5-1987

# Mutagenesis and DNA Repair in Mammalian Cells

Peter M. Glazer
*Yale University.*

# MUTAGENESIS AND DNA REPAIR IN MAMMALIAN CELLS

A Dissertation

Presented to the Faculty of the Graduate School

of

Yale University

in Candidacy for the Degree of

Doctor of Philosophy

by

Peter M. Glazer

May 1987

**Abstract**

MUTAGENESIS AND DNA REPAIR IN MAMMALIAN CELLS

Peter M. Glazer

Yale Universi ty

1987

Two aspects of DNA damage and repair in mammalian cells were investigated. Using a lambda phage shuttle vector, a system was developed to study mutations arising in the DNA of mammalian cells. This system was used to determine the spectrum of mutations induced in cellular DNA by ultraviolet light. Also, the repair of base pair mismatches in DNA was studied by the development of a method to detect a DNA mismatch repair activity in extracts made from cultured human cells.

In order to study mutations arising in mammalian cells, stable mouse L cell lines were established with multiple copies of a lambda phage vector which contains the *supF* gene of *E. coli* as a target for mutagenesis. Rescue of viable phage from high molecular weight mouse cell DNA using lambda *in vitro* packaging extracts was efficient (5 viable phage per μg of cell DNA per lambda genome copy) and yielded a negligible background of phage with mutations in the *supF* gene (zero out of 54,605). From mouse cells exposed to 12 $J/m^2$ of 254 nm ultraviolet (UV) light, 78,510 phage were rescued of which eight were found to have mutant *supF* genes. DNA sequence analysis of the mutants suggests that the primary site of UV mutagenesis in mammalian cells is at pyrimidine-cytosine (Py-C) sequences, and that the most frequent mutation at this site is a C to T transition.

Extension of this system to use the thymidine kinase gene of herpes simplex virus as a target gene was attempted, and the application of this system to the study of mutagenesis in transgenic mice was begun.

A system to study DNA mismatch repair *in vitro* in HeLa cell extracts was also developed. Pre-formed heteroduplex plasmid DNA containing two single base pair mismatches within the *supF* gene of *E. coli* was used as a substrate in a mismatch repair assay. Repair of one or both of the mismatches to the wild type sequence was measured by transformation of a *lac* $^{-}$(Amber) *E. coli* strain in which the presence of an active suppressor gene could be scored. The *E. coli* strain used was constructed to carry mutations in genes associated with mismatch repair and recombination *(mutH, mutU,* and *recA)* so that processing of the heteroduplex DNA by the bacteria was minimal. Extract reactions were carried out by the incubation of the heteroduplex plasmid DNA in the HeLa cell extracts to which ATP, creatine phosphate, creatine kinase, deoxynucleotides, and a magnesium containing buffer had been added. Under these conditions, about 1% of the mismatches were repaired. In the absence of added energy sources or of added deoxynucleotides, the activity in the extracts was significantly reduced. The addition of either aphidicolin or dideoxynucleotides reduced the mismatch repair activity, but only aphidicolin was effective in blocking DNA polymerization in the extracts. It is concluded that mismatch repair in these extracts is an energy-requiring process that is dependent on adequate deoxynucleotide concentrations. The results also indicate that the process is associated with some type of DNA polymerization, but the different effects of aphidicolin and dideoxynucleotides suggest that the mismatch repair activity cannot be accounted for simply by random nick translation activity in the extracts. Further studies are underway to characterize this activity, in particular to determine the effect of DNA methylation on strand selection in the repair of mismatches.

**Acknowledgements**

I am grateful to my supervisor, Dr. William C. Summers, for all the help and encouragement he has given me throughout the course of this work. I am thankful for the help of Dr. Wilma P. Summers, who assisted in the experimental manipulation of tissue culture cells. I thank Dr. Saumyen Sarkar for his willingness to help with many chores. I appreciate the assistance with DNA sequence analysis provided by Jeff Settlemen. Dr. George Chisholm assisted in the preparation of human cell extracts. Drs. Brooks Low and Dan DiMaio provided much guidance. Dr. Efim Golub, along with Dr. Low, provided advice in the genetic manipulation of *E. coli*. I am very grateful to have had the support of Drs. Rich D'Aquila, Marcia Lewis, and Jill Lacy. Susan Baserga was a source of support and assistance throughout this work.

I also wish to thank Drs. Joseph Warshaw and Jim Jamieson for supporting my interest in the Medical Scientist Training Program, and I am grateful to the NIH, which supports this program at Yale.

**Table of Contents**

**List of Figures**

List of Tables

## Chapter 1. Mutagenesis I

### Introduction

The work of this thesis represents an application of molecular genetics to the study of mutagenesis and DNA repair in mammalian cells. Investigation of DNA damage and repair in mammalian cells has been hampered by the difficulty of genetic manipulation of mammalian cells in culture. To circumvent this problem, methods were developed to study DNA metabolism in mammalian cells that would use the tools of molecular biology and exploit the power of bacterial genetic analysis.

Two different aspects of DNA metabolism were studied. The first part of the thesis, described in this chapter, involves an investigation of ultraviolet (UV) light-induced mutagenesis in mouse cells. A target gene, *supF*, the amber suppressor tyrosine tRNA gene of *E. coli* (Brown, et al., 1979), was introduced into mouse cells in culture, and the cells were exposed to UV light. The target gene was rescued from within the mouse cell genome using a shuttle vector system based on the bacteriophage lambda, and the rescued target genes were screened in bacteria for the presence of mutations. The mutations were characterized by DNA sequence analysis, and a representative sample of UV-induced mutations in mammalian cell DNA was determined for the first time.

In further work, the groundwork was laid for extension of this shuttle vector system to transgenic mice. Attempts were also made to improve the system for studying mutagenesis in tissue culture, as discussed in Chapter 2. These included the use of the herpes simplex virus (HSV) thymidine kinase (TK) gene as a target gene whose function can be screened in cell culture.

The second aspect of DNA metabolism in mammalian cells that was studied was the phenomenon of DNA mismatch repair (Chapter 3). Specific DNA molecules were

constructed to carry defined base pair mismatches, and these molecules were used as substrates to study mismatch repair both in mammalian cells in culture and in cell free extracts of mammalian cells. Heteroduplex repair activity was detected for the first time in such extracts, and various factors affecting this activity were studied. An investigation of mismatch repair processes in *E. coli* was also undertaken, resulting in the confirmation and the extension of a recent report of a novel aspect of heteroduplex processing in *E. coli*.

**Introduction to mutagenesis.** The appearance of mutations in the genome of an organism is a process central to evolution, carcinogenesis, and possibly even aging. Since the demonstration of mutagenesis by X-rays by Muller (1927), it has long been a goal of molecular biologists to elucidate the mechanisms involved in mutagenesis. One aspect of mutagenesis that has recently become the focus of considerable study is the determination of the specific kinds of mutations caused by various environmental agents. Much data exist concerning the chemical nature of lesions induced in DNA by various agents (Drake and Baltz, 1976; Singer and Kusmierek, 1982). However, the processing of these lesions, resulting in either repair, mutation, or persistent lesion, is poorly understood, especially in mammalian cells. Mutagens may also cause mutations in DNA not only where a chemical lesion has been created, but also in areas of the genome that have been undamaged. Such mutations are called non-targeted mutations, in contrast to the targeted mutations occuring at specific lesions. In *E. coli*, non-targeted mutations generally occur through the induction by the mutagen of the *recA*- dependent SOS pathway (reviewed by Walker, 1984), but in mammalian cells little is known about such mutations except that they do occur (Dasgupta and Summers, 1978).

Analysis of the mutagenic spectrum of an agent may provide clues to help elucidate mechanisms of DNA repair and of targeted and non-targeted mutagenesis. It can provide insight into which lesions in DNA are mutagenic, and it continues on a molecular level investigation into the well established link between mutagenesis and carcinogenesis (Ames et

al., 1973). This type of analysis is especially useful in the study of mammalian cells, since the genetic approach to the study of DNA repair that has been used so successfully with *E. coli* cannot be easily applied to mammalian cells.

**Mutagen specificity in *E. coli*.** Jeffrey Miller, in reviewing mutagen specificity in *E. coli* (Miller, 1983), describes three basic mutagenic pathways. The mutagen may be incorporated into the DNA in place of the normal bases. This can cause mispairing with other bases, leading to mutations. It is thought that 5-bromouracil may be incorporated opposite adenine residues and subsequently mispair with guanine, a model consistent with the report that 80 of 83 mutations induced in the cI gene of lambda by 5-bromouracil were A:T to G:C transitions (Skopek and Hutchinson, 1982). Similarly, 2-aminopurine, an adenine analog, can pair with thymine but also mispairs with cytosine. Coulondre and Miller (1977) found the expected preponderance of transitions in the *lacI* gene of *E. coli* exposed to this compound.

Another group of mutagens acts by chemically altering bases, resulting in mispairing. Alkylating agents, such as ethyl methanesulfonate (EMS) and N-methyl-N'-nitro-N-nitrosoguanidine (MNNG), act in this way. Of the lesions generated by these agents, $O^6$-alkyl guanine may be the most mutagenic, leading to mispairing with thymine (Drake and Baltz, 1976). Consistent with this, Coulondre and Miller (1977) demonstrated that these agents tend to produce G:C to A:T transitions.

A third mechanism of mutagenesis occurs where the mutagen alters the bases in DNA so that normal base pairing is prevented, generating a block to replication. In *E. coli*, induction of the *recA*- dependent SOS pathway is thought to allow lesion bypass and the production of mutations (Witkin, 1976). In some cases, mutagens may generate lesions which are metabolized by the cell, leading to the removal of the entire base and generating apurinic and apyrimidinic (AP) sites. Such sites are themselves mutagenic, especially after SOS induction in *E. coli* (Schaaper and Loeb, 1981). Kunkel (1984) and Schaaper et al.

(1983) have shown that apurinic sites cause mostly G:C to T:A transversions in *E. coli*. This result is consistent with the model of preferential adenine insertion opposite apurinic sites suggested by the *in vitro* work of Strauss et al. (1981). The mutagens benzo[a]pyrene (BP) (Eisenstadt et al, 1982), 2-acetylaminofluorene (AAAF) (Foster et al., 1983), and aflatoxin B1 (AFB1) (Foster et al., 1983), when metabolically activated, all cause primarily G:C to T:A transversions. Guanine is the most common target base for these compounds. Although they each form different adducts with guanine and although these compounds show specificty for different guanine residues in a given gene, it has been suggested that they all induce mutations via apurinic sites that have come from the removal of the damaged guanine residues (Foster et al., 1983). Apurinic sites have also been invoked to explain the spectrum of mutagenesis seen by Miller and Low (1984) in *E. coli* in which the SOS system is induced by growing *tif-1* mutants at 42°C. They found mostly G:C to T:A transversions in the *lacI* gene, although these mutations occured at different sites from those seen with BP, AAAF, or AFB1. They suggested that pre-existing apurinic sites which occur natually in cells by spontaneous depurination and by the action of N-glycosylases could constitute cryptic sites which do not induce the SOS system themselves but which are SOS mutable.

Other SOS dependent mutagens which have been well studied include 4-nitroquinoline-1-oxide (NQO), which causes mostly G:C to A:T transitions, but also some G:C to T:A transversions at different sites. The antitumor drug, cis-diamminedichloroplatinum (II), (cis-platinum), forms adducts at guanine residues, including crosslinks between nearby guanines from the $O^6$ to the $N^7$ positions of the residues. Brouwer et al. (1981) found that most base substitutions caused by cis-platinum were G:C to T:A transversions, with a small number of G:C to A:T transitions, both occuring mainly at GAG or GCG sequences. They suggested that intrastrand cross links between guanines separated by a single base may be the pre-mutagenic lesion. The $N^7$ adduct may lead to depurination and subsequent transversion mutations, while the $O^6$ adduct

may cause mispairing and therefore transition mutations.

Ultraviolet light (UV) as a mutagen has been the subject of considerable study in *E. coli*, and in *E. coli* it has been associated with a variety of mutations, arising from several photoproducts. Initially, it was thought that the pyrimidine-pyrimidine cyclobutane dimers (Beukers and Berends, 1960) were the most important pre-mutagenic lesion induced by UV. Brash and Haseltine (1982), however, suggest that for *E. coli* the main mutagenic lesion may be the pyrimidine-pyrimidone-(6-4) photoproduct, which occurs predominantly at TC and CC sequences (but not at CT sequences), rather than thymidine dimers, even though thymidine dimers occur more frequently in response to UV light. They analyzed the distribution of UV-induced photoproducts in the *lacI* gene in *E. coli* in comparison with the distribution of UV-induced nonsense mutations in the same gene as reported by Coulondre and Miller (1977). They found an excellent correlation between the frequency of nonsense mutations and the frequency of UV-induced damage at TC and CC sites, but no correlation with UV-induced damage at TT sites. The correlation at TC and CC sites was better for the (6-4) product than for the cyclobutane dimer. At most pyrimidine-pyrimidine pairs, cyclobutane dimers generally occur at a several-fold higher frequency, but at sites where mutations tend to occur, the frequency of (6-4) photoproducts was found to be significantly elevated, usually higher than the level of the cyclobutane dimers themselves. Brash and Haseltine noted that the (6-4) photoproduct is not formed at CT sequences. At these sites, only cyclobutane dimers are found. Arguing that cyclobutane dimers are less mutagenic, they therefore predicted that mutations at CT sites such as A:T to G:C transitions should be less frequent than mutations at CC or TC sites, such as G:C to A:T transitions. In the nonsense mutation assay system of Coulondre and Miller (1977), however, A:T to G:C transitions are not detectable, and so the hypothesis of Brash and Haseltine was not fully tested. Miller (1985) went on to analyze UV induced mutations in the *lacI* gene carried on an F' episome of *E. coli* by direct DNA sequencing rather than by a nonsense mutation

suppression assay. He detected 60% G:C to A:T transitions and only 8% A:T to G:C transitions, supporting the hypothesis of Brash and Haseltine.

Several groups have contributed further data concerning UV specificity in *E. coli*. Leclerc and Istock (1982) observed mostly transitions at pyrimidine-pyrimidine sites in the *lac* gene promoter in phage M13mp2. Wood et al. (1984) examined a series of mutations induced in the cI gene of lambda phage. Prior to infection of *E. coli*, the phage were either exposed to 254 nm UV light, which induces the usual variety of photoproducts, or were treated with acetophenone, followed by exposure to 313 nm light, which induces mostly thymidine dimers and no detectable (6-4) photoproducts. They found that with 254 nm light, 60 of the 62 mutations occured at pyrimidine-pyrimidine sites, and that 63% of these were transitions. In contrast, the acetophenone plus 313 nm light treatment generated only one transition among 22 sequenced mutations. They concluded that thymidine dimers do not induce transitions efficiently, and so they inferred that the main pre-mutagenic lesion induced by UV light in *E. coli* is normally not thymidine dimers but rather the pyrimidine-pyrimidone-(6-4) product. Wood and Hutchinson (1984) also looked at non-targeted mutagenesis of unirradiated lambda phage in *E. coli* hosts irradiated with UV light. They found that 10 of 13 mutations sequenced were frameshift mutations, in contrast to the preponderance of transitions they observed with targeted UV mutagenesis. This suggested that non-targeted mutagenesis plays only a small role in the spectrum of UV-induced targeted mutations.

**Mutagen specificity in mammalian cells.** In contrast to the large amount of data available for *E. coli* , little is known about mutagen specificity in mammalian cells. This is due in part to the difficulty of genetic manipulation of mammalian cells. It was the purpose of this thesis work to circumvent this difficulty by developing a system to analyze a target gene present within the DNA of mammalian cells.

In early attempts to study mutagenesis in mammalian cells, the usual approach was to

examine the effect of mutagens on endogenous mammalian genes. Only genes whose function can be selected for or against in cell culture are candidates for such analysis. These studies are limited to Southern blot analysis of the mutant target genes (Southern, 1975), since cloning and sequencing these extremely large genes is a difficult task.

In one study, Fuscoe et al. (1983) looked at mutants in the 33 kb hypoxanthine-guanine phosphoribosyl transferase *(hprt)* gene of V79 Chinese hamster cells. Selecting for HPRT-deficient clones, they isolated 10 spontaneous and 9 UV-induced mutants. Of these 19 mutants, 17 showed no change in the mobility of any of the several restriction enzyme fragments detected on Southern blot analysis. One spontaneous mutant and one induced mutant had evidence of major deletions involving the *hprt* locus. Graf and Chasin (1982) investigated γ-ray induced mutations at the dihydrofolate reductase *(dhfr)* locus in chinese hamster ovary (CHO) cells. They looked at 9 mutants, of which 2 showed major rearrangements involving the *dhfr* gene. Meuth and Arrand (1982) studied EMS-induced mutations in the adenosine phosphsribosyl transferase gene *(aprt)* of CHO cells, and they again found few major rearrangements among the mutated *aprt* genes. Most of the mutations observed may have been point mutations, but small rearrangements could not be ruled out. Such studies, furthermore, do not allow for the determination of the exact molecular nature of the mutations.

Other studies that are basically similar in concept have used exogenous rather than endogenous genes as target genes. In these studies, gene transfer methods were used to introduce selectable genes into mammalian cells, and these genes were used to study mutagenesis by the same kind of analysis used for the endogenous *hprt* and *aprt* genes. The advantage of these studies was that small genes, without introns and with simple restriction enzyme cleavage patterns, could be used. The disadvantage is that such introduced genes tend to be unstable. Spontaneous mutation rates were high, usually resulting from frequent deletions of the exogenous gene (Perucho et al., 1980). In some

cases, apparent mutants arose without deletions or other alterations of the target genes; rather, in these clones the target gene was inactivated by an epigenetic phenomenon such as hypermethylation (Christy and Scangos, 1982 and 1984; Ostrander et al., 1982). Such apparent mutant cell lines gave rise to revertant subclones at extremely high frequencies, and the appearance of revertants was associated with a pattern of demethylation of the target gene (usually the HSV TK gene in these studies). This effect remains unexplained, as different clones derived from transformation with the same gene showed considerable variability in regard to this phenomenon (Hardies et al., 1983).

Two groups have succeeded in establishing hamster cell lines with just a single, relatively stable copy of the bacterial guanine phosphoribosyl transferase *(gpt)* gene driven by the simian virus 40 (SV40) early promoter. The cell line produced by Thacker et al. (1983) showed a spontaneous mutation frequency of $3x10^{-4}$. Increases in mutant frequency were induced by x-rays and other mutagens, but such mutation frequencies were about 20-fold higher than those observed with the native *hprt* gene in the studies cited above. Southern blot analysis of the mutations revealed that almost all were due to the deletion of the integrated *gpt* gene and of neighboring sequences. A similar system using the *gpt* gene in CHO cells was developed by Tindall and co-workers (1984). They found that most spontaneous and most X-ray induced mutations involved major deletions, but they noted that most mutations generated by EMS, ICR 191, and UV caused no detectable alterations in their restriction pattern.

Another approach to the study of mutagenesis in mammalian cells has been the use of mammalian viruses as probes for mutagenesis. This approach was taken by Dasgupta and Summers (1978), who used HSV to demonstrate in mammalian cells a phenomenon similar to Weigle reactivation in bacteria (Weigle,1953). They found that the survival of UV-irradiated HSV on UV-irradiated Vero cells was increased over that on unirradiated cells. They also found that mutations in the TK gene of the virus were increased in frequency when

the virus was grown on UV-irradiated host cells as opposed to unirradiated cells. They concluded that an "error-prone" inducible UV-reactivation phenomenon exists in mammalian cells. A similar reactivation of UV-irradiated SV40 in cells pre-treated with various carcinogens was observed by Sarasin and Hanawalt (1978).

**Shuttle vectors.** The above studies provide important information, but they are limited in that they do not define precisely the nature of the mutations occuring in the cells. To address this limitation, several groups have begun to develop systems to study mutagenesis which employ shuttle vectors that can replicate both in mammalian cells and in bacteria. In this way, the power of prokaryotic genetic analysis can be exploited in analyzing target genes that have been mutated in mammalian cells. Once carried on a replicative vector in bacteria, the mutant target genes are accessible to further examination by, for example, DNA sequence analysis.

The various shuttle vectors used initially were similar in design. Most were recombinant plasmids containing, in addition to a target gene, part of the SV40 genome, including its origin of replication, and part of the plasmid pBR322, including its origin of replication and the ampicillin resistance gene. The target genes that have been used are genes that can easily be screened in suitable bacteria, including *lacI* (Calos et al., 1983), galactokinase (Razzaque et al., 1983; Razzaque et al., 1984), *supF* (Sarkar et al., 1984; Seidman et al., 1985), and *gpt* (Ashman abnd Davidson, 1984). Experimentally, the shuttle vectors are introduced into mammalian cells under various circumstances and are allowed to replicate. The episomal DNA is then harvested and used to transform *E. coli* to ampicillin resistance. The bacterial colonies can be screened for the activity encoded by the target gene, and the plasmid DNA in these colonies can be analyzed.

Initial work by all these groups, except Sarkar et al., was plagued by a high mutation frequency (1-2%) in the absence of any mutagenic treatment. Sarkar et al. found a background mutation frequency of only 0.2%, which may be explained by the construction

of their vector, p3AC, in which the *supF* target gene is placed between the ampicillin resistance gene and the SV40 origin. Since most of the spontaneous mutations observed involve rearrangements, the fact that the *supF* gene in p3AC is placed between two genes essential for plasmid survival in these assays means that many of the spontaneous mutations will generate plasmids that are not viable in the assay and so will not be detected. Sarkar and co-workers used their plasmid shuttle vector to extend their earlier work involving UV reactivation of HSV (Dasgupta and Summers, 1978). After pre-treatment of monkey COS-7 cells (Gluzman, 1981) with EMS, they introduced their shuttle vector into the the cells and allowed it to replicate. With plasmid replication in EMS-treated cells, they found a 10-fold increase in *supF* mutation frequency relative to plasmid replication in untreated control cells. They also found that most of the mutant plasmids derived from replication in untreated cells were due to rearrangements, while most of those in the case of EMS pre-treated cells were small changes, undetectable by restriction enzyme analysis. They suggest that this is further evidence for a damage-inducible mutagenic pathway, or "SOS" pathway, in mammalian cells.

Other workers have begun to refine their systems to lower the background level of mutagenesis in order to detect mutations induced by the agents they seek to study. Calos and co-workers at first detected very high background mutation frequencies (up to 2%) in monkey, mouse and human cell lines using SV40- and polyoma-based vectors (Calos et al., 1983; Lebkowski et al., 1984). They suggested that the transfection process itself might be mutagenic. However, they found that the human cell line 293 (Graham et al., 1973) replicates SV40-based vectors with a spontaneous mutation frequency as low as $2x10^{-5}$ (Lebkowski et al., 1985). Using their *lacI* -containing SV40-based vector shuttle vector, they have gone on to study mutagenesis by UV and EMS in the 293 cells. Recently, they have begun experimenting with Epstein-Barr virus (EBV)-derived vectors (M. Calos, personal communication). In their protocol they first transfect cells with the shuttle vector,

and then they expose the cells to the mutagen. The results of their work will be discussed in conjunction with the results of this thesis.

Seidman and co-workers have also developed an SV40-based shuttle vector that has been useful in the study of mutagenesis in mammalian cells (Seidman et al., 1985). Using the *supF* gene as a target gene, they have studied UV-induced mutagenesis in repair proficient monkey and human cells and in repair deficient, xeroderma pigmentosum (XP)-derived human cell lines (Hauser et al., 1986; Bredberg et al., 1986; Protic-Sabljic et al., 1986). In contrast to the Calos group, they treated the shuttle vector DNA with the mutagen *in vitro* before transfection into the cells. Their results will also be discussed below.

Other groups have also been working on systems to study mutagenesis. Yang et al. (1987) have used the system of Sarkar et al. (1984) to study mutagenesis by benzo[a]pyrene-7,8-diol-9,10-epoxide in monkey cells. Ashman and Davidson (1984) tried an SV40 shuttle vector but were faced with a high background problem, while Drinkwater and Klinedinst (1986) have worked with an EBV-derived vector. Bourre and Sarasin (1982) did some early but limited work with a reversion assay for a temperature sensitive SV40 virus.

The work with these vectors has generated considerable data, but the the relevance of these systems to the mutagenesis of actual mammalian cell DNA has been questioned because these vectors are extra-chromosomal and replicate independently from the cell DNA. It is not known if the DNA repair processes which act on plasmids which exist outside the chromosomes are the same as those processes which act on the chromosomal DNA, itself. Different repair enzymes may be involved, or the same enzymes may operate with different efficiencies on extrachromosomal versus chromosomal sequences. Indeed, it has been observed that some characteristics of DNA recombination among extrachromosomal sequences are different from those among integrated sequences in mammalian cells (R. M.

Liskay, personal communication). For this reason, it is important to study the mutagenesis of genes contained within the DNA of the mammalian cell. Therefore, at the same time that the work with extra-chromosomal shuttle vectors described above was beginning, the work of this thesis was initiated toward the goal of developing a system of studying the mutagenesis of a gene contained within mammalian cell DNA.

Two approaches were considered. One was the use of a vector based on a retrovirus, since part of the replicative cycle of retroviruses involves integration of retroviral cDNA into the genome of the host cell. After integration and mutagenesis, however, the retrovirus vector containing the target gene must be retrieved from the cell DNA and transferred to bacteria for analysis. The most straightforward way of doing this is to include in the vector an SV40 origin of replication. When cells containing such a hybrid vector are fused with cells permissive for SV40 replication, such as COS cells, the SV40 DNA and sequences surrounding it are replicated rapidly, generating many episomal copies of these sequences (Conrad et al., 1982). The DNA can be recovered by the Hirt procedure (Hirt, 1967), as with the SV40 shuttle vectors.

This approach was not chosen because the vector must pass through a stage of replication outside the chromosome as essentially an SV40 vector. In this process, it might be subject to a high spontaneous mutation frequency, and it might be exposed to non-chromosomal repair processes. Ashman and co-workers (1986), however, have pursued this approach. They introduced the *gpt* gene into the retroviral vector of Cepko et al. (1984), and with it they infected an HPRT-deficient mouse cell line. They obtained a line with a single copy of the vector, and vectors rescued from this cell line by COS cell fusion showed a low spontaneous mutation frequency ($2\text{-}4 \times 10^{-5}$). With EMS treatment of the cells, they were able to induce a 10-fold higher frequency of mutations, and analysis of these mutations is underway.

**Lambda phage as a shuttle vector.** In the work of this thesis, a shuttle vector

system based on the bacteriophage lambda was developed in order to study the mutagenesis of a gene contained within mammalian cell DNA. In this system, the target gene, *supF*, plus a gene whose function can be selected in tissue culture, the neomycin resistance gene *(neo)*, are carried within a lambda phage vector, λsupF-neo. This vector DNA is introduced into mammalian cell DNA by standard gene transfer methods, and cell lines which have incorporated copies of the vector DNA into their genome are selected in the presence of the neomycin analog, G418. The phage vector DNA should therefore behave with respect to replication and mutagenesis as does chromosomal DNA. The cells are then subjected to treatment with a mutagen, and high molecular weight DNA is prepared from these cells. This DNA is used as a substrate in lambda *in vitro* packaging reactions (Hohn, 1979), in which the lambda packaging enzymes are able to cut out the lambda vector DNA from within the mammalian cell DNA and to package it into viable phage particles. These phage can then be grown on a lawn of a suitable *E. coli* host such as strain SY204, which carries an amber nonsense mutation in the ß-galactosidase gene so that the function of the *supF* gene can be assayed (Sarkar et al., 1984). Forward mutations to the non-suppressor phenotype are scored, and the identified mutant genes are analyzed further by DNA sequencing.

Crucial to this system is the ability of the lambda packaging extracts to cut out and package the lambda DNA contained within the cell DNA. The feasability of this process was demonstrated by Lau and Kan (1983 and 1984). They used phage DNA containing a functional TK gene to transform TK-deficient mouse cells to aminopterin resistance in the presence of thymidine. They were able to rescue phage DNA carrying the intact TK gene using the lambda packaging extracts, and they suggested that this technique might be useful for isolating genes that confer a selectable phenotype on mammalian cells in culture. They also tried cosmid constructs, but these were prone to rearrangement at a high frequency.

## Materials and methods

**Cells.** The construction of *E. coli* SY204 (*lacZ125* (Am) *trp-49, hsdR2::Tn10)* was performed by Sarkar et al.(1984). *E. coli* SY301 was constructed by introduction into SY204 of a mutation in ultraviolet mutability at the *umuC* locus by P1 transduction with *Tn5* kanamycin resistance from *E. coli* GW2100, which carries the *umuC122::Tn5* mutation (Elledge and Walker, 1983). *E. coli* N1066 is a triple tandem lambda lysogen obtained from M. Gottesman. Mouse LTK$^-$ cells (Kit et al., 1963) and human 143TK$^-$ cells were grown in modified Eagle's medium (MEM) supplemented with 10% fetal calf serum.

**Phage.** λsupF-neo was constructed from the phage vector λgt2 (Panasenko et al., 1977). The *supF* suppressor gene was isolated as a 200 base pair *EcoRI* fragment from plasmid p3AC (Sarkar et al., 1984). This gene is the amber suppressor tyrosine tRNA gene synthesized by Brown et al. (1970). The wild type gene, designated *supF*$^+$ (Su$^-$), is not an amber suppressor but is simply a tyrosine tRNA gene. The suppressor tRNA gene, which is the gene contained in the vector, is a mutant version of the *supF*$^+$ gene, and it may be designated as *supF*$^-$ (Su$^+$). Mutations in the suppressor *supF* gene which reduce or eliminate its function as a nonsense suppressor but which do not constitute reversion to the original wild type sequence (i.e. these are mutations which affect the tRNA function apart from reversion of the anti-codon sequence to that for tyrosine) may be designated *supF*$^-$ (Su$^-$). For purposes of clarification, these designations are sometimes used in this thesis. In general, however, the use of *supF* in this thesis indicates the amber suppressor tyrosine tRNA gene [*supF*$^-$ (Su$^+$)] unless otherwise indicated. Following attachment of *SalI* linkers, the *supF* gene [*supF*$^-$ (Su$^+$)] was ligated into the unique *SalI* site of the plasmid pdBPV-MMTneo (Law et al., 1983) from which the 8000 base-pair *BamH1* fragment containing bovine papilloma virus sequences had been previously eliminated. The resulting plasmid, pPG2, was linearized at the *EcoR1* site and ligated into the *EcoRI* cloning site of

λgt2 (Fig. 1.1).

**DNA transfection.** LTK$^-$ cells 143TK$^-$ cells were transfected with λsupF-neo DNA in the absence of carrier by the calcium phosphate method (Graham and van der Eb, 1973). Prior to transfection, the phage DNA was self-ligated at a concentration of 300 μg/ml in order to generate concatamers. Transformed cells were selected at an initial density of 4000 cells/cm$^2$ in the presence of 400 μg/ml of the neomycin analogue, G-418.

**Phage rescue.** DNA was prepared from *E. coli* N1066 by standard methods (Maniatis et al., 1982). High molecular weight DNA was prepared from transformed mouse cells and transformed human cells by lysing the monolayer of cells with 1 ml per 25 cm$^2$ of a solution of 10 mM Tris pH 8, 0.1% sodium dodecyl sulfate (SDS), and 1mM EDTA, after washing the monolayer with phosphate buffered saline (PBS). The resulting viscous solution was subjected to gentle phenol extraction, and the DNA was precipitated out of the aqueous phase by the addition of 2 volumes of 95% ethanol. The precipitated DNA was spooled around a glass rod, dried, and redissolved in water to yield a concentration of 0.2 mg/ml. This preparation of DNA is contaminated with RNA, but treatment of the solution with ribonuclease is not necessary. The DNA was used as a substrate for λ *in vitro* packaging extracts, which were prepared exactly as described by Hohn (1979). Only extracts efficient enough to package at least 5x10$^8$ plaque forming units (pfu) per μg of purified lambda DNA were used. For cell DNA, up to 1 μg of DNA in up to 5 μl of water was added per 50 μl packaging reaction. Packaged phage were plated on Luria broth plates in top agar containing 200 μg/ml of 5-bromo-4-chloro-3-indolyl ß-D-galactopyranoside (X-gal) and 75 μg/ml of isopropyl ß-D-thiogalactopyranoside (IPTG). Either SY204 or SY301 was used as a host. Since both strains contain *lacZ125* , an amber nonsense mutation in the ß-galactosidase gene, phage carrying a suppressor gene [*supF*$^-$ (Su$^+$)] produce blue plaques in this assay, while colorless plaques arise from phage which have lost suppressor activity.

**DNA blots.** DNA dot blots were performed using a Schleicher and Schuell microsample filtration manifold as described by the manufacturer. Southern blots and filter hybridizations were performed by standard methods (Southern, 1975).

**DNA sequence analysis.** Newly mutant *supF* genes [*supF*$^-$ ($Su^-$)] were subcloned from λsupF-neo into the *SalI* site of phage M13mp10 (Messing, 1983). Sequence analysis by the chain termination method (Sanger et al., 1977) was carried out using $^{35}$S-dATP (Amersham) and buffer gradient gels (Biggin et al., 1983).

## Results

**Transfection of lambda phage shuttle vector.** Fig. 1.1 schematically depicts the lambda shuttle vector used in these studies. This vector carries a marker selectable in mammalian cells, the neomycin resistance gene driven by the mouse metallothionein-I promoter. It also carries the *supF* gene [*supF*$^-$ ($Su^+$)] of *E. coli* as the target gene for mutagenesis experiments. This phage can lysogenize *E. coli* and confers on such lysogens resistance to ampicillin. Hence phage which lack suppressor activity can be screened quickly to determine if they represent background from the *in vitro* packaging extract, in which case they cannot form ampicillin resistant lysogens, or if they are *bona fide* mutants rescued from mammalian cell DNA. This construct carries no origin of replication that can function in mammalian cells, and so it is unlikely to be maintained outside of the chromosomes.

Prior to transfection, the phage DNA was ligated to form concatamers to protect the lambda cohesive ends, which are necessary for efficient packaging of the DNA into phage particles (Hohn, 1979; Hohn, 1983; Lindenmaier et al., 1982). Transfection of the human 143TK$^-$ cells yielded only one G418-resistant colony containing less than 5 copies of the lambda vector DNA per cell equivalent of DNA. No phage were rescued from the DNA of these cells, and no further work was done with them. In contrast, transfection of LTK$^-$ cells with λsupF-neo DNA and selection in the presence of 400 μg/ml of G418 yielded 15 colonies of resistant cells designated LN1 through LN15. Fig. 1.2 shows a DNA dot blot done to determine the number of copies of the λsupF-neo genome per cell. As standards for quantitation, DNA from the parental cell line, LTK$^-$, was mixed with appropriate amounts of purified λsupF-neo DNA to approximate 0, 1, 5, 10 and 50 phage genome copies per cell. For each standard or cell line, aliquots containing 1 and 5 μg of DNA were applied to the nitrocellulose filter, which was exposed to the $^{32}$P-labelled internal 6800 base pair *EcoRI* fragment of λsupF-neo DNA under hybridization conditions. As can be seen, there is a

wide range in copy number (1 to approximately 100).

Southern blot analysis of *EcoRI* digested DNA from the 15 cell lines and from LTK$^-$ cells, along with digested DNA from the standards described above, is shown in figures 1.3 and 1.4. Again, the hybridization probe consisted of the 6800 base pair *EcoRI* fragment of λsupF-neo. This analysis shows that the internal 6800 base pair fragment of λsupF-neo DNA containing the *supF* gene is essentially unrearranged in most of the cell lines regardless of copy number.

**Phage rescue.** As a preliminary test of the feasability of λ rescue from within foreign DNA, high molecular weight DNA was prepared from *E. coli* N1066, a triple tandem lambda lysogen, and this DNA was used as a substrate for the lambda *in vitro* packaging reaction. From 1 μg of bacterial DNA, 2 x $10^5$ phage were rescued. This represents an efficiency of phage rescue of 6.7 x $10^4$ pfu per μg bacterial cell DNA per copy of lambda. DNA from N1066 was diluted 1000-fold with mammalian cell DNA to approximate the situation in which 3 copies of lambda phage DNA are present per mammalian cell genome. With this DNA mixture as a substrate, the *in vitro* packaging reactions yielded 300 to 600 pfu per μg, an efficiency of 100 to 200 pfu per μg of total DNA per copy of lambda.

Rescue from mouse cell DNA, however, was not as efficient. High molecular weight DNA from each cell line was used as a substrate in lambda *in vitro* packaging reactions. In Fig. 1.5, the number of pfu rescued from 1 μg of cell DNA is plotted on a logarithmic scale against the number of copies of λsupF-neo per cell. As expected, there is a direct relationship between these variables. On average, the yield is about 1 to 5 pfu per μg of cell DNA per copy of λ.

Hence, the process of lambda rescue from mouse cell DNA seems to be less efficient than rescue from *E. coli* DNA, even when the *E. coli* DNA is in the presence of an excess of mammalian cell DNA. One explanation may be that these lambda packaging extracts,

which were made by the method of Hohn (1979), retain to some extent the *E. coli* K restriction system. Thus, there is some restriction of the unmodified mouse cell DNA, but no such restriction of the appropriately modified *E. coli* N1066 DNA. Rosenberg et al. (1985) have constructed an *E. coli* C strain from which lambda packaging extract can be made. This strain lacks a restriction system, and is reported to give higher yields of lambda from unmodified DNA (Rosenberg, 1985). Future work with this system will use packaging extracts made from the strain of Rosenberg and co-workers.

**Analysis of rescued phage.** The cell line yielding the greatest number of rescued phage was LN12. From this cell line, which carries about 100 copies of the phage genome per cell, 500 phage could be rescued per μg of cell DNA. Phage rescued from this cell line were analyzed for *supF* activity by growth on a *lac* $^-$(amber) strain of *E. coli*, either SY204 or SY301, in the presence of the chromogenic ß-galactosidase indicator, X-gal. 54,605 phage were screened, and all were found to produce the blue plaques indicative of a functional suppressor gene [*supF*$^-$ (Su$^+$)] in the assay used. No mutant phage were rescued from cells in the absence of exposure to mutagens. Therefore, the background level of mutants in this system may be less than one in 54,605.

**Mutagenesis by ultraviolet light.** LN12 cells at a density of 2 to 4 x$10^5$/cm$^2$ were exposed to 12 J/m$^2$ of 254 nm UV light. This dose reduced the survival of the cells with respect to their ability to form viable colonies to about 10% of that of untreated cells. Forty-eight hours after exposure to UV, DNA was harvested from the cells. Screening was performed using two *lac* (Am) *E. coli* hosts which differed only in that one (SY301) was a *umuC*- deficient derivative of the other (SY204). *umuC*- deficient strains show decreased UV-mutability with respect to both their own genome and that of UV-irradiated phage for which they are host (Kato and Shinoura, 1977). If the phage rescued from irradiated LN12 cells carry photochemical lesions in their DNA that have not been processed into mutations within the mouse cells (possibly because of radiation damage to the cells), then screening of

phage on SY301 may yield a lower frequency of mutants. If, however, the mutations in the rescued phage were fully processed in the mouse cells, then the mutation frequency should be the same as measured in the two *E. coli* strains. Table 1.1 lists the number of phage screened and the number of mutants detected using the two strains. As can be seen, the mutation frequency after irradiation is not affected by the strain used for screening, implying that the mutations arose from the processing within the mouse cells of UV-induced lesions. If the data from the two bacterial strains are considered together, the total mutation frequency of $1.02 \times 10^{-4}$ appears higher than that derived from untreated LN12 cells ($\leq 0.21 \times 10^{-4}$), and so it is reasonable to assume that the mutants collected are the result of exposure of the cells to UV light.

It should be noted that the cells used to study mutagenesis in the absence of UV-irradiation generally were passaged fewer times in culture than were the cells used to study UV-induced mutagenesis. However, several samples of unirradiated cells were collected at varying passage times, and neither the early passage cells nor the late passage cells yielded phage with *supF* gene mutations. It is felt therefore that the minor differences in passage time between UV-irradiated cells and unirradiated cells do not account for the differences in mutation frequency between phage rescued from the two groups of cells.

The *supF* genes from the eight mutant phage were subcloned into M13mp10 for sequence analysis. Table 1.2 presents the results of this analysis. Seven of the eight mutations consisted of single base changes. One involved two base changes separated by a single nucleotide. No rearrangements, insertions or deletions were detected. Of the nine base changes, eight occured at the 3' cytosine of a Py-C sequence. Seven of these eight were C to T transitions, while one was a C to G transversion. The remaining mutation consisted of a T to C transition at a C-T-C triplet, which may be interpreted as a mutation at the 5' pyrimidine of a Py-C pair, although it may represent a mutation at the 3' side of a C<>T cyclobutane dimer. All the mutations occured within runs of at least four and up to ten

pyrimidines. These conclusions are summarized in Fig. 1.6.

The position of each of the mutations can be seen better in Fig. 1.7, in which the mutated nucleotides are identified on the cloverleaf *supF* tRNA structure. The mutations seem to cluster in two regions, and there is some overlap among them. One of the two changes in mutant number 3 constitutes the single change in mutant number 6, and mutants 1, 2 and 8 involve identical mutations. Mutants 1 and 2 came from the same mutagenesis experiment, but this experiment and the ones which yielded mutants 3, 6 and 8 were all different. Thus all except possibly numbers 1 and 2 were independent mutants.

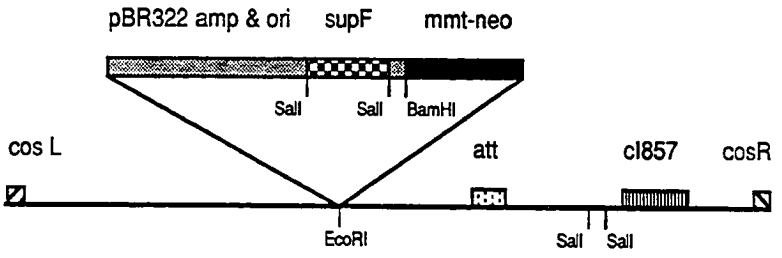


Fig. 1.1. Diagram of the λsupF-neo vector genome. This phage is a derivative of λgt2 (Panasenko et al., 1977). At the unique *EcoRI* site of the phage, a pBR322-derived plasmid, pPG2, was inserted. This plasmid contains the *E. coli* ampicillin resistance gene (amp), the amber suppressor tRNA gene, *supF* [*supF*$^{-}$ ($Su^{+}$)], in the unique *SalI* site of the plasmid, and the mouse metallothionein-I promoter attached to the *E. coli* coding sequence for neomycin resistance (mmt-neo). For orientation purposes, the right and left lambda cohesive ends (cosL and cosR) and the lambda *att* site are indicated, along with the locus of the *cI857* gene from λgt2.

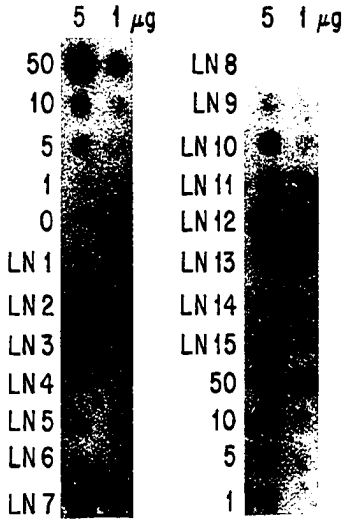


Fig. 1.2. DNA dot blot analysis of the number of copies of the λsupF-neo vector genome contained in transfected mouse LTK$^-$ cells. DNA was prepared from 15 G-418-resistant cell lines (LN1 through LN15) made by transfection of LTK$^-$ cells with λsupF-neo vector DNA. As standards for quantitation, DNA from untransfected LTK$^-$ cells was mixed with purified λsupF-neo DNA to approximate 0, 1, 5, 10 and 50 phage genome copies per cell. For each sample as indicated, aliquots of 1 and 5 μg of total DNA were applied to a nitrocellulose filter and were hybridized to the $^{32}$P-labelled 6800 base pair internal *EcoRI* fragment of λsupF-neo DNA. The autoradiograph of the hybridized filter is shown.

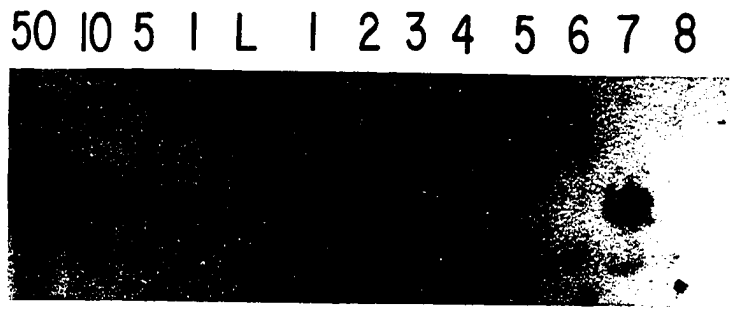


Fig. 1.3. Southern blot analysis of λsupF-neo vector DNA contained in transfected mouse LTK$^-$ cells. DNA was prepared from eight G-418-resistant cell lines (LN1 through LN8) made by transfection of LTK$^-$ cells with λsupF-neo vector DNA. As standards for quantitation, DNA from untransfected LTK$^-$ cells was mixed with purified λsupF-neo DNA to approximate 0 (lane "L"), 1, 5, 10 and 50 phage genome copies per cell. For each sample as indicated, 10 μg of DNA was digested with *EcoRI*, subjected to agarose gel electrophoresis, transferred to a nitrocellulose filter, and hybridized to the $^{32}$P-labelled 6800 base pair internal *EcoRI* fragment of λsupF-neo DNA. The autoradiograph of the hybridized filter is shown.

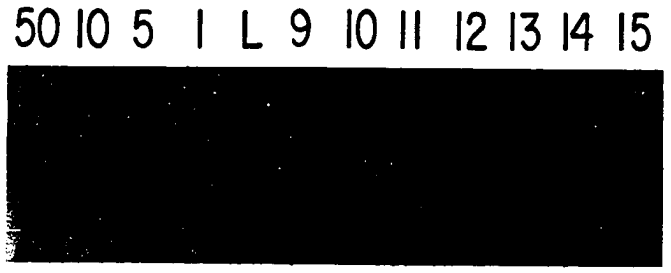


Fig. 1.4. Southern blot analysis of λsupF-neo vector DNA contained in transfected mouse LTK$^-$ cells. DNA was prepared from seven G-418-resistant cell lines (LN9 through LN15) made by transfection of LTK$^-$ cells with λsupF-neo vector DNA. As standards for quantitation, DNA from untransfected LTK$^-$ cells was mixed with purified λsupF-neo DNA to approximate 0 (lane "L"), 1, 5, 10 and 50 phage genome copies per cell. For each sample as indicated, 10 μg of DNA was digested with *EcoRI*, subjected to agarose gel electrophoresis, transferred to a nitrocellulose filter, and hybridized to the $^{32}$P-labelled 6800 base pair internal *EcoRI* fragment of λsupF-neo DNA. The autoradiograph of the hybridized filter is shown.

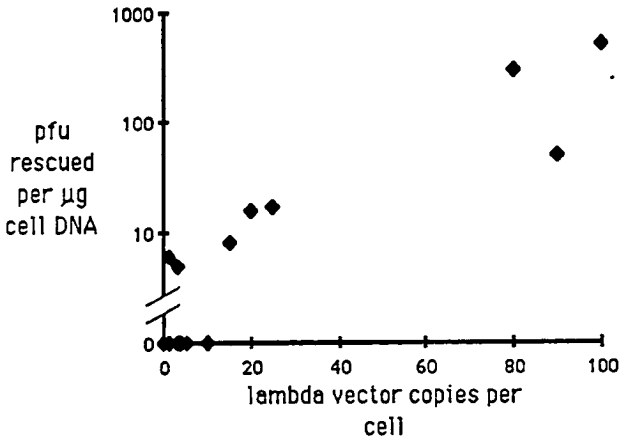


Fig. 1.5. Efficiency of phage rescue. High molecular weight DNA from LTK$^-$ cells and from cell lines LN1 through LN15 was used as a substrate in lambda in vitro packaging reactions. The resulting number of plaque forming units (pfu) rescued per µg of cell DNA for each cell line is plotted on a logarithmic scale against the number of copies of the λsupF-neo genome estimated to be present per cell on the basis of the DNA dot blot shown in Fig. 1.2. Note that the vertical axis is broken and that the horizontal axis denotes the value zero. Points displayed on the horizontal axis represent the 7 cell lines (LN1, LN5, LN6, LN9, LN10, LN13 and LN14) plus the LTK$^-$ cells from which no phage have been rescued.

Table 1.1. Frequency of mutant phage rescued from mouse LN12 cells.

| | Control cells | | | UV-irradiated cells | | |
|---|---|---|---|---|---|---|
| *E. coli* host | Phage, number | Mutants, number | Mutation frequency | Phage, number | Mutants, number | Mutation frequency |
| SY204 *(wt)* | 41,105 | 0 | $<0.24 \times 10^{-4}$ | 47,510 | 5 | $1.05 \times 10^{-4}$ |
| SY301 *(umuC)* | 13,500 | 0 | $<0.74 \times 10^{-4}$ | 31,000 | 3 | $0.97 \times 10^{-4}$ |
| Totals | 54,605 | 0 | $<0.21 \times 10^{-4}$ | 78,510 | 8 | $1.02 \times 10^{-4}$ |

Phage were rescued by in vitro packaging of DNA made from either unirradiated mouse LN12 cells or from LN12 cells treated with 12 joules/$m^2$ of UV radiation. Phage were screened for mutations in the *supF* gene by an assay for suppression using two *lac* (amber) *E. coli* hosts. either *E. coli* SY204 or its umuC-deficient derivative, *E. coli* SY301.

Table 1.2. Sequences of *supF* gene mutations in rescued phage.

| Mutant number | *supF* sequence | mutant sequence |
|---|---|---|
| 1,2 & 8 | ACTT**C**GAA<br>TGAA**G**CTT | ACTT**T**GAA<br>TGAA**A**CTT |
| 3 | AAAG**G**G**A**G<br>TTTC**C**C**T**C | AAAG**A**G**G**G<br>TTTC**T**C**C**C |
| 4 | ACTTC**G**AA<br>TGAAG**C**TT | ACTTC**A**AA<br>TGAAG**T**TT |
| 5 | AAT**C**CTTC<br>TTC**G**GAAT | AAT**T**CTTC<br>TTA**A**GAAG |
| 6 | AAAG**G**GAA<br>TTTC**C**CTC | AAAG**A**GAG<br>TTTC**T**CTC |
| 7 | AGACT**C**TA<br>TCTGA**G**AT | AGACT**G**TA<br>TCTGA**C**AT |

The *supF* genes of mutant phage rescued from irradiated mouse cells were isolated and sequenced. For each *supF* mutant [*supF*$^-$ ($Su^-$)] as indicated, an eight base pair region of DNA sequence, including the actual mutation, is listed opposite to the corresponding sequence of the suppressor *supF* gene [*supF*$^-$ ($Su^+$)].

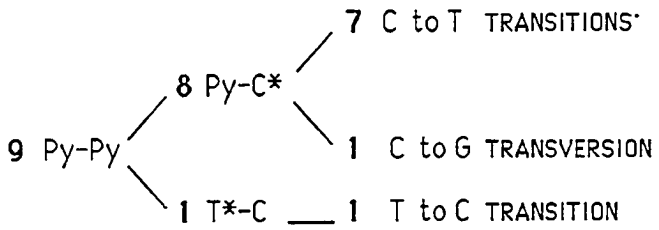


Fig. 1.6. Summary of UV-induced mutations in LN12 cells. The *supF* genes of mutant phage rescued from LN12 cells were isolated and sequenced. A total of 9 single base-pair changes were observed in the 8 mutant genes sequenced. The diagram indicates the number of times a mutation was found at each particular kind of site. An asterisk (*) marks the nucleotide involved in the mutation. Nucleotide pairs are written in the 5' to 3' direction. (Py, pyrimidine).

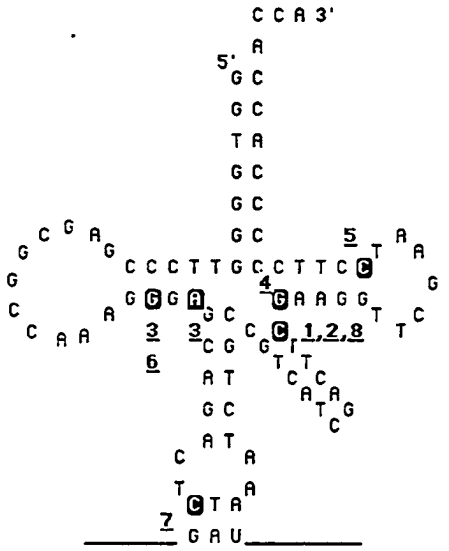

Fig. 1.7. Location of the UV-induced mutations on the cloverleaf structure of the *supF* tRNA. Shadowed letters indicate the position on the tRNA structure of nucleotides observed to be mutated in the *supF* genes of mutant phage rescued from UV-irradiated LN12 cells. Also noted are the numerical designations of the mutant phage which bear the particular mutations. Note that mutant number 3 involves two nearby base changes.

## Discussion.

Much of the current understanding of DNA repair and mutagenesis in mammalian cells depends on extrapolations from mechanisms elucidated in bacteria. We have developed a system to study directly mutagenesis in mammalian cells by exploiting the power of procaryotic genetics. Using a lambda phage vector that is stably integrated into the mammalian cell genome, we have avoided the problem of a high background level of mutagenesis that has been observed in other systems. Also, our results may be more representative of the processes that affect the normal cellular DNA in contrast to the results obtained with extrachromosomal, independently replicating SV40-related vectors. The use of a phage vector is preferable to the use of a cosmid, since Lau and Kan (1983; 1984) report a high frequency of rearrangements among cosmids rescued from mammalian cells.

The efficiency of gene rescue using this system renders it suitable for the study of mutagen specificity. However, future improvements on this system would include the incorporation of an M13 phage origin of replication in the lambda phage vector next to the target gene. Once a lambda phage bearing a mutant *supF* gene is isolated, the plasmid insert can be cut out, self-ligated, and used to transform $F^+$ *E. coli* to ampicillin resistance. Infection of the resulting bacteria with wild type M13 phage leads not only to the production of single-stranded phage genomes but also to the generation of single-stranded copies of the plasmid DNA for easy DNA sequencing (Levinson et al., 1984). Another useful modification would be to use a truncated TK gene, lacking its promoter, as a selectable marker in the phage vector. Transfection of $LTK^-$ cells with such a vector and selection of the cells for the presence of a functional TK gene tends to yield colonies with many copies of the vector DNA (W. P. Summers, personal communication). With more copies of the lambda vector contained in the mouse cells, the system will be more efficient. In future work with any of these systems, packaging extract made from the *E. coli* C strain developed by

Rosenberg et al. (1985) will be used in order to avoid the effects of the *Ecoli* K restriction system active in the extracts made by the method of Hohn (Hohn, 1979; Rosenberg, 1985).

In this study, it was demonstrated that the mutation frequency is not greatly affected by the *umuC* mutation in the *E. coli* strain used for growth of the rescued phage. This result suggests that the UV-induced mutations detected in this system are actually produced in mammalian cells and do not result from bacterial repair of unprocessed lesions persisting in mouse cell DNA.

The use of the *supF* gene as a target gene in these experiments provides the benefit not only of a small gene which can be easily sequenced, but also of a gene which has been studied with regard to the functional implications of a variety of mutations (Celis and Piper, 1982; Rich and RajBhandary, 1976). The results of work that has examined tRNA metabolism, coupled with the data accumulated by other groups studying mutagenesis with the *supF* gene (Hauser et al., 1986; Bredberg et al., 1986; Protic-Sabljic et al., 1986), have demonstrated that a variety of mutations can inactivate the *supF* and other bacterial tRNA genes. In these studies and in the work of this thesis, 443 *supF* mutants have been detected, 307 of which were transition mutations, and 128 of which consisted of the 4 types of transversion mutations. The sites at which mutations have been observed are also quite numerous. Although most mutations that have been found in the *supF* gene are confined to the 90 nucleotide region that corresponds to the tRNA cloverleaf molecule, mutations within this region affecting the tRNA function have been detected at almost 60 of the 90 nucleotides. Hence, while there may be some as yet unknown constraints on the nature of mutations which can be detected using this system, all of the possible types of point mutations have been observed to inactivate the *supF* gene.

The analysis of UV-induced mutations in mouse cells presented here suggests that such mutations tend to occur at Py-C pairs and that the predominant change is a C to T transition at these sites. This specificity is similar to that reported in most studies with *E.*

*coli* (Brash and Haseltine, 1982; Wood et al., 1984; Miller, 1985), supporting the validity of cautious extrapolation from *E. coli* to mammalian cells. Interestingly, no mutations were found at T-T pairs even though T-T pairs are well represented in the *supF* gene, with 31 T-T sites in addition to 59 Py-C sites in the 200 nucleotide-long gene. Brash and Haseltine (1984) suggest that for *E. coli* the main mutagenic lesion may be the Py(6-4)Pyo photoproduct (which occurs frequently at T-C and C-C sites but rarely at C-T and T-T sites) rather than thymidine dimers even though thymidine dimers may occur more frequently in response to UV light. Our data indicate that such may be the case for mammalian cells as well.

For *E. coli*, it has been suggested that when the polymerase encounters an unrepaired photoproduct, it preferentially inserts adenine (Howard and Tesman, 1964). This has been demonstrated with DNA polymerase I *in vitro* (Rabkin et al., 1983). The results presented here suggest that a similar preference may exist for mammalian polymerase(s) as well, since the insertion of adenine opposite cytosine-derived photoproducts residues would account for the predominance of G:C to A:T transitions observed in these experiments. The results also are consistent with the notion that at least one mammalian polymerase can replicate DNA that has been damaged by UV light, although it may do so in an error-prone fashion. Error-free excision repair or lesion repair may go on in the cells, but the detection of mutagenesis by UV light, which generates lesions with impaired base pairing properties, suggests that there is some replication system that tolerates DNA damage. That this system exists and may be inducible by UV light was first suggested by Dasgupta and Summers (1978).

The UV-induced mutations which have been detected so far in this system appear at sites where the two most common UV lesions, cyclobutane dimers and (6-4) photoproducts, occur. Whether the main premutagenic lesion is the dimer or the (6-4) product, this association of the common lesion sites with the sites of mutations suggests that the UV-induced mutagenesis in the mouse cells is targeted to pre-mutational lesions generated by

the UV light. Preliminary results of Sarkar and Summers (personal communication) lends further support to the argument that these mutations are targeted. They have begun to determine the sequence changes among a collection of non-targeted mutations generated in the *supF* gene contained in an SV40 vector replicating in damaged monkey cells (Sarkar et al., 1984). They find a relatively high frequency of transversions in contrast to the predominance of transitions seen here, so their spectrum of non-targeted mutations is different from the spectrum of UV-induced mutations that we have detected.

Similar results were found by Hauser et al. (1986) and of Lebkowski et al. (1985). Hauser and co-workers, using an SV40-based vector, examined mutants generated by passage of the UV-irradiated vector through monkey cells. They found about a 20-fold increase in mutation frequency with UV treatment of the DNA, although the total yield of colonies was down 5-fold. Of their spontaneous mutations, slightly over 50% were large or small deletions, while the remainder were point mutations, split evenly between G:C to A:T transitions and G:C to T:A transversions. Among their UV-induced mutations, over 90% were point mutations, with 61% G:C to A:T and 7% A:T to G:C. The rest of the point mutations were split among the 4 types of tranversions.

Lebkowski and co-workers examined spontaneous and UV-induced mutations occuring in an SV40 shuttle vector replicating in human 293 cells. They used the genetic system of Coulondre and Miller (1977) to deduce the sequence of nonsense mutations occuring in the *lacI* gene in their vector. Their genetic screen detects *lacI* mutants of all types, including insertions, deletions, and missense mutations, but they actually analyze only the nonsense mutations. They estimate that only 25% of the point mutations are nonsense mutations, so the majority of mutations generated in their experiments are not analyzed. Furthermore, because of the limitations of their assay, only 5 of the 6 possible types of point mutations can counted; A:T to G:C transitions are not measured. Miller (1985), however, collected UV-induced mutants in the *lacI* gene in *E. coli*, and he analyzed them both by his

nonsense suppression assay and by DNA sequencing. He showed that the methods yield similar results, in spite of the limitations of the purely genetic system. In any case, the results of Lebkowski and co-workers are similar to those presented here and to those of Hauser et al. They found that 84% of the UV-induced point mutations were G:C to A:T transitions, mostly at TC or CC sites. In contrast, the spontaneous mutations they observed included about as many G:C to T:A transversions as G:C to A:T transitions.

In an early attempt to study UV mutagenesis in mammalian cells, Sarasin and co-workers (Bourre and Sarasin, 1982; Sarasin et al., 1985) examined a collection of UV-induced reversion mutations in a temperature sensitive strain of SV40. About half of these mutations were transversions and half were transitions, in contrast to our results, although, as we observed, most mutations occured at Py-C sites. The reason they detected relatively more transversions may be that they were looking for reversion mutations rather than forward mutations, and so the spectrum of mutations they observed may be biased by the particular requirements for intragenic suppression of the initial mutation.

As a further application of their shuttle vector system, Bredberg et al. (1986) have examined the mutational spectrum of a UV-irradiated shuttle vector propagated in xeroderma pigmentosum cells. Plasmid survival was about 10-fold lower in the XP cells, but the mutation frequency among the survivors was about the same as with the repair proficient human fibroblasts. There was a lower frequency of multiple change mutations, and relatively more single base substitutions. Among the point mutations, there were significantly fewer transversions produced in the XP cells; instead, G:C to A:T transitions accounted for 93% of all mutations, in contrast to 73% for this type of mutation in the control cells. Bredberg and co-workers suggest that the decrease in multiple change mutations may be attributed to the defect in excision repair in the XP cells. They propose that the mutations which involve multiple non-contiguous base changes arise from an error-prone excision repair process (present in normal but not XP cells) that produces mutations at sites that did not necessarily

have pre-mutational lesions but that are near to sites of lesions. They also suggest that transversion mutations, which are under-represented in XP cells, may be related to this excision repair process. In contrast, since XP cells are capable of DNA replication, the G:C to A:T transition mutations that are common in XP cells may be related to adenine insertion during replication of DNA containing TC and CC pyrimidine-pyrimidone (6-4) photoproducts. Because these G:C to A:T mutations are so common to XP cells, they suggest further that these mutations may be associated with the high incidence of cancer seen in XP patients.

Protic-Sabljic et al. (1986) used the same shuttle vector system to investigate the effect of *in vitro* photoreactivation of irradiated vector DNA with *E. coli* photolyase prior to transfection into monkey cells. They found that photoreactivation, which removed about 90% of the cyclobutane dimers but none of the (6-4) photoproducts, lowered the mutation frequency slightly and altered the mutational spectrum. They collected fewer multiple change mutations and fewer G:C to A:T transitions. They conclude that cyclobutane pyrimidine dimers are mutagenic in mammalian cells, causing multiple mutations and transitions. It should be noted however, that with or without photoreactivation, the sites of mutations induced by UV remained the same. Hence, although removal of most of the cyclobutane dimers changes the the frequency and slightly alters the spectrum of mutations, it does not alter the sites of the mutations. For this reason, the suggestions put forward by Protic-Sabljic do not clearly fit all of the results, and the implications of their data regarding the mechnism of UV mutagenesis are not clear cut. They go on to suggest that the (6-4) photoproducts are more likely to cause transversions, but this proposal is questionable because their own data shows that, with or without photoreactivation, transitions (especially G:C to A:T transitions) are the most common mutations produced by UV light.

Lebkowski et al. (1986) also used their system to study sequence changes induced by EMS. They collected 54 mutations, and 53 of these were G:C to A:T transitions. This

predominance of transitions is similar to the effects of UV light, although UV does produce a greater range of mutations. A further difference is that only 28 of these (53%) occurred at pyrimidine-pyrimidine sequences, in contrast to the specificity of UV for these sequences. This difference in mutational spectrum is further evidence that targeted mutations are being studied in these systems.

**Future work: Transgenic mice.**

In this thesis, mutagenesis in mammalian tissue culture cells was studied by taking advantage of the capacity of lambda *in vitro* packaging extracts to cut out and package phage vector DNA from within the DNA of mammalian cells. This approach can in principle be extended from tissue culture cells to the intact animal by the use of techniques to generate transgenic mice. In experiments to study gene expression in transgenic mice, it has already been demonstrated that multiple copies of phage vector DNA injected into mouse embryos can persist in the grown animals (Constantini and Lacy, 1981; Brinster et al., 1981; Lacy et al., 1983). There is, furthermore, no reason to expect that phage recovery from the DNA of an intact animal should be any less efficient than rescue from the DNA of tissue culture cells.

For the transgenic mice, it is necessary to use target genes such as *supF* and *lacI* that can be screened in *E. coli* following phage rescue, since no selection of mutations within the mice is possible. Our experience with λsupF-neo shows that it is desirable to have as many copies of the vector DNA in the cells as possible. Two procedures can be used to enhance the phage vector copy number in the mice. Repeated cycles of microinjection of embryos from already transgenic mice can be undertaken, and selective interbreeding of different transgenic mice containing copies of the phage DNA integrated at different sites can be performed.

Once a transgenic mouse strain with a suitably high number of phage vector copies is

isolated, many experiments that are beyond the scope of the cell culture systems can be tried. Tissue-related and age-related differences in mutagenesis can be studied. Germ line and somatic tissue can be compared. In the whole animal, transplacental mutagenesis by chemicals and intrauterine mutagenesis by radiation can be examined. With the transgenic mouse system, furthermore, physiologically relevant estimates of mutation rates can be obtained.

This project was conceived and initiated in the final stages of this thesis, and so complete development of this system has not been achieved. As a first step, however, a new phage vector for use in the transgenic mice was constructed. The structure of this phage vector, λsI13, is presented in Fig.1.7. Both the *supF* gene and the *lacI* gene were included so that each rescued phage can be screened for the presence of mutations in two target genes. The inclusion of the *lacI* gene also makes the transgenic mouse system useful to workers who prefer to study the *lacI* gene because of the genetic system available for its analysis (Coulondre and Miller, 1977). The M13 origin will facilitate sequencing of mutant genes, as discussed above.

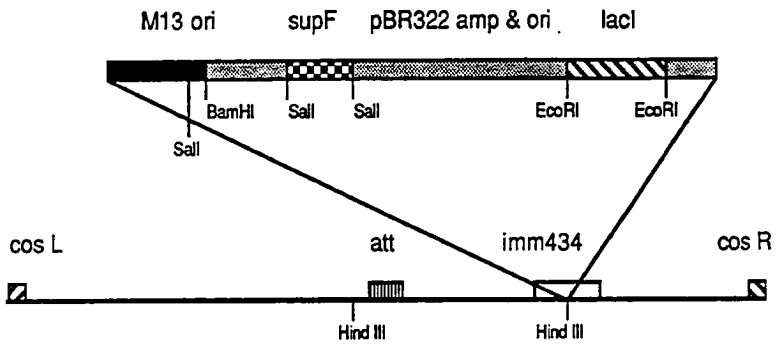


Fig. 1.8. Diagram of the λsI13 vector genome. This phage is a derivative of λL47.1 (Loenen and Brammar, 1980). At one *HindIII* site of the phage, a pBR322-derived plasmid, psI13, was inserted. The plasmid psI13 includes the plasmid pPG2 (see materials and methods, chapter 1) from which the *HindIII* to *BamHI* fragment containing the neomycin resistance gene was removed. The remaining plasmid includes the *supF* gene, the ampicillin resistance gene (amp), and pBR322 origin of replication (ori). To this plasmid, the 0.7 kb *HindIII* to *BamHI* piece from the plasmid psD-Xho (Levinson et al., 1984) containing the M13 origin of replication was added (M13 ori). The 1.7 kb *EcoRI* piece containing the *lacI* gene and the *lacZα* region from the plasmid pMC9 (Calos et al., 1983) was inserted at the *EcoRI* site (lacI). For orientation purposes, the left and right lambda cohesive ends (cosL and cosR) and the lambda *att* site are indicated, along with the locus of the *imm434* gene from λL47.1.

## Chapter 2. Mutagenesis II

### Introduction

One drawback to the system with λsupF-neo is that the isolation of mutants is a tedious process. As an alternative approach, work was undertaken to develop a system in which mutations in the target gene could be selected in cell culture and in which the mutant target genes could still be retrieved by the phage rescue process. In this way, millions of target genes contained in millions of cells could be screened very quickly using selective methods in tissue culture, as opposed to screening thousands of target genes in thousands of rescued phage.

For this approach, λTK was constructed (Fig.2.1), and it was used to transfect mouse $LTK^-$ cells. The vector contains the HSV TK gene, along with the *supF* gene. The TK gene was chosen as a target gene in these experiments because selective conditions have been determined which will select for and against the presence of a functioning TK gene in cell culture. For this system, a cell line with a single copy and only a single copy of the λTK DNA is necessary; in a cell line with multiple copies of the TK gene, a mutation induced in one TK gene could not be detected while the others were functional. This single copy must also be stable, so that a high frequency of spontaneous mutations to the TK-deficient phenotypoe does not obscure the induced mutagenesis being studied. The difficulty in establishing a stable, single copy line is a major obstacle to this approach. The single copy of the λTK vector DNA must also be in a configuration so that it can be cut out and packaged by the lambda packaging extract. The lambda cohesive ends and several hundred base pairs of lambda sequences on either side of them must therefore remain intact, since these sequences are necessary for packaging (Hohn, 1983).

## Materials and Methods

**Phage.** λTK was constructed from the phage λL47.1 (Loenen and Brammar, 1980). The *supF* gene was isolated as a 200 base pair *SalI* fragment from λsupF-neo. It was ligated into the *SalI* site of the plasmid pX1 (Enquist et al., 1979). The resulting plasmid, pXF, was linearized at the *HindIII* site and was ligated into λL47.1 at the rightward of the two *HindIII* sites, in the orientation shown in Fig. 2.1.

**DNA transfection.** In one experiment with λTK, LTK$^-$ cells were transfected with λTK DNA to which a 100-fold molar excess of λgt11 DNA (Young and Davis, 1983) had been added. The phage DNA was self-ligated at a concentration of 300 μg/ml. Transformed cells were selected for the presence of a functional TK gene by MAGGT selection (growth medium containing 0.6 μM methotrexate/ 50 μM adenosine/ 50 μM guanosine/ 0.1 μM glycine/ 16 μM thymidine). In a second experiment with λTK, the λTK DNA was ligated to a 10-fold molar excess of λgt11 *HindIII* end fragments. The 4500 base pair and the 5500 base pair end fragments containing both *cos* sites were gel-purified from a *HindIII* digest of λgt11. Before loading on the preparative agarose gel, the digest was heated to 50°C in order to separate the end fragments that may migrate together in an anomalous band because of annealing at the *cos* sites. The resulting ligation mixture was used to transfect LTK$^-$ cells as above without any purification of the desired ligated molecules.

**Selection for TK mutants.** Mouse L cells containing the λTK vector were distributed into dishes at $10^4$ cells/cm$^2$. They were grown in the presence of MEM supplemented with 10% calf serum and with trifluoro-thymidine (TFT) at a concentration of either 4 or 8 μg/ml. Surviving colonies that were evident 5 days after the imposition of selective conditions were considered to be functionally TK-deficient.

Other methods were as described in Chapter 1.

## Results

**λTK.** The structure of λTK is presented in Fig. 2.1. This vector was made by insertion of a recombinant plasmid, pXF, into λL47.1 at the rightward of the two *HindIII* sites in the phage. This recombinant plasmid was made up of the entire plasmid pBR322 containing two inserts: (1) the 3.5 kb HSV *BamHI* fragment, which includes the TK gene, inserted at the *BamHI* site, and (2) the 200 bp *SalI* fragment from λsupF-neo, containing the *supF* gene, inserted at the *SalI* site. The inclusion of *supF* along with TK in this vector was done to provide a marker in the phage closely linked to the TK gene whose presence could be rapidly detected in rescued phage. Rescued phage that are supposed to carry TK mutations can be screened for the presence of a functional *supF*. Rescued phage lacking *supF* genes could be identified as either contaminating phage from the packaging extract or as phage which had suffered deletions or extensive mutations affecting the *supF* locus as well as the TK locus.

**Transfections.** Mouse LTK$^-$ cells were transfected with λTK DNA in two different ways. Both methods were designed to protect the lambda *cos* sites and flanking packaging signals (Hohn, 1983) so that the vector DNA could be subsequently rescued from the cells. The other goal was to produce cell lines carrying only a single copy of the vector DNA, so that mutations in the TK gene could be detected by selection of the cells in culture.

In the first approach, λTK DNA was mixed with a 100-fold molar excess of λgt11 DNA (Fig. 2.1), and *cos* site ligation of the lambda DNA was carried out prior to transfection by the calcium phosphate co-precipitation method (Graham and van der Eb, 1973). The ligation to λgt11 protects the lambda *cos* site packaging signals, and the choice of the 100:1 ratio is based on the transfection efficiency of the LTK$^-$ cells with λsupF-neo, in which the highest copy number cell line had 100 copies of the vector DNA. It was hoped that although a given cell might receive many copies of lambda DNA, few of these would be

λTK. In the second transfection, the end fragments of λgt11, containing the *cos* sites and the packaging signals, were isolated, and these fragments were mixed with λTK DNA in a ligation reaction prior to transfection. TK-positive colonies were selected following transfection. Of the TK-positive colonies produced in the first transfection, 12 were chosen for further analysis. These were designated LTK1 through LTK12. From the second, 20 TK-positive colonies, designated LTK13 through LTK32, were chosen for study.

**Analysis of cell lines LTK1 through LTK12.** DNA was isolated from cell lines LTK1 through LTK12 and was used to determine the vector copy number in each cell line. The DNA was digested with *BamHI* in order to produce the 3.5 kb *BamHI* fragment containing the TK gene. As standards for quantitation, two cell lines provided by R. M. Liskay which were previously shown to carry single intact copies of the TK gene were included in the analysis. These cell lines were 20-4-28 and Rec2. These cells contain the TK gene within a plasmid construct in which the sequences flanking the TK gene are slightly different from those in λTK (R. M. Liskay, personal communication). Hence, DNA from these cells was cut with *BamHI* and *HindIII* to yield a fragment representing the TK gene of reduced mobility relative to the 3.5 kb *BamHI* fragment from the LTK cell lines. In these cells there are additional fragments of the TK gene of varying sizes. These fragments account for the extra bands of increased moblity seen in the Southern blot. Nonetheless, the slowest mobility bands in the Rec2 and 20-4-28 lanes correspond to the single intact copies of the TK gene in these cell lines. The digested DNA was separated by agarose gel electrophoresis and analyzed by Southern blot using the *EcoRI* fragment of the TK gene as a probe (Wagner et al., 1981). The resulting autoradiogram is pictured in Fig. 2.2. The results indicate that LTK2, LTK7, and LTK8 contain unrearranged TK genes in low copy number. Only these cell lines were studied further.

DNA prepared from LTK2, LTK7, and LTK8 was used as a substrate in lambda *in vitro* packaging reactions. Phage were rescued from all three cell lines, but, as indicated in

Table 2.1, some of the phage rescued had phenoypes different from either λTK or λgt11. λTK has no mutation in the gene for bacterial lysis (the S gene), and it carries the suppressor *supF* gene [*supF*$^-$ (Su$^+$)]; hence, it forms large, light blue plaques on a lawn of *E. coli* SY204 *lac* (Am) in the presence of X-gal and IPTG. On the other hand, λgt11 bears the S100 lysis mutation, and it carries the *lacZ* gene. It forms small, dark blue plaques. The large, colorless plaques and the large, dark blue plaques produced by some of the rescued phage represent novel phenotypes, and these may be the result either of mutation or of recombination between the phage genomes. These phages were not investigated further, so the structures of their genomes are not known. None of the phage rescued from LTK7 had the phenotype of λTK. LTK2 yielded a mixture of phage, while 3 of the 5 phage rescued from LTK8 had the phenotype of λTK.

The frequencies at which the three cell lines spontaneously give rise to functionally TK-deficient subclones were determined. The Rec2 cell line, previously determined to carry a single, stable copy of the TK gene (R. M. Liskay, personal communication), was included in the analysis as a standard of comparison. The cells were seeded into dishes at $10^4$ cells/cm$^2$ in the presence of TFT at either 4 or 8 μg/ml. Within 5 days, TFT resistant colonies were visible. Table 2.2 presents the frequencies of TFT resistant colonies. LTK8 showed the lowest frequency of spontaneous TFT resistant sublones, but it yielded such colonies at a frequency 40 to 70 times higher than did Rec2.

It was thought that this high frequency of spontaneous TFT resistant colonies found with the LTK2, 7, and 8 cell lines might be the result of some heterogeneity in the cell lines. To circumvent this potential problem, 30 subclones of LTK8 were isolated, designated LTK8-1 through LTK8-30, and these were tested in TFT. Six of these sublines ( LTK8-3, 15, 16, 20, 22, and 28) showed a low (~$10^{-4}$) frequency of spontaneous mutation to TFT resistance. DNA for lambda packaging reactions was prepared from these six cell lines and from two others with a high spontaneous mutation rate (LTK8-6 and 11). From the 8

LTK-8 sublones, 38 phage were rescued. All formed colorless plaques, indicative of the loss of the *supF* gene (and probably of the TK gene as well). By plaque hybridization (not shown), it was directly determined that none of the plaques contained the TK gene. As a control, phage were rescued from LTK8 DNA; 3 blue and 2 colorless plaques were obtained. Only the blue plaques were positive for the TK gene by plaque hybridization. Although some of the LTK8 subclones contained stable TK genes, the TK genes were no longer recoverable within a lambda phage. It is likely that rearrangements which may have led to increased stability of the TK genes may have also rendered these genes inaccessible to phage rescue.

**Analysis of cell lines LTK13 through LTK32.** Rearrangements in the LTK8 cell lines may have occured because of recombination between λTK genomes and the multiple λgt11 genomes with which λTK shares considerable homology. To reduce the amount of homologous DNA with which the λTK DNA might recombine, an alternative strategy was employed in which the λTK *cos* sites were protected not by ligation to entire λgt11 genomes but by ligation to just the λgt11 end fragments.

From the transfection of LTK$^-$ cells with λTK DNA ligated to λgt11 end fragments, 20 colonies, designated LTK13 through LTK32, were chosen for further study. One subclone from each cell line was isolated before any further experiments were done in order to avoid the problem of possible heterogeneity of the original colonies.

The cells were placed in medium containing 4 μg/ml of TFT to determine the frequency of spontaneous mutation to TFT resistance, as an index of the functional instability of the TK genes in the cells. Of the 20 cell lines, 9 were found to give off TFT resistant colonies at a low frequency. These were LTK13, 18, 20, 21, 23, 24, 26, 28, and 29. Southern blot analysis of the TK genes in these cell lines was performed, and the results presented in Fig. 2.3 show that all of these cell lines carry multiple copies of the TK gene. This conclusion is based on comparison of the bands representing the TK genes in the LTK

cell lines with the band representing the single intact TK gene copy in the cell line Rec2. Clearly, all the bands from the LTK cell lines are darker than the one from the Rec2 cells. Because these cell lines were judged to contain multiple copies of the TK gene, they were not considered useful and were not studied further.

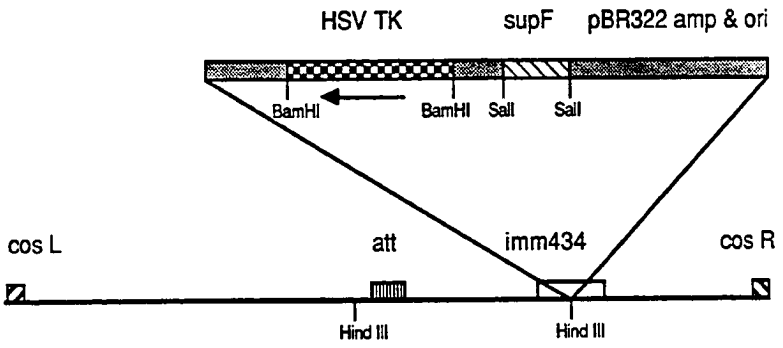


Fig. 2.1. Diagram of the λTK vector genome. This phage is a derivative of λL47.1 (Loenen and Brammar, 1980). At one HindIII site of the phage, a pBR322-derived plasmid, pXF, was inserted. The plasmid pXF consists of the entire pBR322 plasmid with its ampicillin resistance gene (amp), plus the herpes simplex virus thymidine kinase gene (HSV TK) inserted at the *BamHI* site and the *supF* gene [*supF*$^-$ ($Su^+$)] inserted in the unique *SalI* site of the plasmid. For orientation purposes, the left and right lambda cohesive ends (cosL and cosR) and the lambda *att* site are indicated, along with the locus of the *imm434* gene from λL47.1. The arrow indicates the direction of transcription of the TK gene.

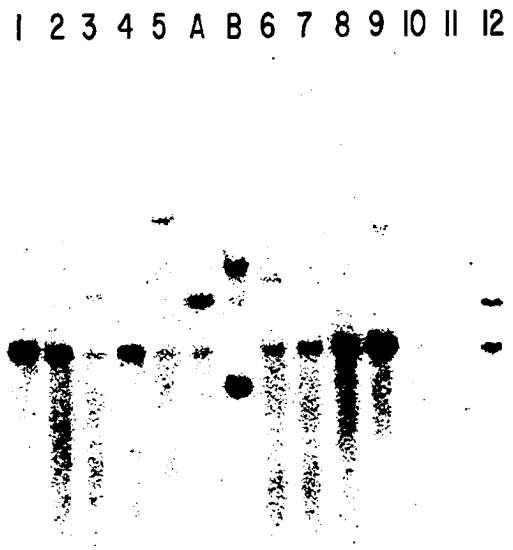


Fig. 2.2. Southern blot analysis of λTK vector DNA contained in transfected mouse LTK$^-$ cells. DNA was prepared from 12 TK-positive cell lines (LTK1 through LTK12) made by transfection of LTK$^-$ cells with λTK DNA. As standards for quantitation, DNA was also prepared from 2 cell lines previously shown to carry single intact copies of the HSV TK gene. These cell lines, Rec2 (lane A) and 20-4-28 (lane B), were supplied by R. M. Liskay. For each cell line LTK1 through LTK12, 10 μg samples of DNA were digested with *BamHI*. DNA samples from Rec2 and from 20-4-28 were digestedwith both *BamHI* and *HindIII* because the sequences flanking the intact TK genes in these cells lack one of the *BamHI* sites but contain a convenient *HindIII* site near the TK gene. The digested DNA was subjected to agarose gel electrophoresis, transferred to a nitrocellulose filter, and hybridized to the $^{32}$P-labelled *EcoRI* fragment of the TK gene. The autoradiograph of the hybridized filter is shown.

Table 2.1. Phenotype distribution of plaques produced by phage rescued from the DNA of mouse LTK2, LTK7, and LTK8 cell lines.

| | | λTK | λgt11 | Novel phenotypes | |
|---|---|---|---|---|---|
| Cell line | Total pfu | Large, Faint blue | Small, Dark blue | Large, Dark blue | Large, Colorless |
| LTK 2 | 41 | 1 | 38 | 0 | 2 |
| LTK 7 | 300 | 0 | 296 | 4 | 0 |
| LTK 8 | 5 | 3 | 0 | 0 | 2 |

Mouse LTK- cells were transfected with a mixture of λTK and λgt11 DNA. Phage were rescued from the DNA of three of the TK-positive cell lines produced in the transfection, and the distribution of phenotypes among the resulting plaques is presented in absolute numbers. The phenotypes of λTK and λgt11 are indicated, as are the novel phenotypes which differ from both λTK and λgt11.

Table 2.2. Frequency of trifluorothymidine resistant colonies arising from TK-positive cell lines in the absence of mutagenic treatment.

| | TFT resistant colonies per $10^5$ cells. | |
|---|---|---|
| | Concentration of trifluorothymidine | |
| Cell line | 4 μg/ml | 8 μg/ml |
| Rec2 | 1 | 4 |
| LTK2 | 93 | 70 |
| LTK7 | 340 | 270 |
| LTK8 | 38 | 31 |

Cells were seeded into dishes at $10^4$ cells per $cm^2$, in the presence of trifluorothymidine at the indicated concentrations. The number of trifluorothymidine (TFT) resistant colonies arising per $10^5$ cells is given.

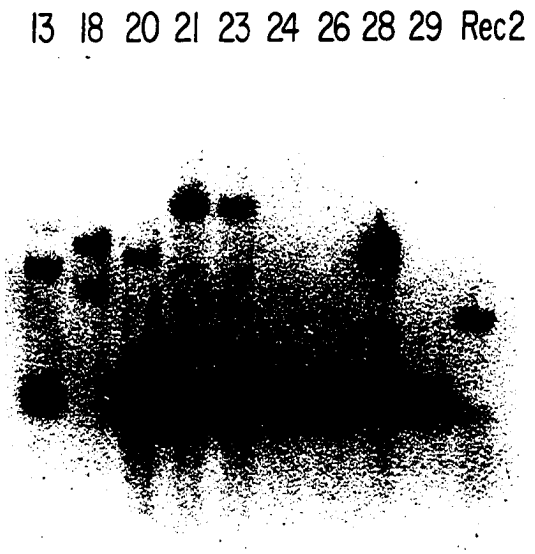


Fig. 2.3. Southern blot analysis of λTK vector DNA contained in transfected mouse LTK$^-$ cells. DNA was prepared from 9 TK-positive cell lines (LTK13, LTK18, LTK20, LTK21, LTK23, LTK24, LTK26, LTK28, and LTK29) made by transfection of LTK$^-$ cells with λTK DNA. As a standard for quantitation, DNA was also prepared from a cell line, Rec2, which was previously shown to carry a single intact copy of the HSV TK gene (R. M. Liskay, personal communication). For each LTK cell line, 10 μg samples of DNA were digested with *BamHI*. The DNA sample from Rec2 was digested with both *BamHI* and *HindIII* because the sequences flanking the TK gene in this cell line lack one of the *BamHI* sites but contain a convenient *HindIII* site near the TK gene. The digested DNA was subjected to agarose gel electrophoresis, transferred to a nitrocellulose filter, and hybridized to the $^{32}$P-labelled *EcoRI* fragment of the TK gene. The autoradiograph of the hybridized filter is shown.

## Discussion

Attempts were made to develop a system in which the TK gene could be used as a target gene for the study of mutagenesis in mammalian cells. A requirement for such a system is the establishment of a cell line in which the TK gene is present in a single copy so that mutations in the gene can be selected in tissue culture. This cell line must have a stable TK-positive phenotype; otherwise, apparent TK-negative cells may arise from the cell line at too high a frequency to detect TK-negative cells produced by the action of mutagens under study. The production of single copy cell lines is possible, but such cell lines are uncommon. By limiting the proportion of TK DNA in a calcium phosphate-DNA precipitate used for transfection, the yield of single copy cells may be enhanced, but many colonies must still be analyzed. In the two gene transfer experiments described, 32 colonies were studied, and of these, only 3 appeared to be single copy lines. An alternate method of gene transfer, such as microinjection of DNA into cells, may generate a higher portion of single copy cell lines. In future experiments microinjection should be tried.

Stability of the single copy TK gene is also a problem. In the absence of mutagen treatment, spontaneous deletions and rearrangements can produce TK-negative cell lines. The factors affecting the stability of genes introduced into mammalian cells are not completely known, but the work of groups attempting to establish single copy *gpt* cell lines indicates that there is considerable variability in the stability of introduced genes (Thacker, 1985). Thus, many single copy cell lines may have to be screened in order to find an acceptably stable one. To address this problem, another selectable gene, such as *neo*, may be linked to the TK gene within the phage vector. Sustained selection of the cells for the presence of a functional *neo* gene at the same time that TFT selection is imposed may help to eliminate the colonies that arise in TFT because of deletion or rearrangement events involving the TK gene.

Another potential problem is that there may be mechanisms in mammalian cells that can generate functionally TK-negative cells without structurally altering the TK sequences. One such mechanism may be inactivation of the TK gene by methylation. The nearby presence of a *neo* gene whose activity is required for cell growth may help to avoid this problem.

In further work with the λTK system, it might be worthwhile to include the M13 origin of replication near to the TK gene in the lambda vector. As suggested for the λsupF-neo system in Chapter 1, this modification would facilitate sequencing of the mutant target genes. A set of thymidine kinase deficient *E. coli* strains containing deletion mutants of the HSV TK gene driven by a bacterial promoter would further streamline mutant gene analysis by allowing rapid mapping of the mutations by recombination and TK selection in *E. coli*.

## Chapter 3. Mismatch repair

### Introduction

The repair of base pair mismatches in the DNA of an organism plays an important role in reducing the frequency of mutations and in preserving the genetic integrity of the organism. Mismatches can occur in several ways. Recombination events can generate heteroduplex regions in DNA, and in some models the process of gene conversion is thought to involve heteroduplex structures as intermediates. In DNA replication, errors can occur which produce mismatched bases that must be corrected in order to avoid a high rate of spontaneous mutagenesis.

In *E. coli*, DNA mismatch repair has been extensively studied. Transfection experiments with heteroduplex bacteriophage DNA (reviewed by Lindahl, 1982) have demonstrated that *E. coli* has an efficient system for mismatch repair. The power of procaryotic genetic analysis has allowed identification of mutants deficient in various aspects of the repair process, and this has led to an understanding of some of the mechanisms involved. The products of the *mutH*, *mutL*, *mutS*, and *mutU* (mutU and uvrD are synonymous) genes all seem to play a role in mismatch repair. Mutations at these loci produce strains which undergo a high rate of spontaneous mutagenesis because they cannot repair DNA mismatches effectively (Lindahl, 1982). Experiments involving the transfection of *E. coli* with hemi-methylated λ DNA coupled with the identification of mutants with altered function in the methylase encoded by the *dam* gene have led to a model of methyl-directed strand selection in *E. coli* mismatch repair requiring the *mutH*, *mutL*, *mutS*, and *mutU* gene products. (Glickman and Radman, 1980; Herman and Modrich, 1981; Marinus et al., 1984; Pukkila et al., 1983). This system is believed to correct replication errors in newly synthesized DNA by recognizing the fully methylated strand as the parental

strand and the unmethylated strand as the newly synthesized daughter strand. Mismatch repair tends to occur on the daughter strand in order to preserve the parental sequence.

Reports by Fishel and Kolodner (1983) and by Fishel et al. (1986) have described efficient *E. coli* repair of mismatches in DNA with no asymmetry in methylation patterns. This process involves the *mutS* and *mutU* but not the *mutH* and *mutL* gene products, and, like the methyl-dependent pathway, this process involves excision and repair tracts several thousand nucleotides long. Fishel and Kolodner (1983) also described a second methyl-independent pathway that requires the *recF* and *recJ* gene products. This less efficient pathway involves excision and synthesis tracts less than 300 nucleotides long. The methyl-independent pathways, as well as the methyl-dependent pathway, are thought to be important in the repair of heteroduplex DNA formed in the process of genetic recombination or by the action of mutagenic agents.

The development by Lu et al. (1983) of a system to detect mismatch repair in cell free extracts of *E. coli* has greatly facilitated study of the enzymology of this process. They used pre-formed heteroduplex molecules with mismatches at an *EcoRI* site in phage f1 R229 DNA as a substrate in an assay for mismatch repair in *E. coli* extracts. The heteroduplex molecule is resistant to *EcoRI* digestion, but the appropriately repaired molecule is sensitive to the enzyme. They found that the *in vitro* activity is dependent on ATP, the state of the *dam* methylation of the heteroduplexes, and on the products of the *mutH*, *mutL*, *mutS*, and *mutU* genes. Crude fractions deficient in these gene products were found to complement in the cell free system. The *in vitro* reaction was accompanied by repair synthesis on the unmethylated strand. Using this system, Su and Modrich (1986) have identified the *mutS* gene product as a 97,000 MW protein that specifically binds to DNA regions containing single base pair mismatches.

The study of mismatch repair in cell free extracts was extended to yeast by Muster-Nassal and Kolodner (1986). Using an assay similar to that of Lu et al., they found

that extracts of mitotic cells of *Saccharomyces cerevisiae* catalyzed the correction of mismatches in a reaction requiring $Mg^{2+}$, ATP, and the four deoxynucleotides. They estimated that mismatch correction was accompanied by specific incorporation of less than 20 nucleotides around the site of the repaired mismatch.

In contrast, much less is known about the mechanisms of heteroduplex repair in mammalian cells. Microinjection experiments have demonstrated that mouse cells can efficiently correct mismatched bases in exogenously prepared heteroduplex DNA (Folger et al., 1985). In this study, linear heteroduplex molecules were prepared from plasmids carrying different amber mutations in the neomycin resistance gene. Injection of these molecules into mouse cells yielded G418 resistant colonies at a 10-fold higher frequency than co-injection of the separate linear mutant homoduplexes. Transformation to G418 resistance with the heteroduplex molecules was as efficent as transformation with plasmid DNA containing the wild type *neo* gene.

Studies by Hare and Taylor (1985) involving the transfection of monkey cells with hemi-methylated SV40 DNA have suggested that strand selection in mismatch repair in these cells may be influenced by methylation patterns. In these experiments, heteroduplex viral DNA was prepared from plasmids containing pBR322 plus either the wild type SV40 genome or the genome of a mutant SV40 virus containing a G:C to A:T transition that destroys a *HpaII* site and generates a *PvuII* site. The heteroduplex DNA was used to transfect monkey CV-1 cells by the DEAE-dextran method, and the plaques were analyzed for the presence of mutant and wild type virus. It should be noted that the authors claimed that each plaque arose from the transfection of a cell with a single molecule of heteroduplex virus DNA , and so single repair events were thereby analyzed. However, they offered no proof of this. In any case, methylation of either the mutant or wild type strands was accomplished prior to the formation of the heteroduplexes using either the *HpaII* methylase or the *HhaI* methylase in *in vitro* reactions or by preparing the DNA in *E. coli* with active

*dam* and *dcm* methylases. The results suggested that hemi-methylation at the two *HhaI* sites bracketing the mismatch directed repair to occur on the unmethylated strand, since among the plaques assayed in this case, 37 of 48 were pure plaques, and all of these had the genotype of the methylated strand. Other experiments showed similar effects associated with *HpaII* methylation and with *dcm* and *dam* methylation, although these effects were not as striking. It was found that nicks in the heteroduplex plasmid DNA, which are the result of the heteroduplex preparation process, have as strong an effect on strand selection as methylation does. This effect may be related to the distance of the nick from the mismatch site.

We have done preliminary experiments to investigate mismatch repair *in vivo* by transfection of monkey cells with an SV40-based shuttle vector containing mismatches in the *supF* gene. The vector DNA was rescued from the monkey cells and used to transform *E. coli* in order to analyze the *supF* genes. These experiments yielded results consistent with the presence of a mismatch repair activity in monkey cells, but the results could also have been explained by post-replicative recombination. Attempts were made to examine the effect of asymmetric methylation of heteroduplex DNA, in both the *E. coli* and mammalian modification patterns, in transfection experiments using suitably constructed SV40 vectors. Conclusive results have not yet been obtained.

The further study of mismatch repair in mammalian cells is hampered by the difficulty of genetic manipulation of these cells. The sort of genetic analysis used with *E. coli* which enabled the discovery and characterization of the several mismatch repair pathways is not currently applicable to mammalian cells. As another approach, we developed an assay to detect DNA mismatch repair in cell free extracts of mammalian cells as a way to study the mechanisms of repair directly *in vitro*. Such an approach to the study of other aspects of mammalian DNA and RNA metabolism has already been described by several groups (Darby and Blattner, 1984; Kucherlapati et al., 1985; Challberg and Kelly, 1979; Li and Kelly, 1984; Manley et al., 1980; Wobbe et al., 1985). In work with extracts from HeLa

cells, we have been able to detect extract-dependent repair of DNA mismatches in specifically constructed substrate heteroduplex DNA. Experiments indicate that this activity is dependent upon ATP, deoxynucleotides, and on DNA polymerization. These experiments were facilitated by the development of a biological assay to detect the repair of the heteroduplex DNA in the extracts. This biological assay required the construction of a strain of *E. coli* deficient in heteroduplex repair, but in which prior repair of heteroduplex DNA could be detected.

## Materials and methods

**Cells.** HeLa cells were obtained from Dr. P. Ghosh, Yale University. The COS-7 derivative of the African green monkey kidney cell line CV-1 contains the T antigen gene of SV40 (Gluzman, 1981). The construction of *E. coli* SY204 *(lacZ125*(Am), *trp-49*, *hsdR2::Tn10)* was performed by Sarkar et al. (1984). *E. coli* EG826 *(lacZ125*(Am), *trp-49*, *hsdR2::Tn10*, *ssb-1*, *malE::Tn10)* was made by Dr. Efim Golub by P1 transduction of the *ssb-1* mutation (Meyer et al., 1979) into SY204. Strain SY208 *(lacZ125*(Am), *hsdR2::Tn10*, *mutH3*, *mutU4*, *strA143)* was constructed from *E. coli* KL874 *($F^-$*, *hisF818*, *leu-3 lacZ498*, *strA143*, *mutH3*, *mutU4)* by first introducing a deletion spanning the *lac* and *pro* loci by conjugation and then introducing the *lacZ125*(Am) mutation by a second conjugation, followed by introduction of the host restriction mutation *hsdR2* by P1 transduction with *Tn10* tetracycline resistance. Strain SY209 was constructed by P1 transduction of the *recA56* mutation from *E. coli* MC136 *($argH^-$*, *trpA36*, *srl-300::Tn10*, *recA56)* into *E. coli* SY208 using cotransduction of *Tn10* as a screen after first curing strain SY208 of tetracycline resistance. *E. coli* SY302 was constructed by P1 transduction of the *recA56* mutation, again from *E. coli* MC136, into SY204 using cotransduction of of *Tn10* as a screen after first curing SY204 of tetracycline resistance. Strain GM272 *($F^-$*, *dcm-6*, *dam-3*, *metB1*, *thi-1?*, *hsdS21*, *lacY1* or *lacZ4*, *galK2*, *mtl-2*, *tonA2* or *tonA31*, *tsx-1* or *tsx-78*, $\lambda^-$, *supE44)* prepared by M. Marinus, was obtained from the *E. coli* genetic stock center at Yale. A summary of the strains used in these experiments is presented in Table 3.1.

**Plasmids.** The plasmid p3AC (Fig. 3.1) was constructed by Sarkar et al. (1984). It consists of the plasmid pBR322 into which was inserted the 200 base pair *EcoRI* fragment from the plasmid πVX which contains the suppressor *supF* gene [*$supF^-$* ($Su^+$)]. A discussion of notation for describing the *supF* gene is included in the Materials and Methods section of Chapter 1. The plasmid was also modified by deletion of the *HaeII* fragment B

and by insertion of the *BamHI* to *HpaII* early region of SV40 at the *ClaI* site. Plasmids p3AC-4 and p3AC-8 differ from p3AC only in that they bear single point mutations in the *supF* gene, rendering this gene non-functional. Plasmids p3AC-4 and p3AC-8 were isolated by passage of the plasmid p3AC through monkey COS cells in the course of the study of mutagenesis by Sarkar et al. (1984). The *supF* genes in these plasmids were sequenced directly from the plasmid DNA by the chain termination method (Sanger et al., 1977) using a pBR322 *EcoRI* site primer (New England Biolabs).

**Heteroduplex preparation.** For the preparation of heteroduplex molecules from p3AC-4 and p3AC-8, 25 μg of p3AC-4 linearized at the *ScaI* site and p3AC-8 linearized at the *BamH1* site were mixed in a total volume of 1 ml of water to which 110 μl of 1N NaOH were added. After 30 minutes at room temperature, 110 μl of 1M $NaH_2PO_4$ and 1280 μl of deionized formamide were added. The solution was incubated overnight at 37° C, followed by overnight dialysis at room temperature against 10 mM Tris and 1 mM EDTA, pH 8. The DNA in the sample was concentrated by ethanol precipitation and examined by agarose gel electrophoresis for succesful generation of the nicked circular duplexes representing heteroduplex molecules, prior to use in the experiments described below.

**Gene transfer methods.** COS-7 cells were propagated as described by Gluzman (1981). Cells were subcultured at a dilution of 1:4 in Dulbecco modified Eagle (DME) medium supplemented with 5% fetal calf serum on plates (diameter, 10 cm) 2 days before transfection. Calcium phosphate-DNA coprecipitates were prepared by the method of Chu and Sharp (1981). After trypsin treatment, cells from four 10-cm plates were suspended in DME medium and pelleted by centrifugation at 1000 rpm for 10 min in a Sorvall GLC-2 centrifuge. The cells were then suspended in 2.5 ml of calcium phosphate-DNA precipitate and incubated at room temperature for 15 min. DME medium (40 ml) supplemented with 12.5 mM $CaCl_2$ and 0.1 x HEPES-buffered saline was then added to the tubes, and the cells were transferred to dishes. The cells were allowed to adhere to the dishes in the incubator

for 6 h. The medium was aspirated, and the cells were shocked with 1 ml of 25% glycerol in DME medium for 1 min at room temperature and rinsed twice with warm medium. Fresh medium was added to the plate, and cells were grown for 48 h. DNA recovered from the COS cells by the method of Hirt (1966) was used to transform *E. coli* SY204 to ampicillin resistance by the method of Hanahan (1983). Bacteria transformed to ampicillin resistance were screened for ß-galactosidase activity by growth in the presence of the chromogenic indicator, 5-bromo-4-chloro-3-indolyl ß-D-galactopyranoside (X-gal) at 50 µg/ml and isopropyl ß-D-thiogalactopyranoside (IPTG) at 20 µg/ml.

**HeLa cell extracts.** Suspension cultures of HeLa cells were grown in RPMI 1640 medium (Gibco) supplemented with 10% fetal calf serum. Whole cell extracts, from 2-3 liters of culture containing 4.5 - 5 x $10^5$ cells/ml, were prepared essentially as described by Manley et al. (1980) with some minor modifications. Cells were washed in PBS, and the cell pellet was resuspended in four packed cell volumes of 0.01 M Tris pH 7.9/ 0.001M EDTA/ 0.005 M dithiothreitol. After 20 min, the cells were lysed by homogenization in a Dounce homogenizer. Four packed cell volumes of 0.05 M Tris pH 7.9/ 0.01M $MgCl_2$/ 0.002 M dithiothreitol/ 25% sucrose/ 50% (vol/vol) glycerol were added with gentle stiring. To this suspension, one packed cell volume of saturated $(NH_4)_2SO_4$ was added dropwise. After 20 min, the extract was centrifuged at 50,000 rpm for 3 hours. The supernatant was decanted, and solid $(NH_4)_2SO_4$ was added to it at 0.33 g/ml. After the $(NH_4)_2SO_4$ was dissolved, 0.01ml of 1N NaOH per 10 g of $(NH_4)_2SO_4$ was added, and the suspension was stirred for 30 min. Cellular material precipitated by the 60% ammonium sulfate was then collected by centrifugation at 15,000 x g for 20 min and resuspended in 0.1 volume of the high speed supernatant of a buffer containing: 50mM Tris pH 7.9, 6mM $MgCl_2$, 0.2 mM EDTA, 40 mM $(NH_4)_2SO_4$, 15% glycerol, and 1 mM DTT. This solution was dialysed for 15 hours at 4° C against two 500 ml volumes of 20 mM HEPES pH 7.9, 100 mM KCl, 12.5 mM $MgCl_2$, 0.1 mM EDTA, 17% glycerol, and 1 mM DTT. Precipitated material was

removed by centrifugation (10 min at 12000 x g at 4° C), and the resulting supernatant was quick-frozen in aliquots and stored at -80° C. Protein concentrations of the extracts were 10-12 mg/ml. These extracts were also used for *in vitro* transcription reactions and were found to be active in this assay, producing predicted, specific, template-dependent products.

**Reaction conditions.** Reactions were carried out at 37° C for 2 hours in a total volume of 25 µl containing 15 µl of extract (150-180 µg of protein), between 100 and 500 ng of DNA in 5 µl of water, and 5 µl of the appropriate buffer. The final concentrations of components in the complete reaction were: 12 mM HEPES pH 7.9, 60 mM NaCl, 60 mM KCl, 7.5 mM $MgCl_2$, 3mM $MgSO_4$, 0.1 mM each of dGTP, dATP, dTTP, and dCTP, 0.9 mM ATP, 10 mM creatine phosphate, 10 µg/ml creatine kinase, 0.2 mM DTT, and 10 % glycerol. Reactions were terminated by the addition of 125 µl of 10 mM Tris pH 8, 5 mM EDTA, 0.1 % SDS, and 200 µg/ml of proteinase K. After 1 hour at 37° C, protein was extracted from the samples with phenol and chloroform, and the DNA was precipitated with ethanol, redissolved, and used for bacterial transformations.

**DNA polymerase assays.** Extract reactions were carried out as described above except that unlabelled dCTP was omitted and [$\alpha$-$^{32}$P]dCTP was added to all the reactions at a concentration of 20 µM and a specific activity of 3000 Curies/millimole. The buffer was adjusted according to the desired experimental conditions. Just prior to the phenol extraction step, aliquots of each sample either were spotted onto filter disks for measurement of radioactivity incorporated into TCA insoluble material or were subjected to agarose gel electrophoresis and autoradiography.

Table 3.1. *E. coli* strains.

| Strain | Markers | Source |
|---|---|---|
| SY204 | *lacZ125, trp-49, hsdR2::Tn10* | S. Sarkar |
| SY208 | *lacZ125, mutH3, mutU4, hsdR2::Tn10 ( F⁻, hisF818, leu3, strA143)* [a] | This thesis. |
| SY209 | SY298, *recA56, srl-300::Tn10* | This thesis. |
| SY294 | SY204, $tet^S$ | This thesis. |
| SY298 | SY208, $tet^S$ | This thesis. |
| SY301 | SY204, *umuC122::Tn5* | This thesis. |
| SY302 | SY294, *recA56, srl-300::Tn10* | This thesis. |
| EG826 | SY294, *ssb-1, malE::Tn10* | E. Golub |
| MC136 | *argH, trpA36, srl-300:Tn10, recA56* | K. B. Low |
| KL874 | *F⁻, hisF818, leu3, lacZ498, strA143,mutH3, mutU4* | K. B. Low |
| GM272 | *F⁻, dcm-6, dam-3, metB1, thi-1?, hsdS21, lacY1* or *lacZ4, galK2, mtl-2, tonA2* or *tonA31, tsx-1* or *tsx-78, λ⁻, supE44* | M. Marinus |

[a] Placement of markers in parentheses indicates that the presence of these markers was not confirmed.

## Results

**Heteroduplex preparation.** Heteroduplex plasmid DNA was constructed from derivatives of the plasmid, p3AC, shown in Fig. 3.1. This plasmid contains the pBR322 origin of replication and ampicillin resistance gene, along with the SV40 replication origin and T-antigen gene. It also contains the *supF* gene [*supF*$^+$ ($Su^+$)], which, as explained in Chapter 1, is an amber suppressor tyrosine tRNA gene of *E. coli*. When a plasmid bearing a suppressor *supF* gene [*supF*$^+$ ($Su^+$)] is introduced into an *E. coli* strain that has an amber nonsense mutation in the ß-galactosidase gene, and the resulting bacterial colonies are grown in the presence of the lac operon inducer IPTG and the chromogenic, ß-galactosidase indicator, X-gal, these colonies are blue. When the suppressor gene is absent or is non-functional, the colonies formed are white. We isolated derivatives of the plasmid, p3AC, in which single point mutations had been introduced into the *supF* gene by passage through monkey COS cells, eliminating its suppressor activity. The sequences of the mutant *supF* genes were determined, and two plasmids with mutant *supF* genes [*supF*$^+$ ($Su^-$)] were chosen for the present study. The mutation contained in the *supF* gene in plasmid p3AC-4 is a T to C transition, while that in p3AC-8 is a C to T transition in a site in the gene 61 base pairs away.

From p3AC-4 and p3AC-8, heteroduplex molecules each containing two single base pair mismatches within the *supF* gene were constructed. The scheme for construction of the heteoduplexes is diagrammed in Fig. 3.2. The circular plasmids were separately converted to linear molecules by digestion with different restriction enzymes which cut each plasmid only at one site. The linear molecules were mixed, denatured, and allowed to renature. In the renaturation step, some strands from p3AC-4 anneal to strands from p3AC-8, since they are homologous except for two base pairs out of about 6200. Because the plasmids were linearized at different sites, annealing between strands from the different plasmids yields

linear molecules with large complementary single stranded overhangs. These single stranded regions can anneal among themselves, combining either with the complementary single strand on the same molecule or with one from another molecule, yielding nicked circular molecules or multimers, respectively. In contrast, reannealing of strands from the same plasmid regenerates the original linear molecules.

The results of the heteroduplex preparation from p3AC-4 and p3AC-8 are illustrated in Fig. 3.3, which shows an analysis by agarose gel electrophoresis of the steps in this process. The first two lanes after the marker lane show the uncut plasmids. These preparartions differ slightly because of different monomer to dimer ratios, but the mobility of the molecules are the same. Furthermore, the plasmids digested with their respective restriction enzymes are shown in lanes 3 and 4, and these show bands of identical mobility. As a control, p3AC-8 was linearized with *BamHI*, denatured, and allow to renature by itself, in the absence of p3AC-4. This is shown in lane 5. The results of the heteroduplex preparation, in which the two linearized plasmids were denatured and renatured together, is presented in lane 6. The new band of reduced mobility (relative to the linear molecules) in lane 6 represents the nicked, circular heteroduplex molecules. The nicked circular molecules are similar to the plasmid p3AC except for the presence of two single base pair mismatches within the *supF* gene. Since these molecules are formed by the combination of either of the two strands in p3AC-4 with its complement in p3AC-8, there are two possible heteroduplex molecules (Fig. 3.4), each with two mismatches, that are presumably formed in equal amounts. In either case, note that both strands bear a base change that inactivates the *supF* gene. Semi-conservative replication of the unrepaired heteroduplexes would simply yield the original mutant plasmids with defective *supF* genes. In the absence of post-replicative recombination, it is only by repair of one or both strands to the normal base prior to replication that a functional, suppressor *supF* gene [*supF*$^{-}$ ($Su^{+}$)] can be generated. For most of the studies, the heteroduplex plasmid DNA was used without separation from the linear

molecules.

**Mismatch repair *in vivo*.** We sought to detect mismatch repair in monkey cells by transfection of the cells with various plasmid constructs containing pre-formed base pair mismatches. The extra-chromosomal plasmid DNA was then harvested by the procedure of Hirt (1967), and the fate of the mismatches was analyzed in two ways. The DNA was either used to transform *E. coli* SY204 to ampicillin resistance in the presence of X-gal in order to examine *supF* gene function, or the DNA was subject to restriction enzyme digestion and Southern blot analysis to examine the fate of the heteroduplex restriction sites in the plasmid substrates.

Heteroduplex DNA made from p3AC-4 and p3AC-8 was used to transfect monkey COS-7 cells. After 48 hours, extrachromosomal DNA was harvested from the cells and was used to transform*E. coli* SY204 to ampicillin resistance in the presence of X-gal and IPTG in order to assess *supF* gene function. The phenotypes of the resulting colonies are given in Table 3.2. It can be seen that passage of the DNA through the monkey cells has produced plasmids of which 4.9% carried functional, suppressor *supF* genes [ *supF*$^-$ ($Su^+$)]. Based on three observations, it can be concluded that most of these functional *supF* genes were produced in the monkey cells. As shown in Table 3.2, direct transformation of SY204 with the heteroduplex DNA yields 2.0% blue colonies. Secondly, it has been shown that over 95% of p3AC DNA recovered from COS-7 cells in transfection experiments such as these has replicated (Sarkar et al., 1984) Since replication resolves the heteroduplexes into homoduplex plasmids, little or no heteroduplex DNA will be present upon rescue from the monkey cells. Thirdly, data presented in Table 3.6, in the context of the study of recombination, show that co-transformation of *E. coli* SY204 with the homoduplex mutant plasmids, p3AC-4 and p3AC-8, yields less than 0.01% blue colonies. Taken together, it can be calculated that at most the replicated 95% portion of the plasmid DNA can yield 95% x 0.01% = 0.0095% blue colonies, and the 5% unreplicated DNA can yield at most 5% x 2%

= .01% blue colonies, for a total of no more than 0.02%, much less than the observed 4.9%.

Although it is clear that the processing of the heteroduplex DNA has occured in the monkey cells, it cannot be said with certainty that all of this processing involves mismatch repair as opposed to recombination. As one measure of extrachromasomal plasmid recombination, plasmids p3AC-4 and p3AC-8 were co-transfected into monkey cells, and the DNA harvested from these cells was analyzed in *E. coli* SY204. The results of this experiment are also presented in Table 3.2, which shows that no blue colonies were detected out of 1800, so that there is less than 0.06% recombination under these circumstances. This result, however, may not be directly comparable to the heteroduplex experiment, since, with co-transfection, the frequency at which both plasmids enter the same cell is not known. With transfection of the heteroduplex, at least one strand from each plasmid must enter each successfully transfected cell. These experiments, therefore, while suggestive of mismatch repair in the monkey cells and consistent with previous observations (Folger et al., 1985; Hare and Taylor, 1985), are not conclusive. An approach similar to that of Folger and co-workers using microinjection would be better able to evaluate the contribution of recombination to the results.

**Mismatch repair *in vitro*.** The study of mismatch repair in mammalian cell extracts depends on a method to detect and measure such repair. We chose to develop a biological assay which would exploit the power of *E. coli* genetics. This entailed generating mutations in a gene which has a discernable phenotype in *E. coli* and using the mutant genes to prepare heteroduplex DNA. This heteroduplex DNA was incubated in the mammalian cell extracts, and the DNA was recovered from the extracts and used to transform a suitable strain of *E. coli*. Analysis of the phenotypes of the transformed *E. coli* allowed detection of heteroduplex repair. A crucial aspect of this assay was the construction of a strain of *E. coli* which was deficient in the metabolism of the heteroduplex DNA. This was needed so that the processing of the heteroduplex DNA could be attributed to reactions occuring in the

mammalian cell extracts rather than to *in vivo* repair in the bacterial cells.

**E. coli strains.** Our experimental goal was to use heteroduplex DNA bearing mismatches in the *supF* gene as a substrate for assaying mismatch repair in mammalian cell extracts. The repair of one or both of the mismatches in the heteroduplex plasmid DNA could be detected by using the DNA, after incubation in the extracts, to transform an *E. coli* strain bearing an amber mutation in the ß-galactosidase gene to ampicillin resistance in the presence of X-gal. The appearance of blue colonies would indicate a newly-generated functional, suppressor *supF* gene [ *supF*$^-$ ($Su^+$)]. *E. coli* cells, however, are themselves able to repair mismatches in DNA. In addition, replication of unrepaired heteroduplex molecules followed by recombination between the resulting mutant plasmids could generate plasmids with functional *supF* genes. Hence, transformation of *E. coli* with heteroduplex DNA in the absence of prior incubation with cell extracts could yield a significant proportion of blue colonies due to DNA metabolism within the bacteria. Since mismatch repair and recombination in *E. coli* are efficient processes, the effect of a mammalian cell extract on the heteroduplex DNA might be too small to detect above the background in the assay from bacterial processing of the heteroduplex.

To circumvent the background problem, we constructed several *E. coli* strains with mutations in some of the genes thought to play a role in mismatch repair, recombination, or both. All strains also contained amber mutations in the ß-galactosidase gene in order to be useful in the assay for *supF* activity. These strains were transformed to ampicillin resistance with heteroduplex, and the number of blue colonies as a percentage of the total in each case was determined. Table 3.3 gives the results of this experiment, along with the relevant genotype of the strains studied. It can be seen that a mutation in the gene for the single strand DNA binding protein of *E. coli*, as in EG826, has a measurable effect on the metabolism of heteroduplex plasmids in *E. coli*. A similar effect is seen as a consequence of a mutation in the *recA* gene, as in SY302. In contrast, the presence of mutations in both the

*mutH* and *mutU* genes, as in SY208, has no detectable affect on the outcome of this assay relative to the wild type. The additional presence of a mutation in the *recA* gene along with mutations in the *mutH* and *mutU* genes, as in SY209, however, reduces the percentage of blue colonies produced by a factor of twelve relative to both the wild type and the *mutH*, *mutU* double mutant, and by a factor of about five relative to the *recA* -deficient strain.

These results suggest that the *recA* protein has a significant role in the processing of heteroduplex plasmid DNA, probably in post-replicative recombination, but perhaps also in the mismatch repair process itself. The *mutH* and *mutU* gene products also play a role in the processing of the heteroduplex plasmid DNA leading to the generation of functional *supF* genes, but, in this assay, their role is manifest only in the presence of a mutation in the *recA* gene. It should be noted that this heteroduplex DNA is derived from plasmids grown in *E. coli* SY204, which is wild type with respect to the *E. coli* DNA methylation systems. Hence, the heteroduplex DNA is fully methylated (according to the *E. coli* pattern) on both strands. Since the mismatch repair system in *E. coli* that is related to the *mutH* and *mutU* genes is normally guided by differences in the methylation patterns on the strands of a heteroduplex, as in newly synthesized, hemi-methylated DNA, the limited effect of the *mutH* and *mutU* mutations in this assay might be related to the fact that the heteroduplex DNA is fully methylated.

The mismatch repair activity in these various *E. coli* strains was further investigated by using hemi-methylated plasmid substrates containing a single base pair mismatch. The heteroduplex plasmids were formed from p3AC containing the functional *supF* gene, and from p3AC-4, containing a mutant *supF* gene due to a G:C to A:T transition, as described above. The plasmids were prepared in either SY204 or in GM272, a *dam*, *dcm*, and *hsdS* mutant in which plasmid DNA is not methylated, as shown in Fig. 3.5. This figure shows an agarose gel analysis of DNA fragments generated by restriction enzyme digestion of the plasmids prepared in the two strains. The enzymes used were *BstNI*, *EcoRII*, *DpnI*, and

*MboI*. The action of *EcoRII* is inhibited by *dcm* methylation at its restriction site, while the action of its isoschizomer, *BstNI*, is independent of methylation. The action of *DpnI* is dependent on *dam* methylation, while its isoschizomer, *MboI* is inhibited by *dam* methylation. The results show that the plasmids prepared in SY204 are almost fully methylated, while those from GM272 have little or no methylation.

Heteroduplexes were made from methylated p3AC and unmethylated p3AC-4, and from unmethylated p3AC and methylated p3AC-4. These two heteroduplex preparations were used to transform the various *E. coli* strains, and the function of the *supF* genes in the resulting colonies were analyzed. Table 3.4 presents the results of this analysis. Three types of colonies were found. Besides the parental blue and white colonies, some colonies were light blue in color. In some cases, these colonies contained sectors of dark blue. When cells from these colonies were re-streaked onto fresh plates, they gave rise almost entirely to light blue colonies, but, at a low frequency ($\sim 10^{-3}$), dark blue colonies were produced. The pure blue and the pure white colonies probably result from mismatch repair followed by replication. The light blue colonies probably come from cells in which replication of the heteroduplex plasmid DNA preceded mismatch repair, generating the two parental plasmids. However, this model has not been fully investigated, as the plasmid DNA from the light blue colonies has not yet been analyzed directly.

The results show that in the wild type, strain SY204, 78% of the heteroduplex molecules underwent repair before replication, and most of the repair events produced colonies that had the phenotype of the methylated strand in the heteroduplex. When heteroduplex from methylated p3AC and unmethylated p3AC-4 was used to transform SY204, almost all of the colonies were blue, the phenotype of the functional *supF* gene in p3AC. When the alternate heteroduplex preparation was used, most of the colonies were white, the phenotype of the mutant *supF* gene in p3AC-4. Similarly, when SY302 *recA56* was used as a host, repair was efficient, and the distribution of phenotypes favored the

phenotype of the methylated strand in the heteroduplex DNA. Repair was efficient with SY208 (*mutH3, mutU4)* and with SY209 (*mutH3, mutU4, recA56)*. However, there was no clear preference for either strand. These results indicate that mismatch repair occured in these cells, but it was carried out without respect to methylation patterns. The results with EG826 *ssb-1* suggest that mismatch repair of hemi-methylated heteroduplex in these cells is inefficient. In contrast, as shown in Table3.3, mismatch repair of fully methylated DNA by EG826 *ssb-1* was closer in efficiency to that in the wild type, SY204. The reason for this observed difference is not known. The mismatch repair activity that did occur on hemimethylated DNA in EG826 (*ssb-1)* cells showed no strand preference.

These results hint at some interesting and perhaps novel aspects of mismatch repair in *E. coli* that are worth further investigation. For the purpose of studying mismatch repair in mammalian cell extracts, however, SY209 proved useful as a host for a biological assay of mismatch repair.

**Activity in HeLa cell extracts.** Heteroduplex DNA, prepared from fully methylated p3AC-4 and p3AC-8 and containing two single base pair mismatches in the *supF* gene as described above, was added to cell-free extracts from Hela cells. The incubation of the DNA in the extracts was carried out either under the conditions of the complete reaction as described in Materials and Methods or with various experimental modifications. These modifications included the following: (1) no added ATP, creatine phosphate, or creatine kinase; (2) no added deoxynucleotides; (3) the addition of aphidicolin at 120 μM; (4) the addition of dideoxynucleotides as follows: ddGTP at 40μM, ddATP at 40μM, ddTTP at 80μM, and ddCTP at 20μM; (5) the addition of dideoxynucleotides at the above concentrations in the absence of added deoxynucleotides; (6) the addition of aphidicolin and of dideoxynucleotides in the above concentrations without added deoxynucleotides; (7) no incubation of the heteroduplex in the extract. The DNA was recovered from the extracts and used to transform SY209 to ampicillin resistance in the presence of X-gal and IPTG, and the

percentage of blue colonies was determined. The results of these experiments are presented in Table 3.5. Pre-incubation of the heteroduplex in the complete extract prior to transformation of SY209 resulted in 1.19% blue colonies, whereas direct transformation of SY209 with the untreated heteroduplex yielded only 0.17% blue colonies. This increase of 7-fold above background is statistically significant ($p < 0.001$), and it indicates that these extracts metabolize the heteroduplex DNA in some way. In the absence of exogenously added energy sources, the activity of the extracts above background is reduced by about 75%. A similar effect is seen in the absence of added nucleotides, resulting in a reduction of extract activity by a factor of five. In the presence of aphidicolin or dideoxynucleotides, the activity of the extracts is again significantly reduced, by about 85% and 75%, respectively. When dideoxynucleotides are added in the absence of added deoxynucleotides, there is even less detectable activity, and when aphidicolin is also added with dideoxynucleotides in the absence of added deoxynucleotides, there is little if any effect above background.

These results show that the process in the extracts being measured is energy dependent, is enhanced by the addition of deoxynucleotides, and is sensitive to aphidicolin and dideoxynucleotides, agents known to inhibit DNA polymerases (Dressler and Lieberman, 1983; Miller and Chinault, 1982; Smith and Okumoto, 1984; Waqar et al., 1984). Taken together, they suggest that at least one aspect of the activity in the extracts involves DNA polymerization. To further investigate this correlation, extract reactions were set up under the same set of conditions as above except that no unlabelled dCTP was added to any of the reactions but instead all reactions (including the "-deoxynucleotides" reaction) received [$\alpha$-$^{32}$P]dCTP at a concentration of 20 $\mu$M (specific activity: 3000 Curies per millimole). The proportion of radioactivity incorporated into TCA insoluble material was determined in each case, as one measure of DNA polymerization. This incorporation was compared with that found in the complete reaction, and in Table 3.5 the relative incorporation under each of the given conditions tested is expressed as a percentage of that in the complete

reaction. As can be seen from Table 3.5, there is little or no correlation between the effect of a given reaction modification on the mismatch repair assay as compared to the assay for dCTP incorporation, especially in the case of the addition of dideoxynucleotides alone.

DNA recovered from the extract reactions carried out in the presence of [$\alpha$-$^{32}$P]dCTP was analyzed by agarose gel electrophoresis and autoradiography. The results, shown in Fig. 3.6, were consistent with the measurements of TCA-insoluble counts and demonstrated that the only labelled high molecular weight DNA species were those corresponding to the input DNA. No complex, reduced mobility forms that might indicate ongoing DNA replication were present, and no discrete degradation products were noted. There is, however, some evidence for what may be ligase activity in the extracts. In addition to the input bands of linear and nicked duplex DNA, there is also a band of mobility greater than that of the nicked duplex that may represent form I DNA, generated by ligation of the nicked duplex.

**Recombination in Hela cell extracts.** An investigation of the activity of the HeLa cell extracts with regard to recombination was done as a control for the mismatch repair assay. These experiments are similar in design to those of Kucherlapati et al. (1985) and are similar in concept to those of Darby and Blattner (1984). Instead of forming heteroduplex molecules from p3AC-4 and p3AC-8, the two plasmids were both added directly to the extracts in either their circular or linear forms. The DNA recovered after incubation in the extracts was used to transform SY209, as above, and the percentage of blue colonies produced, indicative of the formation of a functional *supF* gene [ *supF*$^-$ (Su$^+$)], was determined in each case. The results of these experiments are presented in Table 3.6. No blue colonies were detected when circular plasmids were used to transform SY209, whether or not the mixed plasmids had been pre-incubated in the extracts. When both plasmids were first linearized, very few colonies were produced at all, and none were blue. Only when p3AC-4, present in its circular form, was mixed with the linear form of p3AC-8 were any

blue colonies detected, with 0.044% blue colonies after extract incubation, and about 8-fold fewer, or 0.0062% blue colonies, with no extract treatment. These results, like those reported previously (Kucherlapati et al., 1985; Darby and Blattner, 1984), suggest that there is some activity in the mammalian cell extracts which complements the *recA*- deficiency in SY209. Note that a much lower percentage of blue colonies is produced when the mixture of plasmids is used as a substrate as opposed to when heteroduplex prepared from the two mutant plasmids is used, and so the results with the heteroduplex DNA in the extracts cannot be attributed to recombinational events alone.

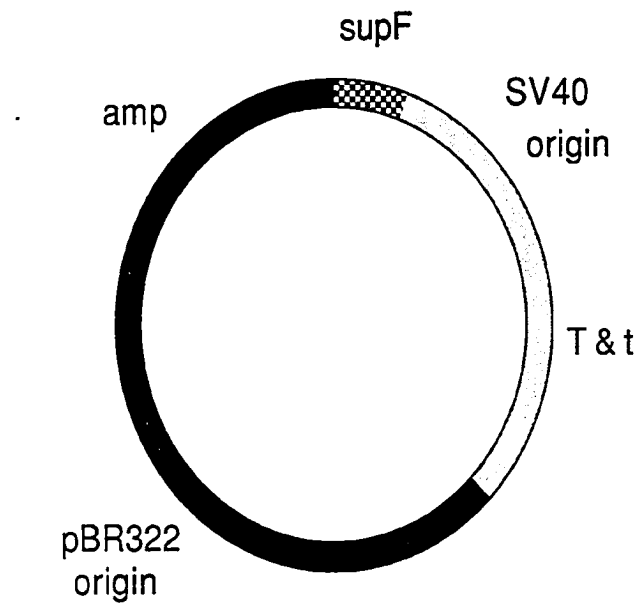


Fig. 3.1. Structure of the plasmid p3AC. The construction of this plasmid is described in Sarkar et al. (1984). It consists of the plasmid pBR322 into which was inserted the 200 base pair *EcoRI* fragment from the plasmid πVX which contains the *supF* gene. The plasmid was also modified by deletion of the *HaeII* fragment B and by insertion of the *BamHI* to *HpaII* early region of SV40 at the *ClaI* site.

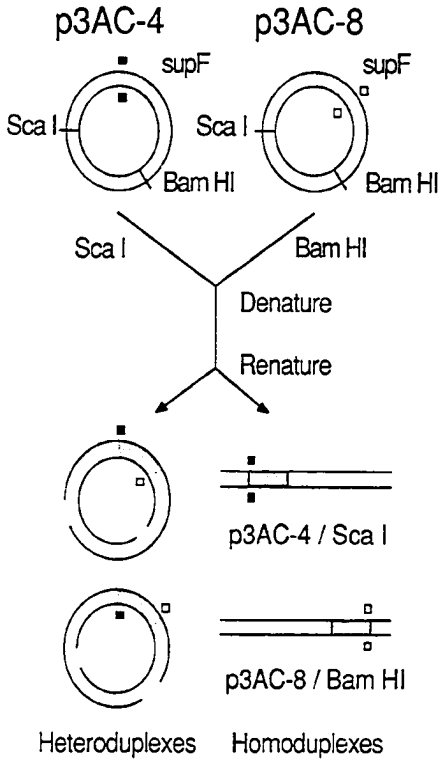


Fig. 3.2. Preparation of heteroduplex plasmid DNA. Plasmids p3AC-4 and p3AC-8, each bearing a single mutation in the *supF* gene, were linearized with *ScaI* and *BamHI*, respectively. The linearized plasmids were mixed, denatured and renatured. In this process, circular molecules are formed which represent heteroduplexes containing two single base pair mismatches in the *supF* gene.

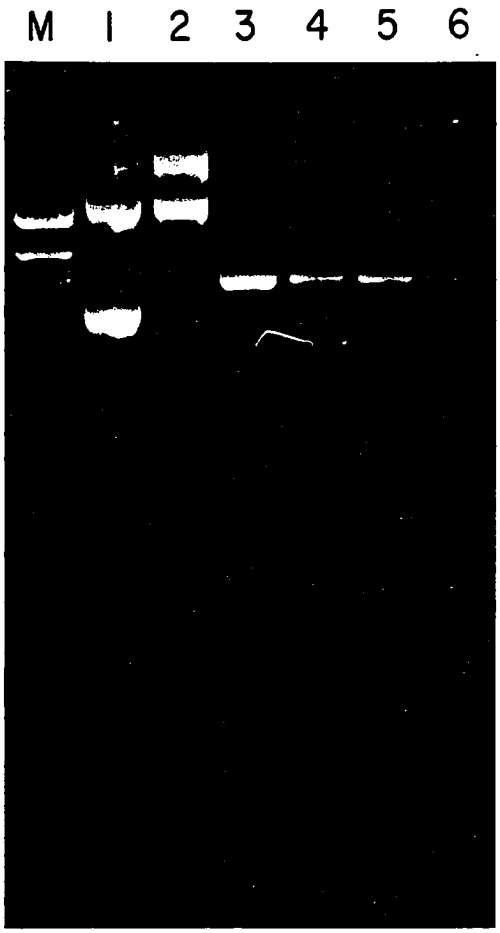


Fig. 3.3. Analysis by agarose gel electrophoresis of the preparation of heteroduplex plasmid DNA. (M) λ / *HindIII* size markers; (1) p3AC-4; (2) p3AC-8; (3) p3AC-4 / *ScaI*; (4) p3AC-8 / *BamHI*; (5) p3AC-8 / *BamHI* denatured and renatured by itself; (6) p3AC-4 / *ScaI* and p3AC-8 / *BamHI* mixed, denatured, and renatured, as illustrated in Fig 3.2.

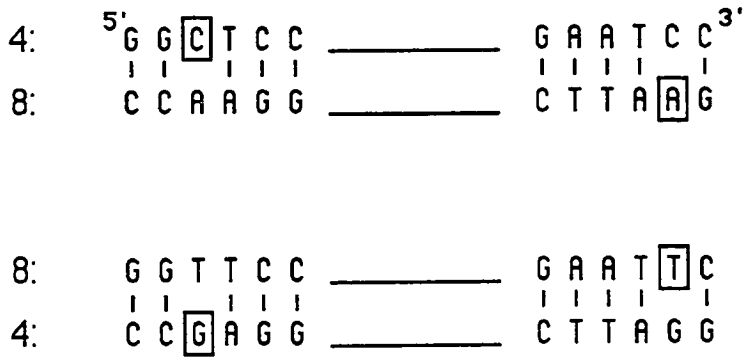


Fig. 3.4. Partial sequence of the heteroduplex plasmids made from p3AC-4 and p3AC-8 showing the base pair mismatches. The nucleotides indicated by the squares represent mutations from the wild type *supF* sequence. The two mismatches within each molecule are separated by 61 base pairs. The strands in the heteroduplexes originating from either p3AC-4 or p3AC-8 are indicated.

Table 3.2. The processing of heteroduplex plasmid DNA in monkey COS cells.

| DNA | Passaged through COS cells prior to transformation of *E. coli.* | *E. coli* colonies | | |
|---|---|---|---|---|
| | | No. blue | Total | % Blue |
| Heteroduplex DNA | yes | 62 | 1,263 | 4.9 |
| Heteroduplex DNA | no | 166 | 8,350 | 2.0 |
| p3AC-4 and p3AC-8 | yes | 0 | 1,800 | 0 |
| p3AC-4 and p3AC-8 | no | 0 | 50,000 | 0 |

Heteroduplex plasmid DNA was prepared from plasmids p3AC-4 and p3AC-8, each of which carry a single base pair mutation in the *supF* gene. The resulting heteroduplex DNA contained two single base pair mismatches in the *supF* gene, as illustrated in Fig. 3.4. It was used to tranfect monkey COS cells. Two days later, extrachromosomal DNA was prepared from the COS cells and was used to transform *E. coli* SY204 to ampicillin resistance in the presence of X-gal and IPTG. A blue colony is indicative of a functional, suppressor *supF* gene [*supF*$^-$ ($Su^+$)]. Alternatively, the heteroduplex DNA was used to transform *E. coli* SY204 directly without passage through the COS cells. For comparison, plasmids p3AC-4 and p3AC-8 were mixed in a 1:1 ratio and were used either to transfect COS cells followed by transformation of SY204 or to transform SY204 directly.

Table 3.3. The processing of heteroduplex plasmid DNA in *E. coli:* phenotypes of colonies formed by transformation of *E. coli* mutants with heteroduplex plasmid DNA.

| *E. coli* strain | Relevant genotype | No. blue | Total | % Blue |
|---|---|---|---|---|
| SY204 | wt | 166 | 8,350 | 2.00 |
| EG826 | *ssb-1* | 17 | 1,970 | 0.86 |
| SY302 | *recA56* | 24 | 2,661 | 0.90 |
| SY208 | *mutH3, mutU4* | 118 | 5,730 | 2.06 |
| SY209 | *mutH3, mutU4, recA56* | 37 | 22,235 | 0.17 |

Heteroduplex plasmid DNA was prepared from plasmids p3AC-4 and p3AC-8, each of which carry a single base pair mutation in the *supF* gene. The resulting heteroduplex DNA contained two single base pair mismatches in the *supF* gene, as illustrated in Fig. 3.4. It was used to transform *E. coli* to ampicillin resistance in the presence of X-gal and IPTG. The phenotypes of the colonies are given; a blue colony is indicative of a functional, suppressor*supF* gene [*supF*$^-$ (Su$^+$)].

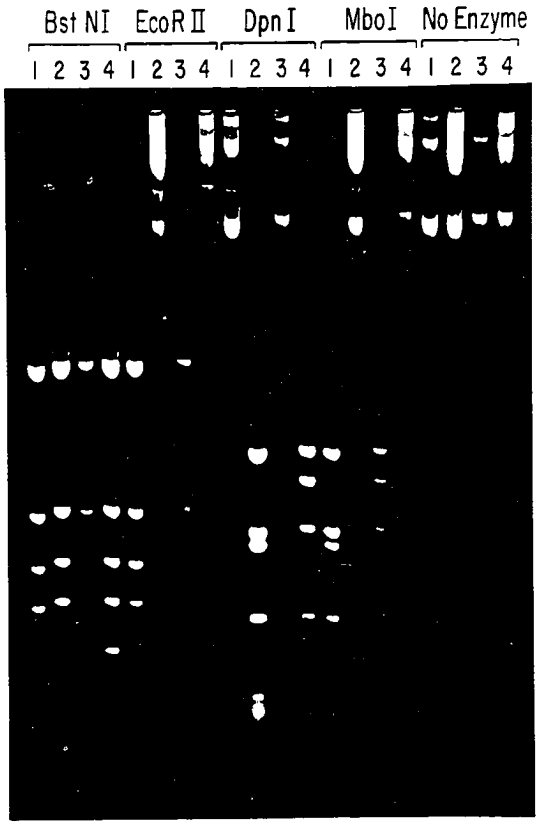


Fig. 3.5. Agarose gel analysis of methylation patterns of plasmid DNA used to make heteroduplex DNA. Plasmids p3AC and p3AC-4 were prepared in either SY204 or in GM272. Strain GM272 has mutations in the *dam*, *dcm* and *hsdS* genes and so is deficient in the methylation of DNA. SY204 is proficient in the methylation of DNA. The samples of DNA assayed by digestion with the indicated restriction enzymes were: (1) p3AC-4 prepared in GM272; (2) p3AC-4 prepared in SY204; (3) p3AC prepared in GM272; (4) p3AC prepared in SY204.

Table 3.4. The processing of hemi-methylated heteroduplex DNA in *E. coli:* phenotypes of colonies formed by transformation of *E. coli* mutants with hemi-methylated heteroduplex plasmid DNA.

| | | | Colonies | | | | |
|---|---|---|---|---|---|---|---|
| Strain | Partial genotype | Heteroduplex DNA[a] | No. blue[b] | No. light blue[b] | No. white[b] | % repair[c] | % blue[d] |
| SY204 | wt | 1 | 1146 | 319 | 5 | 78 | 99 |
| | | 2 | 301 | 140 | 940 | 90 | 24 |
| EG826 | *ssb-1* | 1 | 6 | 190 | 6 | 6 | 50 |
| | | 2 | 25 | 1726 | 28 | 3 | 47 |
| SY302 | *recA56* | 1 | 619 | 185 | 126 | 80 | 83 |
| | | 2 | 42 | 48 | 431 | 91 | 9 |
| SY208 | *mutH3, mutU4* | 1 | 334 | 84 | 321 | 89 | 51 |
| | | 2 | 218 | 55 | 77 | 84 | 74 |
| SY209 | *mutH3, mutU4, recA56* | 1 | 105 | 34 | 143 | 90 | 43 |
| | | 2 | 96 | 12 | 112 | 94 | 46 |

[a]Heteroduplex plasmid DNA was prepared from either (1) fully methylated p3AC and unmethylated p3AC-4 or (2) unmethylated p3AC and methylated p3AC-4 (Fig. 3.5). The plasmid p3AC carries a functional, suppressor *supF* gene [*supF*$^{+}$ (Su$^{+}$)], but in p3AC-4 the *supF* gene is not functional due to a single base pair mutation [*supF*$^{-}$ (Su$^{-}$)].

[b]A blue colony is indicative of a functional, suppressor *supF* gene [*supF*$^{+}$ (Su$^{+}$)], whereas a white colony indicates the absence of a functional gene [*supF*$^{-}$ (Su$^{-}$)]. The light blue colonies may consist of cells in which both functional and non-fuctional *supF* genes are present.

[c]The % repair was calculated by dividing the sum of the blue and white colonies by the total number of colonies.

[d]The % blue was calculated by dividing the number of blue colonies by the sum of the blue and white colonies.

Table 3.5. Mismatch repair and DNA polymerase activity in HeLa cell extracts

| Extract Reaction Conditions | Mismatch repair: Phenotype distribution of SY209 colonies transformed with heteroduplex DNA recovered from extract reactions | | | | DNA Polymerase |
|---|---|---|---|---|---|
| | No. blue | Total | % Blue | % of max mismatch repair[a] | % of maximum [α-$^{32}$P]dCTP incorporation[b] |
| Complete | 186 | 15,639 | 1.19 | 100 | 100 |
| -ATP, creatine kinase, & creatine phosphate | 13 | 3,124 | 0.42 | 25 | 3 |
| -deoxynucleotides | 11 | 3,158 | 0.35 | 18 | 63 |
| +aphidicolin | 16 | 4,805 | 0.33 | 16 | 2 |
| +dideoxynucleotides | 43 | 9,771 | 0.44 | 26 | 103 |
| +dideoxynucleotides -deoxynucleotides | 9 | 2,988 | 0.3 | 13 | NT |
| +dideoxynucleotides +aphidicolin -deoxynucleotides | 8 | 3,950 | 0.2 | 3 | NT |
| No extract | 37 | 22,235 | 0.17 | 0 | NT |

[a] Calculated by taking the percentage of blue colonies produced in each case, subtracting the percentage of blue colonies produced when no extract is used, and then normalizing to the value for the complete reaction.

[b] Calculated by subtracting the background counts and normalizing to the value for the complete reaction. 100% is equivalent to 24,034 cpm.

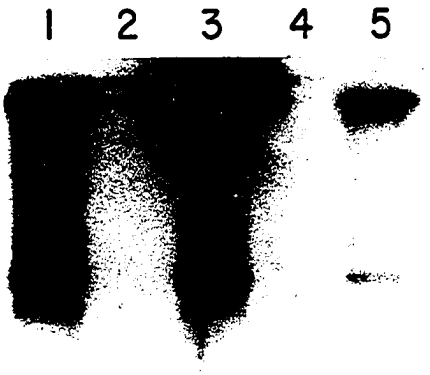


Fig. 3.6. Analysis of DNA polymerase activity in the HeLa cell extracts. DNA recovered from HeLa cell extract reactions carried out in the presence of [$\alpha^{32}$P] dCTP was analyzed by agarose gel electrophoresis and autoradiography. The samples analyzed were from extract reactions carried out under the following conditions: (1) the complete extract reaction; (2) no added ATP, creatine phosphate, or creatine kinase; (3) no added deoxynucleotides; (4) the addition of aphidicolin; (5) the addition of dideoxynucleotides. Details concerning the reaction conditions are given in the text.

Table 3.6 Recombination in HeLa cell extracts

| | | Phenotype distribution of SY209 colonies transformed with DNA recovered from extract reactions | | |
|---|---|---|---|---|
| DNA substrate[a] | Extract[b] | No. blue | Total | % blue |
| 4 & 8 | + | 0 | 39,800 | 0 |
| 4 & 8 | - | 0 | 50,000 | 0 |
| 4 & 8 / *BamHI* | + | 16 | 36,200 | 0.044 |
| 4 & 8 / *BamHI* | - | 4 | 64,000 | 0.006 |
| 4 / *ScaI* & 8 / *BamHI* | + | 0 | 0 | 0 |
| 4 / *ScaI* & 8 / *BamHI* | - | 0 | 50 | 0 |

[a]Abbreviations: 4 (p3AC-4) and 8 (p3AC-8).

[b]Either the complete extract and reaction conditions were used or the samples were incubated in buffer without HeLa cell extract.

## Discussion

We have set up a biological assay in which repair of heteroduplex plasmid DNA by human cell extracts can be subsequently detected by screening bacteria which have been transformed with the plasmid DNA recovered from the extracts. The success of this assay depended on the construction of an *E. coli* strain in which mismatch repair and post-replicative plasmid recombination were minimal so that the background in the assay was low enough to measure the effect of the human cell extracts on the heteroduplex plasmid DNA. The HeLa cell extracts used in these experiments were essentially similar to the *in vitro* transcription extracts first described by Manley et al. (1980), and they have been used in our laboratory for the study of transcription as well as mismatch repair (Read and Summers, 1982).

In the assay, the percentage of blue colonies produced by the transformation of SY209 with the DNA recovered from the extracts is taken as a measure of mismatch repair. The production of blue colonies depends on the conversion of one or both of the single base pair mismatches in the *supF* gene to the normal *supF* base pair. A single correction will give a blue colony if it is followed by plasmid replication, since in this case semi-conservative replication will generate a plasmid with a parental *supF* gene [ $supF^{-}$ ($Su^{+}$) ]in addition to one with a mutant gene [ $supF^{-}$ ($Su^{-}$)]. Some events, however, may convert the mispairs to the mutant base pairs, but these events, since they would yield white colonies, are not specifically counted in this assay. They appear among the many white colonies that arise from the introduction of unrepaired heteroduplex into the SY209 cells. Hence, the assay underestimates the frequency of repair events. It might be expected that conversion of the mispair to either the wild type or the mutant sequence is equally likely, as there were no particular differences between the strands of the heteroduplexes in these experiments, and so the actual frequency of repair events might be at least twice the observed value.

The results demonstrate that the HeLa cell extracts have a significant effect on the heteroduplex DNA. When the heteroduplex DNA is used to transform SY209 directly, only 0.17% of the colonies are blue. In contrast, incubation of the heteroduplex plasmid DNA in the extracts prior to transformation of SY209 produces 1.19% blue colonies, a 7-fold increase. Part of the assay, however, involves propagation of the DNA in *E.coli* before the final results can be determined, and so it is not known if all the steps involved in the repair of the mismatches are carried out and completed in the extracts. Certainly a significant part of the metabolism of the mismatched bases must occur in the extracts, since the results are dependent on the addition to the extracts of ATP, creatine phosphate, and creatine kinase, and of deoxynucleotides. The marked reduction in the activity of the extracts when these species are omitted indicates that some sort of energy-requiring enzymological process is involved which needs a sufficient supply of DNA precursors.

The exact nature of the activity in the extracts is not clear. Theoretically, some type of random nick translation activity might be invoked to account for the repair of mismatches. Evaluation of the data, however, suggests that this explanation cannot account for all of the observed repair activity. The results show that the mismatch repair activity in the extracts is significantly reduced by the addition both of aphidicolin and of dideoxynucleotides. Only aphidicolin, however, had an effect on the incorporation of labelled dCTP into the plasmid DNA. In fact, aphidicolin completely blocked [$\alpha$-$^{32}$P]dCTP incorporation, whereas dideoxynucleotides had no measurable effect at all. The incorporation of [$\alpha$-$^{32}$P]dCTP in these extracts is one measure of DNA polymerase activity. It is also an index of random nick translation activity, since nick translation is a process that is associated with the incorporation of labelled nucleotides into DNA. Semi-conservative DNA replication would also contribute to the results, but analysis of the reactions by gel electrophoresis shows that the pattern of complex, slowly migrating forms indicative of DNA replication is absent. Studies of *in vitro* replication in animal cell extracts have shown also that replication of

plasmids, such as p3AC, which contain SV40 origins of replication, is dependent upon exogenously added T antigen protein, which was not added in these experiments. It is reasonable to assume, therefore, that the assay used to measure DNA polymerization in these experiments is essentially an assay for nick translation. Repair synthesis specifically associated with mismatch repair may also account for some [$\alpha$-$^{32}$P]dCTP incorporation, but to the extent that it does contribute, the trivial explanation of random nick translation cannot be invoked. The data show that the presence of dideoxynucleotides in the extract reactions had no effect on the measured nick translation activity but still reduced the apparent mismatch repair activity by about 75%. This suggests that some part of the observed mismatch repair activity which is sensitive to inhibition by dideoxynucleotides is unrelated to random nick translation. Similarly, in the absence of exogenously added deoxynucleotides, nick translation in the extracts is reduced to 63% of normal, while the apparent mismatch repair activity is just 18% of normal, again demonstrating an incomplete correlation between these activities in the extracts. Thus, although random nick translation cannot be ruled out as a partial explanation for the observed repair of mismatches, it does not appear to be associated with all of the mismatch repair activity in the extracts.

In order to produce functional , suppressor *supF* genes [ *supF*$^-$ ($Su^+$)] in this assay, there must be some independent repair of each of the two mismatches even though they are only 61 base pairs apart. Co-repair of the two mismatches would simply generate the parental mutant plasmids. Whatever the mechanism of processing of the heteroduplex in these extracts, it appears that at least one pathway for heteroduplex processing allows for independent repair of the closely spaced mismatches. If this pathway involves excision and resynthesis (a reasonable assumption, since it is associated with DNA polymerization), then the patch size for this repair pathway is likely to be on the order of 60 base pairs.

There is some precedent for this short patch size in certain mismatch repair reactions in *E. coli* and in *Saccharomyces cerevisiae*. Although the patch size for the *dam* instructed

mismatch repair process in *E. coli* is several thousand nucleotides long (Wagner and Meselson, 1976), as is the patch size for the more efficient of the two known methyl-independent pathways in *E. coli* (Fishel et al.,1986), the patch size for the other methyl independent mismatch repair pathway in *E. coli* is less than 300 base pairs (Fishel and Kolodner, 1983). Experiments with extracts from *Saccharomyces cerevisiae* have detected one mismatch repair pathway with a patch size of only about 15 base pairs (Muster-Nassal and Kolodner, 1986).

Considering the variety of mismatch repair pathways in *E. coli*, it is possible that there are several such pathways in mammalian cells as well. In the experiments described here we may not be detecting all of the possible mismatch repair reactions in mammalian cells. Other mismatch repair pathways may exist; for example, it is possible that there are activities in these extracts or in the HeLa cells which show some strand selectivity in mismatch repair. Because the heteroduplex DNA used as a substrate in these reactions was fully methylated by the *E. coli* K modification system, our experiments thus far were not set up to detect specifically such activity. It could be that the activity we have observed may show strand selectivity when challenged with the appropriate substrate, or it may be that some other as yet undetected pathways may be the ones capable of strand discrimination.

Nonetheless, strand discrimination is an important aspect of mismatch repair, as it is understood in *E. coli*, so that repair can be directed toward the parental sequence when errors occur during replication. The factors which affect strand selectivity in mammalian cells are not known for sure, but the work of Hare and Taylor (1985) suggests that when the substrate DNA is hemi-methylated, there is an apparent bias toward the methylated strand in mismatch correction *in vivo* in monkey cells. We are currently investigating the effect of methylation on strand selection *in vitro* using several approaches. Because these experiments are just underway and have not yielded conclusive results, they have not been discussed thus far. They are mentioned below to indicate the nature of future work related to this thesis.

In order to detect strand discrimination, we sought to develop an assay based on a mismatched restiction enzyme site. We have constructed M13 phage carrying inserts that differ by a single base pair which creates or destroys a restriction site in the double stranded DNA. The inserts were cloned into the M13 vectors in both orientations. Single stranded phage DNA can be made from the vectors, and the DNA from one vector can be annealed to that of another vector carrying an insert in the opposite orientation. The M13 viral strands will not anneal, but the inserts will, generating duplex DNA with the desired mismatch. Using restriction enzymes or single-stranded DNA nucleases, the heteroduplex DNA can be isolated. Since M13 DNA is normally methylated, growth of one vector in a wild type host and one in a *dam*$^-$, *dcm* $^-$host allows production of reasonably pure hemi-methylated heteroduplex for use in mismatch repair assays.

Another approach is to use plasmids with single base mutations at restriction sites. Plasmid DNA can be prepared as described above in wild type and in *dam*,$^-$ *dcm* $^-$hosts. Alternately, if the plasmid contains part of a mammalian virus such as SV40, it is possible to prepare some of the DNA in mammalian cells. Experiments can be designed to determine if mammalian cells or cell extracts can discriminate between DNA prepared in *E. coli* and DNA prepared in mammalian cells.

Several modifications may help to improve upon the system desctribed here to detect mismatch repair in human cell extracts. The efficiency of the process in the extracts is low. This makes it difficult to perform extensive biochemical studies. Although the activity in the extracts does generate results that are reproducibly above the background, they are close enough to the background so that perturbations of the system yield results that are barely different from the background in the assay. This same problem must also be overcome in order to extend this approach to the study of mutant mammalian cell lines that may be abnormal relative to mismatch repair. One approach is to try a nuclear extract rather than a whole cell extract, since DNA repair enzymes may be more concentrated and more active in a

nuclear extract. Secondly, heteroduplex substrates in which the mismatches are further apart than 61 base pairs might be tried. Although the repair patch size in the extracts seems to be small, it may be larger, on average, than 61 base pairs, so that independent repair of these closely spaced mismatches is an infrequent event. Separate repair of mismatches 500 to 1000 base pairs away might be more frequent. Mutations in the *lacI* gene would be suitable for this purpose. Thirdly, in order to reduce the background in the assay, a different *E. coli* strain carrying a mutation in the *recF* gene might prove useful. Fishel and Kolodner (1983) demonstrated that this gene is important both in short patch mismatch repair and in plasmid recombination, the two activities that probably contribute to the background in the assay. Although SY209 *(mutH3, mutU4,* and *recA56)* is suitable for this assay, a *recF* ⁻strain may be better.

The experiments reported here investigated not only mismatch repair in human cell extracts, but also examined some aspects of mismatch repair in *E. coli.* The results presented in Table 3.4 are consistent with the model in which the *mutH* and *mutU* genes are involved in methyl-directed mismatch repair, whereas a mutation in the *recA* gene does not affect methyl-directed mismatch repair. The results in Table 3.3, however, suggest that the recA protein may be involved in methyl-independent mismatch repair in some manner, perhaps as a regulatory element, as shown by the differences between SY208 and SY209 and between SY204 and SY302 in the assay. The *ssb-1* mutation may be involved in both methyl-dependent and methyl-independent mismatch repair, in addition to recombination, because it affects the processing of fully methylated heteroduplex with two mismatches (Table 3.3) and of hemi-methylated heteroduplex with a single mismatch (Table 3.4), as seen in comparisons of SY204 and EG826. These results confirm the observations of Fishel and co-workers (1983; 1986), and they extend the genetic analysis of the various *E. coli* mismatch repair pathways to include an examination of the effect of the *ssb-1* mutation

## Chapter 4. Summary.

The work of this thesis represents an application of molecular genetics to the study of mutagenesis and DNA repair in mammalian cells. Investigation of DNA damage and repair in mammalian cells has been hampered by the difficulty of genetic manipulation of mammalian cells in culture. To circumvent this problem, methods were developed to study DNA metabolism in mammalian cells that would use the tools of molecular biology and exploit the power of bacterial genetic analysis.

Two different aspects of DNA damage and repair in mammalian cells were investigated. Using a lambda phage shuttle vector, a system was developed to study mutations arising in the DNA of mammalian cells. This system was used to study the spectrum of mutations induced in cellular DNA by ultraviolet light. Secondly, the repair of base pair mismatches in DNA was studied by the development of a method to detect a DNA mismatch repair activity in extracts made from cultured human cells.

In order to study mutations arising in mammalian cells, stable mouse L cell lines were established with multiple copies of a lambda phage vector which contains the *supF* gene of *E. coli* as a target for mutagenesis. Rescue of viable phage from high molecular weight mouse cell DNA using lambda *in vitro* packaging extracts was efficient (5 viable phage per μg of cell DNA per lambda genome copy) and yielded a negligible background of phage with mutations in the *supF* gene (zero out of 54,605). From mouse cells exposed to 12 $J/m^2$ of 254 nm ultraviolet (UV) light, 78,510 phage were rescued of which eight were found to have mutant *supF* genes. DNA sequence analysis of the mutants suggests that the primary site of UV mutagenesis in mammalian cells is at pyrimidine-cytosine sequences and not at thymidine-thymidine sequences, and that the most frequent mutation at the pyrimidine-cytosine sequence is a C to T transition.

The application of this system to the study of mutagenesis in transgenic mice was

begun. Once a suitable transgenic mouse strain is isolated, many experiments that are beyond the scope of cell culture systems can be tried. Tissue-related and age-related differences in mutagenesis can be studied. Germ line and somatic tissue can be compared. In the whole animal, physiologically relevant estimates of mutation rates can be obtained, and transplacental and intrauterine mutagenesis can be examined.

Extension of the cell culture system to use the thymidine kinase gene of herpes simplex virus as a target gene was attempted so that cells bearing mutations in the target gene can be selected in cell culture. Such a system requires a cell line with a single, stable copy of the λTK vector in form that can be rescued by the lambda *in vitro* packaging system. Of the 62 cell lines examined, however, none met all these requirements.

The repair of base pair mismatches in DNA was also studied. Because of the difficulties in interpreting the results of the *in vivo* mismatch repair experiments done in monkey COS cells, a system to study DNA mismatch repair *in vitro* in HeLa cell extracts was developed. Pre-formed heteroduplex plasmid DNA containing two single base pair mismatches within the *supF* gene of *E. coli* was used as a substrate in a mismatch repair assay. Repair of one or both of the mismatches to the wild type sequence was measured by transformation of a *lac* $^-$(Amber) *E. coli* strain in which the presence of an active suppressor gene could be scored. The *E. coli* strain used was constructed to carry mutations in genes associated with mismatch repair and recombination *(mutH*, *mutU*, and *recA)* so that processing of the heteroduplex DNA by the bacteria was minimal. Extract reactions were carried out by the incubation of the heteroduplex plasmid DNA in the HeLa cell extracts to which ATP, creatine phosphate, creatine kinase, deoxynucleotides, and a magnesium containing buffer had been added. Under these conditions, about 1% of the mismatches were repaired. In the absence of added energy sources or of added deoxynucleotides, the activity in the extracts was significantly reduced. The addition of either aphidicolin or dideoxynucleotides reduced the mismatch repair activity, but only aphidicolin was effective

in blocking DNA polymerization in the extracts. It is concluded that mismatch repair in these extracts is an energy-requiring process that is dependent on adequate deoxynucleotide concentrations. The results also indicate that the process is associated with some type of DNA polymerization, but the different effects of aphidicolin and dideoxynucleotides suggest that the mismatch repair activity cannot be accounted for simply by random nick translation activity in the extracts. Further studies are underway to characterize this activity, in particular to determine the effect of DNA methylation on strand selection in the repair of mismatches.

References.

Ames, B. N., F. D. Lee, and W. E. Durston. 1973. An improved bacterial test system for the detection and classification of mutagens and carcinogens. Proc. Natl. Acad. Sci. USA **70**:782-786.

Ashman, C. R. and R. L. Davidson. 1984. High spontaneous mutation frequency in shuttle vector sequences recovered from mammalian cell DNA. Mol. Cell. Biol. 4:2266-2272.

Ashman, C. R., P. Jagadeeswaran, and R. L. Davidson. 1986. Efficient recovery and sequencing of mutant genes from mammalian chromosomal DNA. Proc. Natl. Acad. Sci. USA **83**:3356-3360.

Beukers, R. and W. Berends. 1960. Isolation and identification of the radiation product of thymine. Biochim. Biophys. Acta. **41**:550-551.

Biggin, M. D., T. J. Gibson, and G. F. Hong. 1983. Buffer gradient gels and $^{35}S$ label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA **80**:3963-3965.

Bourre, F. and A. Sarasin. 1982. Targeted mutagenesis of SV40 DNA induced by UV light. Nature (London) **305**:68-70.

Brash, D. E. and W. A. Haseltine. 1982. UV-induced mutation hotspots occur at DNA damage hotspots. Nature (London) **298**: 189-192.

Bredberg, A., K. H. Kraemer, and M. M. Seidman. 1986. Restricted ultraviolet mutational spectrum in a shuttle vector propagated in xeroderma pigmentosum cells. Proc. Natl. Acad. Sci. USA **83**:8273-8277.

Brinster, R. L. , H. Y. Chen, M. Trumbauer, A. W. Senear, R. Warren, and R. D. Palmiter. Somatic expression of herpes thymidine kinase in mice following injection of a fusion gene into eggs. Cell **27**:223-231.

Brown, E. L., R. Belagaje, M. J. Ryan, and H. G. Khorana. 1979. Chemical synthesis and cloning of a tyrosine tRNA gene. Methods Enzymol. **68**:109-151.

Brouwer, J., P. van de Putte, A. M. J. Fichtinger-Schepman, and J. Reedijk. 1981. Base pair substitution hot spots in GAG and GCG nucleotide sequences in Escerischia coli K12 induced by cis-diamminedichloroplatinum (II). Proc. Natl. Acad. Sci. U.S.A. **708**:7010-7014.

Calos, M. P., J. S. Lebkowski, and M. R. Botchan. 1983. High mutation frequency in DNA transfected into mammalian cells. Proc. Natl. Acad. Sci. U.S.A. **80**: 3015-3019.

Celis, J. E. and P. W. Piper.1982. Compilation of mutant suppressor tRNA sequences. Nucleic Acids Res. **10**:R83-R91.

Cepko, C. L., B. E. Roberts, and R. Mulligan. 1984. Construction and applications of a highly transmissible retrovirus shuttle vector. Cell **37**:1053-1062.

Challberg. M. D. and T. J. Kelly. 1979. Initiation of adenovirus DNA replication *in vitro*. Proc. Natl. Acad. Sci. U.S.A. **76**:655-659.

Christy, B. A. and G. A. Scangos. 1982. Expression of transferred thymidine kinase gene is controlled by methylation. Proc. Natl. Acad. Sci. U.S.A. **79**:6299-6303.

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

Chu, G. and P. A. Sharp. 1981. SV40 DNA transfection of cells in suspension: analysis of the efficiency of transcription and translation of T-antigen. Gene **13**:197-202.

Constantini, F. and E. Lacy. 1981. Introduction of the rabbit ß-globin gene into the mouse germ line. Nature (London) **294**:92-94.

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

Conrad, S. E, C. P. Liu, and M. R. Botchan. 1982. A fragment spanning the SV40

replication origin is the only DNA sequence required in cis for viral excision. Science **218**:1223-1225.

Darby, V., and F. Blattner. 1984. Homologous recombination catalyzed by mammalian cell extracts *in vitro*. Science **226**:1213-1215.

Dasgupta, U. B. and W. C. Summers. 1978. UV reactivation of herpes simplex virus is mutagenic and inducible in mammalian cells. Proc. Natl. Acad. Sci. USA **75**:2378-2381.

Drake, J. W.and R. H. Baltz. 1976. The biochemistry of mutagenesis. Ann. Rev. Biochem. **45**:11-39.

Dressler, S. L., and M. W. Lieberman. 1983. Identification of DNA polymerases involved in DNA excision repair in diploid human fibroblasts. J. Biol. Chem. **258**:9990-9994.

Drinkwater, N. R. and D. K. Klinedinst. 1986. Chemically induced mutagenesis in a shuttle vector with a low-background mutant frequency. Proc. Natl. Acad. Sci. USA **83**:3402-3406.

Elledge, S. J. and G. C. Walker. 1983. Proteins required for ultraviolet light and chemical mutagenesis: identification of the products of the *umuC* locus of *Escherichia coli*. J. Mol. Biol. **64**:175-192.

Eisenstadt, E., A. J. Warren, J. Porter, D. Atkins, and J. H. Miller. 1982. Carcinogenic epoxides of benzo(a)pyrene and cyclopenta(cd)pyrene induce base substitutions via specific transversions. Proc. Natl. Acad. Sci. U.S.A. **79**:1945-1949.

Enquist, L. W., G. F. Vande Woude, M. Wagner, J. R. Smiley, and W. C. Summers. 1979. Construction and characterization of a recombinant plasmid encoding the gene for the thymidine kinase of Herpes simplex type I virus. Gene **7**:335-342.

Fishel, R.A. and R. Kolodner. 1983. Gene conversion in *Escherichia coli:* The identification of two repair pathways for mismatched nucleotides. UCLA Symp. Mol. Cell. Biol. **11**:309-326.

Fishel, R. A., E. C. Siegel, and R. Kolodner. 1986. Gene coversion in *Escherichia coli:* resolution of heteroallelic mismatched nucleotides by co-repair. J. Mol. Biol. **188**:147-157.

Folger, K. R., K. Thomas, and M. R. Capecchi. 1985. Efficient correction of mismatched bases in plasmid heteroduplexes injected into cultured mammalian cell nuclei. Mol. Cell. Biol. **5**:70-74.

Foster, P. L., E. Eisenstadt, and J. H. Miller. 1983. Base substitution mutations induced by metabolically activated aflatoxin B1. Proc. Natl. Acad. Sci. U.S.A. **80**:2695-2698.

Fuscoe, J. C., R. G. Fenwick, D. H. Ledbetter, and C.T. Caskey. 1983. Deletion and amplification of the HGPRT locus in Chinese hamster cells. Mol. Cell. Biol. **3**:1086-1096.

Glickman, B. W., and M. Radman. 1980. *Escherichia coli* mutator mutants deficient in methylation-instructed DNA mismatch correction. Proc. Natl. Acad. Sci. U.S.A. **77**:1063-1067.

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

Graf, L. H. and L.A. Chasin. 1982. Direct demonstration of genetic alterations at the dihydrofolate reductase locus from cultured mammalian cells. Mol. Cell. Biol. **2**:93-96.

Graham, F. L. and A. J. van der Eb. 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology **52**:456-467.

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

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

Hare, J. T., and J. H. Taylor. 1985. One role for DNA methylation in vertebrate cells is strand discrimination in mismatch repair. Proc. Natl. Acad. Sci. USA **82**:7350-7354.

Hauser, J., M. M. Seidman, K. Sidur, and K. Dixon. 1986. Sequence specificity of point mutations induced during passage of a UV-irradiated shuttle vector plasmid in monkey cells. Mol. Cell. Biol. **6**:277-285.

Herman, G. E., and P. Modrich. 1981. *Escherichia coli* K-12 clones that overproduce dam methylase are hypermutable. J. Bacteriol. **145**:644-646.

Hirt, B. 1967. Selective extraction of polyoma DNA from infected mouse cultures. J. Mol. Biol. **26**:365-369.

Hohn, B. 1979. *In vitro* packaging of λ and cosmid DNA. Methods Enzymol. **68**:299-309.

Hohn, B. 1983. DNA sequences necessary for packaging of bacteriophage λ DNA. Proc. Natl. Acad. Sci. USA **80**:7456-7460.

Howard, B. D. and I. Tessman. 1964. Identification of the altered bases in mutated single-stranded DNA. II. *In vivo* mutagenesis by 5-bromodeoxyuridine and 2-aminopurine. J. Mol. Biol. **30**:364-371.

Kato, T. and Y. Shinoura. 1977. Isolation and characterization of *Escherichia coli* mutants deficient in induction of mutations by ultraviolet light. Molec. Gen. Genet. **156**:121-131.

Kit, S., Dubbs, D. R., Piekarski, L. J. & Hsu, T. C. 1963. Deletion of thymidine kinase activity from L-cells resistant to bromodeoxyuridine. Exp. Cell Res. **31**: 297-312.

Kucherlapati, R. S., J. Spencer, and P. Moore. 1985. Homologous recombination catalyzed by human cell extracts. Mol. Cell. Biol. **5**:714-720.

Lacy, E., S. Roberts, E. P. Evans, M. D. Burtenshaw, and F. D. Constantini. 1983. A foreign ß-globin gene in transgenic mice: integration at abnormal chromosomal positions and expression in inappropriate tissues. Cell **34**:343-358.

Lau, Y. and Y. W. Kan. 1983. Versatile cosmid vectors for the isolation, expression, and

rescue of gene sequences: Studies with the human α-globin gene cluster. Proc. Natl. Acad. Sci. USA **80**:5225-5229.

Lau, Y. and Y. W. Kan. 1984. Direct isolation of the fucnctional human thymidine kinase gene with a cosmid shuttle vector. Proc. Natl. Acad. Sci. USA **81**:414-418.

Law, M., J. C. Byrne, and P. M. Howley. 1983. A stable bovine papillomavirus hybrid that expresses a dominant selective trait. Mol. Cell. Biol. **3**:2110-2115.

Lebkowski, J. S., R. B. DuBridge, E. A. Antell, K. S. Greisen and M. P. Calos. 1984. Transfected DNA is mutated in monkey, mouse and human cells. Mol. Cell. Biol. **4**: 1951-1960.

Lebkowski, J. S., S. Clancy, J. H. Miller, and M. P. Calos. 1985. The *lacI* shuttle: Rapid analysis of the mutagenic specificity of ultraviolet light in human cells. Proc. Natl. Acad. Sci. USA **82**:8606-8610.

Lebkowski, J. S., J. H. Miller, and M. P. Calos. 1986. Determination of DNA sequence changes induced by ethyl methanesulfonate in human cells, using a shuttle vector system. Mol. Cell. Biol. **6**: 1838-1842.

Leclerc, J. E. and N. L. Istock. 1982. Specificity of UV mutagenesis of the *lac* promoter of M13*lac* hybrid phage DNA. Nature (London) **297**: 596-598.

Levinson, A., D. Silver, and B. Seed. 1984. Minimal size plasmids containing an M13 origin for production of single-strand transducing particles. J. Mol. Appl. Genet. **2**:507-517.

Li, J. J., and T. J. Kelly. 1984. Simian virus 40 DNA replication *in vitro*. Proc. Natl. Acad. Sci. USA **81**:6973-6977.

Lindahl, T. 1982. DNA repair enzymes. Ann. Rev. Biochem. **51**:61-87.

Lindenmaier, W., H. Hauser, I. G. de Wilke, and G. Schutz. 1982. Gene shuttling: moving of cloned DNA into and out of eukaryotic cells. Nucleic Acids Res. **10**:1243-1256.

Loenen, W. A. and W. J. Brammar. 1980. A bacteriophage lambda vector for cloning large

DNA fragments made with several restriction enzymes. Gene **10**:249-254.

Lu, A. L., S. Clark, and P. Modrich. 1983. Methyl-directed repair of DNA base-pair mismatches *in vitro*. Proc. Natl. Acad. Sci. USA **80**:4639-4643.

Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular Cloning. Cold Spring Harbor Laboratory, New York.

Manley, J. L., A. Fire, A. Cano, P. A. Sharp, and M. L. Gefter. 1980. DNA-dependent transcription of adenovirus genes in a soluble whole cell extract. Proc. Natl. Acad. Sci. USA **77**:3855-3859.

Marinus, M. G., A. Poteete, and J. A. Arraj. 1984. Correlation of DNA adenine methylase activity with spontaneous mutability in *Escherichia coli* K-12. Gene **28**:123-125.

Meuth, M. and J. E. Arrand. 1982. Alterations of gene structure in ethyl methanesulfonate induced mutants of mammalian cells. Mol. Cell. Biol. **2**:1459-1462.

Messing, J. 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20-79.

Meyer, R. R., J. Glassberg, and A. Kornberg. 1979. An *Escherichia coli* mutant defective in single-stranded binding protein is defective in DNA replication. Proc. Natl. Acad. Sci. USA **76**:1702-1705.

Miller, J. H. 1983. Mutational specificity in bacteria. Ann. Rev. Genet. **17**:215-238.

Miller, J. H. 1985. Mutagenic specificity of ultraviolet light. J. Mol. Biol. **182**:45-68.

Miller, J. H. and K. B. Low. 1984. Specificity of mutagenesis resulting from the induction of the SOS system in the absence of mutagenic treatment. Cell. **37**:675-682

Miller, M. R., and D. N. Chinault. 1982. The roles of DNA polymerases $\alpha$, $\beta$, and $\gamma$ in DNA repair synthesis induced in hamster and human cells by different DNA damaging agents. J. Biol. Chem. **257**:10204-10209.

Muller, H. J. 1927. Artificial transmutation of the gene. Science. **66**: 84-87.

Muster-Nassal, C. and R. Kolodner. 1986. Mismatch correction catalyzed by cell-free extracts of *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **83**:7618-7622.

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

Panasenko, S. M., J. R. Cameron, R. W. Davis, and I. R. Lehman. 1977. Five hundred fold over-production of DNA ligase after induction of a hybrid lambda lysogen constructed *in vitro*. Science **196**:188-189.

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

Protic-Sabljic, M., N. Tuteja, P. Munson, J. Hauser, K. H. Kreamer, and K. Dixon. 1986. UV light-induced cyclobutane pyrimidine dimers are mutagenic in mammalian cells. Mol. Cell. Biol. **6**:3349-3356.

Pukkila, P. J., J. Peterson, G. Herman, P. Modrich, and M. Meselson. 1983. Effects of high levels of DNA adenine methylation on methyl-directed mismatch repair in Escherichia coli. Genetics **104**:571-582.

Rabkin, S. D., P. D. Moore, and B. S. Strauss. 1983. *In vitro* bypass of UV-induced lesions by *Escherichia coli* DNA polymerase I: specificity of nucleotide incorporation. Proc. Natl. Acad. Sci. USA **80**:1541-1545.

Razzaque, A., H. Mizusawa, and M. M. Seidman. 1983. Rearrangement and mutagenesis of a shuttle vector plasmid after passage in mammalian cells. Proc. Natl. Acad. Sci. USA 80:3010-3014.

Razzaque, A., S. Chakrabarti, S. Joffee, and M. M. Seidman. 1984. Mutagenesis of a shuttle vector plasmid in mammalian cells. Mol. Cell. Biol. **4**, 435-441.

Read, G. S., and W. C. Summers. 1982. *In vitro* transcription of the thymidine kinase gene of herpes simplex virus. Proc. Natl. Acad. Sci. USA **79**:5215-5219.

Rich, A. and U. L. RajBhandary. 1976. Transfer RNA: molecular structure, sequence, and properties. Ann. Rev. Biochem. **45**:805-860.

Rosenberg, S. M. 1985. *EcoK* restriction during *in vitro* packaging of coliphage lambda DNA. Gene **39**:313-315.

Rosenberg, S. M., M. M. Stahl, I. Kobayashi, and F. W. Stahl. 1985. Improved *in vitro* packaging of coliphage lambda DNA: a one strain system free from endogenous phage. Gene **38**:165-175.

Sanger, E., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain termination inhibitors. Proc. Natl. Acad. Sci. U.S.A. **74**:5463-5467.

Sarasin, A. R. and P. A. Hanawalt. 1978. Carcinogens enhance survival of UV-irradiated SV40 in treated monkey kidney cells: induction of a recovery pathway? Proc. Natl. Acad. Sci. USA **75**:346-350.

Sarkar, S., U. B. Dasgupta, and W. C. Summers. 1984. Error prone mutagenesis detected in mammalian cells by a shuttle vector containing the *supF* gene of *Escherichia coli*. Mol. Cell. Biol. **4**:2227-2230.

Schaaper, R. M., T. A. Kunkel, and L. A. Loeb. 1983. Infidelity of DNA synthesis associated with bypass of apurinic sites. Proc. Natl. Acad. Sci. U.S.A. **80**:487-491.

Schaaper, R. M. and L. A. Loeb. 1981. Depurination causes mutations in SOS-induced cells. Proc. Natl. Acad. Sci. U.S.A. **78**:1773-1777.

Seidman, M. M., K. Dixon, A. Razzaque, R. Zagursky, and M. L. Berman. 1985. A shuttle vector plasmid for studying carcinogen-induced point mutations in mammalian cells. Gene **38**:233-237.

Singer, B. and J. T. Kusmierek. 1982. Chemical mutagenesis. Ann Rev. Biochem. **52**:655-693.

Skopek, T. R. and F. Hutchinson. 1982. DNA base sequence changes induced by bromouracil mutagenesis of lambda phage. J. Mol. Biol. **159**:19-33.

Smith, C A., and D. S. Okumoto. 1984. Nature of DNA repair synthesis resistant to inhibitors of polymerase α in human cells. Biochemistry **23**:1383-1391.

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

Strauss, B., S. Rabkin., D. Sagher, and P. Moore. 1982. The role of DNA polymerase in base substitution mutagenesis on non-instructional templates. Biochemie. **64**:829-838.

Su, S. and P. Modrich. 1986. *Escherichia coli mutS*-encoded protein binds to mismatched DNA base pairs. Proc. Natl. Acad. Sci. USA **83**:5057-5061.

Thacker, J., P. G. Debenham, A. Stretch, and M. B. T. Webb. 1983. The use of a cloned bacterial gene to study mutation in mammalian cells. Mutat. Res. **111**:9-23.

Thacker, J. 1985. The molecular nature of mutations in cultured mammalian cells: A review. Mutat. Res. **150**:431-442.

Tindall, K. R., Stankowski, L. F. Jr., Machanoff, R., and A. W. Hsie. 1984. Detection of deletion mutations in pSV2gpt-transformed cells. Mol. Cell. Biol. **4**:1411-1415.

Wagner, M., J. A. Sharp, and W. C. Summers. 1981. Nucleotide sequence of the thymidine kinase gene of the herpes simplex virus type I. Proc. Natl. Acad. Sci. USA **78**:1441-1445.

Wagner, R. Jr. and M. Meselson. 1976. Repair tracts in mismatched DNA heteroduplexes. Proc. Natl. Acad. Sci. USA **73**:4135-4139.

Walker, G. C. 1984. Mutagenesis and inducible responses to deoxyribonucleic acid damage in *Escherichia coli*. Microbiol. Rev. **48**:60-93.

Waqar, M. A., M. J. Evans, K. F. Manly, R. G. Hughes, and J. A. Huberman. 1984. Effects of 2',3'-dideoxynucleosides on mammalian cells and viruses. J. Cell Phys. **121**:402-408.

Weigle, J. J. 1953. Induction of mutation in a bacterial virus. Proc. Natl. Acad. Sci. USA **39**:628-636.

Witkin, E. M. 1976. Ultraviolet mutagenesis and inducible DNA repair in *E. coli*. Bacteriol. Rev. **40**:869-907.

Wobbe, C. R., F. Dean, L. Weissbach, and J. Hurwitz. 1985. *In vitro* replication of duplex circular DNA containing the simian virus 40 DNA origin site. Proc. Natl. Acad. Sci. USA **82**:5710-5714.

Wood, R. D. and F. Hutchinson. 1984. Non-taregted mutagenesis of unirradiated lambda phage in *Escherichia coli* host cells irradiated with ultraviolet light. J. Mol. Biol. **173**:293-305.

Wood, R. D., T. R. Skopek, and F. Hutchinson. 1984. Changes in base sequence induced by targeted mutagenesis of lambda phage by ultraviolet light. J. Mol. Biol. **173**:273-291.

Yang, J., V. M. Maher, and J. J. McCormick. 1987. Kinds of mutations formed when a shuttle vector containing adducts of benzo[a]pyrene-7,8-diol-9,10-epoxide replicates in COS7 cells. Mol. Cell. Biol. **7**:1267-1270.

Young, R. A. and R. W. Davis. 1983. Yeast RNA polymerase II genes: isolation with antibody probes. Science **222**:778-782.

**Publications.**

Summers, W. C., S. N. Sarkar, and P. M. Glazer. 1985. Direct and inducible mutagenesis in mammalian cells. Cancer Surveys. 4:517-529.

Glazer, P. M., S. N. Sarkar., and W. C. Summers. 1986. Detection and analysis of UV-induced mutations in mammalian cell DNA using a lambda phage shuttle vector. Proc. Natl. Acad. Sci. USA 83:1041-1044.

Glazer, P. M., S. N. Sarkar., G. E. Chisholm, and W. C. Summers. 1987. DNA mismatch repair detected in human cell extracts. Mol. Cell. Biol. 7:218-224.