A CLAMP LIGATION METHOD FOR POINT MUTATIONAL SPECTROMETRY: MARKED INCREASE IN SCANNING RANGE FOR THE HUMAN GENOME

by

Andrea Seungsun Kim

B.S., Biochemistry
Western Washington University, 1994

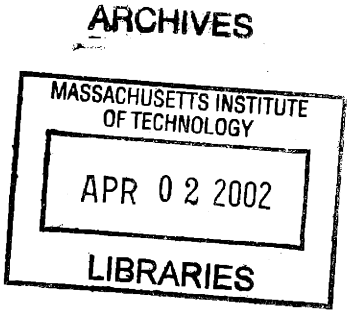
Submitted to the Division of Bioengineering and Environmental Health
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Toxicology

at the

Massachusetts Institute of Technology

February, 2002



Signature of Author:_____
Division of Bioengineering and Environmental Health
December 12, 2001

Certified by:_____
William G. Thilly
Professor of Toxicology
Thesis Advisor

Accepted by:_____
Peter C. Dedon
Professor of Bioengineering and Environmental Health
Chairman, Division Committee on Graduate Studies

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

Professor Peter C. Dedon ___ Chairman

Professor William G. Thilly _____ Thesis Advisor

Professor Bevin P. Engelward ____

Dr. Gerald P. Holmquist _____

Dr. Howard L. Liber _____

A CLAMP LIGATION METHOD FOR POINT MUTATIONAL SPECTROMETRY: MARKED INCREASE IN SCANNING RANGE FOR THE HUMAN GENOME

by Andrea Seungsun Kim

Submitted to the Division of Bioengineering and Environmental Health on December 12, 2001 in Partial Fulfillment of the Requirements for the Degree of Philosophy in Toxicology

ABSTRACT

The study of human mutagenesis requires methods of measuring somatic mutations in normal human tissues and inherited mutations in human populations. Such methods should permit measurement of rare mutations in the presence of abundant wild-type copies and should be general to the human genome.

A sensitivity of 2×10^{-6} for point mutations was recently achieved in human cells using a novel method of target isolation, constant denaturant capillary electrophoresis (CDCE), and high-fidelity polymerase chain reaction (hifi-PCR) (Li-Sucholeiki and Thilly, 2000). This method is applicable to 100-base pair (bp) DNA domains juxtaposed with a naturally occurring domain of a higher melting temperature, or a natural clamp. Such sequence domains represent about 9% of the human genome.

To permit analysis of rare point mutations in the human genome more generally, this thesis developed a procedure in which a clamp can be ligated to any 100-bp sequence of interest. This procedure was combined with the previous method to create a new method of point mutational analysis that is not dependent on a naturally occurring clamp. To demonstrate the new method, a sequence with a natural clamp, a part of the human hypoxanthine-guanine phosphoribosyl transferase (*HPRT*) gene (cDNA-bp 223-318), was analyzed using both the natural and ligated clamps.

A sensitivity of 2×10^{-5} in human cells was demonstrated using the ligated clamp as opposed to 5×10^{-6} using the natural clamp. The sensitivity of the new method using the ligated clamp was demonstrated to be limited by the fidelity of *Pfu* DNA polymerase used for PCR. The sequence of the ligated clamp accounted for the differences in sensitivity as a result of causing a decreased efficiency of mutant enrichment by CDCE.

The new method can be applied to measure somatic mutations in normal human tissues, such as lung tissues, in which point mutations at fractions above 10^{-5} have been observed. This method can also detect predominant inherited mutations even for genes carrying recessive deleterious alleles in pooled samples derived from a large number of individuals.

Thesis Advisor: William G. Thilly
Title: Professor of Toxicology

This thesis is dedicated to all of my family members, especially those who did not get to live up to their potential. Each one of their lives allowed me to learn many valuable lessons in life. Without these lessons, I would not have been able to complete this thesis. I am thankful to you all!

ACKNOWLEDGEMENTS

I would like to begin by thanking my thesis advisor, Professor William G. Thilly. While many years of working with him have been a very challenging experience, these years taught me to reach to my higher self. I envy his high scientific ambitions and his important and necessary work searching for the truth in "what mutates people."

I also thank my thesis committee members, Professor Peter C. Dedon, Professor Bevin P. Engelward, Dr. Gerry P, Holmquist, and Dr. Howard L. Liber. Their guidance and time meant a lot when starting, as well finishing, this thesis.

I thank Dr. Xiao-cheng Li-Sucholeiki and her brilliant technological development. Without her development, this thesis would never have been started.

I would like to acknowledge Dr. Paulo André for teaching me the basics of PCR and CDCE during those first few months in Thilly lab. I am especially thankful for the shared experience of learning the importance of the fidelities of DNA polymerases in mutation research.

Mr. Allen Zhang, a former employee of the Beckman Research Institute, taught me how to ligate DNA molecules. I thank him for his time and effort during those very first few days of working on DNA ligation.

Sensitivity demonstration using the ligated clamp in human cells was achieved after many failures. I thank Ms. Jacklene Goodluck-Griffith and Dr. Aoy Tomita-Mitchell for sharing a copious number of human cells, allowing these failures.

I would like to thank all Thilly lab members, past and present. I am especially thankful for having the opportunities to learn people with diverse backgrounds.

I thank Ms. Debra Luchanin, a former academic administrator of BEH, for her administrative work throughout the years. I also appreciate her caring heart for the students in the division.

Last, but not least, I would like to acknowledge my writing coaches, Ms. Pamela Siska and Ms. Susan Spilecki, at the MIT Writing Center. This thesis would not have been possible without numerous sessions with them. It was the teamwork that made completion of this document possible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREBIATIONS

A	Adenine
AF B_1	Aflatoxin B_1
AP	Apurinic/apyrimidinic
APC	Adenomatous polyposis coli
APS	Ammonium persulfate
bp	Base pair
BP	Benzo[a]pyrene
BPDE	Benzo[a]pyrene 7,8-diol-9,10-epoxides
BSA	Bovine serum albumin
C	Cytosine
CDCE	Constant denaturant capillary electrophoresis
CDGE	Constant denaturant gel electrophoresis
CE	Capillary electrophoresis
DGGE	Denaturing gradient gel electrophoresis
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
G	Guanine
HGMD	Human Gene Mutation Database
hifi-PCR	High-fidelity PCR
HPLC	High-performance liquid chromatography
HPRT	Hypoxanthine-guanine phosphoribosyl transferase

IARC	International Agency for Research on Cancer
id	Inner diameter
kb	Kilobases
LCR	Ligation chain reaction
5mC	5-Methylcytosine
mf	Mutant fraction
MLTP	Mutations leading to truncated proteins
MNNG	N-methyl-N'-nitro-N-nitrosoguanidine
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
Pfu	Pyrococcus furiosus
RFLP	Restriction fragment length polymorphism
rRNAs	Ribosomal RNAs
SDS	Sodium dodecyl sulfate
SNPs	Single-nucleotide polymorphisms
SS	Single strand
SSPE	Sodium chloride, sodium phosphate, EDTA
T	Thymine
Taq	Thermus aquaticus
TBE	Tris-borate, EDTA
TBEB	Tris-borate, EDTA, BSA
TE	Tris-HCl, EDTA

TEMED	N, N, N', N'-tetramethylethylenediamine
6TG	6-Thioguanine
tk	Thymidine kinase
T_m	Melting temperature
tRNAs	Transfer RNAs
UMD	Universal Mutation Database
UV	Ultraviolet light
Vent	Thermococcus litoralis
VHL	Von Hippel-Lindau

1 INTRODUCTION

Many mutations have been found in both inherited and somatic cases of human disease (Bos, 1989; Kogelnik et al., 1998; Béroud et al., 2000; Krawczak et al., 2000). However, causes of human mutation are yet to be determined. Possible sources of human mutation are exogenous agents, endogenous metabolites, and DNA replication errors (Lindahl, 1993; Greenblatt et al., 1994; Coller and Thilly, 1998; Hussain and Harris, 1998; Krawczak et al., 1998; Loeb and Loeb, 2000).

It has been hypothesized that if environmental agents are the primary cause of human mutation, agent-specific distributions of mutations in a given DNA sequence are expected in individuals exposed to these agents (Cariello and Thilly, 1986; Thilly, 1993; Coller and Thilly, 1994). However, testing such a hypothesis is hindered by the lack of genotype-based methods of measuring somatic mutations in normal tissues. Such methods should permit measurement of rare mutations in the presence of their abundant wild-type and should be general to the human genome. Methods that can fulfill these requirements should also allow analysis of inherited mutations in pooled samples derived from a large number of individuals.

The ability to analyze mutations based on their genotype, rather than phenotype, offers a larger pool of target genes and tissues. Most phenotype-based methods are only applicable to cultured cells and to certain tissue types, the cells of which can be grown *in vitro* (e.g., blood). In addition, phenotype-based methods are limited to selectable genes. Thus, genotype-based methods are the methods of choice. However, genotype-based point mutation detection methods for rare mutational analysis offer a small target size of 1- to 6-base pairs (bp) (Aguilar et al., 1994; Hussain et al., 1994; Nakazawa et al., 1994; Ouhtit et al., 1997; Wilson et al., 2000).

For genotype-based mutational analysis of a larger target size of about 100-bp, our laboratory has developed a method called constant denaturant capillary electrophoresis (CDCE) (Khrapko et al., 1994a). CDCE differentiates mutant from wild-type sequences based on differences in their melting temperatures. When combined with additional techniques, CDCE allows detection of point mutations in a variety of human samples (Khrapko et al., 1997a; Li-Sucholeiki and Thilly, 2000; Kim et al., 2001). A recent

technological development in which CDCE is combined with target isolation and high-fidelity polymerase chain reaction (hifi-PCR) allows analysis of mutations at fractions as low as 10^{-6} (Li-Sucholeiki and Thilly, 2000). This development can be applied to only those DNA sequences with a natural clamp, representing about 9% of the human genome.

For genotype-based mutational analysis of a larger target-pool size, this thesis presents a further technological development, a combined method of clamp ligation and target isolation/CDCE/hifi-PCR. This development opens up an additional 89% of the human genome suitable for analysis of rare point mutations. This combined method may allow understanding of human point mutagenesis by performing mutational analysis in a variety of human samples, including normal tissues.

2 LITERATURE REVIEW

2.1 Mutational spectra

Understanding of human mutagenesis requires mutational analysis in humans. Section 2.1.1 defines mutational spectra and discusses how mutational spectra can be used to understand human mutagenesis. Sections 2.1.2 and 2.1.3 review studies on point mutational spectra observed in a variety of human samples.

2.1.1 Definition and purpose of analysis

Benzer and Freese (1958), followed by Benzer (1961), primarily demonstrated mutational spectra, distributions of mutations with regard to position, kind, and frequency in a given DNA sequence. In these studies, each test mutagen was shown to induce a nonrandom distribution of mutations in the *r*II region of the phage T4 genome. These studies also showed that the induced mutations appear more frequently than expected by chance, called hotspot mutations. Since these studies, analysis of mutational spectra has been extended to bacteria, yeast, as well as human cells and tissues (Coulondre and Miller, 1977; Armstrong and Kunz, 1990; Cariello et al., 1990; Khrapko et al., 1997a).

Mutational spectra have been suggested as a tool to understand cause-and-effect relationships (Cariello and Thilly, 1986; Thilly, 1993; Coller and Thilly, 1994). These studies proposed that if environmental agents are the primary cause of human mutation, agent-specific mutational spectra are expected in individuals exposed to these agents. Exogenous agents, endogenous metabolites, and DNA replication errors have been hypothesized as possible sources of human mutation (Lindahl, 1993; Greenblatt et al., 1994; Coller and Thilly, 1998; Hussain and Harris, 1998; Krawczak et al., 1998; Loeb and Loeb, 2000). However, causes of human mutation are yet to be determined. Causes of human mutation may be revealed by a sensitive and general means of measuring mutational spectra in a variety of human samples.

2.1.2 Point mutational spectra in human HPRT gene in vitro

For mutational analysis of a larger pool of target genes and tissues, genotype-based methods are preferred over phenotype-based methods. Most phenotype-based methods are only applicable to cultured cells and to certain tissue types, the cells of which can be grown *in vitro* (e.g., blood). In addition, phenotype-based methods are limited to selectable genes. Despite these limitations, earlier mutational studies have been primarily based on reference to phenotypic selection. This section reviews studies on point mutational spectra observed in cultured human cells using the hypoxanthine-guanine phosphoribosyl transferase (*HPRT*) gene as a selectable marker.

The human *HPRT* gene is comprised of 57,000-bp and consists of nine exons (Edwards et al., 1990). This gene is on the X-chromosome, present as one copy per cell in males (Stout and Caskey, 1985). Many *HPRT* mutations have been found to be associated with inherited cases of human disease, such as Lesch-Nyhan syndrome (Stout and Caskey, 1985; Mohrenweiser and Jones, 1990; Cariello and Skopek, 1993a; Krawczak et al., 2000). This gene is related not only to disease but can be used as a selectable marker. The *HPRT* enzyme is involved in the purine salvage pathway, maintaining purine nucleotide pools in cells (Krenitsky et al., 1969; Stout and Caskey, 1985). Thus, *HPRT* mutants that alter physiological function of the gene product can be selected *in vitro* by their ability to grow in the presence of purine analogs, such as 6-thioguanine (6TG). 6TG is toxic when incorporated into a cell's DNA (Stout and Caskey, 1985).

An *en masse* selection-based method has been developed for analysis of point mutational spectra in selectable genes, such as *HPRT*, in vitro (Keohavong and Thilly, 1992a). Human cells are exposed to each test mutagen of a fixed condition, and mutant cells resistant to a selecting agent (6TG) are *en masse* selected. After polymerase chain reaction (PCR) amplification of a chosen target, selected and amplified mutants are analyzed by denaturing gradient gel electrophoresis (DGGE). DGGE is a DNA separation technique, which differentiates mutant from wild-type sequences based on differences in their melting temperatures (Fischer and Lerman, 1983).

This combined method of *en masse* selection, PCR, and DGGE allows detection of mutations, each representing as low as 1% of selected cells ($\geq 10^{-2}$), with statistical significance (±20% precision) (Keohavong and Thilly, 1992a). The advantage of this *en masse* selection-based method over a clone-by-clone approach is its ability to analyze all detectable point mutations in a chosen target simultaneously. Generating the same outcome by a clone-by-clone approach requires DNA sequencing of numerous mutant clones that are individually selected, thereby making the analysis rather laborious.

The combined method of *en masse* selection, PCR, and DGGE has been applied to analysis of point mutational spectra in a human lymphoblastoid line, TK6. Figure 1 illustrates the distributions of hotspot point mutations with regard to position, kind, and frequency in a part of the human *HPRT* gene. These distributions show that the point mutational spectra in a cultured human cell system are characterized by test mutagens to which the system is exposed. The test mutagens are ICR-191, *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (MNNG), ultraviolet light (UV), and benzo[a]pyrene 7,8-diol-9,10-epoxides (BPDE). Such mutagen-specific mutational spectra are shown to be different from the spontaneous spectrum (see Figure 1).

More importantly, different mutagen-exposure conditions produce different point mutational spectra. As Figure 2 shows, the AHH-1 human lymphoblastoid line exposed to benzo[a]pyrene (BP) under conditions differing in concentration and duration produced dissimilar point mutational spectra.

These results based on phenotypic selection indicate that mutational spectra observed in cultured human cells treated with a mutagen at high dose(s) for a short term cannot be used to predict human *in vivo* mutagenesis. Thus, understanding of human *in vivo* mutagenesis requires mutational analysis in human tissues. However, mutation detection methods based on phenotypic selection are limited to selectable genes, such as *HPRT*, and certain tissue types, such as blood. For these reasons, genotype-based methods are necessary for analysis of mutational spectra in a larger pool of target genes and tissues.

Figure 1. Mutagen-specific point mutational spectra in human cells (TK6)

Human cells, TK6, were exposed to each test mutagen of a fixed condition or left untreated. Mutant cells resistant to 6TG were then *en masse* selected. After PCR amplification of a chosen target, selected and amplified mutants were analyzed by DGGE. DGGE allows analysis of point mutations with regard to position, kind, and frequency in a chosen DNA sequence, thereby generating point mutational spectra.

Sources: ICR-191 and MNNG (Cariello et al., 1990); BPDE (Keohavong and Thilly, 1992b); UV (Keohavong et al., 1991); Spontaneous (Oller and Thilly, 1992).

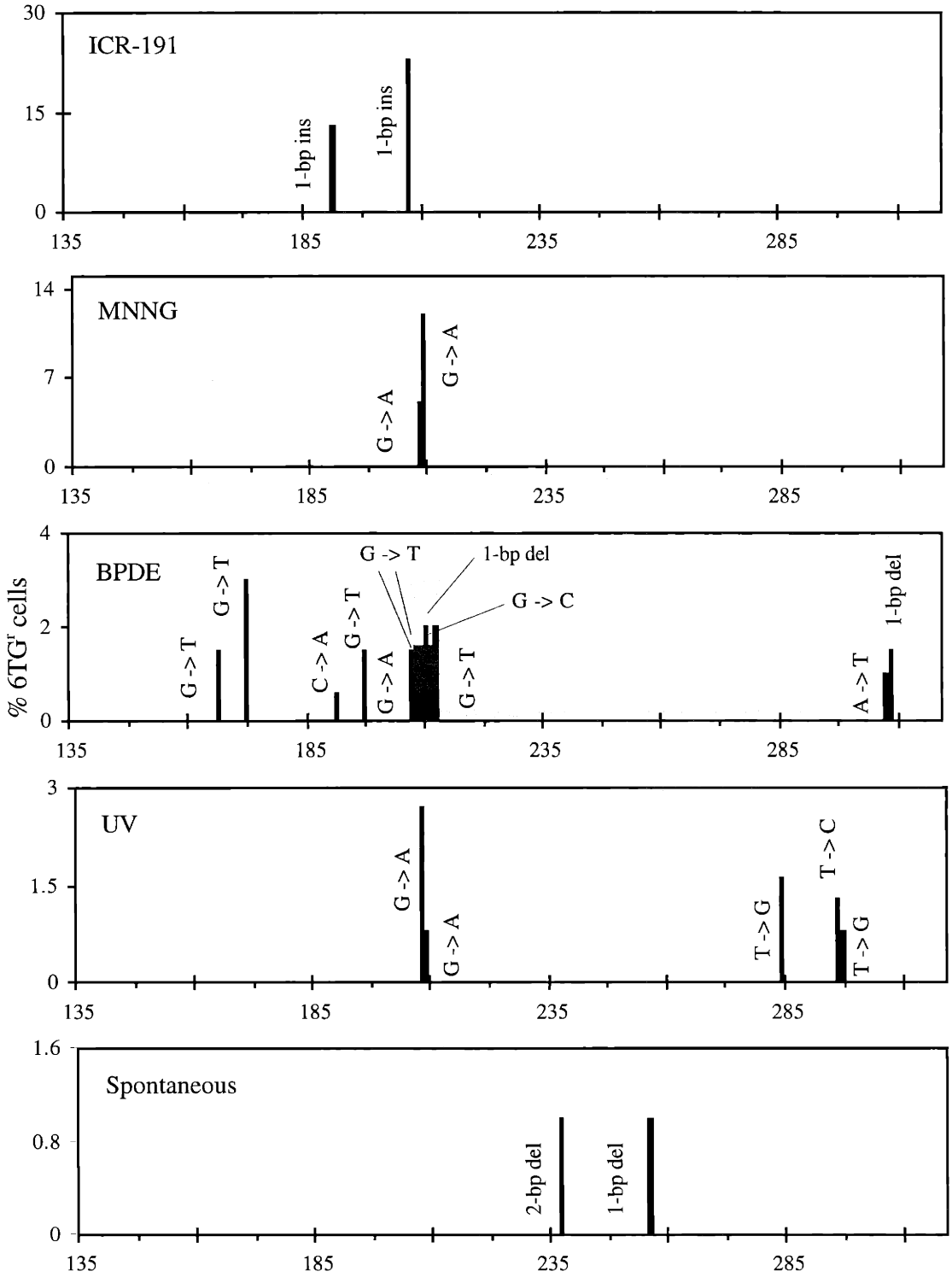
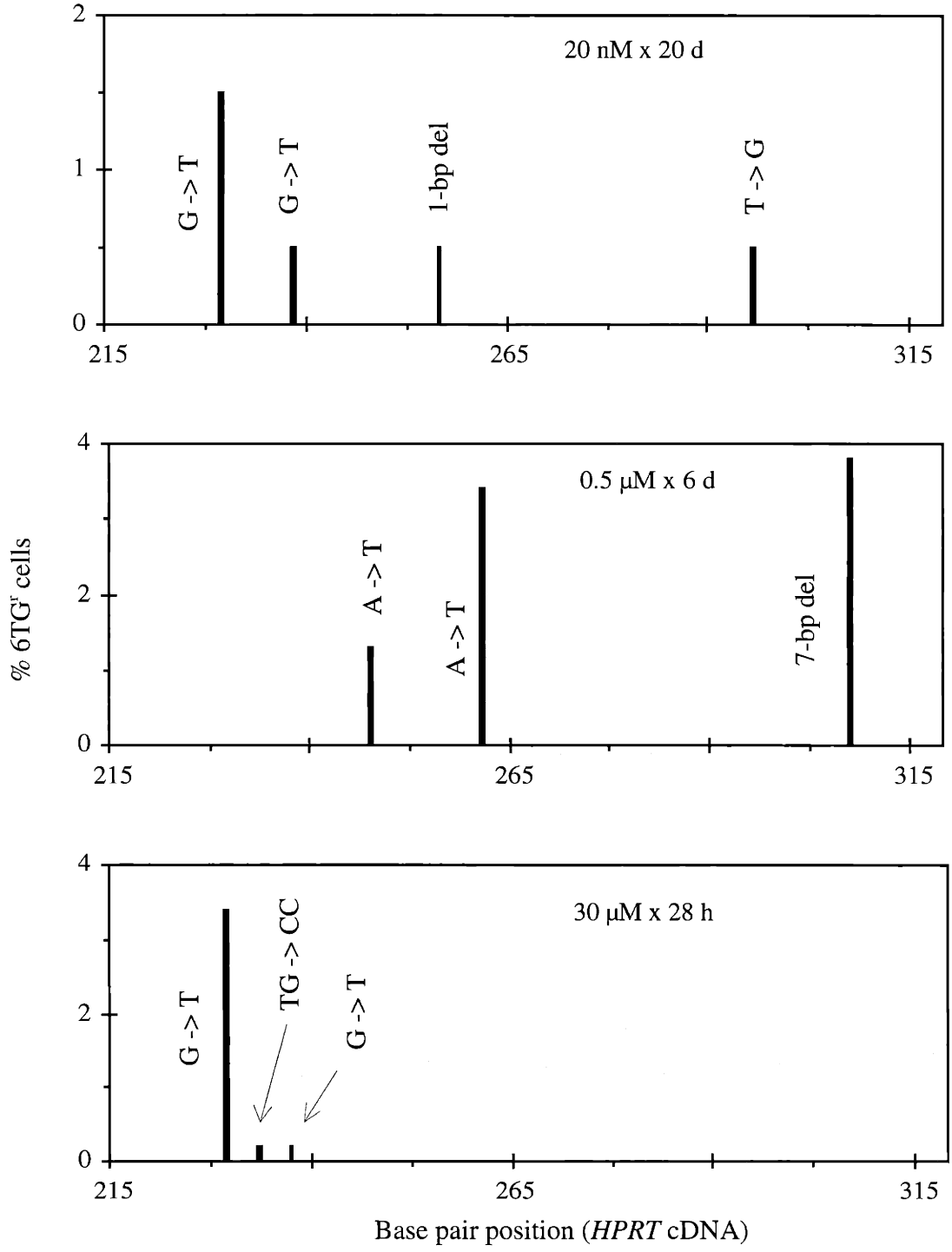

ICR-191
1-bp ins
1-bp ins
MNNG
G -> A
G -> A
BPDE
% 6TG^r cells
G -> T
G -> T
C -> A
G -> T
G -> A
G -> T
G -> T
1-bp del
G -> C
A -> T
1-bp del
UV
G -> A
G -> A
T -> G
T -> C
T -> G
Spontaneous
2-bp del
1-bp del
30
15
0
14
7
4
2
3
1.5
1.6
0.8
135
185
235
285
Base pair position (*HPRT* cDNA)

Figure 2. BP exposure condition-specific point mutational spectra in human cells (AHH-1)

Human cells, AHH1, were treated with BP of different concentrations for different periods of time. Mutant cells resistant to 6TG were then *en masse* selected. After PCR amplification of a chosen target, selected and amplified mutants were analyzed by DGGE. DGGE allows analysis of point mutations with regard to position, kind, and frequency in a chosen DNA sequence, thereby generating point mutational spectra.

Source: Chen and Thilly, 1996.


20 nM x 20 d
G -> T
G -> T
1-bp del
T -> G
0.5 μM x 6 d
A -> T
A -> T
7-bp del
30 μM x 28 h
G -> T
TG -> CC
G -> T
% 6TG' cells
Base pair position (HPRT cDNA)
0
1
2
4
215
265
315

2.1.3 Point mutational spectra in human mitochondrial DNA in vitro and in vivo

Using a genotype-based mutation detection method, point mutational spectra in a part of human mitochondrial DNA have been analyzed in a variety of human samples, including normal tissues (Khrapko et al., 1997a; Coller et al., 1998; Marcelino et al., 1998). This section reviews studies related to such analysis and summarizes the analysis results, allowing understanding of human mitochondrial mutagenesis.

The human mitochondrial genome, distinct from the nuclear genome, is a circular, double-stranded DNA of 16, 569-bp, inherited from the mother (Giles et al., 1980; Anderson et al., 1981; Attardi and Schatz, 1988). It encodes for 13 protein subunits that are involved in synthesizing the respiratory chain complex (Anderson et al., 1981; Attardi and Schatz, 1988). It also encodes for 2 ribosomal RNAs (rRNAs) and 22 transfer RNAs (tRNAs) essential for mitochondrial protein synthesis (Anderson et al., 1981; Attardi and Schatz, 1988). Many mitochondrial mutations have been found to be associated with human aging and disease (Kogelnik et al., 1998; Chinnery and Turnbull, 1999; Wallace, 1999; Cottrell et al., 2000).

When analyzing mutations in human samples, two factors make mitochondrial DNA a more attractive target than nuclear DNA. The first is its multicopy nature: it is present as a few hundred to a thousand copies per cell (Robin and Wong, 1988; Marcelino et al., 1998; Khrapko et al., 1999). This multicopy nature allows detection of mitochondrial mutations using a smaller sample size than that necessary for single copy nuclear mutations. Assuming a thousand mitochondrial DNA copies per cell, a thousand times smaller sample size is expected to be necessary for detection of mitochondrial mutations compared to that required for nuclear mutations of the same fraction. The second factor is the higher mutation rate observed for mitochondrial DNA compared to that for nuclear DNA (Khrapko et al., 1997a). This factor is expected to generate higher mitochondrial mutant fractions in organ tissues. This expectation suggests that a means of analyzing mitochondrial mutations does not have to be as sensitive as is necessary for nuclear mutations.

A genotype-based method has been developed for analysis of point mutations in 100-bp multicopy sequences (e.g., ≈1000 copies per cell) (Khrapko et al., 1997b). The

sensitivity of this method is 10^{-6}, detecting at least 10^2 copies of each mutant in the presence of 10^8 copies of the wild-type. This method is based on constant denaturant gel electrophoresis (CDGE), hifi-PCR, and CDCE. CDGE (Hovig et al., 1991) and CDCE (Khrapko et al., 1994a), both derived from DGGE, separate mutant from wild-type sequences based on differences in their melting temperatures.

For mutational analysis by this genotype-based method, human genomic DNA is restriction digested to liberate a chosen target. Mutations in a chosen target from the wild-type are then separated by CDGE, followed by elution of these separated mutations. These separation and elution procedures allow genotype-based enrichment of mutant sequences relative to the wild-type. Mutant-enriched samples are PCR amplified and are separated by CDCE for further mutant enrichment. PCR is then performed prior to mutational analysis by CDCE.

This combined method of CDGE, PCR, and CDCE has been applied to measure mitochondrial point mutational spectra in a variety of human samples (Khrapko et al., 1997a; Coller et al., 1998; Marcelino et al., 1998). In the study of Khrapko et al. (1997a), normal human tissues, their derived tumors, and human cells grown under pristine conditions were analyzed. In all of these samples, the same set of 17 point mutations with regard to position and kind was observed in a 100-bp mitochondrial sequence. Most of these mutations are G to A and A to G transitions, with the average mutant fractions ranging from 10^{-6} to 4×10^{-4}. When the same mitochondrial sequence was analyzed in human bronchial epithelial cells of smoking and nonsmoking twins, the same set of 17 point mutations as in Khrapko et al. (1997a) was observed (Coller et al., 1998). In addition, Coller et al. (1998) observed no increase in the overall mutant fractions in the smokers' samples compared to those in the nonsmokers'. These studies suggest the spontaneous origin of the observed human mitochondrial point mutations.

DNA replication errors and endogenous damage have been hypothesized as probable spontaneous sources of the set of mitochondrial point mutations observed in Khrapko et al. (1997a), as well as in Coller et al. (1998) (Khrapko et al., 1997a; Marcelino and Thilly, 1999). To date, one of the hypotheses has been tested using human DNA polymerase γ and using the same mitochondrial sequence studied in Khrapko et al. (1997a) and Coller et al. (1998) (Zhang, unpublished results). In this study, about 60% of

polymerase γ-generated hotspot point mutations was determined to be the same as those observed in Khrapko et al. (1997a) and Coller et al. (1998). DNA polymerase γ is the only known polymerase responsible for replicating human mitochondrial DNA (Shadel and Clayton, 1997).

These mutational studies suggest the spontaneous origin of human mitochondrial mutation. While the same type of studies remains to be completed for nuclear sequences, such studies are hindered by the lack of a general and sensitive means of measuring mutations in normal tissues. A desired means should permit measurement of rare nuclear mutations in the presence of their abundant wild-type and should be general to the human genome.

2.2 Nuclear genes associated with human cancer

Mutational analysis of nuclear genes in human tissues is necessary to understand human nuclear mutagenesis. Proto-oncogenes and tumor suppressor genes are the primary nuclear genes of interest since many mutations in these genes have been found to be associated with human cancer (Bos, 1989; Béroud et al., 2000; Krawczak et al., 2000). Cancer is the one of the leading causes of death in the United States (Vital statistics of the United States). This section reviews studies on these primary nuclear genes of interest.

Oncogenes are mutant forms of their normal genes, proto-oncogenes (Diamandis, 1997). The normal gene products are mainly involved in signaling pathways, positively regulating cell division (Wynford-Thomas, 1991; Yarbo, 1992; Baserga et al., 1993). Mutations in proto-oncogenes can deregulate cell growth by expressing an increased level of the gene products, leading to malignant transformation (Wynford-Thomas, 1991; Diamandis, 1997). Such mutations are called gain-of-function, or activating, mutations (Haber and Fearon, 1998). Oncogenes are thought to be 'dominant' genes since one mutated allele is capable of transforming cells (Diamandis, 1997).

Unlike oncogenes, tumor suppressor genes are thought to be 'recessive' genes since mutations in both alleles are necessary for malignant transformation (Brown, 1997; Diamandis, 1997). Most tumor suppressor genes are negative regulators of cell growth and development. Mutations in these genes can free cells from negative growth signals

and lead to malignant transformation (Weinberg, 1991; Wynford-Thomas, 1991; Yarbro, 1992; Brown, 1997). Such mutations are called loss-of-function, or inactivating, mutations (Brown, 1997; Haber and Fearon, 1998).

Tables 1 and 2 summarize the characteristics of selected human proto-oncogenes and tumor suppressor genes, respectively. The major mutation type(s) found in cancer is specific to each gene. In addition, each gene is associated with one, or multiple cancer types or multiple genes are associated with one cancer type. Thus, gene, mutation type, and tissue type of interest should be considered together when planning mutational analysis in human tissues.

2.3 Nuclear mutations associated with human disease

Many mutations in nuclear genes have been found to be associated with a variety of human diseases (Béroud et al., 2000; Krawczak et al., 2000). Studies on mutations associated with human disease can be used as a guide when planning mutational analysis of a variety of human samples, including normal tissues. This section reviews studies on nuclear mutations found in either inherited or somatic cases of human disease.

2.3.1 Inherited mutations

A comprehensive core collection of published germline mutations associated with human disease has been compiled as a database, the Human Gene Mutation Database (HGMD) (Cooper et al., 1998; Krawczak et al., 2000). Somatic gene mutations and mitochondrial genome mutations are not included in this database. To date, the database comprises over 20, 000 different mutations, each found in coding, regulatory, or splicing-relevant regions of about 1,000 different nuclear genes.

When the HGMD entries were sorted by mutation type, a hierarchy was found, as Table 3 summarizes (Cooper et al., 1998; Krawczak et al., 2000). Point mutations are the major mutation type observed. Point mutations include single base-pair substitutions (70.1%), as well as small (≤ 20-bp) insertions and deletions (22.5%). A hierarchy was also found when the single base-pair substitutions found in gene-coding regions were

Table 1. Characteristics of selected human proto-oncogenes

Gene	Function	Major mutation type(s)	Cancer type(s)
K-ras	p21 guanine trinucleotide phosphatase	Point mutation	Pancreatic, colorectal, lung, endometrial, other carcinomas
N-ras	p21 guanine trinucleotide phosphatase	Point mutation	Myeloid leukemia
H-ras	p21 guanine trinucleotide phosphatase	Point mutation	Bladder
EGFR (ERB-B)	Growth-factor receptor	Amplification	Gliomas, squamous and other carcinomas
NEU (ERB-B2)	Growth-factor receptor	Amplification	Breast, ovarian, gastric, other carcinomas
C-myc	Transcription factor	Chromosome translocation, amplification	Burkitts' lymphoma, small-cell carcinoma of the lung, other carcinomas
N-myc	Transcription factor	Amplification	Neuroblastoma, small-cell carcinoma of the lung
L-myc	Transcription factor	Amplification	small-cell carcinoma of the lung
BCL-2	Antiapotosis protein	Chromosome translocation	B-cell lymphoma
CYCD1	Cyclin-D, cell-cycle control	Amplification, chromosome translocation	Breast and other carcinomas, B-cell lymphoma, parathyroid adenomas
BCR-ABL	Chimaeric non-receptor tyrosine kinase	Chromosome translocation	Chronic myelogenous leukemia, acute lymphocytic leukemia
RET	Glial-derived neutropic facotr-receptor tyrosine kinase	Chromosome translocation, point mutation	Thyroid
CDK4	Cyclin-dependent kinase	Amplification, point mutation	Sarcoma
SMO	Transmembrane signaling molecule in sonic hedgehog pathway	Point mutation	Basal-cell skin
β-CAT	Transcriptional co-activator, links E-cadherin to cytoskeleton	Point mutation, in-frame deletion	Melanoma, colorectal
HST	Growth factor	Amplification	Gastric
PML-RAR	Chimaeric transcription factor	Chromosome translocation	Acute promyelocytic leukemia
W2A-PBX1	Chimaeric transcription factor	Chromosome translocation	Pre-B acute lymphocytic leukemia
MDM-2	P53-binding protein	Amplification	Sarcoma
GLI	Transcription factor	Amplification	Sarcoma, glioma
TTG	Transcription factor	Chromosome translocation	T-cell acute lymphocytic leukemia

Adapted from Haber and Fearon, 1998; Verma and Triantafillou, 1998.

Table 2. Characteristics of selected human tumor suppressor genes

Gene	Function	Major mutation type(s)	Cancer type(s)
RB1	Transcriptional regulator	Deletion	Retinoblastoma, osteosarcoma, small-cell lung carcinoma of the lung, breast, and bladder
p53	Transcription factor	Point mutation	50% of all cancers
APC	Regulator of β-catenin function	Truncating mutation	Colorectal
WT1	Transcription factor	Point mutation	Wilms' tumor, renal
NF1	Regulator of G protein-mediated signal transduction	Truncating mutation	Melanoma, neuroblastoma
NF2	Juxta-membrane link to cytoskeleton	Truncating mutation	Schwannomas, meningiomas, ependymomas
p16	Cyclin-dependent kinase inhibitor	Truncating mutation	
VHL	Modulator of RNA polymerase II	Truncating and point mutations	Renal, central nervous system
BRCA1	DNA repair	Truncating mutation	Breast, ovary
BRCA2	DNA repair	Truncating mutation	Breast

Adapted from Brown, 1997; Haber and Fearon, 1998; Bennett et al., 1999; Robertson et al., 1999.

Table 3. Number of HGMD entries by mutation type

Mutation type	Number of entries (% total)
Single base-pair substitutions	16359 (70.1%)
Small deletions (≤ 20-bp)	3849 (16.5%)
Small insertions (≤ 20-bp)	1407 (6.0%)
Small indels (≤ 20-bp)	153 (0.7%)
Repeat variations	30 (0.1%)
Gross insertions and duplications (≥ 20-bp)	137 (0.6%)
Complex rearrangements (including inversions)	244 (1.0%)
Gross deletions (≥ 20-bp)	1166 (5.0%)
Total	23345 (100.0%)

Adapted from Krawczak et al., 2000 (date of update: Oct., 2001).

sorted by nucleotides with respect to their propensity to undergo substitution. Table 4 summarizes this hierarchy: G (38.4%) > C (32.2%) > T (17.0%) > A (12.4%) (Krawczak et al., 1998). Among these substitutions, transitions (62.5%) are dominant over transversions (37.5%) (Krawczak et al., 1998).

Based on analysis of the HGMD entries, it has been hypothesized that these mutations are mainly the result of endogenous mutagenic processes (Krawczak et al., 1998). Such processes include spontaneous deamination of 5-methylcytosine (5-mC) and misalignment mutagenesis during DNA synthesis. These processes can potentially generate point mutations.

Spontaneous deamination of 5-mC has been hypothesized to be a cause of C to T and G to A transitions at CpG dinucleotides (Krawczak et al., 1998). These mutations account for 23% of all HGMD single base-pair substitutions and for 36.9% of HGMD transitions found in gene-coding regions (Krawczak et al., 1998). In vertebrate genomes, about 3 to 4% of all cytosines are methylated (5-mC), with 90 to 100% of these 5-mC occurring in the sequence CpG (Riggs and Jones, 1983; Antequera and Bird, 1993). It has been shown that 5-mC undergoes spontaneous deamination, forming thymine, at a rate much higher than the deamination of cytosine, forming uracil (Shen et al., 1994). The formed uracil can be removed by uracil-glycosylase (Lindahl, 1982). However, the formed thymine, when left unrepaired, basepairs with adenine, converting a methyl CpG sequence into TpG (Antequera and Bird, 1993; Gonzalgo and Jones, 1997). Base excision repair of U:G mismatches has been shown to be up to 6000-fold more efficient than that of T:G mismatches when tested with extracts from human tissues (Schmutte et al., 1995). In brief, the hypermutability of CpG dinucleotides is likely caused by spontaneous deamination of 5-mC combined with inefficient repair of the resulting premutagenic DNA mismatches.

In addition to spontaneous deamination of 5-mC, misalignment mutagenesis during DNA synthesis has been also hypothesized as a cause of point mutations (Krawczak et al., 1998). Eukaryotic DNA polymerases have been shown to generate point mutations *in vitro*, presumably as a result of template-primer misalignments in addition to misincorporation of non-complementary nucleotides (Kunkel, 1990; Kunkel, 1992). Hotspot mutations have been shown to occur in a given DNA sequence, with mutation

Table 4. Spectrum of HGMD single base-pair substitutions in gene-coding regions[a]

Original nucleotide	Number of substitutions by				Total (% total)
	T	C	A	G	
T		654	271	312	1,237 (17.0%)
C	1,632 (940)[b]		371	340	2,343 (32.2%)
A	201	163		538	902 (12.4%)
G	619	453	1,717 (735)[c]		2,789 (38.4%)
Total	2,452	1,270	2,359	1,190	7,271 (100.0%)

[a] Adapted from Krawczak et al., 1998.
[b] Number in parentheses is C to T transitions at CpG dinucleotides.
[c] Number in parenthesis is G to A transitions at CpG dinucleotides.

rates depending on polymerases and neighboring nucleotides (Kunkel, 1985; Kunkel, 1990; Kunkel, 1992).

Point mutations have also been found to be the major mutation type when germline mutations associated with human cancer were sorted by type in individual tumor suppressor genes (Béroud et al., 2000). Among these point mutations, the predominant mutations are small deletions in the adenomatous polyposis coli (*APC*) gene (Laurent-Puig et al., 1998; Muniappan, unpublished results) and single base-pair substitutions in the *p53* gene (Béroud and Soussi, 1998), as well as in the von Hippel-Lindau (*VHL*) gene (Béroud et al., 1998). These predominant point mutations are distributed throughout each gene-coding region, as Figure 3 illustrates.

While point mutations are not the only mutation type associated with human disease, analysis of HGMD entries suggests point mutations as the primary mutation type to be studied. Thus, methods of measuring point mutations are the methods of choice when planning analysis of inherited mutations in a variety of human samples.

2.3.2 Somatic mutations

Studied on gene-specific mutations associated with human disease can be used as a guide when choosing a target gene sequence, in addition to the primary mutation type to be analyzed. This section reviews studies on nuclear gene-specific mutations found in somatic cases of human cancer.

2.3.2.1 Gene-specific mutations

Published somatic mutations associated with human disease have been compiled as a database, the Universal Mutation Database (UMD) (Béroud et al., 2000). This database allows creating and analyzing gene-specific mutation databases. The UMD has been adapted to 13 different nuclear genes, most of which are tumor suppressor genes.

When the UMD somatic mutations in tumor suppressor genes in cancer were analyzed by mutation type, point mutations were observed to be the major type (Béroud et al., 2000). However, differences in gene-specific point mutations have been found when the

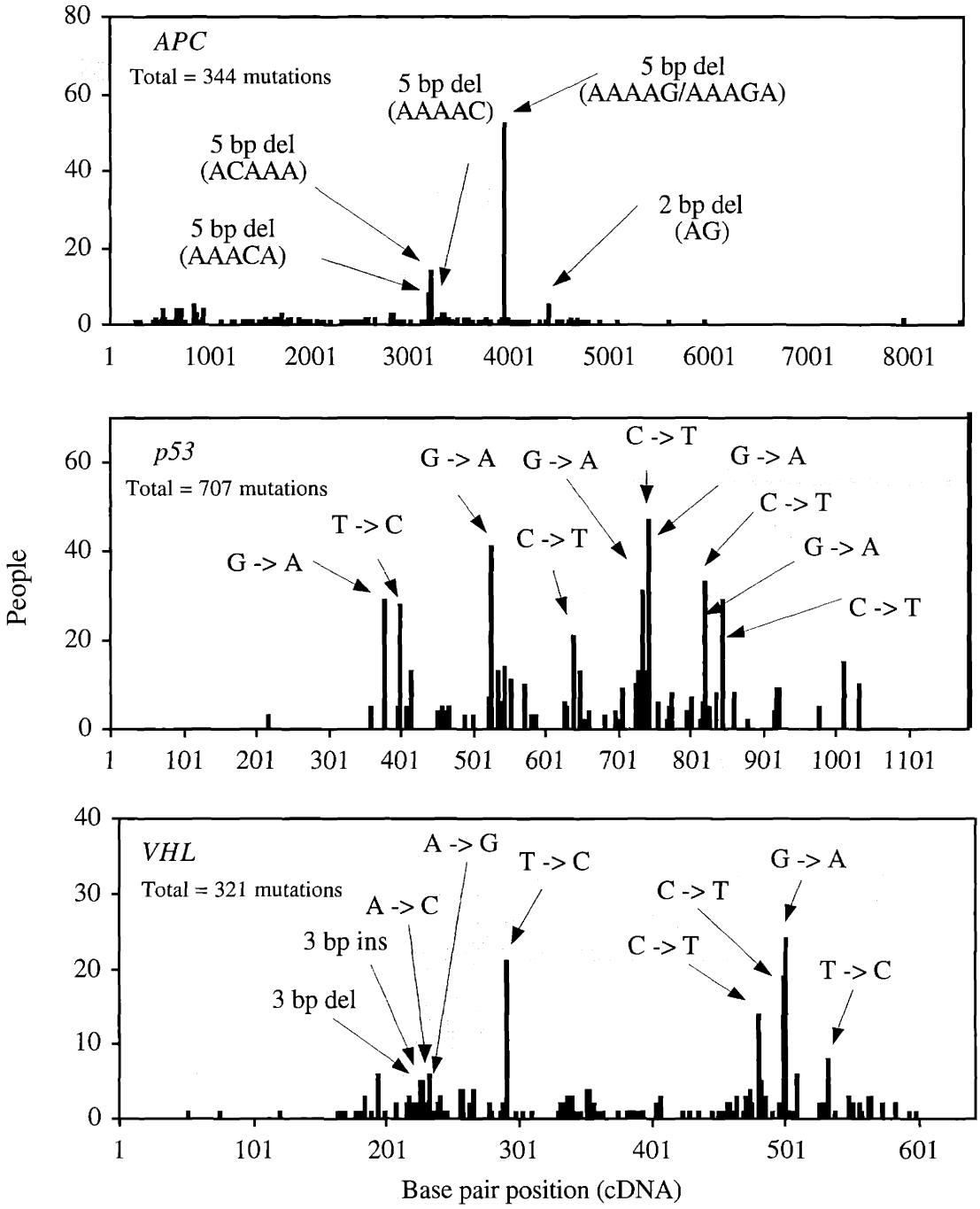


Figure 3. Distribution of germline point mutations in tumor suppressor genes in human cancer

Sources: Béroud et al., 1998; Béroud and Soussi, 1998; Muniappan, unpublished results: collection of published *APC* mutations observed when the entire coding region was analyzed.

UMD somatic entries were sorted by the following two categories: mutations leading to truncated proteins (MLTP) and missense mutations (Béroud et al., 2000) (MLTP includes nonsense mutations, as well as out of frame deletions and insertions). For example, MLTP accounts for 95% of the *APC* mutations, 71% of the *VHL* mutations, and 77% of the *WT1* mutations (Béroud et al., 1998; Jeanpierre et al., 1998; Laurent-Puig et al., 1998; Gallou et al., 1999; Béroud et al., 2000). Missense mutations account for 80% of the *p53* mutations (Béroud and Soussi, 1998; Béroud et al., 2000). These results suggest that differential gene-inactivating mechanisms play a role in tumor formation.

In addition to gene-specific mutations, differential gene-mutation distribution is also observed. For example, the majority of *ras* mutations found in human cancer are clustered in a small number of codons (Bos, 1989). On the other hand, other cancer-related gene mutations, such as *APC*, *p53* and *VHL*, are scattered throughout a larger number of codons (Béroud et al., 1998; Laurent-Puig et al., 1998; Hernandez-Boussard et al., 1999b). Figure 4 illustrates the distribution of somatic mutations in the *APC*, *p53*, and *VHL* genes in human cancer.

Analysis of somatic mutations associated with human cancer suggests point mutations as the primary mutation type to be studied when planning analysis of somatic mutations in a variety of human samples. In addition, a careful selection of target-gene sequences is suggested. For certain genes like *APC*, *p53*, and *VHL*, the entire gene-coding regions can be selected as a target since the observed mutations are scattered throughout each gene-coding region. On the other hand, analysis should be limited to certain parts of gene-coding regions for other genes like *ras* (the observed *ras* mutations are clustered in a small number of codons).

2.3.2.2 *p53* mutations

The *p53* gene has the highest number of mutations among the gene mutations in the UMD (Béroud et al., 2000). For this reason, this section reviews studies related to *p53* somatic mutations associated with human cancer. This section focuses on hypotheses made regarding causes of the *p53* mutations and discusses limitations of these studies in understanding human mutagenesis.

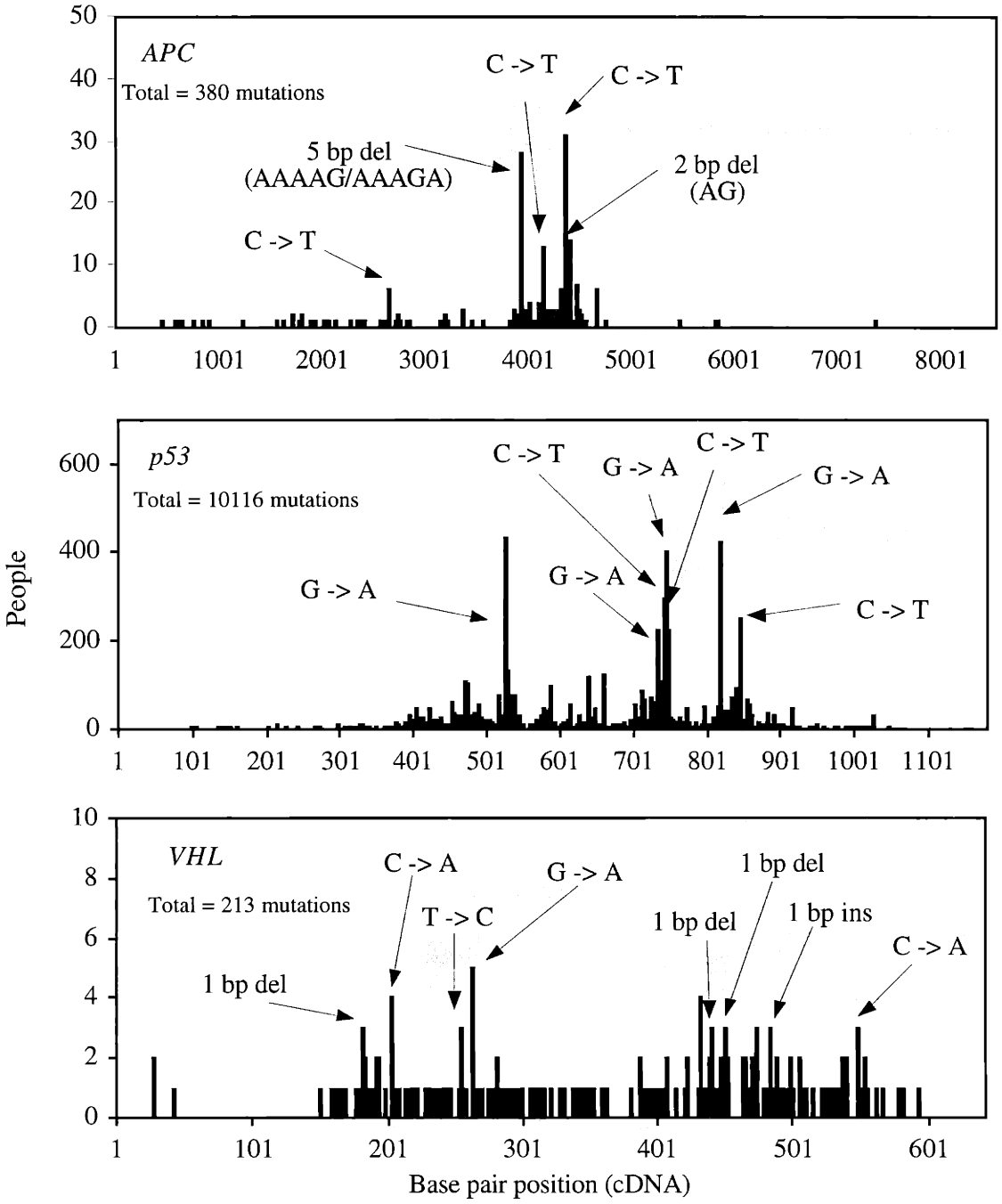


Figure 4. Distribution of somatic point mutations in tumor suppressor genes in human cancer

Sources: Béroud et al., 1998; Hernandez-Boussard et al., 1999b; Muniappan, unpublished results: collection of published *APC* mutations observed when the entire coding region was analyzed.

The human tumor suppressor gene *p53* lies on chromosome 17p13 (Levine et al., 1991; Akashi and Koeffler, 1998). This gene is 20-kb in length and consists of 11 exons, with the gene product being mainly a negative regulator of the cell cycle (Levine et al., 1991; Akashi and Koeffler, 1998). Many *p53* mutations have been found to be associated with various cancer types (Hainaut et al., 1998; Hernandez-Boussard et al., 1999b). For most cancer types, 20 to 50% of cases have been shown to carry a *p53* mutation(s) (Greenblatt et al., 1994).

One of the databases comprising published *p53* mutations associated with human cancer is maintained at the International Agency for Research on Cancer (IARC), the IARC *p53* mutation database (Hainaut et al., 1998; Hernandez-Boussard et al., 1999b). To date, more than 10,000 somatic *p53* mutations have been entered into this database. Most of these mutations are in the gene-coding region. Among IARC mutations 87.2% are single base-pair substitutions, and the remaining 12.8% are small insertions and deletions, as well as complex mutations (Hernandez-Boussard et al., 1999b). Over 90% of all IARC mutations are clustered within the central portion of the *p53* gene-coding region (Soussi et al., 1990; Hernandez-Boussard et al., 1999b). This portion is domains of highly conserved sequences through evolution and consists of the DNA binding domain of the protein, essential to *p53* functional activity (Cho et al., 1994).

Analysis of published *p53* point mutations in all human cancer types has shown that about 50% of the mutations are G to A transitions, with over 50% residing within CpG dinucleotides (Soussi et al., 2000). Spontaneous deamination of 5-mC has been hypothesized to be a cause of such transition mutations in CpG sequences (Greenblatt et al., 1994) (Section 2.3.1 discusses hypermutability of CpG dinucleotides). It has been demonstrated that all of the CpG sites in the human *p53* coding sequence investigated are methylated in all tissue types studied (Tornaletti and Pfeifer, 1995). This demonstration supports the hypermutability of the CpG-dinucleotide hypothesis.

However, analysis by cancer type of these published *p53* point mutations has suggested a link between carcinogen exposure and human cancer (Hussain and Harris, 1998; Soussi et al., 2000). In these studies, tissue-specific *p53* mutations were observed, suggesting the following three hypotheses: codon 249 mutation (AGG -> AGT) and dietary aflatoxin B_1 (AF B_1) exposure in liver cancer, G to T transversions and cigarette

smoke in lung cancer, and CC to TT dipyrimidine mutations and sunlight exposure in skin cancer.

A G to T transversion at the third position of *p53* codon 249 (AGG -> AGT) is the dominant *p53* mutation observed in human liver cancer and has been linked to dietary AFB_1 (Hussain and Harris, 1998; Hernandez-Boussard et al., 1999b; Soussi et al., 2000). This link has been supported by an observation in which a positive dose-response correlation was found between estimated dietary AFB_1 exposure and the frequency of this mutation in normal-appearing liver samples from hepatocellular carcinoma patients with different AFB_1-exposure levels (Aguilar et al., 1994). AFB_1 exposure to human cells *in vitro* has shown to induce *p53* mutations at positions other than the third of codon 249 (Aguilar et al., 1993; Mace et al., 1997). In addition, analysis of formed AFB_1 adducts in the *p53* gene has shown major adduct sites at positions other than codon 249 (Denissenko et al., 1998). These studies suggest that additional mechanisms, such as infection with hepatitis B virus, may be required for selection of the G to T transversion at the third position of *p53* codon 249 in human liver cancer (Denissenko et al., 1998).

The predominant *p53* mutations observed in human lung cancer are G to T transversions (Hernandez-Boussard et al., 1999b). This observation is compatible with the role of exogenous carcinogens present in cigarette smoke, such as BP. The three most frequently reported *p53* mutations of this kind in human lung cancer are at codons 157, 248, and 273 (Hernandez-Boussard et al., 1999b). A G to T transversion at one of these codons, codon 248 (CGG -> CTG), has been shown to be generated as a result of BP exposure to human cells *in vitro* (Cherpillod et al., 1995). In addition to cigarette-smoke exposure, formation of many BPDE-DNA adducts has been shown to be strongly enhanced by methylation of CpG sites in the *p53* gene (Denissenko et al., 1996; Denissenko et al., 1997; Chen et al., 1998). When HeLa cells and human bronchial epithelial cells were treated with BPDE, this agent selectively induced guanine adduct formation at CpG sites in *p53* codons 157, 248, and 273, in accordance with the *p53* mutational hotspot codons observed in human lung cancer (Denissenko et al., 1996). Thus, two different 5-mC pathways appear to play a role in *p53* mutagenesis associated with human cancer: increased affinity of methylated CpG sites for DNA-reactive molecules in addition to spontaneous deamination of 5-mC (Pfeifer et al., 2000).

Analysis of *p53* mutations observed in human lung cancer based on smoking status has supported the hypothesis that cigarette smoking as a direct cause of these mutations in Hernandez-Boussard and Hainaut (1998), as well as in Hainaut and Pfeifer (2001). Another study has supported cigarette smoking as a source of selecting pre-existing mutations (Rodin and Rodin, 2000).

The commonality of tandem dipyrimidine CC to TT mutations in human squamous and basal cell skin carcinoma has been observed (Brash et al., 1991; Ziegler et al., 1993). A particular *p53* mutation of this kind observed in human skin cancer has been analyzed in normal-appearing skin samples from skin cancer patients with different sunlight-exposure levels (Nakazawa et al., 1994; Ouhtit et al., 1997). In these studies, a positive dose-response correlation between the sunlight-exposure level and the frequency of this mutation was observed. In addition to the indicative role of sunlight exposure, preferential DNA-adduct formation and differential DNA repair rates appear to play a role in human skin cancer. Sites where preferential UV-adducts were formed in the *p53* gene in human cells have been shown to correlate with the *p53* mutational hotspot codons observed in human skin cancer (Tornaletti et al., 1993; Tommasi et al., 1997). The majority of these hotspots also have been shown to suffer from slow DNA repair of UV-induced damage in human cells (Tornaletti and Pfeifer, 1994). These studies suggest that preferential DNA-adduct formation and repair rate have contributed to the *p53* mutational hotspots observed in human skin cancer.

However, hypotheses made based on *p53* mutations observed in human cancer should be evaluated carefully. More than 70% of the molecular studies focus on the central region of the *p53* gene, exons 5 though 8 (Hernandez-Boussard et al., 1999b; Soussi et al., 2000). In addition, many factors may influence detection and reporting of mutations, including selection of tumor samples, study design, and choice of methods (Hernandez-Boussard et al., 1999a; Soussi et al., 2000). Inter-individual variability in susceptibility to carcinogens also suggests that a single pathway is highly unlikely (Vineis et al., 1999). For example, carcinogens form adducts with DNA, followed by DNA repair and replication, rather than directly inducing specific point mutations (Vineis et al., 1999). Moreover, mutations found in advanced tumors may not be representative of the type of damage created by an agent in the DNA of the original target cell (Vineis et al., 1999). In

addition, these mutations may not play any role in developing tumors (Strauss, 2000). Thus, mutational analysis in normal tissues must be performed to understand cause-and-effect relationships between environmental agents and human mutation. For such analysis, genotype-based methods of measuring mutations in normal tissues are necessary. Such methods should permit measurement of rare mutations in the presence of their abundant wild-type and should be general to the human genome.

2.4 Somatic nuclear mutations and mutant fractions in normal human tissues

Studies on mutations and mutant fractions observed in normal human tissues can be used as a guide when planning further mutational analysis of normal tissues. This section reviews studies on somatic nuclear mutations, as well as mutant fractions, observed in normal human tissues. The review focuses on point mutations since these mutations are the major mutation type observed in both inherited and somatic cases of human disease.

2.4.1 Phenotype-based analysis of HLA and HPRT mutations

For selectable genes, such as *HLA* and *HPRT*, mutation assays based on phenotypic selection have been developed for analysis of human blood (Albertini et al., 1982; Morley et al., 1983; Albertini et al., 1985; Janatipour et al., 1988). Using such assays, a mutant fraction of about 10^{-5} has been measured in both the *HLA* and *HPRT* genes in T-lymphocytes of middle-aged healthy individuals (Grist et al., 1992; Robinson et al., 1994; Akiyama et al., 1995; Podlutsky et al., 1998). These mutant fractions have been shown to increase linearly with age (Trainor et al., 1984; Vijayalaxmi and Evans, 1984; Davis et al., 1992; Grist et al., 1992; Akiyama et al., 1995; Green et al., 1995). Figure 5 illustrates this increase in the *HPRT* gene.

Assuming a target size of about 1000-bp for both the *HLA* and *HPRT* genes, the expected average mutant fraction per bp in T-lymphocytes of middle-aged healthy individuals is about 10^{-8}. Any hotspot mutations occurring 10 to 100 times more frequently than expected by chance would then appear in these genes at fractions between 10^{-7} and 10^{-6}. Such hotspot mutations in the *HPRT* gene in cultured human cells have

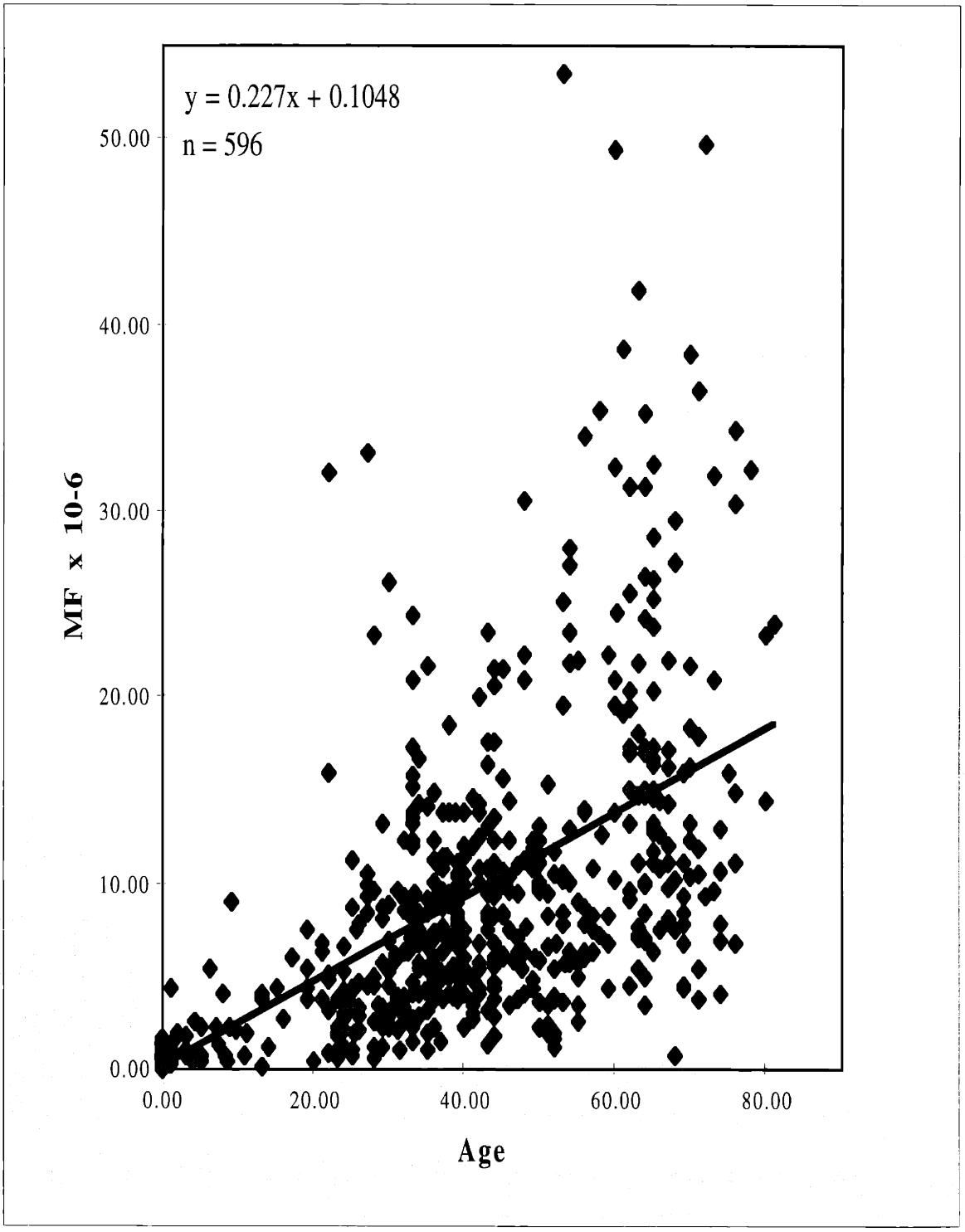


Figure 5. *HPRT* mutant fraction vs. age in T-lymphocytes of healthy individuals

Collection of published *HPRT* mutant fractions (Tomita-Mitchell, unpublished results); based on 6TG selection and clone-by-clone analysis.

been shown to represent between 10 and 50% of the point mutations that affect the physiological function of the gene product (Kat, 1992; Tomita-Mitchell, 1999; Tomita-Mitchell et al., 2000).

Clone-by-clone analysis of 6TG-resistant *HPRT* mutants in T-lymphocytes of healthy individuals has shown that, at birth, the predominant mutation type is large structural alterations (85%), mostly deletions (McGinniss et al., 1989). These deletion mutations are thought to be mediated by V(D)J recombinase acting on sequences within the *HPRT* gene that resemble the V(D)J recombinase signal sequences (Fuscoe et al., 1991; Finette et al., 1996; Fuscoe et al., 1997). As people age, the predominant mutation type changes to point mutations (85%), with the remaining 15% being large structural alterations (Albertini et al., 1990; Lippert et al., 1990; Albertini et al., 1993; Cariello and Skopek, 1993a; Albertini and Hayes, 1997). These observations indicate that factors other than V(D)J recombinase play a role in causing mutations in people as they age.

Published 6TG-resistant *HPRT* mutations observed in T-lymphocytes of healthy individuals have been compiled as a database (Cariello et al., 1998). Table 5 summarizes these mutations by mutation type and corresponding number. Figure 6 illustrates the distribution of these mutations. Single base-pair substitutions account for 76% of all mutations, with the remaining 24% being deletions and insertions. Among single base-pair substitutions, transversions (41%) are slightly more common than transitions (36%). In addition, mutations at GC base pairs (48%) are more common than mutations at AT base pairs (29%). G to A transitions are the most common type of mutation, accounting for 27% of all mutations.

The *HPRT* database does not seem to pinpoint any particular kind of exogenous agent as the cause for these mutations. For example, when kinds and positions of *HPRT* mutations observed in smokers were compared to those of nonsmokers, no significant differences were observed between the two sub-populations (Vrieling et al., 1992; Burkhart-Schultz et al., 1996; Curry et al., 1999; Podlutsky et al., 1999). Thus, these studies imply that *HPRT* mutations in healthy populations are endogenous in origin or induced by ubiquitous environmental exposures (Podlutsky et al., 1998; Podlutsky et al., 1999).

These mutational studies based on phenotypic selection performed on the blood of

Table 5. Summary of somatic mutations in *HPRT* gene in T-lymphocytes of healthy individuals

Mutation type	Number of entries (% total)
G -> A	101 (27%)
A -> G	34 (9%)
G -> C	44 (12%)
G -> T	34 (9%)
A -> C	36 (10%)
A -> T	36 (10%)
Insertion/deletion, 1-bp	53 (14%)
Insertion/deletion, >1-bp	36 (10%)
Total	374 (100%)

Source: Cariello et al., 1998; courtesy of Dr. A. Tomita-Mitchell; based on 6TG selection and clone-by-clone analysis.

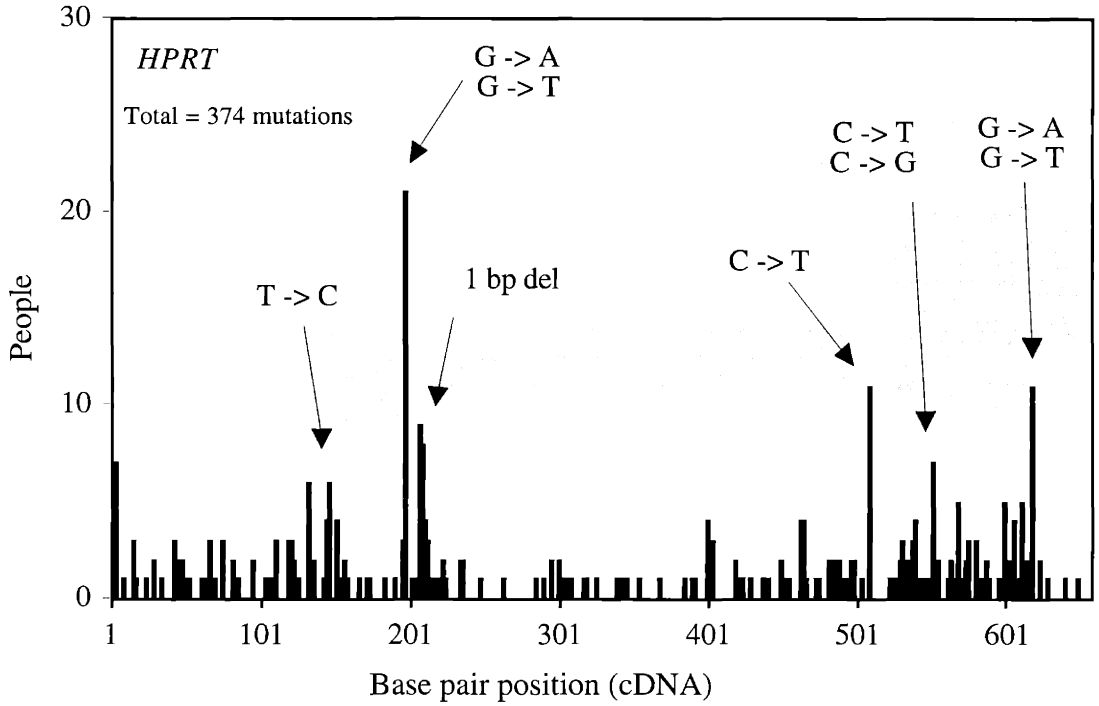


Figure 6. Distribution of somatic mutations in *HPRT* gene in T-lymphocytes of healthy individuals

Source: Cariello et al., 1998; courtesy of Dr. A. Tomita-Mitchell; based on 6TG selection and clone-by-clone analysis.

healthy individuals suggest the spontaneous origin of human *HPRT* mutation. While mutational studies of other nuclear genes in normal human tissues remain to be completed, such studies are hindered by the lack of general and sensitive methods of measuring mutations based on their genotype. Mutation detection methods based on phenotypic selection are limited to selectable genes, such as *HPRT*, and certain tissue types, such as blood. For these reasons, genotype-based methods are the methods of choice.

2.4.2 Phenotype-based analysis of p53 mutations

Mutant forms of the *p53* protein have been detected in skin tissues of healthy individuals (Pontén et al., 1995; Jonason et al., 1996; Tabata et al., 1999). By using antibodies specific to *p53*, the mutant forms can be stained as individual patches (Gannon et al., 1990; Iggo et al., 1990; Baas et al., 1994). In this assay, *p53* mutants are distinguished from the wild-type by their longer half life (Finlay et al., 1988; Harris and Hollstein, 1993). Each compact pattern of the stained patches, a contiguous area of homogeneously stained cells, represents clones derived from a single cell with a *p53* mutation (Jonason et al., 1996; Ren et al., 1996; Ren et al., 1997). Between 50 and 70% of the patches have been shown to contain such clones (Jonason et al., 1996; Pontén et al., 1997; Ren et al., 1997).

Using such a phenotype-based assay, *p53* mutant fractions in skin tissues of healthy individuals can be estimated. Table 6 summarizes the estimated mutant fractions. On average, 33 and 3 *p53*-immunopositive patches per cm^2 of human skin have been observed in sun-exposed and sun-shielded areas, respectively (Jonason et al., 1996). By dividing these observed values by the stem cell number per cm^2 of skin, the *p53* stem cell mutant fractions can be estimated. About 4.5×10^6 keratinocytes have been estimated per cm^2 of normal human skin (Pinkus, 1952; Bergstresser et al., 1978; Weinstein et al., 1984). And among these keratinocytes, between 2 and 10% have been found to be stem cells (Pinkus, 1952; Potten and Hendry, 1973; Potten, 1981; Potten and Morris, 1988). Using the stem cell number of 2.7×10^5 per cm^2 of human skin (median value: $(4.5 \times 10^6) \times 0.06$), the *p53* stem cell mutant fractions can be estimated to be about 1.2×10^{-4} ($33 \div$

Table 6. Summary of *p53* mutant fractions in skin tissues of healthy individuals

Level of sun exposure	Stem cell mutant fraction[a]	Total mutant fraction[b]	Average mutant fraction/bp[c]
Low/sun-shielded	1.1×10^{-5}	10^{-4} to 10^{-3}	10^{-7} to 10^{-6}
High/sun-exposed	1.2×10^{-4}	10^{-3} to 4×10^{-2}	10^{-6} to 4×10^{-5}

[a] On average, 33 and 3 *p53*-immunopositive patches per cm^2 of human skin have been observed in sun-exposed and sun-shielded areas, respectively (Jonason et al., 1996). The stem cell mutant fractions were estimated by dividing these observed values by the stem cell number per cm^2 of human skin, 2.7×10^5.

[b] Jonason et al., 1996.

[c] Assuming a target size of about 1000-bp for the *p53* gene, the average mutant fractions per bp were estimated by dividing the total mutant fractions by 1000.

(2.7×10^5)) and 1.1×10^{-5} $(3 \div (2.7 \times 10^5))$ for sun-exposed and sun-shielded areas, respectively. In this estimation, all immunopositive patches are assumed to be clones, each representing an individual mutant.

The total *p53* mutant fractions are expected to be higher than the stem cell mutant fractions because of the clonally-derived nature of the immuopositive *p53* mutant patches (Jonason et al., 1996; Ren et al., 1996; Ren et al., 1997). The total mutant fractions in skin tissues of healthy individuals have been estimated to be 10^{-3} to 4×10^{-2} and 10^{-4} to 10^{-3} in sun-exposed and sun-shielded areas, respectively (Jonason et al., 1996) (see Table 6). In this estimation, the total mutant fraction was estimated by averaging the measured area of clones and then multiplying by the number of clones per cm^2.

Assuming a target size of about 1000-bp for the *p53* gene, the average mutant fractions per bp in skin tissues of healthy individuals are expected to be 10^{-6} to 4×10^{-5} in sun-exposed areas and 10^{-7} to 10^{-6} in sun-shielded areas (see Table 6). Any hotspot mutations occurring 10 times more frequently than expected by chance would then appear in this gene at fractions 10 times higher than the expected average mutant fractions for both sun-shielded and sun-exposed areas.

When the average *p53* mutant fractions per bp were compared, up to a 400-fold higher mutant fraction has been observed in sun-exposed relative to sun-shielded skin (Jonason et al., 1996). On the other hand, another study showed no such difference (Ouhtit et al., 1998). This discrepancy is perhaps because one study was based on a particular *p53* mutant (Ouhtit et al., 1998), while the other was based on mutations in the entire *p53* gene (Jonason et al., 1996). Further mutational analysis in normal human skin is necessary to validate either study.

2.4.3 Genotype-based analysis of ras and p53 mutations

Several nuclear point mutations have been observed in tissues of healthy individuals without reference to phenotypic selection (Wilson et al., 1999; Ouhtit et al., 1999; Wilson et al., 2000; Li-Sucholeiki et al., unpublished results). Either allele-specific PCR or restriction fragment length polymorphism (RFLP)/PCR (see Section 2.5.3.2) was used as the genotype-based means for such observation. All of these mutations had been

previously observed in human tumors (Bos, 1989; Hainaut et al., 1998; Hernandez-Boussard et al., 1999b).

Table 7 summarizes the position, kind, mutant fraction, and observed number of cases for each mutation detected in normal human tissues. Among these mutations, about a 50-fold higher average mutant fraction was found in lungs compared to that in either blood or skin (5×10^{-5} vs. 10^{-6}). However, this comparison is based on mutations in different genes, as well as mutations in the same gene but at different base-pair positions. When a particular *p53* mutation in normal human skin tissues of different sun-exposure levels was compared, no statistically significant relationship was found between levels of sun exposure and mutant fractions (Ouhtit et al., 1998). In addition, no significant differences were observed in the level of this mutation among people of different ages or genders (Ouhtit et al., 1998).

A similar level of mutant fractions has also been observed in normal-appearing smokers' lungs compared to those of non-smokers' (Li-Sucholeiki et al., unpublished results). In this study, two G to T transversions in the *p53* gene were detected in two smokers' and two nonsmokers' lungs, with an average mutant fraction of 5×10^{-5} for each, in smokers' and nonsmokers' lungs. These results indicate that cigarette smoking does not affect the kind or number of nuclear point mutations in human bronchial epithelial cells. These results confirm a previous observation made in human mitochondrial DNA (Coller et al., 1998).

These genotype-based mutational studies performed on normal human tissues do not pinpoint any particular exogenous agent as a cause for these mutations. However, these studies are few. Therefore, additional studies analyzing more genes, tissue types, and individuals are necessary to validate any observations made in these previous studies.

2.5 Criteria for measuring point mutational spectra in normal human tissues

Analysis of somatic point mutational spectra in normal human tissues allows direct testing of the hypothesis that environmental agents are the primary cause of human mutation. Comparison of mutational spectra in smokers' and non-smokers' lungs reveals effects of cigarette smoke on human mutation. The same type of comparison can be

Table 7. Summary of nuclear point mutations in normal human tissues*

Tissue type	Gene	Position and kind	Mutant fraction (mf)	Number of cases
Blood[a&b]	H-*ras*	Codon 12.2, G -> T	$\geq 10^{-6}$	3
	N-*ras*	Codon 12.1, G -> T	$\geq 10^{-6}$	3
		Codon 12.2, G -> T	$\geq 10^{-6}$	6
		Codon 13.1, G -> A	$\geq 10^{-6}$	1
		Codon 248.1, C -> A	$\geq 10^{-6}$	1
	p53	Codon 248.1 (cDNA bp 742), C -> T	$\geq 10^{-6}$	6
Lung[c]	*p53*	Codon 249.2 (cDNA bp 746) G -> T	$\approx 5 \times 10^{-5}$ [d]	≈200[e]
	p53	Codon 249.3 (cDNA bp 747) G -> T	$\approx 5 \times 10^{-5}$ [d]	≈200[e]
Skin[f]	*p53*	Codon 247.3&248.1 (cDNA bp 741&742), CC -> TT	$< 0.5 \times 10^{-6}$	58
			$0.5 \leq mf \leq 0.99 \times 10^{-6}$	8
			$1.0 \leq mf \leq 1.49 \times 10^{-6}$	4
			$1.5 \leq mf \leq 2.49 \times 10^{-6}$	2
			$\geq 2.5 \times 10^{-6}$	3

* Genotype-based mutational analysis.

[a] Wilson et al., 1999.

[b] Wilson et al., 2000.

[c] Li-Sucholeiki et al., unpublished results: tracheal bronchial epithelial cells of deceased middle-aged individuals, considered as healthy individuals, were analyzed. These individuals died of either stroke or subarachnoid hemorrhage cerebrovascular disease.

[d] Average mutant fraction.

[e] Two smokers' and two nonmokers' lungs were microdissected into about 200 sectors.

[f] Ouhtit et al., 1998.

performed in sun-exposed and un-exposed human skin. However, detection of mutations in normal tissues with statistical significance requires a careful planning since these mutations are present as a much smaller copy number compared to the wild-type copy number (mutations in normal tissues are rare mutations). This section reviews studies on criteria for measuring rare mutations with statistical significance. This review includes stem cell turnover unit size, sample size, and method of analysis.

2.5.1 Turnover unit size

A stem cell gives rise to descendent transition and terminal cells (Potten, 1981; Potten and Morris, 1988; Potten and Loeffler, 1990), generating a stem cell compartment, or a turnover unit. Thus, a mutation in a stem cell leads to a clonally-derived mutant turnover unit. Existence of such mutant turnover units in normal human tissues has been suggested and demonstrated (Jonason et al., 1996; Khrapko et al., 1997a; Coller et al., 1998; Li-Sucholeiki et al., unpublished results).

DNA sequencing analysis of immunopositive *p53* mutant patches has shown that these patches are clonally-derived, each patch representing a particular *p53* mutation, in skin tissues of healthy individuals (Pontén et al., 1995; Jonason et al., 1996; Tabata et al., 1999). The turnover unit size for these *p53* mutations was observed to be between 60 and 3000 cells, with different average unit sizes in tissues of different sun exposure levels: 0.014 mm^2 in sun-shielded tissues and 0.040 mm^2 in sun-exposed tissues (Jonason et al., 1996). These unit sizes are equivalent to 630 cells (0.014 mm^2 x (4.5 x $10^6/cm^2$)) and 1800 cells (0.040 mm^2 x (4.5 x $10^6/cm^2$)) using the cell number of 4.5 x 10^6 per cm^2 of normal human epidermis (Pinkus, 1952; Bergstresser et al., 1978; Weinstein et al., 1984).

Using allele-specific PCR, particular *p53* mutations have been detected as mutant-rich clusters in normal-appearing human bronchial epithelial cells (Li-Sucholeiki et al., unpublished results). These clusters are thought to be clonally derived. In this study, no difference in mutant turnover unit size was observed, determined to be about 32 cells in both smokers' and non-smokers' lungs. These results contradict the study of Jonason et al. (1996) in which, on average, about a 3-fold higher mutant turnover unit size was observed in mutagen-exposed (sun-exposed) relative to un-exposed (sun-shielded) areas

of normal human skin. This study suggests sunlight as the cause for this increase. This contradiction is probably because two different tissue types, lung vs. skin, and two different mutagens, cigarette vs. sunlight, were compared.

While turnover unit size has not been determined for all tissue types, this factor should guide planning somatic mutational analysis in normal human tissues. For example, in a sample containing 10^8 cells with a turnover unit size of 32 cells, the expected number of turnover units is about 3×10^6 ($10^8 \div 32$). Using the observed *HPRT* and *HLA* mutant fractions, 10^{-5} (see section 2.4.1), 30 stem cell mutations ($10^{-5} \times (3 \times 10^6) = 30$) would appear in each gene. Thus, the expected number of mutations per 100-bp sequence is 3, assuming a target size of 1000-bp for these genes. This expectation predicts that to observe 30 mutations, either a 100-bp sequence in 10 tissue samples, each containing 10^8 cells, or 10 sequences, each containing 100-bp in the same sample of 10^8 cells, must be analyzed.

2.5.2 Sample size

When planning detection of somatic mutations in normal tissues, sample size must be examined. Sample size can be limited by choice of mutation detection method, thereby limiting the copy number of any particular mutant to be analyzed. For example, for a sample size of 10^5 diploid human cells containing a mutation at a fraction of 5×10^{-6} in a single copy sequence, one mutant copy ($10^5 \times 2 \times (5 \times 10^{-6})$) would be available for analysis. Based on the Poisson distribution, the probability of not detecting this mutation is 0.37 (e^{-1}). Thus, this sample would generate false negatives 37% of the time, confounding mutational analysis.

Mutation detection with sufficient precision, such as ±20%, requires a statistically significant copy number of any particular mutant, at least 100 copies (Leong et al., 1985; Keohavong and Thilly, 1992a; Chen and Thilly, 1994). Thus, for example, a sample size of at least 10^7 cells (($\geq 10^7$) x 2 x (5×10^{-6}) = ($\geq$100)) is required for a mutation occurring at a fraction of 5×10^{-6} in a single copy sequence. To sum up, reproducible outcomes have been demonstrated in which at least 100 copies of any particular mutant were

detected in various human cell and tissue samples (Keohavong and Thilly, 1992b; Khrapko et al., 1997a; Coller et al., 1998; Li-Sucholeiki and Thilly, 2000).

2.5.3 Methods

Detection of mutations present as abundant copies, representing mutations in diseased tissues, can be as simple as DNA sequencing (Cooper and Krawczak, 1993). However, somatic mutations in normal tissues are rare mutations, and analysis of these mutations requires more sophisticated methods. This section reviews available methods of analyzing rare mutations in human genomic DNA.

2.5.3.1 Hifi-PCR

Introduction of thermostable DNA polymerases to *in vitro* DNA amplification, PCR, has simplified the amplification by allowing it to be automated (Saiki et al., 1988). However, careful selection of a polymerase is necessary for analysis of amplified products since different polymerases generate mutations in a chosen target at various rates during PCR (Keohavong and Thilly, 1989; Cariello et al., 1991; Mattila et al., 1991; Ling et al., 1991; Flaman et al., 1994; Cline et al., 1996). Table 8 summarizes the fidelities of, as well as predominant mutations generated by, three thermostable DNA polymerases commercially available, *Thermus aquaticus* (*Taq*), *Thermococcus litoralis* (*Vent*), and *Pyrococcus furiosus* (*Pfu*).

While for certain applications of PCR products, such as DNA sequencing, the levels and kinds of mutations generated by a polymerase are not a concern, application to analysis of rare mutations requires a polymerase with high fidelity. To date, *Pfu* offers the highest fidelity among the thermostable DNA polymerases available commercially (Flaman et al., 1994; Cline et al., 1996; André et al., 1997). Thus, it is the enzyme of choice for rare event analysis. The fidelity of *Pfu* has been determined to be about 10^{-6}/bp/doubling (see Table 8), a factor which can guide planning of mutational analysis.

A typical template target copy number for PCR is about 10^5 to 10^6 (≈1 to 10 μg) and a chosen target can be amplified up to about 10^{12} copies (10^6- to 10^7-fold amplification or

Table 8. Characteristics of thermostable DNA polymerases

DNA polymerase	Fidelity/bp/doubling*	Predominant mutation type**
Taq	$0.72 - 2.4 \times 10^{-4}$ (8/13)[a]	A -> G (84/157)[d]
Vent	$2.4 - 4.5 \times 10^{-5}$ (5/7)[b]	A -> G (16/18)[e]
Pfu	$0.65 - 2 \times 10^{-6}$ (8/10)[c]	G -> T (15/19)[f]

* Reported fidelities were gathered and each was counted as one. The range indicated is based on the reported fidelities in close agreement. For example, 8 out of 13 were in the range indicated for *Taq*.

** Reported mutations were gathered and each was counted as one. For example, 84 out of 157 were A -> G transitions for *Taq*.

[a] Saiki et al., 1988; Tindall and Kunkel, 1988; Keohavong and Thilly, 1989; Eckert and Kunkel, 1990; Cariello et al., 1991; Chen et al., 1991; Ling et al., 1991; Lundberg et al., 1991; Keohavong et al., 1993; Flaman et al., 1994; Cline et al., 1996; Huang and Keohavong, 1996; Smith and Modrich, 1996.

[b] Cariello et al., 1991; Ling et al., 1991; Mattila et al., 1991; Keohavong et al., 1993; Cline et al., 1996; Smith and Modrich, 1996.

[c] Lundberg et al., 1991; Brail et al., 1993; Cariello and Skopek, 1993b; Barnes, 1994; Flaman et al., 1994; Cline et al., 1996; Smith and Modrich, 1996; André et al., 1997; Parsons and Heflich, 1998; Li-Sucholeiki and Thilly, 2000.

[d] Dunning et al., 1988; Saiki et al., 1988; Keohavong and Thilly, 1989; Ennis et al., 1990; Chen et al., 1991.

[e] Keohavong et al., 1993.

[f] André et al., 1997; Li-Sucholeiki and Thilly, 2000.

20 to 23 doublings) (Saiki et al., 1988; Cha and Thilly, 1993). For this reason, *Pfu*-generated mutations in a chosen target after 20 doublings are discussed in this section. After 20 doublings are performed with a sample using *Pfu*, any hotspot mutations of *Pfu* origin occurring 10 times more frequently than expected by chance would appear in the amplified target at a fraction of 2×10^{-4} (10^{-6}/bp/doubling x 20 doublings x 10). If, for example, the initial sample contained mutations at a fraction of about 2×10^{-4}, the level of these mutations would be indistinguishable from the *Pfu*-hotspot mutations, which would interfere with the sample analysis. Especially if this sample contained mostly G to T transversions, the analysis would be even more confounded since the predominant mutation generated by *Pfu* is a G to T transversion (see Table 8). Thus, either mutations at initial fractions of above 2×10^{-4} should be analyzed or pre-PCR mutant enrichment must be performed when analyzing mutations at initial fractions at or below 2×10^{-4}.

2.5.3.2 Genotype-based point mutation detection

This section reviews studies on genotype-based methods of analyzing rare point mutations, representing somatic point mutations in normal tissues. Unlike phenotype-based methods, which can be applied to selectable genes and certain tissue types, genotype-based methods offer analysis of any gene and tissue type of interest. However, detection of rare mutations requires enrichment of mutant sequences relative to the abundant wild-type sequence. Mutant enrichment can be achieved by selectively amplifying a particular mutant sequence (allele-specific PCR), selectively destroying the wild-type sequence (RFLP/PCR), or spatially separating mutant from wild-type sequences (target isolation/CDCE/hifi-PCR) (Parsons and Heflich, 1997b). Table 9 summarizes sensitivities, target sizes, and analysis limitations of these genotype-based methods of analyzing point mutations at fractions at or below 5×10^{-5} in human genomic DNA. All of these assays use an internal standard to determine the level of mutations in initial samples.

Table 9. Characteristics of genotype-based methods of rare point mutational analysis

Method*	Sensitivity**	Target size (bp)	Analysis limitation
Allele-specific PCR[a]	10^{-6}	1	Small target size
RFLP/PCR[b]	4×10^{-7}	≈ 4-6	Mutations in restriction recognition sequences
Target isolation /CDCE/hifi-PCR[c]	10^{-6}	≈ 100	Mutations in sequences with neighboring natural clamp

* Two point mutation detection methods, MutEx/ACB-PCR (Parsons and Heflich, 1998) and MutEx/PCR/SNuPE (Parsons and Heflich, 1997b), with sensitivities of 10^{-7} and 2×10^{-5}/bp, respectively, are not included in this table since these sensitivities have not been demonstrated in human genomic DNA.

** Sensitivities in human genomic DNA are based on the mutant copy number detected, x, in the presence of the wild-type copy number, y: $x \div y$ ([a]Chen and Zarbl, 1997; [b]Chiocca et al., 1992; [c]Li-Sucholeiki and Thilly, 2000).

2.5.3.2.1 Allele-specific PCR

Detection of a point mutant by allele-specific PCR is based on preferential amplification of a known mutant sequence using mismatched primers. Primers that contain one mismatch to a mutant at the 3'end but two mismatches to the wild-type can preferentially amplify the mutant relative to the wild-type in a PCR reaction mixture (Cha et al., 1992). The amplified mutant in a polyacrylamide gel can then be quantified by comparing its signal intensity to the control PCR samples in which initial mutant fractions are varied. Recently, such analysis has been performed by capillary electrophoresis coupled with laser-induced fluorescence detection, and the initial mutant copy number was measured using an internal standard (Li-Sucholeiki et al., unpublished results).

The sensitivity of allele-specific PCR, 10^{-6}, has been demonstrated in human genomic DNA (≈10 μg) in which 5 copies of a mutant were detected in the presence of 3×10^6 copies of the wild-type (Chen and Zarbl, 1997). This method has been applied to analysis of nuclear point mutations in normal human skin and lung tissues (Nakazawa et al., 1994; Ouhtit et al., 1997; Ouhtit et al., 1998; Li-Sucholeiki et al., unpublished results).

The advantages of allele-specific PCR over RFLP/PCR and target isolation/CDCE/hifi-PCR are its rapidity and simplicity. However, this method offers the smallest target size, 1-bp, and detection of a single mutant at a time. Thus, analysis of various point mutations in the same target of 1-bp requires multiple assays to be optimized and performed, thereby eliminating its advantages especially when a larger target, such as 100-bp, needs to be analyzed.

2.5.3.2.2 RFLP/PCR

RFLP, when coupled with PCR, allows detection of point mutations in restriction recognition sequences (Parry et al., 1990; Zijlstra et al., 1990; Felley-Bosco et al., 1991). Genomic DNA is restriction digested during which mutant sequences in the recognition site of the chosen restriction endonuclease are not cleaved, while the wild-type is cleaved,

enriching for mutants. The uncleaved sequences are then PCR amplified. To generate the majority of the amplified products belonging to the mutants, as opposed to the un-cleaved wild-type, restriction digestion and PCR are repeated 2 to 3 times prior to analysis. The amplified mutants are identified by DNA sequencing or are quantified by oligonucleotide plaque hybridization using sequence-specific probes.

The sensitivity of RFLP/PCR, 4 x 10^{-7}, has been demonstrated in human cells in which 4 copies of each mutant were detected in the presence of 10^7 copies of the wild-type (Chiocca et al., 1992). This method has been applied to analysis of nuclear point mutations in mutagen-treated human cells (Chiocca et al., 1992; Palombo et al., 1992; Pourzand and Cerutti, 1993a, b; Hussain et al., 1994) and normal human tissues (Aguilar et al., 1994; Hussain et al., 1994). The sensitivity of the combined method of RFLP/PCR and ligation chain reaction (LCR) has also been demonstrated in human cells in which 1 copy of each mutant was detected in the presence of 10^7 copies of the wild-type (10^{-7}) (Wilson et al., 1999). RFLP/PCR/LCR has been applied to point mutational analysis in peripheral blood of healthy individuals (Wilson et al., 1999; Wilson et al., 2000).

To date, RFLP/PCR and RFLP/PCR/LCR offer the highest sensitivity among the point mutation detection methods available for analysis of human cells and tissues without phenotypic selection. However, these methods not only have a small target size of about 4 to 6-bp but are prone to generating polymerase-created false positives. Assuming that the efficiency of the first restriction digestion is 100%, the target copy numbers in the initial sample, 1 and 10^7 of each mutant and of the wild-type (initial mutant fraction = 10^{-7}), respectively, change to 1 and 10^5 (enriched mutant fraction = 10^{-5}) after digestion. If, for example, a 10^6-fold amplification (20 doublings) is performed with this sample using *Pfu* DNA polymerase, any *Pfu*-hotspot mutations occurring 10 times more frequently than expected by chance would appear in the amplified target at a fraction of about 2 x 10^{-4} (see Section 2.5.3.1). This mutant fraction is about 10 times higher than the level of mutations in the restriction digested target (2 x 10^{-4} vs. 10^{-5}), which would interfere with the sample analysis. Since such *Pfu*-generated hotspot mutations would not appear in every restriction recognition sequence of about 4- to 6-bp, the sensitivity of RFLP/PCR or RFLP/PCR/LCR would vary for different restriction recognition sequences. Indeed, a 1000-fold difference in sensitivity has been reported using two

different sequences (Felley-Bosco et al., 1991). Thus, the demonstrated sensitivity of about 10^{-7} (Chiocca et al., 1992; Wildson et al., 1999) is valid for certain, but not all, sequences.

2.5.3.2.3 Target isolation/CDCE/hifi-PCR

CDCE is a DNA separation technique by which point mutations in a chosen target can be differentiated from the wild-type based on differences in their melting temperatures. CDCE is in a capillary gel format (Khrapko et al., 1994a), derived from slab gel formats using the same separation principle: DGGE (Fischer and Lerman, 1983) and CDGE (Hovig et al., 1991). The advantages of CDCE over DGGE or CDGE are rapidity (min vs. hr, Khrapko et al., 1994a, b) and higher sensitivity (3×10^{-4} (Khrapko et al., 1994a) vs. 1 to 10×10^{-3} (Cariello et al., 1990; Keohavong et al., 1992b) without prior mutant enrichment). In addition, CDCE offers DNA separation with high resolution (Khrapko et al., 1994b). Especially for the purpose of analyzing rare mutations, CDCE is the method of choice since it avoids interfering factors associated with the slab gel techniques, such as heat-induced DNA damage due to extensive incubation at high temperature and DNA radiolysis caused by radioactive labels used for detection (Hanekamp, 1993).

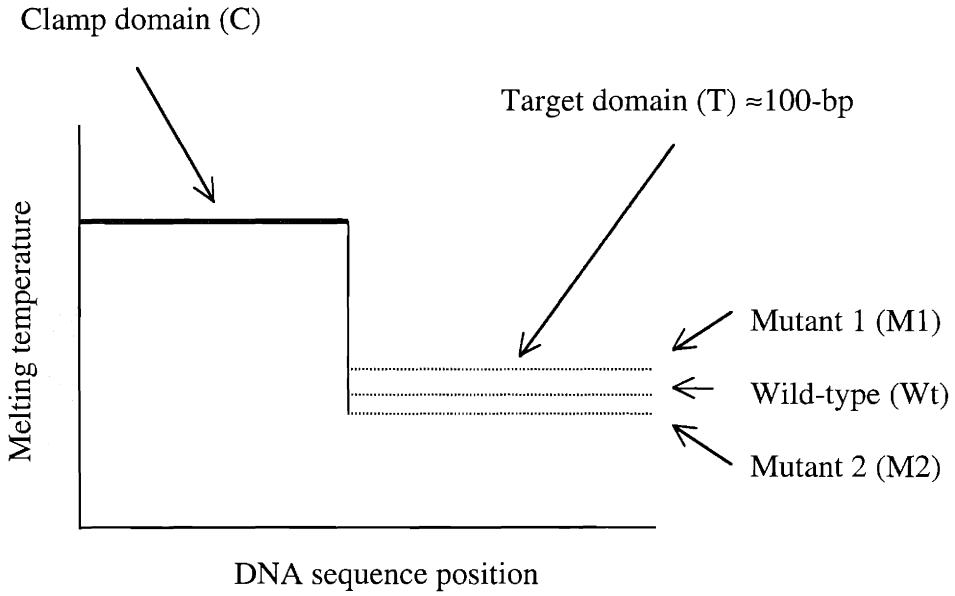
Figure 7 depicts the CDCE separation principle with a hypothetical melting profile of a DNA fragment suitable for CDCE separation. A melting profile is defined as a plot which shows the calculated temperature at which each base pair has an equal chance of being in a helix (un-melted) or random (melted) form (Fischer and Lerman, 1983).

CDCE, when combined with target isolation and hifi-PCR, allows detection of point mutations at fractions as low as 10^{-6} (Li-Sucholeiki and Thilly, 2000). Figure 8 illustrates a flow diagram of this combined method. This method has been applied to analysis of point mutations in a 121-bp nuclear single copy sequence in mutagen-treated human cells (Li-Sucholeiki and Thilly, 2000). In this application, a sensitivity of 10^{-6} was demonstrated in which at least 10^2 copies of each mutant were detected in the presence of 10^8 copies of the wild-type. This sensitivity is limited by the fidelity of *Pfu* DNA polymerase used for PCR.

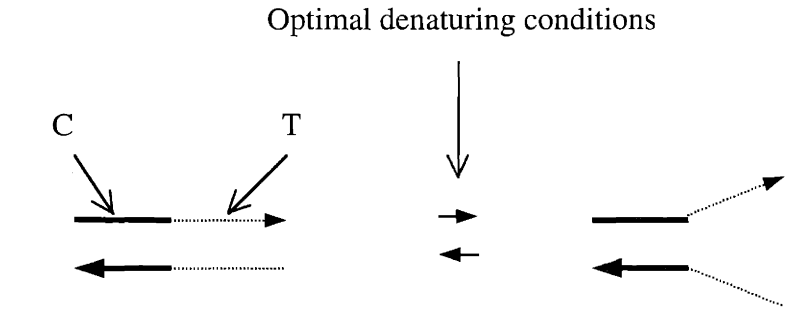
Figure 7. DNA separation by CDCE

A. A fragment suitable for CDCE separation contains a low-melting, or target, domain of about 100-bp juxtaposed with a domain of a higher melting temperature, or a clamp.
B. Since the melting temperature of the clamp is higher than that of the target, a partially-melted intermediate can be formed under a certain range of denaturant concentrations or temperatures. Introducing point mutations to the target generates differential equilibria between partially-melted and un-melted forms, resulting in different electrophoretic mobilities from that of the wild-type. Thus, these point mutations can be CDCE separated from the wild-type under optimal denaturing conditions.

A. Melting profile of DNA fragment suitable for CDCE separation



*B. Separation principle**



Un-melted DNA form, [A] Partially-melted DNA form, [B]

$$[B] / [A] = K_{eq}$$

$$K_{eq} \text{ of M1} \neq K_{eq} \text{ of Wt} \neq K_{eq} \text{ of M2} *$$

↓

Differential electrophoretic mobilities

*Source: Fischer and Lerman (1983).

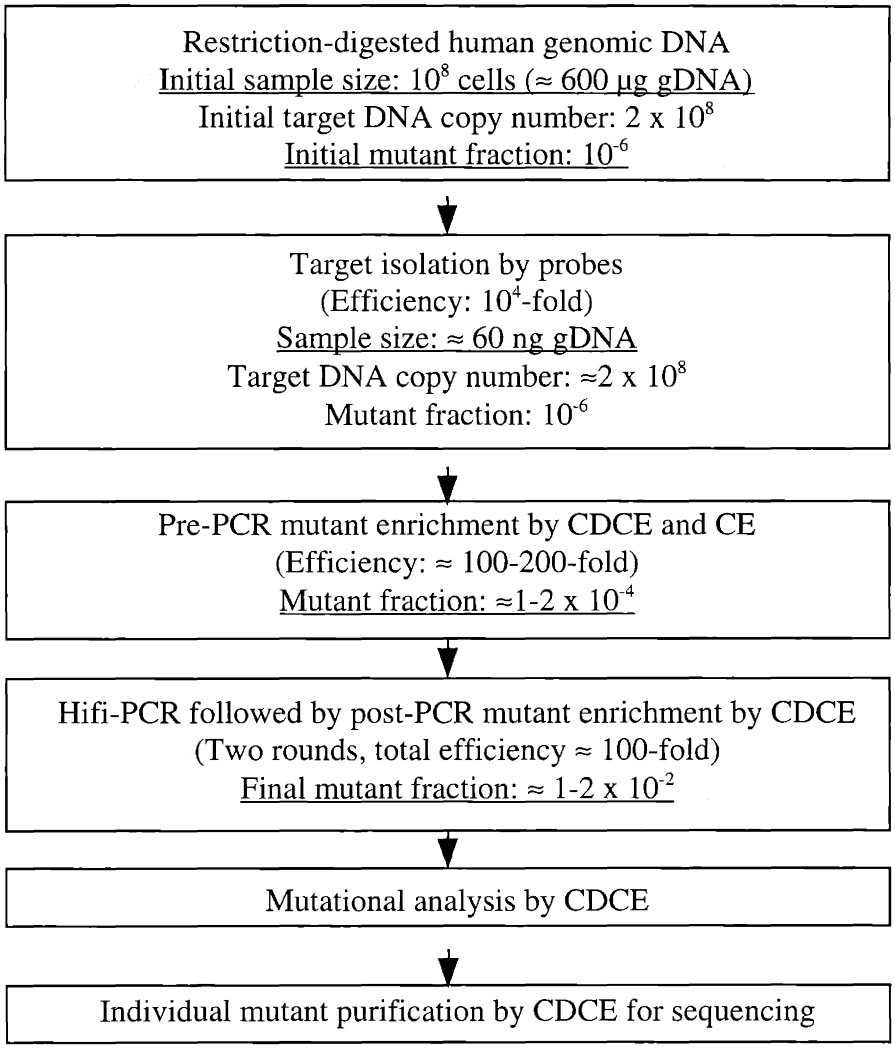


Figure 8. Flow diagram of target isolation/CDCE/hifi-PCR

A chosen target is isolated from a genomic DNA digest using biotin-labeled probes. For this procedure, sequence-specific hybridization is coupled with a biotin-streptavidin capture system. This procedure is followed by pre-PCR mutant enrichment, performed first by separating mutant from wild-type sequences by CDCE and then by eluting these separated mutants. Capillary electrophoresis (CE) is then performed as an additional means to enrich for mutants; CE differentiates mutants from wild-type containing species, such as single strands, in CDCE mutant-enriched samples. Post-PCR mutant enrichment is also performed by CDCE separation and elution. Hifi-PCR and post-PCR mutant enrichment, repeated in two rounds, are performed prior to mutational analysis by CDCE. At this stage, each individual mutant is CDCE purified for sequencing.

Source: Li-Sucholeiki and Thilly, 2000.

To date, target isolation/CDCE/hifi-PCR offers the largest target size over allele-specific PCR and RFLP/PCR. In addition, this method allows detection of point mutations at fractions as low as 10^{-6} with a statistical significance, such as ±20% precision, through its ability to handle large sample sizes of up to about 10^8 human cells (≈600 μg genomic DNA). Analysis of such a large sample size is permitted by target isolation by which an initial sample size of 600 μg of human genomic DNA can be reduced to 60 ng, while retaining the same copy number of a chosen target.

One major drawback of target isolation/CDCE/hifi-PCR is that it can be applied only to DNA sequences with a neighboring natural clamp. To increase the pool of target sequences in the human genome, a means of attaching a clamp to any sequence of interest needs to be included in this method. Available methods of clamp attachment are reviewed in the following section.

2.5.3.3 Clamp attachment

Point mutational analysis by DGGE, CDGE, or CDCE requires DNA sequences with a clamp. While sequences with a natural clamp can be directly analyzed by these methods, clamp attachment is necessary for those sequences without a natural clamp. Three methods are available for attaching a clamp to a DNA sequence of interest (Myers et al., 1985b; Sheffield et al., 1989; Abrams et al., 1990). The principle of each method with limitations is reviewed in this section.

2.5.3.3.1 Cloning-based method

Myers et al. (1985a) first predicted and demonstrated that the melting property of a DNA sequence can be changed by clamp attachment. A clamp, rich in GC bases, is attached to a target sequence of interest by inserting the target into a plasmid vector containing the clamp. For this insertion, linkers of restriction ends are ligated to both ends of the target and the clamp: *Bam*H I (5'GATC3') and *Bgl* II (5'GATC3') to the target, and *Bam*H I and *Hind* III (5'AGCT3') to the clamp. Thus, the clamp can be ligated to the plasmid after digesting the plasmid with *Bam*H I and *Hind* III, and either

end of the target can be ligated to the *Bam*H I end of the clamp after digesting the clamp-ligated plasmid with *Bam*H I.

While point mutations in a chosen target with the attached clamp have been shown to be separated by DGGE, the same demonstration could not be performed using the same target without a clamp (Myers et al., 1985b). This demonstration shows that clamp attachment can make DNA sequences without a natural clamp suitable for point mutational analysis by DGGE, CDGE, or CDCE. However, this cloning-based method of clamp attachment has not been demonstrated with genomic DNA. Thus, this method as it is cannot be applied to mutational analysis in human cells and tissues.

2.5.3.3.2 PCR-based method

Two PCR-based methods of clamp attachment have been developed. The primary method has been demonstrated using genomic DNA as a template and using primers to which an additional 40-45-bp sequence is added at the 5' end (Sheffield et al., 1989). This added sequence is non-complementary to the target and rich in GC bases. A clamp is attached to one end of the amplified sequence by the GC-rich primers during PCR.

The second method is rather an alternative to clamp attachment, called chemiclamping (Girodon et al., 1993; Gille et al., 1998). This method uses PCR primers with a psoralen derivative at the 5'end, combined with UV irradiation of the PCR products. As a result, cross-linking is induced, forming the covalent bond between the two target strands of the PCR products. Depending on whether the cross-linking is desired at one or both ends of a target, the 5' modified primers can be designed for one (Girodon et al., 1993) or both ends (Gille et al., 1998). Although a GC-rich sequence is not added to the primers, point mutations in a cross-linked target have been demonstrated to be suitable for DGGE separation (Girodon et al., 1993; Gille et al., 1998).

The major advantage of the PCR-based methods of clamp attachment over cloning- and probe-based methods is their simplicity. However, these methods cannot be applied to analysis of point mutations at fractions below 2×10^{-4}. Detection of mutations at such low fractions requires pre-PCR mutant enrichment, as Section 2.5.3.1 discusses and

Figure 8 in Section 2.5.3.2.3 illustrates. Therefore, clamp attachment must be performed without PCR.

2.5.3.3.3 Probe-based method

Figure 9 illustrates the probe-based method of clamp attachment (Abrams et al., 1990). This method is the only available method of clamp attachment, which can potentially be applied to target isolation/CDCE/hifi-PCR. Such an application would allow analysis of point mutations at fractions below 2×10^{-4} in any DNA sequence of interest in the human genome.

However, two factors regarding this probe-based method should be noted. The first is the preparation procedure of the probes. The probes are prepared by constructing single-stranded plasmid vectors containing a clamp and inserting a chosen target sequence into the vectors. These vectors are used as a template for generating biotin-labeled probes by primer extension using biotin-labeled primers. These primer-extended vectors are then restriction digested and denatured with NaOH. The strand with the biotin-label is purified by avidin-agarose chromatography, which distinguishes biotin-labeled from unlabeled fragments. This procedure is followed by alkaline-agarose gel electrophoresis, which differentiates single-stranded DNA fragments based on their length (McDonell et al., 1977).

This probe preparation procedure is not only laborious but introduces mutations to the probes at the same time. *Klenow* DNA polymerase, used for primer extension, has been shown to create small deletions and single base-pair substitutions (Scharf et al., 1986; de Boer and Ripley, 1988; Keohavong et al., 1989; Bell et al., 1997) at a rate of about 10^{-4}/bp/doubling (Scharf et al., 1986; Keohavong et al., 1989). Thus, any *Klenow* hotspot mutations occurring 10 times more frequently than expected by chance (1% hotspots) would be generated at a rate of about 10^{-3}/doubling. These 1% hotspot mutations are expected in the probes at a fraction of 10^{-3} after primer extension. Thus, unless the probes are purified for the wild-type, the probes cannot be applied to analysis of point mutations at fractions below 10^{-3}, assuming *Klenow* hotspot mutations are not generated any higher than 1%.

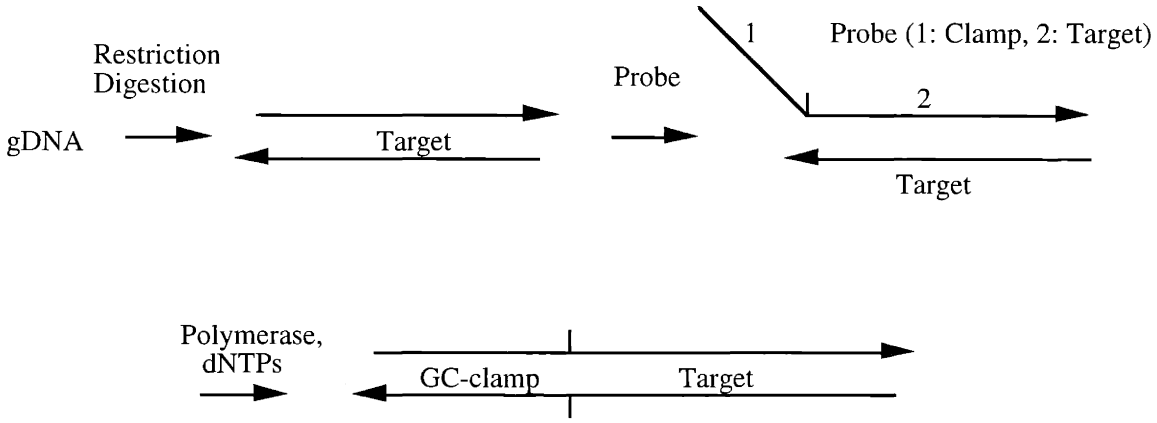


Figure 9. Probe-based method of clamp attachment

Human genomic DNA is restriction digested to liberate a chosen target. This DNA is then denatured, mixed with probes, and allowed to form probe-target hybrids. The probes contain the sequence complementary to the entire target of choice with an additional sequence belonging to a clamp, rich in GC bases, at the 5'end. These hybrids of partial double strands are treated with DNA polymerase and four dNTPs to generate fully double-stranded fragments.

Source: Abrams et al., 1990.

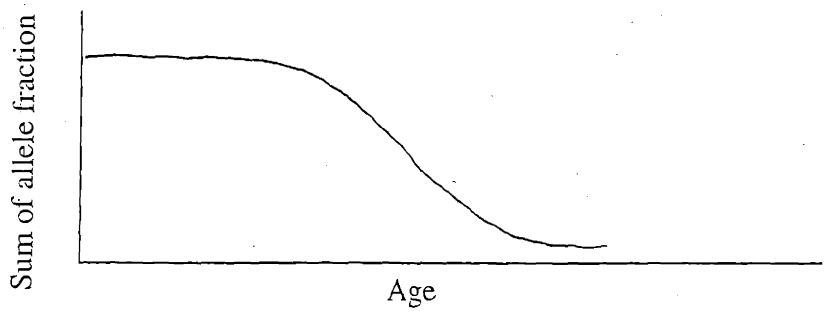
The second factor that should be noted regarding the probe-based method is the efficiency of clamp attachment to a chosen target. Assuming 100% of a chosen target in a genomic DNA digest forms hybrids with the probes, the clamp attachment efficiency would be equal to the efficiency of the probe-target hybrids converting to fully base-paired fragments (see Figure 9). It is desirable to have a clamp sequence rich in GC bases, but increase in the GC content of a template DNA has been shown to reduce the template amplification efficiency (McDowell et al., 1998). Thus, the efficiency of the hybrids converting to fully base-paired fragments may be low. This efficiency can be increased to some degree by introducing additives such as formamide (Sarkar et al., 1990; Varadaraj and Skinner, 1994), DMSO (Winship, 1989; Bookstein et al., 1990; Sun et al., 1993), and betaine (Henke et al., 1997). However, the clamp attachment efficiency of the probe-based method has not been determined. Especially when DNA sample size is a limiting factor for analysis of point mutations at low fractions, achieving high efficiency clamp attachment is critical.

The clamp attachment method developed in this thesis generates a clamp attachment efficiency of greater than 95%. This method is based on enzymatic ligation coupled with mass action. Since clamp attachment is performed without PCR, this method can be combined with target isolation/CDCE/hifi-PCR for analysis of point mutations at fractions below 2×10^{-4}. This combination increases the scanning range for the human genome. Target isolation/CDCE/hifi-PCR alone can be applied to only those DNA sequences with a natural clamp, representing about 9% of the human genome. In addition, none of gene-coding regions for certain genes like *p53* and *BRCA* 1 can be studied by this method. However, the combined method of clamp ligation and target isolation/CDCE/hifi-PCR allows analysis of the majority, if not all, of gene-coding regions.

2.6 Criteria for measuring inherited point mutational spectra in human populations

Studies of inherited point mutational spectra in human populations allow testing of the relationship of genes to diseases and the discovery of causative alleles over the entire gene, as Figure 10 summarizes. However, such studies require analysis of a large

A. Age-specific decline in sum of alleles coding for risk of disease



B. Expected outcome of allele-specific decline

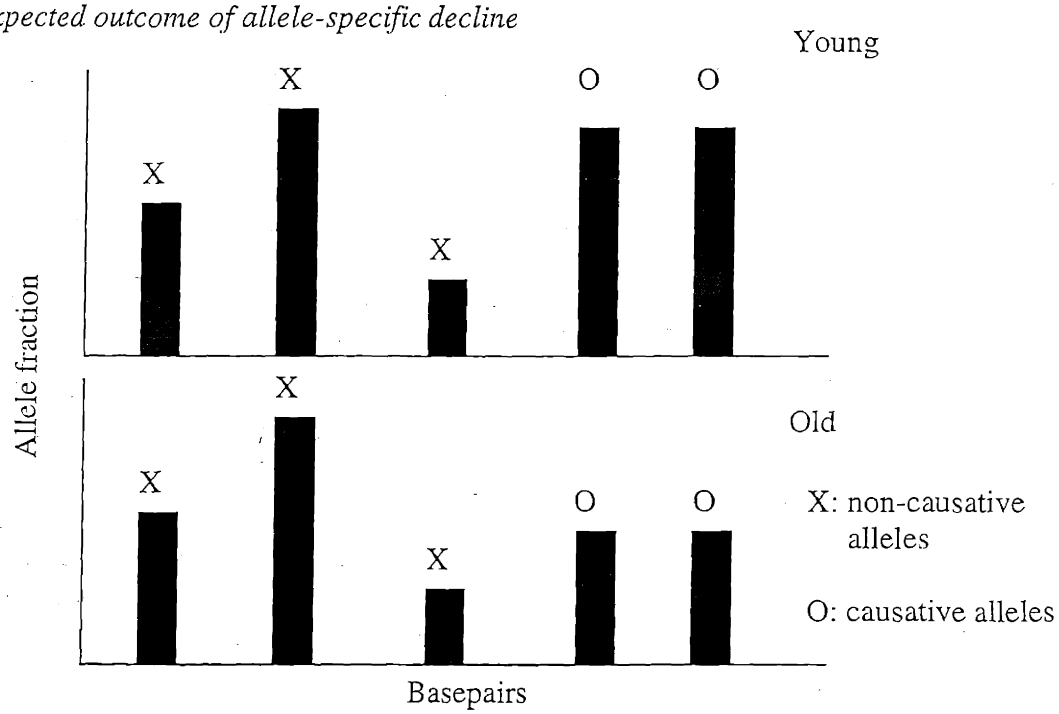


Figure 10. Analysis of inherited point mutational spectra in human populations

A. Age-specific decline in sum of alleles coding for risk of disease is expected in human populations since carriers would die earlier in life than non-carriers.
B. If the hypothesis illustrated in Figure 10 A is true, analysis of the entire set of a gene's alleles (inherited point mutational spectra) in human populations as a function of age can be used as a means to identify causative alleles. Allele-specific decline is expected for causative alleles as people age, while non-causative allele fractions are expected to stay the same.

Source: Tomita-Mitchell et al., 1998.

number of individuals to generate observed mutations with statistical significance. Fulfilling such a requirement demands sensitive and general methods to detect mutations in a large number of individuals.

To analyze alleles at fractions as low as 1% hotspots belonging to non-deleterious genes, methods of detecting mutations at a fraction of 5×10^{-5} in human populations are necessary. For example, in a sample size of 10^7cells (2×10^7 allele copies), derived from 10^5 persons, each person is represented by 100 cells. Alleles at a fraction of 5×10^{-5} in this sample are present as 10^3 copies ($(2 \times 10^7) \div (5 \times 10^{-5})$), which can be detected with statistical significance using mutation detection methods with a sensitivity of 5×10^{-5}. A total gene-inactivating allele fraction of 3×10^{-1} for non-deleterious genes is expected in human populations, and a fraction of 5×10^{-3} for recessive deleterious genes (William G. Thilly, personal communication). 1% hotspot alleles in non-deleterious genes is then expected to appear at a fraction of 5×10^{-5}.

To sum up, sensitive and general methods of measuring mutations based on their genotype can be applied to analysis of inherited mutations in human populations, in addition to analysis of somatic mutations in normal human tissues. These analyses may allow an understanding of human mutagenesis.

3 MATERIALS AND METHODS

3.1 Construction of melting profiles

Sequence-specific melting profiles were constructed using WinMelt™ 2.0 (Medprobe, Norway), software based on a previous work (Lerman and Silverstein, 1987). WinMelt™ 2.0 constructs melting profiles of known sequences as a function of sequence and temperature. This software uses an algorithm for calculating the equilibrium melting transition probability (Fixman and Freire, 1977), combined with the neighbor base-pair doublet parameters (Gotoh and Tagashira, 1981). A close correlation between calculated and experimental measurements has been demonstrated (Myers et al., 1985a, b), ensuring that the algorithm is accurate.

3.2 Proposed method of clamp attachment

3.2.1 Ligation-based methodology

Figure 11 illustrates the proposed method of attaching a clamp to any 100-bp sequence in the human genome. High efficiency clamp ligation to the restriction ends of a chosen target by mass action is expected in a ligation reaction mixture containing at least a 10-fold molar excess of the clamp.

About 3000 restriction enzymes have been found, exhibiting over 200 different specificities (New England BioLabs, 2000-2001; Roberts and Macelis, 2001). Each enzyme recognizes a specific DNA sequence of 4-6-bp (recognition site), cleaves within or next to the recognition site, and generates either blunt or cohesive ends (reviewed in Pingoud and Jeltsch, 1997).

DNA ligase joins DNA by catalyzing the formation of phosphodiester bonds between juxtaposed 5'-phosphoryl and 3'-hydroxyl termini (Kellenberger et al., 1961; Gellert, 1966; Cozzarelli et al., 1967; Gefter et al., 1967; Gellert, 1967; Olivera and Lehman, 1967; Weiss and Richardson, 1967; Zimmerman et al., 1967) (DNA ligase is reviewed in Lehman, 1974; Higgins and Cossarelli, 1979; Engler and Richardson, 1982). Thus,

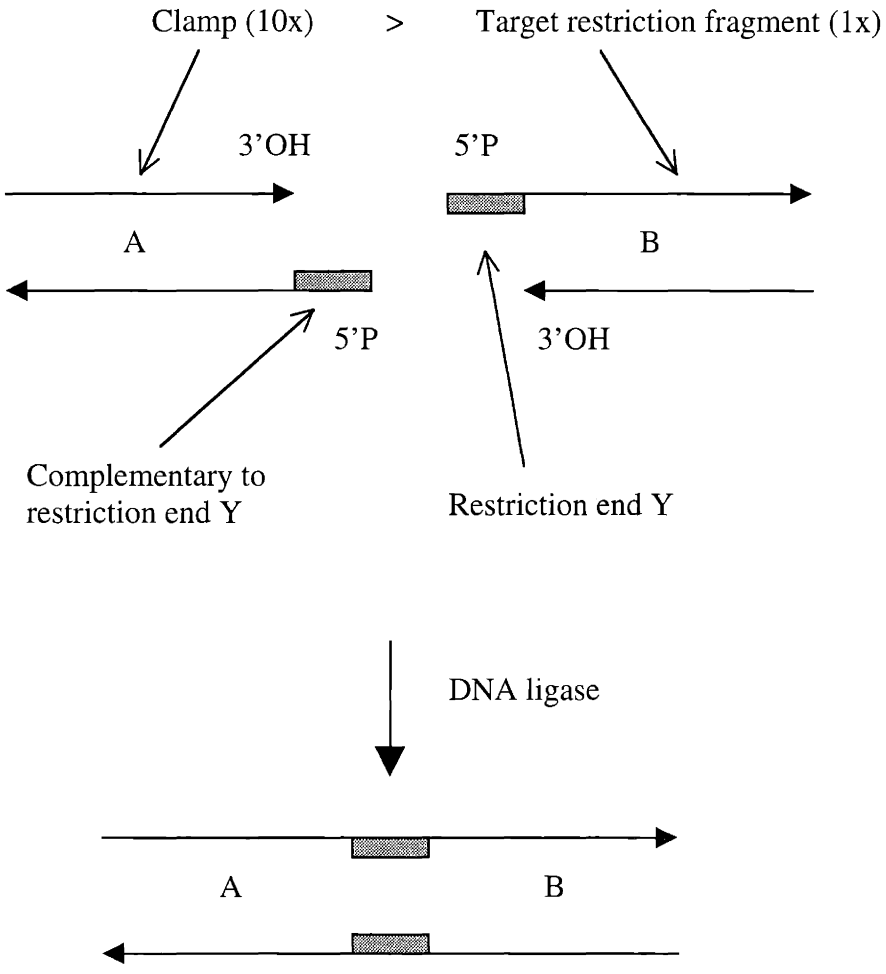


Figure 11. Proposed method of clamp attachment

One end of a desired target is liberated from genomic DNA using a restriction enzyme of choice. For each restriction enzyme chosen, a clamp, with an end complementary to the one generated by the chosen enzyme, is prepared. This clamp is then covalently joined to the target restriction end by DNA ligase. High efficiency clamp ligation by mass action is expected in a ligation reaction mixture containing at least a 10-fold molar excess of the clamp.

Source: Kim et al., 2001.

restriction enzymes produce suitable substrates for ligation since they cleave DNA in both strands by hydrolyzing phosphodiester bonds (Connolly et al., 1984; Grasby and Connolly, 1992; Jeltsch et al., 1992). Ligation of restriction-generated ends was first demonstrated with cohesive ends of 4 and 6 nucleotides (Mertz and Davis, 1972). Since then, ligation of ends as short as one nucleotide (Hung and Wensick, 1984), as well as fully base-paired, or blunt ends (Ehrlich et al., 1977; Mottes et al., 1977; Sgaramella and Ehrlich, 1978), has been demonstrated.

3.2.2 Method of estimating number of chosen restriction ends per human cell

The theoretical number of chosen restriction ends per diploid human cell was estimated based on three factors. First, a haploid human genome contains about 3 x 10^9-bp (Current protocols in molecular biology, 1987; Lander et al., 2001; Venter et al., 2001). Second, two ends are generated each time a restriction endonuclease cleaves its recognition site. Third, depending on the size of the recognition site, the frequency with which an enzyme cleaves would vary, producing fragments with different expected average lengths. For example, an enzyme with a 4-base recognition site would cleave every 256 bases, assuming the human genome is comprised of 50% GC bases. Although the GC content determined for humans is 41% (Marmur and Doty, 1962; Lander et al., 2001), 50% was used instead to simplify the estimation. Thus, the cleavage frequency in bases was estimated as follows: 4^X.

3.2.3 Preparation of clamp for Apo I restriction end

Figure 12 illustrates the characteristics of the clamp prepared for the restriction end generated by *Apo* I restriction endonuclease. *Apo* I has a 6-base recognition sequence and cleaves after the first base pair, generating a 5'-AATT-3' cohesive end (Polisson and Robinson, 1992), as Figure 13 shows.

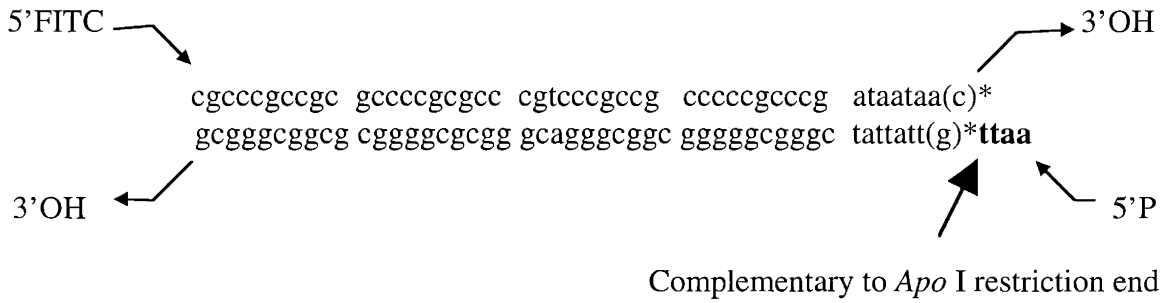
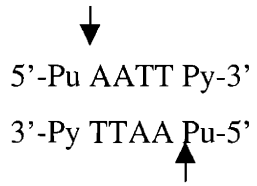


Figure 12. Characteristics of clamp for *Apo* I restriction end

The first 40-bp sequence of the clamp is from a previous study (Sheffield et al., 1989). This clamp has an end that is complementary to, or the same as, the *Apo* I. This end contains 5'-phosphoryl and 3'-hydroxyl residues, allowing clamp self-ligation, as well as clamp ligation to the restriction end generated by *Apo* I.

* The base pair in parentheses was omitted when preparing the clamp using the second approach (see Section 3.2.3.2). This omission was necessary to generate *Apo* I restriction recognition site, where the clamp was ligated to the chosen *HPRT* target.

Figure 13. Recognition sequence of *Apo* I restriction endonuclease



3.2.3.1 First approach: direct hybridization

As a first means of preparing the clamp, a pair of complementary oligonucleotides were synthesized (Synthetic Genetics, San Diego, CA), polyacrylamide gel electrophoresis (PAGE)-purified (Synthetic Genetics), and hybridized to each other. During the synthesis, one strand of the clamp was fluorescein labeled at the 5' end (5'FITC), and a phosphate group was attached at the 5' end (5P') of the other strand, as Figure 12 illustrates. Clamp hybridization was performed at 75°C for 15 min in a buffer containing 0.2 M NaCl, 10 mM Tris-HCl (pH. 7.6), and 2 mM EDTA.

3.2.3.2 Second approach: purification by CDCE

As a second means of preparing the clamp, the chosen *HPRT* target with the ligated clamp was PCR amplified using the GC1/P1 primer pair. This procedure was followed by CDCE purification of the wild-type DNA. This DNA was then restriction digested by *Apo* I prior to CE purification of the target-cleaved clamp.

3.3 Overview of proposed point mutation detection method

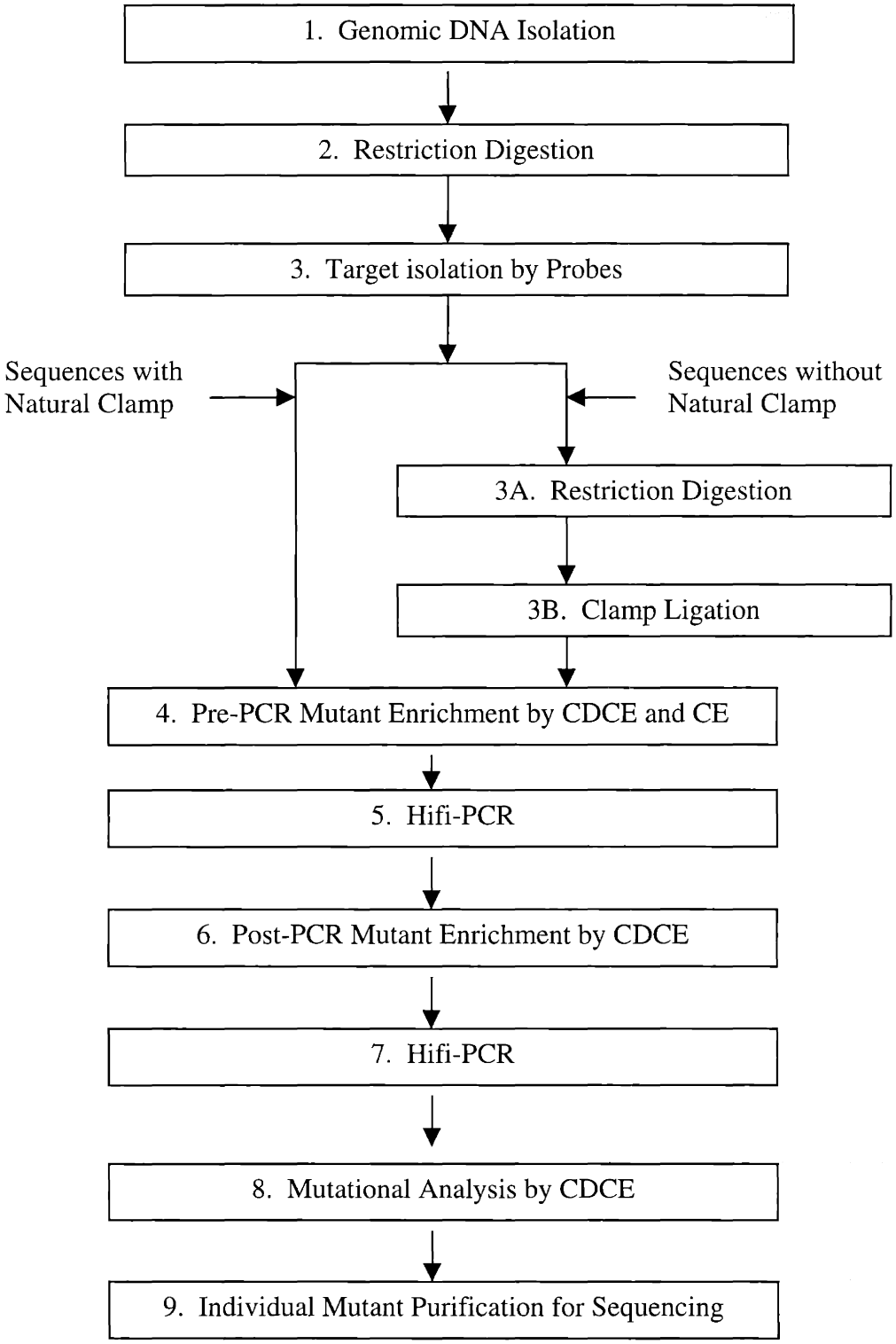
Figure 14 shows a flow diagram of the proposed method, which allows analysis of rare point mutations at fractions below 2×10^{-4} in almost the entire human genome.

Figure 14. Flow diagram of proposed point mutation detection method: natural vs. ligated clamp

Natural clamp: For analysis of DNA sequences with a natural clamp, a chosen target is isolated from a genomic DNA digest using biotin-labeled probes. For this procedure, sequence-specific hybridization is coupled with a biotin-streptavidin capture system. This procedure is followed by pre-PCR mutant enrichment, performed first by separating mutant from wild-type sequences by CDCE and then by eluting these separated mutants. CE is performed as an additional means to enrich for mutants; CE differentiates mutants from wild-type containing species, such as single strands, in CDCE mutant-enriched samples. After hifi-PCR, CDCE is performed for further mutant enrichment. Final mutant-enriched samples are then PCR amplified prior to mutational analysis by CDCE. At this stage, each individual mutant is CDCE purified for sequencing.

Ligated clamp: For analysis of DNA sequences without a natural clamp, restriction digestion and clamp ligation are performed prior to pre-PCR mutant enrichment by CDCE.

Source: Kim et al., 2001.


1. Genomic DNA Isolation
2. Restriction Digestion
3. Target isolation by Probes
Sequences with Natural Clamp
Sequences without Natural Clamp
3A. Restriction Digestion
3B. Clamp Ligation
4. Pre-PCR Mutant Enrichment by CDCE and CE
5. Hifi-PCR
6. Post-PCR Mutant Enrichment by CDCE
7. Hifi-PCR
8. Mutational Analysis by CDCE
9. Individual Mutant Purification for Sequencing

3.4 Test target sequence: *HPRT* exon 3

A part of the human *HPRT* gene (cDNA bp 223-318) was chosen as a test target to discover if the proposed point mutation detection method is as sensitive as the one established for DNA sequences with a natural clamp. This target is comprised of 52% of the *HPRT* exon 3 (cDNA bp 135-318) and is juxtaposed with a domain of a higher melting temperature, a natural clamp (cDNA bp 141-216). Thus, the chosen target allows a direct sensitivity comparison of the proposed mutation detection method using the ligated clamp to that of the established method using the natural clamp.

Figure 15 illustrates the melting profiles of the chosen target. These melting profiles suggest that the chosen target with either the natural or ligated clamp is suitable for CDCE separation.

3.5 Preparation of internal standards

3.5.1 Homoduplex mutant of genomic DNA

A homoduplex mutant of genomic DNA carrying a G -> T transversion at *HPRT* cDNA bp 312 was isolated from $HPRT_{Munich}$ cells. $HPRT_{Munich}$ cells are a lymphoblast line, isolated from a male patient with gout (Wilson and Kelley, 1984). Separation of this mutation in the chosen *HPRT* target from the wild-type has been demonstrated by DGGE (Cariello et al., 1988).

3.5.2 Homoduplex mutant of PCR products (438-bp)

A 438-bp homoduplex mutant carrying a G -> A transition at *HPRT* cDNA bp 309 was constructed by PCR. This mutant is compatible with the chosen *HPRT* target-embedded fragment liberated from genomic DNA by *Bst*N I and *Dra* I restriction endonucleases.

To construct this mutant, three sets of PCR were performed. The first was performed using the IS1/P1-unstable primer pair and using human wild-type genomic DNA as a

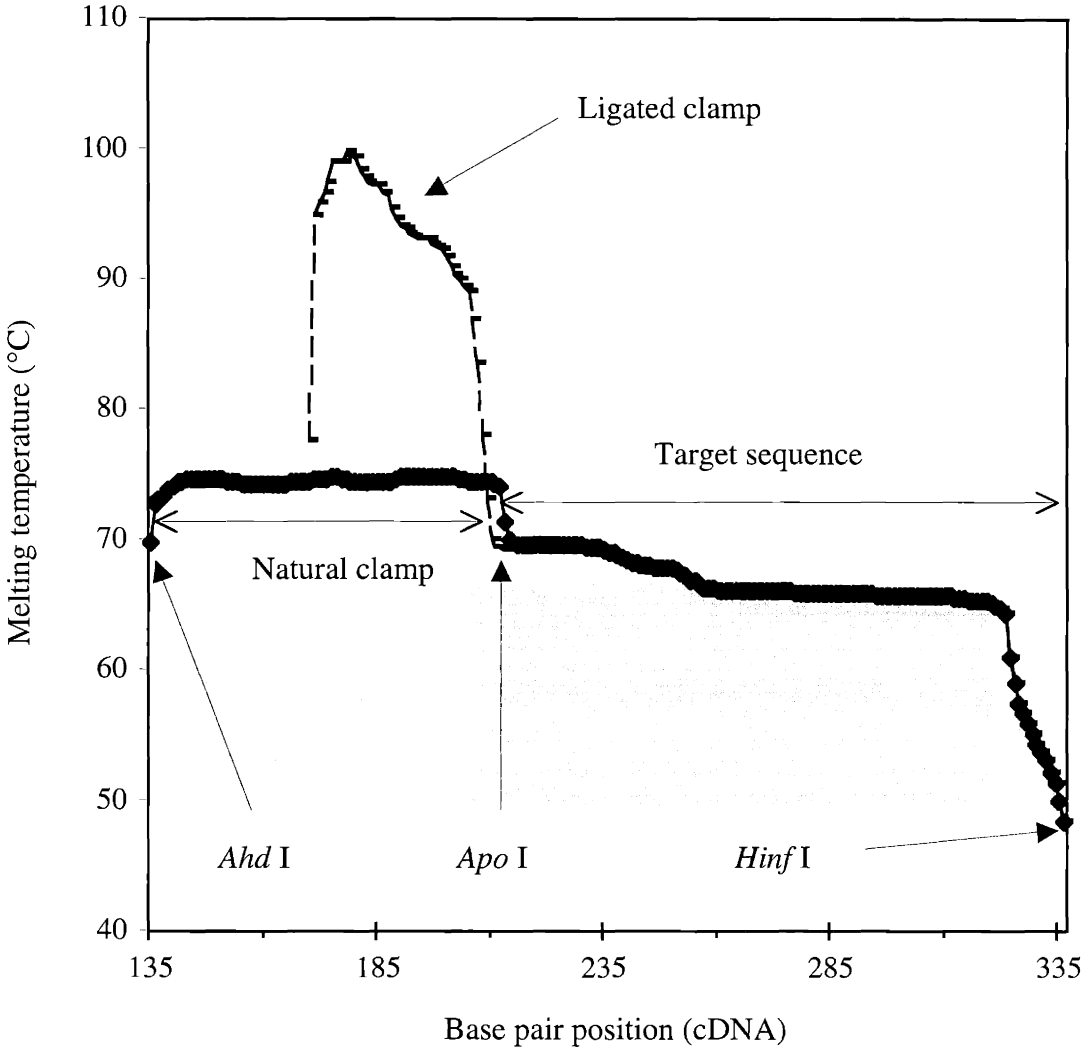


Figure 15. Melting profiles of human *HPRT* exon 3: natural vs. ligated clamp

The solid line represents the melting profile of the target wild-type with the natural clamp, whereas the dotted line represents the melting profile of the same target with the ligated clamp. Restriction digestion by *Ahd* I and *Hinf* I liberates the chosen target with the natural clamp, suitable for CDCE separation. Restriction digestion by *Apo* I liberates the natural clamp from the chosen target with the generation of the *Apo* I restriction end (5'AATT3', *HPRT* cDNA bp 218-221). The prepared clamp for the *Apo* I restriction end can be ligated to this target end, making the *Apo* I-digested target suitable for CDCE separation.

template to generate a 295-bp mutant fragment. The amplified fragment was then used as a primer, together with IS2, for the second set of PCR; human wild-type genomic DNA was used as a template, and the 438-bp mutant fragment was generated. This fragment was purified by CE, performed using a 75-μm id capillary at a constant current of 9 μA. The purified fragment was then used as a template for the third set of PCR; the IS1/IS2 primer pair was used, and enough cycles were applied to convert all the primers into products. Upon this conversion, the expected concentration of the amplified mutant was about 10^{11} copies/μl, equivalent to the initial primer copy number. The known concentrations of the mutant to be used as stock were then made by subsequently diluting the PCR products with dH_20, 10-fold each time.

3.5.3 Homoduplex mutant of PCR products (198-bp/169-bp)

A 198-bp or 169-bp homoduplex mutant carrying a G -> A transition at *HPRT* cDNA bp 309 was constructed by PCR. This mutant is compatible with the chosen *HPRT* target with the natural clamp (198-bp) after *Ahd* I and *Hinf* I restriction digestion or with the ligated clamp (169-bp) after clamp ligation.

The mutant with the natural clamp was amplified using the P3/P1-unstable primer pair and using human wild-type genomic DNA as a template. The same mutant with the ligated clamp was amplified using the GC1/P1-unstable primer pair and using the clamp-ligated wild-type DNA of the chosen *HPRT* target as a template. The known concentrations of the mutant to be used as stock were prepared as described in Section 3.5.2.

3.5.4 Heteroduplex mutant of PCR products (198-bp)

To create a mutant in mutant/wild-type heteroduplexes, first, a homoduplex mutant with the natural clamp (G -> A at *HPRT* cDNA bp 309) was prepared as described in Section 3.5.3. Second, the wild-type and the mutant, each in a homoduplex form at a molar ratio of 100 to 1 were co-amplified using the P3/P1 primer pair. Enough cycles were applied to this PCR in which the homoduplex mutant formed heteroduplexes with

the wild-type by mass action. The known concentrations of the mutant in heteroduplexes to be used as stock were prepared as described in Section 3.5.2.

3.6 PCR

3.6.1 Primers

Figure 16 and Table 10 summarize the sequences and sites of the primers used in this study analyzing the chosen *HPRT* target.

The names and sequences of the primers used in this study analyzing the chosen *p53* target are as follows:

GC2: 5'FITC-GCCGC CTGCA GCCCG CGCCC CCCGT GCCCC CGCCC CGCCG CCGGC CCGGG CGCCT T-3'
BS: 5'-TGACC TGGAG TCTTC C-3'
BS-stable: 5'-TGACC TGGAG TCTTC CAGTG TGACG ATG-3', underlined is an AT -> GC transition at cDNA bp 763
HH: 5'FITC-CTTGC CACAG GTCTC CCCAA-3'
AP: 5'-TATGG AAGAA ATCGG TAAGA-3'

Each primer was synthesized and PAGE-purified by Synthetic Genetics (San Diego, CA).

3.6.2 Reaction conditions

PCR was performed inside closed glass capillaries using an Air Thermo Cycler (Idaho Technologies, Idaho Falls, ID). Each reaction mixture of 10 to 50 µl contained 20 mM Tris-HCl (pH 8.0), 2 mM $MgCl_2$, 10 mM KCl, 6mM $(NH_4)_2SO_4$, 0.1% Triton X-100, 100 µg/ml nuclease-free BSA, 0.2 µM each primer, 0.1 mM dNTPs (Pharmacia, Piscataway, NJ), and 0.1 U/µl *Pfu Turbo*™ DNA polymerase (Stratagene, La Jolla, CA). Each PCR cycle consisted of 7 s of template denaturation at 94°C, 30 s of template-primer annealing

5'- [b]GGTTGGTGTG GAAGTTTAAT GACTAAGAGG [c]tgtttgttat

1

aaagtttaat gtatgaaact TTCTATTAAA TT[d]cctgattt

tatttctgta **ggactga[e]acg tcTTGCTCGA GATGTGATGA**

2

AGGAGATGGG AGGCCATCAC ATTGTAGCCC TCTGTGTGCT

CAAGGGGGGC TATA[f]AATTCT TTGCTGACCT GCTGGATTAC

ATCAAAGCAC TGAATAGAAA TAGTGATAGA TCCATTCCTA

TGACTGTAGA TTTTATCAGA CTGAAGAGCT ATTGTGTGAG

TATATTT AAT ATATG[g]ATTCT TTTTAGTGGC AACAGTAGGT

3

TTTCTTATAT TTTCTTTGAA TCTCTGCAAA CCATACTTGC

TTTCATTTCA CTTGGTTACA GTGAGATTTT TCTAACATAT

TCACTAGTAC TTTACATCAA AGCCAATACT GTTTTTTT[h] –3'

4

Figure 16. Primer sites for chosen *HPRT* target[a]

[a] For each primer, the position of the first base is underlined with a number below (see Table 10 for the name and sequence information).
[b] *Bst*N I restriction end.
[c&d] Two groups of bases in lowercase letters represent sites for Probe 1[c] and Probe 2[d]. These probes were used for target isolation.
[e] *Ahd* I restriction end.
[f] *Apo* I restriction end.
[g] *Hinf* I restriction end.
[h] *Dra* I restriction end.
* The group of bases in bold represents the entire sequence of *HPRT* exon 3 (cDNA bp 135-318).

Table 10. Sequences of primers for chosen *HPRT* target

Name	Site[a]	Sequence
IS1[b]	1	5'-GGTTG GTGTG GAAGT TTAAT GAC-3'
P3[b]	2	5'-ACGTC TTGCT CGAGA TGTGA-3'
GC1[b, c]		5'-CGCCC GCCGC GCCCC GCGCC CGTCC CGCCG CCCCC GCCCG ATAAT AA-3'
P1	3	5'-CATAT ATTAA ATATA CTCAC-3'
P1-stable	3	5'-CATAT ATTAA ATATA CTCAC ACGAT AGCTC TTCAG-3' (underlined is an AT -> GC transition at cDNA bp 316)
P1-unstable	3	5'-CATAT ATTAA ATATA CTCAC ACAAT AGCTT TTC AG-3' (underlined is a GC -> AT transition at cDNA bp 309)
IS2	4	5'-AAAAA AACAG TATTG GCTTT-3'

[a] See Figure 16.

[b] 5'end fluorescein-labeled primers.

[c] This primer, together with P1, was used to amplify the chosen *HPRT* target with the ligated clamp.

at 45°C, and 10 s of primer extension at 72°C. After a desired number of cycles, incubation followed, at 72°C for 2 min and at 45°C for 15 min. Under the reaction conditions specified, about 60% PCR efficiency was observed for the chosen *HPRT* target with either the natural (the P3/P1 primer pair) or ligated (the GC1/P1 primer pair) clamp.

When amplifying the chosen *HPRT* target with the ligated clamp, byproducts, representing between 5 and 20% of the total amplified products, were formed. This is thought to be the result of using the high GC-content primer, GC1 (83%). These byproducts, discovered to be a form of the wild-type, had slower electrophoretic mobilities that that of the wild-type under optimal CDCE-separating conditions, which separated mutant from wild-type sequences. This confounded quantitative analysis of the amplified products by CDCE. However, this problem was minimized by incubating the PCR sample with 0.05 U/μl Exo$^-$ *Pfu* DNA polymerase (Stratagene, La Jolla, CA) for 5 min at 72°C prior to CDCE analysis.

3.7 CDCE and CE

3.7.1 Instrumentation and operating conditions

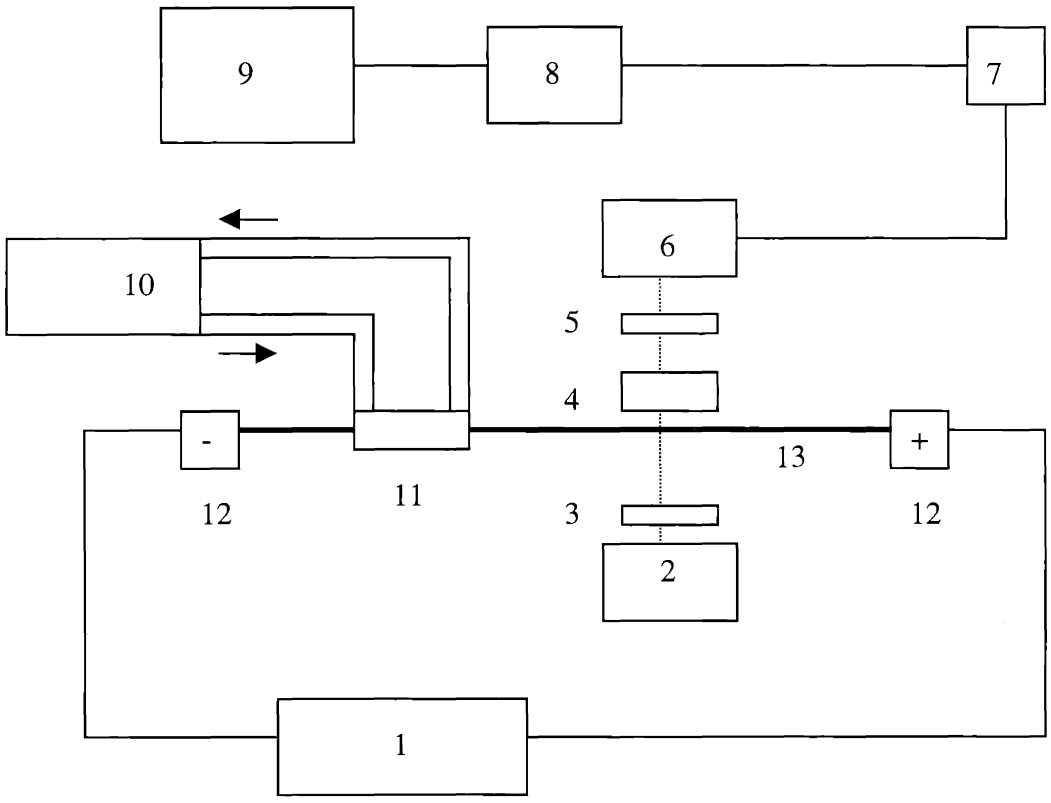
Figure 17 depicts a CDCE apparatus. CDCE uses elevated temperatures to permit separation of mutant from wild-type sequences (Khrapko et al., 1994a). A portion of the capillary is inserted into a water jacket, a regulated-temperature zone where the separation takes place; the temperature of the water jacket is controlled by a constant temperature water circulator. CE is performed at room temperature to separate DNA molecules based on their length and to separate double- from single-stranded DNA of the same length.

Depending on DNA separation efficiency or loading capacity desired, CDCE or CE was performed differently: using a capillary of different inner diameter (id) (e.g., 75 to 540 μm), a capillary of various lengths (e.g., 15 to 30 cm), or a water jacket of various lengths (e.g., 6 to 20 cm). DNA loading capacity has been shown to be increased up to about 100-fold (100 ng vs. 10 μg) as the id of a capillary is increased from 75 to 540 μm

Figure 17. Diagram of CDCE apparatus

CDCE apparatus consists of a CE apparatus (Cohen et al., 1990), a heated zone of constant temperature, and a detector (Khrapko et al., 1994a). Fluorescently-labeled molecules are detected as they pass through a single point of a capillary where an argon laser beam is focused. The emitted fluorescence signal is collected by a microscope objective and directed through filters into a photomultiplier. The signal is then amplified by a current preamplifier and recorded by a computerized data acquisition system. Both ends of the capillary are bathed in buffer reservoirs filled with 1 x TBE (89 mM Tris-borate (pH. 8.0), 1 mM EDTA).

Adapted from Khrapko et al., 1994a; Muniappan and Thilly, 1999; Li-Sucholeiki et al., 1999.



1. Power supply (Model CZE 1000R-2032; Spellman, Hauppauge, NY).
2. Argon ion laser (Model 5425ASL; Ion Laser Technology, Salt Lake City, UT).
3. 488 nm filter (Corion, Franklin, MA).
4. Microscope objective (Newport, Irvine, CA).
5. 540 and 530 nm filters (Corion).
6. Photomultiplier (Oriel Instruments, Stratford, CT).
7. Current preamplifier (Oriel Instruments).
8. Data acquisition system (ACM2-16-8A/T51Bl; Strawberry Tree Computers, Sunnyvale, CA).
9. Computer.
10. Constant temperature water circulator.
11. Water jacket.
12. Buffer reservoirs.
13. Capillary.

(Li and Thilly, 1996). Using a longer water jacket increases separation efficiency of different mutant sequences, as well as of mutant from wild-type sequences. Experiments performed in this study showed that when DNA separation was performed at room temperature, increase in the length of a capillary increased separation efficiency of DNA molecules of different lengths.

Desired CDCE and CE separation conditions were determined by performing test runs with DNA molecules, labeled with fluorescein. For example, desired CDCE separation conditions were determined using PCR products containing both mutant and wild-type sequences.

For purification of desired DNA molecules, electro-elution was performed in 0.1 x to 0.8 x TBEB (0.1 x TBEB: 0.1 x TBE, 0.03 mg/ml BSA) as they reached the anodic end of the capillary after CDCE or CE separation.

3.7.2 Coating of capillaries

The inner surface of capillaries was coated with linear polyacrylamide prior to use. This procedure was necessary to prevent electro-osmotic flow, which otherwise limits separation efficiency of desired molecules (Hjertén 1985). The capillaries were coated as described in previous studies (Hanekamp et al., 1996; Khrapko et al., 1997b; Khrapko et al., 2001) (adapted from Hjertén (1985)):

For each procedure, a fused-silica capillary (≈12 m) of a desired id (Polymicro Technologies, Phoenix, AZ) was filled with 1 M NaOH (using a high-pressure syringe with the needle end inserted into a piece of Teflon tubing). This capillary was incubated at room temperature for 1 hr. The capillary was then washed with dH_20, filled with 1 M HCl, and incubated at room temperature for 10 min. After another washing with dH_2O, the capillary was washed again with 100% methanol. The washed capillary was filled with γ-methacryloyloxypropyltrimethoxysilane (Sigma, St Louis, MO) and was incubated at room temperature overnight. Washing with 100% methanol was repeated, and the capillary was filled with 6% acrylamide solution in 1 x TBE containing 0.1% *N, N, N', N'*-tetramethylethylenediamine (TEMED) and 0.025% ammonium persulfate (APS). The capillary was incubated at room temperature overnight before using.

3.7.3 Preparation of separation matrix

For improved resolution and reproducible DNA separation by CDCE and CE, a replaceable linear polyacrylamide matrix was prepared as described in previous studies (Ruiz-Martinez et al., 1993; Hanekamp et al., 1996; Khrapko et al., 1997b; Khrapko et al., 2001). Preparation of such a sieving matrix was permitted by low concentrations of oxygen, radical initiator (APS), and catalyst (TEMED) in a monomer solution with polymerization allowed at low temperatures (Ruiz-Martinez et al., 1993). This procedure was performed as follows:

5% acrylamide solution in 1 x TBE containing 0.03% TEMED was prepared in a flask with the top sealed with multiple layers of Parafilm. Throughout the procedure, the flask was kept in an ice bath, and a stir bar placed in the flask was left in motion. A stainless steel needle, connected to an argon cylinder by silicone tubing, was dipped into the solution through the Parafilm. This solution was deoxygenated by argon bubbling for 1 hr. The tip of the needle was then taken out of the solution but was kept within the flask during the rest of the procedure. 0.003% of APS was added to the solution using a 100-µl high-pressure syringe through the Parafilm. After a brief mixing, needles were inserted into pre-chilled 10- to 50-ml glass syringes. Each needle was put in the flask through Parafilm but was kept out of the solution. The syringe was then filled with argon gas, which was flushed out into the air. Similarity, the syringe was rinsed with a small volume of the solution. The rinsed syringe was filled with the solution, and the syringe needle was replaced with a 200-µl pipet tip with a sealed end. The filled syringe was placed in vertical position and was kept at 4°C to allow the acrylamide to polymerize and also for storing.

The polymerized matrix was dispensed into 100-µl high-pressure syringes. A piece of Teflon tubing was inserted into each syringe needle end to replace the matrix in a capillary before each CDCE and CE separation. Such a replacement procedure has been shown to be critical in generating reproducibility in DNA separation (Khrapko et al., 1994a).

3.7.4 Sample loading

3.7.4.1 Diluted sample

Each PCR-amplified sample was diluted 10-fold with dH_20 prior to loading into a capillary for CDCE or CE separation. Dilution was necessary since sample injection into a capillary is inhibited by the high salt concentration in the PCR sample. A piece of platinum wire and the cathodic end of the capillary were dipped into the diluted sample solution tube, and the sample was electro-kinetically injected by applying a constant current of 2 μA for 30 s. Typically, about 10^8 copies of a PCR sample are loaded into a 75-μm id capillary by this procedure (Hanekamp et al., 1996; Khrapko et al., 2001).

3.7.4.2 Dialyzed sample

To load the entire DNA sample into a capillary, each desired sample was dialyzed by drop dialysis. Drop dialysis is a simple and rapid method of dialyzing small-volumn samples, smaller than 100 μl (Marusyk and Sergeant, 1980). For this procedure, a 0.025-μm membrane filter (Millipore, Bedford, MA) was floated on the surface of 0.001 to 0.01 x TBE poured into a container (e.g., plastic Petri dish). A desired sample was then transferred to the surface of the floating membrane filter by pipeting and was dialyzed against the buffer by leaving a stir bar in motion for 2 hr. The dialyzed sample was then transferred to a tube by pipeting, and its volume was condensed to about 2 to 4 μl by lyophilizer.

The yield of drop dialysis was determined by comparing the copy number of a chosen sequence before and after the procedure; the copy number was measured by quantitative PCR, followed by CDCE. The range of the yields was between 80 and close to 100% as long as the samples contained 50 mM salt (e.g., NaCl).

The use of TBE at low concentrations as a dialysis buffer is to prevent the ionic strength of the condensed sample solution from being higher than that of 1 x TBE. Such consideration was necessary for electro-kinetic injection of the entire DNA sample into a capillary since ions are preferentially loaded before DNA. For sample injection, the

dialyzed and condensed sample was transferred to a piece of Teflon tubing. One end of the tubing was mounted onto the cathodic end of the capillary, and the other end of the tubing was inserted into a buffer reservoir. By applying a desired current (50 μA for a 320-μm id capillary and 80 μA for a 540-μm id capillary) for 2 min, the sample was injected into the capillary. This injection procedure has been demonstrated to be suitable for loading between 85 and close to 100% of sample DNA into a capillary (Khrapko et al., 1994a; Li-Sucholeiki, 1999); close to 100% sample loading was observed in experiments performed in this study.

3.7.5 Methods of quantitative analysis

Quantitative analyses of fluorescently-labeled DNA molecules were performed by CDCE and CE:

To estimate the copy number of a chosen DNA sequence at any stage throughout the procedure outlined in Section 3.3, quantitative PCR was performed as described in previous studies: a known copy number of an internal standard was co-amplified with an aliquot of each sample desired (Khrapko et al., 1997b; Marcelino et al., 1998; Kim et al., 2001; Lim et al., 2001). This PCR sample was then separated by CDCE. The areas under the separated peaks, which represented the wild-type homoduplex (A_W), mutant homoduplex (A_M), and wild-type/mutant heteroduplexes (A_H), were measured, and the chosen sequence copy number was estimated as follows:

$$([A_W + (A_H \div 2)] \div [A_M + (A_H \div 2)] \times C_{IS}) = \#\text{ of chosen sequence copies/}\mu\text{l} \qquad (1)$$

where C_{IS} is the concentration of the mutant internal standard used for the quantitative PCR.

Similarly, enumeration of individual mutant sequences in initial samples (initial mutant fractions) were measured by CDCE. The areas under the CDCE-separated peaks, which represented individual mutants in an initial sample (A_{MS}) and an added mutant internal standard (A_{MI}) at a known fraction (I_{MF}), were measured. These areas were then compared to measure the initial mutant fractions: $(A_{MS} \div A_{MI}) \times I_{MF}$.

The fidelity of *Pfu* DNA polymerase was estimated as follows:

$$\text{fidelity/bp/doubling} = ([(A_{MS} \div A_{MI}) \times I_{MF}] \div T) \div D \qquad (2)$$

where T is the target size in bases and D is the applied number of doublings used for target amplification.

The areas under the CE-separated peaks, which represented DNA molecules differing in length, were measured for different purposes, such as estimating efficiency of clamp ligation and of restriction digestion.

3.8 Mutational analysis: ligated vs. natural clamp

3.8.1 Procedure of genomic DNA isolation

Genomic DNA from human cells was isolated as described in previous studies (Khrapko et al., 1997b; Kim et al., 2001):

Cells were thawed, centrifuged at 1500 g for 10 min, and washed twice with phosphate-buffered saline (PBS) at a concentration of 10^7 cells/ml. These cells were resuspended in 1 x TE buffer (50 mM Tris-HCl (pH 8.0), 10 mM EDTA) at a concentration of 10^7 cells/ml. 20 mg/ml of Proteinase K (Roche Molecular Biochemicals, Indianapolis, IN) and 10% SDS were then added to the final concentrations of 1 mg/ml and 0.5%, respectively. This sample mixture was incubated in a water-bath shaker (100-200 rpm) at 50°C for 3 hr. 30 mg/ml RNase A (Sigma-Aldrich, Inc. Saint Louis, MO) was then added to a final concentration of 20 μg/ml, followed by incubation in a water-bath shaker (100-200 rpm) at 37°C for 1 hr.

This sample was centrifuged at 10,000 g at 4°C for 15 min, followed by the transfer of the central portion of the supernatant into a new tube by pipeting. This step was repeated two to three times to get as much of the supernatant as possible. To the transferred supernatant, 2.5 M NaCl was added to a final concentration of 250 mM, followed by addition of two volumes of chilled 100% ethanol. A DNA spool was then formed upon mixing of the sample solution and was washed twice with chilled 70% ethanol. After

discarding as much of the 70% ethanol as possible, the spool was air-dried for approximately 30 min. 0.1 x TE buffer was then added to a DNA concentration of 1 to 4 mg/ml, and the spool was dissolved by gently rotating the sample tube.

Absorbencies by a UV spectrophotometer at wavelengths of 260 and 280 nm were measured to estimate the quantity and quality of the isolated DNA. As an additional means to quantify the isolated DNA, the copy number of the chosen *HPRT* target was measured by quantitative PCR, followed by CDCE. For this measurement, an aliquot of each genomic DNA sample, together with a known copy number of the 438-bp mutant internal standard, was restriction digested with *Bst*N I and *Dra* I. An aliquot of this digested sample was then amplified using the P3/P1 primer pair, followed by CDCE separation.

Based on these quantitative analyses, the $HPRT_{\text{Munich}}$ genomic DNA (see Section 3.5.1) was added to the *HPRT* wild-type genomic DNA at a desired fraction prior to restriction digestion by *Bst*N I and *Dra* I.

3.8.2 First restriction digestion: BstN I and Dra I

The reaction mixture of *Bst*N I and *Dra* I restriction enzymes (New England Biolabs, Beverly, MA) contained 10 mM Tris-HCl, 10 mM $MgCl_2$, 50 mM NaCl, and 1 mM dithiothreitol (pH 7.9 at 25°C). 10 mg/ml of BSA was added to a final concentration of 0.1 mg/ml, which was incubated at 37°C for 4 hr with an enzyme/DNA ratio of 1 U/μg.

The efficiency of *Bst*N I and *Dra* I restriction digestion was assessed by 2% agarose gel electrophoresis; human genomic DNA digested with *Bst*N I and *Dra* I multiple times under the specified reaction conditions was used as a reference to indicate complete digestion.

To each restriction-digested sample, the 438-bp mutant internal standard was added at a desired fraction prior to target isolation.

3.8.3 Target isolation

3.8.3.1 Procedure

Isolation of the chosen *HPRT* target was performed with each *Bst*N I and *Dra* I restriction-digested sample. This procedure is based on probe-target hybridization coupled with a biotin-streptavidin capture system, as Figure 18 illustrates. The procedure was performed as described in previous studies with modifications (Li-Sucholeiki and Thilly, 2000; Kim et al., 2001). These modifications made the procedure optimal for the chosen target and minimized sample exposure to heat during the procedure:

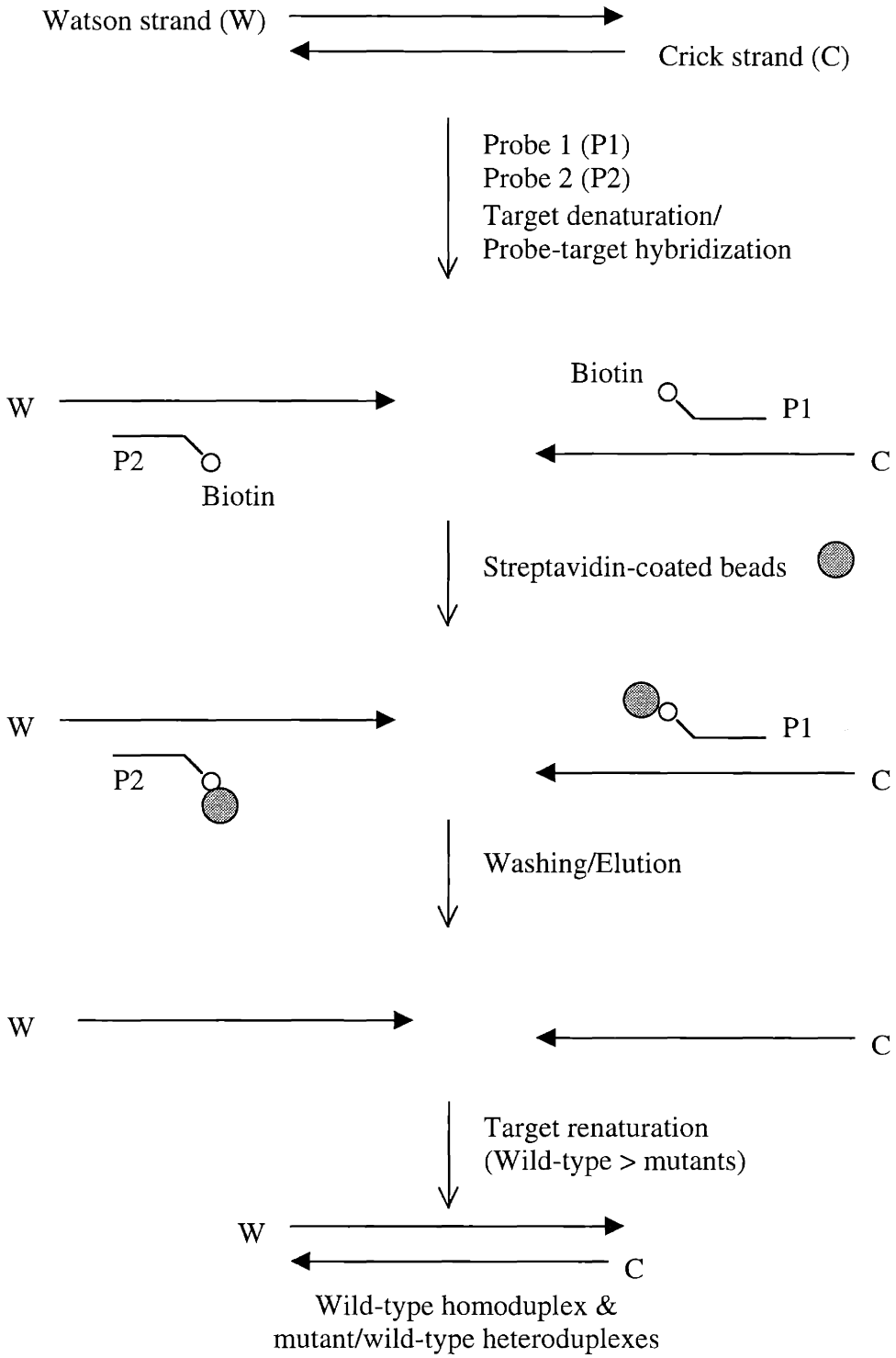
To each *Bst*N I/*Dra* I restriction-digested sample, Probe 1 (5' TGTGG AAGTT TAATG ACTAA GAGGT GTTTG 3') and Probe 2 (5' GACGT TCAGT CCTAC AGAAA TAAAA TCAGG 3') (Synthetic Genetics, San Diego, CA) were added at a probe/target molar ratio of 5 x 10^4 each (the probes were biotinylated at the 5' end by an 18-carbon spacer arm with a 6-carbon linker and were PAGE-purified prior to use) (Figure 15 in Section 3.6.1 indicates the probe sites in the target-embeded *Bst*N I and *Dra* I restriction fragment). This mixture was boiled for 1 min for target denaturation and immediately chilled in an ice bath for 10 min. The chilled sample was incubated at 60°C for 2 hr in 6 x SSPE (0.894 M NaCl, 60 mM sodium phosphate (pH 7.4), 6 mM EDTA) for probe-target hybridization.

Streptavidin-coated glass paramagnetic beads (CPG, Lincoln Park, NJ) were pre-washed twice with 3 x SSPE at a concentration of 10 mg/mL. The washing was performed first by magnetically separating the beads from the supernatant and second by removing the supernatant. The pre-washed beads were added to the probe-target hybridized sample at 0.4 mg of beads/10^8 copies of the target. This mixture was incubated at 50°C for 30 min to allow the binding of the beads to the probe-target hybrids. The beads were then washed with 6 x SSPE at a concentration of 5 mg/mL at room temperature to remove non-specific bead-DNA bindings. This washing step was repeated three times, once with 6 x SSPE and twice with 3 x SSPE. The final washing was performed with a buffer containing 50 mM NaCl, and 1 mM Tris-HCl (pH 7.4) to

Figure 18. Procedure of target isolation

Human genomic DNA is restriction digested to liberate a chosen target. This DNA is then denatured, mixed with biotin-labeled probes, and allowed to form probe-target hybrids. Streptavidin-coated beads are added to this mixture of hybrids, followed by incubation, to allow the binding of the beads to the probe-target hybrids. After washing, the target is eluted in single strands. Target isolation is followed by target renaturation during which mutant sequences formed heteroduplexes with the wild-type by mass action.

Source: Li-Sucholeiki and Thilly, 2000.


Watson strand (W)
Crick strand (C)
Probe 1 (P1)
Probe 2 (P2)
Target denaturation/
Probe-target hybridization
W
P2
Biotin
Biotin
P1
C
Streptavidin-coated beads
W
P2
P1
C
Washing/Elution
W
C
Target renaturation
(Wild-type > mutants)
W
C
Wild-type homoduplex &
mutant/wild-type heteroduplexes

minimize the salt, which could have been carried over by the beads. All the supernatant transferred up to this point was combined for estimating the target DNA yield.

The target was released from the probe-bound beads twice in dH_2O at 20 mg beads/mL at 70°C for 2 min. The released target was magnetically separated from the beads and transferred to a new tube, and its volume was condensed to about 10 µl by lyophilizer prior to target renaturation.

3.8.3.2 Methods of estimating yield and isolation efficiency

The yield and efficiency of target isolation were estimated by quantitative PCR, followed by CDCE. To estimate the target DNA yield, the wild-type copy number of the chosen *HPRT* target in both samples of the final target-isolated (C_E) and combined supernatant (C_S) was measured using the 438-bp mutant internal standard and the P3/P1 primer pair. The ratio of C_E to $C_E + C_S$ was then estimated to determine the target DNA yield ($[C_E \div (C_E + C_S)] \times 100$).

The isolation efficiency of the chosen *HPRT* target relative to non-target sequences in the pool of genomic DNA digest (*Bst*N I/*Dra* I) was estimated using a part of human mitochondrial DNA (DNA bp 10011-10215) as a non-target reference. CDCE separation of mutations in the chosen mitochondrial target from the wild-type has been demonstrated (Khrapko et al., 1994a; Khrapko et al., 1997a, b Coller et al., 1998; Marcelino et al., 1998). Thus, this sequence allows measurement of the sequence copy number by quantitative PCR, followed by CDCE. For this quantitative PCR, an internal standard containing a mutation in the chosen mitochondrial target (A -> G transition at bp 10072) was obtained from a previous study (Khrapko et al., 1997b); CW7 (5' ACCGT TAACT TCCAA TTAAC 3', bp 10,011-10031) and J3 (5'FITC GCGGG CGCAG GGAAA GAGGT 3', complementary to bp 10196-10215) primers (Synthetic Genetic, La Jolla, CA) were used. To estimate the isolation efficiency of the chosen *HPRT* target, the measured wild-type molar ratio of the chosen mitochondrial (M) to *HPRT* (T) targets in the initial sample ($M_I \div T_I = IR$) was compared to that of in the final target-isolated sample ($M_F \div T_F = FR$) (isolation efficiency = $IR \div FR$).

As a second means to estimate the isolation efficiency of the chosen *HPRT* target, absorbencies of desired samples were measured by a UV spectrophotometer at a wavelength of 260 nm. Sample DNA concentrations per a given *HPRT* target copy number before (A_I) and after (A_F) the procedure were then compared to estimate the isolation efficiency ($A_I \div A_F$).

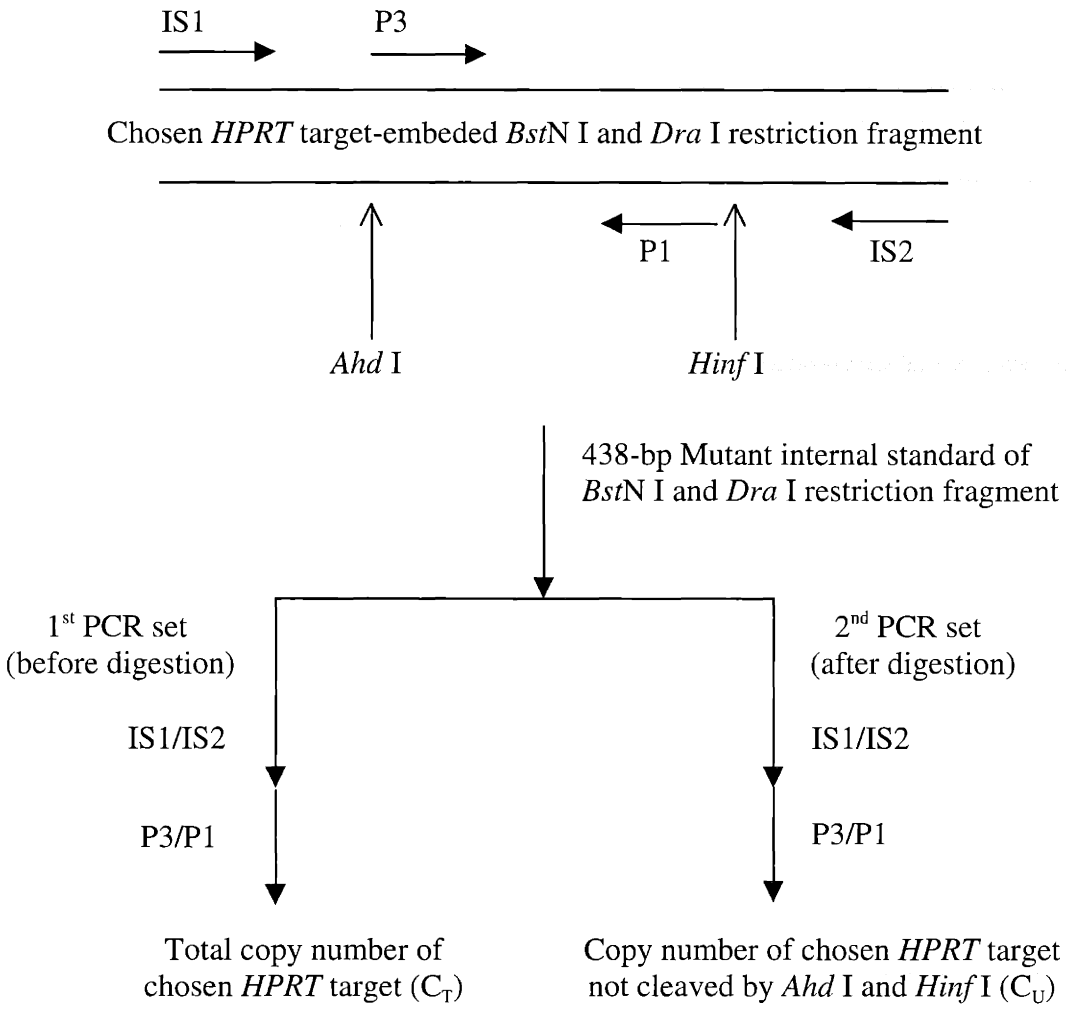
3.8.4 Target renaturation

To allow renaturation of the chosen *HPRT* target, isolated in single strands from the pool of genomic DNA digest, incubation was performed either at 55°C in a renaturation buffer (0.2 M NaCl, 10 mM Tris-HCl (pH 7.6), 2 mM EDTA) or at 28°C in the presence of 40% dimethyl sulfoxide (DMSO) in the renaturation buffer. Incubation was performed for 16 hr during which mutant sequences formed heteroduplexes with the wild-type by mass action. Each renatured sample was desalted by drop dialysis (see Section 3.7.4.2) prior to second restriction digestion.

3.8.5 Second restriction digestion

3.8.5.1 *Ahd* I/*Hinf* I: natural clamp

To liberate the chosen *HPRT* target with the natural clamp (cDNA bp 141-318) from the 438-bp restriction fragment of *BstN* I and *Dra* I, each target-renatured sample was restriction digested with *Ahd* I and *Hinf* I (New England Biolabs, Beverly, MA). The reaction mixture of *Ahd* I and *Hinf* I contained 50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, and 1 mM dithiothreitol (pH 7.9 at 25°C). 10 mg/ml BSA was added to a final concentration of 0.1 mg/ml, which was incubated at 37°C for 2 hr with 10 U of each restriction enzyme. The efficiency of *Ahd* I/*Hinf* I restriction digestion was estimated as Figure 19 illustrates.



Ahd I/*Hinf* I restriction digestion efficiency = $100 - [(C_U \div C_T) \times 100]$

Figure 19. Method of estimating *Ahd* I/*Hinf* I restriction digestion efficiency

The restriction digestion efficiency of *Ahd* I and *Hinf* I is determined by two sets of quantitative PCR, each followed by CDCE. The first set of PCR is performed before the digestion using the IS1/IS2 primer pair and using the 438-bp mutant internal standard. An aliquot of the IS1/IS2-amplified sample is amplified once again using the P3/P1 primer pair to generate an amplified sample suitable for CDCE separation, measuring the total copy number of the chosen *HPRT* target (C_T). These procedures are repeated for the second set of PCR after the digestion to measure the copy number of the chosen target not cleaved by *Ahd* I and *Hinf* I (C_U). The measured copy numbers, C_T and C_U, are then used to determine the digestion efficiency ($[1 - (C_u \div C_t)] \times 100$).

3.8.5.2 *Apo* I: ligated clamp

To generate the chosen *HPRT* target with the ligated clamp, *Apo* I restriction digestion was performed, followed by clamp ligation. The reaction mixture of *Apo* I contained 50 mM Tris-HCl (pH 7.9 at 25°C), 10 mM $MgCl_2$, 100 mM NaCl, and 1 mM dithiothreitol. 10 mg/ml BSA was added to a final concentration of 0.1 mg/ml, which was incubated with 4 to 10 U of *Apo* I (New England BioLabs, Beverly, MA) at 36°C for 2 hr.

The digestion efficiency of *Apo* I was estimated by measuring the copy number of the chosen target with the natural clamp before (C_B) and after the digestion (C_F). To measure the target copy number, quantitative PCR, followed by CDCE, was performed. For this PCR, the P3/P1 primer pair and the 198-bp mutant internal standard were used. C_F represents the target copy number left uncleaved, while C_B represents the total target copy number. The digestion efficiency was estimated as: $100 - [(C_F \div C_B) \times 100]$.

3.8.6 Clamp ligation

To generate the chosen *HPRT* target with the ligated clamp, each *Apo* I restriction digestion was followed by clamp ligation. A desired clamp copy number was added to each ligation reaction mixture containing 66 mM Tris-HCl (pH 7.5 at 20°C), 5 mM $MgCl_2$, 1 mM dithioerythritol, and 1 mM ATP. This mixture was incubated at 16°C for 16 hr with 1 U of T4 DNA ligase (Boehringer Mannheim, Indianapolis, IN).

Additional incubation at 37°C for 2 hr with 10 U of *Hinf* I restriction endonuclease was performed with each sample containing the chosen *HPRT* target that was isolated from the pool of genomic DNA digest (*Bst*N I/*Dra* I). The reaction mixture of *Hinf* I contained 10 mM Tris-HCl, 10 mM $MgCl_2$, 50 mM NaCl, and 1 mM dithiothreitol (pH 7.9 at 25°C).

To measure the copy number of the chosen target with the ligated clamp in each ligation sample, quantitative PCR, followed by CDCE, was performed. For this PCR, the GC1/P1 primer pair and the 169-bp mutant internal standard were used.

3.8.7 Pre-PCR mutant enrichment by CDCE

For pre-PCR mutant enrichment by CDCE, each desired sample of *Ahd* I/*Hinf* I restriction digestion or clamp ligation was dialyzed, condensed to about 2 to 4 μl, and loaded into a capillary. Either 320- or 540-μm id capillaries were used for those samples containing the purified wild-type DNA of PCR products; 540-μm id capillaries were used for those samples containing the wild-type DNA of human cells.

The CDCE conditions were optimized to enrich mutants in mutant/wild-type heteroduplexes since homoduplex mutant sequences in the original sample formed heteroduplexes with the wild-type by mass action after the sample, isolated for the chosen target in single strands, was allowed to renature. At an optimal CDCE separation temperature, determined using a 15 cm-long temperature-regulated zone, the wild-type homoduplex was well separated from mutant/wild-type heteroduplexes, with the different mutant/wild-type heteroduplexes having similar separation degrees. At this temperature, co-elution of the majority, if not all, of these mutant/wild-type heteroduplexes was allowed. Elution was performed in 20 μl of 4 x TBEB for 10 min at a constant current of 50 μA when using a 320-μm id capillary and in 20 μl of 8 x TBEB for 20 min at a constant current of 80 μA when using a 540-μm id capillary.

The efficiency of mutant enrichment by CDCE was estimated based on the ratio of the target wild-type copy number loaded into a capillary (C_T) to the wild-type copy number co-eluted with the mutant/wild-type heteroduplexes (C_{HE}): $C_T \div C_{HE}$; each wild-type copy number was measured by quantitative PCR, followed by CDCE. For this quantitative PCR, either the P3/P1 or GC1/P1 primer pair was used, depending on the clamp type desired, natural or ligated; either the 198- or 169-bp mutant internal standard was used.

As a second means to estimate the mutant enrichment efficiency, the level of a mutant internal standard in test samples before and after the procedure was compared. The test samples, each containing CDCE-purified target wild-type and a mutant internal standard at a fraction of 5×10^{-4}, were prepared. Mutant enrichment by CDCE was performed with these samples, and each eluted sample of mutant/wild-type heteroduplexes was PCR amplified, followed by CDCE. The ratio of the areas under the mutant (A_M) to wild-type (A_W) peaks was then estimated ($F_{MF} = A_M \div A_W$), with the ratio representing the mutant

fraction after the procedure. This ratio was divided by the initial mutant fraction (I_{MF}), 5 x 10^{-4}, to determine the mutant enrichment efficiency: $F_{MF} \div I_{MF}$.

3.8.8 Pre-PCR mutant enrichment by CE

Each desired CDCE-eluted sample of mutant/wild-type heteroduplexes was dialyzed, condensed to about 4 µl, and loaded into a capillary for further mutant enrichment by CE. This procedure, performed at room temperature, used 540-µm id capillaries at a constant current of 80 µA; the chosen *HPRT* target was eluted in 5 µl of 2 x TBEB for 2 min. The eluted DNA represented both wild-type and mutant sequences in double strands, which were separated from the wild-type in single strands, as well as from other wild-type containing species in CDCE-eluted samples. The efficiency of mutant enrichment by CE was estimated as described in Section 3.8.7.

3.8.9 Hifi-PCR and post-PCR mutant enrichment by CDCE

Each pre-PCR mutant-enriched sample was PCR amplified using either the P3/P1 or GC1/P1 primer pair, depending on the clamp type desired, natural or ligated. All the primers were converted into products by applying enough PCR cycles during which mutant sequences formed heteroduplexes with the wild-type by mass action.

Each PCR-amplified sample was then loaded into a capillary for further mutant enrichment by CDCE (post-PCR mutant enrichment), performed using a 75-µm id capillary at a constant current of 9 µA. CDCE-separated mutants were eluted in 10 µl of 1 x TBEB for 4 min. The efficiency of mutant enrichment was estimated as described in Section 3.8.7.

3.8.10 Mutational analysis by CDCE

PCR was performed with 1 to 2 µl of each post-PCR mutant-enriched sample. For this PCR, either the P3/P1 or GC1/P1 primer pair was used, depending on the clamp type desired, natural or ligated. All the primers were converted into products by applying

enough PCR cycles during which mutant sequences formed heteroduplexes with the wild-type by mass action.

The amplified sample was then separated by CDCE using a 75-μm id capillary at a constant current of 9 μA and using a 15 to 20 cm-long temperature-regulated zone. At an optimal separation temperature, mutant and wild-type sequences were separated from one another, allowing for quantitative analysis of these separated mutants. At this stage, individual mutants were purified by CDCE for sequence identification.

3.8.11 Purification of individual mutants by CDCE

To purify individual mutant sequences, CDCE separation was performed as described in Section 3.8.10. This procedure was followed by electro-elution of each separated mutant in 10 μl of 1 x TBEB. An aliquot of this eluted sample (1 to 2 μl) was then PCR amplified for another round of CDCE separation and elution. PCR, followed by CDCE separation and elution, was repeated until a final-eluted sample had close to 100% homogeneity in sequence. An aliquot of the final-eluted sample was PCR amplified to be used as a template for DNA sequencing (MIT, Cambridge, MA).

3.9 Second target sequence: *p53* exon 7

A part of the human *p53* gene (cDNA bp 673-782) was chosen as a second target sequence to demonstrate clamp ligation and CDCE separation. Figure 20 illustrates the melting profile of the chosen target with the ligated clamp.

Figure 21 illustrates the characteristics of the clamp prepared for the restriction end generated by *Avr* II restriction endonuclease. *Avr* II has a 6-base recognition sequence and cleaves after the first base pair, generating a 5'-CTAG-3' cohesive end, as Figure 22 illustrates.

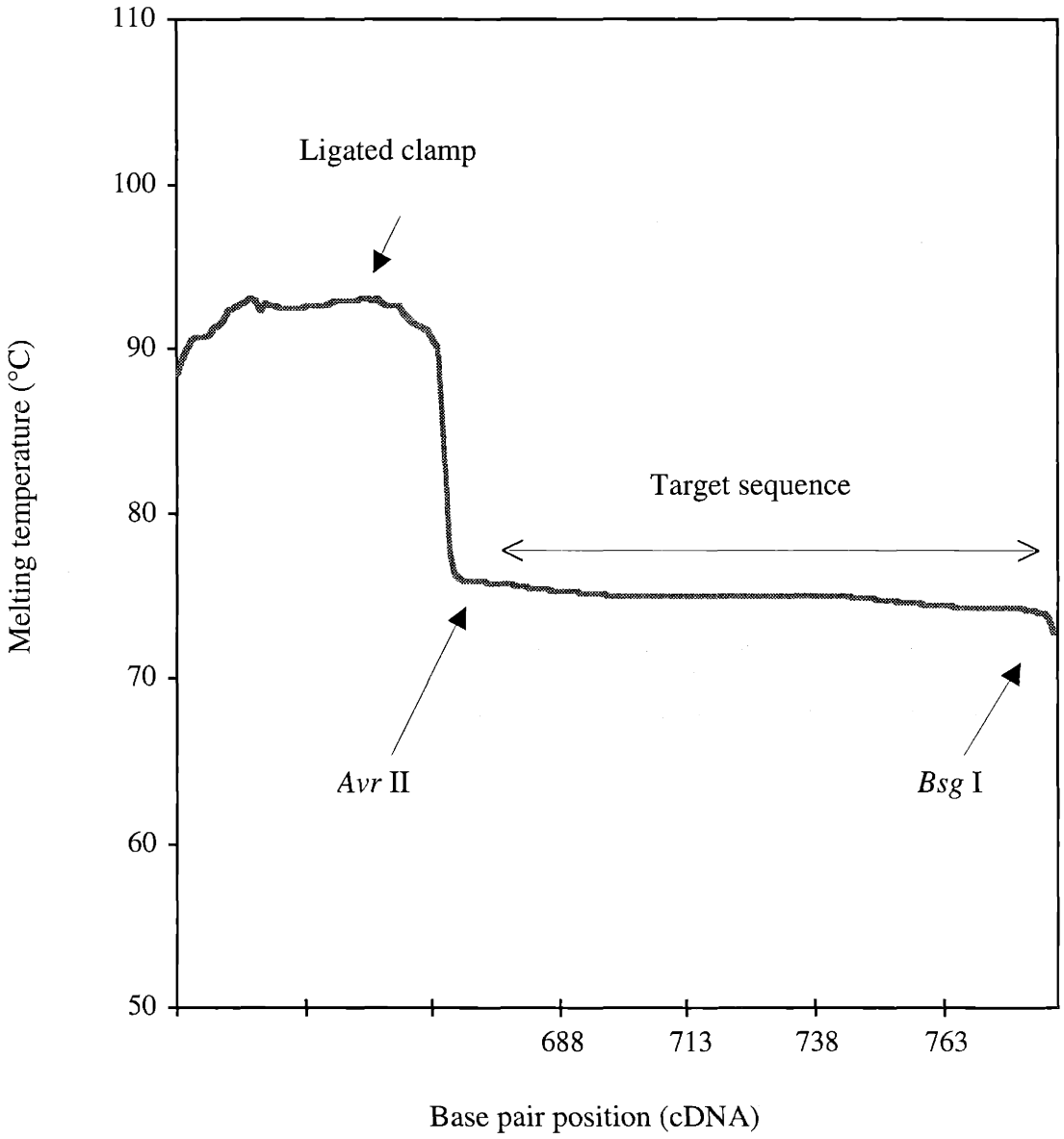


Figure 20. Melting profile of human *p53* exon 7 with ligated clamp

Restriction digestion by *Avr* II and *Bsg* I liberates the chosen *p53* target from genomic DNA. The prepared clamp for the *Avr* II restriction end can be ligated to the *Avr* II target restriction end, making the digested target suitable for CDCE separation.

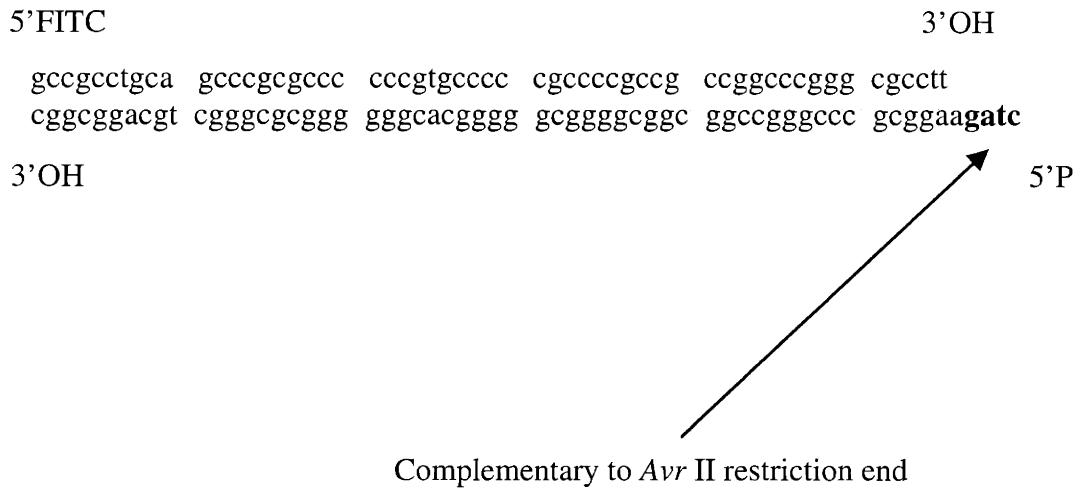
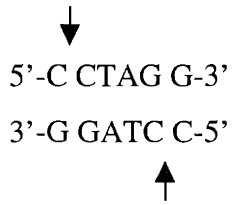


Figure 21. Characteristics of clamp for *Avr* II restriction end

The sequence of the clamp is from a previous study (Cariello et al., 1990). This clamp has an end that is complementary to the *Avr* II restriction end. This end contains 5'-phosphoryl and 3'-hydroxyl residues, allowing clamp ligation to the *Avr* II-generated restriction end.

Figure 22. Recognition sequence of *Avr* II restriction endonuclease



To prepare a substrate containing the chosen *p53* target for *Avr* II restriction digestion, a 266-bp DNA fragment was PCR amplified from wild-type genomic DNA using the HH and AP primer pair. The amplified sequence consisted of exon 7 of the human *p53* gene (cDNA bp 673-782), an *Avr* II restriction recognition site, and flanking intron sequences. The amplified sample was ethanol precipitated and restriction digested with *Avr* II in a reaction mixture containing 10 mM Tris-HCl, 10 mM $MgCl_2$, 50 mM NaCl, and 1 mM dithiothreitol (pH 7.9 at 25°C). This sample mixture was incubated at 36°C for 2 hr prior to clamp ligation.

4 RESULTS AND DISCUSSION

4.1 Melting profile results of human gene sequences

Pre-PCR mutant enrichment is a key procedure in target isolation/CDCE/hifi-PCR, suitable for analysis of rare point mutations at fractions as low as 10^{-6}. Pre-PCR mutant enrichment is based on separation of mutant from wild-type sequences by CDCE and requires that a target sequence have a neighboring natural clamp. This requirement limits the application of the established method to DNA sequences with a natural clamp, represented by X.

To estimate the percentage of X in the human genome, the melting profiles of several human gene sequences were constructed and analyzed. These sequences, totaling 500-kb, belong to five tumor suppressor and two disease-related genes. Figure 23 shows an example of the constructed melting profile belonging to the first 3000-bp of the human *p53* gene. As Table 11 summarizes, 9% of the sequences scanned are X, suitable for target isolation/CDCE/hifi-PCR.

For rare point mutational analysis of a larger target-pool size, a means of attaching a clamp to any 100-bp DNA sequence of interest was devised, as Section 3.2.1 describes. In theory, this clamp attachment technique, when combined with target isolation/CDCE/hifi-PCR, allows analysis of point mutations at fractions as low as 10^{-6} in those sequences without a natural clamp, represented by Y. Such sequences represent 89% of the total gene sequences scanned (see Table 11).

Two percent of the sequences scanned, represented by Z (see Table 11), cannot be studied since separation of point mutations in sequences with melting temperatures greater than 80°C is not yet well established. The use of DGGE, combined with CDGE (Wu et al., 1999), sodium bisulphite-treated DNA (Guldberg et al., 1998), and bipolar clamping (Gille et al., 1998) have been shown to improve DNA separation in such high melting temperature sequences. However, these approaches have not been demonstrated with CDCE. In addition, the latter two require PCR prior to DNA separation, which is not suitable for detection of rare mutations at fractions below 2×10^{-4} (see Section 2.5.3.1).

Figure 23. Melting profile of human *p53* gene (first 3000-bp)

Each X represents a DNA domain (≈100-bp) juxtaposed with a domain (≈100-bp) of a higher melting temperature, or a natural clamp. The melting temperatures of these two domains differ by at least 5°C. Each Y represents a domain of any length without a natural clamp. The melting temperature of each X or Y is below 80°C. Domains of any length with melting temperatures greater than 80°C are marked as Z.

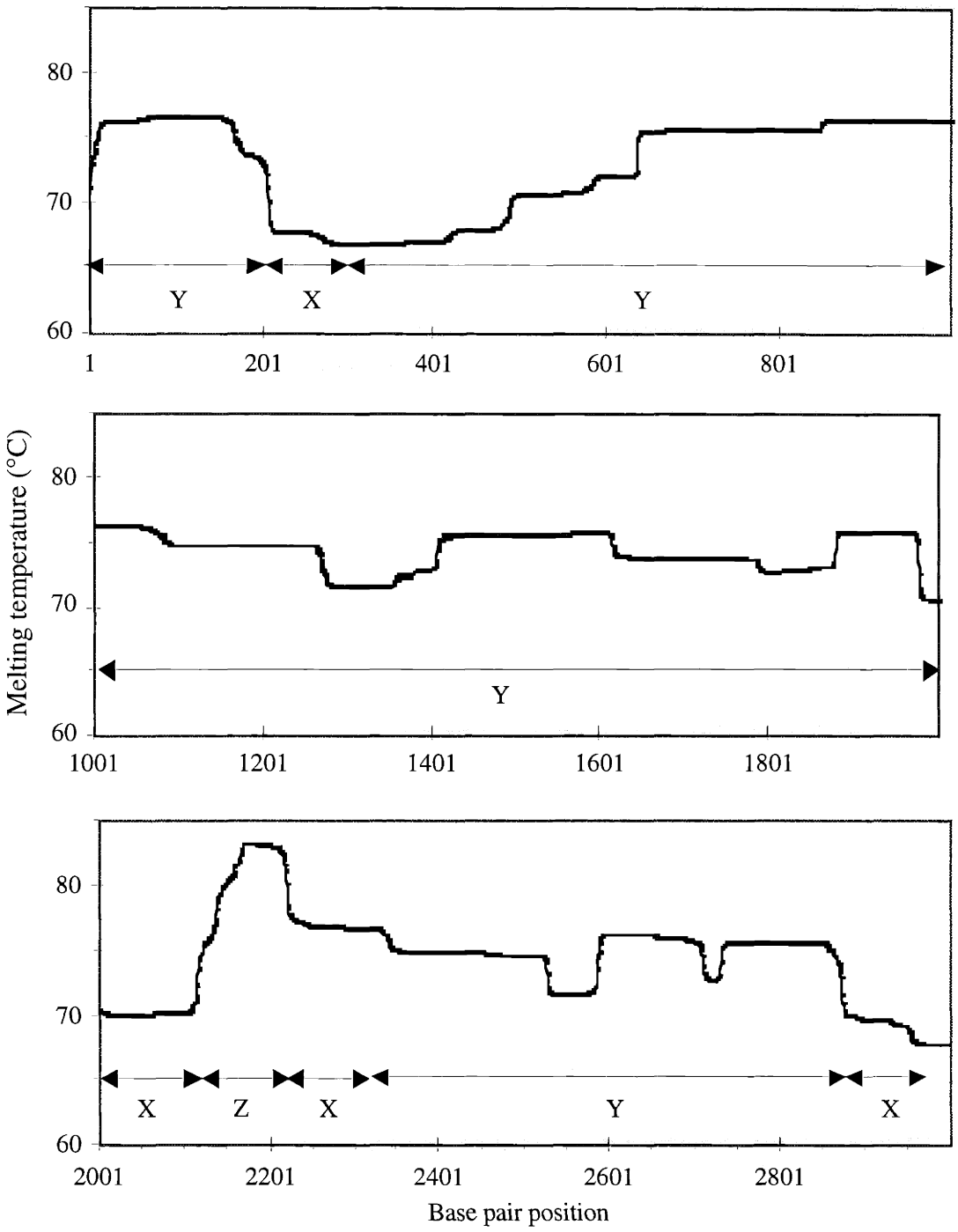

80
70
60
Y
X
Y
1
201
401
601
801
Melting temperature (°C)
80
70
60
Y
1001
1201
1401
1601
1801
80
70
60
X
Z
X
Y
X
2001
2201
2401
2601
2801
Base pair position

Table 11. Melting profile results: entire DNA sequences of selected human genes

Gene (accession number[a])	Size	X[b]	Y[c]	Z[d]
BRCA1 (L78833)	85-kb	11%	88%	1%
NF2 (Y18000)	126-kb	7%	90%	3%
p53 (X54156)	20-kb	10%	90%	<1%
RB (L11910)	180-kb	7%	92%	1%
VHL (AF010238)	15-kb	11%	84%	5%
HPRT (M26434)	57-kb	11%	88%	1%
Mitochondrial DNA (V00662)	16-kb	8%	92%	<1%
% Total	-	9%	89%	2%
Total number of bases	500-kb	45-kb	445-kb	10-kb

[a] Each accession number is from Benson et al., 2000.
[b] Percentage of each gene sequence with a natural clamp, suitable for target isolation/CDCE/hifi-PCR.
[c] Percentage of each gene sequence without a natural clamp, suitable for combined method of clamp ligation and target isolation/CDCE/hifi-PCR.
[d] Percentage of each gene sequence not suitable for CDCE-based mutation detection methods due to its high melting temperature greater than 80°C.

Table 12 summarizes the melting profile results of only the coding regions of six tumor suppressor genes in which many mutations have been found in human cancer. These results show that clamp ligation, coupled with target isolation/CDCE/hifi-PCR, is applicable to the majority, if not all, of each gene sequence scanned. Without clamp ligation, target isolation/CDCE/hifi-PCR alone can be applied to between 0 and 6% of the sequences scanned.

4.2 Clamp attachment/ligation

4.2.1 Estimated number of chosen restriction ends per human cell

The proposed method of clamp attachment ligates a clamp to the restriction ends of a chosen target in a genomic DNA digest by mass action. Thus, the number of chosen restriction ends per diploid human cell was theoretically estimated to determine the clamp copy number necessary for this mass action. Table 13 summarizes the estimated number of ends per cell generated by a chosen restriction endonuclease with a 4-, 5-, or 6-base recognition site.

Isolation of a desired target from a pool of genomic DNA digest precedes clamp ligation in the proposed point mutation detection method (see Section 3.3). Thus, the number of chosen restriction ends per cell after target isolation was estimated. Using the reported efficiency of target isolation, 10^4-fold (Li-Sucholeiki and Thilly, 2000), Table 14 summarizes the estimated number of ends generated by a chosen restriction endonuclease with a 4-, 5-, or 6-base recognition site. This number is per diploid human cell or per 10^8 diploid human cells. 10^8 cells is the sample size necessary for detecting mutations at fractions as low as 10^{-6} with statistical significance. If, for example, a 10-fold higher clamp copy number is necessary for ligating every restriction end of genomic DNA by mass action, this number can be estimated by multiplying each estimated number of ends per 10^8 diploid human cells in Table 14 by 10.

Clamp ligation is not specific to target restriction ends. Out of the estimated number of chosen restriction ends per cell, only two belong to a desired target of a single-copy nuclear gene. Table 14 summarizes the ratio of target to non-target ends generated by a

Table 12. Melting profile results: coding regions of selected human tumor suppressor genes

Gene (accession number[a])	Size	X[c]	Y[d]	Z[e]
APC (M74088)[b]	6578-bp	6%	94%	0%
*BRCA*1 (L78833)	5572-bp	0%	100%	0%
*NF*2 (Y18000)	2069-bb	4%	82%	14%
p53 (X54156)	1182-bp	0%	100%	0%
RB (L11910)	2787-bp	2%	93%	5%
VHL (AF010238)	642-bp	0%	47%	53%

[a] Each accession number is from Benson et al., 2000.

[b] To determine X for coding regions, flanking intron sequences are necessary. Because these sequences had not been discovered for the *APC* gene at the time of the melting profile construction, X was determined for exon 15 only. Exon 15 is comprised of about 6500-bp, representing 77% of the coding region.

[c] Percentage of each gene sequence with a natural clamp, suitable for target isolation/CDCE/hifi-PCR.

[d] Percentage of each gene sequence without a natural clamp, suitable for combined method of clamp ligation and target isolation/CDCE/hifi-PCR.

[e] Percentage of each gene sequence not suitable for CDCE-based mutation detection methods due to its high melting temperature greater than 80°C.

Table 13. Estimated numbers of chosen restriction ends

Size of restriction recognition site (number of bases)	Number of chosen restriction ends per diploid human cell*
4	4.69×10^7
5	1.17×10^7
6	2.93×10^6

* Estimated using Equation 3: $[(6 \times 10^9) \div 4^X] \times 2$, where 6×10^9 is the number of bases per diploid human genome, 4^X is the cleavage frequency of a chosen restriction enzyme in bases (X represents the size of the restriction recognition site), and 2 is the number of restriction ends generated per cleavage.

Table 14. Estimated numbers of chosen restriction ends and ratios of target to non-target restriction ends after target isolation

Size of restriction recognition site (number of bases)	Number of chosen restriction ends per diploid human cell (#/dhc)[a]	Number of chosen restriction ends per 10^8 diploid human cells[b]	Ratio of target to non-target restriction ends[c]
4	4.69×10^3	4.69×10^{11}	$1 : 2.3 \times 10^3$
5	1.17×10^3	1.17×10^{11}	$1 : 5.9 \times 10^2$
6	2.93×10^2	2.93×10^{10}	$1 : 1.5 \times 10^2$

[a] Estimated by dividing Equation 3 by a target isolation efficiency of 10^4-fold. Equation 3: $[(6 \times 10^9) \div 4^X] \times 2$, where 6×10^9 is the number of bases per diploid human genome, 4^X is the cleavage frequency of a chosen restriction enzyme in bases (X represents the size of the restriction recognition site), and 2 is the number of restriction ends generated per cleavage.
[b] Estimated by multiplying #/dhc by 10^8.
[c] Estimated by dividing #/dhc by 2, which represents the number of target restriction ends per cell after target isolation.

chosen restriction endonuclease with a 4-, 5-, or 6-base recognition site after target isolation. Target from non-target restriction fragments with the same ligated clamp is expected to be distinguished by the procedures following clamp ligation, pre-PCR mutant enrichment and PCR, in the proposed point mutation detection method. Thus, the presence of a 10^2-10^3-fold higher copy number of non-target restriction ends in a ligation mixture is not expected to interfere with mutational analysis by the proposed method.

4.2.2 Optimal ligation reaction conditions

Table 15 summarizes the optimal ligation reaction conditions empirically determined, as well as the conditions suggested by the manufacturer of T4 DNA ligase, Boehringer Mannheim (Indianapolis, IN). When comparing these two sets of conditions, two variables should be noted.

The first variable is incubation temperature. The temperature suggested by the manufacturer is 4°C. However, rather a big range, 4 to 22°C, was empirically determined to generate the maximum ligation efficiency of the substrate, the *Apo* I cohesive end. When the temperature was increased to 36°C, the efficiency was reduced by 10% of the maximum efficiency observed.

A big range of optimal incubation temperatures have been reported using as a substrate the *Eco*R I cohesive end (5'AATT3'). This restriction end is the same as the *Apo* I. These temperatures are 4 °C (Ferretti and Sgaramella, 1981), 10 to 15°C (Dugaiczyk et al., 1975), 15 to 25°C (Ehrlich et al., 1977), and 25°C (Sgaramella and Ehrlich, 1978). The melting temperature determined for the *Eco*R I cohesive end is 6°C (Mertz and Davis, 1972).

Substrate concentration is the second variable of ligation reaction conditions to be noted. While the concentration suggested by the manufacturer is 0.015 μM, a higher concentration range, 0.17 to 0.55 μM, was shown to be necessary to generate the maximum ligation efficiency of the substrate. When the concentration was lowered by 10-fold (17 nM), the ligation efficiency was reduced by 10% of the maximum efficiency observed. When lowered by another 10-fold (1.7 nM), the efficiency was further reduced to 50%. Therefore, using a substrate concentration of 0.17 μM or higher was determined

Table 15. Optimal ligation reaction conditions

The optimal ligation reaction conditions were determined empirically by varying incubation temperatures from 4 to 36°C, incubation duration from 1 to 16 hr, units of T4 DNA ligase from 1 to 10, and substrate concentrations from 10^8 to 3.3 x 10^{11}/µl. The prepared clamp for the *Apo* I restriction end was used as a substrate. The criterion used for determining the optimal ligation reaction conditions was the conditions that generated the maximum efficiency of clamp self-ligation. The efficiency of clamp self-ligation for each reaction variable tested was estimated by CE, which separated the self-ligated clamp (C_L, 100-bp) from the un-ligated clamp (C_U, 48-bp): $[C_L \div (C_L + C_U)] \times 100$.

Reaction condition variable	Determined empirically	Suggested by Boehringer Mannheim (Indianapolis, IN)
Incubation temperature	4-22°C	4°C
Incubation duration	16 hr	16 hr
Unit of T4 DNA ligase (1 U = 3.5 x 10^{12} molecules)	1 U	1 U
Substrate concentration	1–3.3 x 10^{11} molecules/µl (0.17-0.55 µM)	9 x 10^9 molecules/µl (0.015 µM)

as an important factor in achieving the maximum ligation efficiency.

4.2.3 Optimal restriction digestion reaction conditions

Certain reaction conditions of restriction digestion limited clamp ligation efficiency, clamp ligating to the digested DNA. The different *Apo* I restriction digestion reaction conditions tested resulted in a similar digestion efficiency of about 95%. However, these test conditions resulted in various clamp ligation efficiencies ranging from 21.5 to 99%.

Table 16 summarizes these results. The restriction digestion incubation temperature recommended by the manufacturer of *Apo* I, New England Biolabs (Beverly, MA), is 50°C. However, the use of such a high temperature should be avoided when planning high efficiency clamp ligation. Lowering the temperature to 37°C resulted in clamp ligation efficiencies higher than those observed at 50°C. In addition, shortening restriction digestion incubation duration from 16 to 4 hr increased the ligation efficiency at both incubation temperatures. Thus, to achieve the maximum clamp ligation efficiency, restriction digestion must be performed at 37°C for a short period of time, such as 4 hr.

A reason behind the inefficiency in clamp ligation with those samples restrict digested at 50°C is yet to be determined. One possible explanation is that such a high temperature degrades the generated restriction ends, lowering the clamp ligation efficiency.

4.2.4 Demonstration of high efficiency clamp ligation

Figure 24 illustrates high efficiency clamp ligation to *Apo* I restriction ends by mass action. In this illustration, the prepared clamp is ligated to a restriction fragment copy number of 10^{10}, 10^{11}, or 10^{12}. This copy number was selected based on the number estimated for a chosen restriction endonuclease with a 4-, 5-, or 6-base recognition site, as Table 14 in Section 4.2.1 summarizes. This selection represents a sample size of 10^8 diploid human cells after target isolation with an efficiency of 10^4-fold.

By using a 10-fold higher clamp copy number, a clamp ligation efficiency of above 95% was estimated for each selected copy number of the restriction ends. This efficiency

Table 16. Reaction conditions of *Apo* I restriction digestion vs. efficiency of clamp ligation

A chosen target was PCR amplified using the P3/IS2 primer pair, ethanol precipitated, and restriction digested under different test reaction condition variables. Using a 10^{12} clamp copy number, clamp ligation was performed with each test sample of 10^{11} copies of the digested DNA. The clamp ligation efficiency was estimated by CE, which differentiated the clamp-ligated (L) from un-ligated (UL) restriction fragments: [L ÷ (L + UL)] x 100.

Test reaction condition variable			Clamp ligation efficiency (%) (mean ± SD, n=2)
Incubation temperature	Enzyme/substrate ratio (U of *Apo* I per µg of substrate, U/µg)	Incubation duration	
37°C	1 (2 U/µg)[a]	4 hr	99.0 ± 0
		16 hr	94.0 ± 1.4
50°C[b]	1 (1U /µg)	4 hr	68.5 ± 7.8
		16 hr	21.5 ± 9.2

[a] This enzyme/DNA ratio is equivalent to 1 U/µg since *Apo* I results in 50% activity at 37°C (New England BioLabs, 2000-2001).

[b] Optimal temperature for *Apo* I, recommended by the manufacturer, New England Biolabs (Beverly, MA).

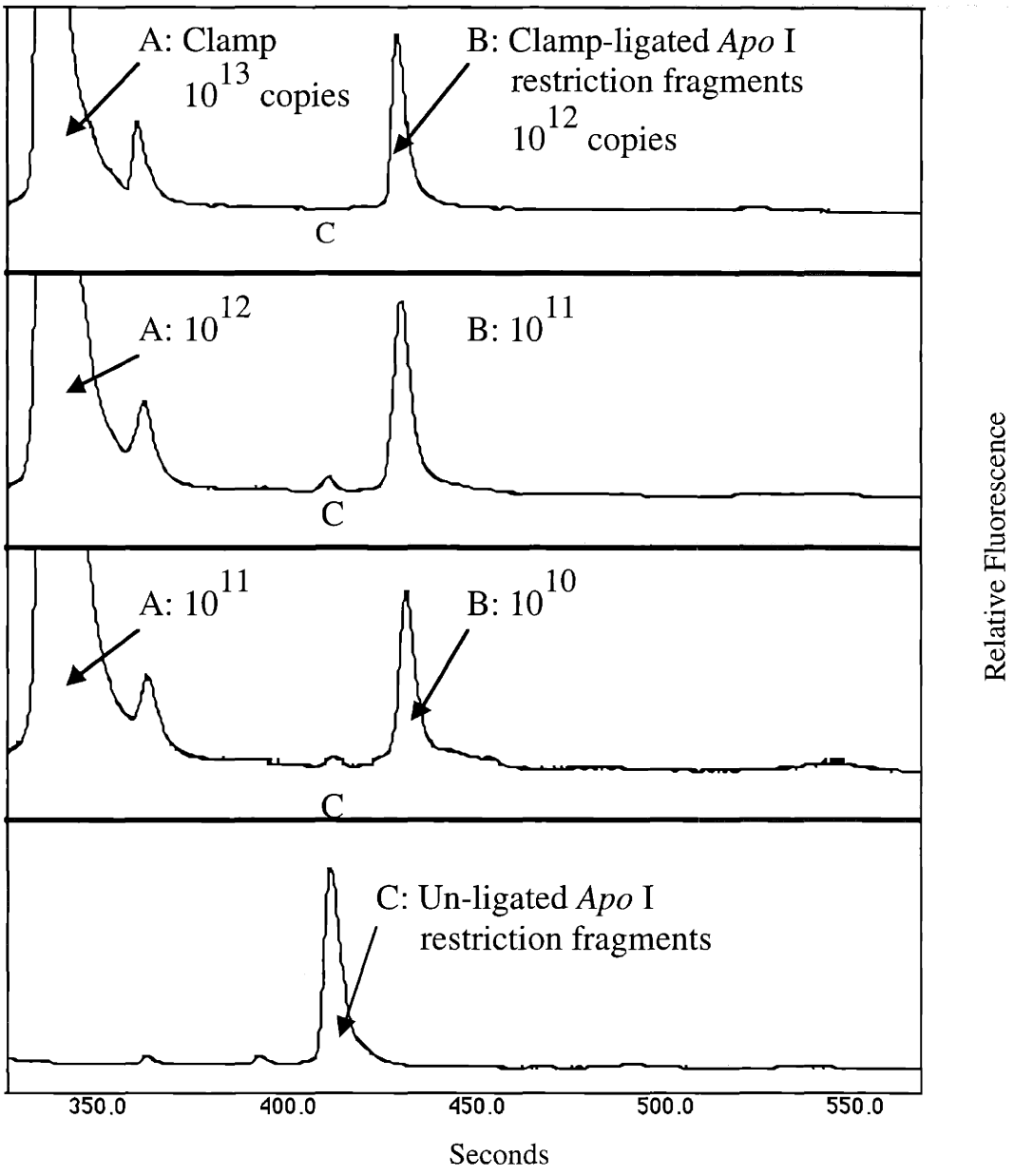


Figure 24. High efficiency clamp ligation*

A chosen target was PCR amplified using the P3/IS2 primer pair, ethanol precipitated, and restriction digested by *Apo* I. Using a 10-fold higher clamp copy number, clamp ligation was performed with the digested sample containing a known copy number of the digested DNA. The clamp ligation efficiency was estimated by CE, which differentiated the clamp-ligated (L) from un-ligated (UL) restriction fragments: [L ÷ (L + UL)] x 100.

* Above 95% clamp ligation efficiency was estimated for each sample.

is thought to be the maximum since increasing the clamp copy number by 10-fold did not increase the ligation efficiency. When the same clamp copy number as the restriction end copy number was used, the ligation efficiency was reduced to 85%. An efficiency of 30% was observed when a 10-fold lower clamp copy number was used.

Section 4.2.2 discusses substrate concentration as an important factor in achieving the maximum ligation efficiency. However, when mass action was coupled with ligation, the substrate concentrations used (3.3×10^9 to 3.3×10^{11} molecules/μl for clamp and 3.3×10^8 to 3.3×10^{10} molecules/μl for *Apo* I restriction end) did not affect the clamp ligation efficiency.

In brief, a clamp ligation efficiency of above 95% by mass action was demonstrated. For this demonstration, a 10-fold higher clamp copy number was used in a ligation reaction mixture containing different restriction end numbers, ranging from 10^{10} to 10^{12}.

4.3 Test target sequence: *HPRT* exon3

4.3.1 Demonstration of CDCE separation: ligated vs. natural clamp

Figures 25 and 26 illustrate CDCE separation of the low T_m mutation in the chosen *HPRT* target from the wild-type. Figure 25 illustrates the separation using the natural clamp, and Figure 26, using the ligated clamp.

These clamps generated a similar separation degree of the mutant from wild-type sequences by CDCE at different temperatures below 71.5°C. The same was true for the high T_m mutant (A -> G transition at *HPRT* cDNA bp 316). However, the target with the natural clamp completely converted to single strands (SS) at temperatures at or above 71.5°C. This conversion happened because the natural clamp lost its ability to hold the two DNA strands together. On the other hand, the ligated clamp was able to separate the mutant from wild-type sequences at temperatures below 85°C. At or above 85°C, the ligated clamp also lost its ability to hold the two strands together. Complete strand dissociation of the same target with the ligated clamp started to appear at a higher temperature than with the natural clamp, as the melting profiles predict (see Figure 15 in Section 3.4).

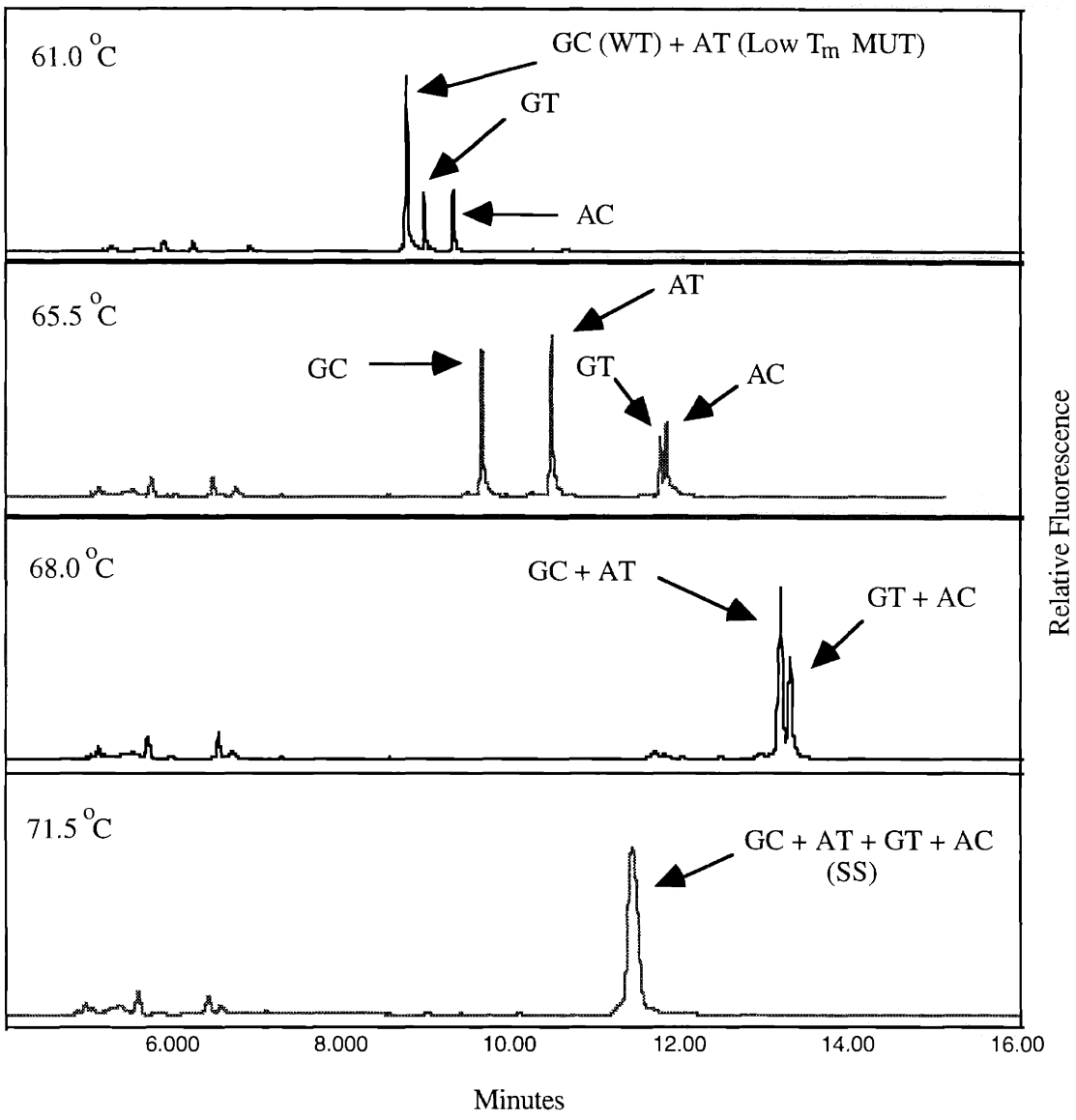


Figure 25. CDCE separation of low T_m mutant from wild-type sequences: *HPRT* exon 3 with natural clamp

Using the P3/P1 primer pair, PCR was performed with a mixture containing the wild-type and the mutant internal standard (low T_m mutant, G -> A transition at *HPRT* cDNA bp 309) at an equal molar ratio. During this PCR, mutant/wild-type heteroduplexes, in addition to the homoduplex of each kind, were formed. The PCR-amplified sample was separated by CDCE, performed using a 6 cm-long temperature-regulated zone of desired temperatures.

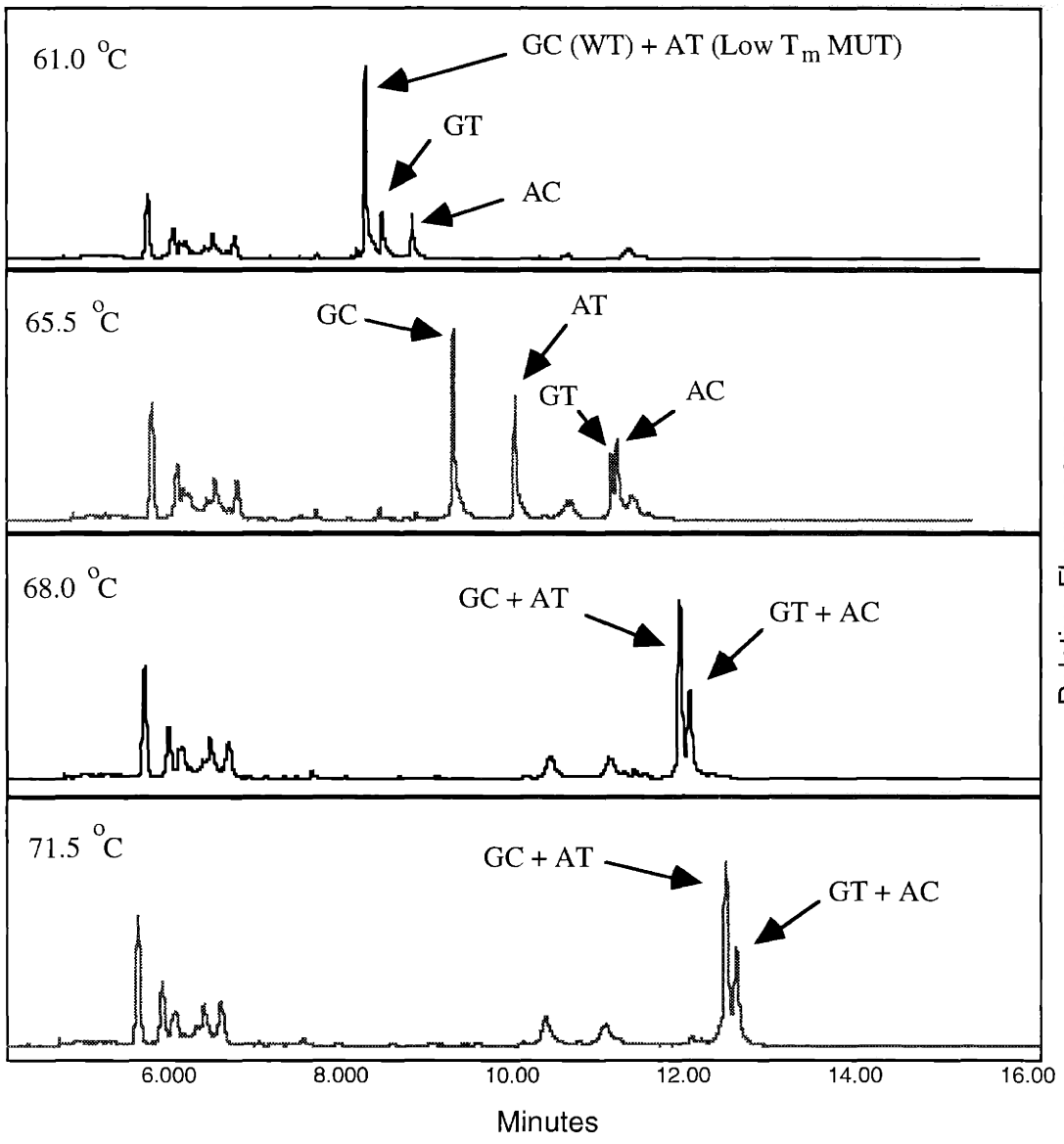


Figure 26. CDCE separation of low T_m mutant from wild-type sequences: *HPRT* exon 3 with ligated clamp

Using the GC1/P1 primer pair, PCR was performed with a mixture containing the wild-type and the mutant internal standard (low T_m mutant, G -> A transition at *HPRT* cDNA bp 309) at an equal molar ratio. During this PCR, mutant/wild-type heteroduplexes, in addition to the homoduplex of each kind, were formed. The PCR-amplified sample was separated by CDCE, performed using a 6 cm-long temperature-regulated zone of desired temperatures.

CDCE separation results demonstrate the ability of the ligated clamp to separate point mutations in the chosen target from the wild-type. These results also demonstrate that a similar degree of CDCE separation is achievable at temperatures below 71.5°C for the chosen target with either the natural or ligated clamp.

4.3.2 Fidelity of Pfu DNA polymerase

As Section 2.5.3.1 discusses, analysis of rare mutations requires a polymerase with high fidelity. For this reason, *Pfu* DNA polymerase was chosen for PCR performed in this study.

The reported fidelity of *Pfu* is about 10^{-6}/bp/doubling (Lundberg et al., 1991; Barnes, 1994; Flaman et al., 1994; Cline et al., 1996; Smith and Modrich, 1996; André et al., 1997; Parsons and Heflich, 1998; Li-Sucholeiki and Thilly, 2000). When the fidelity was estimated using the chosen *HPRT* sequence as a target, about 2×10^{-6}/bp/doubling was observed.

4.3.3 HPRT pseudogenes

Once a human target gene sequence is chosen for mutational analysis, target gene-like sequences, pseudogenes, must be examined. Pseudogenes are linked sequences, which consist of considerable sequence homology with functional genes. However, these pseudogenes are not functional since they contain mutations that inactivate their transcription and translation (Wilde, 1986; Venter et al., 2001). Thus, a mutation assay must be planned carefully so that pseudogenes do not confound analysis of the chosen gene sequence.

Pseudogenes are divided into two groups. The pseudogenes in the first group are linked to parent genes and therefore are thought to be the result of gene duplication (Wilde, 1986; Mighell et al., 2000). The pseudogenes in the second group are dispersed to different chromosomes in the genome (Wilde, 1986; Mighell et al., 2000). The major characteristic of the pseudogenes in the second group is that they are "intron-less," or "processed," and therefore are thought to arise through retrotransposition of processed

mRNA transcripts into the genome (Vanin, 1985; Wilde, 1986; Mighell et al., 2000). The second group of pseudogenes appears to outnumber the first (Vanin, 1985; Wilde, 1986).

One study estimates 2909 processed pseudogenes in the human genome (Venter et al., 2001) while another estimates about a 10-fold higher number, 23,000 to 33, 000 (Goncalves et al., 2000). The first study is based on sequence analysis of the entire human genome, and the second, of human chromosome 22 (Dunham et al., 1999), suggesting that the estimate in the first study is more accurate than that in the second. The number of pseudogenes per gene varies from one to multiple for those genes with a pseudogene(s) (Venter et al., 2001).

Analysis of transcripts that give rise to processed pseudogenes has shown that these transcripts have shorter average lengths compared to those with no pseudogene (Venter et al., 2001). While no significant difference was observed in the overall GC content in one study (Venter et al., 2001), another study observed low GC content for those genes with a processed pseudogene(s) (Goncalves et al., 2000). Additional characteristics of these genes are that they are widely expressed and highly conserved (Goncalves et al., 2000).

The human *HPRT* gene was chosen as a test target gene in this study. This gene has been demonstrated to have *HPRT*-like sequences, thought to be processed pseudogenes (Patel et al., 1983; Gennett and Thilly, 1988). *HPRT* exon-like sequences were also found using the Basic Local Alignment Search Tool (BLAST). Figure 27 illustrates these *HPRT*-like sequences in the human genome. Two additional findings were performed using GenBank: accession numbers U10112 (Sellner and Turbett, 1996) (493-bp comprised of human *HPRT* exon 2- and 3-like sequences) and U43684 (276-bp comprised of human *HPRT* exon 3- and 4-like sequences).

To prevent the *HPRT* pseudogenes in the human genome from confounding mutational analysis of the chosen gene sequence, this analysis was planned carefully. As part of this planning, the primers and probes that do not share the sequence with these pseudogenes were selected.

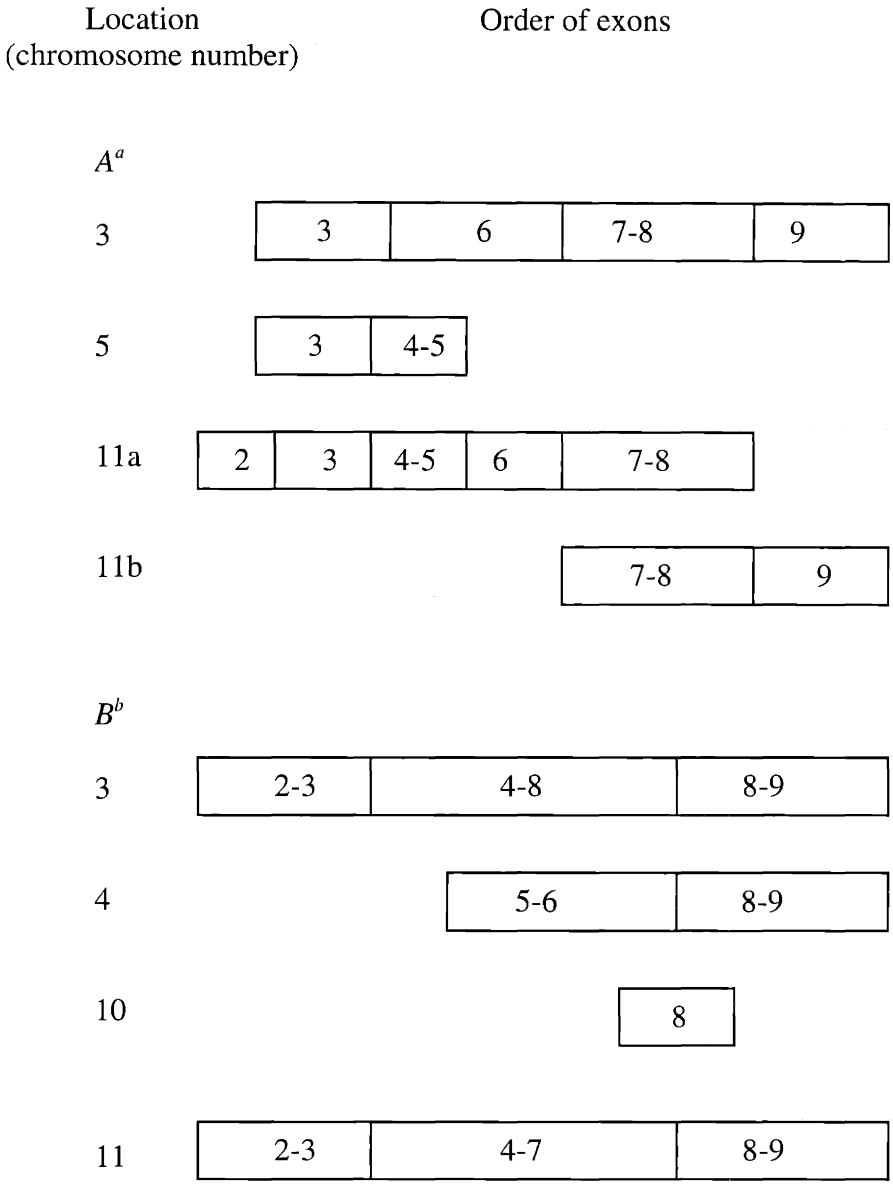


Figure 27. *HPRT* pseudogenes in the human genome

[a] Adapted from Gennett, 1988. Southern blotting and *HPRT* exon-specific probes were used in this study.

[b] Based on the BLAST search of the human genome: > 85% sequence homology for 75% of each exon length was used as a search criterion. BLAST is a tool for searching protein and DNA databases for sequence similarities (Altschul et al., 1990; Altschul et al., 1997).

4.4 Use of internal standards

The internal standards described in Section 3.5 were used for two quantitative analyses in this study. These standards allowed quantitative PCR, measuring the copy number of a chosen sequence in desired samples. They also permitted enumeration of individual mutant sequences in initial samples upon mutational analysis by CDCE.

Quantitative PCR is based on co-amplification of a desired sequence with an internal standard at known concentrations using one set of primers (Reischl and Kochanowski, 1995; Zimmermann and Mannhalter, 1996; Schnell and Mendoza, 1997; Raeymaekers, 2000). When these amplified species are differentiated, by CDCE for example, one can estimate the unknown initial copy number from the known quantities.

Use of an internal standard for quantitative analysis generates more accurate estimation compared to other PCR-based methods that rely on an external standard. These methods measure the quantity of a chosen sequence in the exponential phase using the dilution series of the external standard as a reference (Reischl and Kochanowski, 1995; Zimmermann and Mannhalter, 1996; Schnell and Mendoza, 1997; Raeymaekers, 2000). The advantage of using an internal standard for quantitative analysis is that the analysis is not affected by tube-to-tube variations in amplification efficiency (Raeymaekers, 2000). Moreover, measurement of a chosen sequence copy number is not restricted to the exponential phase (Raeymaekers, 2000).

However, it is a prerequisite that an internal standard has the same PCR-amplification efficiency as the sequence to be quantified. Internal standards of different sequence lengths and compositions have been shown to generate differential PCR-amplification efficiencies (McDowell et al., 1998). In addition, quantitative analysis of co-amplified PCR products being affected by allelic preference in DNA amplification has been suggested (Keohavong et al., 1991). Allelic preference in DNA amplification refers to preferential amplification of mutant sequences relative to the wild-type, or vice versa.

The internal standards used in this study are the same length and sequence as the corresponding sequence to be measured but they differ by one base pair. Such standards can be differentiated from the corresponding sequence by CDCE, allowing quantitative analysis. When the PCR-amplification efficiency of each internal standard was compared

to that of the corresponding sequence, no significant difference was observed. Figure 28 illustrates this comparison. Typically, experiments performed in this study applied about 20 doublings for quantitative PCR and about 40 doublings prior to mutational analysis by CDCE.

These internal standards, when added to the initial samples, allowed enumerating individual mutant sequences upon mutational analysis by CDCE. Analysis of rare mutations requires numerous procedures prior to mutation detection (see Section 3.3). Thus, use of an internal standard is crucial to accurately enumerate individual mutant sequences in the initial samples.

4.5 Mutational analysis: ligated vs. natural clamp

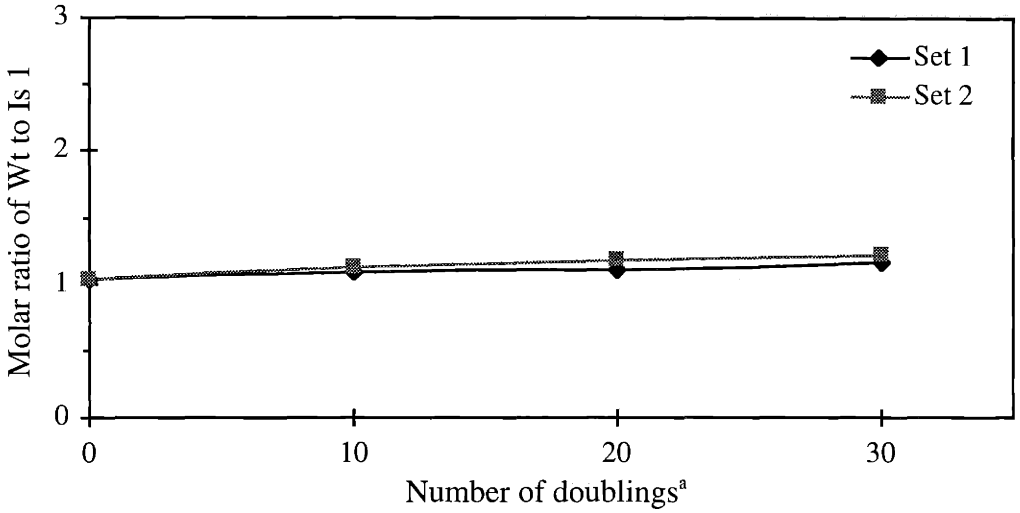
4.5.1 With CDCE-purified wild-type DNA of PCR products

4.5.1.1 Initial DNA samples

To discover the mutation detection sensitivity of the ligated clamp procedure in comparison to the natural clamp procedure, the wild-type DNA of the chosen *HPRT* target was prepared. As an initial study, the wild-type DNA of human cells was replaced with that of PCR products (Section 4.5.2 discusses the sensitivity comparison in human cells). This replacement allowed skipping the first three steps summarized in Section 3.3. However, the use of PCR products required purification of the wild-type. Otherwise, polymerase-generated mutations in the amplified target would have interfered with determining the true mutation detection sensitivity.

To be used as initial samples, the following three components were mixed: the CDCE-purified wild-type DNA, the 198-bp mutant internal standard in heteroduplexes at a known fraction, 5×10^{-6}, 5×10^{-5}, or 5×10^{-4}, and 60 to 600 ng of human genomic DNA per 2×10^{8} copies of the target wild-type. This mixture mimicked the stage after target isolation. It was necessary to use a mutant internal standard in heteroduplexes since homoduplex mutant sequences form heteroduplexes with the wild-type by mass action after a sample isolated for a chosen target in single strands is allowed to renature. When

A. Wild-type (Wt) vs. Internal standard 1 (Is 1: G -> A transition at cDNA bp 309)



B. Wild-type (Wt) vs. Internal standard 2 (Is 2: G -> T transversion at cDNA bp 312)

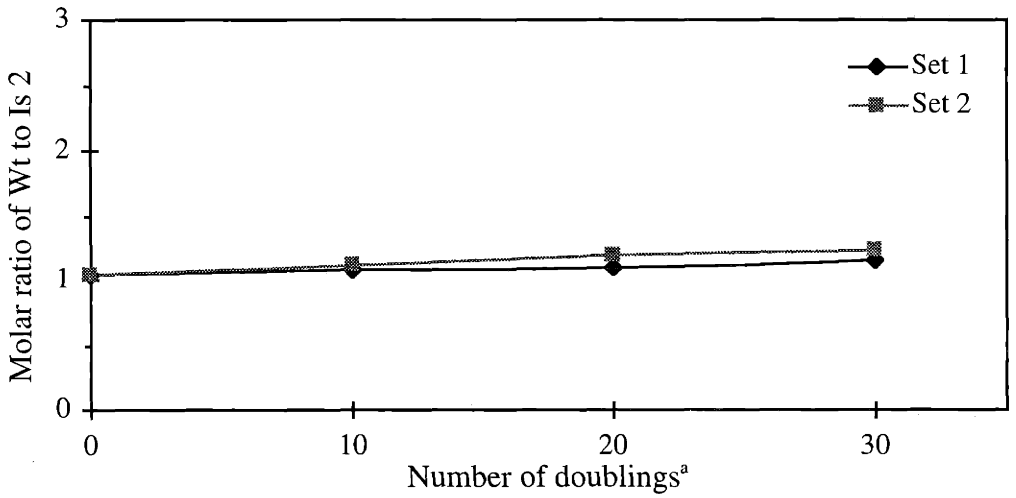


Figure 28. PCR-amplification efficiencies: wild-type vs. internal standards

A mixture containing the wild-type of the chosen *HPRT* target and each internal standard at an equal molar ratio was PCR amplified using the P3/P1 primer pair. After a desired number of doublings, the amplified sample was separated by CDCE to estimate the molar ratio of the amplified species. This ratio served as a reference to compare the PCR-amplification efficiencies of these species.

[a] 10 doublings are equivalent to a 10^3-fold target amplification, 20 doublings, a 10^6-fold, and so forth.

planning this mixture, a target isolation efficiency of 10^3- to 10^4-fold was chosen using 600 μg of human genomic DNA (≈ 10^8 cells) as an initial sample size. 10^8 cells is the sample size necessary for detecting mutations at fractions as low as 10^{-6} with statistical significance (±20% precision). 10^4-fold is the isolation efficiency observed in a previous study (Li-Sucholeiki and Thilly, 2000).

4.5.1.2 Restriction digestion and clamp ligation

Apo I restriction digestion was performed with the initial samples of CDCE-purified wild-type DNA, and the digestion efficiency was estimated to be above 95%. In these digested samples containing the added 60 to 600 ng of human genomic DNA (equivalent to ≈10^4-10^5 cells), between 1.2 x 10^{11} and 1.2 x 10^{12} *Apo* I-generated restriction ends were expected (see Section 4.2.1: *Apo* I is equivalent to a restriction endonuclease with a 5-bp recognition site). Thus, to achieve high efficiency clamp ligation to the expected range of *Apo* I restriction end numbers, either a 10^{13} or 10^{12} clamp copy number was used for each ligation reaction. This number represents a 10-fold higher clamp copy number compared to each expected number of *Apo* I restriction ends. Under such a condition, a clamp ligation efficiency of above 95% was achieved in Section 4.2.4.

To ensure that the added clamp copy number in each ligation reaction did generate the maximum clamp ligation efficiency, clamp dilution experiments were performed, as Table 17 summarizes. A similar clamp-ligated target copy number is observed in all negative control samples of different test clamp copy numbers. On the other hand, the samples of 60 or 600 ng of human genomic DNA show a decrease in the measured target copy number as the clamp copy number decreases. This decrease starts from the 10^{11} clamp copy number for those samples with 60 ng of genomic DNA and from the 10^{12} clamp copy number for those with 600 ng.

The clamp dilution experiments demonstrate that the clamp copy number limits the clamp ligation efficiency. This demonstration confirms an observation made in Section 4.2.4. The clamp dilution experiments also demonstrate that the 10^{13} clamp copy number is large enough to generate the maximum clamp ligation efficiency for a sample size of up to 600 ng of human genomic DNA. In addition, the theoretically estimated number of

Table 17. Summary of clamp dilution experiments

A 2 x 10^8 copy number of the chosen *HPRT* target with the natural clamp was added to each test sample, and those samples without added genomic DNA were used as negative controls. After *Apo* I restriction digestion, followed by clamp ligation, the copy number of the clamp-ligated target in each test sample was measured by PCR and CDCE. For this PCR, the GC1/P1 primer pair was used. The measured copy number served as a reference to compare the clamp ligation efficiencies among different test samples.

Copy number of clamp	Copy number of clamp-ligated *HPRT* target (*HPRT* cDNA bp 223-318) (mean ± SD, n=2)		
	No added gDNA (≈2 x 10^8 *Apo* I restriction ends)	With 60 ng gDNA (≈1.2 x 10^{11} *Apo* I restriction ends[a])	With 600 ng gDNA (≈1.2 x 10^{12} *Apo* I restriction ends[a])
10^{13}	7.5 ± 0.6 x 10^7	7.7 ± 1.1 x 10^7	8.6 ± 0.9 x 10^7
10^{12}	6.0 ± 0 x 10^7	7.8 ± 0.4 x 10^7	4.0 ± 0.1 x 10^7
10^{11}	7.1 ± 0.5 x 10^7	5.1 ± 0.1 x 10^7	5.2 ± 0.4 x 10^6
10^{10}	6.6 ± 0.2 x 10^7	5.2 ± 0.1 x 10^6	6.1 ± 1.1 x 10^5

[a] Theoretically estimated numbers, estimated by multiplying Equation 3 by 10^4 for 60 ng and by 10^5 for 600 ng. Equation 3: $[(6 \times 10^9) \div 4^X] \times 2$, where 6 x 10^9 is the number of bases per diploid human genome, 4^X is the cleavage frequency of a chosen restriction enzyme in bases (X = 5 for *Apo* I restriction enzyme), and 2 is the number of restriction ends generated per cleavage. 60 ng (10^4 cells) and 600 ng (10^5 cells) of human genomic DNA are the sample sizes after target isolation with efficiencies of 10^4- and 10^3-fold, respectively, using an initial sample size of 10^8 cells.

Apo I restriction ends per given cell number is shown to be in close agreement with those determined empirically.

4.5.1.3 Pre-PCR mutant enrichment

4.5.1.3.1 Necessity

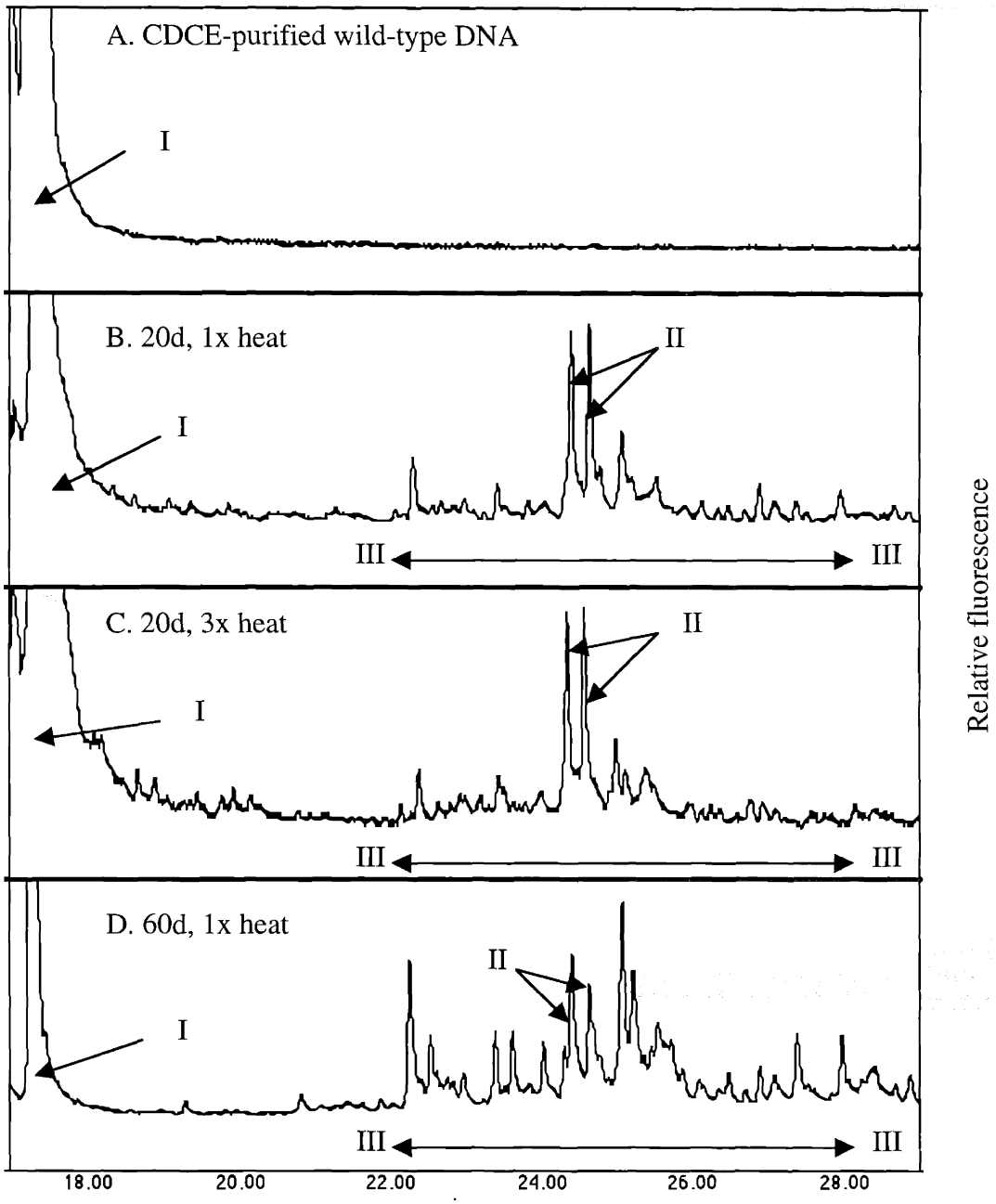
Analysis of rare mutations requires enrichment of mutant sequences relative to their abundant wild-type. This procedure reduces the wild-type copy number while retaining the same mutant copy numbers. As a result, mutant fractions are increased in mutant-enriched samples compared to those in initial samples. CDCE, for example, spatially separates mutant from wild-type sequences, and elution of these separated mutants allows mutant enrichment. Such a procedure can be performed before or after PCR, depending on the mutant copy numbers in desired samples (Kim et al., 2001).

When analyzing mutations at fractions below about 2×10^{-4}, pre-PCR mutant enrichment is an absolute requirement. This necessity is based on the fidelity of *Pfu* DNA polymerase with the applied PCR cycles representing 20 target doublings (see Section 2.5.3.1). Figure 29 illustrates the levels of background noise generated by *Pfu* using the CDCE-purified wild-type DNA of the chosen *HPRT* target as a template. After 20 doublings of target amplification, the level of *Pfu*-generated background noise is about 5 to 2 times lower than the mutant internal standard added at an initial fraction of 10^{-3} (see Figure 29B). This background noise level does not increase as a function of template denaturation time (see Figure 29B vs. C: a 3-fold increase from 14 to 42 s at 94°C per doubling). This observation rules out the possibility of DNA lesions, generated by heat during PCR, contributing as background noise. On the other hand, the level of background noise increases as the number of doublings is increased (see Figure 29B vs. D). These results indicate that the fidelity of *Pfu* determines the level of background noise.

Given the fidelity of *Pfu* and assuming that 20 doublings must be performed prior to mutational analysis by CDCE, only those mutations at fractions above 2×10^{-4} can be directly analyzed. Thus, for analysis of mutations at fractions below 2×10^{-4}, pre-PCR

Figure 29. Levels of mutations generated by *Pfu* DNA polymerase during PCR

The chosen *HPRT* target was PCR amplified using the P3/P1 primer pair, followed by CDCE purification of the wild-type DNA (A). An additional PCR was performed using the purified wild-type DNA as a template. For this PCR, B used 20 doublings and a template denaturation time of 14 s at 94°C per doubling (1x heat), C used 20 doublings and a template denaturation time of 42 s at 94°C per doubling (3x heat), and D used 60 doublings and a template denaturation time of 14 s at 94°C per doubling (1x heat). Post-PCR mutant enrichment, followed by PCR, was performed prior to mutational analysis by CDCE.



I represents the wild-type in homoduplex; II represents the mutant internal standard (IS: G to A transition at *HPRT* cDNA bp 309) added at an initial fraction of 10^{-3}; III represents a region where the majority of *Pfu*-generated mutations in the chosen target migrate; the CDCE-separated mutants are in mutant/wild-type heteroduplexes.

mutant enrichment must be performed. This procedure increases these mutant fractions prior to PCR, allowing mutation detection above the level of background noise generated by *Pfu*.

4.5.1.3.2 Efficiency

Pre-PCR mutant enrichment allows analysis of rare mutations at fractions below 2 x 10^{-4}. However, the sensitivity of a mutation detection method is expected to vary depending on the efficiency of pre-PCR mutant enrichment. This prediction is based on a previous study (Li-Sucholeiki and Thilly, 2000). In this study, the sensitivity of the natural clamp procedure was demonstrated to be limited by the fidelity of *Pfu* DNA polymerase, or equivalently, by the efficiency of pre-PCR mutant enrichment. Thus, the sensitivity of the proposed point mutation detection method can be predicted based on the efficiency of pre-PCR mutant enrichment.

On average, the efficiency of pre-PCR mutant enrichment was estimated to be 32 ± 4-fold for the chosen *HPRT* target with the ligated clamp. This efficiency represents a 32-fold decreased copy number of the wild-type or a 32-fold increased mutant fraction in pre-PCR mutant-enriched samples compared to that in initial samples. An average efficiency of 125 ± 28-fold was estimated for the same target with the natural clamp. These estimations represent the total mutant enrichment efficiencies after CDCE followed by CE.

CE was performed as a means to reduce the copy number of the wild-type in single strands in the CDCE-eluted samples of mutant/wild-type heteroduplexes. As much as 3% target DNA in double strands converted to single strands during sample loading into a capillary. The single-stranded target, regardless of whether mutant or wild-type, co-migrated with mutant/wild-type heteroduplexes under optimal CDCE-separation conditions. To reduce the copy number of single-stranded wild-type in the CDCE-eluted samples, the double-stranded target was separated from the single-stranded target by CE. This procedure was followed by elution of the double-stranded target.

The average efficiency of pre-PCR mutant enrichment estimated for the ligated clamp procedure is about 4-fold lower than that for the natural clamp procedure. This

estimation predicts about a 4-fold lower mutation detection sensitivity for the ligated clamp procedure. The following section discusses the cause of this decrease.

4.5.1.3.3 Cause of decreased efficiency

The efficiency of pre-PCR mutant enrichment is about 4-fold lower for the ligated clamp procedure compared to that for the natural clamp procedure. To determine the cause of this decreased efficiency, the effects of variables introduced by clamp ligation on the mutant enrichment efficiency were investigated. The three variables investigated were: (1) non-target *Apo* I restriction fragments in ligation reaction mixtures, (2) the preparation procedure of the ligated clamp, and (3) the sequence context of the ligated clamp.

(1) Non-target *Apo* I restriction fragments in ligation reaction mixtures: Section 4.2.1 discusses the ratio of target to non-target restriction ends in ligation reaction mixtures. In Section 4.2.1, a 10^2- to 10^3-fold higher copy number of non-target restriction ends was estimated after target isolation. These non-target restriction ends with the ligated clamp may interfere with target amplification and CDCE separation when estimating the mutant enrichment efficiency. As a result, an inaccurate estimation may be generated.

To test this possibility, control samples, each containing 0, 60, or 600 ng of human genomic DNA, were prepared. When the mutant enrichment efficiencies were compared among these samples, no statistically significant difference was observed. Therefore, non-target *Apo* I restriction fragments in ligation reaction mixtures were determined not to be the cause of the decreased efficiency of pre-PCR mutant enrichment.

(2) The preparation procedure of the ligated clamp: As a first means of preparing the clamp for ligation, two complementary oligonucleotides were synthesized, purified by PAGE, and hybridized to each other (see Section 3.2.3.1). The PAGE-purified oligonucleotides were supplied by Synthetic Genetics with a claimed purification efficiency of 95%. Purification of the oligonucleotides was necessary since impure intermediates were expected in the crude synthesis mixtures (Caruthers et al., 1983; Itakura et al., 1984; Warren and Vella, 1995).

High-performance liquid chromatography (HPLC) is effective in purifying synthetic oligonucleotides shorter than about 30 bases in length (Itakura et al., 1984; Warren and Vella, 1995). For oligonucleotides up to 100 bases in length, a number of purification methods have been reported (Johnson et al., 1990; Frenz and Hancock, 1991; Warren and Vella, 1993). Among them, PAGE is the most widely used (Efcavitch, 1990). Regardless of a purification method of choice, some degree of impurity is expected in the purified mixture (e.g., a 5% impurity with a 95% purification efficiency).

An impure clamp, resulting from the impurities of the clamp oligonucleotides, possibly caused the decreased efficiency of pre-PCR mutant enrichment. This clamp, when ligated to the chosen target, may have generated slower electrophoretic mobilities than the pure clamp generated under optimal CDCE-separation conditions. If this is true, the target wild-type with the impure clamp co-migrated with mutant/wild-type heteroduplexes. As a result, the decreased mutant enrichment efficiency was generated.

To test this possibility, a second means of preparing the clamp was devised (see Section 3.2.3.2). This procedure is thought to generate the pureist form of the clamp using current technologies. The mutant enrichment efficiency of the control sample of the first clamp preparation procedure was compared to that of the second; no statistically significant difference was observed. Therefore, the impurities of the clamp oligonucleotides were determined not to be the cause of the decreased efficiency of pre-PCR mutant enrichment.

(3) The sequence context of the ligated clamp: This variable was determined to be the cause of the decreased efficiency of pre-PCR mutant enrichment. The ligated and natural clamps have differences in their sequence contexts, such as 83 and 53% GC contents, respectively. Thus, the effect of this variable on the mutant enrichment efficiency was investigated.

For this investigation, a control clamp comprising the sequence of the natural clamp was prepared using the second means of preparing the clamp. The chosen target with the ligated control clamp is the same as that with the ligated clamp, except that the sequence of the ligated clamp is replaced with that of the natural clamp. On average, the efficiency of mutant enrichment for the sample of control clamp was estimated to be 4 times higher than that for the sample of the ligated clamp. Therefore, the sequence context of the

ligated clamp was determined to be the cause of the decreased efficiency of pre-PCR mutant enrichment.

The ligated clamp with a high GC content is a crucial factor in the proposed point mutation detection method. Such a clamp allows analysis of DNA sequences with melting temperatures below 80°C. However, the sequence of the ligated clamp causes the decreased efficiency of pre-PCR mutant enrichment. Section 4.5.1.3.2 discusses the efficiency of pre-PCR mutant enrichment as being directly related to the sensitivity of a mutation detection method. In brief, the sensitivity of ligated clamp procedure has to be sacrificed to keep the procedure general to the human genome.

4.5.1.4 Hifi-PCR and post-PCR mutant enrichment

The CDCE-mutation detection limit was determined to be about 10^{-3} without prior mutant enrichment. This determination is slightly different from that observed in a previous study, 3×10^{-4} (Khrapko et al., 1994a). This limit of 10^{-3} requires post-PCR mutant enrichment for those PCR-amplified samples containing mutations at fractions at or below 10^{-3}.

Figure 30 illustrates CDCE-mutation detection before and after post-PCR mutant enrichment. The mutant internal standard added at an initial fraction of 10^{-3} is right at the detection limit before the procedure (Figure 30A). After the procedure, the same mutant is detected above the detection limit (Figure 30B). Mutations in a chosen target, created by *Pfu* DNA polymerase during PCR, are expected in post-PCR mutant-enriched samples. Figure 30B illustrates these *Pfu*-generated mutations appearing as background noise.

On average, the efficiency of post-PCR mutant enrichment by CDCE was estimated to be 15-fold for the ligated and natural clamp procedures. This efficiency represents a 15-fold increased mutant fraction in post-PCR mutant-enriched samples compared to that in original PCR-amplified samples. For example, a mutant fraction of 1.5×10^{-2} is generated in post-PCR mutant-enriched samples using a mutant fraction of 10^{-3} in original PCR-amplified samples. As a result, mutations are detected 15 times above the CDCE detection limit of 10^{-3}.

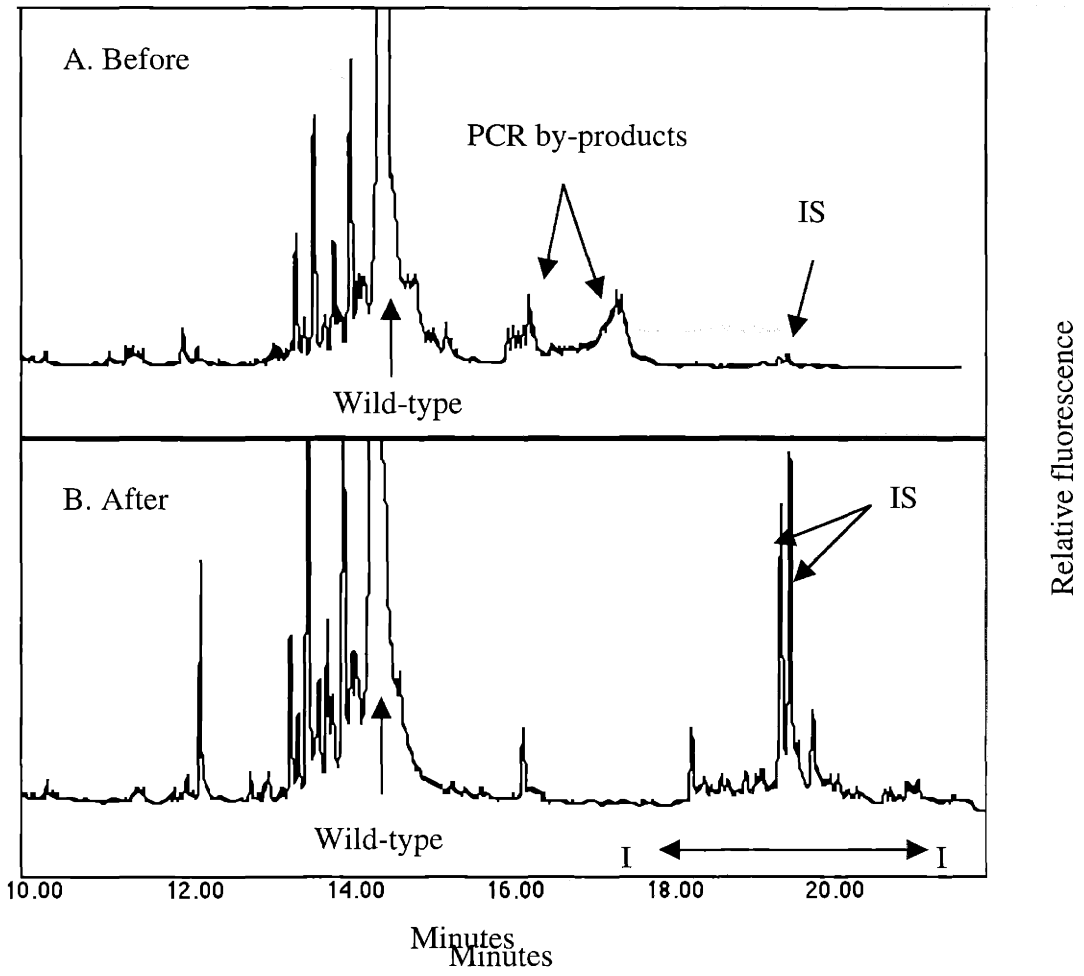


Figure 30. Mutation detection by CDCE: before vs. after post-PCR mutant enrichment*

A. The chosen *HPRT* target with the natural clamp was PCR amplified using the P3/P1 primer pair prior to mutational analysis by CDCE.
B. Post-PCR mutant enrichment by CDCE was performed with the PCR-amplified sample from Figure 30 A. An additional PCR was performed prior to mutational analysis by CDCE.

* IS represents the mutant internal standard (G to A transition at *HPRT* cDNA bp 309) added at an initial fraction of 10^{-3}; I represents a region where the majority of *Pfu*-generated mutations in the chosen target migrate; the CDCE-separated mutants are in mutant/wild-type heteroduplexes; the wild-type peak is shown at 1/20 of its full height.

4.5.1.5 Sensitivity

Table 18 summarizes the efficiencies of pre- and post-PCR mutant enrichment. The combination of these two procedures results in the total mutant enrichment efficiency of 480-fold for the ligated clamp procedure and 1875-fold for the natural clamp procedure. This total efficiency predicts an increased mutant fraction of 2.4×10^{-2} in final mutant-enriched samples using an initial fraction of 5×10^{-5}. An increased fraction of 9.4×10^{-2} is predicted for the natural clamp procedure. Figures 31 and 32 illustrate CDCE detection of the mutant internal standard added at an initial fraction of 5×10^{-4}, 5×10^{-5}, or 5×10^{-6}.

Given that mutations in final mutant-enriched samples are detected above the detection limit of CDCE, analysis of mutations in original samples is limited by the level of background noise. Equivalently, the mutation detection sensitivity of the ligated or natural clamp procedure is limited by this level. Background noise results from DNA modifications generated prior to PCR and mutations generated by DNA polymerase during PCR. Sections 4.5.1.6 and 4.5.2.10 discuss the source of background noise.

The average level of background mutants (M_B) was compared to that of the mutant internal standard (M_I) added at the known initial fraction (F) to estimate the sensitivity: $[(M_B \div M_I)] \times F$. The background mutants compared are six predominant mutations generated by *Pfu* DNA polymerase used for amplifying pre-PCR mutant-enriched samples (see Section 4.5.2.10). A sensitivity of 10^{-5} was estimated for the ligated clamp procedure, and 3×10^{-6} for the natural clamp procedure. This decreased sensitivity was predicted in Section 4.5.1.3.2.

4.5.1.6 Source of background noise

In Section 4.5.1.5, CDCE-purified wild-type DNA was used as initial DNA samples to estimate the mutation detection sensitivities. With these samples *Apo* I restriction digestion was performed at 37°C for 2 hr, and pre-PCR mutant enrichment was performed at ≈ 66°C for 1 to 2 hr. These samples were not exposed to additional heat prior to PCR. These preparation procedures suggest *Pfu*, a DNA polymerase used for

Table 18. Summary of mutant enrichment efficiencies

Mode of mutant enrichment	Average efficiency	
	Ligated clamp	Natural clamp
Pre-PCR	32-fold	125-fold
Post-PCR	15-fold	15-fold
Total	480-fold	1875-fold

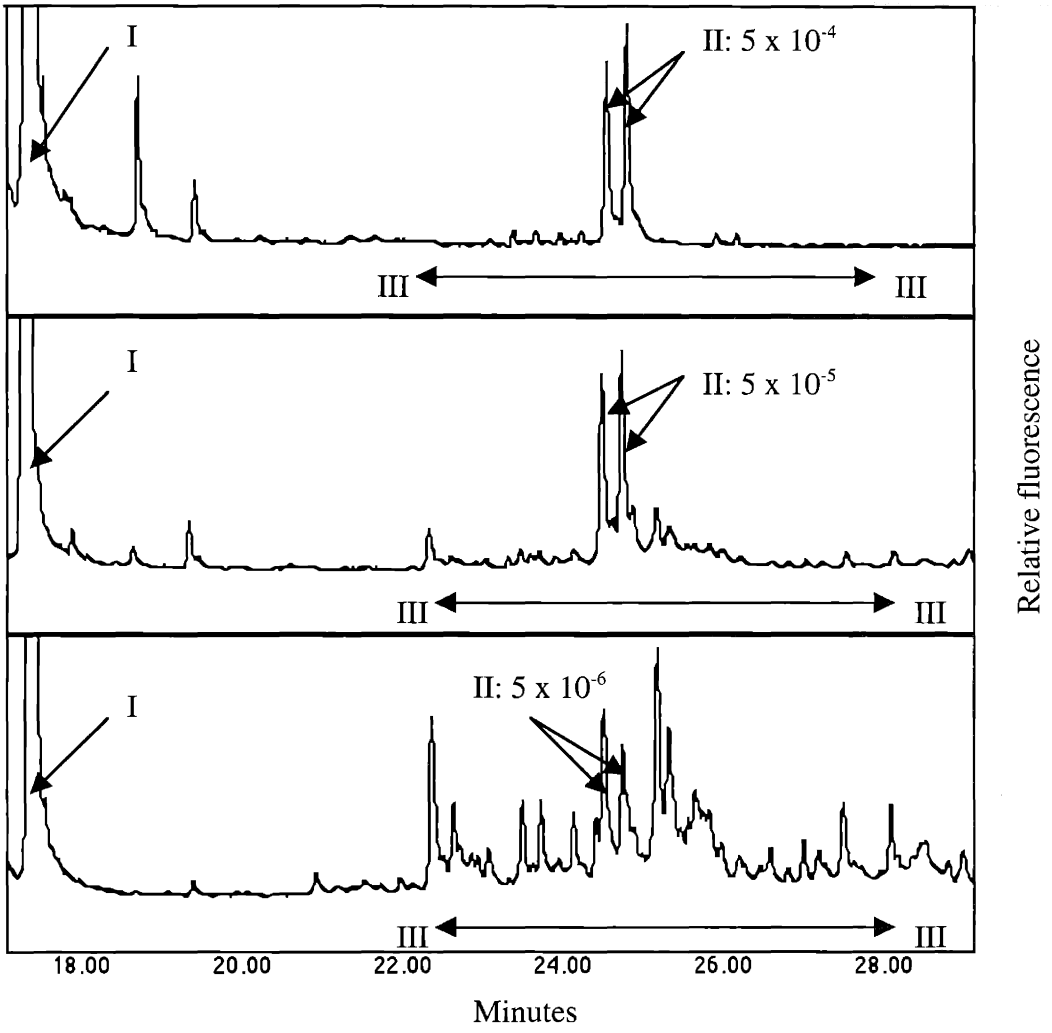


Figure 31. Mutation detection by CDCE: natural clamp (initial samples of CDCE-purified wild-type DNA)*

Pre-PCR mutant enrichment by CDCE, followed by CE, was performed with each initial sample of CDCE-purified wild-type DNA. This procedure was followed by PCR using the P3/P1 primer pair to amplify the chosen *HPRT* target with the natural clamp. Post-PCR mutant enrichment by CDCE, followed by PCR, was performed prior to mutational analysis by CDCE.

*I represents the wild-type in homoduplex; II represents the mutant internal standard (IS: G to A transition at *HPRT* cDNA bp 309) added at an initial fraction of 5×10^{-4}, 5×10^{-5}, or 5×10^{-6}; III represents a region where the majority of the background mutants migrate; the CDCE-separated mutants are in mutant/wild-type heteroduplexes.

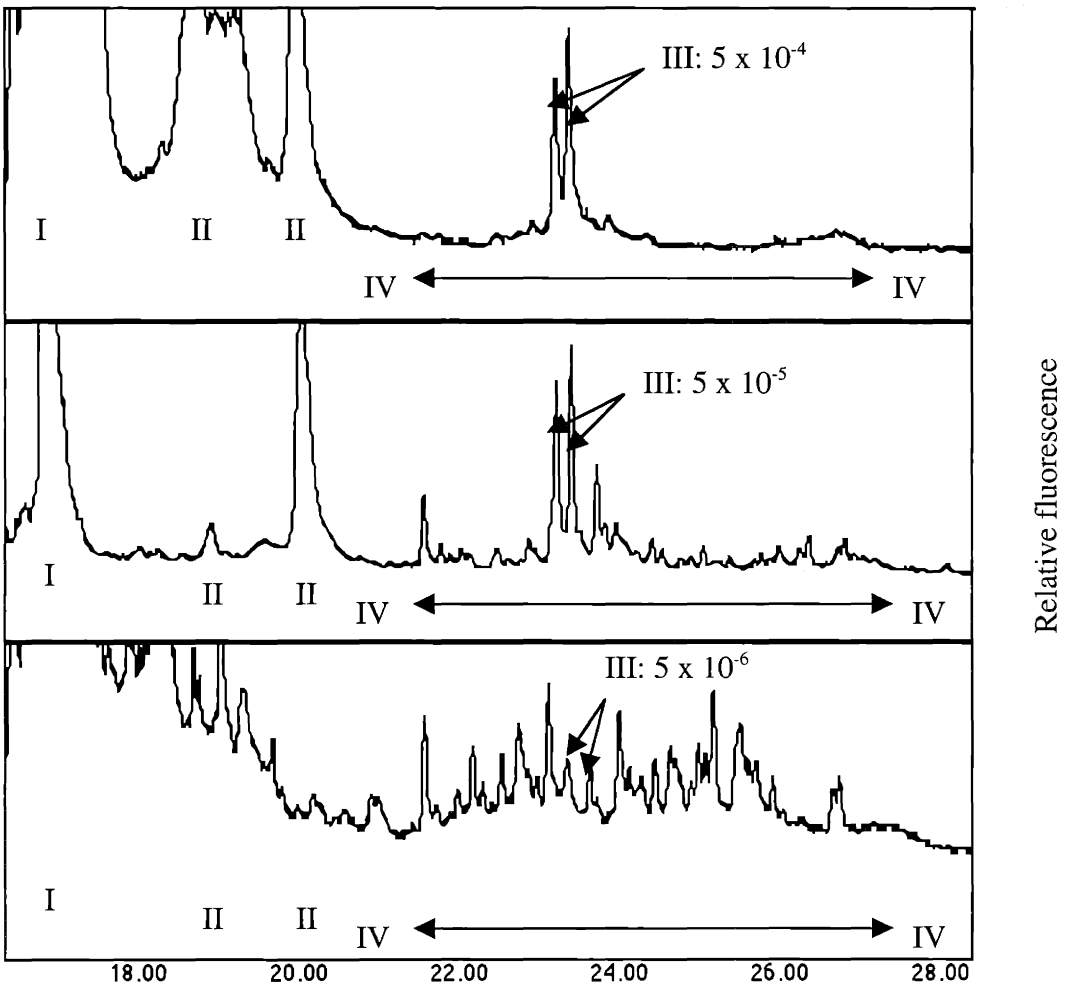


Figure 32. Mutation detection by CDCE: ligated clamp (initial samples of CDCE-purified wild-type DNA)*

Apo I restriction digestion and clamp ligation were performed with each initial sample of CDCE-purified wild-type DNA. These procedures were followed by pre-PCR mutant enrichment by CDCE and CE. Using the GC1/P1 primer pair, the chosen *HPRT* target with the ligated clamp was PCR amplified, followed by post-PCR mutant enrichment by CDCE. Final mutant-enriched samples were PCR amplified prior to mutational analysis by CDCE.

*I represents the wild-type in homoduplex; II represents PCR by-products; III represents the mutant internal standard (IS: G to A transition at *HPRT* cDNA bp 309) added at an initial fraction of 5×10^{-4}, 5×10^{-5}, or 5×10^{-6}; IV represents a region where the majority of the background mutants migrate; the CDCE-separated mutants are in mutant/wild-type heteroduplexes.

PCR, as the source of background noise.

Using an initial sample size of 2×10^8 wild-type copy number, a sample size of 6×10^6 ($(2 \times 10^8) \div 32$) was estimated for the ligated clamp procedure, and 2×10^6 ($(2 \times 10^8) \div 125$) for the natural clamp procedure in pre-PCR mutant-enriched samples. These sample sizes were estimated by directly measuring the wild-type copy number in the mutant-enriched samples. These estimations are equivalent to those calculated using the efficiencies of pre-PCR mutant enrichment. About 20 doublings were applied to each pre-PCR mutant-enriched sample to convert all the primers into products. 20 doublings are equivalent to a 10^6-fold target amplification.

An increased mutant fraction of 1.6×10^{-3} in pre-PCR mutant-enriched samples is expected using an initial mutant fraction of 5×10^{-5}. This expectation is based on the efficiency of pre-PCR mutant enrichment observed for the ligated clamp procedure. After 20 doublings of target amplification, this increased mutant fraction is expected to be 4 times greater than the mutant fraction that belongs to predominant mutations generated by *Pfu* during PCR amplification of pre-PCR mutant-enriched samples ($(2 \times 10^{-6}$/bp/doubling) x 20 doublings x $10 = 4 \times 10^{-4}$). The expected *Pfu*-mutant fraction is based on the fidelity of *Pfu* observed for the chosen *HPRT* target. In addition, this mutant fraction is based on the predominant *Pfu* mutations occurring 10 times more frequently than expected by chance after 20 doublings. The expected level of *Pfu*-mutant fraction is in close agreement with the sensitivity observed for the ligated clamp procedure (1.3×10^{-5}: $[(4 \times 10^{-5}) \div (1.6 \times 10^{-3})] \times 5 \times 10^{-5}$). This agreement is true for the natural clamp procedure.

These agreements indicate that the fidelity of *Pfu* determines the level of background noise upon mutational analysis by CDCE. To verify this indication, the level of background noise as a function of doublings was investigated. As Figure 33 illustrates, the level of background noise increases as a function of doublings. Figure 33A represents a pre-PCR mutant-enriched sample after 20 doublings. Figures 33B and C represent the same sample with an additional 20 doublings.

These results demonstrate that the level of background noise is determined by the fidelity of *Pfu*. Equivalently, this fidelity determines the mutation detection sensitivity of either the ligated or natural clamp procedure. Section 4.5.2.10 discusses the positions

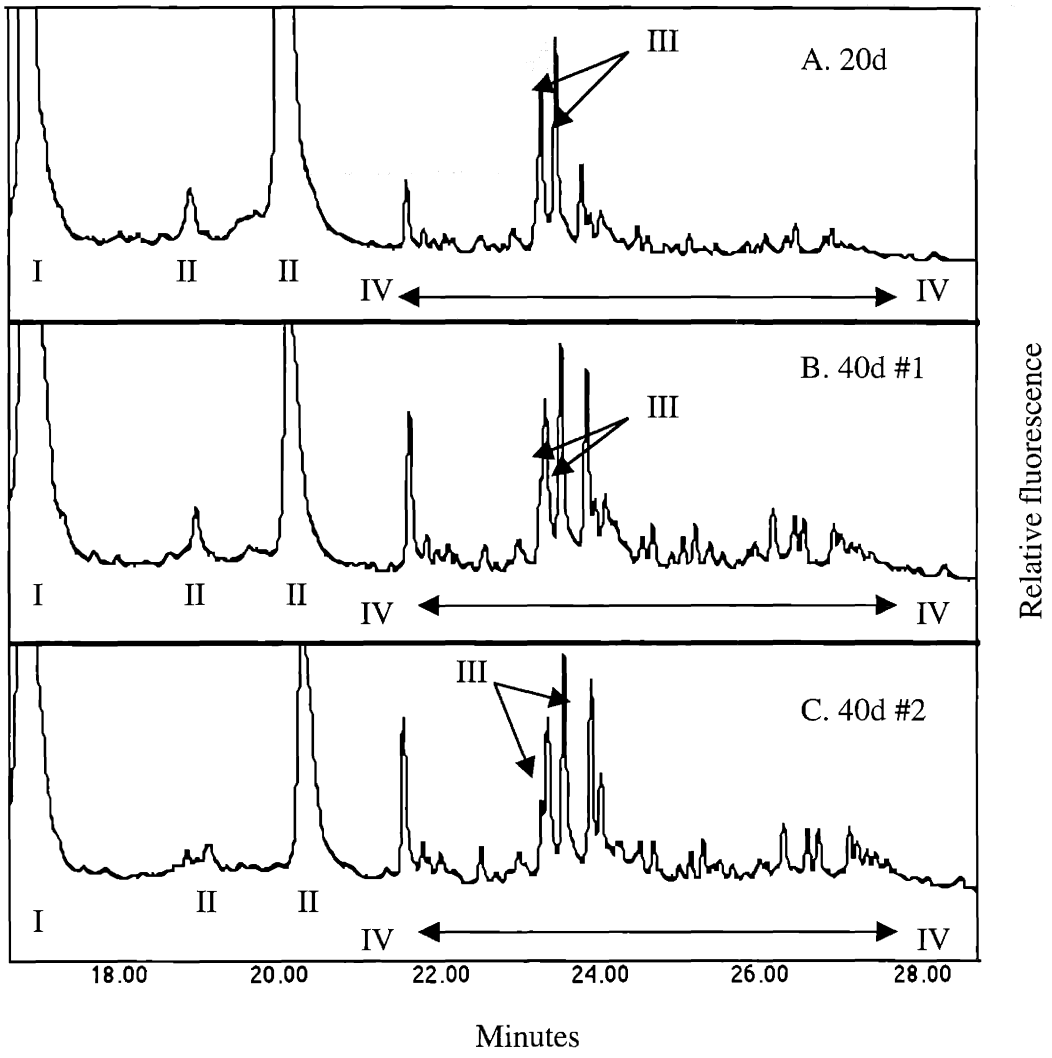


Figure 33. Background mutants vs. doublings*

Apo I restriction digestion and clamp ligation were performed with each initial sample of CDCE-purified wild-type DNA. These procedures were followed by pre-PCR mutant enrichment by CDCE and CE. Using the GC1/P1 primer pair, the chosen *HPRT* target with the ligated clamp was PCR amplified using either 20 (A) or 40 (B and C) doublings. After post-PCR mutant enrichment by CDCE, final mutant-enriched samples were PCR amplified and analyzed by CDCE.

*I represents the wild-type in homoduplex; II represents PCR by-products; III represents the mutant internal standard (IS: G to A transition at *HPRT* cDNA bp 309) added at an initial fraction of 5×10^{-5}; IV represents a region where the majority of the background mutants migrate; the CDCE-separated mutants are in mutant/wild-type heteroduplexes; 20 doublings are equivalent to a 10^6-fold target amplification, and 40 doublings, a 10^{12}-fold.

and kinds of *Pfu* mutations in the chosen *HPRT* target.

4.5.2 With wild-type DNA of human cells

4.5.2.1 Initial DNA samples

To discover the mutation detection sensitivity of the ligated clamp procedure in comparison to the natural clamp procedure in human cells, TK6 was chosen as a target cell line. TK6, an approximately diploid human B cell line, was originally established from a male patient with hereditary spherocytosis (Levy et al., 1968) and was isolated after treatment with ICR-191 (Skopek et al., 1987). TK6 is hemizygous at *HPRT* locus and is heterozygous at *tk* locus. Thus, selection of mutations in these loci can be performed using selecting agents, such as 6TG.

TK6 cells were obtained from a previous study (Tomita-Mitchell, 1999). These cells had been left untreated, maintained in exponential growth by daily dilution, and frozen after 15.5 generations. The expected total *HPRT* mutant fraction in these cells is about 4.2×10^{-6} ($2.7 \times 10^{-7} \times 15.5$). This expectation is based on the spontaneous mutation rate per doubling, 2.7×10^{-7}, observed for the *HPRT* gene (Tomita-Mitchell, 1999). Assuming a target size of about 1000-bp for this gene, *HPRT* hotspot mutations that had occurred 10 to 100 times more frequently than expected by chance, would appear at fractions of 4.2×10^{-8} ($[(4.2 \times 10^{-6}) \div 1000] \times 10$) to 4.2×10^{-7} ($[(4.2 \times 10^{-6}) \div 1000] \times 100$) in these cells. Such *HPRT* hotspot mutations in human cells have been shown to represent about 50% of the point mutations that alter the physiological function of the gene product (Kat, 1992; Tomita-Mitchell et al., 2000). These expected mutant fractions are lower than the mutation detection sensitivities determined for the ligated and natural clamp procedures using the initial samples of CDCE-purified wild-type DNA. To compare the sensitivity of the ligated clamp procedure in human cells to that of the natural clamp procedure, genomic DNA isolated from the TK6 cells of Tomita-Mitchell (1999) was used as initial samples.

4.5.2.2 Genomic DNA isolation

A procedure described in a previous study allows DNA isolation with minimal exposure to potential mutagenic agents, such as phenol (Khrapko et al., 1997b). In this procedure, genomic DNA is isolated from cells by digestion with proteinase K, SDS, and RNase A, followed by ethanol precipitation of the DNA. The DNA isolated by this procedure has been demonstrated to be suitable for analysis of point mutations at fractions as low as 10^{-6} (Khrapko et al., 1997a,b; Coller et al., 1998; Marcelino et al., 1998; Li-Sucholeiki and Thilly, 2000).

The quantity and quality of genomic DNA isolated from human cells was assessed by a UV spectrophotometer. The typical DNA yield was over 90% with the ratio of A_{260} to A_{280} in the range of 1.6 to 2.0.

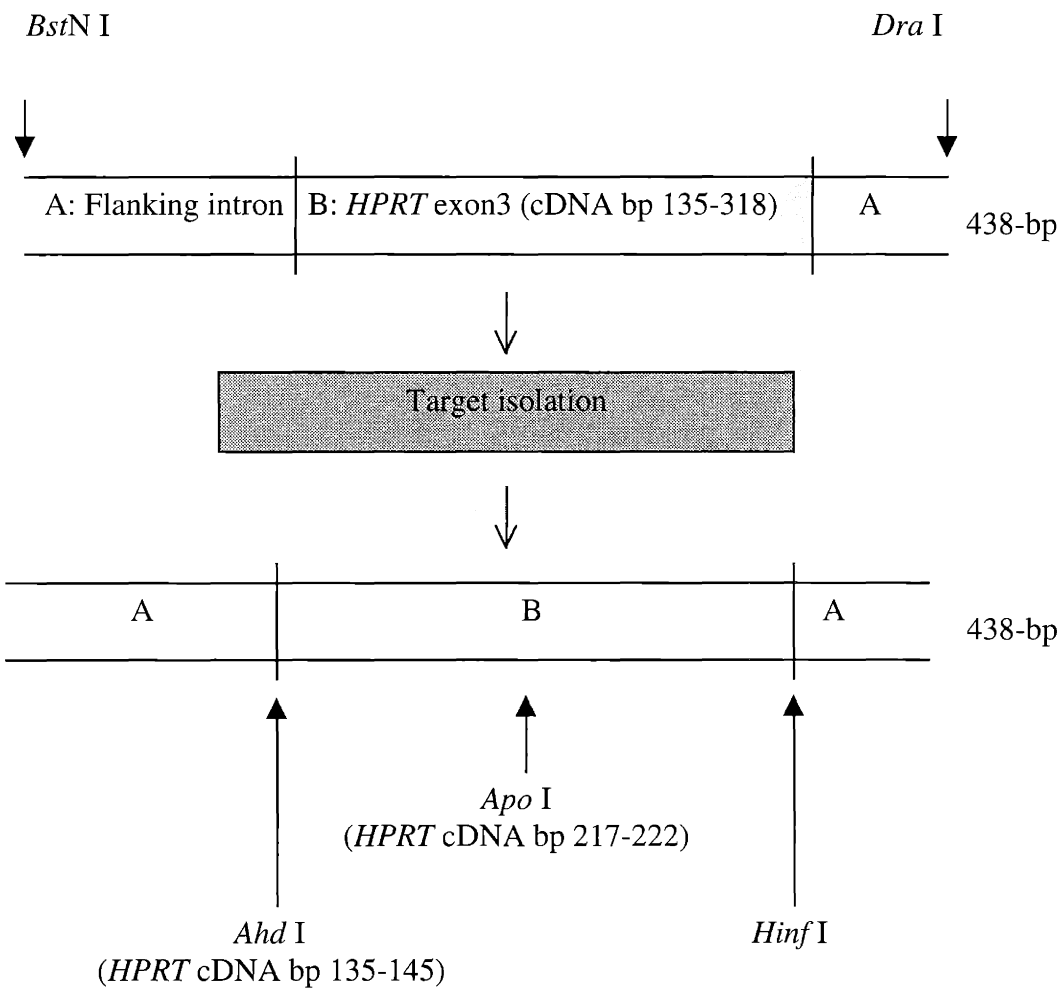
4.5.2.3 Sequential restriction digestion

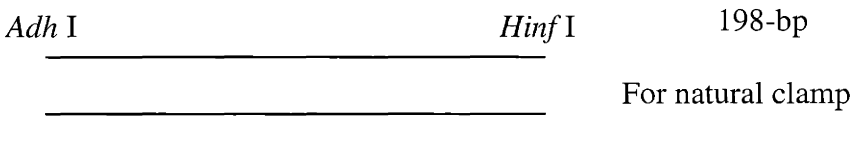
Figure 34 illustrates two sequential restriction digestion procedures, with target isolation in between. The order of these procedures is necessary to generate the chosen *HPRT* target suitable for CDCE separation with a minimum cost. Monetary cost has to be considered since liberating the chosen target from a large quantity of genomic DNA, such as 60 μg ($\approx 10^7$ cells), in many samples can be expensive. 60 μg is the sample size necessary for detecting mutations at fractions as low as 10^{-5} with statistical significance (±20% precision).

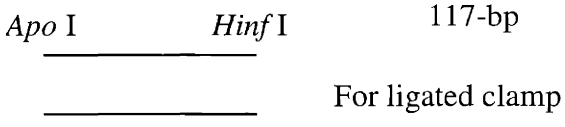
For the first restriction digestion, *Bst*N I and *Dra* I were selected. These enzymes liberate the 438-bp *HPRT* target-embedded fragment from human genomic DNA at a minimum cost. After this target-embedded fragment was isolated from the *Bst*N I and *Dra* I digest, the second restriction digestion followed. This second digestion generated the chosen target suitable for CDCE separation. The cost of restriction endonucleases for the second digestion is no longer a limiting factor since an initial sample size is reduced after target isolation. For example, an initial sample size of 60 μg is reduced to 6 ng using an isolation efficiency of 10^4-fold. For the second restriction digestion, *Ahd* I and

Figure 34. Restriction digestion by *Bst*N I, *Dra* I, *Ahd* I, *Hinf* I, and *Apo* I

The first restriction digestion by *Bst*N I and *Dra* I liberates the 438-bp *HPRT* target-embedded fragment from human genomic DNA. After this target-embedded fragment is isolated from the *Bst*N I and *Dra* I digest, the second restriction digestion followed. This digestion generates the chosen target suitable for CDCE separation. Restriction digestion with *Ahd* I and *Hinf* I is performed for the natural clamp procedure, and with *Apo* I for the ligated clamp procedure.


*Bst*N I
Dra I
A: Flanking intron
B: *HPRT* exon3 (cDNA bp 135-318)
A
438-bp
Target isolation
A
B
A
438-bp
Apo I
(*HPRT* cDNA bp 217-222)
Ahd I
(*HPRT* cDNA bp 135-145)
Hinf I


Adh I
Hinf I
198-bp
For natural clamp


Apo I
Hinf I
117-bp
For ligated clamp

Hinf I were selected for the natural clamp procedure. For the ligated clamp procedure, *Apo* I was chosen.

4.5.2.4 Target isolation

Figure 18 in Section 3.8.3.1 illustrates isolation of a chosen target from a pool of genomic DNA digest. Using the chosen *HPRT* sequence as a target, this procedure generated a 72 ± 2 % target yield. This target yield closely agrees with that observed in a previous study (74%, Li-Sucholeiki and Thilly, 2000).

As a first means of estimating the isolation efficiency of the chosen *HPRT* target relative to non-target sequences in the pool of genomic DNA digest, a part of human mitochondrial DNA (DNA bp 10011-10215) was selected as a non-target reference. About a few hundred to a thousand mitochondrial sequence copies per human cell have been estimated (Robin and Wong, 1988; Marcelino et al., 1998; Khrapko et al., 1999). One *HPRT* sequence copy (X-linked) is expected per male-derived cell, TK6. The molar ratio of the mitochondrial to target sequences in initial samples was estimated to be about 2000/1 (R_I). This ratio was compared to that in target-isolated samples, 1/0.55 ± 0.23 (R_F), to estimate the isolation efficiency: $R_I \div R_F$, 3.6 x 10^3-fold.

As a second means of estimating the efficiency of target isolation, the absorbencies of initial and target-isolated samples were measured by a UV spectrophotometer at a wavelength of 260 nm, and the sample DNA concentrations per a given target copy number were compared. By this method, an efficiency of 700 ± 210-fold was estimated. On average, this efficiency is about 5-fold lower than that estimated by the first method. The second method is thought to be more accurate since the measurement refers to all non-target sequences a sample may have, while the measurement in the first method relies on one non-target sequence.

Using the first means of estimating the target isolation efficiency, one study has reported a 3-fold greater efficiency (Li-Sucholeiki and Thilly, 2000) compared to that observed in this study. Two possible causes of this difference are the use of two different target sequences and a 2-fold difference in the initial ratio of target to non-target

sequences. The *HPRT* target used in this study is expected to be present as one copy per cell. A part of the human *APC* gene, used as a target in the previous study, is expected to be present as two copies per cell.

Based on the target yield and the isolation efficiency observed for the chosen *HPRT* target, a sample size of 86 ng (10^7 cells ÷ 700 = 1.4 x 10^4 cells) is expected in target-isolated samples using an initial sample size of 60 μg human genomic DNA (10^7 target copy number in 10^7 cells). In these target-isolated samples, a target copy number of 7 x 10^6 is expected (10^7 x 0.7 = 7 x 10^6).

4.5.2.5 Target renaturation

4.5.2.5.1 Potential source of background noise

The established protocol for target isolation isolates a chosen target in single strands. Thus, target renaturation is necessary prior to pre-PCR mutant enrichment by CDCE. Substrate concentration, incubation temperature, and buffer composition have been identified as important factors in DNA renaturation (Doty et al., 1960; Marmur and Lane, 1960; Marmur and Doty, 1961; Thrower and Peacocke, 1966; Thrower and Peacocke, 1968). In a previous study, incubation at 55°C for 16 hr in 0.2 M NaCl was performed with each target-isolated sample (Li-Sucholeiki and Thilly, 2000). The substrate concentration used in the study of Li-Sucholeiki and Thilly (2000) was 10^8/μl (personal communication).

The established protocol for target renaturation was investigated since it requires sample exposure to heat, at 55°C, for a long period of time, 16 hr. Heat induces DNA modifications at different rates, with the rates depending on temperature, DNA conformation, pH, and buffer composition (Greer and Zamenhof, 1962; Lindahl and Nyberg, 1972; Lindahl and Karlström, 1973; Lindahl and Nyberg, 1974; Ehrlich et al., 1986). Table 19 summarizes the published rates of heat-induced DNA modifications at various temperatures.

Single-stranded target DNA with modifications, when renatured with the wild-type, is expected to change thermal stability from that without modifications. Indeed,

Table 19. Rates of heat-induced DNA modifications vs. temperatures

Type of modification	Temperature	Rate*	Reference(s)
Deamination of cytosine	95°C	2.0×10^{-7}/sec	Lindahl and Nyberg, 1974; Ehrlich et al., 1986
	70°C	1.2×10^{-8}/sec	Ehrlich et al., 1986; Frederico et al., 1990
	50°C	1.0×10^{-9}/sec	Frederico et al., 1990
Deamination of adenine	110°C	4.0×10^{-8}/sec	Karran and Lindahl, 1980
Depyrimidination	95°C	2.0×10^{-8}/sec	Lindahl and Karlström, 1973
Depurination	95°C	1.5×10^{-7}/sec	Wang et al., 1982

* Each is a published rate for single-stranded DNA.

deamination of cytosine, resulting in G:U pairs, has been shown to reduce the thermal stability by about 2.2°C (Ullman and McCarthy, 1973a). Reduction in the thermal stabiltiy has also been observed with depurination (Ullman and McCarthy, 1973b). These studies suggest heat-induced modifications in a chosen target in the eluted samples of mutant/wild-type heteroduplexes after pre-PCR mutant enrichment by CDCE. These modified species can appear as background noise upon mutational analysis by CDCE, as Section 4.5.2.10 discusses.

DNA modification fractions introduced by each procedure in the proposed point mutation detection method were estimated. Those procedures prior to pre-PCR mutant enrichment were investigated. Those procedures that require incubation at temperatures at or below 37°C were excluded from this investigation. For example, deamination of cytosine has been shown to be negligible in those samples incubated at 37°C for up to 5 days (Wang et al., 1982). The same has been observed for depurination after incubation at 37°C for 2 months (Lindahl and Nyberg, 1972).

As Table 20 summarizes, target renaturation is the procedure that generates deaminated cytosine at the highest fraction. This observation is expected to be true for the other modified species due to the nature of target renaturation requiring sample exposure to heat for 16 hr (see 4.5.2.5.2 for the renaturation efficiencies over different incubation time periods). Thus, available renaturation methods that require incubation at temperatures below 55°C were sought, and three were found.

The first method uses denaturants, such as formamide and urea. Addition of denaturants to renaturation mixtures has been shown to decrease the T_m of the substrate DNA (McConaughy et al., 1969; Blüthmann et al., 1973; Hutton, 1977). However, no simple linear relationship has been established between mean thermal stabilities of substrates and denaturant concentrations (Schmeckpeper and Smith, 1972). In addition, a reduced renaturation rate is expected in mixtures with denaturants compared to those without. This reduction has been explained by the increased solution viscosities as a result of adding denaturants to reaction mixtures (Thrower and Peacocke, 1968; Schmeckpeper and Smith, 1972; Hutton, 1977). Based on this reduced rate, longer sample exposure to heat is expected for those samples with denaturants compared to those without. For this reason, this method was rejected.

Table 20. Expected fractions of heat-induced DNA modifications

Procedure (ss or ds) [a & b]	Temperature/ duration	Expected fractions [c & d]			
		Deamination of cytosine	Deamination of adenine	De-pyrimidination	De-purination
Genomic DNA isolation (ds)	50°C/3 hr				
	100°C/1 min	$>1.2 \times 10^{-5}$	$<2.4 \times 10^{-6}$	$>1.2 \times 10^{-6}$	$>9 \times 10^{-6}$
Target isolation (ss)	70°C/4 min	2.9×10^{-6}			
	60°C/2 hr	$>7.2 \times 10^{-6}$			
	50°C/30 min	1.8×10^{-6}			
Target renaturation (ss)	55°C/16 hr	$>5.8 \times 10^{-5}$			

[a] Listed are those procedures that require incubation at temperatures above 37°C prior to pre-PCR mutant enrichment in the proposed point mutation detection method.
[b] Single- (ss) or double-stranded (ds) DNA conformation in each procedure.
[c] Based on the published rates at each temperature (see Table 19).
[d] Those fractions left blank are due to the lack of published rates at desired temperatures.

The second method, phenol emulsion reassociation technique, allows renaturation at or below room temperature in sample mixtures containing phenol (Kohne et al., 1977). Since phenol is thought to be a potential mutagenic agent, this method was rejected.

The third method uses DMSO and this method was chosen as an alternative to the established method that uses the incubation temperature of 55°C. Addition of 40% DMSO to renaturation mixtures has been shown to decrease the T_m of the substrate DNA by 27°C (Escara and Hutton, 1980). This addition is expected to increase the solution viscosities, lowering the renaturation rate. However, the rate has been shown to increase as the concentration of DMSO is increased (Escara and Hutton, 1980). The effect of DMSO on the solution dielectric constant has been suggested as a cause of this observation contradictory to the expectation (Escara and Hutton, 1980).

DMSO has been used as a cryoprotective agent (Ashwood-Smith, 1979; Greene et al., 1970). A number of studies have been performed to characterize this agent. No toxic effects have been observed on HeLa cells exposed to 5% DMSO for 24 hr at 37°C (Greene et al., 1970). No genetic changes have been observed over extended periods of storage using 10% DMSO (Ashwood-Smith, 1979; Ashwood-Smith, 1985). In addition, it has been shown that DNA cleavage is about 120-fold more frequent in 0% DMSO at T_m+10°C than in 40% DMSO at T_m+10°C (Escara and Hutton, 1980). Depurination has been shown to be about the same in these two variables (Escara and Hutton, 1980). These studies suggest that the renaturation method using DMSO can potentially reduce formation of heat-induced DNA modifications during target renaturation.

As a second means of target renaturation, incubation was performed at 28°C in mixtures containing 40% DMSO. The efficiency of this method over different incubation time periods is discussed in the following section.

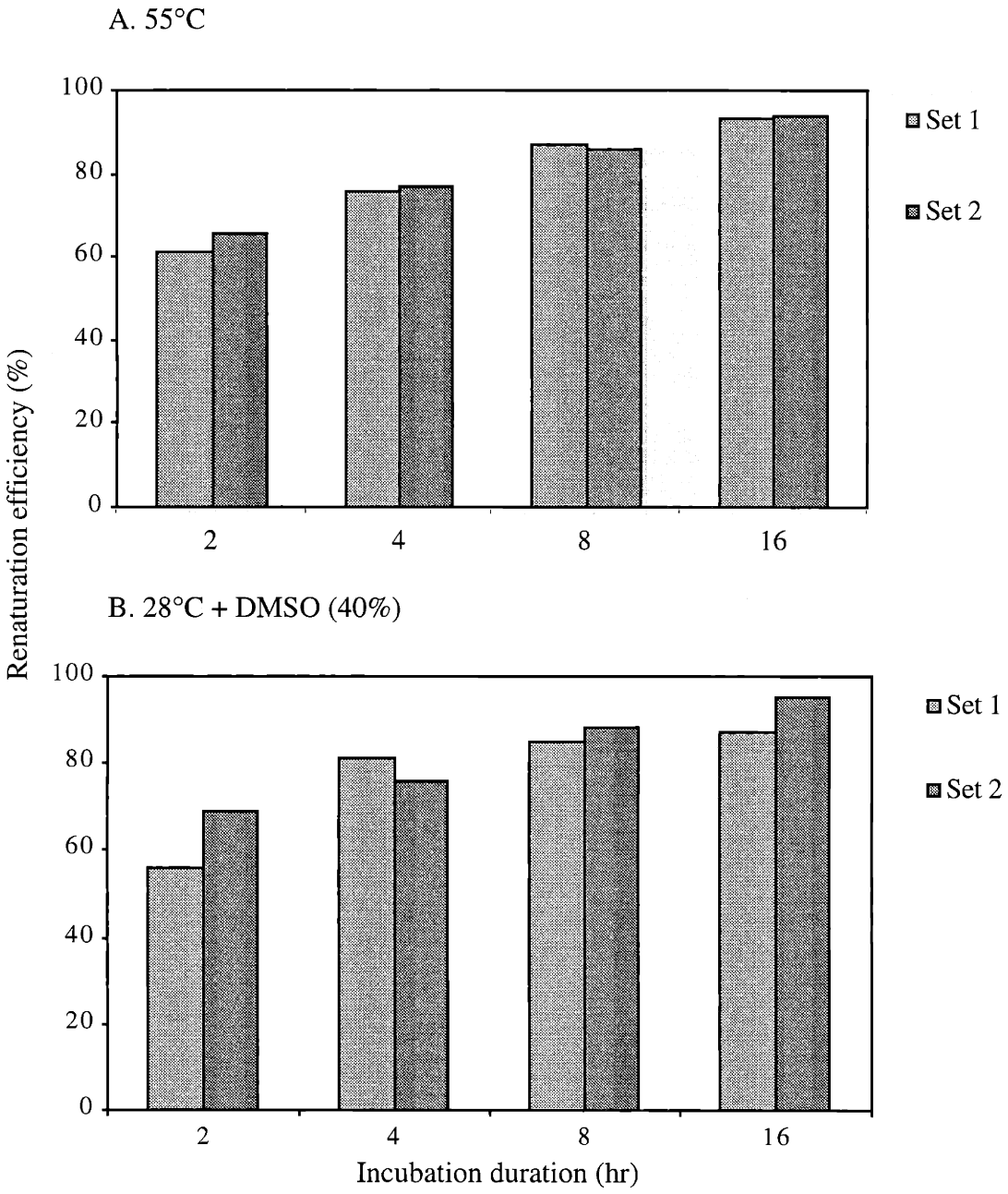
4.5.2.5.2 Efficiency

Figure 35 summarizes efficiencies of two renaturation methods over different incubation time periods. These results demonstrate that a renaturation efficiency of about 90% can be achieved after 16 hr-long incubation using either method. The other 10% is thought to represent the target that never renatured. Some of the substrate DNA has been

Figure 35. Renaturation efficiencies vs. incubation duration*

A chosen target was PCR was amplified using the IS1/IS2 primer pair, and this amplified target at a concentration of 10^8/μl was denatured. Using different test incubation time periods, renaturation was performed at 55°C (A) or at 28°C in the presence of 40% DMSO in the renaturation reaction mixture (B). Each renatured sample was restriction digested by *Apo* I, and the digestion efficiency served as a reference to measure the renaturation efficiencies over different incubation time periods.

* The substrate without denaturation was used as a positive control (*Apo* I digestion efficiency: >95%). The substrate with denaturation and with 0 hr incubation duration was used as a negative control (*Apo* I digestion efficiency: 0%).


A. 55°C
100
80
60
40
20
0
2
4
8
16
Set 1
Set 2
Renaturation efficiency (%)
B. 28°C + DMSO (40%)
100
80
60
40
20
0
2
4
8
16
Set 1
Set 2
Incubation duration (hr)

shown never to renature (Thrower and Peacocke, 1968). The inability to achieve complete renaturation has also been explained by phosphate-ester backbone breakage of substrate DNA during denaturation and renaturation (Marmur and Doty, 1961).

4.5.2.6 Restriction digestion and clamp ligation

The restriction digestion efficiency of *Apo* I was estimated to be about 80% (ligated clamp). The same digestion efficiency was estimated for *Ahd* I and *Hinf* I (natural clamp). These efficiencies are about 10% lower than those observed in another study using the samples of target isolation, followed by target renaturation (Li-Sucholeiki, X.-C., personal communication).

Target isolation is thought to be a cause of this decrease. If this procedure isolates the Watson and Crick strands of a target with unequal efficiencies, a decrease in restriction digestion efficiency is expected. For example, digestion efficiencies of up to 81% (up to 90% of the 0.9 Crick strand) are expected in target-isolated samples containing the Watson and Crick strands of a target at a molar ratio of 1 to 0.9. This expectation is based a previous observation in which up to about 90% renaturation efficiencies were achieved in samples containing the Watson and Crick strands of a target at an equal molar ratio (see Section 4.5.2.5.2).

Clamp ligation was performed with each sample of *Apo* I restriction digestion. A clamp copy number of 10^{11} per 10^{7} target copy number was determined to be necessary to generate the maximum clamp ligation efficiency by mass action. This clamp copy number was determined by clamp dilution experiments, as Section 4.5.1.2 discusses.

4.5.2.7 Pre-PCR mutant enrichment

The efficiency of pre-PCR mutant enrichment was estimated after CDCE, followed by CE. In addition to the wild-type in single strands (see Section 4.5.1.3.2), two more potential wild-type-containing species are expected in the CDCE-eluted samples of mutant/wild-type heteroduplexes using the initial samples of human cells. The first is the wild-type that may be renatured with non-target residual cellular DNA in target-isolated

samples. The second is the wild type-embedded *Bst*N I/*Dra* I fragment left uncleaved after the second restriction digestion. These wild-type-containing species may co-migrate with mutant/wild-type heteroduplexes under optimal CDCE-separation conditions. As a means to reduce the copy number of the wild-type-containing species in the CDCE-eluted samples, CE was performed. CE differentiated the targets of different lengths, allowing this reduction.

Initially, a capillary bore size of 320-μm id was used for pre-PCR mutant enrichment. Based on the loading capacity determined for this bore size (2.5 μg) (Li and Thilly, 1996), a 10^7 target copy number was estimated to be a suitable sample size. This target copy number was estimated to be equivalent to about 86 ng DNA, calculated using the efficiency of target isolation. Using the bore size of 320-μm id, the average efficiency of mutant enrichment was estimated to be 10-fold for the natural clamp procedure. This is about 12-fold lower than that observed in the initial samples of CDCE-purified wild-type DNA. These initial samples contain up to about 600 ng of human genomic DNA.

One possible cause of this decrease is that the bore size of 320-μm id is not able to handle impurities in samples accumulated prior to pre-PCR mutant enrichment. Such impurities, when loaded into a capillary, may cause a "peak-broadening effect," resulting in more of the wild-type co-migrating with mutant/wild-type heteroduplexes under optimal CDCE-separation conditions. When a capillary bore size of 540-μm id (loading capacity: 10 μg (Li and Thilly, 1996)) was used instead, the average mutant enrichment efficiency increased to 60 ± 8.3-fold. An average efficiency of 19 ± 4-fold was estimated for the ligated clamp procedure using the capillary bore size of 540-μm id.

These measured efficiencies using the bore size of 540-μm id are about 2-fold lower compared to those in the initial samples of CDCE-purified wild-type DNA. A similar observation has been reported (between >1- and >2-fold lower efficiencies) (Li-Sucholeiki and Thilly, 2000). These lower efficiencies observed for the ligated and natural clamp procedures predict a 2-fold lower mutation detection sensitivity, as Section 4.5.1.3.2 discusses. This decrease in efficiency is possibly caused by variables introduced by the different initial samples: CDCE-purified wild-type vs. wild-type in human cells.

Using the capillary bore size of 540-μm id and using the initial samples of human cells, the average efficiency of pre-PCR mutant enrichment observed for the natural clamp procedure is 2- to 3-fold lower than that in a previous study (Li-Sucholeiki and Thilly, 2000). This decrease is possibly caused by the use of two different target sequences. Some sequences possibly generate higher efficiencies than other sequences. However, more sequences need to be investigated to validate this assumption.

4.5.2.8 Hifi-PCR and post-PCR mutant enrichment

The average efficiency of post-PCR mutant enrichment by CDCE was estimated to be 15-fold for the ligated and natural clamp procedures. This efficiency agrees with that observed in the initial samples of CDCE-purified wild-type DNA.

4.5.2.9 Sensitivity and accuracy

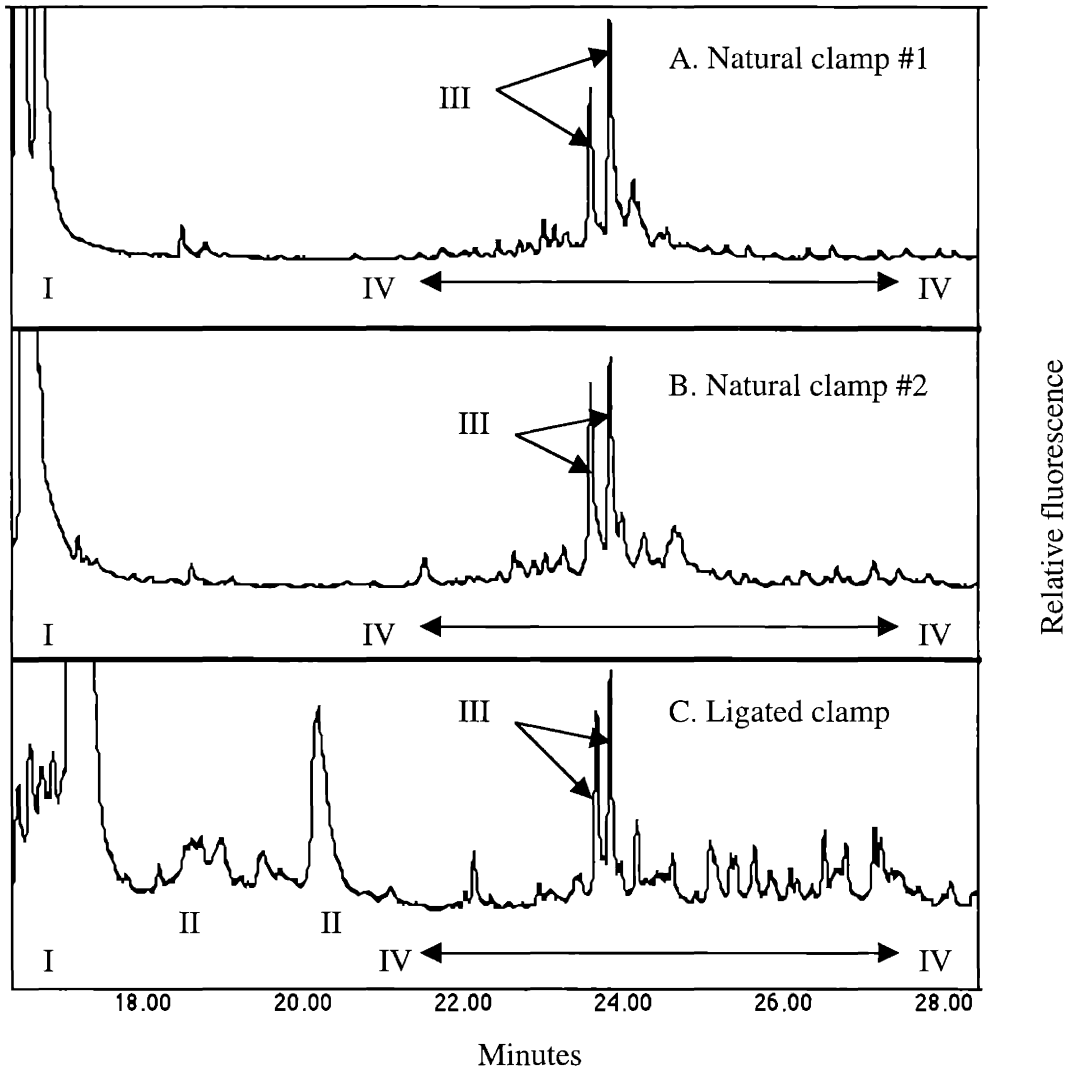
Figure 36 illustrates mutation detection by CDCE. A mutation detection sensitivity of 5×10^{-6} was estimated for the natural clamp procedure, and 2×10^{-5} for the ligated clamp procedure. The sensitivity of the ligated clamp procedure represents detection of as few as 200 copies of each mutant in the presence of 10^7 copies of the wild-type. These sensitivities are slightly lower compared to those in the initial samples of CDCE-purified wild-type DNA. Section 4.5.2.7 predicts this decrease.

Using the initial samples of human cells, the sensitivity estimated for the natural clamp procedure is 5-fold lower that that in a previous study (5×10^{-6} vs. 10^{-6}/bp) (Li-Sucholeiki and Thilly, 2000). However, a direct comparison cannot be made since different target sequences were used. Sensitivity is related to the efficiency of pre-PCR mutant enrichment and to the fidelity of *Pfu* DNA polymerase, as Sections 4.5.1.3.2 and 4.5.1.6 discuss. These relationships can generate various sensitivities using different target sequences. However, more sequences need to be investigated to understand the relationship between the sensitivity and the use of different target sequences.

Figure 37 illustrates CDCE-mutation detection of mutant internal standards added at different initial fractions ranging from 3 to 27×10^{-5} using the ligated clamp procedure.

Figure 36. Mutation detection by CDCE: ligated vs. natural clamp
(initial samples of human cells)

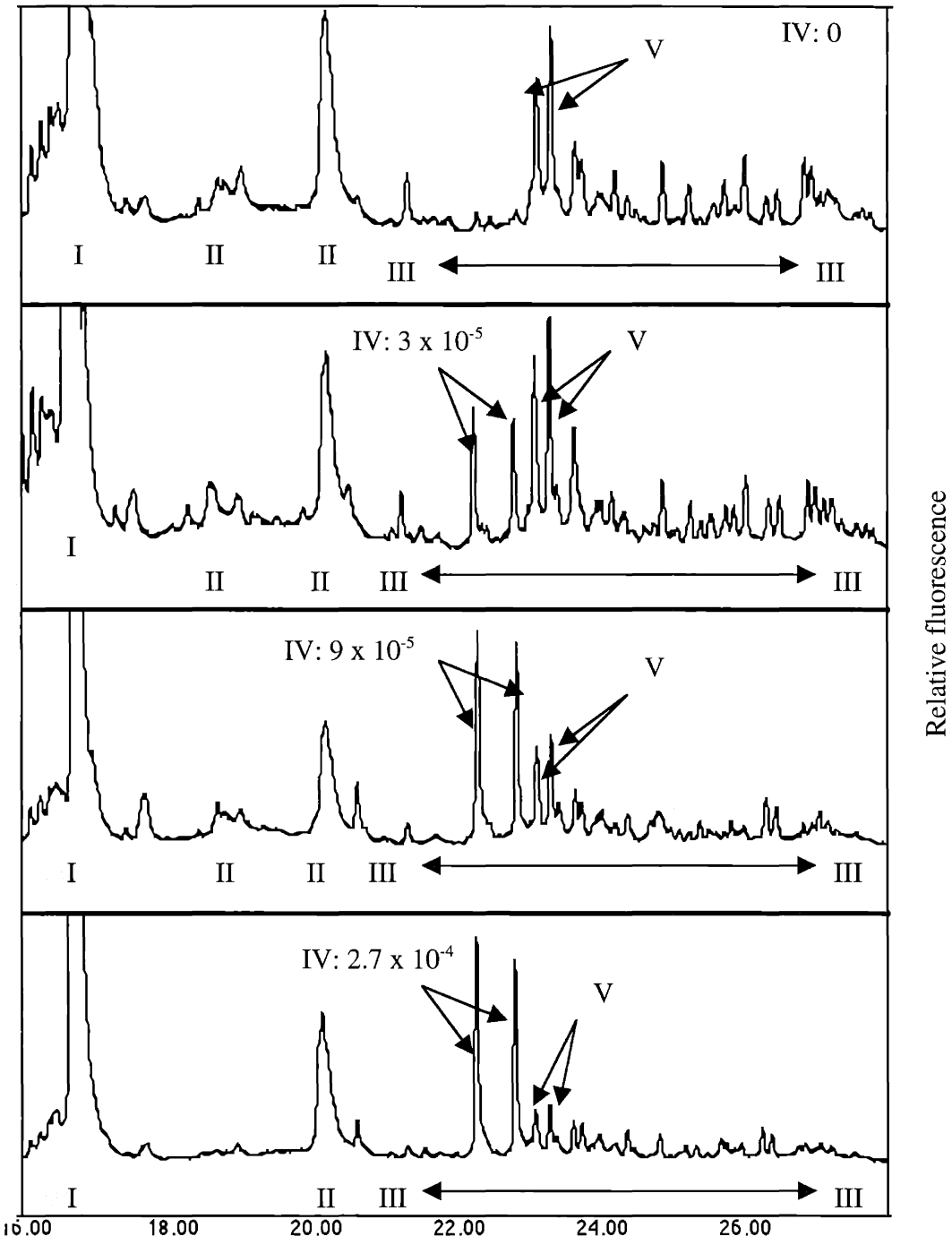
Genomic DNA isolation, restriction digestion with *Bst*N I and *Dra* I, and target isolation were performed with each initial sample of human cells. For the natural clamp procedure (A and B), pre-PCR mutant enrichment by CDCE, followed by CE, was performed. For the ligated clamp procedure (C), *Apo* I restriction digestion and clamp ligation were performed prior to pre-PCR mutant enrichment. PCR was performed with each mutant-enriched sample using either the P3/P1 or GC1/P1 primer pair, depending on the clamp type desired. This procedure was followed by post-PCR mutant enrichment by CDCE. Final-mutant enriched samples were PCR amplified prior to mutational analysis by CDCE.



I represents the wild-type in homoduplex; II represents PCR by-products; III represents the mutant internal standard (IS: G to A transition at *HPRT* cDNA bp 309) added at an initial fraction of 5 x 10^{-5}; IV represents a region where the majority of the background mutants migrate; the CDCE-separated mutants are in mutant/wild-type heteroduplexes; A and B are independently performed experimental results.

Figure 37. CDCE-mutation detection vs. mutant fractions: ligated clamp (initial samples of human cells)

Genomic DNA isolation, restriction digestion with *Bst*N I and *Dra* I, and target isolation were performed with each initial sample of human cells. *Apo* I restriction digestion and clamp ligation were then performed prior to pre-PCR mutant enrichment by CDCE and CE. PCR was performed with each mutant-enriched sample using the GC1/P1 primer pair, followed by post-PCR mutant enrichment by CDCE. Final-mutant enriched samples were PCR amplified and analyzed by CDCE.



I represents the wild-type in homoduplex; II represents PCR by-products; III represents a region where the majority of the background mutants migrate; IV represents the mutant internal standard of genomic DNA (IS1: G -> T transversion at *HPRT* cDNA bp 312) added at an initial fraction of 3 x 10^{-5}, 9 x 10^{-5}, or 2.7 x 10^{-4}; V represents the mutant internal standard of PCR products (IS2: G to A transition at *HPRT* cDNA bp 309) added at an initial fraction of 5 x 10^{-5}; the CDCE-separated mutants are in mutant/wild-type heteroduplexes.

Figure 38 illustrates the quantitative analysis of Figure 37. These results demonstrate the accuracy of the ligated clamp procedure.

4.5.2.10 Source of background noise

Section 4.5.1.6 discusses *Pfu* DNA polymerase as the source of background noise using the initial samples of CDCE-purified wild-type DNA. Using the initial samples of human cells, additional sources of background noise are expected. Section 4.5.2.5.1 discusses heat-induced DNA modifications. In addition, ethanol and UV light have been identified as causes of DNA modifications (Hanekamp, 1993). These modifications in pre-PCR mutant-enriched samples can be PCR-amplified and, as a result, appear as background noise upon mutational analysis by CDCE.

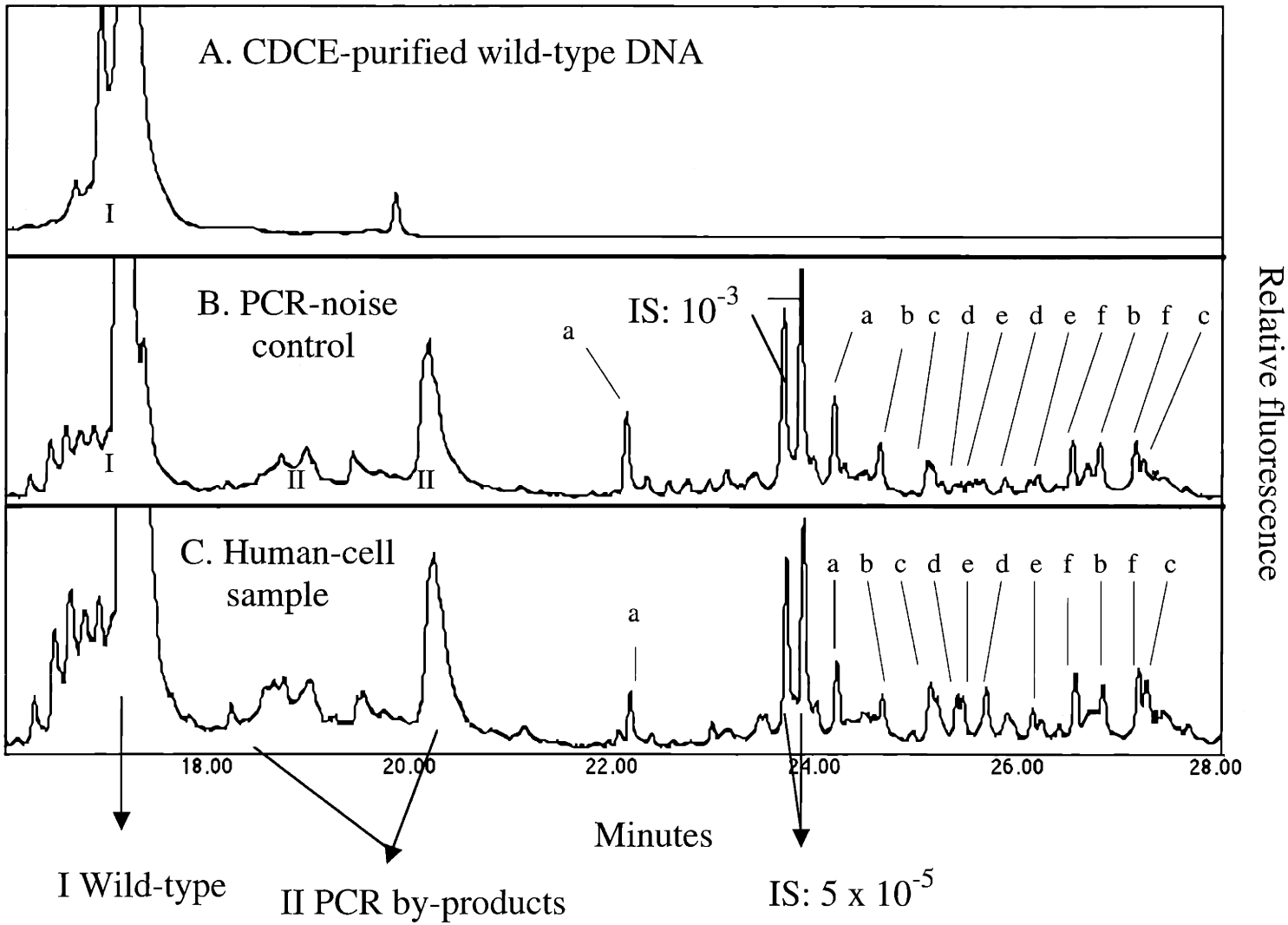
For example, deamination of cytosine, forming uracil, can result in C to T transition. When adenine is converted to hypoxanthine as a result of deamination, this lesion forms a more stable base pair with cytosine than with thymine, resulting in A to G transition (Lindahl, 1979; Lindahl, 1993). Although apurinic/apyrimidinic (AP, abasic) sites can block DNA synthesis (Sagher and Strauss, 1985; Lindahl, 1993), bypassing of these sites during DNA synthesis can be mutagenic (Shearman and Loeb, 1979; Schaaper et al., 1983; Kunkel, 1984; Loeb, 1985; Loeb and Preston, 1986).

Section 4.5.2.5.1 identifies target renaturation as the procedure that generates heat-induced DNA modifications at the highest fractions in the proposed point mutation detection method. For this reason, the level of background noise using the established target renaturation method was compared to that of the alternative method. The established method uses an incubation temperature of 55°C, and the alternative method uses a temperature of 28°C with 40% DMSO in reaction mixtures (see Section 4.5.2.5). These two methods generated no difference in the level of background noise. This observation suggests that the established target renaturation method, requiring incubation at 55°C for 16 hr, is not the source of background noise.

Using the initial samples of human cells, *Pfu* DNA polymerase has been identified as the source of background noise (Li-Sucholeiki and Thilly, 2000). In this present study, three observations were made to confirm the previous study.

Figure 38. Accuracy of mutation detection by CDCE: ligated clamp

Genomic DNA isolation, restriction digestion with *Bst*N I and *Dra* I, and target isolation were performed with each initial sample of human cells. *Apo* I restriction digestion and clamp ligation were then performed prior to pre-PCR mutant enrichment by CDCE and CE. PCR was performed with each mutant-enriched sample using the GC1/P1 primer pair, followed by post-PCR mutant enrichment by CDCE. Final-mutant enriched samples were PCR amplified and analyzed by CDCE. Each output was estimated by comparing the area of IS1 (G -> T transversion at *HPRT* cDNA bp 312) added at an initial fraction of 3×10^{-5}, 9×10^{-5}, or 2.7×10^{-4} to that of IS2 added at an initial fraction of 5×10^{-5} (I): (IS1 ÷ IS2) x I; the average area of the background mutants (B) was compared to the area of IS2 to estimate the average background mutant fraction in each sample: (B ÷ IS2) x I.


A. CDCE-purified wild-type DNA
I
B. PCR-noise control
IS: 10-3
a
a b c d e d e f b f c
I
II
II
C. Human-cell sample
a
a b c d e d e f b f c
18.00
20.00
22.00
24.00
26.00
28.00
Minutes
Relative fluorescence
I Wild-type
II PCR by-products
IS: 5 x 10-5

First, the efficiencies of pre-PCR mutant enrichment and the sensitivities were observed to correlate. Section 4.5.1.6 discusses this correlation as being an indication that the fidelity of *Pfu* determines the level of background noise.

Second, the level of background noise in the initial samples of human cells was observed to increase as the number of doublings increased. This observation was previously demonstrated using the initial samples of CDCE-purified wild-type DNA in Section 4.5.1.6.

Third, the background mutants in a PCR-noise control sample were observed to co-migrate with those in the initial samples of human cells under optimal CDCE-separation conditions. Figure 39A represents CDCE-purified wild-type DNA used as a template to prepare the PCR-noise control. Figure 39B represents the PCR-noise control, prepared by amplifying the chosen target using *Pfu*. Figure 39C represents the initial samples of human cells using the ligated clamp procedure. A set of co-migrating mutants was purified and sequenced. Sequencing analysis showed that each of the co-migrating mutants is the same with regard to position and kind. Figure 40 and Table 21 summarize the sequencing results.

The predominant *Pfu*-mutations in the chosen *HPRT* target were identified to be 4 G to T tranversions, 1 G to A transition, and 1 A to G transition. As Section 2.5.3.1 discusses, G to T transversion has been determined as the predominant mutation type of *Pfu*. The mutations identified in the chosen *HPRT* target represent about 60% of the total *Pfu*-mutations. These mutations appeared at an average fraction equivalent to about 2 x 10^{-5} in original samples. This average mutant fraction limited the sensitivity of the ligated clamp procedure.

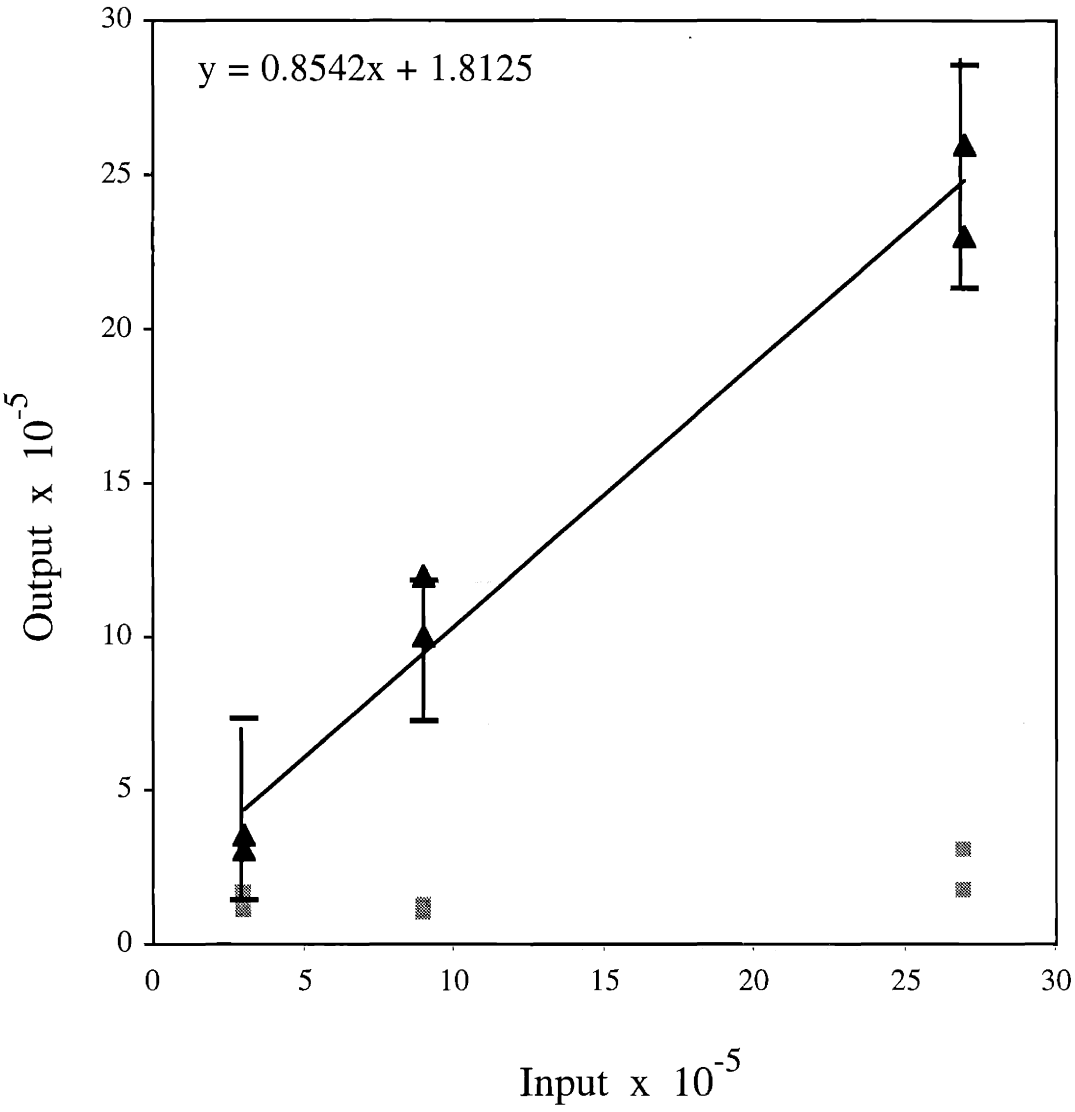
As this section has discussed, heat-induced DNA modifications can appear as background noise. For example, deaminated cytosine is expected to appear in the chosen target at a fraction greater than 5.8 x 10^{-5} in pre-PCR mutant-enriched samples (see Table 20 in Section 4.5.2.5.1). This expectation is greater than the average mutant fraction observed for the predominant *Pfu* mutations.

To date, a limited number of studies are available to explain why *Pfu*, not target renaturation, is the source of background noise. *Pfu* has been shown to specifically recognize the presence of uracil in single-stranded DNA and stall DNA synthesis (Lasken

Figure 39. Source of background noise: *Pfu* DNA polymerase*

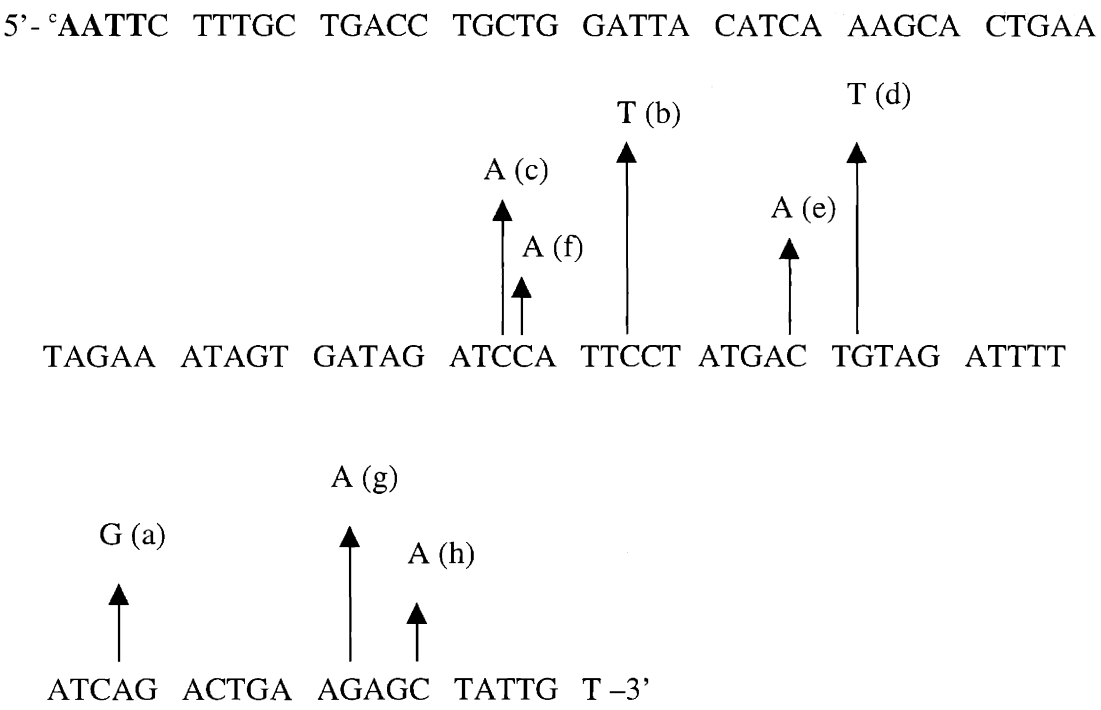
A. PCR was performed using the GC1/P1 primer pair to amplify the chosen *HPRT* target with the ligated clamp, followed by CDCE purification of the wild-type DNA.
B. A mixture containing the CDCE-purified wild-type DNA (Figure 39A) and the mutant internal standard (G to A transition at *HPRT* cDNA bp 309) at a fraction of 10^{-3} was PCR amplified using the GC1/P1 primer pair. Post-PCR mutant enrichment, followed by PCR, was performed prior to mutational analysis by CDCE.
C. Genomic DNA isolation, restriction digestion with *Bst*N I and *Dra* I, and target isolation were performed with the initial sample of human cells. *Apo* I restriction digestion and clamp ligation were performed with the target-isolated sample containing the mutant internal standard (G to A transition at *HPRT* cDNA bp 309) added at an initial fraction of 5×10^{-5}. These procedures were followed by pre-PCR mutant enrichment by CDCE and CE. PCR was performed with the mutant-enriched sample using the GC1/P1 primer pair, followed by post-PCR mutant enrichment by CDCE. The final-mutant enriched sample was PCR amplified and analyzed by CDCE.

* I represents the wild-type in homoduplex; II represents PCR by-products; the CDCE-separated mutants in mutant/wild-type heteroduplexes; Figure 40 and Table 21 summarize the positions and kinds of mutants.



The line represents least squares regression, generated as a result of using 6 independent sets of input and output mutant fractions; the error bars (95% confidence interval) are for the fitted mutant fractions by liner regression; each set of two triangles represents the corresponding output mutant fractions of each input mutant fraction; each set represents duplicate experiments; the squares represent the average background mutant fractions in individual samples.

Figure 40. Positions and kinds of mutations in chosen *HPRT* target



* cDNA bp 218-318; the group of bases in bold represents *Apo* I restriction end to which the clamp was ligated; see Table 21 for the source of each mutation.

Table 21. Summary of mutations in chosen *HPRT* target

Mutant	Position	Kind	Source
c	275	G -> T	*Pfu*
f	276	G -> T	*Pfu*
b	280	G -> A	*Pfu*
e	287	G -> T	*Pfu*
d	289	G -> T	*Pfu*
a	301	A -> G	*Pfu*
g	309	G -> A	Internal standard of PCR products
h	312	G -> T	Internal standard of genomic DNA

et al., 1996; Greagg et al., 1999). As a result, the major primer-extension product was observed to be significantly shorter than an expected full-length product (Greagg et al., 1999). This observation was made for both the 3' to 5' proofreading exonuclease proficient and deficient versions of *Pfu* (Greagg et al., 1999). On the other hand, the primer extension to a full-length product was observed when tested with *Taq* DNA polymerase (Greagg et al., 1999). These results suggest that the shorter primer extension is specific to *Pfu*. More studies investigating the effects of *Pfu* on other heat-induced DNA modifications are necessary. These studies may answer why *Pfu* is the source of background noise for the ligated and natural clamp procedures.

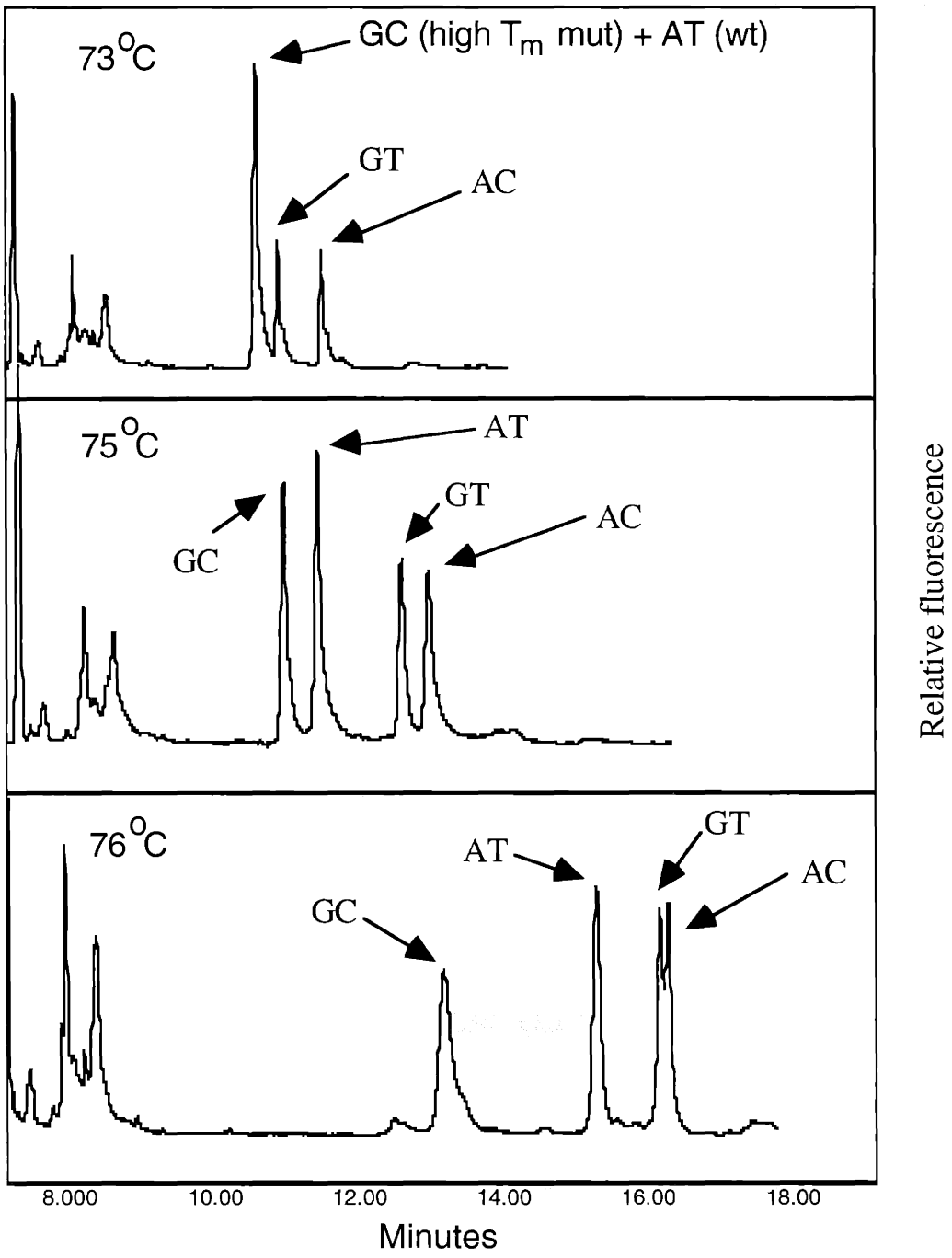
4.6 Demonstration of CDCE separation: *p53* exon 7

Many mutations in the *p53* gene have been found in various types of human cancer (Hainaut et al., 1998; Hernandez-Boussard et al., 1999b). For this reason, a part of the human *p53* gene (cDNA bp 673-782), representing the entire exon 7, was chosen as a second target. Since this target does not have a neighboring natural clamp, clamp ligation is necessary for mutational analysis by CDCE.

Figure 41 illustrates CDCE separation results using the ligated clamp at different temperatures. These results demonstrate the ability of the ligated clamp to separate point mutations in the chosen *p53* target from the wild-type.

Figure 41. CDCE separation of high T_m mutant from wild-type sequences: *p53* exon 7 with ligated clamp

A mutant with a higher melting temperature (high T_m mutant, A -> G transition at *p53* cDNA bp 763) than that of the wild-type was PCR amplified using the GC2/BS-stable primer pair. An additional PCR was performed with a mixture containing the wild-type and the PCR-amplified mutant at an equal molar ratio. The GC2/BS primer pair (*p53* cDNA bp 673-782) was used for this PCR in which mutant/wild-type heteroduplexes, in addition to the homoduplex of each kind, were formed. This PCR sample was separated by CDCE, performed using a 15 cm-long temperature-regulated zone of desired temperatures.


73°C
GC (high T_m mut) + AT (wt)
GT
AC
75°C
AT
GT
AC
GC
76°C
AT
GT
AC
GC
8.000
10.00
12.00
14.00
16.00
18.00
Minutes
Relative fluorescence

5 CONCLUSIONS

For mutational analysis of a larger pool of target genes and tissues, genotype-based methods are preferred over phenotype-based methods. This preference, combined with the ability to measure mutations at low fractions, allows analysis of a variety of human samples, including normal tissues.

For such analysis, a mutation assay was developed. The key development was high efficiency DNA ligation in which a clamp is ligated to any 100-bp sequence of interest by mass action. Clamp ligation was combined with the established CDCE-based mutation detection method developed for DNA sequences with a natural clamp. The established method alone allows analysis of rare point mutations in only about 9% of the human genome. An additional 89% of the human genome can be analyzed by the combined method.

The sensitivity of the combined method was demonstrated to be 2×10^{-5} in human cells. This demonstration represents detection of at least 200 copies of each mutant in the presence of 10^7 copies of the wild-type. This sensitivity was observed to be 4-fold lower than that of the established method, when sensitivities were compared using the same target, a part of the human *HPRT* gene (cDNA-bp 223-318). The GC-rich sequence of the ligated clamp was determined to be the cause of this decrease.

The sensitivity of the combined method was limited by the fidelity of *Pfu* DNA polymerase used for PCR. The polymerase generated 4 G -> T transversions, 1 A -> G transition, and 1 G -> A transition in the chosen *HPRT* target. These mutations appeared at an average fraction equivalent to about 2×10^{-5} in the original samples. This average mutant fraction limited the sensitivity of the combined method.

Future applications of the combined method to analysis of a variety of human samples may allow understanding of human point mutagenesis.

6 SUGGESTED FUTURE STUDIES

6.1. Increase in mutation detection sensitivity

The fidelity of *Pfu* DNA polymerase limits the sensitivity of mutation detection methods. With increased fidelity, the sensitivity can be increased. Studies on mechanisms by which *Pfu* generates mutations in an amplified target may suggest a means to increase the fidelity.

Conventional *Pfu* is the 3' to 5' proofreading exonuclease-proficient version of *Pfu* (exo+ version) (Lundberg et al., 1991). Potentially, a small amount of the exonuclease deficient version of *Pfu* (exo- version) is present in batches of conventional *Pfu* (Thilly, W. G., personal communication). This mixed version may be caused by mutations occurring during transcription and translation of the exo+ version of the enzyme.

A previous observation supports the mixed-version possibility (Kim, unpublished results). In this observation, mutations in a chosen target created by conventional *Pfu* were shown to co-migrate with those created by the exo- version under optimal CDCE-separation conditions. These co-migrating mutations are thought to be the same mutations; however, sequencing analysis needs to be performed for verification.

This observation supports the hypothesis that a small amount of the exo- version present in the mixed version causes all the mutations, determining the fidelity of the mixed version. If this is true, purification of the exo+ version in the mixed version may increase fidelity (Thilly, W. G., personal communication).

6.2. Applications

6.2.1 Inherited mutations

As Section 2.6 describes, the ability to measure mutations at fractions down to 5×10^{-5} allows analysis of inherited mutations with statistical significance in a sample size of 10^7 cells, derived from 10^5 persons. Such analysis can be performed with the mutation assay

developed in this thesis. As a result, the relationships of genes to diseases can be tested and causative alleles over the entire genes can be discovered.

6.2.2 Somatic mutations

As Section 2.4 summarizes, a limited number of somatic mutational studies have been performed on the tissues of healthy individuals. Based on these studies, the mutation assay developed in this thesis is suggested for analysis of *p53* somatic point mutations in the lungs of healthy individuals. Two *p53* point mutations in the gene-coding region at an average fraction of about 5×10^{-5} have been observed in healthy human lungs (Li-Sucholeiki et al., Unpublished results).

The hypothesis that cigarette smoke induces mutations or/and mutant fractions can be directly tested by comparing mutational spectra in 100-bp *p53* sequences in smokers' and non-smokers' lungs. Since none of the human *p53* exons has a natural clamp, clamp ligation is necessary to test such a hypothesis.

7 REFERENCES

Abrams ES, Murdaugh SE, Lerman LS (1990) Comprehensive detection of single base changes in human genomic DNA using denaturing gradient gel electrophoresis and a GC clamp. Genomics 7: 463-75.

Aguilar F, Harris CC, Sun T, Hollstein M, Cerutti P (1994) Geographic variation of p53 mutational profile in nonmalignant human liver. Science 264: 1317-9.

Aguilar F, Hussain SP, Cerutti P (1993) Aflatoxin B1 induces the transversion of G-->T in codon 249 of the p53 tumor suppressor gene in human hepatocytes. Proc Natl Acad Sci U S A 90: 8586-90.

Akashi M, Koeffler HP (1998) Li-Fraumeni syndrome and the role of the p53 tumor suppressor gene in cancer susceptibility. Clin Obstet Gynecol 41: 172-99.

Akiyama M, Kyoizumi S, Hirai Y, Kusunoki Y, Iwamoto KS, Nakamura N (1995) Mutation frequency in human blood cells increases with age. Mutat Res 338: 141-9.

Albertini RJ, Castle KL, Borcherding WR (1982) T-cell cloning to detect the mutant 6-thioguanine-resistant lymphocytes present in human peripheral blood. Proc Natl Acad Sci U S A 79: 6617-21.

Albertini RJ, Hayes RB (1997) Somatic cell mutations in cancer epidemiology. IARC Sci Publ 142: 159-84.

Albertini RJ, Nicklas JA, Fuscoe JC, Skopek TR, Branda RF, O'Neill JP (1993) In vivo mutations in human blood cells: biomarkers for molecular epidemiology. Environ Health Perspect 99: 135-41.

Albertini RJ, Nicklas JA, O'Neill JP, Robison SH (1990) In vivo somatic mutations in humans: measurement and analysis. Annu Rev Genet 24: 305-26.

Albertini RJ, O'Neill JP, Nicklas JA, Heintz NH, Kelleher PC (1985) Alterations of the hprt gene in human in vivo-derived 6-thioguanine-resistant T lymphocytes. Nature 316: 369-71.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403-10.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389-402.

Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG (1981)

Sequence and organization of the human mitochondrial genome. Nature 290: 457-65.

Andre P, Kim A, Khrapko K, Thilly WG (1997) Fidelity and mutational spectrum of Pfu DNA polymerase on a human mitochondrial DNA sequence. Genome Res 7: 843-52.

Antequera F, Bird A (1993) CpG islands. Exs 64: 169-85.

Armstrong JD, Kunz BA (1990) Site and strand specificity of UVB mutagenesis in the SUP4-o gene of yeast [published erratum appears in Proc Natl Acad Sci U S A 1991 Mar 1;88(5):2035]. Proc Natl Acad Sci U S A 87: 9005-9.

Ashwood-Smith MJ (1979) Lack of genetic damage in mammalian cells after cryopreservation at -196 degrees C. Exp Hematol 7: 21-6.

Ashwood-Smith MJ (1985) Genetic damage is not produced by normal cryopreservation procedures involving either glycerol or dimethyl sulfoxide: a cautionary note, however, on possible effects of dimethyl sulfoxide. Cryobiology 22: 427-33.

Attardi G, Schatz G (1988) Biogenesis of mitochondria. Annu Rev Cell Biol 4: 289-333.

Baas IO, Mulder JW, Offerhaus GJ, Vogelstein B, Hamilton SR (1994) An evaluation of six antibodies for immunohistochemistry of mutant p53 gene product in archival colorectal neoplasms. J Pathol 172: 5-12.

Barnes WM (1994) PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. Proc Natl Acad Sci U S A 91: 2216-20.

Baserga R, Porcu P, Sell C (1993) Oncogenes, growth factors and control of the cell cycle. Cancer Surv 16: 201-13.

Bell JB, Eckert KA, Joyce CM, Kunkel TA (1997) Base miscoding and strand misalignment errors by mutator Klenow polymerases with amino acid substitutions at tyrosine 766 in the O helix of the fingers subdomain. J Biol Chem 272: 7345-51.

Bennett WP, Hussain SP, Vahakangas KH, Khan MA, Shields PG, Harris CC (1999) Molecular epidemiology of human cancer risk: gene-environment interactions and p53 mutation spectrum in human lung cancer. J Pathol 187: 8-18.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Rapp BA, Wheeler DL (2000) GenBank. Nucleic Acids Res 28: 15-8.

Benzer S, Freese E (1958) Induction of specific mutations with 5-bromouracil. Proc Nat Acad Sci USA 44: 112-119.

Benzer S (1961) On the topography of the genetic fine structure. Proc Nat Acad Sci USA 47: 403-415.

Bergstresser PR, Pariser RJ, Taylor JR (1978) Counting and sizing of epidermal cells in normal human skin. J Invest Dermatol 70: 280-4.

Beroud C, Collod-Beroud G, Boileau C, Soussi T, Junien C (2000) UMD (Universal mutation database): a generic software to build and analyze locus-specific databases. Hum Mutat 15: 86-94.

Beroud C, Joly D, Gallou C, Staroz F, Orfanelli MT, Junien C (1998) Software and database for the analysis of mutations in the VHL gene. Nucleic Acids Res 26: 256-8.

Beroud C, Soussi T (1998) p53 gene mutation: software and database. Nucleic Acids Res 26: 200-4.

Bluthmann H, Bruck D, Hubner L, Schoffski A (1973) Reassociation of nucleic acids in solutions containing formamide. Biochem Biophys Res Commun 50: 91-7.

Bookstein R, Lai CC, To H, Lee WH (1990) PCR-based detection of a polymorphic BamHI site in intron 1 of the human retinoblastoma (RB) gene. Nucleic Acids Res 18: 1666.

Bos JL (1989) ras oncogenes in human cancer: a review. Cancer Res 49: 4682-9.

Brail L, Fan E, Levin DB, Logan DM (1993) Improved polymerase fidelity in PCR-SSCPA. Mutat Res 303: 171-5.

Brash DE, Rudolph JA, Simon JA, Lin A, McKenna GJ, Baden HP, Halperin AJ, Ponten J (1991) A role for sunlight in skin cancer: UV-induced p53 mutations in squamous cell carcinoma. Proc Natl Acad Sci U S A 88: 10124-8.

Brown MA (1997) Tumor suppressor genes and human cancer. Adv Genet 36: 45-135.

Cariello NF, Douglas GR, Gorelick NJ, Hart DW, Wilson JD, Soussi T (1998) Databases and software for the analysis of mutations in the human p53 gene, human hprt gene and both the lacI and lacZ gene in transgenic rodents. Nucleic Acids Res 26: 198-9.

Cariello NF, Keohavong P, Kat AG, Thilly WG (1990) Molecular analysis of complex human cell populations: mutational spectra of MNNG and ICR-191. Mutat Res 231: 165-76.

Cariello NF, Scott JK, Kat AG, Thilly WG, Keohavong P (1988) Resolution of a missense mutant in human genomic DNA by denaturing gradient gel electrophoresis and direct sequencing using in vitro DNA amplification: HPRT Munich. Am J Hum Genet 42: 726-34.

Cariello NF, Skopek TR (1993a) In vivo mutation at the human HPRT locus. Trends Genet 9: 322-6.

Cariello NF, Skopek TR (1993b) Mutational analysis using denaturing gradient gel electrophoresis and PCR. Mutat Res 288: 103-12.

Cariello NF, Swenberg JA, Skopek TR (1991) Fidelity of Thermococcus litoralis DNA polymerase (Vent) in PCR determined by denaturing gradient gel electrophoresis. Nucleic Acids Res 19: 4193-8.

Cariello NF, Thilly WG (1986) Use of gradient denaturing gels to determine mutational spectrum in human cells. Basic Life Sci 38: 439-52.

Caruthers MH, Beaucage SL, Becker C, Efcavitch JW, Fisher EF, Galluppi G, Goldman R, deHaseth P, Matteucci M, McBride L, et al. (1983) Deoxyoligonucleotide synthesis via the phosphoramidite method. Gene Amplif Anal 3: 1-26.

Cerutti P, Hussain P, Pourzand C, Aguilar F (1994) Mutagenesis of the H-ras protooncogene and the p53 tumor suppressor gene. Cancer Res 54: 1934s-1938s.

Cha RS, Thilly WG (1993) Specificity, efficiency, and fidelity of PCR. PCR Methods Appl 3: S18-29.

Cha RS, Zarbl H, Keohavong P, Thilly WG (1992) Mismatch amplification mutation assay (MAMA): application to the c-H- ras gene. PCR Methods Appl 2: 14-20.

Chen J, Sahota A, Stambrook PJ, Tischfield JA (1991) Polymerase chain reaction amplification and sequence analysis of human mutant adenine phosphoribosyltransferase genes: the nature and frequency of errors caused by Taq DNA polymerase. Mutat Res 249: 169-76.

Chen J, Thilly WG (1994) Use of denaturing-gradient gel electrophoresis to study chromium- induced point mutations in human cells. Environ Health Perspect 102 Suppl 3: 227-9.

Chen J, Thilly WG (1996) Mutational spectra vary with exposure conditions: benzo[a]pyrene in human cells. Mutat Res 357: 209-17.

Chen JX, Zheng Y, West M, Tang MS (1998) Carcinogens preferentially bind at methylated CpG in the p53 mutational hot spots. Cancer Res 58: 2070-5.

Chen ZY, 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: 191-4.

Cherpillod P, Amstad PA (1995) Benzo[a]pyrene-induced mutagenesis of p53 hot-spot codons 248 and 249 in human hepatocytes. Mol Carcinog 13: 15-20.

Chinnery PF, Turnbull DM (1999) Mitochondrial DNA and disease. Lancet 354 Suppl 1: SI17-21.

Chiocca SM, Sandy MS, Cerutti PA (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 U S A 89: 5331-5.

Cho Y, Gorina S, Jeffrey PD, Pavletich NP (1994) Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. Science 265: 346-55.

Cline J, Braman JC, Hogrefe HH (1996) PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. Nucleic Acids Res 24: 3546-51.

Cohen AS, Najarian DR, Karger BL (1990) Separation and analysis of DNA sequence reaction products by capillary gel electrophoresis. J Chromatogr 516: 49-60.

Cole J, Skopek TR (1994) International Commission for Protection Against Environmental Mutagens and Carcinogens. Working paper no. 3. Somatic mutant frequency, mutation rates and mutational spectra in the human population in vivo. Mutat Res 304: 33-105.

Coller HA, Thilly WG (1994) Development and applications of mutational spectra technology. Environ Sci Technol 28: 478A-487A.

Coller HA, Thilly WG (1998) Mutational spectra technology. In Encyclopedia of Environmental Analysis and Remediation, John Wiley & Sons, Inc.. New York, pp. 2967-3005.

Coller HA, Khrapko K, Torres A, Frampton MW, Utell MJ, Thilly WG (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: 1268-77.

Collins FS, Guyer MS, Charkravarti A (1997) Variations on a theme: cataloging human DNA sequence variation. Science 278: 1580-1.

Connolly BA, Eckstein F, Pingoud A (1984) The stereochemical course of the restriction endonuclease EcoRI-catalyzed reaction. J Biol Chem 259: 10760-3.

Cooper DN, Krawczak M (1993) The methodology of mutation detection. In Human Gene Mutation, BIOS Scientific Publishers Limited, Oxford, UK, pp. 61-84.

Cooper DN, Ball EV, Krawczak M (1998) The human gene mutation database. Nucleic Acids Res 26: 285-7.

Costes B, Girodon E, Ghanem N, Chassignol M, Thuong NT, Dupret D, Goossens M (1993) Psoralen-modified oligonucleotide primers improve detection of mutations by denaturing gradient gel electrophoresis and provide an alternative to GC-clamping. Hum Mol Genet 2: 393-7.

Cottrell DA, Blakely EL, Borthwick GM, Johnson MA, Taylor GA, Brierley EJ, Ince PG, Turnbull DM (2000) Role of mitochondrial DNA mutations in disease and aging. Ann N Y Acad Sci 908: 199-207.

Coulondre C, Miller JH (1977) Genetic studies of the lac repressor. III. Additional correlation of mutational sites with specific amino acid residues. J Mol Biol 117: 525-67.

Cozzarelli NR, Melechen NE (1967) Jovin TM, Kornberg A: Polynucleotide cellulose as a substrate for a polynucleotide ligase induced by phage T4. Biochem Biophys Res Commun 28: 578-86.

Current protocols in molecular biology (1987) John Wiley & Sons, New York.

Curry J, Karnaoukhova L, Guenette GC, Glickman BW (1999) Influence of sex, smoking and age on human hprt mutation frequencies and spectra. Genetics 152: 1065-77.

Davies MJ, Lovell DP, Anderson D (1992) Thioguanine-resistant mutant frequency in T-lymphocytes from a healthy human population. Mutat Res 265: 165-71.

de Boer JG, Ripley LS (1988) An in vitro assay for frameshift mutations: hotspots for deletions of 1 bp by Klenow-fragment polymerase share a consensus DNA sequence. Genetics 118: 181-91.

Denissenko MF, Chen JX, Tang MS, Pfeifer GP (1997) Cytosine methylation determines hot spots of DNA damage in the human P53 gene. Proc Natl Acad Sci U S A 94: 3893-8.

Denissenko MF, Koudriakova TB, Smith L, O'Connor TR, Riggs AD, Pfeifer GP (1998) The p53 codon 249 mutational hotspot in hepatocellular carcinoma is not related to selective formation or persistence of aflatoxin B1 adducts. Oncogene 17: 3007-14.

Denissenko MF, Pao A, Tang M, Pfeifer GP (1996) Preferential formation of benzo[a]pyrene adducts at lung cancer mutational hotspots in P53. Science 274: 430-2.

Diamandis EP (1997) Clinical applications of tumor suppressor genes and oncogenes in cancer. Clin Chim Acta 257: 157-80.

Doty P, Marmur J, Eigner J, Schildkraut (1960) Strand separation and specific recombination in deoxyribonucleic acids: physical chemical studies. Proc Nat Acad Sci USA 46: 461-476.

Dugaiczyk A, Boyer HW, Goodman HM (1975) Ligation of EcoRI endonuclease-generated DNA fragments into linear and circular structures. J Mol Biol 96: 171-84.

Dunham I, Shimizu N, Roe BA, Chissoe S, Hunt AR, Collins JE, Bruskiewich R, Beare DM, Clamp M, Smink LJ, Ainscough R, Almeida JP, Babbage A, Bagguley C, Bailey J, Barlow K, Bates KN, Beasley O, Bird CP, Blakey S, Bridgeman AM, Buck D, Burgess J, Burrill WD, O'Brien KP, et al. (1999) The DNA sequence of human chromosome 22. Nature 402: 489-95.

Dunning AM, Talmud P, Humphries SE (1988) Errors in the polymerase chain reaction. Nucleic Acids Res 16: 10393.

Eckert KA, Kunkel TA (1990) High fidelity DNA synthesis by the Thermus aquaticus DNA polymerase. Nucleic Acids Res 18: 3739-44.

Edwards A, Voss H, Rice P, Civitello A, Stegemann J, Schwager C, Zimmermann J, Erfle H, Caskey CT, Ansorge W (1990) Automated DNA sequencing of the human HPRT locus. Genomics 6: 593-608.

Efcavitch JW (1990) The electrophoresis of synthetic oligonucleotides. In Gel Electrophoresis of Nucleic Acids: A practical Approach, IRL Press, New York, pp. 125-149.

Ehrlich M, Norris KF, Wang RY, Kuo KC, Gehrke CW (1986) DNA cytosine methylation and heat-induced deamination. Biosci Rep 6: 387-93.

Ehrlich SD, Sgaramella V, Lederberg J (1977) T4 ligase joins flush-ended DNA duplexes generated by restriction endonucleases. pp. 261-8. In: Vogel HJ, ed. Nucleic acid-protein recognition. New York, Academic Press.

Engler MJ., Richardson CC (1982) DNA ligases. In The Enzymes, Vol. XV, Academic Press, New York, pp. 3-29.

Ennis PD, Zemmour J, Salter RD, Parham P (1990) Rapid cloning of HLA-A,B cDNA by using the polymerase chain reaction: frequency and nature of errors produced in amplification. Proc Natl Acad Sci U S A 87: 2833-7.

Escara JF, Hutton JR (1980) Thermal stability and renaturation of DNA in dimethyl sulfoxide solutions: acceleration of the renaturation rate. Biopolymers 19: 1315-27.

Felley-Bosco E, Pourzand C, Zijlstra J, Amstad P, Cerutti P (1991) A genotypic mutation system measuring mutations in restriction recognition sequences. Nucleic Acids Res 19: 2913-9.

Fernandez E, Bienvenu T, Desclaux Arramond F, Beldjord K, Kaplan JC, Beldjord C (1993) Use of chemical clamps in denaturing gradient gel electrophoresis: application in the detection of the most frequent Mediterranean beta-thalassemic mutations. PCR Methods Appl 3: 122-4.

Ferretti L, Sgaramella V (1981) Temperature dependence of the joining by T4 DNA ligase of termini produced by type II restriction endonucleases. Nucleic Acids Res 9: 85-93.

Finette BA, Poseno T, Albertini RJ (1996) V(D)J recombinase-mediated HPRT mutations in peripheral blood lymphocytes of normal children. Cancer Res 56: 1405-12.

Finlay CA, Hinds PW, Tan TH, Eliyahu D, Oren M, Levine AJ (1988) Activating mutations for transformation by p53 produce a gene product that forms an hsc70-p53 complex with an altered half-life. Mol Cell Biol 8: 531-9.

Fischer SG, Lerman LS (1983) DNA fragments differing by single base-pair substitutions are separated in denaturing gradient gels: correspondence with melting theory. Proc Natl Acad Sci U S A 80: 1579-83.

Fixman M, Freire JJ (1977) Theory of DNA melting curves. Biopolymers 16: 2693-704.

Flaman JM, Frebourg T, Moreau V, Charbonnier F, Martin C, Ishioka C, Friend SH, Iggo R (1994) A rapid PCR fidelity assay. Nucleic Acids Res 22: 3259-60.

Frederico LA, Kunkel TA, Shaw BR (1990) A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy. Biochemistry 29: 2532-7.

Frenz J, Hancock WS (1991) High performance capillary electrophoresis. Trends Biotechnol 9: 243-50.

Fuscoe JC, Vira LK, Collard DD, Moore MM (1997) Quantification of hprt gene deletions mediated by illegitimate V(D)J recombination in peripheral blood cells of humans. Environ Mol Mutagen 29: 28-35.

Fuscoe JC, Zimmerman LJ, Lippert MJ, Nicklas JA, O'Neill JP, Albertini RJ (1991) V(D)J recombinase-like activity mediates hprt gene deletion in human fetal T-lymphocytes. Cancer Res 51: 6001-5.

Gallou C, Joly D, Mejean A, Staroz F, Martin N, Tarlet G, Orfanelli MT, Bouvier R, Droz D, Chretien Y, Marechal JM, Richard S, Junien C, Beroud C (1999) Mutations of

the VHL gene in sporadic renal cell carcinoma: definition of a risk factor for VHL patients to develop an RCC. Hum Mutat 13: 464-75.

Gannon JV, Greaves R, Iggo R, Lane DP (1990) Activating mutations in p53 produce a common conformational effect. A monoclonal antibody specific for the mutant form. Embo J 9: 1595-602.

Gefter ML, Becker A, Hurwitz J (1967) The enzymatic repair of DNA. I. Formation of circular lambda-DNA. Proc Natl Acad Sci U S A 58: 240-7.

Gellert M (1966) Formation of covalent circles of lambda DNA by E. coli extracts. Proc Nat Acad Sci USA 57: 148-155.

Gennett IN, Thilly WG (1988) Mapping large spontaneous deletion endpoints in the human HPRT gene. Mutat Res 201: 149-60.

Gennett IN (1988) Mapping large spontaneous deletion endpoints in the human HPRT gene in vitro and in vivo. Ph.D. Thesis, MIT.

Giles RE, Blanc H, Cann HM, Wallace DC (1980) Maternal inheritance of human mitochondrial DNA. Proc Natl Acad Sci U S A 77: 6715-9.

Gille C, Gille A, Booms P, Robinson PN, Nurnberg P (1998) Bipolar clamping improves the sensitivity of mutation detection by temperature gradient gel electrophoresis. Electrophoresis 19: 1347-50.

Goncalves I, Duret L, Mouchiroud D (2000) Nature and structure of human genes that generate retropseudogenes. Genome Res 10: 672-8.

Gonzalgo ML, Jones PA (1997) Mutagenic and epigenetic effects of DNA methylation. Mutat Res 386: 107-18.

Gotoh O, Tagashira Y (1981) Stabilities of nearest-neighbor doublets in double-helical DNA determined by fitting calculated melting profiles to observed profiles. Biopolymers 20: 1033-1042.

Grasby JA, Connolly BA (1992) Stereochemical outcome of the hydrolysis reaction catalyzed by the EcoRV restriction endonuclease. Biochemistry 31: 7855-61.

Greagg MA, Fogg MJ, Panayotou G, Evans SJ, Connolly BA, Pearl LH (1999) A read-ahead function in archaeal DNA polymerases detects promutagenic template-strand uracil. Proc Natl Acad Sci U S A 96: 9045-50.

Green MH, O'Neill JP, Cole J (1995) Suggestions concerning the relationship between mutant frequency and mutation rate at the hprt locus in human peripheral T-lymphocytes. Mutat Res 334: 323-39.

Greenblatt MS, Bennett WP, Hollstein M, Harris CC (1994) Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis. Cancer Res 54: 4855-78.

Greene AE, Athreya BH, Lehr HB, Coriell LL (1970) The effect of prolonged storage of cell cultures in dimethyl sulfoxide and glycerol prior to freezing. Cryobiology 6: 552-5.

Greer S, Zamenhof S (1962) Studies of depurination of DNA by heat. J Mol Biol 4: 123-141.

Grist SA, McCarron M, Kutlaca A, Turner DR, Morley AA (1992) In vivo human somatic mutation: frequency and spectrum with age. Mutat Res 266: 189-96.

Guldberg P, Gronbak K, Aggerholm A, Platz A, thor Straten P, Ahrenkiel V, Hokland P, Zeuthen J (1998) Detection of mutations in GC-rich DNA by bisulphite denaturing gradient gel electrophoresis. Nucleic Acids Res 26: 1548-9.

Haber DA, Fearon ER (1998) The promise of cancer genetics. Lancet 351 Suppl 2: SII1-8.

Hainaut P, Hernandez T, Robinson A, Rodriguez-Tome P, Flores T, Hollstein M, Harris CC, 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: 205-13.

Hainaut P, Pfeifer GP (2001) Patterns of p53 G-->T transversions in lung cancers reflect the primary mutagenic signature of DNA-damage by tobacco smoke. Carcinogenesis 22: 367-74.

Hanekamp JS (1993) Development of techniques necessary for the measurement and interpretation of mutational spectra in humans. Ph.D. Thesis, MIT.

Hanekamp JS. André P, Coller HA, Li XC, Thilly W G, Khrapko K (1996) Constant denaturant capillary electrophoresis for detection and enrichment of sequence variants. In Laboratory Protocols for Mutation Detection, Oxford University Press, Oxford, pp. 38-41.

Harris CC, Hollstein M (1993) Clinical implications of the p53 tumor-suppressor gene [see comments]. N Engl J Med 329: 1318-27.

Henke W, Herdel K, Jung K, Schnorr D, Loening SA (1997) Betaine improves the PCR amplification of GC-rich DNA sequences. Nucleic Acids Res 25: 3957-8.

Hernandez-Boussard T, Montesano R, Hainaut P (1999a) Sources of bias in the detection and reporting of p53 mutations in human cancer: analysis of the IARC p53 mutation database. Genet Anal 14: 229-33.

Hernandez-Boussard T, Rodriguez-Tome P, Montesano R, Hainaut P (1999b) IARC p53 mutation database: a relational database to compile and analyze p53 mutations in human tumors and cell lines. International Agency for Research on Cancer. Hum Mutat 14: 1-8.

Hernandez-Boussard TM, 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: 385-91.

Higgins N P, Cozzarelli NR (1979) DNA-joining enzymes: a review. Methods Enzymol 68: 50-71.

Hjertén S (1985) High-performance electrophoresis elimination of electroendosmosis and solute adsorption. J Chromatogr 347: 191-198.

Hovig E, Smith-Sorensen B, Brogger A, Borresen AL (1991) Constant denaturant gel electrophoresis, a modification of denaturing gradient gel electrophoresis, in mutation detection [published erratum appears in Mutat Res 1991 May;263(1):61]. Mutat Res 262: 63-71.

Huang H, Keohavong P (1996) Fidelity and predominant mutations produced by deep vent wild-type and exonuclease-deficient DNA polymerases during in vitro DNA amplification. DNA Cell Biol 15: 589-94.

Hung MC, Wensink PC (1984) Different restriction enzyme-generated sticky DNA ends can be joined in vitro. Nucleic Acids Res 12: 1863-74.

Hussain SP, Aguilar F, Cerutti P (1994) Mutagenesis of codon 248 of the human p53 tumor suppressor gene by N-ethyl-N-nitrosourea. Oncogene 9: 13-8.

Hussain SP, Harris CC (1998) Molecular epidemiology of human cancer: contribution of mutation spectra studies of tumor suppressor genes. Cancer Res 58: 4023-37.

Hutton JR (1977) Renaturation kinetics and thermal stability of DNA in aqueous solutions of formamide and urea. Nucleic Acids Res 4: 3537-55.

Iggo R, Gatter K, Bartek J, Lane D, Harris AL (1990) Increased expression of mutant forms of p53 oncogene in primary lung cancer. Lancet 335: 675-9.

Itakura K, Rossi JJ, Wallace RB (1984) Synthesis and use of synthetic oligonucleotides. Annu Rev Biochem 53: 323-56.

Janatipour M, Trainor KJ, Kutlaca R, Bennett G, Hay J, Turner DR, Morley AA (1988) Mutations in human lymphocytes studied by an HLA selection system. Mutat Res 198: 221-6.

Jeanpierre C, Beroud C, Niaudet P, Junien C (1998) Software and database for the analysis of mutations in the human WT1 gene. Nucleic Acids Res 26: 271-4.

Jeltsch A, Alves J, Maass G, Pingoud A (1992) On the catalytic mechanism of EcoRI and EcoRV. A detailed proposal based on biochemical results, structural data and molecular modelling. FEBS Lett 304: 4-8.

Johnson BA, McClain SG, Doran ER, Tice G, Kirsch MA (1990) Rapid purification of synthetic oligonucleotides: a convenient alternative to high-performance liquid chromatography and polyacrylamide gel electrophoresis. Biotechniques 8: 424-9.

Jonason AS, Kunala S, Price GJ, Restifo RJ, Spinelli HM, Persing JA, Leffell DJ, Tarone RE, Brash DE (1996) Frequent clones of p53-mutated keratinocytes in normal human skin [see comments]. Proc Natl Acad Sci U S A 93: 14025-9.

Karran P, Lindahl T (1980) Hypoxanthine in deoxyribonucleic acid: generation by heat-induced hydrolysis of adenine residues and release in free form by a deoxyribonucleic acid glycosylase from calf thymus. Biochemistry 19: 6005-11.

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

Kellenberger G, Zichichi ML, Weigle JJ (1961) Exchange of DNA in the recombination of bacteriophage λ. Proc Nat Acad Sci USA 47: 869-878.

Keohavong P, Ling L, Dias C, Thilly WG (1993) Predominant mutations induced by the Thermococcus litoralis, vent DNA polymerase during DNA amplification in vitro. PCR Methods Appl 2: 288-92.

Keohavong P, Liu VF, Thilly WG (1991) Analysis of point mutations induced by ultraviolet light in human cells. Mutat Res 249: 147-59.

Keohavong P, Thilly WG (1989) Fidelity of DNA polymerases in DNA amplification. Proc Natl Acad Sci U S A 86: 9253-7.

Keohavong P, Thilly WG (1992a) Mutational spectrometry: a general approach for hot-spot point mutations in selectable genes. Proc Natl Acad Sci U S A 89: 4623-7.

Keohavong P, Thilly WG (1992b) Determination of point mutational spectra of benzo[a]pyrene-diol epoxide in human cells. Environ Health Perspect 98: 215-9.

Khrapko K, Andre P, Cha R, Hu G, Thilly WG (1994b) Mutational spectrometry: means and ends. Prog Nucleic Acid Res Mol Biol 49: 285-312.

Khrapko K, Bodyak N, Thilly WG, van Orsouw NJ, Zhang X, Coller HA, Perls TT, Upton M, Vijg J, Wei JY (1999) Cell-by-cell scanning of whole mitochondrial genomes in aged human heart reveals a significant fraction of myocytes with clonally expanded deletions. Nucleic Acids Res 27: 2434-41.

Khrapko K, Coller H, Andre P, Li XC, Foret F, Belenky A, Karger BL, Thilly WG (1997b) Mutational spectrometry without phenotypic selection: human mitochondrial DNA. Nucleic Acids Res 25: 685-93.

Khrapko K, Coller HA, Andre PC, Li XC, Hanekamp JS, Thilly WG (1997a) Mitochondrial mutational spectra in human cells and tissues. Proc Natl Acad Sci U S A 94: 13798-803.

Khrapko K, Hanekamp JS, Thilly WG, Belenkii A, Foret F, Karger BL (1994a) Constant denaturant capillary electrophoresis (CDCE): a high resolution approach to mutational analysis. Nucleic Acids Res 22: 364-9.

Khrapko K, Coller HA, Li-Sucholeiki XC, André PC, Thilly WG (2001) High resolution analysis of point mutations by constant denaturant capillary electrophoresis (CDCE). In Methods in Molecular Biology, Vol. 163: Capillary Electrophoresis of Nucleic Acids, Vol. 2, Humana Press, New Jersey, pp. 57-72.

Kim AS, Li-Sucholeiki XC, Thilly WG (2001) Applications of constant denaturant capillary electrophoresis and complementary procedures: measurement of point mutational spectra. In Methods in Molecular Biology, Vol. 163: Capillary Electrophoresis of Nucleic Acids, Vol. 2, Humana Press, New Jersey, pp. 175-189.

Kim AS, Unpublished results.

Kogelnik AM, Lott MT, Brown MD, Navathe SB, Wallace DC (1998) MITOMAP: a human mitochondrial genome database--1998 update. Nucleic Acids Res 26: 112-5.

Kohne DE, Levison SA, Byers MJ (1977) Room temperature method for increasing the rate of DNA reassociation by many thousandfold: the phenol emulsion reassociation technique. Biochemistry 16: 5329-41.

Krawczak M, Ball EV, Cooper DN (1998) Neighboring-nucleotide effects on the rates of germ-line single-base-pair substitution in human genes. Am J Hum Genet 63: 474-88.

Krawczak M, Ball EV, Fenton I, Stenson PD, Abeysinghe S, Thomas N, Cooper DN (2000) Human gene mutation database-a biomedical information and research resource. Hum Mutat 15: 45-51.

Krenitsky TA, Papaioannou R, Elion GB (1969) Human hypoxanthine phosphoribosyltransferase. I. Purification, properties, and specificity. J Biol Chem 244: 1263-70.

Kunkel TA (1984) Mutational specificity of depurination. Proc Natl Acad Sci U S A 81: 1494-8.

Kunkel TA (1985) The mutational specificity of DNA polymerase-beta during in vitro DNA synthesis. Production of frameshift, base substitution, and deletion mutations. J Biol Chem 260: 5787-96.

Kunkel TA (1990) Misalignment-mediated DNA synthesis errors. Biochemistry 29: 8003-11.

Kunkel TA (1992) DNA replication fidelity. J Biol Chem 267: 18251-4.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, et al. (2001) Initial sequencing and analysis of the human genome. Nature 409: 860-921.

Lasken RS, Schuster DM, Rashtchian A (1996) Archaebacterial DNA polymerases tightly bind uracil-containing DNA. J Biol Chem 271: 17692-6.

Laurent-Puig P, Beroud C, Soussi T (1998) APC gene: database of germline and somatic mutations in human tumors and cell lines. Nucleic Acids Res 26: 269-70.

Lehman IR (1974) DNA ligase: structure, mechanism, and function. Science 186: 790-7.

Leong PM, Thilly WG, Morgenthaler S (1985) Variance estimation in single-cell mutation assays: comparison to experimental observations in human lymphoblasts at 4 gene loci. Mutat Res 150: 403-10.

Lerman LS, Silverstein K (1987) Computational simulation of DNA melting and its application to denaturing gradient gel electrophoresis. Methods Enzymol 155: 482-501.

Levine AJ, Momand J, Finlay CA (1991) The p53 tumour suppressor gene. Nature 351: 453-6.

Levy JA, Virolainen M, Defendi V (1968) Human lymphoblastoid lines from lymph node and spleen. Cancer 22: 517-24.

Li XC, Thilly WG (1996) Use of wide-bore capillaries in constant denaturant capillary electrophoresis. Electrophoresis 17: 1884-9.

Li-Sucholeiki XC, Khrapko K, Andre PC, Marcelino LA, Karger BL, Thilly WG (1999) Applications of constant denaturant capillary electrophoresis/high- fidelity polymerase chain reaction to human genetic analysis. Electrophoresis 20: 1224-32.

Li-Sucholeiki XC (1999) A technology for detecting unselected mutational spectra in human genomic DNA. Ph.D. Thesis, MIT.

Li-Sucholeiki XC, Thilly WG (2000) A sensitive scanning technology for low frequency nuclear point mutations in human genomic DNA. Nucleic Acids Res 28: E44.

Li-Sucholeiki XC, Jing L, Marcelino L, Gruhl A, Sudo H, Gostjeva E, Vatland J, Pan V, Willey J, Zarbl H, Thilly WG, Unpublished results.

Lim EL, Tomita AV, Thilly WG, Polz MF (2001) Combination of competitive quantitative pcr and constant-denaturant capillary electrophoresis for high-resolution detection and enumeration of microbial cells. Appl Environ Microbiol 67: 3897-903.

Lindahl T (1979) DNA glycosylases, endonucleases for apurinic/apyrimidinic sites, and base excision-repair. Prog Nucleic Acid Res Mol Biol 22: 135-92.

Lindahl T (1982) DNA repair enzymes. Annu Rev Biochem 51: 61-87.

Lindahl T (1993) Instability and decay of the primary structure of DNA [see comments]. Nature 362: 709-15.

Lindahl T, Karlstrom O (1973) Heat-induced depyrimidination of deoxyribonucleic acid in neutral solution. Biochemistry 12: 5151-4.

Lindahl T, Nyberg B (1972) Rate of depurination of native deoxyribonucleic acid. Biochemistry 11: 3610-8.

Lindahl T, Nyberg B (1974) Heat-induced deamination of cytosine residues in deoxyribonucleic acid. Biochemistry 13: 3405-10.

Ling LL, Keohavong P, Dias C, Thilly WG (1991) Optimization of the polymerase chain reaction with regard to fidelity: modified T7, Taq, and vent DNA polymerases. PCR Methods Appl 1: 63-9.

Loeb KR, Loeb LA (2000) Significance of multiple mutations in cancer. Carcinogenesis 21: 379-85.

Loeb LA (1985) Apurinic sites as mutagenic intermediates. Cell 40: 483-4.

Loeb LA, Preston BD (1986) Mutagenesis by apurinic/apyrimidinic sites. Annu Rev Genet 20: 201-30.

Lundberg KS, Shoemaker DD, Adams MW, Short JM, Sorge JA, Mathur EJ (1991) High-fidelity amplification using a thermostable DNA polymerase isolated from Pyrococcus furiosus. Gene 108: 1-6.

Mace K, Aguilar F, Wang JS, Vautravers P, Gomez-Lechon M, Gonzalez FJ, Groopman J, Harris CC, Pfeifer AM (1997) Aflatoxin B1-induced DNA adduct formation and p53 mutations in CYP450-expressing human liver cell lines. Carcinogenesis 18: 1291-7.

Marcelino LA, Andre PC, Khrapko K, Coller HA, Griffith J, Thilly WG (1998) Chemically induced mutations in mitochondrial DNA of human cells: mutational spectrum of N-methyl-N'-nitro-N-nitrosoguanidine. Cancer Res 58: 2857-62.

Marcelino LA, Thilly WG (1999) Mitochondrial mutagenesis in human cells and tissues. Mutat Res 434: 177-203.

Marmur J, Doty P (1961) Thermal renaturation of deoxyribonucleic acids. J Mol Biol 3: 585-594.

Marmur J, Doty P (1962) Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature. J Mol Biol, 5: 109-118.

Marmur J, Lane D (1960) Strand separation and specific recombination II deoxyribonucleic acids: biological studies. Proc Nat Acad Sci USA 46: 453-461.

Marusyk R, Sergeant A (1980) A simple method for dialysis of small-volume samples. Anal Biochem 105: 403-4.

Mattila P, Korpela J, Tenkanen T, Pitkanen K (1991) Fidelity of DNA synthesis by the Thermococcus litoralis DNA polymerase--an extremely heat stable enzyme with proofreading activity. Nucleic Acids Res 19: 4967-73.

McConaughy BL, Laird CD, McCarthy BJ (1969) Nucleic acid reassociation in formamide. Biochemistry 8: 3289-95.

McDonell MW, Simon MN, Studier FW (1977) Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels. J Mol Biol 110: 119-46.

McDowell DG, Burns NA, Parkes HC (1998) Localised sequence regions possessing high melting temperatures prevent the amplification of a DNA mimic in competitive PCR. Nucleic Acids Res 26: 3340-7.

McGinniss MJ, Nicklas JA, Albertini RJ (1989) Molecular analyses of in vivo hprt mutations in human T-lymphocytes: IV. Studies in newborns. Environ Mol Mutagen 14: 229-37.

Mertz JE, Davis RW (1972) Cleavage of DNA by R 1 restriction endonuclease generates cohesive ends. Proc Natl Acad Sci U S A 69: 3370-4.

Mighell AJ, Smith NR, Robinson PA, Markham AF (2000) Vertebrate pseudogenes. FEBS Lett 468: 109-14.

Morley AA, Trainor KJ, Seshadri R, Ryall RG (1983) Measurement of in vivo mutations in human lymphocytes. Nature 302: 155-6.

Mottes M, Morandi C, Cremaschi S, Sgaramella V (1977) Restoration by T4 ligase of DNA sequences sensitive to "flush" cleaving restriction enzyme. Nucleic Acids Res 4: 2467-75.

Muniappan BP, Thilly WG (1999) Application of constant denaturant capillary electrophoresis (CDCE) to mutation detection in humans. Genet Anal 14: 221-7.

Muniappan BP, Unpublished results.

Myers RM, Fischer SG, Lerman LS, Maniatis T (1985a) Nearly all single base substitutions in DNA fragments joined to a GC-clamp can be detected by denaturing gradient gel electrophoresis. Nucleic Acids Res 13: 3131-45.

Myers RM, Fischer SG, Maniatis T, Lerman LS (1985b) Modification of the melting properties of duplex DNA by attachment of a GC-rich DNA sequence as determined by denaturing gradient gel electrophoresis. Nucleic Acids Res 13: 3111-29.

Nakazawa H, English D, Randell PL, Nakazawa K, Martel N, Armstrong BK, Yamasaki H (1994) UV and skin cancer: specific p53 gene mutation in normal skin as a biologically relevant exposure measurement. Proc Natl Acad Sci U S A 91: 360-4.

New England BioLabs, Catalog 2000-2001.

Olivera BM, Lehman IR (1967) Linkage of polynucleotides through phosphodiester bonds by an enzyme from Escherichia coli. Proc Natl Acad Sci U S A 57: 1426-33.

Oller AR, Thilly WG (1992) Mutational spectra in human B-cells. Spontaneous, oxygen and hydrogen peroxide-induced mutations at the hprt gene. J Mol Biol 228: 813-26.

Ouhtit A, Nakazawa H, Armstrong BK, Kricker A, Tan E, Yamasaki H, English DR (1998) UV-radiation-specific p53 mutation frequency in normal skin as a predictor of risk of basal cell carcinoma. J Natl Cancer Inst 90: 523-31.

Ouhtit A, Ueda M, Nakazawa H, Ichihashi M, Dumaz N, Sarasin A, Yamasaki H (1997) Quantitative detection of ultraviolet-specific p53 mutations in normal skin from Japanese patients. Cancer Epidemiol Biomarkers Prev 6: 433-8.

Palombo F, Bignami M, Dogliotti E (1992) Non-phenotypic selection of N-methyl-N-nitrosourea-induced mutations in human cells. Nucleic Acids Res 20: 1349-54.

Parry JM, Shamsher M, Skibinski DO (1990) Restriction site mutation analysis, a proposed methodology for the detection and study of DNA base changes following mutagen exposure. Mutagenesis 5: 209-12.

Parsons BL, Heflich RH (1997a) Evaluation of MutS as a tool for direct measurement of point mutations in genomic DNA. Mutat Res 374: 277-85.

Parsons BL, Heflich RH (1997b) Genotypic selection methods for the direct analysis of point mutations. Mutat Res 387: 97-121.

Parsons BL, Heflich RH (1998) Detection of basepair substitution mutation at a frequency of 1 x 10(-7) by combining two genotypic selection methods, MutEx enrichment and allele-specific competitive blocker PCR. Environ Mol Mutagen 32: 200-11.

Patel PI, Nussbaum RL, gramson PE, Ledbetter DH, Caskey CT, Chinault AC (1984) Organization of the HPRT gene and related sequences in the human genome. Somat Cell Mol Genet 10: 483-93.

Pfeifer GP, Tang M, Denissenko MF (2000) Mutation hotspots and DNA methylation. Curr Top Microbiol Immunol 249: 1-19.

Pingoud A, Jeltsch A (1997) Recognition and cleavage of DNA by type-II restriction endonucleases. Eur J Biochem 246: 1-22.

Pinkus H (1952) Examination of the epidermis by the strip method. J Invest Dermatol 19: 431-447.

Podlutsky A, Hou SM, Nyberg F, Pershagen G, Lambert B (1999) Influence of smoking and donor age on the spectrum of in vivo mutation at the HPRT-locus in T lymphocytes of healthy adults. Mutat Res 431: 325-39.

Podlutsky A, Osterholm AM, Hou SM, Hofmaier A, 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: 557-66.

Poland D (1974) Recursion relation generation of probability profiles for specific-sequence macromolecules with long-range correlations. Biopolymers 13: 1859-71.

Polisson C, Robinson D (1992) ApoI, a unique restriction endonuclease from Arthrobacter protophormiae which recognizes 5' RAATTY 3'. Nucleic Acids Res 20: 2888.

Ponten F, Berg C, Ahmadian A, Ren ZP, Nister M, Lundeberg J, Uhlen M, Ponten J (1997) Molecular pathology in basal cell cancer with p53 as a genetic marker. Oncogene 15: 1059-67.

Ponten F, Berne B, Ren ZP, Nister M, Ponten J (1995) Ultraviolet light induces expression of p53 and p21 in human skin: effect of sunscreen and constitutive p21 expression in skin appendages. J Invest Dermatol 105: 402-6.

Potten CS (1981) Cell replacement in epidermis (keratopoiesis) via discrete units of proliferation. Int Rev Cytol 69: 271-318.

Potten CS, Hendry JH (1973) Letter: Clonogenic cells and stem cells in epidermis. Int J Radiat Biol Relat Stud Phys Chem Med 24: 537-40.

Potten CS, Loeffler M (1990) Stem cells: attributes, cycles, spirals, pitfalls and uncertainties. Lessons for and from the crypt. Development 110: 1001-20.

Potten CS, Morris RJ (1988) Epithelial stem cells in vivo. J Cell Sci Suppl 10: 45-62

Pourzand C, Cerutti P (1993a) Genotypic mutation analysis by RFLP/PCR. Mutat Res 288: 113-21.

Pourzand C, Cerutti P (1993b) Mutagenesis of H-ras codons 11 and 12 in human fibroblasts by N-ethyl-N-nitrosourea. Carcinogenesis 14: 2193-6.

Raeymaekers L (2000) Basic principles of quantitative PCR. Mol Biotechnol 15: 115-22.

Reischl U, Kochanowski B (1995) Quantitative PCR. A survey of the present technology. Mol Biotechnol 3: 55-71.

Reischl U, Kochanowski B (1995) Quantitative PCR. A survey of the present technology. Mol biotechnol 3: 55-71.

Ren ZP, Ahmadian A, Ponten F, Nister M, Berg C, Lundeberg J, Uhlen M, Ponten 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: 1791-803.

Ren ZP, Ponten F, Nister M, Ponten J (1996) Two distinct p53 immunohistochemical patterns in human squamous-cell skin cancer, precursors and normal epidermis. Int J Cancer 69: 174-9.

Riggs AD, Jones PA (1983) 5-methylcytosine, gene regulation, and cancer. Adv Cancer Res 40: 1-30.

Roberts RJ, Macelis D (2001) REBASE-restriction enzymes and methylases. Nucleic Acids Res 29: 268-9.

Robertson GP, Huang HJ, Cavenee WK (1999) Identification and validation of tumor suppressor genes. Mol Cell Biol Res Commun 2: 1-10.

Robin ED, Wong R (1988) Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. J Cell Physiol 136: 507-13.

Robinson DR, Goodall K, Albertini RJ, O'Neill JP, Finette B, Sala-Trepat M, Moustacchi E, Tates AD, Beare DM, Green MH, et al. (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: 227-47.

Rodin SN, Rodin AS (2000) Human lung cancer and p53: the interplay between mutagenesis and selection. Proc Natl Acad Sci U S A 97: 12244-9.

Ruiz-Martinez MC, Berka J, Belenkii A, Foret F, Miller AW, Karger BL (1993) DNA sequencing by capillary electrophoresis with replaceable linear polyacrylamide and laser-induced fluorescence detection. Anal Chem 65: 2851-8.

Sagher D, Strauss B (1985) Abasic sites from cytosine as termination signals for DNA synthesis. Nucleic Acids Res 13: 4285-98.

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239: 487-91.

Sarkar G, Kapelner S, Sommer SS (1990) Formamide can dramatically improve the specificity of PCR. Nucleic Acids Res 18: 7465.

Schaaper RM, Kunkel TA, Loeb LA (1983) Infidelity of DNA synthesis associated with bypass of apurinic sites. Proc Natl Acad Sci U S A 80: 487-91.

Scharf SJ, Horn GT, Erlich HA (1986) Direct cloning and sequence analysis of enzymatically amplified genomic sequences. Science 233: 1076-8.

Schmeckpeper BJ, Smith KD (1972) Use of formamide in nucleic acid reassociation. Biochemistry 11: 1319-26.

Schmutte C, Yang AS, Beart RW, Jones PA (1995) Base excision repair of U:G mismatches at a mutational hotspot in the p53 gene is more efficient than base excision repair of T:G mismatches in extracts of human colon tumors. Cancer Res 55: 3742-6.

Schnell S, Mendoza C (1997) Enzymological considerations for a theoretical description of the quantitative competitive polymerase chain reaction (QC-PCR). J Theor Biol 184: 433-40.

Sellner LN, Turbett GR (1996) The presence of a pseudogene may affect the use of HPRT as an endogenous mRNA control in RT-PCR. Mol Cell Probes 10: 481-3.

Sgaramella V, Ehrlich SD (1978) Use of the T4 polynucleotide ligase in the joining of flush-ended DNA segments generated by restriction endonucleases. Eur J Biochem 86: 531-7.

Shadel GS, Clayton DA (1997) Mitochondrial DNA maintenance in vertebrates. Annu Rev Biochem 66: 409-35.

Shearman CW, Loeb LA (1979) Effects of depurination on the fidelity of DNA synthesis. J Mol Biol 128: 197-218.

Sheffield VC, Cox DR, Lerman LS, Myers RM (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 U S A 86: 232-6.

Shen JC, Rideout WM, Jones PA (1994) The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. Nucleic Acids Res 22: 972-6.

Skopek TR, Liber HL, Penman BW, Thilly WG (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: 411-6.

Smith J, Modrich P (1996) Mutation detection with MutH, MutL, and MutS mismatch repair proteins. Proc Natl Acad Sci U S A 93: 4374-9.

Soussi T, Caron de Fromentel C, May P (1990) Structural aspects of the p53 protein in relation to gene evolution. Oncogene 5: 945-52.

Soussi T, Dehouche K, Beroud C (2000) p53 website and analysis of p53 gene mutations in human cancer: forging a link between epidemiology and carcinogenesis. Hum Mutat 15: 105-13.

Stout JT, Caskey CT (1985) HPRT: gene structure, expression, and mutation. Annu Rev Genet 19: 127-48.

Strauss BS (2000) Role in tumorigenesis of silent mutations in the TP53 gene. Mutat Res 457: 93-104.

Sun Y, Hegamyer G, Colburn NH (1993) PCR-direct sequencing of a GC-rich region by inclusion of 10% DMSO: application to mouse c-jun. Biotechniques 15: 372-4.

Tabata H, Nagano T, Ray AJ, Flanagan N, Birch-MacHin MA, Rees JL (1999) Low frequency of genetic change in p53 immunopositive clones in human epidermis. J Invest Dermatol 113: 972-6.

Thilly W G (1993) Measurement of mutation spectra as a molecular dosimeter. In Use of Biomarkers in Assessing Health and Environmental Impacts of Chemical Pollutants, Plenum Press, New York, pp. 47-51.

Thrower KJ, Peacocke AR (1966) The kinetics of renaturation of DNA. Biochim Biophys Acta 119: 652-4.

Thrower KJ, Peacocke AR (1968) Kinetic and spectrophotometric studies on the renaturation of deoxyribonucleic acid. Biochem J 109: 543-57.

Tindall KR, Kunkel TA (1988) Fidelity of DNA synthesis by the Thermus aquaticus DNA polymerase. Biochemistry 27: 6008-13.

Tomita-Mitchell A, Muniappan BP, Herrero-Jimenez P, Zarbl H, Thilly WG (1998) Single nucleotide polymorphism spectra in newborns and centenarians: identification of genes coding for rise of mortal disease. Gene 223: 381-91.

Tomita-Mitchell A (1999) The spontaneous and MNNG-induced mutational spectra in the human HPRT gene in vitro. Ph.D. Thesis. MIT.

Tomita-Mitchell A, Kat AG, Marcelino LA, Li-Sucholeiki XC, Goodluck-Griffith J, Thilly WG (2000) Mismatch repair deficient human cells: spontaneous and MNNG-induced mutational spectra in the HPRT gene. Mutat Res 450: 125-138.

Tomita-Mitchell A, Unpublished results.

Tommasi S, Denissenko MF, Pfeifer GP (1997) Sunlight induces pyrimidine dimers preferentially at 5-methylcytosine bases. Cancer Res 57: 4727-30.

Tornaletti S, Pfeifer GP (1994) Slow repair of pyrimidine dimers at p53 mutation hotspots in skin cancer. Science 263: 1436-8.

Tornaletti S, Pfeifer GP (1995) Complete and tissue-independent methylation of CpG sites in the p53 gene: implications for mutations in human cancers. Oncogene 10: 1493-9.

Tornaletti S, Rozek D, Pfeifer GP (1993) The distribution of UV photoproducts along the human p53 gene and its relation to mutations in skin cancer. Oncogene 8: 2051-7.

Trainor KJ, Wigmore DJ, Chrysostomou A, Dempsey JL, Seshadri R, Morley AA (1984) Mutation frequency in human lymphocytes increases with age. Mech Ageing Dev 27: 83-6.

Ullman JS, McCarthy BJ (1973a) The relationship between mismatched base pairs and the thermal stability of DNA duplexes. II. Effects of deamination of cytosine. Biochim Biophys Acta 294: 416-24.

Ullman JS, McCarthy BJ (1973b) The relationship between mismatched base pairs and the thermal stability of DNA duplexes. I. Effects of depurination and chain scission. Biochim Biophys Acta 294: 405-15.

Vanin EF (1985) Processed pseudogenes: characteristics and evolution. Annu Rev Genet 19: 253-72.

Varadaraj K, Skinner DM (1994) Denaturants or cosolvents improve the specificity of PCR amplification of a G + C-rich DNA using genetically engineered DNA polymerases. Gene 140: 1-5.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, et al. (2001) The sequence of the human genome. Science 291: 1304-51.

Vijayalaxmi, Evans HJ (1984) Measurement of spontaneous and X-irradiation-induced 6-thioguanine-resistant human blood lymphocytes using a T-cell cloning technique. Mutat Res 125: 87-94.

Vineis P, Malats N, Porta M, Real FX (1999) Human cancer, carcinogenic exposures and mutation spectra. Mutat Res 436: 185-94.

Vital statistics of the United States, Vol. II-Mortality Part A. U. S. Department of health and human services, Public health service, Centers for disease control and prevention, National center for health statistics, Maryland.

Vrieling H, Thijssen JC, Rossi AM, van Dam FJ, Natarajan AT, Tates AD, van Zeeland AA (1992) Enhanced hprt mutant frequency but no significant difference in mutation spectrum between a smoking and a non-smoking human population. Carcinogenesis 13: 1625-31.

Wallace DC (1999) Mitochondrial diseases in man and mouse. Science 283: 1482-8.

Wang RY, Kuo KC, Gehrke CW, Huang LH, Ehrlich M (1982) Heat- and alkali-induced deamination of 5-methylcytosine and cytosine residues in DNA. Biochim Biophys Acta 697: 371-7.

Warren WJ, Vella G (1993) Analysis of synthetic oligodeoxyribonucleotides by capillary gel electrophoresis and anion-exchange HPLC. Biotechniques 14: 598-606.

Warren WJ, Vella G (1995) Principles and methods for the analysis and purification of synthetic deoxyribonucleotides by high-performance liquid chromatography. Mol Biotechnol 4: 179-99.

Weinberg RA (1991) Tumor suppressor genes. Science 254: 1138-46.

Weinstein GD, McCullough JL, Ross P (1984) Cell proliferation in normal epidermis. J Invest Dermatol 82: 623-8.

Weiss B, Richardson CC (1967) Enzymatic breakage and joining of deoxyribonucleic acid, I. Repair of single-strand breaks in DNA by an enzyme system from Escherichia coli infected with T4 bacteriophage. Proc Natl Acad Sci U S A 57: 1021-8.

Wilde CD (1986) Pseudogenes. CRC Crit Rev Biochem 19: 323-52.

Wilson JM, Baugher BW, Kelley WN (1984) Hypoxanthine-guanine phosphoribosyltransferase in human lymphoblastoid cells: confirmation of four structural variants and demonstration of a new variant (HPRT Ann Arbor). Adv Exp Med Biol 165: 33-8.

Wilson VL, Wei Q, Wade KR, Chisa M, Bailey D, Kanstrup CM, Yin X, Jackson CM, Thompson B, Lee WR (1999) Needle-in-a-haystack detection and identification of base substitution mutations in human tissues. Mutat Res 406: 79-100.

Wilson VL, Yin X, Thompson B, Wade KR, Watkins JP, Wei Q, Lee WR (2000) Oncogenic base substitution mutations in circulating leukocytes of normal individuals. Cancer Res 60: 1830-4.

Winship PR (1989) An improved method for directly sequencing PCR amplified material using dimethyl sulphoxide. Nucleic Acids Res 17: 1266.

Wu Y, Stulp RP, Elfferich P, Osinga J, Buys CH, Hofstra RM (1999) Improved mutation detection in GC-rich DNA fragments by combined DGGE and CDGE. Nucleic Acids Res 27: e9.

Wynford-Thomas D (1991) Oncogenes and anti-oncogenes; the molecular basis of tumour behaviour. J Pathol 165: 187-201.

Yarbro JW (1992) Oncogenes and cancer suppressor genes. Semin Oncol Nurs 8: 30-9.

Zhang W, Unpublished results.

Ziegler A, Jonason AS, Leffell DJ, Simon JA, Sharma HW, Kimmelman J, Remington L, Jacks T, Brash DE (1994) Sunburn and p53 in the onset of skin cancer [see comments]. Nature 372: 773-6.

Ziegler A, Leffell DJ, Kunala S, Sharma HW, Gailani M, Simon JA, Halperin AJ, Baden HP, Shapiro PE, Bale AE, et al. (1993) Mutation hotspots due to sunlight in the p53 gene of nonmelanoma skin cancers. Proc Natl Acad Sci U S A 90: 4216-20.

Zijlstra J, Felley-Bosco E, Amstad P, Cerutti P (1990) A mammalian mutation system avoiding phenotypic selection: the RFLP/PCR approach. Prog Clin Biol Res 347: 187-200.

Zimmerman SB, Little JW, Oshinsky CK, Gellert M (1967) Enzymatic joining of DNA strands: a novel reaction of diphosphopyridine nucleotide. Proc Natl Acad Sci U S A 57: 1841-8.

Zimmermann K, Mannhalter JW (1996) Technical aspects of quantitative competitive PCR. Biotechniques 21: 268-72, 274-9.