Reflections on the Mechanism of DNA Mismatch Repair

Nicolaas Hermans

Reflections
on the Mechanism of
DNA Mismatch Repair

Nicolaas Hermans

Reflections
on the Mechanism of
DNA Mismatch Repair

Bespiegeling op het mechanisme
van DNA mismatch reparatie

Proefschrift

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

Prof.dr. H.A.P. Pols

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
dinsdag 16 september 2014 om 15:30 uur
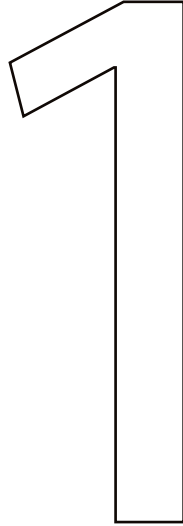door

Nicolaas Hermans
geboren te Amsterdam

ERASMUS UNIVERSITEIT ROTTERDAM

TABLE OF CONTENTS

1

GENERAL INTRODUCTION

1

An apple and an egg

Chicken eggs and duck eggs can look much alike. If someone gave me two eggs, one from a chicken and one from a duck, I would not be surprised if I could not see which one is which. To find the answer, the only thing one has to do is wait, because chicken eggs will always yield chicks, while duck eggs contain ducklings. And even though people have known that a chicken egg will not yield a duck ever since chickens and ducks have been domesticated, why this is the case has only been clarified in the last century.

Deoxyribonucleic acid, better known by its abbreviation DNA, is the carrier of genetic information. A chicken egg will always yield a small chick, because the genetic information carried in the DNA came from its parents, a chicken and a rooster, and the DNA the chick carries is an (almost) exact copy of DNA of its parents. An almost exact copy, but as Darwin explained even before DNA was known to be genetic carrier, mutations do occur (Darwin 1859). Indeed, if enough mutations persist in the genome, the chicken can become a duck, given enough time and selection for the important mutations.

To appreciate the improbability of a chicken egg containing a duckling, it is necessary to put some perspective into how often mutations in DNA occur, how often these mutations persist, and what the implications of these mutations can be. DNA consists of two anti-parallel strands, and each strand contains a chain of nucleotides. These nucleotides are the carrier of the genetic information. DNA polymerases, the enzymes that create copies of DNA molecules, make few errors. Pol δ, one of the polymerases involved in the replication of DNA in human cells, has an error rate of 1 in 100,000 (Fortune, Pavlov et al. 2005). A truly impressive number, if you take into account that the difference between the four nucleotides, the building blocks of DNA, comes down to a few atoms and the ability to form either 2 or 3 hydrogen bonds. At the same time the amount of DNA that needs to be copied before a cell can divide is enormous: the number of DNA base pairs in every single cell in our body capable of cell division is 6×10^9. So, even at this low error rate, thousands of mistakes will occur during every replication cycle. Every single mistake can lead to a mutation if left unrepaired. Theoretically, a certain combination of multiple mutations can change a chicken into a duckling, were it not that the required mutations are unlikely to occur all at once by chance. A good analogy is the infinite monkey theorem: if a monkey behind a typewriter hits random keys, it will, given infinite time, at some point produce the complete works of William Shakespeare[1]. The

---

[1]This theorem has been used to attack the evolution theory, since evolution would surely be as unlikely. However, evolution theory does not deny this improbability, but provides and answer: natural selection. Jorge Luis Borges assay "The total library" ("La biblioteca total") traces back the origins of this theorem, and imagines that the same monkey could produce a library containing all books ever written, and all books that will be written in the future.

chance of doing so however is so small the monkey will probably not have reproduced a single page before the universe ends.

Mutations have an added disadvantage: the chance that a single mutation will damage the genetic information is larger than the chance it will have a beneficial effect. Each mutation potentially contributes to death of the new cell, or towards the development of a cancer cell. It may therefore seem unsurprising that a sizable number of proteins deal with the removal of damaged or wrongly copied DNA, in order to minimize the number of mutations. This process is called "genome maintenance". This chapter aims to introduce the main topic of this thesis, DNA mismatch repair, and put it into the perspective of genome maintenance. Mismatch repair is distinctly different from other repair pathways, because it is the only DNA repair pathway that does not repair damaged DNA. Rather, it can recognize wrongly inserted nucleotides, and remove those nucleotides again while leaving the template intact. This seems straightforward, but at the site of the mismatch it is impossible to know which base is correct and which one needs to be removed. This complicates mismatch repair, and makes it all the more interesting. I will give an overview of the mismatch repair machinery as it was discovered in *Escherichia coli*, a bacterium commonly found in the lower intestine of mammals. *E. coli* serves as the prokaryotic model organism in biological research. Most of the work presented in this thesis also uses the *E.coli* model system, but mismatch repair as it works in our own cells is ultimately more interesting for most of us. Therefore I will also introduce eukaryotic mismatch repair, and explain some of the advantages of looking at the model system in *E.coli*.

DNA damage and repair

DNA metabolism can be categorized in the three R's: Replication (copying of DNA before a cell division), Recombination (the combination of and exchange between separate DNA molecules), and Repair (restoration of DNA lesions) (Friedberg 2003). Several DNA repair pathways exist, each repairing a specific set of lesions. Base excision repair (BER) removes single damaged bases, often caused by oxidative, alkylation, deamination, and depurination /depyrimidination damage (Robertson, Klungland et al. 2009). These damaged nucleotides occur frequently and are relatively simple to repair, often the damaged nucleotide can be cut out of the DNA, and a new base is inserted by a polymerase. Intrastrand crosslinks (crosslinks between bases in the same DNA strand) and bulky adducts are repaired by nucleotide excision repair (NER) (de Laat, Jaspers et al. 1999). These lesions are commonly caused by UV-light. Interstrand crosslinks (crosslinks between bases in opposing DNA strands), which are often introduced by chemo-therapeutic drugs, are processed by the Fanconi anaemia (FA) pathway (Deans

and West 2011) before they can be repaired by other repair pathways. Double stranded breaks in the DNA are repaired by either non-homologous end joining (NHEJ) (Burma, Chen et al. 2006) or Homologous recombination (HR) during replication (Li and Heyer 2008). Double stranded breaks are also often caused by ionizing radiation. Failure or incorrect repair of DNA lesions often results in mutations, chromosomal re-arrangements or cell death. This highlights the importance of DNA repair: every day several thousands of DNA lesions occur in every single cell of our body (Bernstein 2013). One important repair pathway is not yet mentioned, the subject of this thesis: DNA Mismatch repair (MMR). DNA mismatch repair stands out from the other DNA repair pathways because it does not repair damaged DNA. Rather, mismatch repair removes mispaired bases, errors generated by polymerases.

Mismatches consist of small patches of extra helical nucleotides caused by polymerase slippage (referred to as insertion/deletion loops), or nucleotides with non "Watson-Crick" basepairing (Jiricny 2013). DNA consists of 4 basepairs, adenine (abbreviated as A), cytosine (C), guanine (G) and thymine (T). When correctly paired, an A is paired with a T, and a G is paired with a C. This is what is often referred to as "Watson-Crick" basepairing. Any other basepair is thus a mismatch, which means there are 12 different base-base mispairs possible. These mismatches thus consist of otherwise undamaged DNA nucleotides, which can be problematic: it is not obvious which base is correct and which base is incorrect. How mismatch repair proteins solve this riddle will be explained later.

Some mismatches occur more frequent than other mismatches. This is caused by proofreading activity of the replicative DNA polymerases. Almost all replicative polymerases also exhibit exonuclease activity next to their function as polymerase. Mismatches create helical distortions, which make it more difficult for the polymerase to add the next nucleotide. This causes the polymerase to stall briefly, which increases the chance that the wrongly incorporated nucleotide is removed through the exonuclease activity. As a result, this proofreading removes mismatched nucleotides that introduce big helical distortions most efficiently (Arana and Kunkel 2010). Also not every mismatch is recognized in the same way by mismatch repair; for example, a G-T mismatch is repaired more efficiently than a C-C mismatch (Kramer, Kramer et al. 1984, Au, Welsh et al. 1992). It is probably not a coincidence that mismatch repair is most efficient in repairing the mismatches that are left most often by the polymerase (Wu, Clarke et al. , Schaaper and Dunn 1991).

A mismatch is only present when DNA is double stranded. Once the two strands are separated, which happens at every replication cycle, two non-identical copies will be made: a DNA molecule carrying the original sequence,
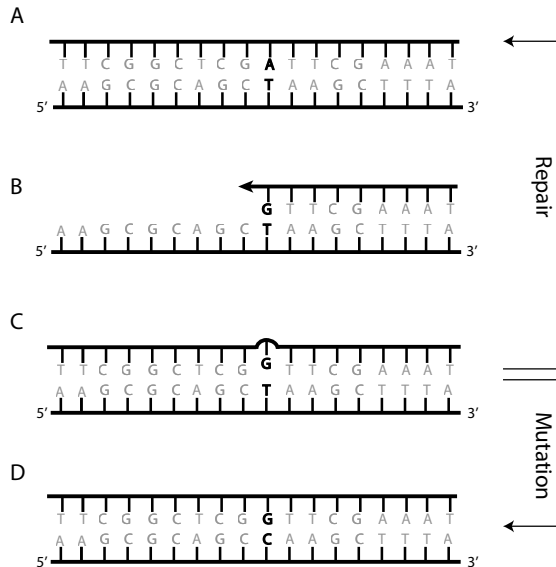
A

Repair

B

C

Mutation

D

Figure 1: Abstract representation of DNA replication and Mismatch Repair. A single DNA strand before replicaton is shown in (A). To copy the DNA, the two complementary strands are separated into two singlestranded DNA molecules. A DNA polymerase can subsequently fill in bases when making a new complementary strand, so that every guanine matches up with a cytosine, and every adenine with a thymine. Sometimes a base is wrongly incorporated, for instance when a guanine is placed opposite of a thymine. This is what is referred to as a mismatched basepair (B). If this mismatch is not removed by the intrinsic proofreading of the polymerase, the mismatch persists in the DNA (C). If not repaired, the mismatch will cause a mutation in one of the strands at the next round of replication (D).

and a mutated version (Figure 1). A single mutation does not necessarily have consequences. In fact, it is estimated that every cell division in humans gives rise to one mutation on average (Iyer, Pluciennik et al. 2006, Hsieh and Yamane 2008). Only when many mutations build up over time big problems arise.

In the last few years, enormous advances in DNA sequencing have provided ample information on the origins of cancer cells. One of the surprises was that thousands of mutations occur before a single cell becomes a tumor cell (Loeb 2011). This indicates that a cell typically has a defect in genomic maintenance, before it can accumulate enough mutations to become a tumor cell. Loss of function of mismatch repair genes is linked to several types of cancer. Lynch syndrome (hereditary nonpolyposis colorectal cancer) is an autosomal dominant genetic condition caused by mutations that impair one of the mismatch repair genes (either MLH1, MSH2, MSH6, PMS2). This means one inherited copy of the dysfunctional gene is sufficient to cause Lynch syndrome. Individuals with Lynch syndrome have a high risk of colorectal cancer, as well as endometrial, ovary, stomach, bladder, skin and brain cancers (Marra and Jiricny 2005, Bellizzi and
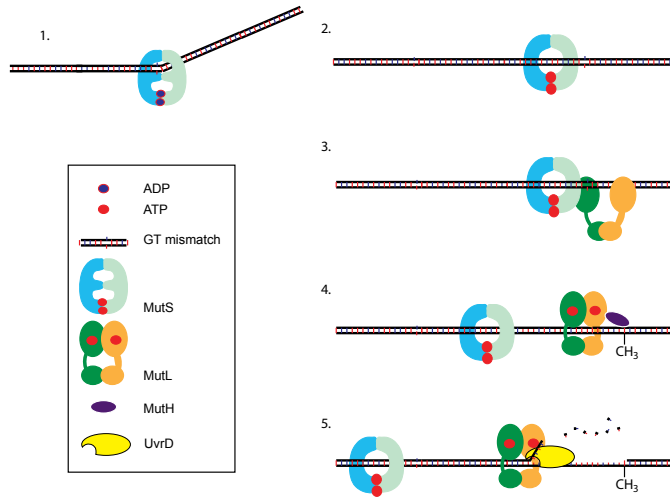
Figure 2: Abstract representation of the *E. coli* MMR pathway. 1) MutS (MutSα or MutSβ in Eukaryotes) recognizes a mismatch by bending the DNA. 2 MutS binds ATP and undergoes a conformational change where it reorganizes the mismatch binding domains. This results in the formation of a MutS sliding clamp which can diffuse laterally over the DNA. 3) The MutS sliding clamp binds MutL (MutLα in Eukaryotes). 4) MutL binds ATP and activates MutH at a transiently hemi-methylated GATC site. In Eukaryotes, the endonuclease MutH is not present, and MutLα has endonuclease activity. This endo-nuclease activity is stimulated by a pre-existing nick, or by PCNA, and is sequence independent. 5) The incised strand is unwound by UvrD, and degraded by exo-nucleases. Resynthesis by a polymerase fills in the gap and the original DNA is restored.

Frankel 2009). Furthermore, 17% of primary colorectal cancer tumors contain defects in MMR (Wimmer and Etzler 2008). This link to cancer has generated a large interest in MMR, and a large body of literature is available. However, the complexity and mechanism of mismatch repair still provide ample unsolved questions, some of which will be addressed in this thesis.

Post Replicative Mismatch Repair

Mismatch repair (MMR) and the two core proteins MutS and MutL are conserved in all kingdoms of life. A short overview of the MMR pathway, as it was discovered *E. coli* and later reconstituted with purified proteins is given in Figure 2. First, MutS recognizes a mismatch. After forming a sliding clamp, MutS can bind MutL. MutL can activate MutH, an endonuclease that can nick double stranded DNA, and thus make an entry point to remove the DNA strand containing the error. Onto this nick a helicase, UvrD, is loaded, which unwinds the DNA and makes it single stranded. Exonucleases can now degrade the nicked strand, after which a polymerase fills the gap. The minimal requirement for *E. coli* mismatch repair in vitro are MutH, MutL, and MutS proteins, DNA helicase II
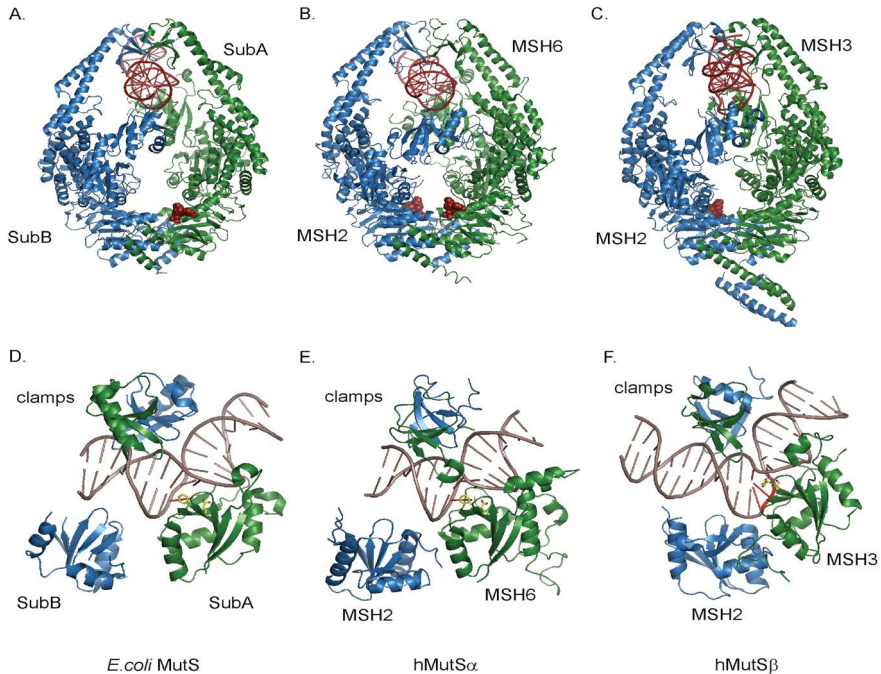
Figure 3: Comparison of the crystal structures of E. coli MutS, MutSα en MutSβ in complex with DNA. A shows a ribbon representation of the structure of E. coli MutS (PDB ID: 1E3M), where one subunit of the homodimer is green, the other subunit is blue. Nucleotide cofactors are shown in dark red. The DNA helix and nucleotide cofactors are shown in red. In B and C, the structure of human MutSα (2O8B.pdb) and MutSβ (3THY.pdb) is shown, respectively where MSH2 is blue, and MSH6 (in MutSα) and MSH3 (in MutSβ) are green. In D, E and F the mismatch binding domains of each protein is shown in detail. Both MutS and MutSα bind a G-T mismatch, which results in a ~60° kink in the DNA . MutSβ binds a 2 nucleotide insertion, and kinks the DNA ~90°. (Figure from (Jiricny 2013).)

(UvrD), single-strand DNA binding protein (SSB), DNA polymerase III holoenzyme, exonuclease I (ExoI), DNA ligase (Lahue, Au et al. 1989). The equivalent proteins used in eukaryotic MMR are given in Table 1, and the mechanistic differences between the *E. coli* and human pathway will be discussed later.

Successful mismatch repair starts with the recognition of a mismatch. The dimeric protein MutS specifically binds mismatched DNA (Su and Modrich 1986). *E. coli* MutS is active as a homodimer (Mendillo, Putnam et al. 2007). The equivalent proteins in mammalian MMR are a heterodimer consisting of MSH2-MSH6 (MutSα), which repairs mismatches and small insertion/deletion loops (IDLs), or MSH2-MSH3 (MutSβ), repairing larger IDLs (Acharya, Wilson et al. 1996). Structures of *E. coli MutS* (Lamers, Perrakis et al. 2000) and *Thermus aquaticus* MutS (Obmolova, Ban et al. 2000) show large structural overlap with their human equivalents MutSα (Warren, Pohlhaus et al. 2007) and MutSβ

(Gupta, Gellert et al. 2012) (Figure 3). Important domains, including the ATP binding domain and the DNA binding domains are present in all proteins, with subtle differences that are related to regulation by nucleotides, and the recognition of specific mismatches or IDLs (Gupta, Gellert et al. 2012).

While searching for a DNA mismatch, MutS diffuses over DNA, while following the helical turns in the DNA (Qiu, DeRocco et al. 2012). A single phenylalanine, conserved in most MutS homologs, protrudes into the minor groove of the DNA (Lamers, Perrakis et al. 2000, Natrajan, Lamers et al. 2003). Once MutS binds to a mismatch, it bends the DNA in a 60° kink (Natrajan, Lamers et al. 2003, Wang, Yang et al. 2003, Cristovao, Sisamakis et al. 2012). This bending is one of the key features of mismatch recognition by MutS, since mismatched DNA bends easier than correctly paired DNA (Gupta, Gellert et al. 2012). Also, the net free energy of unoccupied hydrogen bond donors and acceptors in the mispair itself is very low, so it is unlikely that this is enough for MutS to efficiently find a single mismatch among the correctly paired DNA (Yang 2006). A second highly conserved residue, glutamate 38, makes direct contact with one of the mismatched basepairs (Natrajan, Lamers et al. 2003). Interestingly, even though *E. coli* MutS is a homodimer, MutS is asymmetric once bound to the mismatch, and the ATP binding domains in each monomer have different affinities for ADP (Lamers, Winterwerp et al. 2003). After ATP binding, phenylalanine 36 is retracted from the DNA, and MutS forms a sliding clamp that can leave the mismatch (Acharya, Foster et al. 2003, Lebbink, Georgijevic et al. 2006). This sliding clamp can diffuse over the DNA without following the helical turns of the DNA (Gradia, Acharya et al. 1997, Qiu, DeRocco et al. 2012, Jiricny 2013). While MutS is a sliding clamp, ATP hydrolysis of MutS is inhibited (Antony and Hingorani 2003, Antony and Hingorani 2004).

E. coli	Human	Function
MutS	MutSα (MSH2-MSH6) MutSβ (MSH2-MSH3)	Mismatch recognition
MutL	MutLα (MSH1-PMS2)	Molecular matchmaker
MutH	-	Endonuclease
UvrD	-	Unwinding DNA
ExoI, RecJ, ExoVII, ExoX	EXOI	Degrading DNA
β-Clamp	PCNA	Replication machinery processivity factor
SSB	RPA	Binding single stranded DNA

Table 1: Functions of mismatch repair proteins in *E. coli* and humans.

In the presence of mismatch DNA and ATP, MutL increases the DNase footprint of MutS on the DNA (Grilley, Welsh et al. 1989, Schofield, Nayak et
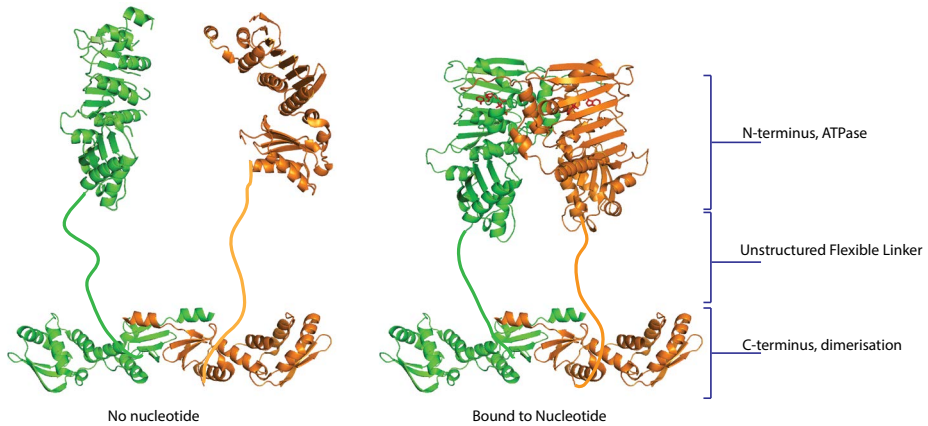
Figure 4: Model for the structure of MutL. Based on the structures of the *E. coli* MutL N-terminal domain and C-Terminal domain (PDB IDs: 1B63 and 1X9Z, respectively). The N-terminal part of MutL contains the ATPase domains, the C-terminal part is used for dimerization. When both monomers are bound to ATP, the N-termini dimerize as well, forming a central cavity within the protein. Sites important for DNA binding are situated along the border of this central cavity. (Figure created using PyMOL [http://www.pymol.org].)

al. 2001), and a ternary complex between MutS-MutL and DNA can be formed (Acharya, Foster et al. 2003, Selmane, Schofield et al. 2003, Winkler, Marx et al. 2011). Formation of the sliding clamp enables MutS to bind MutL (Grilley, Welsh et al. 1989, Habraken, Sung et al. 1998, Schofield, Nayak et al. 2001, Acharya, Foster et al. 2003). In *E. coli*, MutL consists of a homodimer, which dimerizes via its C-terminal domain (Wu, Platt et al. 2003, Guarne, Ramon-Maiques et al. 2004). In humans, several MutL homologs can be found. MutLα, a heterodimer consisting of PMS1 and MLH2 is the most prevalent one and participates in MMR. The N-terminal ATPase domain is conserved among all MutL homologs, and belongs to the GHKL (for Gyrase, Hsp90, histidine Kinase and MutL) ATPase/ kinase superfamily (Ban and Yang 1998, Dutta and Inouye 2000). A proline-rich and probably unstructured linker connects the C-terminal dimerization domain and the N-terminal ATPase containing domain (Guarne, Ramon-Maiques et al. 2004). The ATPase domains can also dimerize in the presence of ATP (Ban and Yang 1998). When MutL binds ATP, the conformational changes in the N-terminal part of MutL change its structure from and open form to a more

compact and closed form (Sacho, Kadyrov et al. 2008, Niedziela-Majka, Maluf et al. 2011), which elutes faster when run on size exclusion column (Ban, Junop et al. 1999, Guarne, Ramon-Maiques et al. 2004). These conformational changes in MutLα were also revealed in Scanning Force Miscroscopy (SFM) images (Sacho, Kadyrov et al. 2008). This closure allows for a central cavity to form in between the C-terminal and N-terminal domains (Guarne 2012)(Figure 4). This cavity is large enough for a DNA double helix to pass through, and residues essential for DNA binding are located in both the N- and C-terminal domains along this central cavity (Guarne, Ramon-Maiques et al. 2004). This leads the way for the hypothesis that MutL can clamp around DNA, much like MutS can, after binding ATP. However, this hypothesis is hard to prove directly, since DNA binding by MutL is dependent on the presence of MutS and mismatch DNA. DNA binding of MutL has been investigated previously using EMSA, nitrocellulose filter binding, Surface Plasmon Resonance spectroscopy and by following the stimulation of the ATPase activity of MutL in the presence of a DNA cofactor (Bende and Grafstrom 1991, Ban, Junop et al. 1999, Mechanic, Frankel et al. 2000, Acharya, Foster et al. 2003, Junop, Yang et al. 2003, Selmane, Schofield et al. 2003, Guarne, Ramon-Maiques et al. 2004, Robertson, Pattishall et al. 2006, Niedziela-Majka, Maluf et al. 2011).

These large conformation changes within MutL, and their dependence on a nucleotide cofactor gave rise to the molecular switch model for MutL: MutL is in a open "inactive" state, and a MutS sliding clamp can induce ATP binding in MutL, which then closes and only then is able to activate downstream factors (Ban, Junop et al. 1999, Acharya, Foster et al. 2003). *E. coli* MutL interacts with MutH (Giron-Monzon, Manelyte et al. 2004) and UvrD (Hall and Matson 1999). Strand incision in *E. coli* MMR is mediated by MutH, an endonuclease which can nick hemi-methylated GATC sites when activated by MutL (Lahue, Su et al. 1987). Strand incision by MutH and stimulation of the helicase activity of UvrD can only occur once MutL is bound to ATP (Junop, Yang et al. 2003).

MMR is different from all other DNA repair pathways in the sense that finding a mismatch is not enough to initiate successful repair. Since neither base of the mismatch is damaged, a separate signal is needed to distinguish the erroneously incorporated base from the template base. In *E. coli*, the adenosine residues in the sequence 'GATC' are methylated by Dam methyltransferase. Dam effectively methylates both strands of most GATC sites, but shortly after replication the sites are methylated only at the template strand. This leaves a short time window during which mismatch repair can utilize the hemi-methylated state of these GATC sites as a signal to distinguish the newly synthesized strand from the template strand. Once this strand discrimination signal disappears MMR loses its strand specificity (Lahue, Su et al. 1987), which is estimated to happen about one to two minutes after replication (Marinus and Casadesus 2009).

Overexpression of Dam methylase in *E. coli* results in a mutator phenotype (Herman and Modrich 1981) by shortening this window of opportunity for MMR. The incision of hemi-methylated GATC sites is done by the endonuclease MutH, which is only able to incise the unmethylated strand of a hemimethylated GATC site (Lahue, Su et al. 1987, Welsh, Lu et al. 1987). The endonuclease function of MutH is stimulated by MutL (Ban and Yang 1998, Hall and Matson 1999), and is thus an integral part of MMR in *E.coli*.

The use of hemi-methylation of GATC sites for strand discrimination is an elegant solution, but not the only way to separate the newly synthesized strand from its template. When mismatch repair was first discovered in *E. coli*, the 3 essential genes were found to be MutS, MutL and MutH (Glickman and Radman 1980). Later, it was found that most organism lack a functional homolog of MutH, and the endonuclease used in mismatch repair in eukaryotes resides in the PMS2 subunit of MutLα (Kadyrov, Dzantiev et al. 2006). The use of the transient hemi-methylation of GATC sites after replication for MMR is confined to a small group of gram negative bacteria, including *E. coli* and *Salmonella*, both often used in research (Jiricny 2013). The strand discrimination signal used in other organisms kept the MMR field busy for over two decades. Pre-existing strand discontinuities, such as at the end of Okazaki fragments, were hypothesize to facilitate MMR on the lagging strand (Claverys and Lacks 1986). Later it was shown that strand directed excision by human MMR proteins could be induced by a single pre-existing nick (Genschel, Bazemore et al. 2002), which supported this hypothesis. However, it did not provide an explanation for MMR in the leading strand, where no Okazaki fragments are available. Only recently it was shown that the orientation of PCNA could induce strand directed incision by MutLα (Pluciennik, Dzantiev et al. 2010). The strand discontinuities introduced by the repair of misincorperated ribonucleotides by RNase H2 were also shown to facilitate strand directed MMR (Ghodgaonkar, Lazzaro et al. 2013, Lujan, Williams et al. 2013).

After the strand incision, the newly synthesized strand can be removed. In *E. coli*, this is mediated by UvrD, an ATP dependent helicase which unwinds double stranded DNA from 3' to 5' as a dimer, and translocates over single stranded DNA 3' to 5' as a monomer (Matson 1986, Lee, Balci et al. 2013). UvrD is thought to have a low processivity measured to be ~45 bp in bulk experiments (Runyon, Wong et al. 1993) and 240 bp in single molecule experiments (Dessinges, Lionnet et al. 2004). Since the repair tracks of mismatch repair in *E. coli* are often larger (Wildenberg and Meselson 1975), multiple rounds of unwinding need to take place, or the processivity of UvrD is increased in the presence of MutS and MutL. The degradation of the unwound DNA can be done by several exonucleases: RecJ, ExoI, ExoVII, and ExoX, which are all partially redundant (Viswanathan and Lovett 1998). After excision of the error-containing strand is

restored by a replicative polymerase (Modrich and Lahue 1996).

Communication between the mismatch and the GATC site

One of the interesting features of mismatch repair is the separation between the mismatch and the strand discriminations site. As soon as mismatch repair was reconstituted, it was shown that a mismatch repair is bi-directional (Cooper, Lahue et al. 1993) and can effectively repair a mismatch even when the nearest GATC site is separated by over 1 kb DNA (Lahue, Au et al. 1989). After the GATC site is incised, the DNA is unwound from the GATC site in the direction of the mismatch (Grilley, Griffith et al. 1993, Dao and Modrich 1998). UvrD unwinds DNA from 3' to 5' (Matson 1986), which implies UvrD is loaded by MutL to the appropriate strand for unwinding in the direction of the mismatch. One of the earliest and most iconic models of how this communication occurs is the looping model. This model was primarily based on Electron Microscopy images (Allen, Makhov et al. 1997) and later Scanning Force Microscopy images (Jia, Bi et al. 2008, Jiang and Marszalek 2011), where MutS was found to form alpha-shaped loops in DNA. Also, the incision of a GATC site that resides on a separate DNA molecule than the mismatch was seen as supportive towards a looping model (Junop, Obmolova et al. 2001, Schofield, Nayak et al. 2001, Selmane, Schofield et al. 2003, Wang and Hays 2004), although strand incision is more efficient when both the mismatch and the GATC site are on the same strand (Pluciennik and Modrich 2007). Several different versions of the looping model have been proposed, and this model is currently the prevalent model for *E. coli* MMR, making its occurrence in textbook The Cell (Alberts, Wilson et al. 2008).

Another hypothesis is that multiple MutL molecules can form filaments on DNA, thus providing a form of communication between the mismatch and the strand discrimination signal. One example of multiple loading of MutL was found in vitro, where MutLα was observed to form short filaments on DNA (Hall, Wang et al. 2001). In vivo, the appearance of MutL foci, and the notable absence of MutS foci (Elez, Murray et al. 2010, Elez, Radman et al. 2012) supports a stoichiometry difference between MutS and MutL. This difference in stoichiometry was also observed with MutSα and MutLα in yeast (Hombauer, Campbell et al. 2011). It is however unclear whether these foci represent successful mismatch repair events. Also, a high cellular concentration of MutL is needed for its function in anti-recombination, but a decrease in cellular MutL has only a small effect on the mutation rate in *E. coli* (Elez, Radman et al. 2007).

Finally there are the models that are solely based on diffusion of MutS along the helix contour as a sliding clamp. The diffusive properties of MutS on DNA were first discovered by experiments that showed MutS can leave DNA by falling off a open DNA end (Schofield, Nayak et al. 2001, Acharya, Foster et al. 2003,

Lebbink, Georgijevic et al. 2006). More recently, several groups visualized the diffusive properties of MutS on the DNA directly in single molecule experiments (Cho, Jeong et al. 2012, Gorman, Wang et al. 2012, Qiu, DeRocco et al. 2012). In combination with this diffusion, loading of multiple MutS sliding clamps can account for a form of directionality (Jeong, Cho et al. 2011), since new clamps will be loaded at the mismatch and diffuse outward.

Other functions of mismatch repair

MMR is coupled to replication via direct interaction with the processivity factors of the replication machinery. Both MutS and MutL can bind the β-Clamp (López de Saro and O'Donnell 2001, Lopez de Saro, Marinus et al. 2006, Simmons, Davies et al. 2008), and in eukaryotic MMR MutSα, Mutβ and MutLα interact with PCNA (Kleczkowska, Marra et al. 2001, Lee and Alani 2006). However, removing errors after replication is not the only function of mismatch repair. One well-studied function of MMR is the inhibition of recombination between divergent (homeologous) DNA substrates (George and Alani 2012). An up to 1000-fold increase in interspecies conjugation between *E. coli* and *Salmonella typhimurium* in MutS, MutL, MutH and UvrD deficient recipients indicates that MMR acts as a barrier to homeologous recombination (Rayssiguier, Thaler et al. 1989, Rayssiguier, Dohet et al. 1991). In eukaryotes individual MMR proteins have roles of varying magnitude in the prevention of homeologous recombination (de Wind, Dekker et al. 1995, Selva, New et al. 1995, Datta, Adjiri et al. 1996) and the prevention of recombination events between divergent repeats (de Wind, Dekker et al. 1999). Recently, a molecular mechanism suggesting how MMR resolves mispaired DNA during homeologous recombination was proposed (Tham, Hermans et al. 2013).

Furthermore, MMR has a rather counterintuitive function in immuno-globulin mutagenesis, where it increases mutagenesis instead of decreasing it (Pena-Diaz, Bregenhorn et al. 2012). Since DNA repair often requires resynthesis of stretches of DNA by polymerases, and PCNA is recruited as a part of the replication machinery, it is not surprising that MMR proteins are often found to interact with repair proteins of other repair pathways. The list of reported interaction partners is long ((Modrich 2006) for review). Also the possible function of MMR in the DNA damage response and apoptosis is controversial (Stojic, Brun et al. 2004). These non-canonical functions of mismatch repair are still under much debate, and outside the scope of this thesis.

SCOPE

When you read the introduction to this thesis, it gives the impression that our knowledge and literature on the workings of a cell is vast, and future scientists just have to fill in the last few blanks. The devil is in the details however, and it is often harder to understand how much we still don't know than it is to see what we already figured out. And while the amount of literature available is indeed immense, our current level of understanding of complex molecules like proteins and DNA and how they can form a working cell is only the tip of the metaphorical iceberg. You could argue we know more about the Big Bang and the interactions between individual atoms than about the workings of our own cells.

This chapter (**chapter 1**) outlines the importance of DNA mismatch repair in maintaining genomic stability and cancer avoidance. As probably all cellular processes, also mismatch repair has plenty of unsolved questions and open ends. Some of the most intriguing questions in MMR revolve around strand discrimination, and the separation between the strand discrimination signal and the mismatch that need to be fixed. This thesis aims to answer two of those questions: 1) how the MMR machinery communicates between the mismatch and the strand discrimination signal and 2) which reaction steps are rate limiting for mismatch removal.

In *E.coli* the strand discrimination signal is a hemi methylated GATC sequence. In **chapter 2**, we assess how the number of GATC sites and their distance to the mismatch influence the excision of the mismatch. We show with *in vitro* assays on mismatched DNA substrates that the distance from the mismatch to the GATC site is relatively unimportant for the efficiency of strand discrimination itself. However, availability of two entry points flanking the mismatch is important for efficient excision of the error containing strand. Additionally, when multiple GATC sites are available, they will be incised rapidly, probably by a single mismatch repair complex. Also in Eukaryotic mismatch repair, the availability for multiple entry points for strand removal is important.

In **chapter 3**, we investigate the communication between the mismatch and the strand discrimination signal in more detail. We compare the strand discrimination by MMR on linear and circular DNA in a series of biochemical assays. With the results we built a quantitative model for strand incision, which incorporates all the currently known parameters from mismatch recognition by MutS, sliding clamp formation, MutL binding and MutH endonuclease activation at the GATC site. From these experiments and the resulting model we discover that diffusion by MutS along the DNA can fully explain the communication between the mismatch and the GATC site, and conformational changes in MutL are the rate limiting step of MMR before excision.

In **chapter 4** we focus on these conformational changes in MutL, and how they

influence the interaction with MutS, DNA, MutH and UvrD. It was shown MutL switches from an open conformation to a closed conformation in the presence of nucleotide cofactors. With MutL mutants that have defects in the nucleotide binding domain, we show MutL binds MutS in the open conformation, and the conformational changes are needed to bind DNA and interact with MutH and UvrD.

Finally, we look at the influence of MMR on the local GC content in the chromosomal DNA of *E.coli* in **chapter 5**. The use of GATC sites as strand discrimination signal in *E.coli* provides the unique opportunity to study the influence of MMR on the genetic makeup of E.coli. The distribution of GATC sites in E.coli is not uniform, and we show that large stretches of DNA without GATC sites (GATC deserts) have a higher GC content than the rest of the genome. This coincides with the mutational bias of the replicative polymerase PolIII, and increased polymorphism of the genes in these GATC deserts.

These new findings enable us to construct a detailed mechanistic model for DNA mismatch repair in *E. coli*. First of all, a diffusion based model of the communication between the mismatch and the strand discrimination signal fits best with our data. Whether only MutS forms a sliding clamp, or also the MutS-MutL complex or MutL by itself can form some kind of sliding clamp on the DNA remains to be determined. Furthermore, the rate limiting steps in MMR are likely conformational changes within MutS and MutL. To make sure the mismatch can be repaired in a timely fashion, multiple loading of both proteins is needed to increase efficiency, and multiple incisions can be made by a single complex. Finally, mismatch repair in *E. coli* can effectively repair mismatches when at least two hemi-methylated GATC sites are flanking the mismatch within a few kb. However, a single GATC site is not enough for efficient repair *in vivo*, which increases the mutation rate in parts of the genome that do not contain any GATC sites within ~3.5 kb.

REFERENCES

Acharya, S., P. L. Foster, P. Brooks and R. Fishel (2003). "The coordinated functions of the E. coli MutS and MutL proteins in mismatch repair." Molecular cell **12**(1): 233-246.

Acharya, S., T. Wilson, S. Gradia, M. F. Kane, S. Guerrette, G. T. Marsischky, R. Kolodner and R. Fishel (1996). "hMSH2 forms specific mispair-binding complexes with hMSH3 and hMSH6." Proc Natl Acad Sci U S A **93**(24): 13629-13634.

Alberts, B., J. H. Wilson and T. Hunt (2008). Molecular biology of the cell. New York, Garland Science.

Allen, D. J., A. Makhov, M. Grilley, J. Taylor, R. Thresher, P. Modrich and J. D. Griffith (1997). "MutS mediates heteroduplex loop formation by a translocation mechanism." The EMBO journal **16**(14): 4467-4476.

Antony, E. and M. M. Hingorani (2003). "Mismatch recognition-coupled stabilization of Msh2-Msh6 in an ATP-bound state at the initiation of DNA repair." Biochemistry **42**(25): 7682-7693.

Antony, E. and M. M. Hingorani (2004). "Asymmetric ATP binding and hydrolysis activity of the Thermus aquaticus MutS dimer is key to modulation of its interactions with mismatched DNA." Biochemistry **43**(41): 13115-13128.

Arana, M. E. and T. A. Kunkel (2010). "Mutator phenotypes due to DNA replication infidelity." Semin Cancer Biol **20**(5): 304-311.

Au, K. G., K. Welsh and P. Modrich (1992). "Initiation of methyl-directed mismatch repair." J Biol Chem **267**(17): 12142-12148.

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

Ban, C. and W. Yang (1998). "Crystal structure and ATPase activity of MutL: implications for DNA repair and mutagenesis." Cell **95**(4): 541-552.

Ban, C. and W. Yang (1998). "Structural basis for MutH activation in E.coli mismatch repair and relationship of MutH to restriction endonucleases." The EMBO journal **17**(5): 1526-1534.

Bellizzi, A. M. and W. L. Frankel (2009). "Colorectal cancer due to deficiency in DNA mismatch repair function: a review." Adv Anat Pathol **16**(6): 405-417.

Bende, S. M. and R. H. Grafstrom (1991). "The DNA binding properties of the MutL protein isolated from Escherichia coli." Nucleic Acids Res **19**(7): 1549-1555.

Bernstein, C. P., AR; Nfonsam, V; Bernstein,H (2013). DNA Damage, DNA Repair and Cancer, New Research Directions in DNA Repair, InTech, DOI: 10.5772/53919. Available from: http://www.intechopen.com/books/new-research-directions-in-dna-repair/dna-damage-dna-repair-and-cancer.

Burma, S., B. P. Chen and D. J. Chen (2006). "Role of non-homologous end joining (NHEJ) in maintaining genomic integrity." DNA Repair (Amst) **5**(9-10): 1042-1048.

Cho, W. K., C. Jeong, D. Kim, M. Chang, K. M. Song, J. Hanne, C. Ban, R. Fishel and J. B. Lee (2012). "ATP alters the diffusion mechanics of MutS on mismatched DNA." Structure **20**(7): 1264-1274.

Claverys, J. P. and S. A. Lacks (1986). "Heteroduplex deoxyribonucleic acid base mismatch repair in bacteria." Microbiol Rev **50**(2): 133-165.

Cooper, D. L., R. S. Lahue and P. Modrich (1993). "Methyl-directed mismatch repair is bidirectional." The Journal of biological chemistry **268**(16): 11823-11829.

Cristovao, M., E. Sisamakis, M. M. Hingorani, A. D. Marx, C. P. Jung, P. J. Rothwell, C. A. Seidel and P. Friedhoff (2012). "Single-molecule multiparameter fluorescence spectroscopy reveals directional MutS binding to mismatched bases in DNA." Nucleic Acids Res **40**(12): 5448-5464.

Dao, V. and P. Modrich (1998). "Mismatch-, MutS-, MutL-, and helicase II-dependent unwinding from the single-strand break of an incised heteroduplex." The Journal of biological chemistry **273**(15): 9202-9207.

Darwin, C. (1859). On the origin of species by means of natural selection. London, J. Murray.

Datta, A., A. Adjiri, L. New, G. F. Crouse and S. Jinks Robertson (1996). "Mitotic crossovers between diverged sequences are regulated by mismatch repair proteins in Saccaromyces cerevisiae." Mol Cell Biol **16**(3): 1085-1093.

de Laat, W. L., N. G. Jaspers and J. H. Hoeijmakers (1999). "Molecular mechanism of nucleotide excision repair." Genes Dev **13**(7): 768-785.

de Wind, N., M. Dekker, A. Berns, M. Radman and H. te Riele (1995). "Inactivation of the mouse Msh2 gene results in mismatch repair deficiency, methylation tolerance, hyperrecombination, and predisposition to cancer." Cell **82**(2): 321-330.

de Wind, N., M. Dekker, N. Claij, L. Jansen, Y. van Klink, M. Radman, G. Riggins, M. van der Valk, K. van't Wout and H. te Riele (1999). "HNPCC-like cancer predisposition in mice through simultaneous loss of Msh3 and Msh6 mismatch-repair protein functions." Nat Genet **23**(3): 359-362.

Deans, A. J. and S. C. West (2011). "DNA interstrand crosslink repair and cancer." Nat Rev Cancer **11**(7): 467-480.

Dessinges, M. N., T. Lionnet, X. G. Xi, D. Bensimon and V. Croquette (2004). "Single-molecule assay reveals strand switching and enhanced processivity of UvrD." Proc Natl Acad Sci U S A **101**(17): 6439-6444.

Dutta, R. and M. Inouye (2000). "GHKL, an emergent ATPase/kinase superfamily." Trends Biochem Sci **25**(1): 24-28.

Elez, M., A. W. Murray, L. J. Bi, X. E. Zhang, I. Matic and M. Radman (2010). "Seeing mutations in living cells." Curr Biol **20**(16): 1432-1437.

Elez, M., M. Radman and I. Matic (2007). "The frequency and structure of recombinant products is determined by the cellular level of MutL." Proc Natl Acad Sci U S A **104**(21): 8935-8940.

Elez, M., M. Radman and I. Matic (2012). "Stoichiometry of MutS and MutL at unrepaired mismatches in vivo suggests a mechanism of repair." Nucleic Acids Res **40**(9): 3929-3938.

Fortune, J. M., Y. I. Pavlov, C. M. Welch, E. Johansson, P. M. Burgers and T. A. Kunkel (2005). "Saccharomyces cerevisiae DNA polymerase delta: high fidelity for base substitutions but lower fidelity for single- and multi-base deletions." J Biol Chem **280**(33): 29980-29987.

Friedberg, E. C. (2003). "DNA damage and repair." Nature **421**(6921): 436-440.

Genschel, J., L. R. Bazemore and P. Modrich (2002). "Human exonuclease I is required for 5' and 3' mismatch repair." J Biol Chem **277**(15): 13302-13311.

George, C. M. and E. Alani (2012). "Multiple cellular mechanisms prevent chromosomal rearrangements involving repetitive DNA." Crit Rev Biochem Mol Biol **47**(3): 297-313.

Ghodgaonkar, M. M., F. Lazzaro, M. Olivera-Pimentel, M. Artola-Boran, P. Cejka, M. A. Reijns, A. P. Jackson, P. Plevani, M. Muzi-Falconi and J. Jiricny (2013). "Ribonucleotides misincorporated into DNA act as strand-discrimination signals in eukaryotic mismatch repair." Mol Cell **50**(3): 323-332.

Giron-Monzon, L., L. Manelyte, R. Ahrends, D. Kirsch, B. Spengler and P. Friedhoff (2004). "Mapping protein-protein interactions between MutL and MutH by cross-linking." J Biol Chem **279**(47): 49338-49345.

Glickman, B. W. and M. Radman (1980). "Escherichia coli mutator mutants deficient in methylation-instructed DNA mismatch correction." Proc Natl Acad Sci U S A **77**(2): 1063-1067.

Gorman, J., F. Wang, S. Redding, A. J. Plys, T. Fazio, S. Wind, E. E. Alani and E. C. Greene (2012).

"Single-molecule imaging reveals target-search mechanisms during DNA mismatch repair." Proceedings of the National Academy of Sciences of the United States of America **109**(45): E3074-3083.

Gradia, S., S. Acharya and R. Fishel (1997). "The human mismatch recognition complex hMSH2-hMSH6 functions as a novel molecular switch." Cell **91**(7): 995-1005.

Grilley, M., J. Griffith and P. Modrich (1993). "Bidirectional excision in methyl-directed mismatch repair." The Journal of biological chemistry **268**(16): 11830-11837.

Grilley, M., K. M. Welsh, S. S. Su and P. Modrich (1989). "Isolation and characterization of the Escherichia coli mutL gene product." J Biol Chem **264**(2): 1000-1004.

Guarne, A. (2012). "The functions of MutL in mismatch repair: the power of multitasking." Prog Mol Biol Transl Sci **110**: 41-70.

Guarne, A., S. Ramon-Maiques, E. M. Wolff, R. Ghirlando, X. Hu, J. H. Miller and W. Yang (2004). "Structure of the MutL C-terminal domain: a model of intact MutL and its roles in mismatch repair." The EMBO journal **23**(21): 4134-4145.

Guarne, A., S. Ramon-Maiques, E. M. Wolff, R. Ghirlando, X. Hu, J. H. Miller and W. Yang (2004). "Structure of the MutL C-terminal domain: a model of intact MutL and its roles in mismatch repair." EMBO J **23**(21): 4134-4145.

Gupta, S., M. Gellert and W. Yang (2012). "Mechanism of mismatch recognition revealed by human MutSbeta bound to unpaired DNA loops." Nat Struct Mol Biol **19**(1): 72-78.

Habraken, Y., P. Sung, L. Prakash and S. Prakash (1998). "ATP-dependent assembly of a ternary complex consisting of a DNA mismatch and the yeast MSH2-MSH6 and MLH1-PMS1 protein complexes." J Biol Chem **273**(16): 9837-9841.

Hall, M. C. and S. W. Matson (1999). "The Escherichia coli MutL protein physically interacts with MutH and stimulates the MutH-associated endonuclease activity." The Journal of biological chemistry **274**(3): 1306-1312.

Hall, M. C., H. Wang, D. A. Erie and T. A. Kunkel (2001). "High affinity cooperative DNA binding by the yeast Mlh1-Pms1 heterodimer." Journal of molecular biology **312**(4): 637-647.

Herman, G. E. and P. Modrich (1981). "Escherichia coli K-12 clones that overproduce dam methylase are hypermutable." Journal of bacteriology **145**(1): 644-646.

Hombauer, H., C. S. Campbell, C. E. Smith, A. Desai and R. D. Kolodner (2011). "Visualization of eukaryotic DNA mismatch repair reveals distinct recognition and repair intermediates." Cell **147**(5): 1040-1053.

Hsieh, P. and K. Yamane (2008). "DNA mismatch repair: molecular mechanism, cancer, and ageing." Mech Ageing Dev **129**(7-8): 391-407.

Iyer, R. R., A. Pluciennik, V. Burdett and P. L. Modrich (2006). "DNA mismatch repair: functions and mechanisms." Chem Rev **106**(2): 302-323.

Jeong, C., W. K. Cho, K. M. Song, C. Cook, T. Y. Yoon, C. Ban, R. Fishel and J. B. Lee (2011). "MutS switches between two fundamentally distinct clamps during mismatch repair." Nat Struct Mol Biol **18**(3): 379-385.

Jia, Y., L. Bi, F. Li, Y. Chen, C. Zhang and X. Zhang (2008). "Alpha-shaped DNA loops induced by MutS." Biochem Biophys Res Commun **372**(4): 618-622.

Jiang, Y. and P. E. Marszalek (2011). "Atomic force microscopy captures MutS tetramers initiating DNA mismatch repair." EMBO J **30**(14): 2881-2893.

Jiricny, J. (2013). "Postreplicative mismatch repair." Cold Spring Harb Perspect Biol **5**(4): a012633.

Junop, M. S., G. Obmolova, K. Rausch, P. Hsieh and W. Yang (2001). "Composite active site of an ABC ATPase: MutS uses ATP to verify mismatch recognition and authorize DNA repair." Mol Cell

7(1): 1-12.

Junop, M. S., W. Yang, P. Funchain, W. Clendenin and J. H. Miller (2003). "In vitro and in vivo studies of MutS, MutL and MutH mutants: correlation of mismatch repair and DNA recombination." <u>DNA repair</u> **2**(4): 387-405.

Kadyrov, F. A., L. Dzantiev, N. Constantin and P. Modrich (2006). "Endonucleolytic function of MutLalpha in human mismatch repair." <u>Cell</u> **126**(2): 297-308.

Kleczkowska, H. E., G. Marra, T. Lettieri and J. Jiricny (2001). "hMSH3 and hMSH6 interact with PCNA and colocalize with it to replication foci." <u>Genes Dev</u> **15**(6): 724-736.

Kramer, B., W. Kramer and H. J. Fritz (1984). "Different base/base mismatches are corrected with different efficiencies by the methyl-directed DNA mismatch-repair system of E. coli." <u>Cell</u> **38**(3): 879-887.

Lahue, R. S., K. G. Au and P. Modrich (1989). "DNA mismatch correction in a defined system." <u>Science (New York, N Y)</u> **245**(4914): 160-164.

Lahue, R. S., S. S. Su and P. Modrich (1987). "Requirement for d(GATC) sequences in Escherichia coli mutHLS mismatch correction." <u>Proceedings of the National Academy of Sciences of the United States of America</u> **84**(6): 1482-1486.

Lamers, M. H., A. Perrakis, J. H. Enzlin, H. H. Winterwerp, N. de Wind and T. K. Sixma (2000). "The crystal structure of DNA mismatch repair protein MutS binding to a G x T mismatch." <u>Nature</u> **407**(6805): 711-717.

Lamers, M. H., H. H. Winterwerp and T. K. Sixma (2003). "The alternating ATPase domains of MutS control DNA mismatch repair." <u>EMBO J</u> **22**(3): 746-756.

Lebbink, J. H., D. Georgijevic, G. Natrajan, A. Fish, H. H. Winterwerp, T. K. Sixma and N. de Wind (2006). "Dual role of MutS glutamate 38 in DNA mismatch discrimination and in the authorization of repair." <u>EMBO J</u> **25**(2): 409-419.

Lee, K. S., H. Balci, H. Jia, T. M. Lohman and T. Ha (2013). "Direct imaging of single UvrD helicase dynamics on long single-stranded DNA." <u>Nat Commun</u> **4**: 1878.

Lee, S. D. and E. Alani (2006). "Analysis of interactions between mismatch repair initiation factors and the replication processivity factor PCNA." <u>J Mol Biol</u> **355**(2): 175-184.

Li, X. and W. D. Heyer (2008). "Homologous recombination in DNA repair and DNA damage tolerance." <u>Cell Res</u> **18**(1): 99-113.

Loeb, L. A. (2011). "Human cancers express mutator phenotypes: origin, consequences and targeting." <u>Nat Rev Cancer</u> **11**(6): 450-457.

Lopez de Saro, F. J., M. G. Marinus, P. Modrich and M. O'Donnell (2006). "The beta sliding clamp binds to multiple sites within MutL and MutS." <u>J Biol Chem</u> **281**(20): 14340-14349.

López de Saro, F. J. and M. O'Donnell (2001). "Interaction of the β sliding clamp with MutS, ligase, and DNA polymerase I." <u>Proceedings of the National Academy of Sciences</u> **98**(15): 8376-8380.

Lujan, S. A., J. S. Williams, A. R. Clausen, A. B. Clark and T. A. Kunkel (2013). "Ribonucleotides are signals for mismatch repair of leading-strand replication errors." <u>Mol Cell</u> **50**(3): 437-443.

Marinus, M. G. and J. Casadesus (2009). "Roles of DNA adenine methylation in host-pathogen interactions: mismatch repair, transcriptional regulation, and more." <u>FEMS Microbiol Rev</u> **33**(3): 488-503.

Marra, G. and J. Jiricny (2005). "DNA mismatch repair and colon cancer." <u>Adv Exp Med Biol</u> **570**: 85-123.

Matson, S. W. (1986). "Escherichia coli helicase II (urvD gene product) translocates unidirectionally in a 3' to 5' direction." <u>J Biol Chem</u> **261**(22): 10169-10175.

Mechanic, L. E., B. A. Frankel and S. W. Matson (2000). "Escherichia coli MutL loads DNA helicase II onto DNA." The Journal of biological chemistry **275**(49): 38337-38346.

Mendillo, M. L., C. D. Putnam and R. D. Kolodner (2007). "Escherichia coli MutS tetramerization domain structure reveals that stable dimers but not tetramers are essential for DNA mismatch repair in vivo." J Biol Chem **282**(22): 16345-16354.

Modrich, P. (2006). "Mechanisms in eukaryotic mismatch repair." The Journal of biological chemistry **281**(41): 30305-30309.

Modrich, P. and R. Lahue (1996). "Mismatch repair in replication fidelity, genetic recombination, and cancer biology." Annu Rev Biochem **65**: 101-133.

Natrajan, G., M. H. Lamers, J. H. Enzlin, H. H. Winterwerp, A. Perrakis and T. K. Sixma (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**(16): 4814-4821.

Niedziela-Majka, A., N. K. Maluf, E. Antony and T. M. Lohman (2011). "Self-assembly of Escherichia coli MutL and its complexes with DNA." Biochemistry **50**(37): 7868-7880.

Obmolova, G., C. Ban, P. Hsieh and W. Yang (2000). "Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA." Nature **407**(6805): 703-710.

Pena-Diaz, J., S. Bregenhorn, M. Ghodgaonkar, C. Follonier, M. Artola-Boran, D. Castor, M. Lopes, A. A. Sartori and J. Jiricny (2012). "Noncanonical mismatch repair as a source of genomic instability in human cells." Mol Cell **47**(5): 669-680.

Pluciennik, A., L. Dzantiev, R. R. Iyer, N. Constantin, F. A. Kadyrov and P. Modrich (2010). "PCNA function in the activation and strand direction of MutLalpha endonuclease in mismatch repair." Proc Natl Acad Sci U S A **107**(37): 16066-16071.

Pluciennik, A. and P. Modrich (2007). "Protein roadblocks and helix discontinuities are barriers to the initiation of mismatch repair." Proceedings of the National Academy of Sciences of the United States of America **104**(31): 12709-12713.

Qiu, R., V. C. DeRocco, C. Harris, A. Sharma, M. M. Hingorani, D. A. Erie and K. R. Weninger (2012). "Large conformational changes in MutS during DNA scanning, mismatch recognition and repair signalling." The EMBO journal **31**(11): 2528-2540.

Rayssiguier, C., C. Dohet and M. Radman (1991). "Interspecific recombination between Escherichia coli and Salmonella typhimurium occurs by the RecABCD pathway." Biochimie **73**(4): 371-374.

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

Robertson, A., S. R. Pattishall and S. W. Matson (2006). "The DNA binding activity of MutL is required for methyl-directed mismatch repair in Escherichia coli." J Biol Chem **281**(13): 8399-8408.

Robertson, A. B., A. Klungland, T. Rognes and I. Leiros (2009). "DNA repair in mammalian cells: Base excision repair: the long and short of it." Cell Mol Life Sci **66**(6): 981-993.

Runyon, G. T., I. Wong and T. M. Lohman (1993). "Overexpression, purification, DNA binding, and dimerization of the Escherichia coli uvrD gene product (helicase II)." Biochemistry **32**(2): 602-612.

Sacho, E. J., F. A. Kadyrov, P. Modrich, T. A. Kunkel and D. A. Erie (2008). "Direct visualization of asymmetric adenine-nucleotide-induced conformational changes in MutL alpha." Mol Cell **29**(1): 112-121.

Schaaper, R. M. and R. L. Dunn (1991). "Spontaneous mutation in the Escherichia coli lacI gene."

Genetics **129**(2): 317-326.

Schofield, M. J., S. Nayak, T. H. Scott, C. Du and P. Hsieh (2001). "Interaction of Escherichia coli MutS and MutL at a DNA mismatch." J Biol Chem **276**(30): 28291-28299.

Selmane, T., M. J. Schofield, S. Nayak, C. Du and P. Hsieh (2003). "Formation of a DNA mismatch repair complex mediated by ATP." J Mol Biol **334**(5): 949-965.

Selva, E. M., L. New, G. F. Crouse and R. S. Lahue (1995). "Mismatch correction acts as a barrier to homeologous recombination in Saccharomyces cerevisiae." Genetics **139**(3): 1175-1188.

Simmons, L. A., B. W. Davies, A. D. Grossman and G. C. Walker (2008). "Beta clamp directs localization of mismatch repair in Bacillus subtilis." Mol Cell **29**(3): 291-301.

Stojic, L., R. Brun and J. Jiricny (2004). "Mismatch repair and DNA damage signalling." DNA Repair (Amst) **3**(8-9): 1091-1101.

Su, S. S. and P. Modrich (1986). "Escherichia coli mutS-encoded protein binds to mismatched DNA base pairs." Proc Natl Acad Sci U S A **83**(14): 5057-5061.

Tham, K. C., N. Hermans, H. H. Winterwerp, M. M. Cox, C. Wyman, R. Kanaar and J. H. Lebbink (2013). "Mismatch Repair Inhibits Homeologous Recombination via Coordinated Directional Unwinding of Trapped DNA Structures." Mol Cell **51**(3): 326-337.

Viswanathan, M. and S. T. Lovett (1998). "Single-strand DNA-specific exonucleases in Escherichia coli. Roles in repair and mutation avoidance." Genetics **149**(1): 7-16.

Wang, H. and J. B. Hays (2004). "Signaling from DNA mispairs to mismatch-repair excision sites despite intervening blockades." EMBO J **23**(10): 2126-2133.

Wang, H., Y. Yang, M. J. Schofield, C. Du, Y. Fridman, S. D. Lee, E. D. Larson, J. T. Drummond, E. Alani, P. Hsieh and D. A. Erie (2003). "DNA bending and unbending by MutS govern mismatch recognition and specificity." Proc Natl Acad Sci U S A **100**(25): 14822-14827.

Warren, J. J., T. J. Pohlhaus, A. Changela, R. R. Iyer, P. L. Modrich and L. S. Beese (2007). "Structure of the human MutSalpha DNA lesion recognition complex." Mol Cell **26**(4): 579-592.

Welsh, K. M., A. L. Lu, S. Clark and P. Modrich (1987). "Isolation and characterization of the Escherichia coli mutH gene product." J Biol Chem **262**(32): 15624-15629.

Wildenberg, J. and M. Meselson (1975). "Mismatch repair in heteroduplex DNA." Proc Natl Acad Sci U S A **72**(6): 2202-2206.

Wimmer, K. and J. Etzler (2008). "Constitutional mismatch repair-deficiency syndrome: have we so far seen only the tip of an iceberg?" Hum Genet **124**(2): 105-122.

Winkler, I., A. D. Marx, D. Lariviere, R. J. Heinze, M. Cristovao, A. Reumer, U. Curth, T. K. Sixma and P. Friedhoff (2011). "Chemical trapping of the dynamic MutS-MutL complex formed in DNA mismatch repair in Escherichia coli." J Biol Chem **286**(19): 17326-17337.

Wu, T. H., C. H. Clarke and M. G. Marinus (1990). "Specificity of Escherichia coli mutD and mutL mutator strains." Gene **87**(1): 1-5.

Wu, X., J. L. Platt and M. Cascalho (2003). "Dimerization of MLH1 and PMS2 limits nuclear localization of MutLalpha." Mol Cell Biol **23**(9): 3320-3328.

Yang, W. (2006). "Poor base stacking at DNA lesions may initiate recognition by many repair proteins." DNA Repair (Amst) **5**(6): 654-666.
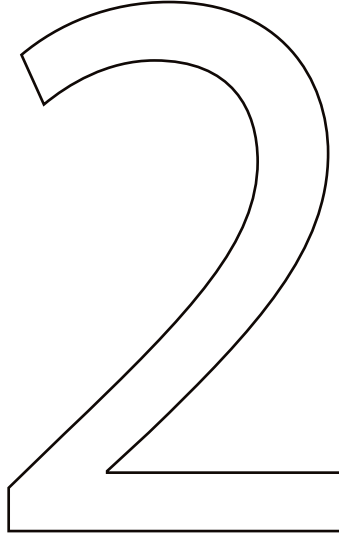
2

ON THE EFFICIENCY OF DAUGHTER STRAND NICKING AND EXCISION DURING DNA MISMATCH REPAIR

Nicolaas Hermans, Michele Cristovão, Mariela Artola-Borån, Herrie H.K. Winterwerp, Claire Wyman, Peter Friedhoff, Josef Jiricny, Roland Kanaar, Joyce H.G. Lebbink.

ABSTRACT

DNA mismatch repair (MMR) is responsible for correcting DNA replication errors, and crucial in maintaining genome stability. The main features of the DNA MMR system are conserved from *Escherichia coli* to human. In *E. coli*, MutS recognizes DNA mismatches and recruits MutL, which subsequently activates MutH, an endonuclease that is able to nick hemi-methylated DNA at the sequence GATC. Transiently hemi-methylated GATC sites provide the signal for distinguishing the newly synthesized DNA from the template strand. The efficiency of MMR in vivo depends on the number of GATC sites and the distance between mismatch and nearest GATC site. To determine at which reaction step the GATC site distribution exerts its effect on repair efficiency, we quantitatively studied the rate of nicking by the MutSLH complex and subsequent strand excision by UvrD and ExoI, while varying the number of GATC sites and their distance from a GT mismatch. When MutS loading onto the mismatch is rate-limiting, the rate of the first nick is independent of the distance between the mismatch and the nearest GATC site, and independent of the number of GATC sites. When only two GATC sites are present, one at 29 bp and one at 1042 bp from the mismatch, there is no preference for either GATC site, and both sites are nicked in rapid succession. This indicates that after recognition of the mismatch, multiple GATC sites can be incised by the same activated complex. Increasing the distance between the mismatch and a single GATC site from 60 to 1042 bp does not have a significant effect on the efficiency of excision of DNA by UvrD and ExoI. However, excision rates are 3 to 6-fold increased when two GATC sites are flanking the mismatch. Interestingly, in human nuclear extracts, the endonuclease activity of MutLα likewise contributed to increased efficiency of repair of mismatched substrates with pre-existing nicks. This indicates that in vitro, multiple nicks increase the efficiency of excision, while strand discrimination remains efficient over distances of 1 kb. These data support a model in which a single activated MMR complex facilitates efficient excision and repair by creating multiple daughter strand nicks.

INTRODUCTION

DNA mismatch repair (MMR) is crucial for maintaining genome stability. MMR deficiency results in a mutator phenotype, increased genetic instability and is associated with a higher risk for developing hereditary nonpolyposis colorectal cancer in humans (Jiricny 2013). Inactivation of MMR proteins leads to a 100 to 1000-fold increase in spontaneous mutation rates (Kunkel and Erie 2005). MMR has many functions that involve the recognition of mispaired basepairs, most notably the strand-specific correction of mismatches and small insertion/deletion loops remaining after replication (Kunkel and Erie 2005). MutS and MutL, two key proteins involved in bacterial MMR, are highly conserved during evolution, underscoring the importance of MMR (Jiricny 2013). The MMR pathway of *E. coli* was the first to be reconstituted *in vitro* with purified components (Lahue, Au et al. 1989). Three proteins are sufficient for recognizing the mismatch and targeting the newly synthesized DNA strand: MutS, MutL and MutH (Lahue, Su et al. 1987). Recognition of mismatches is mediated by MutS, after which MutS forms a sliding clamp on the DNA and is able to bind MutL (Acharya, Foster et al. 2003, Lamers, Georgijevic et al. 2004). MutL then recruits MutH, an endonuclease that can nick DNA at a GATC site (Lahue, Su et al. 1987). Subsequently, DNA helicase II (UvrD) is loaded onto the DNA by MutL and unwinds DNA with a 3' to 5' directionality, starting from the nick towards the mismatch (Dao and Modrich 1998). The exonucleases RecJ, ExoI, ExoVII and ExoX can degrade the displaced strand containing the replication error, and a second round of DNA extension and ligation fills the gap (Viswanathan, Burdett et al. 2001). In Eukaryotes, several homologues of MutS and MutL are present. The two MutS homologues that play a role in MMR are MutSα (a dimer of MSH2 and MSH6 proteins) and MutSβ (a dimer of MSH2 and MSH3 proteins), and are conserved both functionally and structurally to a large extent (for review (Jiricny 2013)). The functional homologue of MutL is MutLα (a dimer of MLH1 and PMS1), which is only partially conserved (Ban, Junop et al. 1999). Excision of the daughter strand from a nick located 5' from the mismatch is accomplished by Exonuclease I, after which resynthesis and ligation can take place.

MMR in *E. coli* differentiates between the nascent DNA strand and the template strand using the methylation status of GATC sites. Dam methylase adds a methyl group to the adenine base at these GATC sites, however, during the first few minutes following replication the nascent strand is not yet modified. MutH nicks a GATC site between the guanine and adenine base on one strand of the DNA but only if the adenine is not methylated, and thus ensures that the template DNA is left intact. The nick in the nascent strand is the entry point for the helicase UvrD. Organisms other than Gram-negative bacteria do not rely on hemi-methylation of GATC sites for strand discrimination. Instead of a separate MutH protein, these organisms harbor a latent endonuclease activity

within MutL/MutLα (Kadyrov, Dzantiev et al. 2006, Pillon, Lorenowicz et al. 2010), which can be activated by β-clamp (in bacteria) or PCNA (in eukaryotes) (Pluciennik, Dzantiev et al. 2010, Guarne 2012). The orientation of PCNA was suggested to direct the nuclease towards the daughter strand (Pluciennik, Dzantiev et al. 2010). In addition, pre-existing strand discontinuities, which can be found at the end of Okazaki fragments or are introduced by the removal of misincorporated ribonucleotides after replication, may play a role in directing MMR to the nascent strand (Ghodgaonkar, Lazzaro et al. 2013, Lujan, Williams et al. 2013).

Previous *in vitro* and *in vivo* experiments have shown that the efficiency of *E. coli* MMR not only depends on the number of hemi-methylated GATC sites, but also on the distance between the mismatch and the nearest GATC site (Lahue, Su et al. 1987, Lu 1987, Bruni, Martin et al. 1988). Increasing the distance from 1 kb to 6 kb abolishes strand-specific repair in cell extracts (Bruni, Martin et al. 1988), while MMR reconstituted from purified components loses its efficiency over distances of 2 kb (Cooper, Lahue et al. 1993). This suggests that the distance between a GATC site and the mismatch greatly influences either the strand discrimination, or the excision of the mismatched base from the DNA, both *in vitro* and *in vivo*. To distinguish between these two possibilities, we quantified the efficiency of MutH activation over several distances between the nearest GATC site and mismatch, ranging from 12 bp to 1042 bp. We did not find any correlation between the number of GATC sites and their distance to the mismatch on the one hand and nicking efficiency on the other hand, and we found that multiple nicks can be made in rapid succession. We also studied the efficiency of excision by UvrD over distances of 60 and 1042 bp from the nick towards the mismatch in the presence of the MMR machinery. Interestingly, while a single nick was sufficient to support strand unwinding and excision, the presence of two nicks flanking the mismatch significantly increased the efficiency of the reaction. Importantly, we extended our finding to mammalian MMR by demonstrating that the endonuclease activity of MutLα contributed to efficient mismatch repair on substrates with pre-existing nicks. Thus, while the molecular mechanisms for daughter strand discrimination and incision differ significantly between *E. coli* and humans, the functional outcome was the same for both organisms; the coordinated introduction of multiple daughter strand nicks increased the overall efficiency of DNA MMR.

RESULTS

Experimental system

Mismatch repair efficiency *in vitro* has been routinely scored using circular and linear DNA substrates containing a single GT mismatch (Langle-Rouault,

Maenhaut-Michel et al. 1987, Bruni, Martin et al. 1988, Su, Lahue et al. 1988, Hall and Matson 1999, Thomas, Pingoud et al. 2002). We produced hemi-methylated DNA substrates using primer extension on phagemid DNA, which allowed the incorporation of a fluorophore into the DNA (Figure 1A). The fluorophore allowed convenient and reliable quantification, circumventing the use of radioactively labeled substrates. We quantified nicking of DNA substrates with a single GT mismatch by the *E. coli* MMR system, and compared it with nicking of control homoduplex DNA (Figure 1B). Nicking was dependent on the presence of the mismatch and MutS, MutL, MutH and ATP (Figure 1B). This indicates that the Alexa647 labeled nucleotide is not recognized as a mismatch by MutS. A small amount of activity was detected using homoduplex DNA. Residual activity on the homoduplex substrate can be due to incorporation of aberrant nucleotides by the T4 polymerase during the generation of the substrates, to modifications occurring during the purification of the substrates, or to intrinsic activity of the MMR proteins on substrates without a mismatch. A small fraction of the mismatched DNA (~10%) is linearized in the full reaction (thin band just below relaxed circle), which is likely caused by incomplete methylation of the ssDNA template.

The endonuclease activity of MutH can be stimulated by MutS-MutL when the mismatch and the hemi-methylated GATC site are on separate molecules at high protein and DNA concentration (Junop, Yang et al. 2003). However, the efficiency of MutH activation is higher when both the mismatch and the GATC site are situated on the same molecule (Au, Welsh et al. 1992, Cooper, Lahue et al. 1993). Also, a strand break or a protein roadblock between the mismatch and the nearest GATC site almost abolishes nicking by MutH (Pluciennik and Modrich 2007). We checked for trans activation of MutH in our experiments by mixing unlabeled mismatched substrate with homoduplex DNA containing an Alexa647. Mismatched DNA was nicked readily, while no increased nicking of homoduplex substrate was observed in the presence of the mismatched substrate (Figure 2). This shows that in our assay, the mismatch needs to be present on the same DNA molecule as the GATC site in order to efficiently activate MutH.

Influence of the number of GATC sites and distance between mismatch and GATC sites on incision rate

First we investigated the influence of the distance between the mismatch and the nearest GATC site on the efficiency of daughter strand nicking. In the original construct pGEM-13Zf(+) a total of 16 GATC sites are present (Baerenfaller, Fischer et al. 2006). In the mismatched substrate, the nearest GATC site is located 12 bp 5' from the GT mismatch (substrate GT-12 in Figure 1A). To increase the distance
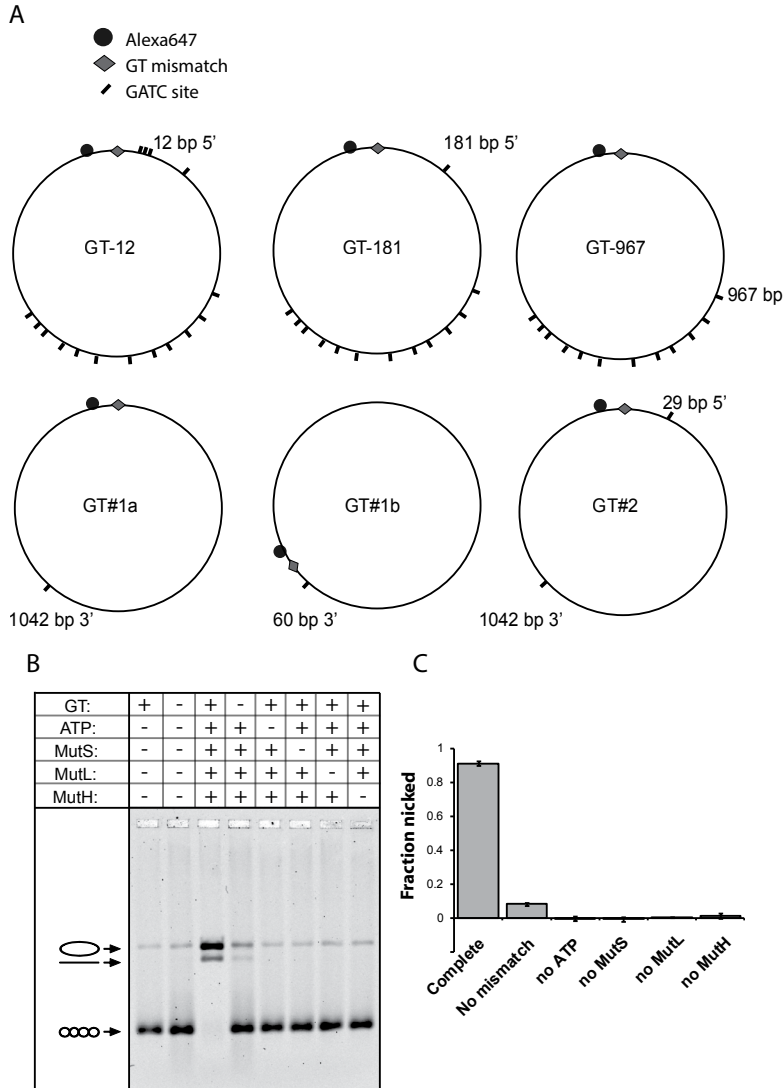
Figure 1: Experimental system. A) Map of the mismatched substrates. The size of GT-12, GT-181, GT-967, GT#1a and GT#1b is 3197 bp; GT#2 is 3199 bp. The mismatch is located at base pair 1. In the first substrate (GT-12), the nearest GATC site is 12 bp from the GT mismatch, and a total of 16 GATC sites is present. In GT-181, 3 GATC sites were removed, and the nearest GATC site is 181 bp from the mismatch. In substrate GT-967, the nearest GATC site is 967 bp from the mismatch, with 12 GATC sites in total. Substrates GT#1a and GT#1b contain only one GATC site, at 1042 or 60 bp from the mismatch, respectively. GT#2 contains 2 GATC sites. B) Agarose gel electrophoresis showing the requirements for the nicking reaction on GT-12 by E.coli MMR proteins. Covalently closed DNA (lower band) is supercoiled and migrates faster through the gel compared to nicked or linearized DNA (upper bands). C) Quantification of the nicking reactions (mean fraction nicked +/- SEM) from 3 independent experiments.
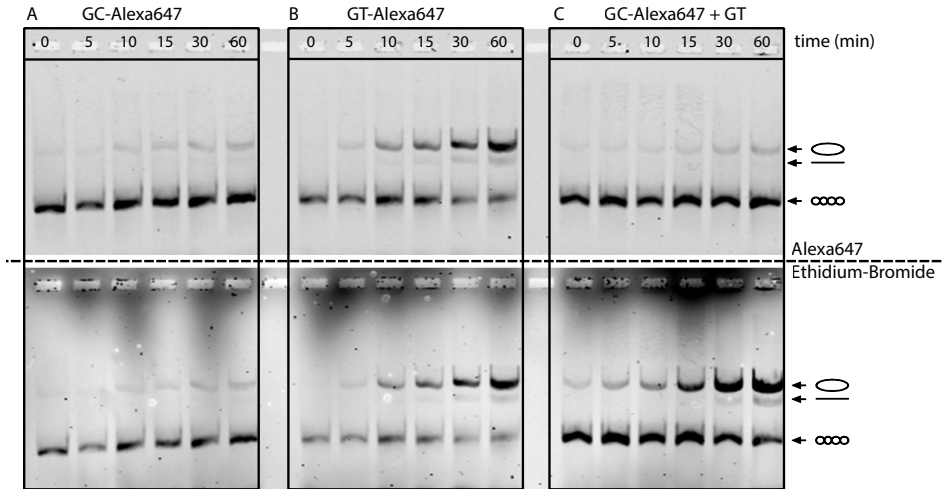
Figure 2: MutS, MutL and MutH do not nick GATC sites in *trans* in our assay conditions. Substrate containing a GT mismatch does not stimulate nicking of homoduplex by MutSLH. (A) Homoduplex substrate labeled with Alexa647 is not significantly nicked by 5 nM MutS, 5 nm MutL, 2.5 nM MutH. (B) Substrate GT-12 containing a single GT mismatch is efficiently nicked under the same assay conditions. (C) When Alexa647 labeled GC-12 and unlabeled GT-12 are mixed, the Alexa647 labeled substrate without a mismatch is not nicked beyond background levels (top panel) while the unlabeled DNA containing the GT mismatch is nicked efficiently (lower panel, ethidium bromide stain).

between the mismatch and the nearest GATC site, GATC sites at 12, 17 and 22 bp from the mismatch were removed using site directed mutagenesis, generating a substrate where the nearest GATC site is located at 181 bp 5' from the mismatch (GT-181). No significant difference in the nicking rate of the GT-181 and GT-12 was found (Figure 3A and C). We also removed the GATC site at 181 bp from the mismatch, which resulted in a substrate where the nearest GATC site is now 967 bp away from the mismatch (GT-967). This further increase in distance did not change the nicking rate significantly (Figure 3A and C). Furthermore, we tested a substrate from which all GATC sites were removed, except for a single site located at 1042 bp 3' of the mismatch (GT#1a). Again, the nicking rate did not change significantly (Figure 3B and C). From these results we conclude that the distance between the GATC site and the mismatch is not rate limiting in our assay, and increasing the number of GATC sites does not increase the efficiency of the strand discrimination. Since we use sub-saturating conditions for MutS binding to the mismatch, the rate-limiting step in these assays is MutS binding to the mismatch, and not communication between mismatch and the GATC site. However, if the distance between the mismatch and a GATC site or the number of GATC sites would influence the success rate of MutSLH, we would expect a
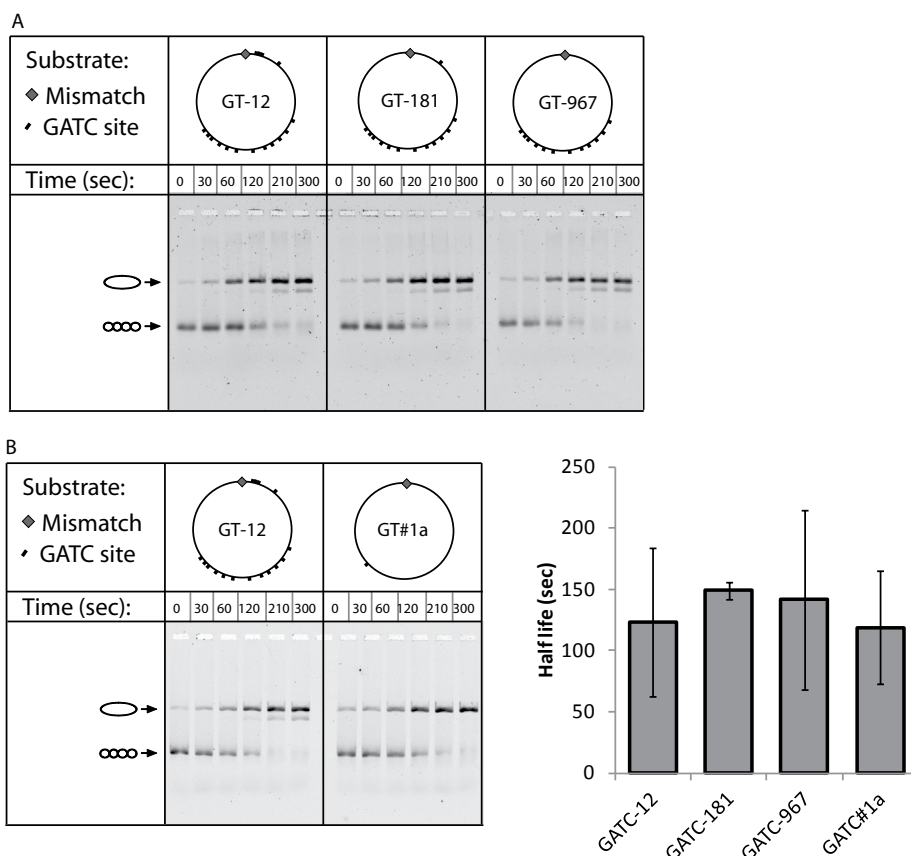
Figure 3: Effect of distance between mismatch and GATC site and number of GATC sites on the nicking of heteroduplex substrates. (A) Time course of the nicking reaction using GT-12, GT-181 and GT-967, substrates that differ in the distance between the mismatch and the nearest GATC site. (B)Time course of nicking the substrate with 16 GATC sites (GT-12) and the substrate with a single GATC site (GT#1a). (C) After quantification of the reaction products and fitting the fraction nicked with a curve describing a single exponential, no significant differences between half-lifes (mean +/- SD of three independent experiments) of the different substrates were found (the variation between independent experiments on the same substrate was larger than the variation on different substrates within one experiment).

sharp decrease in nicking efficiency when the number of GATC sites is decreased from 16 to 1. We therefore conclude that upon successful recognition of the mismatch, the resulting MutS sliding clamp is stable on the DNA, and able to efficiently promote incision of a GATC site at a distance of at least 1042 bp.

Multiple hemi-methylated GATC sites are nicked in rapid succession

To distinguish two separate nicking events on the same molecule, we constructed a substrate with only 2 GATC sites (GT#2); one GATC site at 29 bp from the mismatch, and the other GATC site at 1042 bp from the mismatch (Figure 1A). During gel electrophoresis using denaturing conditions, DNA that is nicked once can be distinguished from DNA that is nicked twice and from closed circular DNA, by quantifying the fluorescence of the Alexa647 fluorophore of each labeled single stranded (ss)DNA fragment in the gel (Figure 4A). This results in the closed circular substrate (top band) being nicked once (middle band) and twice (lower band). If we consider both these nicking events as completely independent, we can model the occurrence of the product with a single nick as an exponential decay of the closed circular DNA, and the product with two nicks as an exponential decay of the substrate with one nick. Since the substrate is not supercoiled during the reaction, introduction of a nick at the first GATC site is not expected to change the conformation of the DNA at the second GATC site, so we assume both single exponentials have the same rate constant. We will refer to this model as a "sequential" model (Figure 4C). MutH, at 50 nM final concentration and in a buffer with only 50 mM KCl, can act as a MutS and MutL-independent endonuclease (Figure 4A). As expected, the endonuclease activity of MutH in the absence of MutS and MutL can be fitted very well with this sequential model (Figure 4E), showing that MutH nicks both sites without a preference for either GATC site. This also indicates that any differences observed upon nicking of both GATC sites in the presence of MutS and MutL are not caused by differences in sequence context of the GATC sites.

At 150 mM KCl the endonuclease activity of MutH is dependent on the presence of MutS, MutL, a mismatch and ATP (Figure 1B). In this case, the nicking rates cannot be fitted with the simple sequential model, because the model over-estimates the amount of product with a single nick, and under-estimates the amount of substrate with two nicks (Figure 5). However, a model that in addition to the independent nicking events, allows the direct conversion of the closed circular substrate into the product with two nicks (Figure 4D) did fit our data (Figure 4F, lines). In this "parallel" model, a single exponential decay describes the conversion of the closed circular DNA either into an intermediate with 1 nick or directly into the product containing 2 nicks. The intermediate with 1 nick then exponentially decays into the product with 2 nicks. This model fits the data correctly when the fraction of DNA converted from closed circular to product with 2 nicks is fixed at 60% (Figure 4F, lines). This implies that on a significant portion of the substrate, the two nicking events are not independent when MutH is activated by MutS and MutL, in contrast to the nicking in the absence of MutS and MutL. Rather, both nicking events occur in rapid succession.

In the abovementioned experiment, MutS and MutL were in excess over DNA substrate (5 nM monomer, thus 2.5 nM dimer of each protein, and 0.5
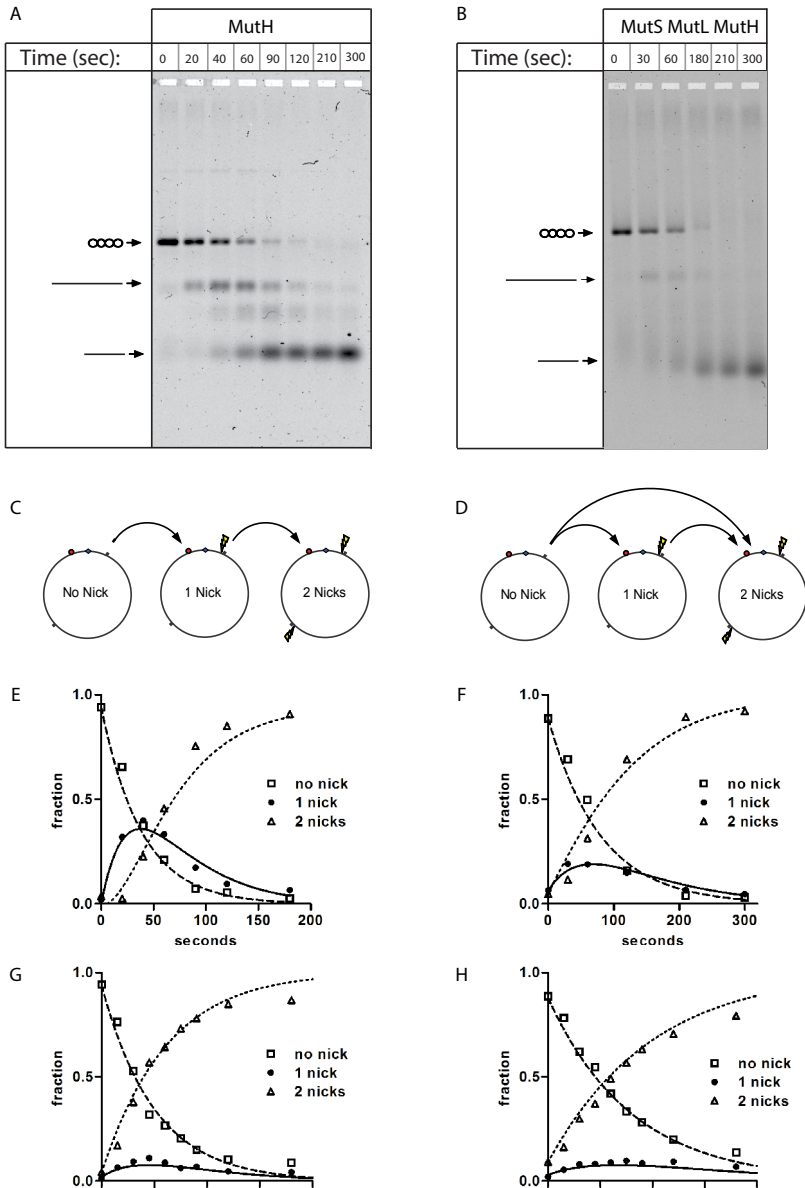
Figure 4: MutS and MutL control iterative nicking of multiple GATC sites by MutH. (A) Denaturing agarose gel showing GT#2 DNA nicked by 50 nM MutH at low salt conditions (50 mM KCl). (B) Denaturing agarose gel showing GT#2 DNA nicked by 5nM MutS, 5 nM MutL, 2.5 nM MutH at 150 mM KCl. (C) Abstract representation of the sequential model. (D) Abstract representation of the parallel model, which in addition to the sequential model allows closed circular DNA to decay into double nicked DNA at once, bypassing the single nicked product. (E-H) Quantification of the fraction of DNA containing no nick (□), 1 nick (●) and 2 nicks (Δ), obtained by nicking with MutH only (n=2, error bars omitted for clarity), fitted by the sequential model (lines). (F-H) Quantification of the mismatch dependent reaction containing MutS, MutL and MutH containing 0.5 nM DNA (Panel F, n=7), 2.5 nM (Panel G, n=3) or 5 nM (Panel H, n=3), fit with the parallel model (lines).
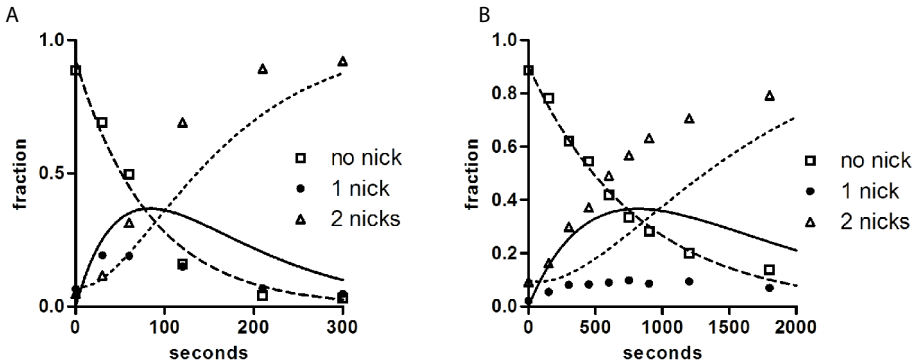
2

Figure 5: Mismatch dependent reaction fit with the sequential model. Mismatched DNA (0.5 nM in panel A, 5 nM in panel B) was nicked in the presence of 5 nM MutS, 5 nM MutL and 2.5 nM MutH and fit with the sequential model. The sequential model properly fits the initial decrease of closed circular DNA, but over-estimates the amount of product with a single nick, and under-estimates the amount of substrate with two nicks.

nM DNA molecules) and multiple loading of MutS dimers might be responsible for the observed iterative nicking. To test this we increased the concentration of mismatched substrate from 0.5 nM to 2.5 nM and 5 nM (Figure 4G and H), such that the effective concentration of MutS and MutL dimers is equimolar to or lower than the concentration of DNA molecules with a single mismatch. This will effectively reduce the frequency of multiple loading event by MutS/MutL on the DNA. At these concentrations, it took longer to nick the available substrate, which is not surprising since protein concentration is limiting. However, in both conditions the appearance of the second nick still occurred rapidly after the first, since the amount of substrate with a single nick was significantly lower than what is predicted by the sequential model (Figure 5B). In fact, on 80% of the closed circular DNA, the nicking events are linked (Figure 4G and H). We therefore conclude multiple nicks are made consecutively, after MutS bound the mismatch.

A single MutS MutL complex can both activate MutH and recruit UvrD

After MutH nicks at a GATC site, the helicase UvrD can be loaded onto the nick by MutL and unwind the DNA (Dao and Modrich 1998). To investigate the influence of UvrD on the consecutive nicking of multiple GATC sites, we added UvrD to the reaction mixture. Our reaction conditions support unwinding by UvrD, however due to the absence of SSB and exonucleases, unwound strands may rapidly reanneal. The addition of 200 nM of wild type UvrD did not significantly influence the progression of the nicking reaction (Figure 6). Likewise 200 nM of UvrD K35M, which has no ATPase activity and fails to
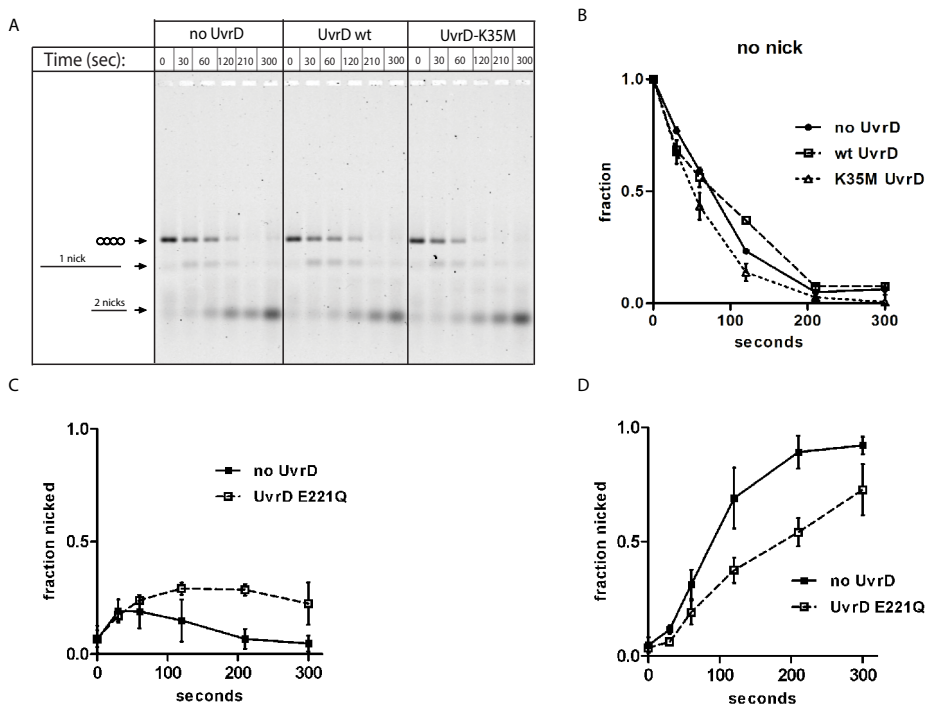
Figure 6: Wild-type UvrD and UvrD-K35M do not inhibit the second nicking event significantly. (A) Denaturing agarose gel showing covalently closed (upper band), linear single nicked (middle band) and double-nicked (lower band) GT#2 DNA. (B-D) Quantification (mean +/- SD, n=3) of the fraction of covalently closed substrate (B), single-nicked intermediate (C) and double-nicked product (D) in the absence of UvrD (●) and presence of wild type UvrD (□) or UvrD K35M (Δ). Setup was the same as with the experiment with UvrD-E221Q (Figure 6).

unwind double stranded DNA, but can still interact with MutL (George, Brosh et al. 1994), did not significantly change the reaction rate (Figure 6). UvrD E221Q, which is another variant lacking helicase activity (George, Brosh et al. 1994, Brosh and Matson 1995), had a minor influence on the disappearance of the covalently closed circle (Figure 7B, top panel). After the first nick is made, the nicked DNA becomes a substrate for UvrD (Runyon and Lohman 1989, Runyon, Bear et al. 1990). Interestingly the influence of UvrD E221Q was more obvious on the nicking of the second GATC site, resulting in a significant increase in the amount of single-nicked reaction intermediate (Figure 7C) and a decrease of substrate that was nicked twice (Figure 7D). This indicates that UvrD E221Q can specifically inhibit the second nicking event, suggesting that a single MutS MutL complex can both activate MutH and recruit UvrD.

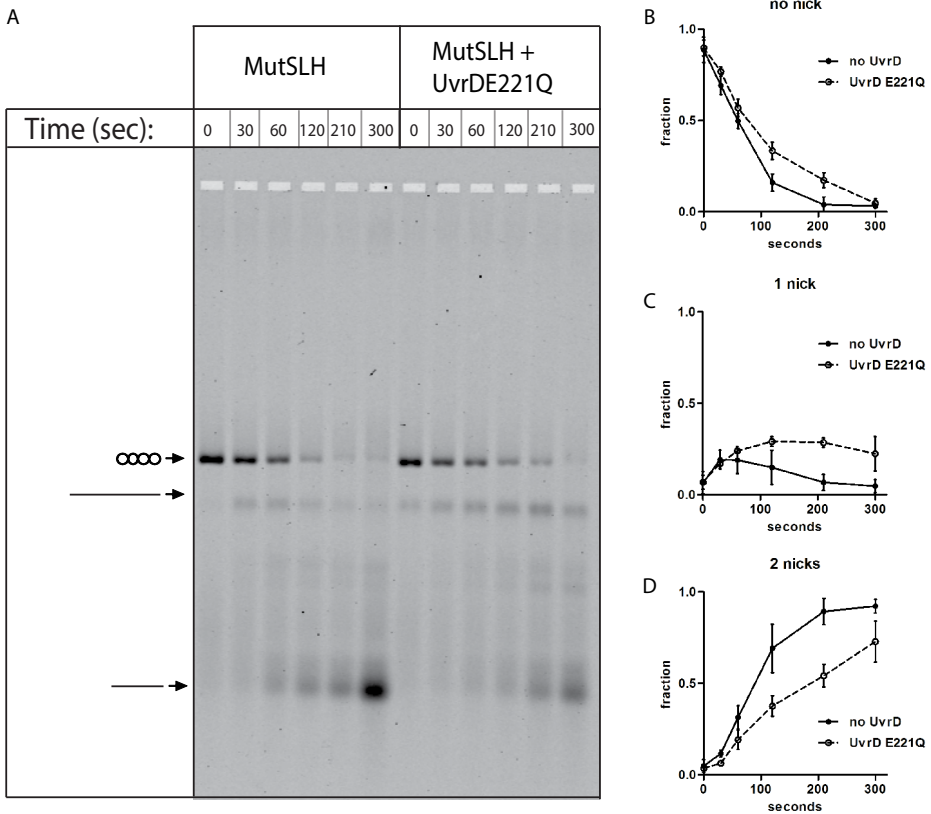**Multiple nicks increase the rate of unwinding by UvrD**

Figure 7: UvrD influences iterative nicking of multiple GATC sites by MutSLH. (A) Denaturing agarose gel with on the left nicking of substrate GT#2 by MutS (5 nM), MutL (5 nM) and MutH (2.5 nM), and on the right the same reaction supplemented with 200 nM UvrD-E221Q. In panels B, C and D the quantification of the fraction of un-nicked substrate, single-nicked intermediate and double-nicked product respectively in the absence (closed circles) and presence (open circles) of UvrD E221Q (mean +/- SD, n=6).

To address whether multiple nicks influence unwinding and excision of the DNA daughter strand, we added UvrD, SSB and ExoI to the nicking reaction (Figure 8). In the first 5 minutes the closed circular DNA (lower band) in converted to open circular DNA (upper band) due to nicking by MutS, MutL and MutH. Subsequently, UvrD can unwind the DNA and ExoI can degrade the unwound ssDNA daughter strand starting from the 3' end. Because we monitor the disappearance of the fluorescent label on the DNA substrates, the observed degradation specifically involves the region around the DNA mismatch. Because ExoI is a 3' to 5' exonuclease, only the nick at the GATC site 3' to the mismatch can be utilized for the degradation along the short path towards the mismatch.
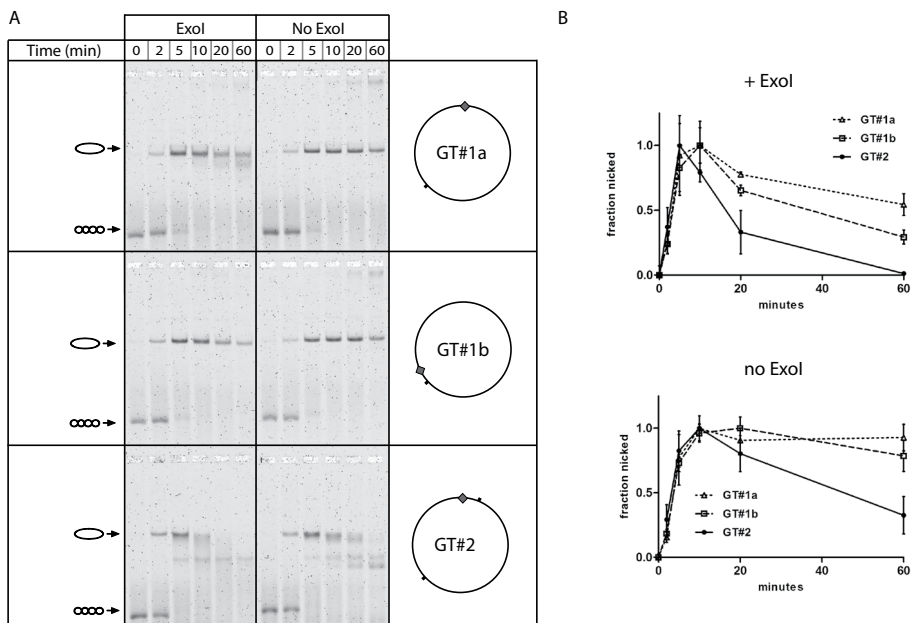
Figure 8: Daughter strand unwinding and degradation on mismatched DNA circles is accelerated by multiple GATC sites. (A) Agarose gel analysis of nicking and unwinding of mismatched DNA circles with different GATC sites by MutS, MutL, MutH, UvrD and SSB in the presence and absence of ExoI. (B) quantification of the fraction of nicked DNA in the presence (top panel) and absence (bottom panel) of ExoI for GT#1a (Δ), GT#1b (□) and GT#2 (●). Data points with error bars represent means +/- SEM of three independent experiments.

When the GATC site is close to the mismatch and the fluorophore (in GT#1b), reduced intensities of the nicked DNA is observed, but no degradation products. When the GATC site is far away from the mismatch and the fluorophore (in GT#1a), smearing can be observed under the nicked DNA. This may be caused by partially degraded substrate in which the fluorophore is still present. The substrate with two GATC sites (GT#2) is nicked with a similar rate to the substrates with only 1 GATC site, but the unwinding and subsequent degradation are significantly faster and an additional reaction product is observed that most likely corresponds to the unwound ssDNA fragment spanned by the two GATC sites that is carrying the Alexa647. The excision reaction was dependent on the presence of UvrD, and to a lesser extent on the presence of ExoI. When UvrD was absent, the DNA was not unwound or degraded (Figure 9). When ExoI was absent, the DNA was unwound but not degraded (Figure 8, compare top left and right panels). Interestingly, the strand that is displaced from GT#2 can be detected as a reaction intermediate below the nicked substrate (Figure 8A lower panel) in the absence of ExoI, while unwinding of the substrate carrying a single GATC site seems to result in the formation of reaction products that
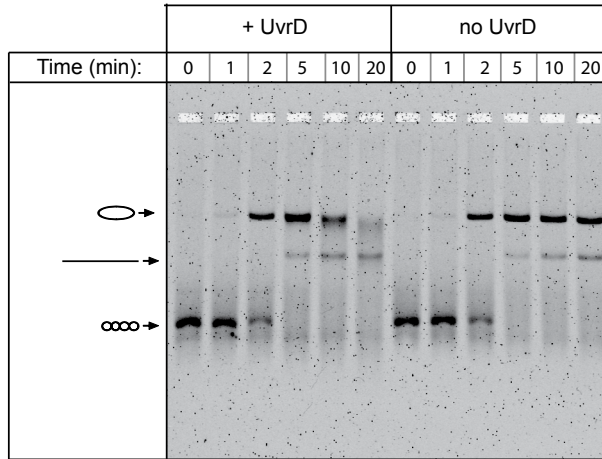
Figure 9: UvrD is needed for unwinding. Agarose gel analysis of MutSLH-induced nicking and consecutive ExoI-mediated degradation of GT#2 in the presence (left) and absence (right) of wild type UvrD. If UvrD is not present, DNA is not degraded by ExoI.

migrate slower through the gel.

When we analyzed the unwinding from a single GATC site, while varying the distance to the GT mismatch from 1042 bp (GT#1a) to 60 bp (GT#1b). Interestingly, we found only a small difference in unwinding rates between both substrates (Figure 8 top and middle panels). However, when a second GATC site was available in GT#2, the unwinding increased about 6 fold (Figure 8, bottom panel). Although UvrD activation can occur both from 3' and 5' nicks (Dao and Modrich 1998), ExoI can only degrade single-stranded DNA from 3' to 5'. Even with two GATC sites, the mismatch and the fluorophore can only be removed from a 3' nick, which is the site at 1042 bp from the mismatch. This implies that the unwinding by UvrD is much more efficient when multiple nicks are present.

Multiple nicks facilitate human mismatch repair

We assessed whether multiple nicks are also utilized in human MMR. In humans, no homologue of MutH has been found, and a latent endonuclease activity resides in MutLα (Kadyrov, Dzantiev et al. 2006). The nuclease activity of MutLα is required for repair of substrates with a single pre-existing nick residing 3' of the mismatch (Zhang, Yuan et al. 2005, Kadyrov, Dzantiev et al. 2006, Kadyrov, Genschel et al. 2009). If a single nick is located 5' of the mismatch, MutLα is not essential for repair in a reconstituted system using purified human proteins (Zhang, Yuan et al. 2005) or in human nuclear extracts (Constantin, Dzantiev et al. 2005). However, the MutLα endonuclease is essential for MMR in nuclear extracts from mouse embryonic fibroblasts (van Oers, Roa et al. 2010). Using an
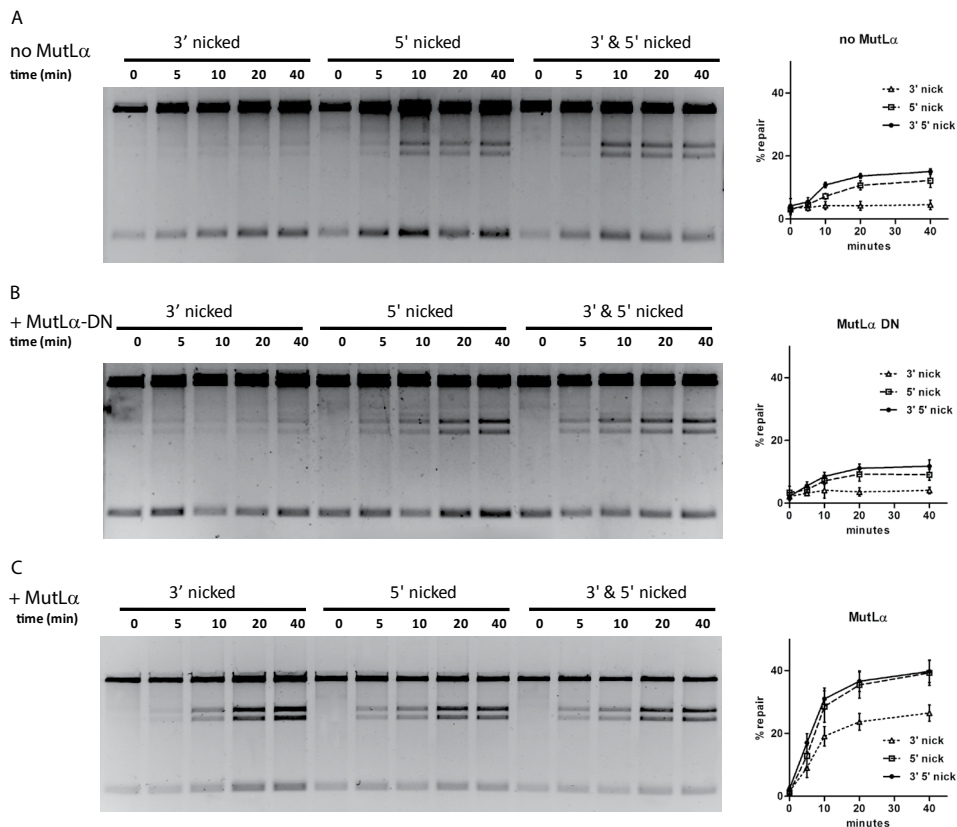
Figure 10: Function of the endonuclease domain in MutLα in 3' repair and 5' MMR. (A) 293T nuclear extracts (which lack MutLα) can repair a GT mismatch modestly only when a 5' nick is present. (B) Supplementing the extracts with MutLα-DN, which is deficient in endonuclease activity, does not restore repair of 3' nicked substrate, nor enhance repair of the 5' nicked substrate. (C) Adding wild type MutLα enables repair on the 3' substrate, and increases activity on the 5' substrate. Panels on the right contain data points with error bars representing means +/- SEM of three independent experiments.

established assay in 293T nuclear extracts lacking MutLα (Baerenfaller, Fischer et al. 2006), we tested whether multiple nicks created by MutLα are used for repair, even when a single 5' nick is available. Repair of a GT mismatch was scored on a circular substrate with a nick 306 bp 3' from the mismatch, or 363 bp 5' from the mismatch, or both 3' and 5' nicks. As expected, the GT mismatch was not repaired in the absence of MutLα when only a single 3' nick was available (Figure 10A). Supplementing the 293T nuclear extracts with MutLα restored repair (Figure 10C), while supplementing the extracts with nuclease-deficient MutLα-DN did not restore repair (Figure 10B). When a 5' nick was present, a modest amount of repair occurred in the absence of MutLα (Figure 10A). This

repair efficiency increased 3-fold when MutLα was present (Figure 10C). This is similar to results obtained by supplementing MutLα-depleted or –deficient extracts with MutLα (Genschel and Modrich 2009). However, MutLα-DN was unable to increase the efficiency of repair (Figure 10B). This indicates that also during human MMR, multiple nicks are made to increase the efficiency of repair, even if a 5' nick is present.

DISCUSSION

During DNA mismatch repair, the crucial step that determines whether a replication error will be correctly repaired or not, is daughter strand discrimination. In all organisms the signal that allows discrimination is only transiently present, providing MMR with a limited window of opportunity to initiate repair. Here we show that the MMR system of *E. coli* introduces multiple daughter strand nicks in rapid succession, which increases the efficiency of daughter strand removal and explains observed dependencies of MMR rates on GATC site distribution (Bruni, Martin et al. 1988, Lahue, Au et al. 1989). In addition we observed that multiple daughter strand nicks introduced by the MutLα endonuclease during human MMR also increase repair efficiency, implying evolutionary functional conservation despite large mechanistic differences in daughter strand discrimination.

In *E. coli* MMR, GATC sites are used as the strand discrimination signal. In the context of the genome, it is unlikely that only a single GATC site is available for initiation of repair. The distance between two GATC sites in the *E. coli* genome rarely exceeds 2 kb (Barras and Marinus 1988), and most often multiple GATC sites are present within 1 kb from the mismatch. We show that at these distances, MMR can efficiently nick GATC sites, independent of the distance between the mismatch and this GATC site. Recently, it has been shown that a MutS sliding clamp is stable on DNA for about 600 seconds, with a diffusion rate of 0.1 μm^2 s^{-1} (Cho, Jeong et al. 2012). In our assay conditions this diffusion rate is fast compared to the rate-limiting step of mismatch binding by MutS, which explains why the distance between the mismatch and the GATC site does not affect the nicking rate. We found that when two GATC sites are present on the substrate, both will be nicked in rapid succession. We think it unlikely that the binding of the first MutS increases the rate of new MutS clamps binding to the DNA, thereby decreasing the delay before the second nick occurs, because lowering the possibility for multiple loading by increasing the DNA concentration did not increase the amount of product with a single nick. Our results thus suggest that the MutS sliding clamp or the MutS-MutL complex does not necessarily dissociate after promoting the first nick, but can move on to activate the endonuclease activity of MutH at the second GATC site.

After nicking by MutH, the DNA strand containing the error is unwound by

UvrD, a helicase that is also activated by MutL (Matson 1986, Dao and Modrich 1998, Mechanic, Frankel et al. 2000). In our reaction, the covalently closed DNA substrate becomes a substrate for UvrD only after the first nick is made. When we add UvrD E221Q to the nicking reaction, the appearance of second nick is delayed. This inhibition is most pronounced with UvrD E221Q, and much less with UvrD and UvrD K35M. UvrD E221Q may form a dead-end complex with MutSL on the DNA, which can explain part of the dominant negative phenotype observed for this mutant (George, Brosh et al. 1994, Brosh and Matson 1995). A possible explanation for this observation is that the same MutSL complex that activates MutH may also load UvrD, and that this loading inhibits the complex from nicking at the second GATC site. This is another indication for the same MutSL complex being able to support multiple downstream activities.

After nicking by MutH, the DNA strand containing the error is unwound by UvrD, a helicase that is also activated by MutL (Matson 1986, Dao and Modrich 1998, Mechanic, Frankel et al. 2000). In our reaction, the covalently closed DNA substrate becomes a substrate for UvrD only after the first nick is made. When UvrD is added to the nicking reaction, the appearance of the nicks on the substrate with 2 GATC sites is unaltered. However, when we add UvrD E221Q to the nicking reaction, the appearance of second nick (but not that of the first nick) is delayed. A possible explanation for this observation is that the same MutSL complex that activates MutH may also load UvrD, and that this loading inhibits the MutSL complex from activating nicking at the second GATC site. This is another indication for the same MutSL complex being able to support multiple downstream activities. The inhibition of the second nicking event is most pronounced with the helicase-dead mutant UvrD E221Q, and not with wild type UvrD and UvrD K35M. Because UvrD E221Q fully retains the ability to bind to DNA and MutL, it may form a dead-end complex with MutSL on the DNA, which can explain part of the dominant negative phenotype observed for this mutant (Brosh and Matson 1995).

UvrD unwinds the DNA from the nick, towards the mismatch, suggesting MutL is loading UvrD onto DNA in a directional manner (Dao and Modrich 1998). Since UvrD unwinds DNA in a 3' to 5' fashion (Matson 1986), MutL loads UvrD onto either the nicked strand, or the template strand. At equimolar concentrations of MutH (2.5 nM) and UvrD (5 nM), assuming UvrD acts as a dimer (Lucius, Maluf et al. 2003, Lee, Balci et al. 2013), the DNA unwinding rate of the substrate with two GATC sites is about 10-fold slower than the nicking rate; half of the closed circular DNA is nicked within 40 seconds, while it takes five minutes for UvrD to unwind half of the nicked DNA. This is in agreement with previous observations that nicked circles are produced fast enough to be kinetically significant reaction intermediates (Au, Welsh et al. 1992) and supports the idea that the distance between mismatch and GATC site may be limiting repair

because of steps after nicking rather than the nicking itself. In two previous studies, the repair of a mismatch has been scored as a function of the number of GATC sites and the distance of these sites from the mismatch (Lahue, Su et al. 1987, Bruni, Martin et al. 1988). In both studies, increasing the distance to the nearest single GATC site to over 2000 bp impaired efficient repair. We found that the rate of daughter strand nicking was not influenced by the distance between mismatch and GATC sites. We also found that the rates for unwinding and excision were decreased only two-fold by increasing the distance between mismatch and GATC site from 60 bp to 1042 bp. The processivity of UvrD is 240 bp (Dessinges, Lionnet et al. 2004). Considering that this processivity is unchanged in the presence of MutS and MutL (Mechanic, Frankel et al. 2000), multiple loading events are needed for unwinding 1042 bp, while a single loading event might be enough to remove the mismatch in the substrate with the GATC site at 60 bp. Therefore, it is surprising that no big difference in unwinding rate were found. A possible explanation is that the first UvrD loading event is relatively slow but that as soon as the DNA is partly unwound, subsequent loading of UvrD is stimulated by the presence of ssDNA or SSB.

Previous studies also show an increase in repair efficiency if more than one GATC site is present (Bruni, Martin et al. 1988, Lahue, Au et al. 1989). We found that the rate of daughter strand nicking is not influenced by the number of GATC sites. However, we did find that unwinding and excision of a substrate with 2 GATC sites is 6-fold faster than processing of a substrate with only a single GATC site. This 6-fold increase is larger than the reported 2-fold increase in repair efficiency, but in our case the GATC sites are flanking the mismatch, while in the repair studies the GATC sites did not flank the mismatch. One reason for the increase in efficiency is that in the presence of two nicked GATC sites, UvrD can be loaded on both sides of the mismatch and can unwind the DNA from two directions simultaneously, basically doubling the unwinding rate. A second reason is that two GATC sites further reduce the chance that the DNA re-anneals before it is degraded by ExoI, because the displaced strand is no longer attached to the DNA circle as it is when only one GATC site is nicked. This is supported by the unwinding reactions in the absence of ExoI, in which the displaced strand from GT#2 is visible as a reaction intermediate while unwinding of GT#1a and GT#1b result in the formation of high-molecular weight material that might correspond to inter-molecular re-annealing of the displaced daughter strands.

E. coli and a number of additional Gram-negative bacteria rely on hemi-methylated GATC sites for strand discrimination during MMR. In most other organisms, MutH is absent and endonuclease-containing MutL/MutLα introduces a nick 5' from the mismatch to enable repair (Zhang, Yuan et al. , Kadyrov, Dzantiev et al. 2006, Jiricny 2013). However, when a 5' nick is already present, MMR can occur in the absence of MutLα, albeit with a lower efficiency when as-

sessed in MutLα-depleted or –deficient extracts (Genschel and Modrich 2003, Genschel and Modrich 2009). Here we report that MutLα does contribute to the efficiency of repair of mismatched substrates containing a pre-existing nick located 5' from the mismatch, but only if it has an intact nuclease active site. Thus, also during human MMR multiple nicks are used for efficient repair, even when a 5' nick is available. Interestingly, the presence of both the 5' and the 3' nick does not significantly increase repair compared to only a single 5' nick, unlike the effect of two nicks on unwinding efficiency by UvrD in the bacterial system. A possible explanation is the distance of 669 bp between the 3' and 5' nick, which may be too large for efficient excision.

In all organisms, replication–coupled strand discrimination signals rapidly disappear. In *E. coli*, Dam methylase fully methylates most GATC sites about one minute after replication (Marinus and Casadesus 2009). For 99.9 percent efficiency during MMR, this time window seems rather short for a relative complicated repair mechanism. In humans, existing nicks get ligated and gaps in Okazaki fragments are rapidly filled in. The rapid introduction of multiple nicks in the daughter strand enlarge the time window in which strand discrimination signals remain available for the MMR system. Furthermore, as we have shown, multiple nicks increase the rates of unwinding and excision. Thus, during *in vivo* MMR, creating multiple nicks in the vicinity of the mismatch may be very important for the fidelity of repair. Since the endonuclease of MutLα is not limited to a defined sequence, like MutH, nicks can be introduced in close proximity to the mismatch (Kadyrov, Dzantiev et al. 2006). The flip side of this increase in efficiency is that introducing multiple single-stranded breaks shortly after finding a mismatch can potentially be mutagenic in itself. Either the increase in efficiency of MMR outweighs this source of possible new mutations, or other factors play a role in limiting the activation of MutH to the direct vicinity of the mismatch. Interestingly, in *Xenopus* egg extracts, the repair tract stays relatively close to the mismatch and is not initiating from the pre-existing nick (Varlet, Canard et al. 1996), suggesting that additional nicks closer to the mismatch, likely introduced by MutLα, are used as entry points for strand excision. This indicates that despite large mechanistic differences in strand discrimination between different organisms, the common feature of MMR is the introduction of multiple nicks to enhance the efficiency of daughter strand removal and repair, implying evolutionary conservation.

MATERIALS AND METHODS

Protein purification

MutS, MutL and UvrD were purified and stored as described (Natrajan, Lamers et al. 2003, Guarne, Ramon-Maiques et al. 2004, Lebbink, Fish et al.

2010, Tham, Hermans et al. 2013). UvrD-E221Q and UvrD-K35M were derived from pET11d-UvrD (George, Brosh et al. 1994) using QuikChange (Stratagene) and purified as wild-type UvrD. MutH was purified as follows: *E. coli* BL21(DE3) cells were transformed with MutH expression plasmid pTX417 (Feng and Winkler 1995) and plated onto LB agar with 50 µg/ml carbenicillin. A colony was picked and cells were grown in LB with 50 µg/ml carbenicillin at 37°C till OD^{600} ~0.6 and induced with 1 mM isopropyl 1-thio-β-D-galactopyranoside for 4 hours. Cells were harvested and resuspended in binding buffer (25 mM Tris pH 8.0, 300 mM KCl, 10 mM imidazole, 0.2 mM DTT) with 1 mM PMSF and protease inhibitors (Roche Diagnostics) and lysed by sonication. The cleared supernatant was incubated with Talon resin (Clonetech Laboratories) for 30 minutes on ice. Beads were washed using binding buffer with 1 M KCl, and MutH was eluted with 250 mM imidazole in binding buffer. The His-tag was removed by cleavage with Thrombin protease (~5 units thrombin/mg MutH; GE Healthcare) while dialyzing against 20 mM Tris pH 8.0, 100 mM KCl, 0.2 mM DTT for 2 hours at 22 °C followed by overnight incubation at 4 °C. The mixture was brought to 20 mM imidazole, incubated with Talon beads to remove uncleaved protein, and loaded onto a heparin column equilibrated in buffer A (25 mM Tris pH 8.0, 0.1 M KCl, 1 mM DTT). MutH was eluted using a gradient of 0.1-1.0 M KCl in buffer A, pooled and diluted 2-fold with buffer A and loaded onto a MonoQ column equilibrated with buffer A. MutH was eluted using the same gradient, pooled and dialyzed overnight against 25 mM MES pH 5.5, 150 mM KCl, 1 mM DTT. MutH was loaded onto a MonoS column equilibrated with 25 mM MES pH 5.5, 0.1 M KCl, 1 mM DTT and eluted using a 0.1-1.0 M KCl gradient. Peak fractions were pooled, concentrated using Centriprep 10 and loaded onto a Superdex 75 column equilibrated with 25 mM Tris pH 8.0, 250 mM KCl, 1 mM DTT. Peak fractions were pooled, concentrated, flash frozen in 25 mM Tris pH 8.0, 250 mM KCl, 1 mM DTT, 50% glycerol and stored at -80°C. Protein concentrations were determined spectrophotometrically ($\varepsilon^{280\ nm}$ = 73,605 $M^{-1}cm^{-1}$ for MutS; 54,270 $M^{-1}cm^{-1}$ for MutL; 38,023 $M^{-1}cm^{-1}$ for MutH; 105,000 $M^{-1}cm^{-1}$ for UvrD (Runyon and Lohman 1989)). ExoI was bought from New England Biolabs and SSB protein was bought from Promega.

DNA substrates

Hemi-methylated DNA substrates were constructed by extending a primer containing a Alexa647 fluorophore on a single-stranded DNA substrate as described (Baerenfaller, Fischer et al. 2006), with the exception that closed circular DNA was purified from gel using a Promega gel purification kit. Oligonucleotides GT14 (5'-CCAGACGTCTGTC-g-ACGTTGGGAAGCT-T*-GAGTATTCTATAGTGTCACCT-3', where the g indicates the nucleotide forming the GT mismatch, and the T* is the Alexa647 labeled nucleotide) and GT28

(5'-GGTAGCTCTTCA-T*-CCGGCAAACAAACC-g-CCGCTGGTAGCG-3') were used to produce mismatched DNA. Homoduplex substrate was generated in the same way, using the primer AT14: (5'-CCAGACGTCTGTC-A-ACGTTGGGAAGCT-T*-GAGTATTCTATAGTGTCACCT-3').

Substrates for human repair

Substrates for MMR in human nuclear extracts were generated as described previously (Baerenfaller, Fischer et al. 2006). Briefly, the heteroduplexes containing a G/T mismatch within an *AclI* restriction site in the 46-bp polylinker of a pGEM13Zf(+) derivative were constructed by primer extension, using the mismatch-containing oligonucleotide (G/T: 5'-AGA CGT CTG TC**G** ACG TTG GGA AGC TTG AG-3') as primer (mispaired residue is highlighted in bold) and the single-stranded phagemid DNA carrying one NtBstNBI nicking site, 3' or 5' from the mismatch, or both as template. After primer extension, ligation, and isolation of the desired supercoiled heteroduplex substrates on CsCl gradients, substrates were nicked with Nt.BstNBI (New England Biolabs) as indicated by the manufacturer. The products were then loaded on a 1% agarose gel and visualized with GelRed.

MutH activation

Nicking activity of MutH in the presence of MutS and MutL was assayed in buffer A (25 mM Hepes KOH [pH 7.5], 150 mM KCl, 1 mg/ml BSA, 5 mM $MgCl_2$, 1 mM DTT) with 1 mM ATP and 0.5 nM DNA molecules (1.6 μM nucleotides). Unless stated otherwise the final concentration MutS and MutL were 5 nM (monomer), and the final concentration MutH was 2.5 nM. Nicking by MutH in the absence of MutS and MutL was performed in buffer A with 50 mM KCl instead of 150 mM KCl. Reactions were started by adding ATP to a mixture of all proteins and DNA. At the indicated times, 10 μL samples of the reaction were stopped with an equal volume of 20% glycerol, 1% SDS and 50 mM EDTA. Samples were analyzed on 1% agarose gels supplemented with 50 μM ethidium bromide, run in 1x TAE. To denature dsDNA circles into ssDNA fragments 10 μl samples of the reactions were stopped by adding 15 μL buffer containing 8 M Urea and 1% SDS. If indicated, samples were denatured by incubating them for 10 minutes at 85°C. Samples were run on a 1.5% agarose gel in 1x TAE and 1M urea, as described (Hegedus, Kokai et al. 2009).

UvrD activation

Unwinding reactions were performed in buffer A, in the presence of 5 nM MutS, 5 nM MutL, 2.5 nM MutH, 5 nM UvrD, 200 nM SSB and 2 mM ATP, in the

absence or presence of 0.1 units of ExoI.

Quantification

Nicking of the closed circular mismatch DNA by MutH over time was analyzed by scanning the fluorescence of the Alexa647 fluorophore using a Typhoon9100 imager (GE Healthcare) and quantifying the intensities of the bands with NIH-ImageJ. The fraction of nicked DNA was determined after background subtraction through dividing the signal of the nicked DNA by the cumulative signal. For the excision data, the signal per lane was normalized to the signal in lane 1, and the fraction of nicked DNA was determined through division of the signal of nicked DNA by the total signal.

Fitting

In the sequential model, the disappearance of the closed circular DNA (N) was fitted with a single exponential decay

$$N(t) = N_0 e^{-kt}$$

where N_0 is the amount of close circular substrate at time 0, t is the time and k is the decay constant. The amount of product with 1 nick (1N) that is produced is inversely related to this function:

$$1N(t) = N_0 - N_0 e^{-kt}$$

In the sequential model, the amount of substrate with 2 nicks was calculated as a single exponential decay of the substrate with 1 nick, using the same rate constant.

In the parallel model, the decay of the closed circular DNA (N) was partitioned into a fraction (A) decaying towards 1 nick product (1N) and a fraction (1-A) decaying towards 2 nick product (2N) with identical rate constants:

$$1N(t) = N_0 - A\, N_0 e^{-kt}$$

$$2N(t) = N_0 - (1 - A)\, N_0 e^{-kt}$$

A was varied with 0.05 increments to select for the value best fitting individual conditions. Again, the same rate constant was used for calculating the exponential decay of the 1 nick substrate into the 2 nick product.

Mismatch repair in human nuclear extracts

MMR reactions were carried out with 275 ng of each DNA substrate and 275

μg of 293T nuclear extracts (Supplemented where indicated with 0.7 pmol of wt or DN mutant MutLα) in a total volume of 70 μl in a buffer containing 20 mM Tris.HCl (pH 7.6), 5 mM MgCl2, 110 mM KCl, 1 mM glutathione, 50 μg/ml BSA, 100 μM dNTPs. The extracts were incubated at 37°C and 12.5 μl aliquots were withdrawn at the indicated time points. The reaction was stopped by adding an equal volume of 2x stop solution containing 1 mM EDTA, 3% SDS and 5 mg/ml proteinase K. The samples were incubated at 55°C for 3 hours, purified on Mini-Clean up columns (Qiagen) and subjected to restriction digests. The digested DNA was resolved on 1% agarose gels.

REFERENCES

Acharya, S., P. L. Foster, P. Brooks and R. Fishel (2003). "The coordinated functions of the E. coli MutS and MutL proteins in mismatch repair." Molecular cell 12(1): 233-246.

Au, K. G., K. Welsh and P. Modrich (1992). "Initiation of methyl-directed mismatch repair." J Biol Chem 267(17): 12142-12148.

Baerenfaller, K., F. Fischer and J. Jiricny (2006). "Characterization of the "mismatch repairosome" and its role in the processing of modified nucleosides in vitro." Methods in enzymology 408: 285-303.

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

Barras, F. and M. G. Marinus (1988). "Arrangement of Dam methylation sites (GATC) in the Escherichia coli chromosome." Nucleic Acids Res 16(20): 9821-9838.

Brosh, R. M., Jr. and S. W. Matson (1995). "Mutations in motif II of Escherichia coli DNA helicase II render the enzyme nonfunctional in both mismatch repair and excision repair with differential effects on the unwinding reaction." Journal of bacteriology 177(19): 5612-5621.

Bruni, R., D. Martin and J. Jiricny (1988). "d(GATC) sequences influence Escherichia coli mismatch repair in a distance-dependent manner from positions both upstream and downstream of the mismatch." Nucleic acids research 16(11): 4875-4890.

Cho, W. K., C. Jeong, D. Kim, M. Chang, K. M. Song, J. Hanne, C. Ban, R. Fishel and J. B. Lee (2012). "ATP alters the diffusion mechanics of MutS on mismatched DNA." Structure 20(7): 1264-1274.

Constantin, N., L. Dzantiev, F. A. Kadyrov and P. Modrich (2005). "Human mismatch repair: reconstitution of a nick-directed bidirectional reaction." J Biol Chem 280(48): 39752-39761.

Cooper, D. L., R. S. Lahue and P. Modrich (1993). "Methyl-directed mismatch repair is bidirectional." The Journal of biological chemistry 268(16): 11823-11829.

Dao, V. and P. Modrich (1998). "Mismatch-, MutS-, MutL-, and helicase II-dependent unwinding from the single-strand break of an incised heteroduplex." The Journal of biological chemistry 273(15): 9202-9207.

Dessinges, M. N., T. Lionnet, X. G. Xi, D. Bensimon and V. Croquette (2004). "Single-molecule assay reveals strand switching and enhanced processivity of UvrD." Proceedings of the National Academy of Sciences of the United States of America 101(17): 6439-6444.

Feng, G. and M. E. Winkler (1995). "Single-step purifications of His6-MutH, His6-MutL and His6-MutS repair proteins of escherichia coli K-12." Biotechniques 19(6): 956-965.

Genschel, J. and P. Modrich (2003). "Mechanism of 5'-directed excision in human mismatch repair." Mol Cell 12(5): 1077-1086.

Genschel, J. and P. Modrich (2009). "Functions of MutLalpha, replication protein A (RPA), and HMGB1 in 5'-directed mismatch repair." J Biol Chem 284(32): 21536-21544.

George, J. W., R. M. Brosh, Jr. and S. W. Matson (1994). "A dominant negative allele of the Escherichia coli uvrD gene encoding DNA helicase II. A biochemical and genetic characterization." Journal of molecular biology 235(2): 424-435.

Ghodgaonkar, M. M., F. Lazzaro, M. Olivera-Pimentel, M. Artola-Boran, P. Cejka, M. A. Reijns, A. P. Jackson, P. Plevani, M. Muzi-Falconi and J. Jiricny (2013). "Ribonucleotides misincorporated into DNA act as strand-discrimination signals in eukaryotic mismatch repair." Mol Cell 50(3): 323-332.

Guarne, A. (2012). "The functions of MutL in mismatch repair: the power of multitasking." Prog Mol Biol Transl Sci 110: 41-70.

53

Guarne, A., S. Ramon-Maiques, E. M. Wolff, R. Ghirlando, X. Hu, J. H. Miller and W. Yang (2004). "Structure of the MutL C-terminal domain: a model of intact MutL and its roles in mismatch repair." The EMBO journal 23(21): 4134-4145.

Hall, M. C. and S. W. Matson (1999). "The Escherichia coli MutL protein physically interacts with MutH and stimulates the MutH-associated endonuclease activity." The Journal of biological chemistry 274(3): 1306-1312.

Hegedus, E., E. Kokai, A. Kotlyar, V. Dombradi and G. Szabo (2009). "Separation of 1-23-kb complementary DNA strands by urea-agarose gel electrophoresis." Nucleic acids research 37(17): e112.

Jiricny, J. (2013). "Postreplicative mismatch repair." Cold Spring Harb Perspect Biol 5(4): a012633.

Junop, M. S., W. Yang, P. Funchain, W. Clendenin and J. H. Miller (2003). "In vitro and in vivo studies of MutS, MutL and MutH mutants: correlation of mismatch repair and DNA recombination." DNA repair 2(4): 387-405.

Kadyrov, F. A., L. Dzantiev, N. Constantin and P. Modrich (2006). "Endonucleolytic function of MutLalpha in human mismatch repair." Cell 126(2): 297-308.

Kadyrov, F. A., J. Genschel, Y. Fang, E. Penland, W. Edelmann and P. Modrich (2009). "A possible mechanism for exonuclease 1-independent eukaryotic mismatch repair." Proc Natl Acad Sci U S A 106(21): 8495-8500.

Kunkel, T. A. and D. A. Erie (2005). "DNA mismatch repair." Annual review of biochemistry 74: 681-710.

Lahue, R. S., K. G. Au and P. Modrich (1989). "DNA mismatch correction in a defined system." Science (New York, N Y) 245(4914): 160-164.

Lahue, R. S., S. S. Su and P. Modrich (1987). "Requirement for d(GATC) sequences in Escherichia coli mutHLS mismatch correction." Proceedings of the National Academy of Sciences of the United States of America 84(6): 1482-1486.

Lamers, M. H., D. Georgijevic, J. H. Lebbink, H. H. Winterwerp, B. Agianian, N. de Wind and T. K. Sixma (2004). "ATP increases the affinity between MutS ATPase domains. Implications for ATP hydrolysis and conformational changes." The Journal of biological chemistry 279(42): 43879-43885.

Langle-Rouault, F., G. Maenhaut-Michel and M. Radman (1987). "GATC sequences, DNA nicks and the MutH function in Escherichia coli mismatch repair." The EMBO journal 6(4): 1121-1127.

Lebbink, J. H., A. Fish, A. Reumer, G. Natrajan, H. H. Winterwerp and T. K. Sixma (2010). "Magnesium coordination controls the molecular switch function of DNA mismatch repair protein MutS." The Journal of biological chemistry 285(17): 13131-13141.

Lee, K. S., H. Balci, H. Jia, T. M. Lohman and T. Ha (2013). "Direct imaging of single UvrD helicase dynamics on long single-stranded DNA." Nat Commun 4: 1878.

Lu, A. L. (1987). "Influence of GATC sequences on Escherichia coli DNA mismatch repair in vitro." J Bacteriol 169(3): 1254-1259.

Lucius, A. L., N. K. Maluf, C. J. Fischer and T. M. Lohman (2003). "General methods for analysis of sequential "n-step" kinetic mechanisms: application to single turnover kinetics of helicase-catalyzed DNA unwinding." Biophys J 85(4): 2224-2239.

Lujan, S. A., J. S. Williams, A. R. Clausen, A. B. Clark and T. A. Kunkel (2013). "Ribonucleotides are signals for mismatch repair of leading-strand replication errors." Mol Cell 50(3): 437-443.

Marinus, M. G. and J. Casadesus (2009). "Roles of DNA adenine methylation in host-pathogen interactions: mismatch repair, transcriptional regulation, and more." FEMS Microbiol Rev 33(3): 488-503.

Matson, S. W. (1986). "Escherichia coli helicase II (urvD gene product) translocates unidirectionally in a 3' to 5' direction." J Biol Chem 261(22): 10169-10175.

Mechanic, L. E., B. A. Frankel and S. W. Matson (2000). "Escherichia coli MutL loads DNA helicase II onto DNA." The Journal of biological chemistry 275(49): 38337-38346.

Natrajan, G., M. H. Lamers, J. H. Enzlin, H. H. Winterwerp, A. Perrakis and T. K. Sixma (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(16): 4814-4821.

Pillon, M. C., J. J. Lorenowicz, M. Uckelmann, A. D. Klocko, R. R. Mitchell, Y. S. Chung, P. Modrich, G. C. Walker, L. A. Simmons, P. Friedhoff and A. Guarne (2010). "Structure of the endonuclease domain of MutL: unlicensed to cut." Molecular cell 39(1): 145-151.

Pluciennik, A., L. Dzantiev, R. R. Iyer, N. Constantin, F. A. Kadyrov and P. Modrich (2010). "PCNA function in the activation and strand direction of MutLalpha endonuclease in mismatch repair." Proc Natl Acad Sci U S A 107(37): 16066-16071.

Pluciennik, A. and P. Modrich (2007). "Protein roadblocks and helix discontinuities are barriers to the initiation of mismatch repair." Proceedings of the National Academy of Sciences of the United States of America 104(31): 12709-12713.

Runyon, G. T., D. G. Bear and T. M. Lohman (1990). "Escherichia coli helicase II (UvrD) protein initiates DNA unwinding at nicks and blunt ends." Proc Natl Acad Sci U S A 87(16): 6383-6387.

Runyon, G. T. and T. M. Lohman (1989). "Escherichia coli helicase II (uvrD) protein can completely unwind fully duplex linear and nicked circular DNA." J Biol Chem 264(29): 17502-17512.

Su, S. S., R. S. Lahue, K. G. Au and P. Modrich (1988). "Mispair specificity of methyl-directed DNA mismatch correction in vitro." The Journal of biological chemistry 263(14): 6829-6835.

Tham, K. C., N. Hermans, H. H. Winterwerp, M. M. Cox, C. Wyman, R. Kanaar and J. H. Lebbink (2013). "Mismatch Repair Inhibits Homeologous Recombination via Coordinated Directional Unwinding of Trapped DNA Structures." Mol Cell 51(3): 326-337.

Thomas, E., A. Pingoud and P. Friedhoff (2002). "An efficient method for the preparation of long heteroduplex DNA as substrate for mismatch repair by the Escherichia coli MutHLS system." Biol Chem 383(9): 1459-1462.

van Oers, J. M., S. Roa, U. Werling, Y. Liu, J. Genschel, H. Hou, Jr., R. S. Sellers, P. Modrich, M. D. Scharff and W. Edelmann (2010). "PMS2 endonuclease activity has distinct biological functions and is essential for genome maintenance." Proc Natl Acad Sci U S A 107(30): 13384-13389.

Varlet, I., B. Canard, P. Brooks, G. Cerovic and M. Radman (1996). "Mismatch repair in Xenopus egg extracts: DNA strand breaks act as signals rather than excision points." Proc Natl Acad Sci U S A 93(19): 10156-10161.

Viswanathan, M., V. Burdett, C. Baitinger, P. Modrich and S. T. Lovett (2001). "Redundant exonuclease involvement in Escherichia coli methyl-directed mismatch repair." J Biol Chem 276(33): 31053-31058.

Zhang, Y., F. Yuan, S. R. Presnell, K. Tian, Y. Gao, A. E. Tomkinson, L. Gu and G. M. Li (2005). "Reconstitution of 5'-directed human mismatch repair in a purified system." Cell 122(5): 693-705.

3

A QUANTITATIVE MODEL FOR DAUGHTER STRAND DISCRIMINATION REVEALS RATE-LIMITING STEPS DURING DNA MISMATCH REPAIR

Nicolaas Hermans, Herrie H.K. Winterwerp, Titia K. Sixma, Roland Kanaar, Peter Friedhoff, Kim Sneppen, Joyce H.G. Lebbink

ABSTRACT

DNA mismatch repair (MMR) is responsible for correcting DNA replication errors, and is crucial in maintaining genomic stability. In Escherichia coli, MMR is initiated by MutS upon recognition of a DNA mismatch, resulting in ATP-dependent recruitment of MutL and activation of MutH. MutH is an endonuclease that is able to nick hemi-methylated DNA at a GATC motif. Transiently hemi-methylated GATC sites provide the signal at which the newly synthesized DNA can be distinguished from the template strand. We quantitatively analyzed the dependence of strand incision by MutSLH on the number and position of GATC sites in circular and linear substrates containing a single DNA mismatch. Using Monte Carlo simulations, we found that a simple diffusive model is able to predict many features of our data, which is a good indication that the incision complex uses a random walk to communicate between the mismatch and GATC sites. Combined with order-of-addition experiments, the simulations suggest that conformational changes in MutL are rate limiting for strand incision. The fast diffusion rate and stability of the MutS sliding clamp, combined with the relatively slow ATP-dependent conformational changes within MutL explain why strand discrimination is independent of the distance between the mismatch and the nearest GATC site. Furthermore, nicking efficiency is increased by multiple loading of MutS sliding clamps, increasing the chance that strand discrimination occurs before the Dam methylase fully restores methylation at GATC sites, and the signal required to discriminate between the parental and daughter strand disappears.

INTRODUCTION

Successful transmission of genetic information is dependent on accurate replication of DNA, and on the removal of errors by DNA repair pathways. DNA mismatch repair (MMR) increases the fidelity of replication by removing base pair mismatches and small insertions and deletions from DNA. Inactivation of MMR results in an elevated mutation rate, and an increased risk to develop certain forms of cancer in mammals (Jiricny 2013). The molecular mechanism of MMR has been studied extensively *in vitro*, and the *E. coli* MMR has been reconstituted over two decades ago (Lahue, Au et al. 1989). MutS and MutL, the two core components of MMR in *E. coli*, are conserved in all kingdoms of life (Jiricny 2013).

The *E. coli* MutS homodimer can recognize base-base mismatches and small insertion and deletion loops left by the replication machinery. While searching for a DNA mismatch, MutS diffuses with a corkscrew-like movement over the DNA (Jeong, Cho et al. 2011, Cho, Jeong et al. 2012), inserting its N-terminal domains into the minor groove of the DNA to scan for mismatches (Lamers, Perrakis et al. 2000, Obmolova, Ban et al.). While diffusing over the DNA, MutS can bind ATP, but will readily hydrolyze it (Heo, Ku et al. 2009). Upon binding a mismatch, the MutS dimer binds in an asymmetric manner with only one of the two subunits making contact with the mismatch (Lamers, Perrakis et al. 2000, Obmolova, Ban et al. 2000). This induces rapid exchange of ADP for ATP, but in this case hydrolysis is inhibited (Gradia, Acharya et al. 1997, Bjornson, Allen et al. 2000, Acharya, Foster et al. 2003, Lebbink, Georgijevic et al. 2006), and a conformational change within the MutS dimer removes the mismatch-binding domain from the minor groove, while retaining a ring-like structure (Lamers, Georgijevic et al. 2004). This conformational change "switches" MutS to a sliding clamp (Acharya, Foster et al. 2003) and allows it to release the mismatch and diffuse laterally over the DNA, without following the helical turns of the DNA (Gradia, Subramanian et al. 1999, Qiu, DeRocco et al. 2012). This sliding clamp can bind MutL (Acharya, Foster et al. 2003).

MutL is also a homo-dimer in solution, dimerizing via an interface on the C-terminus (Ban, Junop et al. 1999, Guarne, Ramon-Maiques et al. 2004). A nucleotide-binding domain resides in the N-terminal part of the protein (Ban, Junop et al. 1999). When both nucleotide binding domains are bound to ATP, these domains also dimerize, resulting in a ring-like MutL structure with a cavity in the middle large enough to fit double-stranded (ds)DNA. Several residues important for DNA binding are located within this cavity (Guarne, Ramon-Maiques et al. 2004), but there is no detailed structural information available for DNA binding to MutL. After the discovery of the nucleotide-binding domains in MutL, it was suggested to act as a molecular switch in MMR (Ban and

Yang 1998, Ban, Junop et al. 1999, Acharya, Foster et al. 2003). ATP binding by MutL results in the conformational change that is required to activate the endonuclease MutH (Junop, Yang et al. 2003).

MutH is an endonuclease that is able to nick DNA at the unmethylated strand of a hemi-methylated GATC motif within dsDNA (Welsh, Lu et al. 1987). This nick is used as an entry point for the helicase UvrD, which can unwind the DNA (Lahue, Su et al. 1987). After specific exonucleases degrade the ssDNA containing the incorrect base, a second round of DNA polymerization fills the gap. GATC sites within the newly synthesized DNA are methylated by Dam methylase within 1 minute after replication (Marinus and Casadesus 2009). Thus, MutH has a narrow time window to locate and incise a GATC site in proximity to a DNA mismatch before the Dam enzyme methylates the newly synthesized DNA strand.

Several models are proposed to explain the communication between the mismatch and the GATC site. The most iconic models involve DNA loop formation by MutS and MutL. These models are primarily based on Electron Microscopy (EM) and Scanning Force Microscopy (SFM) images, in which MMR proteins are shown to form loops in DNA (Allen, Makhov et al. 1997, Jia, Bi et al. 2008, Jiang and Marszalek 2011). Additionally, MutS and MutL are able to activate MutH when the mismatch and the GATC site reside on separate oligonucleotides (Junop, Yang et al. 2003, Selmane, Schofield et al. 2003), albeit less efficiently than when they reside on the same molecule. A second model that accounts for the communication between the mismatch and the strand discrimination signal is based on the observation of short MutLα filaments on DNA (Hall, Wang et al. 2001). This could indicate multiple loading of MutL, thereby forming a structure resembling a protein filament that can span the gap between mismatch and strand discrimination signal. Consistent with the multiple loading of MutL molecules is the local accumulation of GFP tagged MutL *in vivo* in foci, whereas GFP tagged MutS does not accumulate in foci (Elez, Murray et al. 2010, Elez, Radman et al. 2012) Also the absence of MutSα in MutLα foci (Hombauer, Campbell et al. 2011) supports a stoichiometry difference between MutS and MutL. Last, but not least, there is the molecular switch model, in which MutS is activated upon mismatch recognition by nucleotide exchange (analogous to the G-protein switches) resulting in diffusion of ATP-bound MutS along the DNA as a sliding clamp (Gradia, Acharya et al. 1997, Gradia, Subramanian et al. 1999, Acharya, Foster et al. 2003). The diffusion of MutS as a sliding clamp is supported by experiments that show MutS sliding clamps dissociate from DNA by falling off from a free end {Schofield 2001; Acharya 2003; Lebbink 2006}. More recently, several groups showed the diffusive properties of MutS on the DNA directly in single molecule experiments (Cho, Jeong et al. 2012, Gorman, Wang et al. 2012, Qiu, DeRocco et al. 2012). However, these experiments do

not address this communication within a functional context of daughter strand incision or repair, and thus the importance of observed structures or isolated behavior of specific MMR proteins for the overall reaction remains unresolved.

Here we present a quantitative analysis of incision on hemi-methylated circular and linear DNA substrates by MutS, MutL and MutH, and show that a simple diffusive model can predict most aspects of the obtained results. Furthermore, we determine the timeframe in which the individual steps occurring between mismatch recognition and strand incision take place. We use parameters obtained from literature and from our own data to quantitatively simulate strand incision by MutS, MutL and MutH, which puts individual reactions steps into perspective of the complete MMR mechanism, and provides a framework for a more defined mechanistic model of this complex system.

RESULTS

Strand incision by MutSLH at a hemi-methylated GATC site is more efficient on circular than on linear DNA

Strand incision is more efficient on circular substrates than on linear substrates with the strand discontinuity located in either the short or the long path between the DNA mismatch and the unique GATC site (Au, Welsh et al. 1992). To address whether this low activity on linear DNA results from the presence of DNA ends or from the absence of a GATC site in one arm of the linear substrate, we compared incision on circular and linear versions of a substrate in which the mismatch is flanked by hemi-methylated GATC sites on both sides of the mismatch (Figure 1A). Previously, we established that our assay is dependent on the presence of a mismatch, and the proteins MutS, MutL and MutH and ATP (Chapter 2). Because the DNA is labeled with a fluorophore on the unmethylated strand, incision of both individual GATC sites can be monitored when the DNA is linearized (either before or after the nicking reaction) and the labeled nicked fragments are separated from the circular strand under denaturing conditions in agarose gel electrophoresis. We incubated circular DNA with 5 nM MutS, 5 nM MutL and 2.5 nM MutH, and monitored incision of the two GATC sites (Figure 1B). In the shown agarose gel, time courses are shown with circular DNA (lane 1-6), and with linear DNA with increasing protein concentrations (lane 7-24). All the circular DNA is nicked within 300 seconds in the presence of 5 nM MutS, 5 nM MutL, 2.5 nM MutH (lane 1-6). Interestingly, despite the presence of GATC sequences on both sides of the mismatch, there is almost no activity of MutSLH on linear DNA under the same conditions (less than 50% of the linear DNA is nicked after 5 minutes (lane 7-12). Likely, MutS reaches and slides off the end before it can activate MutL and MutH at these protein concentrations on linear DNA. When the concentration of MutS, MutL and MutH is increased 4 fold, the
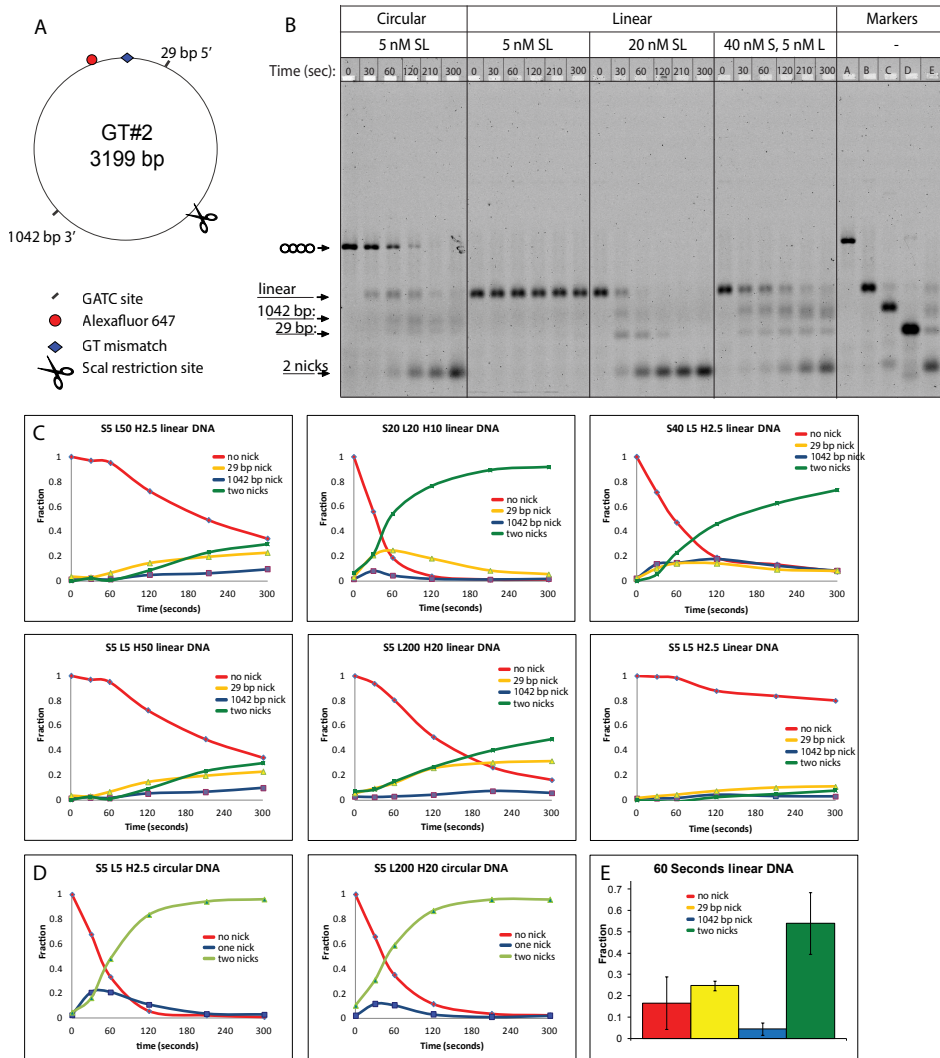
Figure 1: Circular DNA substrate is nicked faster than linear substrate. A) The GT#2 substrate used for these experiments carries a single G.T mismatch and two hemi-methylated GATC sites. In B) an agarose gel with nicking experiments comparing a circular substrate with a linear substrate. The markers on the right side of the gel are closed circular DNA (top band), linearized DNA, DNA nicked at a single GATC site at 1042 bp and linearized with ScaI (marked "1042 bp"), DNA linearized with ScaI and nicked at the GATC site 29bp from the mismatch ("29 bp") and DNA nicked at both GATC sites ("2 nicks"). In C) and D) quantified time courses with linear substrate (C) and circular substrate (D) under a range of conditions are shown. In these graphs, protein concentrations are varied as indicated above the graph. All points are the average of at least 3 independent experiments, lines are added for clarity. The data points represent the fraction closed circular DNA substrate (red), substrate that is nicked once (blue and yellow) and substrate that is nicked twice (green). In E), the fraction closed circular substrate and linear products formed after 60 s, in an experiment with 20 nM MutS, 20 nM MutL and 10 nM MutH. This is the average of 5 experiments ± standard deviation.

linear DNA is nicked efficiently (lane 13-18). Also, an increase in the concentration of either MutH or MutL on linear DNA increased the incision rate, and the GATC site at 29 base pairs from the mismatch is nicked preferentially. When only the MutS concentration is increased to 20 nM, while the MutL concentration and the MutH concentration are kept at 5 and 2.5 nM respectively, this preference disappears and the GATC site at 1042 bp is also utilized (lane 19-24).

The disappearance of the closed circular substrate over time can be fit with a single exponential decay with a half-life of 52 ± 17 seconds (Table 1). This activity is comparable to the activity found previously at similar protein concentrations (Au, Welsh et al. 1992), and higher than typically observed on small substrates (<100 bp) in trans activation assays (Junop, Yang et al. 2003, Selmane, Schofield et al. 2003). We consider it unlikely that the binding of MutS to the mismatch is different on the 3.2 kb circular DNA (which is not supercoiled during the nicking reaction) compared to the linear DNA. Therefore, the reason for the inefficiency of incision of linear DNA must reside in steps beyond mismatch recognition. A possible explanation follows from the diffusive properties of the ATP-bound MutS sliding clamp on DNA. MutS readily dissociates from open DNA ends, while dissociating much slower from circular DNA (Acharya, Foster et al. 2003) or DNA with blocked ends (Lebbink, Fish et al. 2010, Schofield 2001). Since the linear DNA contains two open ends, a MutS clamp can slide off the end, which can drastically decrease the number of MutS molecules on the DNA, and thus the efficiency of nicking. By increasing the concentration of MutL and MutH, the chance of forming the incision complex before MutS slides off increases. On circular DNA, when the concentration of MutL and MutH was increased to 200 nM and 20 nM, the nicking rate did not increase (Figure 1D). We conclude that on circular DNA at these conditions the loading of MutS onto the mismatch is limiting the speed of incision, while on linear DNA the diffusion of the pre-incision complex of the open end limits the reaction.

Preferential nicking of the nearby GATC site on linear DNA

As mentioned earlier, products that have been nicked once can be distinguished from products that have been nicked twice by separating the different ssDNA reaction products using denaturing agarose gel electrophoresis. Furthermore, when the DNA is linearized with the restriction enzym ScaI, the size of the fragments arising from a single-nicking event reveals which of the two GATC sites was nicked (Figure 1B-E). Because one GATC site is located only 29 bp 5' from the mismatch and the other one is located 1042 bp 3' from the mismatch, one might expect preferential nicking of the site close to the mismatch. However, no preference for either GATC site is observed on circular DNA (Figure 2). In contrast, on linear DNA there is a preference for the site
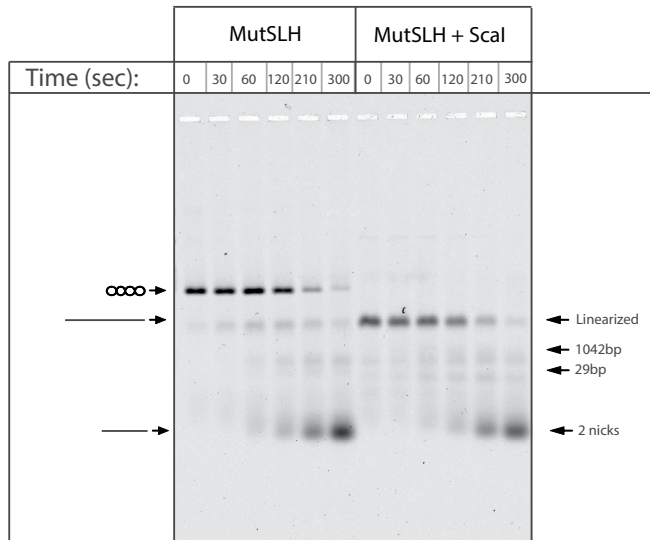
Figure 2: No preferential nicking of GATC sites on circular DNA. On the right side of the gel, DNA was cut with ScaI for 30 minutes after the experiment was stopped. When DNA is circular during the experiment, and linearized only after the end of the experiment, no preferential nicking of the GATC site at 1042 bp or 29 bp from the mismatch was found.

29 bp 5' of the mismatch. Interestingly, this difference manifests itself strongly when MutL and MutH are in excess over MutS (Figure 1B-C) but not so much if MutS is in excess (40 nM MutS over 5 nM MutL) (Figure 1B). *E. coli* MMR is bi-directional (Cooper, Lahue et al. 1993), and the MutS sliding clamp diffuses in both directions without preference (Gorman, Wang et al. 2012), therefore it is unlikely that the preference for the GATC site at 29 bp results from its 5' orientation relative to the mismatch. Also the sequence context of the GATC sites is unlikely to affect the preferential nicking, since no such preference is observed on circular DNA (Figure 2). Additionally, at 50 mM KCl, MutH can nick efficiently in the absence of MutS, and does not show preference for either GATC site under these conditions (Chapter 2). Thus, preferential nicking of the GATC site 29 bp from the mismatch over the GATC site 1042 bp from the mismatch is due to the localization of the GATC sites with respect to the mismatch, but does not manifest itself on a circular substrate.

Incision on circular substrates is not influenced by distance between mismatch and GATC site

To further investigate the possible absence of distance dependence on circular substrates, we designed an experimental setup that can address this in an internally controlled manner on separate substrates. At high protein concentration (200 nM MutS, 200 nM MutL, 50 nM MutH) the delays caused by
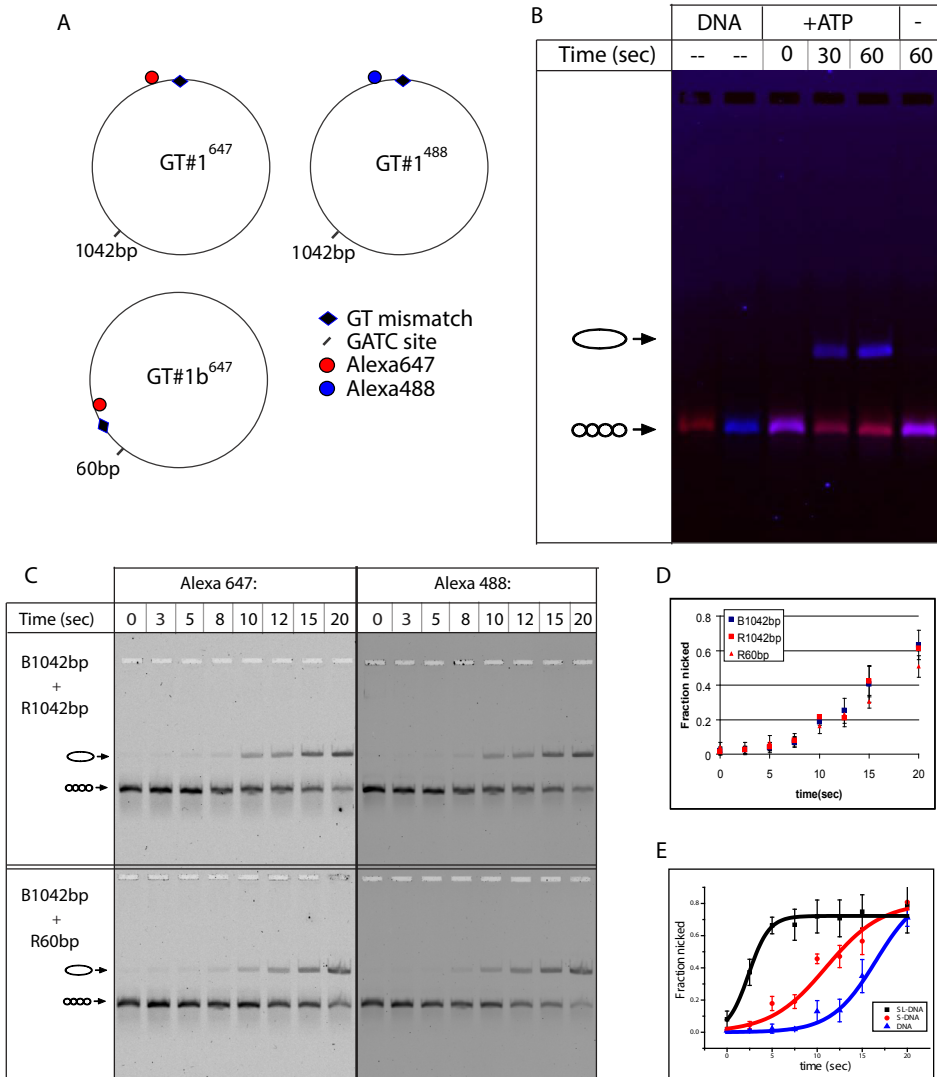
Figure 3: Preloading MutS and MutL increases the nicking rate. Panel A) shows the substrates used for the nicking experiments. All three substrates use the same single stranded phagemid DNA of 3197 bases as backbone, but are produced with different primers. Panel B) is an overlay of two scans, separately exposed for Alexa488 and Alexa647 labeled DNA. Lane 1 shows Alexa647 labeled homoduplex DNA, lane 2 Alexa488 labeled DNA with a GT mismatch, in lane 3, 4 and 5 both substrates were mixed with MutSLH and ATP, and the reaction as stopped at the indicated time. In the reaction mixture displayed in lane 6, ATP was omitted. Panel C) shows a time course of an experiment in which the substrate was a mix of GT#1488 and GT#1b647. The fluorescence of Alexa488 is shown in the top and Alexa647 at the bottom. D) shows the average and the standard deviation of 3 experiments, fit with a logistic function. E) shows the average for the order of addition experiments (dots) fitted with a logistic function (line).

binding steps are minimized, and a possible delay caused by the communication between mismatch and GATC site can be addressed. We created a substrate with a single GATC site, and varied the position of the GT mismatch (Figure 3A). By using two fluorophores, Alexa Fluor 647 and Alexa Fluor 488, differentially labeled substrates can be mixed and the relative amount of product from each substrate can be determined in a single reaction. In this way, each substrate will serve as an internal control for the other to reduce experimental error. MutS, MutL, MutH and 1 nM DNA were mixed with ATP, and reactions were stopped at 2.5 s intervals and the ratio of nicked versus un-nicked DNA was determined with agarose gel electrophoresis (Figure 3B). The nicking rate did not significantly change when the reaction was initiated with DNA and ATP instead of ATP alone (not shown), indicating that at a concentration of 200 nM MutS binding to the mismatch does not contribute significantly to the incision rate.

We first compared nicking rates on differentially labeled substrates with the GATC site at 1042 bp from the mismatch (GT#1, Figure 3C top panels). The half-life of the closed circular DNA was 12.4 ±1.6 seconds for GT#1[647] (GT#1 labeled with Alexa Fluor 647) and 14.3 ±2.1 seconds for GT#1[488] (GT#1 labeled with Alexa Fluor 488) which indicates that the type of fluorophore used did not significantly influence the nicking rate (Figure 3D). We then changed the distance from the mismatch to the nearest GATC site from 1042 bp (substrate GT#1) to 60 bp (substrate GT#1b) and mixed these two substrates to directly compare nicking efficiencies in an internally controlled setup (Figure 3A). Surprisingly, decreasing the distance between the mismatch and the nearest GATC site by almost 1 kb did not significantly change nicking rates (Figure 3C, lower panels); the half-life of the closed circular DNA was 14.1 ±1.7 seconds for GT#1b[647] (Figure 3D). We conclude that on circular DNA the MMR proteins can rapidly locate and incise GATC sites, independent of the distance between the mismatch and the GATC site.

Conformational changes in MutL are rate limiting for daughter strand incision

The conversion of the closed circular substrate into the open circular product shows a "lag-phase" of approximately 7 s before a measurable amount of DNA is nicked (Figure 3D). Because we pre-incubate the proteins and DNA substrate GT#1 in the absence of ATP, and started the reaction with ATP, this delay necessarily includes all reaction steps that occur between binding the mismatch and the actual incision of the GATC site of the DNA. This involves MutS forming a sliding clamp (MutS maturation), MutS binding MutL, MutL binding ATP and the associated conformational change (MutL maturation) and subsequently activating MutH at a GATC site, and MutH nicking the DNA. To investigate the contribution of each of these steps in this lag phase, we changed the order

of addition of the reaction components in the incision assay to allow partial reactions to take place before assembly of the complete incision machinery. We first incubated the substrate GT#1 with MutS and ATP, and after equilibrating the reaction for 5 min, the incision reaction was started by adding MutL and MutH. The initial lag phase of 7 seconds partially disappeared and substrate was nicked with a half-life of 9.7 ±1.8 seconds instead of 14.3 ± 2.1 seconds (Figure 3C). Thus, preloading of MutS on DNA as active sliding clamp in the absence of MutL effectively shortens the half-life by 4.6 seconds. Since the MutS sliding clamp can leave the mismatch (Schofield, Nayak et al. 2001), we expect that multiple clamps can be loaded onto the DNA, analogous to what has been proposed in the molecular switch model (Gradia, Acharya et al. 1997, Gradia, Subramanian et al. 1999, Acharya, Foster et al. 2003). Therefore, the half-life decrease of 4.6 seconds does not represent the average loading time of a single MutS clamp, but rather the increase in incision rate due to the presence of multiple sliding clamps on the DNA.

	[MutS] nM	[MutL] nM	[MutH] nM	Experiment Half-life (seconds)	Model Half-life (seconds)
Circular DNA	5	5	2.5	51.7 ± 16.6[2]	37
	5	200	20	41.4 ± 3[2]	27
	200	200	50	14.3 ± 2.4[2]	13
Linear DNA	5	5	2.5	635[1,2]	830
	5	50	2.5	212[1]	254
	5	5	50	265[1]	710
	20	20	10	37.7 ± 13.1[2]	104
	40	5	2.5	53.9 ± 7.2	240
	5	200	20	172.9 ± 8.1[2]	133

Table 1: Measured half lives of the un-nicked substrate at a range of MutS, MutL and MutH concentrations. Linear DNA is nicked at a much lower rate than circular DNA. Except for the half-lifes marked with [1], all data was fitted with a single exponential decay function. Each half-life is an average of at least three independent experiments ± standard deviation. [1] These half-lifes were fitted with linear regression. [2] These half-lifes were used for the goodness of fit for the modeling.

In order to determine the time it takes to recruit MutH and find the GATC site, we incubated the GT#1 DNA with MutS, MutL and ATP for 5 minutes at 37°C, and started the reaction by adding MutH (Figure 3C). When we pre-incubated the DNA with MutS, MutL and ATP, the half-life decreased to 2.5 ±0.2 seconds (Figure 3D). This shows that 1) MutS and MutL load onto the DNA without MutH, and 2) recruiting MutH, locating a single GATC site in 3.2 kb DNA and
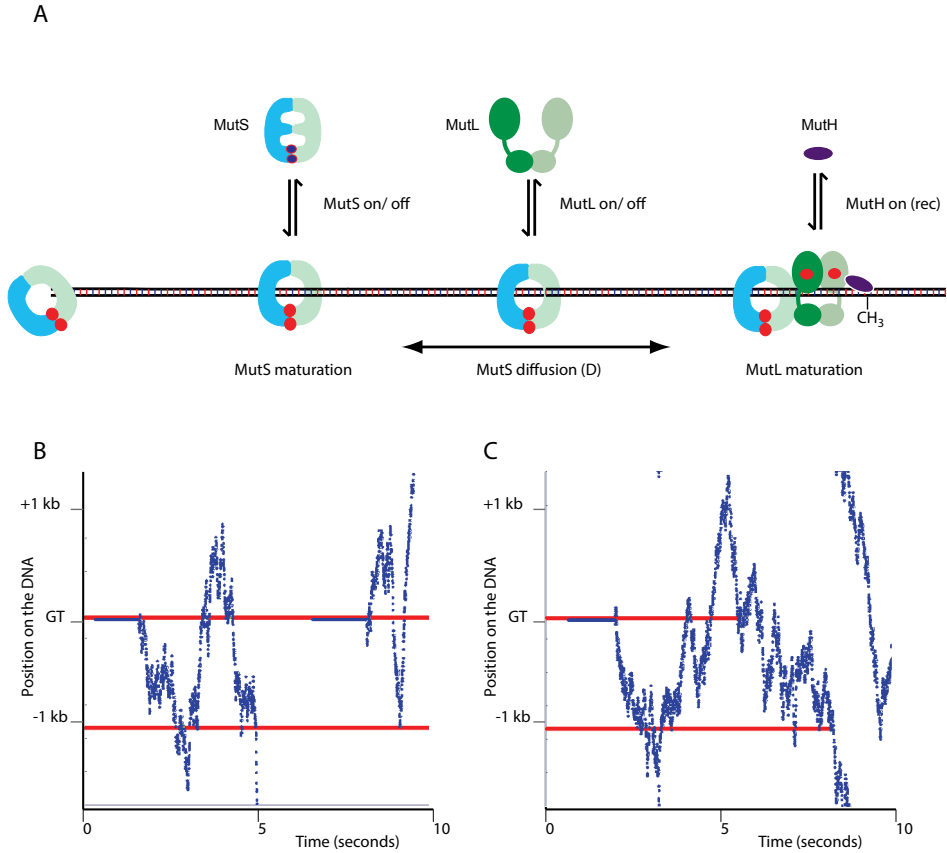
A

Figure 4: Graphical representation of the Monte Carlo simulation. The size and the localization of the mismatch and the GATC sites were similar to the substrates used in the experiments in Figure 1 and 2. The parameters that are used in the model are association and dissociation rate constants for MutS (MutS on/off), a MutS maturation time, association and dissociation rate constants for MutL (MutL on/off), a MutL maturation time and a rate constant for MutH binding and incising at the GATC site (used as a single recognition rate). The final parameters used in the simulations are shown in Table 2. In panel B and C two example simulations are presented as kymographs, with the position of MutS (blue dots) on the DNA as a function of time. The two GATC sites are represented by the two red lines, which disappear when the sites get nicked. B) Simulation on linear DNA, where MutS can fall off the open ends of the DNA. C) Simulation on circular DNA, where MutS will reappear on the other end of the DNA when it reaches the end.

nicking the unmethylated strand takes approximately 2.5 seconds. The search for a GATC site and hydrolysis of the DNA at the GATC site by MutH is thus faster than the binding and maturation of MutS and MutL. On substrates with 12 GATC sites nicking induced by preloaded MutS and MutL seems faster than on the substrate with a single GATC site (data not shown), however at present time resolution is insufficient to properly access this.

Since the recruitment of MutH is relatively fast, and MutS sliding clamp formation takes a maximum of 4.6 seconds, the most time consuming step involves the loading and maturation of MutL. This lasts at least 7.2 seconds (14.3 - 4.6 - 2.5 seconds). However, since multiple MutS sliding clamps have loaded onto the DNA, each of which being able to bind MutL, the average time of a single MutL clamp being loaded and activated is likely slower than 7.2 seconds. Because we use high (saturating) protein concentrations, it is unlikely that binding of MutL to MutS will be rate-limiting. The slow step in the MutL ATPase cycle is the ATP-induced closure of the nucleotide binding domains in the opposing MutL monomers, which can take up to minutes in the absence of MutS, MutH and DNA (Junop, Yang et al. 2003). The closure of MutL is required to activate MutH (Hall, Jordan et al. 1998, Junop, Yang et al. 2003). These results suggest that 1) the slow step in the daughter strand incision can be attributed to conformational changes in MutL, and 2) these changes seem to be considerably accelerated (from a duration of minutes to seconds) in the presence of MutS, MutH and DNA.

Quantitative model reveals timing of events

To better evaluate the observed differences on linear and circular DNA, the lack of distance dependence on circular substrates and the timescale on which conformational changes occur, we constructed a quantitative model. We consider the following interacting components: i) a single linear DNA molecule, with the same size and positioning of the mismatch and GATC sites as our DNA substrate GT#2 used in the experiments, ii) MutS dimers, that can bind the mismatch, mature, bind MutL, diffuse along the DNA, and dissociate, and iii) MutL dimers, that can bind to mature MutS, mature, activate MutH at a GATC site, and dissociate (Figure 4A). The behavior of these molecules was based on a set of assumptions and parameters obtained from literature (Table 2) as well as on our own observations and are discussed in detail below.

DNA is simulated as a single string of 10 bp units. When simulating linear DNA, MutS dissociates when it reaches the end of the DNA molecule. When circular DNA is simulated, the dissociating MutS molecule reappears at the other end of the DNA molecule (Figure 4B). The interaction of MutS with the DNA is governed by several parameters. We use an association rate constant (k_{on}) for MutS binding to a DNA mismatch of 1.25 x 10^7 M^{-1} s^{-1} per monomer. This is only slightly slower than the measured k_{on} for *E. coli* MutS of 4.4×10^7 M^{-1} s^{-1} (Cristovao, Sisamakis et al. 2012). To take into account the observed 20 bp DNaseI footprint of MutS (Grilley, Welsh et al. 1989, Schofield, Nayak et al. 2001), no additional MutS molecules can bind when a MutS molecule is located on the mismatch or on one of the DNA strings directly adjacent to the mismatch.

Parameter	Value
MutS association	$1.25 \times 10^7\ M^{-1}\,s^{-1}$ [this work]
MutS offrate	$0.0017\ s^{-1}$ (Cho, Jeong et al. 2012)
MutS maturation	$0.43\ s^{-1}$ (Jeong, Cho et al.)
MutL association	$2 \times 10^7\ M^{-1}\,s^{-1}$ [P.Friedhoff personal communication]
MutL offrate	$5 \times 10^{-3}\ s^{-1}$ [P.Friedhoff personal communication]
MutL maturation	$0.07\ s^{-1}$ [this work]
MutH recognition	$3 \times 10^7\ M^{-1}$ [this work]
Diffusion MutS clamp	$10^6\ bp^{-2}\,s^{-1}$ (Cho, Jeong et al.)

Table 2: Parameters used for the output of the simulations. The values for the rates of diffusion, MutS maturation and dissociation were taken from literature. The MutL association and dissociation rates were measured by collaborators. The MutS association rate, MutL maturation rate and the MutH recognition rate are fitted with the experimental data. Most of the parameters are initialized with values taken from literature or based on experiments, with exception of the MutH recognition, the MutS association rate, and MutL maturation.

After MutS recognizes the mismatch, a conformational change has to occur in MutS to form a functional sliding clamp. This conformational change is triggered by the exchange of ADP to ATP in at least one of the nucleotide binding sites of MutS (Bjornson, Allen et al. 2000, Lebbink, Georgijevic et al. 2006). In our model, this conformational change is encapsulated in a maturation rate and MutS can only leave the mismatch after maturation into the sliding clamp. A range of numbers has been published for this process, from $0.24\ s^{-1}$ determined with SPR (Lebbink, Fish et al. 2010) to $2.3\ s^{-1}$ measured with FRET (Cristovao, Sisamakis et al. 2012). However, since the MutS dimer has two nucleotide-binding sites, the occupancy of either binding site with ATP or ADP during mismatch recognition is relevant and may determine the rate at which the nucleotide exchange and conformational changes for sliding clamp formation occur. Therefore we choose the rate for MutS sliding clamp formation of $0.43\ s^{-1}$ as reported for *Thermus aquaticus* MutS in the continuous presence of ATP (Jeong, Cho et al. 2011, Qiu, DeRocco et al. 2012). We have analyzed the effect of the MutS maturation rate over the range of values available in literature (Figure 5C,E). A maturation rate between 0.25 and $0.5\ s^{-1}$ can easily be used to fit the data, however, faster maturation rates of $>1\ s^{-1}$ do not.

Once matured, the MutS sliding clamp can diffuse freely over DNA that is not occupied by other MutS molecules. The diffusion along the DNA of the MutS sliding clamp (*D*) is taken from literature as $0.1\ mm^{-2}\ s^{-1}$ (Cho, Jeong et
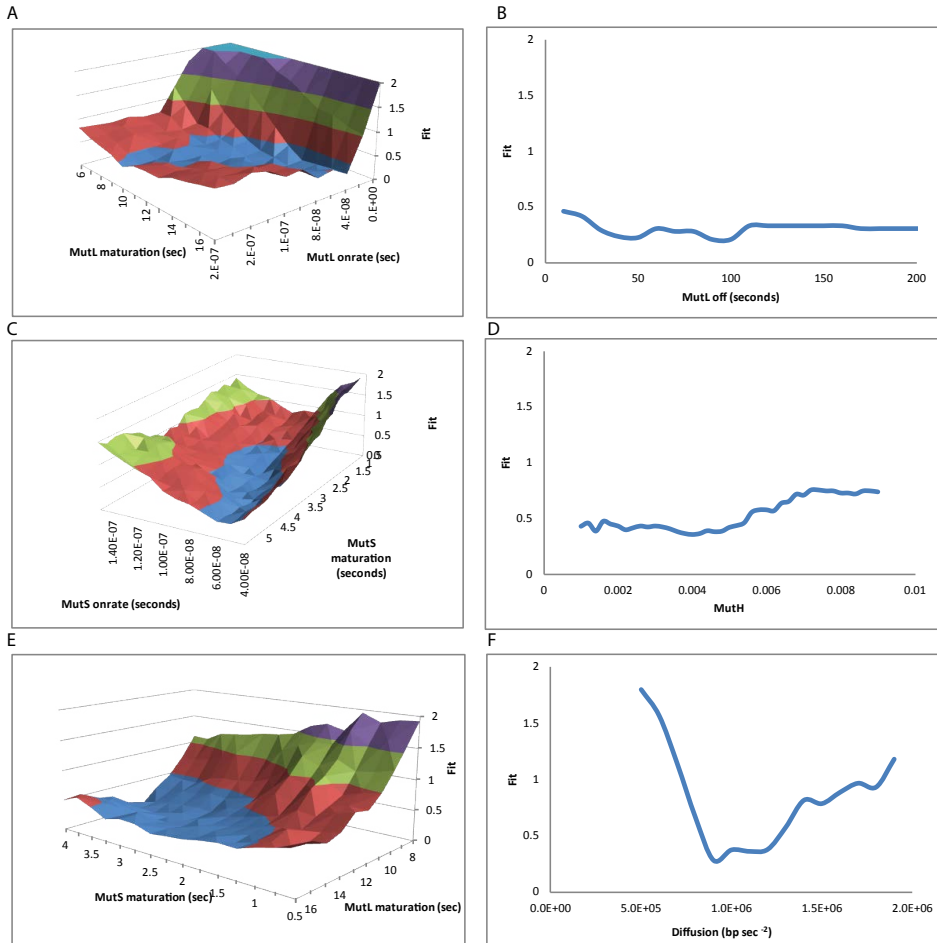
Figure 5: The influence of the MutL maturation, MutS maturation, MutL association and the MutS association on the quality of the fit of the experimental data. In panel A, the effect on the global fit of the data is shown when the MutL maturation and MutS maturation are varied relative to each other. Panel B shows the influence of the MutL dissociation rate on the global fit, D the MutS dissociation rate, and F the diffusion. In C) the MutS association rate is shown, which has an optimum around 1.2×10^7 M^{-1} sec^{-1}.

al. 2012). We do not alter this diffusion rate when MutL is bound to MutS (see discussion). Decreasing the diffusion rate to 0.05 mm^{-2} s^{-1} reduces the fitting power of our model, and an increase also has a detrimental effect (Figure 5F). Only a mature MutS sliding clamp can bind MutL (Acharya, Foster et al. 2003). In the simulation, MutL dimer binds the MutS sliding clamp with a k_{on} of 2×10^7 M^{-1} s^{-1} (P.Friedhoff, personal communication). The k_{on} for MutL could be changed in a range of 1×10^7 M^{-1} s^{-1} to 4×10^7 M^{-1} s^{-1} with minimal influence on the fit of the data (Figure 5A).

Figure 6: Comparison of the output from the simulations with the experimental data. Panel A (circular DNA) and B (linear DNA) sho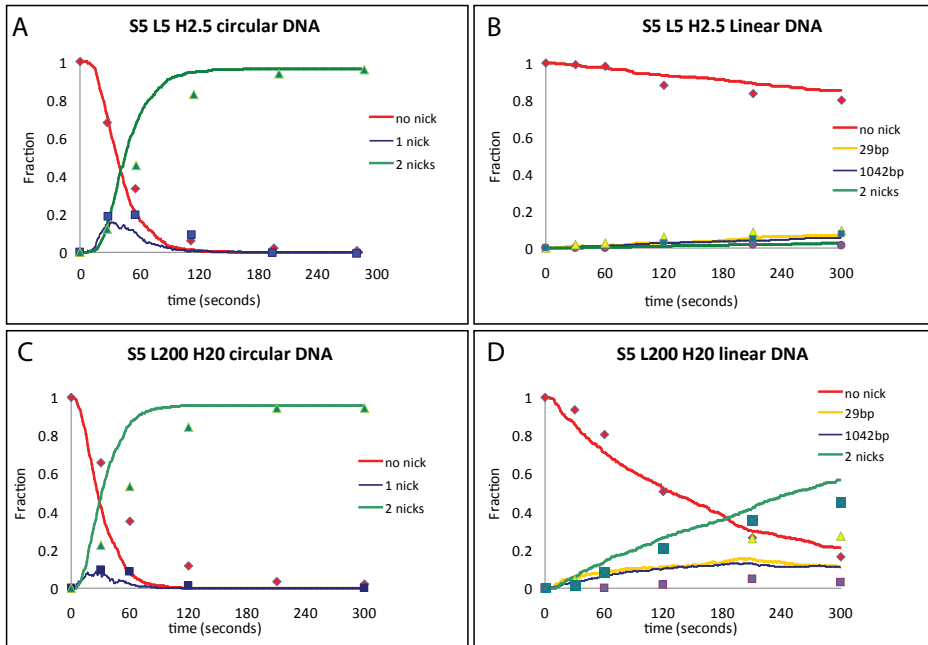w the data (dots) and predictions (lines) from the experiments at 5 nM MutS, 5 nM MutL and 2.5 nM MutH. Panel C shows the incision of circular DNA at 5 nM MutS, 200 nM MutL and 20 nM MutH. Panel D shows the incision of linear DNA at 20 nM MutS, 20 nM MutL and 10 nM MutH. The red line represents the fraction of un-nicked molecules, the yellow line the fraction of molecules nicked only at the GATC situated 29 bp from the mismatch, the blue line represents the fraction of molecules nicked at the GATC site situated at 1042 bp from the mismatch, and the green line represents the fraction of molecules that is nicked at both sites. The dots are the averages of at least three independent experiments, as reported in Figure 3, and the simulations were performed for 1000 molecules.

Nucleoside triphosphate binding by MutL induces ordering of multiple disordered loops, resulting in the dimerization of the nucleotide-binding domains. This is the slow step during the MutL ATPase cycle, taking several minutes to complete in the absence of other ligands (Ban, Junop et al. 1999). In the presence of ssDNA, but not dsDNA, these conformational changes are accelerated (Ban and Yang 1998). Nucleotide binding by MutL is required for MutH activation (Hall, Jordan et al. 1998, Junop, Yang et al. 2003). Our observations that preloading MutS and MutL onto dsDNA in the presence of ATP allows MutH activation within seconds, suggests that also MutS is able to significantly accelerate the disorder-to-order transition in MutL. To our knowledge, the timescale at which these conformational changes occur in the presence of DNA and MutS has not yet been measured directly. Fitting the experimental data with the model resulted in a rate for this conformational change of around 0.07

s^{-1}. Interestingly, a slower k_{on} for MutL cannot compensate for a decrease in the maturation rate of MutL, increasing the MutL maturation rate to 0.25 s^{-1} decreases the fitting potential of the model significantly (Figure 5A,B).

The chance that MutH will nick the GATC site is simulated as 3 x 10^7 M^{-1}, with the prerequisite that a clamp consisting of matured MutS and MutL is positioned on the GATC site. However, since we do not know at what proximity MutL can activate MutH at a GATC site, this number does not have a biological significance. Rather it should be regarded as a combined rate constant for the binding of MutH to the mature MutSL complex, the binding of MutH to its target sequence, and the chemistry of the hydrolysis.

Both MutS and MutL can leave the DNA or the MutS sliding clamp, respectively. The k_{off} for MutL is approximately 0.005 s^{-1} (Peter Friedhoff personal communication), although in our simulations the fit is barely affected by any change in this value below 0.01 per s^{-1} (Figure 5D). The k_{off} for the MutS sliding clamp from circular DNA is 1.7 x 10^{-3} sec^{-1}, measured for the ATP-bound state of *Taq* MutS in single molecule experiments (Cho, Jeong et al. 2012) and estimated from SPR measurements to be in the same order of magnitude for *E. coli* MutS (Lebbink, Fish et al. 2010).

With these parameters, and a simple random walk model, we simulate the movement of the proteins on the DNA and the incision of the GATC sites (Figure 4). Interestingly, the observed difference in incision of circular and linear DNA at 5 nM MutS, MutL and 2.5 nM MutH (Figure 1A) was accurately reproduced by our model (Figure 6A&B), showing that a simple random walk can indeed explain the difference between linear DNA and circular DNA. When the concentration of MutL and MutH in the experiment is increased to 200 nM and 20 nM respectively, the incision of the circular substrate is not changed significantly (Figure 1C). Indeed, the simulations predict only a slight increase in incision rate, which in the data could be masked by experimental variation (Figure 6C). In contrast, on linear DNA, the incision rate in the experiments is significantly increased at these higher protein concentrations, which is correctly predicted by the simulations (Figure 1D, Figure 6D, Figure 7A-D). However, it does not approach the rate on circular DNA. To understand why despite saturating protein concentrations, nicking on linear DNA is never as efficient as on circular DNA, we analyzed the composition and status of the protein complex at the moment of diffusing off from the end of the linear DNA (Figure 8). Clearly at a low MutL concentration, most MutS molecules fall off before binding MutL, and only a small number of maturated MutSL complexes can form. When the concentration of MutL is increased, most MutS molecules will in fact be bound by MutL before they reach the end of the DNA molecule, however MutL will not mature fast enough to activate MutH (Figure 8). We therefore conclude that the
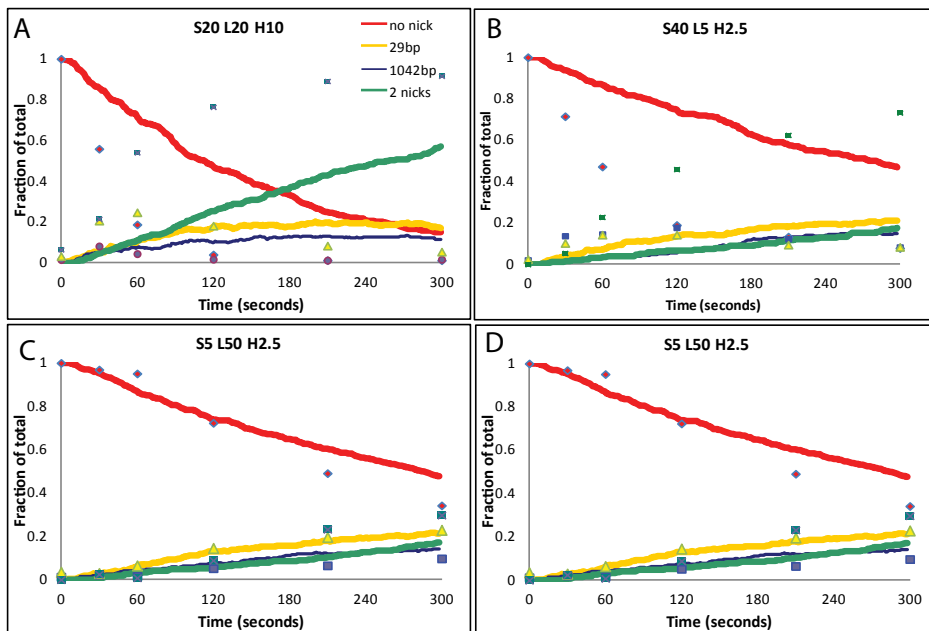
Figure 7: Simulations for higher MutS concentrations on linear DNA. When concentrations of MutS, MutL and MutH are increased to 20 nM for MutS and MutL (panel A), and 10 nM for MutH, the incision (dots) is faster than the model predicts (lines). When only the MutS concentration is increased to 40 nM, the predicted incision rate is also slower than the data (panel B). This suggests cooperativity in MutS binding to the DNA. In panel C and D, the prediction and data are given for 5 nM MutS, 50 nM MutL, 2.5 nM MutH, and 5 nM MutS, 5 nM MutL, 50 nM MutH, respectively.

differences between linear and circular substrates are due to MutS or MutSL complexes sliding off from the open ends.

If the MutS concentration is increased from 5 nM to 40 nM, while the concentration of MutL and MutH is kept constant, the incision rate increases (Figure 1). Interestingly, the simulations underestimate the incision rate under these conditions (Figure 7B). This indicates there is cooperativity in MutS binding on linear DNA, which will be discussed further below.

Quantitatively modeling of mismatched directed nicking at saturating protein concentration

Only a short time window of approximately 1 minute is available for strand-specific MMR in *E. coli* before the hemi-methylation of GATC sites is removed by the Dam methylase (Marinus and Casadesus 2009). The protein concentrations for MutS, MutL and MutH *in vivo* are typically in the range of 100 nM to 1 μM (Feng, Tsui et al. 1996). In our experiments, at saturating protein concentrations

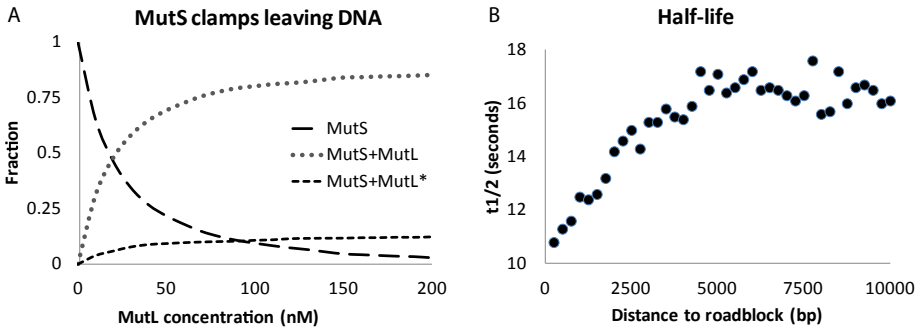**Figure 8: The influence of open ends and blocked ends on the nicking rate.** A) The fraction of MutS clamps that slides off the linear DNA from either end, as a function of the MutL concentration (output of simulation). At low MutL concentration, most MutS sliding clamps slide off before binding MutL (red line). When the MutL concentration is increased, the amount of MutS sliding clamps that slide off the ends bound to non-mature MutL (green) and MutS clamps bound to mature MutL (purple) also increases. B) Roadblocks can increase the nicking rate by decreasing the amount of DNA that is scanned by MutS.

(200 nM MutS, 200 nM MutL, 50 nM MutH), 50% of the circular DNA is nicked in 14 seconds, independent of the distance between the GATC site and the mismatch (Figure 3C). At these protein concentrations, the model predicts a half-life of 11.2 seconds for the GATC site at a distance of 60 bp, and 10.5 seconds for a GATC site at 1042 bp. This is slightly faster than the measured half-life of 14.3 seconds, but some discrepancy is to be expected, since the model does not incorporate any non-specific binding events that may occur at these high protein concentrations. An interesting observation from the simulations is that the nicking rate at these saturating protein conditions is faster than the slowest step in the process, which is MutL maturation. On average, a single MutL dimer will mature in 14 seconds (Table 1). However, since the formation of a MutS sliding clamp takes (on average) 2.3 seconds, multiple MutS sliding clamps will be loaded during the lag phase of the reaction. Since all these sliding clamps can recruit MutL, the average time the first functional MutS-MutL clamp is formed will be shorter than the sum of the average maturation times.

Compared to the relatively small circular and linear DNA substrates used in the *in vitro* assays, chromosomal DNA in cells can basically be considered as continuous in either direction. However, only a limited section of this DNA is expected to be accessible for MutS since diffusion may be blocked by numerous other proteins binding to the DNA directly after replication. To investigate the influence of these possible roadblocks on the efficiency of incision, we introduced reflecting boundaries into our model on both sides of the mismatch.
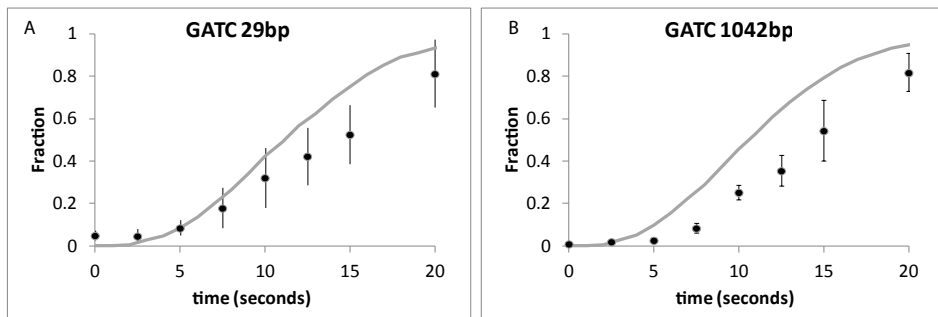
Figure 9: Absence of distance dependence in both the experimental data and the simulations. No significant difference was found between the half life of the substrate with the GATC site 1042 bp from the mismatch (12.4 ±1.6 seconds) compared to the substrate on which the GATC site is located 60 bp away (14.1 ±1.7 seconds). The model (lines in figure) predicts slightly faster rates, with a half-life of the substrate of 10.5 seconds for the GT#1 (GATC site at 1042 bp) substrate, and 11.2 sec for the GT#1b substrate (GATC site at 29 bp).

We evaluated the efficiency of nicking two GATC sites located 130 bp from the mismatch, while varying the position of the reflective barriers from 10 kb to 250 bp from the mismatch. When the reflecting boundaries are moved closer the mismatch, the half-life decreases from ~16 seconds to ~9.5 seconds (Figure 8B). Thus, while roadblocks decrease nicking rates when placed between the mismatch and the GATC site (Pluciennik and Modrich 2007), they increase the efficiency of strand incision when flanking the mismatch by confining the amount of DNA that needs to be scanned in search for a GATC site.

DISCUSSION

We addressed the mechanism of strand incision by MMR biochemically and tried to encapsulate all our results in a deterministic model. With this model we can explain how strand incision works in our *in vitro* experiments, and how it may function *in vivo*. *In vitro*, MMR is more efficient on circular DNA substrates than on linear DNA (Au, Welsh et al. 1992). In our experiments, the rate difference is large: at low protein concentrations (5 nM MutS, 5 nM MutL, 2.5 nM MutH) half of the circular DNA is nicked within 40 seconds, while linear DNA has a half-life longer than 300 seconds. Interestingly, increasing protein concentrations of either MutL or MutH increases the nicking rate on linear DNA, but not on circular DNA (Figure 1). This can be explained by the slow dissociation rate of MutS from closed DNA (Qiu, DeRocco et al. 2012), and the fast dissociation rate of MutS from open-ended (linear) DNA (Schofield, Nayak et al. 2001). Intuitively, a diffusion based model for the MutS sliding clamp can readily explain these differences between linear and circular DNA. If MutS sliding clamp diffusion is indeed used to communicate between the

mismatch and the GATC site, a decrease in nicking rate might be observed when the distance between the mismatch and the nearest GATC site is increased. However, the published diffusion rate of the MutS sliding clamp is 0.1 mm^2 s-1 (Qiu, DeRocco et al. 2012), indicating that MutS can move more than 1 kb within one second. We found that the nicking rate on circular DNA is independent of the distance between the mismatch and the nearest GATC site. When a single GATC site is situated 1042 bp from the mismatch, the average time required for MutH to nick this site is 14.3 ±2.1 seconds. When the distance is decreased to 60 bp, the rate does not change significantly (half-life 14.1 ±1.7 seconds). This indicates that the relative contribution of diffusion to the overall nicking rate is small, implying that the contribution of the assembly of the incision complex is rate limiting.

By changing the order of addition of the reaction components, we allow preloading of MutS, or MutS and MutL onto DNA before the reaction starts. In this way we determined the approximate time required to assemble MutS and MutL on the DNA in the full incision reaction. The order of addition experiments show that MutS can load in the absence of MutL-MutH, effectively reducing the half-life of the substrate to 9.7 ±1.8 seconds, a reduction of 4.6 seconds. Since this decrease in nicking rate is larger than the time required to form a single sliding clamp, multiple loading of MutS sliding clamps is likely responsible for this relatively large increase (Jeong, Cho et al. 2011). Interestingly, the relative decrease in half-life due to preloading MutS (4.6 seconds) is modest compared to the half-life decrease when MutS-MutL (11.6 seconds) is preloaded. This suggests that the formation of an active MutS-MutL complex is rate limiting, and not MutS sliding clamp formation or the incision of the GATC site by MutH. Large conformational changes (ordering of multiple loops at the dimer interface) during the binding of ATP, observed in multiple crystal structures of MutL (Ban and Yang 1998, Ban, Junop et al. 1999) could explain why the formation of an active MutS-MutL complex is rate limiting.

With simulations, we tried to explain our results quantitatively, while incorporating additional published results into a quantitative model. There are several reasons for making a quantitative model: First, simulations can quantitatively explain our experiments. Second, simulations can integrate previously measured parameters on single proteins and reaction steps into the timescales of the complete strand incision. Third, simulations can reveal the implications of changes in parameters in a complex system, and put boundaries on unknown parameters and parameters that cannot be assessed in a dire–ct manner. Our simulations reveal that a diffusive mode of communication between mismatch and GATC sites can readily explain the major differences observed between linear and circular substrates and the lack of distance dependence on circular substrates under a range of protein concentrations. Our

model also qualitatively explains results obtained previously (Pluciennik and Modrich 2007). As showed, a break in the short path between the mismatch and a single GATC site greatly diminishes the efficiency of incision of the GATC site. However, the long path spanning 6 kb is still available. Our simulations suggest that it is not the 6 kb distance between mismatch and GATC site that lowers nicking efficiency but the presence of an open end relatively close to the mismatch and the GATC site, because proteins will slide off the DNA ends before the GATC site can be located.

At high protein concentrations (200 nM MutS, 200 nM MutL), association rates are fast relative to the half-life of 14.3 seconds on the circular substrate (4 and 5 s^{-1} for MutS and MutL, respectively). Furthermore, by preloading only MutS or MutS and MutL prior to the start of the experiment, the half-life can be decreased to 9.7 and 2.5 seconds, respectively (Figure 3C). From this we conclude that conformational changes in MutS and MutL are rate limiting in the overall reaction. Previously, measurements of conformational changes in MutS, which allow it to form a sliding clamp after binding to the mismatch, show them to range from 0.5 to 4.2 seconds (Lebbink, Fish et al. 2010, Cristovao, Sisamakis et al. 2012, Qiu, DeRocco et al. 2012). The timescale of the conformational changes within MutL in the absence of other proteins seems to be very slow, around 5-10 minutes based on the steady-state ATPase rate (Ban, Junop et al. 1999). To our knowledge, the timescales at which conformational changes in MutL occur in the presence of other MMR proteins are not yet measured, but our order-of-addition experiments and the simulations suggest that these are still rate limiting for the incision reaction. In the quantitative model, the maturation of MutL is fitted to be approximately 0.07 s^{-1}.

An increase in MutS concentration in our experiments decreases the half-life of linear DNA more than predicted by the model. The failure of our simulations to properly fit the nicking data at higher MutS concentrations indicates that the relationship between the nicking rate and MutS concentration is not linear. This could be explained by cooperativity in MutS. *E. coli* MutS can bind the mismatch as a dimer, but can also form tetramers (Bjornson, Blackwell et al. 2003). The function of MutS tetramer formation in MMR is currently debated. MutS mutants that are obligate dimers are proficient in MMR *in vitro* and show only a modest repair deficiency *in vivo* (Lamers, Georgijevic et al. 2004, Manelyte, Urbanke et al. 2006, Mendillo, Putnam et al. 2007). However, tetramerization of MutS may play a role in anti-recombination (Calmann, Nowosielska et al. 2005, Tham, Hermans et al. 2013). We do not take MutS tetramers into account in our model, first of all because it is difficult to define when and how MutS forms tetramers, and it is unclear whether, and if so, how this influences daughter strand incision. A possible function for MutS tetramerization could be to keep MutS in the vicinity of the mismatch, consistent with the observed lower

dissociation rates of the MutS tetramer from mismatched DNA compared to the MutS dimer (Groothuizen, Fish et al. 2013). Keeping MutS in close proximity to the mismatch seems important, considering that the MutS sliding clamp can diffuse distances exceeding 1 kb along naked DNA within seconds (Qiu, DeRocco et al. 2012), and is thus likely to miss GATC sites close to the mismatch.

A purely diffusion-based model is not the only available model for communication between the mismatch and the GATC site. An additional model involves DNA loops spanning this distance (Allen, Makhov et al. 1997). Tetramerization of MutS (Calmann, Nowosielska et al. 2005) and DNA loop formation (Tham, Hermans et al. 2013) are important for the anti-recombination function of MMR. However, defining a looping model in such a way that would be suitable for building a quantitative simulation is difficult at best, and a simple diffusive model can readily explain most of the differences found in the nicking rate between linear and circular DNA. One exception is the preferential nicking of a GATC site close to the mismatch on linear DNA. On our GT#2 substrate, two GATC sites are present, one 29 bp from the mismatch, and a second one 1042 bp from the mismatch. When a circular substrate is used, both sites are nicked without preference (Figure 2 and Chapter 2). On linear DNA, the GATC site close to the mismatch is nicked preferentially. Interestingly, this is dependent on the concentration of the individual protein components; the preference for the site close to the mismatch is most pronounced when MutL and or MutH are in excess. When MutS is in excess, the difference between both GATC sites on linear DNA disappears. Further experiments are needed to properly explain these results. These results leave possible functions for factors like DNA loops, or multiple loading of MutL, which are not incorporated in our simulations for reasons stated earlier.

In vivo, hemi-methylated GATC sites are modified by Dam methylase shortly after replication (Marinus and Casadesus 2009). Over expression of Dam methylase in *E. coli* results in a mutator phenotype (Schlagman, Hattman et al. 1986), which is likely caused by failure of strand directional repair by MMR within the shortened time before methylation occurs. This indicates that the ability of MMR to find a hemi-methylated GATC site within a limited timeframe is important for its function. Interestingly, conformational changes in MutS and MutL seem slow compared to the 100-1,000 fold reduction in mutation rate that is achieved within this time window. This suggests that multiple loading of MutS and subsequently MutL plays a large role increasing the repair rate. In addition to multiple loading, our simulations show that roadblocks can limit the target search by MutS clamps, keeping MutS closer to the mismatch. Therefore, it is likely that in genomic regions with fewer or no GATC sites MMR is less efficient than in regions with many GATC sites. DNA stretches exceeding 1 kb that lack GATC sites are relatively sparse in the *E. coli* genome (Chapter 5). It

would be interesting to see whether the absence of GATC sites in large (>5 kb) DNA fragments influences the efficiency of MMR *in vivo*.

EXPERIMENTAL SETUP

Protein purification

MutS, MutL and MutH were expressed and purified as described previously (Tham, Hermans et al. 2013). Protein aliquots were stored at -80°C. Single aliquots of MutS, MutL and MutH were stored in -20°C in storage buffer containing 50% glycerol, and used for several weeks without a noticeable decrease in enzymatic activity.

DNA substrates:

DNA substrates with a single GT mismatch were produced by extending a primer on a circular single stranded DNA (ssDNA) phagemid (derivative of pGEM-13Zf (+)) of 3196 bases, essentially as described previously (Chapter 2 and (Baerenfaller, Fischer et al. 2006)). The primer introduces a single GT mismatch and a fluorophore, either Alexa Fluor 647 (GT#1[647] and GT#1b[647]) or Alexa Fluor 488 (GT#1[488]) (produced by IBE Germany). By mixing DNA labeled with Alexa Fluor 647 and Alexa Fluor 488, we could accurately compare the activity of MutSLH on GT#1b[647] and GT#1[488] within one experiment of 20 seconds duration. When linear DNA was used, the substrate was linearized with the restriction enzyme ScaI, and heat inactivated in the reaction buffer before the initiation of the nicking reaction.

Nicking reactions

Reaction buffer consisted of 25 mM Hepes KOH [pH 7.5], 150 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, 10% glycerol and 100 ng/ml BSA. All reactions were carried out at 37°C. For the gels run under denaturing conditions, reactions were carried out with 5 nM MutS, 5 nM MutL, 2.5nM MutH and 1 mM ATP final concentration, unless stated otherwise. Reactions were started by adding ATP. At the indicated time intervals, 10 µl samples were taken out of the reaction mixture, and the reaction was terminated by adding 15 µl stop buffer containing 8 M urea and 1% SDS. Samples were heated at 85°C for 10 minutes, and run on a 1.5% agarose gel in 1x TEA buffer supplemented with 1 M urea essentially as described in (Hegedus, Kokai et al. 2009). Agarose gels were supplemented with 0.5 µg/ml ethidium bromide to supercoil the DNA.

The order-of-addition experiments were performed with a final concentration of 200 nM MutS (monomer), 200 nM MutL (monomer), 50 nM MutH and 2

nM DNA (6.4 µM nucleotides). DNA substrates, ATP and either MutS, MutS and MutL, or no protein, were incubated at 37°C for 5 minutes. The reaction was started by adding equal volume of buffer including the missing component(s). Reactions were stopped by adding an equal volume of 20% glycerol, 1% SDS, 50mM EDTA, and run on a 1% agarose gel containing or 40 µM chloroquine to supercoil closed circular substrate. Labeled DNA was visualized with a typhoon trio imager (GE Healthcare). The DNA labeled with Alexa Fluor 488 labeled DNA was excited with a 488 nm laser, and the Alexa Fluor 647 DNA was excited at 633 nm. Emission was passed though filters 520BP40 and 670BP30, respectively. The amount of substrate and products was quantified using NIH ImageJ (Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, http://imagej.nih.gov/ij/, 1997-2012). The fraction of closed circular DNA, linear DNA, and nicked DNA substrate was determined, and each curve was fitted independently with a single exponential decay function (for the experiments run at denaturing conditions), or with a sigmoid function (order of addition experiments) in Graphpad Prism(version 5).

Simulations

   All simulations were written in Perl. For the graphs shown, the incision of 1000 DNA virtual DNA molecules were analyzed. Simulations were based on Monte Carlo methods; for each time step, the number and place of MutS/MutL molecules was determined heuristically, and the nicking of the GATC sites recorded. The random walk of each MutS molecule was determined as a 50:50 chance to move to a DNA compartment either to the left or to the right. If MutS was moved into a DNA compartment occupied by another MutS molecule, the move was cancelled. For the DNA, 3 different boundary conditions were used: for linear DNA, MutS falls off when it reaches either end of the DNA. For a circular DNA molecule, MutS is moved to the other end of the DNA when reaching either end. For the simulations calculating the effect of roadblocks (Figure 8), a roadblock was treated as a reflecting boundary: MutS cannot move further when reaching any roadblock on the DNA, but it can move backwards in the opposite direction. To obtain a quantitative description of the goodness of fit, the half-life of the closed circular substrate as predicted by the simulation was divided by the half-life from the experimental data as indicated in Table 1. The root mean squared deviation (rmsd) of the half-lifes predicted by the model compared to the half-life produced by the experiments was used to determine parameters for the MutS association rate, the MutL association rate, the MutL maturation rate and the MutH recognition rate (Table 2). This rmsd is also represented in vertical axis in the Figure 6A-E.

REFERENCES:

Acharya, S., P. L. Foster, P. Brooks and R. Fishel (2003). "The coordinated functions of the E. coli MutS and MutL proteins in mismatch repair." Molecular cell **12**(1): 233-246.

Allen, D. J., A. Makhov, M. Grilley, J. Taylor, R. Thresher, P. Modrich and J. D. Griffith (1997). "MutS mediates heteroduplex loop formation by a translocation mechanism." The EMBO journal **16**(14): 4467-4476.

Au, K. G., K. Welsh and P. Modrich (1992). "Initiation of methyl-directed mismatch repair." J Biol Chem **267**(17): 12142-12148.

Baerenfaller, K., F. Fischer and J. Jiricny (2006). "Characterization of the "mismatch repairosome" and its role in the processing of modified nucleosides in vitro." Methods in enzymology **408**: 285-303.

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

Ban, C. and W. Yang (1998). "Crystal structure and ATPase activity of MutL: implications for DNA repair and mutagenesis." Cell **95**(4): 541-552.

Bjornson, K. P., D. J. Allen and P. Modrich (2000). "Modulation of MutS ATP hydrolysis by DNA cofactors." Biochemistry **39**(11): 3176-3183.

Bjornson, K. P., L. J. Blackwell, H. Sage, C. Baitinger, D. Allen and P. Modrich (2003). "Assembly and molecular activities of the MutS tetramer." J Biol Chem **278**(36): 34667-34673.

Calmann, M. A., A. Nowosielska and M. G. Marinus (2005). "The MutS C terminus is essential for mismatch repair activity in vivo." J Bacteriol **187**(18): 6577-6579.

Cho, W. K., C. Jeong, D. Kim, M. Chang, K. M. Song, J. Hanne, C. Ban, R. Fishel and J. B. Lee (2012). "ATP alters the diffusion mechanics of MutS on mismatched DNA." Structure **20**(7): 1264-1274.

Cristovao, M., E. Sisamakis, M. M. Hingorani, A. D. Marx, C. P. Jung, P. J. Rothwell, C. A. Seidel and P. Friedhoff (2012). "Single-molecule multiparameter fluorescence spectroscopy reveals directional MutS binding to mismatched bases in DNA." Nucleic acids research **40**(12): 5448-5464.

Elez, M., A. W. Murray, L. J. Bi, X. E. Zhang, I. Matic and M. Radman (2010). "Seeing mutations in living cells." Curr Biol **20**(16): 1432-1437.

Elez, M., M. Radman and I. Matic (2012). "Stoichiometry of MutS and MutL at unrepaired mismatches in vivo suggests a mechanism of repair." Nucleic Acids Res **40**(9): 3929-3938.

Feng, G., H. C. Tsui and M. E. Winkler (1996). "Depletion of the cellular amounts of the MutS and MutH methyl-directed mismatch repair proteins in stationary-phase Escherichia coli K-12 cells." J Bacteriol **178**(8): 2388-2396.

Gorman, J., F. Wang, S. Redding, A. J. Plys, T. Fazio, S. Wind, E. E. Alani and E. C. Greene (2012). "Single-molecule imaging reveals target-search mechanisms during DNA mismatch repair." Proceedings of the National Academy of Sciences of the United States of America **109**(45): E3074-3083.

Gradia, S., S. Acharya and R. Fishel (1997). "The human mismatch recognition complex hMSH2-hMSH6 functions as a novel molecular switch." Cell **91**(7): 995-1005.

Gradia, S., D. Subramanian, T. Wilson, S. Acharya, A. Makhov, J. Griffith and R. Fishel (1999). "hMSH2-hMSH6 forms a hydrolysis-independent sliding clamp on mismatched DNA." Mol Cell **3**(2): 255-261.

Grilley, M., K. M. Welsh, S. S. Su and P. Modrich (1989). "Isolation and characterization of the Escherichia coli mutL gene product." J Biol Chem **264**(2): 1000-1004.

Groothuizen, F. S., A. Fish, M. V. Petoukhov, A. Reumer, L. Manelyte, H. H. Winterwerp, M. G. Marinus, J. H. Lebbink, D. I. Svergun, P. Friedhoff and T. K. Sixma (2013). "Using stable MutS dimers

and tetramers to quantitatively analyze DNA mismatch recognition and sliding clamp formation." Nucleic Acids Res.

Guarne, A., S. Ramon-Maiques, E. M. Wolff, R. Ghirlando, X. Hu, J. H. Miller and W. Yang (2004). "Structure of the MutL C-terminal domain: a model of intact MutL and its roles in mismatch repair." The EMBO journal **23**(21): 4134-4145.

Hall, M. C., J. R. Jordan and S. W. Matson (1998). "Evidence for a physical interaction between the Escherichia coli methyl-directed mismatch repair proteins MutL and UvrD." EMBO J **17**(5): 1535-1541.

Hegedus, E., E. Kokai, A. Kotlyar, V. Dombradi and G. Szabo (2009). "Separation of 1-23-kb complementary DNA strands by urea-agarose gel electrophoresis." Nucleic acids research **37**(17): e112.

Heo, S. D., J. K. Ku and C. Ban (2009). "Effect of E. coli MutL on the steady-state ATPase activity of MutS in the presence of short blocked end DNAs." Biochem Biophys Res Commun **385**(2): 225-229.

Hombauer, H., C. S. Campbell, C. E. Smith, A. Desai and R. D. Kolodner (2011). "Visualization of eukaryotic DNA mismatch repair reveals distinct recognition and repair intermediates." Cell **147**(5): 1040-1053.

Jeong, C., W. K. Cho, K. M. Song, C. Cook, T. Y. Yoon, C. Ban, R. Fishel and J. B. Lee (2011). "MutS switches between two fundamentally distinct clamps during mismatch repair." Nat Struct Mol Biol **18**(3): 379-385.

Jia, Y., L. Bi, F. Li, Y. Chen, C. Zhang and X. Zhang (2008). "Alpha-shaped DNA loops induced by MutS." Biochem Biophys Res Commun **372**(4): 618-622.

Jiang, Y. and P. E. Marszalek (2011). "Atomic force microscopy captures MutS tetramers initiating DNA mismatch repair." EMBO J **30**(14): 2881-2893.

Jiricny, J. (2013). "Postreplicative mismatch repair." Cold Spring Harb Perspect Biol **5**(4): a012633.

Junop, M. S., W. Yang, P. Funchain, W. Clendenin and J. H. Miller (2003). "In vitro and in vivo studies of MutS, MutL and MutH mutants: correlation of mismatch repair and DNA recombination." DNA repair **2**(4): 387-405.

Lahue, R. S., K. G. Au and P. Modrich (1989). "DNA mismatch correction in a defined system." Science (New York, N Y) **245**(4914): 160-164.

Lahue, R. S., S. S. Su and P. Modrich (1987). "Requirement for d(GATC) sequences in Escherichia coli mutHLS mismatch correction." Proceedings of the National Academy of Sciences of the United States of America **84**(6): 1482-1486.

Lamers, M. H., D. Georgijevic, J. H. Lebbink, H. H. Winterwerp, B. Agianian, N. de Wind and T. K. Sixma (2004). "ATP increases the affinity between MutS ATPase domains. Implications for ATP hydrolysis and conformational changes." The Journal of biological chemistry **279**(42): 43879-43885.

Lamers, M. H., A. Perrakis, J. H. Enzlin, H. H. Winterwerp, N. de Wind and T. K. Sixma (2000). "The crystal structure of DNA mismatch repair protein MutS binding to a G x T mismatch." Nature **407**(6805): 711-717.

Lebbink, J. H., A. Fish, A. Reumer, G. Natrajan, H. H. Winterwerp and T. K. Sixma (2010). "Magnesium coordination controls the molecular switch function of DNA mismatch repair protein MutS." The Journal of biological chemistry **285**(17): 13131-13141.

Lebbink, J. H., A. Fish, A. Reumer, G. Natrajan, H. H. Winterwerp and T. K. Sixma (2010). "Magnesium coordination controls the molecular switch function of DNA mismatch repair protein MutS." J Biol Chem **285**(17): 13131-13141.

Lebbink, J. H., D. Georgijevic, G. Natrajan, A. Fish, H. H. Winterwerp, T. K. Sixma and N. de Wind (2006). "Dual role of MutS glutamate 38 in DNA mismatch discrimination and in the authorization of repair." EMBO J **25**(2): 409-419.

Manelyte, L., C. Urbanke, L. Giron-Monzon and P. Friedhoff (2006). "Structural and functional analysis of the MutS C-terminal tetramerization domain." Nucleic Acids Res **34**(18): 5270-5279.

Marinus, M. G. and J. Casadesus (2009). "Roles of DNA adenine methylation in host-pathogen interactions: mismatch repair, transcriptional regulation, and more." FEMS Microbiol Rev **33**(3): 488-503.

Mendillo, M. L., C. D. Putnam and R. D. Kolodner (2007). "Escherichia coli MutS tetramerization domain structure reveals that stable dimers but not tetramers are essential for DNA mismatch repair in vivo." J Biol Chem **282**(22): 16345-16354.

Obmolova, G., C. Ban, P. Hsieh and W. Yang (2000). "Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA." Nature **407**(6805): 703-710.

Pluciennik, A. and P. Modrich (2007). "Protein roadblocks and helix discontinuities are barriers to the initiation of mismatch repair." Proceedings of the National Academy of Sciences of the United States of America **104**(31): 12709-12713.

Qiu, R., V. C. DeRocco, C. Harris, A. Sharma, M. M. Hingorani, D. A. Erie and K. R. Weninger (2012). "Large conformational changes in MutS during DNA scanning, mismatch recognition and repair signalling." The EMBO journal **31**(11): 2528-2540.

Schlagman, S. L., S. Hattman and M. G. Marinus (1986). "Direct role of the Escherichia coli Dam DNA methyltransferase in methylation-directed mismatch repair." J Bacteriol **165**(3): 896-900.

Schofield, M. J., S. Nayak, T. H. Scott, C. Du and P. Hsieh (2001). "Interaction of Escherichia coli MutS and MutL at a DNA mismatch." J Biol Chem **276**(30): 28291-28299.

Selmane, T., M. J. Schofield, S. Nayak, C. Du and P. Hsieh (2003). "Formation of a DNA mismatch repair complex mediated by ATP." J Mol Biol **334**(5): 949-965.

Tham, K. C., N. Hermans, H. H. Winterwerp, M. M. Cox, C. Wyman, R. Kanaar and J. H. Lebbink (2013). "Mismatch Repair Inhibits Homeologous Recombination via Coordinated Directional Unwinding of Trapped DNA Structures." Mol Cell **51**(3): 326-337.

Welsh, K. M., A. L. Lu, S. Clark and P. Modrich (1987). "Isolation and characterization of the Escherichia coli mutH gene product." J Biol Chem **262**(32): 15624-15629.
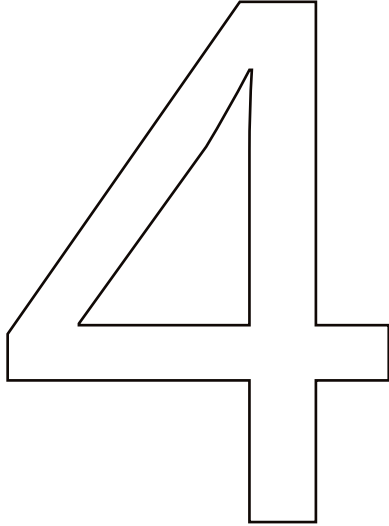
THE ROLE OF NUCLEOTIDE BINDING AND HYDROLYSIS IN MUTL DURING DNA MISMATCH REPAIR

Nicolaas Hermans, Michele Cristovão, Paulien Ikpa, Moara Lieuw Hie,
Aruna Jaddoe, Nora Goossen, Herrie H.K Winterwerp, Joyce H.G. Lebbink

ABSTRACT

DNA mismatch repair increases fidelity of replication by removing mismatches. The initiation of the mismatch repair pathway is done by two proteins, MutS and MutL, which are conserved in all three kingdoms of life. MutS can bind mismatched basepairs and small insertion loops, after which it becomes a sliding clamp on the DNA. This MutS sliding clamp can bind MutL. MutL is often described as a molecular match maker, and is able to activate the endonuclease activity of MutH and loads the helicase UvrD onto the DNA to initiate excision of the error containing strand. Nucleotide binding in MutL was shown to switch a MutL dimer from a open to a closed and more compact conformation. These conformational changes are thought to govern the interaction of MutL with DNA and other proteins in the mismatch repair machinery. We studied the interaction of MutL, and several MutL point mutants deficient in either nucleotide binding, ATP hydrolysis or DNA binding, in functional assays with MutS, MutH, UvrD and DNA. We show that nucleotide binding and hydrolysis by MutL are not needed for the interaction of MutL with MutS. However, nucleotide binding is needed for activation of MutH and UvrD, and greatly increases the affinity of MutL for DNA. Interestingly, when MutL is in a closed nucleotide bound conformation, it can interact with MutH and UvrD, but has a decreased affinity for MutS. We thus come to a model, where MutL binds MutS in an open conformation, but is able to activate downstream factors independently from MutS after being loaded onto the DNA in a closed conformation.

INTRODUCTION

The mismatch repair (MMR) system is responsible for repairing DNA base-base mismatches and small slippage errors introduced by the replication machinery. The basic repair pathway, including the key MMR proteins MutS and MutL, is highly conserved throughout all kingdoms of life (Modrich and Lahue 1996). In *Homo sapiens*, mismatch repair gene silencing and mutations in the MMR proteins MutSα (MSH2/MSH6) and MutLα (MLH1/PMS1) are linked to hereditary nonpolyposis colorectal cancer (Lynch, Lynch et al. 2008).

The first step in the MMR reaction is the recognition of a DNA mismatch by MutS. This is followed by recruitment of MutL and nucleolytic incision of the DNA daughter strand. In Escherichia coli it is the MutH endonuclease that nicks the unmethylated DNA strand at a transiently hemi-methylated GATC site. This strand incision serves as entry site for the UvrD helicase that unwinds the DNA from the nick towards the mismatch. The unwound DNA daughter strand is degraded by several nucleases after which the correct sequence is restored by DNA resynthesis and ligation ((Jiricny 2013) for review).

MutS and its homologs are mismatch-controlled, ATP-operated molecular switches (Gradia, Acharya et al. 1997 2007, Acharya, Foster et al. 2003, Lebbink, Fish et al. 2010) that belong to the ATP binding cassette (ABC) superfamily. These proteins rely on nucleotide-induced conformational changes to fulfill their job. In order to find mismatches, MutS encircles the DNA (Acharya, Foster et al. 2003), testing the DNA for sites of higher flexibility indicative of lesions (Natrajan, Lamers et al. 2003, Gupta, Gellert et al. 2012). Binding induces a sharp kink in the DNA which allows MutS to insert one of its mismatch-binding domains onto the mismatched bases (Lamers, Perrakis et al. 2000, Obmolova, Ban et al. 2000, Lamers, Winterwerp et al. 2003). This triggers uptake of ATP (Gradia, Acharya et al. 1997, Acharya, Foster et al. 2003), resulting in a large conformational change into a stable clamp-like structure that releases the mismatch and diffuses along the DNA (Gradia, Acharya et al. 1997, Blackwell, Martik et al. 1998, Iaccarino, Marra et al. 2000, Junop, Obmolova et al. 2001, Lebbink, Fish et al. 2010, Jeong, Cho et al. 2011, Gorman, Wang et al. 2012, Qiu, DeRocco et al. 2012).

The ATP-bound MutS sliding clamp can recruit MutL (Acharya, Foster et al. 2003, Gorman, Wang et al. 2012, Sharma, Doucette et al. 2013). MutL belongs to the GHKL family of ATPases (Ban, Junop et al. 1999, Dutta and Inouye 2000). The ATPase activity of MutL is necessary for MMR in vitro (Spampinato and Modrich 2000) and in vivo (Aronshtam and Marinus 1996, Ban and Yang 1998, Ban, Junop et al. 1999). Mutations in MutLα that lead to an increase in genome instability and cancer occur mainly in the ATPase domain of PMS2, showing the importance of ATP binding and hydrolysis (Ban and Yang 1998, Raschle, Dufner et al. 2002). Indeed, MutL has also been considered an ATP-operated molecular

switch (Acharya 2003; Ban 1999). MutL dimerizes via its C-terminal domain. The ATPase of MutL resides in its N-terminal domain (Ban, Junop et al. 1999), which is connected to the C-terminal domain by a flexible linker (Ban, Junop et al. 1999, Guarne, Ramon-Maiques et al. 2004). When the N-terminal domain is bound to ATP, conformational changes in the N-terminal part of MutL change its structure from an open form to a more compact and closed form (Ban and Yang 1998, Sacho, Kadyrov et al. 2008, Niedziela-Majka, Maluf et al. 2011). These conformational changes in MutLα were also revealed in Scanning Force Microscopy (SFM) images (Sacho, Kadyrov et al. 2008).

It is evident that the MutL ATPase cycle and the conformational changes that follow ATP binding and release lay at the core of its interaction with MutH and UvrD, and are therefore key to our understanding of strand discrimination and MMR as a whole (Ban, Junop et al. 1999, Acharya, Foster et al. 2003, Junop, Yang et al. 2003). The importance of ATP binding and hydrolysis by MutL for MutH and UvrD activation has been studied extensively (Ban and Yang 1998, Ban, Junop et al. 1999, Acharya, Foster et al. 2003, Junop, Yang et al. 2003, Guarne, Ramon-Maiques et al. 2004, Robertson, Pattishall et al. 2006). However, due to differences in reaction conditions and in some cases absence of controls, the exact role of ATP hydrolysis by MutL in the MutS-dependent reaction remains unresolved.

Here, we address the ATPase activity of MutL from different points of view: recruitment by MutS, conformational changes in MutL, activation of MutH and loading of UvrD. We chose to analyze five key MutL mutants, deficient in at least one of the activities of MutL. To better understand the importance of ATP to this system, we employed different complementary techniques that focused on the different activities of MutS, MutL, MutH and UvrD, while regarding ATP as the major player. We consolidate our findings and those of others into a comprehensive model that helps to understand how MMR is regulated.

RESULTS

Binding and hydrolysis of ATP by MutL control its open/closed conformation

To study in detail the role of nucleotide binding and hydrolysis by MutL in MMR, we purified wild type MutL and five mutant variants that are affected in different steps of the ATPase cycle. These included N33A (presumably defective in ATP binding, (Guarne, Ramon-Maiques et al. 2004)), E29A (defective in ATP hydrolysis, (Ban, Junop et al. 1999, Acharya, Foster et al. 2003), N302A and K307A (decreased ATPase activity, (Ban, Junop et al. 1999, Junop, Yang et al. 2003)) and R266E (ATPase cannot be stimulated by DNA, (Ban and Yang 1998)).

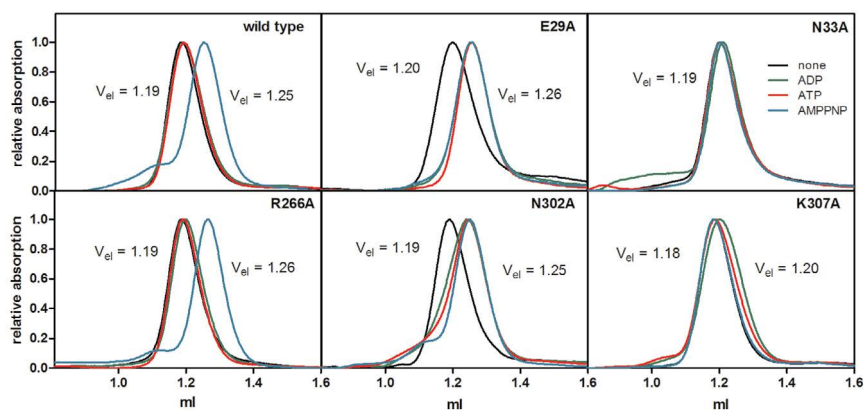We analyzed if the changes in the ATPase activity also influence the

Figure 1: The MutL conformation is regulated by nucleotide binding and hydrolysis. Size exclusion chromatography reveals a fast-eluting MutL species (elution volume (V^{el}) around 1.19 ml, assigned to the 'open' form) and a slow eluting MutL species (V^{el} around 1.25 ml, assigned to the 'closed' form), the distribution of which depends on the bound nucleotide (none, ADP, ATP or AMP-PNP) and the presence of single amino acid substitutions in the nucleotide binding site.

conformation of MutL using size exclusion chromatography. In this assay the closed state of MutL, in which the N-terminal domains have dimerized in response to ATP binding, elutes faster than the open state (Ban, Junop et al. 1999, Guarne, Ramon-Maiques et al. 2004). The MutL variants were incubated overnight with 1 mM ADP, ATP or AMP-PNP (a non-hydrolysable analog of ATP), or in the absence of nucleotide, and their elution profiles on a Superdex 200 size exclusion column was monitored (Figure 1). In agreement with previous reports (Ban, Junop et al. 1999, Guarne, Ramon-Maiques et al. 2004), wild type MutL incubated with AMP-PNP eluted later (around 1.3 ml elution volume) than MutL without nucleotide (around 1.2 ml), in agreement with the nucleotide-bound closed conformation being more compact than the open conformation. Upon incubation with ADP and ATP, MutL eluted as if in the open conformation, indicating the wild type protein is only closed when bound to nucleotide triphosphate which cannot be hydrolysed. Mutation of the residue involved in DNA-dependent modulation of the ATPase activity does not change this, the elution profiles of the R266E variant are identical to that of wild type MutL.

Contrasting this, MutL N33A is always open, consistent with its reported inability to bind nucleotide (Guarne, Ramon-Maiques et al. 2004). The K307A MutL variant behaved exactly like N33A MutL. K307A has a reduced nucleotide binding affinity (Ban, Junop et al. 1999), and while it was reported to close in the presence of AMP-PNP (Junop, Yang et al. 2003), in our assay conditions it
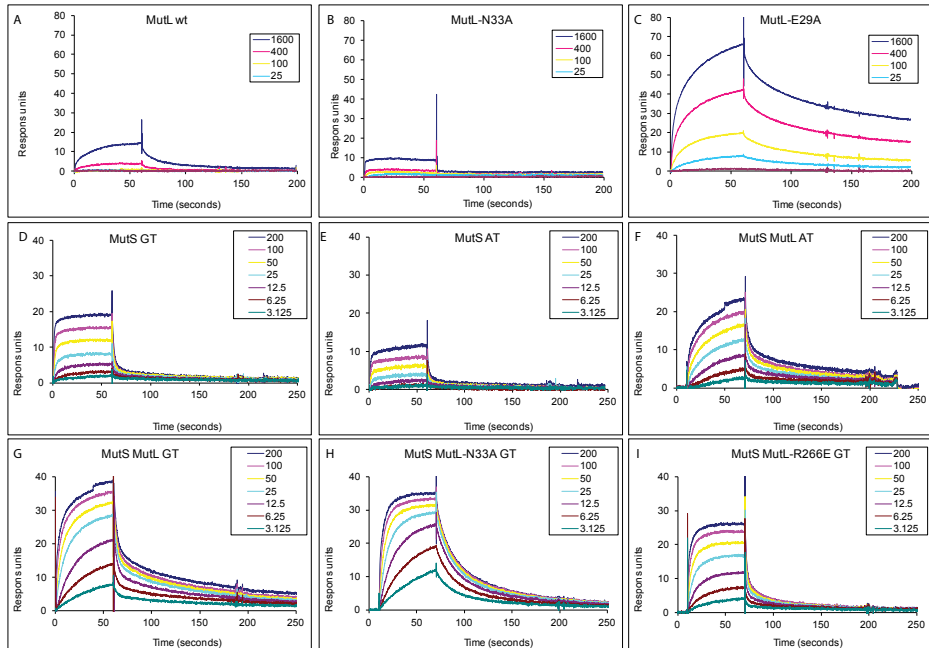
Figure 2: DNA binding by MutL and MutSL complexes. In panel A, B and C the binding of wild type MutL, E29A and N33A respectively to 100 bp DNA homoduplex in the presence of ATP and absence of MutS. Panel D shows the binding of MutS-CF-D835R to 100 bp DNA with a single GT mismatch in the presence of ATP. The concentration of MutS was increased from 6.125 nM to 200 nM (colored lines). In B and C binding of MutS-CF-D835R (E) and MutS-CF-D835R -MutL (F) to a 100 bp homoduplex. The concentration of MutL was kept constant at 200 nM, the concentration MutS-CF-D835R was varied as in Panel A. In (D, E, F) binding of 200 nM MutL, MutL N33A and MutL R266E respectively to 100 bp mismatch-containing DNA in the presence of ATP and MutS-CF-D835R. The concentration of MutL was kept constant at 200 nM, the concentration MutS was varied as in Panel A.

is predominantly in the open conformation. MutL E29A is open in the absence of nucleotide, but closes in the presence of ADP, ATP and AMP-PNP, confirming that this variant can bind but not hydrolyse ATP (Ban and Yang 1998, Acharya, Foster et al. 2003). Interestingly, MutL N302A exhibited a similar profile to MutL E29A, thus despite the ability of this variant to hydrolyze ATP (Junop, Yang et al. 2003) it is predominantly in the closed conformation. Remarkably, both MutL E29A and N302A are also closed in the presence of ADP, indicating it is not the presence of the gamma-phosphate but rather nucleotide in general that is required for closure of the N-terminal domains.

DNA binding by MutL is involved in complex formation with MutS on mismatched DNA

To investigate the interaction of MutS and MutL variants with DNA, we performed surface plasmon resonance (SPR) experiments. We first analyzed the binding of wild type, E29A and N33A MutL to double stranded (ds) DNA in the presence of ATP. To allow direct comparison with other functional assays, we kept physiologically relevant salt concentrations (150 mM KCl). Increasing response units (RU) indicates binding of MutL to DNA. Under the conditions used, wild type MutL has a low affinity for DNA (Figure 2A). Under similar conditions, the apparent binding constant of MutL for an 18-bp duplex possessing a 3'-dT20 tail was found to be relatively low (34 (±4.3) × 10^4 M−1 (Niedziela-Majka, Maluf et al. 2011)). It was reported previously that MutL displays enhanced single stranded DNA (ssDNA) binding activity in the presence of AMP-PNP (Ban and Yang 1998, Mechanic, Frankel et al. 2000, Robertson, Pattishall et al. 2006, Niedziela-Majka, Maluf et al. 2011). Similarly, in the presence of ATP, MutL E29A (which is unable to hydrolyze this nucleotide) has a higher affinity for double stranded DNA (dsDNA) than wild type MutL (Figure 2C). This is unsurprising because of the similar behavior of E29A with ATP and wild type MutL with AMP-PNP during size exclusion chromatography (Figure 1). We conclude that the closed, ATP-bound form of MutL has the highest affinity for both ssDNA and dsDNA.

We also investigated the interaction of MutS and MutL on DNA using the same reaction conditions as for binding of MutL alone. First, we analyzed the binding of MutS-CF-D835R to a 100bp double-stranded duplex containing a central GT mismatch in the presence of ATP (Figure 2D). This obligate dimeric variant (Manelyte, Urbanke et al. 2006, Cristovao, Sisamakis et al. 2012) was used to avoid contributions to the SPR response curves caused by MutS tetramer formation (Groothuizen, Fish et al. 2013). We obtained a K_d of 75 nM by fitting a 1:1 binding model to the data. This affinity constant is only slightly higher than the reported K_d of 37 nM for MutS-CF-D835R binding to a 21 bp oligo with a single mismatch (Groothuizen, Fish et al. 2013), which is probably due to our DNA substrate being substantially longer (to accommodate for MutL binding in subsequent experiments). Next, we analyzed the binding of MutS-CF-D835R to DNA in the presence of 200 nM MutL (Figure 2G-H). MutS-MutL shows increased binding to the homoduplex DNA (Figure 2F), compared to MutS alone (Figure 2B). When a GT mismatch is present, wild type MutL, which by itself does not bind to DNA at this concentration (Figure 2A), increased the binding of protein to the DNA about two- fold compared to MutS alone (Figure 2G), indicating that a specific complex between MutS and MutL is formed. This is further supported by the observed reduced dissociation rate of a fraction of the bound material. Because only a fraction of the bound material reduces more slowly, these experiments cannot be analyzed using a 1:1 binding model. However, it is clear that DNA binding by MutL does play a role in this specific complex for-

mation, because the response during the association phase in the presence of MutL R266E is significantly reduced compared to wild type MutL and the dissociation phase with this variant closely resembles that of MutS alone (Figure 2D). Interestingly, complex formation with MutL N33A is at least as good as for wild type MutL, while the fraction of stably bound material is significantly larger, from which we conclude that ATP binding in MutL is not required to bind MutS but does seem to regulate stability of the complex on DNA (Figure 2F-G). Complex formation between MutS and MutL E29A could not be analyzed in this setup because this MutL variant binds to the DNA by itself under these assay conditions (Figure 2C).

Closed, nucleotide bound MutL can activate MutH in the absence of MutS

We studied the effect of the mutations on the ability of MutL to activate MutH. For this, we generated circular DNA substrates containing a single GT mismatch and two GATC sites (GT#2) using primer extension on phagemid DNA (Chapter 2). Since the initial phagemid DNA was methylated, the resulting DNA substrates used in our assay are hemimethylated. We incorporated a fluorophore, Alexa647, 14 bp from the mismatch, to facilitate detection (Figure 3A). The fluorophore is not recognized as a mismatch, and incision is dependent on the presence of a mismatch, MutS, MutL, MutH and ATP (Chapter 2). Upon activation by MutS and MutL, MutH can nick the closed circular DNA at a hemi-methylated GATC site and converting it to nicked (open) circular DNA.

In the presence of 10 nM MutS, 10 nM MutL, 5 nM MutH and 1mM ATP, most of the closed circular DNA was nicked by MutH after 1 minute of incubation (Figure 3B). After 10 minutes, the substrate in the reactions containing MutL R266E, MutL N302A and to a lesser extent MutL K307A also showed incision. Reactions containing either MutL E29A or MutL N33A did not show MutH activation even after 10 minutes incubation. This indicates that both ATP binding and hydrolysis is required for mismatch- and MutS-dependent MutH activation.

MutL activates MutH in the absence of MutS and a mismatch (Hall and Matson 1999), at relatively low salt concentrations (20 mM NaCl). This stimulation is increased by the addition of ATP. We incubated 50 nM of wild type or mutant MutL with 2 nM MutH and 1 mM ATP at 80 mM KCl, and analyzed incision of covalently closed circles after 10 and 30 minutes (Figure 3C). Under our assay conditions, wild type MutL did not activate MutH in the absence of MutS. Likewise MutL variants N33A and R266E did not activate MutH, indicating that both ATP binding and DNA binding by MutL is required for MutH activation in the absence of MutS. Interestingly, MutL N302A and K307A, and especially MutL E29A, efficiently activated MutH. From this we conclude that in addition to ATP binding, slow or absent ATP hydrolysis is necessary to observe MutS-indepen-
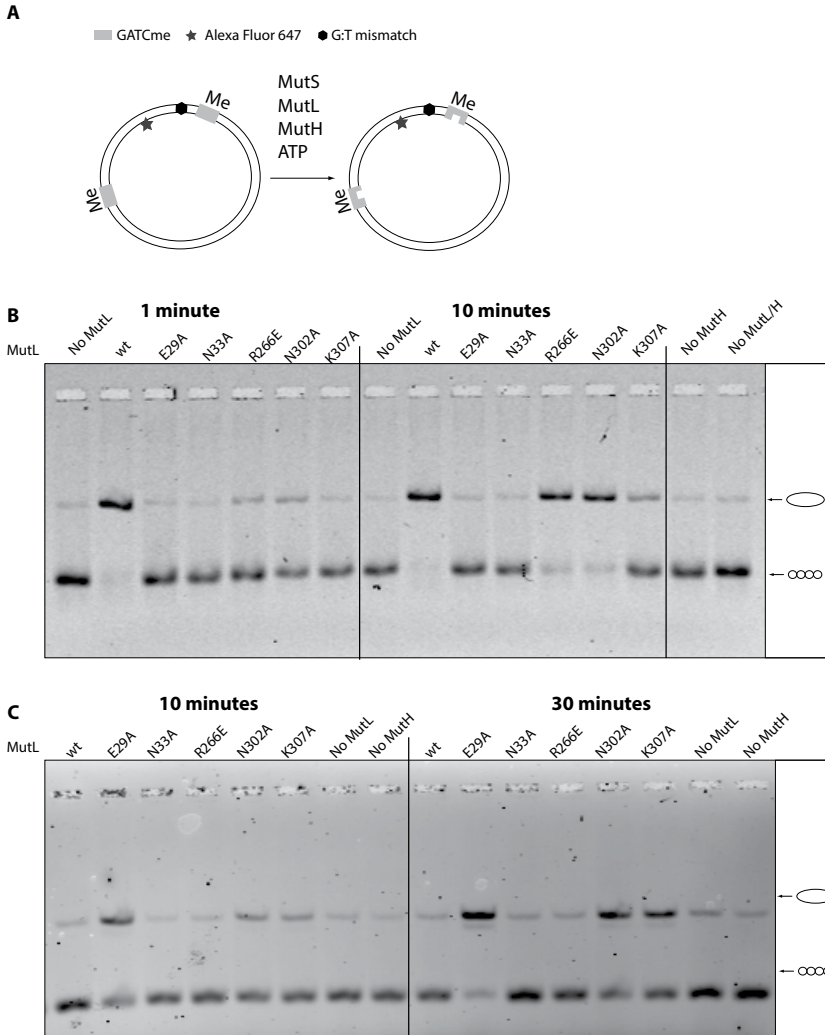
Figure 3: MutS-dependent and -independent activation of MutH by MutL. A. Schematic representation of the DNA nicking reaction. The addition of MutS, MutL, MutH and ATP to circular DNA with two GATC sites converts covalently closed DNA (fast running species in the gels) into an open circular DNA product (slow running species in the gels). B. In the presence of MutS, wild type MutL activates MutH within the 1st minute, whereas MutL R266E, N302A and K307A show activation only after 10 minutes. MutL E29A and N33A do not show significant MutH activation within the time interval tested. C. The nicking reaction was repeated in the absence of MutS with 50 nM MutL and 2 nM MutH at 80 mM KCl instead of 150 mM. Reactions were stopped after 10 minutes (left side) and 30 minutes (right side). In a buffer with lower ionic strength, MutL E29A can activate MutH in the absence of MutS. Also MutL N302A and MutL K307A display some activation of MutH. Wild type MutL does not activate MutH under these conditions.
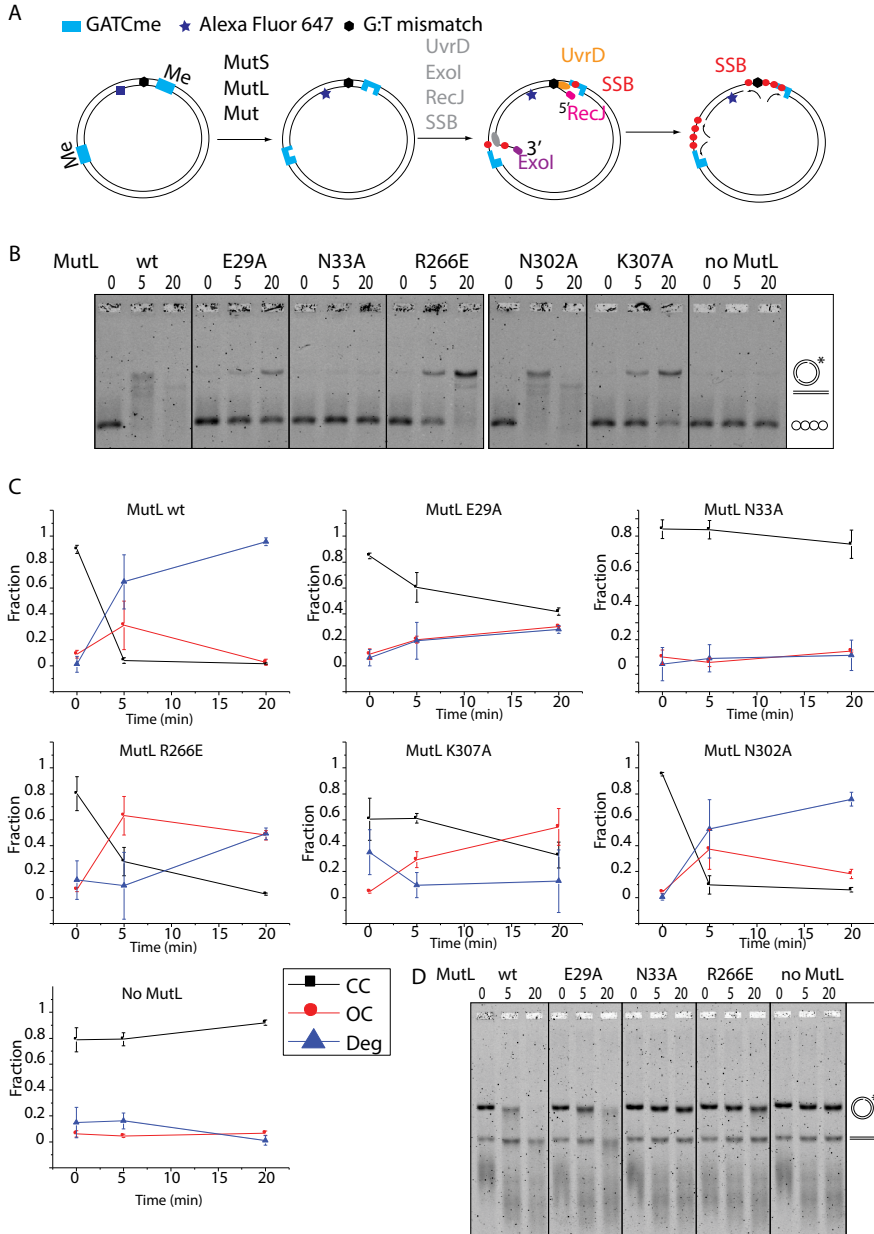
Figure 4: Stimulation of DNA unwinding and excision by MutL. A) Schematic representation of DNA nicking (MutH activation) and subsequent DNA unwinding (UvrD activation) and excision (by RecJ and ExoI). B) The MutL mutants were tested for their ability to activate MutH and UvrD on mismatched DNA within the same reaction in the presence of MutS. The panel on the right shows a control reaction lacking MutL. C) The appearance and disappearance of the DNA bands is plotted over time: closed circular DNA (CC, black line), open circular DNA (OC, red line) and degraded DNA (deg, blue line). D) The unwinding assay repeated in the absence of MutH, on pre-nicked DNA.

dent MutH activation.

Activation of UvrD requires DNA binding by MutL

To study the excision of the mismatch we assessed the activation of UvrD by MutL and the MutL mutants. We extended our MutH activation assay to include UvrD, ExoI, RecJ and SSB. In this UvrD activation assay, MutS and MutL activate MutH to create a nick, and subsequently UvrD is loaded onto the nicks to unwind the DNA in the direction of the mismatch (Figure 4). To prevent re-annealing of the DNA, we included the 3'-5' exonuclease ExoI, 5'-3' exonuclease RecJ and single stranded DNA binding protein (SSB) (all were shown to be involved in MMR, for review see (Jiricny 2013)) which results in degradation of the open circular DNA.

Under our assay conditions MutH activation (disappearance of the closed circular DNA) and subsequent UvrD activation (disappearance of the open circular DNA) by wild type MutL occurs within several minutes (Figure 4B). Because MutH activation by MutL E29A is largely absent and N33A is completely inactive (Figure 3), UvrD activation is not observed for these variants. MutL R266E did activate MutH, but UvrD was not activated by MutL R266E (the nicked product is formed but does not disappear over time). MutL N302 is clearly active in both MutH and UvrD activation, albeit not as active as MutL wt. MutL K307 also shows low activity compared to wild type MutL.

To uncouple MutH and UvrD activation, we pre-nicked the circular DNA substrate with wild type MutSLH and heat inactivated the proteins at 70°C for 10 minutes. We then initiated UvrD unwinding on the nicked substrates in the presence of MutS, ExoI, RecJ, SSB, and MutL wild type and the MutL mutants (Figure 4D). These experiments show MutH is not needed for efficient loading of UvrD, since the substrate is unwound as readily in the presence of MutH as without MutH. Interestingly, we found that MutL E29A is perfectly able to activate UvrD, albeit at a slightly slower rate that wild type MutL. In contrast, MutL N33A did not activate UvrD, which is in line with the failure of MutL N33A to activate MutH. This suggests that the closed nucleotide bound conformation of MutL activates both MutH and UvrD. Unsurprisingly, MutL R266E failed to activate UvrD, which was shown for this MutL mutant before (Guarne, Ramon-Maiques et al. 2004).

DISCUSSION

It is well established that the ATPase function of MutL is indispensable for DNA mismatch repair. A wealth of literature is available on the function of MutL, however the detailed molecular mechanism remains unresolved and controversies exist. To clarify some of these issues, we have performed functional

assays and controls under physiologically relevant conditions that allow direct comparison between different assays and MutL variants defective in different steps of the ATPase cycle. Reported literature data for the ATPase activity of these mutants is summarized in table 1, along with most of our findings. To come to a mechanistic model for the ATPase cycle of MutL, we focused on three major points; 1) the interaction of MutL with MutS, 2) the interaction of MutL with DNA and 3) the interaction of MutL with MutH and UvrD.

MutL variant	k_{cat} (min^{-1})	Form (+ATP)	DNA binding (no MutS)	DNA binding (+ MutS)	MutH activation (-MutS)	MutH activation (+ MutS)	UvrD activation (+MutS)
wt	0.04[1] / 0.4[2]	open	n	y	n	y	y
E29A	0[1,2]	closed	y	n/a	y	n	y
N33A	nd	open	n/a	y	n	n	n
R266E	0.65[2]	open	n	y	n	y	n
N302A	0.007[1] / 0.16[2]	closed	n/a	y	y	y	y
K307A	0.02[1,2]	open	n/a	y	y	y	y

Table 1: overview of activities of wild type and mutant MutL. Steady state rates for ATP hydrolysis (k_{cat}) are taken from literature ([1]Ban, 1999; [2]Junop 2003). Size exclusion chromatography was used to determine whether MutL and its variants are in the open or closed conformation in the presence of ATP. DNA binding in the presence of ATP, and with and without MutS being present simultaneously, was addressed using SPR. MutH activation, with and without MutS, was analyzed by conversion of covalently closed, hemimethylated, mismatched DNA circles into nicked circles. Subsequently, UvrD activation in the presence of MutS was scored by unwinding and degradation of these nicked substrates. n/a = not analyzed; nd = not detectable.

To initiate its task in MMR, MutL first needs to interact with MutS. When MutL was first purified, it was shown that MutL increased the DNaseI footprint of MutS on DNA in the presence of ATP (Lahue, Su et al. 1987, Selmane, Schofield et al. 2003). MutL interacts with the MutS sliding clamp. Also MutL-E32K, which does not bind ATP, can bind to MutS (Spampinato and Modrich 2000). We found that MutL N33A, which is in the open conformation and does not bind ATP, binds MutS similar to wild type MutL. This shows that nucleotide binding by the MutL ATPase domains is not needed for the interaction of MutL with MutS.

DNA binding by MutL is likely only relevant in the presence of MutS, since the affinity of wild-type MutL for DNA is low under physiological salt concentrations (Niedziela-Majka, Maluf et al. 2011). Interestingly, MutL displays enhanced ssDNA binding activity in the presence of the non-hydrolyzable ATP analogue AMP-PNP (Ban, Junop et al. 1999, Mechanic, Frankel et al. 2000, Robertson, Pattishall et al. 2006, Niedziela-Majka, Maluf et al. 2011). In the presence of AMP-PNP, MutL wild type is in a closed conformation (Ban, Junop et al. 1999). We analyzed binding to dsDNA, which is the DNA form that is initially
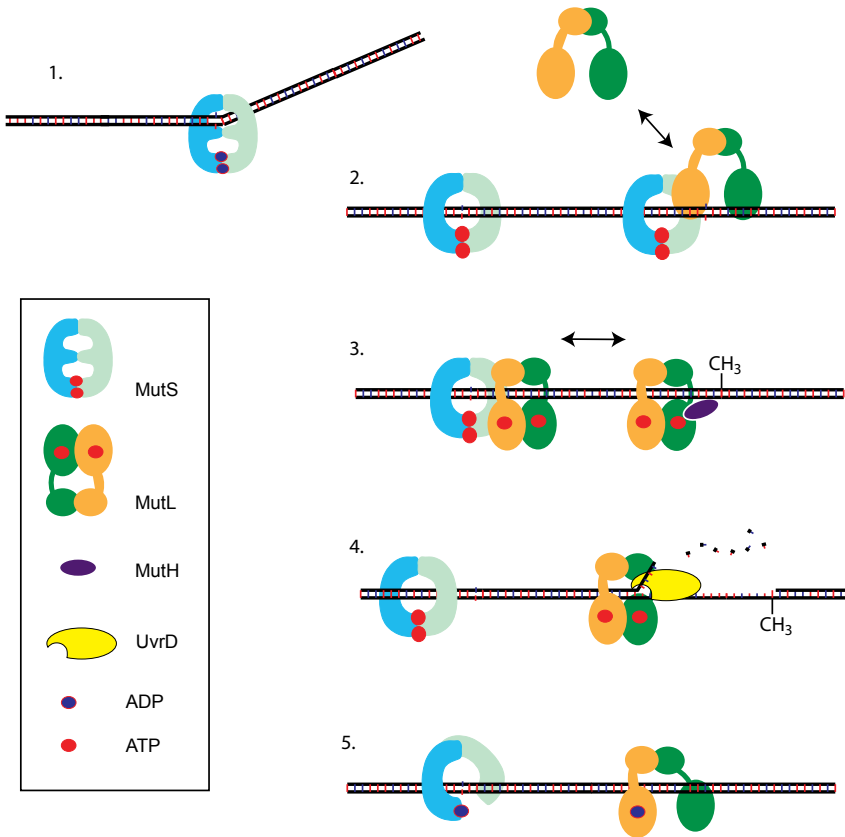
Figure 5: Function of the MutL ATP cycle in MMR. 1) After binding a mismatch, MutS forms a sliding clamp on the DNA. 2) MutL can bind to this MutS sliding clamp in an open conformation. 3) After binding nucleotide, a second dimerization interface forms in the N-terminal domain, and MutL closes forming a clamp around the DNA. 4) MutL can activate MutH at a GATC site, or load UvrD onto a single nick. The MutL sliding clamp can disconnect from MutS without dissociating from the DNA. After ATP hydrolysis, loss of nucleotide from the binding pocket in MutL will disrupt the clamp, and MutL can leave the DNA.

encountered by MutL before MutH activation. We find that MutL E29A in the presence of ATP is in a closed conformation (clearly indicating this variant is able to bind nucleotide (Robertson 2006, and unlike Acharya 2003), has a much higher affinity for dsDNA than wild type MutL. Thus the closed conformation of MutL not only has increased affinity for ssDNA (which will be encountered after UvrD activation) but also for dsDNA (which is encountered upon complex formation with MutS).

After binding MutS, MutL needs to activate MutH and UvrD. In literature, it has been proposed that (depending on assay conditions, especially salt concentrations) the closed, ATP bound form can activate MutH (Junop, Yang et

al. 2003, Guarne, Ramon-Maiques et al. 2004, Robertson, Pattishall et al. 2006). Wild type MutL can activate MutH independently from MutS, in the presence of ATP and AMP-PNP (Ban and Yang 1998, Hall, Jordan et al. 1998). Later, it was shown that MutL E29A, which is always in a closed conformation in the presence of ATP, can activate MutH (Junop, Yang et al. 2003, Guarne, Ramon-Maiques et al. 2004, Robertson, Pattishall et al. 2006). Also MutL N302A, prevalently in the closed conformation, was found to activate MutH (Acharya, Foster et al. 2003). Likewise UvrD can be activated by both MutL E29A and MutL N302A (Guarne, Ramon-Maiques et al. 2004). In our experiments, MutL N33A cannot activate the endonuclease of MutH nor UvrD. All these observations reinforced the notion that the closed, ATP-bound form of MutL activates both MutH and UvrD, and that a defect in ATP binding results in failure of the MutL N-terminal domains to close and activate MutH and UvrD.

However, *in vivo* mutL null strains complemented by MutL E29A or MutL N302A show a 100 fold increase in mutation frequency compared to wild type MutL (Acharya, Foster et al. 2003, Junop, Yang et al. 2003, Robertson, Pattishall et al. 2006). This suggests that even though MutL E29A and MutL N302A are fully functional in activation of MutH and UvrD, they are dysfunctional in other aspects of the MMR mechanism, or are mutagenic themselves by over-activating MutH and/or UvrD. We are in favor of the first explanation, since in our MutS- and mismatch-dependent reaction, MutL wild type is much more efficient than any of the mutant proteins in both incision and unwinding. MutL N302A does show activity, but only at a small fraction of the activity of wild type MutL. MutL E29A does not show any nicking activity under our assay conditions at physiological salt concentrations. This all indicates that none of the mutants is over-active compared to the wild type MutL. The differences between our results and the literature can partially be explained by different assay conditions. Our physiologically relevant salt concentration (150 mM KCl) is relatively high compared to the concentrations used previously (20 mM NaCl (Robertson, Pattishall et al. 2006), 90 mM KCl (Junop, Yang et al. 2003)). Buffer conditions and salt concentration are major determinants for MutL binding to DNA (Niedziela-Majka, Maluf et al. 2011). At relatively low salt concentrations (20 or 80 mM) the closed ATP-bound forms of MutL can interact with DNA and activate MutH. However, at physiological salt concentrations, the affinity for DNA is low and MutL relies on its interaction with MutS to be recruited to the DNA. From our SPR analysis it becomes apparent that the open form of MutL (such as N33A) is recruited efficiently to the DNA by MutS. Likely, MutL E29A has a defect in its interaction with MutS, even though this is hard to measure directly since it has a stronger affinity for DNA. However, this results in the failure to activate MutH at higher salt concentration, also in the presence of MutS, and would explain the mutator phenotype of the E29A and N302A mutations *in vivo*.

In summary we propose the following model (Figure 5). First, MutL binds the MutS sliding clamp in its open conformation. Upon nucleotide binding, MutL undergoes a large conformational change and closes. This closed conformation has a higher affinity for DNA and is needed to activate MutH and UvrD. However, it has a lower affinity for MutS, and might thus dissociate from MutS while remaining on the DNA. This reinforces the interesting hypothesis that MutL can activate MutH and UvrD in the absence of MutS, after being loaded onto the DNA by MutS as a clamp, as has been suggested previously (Ban, Junop et al. 1999, Acharya, Foster et al. 2003). Also, it explains some of the controversy in literature why ATP hydrolysis in MutL is required for functional MMR; ATP hydrolysis by MutL likely serves the same purpose as in MutS by allowing dissociation from the DNA and regeneration of the form capable to rebind to its relevant ligand (*i.e.* the mismatch in the case of MutS and activated MutS in the case of MutL).

Several experiments can be performed to further investigate the possibility of an uncoupling of MutS and MutL after MutL is loaded onto the DNA. It would be interesting to know if a double mutant of MutL E29A and R266E would interact with MutS, and would still be proficient in its activation of MutH. Furthermore, DNA binding residues in MutL N33A might not be required for its complex formation with MutS. To directly prove the uncoupling of a MutL sliding clamp from MutS, single molecule techniques are likely needed. In the last few years, great advances have been made in fluorescently labeling of individual components of the mismatch repair system (Gorman, Plys et al. 2010, Cho, Jeong et al. 2012, Cristovao, Sisamakis et al. 2012, Yokota, Chujo et al. 2013). In mammalian MMR, MutL is structurally and functionally conserved to a large extent, (chapter 2 and (Guarne 2012) for review). It is therefore not unlikely that some of our findings also apply to mammalian MMR.

MATERIALS AND METHODS

Buffers were made with reagent grade chemicals and distilled water that was further deionized by treatment with a Milli-Q water purification system (Millipore Corp., Bedford, MA). For the functional assays, two buffers were used. A buffer containing 25 mM Hepes KOH pH 7.5, 5 mM MgCl, 10% glycerol (v/v), 100 ng / µl BSA, 1 mM DTT and 150 mM KCl (Buffer H) was used for the nicking assay containing MutS, MutL and MutH, and for all the excision assays. A second buffer, containing 25 mM Hepes KOH pH 7.5, 5 mM $MgCl_2$, 10% glycerol (v/v), 1 mg/ml BSA, 1 mM DTT and 80 mM KCl (Buffer L) was used for the nicking assays in the absence of MutS. Higher KCl concentrations decrease MutL-DNA interactions, and prevent MutS and mismatch dependent activation of MutL and MutH. Buffer used in the SPR experiments contained 25 mM Hepes-KOH (pH 7.5), 150 mM KCl, 5 mM $MgCl_2$, 10% glycerol (v/v), 0.05% TWEEN-20, 1mM

DTT, 1 mM ATP. Buffer used for size exclusion chromatography consisted of 25 mM Hepes-KOH (pH 7.5), 150 mM KCl, 5 mM $MgCl_2$, 10% glycerol, 10 mM β-mercaptothanol (Buffer S).

Protein expression and purification

Wild type MutS was expressed and purified as described (Natrajan, Lamers et al. 2003, Tham, Hermans et al. 2013). MutS-CF-D835R was expressed and purified as described (Cristovao, Sisamakis et al. 2012). MutL variant N33A, E29A, K307A, N302A and R266E were constructed from the wild type expression plasmid pTX418 (Feng and Winkler 1995) using QuikChange (Stratagene) and their sequence was verified. MutL, MutL N33A, MutL R266E, MutL N302A and MutL K307A were purified as described (Lebbink et al, 2010, Tham et al 2013). MutL E29A was expressed in BL21(DE3) cells in which the *mutL* gene was inactivated using tn10 transduction (courtesy of Nora Goossen and Moara Lie-Hieuw). MutH was purified using pTX417 (Feng and Winkler 1995) as described in chapter 2. UvrD was purified using pET11d-UvrD (George, Brosh et al. 1994) as described (Guarne, Ramon-Maiques et al. 2004, Tham, Hermans et al. 2013). Proteins were flash-frozen in liquid nitrogen and stored at -80°C in 10 or 50 µl-aliquots.

DNA Substrates

Circular DNA substrates with a single GT mismatch for use in the MutH and UvrD activation assays were produced by extending a primer on a circular single stranded DNA (ssDNA) phagemid of 3196 bases, as described previously (Chapter 2). This substrate contained two hemi-methylated GATC sites, one at 29 bp and the other 1042 bp from the GT mismatch. When nicked substrates were used, mismatched DNA was pre-nicked in buffer H in the presence of 5 nM MutS, 5 nM MutL, 2.5 nM MutH and 1 mM ATP for 10 minutes at 37°C, and subsequently heat inactivated at 80°C for 10 minutes.

Size exclusion chromatography

Wild type MutL or mutant variants (0.25 mg/ml) without nucleotide or mixed with 1 mM ADP, ATP, AMP-PNP were incubated o/n at 4°C in buffer S and subsequently injected onto a Superdex 200 PC 3.2/30 column operated by a SMART system (GE Healthcare) equilibrated in buffer S. Elution profiles monitored at 280 nm were normalized, overlaid and analyzed using Graphpad Prism.

SPR experiments

Surface plasmon resonance (SPR) measurements were performed using a Biacore T100 (GE Healthcare) at 25°C. To monitor MutS MutL complex formation, a 100-bp duplex containing a central GT mismatch was created by annealing the following oligonucleotides: 5'-biotin-AAACAGGCTTAGGCTGGAGCTGAAGCT GAGCTTAGGTTCATCGAGGTTCGAGCTCGGAGCAATTGATCGGTACCCAATTCGCCC CAGCGGCATCCAGGTT and 5'-AACCTGGATGCCG-CTGGGGCGAATTGGGTACCGA TCAATTGCTCCGAGCTCGAACCTCGATGAACCTAAGCTTAGCTTCAGCTCCAGCCTAA GCCTGTTT (nucleotides forming the mismatch indicated in bold). To monitor MutL binding, complementary 70 nt oligonucleotides were annealed (5'-AAC CTGGATGCCGCTGGGGCGAATTGGGTACCGATCAATTGCTCCGAGCTCGAACCTCGA TGAACCTAAGCTCAGCTTCAGCTCCAGCCTAAGCCTGTTT). For each experiment, DNA was immobilized on a Biacore CM5 chip derivatized with streptavidin to a maximum total signal of 10 RU. In between experiments, the chip surface was stripped of DNA substrates using micrococcal Nuclease (NEB). To monitor MutS MutL complex formation, increasing concentrations of MutS D835R (3.1-200 nM final) were mixed with a fixed concentration of MutL (200 nM final) in SPR buffer and injected over the chip surface. To monitor MutL binding, increasing concentrations of MutL (6.1-1600 nM) in SPR buffer were injected. Protein flow was maintained for 60 s, after which only SPR buffer was flown over for 120s. In between different protein injections, the chip surface was regenerated by injection of 0.05% SDS for 60 seconds. Measurements were performed in duplo. The K_d of MutS-D835R was determined using Evilfit software (Gorshkova, Svitel et al. 2008).

Functional assays

Nicking activity of MutH in the presence of MutS and MutL was assayed in buffer H supplemented with 1 mM ATP. Unless stated otherwise, nicking assays were performed using 5 nM MutS and MutL monomers, 2.5 nM MutH, and 1 nM DNA circles (final concentrations). MutS-independent nicking was addressed by omitting MutS from reactions performed in buffer L, on homo-duplexed DNA. Unwinding and excision reactions were performed in reaction buffer H, in the presence of 5 nM MutS, 5 nM MutL, 2.5 nM MutH, 5 nM UvrD, 200 nM SSB (Promega), 0.1 u ExoI (New England Biolabs) and 2 mM ATP, unless stated otherwise. In all cases, reactions were initiated by adding ATP to a mixture containing all proteins and DNA. At the indicated times, 10 μL samples of the reaction were mixed with an equal volume of 20% glycerol, 1% SDS and 50 mM EDTA. Samples were loaded onto 1% agarose gels supplemented with 50 μM ethidium bromide, and run in 1x TAE. Results were visualized by monitoring the fluorescence of the Alexa647 fluorophore on a Typhoon9100 imager (GE Healthcare).

REFERENCES

Acharya, S., P. L. Foster, P. Brooks and R. Fishel (2003). "The coordinated functions of the E. coli MutS and MutL proteins in mismatch repair." Molecular cell 12(1): 233-246.

Aronshtam, A. and M. G. Marinus (1996). "Dominant negative mutator mutations in the mutL gene of Escherichia coli." Nucleic acids research 24(13): 2498-2504.

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

Ban, C. and W. Yang (1998). "Crystal structure and ATPase activity of MutL: implications for DNA repair and mutagenesis." Cell 95(4): 541-552.

Blackwell, L. J., D. Martik, K. P. Bjornson, E. S. Bjornson and P. Modrich (1998). "Nucleotide-promoted release of hMutSalpha from heteroduplex DNA is consistent with an ATP-dependent translocation mechanism." J Biol Chem 273(48): 32055-32062.

Cho, W. K., C. Jeong, D. Kim, M. Chang, K. M. Song, J. Hanne, C. Ban, R. Fishel and J. B. Lee (2012). "ATP alters the diffusion mechanics of MutS on mismatched DNA." Structure 20(7): 1264-1274.

Cristovao, M., E. Sisamakis, M. M. Hingorani, A. D. Marx, C. P. Jung, P. J. Rothwell, C. A. Seidel and P. Friedhoff (2012). "Single-molecule multiparameter fluorescence spectroscopy reveals directional MutS binding to mismatched bases in DNA." Nucleic acids research 40(12): 5448-5464.

Dutta, R. and M. Inouye (2000). "GHKL, an emergent ATPase/kinase superfamily." Trends Biochem Sci 25(1): 24-28.

Feng, G. and M. E. Winkler (1995). "Single-step purifications of His6-MutH, His6-MutL and His6-MutS repair proteins of escherichia coli K-12." Biotechniques 19(6): 956-965.

George, J. W., R. M. Brosh, Jr. and S. W. Matson (1994). "A dominant negative allele of the Escherichia coli uvrD gene encoding DNA helicase II. A biochemical and genetic characterization." Journal of molecular biology 235(2): 424-435.

Gorman, J., A. J. Plys, M. L. Visnapuu, E. Alani and E. C. Greene (2010). "Visualizing one-dimensional diffusion of eukaryotic DNA repair factors along a chromatin lattice." Nat Struct Mol Biol 17(8): 932-938.

Gorman, J., F. Wang, S. Redding, A. J. Plys, T. Fazio, S. Wind, E. E. Alani and E. C. Greene (2012). "Single-molecule imaging reveals target-search mechanisms during DNA mismatch repair." Proceedings of the National Academy of Sciences of the United States of America 109(45): E3074-3083.

Gorshkova, II, J. Svitel, F. Razjouyan and P. Schuck (2008). "Bayesian analysis of heterogeneity in the distribution of binding properties of immobilized surface sites." Langmuir 24(20): 11577-11586.

Gradia, S., S. Acharya and R. Fishel (1997). "The human mismatch recognition complex hMSH2-hMSH6 functions as a novel molecular switch." Cell 91(7): 995-1005.

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

Guarne, A. (2012). "The functions of MutL in mismatch repair: the power of multitasking." Prog Mol Biol Transl Sci 110: 41-70.

Guarne, A., S. Ramon-Maiques, E. M. Wolff, R. Ghirlando, X. Hu, J. H. Miller and W. Yang (2004). "Structure of the MutL C-terminal domain: a model of intact MutL and its roles in mismatch repair." The EMBO journal 23(21): 4134-4145.

Gupta, S., M. Gellert and W. Yang (2012). "Mechanism of mismatch recognition revealed by human MutSbeta bound to unpaired DNA loops." Nat Struct Mol Biol 19(1): 72-78.

Hall, M. C., J. R. Jordan and S. W. Matson (1998). "Evidence for a physical interaction between the Escherichia coli methyl-directed mismatch repair proteins MutL and UvrD." EMBO J 17(5): 1535-1541.

Hall, M. C. and S. W. Matson (1999). "The Escherichia coli MutL protein physically interacts with MutH and stimulates the MutH-associated endonuclease activity." The Journal of biological chemistry 274(3): 1306-1312.

Iaccarino, I., G. Marra, P. Dufner and J. Jiricny (2000). "Mutation in the magnesium binding site of hMSH6 disables the hMutSalpha sliding clamp from translocating along DNA." J Biol Chem 275(3): 2080-2086.

Jeong, C., W. K. Cho, K. M. Song, C. Cook, T. Y. Yoon, C. Ban, R. Fishel and J. B. Lee (2011). "MutS switches between two fundamentally distinct clamps during mismatch repair." Nat Struct Mol Biol 18(3): 379-385.

Jiricny, J. (2013). "Postreplicative mismatch repair." Cold Spring Harb Perspect Biol 5(4): a012633.

Junop, M. S., G. Obmolova, K. Rausch, P. Hsieh and W. Yang (2001). "Composite active site of an ABC ATPase: MutS uses ATP to verify mismatch recognition and authorize DNA repair." Mol Cell 7(1): 1-12.

Junop, M. S., W. Yang, P. Funchain, W. Clendenin and J. H. Miller (2003). "In vitro and in vivo studies of MutS, MutL and MutH mutants: correlation of mismatch repair and DNA recombination." DNA repair 2(4): 387-405.

Lahue, R. S., S. S. Su and P. Modrich (1987). "Requirement for d(GATC) sequences in Escherichia coli mutHLS mismatch correction." Proceedings of the National Academy of Sciences of the United States of America 84(6): 1482-1486.

Lamers, M. H., A. Perrakis, J. H. Enzlin, H. H. Winterwerp, N. de Wind and T. K. Sixma (2000). "The crystal structure of DNA mismatch repair protein MutS binding to a G x T mismatch." Nature 407(6805): 711-717.

Lamers, M. H., H. H. Winterwerp and T. K. Sixma (2003). "The alternating ATPase domains of MutS control DNA mismatch repair." The EMBO journal 22(3): 746-756.

Lebbink, J. H., A. Fish, A. Reumer, G. Natrajan, H. H. Winterwerp and T. K. Sixma (2010). "Magnesium coordination controls the molecular switch function of DNA mismatch repair protein MutS." The Journal of biological chemistry 285(17): 13131-13141.

Lynch, H. T., J. F. Lynch, P. M. Lynch and T. Attard (2008). "Hereditary colorectal cancer syndromes: molecular genetics, genetic counseling, diagnosis and management." Fam Cancer 7(1): 27-39.

Manelyte, L., C. Urbanke, L. Giron-Monzon and P. Friedhoff (2006). "Structural and functional analysis of the MutS C-terminal tetramerization domain." Nucleic Acids Res 34(18): 5270-5279.

Mechanic, L. E., B. A. Frankel and S. W. Matson (2000). "Escherichia coli MutL loads DNA helicase II onto DNA." The Journal of biological chemistry 275(49): 38337-38346.

Modrich, P. and R. Lahue (1996). "Mismatch repair in replication fidelity, genetic recombination, and cancer biology." Annu Rev Biochem 65: 101-133.

Natrajan, G., M. H. Lamers, J. H. Enzlin, H. H. Winterwerp, A. Perrakis and T. K. Sixma (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(16): 4814-4821.

Niedziela-Majka, A., N. K. Maluf, E. Antony and T. M. Lohman (2011). "Self-assembly of Escherichia coli MutL and its complexes with DNA." Biochemistry 50(37): 7868-7880.

Obmolova, G., C. Ban, P. Hsieh and W. Yang (2000). "Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA." Nature 407(6805): 703-710.

Qiu, R., V. C. DeRocco, C. Harris, A. Sharma, M. M. Hingorani, D. A. Erie and K. R. Weninger (2012). "Large conformational changes in MutS during DNA scanning, mismatch recognition and repair signalling." The EMBO journal 31(11): 2528-2540.

Raschle, M., P. Dufner, G. Marra and J. Jiricny (2002). "Mutations within the hMLH1 and hPMS2 subunits of the human MutLalpha mismatch repair factor affect its ATPase activity, but not its ability to interact with hMutSalpha." J Biol Chem 277(24): 21810-21820.

Robertson, A., S. R. Pattishall and S. W. Matson (2006). "The DNA binding activity of MutL is required for methyl-directed mismatch repair in Escherichia coli." The Journal of biological chemistry 281(13): 8399-8408.

Robertson, A. B., S. R. Pattishall, E. A. Gibbons and S. W. Matson (2006). "MutL-catalyzed ATP hydrolysis is required at a post-UvrD loading step in methyl-directed mismatch repair." J Biol Chem 281(29): 19949-19959.

Sacho, E. J., F. A. Kadyrov, P. Modrich, T. A. Kunkel and D. A. Erie (2008). "Direct visualization of asymmetric adenine-nucleotide-induced conformational changes in MutL alpha." Mol Cell 29(1): 112-121.

Selmane, T., M. J. Schofield, S. Nayak, C. Du and P. Hsieh (2003). "Formation of a DNA mismatch repair complex mediated by ATP." J Mol Biol 334(5): 949-965.

Sharma, A., C. Doucette, F. N. Biro and M. M. Hingorani (2013). "Slow conformational changes in MutS and DNA direct ordered transitions between mismatch search, recognition and signaling of DNA repair." J Mol Biol 425(22): 4192-4205.

Spampinato, C. and P. Modrich (2000). "The MutL ATPase is required for mismatch repair." The Journal of biological chemistry 275(13): 9863-9869.

Tham, K. C., N. Hermans, H. H. Winterwerp, M. M. Cox, C. Wyman, R. Kanaar and J. H. Lebbink (2013). "Mismatch Repair Inhibits Homeologous Recombination via Coordinated Directional Unwinding of Trapped DNA Structures." Mol Cell 51(3): 326-337.

Yokota, H., Y. A. Chujo and Y. Harada (2013). "Single-molecule imaging of the oligomer formation of the nonhexameric Escherichia coli UvrD helicase." Biophys J 104(4): 924-933.

5

A POSSIBLE ROLE FOR LOCALLY DIMINISHED DNA MISMATCH REPAIR ACTIVITY IN ESTABLISHING HIGH GC CONTENT PATHOGENICITY ISLANDS

Nicolaas Hermans, Arnold Kuzniar, Joyce H.G. Lebbink

ABSTRACT

    DNA mismatch repair (MMR) increases the fidelity of replication by removing mismatches and insertion/deletion loops left by the replication machinery, and by inhibiting recombination of homeologous DNA fragments. In Escherichia coli and Salmonella the newly synthesized strand is distinguished from its template by the methylation status of a specific sequence, GATC. These GATC sites are methylated by Dam methylase, which restores full methylation shortly after the DNA is copied. The newly synthesized DNA is unmethylated, and this hemimethylated status of GATC sites is used by MMR to remove mismatches. The presence of nearby GATC sites therefore is important for the function of MMR. The distribution of GATC sites in bacterial genomes is not uniform and large stretches of DNA devoid of GATC sites exist. Here we analyzed the genomes of several gram-negative bacteria to test the hypothesis that the genomic segments without GATC sites provide evidence for inefficient MMR. In mutant E. coli cells with non-functional MMR the conversion of A(T) to G(C) is the predominant transition mutation. We show that DNA segments larger than 3000 bp devoid of GATC sites (GATC deserts) have an elevated GC content, which suggests locally diminished MMR efficiency in these segments. Moreover, the GATC deserts often include rearrangement hotspots (rhs) and Rhs genes, which are involved in inter-cellular competition and host-pathogen interaction. The absence of GATC sites in Rhs genes provides an explanation for their increased propensity for rearrangements and polymorphisms.

INTRODUCTION

DNA methylation plays a role in transcriptional regulation and genome main-tenance, as well as in host-pathogen interaction (Marinus and Casadesus 2009). In *Escherichia coli* the adenosine residues in GATC sites are methylated by Dam methyltransferase, which effectively methylates both strands at the adenosine of most GATC sites shortly after replication. However, directly after replication GATC sites are methylated only at the template strand. This provides a short time window for DNA mismatch repair (MMR) to use the hemi-methylated state as a signal to distinguish the newly synthesized strand from the template strand, and to repair base mismatches and small insertion/deletion loops (indels) left by the replication machinery. This strand discrimination is mediated by sev-eral proteins: MutS recognizes the mismatch, MutL functions as a "molecular matchmaker", and MutH functions as a site-specific endonuclease that makes an incision at a hemi-methylated GATC site. The MutH endonuclease is activat-ed by the MutS-MutL complex in a mismatch dependent manner (for review (Jiricny 2013)), and can nick the DNA at unmethylated GATC sites only (Welsh, Lu et al. 1987).

The importance of the methylation of GATC sites in methyl-directed DNA MMR is shown by mutator phenotype of both the dam- and mutH- strains of *E. coli* (Marinus, Poteete et al. 1984, Wu, Clarke et al. 1990). Similarly, the over-expression of Dam methylase in E. coli also gives rise to a mutator phe-notype, by reducing the time window that is available to MMR (Herman and Modrich 1981). In line with these results, methyl-directed DNA MMR *in vitro* is dependent on the presence of hemi-methylated GATC sites (Langle-Rouault, Maenhaut-Michel et al. 1987), and increasing the distance between the mis-match and the nearest GATC site can diminish the repair efficiency (Lahue, Su et al. 1987, Bruni, Martin et al. 1988). Introducing GATC sites into the phage ΦX174, which naturally does not contain any GATC sites, yields a 30-fold de-crease in its mutation rate (Cuevas, Pereira-Gomez et al. 2011). Since the strand discrimination signal is a specific motif, the chromosome of *E. coli* provides a unique opportunity to study the influence of methyl-directed DNA MMR on the mutation frequencies in the chromosome.

The chromosomal DNA of *E. coli* is replicated almost exclusively by DNA polymerase III (Pol III) (Kelman and O'Donnell 1995). The proofreading activity of DNA Pol III removes mismatched nucleotides which introduce large helical distortions most efficiently (Arana and Kunkel 2010), and thus is more prone to leave behind mismatches with a smaller helical distortion. This mutational bias becomes evident in MMR deficient *E. coli* cells. In cells lacking either MutS, MutL or MutH have up to four-fold more frequent transitions from A(T) to G(C) than from G(C) to A(T) (Rewinski and Marinus 1987, Schaaper and Dunn 1991,

Lee, Popodi et al. 2012). The relative occurrence of transversion mutations (interchanges of purine for pyrimidine bases) is increased in these cells as well, but without a bias for one transversion over the other (Rewinski and Marinus 1987, Schaaper and Dunn 1991, Lee, Popodi et al. 2012). It is probably not a coincidence that MMR is most efficient in repairing the mismatches that are left most frequently by DNA Pol III (Wu, Clarke et al. 1990, Schaaper and Dunn 1991). Because the efficiency of MMR depends on the number of GATC sites and distance between them (Bruni, Martin et al. 1988, Lahue, Au et al. 1989), it is likely that in genomic regions with fewer or no GATC sites, MMR is less efficient than in regions with many GATC sites.

The distribution of GATC sites throughout the *E. coli* genome is not uniform (Barras and Marinus 1988, Henaut, Rouxel et al. 1996). The local clustering of GATC sites can be attributed to distinct functions of GATC sites in their methylated, hemi-methylated or unmethylated form. Apart from serving as a strand discrimination signal in MMR, the methylation of GATC sites serves several other functions (reviewed in (Marinus and Casadesus 2009)). Methylation of adenosine residues changes the structure of the DNA, influencing DNA-protein interactions and melting temperature in GATC sites (Barras and Marinus 1989, Marinus and Casadesus 2009). Changes in melting temperature are proposed to play a role in cold shock response (Marinus and Casadesus 2009). The methylation of clusters of GATC sites in the promoter regions of several genes changes their expression during the cell cycle. Similarly, *Salmonella typhimurium* lacking Dam are less virulent, because expression of genes important for their pathogenicity is at least partially regulated by Dam methylase (Heithoff, Sinsheimer et al. 1999, Chatti and Landoulsi 2008). Furthermore, the SeqA protein binds hemi-methylated and unmethylated GATC sites behind the replication fork around the origin of replication, preventing re-initiation of replication (reviewed in (Sanchez-Romero, Busby et al. 2010)). Finally, GATC sites likely increase the efficiency of the anti-recombination function of MMR, and are thus used to prevent the insertion of foreign DNA. The inhibitory effect of MutH (and presumably GATC sites) on interspecies recombination in *E. coli* between DNA from *Salmonella typhimurium* and the circular recipient chromosome is around 20 fold (Stambuk and Radman 1998). The inhibitory effect of MutS in the same study is about 700 fold, thus the function of GATC sites in inhibition of recombination is relatively minor. MutS binds mismatches in both MMR (Jiricny 2013) and homeologous recombination (Rayssiguier, Thaler et al. 1989, Tham, Hermans et al. 2013), however, the role of MutH and GATC sites in anti-recombination is not fully understood.

To better understand the influence of the distribution of GATC sites on the mutational bias of any part of the genome, the dynamic nature of genomic replication and mutation needs to be taken into account. Not all mutations oc-
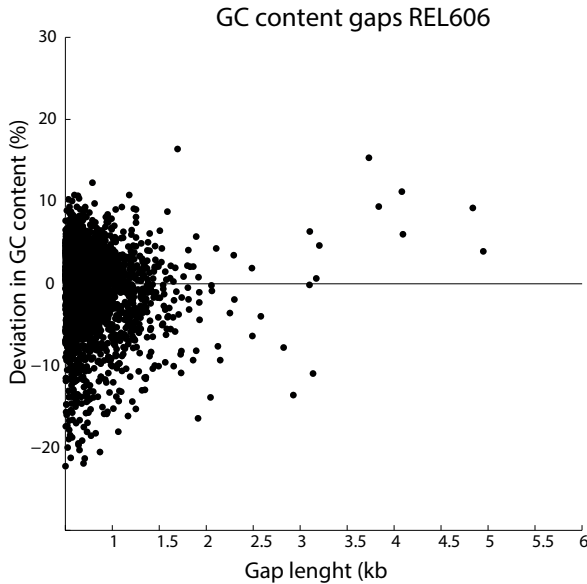
Figure 1: Sequence length-dependent deviation of GC content in DNA segments without GATC sites for *E. coli REL606* (RefSeq accession: NC_001672). With the length of GATC-free DNA segments on horizontal axis while the percent deviation of the GC content in these segments versus adjacent DNA (i.e. 10 kb up- and down-stream) on the vertical axis. In this genome, eleven GATC-free segments are larger than 3000 bp (GATC deserts) were identified, out of which nine have higher GC content (positive deviation) than the adjacent DNA.

cur with similar frequency throughout the genome, so the nucleotide makeup of any part of the genome will reflect both the function (*e.g.* the genes and preferred codon usage) and the local mutational bias. Several non-coding DNA motifs play a important role in the nucleotide make-up of chromosomes (Touzain, Petit et al. 2011). Given the mutational bias of MMR deficient strains, this implies that in genomic regions that lack GATC sites the GC content will tend to be higher. Here we show that stretches of over 3000 bp devoid of GATC sites (further referred to as GATC deserts) in *E. coli* and *Salmonella* have a significantly higher GC content compared to the genomic DNA nascent to these hotspots. These GATC deserts often contain Rhs genes. Rhs genes were discovered in re-arrangement hot-spots *(rhs)*, which were found to be hotspots for rearrangements in *E .coli* K12 under certain selective conditions (Lin, Capage et al. 1984). The absence of GATC sites in these Rhs genes might explain their increased tendency of genomic re-arrangements, as well as provide an explanation for the high GC content and increased polymorphisms in *Rhs* genes.
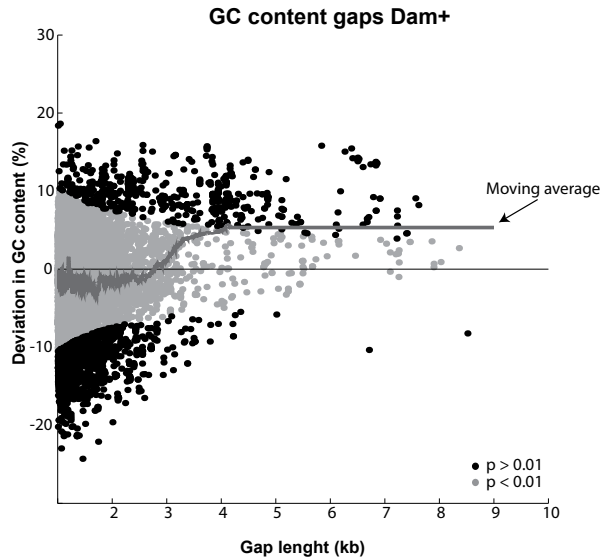
Figure 2: Average GC content of GATC deserts in genomes of *Escherichia, Salmonella* and *Shigella*. The dots represent GATC-free DNA segments, with the distance between two neighboring GATC sites on the x-axis and the deviation of the local genomic GC content on the y-axis. DNA segments with significant deviation in GC content (P < 0.01) are shown in black while the non-significant ones are in gray. The red line represents the moving average (across 300 data points) in the deviation of the GC content. Note: DNA segments shorter than 1 kb are left out for clarity.

RESULTS

GATC deserts have higher GC content

The distribution of GATC sites throughout the E. coli genome is not uniform as the sites often form clusters, e.g. in promoter regions or around the replication origins (Barras and Marinus 1988, Henaut, Rouxel et al. 1996) as the sites often form clusters. Clusters of GATC sites are often found in promoter regions, and around the origin of replication (Marinus and Casadesus 2009). The genome of the *E. coli* strain REL606, a well-studied clone of the *E. coli* B strain used for evolutionary experiments (Daegelen, Studier et al. 2009), contains 19,668 GATC sites. The size distribution of the DNA fragments in between neighboring GATC sites, and the GC content of all these fragments for the *E. coli REL606* genome is shown in Figure 1. Gaps larger than 1000 bp are relatively rare compared to the number of clustered GATC sites, consistent with previous findings that a lower number of these gaps are present in the *E. coli* genome than what might be expected by chance (Barras and Marinus 1988, Henaut, Rouxel et al. 1996). However, we found 62 significant clusters (DNA fragments with a significant higher number of GATC sites, p < 0.05) and 11 significant gaps (large DNA
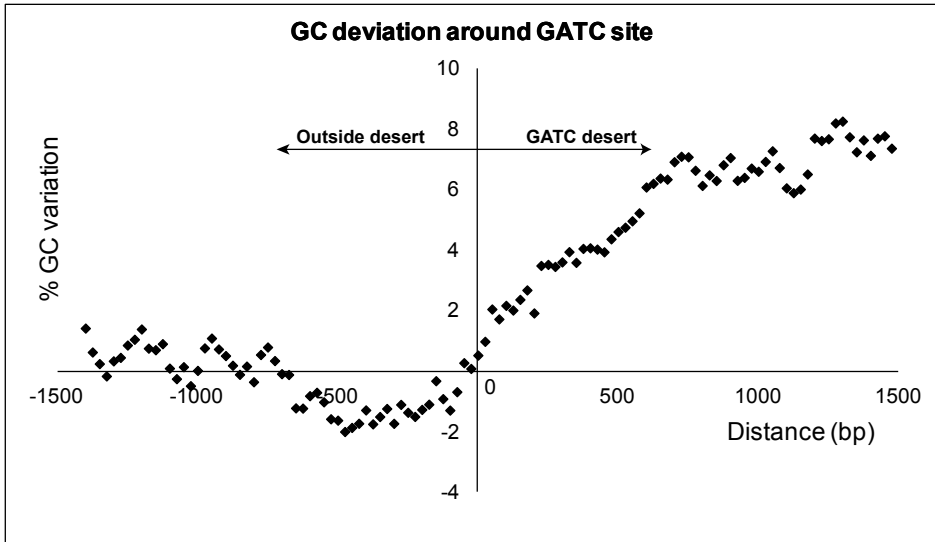
Figure 3: The GC content changes abruptly around the GATC sites at the border of GATC deserts. The dots represent the average GC content of all GC deserts, at a given distance from the GATC sites flanking the GATC desert. The horizontal axis represents the distance in bp from this GATC site. A positive number corresponds with DNA inside the GATC desert (right hand side of the vertical axis), a negative number corresponds with DNA outside of the GATC desert (left of the vertical axis). The GATC site that marks the border of the GATC desert is positioned at 0. The DNA inside the GATC desert has a higher GC content, and the GC content increases abruptly at the bordering GATC site.

fragments with significantly less GATC sites) in the GATC site distribution in a genome-wide scan using the *Ab Initio* Motif Finding Environment (AIMIE) (Mrazek, Xie et al. 2008). In these significant gaps, the efficiency of MMR is presumably lower, and the mutational bias of the replicative polymerase Pol III can increase the GC content of these fragments.

To investigate this, we analyzed the GC content of DNA segments delimited by GATC sites in the *E. coli* REL606 genome. The GC content of the genome can vary locally (Blattner, Plunkett et al. 1997). To account for these local differences, we compared the GC content within fragments devoid of GATC sites to the local GC content 10 kb up- and down-stream from the flanking GATC sites (Figure 1). A direct correlation between the gap size and a deviation in GC content is not immediately obvious, however, of the 11 gaps larger than 3000 bp, 9 have a higher GC content than the average of the genome (Table 1).

Therefore, we analyzed the genomes of Gram-negative bacteria closely related to *E. coli*, including *Escherichia, Shigella* and *Salmonella* genera (see supplementary Table S1). The presence of both MutH and Dam in these bacterial genomes is a strong indication that they depend on methyl-directed
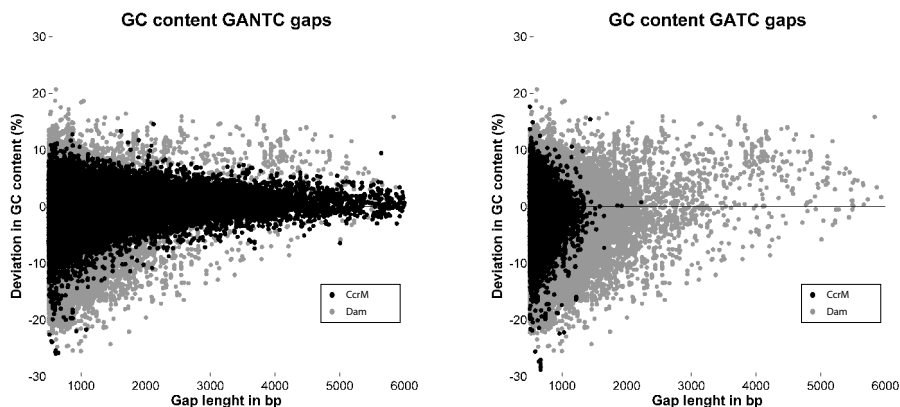
F**igure 4: The GC content of DNA fragments without GANTC or GATC sites.** The GANTC distribution does not cause a deviation in GC content, unlike the GATC distribution. DNA fragments between GANTC sites (black dots) from the organisms that use CcrM methylase, compared to the DNA fragments between GATC sites from the Dam-MutH group (light grey dots, in the background, same as in Figure 2). The GANTC distribution does not show a skew towards higher or lower GC content. In the GATC site distribution in these organisms, no large gaps in the GATC distribution can be found (black dots, compared to grey dots for E. coli group). Additionally, no skew in their GC content is visible.

DNA MMR (Eisen and Hanawalt 1999). All the gaps in the GATC distribution of this "Dam-MutH" group, and the average GC content compared to the local average is plotted in Figure 2. In gaps smaller than 3000 bp, there seems to be no correlation between the average gap size and the GC content. The average GC content of the gaps increases at gaps larger than 3000 bp (Figure 2). Most DNA fragments of over 3500 bp devoid of GATC sites have a significantly higher than average GC content. We will refer to DNA fragments larger than 3000 bp devoid of GATC sites as GATC deserts.

To better understand the changes in GC content around the GATC sites flanking the GATC deserts, we plotted the deviations from the local GC content around each GATC desert (Figure 3). The two GATC sites bordering the GATC desert mark a relative abrupt change in GC content. This shows, that for sequences larger than 3000 bp, there is a strong correlation between the absence of GATC sites, and the GC content of the DNA, and the GATC sites form a direct boundary of this change in GC content.

GANTC sites do not alter GC content

In α-proteobacteria GATNC sites are methylated by the CcrM protein, which has equivalent function as the Dam methylase in *E. coli* (Collier, McAdams et al. 2007). However, these. In contrast to GATC sites, the hemi-methylated GANTC

sites are not used for DNA MMR. This is supported by the presence of MutL endonuclease, which can nick the daughter DNA strand in a non-sequence-specific manner (Guarne 2012). The endonuclease activity is absent in MutL from *E. coli* and *Salmonella* (Pillon, Lorenowicz et al. 2010), and MMR fully relies on the endonuclease MutH for strand discrimination. We compared the GC content of the DNA fragments devoid of GANTC sites from the bacteria with homologs of CcrM methylase to their local genomic GC content, as was done for the Dam-MutH group (Figure 4, genera *Caulobacter, Agrobacterium*, *Sinorhizobium*, and *Brucella*, Table 2). The distribution of GANTC sites in these strains lacks the correlation of GC content as seen in the group of bacteria that uses methyl-directed DNA MMR. Also when these genomes were analyzed for their GATC distribution, no deviation of GC content in GATC deserts were found (Figure 5, right panel). The GATC sites in these genomes are distributed more equally compared to the Gram-negative bacteria in Figure 2, with only few gaps larger than 1 kb. Since the sequence GATC carries as far as we know no specific function in these genomes, we cannot draw any conclusions from this distribution itself.

#	Start..end position	Annotated elements	% deviation in GC
1	494253 . . 499090	rhsD	9.20%
2	1502010 . . 1503939	rhsE	15.30%
3	2100181 . . 2103279	Prophage P2 like (REL606 specific)	-0.10%
4	2107099 . . 2111190	Prophage P2 like (REL606 specific)	6.00%
5	2117950 . . 2121119	Prophage P2 like (REL606 specific)	0.60%
6	2680406 . . 2683540	ECB_02511-2, insA-20, insB-20	-10.90%
7	2999344 . . 3002445	flu_Adhesin_Aida_Type_V_secretion	6.30%
8	3005693 . . 3010641	RadC_and_methyltransferase	3.90%
9	3549705 . . 3553539	rhsB	9.40%
10	3696455 . . 3700536	rhsA	11.20%
11	3783155 . . 3786358	ECB_03522 – ECB_03524	4.60%

Table 1: All GATC deserts (≥3000 bp) detected in the *E. coli REL606* genome. Four of these contain rhs elements (1, 2, 9, 10). Three GATC deserts (3-5) contain putative prophage insertions of the same ancestry whereas one desert (6) contains the gene encoding IS1 transposase (ISB 20).
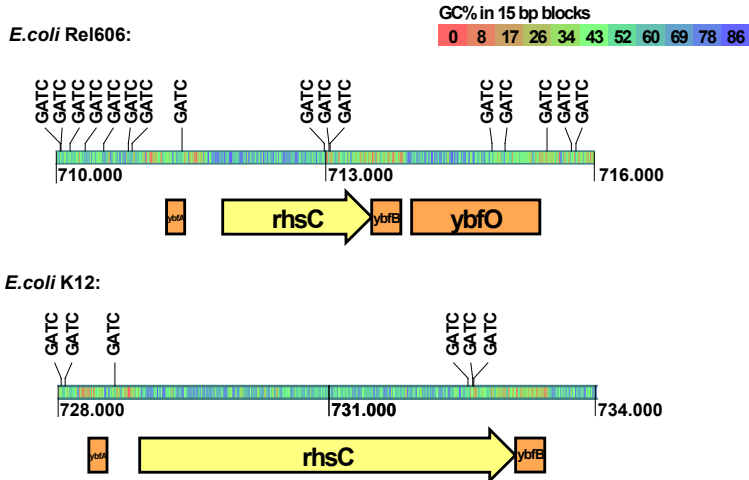
Figure 5: Comparing the rhsC gene of two *E. coli* strains REL606 (RefSeq accession: NC_012967) and K12 (RefSeq accession: NC_000913). AT- and GC-rich DNA regions are indicated in red and blue, respectively. Figure was made using pDRAW32 (www.acaclone.com)

GATC deserts contain *rhs* elements

We investigated what types of genetic elements (e.g. protein-coding genes, transposable elements etc.) are present in the GATC deserts of the *E. coli* REL606 genome. All GATC deserts detected in *E. coli* REL606 are listed in Table 1. Five of the GATC deserts are close to transposable elements, or contain sequences of viral origin. Specifically, the GATC deserts (3-5) are close to each other and contain phage DNA. Phage DNA is often devoid of GATC sites (Cuevas, Pereira-Gomez et al. 2011), which provides an explanation for the presence of the GATC deserts. Furthermore, the GATC desert #6 contains a transposable element. These GATC deserts reinforce the expectation that many GATC deserts are relative recent insertions of foreign DNA, and therefore an elevated GC content cannot be directly attributed to diminished MMR.

Four of the 11 GATC deserts contain *rhs* elements. *Rhs* elements contain genes (Rhs) which consist of conserved and variable domains (Wang, Zhao et al. 1998). These *rhs* elements have been reported in a range of enterobacteria, including *Salmonella* and *Yersina*. The origin of Rhs genes likely predates the emergence of the *Enterobacericeae* family (Jackson, Thomas et al. 2009). We found that the *rhs* elements in *Escherichia* and *Salmonell*a are often found in GATC deserts, and have a high GC content. This implies that the elevated GC content and the absence of GATC sites in these Rhs genes is not caused by a recent insertion of foreign DNA, as is probably the case for the transposable elements and phage insertions. The elevated GC content in these elements thus reflects the function of and selective pressure on the Rhs genes within these

GATC deserts and the local mutational bias.

Interestingly, the RhsC gene is not contained within a GATC desert in the *E. coli* REL606 genome. The RhsC gene of *E. coli* REL606 is substantially smaller (due to deletion of rhs repeats) compared to the homologous RhsC gene of the *E. coli* K-12 (Figure 5). Any effect of this deletion on the putative function of RhsC in REL606 is unknown.

DISCUSSION

DNA MMR plays an important role in maintaining genomic integrity by removing polymerase errors. MMR in *E. coli* utilizes hemi-methylated GATC sites to distinguish the newly synthesized strand from the template strand, and the absence of hemi-methylated GATC sites in the vicinity of a mismatch decreases the efficiency of MMR (Schlagman, Hattman et al. 1986, Bruni, Martin et al. 1988, Lahue, Au et al. 1989). In parts of the genome where MMR is less efficient, the mutational bias of DNA polymerases should have impact on the nucleotide content. In *E. coli* lacking functional MMR, transitions from A(T) to G(C) occur 2-4 fold more often than transitions from G(C) to A(T) (Rewinski and Marinus 1987, Schaaper and Dunn 1991, Lee, Popodi et al. 2012). We found that in the genomes of *E. coli* and *Salmonella,* the GC content is significantly higher in DNA fragments of over 3000 bp that are devoid of GATC sites, dubbed a GATC-deserts. The two GATC sites flanking this GATC desert often mark a relatively sharp transition from DNA just outside of the GATC desert to the GATC desert with an elevated GC content (Figure 3). This shows there is a strong correlation between the absence of GATC sites and an elevated GC content in these genomes. Furthermore, the relative sharp transition would imply a relative sharp decrease in MMR efficiency, even close to the GATC site. This would suggest that mismatch repair loses efficiency in a GATC desert, even if a single GATC site is in close proximity to the mismatch. This is another indication that two flanking GATC sites are utilized for efficiently repairing a single mismatch, as was shown for *in vitro* experiments in Chapter 2.

The correlation between an elevated GC content and GATC deserts does not necessarily mean the absence of GATC sites, and a decrease in efficiency of MMR is also the cause for an elevated GC content of GATC deserts. First, we explored the possibility that the other functions of GATC sites could influence the GC content of GATC deserts. The most likely explanation for a higher GC content for any sequence within a bacterial genome is the introduction of foreign DNA with a high GC content. This can be caused by horizontal gene transfer (HGT) and by insertion of phage DNA and mobile elements like transposons. Heterogeneity in GC content is used to find spots of horizontal gene transfer (HGT) (Lawrence and Ochman 1998, Yoon, Hur et al. 2005). The majority of the *E. coli* genome has been subjected to HGT and HGT is the major source of mutations in *E. coli*

(Studier, Daegelen et al. 2009). Additionally, the frequency of GATC sites in phage DNA is often much lower than in genomic DNA (Marinus and Casadesus 2009). It is therefore unsurprising that a substantial part of the GATC deserts are remnants of inserted phage DNA. Some of these insertions have a lower rather than a higher GC content, which is also the case for one of the detected GATC deserts in *E. coli* REL606 (Table 1). It is unclear what the influence of MMR might be on the GC content of phage DNA that lack GATC sites. On one hand, phage DNA is replicated inside a host, often in a genomic context, and may therefore show a similar mutational bias as the genome of the host. On the other hand, natural selection on the genetic level and replication outside of a genomic context will also influence the nucleotide composition of phage DNA, which could explain the low GC content of the phage DNA devoid of GATC sites. This may explain the heterogeneity of GC content within GATC deserts that find their origin in phage DNA.

An interesting feature of the GATC deserts in *E. coli* REL606 is that many contain rhs elements. *Rhs* elements were found to be hotspots for rearrangements in *E. coli* K12 under certain selective conditions (Lin, Capage et al. 1984). Because of lack of evidence that these *rhs* elements are indeed hotspots for chromosomal rearrangements in bacterial genomes under normal circumstances, *rhs* elements are now often regarded as simple genes encoding a ~1,200 amino acids large protein (Jackson, Thomas et al. 2009, Liu, Knabel et al. 2009). This gene is composed of a GC-rich conserved core region and a variable AT-rich C-terminal tip, and can be subdivided into three subfamilies, RhsA-B-C-F, RhsD-E, and RhsG-H (Wang, Zhao et al. 1998). The function of the Rhs genes is currently unclear, but Rhs genes have been implicated in pathogenicity, intercellular competition (Sisto, Cipriani et al. 2010), and host interactions (van Diemen, Dziva et al. 2005, Poole, Diner et al. 2011, Koskiniemi, Lamoureux et al. 2013). It is clear that the functions of Rhs proteins are at the cell surface or cell envelope and their molecular function may well include a role in carbohydrate binding (Jackson, Thomas et al. 2009). However, the function of Rhs genes is important, since they are present and conserved in almost all genomes of Enterobacteria (Jackson, Thomas et al. 2009) and under strong positive selection pressure (Petersen, Bollback et al. 2007). Also, Rhs genes are not recently transferred from GC-rich organisms, since the origin of Rhs genes likely predates the origin of Enterobactericeae itself (Jackson, Thomas et al. 2009). Therefore, the elevated GC content within GATC desert containing Rhs genes has its origin in the local selective pressure and mutational bias.

Since Rhs genes are often contained within a GATC desert and have an increased GC content, the local efficiency of MMR in these genes could be lower than elsewhere in the genome. This should increase the mutation rate in Rhs genes, likely resulting in a higher frequency of polymorphisms in Rhs genes

between closely related strains. Indeed, substantial genetic polymorphism was found in Rhs genes among strains of a single pathogen (*E. coli* 0157:H7) (Liu, Knabel et al. 2009), and Rhs genes were suggested to be used as markers for multilocus sequence typing. An interesting explanation for the absence of GATC sites in Rhs genes would be to decrease the efficiency of MMR, in order to increase the mutation rate in Rhs genes, thereby increasing polymorphisms. An alternative hypothesis for the function of GATC sites is the inhibition of homeologous recombination. The absence of GATC sites can increase homeologous recombination events with Rhs genes on circular intermediates, as suggested by Jackson and coworkers (Jackson, Thomas et al. 2009). This can provide an explanation why *rhs* elements were found to be rearrangement hotspots under certain conditions (Lin, Capage et al. 1984). For example, a higher local mutation and rearrangement rate could provide an evolutionary advantage during circumstances that require adaptations in inter-cell competition and host-pathogen interactions. Therefore, it can be hypothesized that the absence of GATC sites and the local decrease in MMR efficiency is a feature of these Rhs genes intrinsic to their function.

To further elucidate the possible function of the absence of GATC sites in Rhs genes, several experiments can be done. First, it would be interesting to see whether introducing GATC sites in Rhs genes decreases their potential as rearrangement hotspot. Second, removing GATC sites from other genes should increase their mutation rate, and the mutations should show a similar mutational bias as found in MMR deficient strains. Surprisingly, a direct correlation between the presence of nearby GATC sites and the frequency of a single nucleotide deletion has not been found (Martina, Correa et al. 2012). However, since the function of GATC sites in MMR is well established in *in vitro* assays, we deem it unlikely that GATC distribution has no significant influence on MMR efficiency *in vivo*. The *in silico* analyses presented here suggest that DNA segments larger than 3000 bp without GATC sites then to have increased mutation rates with a specific bias towards A/T to G/C conversions, when compared to genomic regions with GATC sites.

Materials and Methods

The bacterial genomes used for the analyseswere retrieved from the NCBI's Reference Sequence database (RefSeq release 60) (see supplementary Tables S1 and S2). A Perl script was used to find all GA(N)TC sites and output tabulate distances between the adjacent GA(N)TC sites and calculate the GC content. The deviation of the GC content to the local GC content was calculated by subtracting the GC content of a DNA fragment devoid of GATC sites from the GC content of the 10kb up- and down-stream from the two bordering GATC sites.

The statistical significance of the deviating GC content of DNA segments was assessed using the Chi square test. The local GC content from 1500 bp upstream to 1500 bp downstream of the bordering GATC sites was calculated in 50 bp windows, and the window was moved with a 10 bp step size.

Supplementary tables

Accesion	Species name	Size (bp)
NC_010658.1	Shigella boydii CDC 3083-94	4615997
NC_007606.1	Shigella dysenteriae Sd197	4369232
NC_017328.1	Shigella flexneri 2002017	4650856
NC_007384.1	Shigella sonnei Ss046	4825265
NC_017626.1	Escherichia coli 042	5241977
NC_008253.1	Escherichia coli 536	4938920
NC_011748.1	Escherichia coli 55989	5154862
NC_017631.1	Escherichia coli ABU 83972	5131397
NC_008563.1	Escherichia coli APEC O1	5082025
NC_010468.1	Escherichia coli ATCC 8739	4746218
NC_012759.1	Escherichia coli BW2952	4578159
NC_012967.1	Escherichia coli B str. REL606	4629812
NC_004431.1	Escherichia coli CFT073	5231428
NC_017638.1	Escherichia coli DH1	4621430
NC_009801.1	Escherichia coli E24377A	4979619
NC_011745.1	Escherichia coli ED1a	5209548
NC_017633.1	Escherichia coli ETEC H10407	5153435
NC_009800.1	Escherichia coli HS	4643538
NC_011741.1	Escherichia coli IAI1	4700560
NC_017660.1	Escherichia coli KO11FL	5021812
NC_010473.1	Escherichia coli str. K-12 substr. DH10B	4686137
NC_011993.1	Escherichia coli LF82	4773108
NC_017644.1	Escherichia coli NA114	4971461
NC_013364.1	Escherichia coli O111:H- str. 11128	5371077
NC_017646.1	Escherichia coli O7:K1 str. CE10	5313531
NC_017663.1	Escherichia coli P12b	4935294
NC_011415.1	Escherichia coli SE11	4887515
NC_010498.1	Escherichia coli SMS-3-5	5068389
NC_017632.1	Escherichia coli UM146	4993013
NC_007946.1	Escherichia coli UTI89	5065741
NC_017635.1	Escherichia coli W	4900968
NC_017906.1	Escherichia coli Xuzhou21	5386223
NC_011740.1	Escherichia fergusonii ATCC 35469	4588711
NC_015761.1	Salmonella bongori NCTC 12419	4460105
NC_010067.1	Salmonella enterica subsp. arizonae serovar 62:z4,z23	4600800
NC_011149.1	Salmonella enterica subsp. enterica serovar Agona str. SL483	4798660
NC_006905.1	Salmonella enterica subsp. enterica serovar Choleraesuis	4755700
NC_011205.1	Salmonella enterica subsp. enterica serovar Dublin	4842908
NC_011294.1	Salmonella enterica subsp. enterica serovar Enteritidis	4685848

NC_011274.1	Salmonella enterica subsp. enterica serovar Gallinarum	4658697
NC_017623.1	Salmonella enterica subsp. enterica serovar Heidelberg str. B182	4750465
NC_011080.1	Salmonella enterica subsp. enterica serovar Newport str. SL254	4827641
NC_011147.1	Salmonella enterica subsp. enterica serovar Paratyphi A	4581797
NC_011094.1	Salmonella enterica subsp. enterica serovar Schwarzengrund	4709075
NC_016863.1	Salmonella enterica subsp. enterica serovar Typhimurium	4817868
NC_016832.1	Salmonella enterica subsp. enterica serovar Typhi str. P-stx-12	4768352
NC_010658.1	Shigella boydii CDC 3083-94	4615997
NC_007606.1	Shigella dysenteriae Sd197	4369232
NC_017328.1	Shigella flexneri 2002017	4650856
NC_007384.1	Shigella sonnei Ss046	4825265

Table S1: List of 50 bacterial chromosomes from different genera (*Escherichia*, *Salmonella* and *Shigella*) used for the analysis of GATC sites.

Accesion	Species name
NC_015183.1	Agrobacterium sp. H13-3 chromosome
NC_015508.1	Agrobacterium sp. H13-3 chromosome linear
NC_011983.1	Agrobacterium radiobacter K84 chromosome 2
NC_011985.1	Agrobacterium radiobacter K84 chromosome 1
NC_003062.2	Agrobacterium tumefaciens str. C58 chromosome circular
NC_003063.2	Agrobacterium tumefaciens str. C58 chromosome linear
NC_011988.1	Agrobacterium vitis S4 chromosome 2
NC_011989.1	Agrobacterium vitis S4 chromosome 1
NC_016777.1	Brucella abortus A13334 chromosome 2
NC_016795.1	Brucella abortus A13334 chromosome 1
NC_006932.1	Brucella abortus bv. 1 str. 9-941 chromosome I
NC_006933.1	Brucella abortus bv. 1 str. 9-941 chromosome II
NC_010740.1	Brucella abortus S19 chromosome 2
NC_010742.1	Brucella abortus S19 chromosome 1
NC_010103.1	Brucella canis ATCC 23365 chromosome I
NC_010104.1	Brucella canis ATCC 23365 chromosome II
NC_016778.1	Brucella canis HSK A52141 chromosome 1
NC_016796.1	Brucella canis HSK A52141 chromosome 2
NC_012441.1	Brucella melitensis ATCC 23457 chromosome I
NC_012442.1	Brucella melitensis ATCC 23457 chromosome II
NC_007618.1	Brucella melitensis biovar Abortus 2308 chromosome I
NC_007624.1	Brucella melitensis biovar Abortus 2308 chromosome II
NC_003317.1	Brucella melitensis bv. 1 str. 16M chromosome I
NC_003318.1	Brucella melitensis bv. 1 str. 16M chromosome II
NC_017244.1	Brucella melitensis M28 chromosome chromosome 1
NC_017245.1	Brucella melitensis M28 chromosome chromosome 2
NC_017246.1	Brucella melitensis M5-90 chromosome chromosome I
NC_017247.1	Brucella melitensis M5-90 chromosome chromosome II
NC_017248.1	Brucella melitensis NI chromosome chromosome I
NC_017283.1	Brucella melitensis NI chromosome chromosome II
NC_013118.1	Brucella microti CCM 4915 chromosome 2
NC_013119.1	Brucella microti CCM 4915 chromosome 1
NC_009504.1	Brucella ovis ATCC 25840 chromosome II

5

NC_009505.1	Brucella ovis ATCC 25840 chromosome I
NC_015857.1	Brucella pinnipedialis B2/94 chromosome 1
NC_015858.1	Brucella pinnipedialis B2/94 chromosome 2
NC_017250.1	Brucella suis 1330 chromosome II
NC_017251.1	Brucella suis 1330 chromosome I
NC_004310.3	Brucella suis 1330 chromosome I
NC_004311.2	Brucella suis 1330 chromosome II
NC_010167.1	Brucella suis ATCC 23445 chromosome II
NC_010169.1	Brucella suis ATCC 23445 chromosome I
NC_016775.1	Brucella suis VBI22 chromosome II
NC_016797.1	Brucella suis VBI22 chromosome I
NC_002696.2	Caulobacter crescentus CB15 chromosome
NC_011916.1	Caulobacter crescentus NA1000 chromosome
NC_010338.1	Caulobacter sp. K31 chromosome
NC_014100.1	Caulobacter segnis ATCC 21756 chromosome
NC_016812.1	Sinorhizobium fredii HH103
NC_016815.1	Sinorhizobium fredii HH103 plasmid pSfHH103e
NC_018000.1	Sinorhizobium fredii USDA 257 chromosome
NC_009620.1	Sinorhizobium medicae WSM419 plasmid pSMED01
NC_009621.1	Sinorhizobium medicae WSM419 plasmid pSMED02
NC_009636.1	Sinorhizobium medicae WSM419 chromosome
NC_003037.1	Sinorhizobium meliloti 1021 plasmid pSymA
NC_003047.1	Sinorhizobium meliloti 1021 chromosome
NC_003078.1	Sinorhizobium meliloti 1021 plasmid pSymB
NC_015590.1	Sinorhizobium meliloti AK83 chromosome 1
NC_015591.1	Sinorhizobium meliloti AK83 chromosome 3
NC_015596.1	Sinorhizobium meliloti AK83 chromosome 2
NC_017322.1	Sinorhizobium meliloti BL225C chromosome
NC_017323.1	Sinorhizobium meliloti BL225C plasmid pSINMEB02
NC_017324.1	Sinorhizobium meliloti BL225C plasmid pSINMEB01
NC_018683.1	Sinorhizobium meliloti Rm41 plasmid pSYMA
NC_018700.1	Sinorhizobium meliloti Rm41
NC_018701.1	Sinorhizobium meliloti Rm41 plasmid pSYMB
NC_017325.1	Sinorhizobium meliloti SM11 chromosome
NC_017326.1	Sinorhizobium meliloti SM11 plasmid pSmeSM11d
NC_017327.1	Sinorhizobium meliloti SM11 plasmid pSmeSM11c

Table S2: List of bacterial chromosomes using GANTC methylation.

REFERENCES

Arana, M. E. and T. A. Kunkel (2010). "Mutator phenotypes due to DNA replication infidelity." Semin Cancer Biol **20**(5): 304-311.

Barras, F. and M. G. Marinus (1988). "Arrangement of Dam methylation sites (GATC) in the Escherichia coli chromosome." Nucleic Acids Res **16**(20): 9821-9838.

Barras, F. and M. G. Marinus (1989). "The great GATC: DNA methylation in E. coli." Trends Genet **5**(5): 139-143.

Blattner, F. R., G. Plunkett, 3rd, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau and Y. Shao (1997). "The complete genome sequence of Escherichia coli K-12." Science (New York, N Y) **277**(5331): 1453-1462.

Bruni, R., D. Martin and J. Jiricny (1988). "d(GATC) sequences influence Escherichia coli mismatch repair in a distance-dependent manner from positions both upstream and downstream of the mismatch." Nucleic acids research **16**(11): 4875-4890.

Chatti, A. and A. Landoulsi (2008). "GATC sites might regulate virulence of Salmonella typhimurium." Foodborne Pathog Dis **5**(5): 555-557.

Collier, J., H. H. McAdams and L. Shapiro (2007). "A DNA methylation ratchet governs progression through a bacterial cell cycle." Proceedings of the National Academy of Sciences of the United States of America **104**(43): 17111-17116.

Cuevas, J. M., M. Pereira-Gomez and R. Sanjuan (2011). "Mutation rate of bacteriophage PhiX174 modified through changes in GATC sequence context." Infect Genet Evol **11**(7): 1820-1822.

Daegelen, P., F. W. Studier, R. E. Lenski, S. Cure and J. F. Kim (2009). "Tracing ancestors and relatives of Escherichia coli B, and the derivation of B strains REL606 and BL21(DE3)." J Mol Biol **394**(4): 634-643.

Eisen, J. A. and P. C. Hanawalt (1999). "A phylogenomic study of DNA repair genes, proteins, and processes." Mutat Res **435**(3): 171-213.

Guarne, A. (2012). "The functions of MutL in mismatch repair: the power of multitasking." Prog Mol Biol Transl Sci **110**: 41-70.

Heithoff, D. M., R. L. Sinsheimer, D. A. Low and M. J. Mahan (1999). "An essential role for DNA adenine methylation in bacterial virulence." Science **284**(5416): 967-970.

Henaut, A., T. Rouxel, A. Gleizes, I. Moszer and A. Danchin (1996). "Uneven distribution of GATC motifs in the Escherichia coli chromosome, its plasmids and its phages." Journal of molecular biology **257**(3): 574-585.

Herman, G. E. and P. Modrich (1981). "Escherichia coli K-12 clones that overproduce dam methylase are hypermutable." Journal of bacteriology **145**(1): 644-646.

Jackson, A. P., G. H. Thomas, J. Parkhill and N. R. Thomson (2009). "Evolutionary diversification of an ancient gene family (rhs) through C-terminal displacement." BMC Genomics **10**: 584.

Jiricny, J. (2013). "Postreplicative mismatch repair." Cold Spring Harb Perspect Biol **5**(4): a012633.

Kelman, Z. and M. O'Donnell (1995). "DNA polymerase III holoenzyme: structure and function of a chromosomal replicating machine." Annu Rev Biochem **64**: 171-200.

Koskiniemi, S., J. G. Lamoureux, K. C. Nikolakakis, C. t'Kint de Roodenbeke, M. D. Kaplan, D. A. Low and C. S. Hayes (2013). "Rhs proteins from diverse bacteria mediate intercellular competition." Proc Natl Acad Sci U S A **110**(17): 7032-7037.

Lahue, R. S., K. G. Au and P. Modrich (1989). "DNA mismatch correction in a defined system." Science (New York, N Y) **245**(4914): 160-164.

Lahue, R. S., S. S. Su and P. Modrich (1987). "Requirement for d(GATC) sequences in Escherichia coli mutHLS mismatch correction." Proceedings of the National Academy of Sciences of the United States of America **84**(6): 1482-1486.

Langle-Rouault, F., G. Maenhaut-Michel and M. Radman (1987). "GATC sequences, DNA nicks and the MutH function in Escherichia coli mismatch repair." The EMBO journal **6**(4): 1121-1127.

Lawrence, J. G. and H. Ochman (1998). "Molecular archaeology of the Escherichia coli genome." Proc Natl Acad Sci U S A **95**(16): 9413-9417.

Lee, H., E. Popodi, H. Tang and P. L. Foster (2012). "Rate and molecular spectrum of spontaneous mutations in the bacterium Escherichia coli as determined by whole-genome sequencing." Proceedings of the National Academy of Sciences of the United States of America **109**(41): E2774-2783.

Lin, R. J., M. Capage and C. W. Hill (1984). "A repetitive DNA sequence, rhs, responsible for duplications within the Escherichia coli K-12 chromosome." J Mol Biol **177**(1): 1-18.

Liu, K., S. J. Knabel and E. G. Dudley (2009). "rhs genes are potential markers for multilocus sequence typing of Escherichia coli O157:H7 strains." Appl Environ Microbiol **75**(18): 5853-5862.

Marinus, M. G. and J. Casadesus (2009). "Roles of DNA adenine methylation in host-pathogen interactions: mismatch repair, transcriptional regulation, and more." FEMS Microbiol Rev **33**(3): 488-503.

Marinus, M. G., A. Poteete and J. A. Arraj (1984). "Correlation of DNA adenine methylase activity with spontaneous mutability in Escherichia coli K-12." Gene **28**(1): 123-125.

Martina, M. A., E. M. Correa, C. E. Argarana and J. L. Barra (2012). "Escherichia coli frameshift mutation rate depends on the chromosomal context but not on the GATC content near the mutation site." PLoS One **7**(3): e33701.

Mrazek, J., S. Xie, X. Guo and A. Srivastava (2008). "AIMIE: a web-based environment for detection and interpretation of significant sequence motifs in prokaryotic genomes." Bioinformatics **24**(8): 1041-1048.

Petersen, L., J. P. Bollback, M. Dimmic, M. Hubisz and R. Nielsen (2007). "Genes under positive selection in Escherichia coli." Genome Res **17**(9): 1336-1343.

Pillon, M. C., J. J. Lorenowicz, M. Uckelmann, A. D. Klocko, R. R. Mitchell, Y. S. Chung, P. Modrich, G. C. Walker, L. A. Simmons, P. Friedhoff and A. Guarne (2010). "Structure of the endonuclease domain of MutL: unlicensed to cut." Molecular cell **39**(1): 145-151.

Poole, S. J., E. J. Diner, S. K. Aoki, B. A. Braaten, C. t'Kint de Roodenbeke, D. A. Low and C. S. Hayes (2011). "Identification of functional toxin/immunity genes linked to contact-dependent growth inhibition (CDI) and rearrangement hotspot (Rhs) systems." PLoS Genet **7**(8): e1002217.

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

Rewinski, C. and M. G. Marinus (1987). "Mutation spectrum in Escherichia coli DNA mismatch repair deficient (mutH) strain." Nucleic acids research **15**(20): 8205-8215.

Sanchez-Romero, M. A., S. J. Busby, N. P. Dyer, S. Ott, A. D. Millard and D. C. Grainger (2010). "Dynamic distribution of seqa protein across the chromosome of escherichia coli K-12." MBio **1**(1).

Schaaper, R. M. and R. L. Dunn (1991). "Spontaneous mutation in the Escherichia coli lacI gene." Genetics **129**(2): 317-326.

Schlagman, S. L., S. Hattman and M. G. Marinus (1986). "Direct role of the Escherichia coli Dam DNA methyltransferase in methylation-directed mismatch repair." J Bacteriol **165**(3): 896-900.

Sisto, A., M. G. Cipriani, M. Morea, S. L. Lonigro, F. Valerio and P. Lavermicocca (2010). "An Rhs-like genetic element is involved in bacteriocin production by Pseudomonas savastanoi pv. savastanoi." Antonie Van Leeuwenhoek **98**(4): 505-517.

Stambuk, S. and M. Radman (1998). "Mechanism and control of interspecies recombination in Escherichia coli. I. Mismatch repair, methylation, recombination and replication functions." Genetics **150**(2): 533-542.

Studier, F. W., P. Daegelen, R. E. Lenski, S. Maslov and J. F. Kim (2009). "Understanding the differences between genome sequences of Escherichia coli B strains REL606 and BL21(DE3) and comparison of the E. coli B and K-12 genomes." Journal of molecular biology **394**(4): 653-680.

Tham, K. C., N. Hermans, H. H. Winterwerp, M. M. Cox, C. Wyman, R. Kanaar and J. H. Lebbink (2013). "Mismatch Repair Inhibits Homeologous Recombination via Coordinated Directional Unwinding of Trapped DNA Structures." Mol Cell **51**(3): 326-337.

Touzain, F., M. A. Petit, S. Schbath and M. El Karoui (2011). "DNA motifs that sculpt the bacterial chromosome." Nat Rev Microbiol **9**(1): 15-26.

van Diemen, P. M., F. Dziva, M. P. Stevens and T. S. Wallis (2005). "Identification of enterohemorrhagic Escherichia coli O26:H- genes required for intestinal colonization in calves." Infect Immun **73**(3): 1735-1743.

Wang, Y. D., S. Zhao and C. W. Hill (1998). "Rhs elements comprise three subfamilies which diverged prior to acquisition by Escherichia coli." J Bacteriol **180**(16): 4102-4110.

Welsh, K. M., A. L. Lu, S. Clark and P. Modrich (1987). "Isolation and characterization of the Escherichia coli mutH gene product." J Biol Chem **262**(32): 15624-15629.

Wu, T. H., C. H. Clarke and M. G. Marinus (1990). "Specificity of Escherichia coli mutD and mutL mutator strains." Gene **87**(1): 1-5.

Yoon, S. H., C. G. Hur, H. Y. Kang, Y. H. Kim, T. K. Oh and J. F. Kim (2005). "A computational approach for identifying pathogenicity islands in prokaryotic genomes." BMC Bioinformatics **6**: 184.
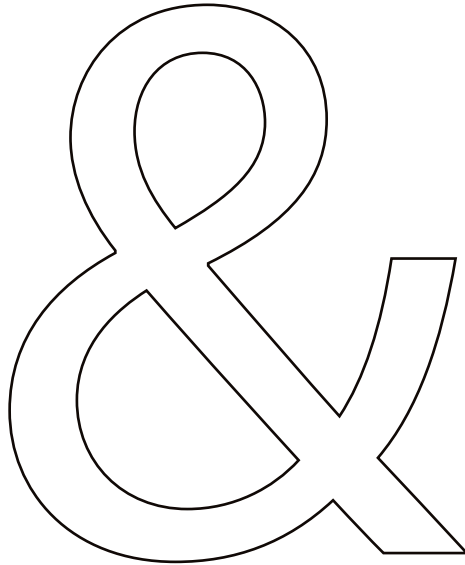
5

&

APPENDIX

SUMMARY

Life can be separated from dead organic matter by looking at two characteristics: growth and reproduction. For both of these, cells at some point need to split into two daughter cells. However, before cell division can take place, all the genetic information, encoded in DNA, needs to be copied. This process is called replication. Failure to replicate DNA correctly leads to mutations. These mutations can cause progenitor cells to have defects and die, or can cause cancer in higher organisms. DNA mismatch repair (MMR), the subject of study in this thesis, increases the fidelity of replication by removing mismatches left by the replication machinery.

Chapter 1 describes the mechanism of MMR, and implications of mutations that arise when mismatches are left uncorrected. Some of the most prevalent forms of hereditary cancers can be traced back to dysfunction of proteins involved in MMR. In *Escherichia coli*, MMR is initiated by MutS upon recognition of a DNA mismatch, resulting in ATP-dependent recruitment of MutL and activation of MutH. MutH is an endonuclease that is able to nick hemi-methylated DNA at a GATC motif, which provides an entry point for MMR to remove the strand with the error. The MMR pathway is conserved in most organisms, and MutS and MutL, the initiators of MMR in *E. coli*, are structurally very similar to their eukaryotic equivalents MutSα and MutLα. This emphasizes the importance of MMR, and sets the stage for consecutive chapters.

Chapter 2 deals with strand discrimination and excision during MMR. In *E. coli*, DNA is methylated by DAM methylase at GATC sites. Transiently hemi-methylated GATC sites provide the signal for distinguishing the newly synthesized DNA from the template strand. The efficiency of MMR *in vivo* depends on the number of GATC sites and the distance between mismatch and nearest GATC site. We quantitatively studied the rate of nicking by MutS, MutL and MutH, and subsequent strand excision by UvrD and ExoI, while varying the number of GATC sites and their distance from a GT mismatch. We find that *in vitro*, multiple nicks increase the efficiency of excision, while strand discrimination remains efficient over distances of 1 kb. Interestingly, we find a similar mechanism in human MMR. We propose a model where a single activated MMR complex facilitates efficient excision and repair by creating multiple daughter strand nicks.

Chapter 3 focuses on the time frame in which the individual steps are carried out. Using Monte Carlo simulations, we found that a simple diffusive model is able to predict many features of our data, which is a good indication that the incision complex uses a random walk to communicate between the mismatch and GATC sites. Combined with order-of-addition experiments, these simulations suggest that conformational changes in MutL are rate limiting for strand incision. Furthermore, the simulations show that the nicking efficiency

can be increased by multiple loading of MutS sliding clamps. This multiple loading could assure that strand discrimination takes place before the signal required to discriminate between the parental and daughter strand disappears.

Chapter 4 investigates the ATPase cycle of MutL. MutL is often described as a molecular matchmaker, and is able to activate the endonuclease activity of MutH and initiate excision of the error containing strand by loading UvrD onto the nick. Nucleotide binding in MutL was shown to switch a MutL dimer from an open to a closed conformation. These conformational changes govern the interaction of MutL with DNA and other MMR proteins. In this chapter we studied the interaction of MutL and several MutL mutants, deficient in either nucleotide binding, ATP hydrolysis or DNA binding, in functional assays with MutS, MutH, UvrD and DNA. We show that nucleotide binding and hydrolysis are not needed for the interaction of MutL with MutS. However, nucleotide binding is needed for activation of MutH and UvrD, and greatly increases the affinity of MutL for DNA. In this chapter we come to a model where MutL binds MutS in an open conformation and is able to activate downstream factors independent of MutS after being loaded onto the DNA in a closed conformation.

Chapter 5 deals with the distribution of GATC sites in the genome of *E. coli*. This distribution is not uniform, and is characterized by clusters of GATC sites and stretches devoid GATC sites. Because we know that the number and distance between GATC sites is important for the efficiency of MMR, we investigated large chromosomal DNA fragments without GATC sites for evidence of inefficient MMR. In *E. coli* cells without functional MMR, conversion of A(T) to G(C) is the predominant mutation. We show that DNA fragments larger than 3000 base pairs devoid of GATC sites have an elevated GC content, which suggests there is a locally diminished MMR efficiency. Also the genomes of other gram negative bacteria have DNA fragments devoid of GATC sites, and show a similar increase in GC content. These GATC deserts often include rearrangement hotspots (rhs) and Rhs genes. The absence of GATC sites in the Rhs genes provides an explanation for the increase of rearrangements and polymorphisms found in Rhs genes, and provides interesting insight in their function in inter-cellular competition and host-pathogen interaction.

SAMENVATTING

Twee van de belangrijkste kenmerken van leven zijn dat al het leven kan groeien en zichzelf kan vermenigvuldigen. Voordat een organisme zichzelf kan vermenigvuldigen moet alle genetische informatie, gecodeerd in DNA, worden gekopieerd. Dit proces wordt ook wel replicatie genoemd. Als er tijdens de replicatie van DNA fouten worden gemaakt, kan dit tot mutaties in het nieuwe DNA leiden. Deze mutaties kunnen er voor zorgen dat de nieuwe cel gebreken vertoont en dood gaat. In meercellige organismen, zoals mensen, kunnen mutaties ook kanker veroorzaken. Het vrijwel foutloos kopiëren van DNA is dus van het allergrootste belang. DNA mismatch repair (MMR), het onderwerp van dit proefschrift, repareert een groot gedeelte van de fouten die achterblijven na replicatie.

Hoofdstuk 1 beschrijft hoe MMR werkt, en wat de gevolgen zijn van mutaties die ontstaan wanneer fouten in het DNA niet gerepareerd worden. Dit laat zien hoe belangrijk MMR is; enkele van de meest voorkomende vormen van erfelijke kanker zijn terug te voeren op het niet correct functioneren van eiwitten die betrokken zijn in MMR. In *Escherichia coli*, een bacterie die veel gebruikt wordt in laboratoria over de hele wereld voor experimentele doeleinden, is de werking van MMR het meest onderzocht. In *E.coli* wordt MMR geïnitieerd door MutS. Dit eiwit kan fouten in DNA opsporen. Als het een fout vindt, verandert MutS van conformatie en initieert de reparatie door een tweede eiwit te binden. Dit tweede eiwit, MutL, activeert op haar beurt MutH. MutH is een eiwit dat in staat is om één van de twee strengen van dubbelstrengs DNA in te knippen. MMR is geconserveerd in de meeste organismen, en MutS en MutL zijn structureel zeer vergelijkbaar met hun equivalenten MutSα en MutLα in menselijk MMR.

Hoofdstuk 2 gaat over het onderscheiden van de originele en de nieuw gekopieerde DNA streng na replicatie, en hoe MMR gebruik maakt van deze informatie om eventuele fouten te repareren. In *E. coli* wordt DNA gemethyleerd door Dam-methylase op de A in de sequentie "GATC". Direct na replicatie zijn deze GATC sites nog niet gemethyleerd, en kunnen dus gebruikt worden om de nieuwe en de originele streng van elkaar te onderscheiden. De efficiëntie van MMR in vitro is afhankelijk van het aantal GATC sites en de afstand tussen de te repareren fout en dichtstbijzijnde GATC site. Dit hoofdstuk beschrijft metingen aan de efficiëntie van het knippen van GATC sites in DNA moleculen met een enkel ingebouwde fout. Hierbij variëren we het aantal GATC sites en de afstand tot de te repareren fout. We vinden dat in vitro het knippen van de DNA streng efficiënt blijft over afstanden tussen de GATC site en de fout van meer dan 1000 basenparen. De daaropvolgende excisie van de fout door twee andere eiwitten, UvrD en ExoI, is echter alleen efficiënt als er meerdere GATC sites rond de fout zijn gesitueerd. Aan de hand van deze resultaten laten we zien dat MMR meer-

dere knippen maakt om efficient DNA te kunnen repareren.

Hoofdstuk 3 focust op de timing waarop de afzonderlijke stappen in MMR worden uitgevoerd. Met behulp van Monte Carlo simulaties laten we zien dat een eenvoudig op diffusie van MutS gebaseerd model in staat is een groot gedeelte van onze experimenten correct te voorspellen. Deze simulaties, gecombineerd met experimenten waar we de tijd meten die het MutS-MutL complex nodig heeft om MutH te activeren suggereren dat het binden van ATP door MutL, en het vervolgens vormen van een actief complex, de belangrijkste factor is in het tijdig activeren van MutH. Bovendien laten de simulaties zien dat meerdere MutS moleculen samen werken om efficient en tijdig een fout te kunnen repareren.

Hoofdstuk 4 beschrijft de ATPase cyclus van MutL. MutL wordt vaak omschreven als een moleculaire koppelaar, omdat MutL als taak heeft de activiteiten van MutS enerzijds (die een fout kan opsporen) en van MutH en UvrD anderzijds (die zorgen dat de fout kan worden verwijderd) te koppelen. Het binden van ATP in MutL kan de conformatie van MutL overschakelen van een "open" naar een "gesloten" conformatie. Deze conformatieverandering ligt ten grondslag aan de interactie van MutL met DNA en andere mismatch repair eiwitten. In dit hoofdstuk bestuderen we de interactie van MutL met MutS , MutH, UvrD en DNA. Door het gebruik van verscheidene MutL mutanten die deficiënties hebben in het binden van ATP, in ATP hydrolyse of in DNA binding, testen we de functie van de conformatie verandering in MutL. We tonen aan dat ATP binding en hydrolyse door MutL niet nodig is voor de interactie van MutL met MutS. Echter, het binden van ATP is nodig voor het activering van MutH en UvrD en verhoogt de affiniteit van MutL voor DNA. Met deze resultaten zijn we instaat om enkele controversiele in het verleden gepubliceerde resultaten te verklaren, en bestaande modellen over de werking van MMR verder uit te breiden en te verfijnen.

Hoofdstuk 5 gaat in op de verdeling van GATC sites in het genoom van *E.coli*. GATC sites zijn niet gelijkmatig verdeeld over het genoom. Hun distributie wordt gekenmerkt door plekken waar clusters GATC sites dicht bij elkaar liggen, en lange stukken DNA waar nauwelijks GATC sites te vinden zijn. Omdat we weten dat het aantal GATC sites en de afstand tussen de fout en de GATC site belangrijk kan zijn voor het succes van MMR hebben we de langste DNA fragmenten in het genoom van *E.coli* zonder GATC sites bestudeerd op zoek naar bewijs van inefficiënte reparatie door MMR. E. coli cellen die een functioneel MMR missen, hebben een sterk verhoogde kans op mutaties. Tussen deze mutaties komt de omzetting van de bases A (T) naar G (C) het meeste voor. In dit hoofdstuk tonen we aan dat DNA fragmenten groter dan 3000 basenparen zonder GATC sites een verhoogd GC-gehalte hebben, wat overeenkomt met de meest voorkomen-

de mutatie bij MMR deficientie. Ook de genomen van andere gram-negatieve bacteriën (een groep bacteriën waar ook *E.coli* toe behoort) tonen een vergelijkbare toename in GC-gehalte in stukken DNA zonder GATC sites. De delen van het genoom zonder GATC sites bevatten vaak virusfragmenten en zogenaamde "Rhs" genen. Deze Rhs genen spelen een belangrijke rol in inter cellulaire competitie en pathogeniciteit van deze bacterien. Het afwezig zijn van GATC sites in deze genen geeft een interessante verklaring voor de grote variatie in deze genen tussen verschillende gerelateerde bacteriën, en heeft mogelijk een functie om de mutatie snelheid op deze plekken te verhogen.

&

CURRICULUM VITAE

Personal information:

Name:	Nicolaas Hermans
Date of birth:	21 October 1981
Place of birth:	Amsterdam, The Netherlands

Education and research:

2008-2014   PhD research in the genetics department,Erasmus MC at the Molecular Radation Biology group (Roland Kanaar) department of Genetics. Erasmus MC Rotterdam, The Netherlands.

2001-2008   Master of science in Molecular Biology, Rijksuniversiteit Groningen, Groningen, The Netherlands

2006   Research internship at the group of Kevin Arrigo on the topic of Iron limitation in Antarctic algae. Standford University, CA, USA

2005   Research internship at the Enzymology group of Bert Poolman on the diffusion of proteins inside bacterial cells. Groningen, The Netherlands

1994-2000   VWO, St. Ignatius gymnasium, Amsterdam, The Netherlands

LIST OF PUBLICATIONS

van den Bogaart G, Hermans N, Krasnikov V, de Vries AH, Poolman B. On the decrease in lateral mobility of phospholipids by sugars. **Biophys. J.** 2007 Mar 1;92(5):1598-605. Epub 2006 Dec 1.

van den Bogaart G, Hermans N, Krasnikov V, Poolman B. Protein mobility and diffusive barriers in Escherichia coli: consequences of osmotic stress. **Mol. Microbiol.** 2007 May;64(3):858-71.

Tham KC, Hermans N, Winterwerp HH, Cox MM, Wyman C, Kanaar R, Lebbink JH. Mismatch repair inhibits homeologous recombination via coordinated directional unwinding of trapped DNA structures. **Mol. Cell** 2013 Aug 8;51(3):326-37. doi: 10.1016/j.molcel.2013.07.008.

Groothuizen FS, Winkler I, Christovao M, Fish A, Winterwerp HH, Reumer A, Marx AD, Hermans N, Nicholls RA, Murshudov GN, Lebbink JHG, Friedhoff P, Sixma TK. Crystal structure of MutL loaded on the MutS sliding clamp. *Submitted.*

Quessada-Vial A, Monnet J, Hermans N, Graves E, Winterwerp HHK, Friedhoff P, Sixma TK, Lebbink JHG, Strick TR. Nicking single DNA molecules to study initiation of mismatch repair. *In progress.*

Hermans N, Kuzniar A, Lebbink JHG. A possible role for locally diminished DNA mismatch repair activity in establishing high GC content pathogenicity islands. *In progress.*

Hermans N, Cristovão M, Artola-Borán M, Winterwerp HHK, Wyman C, Friedhoff P, Jiricny J, Kanaar R, Lebbink JHG. On the efficiency of daughter strand nicking and excision during DNA mismatch repair. *In progress.*

Hermans N, Winterwerp HHK, Sixma TK, Kanaar R, Friedhoff P, Sneppen K, Lebbink JHG. A quantitative model for daughter strand discrimination reveals rate-limiting steps during DNA mismatch repair. *In progress.*

PHD PORTFOLIO

Courses:
Molecular and Cell biology
Radiation safety training (5b)
Laboratory Safety
EHBO / first aid

Workshops:
Annual MGC PhD workshop: *Brugge* (poster presentation), *Koln* (poster presentation), *Maastricht* (presentation).
"Genome Maintenance and Consequences" *Denmark, Kopenhagen* 2010

Conferences and seminars:
MGC DNA repair group meetings
Journal club
MGC Symposium (2008, 2010, 2012)
Presentations at mm2m 7th framework meetings in *Rotterdam* (2008), *Paris* (2009), *Amsterdam* (2009), *Zurich* (2010), *Brussels* (2010), *Egmond aan zee* (2011), *Budapest* (2012), *Amsterdam* (2012).
NWO-CW study group meeting on Protein, DNA and lipids. *Veldhoven* (2008-2010 poster presentation; 2012, presentation).
Annual Dutch Meeting on Molecular and Cellular Biophysics, *Veldhoven* (2009, poster presentation)

International conferences:
Responses to DNA damage: from molecular mechanism to human disease. *Egmond aan zee* (2011, poster presentation)
Gordon Mutagenisis. *Salve Regina, USA* (2012, poster presentation)

Teaching:
Junior science program
Supervision of Bachelor student research project (6 months)

DANKWOORD

Het schrijven van een proefschrift is niet altijd gemakkelijk. Tijdens al het geploeter zou je bijna vergeten dat promoveren ook erg leuk is, het is een voorrecht om in een inspirerende omgeving je eigen kennis uit te kunnen breiden zonder hier direct verantwoording voor af te hoeven leggen. Het beschikbaar en het inzichtelijk maken van die kennis voor anderen is een essentieel deel van het promoveren, en van het werk als wetenschapper in het algemeen, maar tijdens het schrijven is het soms moeilijk in te zien dat het schrijven van een proefschrift meer is dan alleen een vereiste om te kunnen promoveren. Gelukkig zijn er altijd mensen die je op het goede pad houden, door je te helpen, door het goede voorbeeld te geven, of door het verzorgen van de broodnodige afleiding. Ik wil dan ook iedereen bedanken die heeft bijgedragen aan de totstandkoming van dit boekje. Ik zal een aantal personen bij naam noemen, maar behalve de expliciet genoemde personen zijn er natuurlijk nog veel meer.

Joyce, tijdens mijn promotie heb ik veel geleerd. Een positieve instelling is een belangrijke eigenschap van een wetenschapper, hetgeen jij zeker hebt en iets dat mij veel geholpen heeft. Bedankt voor alle dingen die ik van je geleerd heb, onze succesvolle samenwerking, de wetenschappelijke discussies en het plezier! Roland, thanks for the support, the eagerness in reviewing results, and giving the opportunity to explore. Claire, thank you for all the helpful suggestions, also outside of research. I always greatly enjoyed our meetings and your open mindset.

Thanks to all the collaborators in the mm2m consortium. A select group of scientist making international collaboration more than just a pretext to get grant money. The scientific discussions during these meetings were essential for creating the experiments written down in this thesis. I'd explicitly like to thank Titia Sixma for helpful suggestions and for critically reading my thesis, Peter Friedhoff for the suggestion of some key experiments, Terence Strick for ironing out some flaws in my thinking with the logic of a physisist, Kim Sneppen for making the prototype of the model used in chapter 3, and being one of the most hospitable and open scientist I've ever met, and Joe Jiricny for providing the protocol for the DNA substrates so essential for my experiments.

Thanks to all the people in the Molecular Radiation biology lab and other scientist on the sixth floor. I've shared the lab with a number of people during the years, Iztok: I don't envy your struggle with the biacore, great job on making things work, and making some of the nicest protein preparations I've ever seen. KC, you can make the hardest experiment work, thanks for our collaboration. Michele, thank you for our shared experiments, I think we made a great team, and I've enjoyed all our climbing together. Moara, thanks for doing experiments with me, and introducing me to the local radio station. A special thanks to Marcel

for willing to be paranimph and for our nice collaboration, even though the project turned out to be more complicated than originally envisioned. Thanks to Nicole, Paula and Cecile for keeping it real. Thanks to Alex for making nerdy jokes even I can't understand and Joao for keeping me company if I had lunch at 2. Thanks to some former members of the lab, Tommy for the company, Daniel and Berina for PhD advice, and Kostas for introducing me to Perl. Also many thanks to the Vermeulen lab at the 7th floor, without you I'd never have reached the 10th place at the pubquiz.

Ook buiten het lab zijn er een hoop mensen die hebben bijgedragen aan dit boekje: Brigit, Abel, Titus, zonder jullie was ik nooit de persoon geworden die ik nu ben, reuze bedankt voor alle steun en al het plezier bij het huis aan de brug. Verder zijn er nog veel familieleden en vrienden die ik wil bedanken: bedankt allemaal, jullie zijn met teveel om op te noemen, waarschijnlijk een goed teken!

Ymkje, jij hebt mij altijd gesteund en het vertrouwen gehad dat ik het wel kon, ook op momenten dat ik het niet meer zag zitten. Maar minstens zo belangrijk zijn alle lol en avonturen die we de afgelopen jaren samen hebben beleefd, en ik hoop dat er de komende jaren nog minstens evenveel avonturen bij zullen komen. Ik weet zeker dat we die boot nog kunnen halen!

Klaas