6.1 Introduction

6.1.1 Increased spontaneous mutation in MMR deficient cells

Mutations in DNA mismatch repair (MMR) genes are known to affect a number of cellular processes in bacteria, yeast and mammals. The primary role of MMR is to repair mismatched nucleotide pairs generated during DNA replication resulting from DNA polymerase errors. A deficiency in the MMR system will lead to an increased spontaneous mutation rate, known as a "mutator phenotype". Mutator phenotypes reflect deficiencies in several different DNA repair processes, which include nucleotide excision repair (NER), base excision repair (BER) and mismatch repair (MMR) (Charames and Bapat, 2003). The mutator phenotype can be quantified by measuring the mutation rate of reporter genes. For example, ES cells with a mutation in the *Msh2* gene exhibited a 1,000 fold elevated mutation rate in a *HSV-TK* transgene introduced into the genome by gene-targeting (Abuin et al., 2000).

6.1.2 Microsatellite instability as a hallmark of MMR deficiency

Besides an elevated mutation rate, mismatch repair deficiency causes instability in simple sequence repeats (microsatellites) (Levinson and Gutman, 1987). During the replication of a simple sequence repeat, such as mono-,di-, tri- and tetranucleotide repeats, the nascent strand may slip along the template, leading to a bulged mispaired insertion-deletion loop (IDLs), which is a substrate for the DNA mismatch repair system (Parker and Marinus, 1992). Deficiency in the MMR system results in a change in the length of these simple repeats, a phenomena called microsatellite instability (MSI). Simple sequence repeats are abundant in eukaryote genomes, so microsatellite instability has been wildly used as an indicator for MMR deficiency in humans and in mice. In order to check MSI *in vivo*, the endogenous microsatellite repeats are normally recovered by PCR-amplification and sequenced to discover the changes in sequence of the repeats.

However, this method doesn't allow MSI to be quantified and low level of MSI could be easily missed. For example, *Msh6* mutant mice don't exhibit obvious MSI when examined by this method (Edelmann et al., 1997).

An alternative method for examining MSI is by introducing a slippage reporter construct into cells to monitor the MSI. This type of reporter carries a series of short tandem repeats that places the reporter gene out of its reading frame, so that the reporter gene is silent. Slippage mutations that gain or lose repeat units may place the reporter gene in-frame, which can then be identified by selecting for the expression of the reporter gene. This method has been applied to bacteria, yeast and has been used in mammalian cell lines (Farber et al., 1994, Yamada et al., 2003). Abuin et al (2000) constructed a gene-targeting vector carrying a di- nucleotide repeat [poly (CA/GT)] that disrupts the reading frame of a downstream neomycin gene. He targeted this slippage construct into a specific genomic locus and determined the MSI rate by Luria-Delbruck fluctuation analysis and found that this rate was elevated four orders of magnitude in *Msh2*-deficient ES cells and 15 fold in *Msh3*-deficient ES cells (Abuin et al., 2000). Pothof et al (2003) carried out an RNA interference (*RNAi*)–based screen in *C. elegans* for genes that affect the stability of a series of mononucleotides $(A)_{17}$ repeats placed 5' of a fused *gfp/lacZ* (green fluorescence protein/ β-galactosidase) reporter. Many *C.elegans* genes that protect *C.elegans* genome against mutation were identified in this screen, including *C.elegans* homologues of *Msh2, Msh6, Mlh1* and *Pms2*.

6.1.3 Determination of mutation rate by Luria-Delbruck fluctuation analysis

Fluctuation analysis was designed by Luria and Delbruck in 1943 to test whether phage resistant bacteria arose from random mutations or from acquired hereditary immunity. They seeded virus sensitive bacteria at very low density in number of cultures and allowed each to grow to a final culture individually. The number of the phage resistant bacteria in each final culture was then counted.

Based on a random mutation hypothesis, each bacteria has a fixed probability of becoming resistant to phage during each population doubling. This is referred to as the "mutation rate". Thus, the number of the phage resistant bacteria in each culture is the result of a combination of the random mutations that occur at each population doubling and the replication of existing mutants during bacterial growth. The number of mutants will not distribute following Poisson's law, but will vary greatly between each culture, this is called fluctuation. This is actually what Luria and Delbruck observed in their study. By calculating the mean number of mutants, or the proportion of cultures with no mutants, the mutation rate could be obtained by two statistical calculations, which are referred to as Luria and Delbruck method of means, or the P_0 method (luria, 1943). Luria-Delbruck analysis has become one of the most popular methods used to measure the mutation rate in mammalian cells. To determine the mutation rate of cells in a culture by fluctuation analysis, a group of parallel cultures is established with a predetermined small number of cells, ideally at a single cell density. These cells are expanded to a larger number, counted and plated in selective medium to allow the growth of resistant cell colonies. The resistant colonies in each culture are counted and used to calculate the mutation rate according to the mathematical equations provided by Luria-Delbruck.

6.1.4 DNA mismatch repair deficiency leads to increased homologous recombination between diverged sequences.

The DNA mismatch repair machinery is implicated in blocking recombination between diverged sequences (homeologous recombination) in bacteria, yeast and in mice. It has been proposed that the intermediate product of homeogolous recombination contains mismatched nucleotides, which can be recognized by the MMR system (Bailis and Rothstein, 1990, Selva et al., 1995). Evidence that MMR blocks homeologous recombination in the mouse came from gene-targeting experiments performed by homologous recombination. In this experiment, the retinoblastoma (*Rb*) locus in mouse ES cells were targeted using targeting

constructs made with either isogenic genomic DNA or with non-isogenic genomic DNA (te Riele et al., 1992). The non-isogenic targeting construct contained 0.6% base sequence divergence compared to the isogenic targeting construct. In *Msh2*-proficient ES cells, gene-targeting with the non-isogenic targeting constuct was blocked, resulting in a significant decrease in targeting efficiency compared to gene-targeting experiments with the isogenic targeting construct. In an *Msh2*-deficient cell line, homologous recombination with the non-isogenic *Rb* construct was as efficient as with the isogenic constuct (de Wind et al., 1995, Claij and Te Riele, 2002). The effects of mismatch repair proteins Msh2 and Msh3 on preventing homeologous recombination were also assessed by comparing the gene-targeting efficiencies with isogenic and non-isogenic gene-targeting vectors at the *Hprt* locus. In *Msh2*$^{-/-}$ and *Msh2, Msh3* double null cells, the targeting efficiency with non-isogenic vector was comparable to the targeting efficiency with isogenic targeting vectors. However, *Msh3*$^{-/-}$ deficiency itself didn't show a defect in homologous recombination (Abuin et al., 2000).

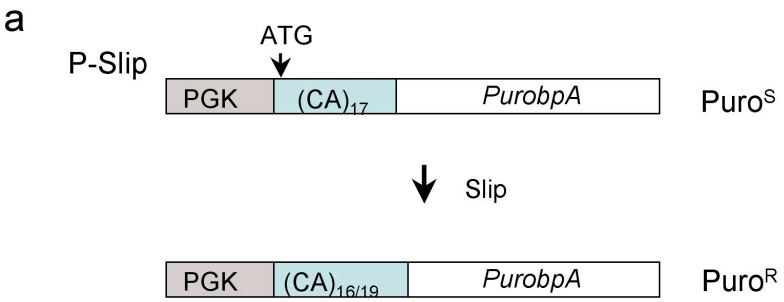
In Chapter 5, potential mismatch repair mutants have been identified in genetic screens for gene trap mutants that are resistant to 6TG, taking advantage of the fact that MMR deficiency causes tolerance to this DNA damaging drug. Many *Msh6* mutants were identified. Other than *Msh6*, novel gene trap mutations were recovered exhibiting tolerance to 6TG. Among these clones is gene trap mutation in *Dnmt1.* The gene trap *Dnmt1* mutant (*Dnmt1-V1*) is a bi-allelic mutant and the expression of *Dnmt1* is fully blocked by the gene trap insertion. The 6TG resistance phenotype was revertible by Cre-mediated deletion of the inserted retrovirus. In this chapter, the function of *Dnmt1* in mismatch repair process is further investigated.

6.2 Results

6.2.1 Construction of the P-Slip slippage cassette

The slippage construct that contains di-nucleotides repeats $(CA)_{17}$ in front of the neomycin reporter gene (*Neo*) has proved to be very useful in examing microsatellite instability in MMR deficiency cells (Farber et al., 1994, Abuin et al., 2000). The gene trap mutants in this study already carry the *βgeo* reporter that renders cells neomycin resistant. To be able to measure MSI activity, a slippage cassette, P-Slip, was designed to contain a series of di-nucleotides repeats following the initiation codon of puromycin phosphotransferase (*puro*) and place it out of it's correct reading frame. Cells harbouring the P-Slip cassette should be sensitive to puromycin due to the frame-shift mutation. If a change in the size of the repeat occurs during the growth of the cells, the reading frame of the P-Slip may be reconstituted, a functional puromycin phosphotransferase protein will be produced and the cell will acquire resistance to puromycin. Therefore, P-Slip allows the determination of the rate of MSI by measuring the number of puromycin resistant colonies.

P-Slip was assembled with a PCR amplified PGK promoter, a *purobpA* fragment lacking the ATG initiation codon, and a strand of synthesized oligonucleotide containing seventeen (CA) repeats ($(CA)_{17}$). The strand of $(CA)_{17}$ repeats was placed just downstream of the ATG start codon, leading to an out of frame mutation in the puromycin cassette (Fig. 6-1). A control vector (P-Slip-ON) was constructed with the same strategy in parallel with P-Slip construct. In contrast to the out of frame P-Slip cassette, P-Slip-ON carries sixteen (CA) repeats, so that the puromycin expression cassette is in frame. P-Slip-ON mimics one type of slippage mutation and thus provides a positive control for the P-Slip cassette. Cells carrying P-Slip-ON cassette should be resistant to puromycin. To check the function of the P-Slip vectors, P-Slip, P-Slip-ON and the parental



b

(CA) 17 repeat

ATGGGGACCGAGTACggatccaacacacacacacacacacacacacaca

cacacacagatacactagtAAGCCCACGGTGCGCCTCGCCACCCGC

GACGA

Figure 6-1. Construction of the slippage vector, P-Slip

a. Schematic representive of P-Slip structure. P-Slip contains a di-nucleotide repeats $(CA)_{17}$ that was inserted in puromycin phosphotransferase(puro) following the ATG , placing the puromycin phosphotransferase out of its normal reading frame. Cells carrying the P-Slip cassette are sensitive to puromycin selection($Puro^S$). Changes in the size of the repeat could place the puromycin phosphotransferase protein in frame and the cell will acquire resistance to Puromycin ($Puro^R$). **b.** The inserted $(CA)_{17}$ repeat (blue colored font) in puromycin coding fragment (black colored font). The $(CA)_{17}$ repeat contains 53 basepairs, which lead to a +2 mutation. Loss of one CA repeat or gain of two repeat units will then restore the reading frame of *Puro*.

PGKpurobpA cassette were linearized and introduced into AB2.2 cells by electroporation. Puromycin resistant ($Puro^R$) cells were selected with 3 μM puromycin for 8 days. $Puro^R$ ES cell colonies were recovered from ES cells with P-Slip-ON and the parental PGKpurobpA, but not from ES cells carrying P-Slip. These results verified that CA repeat in front of the puromycin gene didn't interfere with its activity.

6.2.2 Gene-targeting P-Slip in *ROSA26* locus

In order to provide a precise comparison of the MSI rate between different cell lines, a gene targeting strategy was used to introduce a single copy of P-Slip cassette into a specific genomic locus, *ROSA26*. By using this strategy variation of the MSI rate caused by the copy number of the slippage cassette and the positions effects could be avoided. The *Rosa26* locus was originally identified by retroviral gene trapping in ES cells (Friedrich and Soriano, 1991, Zambrowicz et al., 1997). This locus has been frequently employed for expressing exogenous genes because of the broad spectrum of expression revealed by the gene trap reporter. Also, gene targeting is highly efficient at this locus (Soriano, 1999). Efficient gene-targeting at this locus is important in our study because it allows the P-Slip cassette to be efficiently introduced into multiple cell lines. To make a gene-targeting vector for the *ROSA26* locus (*ROSA26*/Slip-TV), the P-Slip cassette and a PGKBSD selection cassette were inserted between two *ROSA26* genomic arms, of 2.5 kb and 1.9 kb in length. The gene-targeted allele can be identified by Southern analysis using a 5' external probe, which recognizes the targeted allele as a 5.8 kb *Eco*RI fragment and the wild type allele as a 15.5 kb *Eco*RI fragment. Figure 6-2 illustrates the gene-targeting strategy at the *ROSA26* locus and the Southern blot analysis demonstrating three cell lines carrying the targeted allele.

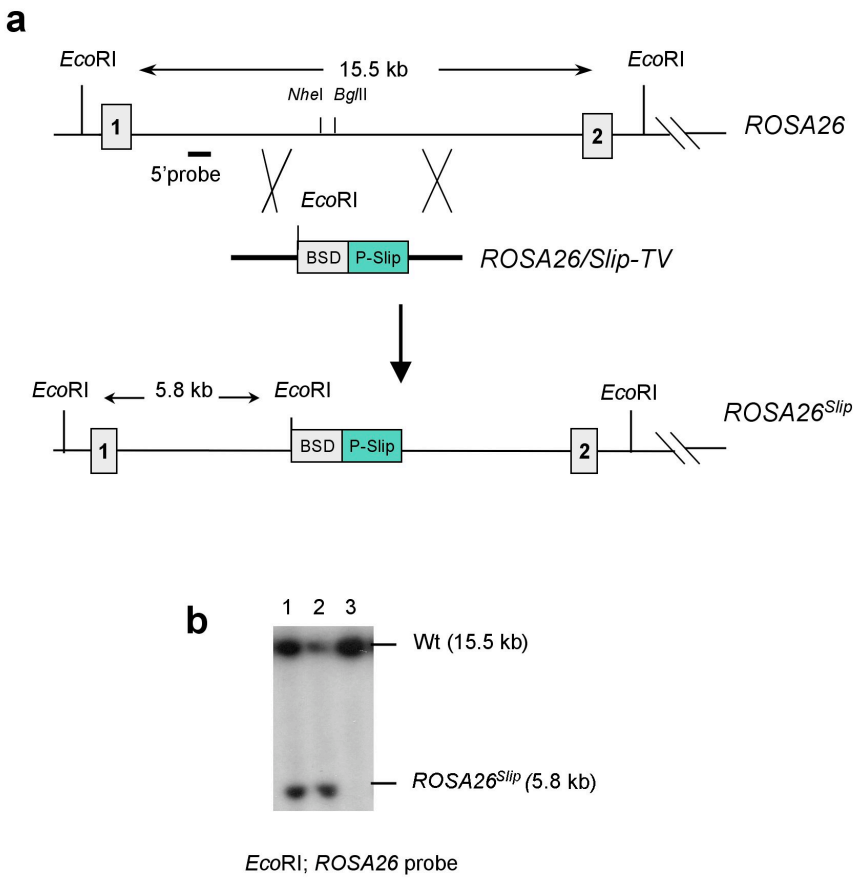


Figure 6-2. Gene-targeting of P-Slip cassette into *ROSA26* locus

a. Schematic of gene-targeting strategy showing the 5' portion of *ROSA26* locus and the gene-targeting vector (*ROSA26/Slip-TV).*
b. Southern-blot showing targeted cells carrying *ROSA26*^{Slip} allele(lane 1 and lane 2).

6.2.3 Determination of MSI rate of the targeted P-Slip cassette in gene trap mutants

The *ROSA26*/Slip gene-targeting vector was targeted into gene-trap mutants that contained the gene-trap cassette in *Msh6, Dnmt1, Tgif,* and *Adprtl2/Rbpsuh* loci. These clones were identified from STB screens (Chapter 5, Table 5-2), and all exhibited resistance to 6TG(2 μM). Both *Msh6* and *Dnmt1* gene-trap mutants are homozygous mutations and the 6TG resistance phenotype is revertible. The *Tgif* gene-trap mutant (*Tgif-V1)* is also a bi-allelic mutant, but the 6TG resistance phenotype is not revertible. The *Parp-2/Rbpsuh* is a complex locus involving two genes, *Parp-2* and *Rbpsuh*. Chromosomal translocation between *Parp-2 and Rbpsuh* loci causes a frame-shift mutation in both genes. The 6TG resistance phenotype is not revertible. The *Blm*-deficient cell line NGG5-3 and a *Blm* wild-type cell line (AB2.2) were used as controls. After confirmation of gene-targeting by Southern analysis using a 5' *ROSA26* external probe, targeted cells from each of the gene-trap cell line were seeded at low density and allowed to form ES cell colones. 12 clones from each cell line were picked into 96 well tissue culture plate and expanded for Luria-Delbruck fluctuation analysis to determine the MSI rate (Methods 2.9). The results are listed in table 6-1. The gene trap *Msh6* mutant exhibited a MSI rate about 10 fold higher than AB2.2 cells. The *Blm*-deficient cells (NGG5-3) exhibited a 2 fold increase in MSI compared to AB2.2 cells. The MSI rate in *Tgif-V1* cells is slightly higher than the *Blm*-deficient cells. However, no obvious change of the MSI rate was observed in the gene trap *Dnmt1-V1*, and *Parp-2/Rbpsuh* cells.

6.2.4 Generation and characterization of Dnmt1-deficient ES cells by gene-targeting

The gene-trap clones were generated on a *Blm*-deficient genetic background. The observation of the microsatellite instability in *Blm*-deficient ES cells raised

Table 6-1. Mutation rate of P-Slip in gene-trap mutants

ES cell lines	Number of selections	Total number of cells	Number of $Puro^R$ clones per select *	Rate of MSI
AB2.2 /#1	12	$7.0x10^7$	16,16,13,5,3,2,0(6)	$2.6 x10^{-7}$
NGG5-3 /#1	12	$4.5 x10^7$	47,11,9.6,5(3),4,3,2(3)	$6.3 x10^{-7}$
Dnmt-V1 /#1	12	$4.6 x10^7$	33,17,15,8,6,4,2,0(5),	$5.6 x10^{-7}$
Tgif-V1/ #1	12	$3.8 x10^7$	72,15,12,11,8,7,6,5(3),2,0	$1.06 x10^{-6}$
STB60/#1	12	$3.5 x10^7$	23(2),5,4,2(4),1(3),0	$6.1 x10^{-7}$
STB20 (*Msh6*) /#1	12	$4.4 x10^7$	152,129,125 100,73,45,44,36,34,30,22,11	$3.6 x10^{-6}$

Table 6-1. Mutation rate of P-Slip in gene-trap mutants

P-Slip was targeted into *ROSA26* locus, the rate of MSI was determined by Luria-Delbruck fluctuation analysis in four gene-trap cell lines, *Dnmt1-V1*, *Tgif*, STB60 (*Parp-l2/Rbpsuh*) and STB20 (a *Msh6*mutant). AB2.2 and NGG5-3 cells were used as controls. " #1" represents the experiment number.

* The order of numbers was arranged arbitrarily from the most to the fewest. The number in parentheses indicates the number of selections giving rise to the corresponding number of drug resistant clones.

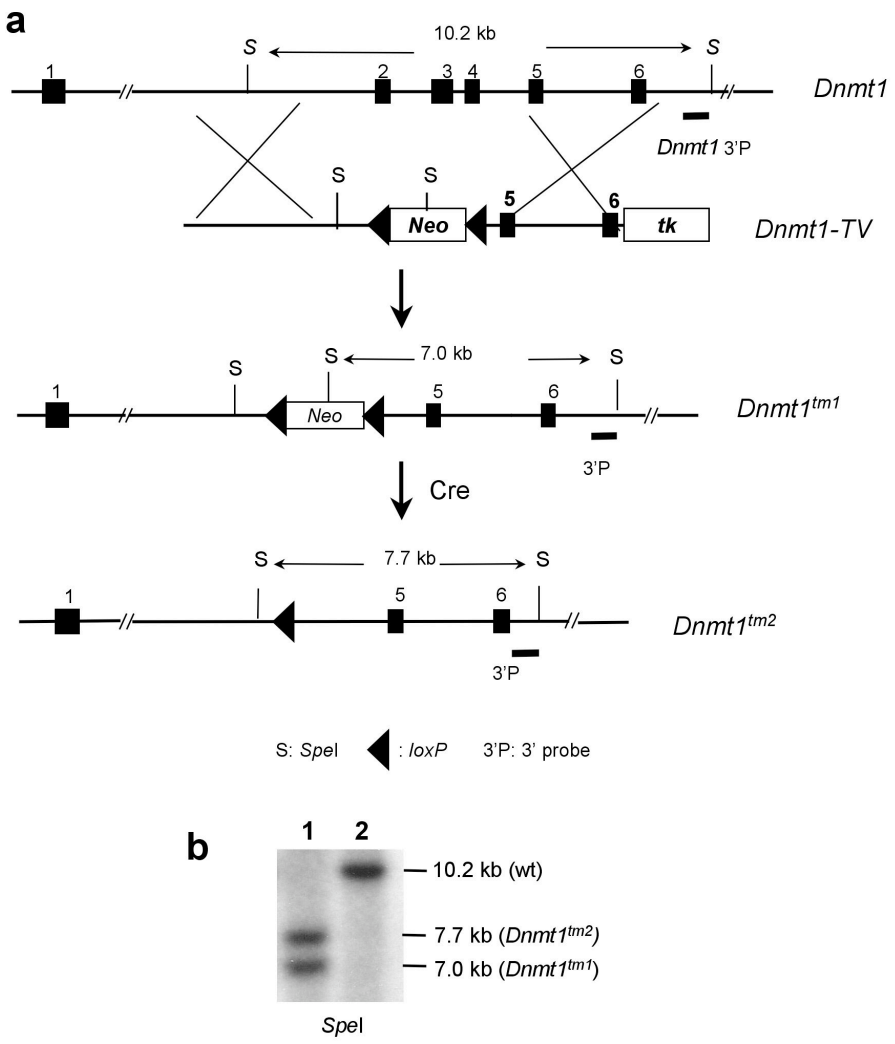


Figure 6-3. Generation *Dnmt1* knockout cells by gene-targeting

a. Schematic of gene-targeting *Dnmt1* locus and the targeted *Dnmt1*^{tm1} and *Dnmt1*^{tm2} alleles. **b**. Southern blot showing a homozyously gene targeted *Dnmt1-KO* cells carrying *Dnmt1*^{tm1} and *Dnmt1*^{tm2} alleles. Lane 1, *Dnmt1*$^{tm1/tm2}$; Lane 2, wild type cells

the concern of a genetic interaction between mismatch repair genes and the *Blm*-deficient background, which could interfere with the analysis of MSI in the gene trap clones. To examine the MSI activity without the effect of *Blm*-deficiency, a *Dnmt1*-deficient cell line was generated on the AB2.2 genetic background by gene-targeting. This cell line was named as *Dnmt1-KO* for *Dnmt1* knock out cell line. *Dnmt1-KO* cells contain two targeted alleles, $Dnmt1^{tm1}$ and $Dnmt1^{tm2}$, in which a 5.5 kb DNA fragment including *Dnmt1* exon 2, exon 3 and exon 4 were deleted (Fig. 6-3).

6.2.4.1 Expression analysis of Dnmt1

Dnmt1 encodes a 5.2 kb messenger RNA which derived from 37 exons. Deletion of exon 2, exon 3 and exon 4 in *Dnmt1-KO* cells removes a 359 bp coding region, which results in a frame-shift mutation (Fig. 6-4). Thus, *Dnmt1-KO* cell line is expected to be a null. A pair of PCR primers was designed to amplify a *Dnmt1* cDNA fragment spanning *Dnmt1* exon 1 to exon 6, which is 615 bp in length. In *Dnmt1-KO* cells, a 256 bp DNA fragment will be amplified because of the 359 bp deletion. RT-PCR analysis revealed the predicted 615 bp and 256 bp fragments in heterozygous gene-targeted $Dnmt1^{+/tm1}$ cells and only a 256 bp fragment in the targeted *Dnmt1-KO* ($Dnmt1^{tm1/tm2}$) cells (Fig. 6-5 a). This result confirmed the knockout mutation created by gene-targeting.

To examine the expression of *Dnmt1*, a cDNA probe was PCR amplified spanning *Dnmt1* exon 8 to exon 39. Northern-blot analysis was carried out on total RNA extracted from $Dnmt1^{+/tm1}$, *Dnmt1-KO*, gene-trap *Dnmt1-V1* and AB2.2 cells. This experiment demonstrated that *Dnmt1* is expressed in the heterozygous and homozygous knockout cells at similar level to the wild type cells. Because *Dnmt1* mRNA is over 5 kb, a 359 bp deletion in the gene-targeted alleles is not resolved by Northern-blot analysis. *Dnmt1* expression could not be detected in the gene-trap *Dnmt1-V1* cells (Fig. 6-5 b).

TCGCGCGAAAAAGCCGGGGTCTCGTTCAGAGCTGTTCTGTCGTCTGC
AACCTGCAAGATGCCAGCGCGAACAGCTCCAGCCCGAGTGCCTGCG
CTTGCCTCCCCGGCAGGCTCGCTCCCGGACCATGTCCGCAGGCGgctc
aaagacttggaaagagatggcttaacagaaaaggagtgtgtgagggagaaattaaacttactgcatg
aattcctgcaaacagaaataaaaagccagttgtgtgacttggaaaccaaattacataaagaggaattat
ctgaggaaggctacctggctaaagtcaagtccctcttaaataaggatttgtccttggagaacggaacac
acactctcactcaaaaagccaacggttgtcccgccaacgggagccggccaacctggagagcagaaa
tggcagactcaaatagatccccaagatccaggcccaagcctcggggacccaggagaagcaagtcg
gacagtgacacccttt**CAGTTGAAACTTCACCTAGTTCCGTGGCTACGAGGAG**
AACCACCAGGCAGACCACCATCACGGCTCACTTCACGAAGGGCCCCA
CTAAACGGAAACCCAAGGAAGAGTCGGAAGAGGGGAACTCGGCTGAG
TCGGCTGCAGAGGAGAGAGACCAG

Figure 6-4. 5' portion of *Dnmt1* coding sequence.

The deletion in the knockout cell line, *Dnmt1-KO,* is represented as lowercase fonts and underlined, which includes 356 base pairs. The ATG codon is in red. The deletion results in a frame shift mutation.

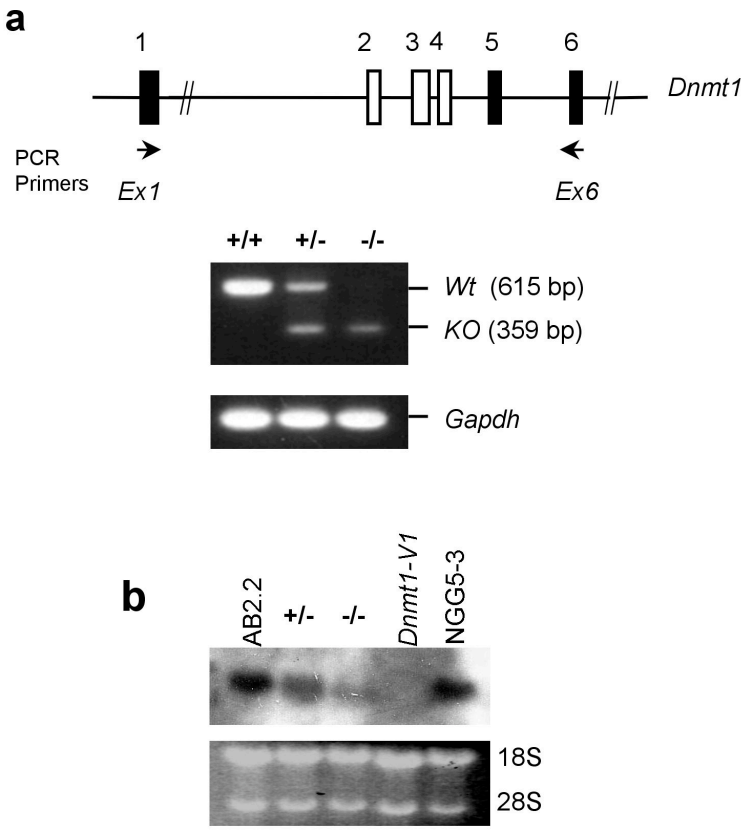


Figure 6-5. Expression analysis of *Dnmt1*

a. RT-PCR amplification of *Dnmt1* cDNA spanning exon 1 to exon 6, demonstrating the deleted transcript (359 bp) in the heterozygous targeted *Dnmt1*$^{+/tm1}$ *(+/-)* and double targeted *Dnmt1*$^{tm1/tm2}$*(-/-)* cells. Note that Exon 2, exon 3 and exon 4 are deleted in the gene-targeted *Dnmt1*^{tm1} and *Dnmt1*^{tm2} alleles (open boxes). *Gapdh,* RT-PCR control. **b.** Northern analysis showing *Dnmt1* expression in the gene trap *Dnmt1-V1* cells, the heterozygous targeted *Dnmt1*$^{+/tm1}$ (+/-) cells, double targeted *Dnmt1*$^{tm1/tm2}$ cells (*Dnmt1-KO*; +/-), and NGG5-3 (*Blm*-deficient) cells. Note that the expression of *Dnmt1* is absent in *Dnmt1-V1* cells. Expression of *Dnmt1* is reduced in *Dnmt1-KO* cells compared to *Dnmt1* wild type AB2.2 and NGG5-3 cells. 18s and 28s, loading controls.

6.2.4.2 Functional analysis of Dnmt1 activity

Dnmt1 protein encodes a cytosine-5 methyltransferase activity, which transfers methyl groups to cytosine at CpG sites (Li et al., 1992). The methylation level at CpG sites can be examined by comparing the restriction digestion pattern generated by restriction enzymes, *Msp* I and *Hpa* II. *Msp* I and *Hpa* II are isoschizomers that recognize the CCGG sites, except that *Hpa* II is sensitive to methylated cytosine at CCGG sites (Fig. 6-6 a). The CpG site in the CCGG sequence is subjected to methylation by Dnmt1. Therefore, genomic DNA with a higher level of CpG methylation will be more resistant to *Hpa* II digestion (Chapman et al., 1984). To measure Dnmt1 activity, genomic DNA was extracted from heterozygous gene-targeted *Dnmt1*$^{+/tm1}$, homozygous gene-targeted *Dnmt1*$^{tm1/tm2}$, gene-trap *Dnmt1-V1* cells and *Dnmt1* wild type control AB2.2 cells and digested with *Msp* I and *Hpa* II, respectively. The digested genomic DNA was then separated on agarose gel and visualized by staining with EtBr (ethidium bromide). Comparison of the *Msp* I and *Hpa* II digestion of each sample revealed that genomic DNA extracted from *Dnmt1*$^{tm1/tm2}$ and *Dnmt1-V1* cells are globally undermethylated compared to AB2.2 cells (Fig. 6-6 b).

A Southern-blot analysis was conducted to assess the methylation of centromeric repeats. Gamma satellite repeats are the major component of centromere repeats composed of a 234 bp repeat unit (Lundgren et al., 2000). Centromeric repetitive sequences are normally highly methylated. To investigate the methylation level at centromeric repeats, *Msp* I and *Hpa* II digested genomic DNA were blotted from an agarose gel to nylon membranes and probed with a gamma satellite repeats probe (Fig. 6-6 b). Southen analysis revealed that gamma satellite repeats were undermethylated in the gene-trap *Dnmt1-V1* cells since *Hpa* II digested genomic DNA exhibited the same defined hybridization pattern as *Msp* I digestion. However, the methylation of gamma satellite repeats

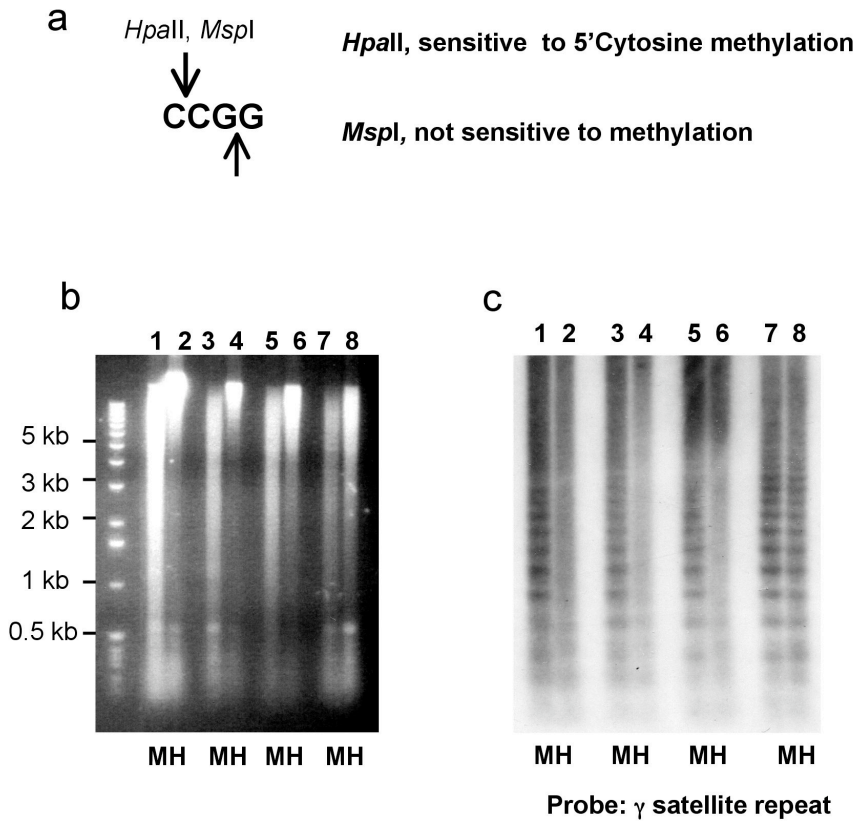


Figure 6-6. Determinatin of DNA methylation.

a. *Hpa*II and *Msp*I are isoschizomers that recognize the same CCGG sites. *Hpa*II is sensitive to the methylated CpG. Methylated CCGG can only be digested with *Msp*I, but not *Hpa*II. **b**. Genomic DNA digested with either *Msp*I or *Hpa*II (left panel) and probed with γ satellite repeat (right panel**).** *Msp*I digested genome: odd numbered lane labed with M; *Hpa*II digested genome: even numbers lanes labled with H. Genomic DNA are from *Dnmt1*-proficient AB2.2 cells (lane 1-2), *Dnmt1*$^{+/tm1}$(lane 3-4), *Dnmt1* $^{tm1/tm2}$ *(Dnmt1-KO*) (lane 5-6*)*, gene-trap *Dnmt1-V1* cells (lane 7-8).

Table 6-2. Mutation rate of P-Slip in *Dnmt1*- deficient cells

ES cell lines	Number of drug selections	Total number of cells	Number of Puro resistant clones per select *	Mutation Rate
AB2.2 /#2	24	$2.8x10^7$	10,3,21,1(3),0(16)	$3.2x10^{-7}$
AB2.2 /#3	24	$3.3x10^7$	29,10,5,2,1,0(19)	$5.0x10^{-7}$
NGG5-3 /#2	24	$1.9x10^7$	7,5,4,3,2(6),1(4),0(10)	$7.0x10^{-7}$
NGG5-3 /#3	19	$1.8x10^7$	4,3,2(5),1(3),0(9)	$5.0x10^{-7}$
Dnmt1-V1 /#2	24	$1.8x10^7$	6,2,(3),1(2),0(18)	$3.7x10^{-7}$
Dnmt1-KO/#2	24	$2.2x10^7$	84,7,7,4,3,2,2,1(5),0(12)	$1.4x10^{-6}$
Dnmt1-KO/#3	23	$3.0x10^7$	128,12(2),5,4(2),3(2),2(2),1(6),0(7)	$1.5x10^{-6}$

Table 6-2. Mutation rate of P-Slip in *Dnmt1*-deficient cells.

P-Slip was targeted into *ROSA26* locus, the rate of MSI was determined by Luria-Delbruck fluctuation analysis in *Dnmt1*-deficient cells. Dnmt1-V1 represents the gene-trap *Dnmt1* mutants. Dnmt1-KO represents the gene-targeted $Dnmt1^{tm1/tm2}$ cells. This experiment was done twice and the data was listed separately and marked as " #2" and " #3 " for the second, and third experiments, respectively. Note "#1" represents the Luria-Delbruck fluctuation analysis experiment listed in table 6-1

* The order of numbers was arranged arbitrarily from the most to the fewest. The number in parentheses indicates the number of selections giving rise to the corresponding number of drug resistant clones.

seemed to be maintained to a similar level in both heterozygous gene-targeted *Dnmt1*$^{+/tm1}$ and homozygous gene-targeted *Dnmt1*$^{tm1/tm2}$ cells as in *Dnmt1* wild type AB2.2 cells, which was demonstrated by the lack of the defined hybridization pattern on *Hpa* II digested genomic DNA (Fig. 6-6 c). These results suggest that the gene-targeted *Dnmt1*^{tm1} and *Dnmt1*^{tm2} alleles generate hypomorphic mutations of *Dnmt1*.

6.2.5 Dnmt1 deficiency in the Dnmt1 knockout cells causes increased MSI

To test the MSI rate in the gene-targeted Dnmt1-deficient cells, the P-Slip cassette was targeted into *ROSA26* locus in *Dnmt1-KO* cells. 20 or 24 single cell clones were obtained by low density plating from the P-Slip targeted *Dnmt1-KO* cells, gene-trap *Dnmt1-V1* cells, *Blm*-deficient cells and the AB2.2 cells. MSI in the targeted P-Slip were determined by Luria-Delbruck fluctuation analysis as described previously. This experiment was repeated once. The results are listed in table 6-2. The MSI activity in *Dnmt1-KO* cells is 3 to 5 fold higher that the MSI rate in AB 2.2 cells and about 2 fold higher than the rate in Blm-deficient cells. The gene-trap *Dnmt1-V1* cell line, which is on the *Blm*-deficient genetic background, however, exhibited a MSI rate lower than *Dnmt1-KO* cells. Because *Dnmt1* expression is fully blocked in *Dnmt1-V1* cells, the difference of MSI rate between *Dnmt1-V1* cells and *Dnmt1-KO* cells are less likely to be derived from residual function in *Dnmt1-V1* cells. Clearly, there is a genetic interaction between *Dnmt*1 and the *Blm gene*. In fact, methylation analysis has shown that *Dnmt1-KO* cells have partial Dnmt1 function. Thus, It would be interesting to know if a null *Dnmt1* mutation in AB2.2 cells would give a higher level of MSI.

6.2.6 The puromycin resistant ES cells carry a single copy of the targeted P-Slip cassette.

Gene-targeted ES cell clones are picked from many ES cells colonies growing on a 90 mm tissue culture plate. Since the gene-targeting efficiency is low, the

majority of the ES cells clones selected in an experiment are those carrying random insertions of the gene-targeting vector. Thus, the gene-targeted ES cell clones may be contaminated with small numbers of cells with randomly inserted transgenes. This contamination is normally ignored because generally less than 10 % of cells represent contaminants. Fluctuation analysis involves plating gene-targeted clones at low density to recover single cell clones, so there is the possibility that a single cell clone may be derived from cells that carry a randomly inserted P-Slip vector. As random insertion is often accompanied by a head-tail concatermerization of the gene-targeting vector, multiple copies of P-Slip cassette will be inserted into genome, logically this clone would have a higher likelihood of being able to restore puromycin expression. To ensure that the puromycin resistant slippage clones do contain the targeted P-Slip cassette at the *ROSA26* locus, 12 puromysin resistant clones were recovered from each cell line and genomic DNA were extracted. Southern analysis using *ROSA26* probe on these cells verified that the P-Slip cassette was targeted into *ROSA26* locus (data not shown).

6.2.7 Mutated CA repeat in Puromycin resistant ES cells

To determine the nature of the mutations in the $(CA)_{17}$ repeats element acquired to produce puromycin resistance, the P-Slip cassette was amplified by PCR from the recovered puromycin resistant ES cell clones and sequenced. Analysis of the sequence data revealed that all clones examined exhibited insertions or deletions in the $(CA)_{17}$ repeat region (Table 6-3). The most frequently identified mutation was the insertion of two extra (CA) repeats, which occurs in all genetic backgrounds, in different cell lines. This observation suggests that a “+2” insertional mutation at dinucleotide repeats may be a common DNA polymerase II error.

Table 6-3. Sequence analysis of P-Slip mutations

ES cells	Mutation
AB2.2	+2
AB2.2	+2
AB2.2	+2
AB2.2	+2
NGG5-3	-7
NGG5-3	+2
Dnmt1-KO	+2
Dnmt1-KO	+2
Dnmt1-KO	+2
Dnmt1-KO	+2
Dnmt1-KO	+2
Dnmt1-KO	+2
Dnmt1-KO	+2
Dnmt1-KO	+2
Dnmt1-KO	-10
Dnmt1-KO	-7
Dnmt1-KO	+2
Dnmt1-KO	+2
Dnmt1-KO	-10

Table 6-3 Sequence analysis of P-Slip mutations.

Individual Puromycin resistant ES cell clones were recovered from AB2.2, NGG5-3, and the *Dnmt1-KO* cells that carrying the targeted P-Slip cassette. The (CA) repeat was recovered by PCR amplification from genomic DNA and sequenced to identify the mutations. The sequence revealed that insertion of two repeats units (+2) are the most frequently mutations. Loss of seven or ten repeats was observed.

6.2.8 Dnmt1 does not block homeologous recombination

The efficiency of recombination between homologous DNA stretches is highly dependent on their sequence identity (Rayssiguier et al., 1991, Shen and Huang, 1986, Nassif and Engels, 1993, Waldman and Liskay, 1988, te Riele et al., 1992). The mismatch repair protein Msh2 has a role in blocking homeologous recombination, which has been demonstrated by comparision of gene-targeting efficiency using isogenic and non-isogenic *Rb* targeting vectors (described previously). The isogeneic vector, 129Rb-puro is derived from a 129 genomic DNA library. The non-isogenic vector B/cRb-puro is derived from BALB/c-derived library. Both vectors contain 10.5 kb genomic arms and use the puromycin phophotransferase gene (puro) as the selection marker. The nonisogenic construct B/cRb-puro contains 0.6% base sequence divergence with respect to the isogenic 129Rb-puro construct (te Riele et al., 1992, Claij and Te Riele, 2002). This experiment revealed that the targeting efficiency with the non-isogenic B/cRb-puro vector is about 14 fold lower than with the isogenic 129Rb-puro vector in *Msh2* proficient cells. However, in *Msh2*-deficient cells, 129Rb-puro and B/cRb-puro displayed similar levels of gene-targeting efficiency (Claij and Te Riele, 2002).

In order to examine if *Dnmt1* is involved in blocking of homeologous recombination, the 129Rb-puro and B/cRb-puro were used to target the *Rb* locus in AB2.2 cells (Dnmt1-proficient), *Dnmt1*-deficient cells *Dnmt1-KO* and a *Msh2*-deficient ES cell line (*Msh2*$^{-/-}$) (Abuin et al., 2000). Abuin et al. (2000) has shown that homologous recombination with non-isogenic *Hprt* gene-targeting vector could occur at similar efficiency as with isogenic *Hprt* gene-targeting vector in

Table 6-4. Homologous recombination with Rb targeting vectors

ES cell line	Gene-targeting frequency		Isogenic versus Non-isogenic
	129Rb-puro(%)	B/cRb-puro(%)	
AB2.2	11 of 89 (12%)	2 of 90(2.2%)	5.5X
Msh2$^{-/-}$	1 of 48(2%)	2 of 81(2.4%)	0.8X
Dnmt1-KO	15 of 82(18%)	0 of 93(<1%)	>18X

Table 6-4. Homologous recombination with isogenic and non-isogenic Retinoblastoma (*Rb*) targeting vectors.

Gene-targeting experiments were carried out with isogenic *Rb* targeting vector (129Rb-puro) and non-isogenic (B/cRb-puro) vectors. The gene-targeted clones were identified by Southern analysis. The targeting frequency was represented as numbers of targeted clones versus total number of Puromycin resistant clones that exhibited clear Southern-hybridization signals.

these *Msh2*$^{-/-}$ cells. Thus, this *Msh2*$^{-/-}$ cell line serves as a positive control for our experiments. The gene-targeting efficiency for each targeting vector is listed in table 6-4. The isogenic targeting vector has a targeting efficiency 5.5 fold higher than the targeting efficiency of the non-isogenic targeting vector in AB2.2 ES cells. In contrast, the isogenic and nonisogenic targeting vector exhibited similar targeting efficiency in *Msh2*$^{-/-}$ ES cells, suggesting the homeolgous recombination experiment itself works. However, the targeting efficiency of the non-isogenic targeting vector in *Dnmt1-KO* cells was much lower than the targeting efficiency with the isogenic targeting vector, suggesting that homeologous recombination is still blocked in *Dnmt-1* deficient ES cells (Table 6-4).

6.3 Discussion

6.3.1 Summary

Simple sequence repeats (microsatellite sequence) are believed to be hot spots for mutations caused by DNA polymerase slippage. Deficiency in the MMR system leads to a significant increase in microsatellite instability (MSI). In this chapter, a puromycin slippage cassette (P-Slip) was developed to investigate MSI. P-Slip contains a strand of di-nucleotides repeats $(CA)_{17}$ placed out of frame in the puromycin phosphotransferase gene(*puro*), rendering the *puro* non-functional. The rate of MSI in P-Slip can be determined by measuring the number of puromycin resistant cells. P-Slip has been introduced by gene-targeting into *ROSA26* locus of various cell lines. By Luria-Delbruck fluctuation analysis, MSI rates were determined in *Msh6, Dnmt1, Tgif*, and *Parp-2/Rbpsuh* gene-trap mutants. Comparison of MSI rates between gene-trap mutants, AB2.2 ES cells and *Blm*-deficienct control cells revealed an elevated MSI rate in the *Msh6* gene-trap clones. However, a 2.5 fold increase in MSI rate was observed in *Blm*-deficient cells compared to wild-type AB2.2 cells, which complicated interpretation of the MSI rate of gene-trap clones. Gene-trap *Dnmt1, Tgif* and *Parp-2/Rbpsuh* exhibited MSI rates similar to *Blm*-deficient ES cells.

To precisely calculate the MSI rate, it is necessary to separate the gene-trap mutations from Blm-deficiency. A *Dnmt1*-deficient cells line *(Dnmt1-KO*) was generated by gene-targeting in AB2.2 ES cells. *Dnmt1-KO* cells carry $Dnmt1^{tm1}$ and $Dnmt1^{tm2}$ alleles, both of which contain a deletion of exon 2, exon 3 and exon 4. Genomic DNA from *Dnmt1-KO* cells was hypomethylated at CpG sites compared to wild type AB2.2 cells. Comparison of the methylation level of gamma centromeric repeats revealed that the deletion mutation created in the *Dnmt1-KO* cells does not abolish *Dnmt-1* function totally, suggesting that the knockout $Dnmt1^{tm1}$ and $Dnmt1^{tm2}$ alleles are hypomorphic alleles. The MSI rate of the targeted P-Slip cassette in the *Dnmt1-KO* cells was determined by Luria-Delbruck fluctuation analysis, which is nearly 5 fold higher than that in AB2.2 cells. The mutations in the P-Slip cassette were recovered and sequenced to confirm the insertions or deletions in the di- nucleotide repeat. This experiment revealed that “+2” insertions could be a common DNA polymerase II error. Most importantly, these experiments pointed out that *Dnmt1* is involved in maintaining the stability of simple sequence repeats, a function of MMR system.

One mismatch protein, Msh2, has a role in suppressing recombination between diverged DNA sequences. This aspect was also studied in the *Dnmt1-KO* cells in this chapter using gene-targeing vectors that was derived either from isogenic or non-isogenic genomic DNA. No increase in the targeting efficiency was observed with non-isogenic targeting vectors in the *Dnmt1-KO* cells, suggesting that *Dnmt-1* may not be vital in blocking homeologous recombination between diverged DNA sequences.

6.3.2 MSI in *Dnmt-1* deficient cells is not a result of changes in MMR gene expression

Dnmt-1 is required for maintenance of DNA methylation patterns during DNA replication, and functions with other *do novo* methylases, such as *Dnmt3A* and *Dnmt3B*, in establishing the DNA methylation pattern during development (Li et

al., 1992). Gene expression is regulated epigenetically by DNA methylation. For example, promoter methylation leads to transcriptional silencing of a gene's expression. It has been known that mismatch repair gene *Mlh1*is silenced in some human tumor samples resulting from promoter hypermethylation (Veigl et al., 1998). Moreover, a hypomethylated genome is associated with a global increase in gene expression (Jackson-Grusby et al., 2001). MSI observed in *Dnmt1* mutants could be caused indirectly by changes in the expression of mismatch repair genes. To investigate this, the expression of four major mismatch repair genes, *Msh2, Msh6, Mlh1* and *Pms2* were examined by RT-PCR in the gene-trap *Dnmt1-V1* mutant. Compared to the *Dnmt1*-proficient NGG5-3 ES cells, the expression of these MMR genes was slightly increased (data not shown). Although RT-PCR cannot provide precise comparison of gene's expression between each sample, the observation of a general increase in MMR genes' expression in *Dnmt-1* deficient ES cells suggests that the increased MSI in *Dnmt1*-deficient ES cell is not likely to be caused by indirect reduction in the expression of MMR genes.

6.3.3 DNA replication, repair and methylation are coordinated processes

In humans, DNA replication, repair and methylation are coordinated by PCNA (proliferating cell nuclear antigen). PCNA forms a ring-shaped trimeric complex that can encircle double-stranded DNA and slide along it. PCNA itself is the DNA polymerase-processivity factor, which is required for DNA replication (Krishna et al., 1994). Moreover, PCNA provides a sliding platform that can mediate the interaction of proteins with DNA in a non-sequence specific manner. Some DNA repair systems, NER (nucleotide-excision repair) and MMR, are linked to the DNA replication process via interaction with PCNA. PCNA can interact with both eukaryote mismatch repair protein complex MutSα and MutSβ (Umar et al., 1996, Bowers et al., 2001, Lau and Kolodner, 2003). Since the primary function of MMR is to repair mismatched nucleotides generated during DNA replicaiton, the coupling of MMR system with DNA replication is believed to be vital. The

DNA methylation protein, Dnmt1, is recruited to the replicating DNA by binding to PCNA. This interaction allows Dnmt1 to perform its function as a maintenance methyltransferase, and methylate the newly synthesized hemimethylated DNA (Chuang et al., 1997, Vertino et al., 2002).

6.3.4 Evidence of links between DNA methylation and DNA mismatch repair

There is evidence suggesting a physical and functional interaction between DNA methylation and MMR systems. Chen et al (1998) observed that *Dnmt1* mutation causes genomic instability, a mutator phenotype. They measured the mutation rate by fluctuation analysis at both the endogenous X-linked *Hprt* (hypoxanthine phosphoribosyltransferase) gene and an integrated viral thymidine kinase (*tk*) transgene in *Dnmt-1* deficient and proficient ES cells. This experiment revealed a 10 fold increase at *Hprt* gene and 6 fold increase at the *tk* gene in the *Dnmt1*-deficient cells. By PCR analysis and Southern analysis, they found that over 60% of the mutants contain genomic rearrangements, which may be explained by aberrant mitotic recombination. However, a 10 fold increase in the mutation rate of the X-linked *Hprt* gene requires further explanation as this mutation rate cannot be explained by mitotic recombination alone. The mutator phenotype displayed by *Dnmt-1* deficient cells in their study is consistent with the observation of increased microsatellite instability in our study.

Recently, a methyl-CpG binding protein MBD4 was linked to DNA mismatch repair and MMR mediated genomic surveillance. MBD4, originally named as MED1 was identified as a protein interacting with the DNA mismatch repair protein, MLH1, in human cells (Bellacosa et al., 1999). Mammalian MBD4 protein contains glycosylase activity that enzymatically removes thymine (T) from a mismatched T/G base pair at CpG sites (Hendrich et al., 1999). Deamination of 5-methylcytosine (m^5C) to T at CpG sites frequently causes T/G mismatches. MBD4 has been shown to be important in suppressing the mutability of the m^5C.

MBD4 deficiency in mice caused an increase in CpG mutability and tumorigenesis (Millar et al., 2002, Wong et al., 2002). However, the function of MBD4 is not limited to repairing T/G mismatch at CpG sites. MED1 can bind to fully and hemimethylated DNA but not to unmethylated DNA *in vitro*. Moreover, transfection of a dominant negative mutated MED1 into cultured cells leads to microsatellite instability in an episomal slippage construct that contains tandem CA repeats (Bellacosa et al., 1999). These observations made MBD4/MED1 an attractive MMR candidate. Further studies revealed that the frameshift mutations in MBD4/MED1 coding sequence occurred frequently in colon, endometrial, pancreatic and gastric tumors exhibiting high microsatellite instability (Riccio et al., 1999, Bader et al., 1999, Yamada et al., 2002). However, MBD4 deficient mice generated by gene-targeting technology do not exhibit MSI (Millar et al., 2002, Wong et al., 2002). The link between MBD4 and the MMR system is supported by recently studies on cultured embryonic fibroblasts (MEFs) derived from *Mbd4*-deficient mice. In this study the apoptosis response to some DNA damaging drugs were examined and revealed that *Mbd4*-deficiency leads to tolerance of simple methylating agents like MNNG. *Mbd4*-deficient MEF cells also displayed similar tolerance to other DNA damaging drugs, such as platinum drugs, which forms intra and interstrand DNA adducts (Cortellino et al., 2003). It is established that the DNA mismatch repair system plays a role in DNA damage surveillance, in which it recognizes the damaged DNA and induces cell cycle arrest and cell death. MMR mutations allow cells to survive treatment with some DNA damaging drugs. The drug tolerance effects exhibited by *Mbd4*-deficient MEFs is phenotypically similar to the DNA damage tolerance exhibited by MMR deficiency, which is characterized by the accumulation of DNA lesions in cells. The DNA damage tolerance effect was also observed in the small intestine in the *Mbd4*-deficient mice (Sansom et al., 2003). These results suggest that the methyl CpG binding protein MBD4 is a multifunctional protein, which is involved in mismatch repair and its related DNA damage surveillance. Since MBD4 displays differences in the binding affinity to fully methylated DNA and hemimethylated DNA, it logically follows that MBD4 may be the molecular obligator that

transduces the DNA methylation signal laid by Dnmt1 along the replicating DNA to the DNA mismatch repair machinery that involves MLH1.

In *Escherichia coli*, the mismatch repair protein MutH distinguishes the newly synthesized DNA strand from the parental DNA strand using the hemi-methylated adenine at GATC site as a signal. Mutation in bacteria *Dam* DNA methylase, the enzyme methylating the GATC sequence, causes a mutator phenotype. Although DNA mismatch repair machinery is highly conserved in evolution, a functional homologue of strand distinguishing protein MutH has not been identified. *Dnmt1* provides an attractive candidate for such an activity since Dnmt1 protein is located to the DNA replication site and it maintains the CpG methylation post DNA replication. Also, our study demonstrated that *Dnmt1* deficiency caused microsatellite instability. However, some evidence suggests that DNA methylation may not be essential for strand discrimination in eukaryotes, especially organisms that are deficient in genomic methylation for example, *S.cerevisiae*, *Drosophila* and *C.elegans* (Modrich and Lahue, 1996). Alternatively, strand discontinuity occurred during DNA replication is thought to be the signal which directs the MMR system to the newly synthesized strand (Holmes et al., 1990). PCNA provides a molecular linker between mismatch repair proteins and DNA polymerase at the replication fork, so that the nicks in the nascent strand, as a result of DNA synthesis, could direct the MMR process into the newly synthesized DNA (Gu et al., 1998, Johnson et al., 1996, Kokoska et al., 1999, Umar et al., 1996).

The MSI exhibited by *Dnmt1*-deficient (Dnmt1-KO) ES cells is about 5 fold higher than methylation proficient ES cells. In contrast, the MSI exhibited by *Msh2*-deficient ES cells is elevated nearly 4 orders of magnitude (Abuin et al., 2000). This result suggests that *Dnmt1* may not provide a vital strand distinguishing signal for MMR system. However, it cannot be ruled out that Dnmt1 may facilitate the strand distinguishing process in higher eukaryotes. In fact, CpG hemimethylation was shown to synergize with single strand nicks in directing

repair to the unmethylated strand in monkey CV1 cells (Hare and Taylor, 1985). It is worthy to point out that the Dnmt1-KO allele generated in this study appeared to be a hypomorph allele. The moderate microsatellite instability exhibited in the Dnmt1-KO cells compared to that in the MMR deficient cells may reflect the remaining DNA methylation activity. To better address this question, it is necessary to measure the rate of MSI in *Dnmt1* null cells.

6.3.5 Blm affects DNA mismatch repair

It was observed in this study that *Blm* deficiency caused mild microsatellite instability. The *Blm* gene was also identified in the *C.elegans* slippage screen for the reversal of a *lacZ* slippage reporter that has a single nucleotide repeats $(A)_{17}$ (Pothof et al., 2003). In this screen 61 genes were identified showing increased genomic instability. To test if the reversal of the *lacZ* reporter depended on the existence of the $(A)_{17}$ repeats, the reversal of an out of frame *lacZ* reporter lacking the single nucleotide repeat was investigated and this experiments revealed that the majority of the genes identified in the slippage screen didn't require the existence of the $(A)_{17}$ repeats to reverse the expression of the *lacZ* reporter. *Blm* is among these genes. It was proposed that reversal of the *lacZ* reporter expression in these mutants was caused by other mechanisms other than MMR, for example, insertion or deletion mutations caused by illegitimate homologous recombination (Pothof et al., 2003). In our study, the sequence analysis of the slippage puromycin resistance clones confirmed the insertion or deletion mutations occurred in the $(CA)_{17}$ repeats. It doesn't appear to be a result of homologous recombination. Thus, MSI in *Blm* deficient ES cells requires further explanation. The human BLM protein associates with the MMR proteins, MSH2, MSH6 and MLH1 in a protein complex, including Brca1 (Wang et al., 2001). Although a functional role of this interaction is not clear, the existence of a physical interaction between BLM and the MMR proteins may indicate a functional connection. In our study, *Dnmt1* and *Blm* double null ES cells (the gene trap *Dnmt1* mutant) exhibited a MSI rate similar to the MSI rated exhibited

in the *Blm* null ES cells, which suggested a genetic interaction between *Blm* and *Dnmt1* (or *Dnmt1* mediated MMR process). To better understand this interaction, it will be important to compare the MSI rate in the MMR deficient and MMR and *Blm* double deficient cells. DNA replication, methylation and repair are coordinated processes, in which Blm, Dnmt1 and MMR proteins play major roles (El-Osta, 2003). Although an interaction between Blm and MMR hasn't been reported, it was observed that deficiency in the MMR process caused an increase in DNA methylation, which may result from the unleashed Dnmt1 activity from MMR related processes (Ahuja et al., 1997). The genetic interaction between Blm, Dnmt1 (MMR) may be a result of the unbalanced DNA replication and the MMR processes.