J. Genet., Vol. 75, Number 2, August 1996, pp. 181–192. © Indian Academy of Sciences

REVIEW ARTICLE

Mismatch repair genes of eukaryotes

R. GEETA VANI and M. R. S. RAO
Department of Biochemistry, Indian Institute of Science, Bangalore 560 012, India

MS received 29 February 1996; revised received 6 June 1996

Abstract. Mismatches that arise during replication or genetic recombination or owing to damage to DNA by chemical agents are recognized by mismatch repair systems. The pathway has been characterized in detail in *Escherichia coli*. Several homologues of the genes encoding the proteins of this pathway have been identified in the yeast *Saccharomyces cerevisiae* and in human cells. Mutations in the human genes *hMSH2*, *hMLH1*, *hPMS1* and *hPMS2* have been linked to hereditary nonpolyposis colon cancer (HNPCC) and to some sporadic tumours. Mismatch repair also plays an antirecombinogenic role and is implicated in speciation.

Keywords. Mismatch repair; eukaryotes; cancer.

1. Introduction

Faithful transmission of genetic information is of paramount importance for the survival of a cell. DNA polymerase, the key enzyme involved in DNA replication, despite its proofreading activity introduces one wrong nucleotide out of every million nucleotides in *in vitro* replication assays (Loeb and Kunkel 1982). It becomes immediately apparent that such an error rate could theoretically result in up to 10 mutations during each generation of a bacterium like *Escherichia coli* and approximately 3000 new mutations in a human cell following DNA synthesis during the S phase. However, the error rate observed *in vivo* is at least 1000-fold lower than that observed *in vitro* (Cheng and Loeb 1993). A number of mechanisms have evolved to ensure high-fidelity transmission of genetic material from one generation to the next since mutations can lead to genotypes that may be deleterious to the cell. In addition to their generation during DNA replication, mismatched nucleotides can also arise owing to physical damage of the DNA (Friedberg 1985). Such mispaired nucleotides in DNA are corrected by the mismatch repair system. To date, the best-defined mismatch repair pathway is the *E. coli mutHLS* pathway (Modrich 1991), and to a lesser extent the *hex* pathway of *Streptococcus pneumoniae* (Claverys and Lacks 1986). The *mutHLS* pathway, apart from ensuring fidelity of chromosome replication, also ensures fidelity of homologous genetic recombination by its antirecombinogenic role which prevents recombination between quasihomologous sequences (Modrich 1991; Radman and Wagner 1993; Selva *et al.* 1995). There have been scattered biochemical studies on mismatch repair in extracts of human (Holmes *et al.* 1990; Thomas *et al.* 1991), *Drosophila* (Holmes *et al.* 1990) and *Xenopus* (Varlet *et al.* 1990) cells. However, the observation that some hereditary colon cancer patients showed instability of microsatellite [$(CA)_n$ repeats] sequences in their DNA (Aaltonen *et al.* 1993) led to identification of the mismatch repair genes of humans and demonstration of a close link between this pathway and hereditary colon cancer. In this article, we discuss recent developments in this area.

2. Mismatch repair system in *E. coli*

As mentioned earlier, the best-defined mismatch repair pathway is the *E. coli mutHLS* pathway, which is dependent on the *mutL*, *mutS*, *mutH* and *mutU* (*uvrD*) gene products (Modrich 1991). The strand specificity that is required for mismatch repair is dictated by the status of adenine methylation at d(GATC) sites, which is a post-replication event. The GATC sequences in newly synthesized DNA are unmethylated and hence serve as a recognition signal for the ensuing mismatch repair. This methyl-directed excision reaction has been reconstituted *in vitro* (Lahue *et al.* 1989). The sequence of events that occur at the region of mismatch is shown in figure 1. Mismatch repair is initiated by recognition of mismatched bases by the MutS protein, followed by

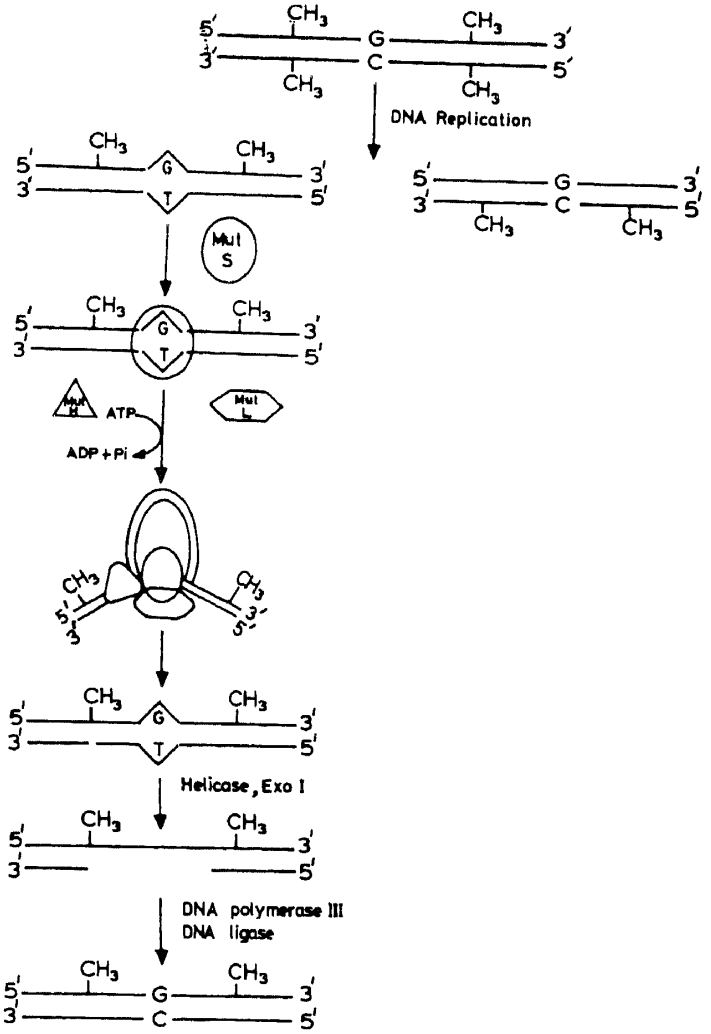

Figure 1. Schematic illustration of the methyl-directed mismatch repair pathway of *E. coli*. Following one round of DNA replication the newly synthesized daughter strands are unmethylated at adenine residues in GATC sites, which provides strand discrimination for repair. The mismatches generated by DNA polymerase during replication are recognized by the MutS protein. MutH, a latent endonuclease, recognizes a GATC site, and micking by MutH at unmethylated GATC is activated by recruitment of MutL protein onto the MutS–DNA complex. The action of DNA helicase unwinds DNA from the nick, and excision of nucleotides is brought about by exonuclease I (Exo I) if the nick is 3′ to the mismatch (as shown in the figure), or by exonuclease VII (Exo VII) or Rec J exonuclease if the nick is 5′ to the mismatch. The gap in the DNA is then filled by DNA polymerase III and sealed by the action of DNA ligase.

recruitment of MutL in an ATP-dependent manner. Assembly of the MutS–MutL complex on the DNA leads to activation of a latent endonuclease activity of MutH which is dependent on hydrolysis of ATP by MutS. MutH cleaves DNA at unmethylated GATC. The incision by MutH is bidirectional in the sense that breakage of the phosphodiester bond in the unmethylated strand occurs at the GATC site present either 5′ or 3′ to the mismatch. The efficient repair reaction requires participation of DNA helicase II (*uvrD* gene product), DNA polymerase III, DNA ligase, the single-stranded-DNA binding protein (SSB) and any of the ssDNA exonucleases (ExoI, ExoVII or RecJ). Exonuclease I (with a 3′→5′ nucleolytic activity) participates if the incision occurs 3′ to the mismatch, while ExoVII or RecJ (with 5′→3′ nucleolytic activity) is involved when the incision occurs 5′ to the mismatch. The gap thus generated is repaired by DNA polymerase III in the presence of SSB and DNA ligase.

In contrast to the mismatch repair pathway in Gram-negative bacteria like *E. coli*, the mismatch repair system of Gram-positive bacteria such as *S. pneumoniae* lacks a MutH homologue and does not use *dam* methylation for strand discrimination. MutS and MutL homologues have been identified in *S. pneumoniae*: HexA and HexB respectively (Claverys and Lacks 1986).

The *mutHLS* pathway of *E. coli* corrects mismatches incorporated during replication. *E. coli* has evolved two other pathways to correct mismatches that arise owing to the action of chemical agents. The very short pathway (VSP) corrects G–T mismatches that are created by deamination of 5-methylcytosine residues to G–C pairs (Jones *et al.* 1987; Freidberg *et al.* 1995). This reaction involves five protein components. The Vsr protein specifically nicks the T-containing strand 5′ to the T in the sequence context of CT(A/T)GG and NT(A/T)GG, and this activity is stimulated by MutS and MutL (Hennecke *et al.* 1991). Excision and resynthesis is carried out by DNA polymerase I and DNA ligase. In addition to repairing G–T mismatches, the VSP system also corrects G–U mispairs although less efficiently.

The second pathway, which is dependent on the *mutY* gene product, corrects mismatches resulting from oxidative damage to guanine residues (Modrich 1991; Freidberg *et al.* 1995). At the nucleotide level, when dGTP is oxidized to 8-oxo-dGTP, the MutT protein degrades the 8-oxo-dGTP to prevent its incorporation into DNA. If, however, 8-oxo-dGTP is incorporated into DNA, it is removed by a reaction catalysed by the *mutM* gene product, Fpg glycosylase (Michaels *et al.* 1992). If the DNA containing 8-oxo-G survives all these protective mechanisms and goes through a round of replication, it results in G–C to T–A change in the second round of replication. To prevent such a change the A–8-oxo-G mispair is recognized by the MutY-dependent repair pathway. The MutY acts both as a glycosylase and as an apurinic/apyrimidinic lyase, creating a nick (Tsai-Wu *et al.* 1992). Excision and resynthesis are carried out by DNA polymerase I and DNA ligase. Interestingly, the MutY-dependent repair reaction also corrects A–G and A–C mispairs to C–G and G–C base pairs respectively, albeit to a lesser extent.

3. Mismatch repair system in yeast

The mismatch repair system was originally invoked to explain gene conversion events during genetic recombination in yeast. Mutations in *Saccharomyces cerevisiae* that resulted in a high frequency of post-meiotic segregation (*pms*) were traced to mutations inactivating a mismatch-repair-associated gene (Williamson *et al.* 1985; Bishop *et al.*

1987; Kramer *et al.* 1989). Five *PMS* genes, namely *PMS1*, *PMS2*, *PMS3*, *PMS5* and *PMS6*, have been identified that affect mismatch repair. Mutations in all these genes increase the frequency of recombination between closely linked markers and decrease the number of gene conversion events relative to post-meiotic segregation events (Williamson *et al.* 1985; Jeyaprakash *et al.* 1994). Subsequently the *PMS1* gene was shown to be homologous to the bacterial *mutL* gene (Kramer *et al.* 1989).

The *E. coli mutS* homologues in *S. cerevisiae* were first cloned by Kolodner and coworkers (Reenan and Kolodner 1992a). Using the degenerate-primer-directed PCR approach, they identified and cloned two *mutS* homologues, namely *MSH1* and *MSH2*. Both appeared to be nucleus-encoded gene products. Transposon mutagenesis was used to decipher their probable functions. Mutations in *MSH2* were shown to affect nuclear DNA mismatch repair. Haploid *msh2* mutants showed an 85-fold increase in the rate of spontaneous mutation over that of the wild type. Homozygous *msh2* diploids, upon sporulation, did not survive. Gene conversion events were also affected in *msh2* mutants; there was a decrease in the ratio of gene conversion events leading to *PMS*. *msh1* mutants showed the petite phenotype, indicating that *MSH1* is probably involved in mitochondrial DNA metabolism. Analysis of mitochondrial DNA from such petite mutants revealed large-scale rearrangements. The *MSH1* gene encodes a protein with a mitochondrial targeting sequence and the protein is localized in mitochondria (Reenan and Kolodner 1992b). Mitochondrial DNA in *msh1/MSH1* strains accumulates point mutations more rapidly than in wild type (Nai-Wen Chi and Kolodner 1994). Purified MSH1 protein hydrolyses ATP and recognizes mismatches in a manner similar to MutS. They suggest that MSH1 plays a role in eliminating errors and homeologous recombination in mitochondrial genome.

4. Mismatch repair gene homologues and human colon cancer

While these developments were occurring in identification and cloning of mismatch repair genes in lower eukaryotes, researchers in the area of the genetics of human colon cancer were encountering a new and challenging puzzle. The DNA isolated from colorectal tumour samples and hereditary nonpolyposis colon cancer (HNPCC) samples showed a phenomenon called microsatellite instability (Aaltonen *et al.* 1993; Thibodeau *et al.* 1993). The human genome contains repeats of mononucleotide, dinucleotide, trinucleotide and tetranucleotide sequences dispersed throughout the genome. When the PCR products of certain tumour samples were analysed, it was observed that alleles showed either expansion or deletion of (CA) repeats in the tumour. Such expansion or deletion of repeat sequences was suspected to arise from a defect in a factor responsible for maintenance of replication fidelity. Loss of function of this factor would cause replication slippage, resulting in such expansion or deletion of (CA) repeats. The association of RER^+ (replication error) phenotype of microsatellites with HNPCC and the observation that *S. cerevisiae* defective in mismatch repair genes manifest instability of dinucleotide repeats (Strand *et al.* 1993) suggested a link between HNPCC and mismatch repair genes in humans. The microsatellite locus that was cosegregating with colon cancer was localized to chromsome 2 (Peltomaki *et al.* 1993).

Kolodner's group had earlier cloned the yeast homologue of *E. coli mutS*. The idea that some factor in the post-replication DNA repair process might be involved in the maintenance of repeat sequences in DNA immediately prompted Kolodner's group to

clone the human homologue of yeast *MSH2*. They again used PCR technology and were successful in obaining the cDNA clone of *hMSH2* (Fishel *et al.* 1993). Expression of *hMSH2* in *E. coli* resulted in an increased rate of accumulation of mutations in the *E. coli* genome, thus establishing that expression of *hMSH2* results in a strong mutator phenotype. A similar observation had been made earlier with expression of *S. pneumoniae hexA* in *E. coli*. The expression of a heterologous gene encoding a mismatch-recognizing gene product in *E. coli* resulted in impairment in the mismatch repair pathway since the heterologous proteins do not interact with the proteins of the native pathway such as MutL and MutH (Prudhomme *et al.* 1991). Under these circumstances mismatches remain uncorrected and hence a strong mutator phenotype results. The *hMSH2* gene maps to human chromosome 2p22–21 near a locus implicated in HNPCC (Fishel *et al.* 1993). Final confirmation that this gene is responsible for colon cancer came from the demonstration that two of the tumour DNAs were found to have a T-to-C transition mutation at position -6 of the splice acceptor site of the intron, located at nucleotide position 2020 of the cDNA sequence.

In a parallel study Vogelstein and coworkers, who had earlier demonstrated close association of microsatellite instability with hereditary colon cancer, succeeded in identifying the *hMSH2* gene as the candidate gene for HNPCC through positional cloning (Leach *et al.* 1993). They also showed that there were germline mutations in three HNPCC kindreds at codon positions 622 (proline → leucine), 265–314 (frame deletion) and 406 (arginine → stop). They also observed mutations in RER^+ cell lines. In an accompanying paper, Vogelstein, Modrich and others demonstrated that mutation rate in $(CA)_n$ repeats in RER^+ tumour cells is at least 100-fold higher than in RER^- cells (Parsons *et al.* 1993). The RER^+ cells were also shown to be defective in strand-specific mismatch repair.

The open reading frame (ORF) of *hMSH2* cDNA contains 2802 nucleotides and encodes a protein of 934 amino acids. The region between amino acids 598 and 789 of the human protein is 85% identical to the corresponding region of the *S. cerevisiae* MSH2 protein. The evolutionary relationship among all the known MutS-related proteins indicates that hMSH2 is a member of a group of MutS homologues. Recently *mutS* gene homologues have been cloned from mouse and *Xenopus* (Varlet *et al.* 1994) and rat (Geeta Vani and Rao 1996).

The *hMSH2* genomic locus has been cloned and shown to encompass a region of 73 kb with 16 exons (Kolodner *et al.* 1994). Availability of the nucleotide sequence of all the intron–exon junctions has enabled design of primer sets in the intron regions to study mutations both within the exon and in the intron–exon junctions.

HNPCC is classically divided into two syndromes. Only colorectal cancer is seen in the Lynch I subtype, while in the Lynch II subtype a variety of additional epithelial-derived tumours are seen. In a third type of syndrome called Muir–Torre syndrome, the disease is characterized by tumours of sebaceous glands and skin in addition to colorectal cancer. Analysis of DNA samples of two large HNPCC kindreds exhibiting features of the Muir–Torre syndrome has revealed inheritance of a frameshift mutation in the *MSH2* gene in one family and a nonsense mutation in exon 12 of *MSH2* in the other family (Kolodner *et al.* 1994).

Liu *et al.* (1994) also analysed 29 HNPCC kindreds for mutations in *hMSH2*. About 40% of them had germline mutations, which were point mutations, truncation, and in-frame and out-of-frame deletions. It is quite likely that the rest of the kindreds carry mutations in other genes of the mismatch repair pathway.

Linkage analysis also showed that in addition to *hMSH2*, localized to chromosome 2, being the candidate gene for HNPCC, a second locus on chromosome 3 p21–23 is also associated with HNPCC (Lindblom *et al.* 1993). It was suspected that one of the other genes of the mismatch repair pathway might be localized to chromosome 3. This prediction proved to be correct. Bronner *et al.* (1994) and Papadopoulos *et al.* (1994) cloned this gene. It has shown to be a *mutL* homologue and was therefore named *hMLH1*. The *hMLH1* cDNA clone contains an ORF of 2268 bp that encodes a protein of 756 amino acids, which has 41% identity with the yeast MLH1. This gene was localized to chromosome 3p 21·3–23 by FISH analysis. Missense mutations in *hMLH1* were detected in DNA samples of HNPCC-affected individuals. Mutations in *hMSH2* account for 60–65% and mutations in *hMLH1* for 30% of all HNPCC patients.

The *hMLH1* genomic locus on chromosome 3p has been cloned and shown to encompass 58 kb of DNA with 19 exons. Sequence information for all the intron–exon junctions has facilitated a PCR approach for analysing each of the *hMLH1* exons for mutations (Kolodner *et al.* 1995).

Two additional *mutL* homologues, namely *hPMS1* and *hPMS2*, have been isolated (Nicolaides *et al.* 1994). *hPMS1* cDNA contains an ORF of 2796 nucleotides encoding a 932-residue protein, which has 27% identity with yeast PMS1. The *hPMS2* cDNA has an ORF of 2586 bp encoding a protein of 862 amino acids, which has 32% identity with yeast PMS1. *hPMS1* and *hPMS2* are localized to chromosomes 2 (2q 31–33) and 7 (7p 22) respectively. Both *hPMS1* and *hPMS2* genes were found to be mutated in the germline of some of the HNPCC patients. These mutations account for 2–5% of all HNPCC cases.

In most HNPCC families, there is a germline mutation of one or more of the mismatch repair genes *hMSH2*, *hMLH1*, *hPMS1* and *hPMS2*. The identification of the second allele in a tumour supports the idea that mutation in both the alleles of a mismatch repair gene probably triggers the onset of cancer.

The instability of microsatellite sequences has thus become a diagnostic tool for detection of possible mutations in the mismatch repair pathway genes. The human data are amply supported by the experiments of Strand *et al.* (1993), who showed that mutations in any one of the three yeast genes *PMS1*, *MLH1* and *MSH2* lead to 100–700-fold increases in poly(GT)-tract instability. In a control experiment they showed that mutations that eliminate the proofreading activity of DNA polymerase had no effect on stability of poly(GT) tracts.

A more direct evidence for the role of *hMSH2* gene in genomic instability and tumorigenesis has come from the recent knockout experiments carried out by two groups of workers. de Wind *et al.* (1995) have generated embryonic stem (ES) cells and mice that are deficient in *mMSH2* gene. The mutant ES cells showed a strong mutator phenotype, hyperrecombination and methylation tolerance. *MSH2*-deficient mice developed normally. However, many of the mutant mice developed lymphomas. In another, independent study, Reitmar *et al.* (1995) also developed homozygous $MSH2^{-/-}$ mice. These mice developed normally through two generations. Some of these mice developed lymphoid tumours which exhibited microsatellite instability. Both these reports provide a direct link between *MSH2* defect and pathogenesis of cancer.

Microsatellite instability, which is a result of mutations in the genes of the mismatch repair pathway, does not explain the pathophysiology of cancer phenotypes. It is quite likely that mutations occur in some of the regulatory genes of the cell cycle, triggering oncogenesis. It is interesting to point out here that the type II TGF-β receptor is

inactivated in colon cancer cells (Markowitz *et al.* 1995); the inactivation was caused by truncation of the 10 adenine repeats by one or two nucleotides within the coding region. TGF-β inhibits growth of multiple epithelial cell types, and hence a mutation in TGF-β receptor relieves this control mechanism and probably initiates tumorigenesis.

Knockout mice carrying disrupted *PMS2* gene were susceptible to developing sarcomas and lymphomas like *MSH2*-deficient mice (Baker *et al.* 1995). Microsatellite instability was observed in testis, tail and the tumours that had developed. A striking phenotype of these mice was that the male mice were sterile and produced abnormal sperms. Cytological observation of the spermatocytes indicated that chromosomal synapsis was severely affected in these *PMS2*-deficient mice.

5. Mismatch recognition by hMSH2

The human *MSH2* gene has been expressed in *E. coli* using the pET expression system with a histidine tag (Fishel *et al.* 1994a). The purified protein binds to single mismatches as well as multiple insertion/deletion loops. hMSH2 also binds to insertion/deletion loops containing $(CA)_n$ repeats (Fishel *et al.* 1994b). The *E. coli* MutS protein on the other hand does not recognize multiple insertion/deletion loops. It remains to be seen how these differences in recognition specificity are manifested at the level of protein–nucleic acid interaction. Even the mechanism of recognition of a single-base mismatch by the MutS family of proteins is a challenging problem particularly since there do not seem to be major alterations in the helical properties of DNA containing a mismatch from that of normal duplex DNA (Patel *et al.* 1984; Hare *et al.* 1986).

Native hMSH2 has been purified from human cells as a heterodimer, termed hMutSα, comprising hMSH2 and a 160-kDa protein also called G:T-binding protein [GTBP] (Drummond *et al.* 1995; Palombo *et al.* 1995). When these proteins were synthesized *in vitro* as individual polypeptides, neither of them showed any mismatch-binding activity. However, when synthesized together, the heterodimer formed showed mismatch-binding activity. Some preliminary studies have been carried out on the relative contribution of each of the two polypeptides in mismatch recognition. The present thinking is that a single-base mismatch and slipped intermediates in single-base runs are recognized by the hMSH2–p160/GTBP complex. On the other hand, slipped intermediates of two or more bases may be recognized by hMSH2 alone or as a heterodimer with a yet unidentified protein. There are no germline mutations in the gene that encodes GTBP in HNPCC cases, and the genomic instability seen in the GTBP mutants is less compared to defects in other genes (Papadopoulos *et al.* 1995). The alterations are predominantly in mononucleotide tracts. The human homologue of MutL has also been purified as a heterodimer, hMutLα, from human cells, and comprises *hMLH1* and *hPMS2* gene products (Li and Modrich 1995). In the yeast *S. cerevisiae* also, MLH1 acts in conjunction with PMS1 in facilitating repair of MSH2-bound mismatches. The involvement of hPMS1 in mismatch repair is not clear at present although mutation in the *hPMS1* gene has been observed in a single patient with family history of colorectal cancer.

6. Other *mutS* homologues

Four additional *mutS* homologues have been identified in *S. cerevisiae*. Crouse and coworkers (New *et al.* 1993) have isolated the *MSH3* gene, which shows high sequence

similarity with bacterial *mutS*. The homology of *MSH3* is more with mouse *Rep3* (Linton *et al.* 1989) or human *DUG* than with yeast *MSH1* or *MSH2*. Mouse *Rep3* and the human *DUG* genes (Fuji and Shimada 1989) are present next to the *DHFR* gene in the respective genomes, and transcription is in opposite orientation from the same *DHFR* promoter. The physiological role of the *MSH3* gene product is not clear at present since disruption of this gene does not severely alter either the spontaneous mutation rates or post-meiotic segregation. Nothing much is known about the functions of *Rep3* and *DUG* genes also.

Another *mutS* homologue, *MSH4*, has been isolated from *S. cerevisiae* by Ross-Macdonald and Roeder (1994). This gene appears to be meiosis-specific and is expressed during meiotic prophase prior to meiosis I chromosome segregation. *MSH4* has been mapped to chromosome VI and encodes a protein of 878 amino acids. The amino acid sequence is 35% identical with *E. coli* MutS and 31–36% identical with the other MutS homologues in yeast. The similarity is maximum in the C-terminal portion encompassing the ATP-binding domain and the DNA-binding motif. The MSH4 protein has amino acid substitutions, compared to other MutS homologues, in both the ATP-binding and the DNA-binding domains. Interestingly, the hydrophobic residues normally present in the DNA-binding domain of other MutS homologues are replaced by hydrophilic amino acids in MSH4. *msh4* null mutation has no effect on DNA mismatch repair. These mutants display normal levels of gene conversion and post-meiotic segregation. However, they show a reduction in crossing-over which results in an increase in nondisjunction of paired homologous chromosomes at meiosis I. These results have been interpreted as suggesting that MSH4 protein interacts with recombination intermediates, facilitating their resolution.

MSH5 is another gene isolated from *S. cerevisiae* belonging to the *mutS* family (Hollingsworth *et al.* 1995). It is different from *MSH4* in sequence but is not involved in DNA mismatch repair, like *MSH4*. The gene has an ORF of 2703 nucleotides and encodes a protein of 901 amino acids. Of all the *mutS* homologues of *S. cerevisiae*, *MSH5* is most homologous to *MSH2* and least to *MSH4*. The phenotype of *msh5* mutants is similar to that of *msh4* mutants in that they show (i) no effect on mismatch repair, (ii) decreased spore viability, (iii) increased levels of chromosome nondisjunction, and (iv) decreased reciprocal exchange between homologous chromosomes. Gene conversion was, however, not affected in *msh5* mutants. *msh4* and *msh5* diploid double mutants revealed that these two genes may belong to the same epistatic group. It has been suggested that MSH4 and MSH5 may interact to form heterodimers similar to human MutSα. Recently Kolodner and coworkers (Kolodner 1995) have identified yet another MutS homologue in *S. cerevisiae*, MSH6. MSH6 is thought to dimerize with MSH2 protein. MSH6 protein may have a function similar to that of human p160/GTBP.

7. Microsatellite instability and sporadic cancer

Although initial studies were directed towards demonstrating a correlation between microsatellite instability and mutations in the mismatch repair genes of HNPCC families, more recently microsatellite instability has been observed in many of the sporadic tumours of colorectal cancer, and those of lung (Merlo *et al.* 1984), breast (Glebov *et al.* 1994) and pancreas (Han *et al.* 1993). It is quite likely that these sporadic tumours have acquired somatic mutations similar to those inherited in HNPCC. There

have been conflicting reports on the occurrence of mutations in the four mismatch repair genes *MSH2*, *MLH1*, *PMS1* and *PMS2* (Liu *et al.* 1995). We have to wait for a much larger database of the analysis of mutations in these genes before we can ask questions regarding origin of these mutations in sporadic tumours.

It is now becoming very clear that tumorigenesis is a result of multigene defects. There could be two possibilities for the involvement of mismatch repair genes in the onset of cancer. First, because of the strong mutator phenotype of mutation in these genes, a loss of function of one of these genes might result in accumulation of mutations in oncogenes and tumour suppressor genes. This hypothesis is strongly favoured by Vogelstein and coworkers (Vogelstein and Kinzler 1994). Second, the loss of mismatch repair functions might cause mutations in some of the key genes involved in control of the cell cycle, as exemplified by the study of type II TGF receptor (Markowitz *et al.* 1995).

8. Mismatch repair system and speciation

The mismatch repair system, in addition to having a key role in maintaining the integrity of DNA sequences in the genome, also plays an important role in providing a barrier to intermixing of genomes between two species. An excellent example of this phenomenon is seen in *E. coli* and *Salmonella typhimurium*. Although drug-resistance extrachromosomal elements can easily move between these two organisms, chromosomal transfer and recombination is very rare. In a classic paper, Radman and coworkers (Rayssiguier *et al.* 1989) showed that this barrier could be removed by inactivating the mismatch repair genes. The barrier to homeologous recombination is likely due to inhibition of branch migration (Worth *et al.* 1994). There is evidence for the presence of a barrier by mismatch repair proteins to homeologous recombination in *S. cerevisiae* (Selva *et al.* 1995). These experiments suggest that mismatch repair enzymes may function as potent inhibitors of interspecies recombination and slow down the rate of evolution (Matic *et al.* 1995).

9. Conclusions and perspectives

In the last five years there have been many advances in identification of genes responsible for mismatch repair in eukaryotic cells. The biochemistry of mismatch repair has been elucidated and reconstituted *in vitro* with purified components synthesized in *E. coli* (Lahue *et al.* 1989). Even the lower eukaryote *S. cerevisiae* has evolved additional mismatch recognition proteins to tackle meiotic recombination events. The meiosis-specific MutS homologues have not been identified in animal systems yet but it is quite likely that they do exist. Although knockout experiments are giving some clues to the function of the mismatch repair genes, we have to wait for detailed biochemical analysis of the mismatch repair pathway in animal cells for a better understanding of the relative contribution of each of the homologues.

The most significant finding in this area over the last few years has been the demonstration of a close link between mismatch repair genes and hereditary colon cancer. Since we do find microsatellite instability in many of the sporadic cancers, a large database on the nature of mutations in the mismatch repair genes might help us in answering some of the questions relating to genetic predisposition of human populations to cancer.

Acknowledgement

Research on the rat homologue of *MSH2* in the authors' laboratory is financially supported by grants from the Department of Science and Technology.

References

Aaltonen A. L., Peltomaki P., Leach R. S., Sistonen P., Pylkkanen L., Mecklin J. P., Jarvinen H., Powell S M., Jen J., Hamilton S. R. *et al.* 1993 Clues to the pathogenesis of familial colorectal cancer. *Science* 260: 812–816

Baker M. S., Bronner E. C., Zhang L., Plug W. A., Robatzek M., Warren G., Elliott A. E., Yu J., Ashley T., Arnheim N., Flavell A. R. and Liskay M. R. 1995 Male mice defective in the DNA mismatch repair gene *PMS2* exhibit abnormal chromosome synapsis in meiosis. *Cell* 82: 309–319

Bishop D. K., Williamson M. S., Fogel S. and Kolodner R. D. 1987 The role of heteroduplex correction in gene conversion in *Saccharomyces cerevisiae*. *Nature* 328: 362–364

Bronner C. E., Baker S. M., Morrison P. T., Warren G., Smith L. G., Lescoe M. K., Kane M., Earabino C., Lipford J., Lindblom A. *et al.* 1994 Mutation in the DNA mismatch repair gene homologue *hMLH1* is associated with hereditary nonpolyposis colon cancer. *Nature* 368: 258–261

Cheng K. C. and Loeb L. A. 1993 Genomic instability and tumor progression: mechanistic considerations. *Adv. Cancer Res.* 60: 121–156

Claverys J. P. and Lacks S. A. 1986 Heteroduplex deoxyribonucleic acid base mismatch repair in bacteria. *Microbiol. Rev.* 50: 133–165

de Wind Niels, Dekker M., Berns A., Radman M. and Riele H. 1995 Inactivation of the mouse *MSH2* gene results in mismatch repair deficiency, methylation tolerance, hyperrecombination and predisposition to cancer. *Cell* 82: 321–330

Drummond T. J., Minli G., Longley J. M. and Modrich P. 1995 Isolation of a hMSH2-p160 heterodimer that restores DNA mismatch repair to tumor cells. *Science* 268: 1909–1912

Fishel R., Lescoe M. K., Rao M. R. S., Copeland M., Jenkins N., Garber J., Kan N. and Kolodner R. D 1993 The human mutator gene homolog *MSH2* and its association with hereditary nonpolyposis colon cancer. *Cell* 75: 1027–1038

Fishel R., Ewel A. and Lescoe M. K. 1994a Purified human MSH2 protein binds to DNA containing mismatched nucleotides. *Cancer Res.* 54: 5539

Fishel R., Ewel A., Lee S., Lescoe K. M. and Griffin J. 1994b Binding of mismatched microsatellite DNA sequences by the human MSH2 protein. *Science* 266: 1403–1405

Freidberg E. C. 1985 *DNA repair* (New York: W. H. Freeman)

Freidberg E. C., Walker G. C. and Siede W. 1995 *DNA repair and mutagenesis* (Washington, DC: ASM Press) pp. 396–399

Fuji H. and Shimada T. 1989 Isolation and characterization of cDNA clones from the divergently transcribed gene in the region upstream from the human dihydrofolate reductase gene. *J. Biol. Chem.* 264: 10057–10064

Geeta Vani R. and Rao M. R. S. 1996 Cloning of the cDNA encoding rat homologue of the mismatch repair gene *MSH2* and its expression during spermatogenesis. (in press)

Glebov O. K., McKenzie K. E., White C. A. and Sukumar S. 1994 Frequent p53 gene mutations and novel alleles in familial breast cancer. *Cancer Res.* 54: 3703–3709

Han H. J., Yanagisawa A., Kato Y., Park J. G. and Nakamura Y. 1993 Genetic instability in pancreatic cancer and poorly differentiated type of gastric cancer. *Cancer Res.* 53: 5087–5089

Hare D., Shapiro L. and Patel D. J. 1986 Wobble dG-dT pairing in right handed DNA: Solution conformation of d(C-G-T-G-A-A-T-T-C-G-C-G) deduced from distance geometry analysis of NOESY spectra. *Biochemistry* 25: 7445–7456

Hennecke F. H., Kolmar K., Brundl K. and Fritz H. J. 1991 The *vsr* gene product of *E. coli* K-12 is a strand and sequence-specific DNA mismatch endonuclease. *Nature* 353: 776–778

Hollingsworth M. N., Ponte L. and Halsey C. 1995 MSH5, a novel MutS homolog, facilitates meiotic reciprocal recombination between homologs in *Saccharomyces cerevisiae* but not mismatch repair. *Genes Dev.* 9: 1728–1739

Holmes J. J., Clark S. and Modrich P. 1990 Strand-specific mismatch correction in nuclear extracts of human and *Drosophila melanogaster* cell lines. *Proc. Natl. Acad. Sci. USA* 87: 5837–5841

Jeyaprakash A., Welch J. W. and Fogel S. 1994 Mutagenesis of yeast MW 104-1B strain has identified the uncharacterized *PMS6* DNA mismatch repair gene locus and additional alleles of existing *PMS1*, *PMS2* and *MSH2* genes. *Mutat. Res.* 325: 21–29

Jones M., Wagner R. and Radman M. 1987 Mismatch repair and recombination in *E. coli*. *Cell* 50: 621–626

Kolodner R. D. 1995 Mismatch repair: Mechanisms and relationship to cancer susceptibility. *Trends Biochem. Sci.* 20: 397–401

Kolodner R. D., Hall N. R., Lipford J., Kane M. F., Rao M. R. S., Morrison P., Wirth L., Finan P. J., Burn J., Chapman P. *et al.* 1994 Structure of the human *MSH2* locus and analysis of two Muir–Torre kindreds for *msh2* mutations. *Genomics* 24: 516–526

Kolodner R. D., Hall N. R., Lipford J., Kane M. F., Morrison P., Finan P. J., Burn J., Chapman P., Earabino C., Merchant E. *et al.* 1995 Structure of the human *MLH1* locus and analysis of a large HNPCC kindred for *MLH1* mutations. *Cancer Res.* 55: 242–248

Kramer W., Kramer B., Williamson M. S. and Fogel S. 1989 Cloning and nucleotide sequence of DNA mismatch repair gene *PMS1* from *Saccharomyces cerevisiae*: homology of PMS1 to procaryotic MutL and HexB. *J. Bacteriol.* 171: 5339–5346

Lahue R. S., Au K. G. and Modrich P. 1989 DNA mismatch correction in a defined system. *Science* 245: 160–164

Leach F. S., Nicolaides N. C., Papadopoulos N., Liu B., Jen J., Parsons R., Peltomaki P., Sistonen P., Aaltonen L. A., Nystrom-Lahti M. *et al.* 1993 Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. *Cell* 75: 1215–1225

Li G. and Modrich P. 1995 Restoration of mismatch repair to nuclear extracts of H6 colorectal tumor cells by a heterodimer of human MutL homologs. *Proc. Natl. Acad. Sci. USA* 92: 1950–1954

Lindblom A., Tannergard P., Werelies B. and Nordenskjold M. 1993 Genetic mapping of a second locus predisposing to hereditary nonpolyposis colon cancer. *Nat. Genet.* 5: 279–282

Linton J. P., Yen J. Y. J., Selby E., Chen Z., Chinsky J. M., Liu K., Kellens R. E. and Crouse G. F. 1989 Dual bidirectional promoters at the mouse *DHFR* locus: cloning and characterization of two mRNA classes of the divergently transcribed *Rep-1* gene. *Mol. Cell. Biol.* 9: 3058–3072

Liu B., Parsons R. E., Hamilton S. R., Petersen G. M., Lynch H. T., Watson P., Markowitz S., Wilson J. K., Green J., Chapelle A. *et al.* 1994 *hMSH2* mutations in hereditary nonpolyposis colorectal cancer kindreds. *Cancer Res.* 54: 4590–4594

Liu B., Nicolaides N. C., Markowitz S., Willson J. K. V., Parsons R. E., Jen J., Papadopoulos N., Peltomaki P., Chapelle A., Hamilton S. R., Kinzler K. W. and Vogelstein B. 1995 Mismatch repair gene defects in sporadic colorectal cancers with microsatellite instability. *Nat. Genet.* 9: 48–55

Loeb L. A. and Kunkel T. A. 1982 Fidelity of DNA synthesis. *Annu. Rev. Biochem.* 51: 429–457

Markowitz S., Wang J., Myreoff L., Parsons R., Sun L., Lutterbaugh J. *et al.* 1995 Inactivation of the type II TGF-β receptor in colon cancer cells with microsatellite instability. *Science* 268: 1336–1338

Matic I., Rayssiguier C. and Radman M. 1995 Interspecies gene exchange in bacteria: The role of SOS and mismatch repair systems in evolution of species. *Cell* 80: 507–515

Merlo A., Mabry M., Gabrielson E., Vollmer R., Baylin S. B. and Sidransky D. 1994 Frequent microsatellite instability in primary small cell lung cancer. *Cancer Res.* 54: 2098–2101

Michaels M. L., Tchou J., Grollman A. P. and Miller J. H. 1992 A repair system for 8-oxo-7,8-dihydrodeoxyguanine. *Biochemistry* 31: 10964–10968

Modrich P. 1991 Mechanisms and biological effects of mismatch repair. *Annu. Rev. Genet.* 25: 229–253

Nai-Wen Chi and Kolodner R. D. 1994 Purification and characterization of MSH1, a yeast mitochondrial protein that binds to DNA mismatches. *J. Biol. Chem.* 269: 29984–29992

New L., Liu K. and Crouse F. G. 1993 The yeast gene *MSH3* defines a new class of eukaryotic *mutS* homologues. *Mol. Gen. Genet.* 239: 97–108

Nicolaides N. C., Papadopoulos N., Liu B., Wei Y. F., Carter K. C., Ruben S. M., Rosen C. A., Haseltine W. A., Fleischmann R. D., Fraser C. M. *et al.* 1994 Mutations of two *PMS* homologues in hereditary nonpolyposis colon cancer. *Nature* 371: 75–80

Palombo F., Gallinari P., Laccarino I., Leltieri T., Hughes M., D'Arrigo A., Truong O., Hsuan J. J. and Jiricny J. 1995 GTBP, a 160 kilodalton protein essential for mismatch-binding activity in human cells. *Science* 268: 1912–1914

Papadopoulos N., Nicolaides N. C., Wei Y.-F., Ruben S. M., Carter K. C., Rosen C. A., Haseltine W. A., Fleischman R. D., Fraser C. M., Adams M. D. *et al.* 1994 Mutation of a *mutL* homolog in hereditary colon cancer. *Science* 263: 1625–1629

Papadopoulos N., Nicolaides N. C., Liu B., Parsons R., Lengauer C., Palombo F., D'Arrigo A., Markowitz S., Wilson J. K. V., Kinzler K. W. *et al.* 1995 Mutations of GTBP in genetically unstable cells. *Science* 268: 1915–1917

Parsons R., Li G. M., Longley J. M., Fang W., Papadopoulos N., Jen J., Chapelle A., Kinzler N. K., Vogelstein B. and Modrich P. 1993 Hypermutability and mismatch repair deficiency in RER$^+$ tumor cells. *Cell* 75: 1227–1236

Patel D. J., Kozlowski S. A., Ikuta S. and Itakurat 1984 Dynamics of DNA duplexes containing G-T, G-A, A-C and T-C pairs: hydrogen exchange at and adjacent to mismatch sites. *Fed. Proc.* 43: 2663–2670

Peltomaki P., Aaltonen A. L., Sistonen P., Pylkkanen L., Mecklin P. J., Jarvinen H., Green S. J., Jass R. J., Weber C. J., Leach F. S., Petersen M. G., Hamilton R. S., Chapelle A. and Vogelstein B. 1993 Genetic mapping of a locus predisposing to human colorectal cancer. *Science* 260: 810–812

Prudhomme M., Mejean V., Martin B. and Claverys J. P. 1991 Mismatch repair genes of *Streptococcus pneumoniae*: HexA confers a mutator phenotype in *Escherichia coli* by negative complementation. *J. Bacteriol.* 173: 7196–7203

Radman M. and Wagner R. 1993 Mismatch recognition in chromosomal interactions and speciation. *Chromosoma* 102: 369–373

Rayssiguier C., Thaler D. S. and Radman M. 1989 The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch repair mutants. *Nature* 342: 396–401

Reenan R. A. G. and Kolodner R. D. 1992a Isolation and characterization of two *Saccharomyces cerevisiae* genes encoding homologs of the bacterial HexA and MutS mismatch repair proteins. *Genetics* 132: 963–973

Reenan R. A. G. and Kolodner R. D. 1992b Characterization of insertion mutations in the *Saccharomyces cerevisiae MSH1* and *MSH2* genes: evidence for separate mitochondrial and nuclear functions. *Genetics* 132: 975–985

Reitmar A. H., Schmits R., Ewel A., Bapat B., Redston M., Mitri A., Waterhouse P., Mittrucker H. W., Wakeham A., Liu B., Thomason A., Griesser H., Gallinger S., Balljausen W. G., Fishel R. and Mak T. W. 1995 MSH2 deficient mice are viable and susceptible to lymphoid tumours. *Nat. Genet.* 11: 64–70

Ross-Macdonald P. and Roeder G. S. 1994 Mutation of a meiosis specific MutS homolog decreases crossing over but not mismatch correction. *Cell* 79: 1069–1080

Selva E. M., New L. and Lahue R. S. 1995 Mismatch correction acts as a barrier to homeologous recombination in *Saccharomyces cerevisiae*. *Genetics* 139: 1175–1181

Strand M., Prolla T. A., Liskay R. M. and Peter T. D. 1993 Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* 365: 274–276

Thibodeau S. N., Bren G., Schaid 1993 Microsatellite instability in cancer of the proximal colon. *Science* 260: 816–819

Thomas D. C., Roberts J. D. and Kunkel T. A. 1991 Heteroduplex repair in extracts of human HeLa cells. *J. Biol. Chem.* 266: 3744–3751

Tsai-Wu J. J., Liu H. L. and Lu A. L. 1992 *E. coli* MutY protein has both N-glycosylase and apurinic-apyrimidinic endonuclease activities on A.C and A.G mispairs. *Proc. Natl. Acad. Sci. USA* 89: 8779–8783

Varlet J., Radman M. and Brooks P. 1990 DNA mismatch repair in *Xenopus* egg extracts: Repair efficiency and DNA repair and synthesis for single base-pair mismatches. *Proc. Natl. Acad. Sci. USA* 87: 7883–7887

Varlet J., Pallard C., Radman M., Moreau J. and Niels de Wind 1994 Cloning and expression of *Xenopus* and mouse MSH2 DNA mismatch repair genes. *Nucl. Acids Res.* 22: 5723–5728

Vogelstein B. and Kinzler K. W. 1994 Colorectal cancer and the intersection between basic and clinical research. *Cold Spring Harbor Symp. Quant. Biol.* 59: 517–521

Willamson M. S., Game J. C. and Fogel S. 1985 Meiotic gene conversion mutants in *Saccharomyces cerevisiae*: isolation and characterization of pms1-1 and pms1-2. *Genetics* 110: 609–646

Worth L., Clark S., Radman M. and Modrich P. 1994 Mismatch repair proteins MutS and MutL inhibit RecA catalyzed strand transfer between diverged DNAs. *Proc. Natl. Acad. Sci. USA* 91: 3238–3241