

# A TECHNOLOGY FOR DETECTING UNSELECTED MUTATIONAL SPECTRA IN HUMAN GENOMIC DNA

by

Xiaocheng Li-Sucholeiki

B.S., Genetics  
Fudan University, 1990

M.S., Cell and Molecular Biology  
Shanghai Institute of Cell Biology, Chinese Academy of Science, 1993

Submitted to the Division of Bioengineering and Environmental Health  
in Partial Fulfillment of the Requirement for the Degree of

Doctor of Philosophy in Toxicology

at the

Massachusetts Institute of Technology

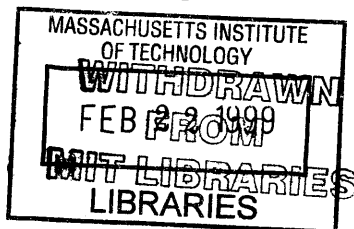
February 1999

© 1999 Massachusetts Institute of Technology  
All right reserved

Signature of Author \_\_\_\_\_  
Division of Bioengineering and Environmental Health

Certified by \_\_\_\_\_  
Professor William G. Thilly  
Thesis Advisor

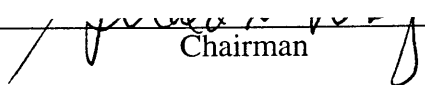
Accepted by \_\_\_\_\_  
Professor Peter C. Dedon  
Chairman, Division Committee on Graduate Studies



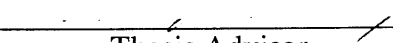
Science

This doctor thesis has been examined by a Committee of the Division of Bioengineering and Environmental Health as follows:

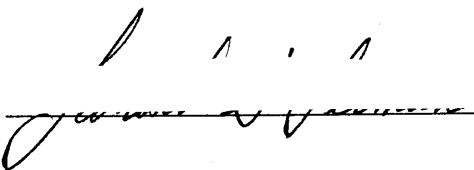
Professor Gerald N. Wogan

*A*  
  
Chairman


Professor William G. Thilly

  
Thesis Advisor


Professor Leonard Lerman



Professor David Schauer



Professor Graham C. Walker



# A TECHNOLOGY FOR DETECTING UNSELECTED MUTATIONAL SPECTRA IN HUMAN GENOMIC DNA

by

Xiaocheng Li-Sucholeiki

Submitted to the Division of Bioengineering and Environmental Health  
in Partial Fulfillment of the Requirement for the Degree of  
Doctor of Philosophy in Toxicology

## ABSTRACT

Most human inherited diseases and cancers are known to be caused by mutations in nuclear genes. But the primary causes of many of these mutations are still unknown. Determination of mutational spectra in disease-related genes in nontumorous tissues may provide direct evidence as to whether a specific mutagenic agent or pathway is involved in a particular human cancer. Such a study, however, requires the detection of a few mutant alleles from a vast excess of wild-type sequences in the presence of large quantities of cellular DNA.

In this research, a means has been developed to detect point mutational spectra in 100-bp single-copy nuclear sequences at mutant fractions equal to or greater than  $10^{-6}$ . The approach is independent of phenotypic selection and can thus be used to scan for point mutations in tissues and organs. The desired sequences are enriched  $10^4$ -fold from the genomic DNA by sequence-specific hybridization coupled with an improved biotin-streptavidin capture system. The rare mutant sequences are enriched 100-fold relative to wild-type sequences by wide-bore constant denaturant capillary electrophoresis (CDCE). The mutant-enriched sample is then amplified using a high-fidelity DNA polymerase without obscuring mutant signals by PCR noise. Amplified mutant sequences are further enriched via two rounds of CDCE coupled with high-fidelity PCR. The desired mutants which are represented as distinct peaks on an electropherogram are then isolated and sequenced.

This approach has been tested by measuring *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (MNNG) induced point mutations in a 121-bp sequence of the adenomatous polyposis coli (*APC*) gene cDNA bp 8543-8663 in human lymphoblastoid MT1 cells. Twelve MNNG-induced mutations were reproducibly observed in MNNG-treated cells at fractions between  $2 \times 10^{-6}$  and  $9 \times 10^{-6}$ . All twelve mutations were GC  $\rightarrow$  AT transitions, eleven of which occurred at guanine residues preceded (5') by a purine. Limit to the sensitivity of this approach arises from mutants created by PCR, which are represented by fourteen GC  $\rightarrow$  TA transversions observed in untreated cells at fractions around  $10^{-6}$ . The approach described herein is general for all DNA sequences suitable for CDCE analysis; such suitable sequences are found in nearly all prokaryotic and eukaryotic genes.

Thesis Advisor: William G. Thilly  
Title: Professor of Civil & Environmental Engineering and Toxicology

## ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my thesis advisor Professor Bill Thilly for his guidance during the course of my graduate career. Bill's enthusiasm, optimism and creative ideas impelled me to keep going despite the many technical challenges that confronted me. I am extremely grateful for his support and encouragement which helped me become a more self-confident and independent scientist.

I would like to acknowledge my thesis committee members, Drs. Leonard Lerman, David Schauer, Graham Walker and Gerald Wogan, for their invaluable insights, comments and suggestions on my work.

I am indebted to Dr. Konstantin Khrapko for teaching me many of the technical skills used in CDCE and his helpful suggestions. I would like to thank Dr. Gengxi Hu for teaching me the techniques of DGGE and "HERA".

I am especially grateful to Jackie Goodluck-Griffith for her indispensable help on the cell culture work; Aoy Tomita-Mitchell, Dr. Luisa Marcelino, Brindha Muniappan and Andrea Kim for their friendship, encouragement and helpful discussions; Drs. Reinhold Wasserkort, Paulo André and John Hanekamp for their help and suggestions during the course of my work.

I would like to thank other current and past members of the Thilly group, Dr. Jia Chen, Dr. Hilary Coller, Rita DeMeo, Dr. Per Olaf Ekestrom, Curtis Glover, Amanda Gruhl, Pablo Herrero, Klaudyn Hong, Paula Lee, Wen Luo, Joey Marquez, Paul Murphy, Kathy Reposo, Beth Ann Turnquist, Janice Vatland and Weiming Zheng for their friendship and support over the years.

I would like to express my special gratitude to my parents for their love, support, and faith in me. This dissertation is affectionately dedicated to them. Special thanks to my brother for his constant encouragement.

Finally, I would like to express my deep appreciation to my beloved husband who has brought so much happiness to my life. Thank you, Irving, for your love and understanding, especially during the many long nights in which you shared me with CDCE.

**TABLE OF CONTENTS**

Title Page	1
Committee Page	2
Abstract	3
Acknowledgments	4
Table of contents	5
List of figures	9
List of tables	11
List of abbreviations	12
1. Introduction	14
2. Literature Review	17
2.1. Genetic changes in humans	17
2.1.1. Germline mutations associated with inherited diseases	17
2.1.2. Somatic mutations associated with tumors: the <i>p53</i> and <i>APC</i> gene	23
2.1.3. Somatic mutations in healthy humans: the <i>hprt</i> gene	35
2.2. Current methods for detection of point mutations	41
2.2.1. Single-strand conformation polymorphism (SSCP)	41
2.2.2. Restriction fragment length polymorphism (RFLP)	44
2.2.3. Allele-specific PCR (ASP)	46
2.2.4. Blocker-PCR	47
2.2.5. Ligase chain reaction (LCR)	48
2.2.6. MutS-based assays	49
2.2.7. Denaturing gradient gel electrophoresis (DGGE) and constant denaturant gel electrophoresis (CDGE)	51
2.2.8. Constant denaturant capillary electrophoresis (CDCE)	54
2.3. MNNG mutagenesis	59
2.3.1. The MNNG-induced DNA lesions and their mutagenic potentials	59

2.3.2. The repair of MNNG-induced DNA lesions	62
2.3.3. The mutational specificity of MNNG	66
3. Materials and Methods	75
3.1. Cell culture and MNNG treatment	75
3.1.1. The human cell lines	75
3.1.2. Cell culture and MNNG treatment	77
3.1.3. Determination of the <i>hprt</i> gene mutant fractions	77
3.2. Isolation of genomic DNA from cultured cells, blood and solid tissues	78
3.3. Enrichment of the desired DNA fragments from genomic DNA	80
3.3.1. Size fractionation by polyacrylamide gel electrophoresis (PAGE)	80
3.3.2. Sequence-specific enrichment	81
3.4. High-fidelity PCR	82
3.5. Constant denaturant gel electrophoresis	83
3.6. Constant denaturant capillary electrophoresis	84
3.6.1. The CDCE instrument	84
3.6.2. Coating of capillaries	87
3.6.3. Preparation of linear polyacrylamide matrices	87
3.6.4. CDCE conditions, sample loading and fraction collection	88
3.6.4. Methods for desalting DNA samples for CE separation	89
3.7. The choice of the <i>APC</i> gene target sequence	90
3.8. Construction of artificial mutants	93
3.9. Measurement of the copy number of the target sequence and the efficiency of restriction digestion	94
4. Results and discussion	98
4.1. Optimization of the CDCE/high-fidelity PCR conditions	98
4.2. Processing of large quantities of genomic DNA from human cells and tissues	102
4.2.1. The phenol-free DNA isolation procedure	102
4.2.2. Sequential release of the desired <i>APC</i> gene fragment	

from genomic DNA	107
4.3. Evaluation of the CDGE-based approach to mutational spectrometry of the <i>APC</i> target sequence in human cells	107
4.3.1. Overview of the approach	107
4.3.2. Enrichment of the target sequence through size fractionation by PAGE	110
4.3.3. Optimization of the CDGE conditions	113
4.3.4. Pre-PCR enrichment of mutants by CDGE	116
4.3.5. High-fidelity PCR and post-PCR enrichment of mutants by CDCE	116
4.3.6. The detection limits of the approach	117
4.4. Development of the means to enrich the desired <i>APC</i> gene fragment from genomic DNA	123
4.4.1. The strategy	123
4.4.2. Optimization of the experimental conditions for the <i>APC</i> gene fragment	123
4.4.3. The enrichment efficiency and yield of the target sequence	128
4.5. Development of the means to enrich cellular mutants by wide-bore CE	130
4.5.1. The necessity and technical problems of using wide-bore capillaries in CDCE	130
4.5.2. Optimization of CDCE conditions for wide-bore capillaries	131
4.5.3. Determination of the DNA loading capacities for capillaries of various sizes	141
4.5.4. Demonstration of the detection sensitivity in reconstruction experiments	144
4.5.5. Enrichment of mutant/wild-type heteroduplexes from target-enriched genomic DNA	147
4.6. The CDCE/high-fidelity PCR approach to mutational spectrometry of the <i>APC</i> target sequence in human cells: application to MNNG-treated MT1 cells	149
4.6.1. Overview of the approach	149
4.6.2. Application to MNNG-treated and untreated MT1 cells	152
4.6.3. Identification of the mutant peaks	158
4.6.4. MNNG-induced point mutational spectrum in the <i>APC</i> target sequence	161
4.6.5. Comparison to previous MNNG mutagenesis studies	165

4.6.6. Sources for the background mutations observed in untreated cells	166
4.7. Single nucleotide polymorphisms in the <i>APC</i> target sequence in human populations	178
4.8. Conclusions	181
5. Suggestions for Future Research	183
5.1. Studies of mutagen-induced mutational spectra <i>in vitro</i>	183
5.2. Studies of mutational spectra in nuclear genes in human tissues	183
5.3. Studies of SNPs in human populations	185
6. References	186



## LIST OF FIGURES

Figure 1. Distribution of the somatic and germline point mutations in the <i>p53</i> gene in human tumors	25
Figure 2. Distribution of the somatic point mutations in the <i>p53</i> gene in colon, liver, skin, and lung cancers	28
Figure 3. Distribution of the somatic and germline mutations in the <i>APC</i> gene in human tumors	33
Figure 4. Distribution of the somatic and germline point mutations in the <i>hprt</i> gene in human T-lymphocytes <i>in vivo</i>	39
Figure 5. Current methods for detection of point mutations	42
Figure 6. The kinds, positions, and frequencies of hotspot mutations observed in the 100-bp mitochondrial DNA target sequence (mitochondrial bp 10031-10131)	57
Figure 7. Comparison of spontaneous and MNNG-induced point mutational spectra in the <i>hprt</i> gene in human lymphoblastoid B-cell lines TK6 and MT1	68
Figure 8. Diagram of the CDCE apparatus	85
Figure 9. Melting map of the <i>APC</i> gene cDNA bp 8434-8704 fragment	91
Figure 10. Quantitative measurement of the <i>Sau3A</i> I and <i>Acc</i> I digestion efficiency	96
Figure 11. Removal of interfering PCR byproducts by a post-PCR incubation procedure	100
Figure 12. Demonstration of CDCE separation of the 243-bp <i>APC</i> gene fragments at various temperatures (I)	103
Figure 13. Demonstration of CDCE separation of the 243-bp <i>APC</i> gene fragments at various temperatures (II)	105
Figure 14. Sequential release of the desired <i>APC</i> gene fragment from large quantities of human genomic DNA	108
Figure 15. Flow diagram of the CDGE-based approach to mutational spectrometry of the <i>APC</i> gene target sequence in human cells	111
Figure 16. CDGE separations of the wild-type and mutant sequences in the low-melting domain of the 271-bp <i>APC</i> gene fragment at different temperatures	114
Figure 17. Demonstration of the sensitivity for the detection of high $T_m$ mutations in the <i>APC</i> gene target sequence in human cells through the CDGE-based approach	118
Figure 18. Demonstration of the sensitivity for the detection of <i>all</i> types	

of point mutations in the <i>APC</i> gene target sequence in human cells through the CDGE-based approach	120
Figure 19. The strategy for sequence-specific enrichment of single-copy nuclear sequences from genomic DNA	124
Figure 20. Schematic representation of the temperature profile as a result of current flow across a capillary	132
Figure 21. CDCE separation performance in a 250 $\mu\text{m}$ i.d. capillary at different electric field strengths	135
Figure 22. Comparison of CDCE separation performance using capillaries of various sizes	137
Figure 23. CDCE separation performance as a function of amounts of DNA loaded on capillaries of various sizes	142
Figure 24. Reconstruction experiments demonstrating the detection sensitivity using wide-bore (540 $\mu\text{m}$ i.d.) CDCE for pre-PCR mutant enrichment	145
Figure 25. Flow diagram of the CDCE/high-fidelity PCR approach to mutational spectrometry of the 121-bp <i>APC</i> gene target sequence in human cells	150
Figure 26. Post-PCR enrichment of the amplified mutant sequences by CDCE coupled with high-fidelity PCR	154
Figure 27. CDCE separations of the mutant homoduplexes derived from the CDCE-purified wild-type DNA, two untreated and two MNNG-treated MT1 genomic DNA samples	156
Figure 28. Distribution of base substitutions in the <i>APC</i> gene target sequence (cDNA bp 8543-8663) in MNNG-treated and untreated MT1 cells	159
Figure 29. MNNG-induced point mutational spectrum in the <i>APC</i> gene target sequence in MT1 cells	163
Figure 30. CDCE separations of the mutant homoduplexes derived from TK6 cells	168
Figure 31. Comparison of CDCE separations of low $T_m$ mutants derived from the CDCE-purified wild-type DNA and untreated MT 1 cells	170
Figure 32. CDCE separations of mutant homoduplexes induced by PCR using <i>Pfu</i> DNA polymerase after 15 and 30 doublings	173
Figure 33. The mutant fraction of all low $T_m$ mutants created by <i>Pfu</i> as a function of the number of doublings	176
Figure 34. CDCE separations of the homoduplexes derived from genomic DNA of TK6 cells and the two sets of pooled blood samples	179

**LIST OF TABLES**

Table 1. Number and type of germline mutations associated with human inherited diseases	18
Table 2. Number and type of germline single base-pair substitutions in gene coding region	19
Table 3. Number and type of somatic mutations in the coding region of the <i>hprt</i> gene in healthy humans	37
Table 4. MNNG-induced DNA methylation adducts <i>in vitro</i> and <i>in vivo</i>	60
Table 5. MNNG-induced mutations observed in <i>E. coli lacI</i> gene, <i>Saccharomyces cerevisiae SUP4-o</i> gene, and human fibroblast <i>hprt</i> gene	67
Table 6. Phenotypic characteristics of TK6 and MT1 cell lines	76
Table 7. Optimized experimental conditions for sequence-specific enrichment of the desired <i>APC</i> gene fragment from digested human genomic DNA	129
Table 8. Experimental and calculated parameters for CDCE in capillaries of various bore widths	140
Table 9. Positions and estimated mutant fractions of the twelve MNNG-induced GC -> AT transitions in the two replicate MT1 cell cultures	162

**LIST OF ABBREVIATIONS**

5-meC	5-methylcytosine
6TG	6-thioguanine
A	Adenine or adenosine or absorbance or area
AFB <sub>1</sub>	Aflatoxin B <sub>1</sub>
AP	Apurinic or Apyrimidinic
APC	Adenomatous polyposis coli
ASA	Allele-specific amplification
ASP	Allele-specific PCR
bp	Base pair
BPDE	Benzo[ $\alpha$ ]pyrene diol epoxide
BSA	Bovine serum albumin
C	Cytosine or cytidine
CDCE	Constant denaturant capillary electrophoresis
CDGE	Constant denaturant gel electrophoresis
CE	Capillary electrophoresis
CPG	Controlled porous glass
DGGE	Denaturing gradient gel electrophoresis
E. coli	Escherichia coli
EDTA	Ethylenediaminetetraacetic acid
G	Guanine or guanosine
gpa	Glycophorin A
GTBP	G/T mismatch-binding protein
Hb	Hemoglobin
hprt	Hypoxanthine phosphoribosyl transferase
LCR	Ligase chain reaction

MAMA	Mismatch amplification mutation assay
MF	Mutant fraction
MGMT	O <sup>6</sup> -methylguanine-DNA methyltransferase
MNNG	<i>N</i> -methyl- <i>N'</i> -nitro- <i>N</i> -nitrosoguanidine
O <sup>6</sup> -meG	O <sup>6</sup> -methylguanine
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PE	Plating efficiency
RFLP	Restriction fragment length polymorphism
Pfu	<i>Pyrococcus furiosus</i>
SDS	Sodium dodecyl sulfate
SNP	Single nucleotide polymorphism
SSCP	Single-strand conformation polymorphism
SSPE	Sodium chloride, sodium phosphate, EDTA
T	Thymine or thymidine
TAE	Tris-acetate, EDTA
TBE	Tris-borate, EDTA
TE	Tris-HCl, EDTA
TEMED	<i>N, N, N', N'</i> -tetramethylethylenediamine
T <sub>m</sub>	Melting temperature
tk	Thymidine kinase
UV	Ultraviolet light

## 1. INTRODUCTION

Seymour Benzer and Ernst Freese (1958) demonstrated that mutagens induce unique distributions of point mutations with regard to kind and position and called these distributions (point) mutational spectra. Since then, mutational spectrometry has become a powerful tool to study mutational mechanisms in prokaryotic and eukaryotic cells (Fuchs et al., 1981; Miller, 1983; Thilly, 1990). However, in all of these studies rare mutant cells were recognized and isolated on the basis of an altered phenotype which conferred them the ability to grow under selective conditions. Such methods preclude the analysis of genes for which no selective conditions have been devised in single cell systems. They also preclude the analysis of any gene in human tissues the cells of which cannot yet be grown *in vitro*.

To overcome these limitations to mutational spectrometry, our laboratory has been developing means to "select" mutants based on differences in the cooperative melting behavior between wild-type and mutant DNA sequences. These efforts are built upon the original demonstration by Fischer and Lerman (1983) that small differences in DNA melting temperatures create clear and discernible differences in the electrophoretic mobility of DNA upon passing through a gel matrix under partially denaturing conditions. Their technique has been extended to constant denaturant capillary electrophoresis (CDCE) which uses laser-induced fluorescence detection and replaceable linear polyacrylamide matrices (Khrapko et al., 1994). In comparison to slab gel electrophoresis, CDCE offers superior performance with high resolution, high sensitivity, fast separation and easy recovery of desired DNA species (Khrapko et al., 1994).

A mutational spectrometry approach has recently been developed in our laboratory in which mutant sequences are first separated from wild-type DNA by constant denaturant gel electrophoresis (CDGE), amplified by high-fidelity polymerase chain reaction (PCR), and further enriched and separated by CDCE (Khrapko et al., 1997a). This approach has enabled us to scan for point mutations in a 100-base pair (bp) mitochondrial DNA sequence

in human cells and tissues. Reproducible mitochondrial point mutational hotspots were discovered at frequencies from  $10^{-5}$  to greater than  $10^{-4}$  (Khrapko et al., 1997b; Coller et al., 1998; Marcelino et al., 1998).

The goal of this thesis project is to extend the CDCE/high-fidelity PCR approach to the detection and measurement of mutational spectra in single-copy nuclear genes in human cells and tissues. Application of this approach to nuclear genes was not straightforward. Both the cellular copy numbers and the point mutant fractions of nuclear genes are several hundred-fold lower than those of the mitochondrial DNA (Robin and Wong, 1988; Grist et al., 1992; Robinson et al., 1994; Khrapko et al., 1997a; Khrapko et al., 1997b). In practical terms this means sorting through about 150 picograms of total cellular DNA to find a typical mitochondrial hotspot but 3 micrograms to find a typical nuclear hotspot. Unfortunately, the exquisite separation capability of capillary electrophoresis (CE) was hampered by a limited loading capacity of DNA (less than 100 nanograms) (Ruiz-Martinez et al., 1993). It was clear that one had to achieve a significant enrichment of the desired nuclear sequences prior to CE separation and/or increase the loading capacity of CE.

The criterion for an assay of sufficient sensitivity to detect nuclear hotspots differs among tissues. Studies in peripheral T-lymphocytes have shown that the average mutant fraction in the nuclear *hprt* and HLA loci is about  $10^{-8}$  per bp in middle aged humans (Grist et al., 1992; Robinson et al., 1994). To obtain reproducible observations of nuclear hotspots in blood requires one to work with  $10^9$  nucleated cells at a sensitivity of  $10^{-7}$  (Leong et al., 1985). Fortunately, in assaying for mutations in most organs, anatomically distinct sectors may be excised and analyzed in series (Coller et al., 1998). Mutations in a stem cell which are of primary interest to us may be found as a colony after the mutations are transmitted to the transitional and terminal cells of that stem cell's turnover unit. Thus a sample consisting of  $5 \times 10^7$  cells with  $10^8$  alleles and a turnover unit size of 128 would have a mutant fraction of  $10^{-6}$  if one stem cell mutated at some previous time. Our experience with dissecting human lungs and colons for mitochondrial mutation studies

suggested that a sample size as large as 0.5 g or  $10^8$  cells for parenchymal epithelia is a practical upper limit for us to set the goal to achieve a detection sensitivity of  $10^{-6}$ .

Two crucial technical developments were made in this project that extended the CDCE/high-fidelity PCR approach to mutational spectrometry of nuclear genes. The first development was a protocol involving sequence-specific hybridization coupled with a biotin-streptavidin capture system to isolate the desired sequences from bulk genomic DNA. The second development was a two-step CE separation procedure using a wide-bore capillary to highly enrich the mutant sequences in the presence of residual cellular DNA prior to high-fidelity PCR.

As an initial sequence for developmental work, a 271-bp fragment of the human adenomatous polyposis coli (*APC*) gene (*APC* cDNA bp 8434-8704) (Grodén et al., 1991) was used. The chosen *APC* fragment has the necessary contiguous high and low isomelting domains for separation of mutants from wild-type (Fischer and Lerman, 1983) as well as conveniently located restriction sites. The target sequence in which mutations can be detected comprises 121 basepairs (*APC* cDNA bp 8543-8663) within the low melting domain. It is located in the untranslated portion of the *APC* gene exon 15 where mutations have not been found associated with colon cancer. This, paradoxically, is a valuable quality for the study of mutagenesis in somatic tissues which requires the use of sequences in which mutations have no *in vivo* phenotypic advantage or disadvantage. The sensitivity and reproducibility of the improved CDCE/high-fidelity approach was demonstrated in measuring the *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (MNNG)-induced point mutations in the *APC* target sequence in nearly diploid human lymphoblastoid MT1 cells.



## 2. LITERATURE REVIEW

### 2.1. Genetic changes in humans

This section reviews the nuclear gene mutations that have been detected in humans *in vivo*, including germline mutations associated with inherited diseases, somatic mutations associated with tumors and somatic mutations in healthy humans. The discussion is focused on point mutations since they provide the most detailed information about the specificity and mechanisms involved in mutagenesis.

#### 2.1.1. Germline mutations associated with inherited diseases

Mutations occurring in germinal cells are the causes of many inherited diseases in humans. To date, nearly 12,000 different germline mutations in a total of 636 nuclear genes have been reported in the literature and collected by The Human Gene Mutation Database (HGMD) (Cooper et al., 1998). Each mutation recorded in the HGMD has been logged only once in order to avoid confusion between recurrent and identical-by-descent lesions. As shown in Table 1, the vast majority of mutations associated with inherited diseases are single base-pair substitutions (70.5%) and deletions (21.6%). The remainder consists of insertions and duplications (6.5%), indels (0.7%), repeat expansions (0.1%) and complex rearrangements not covered by the above categories (0.6%). Among the single base-pair substitutions, 7271 different mutations have been found in the coding regions of 547 different genes (Krawczak et al., 1998). The number and type of these substitutions are shown in Table 2. It can be seen that there is a hierarchy of nucleotides with respect to their propensity to undergo substitution, that is,  $G > C > T > A$ . There is also a highly significant excess of transitions (62.5%) over transversions (37.5%). The excess of transitions observed is partially attributed to those at CpG dinucleotides (i.e.,

Table 1. Number and type of germline mutations associated with human inherited diseases.

Mutation type	Number	% of Total
Single base-pair substitutions	8436	70.5
Small deletion ( $\leq 20$ bp)	1857	15.5
Small insertion ( $\leq 20$ bp)	653	5.5
Small indels ( $\leq 20$ bp)	82	0.7
Repeat expansions	15	0.1
Gross deletion ( $> 20$ bp)	736	6.1
Gross insertions and duplications ( $> 20$ bp)	122	1.0
Complex rearrangements including inversions	71	0.6
Total	11972	100.0

(Adapted from Cooper et al., 1998)

Table 2. Number and type of germline single base-pair substitutions in gene coding regions.

Original nucleotide	No. of substitutions by nucleotide				Total
	T	C	A	G	
T	...	654	271	312	1237
C	1632 (940) <sup>a</sup>	...	371	340	2343
A	201	163	...	538	902
G	619	453	1717 (735) <sup>b</sup>	...	2789
Total	2452	1270	2359	1190	7271

(a) Number in parentheses is the number of CpG → TpG transitions.

(b) Number in parentheses is the number of CpG → CpA transitions.

(Adapted from Krawczak et al., 1998)

CpG → TpG/CpA mutations), which account for 23% of all substitutions and for 36.9% of all transitions. It should be noted that considerable variation exists in the relative proportions of the different types of mutation found for different gene loci, which probably reflects both the gene structure and the selective pressure to generate a disease phenotype (Cooper and Krawczak, 1993; Mohrenweiser, 1994).

The observed preponderance of C → T and G → A transitions at CpG dinucleotides in human germline mutations is believed to be due to the high propensity of 5-methylcytosine (5-meC) to undergo spontaneous deamination to thymine (Duncan and Miller, 1980; Cooper and Youssoufian, 1988). In humans, 70 - 90% of 5-meCs occur in CpG dinucleotides (Cooper, 1983), rendering the CpG dinucleotide a mutational hotspot. To disentangle the effects that mutation and selection have on the observed germline mutation, Cooper and Krawczak (1993) have devised an iterative multivariate procedure to translate the observed proportions into relative single base-pair substitution rates which are corrected for the two confounding effects of differential codon usage and clinical observation likelihood. The relative mutation rate for either CpG → TpG or CpG → CpA has been estimated to be five times higher than the average base substitution rate (Krawczak et al., 1998). Among all of the mutated CpG dinucleotides, there are significant differences in the relative mutation rate depending on the flanking nucleotides (Krawczak et al., 1998). A preference for 5' pyrimidines and 3' purines has been observed among the mutated dinucleotides (Krawczak et al., 1998). This could be due to the sequence-dependent efficiency of both DNA methyltransferase action and G:T mismatch repair (Bolden et al., 1985; Sibghat-Ullah and Day, 1993). Some studies have reported a strand bias in either the CpG deamination frequency or the efficiency of G:T mismatch repair, with CpG → TpG transitions outnumbering CpG → CpA transitions (Skandalis et al., 1994; Leader et al., 1995). This also seems to be true for the collection of germline single base-pair substitutions (Table 2). However, after taking into account the observational bias due to codon usage and magnitude of amino acid exchange, the relative mutation rate of CpG →

CpA is estimated to be 1.4-fold higher than that of CpG  $\rightarrow$  TpG (Krawczak et al., 1998).

5-meC is known to occur at low frequency in non-CpG dinucleotides within triplets of the form CpNpG (Woodcock et al., 1998). However, the relative mutation rates of CpNpG  $\rightarrow$  TpNpG ( $N \neq G$ ) and CpNpG  $\rightarrow$  CpNpA ( $N \neq C$ ) transition in the HGMD database are found to be not significantly higher than the average substitution rate (Krawczak et al., 1998).

In addition to the hypermutability of CpG dinucleotides, some subtle and very localized effects of the surrounding DNA sequence on the relative single base-pair substitution rates have been observed in the germline mutation database (Krawczak et al., 1998). A substantial proportion of substitutions take on the identity of one of the bases immediately flanking the mutated site. This mutational bias toward next neighbors is potentially explicable in terms of misalignment mutagenesis involving highly localized DNA slippage, misincorporation and realignment events at the replication fork, between template and primer (Kunkel, 1992). A moderate correlation between the thermodynamic stability of DNA triplets and the relative mutability of the central nucleotide of a triplet has also been observed. This suggests either inefficient DNA replication in regions of high stability or transient stabilization of misaligned intermediates (Krawczak et al., 1998).

Several other endogenous mutagenic mechanisms have been proposed to explain the remaining 77% germline single base-pair substitutions that are not explicable by methylation-mediated deamination. These include DNA polymerase replication errors, depurination of DNA and DNA damage by oxygen free radicals (Loeb and Cheng, 1990). In mammalian cells, three classes of DNA polymerases,  $\alpha$ ,  $\beta$  and  $\delta$ , are responsible for the replication and repair of nuclear DNA. The error rate and mutational specificity of each polymerase have been determined *in vitro* (Kunkel, 1985; Thomas et al., 1991; Kunkel, 1992). It has been shown that the most frequent types of errors produced by these DNA polymerases are single base-pair substitutions, followed by minus-one-base frameshifts. With respect to single base-pair substitutions, transitions are more frequent than

transversions. When the observed frequencies and types of human germline single base-pair substitutions were compared with the rates and types of errors produced by the DNA polymerases  $\alpha$ ,  $\beta$  and  $\delta$  *in vitro*, a significant correlation was found between the observed substitutions in humans and those from polymerase  $\beta$  (Cooper and Krawczak, 1993). This observation supports the postulate that a substantial proportion of the single base-pair substitutions underlying human inherited diseases are caused by misincorporation during DNA replication.

Depurination of DNA results from cleavage of the *N*-glycosylic bond which is the most frequent spontaneous alteration in the chemical structure of DNA under physiological conditions (Lindahl and Nyberg, 1972). Because most DNA polymerases tend to insert deoxyadenosine opposite the non-instructional apurinic site (Kunkel, 1984; Loeb and Preston, 1986; Hevroni and Livneh, 1988), these lesions could give rise to a high frequency of transversion if they remain unrepaired prior to DNA replication. Transversion has also been observed as the predominant type of base substitution caused by oxidative damage to DNA, presumably from 8-hydroxoguanine (Cheng et al., 1992; Retèl et al., 1993). However, since transversions are under-represented in the database of single base-pair substitutions associated with human inherited diseases, it is concluded that neither depurination nor endogenous oxidative damage is a major mechanism for spontaneous mutation in humans (Cooper and Krawczak, 1993).

In humans, despite numerous studies of the potential genetic consequences of exposure to known *in vitro* mutagens, there has not been a single mutagenic agent being identified as the cause of a genetic disorder (Cooper and Krawczak, 1993). The analysis of the human germline mutation database has shown that the majority of the germline mutations are likely resulted from endogenous mutagenic processes involving chemical degradation, error-prone DNA replication and repair, and possibly, damage by endogenous mutagens. These also include a large proportion of small deletions and insertions in the database which appear to occur as a consequence of replication slippage mediated by the

presence of direct or inverted (palindromic) repeats in the immediate vicinity (Cooper and Krawczak, 1993).

The germline mutation data logged in the HGMD does not allow for a direct estimation of the mutation rates in absolute terms. This is due to a lack of information on the actual numbers of recurrent mutations and the numbers of total meioses screened for their occurrence. In addition, for a mutation to be found associated with disease, it must cause a phenotype sufficiently severe that it comes to clinical attention, yet must not cause embryo lethality. This implies that clinically observed mutations only represent a subset of all mutational events actually occurring in a population.

Several studies have attempted to estimate the germline mutation rates in a small number of gene loci (reviewed by Mohrenweiser, 1994 and Simpson, 1997). Depending on the approach and loci investigated in each study, the average germline mutation rates have been estimated to be between  $0.2 \times 10^{-5}$  and  $2 \times 10^{-5}$  per loci per generation. The nucleotide substitution rate has been estimated to be between  $2 \times 10^{-9}$  and  $1 \times 10^{-8}$  per generation. Although different loci may have different inherent mutation rates, depending on the size and sequence of the gene being screened, the difference in the estimated mutation rates may also reflect differences in the frequency of different molecular events among loci and the analytical strategies employed to detect them. All of the estimates are limited by the small number of events identified in any one screening effort. For these reasons, the estimated germline mutation rate may deviate significantly from the actual mutation rate.

#### 2.1.2. Somatic mutations associated with tumors: the *p53* and *APC* gene

##### *The p53 gene*

The *p53* tumor suppressor gene is localized on chromosome 17p13.1 and

comprises 11 exons. It encodes a nuclear phosphoprotein of 393 amino acids with transcriptional transactivation activity ((Levine et al., 1991). p53 participates in many cellular functions including cell cycle control, DNA repair, apoptosis, genomic plasticity and differentiation (Vogelstein and Kinzler, 1992; Greenblatt et al., 1994). Mutations in the *p53* gene have been found in about half of all somatic cancer types. Germline *p53* mutation is the cause of Li-Fraumeni syndrome, a rare familial autosomal dominant cancer syndrome characterized by early-onset sarcomas, brain tumors, premenopausal breast cancer, leukemias and adrenocortical tumors (Malkin et al., 1990).

A comprehensive *p53* mutation database has been developed at the International Agency of Research on Cancer (IARC). To date, this database contains more than 9000 entries of *p53* somatic mutations and 100 entries of *p53* germline mutations in human tumors (Hainaut et al., 1998). *p53* has a large mutational target size. More than 280 out of the 393 codons are found mutated in tumors, with about 74% of all of these mutations being missense mutations. In this respect, *p53* differs from other tumor suppressor genes (such as *APC*, *BRCA-1*, *p16* and *RB*) which are frequently inactivated by frameshift or nonsense mutations, and the *ras* oncogenes which are activated by mutations at three well-defined codons (Hussain and Harris, 1998). One explanation for the high prevalence of *p53* mutation in cancer is that the missense mutations can cause both a loss of tumor suppressor function and a gain of a transdominant oncogenic activity (Harris, 1996).

The distribution of the tumor-associated *p53* point mutations in the coding region of the gene is not random (Fig. 1). About 75% of all missense mutations cluster in exons 5-8 (codons 126-306), which is the sequence-specific DNA binding region of the protein and highly conserved through evolution (Soussi et al., 1990). Five out of the six dominant mutational hotspots (codons 175, 245, 248, 273 and 282) observed in the somatic spectrum are also hotspots observed in the germline spectrum (Fig. 1). All five hotspots are at methylated CpG dinucleotides. CpG → TpG/CpA transition is the major type of mutation found at these sites in both somatic and germline spectra, suggesting that

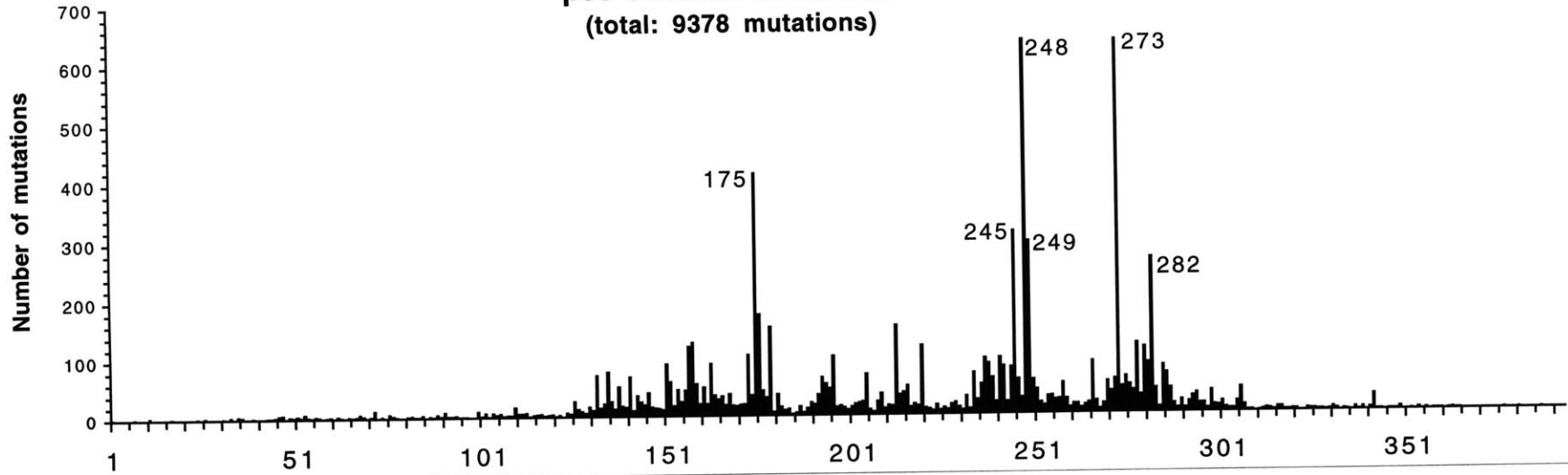


Figure 1. Distribution of the somatic and germline point mutations in the *p53* gene in human tumors.

The horizontal axis represents the *p53* codon sequence. The size of vertical bars represents the number of mutations found at a particular codon. Hotspot codon numbers are indicated. Data were obtained from the IARC *p53* database (Hainaut et al., 1998).

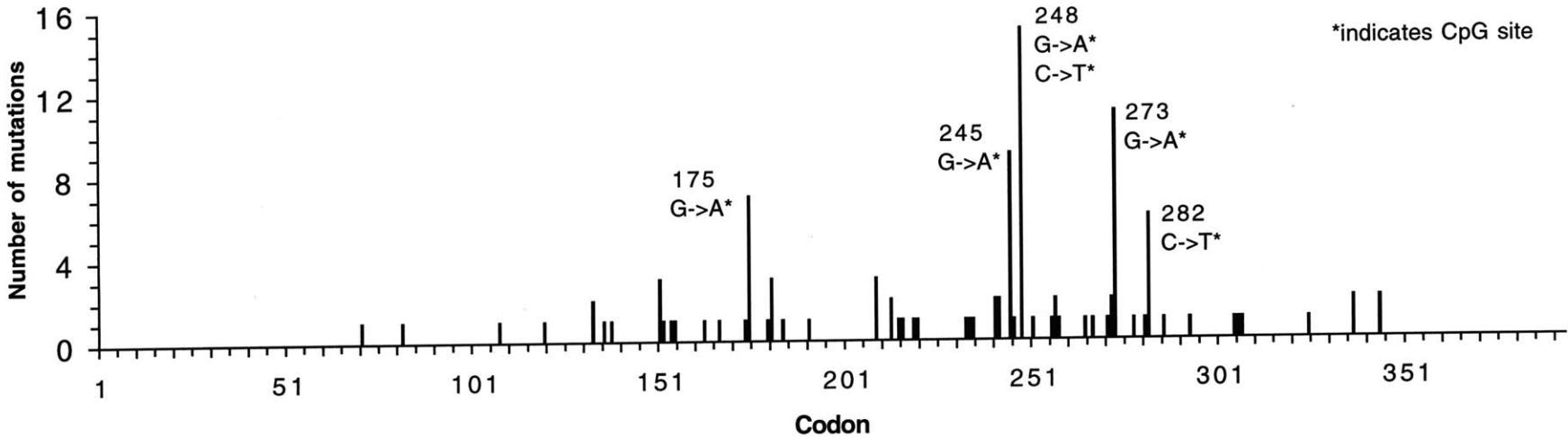
### p53 Somatic Mutations

(total: 9378 mutations)



### p53 Germline Mutations

(total: 111 mutations)



endogenous deamination of 5-meC is the major mechanism responsible. It has been demonstrated that upon exclusion of a small proportion (12%) of tissue- and site-specific point mutations, the spectrum of somatic mutations of the *p53* gene closely resembles that of the overall germline mutations of other genes (Krawczak et al., 1995). This similarity has promoted the hypothesis that the majority of *p53* somatic mutations are of endogenous origin and selected for during the process of cellular transformation (Krawczak et al., 1995).

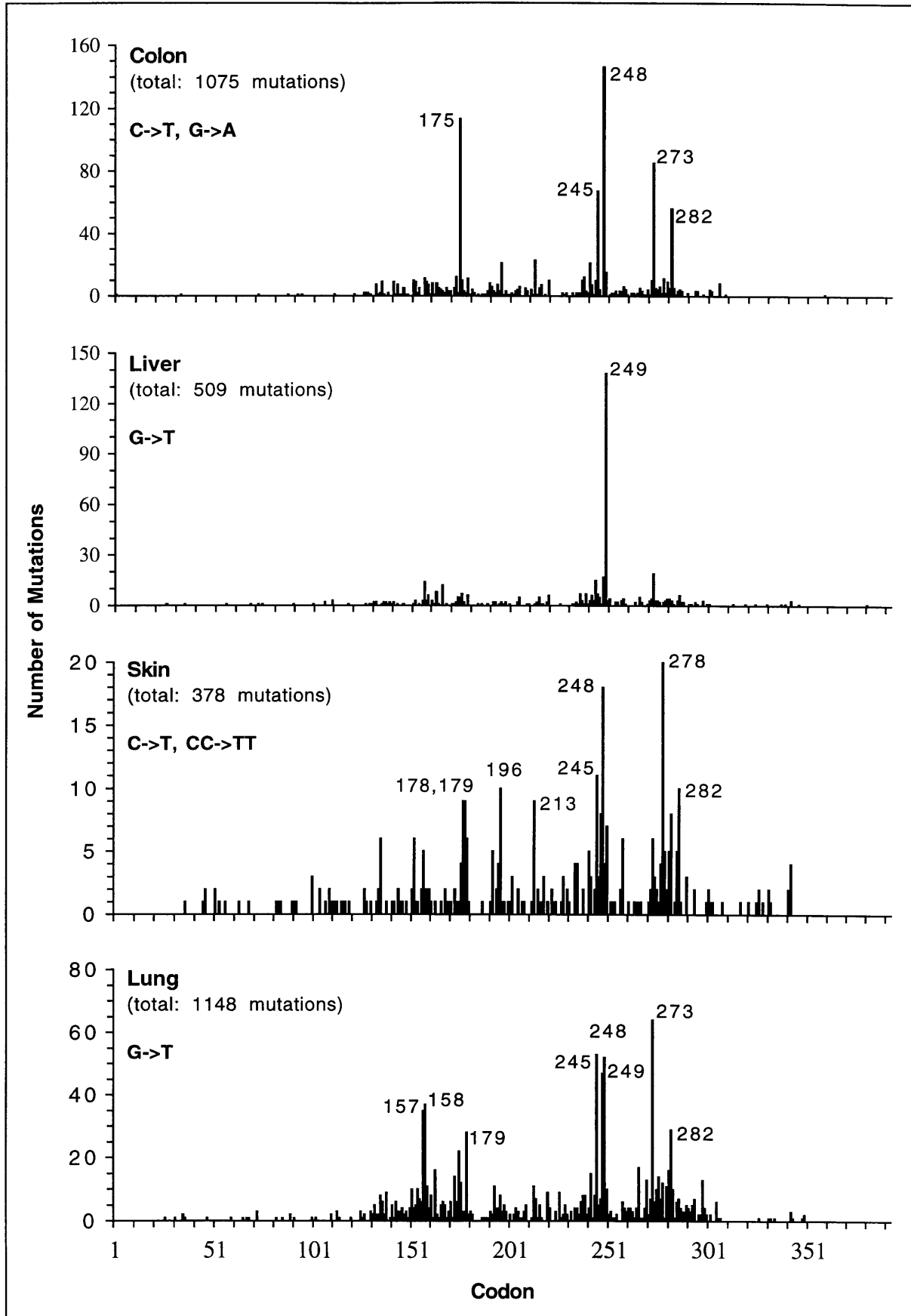
Nevertheless, it has been recognized that the *p53* mutation patterns in tumors are tissue-specific which may provide clues to the etiology of human cancer (Hollstein et al., 1991; Greenblatt et al., 1994; Hussain and Harris, 1998). Fig. 2 compares the *p53* mutation patterns in somatic tumors of colon, liver, skin and lung. The *p53* mutations in colon cancer are characterized by a high proportion (50%) of CpG → TpG/CpA transitions at CpG sites. Most of the transitions occur at the five hotspot positions which are the same as those in the germline spectrum (Fig. 1 and 2). The similarity between the *p53* mutation pattern of somatic colon cancer and that of Li-Fraumeni syndrome indicates that, like germline mutations, the *p53* mutations in colon cancer are primarily caused by endogenous deamination of 5-meC. The *p53* coding region contains 39 CpG dinucleotides (23 in exons 5-8), all of which are methylated in all human tissues and cell types analyzed (Tornaletti and Pfeifer, 1995). The nonrandom distribution of the transitions among these CpG sites may reflect selection bias for particular amino acids or the sequence-dependent efficiency of G:T mismatch repair (Sibghat-Ullah and Day, 1993; Tornaletti and Pfeifer, 1995; Greenblatt et al., 1994). On the other hand, since cytosine methylation has been found to be a major factor determining the strongly preferential formation of bulky DNA adducts at CpG sites in *p53* sequences (Denissenko et al., 1997), there is still a possibility that colon cancer is caused by specific carcinogen(s) that preferentially target methylated CpGs and cause G → A transitions (Pfeifer and Denissenko, 1998).

The *p53* mutation pattern in liver tumors is very unique in that it is dominated by

Figure 2. Distribution of the somatic point mutations in the *p53* gene in colon, liver, skin, and lung cancers.

The horizontal axis represents the *p53* codon sequence. The size of vertical bars represents the number of mutations found at a particular codon. Indicated in each tumor type are the total number of mutations, the major mutation type(s) at the hotspot codons, and hotspot codon numbers. Data were obtained from the IARC *p53* database (Hainaut et al., 1998).

P53 Somatic Mutations in Tumors



one hotspot, a G → T transversion on the nontranscribed strand at the third position of codon 249 (AGG → AGT) (Fig. 2). There is strong evidence to suggest that this mutation is specifically associated with dietary aflatoxin B<sub>1</sub> (AFB<sub>1</sub>) exposure (Wogan, 1992; Aguilar et al., 1994; Montesano et al., 1997; Hussain and Harris, 1998). Although hepatitis B virus (HBV) infection has also been associated with hepatocellular carcinomas, the mechanism is still not clear. One hypothesis postulates that recurrent liver cell proliferation occurring among HBV chronic carriers is the rate-limiting factor favoring the selective clonal development of cells mutated by AFB<sub>1</sub> in certain geographic regions (or other exogenous or endogenous agents in other regions) (Harris and Sun, 1984). Recently, it has been found that codon 249 is not the only damage site caused by AFB<sub>1</sub> in the *p53* gene (Pfeifer and Denissenko, 1998), indicating that a synergistic interaction between aflatoxin exposure and HBV infection and clonal selection may cause the observed *p53* mutation pattern in liver cancer (Greenblatt et al., 1994). It is interesting to note that in lung cancer codon 249 is also a mutational hotspot (Fig. 2), but these mutations occur almost exclusively at the second position of the codon (AGG → ATG). The second position has been found to be a target for benzo[α]pyrene (Denissenko et al., 1996). The different substitutions within the same codon between liver and lung cancer probably reflect differential binding of carcinogens to a given DNA sequence as well as differential repair of the resultant DNA adducts.

The *p53* mutations in skin cancer are characterized by a high prevalence (40%) of C → T and CC → TT transitions at dipyrimidine sites, which incriminates ultraviolet radiation as the major etiology of human skin cancer (Brash et al., 1991; Dumaz et al., 1994; Nakazawa et al., 1994; Jonason et al., 1996). There are 30 dipyrimidine sites (18 on the nontranscribed strand) in the conserved domains of the *p53* gene. The distribution of C → T transitions among these sites is nonrandom. Mutational hotspots are found at codons 178-179, 196, 213, 245, 248, 278 and 282 (Fig. 2), with a strand bias toward the nontranscribed strand. Furthermore, it is found that many of these hotspots (codons 196,

213, 245, 248 and 282) occur at dipyrimidine sites containing 5-meC which also belongs to a CpG site (Tommasi et al., 1997). The nonrandom distribution of transitions in skin cancer has been partially explained by the slow repair of pyrimidine dimers on the nontranscribed strand (Evan et al., 1993), and by the preferential formation and slow repair of pyrimidine dimers at 5-meC bases (Tornaletti and Pfeifer, 1994; Pfeifer and Denissenko, 1998).

The prevalent type of *p53* mutations found in lung cancer is G → T transversion, which comprises one-third of all missense mutations (Hernandez-Boussard and Hainaut, 1998). 93% of the transversions occur at guanine residues on the nontranscribed strand. This mutational specificity is consistent with that of several types of chemical carcinogens found in cigarette smoke such as benzo[α]pyrene. The most frequently mutated codons in lung cancer are 157, 158, 179, 245, 248, 249, 273 and 282 (Fig. 2). Studies using normal human bronchial epithelial cells have demonstrated that benzo[α]pyrene diol epoxide (BPDE) selectively binds to guanines in codons 157, 248 and 273 (Denissenko et al., 1996), and that the selective occurrence of these adduct hotspots is due to an enhancement of adduct formation by 5-meC adjacent to the target guanine at CpG sites (Denissenko et al., 1997). It has also been demonstrated that BPDE adducts on the nontranscribed strand of the *p53* gene are repaired slower than those on the transcribed strand (Denissenko et al., 1998). Recently, it has been shown that the *p53* mutation pattern in lung cancer among nonsmokers significantly differs from that found among smokers by a lower proportion of G → T transversions and a higher proportion of G → C transversions, although the number of nonsmokers analyzed in the study were relatively small (Hernandez-Boussard and Hainaut, 1998).

### *The APC gene*

The *APC* gene is localized on chromosome 5q21-22 and comprises 15 exons

(Kinzler et al., 1991; Groden et al., 1991). The APC protein consists of 2843 amino acids and is located at the basolateral membrane in colorectal epithelial cells. APC plays a central role in controlling the cell death process by integrating signals from a variety of sources and transmitting them to the nucleus (Kinzler and Vogelstein, 1996). The *APC* gene has been identified as the gatekeeper gene of colon cancer, the mutation of which initiates the cascade of colorectal tumorigenesis (Fearon and Vogelstein, 1990).

Mutations in the *APC* gene have been found in over 80% of sporadic colorectal cancer and in some cancer of the stomach, pancreas, thyroid, ovary and other primary sites (Nagase and Nakamura, 1993; Kinzler and Vogelstein, 1996). Germline *APC* mutation is the cause of familial adenomatous polyposis, an autosomal dominant precancerous condition characterized by the appearance of hundreds to thousands of adenomatous polyps throughout the entire colorectum. Although a database containing over 1000 somatic and germline mutations in the *APC* gene is now available (Laurent-Puig et al., 1998), each mutation in this database has been logged only once. In order to identify potential mutational hotspots, the *APC* gene mutations have been re-compiled from the literature taking into account recurrent mutations (B. Muniappan and X.-C. Li-Sucholeiki, unpublished data). The data set which comprises 380 somatic mutations and 288 germline mutations is shown in Fig. 3.

Over 95% of *APC* mutations are either frameshift or nonsense mutations leading to a truncated protein (Laurent-Puig et al., 1998; Hussain and Harris, 1998). The restricted types of mutations found in the *APC* gene do not allow for the kind of extensive epidemiological studies as those done with the *p53* gene. Nevertheless, the observations that most nonsense mutations are C → T transitions at CpG or CpNpG sites and that frameshifts are small deletions at repeats of 2 - 5-bp DNA motifs (Fig. 3) suggest the involvement of endogenous deamination and slipped mispairing mechanisms. Interestingly, the *APC* somatic and germline mutational patterns are not alike (Fig. 3). The somatic mutations are clustered in a region (codons 1286-1513) within exon 15, whereas germline



Figure 3. Distribution of the somatic and germline mutations in the *APC* gene in human tumors.

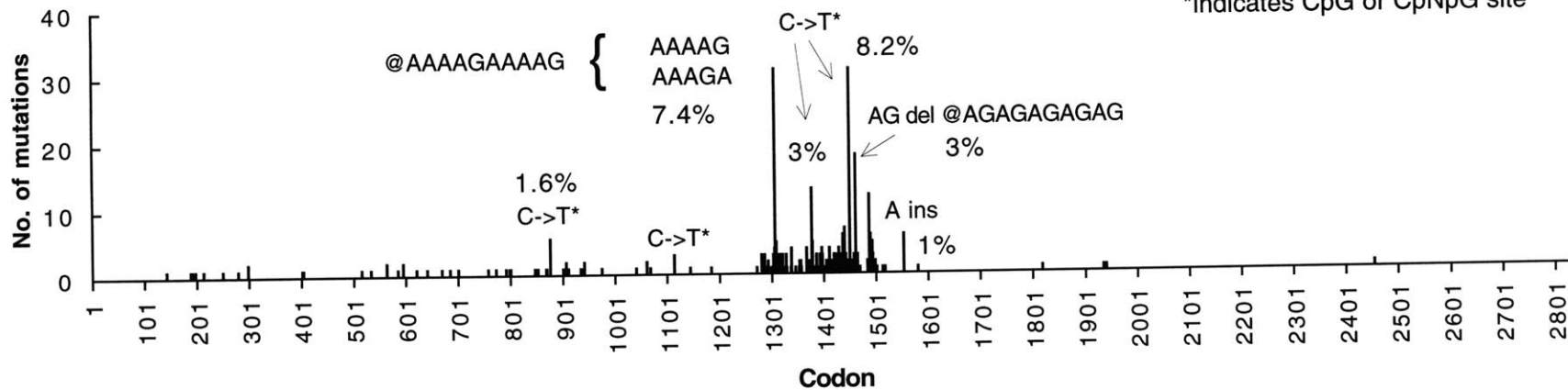
The horizontal axis represents the *APC* codon sequence. The size of the vertical bars represents the number of mutations found at a particular codon. The mutation types at hotspot codons are shown.

(Kindly provided by Brindha Muniappan)

### APC Somatic Tumor Mutations

(total: 380 mutations)

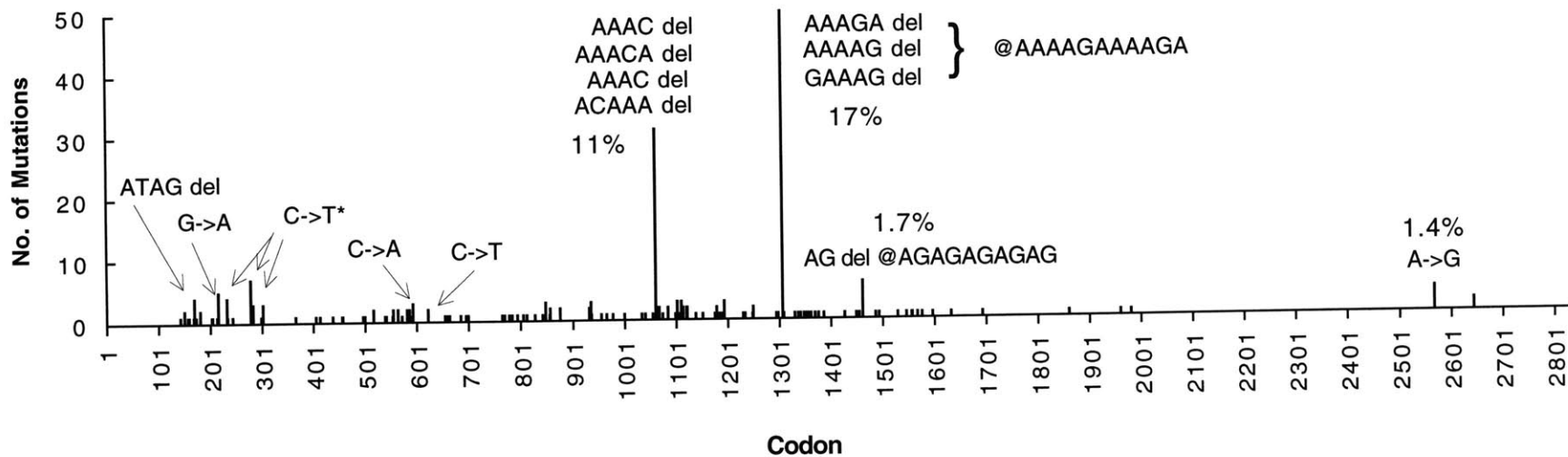
\*indicates CpG or CpNpG site



### APC Germline Tumor Mutations

(total: 288 mutations)

\*indicates CpG or CpNpG site



mutations are scattered throughout the first half of the coding region (Miyoshi et al., 1992; Nagase and Nakamura, 1993). The origin of this difference is unknown. It may reflect differences in the mechanisms of mutation, repair or selection among these two different cell types.

### 2.1.3. Somatic mutations in healthy humans: the *hprt* gene

Somatic mutations in healthy humans have been measured at four gene loci in blood cells on the basis of altered cellular phenotypes: the hemoglobin (*Hb*) and glycophorin A (*gpa*) genes in red blood cells, and the hypoxanthine-guanine phosphoribosyltransferase (*hprt*) and HLA genes in T-lymphocytes (Cole and Skopek, 1994). The mean adult background mutant frequencies at these loci have been determined to be  $5 \times 10^{-8}$  for *Hb*,  $5 - 10 \times 10^{-6}$  for *hprt*,  $1 - 2 \times 10^{-5}$  for *gpa* and  $3 \times 10^{-5}$  for HLA (Albertini et al., 1993; Cole and Skopek, 1994). A significant increase in the mutant frequency with age has been found for all loci (Grist et al., 1992; Cole and Skopek, 1994; Robinson et al., 1994). With respect to detecting a broad spectrum of mutations, the *hprt* gene has been a particularly useful target.

The *hprt* gene is a constitutive but dispensable housekeeping gene localized on the X-chromosome. It comprises nine exons and encodes an enzyme of 218 amino acids that phosphoribosylates hypoxanthine and guanine for purine salvage. HPRT is also able to phosphoribosylate the purine analog 6-thioguanine (6TG) to 6TGMP which is highly cytotoxic. Inactivating mutations in the *hprt* gene allow the mutant cells to grow in the presence of otherwise toxic concentrations of 6TG, providing a convenient means for selection. Detection and characterization of somatic *hprt* mutations in human peripheral blood T-lymphocytes involves direct cloning (Albertini et al., 1982; Morley et al., 1983). T-cells are cultured under limiting dilutions in the presence or absence of 6TG selection. The ratio of cloning efficiency with and without 6TG defines the *hprt* mutant frequency.

Mutant colonies can be isolated and propagated *in vitro* for the molecular characterization of their mutations (Albertini et al., 1982; Morley et al., 1983).

It has been found that gross structural alterations are the predominant type of *hprt* mutations found in newborns (85%), whereas they only account for 15% in adults (Albertini et al., 1993). The most prevalent alterations found in newborns are due to deletions mediated by the V(D)J recombinase. As people age, the *hprt* mutant frequencies increase and the proportion of gross alterations decrease. This is presumably due to an accumulation of endogenous and induced mutations that are not gross alterations.

A comprehensive *hprt* mutation database has been compiled from the literature (Cariello et al., 1997). In combination with a recent study (Podlutzky et al., 1998), there are nearly 500 somatic mutations that have been detected in peripheral T-lymphocytes of healthy humans (including both smokers and non-smokers), and 130 germline mutations in patients with Lesch-Nyhan syndrome (LNS) or gout. Table 3 shows the number and type of somatic mutations found in the coding region of the *hprt* gene. Single base-pair substitutions are the most common mutations observed which constitute 58% of the total. One percent are tandem base substitutions. Deletions and insertions (including frameshifts) account for 37% of the total mutations. The remaining 4% are complex mutations. Among the single base-pair substitutions, mutations at GC base pairs (61%) are more prevalent than mutations at AT base pairs (39%), despite the fact that there are fewer mutable GC base pairs (48%) than AT base pairs (Podlutzky et al., 1998). The majority of the mutations at GC base pairs (71%) occurred at guanines on the non-transcribed strand. G/C → A/T is the most common type of base-pair substitution (35%).

There are eight CpG sites in the *hprt* coding region, the methylation status of which is still unknown (Podlutzky et al., 1998; O'Neill and Finette, 1998). Among the single base-pair substitutions in the *hprt* gene, transitions at CpG sites account for 6.6% (18/271) of all somatic mutations (at basepairs 142, 151 and 508), and 14% (14/96) of all germline mutations (at basepairs 142, 151, 481 and 508) (O'Neill and Finette, 1998; Podlutzky et

Table 3. Number and type of somatic mutations in the coding region of the *hprt* gene in healthy humans.

Type of mutation	No. of mutation	% of total
Single base-pair substitution	271	58 (100) <sup>a</sup>
Transition	127	(47)
G -> A	62	(23)
C -> T	33	(12)
A -> G	14	(5)
T -> C	18	(7)
Transversion	144	(53)
G -> T	28	(10)
C -> A	2	(1)
G -> C	27	(10)
C -> G	13	(5)
A -> T	18	(7)
T -> A	21	(8)
A -> C	4	(1)
T -> G	31	(11)
Mutation at AT	106	(39)
Mutation at GC	165	(61)
Tandem substitution	5	1
Frameshift	35	7
+1 bp	8	1
-1 bp	27	6
Deletion	134	29
3 - 52 bp	12	3
> 52 bp	122	26
Insertion	4	1
Complex	18	4
Total	467	100

(a) Number in parentheses represents the percentage of each type of substitution among the total single base-pair substitutions.

(Data were obtained from Cariello et al., 1997 and Podlutzky, 1998).

al., 1998). Therefore, it has been concluded that transition mutations at CpG dinucleotides are the most frequent *hprt* germline and somatic mutations, implicating endogenous deamination as the major mutagenic mechanism involved in *hprt* mutation *in vivo* (O'Neill and Finette, 1998).

The distribution of the point mutations in the *hprt* gene is non-random (Fig. 4). It has been estimated that there are > 300 mutable sites in the *hprt* coding region where mutations are expected to give rise to a selectable phenotype (Cariello and Skopek, 1993). Among the mutational hotspots (with 4 or more mutations at a single site), base pairs 197, 508 and 617 are the three most prominent ones (Fig. 2). Nearest neighbor analysis has shown that the mutated guanine at basepair 197 is preceded by a CTC-trinucleotide which is the vertebrate topoisomerase I consensus cleavage site (CTC or CTT) (Podlutzky et al., 1998). This consensus sequence has been implicated in single base-pair substitutions at other positions (such as basepairs 143 and 146) (Podlutzky et al., 1998). All mutations at basepair 508 are C → T transitions at a CpG site. Basepair 617 involves a (TG)<sub>2</sub>-repeat.

It has been shown that most of frameshifts and small deletions occur in a sequence context of mononucleotide runs, short di- or trinucleotide repeats, or invert repeats, which resembles certain polymerase arrest sites, symmetric elements, and the consensus cleavage sequence for vertebrate topoisomerase I (Podlutzky et al., 1998).

Taken together, the present spectrum of *hprt* somatic mutations does not seem to pinpoint any particular kind of exogenous agent as the cause for these mutations. The observations of the same set of mutational hotspots in different human populations have led to the postulate that *in vivo* mutations in the *hprt* gene are endogenous in origin or induced by ubiquitous environmental exposures (Burkhart-Schultz et al., 1996; Podlutzky et al., 1998). The spectra of *hprt* somatic mutations in both smokers and non-smokers have been compared. So far, no statistically significant differences have been observed between the two subpopulations (Vrieling et al., 1992; Burkhart-Schultz et al., 1996).

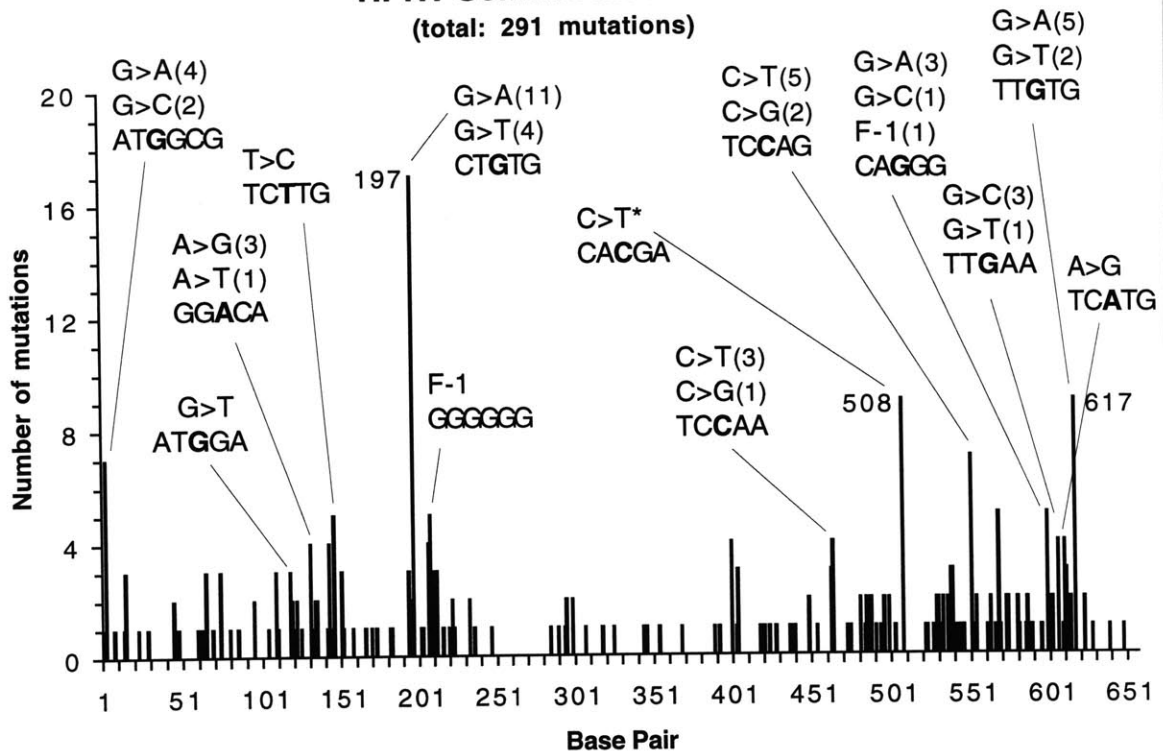
Figure 4. Distribution of the somatic and germline point mutations in the *hprt* gene in human T-lymphocytes *in vivo*.

Data were obtained from Cariello et al., 1997 and Podlutzky et al., 1998. The horizontal axis represents the *hprt* coding sequence. The size of the vertical bars represents the number of mutations found at any one particular basepair. The types of mutations of the mutational hotspots and the neighboring sequence contexts are shown as indicated. \* indicates CpG sites.

(Kindly provided by Aoy Tomita-Mitchell)

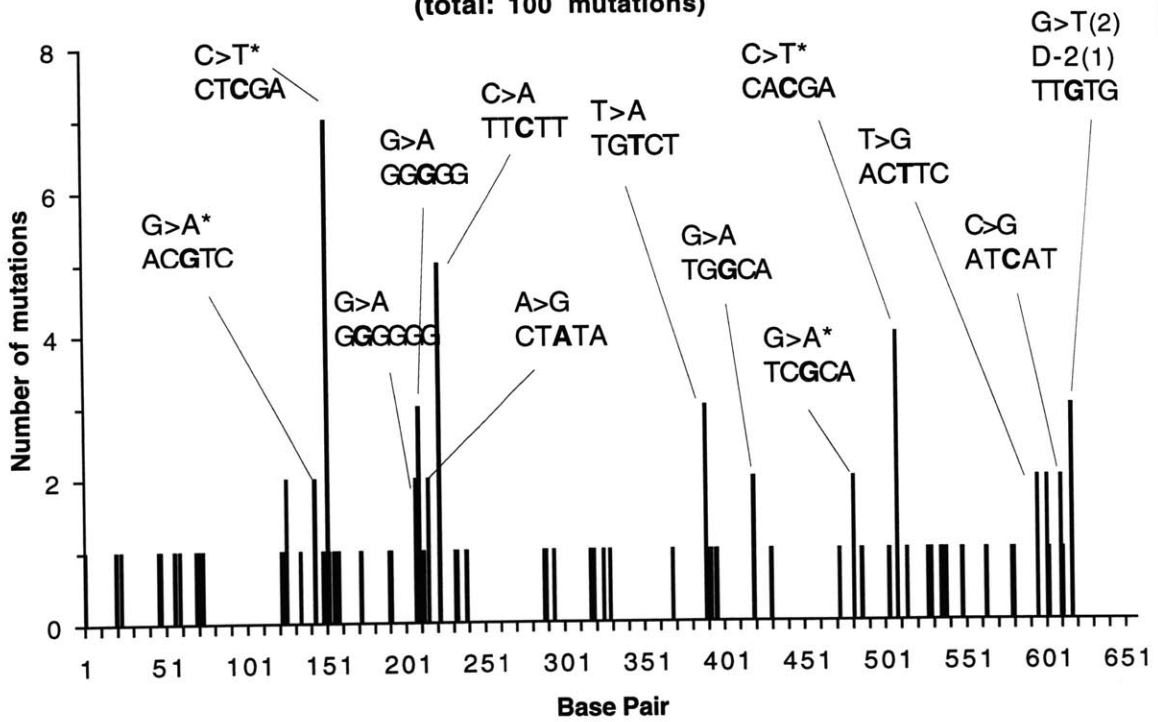
### HPRT Somatic Mutations

(total: 291 mutations)



### HPRT Germline Mutations

(total: 100 mutations)





## **2.2. Current methods for detecting point mutations**

Detection of point mutations has become central to human genetic analysis, such as diagnosis of genetic diseases, discovery of disease-associated genes and identification of the causes of human mutation. These studies require effective methods that can permit rapid screening for mutations in a large number of genes and detection of rare mutants against the background of an excess of wild-type sequence. This section reviews the current methods for detection of point mutations with respect to their principles, sensitivities, applications and limitations (Fig. 5).

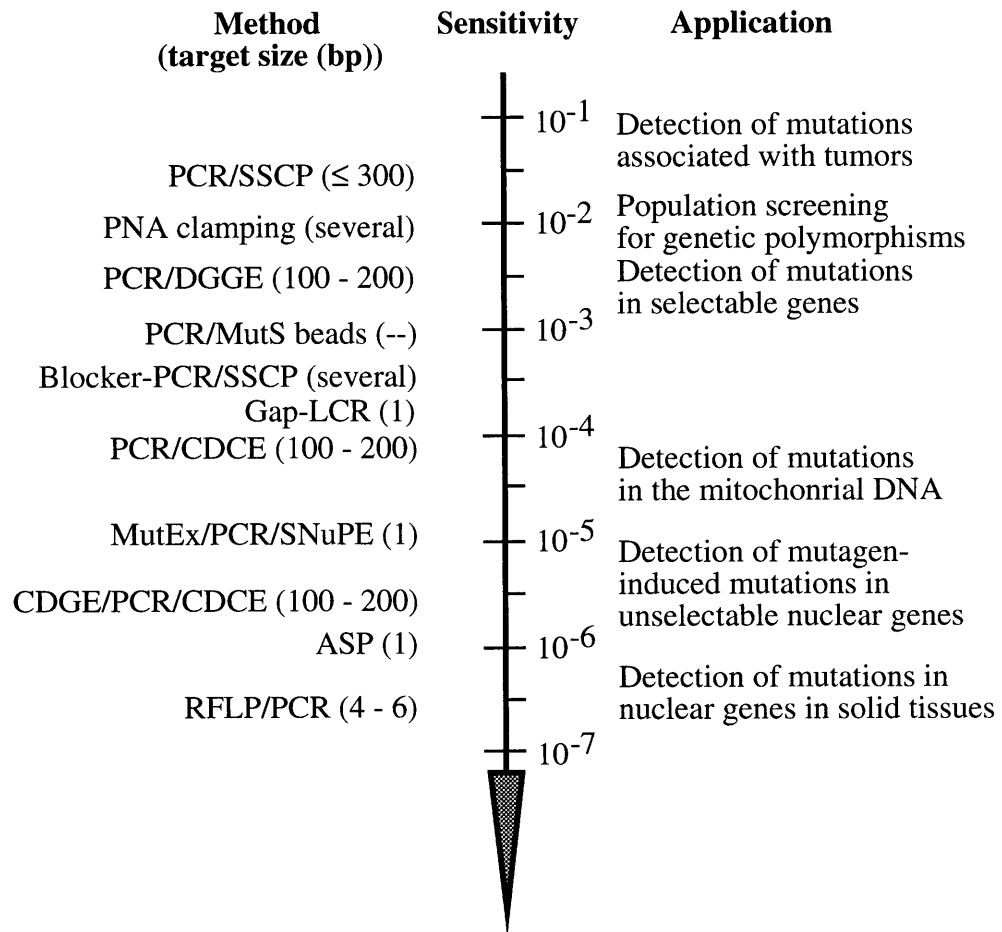
### 2.2.1. Single-strand conformation polymorphism (SSCP)

SSCP identifies most sequence variations in single-strand DNA molecules based on their differential electrophoretic mobilities in a non-denaturing polyacrylamide matrix (Orita et al., 1989). Under nondenaturing conditions, a DNA single-strand will adopt a conformation (presumably through internal base-pairing between short segments) that is uniquely dependent on its sequence composition. This conformation is usually different even if only a single base has been changed. Most conformations seem to alter the physical configuration so that the configuration-to-charge ratio is different enough to cause a mobility difference upon electrophoresis through a polyacrylamide gel.

Typically, PCR is first performed to amplify the analyzed sample with possible genotypic differences. The diluted PCR products are then denatured into single strands and loaded on a nondenaturing polyacrylamide gel. The SSCP single strands are detected using autoradiography of  $^{32}\text{P}$  incorporated during PCR as  $^{32}\text{P}$ -labeled primers or deoxynucleotide triphosphate substrates, or by silver or ethidium bromide staining of the SSCP gels (Sekiya, 1993; Ainsworth et al., 1991). The detection efficiency for single base substitutions by SSCP has been reported to be more than 90% in fragments equal to or less

Figure 5. Current methods for detection of point mutations.

The sensitivities of the methods are here presented in relation to the sensitivities required for various applications. The sensitivity for each method refers to the reported lowest mutant fraction that can be detected by the method as reported in the literature. Abbreviations are given in the text. Dash line (--) indicates data not available.



than 300-bp in length (Sekiya, 1993; Fan et al., 1993). Recently the conventional slab gel analysis of SSCP has been replaced by capillary electrophoresis with laser-induced fluorescence detection using linear polyacrylamide matrices, which has greatly increased the speed and efficiency of this detection method (Katsuragi et al., 1996; Ren et al., 1997; Wenz et al., 1998).

The PCR-SSCP analysis has been widely used in screening for mutations associated with cancers and inherited diseases (Suzuki et al., 1990; Groden et al., 1990; Kishimoto et al., 1992; Mashiyama et al., 1991; Katsuragi et al., 1996; Ren et al., 1997; Wenz et al., 1998; Humphries et al., 1997). The advantage of this technique is the large target size that can be scanned at any one time. The disadvantage is its limited sensitivity in detecting minor mutant alleles among mixed cell populations, which has been reported to be 1 in 4 to 1 in 33 (Suzuki et al., 1990; Wu et al., 1993; Hongya, et al., 1993).

### 2.2.2. Restriction fragment length polymorphism (RFLP)

RFLP (or restriction site mutation assay) detects mutations that alter a restriction endonuclease recognition sequence (Pourzand and Cerutti, 1993a; Steingrimsdottir et al., 1996; Jenkins et al., 1998). Digestion with a chosen restriction enzyme cleaves the wild-type recognition sequence but leaves mutated sites intact. The uncut fragments are then amplified by PCR with primers flanking the restriction sites. Mutants are then identified and quantified by cloning and oligonucleotide-specific hybridization. A known number of copies of "mutant standard" is usually added to the cellular DNA prior to the digestion procedure for later data calibration and estimation of the absolute mutation frequencies (Sandy et al., 1992; Chiocca et al., 1992).

The sensitivity of this approach varies from study to study and the restriction sites analyzed. The highest sensitivity has been reported to be in the range of  $10^{-8}$  to  $10^{-6}$  (Pourzand and Cerutti, 1993a). The primary limitation on the sensitivity of this approach is

the efficiency of the restriction digestion step. Not every restriction site can be analyzed because restriction enzyme digestion does not always approach completeness. In this regard, *Taq* I endonuclease is considered to be particularly well-suited for this assay (Sandy et al., 1992; Chiocca et al., 1992). The thermostability of *Taq* I endonuclease allows for the continuous digestion of the residual wild-type sequence before, during, and after PCR. A secondary limitation on the sensitivity of this approach is the inherent error rate of the DNA polymerase at any particular base pair in PCR. High-fidelity polymerase such as *Pfu* DNA polymerase has been used to increase the sensitivity of the assay (Pourzand and Cerutti, 1993b). Additional factors affecting the performance of the assay include certain types of DNA damage that block restriction digestion and are later converted into mutations during PCR (Steingrimsdottir et al., 1996).

The RFLP/PCR approach has been applied to detecting mutations in two *Msp* I sites in codons 11 and 12 of the human *c-H-ras* 1 gene (Pourzand and Cerutti, 1993b; Amstad and Cerutti, 1995), *Msp* I and *Hae* III sites in codons 247-250 of the human *p53* gene (Aguilar et al., 1993; Hussain et al., 1994a; Hussain et al., 1994b), an *Mnl* I site in rat *H-ras* codon 12 (Lu and Archer, 1992), a *Taq* I site in exon 2 of the human *c-H-ras* 1 gene (Chiocca et al., 1992), and a *Taq* I site at the *aprt* gene of Chinese hamster cells (Steingrimsdottir et al., 1996). In these studies, mutational hotspots were detected at these restriction sites in cells treated with various mutagenic agents including *N*-methyl-*N*-nitrosourea (MNU), *N*-ethyl-*N*-nitrosourea, ethyl methanesulphonate, 4-nitroquinoline-1-oxide, AFB<sub>1</sub> and oxidative radicals. This approach has also been applied toward measuring the frequency of mutations in the *p53* gene in normal human liver samples from different geographical regions that differs in the level of dietary AFB<sub>1</sub> exposure (Aguilar et al., 1994). The frequency of the AGG to AGT mutation at codon 249 in the samples was found to range from 2 - 13 x 10<sup>-7</sup> in the low exposure regions to 0.5 - 1.8 x 10<sup>-5</sup> in the high exposure region. The results support the hypothesis that AFB<sub>1</sub> is a causative mutagen in liver cancer (Aguilar et al., 1994).

Although the RFLP/PCR approach has a very high detection sensitivity, it is constrained in that not every gene sequence of interest encompasses a restriction enzyme cleavage site and that the target size is limited to 4 - 6 base pairs for each assay.

### 2.2.3. Allele-specific PCR (ASP)

ASP detects specific point mutations based on preferential PCR amplification of the rare mutant allele over the wild-type allele. The strategy is to use an oligonucleotide primer that has fewer 3' mismatch(es) with the mutant allele than with the abundant wild-type allele (Bottema and Sommer, 1993). Synonymous terms for this technique in the literature include Allele-Specific Amplification (ASA), PCR Amplification of Specific Alleles (PASA), Amplification Refractory Mutation System (ARMS) and Mismatch Amplification Mutation Assay (MAMA). ASP requires the use of a DNA polymerase, generally *Taq* polymerase, that lacks a 3' → 5' exonuclease activity since such activity would remove the 3'-terminal mismatch(es). The amplified products are then analyzed by gel electrophoresis, and then detected by autoradiography of <sup>32</sup>P incorporated as labeled deoxynucleotide triphosphate substrates in PCR (Cha et al., 1992), or by staining the gel with ethidium bromide or SYBR green (Sarkar et al., 1990; Chen and Zarbl, 1997).

The reported sensitivity of ASP varies from  $2.5 \times 10^{-2}$  to  $2 \times 10^{-6}$  in the literature (Sarkar et al., 1990; Cha et al., 1992; Chen and Zarbl, 1997). To establish an ASP assay with a high sensitivity requires optimization of a variety of parameters including the mismatch primer sequence; the temperature of the primer extension step; the time permitted for extension; and the composition of the reaction mixture, particularly the concentrations of dNTP, MgCl<sub>2</sub>, primer and glycerol (Cha et al., 1992). It has been shown that five mutant alleles among  $3 \times 10^6$  copies of wild-type alleles can be detected through the use of Stoffel fragment coupled with two rounds of PCR (Chen and Zarbl, 1997).

The ASP approach has been widely applied to population screening for single base-

pair substitutions in genes associated with genetic diseases. These genes include the phenylalanine hydroxylase gene associated with phenylketonuria (Sommer et al., 1989), factor IX gene associated with hemophilia (Sarkar et al., 1990), apolipoprotein B gene associated with hypercholesterolemia (Ruzicka et al., 1992), and mitochondrial DNA associated with neuromuscular disease (Münscher et al., 1993; Zhang et al., 1993; Seibel et al., 1993). In these studies, the detection sensitivity has been reported to be around  $10^{-2}$ . Using MAMA, Cha et al. (1994) were able to detect spontaneous mutations at a frequency of  $10^{-5}$  in codon 12 of the rat *H-ras* gene and showed that MNU-induced rat mammary tumors arose from preexisting *H-ras* mutations.

Although it can reach a high level of sensitivity, the major limitation of ASP is that only one specific point mutation can be detected at any one time.

#### 2.2.4. Blocker-PCR

Blocker-PCR detects point mutations based on the preferential amplification of the mutant alleles through the use of a pair of oligonucleotide primers that match both strands of the wild-type sequence but not the mutants (Seyama et al., 1992). These blocker primers contain 3'-terminal dideoxynucleotide so that they cannot be extended during PCR, and also block polymerase elongation of a pair of amplification primers that flank at the 5' side of each of the blocker primers. Conditions are chosen so that binding of the blocker primers to sequences containing a mutation in the binding region is not favored. Therefore the mutant sequence can be preferentially amplified using the pair of amplification primers. By linking this approach to SSCP, Seyama et al. (1992) were able to detect mutations in a human *N-ras* gene at a frequency of  $10^{-4}$  -  $10^{-3}$ . The advantage of blocker-PCR is its relatively large target size. Any base substitution at a number of positions is detectable as long as it destabilizes the hybridization of the blocker-primers.

A modification of the blocker-PCR technique is peptide nucleotide acid (PNA)-

mediated PCR clamping (Thiede et al., 1996). In this approach, a PNA primer is used as a blocker primer to cover the target site. One of the amplification primers overlaps with the PNA primer sequence but does not include the target sequence. Because PNA:DNA duplexes are more stable than DNA:DNA duplexes, the PNA primer will bind to the wild-type sequence more strongly than the amplification primer and suppress the amplification of the wild-type allele during PCR. If the target sequence is mutated, the binding of the amplification primer will be favored over that of the PNA primer and provide selective amplification of mutant alleles. This approach alone has been shown to have a sensitivity of  $5 \times 10^{-2}$  (Thiede et al., 1996). By combining this approach with SSCP analysis, Behn et al. (1998) were able to increase the sensitivity 10 - 50-fold in detecting point mutations at codons 248, 249 and 273 of the *p53* gene.

#### 2.2.5. Ligase chain reaction (LCR)

As ASP and blocker-PCR, LCR also detects point mutations based on preferential amplification of the mutant alleles, but using a different strategy. In LCR, two pairs of synthetic oligonucleotides, which are designed to hybridize at adjacent positions to the complementary strands of the desired mutant allele, are joined by a thermostable ligase. Multiple rounds of denaturation, annealing and ligation result in the exponential amplification of the mutant allele. Since a mismatch at the ligation joint severely reduces the efficiency of ligation, other sequences (including the wild-type) that differ by a single base pair are discriminated (Barany, 1991). By using allele-specific oligonucleotides that differ in length, LCR can be used to detect multiple substitutions at a single position (Barany, 1991). Different approaches are available to detect and quantify the ligated product. Oligonucleotide primers can be labeled with  $^{32}\text{P}$ , fluorescent dyes, digoxigenin, or biotin, and the ligated product monitored by size using gel electrophoresis (Barany, 1991; Kälin et al., 1992; Wiedmann et al., 1993). The sensitivity of LCR is typically less



than  $10^{-2}$  due to the generation of target-independent ligation products. This occurs when complementary primers form duplex molecules having blunt-ends which can then be ligated (Wiedmann et al., 1993).

Several improvements have been made to increase the sensitivity of LCR. One of them is Gap-LCR, which uses complementary probe pairs containing 3' extensions (Abravaya et al., 1995). After hybridization to the target DNA, a gap of one to several bases forms between adjacent probes. A thermostable DNA polymerase (devoid of 3' → 5' exonuclease activity) and the appropriate nucleotide(s) are then used to fill the existing gap so that the extended probes can be jointed by DNA ligase. The use of probe duplexes with non-complementary 3' extensions prevents the generation of target-independent ligation products. In combination with an immunoassay system for detection of the amplification products, Gap-LCR has been applied to the detection of a human immunodeficiency virus (HIV) mutation associated with AZT resistance with a sensitivity of  $10^{-4}$  (Abravaya et al., 1995).

The limitation of the LCR method is that mutations occurring at only one specific base-pair can be detected at any one time.

#### 2.2.6. MutS-based assays

Recently several approaches to mutation detection have been developed based on the selective binding property of the *Escherichia coli* (*E. coli*) mismatch binding protein MutS to DNA molecules containing mismatched bases (Ellis et al., 1994; Lishanski et al., 1994; Wagner et al., 1995; Smith and Modrich, 1996; Parsons and Helflich, 1997a). This binding can be detected utilizing the gel mobility-shift assay (Lishanski et al., 1994), the MutH-, MutL-, and MutS-dependent d(GATC) cleavage reaction (Smith and Modrich, 1996), or the binding-dependent protection from exonuclease digestion assay (Ellis et al., 1994). For example, the MutS/exonuclease (MutEx) assay (Ellis et al., 1994) was initially

developed to map allelic variants in exon 11 of the cystic fibrosis transmembrane regulator gene. Heterozygous or homozygous PCR products were first generated from individuals who were mutant or wild-type for an analyzed gene fragment. The PCR products were then incubated with MutS and treated with T7 DNA polymerase containing 3' → 5' exonuclease activity. The heteroduplexes bound to MutS were blocked from the exonuclease digestion, and thereby the length of the product indicated the position of the mutation.

The MutEx assay was later adapted to screening pools of alleles for a known but rare mutation (C → A) at codon 61 of the mouse *H-ras* gene (Parsons and Helflich, 1997a). The sample was denatured and reannealed to convert the rare mutant alleles into heteroduplexes in the presence of an abundant wild-type allele, and subjected to the MutEx selection step. The MutS-protected sequences were amplified by PCR. A single nucleotide primer extension (SNUPE) was then performed to identify the mutant and wild-type base. The overall sensitivity of this MutEx/PCR/SNUPE approach was determined to be  $2 \times 10^{-5}$ . Since the sensitivity of SNUPE alone is between 1 in 5 and 1 in 50, it was determined that MutEx assay was able to enrich the mutant sequence approximately 1000-fold (Parsons and Helflich, 1997a). The use of SNUPE, however, limits this approach to detecting mutations at one base-pair at a time. Recently, through a combination of mutant enrichment by MutEx (using the thermostable *Thermus aquaticus* MutS protein) and allele-specific competitive blocker-PCR (ACB-PCR), the detection sensitivity of the codon 61 mutation has been reported to reach  $1 \times 10^{-7}$  (Parsons and Helflich, 1998).

It has been shown that MutS immobilized to a solid support exhibits an enhanced ability to discriminate between DNA with and without mismatches as compared to MutS in solution (Wagner et al., 1995). In a reconstruction experiment, mixtures of mutant and wild-type DNA were amplified by PCR using 5'-biotinylated primers. Heteroduplex molecules in the PCR products were then captured by the MutS-coated paramagnetic beads. Using a colorimetric assay based on the biotin-label, it was shown that mutant fractions as low as  $10^{-3}$  could be measured (R. Wagner, unpublished results).

The sensitivity of the MutS-based assays is limited by the background binding of MutS to wild-type homoduplexes. These assays have so far only been carried out in model experiments. The major disadvantage of using MutS in mutation detection is that MutS has different binding affinities for different types of mismatched bases (Su and Modrich, 1986; Babic et al., 1996). This would make it difficult to measure the frequencies of unknown mutations. In addition, it has been shown that MutS binds several types of DNA lesions (Rasmussen and Samson, 1996; Duckett et al., 1996). These lesions, after being enriched by MutS, could be converted into mutations during PCR and give rise to false signals. The reliability and applicability of the MutS-based assays in detecting rare mutations in biological samples remains to be demonstrated.

#### 2.2.7. Denaturing gradient gel electrophoresis (DGGE) and constant denaturant gel electrophoresis (CDGE)

DGGE and CDGE separate double-stranded DNA differing by single base substitutions based on the same theory: DNA cooperative melting equilibrium (Fischer and Lerman, 1983). According to this theory, the melting of a DNA fragment is a stepwise process with each melting domain of the fragment melting as a single unit at a particular denaturing condition. If a DNA fragment consists of two domains, one domain has a higher melting temperature ( $T_m$ ) than the other, then within a certain range of denaturing conditions, the low melting domain of the DNA fragment will melt while the high melting domain remains intact. This partially melted form, which exists in a rapid equilibrium with the unmelted form, has a reduced mobility upon electrophoresis as it travels through a polyacrylamide matrix. The overall mobility of the DNA fragment therefore depends on the stability of its low melting domain, which in turn is extremely sequence dependent. It has been shown that even single base-pair variations in the low melting domain have an altered  $T_m$  allowing for their separations by gel electrophoresis under appropriate denaturing

conditions. The technique is even more effective for separating mutant/wild-type heteroduplexes from wild-type homoduplex DNA because a single mismatch will dramatically destabilize the low melting domain (Thilly, 1985). The addition of a GC-clamp (which is an artificial high melting domain) to the end of a DNA fragment permits analysis of any DNA sequence (Myers et al., 1985b; Sheffield et al., 1989). The advantage of the technique is the large target size (100 - 150-bp) that can be scanned at once.

DGGE employs a linear gradient of urea/formamide in a polyacrylamide gel matrix such that the DNA migrates into an increasing concentration of denaturant. The molecule travels through the gel in the helical form until it reaches the concentration of denaturant at which its low melting domain becomes single-stranded resulting in a great reduction in the mobility of the fragment. After electrophoresis, mutations that have a higher or lower  $T_m$  than that of the wild-type DNA will be found respectively at a higher or lower concentration of denaturant in the gel, i.e., below or above the position of the wild-type DNA.

In combination with PCR, DGGE has been widely used in screening for unknown point mutations associated with tumors and other genetic diseases (Myer et al., 1985a; Kogan and Gitschier, 1990; Higuchi et al., 1991; Weinstein et al., 1990; Newmark et al., 1992). Another important application of the PCR/DGGE approach has been the analysis of entire populations of mutants in a given segment of a phenotypically selectable gene such as the human *hprt* gene (Thilly, 1985; Keohavong and Thilly, 1992a). In this approach,  $10^4$  -  $10^5$  HPRT<sup>-</sup> mutants are generated *in vitro* through the treatment of cultured human cells with a mutagenic agent. The selective agent, 6TG, is added to the bulk culture. Mutants induced by the treatment will eventually overgrow the culture. The genomic DNA is isolated from the complex mutant population. The target sequence is then amplified by PCR and analyzed by DGGE. The DNA bands in the gel are detected using autoradiography of <sup>32</sup>P incorporated during PCR as <sup>32</sup>P-labeled primers. Mutant bands are then individually extracted, purified and sequenced. This approach has been applied to determining the mutational spectra of a variety of genotoxicants, including MNNG, ICR-

191, ultraviolet light, oxygen, hydrogen peroxide, BPDE, cisplatin, X-rays, 4-nitroquinoline-1-oxide, chromium (VI), and AFB<sub>1</sub> (Cariello et al., 1990; Keohavong et al., 1991; Oller and Thilly, 1992; Keohavong and Thilly, 1992a; Keohavong and Thilly, 1992b; Cariello et al., 1992; Okinaka et al., 1993; Cariello and Skopek, 1993; Chen and Thilly, 1994; Cariello et al., 1994). The detection sensitivity of this approach is about  $10^{-2}$ , although detection of a mutant fraction at  $10^{-3}$  has been reported in a reconstruction experiment (Keohavong and Thilly, 1992b).

Several improved protocols of DGGE have been reported to optimize the detection of point mutations. In the double-gradient DGGE technique, a secondary porosity gradient, typically 6.5 - 12% polyacrylamide, is colinearly superimposed to the primary denaturing gradient (Cremonesi et al., 1997; Gelfi et al., 1997). With the help of this secondary sieving gradient, some heteroduplex molecules, which often produce smeared and diffuse bands on conventional DGGE (due to a lack of a sharp melting transition), are recompacted into remarkably narrow bands. Double-gradient DGGE has been also shown to improve the resolution between two homoduplex bands. This technique has been applied to the analysis of point mutations in several exons of the cystic fibrosis transmembrane conductance regulator gene as well as in exons 6 and 8 of the *p53* gene in tumor specimens (Cremonesi et al., 1997; Gelfi et al., 1997).

Other improvements have recently been made to facilitate the analysis of very GC-rich sequences that cannot normally be analyzed by conventional DGGE due to the strand dissociation phenomena (Guldborg et al., 1998). It has been shown that treatment of template DNA with sodium bisulphite drastically lowers the melting temperature of GC-rich sequences and renders them amenable to DGGE analysis. The use of this bisulphite DGGE technique has been demonstrated in the rapid and efficient detection of mutations in the *p16* (INK4/CDKN2) tumor suppressor gene (Guldborg et al., 1998).

CDGE is a modification of DGGE in which a specific denaturing concentration is used allowing for a maximal separation between the wild-type and mutant sequences

(Hovig et al., 1991). In DGGE, the continual movement of partially melted molecules into an increasingly higher concentration of denaturant can give rise to new molecular configurations which may result in loss of resolution as running times are extended. By contrast, in CDGE, DNA molecules which differ by a single base change migrate at consistently different mobilities causing the resolution to behave as a simple function of the distance traveled. It has been shown that when operated under optimal denaturing condition (which corresponds to the steep parts of the profile observed in DGGE), CDGE can provide an increased resolution as compared to DGGE (Hovig et al., 1991; Ridanpää et al., 1995). CDGE in combination with PCR has been applied to the rapid screening for mutations in the *p53* and *BRCA1* genes in a large numbers of cancer patients (Børresen et al., 1991; Andersen et al., 1998).

#### 2.2.8. Constant denaturant capillary electrophoresis (CDCE)

CDCE combines the separation principles underlying CDGE with capillary electrophoresis (CE) which uses laser-induced fluorescence detection and replaceable linear polyacrylamide matrices (Khrapko et al., 1994b). The denaturing conditions in CDCE are achieved through temperature elevation of a section of the capillary which is inserted into a temperature controlled water-jacket. In comparison to slab gel electrophoresis, CDCE offers higher resolution, sensitivity and speed, as well as on-line detection and an amenability to automation.

By coupling high-fidelity PCR with the techniques of CDGE and CDCE for enrichment of mutants, an approach has been developed to measure point mutational spectra in a 100-bp mitochondrial DNA sequence (mitochondrial bp 10,030-10,130) in human cells and tissues (Khrapko et al., 1997a). In this approach, mitochondrial mutant sequences in the digested genomic DNA are first separated from an excess of wild-type sequence by CDGE in their homoduplex forms. CDGE is used at this stage because of its

relatively large sample loading capacity. Through CDGE the high and low  $T_m$  mutants (mutants with melting temperatures higher and lower than that of the wild-type homoduplex, respectively) are enriched separately by extracting DNA from the gel blocks above and below the position of the wild-type homoduplex. The efficiencies of mutant enrichment have been found to be 100 to 200-fold for the high  $T_m$  mutants and 5 to 10-fold for the low  $T_m$  mutants. The mutant-enriched samples are then subjected to high-fidelity PCR using native *Pfu* DNA polymerase. At the end of the PCR reaction all of the rare mutants are converted into heteroduplexes with the wild-type sequence. The PCR products are then separated by CDCE. The heteroduplex fraction is collected to further enrich the mutants over 20-fold and then amplified by a second round of high-fidelity PCR. The mutants are then finally separated on CDCE in their homoduplex forms. Individual mutants are then purified and sequenced. The frequency of a mutant in the original sample is measured by comparing the peak area of the mutant with that of the internal standard. The detection limit of this approach when applied to measuring mitochondrial mutational hotspots in human cells has been determined to be  $10^{-6}$  for the high  $T_m$  mutants and  $10^{-5}$  for the low  $T_m$  mutants (Khrapko et al., 1997a).

The above mutational spectrum analysis has been applied to several healthy human tissues including colon, muscle and lung, their derived tumors, and human lymphoblastoid TK6 cells grown in culture (Khrapko et al., 1997b). In the lung study, bronchial epithelial cells from smoking and nonsmoking twins were also compared (Coller et al., 1998). A set of seventeen hotspots, predominately G:C  $\rightarrow$  A:T and A:T  $\rightarrow$  G:C transitions, have been identified in the 100-bp mitochondrial target sequence with mutant fractions ranging from  $10^{-5}$  to greater than  $10^{-4}$  in the analyzed human tissues (Fig. 6). Remarkably, the same set of point mutational hotspots have been observed in all of the samples investigated. Several tests have been performed to verify the true existence of the 17 hotspots in the analyzed samples (Khrapko et al., 1997a; Khrapko et al., 1997b). The observations of the same set of mutants in different organs and in cultured cells have led to the conclusion that

mitochondrial point mutations in the tissues and sequence studied are spontaneous in origin (Khrapko et al. 1997b).

This same approach has also been applied to the analysis of chemically-induced mutational spectra in the 100-bp mitochondrial sequence in MT1 cells (Marcelino et al., 1998). The TK6-derived MT1 cells, which are highly resistant to MNNG cytotoxicity but remains sensitive to MNNG mutagenicity, were treated with a high dose (4  $\mu\text{M}$ ) of MNNG. A set of seven G:C->A:T transitions at frequencies ranging between  $6 \times 10^{-5}$  and  $3.0 \times 10^{-4}$  were reproducibly observed in the treated cells and were found distinct as compared to spontaneous background mutations (Fig. 6). Interestingly, when MT1 cells were repeatedly treated with a maximum tolerated dose (0.4  $\mu\text{M}$ ) of BPDE, no induced mitochondrial mutations were observed (Marcelino et al., 1998). These results demonstrated that an exogenous mutagen can induce detectable mutations in the human mitochondrial DNA. However, such observations seem only possible when a specific cell line which is resistant to a mutagen's toxicity is treated with an extremely high dose of the given mutagen which far exceeds the tolerance of normal cells. These results further support the previous conclusion that spontaneous mutagenic processes are the predominate causes of human mitochondrial mutation. These studies suggest that the mitochondrial DNA, due to its high level of spontaneous background mutations, may not be an appropriate target for biomonitoring the effect of environmental and occupational exposure on human mutation.

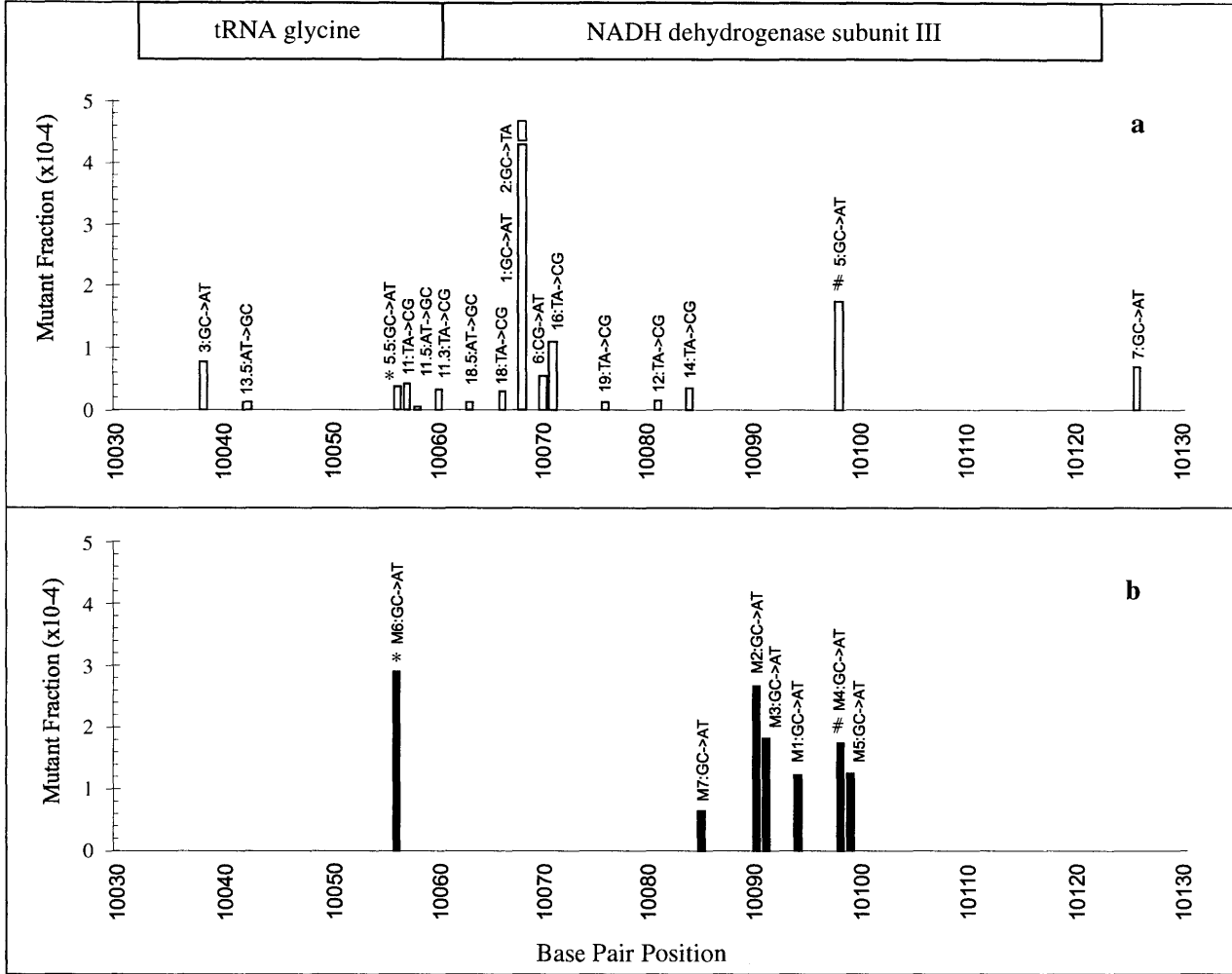
Due to its high sensitivity and ability to scan a large target size, the CDCE/high-fidelity PCR approach has been shown to be the most powerful method currently available for detecting low frequency point mutations.



Figure 6. The kinds, positions, and frequencies of hotspot mutations observed in the 100-bp mitochondrial DNA target sequence (mitochondrial bp 10,031-10,130).

The horizontal axis represents the base pair position of the mitochondrial DNA target sequence. The positions of the two genes encoded by this sequence, tRNA<sup>gly</sup> and NADH dehydrogenase subunit III, are shown as indicated. *a*, The set of hotspot mutations in human tissues. The height of each vertical bar represents the average mutant fraction of each mutant in most of the lung samples analyzed. Note that mutants 1 and 2 are different substitutions at the same bp. *b*, MNNG-induced hotspot mutations in MT1 cells. "\*" and "#" indicate identical sites of mutation in the two mutational spectra.

(Adapted from Marcelino et al., 1998)



### 2.3. MNNG mutagenesis

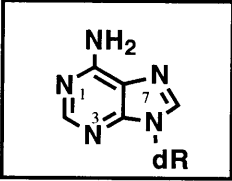
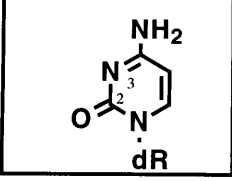
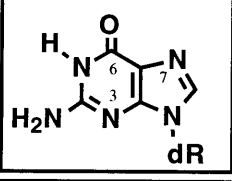
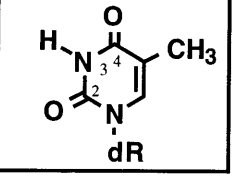
MNNG is a well-studied simple monofunctional alkylating agent. This section briefly reviews: (1) the current knowledge of the mutational consequences of DNA lesions induced by MNNG; (2) the cellular repair mechanisms; and (3) the influence of the local DNA sequence and the efficacy of repair on the distribution of the lesions and the observed mutational spectrum.

#### 2.3.1. The MNNG-induced DNA lesions and their mutagenic potentials

MNNG is not naturally occurring in the environment. It was first synthesized in 1947 through the nitrosation of *N*-methyl-*N'*-nitrosoguanine in cold acid medium (McKay and Wright, 1947). Since the discovery of its mutagenic activity in 1960, MNNG has become one of the most widely used chemical mutagens. MNNG interacts with DNA via cationic methyldiazonium intermediates through an  $S_N1$  reaction mechanism. It methylates cellular DNA at positions N1-, N3-, N7-adenine (A), N3-, N7-, O<sup>6</sup>-guanine (G), N3-, O<sup>2</sup>-cytosine (C), N3-, O<sup>2</sup>-, O<sup>4</sup>-thymine (T), and the phosphate (diester) group (Beranek, 1990). Table 4 summarizes the kinds of DNA methylation adducts that are induced by MNNG and their half-lives *in vitro* and *in vivo* (Goldmacher et al., 1986; Beranek, 1990). Adducts that have only been identified in synthetic homopolymers (at N<sup>6</sup>-A, N<sup>4</sup>-C and N<sup>2</sup>-G) are excluded here.

The N7 position of G represents the principle nucleophilic center in DNA. It is not surprising that N7-meG accounts for about 70 - 80% of the total methylation produced by MNNG in DNA (Lawley and Orr, 1970; Lawley, 1976; Goldmacher et al., 1986). The mutagenic contribution of N7-meG had been overlooked because the lesion was not thought to be misinstructional nor replication arresting (Abbott and Saffhill, 1979; Swann and Magee, 1968; Loveless, 1969). However, alkylation at N7 can increase the rate of

Table 4. MNNG-induced DNA methylation adducts in vitro and in vivo.

Site of methylation	Percentage of total adducts (%)				Half-life (hours)			
	in vitro	in vivo			in vitro	in vivo		
		bacteria	human cells	rat liver		human cells	rat liver	
 <chem>Nc1ncnc2n(cnc12)N</chem> dR	Adenine							
N1-	1.0	–	< 0.2%	–	–	–	–	–
N3-	12.0	2.0	2.9	8.6	17-38	< 1	2-19	
N7-	–	–	3.4	–	–	13	7	
 <chem>Nc1cc[nH]c(=O)n1</chem> dR	Cytosine							
O <sup>2</sup> -	–	–	–	–	–	–	< 5	
N3-	2.0	–	–	–	–	–	–	
 <chem>Nc1nc2[nH]cnc2c(=O)[nH]1</chem> dR	Guanine							
N3-	–	–	9.7	–	–	19	17	
O <sup>6</sup> -	7.0	11.0	9.2	9.2	24	20	5-72	
N7-	67.0	78.0	74.8	82.2	105-192	17	29-72	
 <chem>Cc1c[nH]c(=O)[nH]c1=O</chem> dR	Thymine							
O <sup>2</sup> -	–	–	–	–	–	–	12	
N <sup>3</sup> -	–	–	–	–	–	–	2	
O <sup>4</sup> -	–	–	–	–	–	–	< 5	

depurination by as much as six orders of magnitude through the labilization of the *N*-glycosylic bond (Loeb and Preston, 1986). The half-life of this process *in vitro* (at neutral pH and at 37°C) has been estimated to be between 105 and 192 hours (Singer, 1979). The resultant apurinic (AP) sites are mutagenic (Kunkel, 1984; Loeb and Preston, 1986; Hevroni and Livneh, 1988). In addition, N7-meG is subject to imidazole ring opening to yield a formamidopyrimidine (Fapy) product, which blocks *in vitro* replication (Boiteux and Laval, 1983; O'Connor et al., 1988). Thus, N7-meG alone may be of significant biological importance by virtue of its relative abundance.

N3-meA represents one of the most frequently found minor products of MNNG-induced DNA methylation. Although no miscoding property has been found for N3-meA (Saffhill and Abbott, 1978), the methyl group of N3-meA labilizes the *N*-glycosylic bond so that this base is rapidly lost from the DNA. Its half-life *in vitro* is between 17 and 38 hours (Singer, 1979). In addition, the act of DNA glycosylase on N3-meA produces AP sites as intermediates in alkylation repair (Riazuddin and Lindahl, 1978). The mutagenicity of N3-meA in *E. coli* is SOS dependent and include transversion (Foster and Eisenstadt, 1985).

N3-meG does not appear to block replication (Larson et al., 1985). Although it has yet to be convincingly demonstrated *in vivo*, both O<sup>2</sup>-meC and O<sup>2</sup>-meT have been implicated to have potential miscoding properties (Singer, 1986).

O<sup>6</sup>-meG represents the most important DNA lesion with respect to the induction of gene mutation and cancer by simple alkylating agents. Since the late 1960s, it has been commonly believed that the amount of O<sup>6</sup>-meG in DNA and the rate of its excision decidedly influences the frequency of induced GC → AT transition mutation and organ susceptibility to tumor induction (Loveless, 1969; Schendel and Robins, 1978; Singer, 1979; Newbold et al., 1980; Pegg, 1984). O<sup>6</sup>-meG is capable of mispairing with thymine in both *in vitro* and *in vivo* site-specific mutational systems (Mehta and Ludlum, 1978; Abbott and Saffhill, 1979; Loechler et al., 1984; Eadie et al., 1984; Bhanot and Ray,

1986). It is stable in DNA *in vitro* at physiological pH (Gichner and Veleminsky, 1982).

O<sup>4</sup>-meT is the second most effective premutagenic DNA lesion produced by MNNG. Suggested in studies on methylation induced mutation, O<sup>4</sup>-meT can give rise to AT → GC transition mutations by mispairing with guanine (Preston et al., 1986; Richardson et al., 1987b).

Why O<sup>6</sup>-meG and O<sup>4</sup>-meT prefer to mispair remains unclear. It may reflect preferential hydrogen-bonding capability, or the Watson-Crick alignment of the O<sup>6</sup>-meG:T and O<sup>4</sup>-meT:G mispair may facilitate formation of the phosphodiester links on both the 3' and 5' side of the incoming base during DNA replication (Swann, 1990). Other alternative mechanisms for the mutagenic consequences of O<sup>6</sup>-meG and O<sup>4</sup>-meT have also been suggested. For example, polymerase might mistake O<sup>6</sup>-meG in the template strand for adenine and O<sup>4</sup>-meT for cytosine, due to the physical similarity of these bases (Yamagata et al., 1988). O<sup>6</sup>-meG may protonate its cytosine complement as a consequence of base-pairing. If so, GC → AT transition may result from the increased likelihood of this cytosine to deaminate to uracil (William and Shaw, 1987). DNA glycosylases may repair O<sup>2</sup>-meT resulting in AP sites, which induce preferentially AT → GC transitions (McCarthy et al., 1984; Kunkel, 1984). And lastly, O<sup>6</sup>-meG in the methylated nucleotide pools could incorporate at a position opposite to a thymine residue on the template strand (Toorchen and Topal, 1983; Eadei et al., 1984).

### 2.3.2. The repair of MNNG-induced DNA lesions

Damage induced by DNA methylation in living cells is repaired by two major pathways: direct transfer of the modifying methyl group to the repair protein; and/or removal of the modified base by a glycosylase, leaving an AP site which is then repaired by an excision repair process (Friedberg et al., 1995). Nucleotide excision repair, which is usually involved in repairing adducts that distort the DNA helix, has recently been shown

to be the third pathway playing a backup role in repairing lesions normally repaired by alkyltransferases or glycosylases.

The major pathway for repairing O<sup>6</sup>-meG in both prokaryotic and eukaryotic cells involves the transfer of the methyl group from the DNA to a cysteine acceptor site in the O<sup>6</sup>-methylguanine-DNA methyltransferase (MGMT) (Pegg and Byers, 1992). MGMT brings about this transfer without the need for any cofactor. The DNA is restored completely in a single step, but the cysteine receptor site in the protein is not regenerated. The alkylated form of the protein is unstable in mammalian cells and is degraded rapidly. The presence of MGMT allows for the rapid repair of O<sup>6</sup>-meG following exposure to MNNG as long as the pool of MGMT is not depleted. If the number of O<sup>6</sup>-meG molecules formed in DNA exceeds the content of MGMT, further repair would require either *de novo* synthesis of additional MGMT or the activation of other pathways (Pegg and Byers, 1992). MGMT is not specific for methyl groups. It is also able to remove larger adducts from the O<sup>6</sup> position of DNA, although the rate of such a reaction goes down as the size of the alkyl group increases (Samson et al., 1988).

The amino acid sequences of MGMT have been determined for a variety of species including bacteria, yeast, rodent and human. There is a considerable degree of homology among all of these MGMT sequences (Lindahl et al., 1988; Demple, 1990; Margison et al., 1990; Morohoshi et al., 1990; Tano et al., 1990; Rydberg et al., 1990; Potter et al., 1991; Hakura et al., 1991). Both *E. coli* and *Bacillus subtilis* contain a constitutive MGMT (Ogt and Dat, respectively) and an inducible MGMT (Ada and Ada B, respectively). In eukaryotes, only one MGMT has been detected. The level of the MGMT expression was found to be inducible (three to five-fold more) by MNNG treatment in rat but not human cells (Fritz et al., 1991). The human MGMT has 207 amino acids and its gene is localized on chromosome 10 (Tano et al., 1990; Rydberg et al., 1990). There are striking organ and species differences in the quantities of MGMT found in mammals (Pegg, 1990). In general, human tissues and cells have much higher quantities than their equivalent tissues in

rodents. The liver and spleen have the highest activities of the protein, while the nervous system and mammary gland have the lowest. There are also significant differences in MGMT levels among different cell types within any one particular organ. For example, the level of MGMT is found to be much higher in hepatocytes than in nonparenchymal cells of the liver (Pegg and Byers, 1992). Such differences may be of importance to a cell's sensitivity to carcinogenesis or killing by alkylating agents.

Work with the two forms of MGMT (Ogt and Ada) found in *E. coli* indicates that both of them are able to repair O<sup>4</sup>-meT (Lindahl et al., 1988; Demple, 1990). However, there have been conflicting reports as to the potential of mammalian MGMT to repair O<sup>4</sup>-meT. *In vitro* experiments have demonstrated that purified preparations of human and rodent MGMT can accept a methyl group from O<sup>4</sup>-meT (Zak et al., 1994; Kawate et al., 1995). But the mammalian MGMT does not seem to repair this lesion *in vivo* (Brent et al., 1988; Pegg, 1990). Although there are some observations suggesting that O<sup>4</sup>-meT may be lost from cellular DNA at a relatively slow rate, the enzymatic basis for this reaction has not been established (Wani et al., 1990; Boucheron et al., 1987).

It has been shown that the excinuclease in the nucleotide excision repair pathway can act on O<sup>6</sup>-meG *in vitro* and *in vivo* (Voigt et al., 1989; Samson et al., 1988; Rossi et al., 1989). Although the repair efficiency is relatively poor (Van Houten and Sancar, 1987), it has been suggested that excision repair may provide the main protection against the mutagenicity and toxicity of methylating compounds under conditions in which both the constitutive and inducible MGMTs are saturated (Samson et al., 1988; Rossi et al., 1989). The susceptibility to repair by the *E. coli* *uvrABC* excinuclease has been shown to increase with increasing alkyl chain length (Todd and Schendel, 1983; Samson et al., 1988). One study with human cell lines showed that the cells' capacity for nucleotide excision repair did not affect either the degree of cell killing nor the mutant frequency induced by MNNG. However there was an inverse correlation between them when the cells were treated with ethylnitrosourea (ENU) (Maher et al., 1990). These results indicate that in human cells



nucleotide excision repair is able to repair O<sup>6</sup>-ethylG, but not O<sup>6</sup>-meG.

The major pathway for repairing N-methylpurines is base excision repair. In this pathway, DNA glycosylases catalyze the hydrolytic cleavage of the *N*-glycosylic bond linking an abnormal or damaged base to deoxyribose in the sugar-phosphate DNA backbone, which upon hydrolysis produces an AP site in DNA (Friedberg et al., 1995). The non-instructive AP sites are then repaired starting with the cleavage of phosphodiester bonds at the 5' or 3' position on the AP sites by AP endonucleases or lyases, respectively (Bailly and Verly, 1988). After incision at the AP sites, excision of the damaged sites by exonucleases and resynthesis of the gapped region then follows (Friedberg et al., 1995).

Several N-alkylpurine-specific glycosylases have been identified. *E. coli* expresses one constitutive and one alkylation-induced N-methylpurine DNA glycosylase (MPG) (Lindahl et al, 1988; Shevell et al., 1990). The constitutively expressed Tag glycosylase has a narrow substrate specificity, excising only N3-meA from alkylated DNA. The alkylation-inducible AlkA glycosylase has a somewhat broader substrate specificity, efficiently excising N3-meA, N3-meG, O<sup>2</sup>-meC and O<sup>2</sup>-meT while inefficiently excising N7-methylpurine from alkylated DNA. *E. coli* also possesses a Fapy-DNA glycosylase enzyme, which exhibits a broad substrate specificity for a variety of lesions having ring-opened purines including the imidazole ring-opened form of N7-meG (Laval et al., 1990). cDNAs encoding for the N-methylpurine-DNA glycosylase genes from yeast, rat and human have also been isolated (Chen et al., 1989; Berdal et al, 1990; O'Connor and Laval, 1990; Samson et al., 1991). The mammalian proteins share a significant homology, but are quite different from the two *E. coli* MPGs and the yeast enzyme (MAG protein). The human N-methylpurine glycosylase gene (*MAP*) is located on chromosome 16 and comprises five exons. The purified MAP is able to catalyze the excision of N3-meA, N7-meG and N3-meG from methylated DNA, but is more specific for N3-meA. It is estimated that there are 1000 to 2000 MAP molecules per cell (Friedberg et al., 1995).

Recently it has been shown that nucleotide excision repair may be involved in

removing N-methyl DNA adducts in *Drosophila* (Nivard et al., 1996). After treatment with the methylating agents (MNU and *N*-nitrosodimethylamine), the excision-repair-deficient strains show an increase in the frequency of A/G → T transversions as compared to the wild-type strain. Since the induced transversions are presumably caused by N<sup>3</sup>-meA and N<sup>7</sup>-meG adducts, the results suggest that N-methyl DNA adducts may be the targets for nucleotide excision repair (Nivard et al., 1996).

### 2.3.3. The mutational specificity of MNNG

#### *The types of mutations induced by MNNG*

The spectra of MNNG-induced mutations contain transitions, transversions, deletions and insertions (Coulondre and Miller, 1977; Reed and Hutchinson, 1987; Burns et al., 1987; Richardson et al., 1987a, b; Kohalmi and Kunz, 1988; Gordon et al., 1990; Cariello et al., 1990; Yang et al., 1991; Lukash et al., 1991; Yang et al., 1993; Kat, 1992; Kat, 1993). Table 5 summarizes the kinds and numbers of MNNG-induced mutations observed in the *lacI* gene in bacteria, *SUP4-o* gene in yeast and *hprt* gene in human cells using clone-by-clone analysis (Kohalmi and Kunz, 1988; Gordon et al., 1990; Yang et al., 1991; Lukash et al., 1991; Yang et al., 1993). The MNNG-induced point mutational spectra in the human *hprt* gene have also been determined by *en masse* selection of tens of thousands of 6TG resistant mutants coupled with the PCR/DGGE (or CDCE) approach (Cariello et al., 1990; Kat, 1992; Kat et al., 1993; Tomita-Mitchell et al., unpublished results). The obtained mutational spectra in human lymphoblastoid B-cell lines TK6 and MT1 are shown in Fig. 7.

As shown in Table 5 and Fig. 7, the mutations induced by MNNG in these studies are dominated by single base-pair substitutions, particularly GC → AT transitions. Despite the fact that O<sup>6</sup>-meG is only a minor methylation product, the preponderance of transition at

Table 5. MNNG-induced mutations observed in *E. coli lacI* gene, *Saccharomyces cerevisiae SUP4-o* gene, and human fibroblast *hprt* gene.

Number of mutations (% of the total)			
	<i>E. coli</i> ( <i>lacI</i> ) <sup>a</sup>	<i>S. cerevisiae</i> ( <i>SUP4-o</i> ) <sup>b</sup>	Human fibroblasts ( <i>hprt</i> ) <sup>c</sup>
<b>Substitutions</b>			
Transitions			
GC -> AT	619 (97.9)	160 (96.4)	79 (56)
AT -> GC	10 (1.6)	2 (1.2)	21 (15)
Transversions			
GC -> CG	2 (0.3)	2 (1.2)	4 (2.9)
GC -> TA	0	2 (1.2)	17 (12)
AT -> TA	1 (0.2)	0	13 (9.3)
AT -> CG	0	0	6 (4.3)
Total	633 (100)	166 (100)	140 (100)
<b>Other mutations</b>			
Frameshift (-1)	1	0	1
Deletion	0	0	28 <sup>d</sup>
Insertion	0	0	1
Complex	0	0	13

(a) Data from Gordon et al., 1990.

(b) Data from Kohalmi and Kunz, 1988.

(c) Data from Yang et al., 1991; Lukash et al., 1991; Yang et al., 1993.

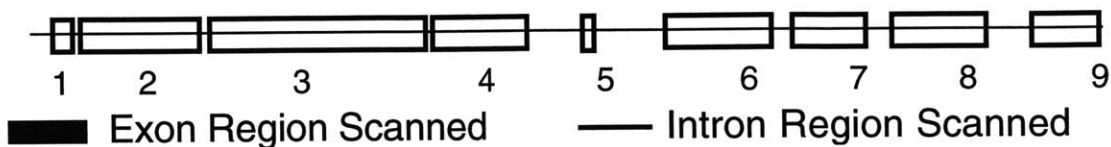
(d) Contains putative splice site mutations

Figure 7. Comparison of spontaneous and MNNG-induced point mutational spectra in the *hprt* gene in human lymphoblastoid B-cell lines TK6 and MT1.

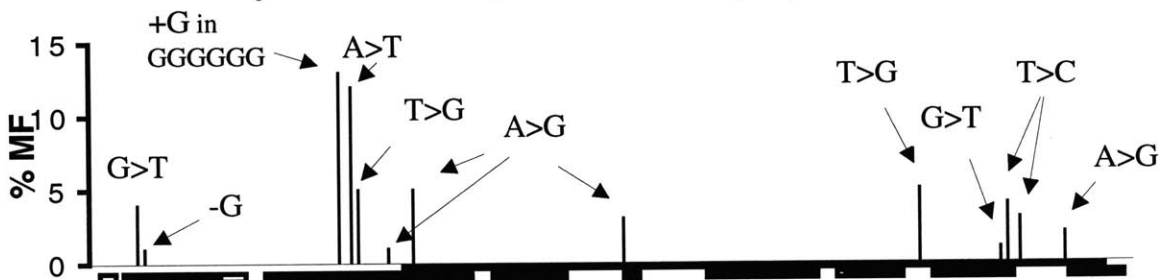
The mutational spectra were obtained through *en masse* selection of 6TG resistant mutants and PCR/DGGE (or CDCE) analysis. The total mutant fraction of the *hprt* gene in the spontaneous MT1, spontaneous TK6, MNNG-treated MT1 (4  $\mu$ M of MNNG, 2x) and MNNG-treated TK6 (0.04  $\mu$ M of MNNG) cultures is  $3.3 \times 10^{-5}$ ,  $1.2 \times 10^{-5}$ ,  $2.7 \times 10^{-2}$  and  $2.5 \times 10^{-5}$ , respectively. The horizontal axis represents the human *hprt* gene with scanned exon and intron regions indicated. The size of each vertical bar represents the frequency of each mutational hotspot relative to the total mutant fraction (MF). The kind of mutation at the corresponding hotspot is shown. "\*" indicates a CpG site. "?" indicates unidentified sequences.

(Kindly provided by Aoy Tomita-Mitchell)

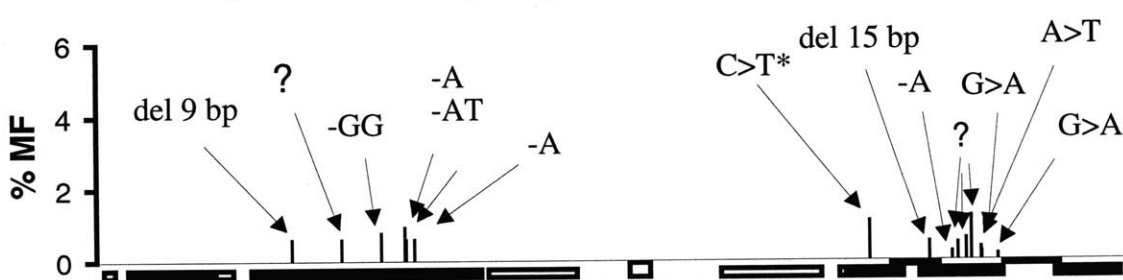
## HPRT Gene



### MT1 (Mismatch repair deficient) Spontaneous

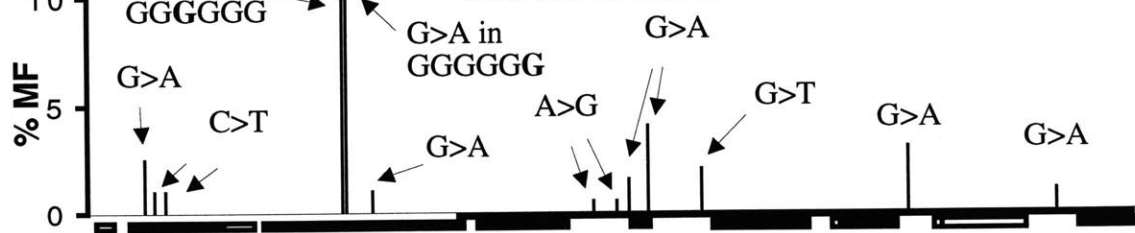


### TK6 (Mismatch repair proficient) Spontaneous



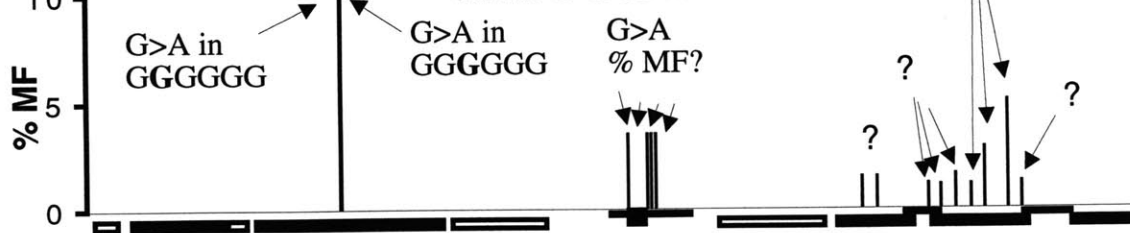
### MT1 (Mismatch repair deficient)

#### MNNG treated



### TK6 (Mismatch repair proficient)

#### MNNG treated



GC basepairs is certainly consistent with the predicted significance of its miscoding potential (Loveless, 1969; Loechler et al., 1984). Among the minor types of MNNG-induced mutations, AT → GC transitions are believed to stem from O<sup>4</sup>-meT lesions (Preston et al., 1986; Richardson et al., 1987b). The AT/GC → TA transversions may be explained by the preferential insertion of adenine through polymerase translesion bypass at abasic or damaged purine sites (Kunkel, 1984; Loeb and Preston, 1986; Hevroni and Livneh, 1988). Other transversions may reflect the use of residues other than adenine during bypass, which may vary from site to site (Randall et al., 1987). Most frameshifts recovered in *lacI* following alkylating treatment consist of the loss of one GC basepair from a contiguous run of GC basepairs (Gordon et al., 1990; Zielenska et al., 1989). Preferential alkylation of N7-guanine has been found to occur within runs of contiguous guanine residues by nitrogen mustards, which correlates with the increased electrostatic potential of these sites (Mattes et al., 1986). It has been postulated that the alkylation damage may promote the formation of stable extrahelical intermediates resulting in template misalignments (Horsfall et al., 1990).

It has been shown that the types and frequencies of mutations induced by MNNG may change depending on the treatment conditions and the cellular background. One study using diploid human fibroblasts showed that at high doses of MNNG, the mutational spectrum of the *hprt* gene was dominated by GC → AT transition, whereas at low doses the AT → GC transition was dominating (Yang et al., 1991). There has been speculation that at low doses, the O<sup>6</sup>-meG adducts can be very efficiently removed by the MGMT making O<sup>4</sup>-meT the major type of premutagenic lesion, while at high doses, the protein is saturated and O<sup>6</sup>-meG becomes dominant (Yang et al., 1991). Another study, however, showed somewhat contradictory results that the level of the MGMT activity did not seem to effect the kinds of mutations induced by MNNG in the *hprt* gene of human diploid fibroblasts (Lukash et al., 1991). An imbalanced intracellular dNTP pool may also effect the types of mutations induced by MNNG. An increased intracellular dCTP:dTTP ratio in

the yeast *S. cerevisiae* was found to reduce the frequency of GC → AT transition but increased that of GC → CG transversion in the *SUP4-o* gene by MNNG (Khalmi et al., 1993). It was postulated that the reduction in GC → AT transition was due to dCTP out competing dTTP for incorporation opposite O<sup>6</sup>-meG, while the increase in GC → CG transversion was due to the increase in the dCTP pool causing error-prone repair of AP sites resulted from N<sup>7</sup>-meG (Khalmi et al., 1993; Kunz et al., 1998).

#### *The site specificity of MNNG-induced transition*

The distribution of GC → AT transitions induced by MNNG in a target sequence is not random. It has been observed that the transitions predominantly occur at guanines preceded by a purine residue (i.e. at 5'-RG-3' sites) in bacteria, yeast and human systems (Burns et al., 1987; Richardson et al., 1987a, b; Reed and Hutchinson, 1987; Kohalmi and Kunz, 1988; Gordon et al., 1990; Yang et al., 1991; Lukash et al., 1991; Kat, 1992). Some studies report that among all of the available 5'-RG-3' sites, there is a preference for 5'-GG-3' sites over 5'-AG-3' sites (Richardson et al., 1987a; Reed and Hutchinson, 1987). But other studies show that both flanking guanines and adenines equally affect the distribution of mutation (Burns et al, 1987; Kohalmi and Kunz, 1988; Gordon et al., 1990). The same 5'-flanking base effect has also been found to influence the distribution of AT → GC transitions (Horsfall et al., 1990).

Based on the premise that GC → AT and AT → GC transitions are the direct result of O<sup>6</sup>-meG and O<sup>4</sup>-meT adducts, several mechanisms have been proposed to explain the observed influence of the 5' flanking base (Burns et al., 1987). The 5' flanking base may influence the mutability of a given guanine by effecting: (1) the distribution of O<sup>6</sup>-meG and O<sup>4</sup>-meT lesions; (2) the efficiency of repair of these lesions; and (3) the fidelity of DNA polymerase bypassing these lesions.

There is evidence to support the notion that the initial deposition of DNA damage is

responsible for the observed site specificity. *In vitro* studies of DNA methylation by MNU show that the frequency of O<sup>6</sup>-guanine methylation depends on the nature of the nucleotides flanking the guanine residue, with a 5' flanking guanine (followed by 5' adenine) providing the most "permissive" condition and a 5' flanking thymine (followed by cytosine) the strongest inhibition (Dolan et al., 1988; Sendowski and Rajewsky, 1991). It has been postulated that the molecular electrostatic potential, an important component of reactivity, of the O<sup>6</sup> position of guanine can be influenced by the adjacent basepairs (Pullman and Pullman, 1981). By quantum analysis, O<sup>6</sup>-guanine exhibits a greater potential and accessibility when preceded by a purine than a pyrimidine (Pullman and Pullman, 1981). However, by comparing the extent of O<sup>6</sup>-meG formation *in vitro* and the mutation frequency *in vivo* at four specific target sites, it was found that the degree of O<sup>6</sup>-meG formation influenced but did not completely account for the distribution of mutagenic hotspots (Richardson et al., 1989).

The MGMT repair can significantly reduce the mutant frequency induced by MNNG (Lukash et al., 1991), and there is some evidence to suggest that this repair mechanism can be influenced by sequence context (Topal et al., 1986). However, the 5'-RG-3' bias has been observed to be independent of the MGMT activity (Richardson et al., 1987a; Kohalmi and Kunz, 1988; Kat, 1992; Schendel and Robins, 1978; Rossi et al., 1989). Therefore, the reparability by MGMT is unlikely to be an important factor in the observed site specificity. The efficiency of nucleotide excision repair of O<sup>6</sup>-meG may also vary from site to site (Jones et al., 1987). In addition, the excinuclease enzyme seems to be able to protect some O<sup>6</sup>-meG lesions from MGMT repair through competitive binding of the lesions and possibly in a sequence-influenced manner (Rossi et al., 1989). One study showed, however, that the 5'-RG-3' site biases of N-ethyl-N-nitrosourea (ENU)-induced transition were independent of the excision repair mechanism (Burns et al., 1988).

Finally, the neighboring base may influence the fidelity of the replication process. Base-stacking interactions at the primer terminus may affect the stability of mispairs, and



subsequently proofreading and processivity of the DNA polymerase (Kunkel et al., 1981). Experimental evidence supporting this mechanism in MNNG mutagenesis has yet to be found.

Several studies including those involving the endogenous *hprt* gene in human cells have reported a strand preference in favor of GC → AT transitions at guanines on the nontranscribed strand (Richardson et al., 1987a; Reed and Hutchinson, 1987; Yang et al., 1991; Lukash et al., 1991; Kat, 1992). But this strand specificity has not been found in other targets (Kohalmi and Kunz, 1988; Gordon et al., 1990). It has been demonstrated that the observed strand bias is independent of both the MGMT activity and the direction of replication (Richardson et al., 1987a; Reed and Hutchinson, 1987; Ito et al., 1994).

Using an assay for forward mutations in the *rpsL* gene in *E. coli*, it was shown that upon very active transcription, MNNG induced GC → AT transitions at guanine residues on the transcribed strand and that the 5'-flanking base of the sites was predominantly thymine; upon decreasing the level of transcription, mutations on the non-transcribed strand were found sharply increased while the 5'-flanking base became predominantly guanine (Ito et al., 1994). Based on the results, it was suggested that the distribution of O<sup>6</sup>-meG might be affected by both the level of transcription and transcription-coupled excision repair (Ito et al., 1994). In contrast to these observations, the transcription level did not seem to affect the strand bias of MNU-induced mutations observed in a gene cloned in shuttle plasmids in human cells (Palombo et al., 1992). However, the target sequence used in this study was an extrachromosomal element, which can not represent endogenous nuclear genes that can be accessed by the DNA repair systems.

To account for the discrepancy in the strand bias of MNNG-induced mutations among the various targets, Gordon and Glickman (1988) argued that the non-random distribution of MNNG-induced mutation in some of the targets could be a reflection of the influence of the mutation on protein structure/function. For example, GC → AT transition in the Gly (5'-GGN-3') and Trp (5'-UGG-3') codon sequences are more likely to be

recovered as a hotspot because Gly residues are often strategic to protein structure and the transition in Trp coding sequence results in translation termination signals. In fact, under non-selective conditions, a GC -> AT mutant, which was not recovered by the phenotypic selection method due to its retained wild-type phenotype, was detected in the *lacI* gene in *E. coli* after treatment with MNNG (Gordon et al., 1990b).

In summary, the observed specificity of MNNG-induced mutational spectra in selectable genes reflects not only the nature and mutagenic consequence of the induced DNA lesions, but also the distribution of the lesions, the influence of cellular repair repertoire and the DNA polymerase fidelity at the lesion sites as well as the effect of mutation on protein function.

### 3. MATERIAL AND METHODS

#### 3.1. Cell culture and MNNG treatment

##### 3.1.1. The human cell lines

Two human B lymphoblastoid cell lines TK6 and MT1 were used in this study. The TK6 cell line is isolated from the HH4 line which is a derivative of the near diploid WI-L2 line originally isolated from a male spherocytosis patient (Skopek et al., 1978). Like its parental lines, TK6 carries one allele of the *hprt* gene which permits its use as a genetic marker in *in vitro* mutation assays using 6TG as the selective agent for mutants. It is also a heterozygote at the *thymidine kinase (tk)* locus (on chromosome 17) which provides a second genetic marker using trifluorothymidine (F<sub>3</sub>TdR) to select for mutants (Skopek et al., 1978).

The MT1 line is a clonal derivative of TK6 treated with ICR-191 and then selected by MNNG (Goldmacher et al., 1986). The phenotypic characteristics of the two cell lines are compared in Table 6. Both cell lines are deficient in MGMT and exhibit similar sensitivities to the mutagenic effect of MNNG. However, MT1 is several hundred times more resistant to the cytotoxicity of MNNG than TK6, and has a spontaneous mutation rate 15 - 30-fold higher than that of TK6 in both the *hprt* and *tk* genes (Goldmacher et al., 1986; Kat et al., 1993).

It has been hypothesized that the distinct phenotypes of MT1 might be related to a defect in its strand-specific mismatch repair system (Goldmacher et al., 1986). *In vitro* biochemical assays demonstrated that the extract of MT1 cells was indeed deficient in strand-specific correction of all eight base-base mismatches (Kat et al., 1993). Later it was found that MT1 cells harbor two missense mutations on separate alleles of the mismatch repair G/T mismatch-binding protein (*GTBP*) gene, including an A -> T transversion at

Table 6. Phenotypic characteristics of TK6 and MT1 cell lines.

Phenotype	TK6	MT1
D <sub>37</sub> (MNNG), 45 min	0.04 $\mu$ M	6 $\mu$ M
Spontaneous mutation rate (6TG <sup>R</sup> )	$1.3 \times 10^{-7}$ mutations/day	$5.8 \times 10^{-6}$ mutations/day
MNNG-induced mutation (6TG <sup>R</sup> , 45 min)	$5.7 \times 10^{-4}$ mutations/cell $\cdot\mu$ M MNNG	$12.5 \times 10^{-4}$ mutations/cell $\cdot\mu$ M MNNG
Absolute clone-forming ability (low density, no feeder cells)	$0.36 \pm 0.07$	$0.014 \pm 0.003$
Initial survivor doubling time after MNNG treatment	18h (unchanged from untreated cells)	$\geq 36$ h (increased from 20 h in untreated cells)
Initial formation of adducts after 8 $\mu$ M MNNG, 45-min treatment	adducts/cell $\times 10^{-4}$	adducts/cell $\times 10^{-4}$
O <sup>6</sup> MeG	8	8
3MeG	24	24
3MeA	5	5
7MeG	125	125
7MeA	3	3
Half-life of adducts		
O <sup>6</sup> MeG	20 h	20 h
3MeG	19 h	17 h
3MeA	$\leq 1$ h	$\leq 1$ h
7MeG	$\sim 17$ h	$\sim 17$ h
7MeA	$\sim 13$ h	$\sim 13$ h
Initial cell cycle progression effect (one generation time, lethal dose)	No effect after treatment with 0.14 $\mu$ M MNNG, 45 min	Marked decrease in S phase fraction after 10 $\mu$ M MNNG, 45 min
Cell cycle block (three generations)	Marked S and G2 + mitosis accumulation	Return to $\sim$ normal distribution
Hydroxyurea-insensitive [ <sup>3</sup> H]thymine uptake into DNA	7-Fold over background (1-h post-treatment)	7-Fold over background (1-h post-treatment)

(Adapted from Goldmacher et al., 1986)

codon 1145 and a G -> A transition at codon 1192 (Papadopoulos et al., 1995). Neither of these mutations are found in TK6 cells (Papadopoulos et al., 1995).

### 3.1.2. Cell culture and MNNG treatment

Both TK6 and MT1 cells were grown in spinner cultures in RPMI 1640 medium supplemented with 10% horse serum (Life Technologies, Gaithersburg, MD) in 37°C humidified incubators with a 5% CO<sub>2</sub> atmosphere. Cells were maintained in exponential growth by daily dilution to 4 x 10<sup>5</sup> cells/ml with the medium. The number of cells was counted on a model B Coulter counter (Coulter Electronics, Hialeah, FL). To freeze down cells, cells were resuspended in fresh medium containing 10% dimethyl sulfoxide (DMSO) at 5 x 10<sup>7</sup> cells/ml and stored immediately at -70°C.

Two cultures of 8 x 10<sup>7</sup> MT1 cells (4 x 10<sup>5</sup> cells/ml) were independently treated once with 4 µM MNNG (Sigma Chemical, St. Louis, MO) for 45 min (Kat, 1992). After treatment, the cells were resuspended in fresh medium and grown for 30 generations with daily dilution to diminish any induced premutagenic lesions which could be mistaken for mutations in the procedure. Untreated MT1 cells grown in parallel served as controls.

### 3.1.3. Determination of the *hprt* gene mutant fractions

Mutant fractions of the *hprt* gene in MT1 cells were measured using a microtiter plate mutation assay (Furth et al., 1981). Cells were properly diluted with fresh medium and plated in 96-well microtiter plates in the presence and absence of 6TG. For the determination of plating efficiency (PE), cells were plated at 2 cells/well in four plates without 6TG. For the determination of *HPRT* mutant fraction, cells were plated at 400 cells/well for MNNG-treated MT1 cells or 40,000 cells/well for untreated MT1 cells in sixteen plates in the presence of 1 µg/ml 6TG. The plates were then kept in 37°C

humidified incubators with a 5% CO<sub>2</sub> atmosphere for 14 - 18 days until visible colonies formed. The number of wells with no colonies in each plate was counted.

The distribution of colonies among wells is considered a Poisson distribution. The average number of colony forming units per well ( $\lambda$ ) can be calculated from  $P(0) = e^{-\lambda}$ , where  $P(0)$  is the probability of any well having no colonies, which can be approximated by the fraction of all wells observed that have no colonies (# wells with no colonies / # wells scored).

The PE of a culture is defined as  $\lambda$  divided by the average number of cells plated per well in the absence of 6TG:

$$PE = \frac{\lambda}{\# \text{ cells per well}} = \frac{-\ln P(0)}{\# \text{ cells per well}} = \frac{-\ln \frac{\# \text{ wells with no colonies}}{\# \text{ wells scored}}}{\# \text{ cells per well}}$$

Only 6TG resistant (6TG<sup>R</sup>) HPRT mutants can form colonies when plated in the presence of 6TG. The *hprt* mutant fraction is determined as:

$$MF = \frac{-\ln \frac{\# \text{ 6TG wells with no colonies}}{\# \text{ 6TG wells scored}}}{\# \text{ cells per 6TG well} \times PE}$$

### 3.2. Isolation of genomic DNA from cultured cells, blood and solid tissues

For cultured cells, the sample was centrifuged at 2500 rpm for 5 min. The supernatant was discarded and cells were resuspended in 1 x TE buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA) at 10<sup>6</sup> - 10<sup>7</sup> cells per ml. For thawed blood, one volume of phosphate-buffered saline (PBS) solution was added to the sample. The sample was centrifuged at 6000 g for 15 min. The supernatant was discarded, and the pellet containing white blood cells was resuspended in PBS and centrifuged again. The procedure was repeated twice. The pellet was finally resuspended in 1 x TE at ~10<sup>6</sup> white blood cells

(equivalent to 1 ml of blood) per ml. For solid tissue biopsy sample, the tissue was cut into small pieces with a scalpel and deeply frozen in liquid nitrogen. Using a mortar and pestle, the tissue pieces were homogenized to a fine powder as thoroughly as possible. The powder was suspended in 1 x TE at 50 mg of tissue per ml.

To the above cell or tissue suspension was then added Proteinase K (20 mg/ml) and sodium dodecyl sulfate (SDS) (10%) to a final concentration of 1 mg/ml and 0.5%, respectively. The sample was mixed by vortexing and incubated in a 50°C water-bath shaker shaking at 200 min<sup>-1</sup> for 3 hr. RNaseA (10 mg/ml) was added to the sample at a final concentration of 0.1 mg/ml. After mixing the sample was incubated in the water-bath shaker at 37°C shaking at 200 min<sup>-1</sup> for 1 hr.

After incubation, the sample was centrifuged at 13,000 rpm at 4°C for 15 min. The central portion of the supernatant was transferred to a new tube using a Pasteur pipette. Care was taken to avoid the bottom pellet and the top layers of white matter (sometimes fat) in tissue samples. The above procedure was repeated two to three times to collect as much of the supernatant as possible.

To the collected supernatant was added 5 M NaCl to a final concentration of 250 mM and two volumes of 100% ethanol. The sample was mixed gently by inverting the tube to form the DNA spool. After it became compact and floated on the surface of the supernatant, the spool was transferred to a new tube and washed twice with cold 70% ethanol. The DNA was air dried for about 30 min. 0.1 x TE was added to a concentration of 2 - 4 mg DNA per ml. The sample was then vortexed vigorously until the DNA was completely dissolved.

The concentration and purity of the isolated genomic DNA was assessed by absorbance (A) measurements at 260 and 280 nm wavelengths. The concentration of the DNA sample was calculated as:  $A_{260} \times 50 \mu\text{g/ml} \times \text{dilution factor}$ . The ratio of  $A_{260}/A_{280}$  was used as an indicator of DNA purity.

The genomic DNA was digested with endonucleases *Hae* III and *Xba* I (New

England Biolabs, Beverly, MA) at a DNA concentration of 2 - 3 mg/ml and an enzyme/DNA ratio of 1 U/ $\mu$ g at 37°C overnight.

### **3.3. Enrichment of the desired DNA fragments from genomic DNA**

#### **3.3.1. Size fractionation by polyacrylamide gel electrophoresis (PAGE)**

Preparative PAGE gels (17 cm x 16 cm x 1.5 mm) contained 6.4% acrylamide and 1/40 bis-acrylamide in 1 x TBE (89 mM Tris-borate, pH 8.3, 1 mM EDTA). The gels were polymerized with 0.1% ammonium persulfate and 0.1% N,N,N',N'-tetramethylethylenediamine (TEMED). Up to 1 mg of genomic DNA digested with *Hae* III and *Xba* I was loaded in the middle preparative lane. About 0.5  $\mu$ g of *pBR322* DNA digested with *Msp* I endonuclease was loaded on the two side marker lanes. Electrophoresis was performed at a constant voltage of 250 V for 1 hr. After electrophoresis, the two side marker lanes were cut out from the gel and stained in ethidium bromide solution (2  $\mu$ g/ml in 1 x TBE) for a few minutes. The *pBR322/Msp I* marker bands were visualized under a UV transilluminator. The position corresponding to the size of the desired genomic DNA fragment was marked by making a notch on the edge of the gel segments. The marked gel segments were then put back to their original positions in the PAGE gel. Based on the position of the two notches, the gel slice containing the desired genomic DNA fragment in the middle lane was cut out.

DNA was electroeluted from the gel slice using a D-gel apparatus (Epigene Corp., Baltimore, MD). Briefly, the gel slices were cut into small pieces and put on top of the electroelution columns pre-packed with DEAE cellulose in 1 x TBE. Electroelution was performed at 125 V for 2 hr. The columns of DEAE cellulose were then drained completely of the TBE buffer, and washed twice with 4.5 ml of 0.2 M NaCl. The NaCl solution was drained, and DNA was eluted from the DEAE cellulose twice with 250  $\mu$ l of 1 M NaCl.



The DNA in the eluate was precipitated with two volume of 100% ethanol at  $-20^{\circ}\text{C}$  overnight. The sample was centrifuged at 14,000 rpm for 10 min. The DNA pellet was washed twice with cold 70% ethanol, air dried and dissolved in 5  $\mu\text{l}$  of ddH<sub>2</sub>O.

### 3.3.2. Sequence-specific enrichment

To the genomic DNA digested with *Hae* III and *Xba* I were added 20 x SSPE and 10% SDS to reach a final concentration of 6 x SSPE (1.08 M NaCl, 60 mM sodium phosphate, pH 7.4, 6 mM EDTA) and 0.1% SDS, and the two 5'-biotinylated 30-mer probes BP1 (5'-CAA AACTGACAGCACAGAATCCAGTGG AAC) (*APC* cDNA bp 8472-8501) and BP2 (5'-AAGACCCAGAATGGCGCTTAGGACTTTGGG) (complementary to *APC* cDNA bp 8501-8530) (Synthetic Genetics, San Diego, CA) to a probe/target molar ratio of  $5 \times 10^4$ . The mixture was dispensed into several 1.5-ml microcentrifuge tubes (~ 0.8 ml / tube), heated in boiling water for 2 min and immediately chilled in an ice bath for 10 min. The sample tubes were incubated at  $58^{\circ}\text{C}$  for 2 hr.

Streptavidin-coated glass paramagnetic beads (CPG, Lincoln Park, NJ) (0.4 mg beads per  $10^8$  copies of the target fragment) were prewashed twice with the washing buffer (1M NaCl, 10 mM Tris-HCl pH 7.6, 2 mM EDTA). The beads were then dispensed into each of the sample tubes and kept suspended in the DNA solution at  $50^{\circ}\text{C}$  for 1 hr using a temperature-controlled rotary thermomixer (Brinkmann Instruments, Westbury, NY) at  $1000 \text{ min}^{-1}$ . After incubation, the beads were magnetically collected to the side of the sample tube by placing a neodymium magnet against the wall of the tube. The solution was removed and all the beads belonging to the same sample were combined.

The beads were washed four times with the washing buffer. At each time the beads were resuspended in the washing buffer at 10 mg beads/ml in the rotating thermomixer at  $50^{\circ}\text{C}$ ,  $1000 \text{ min}^{-1}$  for 5 min. DNA was eluted twice by incubating the beads with ddH<sub>2</sub>O at 20 mg/ml at  $70^{\circ}\text{C}$  for 2 min. The eluate was then concentrated to about 10  $\mu\text{l}$  through

speed-vac centrifugation. 0.1 volume of 10 x reannealing buffer (2 M NaCl, 100 mM Tris-HCl, pH 7.6, 20 mM EDTA) was added to the sample. The sample was then reannealed at 55°C for 16 hr.

The reannealed sample was digested with endonucleases *Sau3A* I and *Acc* I (New England Biolabs) in a 100 - 300  $\mu$ l reaction. The digested sample was desalted and concentrated through ultrafiltration using Microcon-50 devices (Amicon, Beverly, MA) (see Section 3.7.2.). The volume of the desalted samples (20 to 50  $\mu$ l) was further concentrated to 4 - 10  $\mu$ l by speed-vac centrifugation.

### 3.4. High-fidelity PCR

High-fidelity PCR was performed in 10 - 50- $\mu$ l glass capillary tubes using native *Pyrococcus furiosus* (*Pfu*) DNA polymerase (Stratagene, La Jolla, CA) and an Air Thermo-Cycler™ (Idaho Technology, Idaho Falls, ID).

The PCR mixture contained 20 mM Tris-HCl (pH 8.0), 2 mM MgCl<sub>2</sub>, 10 mM KCl, 6 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.1% Triton X-100, 0.1 mg/ml BSA, 0.1 mM dNTPs, 0.2  $\mu$ M each primer, and 0.1 U/ $\mu$ l *Pfu*. The primers used to amplify a 206-bp mitochondrial DNA (mitochondrial DNA bp 10011-10215) are CW7 (5'-ACCGTAACTTCCAATTAAC) (mitochondrial DNA bp 10011-10031) and 5'-fluorescein labeled J3 (5'-GCGGGCGCAG GGAAAGAGGT) (complementary to mitochondrial DNA bp 10196-10215). The primers for amplification of a 243-bp *APC* sequence (*APC* cDNA bp 8441-8683) were 5'-fluorescein labeled AP1 (5'-GAATAACAACACAAAGAAGC) (*APC* cDNA bp 8441-8460) and AP4H (5'-AACAAAAACCCTCTAACAAG) (complementary to *APC* cDNA bp 8664-8683). The PCR mixture was siphoned into a glass capillary tube (Idaho Technology) and the two ends of the tube were then sealed in a flame of a gas burner. The capillary tube was placed in an air thermocycler.

The reaction conditions were as follows: 94°C for 2 min; appropriate cycles, each

cycle with ~10 sec at 94°C for denaturation, ~20 sec at 57°C (for the mitochondrial sequence) or 50°C (for the *APC* sequence) for annealing, ~20 sec at 72°C for extension; and 72°C for 2 min. To create mutant/wild-type heteroduplexes, a sufficient number of PCR cycles was performed in order to deplete most of the primers so that the mutants could reanneal with excess wild-type DNA during the final cycles. To create homoduplexes, aliquots of PCR products were subjected to only 3 - 6 cycles of PCR so that excess primers were still present in the final reactions. After the reaction, the two ends of the capillary tube was cut open with a glass cutter. The PCR reaction was then transferred into a 0.5-ml microcentrifuge tube.

Post-PCR incubations were performed to remove interfering PCR byproducts. The products of the mitochondrial DNA were incubated at 45°C for 30 min and the products of the *APC* sequence were incubated with additional *Pfu* (half of the amount used in the PCR reaction) at 72°C for 5 min.

### **3.5. Constant denaturant gel electrophoresis**

CDGE gels (17 cm x 16 cm x 0.4 mm) contained 7.5% acrylamide and 1/40 bis-acrylamide in 1 x TAE (40 mM Tris-acetate, 1 mM EDTA, pH 8.0). The gels were polymerized with 0.1% ammonium persulfate and 0.1% TEMED. The gels were submerged in a 1 x TAE buffer tank which was pre-heated to a precise temperature around 70°C. After the samples were loaded, electrophoresis was performed at a constant voltage of 160 V for 90 min. Extensive circulation of the buffer was provided by a constant temperature circulator during the runs.

Test runs were performed to determine the optimal separation temperatures using PCR amplified *APC* gene wild-type and artificial mutant fragments and fluorescein-labeled DNA markers which were PCR amplified *hprt* gene exon 3 wild-type and mutant homoduplexes and heteroduplexes. After electrophoresis, the gels were stained in ethidium

bromide solution (2  $\mu\text{g/ml}$  in 1 x TAE) for a few minutes, and photographed using a UV transilluminator.

Separation of cellular DNA was performed at the determined optimal conditions. After electrophoresis, the fluorescein-labeled DNA markers were visualized by illuminating the gel with the 488 nm argon laser. Each marker band was marked on the plate while viewing the gel through a 520 nm low pass glass filter (Oriel, Stratford, CT). Based on the results of test runs, the relative positions of the markers provided information on the exact position to which the desired cellular *APC* mutants migrated in the gel. The portions of the gel below and above the wild-type DNA marker band were cut so as to include the range of mobilities of most mutant fragments.

The gel slices were finely ground between two glass slides and transferred into 1.5 ml microcentrifuge tubes containing 200  $\mu\text{l}$  of 200 mM NaCl. DNA was eluted from the gel by shaking for 20 min at 50°C, 1300  $\text{min}^{-1}$ . The tubes were centrifuged at 14,000 rpm for 10 min to spin down gel particles. The supernatant was collected. DNA was then ethanol precipitated from the supernatant, washed with 70% ethanol and dissolved in 10  $\mu\text{l}$  of ddH<sub>2</sub>O.

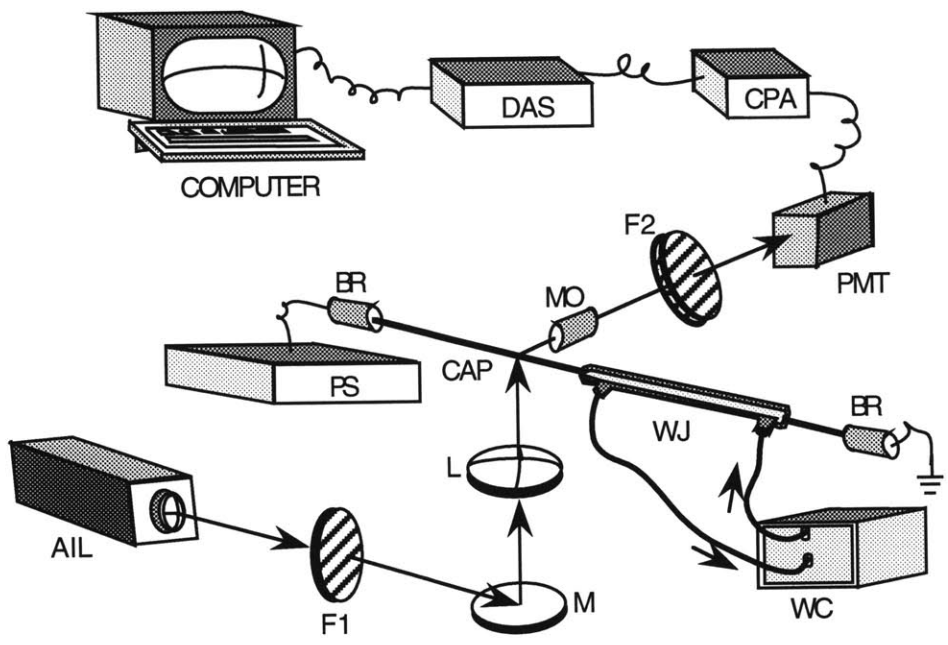
### **3.6. Constant denaturant capillary electrophoresis**

#### **3.6.1. The CDCE instrument**

The diagram of the CDCE apparatus is shown in Fig. 8. Adapted from the CE instrument for separation of DNA sequencing products (Cohen et al., 1990), the key element of CDCE is a heated zone of capillary positioned before the detection window where DNA separation takes place (Khrapko, et al., 1994). The temperature of the heated zone can be controlled by a surrounding water jacket connected to a constant temperature circulator.

Figure 8. Diagram of the CDCE apparatus.

The power supply (PS) is 30 kV DC (Model CZE 1000R-2032, Spellman, Hauppauge, NY). Laser beam from an argon ion laser (AIL) (Model 5425ASL, Ion Laser Technology, Salt Lake City, UT) is filtered through a 488 nm narrow band-pass filter (F1) (10 nm bandwidth) (Corion, Franklin, MA), reflected by a 45° mirror (M) (Newport, Irvine, CA), and focused by a plano-convex glass lens (L) (Newport) onto a detection window of a horizontally positioned fused silica capillary (CAP). Fluorescence emitted from the sample is collected by a 60 x microscope objective (MO) (Newport) and passed through a 540 nm and a 530 nm filter (F2) (Corion) into a photomultiplier (PMT) (Oriel Instruments, Stratford, CT). The signal from the photomultiplier is amplified ( $10^7$  or  $10^8$  V/A) by a current preamplifier (CPA) (Oriel), recorded by a data acquisition system (DAS) (ACM2-16-8A/T51B, Strawberry Tree Computers, Sunnyvale, CA), and transmitted to a Macintosh computer. The buffer reservoirs (BR) are positioned at the two ends of the capillary. At the cathodic end, the capillary is inserted into a water jacket (WJ) the temperature of which is controlled by a constant temperature water circulator (WC).



### 3.6.2. Coating of capillaries

The inner surfaces of fused silica capillaries (Polymicro Technologies, Phoenix, AZ) were coated with 6% linear polyacrylamide via siloxane bonding to prevent electroosmotic flow (Hjertén, 1985). The capillary was treated with 1 M NaOH for 1 hr, washed with dd H<sub>2</sub>O and filled with 1 M HCl for 10 min. The capillary was then washed with 100% methanol, and treated with  $\gamma$ -methacryloxypropyltrimethoxysilane (Sigma, St Louis, MO) overnight. After being washed with 100% methanol, the capillary was filled with a polymerizing solution consisting of 6% acrylamide in 1 x TBE, 0.1% TEMED and 0.025% ammonium persulfate. The capillary was left overnight before use.

### 3.6.3. Preparation of linear polyacrylamide matrices

The linear polyacrylamide matrix used in CDCE was prepared in large quantities and stored at 0 - 4°C (Ruiz-Martinez, et al., 1993; Hanekamp, et al., 1996). The matrix was dispensed into 100- $\mu$ l high pressure syringes to be injected into capillaries before each run. The matrix was prepared as follows.

In a 250-ml flask, 150 ml of 5% acrylamide solution was prepared in 1 x TBE. 45  $\mu$ l of TEMED was then added. The flask was placed in an ice bath on a stir plate and the solution was stirred by a stir bar. The top of the flask was covered with at least 10 layers of Parafilm. Two small holes were punched through the Parafilm with a needle. A long stainless steel needle was inserted into the solution through one of the holes. The needle was connected to an argon cylinder with silicone tubing. Argon gas was then drawn into the acrylamide solution and kept bubbling for 1 hr to deoxygenate the solution.

The tip of the long needle was lifted above the surface of the solution to stop bubbling but maintain the argon pressure in the flask. 45  $\mu$ l of 10% ammonium persulfate was then added into the solution using a 100- $\mu$ l high pressure syringe. The polymerizing

solution was immediately dispensed into several pre-chilled 10 - 50 ml glass syringes. To fill a syringe, a pipetting needle was attached to the syringe and inserted into the flask. The syringe was rinsed twice with argon gas inside the flask and once with small amounts of the polymerizing acrylamide solution. The syringe was then filled with the solution. The pipetting needle on the syringe was replaced with a thin stainless steel needle. The end of the needle was sealed by inserting it into the wall of a short piece of silicone tubing. The syringe was placed in vertical position (needle tip down) at 0 - 4°C while the matrix polymerized.

#### 3.6.4. CDCE conditions, sample loading and fraction collection

Coated capillaries were cut 20 to 33 cm in length with the detection window opened 7 cm from the capillary outlet (the anodic end). DNA samples were introduced into each capillary electrokinetically. Separations of mutant/wild-type heteroduplexes from wild-type sequences in the target-enriched cellular DNA were performed in a 20 cm long, 540  $\mu\text{m}$  i.d. capillary inserted into a 6 cm water jacket at  $\sim 66^\circ\text{C}$ . The desalted sample (about 4  $\mu\text{l}$ ) was injected into a short piece (about 1 cm) of Teflon tubing and tightly mounted onto the capillary inlet. The capillary inlet with attached tubing was then inserted into the inlet buffer reservoir and an electric current was applied at 80  $\mu\text{A}$  for 2 minutes. This process achieved complete loading of the injected sample. The tubing was removed after sample injection, and the capillary inlet was inserted back into the buffer reservoir. Electrophoresis was performed at a constant current of 80  $\mu\text{A}$ . The desired fractions were electro-eluted into 10  $\mu\text{l}$  of 0.8 x TBEB (0.8 x TBE, 0.24 mg/ml) for 10 to 20 min per fraction. The collected fractions were desalted through drop dialysis (see Section 3.6.5) and concentrated before being reloaded onto the capillary for a second separation.

Separations of PCR products were performed in 75  $\mu\text{m}$  i.d. capillaries. 0.5  $\mu\text{l}$  of product solution was diluted to 5  $\mu\text{l}$  with water to reduce the ionic strength of the loading



solution. The inlet buffer reservoir was replaced by the sample vial and an electric field was applied at 2  $\mu\text{A}$  for 30 sec to 2 min. After injection, the capillary inlet was inserted back into the buffer reservoir and electrophoresis was performed. For separation of heteroduplexes from the wild-type homoduplex, CDCE was performed at a constant current of 9  $\mu\text{A}$  in a 21-cm capillary inserted into a 6-cm water-jacket at 64.6°C. The desired heteroduplex fraction was collected into 10  $\mu\text{l}$  of 0.1 x TBEB for about 3 min. For separation of the wild-type and mutant homoduplexes, CDCE was performed at 8  $\mu\text{A}$  in a 33 cm capillary inserted into a 19 cm water jacket at ~65.2°C. The linear polyacrylamide inside the capillary was replaced for each run.

#### 3.6.5. Methods for desalting DNA samples for CE separation

##### *Drop dialysis*

Drop dialysis is a simple and rapid method for dialysis of small-volume samples (< 50  $\mu\text{l}$ ) with a high recovery (Marusyk and Sergeant, 1980). A membrane filter of pore size 0.05  $\mu\text{m}$  in diameter (Millipore, Bedford, MA) was floated on the surface of dialysis buffer containing 0.01 - 0.1 x TBE in a small shallow container (such as a plastic weight boat) placed on a stir plate. A 5 to 50- $\mu\text{l}$  drop of the sample was carefully pipetted on the floating membrane filter. The sample was dialyzed for 1 to 3 hr while the dialysis buffer was slowly stirred by a small stir bar. After dialysis, the droplet was recovered from the membrane with a micropipet.

##### *Ultrafiltration*

Ultrafiltration utilizes centrifugation to force liquid and small molecules (salts, nucleotides, DNA oligo primers, etc.) through a low-binding membrane with an

appropriate molecular weight cut-off. Solutes larger than the nominal cut-off (such as protein and DNA) are retained on top of the membrane. Through repeated centrifugation followed by reconstitution with water, ultrafiltration can be used to efficiently concentrate and desalt samples. In this study, a disposable ultrafiltration device, Microcon-50 (50,000 Dalton MW cut-off) (Millipore) was used to desalt and concentrate 100 - 500  $\mu$ l of restriction digested DNA samples. This method was also used to desalt and remove residual primers from PCR amplified DNA fragments (20 - 50  $\mu$ l) for sequencing.

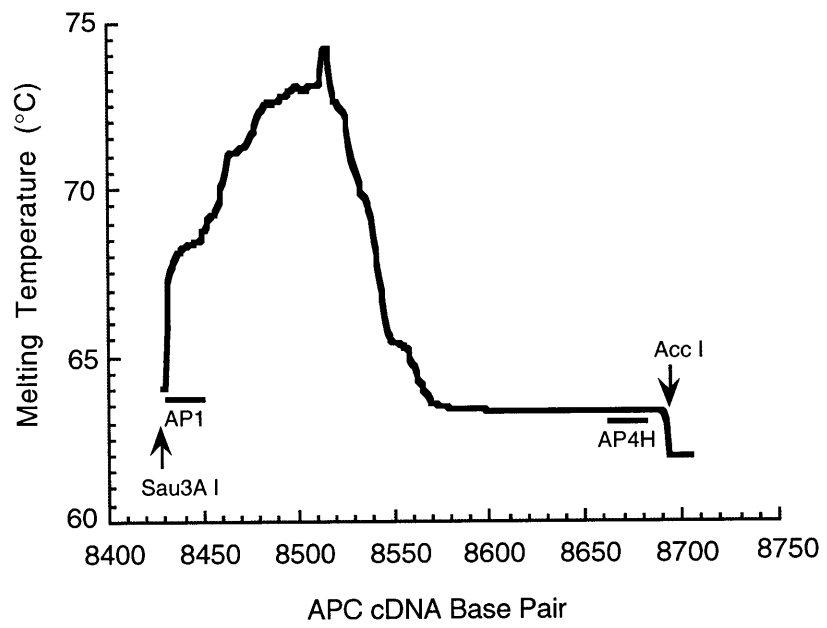
The protocol was as follows. ddH<sub>2</sub>O was added to the sample to a total volume of 500  $\mu$ l. The sample was then pipetted into a Microcon sample reservoir inserted into a vial. The assembly was centrifuged at 14,000 x g for 6 - 7 min. The eluate in the vial was discarded. 400 - 490  $\mu$ l of ddH<sub>2</sub>O was added to the sample reservoir, and the assembly was centrifuged once more at 14,000 x g for 6 - 7 min. This process (called diafiltration) was usually repeated again for desalting restriction digested DNA samples. The sample reservoir was then placed upside down in a new vial, and spun 3 min at 1000 x g to transfer the sample (20 - 40  $\mu$ l) into the vial. If the sample was genomic DNA, it was further concentrated to 4 - 10  $\mu$ l through speed-vac centrifugation.

### **3.7. The choice of the *APC* gene target sequence**

Single-copy nuclear gene fragments suitable for this study should (1) contain two contiguous isomelting domains differing by at least 5°C, a feature permitting their facile analysis by CDCE; and (2) be flanked by endonuclease recognition sites which are not found within the fragments, a feature allowing their release from genomic DNA. As an initial sequence for developmental work, a 271-bp human *APC* gene fragment (*APC* cDNA bp 8434-8704) was chosen for this study. This fragment is composed of a high and low melting domain with a temperature differential of 11°C (Fig. 9). It can be excised from human genomic DNA by endonucleases *Sau3A* I and *Acc* I which recognizes *APC* cDNA

Figure 9. Melting map of the *APC* gene cDNA bp 8434-8704 fragment.

The melting temperature was calculated according to the melting algorithm of Lerman and Silverstein (1987). The positions of the recognition sites of endonucleases *Sau3A* I and *Acc* I are indicated by arrows. The positions of the primers (AP1 and AP4H) used in high-fidelity PCR to amplify this fragment are indicated by bars.



bp 8430-8433 and 8703-8708, respectively. The target sequence where point mutations can be detected comprises 121 base pairs (*APC* cDNA bp 8543-8663) within the low melting domain of the chosen fragment (see Section 4.1).

The human *APC* gene is localized on chromosome 5q21-22. It contains fifteen exons and an 8538-bp open reading frame (Groden et al., 1991). The chosen 121-bp *APC* target sequence is located in exon 15. It is transcribed but not translated *in vivo*. Although *APC* gene mutations have been found in the majority of colorectal cancers as well as in dysplasia of colon mucosa (Nagase and Nakamura, 1993; Powell et al., 1992), no mutation in this particular 121-bp sequence has been reported.

### 3.8. Construction of artificial mutants

To evaluate the CDCE behavior of the chosen *APC* gene fragment, two artificial point mutants were created through PCR amplification of TK6 genomic DNA with the fluorescein-labeled high melting domain primer AP1 (*APC* cDNA bp 8441-8460) and two low melting domain primers (PmutL and PmutH) each containing a mismatched base with the wild-type sequence. The two artificial mutants, mutL and mutH, were designed to carry a G:C → A:T and an A:T → G:C transition at *APC* cDNA bp 8660 and 8652, respectively. The sequences of the two low melting domain primers are as follows:

PmutL (complementary to *APC* cDNA bp 8656-8700):

5'-TCAAATATGGCTTCCAGAACAAAAACCCTCTAACAAGAATTA AAC

where the underlined T forms a G:T mismatch at bp 8660 with the wild-type sequence;

PmutH (complementary to *APC* cDNA bp 8648-8683),

5'-AACAAAAACCCTCTAACAAGAATCAAACCTACTTAC,

where the underlined C forms a A:C mismatch at bp 8652 with the wild-type sequence.

A 492-bp artificial mutant (*APC* cDNA bp 8422-8913) carrying an AT → GC transition at *APC* cDNA bp 8652 was created to be used as an internal standard mutant

compatible with the 482-bp *APC* gene fragment digested from genomic DNA by *Hae* III and *Xba* I (see Section 4.2). To construct this mutant, a 262-bp *APC* fragment (*APC* cDNA bp 8422-8683) containing an AT → GC transition at bp 8652 was first produced by PCR amplification of TK6 genomic DNA. The PCR primers were interior fluorescein-labeled primer IS1 (5'-CCATCTCAG**GATCC**CAACTCC) (*APC* cDNA bp 8422-8441; the *Sau*3A I recognition site underlined and the fluorescein-labeled thymine residue in bold) and primer PmutH containing the desired A:C mismatch with the wild-type DNA. Also, a 492-bp wild-type *APC* fragment was PCR amplified using primers IS1 and IS2 (5'-TATAATCTAGAAATGATTGA) (complementary to *APC* cDNA bp 8894-8913). The 262-bp mutant fragments were then mixed in a large excess with the 492-bp wild-type fragments. The mixture was amplified using IS1 and IS2 to produce the 492-bp mutant internal standard.

### **3.9. Measurement of the copy number of the target sequence and the efficiency of restriction digestion**

To measure the target copy number in a DNA sample, a small aliquot of the sample was mixed with known copies ( $N_m$ ) of the artificial mutant and subjected to PCR using primers AP1 and AP4H. The PCR products were then separated by CDCE. The initial target copy number ( $N_t$ ) in the sample was calculated based on measurement of the area under the wild-type ( $A_w$ ), mutant ( $A_m$ ) and the average of the two heteroduplex ( $A_h$ ) peaks:

$$N_t = N_m * (A_w + A_h) / (A_m + A_h).$$

The internal standard mutant was also used to measure the efficiency of the *Sau* 3AI and *Acc* I digestion that release the desired 271-bp *APC* gene fragment from genomic DNA. Poor digestion efficiency could occur as a result of incomplete DNA reannealing or enzyme dysfunction due to the amounts of salts and EDTA carried over from the previous target enrichment procedures (see Section 4.4). It was therefore necessary to check

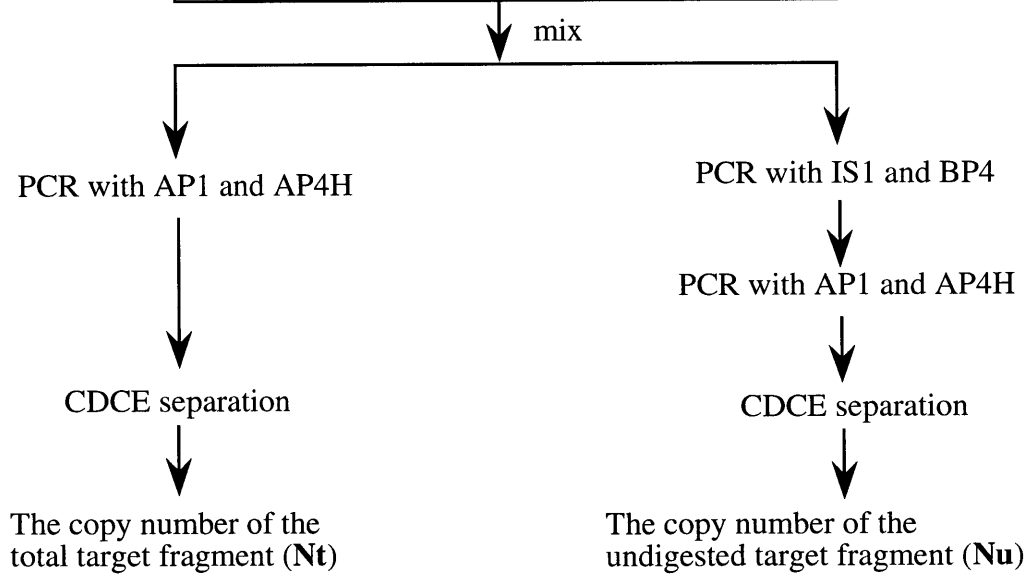
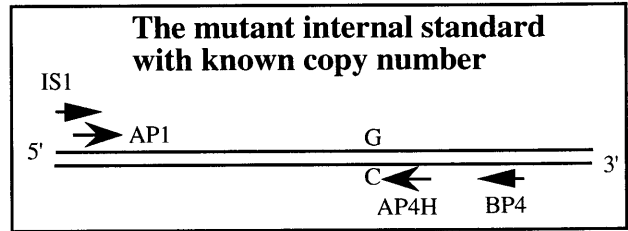
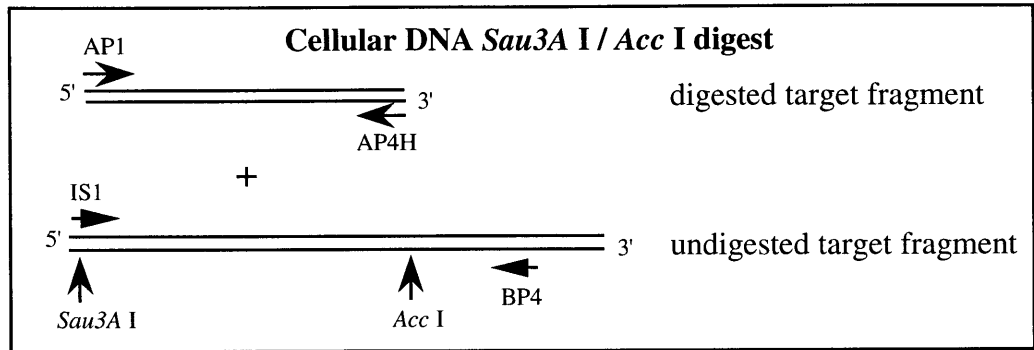
whether the digestion was complete (over 90%) before proceeding to the subsequent steps for enriching the mutant sequences.

Fig. 10 illustrates the method designed to quantitatively measure the *Sau* 3A I and *Acc* I digestion efficiency. A small aliquot of the *Sau*3A I/*Acc* I digested sample was mixed with known copies of the internal standard mutant. One half of the mixture was subjected to PCR using primers AP1 and AP4H by which the mutant internal standard and both the digested and undigested target fragments could be amplified. The PCR products were separated by CDCE and the total target copy number ( $N_t$ ) in the digestion was measured. The other half of the mixture was amplified using primers IS1 and BP4 (5'-GGTGCTACTTTAAATACATG) (*APC* cDNA bp 8787-8796) by which only the undigested target fragments and the mutant internal standard could be amplified. The copy number ( $N_u$ ) of the undigested target fragments was determined by amplifying a small aliquot of the 375-bp PCR products using primers AP1 and AP4H followed by CDCE separation. The *Sau*3A I/*Acc* I digestion efficiency was calculated as  $(N_t - N_u) / N_t$ .

Figure 10. Quantitative measurement of the *Sau3A* I and *Acc* I digestion efficiency.

See text for details.





**The digestion efficiency =  $(Nt - Nu) / Nt$**

## 4. RESULTS AND DISCUSSION

### 4.1. Optimization of the CDCE/high-fidelity PCR conditions

Optimization of the PCR conditions for the chosen 271-bp *APC* gene fragment (*APC* cDNA bp 8434-8704) involved three steps: (1) a suitable position of the low melting domain primer was determined to facilitate CDCE analysis of the amplified fragments; (2) appropriate PCR cycling conditions were determined to obtain an optimal amplification efficiency; and (3) a post-PCR incubation procedure was devised to remove interfering PCR byproducts.

The size of the low melting domain was found to be a critical factor in determining the separation efficiency of the amplified *APC* gene fragments by CDCE. Initially, the position of the low melting domain primer was chosen to be complementary to *APC* cDNA bp 8681-8700 (AP2). Amplification of human genomic DNA using primers AP1 and AP2 produced a 260-bp *APC* gene fragment containing a 157-bp low melting domain. When subjected to CDCE, this fragment appeared as a very broad peak, and the resolution of the wild-type and mutL peaks was poor. Although it was found that the CDCE resolution could be improved by lowering the running current (2  $\mu$ A instead of 9  $\mu$ A) or using longer water-jacket (28 cm instead of 6 cm), the extended separation time (1-2 hr instead of 25 min) at these conditions was prohibitive for routine operations. A new primer AP4H (complementary to *APC* cDNA bp 8654-8683) was then designed so that the size of the low melting domain of the amplified *APC* gene fragment was shortened by 16 basepairs. The resultant 243-bp PCR fragments had a 141-bp low melting domain including a 121-bp measurable target size (*APC* cDNA bp 8543-8663). It was found that the reduction in the size of low melting domain greatly increased the separation efficiency of the amplified fragment.

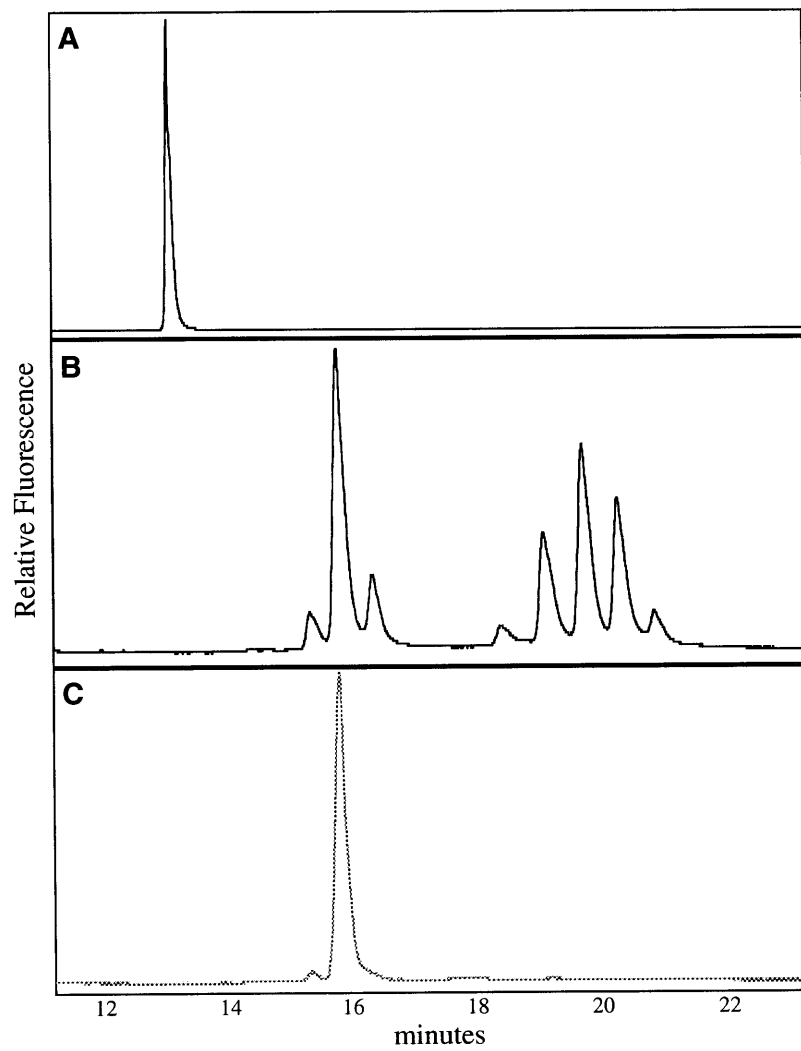
These experimental results are consistent with the notion that peak broadening on

CDCE is caused by an interconversion between the migration states of DNA molecules (i.e. the partially melted and unmelted states) with different electrophoretic mobilities (Khrapko et al., 1996). According to a random walk model (Khrapko et al., 1996), the peak variance ( $\sigma$ ) is given by:  $\sigma = l(n)^{1/2}$ , where  $l$  is the length of an elementary step and  $n$  is the number of steps taken. In CDCE separation,  $l$  is proportional to  $\Delta tE$ , where  $\Delta t$  is the life time of one of the two migration states and  $E$  is the electric field strength; and  $n$  is proportional to  $L / \Delta tE$ , where  $L$  is the migration distance (i.e. the length of the water-jacket). Therefore, the peak variance on CDCE is proportional to  $(\Delta tEL)^{1/2}$  (Khrapko et al., 1996). According to this model, a high efficiency separation (small  $\sigma$ ) on CDCE can be achieved by reducing the field strength ( $E$ ), the length of the water-jacket ( $L$ ) and the life time of one of the two migration states ( $\Delta t$ ). Since  $\Delta t$  is proportional to the length of the low melting domain of the DNA molecule, a reduction in the low melting domain will lead to an increase in the separation efficiency.

The use of primer AP4H in amplifying the chosen *APC* gene fragment, however, introduced several interfering PCR byproducts which migrated slower than the desired fragment on CDCE (Fig. 11A and B). The single-stranded forms of these byproducts appeared to be shorter than the normal fragment by one to four nucleotides upon CE separation. Interestingly, these byproducts were not produced by *exo<sup>-</sup> Pfu* (without the exonuclease activity) or *Taq* polymerase, indicating that they were associated with the exonuclease activity of the native *Pfu*. Several attempts were made to eliminate these byproducts. It was found that their level in the final PCR products were not effected by changing the concentrations of dNTP or PCR buffer components, but could be reduced (but not eliminated) by increasing the amounts of native *Pfu* polymerase used in PCR. Finally, it was discovered that these byproducts could be removed by incubating the final PCR reaction with additional thermostable DNA polymerase (including native and *exo<sup>-</sup> Pfu*, *Taq* and *Vent*) at 72°C for 5 - 10 min (Fig. 11C).

Figure 11. Removal of interfering PCR byproducts by a post-PCR incubation procedure.

About 0.1  $\mu\text{g}$  of TK6 genomic DNA was subjected to PCR amplification using primers AP1 and AP4H in 10- $\mu\text{l}$  reaction for 42 cycles. CE was performed at 9  $\mu\text{A}$  in a 26 cm, 75  $\mu\text{m}$  i.d. capillary with the detection window positioned at 7 cm from the capillary outlet. The horizontal axis represents the time since starting the run, and the vertical axis represents the relative intensity of fluorescence. A, CE separation of the PCR products at room temperature. B, CDCE separation of the PCR products at 64.2°C using a 6-cm water-jacket. C, CDCE separation of the PCR products which had been incubated with additional 0.5 unit of *Pfu* at 72°C for 10 min.



Optimal PCR cycle conditions were determined based on measurement of the PCR efficiency. The PCR efficiency per cycle during the exponential phase of PCR was calculated according to the equation:  $N_p = N_t (1 + e)^n$ , where  $N_t$  and  $N_p$  are the target copy number of the template and product, respectively,  $n$  is the number of cycles and  $e$  is the efficiency per cycle during the exponential phase of PCR. At the optimal PCR cycle conditions (described in Section 3.4.), the efficiency was estimated to be 50 - 60%.

To determine a suitable temperature range for CDCE separation of the 243-bp *APC* gene fragments (*APC* cDNA bp 8441-8683), PCR amplified wild-type and artificial mutant mixtures containing two homoduplexes and two heteroduplexes were subjected to CDCE at various temperatures. The two artificial mutants, mutL and mutH, were designed to carry a G:C → A:T and an A:T → G:C transition at *APC* cDNA bp 8660 and 8652, respectively. As shown in Fig. 12 and 13, the wild-type homoduplex can be well separated from both the low and high  $T_m$  mutants as well as the heteroduplexes at appropriate temperatures. It should be noted that the optimal separation temperatures may be slightly different for different batches of 5% linear polyacrylamide matrices.

## **4.2. Processing of large quantities of genomic DNA from human cells and tissues**

### 4.2.1. The phenol-free DNA isolation procedure

A modified genomic DNA isolation procedure was developed without the use of phenol, a compound which may potentially cause DNA damage. In this procedure, genomic DNA was isolated from the cells or tissues through digestion with proteinase K (1 mg/ml), SDS (0.5%), and RNaseA (0.1 mg/ml) followed by centrifugation and ethanol precipitation. The procedure provided a greater than 90% yield (9 mg) of genomic DNA from  $10^9$  cultured cells or 4 grams of tissues. The  $A_{260}/A_{280}$  ratio was determined to be

Figure 12. Demonstration of CDCE separation of the 243-bp *APC* gene fragments at various temperatures (I).

The DNA samples being separated here are PCR amplified 243-bp *APC* gene fragments (*APC* cDNA bp 8441-8683) including the wild-type homoduplex (GC), the mutL homoduplex (AT), and the two mutant/wild-type heteroduplexes (GT and AC). CDCE was performed at conditions described in Fig. 11 except that a different batch of polyacrylamide matrix was used.

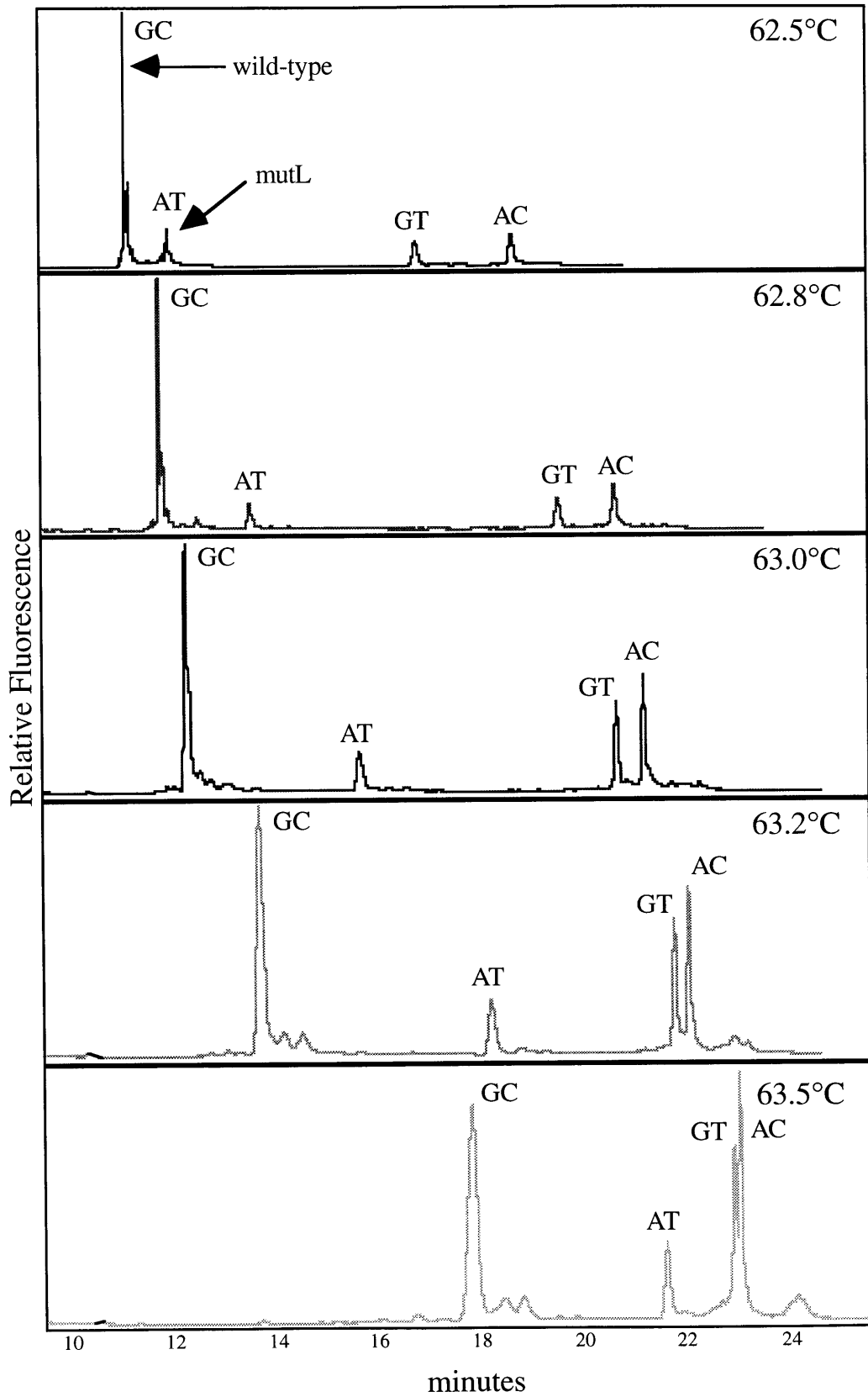
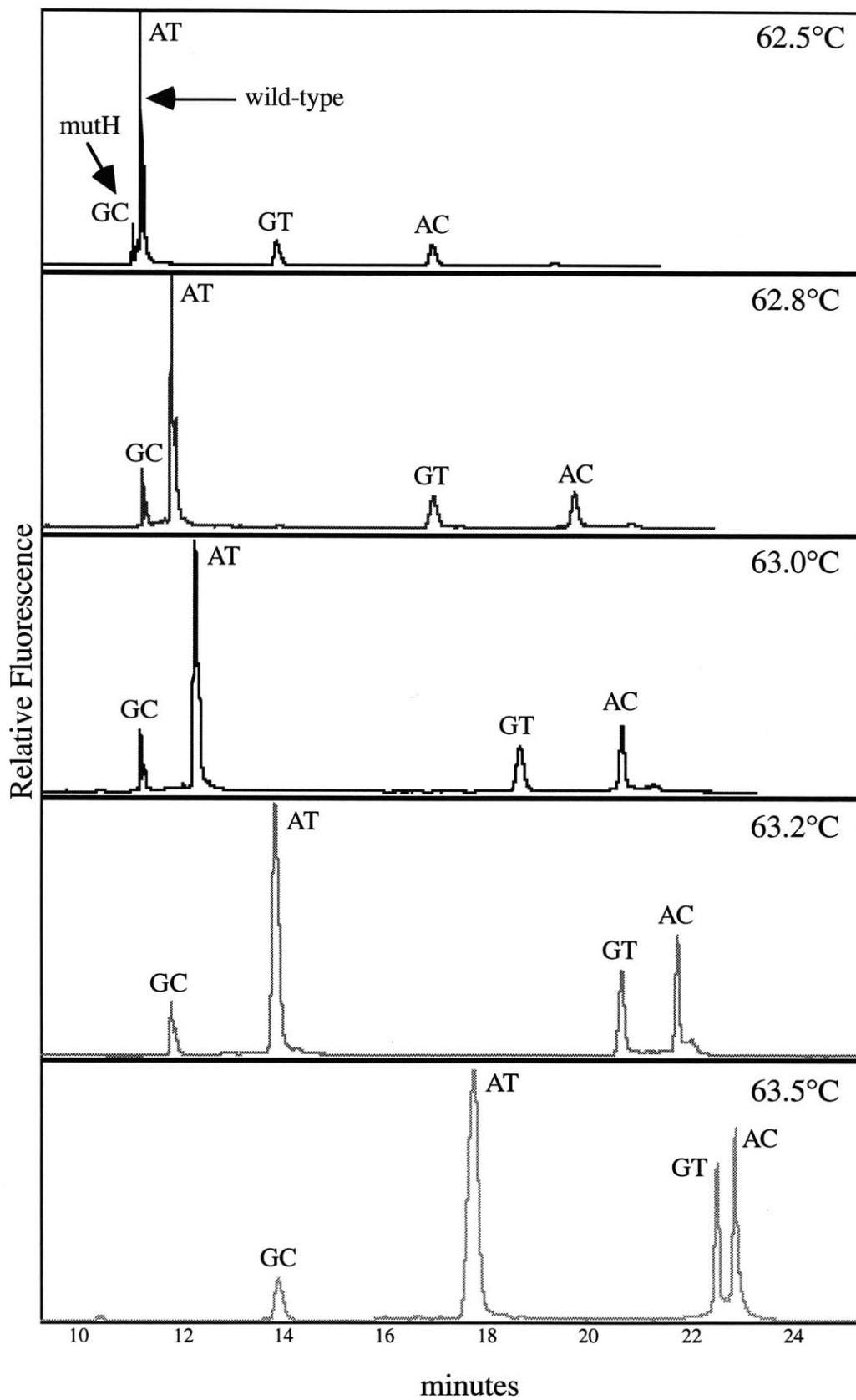




Figure 13. Demonstration of CDCE separation of the 243-bp *APC* gene fragments at various temperatures (II).

The DNA samples being separated here are PCR amplified 243-bp *APC* gene fragments (*APC* cDNA bp 8441-8683) including the wild-type homoduplex (AT), the mutH homoduplex (GC), and the two mutant/wild-type heteroduplexes (GT and AC). CDCE was performed at conditions described in Fig. 11 except that a different batch of polyacrylamide matrix was used.



1.5 - 1.8 and 1.3 - 1.6 for DNA isolated from cultured cells and tissues, respectively. The genomic DNA isolated through this procedure was readily digested by restriction endonucleases.

#### 4.2.2. Sequential release of the desired *APC* gene fragment from human genomic DNA

It was financially prohibitive to use endonucleases *Sau3A* I and *Acc* I to excise directly the 271-bp *APC* gene fragment from the large amounts of genomic DNA required for this study (\$2112 for digestion of 10 mg DNA at 1 unit enzyme/ $\mu$ g DNA). To reduce the monetary cost, a two-step digestion strategy was used (Fig. 14). A 482-bp *APC* gene fragment (*APC* cDNA bp 8422-8903) containing the desired sequence was first excised by digestion with endonucleases *Hae* III and *Xba* I which are 10-fold less expensive than *Sau3A* I and *Acc* I. The digestion was performed overnight at 2 - 3 mg DNA/ml reaction and 1 unit enzyme/ $\mu$ g DNA. The 482-bp *APC* fragment was then enriched from the bulk genomic DNA (see Sections 4.3.2 and 4.4) and digested with *Sau3A* I and *Acc* I to liberate the 271-bp fragment suitable for CDCE separation.

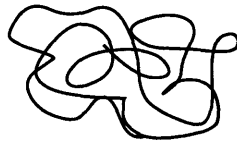
### **4.3. Evaluation of the CDGE-based approach to mutational spectrometry of the *APC* target sequence in human cells**

#### 4.3.1. Overview of the approach

In the previously developed approach to detecting mutational spectra in mitochondrial DNA, mutants were first separated from nonmutant sequences in the presence of several micrograms of cellular DNA through CDGE, and then amplified and enriched by a combination of high-fidelity PCR and CDCE (Khrapko et al., 1997a). CDGE was the method of choice because of its relatively large loading capacity required for

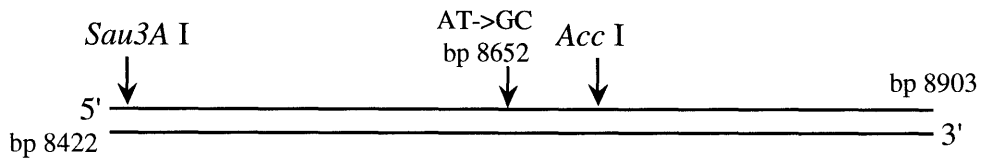
Figure 14. Sequential release of the desired *APC* gene fragment from large quantities of human genomic DNA.

The genomic DNA was first digested with two "inexpensive" endonucleases *Hae III* and *Xba I* to excise the 482-bp *APC* gene fragment (*APC* cDNA bp 8422-8903). A 492-bp internal standard mutant (*APC* cDNA bp 8422-8913) was doped into the digestion at a precise fraction. The *APC* gene fragment was then enriched from the bulk genomic DNA, and finally subjected to digestion with the two endonucleases *Sau3A I* and *Acc I* to liberate the 271-bp fragment (*APC* cDNA bp 8434-8704) suitable for CDCE analysis. The positions of the AT -> GC transition carried by the internal standard mutant and the recognition sites of *Sau3A I* and *Acc I* in the 482-bp fragment are indicated by arrows.

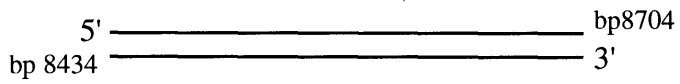


human genomic DNA

digest with *Hae* III and *Xba* I  
add the internal standard mutant



enrich the 482-bp *APC* gene fragment  
digest with *Sau*3A I and *Acc* I



the study.

In the earlier stage of this project, a similar approach was developed for the analysis of the nuclear *APC* gene. As outlined in Fig. 15, the desired *APC* gene fragment was first enriched from genomic DNA through size fractionation using PAGE after two consecutive steps of restriction digestion. The target-enriched sample was then subjected to CDGE to enrich mutant sequences in the presence of residual cellular DNA. The mutant-enriched sample was amplified by high-fidelity PCR. The amplified mutants were further enriched following two rounds of CDCE/high-fidelity PCR, and finally displayed on CDCE. This approach was evaluated through the analysis of TK6 genomic DNA which had been doped with known fractions of an internal standard mutant. This approach and the obtained results are described in detail in the following sections.

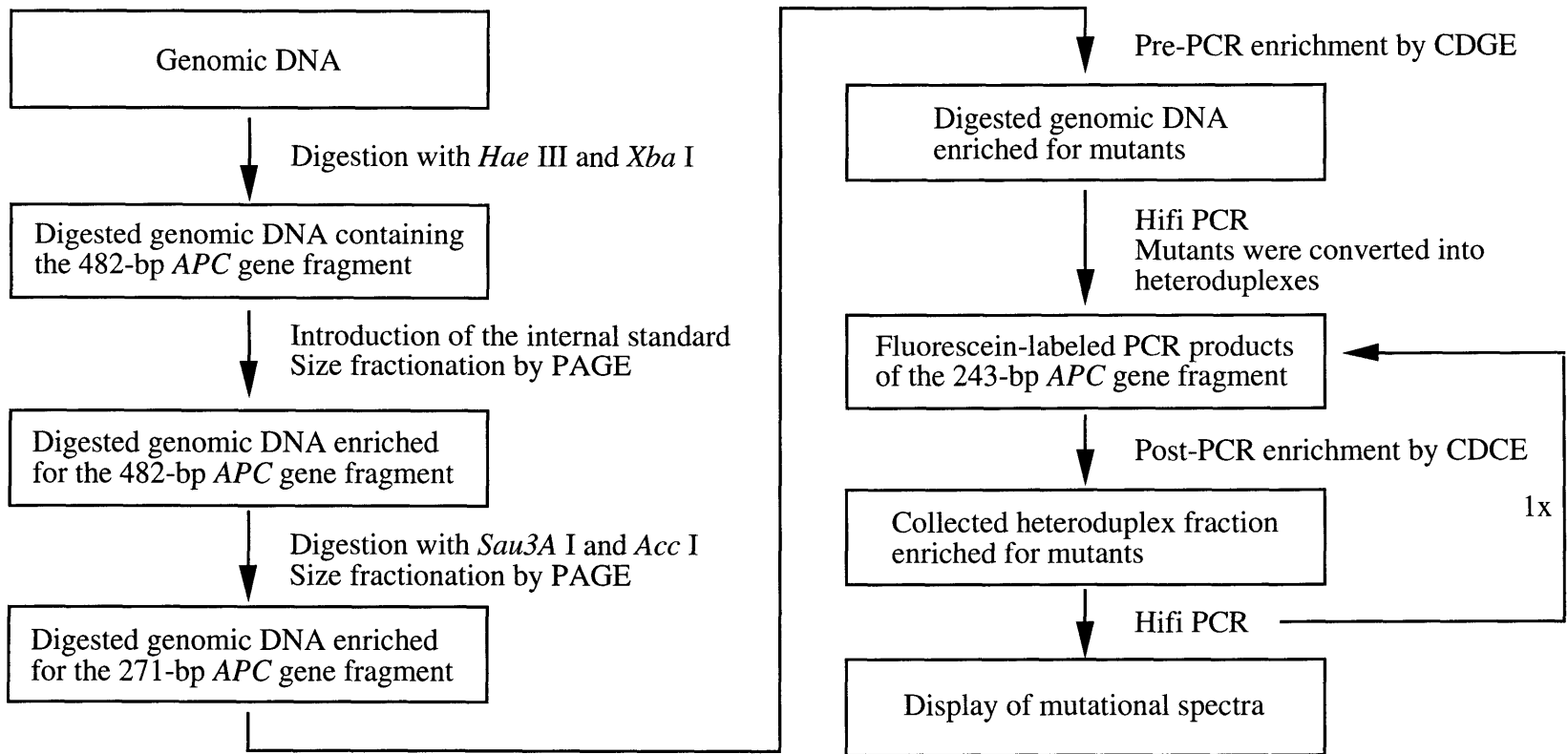
#### 4.3.2. Enrichment of the target sequence through size fractionation by PAGE

Approximately 1 mg of genomic DNA extracted from TK6 cells was digested with *Hae* III and *Xba* I. The copy number of the target sequence in the sample was measured. The internal standard mutant was then doped at a precise fraction ( $10^{-5}$  or  $10^{-4}$ ). The genomic DNA sample was separated by preparative PAGE. A *pBR322* DNA/*Msp* I digest was run in parallel serving as size markers. After electrophoresis, the gel slices containing DNA fragments from 430 to 530 basepairs in length were excised and DNA was electroeluted from the gels. The recovered target sequences were then digested with *Sau3A* I and *Acc* I, and subjected to a second PAGE separation. The gel slices containing fragments of between 230 to 410 basepairs in length were excised and DNA was recovered by electroelution.

To estimate the level of enrichment of the target sequence, aliquots of the digested samples were separated by PAGE, stained in ethidium bromide solution and then photographed under a UV transilluminator. By comparing the fluorescent intensity of the

Figure 15. Flow diagram of the CDGE-based approach to mutational spectrometry of the *APC* gene target sequence in human cells.

See text for details.





entire digest to the portion corresponding to the excised gel slice on the preparative PAGE, the enrichment efficiency was estimated to be 50 - 100-fold in total. Based on measurement of the target copy numbers, the overall yield of the target sequence was estimated to be around 40%. Thus, 1 mg of genomic DNA was reduced to 10 - 20  $\mu$ g containing about  $8 \times 10^7$  copies of the desired 271-bp *APC* gene fragments.

#### 4.3.3. Optimization of the CDGE conditions

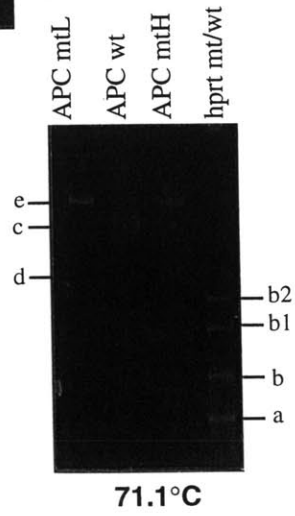
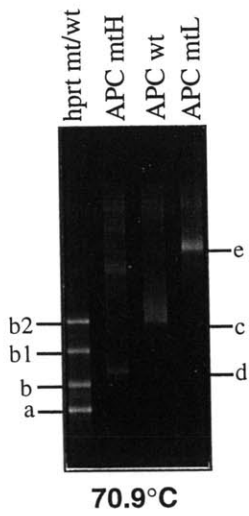
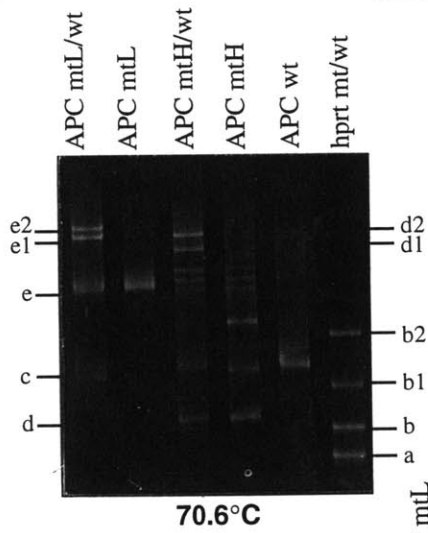
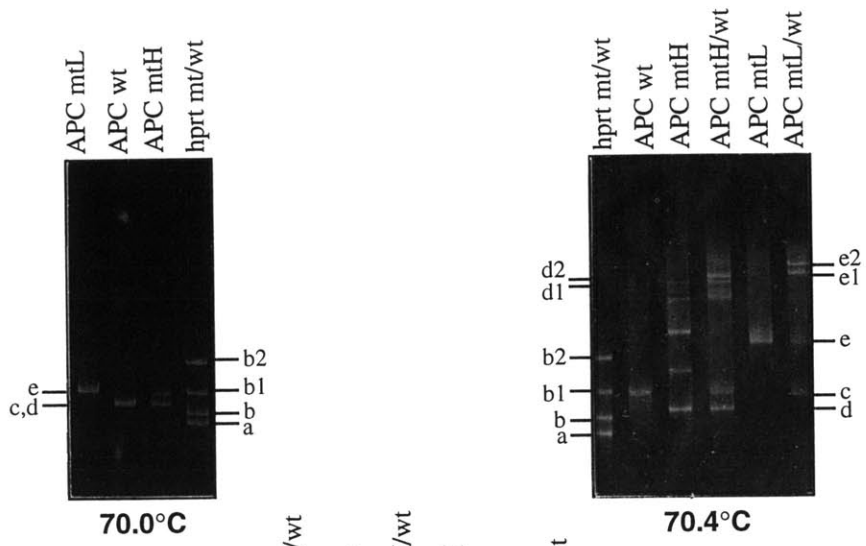
The partially denaturing conditions in CDGE were achieved by controlling the temperature of the TAE buffer tank where the gels were submerged during electrophoresis. To obtain an efficient enrichment of mutants in the target-enriched genomic DNA, an optimal separation temperature had to be determined. This was accomplished using test runs at various temperatures. The test samples contained wild-type and artificial mutant (mutL or mutH) fragments that were PCR amplified with primers IS1 and IS2 and then digested with *Sau3A* I and *Acc* I to resemble the 271-bp cellular *APC* gene fragments. Along with the test samples, PCR amplified fluorescein-labeled *hprt* exon 3 wild-type and mutant mixtures were run in parallel which served as markers.

Fig. 16 shows CDGE separations of the *APC* gene test samples and the *hprt* markers at different temperatures. By comparing the various separation conditions, it was determined that the optimal temperature for separating the wild-type homoduplex from the mutant homoduplexes was 70.9°C, and that from the mutant/wild-type heteroduplexes was 70.4°C. The relative positions of the four *hprt* bands (two homoduplexes and two heteroduplexes), which were visualized by illuminating the gel with a argon laser of 488 nm, would be used as markers indicating the optimal separation conditions for later separating cellular DNA.

It was found that the CDGE separation performance was not only highly sensitive to temperature, but also to the TAE buffer conditions (such as pH) as well. Since it was

Figure 16. CDGE separations of the wild-type and mutant sequences in the low-melting domain of the 271-bp *APC* gene fragment at different temperatures.

CDGE gels (17 cm x 16 cm x 0.4 mm) contained 7.5% acrylamide and 1/40 bis-acrylamide in 1 x TAE (40 mM Tris-acetate, 1 mM EDTA, pH 8.0). The gels were polymerized with 0.1% ammonium persulfate and 0.1% TEMED. The gels were submerged in a 1 x TAE buffer tank which was heated to the precise temperature indicated for each run. Electrophoresis was performed at a constant voltage of 160 V for 90 min. The sample that was separated in each lane is labeled at the top of the lane: the *hprt* gene exon 3 mutant and wild-type mixture (*hprt* mt/wt), the *APC* wild-type sequence (*APC* wt), the *APC* high  $T_m$  mutant (*APC* mtH), the *APC* high  $T_m$  mutant and wild-type mixture (*APC* mtH/wt), the *APC* low  $T_m$  mutant (*APC* mtL), and the *APC* low  $T_m$  mutant and wild-type mixture (*APC* mtL/wt). The DNA fragment corresponds to each band is indicated by a letter: a, the *hprt* exon 3 wild-type homoduplex; b, the *hprt* exon 3 mutant homoduplex; b1 and b2, the two *hprt* exon 3 mutant/wild-type heteroduplexes; c, the *APC* wild-type homoduplex; d, the *APC* high  $T_m$  mutant homoduplex; d1 and d2, the two *APC* high  $T_m$  mutant/wild-type heteroduplexes; e, the *APC* low  $T_m$  mutant homoduplex; e1 and e2, the two *APC* low  $T_m$  mutant/wild-type heteroduplexes.



impractical to refill the 20-liter CDGE tank with freshly prepared TAE after every run, it was necessary to re-optimize the separation temperature using a pre-determined *hprt* marker separation pattern as a reference before the separation of cellular DNA.

#### 4.3.4. Pre-PCR enrichment of mutants by CDGE

Two CDGE separation approaches were tested for enriching mutants in the target-enriched TK6 cellular DNA. One approach separated mutants in their homoduplex forms from the excess of wild-type DNA. By excising the appropriate portions of the gel below and above the wild-type DNA, high and low  $T_m$  mutants were enriched separately. In the other approach, the samples were boiled and reannealed to convert all of the mutants into mutant/wild-type heteroduplexes and then subjected to CDGE. All of the heteroduplexes were enriched as a single preparation by excising the appropriate portion of the gel above the wild-type DNA. DNA was then eluted from the gel slices. The elution efficiency was determined to be about 50%.

The efficiency of mutant enrichment by CDGE was estimated by comparing the copy number of the target sequence eluted from the gel slices that contained the enriched mutants and the wild-type homoduplex, respectively. It was found that the enrichment was ~200-fold for high  $T_m$  mutants and ~5-fold for low  $T_m$  mutants in the "homoduplex" approach, and ~10-fold for *all* types of mutants in the "heteroduplex" approach. The low enrichment efficiency for the low  $T_m$  mutants and heteroduplexes was largely due to the "tailing" phenomena produced by the wild-type DNA upon CDGE separation.

#### 4.3.5. High-fidelity PCR and post-PCR enrichment of mutants by CDCE

The mutant-enriched samples were subjected to high-fidelity PCR using primers AP1 and AP4H to produce  $10^{12}$  copies of the target sequence. During the PCR procedure,

the mutants were converted into heteroduplexes in the presence of an excess of wild-type sequence. The PCR products were separated by CDCE under the appropriate conditions so that all of the heteroduplexes were well separated from the wild-type homoduplex and then collected in a single fraction. The heteroduplex fraction was amplified by PCR and subjected to a second round of CDCE/PCR procedure to further enrich and amplify the mutant sequences. Finally, the mutants were separated in the heteroduplex forms on CDCE (Fig. 17 and 18).

#### 4.3.6. The detection limits of the approach

Fig. 17 shows the final CDCE display of mutant heteroduplexes derived from high  $T_m$  mutants enriched by the CDGE "homoduplex" approach. The heteroduplexes of the mutant internal standard at an initial fraction of  $10^{-5}$  in the TK6 sample were clearly observed. The level of the background PCR noise, which was defined as the detection limit for the high  $T_m$  mutants, was at  $10^{-6}$ . Since the enrichment efficiency for the low  $T_m$  mutants were 20-fold lower than that for the high  $T_m$  mutants, the detection limit for the low  $T_m$  mutants in this approach was estimated to be around  $2 \times 10^{-5}$ . Fig. 18 shows the final CDCE display of mutant heteroduplexes derived from the TK6 genomic DNA using the CDGE "heteroduplex" approach. Although the internal standard at an initial fraction of  $10^{-4}$  was clearly observed, the internal standard at  $10^{-5}$  was unobservable due to the PCR noise. Thus the detection limit for all types of point mutants in this approach was determined to be  $10^{-5}$ .

Based on the above described feasibility studies, two major drawbacks in the CDGE-based approach were recognized. First, the detection sensitivity of the approach was not sufficient to observe *all* types of point mutational hotspots in nuclear genes. Second, the procedure of extracting DNA from gel slices was not only labor intensive, but also produced low yields. The overall yield of the target sequence after the three

Figure 17. Demonstration of the sensitivity for the detection of high  $T_m$  mutations in the *APC* gene target sequence in human cells through the CDGE-based approach.

About 1 mg of genomic DNA isolated from TK6 cells was subjected to the CDGE "homoduplex" approach as described in the text. Depicted are the CDCE separations of the high  $T_m$  mutants (in the heteroduplex form) derived from the CDCE-purified wild-type DNA control and TK6 genomic DNA with the mutant internal standard doped at an initial fraction of  $10^{-5}$  and  $10^{-4}$ , respectively. CDCE was performed under the same conditions as represented in Fig. 11. The wild-type DNA homoduplex (AT), internal standard homoduplex (std. GC), and internal standard/wild-type heteroduplexes (std. GT and AC) are shown as indicated.

CDGE->PCR->2x[CDCE->PCR]

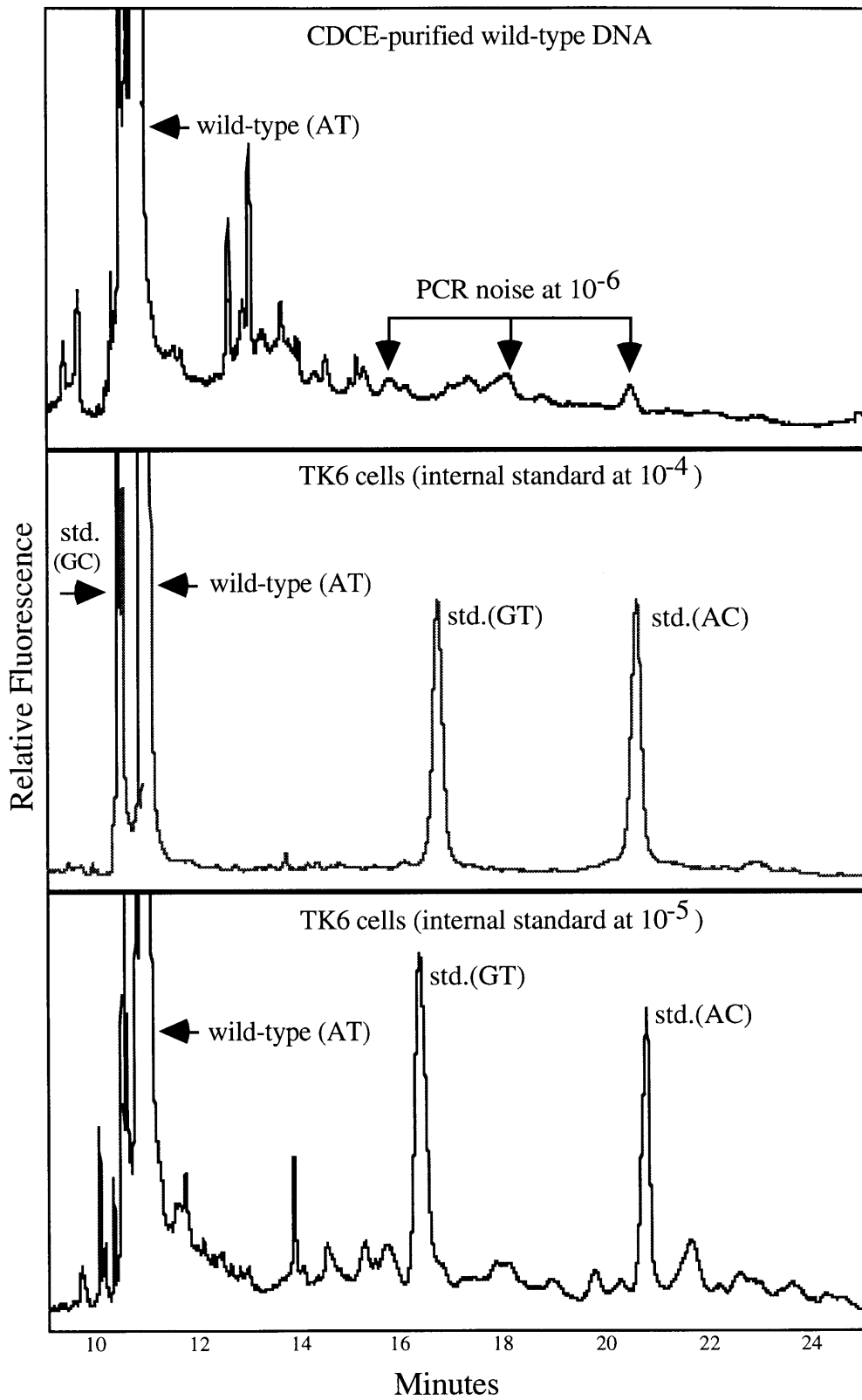
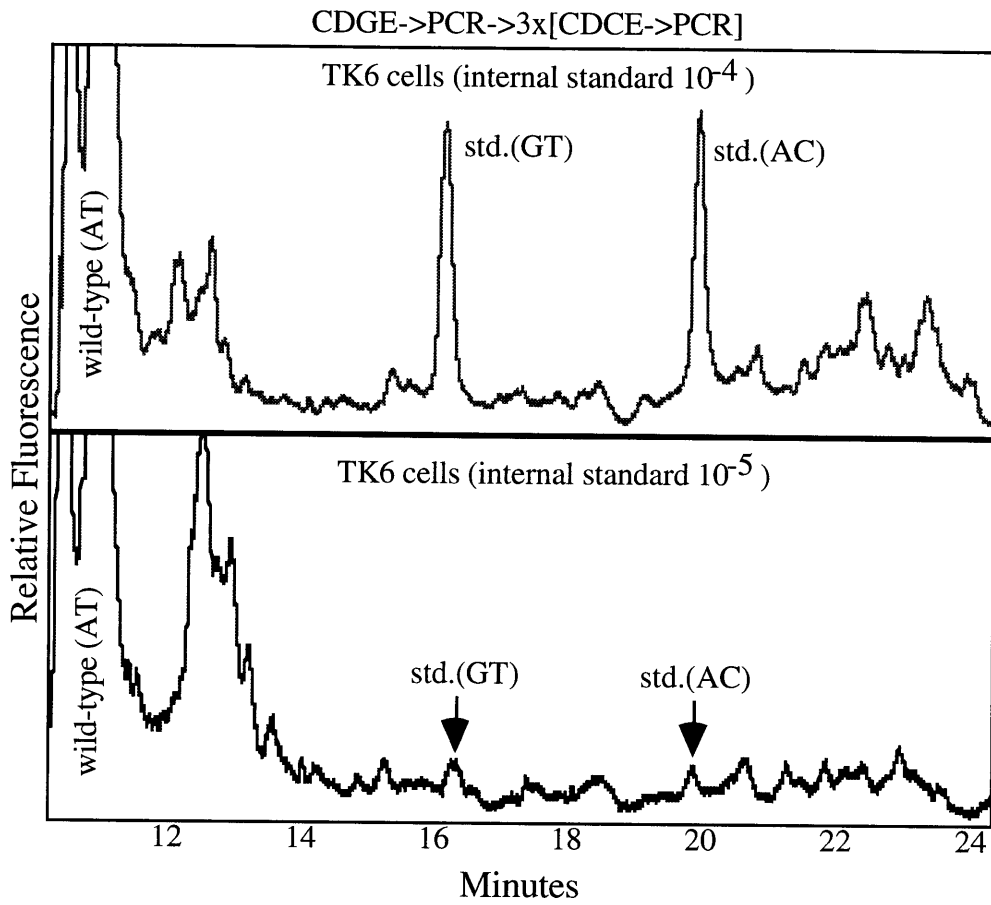


Figure 18. Demonstration of the sensitivity for the detection of *all* types of point mutations in the *APC* gene target sequence in human cells through the CDGE-based approach.

About 1 mg of genomic DNA isolated from TK6 cells was subjected to the CDGE "heteroduplex" approach as described in the text. Depicted are the final CDCE separations of *all* types of mutants (in the heteroduplex form) derived from TK6 genomic DNA with the mutant internal standard doped at an initial fraction of  $10^{-5}$  and  $10^{-4}$ , respectively. CDCE was performed under the same conditions as represented in Fig. 11. The wild-type DNA homoduplex (AT) and the internal standard/wild-type heteroduplexes (std. GT and AC) are shown as indicated.





consecutive slab gel separation procedures (including two PAGE and one CDGE) was only 20% (= 40% x 50%). Therefore, it was concluded that the CDGE-based approach was not suitable for the detection of mutational spectra in nuclear genes. It was necessary to develop new methods which would have an increased efficiency of the mutant enrichment as well as the target sequence yield.

The new strategy which is described in the following sections involved the use of sequence-specific hybridization to obtain a higher efficiency of enrichment of the desired *APC* gene fragment from the genomic DNA and then use of CDCE to achieve a higher efficiency of separation of mutants from nonmutant sequences.

#### **4.4. Development of the means to enrich the desired *APC* gene fragment from genomic DNA**

##### 4.4.1. The strategy

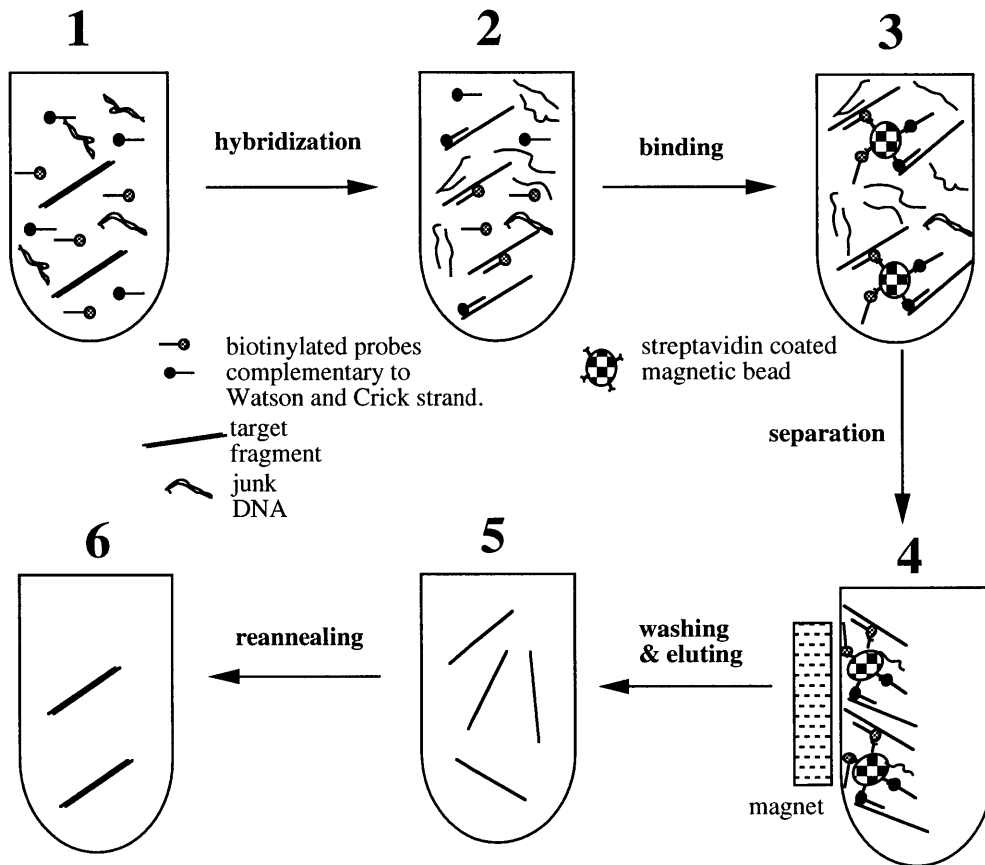
A sequence-specific enrichment strategy was devised to replace the labor-intensive and low-yielding PAGE method for enriching the desired target sequence from genomic DNA. As illustrated in Fig. 19, the restriction digested genomic DNA is heat denatured and hybridized simultaneously with excess biotin-labeled oligonucleotide probes targeted to the Watson and Crick strands of the desired sequence. The hybrids are then captured by streptavidin-coated paramagnetic beads and magnetically separated from the bulk DNA solution. The target fragments are then eluted from the beads by heating, and subsequently reannealed to form double-strands. After reannealing the mutant sequences are converted into heteroduplexes in the presence of an excess of wild-type sequence, which then facilitates their subsequent separation by CDCE.

##### 4.4.2. Optimization of the experimental conditions for the *APC* gene fragment.

The above described strategy was applied to enrich the 482-bp *APC* gene fragment excised by endonucleases *Hae* III and *Xba* I from genomic DNA. The target sequence yield and efficiency of enrichment were monitored during the procedure by measuring the copy numbers of the target sequence and a 205-bp mitochondrial DNA sequence (mitochondrial bp 10011 - 10215) used as a non-target reference. The chosen mitochondrial sequence is present at ~1000 copies per TK6 and MT1 cell and has the necessary biphasic melting profile for CDCE separation used in the copy number measurement (Khrapko et al., 1997b; Marcelino et al., 1998). To obtain a high efficiency of both target sequence yield and enrichment, the following experimental parameters were

Figure 19. The strategy for sequence-specific enrichment of single-copy nuclear sequences from genomic DNA.

See text for details.



optimized:

(1) The biotin-labeled probes: Two 30-mer biotinylated probes, BP1 (5'-CAAAGTACAGCACAGAATCCAGTGGAAAC-3') (*APC* cDNA bp 8472-8501) and BP2 (5'-AAGACCCAGAATGGCGCTTAGGACTTTGGG-3') (complementary to *APC* cDNA bp 8501-8530), were chosen to complement the Watson and Crick sequence in the high melting domain region of the 271-bp *APC* fragment, respectively. The biotin moiety was attached to the 5' end of each probe through a 18-carbon spacer to reduce potential steric hindrance between the probe/target hybrids and the streptavidin-coated beads.

Because the biotin-labeling efficiency of the probes (the percentage of probes that are actually tagged with biotin) is crucial to the recovery of the target sequence, an assay was developed to check and compare probes synthesized by various manufacturers. Each type of probe was paired with a fluorescein-labeled primer (i.e., BP1 with fluorescein labeled AP4H, and BP2 with fluorescein-labeled AP1) and used in PCR of the corresponding *APC* gene sequence. An excess of streptavidin-coated paramagnetic beads was added to the final PCR solution to capture products that were extended from biotin-labeled probes, leaving behind products extended from unlabeled probes. The amounts of fluorescein-labeled products in the PCR solution before and after adding the beads were then measured by CDCE to determine the percentage of products captured on the beads, i.e., the percentage of the probes with biotin tags. Using this assay, it was found that the biotin-labeling efficiency varied from 50 to 93% among the many different manufacturers. The probes with the highest labeling efficiency were then chosen.

(2) The hybridization conditions: The hybridization was carried out in a solution composed of 6 x SSPE buffer (1.08 M NaCl, 60 mM sodium phosphate, pH 7.4, 6 mM EDTA). The hybridization temperature and the molar ratio of the probe to the target were two of the most crucial parameters affecting the yield of the target sequence. Various hybridization temperatures ranging from 37°C to 65°C were investigated. The most optimal temperature was determined to be between 55 and 60°C.

Although an excess of probes relative to the target sequence was required for a high target sequence yield, the excess of unbound probes left after hybridization would compete with the probe-target hybrids for binding to the paramagnetic beads. As a result, large amounts of beads would be required to completely capture the desired hybrids. The use of large amounts of beads was not only costly but also led to an increase in nonspecific binding to non-target DNA. Therefore, the amounts of probes used in hybridization needed to be carefully balanced with consideration to the target sequence yield, the efficiency of enrichment and the monetary cost. It was found that the yield increased as a function of the probe/target molar ratio until it reached approximately  $5 \times 10^4$  where by the yield started to approach a plateau. Suitable amounts of probes for hybridization were therefore determined to be  $5 \times 10^4$ -fold the (molar) amounts of the target sequence.

(3) The type and quantities of paramagnetic beads: The optimal type of beads for this protocol should have a high binding capacity for biotin-tagged DNA and a low background affinity for DNA and proteins. After a comparison study, paramagnetic beads consisting of controlled porous glass encapsulating magnetite (CPG) were chosen. The binding capacity of the CPG bead was found to be 6-fold higher than that of the widely used Dynal bead (Dynal). Furthermore, preliminary results showed that the CPG beads with a silanol (Si-OH) surface had a lower background affinity for non-target DNA than the polystyrene made Dynal beads.

(4) The binding condition: Because a high salt concentration ( $\geq 1$  M NaCl) was recommended for optimal binding between biotin and streptavidin, pre-washed beads were directly added to the hybridization solution to capture the probe-target hybrids. The binding efficiency of the hybrids onto the beads was found optimal at a temperature of 45 - 50°C (instead of room temperature recommended by the manufacture). It is conceivable that such a high temperature might help overcome the steric hindrance that might exist between the beads and the hybrids. It was experimentally determined that a minimum of 1 mg of the CPG beads was required to completely capture  $5 \times 10^8$  copies of the desired hybrids in the

presence of a large excess of competing probes.

(5) The washing and eluting conditions: Once separated from the bulk DNA solution, the beads were washed four times each with 1 M NaCl, 10 mM Tris-HCl, pH 7.6, and 2 mM EDTA buffer at 50°C for 5 min to remove nonspecific bindings of DNA and protein. It has been reported that the nonspecific binding interaction between the DNA and the streptavidin bound to the beads are reduced in buffers having high ionic strengths ( $\geq 1$  M NaCl) (Ito et al., 1992). The beads were then washed twice with 50 mM NaCl, 1 mM Tris-HCl, pH7.4 buffer at room temperature to reduce the amounts of salts carried over by the beads. DNA was eluted from the beads by incubating with ddH<sub>2</sub>O at 72°C for 2 min.

The above described optimized conditions for enriching the 482-bp *APC* gene fragment from genomic DNA are summarized in Table 7.

#### 4.4.3. The enrichment efficiency and yield of the target sequence.

Up to 5 mg of genomic DNA isolated from TK6 and MT1 cells were digested with *Hae* III and *Xba* I and subjected to the above described target enrichment protocol. Through this protocol, the *APC* target sequence was enriched over 10<sup>4</sup>-fold relative to the mitochondrial DNA producing a target sequence yield of between 70 to 80%. Thus, 5 mg of genomic DNA was reduced to less than 1 µg containing 7 - 8 x 10<sup>8</sup> copies of the desired *APC* gene fragments. It was found that the majority (> 95%) of the unrecovered target sequences were still in the genomic DNA solution. These target sequences could not be captured by additional amounts of beads, indicating that they were not hybridized to biotin-labeled probes.

This protocol was also tested in a smaller amount ( $\leq 1$  mg) of genomic DNA isolated from human tissue samples. Although the efficiency of enrichment was more or less the same, the target sequence yield from the genomic DNA obtained from tissues was found to be only ~50%, lower than that obtained from cultured cells. The lower yield is



Table 7. Optimized experimental conditions for sequence-specific enrichment of the desired *APC* gene fragment from digested human genomic DNA.

<b>Experimental parameter</b>	<b>Optimal condition</b>
Hybridization	
Biotin-labeled oligonucleotide probes	> 90% labeling efficiency, 18-carbon spacer
Hybridization buffer	6 x SSCP
Hybridization temperature and time	55 - 60°C, 2 hours
Molar ratio of the probe/target sequence	5 x 10 <sup>4</sup>
Binding	
Type of the paramagnetic bead	controlled porous glass (5 μm, 50 nm pore diameter)
Quantities of the beads	1 mg beads per 5 x 10 <sup>8</sup> copies of the target
Binding temperature and time	50°C, 1 hour
Washing	
Washing buffer	1 M NaCl, 10 mM Tris-HCl, 2 mM EDTA, pH 7.6
Washing temperature and time	50°C for 5 min, repeat four times
Eluting	
Eluting temperature and time	2 mg beads/ml in ddH <sub>2</sub> O, 72°C for 2 min, twice
Reannealing	
Target concentration	10 <sup>8</sup> copies of target/ml
Reannealing buffer	0.2 M NaCl, 10 mM Tris-HCl, 2 mM EDTA, pH 7.6
Reannealing temperature and time	55°C, overnight

most probably due to the poorer purity of the tissue genomic DNA which to some extent effected the efficiency of digestion by *Hae* III and *Xba* I. Nevertheless, it was found that applying one or two additional round(s) of the enrichment procedure to the same genomic DNA sample could be used to further recover the residual target sequence and increase the overall yield to 70%.

The target-enriched samples were reannealed to double-strands and then digested with *Sau3A* I and *Acc* I to release the 271-bp *APC* gene fragments suitable for CDCE analysis. The efficiency of the *Sau3A* I/*Acc* I digestion was then measured (see Section 4.2.4). Occasionally, the digestion efficiency was found to be less than 90% indicating that either the reannealing was not complete or that the digestion reaction was inhibited. In most cases, the incomplete digestion was found to be due to inhibition of the restriction enzymes by high concentrations of salts, EDTA or other impurities carried over from the previous procedures. Simply increasing the digestion volume would usually remedy the problem. After complete digestion, the samples were desalted and concentrated through ultrafiltration.

#### **4.5. Development of the means to enrich cellular mutants by wide-bore CE**

##### 4.5.1. The necessity and technical problem of using wide-bore capillaries in CDCE

It was previously shown that the CDGE separation technique was not sufficient for the detection of mutations in the *APC* gene target sequence due to its low enrichment efficiency for mutant heteroduplexes and low target sequence yield (see Section 4.3). Since CDCE can achieve a higher separation resolution, the substitution of CDGE with CDCE was expected to offer a more effective way to enrich mutant sequences with both a higher efficiency of mutant enrichment and target sequence yield.

CDCE was initially developed using capillaries with 75  $\mu\text{m}$  i.d.. These capillaries

have a very limited loading capacity of no more than 80 ng of DNA. It was found that even after the sequence-specific enrichment procedure, the sample still could not be run on a 75  $\mu\text{m}$  column CDCE due to disturbances in the local potential gradient produced by the residual amounts of DNA and/or other impurities loaded on the column. Current drop was observed a few minutes after starting a run.

One direct approach to increase the loading capacity of CDCE would be to use wider bore capillaries. Theoretically, the sample loading capacity is proportional to the capillary's cross sectional area. However, any increase in the capillary bore decreases its surface-area-volume ratio and consequently leads to a less efficient dissipation of Joule heating generated by the electric current. The decreased heat loss inevitably leads to a rise in the temperature differential between the center of the capillary and its wall. Fig. 20 depicts the profile of the temperature gradient inside a capillary during electrophoresis. Since both the DNA melting kinetics and the gel viscosity are strong functions of temperature, the temperature gradient inside the capillary would cause a DNA velocity gradient and thus loss in CDCE resolution.

A quantitative description of the temperature difference ( $\Delta T$ ) between the center and wall of the capillary is given by (Knox, 1988):

$$\Delta T = Q (\text{i.d.})^2 / (16K) \quad (1)$$

$$Q = E^2 \Lambda c \phi \quad (2)$$

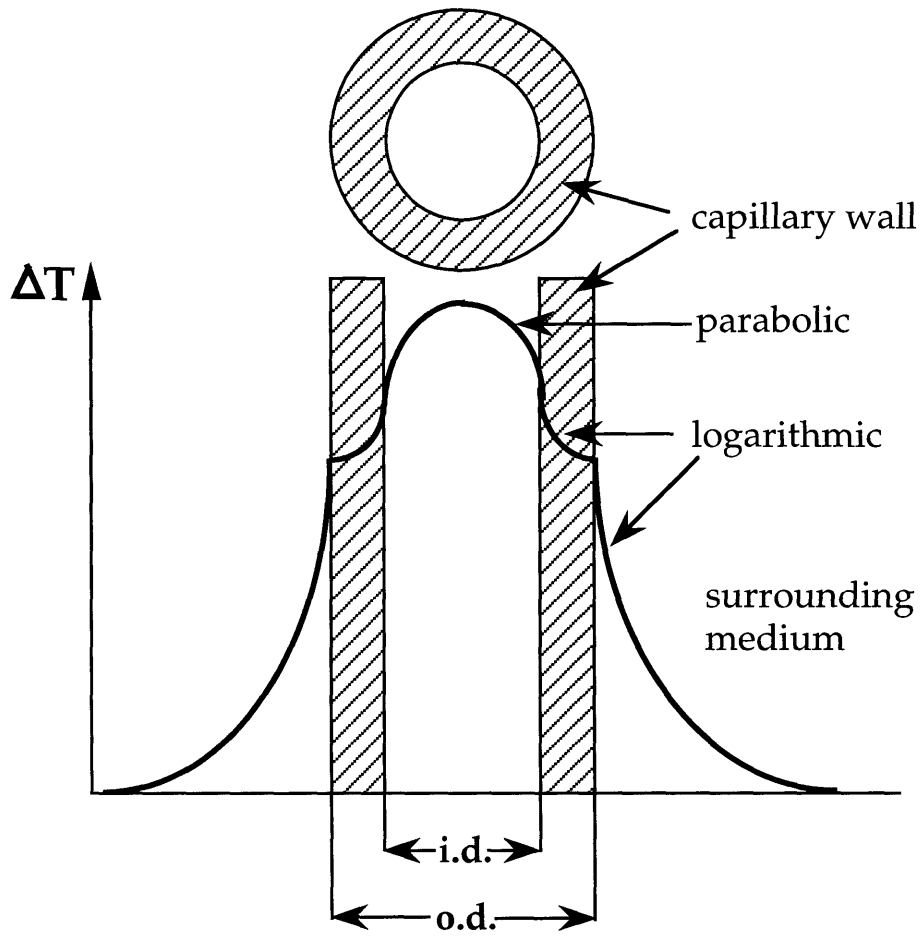
where  $Q$  is the heat output per unit volume [ $\text{W}\cdot\text{cm}^{-3}$ ],  $E$  is the field strength [ $\text{V}\cdot\text{cm}^{-1}$ ],  $\Lambda$  is the equivalent conductance of the electrolyte solution [ $\text{cm}^2\cdot\text{W}^{-1}\cdot\text{mol}^{-1}$ ],  $c$  is the electrolyte concentration [M],  $\phi$  is the total porosity of the medium (=1 for open tube), and  $K$  is the thermal conductance of the medium [ $\text{W}\cdot\text{cm}^{-1}\cdot\text{K}^{-1}$ ].

#### 4.5.2. Optimization of CDCE conditions for wide-bore capillaries.

The feasibility of using wide-bore capillaries of 250, 320 and 540  $\mu\text{m}$  in CDCE

Figure 20. Schematic representation of the temperature profile as a result of current flow across a capillary.

(Adapted from Kuhn and Hoffstetter-Kuhn, 1993)



was investigated. All capillaries were cut 30 cm in length with the detection window positioned 20 cm from the capillary inlet. The length of the water jacket was 6 cm. The test sample used for separation contained fluorescein-labeled *APC* gene wild-type and mutH homoduplexes and their two heteroduplexes that were PCR amplified using primers AP1 and AP4H.

According to equations (1) and (2), in order to maintain a low temperature differential ( $\Delta T$ ) inside a large bore capillary during electrophoresis, one solution is to reduce the applied electric field strength ( $E$ ). To experimentally determine an upper field strength limit with acceptable resolution in each wide-bore capillary, electrophoresis at different electric field strengths was performed. Fig. 21 illustrates such a test in a 250  $\mu\text{m}$  capillary. As shown in Fig. 21 A and B, the resolution of separation was largely maintained at lower electric field strengths of 83 V/cm (30  $\mu\text{A}$ ) and 100 V/cm (40  $\mu\text{A}$ ). When the field strength increased to 127 V/cm (50  $\mu\text{A}$ ), the resolution decreased (Fig. 21 C). At 150 V/cm (60  $\mu\text{A}$ ), distorted peaks appeared, a result of over-heating near the center of the capillary (Fig. 21 D). From an evaluation of both the resolution and the separation times, 100 V/cm was determined to be the most suitable field strength for CDCE in 250  $\mu\text{m}$  capillary.

Note that due to the highly temperature-sensitive nature of CDCE separation (Fig. 11 and 12), an increase in the electric field strength causes a rise in the overall intracapillary temperature and a change in the separation pattern. In order to obtain the same separation pattern at different electric fields, the water bath temperature was adjusted to counteract the intracapillary temperature difference. As indicated in Fig. 21, the water jacket temperature was set at 62.6, 62.5, 62.4 and 62.3°C for CDCE at 30, 40, 50 and 60  $\mu\text{A}$ , respectively.

Suitable electrophoretic conditions (both field strengths and water jacket temperature) were determined in the other two wide-bore capillaries with 320 and 540  $\mu\text{m}$  bores following a similar experimental process. The field strengths were 73 V/cm and 50 V/cm for the 320 and 540  $\mu\text{m}$  capillaries, respectively. The CDCE separations using the 75

Figure 21. CDCE separation performance in a 250  $\mu\text{m}$  i.d. capillary at different electric field strengths.

(A) 83 V/cm, 30  $\mu\text{A}$ , 62.6°C. (B) 100 V/cm, 40  $\mu\text{A}$ , 62.5°C. (C) 127 V/cm, 50  $\mu\text{A}$  constant, 62.4°C. (D) 150 V/cm, 60  $\mu\text{A}$ , 62.3°C. The temperature indicated is that of the water-jacket set for each run. Approximately  $10^9$  copies of fluorescein labeled PCR fragments were loaded onto the capillary.

EFFECT OF FIELD STRENGTH ON  
RESOLUTION IN 250  $\mu\text{m}$  CAPILLARY

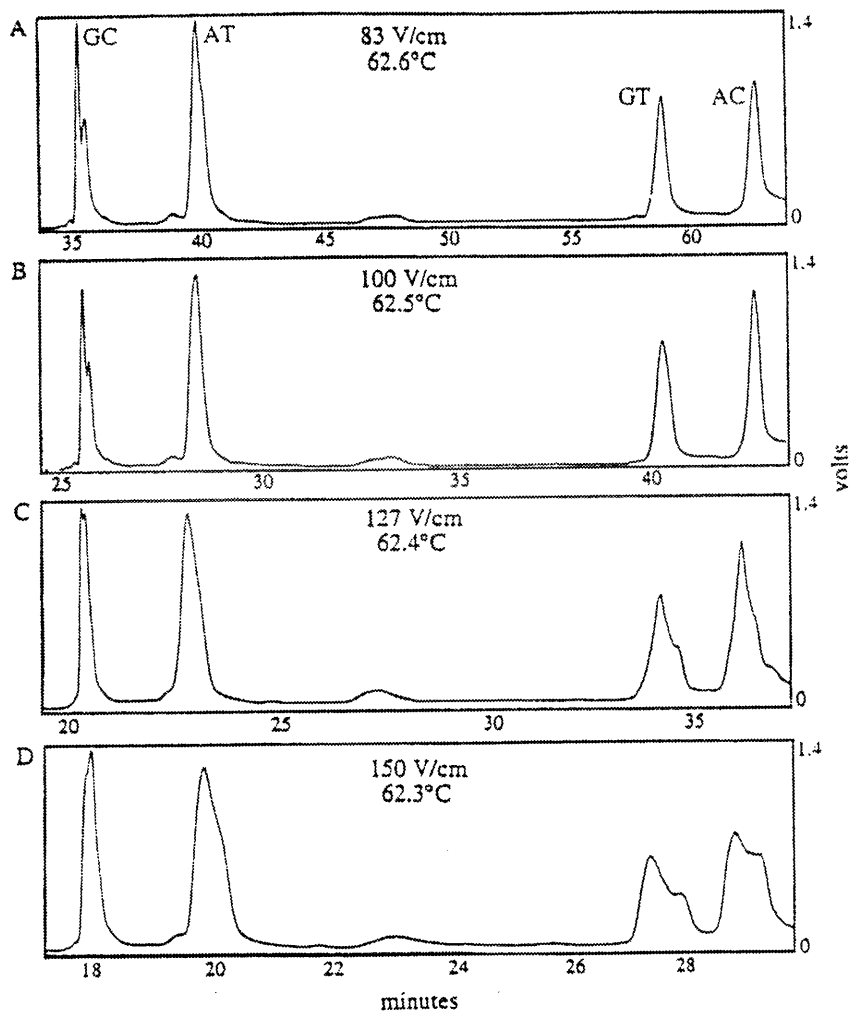
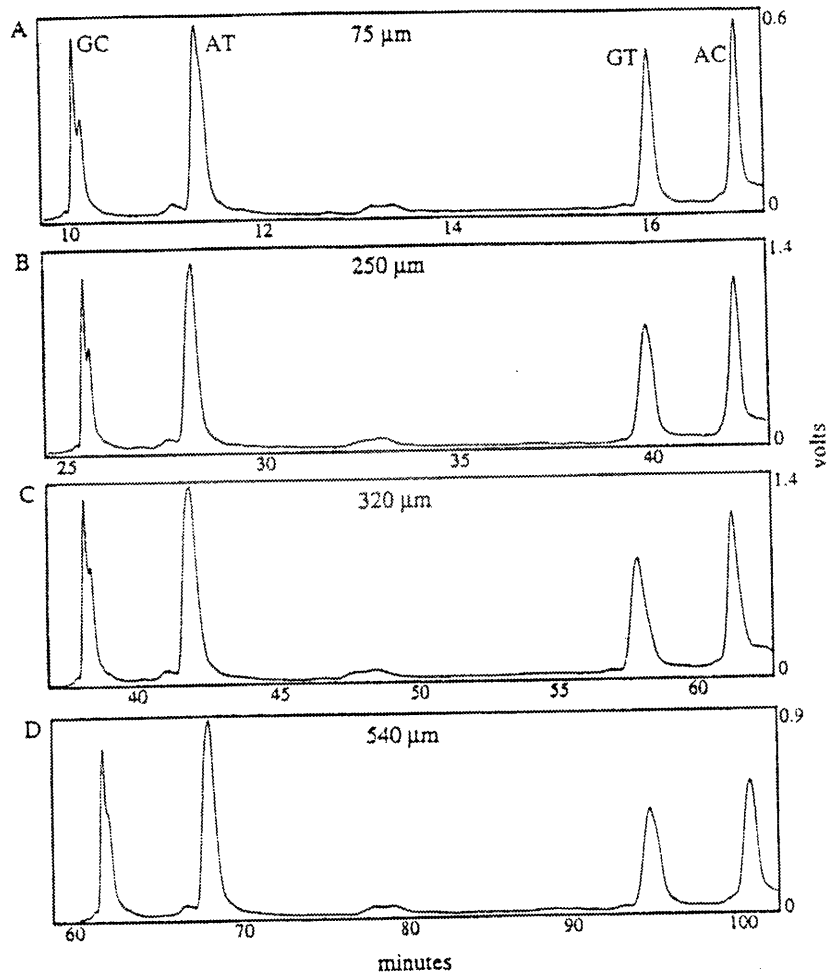




Figure 22. Comparison of CDCE separation performance using capillaries of various sizes.

(A) 75  $\mu\text{m}$ , 260 V/cm, 9.0  $\mu\text{A}$ , 62.5°C. (B) 250  $\mu\text{m}$ , 100 V/cm, 40  $\mu\text{A}$ , 62.5°C. (C) 320  $\mu\text{m}$ , 90 V/cm, 50  $\mu\text{A}$ , 62.5°C. (D) 540  $\mu\text{m}$ , 60 V/cm, 80  $\mu\text{A}$ , 62.6°C.  $10^9$  -  $10^{10}$  copies of fluorescein labeled PCR fragments were loaded onto each capillary.

EFFECT OF COLUMN SIZE ON RESOLUTION



$\mu\text{m}$  narrow-bore and 250, 320, 540  $\mu\text{m}$  wide-bore capillaries were compared in Fig. 22. The results show that the resolution of CDCE separation can be maintained in wide-bore capillaries when the electric field strength is correspondingly reduced.

According to equations (1) and (2) (on Page 131), the ratio of the intracapillary temperature difference of the wide-bore capillary to that of the 75  $\mu\text{m}$  capillary ( $\Delta T_{\text{wb}}/\Delta T_{75}$ ) was calculated based on the experimentally determined suitable electric field strengths (Table 8). Calculations show that  $\Delta T_{\text{wb}}$  of each wide capillary is higher than  $\Delta T_{75}$  ( $\Delta T_{\text{wb}}/\Delta T_{75} > 1$ ) under the given operational conditions. This indicates that low electric field strength has a beneficial effect on the CDCE separation in addition to producing less Joule heat (Khrapko et al., 1996), making separations in wide capillaries relatively more tolerable to the intracapillary temperature gradient.

Theoretical calculations also suggested the use of lower electrolyte concentrations for reducing the heat output (Knox, 1988; Grushka et al., 1989). At lower electrolyte concentrations, higher electric field strengths might be applied without diminishing resolution. However, it was found in this study that as the electrolyte concentrations in both linear polyacrylamide and electrophoretic buffer were reduced by half (from 1 x TBE to 0.5 x TBE), the peaks broadened and the resolution decreased. Since the amount of sample injected (0.1 - 0.01 pmol) was too low to cause inhomogeneity in the electrical field within the capillary (Mikkers et al., 1979; Jorgenson et al., 1983), one reason for the observed peak broadening at low electrolyte concentration could be due to the reduction of cations which through their interaction with the negatively charged phosphate groups in the DNA strands, facilitate the DNA melting equilibrium process on which separation is based. It is clear that in the CDCE system, the undesirable effect of low electrolyte concentration on the separation negates its beneficial effect of producing less heat. One potential solution could be to replace TBE with a buffer having a lower electric conductance but with equal or higher ionic strength (Reijenga et al., 1996).

Table 8. Experimental and calculated parameters for CDCE in capillaries of various bore widths.

Capillary i.d. ( $\mu\text{m}$ )	Current ( $\mu\text{A}$ )	Field Strength ( $\text{V/cm}$ )	Temperature ( $^{\circ}\text{C}$ )	Calculated $\Delta T_{\text{wb}}/\Delta T_{75}^{\text{a}}$
75	9	260	62.5	/
250	40	100	62.5	1.6
320	50	73	62.5	1.4
540	80	50	62.6	1.9

<sup>a</sup> According to equations (1) and (2),  $\Delta T_{\text{wb}}/\Delta T_{75} = \{(\text{i.d. of wb}/75 \mu\text{m}) \cdot (E_{\text{wb}}/E_{75})\}^2$ .

$\Delta T_{\text{wb}}$ : temperature difference inside a wide-bore capillary,

$\Delta T_{75}$ : temperature difference inside a 75  $\mu\text{m}$  capillary,

$E_{\text{wb}}$ : electric field strength applied across a wide-bore capillary,

$E_{75}$ : electric field strength applied across a 75  $\mu\text{m}$  capillary.

#### 4.5.3. Determination of the DNA loading capacities for capillaries of various sizes.

To determine the DNA loading capacities for capillaries as a function of bore diameter, various amounts of restriction digested human DNA containing the set of fluorescein labeled PCR products were introduced into each capillary through electrokinetic injection and the quality of separation was evaluated. Loading was conducted at 50 V/cm for 2 minutes through a Teflon tube tightly mounted onto the end of capillary. This method was able to completely load known copy numbers of fluorescein-labeled DNA fragments (Khrapko et al., 1994). Our expectation was that loading capacity would be approximately proportional to the cross sectional area of the capillary (Rose and Jorgenson, 1988).

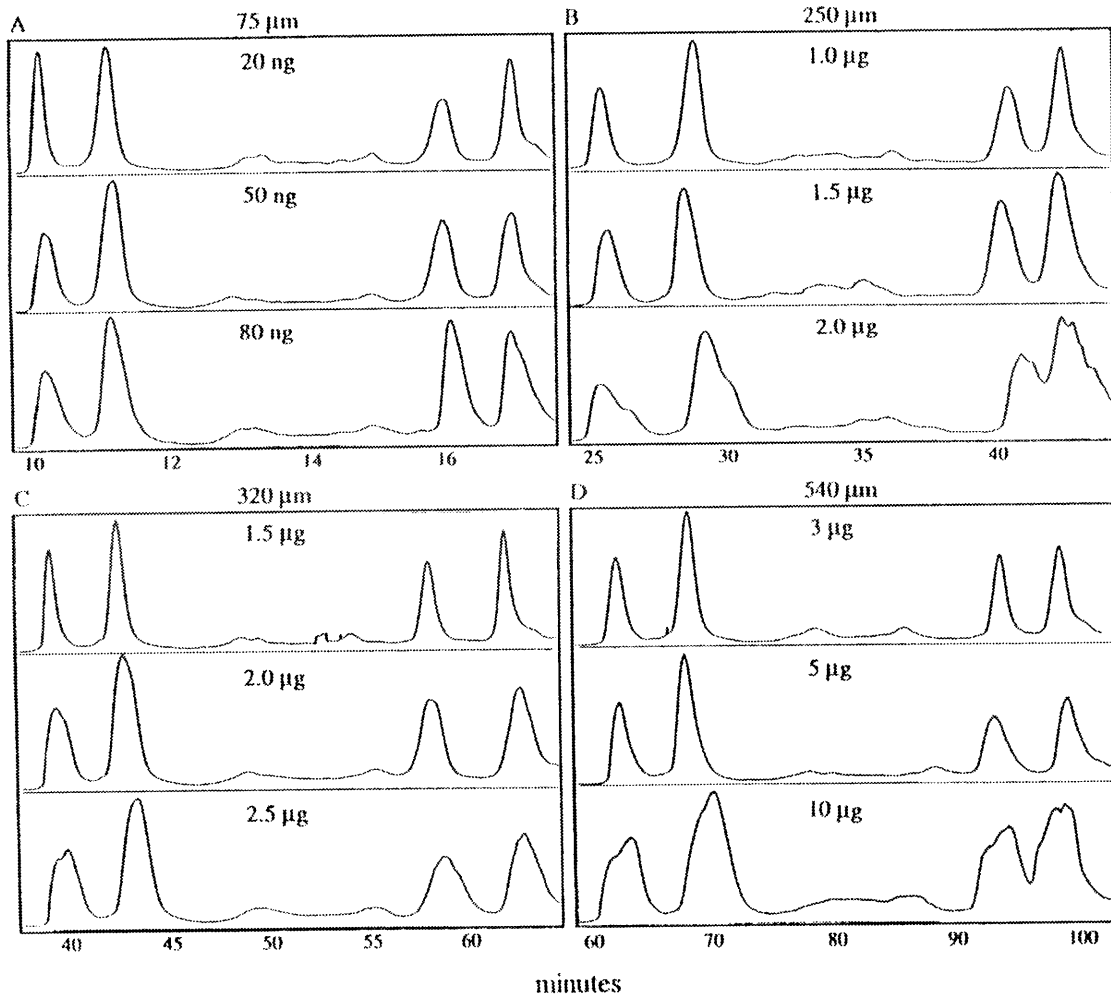
Electrophoresis was performed under the previously defined operational conditions for each capillary bore width (Fig. 22). The electropherograms show that the loading capacity of CDCE was indeed increased with bore width (Fig. 23). A 75  $\mu\text{m}$  capillary shows good separations up to 80 ng (Fig. 23A), a 250  $\mu\text{m}$  bore up to 2.0  $\mu\text{g}$  (Fig. 23B), a 320  $\mu\text{m}$  bore up to 2.5  $\mu\text{g}$  (Fig. 23C) and a 540  $\mu\text{m}$  bore even up to 10  $\mu\text{g}$  (Fig. 23D). It can be seen, however, that at high sample loads, peaks become broader and/or distorted. Larger amounts than those shown resulted in a current drop in all capillaries. The deterioration of resolution and instability of current at excessive sample loading could be explained by the difference in electrical conductivity of the DNA and the electrophoretic medium (Mikkers et al., 1979; Jorgenson et al., 1983; Lauer et al., 1986). High sample concentration in the capillary column can cause disturbances in the local potential gradient. Local over-heating in the sample zone can occur, resulting in broadened and distorted peaks.

One could therefore define DNA loading capacity as the maximum amount of DNA that the capillary can accommodate without observable perturbation to the potential and current while still showing adequate resolution. Using this definition for restriction digested human genomic DNA, the DNA loading capacities for capillaries of the various

Figure 23. CDCE separation performance as a function of amount of DNA loaded on capillaries of various sizes.

(A) 75  $\mu\text{m}$ , 20, 50 and 80 ng. (B) 250  $\mu\text{m}$ , 1.0, 1.5 and 2.0  $\mu\text{g}$ . (C) 320  $\mu\text{m}$ , 1.5, 2.0 and 2.5  $\mu\text{g}$ . (D) 540  $\mu\text{m}$ , 3, 5 and 10  $\mu\text{g}$ .

EFFECT OF AMOUNT OF DNA LOADED ON  
RESOLUTION IN CAPILLARIES OF VARIOUS SIZES



bores were found to be 80 ng for 75  $\mu\text{m}$ , 2.0  $\mu\text{g}$  for 250  $\mu\text{m}$ , 2.5  $\mu\text{g}$  for 320  $\mu\text{m}$ , and 10  $\mu\text{g}$  for 540  $\mu\text{m}$ , respectively.

#### 4.5.4. Demonstration of the detection sensitivity in reconstruction experiments

The efficiency of mutant enrichment by CDCE in a 540- $\mu\text{m}$  capillary was first tested in reconstruction experiments.  $2 \times 10^8$  copies of CDCE-purified wild-type *APC* gene fragments (271-bp) were doped with known copies of artificial mutant/wild-type heteroduplex fragments (271-bp) at a mutant fraction of 0,  $10^{-6}$  and  $10^{-5}$ , respectively. About 2  $\mu\text{g}$  of digested genomic DNA was added to each sample to mimic the presence of residual amounts of cellular DNA known to exist after sequence-specific enrichment of the target sequence.

The samples were separated by CDCE in a 19.5-cm long, 540  $\mu\text{m}$  i.d. column inserted into a 6-cm water-jacket. The electrophoresis was conducted at a constant current of 80  $\mu\text{A}$ . An appropriate temperature was used at which all heteroduplexes coalesced within a single fraction that was well separated from the wild-type homoduplex. The desired heteroduplex fraction was then collected. By comparing the copy number of the target sequence loaded onto the column to that in the heteroduplex collection, the enrichment efficiency of heteroduplexes was estimated to be 200 - 250-fold.

The heteroduplex collections were subjected to high-fidelity PCR. The amplified mutants were further enriched and amplified by two successive rounds of CDCE (in a 75  $\mu\text{m}$  i.d. column) coupled with high-fidelity PCR, and finally displayed on CDCE (Fig. 24). The internal standard heteroduplexes at an initial fraction of  $10^{-6}$  were clearly observable above the background PCR noise (Fig. 24). The results of the reconstruction experiments demonstrated that a highly efficient enrichment of mutant heteroduplexes could be obtained by CDCE separation using a 540  $\mu\text{m}$  i.d. capillary in order to observe mutations at fractions as low as  $10^{-6}$ .

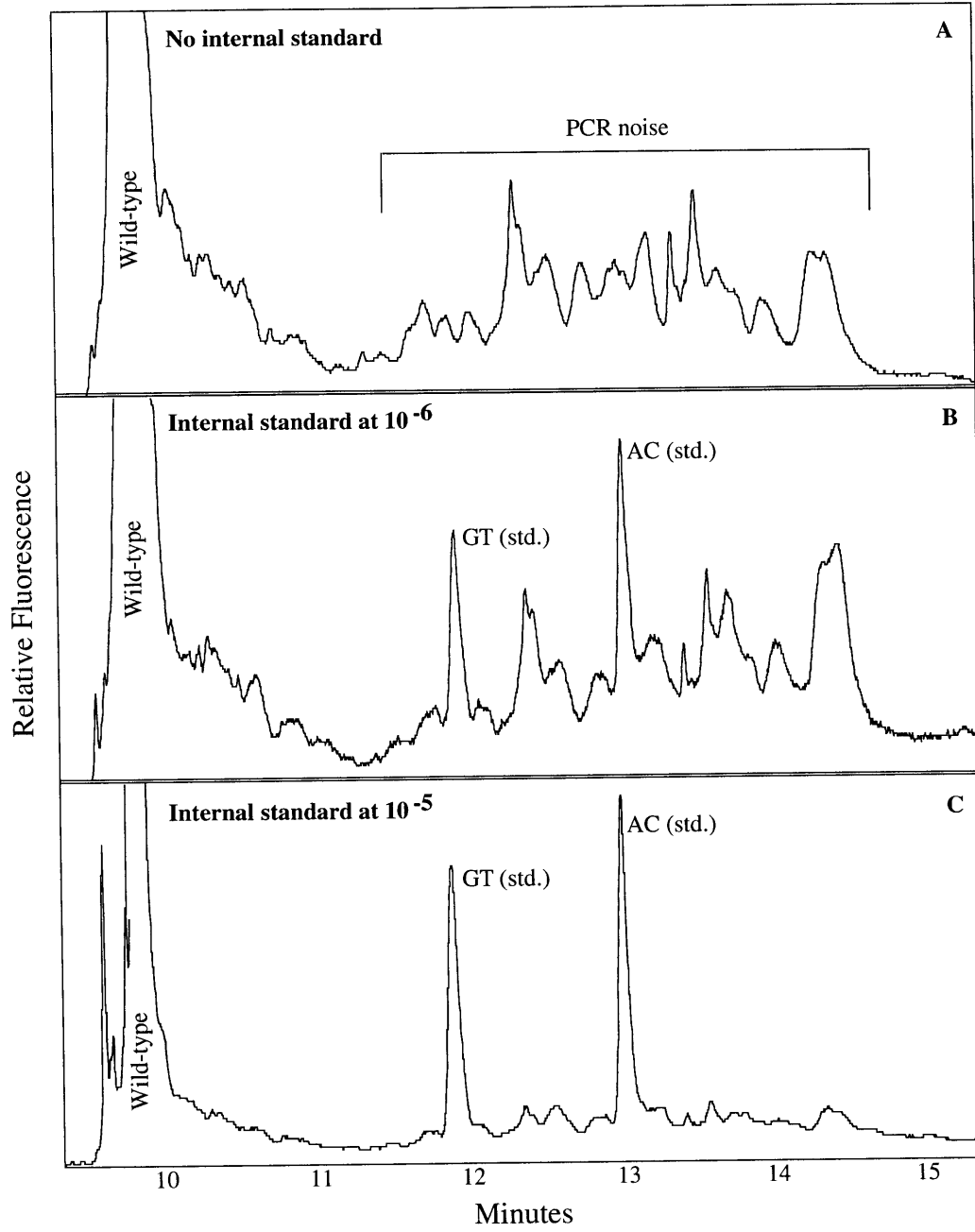


Figure 24. Reconstruction experiments demonstrating the detection sensitivity using wide-bore (540  $\mu\text{m}$  i.d.) CDCE for pre-PCR mutant enrichment.

CDCE-purified wild-type DNA samples doped with internal standard at 0,  $10^{-6}$  and  $10^{-5}$ , respectively, were subjected to the procedure of wide-bore CDCE  $\rightarrow$  high-fidelity PCR  $\rightarrow$  2 x [CDCE-PCR]. Depicted are the final CDCE separations of mutant heteroduplexes. The wild-type homoduplex and the internal standard heteroduplexes (GT and AC) are shown as indicated.

# Reconstruction experiments using CDCE-purified DNA

Wide-bore CDCE -> high-fidelity PCR -> 2 x [CDCE-PCR]



#### 4.5.5. Enrichment of mutant/wild-type heteroduplexes from target-enriched genomic DNA

In order to use CDCE to separate unlabeled cellular DNA, optimal separation temperatures and the collection times for the desired fractions had to be pre-determined using fluorescein-labeled PCR fragments. The same CDCE conditions were then applied to the cellular DNA in "blind" runs. The performance of the separation was then checked by measuring the copy number of the target sequence in each of the collected fractions.

When the cellular DNA enriched for the 271-bp *APC* gene fragment was subjected to CDCE in a 540  $\mu\text{m}$  capillary under optimal conditions, 5% of the wild-type DNA loaded onto the column was found co-eluted in the heteroduplex fraction. That is, the efficiency of enriching mutant heteroduplexes was only 20-fold as compared to over 200-fold as observed in the reconstruction experiments with CDCE-purified wild-type DNA. Based on the level of PCR noise (Fig. 24), a 20-fold pre-PCR enrichment would only allow observation of mutants at fractions at or above  $10^{-5}$ . It was clear that further enrichment of the cellular mutants prior to high-fidelity PCR was necessary. For further CE separation, the collected heteroduplex fraction was desalted by drop dialysis and concentrated through speed-vac.

At first, a second CDCE was attempted to further separate the heteroduplexes from the residual wild-type sequences. However, it was found that the mutants in the first heteroduplex fraction were enriched to less than 2-fold following the second CDCE step. The fact that the second CDCE failed to separate the residual wild-type DNA from the heteroduplexes suggested that the main component of the wild-type DNA in the CDCE heteroduplex was *not* part of the homogenous "tail" of the normal wild-type homoduplex peak. This residual wild-type DNA could exist in other conformations which had a electrophoretic mobility that coincided with the heteroduplexes on CDCE.

CE at room temperature was then used to separate the first heteroduplex fraction. It turned out that only 10 - 20% of the residual wild-type sequences were found in the

fraction containing the expected 271-bp target fragment, while the remainder were found distributed in fractions collected after the 271-bp double-strand peak. This indicated that the majority of the wild-type sequences which migrated within the heteroduplex fraction on CDCE could be of longer length and/or contained single-strand structure. These wild-type sequences could have been generated due to incomplete restriction digestion, heat-induced strand breaks and/or non-specific reannealing during the earlier procedures. The second separation by CE at room temperature thus achieved an additional 5 - 10-fold enrichment of the mutants.

Taken together, it was found that through a combination of CDCE and CE at room temperature in a 540  $\mu\text{m}$  i.d. column, the rare mutant sequences in the target-enriched cellular DNA samples could be enriched 100 - 200-fold against the abundant wild-type DNA. This high efficiency of enrichment prior to PCR was necessary in order to observe mutants at fractions as low as  $10^{-6}$ . As compared to the slab gel CDGE technique, the major advantage of CDCE is its ability to recover the desired fractions through on-column electroelution with a high yield (almost 100%). This high efficiency of recovery permits two consecutive CE separation steps to be carried out for enriching the mutant sequences in the target-enriched cellular DNA.

## **4.6. The CDCE/high-fidelity PCR approach to mutational spectrometry of the *APC* target sequence in human cells: application to MNNG-treated MT1 cells**

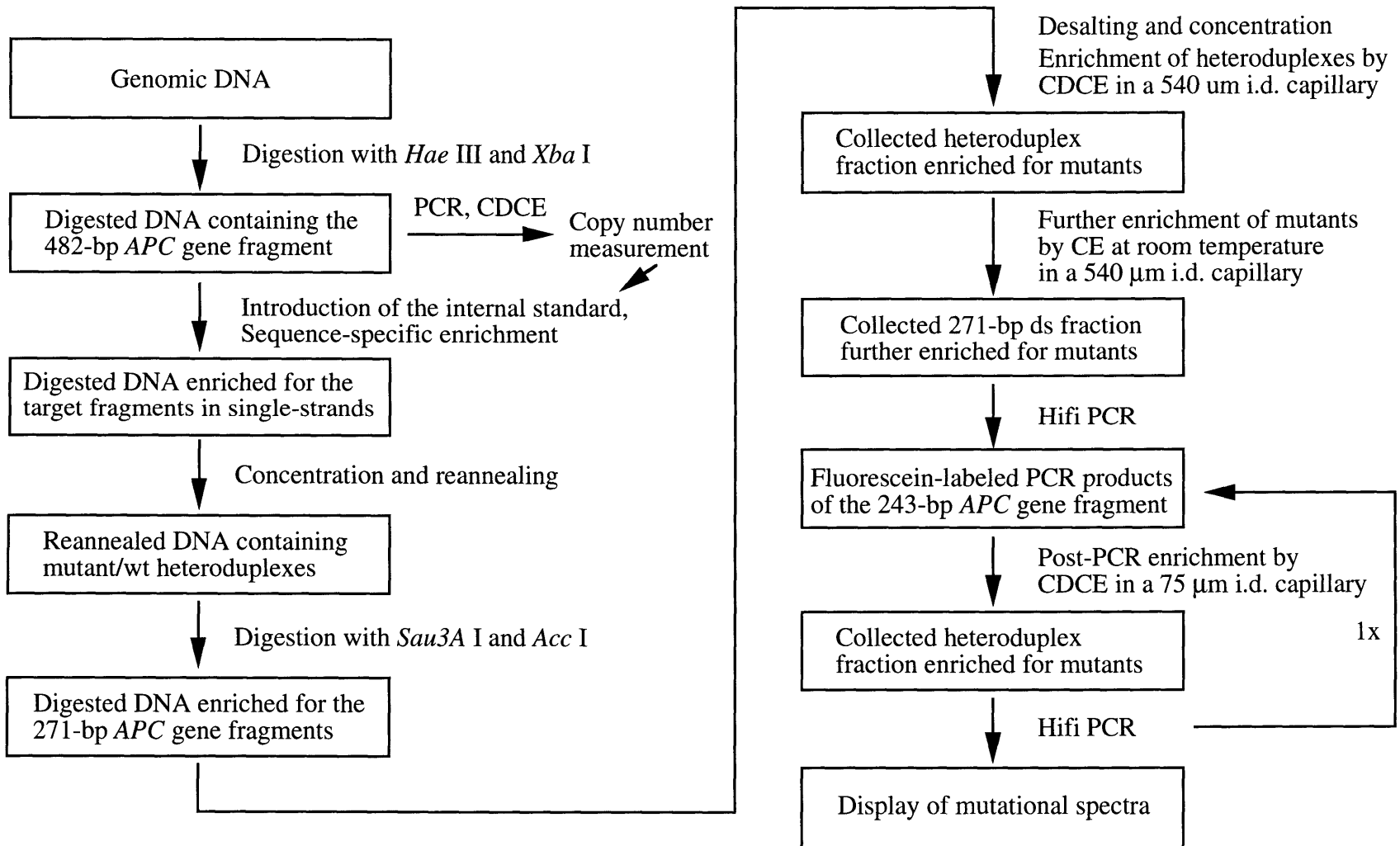
### 4.6.1. Overview of the approach

The improved CDCE/high-fidelity PCR approach incorporated the above described two technical developments which involved the use of sequence-specific hybridization to enrich for the target sequences and wide-bore CE to enrich for mutant sequences prior to PCR. In this approach (Fig. 25), large quantities (up to 10 mg) of genomic DNA were digested with two "inexpensive" endonucleases *Hae* III and *Xba* I to liberate the target sequence embedded in a 482-bp *APC* gene fragment (cDNA bp 8422 - 8903). The copy number of the target sequence was measured by quantitative PCR coupled with CDCE. Known copies of a 492-bp artificial mutant used as an internal standard were then introduced into the sample for determination of the initial fractions of the cellular mutants observed at the end of the procedures. The desired *APC* gene fragments were enriched from the bulk genomic DNA digest through sequence-specific hybridization coupled with magnetic separation. The target-enriched cellular DNA was concentrated and reannealed to form double-strands and resulted in rare mutant sequences being converted into mutant/wild-type heteroduplexes. The sample was then digested with two "expensive" endonucleases *Sau3A* I and *Acc* I to release the desired 271-bp *APC* gene fragment (cDNA bp 8434 - 8704) suitable for CDCE separation.

After desalting and concentration, the sample was subjected to wide-bore CDCE in a 540  $\mu\text{m}$  i.d. capillary to separate the heteroduplexes from an excess of wild-type homoduplex in the presence of residual cellular DNA. CDCE was performed at an optimal temperature at which all heteroduplexes coalesced within a single fraction that was well separated from the wild-type homoduplex. All heteroduplexes were electroeluted into a

Figure 25. Flow diagram of the CDCE/high-fidelity PCR approach to mutational spectrometry of the 121-bp *APC* gene target sequence in human cells.

See text for details.



single fraction. The collected heteroduplex fraction was desalted by drop dialysis and subjected to wide-bore CE at room temperature to further separate the heteroduplexes from the residual wild-type DNA in unknown "cryptic" forms. The fraction containing the normal 271-bp *APC* gene fragments was collected to further enrich the mutant sequences.

The mutant-enriched cellular DNA was amplified by high-fidelity PCR using native *Pfu* DNA polymerase to yield fluorescein-labeled 243-bp *APC* gene fragments (cDNA bp 8441-8683). The PCR products were subjected to CDCE in a 75  $\mu\text{m}$  i.d. capillary to separate the mutant/wild-type heteroduplexes from the wild-type homoduplex. The heteroduplexes were collected in a single fraction and again amplified by high-fidelity PCR. The CDCE/high-fidelity PCR procedure was repeated once more to further enrich the mutants. Finally, the mutant and wild-type homoduplex mixtures were separated on CDCE. Mutant peaks were then individually isolated and sequenced to obtain the mutational spectrum.

#### 4.6.2. Application to MNNG-treated and untreated MT1 cells

The mutational spectrometry approach described above was tested by measuring MNNG induced point mutations in the 121-bp *APC* target sequence in human lymphoblastoid MT1 cells. MT1 cells are highly resistant to the cytotoxicity of MNNG, but remain sensitive to its mutagenicity, a feature that allows treatment of the cells with a large dosage of MNNG to induce high mutant fractions in nuclear genes (Goldmacher et al., 1986; Kat et al., 1993). In this study, exponentially growing MT1 cultures consisting of  $8 \times 10^7$  cells were exposed to 4  $\mu\text{M}$  MNNG for 45 min. This treatment killed 20% of the cells and induced a *hprt* mutant fraction of  $8 \times 10^{-3}$ . After treatment the cells were grown for 30 doublings with daily dilution to dilute unrepaired premutagenic lesions ( $\sim 5 \times 10^4$  *O*<sup>6</sup>-meG/cell) (Goldmacher et al., 1986) which could be mistaken for mutations by the procedures.



Approximately 4.5 mg of genomic DNA was isolated from  $5 \times 10^8$  MT1 cells. After digestion with *Hae* III/*Xba* I, the internal standard mutant was introduced at an initial fraction of  $10^{-5}$  in one MNNG-treated and two untreated MT1 genomic DNA samples, and  $2 \times 10^{-5}$  in the other MNNG-treated MT1 genomic DNA sample. Subsequent sequence-specific enrichment achieved over 10,000-fold enrichment of the target sequences with a yield greater than 70%. Thus the amount of cellular DNA in each sample was reduced to less than 1  $\mu$ g containing  $7 \times 10^8$  copies of the target sequences. Prior to CE separation, the sample was desalted and concentrated through ultrafiltration in order to be electro-injected on the capillary. This procedure resulted in 50% yield of the target sequences.

About  $2 \times 10^8$  copies of the target sequences from each sample including a CDCE-purified wild-type DNA control were subjected to wide-bore CDCE and then CE at room temperature. Based on measurements of the copy number of the target sequence loaded on the column and that eluted in the mutant-containing fraction, the efficiency of mutant enrichment after the two CE separation procedures was determined to be 100 - 200-fold. The mutant-enriched samples which contained approximately  $10^6$  copies of target sequences were amplified by high-fidelity PCR using primers AP1 and AP4H to yield nearly  $2.5 \times 10^{12}$  copies of fluorescein-labeled *APC* gene fragment. The amplified mutant sequences in their heteroduplex forms were further enriched 100-fold by two rounds of CDCE coupled with high-fidelity PCR (Fig. 26). The mutants were converted into homoduplexes following a few cycles of PCR with an excess of remaining primers to prevent the formation of heteroduplexes. In order to separate the complex mixtures of homoduplexes on CDCE at a high resolution, separations were operated at a low electric field strength and using an extended water-jacket to increase the average number of partial meltings and reannealings a molecule undergoes while in the heated zone of the capillary.

Fig. 27 shows the final CDCE separations of mutant homoduplexes derived from the CDCE-purified wild-type DNA, two untreated and two MNNG-treated MT1 cultures. It can be seen that both MNNG-treated MT1 samples contained a distinct set of low  $T_m$

Figure 26. Post-PCR enrichment of the amplified mutant sequences by CDCE coupled with high-fidelity PCR.

Depicted are CDCE separations of fluorescein-labeled PCR products (about  $2 \times 10^8$  copies of the target fragments) derived from the MNNG-treated MT1 cells. The wild-type peak in both electropherograms is shown in 1/10 of its full height. (A) After CDCE and CE at room temperature in a 540  $\mu\text{m}$  i.d. capillary, the mutant-enriched sample was subjected to 40 cycles of high-fidelity PCR. The PCR products were separated by CDCE in a 75  $\mu\text{m}$  i.d. capillary at 9  $\mu\text{A}$  and using a 6-cm water-jacket at 64.6°C. The heteroduplex fraction was collected. The broad peak marked by 'X' is a PCR byproduct generated after direct amplification of genomic DNA. (B) The above collected heteroduplex fraction was subjected to 38 cycles of PCR and separated by CDCE. The heteroduplex fraction was then collected and amplified to further enrich the mutants.

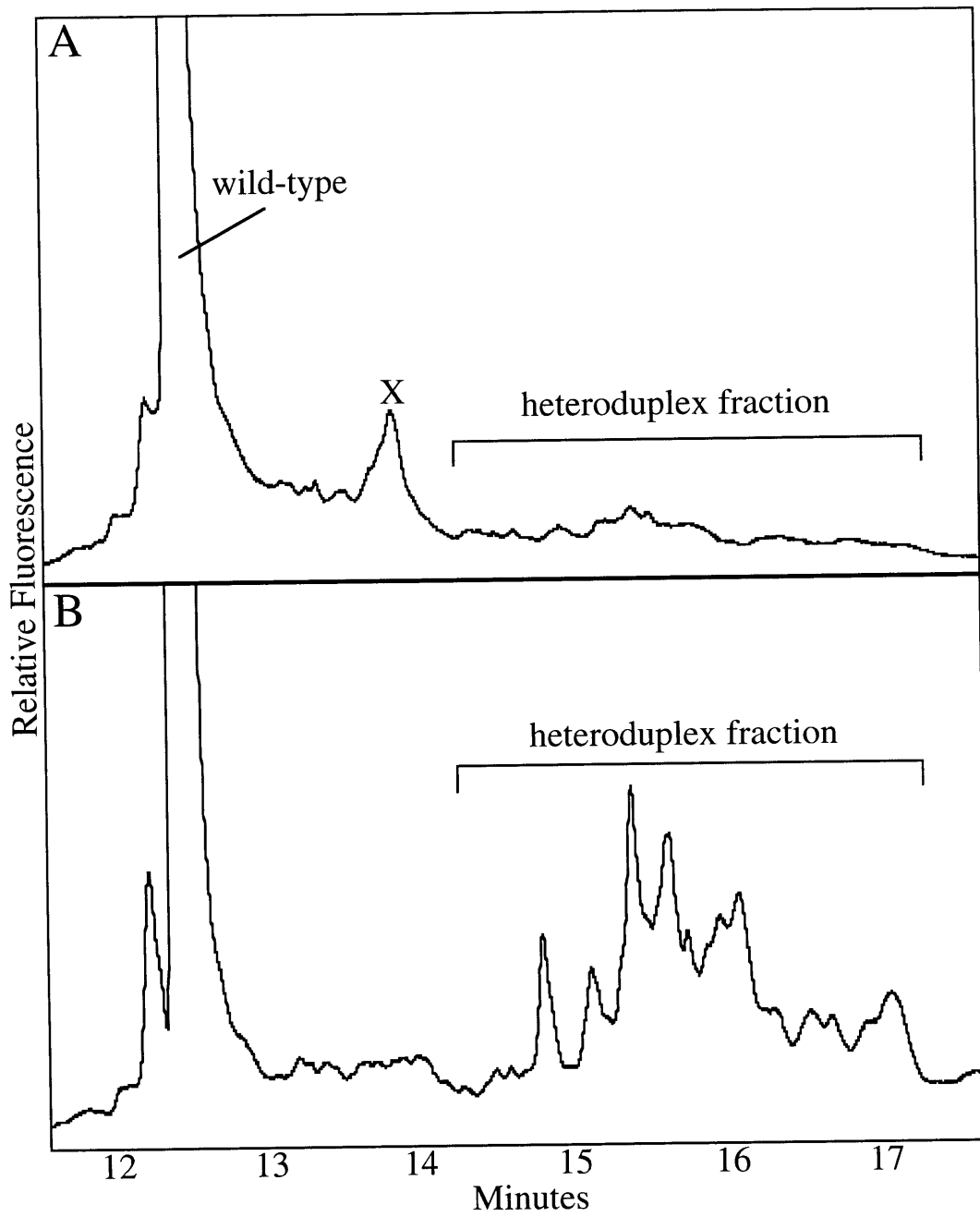
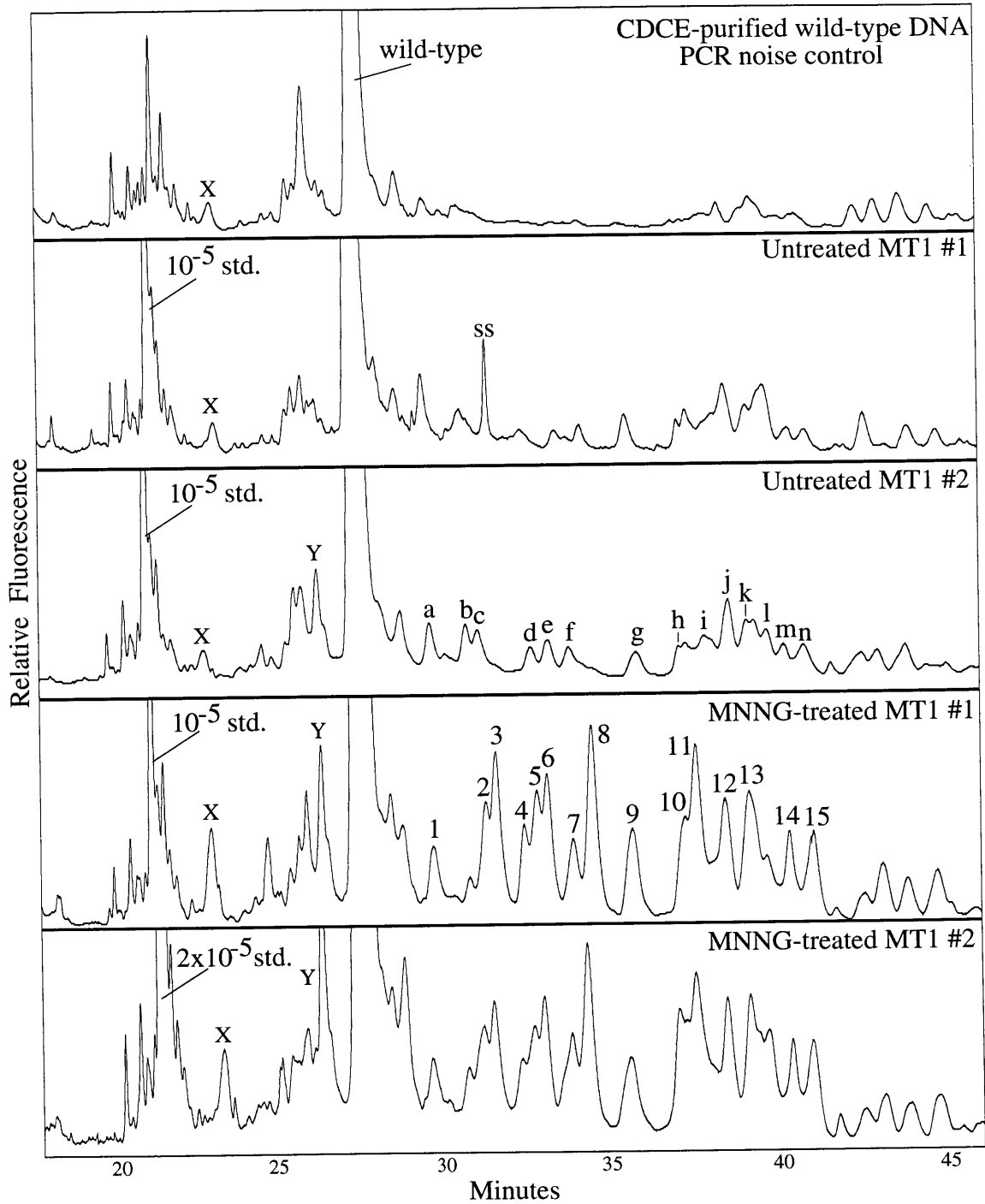


Figure 27. CDCE separations of the mutant homoduplexes derived from the CDCE-purified wild-type DNA, two untreated and two MNNG-treated MT1 genomic DNA samples.

CDCE was performed at 8  $\mu$ A and using a 19-cm water-jacket at 65.20°C. The wild-type peak and the internal standard (std.) at its initial mutant fraction in each cellular sample are indicated. The std. peak is shown at 1/2 of its full height in "Untreated MT1 #1", "Untreated MT1 #2" and "MNNG-treated MT1 #1" samples, and 1/4 of its full height in "MNNG-treated MT1 #2". Mutant peaks that have been isolated and sequenced are designated by letters (*a* to *n*) in the untreated sample and numbers (1 to 15) in the treated sample. "SS" marks single-strand DNA. "X" and "Y" indicate unknown PCR artifacts.

**Wide-bore CDCE -> high-fidelity PCR -> 2 x [CDCE-PCR]**



mutant peaks above the background mutations observed in the CDCE-purified wild-type DNA control and two untreated MT1 samples. The peak pattern exhibited excellent reproducibility in each replicate samples. In the high  $T_m$  mutant region, except for peaks X and Y, no significant difference was observed among the analyzed samples (Fig. 27). Peaks X and Y were proven to be PCR artifacts since they were also observed in PCR products directly amplified from genomic DNA (see Fig. 33).

#### 4.6.3. Identification of the mutant peaks

The individual mutant peaks displayed on the CDCE electropherogram (Fig. 27) were isolated and sequenced. Typically two to four rounds of CDCE collection and PCR were required to adequately purify a mutant peak for DNA sequencing. To confirm the isolated mutant was indeed the desired one, the purified product was co-injected with the original mutant mixture and run on CDCE to verify comigration with the desired peak. The PCR products of the purified mutants were desalted and removed from residual primers by ultrafiltration before being submitted for sequencing.

Fourteen mutations designated by letters *a* to *n* in the untreated MT1 sample and fifteen mutants designated by numbers 1 to 15 in the MNNG-treated MT1 sample (Fig. 27) were purified and sequenced. All of the mutants were single base pair substitutions as shown in the target sequence context in Fig. 28. All of the fourteen mutants in the untreated sample were GC → TA transversions. In contrast, twelve out of the fifteen mutants identified in the treated sample were GC → AT transitions. The three exceptions (peaks 1, 10 and 12) were GC → TA transversions which were identical to peaks *a*, *h* and *j*, respectively, in the untreated control, suggesting that they belong to the set of background mutations.

Figure 28. Distribution of base substitutions in the *APC* gene target sequence (cDNA bp 8543-8663) in MNNG-treated and untreated MT1 cells.

Mutations from the MNNG-treated MT1 cells (labeled with numbers) are shown above the wild-type sequence, while those from the untreated MT1 cells (labeled with letters) are below the sequence. The mutation labels correspond to those in Fig. 27. The base substitution of the internal standard (std.) is also indicated.





#### 4.6.4. MNNG-induced point mutational spectrum in the APC target sequence

The twelve GC → AT transitions which were reproducibly observed in the MNNG-treated samples but absent in the untreated controls were consistent with the miscoding potential of *O*<sup>6</sup>-meG lesions produced by MNNG (Loveless, 1969; Loechler *et al.*, 1984). By comparison to the area under the internal standard peak, the total mutant fraction induced by MNNG in MT1 cells was estimated to be  $8.2 \times 10^{-5}$  after subtracting the background signals derived from the untreated cells.

To obtain an accurate measurement of the area under each mutant peak, CDCE was performed at various temperatures, and for each peak to be measured, the appropriate electropherogram showing the best separation of this peak from its adjacent ones was chosen. By comparison of the area under each mutant peak with that under the internal standard peak, the mutant fractions of the 12 transition mutations were estimated to be between  $2.2 \times 10^{-6}$  and  $9.2 \times 10^{-6}$ , and are highly reproducible in the replicate samples (Table 9). Fig. 29 shows the MNNG-induced point mutational spectrum in the 121-bp *APC* gene target sequence with regard to kind, position and mutant fraction.

The 121 base pairs of the target sequence contain 29 guanine residues; 18 of these sites are preceded (5') by a purine, at which 11 (peaks 2, 3, 4, 6, 7, 8, 9, 11, 13, 14 and 15) were mutated at a total fraction of  $5.9 \times 10^{-5}$ . The remaining 11 guanine sites are preceded (5') by a pyrimidine, at which only one (peak 5) was mutated at a fraction of  $5.6 \times 10^{-6}$ . Assuming all of the GC → AT transitions resulted from *O*<sup>6</sup>-meG, then guanine residues preceded by a purine were on average more than 6 times as likely to be mutated ( $3.3 \times 10^{-6}$ /site) as those preceded by a pyrimidine ( $5 \times 10^{-7}$ /site) in the *APC* target sequence. This difference is highly significant ( $P < 0.001$  based on  $\chi^2$  tests).

An analysis of strand distribution of these transitions showed that 8 of the 12 mutated guanines (peak 2, 3, 4, 5, 7, 8, 9 and 14) resided on the non-transcribed strand which contains 19 guanines with 11 preceded by a purine. The other 4 mutated guanines

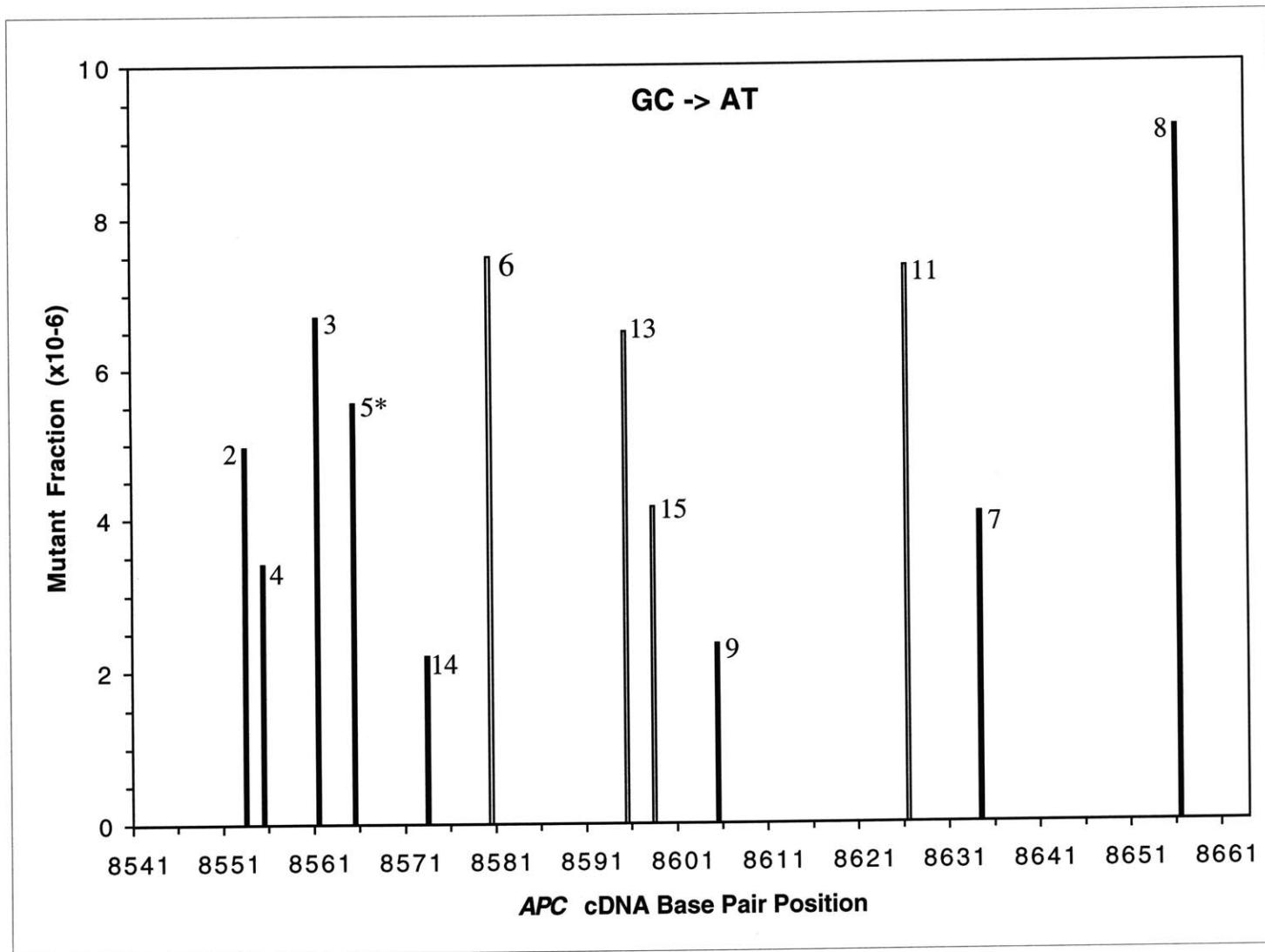
Table 9. Positions and estimated mutant fractions of the twelve MNNG-induced GC -> AT transitions in the two replicate MT1 cell cultures.

Mutation <sup>a</sup>	APC cDNA bp	Estimated mutant fraction ( $\times 10^{-6}$ )		
		MT1 #1	MT1 #2	Average
2	8553	5.1	4.8	5.0
3	8561	7.1	6.3	6.7
4	8555	4.0	2.8	3.4
5	8565	5.5	5.6	5.6
6	8580	7.4	7.6	7.5
7	8634	3.7	4.4	4.1
8	8656	9.1	9.2	9.2
9	8605	2.7	2.0	2.4
11	8626	6.8	7.9	7.4
13	8595	6.3	6.7	6.5
14	8573	2.3	2.1	2.2
15	8598	4.4	3.9	4.2

(a ) The mutant numbers correspond to those in Fig. 27.

Figure 29. MNNG-induced point mutational spectrum in the *APC* gene target sequence in MT1 cells.

The horizontal axis represents the *APC* target sequence from cDNA bp 8543 to 8663. Each mutant is labeled with a number corresponding to that in Fig. 27 and 28. The height of vertical bar represents the average fraction of each mutant in the replicate MNNG-treated MT1 cells. All mutants were GC → AT transitions. Except for one which is indicated by '\*', all transitions occurred at G residues preceded (5') by a purine. Filled bars represent mutants at guanine residues residing on the nontranscribed strand, open bars represent mutants at guanine residues on the transcribed strand.



(peak 6, 11, 13 and 15) were on the transcribed strand which contains 10 guanines with 7 preceded by a purine. The average mutant fraction of GC → AT transitions was  $2.1 \times 10^{-6}$ /site (or  $3.1 \times 10^{-6}$ /site preceded by a purine) on the nontranscribed strand, and  $2.5 \times 10^{-6}$ /site (or  $3.6 \times 10^{-6}$ /site preceded by a purine) on the transcribed strand. The MNNG-induced mutations in the *APC* target sequence therefore did not exhibit a significant strand bias.

#### 4.6.5. Comparison to previous MNNG mutagenesis studies

In our laboratory, MNNG-induced point mutational hotspots were previously investigated in the *hprt* gene in MT1 cells using phenotypic selection. MT1 cells were twice treated with 4  $\mu$ M MNNG which induced a *hprt* mutant fraction of  $1.2 \times 10^{-2}$  (Kat, 1992). Eleven GC → AT hotspots were identified with each representing 1 to 10% of the total 6TGR mutants (Kat, 1992). In this study, treatment of MT1 cells with a single dose of 4  $\mu$ M MNNG induced a *hprt* mutant fraction of  $8 \times 10^{-3}$ . Assuming the target size of the *hprt* gene is 1000-bp, then one induced hotspot at or above 1% of the mutant population would be expected to occur in a given 100-bp nuclear sequence based on Kat's study (1992). The mutant fraction of such a hotspot prior to phenotypic selection would be  $8 \times 10^{-5}$ . The 121-bp *APC* target sequence was vacant for such a prominent hotspot. The vacancy observed in this *APC* particular target sequence was reminiscent of that in the 100-bp *hprt* exon 3 low melting domain (Kat, 1992). Both sequences are low in GC content, which may make them relatively poor targets for MNNG mutagenesis. Nevertheless, owing to the high sensitivity of the CDCE/high-fidelity PCR approach, twelve MNNG-induced "cold" spots at fractions between  $2.2 \times 10^{-6}$  and  $9.2 \times 10^{-6}$  were detected in the *APC* target sequence.

The MNNG mutagenic specificity observed in this study can be compared with that obtained in bacteria, yeast and other human systems based on phenotypic selection

(Horsfall et al., 1990; Gordon et al., 1990; Burns et al., 1987; Richardson et al., 1987; Reed and Hutchinson, 1987; Kohalmi and Kunz, 1988; Yang et al., 1991; Lukash et al., 1991; Kat, 1992). The results in this study agree with the previous observations that MNNG specifically induces GC → AT transitions which predominantly occur at guanine residues preceded (5') by a purine. The kind of mutation is consistent with the miscoding potential of *O*<sup>6</sup>-meG (with thymine) which is believed to be the major premutagenic lesion produced by MNNG (Loveless, 1969; Loechler *et al.*, 1984). It has been suggested that the 5' flanking base may influence the initial deposition of *O*<sup>6</sup>-meG lesions and/or the fidelity of a DNA polymerase bypassing these lesions (Gordon et al., 1990).

An analysis of strand distribution of the twelve MNNG-induced transitions in the *APC* target sequence revealed no strand bias. The absence of a strand specificity has also been observed in other targets (Gordon et al., 1990; Kohalmi and Kunz, 1988). However, several studies including those on the endogenous *hprt* gene in human cells have reported a strand preference in favor of GC → AT transitions at guanines on the nontranscribed strand (Richardson et al., 1987; Reed and Hutchinson, 1987; Yang et al., 1991; Lukash et al., 1991; Kat, 1992). It has been suggested that the non-random distribution of MNNG-induced mutation in some targets could be a reflection of the influence of the mutation on protein structure and function (Gordon and Glickman, 1988). For example, a GC → AT transition in the Gly (5'-GGN-3') and Trp (5'-UGG-3') codon sequences is more likely to be selected as a hotspot because Gly residues are often strategic to protein structure and the transition in Trp coding sequence results in a translation termination signal. The results obtained in the 121-bp *APC* gene sequence without the use of phenotypic selection provide further support to the view that the strand specificity of MNNG-induced mutation in some target sequences may reflect an observation skewed by phenotypic selection.

#### 4.6.6. Sources for the background mutations observed in untreated cells

The background mutational spectrum in untreated MT1 cells consisted of fourteen identified GC → TA transversions ranging in mutant fraction from  $4 \times 10^{-7}$  to  $1.7 \times 10^{-6}$  as compared to the internal standard (Figs 27, 28). These background mutations could have come from several possible sources, including: (1) true spontaneous mutations in MT1 cells; (2) DNA adducts and/or replication mismatch intermediates existing in the cells which were later converted into mutations by *Pfu* DNA polymerase; (3) DNA lesions produced by the procedures; or (4) PCR-introduced noise including newly created mutations by *Pfu* polymerase misincorporation errors or DNA lesions generated during PCR which were converted into mutations.

To discover whether any of these transversions actually represent true spontaneous mutations in MT1 cells, genomic DNA of TK6 cells was subjected to the cascade of the entire procedure. TK6 is the isogenic *GTBP* normal parent line of MT1 with a spontaneous mutation rate that is 15-fold lower than that of MT1 (Goldmacher et al., 1986). Previous studies have shown that these two cell lines exhibit distinct spontaneous mutational spectra in the *hprt* gene (see Fig. 7 on Page 68) (Cariello et al., 1990; Kat, 1992; Kat et al., 1993; Tomita-Mitchell et al., unpublished results). In this study, however, the same set of transversion mutations in MT1 cells was also observed in TK6 genomic DNA (Fig. 30). The results indicate that these background mutations are unlikely to exist as true mutants in the cells.

When the low  $T_m$  mutant regions of the CDCE electropherograms (shown in Fig. 27) were magnified for the PCR noise control and the two untreated MT1 cells (Fig. 31), a great similarity in the peak pattern was observed among the three samples. These observations indicate that PCR-introduced noise could be a significant source for the background mutations. Three of the mutant peaks (designated by the symbols  $\alpha$ ,  $\beta$  and  $\delta$ ) derived from the CDCE-purified wild-type DNA were then isolated and sequenced (Fig. 31). All three mutations were found to consist of GC → TA transversions. Mutations  $\alpha$  and  $\beta$  were found to be identical to mutations *a* and *h*, respectively, isolated from the

Figure 30. CDCE separations of the mutant homoduplexes derived from TK6 cells.

Three milligrams of genomic DNA isolated from TK6 cells were subjected to the previously described mutational spectral analysis. CDCE was performed at 8  $\mu$ A using a 19-cm long water-jacket at a temperature of 64.40°C. The mutant peaks are labeled corresponding to those shown in Fig. 27.



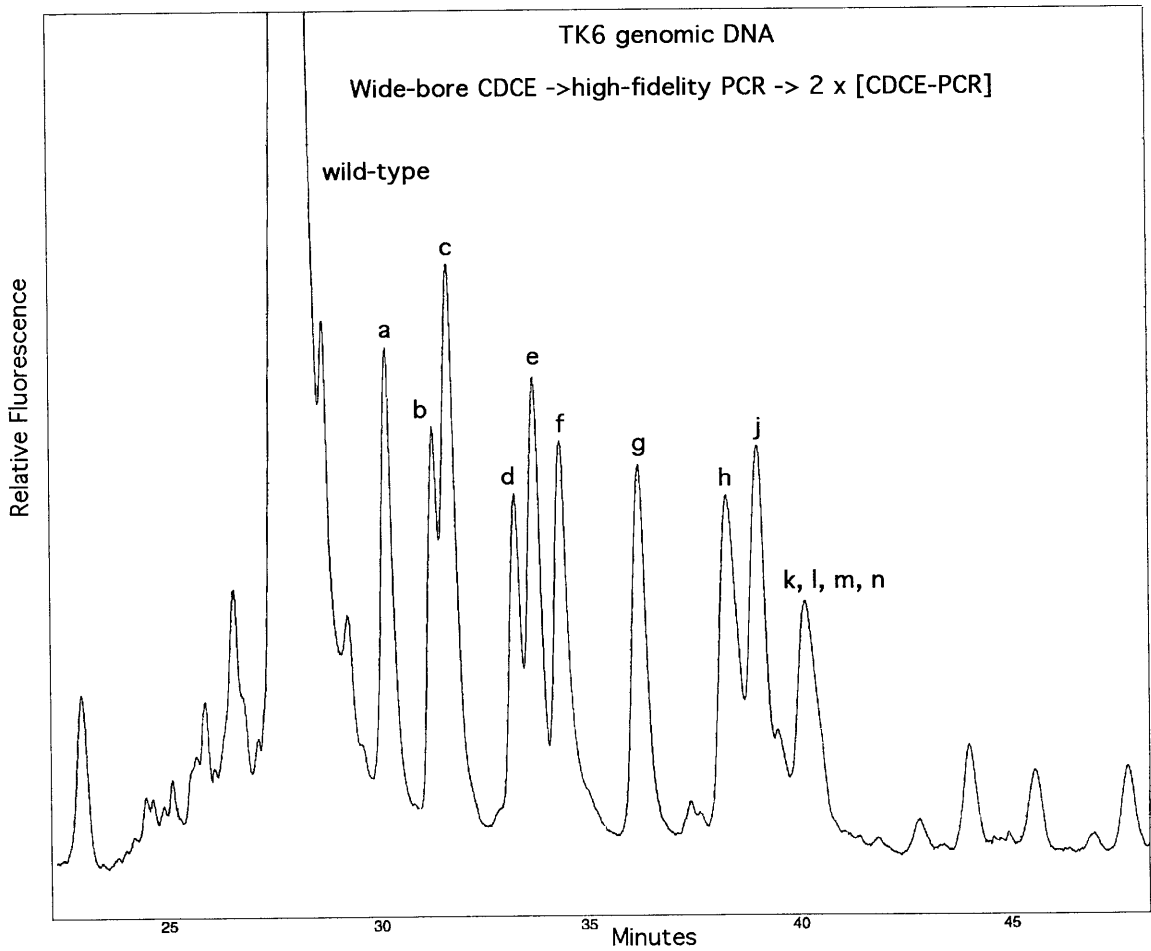
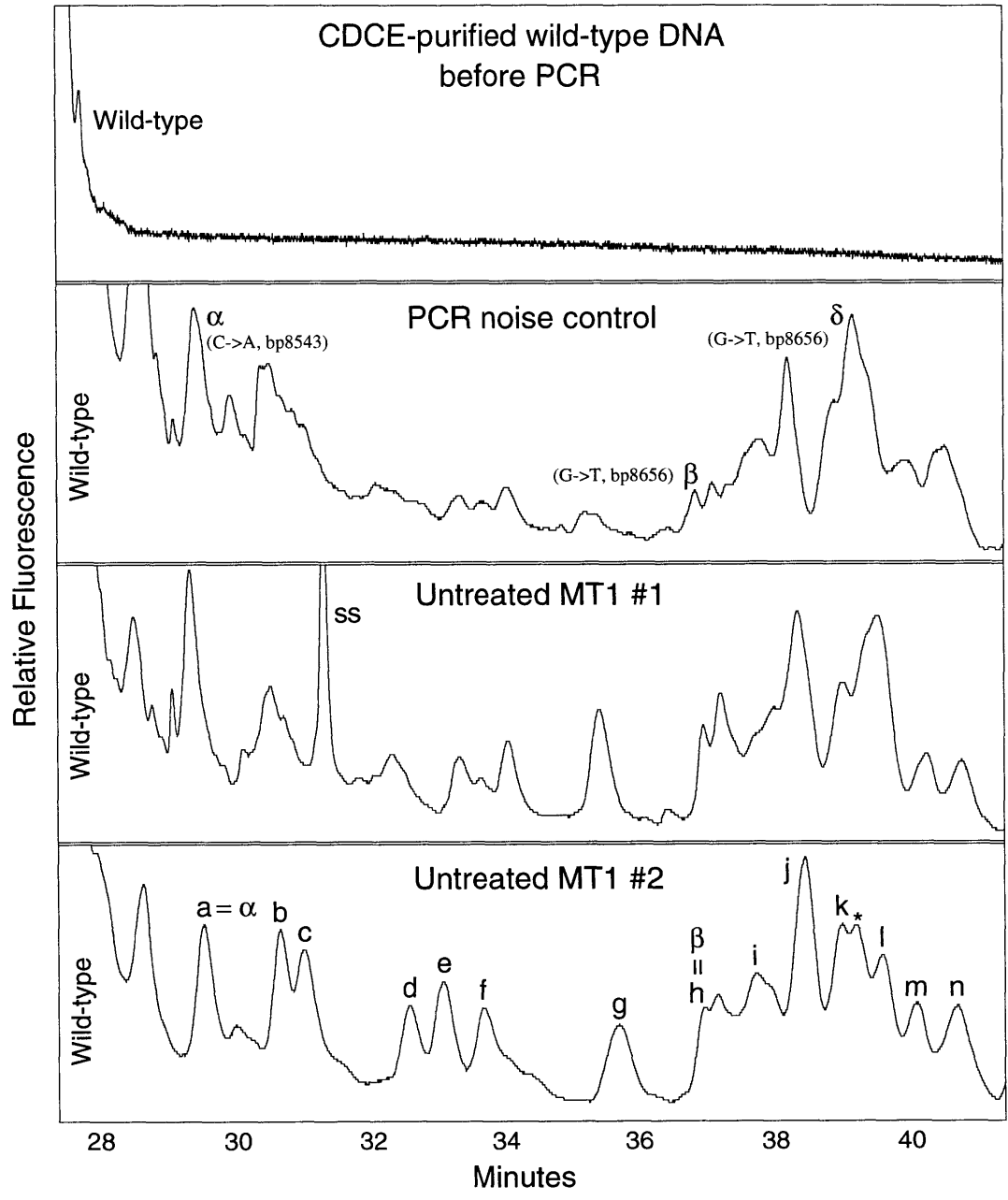


Figure 31. Comparison of CDCE separations of low  $T_m$  mutants derived from the CDCE-purified wild-type DNA and untreated MT 1 cells.

Depicted are the magnification of the regions corresponding to the migration time ranging between 27 and 42 minutes in the CDCE electropherograms shown in Fig. 27 and that of the CDCE-purified wild-type DNA before PCR. The kinds and positions of the three mutant peaks (designated by symbols  $\alpha$ ,  $\beta$  and  $\delta$ ) that have been identified in the PCR control are shown. "\*" indicates an unidentified peak observed in the untreated MT1 #2 sample that comigrated with peak  $\delta$ . "SS" marks single-strand.



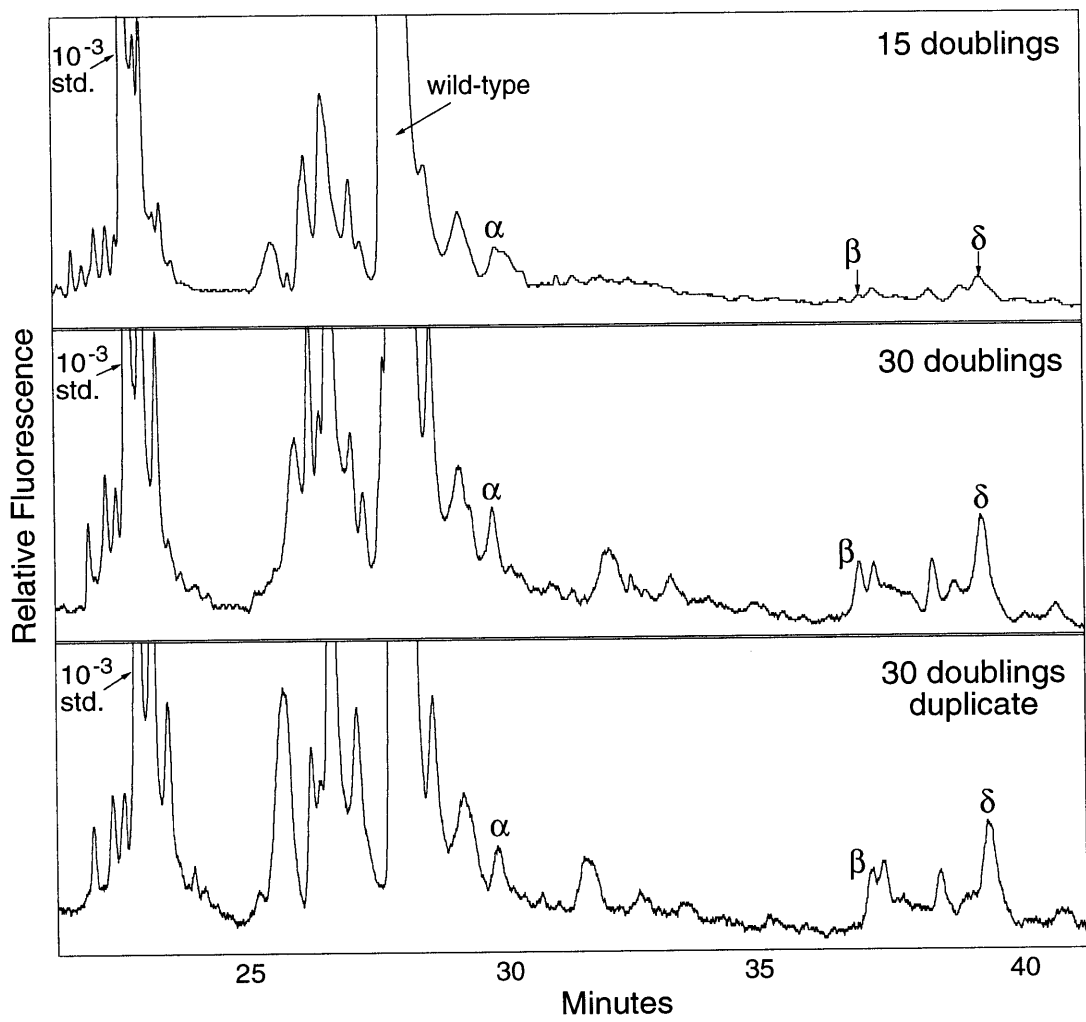
untreated cells. Mutation  $\delta$ , which occurred at the position of *APC* cDNA bp 8638, was not found among the fourteen mutations that were identified from the untreated cells. However, peak  $\delta$  seemed to comigrate with an unidentified peak (\*) located between peaks *k* and *l* in the untreated cells (Fig. 31).

To verify that the mutations derived from the purified wild-type DNA control were indeed created during the PCR process, independent PCR experiments were then performed. The wild-type sequence was amplified from TK6 genomic DNA using primers AP1 and AP4H and purified by CDCE. The purified wild-type DNA was then doped with an internal standard at a mutant fraction of  $10^{-3}$ . Using *Pfu* DNA polymerase,  $4 \times 10^7$  copies of the template were first amplified to a total of  $10^{12}$  copies (15 doublings). Then  $4 \times 10^7$  copies of the diluted PCR products were amplified to  $10^{12}$  copies to obtain another 15 doublings. The mutant sequences in the PCR products (after 15 and 30 doublings) were enriched by two rounds of CDCE coupled with PCR, and then separated on CDCE in their homoduplex forms (Fig. 32). It was observed that most of the mutant peaks (including peaks  $\alpha$ ,  $\beta$  and  $\delta$ ) increased in their mutant fractions from 15 to 30 doublings, proving that these peaks were created by PCR.

Interestingly, a GC  $\rightarrow$  TA transversion was also found to be one of the most prominent *Pfu*-induced mutational hotspots in a 96-bp human mitochondrial DNA (mitochondrial bp 10031-10126), and actually the most prominent hotpot among all of the low  $T_m$  mutant peaks observed in this sequence (André et al., 1997). The *Pfu*-induced GC  $\rightarrow$  TA transversions could be created by several possible mechanisms, including: (1) *Pfu* polymerase misincorporation errors and failure of its proofreading activity, (2) polymerase microheterogeneity (a subset of error-prone polymerase molecules due to mistranscription, mistranslation, or chemical modifications of the normal polymerase), (3) heat-induced mutagenesis (such as depurination), or (4) chemical modifications in the nucleotide pool (such as the spontaneous formation of 8-hydroxyguanine) (André et al., 1997). Preliminary studies showed that variations in denaturation time at 94°C did not affect the

Figure 32. CDCE separations of mutant homoduplexes induced by PCR using *Pfu* DNA polymerase after 15 and 30 doublings.

CDCE was performed at 8  $\mu$ A and using a 19-cm water-jacket at 64.50°C. The wild-type DNA and the internal standard (std.) which was at an initial fraction of  $10^{-3}$  are shown. The std. peak is shown at 1/4 of its full height in all samples. The mutant peaks that have been identified are labeled by the symbols  $\alpha$ ,  $\beta$  and  $\delta$ . The mutant labels correspond to Fig. 30. The duplicate sample for 30 doublings is shown here to demonstrate the reproducibility of the experiments.



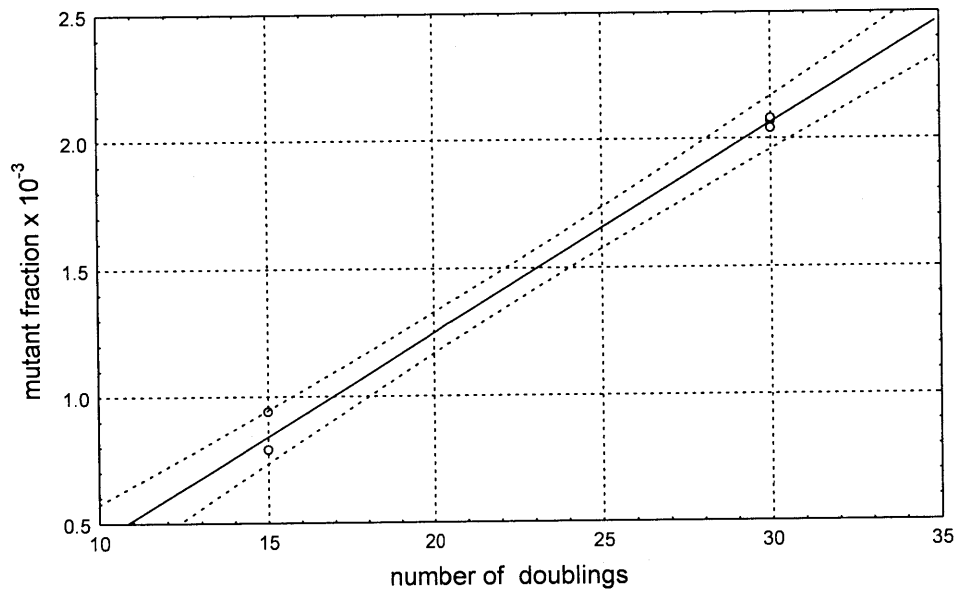
level of the mutant fraction induced by PCR, suggesting that heat-induced DNA lesions are unlikely to be the main origin of PCR noise (Paulo André, personal communication).

The mutant fractions of the low  $T_m$  mutants (between 29 - 41 min) in the independent PCR experiments were plotted as a function of the number of doublings in Fig. 33. Using linear regression analysis, the slope was determined to be  $8.2 \times 10^{-3}$  total low  $T_m$  mutations per doubling (Fig. 33). In comparison, in both untreated MT1 samples, 21 doublings (from  $10^6$  to  $2.5 \times 10^{12}$  copies) was performed to amplify the mutant-enriched cellular DNA. Given that the internal standard was present at a mutant fraction of  $\sim 10^{-3}$  after pre-PCR mutant enrichment, the total mutant fraction of the low  $T_m$  mutants (between 29 - 41 min) was estimated to be  $2.2 - 2.4 \times 10^{-3}$  in the final PCR products (Fig. 27). The observed mutant fraction in the untreated MT1 samples was somewhat higher than that would be expected from 21 doublings of PCR ( $1.3 \pm 0.1 \times 10^{-5}$ ) based on the results of the independent PCR experiments (Fig. 33). The differences in the mutant fraction could be due to the different batches of *Pfu* polymerase used in the two sets of experiments, implicating that polymerase microheterogeneity might play a role. However, the possibility that other sources might also contribute to the observed background mutations in the untreated cells can not be ruled out.

It is important to note that some of these background mutations could have arisen from replication mismatch intermediates or DNA lesions either preexisting in the cells or produced by the experimental procedures themselves. In our previous studies of mitochondrial DNA mutational spectra, evidence was found to suggest that some of the mutations might have arisen from cellular DNA adducts or replication mismatch intermediates existing in the cells prior to DNA isolation (Khrapko et al., 1997a). Resolution of this issue in the nuclear gene study must await application of the previously described procedure in which mutant fractions are independently determined for the original antiparallel DNA strands in the cellular DNA (Khrapko et al., 1997a).

Figure 33. The mutant fraction of all low  $T_m$  mutants created by *Pfu* as a function of the number of doublings. The mutant fractions after 15 and 30 doublings of three independent experiments were estimated by comparing the total area under all low  $T_m$  mutant peaks (migrating between 29 and 42 minutes in Fig. 32) with the area under the internal standard. A linear regression analysis was applied to all data points. The slope of the line was estimated to be  $8.2 \times 10^{-5}$  total low  $T_m$  mutations per doubling. The dash lines represent the 95% confidence interval.





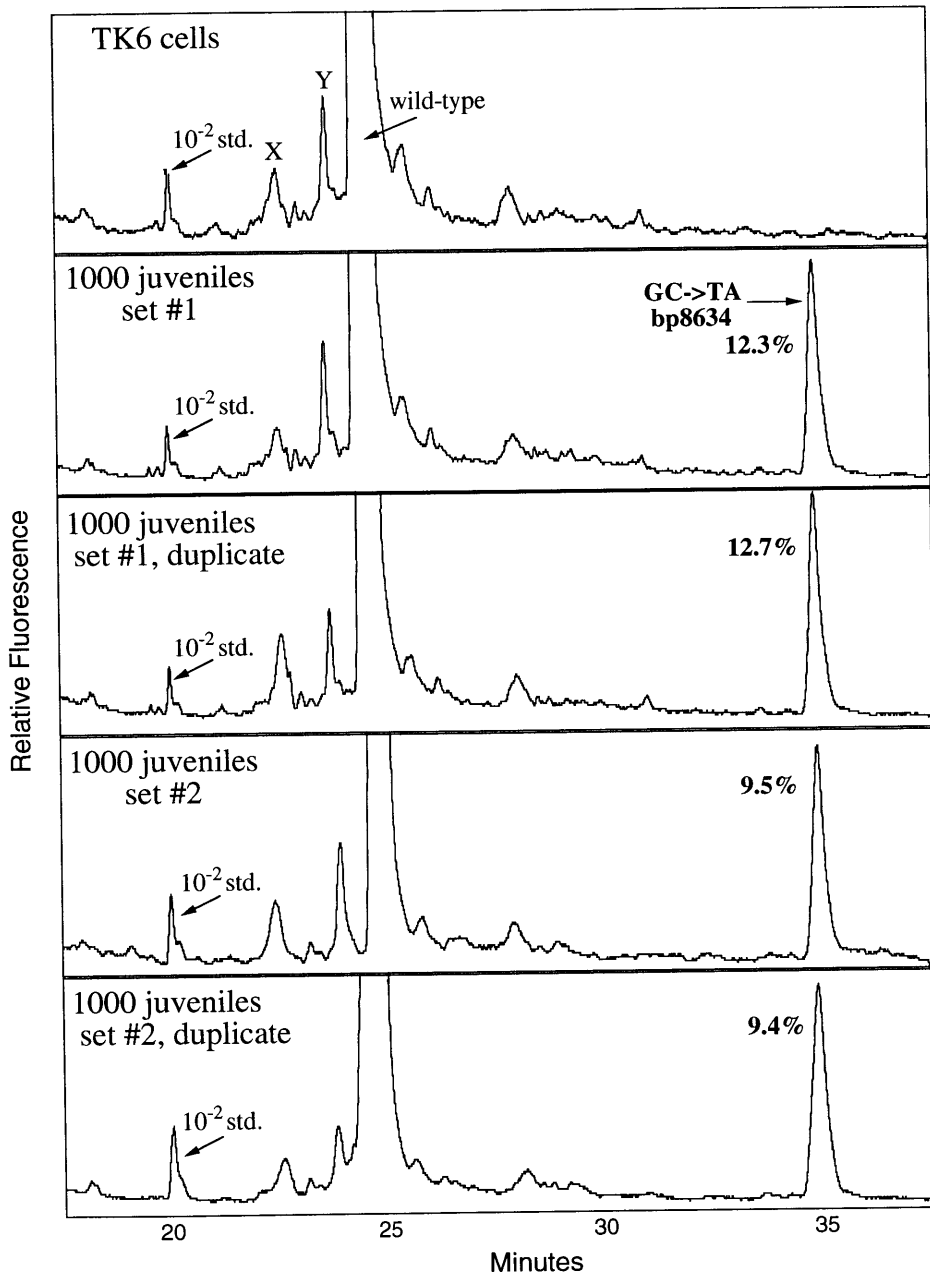
#### 4.7. Single nucleotide polymorphisms in the *APC* target sequence in human populations

The CDCE/high-fidelity PCR approach developed for detecting somatic point mutations in human genomic DNA can be readily employed in screening for single nucleotide polymorphisms (SNPs) in human populations. As a preliminary study, SNPs were searched for in the 121-bp *APC* gene target sequence in 2000 juveniles under the age of six. Two sets of pooled blood samples were created, with 1000 donors in each set. Genomic DNA was isolated from the pooled blood. About 1  $\mu\text{g}$  of the genomic DNA, which contained  $3 \times 10^5$  copies of the *APC* gene alleles (or  $3 \times 10^2$  alleles from each person), was doped with an internal standard at a mutant fraction of  $10^{-2}$  and amplified by high-fidelity PCR. The amplified products were converted into homoduplexes and separated on CDCE. Genomic DNA from TK6 cells was analyzed in parallel to serve as a negative control.

Fig. 34 shows the CDCE separations of the homoduplexes derived from genomic DNA of TK6 cells and the two sets of pooled blood samples, each in duplicate. Distinct from the TK6 negative control, a predominant SNP peak was observed in the blood samples. This peak has been isolated and identified to be a novel GC  $\rightarrow$  TA transversion at the position of *APC* cDNA bp 8613. The allele frequency of this SNP was estimated to be 9.5 and 12.5%, respectively, in the two 1000-juvenile populations. This is the first reported SNP in this particular *APC* gene sequence in human populations. Its biological implication remains to be determined.

Figure 34. CDCE separations of the homoduplexes derived from genomic DNA of TK6 cells and the two sets of pooled blood samples.

CDCE was performed at 9  $\mu$ A and using a 19-cm water-jacket at 64.80°C. Indicated are the wild-type DNA, the internal standard (std.) at  $10^{-2}$ , and the SNP with its frequencies determined in the corresponding blood samples. "X" and "Y" mark unknown PCR artifacts.



#### 4.8. Conclusions

Based on two crucial technical improvements, our previously developed CDCE/high-fidelity PCR approach has been extended to mutational spectrometry of nuclear genes, exemplified here by a 121-bp *APC* gene sequence (Fig. 25). First, sequence-specific hybridization coupled with a biotin-streptavidin capture system was employed to enrich the target sequence from genomic DNA by  $10^4$ -fold. Using this strategy, several milligrams of genomic DNA, which are required for reproducible observation of nuclear gene mutations at fractions as low as  $10^{-7}$  -  $10^{-6}$ , were reduced to less than 1  $\mu\text{g}$  containing the desired target sequence with a yield of greater than 70%. This large reduction in sample size permitted the subsequent procedure for mutant enrichment. The second improvement was a wide-bore capillary electrophoresis technique to enrich for mutant sequences in the presence of the residual cellular DNA. CE using a 540  $\mu\text{m}$  i.d. column provided a large sample loading capacity, a high separation resolution and an excellent target sequence yield (almost 100%). A 100 - 200-fold mutant enrichment was achieved through two consecutive separations of CDCE and CE at room temperature. The high efficiency of mutant enrichment in combination with subsequent high-fidelity PCR using *Pfu* DNA polymerase created a means of sufficient sensitivity to measure nuclear point mutations at fractions as low as  $10^{-6}$ .

The validity of the approach has been tested in the study of MNNG-induced point mutations in the 121-bp *APC* target sequence in MT1 cells. Distinct from the background mutations in the untreated controls, twelve GC  $\rightarrow$  AT transitions were reproducibly observed in the MNNG-treated cells at fractions ranging from  $2.2 \times 10^{-6}$  to  $9.2 \times 10^{-6}$  (Figs. 27 - 29). These transitions predominantly occurred at guanine residues preceded (5') by a purine. The results with regard to the kind of mutation induced by MNNG and its site specificity agree with previous observations in *E. coli*, yeast and other human systems (Burns et al., 1987; Richardson et al., 1987; Kohalmi and Kunz, 1988; Yang et

al., 1991; Kat, 1992).

In the untreated MT1 cells, fourteen background mutations were identified. All fourteen mutations were GC → TA transversions at mutant fraction around  $10^{-6}$ . The observations of the same set of transversion mutations in TK6 cells indicate that they are unlikely to exist as true mutants in the cells. Independent PCR experiments using CDCE-purified wild-type DNA confirmed that *Pfu*-induced mutations were the major source for these transversions. The results suggest that in order to further increase the current detection sensitivity, one would need to improve the mutant enrichment efficiency prior to PCR and/or decrease the level of PCR noise.

The approach described herein can be applied to all DNA sequences suitable for CDCE analysis. It is independent of phenotypic selection and thus permits the direct observation of mutagen-induced mutational spectra in a single cell system without being potentially biased by any "protein filter" effect. Most importantly, it opens the opportunity to study mutational spectra in human tissues and organs. The current detection sensitivity should be sufficient for the observation of stem cell mutations in a tissue sector consisting of  $10^7$  -  $10^8$  cells. Mutations in a stem cell may be found as a colony consisting of 100 - 200 cells after the mutations are transmitted to the transitional and terminal cells of that stem cell's turnover unit. The mutant fraction of a stem cell mutation in a tissue sector would be either zero or  $10^{-6}$  -  $10^{-5}$ , which is above the detection limit of this approach. In addition to detecting somatic mutations, the techniques developed here should also find facile applications in simultaneous screening of large numbers of samples to establish the nature and frequency of rare genetic polymorphisms in human populations.

## 5. SUGGESTIONS FOR FUTURE RESEARCH

### 5.1. Studies of mutagen-induced mutational spectra *in vitro*

Analysis of the site-specificity of a mutational spectrum in a given target sequence may reveal the sequence-specific mechanisms of the mutagenic processes by a given mutagen including initial adduct formation, adduct removal and replication bypass. However, it has been argued that such an analysis is often biased to some extent since the mutation systems used are based on phenotypic selection which confines the analysis to mutations that can produce a functionally altered protein (Gordon and Glickman, 1988; Echols and Goodman, 1990). The CDCE/high-fidelity approach now provides a means to directly address this issue. The mutagen-induced mutational spectra in a selectable gene sequence (such as the *hprt* gene exon 3 low melting domain) can be analyzed and compared under conditions with and without phenotypic selection .

### 5.2. Studies of mutational spectra in nuclear genes in human tissues

The application of the CDCE/high-fidelity PCR approach to human tissue analysis may be first tested in chronically sun-exposed normal skin. Studies of the *p53* gene in sun-exposed, normal human skin have identified UV radiation as the causative agent which produces predominantly C -> T and CC -> TT transitions at dipyrimidine sites (Nakazawa et al., 1994; Jonason et al., 1996; Ren et al., 1997). In these studies, keratinocytes containing mutated p53 protein were observed as immunopositive "patches". Each patch was an independent mutant colony expanded from a mutated stem cell. The average number of the p53 patches per cm<sup>2</sup> was found to be three in sun-shielded skin and thirty in sun-exposed skin (Jonason et al., 1996). Although mutations in other nuclear genes may not cause selective clonal expansion of epidermal cells, it is expected that due to the

constant renewal of the epidermis, mutations in a stem cell will be at least descended to the transitional and terminal cells of that stem cell's turnover unit. Thus, the mutant fraction in the entire epidermis is at least equal to that in the stem cell population.

Based on studies of human epidermal cell proliferation kinetics and colony forming efficiency (Potten, 1981; Weinstein et al., 1984; Potten and Morris, 1988; Jones and Watt, 1993; Heenen and Galand, 1997), it is estimated that under 1 cm<sup>2</sup> surface area of human epidermis, there are a total of  $4.4 \times 10^6$  keratinocytes, among which  $7 \times 10^4$  cells are stem cells. Since there are about thirty p53 patches in sun-exposed skin, the mutant fraction in the stem cell population is  $4.3 \times 10^{-4}$  (= 30/70,000). Assuming the *p53* target size is 1000 base pairs, the average mutant fraction is estimated to be  $4.3 \times 10^{-7}$  per bp in nuclear genes in sun-exposed skin. Mutational hotspots with a 10 - 100-fold increase in the mutant fraction would be at fractions of  $4 \times 10^{-6}$  -  $4 \times 10^{-5}$ , which is above the current detection limit of the CDCE/high-fidelity PCR approach. To reproducibly observe these hotspots, 1 - 10 cm<sup>2</sup> of epidermis ( $4 \times 10^6$  -  $4 \times 10^7$  cells) would be required for the analysis.

The mutant fraction of nuclear genes in normal internal organs is expected to be lower than that in sun-exposed skin. The mutant fraction at the *hprt* and *HLA* loci in peripheral T-lymphocytes has been reported to be about  $10^{-5}$  in humans 50 years of age (Grist et al., 1992; Robinson et al., 1994). Assuming a target size of 1000 base pairs, the average mutant fraction is estimated to be  $10^{-8}$  per bp. Mutational hotspots with a 10 - 100-fold increase in the mutant fraction would be at fractions of  $10^{-7}$  -  $10^{-6}$ . To analyze mutations in internal organs, anatomically distinct sectors may be dissected and analyzed in series (Coller et al., 1998). Mutations in a stem cell may be found in as a colony the size of a turnover unit. Assume that the size of a turnover unit may contain 128 cells (H. Zarbl, personal communication), the mutant fraction of a hotspot in a tissue sector consisting of  $5 \times 10^7$  cells would be either zero or greater than  $10^{-6}$ , which is above the current detection limit. To establish a mutational spectrum in an internal organ such as colon,  $10^8$  -  $10^9$  cells or 24 - 240 cm<sup>2</sup> of colon mucosa (Potten et al., 1992) would be required for the analysis.



### 5.3. Studies of SNPs in human populations<sup>11</sup>

The identification of SNPs has significant importance for mapping and discovering disease-associated genes. Based on the hypothesis that the disease-causing SNPs present in newborns should be significantly decreased in centenarians, it has been proposed that by comparison of the SNP spectra in newborns and centenarians, one may identify SNPs associated with mortal diseases as well as the genes in which they reside (Tomita-Mitchell et al., 1998). Such studies require the examination of a large number of people in order to identify numerically significant SNPs. A population of 10,000 people has been suggested (Tomita-Mitchell et al., 1998). The analytical power of the CDCE/high-fidelity PCR approach offers a cost-efficient means to simultaneously screen such a large population and allow for detection of rare SNPs at frequencies as low as  $5 \times 10^{-4}$  in virtually any ~100-bp DNA sequence.

The experimental plan would be as follows. A pooled DNA sample is created from blood drawn from a cohort of 10,000 persons. To reduce the sampling variation, 100 - 1000 white blood cells (corresponding to about 0.1 - 1  $\mu$ l of blood) per person should be used for each analysis (Tomita-Mitchell et al., 1998). The pooled DNA sample is expected to contain 6 - 60  $\mu$ g of genomic DNA extracted from a total of  $10^6$  -  $10^7$  white blood cells. After restriction digestion, the sequence-specific enrichment technique is employed to enrich the desired target sequences. An appropriate choice of endonucleases used in the digestion should permit sequential enrichment of multiple target sequences from the same pooled sample. The target-enriched sample is amplified by high-fidelity PCR and subjected to a CDCE-PCR-CDCE procedure to enrich, amplify and finally display the SNPs.

## 6. REFERENCES

Abbott, P. J. and Saffhill, R. (1979) DNA synthesis with methylated poly(dC-dG) templates. Evidence for a competitive nature to miscoding by O<sup>6</sup>-methylguanine. *Biochim. Biophys. Acta.* **562**, 51-61.

Abravaya, K., Carrino, J. J., Muldoon, S. and Lee, H. H. (1995) Detection of point mutations with a modified ligase chain reaction (Gap-LCR). *Nucleic Acid Res.* **23**, 675-682.

Aguilar, F., Hussain, S. P. and Cerutti, P. (1993) Aflatoxin B<sub>1</sub> induces the transversion of G → T in codon 249 of the p53 tumor suppressor gene in human hepatocytes. *Proc. Natl. Acad. Sci. USA* **90**, 8586-8590.

Aguilar, F., Harris, C. C., Sun, T., Hollstein, M. and Cerutti, P. (1994) Geographic variation of p53 mutational profile in nonmalignant human liver. *Science* **264**, 1317-1319.

Ainsworth, P. J., Surh, L. C. and Coulter-Mackie, M. B. (1991) Diagnostic single strand conformational polymorphism (SSCP): a simple non-radioisotopic method as applied to a Tay-Sachs B1 variant. *Nucleic Acids Res.* **19**, 405-406.

Albertini, R. J., Castle, K. L. and Borcharding, W. R. (1982) T-cell cloning to detect the mutant 6-thioguanine-resistant lymphocytes present in human peripheral blood. *Proc. Natl. Acad. Sci. USA* **79**, 6617-6621.

Albertini, R. J., Nicklas, J. A. and O'Neill, J. P. (1993) Somatic cell gene mutations in humans: biomarkers for genotoxicity. *Environ. Health Perspect.* **101 (Suppl. 3)**, 193-201.

Amstad, P. A. and Cerutti, P. A. (1995) Ultraviolet-B-light-induced mutagenesis of c-H-ras codons 11 and 12 in human skin fibroblasts. *Int. J. Cancer* **63**, 136-139.

Andersen, T. I., Eiken, H. G., Couch, F., Kaada, G., Skrede, M., Johnson, H., Aloysius, T. A., Tveit, K. M., Tranebjaerg, L., Dorum, A., Moller, P., Weber, B. L. and Borresen-Dale, A. L. (1998) Constant denaturant gel electrophoresis (CDGE) in BRCA1 mutation screening. *Hum. Mutat.* **11**, 166-174.

André, P., Kim, A., Khrapko, K. and Thilly, W. G. (1997) Fidelity and mutational spectrum of *Pfu* DNA polymerase on a human mitochondrial DNA sequence. *Genome Research* **7**, 843-852.

Babic, I., Andrew, S. E. and Jirik, F. R. (1996) MutS interaction with mismatch and alkylated base containing DNA molecules detected by optical biosensor. *Mutat. Res.* **372**, 87-96.

Bailly, V. and Verly, W. G. (1989) AP-endonucleases and AP lyases. *Nucleic Acids Res.* **17**, 3617-3618.

Barany, F. (1991) Genetic disease detection and DNA amplification using cloned thermostable ligase. *Proc. Natl. Acad. Sci. USA* **88**, 189-193.

Behn, M. and Schuermann, M. (1998) Sensitive detection of p53 gene mutations by a

- 'mutant enriched' PCR-SSCP technique. *Nucleic Acids Res.* **26**, 1356-1358.
- Benzer, S. and Freese, E. (1958) Introduction to specific mutations with 5 bromouracil. *Proc. Natl. Acad. Sci. USA* **44**, 112-119.
- Benzer, S. (1961) On the topography of the genetic fine structure. *Proc. Natl. Acad. Sci. USA* **47**, 403-415.
- Beranek, D. T. (1990) Distribution of methyl and ethyl adducts following alkylation with monofunctional alkylating agents. *Mutat. Res.* **231**, 11-30.
- Berdal, K. G., Bjoras, M., Bjelland, S. and Seeberg, E. (1990) *EMBO J.* **9**, 4563
- Bhanot, O. S. and Ray, A. (1986) The in vivo mutagenic frequency and specificity of O<sup>6</sup>-methylguanine in ØX174 RF DNA. *Proc. Natl. Acad. Sci. USA* **83**, 7348-7352.
- Bird, A. P. (1980) DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res.* **8**, 1499-1504.
- Bird, A. P. (1986) CpG-rich islands and the function of DNA methylation. *Nature* **321**, 209-213.
- Boiteux, S. and Laval, J. (1983) Imidazole open ring of 7-methylguanine: an inhibitor of DNA synthesis. *Biochem. Biophys. Res. Commun.* **110**, 552-558.
- Bottema, C. D. K. and Sommer, S. S. (1993) PCR amplification of specific alleles: Rapid detection of known mutations and polymorphisms. *Mutat. Res.* **288**, 93-102.
- Bolden, A. H., Nalin, C. M., Ward, C. A., Poonian, M. S., McComas, W. W. and Weissbach, A. (1985) DNA methylation: sequences flanking C-G pairs modulate the specificity of the human DNA methylase. *Nucleotide Acids Res.* **13**, 3479-3494.
- Boucheron, J. A., Richardson, F. C., Morgan, P. H. and Swenberg, J. A. (1987) Molecular dosimetry of O<sup>4</sup>-ethyldeoxythymine in rats continuously exposed to diethylnitrosamine. *Cancer Res.* **47**, 1577-1581.
- Børresen, A. L., Hovig, E., Smith-Sorensen, B., Malkin, D., Lystad, S., Andersen, T. I., Nesland, J. M., Isselbacher, K. J. and Friend, S. H. (1991) Constant denaturant gel electrophoresis as a rapid screening technique for p53 mutations. *Proc Natl. Acad. Sci. USA* **88**, 8405-8409.
- Brash, D. E., Rudolph, J. A., Simon, J. A., Lin, A., McKenna, G. J., Baden, H. P., Halperin, A. J. and Poter, J. (1991) a role for sunlight in skin cancer: UV-induced p53 mutations in squamous cell carcinoma. *Proc. Natl. Acad. Sci. USA* **88**, 10124-10128.
- Brent, T. P., Dolan, M. E., Fraenkel-Conrat, H., Hall, J., Karran, P., Laval, F., Margison, G. P., Montesano, R., Pegg, A. E., Potter, P. M., singer, B., Swenberg, J. A. and Yarosh, D. B. (1988) Repair of O-alkylpyrimidines in mammalian cells: a present consensus. *Proc. Natl. Acad. Sci. USA* **85**, 1759-1762.
- Burkhart-Schultz, K., Thompson, C. L. and Jones, I. M. (1996) Spectrum of somatic mutation at the hypoxanthine phosphoribosyltransferase (*hprt*) gene of healthy people. *Carcinogenesis* **17**, 1871-1883.

Burns, P. A., Gordon, A. J. E. and Glickman, B. W. (1987) Influence of neighbouring base sequence on *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine mutagenesis in the *lacI* gene of *Escherichia coli*. *J. Mol. Biol.* **194**, 385-390.

Burns, P. A., Gordon, A. J. E., Kunzman, K. and Glickman, B. W. (1988) Influence of neighbouring base sequence on the distribution and repair of ENU-induced lesions in *Escherichia coli*. *Cancer Res.* **48**, 4455-4458.

Cariello, N. F., Cui, L. and Skopek, T. R. (1994) In vitro mutational spectrum of aflatoxin B1 in human hypoxanthine guanine phosphoribosyltransferase gene. *Cancer Res.* **54**, 4436-4441.

Cariello, N. F., Douglas, G. R., Dyaico, M. J., Gorelick, N. J., Provost, G. S. and Soussi, T. (1997) Data-bases and software for the analysis of mutations in the human p53 gene, the human *hprt* gene and both the *lacI* and *lacZ* gene in transgenic rodents. *Nucleic Acids Res.* **25**, 136-137.

Cariello, N. F., Keohavong, P., Kat, A. G. and Thilly, W. G. (1990) Molecular analysis of complex human cell populations: mutational spectra of MNNG and ICR-191. *Mutat. Res.* **231**, 165-176.

Cariello, N. F. and Skopek, T. R. (1993) Mutational analysis using denaturing gradient gel electrophoresis and PCR. *Mutat. Res.* **288**, 103-112.

Cariello, N. F. and Skopek, T. R. (1993) Analysis of mutations occurring at the human *hprt* locus. *J. Mol. Biol.* **231**, 41-57.

Cariello, N. F., Swenberg, J. A. and Skopek, T. R. (1992) In vitro mutational specificity of cisplatin in the human hypoxanthine guanine phosphoribosyltransferase gene. *Cancer Res.* **52**, 2866-2873.

Cha, R. S., Zarbl, H., Keohavong, P. and Thilly, W. G. (1992) Mismatch amplification mutation assay (MAMA): application to the *c-H-ras* gene. *PCR Methods Appl.* **2**, 14-20.

Cha, R. S., Thilly, W. G. and Zarbl, H. (1994) *N*-Nitroso-*N*-methylurea-induced rat mammary tumors arise from cells with preexisting oncogenic *H-ras1* gene mutations. *Proc. Natl. Acad. Sci. USA* **91**, 3749-3753.

Chen, J., Derfler, B., Maskati, A. and Samson, L. (1989) Cloning a eukaryotic DNA glycosylase repair gene by the suppression of a DNA repair defect in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **86**, 7961

Chen, J. and Thilly, W. G. (1994) Mutational spectrum of chromium (VI) in human cells. *Mutat. Res.* **323**, 21-27.

Chen, Z. Y. and Zarbl, H. (1997) A nonradioactive, allele-specific polymerase chain reaction for reproducible detection of rare mutations in large amounts of genomic DNA: application to human *k-ras*. *Anal. Biochem.* **244**, 191-194.

Cheng, K. C., Cahill, D. S., Kasai, H., Nishimura, S. and Loeb, L. A. (1992) 8-Hydroxyguanine, an abundant form of oxidative DNA damage, causes G → T and A → C substitutions. *J. Biol. Chem.* **267**, 166-172.

Chiocca, S. M., Sandy, M. S. and Cerutti, P. A. (1992) Genotypic analysis of *N*-ethyl-*N*-

nitrosourea-induced mutations by *Taq* I restriction fragment length polymorphism / polymerase chain reaction in the *c-H-ras1* gene. *Proc. Natl. Acad. Sci. USA*. **89**, 5331-5335.

Cohen, A. S., Najarian, D. R. and Karger, B. L. (1990) Separation and analysis of DNA sequence reaction products by capillary gel electrophoresis. *J. Chromatogr.* **516**, 49-60.

Cole, J. and Skopek, T. R. (1994) Somatic mutant frequency, mutation rates and mutational spectra in the human population in vivo. *Mutat. Res.* **304**, 33-105.

Coller, H. A., Khrapko, K., Torres, A., Frampton, M. W., Utell, M. J. and Thilly, W. G. (1998) Mutational spectra of a 100-base pair mitochondrial DNA target sequence in bronchial epithelial cells: a comparison of smoking and nonsmoking twins. *Cancer Res.* **58**, 1268-1277.

Cooper, D. N. (1983) Eukaryotic DNA methylation. *Hum. Genet.* **64**, 315-333.

Cooper, D. N., Ball, E. V. and Krawczak, M. (1998) The human gene mutation database. *Nucleic Acids Res.* **26**, 285-287.

Cooper, D. N. and Krawczak, M. (1993) *Human Gene Mutation*. BIOS Scientific Publishers Limited, Oxford, UK.

Cooper, D. N. and Youssoufian, H. (1988) The CpG dinucleotide and human genetic disease. *Hum. Genet.* **78**, 151-155.

Coulondre, C. and Miller, J. H. (1977) Genetic studies of the *lac* repressor IV. Mutagenic specificity in the *lacI* gene of *Escherichia coli*. *J. Mol. Biol.* **117**, 577-606.

Cremonesi, L., Firpo, S., Ferrari, M., Righetti, P. G. and Gelfi, C. (1997) Double-gradient DGGE for optimized detection of DNA point mutations. *Biotechniques* **22**, 326-330.

Demple, B. (1990) Self-methylation by suicide DNA repair enzymes. In *Protein Methylation*, eds. Paik, W. K. and Kim, S. (CRC Press, Boca Ration, Florida), pp285-304.

Denissenko, M. F., Pao, A., Tang, M. S. and Pfeifer, G. P. (1996) Preferential formation of benzo[ $\alpha$ ]pyrene adducts in lung cancer mutational hotspots in *P53*. *Science* **274**, 430-432.

Denissenko, M. F., Chen, J. X., Tang, M. S. and Pfeifer, G. P. (1997) Cytosine methylation determines hot spots of DNA damage in the human *P53* gene. *Proc. Natl. Acad. Sci. USA* **94**, 3893-3898.

Denissenko, M. F., Pao, A., Pfeifer, G. P. and Tang, M. S. (1998) Slow repair of bulky DNA adducts along the nontranscribed strand of the human *p53* gene may explain the strand bias of transversion mutations in cancer. *Oncogene*,

Den Engelse, L., Menkveld, G. J., De Brij, R. J. and Tates, A. D. (1986) Formation and stability of alkylated pyrimidine and purines (including imidazole ring-opened-7-alkylguanine) and alkylphosphotriesters in liver DNA of adult rats treated with ethylnitrosourea or dimethyl nitrosamine. *Carcinogenesis* **7**, 393-403.

- Dolan, M. E., Oplinger, M. and Pegg, A. E. (1988) Sequence specificity of guanine alkylation and repair. *Carcinogenesis* **9**, 2139-2143.
- Duckett, D. R., Drummond, J. T., Murchie, A. I. H., Reardon, J. T., Sancar, A., Lilley, D. M. J. and Modrich, P. (1996) Human MutS $\alpha$  recognizes damaged DNA base pairs containing O6-methylguanine, O4-methylthymine, or the cisplatin-d(GpG) adduct. *Proc. Natl. Acad. Sci. USA* **93**, 6443-6447.
- Dumaz, N., Stary, A., Soussi, T., Daya-Grosjean, L. and Sarasin, A. (1994) Can we predict solar ultraviolet radiation as the causal event in human tumours by analysing the mutation spectra of the *p53* gene? *Mutat. Res.* **307**, 375-386.
- Duncan, B. K. and Miller, J. H. (1980) Mutagenic deamination of cytosine residues in DNA. *Nature* **287**, 560-561.
- Eadie, J. S., Conrad, M., Toorchen, D. and Topal, M. D. (1984) Mechanism of mutagenesis by O<sup>6</sup>-methylguanine. *Nature (Lond.)* **308**, 201-203.
- Echols, H. and Goodman, M. F. (1990) Mutation induced by DNA damage: a many protein affair. *Mutat. Res.* **236**, 301-311.
- Ellis, L. A., Taylor, G. R., Banks, R. and Baumberg, S. (1994) MutS binding protects heteroduplex DNA from exonuclease digestion in vitro: a simple method for detecting mutations. *Nucleic Acids Res.* **22**, 2710-2711.
- Fan, E., Levin, D. B., Glickman, B. W. and Logan, D. M. (1993) Limitations in the use of SSCP analysis. *Mutat. Res.* **288**, 85-92.
- Fearon, E.R., and Vogelstein, B. (1990) A genetic model for colorectal tumorigenesis. *Cell* **61**, 759-767.
- Fischer, S. G. and Lerman, L. S. (1983) DNA fragments differing by single base-pair substitutions are separated in denaturing gradient gels: correspondence with melting theory. *Proc. Natl. Acad. Sci. USA* **80**, 1579-1583.
- Friedberg, E. C., Walker, G. C. and Siede, W. (1995) *DNA repair and mutagenesis*. (ASM Press, Washington, D.C.)
- Fritz, G., Tano, K., Mitra, S. and Kaina, B. (1991) Inducibility of the DNA repair gene encoding O6-methylguanine-DNA methyltransferase in mammalian cells by DNA-damaging treatment. *Mol. Cell Biol.* **11**, 4660-4668.
- Foster, P. L. and Eisenstadt, E. (1985) Induction of transversion mutations in *Escherichia coli* by N-methyl-N'-nitrosoguanidine is SOS-dependent. *J. Bacteriol.* **163**, 213-220.
- Fuchs, R. P. P., Schwartz, N. & Daune, M. D. (1981) Hot spots of frameshift mutations induced by the ultimate carcinogen N-acetoxy-N-2-acetylaminofluorene. *Nature* **294**, 657-659.
- Furth, E. E., Thilly, W. G., Penman, B. W., Liber, H. L. and Rand, W. M. (1981) Quantitative assay for mutation in diploid human lymphoblasts using microtiter plates. *Anal. Biochem.* **110**, 1-8.

- Gelfi, C., Righetti, S. C., Zunino, F., Della Torre, G., Pierotti, M. A. and Righetti, P. G. (1997) Detection of p53 point mutations by double-gradient, denaturing gradient gel electrophoresis. *Electrophoresis* **18**, 2921-2927.
- Gichner, T. and Veleminsky, J. (1982) Genetic effects of *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine and its homologs. *Mutat. Res.* **99**, 129-242.
- Goldmacher, V. S., Cuzick, R. A. and Thilly, W. G. (1986) Isolation and partial characterization of human cell mutants differing in sensitivity to killing and mutation by methylnitrosourea and *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine. *J. Biol. Chem.* **261**, 12462-12471.
- Gordon, A. J. E. and Glickman, B. W. (1988) Protein domain structure influences observed distribution of mutation. *Mutat. Res.* **208**, 105-108.
- Gordon, A. J. E., Burns, P. A. and Glickman, B. W. (1990) *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine induced DNA sequence alteration; non-random components in alkylation mutagenesis. *Mutat. Res.* **233**, 95-103.
- Gordon, A. J. E., Schy, W. E. and Glickman, B. W. (1990b) Non-phenotypic selection of *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine-directed mutation at a predicted hotspot site. *Mutat. Res.* **243**, 145-149.
- Greenblatt, M. S., Bennett, W. P., Hollstein, M. and Harris, C. C. (1994) Mutations in the *p53* tumor suppressor gene: clues to cancer etiology and molecular pathogenesis. *Cancer Res.* **54**, 4855-4878.
- Grist, S. A., McCarron, M., Kutlaca, A., Turner, D. R. and Morley, A. A. (1992) In vivo human somatic mutation: frequency and spectrum with age. *Mutat. Res.* **266**, 189-196.
- Groden, J., Thliveris, J., Samowitz, W., et al. (1991) Identification and characterization of the familial adenomatous polyposis coli gene. *Cell* **66**, 589-600.
- Grushka, E., McCormick, R.M., and Kirkland, J.J. (1989) Effect of temperature gradients on the efficiency of capillary zone electrophoresis separations. *Anal. Chem.* **61**, 241-246.
- Guldberg, P., Gronbak, K., Aggerholm, A., Platz, A., thor Straten, P., Ahrenkiel, V., Hokland, P. and Zeuthen, J. (1998) Detection of mutations in GC-rich DNA by bisulphite denaturing gradient gel electrophoresis. *Nucleic Acid Res.* **26**, 1548-1549.
- Hainaut, P., Hernandez, T., Robinson, A., Rodriguez-Tome, P., Flores, T., Hollstein, M., Harris, C. C. and Montesano, R. (1998) IARC database of p53 gene mutations in human tumors and cell lines: updated compilation, revised formats and new visualisation tools. *Nucleic Acids Res.* **26**, 205-213.
- Hakura, A., Morimoto, K., Sofuni, T. and Nohmi, T. (1991) Cloning and characterization of the *Samonella typhimurium ada* gene, which encodes *O*<sup>6</sup>-methylguanine-DNA methyltransferase. *J. Bacteriol.* **173**, 3663-3672.
- Hanekamp, J. S., André, P., Collier, H. A., Li, X.-C., Thilly, W. G., and Khrapko, K. (1996) CDCE: Constant denaturant capillary electrophoresis for detection and enrichment of sequence variants, in: *Laboratory Protocols for Mutation detection* (Landegren, U., ed.), Oxford University Press, Oxford, pp38-41.

Harris, C. C. (1996) p53 tumor suppressor gene: at the crossroads of molecular carcinogenesis, molecular epidemiology, and cancer risk assessment. *Environ. Health Perspect.* **104** (Suppl. 3), 435-439.

Harris, C. C. and Sun, T. (1984) Multifactorial etiology of human liver cancer. *Carcinogenesis* **5**, 697-701.

Heenen, M. and Galand, P. (1997) The growth fraction of normal human epidermis. *Dermatology* **194**, 313-317.

Hernandez-Boussard, T. M. and Hainaut, P. (1998) A specific spectrum of p53 mutations in lung cancer from smokers: review of mutations compiled in the IARC p53 database. *Environ. Health Perspect.* **106**, 385-391.

Hevroni, D. and Livneh, Z. (1988) Bypass and termination at apurinic sites during replication of single-stranded DNA in vitro: a model for a apurinic site mutagenesis. *Proc. Natl. Acad. Sci. USA* **85**, 5046-5050.

Higuchi, M., Antonarakis, S. E., Kasch, L., Oldenburg, J., Economou-Petersen, E., Olek, K., Arai, M., Inaba, H. and Kazarian Jr., H. H. (1991) Molecular Characterization of mild-to-moderate hemophilia A: detection of the mutation in 25 of 29 patients by denaturing gradient gel electrophoresis. *Proc. Natl. Acad. Sci. USA* **88**, 8307-8311.

Hjertén, S. (1985) High-performance electrophoresis. Elimination of electroendosmosis and solute adsorption. *J. Chromatogr.* **347**, 189-198.

Hollstein, M., Sidransky, D., Vogelstein, B. and Harris, C. C. (1991) p53 mutations in human cancers. *Science* **253**, 49-53.

Hongya, T., Buzard, G. S., Calvert, R. J. and Weghorst, C. M. (1993) 'Cold SSCP': a simple, rapid and non-radioactive method for optimized single-strand conformation polymorphism analysis. *Nucleic Acids Res.* **21**, 3637-3642.

Horsfall, M. J., Gordon, A. J. E., Burns, P. A., Zielenska, M., van der Vliet, G. M. E. and Glickman, B. W. (1990) Mutational specificity of alkylating agents and the influence of DNA repair. *Environ. Mol. Mutagen.* **15**, 107-122.

Hovig, E., Smith-Sørensen, B., Brøgger, A. and Børresen, A.-L. (1991) Constant denaturant gel electrophoresis, a modification of denaturing gradient gel electrophoresis, in mutation detection. *Mutat. Res.* **262**, 63-71.

Humphries, S. E., Gudnason, V., Whittall, R. and Day, I. N. M. (1997) Single-strand conformation polymorphism analysis with high throughput modifications, and its use in mutation detection in familial hypercholesterolemia. *Clin. Chem.* **43**, 427-435.

Hussain, S. P., Aguilar, F. and Cerutti, P. (1994a) Mutagenesis of codon 248 of the human p53 tumor suppressor gene by N-ethyl-N-nitrosourea. *Oncogene* **9**, 13-18.

Hussain, S. P., Aguilar, F., Amstad, P. and Cerutti, P. (1994b) Oxy-radical induced mutagenesis of hotspot codons 248 and 249 of the human p53 gene. *Oncogene* **9**, 2277-2281.

Hussain, S. P. and Harris, C. C. (1998) Molecular epidemiology of human cancer:



contribution of mutation spectra studies of tumor suppressor genes. *Cancer Res.* **58**, 4023-4037.

Ito, T., Nakamura, T., Maki, H. and Sekiguchi, M. (1994) Roles of transcription and repair in alkylation mutagenesis. *Mutat. Res.* **314**, 273-285.

Ito, T., Smith, C. L. and Cantor, C. R. (1992) Sequence-specific DNA purification by triplex affinity capture. *Proc. Natl. Acad. Sci. USA* **89**, 495-498.

Jenkins, G. J. S., Chaleshtori, M. H., Song, H. and Parry, J. M. (1998) Mutation analysis using the restriction site mutation (RSM) assay. *Mutat. Res.* **405**, 209-220.

Jonason, A.S., Kunala, S., Price, G.J., Restifo, R.J., Spinelli, H.M., Persing, J.A., Leffell, D.J., Tarone, R.E. and Brash, D.E. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 14025-14029.

Jones, M., Wagner, R. and Radman, M. (1987) Repair of a mismatch is influenced by the base composition of the surrounding nucleotide sequence. *Genetics* **115**, 605-610.

Jones, P.H. and Watt, F. M. (1993) Separation of human epidermal stem cells from transit amplifying cells on the basis of differences in integrin function and expression. *Cell* **73**, 713-724.

Jorgenson, J.W. and Lukacs, K.D. (1983) Capillary zone electrophoresis. *Science* **222**, 266-272.

Kat, A. (1992) MNNG mutational spectra in the human *hprt* gene. Ph.D. Thesis. MIT.

Kat, A., Thilly, W. G., Fang, W., Longley, M. J., Li, G.-M. and Modrich, P. (1993) An alkylation-tolerant mutator human cell line is deficient in strand-specific mismatch repair. *Proc. Natl. Acad. Sci. USA* **90**, 6424-6428.

Katsuragi, K., Kitagishi, K., Chiba, W., Ikeda, S. and Kinoshita, M. (1996) Fluorescence-based polymerase chain reaction-single-strand conformation polymerase analysis of p53 gene by capillary electrophoresis. *J. Chromatogr. A.* **744**, 311-320.

Kawate, H., Ihara, K., Kohda, K., Sakumi, K. and Sekiguchi, M. (1995) Mouse methyltransferase for repair of *O*<sup>6</sup>-methylguanine and *O*<sup>4</sup>-methylthymine in DNA. *Carcinogenesis* **16**, 1595-1602.

Kälin, I., Shephard, S. and Candrian, U. (1992) Evaluation of the ligase chain reaction (LCR) for the detection of point mutations. *Mutat. Res.* **283**, 119-123.

Keohavong, P., Liu, V. F. and Thilly, W. G. (1991) Analysis of point mutations induced by ultraviolet light in human cells. *Mutat. Res.* **249**, 147-159.

Keohavong, P. and Thilly, W. G. (1992a) Mutational spectrometry: a general approach for hot-spot point mutations in selectable genes. *Proc. Natl. Acad. Sci. USA* **89**, 4623-4627.

Keohavong, P. and Thilly, W. G. (1992b) Determination of point mutational spectra of benzo[ $\alpha$ ]pyrene-diol epoxide in human cells. *Environ. Health Perspect.* **98**, 215-219.

Khrapko, K., André, P., Cha, R., Hu, G. and Thilly, W. G. (1994a) Mutational

spectrometry: means and ends. *Prog. Nucl. Acids. Res. Mol. Biol.* **49**, 285-312.

Khrapko, K., Hanekamp, J. S., Thilly, W. G., Belenkii, A., Foret, F. and Karger, B. L. (1994b) Constant denaturant capillary electrophoresis (CDCE): a high resolution approach to mutational analysis. *Nucleic Acids Res.* **22**, 364-369.

Khrapko, K., Collier, H. and Thilly, W. G. (1996) Efficiency of separation of DNA mutations by constant denaturant capillary electrophoresis is controlled by the kinetics of DNA melting equilibrium. *Electrophoresis* **17**, 1867-1874.

Khrapko, K., Collier, H., André, P., Li, X.-C., Foret, F., Belenky, A., Karger, B. L. and Thilly, W. G. (1997a) Mutational spectrometry without phenotypic selection: human mitochondrial DNA. *Nucleic Acids Res.* **25**, 685-693.

Khrapko, K., Collier, H. A., André, P. C., Li, X.-C., Hanekamp, J. S. and Thilly, W. G. (1997b) Mitochondrial mutational spectra in human cells and tissues. *Proc. Natl. Acad. Sci. USA* **94**, 13798-13803.

Kinzler, K. W., Nilbert, M. C., Su, L. K., Vogelstein, B., Bryan, T. M., Levy, D. B., Smith, K. J., Preisinger, A. C., Hedge, P., McKechnie, D. et al. (1991) Identification of FAP locus genes from chromosome 5q21. *Science* **253**, 661-665.

Kinzler, K. W. and Vogelstein, B. (1996) Lessons from hereditary colorectal cancer. *Cell* **87**, 159-170.

Kishimoto, Y., Murakami, Y., Shiraishi, M., Hayashi, M. and Sekiya, T. (1992) Aberrations of the p53 tumor suppressor gene in human non-small cell carcinomas of the lung. *Cancer Res.* **52**, 4799-4804.

Knox, J. (1988) Thermal effects and band spreading in capillary electro-separation. *Chromatographia* **26**, 329-337.

Kogan, S. and Gitschier, J. (1990) Mutations and a polymorphism in the factor VIII gene discovered by denaturing gradient gel electrophoresis. *Proc. Natl. Acad. Sci. USA* **87**, 2092-2096.

Kohalmi, S. E. and Kunz, B. A. (1988) Role of neighbouring bases and assessment of strand specificity in ethylmethanesulphonate and *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine mutagenesis in the *SUP4-o* gene of *Saccharomyces cerevisiae*. *J. Mol. Biol.* **204**, 561-568.

Kohalmi, S. E., Roche, H. M. and Kunz, B. A. (1993) Elevated intracellular dCTP levels reduce the induction of GC → AT transitions in yeast by ethyl methanesulfonate of *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine but increase alkylation-induced GC → CG transversions. *Mutagenesis* **8**, 457-465.

Krawczak, M., Ball, E. V. and Cooper, D. N. (1998) Neighboring-nucleotide effects on the rates of germline single-base-pair substitution in human genes. *Am. J. Hum. Genet.* **63**, 474-488.

Kunkel, T. A., Schaaper, R. M., Beckman, R. A. and Loeb, L. A. (1981) On the fidelity of DNA replication. Effect of the next nucleotide on proofreading. *J. Biol. Chem.* **256**, 9883-9889.

Kunkel, T. A. (1984) Mutational specificity of depurination. *Proc. Natl. Acad. Sci. USA* **81**, 1494-1498.

Kunkel, T. A. (1985) The mutational specificity of DNA polymerase- $\alpha$  during in vitro DNA synthesis. *J. Biol. Chem.* **260**, 5787-5796.

Kunkel, T. A. (1992) DNA replication fidelity. *J. Biol. Chem.* **267**, 18251-18254.

Kunz, B. A., Henson, E. S., Karthikeyan, R., Kuschak, T., McQueen, S. A., Scott, C. A. and Xiao, W. (1998) Defects in base excision repair combined with elevated intracellular dCTP levels dramatically reduce mutation induction in yeast by ethyl methanesulfonate and N-methyl-N'-nitro-N-nitrosoguanidine. *Environ. Mol. Mutagen.* **32**, 173-178.

Larson, K., Sahm, J., Shenkar, R. and Strauss, B. (1985) Methylation-induced blocks to in vitro DNA replication. *Mutat. Res.* **150**, 77-84.

Lauer H.H. and McManigill, D. (1986) Zone electrophoresis in open-tubular capillaries - recent advances. *Trends Anal. Chem.* **5**, 11-15.

Laurent-Puig, P., Bérout, C. and Soussi, T. (1998) APC gene: database of germline and somatic mutations in human tumors and cell lines. *Nucleic Acids Res.* **26**, 269-270.

Laval, J., Boiteux, S. and O'Connor, T. R. (1990) Physiological properties and repair of apurinic/a[apyrimidinic sites and imidazole ring-opened guanines in DNA. *Mutat. Res.* **233**, 73-79.

Lawley, P. D. (1976) Methylation of DNA by carcinogens, some applications of chemical analytical methods. In: R. Montesano, H. Bartsch and L. Tomatis (Eds.), Screening tests in chemical carcinogenesis. IARC Sci. Publ. No. 12, Lyon, pp. 181-208.

Lawley, P. D. and Orr, D. J. (1970) Specific excision of methylation products from DNA of *Escherichia coli* treated with N-methyl-N'-nitro-N-nitrosoguanidine. *Chem.-Biol. Interact.* **2**, 154-157.

Leadser, D. P., Peter, B. and Ehmer, B. (1995) Analysis of CpG dinucleotide frequency in relationship translational reading frame suggests a class of genes in which mutation of this dinucleotide is asymmetric with respect to DNA strand. *FEBS Lett.* **376**, 125-129.

Leong, P.-M., Thilly, W. G. and Morgenthaler, S. (1985) Variance estimation in single-cell mutation assays: comparison to experimental observations in human lymphoblasts at 4 gene loci. *Mutat. Res.* **150**, 403-410.

Lerman, L.S. and Silverstein, K. (1987) Computational simulation of DNA melting and its application to denaturing gradient gel electrophoresis. *Meth. Enzymol.* **155**, 482-501.

Levine, A. J., Momand, J. and Finlay, C. A. (1991) The p53 tumor suppressor gene. *Nature* **351**, 453-456.

Li, X.-C. and Thilly, W. G. (1996) Use of wide-bore capillaries in constant denaturant capillary electrophoresis. *Electrophoresis* **17**, 1884-1889.

Lindahl, T. and Nyberg, B. (1972) Rate of depurination of native deoxyribonucleotide

acid. *Biochemistry* **11**, 3610-3618.

Lindahl, T., Sedgwick, B., Sekiguchi, M. and Nakabeppu, Y. (1988) Regulation and expression of the adaptive response to alkylating agents. *Annu. Rev. Biochem.* **57**, 133-157.

Lishanski, A., Ostrander, E. A. and Rine, J. (1994) Mutation detection by mismatch binding protein, MutS, in amplified DNA: application to the cystic fibrosis gene. *Proc. Natl. Acad. Sci. USA* **91**, 2674-2678.

Loeb, L. A. and Cheng, K. C. (1990) Errors in DNA synthesis: a source of spontaneous mutations. *Mutat. Res.* **238**, 297-304.

Loeb, L. A. and Preston, B. D. (1986) Mutagenesis by apurinic/apyrimidinic sites. *Ann. Rev. Genet.* **20**, 201-230.

Loechler, E. L., Green, C. L. and Essigmann, J. M. (1984) In vivo mutagenesis by *O*<sup>6</sup>-methylguanine built into a unique site in a viral genome. *Proc. Natl. Acad. Sci. USA* **81**, 6271-6275.

Loveless, A. (1969) Possible relevance of *O*<sup>6</sup>-alkylation of deoxyguanosine to the mutagenicity and carcinogenicity of nitrosamines and nitrosamides. *Nature (Lond.)* **223**, 206-207.

Lu, S.-J. and Archer, M. C. (1992) Ha-*ras* oncogene activation in mammary glands of *N*-methyl-*N*-nitrosourea-treated rats genetically resistant to mammary adenocarcinogenesis. *Proc. Natl. Acad. Sci. USA.* **89**, 1001-1005.

Lukash, L. L., Bolou, J., Pegg, A. E., Dolan, M. E., Maher, V. M. and McCormick, J. J. (1991) Effect of *O*<sup>6</sup>-alkylguanine -DNA alkyltransferase on the frequency and spectrum of mutations induced by *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine in the *HPRT* gene of diploid human fibroblasts. *Mutat. Res.* **250**, 397-409.

Lundberg, K.S., Shoemaker, D.D., Adams, M.W., Short, J.M., Sorge, J.A. and Mathur, E.J. (1991) High-fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus*. *Gene* **108**, 1-6.

Maher, V. M., Domoradzki, J., Bhattacharyya, N., Tsujimura, T., Corner, R. C. and McCormick, J. J. (1990) Alkylation damage, DNA repair and mutagenesis in human cells. *Mutat. Res.* **233**, 235-245.

Malkin, D., Li, F. P., Strong, L. C., Fraumeni, J. F., Nelson, C. E., Kim, D. H., Kassel, J., Gryka, M. A., Bischoff, F. Z., Tainsky, M. A. and Friend, S. H. (1990) Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science* **250**, 1233-1238.

Marcelino, L. A., Andre, P. C., Khrapko, K., Collier, H. A., Griffith, J. and Thilly, W. G. (1998) Chemically induced mutations in mitochondrial DNA of human cells: mutational spectrum of methylnitrosoguanidine in mitochondrial DNA. *Cancer Res.* **58**, 2857-2862.

Margison, G. P., Cooper, D. P. and Potter, P. M. (1990) The *E. coli* *ogt* gene. *Mutat. Res.* **233**, 15-21.

Marusyk, R. and Sergeant, A. (1980) A simple method for dialysis of small-volume samples. *Anal. Biochem.* **105**, 403-404.

Mashiyama, S., Murakami, Y., Yoshimoto, T., Sekiya, T. and Hayashi, K. (1991) Detection of p53 gene mutations in human brain tumors by single-strand conformation polymorphism of polymerase chain reaction products. *Oncogene* **6**, 1313-1318.

Mattes, W. B., Hartley, J. A. and Kohn, K. (1986) DNA sequence selectivity of guanine-N<sup>7</sup> alkylation by nitrogen mustards. *Nucleic Acids Res.* **14**, 2971-2987.

McCarthy, T. V., Karran, P. and Lindahl, T. (1984) Inducible repair of O-alkylated DNA pyrimidines in *Escherichia coli*. *EMBO J.* **3**, 545-550.

McKay, A. F. and Wright, G. F. (1947) Preparation and Properties of N-methyl-N-nitroso-N'-nitroguanidine. *J. Am. Chem. Soc.* **69**, 3028-3030.

Mehta, J. R. and Ludlum, D. B. (1978) Synthesis and properties of O<sup>6</sup>-methyldeoxyguanylic acid and its copolymers with deoxycytidylic acid. *Biochim. Biophys. Acta.* **521**, 770-778.

Mikkers, F.E.P., Everaerts, F.M. and Verheggen, Th. P.E.M. (1979) High-performance zone electrophoresis. *J. Chromatogr.* **169**, 11-20.

Miller, J. H. (1983) Mutational specificity in bacteria. *Annu. Rev. Genet.* **17**, 215-238.

Miyoshi, Y., Nagase, H., Ando, H., Horii, A., Ichii, S., Nakatsuru, S., Aoki, T., Miki, Y., Mori, T. and Nakamura, Y. (1992) Somatic mutations of the APC gene in colorectal tumors: mutation cluster region in the APC gene. *Hum. Mol. Genet.* **1**, 229-233.

Mohrenweiser, H. (1994) Impact of the molecular spectrum of mutational lesions on estimates of germinal gene-mutation rates. *Mutat. Res.* **304**, 119-137.

Morley, A. A., Trainor, K. J., Seshadri, R. and Ryall, R. G. (1983) Measurement of in vivo mutations in human lymphocytes. *Nature* **302**, 155-156.

Montesano, R., Hainaut, P. and Wild, C. P. (1997) Hepatocellular carcinoma: from gene to public health. *J. Natl. Cancer Inst.* **89**, 1844-1851.

Morohoshi, F., Hayashi, K. and Munakata, N. (1990) *Bacillus subtilis ada* operon encodes two DNA alkyltransferases. *Nucleic Acids Res.* **18**, 5473-5480.

Munemitsu, S., Souza, B., Muller, O., Albert, I., Rubinfeld, B., and Polakis, P. (1994) The APC gene product associates with microtubules in vivo and promotes their assembly in vitro. *Cancer Res.* **54**, 3676-3681.

Münscher, C., Rieger, T., Müller-Höcker, J. and Kadenbach, B. (1993) The point mutation of mitochondrial DNA characteristic for MERRF disease is found also in healthy people of different ages. *FEBS Lett.* **317**, 27-30.

Myers, R. M., Lumelsky, N., Lerman, L. S. and Maniatis, T. (1985a) Detection of single base substitution in total genomic DNA. *Nature* **313**, 495-498.

Myers, R. M., Fischer, S. G., Lerman, L. S. and Maniatis, T. (1985b) Nearly all single

base substitutions in DNA fragments joined to a GC-clamp can be detected by denaturing gradient gel electrophoresis. *Nucleic Acid Res.* **13**, 3131-3145.

Nagase, H., and Nakamura, Y. (1993) Mutations of the APC (adenomatous polyposis coli) gene. *Hum. Mutat.* **2**, 425-434.

Nakazawa, H., English, D., Randell, P.L., Nakazawa, K., Martel, N., Armstrong, B.K. and Yamasaki, H. (1994) UV and skin cancer: specific *p53* mutation in normal skin as a biologically relevant exposure measurement. *Proc. Natl. Acad. Sci. USA* **91**, 360-364.

Newbold, R. F., Warren, W., Medcalf, A. S. C. and Amos, J. (1980) Mutagenicity of carcinogenic methylating agents is associated with a specific DNA modification. *Nature (Lond.)* **283**, 596-599.

Newmark, J. R., Hardy, D. O., Tonb, D. C., Carter, B. S., Epstein, J. I., Isaacs, W. B., Brown, T. R. and Barrack, E. R. (1992) Androgen receptor gene mutations in human prostate cancer. *Proc. Natl. Acad. Sci. USA* **89**, 6319-6323.

Nivard, M. J. M., Pastink, A. and Vogel, E. W. (1996) Mutational spectra induced under distinct excision repair conditions by the 3 methylating agents *N*-methyl-*N*-nitrosourea, *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine and *N*-nitrosodimethylamine in postmeiotic male germ cells of *Drosophila*. *Mutat. Res.* **352**, 97-115.

O'Connor, T. R., Boiteux, S. and Laval, J. (1988) Ring-opened 7-methylguanine residues in DNA are a block to in vitro DNA synthesis. *Nucleic Acids Res.* **16**, 5879-5894.

O'Connor, T. R. and Laval, J. (1990) Isolation and structure of a cDNA expressing a mammalian 3-methyladenine-DNA-glycosylase. *EMBO J.* **9**, 3337-3342.

O'Neill, J. P. and Finette, B. A. (1998) Transition mutations at CpG dinucleotides are the most frequent in vivo spontaneous single-base substitution mutation in the human *HPRT* gene. *Environ. Mol. Mutagen.* **32**, 188-191.

Oller, A. R. and Thilly, W. G. (1992) Mutational spectra in human B-cells, spontaneous, oxygen and hydrogen peroxide-induced mutations at the *hprt* gene. *J. Mol. Biol.* **228**, 813-826.

Okinaka, R. T., Anzick, S. L., Oller, A. and Thilly, W. G. (1993) Analysis of large X-ray-induced mutant populations by denaturing gradient gel electrophoresis. *Radiat. Res.* **135**, 212-221.

Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K. and Sekiya, T. (1989) Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc. Natl. Acad. Sci. USA* **86**, 2766-2770.

Palombo, F., Kohfeldt, E., Calcagnile, A., Nehls, P. and Dogliotti, E. (1992) *N*-methyl-*N*-nitrosourea-induced mutations in human cells. Effects of the transcriptional activity of the target gene. *J. Mol. Biol.* **223**, 587-594.

Papadopoulos, N., Nicolaidis, N. C., Liu, B., Parsons, R., Lengauer, C., Palombo, F., D'Arrigo, A., Markowitz, S., Willson, J. K. V., Kinzler, K. W., Jiricny, J. and Vogelstein, B. (1995) Mutations of *GTBP* in genetically unstable cells. *Science* **268**, 1915-1917.

- Parsons, B. L. and Heflich, R. H. (1997a) Evaluation of MutS as a tool for direct measurement of point mutations in genomic DNA. *Mutat. Res.* **374**, 277-285.
- Parsons, B. L. and Heflich, R. H. (1997b) Genotypic selection methods for the direct analysis of point mutations. *Mutat. Res.* **387**, 97-121.
- Parsons, B. L. and Heflich, R. H. (1998) Detection of basepair substitution mutation at a frequency of  $1 \times 10^{-7}$  by combining two genotypic selection methods, MutEx enrichment and allele-specific competitive blocker PCR. *Environ. Mol. Mutagen.* **32**, 200-211.
- Pegg, A. E. (1984) Methylation of the O<sup>6</sup> position of guanine in DNA is the most likely initiating event in carcinogenesis by methylating agents. *Cancer Invest.* **2**, 223-231.
- Pegg, A. E. (1990) Mammalian O<sup>6</sup>-alkylguanine-DNA transferase: regulation and importance in response to alkylating carcinogenesis and therapeutic agents. *Cancer Res.* **50**, 6119-6129.
- Pegg, A. E. and Byers, T. L. (1992) Repair of DNA containing O<sup>6</sup>-alkylguanine. *FASEB J.* **6**, 2302-2310.
- Pfeifer, G. P. and Denissenko, M. F. (1998) Formation and repair of DNA lesions in the *p53* gene: relation to cancer mutations? *Environ. Mol. Mutagen.* **31**, 197-205.
- Podlutzky, A., Osterholm, A. M., Hou, S. M., Hofmaier, A. and Lambert, B. (1998) Spectrum of point mutations in the coding region of the hypoxanthine-guanine phosphoribosyltransferase (*hprt*) gene in human T-lymphocytes *in vivo*. *Carcinogenesis* **19**, 557-566.
- Potter, P. M., Rafferty, J. A., Cawkwell, L., Wilkinson, M. C., Cooper, D. P., O'Connor, P. J. and Margison, G. P. (1991) Isolation and cDNA cloning of a rat O<sup>6</sup>-alkylguanine-DNA-alkyltransferase gene, molecular analysis of expression in rat liver. *Carcinogenesis* **12**, 727-733.
- Potten, C.S. (1981) Cell replacement in epidermis (keratopoiesis) via discrete units of proliferation. *Int. Rev. Cytol.* **69**, 271-318.
- Potten, C.S. and Morris, R.J. (1988) Epithelial stem cells *in vivo*. *J. Cell Sci. (Suppl)* **10**, 45-62.
- Potten, C. S., Kellett, M., Roberts, S. A., Rew, D. A. and Wilson, G. D. (1992) Measurement of *in vivo* proliferation in human colorectal mucosa using bromodeoxyuridine. *Gut* **33**, 71-78.
- Pourzand, C. and Cerutti, P. (1993a) Genotypic mutation analysis by RFLP/PCR. *Mutat. Res.* **288**, 113-121.
- Pourzand, C. and Cerutti, P. (1993b) Mutagenesis of H-*ras* codons 11 and 12 in human fibroblasts by N-ethyl-N-nitrosourea. *Carcinogenesis* **14**, 2193-2196.
- Powell, S.M., Zilz, N., Beazer-Barclay, Y. et al. (1992) APC mutations occur early during colorectal tumorigenesis. *Nature* **359**, 235-237.
- Preston, B. P., Singer, B. and Loeb, L. A. (1986) Mutagenic potential of O<sup>4</sup>-

methylthymine in vivo determined by an enzymatic approach to site-specific mutagenesis. *Proc. Natl. Acad. Sci. USA* **83**, 8501-8505.

Pullman, A. and Pullman, B. (1981) Molecular electrostatic potential of the nucleic acids. *Quart. Rev. Biophys.* **14**, 289-380.

Randall, S. K., Eritja, R., Kaplan, B. E., Petruska, J. and Goodman, M. F. (1987) Nucleotide insertion kinetics opposite abasic lesions in DNA. *J. Biol. Chem.* **262**, 6864-6870.

Rasmussen, L. J. and Samson, L. (1996) The *Escherichia coli* MutS DNA mismatch binding protein specifically binds *O*<sup>6</sup>-methylguanine DNA lesions. *Carcinogenesis* **17**, 2085-2088.

Reed, J. and Hutchinson, F. (1987) Effect of the direction of DNA replication on mutagenesis by *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine in adapted cells of *Escherichia coli*. *Mol. Gen. Genet.* **208**, 446-449.

Reijenga, J. C., Verheggen, T. P. E. M., Martens, J. H. P. A. and Everaerts, F. M. (1996) *J. Chromatogr. A* **744**, 147-153.

Ren, J., Ulvik, A., Ueland, P. M. and Refsum, H. (1997) Analysis of single-strand conformation polymorphism by capillary electrophoresis with laser-induced fluorescence detection using short-chain polyacrylamide as sieving medium. *Anal. Biochem.* **245**, 79-84.

Ren, Z.-P., Ahmadian, A., Pontén, F., Nistér, M., Berg, C., Lundeberg, J., Uhlén, M. and Pontén, J. (1997) Benign clonal keratinocyte patches with p53 mutations show no genetic link to synchronous squamous cell precancer or cancer in human skin. *Am. J. Pathol.* **150**, 1791-1803.

Retèl, J., Hoebee, B., Braun, J. E. F., Lutgerink, J. T., van den Akker, E., Wanamarta, A. H., Joenje, H. and Lafleur, M. V. M. (1993) Mutagenic specificity of oxidative DNA damage. *Mutat. Res.* **299**, 165-182.

Riazzudin, S. and Lindahl, T. (1978) Properties of 3-methyladenine-DNA glycosylase from *E. coli*. *Biochemistry* **17**, 2110-2118.

Richardson, K. K., Crosby, R. M., Richardson, F. C. and Skopek, T. R. (1987a) DNA base changes induced following in vivo exposure of unadapted, adapted or Ada<sup>-</sup> *Escherichia coli* to *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine. *Mol. Gen. Genet.* **209**, 526-532.

Richardson, K. K., Richardson, F. C., Crosby, R. M., Swenberg, J. A. and Skopek, T. R. (1987b) DNA base changes and alkylation following in vivo exposure of *Escherichia coli* to *N*-methyl-*N*-nitrosourea or *N*-ethyl-*N*-nitrosourea. *Proc. Natl. Acad. Sci. USA* **84**, 344-348.

Richardson, F., Boucheron, J. A., Skopek, T. R. and Swenberg, J. A. (1989) Formation of *O*<sup>6</sup>-methyldeoxyguanosine at specific sites in a synthetic oligonucleotide designed to resemble a known mutagenic hotspot. *J. Biol. Chem.* **264**, 838-841.

Ridanpää, M., Burvall, K., Zhang, L. H., Husgafvel-Pursiainen, K. and Onfelt, A.



- (1995) Comparison of DGGE and CDGE in detection of single base changes in the hamster *hprt* and human *N-ras* genes. *Mutat. Res.* **334**, 357-364.
- Robin, E. D. and Wong, R. (1988) Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. *J. Cell Physiol.* **136**, 507-513.
- Robinson, D. R., Goodall, K., Albertini, R. J., O'Neill, J. P., Finette, B., Sala-Trepat, M., Moustacchi, E., Bates, A. D., Beare, D. M., Green, M. H. L. and Cole, J. (1994) An analysis of *in vivo hprt* mutant frequency in circulating T-lymphocytes in the normal human population: a comparison of four datasets. *Mutat. Res.* **313**, 227-247.
- Rose, D.J. and Jorgenson, J.W. (1988) Characterization and automation of sample introduction methods for capillary zone electrophoresis. *Anal. Chem.* **60**, 642-648.
- Rossi, S. C., Conrad, M., Voigt, J. M. and Topal, M. D. (1989) Excision repair of *O*<sup>6</sup>-methylguanine synthesized at the rat *H-ras N*-methyl-*N*-nitrosourea activation site and introduced into *Escherichia coli*. *Carcinogenesis* **10**, 373-377.
- Rubinfeld, B., Souza, B., Albert, I., et al. (1993) Association of the APC gene product with  $\beta$ -catenin. *Science* **262**, 1731-1734.
- Ruiz-Martinez, M. C., Berka, J., Belenkii, A., Foret, F., Miller, A. W. and Karger, B. L. (1993) DNA sequence by capillary electrophoresis with replaceable linear polyacrylamide and laser-induced fluorescence detection. *Anal. Chem.* **65**, 2851-2858.
- Ruzicka, V., März, W., Russ, A. and Gross, W. (1992) Apolipoprotein B (Arg<sup>3500</sup> -> Gln) allele specific polymerase chain reaction: large-scale screening of pooled blood samples. *J. Lipid Res.* **33**, 1563-1567.
- Rydberg, B., Spurr, N. and Karran, P. (1990) cDNA cloning and chromosomal assignment of the human *O*<sup>6</sup>-methylguanine-DNA methyltransferase. *J. Biol. Chem.* **265**, 9563-9569.
- Saffhill, R. and Abbott, P. J. (1978) Formation of *O*<sup>2</sup>-methylthymine in poly(dA-dT) on methylation with *N*-methyl-*N*-nitrosourea and dimethyl sulphate. Evidence that *O*<sup>2</sup>-methylthymine does not miscode during DNA synthesis. *Nucleic Acids Res.* **5**, 1971-1978.
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Samson, L., Derfler, B., Boosalis, M. and Call, K. (1991) Cloning and characterization of a 3-methyladenine DNA glycosylase cDNA from human cells whose gene maps to chromosome 16. *Proc. Natl. Acad. Sci. USA* **88**, 9127-9131.
- Samson, L., Thomale, J. and Rajewsky, M. F. (1988) Alternative pathways for the *in vivo* repair of *O*<sup>6</sup>-alkylguanine and *O*<sup>4</sup>-alkylthymine in *Escherichia coli*: the adaptive response and nucleotide excision repair. *EMBO J.* **7**, 2261-2267.
- Sandy, M. S., Chiocca, S. M. and Cerutti, P. A. (1992) Genotypic analysis of mutations in *Taq* I restriction recognition sites by restriction fragment length polymorphism / polymerase chain reaction. *Proc. Natl. Acad. USA.* **89**, 890-894.

Sarkar, G., Cassady, J., Bottema, C. D. K. and Sommer, S. S. (1990) Characterization of polymerase chain reaction amplification of specific alleles. *Anal. Biochem.* **186**, 64-68.

Schendel, P. F. and Robins, P. E. (1978) Repair of O<sup>6</sup>-methylguanine in adapted *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **75**, 6017-6020.

Seibel, P., Flierl, A., Kottlors, M. and Reichmann, H. (1994) A rapid and sensitive PCR screening method for point mutations associated with mitochondrial encephalomyopathies. *Biochem. Biophys. Res. Commun.* **200**, 938-942.

Sekiya, T. (1993) Detection of mutant sequences by single-strand conformation polymorphism analysis. *Mutat. Res.* **288**, 79-83.

Sendowski, K. and Rajewsky, M. F. (1991) DNA sequence dependence of guanine-O<sup>6</sup> alkylation by the *N*-nitroso carcinogens *N*-methyl- and *N*-ethyl-*N*-nitrosourea. *Mutat. Res.* **250**, 153-160.

Seyama, T., Ito, T., Hayashi, T., Mizuno, T., Nakamura, N. and Akiyama, M. (1992) A novel blocker-PCR method for detection of rare mutant alleles in the presence of an excess amount of normal DNA. *Nucleic Acid Res.* **20**, 2493-2496.

Sheffield, V. C., Cox, D. R., Lerman, L. S., and Myers, R. M. (1989) Attachment of a 40-base-pair G+C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. *Proc. Natl. Acad. Sci. USA* **86**, 232-236.

Shevell, D. E., Friedman, B. M. and Walker, G. C. (1990) Resistance to alkylation damage in *Escherichia coli*: role of the Ada protein in induction of the adaptive response. *Mutat. Res.* **233**, 57-72.

Sibghat-Ullah and Day, R. S. (1993) DNA-substrate sequence specificity of human G:T mismatch repair activity. *Nucleic Acids Res.* **21**, 1281-1287.

Simpson, A. J. (1997) The natural somatic mutation frequency and human carcinogenesis. *Adv. Cancer Res.* **71**, 209-240.

Singer, B. (1979) *N*-nitroso alkylating agents: formation and persistence of alkyl derivatives in mammalian nucleic acids as contributing factors in carcinogenesis. *J. Natl. Cancer Inst.* **62**, 1329-1339.

Singer, B. (1986) O-alkyl pyrimidines in mutagenesis and carcinogenesis: occurrence and significance. *Cancer Res.* **46**, 4879-4885.

Skandalis, A., Ford, B. N. and Glickman, B. W. (1994) Strand bias in mutation involving 5-methylcytosine deamination in the human *hprt* gene. *Mutat. Res.* **314**, 21-26.

Skopek, T. R., Liber, H. L., Penman, B. W. and Thilly, W. G. (1978) Isolation of a human lymphoblastoid line heterozygous at the *thymidine kinase* locus: possibility for a rapid human cell mutation assay. *Biochem. Biophys. Res. Commun.*, **84**, 411-416.

Smith, C.A. and Mellon, I. (1990) Clues to the organization of DNA repair systems gained from studies of intragenomic repair heterogeneity. G. Obe (Ed.), *Advances in Mutagenesis*, Vol. 1, Springer, Berlin, pp. 153-194.

- Smith, J. and Modrich, P. (1996) Mutation detection with MutH, MutL, and MutS mismatch repair proteins. *Proc. Natl. Acad. Sci. USA* **93**, 4347-4379.
- Sommer, S. S., Cassady, J. D., Sobell, J. L. and Bottema, C. D. K. (1989) A novel method for detecting point mutations or polymorphisms and its application to population screening for carriers of phenylketonuria. *Mayo Clin. Pro.* **64**, 1361-1372.
- Soussi, T., Caron de Fromentel, C. and May, P. (1990) Structural aspects of the p53 protein in relation to gene evolution. *Oncogene* **5**, 945-952.
- Steingrimsdottir, H., Beare, D., Cole, J., Leal, J. F. M., Kostic, T., Lopez-Barea, J., Dorado, G. and Lehmann, A. R. (1996) Development of new molecular procedures for the detection of genetic alterations in man. *Mutat. Res.* **353**, 109-121.
- Su, L.-K., Vogelstein, B. and Kinzler, K.W. (1993) Association of the APC tumor suppressor protein with catenin. *Science* **262**, 1734-1737.
- Su, S.-S. and Modrich, P. (1986) *Escherichia coli* mutS-encoded protein binds to mismatched DNA base pairs. *Proc. Natl. Acad. Sci. USA* **83**, 5057-5061.
- Suzuki, Y., Orita, M., Shiraishi, M., Hayashi, K. and Sekiya, T. (1990) Detection of ras gene mutations in human lung cancers by single-strand conformation polymorphism analysis of polymerase chain reaction products. *Oncogen* **5**, 1037-1043.
- Swann, P. F. and Magee, P. N. (1968) Nitrosamine-induced carcinogenesis. The alkylation of nucleotide acids of the rat by N-methyl-N-nitrosourea, dimethylnitrosamine, dimethyl sulphate and methyl methanesulphonate. *J. Biochem.* **110**, 39-47.
- Swann, P. F. (1990) Why do O<sup>6</sup>-alkylguanine and O<sup>4</sup>-alkylthymine miscode? The relationship between the structure of DNA containing O<sup>6</sup>-alkylguanine and O<sup>4</sup>-alkylthymine and the mutagenic properties of these bases. *Mutat. Res.* **233**, 81-94.
- Tano, K., Shiota, S., Collier, J., Foote, R. S. and Mitra, S. (1990) Isolation and structural characterization of a cDNA clone encoding the human DNA repair protein for O<sup>6</sup>-alkylguanine. *Proc. Natl. Acad. Sci. USA* **87**, 686-690.
- Thiede, C., Bayerdörffer, E., Blasczyk, R., Wittig, B. and Neubauer, A. (1996) Simple and sensitive detection of mutations in the *ras* proto-oncogenes using PNA-mediated PCR clamping. *Nucleic Acid Res.* **24**, 983-984.
- Thilly, W. G. (1981) Chemicals, genetic damage, and the search for truth. *Technol. Rev.* **83**, 36-41.
- Thilly, W. G. (1985) Potential use of gradient denaturing gel electrophoresis in obtaining mutational spectra from human cells. In *Carcinogenesis*, eds. Huberman, E. & Barr, S. H. (Raven Press, New York), vol. 10, pp. 511-528.
- Thilly, W. G. (1990) Mutational spectrometry in animal toxicology testing. *Annu. Rev. Pharmacol. Toxicol.* **30**, 369-385.
- Thomas, D. C., Roberts, J. D., Sabatino, R. D., Myers, T. W., Tan, C. K., Downey, K. M., So, A. G., Bambara, R. A. and Kunkel, T. A. (1991) Fidelity of mammalian DNA replication and replicative DNA polymerases. *Biochemistry* **30**, 11751-11759.

- Todd, M. L. and Schendel, P. F. (1983) Repair and mutagenesis in *Escherichia coli* K-12 after exposure to various alkyl-nitrosoguanines. *J. Bacteriol.* **156**, 6-12.
- Tomita-Mitchell, A., Muniappan, B. P., Herrero-Jimenez, P., Zarbl, H. and Thilly, W. G. (1998) Single nucleotide polymorphism spectra in newborns and centenarians: identification of genes coding for risk of mortal disease. *Gene*, in press.
- Tommasi, S., Denissenko, M. F. and Pfeifer, G. P. (1997) Sunlight induces pyrimidine dimers preferentially at 5-methylcytosine bases. *Cancer Res.* **57**, 4727-4730.
- Toorchen, D. and Topal, M. D. (1983) Mechanisms of chemical mutagenesis and carcinogenesis: effects on DNA replication of methylation at the O<sup>6</sup>-guanine position of dGTP. *Carcinogenesis* **4**, 1591-1597.
- Topal, M. D., Eadie, J. S. and Conrad, M. (1986) O<sup>6</sup>-methylguanine mutation and repair is nonuniform. *J. Biol. Chem.* **261**, 9879-9885.
- Tornaletti, S. and Pfeifer, G. P. (1994) Slow repair of pyrimidine dimers at p53 mutation hotspots in skin cancer. *Science* **263**, 1436-1438.
- Van Houten, B. and Sancar, A. (1987) Repair of *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine-induced DNA damage by ABC excinuclease. *J. Bacteriol.* **169**, 540-545.
- Vireling, H., Thijssen, J. C. P., Rossi, A. M., van Dam, F. J., Natarajan, A. T., Tates, A. D. and van Zeeland, A. A. (1992) Enhanced *hprt* mutant frequency but no significant difference in mutation spectrum between a smoking and a non-smoking human population. *Carcinogenesis* **13**, 1625-1631.
- Vogelstein, B. and Kinzler, K. W. (1992) p53 function and dysfunction. *Cell* **70**, 523-526.
- Voigt, J. M., Van Houten, B., Sancar, A. and Topal, M. D. (1989) Repair of O<sup>6</sup>-methylguanine by ABC excinuclease of *Escherichia coli* *in vitro*. *J. Bio. Chem.* **264**, 5172-5176.
- Wagner, R., Debbie, P. and Radman, M. (1995) Mutation detection using immobilized mismatch binding protein (MutS). *Nucleic Acid Res.* **23**, 3944-3948.
- Walker, G. C. (1984) Mutagenesis and inducible responses to deoxyribonucleic acid damage in *Escherichia coli*. *Microbiol. Rev.* **48**, 60-93.
- Wani, A. A., Wani, G. and D'Ambrosio, S. M. (1990) A human DNA repair activity specific for O<sup>4</sup>-ethylthymine: identification and partial characterization. *Carcinogenesis* **11**, 1419.
- Weinstein, G.D., McCullough, J.L. and Ross, P. (1984) Cell proliferation in normal epidermis. *J Invest Dermatol.* **82**, 623-628.
- Weinstein, L. S., Geiman, P. V., Friedman, E., Kadowake, T., Collins, R. M., Gershon, E. S. and Spiegel, A. M. (1990) Mutations of the G<sub>s</sub>  $\alpha$ -subunit gene in Albright hereditary osteodystrophy detected by denaturing gradient gel electrophoresis. *Proc. Natl. Acad. Sci. USA* **87**, 8287-8290.

Wenz, H. M., Ramachandra, S., O'Connell, C. D. and Atha, D. H. (1998) Identification of known p53 point mutations by capillary electrophoresis using unique mobility profiles in a blinded study. *Mutat. Res.* **382**, 121-132.

Wiedmann, M., Barany, F. and Batt, C. A. (1993) Detection of *Listeria monocytogenes* with a nonisotopic polymerase chain reaction-coupled ligase chain reaction assay. *Appl. Environ. Microbiol.* **59**, 2743-2745.

William, L. D. and Shaw, B. R. (1987) Protonated base pairs explain the ambiguous pairing properties of O<sup>6</sup>-methylguanine. *Proc. Natl. Acad. Sci. USA* **84**, 1779-1783.

Wogan, G. N. (1992) Aflatoxins as risk factors for hepatocellular carcinoma in humans. *Cancer Res.* **52** (Suppl.), 2114s-2118s.

Woodcock, D. M., Crowther, P. J., Jefferson, S. and Diver, W. P. (1988) Methylation at dinucleotides other than CpG: implications for human maintenance methylation. *Gene* **74**, 151-152.

Wu, J. K., Ye, Z. and Darras, B. T. (1993) Sensitivity of single-strand conformation polymorphism (SSCP) analysis in detecting p53 point mutations in tumors with mixed cell populations. *Am. J. Hum. Genet.* **52**, 1273-1275.

Yamagata, Y., Kohda, K. and Tomita, K. (1988) Structural studies of O<sup>6</sup>-methyldeoxyguanosine and related compounds, a promutagenic DNA lesion by methylating carcinogens. *Nucleic Acids Res.* **16**, 9307-9321.

Yang, J.-L., Hu, M.-C. and Wu, C.-W. (1991) Novel mutational spectrum induced by N-methyl-N'-nitro-N-nitrosoguanidine in the coding region of the hypoxanthine (guanine) phosphoribosyltransferase gene in diploid human fibroblasts. *J. Mol. Biol.* **221**, 421-430.

Yang, J.-L., Lin, J.-G., Hu, M.-C. and Wu, C.-W. (1993) Mutagenicity and mutational spectrum of N-methyl-N'-nitro-N-nitrosoguanidine in the hprt gene in G1-S and late S phase of diploid human fibroblasts. *Cancer Res.* **53**, 2865-2873.

Zak, P., Kleibl, K. and Laval, F. (1994) Repair of O<sup>6</sup>-methylguanine and O<sup>4</sup>-methylthymine by the human and rat O<sup>6</sup>-methylguanine-DNA transferase. *J. Biol. Chem.* **269**, 730-733.

Zhang, C., Linnane, A. W. and Nagley, P. (1993) Occurrence of a particular base substitution (3243 A to G) in mitochondrial DNA of tissues of aging humans. *Biochem. Biophys. Res. Commun.* **195**, 1104-1110.

Zielenska, M., Horsfall, M. J. and Glickman, B. W. (1989) The dissimilar mutational consequences of SN1 and SN2 DNA alkylation pathways: clues from the mutational specificity of dimethyl sulfate in the *lacI* gene of *Escherichia coli*. *Mutagenesis* **4**, 230-234.