

Structure of the MutL α C-terminal domain reveals how Mlh1 contributes to Pms1 endonuclease site

Emeric Gueneau^{1,2}, Claudine Dherin^{3,4}, Pierre Legrand⁵, Carine Tellier-Lebegue^{1,2}, Bernard Gilquin^{1,2}, Pierre Bonnesoeur^{1,2}, Floriana Londino^{1,2}, Cathy Quemener^{1,2}, Marie-Hélène Le Du^{1,2}, Josan A Márquez⁶, Mireille Moutiez⁷, Muriel Gondry⁷, Serge Boiteux^{3,4} & Jean-Baptiste Charbonnier^{1,2}

Mismatch-repair factors have a prominent role in surveying eukaryotic DNA-replication fidelity and in ensuring correct meiotic recombination. These functions depend on MutL-homolog heterodimers with Mlh1. In humans, MLH1 mutations underlie half of hereditary nonpolyposis colorectal cancers (HNPCCs). Here we report crystal structures of the MutL α (Mlh1-Pms1 heterodimer) C-terminal domain (CTD) from *Saccharomyces cerevisiae*, alone and in complex with fragments derived from Mlh1 partners. These structures reveal structural rearrangements and additional domains in MutL α as compared to the bacterial MutL counterparts and show that the strictly conserved C terminus of Mlh1 forms part of the Pms1 endonuclease site. The structures of the ternary complexes between MutL α (CTD) and Exo1 or Ntg2 fragments reveal the binding mode of the MIP-box motif shared by several Mlh1 partners. Finally, the structures provide a rationale for the deleterious impact of MLH1 mutations in HNPCCs.

The DNA mismatch-repair (MMR) system corrects replication errors and is thus essential for maintaining genome integrity^{1,2}. MMR factors are also involved in other DNA-metabolic processes, ensuring the quality of meiotic recombination events and activating checkpoints and apoptotic responses to several classes of DNA damage, as well as regulating somatic hypermutation during B-cell maturation³. Defective MMR function is associated with deleterious effects on human health, such as those observed in HNPCCs, also called Lynch syndrome, and the development of sporadic tumors in a number of tissues^{4,5}.

In eukaryotes, MMR is initiated by either binding of the MutS homolog MutS α (Msh2–Msh6 heterodimer) to base-base mismatches or binding of the MutS β homolog (Msh2–Msh3 heterodimer) to insertion and deletion loops. MutS α forms a sliding clamp structure that can freely diffuse on the DNA duplex and that binds tightly to mismatch-containing structures^{6–8}. In a second step, a MutL homolog is recruited to a lesion-bound MutS homolog in an ATP-dependent manner, and it coordinates downstream steps in the repair reaction⁹. The eukaryotic MutL α heterodimer (constituted by Mlh1–Pms1 in *S. cerevisiae* and MLH1–PMS2 in humans) is the predominant MutL homolog active in MMR. Recent studies have shown that MutL α has a latent endonuclease that is dependent on the integrity of a conserved DQHA(X₂)E(X₄)E motif located in the C-terminal region of yeast Pms1 or human PMS2 subunits^{10–12}. An *S. cerevisiae* mutant containing a variant of MutL α with a compromised endonuclease activity (the Pms1(E707K) mutation in the metal-binding

motif) exhibits a strong mutator phenotype as seen in Pms1 knockout (*pms1Δ*), which demonstrates that in yeast the endonuclease activity of MutL α is essential for MMR¹¹. Similarly, endonuclease-deficient *Pms2*^{E702K} knock-in mice display increased genomic mutation rates and a strong cancer predisposition¹³.

MutL α is a heterodimer composed of Mlh1 and Pms1 in yeast and of MLH1 and PMS2 in humans, whereas bacterial MutLs are homodimeric^{14,15}. Biochemical and structural data show that MutL α is formed by an N-terminal domain (NTD) and a CTD connected with long linker regions that confer highly dynamical properties to the heterodimer (Fig. 1a)^{16,17}. The NTD of MutL α is composed of the NTDs of Mlh1 and Pms1, which possess an ATPase activity that is necessary for MMR^{18–20}. The CTD of MutL α is composed of the CTDs of Mlh1 and Pms1, and it is necessary for heterodimerization²¹. As mentioned above, the CTD of Pms1 in yeast (and of PMS2 in humans) possesses a conserved metal-binding motif essential for the endonuclease activity of MutL α ^{10,11,22,23}. To understand the role of the eukaryotic MutL α in the MMR pathway, it is critical to examine structural information. The crystal structures of NTD fragments of yeast Pms1 and human PMS2 showed that their structures closely resemble the ATPase fragment of *Escherichia coli* MutL-NTD^{18,19}. In contrast, no information about the crystal structure of the eukaryotic MutL α (CTD) heterodimer has been reported, to our knowledge.

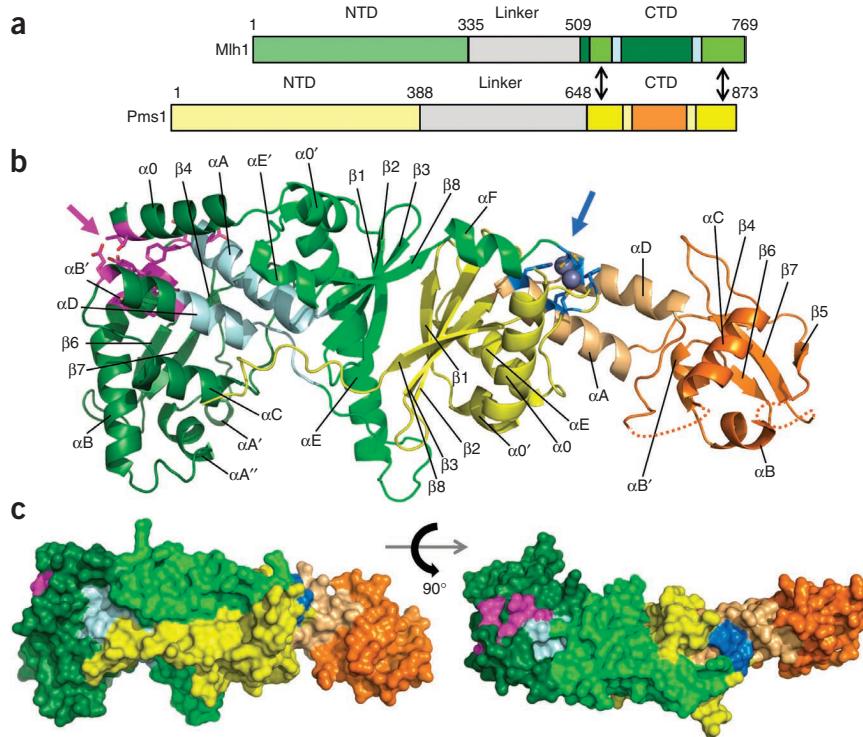
The structures of homodimeric bacterial MutL C-terminal and N-terminal regions provided insights into the overall arrangement

¹Unité Mixte de Recherche 8221, Commissariat à l'Energie Atomique, Centre National de la Recherche Scientifique, Institut de Biologie et Technologies de Saclay, Gif-sur-Yvette, France. ²Université Paris Sud, Orsay, France. ³Unité Mixte de Recherche 217, Institut de Radiobiologie Cellulaire et Moléculaire, Fontenay-aux-Roses, France. ⁴Unité Propre de Recherches 4301, Centre de Biophysique Moléculaire, Centre National de la Recherche Scientifique, Orléans, France. ⁵SOLEIL Synchrotron, L'Orme de Merle, Saint Aubin, Gif-sur-Yvette, France. ⁶High Throughput Crystallization Laboratory, European Molecular Biology Laboratory, Grenoble, France.

⁷Service d'Ingénierie Moléculaire des Protéines, Commissariat à l'Energie Atomique, Institut de Biologie et Technologies de Saclay, Gif-sur-Yvette, France. Correspondence should be addressed to J.-B.C. (jb.charbonnier@cea.fr) or S.B. (serge.boiteux@cnrs-orleans.fr).

Received 18 July 2012; accepted 4 January 2013; published online 24 February 2013; doi:10.1038/nsmb.2511

Figure 1 Crystal structure of MutL α (CTD). (a) Schematic diagram of the *S. cerevisiae* Mlh1 and Pms1 proteins. The limits of the CTDs correspond to the limits of the folded domains observed in the present crystal structures. (b) Overall view of the Mlh1(CTD)-Pms1(CTD) heterodimer. The Mlh1 dimerization and regulatory domains are colored respectively in green and dark green. The Pms1 dimerization and regulatory domains are colored respectively in yellow and orange. The Mlh1 regulatory domain includes the S2 site (magenta, indicated by magenta arrow), involved in the interactions with Exo1, Ntg2 and other proteins containing a MIP-box motif. The connector regions between the Mlh1 and Pms1 domains are colored respectively in pale cyan and light orange. The Pms1 connector contains the endonuclease site (blue, indicated by blue arrow), and two zinc atoms (gray spheres). Mlh1 and Pms1 secondary structure elements are numbered according to *B. subtilis* MutL numeration (0 and ' indicate additional secondary structure elements compared to bacterial proteins). (c) Surface representation of the MutL α (CTD) heterodimer. The left figure has the same orientation as in b, and the right figure is rotated 90° around the horizontal axis.



of these regions in prokaryotes^{14,24–26}. The *E. coli* MutL(CTD) regions are organized in two distinct subdomains: a dimerization and a regulatory subdomain^{24,27}. As with eukaryotic MutL heterodimers, the MutL homodimers of most eubacteria, with the notable exception of *E. coli*, have an endonuclease activity, resident in the CTD, that creates a nick used for mismatch excision²⁸. The prokaryotic MutL(CTD)s share only limited sequence identities (10–18%) with the eukaryotic MutL α (CTD)s. These low homologies limit the use of prokaryotic structures in accurate modeling of eukaryotic MutL α (CTD) and analysis of the effects of cancer-causing mutations^{29–31}. Even the location of the dimeric interface in the MutL α (CTD) was recently debated^{30,32}.

In *S. cerevisiae*, the Mlh1(CTD) physically interacts with Mlh2 and Mlh3 to form the MutL β and MutL γ heterodimers, respectively¹⁵. In addition, Mlh1(CTD) interacts with proteins such as Exo1, Sgs1 (DNA helicase of the RecQ family) and Ntg2 (DNA N-glycosylase and apurinic or apyrimidinic lyase)^{33–35}. Mlh1 interacts with the latter proteins through a common (R/K)SK(Y/F)F motif, called the MIP box (for Mlh1 interacting protein box)^{33,36}. The Mlh1 MIP binding site was named the S2 site, to differentiate it from the Mlh1-Pms1 heterodimerization region, called the S1 site. An allele that disrupts interaction at S2 site, *mlh1*-E682A, was shown to affect Exo1-dependent functions such as mutation avoidance and meiotic recombination^{36,37}.

We set out to determine the crystal structures of the *S. cerevisiae* MutL α (Mlh1-Pms1 heterodimer) CTD region alone and in complex with fragments containing the MIP-box motifs of the Exo1 and Ntg2 proteins. We reveal the heterodimerization interface between Mlh1 and Pms1 (S1 site). These structures highlighted the unanticipated direct participation of the highly conserved last residue in Mlh1, Cys769, in the formation of the Pms1 endonuclease site and also showed the interaction surfaces between the MutL α (CTD) and peptides derived from the Exo1 and Ntg2 proteins (S2 site). Finally, a model of the human MutL α (CTD) based on the present structures provided a rationale for human HNPCC-causing mutations²⁹.

RESULTS

Crystal structure of the MutL α (CTD) from *S. cerevisiae*

The MutL α (CTD) preparation used for crystallization had no detectable ATP-Mn²⁺-dependent endonuclease activity on double-stranded plasmid DNA (Online Methods and **Supplementary Fig. 1a,b**). We observed a lack of endonuclease activity with concentrations of MutL α (CTD) up to 5 μ M. Therefore, the endonuclease activity of the MutL α (CTD) was less than 1% of that displayed by the full-length yeast MutL α ¹¹. This is coherent with results from studies of the MutL(CTD) from *Bacillus subtilis*, which has no detectable endonuclease activity²⁶, or the MutL(CTD) from *Aquifex aeolicus*, which has a residual endonuclease activity³⁸. To investigate the DNA-binding activity of the MutL α (CTD) from *S. cerevisiae*, we performed electrophoretic mobility shift assays. The results showed that the yeast MutL α (CTD) can bind to 59-base pair (bp) homoduplex DNA with an apparent dissociation constant (K_d^{app}) of 2.2 μ M (**Supplementary Fig. 1c**). For comparison, an apparent K_d value of 0.16 μ M was determined for the NTD of Pms1 from *S. cerevisiae* by using a 41-bp DNA probe¹⁸.

The crystal structure of MutL α (CTD) was determined at 2.5-Å resolution by using the single-wavelength anomalous diffraction (SAD) method with selenomethionine-labeled proteins (**Table 1**). The CTD regions of Mlh1 and Pms1 are composed of two domains with a general organization similar to that of the dimerization and regulatory domains of prokaryotic MutL(CTD)s^{24–26} (**Fig. 1b,c** and **Supplementary Fig. 2a,b**). Both Mlh1 and Pms1 dimerization domains share a similar structure and can be superimposed (r.m.s. deviations of 2.5 Å over 63 C α positions; **Supplementary Fig. 2c**). They are constituted by the N- and C-terminal ends of Mlh1(CTD) (residues 525–560 and 702–769) and Pms1(CTD) (residues 649–701 and 822–873). They form a core with a four-stranded β -sheet (β 1, β 2, β 3 and β 8) and a helix (α E). Compared to bacterial MutL(CTD)s, Mlh1 and Pms1 present two additional helices at their N termini, which we call α 0 and α 0'. The Pms1 helices α 0 and α 0' are located in the Pms1 dimerization domain. The Mlh1 helices α 0 and α 0' are located, respectively, in the Mlh1 regulatory and dimerization domains. Mlh1 has

a helix, α E, that is 13 amino acids (aa) longer than its equivalent in Pms1 and other bacterial MutL(CTD)s. Mlh1 thus presents a 22-Å protrusion compared to Pms1(CTD) and MutL(CTD)s.

The regulatory domains are mainly constituted by the inner region of the Mlh1(CTD) and its first helix α 0 and by the inner region of Pms1(CTD) (residues 509–524 and 578–689 for Mlh1 and residues 718–805 for Pms1; Fig. 1b,c). They contain a four-stranded β -sheet (β 4, β 5, β 6 and β 7) and two helices (α B and α C). These regions of Mlh1 and Pms1 superimpose with an r.m.s. deviation of 3.1 Å for 69 C α atoms, and they superimpose with the equivalent regions of the bacterial MutL(CTD)s with r.m.s. deviations between 2.2 Å and 3.2 Å (Supplementary Fig. 2d). Mlh1 presents two additional small helices, α A' and α A'', in place of strand β 5 in Pms1(CTD). The dimerization and regulatory regions of Mlh1(CTD) and Pms1(CTD) are linked by two connectors constituted by the helices α A and α D and the linker between the helices α D and α E. Superimposition of the Mlh1(CTD) and Pms1(CTD) regions shows a large reorientation of the Pms1 regulatory domain, compared to that of Mlh1, caused by different orientation of helix α C (Supplementary Fig. 2e). Pms1 thus presents an 'open' orientation that results in the disruption of the packing between helices α C and α D observed in Mlh1(CTD) and other bacterial MutL(CTD)s. We extended this analysis to the other structures of MutL(CTD)s deposited in the RCSB Protein Data Bank.

We observed that the regulatory domains adopt various positions relative to the dimerization domains after superimposition of the dimerization domains of all MutL(CTD)s (Supplementary Fig. 2f). The positions of the regulatory domains vary from a 'closed' conformation relative to the endonuclease site, in *Neisseria gonorrhoeae* MutL(CTD) (PDB 3NCV) or *S. cerevisiae* Mlh1(CTD), to an open conformation, in *S. cerevisiae* Pms1(CTD) or in one crystal form of *B. subtilis* MutL(CTD) (PDB 3KDG).

MutL α (CTD) has a large heterodimerization interface

The Mlh1(CTD)–Pms1(CTD) heterodimer buries an accessible surface area of 3,700 Å² (1,800 Å² from Mlh1 and 1,900 Å² from Pms1). This area is two times larger than the buried area observed in the bacterial MutL(CTD) homodimers^{24–26}. The heterodimerization interface involves 53 residues from Mlh1 and Pms1 and is spread over three main patches. The first patch (patch I) is constituted by the packing of the four-strand β -sheet (β 1, β 2, β 3 and β 8) of Mlh1 with the equivalent β -sheet of Pms1. The core of this patch involves mainly six Mlh1 and six Pms1 hydrophobic residues (Fig. 2a) and accounts for 36% of the whole buried area of the heterodimer (1,330 Å² buried area). Patch I has a pseudosymmetric character involving the same positions in the sequence alignment of Mlh1 and Pms1 (Supplementary Fig. 2a).

A second patch (patch II) in the heterodimerization S1 site involves the last 12 residues of Pms1 (aa 862–873), which adopt a well-structured

Table 1 Data collection and refinement statistics

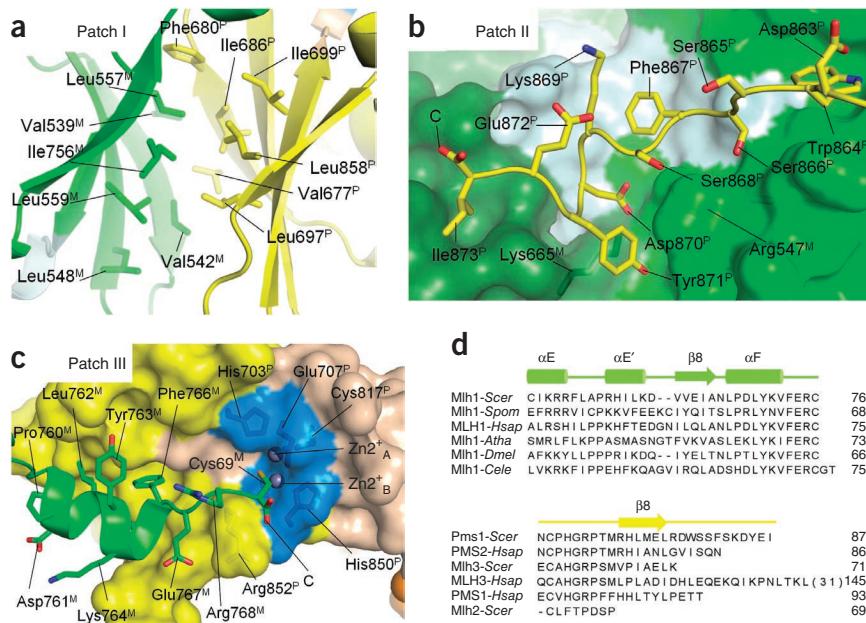
	MutL α (CTD) native	MutL α (CTD) SAD	MutL α (CTD)–Ntg2 molecular replacement	MutL α (CTD)–Exo1 molecular replacement
Data collection				
Space group	<i>C</i> 2	<i>C</i> 2	<i>C</i> 2	<i>C</i> 2
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	187.2, 66.1, 74.0	190.9, 66.2, 74.5	190.9, 66.3, 74.1	193.7, 66.1, 74.5
α , β , γ (°)	90.0, 90.8, 90.0	90.0, 90.8, 90.0	90.0, 90.2, 90.0	90.0, 91.3, 90.0
	Se-Edge			
Wavelength	0.9800	0.9791	0.9801	0.9801
Resolution (Å) ^a	50–2.50 (2.65–2.50)	45–3.12 (3.31–3.12)	50–2.7 (2.86–2.7)	50–3.04 (3.23–3.04)
R_{sym}	9.80 (73.7)	16.1 (61.0)	9.8 (76.8)	13.6 (79.3)
$I/\sigma I$	9.8 (2.0)	12.3 (3.7)	12.2 (2.3)	10.0 (2.2)
Completeness (%)	99.5 (98.4)	98.7 (96.0)	99.3 (96.8)	99.5 (98.1)
Redundancy	3.7 (3.7)	9.0 (9.0)	5.6 (5.6)	5.6 (5.5)
Refinement				
Resolution (Å)	39.8–2.5		31.6–2.7	48.4–3.04
No. reflections	31,477		25,738	18,219
$R_{\text{work}} / R_{\text{free}}$	17.0 / 21.0		16.9 / 19.2	18.0 / 19.6
No. atoms				
Protein	3,808		3,942	3,877
Ligand/ion ^b	224		188	59
Water	180		129	56
<i>B</i> factors				
Protein	77.0		85.8	99.0
Ligand/ion	71.4		90.6	65.4
Water	53.6		63.6	57.6
r.m.s. deviations				
Bond lengths (Å)	0.01		0.01	0.01
Bond angles (°)	1.2		1.18	1.18

^aThe values in parentheses are for the highest-resolution shell. ^bLigand/ion includes two zinc atoms in the Pms1 endonuclease site of each structure.

extended conformation that spreads from the dimerization region to the linker and regulatory domains of Mlh1 (Figs. 1b,c and 2b). This second patch accounts for 1,400 Å² (38% of the total buried area) and involves two salt bridges (Asp870^P–Arg547^M and Asp870^P–Lys665^M; the superscripts M and P stand for Mlh1 and Pms1). Finally, patch III involves the last 11 residues of Mlh1 (residues 759–769), which are structured into the helix α F (aa 759–766), and an extended structure formed by three amino acids, Glu767, Arg768 and Cys769. Patch III accounts for 970 Å² (26% of the total buried area; Fig. 2c). The C terminus of Mlh1 extends farther than the corresponding region of bacterial MutL(CTD), and the last residue of Mlh1, Cys769^M, reaches the Pms1 metal-binding site. The Mlh1 C terminus is highly conserved in eukaryotes, in particular the last four invariant residues 766–FERC-769 (Fig. 2d).

In yeast, Mlh1 can form heterodimers with Pms1 (MutL α) but also with Mlh2 (MutL β) and Mlh3 (MutL γ)¹⁵. The yeast Mlh3 protein is 11 aa shorter than Pms1 and stops at the end of strand β 8 so that all interactions described on patch II are not present in the Mlh1–Mlh3 heterodimer. To evaluate the role of patch II in the integrity of the Mlh1(CTD)–Pms1(CTD) complex, we performed yeast two-hybrid (Y2H) experiments (Supplementary Table 1). We showed that a Pms1(CTD) variant lacking the last ten C-terminal residues does not interact with the Mlh1(CTD) region, which indicated the importance of patch II for the yeast Mlh1–Pms1 heterodimer's formation and stability. A Y2H assay also revealed the central

Figure 2 MutL α heterodimerization patches. (a) Patch I of the MutL α (CTD) interface. Patch I is constituted by packing between the Mlh1(CTD) and the Pms1(CTD) β -sheets located in the dimerization domain. The core of patch I mainly consists of contacts between six Mlh1 (green) and six Pms1 (yellow) hydrophobic residues conserved in bacterial MutL(CTD)s. (b) Patch II of the MutL α (CTD) interface. Patch II consists of interactions between the extended Pms1 C terminus (stick) and the Mlh1(CTD) dimerization and regulatory domains (surface). (c) Patch III of the MutL α (CTD) interface. Patch III consists of the interactions between the Mlh1 C terminus with the Pms1(CTD). The two zinc atoms observed in the Pms1 endonuclease site are represented by gray spheres. The blue patch corresponds to the Pms1 residues involved in the metal-binding site. (d) Multiple sequence alignments. Top, sequence alignments of the Mlh1 C terminus. Bottom, sequence alignments of the C termini of *S. cerevisiae* Pms1, Mlh3 and Mlh2 proteins and human PMS2, MLH3 and PMS1 proteins. Scer, *S. cerevisiae*; Spom, *Schizosaccharomyces pombe*; Hsap, *Homo sapiens*; Atha, *Arabidopsis thaliana*; Dmel, *Drosophila melanogaster*; Cele, *Caenorhabditis elegans*. For MLH3-Hsap, (31) indicates that there are 31 additional amino acids preceding the C terminus in this protein.



role for the salt bridge Asp870^P-Arg547^M, as Pms1(D870A) and Mlh1(R547A) mutants do not interact with wild-type Mlh1(CTD) and wild-type Pms1(CTD), respectively. However, Mlh1(R547A) partially suppressed the spontaneous mutagenesis phenotype of *mlh1* Δ yeast strains, which suggested a residual interaction between Mlh1(R547A) and Pms1 *in vivo*³⁶. In agreement with the absence

of patch II in the Mlh3 sequence, the mutant Mlh1(R547A) interacted with Mlh3 as the wild type did. Y2H analyses also showed that deletion of the ten final residues of Mlh1 (Δ C10) did not affect interactions with Pms1, Mlh3 and other protein partners. In summary, the MutL α heterodimerization (S1 site) is mainly mediated by patches I and II.

The last Mlh1 residue participates in Pms1's endonuclease site

The eukaryotic MutL α heterodimers present a latent endonuclease activity confined in the CTD of the Pms1 subunit (PMS2 in humans) and dependent on the integrity of a conserved DQHA(X)₂E(X)₄E motif. Three additional conserved motifs of the Pms1(CTD) region (ACR, C(P/N)HGR and F(X)R motifs) have been proposed to form the metal-binding site of the endonuclease, on the basis of sequence alignments¹². In the crystal structure of the MutL α (CTD), we observed two metal ions in the site defined by the first three conserved motifs in Pms1 (Supplementary Fig. 3a–c). The X-ray fluorescence spectra of the MutL α (CTD) crystals produced a strong signal characteristic of zinc atoms. To confirm this, we collected diffraction data on a unique crystal both at the Zn²⁺ absorption peak and at a pre-edge wavelength. The resulting anomalous difference electron density maps presented a strong peak at the position occupied by the

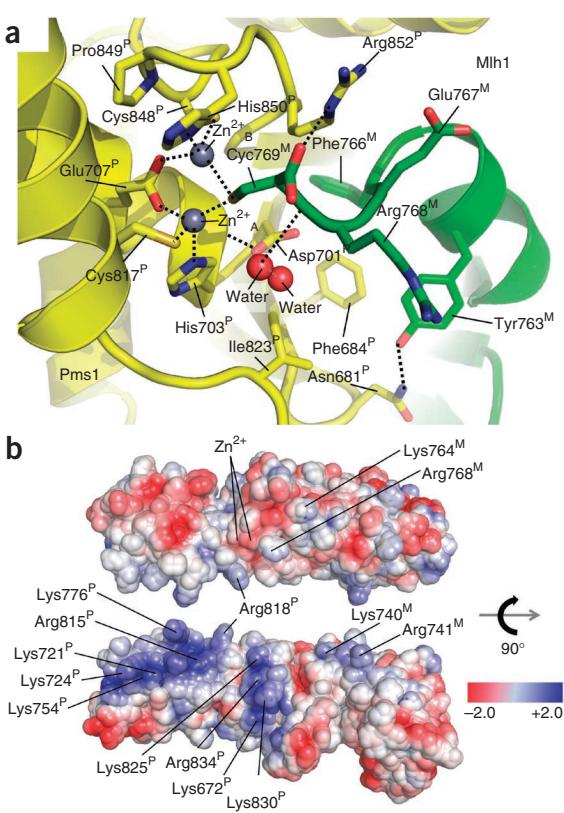


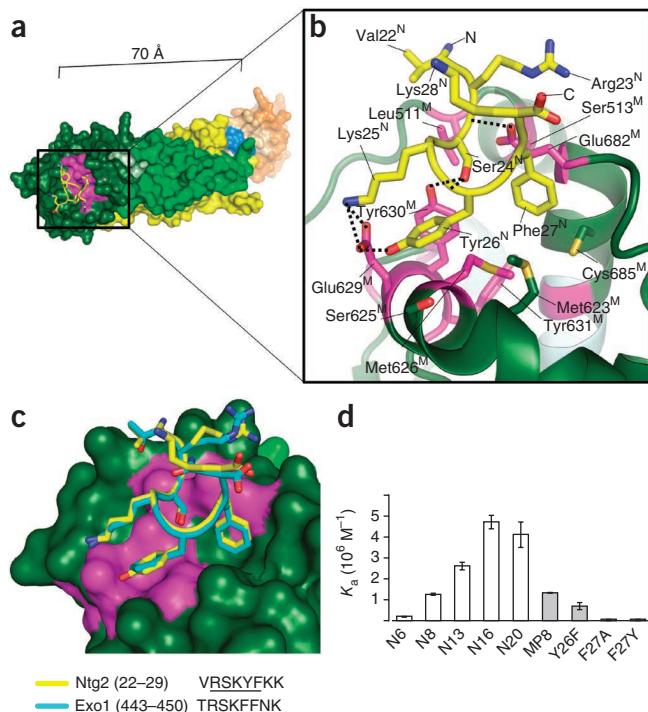
Figure 3 MutL α endonuclease site. (a) The Pms1(CTD) endonuclease site. The Pms1(CTD) endonuclease site (yellow) is constituted by five residues, His703^P and Glu707^P from helix α A, Cys817^P from helix α D and Cys848^P and His850^P from the loop α E– β 8, that stabilize two zinc atoms (gray spheres). The Mlh1(CTD) (green) interacts with the metal-binding site through the thiol group of its last residue Cys769^M. Both zinc ions have tetrahedral coordination with MutL α atoms. The last 11 residues of Mlh1 have a central role in positioning the Mlh1 Cys769 residue in the Pms1 endonuclease site. (b) Electrostatic potential surface of MutL α (CTD). Molecular graphic views of the MutL α (CTD) complex with the surface electrostatic potentials generated with PBEQ-Solver (+2 kcal (mol \times e)⁻¹ in blue to -2 kcal (mol \times e)⁻¹ in red)⁴⁸. The top view corresponds to an orientation close to that in a and the bottom view to a 90° rotation along the horizontal axis.

Figure 4 Structures of ternary complexes between MutL α (CTD) and MIP box-containing fragments. (a) Overall view of the position of an Ntg2 fragment containing the MIP-box motif in complex with the MutL α (CTD) heterodimer. The distance between the Pms1 endonuclease site (blue) and the Mlh1 S2 site (magenta) is about 70 Å. (b) Detailed view of the Mlh1 S2 site with an 8-aa Ntg2 fragment. Magenta residues correspond to the Mlh1(CTD) residues identified by Y2H as important for the interaction with Ntg2 and Exo1 (in ref. 36 and this study). Ntg2 residue Ser24^N makes an intramolecular hydrogen bond that stabilizes the turn structure of the fragment and a hydrogen bond with Mlh1 residue Tyr630^M. The two aromatic residues Tyr26^N and Phe27^N are buried on the Mlh1 S2 site. (c) Superimposition of the MutL α (CTD)-Ntg2-fragment complex with the MutL α (CTD)-Exo1-fragment complex. Mlh1 is shown in surface representation to highlight the cavity occupied by the second aromatic residue of the MIP motif. The two fragments superimpose well, and Exo1's first aromatic residue, Phe447^E, occupies the same position as Ntg2's first aromatic, Tyr26^N. The sequences of the fragments used in the crystal structures are reported, and the MIP-box motifs are underlined. (d) Association constants (K_a), measured by isothermal titration calorimetry of peptides containing a MIP box with Mlh1(CTD) or with MutL α (CTD). The values reported correspond to Ntg2 fragments with lengths varying from 6 to 20 aa (N6, N8, N13, N16 and N20) with Mlh1(CTD) alone, to an 8-aa Ntg2 fragment with MutL α (CTD) heterodimer (MP8), or to 13-aa Ntg2 fragments with single mutations (Y26F, F27A and F27Y) with Mlh1(CTD) alone. Thermodynamic values are presented in Table 2. Measurements were made in duplicate, except for N8 and N20, which were made in triplicate. The error bars represent the range of values for duplicates and s.d. for triplicates.

Pms1 metal ions, confirming that the Pms1 endonuclease site contains two Zn²⁺ ions in the crystal (Supplementary Fig. 3c).

The Glu707^P of the DQHA(X)₂E(X)₄E conserved motif of Pms1 interacts with both zinc atoms (Fig. 3a). The first zinc atom, Zn²⁺_A, is stabilized by His703^P and Cys817^P, whereas the second zinc atom, Zn²⁺_B, is stabilized by the residues His850^P and Cys848^P. Consistently, a MutL α variant composed of Mlh1 and Pms1(E707K) mutant has no detectable endonucleolytic activity *in vitro* and no functional MMR activity *in vivo*¹¹. The first conserved aspartate, Asp701^P from the DQHA(X)₂E(X)₄E motif, is not directly involved in metal binding but makes an N-terminal capping interaction³⁹ in helix α A, which suggests a role of this residue in the stability of the Pms1 endonuclease site. This observation is in agreement with the loss of endonuclease and MMR activities observed with the mutation on the human PMS2(D699N), at the position equivalent to yeast Asp701 (ref. 10). A striking feature of the MutL α endonuclease site, revealed by our structural analyses, is the participation of the last residue of Mlh1, Cys769^M, in the Pms1 metal-binding site and in particular of the thiol group of Cys769^M, which is well positioned to directly interact with the two zinc atoms (Fig. 3a). In bacterial homodimers, the second subunit does not participate directly with the metal-binding site as in eukaryotic MutL α . The structure of *B. subtilis* MutL(CTD) shows that two water molecules occupy the position of the thiol of Cys769^M in the bacterial protein (Supplementary Fig. 3d). A structure of the nonfunctional human MLH1(CTD) homodimer has been deposited in the PDB (3RBN). In this structure the FERC motif is not observed (Supplementary Fig. 3e). The analysis of the surface electrostatic potential around the Pms1 endonuclease site is also reported (Fig. 3b). We observed two patches of basic residues beside the endonuclease site, which could interact with DNA substrate by forming salt bridges with the phosphates of the DNA backbone (Fig. 3b and Supplementary Fig. 1c).

To further investigate the role of the Mlh1 C-terminal conserved sequence, we analyzed, by Y2H and by mutation rates in *Mlh1p* Δ cells, a version of Mlh1 carrying a C-terminal deletion, Δ C1, that is, lacking the C-terminal cysteine Cys769. This Mlh1 variant proved to



bind efficiently to Pms1, Mlh3 and Ntg2 in the Y2H assay but failed to complement the mutator phenotype of Δ Mlh1p yeast, thus confirming the functional role of the Mlh1 C terminus (Supplementary Table 1b,c). This data are in agreement with the mutator phenotype reported for a Mlh1 Δ C10 variant²¹. These results are also in agreement with the absolute conservation of the Mlh1 C terminus (FERC) sequence (Fig. 2d and ref. 36).

Structures of MutL α (CTD) in complex with two MIP-box motifs

We previously identified, in addition to the heterodimerization S1 site, an S2 site in yeast Mlh1(CTD) that is involved in the recruitment of the exonuclease Exo1, the DNA N-glycosylase and lyase Ntg2 and the helicase Sgs1 (ref. 36). These proteins share a MIP-box motif with the consensus sequence, (R/K)SK(Y/F)F^{33,36}, that is required for this interaction. Here, we determined the crystal structures of the MutL α (CTD) complexed with two fragments containing the Exo1 and the Ntg2 MIP-box motifs at, respectively, 2.7-Å and 3.0-Å resolution (Table 1 and Supplementary Fig. 4a,b). In the ternary complexes, the Ntg2 and Exo1 MIP-box fragments adopt an ST-turn conformation⁴⁰ with the side chain of the serine of the MIP box forming an intramolecular hydrogen bond with the main chain NH group of the residue in position (i+2) in the peptide. A superimposition of both ternary complexes shows that the binding modes of Ntg2 and Exo1 fragments are similar (r.m.s. deviation over the residues of the interface (r.m.s. deviation-i) of 0.53 Å over 155 atoms; Fig. 4a–c). Mlh1 exhibits no conformational change of the S2 site upon Ntg2 or Exo1 peptide binding. Analysis of the interface between Ntg2 or Exo1 fragments and Mlh1 highlights the role of ten residues in Mlh1 that we previously pinpointed as important for the interaction³⁶. In particular, Mlh1 amino acid Glu682 was previously identified as critical for the interaction and was used as an Mlh1 separation-of-function mutant for *in vivo* studies^{36,37}. The side chain of Glu682^M makes a hydrogen bond with the main chain NH of the MIP-box serine. To complete the mutational mapping over the S2 site, four Mlh1 mutants (L511A, M623A, S625A and E680A) were analyzed by Y2H. We found that

Table 2 Calorimetry analyses of the interactions at the Mlh1 S2 site

Protein	Peptide	Sequence	K_a (10^6 M $^{-1}$)	K_d (μ M)	ΔH° (kcal mol $^{-1}$)	$-T\Delta S^\circ$ (kcal mol $^{-1}$)	ΔG° (kcal mol $^{-1}$)
Mlh1(CTD)	pNtg2 (6-mer)	20-RSKYFK	0.19 ± 0.02	5.1 ± 0.46	-13.6 ± 0.4	+6.8 ± 0.45	-6.8 ± 0.1
	pNtg2 (8-mer)*	20-VRSKYFKK	1.25 ± 0.03	0.80 ± 0.02	-14.1 ± 1.7	+6.3 ± 1.7	-7.8 ± 0.1
	pNtg2 (13-mer)	20-VEVRSKYFKKKNER	2.61 ± 0.19	0.38 ± 0.03	-18.9 ± 0.8	+10.7 ± 0.8	-8.2 ± 0.1
	pNtg2 (16-mer)	18-EEVEVRSKYFKKKNERT	4.72 ± 0.32	0.21 ± 0.02	-21.5 ± 1.5	+13.0 ± 1.5	-8.5 ± 0.1
	pNtg2 (20-mer)*	16-DIEEVEVRSKYFKKKNERTVE	4.1 ± 0.61	0.24 ± 0.03	-18.2 ± 1.6	+9.7 ± 1.5	-8.5 ± 0.1
Mlh1(CTD)-Pms1(CTD)	pNtg2 (8-mer)	20-VRSKYFKK	1.33 ± 0.02	0.75 ± 0.02	-29.2 ± 0.8	+21.4 ± 0.8	-7.8 ± 0.1
Mlh1(CTD)	pExo1 (8-mer)	443-TRSKFFNK	0.34 ± 0.01	2.94 ± 0.03	-15.2 ± 1.7	+8.1 ± 1.7	-7.1 ± 0.1
Mlh1(CTD)	pSgs1 (8-mer)	1381-TKSKEFGA	0.44 ± 0.13	2.3 ± 0.9	-25.2 ± 4.3	+18.0 ± 4.2	-7.2 ± 0.2
Mlh1(CTD)	pNtg2 (13-mer-Y26F)	20-VEVRSKYFKKKNER	0.69 ± 0.17	1.45 ± 0.46	-24.3 ± 0.6	+16.9 ± 0.5	-7.4 ± 0.2
	pNtg2 (13-mer-F27A)	20-VEVRSKYAKKKNER	NI	NI	NI	NI	NI
	pNtg2 (13-mer-F27Y)	20-VEVRSKYKKNER	NI	NI	NI	NI	NI

The numbers before the peptide sequences correspond to the position of the first amino acid of the peptide in the protein sequence. The positions substituted for other amino acids are in bold. Experiments were performed in duplicate or in triplicate (indicated by an asterisk), with data shown as means ± range or ± s.d., respectively. NI, no interaction (that is, it indicates weak and constant signal under the condition used).

position Leu511 is an additional contributor to the interaction in the S2 site (**Supplementary Table 2**).

Our previous studies showed that the key positions within the MIP-box motif (R/K)SK(Y/F)F were the serine and the two aromatic residues³⁶. The serine, Ser24 in Ntg2 and Ser445 in Exo1, contributes to structure the peptide in an ST-turn conformation, as previously mentioned, and it also makes a hydrogen bond with the hydroxyl moiety of Mlh1 Tyr630. The first aromatic, Tyr26 in Ntg2 or Phe447 in Exo1, is in contact with Mlh1 residues Met626 and Tyr630. The hydroxyl group of Tyr26 in Ntg2 makes a hydrogen bond with the Mlh1 Glu629 side chain. We observed by calorimetry a 26% residual association constant, K_a , of an Ntg2 fragment with the mutation Y26F compared to the wild-type Ntg2 fragment, in agreement with the hydrogen bond mediated by Ntg2's Tyr26 (**Table 2**). The second aromatic residue in the MIP-box motif is a phenylalanine, Phe27 for Ntg2 and Phe449 for Exo1, which is fully buried in a cavity surrounded by residues Met626, Tyr630, Cys685 and Leu686 of Mlh1. We analyzed by calorimetry the interaction between Mlh1 and Ntg2 fragments with mutations F27A or F27Y and observed that these mutations disrupt the interaction (**Fig. 4d**).

We also analyzed by calorimetry the interaction between the Mlh1 subunit and Ntg2 fragments of different lengths (**Table 2**). We measured, in comparison to the MIP box (8 aa) used in the cocrystal, a 17% residual K_a with a 6-aa Ntg2 fragment reduced to the minimal MIP-box motif and a four-fold higher K_a with a 16-aa or a 20-aa Ntg2 fragment, which suggests some participation of neighboring residues. The MutLo(CTD) heterodimer bound an 8-aa Ntg2 fragment

with a similar affinity to that of the Mlh1(CTD) subunit alone (**Supplementary Fig. 4c,d**). Our results give a rationale for the key role of the serine and the two aromatic residues within the MIP-box motif. They also support the possibility of formation of ternary complexes, such as Mlh1-Pms1-Exo1, Mlh1-Pms1-Ntg2 or Mlh1-Pms1-Sgs1, endowed with specific biological functions and confirm the competitive interaction of protein with the MIP-box motif on a unique Mlh1 S2 site.

MLH1 missense variants associated with HNPCC

Each year, approximately 30,000 patients manifest a HNPCC syndrome, the most common being hereditary colorectal cancer-predisposing syndrome⁵. MLH1 and MSH2 mutations underlie most of these HNPCC cases, and about 400 missense variants have been found on the genes encoding MLH1 or MSH2, including 58 missense variants for the sole MLH1(CTD)^{5,29}. Using the MutLo(CTD) structure reported in this study, we modeled the human MutLo(CTD) with Modeller⁴¹ (Online Methods). This approach was straightforward, thanks to the high sequence identity between yeast and human

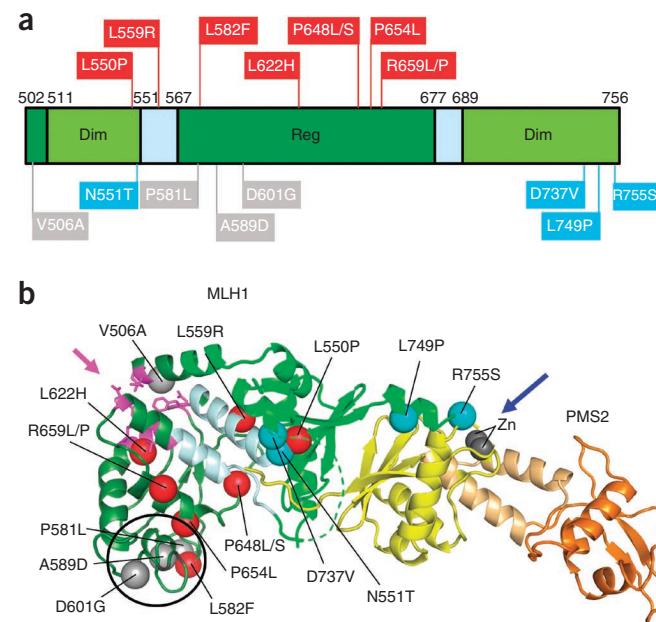


Figure 5 Mapping of deleterious mutations associated with HNPCC cancers. (a) Schematic representation of the positions of the deleterious mutations on MLH1(CTD) associated with HNPCC cancers. Slashes indicate that both mutations have been identified in patients with HNPCC. Dim and Reg indicate dimerization and regulatory domains, respectively. (b) Mapping of the deleterious mutations on a model of the human MLH1(CTD)-PMS2(CTD) heterodimer. The positions of the deleterious mutations are represented with spheres on the $\text{Ca}\alpha$. The variants are colored according to the potential impact of the missense mutation: red, mutations located in MLH1 hydrophobic core that probably induce a destabilization of MLH1; cyan, mutations close to the PMS2 heterodimerization interface that probably destabilize heterodimer formation; gray, mutations located at the surface of the MLH1(CTD). Three mutations (circled) clustered in the same region of MLH1 and may correspond to an additional MLH1 interaction site. The blue and magenta arrows indicate respectively the position of the MLH1-PMS2 endonuclease site and of the MLH1 S2 site. A superimposition of the yeast Mlh1(CTD) and human MLH1(CTD) structures is presented in **Supplementary Figure 6e**.

sequences of the MutLo(CTD) (39% sequence identity and 57% similarity). All the amino acids involved in the interaction sites and in the endonuclease site, described above, are conserved between the yeast and human MutLo(CTD) heterodimers. We used this model to locate the positions of 58 MLH1 mutations and in particular 15 classified as deleterious (Fig. 5a)²⁹. We classified the deleterious variants into three groups: (i) Group 1 variants (in red in Fig. 5b) are located in the Mlh1(CTD) hydrophobic core and are very likely to affect Mlh1 stability; (ii) Group 2 variants (in cyan in Fig. 5b) are located close to the heterodimerization S1 site; (iii) Group 3 variants (in gray in Fig. 5b) correspond to MLH1 residues exposed on the surface of the protein. Three variants of Group 3 (P581L, D601G and A589D) cluster in the same region of MLH1. Several studies have reported direct protein-protein interactions between human MLH1 and proteins involved in various cellular pathways. Direct interactions have thus been described with PCNA⁴², with MED1 (MBD4)⁴³, with FAN1, with angiogenin and with BRIP1 (BACH1)⁴⁴. We propose that the MLH1 exposed patch defined by Group 3 deleterious mutants could be involved in one of these additional interactions.

DISCUSSION

Roles of Mlh1 and Pms1 C termini in heterodimer formation

The structures presented here reveal that, contrary to bacterial MutL(CTD) homodimers, eukaryotic MutLo heterodimers are stabilized by three distinct patches. Patch I, which constitutes the core of the MutLo heterodimerization interface, probably also constitutes the core of the MutLoβ (Mlh1–Mlh2) and MutLoγ (Mlh1–Mlh3) heterodimerization sites. Patch II was found to be crucial to the Mlh1–Pms1 complex because the deletion of the last ten residues of Pms1 abolishes the interaction as monitored by Y2H (Supplementary Table 1a). Similarly, the disruption of the salt bridge between Arg547 in Mlh1 and Asp870 in Pms1 caused by R547A or D870A mutants strongly affects the Mlh1–Pms1 interaction in Y2H assays. In a previous study, a genetic screen was performed to identify mutations that cause a strong mutator phenotype only in *exo1Δ* mutant strains⁴⁵. Out of the 19 mutations identified, 13 mutations were localized on the Mlh1 or Pms1 proteins. Among them, Pms1(D870N) and Mlh1(R547K) match precisely the positions involved in the critical salt bridge identified here. However, in both Mlh1(R547K) and Mlh1(R547A) mutants, MMR remains partially functional^{36,45}. Taken together, these results suggest that mutations at patch II destabilize MutLo with an impact that is synergetic with an absence of Exo1. Patch III, involving the C-terminal end of Mlh1, has a minor role in the heterodimerization, because deletion of the last ten residues of Mlh1 did not affect the Mlh1–Pms1 interaction^{21,36} (Supplementary Table 1b).

Role of Mlh1 C terminus in MutLo endonuclease site

The heterodimerization patch III overlaps the Mlh1 C-terminal homology (CTH) motif defined previously²¹ that is constituted by the last 13 aa of Mlh1's extreme C terminus. The CTH motif includes the LY(X)₂FERC-stop sequence, which is strictly conserved in eukaryotic Mlh1(CTD) sequences³⁶. The Mlh1-CTH motif is a hallmark of eukaryotes because bacterial MutL homodimers, even among endonuclease MutLs, do not present a conserved C-terminus motif. The crystal structure of the *S. cerevisiae* MutLo(CTD) reveals that the CTH motif packs against the Pms1 subunit and that the last Mlh1 residue, Cys769, participates in the Pms1 metal-binding site. The functional role of this CTH motif is confirmed by the lack of complementation of the *mlh1Δ* mutant by Mlh1 constructs with deletions of Cys769 (this study) or of the last ten residues²¹ despite the capacity of these constructs to interact with Pms1. Our results are in agreement

with a strong mutator phenotype reported for Mlh1 mutants F766A and C769S in the CTH motif that presents no effect on the Mlh1–Pms1 interaction²¹.

The absence of endonuclease activity observed for the MutLo(CTD) suggests that the conformation observed may correspond to an apo conformation of the enzyme and that some conformational rearrangements should occur in the presence of the NTD and the linker regions of the MutLo heterodimer, as well as in the presence of DNA and Mn²⁺ or Mg²⁺ cations, as suggested by *in vitro* assays^{10,11}. The crystal structure of an eukaryotic full-length MutLo with DNA substrate will help to determine the active conformation of the Pms1 endonuclease site.

Recruitment of Exo1 and Ntg2 through their MIP-box motifs

The crystal structures of MutLo(CTD) with the Exo1 or the Ntg2 fragments give the first molecular view, to our knowledge, of the interaction between a MIP-box motif, (R/K)SK(Y/F)F and the Mlh1 S2 site³⁶. The structure of a nonfunctional human MLH1(CTD) homodimer was deposited in the RCSB Protein Data Bank (3RBN, not published). The protein was cloned with a 7-aa linker (sequence GRENLKYF). Notably, the YF aromatic dyad of this linker occupies the Mlh1 S2 site in a similar way to the Ntg2 and Exo1 (Y/F)F dyads (Supplementary Fig. 4e). The MIP-box motif of Exo1 is separated from its exonuclease domain (1–335) by a linker of ~110 aa (Supplementary Fig. 5). This suggests that the distance between the Exo1 exonuclease site and the Pms1 endonuclease site may vary from direct contact to up to 200 Å. The flexibility inherent to this binding mode is in good agreement with the multiple tasks supported by Mlh1–Exo1 complexes. In DNA MMR, the Exo1 interaction with MutLo on the S2 site is important for its *in vivo* role in MMR mutation avoidance^{36,46}. During meiosis, the Exo1 interaction with MutLo on the S2 site is also important for its *in vivo* role in resolving double Holliday junctions and promoting crossing over³⁷. These two scenarios involve different environments that may benefit from a large variability of the Exo1 position relative to Mlh1 heterodimers.

It has been reported that yeast full-length Mlh1 interacts with PCNA in Y2H analysis⁴², and a recent study⁴⁷ has proposed that this interaction would involve a degenerated PIP-box motif in the yeast Mlh1(CTD) (572–578 QxxLxxFA). The crystal structures of the yeast MutLo(CTD) and of the human MLH1(CTD) homodimer show that this motif is deeply buried in all these structures and is thus a weak candidate for an interaction with PCNA unless Mlh1(CTD) undergoes large structural rearrangements (Supplementary Fig. 6a–d).

Mapping of MutLo missense variants in HNPCC

The mapping of HNPCC missense mutations on MLH1(CTD) allows classification of the variants according to their potential impact on the protein in three classes: MLH1 destabilization, MLH1–PMS2 heterodimer destabilization and exposed mutations involved in potential protein-protein interaction sites. We compared our model of the human MutLo(CTD) with the crystal structure of the nonfunctional human MLH1(CTD) homodimer to refine our analysis of the impact of the deleterious mutants (Supplementary Fig. 6e). Some deleterious mutations are located at the heterodimerization interface in our crystal structure (positions in cyan in Fig. 5b). We propose that our structure of the *S. cerevisiae* MutLo(CTD) is closer to the human MutLo(CTD) heterodimer than to the nonfunctional human MLH1(CTD) homodimer. Human PMS2 indeed has a higher sequence homology with yeast Pms1 than with human MLH1 (44% and 10% sequence identities, respectively). The structures of *S. cerevisiae* MutLo(CTD) reported in this study thus provide new tools to explore functional and structural properties of MutL homologs in human cells in the context of HNPCC.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Coordinates and structure-factor files have been deposited in the Protein Data Bank, with accession codes **4E4W**, **4FMN** and **4FMO** for MutLo(CTD), MutLo(CTD)–(Ntg2 peptide) and MutLo(CTD)–(Exo1 peptide) complexes, respectively.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

This work was supported by Association pour la Recherche contre le Cancer grant numbers 4995 and 8006 to S.B. and J.-B.C. E.G. has received a Ministère de l'Education Nationale de la Recherche et de la Technologie (University Paris Sud) and Fondation pour la Recherche Médicale PhD fellowships. We thank the European Commission Infrastructure Protein Production Platform (P-CUBE) project (FP7/2007–2013; grant no. 227764) for access to the High Throughput Crystallization platform at the Grenoble Outstation of the European Molecular Biology Laboratory (EMBL). We thank synchrotrons SOLEIL (beamline Proxima 1; Gif, France) and European Synchrotron Radiation Facility (beamlines ID29, ID23-1, ID23-2; Grenoble, France) for access to beamlines. We thank R.D. Kolodner (Ludwig Institute for Cancer Research, La Jolla, California, USA) and P. Bertrand (Commissariat à l'Energie Atomique, Fontenay-aux-Roses, France) for yeast strains and X. Veautre for help in DNA purification. We thank A. Thompson, S. Zinn-Justin and R. Guerois for careful reading of the manuscript.

AUTHOR CONTRIBUTIONS

S.B. and J.-B.C. designed the research; E.G., C.D., P.L., C.T.-L., B.G., P.B., F.L., C.Q., M.-H.L.D., J.A.M., M.M., M.G., S.B. and J.-B.C. carried out the experiments; E.G., C.D., P.L., C.T.-L., S.B. and J.-B.C. analyzed the data; S.B. and J.-B.C. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Iyer, R.R., Pluciennik, A., Burdett, V. & Modrich, P.L. DNA mismatch repair: functions and mechanisms. *Chem. Rev.* **106**, 302–323 (2006).
- Kunkel, T.A. & Erie, D.A. DNA mismatch repair. *Annu. Rev. Biochem.* **74**, 681–710 (2005).
- Jiricny, J. The multifaceted mismatch-repair system. *Nat. Rev. Mol. Cell Biol.* **7**, 335–346 (2006).
- Lagerstedt Robinson, K. et al. Lynch syndrome (hereditary nonpolyposis colorectal cancer) diagnostics. *J. Natl. Cancer Inst.* **99**, 291–299 (2007).
- Lynch, H.T. et al. Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. *Clin. Genet.* **76**, 1–18 (2009).
- Gorman, J. et al. Dynamic basis for one-dimensional DNA scanning by the mismatch repair complex Msh2-Msh6. *Mol. Cell* **28**, 359–370 (2007).
- Jeong, C. et al. MutS switches between two fundamentally distinct clamps during mismatch repair. *Nat. Struct. Mol. Biol.* **18**, 379–385 (2011).
- Warren, J.J. et al. Structure of the human MutS α DNA lesion recognition complex. *Mol. Cell* **26**, 579–592 (2007).
- Acharya, S., Foster, P.L., Brooks, P. & Fishel, R. The coordinated functions of the *E. coli* MutS and MutL proteins in mismatch repair. *Mol. Cell* **12**, 233–246 (2003).
- Kadyrov, F.A., Dzantiev, L., Constantin, N. & Modrich, P. Endonucleolytic function of MutLo in human mismatch repair. *Cell* **126**, 297–308 (2006).
- Kadyrov, F.A. et al. *Saccharomyces cerevisiae* MutLo is a mismatch repair endonuclease. *J. Biol. Chem.* **282**, 37181–37190 (2007).
- Kosinski, J., Plotz, G., Guarne, A., Bujnicki, J.M. & Friedhoff, P. The PMS2 subunit of human MutLo contains a metal ion binding domain of the iron-dependent repressor protein family. *J. Mol. Biol.* **382**, 610–627 (2008).
- van Oers, J.M. et al. PMS2 endonuclease activity has distinct biological functions and is essential for genome maintenance. *Proc. Natl. Acad. Sci. USA* **107**, 13384–13389 (2010).
- Ban, C., Junop, M. & Yang, W. Transformation of MutL by ATP binding and hydrolysis: a switch in DNA mismatch repair. *Cell* **97**, 85–97 (1999).
- Wang, T.F., Kleckner, N. & Hunter, N. 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. USA* **96**, 13914–13919 (1999).
- Gorman, J., Plys, A.J., Visnapuu, M.L., Alani, E. & Greene, E.C. Visualizing one-dimensional diffusion of eukaryotic DNA repair factors along a chromatin lattice. *Nat. Struct. Mol. Biol.* **17**, 932–938 (2010).
- Sacho, E.J., Kadyrov, F.A., Modrich, P., Kunkel, T.A. & Erie, D.A. Direct visualization of asymmetric adenine nucleotide-induced conformational changes in MutLo. *Mol. Cell* **29**, 112–121 (2008).
- Arana, M.E. et al. Functional residues on the surface of the N-terminal domain of yeast Pms1. *DNA Repair (Amst.)* **9**, 448–457 (2010).
- Guarné, A., Junop, M.S. & Yang, W. Structure and function of the N-terminal 40 kDa fragment of human PMS2: a monomeric GHL ATPase. *EMBO J.* **20**, 5521–5531 (2001).
- Tran, P.T. & Liskay, R.M. Functional studies on the candidate ATPase domains of *Saccharomyces cerevisiae* MutLo. *Mol. Cell Biol.* **20**, 6390–6398 (2000).
- Pang, Q., Prolla, T.A. & Liskay, R.M. Functional domains of the *Saccharomyces cerevisiae* Mlh1p and Pms1p DNA mismatch repair proteins and their relevance to human hereditary nonpolyposis colorectal cancer-associated mutations. *Mol. Cell Biol.* **17**, 4465–4473 (1997).
- Deschênes, S.M. et al. The E705K mutation in hPMS2 exerts recessive, not dominant, effects on mismatch repair. *Cancer Lett.* **249**, 148–156 (2007).
- Erdeniz, N., Nguyen, M., Deschenes, S.M. & Liskay, R.M. Mutations affecting a putative MutLo endonuclease motif impact multiple mismatch repair functions. *DNA Repair (Amst.)* **6**, 1463–1470 (2007).
- Guarné, A. et al. Structure of the MutL C-terminal domain: a model of intact MutL and its roles in mismatch repair. *EMBO J.* **23**, 4134–4145 (2004).
- Namadurai, S. et al. The C-terminal domain of the MutL homolog from *Neisseria gonorrhoeae* forms an inverted homodimer. *PLoS ONE* **5**, e13726 (2010).
- Pillon, M.C. et al. Structure of the endonuclease domain of MutL: unlicensed to cut. *Mol. Cell* **39**, 145–151 (2010).
- Kosinski, J., Steindorf, I., Bujnicki, J.M., Giron-Monzon, L. & Friedhoff, P. Analysis of the quaternary structure of the MutL C-terminal domain. *J. Mol. Biol.* **351**, 895–909 (2005).
- Fukui, K. DNA mismatch repair in eukaryotes and bacteria. *J. Nucleic Acids* **2010**, 1–16 (2010).
- Chao, E.C. et al. Accurate classification of MLH1/MSH2 missense variants with multivariate analysis of protein polymorphisms-mismatch repair (MAPP-MMR). *Hum. Mutat.* **29**, 852–860 (2008).
- Kosinski, J., Hinrichsen, I., Bujnicki, J.M., Friedhoff, P. & Plotz, G. Identification of Lynch syndrome mutations in the MLH1–PMS2 interface that disturb dimerization and mismatch repair. *Hum. Mutat.* **31**, 975–982 (2010).
- Takahashi, M. et al. Functional analysis of human MLH1 variants using yeast and *in vitro* mismatch repair assays. *Cancer Res.* **67**, 4595–4604 (2007).
- Cutalo, J.M., Darden, T.A., Kunkel, T.A. & Tomer, K.B. Mapping the dimer interface in the C-terminal domains of the yeast MLH1–PMS1 heterodimer. *Biochemistry* **45**, 15458–15467 (2006).
- Gellon, L., Werner, M. & Boiteux, S. Ntg2p, a *Saccharomyces cerevisiae* DNA N-glycosylase/apurinic or apyrimidinic lyase involved in base excision repair of oxidative DNA damage, interacts with the DNA mismatch repair protein Mlh1p. Identification of a Mlh1p binding motif. *J. Biol. Chem.* **277**, 29963–29972 (2002).
- Pedraza, G. et al. Direct association of Bloom's syndrome gene product with the human mismatch repair protein MLH1. *Nucleic Acids Res.* **29**, 4378–4386 (2001).
- Tran, P.T., Simon, J.A. & Liskay, R.M. Interactions of Exo1p with components of MutLo in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **98**, 9760–9765 (2001).
- Dherin, C. et al. Characterization of a highly conserved binding site of Mlh1 required for exonuclease I-dependent mismatch repair. *Mol. Cell Biol.* **29**, 907–918 (2009).
- Zakharevich, K. et al. Temporally and biochemically distinct activities of Exo1 during meiosis: double-strand break resection and resolution of double Holliday junctions. *Mol. Cell* **40**, 1001–1015 (2010).
- Mauris, J. & Evans, T.C. Adenosine triphosphate stimulates *Aquifex aeolicus* MutL endonuclease activity. *PLoS ONE* **4**, e7175 (2009).
- Aurora, R., Creamer, T.P., Srinivasan, R. & Rose, G.D. Local interactions in protein folding: lessons from the α -helix. *J. Biol. Chem.* **272**, 1413–1416 (1997).
- Duddy, W.J., Nissink, J.W., Allen, F.H. & Milner-White, E.J. Mimicry by asx- and ST-turns of the four main types of β -turn in proteins. *Protein Sci.* **13**, 3051–3055 (2004).
- Fiser, A. & Sali, A. Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol.* **374**, 461–491 (2003).
- Umar, A. et al. Requirement for PCNA in DNA mismatch repair at a step preceding DNA resynthesis. *Cell* **87**, 65–73 (1996).
- Bellacosa, A. et al. MED1, a novel human methyl-CpG-binding endonuclease, interacts with DNA mismatch repair protein MLH1. *Proc. Natl. Acad. Sci. USA* **96**, 3969–3974 (1999).
- Cannava, E., Gerrits, B., Marra, G., Schlapbach, R. & Jiricny, J. Characterization of the interactome of the human MutL homologues MLH1, PMS1, and PMS2. *J. Biol. Chem.* **282**, 2976–2986 (2007).
- Amin, N.S., Nguyen, M.N., Oh, S. & Kolodner, R.D. *exo1*-Dependent mutator mutations: model system for studying functional interactions in mismatch repair. *Mol. Cell Biol.* **21**, 5142–5155 (2001).
- Tran, P.T. et al. A mutation in EXO1 defines separable roles in DNA mismatch repair and post-replication repair. *DNA Repair (Amst.)* **6**, 1572–1583 (2007).
- Lee, S.D. & Alani, E. Analysis of interactions between mismatch repair initiation factors and the replication processivity factor PCNA. *J. Mol. Biol.* **355**, 175–184 (2006).
- Jo, S., Vargyas, M., Vasko-Szedlar, J., Roux, B. & Im, W. PBEQ-Solver for online visualization of electrostatic potential of biomolecules. *Nucleic Acids Res.* **36**, W270–W275 (2008).

ONLINE METHODS

Purification, endonuclease and DNA-binding activities of the Mlh1(CTD)–Pms1(CTD) complex. The Mlh1(CTD) fragment, containing residues 485 to 769, was cloned into a vector derived from pETM30, with a p15 origin, a kanamycin-resistance sequence, a His₆-GST tag in the N terminus and a TEV protease site. The Pms1(CTD) fragment, containing residues 635 to 873, was cloned into a pDEST HisPKM 596 (AFMB) with a pBR322 origin, an ampicillin-resistance sequence, a His₆-MBP tag in the N terminus and a TEV protease site. BL21(DE3) star cells (Invitrogen) were transformed with MBP-Pms1 plasmid. Competent cells with the MBP-Pms1 plasmid were prepared⁴⁹ and transformed with the GST-Mlh1 plasmid, grown at 37 °C until an optical density of 1.2 at 600 nm and induced with 0.5 mM IPTG (isopropyl-β-D-thiogalactopyranoside) overnight at 20 °C. The cells were lysed by sonication in phosphate-buffered saline (PBS 1×), 1 mM phenylmethylsulfonyl fluoride (PMSF). Benzonase (100 U/ml lysate) (Novagen) was added, and MgCl₂ was added to a final concentration of 10 mM. After centrifugation, the supernatant fraction was loaded on dextrin-Sepharose resin (GE Healthcare) equilibrated with lysis buffer. MBP-tagged proteins were eluted with PBS buffer containing 10 mM maltose. The GST-Mlh1 protein coelutes with the MBP-Pms1 protein bound on the dextrin resin. The quantity of fusion proteins was estimated by the Bradford assay, and 1% (wt/wt) TEV protease was added directly to the eluate overnight at 4 °C. The cleaved products were applied to a HisTrap column (GE Healthcare), and the Mlh1(CTD)–Pms1(CTD) heterodimer was collected in the flow through. The complex was dialyzed against buffer A (20 mM NaH₂PO₄, pH 7.5, 50 mM NaCl, 10 mM β-mercaptoethanol) and applied to a RESOURCE Q anion-exchange column (GE Healthcare) equilibrated with buffer A. The Mlh1–Pms1 complex was eluted at 200 mM NaCl with a gradient of 0.1–1 M NaCl over 20 column volumes. The production of the selenomethionine Mlh1(CTD)–Pms1(CTD) complex is described in the **Supplementary Note**. The endonuclease and DNA-binding activity measurements of the MutLox(CTD) are reported in **Supplementary Note**.

Crystallization and structure determination. First, crystals of the Mlh1(CTD)–Pms1(CTD) were obtained on an HTX platform (EMBL, Grenoble). The crystals were then grown by vapor diffusion in sitting drops at 20 °C. Optimized conditions were realized by mixing 1.5 μl of protein sample at 4 mg/ml with 1.5 μl reservoir solution containing 10% wt/wt of MPEG 5000, 100 mM Na HEPES, pH 7 and 5 mM MgSO₄. The native crystals diffracted at 100 K on protein-crystallography beamline Proxima1 (Soleil, France) up to 2.3 Å in the best direction but with a strong anisotropy in one direction. The crystal structure was determined with a Se-Met–substituted complex by the SAD method, with data collection at 0.9800 Å and 0.9791 Å for native and Se-edge data, respectively. The Se-edge data have an anomalous signal characterized by a SigAno value of 1.068 and an anomalous correction of 29%. The anomalous correction is the correlation coefficient between signed anomalous differences, $|F_{hkl}| - |F_{h-k-l}|$ as calculated by the XDS program⁵⁰. The crystals of the Mlh1(CTD)–Pms1(CTD) complex belong to the space group C2, with a single copy of the heterodimer in the asymmetric unit. The positions of eleven selenium atoms, two zinc atoms and one sulfur atom were determined by SHELXD⁵¹. Experimental phases were calculated in Phaser⁵². The tracing of an initial model was realized in COOT⁵³. The model was subsequently completed by automatic building in Buccaneer⁵⁴ by using first the Se-Met–substituted data set then the native data set. Crystals used for zinc identification are Se-Met substituted. Data collection were made at the zinc peak ($\lambda = 1.28149$ Å) and at the pre-edge ($\lambda = 1.28614$ Å). Diffraction data were processed with XDS and were coscaled in XSCALE. Figures were prepared with PyMOL (PyMOL Molecular Graphics System, Version 0.99rc6, Schrödinger, LLC). The final model was refined with BUSTER⁵⁵. Twenty-four residues in the N terminus of Mlh1(CTD) (485–508) and 26 residues of Pms1 (635–648, 735–737, 758–764 and 796–797) have not been modeled, owing to absence of electron density. The crystal structures of the ternary complexes between the MutLox(CTD) and the two peptides containing the MIP-box motifs

(Ntg2 fragment 2229, VRSKYFKK, and Exo1 fragment 444–451, TRSKFFNK) were obtained by soaking the MutLox(CTD) crystals with 0.25-mM solutions of Ntg2 or of Exo1 fragments. Data were collected at Proxima 1 (SOLEIL) at 100 K and at wavelength of 0.9801 Å. The two ternary complexes were solved by molecular replacement with the MutLox native structure with Molrep program⁵⁶. The complexes with Ntg2- and Exo1-fragments were refined to 2.7 Å and 3.0 Å, respectively. The three structures presented here have no residues in outlier regions of the Ramachandran plots⁵⁷. The last residues of Exo1 and Ntg2 fragments, Lys450 and Lys29, are not observed in the electron density maps.

Yeast strains, plasmids, microbiological methods and yeast two-hybrid assays. *S. cerevisiae* strains and plasmids used in the present study are listed in **Supplementary Note**. Microbiological methods and yeast two-hybrid assays are similar to those already reported³⁶ and are detailed in **Supplementary Note**.

Isothermal titration calorimetry. The Mlh1(CTD) alone was purified for ITC analyses as described previously³⁶. The synthetic peptides used in this study were purchased from Genecust at 95% purity. The binding between Mlh1(CTD) and different peptides was measured by using a VP-ITC (GE-Healthcare). Prior to the measurements, all the solutions were degassed under vacuum for elimination of bubbles. The reaction cell was loaded with 15 μM Mlh1 protein, and the syringe contained 150 μM peptide solutions. Thirty injections of 10-μl volume were made at 210-s intervals. To correct for the heat effects of dilution, control experiments were performed at the same concentration of peptides with buffer in the cell. Thermodynamic parameters ΔH° (enthalpy change), n (stoichiometry) and K_a (association constant) were obtained by nonlinear least-squares fitting of the experimental data by using the single set of independent-binding-sites model of the Origin software provided with the instrument. The free energy of binding (ΔG°) and entropy (ΔS°) were obtained by using the classical thermodynamic formulae: $\Delta G^\circ = -RT \times \ln K_a$ and $\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$, where R is the gas constant and T is the absolute temperature in K. The titrations were carried out at 6 °C in 50 mM Na₂HPO₄, pH 7.7, 300 mM NaCl, 10 mM β-mercaptoethanol.

Modeling of the human MLH1(CTD)–PMS2(CTD) complex. The amino acid sequences of *S. cerevisiae* Mlh1 and Pms1 were used as a queries to search the nonredundant database (nr70) by using Psi-Blast⁵⁸. Twenty-five Mlh1 sequences and 42 Pms1 sequences from eukaryotic organisms including *H. sapiens* were retrieved and realigned by using MUSCLE⁵⁹. These alignments and the crystal structure of *S. cerevisiae* MutLox were used to build a model of the MLH1(CTD)–PMS2(CTD) complex from Asn502 to Cys756 for MLH1 and from Pro654 to Asn862 for PMS2 with the Modeller program⁴¹.

49. Inoue, H., Nojima, H. & Okayama, H. High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96**, 23–28 (1990).
50. Kabsch, W. Xds. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
51. Sheldrick, G.M. A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122 (2008).
52. McCoy, A.J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D Biol. Crystallogr.* **63**, 32–41 (2007).
53. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
54. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1002–1011 (2006).
55. Smart, O.S. *et al.* Exploiting structure similarity in refinement: automated NCS and target structure restraints in Buster. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 368–380 (2012).
56. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 22–25 (2010).
57. Laskowski, R.A., Mac Arthur, M.W., Moss, D.S. & Thornton, J.M. PROCHECK: a program to check the stereochemical quality of protein structure. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
58. Altschul, S.F. & Koonin, E.V. Iterated profile searches with PSI-BLAST—a tool for discovery in protein databases. *Trends Biochem. Sci.* **23**, 444–447 (1998).
59. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

Supplementary Information

Supplementary Figures :

Suppl Fig. 1 : Purification, endonuclease and DNA-binding activities

- (a) Purification of the MutL α (CTD) heterodimer.
- (b) The MutL α (CTD) from *S. cerevisiae* has no detectable ATP-Mn $^{2+}$ -dependent endonuclease activity.
- (c) DNA binding activity of the MutL α (CTD) from *S. cerevisiae*.

Suppl Fig. 2 : Multiple sequence alignments and superimpositions

- (a) Structure based sequence alignments of yeast and human Mlh1(CTD)s, Pms1(CTD)s with bacterial MutL(CTD)s.
- (b) Superimposition of the *S. cerevisiae* MutL α (CTD) with the *B. subtilis* MutL homodimer.
- (c-e) Superimpositions of yeast Mlh1(CTD), Pms1(CTD) and *B. subtilis* MutL subdomains.
- (f) Superimpositions of the bacterial and eukaryotic MutL(CTD)s.

Suppl Fig. 3 : The endonuclease site of MutL α (CTD)

- (a) The fluorescence spectra.
- (b) Statistics of the data collection at Zn peak.
- (c) Electron density of Fourier-difference anomalous maps.
- (d) The *B. subtilis* endonuclease site.
- (e) Superimposition of the *S. cerevisiae* Mlh1(CTD) with the human MLH1(CTD) homodimer.

Suppl Fig. 4 : The Mlh1 S2 site

- (a-b) Stereo view of the omit maps on (a) Ntg2 and (b) Exo1 peptides.
- (c-d) ITC thermograms.
- (e) Superimposition of human MLH1(CTD) homodimer with its linker on the yeast MutL α (CTD)-Ntg2 complex.

Suppl Fig. 5 : Model of the MutL α heterodimer in complex with Exo1

- (a) Schematic representations of the *S. cerevisiae* Exo1, Mlh1 and Pms1 sequences.
- (b) 3D model of the full length MutL α heterodimer in complex with Exo1.

Suppl Fig. 6 : Putative Mlh1 PIP-box motifs and superimposition of yeast Mlh1(CTD) with human MLH1(CTD)

- (a) Sequences of some canonic PIP-box motifs and of the putative Mlh1 PIP-box motifs proposed.
- (b) Mapping of the Mlh1-1 and Mlh1-2 putative PIP-box motifs in the yeast MutL α (CTD) structure.
- (c) The Mlh1-1 PIP-box motif.
- (d) The Mlh1-2 PIP-box motif.
- (e) Superimposition of the human MLH1(CTD) on the 3D model of the human MLH1(CTD) made from the crystal structure of the *S. cerevisiae* MutL α (CTD).

Supplementary Tables

Suppl Table 1 Two-hybrid and complementation assays using mutants of Mlh1 and Pms1 at the S1 site on Patch II and Patch III.

- (a-b) Y2H assays
- (c) Complementation assays

Suppl Table 2 : Two-hybrid assays to refine the mapping of the Mlh1 S2 site.

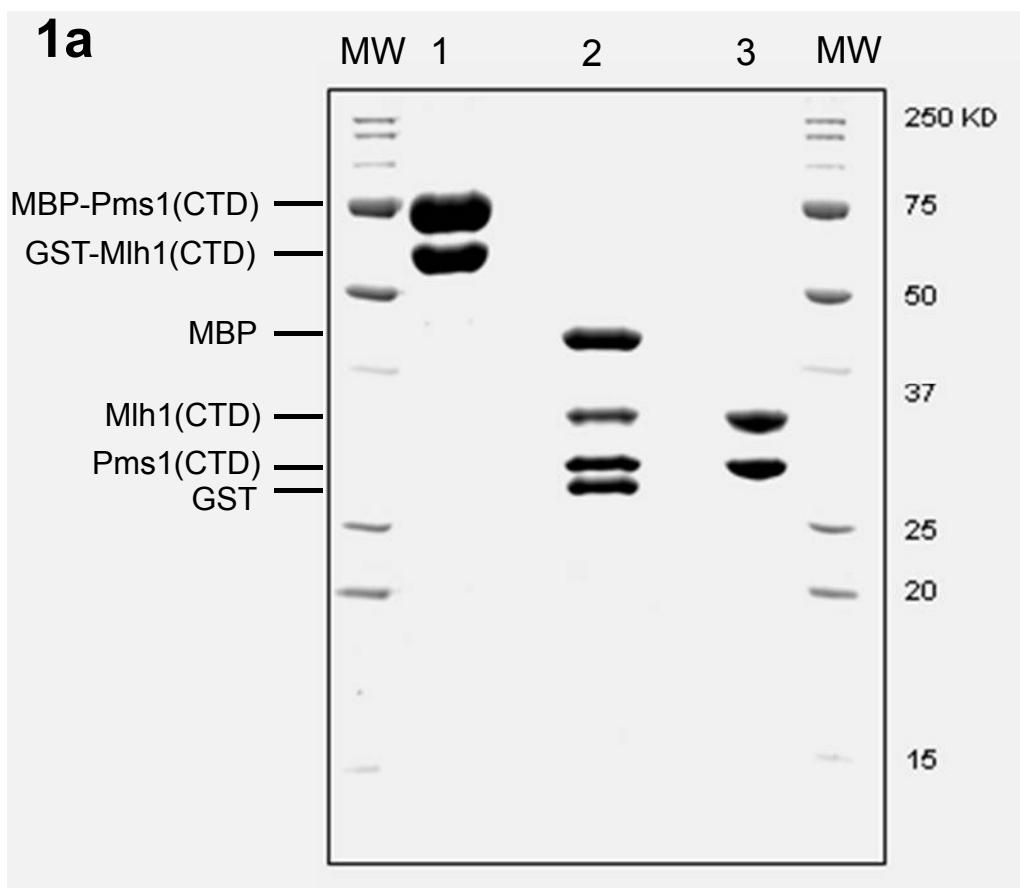
Supplementary Notes :

- (a) Experimental Procedures.
- (b) Additional information on the interactions involved in the three patches of the heterodimer interface.

References of the Supplementary Information

Supplementary Information - page 1 - Gueneau et al

Suppl Fig. 1 : Purification, endonuclease and DNA-binding activities

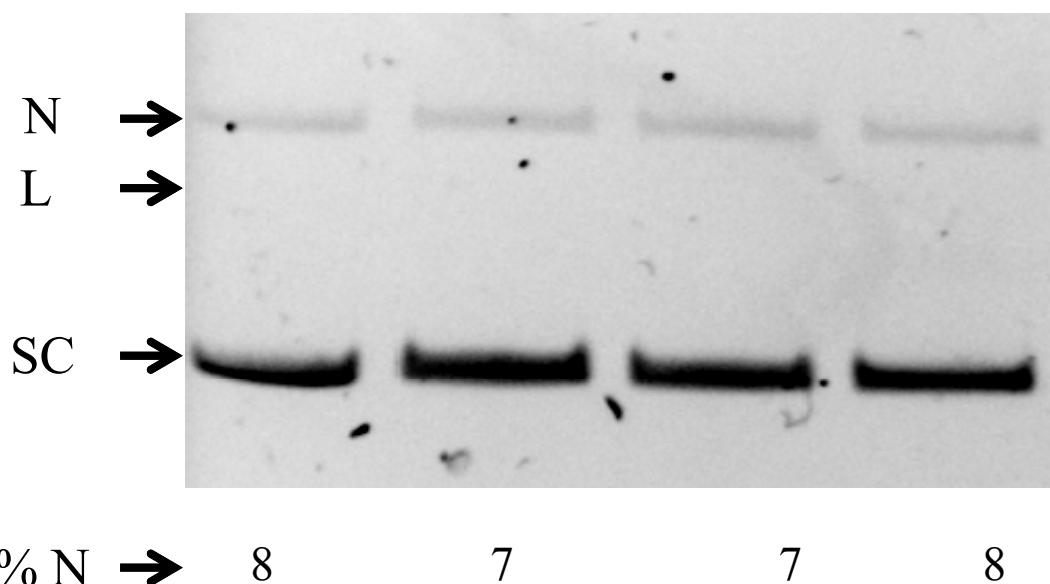


Supplementary Fig. 1a: Purification of the MutL α (CTD) heterodimer. (1) The fusion proteins, MBP-Pms1 (71.1 kDa) and GST-Mlh1 (62 kDa), are co-purified on an amylose resin. (2) The fusion proteins are cleaved by the TEV protease. (3) Mlh1(CTD) (33.2 kDa) and Pms1(CTD) (27.9 kDa) are co-purified on an Hi-Trap column (complex in the flow-through) and then on a Resource Q anion exchange column. (MW : molecular weights).

Suppl Fig. 1 : Purification, endonuclease and DNA-binding activities

1b

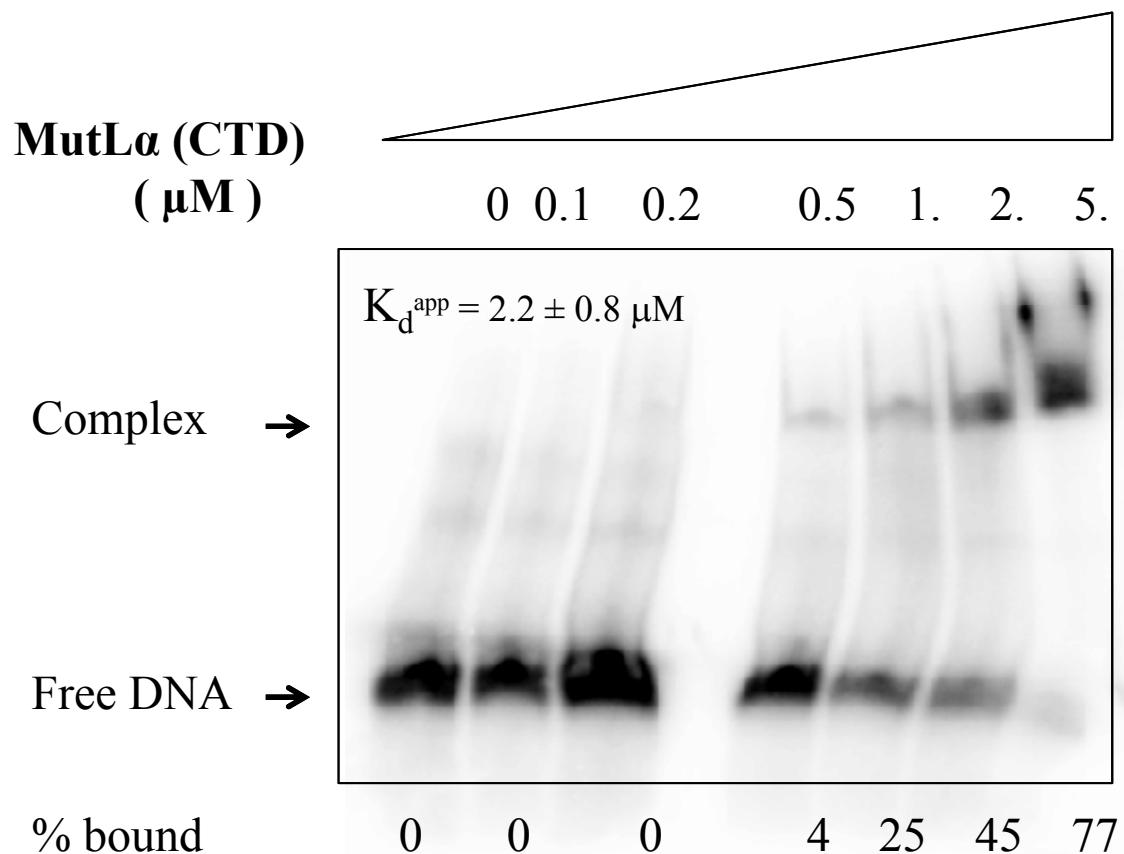
MutLα (CTD)	-	+	+	+
1 mM MgCl₂	-	+	-	-
1 mM MnSO₄	+	-	+	+
0.5 mM ATP	+	+	+	-



Supplementary Figure 1b: The MutL α (CTD) from *S. cerevisiae* has no detectable ATP-Mn²⁺ -dependent endonuclease activity. Endonuclease activity of the yeast MutL α (CTD) on supercoiled homoduplex plasmid pVil1 DNA (5303 bp) was determined as described under **Supplementary Notes paragraph (a5)**, except that ATP, MnSO₄ and MgCl₂ were varied as shown. The MutL α (CTD) concentration used was 1 μ M. Migration of supercoiled (SC), nicked (N), and linear (L) DNA is indicated. %N corresponds to percentage of nicked DNA

Suppl Fig. 1 : Purification, endonuclease and DNA-binding activities

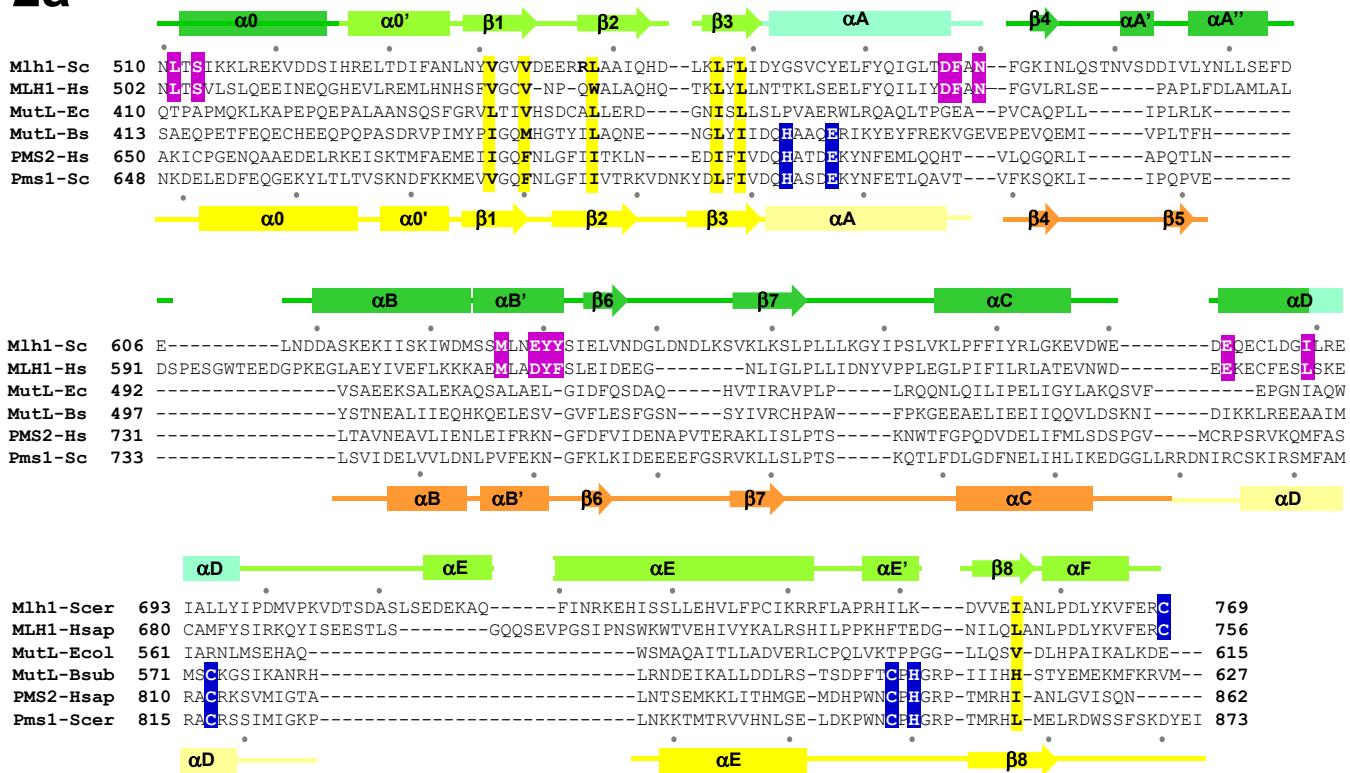
1c



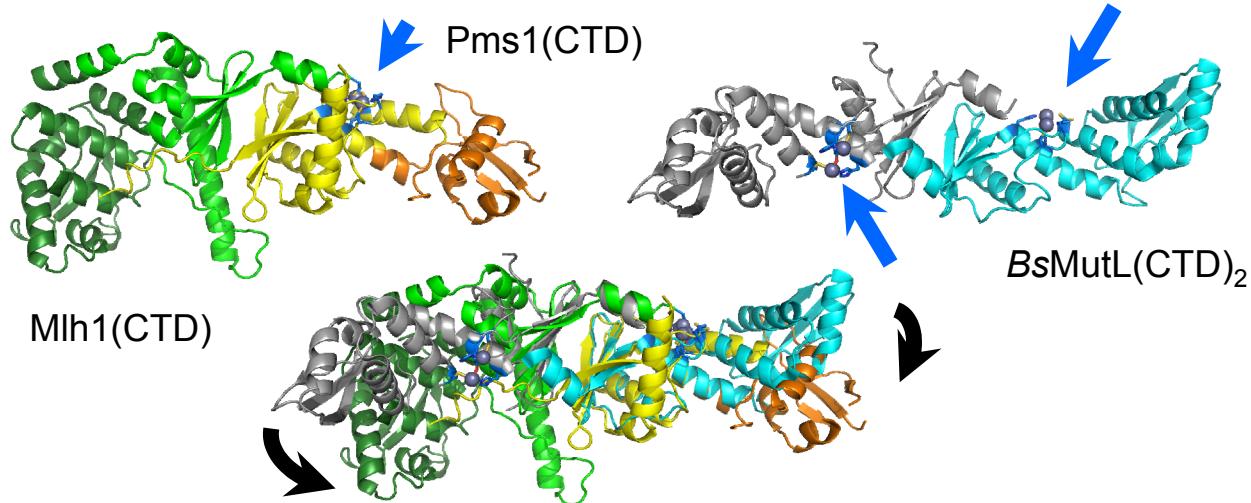
Supplementary Figure 1c : DNA binding activity of the MutLa(CTD) from *S. cerevisiae*.
The DNA binding affinity was determined using a 59-bp homoduplex DNA as substrate in the presence of MutLa(CTD) (0.1 to 5 μM). EMSA was performed as described in **Supplementary Notes paragraph (a6)**. EMSA shown is a representative experiment and the apparent K_d value reported is the average of four independent determinations.

Suppl Fig 2: Multiple sequence alignments and superimpositions

2a

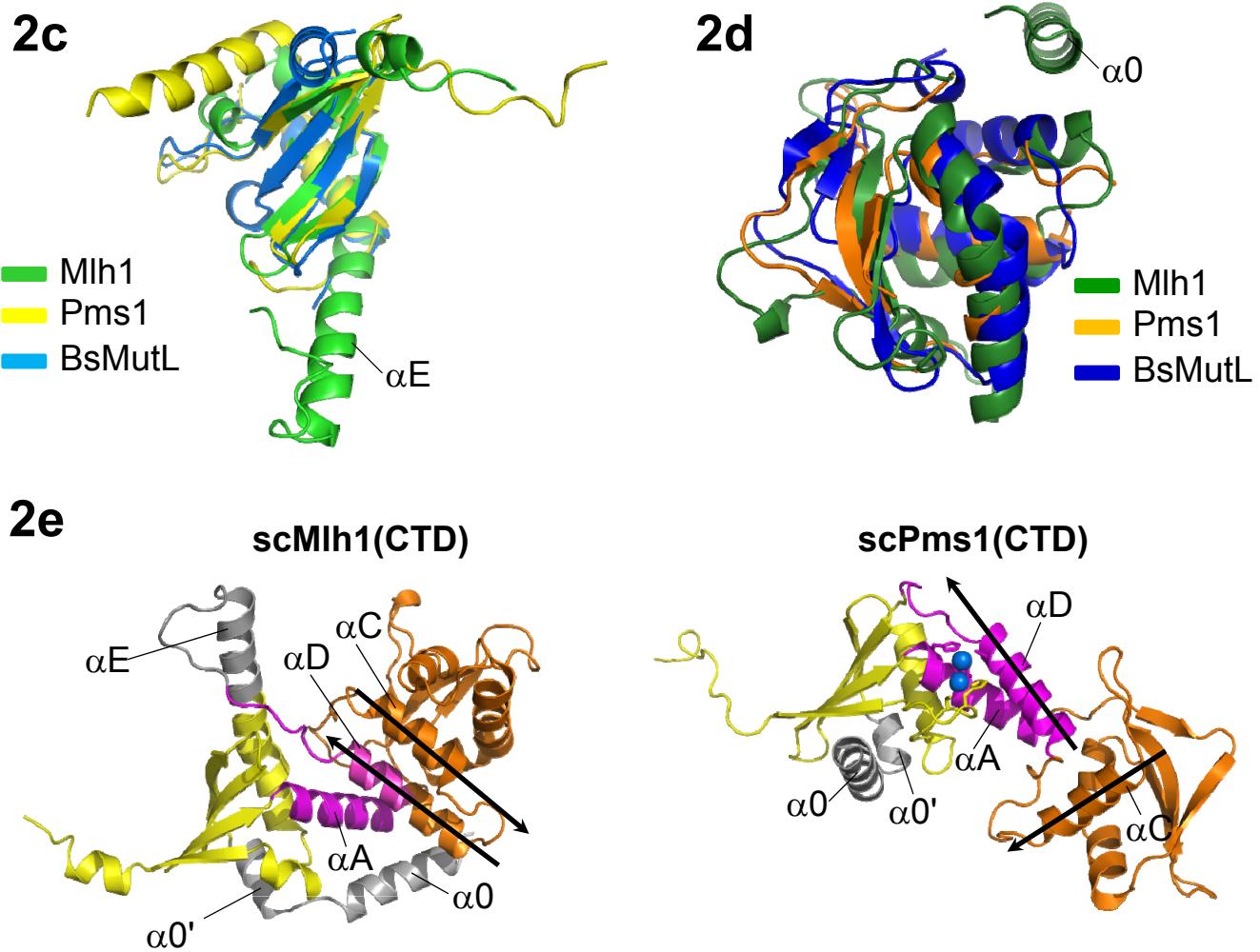


2b



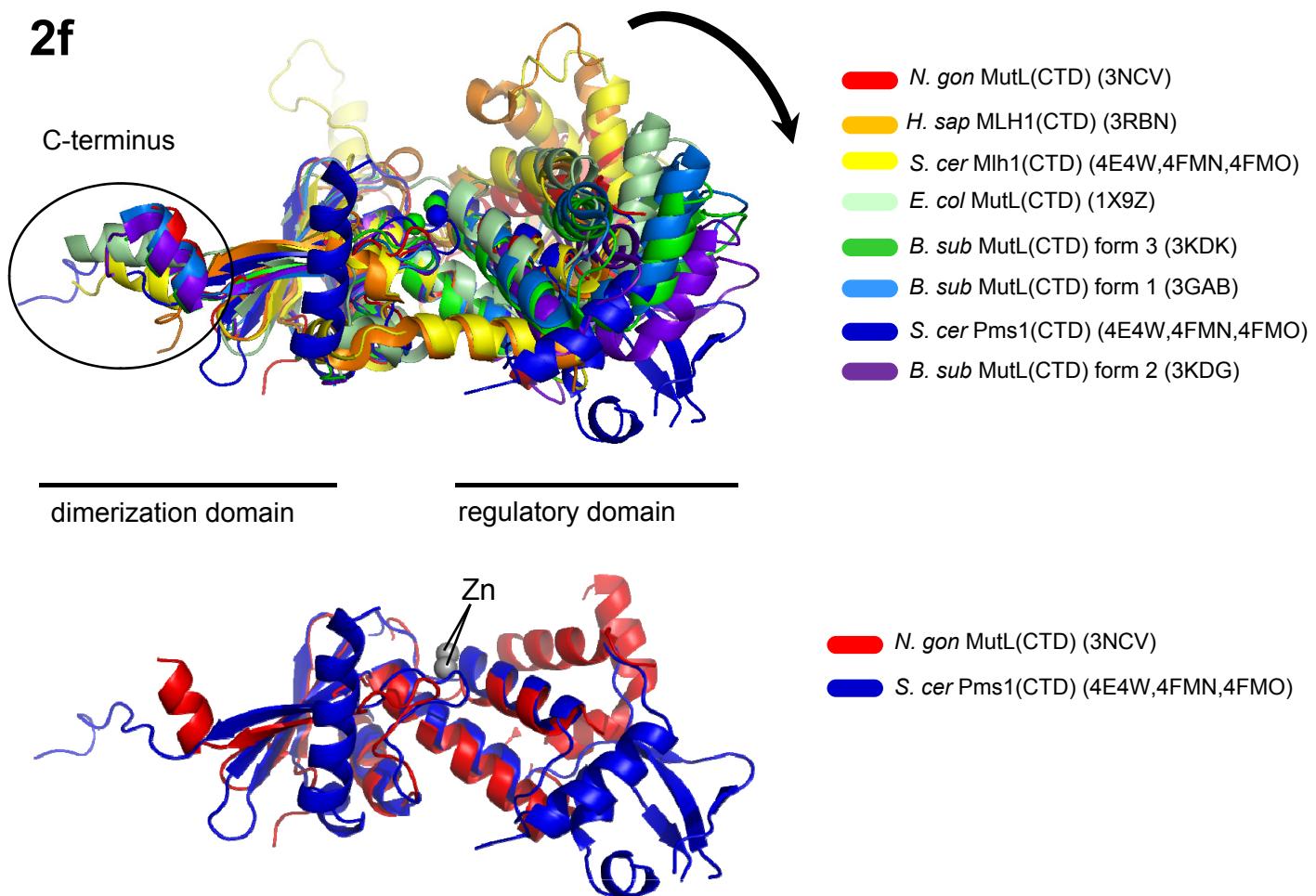
Supplementary Fig. 2 : (a) Structure based sequence alignments of yeast and human Mlh1(CTD)s, Pms1(CTD)s with bacterial MutL(CTD)s. Residues involved in the endonuclease site and in the Mlh1 S2 site are highlighted in blue and in magenta respectively. Hydrophobic Mlh1 and Pms1 residues involved in the heterodimerization S1 site and conserved in the bacterial MutL(CTD)s counterparts are highlighted in yellow. (To fit Pms1 β strand numbering, Mlh1 β strand following β4 is called β 6). **(b) Superimposition of the *S. cerevisiae* MutLa(CTD) with the *B. subtilis* MutL homodimer.** The superimposition is shown in the middle. The blue arrows indicate the Pms1 and *B. subtilis* MutL homodimers endonuclease sites. The *B. subtilis* MutL homodimer is superimposed on yeast MutLa(CTD) through the dimerization domains of both complex. Black arrows underline the differences of positions of the Mlh1 and Pms1 regulatory domains compare to *B. subtilis* MutL(CTD) regulatory domains.

Suppl Fig 2: Multiple sequence alignments and superimpositions



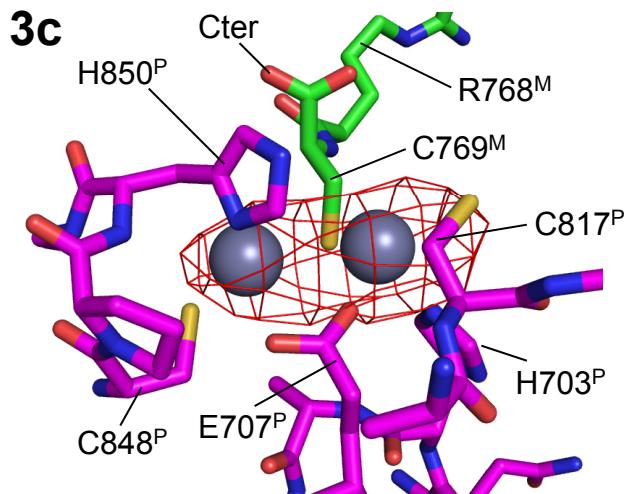
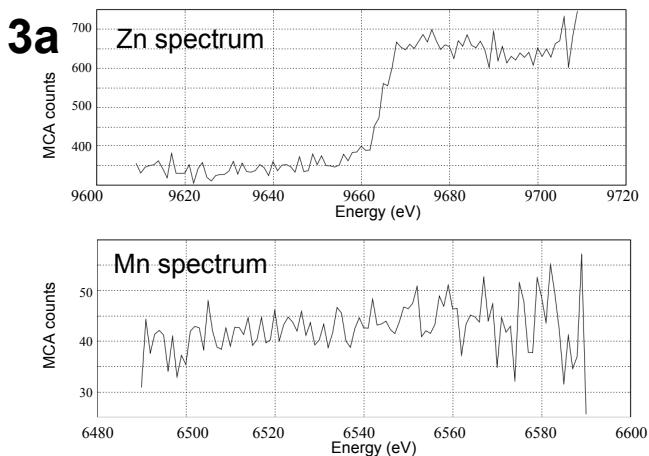
Supplementary Fig. 2: (c-e) Superimpositions of yeast Mlh1(CTD), Pms1(CTD) and *B. subtilis* MutL subdomains. (c) Superimposition of the dimerization domains. (d) Superimposition of the regulatory domains. (e) Relative positions of the dimerization and regulatory domains in yeast Mlh1(CTD) and Pms1(CTD). The C-terminal domain of *S. cerevisiae* Mlh1(CTD) (left) and Pms1(CTD) (right) have been superimposed on their dimerization domain. Dimerization, regulatory and linkers between both domains are colored in yellow, orange and magenta respectively. Zinc atoms observed in scPms1 are colored in blue. Additional secondary structure elements compare to bacterial proteins present in scMlh1 and scPms1 are coloured in grey. The direction of the helices α C and α D of the CTDs are indicated by an arrow to underline the disruption of the contacts between these two helices in the Pms1(CTD) and the consecutive reorientation of the regulatory domain in Pms1.

Suppl Fig 2: Multiple sequence alignments and superimpositions



Supplementary Fig. 2: (f) Superimposition of the bacterial and eukaryotic MutL(CTD)s. Five bacterial MutL(CTD)s including 3 crystal forms of *B. subtilis*, our crystal structure of the *S. cerevisiae* Mlh1(CTD) and Pms1(CTD) and the non-published human MLH1(CTD) were superimposed on their dimerization domains. These superimposition underlines the various positions of the regulatory domains from a « closed » conformation (relative to the endonuclease site) in *N. gonorrhoeae* MutL(CTD), human MLH1(CTD) or *S. cerevisiae* Mlh1(CTD) to an « open » conformation like in *S. cerevisiae* Pms1(CTD) or one crystal form of *B. subtilis* MutL(CTD) (3KDG). (Top) The closed to open conformation are coloured from red to purple. The superimpositions highlight the diversity of C-terminus conformation in all these CTDs (black circle). (Bottom) The extreme « closed » and « open » conformations of the 8 MutL(CTD)s structures are shown.

Suppl Fig 3 : The endonuclease site of the MutL α (CTD)



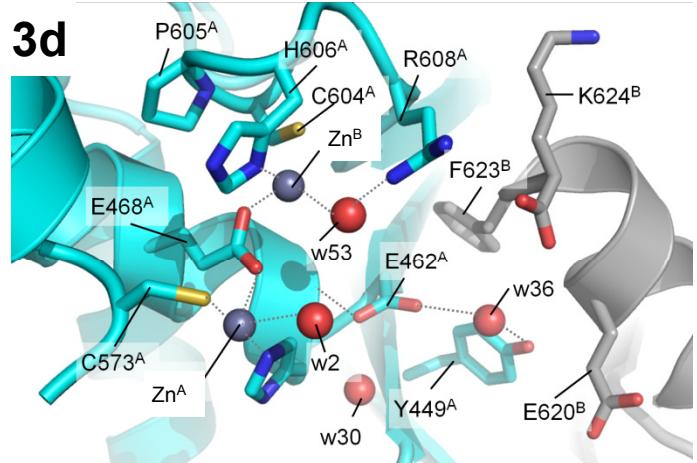
Supplementary Fig. 3: (a) The fluorescence spectra. The fluorescence spectra around the K absorption edge of Zn (up) and Mn (bottom) reveal the presence of Zn atoms in the crystal and the absence of Mn. (b) Statistics of the data collection at the Zn peak. Statistics on a MutL α (CTD) crystal at the Zn peak ($\lambda=1.2815\text{\AA}$) and at a pre-edge wavelength. (c) Electron density of Fourier-difference anomalous maps. Maps are contoured at 4σ shows the presence of two Zn atoms in the Pms1 metal binding site (d) The *B. subtilis* endonuclease site. View of *B. subtilis* active site with the same orientation than the Mlh1(CTD)-Pms1(CTD) complex in Fig 3a of the main article. Two waters molecules (w2 and w53) chelate the Zn in the bacterial MutL in place of Mlh1 C769 last residue. In *B. subtilis* MutL, the C-terminal end is structured until position K624 and the last three amino acids of the bacterial sequence (625-627, RVM) are not observed. (e) Superimposition of the *S. cerevisiae* Mlh1(CTD) with the human MLH1(CTD) homodimer. The C-terminus of hMLH1 (749-751) has a different conformation from the equivalent residues in yMlh1 (L762-K763) and the last five residues, VFERC are not observed in hMLH1(CTD) homodimer (3RBN).

3b

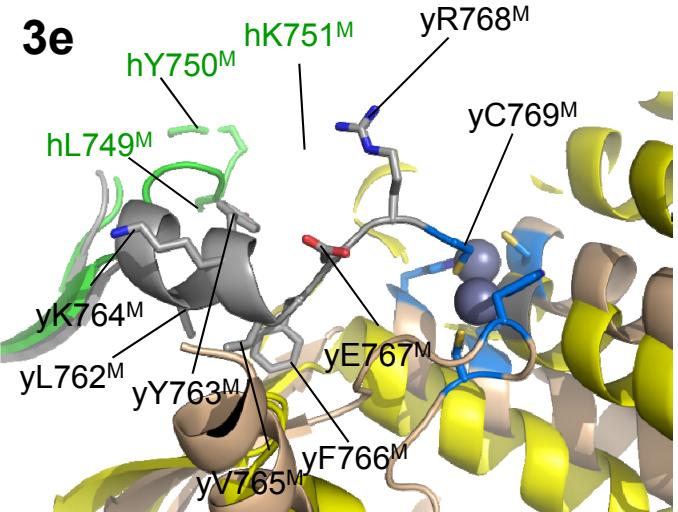
	Data collection	
	Zn(peak)	Zn(pre-edge)
Unit Cell	190.9 \AA , 66.2 \AA , 74.7 \AA 90.0°, 90.8°, 90.0°	
Wavelength (Å)	1.2815	1.2861
Resolution (Å) ^a	35.0-4.6(4.97-4.6)	35.0-4.5
Completeness (%) ^a	98.2 (98.9)	98.4 (98.9)
Redundancy ^a	2.0 (2.0)	2.0 (2.0)
Rmerge (%) ^a	18.7 (24.9)	30.2 (51.9)
I/s(I) ^a	6.1 (4.0)	4.06 (1.97)

^a Data in the highest-resolution shell are shown

3d

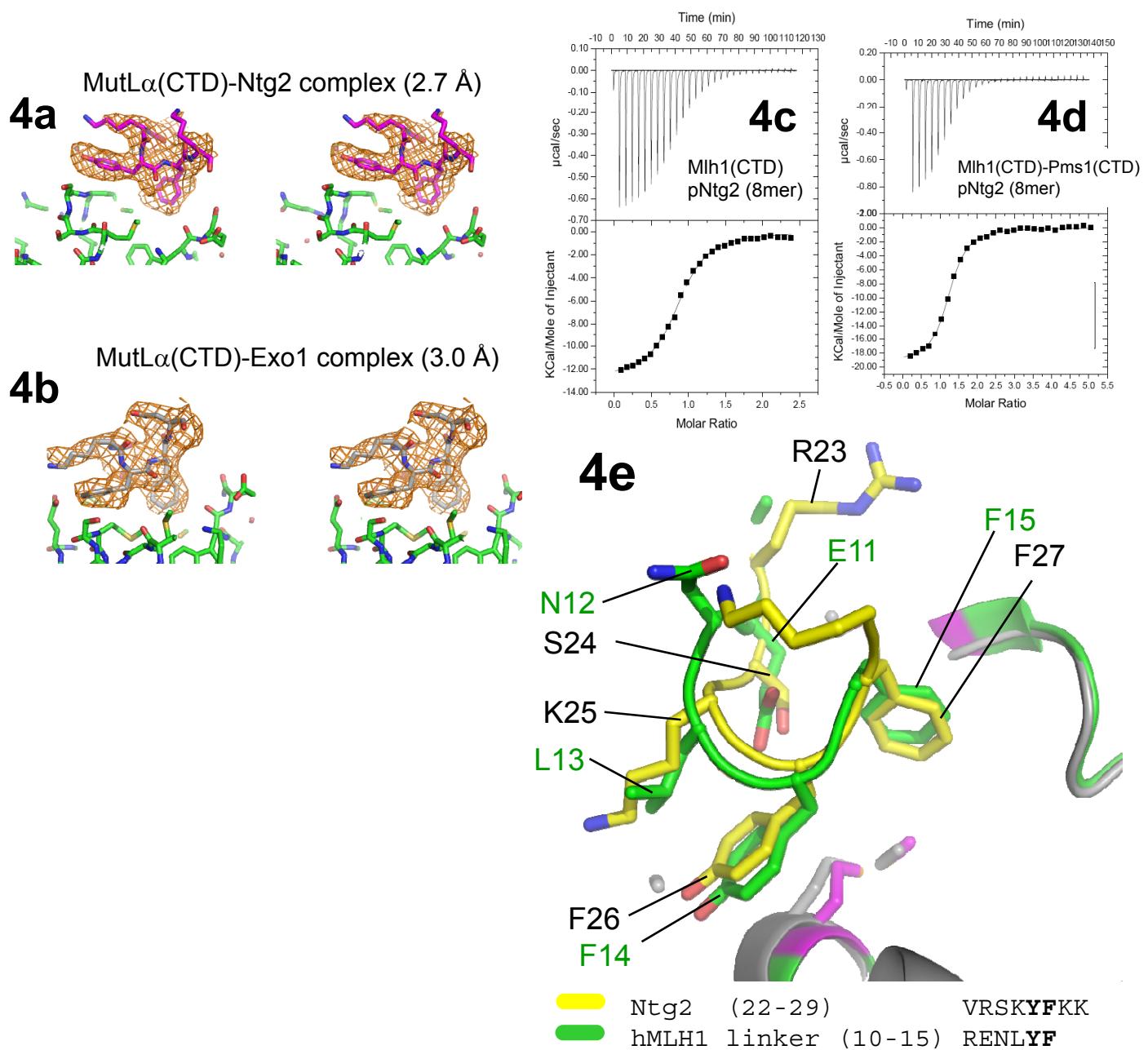


3e



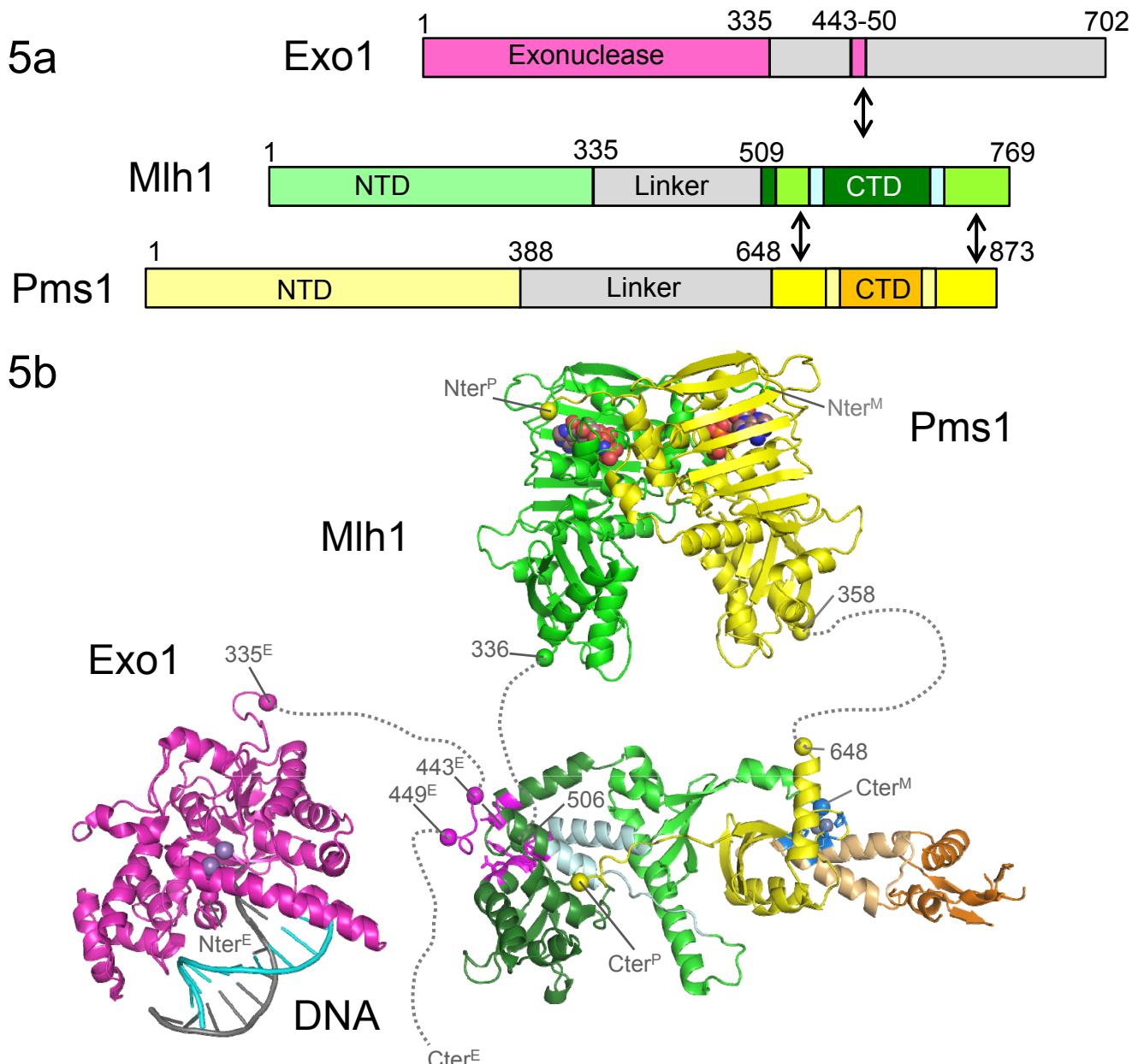
■ yMlh1 (762-769) LYKVFERC
■ hMLH1 (749-756) LYKVFERC

Suppl Fig 4: The Mlh1 S2 Site



Supplementary Fig. 4: (a-b) Stereo view of the omit maps on (a) Ntg2 and (b) Exo1 peptides. The electron density maps correspond to the Fo-Fc maps contoured at 2σ and generated before building the peptides in the ternary complexes MutL α (CTD)-Ntg2 and MutL α (CTD)-Exo1. The final model of the Ntg2, Exo1 and Mlh1 are represented in magenta, grey and green respectively. **(c-d) ITC thermograms.** Binding isotherms and integrated heats are shown in upper and lower panels. ITC thermograms with Mlh1(CTD) alone (c) or Mlh1(CTD)-Pms1(CTD) heterodimer (d). **(e) Superimposition of human MLH1(CTD) homodimer with its linker on the yeast MutL α (CTD)-Ntg2 complex.** A linker GRENL \mathbf{Y} FF (green) used during the cloning of the human MLH1(CTD) occupies the same Mlh1 S2 site as reported for the peptides of Ntg2 (yellow) and Exo1. yMlh1 is shown in grey and hMLH1 is shown in green with residues of S2 site in magenta

Suppl Fig 5 : Model of the MutL α heterodimer in complex with Exo1



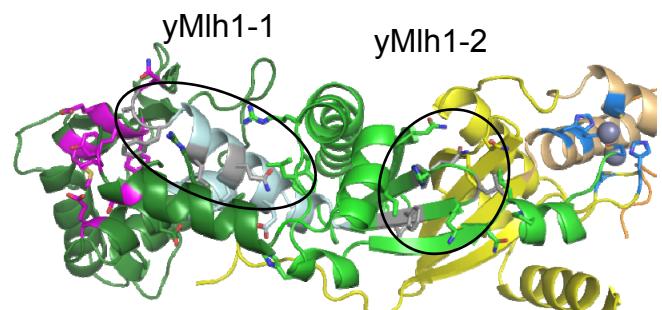
Supplementary Fig. 5 : (a) Schematic representations of the *S. cerevisiae* Exo1, Mlh1 and Pms1 sequences. The grey boxes correspond to regions predicted as non-structured **(b) 3D model of the full length MutL α heterodimer in complex with Exo1.** The MutL α (NTD) heterodimer built from *E. coli* MutL(NTD) homodimer complexed with ADPnPs (the nucleotides are represented in sphere mode) (RCSB 1B63) (Ban *et al.*, 1999) was manually positioned over the MutL α -CTD structure. The MutL α is represented in a semi-condensed conformation as observed by AFM (Sacho *et al.*, 2008). The Exo1-DNA complex (RCSB 3QEB) is manually positioned closed to the Mlh1 S2 site. The grey dashed lines represent the linkers between the Pms1(NTD) and Pms1(CTD) regions (291aa), between the Mlh1(NTD) and Mlh1(CTD) regions (171aa), between the Exo1 nuclease domain and its MIP-box (109aa). The endonuclease and S2 sites of MutL α are coloured in blue and magenta respectively. The N- and C-termini of Mlh1, Pms1 and Exo1 domains are represented by a sphere on the corresponding C α (M,P,E superscript stand for Mlh1, Pms1 and Exo1).

Suppl Fig 6 : Putative Mlh1 PIP-box motifs and superimposition of yeast Mlh1(CTD) with human MLH1(CTD).

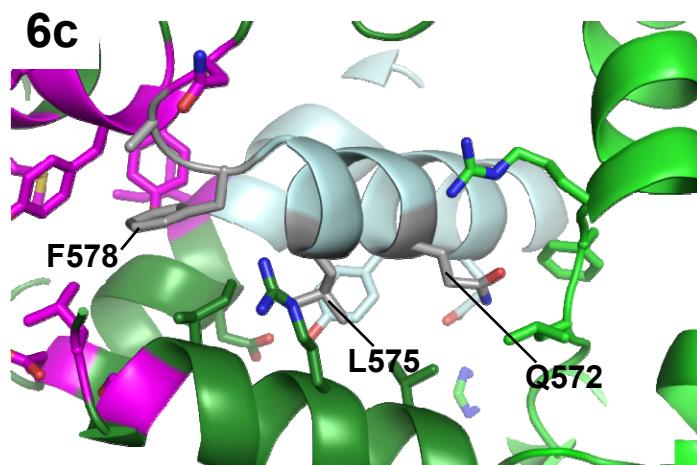
6a

		Qxxhxxaa
hP21	(144)	KRR <u>Q</u> TSM <u>TDFYHS</u>
hP66	(456)	ANR <u>QVSITGFFQR</u>
hFen1	(337)	GST <u>QGRLLDDFFKV</u>
hMSH6	(4)	MSR <u>QSTLYSFFPL</u>
yMlh1-1	(572)	LFY <u>QIGLTD<u>FANF</u></u>
hMLH1-1	(562)	LFY <u>QILLYD<u>FANF</u></u>
yMlh1-2	(552)	AAI <u>QHDL<u>KLFLID</u></u>
hMLH1-2	(542)	ALA <u>QHTKLYLLN</u>

6b

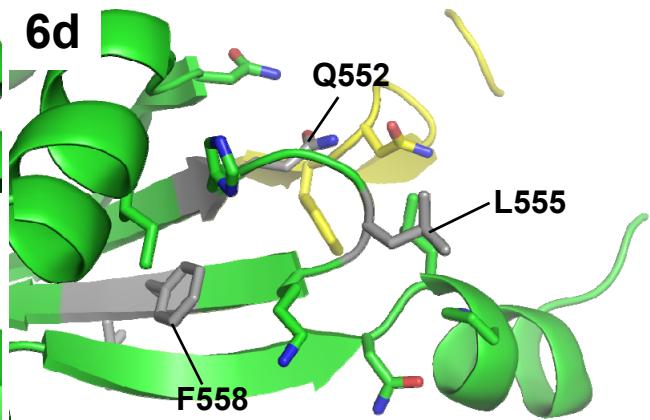


6c



yMlh1-1 (572) LFYQIGLTDFANF

6d

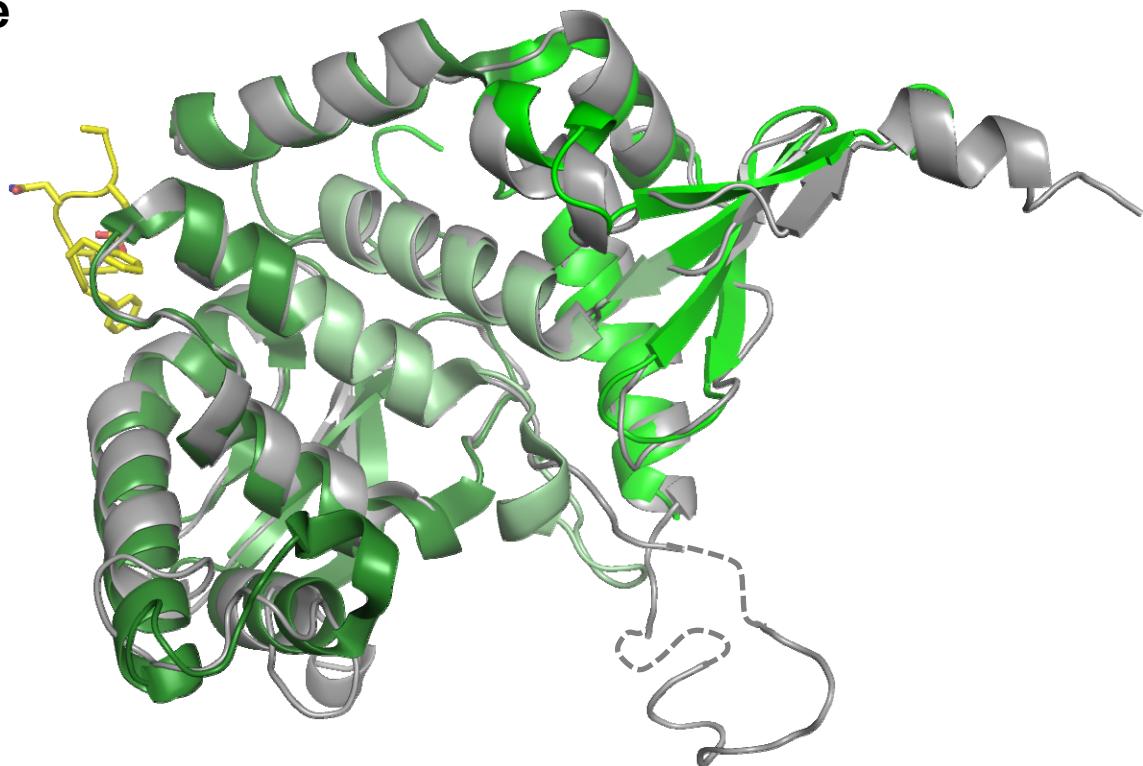


yMlh1-2 (552) AAIQHDLKLFLID

Supplementary Fig. 6 : (a) Sequences of some canonic PIP-box motifs and of the putative Mlh1 PIP-box motifs proposed. Lee et al have proposed two potential degenerated PIP-box motifs in the yMlh1(CTD) region (Mlh1-1 and Mlh1-2) (Lee et al, 2006). **(b) Mapping of the Mlh1-1 and Mlh1-2 putative PIP-box motifs in the yeast MutL α (CTD) structure.** **(c) The Mlh1-1 PIP-box motif.** This motif is deeply buried in the regulatory region, close to the yMlh1 S2 site and present a structure different from the classical PIP-box structure. **(d) The Mlh1-2 PIP-box motif.** This motif is also buried in yMlh1 and is located close to the Pms1 interface. This suggests that Mlh1-1 and Mlh1-2 motifs are not functional PIP-box motifs without any major conformational rearrangement of the Mlh1(CTD) regions to expose these motifs.

Suppl Fig 6 : Putative Mlh1 PIP-box motifs and superimposition of yeast Mlh1(CTD) with human MLH1(CTD).

6e



Supplementary Fig. 6 : (e) Superimposition of the human MLH1(CTD) on the 3D model of human MLH1(CTD) made from the crystal structure of the *S. cerevisiae* MutL α (CTD). The MLH1(CTD) homodimer structure is in dark and light green (3RBN, not published) and the model of MLH1(CTD) made from the yeast structure is in grey. The model from yeast Mlh1 superimposes well with the human MLH1(CTD) (rmsd of 1.71 Å over 205C α). A linker introduced by the cloning and present in the human MLH1(CTD) homodimer crystal form is represented in yellow. It occupies the S2 site due to the presence of a YF dyad in this peptide reminiscent of the [Y/F]F dyad of the MIP-box motif

Suppl Tables 1: Two hybrid and complementation assays using mutants of Mlh1 and Pms1 at the S1 site on Patch II and Patch III.

(a)	^a Mlh1	(b)	^b Ntg2	^b Exo1	^b Pms1	^b Mlh3
^b Pms1-WT	+	^a Mlh1-WT	+	+	+	+
^b Pms1-D870A	-	^a Mlh1-R547A	+	+	-	+
^b Pms1-ΔC10	-	^a Mlh1-ΔC10	+	+	+	+
		^a Mlh1-ΔC1	+	+	+	+

(c)	Mutation Rate (10^{-7})
Empty	59.7 (46.4-63.9)
Mlh1-WT	6.3 (4.9-6.9)
Mlh1-R547A	25.1 (21.5-28.8)
Mlh1-ΔC1	69.2 (50.0-88.9)

Supplementary Table 1: (a-b) Y2H assays. ^aPrey constructs and ^bBait constructs (see Supplementary Notes (a3)). (+): blue colonies, (-): white colonies. **(c) Complementation assays.** BSL124 *Δmlh1* was transformed using pGBT9 vector either empty or expressing Mlh1 WT or mutant. Mutation rates were determined from the number of Can^R mutants by the method of the median. Each value represents ≥ 22 independent cultures. The numbers in parentheses indicate the low and high values for the 95% confidence interval for each rate. (see Supplementary Notes (a4)).

Suppl Table 2 : Two hybrid assays to refine the mapping of the Mlh1 S2 site.

	^b Ntg2	^c Exo1	^d Pms1	^e Mlh3
^a Mlh1-WT	+	+	+	+
Mlh1-L511A	-	-	+	+
Mlh1-M623A	+	+	+	+
Mlh1-S625A	+	+	+	n.d.
Mlh1-E680A	+	+	+	+

Supplementary Table 2: Interactions were monitored using the following partners from *Saccharomyces cerevisiae*. ^apACT2-Mlh1(483-769) CTD-WT and mutants (L511A, M623A, S625A and E680A) were used as prey constructs. ^bpGBT9-Ntg2(1-380), ^cpAS2ΔΔ-Exo1(400-702), ^dpAS2ΔΔ-Pms1(661-873), ^epGBT9-Mlh3(460-715) were used as bait constructs. (+): blue colonies, (-): white colonies, (n.d.) not determined.

Supplementary Notes (1/3)

(a) Experimental Procedures

(a.1) Production of Seleno-methionine Mlh1(CTD)-Pms1(CTD): The production of the Se-Met Mlh1-Pms1 complex was performed with the same protocol than the non-labelled complex, except that the cellular cultures were performed in minimal medium supplemented with most amino acids and selenomethionine as detailed below. For 2 liters medium, the following solutions are prepared: (i) amino acids solutions: 17 solutions with 400 mg of 17 amino acids (Sigma Aldrich), prepared separately, are dissolved in 40 mL sterilized H₂O (17 amino acids are all amino acids except methionine, tyrosine, and cysteine); (ii) seleno-methionine solution: 250 mg of seleno-methionine (Acros Organic) are dissolved in 200 ml sterilized H₂O; (iii) salt solution: a solution with 4 g KH₂PO₄, 16 g NaHPO₄, 1 g NaCl dissolved in 1,100 mL; (iv) stock oligo solution: prepared with 0.5 g EDTA, 415 mg FeCl₃, 42 mg ZnCl₂, 6.5 mg CuCl₂:2H₂O, 5 mg CoCl₂:6H₂O, 5 mg H₃BO₃ dissolved in 500 mL H₂O and filtered on 0.22µm filter. The medium is prepared by mixing the salt solution (iii), 20 ml of stock oligo solution (iv), 2 ml of thiamine at 1 mg/mL, 2 mL of biotine at 1mg/mL. We add 0.49 g MgSO₄ and 0.1 g CaCl₂ and heat a little at 37 °C if necessary. We add 4 g glucose and 1 g (NH₄)₂SO₄ and 2 ml glycerol 50%. This intermediate solution is filtered on 0.22 µm filter. We finally add the 17 solutions of amino acids (i) and the solutions of seleno-methionine (ii) and the antibiotic required.

(a.2) Yeast strains, plasmids and microbiological methods : *S. cerevisiae* strains used are BSL120 *MATa*, *ura3-52*, *leu2Δ1*, *trp1Δ63*, *his3Δ200*, *lys2ΔBgl*, *hom3-10*, *ade2Δ1*, *ade8*, *met* -, BSL124 like BSL120 but *mlh1::hisG* (R. Kolodner and P. Bertrand) and Y190 *MATa*, *gal4*, *gal80*, *his3*, *trp1*, *ade2*, *ura3*, *leu2*, *URA3::GAL1::lacZ*, *LYS2::GAL4(UAS)::HIS3 cyhR*. Yeast cells were grown at 30°C in YPD medium (1% yeast extract, 1% bactopeptone, 2% glucose, with 2% agar for plates) or YNBD medium (0.17% yeast nitrogen base without amino acids, 2% glucose, with 2% agar for plates) supplemented with appropriate amino acids and bases. Point mutations on plasmid DNA were performed by PCR-mediated mutagenesis (QuickChange site-directed mutagenesis kit, Stratagene, La Jolla, CA) using relevant plasmids and specific primers. Yeast transformation assays were performed using the polyethylene glycol/lithium chloride method (Gietz et al, 1992). Bacterial strains XL-10, DH5α and JM105 were used for plasmid construction, preparation and mutagenesis. Gene disruptions and mutations were confirmed by PCR on plasmid DNA and sequencing analysis. Details of strains, plasmids and primers are available upon request.

(a.3) Yeast two-hybrid assays. The Y2H assays were performed using strain Y190 *MATa*, *gal4*, *gal80*, *his3*, *trp1*, *ade2*, *ura3*, *leu2*, *URA3::GAL1::lacZ*, *LYS2::GAL4(UAS)::HIS3 cyhR*. Yeast two-hybrid plasmid vectors are pGBT9 (2μ *pADH1Δ GAL4-BD TRP1 Amp^R*) (Stratagene) and pAS2ΔΔ (2μ *pADH1 GAL4-BD TRP1 Amp^R*) (Gellon et al, 2002) for bait constructs and pACT2 (2μ *pADH1Δ GAL4-AD LEU2 Amp^R*) (Stratagene) for prey constructs (Gellon et al, 2002). Plasmids pACT2-Mlh1(483-769)-CTD and mutant (K547A), pACT2-Mlh1(483-759)-CTD (ΔC10), and pACT2-Mlh1(483-768)-CTD (ΔC1) were used as prey constructs. Plasmids, pAS2ΔΔ-Pms1(661-873)-CTD and mutant (D870A), pAS2ΔΔ-Pms1(661-863)-CTD (ΔC10), pAS2ΔΔ-Exo1(400-702)-CTD, pGBT9-Ntg2(1-380), pGBT9-Mlh3(460-715)-CTD were used as bait constructs. Y190 was transformed by a pair of plasmids expressing [Gal4-BD-fusion TRP1 (bait) / Gal4-AD-fusion LEU2 (prey)] and plated onto selective media for 3-4 days at 30°C and tested for β-galactosidase production in an overlay plate assay (Gellon et al, 2002).

Supplementary Notes (2/3)

(a.4) Complementation assays. The spontaneous mutation rates at *CAN1* were determined using yeast strains, BSL120 (WT) and BSL124 (*Dmlh1*). BSL120 and BSL214 were transformed with pGBT9-derived plasmids either empty or expressing Mlh1(1-769)(WT), Mlh1(1-769) (R547A) or Mlh1(1-768) (DC1). Yeast strains were grown at 30°C in YNBD medium supplemented with the appropriate bases and amino acids without tryptophan. For each strain, 11 independent cultures of 2 ml were inoculated with about 5x10² cells and grown at 30°C for 2-3 days. The cell density was measured by plating dilutions on YNBD plates without tryptophan and counting the colonies after 2-3 days at 30°C. The quantification of canavanine-resistant mutants (Can^R) was determined after plating aliquots of each culture onto YNBD plates lacking tryptophan but containing 60 mg/L L-canavanin (Sigma). Colonies were counted after 4-5 days at 30°C. All experiments were carried out independently 2-5 times. Mutation rates were determined from the number of Can^R colonies by the method of the median.

(a.5) Endonuclease assay: The MutLα(CTD) nicking activity was assayed as previously described for the wild-type MutLα from *S. cerevisiae* (Kadyrov et al, 2007). Reactions (20 µl-final volume) contained 20 mM Hepes-NaOH (pH 7.6), 20 mM NaCl, 1 mM DTT, 0.2 mg/ml bovine serum albumin, 1% glycerol (w/v), 1 mM MnSO₄, 0.5 mM ATP, 1 nM pVII supercoiled DNA (Veautre and Sarasin, 1997) and MutLα(CTD) (0.1 to 5 µM). Incubation was 20 min at 30°C. Reaction was terminated by addition of SDS (0.1%) and incubation for 5 min at 50°C. Then, 5% glycerol (w/v) was added and products were resolved by electrophoresis through 0.8% agarose in TAE buffer. After staining with 0.5 µg/ml ethidium bromide, DNA species were quantified using Molecular Imager and Image 3.0 software by BIORAD.

(a.6) DNA binding assay and apparent *Kd* determination: Electrophoretic mobility shift assay (EMSA) mixtures (20 µl-final volume) contained 20 mM Hepes-NaOH (pH 7.6), 20 mM NaCl, 1 mM DTT, 0.2 mg/ml bovine serum albumin, 5% glycerol (w/v), 1 mM MnSO₄, 0.5 mM ATP, 50 fmoles (2.5 nM) [5'-³²P]-end labelled 59-bp homoduplex substrate (Guibourt et al. 2000) and MutLα(CTD) (0.1 to 5 µM). Incubation was 30 min at 4°C. Half of the reaction (10 µl) was loaded onto a 6% non-denaturing polyacrylamide gel containing 0.25 x TBE and electrophoresed at 95 V for 2 hours at 4°C. Gels were dried on 3MM Whatman paper and analysed by PhosphorImaging Typhoon Trio (Amersham Biosciences) and ImageQuant software. Free DNA and bound DNA were quantified. Assuming a 1:1 stoichiometry between MutLα(CTD) and the DNA probe, the apparent dissociation constant (*K_d*) can be determined from the concentration of free MutLα(CTD), free DNA and DNA-MutLα(CTD) complex. If the total concentration of MutLα(CTD) is close to that of the free MutLα(CTD) at equilibrium: *K_d*= MutLα(CTD) total concentration, when 50% of the DNA is bound by the MutLα(CTD) heterodimer.

Supplementary Notes (3/3)

(b) Additionnal information on the interactions involved in the three patches of the heterodimer interface

These information correspond to interactions not described in the main text.

Patch I: The hydrophobic residues involved in the Patch I are also conserved in the bacterial MutL homodimerization interfaces. The Patch I is bordered by three salt-bridges (D706^M-K673^P, K724^M-E675^P and, E725^M-K829^P) and one hydrogen bond between (V542^M carbonyl and Y695^P hydroxyl).

Patch II: The patch II also involves three hydrogen bonds (S866^P-R546^M (carbonyl), D870^P(carbonyl)-R672^M and Y871^P-M701^M(amide nitrogen)) and van der Waals interactions with hydrophobic residues W864^P; F867^P and I873^P.

Patch III: It involves three hydrogen bonds between Mlh1 end and Pms1(CTD) (Y763^M-N681^P, E767^M-R852^P and R768^M(carbonyl)-R852^P).

References of the Supplementary Informations

- Ban, C., Junop, M., and Yang, W. 1999. Transformation of MutL by ATP binding and hydrolysis: a switch in DNA mismatch repair. *Cell* **97**(1): 85-97.
- Cowtan, K. 2006. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* **62**(Pt 9): 1002-1011.
- Dherin, C., Gueneau, E., Francin, M., Nunez, M., Miron, S., Liberti, S.E., Rasmussen, L.J., Zinn-Justin, S., Gilquin, B., Charbonnier, J.B., and Boiteux, S. 2009. Characterization of a highly conserved binding site of Mlh1 required for exonuclease I-dependent mismatch repair. *Mol Cell Biol* **29**(3): 907-918.
- Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**(5): 1792-1797.
- Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. 2010. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**(Pt 4): 486-501.
- Fiser, A. and Sali, A. 2003. Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol* **374**: 461-491.
- Gellon, L., Werner, M., and Boiteux, S. 2002. Ntg2p, a *Saccharomyces cerevisiae* DNA N-glycosylase/apurinic or apyrimidinic lyase involved in base excision repair of oxidative DNA damage, interacts with the DNA mismatch repair protein Mlh1p. Identification of a Mlh1p binding motif. *J Biol Chem* **277**(33): 29963-29972.
- Gietz, D., St Jean, A., Woods, R.A., and Schiestl, R.H. 1992. Improved method for high efficiency transformation of intact yeast cells. *Nucleic Acids Res* **20**(6): 1425.
- Guibourt, N., Castaing, B., Van Der Kemp, P.A., and Boiteux, S. 2000. Catalytic and DNA binding properties of the ogg1 protein of *Saccharomyces cerevisiae*: comparison between the wild type and the K241R and K241Q active-site mutant proteins. *Biochemistry* **39**(7): 1716-1724.
- Inoue, H., Nojima, H., and Okayama, H. 1990. High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96**(1): 23-28.
- Kadyrov, F.A., Holmes, S.F., Arana, M.E., Lukianova, O.A., O'Donnell, M., Kunkel, T.A., and Modrich, P. 2007. *Saccharomyces cerevisiae* MutLalpha is a mismatch repair endonuclease. *J Biol Chem* **282**(51): 37181-37190.
- Lee, S.D. and Alani, E. 2006. Analysis of interactions between mismatch repair initiation factors and the replication processivity factor PCNA. *J Mol Biol* **355**(2): 175-184.
- McCoy, A.J. 2007. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr D Biol Crystallogr* **63**(Pt 1): 32-41.
- Sacho, E.J., Kadyrov, F.A., Modrich, P., Kunkel, T.A., and Erie, D.A. 2008. Direct Visualization of Asymmetric Adenine Nucleotide-Induced Conformational Changes in MutLalpha. *Mol Cell* **29**(1): 112-121.
- Veaute, X. and Sarasin, A. 1997. Differential replication of a single N-2-acetylaminofluorene lesion in the leading or lagging strand DNA in a human cell extract. *J Biol Chem* **272**(24): 15351-15357.