1.1 The mouse as a genetic model

Mice are similar to humans in both anatomy and physiology. As a mammalian model system, the mouse has advantages of a small body size and short generation time. In addition to studies of basic biological processes such as DNA metabolism, mice have served a model for studying mammalians aspects of development, immunology and behavior.

Mice and humans diverged from a common ancestor about 65 millions years ago. Mice have a genome of $2.5x10^9$ bases, which is 14% smaller than the genome of humans, $2.9x\ 10^9$ bases. 99% of human genes are represented by an identifiable mouse homologue, and 80% of mouse genes have a single human orthologue. More than 90% of the mouse and human genomes can be clustered into chromosomal segments of conserved synteny, reflecting the conservation of gene organization (Waterston et al., 2002). Based on the analysis of 67,000 mouse cDNA sequences and the comparative study of the human and mouse genomes, it is predicted that the mouse and human genomes contain about 30,000 protein coding cDNA and 15,000 non-coding cDNA including alternative spliced products (Okazaki et al., 2002).

The wild spread use of the mouse for biomedical research is due to the development of many genetic and genomic tools. One of the landmarks in mouse genetics was the isolation of pluripotent mouse embryonic stem cells (ES) (Evans and Kaufman, 1981) and the demonstration that cultured ES cells can be reintroduced into host blastocysts and repopulate somatic tissues as well as germ line during embryogenesis (Bradley et al., 1984). Importantly, cultured ES cells maintain their pluripotency after modification of their genome enabling these modifications to be established in mice (Robertson et al., 1986). Further more, targeted mutations could be introduced into ES cells through homologous recombination (Capecchi, 1989). These findings initiated a new era in mouse genetics where precise loss or gain of function mutations can be established in

the mouse through *in vitro* manipulation of ES cells. These approaches, together with the traditional transgenic technique via zygote injection, are classified as reverse genetics (Landel et al., 1990). Many new genomic tools have been subsequently developed to help to decipher the functions coded in the mouse genome, for example balancer chromosomes, chromosome deletions and duplications (Yu and Bradley, 2001).

1.2 Mouse ES cells as a genetic tool

Whereas the mouse has unique advantages as a model for humans, it has limitations both technically and economically in studies that require large numbers of animals, for example, genetic screens. A saturating genome screen covering all 30,000 genes is extremely difficult to conduct *in vivo*. An alternative approach is to conduct assays and screens on cultured cells. Mouse ES cells offer unique advantages in cell-based screens. ES cells exhibit unlimited growth in culture, which allows genetic and molecular manipulation of the cells and the manifestation of the phenotypic consequences. In contrast to other cells lines that are capable of long periods of growth such as somatic cell lines established from tumors and transformed cell lines, ES cells more precisely reflect a normal biological and physiological status. It is notable that homologous recombination in ES cells is much more efficient than that in the somatic cell lines, which allows the use of gene-targeting approaches. More than 10,000 genes are expressed in ES cells (Sharov et al., 2003). These genes are required to elaborate the fundamental components required for a mammalian cell such as structural components and physiological systems for essential functions like metabolism, cell division and DNA repair (Sharov et al., 2003). Another important aspect is that ES cells can differentiate in a defined cell culture system. *In vitro* ES cell differentiation recapitulates many *in vivo* developmental processes such as adipogenesis, cardiogenesis, haematopoiesis, myogenesis, neurogenesis and chondrogenesis (Wobus, 2001). Thus, many developmentally regulated genes and processes can be studied in ES cell. In brief, cultured ES cells can provide

access to a significant fraction of the genes in mouse genome and can serve as a genetic tool for a large variety of *in vitro* studies.

1.3 Approaches to generate homozygous mutations in ES cells

1.3.1 Sequential gene-targeting

Loss of function mutations are generated most frequently through gene-targeting in ES cells. Gene-targeting is achieved by homologous recombination, a process in which a DNA sequence recombines with its endogenous homologous genomic locus. For gene-targeting, the vectors are built to contain a drug resistance marker along with the homologous arms so that ES cells with an integrated gene-targeting vector can be selected by drug resistance. To generate ES cells with homozygous autosomal mutations, two alleles can be disrupted sequentially by gene-targeting (te Riele et al., 1990). Abuin et al (1996) reported the construction of an ES cell line carrying homozygous mutations at two different genes by sequential gene-targeting. He used a marker-recycling method, in which the neomycin drug selection marker (*Neo*) is flanked by two *loxP* sites, so that it can be removed by Cre-mediated recombination. Then, the same gene-targeting vector could be used to target the second allele (Fig. 1-1 a). Although gene-targeting allows the generation of defined mutations precisely at any gene locus, the sequential gene-targeting method is relatively time-consuming and can only be applied on a gene-by-gene basis.

1.3.2 High concentration G418 selection

It has been known for several decades that homozygous mutated cells and/or wild type cells can be generated spontaneously from cultured mammalian somatic cells containing a heterozygous mutation, a phenomena known as loss of heterozygosity (LOH). LOH can occur by many mechanisms including regional or whole chromosome loss, mitotic recombination and gene inactivation.

Evidence suggested that the mechanism causing LOH varies in different cell lines. For example, in CHO (Chinese hamster ovary) cells, LOH occurs most commonly through loss of a whole chromosome followed by chromosomal duplication (Campbell et al., 1981, Wasmuth and Vock Hall, 1984). In murine lymphoid cell lines, mitotic recombination between homologous non-sister chromatids is believed to be the major cause of LOH (Nelson et al., 1989, Potter et al., 1987, Rajan et al., 1983). Mortensen et al (1992) explored the potential of producing homozygous gene-targeted mouse ES cells from heterozygous gene-targeted ES cells via LOH. He created four different heterozygous gene-targeted ES cell lines carrying the *Neo* cassette. By culturing these heterozygous gene-targeted ES cells in high concentrations of G418, many homozygous gene-targeted cells were recovered containing two copies of the *neo* cassette. The existence of two copies of *Neo* cassette suggests that LOH has occurred either via mitotic recombination between homologous non-sister chromatids or by chromosomal loss followed by chromosomal duplication. The LOH rate in these studies was estimated to be about $1x10^{-5}$ per locus/ cell/ generation. Compared to the sequential targeting method, this high drug concentration selection approach doesn't require two or more cycles of gene-targeting, therefore, providing an easy method for obtaining homozygous mutations in ES cells (Dufort et al., 1998, Carmeliet et al., 1996, Reaume et al., 1995). Lefebvre et al (2001) investigated the mechanism of LOH in ES cells by gene-targeting the *Neo* cassette into a hybrid ES cell line (R1 cell line) that was established from F1 hybrid embryos obtained from a cross between mice of two different inbred 129 substrains. Use of a hybrid cell line allows tracking the origin of two homologous chromosomes by analyzing polymorphic DNA markers. In this study, they targeted the *Neo* cassette into six different genomic loci on four different chromosomes, Chr 2, Chr 5, Chr10 and Chr17, and showed that all of the homozygous gene-targeted clones recovered by high concentration G418 selection had lost the heterozygosity of distant linked DNA markers, which is consistent with a mechanism of chromosomal loss and duplication. Recovery of homozygous mutants from various chromosomes implied that homozygous

mutations could be generated on a genome-wide basis through high concentrations of G418 selection.

1.3.3 Induced mitotic recombination

Mitotic recombination has been used extensively in *Drosophila* to generate "genetic mosaics", a term for an individual with cells with more than one genotype (Perrimon, 1998). Mitotic recombination is also known as somatic recombination, during which chromosomal crossover occurs between two homologous non-sister chromatids during mitosis. A crossover in the G2 phase of the cell cycle between two homologous non-sister chromatids can be segregated either in a way that the recombinant chromatids segregate to opposite poles, so they separate to two daughter cells (X-segregation) or in a way that the recombinant chromatids segregate to the same pole in the same daughter cell (Z-segregation). Genetic mosaic occurs as a result of G2 crossover followed by X-segregation, in which single allelic genetic variation is localized to two sister-chromatids in the G2 phase and segregated into the same daughter cell (Stern, 1936).

In *Drosophila* the mitotic recombination system has been combined with site-specific recombination systems, such as the FLP/FRT system. The mitotic recombination in *Drosophila* can be induced by FLP-mediated recombination between two FRT sites that have been inserted into the same genomic locus on homologous chromosomes. By controlling the expression of FLP enzyme, mitotic recombination can be induced with spatial and temporal specificity. Notably, it has been shown that after FLP-mediated homologous recombination in the G2 phase of the cell cycle, recombinant chromatids are directed to consistently segregate opposite to each other (X-segregation) (Beumer et al., 1998). Recently, Liu et al (2002) adopted the concept of mitotic recombination and demonstrated that mitotic recombination could be induced in mouse ES cells via Cre-mediated recombination between targeted *loxP* sites. The mitotic

recombination frequency varies between different genomic loci and chromosomes with a range from 10^{-5} to 10^{-2} after transient Cre expression. In Liu's study, X-segregation was also observed as the major event followed by G2 crossover compared to Z-segregation, which is consistent with the results shown by FLP/FRT induced mitotic recombination in *Drosophila*. This pioneering work signals that induced mitotic recombination will be a powerful tool in mouse genetics for generating homozygous mutations in ES cells and for mosaic analysis in mice.

1.3.4 Elevated mitotic recombination in *BLM*-deficient cells

Mitotic recombination can occur spontaneously, leading to the LOH in some cell lines, for example, in cultured murine lymphoid cell lines (Nelson et al., 1989). However, the spontaneous LOH rate is very low in normal cells and cannot be used as an efficient tool for generating homozygous mutations. Recently, it has been shown that mitotic recombination rate is increased in human and mouse cells that lack the function of a DNA helicase, BLM. This opens up the possibility of generating homozygous mutations based on the enhanced LOH rate in *BLM*-deficient cells (Fig. 1-1 d). *BLM*-deficient cells constitute the foundation of my study. I used *BLM*-deficient cells as a tool to generate homozygous recessive mutations in order to conduct a genetic screen. The current knowledge about *BLM* gene and its' functions are presented in the following sections.

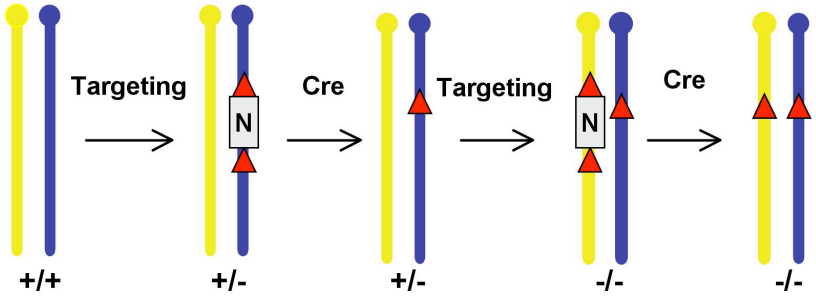

Figure 1-1 a. Sequential gene-targeting, showing the targeting of a genomic locus with *loxP* flanked *Neo* selection cassette. Cre-mediated recombination enables the construction of a marker free homozygous mutated cell. The red triangle represents the *loxP* site; "N" represents the *Neo* cassette.

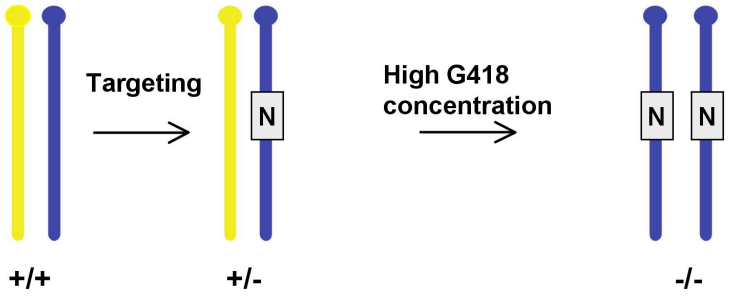

Figure 1-1 b. High concentration G418 selection, showing that homozgous mutations can be obtained by LOH involving loss and gain of whole chromosomes. "N" represents the *Neo* cassette.

Figure 1-1. Approaches to generate homozygous mutations in ES cells

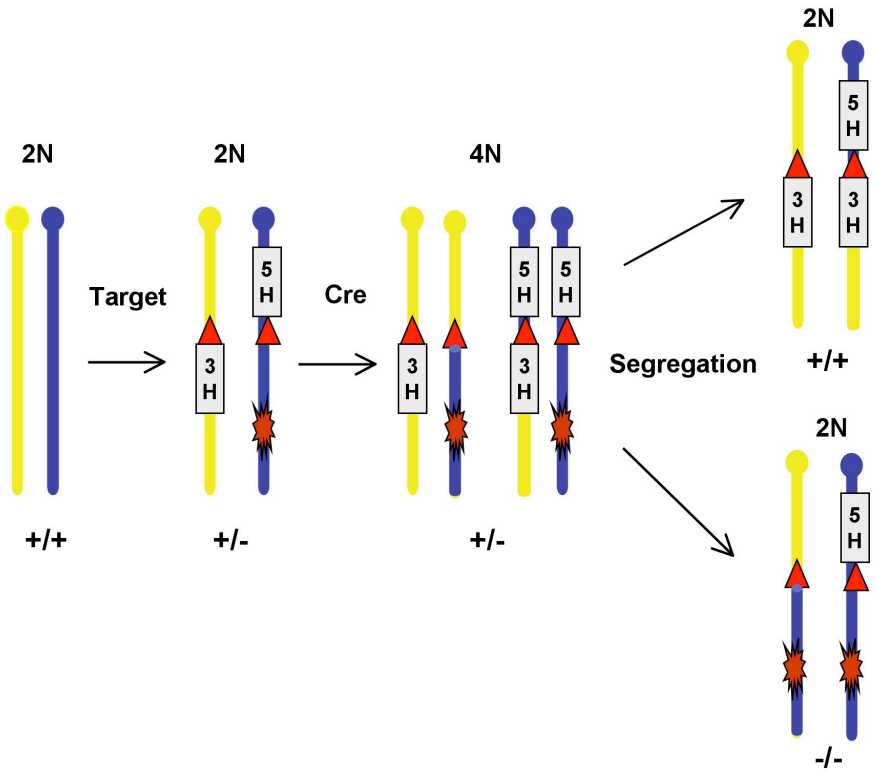

Figure 1-1 c. Induced mitotic recombination between two *loxP* sites that are targeted into two homologous chromosomes. The Cre-*loxP* mediated recombination occurs predominantly at the G2 phase of a dividing cell. "5H" represents the 5' *Hprt* cassette; 3'H represents the 3' *Hprt* cassette. The Red star represents a mutation.

Figure 1-1. Approaches to generate homozygous mutations in ES cells

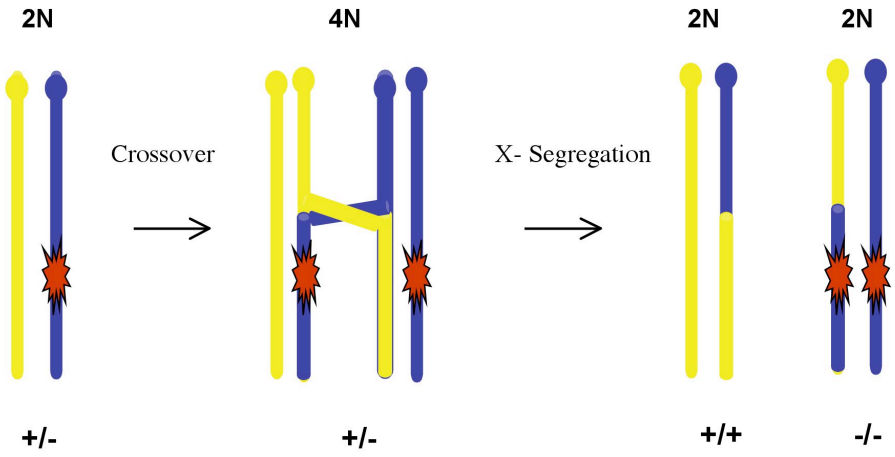

Fig. 1-1 d. *Blm*-deficient cells.

The deficiency in the *Blm* gene leads to increased LOH, which occurs predominantly via mitotic recombination. G2 cross-over between non-sister homologous chromatids followed by X-segregation results in homozygous mutant or wild-type cells.

Figure1-1. Approaches to generate homozygous mutations in ES cells

1.3.4.1 Bloom's syndrome

BLM is the gene responsible for the human disease Bloom's syndrome (BS). BS is a rare autosomal recessive transmitted disorder. German (1993) reviewed the major clinical features of Bloom's syndrome, which includes proportional dwarfism, sun-sensitive erythematous lesions of facial skin, immunodeficiency, a broad spectrum of neoplasm of multiple tissue types with early incidence and reduced fertility. Cells from BS patients exhibited excessive genomic instability including an increased spontaneous mutation rate at specific genomic loci and large microscopically visible genomic rearrangements such as chromosomal gaps, breaks and translocations. The hallmark feature of BS cells is the enormously elevated rate of sister-chromatid exchange (SCE), which is illustrated in bromodeoxyuridine (Brdu)-labeled BS cells. The increase in SCE has been shown to represent the loss of suppression of homologous recombination events in BS cells (Groden et al., 1990, Sonoda et al., 1999).

1.3.4.2 *BLM*, Bloom's syndrome gene

BLM was mapped by genetic linkage analysis to a position about 1 cM away from the gene, *FES,* on human chromsome 15 (German et al., 1994). Ellis et al (1995 a) localized and cloned the *BLM* gene using an unusual method, somatic crossover point (SCP) mapping method. The SCP mapping method was based on the observation that in some BS patients a small population of low SCE lymphocytes exists in the blood in spite of the fact that somatic cells from BS patients exhibited the characteristic high SCE rate. By examination of polymorphic DNA markers around the *BLM* locus, it was revealed that low SCE lymphocytes arose from somatic recombination within the *BLM* locus in individuals who had inherited *BLM* alleles mutated at different sites. Somatic recombination in such compound heterozygotes may reconstitute a functional *BLM* gene. Lymphocytes derived from stem cells which have undergone this

recombination event will then show a low SCE rate (Ellis et al., 1995). Therefore, the position of *BLM* gene can be precisely mapped by identification of the somatic crossover site in low SCE lymphocyte cell lines through DNA polymorphic marker analysis.

BLM encodes a member of the RecQ family of DExH box DNA helicases, comprising 1417 amino acids. The central region of BLM, the helicase domain, contains seven conserved motifs (Ellis et al., 1995). RecQ helicases have DNA-dependent ATPase activity and ATP-dependent DNA helicase activity with a 3' to 5' polarity (Wu et al., 2001). In general, DNA helicases are required to alter DNA topology in processes, in which a single strand DNA needs to be generated such as for DNA replication, repair and transcription. Besides mammalian BLM helicase, other members of RecQ famlily have been identified including the RecQ helicase in *E.coli*, the *BLM* homologue in yeast, fruit fly, worms and mammalian *RECQL1*, *WRN*, *RECQL4* and *RECQL5*. It is notable that *E.coli* and the unicellular eukaryotes, *Saccharomyces cerevisiae*, contain a single member of RecQ helicase, but in humans and mice multiple RecQ homologues exist. The mutations of *WRN* and *RECQL4* in humans lead to Werner's and Rothmund-Thomson syndromes respectively. In addition to the conserved helicase domain, BLM and other mammalian RecQ helicases contain extended N-terminal and C-terminal regions that vary greatly in length with a low degree of sequence conservation (Nakayama, 2002). This sequence variation in mammalian RecQ helicase implies the functional specificity of each protein.

1.3.4.3 Enzyme activity, an untypical DNA helicase

The studies of enzymatic activity of BLM helicase and its homologue in bacteria and yeast reveal that BLM can unwind a variety of DNA constructs *in vitro*. In contrast to a typical DNA helicase that binds and unwind a standard B-form DNA duplex, BLM protein does not bind or unwind blunt-ended, fully duplex DNA molecules. It prefers some atypical substrates. One of the preferred substrates of

BLM *in vitro* is a synthetic X-structured DNA molecule, which is used as a model for Holliday junction formed during homologous recombination. This finding suggests a role of BLM in homologous recombination (Karow et al., 2000). Moreover, BLM, WRN and the yeast RecQ helicase homologue, SGS1 can efficiently unwind an unusual G-G paired G-quadruplex structure *in vitro*. A G-quadruplex structure forms *in vivo* within G-rich DNA sequences, for example in G-rich telomeres and rDNA. This activity suggests a potential role of RecQ helicase in DNA replication during which the G-quadruplex formed in some specific G-rich sequence needs to be disrupted to allow the progression of DNA replication forks.

1.3.4.4 BLM in DNA replication

Evidence from biochemical and genetic studies in both prokaryotes and eukaryotes suggest that BLM is a multifunctional protein, which has a major role in DNA replication. For example, BS cells exhibit a protracted S phase and accumulate DNA replication intermediates of abnormal sizes (Lonn et al., 1990). *S.cerevisiae* or *S. pombe* RecQ helicase mutants fail to execute normal cell cycle progression following recovery from a S-phase cell cycle arrest caused by DNA replication inhibitor HU (Hydroxyurea), and these mutant strains are hypersensitive to HU (Frei and Gasser, 2000, Stewart et al., 1997). Consistent with this result, it has been shown that BS cells are hypersensitive to HU. BLM has also been shown to be a substrate of the protein kinase, ATR (ataxia telangiectasia and rad3+ related), which is activated in response to a DNA replication block. Blocking of BLM phosphorylation results in the failure of recovery from HU-induced replication inhibition (Davies et al., 2004). A role of BLM in DNA replication is also supported by the studies of the expression pattern of BLM and the cellular localization of BLM protein. BLM localizes to promyelocytic leukemia protein (PML) nuclear bodies and accumulates during the late S and G2 phase during the cell division cycle (Dutertre et al., 2000, Bischof et al., 2001). Following the inhibition of DNA replication by HU, BLM is

found to be localized to the DNA replication foci in S-phase (Sengupta et al., 2003). These results suggest that BLM is translocated to arrested replication forks to assist the progression of DNA replication.

The exact role of BLM in DNA replication is unclear. BLM may act in two non-exclusive processes. One possible role of BLM is to function as a "roadblock remover", in which BLM removes unusual DNA secondary structures, such as the G-Quadruplex or other obstacles, to prevent potential replication stalling or the collapse of the replication fork. This view is supported by the observation that BLM protein can promote branch migration of a Holiday junction (Karow et al., 2000, Hickson, 2003). Another possible role is that BLM is involved in re-starting DNA replication after the demise of a replication fork, a process involving homologous recombination-mediated double strand break (DSB) repair. It has been shown that BLM interacts with proteins required for DNA replication, for example, RPA (replication protein A), the major single strand DNA binding protein which is required in DNA replication and recombination (Brosh et al., 2000). BLM forms a complex with RAD51. RAD51 catalyzes DNA strand invasion and exchange in homologous recombination (Wu et al., 2001).

1.3.4.5 A model for sister-chromatid exchange caused by *BLM*-deficiency

The characteristic phenotypic consequence of BLM mutation is excessive somatic recombination and SCE. It has been shown that the occurrence of SCE requires the homologous recombination system. Cells deficient in homologous recombination protein RAD51 and RAD54 exhibited a significantly reduced level of SCE (Sonoda et al., 1999). A model involving DNA replication and homologous recombination has been proposed to explain why *BLM*-deficiency during DNA replication will lead to SCE (Nakayama, 2002). In brief, DNA strand breaks or gaps may exist under various physiological conditions. When the replication fork encounters a single-strand nick or a gap on the template strand, it will collapse and a double strand break is created. Then, a repair process is

initiated, leading to the formation of a Holiday junction through homology-directed strand invasion and exchange, which is a process mediated by the homologous recombination machinery. The Holliday junction can be simply unwound by a RecQ helicase (BLM in mammals) to re-establish the replication fork (Karow et al., 2000). In this case, the repair process is error-free and no SCE is generated. In the absence of the RecQ helicase, the Holliday junction may be resolved by recombination pathways that cause chromatid exchanges.

1.3.4.6 Proteins interacting with BLM

The proteins interacting with BLM may provide additional clues to the functions of BLM. BLM has been shown to direct interact or co-localize with many proteins. Topoisomerase IIIα is one of the BLM interacting proteins. This interaction is a direct one and exists in both prokaryotes and eukaryotes. The function of this interaction is not clear yet. It has been suggested that topoisomerase IIIα is required for resolving abnormal recombination intermediates (Wu et al., 1999). BLM also associates with RAD51 and RPA, which is consistent with the role of BLM in DNA replication (discussed above).

Recently, BLM has been found to be a component of a large protein complex including BRCA1, which is referred to as BASC (BRCA1-associated genome surveillance complex) (Wang et al., 2000). This complex includes many proteins involved in DNA repair or DNA damage response such as MSH2, MSH6, MLH1, ATM, BLM, the RAD50-MRE11-NBS complex and DNA replication factor C (RFC), a protein complex that facilitates the loading of PCNA (proliferating cell nuclear antigen) onto DNA. MSH2, MSH6 and MLH1 are major components of the DNA mismatch repair (MMR) system. MMR system removes mismatched nucleotides generated during DNA duplication. MMR also plays a role in the genome surveillance process, in which certain types of DNA lesion are recognized and signal cell death or cell cycle arrest. ATM (Ataxia-telangiectasia mutated) is a serine/threonine protein kinase that plays a central role in sensing

and transducing cellular signals in response to DNA damage. The RAD50-MRE11-NBS (RMN) complex and BRCA1 are critical in repairing DNA double strand breaks (DSB) via homologous recombination. The precise roles of BASC complex haven't been established yet. The existence of multiple proteins involved in DNA damage repair and signaling processes suggest that BASC complex plays a role in the DNA damage response. Current studies suggest that the function of BRCA1 in DSB repair doesn't require BLM. BS cells don't exhibit obvious sensitivity to γ-irradiation that induces DSB, while the deficiency of BRCA1 leads to a hypersensitive γ-irradiation response (Franchitto and Pichierri, 2002).

p53 has also been reported to be a binding partner of BLM and another mammalian RecQ helicase WRN (Blander et al., 1999, Spillare et al., 1999, Wang et al., 2001). p53 is a transcription factor that plays a central role in cell cycle arrest and apoptosis (Oren and Rotter, 1999). However, the interaction between BLM and p53 is not required in either cell cycle arrest or apoptosis, the majority of BS cells appear to have normal p53 accumulation and undergo cell cycle arrest and apoptosis in response to certain type of DNA damage (Lu and Lane, 1993, Ababou et al., 2002). In contrast, p53 may have a role in repairing stalled replication forks, a process involving BLM. This view is based on a recent finding that p53 modulates the frequency of homologous recombination and SCE in *BLM*-deficient cells (Sengupta et al., 2003). In this study, it was reported that p53, BLM and RAD51 co-localized to sites of stalled DNA replication forks in response to DNA replication inhibition induced by HU treatment. Loss of p53 function enhanced synergistically the homologous recombination and SCE frequency in *BLM*-deficient cells derived from Bloom's syndrome patients. Consistent with this observation, it has been reported that p53 can bind to Holiday junctions and facilitates their resolution (Lee et al., 1997). p53 can modulate the procession of Holiday junctions by BLM *in vitro* (Yang et al., 2002). Mutation in p53 also results in elevated homologous recombination (Susse et al., 2000, Slebos and Taylor, 2001, Saintigny and Lopez, 2002), and this activity of

p53 appears to be independent of its transcriptional activation function (Willers et al., 2000).

1.3.4.7 Mouse models of Bloom's syndrome

The BLM homologue in mice is located on chromosome 7. The gene is approximately 88 kb in length and consists of 23 exons. The first exon is non-coding 5'UTR and is represented differently in testis and somatic cells as the result of alternative splicing (McDaniel et al., 2003). Six different *Blm* knockout alleles have been described, including $Blm^{tm1ches}$, $Blm^{tm1/Brd}$, $Blm^{tm2/Brd}$, $Blm^{tm3/Brd}$, $Blm^{tm1/Gos}$ *and* $Blm^{tm3/ches}$ (Chester et al., 1998, Luo et al., 2000, Goss et al., 2002, McDaniel et al., 2003). Four *Blm* alleles, $Blm^{tm1ches}$, $Blm^{tm1/Brd}$, $Blm^{tm1/Gos}$, $Blm^{tm3/ches}$, were generated by gene-targeting with replacement targeting vectors, resulting in deletion of coding exons. These alleles have been shown to be null by Western-blot analysis of Blm protein expression and homozygous knockout mice with these alleles appear to be embryonic lethal. *Blm*-deficiency doesn't have overt effect on the growth and survival of ES cells. Two alleles, $Blm^{tm2/Brd}$ and $Blm^{tm3/Brd}$, were generated by insertional gene-targeting events, which results in the duplication of exon3. This duplication caused a frame-shift mutation. The $Blm^{tm2/Brd}$ allele is homozygous lethal while the derived $Blm^{tm3/Brd}$ is viable. The homozygous mice ($Blm^{tm2/Brd}$, $Blm^{tm3/Brd}$) exhibited genomic instability and tumor susceptibility, a phenotype mimicking the human Bloom's syndrome. Thus $Blm^{tm3/Brd}$ mice serve as a better animal model for human Bloom's syndrome. Moreover, a significant increase in the SCE was observed in $Blm^{tm1/Brd}$ / $Blm^{tm3/Brd}$ ES cells (Luo et al., 2000). Recently, it was shown that $Blm^{tm3/Brd}$ could rescue the embryonic lethality of $Blm^{tm3/ches}$ alleles. *Blm*-deficient cells carrying Blm^{tm3Brd} and $Blm^{tm3/ches}$ alleles have been reported to have a SCE rate about two fold lower than the cells with two $Blm^{tm3/ches}$ alleles, which the authors suggest reflecting the hypomorphic activity of the Blm^{tm3Brd} allele (McDaniel et al., 2003).

1.3.4.8 Elevated LOH rate in *Blm*-deficient mouse ES cells

The direct phenotypic consequence of increased somatic recombination is loss of heterozygosity (LOH) of single allelic mutations. Luo et al (2000) determined the LOH rate in $Blm^{tm1/Brd}$ / $Blm^{tm3/Brd}$ ES cells by measuring the loss of a single copy *Hprt* minigene that was gene-targeted into an autosomal genomic locus. Cells that have lost the *Hprt* minigene become resistant to the drug, 6-thioguanine. By Luria-Delbruck fluctuation analysis the rate of LOH was determined to be 2.3X 10^{-5} (locus/cell/generation) in wild type ES cells and 4.2X10^{-4} (locus/cell/generation) in *Blm*-deficient ES cells, respectively (Luo et al., 2000). Although the *Hprt* gene can be lost by several mechanisms, for example, loss of whole chromosome, spontaneous mutation and deletion, mitotic recombination between homologous chromosomes is believed to be the major cause of LOH in *Blm*-deficient cells (Sonoda et al., 1999).

In summary, the biochemical and genetic studies point out that BLM plays a critical role in repairing DNA replication fork abnormalities. BLM facilitates the smooth progressing of DNA replication by preventing stall of the replication forks or facilitating stalled DNA replication forks to restart in an error-free way. The increased rate of SCE and mitotic recombination exhibited in *BLM*-deficient cells is the result of the switching from BLM-dependent error-free repair to BLM-independent error-prone repair of the stalled replication forks. BLM is not required for cell growth or survival in culture. *Blm*-deficiency in mouse ES cells caused a 20-fold increase in the rate of LOH, which provides the basis for deriving homozygous autosomal mutations from a single allele mutation.

1.4 Recessive genetic screens in mammalian cells

Recessive genetic screens in a diploid genome require a strategy to generate homozygous mutations. In the early 70's, it has been shown that recessive mutations could be recovered from cultured mammalian cells that are partially

hemizygous (Siminovitch, 1976). The most frequently used cell line for deriving recessive mutations were CHO (Chinese hamster ovary) cells. Functional hemizygosity could be caused by several possible mechanisms, such as gene inactivation and genomic rearrangements, which results in loss of function of one copy of certain autosomal genes, therefore rendering phenotypic hemizygosity to these cells. Since genomic rearrangement and gene inactivation occur randomly, each CHO cell line may have accumulated mutations in different sets of genes (Deaven and Petersen, 1973, Worton, 1978, Gupta et al., 1978). It has been shown that multiple recessive mutations from CHO-CHO hybrids segregated randomly, suggesting that the functional hemizygosity in CHO cells is not only restricted to one or a few chromosomal regions but appears to be wildly distributed (Gupta, 1980). In the early 80's, CHO cells were used to isolate recessive mutants that are sensitive to killing by ultraviolet radiation (UV) (Thompson et al., 1980, Busch et al., 1980). 44 UV sensitive mutant clones were classified into 4 different complementation groups (Thompson et al., 1981). The genes mutated in the second complementation group were determined by a genetic rescue method using cloned human genomic DNA, which led to the identification of the important nucleotide excision repair gene, *ERCC1* (excision repair complementing defective repair in Chinese hamster cells) (Westerveld et al., 1984). These data demonstrated the application of genetic aneuploidy in a recessive genetic screen. However, this strategy is greatly restricted by the fact that CHO cell lines contain partial functional hemizygous genomes. Therefore, recessive mutations located in the functional diploid regions can't be recovered. In this regard, *Blm*-deficiency has an overt advantage. Since homozygous mutations in *Blm*-deficient cells are generated preferentially by mitotic recombination, in principle all genes on an autosomal chromosome can be accessed. Therefore, *Blm*-deficient cells will allow broader genome coverage in a genetic screen than CHO cells.

1.5 DNA mismatch repair

The proper functions of DNA metabolism processes including DNA replication and DNA repair are crucial for the integrity of the genetic materials. The integrity of DNA is constantly challenged by many environmental or physiological factors. Accordingly, many proteins and DNA repair systems have been identified acting coordinately to prevent and eliminate the errors in DNA. The DNA mismatch repair (MMR) system plays a critical role in guarding genome integrity in virtually all organisms from bacteria to human. The primary function of MMR system is to recognize and correct base-base mismatches and small insertion and deletion (I/D) loops that arise during DNA replication (replicative mismatch repair). Defects in the MMR system will result in an elevated spontaneous mutation rate, a mutator phenotype and expansion or deletion of simple repeat sequences, known as microsatellite instability (MSI). The importance of MMR in guarding genome stability has been highlighted by the association of defects in MMR with cancer. Besides its function in repairing DNA replication errors, the MMR system has been linked to general DNA recombination processes including meiosis and homologous recombination. Moreover, a role of MMR in processing chemically damaged DNA, also known as DNA damage surveillance has been documented. The MMR system has been extensively reviewed by others (Modrich, 1991, Modrich and Lahue, 1996, Buermeyer et al., 1999, Hsieh, 2001). In this section I would like to provide an overview of the basics of the MMR system and emphasize the functions of eukaryotic MMR.

1.5.1 DNA mismatch repair in bacteria

The first studies of the MMR system started more than three decades ago in bacteria when genetic screens were conducted to isolate mutants with elevated spontaneous mutation rates. This research resulted in the identification of four central components of MMR system, MutS, MutL, MutH, and MutU. MMR in bacteria has been most thoroughly investigated and serves as the model for other organisms (Modrich, 1991). For the simplicity, it is separated into three major steps (Fig. 1-2): Step 1: Mismatch recognition, in which MutS proteins form

a homodimer complex which binds to the mismatched nucleotides. Step 2: Strand discrimination and excision. In an ATP dependent manner, the MutS homodimer complexes with dimerized MutL protein and stimulates the endonuclease activity of MutH. Consequently, a nick is generated in the newly synthesized DNA strand by the activated MutH using the semi-methylated GATC as the strand discrimination signal. Then, MutU, a DNA helicase, is loaded to the MutH induced nick to unwind the duplex DNA molecule. Then, with the help of exonucleases, the newly synthesized DNA strand containing the mismatched nucleotide is removed to leave a single strand DNA gap. Step 3: Resynthesis and ligation. Single-strand DNA binding protein (SSB), DNA polymerase III and DNA ligase are required for the resynthesis and ligation, which fills in the gap created by strand excision.

1.5.2 DNA mismatch repair in eukaryotes

The MMR system has been highly conserved throughout evolution. Compared to the MMR system in bacteria, MMR systems in higher eukaryotes have evolved more specificity and functions, reflected by the existence of multiple MutS and MuL homologues in yeast and mammals (Table 1-1). In yeast and mammals, mismatch recognition is conducted by three MutS homologues, MSH2, MSH3 and MSH6. MSH2 can complex with either MSH6 or MSH3, forming two protein heterodimers, MutSα (MSH2/MSH6) or MutSβ (MSH2/MSH3). MutSα and MutSβ exhibit different binding preferences for DNA substrates. MutSα predominantly binds single base mismatches and single insertion/deletion loops while MutSβ binds single and larger insertion and deletion loops (Fig. 1-3 a) (Acharya et al., 1996, Drummond et al., 1995). The function of MutSα and MutSβ overlaps with respect to the recognition of small insertion/deletion mismatches. Consistent with the role of MSH2 in both complexes, MSH2 mutations cause the highest level of mutator and MSI phenotypes, while MSH6 and MSH3 mutants exhibit mild or modest ones (Fig. 1-3 b). The functional homologue of bacterial MutL in yeast and humans is MLH1 (MutL homologue). In mammals, MLH1 forms a heterodimer protein complex with

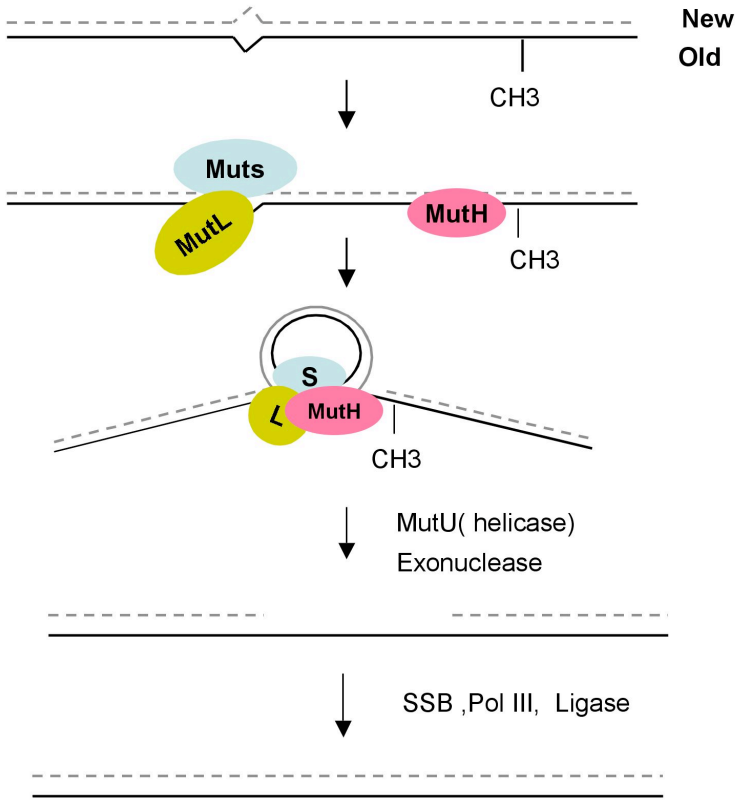

Figure 1-2. DNA mismatch repair in *E.coli*

DNA Mismatch repair (MMR) initiates when a homodimer of MutS protein recognizes mismatched nucleotides. Then, the MutS homodimer complexes with a homodimer of MutL and stimulates the endonuclease activity of MutH. A nick is made in the newly synthesized DNA strand by activated MutH using the semi-methylated GATC as the strand discrimination signal. Then, the MutU DNA helicase is loaded to the MutH induced nick to unwind the duplex DNA molecule. With the help of exonucleases, the newly synthesized DNA strand with the mismatched nucleotides is removed to leave a single strand DNA gap. Finally, the single strand gap is filled in with the help of single-strand DNA binding protein (SSB), DNA polymerase III and DNA ligase.

E.coli	*S.cerevisiae*	Human/mouse
MutS	*MSH1*	-
	MSH2	*MSH2*
	MSH3	*MSH3*
	MSH6	*MSH6*
	MSH4	*MSH4*
	MSH5	*MSH5*
MutL	*MLH1*	*MLH1*
	PMS1	*PMS2*
	MLH2	*PMS1*
	MLH3	*MLH3*
MutH	-	-
MutU(UvrD)	-	-

Table 1-1. Homologues of bacterial MMR genes in *S.cerevisiae* and mammals.

Multiple *MutS and MutL* homologs exist in eukaryotes.

Mammalian *MSH2, MSH3, MSH6, MLH1, PMS2* (yeast *PMS1*), *PMS1*(yeast *MLH2*) are involved in replication repair.

MSH4, MSH5, MLH1 and *MLH3* function in meiotic processes.

MSH2 has a role in homologous recombination.

MSH2, MSH3, MSH6, MLH1,PMS2 are invovled in DNA damage surveillance. *MSH4* and *MSH5* also exhibit a minor role.

MSH1 in *S. cerevisiae* is required for normal mitochondria function in yeast. A mammalian homologue of *MSH1* has not been identified.

Mut H and *MutU* homologues have not been identified in eukaryotes.

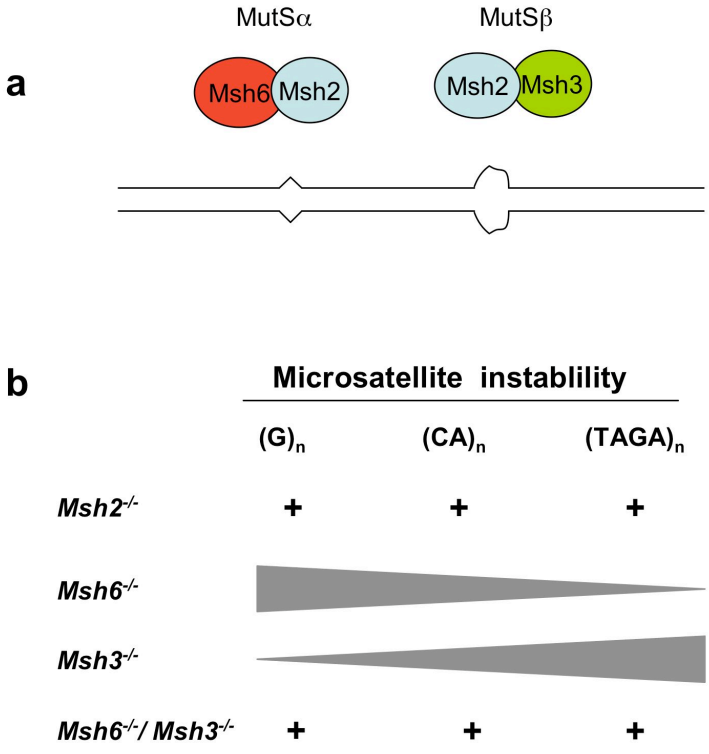

Figure 1-3. DNA mismatch recognition in Eukaryotes

a. Three MutS homologues, Msh2, Msh3 and Msh6 are involved in mismatch recognition in eukaryotes. They form two protein complexes, MutSα (Msh2/Msh6) and MutSβ (Msh2/Msh3). MutSα and MutSβ exhibit different mismatch binding abilities. MutSα predominantly binds single base mismatches and single insertion/deletion loops while MutSβ binds single and larger insertion and deletion loops. **b**. Defects in different MMR recognition proteins results in a different MSI phenotype. *Msh2* deficiency and *Msh3/Msh6* double mutants cause MSI in single nucleotide repeats ($(G)_n$), dinucleotide repeats ($(CA)_n$) and tetranucleotide repeats ($(TAGA)_n$), whereas *Msh6* or *Msh3* single mutants exhibit mild or modest MSI that varies depending on the repeat unit length. The gradient triangles represent the variation of the MSI activity.

PMS2 (post-meiotic segregation 2), which is designated as MutLα. MutLα complex binds to either MutSα or MutSβ protein complex (Li and Modrich, 1995). It has been shown that MLH1 plays a central role in MMR similar to MSH2. Mutations of either MSH2 or MLH1 fully abolish mismatch repair (Prolla et al., 1998). Besides mammalian PMS2, MLH1 can also complex with PMS1 and MLH3. These protein complexes appear to have a minor role in DNA mismatch repair. Mutations of PMS1or MLH3 result in less severe mutator or MSI phenotypes compared with MLH1 (Papadopoulos et al., 1994, Lipkin et al., 2000). In yeast, the functional homologue of MutLα is composed of Mlh1 and Pms1 (Prolla et al., 1994). Yeast Mlh1 can also complex with Mlh2, Mlh3. Both complexes have been shown to have a role in inhibition of mutation of simple sequence repeats (Harfe and Jinks-Robertson, 2000, Flores-Rozas and Kolodner, 1998).

Similar to the bacteria MMR system, mismatch repair in eukaryotes is also directed to the newly synthesized DNA strand. However, a functional MutH homologue has not been identified in either yeast or mammals. Instead, strand discontinuity has been hypothesized to serve as the strand discrimination signal in eukaryotes. Nicks or gaps that exist between neighboring Okazaki fragments in the nascent DNA strand during DNA replication could direct MMR to the newly synthesized strand. This hypothesis is supported by an *in vitro* mismatch repair experiment in which MMR process was directed by nicks situated in the proximity of mismatched nucleotides and this substrate can be efficiently corrected in a directional manner in extracts of *E.coli* mutH mutant (Modrich, 1997). Another hypothesis is that DNA mismatch repair is directly coupled to the DNA replication fork by DNA polymerase associated factor, PCNA (proliferating cell nuclear antigen), therefore alleviating the necessity of a MutH homologue in eukaryotes. It has been showed that PCNA associates with eukaryotic MutS and MutL protein complexes and mutation of PCNA causes a mutator and MSI phenotype in yeast (Gu et al., 1998, Johnson et al., 1996, Umar et al., 1996, Kokoska et al., 1999).

Several DNA nucleases are implicated in eukaryotic MMR, which may act in excision of the mismatched nucleotides. The major player is exonuclease 1 (EXO1). EXO1 protein has 5' to 3' exonuclease activity and it interacts with MSH2 in yeast and mammals (Tishkoff et al., 1997a). Mutation of *Exo*1 causes an increased spontaneous mutation rate in yeast (Amin et al., 2001). *Exo1* knockout mice were created recently, and *Exo1* null cells have a MMR deficiency with an increased spontaneous rate and microsatellite instability (Wei et al., 2003). Recently it was shown in an *in vitro* MMR reconstitution experiment that hRPA (human replication protein A), a single strand DNA binding protein, plays multiple roles in MMR, protecting the template DNA strand from degradation *in vitro,* enhancing DNA excision by Exo1 and facilitating the repair synthesis (Ramilo et al., 2002, Genschel and Modrich, 2003).

In yeast, Rad27 (the yeast flap exonuclease homologue (FEN1)), DNA polymerase delta and DNA plolymerase zeta were thought to play a role in DNA mismatch repair based on the mutator phenotypes of mutations in these components (Kolodner and Marsischky, 1999). However, it was shown recently that FEN1 plays a critical role in processing Okazaki fragments and homologous recombination mediated DNA repair processes. Thus, the mutator phenotype exhibited in yeast rad27 mutants and mammalian FEN1 mutants may be an indirect result of abnormalities in DNA replication, which argues against a direct role of Rad27/FEN1 in MMR (Tishkoff et al., 1997b, Ruggiero and Topal, 2004, Liu et al., 2004). Until now the function of DNA polymerase delta and DNA polymerase zeta in mammalian DNA mismatch repair has not been reported.

1.5.3 MMR in homologous recombination

The MMR system also plays a role in DNA recombination. DNA recombination involves annealing of complementary DNA strands, which often will contain imperfectly matched sequences. These strands form heteroduplex DNA intermediates, which are the substrates for MMR. In yeast, it has been shown that

the MMR system can repair the mismatched nucleotides in heteroduplex recombination intermediates. Mutations in yeast *MSH2, MLH1* and *PMS1* (post meiotic segregation 1) genes caused an increase in post meiotic segregation, which is the result of lack of repair of the heteroduplex intermediates generated during the first mitotic division following meiosis (Alani et al., 1994, Prolla et al., 1994).

Studies in bacteria, yeast and mammals have all revealed that the MMR system acts as a barrier to homologous recombination, in which the binding of MMR to the heteroduplex intermediates elicits a yet unclear downstream process that prevents the occurrence or the progression of homologous recombination between diverged DNA sequences. It has long been known that homologous genes in two closely related bacteria, *Escherichia coli* and *Salmonella typhimurium*, generally will not recombine, although their nucleotide sequences are 80% identical. Mutations in mutH, mutL, mutS and mutU result in a 50 to 3000-fold increase in such interspecies recombination (Rayssiguier et al., 1989). Recombination between two 405 bp substrates in *E.coli* is reduced 240 fold when the sequence homology was decreased from 100% to 89%. While in a MutS deficient strain, the decrease was only about 9 fold (Shen and Huang, 1989). The role of yeast *MSH2, MSH3, MSH6, MLH1* and *PMS1* on homologous recombination have been tested in mitotic recombination assays, in which a homologous recombination event was required to reconstitute a functional selection marker gene on a yeast chromosome. These experiments revealed that mutations in *MSH2* significantly increased homologous recombination between diverged DNA sequences. However, mutation of *MSH3, MSH6, MLH1* or *PMS2* exhibited a modest or minor effect in this assay (Selva et al., 1995, Selva et al., 1997, Nicholson et al., 2000). This result suggests that mismatch recognition protein complexes involving *MSH2* play an important role in recombination between diverged sequences. Consistent with this result, *MSH2*-deficient mouse cells exhibited a significant increase in homologous recombination between diverged DNA sequences, while *MSH3*-deficient cells exhibited a minor effect (de Wind et al., 1995, Abuin et al., 2000). The effects of *Mlh1, Msh6* and

other mammalian *MutS* and *MutL* homologs on homologous recombination between diverged sequences have not been directly examined.

1.5.4 DNA mismatch repair in meiosis

In eukaryotic meiosis, each pair of homologous chromosomes physically interacts and forms chromosomal crossovers as a result of homologous recombination. The connection of the aligned homologous chromosomes can be visualized with an eletron microscope as discernable structure called synaptonemal complex. Two MutS homologues, MSH4 and MSH5 play a role in meiosis. In yeast, Msh4 and Msh5 form a heterodimer protein complex. Mutation of either *MSH4* or *MSH5* gene causes a reduction in meiotic crossover and increased levels of meiosis I chromosome nondisjunciton. *msh4* and *msh5* mutant strains display normal DNA mismatch repair function, suggesting they are not involved in replicative DNA repair (Ross-Macdonald and Roeder, 1994, Hollingsworth et al., 1995). Consistent with this observation, Mammalian homologues of MSH4 and MSH5 exhibit the same effect on meiosis. Human MSH4 and MSH5 form a heterodimer (Bocker et al., 1999). Mice lacking Msh4 or Msh5 are sterile in both males and females, and show abnormalities in chromosome pairing and synapsis at the meiosis prophase 1 (Edelmann et al., 1999, Kneitz et al., 2000). The major mismatch recognition protein in replication repair, Msh2, Msh3, and Msh6, are not involved in meiosis (de Wind et al., 1995, Edelmann et al., 1997). In contrast, Mlh1 acts in both replication repair and meiosis. Mutation in *MLH1* gene caused reduced meiotic crossovers in yeast. In mice, *Mlh1*-deficiency leads to sterility in both male and females (Hunter and Borts, 1997). In addition, *Pms2*-deficiency causes infertility in male mice with abnormal chromosome synapsis, suggesting a role of mammalian Pms2 in meiosis (Baker et al., 1995). Recently, it was shown that the eukaryotic MutL homologue, Mlh3, possesses a distinct function in meiosis. *Mlh3* mutant mice are viable but sterile with reduced meiosis crossovers and a meiotic block. Mlh3 protein is required for Mlh1 binding to meiotic chromosomes and is found to localize to meiotic chromosomes. *Mlh3* mutation in mice doesn't cause discernable

microsatellite instability (Lipkin et al., 2002). The exact role of the MMR proteins in meiosis is still unclear.

1.5.5 MMR in DNA damage surveillance

1.5.5.1 MMR deficiency causes DNA methylation damage tolerance

Studies of MMR deficient cell lines have identified altered responses to DNA methylation damages. The MMR system appears to recognize DNA damage and trigger downstream cell cycle arrest (G2/M) and apoptotic cell death. This function prevents the accumulation of mutagenic DNA lesions and is therefore called MMR-mediated DNA damage surveillance. The function of MMR in DNA damage surveillance was first reported in bacteria. The hypersensitivity of *dam*$^-$ bacteria to simple methylating agents, such as methyl-nitrosourea (MNU) and N-methy-N'-nitro-N-nitrosoguanidine (MNNG) could be rescued by additional mutation in mutS or mutL (Karran and Marinus, 1982). Later, it was demonstrated that cell lines which were tolerant to DNA alkylating agents, such as MNNG, were deficient in mismatch recognition *in vitro*, which implies a link between eukaryote MMR system and DNA methylation damage (Kat et al., 1993, Branch et al., 1993). Clear evidence of a link between MMR and methylation damage came from two human cell lines with mutations in MMR genes, *MLH1* or *MSH2*. The human colorectal adenocarcinoma cell line, HCT116, has a MLH1 mutation and displays microstatellite instability and tolerance to MNNG. Transfer of chromosome 3 that contains the *MLH1* gene to this cell line restored the mismatch repair activity and made the cells sensitive to MNNG (Boyer et al., 1995, Koi et al., 1994). Similar chromosome transfer experiments confirmed that mutations in hMSH2 and hMSH6 caused a MNNG tolerance phenotype in two human endometrial adenocarcinomal cell lines (Umar et al., 1997). These observations suggest that the MMR system is required to trigger cell death in response to DNA methylation damage.

1.5.5.2 Two models for the function of MMR in DNA damage surveillance

The major cytotoxic activity of MNNG is to methylate guanine (G) at the O^6 position, generating a modified nucleotide, O^6-methyguanine (O^6-meG). O^6-meG is repaired by methyguanine methyltransferase (MGMT), which removes the methy-group from O^6-meG. The MNNG tolerance exhibited in MMR deficient cells is not a result of increased MGMT activity because in MMR deficient cells, O^6-meG persists in cells instead of being cleared by MGMT, and the cells are overloaded with G-A transitional mutations (Karran and Bignami, 1992). It was later demonstrated that O^6-meG can pair with either thymidine (T) or cytosine (C) during DNA replication and form imperfect O^6-meG/T or O^6-meG/C basepairs. Both O^6-meG/T and O^6-meG/C basepairs can be bound by the mismatch repair recognition protein complex, MutSα, in MMR proficient cell extracts but not in MNNG tolerant cell extracts (Griffin et al., 1994, Duckett et al., 1996). This data suggests that the mismatch binding ability of MMR proteins is involved in the cytotoxic pathway of MNNG. The exact mechanistic link between MMR deficiency and DNA methylation damage tolerance has not been fully established. Two models have been proposed. In one model, the binding of the mismatched nucleotides and the subsequent repair are thought to be essential. During the MMR process, the newly synthesized DNA strand containing the mismatched T of the O^6-meG/T base pair is removed by DNA exonuclease. However, a thymidine will again be incorporated and pairs with O^6-meG, which will initiate another round of mismatch repair. This “futile” repair process could stall DNA replication and create double strand breaks, both of which may serve as a signal for cell cycle arrest and cell death. It has been shown that MNNG could only trigger apoptosis in dividing cells and in these cells the apoptosis was preceded by a wave of DNA double strand breaks (Roos et al., 2004). Giving the established function of MMR in repairing mismatched nucleotides, this model provides a simple explanation for the DNA methylation tolerance.

In the other model, it was proposed that MMR components serve as a general DNA damage sensor. The binding of MMR proteins to damaged DNA could trigger a downstream signaling cascade that signals cell death and cell cycle arrest. In this model, DNA mismatch repair is not required and thus the MMR system is expected to be able to sense a broad spectrum of DNA damage besides DNA methylation (Fink et al., 1998, Karran, 2001). Indeed, it has been shown that tumor cells lines with defects in MSH2, or MLH1 exhibited modest but significant tolerance to many chemotherapeutic drugs, which induce various types of DNA damage, for example, Cisplatin and Doxorubincin. Cisplatin forms bulky intra or inter DNA strand crosslinks and Doxorubincin is a DNA topoisomerase inhibitor. MutSα protein complex is able to bind to the DNA lesion caused by cisplatin (Aebi et al., 1996). The depletion of DNA topoisomerase activity by topoisomerase inhibitors will stall DNA replication, and the arrested DNA replication may trigger the MMR system-mediated cell cycle arrest and cell death pathways (Fedier et al., 2001). Consistent with the general DNA damage sensor model, it was shown that overexpression of human *MSH2* or *MLH* genes can trigger apoptosis in either mismatch repair-proficient or -deficient cells without DNA damage (Zhang et al., 1999).

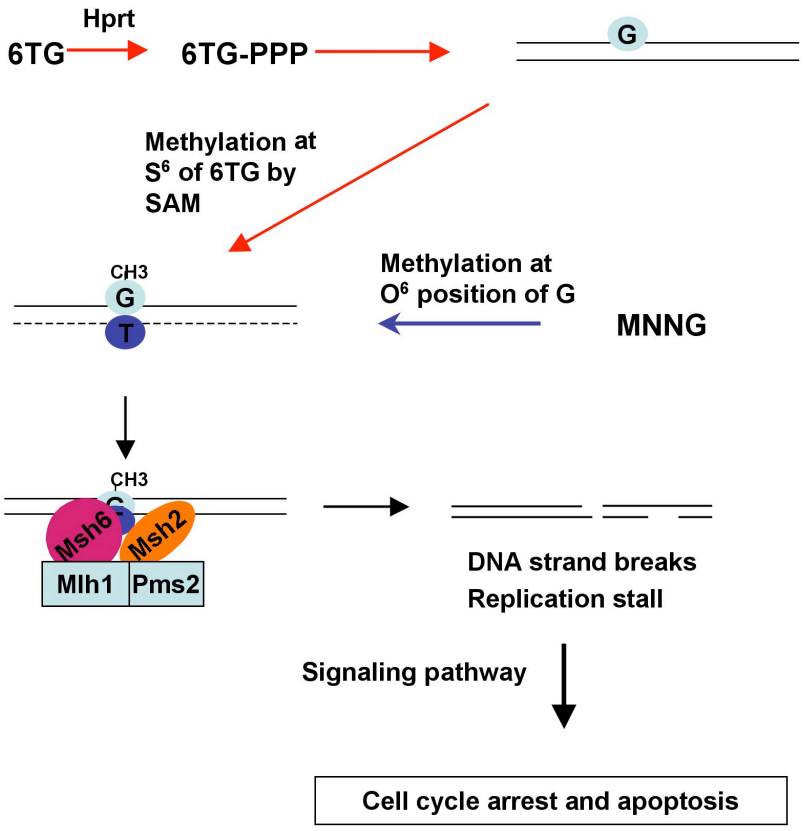

Figure 1-4. DNA mismatch repair in MNNG and 6TG cytotoxity

The cytotoxic mechanism of 6TG is initiated in cells by Hprt (hypoxanthine-guanine phosphoribosyltransferase), which converts 6TG to 2-deoxy-6-thioguanosine triphosphate (6TG-PPP). 6TG-PPP is incorporated into DNA as a guanine analogue during DNA synthesis. S-adenosylmethionine (SAM) methylates 6TG at the S^6 position to form S^6-methylthioguanine (S^6-mG). S^6-mG pairs with T during DNA replication, forming S^6-mG/T mismatched basepairs . MNNG methylates guanine (G) at the O^6 position in DNA to form O^6-mG , which also forms an imperfect basepair with T. The binding of the MutSα to S^6-mG/T or O^6-mG initiates multiple cycles of mismatch repair that causes single and double strand DNA breaks. The DNA strand breaks and the subsequent DNA replication stall signals cell cycle arrest and cell death.

1.5.5.3 MMR deficiency causes tolerance to 6-thioguanine (6TG)

6TG has long been used as a purine anti-metabolite drug in the treatment of acute leukemia (Elion, 1989). The cytotoxity of 6TG requires the activity of hypoxanthine-guanine phosphoribosyl transferase (HPRT), which transfers the sugar phosphate group to 6TG to form 2'-deoxy-6-thioguanosine-triphosphate, the active guanine nucleotide analogue in DNA synthesis. Cells that lack the *HPRT* gene are fully resistant to 6-TG killing. In the mid 90's, it was found that 6TG cytotoxity requires MMR activity. Cells with defects in *MSH2, MSH6, or MLH1* genes are tolerant to 6TG (Aebi et al., 1997). It is believed that the cytotoxic mechanism of 6TG is similar to MNNG. Both drugs show delayed cytotoxity and elevated SCE. Notably, MMR deficient cells that are tolerant to MNNG are also tolerant to 6-TG (Tidd and Paterson, 1974). The direct link betwwn MMR and 6TG cytotoxity was established by two studies. One shows that 2'-deoxy-6-thioguanosine-triphosphate in DNA can be methylated by S-adenosylmethionine (SAM) to form S^6-methylthioguanine (S^6-mG). S^6-mG can pair with either thymidine (T) or cytosine(C) in the growing DNA strand (Swann et al., 1996) and the S^6-mG/T basepair is the binding substrate of mismatch repair complex MutSα (Waters and Swann, 1997). Based on the structural similarity of O^6-methyguanine (O^6-meG, generated by MNNG) and S^6-thioguanine, it is conceivable that MNNG and 6-TG share similar cytotoxic processes involving MMR damage surveillance. The MMR-mediated 6TG cytotoxic mechanism is illustrated (Fig. 1-4).

1.5.5.4 Molecular basis of MMR in DNA damage surveillance

The molecular basis of MMR mediated DNA damage surveillance is poorly defined. The p53 pathway may be involved, which is suggested from the comparison of p53 activity between MMR proficient and deficient cells following DNA methylation damage (O^6-meG) induced by Temozolomide (D'Atri et al.,

1998). The expression of p53 and p21/waf-1 (p21/waf-1 is induced by p53) was up regulated following Temozolomide treatment in MMR proficient lymphoblast cells, which is coincident with a G2/M cell cycle arrest and apoptosis. However, in the MMR deficient cells, the cell cycle arrest and apoptosis phenotype was attenuated and no P53 induction was detected (D'Atri et al., 1998). This relationship was supported by a recent study, which showed that MNNG triggered apoptosis was accompanied by p53 and Fas receptor up regulation. Inhibition of Fas receptor activity attenuated MNNG-induced cell death in a lymphoblast cell line (Roos et al., 2004). Although the p53 pathway appears to be involved in the cell cycle arrest and apoptosis in lymphoblast, the involvement of p53 in other cell lines is less certain. In the human kidney derived fibroblast cell line, 293T, p53 is not essential. Although 293T cells lack p53 activity, they can undergo G2/M cell cycle arrest and apoptosis following MNNG treatment (Cejka et al., 2003, di Pietro et al., 2003). A link between MMR surveillance and the ATR signaling pathway was established recently. ATR (ataxia telangiectasia and rad3+ related) is an important cell cycle checkpoint protein kinase, which is activated in response to a block in DNA replication. MSH2 protein interacts with ATR, which regulates the phosphorylation of downstream effectors including CHK1 and SMC1 (structure maintenance of chromosome 1) (Wang and Qin, 2003). The ablation of ATR or the inhibition of CHK1 attenuates the MNNG and 6TG induced G2/M cell cycle arrest (Stojic et al., 2004, Yamane et al., 2004).

1.5.5.5 MED1/MBD4, a methyl-CpG binding protein involved in DNA damage surveillance

MED1 was identified as a protein interacting with MLH1 in human cells (Bellacosa et al., 1999). MED1 is a member of a group of methyl-CpG binding proteins, which is also referred to as MBD4 (methyl-binding domain 4) in some publications. MBD4 (MED1) binds to fully and hemimethylated DNA but not to unmethylated DNA *in vitro* (Bellacosa et al., 1999). Deamination of 5-methylcytosine (m^5C) to T occurs frequently at CpG sites, which causes T:G

mismatch. Mammalian MBD4 has glycosylase activity that enzymatically removes thymine (T) from a mismatched T:G basepair at CpG sites (Hendrich et al., 1999). MBD4 has been shown to be important in suppressing the mutational load caused by deamination of the m^5C. *Mbd4*-deficient mice have an increased rate of CpG mutability and tumorigenesis (Millar et al., 2002, Wong et al., 2002). However, the function of MBD4 is not limited to repairing T:G mismatches at CpG sites. Transfection of a dominant negative form of MBD4 into cultured cells leads to MSI in an episomal slippage construct that contains tandem CA repeats (Bellacosa et al., 1999). This observation and the fact that MBD4 interacts with MLH1 suggest that MBD4 plays a role in MMR. Further studies have revealed that frameshift mutations in *MBD4* coding sequence occur frequently in colon, endometrial, pancreatic and gastric tumors with high rates of MSI (Riccio et al., 1999, Bader et al., 1999, Yamada et al., 2002). However, *Mbd4*-deficient mice generated by gene-targeting do not exhibit MSI (Millar et al., 2002, Wong et al., 2002). A link between MBD4 and MMR surveillance was demonstrated recently by studies on cultured mouse embryonic fibroblasts (MEF) derived from *Mbd4*-deficient mice. In this study the response to DNA damaging drugs were examined and revealed that *Mbd4*-deficient MEFs failed to undergo G2/M cell cycle arrest and apoptosis in response to the treatment of simple methylating agents like MNNG. Moreover, the cytotoxic response to other DNA damaging drugs such as the DNA crosslinking platinum drugs was also attenuated in *Mbd4*-deficient MEF cells (Cortellino et al., 2003). The drug tolerance exhibited by *Mbd4*-deficient MEFs is similar to the DNA damage tolerance exhibited by cells with deficient MMR, which is characterized by the accumulation of DNA lesions in cells (Cortellino et al., 2003). The function of MBD4 in DNA damage surveillance was also observed in the small intestine in *Mbd4*-deficient mice. Mice deficient for *Mbd4* showed significantly reduced apoptotic responses following treatment with a range of cytotoxic agents including cisplatin and 5-fluorouracil (5-FU), a DNA replication inhibitor. Mice lacking both Mlh1 and Mbd4 functions didn't show synergistic effect on DNA damage induced apoptosis, suggesting that MBD4 and MLH1 act in the same pathway (Sansom et al., 2003).

1.5.6 A genetic screen for genes that protect the *C. elegans* genome against mutations

Research on the MMR system was greatly stimulated when the major mismatch repair components were isolated in genetic screens in bacteria. However, genetic screens in dipoid organism like mice is extremely restricted by the difficulty of obtaining homozygous mutations. Taking advantage of the recently developed RNAi (RNA interference) technology in *C.elegans*, a genetic screen has been conducted for genes that protect the *C. elegans* genome against mutations, which includes MMR genes. For simplicity this screen will be referred to as the *C.elegans* MMR screen (Pothof et al., 2003). RNAi technology in *C. elegans* was developed based on the phenomena that the double strand RNA is able to knockdown the expression of the endogenous genes that are homolgous to it (Hannon, 2002). The *C.elegans* MMR screen was based on the *C.elegans* RNAi library, which contains bacterial strains that each produce double-stranded RNAs (dsRNA) for an individual nematode gene. This library is able to target ~86% of predicted *C. elegans* genes. Loss-of-function phenotypes when performing systemic RNAi on a genome-wide scale is estimated to be ~65% (Fraser et al., 2000). To provide a readout for potential MMR mutations (leading to increased DNA genomic instability), a *gfp-LacZ* reporter construct was put out of frame by an A_{17} mononucleotide DNA repeat cloned directly between the initiation ATG and the *gfp-LacZ* open reading frame. Genomic instability mutations that restore the *gfp-LacZ* reading frame can be identified by inspecting the expression of GFP and/or LacZ. The presence of mono- nucleotide repeat sensitizes the reporter system for frameshift mutations. In this screen, several well-known MMR genes were identified including *C. elegans* homologues of human *MLH1, PMS2, MSH2* and *MSH6*. In addition, many genes were recovered with functions in DNA replication, repair, chromatin organization and cell cycle control (Pothof et al., 2003).

1.5.7 MMR deficiency in Cancer

Carcinogenesis is a multi-step genetic process, during which several mutations must be acquired before a normal cell develops into a tumor. The putative tumor cell has to override normal cell cycle control, genetic programs of differentiation, senescence and apoptosis. Each of these steps requires alteration of one or several genes. It has been hypothesized that genomic instability is fundamental during tumorigenesis because elevated mutation rates facilitate the accumulation of multiple mutations (Schmutte and Fishel, 1999). The finding that MMR deficiency is associated with hereditary non-polyposis colorectal cancer (HNPCC) and several sporadic tumors illustrate the importance of the DNA mismatch repair system in maintaining genomic stability (Peltomaki, 2001). HNPCC accounts for nearly 8% of all colon cancers. HNPCC shows an autosomal dominant mode of inheritance, high penetrance and an early onset of tumorigenesis. The molecular hallmark of HNPCC is a high or low level of microsatellite instability (MSI), a feature that is characteristic of MMR deficiency. Indeed, the first human MutS homologue, *MSH2*, was cloned because of its linkage with HNPCC (Fishel et al., 1993). Besides *hMSH2*, germ line mutations in *hMLH1, hPMS2, hMSH6* and *hPMS1* have also been found in HNPCC patients (Wei et al., 2002). HNPCC patients usually inherit one mutated allele from one parent. The other normal allele is mutated in somatic tissues either by loss of heterozygosity, point mutation or hypermethylation. Cells with homozygous mutated MMR genes will then be predisposed to tumorigenesis. MSH2 and Mlh1 are the core components of mismatch protein complex, MutSα and MutLα. Consistent with this role, mutations in *MSH2* and *MLH1* are responsible for 50% and 40% of HNPCC respectively, while mutations in *MSH6*, *PMS2* and *PMS1* are less frequently found in HNPCC (Peltomaki, 2001).

As a consequence of MSI, those genes having simple repeat sequences in their coding region have a greater chance of acquiring a mutation and could be important targets of the MMR deficiency phenotype. Several genes that regulate

cellular growth, cell cycle control, DNA repair and apoptosis lie in this category, including *TGFβ RII, IGF2R, MSH3, MSH6, p53* and *BAX* (Peltomaki, 2001).

1.6 Mutagenesis in mice and embryonic stem cells

1.6.1 Forward genetics, phenotype-based screens

Mutagenesis followed by phenotypic screening is one of the most powerful genetic approaches to elucidate the molecular basis of complex biological phenomena. Such strategies are referred to as “forward genetics”. In the last several decades, forward genetic screens have been conducted in several “model” organisms, including mice. In mice, mutations occur spontaneously at a low efficiency (about $5x10^{-6}$ per locus per gamete). However, the spontaneous mutation rate is far too low for a genetic screen. Highly efficient mutagenesis can be achieved using DNA damaging agents such as irradiation (X-rays, γ-irradiation) or chemicals such as *N*-ethyl-*N*-nitrosourea (ENU). X-ray and gamma irradiation can induce mutations with a rate 20-100 times higher than the spontaneous mutation rate. However, irradiation can break DNA strands, resulting in chromosome rearrangements such as deletions and translocations. The complexity and size of DNA lesions induced by irradiation limits its use as a mutagenesis method. However, γ-irradiation has been used to generate mice containing regional deletions. These deletion mice have been shown to be useful in combination with single gene mutations (You et al., 1997, Goodwin et al., 2001, Bergstrom et al., 1998, Chao et al., 2003).

Chemical mutagens, such as ENU, are one of the most potent mutagens in mice. ENU mutates mouse spermatogonial stem cells with a frequency of $1.5\text{-}6x10^{-3}$ per locus per gamete (Bode, 1984, Hitotsumachi et al., 1985). ENU alkalizes oxygen of DNA nucleotides, which, if not repaired, causes predominately single nucleotide mutations including A/T to G/C, A/T to C/G, A/T to T/A, G/C to C/G and G/C to T/A transitions and transversions in mice. In a typical ENU

mutagenesis screen, male mice (G_0) are injected with ENU to generate mutated gametes. Mating the ENU-treated founder males to unaffected wild type females will then produce G1 offspring. Each of the G_1 animals will carry a unique set of mutated alleles. These G_1 animals can be used directly to screen for dominant phenotypes, or they can be backcrossed to wild type animals to establish lines of mice that carry the same set of mutations. By inter-crossing mice from the same line, some of the descendants will carry homozygous mutations that can be used to screen for recessive phenotypes. Because of the comparative simplicity of a dominant screen, most of the genetic screens that have already been performed were set up to identify dominant phenotypes. Approximately, 2% of G_1 animals exhibit a heritable dominant phenotype (Justice et al., 1999, Brown and Balling, 2001, Balling, 2001).

However, the majority of mutations induced by ENU are recessive. The requirement for a complicated and expensive breeding program, and the difficulty to genotype mice with point mutations limits the application of genome-wide recessive ENU mutagenesis screens. To circumvent this, mice with defined regional chromosome deletions and inversions have been generated via Cre-*loxP* mediated chromosome engineering techniques. These deletion and inversion mice provide essential genetic tools to maximize the efficiency of ENU mutagenesis because the homozygous mutant mice can be identified and the mutation mapped (Yu and Bradley, 2001). By crossing the ENU mutated G_1 mice with mice carrying chromosomal deletions, recessive phenotypes can be manifested in G_2 animals if the mutation lies in the deletion intervals. The use of chromosome deletion mice has been demonstrated by a few pioneering experiments, which located and identified molecular lesions using mice containing a set of overlapped small deletions generated either by gamma-irradiation or by chromosome engineering techniques (Bergstrom et al., 1998, Su et al., 2000, Lindsay et al., 2001, Chao et al., 2003).

Although mice that carry small deletions are very useful, larger deletions may sometimes lead to reduced fitness, infertility or even lethality because of haploid insufficiency in some genes. Balancer chromosomes were developed originally in *Drosophila* to maintain recessive lethal mutations and have been proved to be an important tool for stock maintenance. A balancer chromosome carries one large inversion or a set of inversions along a chromosome. Productive meiotic crossovers between an inversion chromosome and a normal chromosome are efficiently suppressed because this type of crossover leads to inviable germ cells, harboring dicentric or acentric chromosome. Recently, balancer chromosomes in mice carrying large chromosomal inversions were created with Cre-*loxP* mediated chromosome engineering (Zheng et al., 1999). These balancer chromosomes were engineered to carry a visible dominant marker and a recessive lethal mutation on the chromosomal inversion. Mice with homozygous balancer chromosomes are automatically eliminated from crosses between heterozygous mice with one balancer chromosome because of the recessive lethal mutation. Mutations can be maintained without recombinational loss in the balanced heterozygotes and tracked by the visible dominant marker. The utility of balancer chromosomes has been demonstrated recently in an ENU mutagenesis screen for recessive mutations along a 24 cM balanced chromosomal region on mouse chromosome 11 (Kile et al., 2003).

Because of the lack of an overt molecular tag, the identification of an ENU induced mutation normally starts with linkage analysis in order to locate the mutation of interest within a small chromosomal region of several centimorgens (cM), which often requires analyzing hundreds of meiotic events. Candidate genes within that region can be assessed based on their expression pattern, structure, and functional domains. Sequencing of candidate genes will determine the molecular change. Confirmation of the mutation can be acquired by phenotype-complementation, for example using a cDNA construct or a large genomic DNA fragment such as bacterial artificial chromosomes (BACs) that harbor one or multiple candidate genes (King et al., 1997, Allen et al., 2003,

Floyd et al., 2003, Swing and Sharan, 2004, Zhang et al., 1994). As discussed above, mice with chromosome deletions and inversions provide essential tools for ENU mutagenesis in term of mutation identification and maintenance.

ENU mutagenesis is a useful technique which can quickly create a large number of mutations at random to allow screens for any phenotypic abnormality provided that the phenotype of interest is visible or detectable. Moreover, because ENU induces point mutations in a random fashion, independent mutations in the same gene can be generated which may act as hypermorph and neomorph alleles in addition to the common loss of function alleles. ENU mutagenesis screens have been conducted on many thousands of mutant mice and a number of mutations have been characterized which mimic human diseases.

1.6.2 Reverse genetics, transgenic animals and gene-targeting

In contrast to applying random mutagenesis and conducting forward genetic screen, reverse genetics can be used to directly obtain functional information of a gene by mutating or over expressing the gene and examining the consequence in the resultant transgenic or knockout mice. Transgenic mice were originally generated by directly injecting exogenous DNA into fertilized zygotes or by retroviral infection of early development stage embryos (Jaenisch, 1988). The injected DNA or retroviral vector could stably integrate into the host genome and was transferred into the mouse germ line. These methods usually produce gain of function alleles that are useful in studying the biological effect of over-expressed genes *in vivo* (Berns, 1991). However, these methods have shown some major limitations. First, the integration of injected DNA is random and uncontrollable. The injected DNA often forms head to tail concatemers before integration and integration is often accompanied by chromosomal rearrangements in the flanking DNA around the integration site. Furthermore, the expression of genes varies between different integration sites, cell and tissue types. Expression from a retroviral vector can even be totally abolished by DNA

methylation (Jaenisch, 1988). Recently, gene-targeting technology has been developed as a new version of transgenic technology, which is capable of introducing a single copy DNA fragment into a specific genomic locus in a predicable manner.

Taking advantage of mouse embryonic stem cell (ES) technology and homologous recombination, gene-targeting provides a more powerful means to generate transgenic mice harboring precise mutations in the gene of choice. ES cells are pluripotent cells, established from inner cell mass (ICM) of pre-implantation blastocysts. ES cells in culture maintain unlimited self-renewal ability. Most importantly, ES cells, even after modification in culture, can contribute to all somatic tissues as well as the germ line of chimaeras when they are injected into host blastocysts (Evans and Kaufman, 1981, Bradley et al., 1984, Robertson et al., 1986, Kuehn et al., 1987). Targeted mutations can be achieved by homologous recombination between endogenous genes and a targeting vector (Doetschman et al., 1987, Thomas and Capecchi, 1987). In its simplest form a gene-targeting vector is constructed to carry a DNA fragment homologous to the targeted gene and a positive selection marker. ES cells can be directly selected with the integrated targeting vector and the subset with the engineered mutation at the targeted site can be identified by Southern-blotting or PCR. Since the advent of gene-targeting technology in the late 80's, it has quickly evolved to be one of the most frequently practiced approaches in mouse genetics. With improved molecular cloning technologies like *E.coli* recombineering (Copeland et al., 2001), gene-targeting vectors can be quickly constructed with long homologous arms to obtain better gene-targeting efficiencies. Nowadays, gene-targeting vectors can be engineered at will to target any genes, generating all possible classes of mutations like loss of function, gain of function, point mutations and knockin alleles. Combined with the Cre-*loxP* technology, the "expression" of a mutation can also be made controllable or inducible in a temporal or spatial manner (Ramirez-Solis et al., 1993). Despite the power of creating mutations in targeted genes, the gene-targeting approach

requires the prior knowledge of genes to design a gene-targeting vector. Thus, novel phenotypic information about a gene is often missed. Gene-targeting can only be applied on a gene-by-gene bases and it has not yet been employed for genetic screens.

1.6.3 Insertional mutagenesis, the gene trap approach

Random insertional mutagenesis can also be used to mutate genes. The integrated DNA molecule provides a sequence tag for identifying the mutated gene using a PCR-based method. Retroviruses have been used as insertional mutagens since the late 70's. The integration of a retrovirus may produce a loss of function mutation when it inserts into the coding region of a gene. Retroviruses can also generate gain of function mutations, in which expression of a gene is increased by the viral enhancer element (Jaenisch et al., 1981, Lund et al., 2002, Mikkers et al., 2002). The mutational efficiency of a randomly integrated retroviral vector is very low. Over 95% of mouse genome is non-coding sequences. Retroviral integrations in these regions are often phenotypically "neutral" to cells. The availability of ES cell technology in the mid 80's expedited the design of better insertional mutagens, the gene trap vectors, in the following years. Gene-trap mutagenesis predominantly produces loss of function mutations in a random fashion. The gene trap vector serves as a molecular tag for cloning of the mutation. Combined with the ES cell technology, gene trap offer a valuable tool for rapidly creating large numbers of loss of function mutations for functional genomic studies in mice (Stanford et al., 2001, Evans et al., 1997)

1.6.3.1 Gene trap methods

During past 10 years, various gene trap vectors have been designed for individual experiments. These vectors contain a non-functional reporter gene cassette and the expression of the reporter gene requires the cis-elements of an endogenous gene. The basic gene trap designs include enhancer traps,

promoter traps and polyadenylation signal (PolyA) traps (Fig. 1-5) (Zambrowicz and Friedrich, 1998).

The enhancer trap vectors were originally used to study the effect of host genes on the expression of transgenic reporters in mice. For this purpose, the enhancer trap vectors were built to contain *E.coli lacZ* gene with a minimal promoter sequence. The expression of the reporter requires the vector to insert near a cis-acting enhancer element (Allen et al., 1988, Kothary et al., 1988, Gossler et al., 1989). Similar designs were adopted in *Drosophila* in genetic screens for cis-elements that were able to drive the expression of a minimal promoter fused to a *lacZ* reporter gene. These early experiments established that the expression of the reporter gene is regulated by the flanking cis-elements, and the reporter expression often displays a spatially or temporally restricted pattern (Bellen et al., 1989, Bier et al., 1989). Enhancer trap vectors haven't been extensively used because they are not efficient mutagens. The enhancer elements of a gene are often a large distance away from the coding elements so that the insertion of the enhancer trap vector does not normally disrupt the expression of the gene. Promoter traps and PolyA traps are much better mutagens.

The essential component of a promoter trap is a reporter gene that possesses a strong splicing acceptor (SA) but lacks a promoter. Therefore, the reporter can only be transcribed from the endogenous gene into which the promoter trap reporter integrates, generating a fused transcript containing a 5' portion of the endogenous gene and the coding sequence of the reporter. Consequently, the transcription of the endogenous genes is disrupted, creating a loss of function mutation. The expression of the mutated gene can be assessed by inspecting the expression of the reporter gene. Because of the nature of the promoter gene trap design, these vectors can only mutate genes expressed in the experimental cell line.

A PolyA trap vector utilizes a reporter gene lacking a polyadenylation signal, but possessing a “strong” splice donor (SD). The reporter gene has its own promoter but can only generate a stable transcript if the PolyA trap vector inserts into an endogenous gene and downstream PolyA signal is provided. In contrast to promoter traps, the PolyA trap vector can be used to mutate any gene regardless of its expression status in the experimental cell line since the reporter is expressed from an exogenous active promoter.

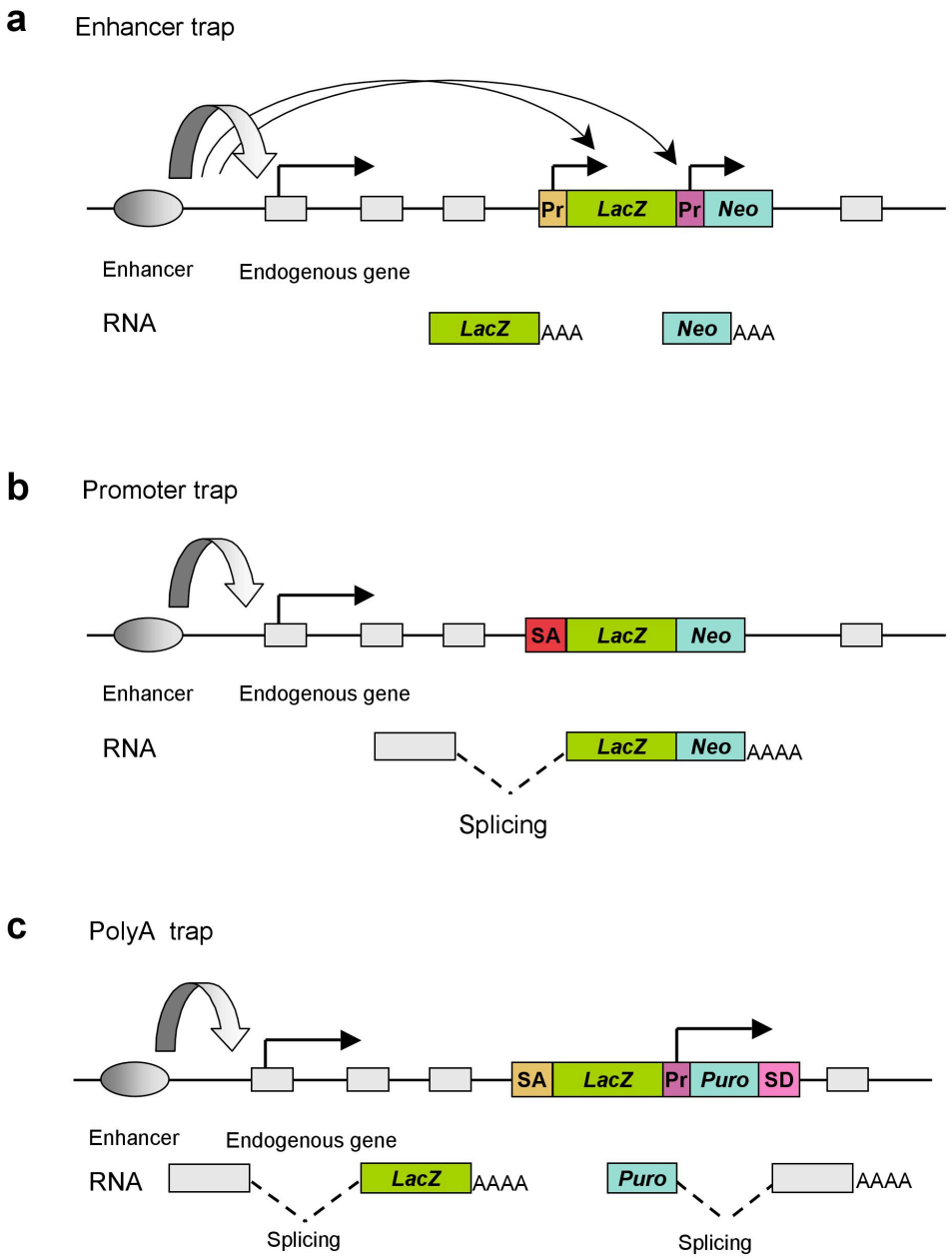
a
Enhancer trap
Pr
LacZ
Pr
Neo
Enhancer
Endogenous gene
RNA
LacZ
AAA
Neo
AAA
b
Promoter trap
SA
LacZ
Neo
Enhancer
Endogenous gene
RNA
LacZ
Neo
AAAA
Splicing
c
PolyA trap
SA
LacZ
Pr
Puro
SD
Enhancer
Endogenous gene
RNA
LacZ
AAAA
Puro
AAAA
Splicing
Splicing

Figure 1-5. Schematic representation of basic gene trap strategies, showing the structure and the expression of the gene trap cassettes integrated in an endogenous gene**. a.** Enhancer trap. *LacZ* and *Neo* reporter genes are driven by minimal promoters (Pr). The exression of reporter genes are enhanced by the endogenous enhancer. **b.** Promoter trap, showing the *SAβgeo* gene trap cassette, a fused *lacZ/Neo* gene with a consensus splicing acceptor (SA). **c**. PolyA trap. *Puro* is transcribed from an autonomous promoter (Pr) and spliced from the splice donor (SD) in the gene trap cassette into endogenous gene. Note that in some cases, a fused *SAβgeo* promoter gene trap cassette is combined with the PolyA trap vector, which can provide a color reporter for monitoring the expression of endogenous genes (Zambrowicz et al*.,* 1998). Gray rectangles represent exons of an endogenous gene.

1.6.3.2 Gene trap mutagenesis in genetic screens

1.6.3.2.1 Expression screens

A genetic screen is an essential approach to establish relationships between genes and functions. Phenotype-driven screens in mice have been extremely restricted because of the difficulty in obtaining homozygous mutations. One fact of a gene's function can often be obtained by assessing its expression. For example, the expression of developmentally important genes often exhibited highly restricted patterns during development. The expression of genes involved in cell signaling pathways can be induced or repressed by physiological molecular signals. Because the promoter trap approach allows a quick examination of the expression of an endogenous gene, this vector type has been used for expression screens. Wurst et al. (1995) performed an expression screen in mouse embryos for genes involved in embryogenesis based on the hypothesis that such genes will exhibit temporally or spatially restricted expression patterns. They mutated ES cells with a promoter trap vector containing the *lacZ* reporter gene. 279 gene trap clones were assessed in chimeric embryos, and by X-gal staining the expression patterns of the mutated genes were examined. Approximately, one third of genes expressed in ES cells are either temporally or spatially regulated during embryogenesis (Wurst et al., 1995). This work demonstrated the feasibility of the use of the promoter gene trap in an expression screen. However, the generation of a large quantity of chimeras or mice requires significant time, labor and animal resources and is not practical for many laboratories.

ES cells are pluripotent cells. They can not only contribute to all tissues in mice but also differentiate into many cell lineages *in vitro*, therefore allowing prescreening of genes which function in specific types of cells. The gene-trap mutagenesis combined with various *in vitro* ES cell differentiation conditions has

been applied to screen for genes expressed in chondrocytes, cardiomyocytes, skeletel muscle cells, haematopoietic cells, endothelial cells and neurons (Baker et al., 1997, Stanford et al., 1998, Hirashima et al., 2004, Muth et al., 1998, Hidaka et al., 2000, Shirai et al., 1996, Thorey et al., 1998, Stuhlmann, 2003).

Chromatin or chromosomal proteins normally show restricted cellular localization within nuclear compartments or sub-nuclear compartments, thus the genes encoding these proteins may be identified by examining the localization of the gene trap reporter protein in cells (Tate et al., 1998). Genes with altered expression levels in response to many physiological stimuli or signals, such as retinoic acid and gamma-irradiation have also been screened in culture using the promoter gene trap approach (Forrester et al., 1996, Vallis et al., 2002, Mainguy et al., 2000).

Gene trap vectors can also be specially designed to suit individual screens. A secretory trap vector was designed to identify secreted and transmembrane proteins. In this screen, a transmembrane signal sequence was placed adjacent to the *βgeo* gene trap reporter. The transmembrane signal will place the βgeo protein inside the endoplasmic reticulum (ER) so that it doesn't function. To allow the detection of the *βgeo* expression, an additional N-terminal signal sequence from the trapped gene is required to place it outside the ER (Skarnes, 2000). Hundreds of secreted and transmembrane proteins have been identified by this approach (Mitchell et al., 2001). Recently, this secretary trap was modified to identify genes controlling neuronal axon guidance. In this design, an axonal marker is co-expressed with the *LacZ* gene trap reporter to label the neuronal axons. By staining the expression of the *LacZ* reporter and the axon marker in mice, genes with restricted expression patterns in neuronal axons were identified (Leighton et al., 2001).

The gene trap approach is not restricted to ES cells, it can be applied to other cultured cell types to study genes with unique features in these cell lines. For

example, a gene trap screen has been conducted in hematopoietic cells which were induced to undergo apoptosis by growth factor deprivation. Genes with potential survival functions in hematopoietic lineages could be identified based on their induced expression following growth factor deprivation (Wempe et al., 2001). To better understand the complex signaling networks involved in germ cell maturation, gene trap screens have been conducted in Sertoli cells, the somatic cells supporting and controlling male germ cell development (Vidal et al., 2001). Differentiating germ cells were then added to the mutated Sertoli cells to screen for cells showing changes in the expression of the trapped genes. Gene trap strategies have been used in NIH3T3 fibroblasts to identify inhibitors of oncogenic transformation, in cultured B-cells to identify lipopolysaccharide (LPS) responsive genes and in human lung carcinoma cells to identify TGF-beta-responsive genes (Kerr et al., 1996, Andreu et al., 1998, Akiyama et al., 2000). Taken together, these experiments show that gene trap approach is a powerful mutagenesis method with versatile and broad applications in genetic screens.

1.6.3.2.2 Gene trap in phenotype-driven screens

Phenotype-driven screens in diploid genome require a strategy to obtain homozygous mutations. Chinese hamster ovary (CHO) cells contain a partial hemizygous genome. Therefore, recessive mutations within the hemizygous regions can be phenotypically accessed. Screens in CHO cells have been successfully applied to isolate recessive mutations that are sensitive to UV radiation, for example the base excision repair (BER) protein *ERCC1* (Westerveld et al., 1984). A Gene trap screen has also been conducted in CHO cells to identify mutations in glycosylation. Cells with defects in glycosylation are resistant to wheat germ agglutinin. Four individual mutants were isolated in this experiment. By Southern-blot analysis, four gene trap insertion sites were mapped to different positions in a 796 base pair region (Hubbard et al., 1994). The localization bias of the gene trap mutations identified in this screen may reflect the limited hemizygous genome of CHO cells or gene trap vector insertion

“hot spots”. Unfortunately, information about the efficiency of the screen was not provided, nor the identify of the mutated gene was isolated.

1.6.3.3 Methods for introducing gene trap vectors into cells

1.6.3.3.1 Electroporation

The simplest way to perform gene trap mutagenesis is to electroporate the linearized gene trap vector into cells as “naked” DNA. The gene trap vector can integrate into genome randomly, which is normally accompanied by DNA concatermerization. Thus many copies of electroporated linear DNA molecules form head to tail arrays and integrate into host genome together through a process mediated by a DNA repair process known as non-homologous end joining DNA repair (NHEJ) (Brinster et al., 1985, Skarnes, 2000). This method has several limitations. First, multiple copies of the gene trap vector in one locus complicates the identification of the gene trap mutations. Second, the gene trap vector can be truncated during electroporation.

1.6.3.3.2 Retroviral based gene transfer

1.6.3.3.2.1 Retroviral life cycle

The typical retrovirus genome consists of two copies of a single-stranded RNA molecule of about 8-12 kb, depending upon the retroviral species. The genome encodes three major proteins, Gag, Pol and Env. Gag is processed to make the core proteins. Pol has the reverse transcriptase, Rnase H and integrase activities. Env is the viral envelope protein that resides in the lipid layer and mediates the viral-host cell interaction during viral infection. The viral particle consists mostly of gag-derived proteins, genomic RNA, and the reverse transcriptase protein as the virus nucleoprotein core, which are enclosed by the outer lipid-protein shell of the viral envelope. Viral particles infect host cells by

binding to cell surface receptors, a process determined largely by the envelope proteins of the retrovirus. Infection leads to injection of the virus nucleoprotein core. Once inside the cell, a double-stranded DNA is generated from the viral genomic RNA by the reverse transcriptase. Catalyzed by the viral integrase, the double strand viral DNA integrates stably into the host chromosome. The integrated viral DNA is known as proviral DNA. At this stage, the virus is now prepared to initiate a new round of replication. Full-length genomic mRNA is transcribed from proviral DNA by the host cell RNA polymerase, initiated at the beginning of the R region of the 5' LTR (Long Terminal Repeat) and terminating at the end of the R region at the 3'LTR. Full length genomic RNA can be spliced and provides messenger RNAs, from which the viral proteins are synthesized. The full length genomic RNA and the viral nucleoproteins are packed into new viral particles and released from the host cell by budding from the plasma membrane (Coffin J M and E, 1996).

1.6.3.3.2.2 Recombinant retroviral, viral packaging cell lines

Recombinant retroviral vectors have been developed and been wildly used to transfer genes into eukaryote cells because of the capability of retroviral integration, allowing constitutive expression of exogenous genes carried by the retrovirus. Recombinant retroviral vectors have been constructed. The major components of recombinant retroviral vectors include the 5' long terminal repeat, the 3' long terminal repeat and cis -elements essential for viral RNA packaging, such as viral packaging sequence Ψ. The viral proteins can be produced in trans and are thus deleted from the viral genome to accommodate exogenous DNA. Deletion of the trans-elements in a recombinant retrovirus leads to replication

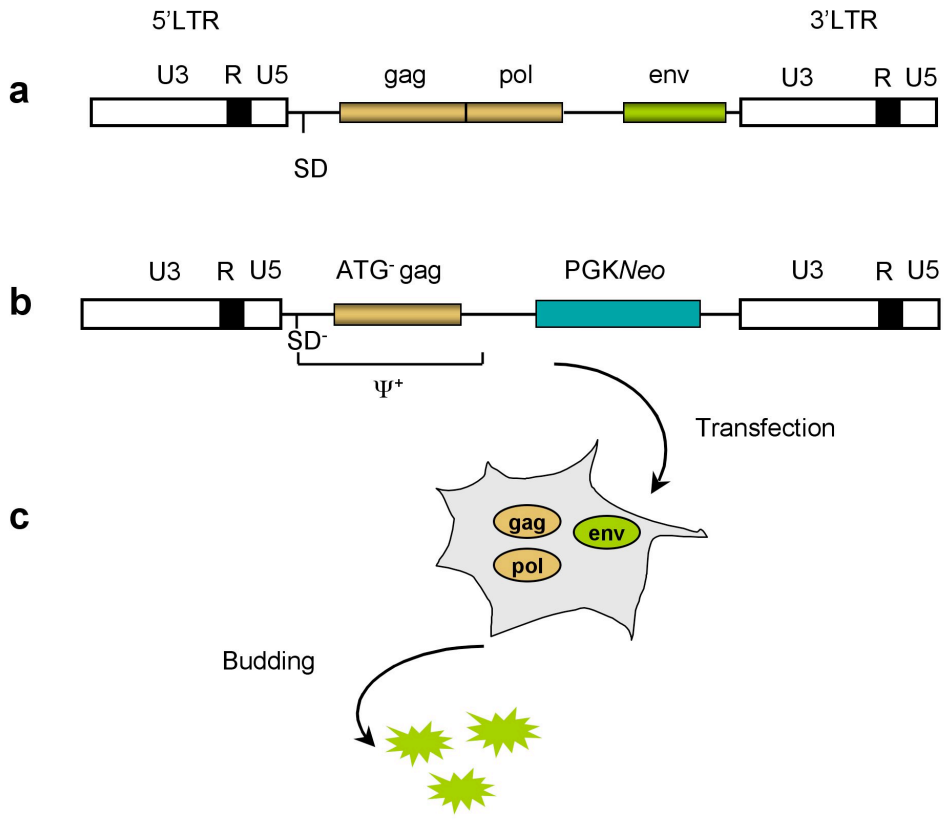

Figure 1-6. Recombinant retroviral vectors and viral production.

a. The structure of wild type retrovirus with a simple genome, illustrated as a provirus containing the long terminal repeat (LTR), genes that encode viral protein core (*gag*), reverse transcriptase (*pol*) and the envelope protein (*env*). SD, viral splice donor. **b.** Schematic of a representative recombinant retroviral vector based on Moloney murine leukemia virus (Mo MuLV). DNA sequences including *pol* and *env* genes are deleted. The splice donor and *gag* sequence remain to facilitate viral packaging, which is indicated as Ψ^+, representing an extended viral packaging signal. To avoid the interference of internal gene expression by viral mRNA splicing and protein translation, the viral splice donor is mutated (SD^-) and the initiation codon of the *gag* gene is deleted (ATG^-gag). Figure is adapted from retroviral vector pBabe (Morgenstern *et al.*, 1990). **c.** To produce infectious virus, proteins that are required for viral reproduction, Gal/Pol and Env, are expressed in a mammalian cell line (viral packaging cell line). The recombinant retroviral vector DNA is transfected into the viral packaging cell line and infectious viral particles are packaged and released from the cells.

deficiency. To produce infectious virus, proteins that are required for viral reproduction, Gal/Pol and Env, are expressed in a mammalian cell line, so called viral packaging cell lines. Once the recombinant retroviral vector DNA is transfected into the viral packaging cell line, infectious viral particles can be produced and released (Fig. 1-6)(Somia, 2004).

1.6.3.3.2.3 Self-inactivating (SIN) retroviral vector

A more recent development is the self-inactivating (SIN) retrovirus that lacks the enhancer or both enhancer and promoter sequences in the integrated provirus. The viral U3 regions of the LTRs possess strong enhancer and promoter activity, which can interfere with the expression of exogenous genes from the internal promoter. Viral enhancers in integrated provirus can activate surrounding cellular genes, such as oncogenes. In some cell lines, these enhancers are targets for epigenetic silencing. A SIN retroviral vector will produce an integrated provirus lacking the viral enhancer and/or promoter. In a typical SIN vector, the enhancer sequence in the U3 region in the viral 3'LTR is removed, while the enhancer in the 5'LTR remains intact. Thus, a full length genomic RNA can be generated by the functional 5'LTR and an infectious viral particle can be produced. However, as a consequence of reverse transcription and second strand synthesis the U3 region of the 5'LTR is copied from the U3 region of the 3'LTR. Thus the integrated provirus will contain the deleted U3 region in both of the 5' and 3'LTRs, leading to an inactivated provirus lacking the enhancer (Fig. 1-7) (Yu et al., 1986, Yee et al., 1987, Soriano et al., 1991).

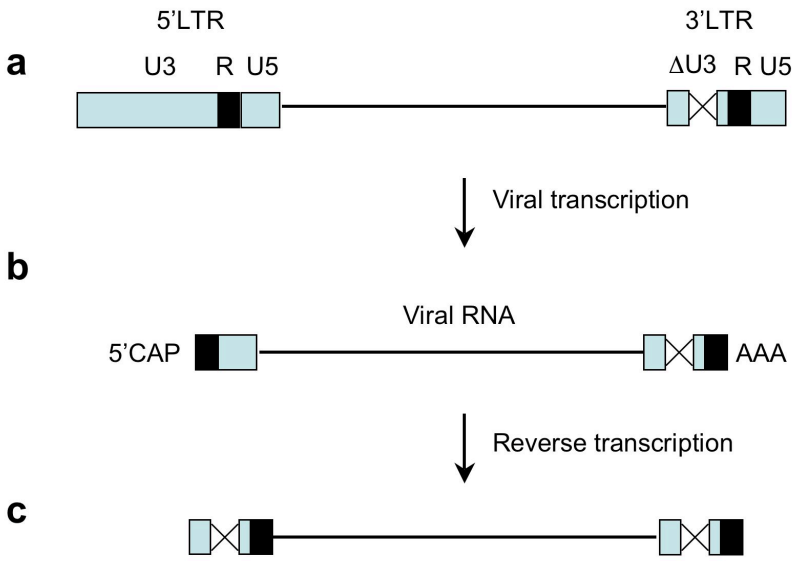

Figure 1-7. Principle of self inactivating (SIN) retroviral vectors.

a. The structure of a recombinant retroviral vector, in which the enhancer fragment in the U3 region of viral 3'LTR is deleted. **b**. Viral RNA is transcribed, initiating from the R region of the 5' LTR and ending at the end of R region of the 3'LTR. **c.** Reverse transcription of viral RNA in host cells generates a linear DNA duplex. The U3 region of the 5'LTR is copied from the U3 region of the 3'LTR in viral genome RNA during this process. The resultant virus will contains a non-functional 5'LTR harboring the deletion, thus it is inactivated.

1.6.3.3.2.4 Retroviral based gene traps

Von Melchner and Ruley developed the first retroviral gene trap vector (von Melchner and Ruley, 1989). In this design, the gene trap cassette is inserted in the U3 region of 3'LTR and replaces the viral enhancer. After viral replication and integration, the provirus carries a duplicated gene trap cassette in both of the 5' and 3'LTRs (von Melchner et al., 1992). Friedrich and Soriano (Friedrich and Soriano, 1991) constructed another version of retroviral gene trap vector, ROSA (reverse orientation splice acceptor) gene trap vector. In this ROSA vector, the gene trap cassette was placed between viral LTRs of a SIN vector in the opposite orientation relative to viral transcription. This reverse orientation was essential in order to avoid removal of the viral packageing sequence Ψ from the full length genomic RNA by splicing from the upstream viral splice donor sequence to the splice acceptor in the gene trap cassette.

Retroviral gene trap vectors can also be made revertible by inserting a *loxP* site into viral U3 region in the 3' LTR. The *loxP* site will be duplicated to the 5'LTR in the integrated provirus, resulting in a provirus flanked by *loxP* sites. By Cre-*loxP* mediated recombination the, *loxP*-flanked provirus can be removed, leaving only a single LTR with a *loxP* site in the genome (Ishida and Leder, 1999).

Gene trap mutagenesis using a retroviral vector has advantages and limitations First, in contrast to electroporation, only a single copy of retrovirus integrates into one genomic locus. Second, by controlling the viral multiplicity, most of cells will contain a single gene trap mutation. Another advantage of this method is that once a stable virus producing cell line is made, large amount of gene trap virus can be produced easily, which significantly improves the throughput of the gene trap mutagenesis method. The major limitation of the retroviral based gene trap method is gene trap "hot spots" caused by non-random retroviral integrations (discussed in the following section).

1.6.3.4 Gene trap "hot spots"

Although gene trap mutagenesis was originally designed as a random method, it has been noticed that some genes appear to be mutated more frequently by a gene trap vector than others (Skarnes, 2000). Recently, the German Gene Trap Consortium (GGTC) has reported a systematic analysis of gene trap "hot spots" by collecting over 10,000 gene-trapped ES clones using four different gene trap vectors, including both electroporation-based and retroviral-based vectors (Hansen et al., 2003). They found that the gene trap insertion sites were dispersed throughout the genome and occurred more frequently in chromosomes with high gene density, which suggests that there is no obvious bias to a single chromosome. 75% of the gene trap mutations appeared only once in the gene trap database, while 25% were "hit" multiple times, suggesting that most genes are accessible to gene trap mutagenesis and "hot spots" (25%) are relatively minor targets. By comparing gene trap "hot spots" arising from different gene trap vectors, they found that some of the "hot spots" (nearly 50%) are common for all vectors, suggesting that the gene trap efficiency could be affected by locus-specific factors. Notably, more than 50% of the hot spots are vector-specific, suggesting that each gene trap vector design will have limited genome coverage. Therefore, it is recommendable to utilize multiple vectors in order to obtain broader genome coverage in gene trap mutagenesis.

The factors that cause gene trap hot spots have not been clearly demonstrated, especially for those specific for individual vectors. Some general factors have been recognized, for example, chromatin structure is expected to affect the gene trap efficiency. Open euchromatic regions that contain transcriptionally active gene are believed to be more permissible to the integration of gene trap vectors. The recovery of cells with gene trap mutations requires that the reporter gene is stably expressed. Therefore, factors that affect the expression or the stability of the gene trap reporter could contribute to a bias of the gene-trap vector, for example, the gene structure and the reading frame. If a fused gene trap

transcript has the endogenous protein translation initiation codon (ATG) before the ATG codon of the gene trap reporter, the translation machinery will prefer the first one and translate a fused protein. In this case, a truncated endogenous protein will be produced if the endogenous ATG and the reporter gene are not in the same reading frame. Consequently, this gene trap mutation will not be recovered. To solve this problem, an IRES (internal ribosome entry site) fragment may be placed between the splicing acceptor and the gene trap reporter. The IRES sequence allows the CAP-independent translation of the reporter from an internal ATG site. Therefore, translation initiation of the reporter will be independent of the reading frame of the trapped gene. In addition, the IRES sequence is able to enhance protein translation, allowing detection of the genes expressed at low levels (Bonaldo et al., 1998).

1.7 Thesis project

The primary goal of the project was to explore the possibility of generating homozygous mutations in *Blm*-deficient mouse ES cells and to investigate the application of a recessive genetic screen for genes involved in MMR surveillance. In this introduction, the function and phenotypic consequences caused by *Blm*-deficiency has been discussed. In mouse ES cells it has been shown that a single allele mutation on an autosomal chromosome can be lost frequently via LOH, generating a bi-allelic (homozygous) mutation. This feature of *Blm*-deficient cells was explored and used as a genetic tool to generate homozygous mutations. Another aim of the study was to identify new MMR components. Although the key players of the MMR system have been identified and their role in repairing DNA replication errors have been studied in detail, knowledge of the MMR system is incomplete, for example, how does the eukaryotic MMR system distinguish the nascent DNA strand in replicating DNA? What is the molecular basis of MMR surveillance? Finally, the knowledge of the MMR system was largely obtained from studies in bacteria and yeast. Although the MMR system seems highly conserved, in higher eukaryotes the MMR system has evolved

more specific functions. Some of these functions have been elucidated (for example in meiosis) while others have not yet been fully defined, for example, the function of MED1/MBD4 in MMR mediated DNA damage surveillance. MED1/MBD4 is a methyl-CpG binding protein. It is notable that this aspect is only found in mammals, but not in yeast, worms or fruit flies since their genomes are deficient in DNA methylation. Therefore, it is important to identify mammalian specific MMR genes. In this regard, performing a genetic screen in a mammalian system for MMR genes is essential to identify elements of this system that can't be identified based on evolutionary conservation.