

- ethylenediaminetetraacetate and conjugation to cysteine side chains. *Bioconjugate Chem.* 8, 44–48
- 23 Rana, T.M. and Meares, C.F. (1991) Transfer of oxygen from an artificial protease to peptide carbon during proteolysis. *Proc. Natl. Acad. Sci. U. S. A.* 88, 10578–10582
- 24 Platis, I.E. *et al.* (1993) Oxidative polypeptide cleavage mediated by EDTA–Fe covalently linked to cysteine residues. *Biochemistry* 32, 12761–12767
- 25 Chen, C.H. and Sigman, D.S. (1987) Chemical conversion of a DNA-binding protein into a site-specific nuclease. *Science* 237, 1197–1201
- 26 Aki, T. and Adhya, S. (1997) Repressor induced site-specific binding of HU for transcriptional regulation. *EMBO J.* 16, 3666–3674
- 27 Young, R. (1991) RNA polymerase II. *Annu. Rev. Biochem.* 60, 689–715
- 28 Chamberlin, M. (1976) RNA polymerase – an overview. In *RNA Polymerase* (Losick, R. and Chamberlin, M., eds), pp. 17–67. Cold Spring Harbor Laboratory Press
- 29 Ishihama, A. (1997) Promoter selectivity control of RNA polymerase. In *Nucleic Acids and Molecular Biology* (Vol. 11) (Eckstein, F. and Lilley, D.M.J., eds), pp. 53–70, Springer-Verlag
- 30 Gill, S.C. *et al.* (1991) *Escherichia coli*  $\sigma 70$  and NusA proteins, I. Binding interactions with core RNA polymerase in solution and within the transcription complex. *J. Mol. Biol.* 220, 307–324
- 31 Hillel, Z. and Wu, C.W. (1977) Subunit topography of RNA polymerase from *Escherichia coli*. A cross-linking study with bifunctional reagents. *Biochemistry* 16, 3334–3342
- 32 Arthur, T.M. and Burgess, R.R. (1998) Localization of a  $\sigma 70$  binding site on the N terminus of the *Escherichia coli* RNA polymerase  $\beta'$  subunit. *J. Biol. Chem.* 273, 31381–31387
- 33 Malhotra, A. *et al.* (1996) Crystal structure of a  $\sigma 70$  subunit fragment from *E. coli* RNA polymerase. *Cell* 87, 127–136
- 34 Zhang, G. *et al.* (1999) Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* 98, 811–824
- 35 Polyakov, A. *et al.* (1995) Three-dimensional structure of *E. coli* core RNA polymerase: promoter binding and elongation conformations of the enzyme. *Cell* 83, 365–373
- 36 McMahan, S.A. and Burgess, R.R. (1999) Mapping protease susceptibility sites on the *Escherichia coli* transcription factor  $\sigma 70$ . *Biochemistry* 38, 12424–12431
- 37 Naryshkin, N. *et al.* (2000) Structural organization of the RNA polymerase–promoter open complex. *Cell* 101, 601–611
- 38 Ebright, R.H. (1998) RNA polymerase–DNA interaction: structures of intermediate, open, and elongation complexes. *Cold Spring Harbor Symp. Quant. Biol.* 63, 11–20
- 39 Helmann, J.D. and deHaseth, P.L. (1999) Protein–nucleic acid interactions during open complex formation investigated by systematic alteration of the protein and DNA binding partners. *Biochemistry* 38, 5959–5967
- 40 Allison, L.A. *et al.* (1985) Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. *Cell* 42, 599–610
- 41 Helmann, J.D. (1994) Bacterial sigma factors. In *Transcription: Mechanisms and Regulation* (Conaway, R.C. and Conaway, J.W., eds), pp. 1–17, Raven Press, New York, NY, USA
- 42 Wigneshweraraj, S.R. *et al.* (2000) Conservation of sigma–core RNA polymerase proximity relationships between the enhancer-independent and enhancer-dependent sigma classes. *EMBO J.* 19, 3038–3048
- 43 Culver, G.M. *et al.* (1999) Identification of an RNA–protein bridge spanning the ribosomal subunit interface. *Science* 285, 2133–2136
- 44 Newcomb, L.F. and Noller, H.F. (1999) Directed hydroxyl radical probing of 16S rRNA in the ribosome: spatial proximity of RNA elements of the 3' and 5' domains. *RNA* 5, 849–855

## Dual recognition–incision enzymes might be involved in mismatch repair and meiosis

Harmit S. Malik and Steven Henikoff

Mismatch repair in many organisms depends on three proteins: the mismatch-recognition protein MutS, a nicking endonuclease MutH, and MutL, which acts as a scaffold between these. However, many genomes lack *MutL* but possess *MutS*. In one of these cases, in a coral mitochondrial genome, a gene is present that encodes a MutS protein fused to an HNH nicking endonuclease, potentially eliminating the requirement for MutL. Likewise, many prokaryotes could operate similarly, independently of MutL by encoding a fused MutS–Smr (*MutS2*) protein. Smr, which is proposed to be a nicking endonuclease, can also be found separately in many eukaryotes, where it might play a role in mismatch repair or meiotic chromosome crossing-over.

**DNA REPLICATION CAN** be an error-prone process. Errors introduced during replication can be corrected by the mismatch repair (MMR) system found in most organisms. Mutations in the MMR system have been associated with many hereditary diseases, including nonpolyposis colon cancer<sup>1</sup>. The most extensively studied MMR system is the

*Escherichia coli* MutS–MutL–MutH pathway (Fig. 1a, Ref. 2). A mismatch is recognized by the MutS dimer, which then recruits MutL and MutH dimers, with MutL acting as a scaffold between MutS and the endonuclease MutH. MutH nicks the DNA at a hemi-methylated GATC site, where the Dam methyltransferase has introduced a methylated adenine that provides the signal to distinguish the parental strand from the newly synthesized and unmethylated daughter strand. Because MutS can recognize a variety of base mismatches, as well as loops, MutLSH systems can correct a

variety of replication errors, including frameshifts caused by strand slippage in microsatellites.

Eukaryotic MutS proteins have been best characterized in the yeast *Saccharomyces cerevisiae*. This yeast has six MutS homologs (or MSH). MSH1 is found only in fungi and is involved in mitochondrial MMR (Ref. 3). MSH2–MSH3 and MSH2–MSH6 heterodimers are responsible for repairing short (1–2 bases) and larger loops of mismatches, respectively<sup>4,5</sup>. MSH4 and MSH5 do not appear to play a role in MMR at all. Instead, a heterodimer of MSH4–MSH5 is involved in meiotic crossing-over and chromosome segregation<sup>6–8</sup>, where it is believed to recognize and resolve recombination intermediates that involve segments of mismatched bases brought together from the two parental chromosomes during meiosis. There are four MutL homologs (MLH) in *S. cerevisiae*. MLH1 heterodimerizes with either Pms1p (also an MLH), MLH2 or MLH3. MLH1–Pms1p helps correct single-base mismatches and small insertions and deletions, whereas MLH1–MLH3 heterodimers are involved in meiotic recombination, along with MSH4 and MSH5 (Refs 9,10). The role of MLH2 is still unclear. No genes for putative MutH homologs have been identified in the *S. cerevisiae* genome so far (Fig. 1b).

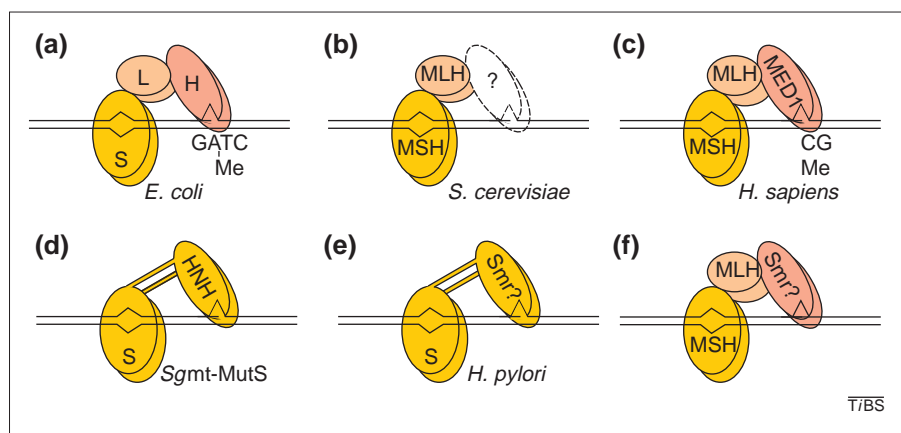
By virtue of their sequence similarity, it is relatively simple to identify MutS homologs from a variety of species. Most genomes that contain *MutS* also possess genes for MutL homologs. However, very few of these genomes also contain genes encoding proteins

H.S. Malik and S. Henikoff are at the Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA.  
Email: steveh@fhcrc.org

homologous to MutH (Ref. 11). From sequence comparisons, it appears likely that Dam methyltransferase and MutH represent remnants of an ancient restriction–modification (R–M) system that was recruited by a eubacterial genome for DNA repair<sup>12</sup>; only a few bacteria possess both Dam and MutH. Therefore, most organisms can use a different nicking endonuclease. This was demonstrated recently by the discovery of MED1, a nicking endonuclease with homology to the glycosylase–lyase family. Although MED1 (also known as MBD4) is not homologous to MutH, it appears to carry out an analogous role to MutH in humans<sup>13</sup>. Mutations in MED1 are found in human carcinomas with microsatellite instability<sup>14</sup>. MED1 bears a methyl–CpG-binding domain, and it has been suggested that cytosine methylation is at least one of the signals used to distinguish parental from daughter strands in the human genome (Ref. 13, Fig. 1c). However, DNA modification is not ubiquitous. Many genomes lacking any DNA modification (e.g. *Drosophila melanogaster* and *Caenorhabditis elegans*) also contain genes for MutS and MutL homologs, and are probably subject to MMR. In the absence of a DNA modification signal, how do MMR systems distinguish between parental and daughter strands? The identification of alternative MutH analogs should help delineate how this strand recognition occurs in other systems.

### The coral mitochondrial MutS

A gene that encodes a MutS homolog has been identified in the mitochondrial genome of the coral *Sarcophyton glaucum* (*Sgmt-MutS*, Ref. 15). A subsequent report suggested that *Sgmt-MutS* might have been acquired via a ‘reverse’ horizontal transfer from the nuclear genome to the mitochondria<sup>16</sup>. However, a more comprehensive analysis<sup>11</sup> concluded that *Sgmt-MutS* was part of a relatively uncharacterized group of prokaryotic MutS homologs, termed MutS2, and was unlikely to be of nuclear origin. We identified a putative HNH endonuclease domain in the *Sgmt-MutS* protein using PSI-BLAST searches<sup>19</sup>, with embedded sequences<sup>20</sup> as a query (Fig. 2). HNH endonuclease domains have been previously identified in proteins conferring mobility to Group I and Group II introns, and inteins, as well as in bacteriocins<sup>17,18</sup>. The *Sgmt-MutS* HNH was detected at significant levels (*E* values of  $10^{-4}$  to  $10^{-5}$ ) using the Group-II intron consensus. This level of significance is



**Figure 1**

Various configurations of mismatch repair (MMR) systems. **(a)** *Escherichia coli* MutLSH system. MutS (S) proteins recognize mismatches and recruit MutL (L), which acts as a scaffold between MutS and MutH (H). MutH binds and cleaves 5' of an unmethylated GATC site paired to a methylated (Me) site. **(b)** In *Saccharomyces cerevisiae*, six MutS and four MutL homologs (MSH and MLH, respectively) are known that function in MMR and chromosome crossing-over and segregation. No MutH-like genes have been identified, and the signal to distinguish between the parental and the daughter strands is also unknown. **(c)** MED1 is an analog of MutH in *Homo sapiens* that has been proposed to recognize CG hemimethylated sites. **(d)** The *Sgmt-MutS* (a MutS homolog present in the mitochondrial proteome of the coral *Sarcophyton glaucum*) is proposed to be a dual recognition–incision enzyme, which uses an HNH domain for MutL-independent nicking. *Sgmt-MutS* might recognize a specific DNA sequence rather than a particular DNA modification. **(e)** MutS2 proteins (in *Helicobacter pylori*, for example) encode a fusion between MutS proteins and Smr proteins; it is proposed, by analogy to *Sgmt-MutS*, that Smr is also a nicking endonuclease. **(f)** In some genomes, interactions between MutS and Smr might be mediated by MutL proteins.

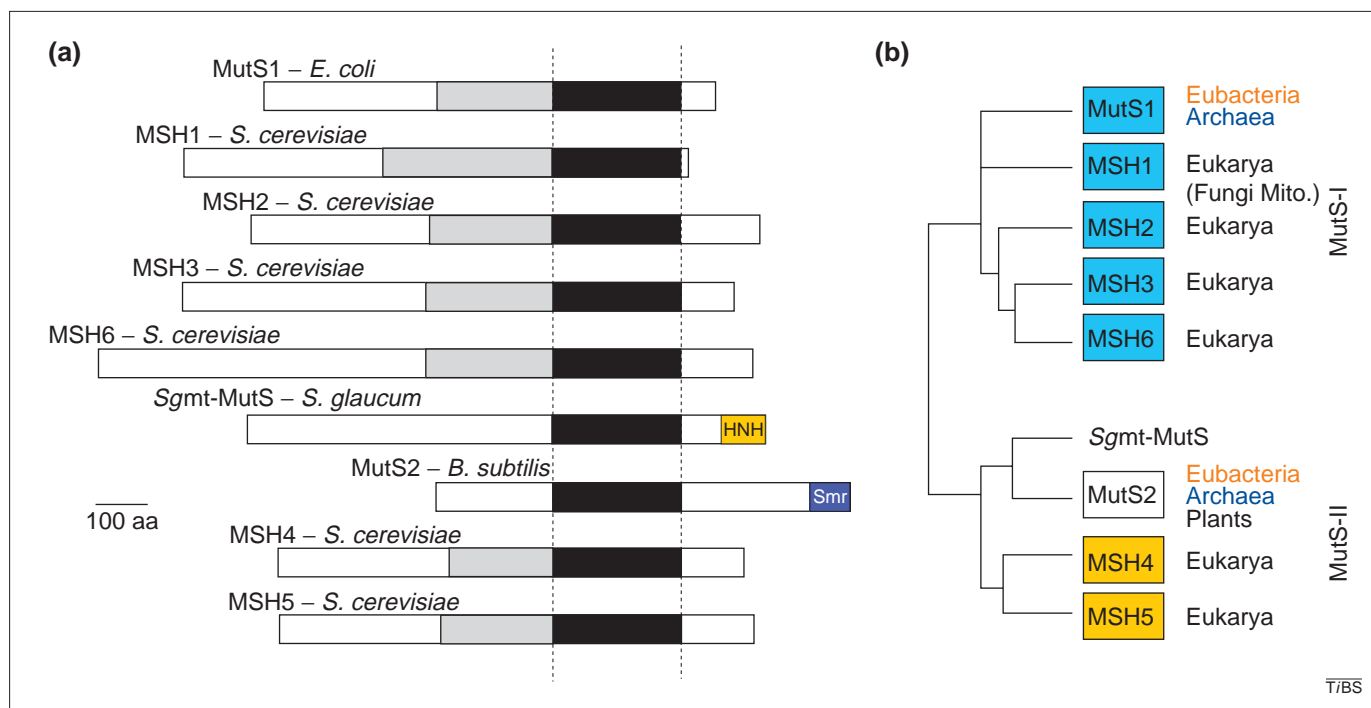
	C	Q	I	CG	..(4)..	L	E	V	H	H	V	R	..(21)..	N	R	K	Q	I	T	C	K	T	C	H	F	K	V	H			
<i>Saccharomyces cerevisiae</i>	C	Q	I	CG	..(4)..	L	E	V	H	H	V	R	..(21)..	N	R	K	Q	I	T	C	K	T	C	H	F	K	V	H			
<i>Neurospora crassa</i>	C	I	I	CG	..(4)..	V	E	M	H	H	V	R	..(22)..	N	R	K	Q	V	P	L	C	K	T	H	I	G	L	H			
<i>Podospira anserina</i>	C	W	V	CG	..(4)..	I	E	M	H	H	V	K	..(19)..	N	R	K	Q	I	P	V	C	K	G	C	H	V	K	I	H		
<i>Allomyces macrogynus</i>	C	W	V	CG	..(4)..	I	E	M	H	H	V	R	..(20)..	N	R	K	Q	I	P	V	C	R	S	C	H	L	K	I	H		
<i>Schizosaccharomyces pombe</i>	C	A	A	C	..(4)..	V	E	M	H	H	V	R	..(19)..	N	R	K	Q	I	P	L	C	R	S	C	H	M	K	T	H		
<i>Sphingomonas</i> sp.	C	E	R	CG	..(5)..	C	E	V	H	H	V	R	..(17)..	R	Q	R	K	R	M	V	L	C	S	R	C	H	N	D	V	H	
<i>Lactococcus lactis</i>	C	E	L	CG	..(7)..	Y	E	I	H	H	V	N	..(18)..	Q	R	K	T	L	V	V	C	F	H	C	H	R	H	V	I	H	
<i>Marchantia polymorpha</i>	C	V	A	D	C	P	..(3)..	I	K	L	H	H	I	R	..(32)..	S	R	K	Q	I	P	L	C	P	K	H	H	S	D	F	H
<i>Sgmt-MutS</i>	C	E	I	CG	..(3)..	D	A	V	H	H	I	K	..(17)..	N	..	L	V	P	V	C	S	S	C	H	L	D	I	H			
<i>McrA - Escherichia coli</i>	C	E	N	CG	..(13)..	L	E	V	H	H	V	I	..(11)..	N	..	C	V	A	L	C	P	N	C	H	R	E	L	H			

**Figure 2**

*Sgmt-MutS* (a MutS homolog present in the mitochondrial proteome of the coral *Sarcophyton glaucum*) encodes an HNH endonuclease domain. A consensus of HNH endonucleases from a variety of sources<sup>17,18</sup> – Group-I and Group-II introns, bacteriocins and inteins – was constructed using embedding techniques<sup>20</sup>. In this method, a single sequence is selected from a set of blocks and enriched by replacing the conserved regions delineated by the blocks by consensus residues derived from the blocks. This consensus was used for iterative database searches using PSI-BLAST (Ref. 19). Presentation of the alignment of selected HNH endonuclease domains is made possible using MACBOXSHADE. HNH domains presented are from Group-II introns: *S. cerevisiae* (V00694), *N. crassa* (S07649), *P. anserina* (E48327), *A. macrogynus* (S63652), *S. pombe* (S10070), *S. aromaticivorans* (AF079317), *L. lactis* (X89922), *M. polymorpha* (S25952), the methyl-cytosine specific restriction endonuclease from *Escherichia coli*, *McrA* (Z19104), and *Sgmt-MutS*. The three-dimensional structure of HNH endonucleases identifies three histidine residues (black arrowheads) that comprise the active site<sup>31</sup>, whereas the white arrowhead highlights the asparagine residue previously thought to be part of the active site<sup>17</sup>. Thus, *Sgmt-MutS* is inferred to encode an active HNH endonuclease.

better than for other groups of previously identified HNH motifs, in bacteriocins, for example. In addition, *Sgmt-MutS* has a well-conserved di-cysteine motif upstream of the HNH endonuclease motif (Fig. 2). Therefore, the

HNH domain from *Sgmt-MutS* bears the closest resemblance to the endonuclease domains from Group-II introns. In Group-II introns, this HNH endonuclease domain is responsible for nicking a target site only on one strand within a



**Figure 3**

Phylogenomics of the MutS family of proteins. **(a)** The different MutS subfamilies, as proposed by Eisen<sup>11</sup>. The black boxes refer to regions of homology identified among all the MutS homologs (MSH) that include the predicted ATPase motifs. In the MutS2 lineage, additional domains are found at the C-terminal end, as represented in Sgmt-MutS (a MutS homolog present in the mitochondrial proteome of the coral *Sarcophyton glaucum*) (HNH) and MutS2 from *Bacillus subtilis* (Smr). In all other lineages, N-terminal extensions (gray boxes) bear regions of homology<sup>23</sup> that are proposed to constitute MutL-binding determinants. **(b)** Phylogenetic analysis of the MutS family of proteins. Although the general conclusions are the same as in a previous analysis<sup>11</sup>, additional sequence data allow better resolution in the MutS-I lineage. MutS-I proteins are involved in mismatch repair (blue), whereas MutS-II proteins are proposed to be exclusively involved in chromosome crossing-over and segregation (yellow). The biological function of Sgmt-MutS and MutS2 is unknown.

specific DNA sequence to initiate reverse transcription<sup>21</sup>. The presence of an HNH endonuclease domain in Sgmt-MutS, similar to the nicking endonucleases of Group-II introns, is satisfying because an endonuclease that introduces double-stranded breaks would be useless in MMR. It raises the intriguing possibility that Sgmt-MutS might also recognize a specific DNA sequence rather than a particular DNA modification. Sgmt-MutS thus appears unique in that both mismatch recognition and nicking activities appear to be present in the same protein (Fig. 1d). Sgmt-MutS might thus represent a compact MMR system.

#### Phylogenomics of the MutS family of proteins

With the advent of large-scale genome sequencing efforts, it has become possible now to assign (putative) functions to unknown genes by evolutionary analysis (i.e. phylogenomics)<sup>22</sup>. The MutS family of proteins is well conserved and possesses several 'blocks' of highly conserved amino acids that include the ATPase domain (Fig. 3a). A phylogenomic analysis (Fig. 3b) based on an

alignment of this segment divides MutS homologs into eight subfamilies<sup>11</sup> and provides insights into proposed functions for these subfamilies by extrapolating across orthologous groups. By virtue of greater sequence coverage, it has been possible to make subtle refinements to the original analysis<sup>11</sup>. MutS families are divided into two general groups. The MutS-I group includes the MutS1 from *E. coli*, as well as some eubacterial and archaeal homologs, and the eukaryotic MutS homologous groups, MSH1, MSH2, MSH3 and MSH6, all of which are involved in MMR. The proposed MutS-II group includes MutS2 representatives from eubacterial, archaeal and plant sources, as well as the MSH4 and MSH5 eukaryotic groups, which are not involved in mismatch repair but in chromosome crossover and segregation. Plant MutS2 proteins are clearly derived from a cyanobacterial source and probably represent a horizontal transfer from their plastids. Some sequence similarity can be found between MSH1–6 and MutS1 outside the proposed ATPase domain (Fig. 3a) but not within the MutS2 lineage<sup>23</sup>.

From an evolutionary standpoint, it seems likely that a prokaryotic MutS1-like

ancestor gave rise to MSH1 (fungal mitochondrial) and MSH2, with the latter subsequently diversifying into the MSH3 and MSH6. Note that in eukaryotic MMR, MSH2–MSH3 and MSH2–MSH6 heterodimers are the functional units. This suggests that in eukaryotes, one would always expect to find MSH2 but not necessarily MSH3 or MSH6. Lack of MSH3 or MSH6 would either indicate that MSH3 and MSH6 did not diversify from MSH2 (in early-branching eukaryotes) or that these were subsequently lost, because of functional redundancy. Both of these scenarios might be true. In *C. elegans* and in the diplomonad *Giardia lamblia*, sequence coverage is quite extensive across the entire genome. In both cases, the MSH2, MSH4 and MSH5 homologs, but neither MSH3 nor MSH6, are readily detected by BLAST searches (confirmed by phylogenetic analyses). In these species, we predict that MSH2 homodimers act as the functional subunits in MMR.

#### MutS without MutL?

Although it is clear that most organisms that possess MutS homologs do not contain a MutH homolog, it is puzzling that some of these do not have MutL homologs either. It has been suggested

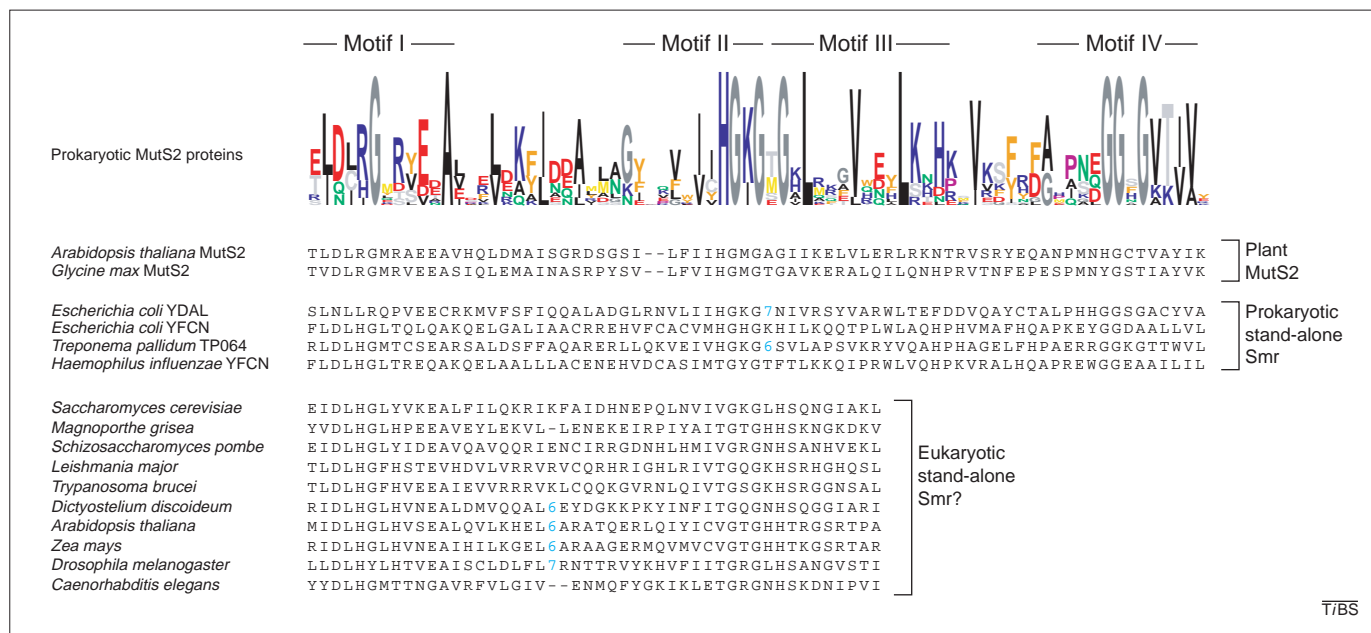


Figure 4

Smr domains in prokaryotic and eukaryotic species, and identification of Smr domains in various organisms. An alignment of Smr domains from prokaryotic MutS2 proteins<sup>26</sup> is shown in Logos format<sup>32</sup> with previously identified motifs highlighted. In Logos format, the size of each residue at a particular position is proportional to its incidence in that position, and the total height of all the residues in the position is proportional to the relative conservation (information content) of that position. PSI-BLAST (Ref. 19) searches were used to identify additional Smr homologs shown below the Logos (numbers of amino acids deleted for purposes of alignment are shown in light blue). This approach is essentially similar to that used in the initial identification of Smr (Ref. 26), except that the eukaryotic Smr homologs identified in TBLASTN searches versus various nucleotide databases (dbest, etc.) were subsequently used to obtain the various putative eukaryotic homologs. Note that in the proposed eukaryotic Smr homologs, Motif IV cannot be found at statistically significant levels.

previously that these organisms lack an MMR system<sup>24</sup>. However, all the known *MutS*-bearing genomes that lack *MutL* contain a MutS2-encoding gene<sup>25,26</sup>. In cases where two genes encoding MutS homologs, MutS1 and MutS2, are present in the same genome, only one gene for a MutL homolog is present. The simplest explanation for these observations is that the MutS2 lineage operates in a MutL-independent fashion. Because MutL acts as a scaffold between MutS and the nicking endonuclease<sup>2</sup>, direct linking of the two domains would obviate the necessity for a scaffold linker. Sgmt-MutS could represent an example of such an arrangement in the mitochondrial proteome of the coral *S. glaucum*, where MutS is predicted to be linked directly to a nicking HNH endonuclease (Fig. 1c), and MutL homologs are absent.

Extending the analogy of the Sgmt-MutS protein to other members of the MutS2 lineage, all other MutS2 proteins represent a fusion of two domains, the MutS domain and a smaller C-terminal domain that has been referred to as the small MutS-related (Smr) domain (Fig. 4, Ref. 26). We speculate that the Smr domain acts in analogous fashion to the HNH domain as a nicking endonuclease (Figs 1e, 3a). Although biochemical proof of the enzymatic activity of Smr is still

lacking, the fusion of two proteins in evolutionary history is generally considered suggestive of them being involved in the same pathway (see Ref. 27, for example). Sequences encoding Smr domains are found in many lineages of eukaryotic genomes (Fig. 4), and its ubiquity suggests an important role. In many species, Smr is present as a stand-alone open reading frame (i.e. not fused to MutS) and would thus require a scaffold linker to interact with MutS, a role that might be carried out by MutL (Fig. 1f). It is not known whether prokaryotic MutS2 proteins function in mismatch repair or chromosome segregation, or both. The function of Smr is also unknown but might be restricted to chromosome crossing-over and segregation, because MutS2 proteins fall within the MutS-II lineage (Fig. 3b). Other members of the MutS-II lineage, the eukaryotic MSH4 and MSH5 proteins, appear to act in a MutL-dependent manner<sup>9,10</sup> in meiotic chromosome crossing-over. The role of Smr in chromosome segregation is under investigation.

MutS2 proteins represent the only MutS lineage that operates independently of MutL. Other MutS lineages are homologous beyond the core ATPase domain, and it is possible that these 'extra' regions might represent binding determinants for MutL (Ref. 23, Fig. 3a).

This argues against the possibility of MutS2 being an ancestral lineage, consistent with the framework in Fig. 3b (Ref. 11). Thus, a prokaryotic MutL-dependent MutS2 ancestor probably gave rise to the eukaryotic MSH4 and MSH5 lineages that now play important roles in chromosome crossing-over and segregation, possibly with the aid of Smr. After a fusion of MutS2 and Smr, the MutL requirement might have been lost. Smr might have been functionally replaced by a nonhomologous HNH endonuclease domain in Sgmt-MutS.

Although the presence of MutS is a good indicator for the ability of the genome to recognize mismatches, no other ubiquitous component is apparent. Thus, MutS-bearing species can lack either MutL (*Helicobacter pylori*, *Campylobacter jejuni* – both of which possess MutS2 only) or Smr (*Chlamydia* species, *Rickettsia prowazekii*). In species lacking MutL, MutS is fused to another domain that might represent a nicking endonuclease (Sgmt-MutS, MutS2-Smr). In all other cases, MutL might act as a scaffold between MutS and Smr or other (unknown) nicking endonucleases.

#### What is the signal?

It is crucial for the MMR machinery to be able to distinguish between the

parental strand and the newly synthesized daughter strands in DNA replication. Although MutH and MED1 rely on DNA methylation to distinguish between parental and daughter strands (Fig. 1a,b), it is not clear how the HNH domain in Sgmt-MutS makes this distinction. One possibility is that, like other members of the MutS-II lineage, the only function of Sgmt-MutS is in chromosome crossing-over and segregation<sup>11</sup>; nicking a strand in a DNA-sequence-specific manner might suffice to resolve the Holliday-junction-like structures<sup>8</sup>. If, however, Sgmt-MutS plays a role in MMR, the DNA target sites (typically nonpalindromic) recognized by the HNH endonuclease would be under selective pressure to be properly oriented in the mitochondrial genome relative to the origin of replication.

We have suggested that Smr might represent the more common form of a 'resolving endonuclease'. However, the general nicking endonuclease involved in mismatch repair is still elusive. Like Dam-MutH, the recruitment of R-M enzymes for DNA repair function might be generally true for prokaryotic genomes, in which multiple R-M systems as well as HNH endonucleases (of unassigned function) are present. For example, *visB*, which is a stand-alone HNH endonuclease phylogenetically related to Group-II introns and Sgmt-MutS, might be a good candidate for a nicking endonuclease in *Bacillus subtilis*. Successive nonhomologous recruitments of different R-M systems in mismatch repair (where the former restriction endonuclease is now serving as a nicking endonuclease) might have blurred the identification of MutH analogs in different prokaryotic species.

In eukaryotic genomes, it is relatively rare to encounter genes encoding such 'nicking endonucleases', except in retrotransposable elements<sup>28</sup>. To account for the lack of candidate endonucleases, a previous model<sup>29</sup> has suggested that mismatch repair might proceed independently of a nicking endonuclease, using the tethering of MMR complexes to the DNA replication machinery to stall the replication fork until errors are corrected. As DNA synthesis is discontinuous on the lagging strand, it is likely that there is no requirement for a (hypothetical) nicking endonuclease to recognize this strand. However, leading-strand synthesis is continuous, and it is more likely that the tethering of MMR and replication machineries via the sliding clamp protein PCNA (Ref. 30) might

be involved in strand recognition in conjunction with an as-yet-unknown nicking endonuclease. The identification of MED1 reiterates the likelihood of finding such additional eukaryotic endonucleases involved in MMR. The identification of Smr (Ref. 25) and Sgmt-MutS-HNH might provide a new paradigm for identification of additional components of the MutS-containing pathways, and their roles in repair and meiosis.

### Conclusions

The unexpected finding of a gene encoding a 'fused' MutS2-HNH (mismatch recognition-nicking endonuclease) protein in a coral mitochondrial genome (Sgmt-MutS) has raised the possibility that the MutS2 group operates independently of MutL. The more general MutS2 members possess a fused MutS-Smr domain, leading to the hypothesis that Smr also represents a nicking endonuclease likely to be involved in meiotic chromosome crossing-over and segregation.

This study suggests the possibility of a MutL-independent mode of MMR. The proposed existence of dual recognition-incision (MutS-'nickase') enzymes also suggests potential strategies for complementing defects in the MMR pathway, perhaps by introducing a fused *MutS-MED1* gene into mammalian genomes. MMR is a crucial process for an organism to keep its mutational load in check. Resolution of mismatched intermediates is also essential in meiosis. The myriad MMR-like systems highlighted in this report reflects the evolutionary opportunism shown by different genomes to accomplish these important processes. Thus, lack of homologous proteins might not suffice to rule out the presence of alternative means to the same end.

### Acknowledgements

We thank G. Smith, J. Eisen, J. Smothers, P. Ng, J. Farah, D. Vermaak and K. Ahmad for helpful discussions and their comments on this article. This work was supported by N.I.H. grant GM-29009 to S.H. H.S.M. is a postdoctoral fellow of the Helen Hay Whitney Foundation.

### References

- Kolodner, R.D. *et al.* (1994) Human mismatch repair genes and their association with hereditary non-polyposis colon cancer. *Cold Spring Harbor Symp. Quant. Biol.* 59, 331–338
- Modrich, P. (1989) Methyl-directed DNA mismatch correction. *J. Biol. Chem.* 264, 6597–6600
- Chi, N.W. 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

- Sia, E.A. *et al.* (1997) Genetic control of microsatellite stability. *Mutat. Res.* 383, 61–70
- Marsischky, G.T. *et al.* (1996) Redundancy of *Saccharomyces cerevisiae* MSH3 and MSH6 in MSH2-dependent mismatch repair. *Genes Dev.* 10, 407–420
- Pochart, P. *et al.* (1997) Conserved properties between functionally distinct MutS homologs in yeast. *J. Biol. Chem.* 272, 30345–30349
- 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
- Zalavsky, J. *et al.* (1999) Crossing over during *Caenorhabditis elegans* meiosis requires a conserved MutS-based pathway that is partially dispensable in budding yeast. *Genetics* 153, 1271–1283
- Prolla, T.A. *et al.* (1994) MLH1, PMS1, and MSH2 interactions during the initiation of DNA mismatch repair in yeast. *Science* 265, 1091–1093
- Wang, T.F. *et al.* (1999) Functional specificity of MutL homologs in yeast: evidence for three Mlh1-based heterocomplexes with distinct roles during meiosis in recombination and mismatch correction. *Proc. Natl. Acad. Sci. U. S. A.* 96, 13914–13919
- Eisen, J.A. (1998) A phylogenomic study of the MutS family of proteins. *Nucleic Acids Res.* 26, 4291–4300
- Ban, C. and Wang, W. (1998) Structural basis for MutH activation in *E. coli* mismatch repair and relationship of MutH to restriction endonucleases. *EMBO J.* 17, 1526–1534
- Bellacosa, A. *et al.* (1999) MED1, a novel human methyl-CpG-binding endonuclease, interacts with DNA mismatch repair protein MLH1. *Proc. Natl. Acad. Sci. U. S. A.* 96, 3969–3974
- Bader, S. *et al.* (1999) Somatic frameshift mutations in the MBD4 gene of sporadic colon cancers with mismatch repair deficiency. *Oncogene* 18, 8044–8047
- Pont-Kingdom, G.A. *et al.* (1995) A coral mitochondrial mutS gene. *Nature* 375, 109–111
- Pont-Kingdom, G.A. *et al.* (1998) Mitochondrial DNA of the coral *Sarcophyton glaucum* contains a gene for a homologue of bacterial MutS: a possible case of gene transfer from the nucleus to the mitochondrion. *J. Mol. Evol.* 46, 419–431
- Shub, D.A. *et al.* (1994) Amino acid sequence motif of group I intron endonucleases is conserved in open reading frames of group II introns. *Trends Biochem. Sci.* 19, 402–404
- Gorbalenya, A.E. (1994) Self-splicing group I and group II introns encode homologous (putative) DNA endonucleases of a new family. *Protein Sci.* 3, 1117–1120
- Altschul, S.F. *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402
- Henikoff, S. and Henikoff, J.G. (1997) Embedding strategies for effective use of information from multiple sequence alignments. *Protein Sci.* 6, 698–705
- Zimmerly, S. *et al.* (1995) Group II intron mobility occurs by target DNA-primed reverse transcription. *Cell* 82, 545–554
- Eisen, J.A. (1998) Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. *Genome Res.* 8, 163–167
- Aravind, L. *et al.* (1999) Conserved domains in DNA repair proteins and evolution of repair systems. *Nucleic Acids Res.* 27, 1223–1242
- Eisen, J.A. *et al.* (1997) Gastrogenomic delights: a movable feast. *Nat. Med.* 3, 1076–1078
- Eisen, J.A. and Hanawalt, P.C. (1999) A phylogenomic study of DNA repair genes, proteins, and processes. *Mutat. Res.* 435, 171–213
- Moreira, D. and Philippe, H. (1999) Smr: a bacterial and eukaryotic homologue of the C-terminal region of the MutS2 family. *Trends Biochem. Sci.* 24, 298–300
- Marcotte, E.M. *et al.* (1999) A combined algorithm for genome-wide prediction of protein function. *Nature* 402, 83–86
- Yang, J. *et al.* (1999) Identification of the endonuclease domain encoded by R2 and other site-specific, non-long terminal repeat retrotransposable elements. *Proc. Natl. Acad. Sci. U. S. A.* 96, 7847–7852
- Jiricny, J. (1998) Replication errors: cha(lle)nging the genome. *EMBO J.* 17, 6427–6436
- Umar, A. *et al.* (1996) Requirement for PCNA in DNA mismatch repair at a step preceding DNA resynthesis. *Cell* 87, 65–73
- Ko, T.P. *et al.* (1999) The crystal structure of the DNase domain of colicin E7 in complex with its inhibitor Im7 protein. *Struct. Fold Des.* 7, 91–102
- Schneider, T.D. and Stephens, R.M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* 18, 6097–6100