STUDYING THE ACTIVATION STATES OF MUTS AND MUTL IN DNA MISMATCH REPAIR

STUDYING THE ACTIVATION STATES OF MUTS AND MUTL IN DNA MISMATCH REPAIR

DORETH BHAIROSING-KOK

DORETH BHAIROSING-KOK

NSL

551500-L-os-Bhairosing

Studying the Activation States of MutS and MutL in DNA Mismatch Repair

Het bestuderen van de activatietoestanden van MutS en MutL in DNA mismatch herstel

Doreth Bhairosing-Kok

551500-L-bw-Bhairosing
Processed on: 7-12-2020

ISBN: 9789464190946

Publisher: Het Nederlands Kanker Instituut – Antoni van Leeuwenhoek Ziekenhuis
Cover & Lay-out: Eduard Boxem, persoonlijkproefschrift.nl
Printing: Gildeprint Enschede, gildeprint.nl

This thesis was printed with financial support from Erasmus University Rotterdam and the Netherlands Cancer Institute

Studying the Activation States of MutS and MutL in DNA Mismatch Repair

Het bestuderen van de activatietoestanden van MutS en MutL in DNA mismatch herstel

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof.dr. F.A. van der Duijn Schouten

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
vrijdag 29 januari om 13.30 uur
door

Doreth Bhairosing-Kok
geboren te Stede Broec

Erasmus University Rotterdam

Promotiecommissie

Promotor	Prof. dr. T.K. Sixma
Overige leden	Dr. M.H. Lamers Prof. dr. H. te Riele Dr. Ir. J.H.G. Lebbink Prof. dr. A. Perrakis

TABLE OF CONTENTS

CHAPTER 1

Introduction

THE ENCYCLOPEDIA OF THE CELL

Our body is composed of many cells and each individual cell contains DNA. The DNA can be seen as the complete encyclopedia for the cell to know how to function and how to survive. For this reason, it is important that a cell contains DNA. For each process, different parts of DNA can be essential. Nevertheless, each cell contains the full set DNA. This might sound inefficient, but it actually makes a lot of sense knowing that all cells in our body arise from one zygote, a fertilized ovum. Containing exactly one copy of the DNA; a combination of DNA of the mother and the father, merged into one encyclopedia. A full encyclopedia has chapters. For the cell we know these chapters as chromosomes.

DNA is important for every cell, but what does it look like? In 1953, James Watson and Francis Crick revealed the structure of DNA [1]. This structure showed that DNA is composed of two chains that we now know as polynucleotides. These two chains form a double helix. The two polynucleotides are a repetition of four building blocks, called nucleotides each containing a specific base; Cytosine (C), Guanine (G), Adenine (A) and Thymine (T), often referred to by only their first letter. These nucleotides always pair in the opposite strands (polynucleotides) in the same manner; C pairs with G and A pairs with T. Therefore, one knows the order of nucleotides in the complementary polynucleotide once the sequence of the other polynucleotide is known. So, in brief, DNA is an encyclopedia with one long sequence of C, G, T and A letters, divided into books, called chromosomes.

DNA replication & DNA mismatch repair

During cell division, a mother cell splits into two daughter cells with each the exact same content as the mother cell. Therefore, the DNA needs be duplicated, so that each cell contains a copy. The cell nicely uses the complementary nature of the DNA helix. The duplication of DNA is called DNA replication and this is performed by the replication machinery. This machine figuratively zips open the double helix and inserts the matching nucleotides on the complementary sides of each of the single stranded polynucleotides. This results in two new double stranded DNA helices, each composed of two complementary polynucleotides. This is a very accurate process, given the fact that human DNA is composed of 3.2 billion nucleotides (single stranded) [2, 3].

One major player of the replication machinery is the polymerase. This enzyme is responsible to insert the correct nucleotide at the correct position, opposite to the other nucleotide. Given the size of the whole genome, it is known and understandable that polymerases occasionally make mistakes and insert an incorrect nucleotide. This happens in every 1 in 100-1000 nucleotides, resulting in an error in the DNA code, also known as a mismatch [4]. Luckily, many polymerases have an error-correcting ability called proofreading [5]. When an incorrect nucleotide is inserted, for example a G opposite to a T, the replication machinery excises the nucleotide in the newly synthesized polynucleotide by going in reverse direction and removing the nucleotide and replacing it with the correct nucleotide.

Proofreading is a very important feature of polymerases since it minimizes the error rate during DNA replication to 1 in 10^{-4} to 10^{-5} [4]. However, polymerases are, like humans, not perfect. Sometimes one or multiple incorrectly incorporated nucleotides escape the proofreading ability of the polymerase and this results in a mismatch in the new DNA sequence of the daughter cell. Whenever this cell will divide into two new daughter cells, this mismatch will be present in each cell and every other generation of cells that arise from this. One can imagine the consequences.

In line with this, DNA errors also occur due to polymerase strand slippage (or replication slippage) [6-8]. During replication, the polymerase can face a secondary element and/or repeat. It will stall, suspends replication and detaches from the template strand. The newly synthesized polynucleotide dissociates and pairs somewhere upstream in the same template strand. This piece is the starting point for the polymerase. Instead of continuing normal replication, it also backtracks and inserts the nucleotides again, resulting in an expansion of the region which is undesirable.

Fortunately, most organisms have another back-up system for repairing mismatches and errors made by polymerases; DNA Mismatch Repair (MMR). This mechanism, as the name implies, repairs mismatches but also insertions and deletions [9] that arise during DNA replication [9]. As a result, there will be limited mismatches or IDLs present in the DNA of daughter cells.

MMR is essential for the cell to cope with the mismatches that have escaped proofreading by the polymerase. These mismatches can arise in every piece of DNA. DNA is also divided into genes and intergenic regions. Human DNA contains roughly about 20.000 to 25.000 genes [3]. Genes are coding for proteins. So, the presence of a mismatch in a protein-coding gene can have a different outcome than an error in the non-protein coding genes. Errors or mismatches cause mutations and in protein coding genes these can lead to altered activity of that particular protein. More specific, mutations in proto-oncogenes lead to a distortion in cell growth and proliferation, which often leads to tumor development. Non-functional MMR is therefore often linked to a high mutation rate. MMR deficiency in human is caused by germline mutations in MMR genes. These patients suffer from Lynch syndrome (LS), also known as hereditary non-polyposis colon cancer (HNPCC) [10, 11]

MMR deficiency – Lynch Syndrome

Lynch syndrome (LS) is a genetic disease that is characterized by a germline mutation in one of the MMR genes; *MSH2*, *MSH6*, *MLH1* and *PMS2* [10, 11]. LS patients have an increased risk to develop colorectal cancer (CRC) and other cancers [12]. When a secondary acquired (non-germline) mutation in the other allele arises, loss of function of this particular gene leads to a deficiency in MMR. Mutations in MSH2 or MSH6 lead to misfunction of the protein MutSα. This protein detects mismatches and interacts with the replication machinery. MLH1 and PMS2 are coding for the MutLα protein, this protein interacts with MutSα to repair the mismatch. Mutations in MLH1 and MSH2 are most common for LS patients, followed by MSH6 and rarely in PMS2 [11, 12].

In general, mutations appear more frequent in repetitive DNA sequence (microsatellites) than random sequences, due to polymerase strand slippage [8, 9, 13, 14]. Loss of MMR leads to many mutations in microsatellite regions which causes microsatellite instability (MSI). MSI is a molecular hallmark for LS tumors but also for non-LS-CRC.

Mutations in protein-coding genes cause frame shift peptides (FSP). These FSPs act as novel antigens and therefore elicit immune responses. Cells cope with this by attacking these neoantigens with cytotoxic T-cells, resulting in elimination of the antigen. The mutation rate in LS-CRC patients is higher than for non-LS-

CRC patients, due to the hypermutability of microsatellites [12]. Therefore, the frequency of neoantigens is also higher for LS-CRC patients.

Recent developments have shown that immunotherapy is very effective for treating various cancers. Several types of immunotherapy exist, such as T-cell transfer therapy and immune checkpoint inhibitors such as the PD-1 inhibitor. Targeting the PD-1 checkpoint has been shown to be very promising for LS-CRC patients and other MMR-deficient cancers that show a high MSI load [15-18]. Cancer cells express PD-L1 on the surface which is needed for cell survival. It hacks the T cell scanning system, by interacting with PD-1 receptor and thereby inactivating the cytotoxic T cell, which in turn prevents the cancer cell from being killed. A PD-1 checkpoint inhibitor, Pembrolizumab, prevents this interaction and is shown to be an effective treatment for high-MSI tumors [17]. As patients with this cancer hallmark benefit especially from this treatment, proper screening and diagnosis is important to ensure effective treatment.

DNA Mismatch repair

Overview

MMR corrects mismatches and IDLs that arise during DNA replication [19-21]. This mechanism is conserved from prokaryotes to higher eukaryotes. MMR is initiated by a protein called MutS (Fig 1). MutS is an oval-shaped protein that can encircle DNA. During DNA replication, MutS interacts with the replication machine and checks the DNA if any mismatches, insertions or deletions were integrated. Once one of these errors is detected by MutS, the DNA kinks at the position of the error. As a result, MutS binds ATP and undergoes a series of ATP-induced conformational changes while it is still bound to the DNA [22, 23]. This allows MutS to adopt a defined sliding clamp conformation. As the name implies, sliding clamp MutS can slide on the DNA, away from the mismatch (Fig 1).

Then a second protein can be recruited; MutL. MutL will be loaded onto the DNA by MutS, only when MutS is in the sliding clamp conformation [23-26] (Fig 1). Until this point, the initiation of MMR is similar for all known species that have MMR. In a subset of Gram-negative prokaryotes, like *Escherichia coli (E.coli)*, a third protein is then recruited; MutH. This protein is an endonuclease and it nicks the newly synthesized strand at a specific sequence [27]. In other species, such as eukaryotes, this task is performed by the endonuclease function MutL, MutLα.

Nevertheless, the result is exactly the same; the nick functions as a starting point to unwind the DNA by the helicase UvrD. This unwinding allows exonucleases to remove the newly synthesized strand. Next, the polymerase will resynthesize the new strand and the ligase will seal the nick [20] (Fig 1).

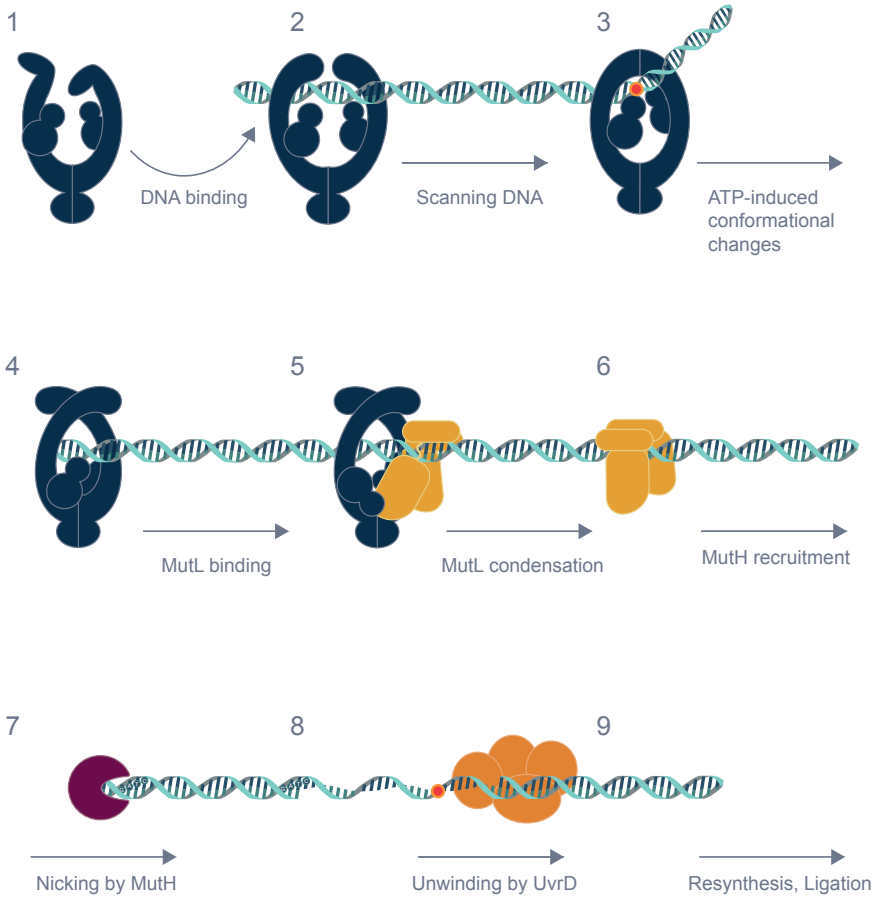


Figure 1. Initiation of *E.coli* DNA mismatch repair. MutS can adopt several conformations in its apo-state, including kinking the lever domains (1). This state can proceed towards DNA binding, and MutS will scan the DNA (2) in a more open conformation. Upon mismatch detection (3), the DNA is bent and the mismatch binding domains are interacting with the mismatch via monomer A sticking the Phe36 on the mismatched base. Then, upon ATP binding, MutS undergoes a conformational change towards a transition state where the mismatch binding domain of monomer A still interacts with the DNA but monomer B not (4). Next, MutS adapts a sliding clamp conformation and loads MutL on DNA (5), followed by release of MutL (6). MutL will recruit endonuclease MutH (7), MutH will nick hemimethylated GATC sites which will be an entry point for UvrD to unwind the DNA (8). Afterwards, DNA will be resynthesized and ligated (9).

MMR proteins in other processes

MutS and MutL can also interact with the replication machinery; the β-clamp in prokaryotes and PCNA in eukaryotes [28-33]. These interactions play an important role in MMR for strand discrimination in bacteria lacking MutH and eukaryotes. In addition to these roles in post-replicative MMR, the eukaryotic MMR proteins also play essential roles in apoptosis, meiotic and mitotic recombination. It also has roles in cell-type-specific processes such as somatic hypermutation and triplet-repeat extension [21]. During recombination, MMR proteins prevent strand exchange between homeologous sequences. As a result, a defect in MMR leads to an increase in homeologous recombination [34, 35].

Besides the protecting roles of MMR, the presence of MMR in cells can also have a mutagenic effect, especially during triplet repeat expansion. An increase of triplet sequences occurs during DNA replication due to replication slippage. MMR will try to repair these damaged trinucleotide repeat sequences, and this increases the change for replication slippage and therefore trinucleotide repeat expansion [20, 36, 37]. The number of trinucleotide repeats is related to the progression of trinucleotide repeat disorders, including Huntington's disease and myotonic dystrophy [38, 39]. Finally, MMR is also involved in promoting somatic hypermutation during the formation of immunoglobulins in B lymphocytes [40-42]

Because of the conservation in MMR initiation between prokaryotes and eukaryotes, MMR is often studied using proteins from model organisms such as *Thermus aquaticus* (*Taq*) and *E.coli*. The work in this thesis is all performed on the *E.coli* system. However, when indicated, results of studies in other species have been used for discussion.

MMR proteins

MutS

MutS is the initiator of MMR. It is an ATPase and belongs to the ATP binding cassette (ABC) family of ATPases [43]. Other known members of this family are Rad50 and structural maintenance of chromosomes (SMC) proteins, and ABC membrane transporters [43-47]. Like other ABC ATPases, MutS acts as a dimer although it can form tetramers at high concentration which is important for the

recombination functions of MutS [48], but not for MMR [49, 50]. In contrast to prokaryotes, tetramerization has not been observed for eukaryotic homologs.

Crystal structures have elucidated the shape of the MutS dimer bound to DNA that contains a mismatch. The first structures solved were homodimeric *Taq* and *E.coli* MutS [51, 52] (Fig 2). Afterwards, many structures followed of *E.coli* and different species, showing a high similarity among MutS homologs [23, 50, 52-58] (Chapter 2 & 3). Human MutS homologs are heterodimers; MutSα is composed of MutS Homolog 2 (Msh2) and Msh6 and MutSβ is formed by Msh2 and Msh3. Each heterodimer has its unique features, with the main difference that MutSa binds single base mismatches and small IDLs while MutSβ interacts with larger IDL up to 13 base pairs long [59].

For all known MutS homologs, two monomers are encircling the DNA by forming an asymmetric dimer. Each MutS monomer is composed of six domains: 1) Mismatch binding, 2) Connector, 3) Core & Lever, 4) Clamp, 5) ATPase and 6) Tetramerization (Fig 2) [51], with the clamp domain on the 'top' and the tetramerization domain at the 'bottom'. The tetramerization domain was not present in the first *E.coli* MutS crystal structure, but was characterized later [49, 50]. MutSα has an additional N-terminal disordered domain in Msh6 that is predicted to be unstructured and has not been structurally solved so far.

Each domain has an assigned function. The mismatch binding domain of monomer A interacts specifically with the mismatch, via stacking its Phe36 on the mismatched base. The equivalent domain in monomer B does not interact directly with the mismatch, but has weak interactions with the DNA backbone [51] (Chapter 3). The Phe36 residue in monomer A, together with two negatively charged residues Glu38 and Asp35, plays a crucial role in mismatch recognition and positioning the mismatch binding domain [51, 60]. Glu38 forms hydrogen bonds to one of the mismatched bases and together with the negatively charged carboxylate group, it discriminates between matched and mismatched DNA [60-62]. Mutant studies show that this hydrogen bond initiates intramolecular signaling which results in inhibition of ATP hydrolysis.

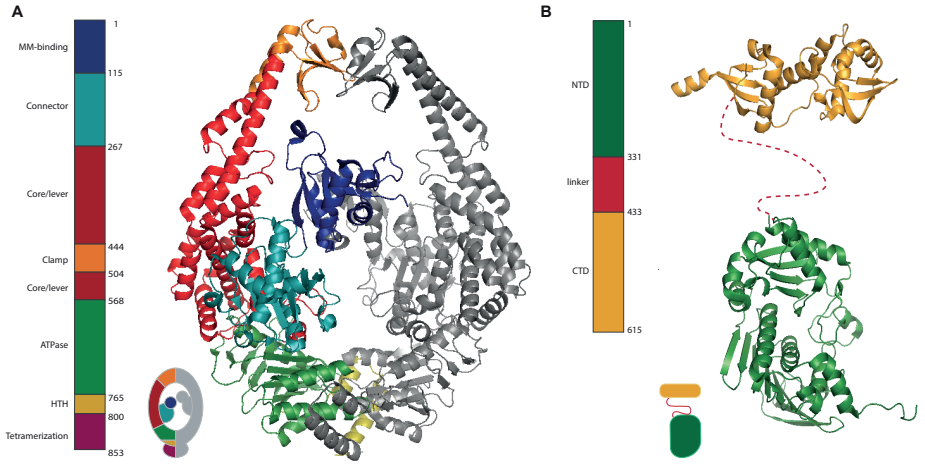


Figure 2. Domains of MutS and MutL. A) MutS in mismatch bound conformation (DNA not shown) (PDB ID: 1E3m). All domains are colored as in the schematic representation (tetramerization domain not shown). **B)** MutL N-terminal domain and C-terminal domain with the unstructured (flexible) linker (PDB ID: 1NHJ (NTD) & 1X9Z (CTD)). Domains and linker are colored as in the schematic representation.

The connector domain links the mismatch binding domain to the core of MutS. MutS undergoes many conformational changes upon MMR initiation. During these rearrangements, the connector domain adopts different orientations which are essential for proceeding towards next steps in MMR initiation, including recruitment of MutL [23] (Chapter 3). The core and lever domains are forming the 'backbone' of the dimer. The lever consists of two a-helices that are forming a coiled-coil. Although rather unusual for a coiled-coil, MutS is able to kink the lever domain at a defined hinge point within the coiled coil. This allows MutS to rotate the top domains both inward and outward at a range of angles. This plays an essential role for DNA binding and conformational changes after mismatch recognition [57] (Chapter 2). The clamps are forming the top part of the MutS where it interacts with the DNA in certain conformations along the MMR cascade. The clamp domains are interacting with each other during scanning and detection of the mismatch, but this interaction is lost after mismatch recognition.

Lastly, the bottom part of the monomer is composed of the C-terminal ATPase domain, the Helix-Turn-Helix (HTH) domain and the tetramerization domain. MutS is able to bind ATP and ADP in the nucleotide-binding pocket, which is formed

by both subunits in the dimer interface. The nucleotide-binding state of MutS is related to its mode of action. Since MutS has two nucleotide binding sites, they can be occupied both symmetric and asymmetric [63-68]. This makes the MutS dimer an asymmetric homodimer. Various occupancies are possible depending on the presence and timing of mismatched DNA and ATP [66, 69]. Prior to mismatch binding, there is already asymmetry in MutS since one monomer is bound to ADP and one is not. In addition, when MutS binds mismatched DNA, more asymmetry is induced since one of the monomers interacts with DNA directly and only this monomer is nucleotide-bound.

As mentioned, MutS belongs to the family of ABC ATPases. Besides the asymmetry that we see in MutS, asymmetry in ATP binding sites is found in many members of this family [70-74]. One of the family members is condensin, a conserved SMC complex composed of SMC2 and SMC4. Condensin harbors two ATPase sites. ATP is buried between the signature motif of one SMC unit and the Walker A and B motifs of the other [75, 76]. Recently, it was shown that condensin's ATPase has a dual role in chromosome condensation. Mutating one ATPase site impairs condensation, while mutation of the ATPase of the second monomer results in the opposite effect, namely hyperactive condensing that compacts DNA faster than wild type [74]. These alternating ATPase domains are also observed for MutS [65], where there are different affinities for nucleotides per ATPase domain and the occupancy of each domain has a different effect in MMR.

MutL

MutL is the second activated protein in the MMR machinery. MutL is known as a 'molecular matchmaker'. It is recruited by MutS after mismatch detection and it also regulates the recruitment of the downstream players of MMR [23, 77, 78]. MutL is also a homodimer and is a weak ATPase belonging to the GHKL (*G*yrase, *H*SP90, histidine *K*inase, Mut*L*) superfamily. Family members share the nucleotide binding Bergerat-fold [79]. Almost all members of the GHKL family form a dimer and use ATP binding and hydrolysis to induce large conformational changes due to transient dimerization of the ATPase domains [80].

MutL is composed of two rigid domains; an N-terminal domain (NTD) and a C-terminal domain (CTD), also known as LN40 (NTD) and LC20 (CTD) (Fig 2). These domains are linked via a flexible linker which has various length among homologs. The CTD forms the primary dimer interface and the ATP-binding sites

are located in the NTD. In addition to the CTD, the NTD domains can also dimerize under defined conditions, resulting in a closed ring-shaped MutL dimer. The NTD is highly conserved from *E.coli* to humans [81, 82], and also the CTD is structurally conserved for most MutL homologs despite the poor sequence conservation [83-85].

Likely due to the flexible and unstructured linker, no full-length structure at high resolution is solved. Attempts have been done and are ongoing, these results are being discussed in Chapter 4. Nevertheless, crystal structures of NTD and CTD of various species have been solved [23, 32, 82-84, 86-89]. This, together with other data, showed us that MutL can adopt different conformations depending on the nucleotide state of MutL. Size exclusion chromatography, atomic force microscopy (AFM) and crystal structures show that upon ATP binding MutL adopts various conformations [83, 86, 90-92].

ADP-bound MutL is dimerized via the CTDs, but with open and flexible NTDs. ATP binding by the NTD stimulates dimerization and compaction. Given the fact that MutL is a weak ATPase, it is thought that MutL goes through a cycle of dimerization and then becomes more compact, which can also occur in a two-step mechanism with having one NTD compact with the CTDs while the second NTD is still flexible. Just like MutS, MutL can also be an asymmetric homodimer with different nucleotides occupied in the ATP-binding sites. Upon ATP hydrolysis, the compaction is lost and the NTDs are not interacting anymore resulting in an open MutL conformation [80]. We can state that MutL is in an equilibrium between open, semi-condensed and condensed, depending on the concentration and type of nucleotide [92], and the presence of MutS and (mismatched) DNA.

MMR initiation in detail

Scanning and detecting mismatches

Prior to DNA loading, MutS can adopt several conformations in its apo-state [57, 58] (Chapter 2) (Fig 3). The crystal structure of *E.coli* MutS in its apo-state showed bent lever domains, this in contrast with the straight conformation when MutS interacts with DNA (Chapter 2 & 3). In the absence of DNA, MutS binds one ADP with high affinity and the second ADP with low affinity [69]. The ATPase domains of MutS can be occupied by ADP or ATP, or be empty. All combinations between these three options are possible, although the hydrolyzing conditions *in vivo* result

in a MutS dimer that is mostly occupied with one ADP molecule with the second site being empty or bound to ADP or ATP.

Upon DNA binding, MutS scans for mismatches. Mismatches are rare so MutS needs an efficient mechanism to do this; using a minimal amount of energy but not missing any distortion in the DNA sequence and conformation. A cryo-EM map of MutS scanning DNA showed that the clamp domains embrace the DNA and the mismatch binding domains interact with DNA from the other side (Chapter 3). During processing of this cryo-EM data, several 3D classes showed that MutS could open up to various extents. Together with the observed bending in DNA, we can assume that MutS actively tries to distort DNA during scanning. This means that MutS tries to sense the mismatches with the clamp domains, making the scanning process rather efficient.

When the distortion is found by MutS, several conformational changes take place both for DNA and MutS. Because MutS tries to find weaknesses in the DNA helix, it can easily induce a bend in the DNA at the mismatch location. As a result, DNA is bent at a 60° angle and MutS interacts with its mismatch binding domain and stacks a conserved phenylalanine (Phe36) on top of the mismatch [51, 52, 93]. Besides the interaction of Phe36 with the mismatched base, Glu38 forms hydrogen-bonds with the same base. This interaction helps MutS to discriminate between matched and mismatched bases. Moreover, Glu38 authorizes downstream repair signaling by inducing intramolecular signaling which leads to inhibition of ATP hydrolysis bound by MutS. In theory, every mismatch can be induced by the polymerase, however, some mismatches occur more often than others [4]. Nevertheless, a range of mismatches was tested and all mismatches showed a common recognition mode by MutS, except for a C:C mismatch [53]. In addition, the type of mismatch is not correlated with the ability to undergo the conformational change towards a sliding clamp [50].

Next to mismatches, DNA ends can also be interpreted by MutS as a DNA lesion. It was shown that both *E.coli* and *T. aquaticus* MutS have a strong affinity for DNA ends compared to homoduplex DNA [26, 94], and this was also shown by AFM [95]. Likely, this high affinity for DNA ends is because MutS recognize this end as a DNA distortion and can undergo a conformational change.

Figure 3. Conformations of MutS. From left to right; Apo state (PDB ID: 6I5F), scanning state (PDB ID: 7AIF), MM bound (PDB ID: 1E3M and 7AI6), transition state (PDB ID: 7AI7) and Sliding clamp bound to MutL$^{\Delta N40}$ (PDB ID: 7AIB).

Transition from mismatch bound towards sliding clamp

After the interaction with the mismatched base, MutS undergoes an ATP-induced conformational change towards a state where MutS can only release DNA at its ends, forming a sliding clamp [23, 96, 97]. This sliding clamp is necessary to allow MutL binding [23, 26, 98]. In this state, the DNA is pushed downwards and the clamp and lever domains are crossing over each other and over the DNA. Although this conformation is well known and validated, the transition towards this conformation was not fully understood. There are indications that MutS transforms from mismatch-bound to sliding clamp via an intermediate step, indicating the existence of a transient intermediate state. This was studied by Förster Resonance Energy Transfer (FRET) experiments between MutS and open/closed DNA. It was found that if MutS, after mismatch recognition, binds both ADP and ATP, it undergoes a conformational change leading to the 'post-recognition sliding clamp state'. This state is characterized by a medium FRET between Domains I (mismatch binding domain), which means that these domains are partially open [99]. In Chapter 3, we show a cryo-EM map of a state that may represent the transition from the mismatch bound towards the clamp state of MutS (Chapter 3) (Fig 3). The transition state resembles the MutL-bound sliding clamp with small but important differences, making it a unique conformation. This cryo-EM map is obtained on open end DNA, where MutS has a high affinity for. Likely due to the fact that it can start transitioning from scanning or mismatch bound, towards a sliding clamp.

Sliding clamp formation and recruitment of MutL.

After mismatch detection and the conformational change towards the sliding clamp, MutS adopts the sliding clamp conformation that can bind MutL. It is known that, under physiological conditions, MutL cannot bind DNA independently but needs the presence of MutS bound to DNA. This could indicate that MutL binds DNA via MutS. Indeed, this is the case as was shown by many groups using various types of experiments [96, 97, 99, 100]. In addition, the crystal structure and cryo-EM map of MutS bound to $MutL^{LN40}$ and DNA showed the trimeric complex of MutS-$MutL^{LN40}$ on DNA [23] (Chapter 3).

Since *E. coli* MutS is a homodimer, there are two possible docking sites for $MutL^{LN40}$. Disruption of one of these two docking sites, still allows MutH activation *in vitro* (Chapter 3). In addition, the cryo-EM structure of crosslinked MutS-Mut^{LN40}

shows only binding of one MutL^{LN40} domain to MutS in contrast to the symmetric crystal structure where both docking sites on MutS were occupied by a MutL^{LN40} molecule. The structure shows that it is the DNA itself that prevents a second MutL from binding, even if it is present.

It is shown that the NTD of MutL interacts with MutS and DNA. However, the loading of full length MutL on DNA is not fully understood. As discussed previously, MutL can adopt various conformations depending on the nucleotide state. It is not known if the CTDs of MutL dimerizes prior to DNA loading and MutS interaction or afterwards, although we know that MutL forms a dimer in solution. However, dimerization via the NTDs is not possible since this would prevent MutL^{LN40} to interact with MutS and DNA (Chapter 3). Therefore, one could assume that the lifetime of MutS-MutL on DNA is rather short if MutL needs to dissociate from MutS to dimerize via the NTDs and make a closed circle around the DNA. However, various groups have shown that MutS-MutL form a stable, long-lived complex on DNA [23, 101, 102], and maybe dimerization is not essential during interaction with MutS. A 3D structure of full length MutS-MutL would allow us to understand the loading of MutL. Chapter 4 shows the results on obtaining a full length structure of MutS-MutL on DNA by cryo-EM.

Strand discrimination, nicking and unwinding

Scanning for mismatches and activation of MutS is similar from *E.coli* to humans. However, strand discrimination and further MMR activation is different for gram-negative bacteria such as *E.coli* than for other species. *E.coli* MMR uses the methylation of GATC sites as a discrimination signal since the newly replicated strand remains unmethylated temporarily [27]. Therefore, the system can sense which base of the mismatch is wrongly incorporated and needs to be replaced. The *E.coli* endonuclease, MutH, nicks the newly synthesized strand at hemi-methylated GATC sites. This nick then functions as a starting point for the helicase (Fig 1). MutH is a member of the type II family of restriction endonucleases [103]. MutH is recruited by MutL by forming a transient complex on the DNA. *In vitro* experiments show that, at physiological salt conditions, MutS needs to be present to recruit MutH to the DNA. Also, there are indications that a (transient) trimeric complex of MutS-MutL-MutH exist but this needs to be studied in more detail [104]. With the use of site-directed mutagenesis and site-specific crosslinking, the interaction sites of MutL with MutH were studied [105]. MutH can interact with both the CTD

and NTD of MutL although it is not known whether MutH contacts both the NTD and CTD domains of MutL prior to or after DNA binding by MutH.

Species lacking this DNA methylation signal are discriminating the new strand in a different manner. How individual species do this is not fully understood, although it is clear that the processivity clamp and MutL homologs play an important role [32, 59, 106, 107]. MutL homologs, like human MutLα, are latent endonucleases and their nuclease activity is activated by PCNA/β-clamp. Also, more importantly, the loading orientation of the clamp by Replication Factor C plays a role in discriminating between the parent and the newly replicated strand [106, 108]. But many questions are still unanswered and the precise mode of action of MutLα with the processivity clamps needs be studied in more detail.

After nicking of the DNA at GATC sites, the DNA needs to be unwound. This task is fulfilled by UvrD, belonging to the Helicase II family. UvrD is recruited by MutL [109]. Although a monomer of UvrD can translocate along single-stranded DNA, it is the UvrD dimer that is necessary for helicase activity *in vitro* [110]. In addition, MutL enhances the unwinding processivity of UvrD [110, 111]. For eukaryotes, there is no UvrD homolog known and it is exonuclease EXO1 that degrades the DNA strand in both 5'--> 3' and 3'--> 5', depending on the orientation of the nick related to the mismatch. Finally, for both prokaryotes and eukaryotes, the gap will be filled and sealed.

Structural Biology

Several methods are being used to study proteins and protein complexes, with some methods being more suitable for biomolecules of a certain size range. Nucleic Magnetic Resonance (NMR) is most suitable for relatively small biomolecules. X-ray crystallography can be used to study small and large biomolecules, with the note that crystals needs to be obtained of the molecule of interest. Cryo electron microscopy (cryo-EM) started out as a method to study the 3D structure of relatively large biomolecules. Due to the rapid developments in the field, smaller and smaller molecules can be studied. Besides the recent developments, cryo-EM has the advantage that growing crystals becomes unnecessary, which is the bottleneck in X-ray crystallography. Crystals tend to select for a specific conformation, and this is not the case for cryo-EM, which allows elucidation of transient conformations that could never be captured in a crystal lattice.

Single particle cryo-EM is a technique that was developed in the 1970s [112-116]. The technique is characterized by freezing your sample of interest at cryogenic temperatures; by plunge freezing the aqueous sample in liquid ethane, the sample gets a cover of vitrified ice [114]. This sample will then be analyzed by an electron beam and a significant amount of data is recorded. This is used to determine the 3D structure model of the biomolecule or biomolecules of interest by using 2D projections. Over the last 50 years, many developments took place. Firstly, the development of a protective vitrified ice layer around the sample to protect it to the high vacuum at the electron microscope [114]. Secondly, the development and commercial availability of direct electron detector cameras [117]. These cameras can localize the location of an electron much more precisely and the sensors can run at high frame rates, enabling cryo-EM images to be recorded as a stack of movie frames in a short time period. This increases the maximum resolution and allows to correct for beam induced motion [118]. The last major breakthrough is the analysis of cryo-EM data. Maximum likelihood-based approaches were developed in the past. Later, a Bayesian approach to cryo-EM reconstruction was described and implemented in RELION, a software used for analysis of single particle cryo-EM image processing [119, 120]. These major breakthroughs caused a revolution in cryo-EM in the last decade [119-122]. All this lead to winning the Nobel Prize in 2017 [123].

In this thesis, X-ray crystallography has been used to determine the structure of apo-MutS (Chapter 2). But most of the work in this thesis is based on data recorded using cryo-EM (Chapter 3 & 4). Since MutS is a very dynamic protein, it also adopts a range of conformations. Our biggest interest is to capture all these conformations which would allow us to understand the transition that MutS is undergoing along the mismatch repair cascade. Because MutS can slide on DNA in the presence of ATP, we had to use blocks at the end of the DNA to prevent MutS from sliding of the DNA. This was done by purifying monovalent streptavidin [124] and bind this to 5' biotinylated DNA. While this set-up allowed us to trap MutS and MutL together on DNA, it also complicated the analysis since both MutS and MutL were in very close proximity of the streptavidin moiety. This will be discussed in more detail in Chapter 4.

OUTLINE OF THESIS

This work describes the studies on the *E.coli* MMR proteins MutS and MutL. We studied how MutS undergoes several transitions and some of these transitions we have been able to trap using cryo-EM. In addition to the structural studies, we validated these conformations using biophysical experiments and crosslinking. We studied the mechanism of how MutS adopts different conformations. Furthermore, attempts were done to study MutL on DNA and together with MutS on DNA.

Chapter 2 describes the work on the apo-state of MutS. In this Chapter we show that MutS likely proceed to DNA binding via opening of the lever domains This opening is facilitated by kinking the lever domains at a defined hinge point. We show via mutational studies, crosslinking and DNA binding assays that kinking of the lever domains plays both a role in DNA binding and the transition to the sliding clamp.

Chapter 3 shows our work in collaboration with scientists in Madrid, Leiden, Giessen and Rotterdam. Together we show four cryo-EM maps of MutS in various conformations and the validation of these structures. We show that MutS scans DNA in a more open conformation and switches to the more closed mismatch-bound conformation. Then, MutS transforms to the transitions state, likely by binding and/or hydrolysis of ATP. Then, MutS adopts the sliding clamp conformation where it can bind MutL. Also here kinking of the lever domains plays an important role.

Chapter 4 explains all the work that has been done to study the complex of MutS and MutL on DNA, and MutL alone on DNA. All methods will be discussed and the processed cryo-EM data shows what we have learned from these studies.

Chapter 5 discusses all results against the broader picture of MMR.

REFERENCES

1. Watson, J., D., Crick, H.F., *Genetical Implications of the structure of Deoxyribonucleic acid.* Nature, 1953. **171**: p. 964-7.
2. Gregory, T.R., et al., *Eukaryotic genome size databases.* Nucleic Acids Res, 2007. **35**(Database issue): p. D332-8.
3. Collins FS, M.M., Patrinos A. , *The Human Genome Project: lessons from large-scale biology.* Science, 2003. **300**(5617): p. 286-290.
4. Kunkel, T.A., *Evolving views of DNA replication (in)fidelity.* Cold Spring Harb Symp Quant Biol, 2009. **74**: p. 91-101.
5. Hopfield, J.J., *The energy relay: a proofreading scheme based on dynamic cooperativity and lacking all characteristic symptoms of kinetic proofreading in DNA replication and protein synthesis.* Proc Natl Acad Sci U S A, 1980. **77**(9): p. 5248-52.
6. Kunkel, T.A., *Nucleotide repeats. Slippery DNA and diseases.* Nature, 1993. **365**(6443): p. 207-8.
7. Viguera, E., D. Canceill, and S.D. Ehrlich, *Replication slippage involves DNA polymerase pausing and dissociation.* Embo j, 2001. **20**(10): p. 2587-95.
8. De' Angelis, G.L., et al., *Microsatellite instability in colorectal cancer.* Acta Biomed, 2018. **89**(9-s): p. 97-101.
9. Pfuderer, P.L., et al., *High endothelial venules are associated with microsatellite instability, hereditary background and immune evasion in colorectal cancer.* Br J Cancer, 2019. **121**(5): p. 395-404.
10. Lynch, H.T. and A. de la Chapelle, *Genetic susceptibility to non-polyposis colorectal cancer.* J Med Genet, 1999. **36**(11): p. 801-18.
11. Lynch, H.T., et al., *Milestones of Lynch syndrome: 1895-2015.* Nat Rev Cancer, 2015. **15**(3): p. 181-94.
12. Cerretelli, G., et al., *Molecular pathology of Lynch syndrome.* J Pathol, 2020. **250**(5): p. 518-531.
13. Llosa, N.J., et al., *The vigorous immune microenvironment of microsatellite instable colon cancer is balanced by multiple counter-inhibitory checkpoints.* Cancer Discov, 2015. **5**(1): p. 43-51.
14. Seth, S., et al., *Lynch syndrome - cancer pathways, heterogeneity and immune escape.* J Pathol, 2018. **246**(2): p. 129-133.
15. Agata, Y., et al., *Expression of the PD-1 antigen on the surface of stimulated mouse T and B lymphocytes.* Int Immunol, 1996. **8**(5): p. 765-72.
16. Francisco, L.M., P.T. Sage, and A.H. Sharpe, *The PD-1 pathway in tolerance and autoimmunity.* Immunol Rev, 2010. **236**: p. 219-42.
17. Kwok, G., et al., *Pembrolizumab (Keytruda).* Hum Vaccin Immunother, 2016. **12**(11): p. 2777-2789.

18. Marcus, L., et al., *FDA Approval Summary: Pembrolizumab for the Treatment of Microsatellite Instability-High Solid Tumors.* Clin Cancer Res, 2019. **25**(13): p. 3753-3758.

19. Modrich, P. and R. Lahue, *Mismatch repair in replication fidelity, genetic recombination, and cancer biology.* Annu Rev Biochem, 1996. **65**: p. 101-33.

20. Kunkel, T.A. and D.A. Erie, *DNA mismatch repair.* Annu Rev Biochem, 2005. **74**: p. 681-710.

21. Jiricny, J., *The multifaceted mismatch-repair system.* Nat Rev Mol Cell Biol, 2006. **7**(5): p. 335-46.

22. Allen, D.J., et al., *MutS mediates heteroduplex loop formation by a translocation mechanism.* Embo j, 1997. **16**(14): p. 4467-76.

23. Groothuizen, F.S., et al., *MutS/MutL crystal structure reveals that the MutS sliding clamp loads MutL onto DNA.* Elife, 2015. **4**: p. e06744.

24. Grilley, M., et al., *Isolation and characterization of the Escherichia coli mutL gene product.* J Biol Chem, 1989. **264**(2): p. 1000-4.

25. Drotschmann, K., et al., *The Escherichia coli MutL protein stimulates binding of Vsr and MutS to heteroduplex DNA.* Nucleic Acids Res, 1998. **26**(4): p. 948-53.

26. Acharya, S., et al., *The coordinated functions of the E. coli MutS and MutL proteins in mismatch repair.* Mol Cell, 2003. **12**(1): p. 233-46.

27. Au, K.G., K. Welsh, and P. Modrich, *Initiation of methyl-directed mismatch repair.* J Biol Chem, 1992. **267**(17): p. 12142-8.

28. Kleczkowska, H.E., et al., *hMSH3 and hMSH6 interact with PCNA and colocalize with it to replication foci.* Genes Dev, 2001. **15**(6): p. 724-36.

29. Lopez de Saro, F.J., et al., *The beta sliding clamp binds to multiple sites within MutL and MutS.* J Biol Chem, 2006. **281**(20): p. 14340-9.

30. Lopez de Saro, F.J. and M. O'Donnell, *Interaction of the beta sliding clamp with MutS, ligase, and DNA polymerase I.* Proc Natl Acad Sci U S A, 2001. **98**(15): p. 8376-80.

31. Pillon, M.C., et al., *The sliding clamp tethers the endonuclease domain of MutL to DNA.* Nucleic Acids Res, 2015.

32. Almawi, A.W., et al., *Binding of the regulatory domain of MutL to the sliding β-clamp is species specific.* Nucleic Acids Res, 2019. **47**(9): p. 4831-4842.

33. Liu, J., et al., *MutL sliding clamps coordinate exonuclease-independent Escherichia coli mismatch repair.* Nat Commun, 2019. **10**(1): p. 5294.

34. de Wind, N., et al., *Inactivation of the mouse Msh2 gene results in mismatch repair deficiency, methylation tolerance, hyperrecombination, and predisposition to cancer.* Cell, 1995. **82**(2): p. 321-30.

35. Rayssiguier, C., D.S. Thaler, and M. Radman, *The barrier to recombination between Escherichia coli and Salmonella typhimurium is disrupted in mismatch-repair mutants.* Nature, 1989. **342**(6248): p. 396-401.

36. Manhart, C.M. and E. Alani, *Roles for mismatch repair family proteins in promoting meiotic crossing over.* DNA Repair (Amst), 2016. **38**: p. 84-93.

37. McMurray, C.T., *Mechanisms of trinucleotide repeat instability during human development.* Nat Rev Genet, 2010. **11**(11): p. 786-99.

38. Den Dunnen, W.F.A., *Trinucleotide repeat disorders.* Handb Clin Neurol, 2017. **145**: p. 383-391.

39. Schmidt, M.H.M. and C.E. Pearson, *Disease-associated repeat instability and mismatch repair.* DNA Repair (Amst), 2016. **38**: p. 117-126.

40. Cascalho, M., et al., *Mismatch repair co-opted by hypermutation.* Science, 1998. **279**(5354): p. 1207-10.

41. Chahwan, R., et al., *Mismatch-mediated error prone repair at the immunoglobulin genes.* Biomed Pharmacother, 2011. **65**(8): p. 529-36.

42. Schanz, S., et al., *Interference of mismatch and base excision repair during the processing of adjacent U/G mispairs may play a key role in somatic hypermutation.* Proc Natl Acad Sci U S A, 2009. **106**(14): p. 5593-8.

43. Hopfner, K., *Rad50/SMC proteins and ABC transporters: unifying concepts from high-resolution structures.* Current Opinion in Structural Biology, 2003. **13**(2): p. 249-255.

44. Melby, T.E., Ciampaglio, C.N., Briscoe, G., Erickson, H.P., *The Symmetrical Structure of Structural Maintenance of Chromosomes (SMC) and MukB Proteins: Long, Antiparallel Coiled Coils, Folded at a Flexible Hinge.* The Journl of Cell biology, 1998. **142**(6): p. 1595–1604.

45. Kamada, K., Miyata, M., Hirano, T, *Molecular Basis of SMC ATPase Activation: Role of Internal Structural Changes of the Regulatory Subcomplex ScpAB.* Cell, 2013. **21**: p. 581–594.

46. Diebold-Durand, M., Lee, H., Ruiz Avila, L.B., Noh, H., Shin, H., Im, H., Bock, F.P., Burmann, F., Durand, A., Basfeld, A., Ham,S., Basquin, J., Oh, B.,Gruber, S, *Structure of Full-Length SMC and Rearrangements Required for Chromosome Organization.* Cell, 2017. **67**: p. 334–347.

47. Bürmann, F., et al., *A folded conformation of MukBEF and cohesin.* Nature Structural & Molecular Biology, 2019. **26**(3): p. 227-236.

48. Tham, K.C., et al., *Mismatch repair inhibits homeologous recombination via coordinated directional unwinding of trapped DNA structures.* Mol Cell, 2013. **51**(3): p. 326-37.

49. Mendillo, M.L., C.D. Putnam, and R.D. Kolodner, *Escherichia coli MutS tetramerization domain structure reveals that stable dimers but not tetramers are essential for DNA mismatch repair in vivo.* J Biol Chem, 2007. **282**(22): p. 16345-54.

50. Groothuizen, F.S., et al., *Using stable MutS dimers and tetramers to quantitatively analyze DNA mismatch recognition and sliding clamp formation.* Nucleic Acids Res, 2013. **41**(17): p. 8166-81.

51. Lamers, M.H., et al., *The crystal structure of DNA mismatch repair protein MutS binding to a G x T mismatch.* Nature, 2000. **407**(6805): p. 711-7.

52. Obmolova, G., et al., *Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA.* Nature, 2000. **407**(6805): p. 703-10.

53. Natrajan, G., et al., *Structures of Escherichia coli DNA mismatch repair enzyme MutS in complex with different mismatches: a common recognition mode for diverse substrates.* Nucleic Acids Res, 2003. **31**(16): p. 4814-21.

54. Warren, J.J., et al., *Structure of the human MutSalpha DNA lesion recognition complex.* Mol Cell, 2007. **26**(4): p. 579-92.

55. Gupta, S., M. Gellert, and W. Yang, *Mechanism of mismatch recognition revealed by human MutSbeta bound to unpaired DNA loops.* Nat Struct Mol Biol, 2011. **19**(1): p. 72-8.

56. Lamers, M.H., et al., *ATP increases the affinity between MutS ATPase domains. Implications for ATP hydrolysis and conformational changes.* J Biol Chem, 2004. **279**(42): p. 43879-85.

57. Bhairosing-Kok, D., et al., *Sharp kinking of a coiled-coil in MutS allows DNA binding and release.* Nucleic Acids Res, 2019. **47**(16): p. 8888-8898.

58. Nirwal, S., et al., *Mechanism of formation of a toroid around DNA by the mismatch sensor protein.* Nucleic Acids Res, 2018. **46**(1): p. 256-266.

59. Kunkel, T.A. and D.A. Erie, *Eukaryotic Mismatch Repair in Relation to DNA Replication.* Annu Rev Genet, 2015. **49**: p. 291-313.

60. Lebbink, J.H., et al., *Dual role of MutS glutamate 38 in DNA mismatch discrimination and in the authorization of repair.* Embo j, 2006. **25**(2): p. 409-19.

61. Junop, M.S., et al., *Composite active site of an ABC ATPase: MutS uses ATP to verify mismatch recognition and authorize DNA repair.* Mol Cell, 2001. **7**(1): p. 1-12.

62. Schofield, M.J., et al., *Interaction of Escherichia coli MutS and MutL at a DNA mismatch.* J Biol Chem, 2001. **276**(30): p. 28291-9.

63. Bjornson, K.P., D.J. Allen, and P. Modrich, *Modulation of MutS ATP hydrolysis by DNA cofactors.* Biochemistry, 2000. **39**(11): p. 3176-83.

64. Blackwell, L.J., et al., *Distinct MutS DNA-binding modes that are differentially modulated by ATP binding and hydrolysis.* J Biol Chem, 2001. **276**(36): p. 34339-47.

65. Lamers, M.H., H.H. Winterwerp, and T.K. Sixma, *The alternating ATPase domains of MutS control DNA mismatch repair.* Embo j, 2003. **22**(3): p. 746-56.

66. Hingorani, M.M., *Mismatch binding, ADP-ATP exchange and intramolecular signaling during mismatch repair.* DNA Repair (Amst), 2016. **38**: p. 24-31.

67. Iaccarino, I., et al., *hMSH2 and hMSH6 play distinct roles in mismatch binding and contribute differently to the ATPase activity of hMutSalpha.* Embo j, 1998. **17**(9): p. 2677-86.

68. Jacobs-Palmer, E. and M.M. Hingorani, *The effects of nucleotides on MutS-DNA binding kinetics clarify the role of MutS ATPase activity in mismatch repair.* J Mol Biol, 2007. **366**(4): p. 1087-98.

69. Monti, M.C., et al., *Native mass spectrometry provides direct evidence for DNA mismatch-induced regulation of asymmetric nucleotide binding in mismatch repair protein MutS.* Nucleic Acids Res, 2011. **39**(18): p. 8052-64.

70. Zhou, Y., et al., *Toward Determining ATPase Mechanism in ABC Transporters: Development of the Reaction Path-Force Matching QM/MM Method.* Methods Enzymol, 2016. **577**: p. 185-212.

71. Mittal, A., et al., *Asymmetry in the homodimeric ABC transporter MsbA recognized by a DARPin.* J Biol Chem, 2012. **287**(24): p. 20395-406.

72. Procko, E., et al., *Distinct structural and functional properties of the ATPase sites in an asymmetric ABC transporter.* Mol Cell, 2006. **24**(1): p. 51-62.

73. Zaitseva, J., et al., *A structural analysis of asymmetry required for catalytic activity of an ABC-ATPase domain dimer.* Embo j, 2006. **25**(14): p. 3432-43.

74. Elbatsh, A.M.O., et al., *Distinct Roles for Condensin's Two ATPase Sites in Chromosome Condensation.* Mol Cell, 2019. **76**(5): p. 724-737.e5.

75. Hirano, T., *Condensin-Based Chromosome Organization from Bacteria to Vertebrates.* Cell, 2016. **164**(5): p. 847-57.

76. Hopfner, K.P., *Invited review: Architectures and mechanisms of ATP binding cassette proteins.* Biopolymers, 2016. **105**(8): p. 492-504.

77. Sancar, A. and J.E. Hearst, *Molecular matchmakers.* Science, 1993. **259**(5100): p. 1415-20.

78. Mendillo, M.L., et al., *A conserved MutS homolog connector domain interface interacts with MutL homologs.* Proc Natl Acad Sci U S A, 2009. **106**(52): p. 22223-8.

79. Dutta, R. and M. Inouye, *GHKL, an emergent ATPase/kinase superfamily.* Trends Biochem Sci, 2000. **25**(1): p. 24-8.

80. Groothuizen, F.S. and T.K. Sixma, *The conserved molecular machinery in DNA mismatch repair enzyme structures.* DNA Repair (Amst), 2016. **38**: p. 14-23.

81. Ban, C., M. Junop, and W. Yang, *Transformation of MutL by ATP binding and hydrolysis: a switch in DNA mismatch repair.* Cell, 1999. **97**(1): p. 85-97.

82. Guarne, A., M.S. Junop, and W. Yang, *Structure and function of the N-terminal 40 kDa fragment of human PMS2: a monomeric GHL ATPase.* Embo j, 2001. **20**(19): p. 5521-31.

83. Guarne, A., et al., *Structure of the MutL C-terminal domain: a model of intact MutL and its roles in mismatch repair.* Embo j, 2004. **23**(21): p. 4134-45.

84. Pillon, M.C., et al., *Structure of the endonuclease domain of MutL: unlicensed to cut.* Mol Cell, 2010. **39**(1): p. 145-51.

85. Kosinski, J., et al., *Analysis of the quaternary structure of the MutL C-terminal domain.* J Mol Biol, 2005. **351**(4): p. 895-909.

86. Ban, C. and W. Yang, *Crystal structure and ATPase activity of MutL: implications for DNA repair and mutagenesis.* Cell, 1998. **95**(4): p. 541-52.

87. Namadurai, S., et al., *The C-terminal domain of the MutL homolog from Neisseria gonorrhoeae forms an inverted homodimer.* PLoS One, 2010. **5**(10): p. e13726.

88. Gueneau, E., et al., *Structure of the MutLalpha C-terminal domain reveals how Mlh1 contributes to Pms1 endonuclease site.* Nat Struct Mol Biol, 2013. **20**(4): p. 461-8.

89. Wu, H., et al., *Structure of the human MLH1 N-terminus: implications for predisposition to Lynch syndrome.* Acta Crystallogr F Struct Biol Commun, 2015. **71**(Pt 8): p. 981-5.

90. Hermans, N., et al., *Dual daughter strand incision is processive and increases the efficiency of DNA mismatch repair.* Nucleic Acids Res, 2016. **44**(14): p. 6770-86.

91. Mardenborough, Y.S.N., et al., *The unstructured linker arms of MutL enable GATC site incision beyond roadblocks during initiation of DNA mismatch repair.* Nucleic Acids Res, 2019. **47**(22): p. 11667-11680.

92. Sacho, E.J., et al., *Direct visualization of asymmetric adenine-nucleotide-induced conformational changes in MutL alpha.* Mol Cell, 2008. **29**(1): p. 112-21.

93. Yamamoto, A., et al., *Requirement for Phe36 for DNA binding and mismatch repair by Escherichia coli MutS protein.* Nucleic Acids Res, 2000. **28**(18): p. 3564-9.

94. Yang, Y., et al., *Determination of protein-DNA binding constants and specificities from statistical analyses of single molecules: MutS-DNA interactions.* Nucleic Acids Res, 2005. **33**(13): p. 4322-34.

95. Tessmer, I., et al., *Mechanism of MutS searching for DNA mismatches and signaling repair.* J Biol Chem, 2008. **283**(52): p. 36646-54.

96. Gradia, S., et al., *hMSH2-hMSH6 forms a hydrolysis-independent sliding clamp on mismatched DNA.* Mol Cell, 1999. **3**(2): p. 255-61.

97. Jeong, C., et al., *MutS switches between two fundamentally distinct clamps during mismatch repair.* Nat Struct Mol Biol, 2011. **18**(3): p. 379-85.

98. Prolla, T.A., et al., *MLH1, PMS1, and MSH2 interactions during the initiation of DNA mismatch repair in yeast.* Science, 1994. **265**(5175): p. 1091-3.

99. Qiu, R., et al., *Large conformational changes in MutS during DNA scanning, mismatch recognition and repair signalling.* Embo j, 2012. **31**(11): p. 2528-40.

100. Cho, W.K., et al., *ATP alters the diffusion mechanics of MutS on mismatched DNA.* Structure, 2012. **20**(7): p. 1264-74.

101. Liu, J., et al., *Cascading MutS and MutL sliding clamps control DNA diffusion to activate mismatch repair.* Nature, 2016. **539**(7630): p. 583-587.

102. Lee, J.B., et al., *Single-molecule views of MutS on mismatched DNA.* DNA Repair (Amst), 2014. **20**: p. 82-93.

103. Welsh, K.M., et al., *Isolation and characterization of the Escherichia coli mutH gene product.* J Biol Chem, 1987. **262**(32): p. 15624-9.

104. Ahrends, R., et al., *Identifying an interaction site between MutH and the C-terminal domain of MutL by crosslinking, affinity purification, chemical coding and mass spectrometry.* Nucleic Acids Res, 2006. **34**(10): p. 3169-80.

105. Giron-Monzon, L., et al., *Mapping protein-protein interactions between MutL and MutH by cross-linking.* J Biol Chem, 2004. **279**(47): p. 49338-45.

106. Pluciennik, A., et al., *PCNA function in the activation and strand direction of MutLalpha endonuclease in mismatch repair.* Proc Natl Acad Sci U S A, 2010. **107**(37): p. 16066-71.

107. Genschel, J., et al., *Interaction of proliferating cell nuclear antigen with PMS2 is required for MutLα activation and function in mismatch repair.* Proc Natl Acad Sci U S A, 2017. **114**(19): p. 4930-4935.

108. Kadyrova, L.Y. and F.A. Kadyrov, *Endonuclease activities of MutLα and its homologs in DNA mismatch repair.* DNA Repair (Amst), 2016. **38**: p. 42-49.

109. Yamaguchi, M., V. Dao, and P. Modrich, *MutS and MutL activate DNA helicase II in a mismatch-dependent manner.* J Biol Chem, 1998. **273**(15): p. 9197-201.

110. Ordabayev, Y.A., et al., *Regulation of UvrD Helicase Activity by MutL.* J Mol Biol, 2018. **430**(21): p. 4260-4274.

111. Ordabayev, Y.A., et al., *UvrD helicase activation by MutL involves rotation of its 2B subdomain.* Proc Natl Acad Sci U S A, 2019. **116**(33): p. 16320-16325.

112. De Rosier, D.J. and A. Klug, *Reconstruction of three dimensional structures from electron micrographs.* Nature, 1968. **217**(5124): p. 130-4.

113. Taylor, K.A. and R.M. Glaeser, *Electron diffraction of frozen, hydrated protein crystals.* Science, 1974. **186**(4168): p. 1036-7.

114. McDowall, A.W., et al., *Electron microscopy of frozen hydrated sections of vitreous ice and vitrified biological samples.* J Microsc, 1983. **131**(Pt 1): p. 1-9.

115. Nogales, E., S.G. Wolf, and K.H. Downing, *Structure of the alpha beta tubulin dimer by electron crystallography.* Nature, 1998. **391**(6663): p. 199-203.

116. Nogales, E., *The development of cryo-EM into a mainstream structural biology technique.* Nat Methods, 2016. **13**(1): p. 24-7.

117. McMullan, G., A.R. Faruqi, and R. Henderson, *Direct Electron Detectors.* Methods Enzymol, 2016. **579**: p. 1-17.

118. Li, X., et al., *Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM.* Nat Methods, 2013. **10**(6): p. 584-90.

119. Scheres, S.H., *RELION: implementation of a Bayesian approach to cryo-EM structure determination.* J Struct Biol, 2012. **180**(3): p. 519-30.

120. Cheng, Y., *Single-particle cryo-EM-How did it get here and where will it go.* Science, 2018. **361**(6405): p. 876-880.

121. Bai, X.C., et al., *An atomic structure of human γ-secretase.* Nature, 2015. **525**(7568): p. 212-217.

122. Zivanov, J., et al., *New tools for automated high-resolution cryo-EM structure determination in RELION-3.* Elife, 2018. **7**.

123. Zhang, K. and Z. Liu, *A brief introduction of cryo-EM revolution-the Nobel Prize in Chemistry 2017.* Sci China Life Sci, 2018. **61**(3): p. 368-370.

124. Fairhead, M., et al., *Plug-and-play pairing via defined divalent streptavidins.* J Mol Biol, 2014. **426**(1): p. 199-214.

CHAPTER 2

Sharp kinking of a coiled-coil in MutS allows DNA binding and release

Doreth Bhairosing-Kok[1*]
Flora S. Groothuizen[1*]
Alexander Fish[1*]
Shreya Dharadhar[1]
Herrie H.K. Winterwerp[1]
Titia K. Sixma[1]

* These authors contributed equally

[1]*Division of Biochemistry and Oncode Institute, Netherlands Cancer Institute, Amsterdam, the Netherlands*

Nucleic Acids Research, 47, 8888-8898 (2019)

ABSTRACT

DNA mismatch repair (MMR) corrects mismatches, small insertions and deletions in DNA during DNA replication. While scanning for mismatches, dimers of MutS embrace the DNA helix with their lever and clamp domains. Previous studies indicated generic flexibility of the lever and clamp domains of MutS prior to DNA binding, but whether this was important for MutS function was unknown.

Here we present a novel crystal structure of DNA-free *E.coli* MutS. In this apo-structure, the clamp domains are repositioned due to kinking at specific sites in the coiled-coil region in the lever domains, suggesting a defined hinge point. Mutations at the coiled-coil hinge point show that a disruption of the helical fold at the kink site diminishes DNA binding and that increased stability of coiled-coil results in stronger DNA binding. We conclude that the site-specific kinking of the coiled-coil in the lever domain is important for loading of this ABC-ATPase on DNA.

INTRODUCTION

The DNA mismatch repair (MMR) pathway is responsible for maintaining genetic information by correcting base-substitution and insertion-deletion mismatches, generated during DNA replication (1,2). MMR deficiency results in a mutator phenotype and in humans it can predispose to cancer, referred to as HNPCC or Lynch syndrome (3). MutS initiates the repair by scanning the DNA for mismatches. Upon mismatch detection, it signals for repair by forming a sliding clamp that activates MutL (4,5). MutL then activates the downstream repair, which includes nicking the newly replicated strand, unwinding the DNA and resynthesis of the daughter strand.

MutS proteins belong to the ATP-binding cassette (ABC) ATPases and are evolutionarily conserved from bacteria to mammals. MutS forms constitutive dimers, while some prokaryotic MutS homologs can also tetramerize through their C-terminal domain, but this is not required for MMR (6-8). In eukaryotic cells, the MutS homologs that are active in MMR form heterodimers (MSH2/MSH6 or MSH2/MSH3) (9). In both prokaryotes and eukaryotes, MutS acts as a heterodimer during MMR. During mismatch binding, both monomers embrace the DNA helix with their lever and clamp domains (10-13) but only one of the monomers recognizes the DNA mismatch through its mismatch-binding domain.

The binding of the MutS clamp domains around the DNA helix is expected to be a general feature, regardless of the presence of a mismatch. The crystal structure of DNA-free *Thermus aquaticus* MutS indicated disorder of large portions of MutS in the absence of DNA. The clamp, lever and mismatch domains were not ordered while the dimer itself was kept intact (11). Conformational freedom of the clamp domains was also observed in SAXS studies with the DNA-free protein (8). Both studies indicate that in the absence of DNA, the clamp and lever domains of MutS dimers are flexible, but how such flexibility is achieved was unclear. The lever domains of MutS are composed of two α-helices, forming a left-handed antiparallel coiled-coil (14). Other ABC-ATPase family members, such as Rad50 and the Structural Maintenance of Chromosomes (SMC) family have coiled-coil levers that are even more extended. The latter proteins were shown to have conformational flexibility in their coiled-coils in Atomic Force Microscopy (AFM) and rotary shadowing Electron Microscopy (EM) analyses (15-18).

Here we present a novel crystal structure of a DNA-free *Escherichia coli* MutS dimer, where the lever domains display an unexpected kink, resulting in a displacement of the clamp domains relative to the mismatch-bound state (10). In our structure, the conformational freedom originates from hinging of specific regions in the lever domains. We wondered if the ability to kink rather than generic flexibility could be important for MutS function. We studied the effect of mutations at the kinking site on different steps in MutS activation, including DNA binding, mismatch recognition, ATP-dependent clamp formation and loading of MutL. The data show that the defined hinge is important, particularly for DNA binding by this ABC-ATPase.

MATERIALS AND METHODS

Proteins

MutS mutants were created in the *mutS* gene in vector pET-3d and LOCK1, a MutS double cysteine variant E435C R449C, was created in a *mutS* cysteine free vector (19,20), using the QuikChange Site-Directed Mutagenesis Kit (Stratagene) and appropriate primers (IDT). All MutS and MutL proteins were expressed and purified as described (21), except for mutants MutS FLEX2 and MutS FLEX3 where the lysis buffer contained 10% glycerol and an increased salt concentration of 400 mM.

Crystallography

Crystallization of full-length DNA-free MutS P839E was performed using 50 μM MutS (in 25 mM Hepes 7.5, 150 mM NaCl, 10 mM $MgCl_2$) mixed with ADP (final concentration 50 μM) in 200 nL. The protein was crystallized using vapor diffusion with 3-8% dioxane, 1.4-1.7 mM $(NH_4)_2SO_4$ and 100 mM Hepes pH 7.0. The crystal was transferred to mother liquor supplemented with 30% glycerol before flash cooling it in liquid nitrogen.

Crystallographic data were collected at ESRF beamline ID14-4 and was processed using iMosflm (22) and Scala (23). The initial structure was solved using molecular replacement for the ATPase domain in Phaser (24) with part of chain A of PDB entry 1WB9 as search model and stepwise addition of domains. Structure refinement was performed using Buster (25), Refmac5 (26) and PDB-

REDO (27). Residues 12-80 (mismatch binding domain) in chain A and residues 442-503 (clamp domain) in chain B have density that is not very well defined, indicating some flexibility within the crystal. They were placed in the structure as rigid bodies, using the conserved fold of the mismatch binding domain and clamp domain, respectively. See Table 1 for crystallographic statistics.

Table 1. Crystallographic data and refinement statistics

Data collection	
λ (Å)	0.934
Resolution range (Å)[a]	80.2-2.6 (2.67-2.6)
Completeness (%)	98.1 (94.2)
I/σ(I)	4.6 (1.4)
R_{merge} (%)	0.30 (1.2)
Space group	P 2_1 2_1 2_1
Cell dimensions (Å)	113.38 113.53 158.90
Total no. of observations	253695 (17401)
Total no. of unique reflections	62205 (4381)
Multiplicity	4.1 (4.0)
Wilson's B-factor ($Å^2$)	32.2
Refinement	
No. of atoms (protein/solvent)	12198/254
Average B-factor ($Å^2$)	59
R_{free} reflections	3051 (4.91%)
R_{work}/R_{free} (%)	20.6/25.7
Ramachandran[b]	1450/73/0
Bond r.m.s.z/r.m.s.d. (Å)	0.53/0.0106
Angle r.m.s.z/r.m.s.d.(°)	0.78/1.27

[a] Numbers within parentheses refer to the highest resolution shell

[b] Number of residuals favoured/allowed/outliers

Protein stability measurements

Protein stability was assayed using a Prometheus NT.48 (Nanotemper). WT or mutant MutS proteins were diluted to 1 mg/mL and subjected to a temperature gradient to determine melting temperatures (T_m), which were read out by changes in tryptophan fluorescence. Analytical gel filtration was performed for MutS WT

and all mutants on a S200 5/100 (GE Healthcare) in 25 mM Hepes 7.5, 250 mM KCl and 1 mM DTT on an Akta Micro system.

Equilibrium DNA binding

Fluorescence polarization measurements to assess DNA-binding affinities of WT and mutant MutS proteins were performed in buffer consisting of 25 mM Hepes pH 7.5, 150 mM KCl, 5 mM $MgCl_2$, 0.05% Tween-20. A concentration of 0.5 nM of 5' labeled TAMRA-21-bp DNA with a mismatch at position 9 (5'-TAMRA-AGCTGCCAGGCACCAGTGTCA annealed with TGACACTGGTGCTTGGCAGCT) was used as fluorescent probe. MutS proteins were serially diluted in black flat-bottomed 384 well plates (Corning) in 30 µL volumes. The plate was equilibrated at RT for 5 minutes, after which polarization of the TAMRA label was read out in a PHERAstar FS machine (BMG Labtech) with a 540/590 (excitation/emission) FP module. K_d^{app} values were determined using GraphPad Prism 7.02.

Surface plasmon resonance analysis of DNA-binding kinetics

Kinetics of MutS binding to 21-bp DNA containing a GT mismatch (sequence, see above), which was attached to a streptavidin chip via a biotin-conjugated $(dT)_{20}$ linker, were determined using surface plasmon resonance (SPR). The measurements were performed in a Biacore T200 system (GE Healthcare) at 25 °C with the same setup as described previously (8,21). Data analysis of FP and SPR experiments was done using GraphPad Prism 7.02 (GraphPad Software Inc.). In order to calculate apparent equilibrium binding constants (K_d^{app}) binding responses were plotted as function of protein concentration and fitted with one site binding equation:

$$Y = \frac{Bmax \times X}{Kd + X} + Background$$

Where Y = binding response; X = protein concentration; Background = background response, B_{max} = Maximum binding response; K_d = apparent equilibrium binding constants.

In order to calculate apparent dissociation rate constants (k_{off}^{app}) the dissociation phases of the SPR sensograms were fitted with one phase decay equation:

$$Y = (Y_0 - Plateau) * e^{-k \times (t - t_0)} + Plateau$$

Where Y = binding response; Y_0 = initial binding response; Plateau = final plateau after dissociation; k = k_{off}^{app} ; t_0 = dissociation starting time.

SPR to analyze MutL binding

Binding kinetics of MutS-MutL to 100-bp containing a GT mismatch, were performed in the same set-up (8,21), with the addition 1 mg/ml BSA and 1 mM ATP. Biotinylated DNA was immobilized on a streptavidin chip to a signal of ~ 15.0 RU. Double biotinylated DNA (both 5' ends) were used as a blocked end oligo. MutS injections of 1200 nM for 60 seconds were followed by MutL injections for 120 seconds with concentrations ranging from 0-2048 nM. High concentration for MutS was chosen to achieve a full binding for all MutS mutants. In between individual cycles, SA chips were regenerated with 0.5% SDS. MutL binding analysis involved normalization on MutS binding at 60 seconds and subtracted for MutS binding, therefor making it possible to fit K_d^{app} for MutL binding using GraphPad Prism 7.

For titration of ATP in this context, the same DNA was immobilized. Experiments were performed using the same buffer but lacking ATP. Again 1200 nM MutS was injected for 60 seconds, followed by a second injection of 120 seconds containing buffer (control), 0.5 µM, 4 µM or 32 µM ATP, or 400 nM MutL supplemented with 0.5 µM, 4 µM or 32 µM ATP.

Crosslinking of LOCK1

A double cysteine mutant E435C R449C (LOCK1) was reduced at 80 µM with 10 mM DTT for 30 min and dialyzed o/n at 4 °C into binding buffer (25 mM Hepes 7.5, 150 mM KCl, 5 mM $MgCl_2$), to remove DTT. Crosslinkers $BM(PEG)_2$ (linker length 14.7Å) and $BM(PEG)_3$ (linker length 17.8Å) (Sigma) were dissolved in DMSO to a final concentration of 50 mM. Crosslinker was added in 30-fold molar excess over the MutS-monomer concentration and incubated for 3 h at 4 °C. Excess of crosslinker was removed using Zeba spin column (7K MWCO, 2 mL, ThermoFisher) and crosslinked MutS was used for DNA-binding kinetics using SPR. SDS PAGE gel analysis was used to confirm the absence of MutS inter-crosslink dimers (not shown).

Crosslinked LOCK1 was added to 100-bp containing a GT mismatch (21) in 10:1 ratio ([DNA]:[MutS monomer]), incubated for 30 min at RT and analyzed on an Akta Micro using Superdex 200 5/150 (in binding buffer). Using crosslinked

LOCK1 and LOCK1 we could examine the binding affinity of the mixed sample. Both proteins were mixed in a 1:1 ratio and incubated for 30 min at RT to form a new equilibrium of heterodimers and homodimers. After 30 min, dilutions were made for SPR experiments (see details above). For all SPR experiments regarding LOCK1, the binding buffer was supplemented with 1 mg/ml BSA.

RESULTS

Crystal structure of DNA-free MutS

Full length *E. coli* MutS P839E, that does not tetramerize (8), was crystallized in the absence of DNA and its structure was solved at a resolution of 2.6 Å (Table 1, Figure 1A). Structure solution required a step-wise approach, as domains were rearranged with respect to each other and to other previously determined structures, preventing straightforward molecular replacement. The structure was solved by first placing the ATPase domains, followed by stepwise addition of other domains. In the final structure, the connector and mismatch binding domains can be positioned, but they are poorly resolved. In contrast to most other structures, the full-length protein was used for crystallization. However, the C-terminal tetramerization domains were not resolved in the density, indicating that their position is not stabilized by crystal contacts in this crystal form. This flexible positioning of the C-terminal domains is in line with SAXS analysis (8).

In this structure, the homodimer of MutS adopts an asymmetric conformation, where domains are arranged differently between the two monomers; this is true for lever, clamp connector and mismatch binding domains. However, the ATPase domains are symmetric, in contrast to mismatch-bound *E. coli* MutS structures. In the presence of excess nucleotides, binding of two ADP nucleotides was the most abundant state in native mass-spectrometry measurements (28). Both monomers bind ADP, but no density can be seen for magnesium. This lack of magnesium was previously observed in other DNA-free structures of MutS homologs (11,29).

Kinking of lever domains MutS

A remarkable feature in this DNA-free crystal structure is the position and orientation of the clamp and lever domains of both monomers (Figure 1). One of the clamp domains has moved towards the core of the dimer (inward kink, 23°),

relative to mismatch-bound *E.coli* MutS, whereas the clamp domain in the other subunit moved outward (outward kink, 66°). Strikingly, both motions originate from a sharp kink in the same region; residues 440-443 and residues 515-518 in helix 1 and 2 of the coiled-coil respectively, within the lever domain (Figure 1B).

The two lever domains adopt an orientation and conformation that is changed relative to other MutS structures (8,10,29,30). Interestingly, these parts of the helices showed relatively high B-factors in the mismatch-bound MutS structures, already suggesting some degree of disorder (10). The presence of these defined hinge points suggests a way for MutS dimers to 'open up' and allow a DNA helix to enter the DNA-binding site, while DNA could also be released in this manner if no mismatch is found and/or sliding clamp is formed. We therefore hypothesize that this crystal structure represents a conformation of MutS that precedes DNA binding.

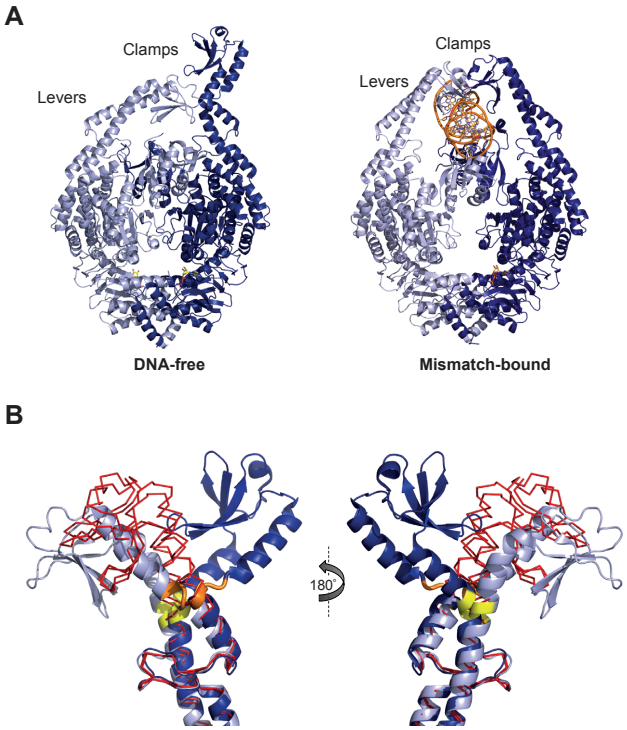


Figure 1. DNA-free crystal structure of MutS with differently positioned clamp domains. A) Cartoons of crystal structure of apo-structure and mismatch-bound (PDB ID: 1E3M) *E. coli* MutS. The two monomers are shown as dark blue and light blue cartoons, and DNA and ADP are shown in orange. **B)** Superposition to show the kinking that reorients the clamp domains of both monomers in the apo-structure (color as in A) relative to mismatch-bound MutS (shown as red ribbons).

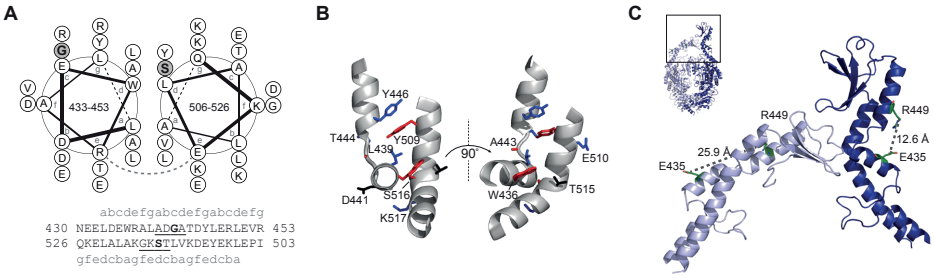
Mutations to influence the kinking in the lever domain

We wondered whether the kink-movements in the lever domains were essential for DNA binding. Therefore, we studied the kinking of the lever domains in two ways; 1) by making several mutants in the coiled-coil region that would affect the kinking-ability (Figure 2A, B) and 2) by locking MutS in a kinked-outward conformation via crosslinking of two cysteines in one of the helices of the coiled-coil (Figure 2C).

We made mutants to affect the coiled-coil stability in the hinge-region of both helices. The lever domains of MutS form a left-handed antiparallel coiled-coil arrangement (14) (Figure 2A). Coiled-coil structures are common structural motifs in which helices wrap around each other to form a super helix. While the effect of sequence variation on coiled-coil stability is not fully understood, it is generally accepted that hydrophobic residues such as leucines and alanines at the *a* and *d* positions of the heptad repeats facilitate dimer interaction (31). Similarly, charged residues such as glutamate or lysine at the *e* and *g* positions facilitate interhelical electrostatic interactions (31) (Figure 2A). Both helices of the coiled-coil are kinked in the DNA-free structure. Therefore, we hypothesized that if we can create a MutS mutant that has less ability to kink, MutS will be more stable in the closed conformation and will remain a stable complex with DNA. Inversely, the more the kink will be promoted, the less stable the coiled-coil will be and the DNA binding ability will decrease. Using this hypothesis, we designed several mutants aimed to stabilize or perturb the coiled-coil, to investigate the importance of the kinking that we observed in our crystal structure.

Stabilization of the coiled-coils was promoted by introducing hydrophobic residues at position *a* and *d*, resulting in MutS mutants STAB1-3 (Figure 2). In addition, STAB4 was made with the additional charged residues at position *e* and *g* to allow more ionic interactions. We also aimed to perturb the helical fold by introducing Pro/Gly/Pro motifs at the hinge region. These destabilizing mutants will weaken the coiled-coil stability, thereby making it more flexible. We created MutS mutants FLEX1 and FLEX2 and a combination of these, FLEX3 (Figure 2).

The second approach was to generate a double cysteine mutant of MutS (LOCK1), that allowed us to lock the MutS dimer in the outward-kinked conformation in both monomers by crosslinking the two sides of the helix using a bismaleimide crosslinker with a suitable length (Figure 2C).



Protein			Mutations	Description
Aimed to promote stable coiled-coil				
STAB1	430 NEELDEWRALAPAVTDYLERLEVR 453 526 QKELALAKGKSTLVKDEFEKLEPI 503		D441P/G442A/A443V Y509F	Promote helical fold in *h1*, increase *cc* packing with A443V/Y509F
STAB2	430 NEELDEWRALADGATDYLERLEVR 453 526 QKELALAKGKATLVKDEYEKLEPI 503		S516A	Promote helical fold at kink in *h2*
STAB3	430 NEELDELRALADGATDYLERLEVR 453 526 QKELALAKGKATLVKDELEKLEPI 503		W436L/Y509L/S516A	Promote quality of hydrophobic core of *cc*
STAB4	430 NEELDELRAEADGAEDELERLEVR 453 526 QKELALAKGKATLVKDKLEKLEPI 503		W436L/L439E/T444E/Y446E Y509L/E510K/S516A	Promote quality of hydrophobic core and salt bridges of *cc*
Aimed to promote locked kink-out conformation				
LOCK1	430 NEELDCWRALADGATDYLECLEVR 453 526 QKELALAKGKSTLVKDEYEKLEPI 503		E435C/R449C	Two cysteines residues for crosslinking
Aimed to promote destabilization coiled coil				
FLEX1	430 NEELDEWRALAPGPTDYLERLEVR 453 526 QKELALAKGKSTLVKDEYEKLEPI 503		D441P/A443P	Pro/Gly/Pro motif in *h1*
FLEX2	430 NEELDEWRALADGATDYLERLEVR 453 526 QKELALAKGPGPLVKDEYEKLEPI 503		T515P/S516G/K517P	Pro/Gly/Pro motif in *h2*
FLEX3	430 NEELDEWRALAPGPTDYLERLEVR 453 526 QKELALAKGPGPLVKDEYEKLEPI 503		D441P/A443P T515P/S516G/K517P	Pro/Gly/Pro motif in *h1* and *h2*

h1 = helix 1 (kinks around residue 442); *h2* = helix2 (kinks around residue 516); *cc* = coiled coil

Figure 2. The coiled-coil in the lever domain of MutS. **A)** Helical wheel representation of the coiled-coil in the lever domain of WT MutS. Sequences of the helices and their corresponding heptad assignment are written below. The residues around which the kinking takes place are underscored. Table below shows all mutants that were made to study the coiled-coil stability. Positions on helical wheel were chosen first and suitable mutation were made on specific positions. **B)** Residues mutated on hydrophobic core positions in coiled-coil (*a* and *d*) are shown in red. Residues that could contribute to stability due to ionic interactions (*e* and *g*) are shown in blue. Remaining mutants are colored in black. **C)** MutS can be locked in a kink-open conformation by creating a double cysteine MutS and creating a bridge between position 435C and 449C using bismaleimide-activated PEG compound. This is unique for the kink-open conformation since the kink-in and mismatch-bound conformation (not shown) will not allow this due to distance and sterical hindrance.

Since mutations affect the secondary structure, they can also influence overall protein stability. Therefore, all mutant proteins were subjected to thermostability measurements and compared to WT (Table 2). We found that the mutations had only marginal effects on the stability of the protein, as indicated by melting temperature for all variants, with the exception of LOCK1 and FLEX3. In addition, the purification of MutS FLEX2 and FLEX3 required higher salt and glycerol during lysis, indicating lower solubility, but this was not observed for the LOCK1 mutant. Results on these three mutants must therefore be interpreted with caution. In addition, analytical gel filtration profiles of all variants show similar elution profiles, indicating that no major conformational changes have occurred (Supplemental Figure S1)

DNA binding is affected by mutations in the kinking regions

To investigate whether the mutations in the coiled-coil influence DNA binding abilities, we used fluorescence polarization (FP) to measure equilibrium binding to a TAMRA-labeled DNA oligomer containing a GT mismatch (Table-2) (Figure 3). As predicted, all STAB-mutants showed comparable or stronger binding than WT. All FLEX-mutants that were designed to decrease the coiled-coil stability showed weaker DNA binding.

To investigate whether these differences in affinities for DNA originated from changes in binding kinetics, we performed surface plasmon resonance (SPR) assays in which DNA binding kinetics can be assessed. The SPR measurements occur under flow conditions resulting in somewhat different affinity values compared to the FP measurements. However, the variations between mutants remained consistent between the two techniques. The STAB mutants show a similar K_d^{app} as WT except for STAB4. However, their apparent off rates are different (Table 2, Figure 3C).

Destabilizing mutants, FLEX1, FLEX2 and FLEX3 all show faster dissociation rate, resulting into a weaker DNA affinity (Table 2, Figure 3D). We also observed that mutating residues in helix 2 (FLEX2) has a larger effect on DNA affinity than mutating residues in helix 1 (FLEX1).

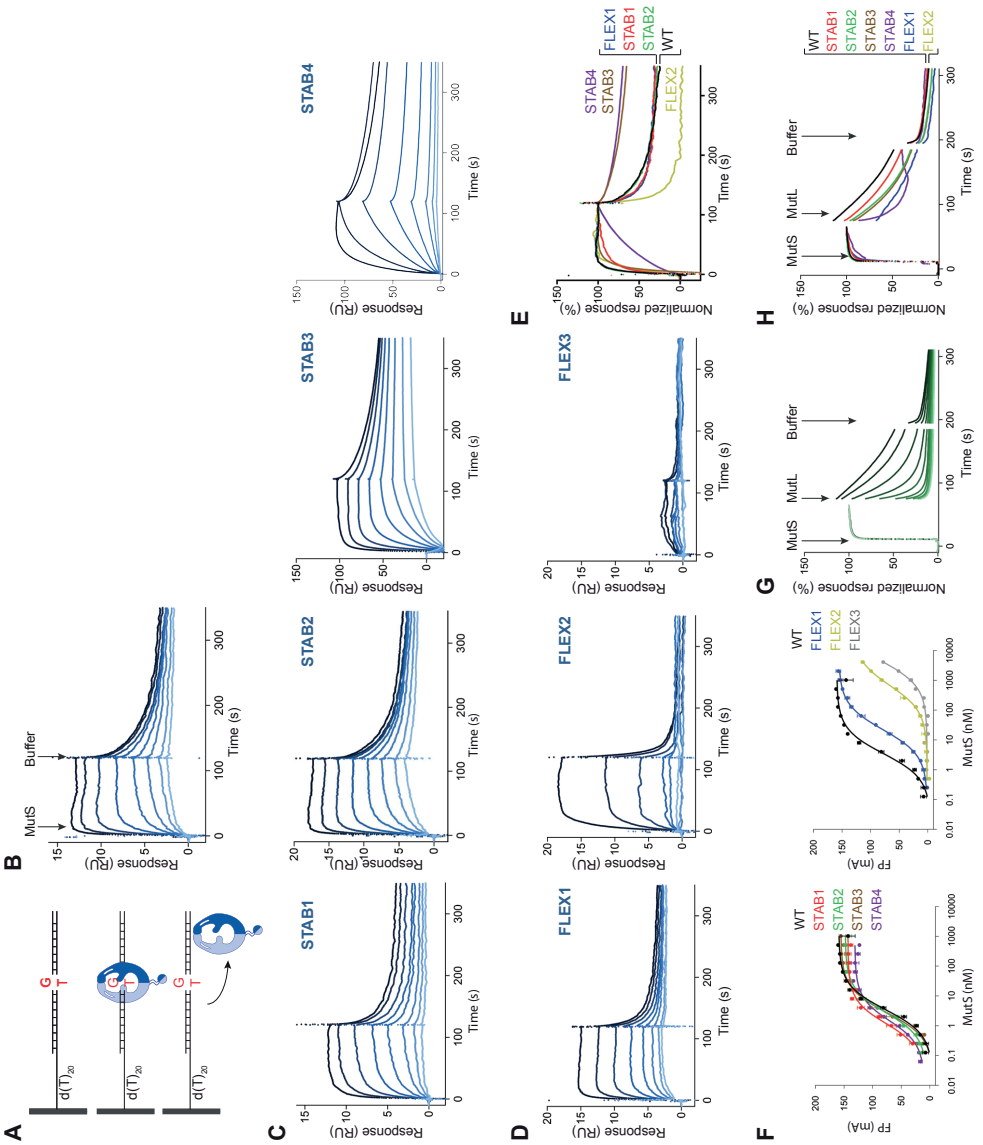

A
B
C
D
E
F
G
H
MutS
Buffer
MutL
Response (RU)
Time (s)
STAB1
STAB2
STAB3
STAB4
FLEX1
FLEX2
FLEX3
WT
Normalized response (%)
FP (mA)
MutS (nM)

Figure 3. Binding of MutS mutants to mismatched DNA and MutL. **A)** Schematic representation of SPR set-up. Biotinylated DNA consisting of 21 basepairs containing a GT mismatch and a Thymine-linker, immobilized on a streptavidin chip (top). MutS binds to the mismatch (middle) and will dissociate (bottom). **B)** MutS WT binding curve with concentrations ranging from 5-640 nM. Starting with 120 seconds injection of MutS, followed by buffer injection for 240 seconds. **C)** Binding curves for stabilizing mutants STAB1, STAB2, STAB3 and STAB4. **D)** Binding curves for flexible mutants FLEX1, FLEX2 and FLEX3. **E)** Binding curve of all MutS mutants and WT at 160 nM, normalized on maximum response at t=120 s. **F)** Equilibrium binding measurements using TAMRA-labeled 21-bp DNA containing a GT mismatch. Data points are averages between three measurements and error bars represent SEM. **G)** Binding of MutS and MutL on 100-bp oligo, normalized on maximum MutS binding. MutL WT (concentrations ranging from 0-2048 nM) binding curve to MutS WT (1200 nM). The signal at t=65-75 sec and t=185-195 sec was removed due to the signal noise at the start and end of MutL injection. **H)** Binding curves of MutL WT at concentration 2048 nM to all MutS constructs, normalized on maximum MutS binding (Supplemental Figure S4 and S5 for individual runs).

Analysis of sliding clamp formation and MutL recruitment

Next, we tested whether the change of the coiled-coil stability of MutS would affect clamp formation and MutL loading on DNA (21). This analysis was not performed for MutS FLEX3 as it did not bind DNA sufficiently well for these assays. First, we analyzed ATP-induced release of blocked-end DNA as proxy for sliding clamp formation. We did this for all MutS STAB and FLEX mutants by performing SPR experiments with 100-bp heteroduplex DNA in the presence of ATP. All MutS mutants show slower dissociation on closed-end DNA than on open-end DNA (Supplemental Figure S2), indicating that all mutants are able to form the ATP-induced sliding clamp, which is required for MutL binding.

Next, MutL binding experiments were performed using the same oligo and ATP concentration as mentioned above, but now with titrations of MutL (Figure 3G). The binding response of the highest MutL concentration was plotted (Figure 3H) (individual experiments, Supplemental Figure S3 and S4).

We observed that WT, STAB1, STAB2, STAB3 and FLEX1 have a similar K_d^{app} for MutL binding (Supplemental Figure S4), showing that loading of MutL is not affected for these mutants. MutS FLEX2 has little MutL binding, likely due to poor DNA binding (Figure 3D). Finally, MutS STAB4 shows poor MutL binding on open heteroduplex DNA, compared to MutS WT, but it is not impaired on blocked DNA (Supplemental Figure S4).

Table 2. Properties of the MutS variants used in this study.

	MutS Binding				MutL binding
Protein	Tm* (°C)	K_d^{app} in FP# (nM)	K_d^{app} in SPR # (nM)	k_{off}^{app} in SPR # (s^{-1})	K_d^{app} in SPR # (nM)
WT MutS	43.2 ± 0.6	3.2 ± 0.5	69 ± 4	0.016 ± 2.7 *10^{-5}	15 ± 1
Aimed to promote stable coiled-coil					
STAB1	42.1 ± 0.9	0.95 ± 0.2	76 ± 6	0.027 ± 5.1 *10^{-5}	18 ± 2
STAB2	43.2 ± 0.7	2.5 ± 0.3	62 ± 9	0.018 ± 2.8 *10^{-5}	16 ± 2
STAB3	44.4 ± 1.0	2.5 ± 0.3	52 ± 11	0.009 ± 1.3 *10^{-5}	17 ± 2
STAB4	43.0 ± 0.5	1.2 ± 0.2	110 ± 23	0.009 ± 1.4 *10^{-5}	11 ± 2
Aimed to promote instable coiled coil					
FLEX1	42.4 ± 1.0	22 ± 0.7	188 ± 17	0.029 ± 7.4 *10^{-5}	8.0 ± 1
FLEX2	41.7 ± 1.6	>670	978 ± 236	*N/A*	>1500
FLEX3	39.9 ± 0.8	>4600	N/A	N/A	N/A
Aimed to promote locked kink-out conformation					
LOCK1 – Crosslinked$	ND	ND	N/A	N/A	N/A
LOCK1$	38.6 ± 4.9	ND	28 ± 1	0.022 ± 3.8 *10^{-5}	N/A
WT$	ND	ND	30 ± 4	0.017 ± 2.4 *10^{-5}	
MIXED LOCK1$	ND	ND	93 ± 7.5	0.009 ± 1.2 *10^{-5}	N/A

* Standard deviation of three measurements are indicated after the ± sign;

Standard errors of fitting are indicated after the ± sign;

$ BSA was added to SPR running buffer at 1 mg/ml concentration

N/A: parameters could not be determined, *ND*: parameters were not determined.

As STAB4 binds mismatched DNA with an affinity in the same range as WT (respectively 110 ± 23 and 69 ± 4 nM), this could not explain the poor MutL recruitment on open DNA. We wondered if MutL binding itself was impaired, or whether the effect could be due to earlier steps.

To address this question, we further analyzed STAB4 behavior as a function of ATP. At a relatively high ATP concentration (32 μM) MutL stabilizes the WT protein on open-end DNA (Figure 4A). STAB4 is clearly impaired in this MutL-dependent stabilization, but we also noticed that STAB4 dissociates less than WT MutS (Figure 4A). Therefore, we analyzed dissociation from open-end DNA at different ATP concentrations (Figure 4B-E). We observed that STAB4 responds slower to ATP than WT (Figure 4B, C) indicating a defect in sliding clamp formation. Since sliding clamp formation and MutL binding are not impaired on blocked-end DNA we can conclude that STAB4 may be defective in the transition towards the MutL-activating state and that this affects MutL loading on open-end DNA.

Both monomers are able to kink-out and required for full DNA binding

In our crystal structure we observe two types of kink; one monomer kinking 'in' and one monomer kinking 'out'. We wondered whether this is a state that can bind DNA, or if this is only a conformation to 'open up' for DNA binding. To test this, we designed a mutant that can lock each monomer in the "kinked-out" state by crosslinking of two cysteines irreversibly at either side of the hinge within the monomer. We introduced cysteines E435C and R449C into the cysteine free variant of MutS (19,20), creating MutS LOCK1 (Figure 2C). Addition of $BM(PEG)_3$, to LOCK1 generates the crosslink within the monomer. Binding abilities of LOCK1 and crosslinked LOCK1 to a 100-bp oligo were first verified via gel filtration analysis (Supplemental Figure S5). A peak shift, between 100GT50 and LOCK1-100GT50, shows binding of LOCK1. However, this peak shift is absent when LOCK1 is crosslinked, indicating that crosslinked LOCK1 cannot bind 100GT50. To determine the kinetic parameters of LOCK1, with and without a crosslinker, to heteroduplex DNA, SPR experiments were performed. These data showed that crosslinked LOCK1 has a poor DNA affinity (Figure 5B) too poor for further analysis. Unmodified LOCK1 shows DNA-binding kinetics similar to MutS WT (compare Figure 5A and 5B) (Table 2).

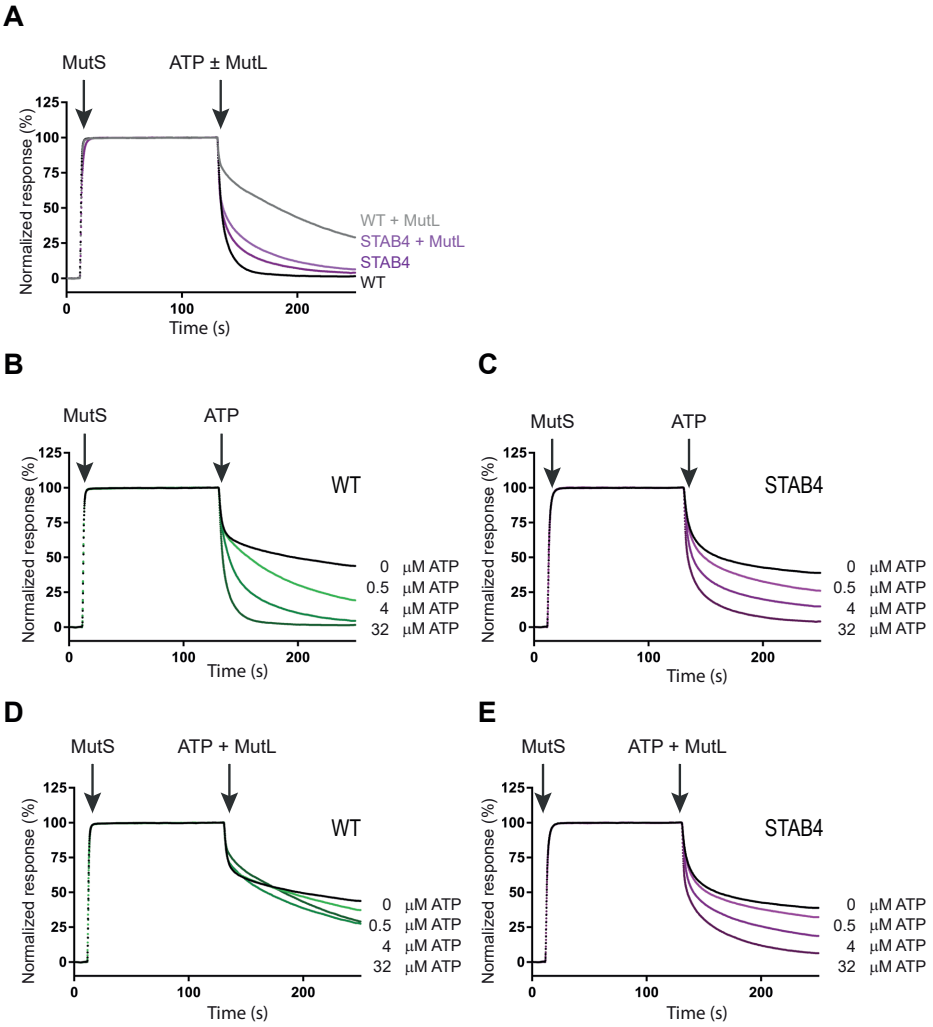


Figure 4. STAB4 is affected in ATP response and MutL binding. A) MutS (1200 nM) was injected for 120 seconds in the absence of ATP on 100GT50, followed by a second injection containing either 32 µM ATP or 32 µM ATP + 400 nM MutL. MutS WT followed by ATP induced release (black) and bound by ATP and MutL (gray). MutS STAB4 with ATP (purple) and co-injected with MutL (pink). Binding is normalized on maximum MutS response**. B)** MutS injection for 120 seconds (1200 nM), followed by 120 second injection of buffer (black) or 0.5 µM, 4 µM or 32 µM ATP respectively in light to dark curves. **C)** MutS STAB4 binding and dissociation, details see B). **D)** MutS injection for 120 seconds, followed by 120 second injection of buffer (black) or 0.5 µM, 4 µM or 32 µM ATP supplemented with 400 nM MutL, respectively in light to dark curves. **E)** MutS STAB4 and MutL binding and dissociation, details see D).

To test the importance of both monomers in DNA binding, we generated heterodimers of unmodified and crosslinked LOCK1 by creating a mixed sample of unmodified LOCK1 and crosslinked LOCK1 in equimolar ratios to allow the heterodimers to form. We used two-fold higher monomer concentrations, to account for an expected mixture of unmodified homodimer LOCK1, heterodimer and homodimer crosslinked LOCK1, distribution of 1:2:1. Although the actual distribution is unknown, the resulting mixture shows kinetic behaviour of the mixture that is different from either of the homodimers (Figure 5C, 5B) indicating that heterodimers were produced and that the crosslinked coiled coil affects the DNA binding properties.

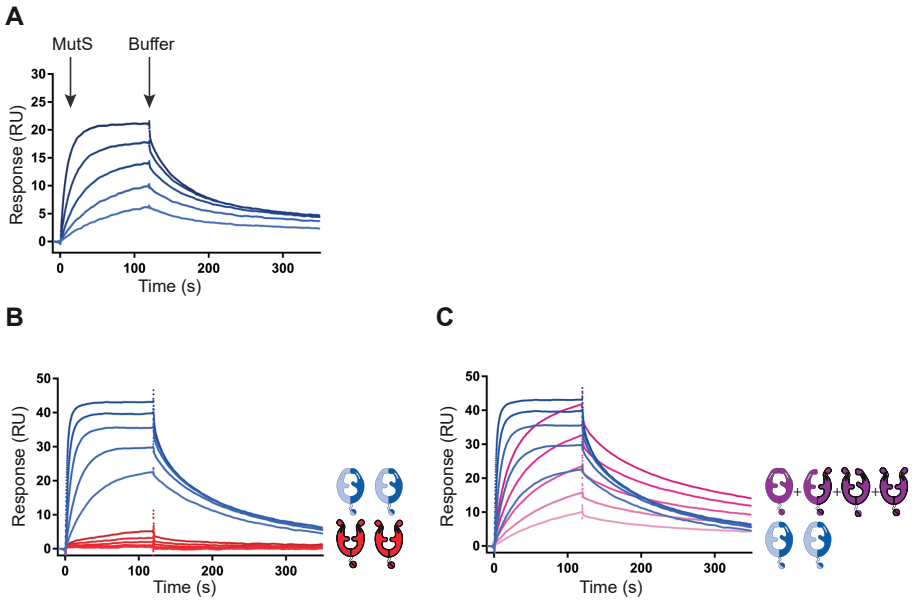


Figure 5. Stabilized kink (LOCK1) slows down kinetics of mismatched DNA binding. All SPR curves were performed in the presence of 1 mg/ml BSA. All runs started with 120 seconds injection of MutS, followed by buffer injection for 240 seconds **A)** SPR titration curve of MutS WT onto mismatched DNA with concentrations 5-80 nM. **B)** SPR titration curve of MutS LOCK1 onto mismatched DNA with concentrations 8-128 nM.5 In blue MutS LOCK1, in red LOCK1 crosslinked with $BM(PEG)_3$. **C)** Homodimer LOCK1 (see panel B, blue) together with a 1:1 mixture of non-crosslinked and crosslinked (purple). Both constructs have a titration of 8-128 nM, making the total MutS concentration 16-256 nM.

We conclude that kinking-movements are important in DNA loading since locked outward-kink LOCK1 shows very poor DNA binding and that heterodimers bind mismatched DNA with a weaker K_d^{app} but a slower k_{off}^{app} compared to non-crosslinked LOCK1 indicating that both monomers are important for DNA binding.

DISCUSSION

MutS is known to possess conformational freedom of its clamp domains in solution, as observed in SAXS studies (8) and in a previous crystal apo-structure of *Taq* MutS (11). However, the nature of this disorder remained unclear. In this work, we have shown that kinking of the helices in the lever domains of MutS can allow for movement of the clamp domains, as observed in the new apo-structure. Our measurements show that perturbation of the helical fold in these regions influences DNA binding properties of MutS. Therefore, we hypothesize that kinking of coiled-coils as observed in this new crystal structure precedes DNA binding. Finally, crosslinked LOCK1 shows that both monomers can kink outward and are involved in full DNA binding.

Our structure is the first apo-structure for *E.coli* MutS. Recently, an apo-structure of MutS *N. gonorrhoeae* was reported (29). The conformation of the clamp domains in ADP-bound MutS *N. gonorrhoeae* is rather different than in our structure. It has straight coiled-coils, without any kink in the lever domains, but with a somewhat more open conformation compared to mismatch-bound MutS.

All three MutS apo-structures are symmetrical in their nucleotide-state. DNA-free *Taq* MutS is lacking any nucleotide, our *E.coli* MutS structure has two ADP bound, as expected at this concentration of ADP (28,32). Both apo-structures of *N. gonorrhoeae* also have two nucleotides bound. *N. gonorrhoeae* MutS has been crystallized in the presence of ADP (pdb: 5yk4). Surprisingly, on close inspection of the electron density maps, there is additional density visible, which could be explained by a third phosphate group. Possibly ATP or AMP-PNP were present, as in the second *N. gonorrhoeae* published structure (pdb: 5x9w) (29). It is possible that this nucleotide difference explains the differences in the lever domain conformation, relative to our ADP-bound kinked structure. Alternatively, the

kinking is a stochastic process and the open form switches continually between straight, kinked-in and kinked out states and the crystal traps a defined state.

Surprisingly, despite the difference in nucleotide state, all three apo-structures of MutS are lacking Mg^{2+} while it was present in each crystallization set-up. It may be explained by the fact that magnesium does not play a role in the equilibrium of ADP binding (32).

We made a series of mutants to analyze the importance of the defined kink for MutS function. Several mutants that were designed to improve the stability of the coiled-coil (STAB3, 4), resulted in DNA binding with slower kinetics. This could indicate that 'opening' up of the dimer indeed requires the ability to kink the helices. The Pro/Gly/Pro motifs at the kinking sites in the FLEX-mutants were designed to destabilize the coiled-coil. These mutations had a very clear effect on DNA binding, where mutations in either of the two helices resulted in weaker binding to DNA, mainly due to faster release. In contrast to STAB3 and STAB4, the effect of Pro/Gly/Pro motif on DNA release is not balanced out by a comparable effect on association. Possibly, the effect on the helical stability is too big to form a stable MutS:DNA complex. The effect of introducing the Pro/Gly/Pro motif is less for residues 441-443 than for residues 515-517, in-line with the smaller sequence change since residue 442 is already a glycine in the native protein.

We found that an intra-domain crosslink in the LOCK1 mutant severely affects DNA binding. This shows that kinking of the lever domains is required for DNA binding. Since the dissociation rate of mixed LOCK1 is slower, it seems that both monomers need to be flexible for normal DNA binding and release.

All our mutants were able to form a sliding clamp on blocked DNA and able to recruit MutL. However, on open heteroduplex DNA, STAB4 did no longer bind MutL, although it still slides to some extent. This seems to be primarily caused by its weaker response to ATP, which may have slowed down the ability to arrive at the conformation required for MutL binding. Since it was still able to recruit MutL on closed-end DNA the slower sliding clamp formation may be the main defect in this mutant.

Our data show that the ability to rearrange the coiled-coil levers is important for MutS function. This is not necessarily a surprise, as flexibility of the coiled-coil-region was already shown to be important for related ABC-ATPases, such as the

SMCs and RAD50 (15-18). What is surprising though, is that this flexibility appears as a sharp kink, rather than a gradual bend, although we believe that the precise kink angles may vary in the absence of crystal contacts. Sharp kinks have been observed in the coiled-coils of the SMC proteins by AFM and rotary shadowing EM (15,33). During revision of this manuscript, a paper appeared that revealed the presence of a defined kink in the structure of E.coli MukBEF and demonstrate its presence in *S. cerevisiae*'s cohesin (34). The authors then show that discontinuity in the coiled coils of related SMC proteins is a conserved feature in the SMC family suggesting that the so-called elbow movement is important for function. Together with our findings, this could open up a new mode of action for ABC-ATPase.

An interesting question is whether the hinge can respond by internal (e.g. ATP binding) or external (e.g. DNA binding) forces. Literature analysis does not give much precedent, with the exception of the 'buckling' observed in thermosensitive K(2P) channels coiled-coils and long α-helix filaments as a response to external stress to the ends of the helices (35,36). However, this type of force seems unlikely for MutS. Although it is theoretically possible that crystal contacts contributed to the kinked MutS state, our data indicate that the ability to move these regions affects DNA binding. This suggests that the flexibility of the hinge region is used in the binding process. How this is organized will need further research.

In conclusion, we have found unexpected defined hinge points in the coiled-coil helices in the lever domains of MutS. It suggests a manner in which this protein can subtly achieve flexibility before adapting to the more ordered and probably favorable DNA-bound state. Some mutations in the hinging region have clear effects on DNA binding properties of MutS, suggesting that there is a fine balance of the helical stability of the lever domains for proper DNA mismatch repair.

Data Availability

Access PDB code for the crystal structure is 6I5F.

Supplementary Data statement

Supplementary Data are available at NAR Online.

Funding

This work was supported by the Netherlands Organisation for Scientific Research (NWO) 714.016.002 (TOP) and 711.011.011 (ECHO), CGC.nl as well as EU FP7 project MM2M.

Acknowledgements

We thank Robbie Joosten for critical analysis of the structure of MutS *N. gonorrhoeae* (PDB: 5yk4) and assistance in the refinement and Susanne Bruekner for her suggestion to lock MutS using a double cysteine mutant. Joyce Lebbink and Meindert Lamers provided useful comments on the manuscript.

Author contributions

The crystal structure was solved and refined by AF, with assistance from TKS. FSG, SD and DBK performed mutant design and cloning. HHKW, DBK and SD performed purification of mutants. SPR experiments were performed by DBK and AF. Protein stability experiments were performed by AF. FP experiments were performed by FSG and SD. Experiments regarding double cysteine mutants were performed by DBK. TKS initiated and supervised the study and FSG and DBK wrote the manuscript with contributions of TKS and AF and critical input from all authors.

REFERENCES

1. Kunkel, T.A. and Erie, D.A. (2005) DNA mismatch repair. *Annual review of biochemistry*, **74**, 681-710.
2. Jiricny, J. (2013) Postreplicative mismatch repair. *Cold Spring Harbor perspectives in biology*, **5**, a012633.
3. Lynch, H.T. and de la Chapelle, A. (1999) Genetic susceptibility to non-polyposis colorectal cancer. *Journal of medical genetics*, **36**, 801-818.
4. Yamaguchi, M., Dao, V. and Modrich, P. (1998) MutS and MutL activate DNA helicase II in a mismatch-dependent manner. *The Journal of biological chemistry*, **273**, 9197-9201.
5. Acharya, S., Foster, P.L., Brooks, P. and Fishel, R. (2003) The coordinated functions of the E. coli MutS and MutL proteins in mismatch repair. *Molecular cell*, **12**, 233-246.
6. Calmann, M.A., Nowosielska, A. and Marinus, M.G. (2005) Separation of mutation avoidance and antirecombination functions in an Escherichia coli mutS mutant. *Nucleic acids research*, **33**, 1193-1200.
7. Mendillo, M.L., Hargreaves, V.V., Jamison, J.W., Mo, A.O., Li, S., Putnam, C.D., Woods, V.L., Jr. and Kolodner, R.D. (2009) A conserved MutS homolog connector domain interface interacts with MutL homologs. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 22223-22228.
8. Groothuizen, F.S., Fish, A., Petoukhov, M.V., Reumer, A., Manelyte, L., Winterwerp, H.H., Marinus, M.G., Lebbink, J.H., Svergun, D.I., Friedhoff, P. *et al.* (2013) Using stable MutS dimers and tetramers to quantitatively analyze DNA mismatch recognition and sliding clamp formation. *Nucleic acids research*, **41**, 8166-8181.
9. Acharya, S., Wilson, T., Gradia, S., Kane, M.F., Guerrette, S., Marsischky, G.T., Kolodner, R. and Fishel, R. (1996) hMSH2 forms specific mispair-binding complexes with hMSH3 and hMSH6. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 13629-13634.
10. Lamers, M.H., Perrakis, A., Enzlin, J.H., Winterwerp, H.H., de Wind, N. and Sixma, T.K. (2000) The crystal structure of DNA mismatch repair protein MutS binding to a G x T mismatch. *Nature*, **407**, 711-717.
11. Obmolova, G., Ban, C., Hsieh, P. and Yang, W. (2000) Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. *Nature*, **407**, 703-710.
12. Warren, J.J., Pohlhaus, T.J., Changela, A., Iyer, R.R., Modrich, P.L. and Beese, L.S. (2007) Structure of the human MutSalpha DNA lesion recognition complex. *Molecular cell*, **26**, 579-592.
13. Gupta, S., Gellert, M. and Yang, W. (2011) Mechanism of mismatch recognition revealed by human MutSbeta bound to unpaired DNA loops. *Nature structural & molecular biology*, **19**, 72-78.
14. Lupas, A. (1996) Coiled coils: new structures and new functions. *Trends in biochemical sciences*, **21**, 375-382.

15. Noort van, J., Heijden van der, T., Jager de, M., Wyman, C., Kanaar, R., Dekker, C. (2003) The coiled-coil of the human Rad50 DNA repair protein contains specific segments of increased flexibility. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 7581–7586.

16. Eeftens, J.M., Katan, A.J., Kschonsak, M., Hassler, M., Wilde de, L., Dief, E.M., Haering, C.H., Dekker, C. (2016) Condensin Smc2-Smc4 Dimers Are Flexible and Dynamic. *Cell reports*, **14**, 1813-1818.

17. Kamada, K., Su'etsugu, M., Takada, H., Miyata, M., Hirano, T. (2017) Overall Shapes of the SMC-ScpAB Complex Are Determined by Balance between Constraint and Relaxation of Its Structural Parts. *Cell*, **25**, 603–616.

18. Kamada, K., Miyata, M., Hirano, T. (2013) Molecular Basis of SMC ATPase Activation: Role of Internal Structural Changes of the Regulatory Subcomplex ScpAB. *Cell*, **21**, 581–594.

19. Manelyte, L., Urbanke, C., Giron-Monzon, L. and Friedhoff, P. (2006) Structural and functional analysis of the MutS C-terminal tetramerization domain. *Nucleic acids research*, **34**, 5270-5279.

20. Winkler, I., Marx, A.D., Lariviere, 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. *The Journal of biological chemistry*, **286**, 17326-17337.

21. Groothuizen, F.S., Winkler, I., Cristovao, M., Fish, A., Winterwerp, H.H., 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.

22. Battye, T.G., Kontogiannis, L., Johnson, O., Powell, H.R. and Leslie, A.G. (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta crystallographica. Section D, Biological crystallography*, **67**, 271-281.

23. Evans, P. (2006) Scaling and assessment of data quality. *Acta crystallographica. Section D, Biological crystallography*, **62**, 72-82.

24. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. (2007) Phaser crystallographic software. *J Appl Crystallogr*, **40**, 658-674.

25. Bricogne G, B.E., Brandl M, Flensburg C, Keller P, Paciorek P, Roversi P, Sharff A, Smart O, Vonrhein C, Womack T. (2010) BUSTER version 2.9. Cambridge, United Kingdom: Global Phasing Ltd.

26. Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F. and Vagin, A.A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta crystallographica. Section D, Biological crystallography*, **67**, 355-367.

27. Joosten, R.P., Salzemann, J., Bloch, V., Stockinger, H., Berglund, A.C., Blanchet, C., Bongcam-Rudloff, E., Combet, C., Da Costa, A.L., Deleage, G. *et al.* (2009) PDB_REDO: automated re-refinement of X-ray structure models in the PDB. *J Appl Crystallogr*, **42**, 376-384.

28. Monti, M.C., Cohen, S.X., Fish, A., Winterwerp, H.H., Barendregt, A., Friedhoff, P., Perrakis, A., Heck, A.J., Sixma, T.K., van den Heuvel, R.H. *et al.* (2011) Native mass spectrometry provides direct evidence for DNA mismatch-induced regulation of asymmetric nucleotide binding in mismatch repair protein MutS. *Nucleic acids research*, **39**, 8052-8064.

29. Nirwal, S., Kulkarni, D.S., Sharma, A., Rao, D.N. and Nair, D.T. (2018) Mechanism of formation of a toroid around DNA by the mismatch sensor protein. *Nucleic acids research*, **46**, 256-266.

30. Natrajan, G., Lamers, M.H., Enzlin, J.H., Winterwerp, H.H., Perrakis, A. and Sixma, T.K. (2003) Structures of Escherichia coli DNA mismatch repair enzyme MutS in complex with different mismatches: a common recognition mode for diverse substrates. *Nucleic acids research*, **31**, 4814-4821.

31. Arndt, K.M., Pelletier, J.N., Muller, K.M., Pluckthun, A. and Alber, T. (2002) Comparison of in vivo selection and rational design of heterodimeric coiled coils. *Structure (London, England : 1993)*, **10**, 1235-1248.

32. Lebbink, J.H., Fish, A., Reumer, A., Natrajan, G., Winterwerp, H.H. and Sixma, T.K. (2010) Magnesium coordination controls the molecular switch function of DNA mismatch repair protein MutS. *The Journal of biological chemistry*, **285**, 13131-13141.

33. Diebold-Durand, M., Lee, H., Ruiz Avila, L.B., Noh, H., Shin, H., Im, H., Bock, F.P., Burmann, F., Durand, A., Basfeld, A., Ham,S., Basquin, J., Oh, B.,Gruber, S. (2017) Structure of Full-Length SMC and Rearrangements Required for Chromosome Organization. *Cell*, **67**, 334–347.

34. Bürmann, F., Lee, B.-G., Than, T., Sinn, L., O'Reilly, F.J., Yatskevich, S., Rappsilber, J., Hu, B., Nasmyth, K. and Löwe, J. (2019) A folded conformation of MukBEF and cohesin. *Nature structural & molecular biology*, **26**, 227-236.

35. Lolicato, M., Riegelhaupt,P.M., Arrigoni,C., Clark, K.A., Minor, D.L. (2014) Transmembrane Helix Straightening and Buckling Underlies Activation of Mechanosensitive and Thermosensitive K2P Channels. *Neuron*, **84**, 1198–1212.

36. Palenčár, P.a.B., T. (2014) Buckling transition in long α-helices. *The Journal of Chemical Physics*, **141**, 174901-174912.

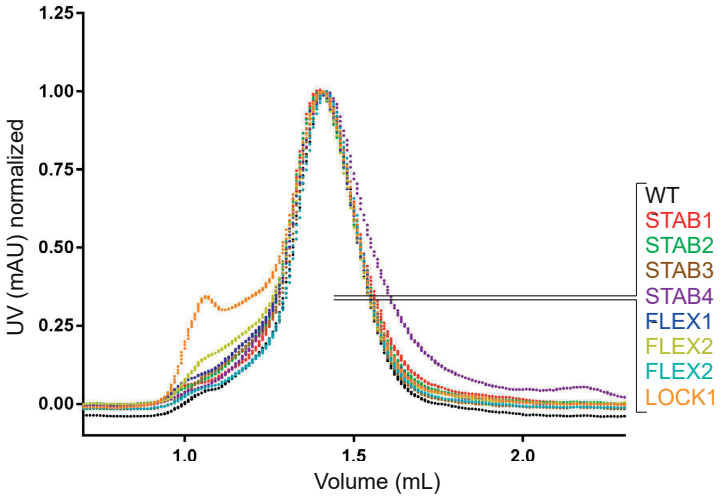


Figure S1. Analytical gel filtration analysis for MutS WT and mutants. All MutS mutants were analyzed over an S200 increase analytical column. The similar elution profiles indicate that the mutants are well-folded.

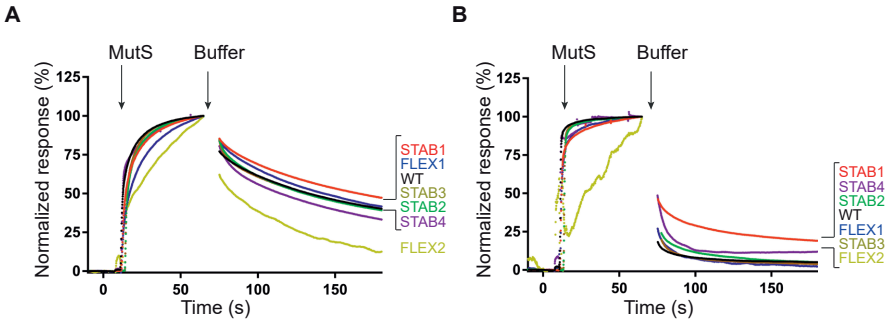


Figure S2. ATP-induced release of wild type and mutant MutS from a mismatch. SPR experiments with 100bp-oligo heteroduplex DNA. MutS injection (60 s) followed by buffer injection (120 s) all in the presence of 1 mM ATP. The ATP allows MutS WT and mutants to undergo a conformational change, which results sliding of MutS. The signal at t=65-75 s was removed, to remove excessive noise at the end of the injection from the graphs. **A)** Blocked-end oligo, showing a slower dissociation indicating sliding. **B)** Open-end oligo.

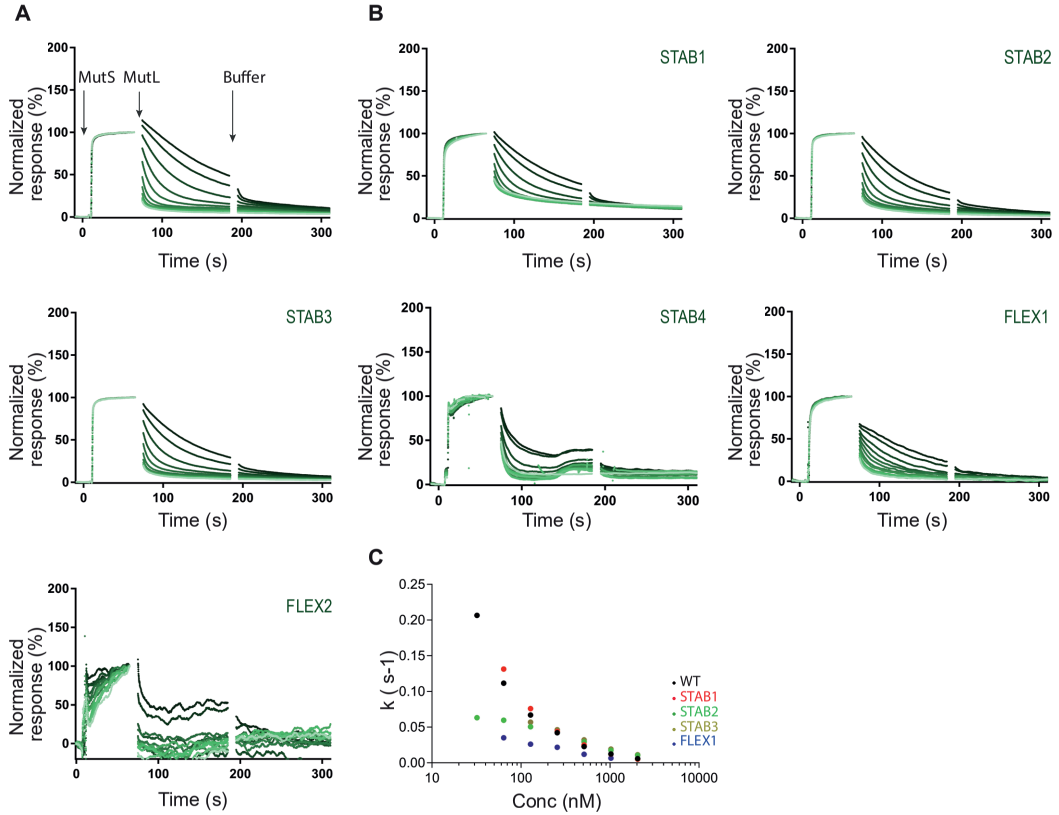


Figure S3. Binding of MutS mutants to MutL on open-end DNA. A) MutS WT with MutL titrations. MutS injection (1200 nM) for 60 seconds, followed by MutL injection 120 seconds in titration concentrations (light to dark respectively 0 nM – 2048 nM). The curves are normalized on MutS binding. The signal at t=65-75 seconds and t=185-195 seconds was removed due to the noise at the beginning and at the end of MutL injection. **B)** MutL binding for MutS mutant STAB1, STAB2, STAB3, STAB4, FLEX1 and FLEX2. For details, see A. **C)** Non-linear one phase decay fit of the MutL injection phase. MutS WT, STAB1, STAB2, STAB3 and FLEX1 all show a comparable dissociation rate as a function of MutL concentration. MutL binding for FLEX2 is not plotted since binding of FLEX2, and therefore MutL, were too poor. STAB4 could not be plotted since the dissociation was too fast and time resolution of SPR doesn't allow a correct fitting.

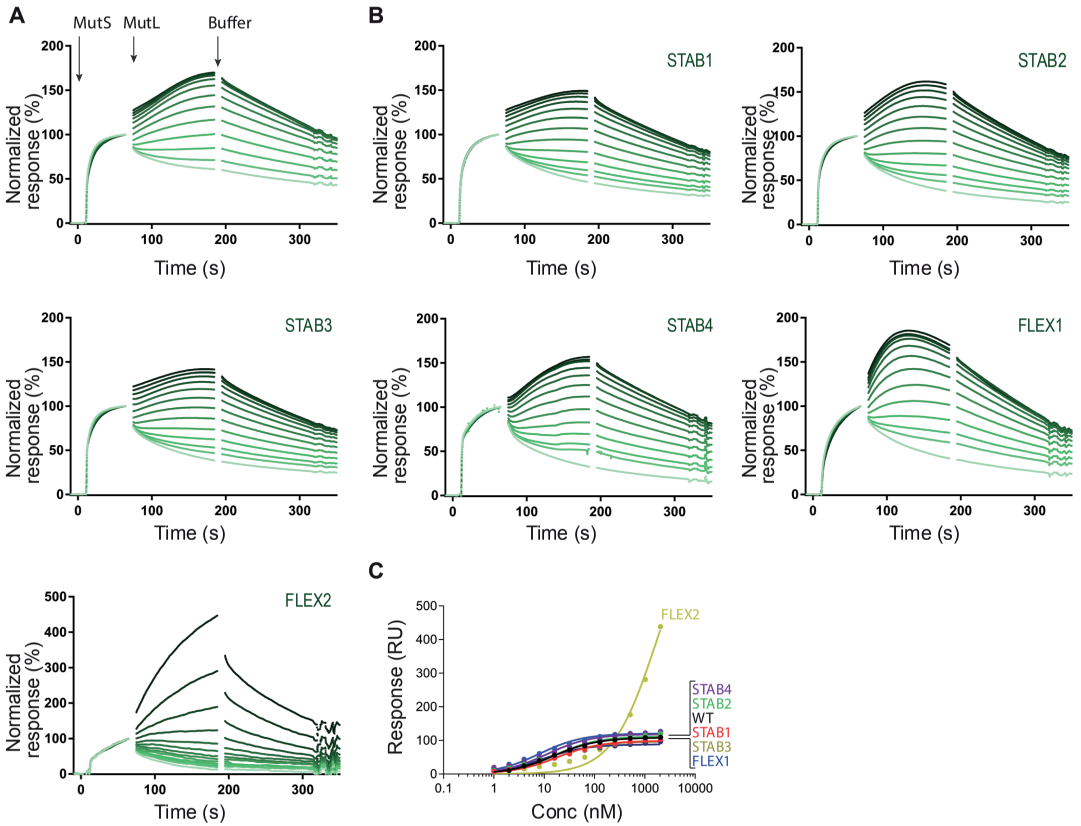


Figure S4. Binding of MutS mutants to MutL on closed-end DNA. **A** MutS WT with MutL titrations. MutS injection (1200 nM) for 60 seconds, followed by MutL injection 120 seconds in titration concentrations (light to dark respectively 0 nM – 2048 nM). The curves are normalized on MutS binding. The signal at t=65-75 seconds and t=185-195 seconds was removed due to the noise at the beginning and at the end of MutL injection **B)** MutL binding for MutS STAB1, STAB2, STAB3, STAB4, FLEX1 and FLEX2. All mutants show binding properties compared to WT except MutS FLEX2. **C)** After normalization on MutS binding, MutL binding responses were plotted as function of the concentration and fitted with one site equation (Table 2). Almost all mutants have a similar as WT, however, MutS FLEX2 is an outlier due to its poor DNA binding.

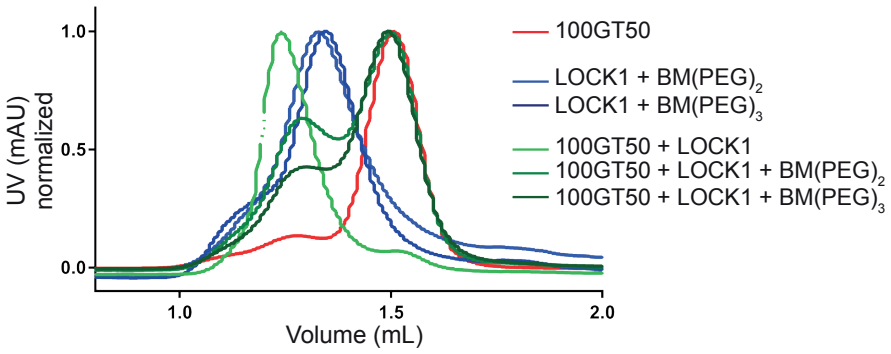


Figure S5. **Merged analytical gel filtration runs for crosslinked MutS LOCK1**. In red 100GT50. In blue and dark blue MutS LOCK1 crosslinked with respectively BM(PEG)2 and BM(PEG)3. The green curves are MutS LOCK1 co-injected with 100GT50. From light green to dark green are respectively MutS LOCK1, MutS LOCK1 crosslinked with BM(PEG)2 and MutS LOCK1 crosslinked with BM(PEG)3.

CHAPTER 3

The selection process of licensing a DNA mismatch for repair

Rafael Fernandez-Leiro[1,2]
Doreth Bhairosing-Kok[3]
Vladislav Kunetsky[4]
Charlie Laffeber[5]
Herrie H. Winterwerp[3]
Flora Groothuizen[3]
Alexander Fish[3]
Joyce H. G. Lebbink[5,6]
Peter Friedhoff[4]
Titia K Sixma[3]
Meindert H. Lamers[1,7]

[1]MRC Laboratory of Medical Research, Cambridge, United Kingdom.
[2]Spanish National Cancer Research Centre, Madrid, Spain.
[3]Division of Biochemistry and Oncode Institute, Netherlands Cancer Institute, Amsterdam, The Netherlands.
[4]Institute for Biochemistry, Justus-Liebig University, Giessen, Germany.
[5]Department of Molecular Genetics and Oncode Institute, Erasmus University Medical Center, Rotterdam, the Netherlands.
[6]Department of Radiation Oncology, Erasmus University Medical Center, Rotterdam, The Netherlands,
[7]Department of Cell and Chemical Biology, Leiden University Medical Center, Leiden, The Netherlands.

Accepted in Nature Structural & Molecular Biology

ABSTRACT

DNA mismatch repair detects and removes mismatches from DNA by a conserved mechanism, reducing the error rate of DNA replication a 100-1000 fold. In this process, MutS homologs scan DNA, recognize mismatches and initiate repair. How the MutS homologs selectively license repair of a mismatch among millions of matched base pairs is not understood. Here we present four cryo-EM structures of *E. coli* MutS that provide snapshots from scanning homoduplex DNA, to mismatch binding and MutL activation, via an intermediate state. During scanning the homoduplex DNA forms a steric block that prevents MutS from transitioning into the MutL-bound clamp-state, which can only be overcome through kinking of the DNA at a mismatch. Structural asymmetry in all four structures indicates a division of labour between the two MutS monomers. Together, these structures reveal how a small conformational change from the homoduplex to heteroduplex-bound MutS acts as a licensing step that triggers a dramatic conformational change that enables MutL binding and initiation of the repair cascade.

INTRODUCTION

The evolutionary conserved DNA mismatch repair (MMR) process protects against incorporation of errors during replication, which is critical for genome maintenance. It also plays roles in regulation of recombination, triplet repeat expansion and DNA damage signaling (Jiricny, 2013; Li et al., 2016). The initiation of mismatch repair only starts when MutS recognizes a single mismatch among millions of correctly matched base pairs (bp).

Existing structures of MutS homologs bound to mismatched DNA present two MutS monomers forming an oval-shaped dimer that encircles the DNA (Gupta et al., 2011; Lamers et al., 2000; Natrajan et al., 2003; Obmolova et al., 2000; Warren et al., 2007). The mismatch is exposed through a 60° kink of the DNA at the site of the mismatch. Only one of the two MutS monomers interacts with the mismatch and inserts a conserved phenylalanine into the DNA adjacent to the mismatched base. The second monomer interacts with the DNA ~6 bp away from the mismatch in a sequence-independent manner. This mode of mismatch recognition is conserved between different mismatches (Natrajan et al., 2003), and species (Gupta et al., 2011; Warren et al., 2007). Upon mismatch binding, MutS undergoes an ATP-dependent conformational change that transforms MutS into a sliding clamp that can only dissociate from open DNA ends (Blackwell et al., 2001; Gradia et al., 1999) or ssDNA (Heo et al., 2007; Jeong et al., 2011). Only in this sliding clamp conformation will MutS recruit MutL and initiate repair (Acharya et al., 2003; Schofield et al., 2001). In *E. coli*, MutL subsequently activates the endonuclease MutH that creates a nick in the newly synthesized strand on hemi-methylated GATC sites (Hall and Matson, 1999; Junop et al., 2003). In other species, including eukaryotes, the endonuclease activity resides in the MutL homologs themselves (Fukui et al., 2008; Kadyrov et al., 2006) and is directed to the newly synthesized strand by the DNA sliding clamp (b or PCNA) (Pillon et al., 2011; Pluciennik et al., 2010). The newly synthesized strand is then excised past the mismatch after which a DNA polymerase can resynthesize the removed stretch of DNA.

Before MutS finds a mismatch, it scans millions of matched base pairs without initiating repair. How this extreme high-fidelity of mismatch repair initiation is achieved is not known. In this work we present four cryo-EM structures of MutS in the scanning state (homoduplex-bound), in the mismatch-recognition state (heteroduplex-bound), an intermediate state and the mismatch repair initiation

state (MutL-bound clamp state). Together these structures reveal how homoduplex DNA forms a steric block that prevents MutS from transitioning into the activated, MutL recruiting clamp state, and that mismatch binding overcomes this steric block. The structures furthermore generate a 'movie' that reveals the mechanisms that enable MutS to exert its many roles, from searching for and recognition of a mismatch to recruitment of MutL and activation of the repair cascade.

Four cryo-EM structures provide a movie of Mismatch repair initiation.

We determined four cryo-EM structures of MutS in sequential states of the mismatch repair process: 1) bound to perfectly matched DNA (homoduplex DNA), 2) bound to DNA carrying a single mismatch (heteroduplex DNA), 3) an intermediate state between mismatch-bound and the clamp state, and 4) the clamp state, where MutS is bound to DNA and the N-terminal domain of MutL ($MutL^{LN40}$: residues 1-331 (Ban and Yang, 1998)) (Figure 1A-H). The structure of MutS on homoduplex DNA was determined using a linearized plasmid (Figure 1I), on which the MutS molecules can be seen like beads on a string (Figure 1M). The structures of both mismatch-bound MutS and $MutL^{LN40}$-bound MutS were derived from a single dataset using crosslinked MutS-$MutL^{LN40}$ (Groothuizen et al., 2015) (Figure 1N) and a 61 bp mismatched DNA substrate, which was end-blocked with streptavidin (Figure 1J, L). $MutL^{LN40}$ is well defined in the clamp state but it is not visible in the mismatch-bound structure, indicating that $MutL^{LN40}$ is flexible in this structure. Finally, the intermediate-state structure was obtained using a 50 bp homoduplex DNA substrate that was not end-blocked (Figure 1K) and where MutS was found located near the end of the DNA substrate, trapped in what appears to be an intermediate state between mismatch-bound and MutL-bound. For details on data collection, processing and refinement, see Figure S1-S4.

In all four structures, the DNA is clamped between the two MutS monomers, but adopts different conformations and positions. On homoduplex DNA, the DNA is straight and held between the mismatch binding domains (residues 1-115) and clamp domains (residues 116-266) of both MutS monomers (for domain organization see Figure 1P). In the mismatched-bound MutS structure too, the DNA is held between the mismatch and clamp domains, but the DNA is kinked by 60° and DNA binding domains have rearranged with respect to one another (see Figure 2 for more details). This state corresponds to the crystal structure

of *E. coli* MutS on mismatched DNA (Lamers et al., 2000) and the latter can be placed into the cryo-EM map with only minor adjustments (Figure S2A-B). In the $MutL^{LN40}$-bound structure the DNA is straight again but has moved ~25 Å (i.e. more than one width of DNA) downwards (in the orientation of Figure 1A-D) towards the ATPase domains. The mismatch and connector domain show a dramatic repositioning in the $MutL^{LN40}$-bound structure: the connector domain has rotated ~180° downwards, while the mismatch domain has become flexible and is not observed in the cryo-EM map (Figure 1H). The intermediate state structure adopts a conformation in-between the mismatch-bound and $MutL^{LN40}$-bound structure. In one monomer the mismatch domain remains in contact with the DNA and is half-way between the mismatch-bound and $MutL^{LN40}$-bound structure (Figure 1G), while in the other monomer the mismatch and connector domains have moved to a position identical to the $MutL^{LN40}$-bound structure.

Between the four structures, the distance between the two MutS monomers changes. This can be measured by the distance between residues phenylalanine 400 that are located on the outside of the MutS dimer (marked in red spheres in Figure 1A-D), which decreases from a maximum of 98 Å (in the scanning state structure) to 63 Å (in the intermediate structure). Finally, in response to the movement of the DNA, the clamp domains move up and down by ~ 9Å with respect to the ATPase domains located at the other end of the MutS dimer (Figure 1O). Combined, these four structures provide a 'movie' of how MutS scans for mismatches, licenses for repair and transforms into a novel conformation that recruits MutL (Supplementary Movie 1). The molecular mechanism and implications of the transitions between the four states will be described below.

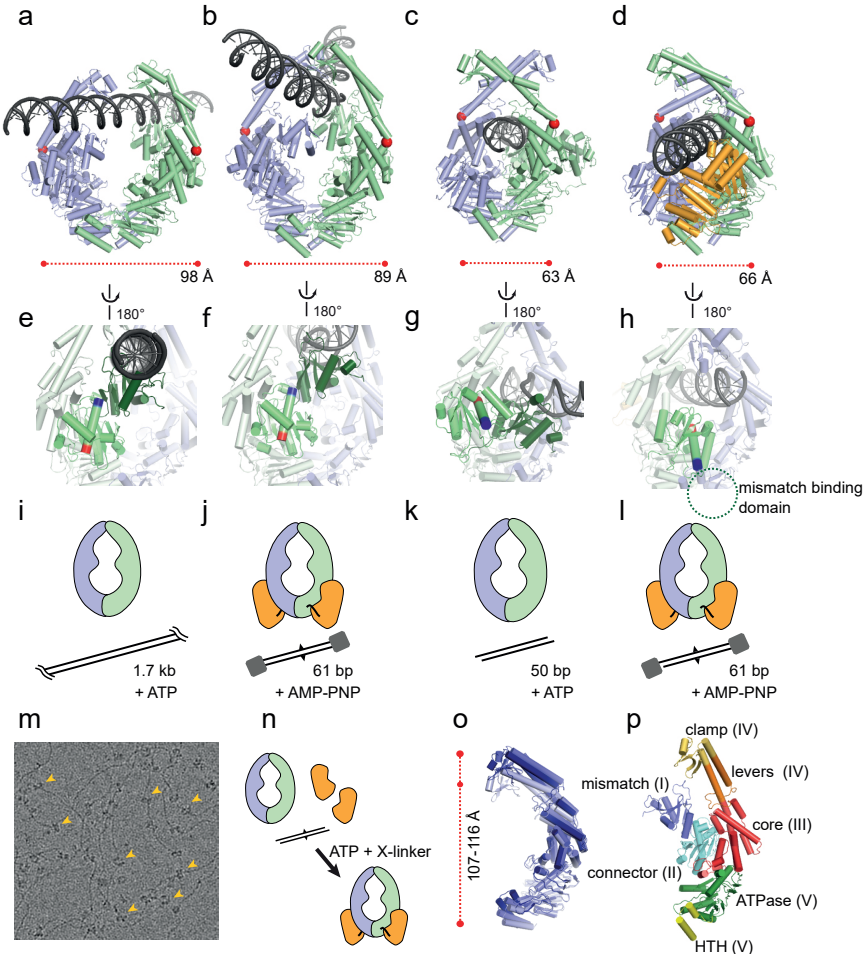


Figure 1. Structures of MutS at consecutive steps of the repair cascade. Structures of **a**, scanning MutS, **b**, mismatch-bound MutS, **c**, transition state MutS and **d**, MutL^{LN40}-bound MutS. Monomer A is shown in green, monomer B in blue, MutL^{LN40} in orange and DNA in black. Residues 400 are marked in red spheres, and their distance shown below with a dashed red line. **e-h**, Reverse view highlighting the conformational changes of mismatch and connector domains of monomer A. Mismatch domain is shown in dark-green, connector domain in lighter green. The N- and C-termini of the central helix (residues 231-248) are shown in red and blue, respectively. **i-l**, Schematic representation of the assembly of the four structures. MutS monomers show in in blue and green, MutL^{LN40} in orange and DNA in double black lines. The mismatch is indicated with black triangles and the crosslinker between MutS and MutL^{LN40} with a short black line. **m**, Micrograph of scanning MutS showing 'beads on a string' of MutS on DNA. Yellow arrows highlight some of the particles. **n**, Schematic representation of the assembly of cross-linked MutS-MutL^{LN40}. **o**, Overlay of monomer B from four structures showing the vertical expansion/contraction of the MutS molecule (mismatch and connect domain are omitted for clarity). **p**, The domains of MutS marked in different colors and indicated by name and number.

Scanning MutS shows dynamic interaction with DNA

On average MutS scans millions of base pairs before it encounters a mismatch. Single molecule analysis of yeast MutSa (Gorman et al., 2012) and Taq MutS (Cho et al., 2012; Jeong et al., 2011) reveals that it scans the DNA in stretches of thousands of base pairs (kb) per binding event, at speeds of 700 bp/s. Due to this high mobility of MutS during scanning of the DNA it is a challenging target for structural characterization. Here, we used a 1.7 kb linearized plasmid DNA (see methods for details) on which MutS molecules can freely diffuse and be subsequently frozen for cryo-EM analysis (Figure 1M). Due to the long DNA molecules that are stretched across the holes of the cryo-EM grid, the MutS molecules show a preferential orientation. Therefore, data was collected both at 0° and 25° tilt-angle of the sample grid in separate data collections and subsequently combined, resulting in a final cryo-EM map to 4.5 Å resolution (see Figure S1 for details). Similar to the structure of mismatch-bound MutS, the DNA is held between the mismatch binding domains and clamp domains (Figure 2A). The DNA, however, is straight (Figure 1A, 2A), unlike the mismatch-bound structure where the DNA is kinked by 60° (Figure 1B, 2B). Although ~30 base pairs can be discerned, the density for the DNA is less well defined than that for the protein part (Figure 2A) consistent with the movement of MutS on DNA during the search for a mismatch. Multi-body refinement and principle component analysis (Nakane et al., 2018) further shows that the interaction between the clamp domains and DNA is dynamic (Figure 2C). The three main principle components that represent the majority of the conformational sampling (Figure S1) show a movement of monomer B that is located to the left to monomer A (as seen in Figure 2C-D), but never passes over to the other side (Figure 2C-D, and Supplementary Movie 2). This movement of monomer B coincides with a modest bending of the DNA suggesting that MutS is actively trying to distort the DNA while scanning for a mismatch.

Mismatch binding overcomes a steric block that prevents transition into the sliding clamp

The cryo-EM structure of mismatch-bound MutS shows the canonical kinking of the DNA and only differs from the various crystal structures of this state by a few details. In the cryo-EM structure, 48 bp of the 61 bp DNA duplex are visible, three-fold more than in any MutS crystal structure (Gupta et al., 2011; Lamers et al.,

2000; Natrajan et al., 2003; Obmolova et al., 2000; Warren et al., 2007), including that of *E. coli* MutS (Figure S5B). Consequently, the mismatch binding domain of monomer B is better defined than in the crystal structure (Figure S5C) as in the latter case one arm of the DNA was too short for this domain to fully interact. In the mismatch-bound conformation MutS is more stably bound to the DNA than in the scanning state, as indicated by the better resolved density of the DNA (Figure 2B).

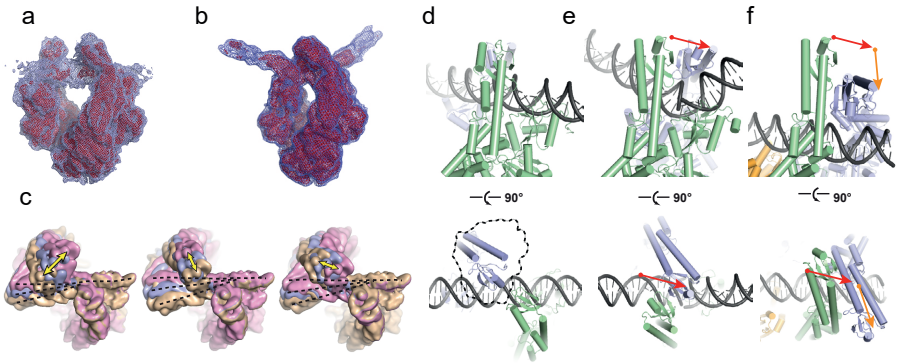


Figure 2. Licensing of DNA mismatch repair. a, Cryo-EM maps of scanning MutS shown as mesh at different contour levels (blue: lower; red: higher) revealing the poorly defined density for DNA at higher contour level. **b**, Cryo-EM maps of mismatch-bound MutS showing well-defined density of DNA at the higher contour level. **c**, Top view of multi-body refine-ment maps of scanning MutS showing the three first principle components of the clamp domain movement. Three maps are shown in blue, pink and beige. Dashed black lines indicates the position of the DNA in the three maps. **d-f**, Side (top row) and top view (bottom row) of **d,** MutS in scanning mode, **e**, mismatch binding mode and **f**, MutL^{LN40} binding mode. The dashed line in the bottom panel of d represents the area explored by the clamp domain in monomer B during the scanning mode. Red arrow in e-f indicates the movement of clamp domain of monomer B with respect to monomer A in the scanning to mismatch transition. Orange arrow in f indicates the movement from mismatch-bound to LN40-bound.

Once MutS has bound a mismatch, it will undergo an ATP-dependent conformational change that turns MutS into a sliding clamp that will bind MutL and initiate the repair cascade (Schofield et al., 2001). Importantly, comparison of the structure of MutS on homoduplex and mismatched DNA reveals how MutS is prevented from erroneously initiating repair during scanning of the DNA (Figure 2D-F). In the scanning state, the two clamp domains are directly opposed to each other while probing the DNA with a movement of monomer B towards the left of

monomer A as displayed in Figure 2C-D. In contrast, in the mismatch-bound state the clamp domain of monomer B has moved towards the right of monomer A and translated by ~35 Å along the length of the DNA (Figure 2E). This new position enables the closing of the clamp domains observed in the MutL-bound state where the clamp domains move an additional ~50 Å (Figure 2F, Supplementary Movie 2). Thus, the subtle rearrangement of the clamp domains upon mismatch binding appears to be the licensing step that enables MutS to transform into the sliding clamp that recruits MutL and initiates repair.

Structure of an intermediate state in MutS

Unexpectedly, we also obtained a structure of MutS in an intermediate state between the mismatch-bound and $MutL^{LN40}$-bound structure (Figure 3). In our initial attempt to obtain a structure of MutS on homoduplex DNA, we used a 50 bp perfectly matched DNA substrate with open ends in the presence of ATP. To our surprise, the resulting structure shows the MutS dimer located near the end of the DNA substrate (Figure 3A-B). This structure is in an intermediate conformation between the mismatch-bound state and the $MutL^{LN40}$-bound state (Supplementary Movie 3). The lever domains have rotated inwards pushing the DNA down by ~25 Å similar to the $MutL^{LN40}$-bound structure. Monomer B is in a conformation identical to that of the $MutL^{LN40}$-bound structure with the connector domain rotated by ~180° and the mismatch domain not visible in the cryo-EM map. In contrast, the connector domain of monomer A has only made a partial transition towards the 180° rotated conformation while the mismatch binding domain has moved away from the main body of MutS (Figure 3C-E) and remains in contact with the DNA where it inserts the conserved phenylalanine 36 between two adjacent base pairs analogous to the mismatch bound state (Figure 3F). Although the density of the DNA is not of high enough resolution to identify individual bases, larger features such as the major and minor groove and the position of the DNA backbone are well resolved and can be used to build a model. The DNA is bent in the same direction of mismatch-bound MutS yet without the pronounced kink (Figure 3G). In addition, the DNA is overwound, induced by the closing of the lever domains and the continued contact of one mismatch binding domain with the DNA. The distortion of the DNA provides a possible explanation for the trapping of this intermediate state at the end of the DNA substrate as it would not be possible on a continuous stretch of DNA where the rigidity of the DNA would create an opposing force

against the distortion. Indeed, previous studies have also observed MutS to bind at or near DNA ends (Acharya et al., 2003; Tessmer et al., 2008; Yang et al., 2005). It is therefore likely that this intermediate state is the result of the propensity of MutS to distort the DNA, which can only be achieved at a mismatch or, as in this case, near the end of a DNA substrate.

The intermediate position of the connector and mismatch binding domains suggests that this structure could be a structural mimic of a short-lived state between the mismatch-bound and MutL-bound form of MutS that was recently revealed by single molecule studies and termed "preceding mobile" state (or "bent state 3") (LeBlanc et al., 2018; Qiu et al., 2015). Therefore, to assess if the intermediate state could represent the "preceding mobile" state we used cysteine free MutS (Winkler et al., 2011), in which we introduced a single cysteine at position 220 ($MutS^{R220C}$). This position comes close to the DNA in the intermediate state (distance 220Cγ -DNA = 8.0 Å), but not in the mismatch-bound or $MutL^{LN40}$-bound conformations (43 and 28 Å respectively, Figure 3H). The DNA substrate was modified with a sulphur reactive crosslinker 4 bp away from the G:T mismatch in the middle of a 59 bp DNA oligo. In presence of ATP and a mismatch, which are needed to transform MutS into the clamp state, we obtain significant crosslinking, but not in the absence of ATP or when using a matched DNA substrate (Figure 3I). Since the crosslinking occurs 27 bp away from the end of the DNA, this crosslinking is consistent with the notion that our intermediate structure is similar, if not identical, to the "preceding mobile" state that MutS adopts when switching from the mismatch-bound to MutL-bound state.

Structure of $MutL^{LN40}$-bound MutS on DNA

Recently, the crystal structure of MutS bound to $MutL^{LN40}$ was determined (Groothuizen et al., 2015). However, despite being present during crystal growth, DNA was not visible in the electron density. Therefore, we determined the structure of MutS-$MutL^{LN40}$ and DNA by cryo-EM. The site-specific cross-linked complex of $MutS^{N246C}$ and $MutL^{LN40\text{-}N131C}$ was prepared as described, with each MutS monomer covalently bound to an $MutL^{LN40}$ monomer (Groothuizen et al., 2015) (Figure 1N). This cross-linked complex was subsequently incubated with a 61 bp DNA duplex with a G:T mismatch and streptavidin blocked ends in the presence of 1 mM of the non-hydrolysable ATP analogue AMP-PNP (Figure 1L).

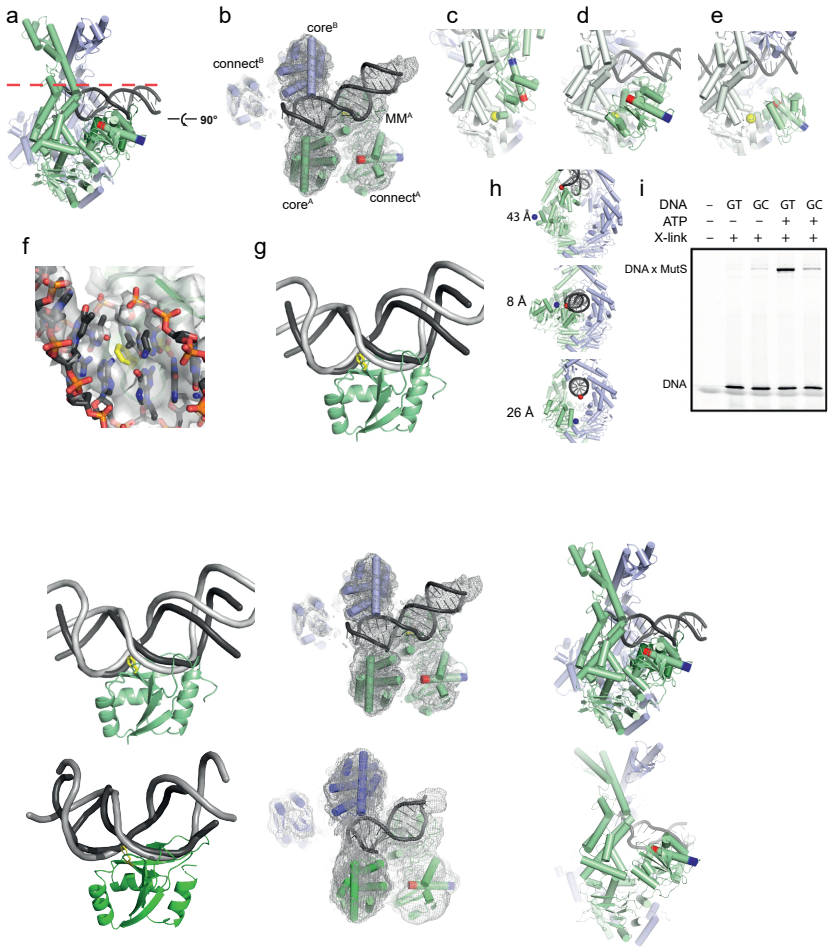


Figure 3. An intermediate state between mismatch and MutL binding. a, Side view of the intermediate state MutS. Dashed line indicates view point in panel b. **b**, Cut-through of transition state MutS with cryo-EM map shown at contour level 4 (grey mesh). Lever and clamp domains are omitted for clarity. **c-e**, Rotation of the connector domain from mismatch-bound, to intermediate state, to MutL-bound state. Connector domain in bright green with N- and C-termini of the central helix colored in red and blue, respectively. Rotation point at residue 267 is marked with a yellow sphere. **f**, Close up of mismatch binding domain interacting with DNA. Intercalating phenylalanine 36 (as in 1E3M) is coloured in yellow. Cryo-EM map in grey transparent surface **g**, Overwound DNA in the intermediate state shown in dark grey, DNA from mismatch-bound MutS shown in light grey. Mismatch binding domain of monomer A shown in light green, phenylalanine 36 in yellow. **h**, Crosslink distance between residue 220 and closest DNA marked by blue and red spheres in mismatch-bound (top), intermediate state (middle) and MutL^{LN40}-bound (bottom). **i**, SDS-page analysis of crosslinking experiment between residue 220 and DNA in different states of the repair process.

The complex was subsequently purified by gel filtration before preparation of cryo-EM grids. The resulting cryo-EM map shows the intact complex of MutS, $MutL^{LN40}$ and DNA to a final resolution of 4.7 Å (Figure 4A and Figure S4). The conformation of MutS-$MutL^{LN40}$ is similar to that observed in the crystal structure but now the DNA is clearly visible in the cryo-EM map, passing through the central pore of the complex (Figure 4A-C) consistent with the sliding clamp that MutS becomes upon mismatch and ATP binding (Blackwell et al., 2001; Gradia et al., 1999). Strikingly, there is pronounced asymmetry in the complex, with each MutS monomer interacting differently with the DNA. In addition, only one $MutL^{LN40}$ monomer is well-defined in the cryo-EM map, while the second $MutL^{LN40}$ monomer shows only residual density (Figure 4B). This structural asymmetry is due to a difference in the spacing of the two MutS monomers and the pitch of the DNA helix, such that monomer A interacts with the minor groove of the DNA, while monomer B predominantly interacts with the major groove (Figure 3B). As a result, $MutL^{LN40}$ is well positioned to interact with the DNA at only one end. Modelling of the second $MutL^{LN40}$ on the opposite end reveals clashes with the DNA, explaining the lack of density for this monomer (not shown).

Large-scale conformational changes have occurred between the mismatch-bound structure and the $MutL^{LN40}$-bound structure (Figure 1D, Supplementary Movie 1). The two MutS monomers have rotated inwards by ~23 Å around a pivot point located in the centre of the interface of the two ATPase domains, at the position of the gamma-phosphate. This tilts the MutS monomers towards each other, such that the lever arms cross, acting like a tourniquet that pushes the DNA downwards by ~25 Å. To accommodate the downward movement of the DNA, the connector and mismatch binding domains have repositioned: the connector domains have rotated by ~180 degrees downwards while the mismatch binding domains have become mobile and are not visible in the cryo-EM map (Figure 1H). The downward rotation of the connector domain creates a bi-partite binding site for $MutL^{LN40}$ consisting of the repositioned connector domain of monomer B and the ATPase domain of monomer A (Figure 4C), identical to that observed in the crystal structure of MutS-$MutL^{LN40}$ (Groothuizen et al., 2015) and consistent with the crosslinking result between yeast MSH2/6 and Mlh1/Pms1 (Mendillo et al., 2009).

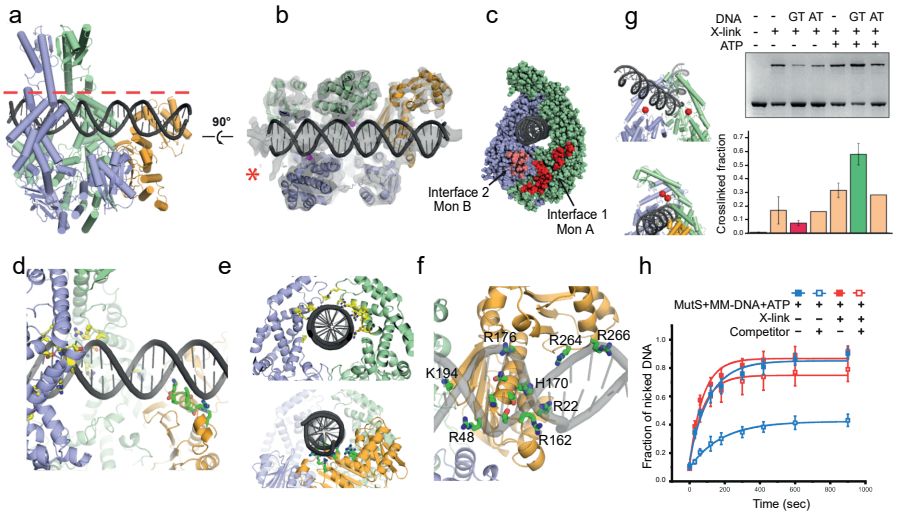


Figure 4. MutS-DNA-MutLLN40 complex. a, Side view of MutL^{LN40}-bound MutS. Dashed line indicates view point in panel b. **b**, Cut-through of MutL^{LN40}-MutS with cryo-EM map shown at contour level 5 (grey transparent surface). Red star marks the predicted position of MutL^{LN40} in monomer B. Lever and clamp domains are omitted for clarity. **c**, Composit binding surfaces for MutL, comprising the ATPase domain of monomer A (Interface 1, shown in dark red) and the rotated connector domain of monomer B (Interface 2, shown in light red). **d**, DNA interaction in the MutS-MutL^{LN40} complex, showing MutS residues in yellow sticks, and MutL^{LN40} residues in green sticks. **e**, Front view showing the residues from MutS (top) and MutL^{LN40} (bottom) in contact with DNA. **f**, Topview of MutL^{LN40} DNA interacting residues. g, SDS-page analysis of crosslinking experiment between residues 420 of MutS monomers A and B. **h**, MutH activation assay, requiring the action of MutS and MutL. Native and crosslinked MutS show similar activities, but only crosslinked MutS withstand competitor DNA (red open squares).

The DNA is located in the central pore and shows well defined major and minor grooves (Figure 4B). 40 of the 61 bp of the DNA substrate are visible. The DNA is straight B-form in which the mismatch cannot be discerned. Although DNA has moved towards the ATPase domains, it does not interact with them, unlike the related ABC ATPase RAD50 where the DNA is in direct contact with the ATPase domains (Käshammer et al., 2019; Liu et al., 2016; Seifert et al., 2016). Instead, the DNA is held between the lever domains of MutS that form an arch that holds the DNA down, while ~10 bp away, the MutL^{LN40} is positioned below the DNA, pushing it upwards (Figure 4D-F). MutL^{LN40} predominantly interacts with the DNA through residues 162-170 that form an arc that protrudes into the major groove of the DNA (Figure 4F). Comparison of the intermediate state and MutL-bound state suggests that the DNA may also have to translate by ~8 Å along its

axis such that the major groove is optimally aligned for this interaction to take place (Supplementary Movie 1). Additional contacts between $MutL^{LN40}$ and the DNA are made through a patch of residues that interact with the DNA backbone: arginines 22, 48, 176, 264, 266 and lysine 194. In agreement, mutation of arginine 266 reduces DNA binding (Groothuizen et al., 2015; Junop et al., 2003; Robertson et al., 2006), while additional mutation of arginines 162 and 316 results in a complete loss of DNA binding (Groothuizen et al., 2015). Of these, R162 is in contact with the DNA, whereas R316 is 14 Å away from the DNA and likely not to contribute to DNA binding in this conformation. The interaction surface between MutL and the DNA is small (280 $Å^2$) compared to that of MutS in the mismatch-bound state (1720 $Å^2$), consistent with the low affinity of MutL for DNA (Ban et al., 1999; Bende and Grafström, 1991) and the need for MutS to bring MutL to the DNA.

While it is established that the MutS clamp state is able to load MutL onto the DNA (Acharya et al., 2003; Groothuizen et al., 2015; Schofield et al., 2001), it is unknown if this state is also required for downstream reaction steps. We therefore investigated if a trapped MutS clamp would allow for activation of MutH. For this, we introduced a single cysteine at position 420 in a cysteine free MutS (Winkler et al., 2011). This $MutS^{R420C}$ can be crosslinked across the dimer interface in the clamp state of MutS (distance 420Cγ - 420Cγ = 5.5 Å), but not in the mismatched-bound form of MutS (420Cγ - 420Cγ = 40 Å) (Figure 4G). Indeed, in the presence of mismatched DNA and ATP we obtain crosslinking efficiencies of up to 60%, compared to 8% with mismatched DNA alone, or 16% in the absence of DNA and ATP. Next, we determined the effect of the crosslinking on the activation of MutH via a DNA nicking assay using a circular substrate containing a single mismatch and hemi-methylated GATC site (Hermans et al., 2016). Here, $MutS^{R420C}$ is first crosslinked on the mismatched circular DNA in the presence of ATP, followed by the addition of MutL and MutH (Figure 4H). The nicking activity of MutH is unaltered by the presence of the crosslinker, indicating that the crosslinked MutS clamp is fully capable of binding MutL and activating MutH. Finally, we added an excess of competitor mismatched DNA, after crosslinking but before addition of MutL and MutH. In the absence of crosslinker, the nicking activity is reduced to less than 20%, due to the release of MutS from the circular substrate DNA and rebinding to the competitor. In contrast, in the presence of the crosslinker, the MutH nicking activity is retained at ~90%, indicating that the MutS is retained

on the DNA. This shows the clamp state of MutS is both necessary and sufficient for MutL recruitment and subsequent activation of MutH.

The lever domain of MutS is kinked in the clamp state

In a subset of $MutL^{LN40}$-bound MutS particles (42%), the lever domain of the monomer B (furthest away from $MutL^{LN40}$) is kinked by ~45° and wrapped around the major groove of the DNA (Figure 5A-B). Interestingly, a similar kink in the lever domain was also observed in the recent crystal structure of MutS without DNA (Figure 5C), where one monomer has the lever domain kinked inwards while the second monomer has the lever domain kinked outwards (Bhairosing-Kok et al., 2019). Mutations in the kink region affected initial loading of MutS onto the DNA. In our cryo-EM structure of $MutL^{LN40}$-bound MutS, the kink appears to also play a role in the later stages of the repair process as it increases the contact area to the DNA in the MutS clamp state. To investigate this potential role, we determined the effects of a glutamine 476 to alanine mutation ($MutS^{Q476A}$) that is poised to interact with the DNA in the kinked clamp state (Figure 5D).

Using surface plasmon resonance, we find that the steady-state binding level of $MutS^{Q476A}$ amounts up to ~65% of wild type, (Figure 5E-F) which is consistent with its position close to the DNA also during mismatch binding. In the presence of ATP, wild type and mutant MutS increase their binding to DNA, indicating that both are able to transform into the clamp state and move away from the mismatch to allow more MutS to bind to the DNA substrate. However, the rate at which $MutS^{Q476A}$ increases binding is reduced relative to wild type MutS (Figure 5G), indicating that the DNA interaction of glutamine 476 enabled by the kinking of the lever domain promotes optimal clamp formation.

We therefore conclude that MutS lever domain kinking is important both in the initial stages where the outward kinking facilitate loading of MutS on the DNA (Bhairosing-Kok et al., 2019), as well as later where inward kinking facilitates the transformation of MutS into the sliding clamp that recruits MutL and allows for the continuation of the repair cascade.

The role of asymmetry in mismatch repair

In all four structures of MutS presented here, there is clearly defined structural asymmetry in the MutS homodimer, analogous to the eukaryotic MutS homologs that form true heterodimers. We therefore wondered if the structural asymmetry

in the *E. coli* MutS dimers is required for mismatch repair. Specifically, we wondered if initial mismatch binding dictates which side of the MutS dimer will bind MutL. To this end, we created obligate MutS heterodimers by making use of the bi-partite MutL binding site on MutS shared between the ATPase domain of one monomer (Interface 1) and the connector domain of the other monomer (Interface 2) (Figure 6A). In addition, we included a mutation of phenylalanine 36 ($MutS^{F36A}$) (Yamamoto et al., 2000) to disrupt mismatch recognition in combination with either Interface 1 or Interface 2. Mutation of either Interface 1 or Interface 2 in MutS homodimers does not affect mismatch binding but leads to loss of MutL binding (Figure 6B) and loss of activation of MutH in the DNA nicking assay (Figure 6C). Mutation of F36A in the MutS homodimer disrupts mismatch binding, and consequently MutL binding and MutH activation.

Mixing of the two Interface mutants 1 and 2 to create heterodimeric MutS partially restores MutL binding and MutH activation, indicating that a single MutL binding site on the MutS dimer is sufficient for mismatch repair. Next, we combined the mismatch binding mutation F36A with either the Interface 1 or Interface 2 mutation, to create functional MutS heterodimers in which mismatch binding occurs on the opposing side or on the same side of the MutL binding, respectively. Here we find that these functional heterodimers of MutS still retain MutL recruitment and MutH activation up to maximally 50% of wild type levels. Since the protein mixtures contain 50% heterodimeric MutS and 50% of inactive homodimers, these results indicate that the functional heterodimers can reach near wild type levels of activity. This shows that in the bacterial system a single active side of the MutS dimer is sufficient for MutL recruitment and activation of MutH.

The structural and functional asymmetry that we observe in the bacterial MutS dimer suggests it may have been an evolutionary precursor of the eukaryotic true heterodimers MSH2/6 and MSH2/3. However, unlike the eukaryotic system, there is no strict division of labour, as is does not matter on which subunit the MutL binding site is left intact with respect to mismatch binding. A small difference between the two heterodimers is appreciable in both MutL binding and MutH activation, yet the details of this preference are beyond the scope of this work.

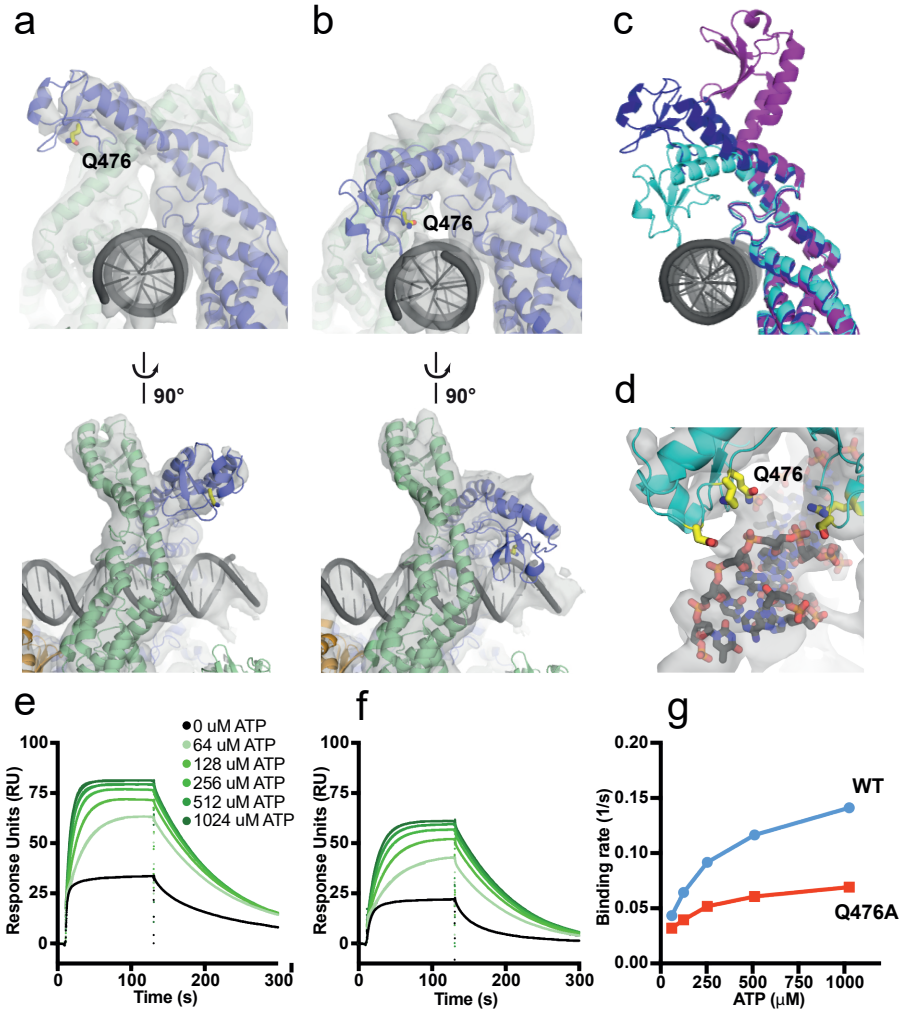


Figure 5. Kinking of the lever domain promotes clamp formation. a, MutS-MutL^{LN40}-complex with straight lever domain: front (top row) and side view (bottom row). MutS monomers A and B in green and blue with cryo-EM map in grey transparent surface. Glutamine 476 indicated in yellow sticks. **b**, MutS-MutL^{LN40}-complex with kinked lever domain. **c**, Comparison of inward kinked (cyan), straight (blue) and outward kinked (purple) lever domain (see main text for more details). **d**, Close up view of the DNA interaction in the kinked lever domain. Residues close to DNA are highlighted in yellow sticks. **e**, Mismatch DNA binding by wild-type MutS and **f**, MutS^{Q476A} as measured by Surface Plasmon Resonance. MutS was added to an end-blocked 100 bp DNA substrate with a centrally located GT mismatch with increasing concentrations of ATP (0-1024 μM). **g**, Comparison of clamp formation - represented as the binding rate - by MutS^{WT} (blue circles) and MutS^{Q476A} (red squares) derived from the binding profiles in panels e and f.

Discussion

DNA mismatch repair is an energy consuming and potentially dangerous repair process that excises up to thousands of nucleotides from the newly synthesized DNA strand. Therefore, a highly accurate validation system needs to be in place that prevents MutS from erroneously initiating repair during scanning of millions of base pairs. Our structures reveal a small conformational change from scanning MutS to mismatch-bound MutS that acts as a licensing step for the transition into the sliding clamp state that recruits MutL and initiates the repair cascade. In this licensing step, the straight homoduplex DNA acts as a steric block during scanning that can only be overcome through kinking of the DNA at a mismatch. This licensing enables the transformation into the MutL-binding sliding clamp via an intermediate state. The structural asymmetry observed in all four structures is formalized in the true asymmetric MutS heterodimers in eukaryotes. Indeed, engineered functional heterodimers of *E. coli* MutS retain mismatch repair initiation activity, indicating that the bacterial homodimers are only a few mutations away from becoming heterodimers. Thus, the structural asymmetry in the MutS homodimers may be an evolutionary precursor to the true heterodimers found in eukaryotes.

METHODS

Site-directed mutagenesis

MutS mutants were made using site directed mutagenesis in different background plasmids. $MutS^{R220C}$ and $MutS^{R420C}$ were created in a cysteine free version of MutS (Groothuizen et al., 2013). $MutS^{Q476A}$ was made in the background of $MutS^{D835R}$, a version of MutS that does not form tetramers (Groothuizen et al., 2013). $MutS^{F36A}$ was made in a background of both Interface Mutant 1 and Interface Mutant 2 (Groothuizen et al., 2015).

Protein purification

MutS, MutH, and monovalent streptavidin were purified as described previously (Groothuizen et al., 2015; Howarth et al., 2006; Lamers et al., 2000).

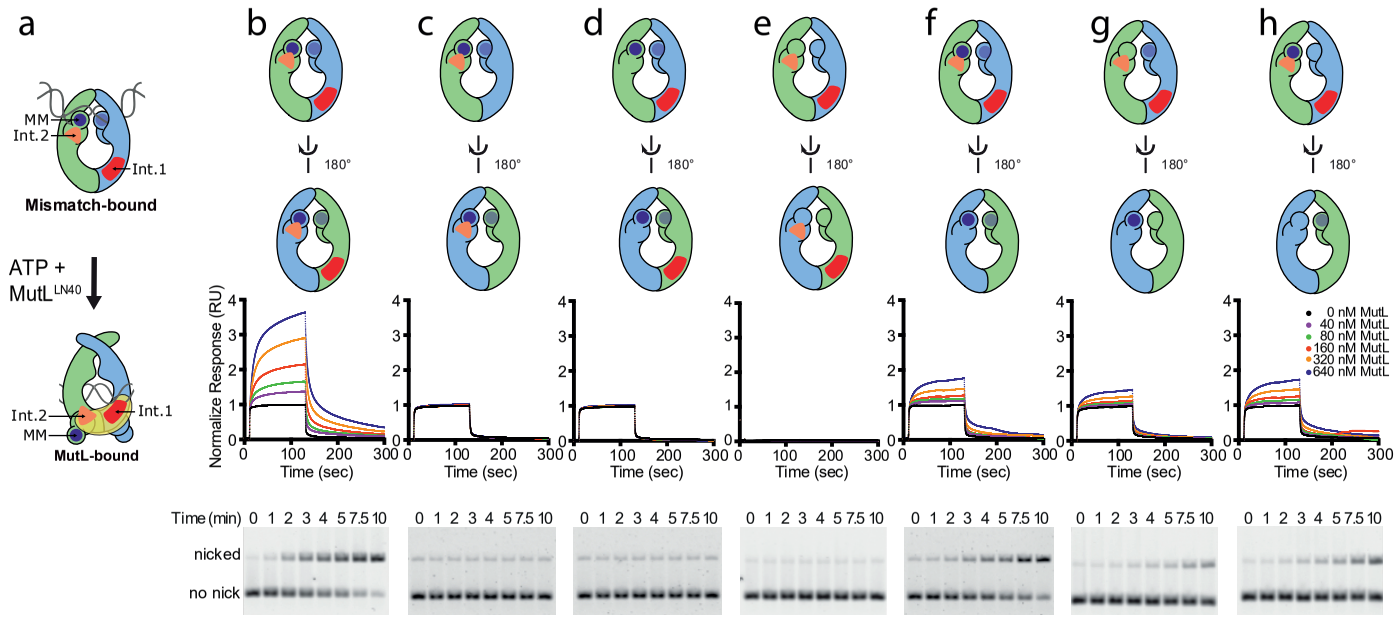


Figure 6. MutS heterodimers are mismatch repair competent. a, Schematic representation of the mismatch and MutL binding regions in MutS during mismatch binding (top) and MutL binding (bottom). Monomer A in blue, monomer B in green, mismatch binding area indicated with a blue dot, MutL Interface 1 with red rectangle, MutL interface 2 with orange triangle. For more detail of the mismatch and MutL binding sites, see Extended Data Fig. 6-7. **b-h**, Biochemical analysis of the MutS homodimeric and heterodimeric variants. Top two rows: schematic respresentation of the different MutS homo- and heterodimers. Third row: surface plasmon resonance binding curves of MutS and MutL binding to open-ended 100 bp DNA subtrate with a centrally located GT mismatch. Bottom row: MutH-dependent nicking assays using an engineered circular DNA substrate with a single mismatch and a single hemi-methylated GATC-site. Nicking of the circular DNA substrate by MutH requires interaction of mismatch, MutS, MutL and MutH (see methods for details).

MutL was expressed and purified as follows. *E. coli* BL21(DE3) cells were transformed with plasmid pTX418 (Feng and Winkler, 1995) coding for MutL and plated onto LB agar plates and incubated at 37 °C o/n. Single colony was picked and cells grew in LB supplemented with 1% glucose, 10mM KH_2PO_4 pH 7.4 and 50 µg/mL carbencillin at 37 °C to OD600 ~0.6 and induced with 1mM 1-thio-β-D-galactopyranoside for 3.5 hours. All purification steps were performed at 4 °C. Cells were resuspended in lysis buffer (50 mM KH_2PO_4 pH 7.4, 100 mM KCl, 5 mM imidazole, 2 mM 2-mercaptoethanol, 1 mM PMSF) supplemented with protease inhibitors (Thermo Fischer) and lysed by sonication for 7 minutes using intervals (45 seconds on and 15 seconds off). After centrifugation, supernatant was incubated with Talon resin (Clonetech Laboratories, Takara holdings inc, Japan) for 30 min at 4 °C. Talon resin was washed with washing buffer (50 mM KH_2PO_4, 1M KCl, 5 mM imidazole, 2 mM 2-mercaptoethanol) and MutL was eluted with elution buffer (50 mM KH_2PO_4, 100 mM KCl, 300 mM imidazole, 2 mM 2-mercaptoethanol). Elution was loaded on a pre-equilibrated heparin column (24.3 KH_2PO_4, 5.7 KH_2PO_4, 100 mM KCl, 5 mM imidazole, 2 mM 2-mercaptoathanol) and was eluted using a gradient from 0.1-1.0 M KCl. Fractions were pooled, concentrated to ~1-2 mL and loaded onto a pre-equilibrated Superdex S200 16/60 with (40.5 mM KH_2PO_4, 9.5 mM KH_2PO_4, 100 mM KCl, 10 mM EDTA, 1 mM 2-mercaptoethanol). Fractions were analyzed on SDS-PAGE and full-length MutL fractions were pooled, concentrated, flash frozen and stored at -80°C.

DNA substrates

For the mismatch-bound and MutL-bound structures a 61 bp dsDNA with a centrally located G:T mismatch and 5' biotinylated ends was used (Table S3). Monovalent streptavidin was added and purified over gel filtration to remove excess streptavidin. For the intermediate (end-bound) structure a perfectly matched 50 bp dsDNA (Table S3) was used were purified on Superdex 200 gel filtration column.

For the scanning state MutS, a DNA mini plasmid of 1765 bp (pRC1765) was created by deleting unwanted sequences from a pET plasmid, leaving only the origin of replication and ampicillin resistance gene and its promoter. Two BbvCI nicking sites were introduced 33 nucleotides apart in order to allow for introduction of modified DNA through annealing and ligation (e.g. mismatch, DNA lesions, DNA flap, etc.). pRC1765 is available via AddGene (ID-141346).

For protein DNA-crosslinking experiments a 59 base pair substrate with a G:T mismatch in position 24 and a amino modifier C6 dT (Eurogentec) at position 28 was used (See Table S2). DNA oligos were modified as described previously (Monakhova et al., 2015). In brief, cross-linker SPDP (6.8 Å) was dissolved in DMSO to 500 mM, aliquoted and stored at -20 °C. Cross-linker solution (5 µl) was added to 45 µl oligonucleotide (100 µM) and incubated overnight at room temperature. Excess of unbound cross-linker was removed by gel filtration with Zeba Desalt Spin Columns (Pierce, Thermo Scientific). Final DNA substrate was generated by annealing to a complementary strand labelled with Cy5 at the 5' end.

Protein-protein and protein-DNA crosslinking

For cryo-EM studies of the MutS-MutL^{LN40}-DNA complex, MutS$^{\Delta C800}$ D246C (MutS$^{\Delta C800}$) and the 40 kDa N-terminal MutL^{LN40} domain (Ban and Yang, 1998) of MutL N131C (MutL^{LN40}) were crosslinked with BM(PEO)$_3$ in the presence of mismatched DNA and ATP as described (Groothuizen et al., 2015). This procedure saturates MutS with MutL, such that each individual MutS monomer has a MutL monomer loaded. The resulting complex was subsequently purified by gel filtration.

Analytical crosslinking of single cysteine MutS^{R420C} dimer variant (2 µM) was performed on 59 bp DNA (1 µM) containing digoxigenin ends blocked with anti-digoxigenin Fab fragments (4 µM) at room temperature in buffer FB150 (25 mM Hepes-KOH pH 7.5, 150 mM KCl, 5 mM $MgCl_2$, 0.005% Tween 20) in the presence or absence of ATP (100 µM) followed by the addition of crosslinker MTS-4-MTS (50 µM).

Protein-DNA crosslinking was performed by adding single cysteine MutS^{R220C} (2 µM) to the modified DNA duplex (1 µM) pre-incubated in buffer FB150 (25 mM HEPES/KOH, pH 7.5, 150 mM KCl, 5 mM MgCl2) with 1 mM ATP. The mixture was kept on ice overnight. The crosslinking was performed in the volume of 10 µl. To separate cross-linked MutS/DNA complex from uncross-linked proteins 8 % SDS-polyacrylamide gels electrophoresis (SDS-PAGE) were used. DNA was observed by fluorescence of Cy5 dye.

Cryo-electron microscopy

The crosslinked MutS-MutL^{LN40} complex was bound to DNA in the presence of AMP-PNP and purified over a gel filtration column (Superdex200) on an Akta

Micro system (GE healthcare). The peak fraction was supplemented with 0.006% (w/v) Tween-20, applied to a glow-discharged Quantifoil 1.2/1.3 400-mesh grid, and vitrified using a manual plunger at 4ºC. For the MutS-transition and MutS-scanning complexes, MutS-FL^{D835R} was purified over a gel filtration column (Superdex200) on an Akta Micro system (GE healthcare) and the peak fraction was supplemented with 0.006% (w/v) Tween-20, 1 mM ATP, and mixed with DNA (5 µM final concentration of 50mer homoduplex or 50 nM final concentration of Kpn1-linearized plasmid pRC1765) prior to vitrification using similar settings as described above.

For the MutS-MutL^{LN40} complex, data were manually collected on a Titan Krios at MRC-LMB equipped with Gatan Quantum GIF with K2 direct electron detector in 6 separate sessions. For the MutS-transition state, data was collected on a Titan Krios at MRC-LMB equipped with Gatan Quantum GIF with K2 direct electron detector in a single session using automated data collection procedures with Serial-EM (Mastronarde, 2005). Data for MutS-scanning complex was collected on a Titan Krios at MRC-LMB equipped with Gatan Quantum GIF with K2 direct electron detector using EPU (ThermoFixer) in 4 separate sessions. During the 3rd and 4th sessions the stage was tilted to 25º and 40º respectively to increase the orientation spread of the particles. For details on microscope setup see Data table S1 and S2.

Data processing followed a similar scheme using Relion 3.0 (Zivanov et al., 2018) and Relion 3.1 beta (for CTF-refinement) for all datasets unless stated otherwise. For the MutS-MutL^{LN40} complex, a total of 1608 movies were pre-processed using the MotionCorr implementation in Relion 3.0 and gCTF (Zhang, 2016). Reference-free picking was initially performed in Relion, yielding initial 2D classes that were subsequently used for extensive autopicking. After satisfactory picking, a thorough clean-up of particles was done using 2D classification. For this dataset, initial 3D classification was performed using two different initial models simultaneously in order to separate mismatch-bound-MutS particles from MutS-MutL^{LN40}-clamp particles. After this initial splitting of the dataset, subsequent 3D classification runs were done in order to further clean-up the data. Particles discarded from 3D classification runs of the MutS-mismatch-bound data were fed to the MutS-MutL^{LN40}-clamp data for 3D classification and vice versa. After several runs of 3D classification, we were left with two clean datasets for both mismatch-bound and clamp states with 42,433 particles and 77,446 particles respectively. Data

processing continued for both states independently including particle polishing, CTF-refinement and 3D refinement using common procedures in Relion. See Figure S2 and S4 for details on data processing and validation. For the MutS-scan state, the 4 sessions were pre-processed individually using WARP (Tegunov and Cramer, 2019) to perform motion correction, CTF estimation, micrograph curation and particle picking. Data processing continued in Relion 3.0 following standard procedures, including multibody refinement to study the dynamics in between MutS monomers in this conformation. See Figure S1 for details on data processing scheme and validation. Pre-processing of the MutS-transition state data was performed in Relion. The final selection of particles was achieved by using focused 3D classification after performing signal subtraction in Relion 3.1 in order to keep only the signal from DNA and mismatch domain. After selecting the particles with good occupancy for this region, we performed a final refinement on the original particle set for this selection that provided a more defined and complete map. See Figure S3 for details on data processing scheme and validation.

Model building was performed using Coot (Emsley et al., 2010), REFMAC5 (Murshudov et al., 2011), the CCPEM-suite (Nicholls et al., 2018) and Phenix (Liebschner et al., 2019). With the exception of the intermediate MutS state, the resolution and quality of the cryo-EM maps didn't allow for detailed model building. Therefore, we relied on the previous models available. For details on model refinement and validation see Table S2. In brief, model building started by rigid-body fitting known crystal structures (1E3M and 5AKB) into the different maps using Coot. Next, we used iterative rounds of real-space-refinement in Phenix and manual building in Coot. After initial refinement, maps were sharpened using Loc-Scale (Jakobi et al., 2017), which aided in model building and refinement. Final validation of the model and data was performed using CCPEM, Refmac and Molprobity.

Surface plasmon resonance

Binding kinetics of MutS and MutS/MutL to DNA were determined using surface plasmon resonance (SPR). The measurements were performed in a Biacore T200 system (GE Healthcare) at 25◦C with the same setup as described previously (Bhairosing-Kok et al., 2019). Data analysis was done using GraphPad Prism 7.02 (GraphPad Software Inc.). To measure the kinetics of MutS binding to DNA, we used a double-biotinylated 100-bp DNA substrate containing a centrally located

GT mismatch that was attached to a streptavidin-coated chip and blocked with mono-valent streptavidin on the other end. Experiments were performed in the presence of increasing concentrations of ATP (64 to 1024 μM). In order to measure the binding of MutS and MutL, we used an open-ended 100bp DNA substrate with a centrally located mismatch bound to a streptavidin chip as described above. MutS injections of 400 nM for 60 s were followed by MutL injections for 120 s with concentrations ranging from 0 to 1280 nM in the presence of 1 mM ATP.

MutH nicking assays

Nicking reactions were carried out at indicated MutS, MutL and MutH concentrations and 0.5 nM of a 3.2 kb circular DNA substrate with a single mismatch, Alexa^{647} fluorophore and hemi-methylated GATC site (GT#1^{647}) (Hermans et al., 2016) and 1 mM ATP at 37◦C. Reactions were stopped by adding an equal volume of 20% glycerol, 1% SDS, 50 mM EDTA and run on a 0.8 % agarose gel containing 40 μM chloroquine. Gels were scanned on a Typhoon FLA imager (GE Healthcare) with excitation at 633 nm and emission collected via the 670BP30 filter. For analysis of MutS heterodimers (Figure 6) 0.5 mM of each MutS variant was mixed and incubated for 60 minutes at 4˚C prior to dilution into the reaction mix with final concentration of 10 nM MutS, 10 nM MutL and 5 nM MutH. For crosslinking and competition assays (Figure 4H), MutS^{420C} (40 nM), ATP (50 μM) and GT#1^{647} DNA (1 nM) were mixed, followed by addition of 1 μM of MTS-4-MTS crosslinker (Toronto Research Chemicals) for 2.5 minutes. Where indicated, competitor DNA (0.2 mM, 30 bp with single GT mismatch, DIG-label at both ends and blocked with anti-DIG Fab fragment (Sigma-Aldrich) was added for 60 minutes prior to dilution into the reaction mix at final concentrations of 20 nM MutS, 20 nM MutL and 10 nM MutH.

Data availability

Cryo-EM maps and atomic models have been deposited in the Electron Microscopy Database and Protein Data Bank. PDB codes: 7AI6, 7AI5, 7AI7, 7AIC and 7AIB. pRC1765 mini plasmid is available via AddGene (ID-141346). Figures 3, 4, 5 and 6 have associated raw data.

Acknowledgement

All cryo-EM data was collected at the LMB cryo-EM facility. We thank Christos Savva for cryo-EM support; Jake Grimmett, Toby Darling, Jose Luís Fernández and Pablo Reviriego for IT support. This work has been supported by a UK Medical Research Council grant U105197143 and LUMC Research Fellowship to M.H.L., Oncode Institute, NWO-TOP 714.016.002 and NWO-Gravity CGC.nl to T.K.S., a Spanish Ministry of Economy and Competitiveness Grant BFU2017-87316 to R.F.L., NWO-Gravity CGC.nl and Oncode Institute to J.H.G.L., and European Community's Horizon2020 Marie Skłodowska Curie grant [722433] to P.F.

Author contributions

R.F.L., T.K.S. and M.H.L. conceived the overall experimental design; R.F.L. prepared samples, collected and processed cryo-EM data; D.B.K. purified proteins and performed SPR experiments. A.F. helped with SPR data analysis. V.K. performed cross-linking experiments and MutH nicking assays on crosslinked MutS; P.F. designed crosslinking experiments; J.H.G.L and C.L. performed MutH nicking assays on MutS heterodimers; F.G. and H.H.W. prepared MutS-MutL^{LN40} crosslinked complex; R.F.L., T.K.S and M.H.L. wrote the manuscript with contributions from all authors.

Declaration of Interests

The authors declare no competing interest.

Material & Correspondence

Correspondence and requests for materials should be addressed to: Meindert Lamers (m.h.lamers@lumc.nl), Titia K. Sixma (t.sixma@nki.nl) or Rafael Fernández-Leiro (rfleiro@cnio.es).

REFERENCES

Acharya, S., Foster, P.L., Brooks, P., and Fishel, R. (2003). The coordinated functions of the E. coli MutS and MutL proteins in mismatch repair. *12*, 233–246.

Ban, C., and Yang, W. (1998). Crystal structure and ATPase activity of MutL: implications for DNA repair and mutagenesis. Cell *95*, 541–552.

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.

Bende, S.M., and Grafström, R.H. (1991). The DNA binding properties of the MutL protein isolated from Escherichia coli. Nucleic Acids Res. *19*, 1549–1555.

Bhairosing-Kok, D., Groothuizen, F.S., Fish, A., Dharadhar, S., Winterwerp, H.H.K., and Sixma, T.K. (2019). Sharp kinking of a coiled-coil in MutS allows DNA binding and release. Nucleic Acids Res. *74*, 681.

Blackwell, L.J., Bjornson, K.P., Allen, D.J., and Modrich, P. (2001). Distinct MutS DNA-binding modes that are differentially modulated by ATP binding and hydrolysis. *276*, 34339–34347.

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.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. *66*, 486–501.

Feng, G., and Winkler, M.E. (1995). Single-step purifications of His6-MutH, His6-MutL and His6-MutS repair proteins of escherichia coli K-12. BioTechniques *19*, 956–965.

Fukui, K., Nishida, M., Nakagawa, N., Masui, R., and Kuramitsu, S. (2008). Bound nucleotide controls the endonuclease activity of mismatch repair enzyme MutL. J. Biol. Chem. *283*, 12136–12145.

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. USA *109*, E3074–E3083.

Gradia, 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. *3*, 255–261.

Groothuizen, F.S., Fish, A., Petoukhov, M.V., Reumer, A., Manelyte, L., Winterwerp, H.H.K., Marinus, M.G., Lebbink, J.H.G., Svergun, D.I., Friedhoff, P., et al. (2013). Using stable MutS dimers and tetramers to quantitatively analyze DNA mismatch recognition and sliding clamp formation. Nucleic Acids Res. *41*, 8166–8181.

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.

Gupta, S., Gellert, M., and Yang, W. (2011). Mechanism of mismatch recognition revealed by human MutSβ bound to unpaired DNA loops. Nat. Struct. Mol. Biol. *19*, 72–78.

Hall, M.C., and Matson, S.W. (1999). The Escherichia coli MutL protein physically interacts with MutH and stimulates the MutH-associated endonuclease activity. J. Biol. Chem. *274*, 1306–1312.

Heo, S.-D., Cho, M., Ku, J.K., and Ban, C. (2007). Steady-state ATPase activity of E. coli MutS modulated by its dissociation from heteroduplex DNA. Biochem Biophys Res Commun *364*, 264–269.

Hermans, N., Laffeber, C., Cristóvão, M., Artola-Borán, M., Mardenborough, Y., Ikpa, P., Jaddoe, A., Winterwerp, H.H.K., Wyman, C., Jiricny, J., et al. (2016). Dual daughter strand incision is processive and increases the efficiency of DNA mismatch repair. Nucleic Acids Res. *44*, 6770–6786.

Howarth, M., Chinnapen, D.J.-F., Gerrow, K., Dorrestein, P.C., Grandy, M.R., Kelleher, N.L., El-Husseini, A., and Ting, A.Y. (2006). A monovalent streptavidin with a single femtomolar biotin binding site. Nat. Methods *3*, 267–273.

Jakobi, A.J., Wilmanns, M., and Sachse, C. (2017). Model-based local density sharpening of cryo-EM maps. Elife *6*, 213.

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.

Jiricny, J. (2013). Postreplicative mismatch repair. Cold Spring Harb Perspect Biol *5*, a012633.

Junop, M.S., Yang, W., Funchain, P., Clendenin, W., and Miller, J.H. (2003). In vitro and in vivo studies of MutS, MutL and MutH mutants: correlation of mismatch repair and DNA recombination. DNA Repair (Amst) *2*, 387–405.

Kadyrov, F.A., Dzantiev, L., Constantin, N., and Modrich, P. (2006). Endonucleolytic function of MutLalpha in human mismatch repair. Cell *126*, 297–308.

Käshammer, L., Saathoff, J.-H., Lammens, K., Gut, F., Bartho, J., Alt, A., Kessler, B., and Hopfner, K.-P. (2019). Mechanism of DNA End Sensing and Processing by the Mre11-Rad50 Complex. Mol. Cell *76*, 382–394.e386.

Lamers, M.H., Perrakis, A., Enzlin, J.H., Winterwerp, H.H., de Wind, N., and Sixma, T.K. (2000). The crystal structure of DNA mismatch repair protein MutS binding to a G x T mismatch. Nature *407*, 711–717.

LeBlanc, S.J., Gauer, J.W., Hao, P., Case, B.C., Hingorani, M.M., Weninger, K.R., and Erie, D.A. (2018). Coordinated protein and DNA conformational changes govern mismatch repair initiation by MutS. Nucleic Acids Res. *46*, 10782–10795.

Li, Z., Pearlman, A.H., and Hsieh, P. (2016). DNA mismatch repair and the DNA damage response. DNA Repair (Amst) *38*, 94–101.

Liebschner, D., Afonine, P.V., Baker, M.L., Bunkóczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A.J., et al. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. Acta Crystallogr D Struct Biol *75*, 861–877.

Liu, Y., Sung, S., Kim, Y., Li, F., Gwon, G., Jo, A., Kim, A.-K., Kim, T., Song, O.-K., Lee, S.E., et al. (2016). ATP-dependent DNA binding, unwinding, and resection by the Mre11/Rad50 complex. Embo J. *35*, 743–758.

Mastronarde, D.N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. J. Struct. Biol. *152*, 36–51.

Mendillo, M.L., Hargreaves, V.V., Jamison, J.W., Mo, A.O., Li, S., Putnam, C.D., Woods, V.L., and Kolodner, R.D. (2009). A conserved MutS homolog connector domain interface interacts with MutL homologs. Proc. Natl. Acad. Sci. USA *106*, 22223–22228.

Monakhova, M., Ryazanova, A., Hentschel, A., Viryasov, M., Oretskaya, T., Friedhoff, P., and Kubareva, E. (2015). Chromatographic isolation of the functionally active MutS protein covalently linked to deoxyribonucleic acid. J Chromatogr A *1389*, 19–27.

Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr. D Biol. Crystallogr. *67*, 355–367.

Nakane, T., Kimanius, D., Lindahl, E., and Scheres, S.H. (2018). Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. Elife *7*, 1485.

Natrajan, G., Lamers, M.H., Enzlin, J.H., Winterwerp, H.H.K., Perrakis, A., and Sixma, T.K. (2003). Structures of Escherichia coli DNA mismatch repair enzyme MutS in complex with different mismatches: a common recognition mode for diverse substrates. Nucleic Acids Res. *31*, 4814–4821.

Nicholls, R.A., Tykac, M., Kovalevskiy, O., and Murshudov, G.N. (2018). Current approaches for the fitting and refinement of atomic models into cryo-EM maps using CCP-EM. Acta Crystallogr D Struct Biol *74*, 492–505.

Obmolova, G., Ban, C., Hsieh, P., and Yang, W. (2000). Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. Nature *407*, 703–710.

Pillon, M.C., Miller, J.H., and Guarné, A. (2011). The endonuclease domain of MutL interacts with the β sliding clamp. DNA Repair (Amst) *10*, 87–93.

Pluciennik, A., Dzantiev, L., Iyer, R.R., Constantin, N., Kadyrov, F.A., and Modrich, P. (2010). PCNA function in the activation and strand direction of MutLα endonuclease in mismatch repair. Proc. Natl. Acad. Sci. USA *107*, 16066–16071.

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. USA *112*, 10914–10919.

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.

Schofield, M.J., Nayak, S., Scott, T.H., Du, C., and Hsieh, P. (2001). Interaction of Escherichia coli MutS and MutL at a DNA mismatch. *276*, 28291–28299.

Seifert, F.U., Lammens, K., Stoehr, G., Kessler, B., and Hopfner, K.-P. (2016). Structural mechanism of ATP-dependent DNA binding and DNA end bridging by eukaryotic Rad50. Embo J. *35*, 759–772.

Tegunov, D., and Cramer, P. (2019). Real-time cryo-electron microscopy data preprocessing with Warp. Nat. Methods *16*, 1146–1152.

Tessmer, I., Yang, Y., Zhai, J., Du, C., Hsieh, P., Hingorani, M.M., and Erie, D.A. (2008). Mechanism of MutS searching for DNA mismatches and signaling repair. *283*, 36646–36654.

Warren, J.J., Pohlhaus, T.J., Changela, A., Iyer, R.R., Modrich, P.L., and Beese, L.S. (2007). Structure of the human MutSalpha DNA lesion recognition complex. Mol. Cell *26*, 579–592.

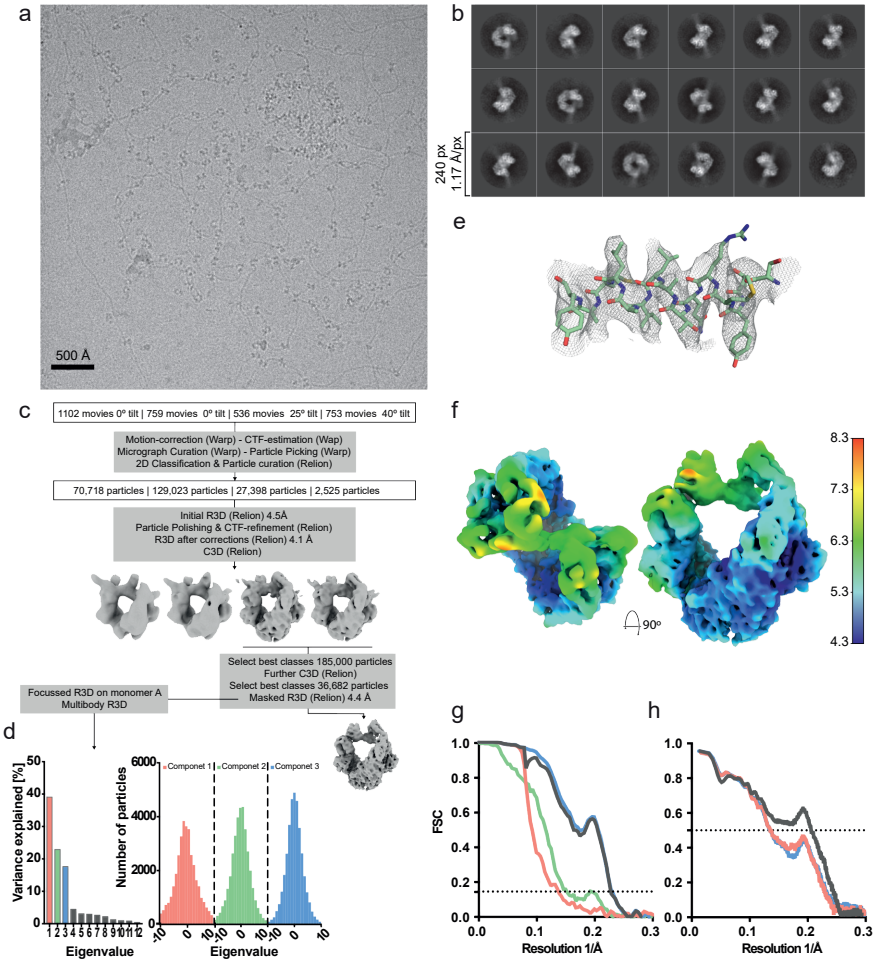
Winkler, I., Marx, A.D., Lariviere, D., Heinze, R.J., Cristovao, M., Reumer, A., Curth, U., Sixma, T.K., and Friedhoff, P. (2011). Chemical trapping of the dynamic MutS-MutL complex formed in DNA mismatch repair in Escherichia coli. Journal of Biological Chemistry *286*, 17326–17337.

Yamamoto, A., Schofield, M.J., Biswas, I., and Hsieh, P. (2000). Requirement for Phe36 for DNA binding and mismatch repair by Escherichia coli MutS protein. Nucleic Acids Res. *28*, 3564–3569.

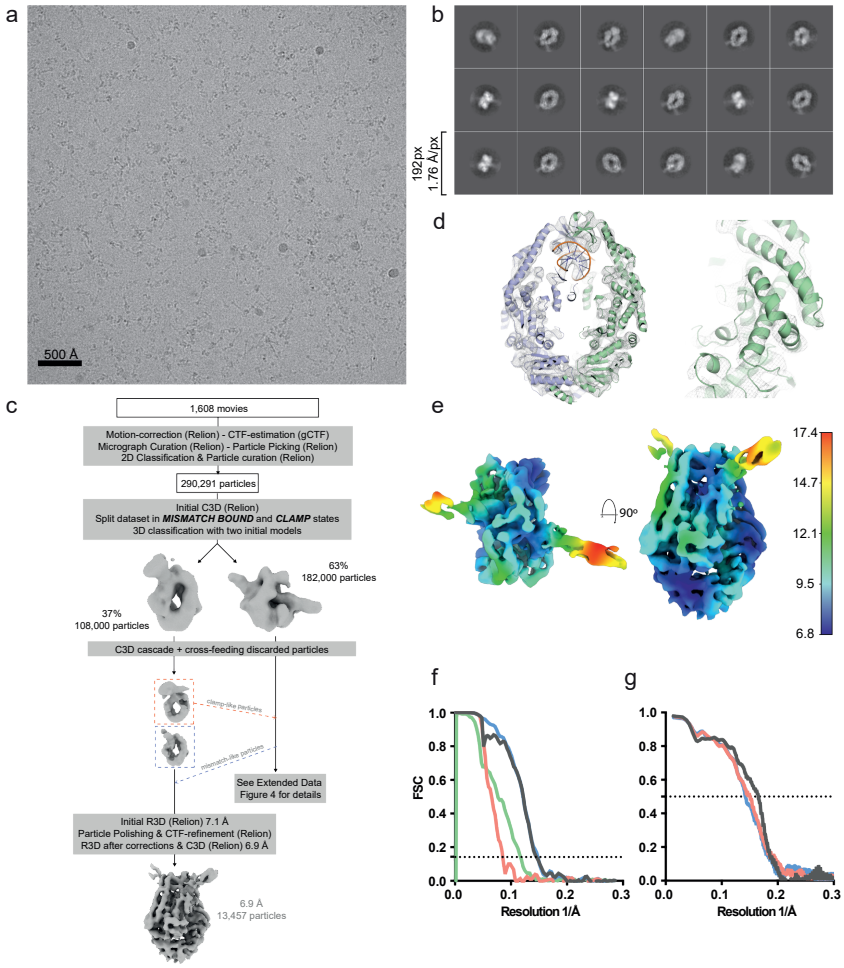
Yang, Y., Sass, L.E., Du, C., Hsieh, P., and Erie, D.A. (2005). Determination of protein-DNA binding constants and specificities from statistical analyses of single molecules: MutS-DNA interactions. Nucleic Acids Res. *33*, 4322–4334.

Zhang, K. (2016). Gctf: Real-time CTF determination and correction. J. Struct. Biol. *193*, 1–12.

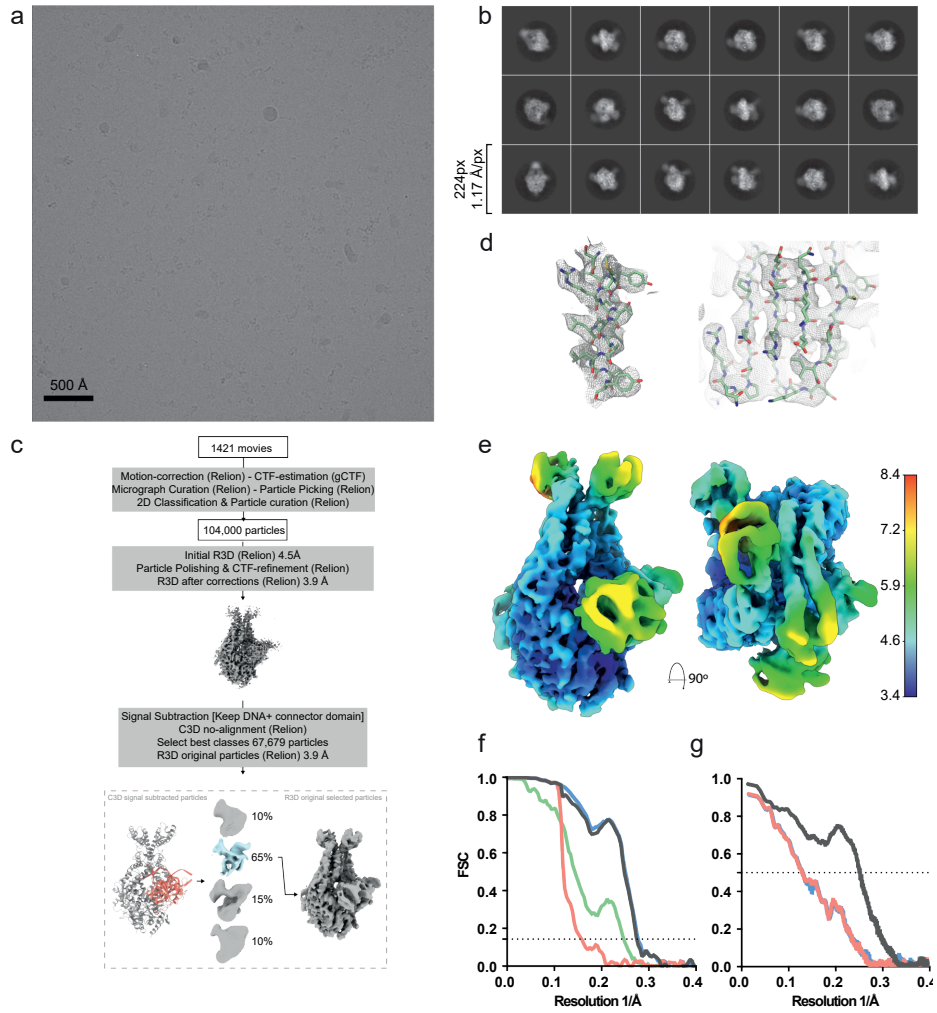
Zivanov, J., Nakane, T., Forsberg, B.O., Kimanius, D., Hagen, W.J., Lindahl, E., and Scheres, S.H. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. Elife *7*, 163.



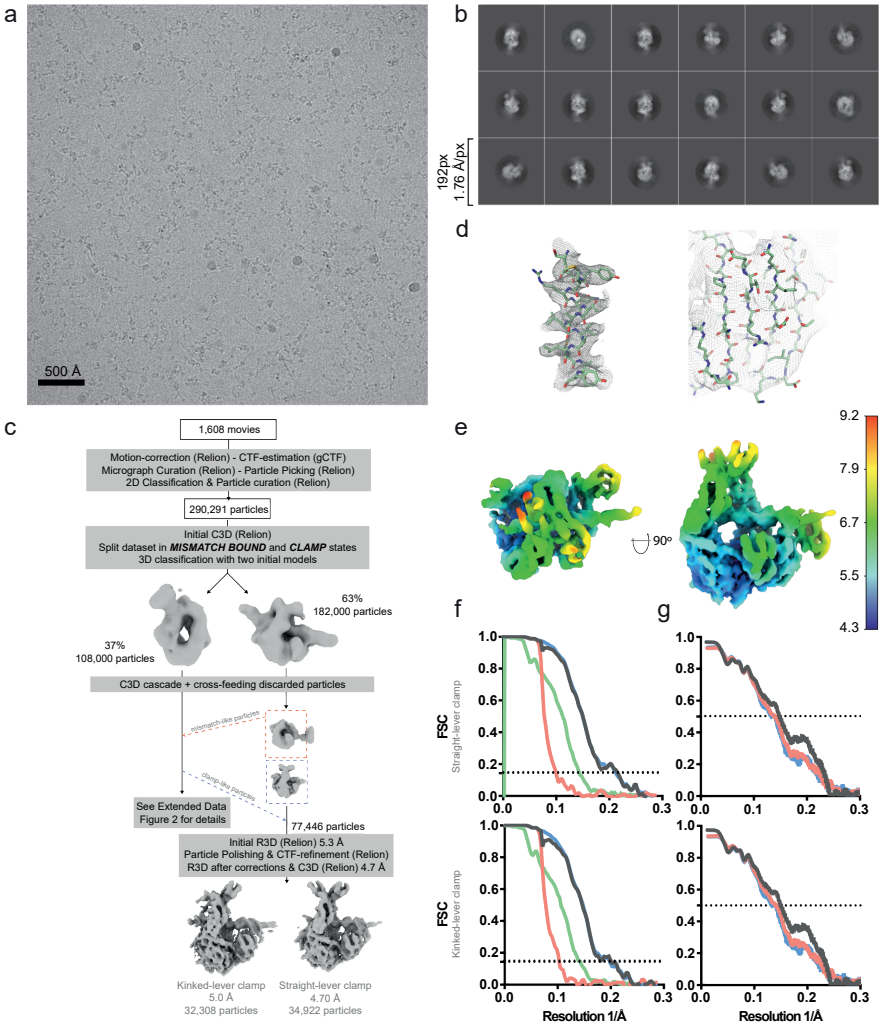
Supplemental Figure 1. CryoEM data analysis of scanning state MutS. a, Representative micrograph (0º tilt). Plasmid DNA can be observed on the micrograph with MutS dimers bound like beads-on-string. **b**, 2D class averages from full dataset. **c**, Schematic representation of main data processing procedures. See methods section for more details. **d**, Principle Component Analysis of multibody refinement data. **e**, Detail of model fit to map. **f**, Final refinement map colored by local resolution. **g**, Fourier Shell Correlation between half-maps from final refinement. Green line: unmasked. Blue line: masked. Red line: phase randomised. Grey line: corrected **h**, Model-vs-map Fourier Shell Correlation. Grey line: model vs full-map. Blue line: model refined against half-map 1, FSC against half-map 1. Red-line: shaked model refined against half-map 1, FSC against half-map 2.



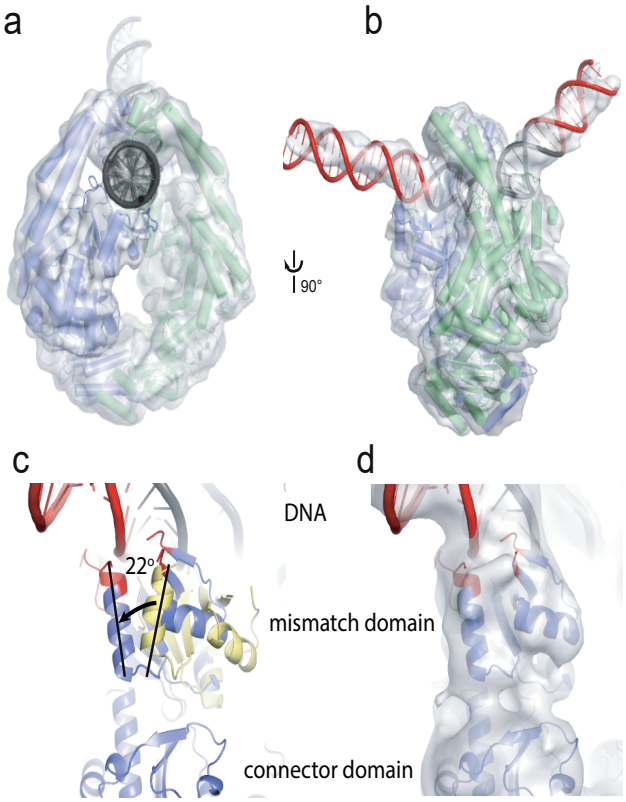
Supplemental Figure 2. CryoEM data analysis of mismatch-bound MutS. a, Representative micrograph. **b**, 2D class averages from full dataset. **c**, Schematic representation of main data processing procedures. See methods section for more details. **d**, Detail of model fit to map. **e**, Final refinement map colored by local resolution. **f**, Fourier Shell Correlation between half-maps from final refinement. Green line: unmasked. Blue line: masked. Red line: phase randomised. Grey line: corrected **g**, Model-vs-map Fourier Shell Correlation. Grey line: model vs full-map. Blue line: model refined against half-map 1, FSC against half-map 1. Red-line: shaked model refined against half-map 1, FSC against half-map 2.



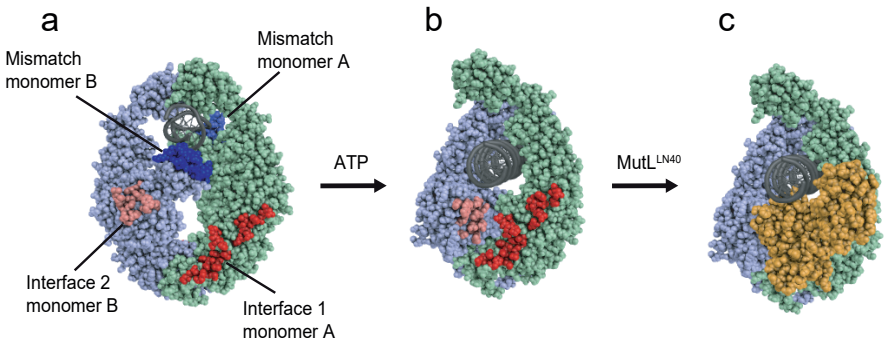
Supplemental Figure 3. CryoEM data analysis of intermediate state MutS. **a**, Representative micrograph. **b**, 2D class averages from full dataset. **c**, Schematic representation of main data processing procedures. See methods section for more details. **d**, Detail of model fit to map. **e**, Final refinement map colored by local resolution. **f**, Fourier Shell Correlation between half-maps from final refinement. Green line: unmasked. Blue line: masked. Red line: phase randomised. Grey line: corrected **g**, Model-vs-map Fourier Shell Correlation. Grey line: model vs full-map. Blue line: model refined against half-map 1, FSC against half-map 1. Red-line: shaked model refined against half-map 1, FSC against half-map 2.



Supplemental Figure 4. CryoEM data analysis of MutLLN40-bound MutS. a, Representative micrograph. **b**, 2D class averages from full dataset. **c**, Schematic representation of main data processing procedures. See methods section for more details. **d**, Detail of model fit to map. **e**, Final refinement map colored by local resolution. **f**, Fourier Shell Correlation between half-maps from final refinement. Green line: unmasked. Blue line: masked. Red line: phase randomised. Grey line: corrected g, Model-vs-map Fourier Shell Correlation. Grey line: model vs full-map. Blue line: model refined against half-map 1, FSC against half-map 1. Red-line: shaked model refined against half-map 1, FSC against half-map 2.

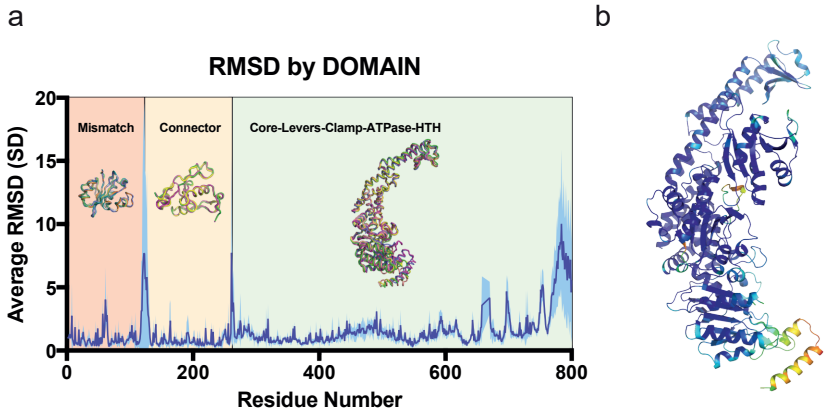


Supplemental Figure 5. Comparison of crystal and cryo-EM structure of mismatch-bound MutS. **a**, Front view of the crystal structure of mismatch-bound MutS (1E3M) fitted into in the cryo-EM map of mismatch-bound MutS. **b**, Side view, with additional stretches of DNA shown in red. **c**, Close up of the mismatch binding domain of monomer B with former position of the mismatch binding domain shown in yellow and the ~22° domain rotation indicated by a black arrow. Additional visible loops, modeled after the mismatch binding of monomer A, are highlighted in red. **d**, Same view as in panel c, showing the cryo-EM map for the DNA, mismatch binding domain and connector domain.



Supplemental Figure 6. Mismatch and MutL binding surfaces in MutS. a, Mismatch-bound MutS showing the mismatch and MutL binding surfaces. MutS monomer A in green, monomer B in light-blue, mismatch binding in dark blue, MutL Interface 1 in dark red, and MutL Interface 2 in light red. **b**, Binding of ATP transforms MutS into a sliding clamp that brings MutL-Interface 1 and 2 together, creating the binding site for MutL **c**, MutS bound to MutL^{LN40}.

Supplemental Figure 7. Engineered MutS homo- and heterodimers. Front view (top row) and reverse view (bottom row) of the MutS dimer **a**, WT MutS showing mismatch and MutL binding surface on both sides of the dimer **b**, Interface 1 mutant lacks one half of MutL-binding site on either side of the MutS dimer. **c**, Similarly, Interface 2 mutant also lacks one half of the MutL-binding site on either side of the MutS dimer. **d**, Mismatch mutant lacks the mismatch binding site in both monomers. **e**, Mixing of Interface 1 and Interface 2 mutant creates a heterodimer with an intact MutL binding site on one side, but lacks the MutL-binding interface on the reverse side. **f**, A double Mismatch-Interface 1 mutant mixed with an Interface 2 mutant generates a MutS dimer with an intact MutL binding site on one side and the mismatch binding site on the reverse side. This configuration is similar to the eukaryotic MSH2-MSH6 and MSH2-MSH3 hetero dimers. **g**, A double Mismatch-Interface 2 mutant mixed with an Interface 1 mutant generates a MutS dimer with both MutL and mismatch binding sites on the same face of the MutS dimer.



Supplemental Figure 8. Maps and models. a, Graph showing average by-residue RMSD from all structures presented (Scanning, Mismatch, Intermediate and MutS-MutL^{LN40} Clamp) when compared to the mismatch-bound MutS crystal structure 1EM3. Chain A from all structures were superimposed. Mismatch and conector domains where superimposed indepenntly from the rest of the molecule. **b**, MutS monomer (1E3M) colored based on average RMSD values calculated as described in a.

Extended Data Table 1: Cryo-EM data collection, refinement and validation statistics

	MutS-scan PDB 7AI6	MutS-mismatch PDB 7AI5	MutS-intermediate PDB 7AI7	MutS-clamp PDB 7AIB	MutS-kink-clamp PDB 7AIC
Data collection and processing					
Magnification	105,000	64,000	105,000	64,000	64,000
Voltage (kV)	300	300	300	300	300
Electron exposure ($e^-/Å^2$)	40	40	40	40	40
Defocus range (μm)	0.5 – 4.0	1.5 – 3.0	1.5 – 3.0	1.5 – 3.0	1.5 – 3.0
Pixel size (Å)	1.17	1.76	1.17	1.76	1.76
Symmetry imposed	C1	C1	C1	C1	C1
Initial particle images (no.)	229,664	290,291	104,000	290,291	290,291
Final particle images (no.)	36,682	13,457	67,679	34,922	32,308
Map resolution (Å)					
FSC threshold=0.143	4.4	6.9	3.9	4.7	5
Map resolution range (Å)	4.3-8.3	6.8-17.4	3.4-8.3	4.3-9.2	4.5-9.4
Refinement					
Initial model used	PDB 1E3M	PDB 1E3M	PDB 1E3M PDB 5AKB	PDB 5AKB	PDB 5AKB
Model resolution (Å)					
FSC threshold=0.5	4.9	7.3	4.0	6.7	6.7

Extended Data Table 1. Continued

	MutS-scan PDB 7AI6	MutS-mismatch PDB 7AI5	MutS-intermediate PDB 7AI7	MutS-clamp PDB 7AIB	MutS-kink-clamp PDB 7AIC
Map sharpening *B* factor (Å^2)	Loc-Scale	Loc-Scale	Loc-Scale	Loc-Scale	Loc-Scale
Model composition					
Non-hydrogen atoms	13,378	13,474	12,003	13841	13,826
Protein residues	1,576	1,577	1,460	1,591	1,589
Ligands	46	51	23	60	60
B factors (Å^2)					
Protein	125.63	104.64	97.52	139.72	108.68
Ligand	140.00	125.01	119.47	140.0	129.94
R.m.s. deviations					
Bond lengths (Å)	0.014	0.014	0.015	0.007	0.013
Bond angles (°)	1.783	1.781	1.906	0.903	1.615
Validation					
MolProbity score	1.61	1.44	1.51	1.71	1.99
Clashscore	4.55	3.46	4.95	4.18	6.21
Poor rotamers (%)	0.80	0.60	0.20	0.60	1.50
Ramachandran plot					
Favored (%)	97.07	95.92	96.22	93.84	95.80
Allowed (%)	2.93	4.08	3.78	6.16	4.14
Disallowed (%)	0.00	0.00	0.00	0.00	0.06

Extended Data Table 2: DNA substrates

DNA oligonucleotide	length	Sequence	Modification
Heteroduplex top	61	5′-**/bio/**CTG AAG CTT AGC TTA GGA TCA TCG AGG ATC **G**AG CTC GGT GCA ATT CAG CGG TAC CCA ATT C-3′	5′ biotin
Heteroduplex bottom	61	5′-**/bio/**GAA TTG GGT ACC GCT GAA TTG CAC CGA GCT **T**GA TCC TCG ATG ATC CTA AGC TAA GCT TCA G-3′	5′ biotin
Homoduplex top	50	5′-CTT AGC TTA GGA TCA TCG AGG ATC GAG CTC GGT GCA ATT CAG CGG TAC CC-3′	n.a.
Homoduplex bottom	50	5′-GGG TAC CGC TGA ATT GCA CCG AGC TCG ATC CTC GAT GAT CCT AAG CTA AG-3′	n.a.
Amino-modified top	59	5′-GTG CGC AAA TCC AGA CGT CTG TC**G** ACG **T**TG GGA AGC TTG AGT ATT CTA TAG TGT CAC CT-3′	Amino modified T at position 28
Amino-modified bottom	59	5′-**/Cy5/**AGG TGA CAC TAT AGA ATA CTC AAG CTT CCC AAC GT**T** GAC AGA CGT CTG GAT TTG CGC AC-3′	5′ Cy5 fluorophore

Mismatched base pairs are highlighted in bold-red font.

CHAPTER 4

Studying Full length MutS-MutL on DNA by Cryo-Electron Microscopy

Doreth Bhairosing-Kok[1]
Rafael Fernández-Leiro[2,3]
Alexander Fish[1]
Herrie H. Winterwerp[1]
Meindert H. Lamers[2,4]
Titia K. Sixma[1]

[1] *Division of Biochemistry and Oncode Institute, Netherlands Cancer Institute, Amsterdam, The Netherlands*
[2] *MRC Laboratory of Medical Research, Cambridge, United Kingdom*
[3] *Spanish National Cancer Research Centre, Madrid, Spain*
[4] *Department of Cell and Chemical Biology, Leiden University Medical Center, Leiden, The Netherlands*

ABSTRACT

DNA mismatch repair (MMR) is responsible for correcting errors that are formed during DNA replication. MMR proteins MutS and MutL are initiating the repair. *E.coli* MutS scans newly-replicated DNA and recruits MutL after it has detected a mismatch. Molecular matchmaker MutL in turn recruits other MMR proteins to repair the mismatch. The recruitment of MutL by MutS has been of great interest to the MMR field. The structure of MutS-MutL^{LN40} revealed the bi-partite interaction site of MutL^{LN40} with both MutS monomers. However, the recruitment of full length MutL is still not fully understood. Here, we study complex formation of MutS and MutL and analyze full length MutS-MutL on DNA and MutL on DNA using cryo-electron microscopy (cryo-EM). Processing these data showed that MutS and MutL form individual complexes with DNA and the heterotrimeric complex of MutS-MutL-DNA is very transient. Options for future experiments are discussed.

INTRODUCTION

DNA mismatch repair (MMR) is an evolutionary conserved pathway that prevents the presence of mismatches in newly synthesized DNA by repairing these mismatches. Among all DNA repair pathways, MMR is unique by focusing on mismatches, insertions and deletions [1], that are incorporated incorrectly during DNA replication.

MMR is important to keep mutation rates low and prevent microsatellite instabilities (MSI). Together with the proofreading mechanism of the polymerase, MMR keeps the mutation rate to a minimum. Defects in MMR can lead to a mutator phenotype, which results into a predisposition of cancer development in humans, also known as Lynch Syndrome or HNPCC [2, 3].

The first MMR protein in the pathway, MutS, interacts with the replication machinery, scanning for mismatches that are present in the newly synthesized DNA [4]. *E.coli* MutS forms a dimer around the DNA helix and it scans the DNA using its DNA-interaction domains. Upon mismatch recognition, MutS stacks a conserved phenylalanine on the mismatched base. This movement coincides with kinking of the DNA at the region where the mismatch is located, allowing MutS to intrude with its phenylalanine residue [5, 6]. Next, MutS undergoes a series of ATP-induced conformational changes resulting in a so-called sliding clamp formation [7, 8] (Chapter 3). The MutS sliding clamp allows loading of the second MMR-protein, MutL.

MutL and their homologs are composed of two domains; an N-terminal domain (NTD) and a C-terminal domain (CTD) with the first one often referred as LN40, given its molecular weight. MutL also forms a dimer and dimerizes via its CTD. In addition, upon ATP binding, the N-terminal ATPase domains can dimerize too, forming a closed system. As mentioned, MutS can load MutL on the DNA, which will lead to the next stage in the MMR cascade. Depending on the species, MMR continues by nicking the DNA in the newly synthesized daughter strand by MutL (eukaryotes). Alternatively, in *E.coli* and other gamma-bacteria, this task is performed by a separate protein called MutH. The nuclease function of MutH or MutL results in a nick in the newly synthesized strand of DNA, which will be the start of unwinding of the DNA and partial removal of the new strand. Finally, the new strand will be resynthesized and ligated by the replication machinery.

Understanding the role and activation of MutS and MutL and their co-operative mechanism is our main goal. How information on the presence of a mismatch is sensed and transmitted selectively in a pool of homoduplex DNA remains of great interest. We have information how MutS identifies the mismatch [5] and on its conformational changes that leads to MutL binding [7-12]. However, we know very little about the subsequent steps, that lead to MutL conformational changes that will allow nuclease activation.

These studies are complicated by the fact that both MutS and MutL adopt multiple conformations during MMR initiation. Förster resonance energy transfer (FRET) and Atomic force Microscopy (AFM) experiments, and X-ray crystallography structures showed that both MutS and MutL can adopt several conformations independently of each other [5, 13-16]. Single molecule, Surface Plasmon Resonance (SPR), FRET and crosslinking experiments show that MutS and MutL are interacting with each other on DNA [8, 12, 17-19]. Regardless of these efforts, our understanding of how full length MutS activates and interacts with MutL is not complete.

Recently, a crystal structure of *E.coli* $MutS^{DC800}$ crosslinked to $MutL^{LN40}$ was solved, showing the interaction of $MutL^{LN40}$ with the MutS dimer [8]. This showed the large conformational changes that take place between DNA mismatch recognition and clamp formation and how this creates a bipartite interaction site for MutL. In chapter 3, we presented follow-up on this work, revealing how the MutS clamp interacts with DNA and how the $MutL^{LN40}$ is positioned. This latter study made use of the crosslinked material, but now employed by cryo-electron microscopy (cryo-EM).

In this chapter, we took different approaches to trap full length *E.coli* MutS and MutL on DNA to capture relevant states of MMR initiation for high-resolution structural analysis. We first analyzed MutS-MutL complex stability on DNA, using SPR. Then we describe a series of cryo-EM datasets collected under different conditions of DNA/MutS/MutL complexes as well as single MutL/DNA data set. Table 1 shows an overview of these four datasets. Cryo-EM analysis of MutS-MutL on DNA was difficult and compositional and conformational heterogeneity

prevented us from structure solution. Therefore, we analyse the cryo-EM analysis and we end with some recommendations for future studies.

Table 1

Dataset	Proteins	DNA construct	Microscope & camera	Location
A0	$MutS^{WT/D835R}$ + $MutL^{WT/N33A}$	100GT50 + MonoStrep	Tecnai CCD	LMB, Cambridge
A1	$MutS^{D835R}$ + $MutL^{WT}$	100GT50 + MonoStrep	Titan Krios K2	LMB, Cambridge
A2	$MutS^{D835R}$ + $MutL^{WT}$	100GT50 + MonoStrep	Titan Krios K2	NeCEN, Leiden
B	$MutS^{D835R}$ + $MutL^{WT}$	65GT33	Titan Krios K2	NeCEN, Leiden
C	Single cys $MutS^{D246C\_D835R}$ + Single cys $MutL^{N131C}$	65GT33	Talos Arctica F3	CNB, Madrid
D	$MutL^{WT}$	50GC25	Titan Krios F3	RMS, Leicester

MATERIALS AND METHODS

1 – Expression constructs and protein purification

$MutS^{D835R}$ was created in the MutS gene in vector pET-3D [5, 13]. Single cysteine $MutS^{D246C}$ was obtained as described previously [19]. D835R mutation was introduced in this background using the QuikChange Site-Directed Mutagenesis Kit (Stratagene) and appropriate primers (IDT). $MutL^{N33A}$ was created in the MutL gene in plasmid pTX418 [20]. Single cysteine $MutL^{N131C}$ was obtained as described previously [21].

MutS WT and mutants were purified as described previously, except that the final gel filtration buffer contains 250 mM KCl instead of 150 mM KCl [8, 13]. MutL WT and mutants were purified as described in Chapter 3. All proteins were purified and stored at -80°C and thawed before experiments.

1.1 – Dataset A

Monovalent streptavidin (Streptavidin Alive 1 – Dead 3) (from now: Monostrep) was used to block biotinylated oligos. Monostrep is a variant where in three of the four monomers the binding activity was inactivated (Streptavidin-Dead (D) [22]. It is made by separate expression of differentially tagged Streptavidin-Alive (SAe) and Streptavidin-dead (D) in a 3:1 ratio of the two proteins. Both plasmids for SAe and D were transformed separately into *E.coli* BL21(DE3), plated onto LB

agar plates and incubated at 37 °C o/n. A single colony of SAe was used to grow in 4 liter LB media. A singe colony of D was used to grow 2 liter LB media. Both cell cultures grew at 37°C. Both proteins were purified separately and combined when indicated. Bacteria were grown until OD_{600} reached 1.0, followed by induction by 1 mM IPTG and cells continued growing for 2h at 30 °C. All purification steps were performed at 4 °C. Cells were resuspended in lysis buffer (50 mM Hepes 7.5, 300 mM NaCl) (no need for reducing agents) and lysed by sonication for 7 minutes using intervals (45 seconds on and 15 seconds off). Lysed cells were pelleted and resuspended in 60 mL milliQ, followed by homogenization. Cells were spun again for 30 min at 27.220 x g, and the pellet was dissolved in 16 mL 6M Guanidinium Chloride to unfold the protein. Concentrations were estimated and D and SAe were mixed in a 4:1 ratio relatively (to ensure for highest yield of 3:1 possible). The mixture was added dropwise to 1 L refolding buffer (50 mM Hepes 7.5, 100 mM NaCl), and a new equilibrium of tetramers was formed overnight at 4 °C while stirring. Next day, 15 mL equilibrated talon beads were added to the refolded protein to bind the his-tag of SAe. The mixture was transferred to a gravity column and beads were washed with wash buffer (50 mM Hepes 7.5, 100 mM NaCl and 5 mM Imidazole), followed by a wash with lysis buffer and eluted in elution buffer (50 mM Hepes 7.5, 100 mM NaCl, 350 mM Imidazole). Eluted fraction was loaded onto an equilibrated Resource Q column and eluted in a 30 column volumes gradient ranging from 5 mM to 1 M NaCl. Fractions were analyzed on SDS-page gel and fractions containing streptavidin in a 3:1 ratio (D:SAe) were pooled and concentrated before loading onto a Superdex S200 16/60 column. Fractions were analyzed on SDS-page gel, concentrated and plunge frozen and stored at -80°C.

2 – Crosslinking MutS-MutL (Dataset C)

Crosslinking procedure of $MutS^{D246C\text{-}D835R}$ with $MutL^{N131C}$ was based on the crosslinking protocol for $MutS^{\Delta C800}$-$MutL^{LN40}$ [8], with some adjustments. We started with a molar ratio of 1:2 for MutS:MutL to ensure full saturation of the MutS monomer with MutL dimers. MutL was diluted to 2 µM MutL before mixing with the crosslinker. Instead of mixing equimolar DNA and MutS, a ratio of 2:1 was used (DNA:MutS) and incubated for 20 min instead of 10 min. Final gel filtration patterns of the full length crosslinked complex looked different than for $MutS^{DC800}$-$MutL^{LN40}$, given its larger size. All fractions were analyzed on SDS-page and correct fractions were plunge frozen and stored at -80°C.

3 – Analytical gel filtration – Äkta Micro

Protein complexes were purified and brought to homogeneity using analytical size exclusion chromatography at microscale using an Äkta Micro system (GE Healthcare). This was done both in test settings, to define the best order of assembly, and for data collection, where it allowed quick buffer exchange. For several datasets, this has been performed on different columns (see below). Binding buffer for test runs was 25 mM Hepes 7.5, 150 mM KCl and 5 mM $MgCl_2$.

3.1 – Dataset A

Annealed DNA oligos 100GT50 were purified over Mono Q (1.6/5) 0.1 ml column (GE Healthcare). Gel filtration column Superdex S200 5/150 (GE Healthcare) was used to test binding of Monovalent Streptavidin to biotinylated 100GT50, and binding of MutS to blocked 100GT50. Assembly of MutS and MutL on blocked DNA was tested using Superose 6 increase 3.2/300 (GE Healthcare).

3.2 – Dataset C

Crosslinked MutS-MutL (MutSxMutL) on 65GT33 was purified over a 2.4 mL S200 increase 3.2/300 (GE Healthcare) gel filtration column.

4 – DNA binding experiments by surface plasmon resonance

Surface plasmon resonance (SPR) experiments were performed using a Biacore T200 system (GE Healthcare) at 25°C, with the same set-up as described previously [13]. All details regarding experiments are listed in figure legend.

5 – Protein solubility screen

A Mosquito (TTP Labtech) Microsyringe-based nanoliter dispenser was used to set up droplets; 100 nL $MutL^{N33A}$ and 300 nL buffer, and incubated for 18 h at 20°C. JBS solubility screen (Jena Bioscience) was used to test various buffer conditions and 96-well MRC 2-drop plates were used (Swissci). Solubility scoring was done by checking precipitation under the microscope.

6 – Grid preparation & Data collection

6.1.1 – Dataset A0 to determine conditions of grid preparation

To decide which combination of proteins to use for large scale data collection, several test datasets have been collected with combinations of $MutS^{WT}$/$MutS^{D835R}$

with $MutL^{WT}$/$MutL^{N33A}$. Blocked heteroduplex DNA was made by mixing 100GT50 (sequence see [13]) and monovalent streptavidin, followed by purification over a Superose 6 increase 3.2/300 gel filtration column on an Äkta Micro system (GE healthcare). 46 mM 100GT50 was mixed with 111 mM Monostrep. Then, 90 mL was loaded and the peak fractions were used for further analysis (concentration ±15 mM).

The binding buffer for MutS and MutL on DNA, was optimized by a protein solubility screen and was used for all grids (20 mM KH_2PO_4 pH 7.4, 30 mM KCl, 5 mM $MgCl_2$, 1 mM DTT). For grid type 1, blocked DNA was mixed with MutS and incubated for five minutes at room temperature. Next, 50 mL of the mixture (12 mM DNA_Strep + 33 mM MutS) was purified using Superose 6 increase column and the peak fraction was used. During purification, a mixture of 4 mM MutL and 1 mM ATP was prepared and stored at 4°C. When the mixture of MutS and DNA came off the column, Tween-20 was added (0.0066%). Then, the MutS-DNA-Tween mixture was mixed with the MutL-ATP mixture (1:1) and used for grid preparation. For grid type 2, DNA_Strep and MutS were mixed and incubated for 5 minutes at room temperature. Separately, a mixture was made of MutL and ATP. These mixtures were mixed (1:1) with a final concentration of 11 mM blocked DNA, 22 mM MutS, 20 mM MutL and 1 mM ATP.

All test grids were glow-discharged Quantifoil 1.2/1.3 400-mesh grids, and vitrified using a manual plunger at 4°C. Test datasets were collected on a 120 kV Tecnai (FEI) microscope with a CCD camera.

6.1.2 – Dataset A1 - LMB

Grids were prepared as described above (grid type 1), except a Vitrobot (FEI, Hillsboro, OR) at room temperature was used instead of a manual plunger. Grid type was Cupper Quantifoil 1.2-1.3 300 mesh. All data was collected using a Titan Krios electron microscope (FEI) operated at 300 kV equipped with a K2 summit direct electron detector (Gatan, Pleasaston, CA) at the Laboratory of Molecular Biology (LMB, Cambridge, UK). A total of 1288 movies were manually collected in single-electron counting mode at 1.76 Å/pixel, using a total dose of 40 $e^-/Å^2$ over a total of 20 frames.

6.1.3 –Dataset A2 - NeCEN

Grids were prepared as described above (2.6.1.2). All data was collected using a Titan Krios electron microscope (FEI) operated at 300 kV equipped with a K2 summit direct electron detector (Gatan, Pleasaston, CA) at the Netherlands Centre for Electron Nanoscopy (NeCEN, Leiden, NL). In total 1519 micrographs were collected using EPU (ThermoFisher Scientific), in super resolution mode at 1.76 Å/pixel, using a total dose of 40 $e^-/Å^2$ over a total of 20 frames. Micrographs were collected with a defocus range of 1.4-3.0 mm.

6.2 – Dataset B

Grids were prepared as described above (2.6.1.2) except 65GT33 DNA was used instead of 100GT50 and the oligos were not blocked (5'AGCTGAAGCTTAGCTTAGGATCATCGAGGATCGAGCTCGGTGCAATTCAGCGGT ACCCAATTCGC-3' annealed with 5' GCGAATTGGGTACCGCTGAATTGCACCGAGCT TGATCCTCGATGATCCTAAGCTAAGCTTCAGCT 3'), and grids had 300 mesh instead of 400. All data was collected using a Titan Krios electron microscope (FEI) operated at 300 kV equipped with a K2 summit direct electron detector (Gatan, Pleasaston, CA) at the Netherlands Centre for Electron Nanoscopy (NeCEN, Leiden, NL). In total 1730 movies were collected using EPU (ThermoFisher Scientific). Movies were collected at 1.086 Å/pixel, two micrographs per hole and a total dose of 40 $e^-/Å^2$ over a total of 21 frames. Micrographs were collected with a defocus range of 1.8-3.0 mm.

6.3 – Dataset C

In this dataset the crosslinked MutSxMutL was combined with a 65GT33 DNA oligomer. Vitrobot was prepared at 4°C and 85-90% humidity. Cupper Quantifoil 0.6-1.0 300 mesh grids were used. Grids were glow discharged (30", 25 mAmp, 0.1 bar) and MutSxMutL was thawed and loaded on an equilibrated Superdex S200 increase 3.2/300 (GE Healthcare) on an Äkta Micro system (GE healthcare) in buffer 25 mM Hepes 7.5, 150 mM KCl and 5 mM $MgCl_2$, for buffer exchange. Peak fraction was mixed with 65GT33, with a final concentration of ± 0.7 mg/mL MutSxMutL and 5 mM DNA. The MutSxL-DNA sample was mixed with ATP and 3 mL were loaded on the grid. Final concentration of ATP is 1 mM. Grids were Cupper Quantifoil 0.6/1.0 300 mesh.

Data was collected using a Talos Artica operated at 200kV equipped with a FalconIII direct electron detector at the Centro Nacional de Biotecnología (Madrid, Spain). In total 666 movies were recorded using EPU at 0.86 Å/pixel. Total dose of 35 $e^-/Å^2$ over a total of 70 frames. Micrographs were collected at a defocus range of 2.0-3.6 mm.

6.4 – Dataset D (MutL)

In this dataset MutL alone was mixed with homoduplex DNA (50GC25) at low salt conditions (25 mM Hepes 7.5, 50 mM KCl, 5 mM $MgCl_2$). to promote MutL binding to DNA [17, 23]. The Vitrobot was prepared at 4°C and 85-90% humidity and grids were glow discharged (45", 25 mAmp, 0.1 bar). 50GC25 (5' CTTAGCTTAGGATCATCGAGGATCGAGCTCGGTGCAATTCAGCGGTACCC 3' was annealed with 5' GGGTACCGCTGAATTGCACCGAGCTCGATCCTCGATGATCCTAAG CTAAG 3'). MutL was diluted to 1 mM, 7.5 mM 50GC25 was used, combined with 1 mM ATP. MutL and DNA were pre-mixed and ATP was added prior to grid freezing.

All grids were glow-discharged Copper Quantifoil 0.6/1.0 300 mesh grids. Dataset was collected on a Titan Krios electron microscope (FEI) operated at 300 kV equipped with a K2 summit direct electron detector (Gatan, Pleasaston, CA) and a Volta Phase plate (FEI) [24, 25] at the Midlands Regional cryo-EM facility (Leicester, UK). In total 1583 micrographs were collected at 1.4 Å/pixel, and a total dose of 35.5 $e^-/Å^2$ over a total of 75 frames. Defocus set-up was 0.6 and 0.8 mm.

7 – Data processing

All processing was performed in Relion 3.0 [26]. For all movies for all datasets, pre-processing was done by MotionCorr [27] and gCTF [28]. Particle picking was done as indicated and 2D and 3D classification was performed in Relion, unless stated otherwise.

7.1.1 Dataset A1 – LMB

In total 1288 movies were collected, after micrograph inspection 1153 remained. Of this set, 50 micrographs were used to pick particles (box 140 px) using Relion Autopick. 2D classification (mask 160 Å) with this subset was performed to generate 2D classes that served as reference for Autopick with reference. This was

then performed on all 1153 micrographs. Particles were extracted (1.849.097) with a box of 140 pixels. Several rounds of 2D and 3D classification were performed.

7.1.2 Dataset A2 – NeCEN

During a 72-hour data collection, 1519 micrographs were collected. After micrograph inspection, 338 were removed primarily due to a high number of movies that were too close to focus and 1081 were retained for processing. Particle picking showed difficulties due to the large differences in defocus values between micrographs (1.4-3.0 mm). For this reason, micrographs were split in two groups based on the defocus value (cut off 2.0 mm). Particle picking was performed using Relion Autopick using 2D references from Dataset A obtained at the LMB, with different autopick settings for the two sets of micrographs; For the low defocus micrographs we used a higher picking threshold (0.7) and for high defocus micrographs we used a low threshold (0.05). Particles were extracted (box 140 px) and joined, having a total number of 1.656.460 particles picked. Due to the mild cut-off in autopicking, relatively many of the picked particles were noise, resulting in so called Einstein-from-noise classes [29]. Several rounds of 2D classification (mask 160 Å) were needed to eliminate the majority of these particles. After this, class averages (particles) were divided into categories based on the view; top, front, side and other, in order to remove false particles optimally. After these rounds, 180.074 particles remained and were used for 3D classification. A low-pass model of the cryo-EM map of MutS-MutL^{LN40} (Chapter 3) was used as an initial model for the first round of 3D classification. Several 3D classification runs followed to obtain a better 3D model.

7.2 – Dataset B

During a 72-hour data collection, 2016 micrographs were collected After micrograph inspection 1728 remained. Particles were picked using Gautomatch. First, particles were extracted from 50 micrographs (12.672 particles) and two rounds of 2D classifications were done. After satisfactory initial 2D classification analysis, we decided to proceed with the full data set. All picked particles with Gautomatch were extracted from all 1728 micrographs (461.010 particles, box 110 pixels (binned twice). After five rounds of 2D classification, particles were re-extracted without binning (187.028 particles, box 220 pixels). Several attempts were done to obtain classes of MutS-MutL on DNA; splitting on views/orientations

and on size. Different mask sizes were tried in the range of 140 Å – 200 Å depending on the size of the particles.

7.3 – Dataset C

In total, 666 movies were collected and 577 remained after micrograph inspection. Using Autopick (Relion), 325.165 particles were picked and extracted (box 200 px). Several rounds of 2D classification did not show any MutS-like 2D classes. Manual picking was tried by looking for MutS-shaped particles. This did not show any MutS-like 2D classes. Focusing on a selection of DNA-bound classes for 2D classification was done, to stimulate the formation of MutS-MutL-DNA classes. Finally, 2D classification was done in a more standard way; keeping good particle and discard bad particles. a few representative good classes were used as a reference for re-picking particles by Autopick using a 2D-reference. Again, several rounds of 2D classification were done. Different mask sizes were tried in the range of 140 Å – 160 Å depending on the size of the particles.

7.4 – Dataset D

This dataset has been processed twice; first at the NKI and secondly at the CNIO.

7.4.1 – Processing Dataset D at NKI

In total, 1583 movies were collected. First selection was based on motion correction metadata were 1302 remained. After this selection, micrographs were checked manually and 1195 micrographs were selected for further analysis. Particle picking was done using Gautomatch and particles were extracted (382.554). Several box sizes were tested (125 px, 150 px, 180 px, 200, 240 px) but box size of 240 px seemed the best size. Several rounds of 2D classification were done (mask 180 Å). Good, representative classes were used for Autopick with reference in Relion, but several rounds of 2D classification showed worse classes than Gautomatch-picked particles. After several rounds of 2D, we categorized the particles in three groups: MutL+DNA, only MutL and small looking classes. After these 2D-runs, particles were selected into two groups: 'single' MutL with DNA (27.790 particles) and 'double' MutL with DNA (14.267 particles). Each subgroup was used for 3D classification.

7.4.2 – Processing Dataset D at CNIO

CTF correction was done for all 1583 movies using Warp [30]. Based on CTF estimated resolution, two selections were made; all movies up to 4.0 Å (1134 movies) and up to 7.0 Å (1473 movies). Picking was also done by Warp and coordinates were imported and extracted in Relion (644.733 particles) (box 240 px). Several rounds of 2D classification showed good classes (mask 180 Å) but also some classes that were off-center. These classes were separated, centered manually by using a custom script and joined with the remaining particles again. Final selection of particles was used (67.152 particles) to obtain 3D-models using Cryosparc [31]. Three models were obtained and discussed.

8 – Fitting cryo-EM data

The 3D cryo-EM map of dataset D was used to fit two single crystal structure models of $MutL^{LN40}$ and a straight DNA molecule. Chimera [32] was used to fit these models in the cryo-EM map.

RESULTS

Dataset A1 & A2

Optimization of complex stability and assembly of MutS-MutL on DNA

To obtain a stable MutS-MutL complex on DNA, several parameters were tested and optimized. Since it is known that MutS-MutL can slide off the DNA, we blocked the DNA end. Double stranded 100mer DNA with a GT mismatch at position 50 (100GT50) was 5' biotinylated and blocked with Monostrep on either side (Fig 1A). Analytical gel filtration showed a peak shift from DNA to DNA bound with Monostrep. This corresponds to end-blocked oligo's and could be used for further experiments (Fig 1B).

Previous cryo-EM work showed that the presence of glycerol disturbs the contrast on the micrographs [33]. Therefore, MutL was purified in a phosphate buffer, instead of the usual buffer containing glycerol [34] (see Materials and methods). A solubility screen for MutL showed high solubility in $KHPO_4$-buffer, pH 7.5 (Supplemental figure 1). Therefore, we chose to work with a binding buffer containing $KHPO_4$. The KCl concentration was carefully chosen as such that the ionic strength, together with $KHPO_4$, is optimal for MutS ATPase activity [23].

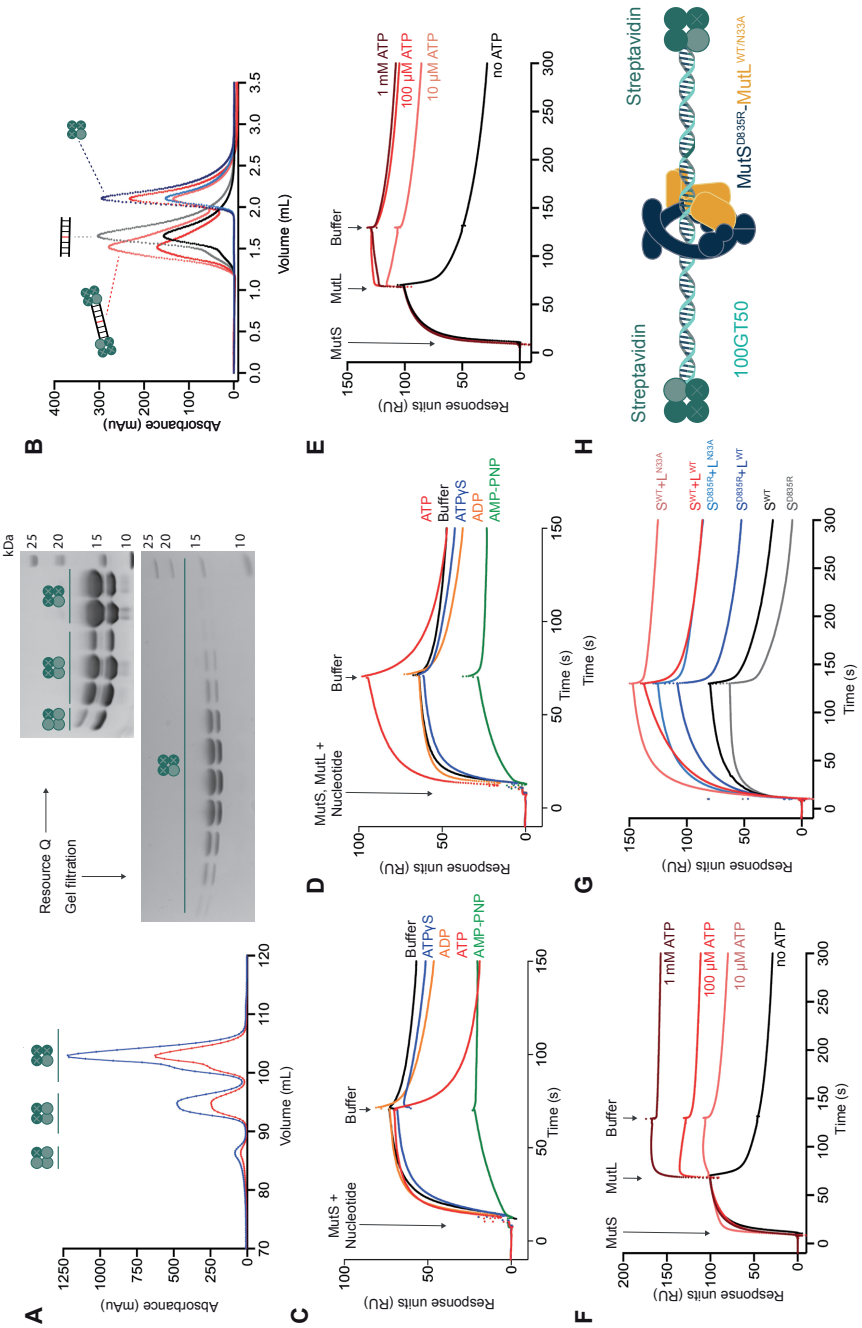

A
Absorbance (mAu)
Volume (mL)
kDa
Resource Q
Gel filtration
B
Absorbance (mAu)
Volume (mL)
C
MutS + Nucleotide
Buffer
ATPγS
ADP
ATP
AMP-PNP
Response units (RU)
Time (s)
D
MutS, MutL + Nucleotide
Buffer
ATP
Buffer
ATPγS
ADP
AMP-PNP
Response units (RU)
Time (s)
E
MutS
MutL
Buffer
1 mM ATP
100 µM ATP
10 µM ATP
no ATP
Response units (RU)
F
MutS
MutL
Buffer
1 mM ATP
100 µM ATP
10 µM ATP
no ATP
Response units (RU)
Time (s)
G
Response units (RU)
Time (s)
H
Streptavidin
Streptavidin
100GT50
MutS D835R-MutL WT/N33A

Figure 1. Optimization complex stability MutS-MutL on DNA for dataset A. **A)** Chromatogram of refolded Streptavidin and corresponding gel (top). The Asp-loop in Monovalent-Dead allows to seperate the different streptavidin tetramers on ionic strength. Each peak corresponds with a different tetramer; 3A-1D (left), 2D-2A (middle) and 3D-1A (right), as shown on the gel (top right). Final result of gel filtration fractions (bottom gel) showing 3:1 ratio of respectively monovalend Dead:Alive. **B)** Analytical gel filtration profiles of Streptavidin (blue), 100GT50 (black) and 100GT50 + Streptavidin (red). Dark curves are absorbance values measured at 280 nm, lighter curves at 260 nm. **C)** SPR binding curves of 200 nM MutS co-injected with 1 mM nucleotide for 60 s, followed by buffer injection without nucleotide. **D)** See C, only co-injection of MutS, MutL and nucleotide. **E)** Merged SPR curves of 4 individual experiments with 10 uM, 100 uM or 1mM ATP, or buffer (no ATP). First injection, 60 s 200 nM MutS, second injection 60 s 1200 nM $MutL^{WT}$ and followed by buffer injection (with corresponding ATP concentration). **F)** See E), but $MutL^{N33A}$ instead of $MutL^{WT}$. **G)** Merged SPR curves to summarize the combinations of $MutS^{WT/D835R}$ and $MutL^{WT/N33A}$. All dissociation curves are similar, indicating that the stability of the complexes are comparable. **H)** Schematic set-up of the complex; 100GT50 blocked with streptavidin with $MutS^{D835R}$ and $MutL^{WT}$.

MutS and MutL are ATPases and complex stability is correlated to the nucleotide state of both MutS and MutL (Chapter 1). Since MutS, MutL and DNA are mixed before grid preparation, using one type of nucleotide would be the most feasible option. Complex formation under varying nucleotide conditions was tested by performing Surface Plasmon Resonance (SPR) experiments. MutS binding was performed in the presence of nucleotides with and without MutL (Fig 1C, D). These data showed that to obtain a complex of MutS-MutL, ATP is required, in line with the need for clamp formation by MutS before MutL is bound [7, 10, 35].

Secondly, we needed to find the optimal ATP concentration. One hypothesis, was that low concentration of ATP (0-10 μM) would be sufficient for MutS conformational change, but would allow MutL to stay in the open conformation, while a high concentration of ATP (1-5 mM) results in fully condensed MutL [15, 34] (Personal communication P. Friedhoff). As an alternative approach, a mutant of MutL was tested; $MutL^{N33A}$, which is known for its inability to bind ATP and therefore is expected to remain in the open conformation. MutS-MutL binding was tested by performing SPR experiments, in the presence of various ATP concentrations (Fig 1E-F). The various ATP concentrations gave different outcomes; 1 mM ATP gave the highest binding response. For this concentration, there was only a mild difference observed between MutL and $MutL^{N33A}$. Therefore, this concentration was chosen for $MutL^{WT}$, because we predicted to have a fully closed conformation of MutL and therefore no heterogeneity in (MutS-)MutL conformations.

All combinations of $MutS^{WT/D835R}$ and $MutL^{WT/N33A}$ were tested (Fig 1G). Dissociation patterns of all combinations were comparable. We chose $MutS^{D835R}$ instead of $MutS^{WT}$ to avoid the formation of tetramers [36], to simplify the cryo-EM analysis. Finally, both MutL and $MutL^{N33A}$ were tested for grid formation (Fig 1H)

Grid preparation

We know that the complex of MutS-MutL on DNA is transient. Therefore, it is important to find the optimal way to assemble the complex. Besides obtaining a stable complex, we also aim for a homogeneous sample to make data processing feasible.

For this reason, we tested two methods to assemble the complex (Supplemental figure 2). The first one is to mix MutS with blocked heteroduplex DNA and run this over an analytical gel filtration column. The peak fraction should contain MutS bound to blocked DNA. Next, a mixture of MutL, ATP and Tween-20 was added to MutS-DNA before freezing the grid (grid type 1) (Supplemental figure 2A). Alternatively, the mixture of MutS and DNA was mixed with the mixture of MutL and ATP, and be analyzed by analytical gel filtration. Tween-20 and extra ATP were added to the peak fraction before freezing grids (grid type 2) (Supplemental figure 2A). Both grid types were tested on the microscope, as were combinations of the both WT and mutant MutS and MutL (Supplemental figure 2B-F). Grid type 1 with $MutS^{D835R}$ and $MutL^{WT}$ were chosen to collect a full dataset (Fig 2).

Data collection and processing

Two full datasets were collected on the same type of grid; one at the LMB and one at NeCEN (Fig 2) (Materials & Methods). Both datasets were processed separately. Processing of both datasets showed MutS in a clamp-state on DNA bound by extra density on only one side of MutS. This was expected since $MutSxMutL^{LN40}$ also showed only one $MutL^{LN40}$ bound to the MutS dimer, even when two $MutL^{LN40}$ modules were covalently crosslinked to the MutS dimer (Chapter 3). However, 3D classification of the data only showed a small volume adjacent to MutS. This could fit a single $MutL^{LN40}$, but not more, despite the fact that full length MutL was used. This indicates that the rest of the protein is flexible in this state. Although the presence of free MutL would in theory allow the crosslinked MutL to form a MutL dimer, we did not observe the $MutL^{LN40}$ dimer. This was seen after processing the LMB-dataset and no additional complexes or conformations were visualized with extra data that was obtained at NeCEN (Fig 2C).

Besides missing density for full length MutL, 2D classification showed some other limitations of these datasets. Many 2D classes were MutS(-MutL) on DNA projections seen from the top (Fig 2C). These preferred orientations were obtained

due to the long oligo that was used. Other states of MutS, collected on shorter oligos, resulted in less preferred orientations (Chapter 3).

Besides the preferred orientations, the presence of the Streptavidin block disturbed also the processing. Many particles, and therefore 2D classes, showed that MutS and/or MutL were very close to streptavidin. These particles were both included and excluded to see if they could contribute or disturb 3D classification of the MutS-MutL complex. They did not contribute in any way so these particles were left out prior to 3D classification.

Both datasets showed that a shorter oligo was necessary to prevent preferred orientations during freezing and that the presence of a DNA block hindered the processing of the mobile MutS-MutL complex.

Dataset B

Optimization of complex stability and assembly of MutS-MutL on DNA

To avoid the preferred orientations of MutS and DNA, that was obtained in dataset A, we studied the most optimal oligo length. The oligo should be as short as possible but still long enough to fit MutS-MutL on DNA. SPR experiments were performed on various oligo lengths, always with a GT mismatch in the middle (Fig 3A). This analysis showed that double stranded DNA of 65 base pairs is sufficient to fit a MutS-MutL complex (Fig 3A, B).

Besides the oligo length, the presence of the streptavidin blocks was complicating the cryo-EM analysis of dataset A. For this reason, we chose to collect new data on a 65bp oligo, without the presence of DNA blocks (Fig 3C).

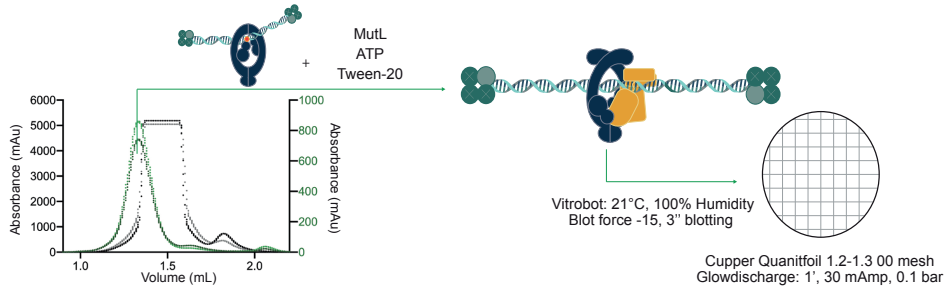
Grid preparation

Because the overall grids and data looked good, nothing was changed in the protocol of grid preparation, except the replacement of blocked 100GT50 by unblocked 65GT33 (Fig 4).

Data collection and processing

Data was collected and processed (Fig 4). 2D Classification of the picked particles showed indeed more orientations of MutS, likely due to the shorter oligo that was used. More front and side views were obtained instead of only top views (Fig 4C).

A



B

	Dataset A-1	Dataset A-2
Microscope	Titan Krios	Titan Krios
Laboratorium	Laboratory of Molecular Biology (LMB)	Netherlands Center for Nanoscopy (NeCEN)
Camera	K2	K2
Å/px	1.76	1.76
Voltage (kV)	300	300
Exposure time (s)	*25*	16
Number of frames	20	20
Total exposure (e⁻/Å²)	40	40
Defocus range (μm)	N.A.	1.4-3.0

C

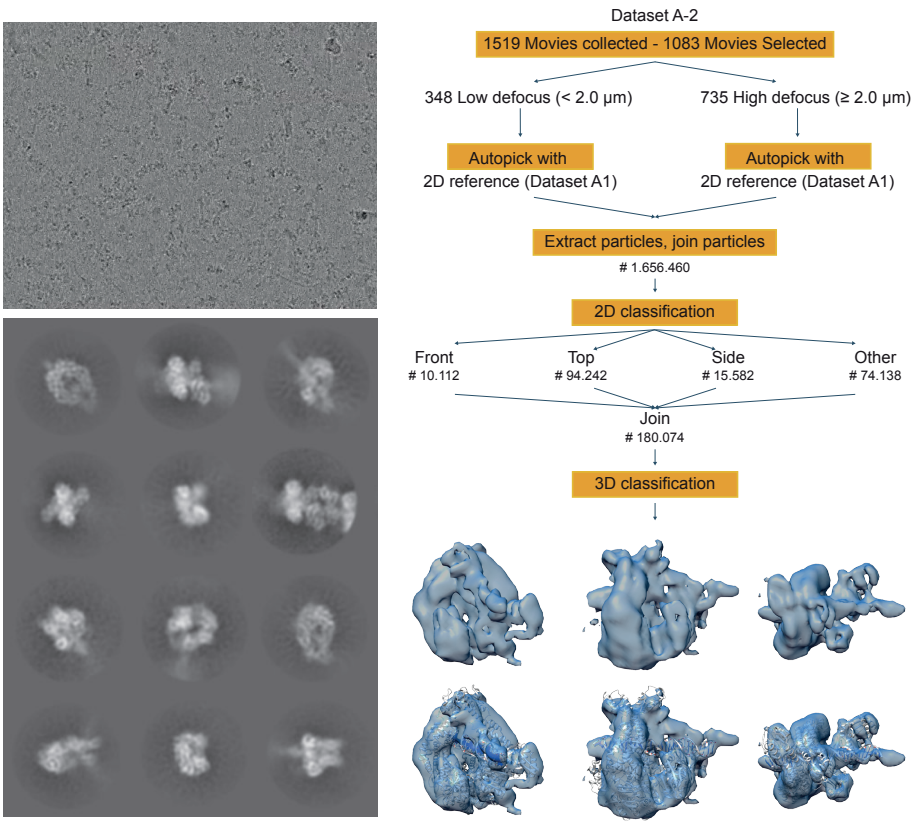


Figure 2. Data collection and processing of Dataset A. A) Grid preparation of MutS-MutL on 100GT50. 100GT50-Streptavidin (black curve) was purified and peak fractions were collected. MutS-^{D835R} was mixed with 100GT50 and purified by analytical gel filtration (green curve). Peak fraction was collected and mixed with a mixture of MutL, ATP and Tween-20. For concentrations see Materials and Methods. Grids were glow discharged and frozen as indicated in the figure. **B)** Details of data collection. **C)** For details see Materials and Methods. Top left is a micrograph of the NeCEN dataset and bottom left examples of 2D classes.

However, 2D classes of MutS-MutL together were very rare compared to dataset A. The removal of the block in combination with the presence of 1 mM ATP likely resulted in sliding off of the complex or individual components.

Because a stable complex could not be obtained without blocks, and the presence of blocks complicates the processing, we decided to crosslink the sample. By crosslinking MutS and MutL, we can force the proteins to be in close proximity on DNA.

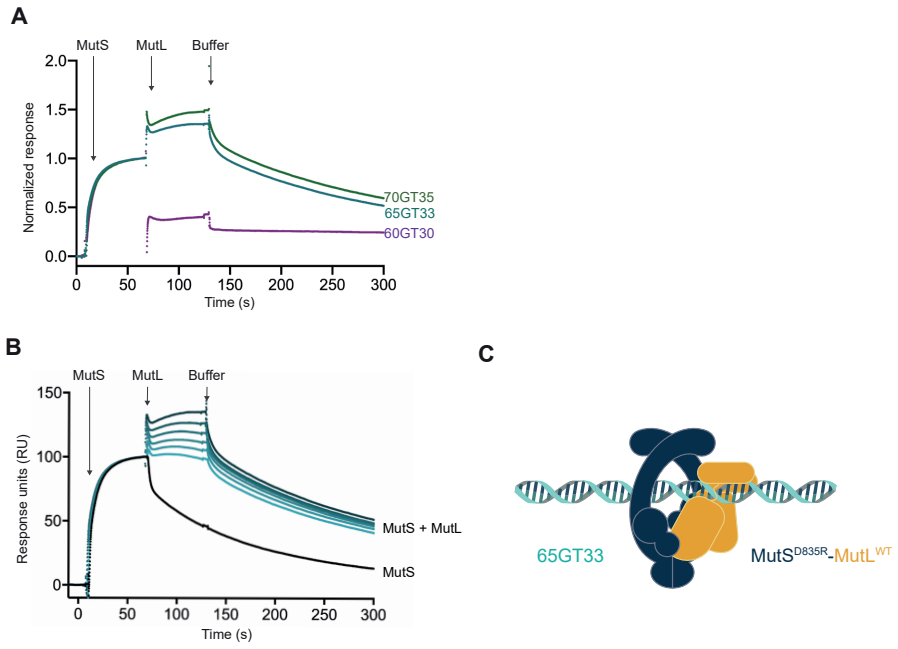
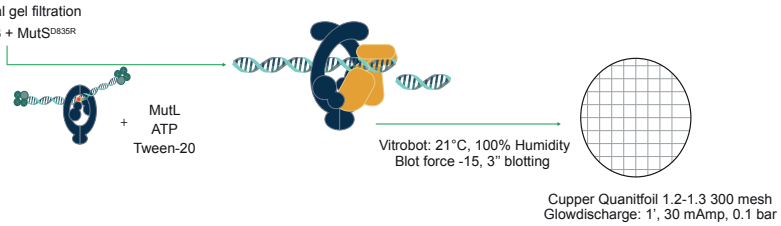


Figure 3. Optimization complex stability MutS-MutL on DNA for dataset B. **A)** Normalized SPR curves from individual experiments on 3 blocked oligo's; 60GT30 (purple), 65GT33 (turqoise) and 70GT35 (green). First injection, 60 s 200 nM MutS, followed by 60 s 1600 nM MutL and finally buffer injection for dissociation of the complex. **B)** SPR binding curves of MutS and MutL on blocked 65GT33. First injection MutS (60s, 200 nM) followed by a second injection of 0-1600 nM MutL for 60s. **C)** Schematic set-up of the complex; 65GT33 with MutS^{D835R} and MutL^{WT}.

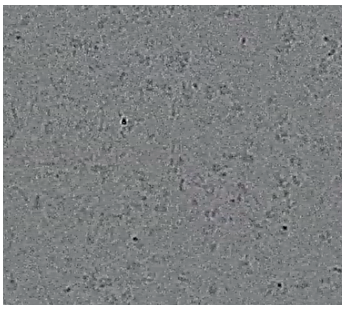
A

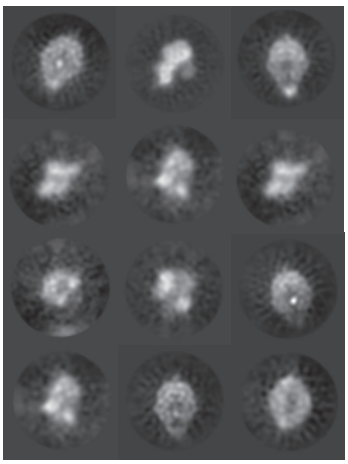


B

	Dataset B
Microscope	Titan Krios
Laboratorium	Netherlands Center for Nanoscopy (NeCEN)
Camera	K2
Å/px	1.086
Voltage (kV)	300
Exposure time (s)	6.3
Number of frames	21
Total exposure (e$^-$/Å^2)	40
Defocus range (μm)	1.8-3.0

C





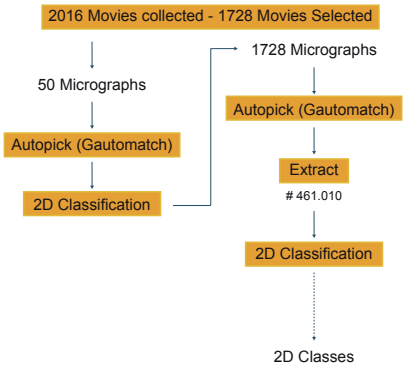


Figure 4. Data collection and processing of Dataset B. **A)** Grid preparation of MutS-MutL on 65GT33. DNA with MutS^{D835R} was mixed and purified. Peak fraction was used and mixed with a mixture of MutL, ATP and Tween-20. For concentrations see Materials and Methods. Grids were glow discharged and frozen as indicated in the figure. **B)** Details of data collection. **C)** For details see Materials and Methods. Top left is a micrograph and bottom left examples of 2D classes.

Dataset C

Optimization of complex stability and assembly of MutS-MutL on DNA

In order to trap the transient MutS-MutL complex more efficiently on DNA, we crosslinked full-length single-cysteine MutS D246C D835R with full-length single-cysteine MutL N131C (MutSxMutL) [19]. This same crosslinking strategy was shown to be successful for MutS-MutL^{LN40} [8] (Chapter 3). For structure determination by X-ray crystallography at that time, MutL was truncated in order to obtain sufficient amounts of a stable and homogeneous complex. However, cryo-EM analysis is more suitable to study flexible and heterogeneous complexes and only needs limited material.

Crosslinking of MutS and MutL was based on the protocol of crosslinking MutS and MutL^{LN40} (see Materials and methods). The crosslinking protocol and its results are shown in Figure 5. Binding of MutSxMutL on DNA was tested by SPR on various oligo lengths and compared to a mixture of MutS and MutL without crosslinker. MutSxMutL reaches the binding equilibrium slower than MutS+MutL, but more important, the dissociation phase is far slower for MutSxMutL than MutS+MutL. This indicated that MutSxMutL forms a stable complex on open DNA in the presence of ATP and could be studied with cryo-EM (Fig 5).

Based on the results in chapter 3, we know that when MutL interacts with MutS via one binding interface of MutS, full MMR initiation can still be initiated. In addition, we showed that, regardless of MutS being double crosslinked to MutL^{LN40}, only one MutL^{LN40} moiety interacts with MutS and DNA. Based on these results, we aimed to saturate the crosslinking and therefore crosslink all MutS monomers with one MutL monomer, and the presence of free MutL to form MutL dimers. As seen on the SDS-page gel from the final gel filtration column, free MutL is present to complement MutSxMutL (Fig 5A). To see whether our purified MutSxMutL sample is saturated with MutL, we tested if the sample could bind more MutL, using SPR. MutL injections were not showing an increase in binding, indicating that there is enough free MutL present to form MutL dimers with MutSxMutL (Fig 5C).

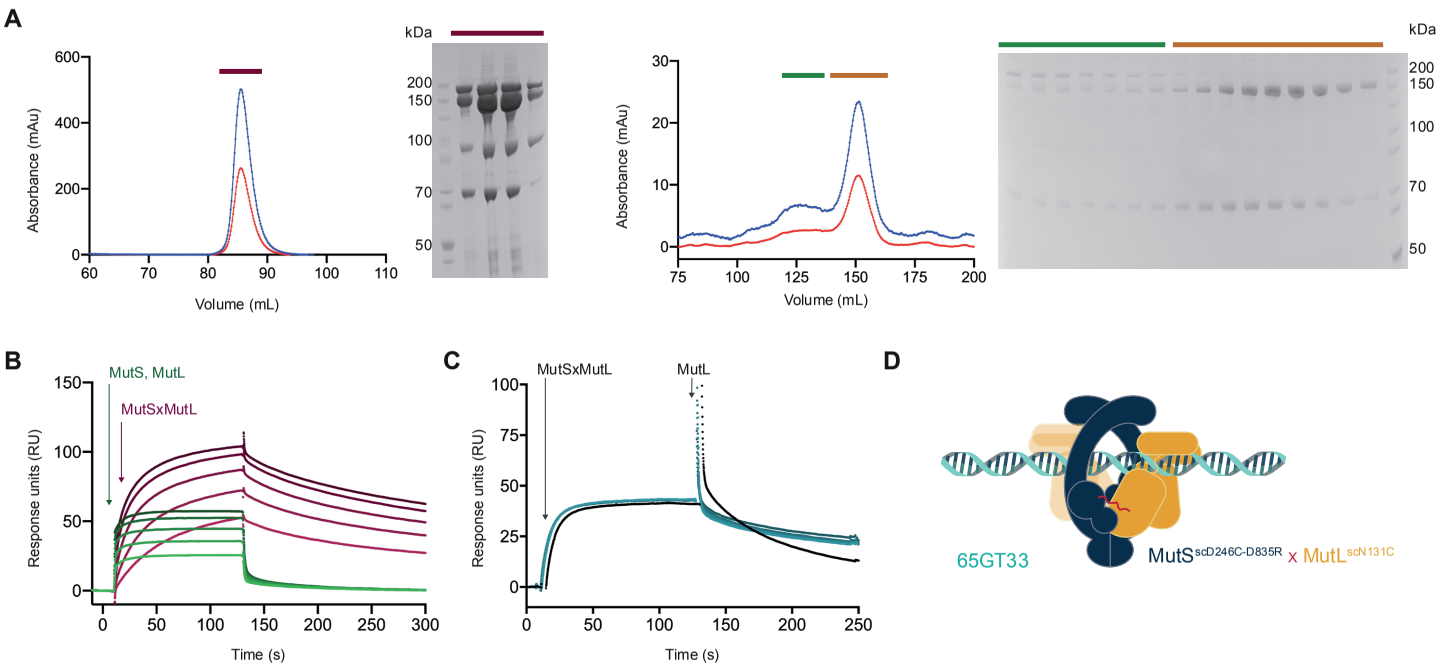
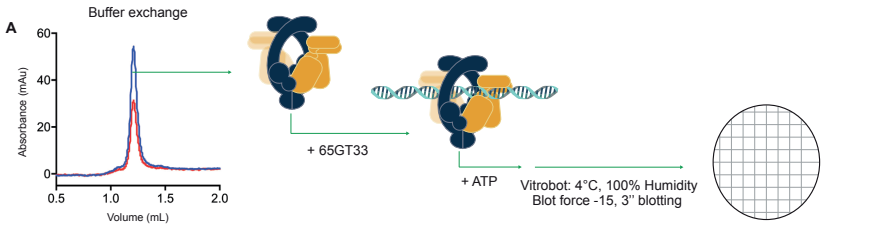


Figure 5. Optimization complex stability MutS-MutL on DNA for dataset C. A) Purification of crosslinked MutSxMutL (See materials and methods for details). Left chromatogram and SDS-page gel are results of purified sample over a heparin column. The main four bands visbile (from top to bottom) are corresponding to MutSxMutS (190 kDa), MutSxMutL (162 kDa), MutS (95 kDa) and MutL (67 kDa). Single MutS and single MutL are present due to dimerization in solution with a crosslinked product. The pooled heparin fractions are purifed over a tandem S200 16/60 gel filtration column and results are shown on the right chomatogram and corresponding SDS-page gel. The first small peak (green bar) corresponds to MutSxMutS and free MutS and MutL, while the second peak (orange) corresponds to MutSxMutL and free MutL. These fractions were used for cryo-EM analysis and SPR experiments. **B)** SPR curves of MutSxMutL (red) and MutS + MutL (mixed) (green) on open 65GT33. Injection for 120 s (0-1024 nM). **C)** SPR experiment to test the abbility of MutSxMutL to bind extra free MutL. First injection 200 nM MutSxMutL for 120 s, followed by 200 - 800 nM MutL (turqoise) or buffer (black). **D)** Schematic set-up of the complex; 65GT33 with MutS$^{scD246C-D835R}$ crosslinked to MutL^{scN131C}.



B

	Dataset C
Microscope	Talos Artica
Laboratorium	Centro Nacional de Biotecnología
Camera	F3
Å/px	0.855
Voltage (kV)	300
Exposure time (s)	35
Number of frames	70
Total exposure (e⁻/Å²)	35
Defocus range (μm)	2.0-3.6

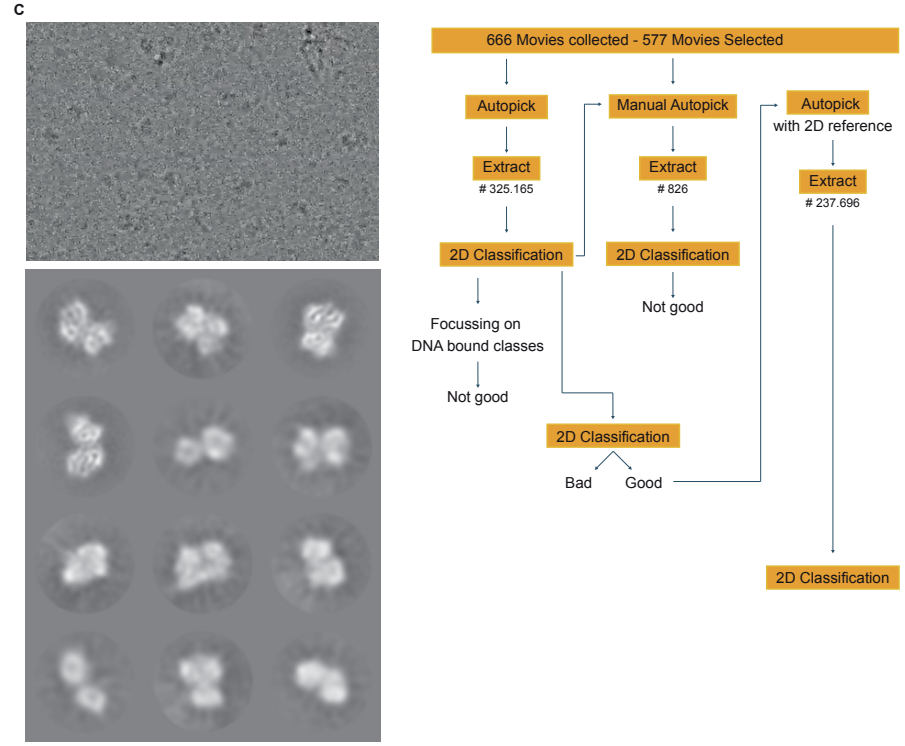


Figure 6. Data collection and processing of Dataset C. A) Grid preparation of MutSxMutL on 65GT33. MutSxMutL was purified over analytical gel filtration column for a buffer exchange. After purification, DNA was added immediatly. This mixture was mixed with ATP and frozen afterwards. For concentrations see Materials and Methods. Grids were glow discharged and frozen as indicated in the figure. **B)** Details of data collection. **C)** For details see Materials and Methods. Top left is a micrograph and bottom left examples of 2D classes.

Grid preparation

MutSxMutL is purified in a buffer containing 10% glycerol, which is not recommended for cryo-EM analysis. Therefore, a buffer exchange was necessary prior to grid freezing. To optimize the grid quality and stability of the complex, we tried several ways to assemble the complex prior to grid freezing. The best result was obtained when MutSxMutL was analyzed over an analytical gel-filtration without the addition of DNA, this in contrast with dataset A and B. After buffer exchange, DNA was added to MutSxMutL. Finally, this mixture is mixed with ATP, loaded on the grid and plunge frozen (Fig 6).

Data collection and processing

Details regarding data collection and processing are shown in Fig 6. First rounds of 2D classification showed unexpected 2D classes. After several rounds of 2D classification and subset selections, MutS-like classes were still absent. Even manual particle picking of particles that had MutS-views did not lead to any MutS-like 2D classes. Nevertheless, the obtained classes had a good overall quality and looked like MutL 2D classes. Because of this, this attempt was stopped and we continued focusing on MutL on DNA.

Dataset D - MutL

Complex optimization

Complex stability was not optimized since it is known that MutL can bind DNA *in vitro* in the absence of MutS and a mismatch, in the presence of low salt [17, 23]. Therefore, the binding buffer was adjusted (25 mM Hepes 7.5, 50 mM KCl, 5 mM $MgCl_2$). Instead of mismatched DNA, matched DNA was used. We choose not to use DNA blocks due to the chance of obtaining preferred orientations as was seen in dataset A1 and A2.

Grid preparation

From personal experience we know that MutL behaves difficult on an analytical gel filtration column when in it is in a diluted concentration. Therefore, MutL was purified in the absence of glycerol and grids were made without the use of a purification step using analytical gel filtration column. Instead, MutL, DNA and ATP were mixed and grids were prepared afterwards (Fig 7A).

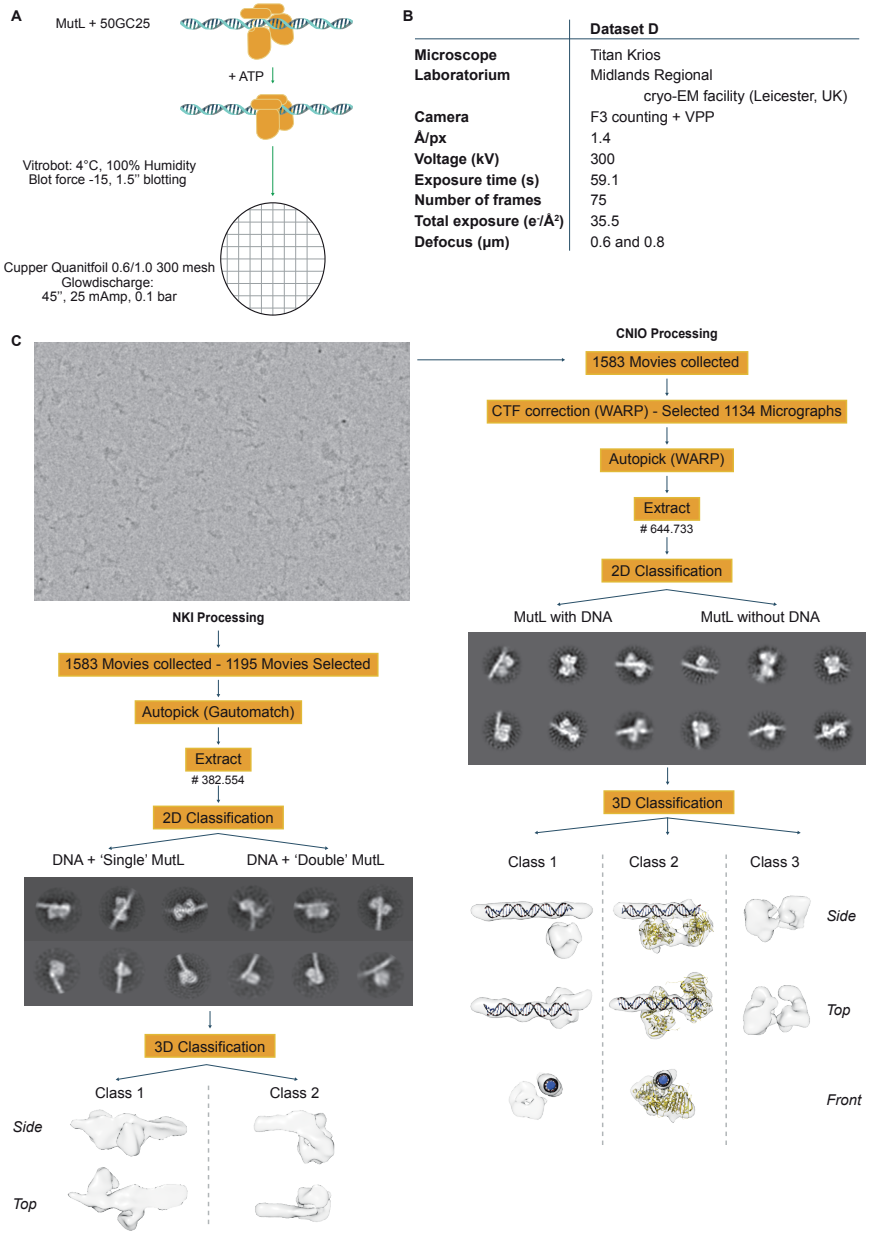


Figure 7. Data collection and processing of Dataset D. **A)** Grid preparation of MutL on DNA. MutL was mixed with DNA (50GC25). Prior to grid freezing, ATP was added. Grids were glow discharged and frozen as indicated. **B)** Details of data collection. **C)** Processing overview. For details see Materials and Methods. Left column is first processing attempt at NKI. Right column is second attempt at CNIO using WARP and Cryosparc. Several 3D classes were obtained but the resolution was not high enough to fit existing crystal stuctures of MutL with 100% succes. Class 2 (right column) nicely fits a piece of DNA and two MutL^{LN40} domains. The fitting is a hypothesis and therefore needs to be interpretated with caution.

Data Collection and processing

Grids were collected on a Titan Krios electron microscope with a K2 detector and a Volta Phase plate (VPP) [24, 25]. Details regarding data collection and processing are shown in Fig 7B. Data was first processed at the NKI. When processing was not successful, a different approach was tried at the CNIO.

First attempt in processing (Fig 7C, left column) was to use Gautomatch for picking particles and to select on three groups during 2D classification; MutL + DNA, MutL alone and small (undefined) classes. This was followed by merging all good classes and then select on two subtypes; DNA with a small density ('single' MutL) and DNA with two small densities ('double' MutL). The representative classes are shown in Figure 7C, left column. Then, we ran 3D classification for these particles. The 3D classes showed similar views/orientations as was observed in 2D classification but MutL could not be fitted properly in the density. Class 1 looks like two domains interacting with DNA, similar to the top row of 2D classes. Unfortunately, both the NTD and CTD don't give a proper fit in the density in both class 1 and 2. Orientations and views were missing and the resolution was not high enough for placing the MutL domains in the density.

Processing at the CNIO was done using software WARP [30], Cryosparc [31] and Relion (Fig 7C, right column). Picked particles by the machine learning software WARP looked more promising than picked particles by Gautomatch. 2D classification was done in Relion and Cryosparc and a final selection was made and used for 3D classification. The best three classes are shown (Fig 7C, right column). Class 1(23.163 particles) shows similarities with class 2 of the NKI processing, however there are less DNA interactions between the DNA density and the extra MutL density. Again, the resolution does not allow a proper fit of MutL NTD and CTD. Class 2 (27.056 particles) shows the best fit with DNA and with two domains of MutL NTD. This class is comparable as class 1 of the NKI processing and also corresponds to the obtained 2D classes. Finally, class 3 (16.933 particles) is the only class obtained without DNA given the absence of a long and elongated density. However, also here, it is not possible to fit MutL NTD or CTD alone in the density. Possibilities are two NTDs or two CTDs or a full MutL monomer with a NTD and CTD. Although the data looks promising, no high-resolution structure could be

obtained, and more work needs to be done to obtain a full length structure of MutL on DNA.

Class 2 in Figure 7C showed the low-resolution model with the fit of two $MutL^{LN40}$ domains bound to a straight DNA molecule. The crystal structures of MutL bound to an ATP analog (PDB ID: 1NHJ) were used to fit in the electron density. The resolution is not high enough to build MutL, so the conformation of the crystal structure was placed in the density without further changes in the conformation. Therefore, the fit must be interpreted with caution. In the cryo-EM map of MutS-$MutL^{LN40}$, MutL interacts with DNA. The stretch of residues 162-170 are in close proximity with DNA, but also residues R22, R48, R176, R264, R266 and K194 (Chapter 3). In the crystal structure of MutS-$MutL^{LN40}$, mutation of residues R266, R162 and R316 together resulted in a decrease in DNA binding, [8]. Recently, a paper was published where they probed the binding center of *E.coli* MutL to DNA, using FRET [37]. They showed that residues T218 and A251 are in close proximity to DNA in the MutL-DNA complex. All these residues from different groups and experiments are shown in our fit of DNA with two $MutL^{LN40}$ domains (Fig 8). Based on this model, residues R162 and K165 are closest to the DNA backbone. Residues T218 and A251 are close, but in this conformation, they cannot form any interactions with DNA. Rearrangements in $MutL^{LN40}$ could bring these residues in closer proximity. Residue R316 (not shown) is not near the DNA backbone, as was the case in the cryo-EM map of MutS-$MutL^{LN40}$.

4. DISCUSSION

MutS and MutL are the first two proteins in the mismatch repair process. When MutS finds the mismatch, it recruits MutL in an ATP-dependent manner. The interaction of MutS with MutL on DNA has been of interest for many years, since the precise mechanism of MutL recruitment is not fully understood. With our work, we tried to solve the structure of full length MutS-MutL bound do DNA. This work will be discussed and future plans are proposed.

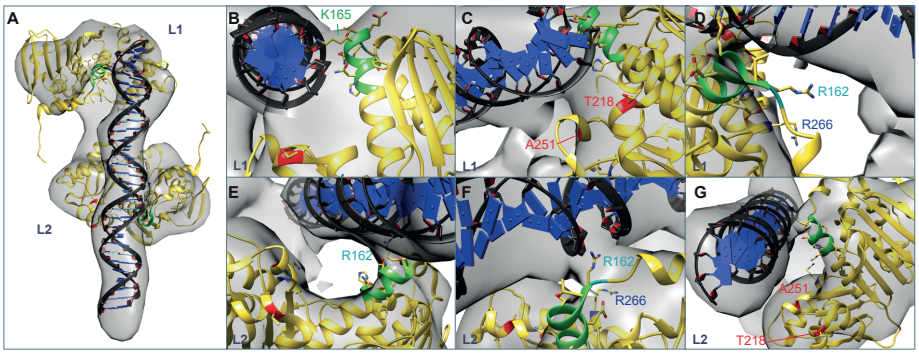


Figure 8. Fit of two MutL^{LN40} X-ray strucures in cryo-EM map Dataset D. A) Top view of two MutL^{LN40} molecules and DNA fitted in map (not shown, see Fig7C). Top MutL is named L1, bottom MutL is named L2. **B-D)** Zoom in of residues. Residues close enough for interaction are K165 (green) and R162 (Cyan). Residues that are not close enough are R266, T218 and A251 to interact with DNA. **E-G)** Zoom in of residues. R162 (cyan) is close enough to DNA to interact. Residues R266, T218 and A251 are not cleose enough for interaction with DNA (see main text for discussion).

Cryo-EM datasets A, B, and C

Three different datasets for MutS-MutL have shown different results. MutS-MutL on blocked 100GT50 showed preferred orientations, on open 65GT33 we lost complex formation and crosslinked MutSxMutL on 65GT33 trapped mostly MutL instead of MutS-MutL. We can conclude that trapping this transient complex by cryo-EM needs to be studied in more detail. A relatively short oligo will work best. It needs to be long enough to have the full complex bound but as short as possible to avoid preferred orientations. MutS and MutL are suitable for cryo-EM analysis given the size and classes that were obtained. However, the stability of MutS-MutL on DNA needs to be further optimized.

Complex optimization by SPR has shown that MutS and MutL have a long lifetime on closed DNA or when they are crosslinked (Fig 1E, Fig 5B purple curves). Since non-crosslinked MutS-MutL on open DNA shows fast dissociation (Fig 5B green curves), we can assume that crosslinked MutS-MutL shows binding of both proteins. Otherwise, the co-injection of MutS and MutL should show binding of at least one of the two proteins. Therefore, it is rather odd that cryo-EM analysis only shows particles of MutL and not MutS. This was not expected after the SPR

experiment with MutSxMutL, given the fact that the conditions are the similar for SPR and cryo-EM (buffer, ATP concentration and proteins). For this reason, it is possible that the complex is stable, but difficult to trap with cryo-EM.

Is the complex of MutS-MutL too transient?

Despite the differences between datasets and the questions that are rising here, there is one main question; Why did we not manage to trap full length MutS-MutL on DNA? Together with us, various groups have studied this complex using several techniques and different model organisms. Single molecule studies show that the lifetime of MutS-MutL on long DNA in the presence of ATP, is roughly 43 ± 3 seconds [17]. This should be long enough since there are only a few seconds between complex assembly and grid freezing, although we use short DNA. This publication also showed that after mismatch binding, MutS and MutL (in the presence of ATP) can follow different scenarios. In one of these, MutL dissociates from MutS and forms stable complexes on DNA that can diffuse with a high speed and bypass MutS clamps. It was also shown by other groups that MutL can pass Cas9-roadblocks [17, 38]. In the context of our data, this suggests that the conditions we used in dataset C, were maybe pushing MutL to form stable MutL clamps and dissociating from MutS. Also, since MutL can pass roadblocks and MutS, it could theoretically bypass a streptavidin block. Streptavidin is roughly 60 Å width while the smallest side of Cas9 is 87 Å. This can be a reason that we did not see single MutL clamps on 100GT50 in dataset A.

There is a wide selection of published results that suggest that the complex of MutS-MutL exist [7, 8, 12, 14, 39-42]. However, it could be that MutS and MutL interact transiently to load MutL on DNA, and then are in close proximity for a long time, rather than forming a stable complex. Data supporting this option, is the cryo-EM structure of MutS-MutL^{LN40} on DNA (Chapter 3). MutL^{LN40} interacts with the MutS dimer and DNA, but would not allow full dimerization with a second MutL^{LN40}. This could suggest that, rearrangements of MutL are needed after loading by MutS, to dimerize with the second MutL^{LN40} domain and these could involve a release of the MutS interaction.

Finally, there is a possibility that the complex of MutS-MutL on DNA is formed, but not detectable or destroyed by the process of freezing for cryo-EM. During grid preparation, proteins are spread over a glow-discharged grid and this is rapidly

plunged into liquid ethane whereby the thin liquid layer converts into vitreous ice. The created air-water interface has the potential to absorb proteins and to damage them, or, disrupt the complex formation [43]. Possibly this happened for the MutS-MutL analysis. Nevertheless, the MutS-MutL^{LN40} sample did not face any detectable problems but it is possible that full length MutL behaves differently.

Recently, a study on human MMR showed that MutSα and MutLα assemble into multimeric complexes comprising three to eight proteins around a DNA mismatch [44]. This is in contrast with the assumptions of the existence of a single MutS-MutL complex. In our cryo-EM data we actually see both situations. Dominantly, we obtain single classes of MutS-MutL together (Dataset A). However, we also observed 2D classes where multiple MutS and MutL molecules were aligned on the DNA oligo (Fig 2C, second row, right class), indicating the same possibilities for the prokaryotic system. Nevertheless, these classes were a minority in the pool of particles and classes, although we do not know whether this is oligomeric species, that the freezing conditions disturb these selectively or that the DNA molecule is too short.

How to continue?

For future continuation of the project, there are some possibilities that could be considered. A common problem from all datasets was that MutS and MutL are dynamic proteins and oligos were maybe not the most optimal substrate to bind to. DNA blocks to the long oligo induced preferred orientations and the absence of the block made complex formation challenging. As an alternative DNA substrate, a minicircle or plasmid could be used. A plasmid was also used when studying the scanning state of MutS (Chapter 3). One thing to keep in mind is that a plasmid is far bigger than a single oligo. Thus, the plasmid will spread out over the grid in such a way, that you likely obtain preferred orientations. This could be overcome by collecting data at a tilted angle to obtain good data for full 3D reconstruction. As an alternative to a plasmid, nanocircles could be made [45]. One thing to take into account is that the circle should not be too small. A small circle will have tension among the DNA bases. This may induce other conformations given the fact that MutS tries to induce a bend in the DNA while searching for mismatches. However, the use of a plasmid or circle could answer the question whether MutS and MutL form oligomeric classes as was observed for human MutSα and MutLα [44].

Another option would be to crosslink MutS to DNA. Options to achieve this have been shown previously [46]. The position of the cysteine should be considered carefully to ensure that MutS can undergo all the conformational changes necessary towards the MutL-recruiting sliding clamp conformation. We hypothesize that a single cysteine at position 420 could be used for such crosslinking, given the fact that R420 is in close proximity to the DNA in all states from scanning towards MutL-bound sliding clamp. In this way, crosslinked $MutS^{R420C}$ to DNA could be purified and MutL and ATP could be added prior to grid freezing to trap MutS-MutL on DNA. Single cysteine MutS R420C was purified for this purpose but also used for the validation experiments in Chapter 3.

In line with site-specific crosslinking, GraFix or AgarFix can also be tried [33, 47, 48]. This method can be used to stabilize macromolecular complexes. Macromolecules undergo a weak intramolecular crosslinking while being purified by density gradient ultracentrifugation. After a buffer exchange, the sample can be used for cryo-EM. This buffer exchange can lead to dilution of the sample or sample loss in case the stability is poor after the exchange. An alternative method is Agarfix. Gentle crosslinking is accomplished by diffusion of the crosslinker into the agarose drop containing your sample. The sample will be recovered by diffusion or electroelution and can be used directly for cryo-EM grid preparation.

A last option is to use a mutant of MutL where the linker is truncated. As discussed, MutL can bypass objects on DNA. When the linker of the MutL monomer is truncated, this ability is reduced or absent, depending on the length of the linker [38]. Maybe this prevents MutL from forming an independent stable sliding clamp or from falling of the DNA.

Cryo-EM Dataset D and how to continue?

Processing dataset D showed promising 2D and 3D classes of MutL, both alone and on DNA. The fit of two $MutL^{LN40}$ molecules on DNA (Fig 7, Fig 8), indicates that the C-terminal domains of MutL are flexible and not forming a stable interaction with DNA. Some data shows limited interaction between the CTD of MutL with DNA [37], while other data shows that mutating several positively charged residues affects DNA binding [49]. It is known that the CTD of MutL interacts with the processivity clamp [50, 51]. This could explain the limited interaction of MutL with DNA that was seen in mutational studies. However, high concentration of ATP (1 mM) also

results in compaction of MutL. This was observed both in the absence and presence of DNA [34, 49]. This would mean, that both the NTD and CTD of MutL should be visible in our cryo-EM map. Since we did not observe this, it is still the question what the best conditions are to trap full length MutL on DNA.

For this reason, more optimization is needed to trap full length MutL on DNA in a compact conformation. Little optimization was done prior to grid preparation and data collection, since it is known that MutL binds homoduplex DNA at a low salt concentration. Given the fact that we obtained multiple conformations of MutL in Dataset D, more optimization could be done to trap one state and therefore obtain a homogenous set of particles. A range of ATP concentrations could be tested by both SPR and cryo-EM, to see if this can push the sample to a homogenous state. In addition, other nucleotides, including AMP-PNP could be tested which could maybe trap MutL on DNA in a better, stable manner. Alternatively, one can test truncated constructs of MutL with various linker lengths to see if this improves the stability of MutL on DNA. Finally, a different DNA substrate can be used, as was suggested for MutS-MutL.

The dataset of MutL [52] was of high quality but was collected at a high magnification. Also, the particles showed several conformations in 3D classifications. Therefore, a relatively small number of particles was left in each class. If a more homogenous sample can be frozen, and more micrographs collected, the resolution of the 3D cryo-EM maps of MutL on DNA may improve and allow full structure solution.

In conclusion, this work shows initial cryo-EM studies of *E.coli* MutS-MutL and MutL on DNA. Several attempts and methods have shown that the size of MutS-MutL is feasible for cryo-EM analysis. However, since the complex is transient, we have not been able to successfully solve the structures at high resolution.

REFERENCES

1. Kunkel, T.A. and D.A. Erie, *DNA mismatch repair.* Annu Rev Biochem, 2005. **74**: p. 681-710.
2. Lynch, H.T. and A. de la Chapelle, *Genetic susceptibility to non-polyposis colorectal cancer.* J Med Genet, 1999. **36**(11): p. 801-18.
3. Cerretelli, G., et al., *Molecular pathology of Lynch syndrome.* J Pathol, 2020. **250**(5): p. 518-531.
4. Lahue, R.S., K.G. Au, and P. Modrich, *DNA mismatch correction in a defined system.* Science, 1989. **245**(4914): p. 160-4.
5. Lamers, M.H., et al., *The crystal structure of DNA mismatch repair protein MutS binding to a G x T mismatch.* Nature, 2000. **407**(6805): p. 711-7.
6. Malkov, V.A., et al., *Photocross-linking of the NH2-terminal region of Taq MutS protein to the major groove of a heteroduplex DNA.* J Biol Chem, 1997. **272**(38): p. 23811-7.
7. Gradia, S., et al., *hMSH2-hMSH6 forms a hydrolysis-independent sliding clamp on mismatched DNA.* Mol Cell, 1999. **3**(2): p. 255-61.
8. Groothuizen, F.S., et al., *MutS/MutL crystal structure reveals that the MutS sliding clamp loads MutL onto DNA.* Elife, 2015. **4**: p. e06744.
9. Bowers, J., et al., *MSH-MLH complexes formed at a DNA mismatch are disrupted by the PCNA sliding clamp.* J Mol Biol, 2001. **306**(5): p. 957-68.
10. Jeong, C., et al., *MutS switches between two fundamentally distinct clamps during mismatch repair.* Nat Struct Mol Biol, 2011. **18**(3): p. 379-85.
11. Pillon, M.C., et al., *The sliding clamp tethers the endonuclease domain of MutL to DNA.* Nucleic Acids Res, 2015.
12. Qiu, R., et al., *MutL traps MutS at a DNA mismatch.* Proc Natl Acad Sci U S A, 2015. **112**(35): p. 10914-9.
13. Bhairosing-Kok, D., et al., *Sharp kinking of a coiled-coil in MutS allows DNA binding and release.* Nucleic Acids Res, 2019. **47**(16): p. 8888-8898.
14. LeBlanc, S., et al., *Using Atomic Force Microscopy to Characterize the Conformational Properties of Proteins and Protein-DNA Complexes That Carry Out DNA Repair.* Methods Enzymol, 2017. **592**: p. 187-212.
15. Sacho, E.J., et al., *Direct visualization of asymmetric adenine-nucleotide-induced conformational changes in MutL alpha.* Mol Cell, 2008. **29**(1): p. 112-21.
16. Nirwal, S., et al., *Mechanism of formation of a toroid around DNA by the mismatch sensor protein.* Nucleic Acids Res, 2018. **46**(1): p. 256-266.
17. Liu, J., et al., *Cascading MutS and MutL sliding clamps control DNA diffusion to activate mismatch repair.* Nature, 2016. **539**(7630): p. 583-587.
18. Hanne, J., et al., *MutS homolog sliding clamps shield the DNA from binding proteins.* J Biol Chem, 2018. **293**(37): p. 14285-14294.

19. Winkler, I., et al., *Chemical trapping of the dynamic MutS-MutL complex formed in DNA mismatch repair in Escherichia coli.* J Biol Chem, 2011. **286**(19): p. 17326-37.

20. Ban, C. and W. Yang, *Crystal structure and ATPase activity of MutL: implications for DNA repair and mutagenesis.* Cell, 1998. **95**(4): p. 541-52.

21. Giron-Monzon, L., et al., *Mapping protein-protein interactions between MutL and MutH by cross-linking.* J Biol Chem, 2004. **279**(47): p. 49338-45.

22. Fairhead, M., et al., *Plug-and-play pairing via defined divalent streptavidins.* J Mol Biol, 2014. **426**(1): p. 199-214.

23. Acharya, S., et al., *The coordinated functions of the E. coli MutS and MutL proteins in mismatch repair.* Mol Cell, 2003. **12**(1): p. 233-46.

24. Danev, R., et al., *Volta potential phase plate for in-focus phase contrast transmission electron microscopy.* Proc Natl Acad Sci U S A, 2014. **111**(44): p. 15635-40.

25. Danev, R. and W. Baumeister, *Cryo-EM single particle analysis with the Volta phase plate.* Elife, 2016. **5**.

26. Zivanov, J., et al., *New tools for automated high-resolution cryo-EM structure determination in RELION-3.* Elife, 2018. **7**.

27. Li, X., et al., *Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM.* Nat Methods, 2013. **10**(6): p. 584-90.

28. Zhang, K., *Gctf: Real-time CTF determination and correction.* J Struct Biol, 2016. **193**(1): p. 1-12.

29. Henderson, R., *Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise.* Proc Natl Acad Sci U S A, 2013. **110**(45): p. 18037-41.

30. Tegunov, D. and P. Cramer, *Real-time cryo-electron microscopy data preprocessing with Warp.* Nat Methods, 2019. **16**(11): p. 1146-1152.

31. Punjani, A., et al., *cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination.* Nat Methods, 2017. **14**(3): p. 290-296.

32. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis.* J Comput Chem, 2004. **25**(13): p. 1605-12.

33. Stark, H., *GraFix: stabilization of fragile macromolecular complexes for single particle cryo-EM.* Methods Enzymol, 2010. **481**: p. 109-26.

34. Niedziela-Majka, A., et al., *Self-assembly of Escherichia coli MutL and its complexes with DNA.* Biochemistry, 2011. **50**(37): p. 7868-80.

35. Allen, D.J., et al., *MutS mediates heteroduplex loop formation by a translocation mechanism.* Embo j, 1997. **16**(14): p. 4467-76.

36. Groothuizen, F.S., et al., *Using stable MutS dimers and tetramers to quantitatively analyze DNA mismatch recognition and sliding clamp formation.* Nucleic Acids Res, 2013. **41**(17): p. 8166-81.

37. Monakhova, M., et al., *Probing the DNA-binding center of the MutL protein from the Escherichia coli mismatch repair system via crosslinking and Förster resonance energy transfer.* Biochimie, 2020. **171-172**: p. 43-54.

38. Mardenborough, Y.S.N., et al., *The unstructured linker arms of MutL enable GATC site incision beyond roadblocks during initiation of DNA mismatch repair.* Nucleic Acids Res, 2019. **47**(22): p. 11667-11680.

39. Cho, W.K., et al., *ATP alters the diffusion mechanics of MutS on mismatched DNA.* Structure, 2012. **20**(7): p. 1264-74.

40. Mendillo, M.L., D.J. Mazur, and R.D. Kolodner, *Analysis of the interaction between the Saccharomyces cerevisiae MSH2-MSH6 and MLH1-PMS1 complexes with DNA using a reversible DNA end-blocking system.* J Biol Chem, 2005. **280**(23): p. 22245-57.

41. Qiu, R., et al., *Large conformational changes in MutS during DNA scanning, mismatch recognition and repair signalling.* Embo j, 2012. **31**(11): p. 2528-40.

42. Kunkel, T.A. and D.A. Erie, *Eukaryotic Mismatch Repair in Relation to DNA Replication.* Annu Rev Genet, 2015. **49**: p. 291-313.

43. Noble, A.J., et al., *Reducing effects of particle adsorption to the air-water interface in cryo-EM.* Nat Methods, 2018. **15**(10): p. 793-795.

44. Bradford, K.C., et al., *Dynamic human MutSα-MutLα complexes compact mismatched DNA.* Proc Natl Acad Sci U S A, 2020.

45. Xiao, Y., et al., *Generation of DNA nanocircles containing mismatched bases.* Biotechniques, 2011. **51**(4): p. 259-62, 264-5.

46. Monakhova, M., et al., *Chromatographic isolation of the functionally active MutS protein covalently linked to deoxyribonucleic acid.* J Chromatogr A, 2015. **1389**: p. 19-27.

47. Kastner, B., et al., *GraFix: sample preparation for single-particle electron cryomicroscopy.* Nat Methods, 2008. **5**(1): p. 53-5.

48. Adamus, K., et al., *AgarFix: Simple and accessible stabilization of challenging single-particle cryo-EM specimens through crosslinking in a matrix of agar.* J Struct Biol, 2019. **207**(3): p. 327-331.

49. Guarne, A., et al., *Structure of the MutL C-terminal domain: a model of intact MutL and its roles in mismatch repair.* Embo j, 2004. **23**(21): p. 4134-45.

50. Genschel, J., et al., *Interaction of proliferating cell nuclear antigen with PMS2 is required for MutLα activation and function in mismatch repair.* Proc Natl Acad Sci U S A, 2017. **114**(19): p. 4930-4935.

51. Pillon, M.C., J.H. Miller, and A. Guarne, *The endonuclease domain of MutL interacts with the beta sliding clamp.* DNA Repair (Amst), 2011. **10**(1): p. 87-93.

52. Soufi, B., et al., *Characterization of the E. coli proteome and its modifications during growth and ethanol stress.* Front Microbiol, 2015. **6**: p. 103.

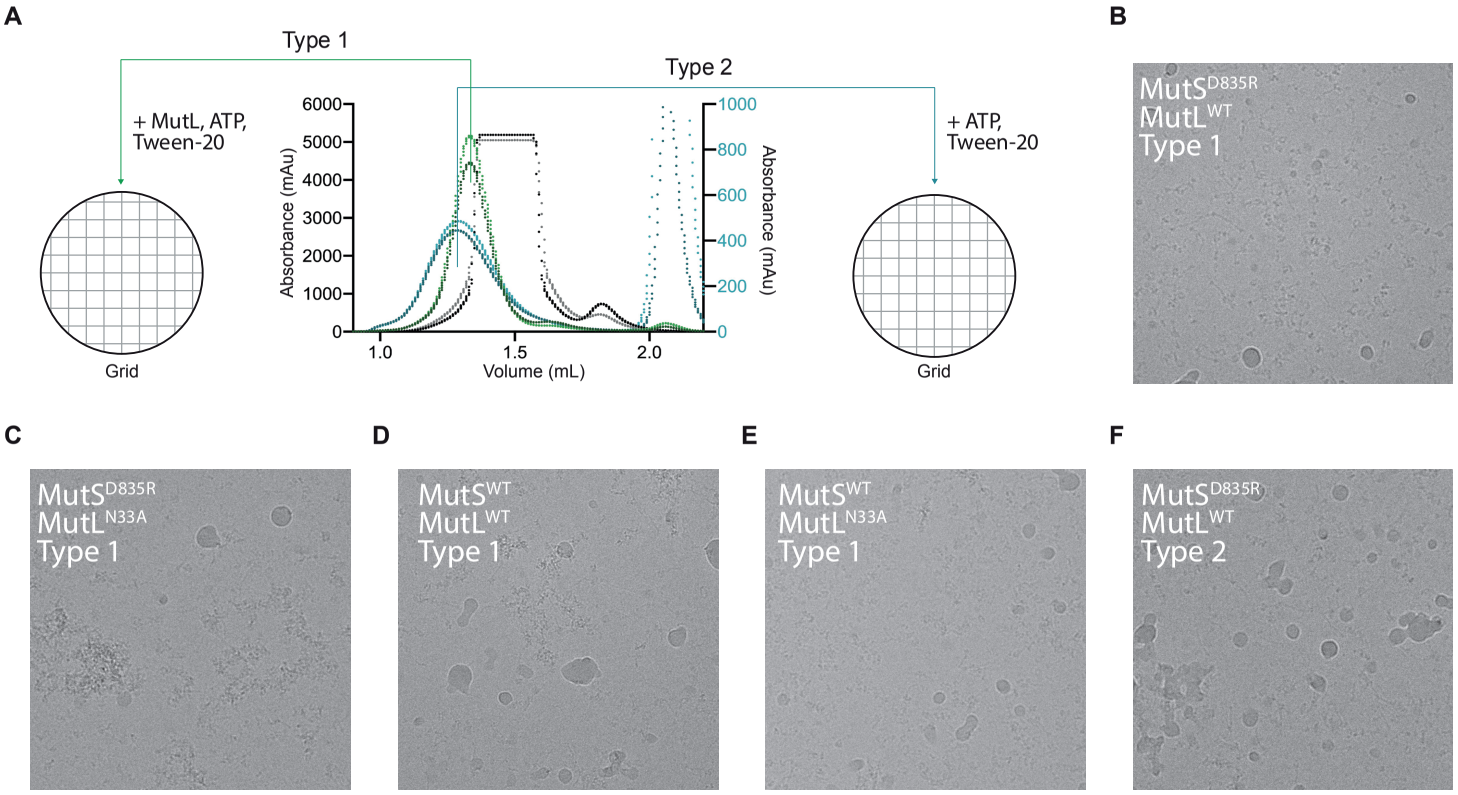
MutL^{N33A} in glycerol buffer

		Well			
		A1-B12	C1-D12	E1-F12	G1-H12
pH	Buffer (100mM)	No Additive	NaCl (150mM)	Glycerol (5%V/V)	NaCl & Glyc
5.0	K Acetate	A1	C1	E1	G1
5.5	MES	A2	C2	E2	G2
6.0	ADA	A3	C3	E3	G3
6.5	MES	A4	C4	E4	G4
6.5	BisTris	A5	C5	E5	G5
6.5	$KHPO_4$	A6	C6	E6	G6
7.0	PIPES	A7	C7	E7	G7
7.0	Imidazole	A8	C8	E8	G8
7.0	ADA	A9	C9	E9	G9
7.0	MOPS	A10	C10	E10	G10
7.5	HEPES	A11	C11	E11	G11
7.5	$KHPO_4$	A12	C12	E12	G12
7.5	DIPSO	B1	D1	F1	H1
7.5	Tricine	B2	D2	F2	H2
8.0	HEPES	B3	D3	F3	H3
8.0	Bicine	B4	D4	F4	H4
8.0	Tris	B5	D5	F5	H5
8.5	Tris	B6	D6	F6	H6
8.5	TABS	B7	D7	F7	H7
8.5	TAPS	B8	D8	F8	H8
9.0	CAPSO	B9	D9	F9	H9
9.0	CHES	B10	D10	F10	H10
9.5	Glycine	B11	D11	F11	H11
10.0	CAPS	B12	D12	F12	H12

MutL^{N33A} in Phosphate buffer

	Well			
	A1-B12	C1-D12	E1-F12	G1-H12
	No Additive	NaCl (150mM)	Glycerol (5%V/V)	NaCl & Glyc
	A1	C1	E1	G1
	A2	C2	E2	G2
	A3	C3	E3	G3
	A4	C4	E4	G4
	A5	C5	E5	G5
	A6	C6	E6	G6
	A7	C7	E7	G7
	A8	C8	E8	G8
	A9	C9	E9	G9
	A10	C10	E10	G10
	A11	C11	E11	G11
	A12	C12	E12	G12
	B1	D1	F1	H1
	B2	D2	F2	H2
	B3	D3	F3	H3
	B4	D4	F4	H4
	B5	D5	F5	H5
	B6	D6	F6	H6
	B7	D7	F7	H7
	B8	D8	F8	H8
	B9	D9	F9	H9
	B10	D10	F10	H10
	B11	D11	F11	H11
	B12	D12	F12	H12

Supplemental Figure 1. Dynamic Light Scattering solubility screen of MutL^{N33A}. Results of solubility screen for MutL^{N33A} in buffer containing glycerol (left) and phosphate buffer (right) (see Materials and Methods). Red boxes indicate precipitation, orange indicates mild precipitation and green coresponds to clear wells.



Supplemental Figure 2. Test datasets MutS-MutL on 100GT50. **A)** Testing two methods to make grids. Type 1 is made by taking the peak fraction of 100GT50 + streptavidin (black curve), mix this with MutS and purify this by gel filtration. Take this peak fraction and mix this with a pre-made mixture of MutL, ATP and Tween-20 prior to grid freezing. Type 2 is made by mixing 100GT50-streptavidin (black curve) with MutS, MutL and ATP. Next, this peak fraction is mixed with a pre-made mixture of ATP and Tween-20, and grids were made after mixing. B) Micrograph of grid of $MutS^{D835R}$ and $MutL^{WT}$, type 1**. C)** Micrograph of grid of $MutS^{D835R}$ and $MutL^{N33A}$, type 1. **D)** Micrograph of grid of $MutS^{WT}$ and $MutL^{WT}$, type 1. **E)** Micrograph of grid of $MutS^{WT}$ and $MutL^{N33A}$, type 1. **F)** Micrograph of grid of $MutS^{D835R}$ and $MutL^{WT}$, type 2.

CHAPTER 5

General discussion

GENERAL DISCUSSION

The conserved MMR mechanism is important for the cell to keep the mutation level low. The MMR initiator, MutS, scans the DNA and finds disruptions in the DNA helix where it signals for repair via downstream proteins, including MutL. Both MutS and MutL undergo a series of events that coincide with conformational changes that can be induced by external factors such as the presence of a DNA mismatch or binding of ATP.

In chapter 2, 3 and 4, we have shown that MutS adopts several conformations during scanning and detecting of mismatches and during downstream signaling. These steps lead to the recruitment of MutL. As MutS, MutL undergoes a series of structural rearrangements too. Our findings regarding these transformations that MutS and MutL undergo during MMR initiation will be discussed in this chapter.

MutS and its ability to kink the lever domains

In chapter 2 we have shown the first 3D structure of the apo-state of *E.coli* MutS. This state of MutS showed flexibility in the lever domains of both monomers and we showed that this flexibility is required for both DNA binding and downstream signaling [1]. The flexibility of the lever domains was also observed in our cryo-EM data (Chapter 3). The data of crosslinked MutS-MutL^{LN40} on DNA, also showed crosslinked kinked MutS-MutL^{LN40}, with one of the two monomers kinked at the defined hinge point as observed in the apo-state of MutS. We tested whether the ability to kink was important, aiming to either prevent kink conformation or the interactions that the MutS clamp domain made with DNA in this conformation. Both disruptions resulted in poor recruitment of MutL on open end DNA (Chapter 3). The kink and its importance were also found for the *E.coli* homolog of cohesin, MukBEF [2]. X-ray crystallography and cryo-EM data showed that MukBEF could alternate between a folded and extended state. A defined disruption in the long coiled-coils arms, a so-called elbow, functions as hinge point. The fact that an ATPase protein can alternate between two conformations raises the possibility that this is coupled to an ATP hydrolysis cycle. For MukBEF and other SMC-proteins there are indications that this is indeed the case [3, 4].

Besides our 3D structure of *E.coli* apo-MutS, two other apo states of homologous MutS proteins are solved. One is full length *N. gonorrhoeae* MutS and the other is *T. aquaticus* (*Taq*) MutS, with the latter one missing the clamp and lever domains [5,

6]. *N. gonorrhoeae* MutS has a more open conformation and shows strong overlap to the scanning conformation of MutS (Chapter 3), with only a different angle for the lever and clamp domains of *E.coli* MutS due to DNA binding. Apo-state of *Taq* MutS shows more similarities with the apo state of *E.coli* MutS with both having a more closed conformation, except the defined kink in the lever domains. The nucleotide state of these three different apo-states is also unique. This raises the possibility that the ATPase cycle is linked to the conformation of MutS in its DNA-free state.

The MutS-MutL complex and the attempts to trap it

The 3D structure of MutS bound to mismatched DNA was already solved for several species [6-10]. Besides this conformation of MutS, where there also many findings confirming the existence of a MutS sliding clamp [11, 12]. Fifteen years after the first MutS structure, the 3D structure of crosslinked MutS-MutL^{LN40} in sliding clamp conformation was solved [13]. Although this structure confirmed many hypothesizes, this structure lacked the DNA molecule that MutS and MutL were bound to. Cryo-EM allows resolving the structure of MutS-MutL^{LN40} bound to DNA. The results are shown and discussed in Chapter 3.

MutL^{LN40} interacts with the bi-partite interaction site on MutS [13]. The bi-partite interaction site is formed by the connector domain of monomer A and the ATPase domain of monomer B. The MutS sliding clamp has two bi-partite interaction sites, which would in theory be occupied by two MutL^{LN40} domains. However, we have shown that MMR initiation *in vitro* is fully functional when only one interaction site is functional. In addition, the cryo-EM map only showed one MutL^{LN40} bound, confirming the hypothesis (Chapter 3). One of our questions is, how this asymmetry is introduced in MutS? Which interaction site will be bound by MutL? In human MutSα, this asymmetry is well defined since MSH6 interacts directly with the mismatch while MSH2 does not. Also, the connector domain is well conserved in MSH2. This could indicate that the connector domain of MSH2 and the ATPase domain of MSH6 forms the relevant bi-partite interaction site for MutLα.

Since *E.coli* MutS is a homodimer, asymmetry is induced rather than defined beforehand. The conformation of the MutS sliding clamp is symmetrical. However, the transition from mismatch bound towards sliding clamp takes place in steps.

Here, asymmetry is present since the intermediate MutS conformation shows two different conformations for each monomer. Also, several cryo-EM maps showed one kinked monomer in both the intermediate conformation as for the sliding clamp (data not shown). Unfortunately, in both the intermediate and sliding clamp cryo-EM maps, one cannot track down which monomer interacted directly with the mismatch and which monomer did not. But there are many indications to believe that these asymmetry states are linked throughout the MMR pathway.

In Chapter 4, we showed the attempts on obtaining a full length cryo-EM structure of MutS and MutL on DNA. We made progress in studying the assembly and stability of the complex, and what the best way could be to trap the proteins together on DNA, suitable for cryo-EM analysis. We have seen that MutS and MutL are very dynamic and trapping them without the presence of a crosslinker or DNA block is challenging. Similar findings were seen by other groups. Single molecule studies have shown that MutS and MutL, after mismatch detection, can divide into several populations of complex composition and dissociation patterns [14]. It is not known what the next step would be in MMR for each of these populations. What we do know, is that MutH will be recruited after MutL is loaded on DNA. There are experiments published that indicated that MutH would be recruited by MutS-MutL, forming a transient trimeric complex [15, 16]. However, other findings also show that this doesn't have to be the default situation and MutH will interact with MutL alone, rather than MutS-MutL [14, 17]. If we look for the answer in our own data, we can hypothesize that indeed MutL releases from MutS before activating MutH (Chapter 4), making a trimeric complex unlikely. In conclusion, the study of the *E.coli* MutS-MutL complex is complicated and ongoing, but it has taken the field a step further in the MMR field by the results we have shown in Chapter 3 and 4.

MutL on DNA

MutL plays many important roles in MMR, and therefore it is called a molecular matchmaker. It interacts with MutS, MutH, UvrD and the β-clamp in *E.coli*, and with MutSα, PCNA and other exonucleases in eukaryotes [15, 18-24]. The function of MutL in *E.coli* is different than for species lacking the endonuclease MutH. For these species, MutL also fulfills this role by nicking the newly synthesized strand after a mismatch is detected. MutL is a diverse and complicated protein to study. The MutL dimer, that is composed of four domains and two linkers, and can therefore adopt

many conformations [25, 26]. It is known that both NTD and CTD can interact with DNA, but the NTDs are forming a tighter complex with DNA [27], and the CTD is involved in interaction with the processivity clamps [19, 28, 29]. The cryo-EM analysis of full length MutL also showed that two NTD molecules are interacting with DNA and the CTD was not visible, was similar to what we observed in the presence of MutS (Chapter 4, dataset B).

But what does this mean regarding the movement of full length MutL on DNA? Unfortunately, that is still a question that does not have a well-defined answer. There are indications that MutL can make a closed system around the DNA by dimerizing both the NTDs and CTDs of a MutL dimer [30-33], and that MutL can form a fast and stable diffusive sliding clamp [14]. Under our structural conditions we did not observe this, and only found LN40 well defined on DNA (Chapter 4)

For the future many questions remain: Does the conformation of MutL and MutL homologs, depend on the next step that is necessary in the MMR pathway? Which protein needs to be recruited? Is the nick in the DNA already made? These are signals that MutL needs to 'know' before the next step is activated. Studies have shown the interaction sites between MutL and MutH, MutL and UvrD and MutL and the β-clamp. However, what signals both internal and external, and what stimulates MutL to act properly, is not clear and these are important to study to increase our understanding of MMR. Therefore, more research needs to be done to study MutL, and partners, on DNA.

MutL and MutS are dynamic proteins. As shown, they interact with multiple downstream effector proteins, which makes these ATPases an interesting subject to study. In this thesis, we have shown novel conformations of both MutS and MutL, both alone and on DNA, using X-ray crystallography and cryo-EM. Together with all the validation work that we showed in the chapters, it increased our understanding of MMR initiation. The scanning conformation of MutS gave us new insights how MutS scans and initiates the bent in DNA. The intermediate state of MutS showed how MutS transitions from MM-bound towards a ATP-induced sliding clamp. The sliding clamp of MutS bound to $MutL^{LN40}$ showed the position of the DNA and that a single $MutL^{LN40}$ bound to MutS is sufficient. We also showed that MutL interacts independently of MutS, with DNA, via its N-terminal domains which was predicted but not shown before.

However, more research needs to be performed to have a better understanding of the transitions MutS and MutL need to make from one step towards the other. Knowing how full length MutL is loaded and released from MutS and how the next step is initiated by MutL would be of great interest. This would contribute to the MMR initiation cascade movie.

REFERENCES

1. Bhairosing-Kok, D., et al., *Sharp kinking of a coiled-coil in MutS allows DNA binding and release.* Nucleic Acids Res, 2019. **47**(16): p. 8888-8898.
2. Bürmann, F., et al., *A folded conformation of MukBEF and cohesin.* Nature Structural & Molecular Biology, 2019. **26**(3): p. 227-236.
3. Minnen, A., et al., *Control of Smc Coiled Coil Architecture by the ATPase Heads Facilitates Targeting to Chromosomal ParB/parS and Release onto Flanking DNA.* Cell Rep, 2016. **14**(8): p. 2003-16.
4. Soh, Y.M., et al., *Molecular basis for SMC rod formation and its dissolution upon DNA binding.* Mol Cell, 2015. **57**(2): p. 290-303.
5. Nirwal, S., et al., *Mechanism of formation of a toroid around DNA by the mismatch sensor protein.* Nucleic Acids Res, 2018. **46**(1): p. 256-266.
6. Obmolova, G., et al., *Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA.* Nature, 2000. **407**(6805): p. 703-10.
7. Lamers, M.H., et al., *The crystal structure of DNA mismatch repair protein MutS binding to a G x T mismatch.* Nature, 2000. **407**(6805): p. 711-7.
8. Natrajan, G., et al., *Structures of Escherichia coli DNA mismatch repair enzyme MutS in complex with different mismatches: a common recognition mode for diverse substrates.* Nucleic Acids Res, 2003. **31**(16): p. 4814-21.
9. Warren, J.J., et al., *Structure of the human MutSalpha DNA lesion recognition complex.* Mol Cell, 2007. **26**(4): p. 579-92.
10. Gupta, S., M. Gellert, and W. Yang, *Mechanism of mismatch recognition revealed by human MutSbeta bound to unpaired DNA loops.* Nat Struct Mol Biol, 2011. **19**(1): p. 72-8.
11. Allen, D.J., et al., *MutS mediates heteroduplex loop formation by a translocation mechanism.* Embo j, 1997. **16**(14): p. 4467-76.
12. Qiu, R., et al., *Large conformational changes in MutS during DNA scanning, mismatch recognition and repair signalling.* Embo j, 2012. **31**(11): p. 2528-40.
13. Groothuizen, F.S., et al., *MutS/MutL crystal structure reveals that the MutS sliding clamp loads MutL onto DNA.* Elife, 2015. **4**: p. e06744.
14. Liu, J., et al., *Cascading MutS and MutL sliding clamps control DNA diffusion to activate mismatch repair.* Nature, 2016. **539**(7630): p. 583-587.
15. Acharya, S., et al., *The coordinated functions of the E. coli MutS and MutL proteins in mismatch repair.* Mol Cell, 2003. **12**(1): p. 233-46.
16. Hays, J.B., P.D. Hoffman, and H. Wang, *Discrimination and versatility in mismatch repair.* DNA Repair (Amst), 2005. **4**(12): p. 1463-74.
17. Ahrends, R., et al., *Identifying an interaction site between MutH and the C-terminal domain of MutL by crosslinking, affinity purification, chemical coding and mass spectrometry.* Nucleic Acids Res, 2006. **34**(10): p. 3169-80.
18. Lopez de Saro, F.J., et al., *The beta sliding clamp binds to multiple sites within MutL and MutS.* J Biol Chem, 2006. **281**(20): p. 14340-9.

19. Almawi, A.W., et al., *Binding of the regulatory domain of MutL to the sliding β-clamp is species specific.* Nucleic Acids Res, 2019. **47**(9): p. 4831-4842.

20. Liu, J., et al., *MutL sliding clamps coordinate exonuclease-independent Escherichia coli mismatch repair.* Nat Commun, 2019. **10**(1): p. 5294.

21. Grilley, M., et al., *Isolation and characterization of the Escherichia coli mutL gene product.* J Biol Chem, 1989. **264**(2): p. 1000-4.

22. Ordabayev, Y.A., et al., *Regulation of UvrD Helicase Activity by MutL.* J Mol Biol, 2018. **430**(21): p. 4260-4274.

23. Hall, M.C. and S.W. Matson, *The Escherichia coli MutL protein physically interacts with MutH and stimulates the MutH-associated endonuclease activity.* J Biol Chem, 1999. **274**(3): p. 1306-12.

24. Pluciennik, A., et al., *PCNA function in the activation and strand direction of MutLalpha endonuclease in mismatch repair.* Proc Natl Acad Sci U S A, 2010. **107**(37): p. 16066-71.

25. Yamamoto, T., et al., *Evidence for ATP-dependent structural rearrangement of nuclease catalytic site in DNA mismatch repair endonuclease MutL.* J Biol Chem, 2011. **286**(49): p. 42337-48.

26. Sacho, E.J., et al., *Direct visualization of asymmetric adenine-nucleotide-induced conformational changes in MutL alpha.* Mol Cell, 2008. **29**(1): p. 112-21.

27. Monakhova, M., et al., *Probing the DNA-binding center of the MutL protein from the Escherichia coli mismatch repair system via crosslinking and Förster resonance energy transfer.* Biochimie, 2020. **171-172**: p. 43-54.

28. Genschel, J., et al., *Interaction of proliferating cell nuclear antigen with PMS2 is required for MutLα activation and function in mismatch repair.* Proc Natl Acad Sci U S A, 2017. **114**(19): p. 4930-4935.

29. Pillon, M.C., J.H. Miller, and A. Guarne, *The endonuclease domain of MutL interacts with the beta sliding clamp.* DNA Repair (Amst), 2011. **10**(1): p. 87-93.

30. Ban, C., M. Junop, and W. Yang, *Transformation of MutL by ATP binding and hydrolysis: a switch in DNA mismatch repair.* Cell, 1999. **97**(1): p. 85-97.

31. Hall, M.C., et al., *DNA binding by yeast Mlh1 and Pms1: implications for DNA mismatch repair.* Nucleic Acids Res, 2003. **31**(8): p. 2025-34.

32. Guarne, A., et al., *Structure of the MutL C-terminal domain: a model of intact MutL and its roles in mismatch repair.* Embo j, 2004. **23**(21): p. 4134-45.

33. Kosinski, J., et al., *Analysis of the quaternary structure of the MutL C-terminal domain.* J Mol Biol, 2005. **351**(4): p. 895-909.

ABBENDUM

SUMMARY

DNA mismatch repair (MMR) is a DNA repair mechanism that corrects small insertions, deletions and mismatches. Almost every organism has MMR and therefore this mechanism is highly conserved. For this reason, we studied MMR initiation in the model organism *Escherichia coli* (*E.coli*). The findings of the conformational changes that *E.coli* MMR proteins undergo and what this means for our understanding of MMR, are discussed in this dissertation.

Chapter 1 discusses the current understanding of MMR in general. The mechanism is explained in great detail. MMR initiation is conserved from prokaryotes to eukaryotes, but follow-up steps are different in several species. These similarities and differences are discussed. There are many proteins involved in the MMR cascade, but in this dissertation and this chapter we discuss mainly two MMR proteins: MutS and MutL. These are the initiators of MMR. MutS scans DNA and detects mismatches. As a result, MutS binds ATP and recruits MutL. During MMR initiation, MutS and MutL undergo several conformations. These conformations are studied in great detail by us and other groups. All these steps and conformations are being discussed in great detail. Finally, structural biology methods cryo-Electron Microscopy (cryo-EM) and X-ray crystallography are being discussed briefly since these methods are being used in all chapters.

In **Chapter 2** we show the crystal structure of *E.coli* MutS in its apo-state. We discuss this new conformation of MutS and compare this with other known conformations of MutS (homologs). In addition, we also validated this structure with mutational studies, biophysical experiments and crosslinking. The apo-state of MutS showed a kink in both lever domains, with one monomer kinking inwards and the other facing outwards. In both monomers the kink is originating from a defined hinge point. Mutational studies in and around this hinge point show that kinking of the lever domains is required for MutS to proceed to DNA binding. In addition, we showed with intra-crosslinking experiments that both monomers are able to kink outwards. Finally, it was found that MutL loading also depends on the ability of MutS to kink the levers.

In **Chapter 3** we show the results of a shared project with our collaborators on studying several novel conformations of MutS. By using cryo-EM, we solved four MutS cryo-EM structures; MutS in scanning mode, MutS bound to mismatched DNA, MutS in an intermediate conformation that transitions towards a sliding clamp, and the MutS sliding clamp bound to DNA and the N-terminal domain of MutL ($MutL^{LN40}$). These conformations were then validated using biochemical and biophysical techniques. We showed that, during scanning, MutS actively tries to distort the DNA by inducing a bend in the DNA via the clamp domains. At the mismatch, this weak point in the DNA allows MutS to induce the kink in the DNA that has been observed before. The novel intermediate conformation of MutS shows both similarities with MutS bound to a mismatch and MutS sliding clamp, showing the step-wise transition that MutS undergoes after mismatch detection. The sliding clamp conformation of MutS was obtained by crosslinking single cysteine MutS to single cysteine $MutL^{LN40}$. The cryo-EM map showed the MutS dimer bound to one $MutL^{LN40}$ and DNA. This, together with validation experiments shows that MMR *in vitro* is functional if MutL only binds one of the two bi-partite interaction sides rather than both, as was observed in the symmetric crystal structure.

Chapter 4 shows the work regarding the attempts and strategies to obtain a full length cryo-EM structure of MutS and MutL on DNA. It is known that MutS and MutL form a transient complex on DNA and that both proteins are dynamic and can adopt multiple conformations. For this reason, we studied the optimal conditions to trap the complex of MutS and MutL on DNA and to obtain a homogeneous sample. Several cryo-EM datasets have been collected with different settings. First dataset showed good 2D classes but no 3D reconstruction was possible due to preferred orientations that were induced by a too long DNA molecule and streptavidin blocks at the end of the DNA oligo. The second dataset was obtained on a shorter DNA molecule and open ends DNA, which resulted in minimal complex formation and therefore no 2D and 3D classes could be obtained of MutS-MutL. The third attempt was crosslinking full length single cysteine MutS with full length single cysteine MutL. This complex showed a long lifetime in Surface Plasmon Resonance experiments and were therefore used for cryo-EM analysis. Strikingly, only MutL particles and classes were obtained instead of the complex of MutS-MutL. In conclusion, the study is still ongoing and all observations are being discussed in the context of known literature.

In addition to the work on MutS-MutL, we also studied the complex of MutL on DNA using cryo-EM. This showed several 3D classes, with one of these being the most dominant and well-characterized. This showed the binding of two $MutL^{LN40}$ molecules to a straight DNA molecule. In this reconstruction the C-terminal domains are not visible, indicating that they are disordered.

In **Chapter 5** we discuss the results of chapter 2, 3 and 4 in context of the current state of the MMR field. What did we learn from our efforts, and what information is still missing? This chapters discusses the conclusions and makes suggestions for new experiments.

SAMENVATTING

DNA mismatch reparatie (MMR) is een DNA herstel mechanisme dat kleine invoegingen, verwijderingen of foutjes (*mismatches*) corrigeert in het DNA. Bijna ieder organisme heeft MMR en het is erg geconserveerd. Om te begrijpen hoe het werkt, bestuderen wij de start van MMR in een model organisme, Escherichia coli (*E.coli*). De bevindingen omtrent de conformationele veranderingen van de MMR-eiwitten en wat deze betekenen voor het beter begrip van MMR, worden besproken in dit proefschrift.

In **Hoofdstuk 1** bespreken we de huidige status van het MMR-veld. Het mechanisme wordt hier in detail besproken. De start van MMR is geconserveerd van prokaryoten tot eukaryoten, maar het verdere verloop van MMR is anders in verschillende soorten organismen. Deze overeenkomsten en verschillen worden besproken. Er zijn veel eiwitten betrokken bij MMR maar in dit proefschrift bespreken we voornamelijk twee MMR-eiwitten: MutS en MutL. Deze twee eiwitten zijn de initiatiefnemers van het traject. MutS scant het DNA en detecteert de *mismatches*. Als gevolg hiervan, bindt MutS ATP en rekruteert het MutL. Tijdens dit proces, nemen MutS en MutL verscheidene conformaties aan. Deze conformaties worden door veel onderzoeksgroepen, inclusief ons, bestudeerd. Tenslotte worden de structurele biologie methoden cryoelektronenmicroscopie (cryo-EM) en röntgenkristallografie besproken aangezien deze twee methoden breed worden toegepast in de hoofdstukken in dit proefschrift.

In **Hoofdstuk 2** tonen wij de kristalstructuur van *E.coli* MutS *apo*-staat. We bespreken deze nieuwe conformatie van MutS en vergelijken deze met andere bekende conformaties van MutS en MutS-homologen. Daarnaast hebben we de structuur gevalideerd met mutatieproeven, biofysische experimenten en crosslinken. De *apo*-status van MutS laat een knik zien in beide *lever* domeinen, waar de ene monomeer een knik naar binnen laat zien en de ander een knik naar buiten. In beide monomeren is de knik ontstaan vanuit hetzelfde specifieke scharnierpunt. Mutatieproeven laten zien dat deze knik-bewegingen nodig zijn om MutS aan DNA te binden. Daarnaast laten we ook zien dat beide monomeren naar buiten kunnen knikken. Tenslotte zien we dat de knik nodig is om MutS in staat te stellen om MutL te laden op het DNA.

In **Hoofdstuk 3** laten wij de resultaten zien van een samenwerking waarbij we verschillende conformaties van MutS gevonden hebben. Door gebruik te maken van cryo-EM hebben we vier MutS cryo-EM-structuren opgelost; MutS in de scanning-modus, MutS gebonden aan DNA met een mismatch, MutS in de overgang naar de *sliding clamp* en MutS in de *sliding clamp* conformatie gebonden aan het N-terminale domein van MutL ($MutL^{LN40}$). Deze structuren zijn gevalideerd met behulp van biochemische en biofysische technieken. We laten zien dat MutS, tijdens het scannen van DNA, probeert om het DNA te vervormen en een buiging te introduceren. Bij de mismatch is er een zwak punt in het DNA en dit stelt MutS in staat het DNA te knikken, zoals we al eerder hebben gezien. De nieuwe conformatie waar MutS in de overgang is naar de *sliding clamp,* lijkt zowel op de mismatch-gebonden conformatie als op de *sliding clamp* conformatie. Dit geeft duidelijk aan dat MutS dit in een stapsgewijs proces ondergaat. De structuur van MutS in de *sliding clamp* conformatie was verkregen door $MutL^{LN40}$ aan MutS te crosslinken. De cryo-EM-structuur laat zien dat MutS één $MutL^{LN40}$ molecuul op DNA laadt. Dit is gevalideerd door heterodimeren te maken van MutS waarbij maar een binding site van MutS beschikbaar is. Daarmee toonden we aan dat MMR *in vitro* functioneel kan zijn als MutL maar aan één van de tweedelige bindingoppervlaktes bindt in plaats van aan beide zoals was geobserveerd in de kristalstructuur.

Hoofdstuk 4 laat alle bevindingen en strategieën zien om een cryo-EM-structuur te verkrijgen van *full length* MutS en MutL op DNA. Het is bekend dat MutS en MutL een kortdurend complex vormen en dat beide eiwitten dynamisch zijn en meerdere conformaties kunnen aannemen. Om deze reden, hebben we gezocht naar optimale condities, waar MutS en MutL stabiel op DNA geladen zijn. Verscheidene cryo-EM-datasets zijn opgenomen onder verschillende condities. De eerste dataset laat goede 2D-klassen zien, maar een 3D-reconstructie was niet mogelijk vanwege voorkeur oriëntaties die kwamen door een te lang DNA-molecuul en streptavidine blokken aan de DNA-uiteindes. De tweede dataset was opgenomen op een korter DNA-molecuul en met open uiteindes. Dit leidde tot minimale complexvorming en daarom konden geen 2D- en 3D-klassen worden waargenomen voor MutS-MutL. De derde poging was het crosslinken van MutS aan MutL. Met behulp van *DNA binding* experimenten, zagen we dat dit complex langdurig bij elkaar bleef, en daarom perfect zou zijn voor cryo-EM. Merkwaardig genoeg zagen we alleen MutL

in plaats van MutS en MutL samen. Op basis hiervan kunnen wij geen concrete conclusies trekken. Het onderzoek is nog gaande en voorlopige resultaten worden besproken.

Naast het bestuderen van MutS en MutL op DNA, hebben we ook gekeken naar het complex van MutL en DNA met cryo-EM. Hier vonden we verschillende 3D-klassen, waarvan één het meest detail toonde. Dit liet de binding zien van twee $MutL^{LN40}$ moleculen gebonden aan een recht DNA-molecuul. Kennelijk was het C-terminale domein niet voldoende geordend om zichtbaar te zijn.

In **hoofdstuk 5** is een discussie waar we de resultaten van hoofdstuk 2, 3 en 4 plaatsen in de context van de status van het MMR-veld. Wat hebben we geleerd van deze experimenten, en welke informatie mist nog? Kunnen we nieuwe conclusies trekken en welke nieuwe experimenten zijn nodig? Dit wordt allemaal besproken in het laatste hoofdstuk.

STELLINGEN

1. MutS scant DNA tijdens DNA-replicatie. MutS probeert zwakke plekken in het nieuwe DNA te detecteren door een buiging in het DNA op te wekken. MutS doet dit met behulp van de *clamp* en *lever* domeinen die tijdens het scannen aan het DNA binden. Bij een zwakke plek lukt het MutS om DNA te buigen en kan de volgende stap plaatsvinden. (dit proefschrift)

2. MutS is in staat de lever domeinen te buigen op een speciaal gedefinieerd scharnierpunt. Daarmee kan MutS het bovenste deel van dit domein flexibel bewegen. Deze flexibiliteit is nodig voor het laden op DNA en de transitie naar de *sliding-clamp* conformatie. (dit proefschrift)

3. Na het detecteren van de mismatch creëert MutS twee bindingsvlakken voor MutL, maar MutL heeft maar één bindingsvlak nodig om MMR te starten *in vitro*. Het maakt niet uit welke van de twee vlakken MutL kiest. (dit proefschrift)

4. MutS en MutL vormen een complex maar dit complex is erg tijdelijk en moeilijk vast te leggen met structureel biologische methoden. (dit proefschrift)

5. Het N-terminale domein van MutL bindt individueel aan het DNA terwijl het C-terminale domein flexibel is. (dit proefschrift)

6. Science and everyday life cannot and should not be separated. (Rosalind Franklin)

7. Have no fear of perfection; you'll never reach it. (Salvador Dali)

8. Gezuiverde eiwitten kunnen het best bewaard blijven bij -80°C en niet op kamertemperatuur

9. Racism is still with us. But it is up to us to prepare our children for what they have to meet, and, hopefully, we shall overcome. (Rosa Parks)

10. Een proefschrift schrijven in corona-tijd met een baby thuis vergt discipline, doorzettingsvermogen en een lieve partner

11. Brutalen hebben de halve wereld. (Erik Kok)

CURRICULUM VITAE

Doreth Kok werd geboren op 12 april 1991 te Stede Broec en groeide op in Bovenkarspel. In 2009 behaalde ze daar haar VWO-diploma met een dubbel profiel (Natuur & Gezondheid en Natuur & Techniek). Hierna begon zij aan de bachelor Moleculaire Levenswetenschappen aan de Radboud Universiteit in Nijmegen. Deze werd in 2009 afgerond met een stage in het lab van prof. dr. Bé Wieringa op de afdeling Celbiologie aan het Radboud Institute for Molecular Life Sciences (RIMLS). Na het afronden van de bachelor in 2009 is ze gestart met de onderzoeksmaster Biomolecular Sciences aan de Vrije Universiteit te Amsterdam. De eerste masterstage heeft ze voltooid in het lab van prof. dr. Martine Smit onder begeleiding van dr. Folkert Verkaar. Tijdens deze stage onderzocht ze een nieuwe regulator van het Wnt-Pathway. Haar enthousiasme voor dit onderzoeksveld heeft ervoor gezocht om ook tijdens de tweede stage aan Wnt-signaling te werken, in het lab van prof. dr. William Weis aan Stanford University in Amerika, onder leiding van dr. Jayanth Chodaparambil. Hier bestudeerde zij de rol van een andere regulator van het Wnt-pathway en deed zij ook structuur biologisch onderzoek. Doreth besloot dat zij in dit soort onderzoek promotieonderzoek wilde doen. In september 2014 begon ze als onderzoeker in opleiding in de groep van prof. dr. Titia Sixma op de afdeling Biochemie aan het Nederlands Kanker Instituut. Een deel van de werkzaamheden in dit proefschrift werd uitgevoerd op het MRC Laboratory of Molecular Biology in Cambridge (Engeland) en het Centro Nacional de Investigaciones Oncológicas in Madrid (Spanje), onder supervisie van dr. Meindert Lamers en dr. Rafael Fernández-Leiro. De resultaten van dit promotieonderzoek staan beschreven in dit proefschrift.

DANKWOORD

En dan het laatste hoofdstuk van dit proefschrift, het dankwoord. Het lijkt niet zo belangrijk na alle andere hoofdstukken, maar dat is het eigenlijk wel. Want promoveren doe je niet alleen. Van het begin tot het eind zijn er veel mensen bij betrokken en die ik wil bij deze graag bedanken.

Als eerste wil ik **Titia** bedanken. Bedankt voor de kans dat ik onderzoek mocht doen in jouw groep. Vanaf de eerste weken wist ik waar ik aan toe was. Je bent duidelijk in je mening en daar houd ik van. Ik ben dat namelijk ook, en dit werkte voor ons beide. Ik waardeer het heel erg dat je me op mijn tempo hebt laten ontwikkelen en me alleen een zet gaf wanneer dit nodig was. Hierdoor kon ik mijn draai vinden in mijn PhD. Je hebt me de kans gegeven om cryo-EM te leren toen dit nog nieuw was voor ons lab. Dit was spannend maar je zorgde voor goede collaborations met Rafa en Meindert, waardoor ik kon groeien. Dankjewel.

A big thanks to **Meindert** and **Rafa**. In my first week we started our collaboration and in my second week I visited you in Cambridge at the LMB. It felt good from the first moment. The visits became longer and more frequent, more Krios sessions were booked and the hours of sleep became less (pre-EPU time). Rafa, I admire your enthusiasm and the fact that you always (!) took time for me to explain me everything I asked you, despite how busy you were with your work and family. From the simple things of making a new folder in the terminal, to making good quality cryo-EM grids and analyzing cryo-EM data. You always took the effort to Whatsapp, Skype, email or call with me when I was lost in the world of processing. I learned so much from you. Also, you visited us many times when we needed your help as cryo-EM newbies at the NKI. It sounds cheesy, but without you, my PhD and my chapters would be very empty and boring. I also enjoyed my year at your lab. It was a warm welcome, I will never forget that. Again, thanks a lot for everything, *muchas gracias por todo*. Meindert, bedankt voor het warme welkom in jouw lab in Cambridge toen ik net begon aan mijn PhD. Je liet me niet aan mijn lot over en dat was heel fijn. Dankjewel dat je mijn co-promotor wilt zijn.

Tassos, my second favorite Greek at the floor and ex-MMR member. Thank you for your guidance during my PhD and your input during the committee meetings.

This was useful. I hope the new microscope brings your group many structures. Also, big thanks for facilitating the RHPC which made processing a bit easier these last years.

Hein en **Heinz**, bedankt dat jullie in mijn PhD commissie zaten. Ik waardeerde jullie kritische vragen tijdens de jaarlijkse meetings. Naast Titia en Tassos was het goed om ook met jullie samen kritisch te kijken naar mijn projecten, en hier en daar keuzes te maken. Dit was nuttig en nogmaals bedankt daarvoor.

MMR is a topic that is being studied in the lab for more than two decades. Often this was done in collaboration with other labs. **Joyce** and **Peter**, I enjoyed the MMR meetings with you and the other lab members a lot. Often, I was amazed of how many details you knew regarding papers and performed experiments, even after a 3-course dinner and drinks. Thank you for these inspirational meetings, both online and in real life. Also, for all your input on my work, this was always very much appreciated. And a big tanks to the other members of the DNARepairMan group: **Terence**, **Mark** and **Nigel** and all the PhD students.

En dan komen we aan bij B8. En wat was B8 zonder **Herrie** en **Pim**. **Herrie**, op de valreep voor je pensioen hebben we toch nog intensief samengewerkt. Het lukte je om full length MutS en MutL aan elkaar te crosslinken. Met een kleine optimalisatie werd dit een perfect sample voor mij om te analyseren met cryo-EM in Madrid. Gelukkig heb je het mij ook goed geleerd en kon ik het na je pensioen overnemen. Dankjewel voor de continuïteit in ons MMR groepje, en bedankt voor alle preps. **Pim**, wij werkten niet samen, maar de ochtenden samen als vroege vogels was een goede start van mijn dag (jouw dag was dan al 2 uur bezig). Bedankt voor de leuke gesprekken, adviezen en het starten en induceren van mijn kweek na dat Herrie met pensioen was. Ik wens je een mooi pensioen met je kleinkinderen en hopelijk kun je volgend jaar toch de fietstocht maken die op de planning staat. **Sasha**, I remembered you as the guy that asked me a few critical questions during my job interview. Later on, I realized that this was nothing personal and that I was super lucky to work with you in the MMR group. I learned so much from you, in biophysics, but also in life. Thank you! **Nassos**, where do I need to start? I got to know you as a shy, hardworking guy in my office who didn't say much. But spending New Years together because you got stuck in the Netherlands because of your visa,

was the start of a good friendship. I appreciate that you are always willing to help, and man, you are good in cloning. And thank you for joining me to drovers dog to eat burgers. **Shreya**, my office and bench neighbor, Indian tour guide and personal shopper for lehengas (for both me and Léla), and paranymph! It is always fun and relaxed to have you around. Unfortunately, we could not spend the last months in our office together, but we managed anyhow to finish all together. Super proud on all of us! **Bart**, jij begon in hetzelfde jaar als Nassos en ik, maar jij was blijkbaar een stukje slimmer en bent vorig jaar al gepromoveerd. Ik vond het heel relaxed om met jou in het kantoor te zijn. Altijd gezellig en nuchter, dat paste goed bij me. Bedankt dat ik jouw paranimf mocht zijn! **Andrea**, the expert in BRCA and workstations (which is not the same as a PC!). You helped me with setting up the processing facilities at the RHPC. I can imagine you were not always excited when I came to ask for your help when something was not working for me on the workstation. But, you always took time and you were patient. And I very much appreciated this. Thank you Andrea, **Torben**, **Ismail**, **Robbie** and **Tassos** for setting up the RHPC. **Torben**, thank you for all your help regarding data transfers, login problems from home and Madrid. Thanks to **Patrick**, **Magda**, **Justina**, **John**, **Yvette** and **Tati**. You guys are a constant factor in the lab and that's super important to keep the lab ongoing. Yvette, you did a great job by becoming our new lab manager. Tati, thanks for all your help, I hope you enjoy your new job. I wish everyone the best. **Susanne**, you joined the MMR group halfway my PhD which was really nice to discuss our results. I hope you can get the structures of the MMR proteins, good luck! Thanks **Luca** for bringing good vibes to B8 and the lab, and for allowing me to take your spot in Oficina de Frustracion. Also, thanks for all the other B8 members for making my PhD a great period. **Mirna**, waar waren we zonder je geweest op B8. Bedankt voor je interesse en snelle hulp, waar ik ook was. **Patty**, je hebt een tijd ons geholpen op B8 vanuit B7, dankjewel daarvoor.

Besides my colleagues at the NKI, I spent quite some time at the CNIO. **Ana**, you are so much fun! Please always stay yourself, because I love your energy and kindness. Thanks for making my time at the CNIO a lot of fun, because drinking *té con leche* with you is much more fun than 'drinking' alone. Thank you **Jaska**, **Carmen**, **Anna**, **Ara**, **Samu, Ángel** and all the others for helping me and making my stay at the CNIO memorable. I also want to thank **Jay**, my master supervisor in Stanford. You introduced me to structural biology and gave me the confidence during my master

to start a PhD. And I was even more lucky to see your dancing skills at my wedding. Thanks, and hopefully I can visit you soon in Boston.

Anne, dankjewel voor alle lab (nerd) memes en de gezellige bezoekjes in Londen, Hilversum en Amersfoort. Het kan helaas niet zo vaak omdat je in Londen woont, maar gelukkig hebben ze daar nu in ieder geval een goede biologie docent er bij. Ik hoop je snel weer te bezoeken in Londen. **Femke**, je hebt altijd interesse gehad in alles wat ik doe en meemaak. Jouw luisterend oor is altijd fijn. Je bent een lieverd, en ben heel blij dat je mijn paranimf bent. **Alwin**, ook wil ik jouw bedanken. Je moest het allemaal maar aanhoren tijdens onze etentjes maar je bleef altijd de goede vragen stellen en dat waardeer ik. Wie weet gaan we samen ooit nog bij 'een bedrijf' werken. **Kirsten**, **Simone** en **Mariska**, ik blijf natuurlijk een half BMW-kindje, maar ik ben toch blij dat ik er altijd weer bij ben. Zelfs nu we allemaal druk zijn en verspreid over het land wonen, lukt het ons toch om te blijven afspreken. Thanks voor de gezelligheid de afgelopen tijd!

Je partner kies je, maar je schoonfamilie krijg je er bij. En hoewel dat voor sommige niet altijd goed uitpakt, mag ik mij gelukkig prijzen met mijn schoonfamilie. **Rosita**, bedankt voor alles. Dat klinkt een beetje breed, maar zo is het. Met allerlei dingen heeft u me wel geholpen. Extra koken voor de late avonden, een jaar mijn thuis zijn omdat we bij u woonden, of extra oppassen op Léla toen ik druk was met afronden. **Tante Shyama**, dankuwel voor de gezelligheid en alle goede zorgen. De etentjes met kerst waren ieder jaar weer iets om naar uit te kijken, maar ook alle restaurants die we tussendoor samen bezochten was altijd fijn en gezellig.

Mijn zussen **Karin** en **Ellen**, en ook **Mark** en **Niek**, bedankt voor alles de afgelopen jaren. Voor het begrip als ik weer weg moest, en we de familieplanning toch weer even om moesten gooien. En Karin en Niek, voor het oppassen in Madrid toen ik zelf toch weer terug naar Amsterdam moest. **Mama** en **Papa**, bedankt voor alles. We moesten altijd een bijbaantje hebben, en dat vond ik soms wel eens stom. Maar het heeft me gevormd tot wie ik ben en daar ben ik heel dankbaar voor. Papa, bedankt dat je me hebt laten in zien dat je voor jezelf mag opkomen. Brutalen hebben de halve wereld, dat is iets wat ik vaak in mijn achterhoofd heb. Mama, bedankt voor alle goede zorgen, ook toen we weer een jaartje thuis kwamen wonen. Papa boft met zo'n goed thuis.

En dan als laatste, mijn gezin. **Pat**, al ruim 15 jaar zijn we samen en maken we van alles mee. Je bent altijd lief en zorgzaam voor mij en voor Léla. Je hebt me altijd gesteund en zelfvertrouwen gegeven als ik het heel eventjes miste. Ook tijdens alle bezoeken naar Cambridge en Madrid was je er voor me, niks was je te gek. Tijdens de lockdown zorgde je ervoor dat de weg naar het einde van mijn PhD vrij was en dat zal ik nooit vergeten. Never change a winning team, te quiero mucho. **Léla**, je bent nog klein, maar zo perfect. Papa en ik houden van je.