) single insertion of pSVL; (ii) corresponding restriction endonuclease pattern of total cellular DNA from SVLH-6 hybridized with the *neo* probe; (iii) corresponding restriction endonuclease pattern of total cellular DNA from SVLH-6 hybridized with the *bla* probe. The sizes (in kb) indicate the bands obtained in the blots. M, marker DNA: pSV2*neotk* plasmid digested with *Bam*HI, *Bgl*II, and *Nde*I and hybridized with the *neo* probe; B, *Bgl*II; H, *Hind*III; Hp, *Hpa*I; S, *Sac*I.

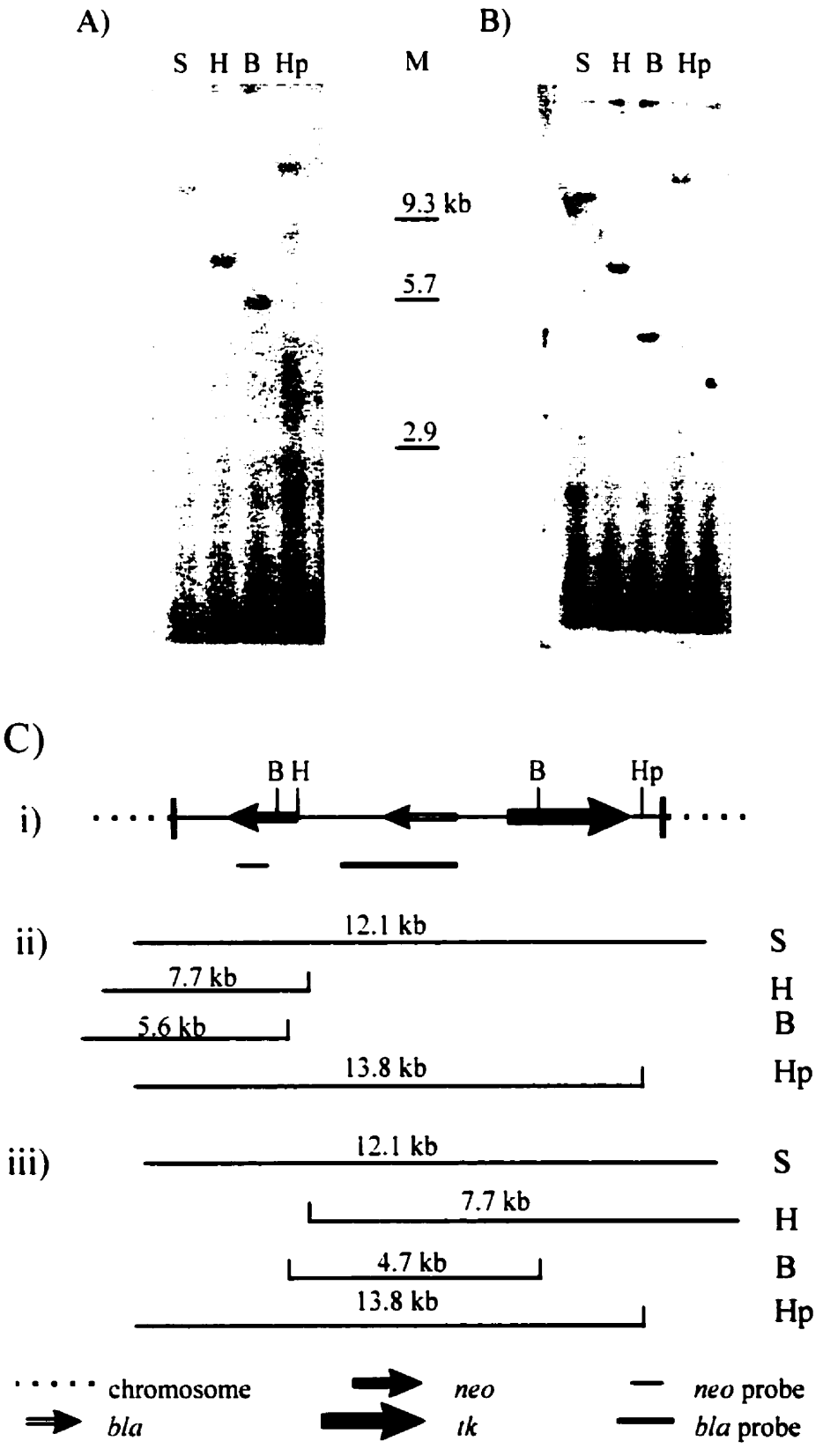
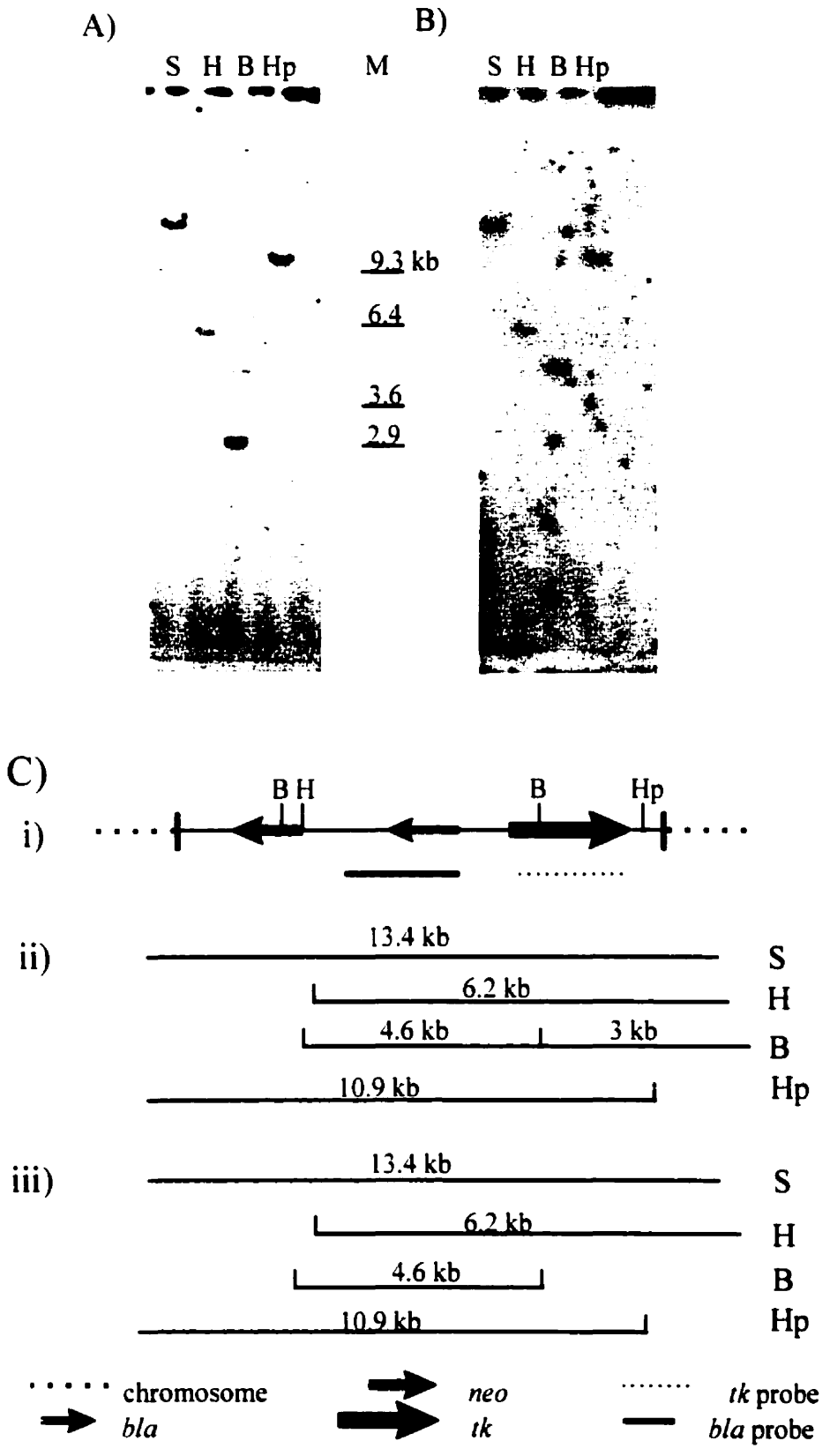

A)
S H B Hp
M
9.3 kb
5.7
2.9
B)
S H B Hp
C)
i)
B H
B
Hp
ii)
12.1 kb
S
7.7 kb
H
5.6 kb
B
13.8 kb
Hp
iii)
12.1 kb
S
7.7 kb
H
4.7 kb
B
13.8 kb
Hp
chromosome
bla
neo
tk
neo probe
bla probe

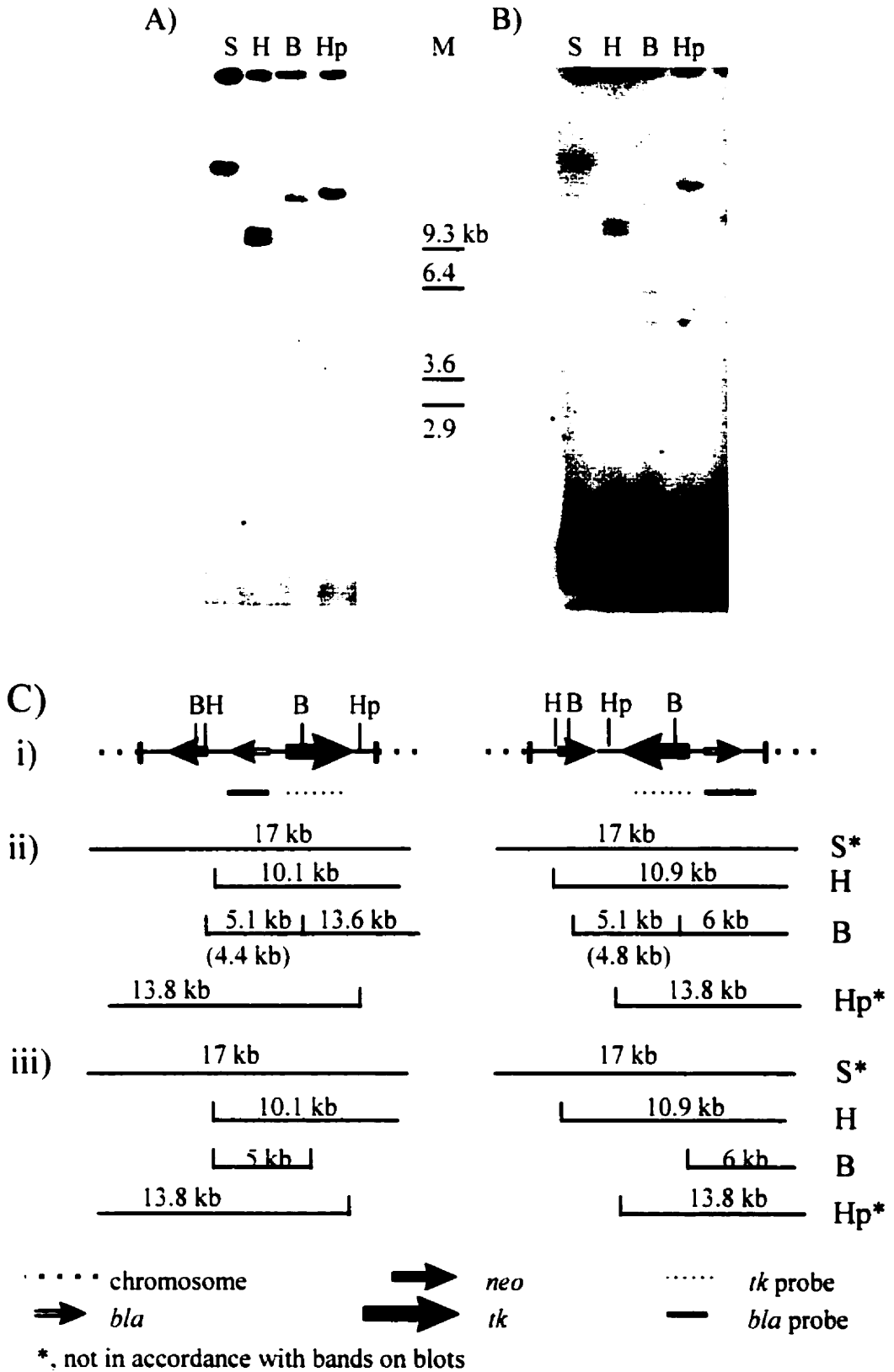
Figure 3.9. Southern blot analysis of total cellular DNA from the SVLHG-2 cell line. (A) Hybridization of the Southern blot with the *tk* probe. (B) Hybridization of the Southern blot with the *bla* probe. (C) Putative integration pattern of the pSVL plasmid within the SVLHG-2 cellular DNA: (i) single insertion of pSVL; (ii) corresponding restriction endonuclease pattern of total cellular DNA from SVLHG-2 hybridized with the *tk* probe; (iii) corresponding restriction endonuclease pattern of total cellular DNA from SVLHG-2 hybridized with the *bla* probe. The sizes (in kb) indicate the bands obtained in the blots. M, marker DNA: pSV2*neotk* plasmid digested with *Bam*HI, *Bgl*II, and *Nde*I and hybridized with the *tk* probe; B, *Bgl*II; H, *Hind*III; Hp, *Hpa*I; S, *Sac*I.


A)
B)
S H B Hp
M
S H B Hp
9.3 kb
6.4
3.6
2.9
C)
i)
B H
B
Hp
ii)
13.4 kb
S
6.2 kb
H
4.6 kb
3 kb
B
10.9 kb
Hp
iii)
13.4 kb
S
6.2 kb
H
4.6 kb
B
10.9 kb
Hp
chromosome
bla
neo
tk
tk probe
bla probe

with either the *tk* or the *bla* probe, also produced a single band (10.9 kb) (lane Hp), as did hybridization with the *bla* probe following *Bgl*II digestion (4.6 kb) (Figure 3.9B, lane B). These results suggested that the plasmid has not integrated at the region of the plasmid containing the *bla* gene. Instead, it was hypothesized that integration had occurred within the region between the 3' ends of the *tk* and *neo* genes as illustrated in Figure 3.9C (i). This hypothesis was supported by the identical size of the bands obtained upon hybridization with either probe following *Hin*dIII (6.2 kb) and *Hpa*I (10.9 kb) digestions [Figure 3.9C (ii) (iii)]. The band obtained following *Bgl*II digestion and hybridization with the *bla* probe (Figure 3.9B, lane B) had a size of 4.6 kb which is similar to the size of the *Bgl*II fragment in the plasmid (4.4 kb) [Figure 3.9C (iii)]. One of the two bands resulting from hybridization with the *tk* probe (4.6 kb) following *Bgl*II digestion (Figure 3.9A, lane B) also corresponded in size to the band observed following hybridization with the *bla* probe [Figure 3.9C (ii)], thus reinforcing the proposed plasmid integration pattern [Figure 3.9C (i)].

Analysis of SVLHG-4 DNA, from another stable SVLHG cell line, is shown in Figure 3.10. The two bands obtained following *Hin*dIII digestion (10.1 and 10.9 kb) suggested that in this cell line pSVL has integrated in two copies (lane H, Figures 3.10A and B). The hybridization pattern with *Bgl*II suggested that the two pSVL plasmids have integrated at different sites within the chromosomal DNA, with one plasmid being inserted between the 3'ends of the *tk* and *neo* genes, and the other between the *bla* and *neo* genes [Figure 3.10C (i)]. Three bands were observed from digestion with *Bgl*II, following hybridization with the *tk* probe (5.1, 6 and 13.6 kb) (Figure 3.10A, lane B). This pattern was in accordance with the proposed sites of integration, since two of the fragments

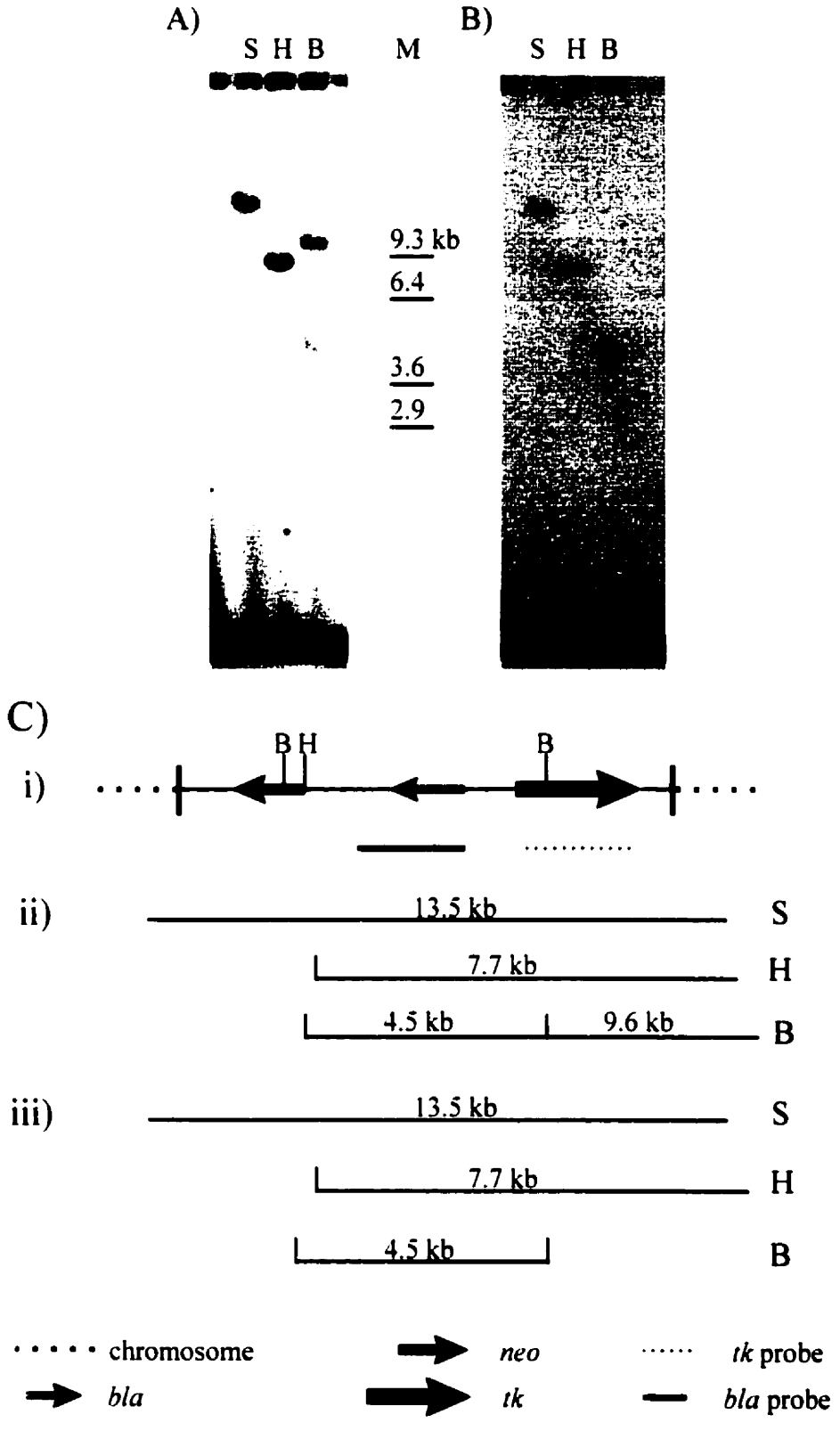
Figure 3.10. Southern blot analysis of total cellular DNA from the SVLHG-4 cell line. (A) Hybridization of the Southern blot with the *tk* probe. (B) Hybridization of the Southern blot with the *bla* probe. (C) Putative integration pattern of the pSVL plasmid within the SVLHG-4 cellular DNA: (i) double insertion of pSVL; (ii) corresponding restriction endonuclease pattern of total cellular DNA from SVLHG-4 hybridized with the *tk* probe; (iii) corresponding restriction endonuclease pattern of total cellular DNA from SVLHG-4 hybridized with the *bla* probe. The sizes (in kb) indicate the bands obtained in the blots. M, marker DNA: pSV2*neotk* plasmid digested with *Bam*HI, *Bgl*II, and *Nde*I and hybridized with the *tk* probe; B, *Bgl*II; H, *Hind*III; Hp, *Hpa*I; S, *Sac*I.


A)
S H B Hp
M
9.3 kb
6.4
3.6
2.9
B)
S H B Hp
C)
BH
B
Hp
HB
Hp
B
i)
ii)
17 kb
10.1 kb
5.1 kb
13.6 kb
(4.4 kb)
13.8 kb
17 kb
10.9 kb
5.1 kb
6 kb
(4.8 kb)
13.8 kb
S*
H
B
Hp*
iii)
17 kb
10.1 kb
5 kb
13.8 kb
17 kb
10.9 kb
6 kb
13.8 kb
S*
H
B
Hp*
chromosome
bla
neo
tk
tk probe
bla probe
*, not in accordance with bands on blots

produced by digestion of the plasmid with *Bgl*II are close in size (4.4 kb and 4.8 kb) and probably hybridized as one band [Figure 3.10C (ii)]. Hybridization with the *bla* probe following *Bgl*II digestion produced two bands (5 and 6 kb) (Figure 3.10B, lane B), with the same size as two of the bands from hybridization with the *tk* probe [Figure 3.10C (ii) (iii)]. This hybridization pattern correlates with the postulated plasmid integration pattern. Two bands would be expected from hybridization with either the *tk* or *bla* probe following *Sac*I digestion, according to the proposed pattern of integration [Figure 3.10C (ii) (iii)]. However, one strong band and two faint bands were observed (Figures 3.10A and B, lane S) which led to the suggestion that the strong band was due to incomplete digestion. On the other hand, the two faint bands could be products of overdigestion. In this instance, the bands resulting from the inserted plasmids would coincide, since one band was observed (17 kb). The *Hpa*I digestion produced a single band (13.8 kb) following hybridization with either probe (Figures 3.10A and B, lane Hp), when two bands would be expected from the proposed pattern of integration [Figure 3.10C (ii) (iii)]. The *Hpa*I hybridization pattern obtained would suggest that the two expected bands are close in size and therefore coincide. Despite the discrepancies in the *Sac*I and *Hpa*I digestions, the model proposed is at this time the most plausible.

The hybridization pattern obtained for yet another stable SVLHG cell line DNA, SVLHG-5, suggested that one copy of the pSVL plasmid has integrated into the chromosomal DNA (Figures 3.11A, B). A single band was obtained following hybridization with the *tk* and *bla* probes following digestion with *Sac*I (13.5 kb) or *Hin*dIII (7.7 kb) (lanes S and H, respectively). The size of the band resulting from hybridization with the *bla* probe following *Bgl*II digestion (4.5 kb) (Figure 3.11B, lane B), suggested that

Figure 3.11. Southern blot analysis of total cellular DNA from the SVLHG-5 cell line. (A) Hybridization of the Southern blot with the *tk* probe. (B) Hybridization of the Southern blot with the *bla* probe. (C) Putative integration pattern of the pSVL plasmid within the SVLHG-5 cellular DNA: (i) single insertion of pSVL; (ii) corresponding restriction endonuclease pattern of total cellular DNA from SVLHG-5 hybridized with the *tk* probe; (iii) corresponding restriction endonuclease pattern of total cellular DNA from SVLHG-5 hybridized with the *bla* probe. The sizes (in kb) indicate the bands obtained in the blots. M, marker DNA: pSV2*neotk* plasmid digested with *Bam*HI, *Bgl*II, and *Nde*I and hybridized with the *tk* probe; B, *Bgl*II; H, *Hind*III; S, *Sac*I.

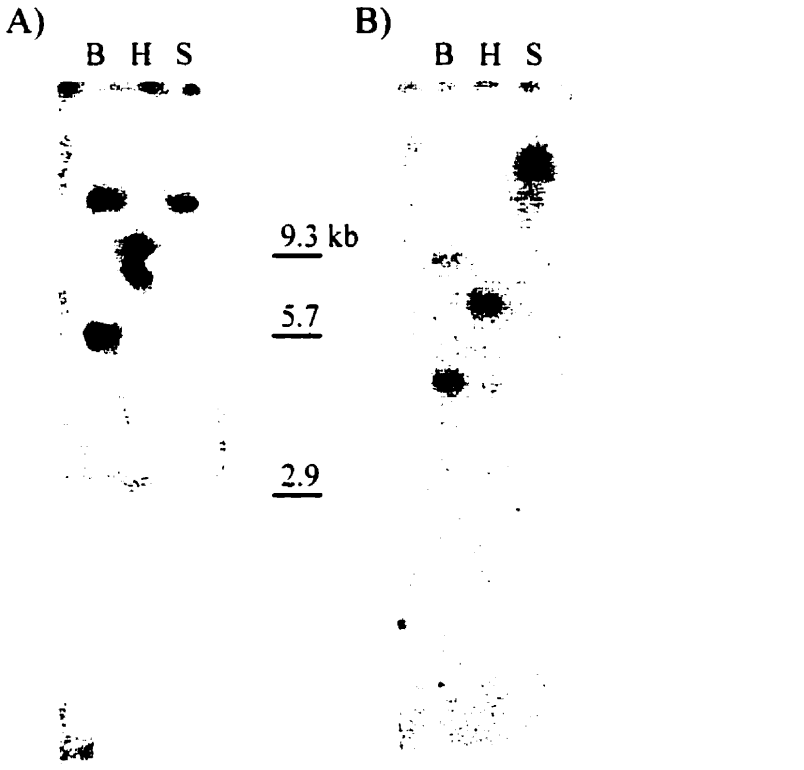

A)
S H B
M
9.3 kb
6.4
3.6
2.9
B)
S H B
C)
i)
B H
B
ii)
13.5 kb
S
7.7 kb
H
4.5 kb
9.6 kb
B
iii)
13.5 kb
S
7.7 kb
H
4.5 kb
B
chromosome
bla
neo
tk
tk probe
bla probe

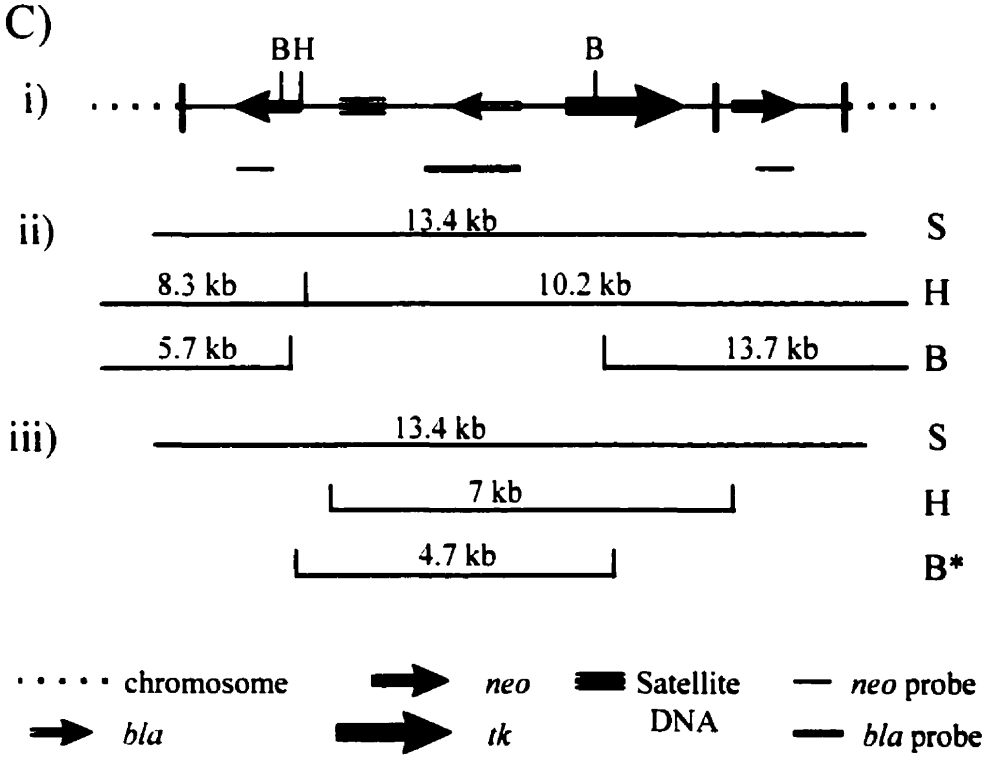
the plasmid has integrated between the 3' ends of the *tk* and *neo* genes [Figure 3.11C (i) (iii)]. Hybridization with the *tk* probe following *Bgl*II digestion produced two bands (9.6 kb and 4.5 kb) (Figure 3.11A, lane B), which was expected from the proposed pattern of integration [Figure 3.11C (ii)]. The 4.5 kb band coincided with the band obtained with the *bla* probe [Figure 3.11C (ii) (iii)].

3.2.2 Cells Containing the pCFD1 Plasmid

Stable cell lines transfected with the pCFD1 plasmid, that contained the 1.797 kb satellite DNA sequence next to the 5' end of the *neo* gene, were also studied. The pCFD1 plasmid has a total size of 11.1 kb. Cell lines grown in the presence of G418 alone are referred to as D1G. Total cellular DNA isolated from D1G-2 cells was studied through hybridization with the *neo* and *bla* probes (Figures 3.12A and B, respectively). The DNA used as a marker (M) was a mixture of fragments resulting from digestion of the pSV2*neotk* plasmid with *Bam*HI, *Bgl*II and *Nde*I, and hybridized with the *neo* probe. As suggested, by the size of the single band obtained following *Sac*I digestion and hybridization with either probe (13.4 kb), only one copy of the pCFD1 plasmid was integrated into the chromosomal DNA (lane S, Figures 3.12A and B). The two bands, resulting from hybridization with the *neo* probe following *Hin*dIII digestion (8.3 and 10.2 kb) (Figure 3.12A, lane H), suggested that a duplication of part of the *neo* gene may have occurred [Figure 3.12C (i) (ii)]. A single band (7 kb) was obtained following hybridization with the *bla* probe (Figure 3.12B, lane H). The size of this band did not coincide with any of the two bands resulting from hybridization with the *neo* probe, suggesting that plasmid integration has occurred into the region between the 3' ends of the *neo* and *tk* genes [Figure 3.12C (i) (ii) (iii)]. Based on this

Figure 3.12. Southern blot analysis of total cellular DNA from the D1G-2 cell line. (A) Hybridization of the Southern blot with the *neo* probe. (B) Hybridization of the Southern blot with the *bla* probe. (C) Putative integration pattern of the pCFD1 plasmid within the D1G-2 cellular DNA: (i) single insertion of pCFD1; (ii) corresponding restriction endonuclease pattern of total cellular DNA from D1G-2 hybridized with the *neo* probe; (iii) corresponding restriction endonuclease pattern of total cellular DNA from D1G-2 hybridized with the *bla* probe. The sizes (in kb) indicate the bands obtained in the blots. M, marker DNA: pSV2*neotk* plasmid digested with *Bam*HI, *Bgl*II, and *Nde*I and hybridized with the *neo* probe; B, *Bgl*II; H, *Hind*III; S, *Sac*I.



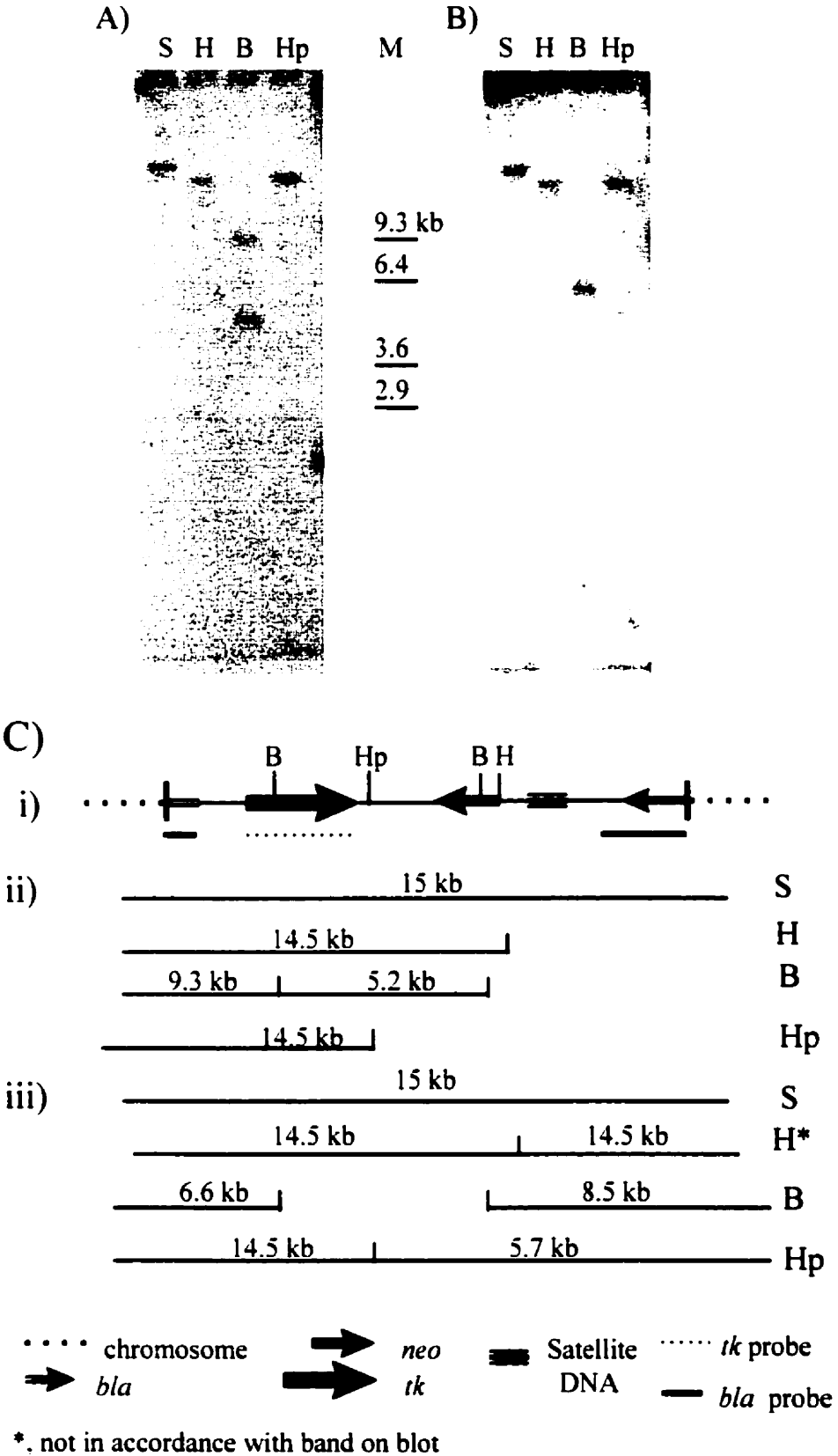


*, not in accordance with band on blot

pattern of integration, two bands would be obtained following *Bgl*II digestion by hybridization with the *neo* probe [Figure 3.12C (ii); Figure 3.12A, lane B] (5.7 and 13.7 kb bands). Only one band should result from hybridization with the *bla* probe [Figure 3.12C (iii)], however two bands of 4.7 and 9.3 kb were observed (Figure 3.12B, lane B). The 9.3 kb faint band could result from non-specific hybridization. Moreover, the 4.7 kb band should have a size of about 6.2 kb because of the presence, in the plasmid, of the satellite DNA sequence. Although the proposed model of integration appeared the most plausible, the discrepancy observed with the *Bgl*II hybridization pattern suggested that another event had likely occurred.

The hybridization pattern of DNA from the D1G-3 stable cell line is shown in Figures 3.13A and B. The single band, obtained following *Sac*I digestion and hybridization with either the *tk* or *bla* probe (15 kb), suggested that only one copy of the pCFD1 plasmid was integrated (lane S, Figures 3.13A and B). Two bands appeared to be present following *Bgl*II digestion and hybridization with the *bla* probe, suggesting that integration may have occurred within the *bla* gene or the region adjacent to it [Figure 3.13B, lane B; Figure 3.13C (i)]. Following this integration model, digestion with *Bgl*II, and hybridization with the *tk* or *bla* probe, would produce two bands [Figure 3.13C (ii) (iii)]. Of the two bands observed following hybridization with the *tk* probe, the 5.2 kb band, is close in size to the 4.8 kb *Bgl*II band of the plasmid [Figure 3.13C (ii)]. The band produced with the *bla* probe following *Hin*dIII digestion (14.5 kb) (Figure 3.13B, lane H) had an identical size to the band obtained with the *tk* probe (Figure 3.13A, lane H). However, two bands would be expected, from the proposed model of integration, following hybridization with the *bla* probe [Figure 3.13C (iii)]. The two bands might therefore be similar or close in size. Digestion with *Hpa*I, and

Figure 3.13. Southern blot analysis of total cellular DNA from the D1G-3 cell line. (A) Hybridization of the Southern blot with the *tk* probe. (B) Hybridization of the Southern blot with the *bla* probe. (C) Putative integration pattern of the pCFD1 plasmid within the D1G-3 cellular DNA: (i) single insertion of pCFD1; (ii) corresponding restriction endonuclease pattern of total cellular DNA from D1G-3 hybridized with the *tk* probe; (iii) corresponding restriction endonuclease pattern of total cellular DNA from D1G-3 hybridized with the *bla* probe. The sizes (in kb) indicate the bands obtained in the blots. M, marker DNA: pSV2*neotk* plasmid digested with *Bam*HI, *Bgl*II, and *Nde*I and hybridized with the *tk* probe; B, *Bgl*II; H, *Hin*dIII; H, *Hpa*I; S, *Sac*I.

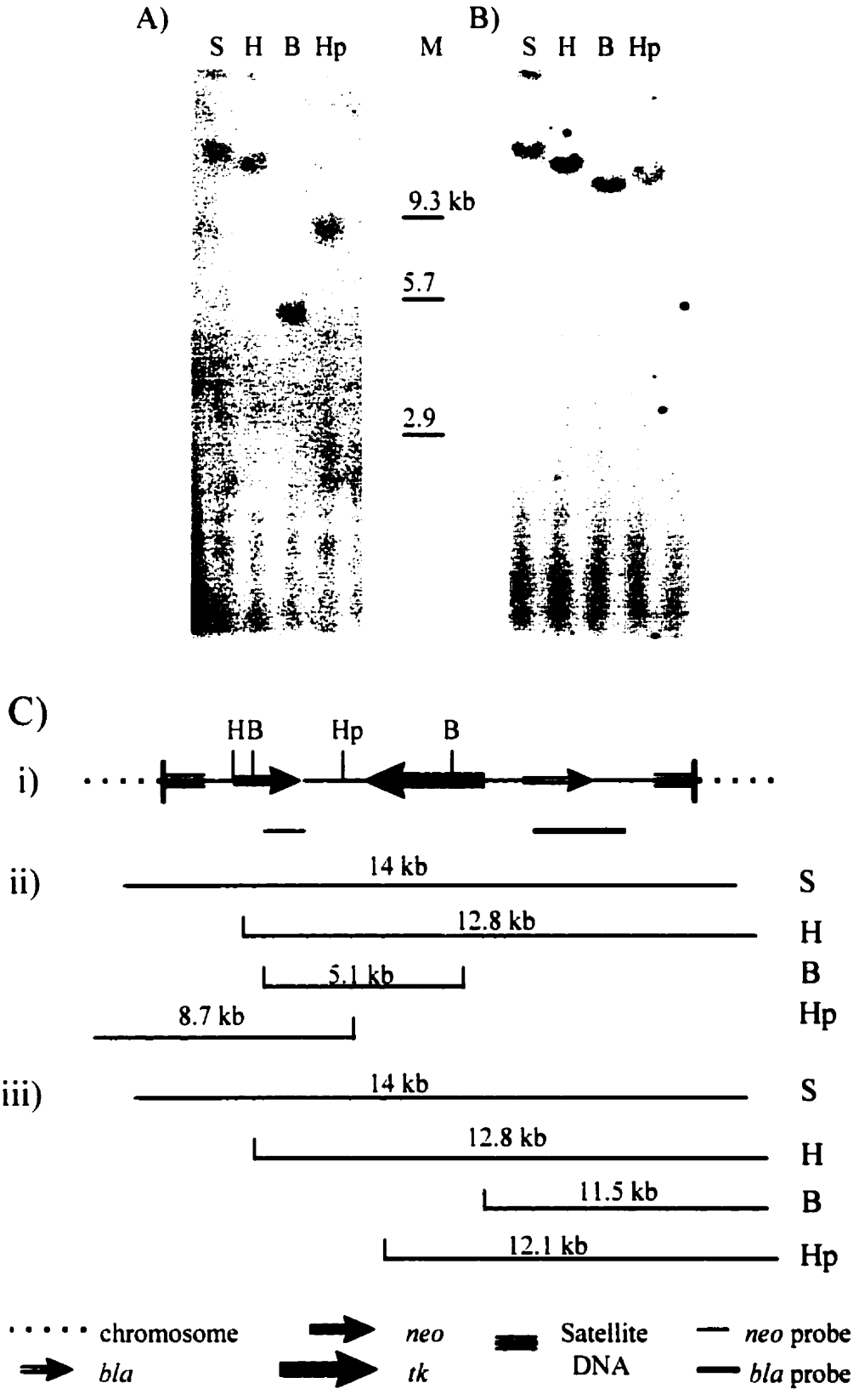


*, not in accordance with band on blot

hybridization to the *tk* probe, produced a 14.5 kb band (Figure 3.13A, lane Hp). Hybridization with the *bla* probe also produced a 14.5 kb band, but a second band (5.7 kb), though faint, also appeared to be present (Figure 3.13B, lane Hp). The presence of these two bands correlated with the proposed pattern of integration [Figure 3.13C (iii)]. The faint bands observed from hybridization with the *bla* probe following *Bgl*II and *Hpa*I digestion could result from weak hybridization (Figure 3.13B, lanes B and Hp, respectively).

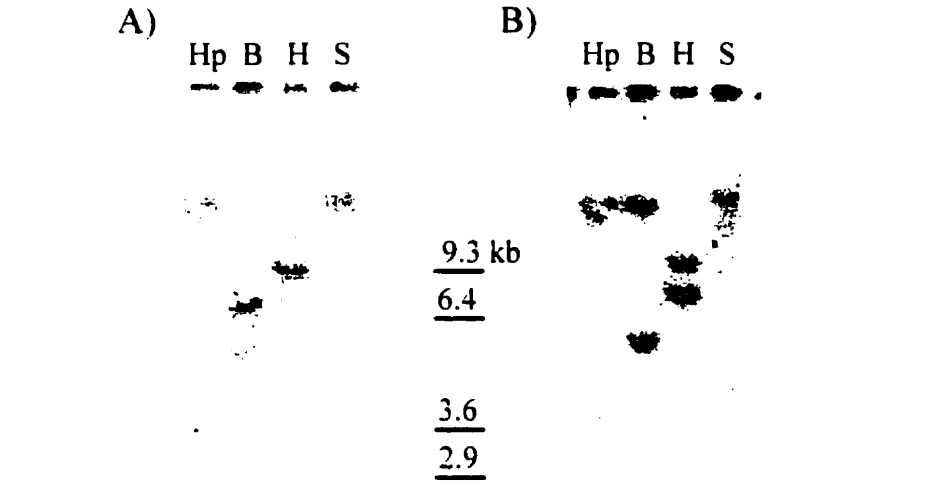
Stable cell lines containing the pCFD1 plasmid and grown in HAT-containing medium are designated D1H. DNA from cell line D1H-2 was hybridized with either the *neo* (Figure 3.14A) or the *bla* probe (Figure 3.14B). The hybridization pattern shown in Figures 3.14A and B suggested that a single copy of the pCFD1 plasmid has integrated. A single band was obtained with both probes following digestion with either *Sac*I or *Hin*dIII (14 and 12.8 kb, respectively) (lanes S and H, Figures 3.14A, and B). The size of the hybridization bands obtained following *Hin*dIII digestion was identical with both probes (12.8 kb). However, the sizes of the bands following *Hpa*I digestion and hybridization were different with the *neo* (8.7 kb) and *bla* (12.1 kb) probes, suggesting that plasmid integration has occurred within the satellite DNA sequence of the plasmid [Figure 3.14C (i) (ii) (iii)]. The hybridization pattern and size of bands obtained following *Bgl*II digestion with the *neo* and *bla* probes (5.1 and 11.5 kb, respectively) supported the suggested site of integration [Figure 3.14C (ii) (iii)]. Hybridization following *Hpa*I digestion produced a single band of 8.7 kb with the *neo* probe (Figure 3.14A, lane Hp), and 12.1 kb with the *bla* probe (Figure 3.14B, lane Hp). These bands of different sizes correlated with the proposed integration pattern [Figure 3.14C (ii) (iii)].

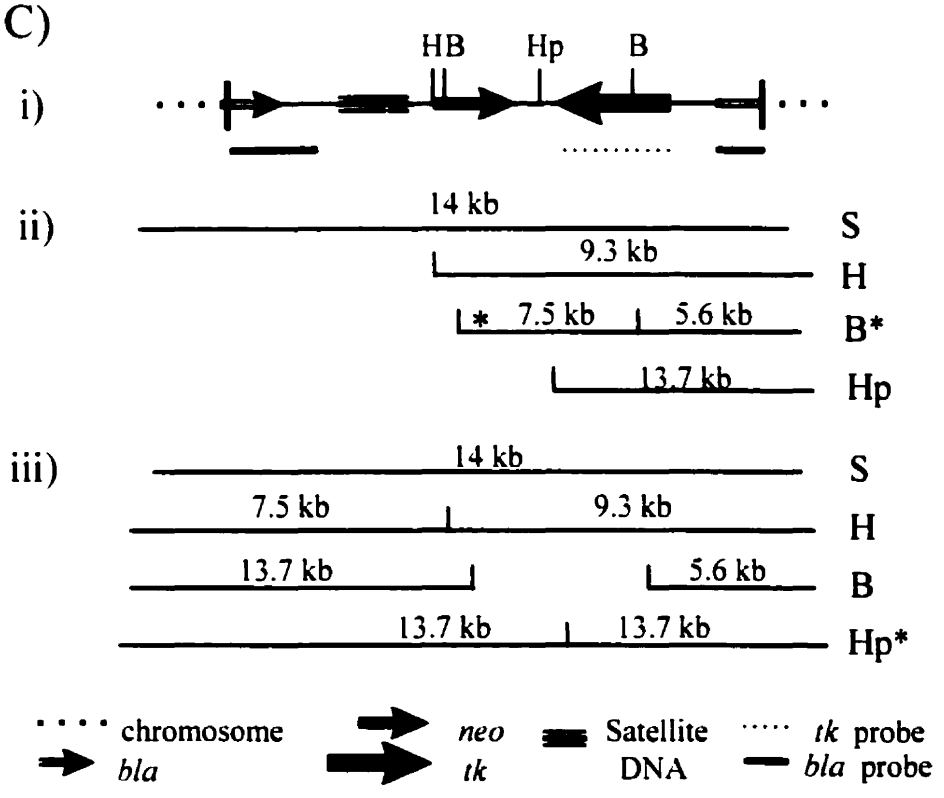
Figure 3.14. Southern blot analysis of total cellular DNA from the D1H-2 cell line. (A) Hybridization of the Southern blot with the *neo* probe. (B) Hybridization of the Southern blot with the *bla* probe. (C) Putative integration pattern of the pCFD1 plasmid within the D1H-2 cellular DNA: (i) single insertion of pCFD1; (ii) corresponding restriction endonuclease pattern of total cellular DNA from D1H-2 hybridized with the *neo* probe; (iii) corresponding restriction endonuclease pattern of total cellular DNA from D1H-2 hybridized with the *bla* probe. The sizes (in kb) indicate the bands obtained in the blots. M, marker DNA: pSV2*neotk* plasmid digested with *Bam*HI, *Bgl*II, and *Nde*I and hybridized with the *neo* probe; B, *Bgl*II; H, *Hin*dIII; H, *Hpa*I; S, *Sac*I.


A)
S H B Hp
B)
M
S H B Hp
9.3 kb
5.7
2.9
C)
HB
Hp
B
i)
ii)
14 kb
S
12.8 kb
H
5.1 kb
B
8.7 kb
Hp
iii)
14 kb
S
12.8 kb
H
11.5 kb
B
12.1 kb
Hp
chromosome
neo
Satellite DNA
neo probe
bla
tk
bla probe

DNA from a different D1H cell line, D1H-3, was hybridized with the *tk* and *bla* probes (Figures 3.15A and B). Hybridization with either probe, following *Sac*I digestion, revealed that there was likely only one integrated copy of the pCFD1 plasmid, since a single band of 14 kb was obtained (lane S, Figures 3.15A and B). A single band was also observed following *Hin*dIII digestion and hybridization with the *tk* probe (9.3 kb) (Figure 3.15A, lane H). However, hybridization with the *bla* probe produced two bands (7.5 and 9.3 kb) (Figure 3.15B, lane H), which suggested that integration has occurred within the *bla* gene or the region adjacent to it [Figure 3.15C (i) (iii)]. According to this pattern of integration, two bands would be obtained following *Bgl*II digestion and hybridization with the *tk* probe [Figure 3.15C (ii)]. Two bands were observed in the blot (5.6 and 7.5 kb) (Figure 3.15A, lane B). Hybridization with the *bla* probe following *Bgl*II digestion also produced two bands (5.6 and 13.7 kb) (Figure 3.15A, lane B). Of the two bands observed, the 5.6 kb band coincided with the 5.6 kb band from the *tk* hybridization (Figure 3.15A, lane B). It therefore appeared that the 7.5 kb band observed after hybridization with the *tk* probe corresponds to the *Bgl*II fragment of the plasmid [Figure 3.15C (ii)]. However, in the plasmid, this fragment has a size of only 4.8 kb. Despite this discrepancy, the proposed pattern of plasmid insertion was the most probable. Hybridization with the *tk* probe following *Hpa*I digestion appeared to produce a single band (13.7 kb) (Figure 3.15A, lane Hp). This is the pattern expected from the proposed integration event [Figure 3.15C (ii)]. Two bands would also be expected following hybridization with the *bla* probe [Figure 3.15C (iii)]. There appeared to be a single band on the blot (13.7 kb) (Figure 3.15B, lane Hp), which could result from two bands that are close or identical in size.

Figure 3.15. Southern blot analysis of total cellular DNA from the D1H-3 cell line. (A) Hybridization of the Southern blot with the *tk* probe. (B) Hybridization of the Southern blot with the *bla* probe. (C) Putative integration pattern of the pCFD1 plasmid within the D1H-3 cellular DNA: (i) single insertion of pCFD1; (ii) corresponding restriction endonuclease pattern of total cellular DNA from D1H-3 hybridized with the *tk* probe; (iii) corresponding restriction endonuclease pattern of total cellular DNA from D1H-3 hybridized with the *bla* probe. The sizes (in kb) indicate the bands obtained in the blots. M, marker DNA: pSV2*neotk* plasmid digested with *Bam*HI, *Bgl*II, and *Nde*I and hybridized with the *tk* probe; B, *Bgl*II; H, *Hin*dIII; H, *Hpa*I; S, *Sac*I.





*, not in accordance with band on blot

3.3. Analysis of Short-Term Expression of the Transfected *tk* Gene

In an attempt to determine if the negative effect observed following transfection of pSVL-containing satellite II DNA occurred prior to or after integration into the chromosomal DNA, the short-term expression of the transfected genes was analyzed through Northern blotting. Hybridization, with the *tk* probe, of the total RNA extracted from cells over a short period of time following transfection (0 to 72 hours) is shown in Figure 3.16. Three different cell lines were analyzed which had been transfected with plasmids pSVL, pCFD1 and pDSD2. Plasmids pCFD1, containing the 1.797 kb satellite II DNA sequence, and pDSD2, containing an 813 bp fragment of the satellite II DNA, had produced a very low frequency of stable transfectants (<1%) (Tables 3.1 and 3.2). Total cellular RNA from mock-transfected AK143B cells was used as a control (Figure 3.16A, lanes 13-16). As seen with plasmids pSVL, pCFD1 and pDSD2, no *tk* RNA was detected at 0 hours, as expected (lanes 1, 5, 9). At 24, 48 and 72 hours post-transfection, a smear was observed for all three plasmids (lanes 2-4, 6-8, 10-12). Hybridization of the same membranes with a metallothionein-II gene present in the genome had indicated that the smears were not due to degradation of the total cellular RNA extracted from cells transfected with any of the plasmids, as a single band was observed (Figure 3.16B, lanes 1-16).

In order to try to enhance the sensitivity of the hybridization reaction, the Northern analysis was repeated using only poly(A)$^+$ RNA (Figure 3.17). No signal was observed following hybridization with the *tk* probe (Figure 3.17A). Hybridization with the metallothionein-II gene (Figure 3.17B) showed that the poly(A)$^+$ RNA, at various time points, was not present in sufficient quantity (lanes 3-6, 8-13, 15, and 16), for several

Figure 3.16. Northern blot analysis of total RNA from cells transfected with plasmids pSVL, pCFD1 or pDSD2. The (-) sign indicates the mock-transfected AK143B cells. Times of RNA extraction are indicated in hours. (A) Hybridization of total RNA with the *tk* probe. (B) Hybridization of total RNA with a probe from the metallothionein-II gene. M represents the RNA ladder used as a marker.

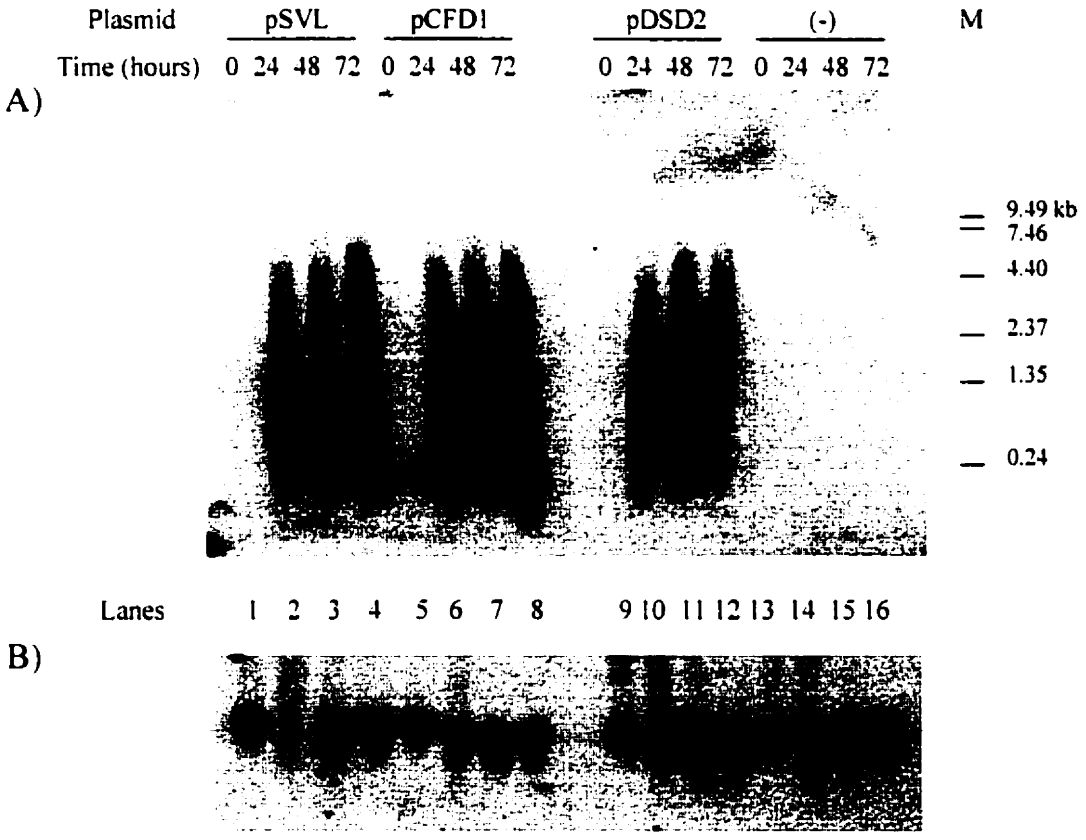
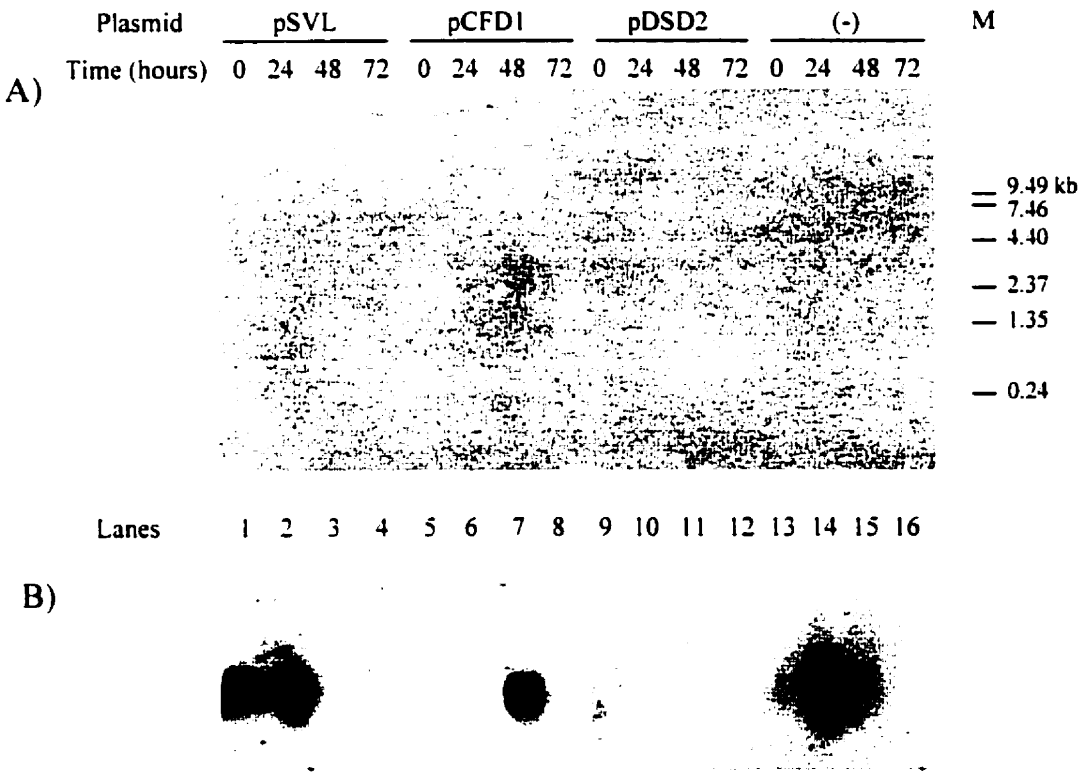

Plasmid
pSVL
pCFD1
pDSD2
(-)
M
Time (hours) 0 24 48 72 0 24 48 72 0 24 48 72 0 24 48 72
A)
9.49 kb
7.46
4.40
2.37
1.35
0.24
Lanes 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
B)

Figure 3.17. Northern blot analysis of poly(A)$^+$ RNA from cells transfected with plasmids pSVL, pCFD1 or pDSD2. The (-) sign indicates the mock-transfected AK143B cells. Times of RNA extraction are indicated in hours. (A) Hybridization of poly(A)$^+$ RNA with the *tk* probe. (B) Hybridization of poly(A)$^+$ RNA with a probe from the metallothionein-II gene. M represents the RNA ladder used as a marker.


Plasmid
pSVL
pCFD1
pDSD2
(-)
M
Time (hours)
0 24 48 72 0 24 48 72 0 24 48 72 0 24 48 72
A)
9.49 kb
7.46
4.40
2.37
1.35
0.24
Lanes
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
B)

plasmid-transfected cells and two of the mock-transfected AK143B. In the other cases, plasmid pSVL at 0 and 24 hours, pCFD1 at 48 hours, and AK143B at 24 and 48 hours (lanes 1, 2, 7, 14, and 15), a signal was detected following hybridization with the metallothionein-II gene.

CHAPTER 4

DISCUSSION

A cloned 1.797 kb human satellite II DNA sequence (pCFD1) was previously observed to exert a very severe negative effect on the recovery of stable *tk*$^-$/*neo*^{r} transfectants (<1% transfection efficiency) (Fouquet and DuBow, 1992b). This effect was observed with the satellite II DNA sequence positioned next to the 5' end of the *neo* gene in the plasmid vector in which it was inserted (pSVL). The present study has confirmed that the observed negative effect is in fact related to the position of the satellite II DNA sequence relative to the *neo* and *tk* marker genes in the plasmid. For example, plasmid pDH1, which was constructed with the satellite II DNA sequence between the 3' ends of the *neo* and *tk* genes, did not exert as negative an effect on the transfection efficiency as plasmid pCFD1 (28% as opposed to <1%) (Table 3.1). The location of the satellite II DNA sequence away from the 5' end of the *neo* gene therefore appeared to attenuate the negative effect on recovery of stable *tk*$^-$/*neo*^{r} transfectants. However, the proportion of transfectants recovered was still lower as compared to the vector pSVL. The size control pCF1.8, which contains a fragment from bacteriophage λ DNA of similar size to the satellite DNA, has shown that the larger plasmid size did not affect transfection efficiency to the same extent (Table 3.1). Therefore, there appears to be an event specific to the satellite II DNA sequence, that is involved in affecting the recovery of stable transfectants. The effect seemed nonetheless to be predominant with the satellite II DNA sequence in close juxtaposition to the *neo* gene.

When satellite II DNA fragments of 813 bp and 983 bp were used in place of the 1.797 kb fragment, a negative effect on transfection efficiency was still observed (Table

3.2). These smaller derivatives of the 1.797 kb satellite DNA sequence displayed an orientation-dependent effect, with the two different orientations of the fragments relative to the *neo* gene producing different effects. The linearization of the plasmids containing these satellite II DNA fragments had not eliminated the severe effect observed with pDSD2 and pDSD3, each containing a 813 bp or a 983 bp DNA fragment, respectively. However, it did produce an increase in the proportion of transfectants with the pDSD1 and pDSD4 plasmids, which were producing higher values than pDSD2 and pDSD3 in the circular conformation. The increase in transfection efficiency observed with linear plasmids (Chu *et al.*, 1987; Colbere-Garapin *et al.*, 1979) has been attributed to the fact that free ends of DNA molecules are more recombinogenic (Subramanian, 1979; Wilson *et al.*, 1982). However, transfection studies with linear pCFD1, which contains the full size 1.797 kb satellite II DNA sequence, previously demonstrated only a very moderate increase in transfection efficiency (from <1% to 1%) (Fouquet and DuBow, 1992b), as was observed with pDSD2.

Using smaller satellite II DNA fragments ranging from 354 bp to 620 bp eliminated the severe negative effect observed with the larger sequences (Table 3.3). However, in the case of three of these satellite DNA fragments (354, 458, and 620 bp) an orientation-dependent effect was still observed, with lower transfection efficiencies observed with one of the two orientations of the satellite DNA fragments. Again, a control size plasmid, pDSL1 demonstrated that the larger plasmid size did not affect transfection efficiency. There is therefore something particular to the satellite DNA fragments, or their specific orientation, which produces an effect on recovery of stable transfectants. The particularity of the 620 bp satellite DNA fragment consists in the presence of a 49 bp region devoid of

the pentameric repeat unit 5'-TTCCA-3' (Figure 2.2). The other small fragments do not present specific features, other than their similar AT base pair content (~60%).

This study has also shown that the recovery of stable *neo*^{r} transfectants, from plasmids containing various sizes of the satellite DNA sequence, was particularly affected, as shown by the lower transfection efficiencies observed in the presence of G418 alone, as compared to HAT (Tables 3.4 and 3.5). These results suggested that the *neo* gene was affected the most from the presence of the satellite II DNA. In addition, the frequency of transfection, in the presence of single drugs, was higher than the value observed in the presence of both HAT and G418. The presence of both drugs in the selective media likely poses a greater constraint on the transfected cells, as it selects for cells that have successfully integrated the transfected plasmids and are stably expressing both the *neo* and *tk* genes.

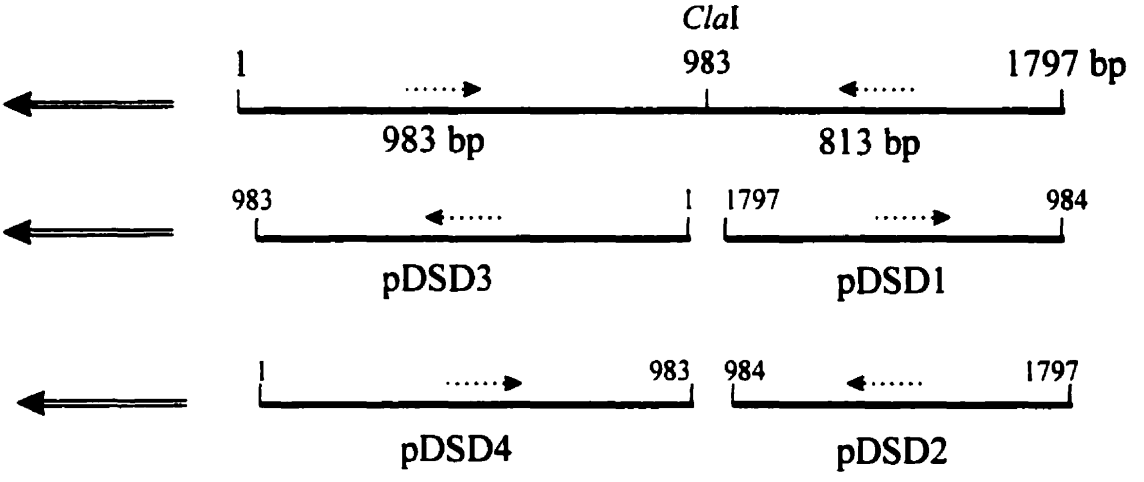
The event(s) responsible for the severe negative effect observed with the cloned satellite II DNA may reside in the nature of the satellite sequence itself. Their original location within heterochromatin, the very compact region of chromosomes, suggest that the satellite II DNA sequences used in these experiments are prone to heterochromatinization. This process would cause the binding of specific proteins to the repetitive DNA sequences present in the satellite DNA, thus leading to condensation of the DNA. A heterochromatinization event could have taken place either prior to, or after, chromosomal integration of the transfected plasmids, thus causing lower expression of either the *neo* or *tk* genes, which would ultimately lead to decreased transfection efficiency.

Satellite II DNA sequences may, themselves, possess recognition sites for proteins involved in heterochromatinization. The presence of initiator sites in heterochromatic domains was proposed in the boundary model of Tartof *et al.* (1984). If those recognition

sites are present in the satellite II DNA sequences located in the plasmids, they would be able to initiate a heterochromatinization process, as has been observed for the phenomenon of position-effect variegation (PEV). These sites could be present in both the 813 and 983 bp fragments of the satellite II DNA sequence, since both fragments produced a severe negative effect on the recovery of stable transfectants. Due to their repetitive nature, these satellite DNA sequences could be recognized by proteins involved in heterochromatin formation. If such a heterochromatinization process initiates at the satellite DNA, it could possibly spread to the *neo* gene, due to its close juxtaposition, interfering with the access of transcriptional factors. Therefore, the *neo* gene would appear to be part of heterochromatin and not be efficiently expressed. As explained in Chapter 1, expression of euchromatic genes placed in the vicinity of heterochromatin components can be negatively affected. Impaired expression of the *neo* gene could thus explain the low recovery of stable *neo*$^{+}$ transfectants. Because of the presence of satellite II DNA sequences adjacent to the *neo* gene in the transfected plasmids, heterochromatinization of this gene could even take place at euchromatic chromosomal sites of integration. As the 813 and 983 bp fragments of the satellite DNA sequence showed an orientation-dependent effect, it is possible that there is direction to the heterochromatinization process. As proposed by Tartof *et al.* (1984), the compaction of heterochromatic domains may occur by polarized spreading. Recognition sites for heterochromatinization could be present in both the 813 and 983 bp fragments, but initiate spreading in opposite directions (Figure 4.1).

There appeared to be a limit to the size at which satellite II DNA sequences can exert a negative effect, since the recovery of transfectants seemed to be less affected with fragments of less than 813 bp. However, as previously mentioned, the orientation of the

Figure 4.1. Proposed direction of heterochromatinization in the 983 and 813 bp fragments of the 1.797 kb satellite DNA sequence. The direction changes depending on the orientation of the fragments in the plasmid relative to the 5' end of the *neo* gene.



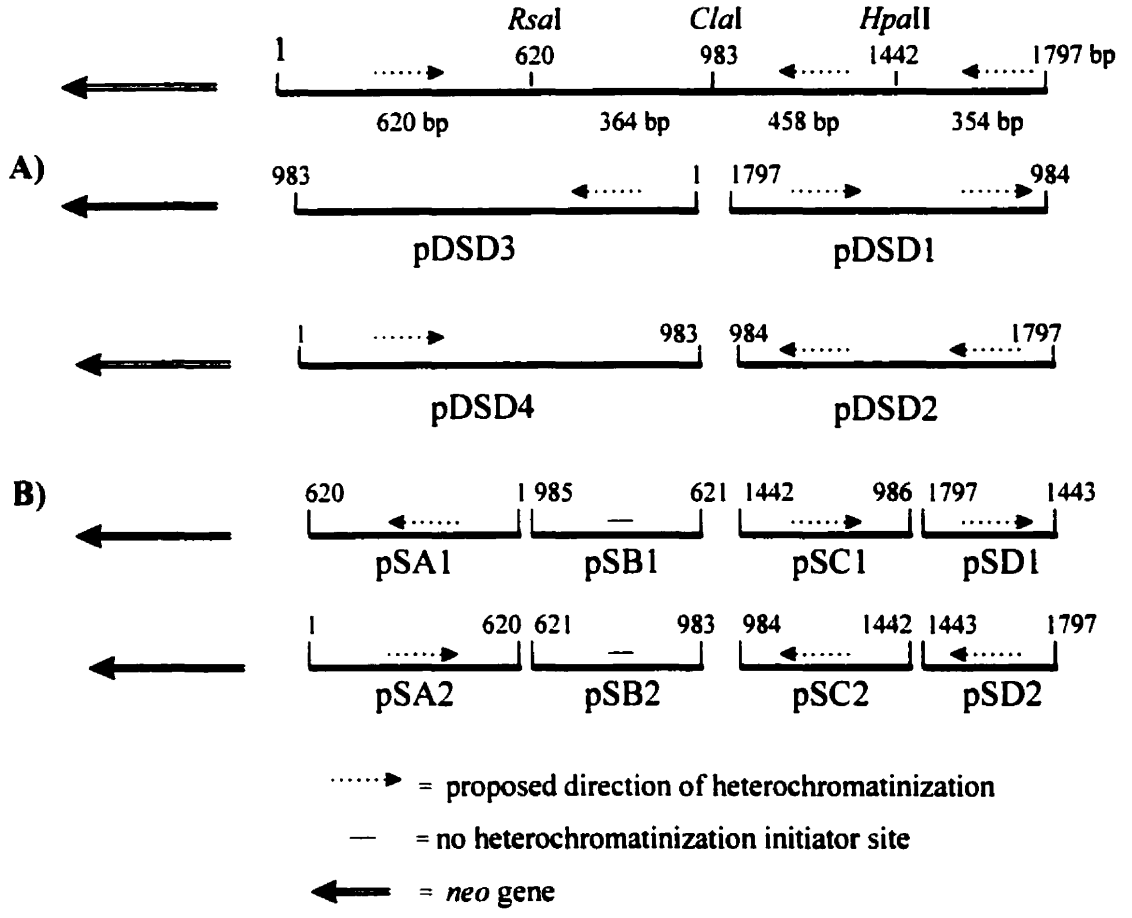
⋯⋯▶ = proposed direction of heterochromatinization

⟸ = *neo* gene

satellite DNA fragments, relative to the genes in the plasmid still appeared to be an important factor. Therefore, the theory of direction of spreading for heterochromatinization could still hold true with these smaller satellite II DNA sequences. As the 354, 458 and 620 bp fragments were associated with lower transfection efficiencies in only one of the two orientations (Table 3.3), initiator sites could be present in these three sequences (pSA1 and pSA2, pSC1 and pSC2, pSD1 and pSD2 plasmids) (Figure 4.2B). Conversely, since the 364 bp DNA fragment did not induce low transfection efficiencies (Table 3.3), there may not be initiator sites in this sequence (pSB1 and pSB2 plasmids) (Figure 4.2B). The direction of heterochromatinization in pSC2 (458 bp satellite DNA fragment) and pSD2 (354 bp satellite DNA fragment) would correlate with the direction in pDSD2 (813 bp satellite DNA fragment), as shown in Figure 4.2A, B. However, the direction would not seem to correlate in pSA2 (620 bp satellite DNA fragment) and pDSD3 (983 bp satellite DNA fragment) (Figure 4.2A, B), which both produced a low transfection efficiency (Tables 3.2 and 3.3). Therefore, heterochromatinization might not be solely responsible for the effect observed with these satellite DNA sequences. A higher order structure of the satellite DNA sequences may also play a role.

The search for curved regions in the 1.797 kb satellite II DNA sequence was performed using the Bend-it program of Gabrielian *et al.* (1998). Curved motifs usually have curvature values between 5 and 25° per helical turn, and straight motifs give values below 5° per helical turn. The degree of curvature was not found to be very pronounced in the different regions of the 1.797 kb satellite II DNA sequence. The most curved region, located within the 620 bp DNA sequence, reached 9° per helical turn, while the other curved motifs, scattered throughout the 1.797 kb DNA sequence, were approximately 6.5° per

Figure 4.2. Proposed direction of heterochromatinization in the various fragments of the 1.797 kb satellite DNA sequence. The direction changes depending on the orientation of the fragments relative to the 5' end of the neo gene in the plasmids.

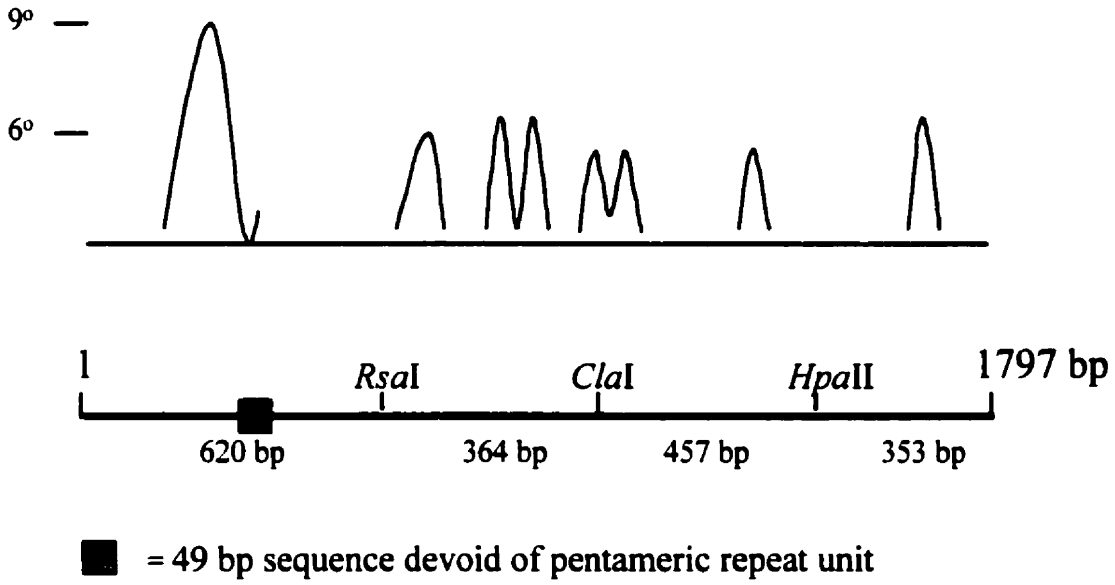

RsaI
ClaI
HpaII
1
620
983
1442
1797 bp
620 bp
364 bp
458 bp
354 bp
A)
983
1
1797
984
pDSD3
pDSD1
1
983
984
1797
pDSD4
pDSD2
B)
620
1
985
621
1442
986
1797
1443
pSA1
pSB1
pSC1
pSD1
1
620
621
983
984
1442
1443
1797
pSA2
pSB2
pSC2
pSD2
= proposed direction of heterochromatinization
— = no heterochromatinization initiator site
= neo gene

helical turn (Figure 4.3). The 49 bp region, devoid of the 5 bp satellite DNA repeat unit, could be implicated in the higher degree of curvature observed for the 620 bp sequence. The various effects seen with different orientations of the satellite II DNA sequences could relate to the positioning of the curved motifs, creating different conformations of the fragments. These different conformations could have distinct effects on heterochromatinization of the fragment, or on the binding of transcription factors. However, the particular conformations of repetitive DNA sequences involved in the binding of hetorochromatin proteins have not been identified to date.

Whether or not the size of the satellite II DNA sequence present in the plasmids is sufficient to produce the negative effects observed should also be considered. A previous study indicated that a 1Mb heterochromatic segment was insufficient to induce detectable PEV on adjacent genes located far from heterochromatin (Tolchkov *et al.*, 1997). However, a study implicating telomeric satellite DNA, as low as 1356 bp in total, has shown that this amount of DNA can greatly affect the expression of a flanking gene (Maiorano *et al.*, 1997). It was also shown previously (Fouquet and DuBow, 1992b), and in this study that a 1.797 kb satellite II DNA sequence was able to produce severely negative effects on stable transfection of adjacent genes. Our results suggest that satellite DNA sequences, as small as 813 bp are able to exert the same negative effect.

The establishment of stable cell lines involves the integration of the transfected plasmids into chromosomal DNA (Scangos *et al.*, 1981). The presence of satellite II DNA sequences in the plasmids may favor a mechanism for targeting integration from a particular site in the plasmid. Examination of cell lines transfected with plasmid pCFD1 indicated that the satellite II DNA sequence in the plasmid is not necessarily a target for integration, but is

Figure 4.3. Representation of the position of curved regions in the various fragments of the 1.797 kb satellite DNA sequence. The curvature is expressed in degrees (°) per helical turn.



■ = 49 bp sequence devoid of pentameric repeat unit

used as an insertion site for the D1H-2 cell line (Table 4.1). Due to the very low transfection frequency in the presence of G418, the *neo* gene could be expected to be a target site for the integration event. Cell lines grown in medium containing G418 alone, however, would indicate the stable transfectants capable of expressing the *neo* gene. In our studies, cell lines grown in the presence of HAT alone, such as D1H-2 and D1H-3, did not seem to possess a pCFD1 plasmid integrated within the *neo* gene. In these cell lines, integration appeared to have occurred within the satellite II DNA, for D1H-2, or within the region encompassing the *bla* gene, for D1H-3 (Table 4.1). For the D1G-2 cell line, duplication of part of the *neo* gene may have resulted during integration of the pCFD1 plasmid [Figure 3.12C (i)]. This indicated that rearrangements may be associated with the presence of the satellite II DNA. Moreover, the postulated pattern of integration was difficult to ascertain for cell lines D1G-2 and D1H-3, suggesting that rearrangements may in fact occur with the integration of satellite DNA-containing plasmids. For example, in the D1H-3 cell line, a band that was larger than expected was observed, indicating that amplification or insertion of a sequence may have occurred [Figure 3.15C (ii)]. DNA transfection, however, is not usually associated with a high frequency of spontaneous DNA rearrangements (Perucho and Wigler, 1981), although, DNA rearrangements have been observed following integration of transfected sequences into mice cells (Butner and Lo, 1986). The nature of these rearrangements was postulated to consist of amplification and deletion events. In addition, interspersed between copies of integrated DNA, unidentified sequences were observed.

The most common site of integration among the cell lines examined appeared to be between the 3' ends of the *neo* and *tk* genes in the transfected plasmids (Table 4.1). Results

Table 4.1. List of the copy number and sites of integration of integrated plasmids in various cell lines.

	Copy number of integrated plasmids	Plasmid site of integration
D1G-2	1	between 3' ends of *neo* and *tk*
D1G-3	1	*bla* or adjacent region
D1H-2	1	Sat DNA
D1H-3	1	*bla* or adjacent region
SVLG-1	2 (adjacent)	*bla* or adjacent region
SVLG-6	1	*tk* gene
SVLH-6	1	between 3' ends of *neo* and *tk*
SVLHG-2	1	between 3' ends of *neo* and *tk*
SVLHG-4	2 (separate)	between 3' ends of *neo* and *tk*; between *bla* and *neo*
SVLHG-5	1	between 3' ends of *neo* and *tk*

from the cell lines analyzed indicated that integration in the marker genes in the plasmids was avoided. This is to be expected for cells grown in medium containing both HAT and G418, such as the SVLHG cell lines. Integration outside the *neo* and *tk* genes was also seen for cell lines grown in the presence of one drug or the other, to the exception of the SVLG-6 cell line which was postulated to contain the pSVL plasmid integrated through the *tk* gene [Figure 3.7C (i)]. If integration of the transfected plasmids did not occur through the marker genes, the negative effect observed with the presence of satellite DNA sequences could be of an epigenetic nature. The satellite II DNA sequence could have greatly affected the expression of the transfected genes (more specifically the *neo* gene as indicated above) by directing integration of the transfected plasmids into repetitive DNA sites in the chromosome, as was seen in previous studies (Talarico *et al.*, 1988). If integration occurred into highly repetitive DNA or heterochromatin, a heterochromatinization process would have impaired the expression of the transfected genes, as described above. However, if *neo* gene expression was affected in this manner, the *tk* gene would likely also have suffered because of the proximity of the two genes. Transfection results indicated, though, that the *tk* gene was not significantly affected (Table 3.4). Another hypothesis could involve insertion of the transfected plasmids in a euchromatic chromosomal site. A process of *trans*-inactivation, as described in section 1.6.1 (Henikoff and Comai, 1998), may then be responsible for the impaired expression of the *neo* gene. The satellite II DNA sequence in the plasmids may have associated with a heterochromatic site on a different chromosome, or in a region of the genome rich in heterochromatin. Gene expression might therefore have become affected due to the proximity of this gene next to this heterochromatic environment, leaving the more distally-located *tk* gene less affected.

The copy number of plasmids inserted into chromosomal DNA did not necessarily reflect the level of transfection efficiency. Cell lines containing the pSVL plasmid, which did not contain satellite II DNA sequences, did not necessarily possess more copies of integrated plasmids. However, the SVLHG-4 cell line seemed to contain two separate copies of pSVL (Table 4.1), and SVLG-1 contained two adjacent copies of this plasmid. Ligation of exogenous transfected DNA sequences before integration has been observed previously (Perucho *et al.*, 1980). Nonetheless, single plasmid insertions seemed to prevail among the cell lines studied. In our studies, electroporation was used as a method of transfection because it can result in less copies of integrated DNA per recipient cell than the calcium phosphate transfection procedure (Ausubel *et al.*, 1990). Another advantage of this method is the production of a high frequency of permanent transfectants (Chu *et al.*, 1987). The evaluation of expression of the transfected *neo* and *tk* genes, through transfection efficiency, was therefore generally made based on a similar number of genes.

Analysis of the short-term expression of transfected genes was undertaken to try to determine whether the satellite II DNA sequences were able to induce their effect prior to integration. Total RNA from AK143B cells transfected with pSVL, pCFD1 or pDSD2 produced a smear upon hybridization with the *tk* gene (Figure 3.16). Since hybridization with the metallothionein-II gene showed that the smears were not due to degradation of the RNA, low specificity of the *tk* probe to the total RNA could be responsible for the smears observed. In this case, the actual *tk* poly(A)$^+$ RNA would be masked by the observed smears. However, lighter exposure of the blots did not show the presence of a band at the *tk* poly(A)$^+$ RNA region (1.4 kb) (data not shown). Alternatively, transcripts of various sizes, containing the *tk* region, may have been produced from the transfected plasmids. This may

indicate that prior to integration, transcription of the genes in the plasmids proceeded beyond the transcriptional stop sites. Integration of the plasmids may be required to help stabilize the expression of these genes. No differences in the level of expression of the *tk* gene from either the plasmid vector (pSVL) or the plasmids containing satellite II DNA sequences (pCFD1 and pDSD2) could be detected.

Hybridization of the poly(A)$^+$ RNA, with the *tk* gene, from cells transfected with pSVL, pCFD1 or pDSD2 was also attempted. In this case, the presence of a signal was not detected (Figure 3.17). Although the isolation of poly(A)$^+$ RNA was not successful in all cases, the presence of metallothionein-II poly(A)$^+$ RNA in some of the samples suggested that, in these cases, there was not a large amount of *tk* mRNA being produced. There may not have been enough *tk* mRNA produced, or this mRNA may not have been present in sufficient quantity to allow it to be detected. Transient HSV-1 *tk* mRNA levels have previously been examined following transfection of HeLa cells with the *tk* gene (Gelman and Silverstein, 1985). It was shown that transfected cells accumulated very little *tk* mRNA 48 hours post-transfection. The amount of *tk* mRNA produced in our assay may therefore be below the limit of detection. In addition, it has also been observed that the *neo* gene can act as a transcriptional silencer, by exerting a *cis*-acting negative effect on expression from promoters in plasmid or retroviral vectors transfected into mammalian cells (Artlet *et al.*, 1991). This observation may also explain the low level of short-term *tk* mRNA expression. Another explanation may be that the method used to isolate the poly(A)$^+$ RNA was not successful in purifying the *tk* poly(A)$^+$ RNA. These short-term expression studies did not help in determining whether the satellite II DNA exerts negative effects on gene expression prior to plasmid integration into the chromosome. These analyzes were nonetheless

attempted in the hope that they would help in gaining a better understanding of the phenomenon observed with satellite II DNAs.

Future experiments could include analysis of the pattern of integration of plasmids such as pDSD2 and pDSD3, which had also displayed a severe effect on the recovery of stable transfectants. This analysis would provide more information on plasmid integration sites. In addition, the chromosomal DNA site into which transfected plasmids have integrated could be examined, in order to determine whether integration occurred into a site rich in repetitive DNA sequences or in a euchromatic site. This would help in obtaining a better understanding of the effects sustained by transfected genes. *In situ* hybridization would provide a direct way of identifying the chromosomal site of insertion of the transfected plasmids. It would also be interesting to examine the sequence of the integrated genes, to determine if mutational events are involved in the negative effects observed. In addition, a change in the satellite DNA sequence present in the transfected plasmids would help determine whether it is the particular DNA sequence used which creates the negative effects observed on the marker genes. Another human satellite DNA sequence containing a different base composition, and a different periodicity, could be used. Transfections with the new constructs would hopefully reveal whether the effects observed are due to the DNA sequence of the 1.797 kb human satellite II DNA. Satellite DNA sequences from other mammalian species could also be used to determine if there is a species-specificity to the negative effect observed, or if any satellite DNA sequence positioned next to the *neo* gene acts in the same manner.

The studies presented in this study support the observation that the presence of satellite II DNA sequences, part of the repetitive fraction of the human genome, can be

disruptive to the expression of euchromatic genes. Whether this behavior is correlated with genetic or epigenetic events is not clear. Further studies will hopefully help in a better understanding of the effects observed with satellite II DNA sequences.

REFERENCES

Adams RLP, Burdon RH, and Fulton J (1983). Methylation of satellite DNA. Biochem. Biophys. Res. Comm. *113*: 695-702.

Adams JW, Kaufman RE, Kretschmer PJ, Harrison M, and Nienhaus AW (1980). A family of long reiterated DNA sequences, one copy of which is next to the human beta globin gene. Nucl. Acids Res. *8*: 6113-6128.

Alfageme CR, Rudkin GT, and Cohen LH (1980). Isolation, properties and cellular distribution of D1, a chromosomal protein of *Drosophila*. Chromosoma *78*: 1-31.

Artlet P, Grannemann R, Friel J, Bartsh J, and Hauser H (1991). The prokaryotic neomycin-resistance-encoding gene acts as a transcriptional silencer in eukaryotic cells. Gene *99*: 249-254.

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, and Struhl K (ed.) (1990). Current Protocols in Molecular Biology, Vol. 1, Unit 9.3.

Bachmann L, Venanzetti F, and Sbordoni V (1996). Tandemly repeated satellite DNA of *Dolichopoda schiavazzii*: a test for models on the evolution of highly repetitive DNA. J. Mol. Evol. *43*: 135-144.

Baldwin L, and Macgregor HC (1985). Centromeric satellite DNA in the newt *Triturus cristatus karelinii* and related species: Its distribution and transcription on lampbrush chromosomes. Chromosoma *92*: 100-107.

Barceló F, Gutierrez F, Barjau I, and Portugal J (1998). A theoretical perusal of the satellite DNA curvature in tenebrionid beetles. J. Biomol. Struct. Dyn. *16(1)*: 41-50.

Barceló F, Pons J, Petitpierre E, Barjau I, and Portugal J (1997). Polymorphic curvature of satellite DNA in three subspecies of the beetle *Pimelia sparsa*. Eur. J. Biochem. *244*:318-324.

Barr ML, and Bertram EG (1949). A morphological distinction between neurones of the male and female and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis. Nature *163*: 676-677.

Belyaev ND, Keohane AM, and Turner BM (1996). Differential underacetylation of histones H2A, H3 and H4 on the inactive X chromosome in human female cells. Hum. Genet. *97*, 573-578.

Beridze TG (1972). DNA nuclear satellites of the genus *Phaseolus*. Biochim. Biophys. Acta *262*: 393-396.

Beridze TG (1986). *Satellite DNA*. Springler-Verlag, Berlin.

Beridze TG, Odintsova MS, and Sissakyan NM (1967). Distribution of DNA components of bean leaves in the fractions of cell structures. Mol. Biol. *1*: 142-153.

Bernat RL, Borisy GG, Rothfield NF, and Earnshaw WC (1990). Injection of anticentromere antibodies in interphase disrupts events required for chromosome movement in mitosis. J. Cell Biol. *111*: 1519-1533.

Bird AP (1987). CpG islands as gene markers in the vertebrate nucleus. Trends Genet. *3*: 342-347.

Birnstiel ML, Speirs J, Purdom I, Jones K, and Loening UE (1968). Properties and composition of the isolated ribosomal DNA satellite of Xenopus laevis. Nature *219*: 454-463.

Bischoff FR, Maier G, Tilz G, and Ponstingl H (1990). A 47-kDa human nuclear protein recognized by antikinetochore autoimmune sera is homologous with the protein encoded by RCC1, a gene implicated in onset of chromosome condensation. Proc. Natl. Acad. Sci. USA *87*: 8617-8621.

Blennow E, Telenius H, de Vos D, Larsson C, Henriksson P, Johansson O, Carter NP, and Nordenskjold M (1994). Tetrasomy 15q: two marker chromosomes with no detectable alpha-satellite DNA. Am. J. Hum. Genet. *54*: 877-883.

Bloom KS, and Carbon J (1982). Yeast centromere DNA is in a unique and highly ordered structure in chromosomes and small circular minichromosomes. Cell *29*: 305-317.

Braunstein M, Sobel RE, Allis CD, Turner BM, and Broach JR (1996). Efficient transcriptional silencing in Saccharomyces cerevisiae requires a heterochromatin histone acetylation pattern. Mol. Cell. Biol. *16*: 4349-4356.

Britten RJ, and Khone DE (1966). Nucleotide sequence repetition in DNA. Carnegie Inst. Wash. Year Book *65*: 78-106.

Britten RJ, and Khone DE (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.

Brown SW (1966). Heterochromatin. Science *151*: 417-425.

Brown W, and Tyler-Smith C (1995). Centromere activation. Trends Genet. *11*: 337-339.

Butner KA, and Lo CW (1986). High frequency DNA rearrangements associated with mouse centromeric satellite DNA. J. Mol. Biol. *187*: 547-556.

Callan HG (1982). Lampbrush chromosomes. Proc. R. Soc. Lond. (Biol.) *214*: 417-448.

Cavalier-Smith T (1978). Nuclear volume control by nucleoskeletal DNA, selection for cell volume and cell growth rate, and the solution of the DNA C-value paradox. J. Cell. Sci. *34*: 274-278.

Cavalier-Smith T (1980). How selfish is DNA? Nature *285*: 617-618.

Charlesworth B, Langley CH, and Stephan W (1986). The evolution of restricted recombination and the accumulation of repeated DNA sequences. Genetics *112*: 947-962.

Charlesworth B, Sniegowski P, and Stephan W (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. Nature *371*: 215-220.

Chu G, Hayakawa H, and Berg P (1987). Electroporation for the efficient transfection of mammalian cells with DNA. Nucl. Acids Res. *15*: 1311-1326.

Clarke L, and Carbon J (1980). Isolation of a yeast centromere and construction of a functional small circular chromosome. Nature *287*: 504-509.

Cléard F, Delattre M, and Spierer P (1997). SU(VAR)3-7, a *Drosophila* heterochromatin-associated protein and companion of HP1 in the genomic silencing of position-effect variegation. EMBO J. *16*: 5280-5288.

Coats SR, Zhang Y, and Epstein LM (1994). Transcription of satellite 2 DNA from the newt is driven by a snRNA type of promoter. Nucl. Acids Res. *22*: 4697-4704.

Colbere-Garapin F, Chousterman S, Horodniceann F, Kourilsky P, and Garaoin A (1979). Cloning of the active thymidine kinase gene of herpes simplex virus type I in *Escherichia coli* K-12. Proc. Natl. Acad. Sci. USA *76*: 3755-3759.

Cooke CA, Bernat RL, and Earnshaw WC (1990). CENP-B: a major human centromere protein located beneath the kinetochore. J. Cell Biol. *110*: 1475-1488.

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

Corneo G, Zardi L, and Polli E (1972). Elution of human satellite DNAs on a methylated albumin kieselguhr chromatographic column: isolation of satellite DNA. IV. Biochim. Biophys. Acta: *269*: 201-204.

Csink AK, and Henikoff S (1996). Genetic modification of heterochromatic association and nuclear organization in Drosophila. Nature *381*: 529-531.

Davis C, and Wyatt GR (1989). Distribution and sequence homogeneity of an abundant satellite DNA in beetle, *Tenebrio molitor*. Nucl. Acids Res. *17*: 5579-5586.

Diaz MO, Barsacchi-Pilone G, Mahon KA, and Gall JG (1981). Transcripts of both strands of a satellite DNA occur on lampbrush chromosome loops of the newt Notophtalamus. Cell *24*: 649-659.

Disney JE, Johnson KR, Magnuson NS, Sylvester SR, and Reeves R (1989). High-mobility group protein HMG-I localizes to G/Q- and C-bands of human and mouse chromosomes. J. Cell Biol. *109*: 1975-1982.

Doheny KF, Sorger PK, Hyman AA, Tugendreich S, Spencer F, and Hieter P (1993). Identification of essential components of the *S. cerevisiae* kinetochore. Cell *73*: 761-774.

Dombroski BA, Mathias SL, Nanthakumar E, Scott AF, and Kazazian HH Jr. (1991). Isolation of an active human transposable element. Science *254*: 1805-1808.

Dover GA, Brown S, Coen E, Dallas J, Strachan T, and Trick M (1982). The dynamics of genome evolution and species differentiation. In Dover GA, Flavell RB (eds), *Genome Evolution.* Academic Press, San Diego.

Dower WJ, Miller JF, and Ragsdale CW (1988). High efficiency transformation of *E.coli* by high voltage electroporation. Nucl. Acids Res. *16*: 6127-6145.

Drew HR, and Travers AA (1985). DNA bending and its relation to nucleosome positioning. J. Mol. Biol. *186*: 773-790.

Dupraw EJ (1970). *DNA and Chromosomes*. Holt, Rinehart and Winston, New York.

Earnshaw WC, Sullivan KF, Machlin PS, Cooke CA, Kaiser DA, Pollard TD, Rothfield NF, and Cleveland DW (1987). Molecular cloning of cDNA for CENP-B, the major human centromere autoantigen. J. Cell Biol. *104*: 817-829.

Earnshaw WC, Ratrie H, and Stetten G (1989). Visualization of centromere proteins CENP-B and CENP-C on a stable dicentric chromosome in cytological spreads. Chromosoma *98*: 1-12.

Eberl DF, Duyf BJ, and Hilliker AJ (1993). The role of heterochromatin in the expression of a heterochromatic gene, the *rolled* locus of *Drosophila melanogaster*. Genetics *134*: 277-292.

Ehrlich M, and Wang RYH (1981). 5-Methylcytosine in eukaryotic DNA. Science *212*: 1350-1357.

Eissenberg JC, Morris GD, Reuter G, and Hartnett T (1992). The heterochromatin-associated protein HP-1 is an essential protein in Drosophila with dosage-dependent effects on position effect variegation. Genetics *131*: 345-352.

Ekwall K, Olsson T, Turner BM, Cranston G, and Allshire RC (1997). Transient inhibition of histone deacetylation alters the structural and functional imprint at fission yeast centromeres. Cell *91*: 1021-1032.

Elder JF Jr., and Turner BJ (1995). Concerted evolution of repetitive DNA sequences in eukaryotes. Q. Rev. Biol. *70*: 297-320.

Elgin SCR (1997). Heterochromatin and gene regulation in Drosophila. Curr. Opin. Genet. Dev. *6*: 193-202.

Enukashvily NI, Lobov IB, and Podgornaya OI (1999). Human nuclear matrix contains a protein complex which can specifically recognize alpha-satellite DNA in vitro. Biochemistry (Mosc.) *64(4)*: 473-480.

Epstein LM, and Coats SR (1991). Tissue-specific permutations of self-cleaving newt satellite-2 transcripts. Gene *107*: 213-218.

Fanning TG, and Singer MF (1987). LINE-1: a mammalian transposable element. Biochim. Biophys. Acta *910*: 203-212.

Feinstein SI, Racaniello VR, Ehrlich M, Gehrke CW, Miller DA, and Miller OJ (1985). Pattern of undermethylation of the major satellite DNA of mouse sperm. Nucl. Acids Res. *13*: 3969-3978.

Felsenfeld G (1985). DNA. Sci. Am. *253(4)*: 58-67.

Feng Q, Moran JV, Kazazian HH Jr, and Boeke JD (1996). Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell *87*: 905-916.

Festenstein R, Tolaini M, Corbella P, Mamalaki C, Parrington J, Fox M, Miliou A, Jones M, and Kioussis D (1996). Locus control region function and heterochromatin-induced position effect variegation. Science *271*: 1123-1125.

Fouquet C, and DuBow MS (1992a). Detection of human satellite II DNA-binding polypeptides in HeLa cellular and nuclear extracts. Biochem. Int. *27(3)*: 457-464.

Fouquet C, and DuBow MS (1992b). Effect of *cis*-located human satellite DNA on the electroporation efficiency of *neo* and HSV-1 *tk* containing plasmids. Mut. Res. *284*: 321-328.

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

Gall JG, Cohen EH, and Polan ML (1971). Repetitive DNA sequences in *Drosophila*. Chromosoma *33*: 319-344.

Gasser SM, Amati BB, Cardenas ME, and Hofmann JFX (1989). Studies on scaffold attachment sites and relation to genome function. Int. Rev. Cytol. *119*: 57-96.

Gatti M, and Pimpinelli S (1992). Functional elements in *Drosophila melanogaster* heterochromatin. Ann. Rev. Genet. *26*: 239-275.

Gelman IH, and Silverstein S (1985). Identification of immediate early genes from herpes simplex virus that transactivate the virus thymidine kinase gene. Proc. Natl. Acad. Sci. USA *82*: 5265-5269.

Gluzman Y (1981). SV40-transformed simian cells support the replication of early SV40 mutants. Cell *23*: 175-182.

Goh P, and Kilmartin JV (1993). NCD10: a gene involved in chromosome segregation in Saccharomyces cerevisiae. J. Cell Biol. *121*: 503-512.

Goring DR, and DuBow MS (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. Commun. *133*: 195-201.

Goring DR, Gupta K, and DuBow MS (1987). Analysis of spontaneous mutations in a chromosomally located HSV-1 thymidine kinase (TK) gene in a human cell line. Somat. Cell. Mol. Genet. *13(1)*: 47-56.

Gowen JW, and Gay EH (1933). Eversporting as a function of the *Y* chromosome in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA *19*: 122-126.

Grady DL, Ratliff RL, Robinson DL, McCanlies EC, Meyne J, and Moyzis RK (1992). Highly conserved repetitive DNA sequences are present at human centromeres. Proc. Natl. Acad. Sci. USA *89*: 1695-1699.

Grant SG, and Chapman VM (1988). Mechanisms of X-chromosome regulation. Ann. Rev. Genet. *22*: 199-233.

Grigliatti T (1991). Position-effect variegation – an assay for nonhistone chromosomal proteins and chromatin assembly and modifying factors. Methods Cell Biol. *35*: 587-627.

Guzzo A, Karatzios C, Diorio C, and DuBow MS (1994). Metallothionein-II and ferritin H mRNA levels are increased in arsenite-exposed HeLa cells. Biochem. Biophys. Res. Commun. *205(1)*: 590-595.

Harrington JJ, Van Bokkelen G, Mays RW, Gustashaw K, and Willard HF (1997). Formation of *de novo* centromeres and construction of first-generation human artificial microchromosomes. Nat. Genet. *15*: 345-355.

Heitz E (1928). Das heterochromatin der moose. I. Jb. Wiss. Bot. *69*: 762-818.

Heller R, Brown KE, Burgtorf C, and Brown WRA (1996). Minichromosomes derived from the human Y chromosome by telomere-directed chromosome breakage. Proc. Natl. Acad. Sci. USA *93*: 7125-7130.

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

Henikoff S, and Comai L (1998). *Trans*-sensing effects: the ups and downs of being together. Cell *93*: 329-332.

Henikoff S, and Dreesen TD (1989). Trans-inactivation of the *Drosophila* brown gene: Evidence for transcriptional repression and somatic pairing dependence. Proc. Natl. Acad. Sci. USA *86*: 6704-6708.

Hennig KA, Novotny EA, Compton ST, Guan XY, Liu PP, and Ashlock MA (1999). Human artificial chromosomes generated by modification of a yeast artificial chromosome containing both human alpha satellite and single-copy DNA sequences. Proc. Natl. Acad. Sci. USA *96(2)*: 592-597.

Hilliker AJ (1976). Genetic analysis of the centromeric heterochromatin of chromosome 2 of *Drosophila melanogaster*: deficiency mapping of EMS-induced lethal complementation groups. Genetics *83*: 765-782.

Hilliker AJ, Appels R, and Schalet A (1980). The genetic analysis of *D. melanogaster* heterochromatin. Cell *21*: 607-619.

Holland KA, Kereso J, Zákány J, Pravnovskzy T, Monostori E, Belyaer N, and Hadlaczky G (1995). A tightly bound chromosome antigen is detected by monoclonal antibodies in a ring-like structure on human centromeres. Chromosoma *103*: 559-566.

Holmquist GP (1987). Role of replication time in the control of tissue-specific gene expression. Am. J. Hum. Genet. *40*: 151-173.

Homberger HP (1989). Bent DNA is a structural feature of scaffold-attached regions in *D. melanogaster* interphase nuclei. Chromosoma *98*:99-104.

Hsieh TS, and Brutlag DL (1979a). A protein that preferentially binds *Drosophila* satellite DNA. Proc. Natl. Acad. Sci. *76*: 726-730.

Hsieh TS, and Brutlag DL (1979b). Sequence and sequence variation within the 1.688 g/cm^3 satellite DNA of *Drosophila melanogaster*. J. Mol. Biol. *135*: 465-481.

Ikeno M, Grimes B, Okazaki T, Nakano M, Saitoh K, Hoshino H, McGill NI, Cooke H, and Masumoto H (1998). Construction of YAC-based mammalian artificial chromosomes. Nat. Biotech. *16*: 431-439.

Jabs EW, Goble CA, and Cutting GR (1989). Macromolecular organization of human centromeric regions reveals high-frequency, polymorphic macro DNA repeats. Proc. Natl. Acad. Sci. USA 86: 202-206.

James TC, Eissenberg JC, Craig C, Dietrich V, Hobson A, and Elgin SC (1989). Distribution patterns of HP1, a heterochromatin associated nonhistone chromosomal protein of Drosophila. Eur. J. Cell Biol. *50*: 170-180.

Jamrich M, Warrior R, Steele R, and Gall JG (1983). Transcription of repetitive sequences on Xenopus lampbrush chromosomes. Proc. Natl. Acad. Sci. USA *80*: 3364-3367.

Jeppesen P, Mitchell A, Turner B, and Perry P (1992). Antibodies to defined histone epitopes reveal variations in chromatin conformation and underacetylation of centric heterochromatin in human metaphase chromosomes. Chromosoma *101*: 322-332.

Jeppesen P, and Turner BM (1993). The inactive X chromosome in female mammals is distinguished by a lack of histone H4 acetylation, a cytogenetic marker for gene expression. Cell *74*: 281-289.

Jurka J (1997). Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. Proc. Natl. Acad. Sci. USA *94*: 1872-1877.

Jurka J, Zietkiewicz E, and Labuda D (1995). Ubiquitous mammalian-wide interspersed repeats (MIRs) are molecular fossils from the mesozoic era. Nucl. Acids Res. *23*: 170-175.

Kapoor M, Montes de Oca Luna R, Liu G, Lozano G, Cummings C, Mancini M, Ouspenski I, Brinkley BR, and May GS (1998). The *cenpB* gene is not essential in mice. Chromosoma *107(8)*: 570-576.

Karin M, and Richards RI (1982). Human metallothionein genes: molecular cloning and sequence analysis of the mRNA. Nucl. Acids Res. *10(10)*: 3165-3173.

Karpen GH, and Allshire RC (1997). The case for epigenetic effects on centromere identity and function. Trends Genet. *13*: 489-496.

Kato S, Anderson RA, and Camerini-Otero RD (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.

Kedes LH (1979). Histone genes and histone messengers. Ann. Rev. Biochem. *48*: 837-870.

Kellum R, and Schedl P (1991). A position-effect assay for boundaries of higher order chromosomal domains. Cell *64*: 941-950.

Kellum R, Raff JW, and Alberts BM (1995). Heterochromatin protein 1 distribution during development and during the cell cycle in Drosophila embryos. J. Cell Sci. *108*: 1407-1418.

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

Kitagawa K, Masumoto H, Ikeda M, and Okazaki T (1995). Analysis of protein-DNA and protein-protein interactions of centromere protein B (CENP-B) and properties of the DNA-CENP-B complex in the cell cycle. Mol. Cell. Biol. *15*: 1602-1612.

Kiyama R, Matsui H, Okumura K, and Oishi M (1987). A group of repetitive human DNA families that is characterized by extrachromosomal oligomers and restriction-fragment length polymorphism. J. Mol. Biol. *193*: 591-597.

Koo HS, and Crothers DM (1988). Calibration of DNA curvature and a unified description of sequence-directed bending. Proc. Natl. Acad. Sci. USA *85*: 1763-1767.

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

La Volpe A, Ciaramella M, and Bazzicalupo P (1988). Structure, evolution and properties of a novel repetitive DNA family in *Caenorhabditis elegans*. Nucleic Acids Res. *1*: 8213-8231.

Lica LM, Narayanswami S, and Hamkalo BA (1986). Mouse satellite DNA, centromere structure, and sister chromatid pairing. J. Cell Biol. *103*: 1145-1151.

Lima-de-Faria A (1969). In *Handbook of Molecular Cytology*. Lima-de-Faria (ed.). American Elsevier, New York, p227.

Lindsley DL, Edington CW, and Von Halle ES (1960). Sex-linked recessive lethals in Drosophila whose expression is suppressed by the Y chromosome. Genetics *45*: 1649-1670.

Littlefield JW (1964). Selection of hybrids from mating of fibroblasts *in vitro* and their presumed recombinants. Science *145*: 709-710.

Lo CW (1983). Transformation by iontophoretic microinjection of DNA: multiple integrations without tandem insertions. Mol. Cell. Biol. *3*: 1803-1814.

Lombillo VA, Nislow C, Yen TJ, Gelfand VI, and McIntosh JR (1995). Antibodies to the kinesin motor domain and CENP-E inhibit microtubule depolymerization-dependent motion of chromosomes *in vitro*. J. Cell Biol. *128*: 107-115.

Long EO, and Dawid IB (1980). Repeated genes in eukaryotes. Ann. Rev. Biochem. *49*: 727-764.

Lund T, Hotlund J, Fredrickson M, and Laland SG (1983). On the presence of two new mobility group-like proteins in HeLa S3 cells. FEBS Lett. *152*: 163-167.

Luzi E, Eckstein F, and Barsacchi G (1997). The newt ribozyme is part of a riboprotein complex. Proc. Natl. Acad. Sci. USA *94*: 9711-9716.

Maiorano D, Cece R, and Badaracco G (1997). Satellite DNA from the brine shrimp *Artemia* affects the expression of a flanking gene in yeast. Gene *189*: 13-18.

Marahrens Y, and Stillman B (1992). A yeast chromosomal origin of DNA replication defined by multiple functional elements. Science *255*: 817-823.

Maraia RJ, and Sarrowa J (1995). In *The impact of short interspersed elements (SINEs) on the host genome*. Maraia R (ed.). Springer, New York, NY, pp 163-196.

Martin SL (1991). Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. Mol. Cell. Biol. *11*: 4804-4807.

Martin-Morris LM, Csink AK, Dorer DR, Talbert PB, and Henikoff S (1997). Heterochromatic trans-inactivation of Drosophila white transgenes. Genetics *147*: 671-677.

Martinez-Balbas A, Rodriguez-Campos A, Garcia-Ramirez M, Sainz J, Carrera P, Aymami J, and Azorin F (1990). Satellite DNAs contain sequences that induce curvature. Biochem. *29*: 2342-2348.

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

Masumoto H, Sugimoto K, and Okazaki T (1989b). Alphoid satellite DNA is tightly associated with centromere antigens in human chromosomes throughout the cell cycle. Exp. Cell Res. *181*: 181-196.

Mathias SL, Scott AF, Kazazian HH Jr., Boeke JD, and Gabriel A (1991). Reverse transcriptase encoded by a human transposable element. Science *254*: 1808-1810.

McGhee JD, and Felsenfeld G (1980). Nucleosome structure. Annu. Rev. Biochem. *49*: 1115-1156.

McGhee JD, Wood WI, Dolan M, Engel JD, and Felsenfeld G (1981). A 200 base pair region at the 5' end of the chicken adult beta-globin gene is accessible to nuclease digestion. Cell *27*: 45-55.

Meselson M, Stahl FW, and Vinograd J (1957). Equilibrium sedimentation of macromolecules in density gradients. Proc. Natl. Acad. Sci. USA *43*: 581-588.

Mestrovic N, Plohl M, Mravinac B, and Ugarkovic D (1998). Evolution of satellite DNAs from the genus Palorus--experimental evidence for the 'library' hypothesis. Mol. Biol. Evol. *15(8)*: 1062-1068.

Miklos GLG (1985). Localized, highly repetitive DNA sequences in vertebrate and invertebrate genomes. In *Molecular Evolutionary Genetics*. RJ MacIntyre (ed.). Plenum Press, New York, pp 241-313.

Miller OJ, Schnedl W, Allen J, and Erlanger BF (1974). 5-Methylcytosine localised in mammalian constitutive heterochromatin. Nature *251*: 636-637.

Milman G, and Herzberg M (1981). Efficient DNA transfection and rapid assay for thymidine kinase activity and viral antigenic determinants. Som. Cell. Genet. *7*: 161-170.

Miniou P, Jeanpierre M, Blanquet V, Sibella V, Bonneau D, Herbelin C, Fischer A, Niveleau A, and Viegas-Péquignot (1994). Abnormal methylation pattern in constitutive and facultative (X inactive chromosome) heterochromatin of ICF patients. Hum. Mol. Genet. *3*: 2093-2102.

Miyazaki WY, and Orr-Weaver TL (1994). Sister-chromatid cohesion in mitosis and meiosis. Annu. Rev. Genet. *28*: 167-187.

Moritz KB, and Roth GE (1976). Complexity of germline and somatic DNA in *Ascaris*. Nature *259*: 55-57.

Moroi Y, Peebles C, Fritzler MJ, Steigerwald J, and Tan EM (1980). Autoantibody to centromere (kinetochore) in scleroderma sera. Proc. Natl. Acad. Sci. USA *77*: 1627-1631.

Muller HJ (1930). Types of visible variations induced by X-rays in Drosophila. J. Genet. *22*: 299-334.

Muro Y, Masumoto H, Yoda K, Nozaki N, Ohashi M, and Okazaki T (1992). Centromere protein B assembles human centromeric alpha-satellite DNA at the 17-bp sequence, CENP-B box. J. Cell Biol. *116*: 585-596.

Murphy TD, and Karpen GH (1995). Localization of centromere function in a Drosophila minichromosome. Cell *82*: 599-609.

Murphy TD, and Karpen GH (1998). Centromeres take flight: alpha satellite and the quest for the human centromere. Cell *93*: 317-320.

Murray AW, and Szostak JW (1983). Construction of artificial chromosomes in yeast. Nature *305*: 189-193.

Nickerson JA, Krochmalnic A, Wan KM, and Penman S (1989). Chromatin architecture and nuclear RNA. Proc. Natl. Acad. Sci. USA *86*: 177-181.

Nowalk R (1994). Mining treasures from junk DNA. Science *263*: 608-610.

Ohshima K, Hamada M, Terai H, and Okada N (1996). The 3' ends of tRNA-derived short interspersed repetitive elements are derived from the 3' ends of long interspersed repetitive elements. Mol. Cell. Biol. *16*: 3756-3764.

Ohshima K, Koishi R, Matsuo M, and Okada N (1993). Several short interspersed repetitive elements (SINEs) in distant species may have originated from a common ancestral retrovirus: characterization of a squid SINE and a possible mechanism for generation of tRNA-derived retroposons. Proc. Natl. Acad. Sci. USA *90*: 6260-6264.

Okada N (1991). SINEs: short interspersed repeated elements of the eukaryotic genome. Trends Ecol. Evol. *6*: 358-361.

Okada N, Hamada M, Ogiwara I, and Ohshima K (1997). SINEs and LINEs share common 3' sequences: a review. Gene *205*: 229-243.

O'Neill TE, Smith JG, and Bradbury EM (1993). Histone octamer dissociation is not required for transcript elongation through arrays of nucleosome cores by phage T7 RNA polymerase *in vitro*. Proc. Natl. Acad. Sci. USA *90*: 6203-6207.

O'Neill LP, and Turner BM (1995). Histone H4 acetylation distinguishes coding regions of the human genome from heterochromatin in a differentiation-dependent but transcription-independent manner. EMBO J. *14*: 3946-3957.

Orgel LE, and Crick FHC (1980). Selfish DNA: the ultimate parasite. Nature *284*: 604-607.

Pableo EC, and Triantaphyllou AC (1989). DNA complexity in the root-knot nematode (*Meloidogyne* spp.) genome. J. Nematol. *21*: 260-263.

Page SL, Earnshaw WC, Choo KHA, and Shaffer LG (1995). Further evidence that CENP-C is a necessary component of active centromeres: studies of a dic(X;15) with simultaneous immunofluorescence and FISH. Hum. Mol. Genet. *4*: 289-294.

Peacock WJ, Lohe AR, Gerlach WL, Dunsmuir P, Dennis ES, and Appels R (1978). Fine structure and evolution of DNA in heterochromatin. Cold Spring Harbor Symp. Quant. Biol. *42*: 1121-1135.

Pech M, Igo-Kemenes T, and Zachau HG (1979a). Nucleotide sequence of a highly repetitive component of rat DNA. Nucl. Acids Res. *7*: 417-432.

Pech M, Streeck RE, and Zachau HG (1979b). Patchwork structure of a bovine satellite DNA. Cell *18*: 883-893.

Pellicer A, Robins D, Wold B, Sweet R, Jackson J, Lowy I, Roberts JM, Sim GK, Silverstein S, and Axel R (1980). Altering genotype and phenotype by DNA mediated gene transfer. Science *209*: 1414.

Perucho M, Hanahan D, and Wigler M (1980). Genetic and physical linkage of exogenous sequences in transformed cells. Cell *22*: 309-317.

Perucho M, and Wigler M (1981). Linkage and expression of foreign DNA in cultured animal cells. Cold Spring Harbor Symp. Quant. Biol. *45*: 829-838.

Pietras DF, Bennett KL, Siracusa LD, Woodworth-Gutai M, Chapman VM, Gross KW, Kane-Haas C, and Hastie ND (1983). Construction of a small *Mus musculus* repetitive DNA library; identification of a new satellite sequence in *Mus musculus*. Nucl. Acids Res. *11*: 6969-6983.

Piotte C, Castagnone-Sereno P, Bongiovanni M, Dalmasso A, and Abad P (1994). Cloning and characterization of two satellite DNAs in the low-C-value genome of the nematode *Meloidogyne* spp. Gene *138*: 175-180.

Plohl M, and Ugarkovic D (1994). Characterization of two abundant satellite DNAs from the mealworm *Tenebrio obscurus*. J. Mol. Evol. *39*: 489-495.

Pons J, Petitpierre E, and Juan C (1993). Characterization of the heterochromatin of the darkling beetle *Misolampus goudoti*: cloning of two satellite DNA families and digestion of chromosomes with restriction enzymes. Hereditas *119*: 179-185.

Qu G, Dubeau L, Narayan A, Yu MC, and Ehrlich M (1999a). Satellite DNA hypomethylation vs. overall genomic hypomethylation in ovarian epithelial tumors of different malignant potential. Mut. Res. *423(1-2)*: 91-101.

Qu GZ, Grundy PE, Narayan A, and Ehrlich M (1999b). Frequent hypomethylation in Wilms tumors of pericentromeric DNA in chromosomes 1 and 16. Cancer Genet. Cytogenet. *109(1)*: 34-39.

Radic MZ, Lundgren K, and Hamkalo BA (1987). Curvature of mouse satellite DNA and condensation of heterochromatin. Cell *50*: 1101-1108.

Renault S, Rouleux-Bonnin F, Periquet G, and Bigot Y (1999). Satellite DNA transcription in Diadromus pulchellus (Hymenoptera). Insect Biochem. Mol. Biol. *29(2)*: 103-111.

Reuter G, Gausz J, Gyurkovics H, Friede B, Bang R, Spierer A, Hall LM, and Spierer P (1987). Modifiers of position-effect variegation in the region from 86C to 88B of the *Drosophila melanogaster* third chromosome. Mol. Gen. Genet. *210*: 429-436.

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.

Rhim JS, Cho HY, and Huebner RJ (1975). Non-producer human cells induced by murine sarcoma virus. Int. J. Cancer *15*: 23-29.

Roizes GP, and Pages M (1982). Conserved and divergent sequences of bovine satellite DNAs. In *Genome evolution*. Dover GA, Flavell RB (eds). Academic Press, San Diego, pp 95-111.

Royle NJ, Clarkson RE, Wong Z, and Jeffreys AJ (1988). Clustering of hypervariable minisatellites in the proterminal regions of human autosomes. Genomics *3*: 352-360.

Russell LB, and Cacheiro NL (1978). The use of X-autosome translocations in the study of X-inactivation pathways and nonrandomness. Basic Life Sci. *12*: 393-416.

Russell LB, and Montgomery CS (1970). Comparative studies on X-autosome translocations in the mouse. II. Inactivation of autosomal loci, segregation, and mapping of autosomal breakpoints in five T (X;1) S. Genetics *64*: 281-312.

Saint-Dic D, and DuBow MS (1995). Effect of cis-located human satellite DNA on electroporation efficiency. Methods Mol. Biol. *48*: 199-210.

Saitoh H, Tomkiel J, Cooke CA, Ratrie H, Maurer M, Rothfield NF, and Earnshaw WC (1992). CENP-C, an autoantigen in Scleroderma, is a component of the human inner kinetochore plate. Cell *70*: 115-125.

Sambrook J, Fritsch EF, and Maniatis T (1989). Molecular cloning: a laboratory manual. Second edition. Cold Spring Harbor Press, Cold Spring Harbor, New York.

Scangos G, Huttner KM, Juricek DK, and Ruddle FH (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.

Schildkraut CL, Marmur J, and Doty P (1962). Determination of the base composition of deoxyribonucleic acid from its buoyant density in CsCl. J. Mol. Biol. *4*: 430-443.

Schlessinger D, and Nagaraja R (1998). Impact and implications of yeast and human artificial chromosomes. Ann. Med. *30*: 186-191.

Schmid CW (1996). Alu: structure, origin, evolution, significance and function of one-tenth of human DNA. Prog. Nucleid Acid Res. Mol. Biol. *53*: 283-319.

Schnedl W, Dev VG, Tantravahi R, Miller DA, Erlanger BF, and Miller OJ (1975). 5-Methylcytosine in heterochromatic regions of chromosomes: chimpanzee and gorilla compared to the human. Chromosoma *52*: 59-66.

Shampay J, Szostak JW, and Blackburn EH (1984). DNA sequences of telomeres maintained in yeast. Nature *310*: 154-157.

Singer MF (1982). SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes. Cell *28*: 433-434.

Smit AF (1996). The origin of interspersed repeats in the human genome. Curr. Opin. Genet. Dev. *6*: 743-748.

Smith GP (1976). Evolution of repeated DNA sequences by unequal crossover. Science *191*: 528-535.

Sol K, and DuBow MS (1992). Structure and organization of a human EcoRI satellite II DNA family. Mol. Biol. *11*: 159-168.

Sol K, Lapointe M, MacLeod M, Nadeau C, and DuBow MS (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 *Kpn* I fragments. Biochim. Biophys. Acta *868*: 128-135.

Southern EM (1970). Base sequence and evolution of guinea-pig α-satellite DNA. Nature *227*: 794-798.

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

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

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

Spofford JB (1976). Position-effect variegation in *Drosophila*. In *Genetics and Biology of Drosophila*. Ashburner M and Novitski E (eds). Academic Press, London, pp 955-1019.

Subramanian K (1979). Segments of Simian virus 40 DNA spanning most of the leader sequence of the major late viral messenger RNA are dispensable. Proc. Natl. Acad. Sci. USA *76*: 2556-2560.

Sugimoto K, Hagishita Y, and Himeno M (1994). Functional domain structure of human centromere protein B. J. Biol. Chem. *269*: 24271-24276.

Sullivan KF, and Glass CA (1991). CENP-B is a highly conserved mammalian centromere protein with homology to the helix-loop-helix family of proteins. Chromosoma *100*: 360-370.

Sullivan B, and Schwartz S (1995). Identification of centromeric antigens in dicentric Robertsonian translocations: CENP-C and CENP-E are necessary components of functional centromeres. Hum. Mol. Genet. *4*: 2189-2197.

Talarico D, Peverali AF, 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.

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

Tautz D (1993). Notes on the definition and nomenclature of tandemly repetitive DNA sequences. EXS *67*: 21-28.

Tazi J, and Bird AP (1990). Alternative chromatin structure at CpG islands. Cell *60*: 909-920.

Telenius H, Szeles A, Kereso J, Csonka E, Praznovszky T, Imreh S, Maxwell A, Perez CF, Drayer JI, and Hadlaczky G (1999). Stability of a functional murine satellite DNA-based artificial chromosome across mammalian species. Chromo. Res. *7(1)*: 3-7.

Thrower DA, Jordan MA, Schaar BT, Yen TJ, and Wilson L (1995). Mitotic HeLa cells contain a CENP-E associated minus end-directed microtubule motor. Embo J. *14*: 918-926.

Tolchkov EV, Kramerova IA, Lavrov SA, Rasheva V, Bonaccorsi S, Alatortsev VE, and Gvozdev VA (1997). Position-effect variegation in *Drosophila melanogaster X* chromosome inversion with a breakpoint in a satellite block and its suppression in a secondary rearrangement. Chromosoma *106*: 520-525.

Tomkiel J, Cooke CA, Saitoh H, Bernat RL, and Earnshaw WC (1994). CENP-C is required for maintaining proper kinetochore size and for a timely transition to anaphase. J. Cell Biol. *125*: 531-545.

Toneguzzo F, Hayday AC, 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.

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

Vance GH, Curtis CA, Heerema NA, Schwartz S, and Palmer CG (1997). An apparently acentric marker chromosome originating from 9p with a functional centromere without detectable alpha and beta satellite sequences. Am. J. Med. Genet. *71*: 436-442.

Varley JM, Macgregor HC, Nardi I, Andrews C, and Erba HP (1980). Cytological evidence of transcription of highly repeated DNA sequences during the lampbrush stage in Triturus cristatus carnifex. Chromosoma *80*: 289-307.

Vig BK, and Willcourt M (1998). Decondensation of pericentric heterochromatin alters the sequence of centromere separation in mouse cells. Chromosoma *107*: 417-423.

Vogt P (1992). Code domains in tandem repetitive DNA sequence structures. Chromosoma *101*: 585-589.

Voullaire LE, Slater HR, Petrovic V, and Choo KH (1993). A functional marker centromere with no detectable alpha-satellite, satellite III, or CENP-B protein: activation of a latent centromere? Am. J. Hum. Genet. *52*: 1153-1163.

Wakimoto BT (1998). Beyond the nucleosome: epigenetic aspects of position-effect variegation in *Drosophila*. Cell *93*: 321-324.

Walker PMB (1971). Repetitive DNA in higher organisms. Prog. Biophys. Mol. Biol. *23*: 147-190.

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

Warburton PE, Waye JS, and Willard HF (1993). Nonrandom localization of recombination events in human alpha satellite repeat unit variants: implications for higher-order structural characteristics within centromeric heterochromatin. Mol. Cell. Biol. *13*: 6520-6529.

Waring M, and Britten RJ (1966). Nucleotide sequence repetition: a rapidly reassociating fraction of mouse DNA. Science *154*: 791-794.

Weiler KS, and Wakimoto BT (1995). Heterochromatin and gene expression in Drosophila. Ann. Rev. Genetics *29*: 577-605.

Weiner AM (1980). An abundant cytoplasmic 7S RNA is complementary to the dominant interspersed middle repetitive DNA sequence family in the human genome. Cell *22*: 209-218.

Wilson AC (1976). In *Molecular Evolution*. Ayala FJ (ed.). Sinauer, Sunderland, pp 225-236.

Wilson JH, Berget PB, and Pipas JM (1982). Somatic cells efficiently join unrelated DNA segments end to end. Mol. Cell Biol. *2*: 1258-1269.

Willard HF (1990). Centromeres of mammalian chromosomes. Trends Genet. *6*: 410-416.

Willard HF (1998a). Centromeres: the missing link in the development of human artificial chromosomes. Curr. Opin. Genet. Dev. *8*: 219-225.

Willard HF (1998b). Human artificial chromosomes coming into focus. Nat. Biotech. *16*: 415-416.

Williams BC, Murphy TD, Goldberg ML, and Karpen GH (1998). Neocentromere activity of structurally acentric mini-chromosomes in Drosophila. Nat. Genet. *18(1)*: 30-37.

Wolffe AP (1992). *Chromatin: Structure and Function.* Academic Press.

Yen TY, Schaar BT, Szilak I, and Cleveland DW (1992). CENP-E is a putative kinetochore motor that accumulates just before mitosis. Nature *359*: 536-539.

Yoda K, Ando S, Okuda A, Kikuchi A, and Okazaki T (1998). In vitro assembly of the CENP-B/alpha-satellite DNA/core histone complex: CENP-B causes nucleosome positioning. Genes Cells *3(8)*: 533-548.

Yoda K, Kitagawa K, Masumoto H, Muro Y and Okazaki T (1992). A human centromere protein, CENP-B, has a DNA binding domain containing four potential α helices at the NH_2 terminus, which is separable from dimerizing activity. J. Cell Biol. *119*: 1413-1427.

Yunis JJ, and Yasmineh WG (1971). Heterochromatin, satellite DNA and cell function. Science *174*:1200-1209.

Zimmer EA, Martin SL, Beverly SM, Kan YW, and Wilson AC (1980). Rapid duplication and loss of genes coding for the α chains of hemoglobin. Proc. Natl. Acad. Sci. USA *77*: 2158-2162.

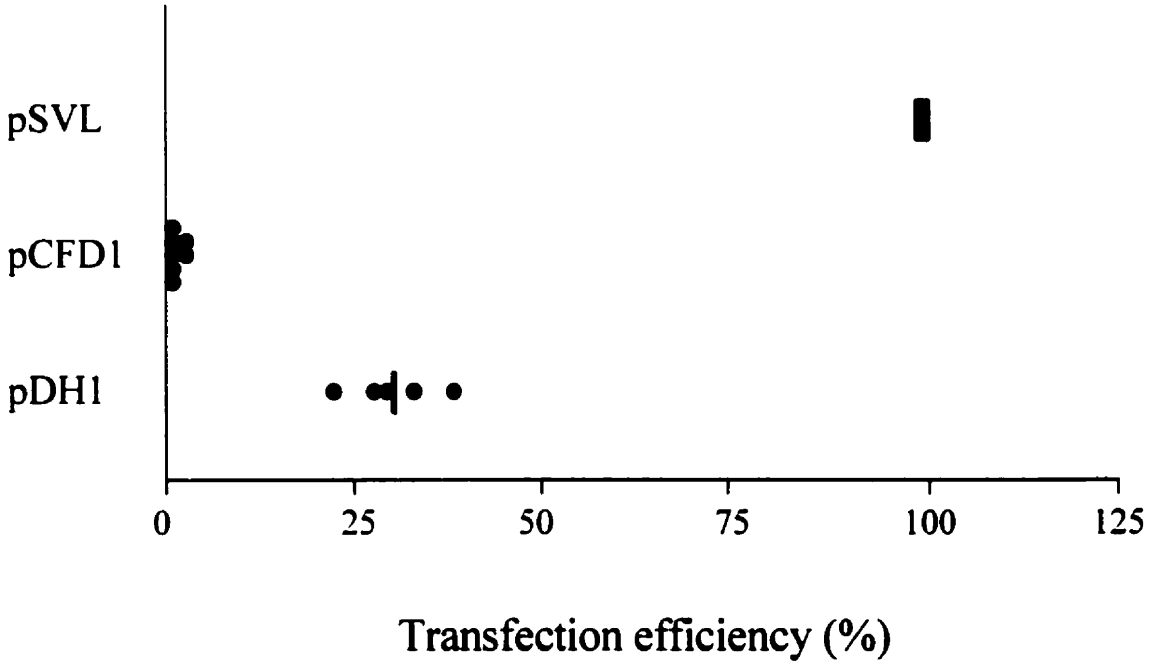
Zlatanova J, and Van Holde K (1992). Histone H1 and transciption: still an enigma? J. Cell. Sci. *103*: 889-895.

Zuckerland E, and Hennig W (1995). Tracking heterochromatin. Chromosoma *104*: 75-83.

APPENDIX

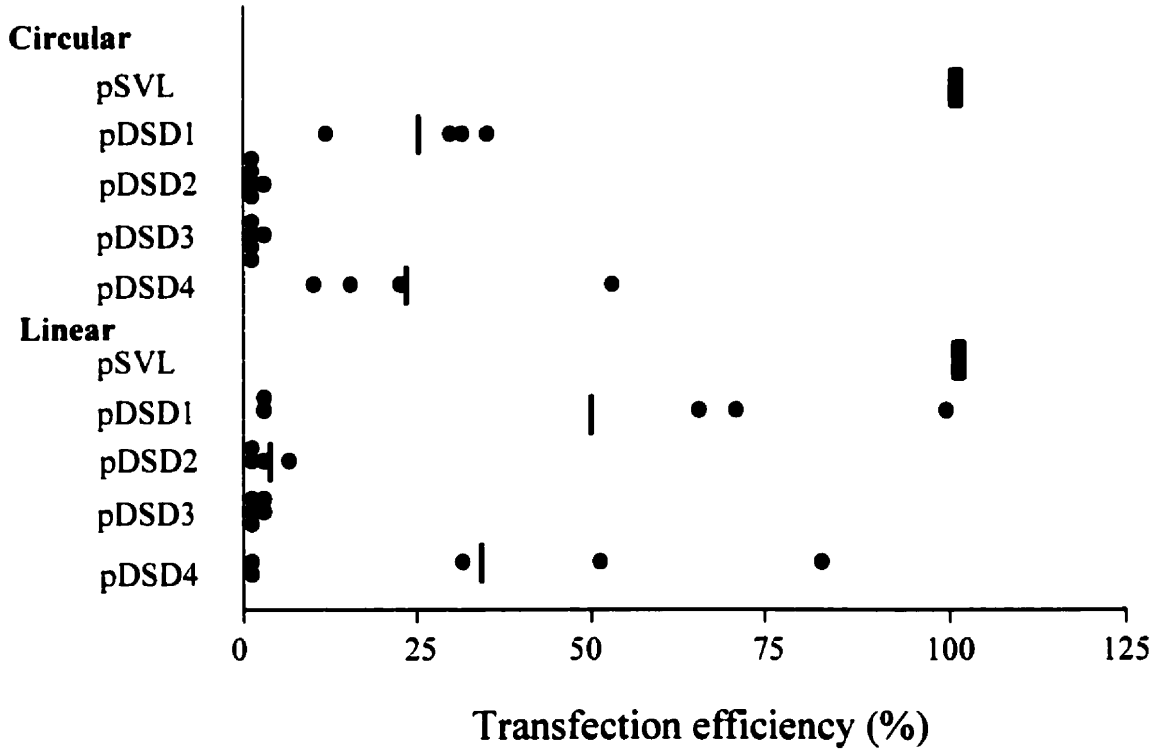
Appendix 1

Frequency of transfection using plasmids containing the 1.797 kb human satellite II DNA sequence at the *Stu*I (pCFD1) or *Hpa*I (pDH1) sites, in media containing HAT and G418. Plasmid pSVL is the plasmid vector used as a reference (100%). The vertical line (|) indicates the average.


pSVL
pCFD1
pDH1
0
25
50
75
100
125
Transfection efficiency (%)

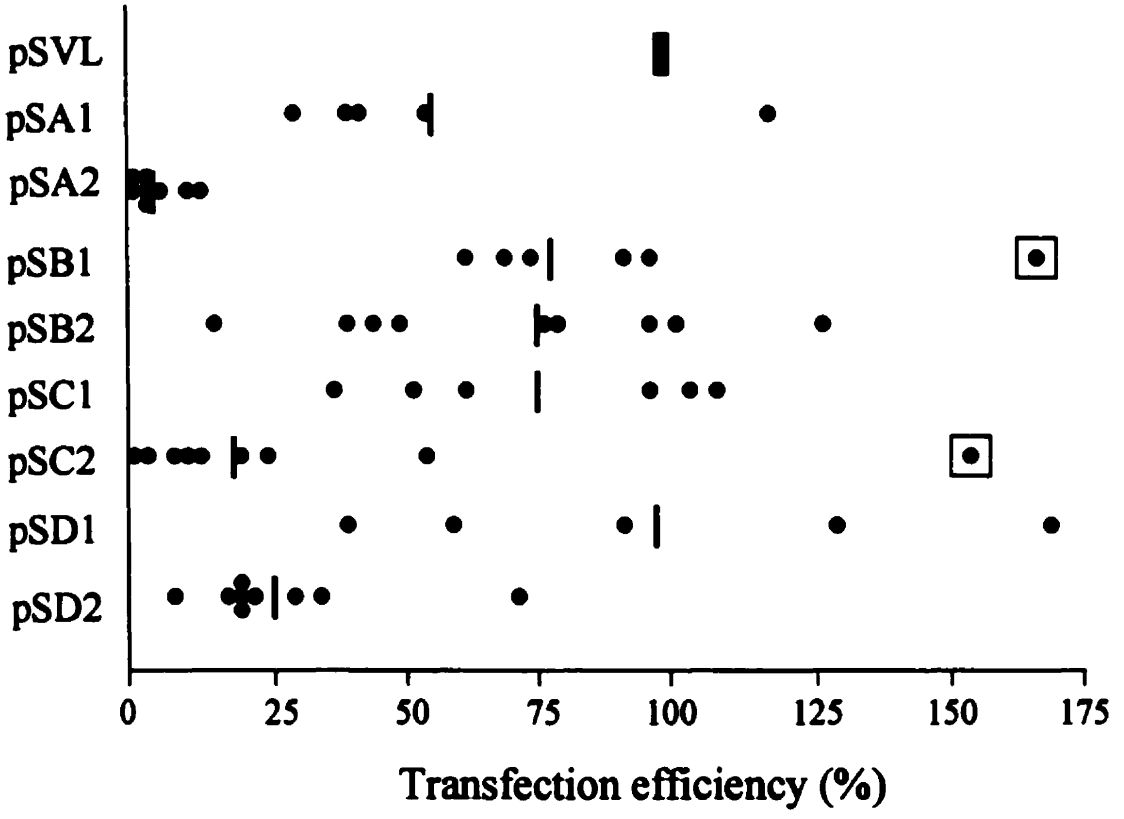
Appendix 2

Frequency of transfection using plasmids containing fragments of 813 bp (pDSD1 and pDSD2) and 983 bp (pDSD3 and pDSD4) from the 1.797 kb human satellite II DNA sequence in media containing HAT and G418. Circular and linear conformations of the plasmids were used. Plasmid pSVL is the plasmid vector, used as a reference (100%). The vertical line (|) indicates the average.


Circular
pSVL
pDSD1
pDSD2
pDSD3
pDSD4
Linear
pSVL
pDSD1
pDSD2
pDSD3
pDSD4
0
25
50
75
100
125
Transfection efficiency (%)

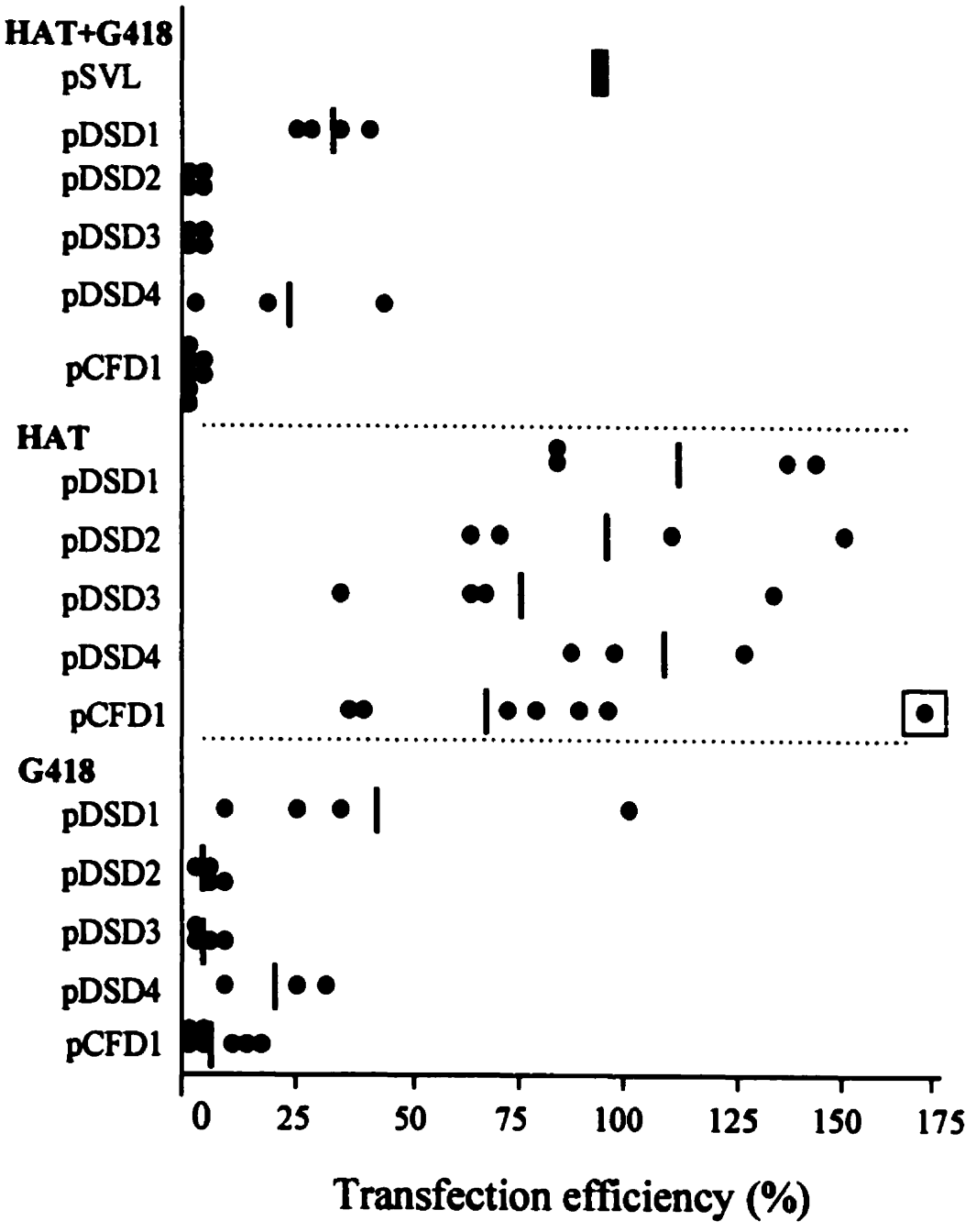
Appendix 3

Frequency of transfection using plasmids containing fragments of 354 bp (pSD1, pSD2), 458 bp (pSC1, pSC2), 364 bp (pSB1, pSB2), and 620 bp (pSA1, pSA2) from the 1.797 kb human satellite II DNA sequence in media containing HAT and G418. Plasmid pSVL is the plasmid vector, used as a reference (100%). The vertical line (|) indicates the average. The boxed values were excluded as atypical. Averages exclude the boxed values.


pSVL
pSA1
pSA2
pSB1
pSB2
pSC1
pSC2
pSD1
pSD2
0
25
50
75
100
125
150
175
Transfection efficiency (%)

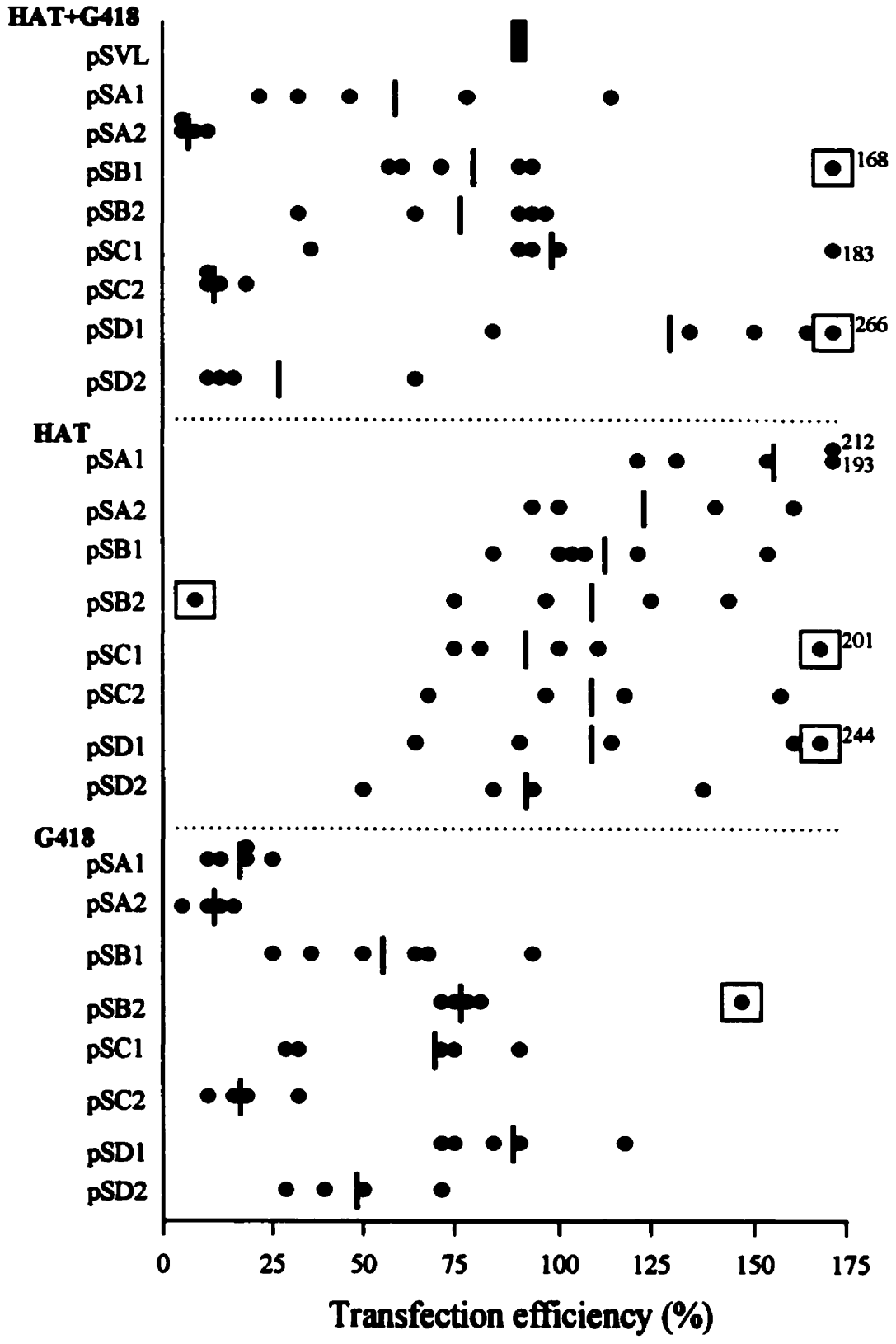
Appendix 4

Frequency of transfection using plasmid pCFD1 containing the 1.797 kb human satellite II DNA sequence or plasmids containing fragments of 813 bp (pDSD1, pDSD2) and 983 bp (pDSD3, pDSD4) from the satellite II DNA sequence, in media containing HAT plus G418, HAT, or G418. Plasmid pSVL is the plasmid vector, used as a reference (100%). The vertical line (|) indicates the average. The boxed values were excluded as atypical. Averages exclude the boxed values.


HAT+G418
pSVL
pDSD1
pDSD2
pDSD3
pDSD4
pCFD1
HAT
pDSD1
pDSD2
pDSD3
pDSD4
pCFD1
G418
pDSD1
pDSD2
pDSD3
pDSD4
pCFD1
0
25
50
75
100
125
150
175
Transfection efficiency (%)

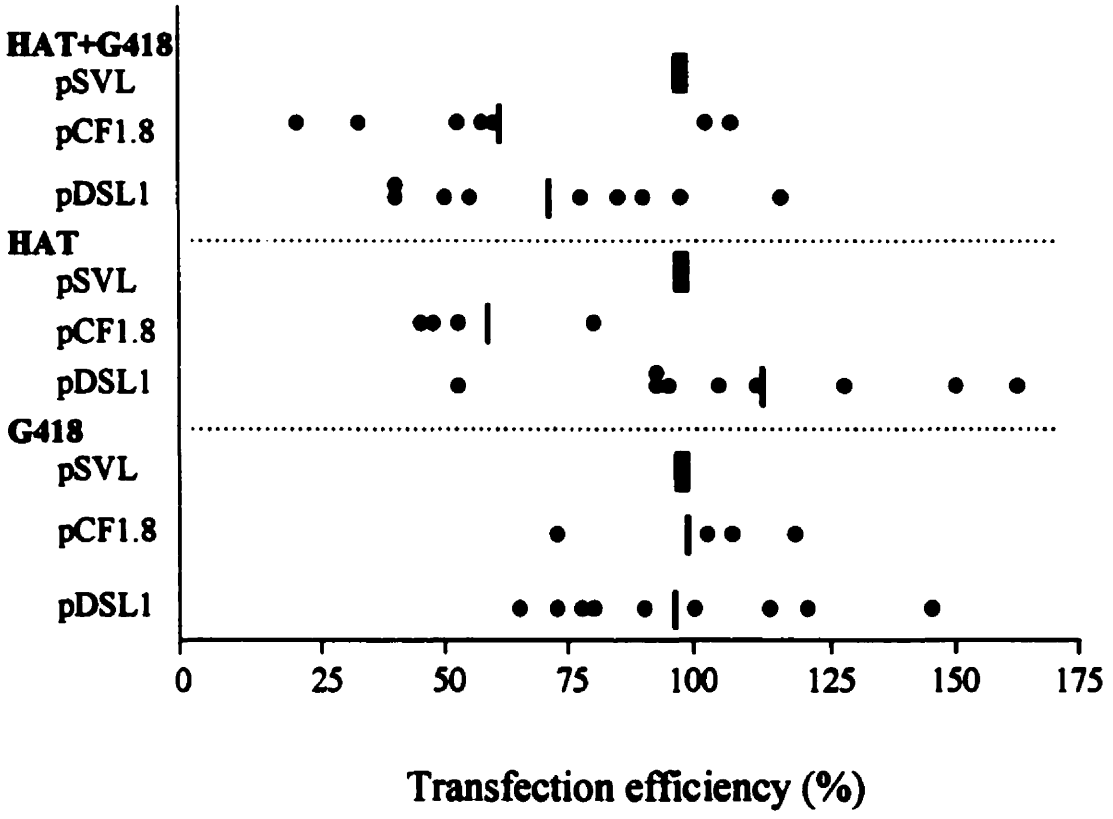
Appendix 5

Frequency of transfection using plasmids containing fragments of 354 bp (pSD1, pSD2), 458 bp (pSC1, pSC2), 364 bp (pSB1, pSB2), and 620 bp (pSA1, pSA2) from the 1.797 kb human satellite II DNA sequence in media containing HAT plus G418, HAT, or G418. Plasmid pSVL is the plasmid vector, used as a reference (100%). The vertical line (|) indicates the average. The boxed values were excluded as atypical. Averages exclude the boxed values.


HAT+G418
pSVL
pSA1
pSA2
pSB1
pSB2
pSC1
pSC2
pSD1
pSD2
168
183
266
HAT
pSA1
pSA2
pSB1
pSB2
pSC1
pSC2
pSD1
pSD2
212
193
201
244
G418
pSA1
pSA2
pSB1
pSB2
pSC1
pSC2
pSD1
pSD2
0
25
50
75
100
125
150
175
Transfection efficiency (%)

Appendix 6

Frequency of transfection using control plasmids containing fragments of 1.8 kb (pCF1.8) or 604 bp (pDSL1) from bacteriophage λ DNA in media containing HAT+G418, HAT, or G418. Plasmid pSVL is the plasmid vector, used as a reference (100%). The vertical line (|) indicates the average.


HAT+G418
pSVL
pCF1.8
pDSL1
HAT
pSVL
pCF1.8
pDSL1
G418
pSVL
pCF1.8
pDSL1
0
25
50
75
100
125
150
175
Transfection efficiency (%)