A personal historical view of DNA mismatch repair with an emphasis on eukaryotic DNA mismatch repair

Permalink
https://escholarship.org/uc/item/34t0k1dx

Author
Kolodner, RD

Publication Date
2015-10-23

DOI
10.1016/j.dnarep.2015.11.009

Peer reviewed
A personal historical view of DNA mismatch repair with an emphasis on eukaryotic DNA mismatch repair

By

Richard D. Kolodner

From

Ludwig Institute for Cancer Research, Department of Cellular and Molecular Medicine, Moores-UCSD Cancer Center and Institute for Molecular Medicine, University of California, San Diego School of Medicine. La Jolla, CA 92093-0669

Address correspondence to:

Richard D. Kolodner
rkolodner@ucsd.edu
(858) 534-7804 (phone)
(858) 534-7750 (fax)
1. The early days of mismatch repair

I first learned about the existence of DNA mismatch repair (MMR) in the early 1970's as a graduate student from Robert "Bob" C. Warner, who was one of the few people in our department whose lab worked on DNA. Bob was interested in isolating and proving the existence of Holliday junction-containing DNA molecules, the homologous recombination (HR) intermediate Robin Holliday had postulated in 1964 to form mispaired bases that could then account for some aspects of gene conversion when they were acted on by MMR [1]. I only later became aware of parallel studies on the possibility that MMR functioned in mutagenesis and replication fidelity [2]; the roles of mismatch repair in suppressing HR between divergent DNA and in the DNA damage response were discovered many years later (Discussed below). My many stimulating discussions with Bob led me to purify the only restriction endonuclease known at the time, the Hind II+III mixture, for use by his lab in their studies of HR intermediates. This experience led me away from my graduate studies with Krishna Tewari with whom I had worked performing quite challenging studies on the molecular characterization of chloroplast DNA to doing postdoctoral work on the enzymology of DNA replication. While I really enjoyed my postdoctoral studies on bacteriophage T7 DNA replication proteins and felt I was learning about protein purification from a true master, Charles Richardson, I also felt this was not the area I wanted to establish my own laboratory in. This rekindled my interest in HR and MMR and Charles was extremely generous to let me spend
part of my time using new methods of site-directed mutagenesis to make mutations that altered restriction endonuclease cleavage sites in plasmid DNAs for use in making DNA substrates that I planned to use to study the genetics and biochemistry of HR and MMR. I cannot emphasize enough that I had three great mentors and how they each helped to educate me and aided in the advancement of my career. In this vein, I am also pleased that so many of my own trainees have gone on to establish their own successful labs.

When I moved to the Dana-Farber Cancer Institute in 1978 to start my own lab, I began studying plasmid recombination and, to a lesser extent, MMR in Escherichia coli, areas that I really had never worked in before. Given this, it seems amazing that I was able to obtain an academic position and obtain funding for my lab; I think there is a lesson in my career experiences that I hope current day research institutions and funding agencies will reflect on leading them to adjust their ways to make it easier than presently seems possible for young scientists to go in new directions. During the 1960’s, 1970’s and early 1980’s the genetic basis for the model for E. coli methyl-directed MMR was worked out, to a great extent through the efforts of the Meselson, Marinus and Radman labs as well as others [3-8]. This system was thought to act on mispairs that were formed as a result of DNA replication errors as well as on mispairs that were formed in heteroduplex HR intermediates when mutant and wild-type single strands were paired with each other [9, 10]. The key MMR genes including mutS, mutL, mutH, uvrD and dam were identified through genetic studies [9, 10].
Transient under-methylation of the newly synthesized DNA strand was proposed to be the strand specificity mark during post-replication MMR and some studies suggested that MutS might be involved at the mispair recognition step and that MutL and MutH might be involved in strand discrimination [9-12]. These studies as well as the availability of the replication proteins and exonucleases set the stage for the subsequent reconstitution studies from the Modrich lab [13, 14] and mechanistic studies of individual *E. coli* MMR proteins from the Modrich lab and many other labs (For representative reviews see [15, 16]). In retrospect, it is ironic that the studies of *E. coli* MMR have been so influential in the MMR field given that methyl-directed post-replication MMR only occurs in a small group of bacteria (see the minireview by Putnam in this issue). Below I provide a personal overview of eukaryotic MMR including my views of the history of MMR, what we know about MMR mechanisms, the role of MMR defects in cancer and questions that remain to be answered. However, I refer the reader to the work of the authors contributing articles to this volume as well as other papers for the details and the many aspects of MMR, including *E. coli* MMR, that I have not had space to cover in detail.

The nomenclature for MMR genes/proteins is confusing because the human homolog of *Saccharomyces cerevisiae* PMS1 was named PMS2 and the human homolog of *S. cerevisiae* MLH2 was named PMS1. In this article, I have generally used *S. cerevisiae* gene/protein names. However, in places where it was important to specifically refer to a human or mouse gene/protein, I have
included the suffixes sc (S. cerevisiae), h (human) and mus (mouse) to help prevent confusion. A useful summary table of MMR genes, proteins and protein complexes has been published recently [17] (see the minireview by Putnam in this issue).

2. Identification of eukaryotic mismatch repair genes and proteins

Having already developed studies of HR in the yeast S. cerevisiae, my lab initiated studies on S. cerevisiae MMR because of two observations suggesting that eukaryotic MMR might be quite different from E. coli MMR. First, S. cerevisiae was found to have no DNA methylation and therefore could not have methyl-directed MMR [18, 19]. Second, with the development of the double strand break repair (DSB/DSBR) model of HR [20], it was proposed that DSBR could account for the relationship between gene conversion and post-meiotic segregation of genetic markers without the need for classical MMR [20]. Thus, it was proposed that gene conversion would occur when a genetic marker was located in the region near the DSB that was degraded during DSBR to form a double strand gap eliminating the mutant or wild-type allele, and post-meiotic segregation would occur when the genetic marker was located in the region of heteroduplex DNA flanking the DSB during DSBR, thus resulting in a mispair that was then not acted on by MMR. Further, post-replication MMR could be explained if an endonuclease were to make a DSB at the site of a mispair followed by DSBR with the sister DNA serving as a donor during repair [21];
endonucleases that could cleave at mispairs had been discovered [22]. This model was in contrast to more classical models where heteroduplex HR intermediates containing mispaired bases were formed and were then either repaired by MMR resulting in gene conversion or escaped repair resulting in post-meiotic segregation of the genetic marker [20]. A model illustrating these views of HR is presented in Figure 1.

To investigate whether *S. cerevisiae* had classical MMR, we adapted our previously developed *E. coli* heteroduplex plasmid transformation system [11, 12] to the study of MMR in *S. cerevisiae*. On transformation of *S. cerevisiae* with heteroduplex plasmids, we recovered plasmid products that were consistent with MMR having acted on these plasmids [23, 24]. We also showed that cell-free extracts of *S. cerevisiae* were able to catalyze repair of mispair-containing M13 phage-based heteroduplex substrates [25]; other eukaryotic cell-free MMR systems were subsequently developed by others [26-28]. At this time we were aware of mutations identified in Seymour Fogel’s lab named *pms* mutations, for Post-Meiotic Segregation, which caused a decrease in gene conversion events and a concomitant increase in post-meiotic segregation events [29]. Interpreted in the context of classical MMR, these results were consistent with *pms* mutations causing a MMR defect. However, interpreted in the context of the DSBR model, these results were consistent with *pms* mutations causing a decrease in the size of the double strand gap and an increase in the length of the heteroduplex regions flanking the double strand gap increasing the chance that
the position of the allele being followed would fall in the heteroduplex region flanking the DSB. We obtained pms1 mutations from the Fogel lab and used our heteroduplex plasmid transformation assays to show that pms1 mutations caused increased recovery of plasmid products that had escaped repair, thus establishing pms1 mutations as causing a MMR defect and defects in processing mispairs formed during meiotic recombination to gene conversion events resulting in increased post-meiotic segregation events [30]. Our subsequent studies of Seymour’s pms1 through pms6 mutations were of great value in helping to identify and study MMR genes. As a philosophical note, I have found trying to address this type of unanswered, possibly even somewhat controversial scientific question to be the most intellectually rewarding part of being a scientist.

Seymour’s lab went on to clone the PMS1 gene and discovered that it encoded a MutL homologue, a critical protein in E. coli MMR [31]. This was extremely surprising to me as we thought that eukaryotes did not have methyl-directed MMR. This motived my lab to use newly developed methods of degenerate PCR to search for S. cerevisiae genes encoding MutS homologues. We were able to initially identify two such MutS Homologue genes, MSH1 and MSH2 [32, 33]. Genetic studies suggested that MSH1 encoded a protein that functioned in the repair of mitochondrial DNA [33]; this was surprising because the prevailing view at the time, now no longer held, was that mitochondria lacked DNA repair systems. In contrast, genetic studies showed that msh2 defects caused increased nuclear mutation rates and increased post-meiotic segregation of
genetic markers during meiosis establishing Msh2 as the first identified eukaryotic MutS related protein [33]; MSH2 was later found to be the gene mutated in Fogel's pms5 mutant [34]. The Crouse laboratory identified MSH3, which was also known in humans and mouse, but msh3 defects were not initially found to cause phenotypes consistent with MMR defects [35]. We also identified MSH6 by analyzing the S. cerevisiae genome sequence and found that it was the gene mutated in Fogel's pms3 and pms6 mutants, which had not yet been extensively studied [36, 37]. The MSH4 and MSH5 genes (see the minireview by Manhart and Alani in this issue), encoding proteins that function in meiotic recombination but not MMR, were described by others [38, 39] and will not be discussed in this article. A key clue to the Msh proteins came from studies of human cell-free extracts that catalyzed MMR performed in the Modrich and Jiricny labs, that identified the in vitro human msh2 MMR defect-complementing activity as a two subunit protein complex containing Msh2 and a fragment of GTBP (also called Msh6) [40, 41]. My lab performed two-hybrid interaction studies showing that Msh2 was the common subunit of the Msh2-Msh3 and Msh2-Msh6 complexes [36]. We further showed that these complexes were partially redundant with each other, consistent with the idea based on the function of bacterial MutS that these complexes were mispair recognition proteins with partially overlapping mispair-binding specificity [36]. Other genetic studies provided additional details about the predicted mispair recognition specificity of these two complexes [42]. Collaborative biochemical studies with the Fishel lab as well as biochemical studies by others quickly extended this model to the
human MSH complexes [43, 44] and studies of mutant mice and mouse cell lines as well as human tumor cell lines provided supporting genetic evidence [45-49].

The discovery and characterization of the MutL Homologue or MLH complexes was similar to that of the MSH complexes. *S. cerevisiae MLH1* was discovered by the Liskay lab using degenerate PCR similar to the discovery of *MSH2* [50] and later found to be the gene mutated in Fogel’s *pms2* and *pms4* mutants [51]. Genetic studies showed that defects in *MLH1* and *PMS1* caused very similar mutator phenotypes [50] and, consistent with this early biochemical studies, established the existence of a Mlh1-Pms1 complex that could be recruited by Msh2 [52]. Protein fractionation studies identified the human *in vitro* *mlh1* MMR defect-complementing activity as a two-subunit complex containing Mlh1 and Pms2 (Pms2 is the human homologue of *S. cerevisiae* Pms1) [53]. The other two *S. cerevisiae MLH* genes, *MLH2* and *MLH3* were identified through the analysis of the *S. cerevisiae* genome sequence as well as using two-hybrid screens [54-56]. Like Pms1, the Mlh3 protein forms a complex with Mlh1, and genetic studies have indicated that the Mlh1-Mlh3 complex plays a minor role in MMR by substituting for Mlh1-Pms1 in the repair of some types of mispairs [55, 56], as have biochemical studies of hMlh1-Mlh3 [57]. The Mlh1-Mlh3 complex also plays a role in meiotic recombination in conjunction with the Msh4-Msh5 complex [54, 58, 59]. The Mlh1-Mlh2 complex has remained enigmatic. Genetic studies in both *S. cerevisiae* and mouse as well as biochemical complementation studies utilizing human *in vitro* MMR reactions initially provided little evidence for
a role in MMR [56, 60, 61] although some studies have suggested Mlh2 might play a minor role in meiotic recombination [54, 59]. However, recent studies have indicated that Mlh1-Mlh2 is more important for MMR when Mlh1-Pms1 levels are reduced suggestive of a different role in MMR than the other MutL homologue complexes [62].

The other genes/proteins thought to be involved in MMR include Exonuclease 1 (Exo1), Proliferating Cell Nuclear Antigen (PCNA), Replication Factor C (RFC), Replication Factor/Protein A (RPA) and DNA polymerase δ (DNA Pol δ). Exo1 was first described as a meiotic HR 5’ to 3’ exonuclease in S. pombe [63, 64]. A role for Exo1 in MMR in S. cerevisiae was discovered through the isolation of EXO1 as encoding an Msh2-interacting protein in a two-hybrid screen [65] and was subsequently found to also interact with Mlh1 [66, 67]. While Exo1 is thought to play a key role in the excision step of MMR, genetic studies have shown that Exo1 is not essential for MMR and that other excision mechanisms must exist [65, 68, 69]. PCNA was implicated in MMR through its role in the resynthesis step of MMR, through protein-protein interaction studies, and the identification of mutations in the S. cerevisiae POL30 gene encoding PCNA, which cause MMR defects and a mutator phenotype [70-73]. Subsequently, PCNA was shown to play a role in coupling MMR to DNA replication through interactions with PIP (PCNA-interacting peptide)-box motifs located in the N-termini of Msh3 and Msh6 [74-77]. PCNA also interacts with scMlh1-Pms1 and hMlh1-Pms2 [78, 79], and this interaction could play a role in activating the
endonuclease activity of these 2 complexes [80, 81]. The PCNA loading factor RFC was implicated in MMR through the requirement for PCNA, and the identification of a MMR-defective mutation in one of the genes encoding RFC has provided genetic support for this requirement [82]. DNA Pol δ was implicated in MMR through in vitro reconstitution studies [83], but a possible role for other DNA polymerases has not been ruled out. Similarly, RPA was implicated in MMR through in vitro reconstitution studies [84-86]. Some studies have suggested that HMGB, PARP1, RFX and most recently the MCM8-MCM9 complex may play a role in MMR, but the evidence for these is in my view not yet well established [87-90]. A model summarizing the known MMR proteins is presented in Figure 2.

3. Biochemical activities of MMR proteins

The presently known MMR proteins fall into two groups. One consists of proteins that function in many aspects of DNA metabolism including DNA replication and recombination, such as Exo1, RPA, PCNA, RFC and DNA Pol δ. These proteins have been extensively characterized in a variety of contexts and will not be generally commented on here. The second consists of proteins that are more specific to MMR, including the MutS-related heterodimers Msh2-Msh6 and Msh2-Msh3 and the MutL-related heterodimers Mlh1-Pms1 (Mlh1-Pms2 in human), Mlh1-Mlh3 and Mlh1-Mlh2 (Mlh1-Pms1 in human) (see the minireview by Putnam in this issue). Many of these proteins have been extensively characterized using a diversity of techniques beyond more standard biochemical studies including X-
ray crystallography, small angle X-ray scattering, deuterium-exchange mass spectrometry, imaging proteins in vivo and single molecule biochemistry approaches, many of which have been coupled to different types of genetics studies [91, 92](see the minireviews by Groothuizen and Sixma, by Friedhoff and Gotthardt and by Schmidt and Hombauer in this issue). The application of these approaches has been exciting to follow and has brought many new insights to the field.

The Msh2-Msh6 and Msh2-Msh3 mispair-recognition complexes have been shown to form a ring around DNA and under certain conditions form a sliding clamp structure, a conformation that the Fishel lab has extensively studied and incorporated into specific models for MMR mechanisms [93, 94] (see the minireview by Hingorani in this issue). In the absence of nucleotide or in the presence of ADP, these mispair recognition factors appear to form an unstable ring on base paired DNA and similarly form an unstable ring at the site of the mispair in mispaired DNA [95]. On addition of ATP, the proteins bound to base paired DNA dissociate whereas the mispair-bound proteins are converted to a stable clamp that then slides along the DNA and, on linear DNAs, dissociates from the ends of the DNA [93, 96-98]. The pioneering X-ray crystallography studies of the bacterial MutS proteins performed by Sixma [99] as well as Hsieh and Yang [100], followed later by structural studies of the eukaryotic Msh complexes [101, 102] as well as structure-function genetics studies have contributed enormously to the understanding of the mispair-bound forms of these
proteins. The Msh2-Msh6 and Msh2-Msh3 complexes can also recruit Mlh1-Pms1 and Mlh1-Mlh2, and most likely Mlh1-Mlh3, in a mispair-dependent, ATP-dependent reaction corresponding to the conditions where sliding clamps form [62, 95, 97, 103-107]. However, while mutations that disrupt sliding clamp formation also disrupt MMR [37, 108, 109], it is not clear at which step of MMR sliding clamps are required because at least one sliding clamp-defective Msh2-Msh6 mutant is proficient for recruiting and activating the Mlh1-Pms1 endonuclease and promotes Mlh1-Pms1 endonuclease-dependent MMR reactions in vitro [110]. It is also possible that sliding clamps reflect an ATP-induced conformational change required for MMR with sliding clamps being predominantly seen in the absence of other MMR proteins. Many studies have suggested that a stable 1:1 complex of an Msh complex with an Mlh complex plays a role in MMR (for representative papers see [97, 103, 104, 106, 107]). In contrast, imaging of MMR proteins in living cells has suggested that Msh2-Msh6, and likely Msh2-Msh3, can act to catalytically load multiple molecules of Mlh1-Pms1 and Mlh1-Mlh2 (and possibly Mlh1-Mlh3) onto DNA in a mispair-dependent reaction [62, 77]. Determining if the Msh complexes can catalytically load Mlh complexes onto DNA in reconstituted biochemical reactions and then elucidating the biochemical activities of the Msh complex-free Mlh complexes will be critical to decipher the mechanism of the early steps in MMR.

The function of the eukaryotic MutL-related complexes, Mlh1-Pms1, Mlh1-Mlh3 and Mlh1-Mlh2 remained poorly understood for approximately twelve years after
the first such complex, Mlh1-Pms1, was discovered. Then the Modrich lab demonstrated that both the human hMlh1-Pms2 complex and its S. cerevisiae homolog scMlh1-Pms1 have an endonuclease activity [70, 111] (see the minireview by Kadyrova and Kadyrov in this issue). The endonuclease active site has been extensively characterized by structure-function genetics and X-ray crystallography [110, 112, 113]. This endonuclease has very weak activity on its own but can be activated to nick supercoiled, base paired DNA by PCNA and RFC and can be more strikingly activated by Msh2-Msh6 (or Msh2-Msh3), PCNA and RFC to nick mispaired DNA containing a pre-existing nick. Remarkably, in this latter 4-protein nicking reaction, the Mlh1-Pms1 complex is highly specific for nicking the already nicked strand [70, 80, 110, 111]. Not only is the molecular mechanism for this strand-specific nicking unknown, a role for such a mechanism in MMR seems paradoxical given that pre-existing nicks are often present in newly replicated DNA, in particular during lagging strand replication and these can potentially serve as sites for the initiation of excision reactions. It is possible that such a mechanism simply increases the probability of nicks being present 5’ to mispairs to act as substrates for Exo1 or other excision processes, or it is possible that this mechanism reflects other excision mechanisms possibly mediated by high-density nicking at or near the mispair [81, 114]. Sequence homology relationships suggested that Mlh1-Mlh3 would have an endonuclease activity (see the minireview by Putnam in this issue) and this has been demonstrated, although the properties of the Mlh1-Mlh3 endonuclease thus far reported are strikingly different from those of the Mlh1-Pms1 endonuclease [115,
The Mlh1-Mlh2 complex is not predicted to have an endonuclease activity (see the minireview by Putnam in this issue) and its possible role in MMR has remained unclear. Surprisingly, Mlh1-Mlh2 shares many properties with Mlh1-Pms1 including mispair-dependent recruitment by Msh2-Msh6 and Msh2-Msh3 and mispair- and Msh2-Msh6-dependent formation of foci in living cells, suggesting that Mlh1-Msh2 may play some type of accessory role in MMR [62]. Much remains to be learned about Mlh1-Mlh2 and Mlh1-Mlh3 in comparison to what we know about Mlh1-Pms1.

4. Reconstitution of MMR in vitro

The demonstration that extracts of S. cerevisiae, Xenopus, Drosophila, human and mouse cells could catalyze repair of mispair-containing DNAs [25-28, 69] set the stage for biochemical fractionation and reconstitution studies demonstrating that eukaryotic MMR proteins can catalyze a nick-directed MMR reaction similar to the reaction catalyzed by E. coli MMR proteins. The most rapid progress was made in the human system, largely due to work performed in the Modrich and Li labs. However, I am pleased that my lab was ultimately able to reconstitute MMR reactions using purified S. cerevisiae MMR proteins, making it possible to apply the wealth of insights from genetics studies of MMR in S. cerevisiae to the biochemistry of reconstituted MMR reactions. To date, reconstitution studies have defined two types of mispair-directed excision/repair reactions. In one reaction, a combination of Msh2-Msh6 or Msh2-Msh3, Exo1, DNA Pol δ, the
single-stranded DNA binding protein RPA, PCNA and RFC promotes the repair of a circular mispaired substrate containing a nick on the 5' side of the mispair. In this reaction, the mispair recognition factors and other proteins promote excision by Exo1 past the mispair, followed by repair DNA synthesis to fill in the resulting gap repairing the initial mispair [85, 86, 117]. In a second reaction, a combination of Msh2-Msh6 or Msh2-Msh3, Mlh1-Pms1 (hMlh1-Pms2), Exo1, DNA Pol δ, RPA, PCNA and RFC promotes the repair of a circular mispaired substrate containing a nick on the 3' side of the mispair. In this reaction, the Mlh1-Pms1 endonuclease is activated to generate nicks 5' to the mispair, which then allows Exo1-dependent 5' excision and subsequent gap filling to occur essentially as in the 5'-nick-directed in vitro MMR reactions [86, 110]. In both of these reactions, mispair recognition by Msh2-Msh6 and Msh2-Msh3 appears to stimulate excision by Exo1, in part by overcoming inhibition of Exo1 by RPA [118]. Whether the ability of Exo1 to interact with Msh2, and potentially with the Msh2-Msh6 and Msh2-Msh3 sliding clamps, plays a role in this stimulation of Exo1 is unclear. Mlh1-Pms1 does not stimulate excision in reconstituted 5'-nick-directed MMR reactions even though hMlh1-Pms2 appears to be required for maximal 5'-nick-directed excision in human extract-based MMR reactions under conditions where the hMlh1-Pms2 endonuclease should not be required [119].

A critical question about the reconstituted MMR reactions discussed above is how well they correspond to MMR in vivo. It seems likely that these reactions recapitulate nick-directed Exo1-mediated mispair excision and coupled gap-filling
by DNA Pol δ that functions in MMR, particularly on the lagging strand, which is expected to have a high density of nicks due to discontinuous DNA synthesis. It is not clear whether other DNA polymerases besides DNA Pol δ can participate in the gap-filling reaction. In contrast, genetic studies in S. cerevisiae and mice have shown that Exo1 defects only cause small defects in MMR and that redundant excision mechanisms exist [65, 69, 120]. Strikingly, all known mutations that specifically inactivate Exo1-independent MMR but not Exo1-dependent MMR that have been characterized to date are highly but not completely defective for activation of the Mlh1-Pms1 endonuclease [68, 81, 110]. These mutations all cause defects in reconstituted Mlh1-Pms1-dependent MMR reactions suggesting that these reactions could represent coupling of the Exo1-independent MMR Mlh1-Pms1 activation pathway to an Exo1-dependent excision system [110]. These results also suggest that high levels of Mlh1-Pms1 endonuclease activation are required for Exo1-independent MMR, raising the possibility that high levels of Mlh1-Pms1 endonuclease activation could represent an endonuclease-promoted mispair excision pathway [81, 114]. Other possible Exo1-independent mispair excision mechanisms have also been proposed [120, 121]. Finally, MMR is coupled to DNA replication in vivo [77, 122], but little is known about how the MMR mechanisms defined in vitro are coupled to DNA replication.

5. Genetics of mammalian MMR defects
My initial interest in a possible relationship between MMR defects and human disease came from the similarity between the phenotypes caused by defects in *S. cerevisiae MSH1* and some of the human mitochondrial myopathies; this idea did not prove to be relevant as mammals do not appear to contain a gene encoding an Msh1 protein. Regardless, I set out to try to clone human and mouse *MSH2* in collaboration with Rick Fishel and later to clone human *MLH1* and *PMS1* (known in the literature as *PMS2*) by extending the collaboration to include Mike Liskay. I knew that defects in MMR genes caused microsatellite instability due to the published work on *E. coli* MMR and because I had given my *msh2* mutations to Tom Petes for his studies on microsatellite instability in *S. cerevisiae* [123, 124]. We had already cloned the human *MSH2* gene when reports appeared on microsatellite instability in human tumors and in tumors associated with Hereditary Non-polyposis Colorectal Cancer (HNPCC, now called Lynch Syndrome) [125, 126], causing considerable excitement. We rapidly mapped the *MSH2* gene to the Chromosome 2 HNPCC locus, setting off our search for evidence that *MSH2* was mutated in HNPCC [127, 128]. The chromosome 3 HNPCC locus was reported a few months later [129] leading to our subsequent studies demonstrating that *MLH1* was the chromosome 3 HNPCC gene [130, 131]. We focused on these two genes because they were the only two genes that could be linked to HNPCC by mapping studies. Similar studies were performed de la Chapelle and Vogelstein [132, 133]. It was exciting for me to learn how to clone and determine the structure of mammalian genes and to set up molecular diagnostics assays. I was particularly grateful to my colleagues at
the Dana-Farber Cancer Institute who worked very hard to help identify collaborators who could provide samples from HNPCC families for analysis. I was also grateful to be able to work with these clinical collaborators, whose efforts were essential in ultimately showing that *MSH2* and *MLH1* were the major HNPCC genes and in bringing these genes into the area of molecular diagnostics of cancer predisposition. A particularly striking moment occurred when a mutation carrier identified in our initial studies who did not have cancer was re-examined by their physician and found to have a previously undetected cancer, thereby potentially saving her life; this impressed on me the true importance of finding ways to use basic science discoveries to impact human health.

Many laboratories rapidly started working on the relationship between MMR defects and cancer susceptibility, both through studies of human patients and the establishment of mouse models (see the minireviews by Heinen, by Begum and Martin, by Peña-Diaz and Rasmussen, by Sijmons and Hofstra and by Lee, Tosti and Edelmann in this issue). My lab focused on further developing the genetics and biochemistry of *S. cerevisiae* MMR, which has provided an excellent foundation for studies of the genetics of MMR in human and mouse, the development of molecular diagnostics for defects in *MSH2* and *MLH1* as these were clearly the major Lynch Syndrome genes (they were also the only cancers we had good access to), and providing the first evidence for silencing of *MLH1* in sporadic MMR-defective cancers and for *MLH1* epimutations [128, 131, 134, 135]. Studies from other labs established that *hPMS2* and *MSH6* were also
defective in cancer, albeit in cancers/syndromes that seem somewhat different from classical Lynch syndrome, whereas defects in hPMS1 (scMLH2), MLH3 and EXO1 were not convincingly demonstrated or were very rare, consistent with the weak mutator phenotypes caused by defects in these genes in model organisms (For representative papers see [136-140]). Studies of human cancer cell lines containing MMR defects initially characterized in the Karran and Kunkel labs as well as by others also helped establish the nature of human MMR defects and the relationships between MMR and DNA damage responses [141-145]. The first mouse models of MMR defects were produced by the Liskay (mPMS2) and Mak (MSH2) labs and de Wind working with te Riele (MSH2); these were rapidly followed by studies of other mouse MMR genes by Edelmann and Kucherlapati (MLH1, MSH6, MSH3, EXO1), Liskay (MLH1), de Wind and te Riele (MSH6, MSH3) and Lipkin (MLH3) [45-47, 60, 69, 146-151] (see the minireview by Lee, Tosti and Edelmann in this issue). The extensive efforts in human and mouse genetics, most recently incorporating data generated by cancer genome sequencing projects such as The Cancer Genome Atlas Project have provided a broad picture of the contribution of human MMR defects to cancer susceptibility; it is gratifying to me that studies in S. cerevisiae have provided such an accurate road map for the study of the mammalian MMR genes attesting to the importance of studies in model organisms. In the long run, I hope that understanding MMR defects in cancer will lead to treatments for MMR-defective cancers and in this regard, recent reports that MMR defects may predict responses to immunotherapy are cause for optimism [152, 153].
6. Unanswered or incompletely answered big-picture questions about MMR mechanisms

While we have learned a great deal about eukaryotic MMR, there is still much that we do not know. The unanswered questions about MMR appear to fall in three general areas. First, what are the detailed reaction mechanisms of individual MMR proteins and key groups of MMR proteins that functionally interact? Second, how do we reconstitute complete MMR reactions that reflect coupling to DNA replication, processing of chromatin substrates and how are strand-discrimination signals produced and recognized? Third, how do MMR proteins function in other processes beyond post-replication MMR such as preventing recombination between divergent DNA sequences and the DNA damage response?

The MMR proteins are a fascinating group of proteins and individual activities have been linked to each protein and studied in almost all cases. A striking property of virtually all known MMR proteins is that their function in MMR is driven by different conformational changes and different protein-protein interactions. Yet how these interactions then modulate the activities of each MMR protein seems less clear. Some examples include: 1) do interactions of Exo1 with Msh2-Msh6 (or Msh2-Msh3) and/or Mlh1-Pms1 play a role in recruiting and activating Exo1; 2) how does Mlh1-Pms1 interact with its activating proteins
so that its nicking activity is directed to the already nicked strand; 3) does Mlh1-Pms1 function as a stable complex with Msh2-6 (or Msh2-Msh3) or is it loaded onto the DNA and then acts independently of the mispair recognition proteins; 4) does Mlh1-Mlh2 stimulate the activity of Mlh1-Pms1 and if so, how; and 5) how does the Mlh1-Mlh3 endonuclease function in MMR given that its activity seems so different from that of Mlh1-Pms1. These types of questions as well as whether post-translational modifications modulate the activity of MMR proteins and whether there are MMR proteins yet to be identified will likely be addressed over the next few years using a combination of genetics and biophysical methods.

Tremendous progress has been made in identifying eukaryotic MMR proteins, elucidating their activities and reconstituting MMR reactions in vitro, although many questions about MMR mechanisms remain to be answered. These reconstituted MMR reactions in many ways mimic the mechanisms of E. coli MMR reactions, despite the greater number of MutS-related and MutL-related heterodimeric complexes involved in eukaryotic MMR. However, eukaryotic MMR does not use post-replication DNA methylation asymmetries as a strand specificity signal, and the eukaryotic strand specificity signal has not been definitively established. An attractive hypothesis that has existed for many years due to the fact that MMR proteins catalyze nick-directed MMR is that nicks in the newly replicated strand due to discontinuous DNA synthesis serve as a strand specificity signal [27, 154]. In addition, it has been suggested that PCNA loaded at these nicks or remaining at these nicks after DNA synthesis [155] plays a role
in strand specificity [74, 156], which is especially attractive given the activation of the nick-directed Mlh1-Pms1 (hMLH1-PMS2) endonuclease by PCNA loaded at nicks [80]. This hypothesis is most applicable to the lagging strand where nicks are expected to be prevalent; however, there is currently little direct evidence for this mechanism of strand recognition and it raises the question of what is the exact role of the Mlh1-Pms1 (hMLH1-PMS2) endonuclease, given that nicks are already present in the DNA. This hypothesis is likely less applicable to the leading strand that is synthesized more continuously and where PCNA is less important in DNA synthesis compared to its role in Pol δ-dependent lagging strand DNA synthesis [157]. Nicking of misincorporated ribonucleotides by RNase H2 has been postulated to provide a strand specificity signal that could act on the leading strand [158, 159]; however, the mutator phenotype caused by RNase H2 defects seem too weak for this to be a major strand discrimination mechanism [160]. MMR is also coupled to DNA replication, which could aid in the recognition of rarely occurring mispairs in vivo [77, 122][(for a review see [161]). This coupling occurs by at least two different mechanisms, one of which utilizes the interaction between PCNA and the N-terminus of either Msh6 or Msh3 whereas the other mechanism(s) has not been identified [74, 75, 77]. There is some evidence that an interaction between the deuterostome-specific Msh6 PWWP domain (not present in Msh3) and a histone modification can help target Msh2-Msh6 to chromatin during S-phase, although PWWP domain mutations have not been tested to verify the significance of this interaction [162]. Finally, MMR must act on chromatin and interact with chromatin dis-assembly
and assembly pathways as well as other pathways that process replication intermediates [163-166]. It is possible that studies focused on using purified proteins to reconstitute MMR reactions that are coupled to DNA replication and chromosome dis-assembly and assembly in vitro could provide new insights into MMR mechanisms and how MMR is targeted to the newly synthesized DNA strands.

Much of the investigation of MMR has focused on the enzymatic mechanisms of post-replication MMR and mutagenesis mechanisms involving MMR. Yet MMR proteins function in other processes. One example is the role of MMR proteins along with other proteins such as the Sgs1 DNA helicase in preventing recombination between divergent DNA sequences, a reaction that was first described by Radman and Huang [167-170] (see the minireview by Tlam, Kanaar and Lebbnik in this issue). This process seems likely to play an important role in preventing non-allelic homologous recombination, which might otherwise lead to genome rearrangements [171]. Little is yet known about the enzymatic mechanisms of this process, although the recent demonstration that Msh2-Msh6 can recognize mispairs in the context of Rad51-mediated D-loops represents important progress [172]. Similarly, the mechanism by which MMR proteins act on mispairs in the intermediates formed during normal homologous recombination increasing gene conversion and decreasing post-meiotic segregation has not been extensively studied. MMR proteins also act in the processes that promote the expansion of trinucleotide repeats [173, 174], an
important human disease process; however, the role of MMR proteins in this process has not been fully elucidated (see the minireview by Schmidt and Pearson in this issue). Finally, MMR proteins play a role in DNA damage signaling, in promoting cell death in response to DNA damaging agents such as chemotherapeutic agents and in suppressing mutations induced by DNA damage as well as promoting mutagenesis (For representative reviews see [144, 175-177])(see the minireviews by Li, Pearlman and Hsieh, by Crouse, and by Zanotti and Gearhart in this issue). Considerable data have been accumulated on the interaction between MMR proteins and signaling proteins, and many cell biology studies have documented defects in cell death mechanisms in MMR-defective mutants. However, the MMR-dependent cell death mechanisms such as direct signaling to initiate cell death or the induction of futile cycles of DNA repair have not been fully elucidated. These important MMR-related pathways all warrant further study.

**Acknowledgements**

I would like to thank Drs. Eva Goellner, Chris Putnam and Anjana Srivatsan for helpful discussions, help with figures and comments on the manuscript, Dr. Rodney Rothstein for advice on Figure 1, and Drs. Rick Fishel and Chris Pearson for helpful discussions. Work on MMR in my lab was supported by the Ludwig Institute for Cancer Research and NIH grant GM50006.
Figure Legends

Figure 1. Double strand break repair models illustrating potential roles of mismatch repair in recombination and potential roles of double strand break repair in mismatch repair. **A.** DSBR model that explains how gene conversion can occur without the need to invoke MMR. One (the a allele) of the two alleles (a allele-containing chromosome in blue and A allele-containing chromosome in red) whose segregation during DSBR is illustrated is located on the recipient chromosome at the location where the DSB occurs and is then expanded into a gap with 3’ single strand overhangs thus eliminating the a allele. Strand invasion to form a D-loop followed by priming of repair synthesis (dashed lines) and strand displacement, capture of the other end of the DSB and additional repair DNA synthesis result in the formation of a double Holliday junction intermediate. Resolution of the two Holliday junctions in the non-crossover directions yields two A allele-containing chromosomes and hence a gene conversion event. Other resolution directions are possible, but these do not alter the illustrated gene conversion event. **B.** An alternative of the DSBR model illustrated in A where one (the a allele) of the two alleles whose segregation during DSBR is illustrated is located on the recipient chromosome in the region that is converted to a 3’ single strand overhang. This then results in the formation of a heteroduplex region containing an A:a mispair on one of product chromosomes whereas repair synthesis recreates the starting allele on both strands of the other produce chromosome. If DNA replication then occurs
without MMR (Left) Post-Meiotic Segregation results whereas if MMR occurs prior to DNA replication as illustrated (Right), then gene conversion occurs. 

C. The DSBR model can account for post-replication repair of mispairs if a mispair (illustrated as a:A) generated as the result of a replication error is simply cleaved by a mismatch-specific endonuclease to generate a DSB. Then DSBR will result in copying the wild-type allele into the broken chromosome effectively producing the same outcome as classical excision repair mechanisms of MMR.

Figure 2. Illustration summarizing the known eukaryotic proteins that function in mismatch repair. An MMR pathway involving the Msh2-Msh6 mispair recognition complex, the Mlh1-Pms1 endonuclease, Exo1, PCNA, the RFC clamp loader, RPA and DNA Pol δ is illustrated on the left side of the figure. Interactions between Msh2 and Mlh1, between Exo1 and both Msh2 and Mlh1, and between PCNA and the N-terminus of Msh6 are indicated. The Mlh1-Mlh2 complex, which is not an endonuclease, is indicated because it appears to have a role in MMR when the level of Mlh1-Pms1 is experimentally reduced. Substitution of the Msh2-Msh3 mispair recognition complex for Msh2-Msh6 is illustrated in the center and on the right, and substitution of the Mlh1-Mlh3 complex for the Mlh1-Pms1 complex in MMR reactions involving Msh2-Msh3 is illustrated to the right (note that it has not been resolved whether Mlh1-Mlh3 can also function in conjunction with Msh2-Msh6). The thickness of the arrows connecting each pathway to the three classes of mispairs illustrated reflects the relative importance of each pathway in repair of the different classes of mispairs.
The color-coding indicates whether the different genes are essential or are redundant/partially redundant with each other and hence whether loss-of-function mutations in the different genes result in complete or partial loss of MMR or inviability. All gene/protein names are S. cerevisiae gene/protein names.

References


[23] D.K. Bishop, J. Andersen, R.D. Kolodner, Specificity of mismatch repair following transformation of Saccharomyces cerevisiae with heteroduplex plasmid DNA,


[34] A. Jeyaprakash, J.W. Welch, S. Fogel, Mutagenesis of yeast MW104-1B strain has identified the uncharacterized PMS6 DNA mismatch repair gene locus and
additional alleles of existing PMS1, PMS2 and MSH2 genes, Mutation research, 325 (1994) 21-29.


[51] A. Jeyaprakash, R. Das Gupta, R. Kolodner, Saccharomyces cerevisiae pms2 mutations are alleles of MLH1, and pms2-2 corresponds to a hereditary nonpolyposis colorectal carcinoma-causing missense mutation, Molecular and cellular biology, 16 (1996) 3008-3011.


[82] Y. Xie, C. Counter, E. Alani, Characterization of the repeat-tract instability and mutator phenotypes conferred by a Tn3 insertion in RFC1, the large subunit of the yeast clamp loader, Genetics, 151 (1999) 499-509.


Complete loss of MMR  
Loss of base:base MMR  
Redundant/weak effects  
Essential for viability

Base:Base Mismatch
Small Insertion:Deletion Mismatch
Large Insertion:Deletion Mismatch

Mlh1
Pms1
Exo1
PCNA
POLδ
RFC
POLδ
RPA

Mlh2
Mlh1
Msh2
Msh6

Mlh3
Mlh1
Msh2
Msh3

Mlh2
Mlh1
Msh2
Msh3

Mlh1
Mlh2
Msh2
Msh3

Mlh1
Mlh2
Msh2
Msh3