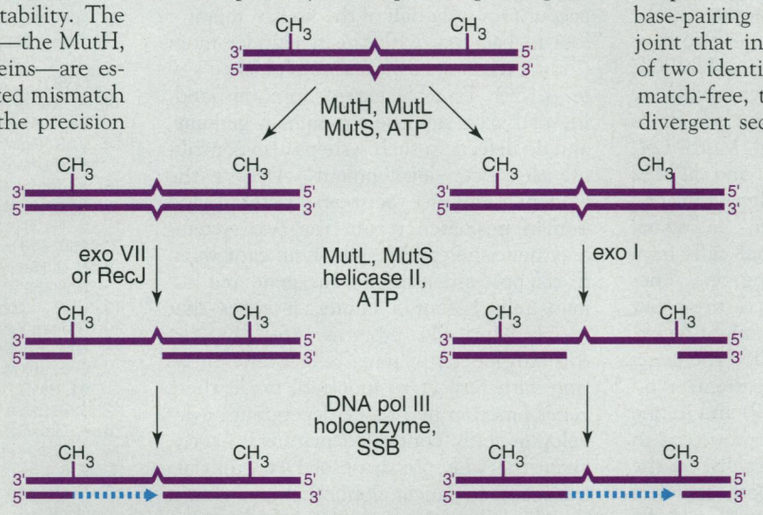
Mismatch Repair, Genetic Stability, and Cancer

Paul Modrich

Because mutations are the primary cause of heritable disease and cancer and may also contribute to the aging process, cells go to great lengths to preserve the integrity of their genetic material. Study of the molecular systems responsible for maintenance of genetic integrity became available more than 20 years ago with the identification of *Escherichia coli* mutants that display a marked increase in the rate of spontaneous mutation (*1*). The genetic defects in these strains are now known to inactivate mutation avoidance systems that are critical for the maintenance of genetic stability. The products of four of these genes—the MutH, MutL, MutS, and MutU proteins—are essential for *E. coli* methyl-directed mismatch repair, a system that ensures the precision of both chromosome replication and genetic recombination. The importance of this fidelity device has been dramatically illustrated during the past year with the demonstration that inactivation of the corresponding human pathway is the primary cause of certain types of cancer.

Replication and recombination errors produce base-pairing anomalies within the DNA helix, mismatches that violate the Watson-Crick pairing rules—which specify that the purine bases A or G on one strand pair with the pyrimidine bases T or C, respectively, on the other. The process of DNA biosynthesis, although highly precise, is intrinsically imperfect. Relatively common DNA biosynthetic errors include insertion of an incorrect base; for example, T opposite G, or the addition of an extra nucleotide or two, resulting in unpaired bases within the helix. It is the job of the cellular mismatch repair system to recognize such mispairs and to eliminate biosynthetic mistakes from newly synthesized DNA strands.

Because mismatches consist of normal Watson-Crick bases, mismatch repair systems rely on secondary signals within the helix to identify the newly synthesized DNA strand, which by definition contains the replication error. The requisite strand-specificity for processing of replication errors in *E. coli* is provided by patterns of adenine methylation at GATC sequences. Because GATC modification occurs after DNA strand synthesis, newly synthesized DNA exists briefly in an unmethylated state, and it is this transient absence of modification that targets repair to the new DNA strand (*2*). The mechanism of replication error correction by the methyl-directed pathway is complex, depending on 10 activities (see figure). Repair is initiated by binding of MutS to the mismatch, followed by the addition of MutL (*3*). Assembly of this complex leads to activation of a latent GATC endonuclease associated with the MutH protein, which incises the unmodified strand at a hemimethylated d(GATC) sequence (*4*). The resulting strand break can occur on either side of the mismatch. The ensuing excision reaction, which depends on MutS, MutL, and the cooperative action of DNA helicase II (the MutU protein) with an appropriate exonuclease, removes that portion of the unmodified strand spanning the GATC site and the mismatch. This reaction is strictly exonucleolytic, initiating at the strand break and proceeding toward the mispair without regard to location of the strand break (*5*). This unusual bidirectional excision capability implies that the methyl-directed system keeps track of which side of the mispair the strand-signal is located.



The mechanism of *E. coli* methyl-directed mismatch repair. Although not shown, DNA ligase is required to restore covalent integrity to the repaired DNA strand. [Reproduced with permission from M. Grilley *et al.* (*5*)]

In addition to their role in correction of DNA biosynthetic errors, MutS and MutL also ensure the fidelity of genetic recombination by blocking crossovers between sequences that have diverged genetically (*6*). Action in this manner prevents crossovers between related sequence elements that are present in multiple copies in the bacterial genome and hence the duplication and deletion mutations that result from such events (*7*). The molecular mechanism underlying these recombination effects of MutS and MutL is less well understood, but undoubtedly involves interaction of the proteins with mispairs that can occur in the heteroduplex joint, a key recombination intermediate generated by transfer of a strand from one helix into a region of homology of a second where it pairs with its complement according to Watson-Crick base-pairing rules. While a heteroduplex joint that involves corresponding regions of two identical chromosomes will be mismatch-free, the structure formed between divergent sequences will contain multiple mispairs. Indeed, in vitro experiments have shown that MutS and MutL block the progression of heteroduplex formation between two DNAs that have diverged by several percent at the sequence level (*8*).

Given the importance of mismatch repair in stabilizing the bacterial genome, it is not surprising that homologous systems have been identified in higher cells. Yeast (*9*, *10*) and (as described below) human cells encode homologs of bacterial MutS and MutL. Furthermore, nuclear extracts of human cells support strand-specific mismatch correction in a reaction that is remarkably similar to methyl-directed repair with respect to both mismatch specificity (*11*, *12*) and bidirectional excision capability (*13*). Initial evidence that this strand-specific pathway functions in mutation avoidance in human cells was provided by the demonstration that a hypermutable, cultured cell line is defective in mismatch repair (*14*). This mutator cell line was isolated in vitro by virtue of its ability to survive the presence of DNA lesions resulting from exposure to simple alkylating agents, lesions that otherwise kill normal human cells (*15*). In addition to indicating a general role for mismatch repair in stabilization of the human genome, these findings thus suggest that the system recognizes lesions other than

The author is in the Department of Biochemistry and the Howard Hughes Medical Institute, Duke University Medical Center, Durham, NC 27710, USA.

conventional mispairs and implicate the pathway in the death of those cells that have suffered unacceptable levels of certain types of DNA chemical damage.

The clinical ramifications of genetic destabilization have been dramatically demonstrated by the finding that certain sporadic cancers (*16–18*), and virtually all tumors associated with hereditary nonpolyposis colorectal cancer (HNPCC) (*19*) (one of the most common syndromes that causes cancer predisposition), are highly prone to mutation. Like bacterial mismatch repair mutants, cell lines derived from such tumors accrue mutations at rates that can be more than a hundred times that of normal human cells (*20–22*), and biochemical analysis of a number of these lines has consistently revealed an associated deficiency in strand-specific mismatch repair (*20, 23*).

Compelling evidence that loss of mismatch repair proficiency is the primary step in development of HNPCC tumors has been provided during the past year by the spectacular elucidation of the genetic basis of the disease. The majority of HNPCC cases are attributable to a defect at any one of four loci: The *hMSH2* gene encodes a protein homolog of bacterial MutS (*24, 25*), while *hMLH1*, *hPMS1*, and *hPMS2* specify distinct MutL homologs (*26–28*). HNPCC is inherited in an autosomal dominant fashion, with normal cells from affected individuals containing one functional and one defective copy of the repair gene in question (*25–28*). As might be expected from presence of a wild-type gene, normal cells from HNPCC patients typically display low mutability (*19*) and in one case have been shown to be proficient in mismatch repair (*20*). Tumor cells, on the other hand, are defective in both copies of the affected gene, with inactivation of the wild-type allele as a result of somatic mutation (*25, 28*). Coupled with the demonstrable repair defect in cancer cells, this key observation implies that the initial event in development of HNPCC tumors is the functional loss of a critical mismatch repair activity, with the resultant genetic destabilization presumably leading to mutations that circumvent the regulatory systems that control cell proliferation.

The study of fidelity devices in bacteria provided the groundwork for this rapid progress, and the similarities between bacterial and human systems raise several interesting questions. Characterization of mutations that occur in mismatch repair–deficient tumor cells has been limited, with those identified to date probably caused by replication errors (*19, 21*). Because bacterial strains deficient in MutS or MutL are also prone to recombination errors, do HNPCC tumor cells display a similar instability with respect to illegitimate recombination and do such events contribute to tumor development? Mismatch repair genes account for only half of the known mutator loci in bacteria, with the remainder functioning in distinct mutation avoidance systems (*29*). To what extent do corresponding pathways stabilize the human genome, and do defects in such systems also contribute to cancer development? Despite the evident similarity between bacterial and human mismatch repair, the two systems may nevertheless differ in significant ways. *E. coli* possesses a single *mutL* gene, but human cells harbor a family of genes that specify MutL-like proteins (*26–28*). Does this multiplicity of genes reflect differentiation with respect to function, or do these genes function in a tissue-dependent or developmentally controlled manner? Lastly, given the widespread use of DNA alkylating agents in cancer chemotherapy, what is the clinical significance of the finding that mismatch repair–deficient cells are resistant to killing by at least one class of such agents? With the current renaissance in DNA repair, it seems likely that answers to these and related questions will be forthcoming in the near future.

References

1. E. C. Cox, *Annu. Rev. Genet.* **10**, 135 (1976).
2. M. Meselson, in *Recombination of the Genetic Material*, K. B. Low, Ed. (Academic Press, San Diego, 1988), pp. 91–113.
3. M. Grilley, K. M. Welsh, S.-S. Su, P. Modrich, *J. Biol. Chem.* **264**, 1000 (1989).
4. K. G. Au, K. Welsh, P. Modrich, *ibid.* **267**, 12142 (1992).
5. M. Grilley, J. Griffith, P. Modrich, *ibid.* **268**, 11830 (1993).
6. C. Rayssiguier, D. S. Thaler, M. Radman, *Nature* **342**, 396 (1989).
7. M.-A. Petit, J. Dimpfl, M. Radman, H. Echols, *Genetics* **129**, 327 (1991).
8. L. Worth, S. Clark, M. Radman, P. Modrich, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 3238 (1994).
9. W. Kramer, B. Kramer, M. S. Williamson, S. Fogel, *J. Bacteriol.* **171**, 5339 (1989).
10. R. A. G. Reenan and R. D. Kolodner, *Genetics* **132**, 963 (1992).
11. J. Holmes, S. Clark, P. Modrich, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 5837 (1990).
12. D. C. Thomas, J. D. Roberts, T. A. Kunkel, *J. Biol. Chem.* **266**, 3744 (1991).
13. W.-H. Fang and P. Modrich, *ibid.* **268**, 11838 (1993).
14. A. Kat *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6424 (1993).
15. V. S. Goldmacher, R. A. Cuzick, W. G. Thilly, *J. Biol. Chem.* **261**, 12462 (1986).
16. Y. Ionov, M. A. Peinado, S. Malkhosyan, D. Shibata, M. Perucho, *Nature* **363**, 558 (1993).
17. S. N. Thibodeau, G. Bren, D. Schaid, *Science* **260**, 816 (1993).
18. H.-J. Han, A. Yanagisawa, Y. Kato, J.-G. Park, Y. Nakamura, *Cancer Res.* **53**, 5087 (1993).
19. L. A. Aaltonen *et al.*, *Science* **260**, 812 (1993).
20. R. Parsons *et al.*, *Cell* **75**, 1227 (1993).
21. N. P. Bhattacharyya, A. Skandalis, A. Ganesh, J. Groden, M. Meuth, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 6319 (1994).
22. J. R. Eshleman *et al.*, *Oncogene*, in press.
23. A. Umar *et al.*, *J. Biol. Chem.* **269**, 14367 (1994).
24. R. Fishel *et al.*, *Cell* **75**, 1027 (1993).
25. F. S. Leach *et al.*, *ibid.*, p. 1215.
26. N. Papadopoulos *et al.*, *Science* **263**, 1625 (1994).
27. C. E. Bronner *et al.*, *Nature* **368**, 258 (1994).
28. N. C. Nicolaides *et al.*, *ibid.* **371**, 75 (1994).
29. H. Echols, *Annu. Rev. Biochem.* **60**, 477 (1991).