

# A conserved motif in the disordered linker of human MLH1 is vital for DNA mismatch repair and its function is diminished by a cancer family mutation

Karla Wolf<sup>1</sup>, Jan Kosinski<sup>2</sup>, Toby J. Gibson<sup>3</sup>, Nicole Wesch<sup>4</sup>, Volker Dötsch<sup>4</sup>, Maurizio Genuardi<sup>5</sup>, Emanuela Lucci Cordisco<sup>6</sup>, Stefan Zeuzem<sup>1</sup>, Angela Brieger<sup>1</sup> and Guido Plotz<sup>1,\*</sup>

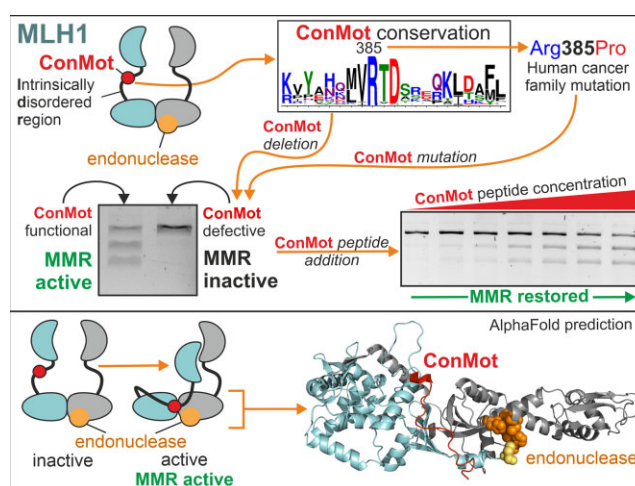
<sup>1</sup>Department of Internal Medicine 1, University Hospital, Goethe University, Frankfurt am Main, 60590, Germany, <sup>2</sup>European Molecular Biology Laboratory (EMBL), Centre for Structural Systems Biology (CSSB), Hamburg, 22607, Germany, <sup>3</sup>European Molecular Biology Laboratory (EMBL), Structural and Computational Biology Unit, Heidelberg, 69117, Germany, <sup>4</sup>Institute of Biophysical Chemistry and Center for Biomolecular Magnetic Resonance, Goethe University, Frankfurt am Main, 60438, Germany, <sup>5</sup>UOC Genetica Medica, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome 00168, Italy and <sup>6</sup>Dipartimento di Scienze della Vita e di Sanità Pubblica, Università Cattolica del Sacro Cuore, Rome 00168, Italy

Received December 12, 2022; Revised April 26, 2023; Editorial Decision April 28, 2023; Accepted May 17, 2023

## ABSTRACT

DNA mismatch repair (MMR) is essential for correction of DNA replication errors. Germline mutations of the human MMR gene MLH1 are the major cause of Lynch syndrome, a heritable cancer predisposition. In the MLH1 protein, a non-conserved, intrinsically disordered region connects two conserved, catalytically active structured domains of MLH1. This region has as yet been regarded as a flexible spacer, and missense alterations in this region have been considered non-pathogenic. However, we have identified and investigated a small motif (ConMot) in this linker which is conserved in eukaryotes. Deletion of the ConMot or scrambling of the motif abolished mismatch repair activity. A mutation from a cancer family within the motif (p.Arg385Pro) also inactivated MMR, suggesting that ConMot alterations can be causative for Lynch syndrome. Intriguingly, the mismatch repair defect of the ConMot variants could be restored by addition of a ConMot peptide containing the deleted sequence. This is the first instance of a DNA mismatch repair defect conferred by a mutation that can be overcome by addition of a small molecule. Based on the experimental data and AlphaFold2 predictions, we suggest that the ConMot may bind close to the C-terminal MLH1-PMS2 endonuclease and modulate its activation during the MMR process.

## GRAPHICAL ABSTRACT



## INTRODUCTION

DNA mismatch repair (MMR) is a highly conserved repair system; it is active predominantly in the correction of base-base mismatches and insertion-deletion loops that occur during replication. It increases replication fidelity by three orders of magnitude (1). Its heterozygous inactivation in the germline causes Lynch syndrome (OMIM #120435), a dominantly heritable cancer predisposition which affects mainly endometrium and colorectum, but also other organs. Lynch syndrome accounts for 2–5% of all colorectal cancer (CRC) cases, but approximately 15% of sporadic

\*To whom correspondence should be addressed. Tel: +49 69 6301 87668; Fax: +49 69 6301 84441; Email: plotz@med.uni-frankfurt.de

CRC patients also show somatic inactivation of MMR, mostly due to hypermethylation of the MLH1 promoter (2).

Two protein dimers are the major components of DNA mismatch repair: both MutS and MutL form homodimers in prokaryotes and heterodimers in eukaryotes. In humans, MSH2-MSH6/3 and MLH1-PMS2 are the major repair proteins. Genetic inactivation of MLH1, MSH2, MSH6 and PMS2 underlies Lynch syndrome (3–6).

MutS proteins are ATPases and versatile detectors of base mismatches and insertion-deletion loops in the DNA duplex (5,7–10). Their mode of mismatch detection has been studied in several crystal structures and involves kinking of the DNA duplex at the mismatched site (for review, see (11)). Minimal MMR reactions have been reconstituted *in vitro* from purified components and, in case of the human system, rely on MutS $\alpha$ , MutL $\alpha$  and some additional factors (RPA, RFC, PCNA, exonuclease I and DNA polymerase) (12–14). However, the mechanism of the repair reaction is still under discussion (15–17). It is difficult to assess experimentally since both protein dimers can assume a multitude of conformational states, depending on the loading of their ATP binding sites and on interactions with homoduplex DNA, heteroduplex DNA or with each other (18–21). Moreover, MutS proteins may utilize different modes of movement on DNA (for mismatch search and for repair initiation) (for review, see (22)). The repair reaction is initiated by mismatch recognition, which authorizes binding of ATP by MutS (21,23). This induces a comprehensive conformational transition that has been found to transform MutS proteins into stably sliding DNA clamps, and sliding of MutS proteins as well as of MutS-MutL complexes is considered to be an important intermediate of the DNA mismatch repair reaction (24–28). Alternatively, it has also been suggested that MutL directs MutS to remain stationary in the vicinity of the mismatch after ATP binding, suggesting that repair is initiated by mispair-bound MutS-MutL complexes (29,30). ATP binding allows interaction of MutS with MutL dimers via their N-terminal domain (19,21,31,32). Before initiating DNA strand removal, MMR has to identify the newly replicated (and therefore erroneous) strand. While hemi-methylated GATC sites are used in some gammaproteobacteria for this task (33), interaction with PCNA at the replication fork is considered to allow strand discrimination in other organisms (34–38). After strand identification, the endonucleolytic activity of MutL proteins is activated (39,40). This activation is PCNA-dependent and introduces nicks in the wrongly replicated DNA strand in vicinity to the mismatch, possibly through direct interactions of MutL N-terminal domains (NTD) and C-terminal domains (CTD) (41–46). The nicks are introduced on both 5'- and 3'-sides of the mismatch in the newly replicated strand to facilitate its digestion (46–49).

Therefore, MutL proteins are mediators between mismatch recognition and removal of the faulty DNA strand. Their precise mode of function is still under discussion. MutL proteins contain ATPase activities in their structured NTD (50). Furthermore, their structured CTDs confer constitutive dimerization (51,52). ATP binding has been found to cause transient dimerization of the NTD (53) and conformational condensation of the protein (44). MutL proteins possess DNA binding capability (49,54–57). They have been

observed to be able to move on DNA duplexes and even traverse obstacles like nucleosomes (27,58,59), presumably by formation of a DNA ring using their flexible linkers and the dimerization of the N-terminal domains (11). Additionally, there is evidence that MutL proteins cooperatively bind DNA and can form active polymers on DNA (60,54).

Eukaryotic MutL proteins interact with DNA and mismatch repair components like MutS proteins and PCNA (31,32,54,61–64), but also with further proteins, conferring effects on proliferation, metastasis and apoptosis control (65–69), therefore they can be regarded as versatile adaptors for diverse purposes exceeding their immediate role in mismatch repair.

While the catalytic activities of both structured domains of MutL (NTD and CTD) have been investigated, it is unclear how both domains communicate to facilitate the repair reaction and which role the MutL linkers play. The linkers represent intrinsically disordered regions (IDR) (70). Although unstructured, such regions occur in about one third of eukaryotic proteins, and they often fulfil specific functions, frequently involving protein interactions, either within the same or with other proteins (71–73). Small motifs or modules located within these IDRs may facilitate such tasks (74). The linker regions in MLH1 and PMS2 show no general evolutionary conservation in sequence and have significant differences in length, being 100–300 amino acids long, in case of the PMS2 homologue MLH3 even 800 amino acids. Small coding variations in these regions found in patients are considered non-pathogenic because practically all pathogenic variants have as yet been observed in the structured N- and C-terminal domains (15). Although they are flexible and of significant length, the linkers are required for repair since proteolytic cleavage of the linker *in vitro* reduces MutL DNA binding capacity, and cleavage *in vivo* abrogates MMR (75). Deletions in the yeast MLH1 linker conferred effects on MMR capacity, DNA-stimulated ATPase activity, movement on DNA and nicking (76). It was concluded that the linkers are relevant for completing the ATPase cycle and therefore for both movements on DNA and efficient endonuclease activation. A recent study using ‘handcuffing’ of the linkers via insertion of rapamycin-dependent FRB and FKBP dimerization domains in specific positions of the MLH1 and PMS2 linkers showed that inter-dimer handcuffing close the C-terminal dimerization domain caused none or weak effects, while N-terminal handcuffing conferred decreased MMR activity and DNA binding (77). However, ATPase- and endonuclease activities were increased, suggesting an inappropriate activation of MutL $\alpha$ . Handcuffing the MLH1 linker with itself displayed the most pronounced MMR defect: while no effect on ATPase activity was detectable, DNA binding and endonuclease activity were reduced, suggesting that the MLH1 linker may be important for conformational rearrangements bringing the DNA strand to the endonuclease site.

We have investigated the human MLH1 linker region and identified a small motif (subsequently termed ConMot) that, in contrast to the rest of the linker, displays a high degree of conservation exclusively in eukaryotes. Genetic variants of unclear clinical significance have been reported in this motif in cancer patients. We have analyzed the role

of the ConMot and the variants located therein concerning their effects on protein expression and mismatch repair activity.

## MATERIALS AND METHODS

### Cell culture, reagents, vectors and variant synthesis

HEK293 and HEK293T cells were purchased at DSMZ German Collection of Microorganisms and Cell Culture (Braunschweig, Germany) and maintained in D-MEM containing 10% FCS and antibiotics/antimycotics.

The pcDNA3 expression vector containing the entire open reading frame of human MLH1 was a gift of Dr Hong Zhang (Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, USA). The pSG5 expression vector containing full-length human PMS2 cDNA was provided by Dr Bert Vogelstein (Johns Hopkins Oncology Center, Baltimore, MD, USA). Amino-acid positions in MLH1 refer to the 756 aminoacid reference MLH1 sequence (NP\_000240.1).

Sequence variants of the MLH1 expression vector were generated using the Q5 site directed mutagenesis system (New England Biolabs, Ipswich, MA, USA) with the appropriate primers (Supplementary Table S1).

Four peptides were used for the complementation assays: a 32-mer peptide containing the sequence of the MLH1-ConMot (ConMot-32, SGSSDKVYAHQMVRTDSRE-QKLDAFLQPLSK), a 32-mer peptide with a scrambled sequence of the MLH1-ConMot (Scramble-32, SGSSD-MVQFEKLKDRRSQDATYVAHLQPLSKP), a 20-mer peptide which contains the conserved sequence of the ConMot (ConMot-20, KVYAHQMVRTDSREQKLDAF). These three peptides were synthesized by GenScript (Leiden, Netherlands). Furthermore, another 32-mer control peptide (H3) with the Histone H3-N-terminal sequence (ARTKQTARKSTGGKAPRKQLATKAARKSAPAT) purchased from EpiCypher, Durham, USA, was used.

### Transfection, protein extraction and expression analysis

MLH1-deficient HEK293T cells (78) were transiently transfected with 5 mg of vector DNA and 20 mL of polyethyleneimine (1 mg/mL, 'Max' linear, 40 kDa, Polysciences, Warrington, PA) and extracted as described previously (79). The extracts were analyzed by SDS-PAGE and immunoblotting (using anti-MLH1, G168-728, BD Biosciences, and anti-PMS2, E-19, and anti-beta-Actin, C2, from Santa Cruz Biotechnologies). Chemiluminescence signals (Immobilon, Millipore) were detected in an LAS-4000 mini camera (Fuji) and quantified using Multi Gauge v3.2.

For determining MLH1 concentration, pure MutL $\alpha$  (MLH1-PMS2) was used, which was a kind gift of Josef Jiriny, Zürich, Switzerland (80).

### Mismatch repair assay

Mismatch repair activity was assessed as described in detail before (79,81) (Supplementary Figure S3). In short, 50  $\mu$ g of nuclear extract of HEK293T cells, which do not express MLH1-PMS2 due to a hypermethylation of the MLH1 promoter (78), was supplemented with 5  $\mu$ g native extract

of HEK293T cells co-transfected with plasmids encoding MLH1 and PMS2; the MLH1 plasmid either encoded the wildtype sequence or the variant to be investigated. Mismatched DNA plasmid substrate (Supplementary Figure S3) prepared according to previously published procedures (81–83) was added, and reactions in a final volume of 25  $\mu$ L were incubated at 37°C for 15 min. Repair efficiency was scored by purifying the mismatched DNA plasmid from the reaction and assessing the fraction of homoduplex sequence by restriction digestion of the mismatched site using EcoRV. The substrate was additionally linearized with AseI.

For complementation assays with peptide, the indicated peptide(s) were used at stock concentrations of 50 or 500  $\mu$ M in HEPES–KOH 20 mM pH 7.6 and supplemented to the MMR reactions.

### Circular dichroism (CD) and melting analysis of ConMot peptide

For the CD analysis, 33  $\mu$ M ConMot peptide was used in 20 mM Na-phosphate buffer pH 7.2, 20°C. Three scans were performed with 190–300 nm wavelength, using a 1 mm light path. A blank was measured with buffer. For the melting analysis, 33  $\mu$ M ConMot peptide was used in 20 mM Na-phosphate buffer pH 7.2. Detection was at 222 nm, temperature was scanned from 25–95°C. A blank with buffer only was subtracted.

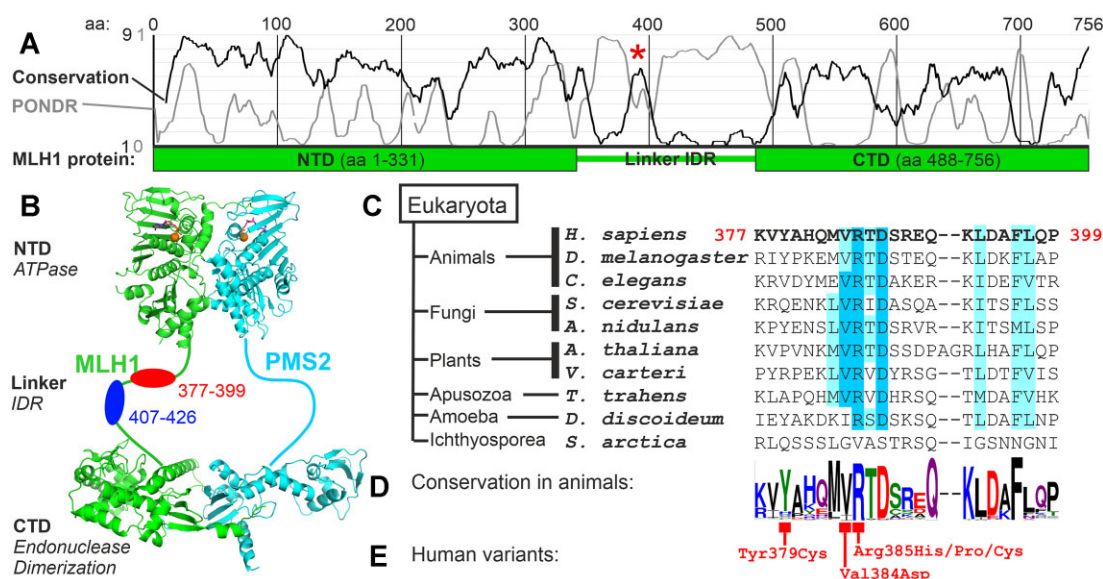
### Bioinformatic analyses

Sequences for alignments were retrieved using BLAST with the human MLH1 protein sequence (NM\_000240) as query. Retrieved sequences were specified to organisms' classes or kingdoms in question by restricting the search on certain TaxIDs. The resulting hits were manually curated according to established procedures (84). Only one MLH1 sequence per organism was retained, removal of incomplete sequences, and verification that the retrieved sequence corresponds to an MLH1 (and not PMS2 or other) protein, which can easily be identified by the highly conserved, C-terminal FERC sequence of eukaryotic MLH1 proteins (48,85). By that procedure, 567 sequences from animals, 348 from fungi, 117 from embryophyta could be identified, while few sequences were available from other eukaryotic kingdoms (3 from red algae, 9 from green algae, 8 from amoebozoae, 1 apusozoa and 2 choanoflagellata). Sequences were handled in JalView (86). Alignments were performed with Muscle (87). Conservation scores were calculated with these alignments using ConSeq (88). Conservation images of motifs were created using WebLogo 3 (89).

The model of MLH1-PMS2 complex was built using AlphaFold (90) version 2.2.2 with all options set to default except the max\_recycles parameter which was increased from 3 to 12 maximise model accuracy. AlphaFold was run using AlphaFold pipeline (91).

Peptide folding predictions were performed using PEP-FOLD3 (92).





**Figure 1.** Location, conservation and human variation of a conserved MLH1 linker motif. (A) Conservation scores of the MLH1 protein sequence were calculated using ConSeq, resulting in no conservation (score 1) to very high conservation (score 9). The PONDR score was calculated using the human MLH1 sequence; scores above 0.5 are associated with intrinsically disordered regions (IDRs). The linker region displays low conservation and high PONDR scores except for a small area (marked by an asterisk). (B) Overview of the MutLα heterodimer consisting of MLH1 and PMS2. Both contain structured, N-terminal ATPase domains (NTD) capable of transient dimerization and structured C-terminal endonuclease domains (CTD) conferring constitutive dimerization. NTD and CTD are connected by a linker IDR. The conserved area (aa 377–396) is located in this linker (red). A control region deleted for subsequently described experiments is also shown (blue). (C) Conservation of the conserved motif in organisms of different eukaryotic kingdoms. Dark blue shading represents high conservation, light blue shading weaker conservation. (D) WebLogo presentation of conservation in animals. Degree of conservation is represented by letter size, variety of residues in the respective position is given by letters of different sizes. (E) Human variants have been observed in the conserved area and are annotated in the WebLogo presentation.

## RESULTS

### Eukaryotes have a conserved motif within the intrinsically disordered MLH1 linker region

MutL proteins, including human MLH1 and PMS2, possess two structured, highly conserved functional domains: the ATPase function is located in the N-terminus (NTD), while an endonucleolytic activity is located in the C-terminus (CTD) of PMS2 (Figure 1A and B). The CTDs also confer a constitutive dimerization of both proteins (51,52), while the N-terminal ATPase domain can transiently dimerize, controlled by the binding and hydrolysis of ATP (53).

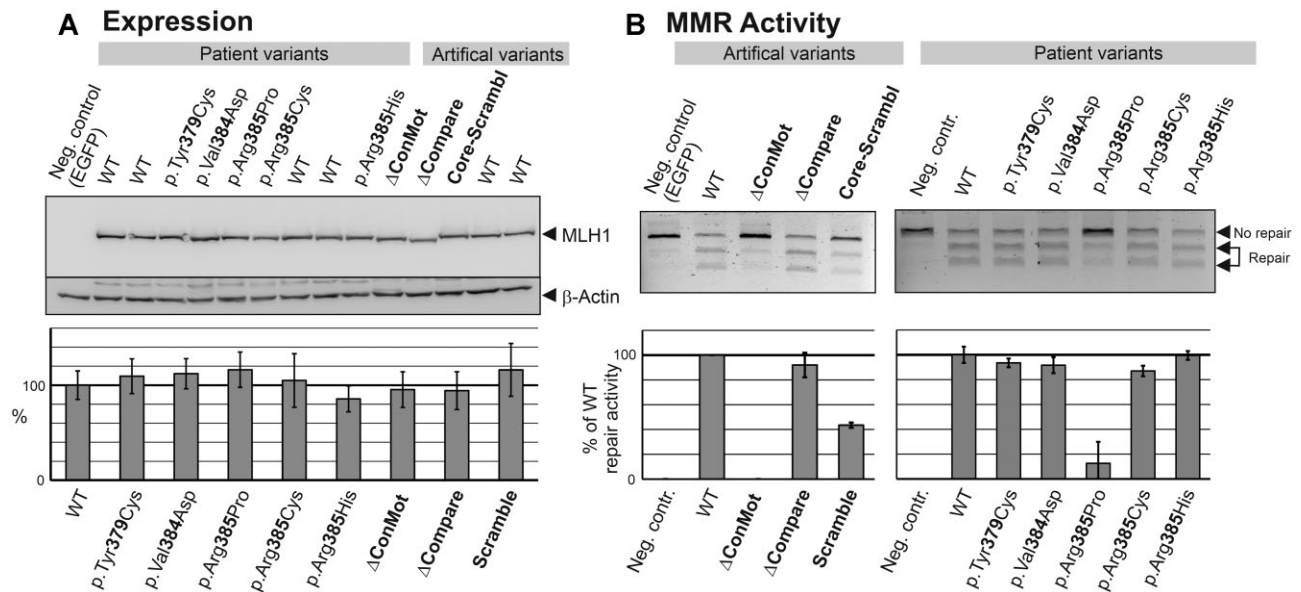
Both structured domains are connected by a non-conserved linker of variable length, which has been suggested to contribute to DNA binding (75). This sequence represents an intrinsically disordered region (IDR) of the protein, as is illustrated by a PONDR analysis (Figure 1A) (93). Consistent with the low sequence and length conservation of this IDR, small coding alterations in this region that have been identified in humans are usually considered non-pathogenic (15).

Depending on the composition of sequences analyzed in MutL alignments, a small conserved area becomes evident within the disordered linker region (Figure 1A). Along with increased conservation, the PONDR score is reduced, suggesting that this area may fold into a secondary structure (Figure 1A). A comprehensive alignment analysis demonstrated that the peak of conservation within the linker corresponds to a highly conserved MVRTD motif found exclu-

sively in MLH1 proteins of the eukaryotic domain, which is embedded in a less conserved motif embracing 22–24 amino acids (Figure 1C and D). Kingdom-specific alignments show that these flanking sequences comprise residues highly conserved within their respective eukaryotic kingdom but showing some variation between kingdoms (Supplementary Figure S1).

The strong sequence conservation suggests a biological significance of this motif which is, for ease of reading flow, termed ConMot subsequently. Notably, within the ConMot sequence, human variants have been identified in cancer patients (Figure 1E) (94).

Further sequence evaluation demonstrated that not only a reduction in PONDR score is predicted for the ConMot, but structure prediction algorithms suggest an elevated propensity to form a secondary structure. The peptide folding algorithm PEP-FOLD3 predicts a potential  $\alpha$ -helical structure in two short sequences of the human ConMot, interrupted by the highly conserved MVRTD motif (Supplementary Figure S2A). Therefore, the sequence may provide a potential for folding into a secondary structure (Supplementary Figure S2B). We have analyzed a 32-mer peptide containing the human ConMot sequence by circular dichroism and melting analyses (Supplementary Figure 2C and D). Neither investigation provided evidence for secondary structure formation of the ConMot peptide. However, the investigated 32-mer peptide is rather short for forming stable secondary structures and may lack protein or DNA interactions that under biological conditions would confer secondary structure stabilization.



**Figure 2.** Expression and DNA mismatch repair activity of human and artificial ConMot variants. Expression vectors encoding either WT or variant MLH1 and PMS2 were co-transfected in HEK293T cells which do not express endogenous MLH1-PMS2. As a negative control, one sample was only transfected with an EGFP. After 48 h, native whole cell protein extracts were prepared. Expression was assessed by SDS-PAGE and western blotting. (A) Representative western blot of MLH1 wildtype and variant expressions. Average expression levels of MLH1 of > 5 independent transfections were determined and are given in the bar diagram below. (B) The extracts of the transfections were used to assess the DNA mismatch repair activities of the variant MLH1 proteins in comparison with wildtype (WT) MLH1-PMS2 as detailed in Materials and Methods and as shown in Supplementary Figure S3. The EGFP-transfected extract sample served as negative control. Representative agarose gels are shown in the top panel. The bottom panel shows average repair values of > 3 independent experiments.

### Deletion and scrambling of the ConMot and one human variant abolish DNA mismatch repair

A prominent task of MLH1 is its contribution to the DNA mismatch repair reaction. In order to assess the relevance of the ConMot for MLH1 function, we therefore introduced artificial and human variants in MLH1, expressed the MLH1-PMS2 heterodimers in MLH1-deficient HEK293T-cells and investigated their capacity to repair a G-T mismatch in a test plasmid.

We constructed a deletion variant ( $\Delta$ ConMot) in which the conserved ConMot sequence comprising 20 residues was removed (red in Figure 1B). For comparing if the shortening of the MLH1 linker region can disturb mismatch repair by itself, we generated a deletion-comparison variant in which a non-conserved fraction of the linker comprising also 20 residues was removed ( $\Delta$ Compare) (blue in Figure 1B). For assessing the relevance of the central conserved motif sequence, a scrambled variant was investigated, in which the MVRTD sequence was substituted by the sequence DTMVR.

Wildtype and variant MLH1-PMS2 proteins were well expressed (Figure 2A). Although most human MLH1 missense variants confer functional defects as well as pathogenicity by destabilizing the MLH1 protein (79), neither the human variants nor the artificial variants in the linker region investigated here had any effect on expression, suggesting that even gross alterations in the linker region do not significantly destabilize the MLH1 protein. This is consistent with the absence of large secondary structures in the linker region, and also suggests that potential interactions of the ConMot with the MLH1 NTD or CTD regions do

not significantly contribute to protein stability either. The shortening conferred by the deletion variants was visible by corresponding size shifts (Figure 2A).

We next tested the DNA mismatch repair capacity of all variants by assessing the restoration of a mismatched restriction site in a test plasmid (Supplementary Figure S3). This demonstrated that deletion of the ConMot from MLH1 fully abrogated mismatch repair ( $\Delta$ ConMot), while the more C-terminally located deletion of identical size ( $\Delta$ Compare) had no detectable effect on repair activity (Figure 2B, left panel). The scrambling of the highly conserved MVRTD motif also significantly reduced repair capacity.

The applied MMR assay procedure is not only suitable for research purposes but can also be applied for pathogenicity clarification of human variants (79,81,95). We therefore tested human variants of the ConMot (Figure 1E). Of these, the proline substitution of Arg385 located within the MVRTD motif displayed a loss of mismatch repair activity (Figure 2B, right panel). The catalytic activities of other substitutions, including a cysteine and a histidine substitution of the same residue, were similar as the wildtype sequence.

The core MVRTD motif is highly conserved over the complete eukaryotic domain, while the flanking sequences are also conserved, albeit in a kingdom-specific manner (Supplementary Figure S1). This may suggest that the core motif may be invariant for mechanistical reasons, while the flanking sequence may fulfil other functions that have allowed alternative sequences to evolve. To directly compare the contribution of core sequence versus flanking

sequences, we used the core-scramble variant and a hybrid variant with a plant ConMot as it is, for example, present in walnut and mungo bean. Exclusively in the plant kingdom, a two-residue (alanine-glycine) insertion in the ConMot flanking sequence is common (Supplementary Figure S1), and, while conservation is high in plants, the pattern of conservation is different from animals (Figure 3B). Therefore, the in the human-plant hybrid, the ConMot contains a functional flanking sequence, albeit from a different organism.

We compared the core-scramble and human-plant-hybrid variants with the ConMot deletion variant and a full scramble variant. All variants were well expressed (indistinguishable from wildtype) (Figure 3C). The core scramble variant again had approximately 40% activity, which was comparably active as the human-plant hybrid (Figure 3D). This shows that the flanking sequence, although variant in inter-kingdom comparisons, is only fully functional within its species context, and that both core and flanking sequence contribute to a similar degree to the functionality of the ConMot.

A ConMot peptide restores mismatch repair capacity of ΔConMot

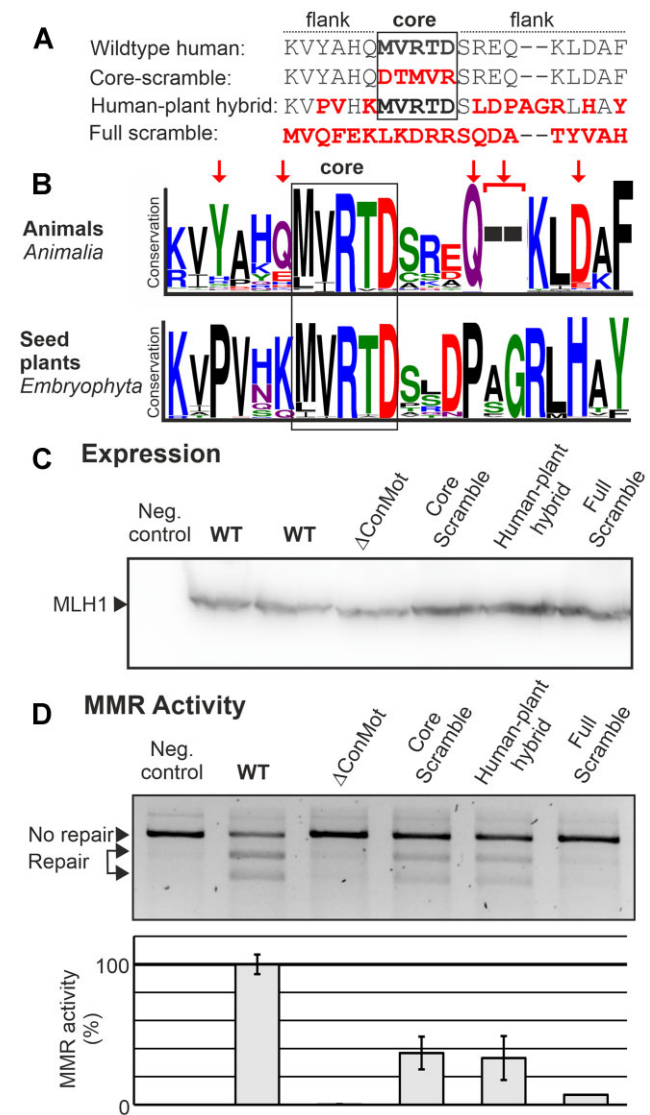
Since the ConMot is localized in the flexible MLH1 linker region, it potentially confers a protein-protein interaction or an interaction with DNA; both could be reconciled with current assumptions on the function of the MLH1 linker.

We performed competition/supplementation tests to gain information on the contribution of the ConMot to the DNA mismatch repair reaction using a 32mer peptide containing the human ConMot sequence.

First, we attempted to interfere with a regular DNA mismatch repair reaction by adding the ConMot peptide to a reaction using extract of MMR-proficient HEK293 cells (Figure 4A). Even 75 μM ConMot peptide did not detectably interfere with the catalytic repair activity of the cell extract, suggesting that the ConMot peptide does not significantly compete with the MLH1-ConMot in a manner causing a defect in mismatch repair.

Second, we used extract of the repair-deficient cell line HEK293T supplemented with either wildtype MutLα (MLH1-PMS2) or the ΔConMot deletion variant, corresponding to experiment shown in Figure 2B. In this experiment, we again added ConMot peptide. Interestingly, the DNA mismatch repair-defect of the ΔConMot variant could be overcome by the addition of exogenous ConMot-peptide (Figure 4B). In contrast, a scrambled control peptide did not restore MMR activity. A smaller peptide (20 aa) containing only the highly conserved ConMot sequence was not active either, suggesting that there is a requirement for additional residues neighboring the ConMot sequence for activity. Since the presence of the 20mer-ConMot did not affect MMR restauration by the 32mer-ConMot, it is likely that target binding, not activation, is compromised in the smaller peptide (Figure 4B).

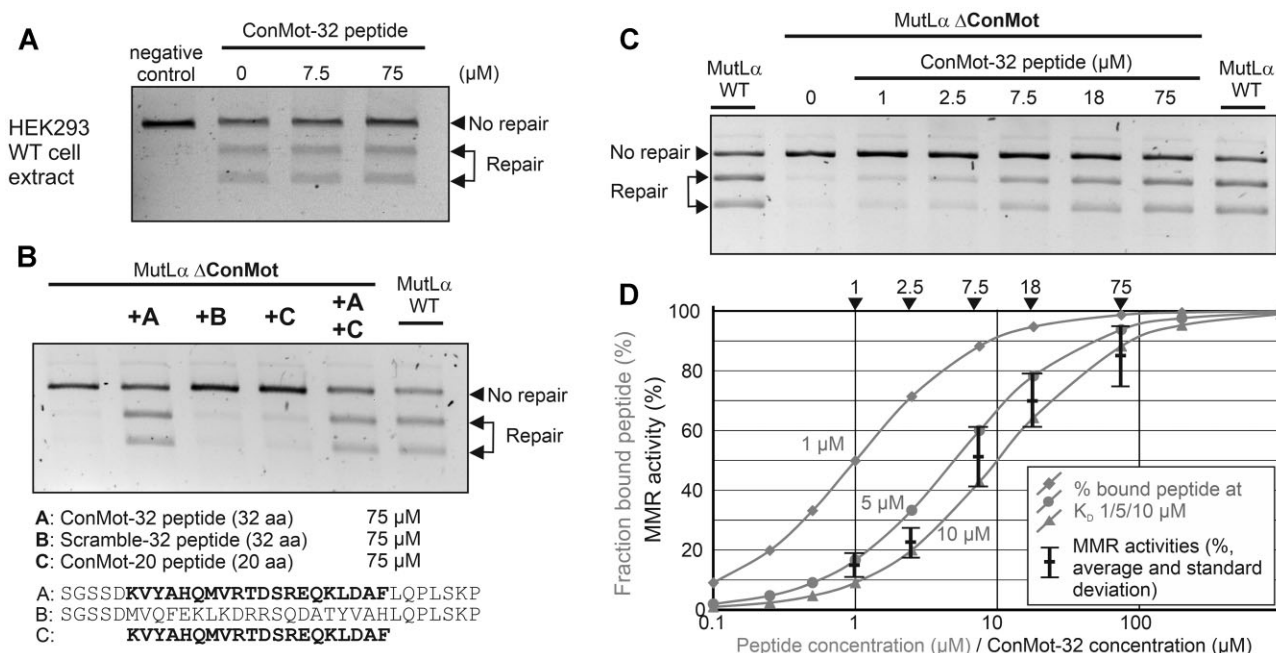
We investigated the effect of the ConMot peptide more closely by titrating the required peptide concentrations for activation. We determined that the mismatch repair reactions performed in this study contained approximately 280



**Figure 3.** Analysis of the relevance of ConMot core and surrounding sequences. MLH1 proteins containing either the wildtype human ConMot motif sequence or variants were directly compared concerning expression and MMR activity. (A) Comparison of the ConMot variants used: the core-scramble variant sequence, the full-scramble variant sequence and an animal-plant hybrid that contains the ConMot consensus sequence of plants as it exists, for example, in walnut and mungo bean (*Vigna radiata* and *Juglans regia*). (B) Comparison of conservation in the animal versus plant ConMot in WebLogo representation. Specific differences in conservation within the sequences flanking the core MVRTD sequence are marked by red arrows above the animal sequence. (C) Western blot of the expression of the different MLH1-ConMot wildtype (WT) and the variants shown in A after transfection in HEK293T cells; negative control: transfection of EGFP. (D) MMR activities were assessed in direct comparison to negative control and wildtype human MLH1 as in Figure 2B. Average activities and standard deviations were calculated from >3 independent experiments.

fmol MLH1 (Supplementary Figure S4A and B), which is in good agreement with previous quantifications of MLH1-PMS2 in cells and extracts (96). An increase in MMR activity became detectable at a ConMot peptide concentration of 1 μM, 50% activation was achieved at 7.5 μM, and almost complete activation at 75 μM (Figure 4C and D).





**Figure 4.** Reaction of DNA mismatch repair to addition of ConMot peptide. (A) Nuclear extract of MLH1-proficient HEK293 cells was used for performing the DNA mismatch repair reaction. Four reactions were undertaken, with two reactions supplemented with either 7.5 μM or 75 μM ConMot-32 peptide. A negative control reaction was performed by omitting the 37°C repair incubation. (B) MMR reactions were performed with nuclear extract of MLH1-deficient HEK293T cells supplemented with extract of cells transfected either with wildtype MLH1-PMS2 (positive control) or with MLH1-ΔConMot-PMS2. To these MMR reactions, different peptides were added at a concentration of 75 μM: A ConMot-32 peptide (A), a Scramble-32 peptide (B) in which the conserved ConMot sequence is scrambled, and a short ConMot-20 peptide (C) that only contains the twenty highly conserved ConMot residues. The peptide sequences are given below the gel image. (C) The concentration-dependence of the MMR-activating effect of the ConMot-32 peptide observed in (B) was measured over a concentration range from 1–75 μM. (D) Theoretical binding curves of a ligand (peptide) to a protein were calculated for dissociation constants ( $K_D$ ) of 1, 5 and 10 μM for a protein concentration as present for MLH1-PMS2 in the MMR assay reactions (see also Supplementary Figure S3). Into this diagram, the average MMR activities (and standard deviations) measured in >3 of the experiment exemplarily shown in C were introduced as bars at ConMot-32 peptide concentrations of 1, 2.5, 7.5, 18 and 75 μM.

Dissociation constants ( $K_D$ ) for peptides to proteins are typically in the range of 1–10 μM (97). We compared the measured concentration-dependent MMR activities with theoretical binding curves of (peptide) ligands to (protein) substrates for  $K_D$  values of 1, 5 and 10 μM (Supplementary Figure S4C) at the determined MLH1 concentration. The degree of MMR restoration corresponded very well to the theoretical fraction of peptide binding to a protein for a  $K_D$  of 7.5 μM (Figure 4D). It is therefore plausible that a binding reaction of the ConMot peptide (to a protein) underlies the restoration of MMR activity.

In an identical way, we analyzed the effects of the ConMot peptide on MMR activity of the defective human MLH1 missense variant p.Arg385Pro (Figure 5). Restoration of MMR activity was again observed, and required identical peptide concentrations as with the ConMot-deletion variant, suggesting that the human p.Arg385Pro variant abolishes the interaction of the MLH1-ConMot with its target binding site, and that addition of ConMot-peptide can replace the function of a mutated ConMot.

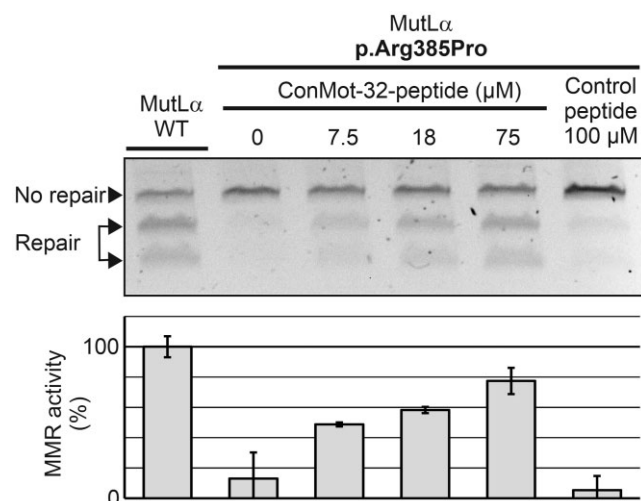
#### The ConMot variant p.arg385tyr partly restores MMR deficiency of the endonuclease variant p.tyr750arg

In our AlphaFold2 model of the MLH1-PMS2 heterodimer, the ConMot is predicted with high confidence to bind to the CTD of MLH1 (Supplementary Figure S5,

Figure 6A). The p.Arg385 residue of the ConMot is predicted to be in sufficient proximity to the p.Tyr750 to engage in a side-chain-side-chain interaction (Figure 6B). Both p.Arg385 and p.Tyr750 are highly conserved, with p.Arg385 being the most highly conserved residue of the ConMot (Figure 6B).

We speculated that a putative p.Arg385-p.Tyr750 side chain interaction may be accessible for experimental verification. We performed an exchange of these residues and created the variants p.Arg385Tyr, p.Tyr750Arg and a double variant with interchanged side-chains (p.Arg385Tyr + p.Tyr750Arg). As controls, we generated p.Arg385Ala and p.Arg385Glu. All variant proteins were well expressed (Figure 6C).

While the p.Arg385Tyr substitution was well tolerated in terms of MMR activity, the p.Tyr750Arg alteration caused a comprehensive defect of MMR (Figure 6D). Intriguingly, this defect was partially reverted in the double variant (p.Arg385Tyr + p.Tyr750Arg): repair activity of this double variant was more than twice as high as that of p.Tyr750Arg (it increased from 22.4% to 50.2% on average). This observation was highly reproducible and significant ( $P = 0.0028$ ). This was not the case for the two control alterations of p.Arg385 to alanine and to glutamate: these had no reverting effect on the defect of the p.Tyr750Arg variant, but instead acted additively to result in a more pronounced MMR defect (Figure 6D).



**Figure 5.** ConMot peptide restores the MMR defect of the p.Arg385Pro patient variant. MMR reactions (Supplementary Figure S3) were performed identically as shown in Figure 4C with nuclear extract of MLH1-PMS2-deficient HEK293T cells, supplemented either with extract of cells transfected with either wildtype (WT) MLH1-PMS2 or with the MLH1 variant p.Arg385Pro and PMS2 as detailed in Materials and Methods. ConMot-32-peptide was added in concentrations of 0, 7.5, 18 and 75 μM. For comparison, a control-peptide (H3) of identical length, but with non-specific sequence, was added at a concentration of 100 μM. Several MMR assays were performed; a representative agarose gel is shown (top), and average MMR activity values and standard deviations of three independent experiments were calculated (bottom).

Following the concept of co-evolution, detrimental effects of missense substitutions may potentially be restored by subsequent adaptive substitutions of another residue that is typically in spatial and/or functional contact with the first one (98). Detection of such coupled substitutions in multiple sequence alignments therefore is informative for identifying or confirming direct side chain contacts in three-dimensional structure (99) and is exploited by prediction algorithms as AlphaFold (90).

The expected result of a simultaneous substitution of two independent residues is that a potential defect has at least the same extend as the more deficient single variant. More likely, both may exert an additive effect and result in a more pronounced defect of the combined variant, similar as we have reported before for two small coding variants in MLH1 that conferred an additive defect on protein stability and resulted in a disease phenotype (95). Likewise, in this experiment, the two control variants (p.Arg385Ala and p.Arg385Glu) also reacted in this manner and both additionally aggravated the defect of p.Arg750Tyr.

Therefore, although restauration of repair was incomplete in the p.Arg385Tyr + p.Tyr750Arg double variant, the finding that the second alteration (p.Arg385Tyr) relieved the functional defect of the primary mutation (p.Tyr750Arg) most likely is related to a spatial or functional interaction of both side chains, specifically since both residues were only interchanged. While further investigations are required to confirm and characterize ConMot binding to the MLH1-CTD, this result provides evidence that the interaction of p.Arg385 with p.Tyr750 was correctly predicted by AlphaFold2.

## DISCUSSION

The MutL linker regions have recently attracted increased attention concerning their functional role in the mismatch repair reaction (15,59,76,77). Due to their low degree of sequence and length conservation, human missense alterations are commonly considered innocuous (15). Our investigations demonstrate the existence of a small motif (ConMot) within the human MLH1 linker that is vital for MLH1 function. For a corresponding sequence in yeast *mlh1*, a similar significance has very recently been reported (100). The ConMot motif is exclusively conserved in eukaryotes and comprises 20 amino acids (22 in plants).

### The ConMot sequence and alteration tolerance

The ConMot consists of a central, universally conserved core motif of five amino acids (MVRTD) and of flanking sequence whose conservation is kingdom-specific. While deletion of the ConMot and complete scrambling rendered MLH1 fully deficient in MMR, gross variants of the core motif (core-scramble) and of the flanking sequence (a human chimera containing the flanking sequence of plants) retained approximately 40% catalytic activity. This suggests that its contribution to MMR is not focused on single residues. This is confirmed by the observation that, of all tested missense alterations, only some of the most highly conserved residue Arg385 reduced MMR activity. While three tested substitutions were tolerated (p.Arg385Cys/His/Tyr), three other caused a defect similarly strong as the core-scramble variant (p.Arg385Pro/Ala/Glu). The most pronounced effect was observed for the human patient variant p.Arg385Pro, possibly because the sterically inflexible proline confers a more general strain on target binding by the ConMot than other missense variants and even the core scramble variants. While exact determination of ConMot target binding will be required to explain these results satisfactorily, it seems plausible that ConMot activity primarily depends on its binding to its target, and that this binding has a strongly co-operative nature and is therefore rather tolerant to a number of alterations.

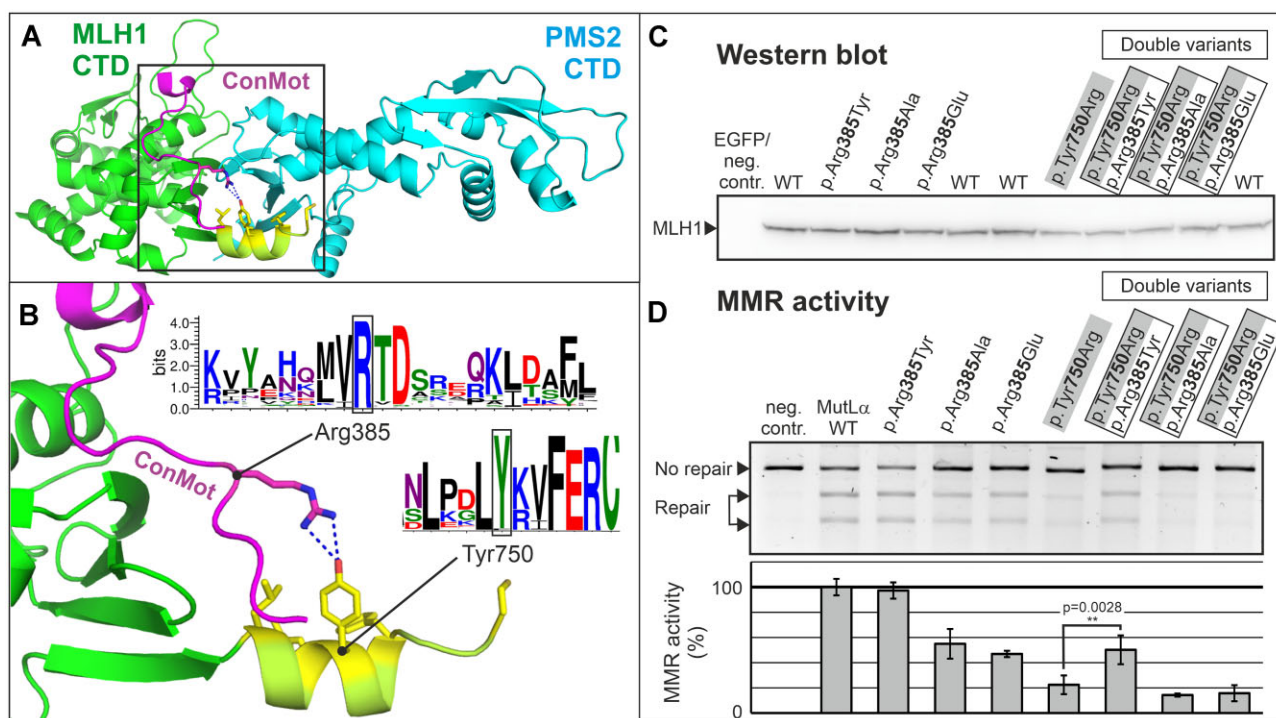
### The potential role of the ConMot in MLH1 function

Since the ConMot is a short motif, it is unlikely to perform any specific activity by its own; rather, it will need to bind a target to perform its function and facilitate MLH1-PMS2 activity. Since the ConMot is situated in the flexible linker, it enjoys some spatial liberty (Figure 7A-a). Our finding that an isolated ConMot peptide can replace the MLH1-ConMot confirms that the exact position of the ConMot within the protein is not of significant importance, suggesting that it moves to its target during the repair process. This is consistent with the observation that the corresponding yeast motif was also active when moved within the yMLH1 linker or even when transferred to the yPMS1 linker (100).

The current results in the peptide complementation experiments allow interesting conclusions concerning the target binding site of the ConMot:

In general, the target may either be another protein or DNA (a third partner, #1, Figure 7A-b), or alternatively, it





**Figure 6.** The MMR defect of p.Tyr750Arg is partly compensated by p.Arg385Tyr. (A) The AlphaFold2 prediction of the MLH1-PMS2 dimer (MLH1: green; PMS2: cyan) suggests that the ConMot (purple) binds to the CTD of MLH1. The prediction suggests an interaction with the C-terminal helix of MLH1 (light green, hydrophobic residues coloured in yellow), which is involved in  $Zn^{2+}$ -binding at the PMS2 endonuclease site. (B) The side chain of the highly conserved Arg385 is predicted to be in proximity to the side chain of the highly conserved Tyr750 residue. (C) MLH1-PMS2-deficient HEK293T cells were co-transfected with the indicated single- and double variants of MLH1 and PMS2. Protein extracts were prepared, separated by SDS-PAGE, blotted and probed for MLH1 protein expression. (D) MMR assays were performed to test DNA mismatch repair activity as detailed in Materials and Methods. One representative agarose gel is shown, and average repair values in comparison to wildtype and standard deviations of >3 independent experiments are shown in a bar diagram. A two-sided *t*-test was performed for testing significance of the difference in repair activity of p.Tyr750Arg in comparison to the double variant p.Arg385Tyr + Tyr750Arg.

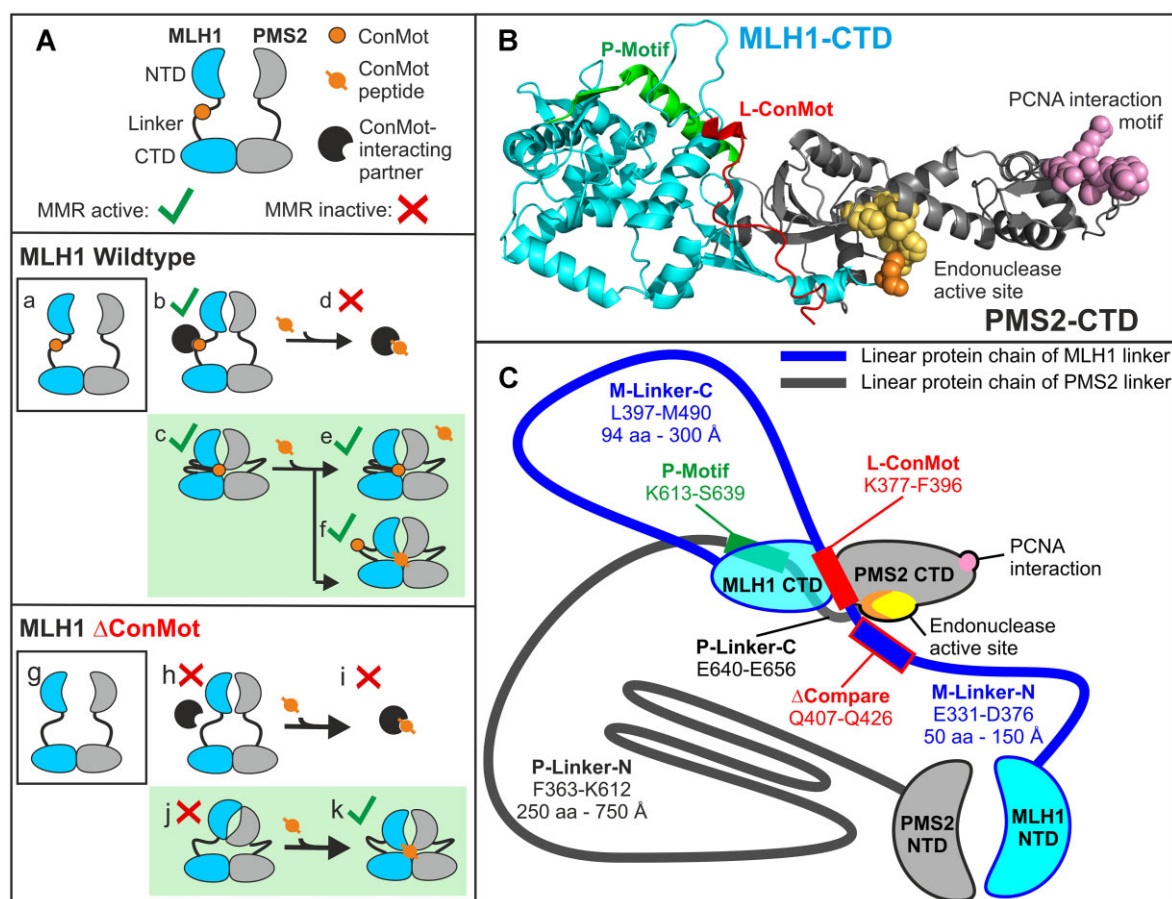
may perform an interaction within the MLH1-PMS2 dimer (#2) (Figure 7A-c).

We have observed that DNA mismatch repair is not attenuated by addition of even high concentrations of ConMot-peptide (Figure 4A). This finding is supportive of an intramolecular interaction (Figure 7A-c), since formation of a ternary complex with another partner (Figure 7A-b) would be likely impaired in the presence of excess ConMot-peptide by out-competing the MLH1-ConMot for binding to this partner and thereby displacing it from the ternary complex (Figure 7A-d). In contrast, if the MLH1-ConMot is involved in an intramolecular interaction (Figure 7A-c), addition of ConMot-peptide in excess is unlikely to interfere with the DNA repair reaction. Either the ConMot-peptide may be excluded from the complex because it cannot compete with the much more efficient intramolecular interaction of the MLH1-ConMot (Figure 7A-e), or it does displace the MLH1-ConMot without interfering with the repair reaction (which is possible, see below) (Figure 7A-f).

The results observed with the ConMot-deletion variant of MLH1 ( $\Delta$ ConMot, Figure 7A-g) also suggest an interaction of the MLH1-ConMot within the MLH1-PMS2 dimer. In these experiments, it was possible to overcome the MMR-deficiency of the MLH1  $\Delta$ ConMot variant by addition of exogenous ConMot-peptide. If the MLH1-ConMot

were involved in a ternary complex formation, the effect of peptide addition (again) would have been the displacement of the ternary partner, without re-constituting repair activity (Figure 7A-hi). In contrast, the finding that addition of ConMot-peptide restores mismatch repair activity suggests that the ConMot-peptide has re-established an intramolecular or inter-subunit interaction within MLH1-PMS2 (Figure 7A-jk).

It is relevant to note that, in contrast to our results, peptide addition in yeast *inhibited* activity of wildtype yMLH1-yPMS1 (100). However, in these experiments, only a fraction of the complete DNA mismatch repair reaction (non-specific endonuclease activation in the absence of a mismatch) was measured under rather artificial conditions (presence of manganese instead of magnesium), making it possible the observed inhibition arises only in this experimental set-up (see Supplementary Table S2 for a detailed comparison of reaction conditions). Additionally, the use of a shorter peptide (25 aa instead of 32 aa) may have contributed to the conflicting results, since our experiments showed that peptide length is a relevant factor in activation (Figure 4B). Since the experiments described here reflected the behavior of the complete MMR reaction under largely biological conditions and consistently demonstrated non-inhibition of the wildtype reaction and re-constitution of repair in ConMot-mutant variants, we assume that this



**Figure 7.** Potential targets and functions of the ConMot motif. (A) Potential effects in competition-/supplementation experiments of MLH1-PMS2 proteins with ConMot peptide dependent on the nature of the ConMot-interaction. The scenarios consistent with the experimental findings are shaded green. The MutL $\alpha$ -heterodimer (MLH1-PMS2) is shown with its constitutively dimerized C-terminal domains (CTD) and the ATP-dependently dimerizing N-terminal domains (NTD), and the unstructured linker region in between (a). The ConMot is located in the MLH1 linker. The wildtype MutL $\alpha$  heterodimer (WT, a) could either use the ConMot for an interaction with a third partner (b), in this case the addition of ConMot peptide is likely to interfere with DNA mismatch repair (d). Alternatively, the ConMot may confer an interaction within the dimer (c), in this case the addition of exogenous ConMot peptide remains without effect (e/f). The MLH1- $\Delta$ ConMot deletion variant (g) cannot support the interaction with a putative third partner (protein or DNA) (h), resulting in a loss of MMR activity, and addition of exogenous ConMot peptide would not re-establish this interaction or MMR activity (i). An interaction of the ConMot within the dimer (j) would also be abrogated, leading to a MMR defect, but addition of ConMot peptide could re-establish this interaction and render the complex MMR-proficient (k). (B) AlphaFold2-prediction of the MLH1-PMS2-dimer: the CTDs and fractions of the linkers binding to the CTD according to this prediction are shown. Residues of the endonuclease active site are shown in spheres (MLH1 Cys756 in orange and PMS2 His703, Glu707, Cys817, Cys 848 and His850 in yellow), residues of the PCNA interaction motif (PMS2 <sup>723</sup>QKLIIP) in pink spheres. The ConMot (red) is predicted to form a short  $\alpha$ -helix and a larger linear fraction and to be bound in full length to the MLH1-CTD close to the dimer interface, with its N-terminal residues touching the last C-terminal MLH1 helix that contributes to the endonuclease active site via Cys756 (orange). The P-motif (green) is composed of a fraction of the PMS2 linker which is predicted to form an  $\alpha$ -helix and a short  $\beta$ -sheet that are bound to the MLH1-CTD. (C) Schematic representation of size relations of the MLH1-PMS2 structured domains and linkers according to the AlphaFold2-prediction. Lengths and sizes are roughly to scale, corresponding to the MLH1 and PMS2 CTDs which cover approximately 50 Å and 65 Å in width, respectively, and the NTDs which are approximately 65 Å in height. The predicted binding of the ConMot and the P-motif to the MLH1-CTD causes the linkers to form two ring structures: a big one (900 aa) is composed of the N-terminal fractions of both MLH1 linker (M-Linker-N) and PMS2 linker (P-Linker-N) and can open and close by dimerization of the NTDs. The second, smaller one is formed by the C-terminal fraction of the MLH1 linker (M-Linker-C). Both ring structures are potentially big enough to accommodate a DNA double helix.

activating effect better reflects the biological function of the ConMot.

A further aspect corroborating an internal interaction of the ConMot is the observation that the AlphaFold2-prediction, which suggested a binding of the ConMot to the MLH1-CTD with high confidence (Supplementary Figure S5) also predicted a functional interaction of the ConMot residue p.Arg385 with the endonuclease residue p.Tyr750, for which we provided experimental evidence (Figure 6).

In summary, the evidence suggests that the ConMot interacts with a site located within the MLH1-PMS2 dimer (Figure 7B). There, it would be in contact with p.Tyr750 in the C-terminal helix of MLH1 which contributes an essential Zn<sup>2+</sup> binding Cys residue (Cys765) to the composite MLH1-PMS2 endonuclease site (46,101). Therefore, the ConMot could potentially modulate endonuclease activity, or even contribute catalytically relevant residues, from this position, consistent with the observed MMR defect and

with the observation that mutation of the corresponding motif in yeast inactivated endonuclease activity (100). Binding of the ConMot to this site would also be consistent with the observation that ConMot residues were found to be close to the DNA backbone in FeBABE experiments (102), since it is simultaneously located close to the DNA in this proximity to the endonuclease site.

Besides a potential involvement in endonuclease activation, the proposed interaction of the ConMot with the MLH1-CTD would subdivide the central cavity of MLH1-PMS2 that is formed by N-terminal dimerization (Figure 7C). A smaller linker ring of approximately 300 Å circumference would be formed simultaneously with a larger ring comprising the residual fraction of the MLH1-CTD combined with both MLH1- and PMS2-NTD and the major part of the PMS2 linker. This may contribute to a controlled and targeted activity of the endonuclease function to the newly synthesized DNA strand whose identity is communicated by PCNA at the replication fork (38). Alternatively, the ConMot may be involved in formation of MutL active polymers that have been observed to occur on DNA by binding neighboring subunits (60).

Taken together, the ConMot is a conserved, movable, short motif located in an IDR, and likely to confer interaction to a structured protein domain. Therefore, it has features attributed to so-called ‘Short Linear Motifs’ (SLiM), which frequently confer weak and/or transient protein interactions (74,84,103,104).

Besides the interaction of the ConMot with the MLH1-CTD, a helical motif in the PMS2 linker is present in our AlphaFold2-predictions and in previous predictions for yeast and other organisms (100). This P-motif is predicted to bind in the MLH1-CTD (Figure 7B and C). This interaction would tether the most C-terminal portion of the PMS2 linker to the structured MLH1 CTD, causing the PMS2 linker to actually be ‘connected’ to the MLH1-CTD instead of the PMS2-CTD (P-Linker-N in Figure 7C). This has plausibly been suggested to evade sterical problems in interactions during the repair process (100), since the PMS2 linker thereby is at greatest possible distance to the PMS2 PCNA interaction motif (64) and the endonuclease domain (48) (Figure 7C).

### Human variants in the ConMot

As yet, pathogenic small coding variants causing Lynch syndrome have almost exclusively been identified either in the conserved NTD or in the CTD, where they can suppress diverse aspects of MLH1 function (105). In the NTD, they may interfere with the ATPase cycle (50,80) and/or with MLH1-MSH2 interaction (83). In the CTD, they may disturb dimerization (51), frequently affect protein stability (79) and/or may impair endonuclease activity (47,85). In contrast, the MLH1 linker region appeared largely devoid of pathogenic small coding variants (15). Our present data demonstrate that the ConMot represents an exception to this rule.

We investigated several human alterations reported within the ConMot. Of these, the p.Arg385Pro alteration displayed a mismatch repair defect. It is located in the core of the strongly conserved MVRTD motif. The substitution to proline likely introduces a sterical strain that

interferes with the ability of the ConMot motif to adopt a proper conformation for target binding. This alteration has originally been identified in a patient with a large adenoma with focal high grade dysplasia and a family history of CRC (family ID R-RM6) (106) (Supplementary Figure S6). Microsatellite instability (MSI) in the tumor tissue is a hallmark of MMR-defective tumors (107), but MSI testing has only been performed in the adenoma tissue of the index patient and therefore was not informative (106). The Arg385Pro allele (rs63750430) has been identified in low frequencies in Asian and European populations ([https://gnomad.broadinstitute.org/variant/3-37067243-G-A?dataset=gnomad\\_r2.1](https://gnomad.broadinstitute.org/variant/3-37067243-G-A?dataset=gnomad_r2.1)) (108). As with many small coding alterations, insufficient information is available for this alteration, therefore it has remained unclear if this variant is causative for cancer diseases. Correspondingly, it is listed as an ‘variant of uncertain significance’ (class 3) in the clinical reference database for variant classifications (<https://www.insight-database.org/classifications/>). The MMR defect of this variant confirms that this variant is likely pathogenic and therefore causative for Lynch syndrome in this family by affecting the proper function of the MLH1 ConMot.

While the sample size is too small for a final evaluation, it is interesting to note that the average age at cancer diagnosis is rather high in the family with the p.Arg385Pro variant. We have shown before that incomplete inactivation of MLH1 by missense alterations prompts a milder Lynch syndrome phenotype with lower penetrance and higher average age of cancer onset (79), and a similar observation has recently been published for an MSH2 variant (109). In case of the ConMot, a complete loss of MMR activity was observed only in the deletion and full scramble variants, while other alterations like the ‘Core-Scramble’ and the ‘Animal-plant hybrid’ variants, still retained some activity (Figures 2 and 3). It is therefore possible that small ConMot alterations retain partial activity entailing a milder Lynch syndrome phenotype.

Intriguingly, the DNA mismatch repair defect of the p.Arg385Pro variant could be overcome by addition of ConMot peptide. To our knowledge, this is the first example of restoration of mismatch repair activity to a MMR-deficient human MLH1 mutant by addition of a small compound.

In contrast to p.Arg385Pro, conservative substitutions of this residue to histidine and cysteine (p.Arg385His and p.Arg385Cys) did not affect catalytic activity. For neither substitution, pathogenicity classifications exist due to a lack of clinical data. It is interesting to note that homologous mutations in yeast of both substitutions displayed defective MMR (100). This may reflect different sensitivity of the motif to substitutions in the yeast and the human system.

The p.Tyr379Cys variant has been clinically classified as ‘likely not pathogenic’ (class 2) based on co-segregation data (<https://www.insight-database.org/classifications/>). Consistent with this, we did not find a catalytic defect for this variant. Although tyrosine 379 is a conserved residue of the ConMot, the substitution is rather conservative, suggesting that the ConMot-target interaction is not significantly disturbed.



Although the p.Val384Asp variant is a non-conservative exchange of a highly conserved, small lipophilic residue, there was no detectable impairment of MMR. Consistent with this experimental result, this alteration is a polymorphism (rs63750447, up to 2.7% of controls) in East Asia, speaking against a causative role in cancer; on the other hand, it is the most prevalent somatic alteration reported in MLH1 in different tumor entities (110). According to our data, the variant is unlikely to be causative for Lynch syndrome; further investigations may show if a causative association with other diseases exists.

In summary, we have described identification and preliminary characterization of a conserved eukaryotic protein motif located in the otherwise unstructured, unconserved linker region of MLH1 proteins. This motif is indispensable for human mismatch repair and most likely performs an intramolecular interaction with the CTD of MLH1 during the repair process. While showing a certain degree of tolerance to substitutions, we have shown that variants in the ConMot can disable mismatch repair and therefore can be causative for Lynch syndrome.

## DATA AVAILABILITY

The data underlying this article are available in the article and in its online supplementary material.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We are grateful to Ricardo M. Biondi for helpful discussions on peptide biochemistry and analytics.

## FUNDING

Deutsche Forschungsgemeinschaft (DFG) [PL688 2/1 to G.P.]. Funding for open access charge: Institutional budget. *Conflict of interest statement*. None declared.

## REFERENCES

- Kunkel, T.A. and Erie, D.A. (2005) DNA mismatch repair. *Annu. Rev. Biochem.*, **74**, 681–710.
- Cohen, S.A., Pritchard, C.C. and Jarvik, G.P. (2019) Lynch syndrome: from screening to diagnosis to treatment in the era of modern molecular oncology. *Annu. Rev. Genomics Hum. Genet.*, **20**, 293–307.
- Palombo, F., Gallinari, P., Iaccarino, I., Lettieri, T., Hughes, M., D'Arrigo, A., Truong, O., Hsuan, J.J. and Jiricny, J. (1995) GTBP, a 160-kilodalton protein essential for mismatch-binding activity in human cells. *Science*, **268**, 1912–1914.
- Palombo, F., Iaccarino, I., Nakajima, E., Ikejima, M., Shimada, T. and Jiricny, J. (1996) hMutSbeta, a heterodimer of hMSH2 and hMSH3, binds to insertion/deletion loops in DNA. *Curr. Biol.*, **6**, 1181–1184.
- Fishel, R. and Wilson, T. (1997) MutS homologs in mammalian cells. *Curr. Opin. Genet. Dev.*, **7**, 105–113.
- Fishel, R. and Kolodner, R.D. (1995) Identification of mismatch repair genes and their role in the development of cancer. *Curr. Opin. Genet. Dev.*, **5**, 382–395.
- Su, S.S., Lahue, R.S., Au, K.G. and Modrich, P. (1988) Mismatch specificity of methyl-directed DNA mismatch correction in vitro. *J. Biol. Chem.*, **263**, 6829–6835.
- Su, S.S. and Modrich, P. (1986) Escherichia coli mutS-encoded protein binds to mismatched DNA base pairs. *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 5057–5061.
- Marsischky, G.T., Filosi, N., Kane, M.F. and Kolodner, R. (1996) Redundancy of Saccharomyces cerevisiae MSH3 and MSH6 in MSH2-dependent mismatch repair. *Genes Dev.*, **10**, 407–420.
- Genschel, J., Littman, S.J., Drummond, J.T. and Modrich, P. (1998) Isolation of MutSbeta from human cells and comparison of the mismatch repair specificities of MutSbeta and MutSalpha. *J. Biol. Chem.*, **273**, 19895–19901.
- Groothuizen, F.S. and Sixma, T.K. (2016) The conserved molecular machinery in DNA mismatch repair enzyme structures. *DNA Repair (Amst.)*, **38**, 14–23.
- Zhang, Y., Yuan, F., Presnell, S.R., Tian, K., Gao, Y., Tomkinson, A.E., Gu, L. and Li, G.M. (2005) Reconstitution of 5'-directed human mismatch repair in a purified system. *Cell*, **122**, 693–705.
- Constantin, N., Dzantiev, L., Kadyrov, F.A. and Modrich, P. (2005) Human mismatch repair: reconstitution of a nick-directed bidirectional reaction. *J. Biol. Chem.*, **280**, 39752–39761.
- Dzantiev, L., Constantin, N., Genschel, J., Iyer, R.R., Burgers, P.M. and Modrich, P. (2004) A defined human system that supports bidirectional mismatch-provoked excision. *Mol. Cell*, **15**, 31–41.
- London, J., Martín-López, J., Yang, I., Liu, J., Lee, J.-B. and Fishel, R. (2021) Linker domain function predicts pathogenic MLH1 missense variants. *Proc. Natl. Acad. Sci. U.S.A.*, **118**, e2019215118.
- Ortega, J., Lee, G.S., Gu, L., Yang, W. and Li, G.-M. (2021) Mismatch-bound human MutS-MutL complex triggers DNA incisions and activates mismatch repair. *Cell Res.*, **31**, 542–553.
- Bradford, K.C., Wilkins, H., Hao, P., Li, Z.M., Wang, B., Burke, D., Wu, D., Smith, A.E., Spaller, L., Du, C. et al. (2020) Dynamic human MutSα-MutLα complexes compact mismatched DNA. *Proc. Natl. Acad. Sci. USA*, **117**, 16302–16312.
- Sixma, T.K. (2001) DNA mismatch repair: MutS structures bound to mismatches. *Curr. Opin. Struct. Biol.*, **11**, 47–52.
- Groothuizen, F.S., Winkler, I., Cristóvão, M., Fish, A., Winterwerp, H.H.K., Reumer, A., Marx, A.D., Hermans, N., Nicholls, R.A., Murshudov, G.N. et al. (2015) MutS/MutL crystal structure reveals that the MutS sliding clamp loads MutL onto DNA. *Elife*, **4**, e06744.
- Qiu, R., DeRocco, V.C., Harris, C., Sharma, A., Hingorani, M.M., Erie, D.A. and Wenginger, K.R. (2012) Large conformational changes in MutS during DNA scanning, mismatch recognition and repair signalling. *EMBO J.*, **31**, 2528–2540.
- Fernandez-Leiro, R., Bhairasing-Kok, D., Kunetsky, V., Laffey, C., Winterwerp, H.H., Groothuizen, F., Fish, A., Lebbink, J.H.G., Friedhoff, P., Sixma, T.K. et al. (2021) The selection process of licensing a DNA mismatch for repair. *Nat. Struct. Mol. Biol.*, **28**, 373–381.
- Gorman, J. and Greene, E.C. (2013) Target search dynamics during post-replicative mismatch repair. *Cell Cycle*, **12**, 537–538.
- Lebbink, J.H., Georgijevic, D., Natrajan, G., Fish, A., Winterwerp, H.H., Sixma, T.K. and Wind, N.d. (2006) Dual role of MutS glutamate 38 in DNA mismatch discrimination and in the authorization of repair. *EMBO J.*, **25**, 409–419.
- Jeong, C., Cho, W.-K., Song, K.-M., Cook, C., Yoon, T.-Y., Ban, C., Fishel, R. and Lee, J.-B. (2011) MutS switches between two fundamentally distinct clamps during mismatch repair. *Nat. Struct. Mol. Biol.*, **18**, 379–385.
- Gradiš, S., Subramanian, D., Wilson, T., Acharya, S., Makhov, A., Griffith, J. and Fishel, R. (1999) hMSH2-hMSH6 forms a hydrolysis-independent sliding clamp on mismatched DNA. *Mol. Cell*, **3**, 255–261.
- Cho, W.-K., Jeong, C., Kim, D., Chang, M., Song, K.-M., Hanne, J., Ban, C., Fishel, R. and Lee, J.-B. (2012) ATP alters the diffusion mechanics of MutS on mismatched DNA. *Structure*, **20**, 1264–1274.
- Gorman, J., Wang, F., Redding, S., Plys, A.J., Fazio, T., Wind, S., Alani, E.E. and Greene, E.C. (2012) Single-molecule imaging reveals target-search mechanisms during DNA mismatch repair. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, E3074–E3083.
- Liu, J., Hanne, J., Britton, B.M., Bennett, J., Kim, D., Lee, J.-B. and Fishel, R. (2016) Cascading MutS and MutL sliding clamps control DNA diffusion to activate mismatch repair. *Nature*, **539**, 583–587.
- Hao, P., LeBlanc, S.J., Case, B.C., Elston, T.C., Hingorani, M.M., Erie, D.A. and Wenginger, K.R. (2020) Recurrent mismatch binding

- by MutS mobile clamps on DNA localizes repair complexes nearby. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 17775–17784.
30. Qiu, R., Sakato, M., Sacho, E.J., Wilkins, H., Zhang, X., Modrich, P., Hingorani, M.M., Erie, D.A. and Weninger, K.R. (2015) MutL traps MutS at a DNA mismatch. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 10914–10919.
  31. Plotz, G., Raedle, J., Brieger, A., Trojan, J. and Zeuzem, S. (2003) N-terminus of hMLH1 confers interaction of hMutLalpha and hMutLbeta with hMutSalpha. *Nucleic Acids Res.*, **31**, 3217–3226.
  32. Plotz, G., Raedle, J., Brieger, A., Trojan, J. and Zeuzem, S. (2002) hMutSalpha forms an ATP-dependent complex with hMutLalpha and hMutLbeta on DNA. *Nucleic Acids Res.*, **30**, 711–718.
  33. Modrich, P. (1989) Methyl-directed DNA mismatch correction. *J. Biol. Chem.*, **264**, 6597–6600.
  34. Clark, A.B., Valle, F., Drotschmann, K., Gary, R.K. and Kunkel, T.A. (2000) Functional interaction of proliferating cell nuclear antigen with MSH2-MSH6 and MSH2-MSH3 complexes. *J. Biol. Chem.*, **275**, 36498–36501.
  35. Flores-Rozas, H., Clark, D. and Kolodner, R.D. (2000) Proliferating cell nuclear antigen and Msh2p-Msh6p interact to form an active mispair recognition complex. *Nat. Genet.*, **26**, 375–378.
  36. 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**, 175–184.
  37. Hombauer, H., Campbell, C.S., Smith, C.E., Desai, A. and Kolodner, R.D. (2011) Visualization of eukaryotic DNA mismatch repair reveals distinct recognition and repair intermediates. *Cell*, **147**, 1040–1053.
  38. Putnam, C.D. (2021) Strand discrimination in DNA mismatch repair. *DNA Repair (Amst.)*, **105**, 103161.
  39. 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**, 37181–37190.
  40. Kadyrova, L.Y. and Kadyrov, F.A. (2016) Endonuclease activities of MutL $\alpha$  and its homologs in DNA mismatch repair. *DNA Repair (Amst.)*, **38**, 42–49.
  41. Kunkel, T.A. and Erie, D.A. (2015) Eukaryotic mismatch repair in relation to DNA replication. *Annu. Rev. Genet.*, **49**, 291–313.
  42. Modrich, P. (2006) Mechanisms in eukaryotic mismatch repair. *J. Biol. Chem.*, **281**, 30305–30309.
  43. Modrich, P. (2016) Mechanisms in *E. coli* and human mismatch repair (nobel lecture). *Angew. Chem. Int. Ed. Engl.*, **55**, 8490–8501.
  44. 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 MutL alpha. *Mol. Cell*, **29**, 112–121.
  45. Yamamoto, T., Iino, H., Kim, K., Kuramitsu, S. and Fukui, K. (2011) Evidence for ATP-dependent structural rearrangement of nuclease catalytic site in DNA mismatch repair endonuclease MutL. *J. Biol. Chem.*, **286**, 42337–42348.
  46. Pillon, M.C., Lorenowicz, J.J., Uckelmann, M., Klocko, A.D., Mitchell, R.R., Chung, Y.S., Modrich, P., Walker, G.C., Simmons, L.A., Friedhoff, P. et al. (2010) Structure of the endonuclease domain of MutL. unlicensed to cut. *Mol. Cell*, **39**, 145–151.
  47. Kadyrov, F.A., Dzantiev, L., Constantin, N. and Modrich, P. (2006) Endonucleolytic function of MutLalpha in human mismatch repair. *Cell*, **126**, 297–308.
  48. Gueneau, E., Dherin, C., Legrand, P., Tellier-Lebegue, C., Gilquin, B., Bonnesoeur, P., Londino, F., Quemener, C., Le Du, M.-H., Márquez, J.A. et al. (2013) Structure of the MutL $\alpha$  C-terminal domain reveals how Mlh1 contributes to Pms1 endonuclease site. *Nat. Struct. Mol. Biol.*, **20**, 461–468.
  49. Schorzman, A.N., Perera, L., Cutalo-Patterson, J.M., Pedersen, L.C., Pedersen, L.G., Kunkel, T.A. and Tomer, K.B. (2011) Modeling of the DNA-binding site of yeast Pms1 by mass spectrometry. *DNA Repair (Amst.)*, **10**, 454–465.
  50. Ban, C. and Yang, W. (1998) Crystal structure and ATPase activity of MutL. implications for DNA repair and mutagenesis. *Cell*, **95**, 541–552.
  51. Kosinski, J., Hinrichsen, I., Bujnicki, J.M., Friedhoff, P. and Plotz, G. (2010) Identification of Lynch syndrome mutations in the MLH1-PMS2 interface that disturb dimerization and mismatch repair. *Hum. Mutat.*, **31**, 975–982.
  52. Guarne, A., Ramon-Maiques, S., Wolff, E.M., Ghirlando, R., Hu, X., Miller, J.H. and Yang, W. (2004) Structure of the MutL C-terminal domain. a model of intact MutL and its roles in mismatch repair. *EMBO J.*, **23**, 4134–4145.
  53. Ban, C., Junop, M. and Yang, W. (1999) Transformation of MutL by ATP binding and hydrolysis. a switch in DNA mismatch repair. *Cell*, **97**, 85–97.
  54. Hall, M.C., Wang, H., Erie, D.A. and Kunkel, T.A. (2001) High affinity cooperative DNA binding by the yeast Mlh1-Pms1 heterodimer. *J. Mol. Biol.*, **312**, 637–647.
  55. Hall, M.C., Shcherbakova, P.V., Fortune, J.M., Borchers, C.H., Dial, J.M., Tomer, K.B. and Kunkel, T.A. (2003) DNA binding by yeast Mlh1 and Pms1. implications for DNA mismatch repair. *Nucleic Acids Res.*, **31**, 2025–2034.
  56. Robertson, A., Pattishall, S.R. and Matson, S.W. (2006) The DNA binding activity of MutL is required for methyl-directed mismatch repair in Escherichia coli. *J. Biol. Chem.*, **281**, 8399–8408.
  57. Borsellini, A., Lebbink, J.H.G. and Lamers, M.H. (2022) MutL binds to 3' resected DNA ends and blocks DNA polymerase access. *Nucleic Acids Res.*, **50**, 6224–6234.
  58. Gorman, J., Plys, A.J., Visnapuu, M.-L., Alani, E. and Greene, E.C. (2010) Visualizing one-dimensional diffusion of eukaryotic DNA repair factors along a chromatin lattice. *Nat. Struct. Mol. Biol.*, **17**, 932–938.
  59. Mardenborough, Y.S.N., Nitsenko, K., Laffey, C., Duboc, C., Sahin, E., Quessada-Vial, A., Winterwerp, H.H.K., Sixma, T.K., Kanaar, R., Friedhoff, P. et al. (2019) The unstructured linker arms of MutL enable GATC site incision beyond roadblocks during initiation of DNA mismatch repair. *Nucleic Acids Res.*, **47**, 11667–11680.
  60. Manhart, C.M., Ni, X., White, M.A., Ortega, J., Surtees, J.A. and Alani, E. (2017) The mismatch repair and meiotic recombination endonuclease Mlh1-Mlh3 is activated by polymer formation and can cleave DNA substrates in trans. *PLoS Biol.*, **15**, e2001164.
  61. Galio, L., Bouquet, C. and Brooks, P. (1999) ATP hydrolysis-dependent formation of a dynamic ternary nucleoprotein complex with MutS and MutL. *Nucleic Acids Res.*, **27**, 2325–2331.
  62. Pillon, M.C., Babu, V.M.P., Randall, J.R., Cai, J., Simmons, L.A., Sutton, M.D. and Guarné, A. (2015) The sliding clamp tethers the endonuclease domain of MutL to DNA. *Nucleic Acids Res.*, **43**, 10746–10759.
  63. Winkler, I., Marx, A.D., Larivière, D., Heinze, R.J., Cristovao, M., Reumer, A., Curth, U., Sixma, T.K. and Friedhoff, P. (2011) Chemical trapping of the dynamic MutS-MutL complex formed in DNA mismatch repair in Escherichia coli. *J. Biol. Chem.*, **286**, 17326–17337.
  64. Genschel, J., Kadyrova, L.Y., Iyer, R.R., Dahal, B.K., Kadyrov, F.A. and Modrich, P. (2017) Interaction of proliferating cell nuclear antigen with PMS2 is required for MutL $\alpha$  activation and function in mismatch repair. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 4930–4935.
  65. Brieger, A., Adryan, B., Wolpert, F., Passmann, S., Zeuzem, S. and Trojan, J. (2010) Cytoskeletal scaffolding proteins interact with Lynch-Syndrome associated mismatch repair protein MLH1. *Proteomics*, **10**, 3343–3355.
  66. Cannavo, E., Gerrits, B., Marra, G., Schlapbach, R. and Jiricny, J. (2007) Characterization of the interactome of the human MutL homologues MLH1, PMS1, and PMS2. *J. Biol. Chem.*, **282**, 2976–2986.
  67. Hidaka, M., Takagi, Y., Takano, T.Y. and Sekiguchi, M. (2005) PCNA-MutSalpha-mediated binding of MutLalpha to replicative DNA with mismatched bases to induce apoptosis in human cells. *Nucleic Acids Res.*, **33**, 5703–5712.
  68. Zhang, H., Richards, B., Wilson, T., Lloyd, M., Cranston, A., Thorburn, A., Fishel, R. and Meuth, M. (1999) Apoptosis induced by overexpression of hMSH2 or hMLH1. *Cancer Res.*, **59**, 3021–3027.
  69. Ackermann, A., Schrecker, C., Bon, D., Friedrichs, N., Bankov, K., Wild, P., Plotz, G., Zeuzem, S., Herrmann, E., Hansmann, M.-L. et al. (2019) Downregulation of SPTAN1 is related to MLH1 deficiency and metastasis in colorectal cancer. *PLoS One*, **14**, e0213411.
  70. Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P., Brown, C.J. and Dunker, A.K. (2003) Predicting intrinsic disorder from amino acid sequence. *Proteins*, **53**(Suppl. 6), 566–572.

71. Oldfield, C.J. and Dunker, A.K. (2014) Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu. Rev. Biochem.*, **83**, 553–584.
72. Ruff, K.M. and Pappu, R.V. (2021) AlphaFold and implications for intrinsically disordered proteins. *J. Mol. Biol.*, 167208.
73. Tompa, P. (2005) The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett.*, **579**, 3346–3354.
74. Tompa, P., Davey, N.E., Gibson, T.J. and Babu, M.M. (2014) A million peptide motifs for the molecular biologist. *Mol. Cell*, **55**, 161–169.
75. Plys, A.J., Rogacheva, M.V., Greene, E.C. and Alani, E. (2012) The unstructured linker arms of Mlh1-Pms1 are important for interactions with DNA during mismatch repair. *J. Mol. Biol.*, **422**, 192–203.
76. Kim, Y., Furman, C.M., Manhart, C.M., Alani, E. and Finkelstein, I.J. (2018) Intrinsically disordered regions regulate both catalytic and non-catalytic activities of the MutL $\alpha$  mismatch repair complex. *Nucleic Acids Res.*, **47**, 1823–1835.
77. Furman, C.M., Wang, T.-Y., Zhao, Q., Yugandhar, K., Yu, H. and Alani, E. (2021) Handcuffing intrinsically disordered regions in Mlh1-Pms1 disrupts mismatch repair. *Nucleic Acids Res.*, **49**, 9327–9341.
78. Trojan, J., Zeuzem, S., Randolph, A., Hemmerle, C., Brieger, A., Raedle, J., Plotz, G., Jiricny, J. and Marra, G. (2002) Functional analysis of hMLH1 variants and HNPCC-related mutations using a human expression system. *Gastroenterology*, **122**, 211–219.
79. Hinrichsen, I., Brieger, A., Trojan, J., Zeuzem, S., Nilbert, M. and Plotz, G. (2013) Expression defect size among unclassified MLH1 variants determines pathogenicity in Lynch syndrome diagnosis. *Clin. Cancer Res.*, **19**, 2432–2441.
80. Raschle, M., Dufner, P., Marra, G. and Jiricny, J. (2002) Mutations within the hMLH1 and hPMS2 subunits of the human MutL $\alpha$  mismatch repair factor affect its ATPase activity, but not its ability to interact with hMutS $\alpha$ . *J. Biol. Chem.*, **277**, 21810–21820.
81. González-Acosta, M., Hinrichsen, I., Fernández, A., Lázaro, C., Pineda, M., Plotz, G. and Capellá, G. (2020) Validation of an in vitro mismatch repair assay used in the functional characterization of mismatch repair variants. *J. Mol. Diagn.*, **22**, 376–385.
82. Wang, H.X. and Hays, J.B. (2002) Mismatch repair in human nuclear extracts - Quantitative analyses of excision of nicked circular mismatched DNA substrates, constructed by a new technique employing synthetic oligonucleotides. *J. Biol. Chem.*, **277**, 26136–26142.
83. Plotz, G., Welsch, C., Giron-Monzon, L., Friedhoff, P., Albrecht, M., Piiper, A., Biondi, R.M., Lengauer, T., Zeuzem, S. and Raedle, J. (2006) Mutations in the MutS $\alpha$  interaction interface of MLH1 can abolish DNA mismatch repair. *Nucleic Acids Res.*, **34**, 6574–6586.
84. Gibson, T.J., Dinkel, H., van Roey, K. and Diella, F. (2015) Experimental detection of short regulatory motifs in eukaryotic proteins: tips for good practice as well as for bad. *Cell Commun. Signal.*, **13**, 42.
85. Smith, C.E., Mendillo, M.L., Bowen, N., Hombauer, H., Campbell, C.S., Desai, A., Putnam, C.D. and Kolodner, R.D. (2013) Dominant mutations in *S. cerevisiae* PMS1 identify the Mlh1-Pms1 endonuclease active site and an exonuclease 1-independent mismatch repair pathway. *PLoS Genet.*, **9**, e1003869.
86. Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M. and Barton, G.J. (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**, 1189–1191.
87. Edgar, R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinf.*, **5**, 113.
88. Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T. and Ben-Tal, N. (2016) ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.*, **44**, W344–W50.
89. Crooks, G.E., Hon, G., Chandonia, J.M. and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.*, **14**, 1188–1190.
90. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A. *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583–589.
91. Yu, D., Chojnowski, G., Rosenthal, M. and Kosinski, J. (2023) AlphaPullDown—a python package for protein-protein interaction screens using AlphaFold-Multimer. *Bioinformatics*, **39**, btac749.
92. Shen, Y., Maupetit, J., Derreumaux, P. and Tufféry, P. (2014) Improved PEP-FOLD approach for peptide and miniprotein structure prediction. *J. Chem. Theory Comput.*, **10**, 4745–4758.
93. Xue, B., Dunbrack, R.L., Williams, R.W., Dunker, A.K. and Uversky, V.N. (2010) POND-R-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim. Biophys. Acta*, **1804**, 996–1010.
94. Fokkema, I.F.A.C., Taschner, P.E.M., Schaafsma, G.C.P., Celli, J., Laros, J.F.J. and Dunnen, J.T.d. (2011) LOVD v.2.0: the next generation in gene variant databases. *Hum. Mutat.*, **32**, 557–563.
95. Hinrichsen, I., Schäfer, D., Langer, D., Köger, N., Wittmann, M., Aretz, S., Steinke, V., Holzapfel, S., Trojan, J., König, R. *et al.* (2015) Functional testing strategy for coding genetic variants of unclear significance in MLH1 in Lynch syndrome diagnosis. *Carcinogenesis*, **36**, 202–211.
96. Chang, D.K., Ricciardiello, L., Goel, A., Chang, C.L. and Boland, C.R. (2000) Steady-state regulation of the human DNA mismatch repair system. *J. Biol. Chem.*, **275**, 18424–18431.
97. Neduva, V. and Russell, R.B. (2005) Linear motifs: evolutionary interaction switches. *FEBS Lett.*, **579**, 3342–3345.
98. Juan, D., Pazos, F. and Valencia, A. (2013) Emerging methods in protein co-evolution. *Nat. Rev. Genet.*, **14**, 249–261.
99. Morcos, F., Pagnani, A., Lunt, B., Bertolino, A., Marks, D.S., Sander, C., Zecchina, R., Onuchic, J.N., Hwa, T. and Weigt, M. (2011) Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, E1293–E301.
100. Torres, K.A., Calil, F.A., Zhou, A.L., DuPrie, M.L., Putnam, C.D. and Kolodner, R.D. (2022) The unstructured linker of Mlh1 contains a motif required for endonuclease function which is mutated in cancers. *Proc. Natl. Acad. Sci. U.S.A.*, **119**, e2212870119.
101. Kosinski, J., Plotz, G., Guarne, A., Bujnicki, J.M. and Friedhoff, P. (2008) The PMS2 subunit of human MutL $\alpha$  contains a metal ion binding domain of the iron-dependent repressor protein family. *J. Mol. Biol.*, **382**, 610–627.
102. Claeys Bouuaert, C. and Keeney, S. (2017) Distinct DNA-binding surfaces in the ATPase and linker domains of MutL $\gamma$  determine its substrate specificities and exert separable functions in meiotic recombination and mismatch repair. *PLoS Genet.*, **13**, e1006722.
103. van Roey, K., Uyar, B., Weatheritt, R.J., Dinkel, H., Seiler, M., Budd, A., Gibson, T.J. and Davey, N.E. (2014) Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem. Rev.*, **114**, 6733–6778.
104. Kumar, M., Michael, S., Alvarado-Valverde, J., Mészáros, B., Sámano-Sánchez, H., Zeke, A., Dobson, L., Lazar, T., Örd, M., Nagpal, A. *et al.* (2021) The Eukaryotic Linear Motif resource: 2022 release. *Nucleic Acids Res.*, **50**, D497–D508.
105. Raevaara, T.E., Korhonen, M.K., Lohi, H., Hampel, H., Lynch, E., Lonnqvist, K.E., Holinski-Feder, E., Sutter, C., McKinnon, W., Duraisamy, S. *et al.* (2005) Functional significance and clinical phenotype of nontruncating mismatch repair variants of MLH1. *Gastroenterology*, **129**, 537–549.
106. Genuardi, M., Carrara, S., Anti, M., Ponz de Leòn, M. and Viel, A. (1999) Assessment of pathogenicity criteria for constitutional missense mutations of the hereditary nonpolyposis colorectal cancer genes MLH1 and MSH2. *Eur. J. Hum. Genet.*, **7**, 778–782.
107. Ionov, Y., Peinado, M.A., Malkhosyan, S., Shibata, D. and Perucho, M. (1993) Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature*, **363**, 558–561.
108. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P. *et al.* (2020) The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, **581**, 434–443.
109. Bruekner, S.R., Pieters, W., Fish, A., Liaci, A.M., Scheffers, S., Rayner, E., Kaldenbach, D., Drost, L., Dekker, M., van Hees-Stuivenberg, S. *et al.* (2023) Unexpected moves: a conformational change in MutS $\alpha$  enables high-affinity DNA mismatch binding. *Nucleic Acids Res.*, **51**, 1173–1188.
110. Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H., Cole, C.G., Creatore, C., Dawson, E. *et al.* (2019) COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.*, **47**, D941–D947.