 M.G. Marinus, LRB823, x63330, martin.marinus@umassmed.edu

Mutation

References
Snyder, L. and Champness, W. "Molecular Genetics of Bacteria", 2nd edition, 2003, ASM Press, Washington DC. Chapter 3. - *Good introductory coverage.*
Horst, J.P. et al. 1999. *Escherichia coli* mutator genes. Trends Microbiol. 7, 29-36. -*A recent review.*
Slechta ES, et al 2002. Evidence that selected amplification of a bacterial *lac* frameshift allele stimulates Lac(+) reversion (adaptive mutation) with or without general hypermutability. Genetics. 161, 945-56. -*Adaptive mutation in detail*.
Ellis et al. 2001. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. PNAS 98, 6742-6746. -*Mutagenesis made easy.*

A1. Definition
A mutation is any heritable change in DNA sequence. This may, or may not, affect the phenotype of the organism. Such a change may occur either spontaneously or be induced as a consequence of treatment with a DNA damaging agent (X-rays, ultraviolet light, chemicals, etc.). In *E. coli*, it is also possible to target mutations at any desired base pair by molecular genetic techniques using single-stranded oligonucleotides (Ellis et al., 2001).

A2. Types of Mutation
Missense, nonsense, base pair, transition, transversion, frameshift, deletion, insertion, duplication, inversion, translocation, silent, suppressor, etc. These are described in the Snyder and Champness chapter on mutation. This chapter also briefly describes the process of translation and its role in suppression of nonsense mutations.

A3. Mutation Rates are Low
In laboratory populations of growing *E. coli*, the rate of mutation per base pair is about 10^{-10} -10^{-11}. These changes are the result of errors from DNA replication or from lack of DNA repair. Given that the error rate of base selection by the DNA polymerase III catalytic subunit (alpha) is about 10^{-4} -10^{-5}, other backup systems must exist to remove these errors. Two such systems are "proofreading" and "mismatch repair", which are described below. Proofreading and mismatch repair each contributes about 100 to 1000-fold to fidelity of chromosome duplication.

In non-essential *E. coli* genes, the most frequent spontaneous alteration is due to insertion sequence (IS) elements. These elements are about a thousand base pairs long and encode proteins for their own maintenance and transposition. They are similar to Ty elements in yeast, P and *copia* elements in *Drosophila* and *Ds* elements in maize.

High mutation rates (10,000-fold above spontaneous) can be achieved under certain conditions in mutant *E. coli* bacteria; however, such populations die out eventually due to "error catastrophe".

In growing *E. coli* cells, spontaneous mutations occur at random throughout growth of the culture. This is the basis for the Luria-Delbruck fluctuation test. In non-growing cells, mutations also occur but do not follow the Luria-Delbruck distribution and appear to arise by a different

mechanism that is dependent, in part, on certain recombination proteins and may involve DNA amplification.

The low rate of mutation has been interpreted as a balance between genome stability and the necessity to produce mutations for evolution. In natural bacterial populations, however, the frequency of mutator bacteria is very high (about 10%) as compared to laboratory strains (about 10^{-6}).

A4. Deamination of 5-Methylcytosine
5-Methylcytosine is widely distributed in pro- and eukaryotes and is an important cause of mutation by the spontaneous deamination of 5-methylcytosine to thymine. This produces T/G mispairs, which, if replicated, cause GC to AT transitions. These mutational "hotspots" were first detected in the *lacI* gene but have been found in many other genes including that for human p53 where about one third of the mutations are generated by this mechanism.

B. Target Genes
Although many genes have been used as targets for mutational analysis, the *lacI* gene is a popular target especially the first 250 bp that encode the DNA binding domain of the Lac repressor and can be selected as dominant-negative mutations (*lacI^{-d}*). The proximal part of the *lacZ* gene is also a popular target. Mutation to rifampicin-resistance occurs in the gene encoding the beta-subunit of RNA polymerase and since this is highly conserved in bacteria, it is becoming the experimental system of choice for comparative studies.

C. Mutator Genes (Table 1)
As mentioned above, the DnaE subunit of DNA polymerase III holoenzyme makes mistakes at a low frequency. It should, therefore, be possible to isolate *dnaE* mutants that lead to either increased or decreased mutagenesis. It should also be possible to isolate mutations in the repair systems that remove replication errors. These should show an increased mutation frequency.

C1. Isolation of mutator genes. The selection is for increased mutation frequency using various reporter systems. The first mutator gene (*mutT*) was a spontaneous isolate showing instability for phage resistance. Several others were isolated by "brute force" methods where mutagenized survivors were monitored for increased frequency to antibiotic resistance (e.g., streptomycin). Others were isolated by screening for specific base pair alterations. In this case a panel of *lacZ* mutants was constructed which had defined mutations affecting the catalytic glutamate-461 in the β-galactosidase active site. No other amino acid can substitute for glutamate. These mutants have a Lac^{-} phenotype.

-GGG AAT GAG TCA GGC- β-galactosidase coding sequence
E461

Sequence at E461	Type of mutation scored in Lac^{+} revertant
GAG	None
TAG	AT to CG
GGG	GC to AT
CAG	GC to CG
GCG	GC to TA
GTG	AT to TA
AAG	AT to GC

A particular Lac^{-} strain can be mutagenized and the survivors screened for an increase in Lac^{+} reversion. This can be done easily by inspection of colonies on the appropriate medium.

C2. Holoenzyme mutators. These map to *dnaE* (*polC*) and *dnaQ*. Mutator mutations in *dnaE* have either increased or decreased mutagenesis but this does not correlate with correct nucleotide selection. For at least one of these alleles, the effect is to modify the proofreading activity of the DnaQ subunit. For historical reasons, a powerful mutator mutation in *dnaQ* was designated *mutD*. MutD proteins have decreased 3' to 5' exonuclease activity thereby preventing removal of the mismatched base at the primer terminus. The massive number of errors produced in the *mutD* strain results in saturation of the mismatch repair system (see below).

C3. MutT. The *mutT* gene encodes a nucleoside triphosphatase that preferentially hydrolyzes 8-oxodGTP and 8-oxoGTP, thereby preventing incorporation of the oxidized base into DNA and RNA (see Fig. 2).

C4. Mismatch repair mutators. This group includes the *dam*, *mutS*, *mutL, uvrD* and *mutH* mutators. A several hundred-fold increase in frameshift and transition mutations are found in these strains. The pathway is shown in Fig. 1. A brief description of the pathway is also given in Snyder and Champness, p.28. The system is conserved in other bacteria, archaea, and eukaryotes. In humans, loss of mismatch repair is the basis for hereditary non-polyposis colon cancer and some sporadic cancers.

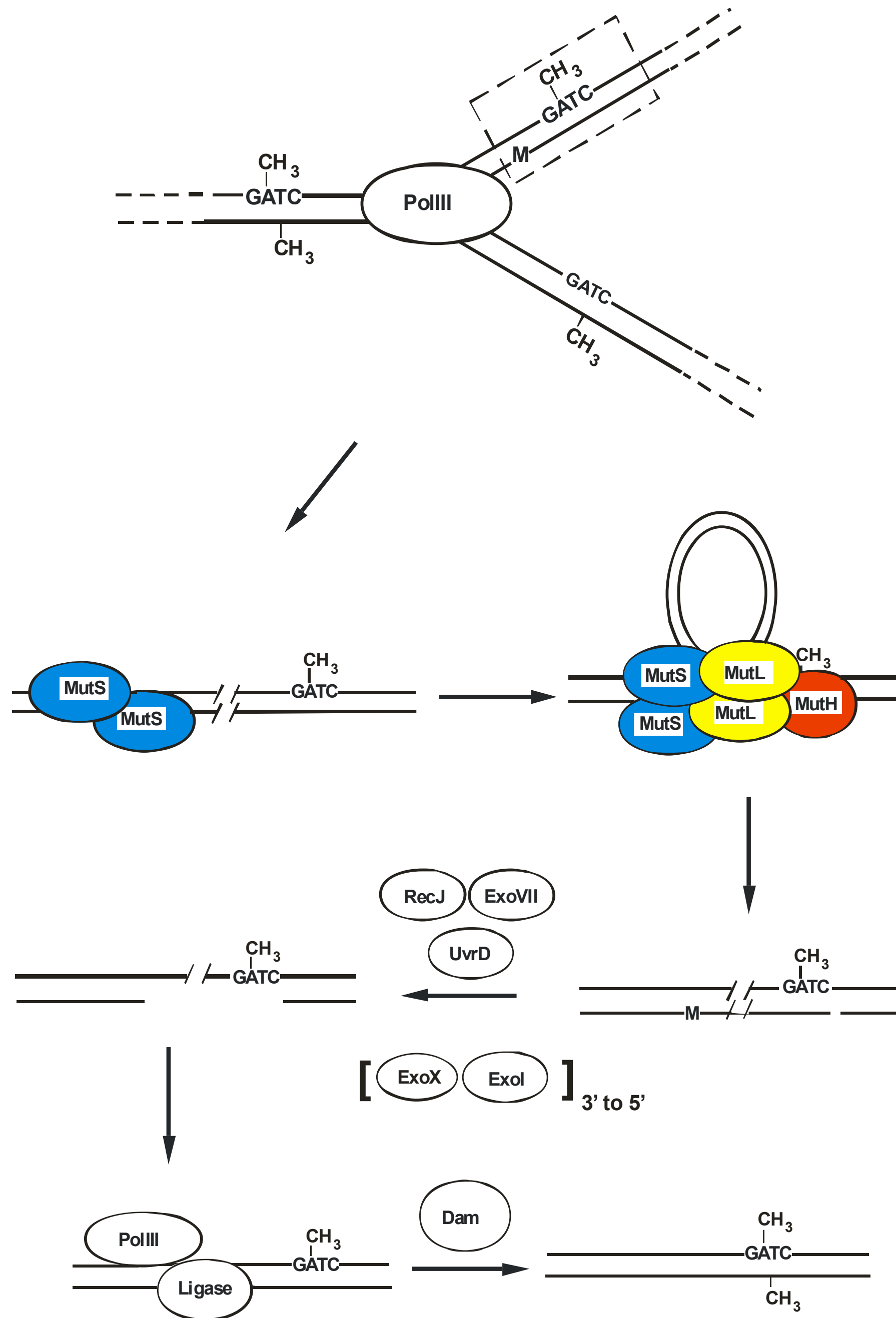


Fig. 1. Model for *dam*-directed mismatch repair. The model is based on the in vitro properties of the MutS, MutL and MutH proteins. The stoichiometry and the nature of protein-protein interactions are not understood in detail. A replication fork is shown at the top of the figure with DNA polymerase III holoenzyme (PolIII), methylated or unmethylated GATC (*dam*) sequences and a replication mistake (M) generating a base-base or deletion/insertion mismatch. The boxed area is expanded in the rest of the figure. The mismatch is bound by MutS and, in an ATP-dependent reaction, a looped structure is formed with MutS, MutL and MutH proteins at the base of the loop. Incision by activated MutH occurs in the newly synthesized strand at an unmethylated GATC sequence. A gap is formed by the action of exonucleases including exonuclease I (ExoI), ExoVII, ExoX and RecJ and the direction of excision is determined by the UvrD helicase. Excision can be in either the 5' to 3' or 3' to 5' direction; only one is shown. Resynthesis is accomplished by DNA polymerase III holoenzyme and the nick sealed by DNA ligase. Subsequent methylation by DNA adenine methyltransferase (Dam) completes the process.

D. Mutator Genes and DNA Repair.

D1. MutM and MutY. MutM and MutY are glycosylases that act on oxidized guanine residues in DNA (See Fig. 2). In the absence of both MutM and MutY, a several thousand-fold increase in the frequency of CG to AT mutations is observed, whereas a much lower frequency is detected in cells lacking either *mutM* or *mutY*. Genetic experiments have shown that overproduction of MutM suppresses the mutability of a *mutY* mutant, indicating that MutM acts before MutY (Fig. 2).

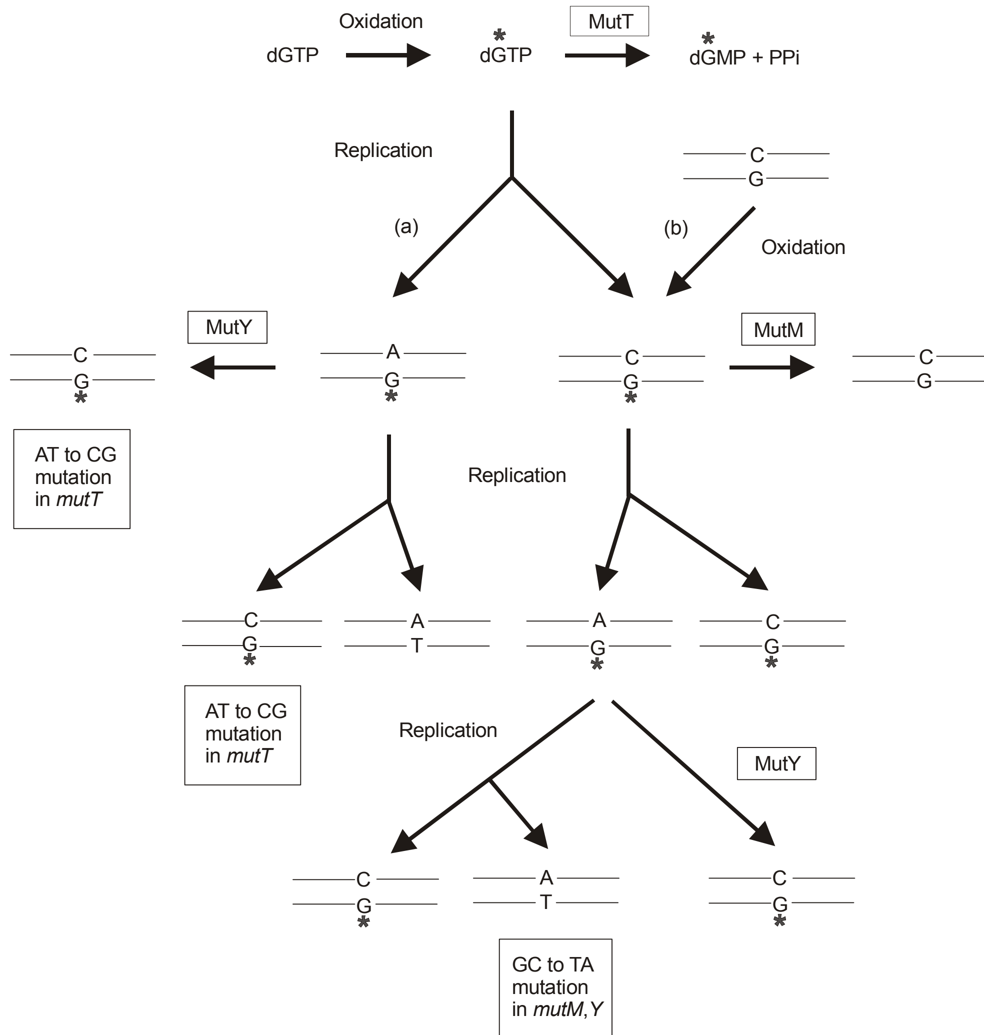
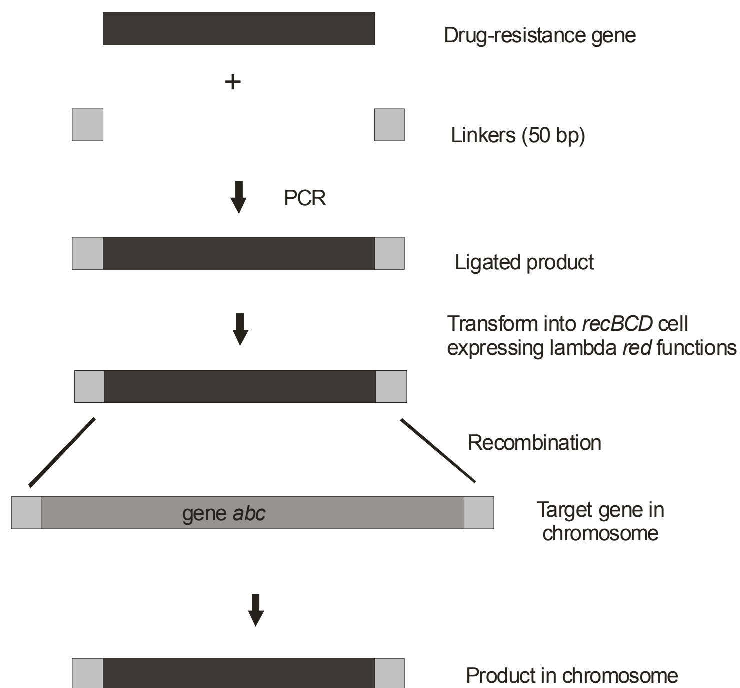


Fig. 2. Model for the proposed action of MutT, MutM and MutY. *G represents 8-oxo-G. The top line indicates that MutT hydrolyses d*GTP thereby preventing its incorporation. The next line shows that oxidation of G in DNA can occur and its removal by MutM. Incorporation of d*GTP can form *GC or *GA pairs. Replication of *GC produces *GC and *GA. MutY converts *GA to *GC. Loss of MutT increases formation of *GA base pairs eventually leading to AT to CG mutations. Loss of MutM and MutY increases the probability of *GC replication yielding *GC and *GA; the latter in turn produces TA base pairs leading to GC to TA mutations.

D2. Other Mutator Genes.
The mutator genes we have discussed so far are the most important in terms of their effects on spontaneous mutation rates. There are other "weaker" mutator genes whose products are involved in repairing DNA damage and these are listed in Table 1 together with the "strong" mutator genes.

The *recG* mutator effect (several hundred fold) is seen only in non-growing cells ("adaptive mutation"). At present the only known functions for RecG are in processing recombination intermediates and in repair of stalled DNA replication forks.

E. Induced Mutation
There are various ways to induce mutations into the *E. coli* genome. A traditional way is the use of agents such as ultraviolet (UV) light and chemicals (e.g., EMS [ethyl methane sulfonate] and MNNG [methyl-nitro-nitrosoguanidine]) to induce mutations at random sites followed by a genetic selection and/or screening. These agents produce mainly GC to AT transitions. An alternative is to use transposons with associated drug markers (tetracycline, kanamycin, etc.) to allow easy mapping of the insertion site. The latest method allows for transformation of a particular strain (expressing the phage lambda recombination system) with a mutagenic single-stranded oligonucleotide or double-stranded PCR product (Ellis et al., 2001).



Scheme to replace gene *abc* on the chromosome with a drug-resistance marker (after Murphy, K).

There are also a variety of techniques for introducing mutations at specific base pairs in genes on plasmids (many commercially available kits). Alternatively, mutations (frequently more than one) can be produced in such genes by amplifying the gene sequence by mutagenic PCR followed by sub-cloning into the vector plasmid. Yet another alternative is to propagate the plasmids through mutator strains.

Table 1. Mutator genes in *E. coli*

Gene	Gene product	Function	Specificity
dnaQ *(mutD)*	ἐ subunit of DNA polymerase III, 3'-> 5' exonuclease	Removes incorrectly paired nucleotides during replication (proofreading)	Mostly transversions
polC *(dnaE)*	ά subunit of DNA polymerase III	Correct nucleotide selection and proofreading	All base substitutions and frameshifts
polA	DNA polymerase I	General repair functions	Deletions, frameshifts
mutT	Nucleoside triphosphatase	Prevents mispairing of 8-oxoG with template A during replication	AT -> CG
dam	DNA adenine methyltransferase	Methylation imparts strand specificity	GC -> AT, AT -> GC frameshifts
mutS	DNA mismatch recognition, ATPase	Binds DNA mismatches	GC -> AT, AT -> GC frameshifts
mutL	ATPase	Stimulates MutS, MutH and Vsr activity	GC -> AT, AT -> GC frameshifts
mutH	Endonuclease	Nicks hemi-methylated GATC sequences	GC -> AT, AT -> GC frameshifts
uvrD	DNA helicase II	Strand displacement	GC -> AT, AT -> GC frameshifts
mutY	DNA glycosylase	Removes A from 8-oxoG-A or A-G mispairs	GC -> TA
mutM (*fpg*)	DNA glycosylase	Removes 8-oxoG from 8-oxoG-C mispair	GC -> TA
sodA, sodB	Superoxide dismutases	Removes superoxide radicals	
oxyR	Regulatory protein	Regulates hydrogen peroxide-inducible genes	AT -> TA
nth, nei	Glycosylase and AP-lyase activity	Removal of oxidized pyrimidine bases	GC -> AT
xthA, nfo	Nucleases	5' AP-endonuclease activity	AT -> TA
ung	Uracil glycosylase	Removes U from U-G mispairs	GC -> AT
vsr	Endonuclease	Cleaves adjacent to T-G mismatches	GC -> AT
ada, ogt	Methyltransferase	Removes methyl groups from O^6-methylguanine in DNA	GC -> AT
recA	DNA binding protein	Catalyzes strand pairing and exchange in general recombination; co-protease activity on LexA, UmuD, etc.	GC -> TA AT -> TA
recG	DNA helicase	Branch migration of Holliday junctions	Frameshifts
hns (*bglY*)	DNA binding protein	Histone-like protein involved in chromosome organization	Deletions
topB (*mutR*)	DNA topoisomerase III	Appears to decatenate chromosomes	Deletions between small repeats
ssb	Single-stranded DNA binding protein	Protects single-stranded DNA	Point mutations, re-arrangements