

ATOMIC FORCE MICROSCOPY STUDIES OF EUKARYOTIC CLAMP/CLAMP
LOADER COMPLEX AND MISMATCH REPAIR INITIATION COMPLEX

Na Zhang

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements for the degree Doctorate of Philosophy of the
Department of Chemistry

Chapel Hill
2009

Approved by: Sergei Sheiko

Mathew Redinbo

Linda Spremulli

Thomas A. Kunkel

Dorothy A. Erie

ABSTRACT

Na Zhang

ATOMIC FORCE MICROSCOPY STUDIES OF EUKARYOTIC CLAMP-CLAMP LOADER COMPLEX AND MISMATCH REPAIR INITIATION COMPLEX (Under the direction of Dr. Dorothy A. Erie)

As an advanced single molecule technique, atomic force microscopy (AFM) is a powerful and versatile tool for high resolution surface imaging and probing physical properties of soft, nonconductive bio-materials *in vitro*. Imaging of protein-protein and protein-DNA complexes provides structural and conformational information about the interactions of these biomolecular assemblies. In this study, we have used AFM to examine two different protein complexes: the eukaryotic RFC complex function in loading PCNA clamp onto different DNA substrate and eukaryotic MutS homologs function in the initiation of DNA mismatch repair (MMR).

In the study of clamp loader RFC complex, we investigated the effect of nucleotide cofactors on the oligomerization states of RFC interacting with PCNA and DNA substrate. We observed that ATP binding induces a conformational change of RFC and that ATP hydrolysis causes RFC dissociation into small subcomplexes. However, PCNA inhibits the ATP-induced disassembly of RFC. Intriguingly, we found in the presence of ATP, some of the RFC subunits are ejected from DNA substrate, leaving RFC

subcomplex bound to the DNA, and it appears that these subcomplexes form stable interaction with PCNA on the DNA. We proposed that this DNA-bound RFC subcomplex tethers PCNA ring at the single strand/double strand junction of primer-template DNA or nick DNA. We further suggest that dissociation of RFC subcomplex from PCNA and DNA substrate is promoted by downstream PCNA-interacting proteins, such as DNA polymerase. In addition to these insights into the complicated potential loading mechanism of PCNA, we observed other RFC-DNA complexes such as RFC-DNA filaments with nicked DNA without nucleotide cofactor and RFC-DNA spider-like complexes containing multiple RFCs and DNAs in the presence of ATP. Although we do not know the physiological role, if any, of such RFC-DNA complexes, these complexes suggest RFC can possess other functions besides as clamp loader, such as helicase.

In the study of MMR initiation complexes, eukaryotic MutS homologs (MutS α and MutS β), we found, unlike their prokaryotic homologs, eukaryotic MutS homologs bind different DNA substrates with similar conformation. MutS α and MutS β both exhibits weak binding specificity to their specific DNA substrates, which makes it more complicated to analyze their specific complexes. However, it appears that eukaryotic MutS homologs do not recognize mismatched bases simply depending on the formation of unbent complexes as seen in the prokaryotic MutS. It is possible they employ other high class mechanism in which the event of recognition of different mismatched DNA substrates happens downstream of mismatch binding.

ACKNOWLEDGMENT

I would like to acknowledge my supervisor, Erie A. Dorothy, who has given me an incredible amount of help and guidance; not only in science but also as a mentor. Under her direction I have become a dedicated scientist. I appreciate all the time she took to instruct and guide me.

I am truly grateful for the help and encouragement of my previous lab mates: Dr. Ingrid Tessmer, Dr. Elizabeth Sacho and Susan Doyle. They taught me their precious expert knowledge about atomic force microscopy. I especially appreciate Dr. Yu Xue, Dr. Scott Kennedy, and Dr. Erika Pearson who led me into the field of molecular biology.

I would like thank to my current lab mates: Vanessa Van Vranken, Cherie Lanyi, and Dan Burke. I appreciate their encouragement and advice during my writing of this dissertation.

During the past six years, I cooperated with the people from Manju M Hingorani lab. Without their continued supply of protein, the project would not be complete.

My special gratitude goes to Charles Kim Stapleton, who is always infinitely patient, supportive, caring and lovely.

TABLE OF CONTENTS

	Page
LIST OF TABLES-----	vii
LIST OF FIGURES-----	viii
LIST OF ABBREVIATIONS-----	xi
Chapter	
1. Introduction of DNA Mismatch Repair and Clamp-Clamp Loader Complex	
1.1 Introduction of DNA Mismatch Repair (MMR)-----	1
1.2 Introduction of Clamps and Clamp Loaders-----	15
2. AFM Study of Eukaryotic RFC Loading PCNA Process	
2.1 Introduction-----	68
2.2 Materials and Methods-----	72
2.3 Results-----	73
2.4 Discussion-----	87
3 AFM study of Mismatch Recongnition by Eukaryotic MutS Homologs	

3.1 Introduction-----	133
3.2 Materials and Methods-----	137
3.3 Results-----	140
3.4 Discussion-----	146

LIST OF TABLES

Table 1.1 Identity and function of <i>E. coli</i> and eukaryotic proteins involved in MMR--	36
Table2.1. Relationship of <i>S.cerevisiae</i> RFC' MW with AFM volume-----	101
Table2.2 Statistical conformation analysis of RFC on 1077Nick DNA substrate in absence of nucleotide-----	113
Table 2.3 Statistical conformation analysis of RFC on PUC18Nick DNA substrate in absence of nucleotide-----	121

LIST OF FIGURES

Figure 1.1 Mechanism of <i>E. coli</i> methyl-directed mismatch repair-----	35
Figure 1.2 Reconstitute eukaryotic mismatch-provoked excision systems-----	37
Figure 1.3 Crystal structure of <i>E. coli</i> MutS binding to G·T mismatch-----	49
Figure 1.4 Crystal structure of Mut α binding with G·T mispair complex -----	41
Figure 1.5. Models for signaling downstream MMR events following mismatch recognition and mismatch recognition mechanism of <i>E. coli</i> MutS-----	43
Figure 1.6 Crystal structure of <i>E. coli</i> γ -complex and schematic view of the mechanism of opening β -clamp-----	46
Figure 1.7 Crystal structure of RFC clamp loader complex and schematic of the clamp loading process-----	58
Figure 1.8 Schematic view of AAA+ ATPase family -----	52
Figure 2.1 AFM images of RFC in the presence of different nucleotide cofactors-----	99
Figure 2.2 Histograms of RFC volume distributions with different nucleotide cofactors-----	102
Figure 2.3 Nucleotide binding introduces RFC conformational change-----	104
Figure 2.4 AFM images of RFC·PCNA complex in the presence of ATP and ATP γ S respectively-----	105
Figure 2.5 Histograms of volume analysis of RFC-PCNA complex in the presence of ATP and ATP γ S and histogram of volume analysis of PCNA-----	106
Figure 2.6 AFM images and histograms of volume distribution of RFC–primer template DNA complex with ATP γ S in the presence or absence of PCNA-----	108
Figure 2.7 Schematic view of DNA substrates used in AFM study-----	110
Figure 2.8 Representative AFM image of filament formation as RFC incubated with 1077Nick in the absence of nucleotide-----	111

Figure 2.9 AFM images of RFC and DNA complex in presence of ATP or ADP-----	114
Figure 2.10 Representative AFM image and the bending angle and binding position analysis of RFC in the presence of ATP γ S-----	116
Figure 2.11 Schematic view of RFC binding to the 1077Nick dsDNA in the presence of ATP γ S and histogram of volume analysis results-----	118
Figure 2.12 Representative AFM image and histogram of bent angle analysis of RFC on pUC18Nick DNA substrate in the absence of nucleotide-----	120
Figure 2.13 AFM image RFC with pUC18Nick DNA complex in the presence of ATP in high salt buffer. Histograms of bend angle and volume analysis of RF pUC18Nick DNA complex in the presence of ATP-----	122
Figure 2.14 AFM image RFC+PCNA with pUC18Nick DNA complex in the presence of ATP in high salt buffer. Histograms of bend angle and volume analysis of RFC –PCNA-pUC18Nick DNA complex in the presence of ATP-----	124
Figure 2.15 Schematic view of the hypothesis mechanism of RFC loading PCNA onto primer-template DNA substrate-----	126
Figure 3.1 Representative AFM image of Msh2Msh3 and histogram of volume distribution of Msh2Msh3-----	151
Figure 3.2: Schematic view of the DNA substrate used in AFM study-----	152
Figure 3.3 Representative AFM image of Msh2Msh3 in the presence homoduplex dsDNA substrate, the distribution of position Msh2Msh6 on homoduplex DNA and histogram of DNA bend angle introduced by Msh2Msh3 binding-----	153
Figure 3.4 Representative AFM image of Msh2Msh3 in the presence of T-bulge DNA substrate, position distribution of Msh2Msh3 on T-bulge heteroduplex DNA and histogram of DNA bend angle introduced by Msh2Msh3 binding-----	155
Figure 3.5 Representative AFM image of Msh2Msh3 in the presence of TT-bulge DNA substrate, position distribution of Msh2Msh3 on TT-bulge DNA and histogram of bend angle introduce by Msh2Msh3 bound at double T-bulge site and homoduplex site on 1125TTbulge DNA-----	157
Figure 3.6 Representative AFM image of human Msh2Msh3 in the presence of TT-bulge	

DNA substrate, position distribution of human Msh2Msh3 on TT-bulge DNA and histogram of bend angle introduced by human Msh2Msh3 bound at double T-bulge site and homoduplex site on 1125TTbulge-----159

Figure 3.7 Representative AFM image of Msh2Msh6 in presence of homoduplex DNA position distribution of Msh2Msh6 on homoduplex DNA and histogram of DNA bend angles introduced by Msh2Msh6 binding-----161

Figure 3.8 Representative AFM image of Msh2Msh6 in the presence of G/T mismatch DNA, position distribution of Msh2Msh6 on 982GT heteroduplex DNA and histogram of DNA bend angles introduced by Msh2Msh6 bound at G/T mismatch site and homogenous site on 982GT DNA substrate-----163

Figure 3.9 Representative AFM image of Msh2Msh6 in the presence of T-bulge DNA substrate, position distribution of Msh2Msh6 on T-bulge DNA and histogram of bend angle introduced by Msh2Msh6 bound at single T-bulge site and homoduplex site on 1124Tbulge DNA-----165

LIST OF ABBREVIATION

%	percent
α	alpha
β	beta
γ	gamma
δ	Delta
θ	theta
ϕ	phi
χ	chi
μL	microliter
~	approximately
$^{\circ}\text{C}$	degrees Celsius
3'	three prime end
5'	five prime end
A	adenine
\AA	angstrom

ADP	adenosine diphosphate
AFM	atomic force microscopy
Arg	arginine
ASCE	additional strand conserved E family
ATP	adenosine-5'-triphosphate
ATPase	adenosine triphosphatase
C	cytosine
cal	calorie
DNA	deoxyribonucleic acid
dsDNA	double stranded deoxyribonucleic acids
ssDNA	single stranded deoxyribonucleic acids
<i>E.coli</i>	<i>Escherichia coli</i>
EM	electron microscopy
<i>et. al</i>	and others
G	guanine
Glu	glutamate

IDL	insertion-deletion loop
<i>in vitro</i>	outside a living system
<i>in vivo</i>	inside a living system
IRC	initial recognition complex
URC	ultimate recognition complex
Kd	dissociation constant
kDa	kilo Dalton
m	meter
MMR	mismatch repair
mol	molar
nM	nanomolar
PCNA	proliferating cell nuclear antigen
PCR	polymerase chain reaction
Phe	phenylalanine
RFC	replication factor C
SRC	Serine-Arginine-Cysteine

SV40	simian virus40
T	thymine
<i>Taq</i>	Thermus aquaticus
Tyr	Tyrosine

CHAPTER ONE: INTRODUCTION OF DNA MISMATCH REPAIR AND CLAMP-CLAMP LOADER COMPLEX

1.1 Introduction of DNA mismatch repair (MMR)

Mismatched nucleotides arise from polymerase misincorporation errors, recombination between heteroallelic parental DNAs, and chemical or physical DNA damage. Mismatch repair (MMR) corrects DNA biosynthetic errors, increasing replication fidelity ~1000 fold, ensures the fidelity of recombination, and participates in the earliest steps of checkpoint, and apoptotic responses to several classes of DNA damage (Iyer, Pluciennik, et al, 2006; Kunkel, and Erie, 2005; Li, 2008; Modrich, 2006). Inactivation of MMR in humans is associated with hereditary nonpolyposis colon cancer and 15-25% sporadic tumors that occur in a number of tissues (Peltomaki, 2003; Peltomaki, 2005).

The *E.coli* methyl-directed Mismatch Repair Reaction

E. coli MMR is initiated when MutS recognizes base-base mismatches and small nucleotide insertion/deletion (IDL) mispairs in DNA (Figure 1.1). MutL interacts physically with MutS, enhances mismatch recognition, and activates MutH to nick the daughter strand. Because MutL stimulates the loading and processivity of helicase II (or UvrD) at the MMR initiation site, MutL may play a role as a molecular matchmaker that

facilitates the assembly of a functional MMR complex (Dao, and Modrich, 1998; Guarne, Ramon-Maiques, et al, 2004). Both MutS and MutL function as homodimers and possess ATPase activity (Ban, and Yang, 1998a). In *E. coli*, DNA is methylated at the N6 position of adenine in dGATC sequences. In replicating DNA, the daughter strand is transiently unmethylated, and it is the presence of hemimethylated dGATC sequence that molecularly distinguishes the newly synthesized daughter strand from the parental DNA strand. In MMR, MutH recognizes hemimethylated dGATC sequences and functions as monomer. Upon its activation by MutS and MutL in the presence of ATP and a mismatch, MutH specially incises the unmethylated daughter strand of hemimethylated dGATC, and the strand specific nick is a starting point for mismatch-provoked excision (Ban, and Yang, 1998b; Lee, Chang, et al, 2005; Ramilo, Gu, et al, 2002). In the presence of MutL, helicase II (UvrD) loads at the nick and unwinds the duplex from the nick towards the mismatch, generating single-strand DNA. The parent strand is rapidly bound by single-stranded DNA-binding protein (SSB) and protected from nuclease attack (Dao, and Modrich, 1998; Ramilo, Gu, et al, 2002). Depending on the position of the strand break relative to the mismatch, ExoI or ExoX(3'-5' exonuclease), or ExoVII or RecJ (5'-3' exonuclease) excises the nicked strand from the nicked site (the dGATC site) up to and slightly past the mismatch. The resulting single gap undergoes DNA resynthesis and ligation by DNA polymerase III holoenzyme, SSB and DNA ligase (Modrich, and Lahue, 1996).

Proteins Homolog in Eukaryotic Mismatch Repair

Several eukaryotic MMR proteins have been identified based on their homology to *E.coli* MMR proteins. These include eukaryotic homologs MutS, MutL, EXOI, single strand DNA-binding protein RPA, proliferation cellular nuclear antigen (PCNA), DNA polymerase δ or ϵ and DNA ligase I (Table 1.1).

-MutS homologs-

Eukaryotic cells possess two MutS homologs that function as heterodimers and share Msh2 as common subunit: MutS α (Msh2·Msh6 heterodimer) and MutS β (Msh2·Msh3 heterodimer). MutS α , which represents 80-90% of the cellular Msh2, preferentially recognizes base-base mismatches and insertion/ deletion (IDL) mispairs in which one strand contains 1 or 2 unpaired nucleotides, but it is also capable of recognizing of large IDL heterologies with reduced affinity (Drummond, Li, et al, 1995; Genschel, Littman, et al, 1998; Palombo, Gallinari, et al, 1995; Palombo, Iaccarino, et al, 1996). MutS β recognizes IDL mismatches of 2-to ~10 nucleotides, weakly recognizes single-nucleotide IDL mispairs, and is essentially inert on base-base mismatches (Genschel, Littman, et al, 1998; Palombo, Gallinari, et al, 1995).

-MutL homologs-

Three eukaryotic MutL homologs have been identified and like eukaryotic MutS homologs, they function as heterodimeric complexes with MLH1 as a common subunit. MutL α , a heterodimer of MLH1 and PMS2 (PMS1 in yeast), is the primary MutL homolog, acting in human mitotic cells and supports repair initiated by either MutS α or

MutS β (Li, and Modrich, 1995). In a reconstituted human MMR system, human MutL α regulates termination of mismatch-provoked excision (Zhang, Yuan, et al, 2005). Recent studies show that MutL α possesses a PCNA/RFC-dependent endonuclease activity that plays a critical role in 3' nick –directed MMR involving EXO1 (Kadyrov, Dzantiev, et al, 2006). Although MutL α accounts for ~90% of the MLH1 in human cells, two low abundance complexes involving MLH1 have also been identified. Human heterodimer MLH1·PMS1 (MutL β) has been isolated, but its molecular activity have not been ascertained (Raschle, Marra, et al, 1999). MLH1 heterodimerizes with MLH3 to form MutL γ , and the MutL γ complex has been reported to support modest levels of base-base and single nucleotide IDL mismatch repair in vitro, events that are presumably initiated by MutS homologs in vitro (Cannavo, Marra, et al, 2005).

-PCNA and RFC-

PCNA plays multiple roles in MMR. PCNA interacts with Msh2 and MLH1 and is thought to play roles in the initiation steps of MMR (Clark, Valle, et al, 2000; Gu, Hong, et al, 1998; Umar, Buermeier, et al, 1996). PCNA also interacts with Msh2·Msh3 and Msh2·Msh6 via a conserved PCNA interaction motifs in Msh3 and Msh6, termed the PIP box (Clark, Valle, et al, 2000; Warbrick, 2000). PCNA increases the mismatch –binding specificity of Msh2-Msh6, and it can assist in delivering of Msh2-Msh6 to mismatched DNA (Flores-Rozas, Clark, and Kolodner, 2000; Lau, and Kolodner, 2003). PCNA and polymerase δ also have been implicated in repair DNA resynthesis, as PCNA confers processivity on DNA polymerase δ (Genschel, and Modrich, 2003; Gu, Hong, et al, 1998;

Umar, Buermeyer, et al, 1996). PCNA is also required for mismatch-provoked excision. The most compelling evidence for PCNA involvement in the excision step of MMR has been provided by p21 inhibition study. p21 forms a stable complex with DNA-bound PCNA so that p21 interferes with downstream PCNA-dependent events after PCNA loading in MMR. p21 is found to abolish 3'-directed mismatch-provoked excision in HeLa cell extracts, however only 40-50% of 5'-directed excision events are subjected to p21 inhibition, implying occurrence of at least two hydrolytic modes of 5'-directed excision (Genschel, and Modrich, 2003; Guo, Presnell, et al, 2004; Umar, Buermeyer, et al, 1996).

RFC provides at least two functions in the reconstituted bidirectional excision system: loading of PCNA onto helix and suppression of the 5' to 3' hydrolysis from a 3' strand nick. In vital, the amino terminal ligase homology domain of the large RFC subunit, which is not required for PCNA loading on the helix, is important for suppression of 5' to 3' hydrolysis from a 3'-strand break but is not required for activation of 3' directed excision. Domain II of the large RFC subunit (Figure 1.6, a), which functions in PCNA loading, is not necessary for suppression of 5' to 3' hydrolysis from the 3'-nick but it is required for activation of 3'-directed excision. These results suggest a mechanism in which face of PCNA would be oriented toward the mismatch in 3' and 5' heteroduplexes and the interaction of PCNA with ExoI will determine the direction of excision (Dzantiev, Constantin, et al, 2004).

-ExoI-

ExoI is a 5'-3' exonuclease, and is involved in both 5' and 3' directed MMR. ExoI can readily carry out 5' directed mismatch excision in the presence of MutS α or MutS β and RPA (Genschel, and Modrich, 2003; Zhang, Yuan, et al, 2005); however, its role in catalyzing 3' nick-directed excision requires the MutL α endonuclease activity, which is activated by PCNA and RFC. After recognition of the 3' nick and the MutS- mismatch complex, MutL α endonuclease makes incisions 5' and 3' to the mismatch in a manner dependent on PCNA and RFC. Exo I performs 5' to 3' excision from the MutL α incision site through and beyond the site of the mismatch (Dzantiev, Constantin, et al, 2004; Kadyrov, Dzantiev, et al, 2006).

-RPA-

RPA seems to be involved in all stages of MMR. It appears to bind to nicked heteroduplex DNA before MutS α and MutL α and stimulates mismatched-provoked excision. In addition, it protects the ssDNA gapped region generated during excision and facilitates DNA resynthesis (Dzantiev, Constantin, et al, 2004; Ramilo, Gu, et al, 2002; Zhang, Yuan, et al, 2005). Furthermore, RPA is phosphorylated after polymerase δ is recruited to the gapped DNA substrate. Recent studies indicate that (i) phosphorylation reduces the affinity of RPA to the DNA, (ii) Unphosphorylated RPA stimulates mismatch-provoked DNA excision more efficiently than phosphorylated RPA, and (iii) phosphorylated RPA facilitates MMR-associated DNA resynthesis more efficiently than unphosphorylated RPA. These results are consistent with the idea that a high affinity RPA-DNA complex might be required to protect nascent ssDNA and to displace DNA

bound MutS α /MutL α , while a lower affinity RPA-DNA complex might facilitate DNA resynthesis by Pol δ (Guo, Zhang, et al, 2006).

-Polymerase δ and ϵ -

The editing exonuclease functions of DNA polymerase δ and ϵ have been postulated to provide hydrolytic functions in mismatch repair (Tran, Gordenin, and Resnick, 1999; Wang, and Hays, 2002), however this suggestion has been questioned (Datta, Schmeits, et al, 2000). The evidence for polymerase δ being involved in MMR resynthesis comes from the depleted in DNA polymerase extracts system that sustains mismatch-provoked excision but fails to support complete repair reaction is restored repair with highly purified fraction of polymerase δ devoid of datable pol α and pol ϵ . Thus polymerase δ is required for eukaryotic mismatch correction, but supporting roles for polymerase α and ϵ have not been ruled out (Longley, Pierce, and Modrich, 1997).

-HMGB1-

HMGB1, a non-histone chromatin protein, binds to certain type of DNA damage, Interacts with MutS α and may play an important role in early steps of the reaction prior to excision (Zhang, Yuan, et al, 2005).

Reconstituted Eukaryotic Mismatch Excision/Repair System

Several systems have been reconstituted to support mismatch-provoked excision by purified human proteins (Figure, 1.2). The simplest one is comprised of only four proteins, MutS α , MutL α , ExoI and RPA, which supports mismatch-dependent 5' to 3'

hydrolysis from a 5' strand break. MutS α activates ExoI hydrolysis on a 5'-heteroduplex in a mismatch- and ATP- dependent manner. In the absence of RPA, MutS α renders the ExoI highly processive, resulting removal of ~2000 nucleotides prior to dissociation, an effect attributed to formation of a MutS α ·ExoI complex. However, RPA modulates the behavior of this complex, reducing the processivity of ExoI to ~250 nucleotides. A single strand gap coated with RPA is an extremely poor substrate for ExoI, but MutS α promotes ExoI initiation at such sites provided that the DNA contains a mismatched base pair. It has been suggested that hydrolysis is dramatically attenuated upon mismatch removal because of the absence of MutS α . MutL α also plays an important role of acting with MutS α to suppress ExoI hydrolysis on DNA that lacks a mispair (Genschel, and Modrich, 2003; Zhang, Yuan, et al, 2005).

This four protein system also supports mismatch-provoked excision on 3'-heteroduplex. However, the hydrolysis precedes 5' to 3' from 3' strand break, which is the wrong polarity for mismatch removal. Adding PCNA and RFC to the reaction yields a system that supports mismatch removal from both 5' and 3' heteroduplex. The nonspecific 5' to 3' hydrolysis activity initiating at the 3' nick is largely suppressed by RFC, and excision occurs with apparent 3' to 5' polarity resulting in mismatch removal (Dzantiev, Constantin, et al, 2004). Furthermore, other research shows that MutS α , RFC, and PCNA activate a latent endonuclease activity in MutL α in an ATP and mismatch-dependent manner. Incision by activated MutL α endonuclease occurs both 3' and 5' to the mismatch on the nicked strand. In the case of 3'-heteroduplex which the nick is on the 3' side of mismatch, incision distal to the mismatch provides an initiation site for mismatch

removal by the 5' to 3' action of MutS α -ExoI complex. The PCNA-dependent endonucleolytic system also incises 5'-heteroduplex strand (Kadyrov, Dzantiev, et al, 2006). This model posits that the nick serves as a strand signal but not as site for excision initiation, which actually occurs at strand break produced by mismatch-activated MutL α endonuclease. This result is consistent with the result found in *Xenopus* egg extracts, which demonstrated a significantly higher specific radioactivity in the vicinity of the mismatch relatively near the strand break, through analyzing the repair products incorporated with radiolabeled nucleotides (Varlet, Canard, et al, 1996).

The established systems to date are minimal systems. The fact the excision products in purified systems are more dispersed than those observed in nuclear extracts indicates the probable existence of other excision activities that may function in a redundant manner with respect to ExoI (Genschel, Bazemore, and Modrich, 2002). Furthermore, the mismatch dependence of reconstituted 5'-directed excision is not as dramatic as that observed in nuclear extracts indicating that one or more specificity factors may be lacking (Dzantiev, Constantin, et al, 2004; Genschel, and Modrich, 2003).

Coupling of Mismatch Recognition and Strand Discrimination

A major question of MMR is how the MMR proteins facilitate the communication between two physically distant DNA sites: the mismatch and the strand discrimination signal. The role of hemi-methylated dGATC sites as signal for strand discrimination is not conserved from *E. coli* MMR to eukaryotic MMR. However, because the hemi-methylated dGATC directs MutH-dependent nicking as a starting point for mismatch-

provoked excision, human MMR is presumed to be nick directed in vital, it is generally agreed the strand discrimination signal is a strand specific nick in both prokaryotic and eukaryotic systems, but how the nick is generated in eukaryotes remains unknown.

Previous studies have proposed several alternative models for the mismatch and strand discrimination signal process, which can be classified into a “stationary” model, a “translocation” model and a “sliding clamp” model (Figure 1.5, a). The stationary model proposes that interaction among MMR proteins induces DNA bending or looping that brings these two distant sites together, while MutS remains bound at or near mismatch. In this model, the MutS ATPase activity acts in a proof-reading role to verify mismatch binding and authorize the downstream excision (Guarne, Ramon-Maiques, et al, 2004; Junop, Obmolova, et al, 2001). In the translocation model, ATP reduces the mismatch binding affinity of MutS or MSH heterodimer, and ATP hydrolysis drives bidirectional translocation of MutS proteins along the DNA helix. DNA is threaded through the protein complex until the latter reaches a strand discrimination signal in either orientation, resulting formation of a loop (Allen, Makhov, et al, 1997). In the “sliding clamp” model, MutS or MSH heterodimer binds to DNA in an ADP-bound state. The mismatch binding triggers a MutS or MSH herterodimer conformational change that allows an ADP to ATP exchanges. The ATP binding, not ATP hydrolysis signals downstream events, including formation of a ternary sliding clamp with MutL and sliding from mismatch to the nick (Fishel, 1998; Genschel, and Modrich, 2003; Jiang, Bai, et al, 2005; Mendillo, Mazur, and Kolodner, 2005).

Structure and Function of *E.coli* MutS and Human MutSa

The crystal structures of *E. coli* MutS with G-T, C-A, A-A, G-G and an unpaired T heteroduplex DNA, and *Taq* MutS with an unpaired T heteroduplex are strikingly similar (Lamers, Perrakis, et al, 2000; Natrajan, Lamers, et al, 2003; Obmolova, Ban, et al, 2000). MutS has domains (Figure 1.3). Domains I (mismatch binding domain) and VI (clamp) bind to DNA, and domain V (ATPase domain) contains the dimerization interface and nucleotide-binding site. The DNA and nucleotide binding sites are widely separated but connected by domain III (lever), which interacts directly with domain IV and indirectly with domain I via domain II (connector). The homodimer of MutS forms a clamp-like structure similar to the Greek letter “Θ” with two large channels. The upper channel is large enough and has an electrostatic potential that could bind DNA, leading to the idea that it may function in DNA recombination or in the search for the strand discrimination signal; however, no evidence of DNA binding in the upper channel exists (Yang, Junop, et al, 2000). The heteroduplex DNA is kinked $\sim 60^\circ$ and bound to the lower channel of MutS dimer. In crystals without DNA, the DNA binding domains and part of the connector domain are disordered. These results indicate that mismatched DNA binding induces large conformational changes both the DNA and in MutS. Although MutA is comprised two identical monomers, the heteroduplex bound dimer is structurally asymmetric with mismatch recognition contacts provided by only one subunit. Two residues of Phe-X-Glu motif in the same subunit make mismatch base specific contacts. In both the *E.coli* and *Taq* structures, the unpaired or mispaired base is rotated into the minor groove by $\sim 3\text{\AA}$ and the phenylalanine of the Phe-X-Glu motif stacks with

mispaiored base. The glutamic acid hydrogen bounds to the same base (to N7 if the base is a purine and to N3 if it is pyrimidine). Most contacts from two subunits are to the DNA backbone and are, therefore, DNA sequence nonspecific, as expected given the need to repair replication errors in a variety of different sequence contexts.

The crystal structure of human MutS α , with Msh6 lacking the first 340 amino acids, in complex with G: T mismatch DNA shares common domain architecture with prokaryotic MutS. Each protein can be divided into five domains, referred as mismatch binding domain, connector, levers, clamps and ATP binding domains (Figure 1.4). Msh2·Msh6 heterodimer forms an asymmetric oval disc, pierced by two channel, like the letter “ Θ ”. A DNA helix containing a single mispair is bent by $\sim 45^\circ$ and bound in the lower channel. Only Msh6 makes specific contacts with mispair through conserved Phe-X-Glu motif. Furthermore, the N-terminus of Msh6 (residues 360-398) contains additional nonspecific contacts with the DNA backbone. By contrast, the DNA binding domain of Msh2 is rotated up and away from backbone and makes only one contact with the DNA, which is different from the prokaryotic homodimer. Interactions between Msh6 and the DNA substrate bury 1142 \AA^2 (with domain I contributing 856 \AA^2), which is probably sufficient to bend the DNA without any contribution from Msh2. The upper channel is surrounded by disordered loops and is too small to contain a second DNA double strand as been proposed in prokaryotic analog (Warren, Pohlhaus, et al, 2007).

The ATPase domain (domain 5, Figure1.4, a) is the most highly conserved region of the MutS homologs. Domain 5 from *E.coli* and hMsh2 are 48% identical. As member

of ABC-transporter ATPase family, MutS homologs have two composite ATP sites, each consisting of a Walker A and Walker B motif from one monomer and ABC signature motif from alternative monomer. In addition, the C-termini of both ATPase domains form conserved helix-turn-helix motifs that interact with ATPase domain of the opposed protomer. This interaction stabilizes the ABC-ATPase dimer interface in the absence of ATP binding, and is required for dimerization of canonical ABC-transporter ATPase domains (Lamers, Perrakis, et al, 2000; Obmolova, Ban, et al, 2000; Warren, Pohlhaus, et al, 2007). The ATP binding sites are asymmetric as is the mismatch binding sites. In the absence of DNA, one site has a high affinity for ADP and the other has a high affinity for ATP (Antony, and Hingorani, 2004; Bjornson, and Modrich, 2003; Blackwell, Bjornson, et al, 2001). In eukaryotes, the Msh6 subunit contains a high-affinity ATP binding sites and Msh2 contains a high-affinity ADP binding site (Antony, and Hingorani, 2003; Blackwell, Martik, et al, 1998; Gradia, Acharya, and Fishel, 2000). Stable binding of ATP to Msh6 appears to cause a decreased affinity of Msh2 for ADP (Mazur, Mendillo, and Kolodner, 2006). Furthermore, the two classes of sites can be simultaneously occupied by ADP and an ATP analogue, suggesting that the ADP·MutS·ATP species may be highly populated in solution. However, the high concentrations of ATP or nonhydrolyzable ATP analogs can apparent compete out the ADP, suggesting that a complex of MutS with two ATP molecules bound can form. ADP can also bind to both ATPase sites (Antony, and Hingorani, 2003; Antony, and Hingorani, 2004; Bjornson, and Modrich, 2003; Martik, Baitinger, and Modrich, 2004).

Allosteric coupling between the ATPase sites and the DNA binding sites plays a central role in the MutS homologs mechanism. DNA substrate binding is known to modulate the ATPase activity: both hetero and homoduplex DNA stimulate ADP-ATP exchange, but only heteroduplex DNA appears to change the rate limiting step for turnover of ATP. In the presence of ATP alone or ATP and homoduplex DNA, there is burst of hydrolysis of one ATP equivalent per dimer; whereas, in the presence of mismatch, the burst of hydrolysis is suppressed (Acharya, Foster, et al, 2003; Antony, and Hingorani, 2003; Bjornson, Allen, and Modrich, 2000). Similarly, ATP binding modulates DNA substrate binding. ADP has little effect on the affinity of MutS proteins for mismatch DNA. However, ATP and ATP analogs decrease mismatched DNA-binding affinity, although substantial mismatch specificity is retained under conditions that support ATP hydrolysis (Blackwell, Martik, et al, 1998; Blackwell, Bjornson, et al, 2001; Gradia, Acharya, and Fishel, 1997; Hess, Gupta, and Kolodner, 2002; Joshi, Sen, and Rao, 2000; Martik, Baitinger, and Modrich, 2004).

Like other DNA repair enzymes, MutS homologs must locate a subtle base pair anomaly within a vast excess of nonsubstrate, correctly paired DNA. Insights into repair specificity come from recent AFM studies that directly visualize a MutS bound to mismatched and to homoduplex DNA (Wang, Yang, et al, 2003). MutS- homoduplex complexes are bent at homoduplex site, but MutS –heteroduplex complexes are bimodal with a significant fraction of the complexes unbent. These results indicate the unbent state is the result of unique interactions between the mismatch base and MutS, and suggest that the bent conformation may be an intermediate in the formation of the unbent

state. These data led to the proposal that MutS binds to DNA nonspecifically and bends in search of mismatch (Figure 1.5, b), and upon specific recognition of a mismatch, MutS undergoes a conformational change to an initial recognition complex (IRC) in which the DNA is kinked, with interactions similar to those in the crystal structures. MutS then undergoes further conformational changes to the ultimate recognition complex (URC) in which the DNA is unbent with the mismatched base possibly being flipped out.

Similar AFM studies have not been conducted on the eukaryotic homologs. Strikingly, the crystal structure of human MutS α reveals that it recognizes different DNA substrates in a similar manner (Warren, Pohlhaus, et al, 2007) . MutS α recognizes the G·T mispair or a single base T insertion /deletion loop in the DNA MMR , as well as an O⁶-methyl guanine·T mispair in a similar manner although mismatch initiates repair and O⁶-methyl initiates apoptosis. MutS α also recognizes a G·dU mispair, a putative immediate in somatic hypermutation. All the crystal structures of MutS α -DNA complexes are virtually identical. This result is surprising given the significant differences in DNA base pairing, conformation, and thermal stability of these mismatches (Kramer, Kramer, and Fritz, 1984; Kramer, Kramer, et al, 1989). This observation led to the suggestion that diversity of MutS α dependent responses to DNA lesions is generated in events downstream of the lesion recognition step or they all crystallize in the same conformation.

1.2 **Introduction of clamps and clamp loaders**

Chromosome replication requires a DNA polymerase that can rapidly duplicate thousands of nucleotides. In all cells, the replicative polymerase is tethered to DNA by a ring-shaped clamp that encircles DNA and slides freely along it. This interaction confers a high degree of processivity to the replicative polymerase. Ring-shaped sliding clamps are constructed from either two or three identical crescent-shaped protomers and their closed circular structure requires an active mechanism to open them for assembly onto DNA. This loading process is carried out by multiprotein machines known as clamp loaders, which crack open the ring and place it around the primed DNA in an ATP driven reaction.

Brief introduction of AAA+ proteins family

The defining feature of AAA+ proteins is a structurally conserved ATP-binding module that oligomerizes into active arrays. ATP binding and hydrolysis events at the face of neighboring subunits drive conformational changes within the AAA+ assembly that direct translocation or remodeling of target substrates (Erzberger, and Berger, 2006). AAA+ proteins are involved in a myriad of biological process, including ATP-dependent proteolysis, membrane fusion, protein trafficking, and DNA replication, recombination, and repair.

As a subfamily of additional strand conserved E family (ASCE), AAA+ proteins have conserved ASCE core nucleotide-binding pocket lying at the apex of three adjacent, parallel β -strands in a compact $\alpha\beta\alpha$ -fold (Figure, 1.8 a). The primary distinguishing feature of AAA+ proteins is a β -strand addition to the ASCE core, together with a small

helical bundle fused to the C terminus of the central $\alpha\beta\alpha$ -fold. Several other features are characteristic of this family (Figure 1.8, b). For example, the conserved arginine residues (arginine finger) of homology SHC or SRC motif interact with nucleotide-binding pocket of neighboring subunit. In addition, there are two other nucleotide-interaction motifs termed sensor I and sensor II elements. The sensor I motif, which resides at the top of strand β_4 , is typically an Asn, although other polar residues, such as Ser, Thr, or Asp are also found. This residue is believed to act in concert with the second acidic residue of the Walker-B motif to properly orient a water molecule for nucleophilic attack on the γ -phosphate of ATP. Sensor II usually contains an arginine at the base of helix α_7 that interacts with the γ -phosphate of ATP (Guenther, Onrust, et al, 1997).

Seven major clades of AAA+ proteins have been defined on the basis of sequence alignments and structural information (Figure 1.8,c). The differences between clades arise from the insertion of secondary structural elements at defined places within and around the core AAA+ fold (Iyer, Leipe, et al, 2004).

Clade 1 is represented by the clamp loader family. The pentameric complex recognizes the primer-template junctions and promotes the opening and loading of the polymerase processivity clamp onto DNA.

Clade 2, the initiator clade, includes all cellular origin-processing proteins, as well as the bacterial, archaeal, and eukaryal helicase-loading proteins. This group is defined by the presence of an extra α -helix inserted between the second and third strands of the central β sheets. DNA replication is initiated at distinct chromosomal sites known as

replication origins. The initiators are responsible for properly recognizing replication origins and for enabling the subsequent assembly of the replication machinery (Messer, 2002). Most of AAA+ initiators are monomeric in solution and assemble into oligomeric structures when bound to specific target DNA sequences at origins (Cunningham, and Berger, 2005). Remarkably, DnaA has been revealed to form a helical filament on DNA in presence of ATP. The filament arrangement places DNA-binding domains appended to the C-terminus of the AAA+ module on the outside of the filament. This orientation suggests that the interaction with origin DNA sequences may lead to the introduction of right-handed helical wrap to this region, destabilizing adjacent DNA elements as a prelude to origin melting and replisome assembly. In addition, the assembly at the boundaries of the filament, exposing an arginine finger at one end and a nucleotide-binding pocket at the other end, may serve as a docking sites for auxiliary replication factors that also contain AAA+ domains (Erzberger, Mott, and Berger, 2006).

Clade 3 is a group of proteins that form closed hexameric assemblies. These classic clade proteins are involved in processes such as protein degradation, microtubule serving, membrane fusion, and peroxisome biogenesis. Classic AAA+ modules share a small helical insertion before helix $\alpha 2$ of the ASCE fold. The classic clade is unique for possessing two arginine residues instead of a single residue commonly see in other AAA+ proteins (Erzberger, and Berger, 2006).

Clades 4, 5, 6 and 7 share a characteristic β -hairpin insertion between helix 3 and strand 4. They are composed of Pre-sensor I insertion superclade. The β -hairpin

insertion appears to link ATP turnover to DNA translocation. Clade5, composed of the HslU, ClpAB-CTD, LonAB, and RuvB families (HCLR clade), represents the central group of the PS-I superclade. Clade 4 differs from clade 5 with an unusual bundle formed by elements N- and C- terminal to the ASCE core instead of AAA+ lid. Clade 6 and Clade 7 contain an additional β -hairpin insertion that disrupts helix $\alpha 2$.

There is little correlation between AAA+ protein subtypes and a specific remodeling activity. Nevertheless, Clade 1 and Clade 2 are characterized by open-ring assemblies that differ significantly from other AAA+ proteins, most of which form closed hexameric rings. These kinds of assemblies appear to function on as a single turnover allosteric switches, rather than processive molecular machines. ATP binding appears to flip these proteins into an activated state by influencing the conformation of the AAA+ modules and enabling them to remodel respective proteins.

***E.coli* and bacteriophage T4 clamp loader**

In *E. coli*, the β - clamp is a ring-shaped homodimer that encircles the double-stranded DNA and tethers polymerase (Pol III) to the template for high processivity (Kong, Onrust, et al, 1992; Stukenberg, Studwell-Vaughan, and O'Donnell, 1991). The clamp loader is required to open the closed ring of the β clamp and load it onto the template in an ATP-driven process.

In *E. coli*, the clamp loader is known as γ -complex, which comprises five subunits: δ , δ' and three copies of γ . Two other subunits associated with clamp loader χ and ψ , are

not required for clamp loading *in vitro* (Onrust, and O'Donnell, 1993). However, the $\psi\cdot\chi$ complex provides the linkage between the clamp loader and single strand binding proteins (SSB), and this linkage increases the replication activity. The ψ protein binds tightly to the collar domain of the γ complex and the χ tucks onto the C-terminal tail segment of SSB, thus bridging the clamp loader and SSB (Glover, and McHenry, 1998; Kelman, Yuzhakov, et al, 1998; Olson, Dallmann, and McHenry, 1995; Simonetta, Kazmirski, et al, 2009; Witte, Urbanke, and Curth, 2003).

The crystal structure of the γ -complex revealed that the five subunits are arranged in a spiral (Figure 1.6, a), with δ and δ' bracketing the three γ subunits. The C-terminal domain of each subunit mediates oligomerization into a spiral heteropentamer. The primer-template DNA is thought to get out of the gap between δ and δ' subunit. The three γ subunits are the only subunits of the clamp loader that bind ATP, and therefore comprise the 'motor' that drives the clamp loading reaction. The δ subunit cracks open the β -dimer, even in the absence of γ and δ' , at one interface, and thus is termed the 'wrench'. The δ' seems to be a rigid protein that lacks the flexible motions that occur in the δ and γ subunits. The rigidity allows δ' to act as a stator or a backboard, directing ATP-induced conformational changes, propagating from γ_1 through γ_2 , γ_3 and finally through δ , pulling δ away from δ' (Jeruzalmi, Yurieva, et al, 2001).

The γ complex belongs to AAA+ family of ATPases, although only the γ subunits have functional ATPase sites. The ATP-binding sites are at the interface of the pairs of subunits, resulting in three ATP sites within the γ complex at the δ'/γ_1 , γ_1/γ_2 and γ_3 . The

arginine finger extends toward to the ATP binding site in the adjacent subunit and senses ATP binding and hydrolysis. Crystal structures of the nucleotide free and ATP γ S bound states revealed that only two of three ATP binding sites (γ 1 and γ 3) are occupied by ATP γ S, with the γ 2 subunit binding site being physically blocked by the γ 1 subunit (Jeruzalmi, O'Donnell, and Kuriyan, 2001; Kazmirski, Podobnik, et al, 2004). The sites that bind ATP do so with equal affinity and there appears to be little or no cooperativity in ATP binding. The conformation of γ complex with two ATP-bound is likely to represent a stable, but inactive state of the clamp loader. Modeling the kinetics of ATP hydrolysis in the absence of the clamp suggest that the clamp loader binds and hydrolyzes three molecules of ATP (Williams, Snyder, et al, 2004).

On binding ATP, the γ complex undergoes conformational changes that increase the affinity of the clamp loader for the clamp and DNA (Turner, Hingorani, et al, 1999) (Hingorani, and O'Donnell, 1998; Naktinis, Onrust, et al, 1995; Turner, Hingorani, et al, 1999) (Bertram, Bloom, et al, 1998; Hingorani, and O'Donnell, 1998). Within the γ complex, δ is blocked from interaction with β by δ' . However, ATP binding introduces a conformational change in the γ complex, exposing δ for interaction with β and subsequent ring opening (Naktinis, Onrust, et al, 1995). The δ subunit opens the β clamp using the energy from protein-protein interactions and does not need ATP. The binding of δ subunit relaxes the spring tension of α helix at the interface with the β clamp (Figure 1.6 b), thus disrupting the structure of dimer interface (Jeruzalmi, Yurieva, et al, 2001; Turner, Hingorani, et al, 1999).

ATP binding is sufficient for the formation of a ternary clamp loader–clamp- DNA complex; nevertheless, ATP hydrolysis is required for releasing the clamp onto the DNA (Bertram, Bloom, et al, 1998; Bloom, 2006; Turner, Hingorani, et al, 1999). Template DNA triggers three ATP sites in γ complex to hydrolyze ATP at the same global stage of reaction, releasing the clamp on DNA (Bertram, Bloom, et al, 2000). Although the β clamp does not induce the γ complex to hydrolyze ATP in the absence of DNA, the β clamp does increase the overall rate of hydrolysis in the presence of DNA. In the absence of the β clamp, the γ complex hydrolyzes two molecules of ATP more rapidly than the third one. In the presence of the β clamp, all three molecules of ATP hydrolyze at the same rapid rate as the two molecules of ATP hydrolyzed in the absence of the β clamp (Williams, Snyder, et al, 2004). The results from studies of mutation of arginine finger of the clamp loader suggest that ATP binding at one subset, γ_2 and γ_3 , is largely responsible for increasing the affinity of the complex for DNA, and ATP binding at γ_1 is largely responsible for increasing the affinity of complex for β clamp (Bloom, 2006; Snyder, Williams, et al, 2004).

The recently published crystal structure of *E.coli* γ complex with primer template DNA provides insight into clamp loader and DNA interactions (Figure 1.6, c). In the crystal structure, the AAA+ modules of the γ and δ' subunits form a right-handed spiral around the double-stranded portion of the DNA with a uniform rise and rotation around a common helical axis (Figure 1.6, c). The symmetry in the quaternary arrangement of the ATPase domains of the γ complex results in the configuration of each interfacial ATP-binding site being essentially the same. It has been suggested that the symmetrical

formation of catalytically competent ATP-binding sites facilitates release the clamp loader upon recognition of primer-template DNA. The positive ends of the helix dipoles of helices $\alpha 4$ and $\alpha 5$ within each subunit are positioned close to negatively charged phosphate groups, and the tips of these two helices are bisected by the backbone of the template strand. Domain I of the δ subunit is disengaged from DNA. Domain III of the δ subunit positions the side chain containing Tyr 316 so that it stacks on the nucleotide base at the 3' end of the primer strand, resulting in termination of the primer strand and a sharp bend in the template strand as it exits the clamp loader chamber. The collar domains encircle a tunnel that leads into the site within the central chamber where the last nucleotide of 3' primer strand is bound. The blocking function of the Tyr 316 side chain is reminiscent of the role of an aromatic side chain in the UvrD helicase that serves as a "separation pin" by splitting the path of DNA (Lee, and Yang, 2006). The single-stranded 5' overhang of the template strand exits the central chamber of the clamp loader and binds to the exterior surface of the domain III of δ subunit (Simonetta, Kazmirski, et al, 2009).

T4 bacteriophage encodes its own replication machinery, including a clamp (gp45 trimer) and clamp loader (gp44/gp62 complex). The T4 clamp loader is composed of five subunits; four gp44 subunits and one gp62 subunit. The gp44 subunit is an AAA+ protein and contains both SRC and P-loop motifs, which bind and hydrolyze ATP. The gp62 subunit shares no significant sequence homology with AAA+ proteins. The ATPase activity of a tetramer of gp44 subunits is not responsive to the addition of gp45 clamp in the absence of the gp62 subunit, whereas the clamp does stimulate the ATPase activity of

the complete gp44/62 loader complex. This result suggests that the gp62 subunit may be analogous to the δ subunit of *E.coli* and this serves as the wrench of the T4 clamp loader complex (Rush, Lin, et al, 1989).

The T4 gp44/gp62 clamp loader initially binds four ATPs with equal affinity (Young, Weitzel, and von Hippel, 1996). On binding gp45, two of four molecules are hydrolyzed to open the clamp wide enough so that it could be loaded on dsDNA. The remaining two molecules of ATP are hydrolyzed to close the clamp on the DNA. The clamp loader gp44/gp62 plays an additional chaperone-like role in the formation of the DNA polymerase holoenzyme. Gp44/gp62 mediates the interaction between the gp45-DNA complex and gp43 (T4 bacteriophage DNA polymerase). The transient formation of a multiple proteins complex of gp45·gp44/gp62·DNA·gp43 is observed before release of gp44/gp62 (Sexton, Kaboord, et al, 1998; Trakselis, Berdis, and Benkovic, 2003).

Gp45 is partially open in solution with a separation distance of 42 Å at one subunit interface and 17 Å at the other two interfaces (Alley, Shier, et al, 1999). Moreover, the T4 clamp is much less stable than γ complex with a K_d of 250 nM for the dimer compared to a K_d of < 60 pM for the β clamp (Yao, Turner, et al, 1996). It is intriguing that the partially open gp45 clamp requires the hydrolysis of ATP by the clamp loader for further opening to facilitate its loading onto DNA. Another study suggests activation of the clamp loading process requires only ATP binding, and that the hydrolysis of a single ATP was required only to release the activated clamp from the clamp loader and

complete its loading onto the primer-template DNA (Pietroni, Young, et al, 2001; Pietroni, and von Hippel, 2008).

The result that gp44/gp62 forms relatively stable complex with DNA in the presence of ATP is remarkably contrasted to the results from *E.coli* γ complex and eukaryotic RFC , in which ATP binding and hydrolysis destabilizes the clamp loader-DNA complex. In addition, this gp44/gp62-DNA binary complex is active in loading gp45 onto DNA. These results suggest that gp44/gp62 binds the clamp after binding DNA, whereas, it's generally believed that for *E.coli* and eukaryotic RFC that the clamp must bind the clamp loader before binding DNA. The binding of ATP to gp44/gp62 induces a conformational change in gp44/gp62 that configures the clamp loader for effective DNA binding. A gp44/gp62-DNA complex is formed with the hydrolysis of one equiv of ATP to interact with the opened form of gp45 on primer-template DNA and orient the clamp properly (Zhuang, Berdis, and Benkovic, 2006).

RFC Structure

RFC was shown to be essential for simian virus40 (SV40) DNA replication *in vitro* (Tsurimoto, Stillman, 1989). It is involved in polymerase switching from Pol α to Pol δ during initiation of leading strand DNA replication at the SV40 origin and for the synthesis of Okazaki fragments during lagging strand DNA synthesis. Clamp loaders have to rapidly recognize primed sites on template DNA and assemble clamps on them for efficient initiation.

The eukaryotic clamp PCNA is a trimer of three identical subunits arranged head to tail to generate a ring with a large central cavity for encircling DNA (Gulbis, Kelman, et al, 1996; Krishna, Kong, et al, 1994). Both DNA polymerase and clamp loader bind to the C-terminal face of PCNA in a hydrophobic pocket between two domains (Figure 1.7 f). Studies in the *E. coli* system have shown that clamp loader and DNA polymerases compete for binding the clamp, and therefore the clamp loader must eject from the clamp for DNA polymerase to use it (Naktinis, Turner, and O'Donnell, 1996). Likewise, pol δ also competes for binding to PCNA in eukaryotes (Mossi, Jonsson, et al, 1997; Oku, Ikeda, et al, 1998). Therefore after PCNA is linked to DNA, RFC is thought to eject from the clamp to allow the polymerase access to the clamp (Podust, Tiwari, et al, 1998).

Yeast RFC consists of a large subunit RFC A (RFC1) with a molecular mass of 95 kDa and four smaller subunits RFC B (RFC4), RFC C (RFC3), RFC D (RFC 2), RFC E (RFC5) of 36-40 kDa (Figure ,1.7 a). The five subunits each contain three domains (domains I-III), and the primary sequence of these domains is homologous among the five subunits; only the RFCA (RFC1) subunit has additional N- and C-terminal regions (Bowman, O'Donnell, and Kuriyan, 2004; Cullmann, Fien, et al, 1995). The C termini of all five RFC subunits consisting of domain III, pack together to form a stable cylindrical structure, referred to as the “collar” (Figure 1.7, d), and are required for RFC complex formation. The N –terminal AAA+ ATPase modules, composed of domains I and II, are assembled into a right-handed spiral in the order of RFC1-4-3-2-5 or RFC A-B-C-D-E (Bowman, O'Donnell, and Kuriyan, 2004; Yao, Coryell, et al, 2003). Peptide sequence alignment of all RFC subunits from human RFC and yeast RFC, compared with their

prokaryotic functional homologs, gp44 from phage T4 and γ subunit from *E. coli*, shows an overall sequence similarity (Figure 1.7, c). Stillman's group named the similar regions RFC box I-VIII, numbered from N-terminal to C-terminal. RFC box I is presented only in large RFC subunits and shows similarity to a region in prokaryotic DNA ligases and in procyclic acidic repetitive protein (PARPs) from eukaryotes. The RFC large subunit has DNA binding activity by itself and the DNA-binding activity was mapped to a region containing the ligase homology domain (Tsurimoto, and Stillman, 1991). In fact, the homology domain increases the nonspecific DNA binding activity of RFC and deletion of N-terminal domain containing box I increases RFC clamp loading activity (Gomes, Gary, and Burgers, 2000; Uhlmann, Cai, et al, 1997). RFC box II, III, IV and VI are conserved as ATP/GTP binding regions. The most conserved region is the phosphate-binding loop (P-loop) within box III. The "arginine finger", Ser-Arg-Cys (SRC) motif is conserved within small RFC subunits, but only Cys is present in the large RFC subunit (RFC_A). RFC box V bears similarity to the DEAD-Box proteins, a family of putative RNA helicases which also have P-loops and are ATPases. In addition to these functionally related proteins, a search of the sequence databases revealed sequence similarity to a predicted protein sequence from *CHL 12* gene of *S.cerevisia*. There is significant similarity between the *CHL 12* gene product and the RFC subunits, ranging from 20 to 25% amino acid identity and up to 50% similarity. *CHL 12* was isolated as a chromosome loss mutation and is identical to *CTF 18*. Mutants in *CHL 12* are incapable of stable maintenance of circular and linear artificial chromosomes (Cullmann, Fien, et al, 1995).

Recent crystal structure results revealed a stabilized RFC complex (Figure 1.7, d) with a closed PCNA clamp in the presence of ATP γ S (Bowman, O'Donnell, and Kuriyan, 2004). Because ATP hydrolysis weakens the interaction between clamp and clamp loader, they replaced the arginine finger residues that are present in the highly conserved Ser-Arg-Cys (SRC) motifs of four RFC subunits (RFC B, C, D, and E) with glutamine, which resulted in a significant reduction in ATPase activity. In the structure of RFC: PCNA complex, the five AAA+ modules of RFC assemble into a right-handed spiral, leaving a gap between RFC-A and RFC-E. This gap serves the purpose of allowing DNA to enter the central chamber in the clamp loader and thereby positions DNA into the open clamp, which is docked underneath the clamp loader. The right-handed spiral arrangement of the five AAA+ domains of RFC displays roughly the same pitch as that of double-stranded B-form DNA, which suggests a mechanism of DNA recognition. Only three of the RFC subunits (RFC-A, RFC-B and RFC-C) make contact with closed PCNA clamp. RFC-A and RFC-C interact with two of PCNA conserved hydrophobic grooves. RFC-B which is located between RFC-A and RFC-C, makes limited and primarily polar interactions with the intersubunit region of the clamp. All three RFC subunits interact with PCNA primarily through the C-terminal end of the clamp-interaction helix (α 4).

The PCNA ring is closed in this crystal structure, yet ATP γ S enables the RFC: PCNA complex to bind DNA and therefore presumably promotes PCNA opening. An EM structure of RFC-PCNA-primer DNA ternary complex revealed an open spiral form of the RFC: PCNA: ATP γ S complex and it is likely all the subunits of RFC have contacts with PCNA (Miyata, Suzuki, et al, 2005). Molecular dynamics simulations suggest that

the PCNA ring could open out of plane and have a tendency to form a right handed spiral (Kazmirski, Zhao, et al, 2005). A fluorescence resonance energy transfer study indicates that PCNA is held open by RFC in response to bound nucleotide (Figure 1.7d) (Zhuang, Yoder, et al, 2006). Therefore this closed conformation of PCNA may be due to the specific point mutation that prevents nucleotide hydrolysis during crystal growth.

In this crystal structure, only two of three PCNA hydrophobic pockets interact with RFC subunits. RFC-D and RFC-E are suspended above PCNA with RFC-E positioned above the third hydrophobic pocket of PCNA. However, a surface plasmon study demonstrated that RFC-E and RFC-D both interact with PCNA, and a PCNA unloading experiment suggests that the RFC-(D,E) subcomplex is primarily responsible for the PCNA-clamp-opening function of RFC (Yao, Johnson, et al, 2006). Therefore, the crystal structure of RFC: PCNA (closed): ATP γ S is suspected to represent the intermediate conformation after RFC loads PCNA onto DNA. Ring closure is presumed to lead to loss of RFC affinity for PCNA and dissociation of RFC from the PCNA-DNA complex.

ATP utilization by yeast RFC

RFC belongs to the AAA+ ATPase family. This family includes a wide variety of factors that couple ATP binding and hydrolysis to remodel protein and substrates (Erzberger, and Berger, 2006). The ATP sites of AAA+ multimers are located at subunit interfaces, and the adjacent subunit contributes a catalytic Arg residue which is part of Ser-Arg-Cys (SRC) motif to facilitate ATP hydrolysis (Figure 1.7 b). This arginine side

chain is referred to as an “arginine finger” by analogy to an arginine residue inserted at the active site of Ras protein by the GTPase-activating proteins. Interactions between AAA+ modules of RFC subunits create four interfacial ATP sites that are competent for hydrolysis (Bowman, O'Donnell, and Kuriyan, 2004; Yao, Coryell, et al, 2003). The four ATP sites are referred to as ATP sites A-D, with site A at the RFC-A/B interface and site D at the RFC-D/E interface. The RFC-B arginine finger functions in ATP site A, and the arginine finger of RFC-C functions in ATP site B and so on. RFC-E (RFC5) also binds nucleotide, but the nucleotide-binding pocket of RFC-E is not competent for hydrolysis nor is it needed for RFC activity (Bowman, O'Donnell, and Kuriyan, 2004; Podust, Tiwari, et al, 1998).

The process of RFC loading PCNA is an ATP driven pathway. RFC can load and release PCNA onto a primer-template site at a rate of $1-2 \text{ s}^{-1}$, which is compatible with the estimated rate of Okazaki fragment synthesis in vivo ($0.5-1 \text{ s}^{-1}$). There are four key steps of clamp loading, which are manipulated by ATP binding and hydrolysis, PCNA and primer DNA template binding (Figure 1.7, e).

- I. ATP binding induces activation of RFC such that it can bind PCNA with high affinity. RFC has five nucleotide binding sites and shows weak ATPase activity (0.024 s^{-1}). Full length RFC alone can bind three molecules of $\text{ATP}\gamma\text{S}$, and when PCNA or primer DNA or both are present in the reaction, full length RFC can bind five molecules of $\text{ATP}\gamma\text{S}$. ATP binding initiates a slow conformational

change in RFC, enabling it to bind and open PCNA, and subsequently bind DNA (Chen, Levin, et al, 2009).

- II. PCNA opening locks RFC into an active state, and the resulting RFC: ATP: PCNA (open) intermediate is ready for DNA entry into the clamp (Chen, Levin, et al, 2009; Gomes, Schmidt, and Burgers, 2001; Zhuang, Yoder, et al, 2006). Activated RFC has high affinity for both PCNA and primer-template DNA. However, RFC forms a stable complex with primer-template DNA only in the presence of ATP γ S. Interaction with primer-template DNA leads to rapid ATP hydrolysis and disassembly of RFC from the DNA ($\sim 25\text{s}^{-1}$). However, RFC and PCNA can form a relatively stable complex which slowly disassembles ($\sim 1.7\text{s}^{-1}$). Inclusion of PCNA results in a faster rate of RFC activation (1.5s^{-1} - 4.6s^{-1}), and PCNA stimulates RFC-DNA-dependent ATPase activity downstream of DNA binding ($\sim 11\text{s}^{-1}$ to $\sim 53\text{s}^{-1}$). The conformation of PCNA docked underneath RFC is predicted to open into a spiral conformation complementary to that adopted by the RFC subunits. Conversely, the open conformation of PCNA locks RFC in an activated state (Chen, Levin, et al, 2009; Kazmirski, Zhao, et al, 2005; Miyata, Suzuki, et al, 2005; Zhuang, Yoder, et al, 2006).
- III. The RFC: ATP: PCNA (open) complex binds specially to primer-template DNA. It is RFC, not PCNA that provides the grip and holds RFC: ATP: PCNA (open) complex to DNA. RFC has at least two types of DNA binding domains. The ligase homolog domain at N-terminal of RFC-A subunit is ATP independent and

confers RFC nonspecific DNA binding properties (Gomes, Gary, and Burgers, 2000; Schmidt, Gomes, and Burgers, 2001; Uhlmann, Cai, et al, 1997). Deletion of the ligase homolog domain increases the clamp loading activity. The DNA binding domain at the C-terminal site of RFC-A and other RFC subunits may function as coordinate units that require ATP for binding with DNA. The RFC A-D subunits contain several conserved positively charged and polar side chains that are oriented toward the central chamber and may function to bind DNA (Bowman, O'Donnell, and Kuriyan, 2004; Yao, Johnson, et al, 2006). Mutation of conserved basic residues in RFC subunit B, C, D, or all three subunits together dramatically decreases affinity to DNA (Johnson, Yao, et al, 2006; Yao, Johnson, et al, 2006). In addition, mutation analysis in yeast RFC has shown that ATP binding to site C and D of RFC is essential for DNA binding and the argining finger of RFC-D is likely needed for the ATP in site C to promote the conformation change in RFC required for DNA binding. RFC with an argining finger mutation of RFC-D subunit loses the affinity to primer-template DNA (Johnson, Yao, et al, 2006; Schmidt, Gomes, and Burgers, 2001)

- IV. ATP hydrolysis triggered by recognition of a DNA primer-template junction results in the dissociation of RFC from the clamp and PCNA closure (Chen, Levin, et al, 2009; Gomes, and Burgers, 2001; Yao, Johnson, et al, 2006). RFC alone shows a slow ATPase activity ($\sim 0.024\text{s}^{-1}$). PCNA has a small but reproducible effect, increasing the rate to $\sim 0.046\text{s}^{-1}$. In contrast, primer template DNA triggers rapid ATP hydrolysis of RFC $\sim 54\text{s}^{-1}$ in the presence of PCNA, and

$\sim 11\text{s}^{-1}$ in the absence of PCNA. RFC hydrolyzes three molecules of ATP in the presence of PCNA, and two molecules of ATP in the absence of PCNA. Studies of the arginine finger mutation suggest that DNA triggers ATP hydrolysis in site C which in turn, may drive hydrolysis in additional ATP sites. ATP hydrolysis in site D is specially triggered by PCNA, and it has been proposed that this site leads to closure of PCNA around DNA. It is not clear how the clamp loader is released from DNA for another cycle of clamp loading. (Chen, Levin, et al, 2009; Johnson, Yao, et al, 2006; Schmidt, Gomes, and Burgers, 2001).

Alternative clamp loaders

There are three additional clamp loaders or putative clamp loaders and one clamp identified in eukaryotic cells. These three clamp loaders consist of four RFC small subunits in common and one pathway-specific large subunit instead of RFC-A (Majka, and Burgers, 2004).

Rad24-RFC clamp loader functions in DNA damage check-point pathway, loading Rad17-Mec3-Ddc1 PCNA-like clamp onto DNA (Kondo, Matsumoto, and Sugimoto, 1999; Majka, and Burgers, 2003). Ctf18-RFC clamp loader is involved in establishment of sister chromatid cohesion (Hanna, Kroll, et al, 2001; Mayer, Gygi, et al, 2001). Elg1-RFC is essential for maintenance of chromosome stability. Yeast Elg1 mutants show elevated levels of recombination, chromosome loss and gross chromosome rearrangement (Bellaoui, Chang, et al, 2003; Kanellis, Agyei, and Durocher, 2003; Smith, Hwang, et al, 2004). Recently, Elg1-RFC has been found to be involved in sister

chromatin cohesion too. However, Elg1-RFC cohesion activity is distinctive from that of Ctf18-RFC (Maradeo, and Skibbens, 2009; Parnas, Zipin-Roitman, et al, 2009).

Although it is well established from genetic studies that Rad24, Ctf18, and Elg1 RFC-like clamps function in separate pathways, there is also cross-talk between these pathways. All three RFC-like clamp loaders interact with PCNA. Ctf18-RFC can load and unload PCNA onto a primer-template DNA substrate (Bermudez, Maniwa, et al, 2003). A physical interaction between Elg and PCNA suggest Elg1-RFC may also open and thus load PCNA onto DNA (Kanellis, Agyei, and Durocher, 2003). Although Rad24-RFC could not load PCNA, it could open PCNA and unload it from DNA (Yao, Johnson, et al, 2006). PCNA is crucially involved in establishment of sister chromatid cohesion in S phase (Moldovan, Pfander, and Jentsch, 2006). Sister chromatid cohesion is essential for the equal segregation of replicated chromosomes to the daughter cell. Appropriate cohesion involves identifying the products of chromosome replication as sisters, depositing cohesions onto each sister and then modifying those cohesions to form structural bridges that tether together sister chromatids until anaphase onset. Ctf7/ Eco1 is an acetyltransferase that activates cohesion during S-phase and is essential for sister chromatid pairing (Skibbens, Corson, et al, 1999). Ctf7/Eco1 is found associated with all four clamp loading complexes, RFC-A, Rad24, Ctf18 and Elg1 (Kenna, and Skibbens, 2003; Maradeo, and Skibbens, 2009). In addition, RFCs often compensate for one another during DNA replication checkpoint activation and repair (Bermudez, Maniwa, et al, 2003; Naiki, Kondo, et al, 2001). It is likely that RFC and any other alternative clamp loader will similarly participate in cohesion.

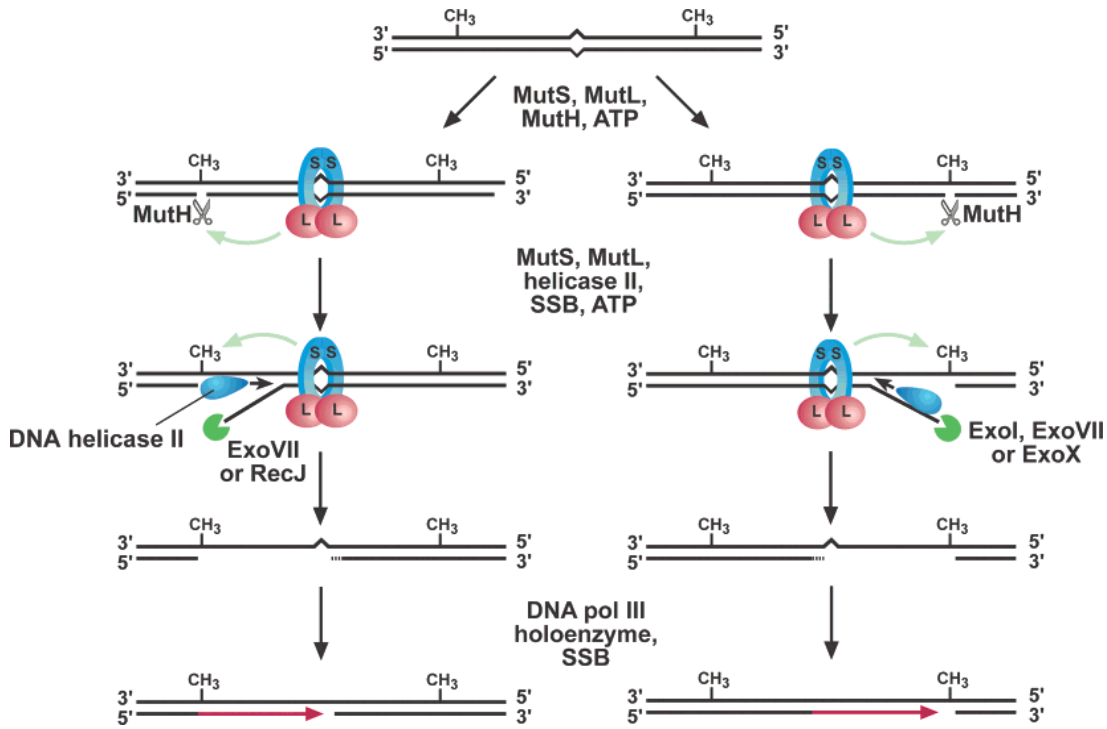


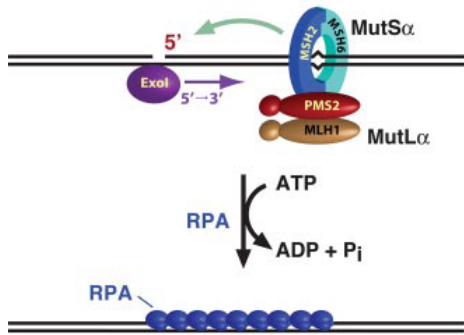
Figure 1.1 Mechanism of *E. coli* methyl-directed mismatch repair (reproduced from G. Li, *Cell Research*, 2008). *E. coli* mismatch repair is initiated when MutS specifically recognizes mismatched DNA. MutS interaction with MutL activates the latent endonuclease activity of MutH in an ATP-dependent manner, which cleaves the newly synthesized daughter strand at hemimethylated GATC sites. The resulting nick, which can be either 3' or 5' to the mismatch, is the entry point for MutL-dependent loading of DNA helicase II and binding of single-strand DNA binding protein. Depending on the location of the nick to the mismatch, the generated single strand is digested by a 3' or 5' exonuclease. The excision removes the error and allows highly accurate DNA polymerase III to correctly resynthesize the strand. DNA ligase seals the nick to complete MMR.

Table 1.1 Identity and function of *E. coli* and eukaryotic proteins involved in MMR

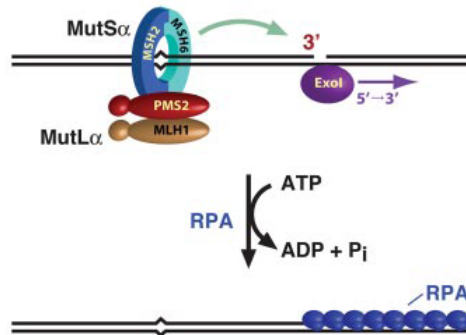
<i>E. coli</i>	Function	Homologs	Function
MutS	Binds mismatch	Msh2·Msh6 (MutS α)	Repairs single base-base and 1-2 base IDL mismatches
		Msh2·Msh3 (MutS β)	Repairs mismatches of 2 to ~10 nucleotides and some small single IDLs
MutL	Matchmaker that coordinates multisteps in MMR	MLH1·PMS2 (yPms1) (MutL α)	Matchmaker, endonuclease, termination of mismatch-provoked excision
		MLH1·MLH2 (hPMS1) (MutL β)	Unknown function in MMR
		MLH1·MLH3(MutL γ)	Supports modest levels of base-base and single IDL
MutH	Strand discrimination, endonuclease	None	
UvrD	DNA helicase	None	
SSB	Participates in excision and DNA synthesis	RPA	ssDNA binding/protection, stimulates mismatch excision, termination of DNA excision, promoting DNA synthesis
β -clam	May recruit MutS to mismatch/ replication foci, enhances processivity of pol III	PCNA	Recruits MMR proteins to mismatch, participates in excision, activation of MutL α endonuclease, DNA resynthesis
γ -complex	β -clamp loader	RFC	PCNA clamp loader, modulates excision polarity, activation of MutL α endonuclease
ExoI or ExoX	3' to 5' excision of ssDNA	ExoI	5' to 3' excision of ssDNA
RecJ or ExoVII	5' to 3' excision of ssDNA		
DNA pol III	DNA resynthesis	DNA pol δ and pol ϵ	DNA resynthesis
DNA ligase	Nick ligation	DNA ligase	Nick ligation

a

5'-heteroduplex:

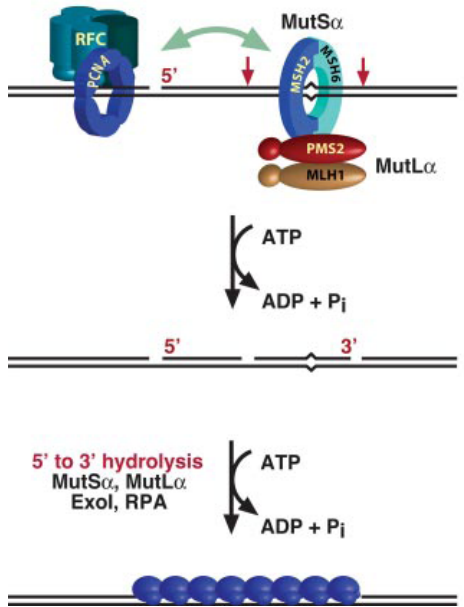


3'-heteroduplex:



b

5'-directed



3'-directed

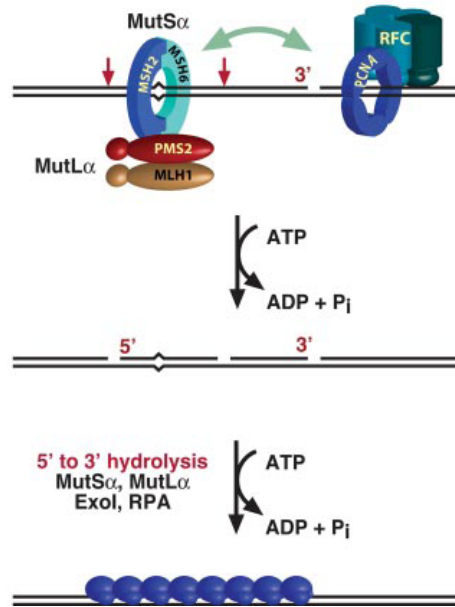


Figure 1.2 Reconstituted eukaryotic mismatch-provoked excision systems

(reproduced from P.Mordrich, *JBC*,2006). a. 5'-heteroduplex excision depending on MutS α , MutL α , RPA and ExoI. MutS α activates ExoI 5' to 3' hydrolysis on 5' heteroduplex and renders ExoI highly processive about ~2000 nucleotides, an effect attributed to formation of MutS α ·ExoI complex. RPA reduces the high processivity of MutS α ·ExoI complex to ~250 nucleotides. MutL α acts in concert with MutS α to suppress ExoI hydrolysis on DNA that lacks mismatch. MutS α also activates ExoI hydrolysis on 3'-heteroduplex, but in this case hydrolysis proceeds in the wrong direction. b. Reconstituted bidirectional eukaryotic mismatch-provoked excision system is comprised of MutS α , MutL α , RPA, ExoI, PCNA and RFC. In this case, PCNA and RFC activate MutL α latent endonuclease. MutL α incises 5' or 3'- heteroduplex strand in an ATP-dependent manner, generating a strand break as the entry site of MutS α activated ExoI, which removes the error by the 5' to 3' hydrolysis. In this model, the original nick serves as a strand discrimination signal, but not as site of excision initiation.

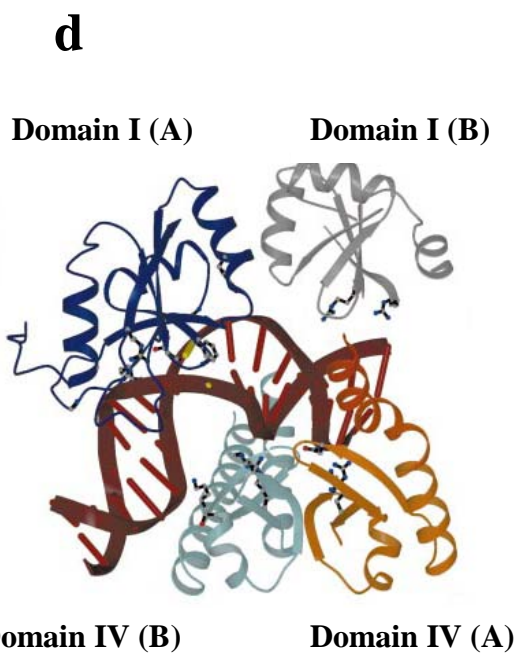
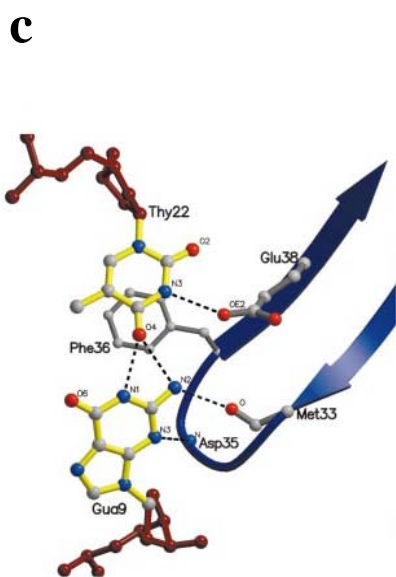
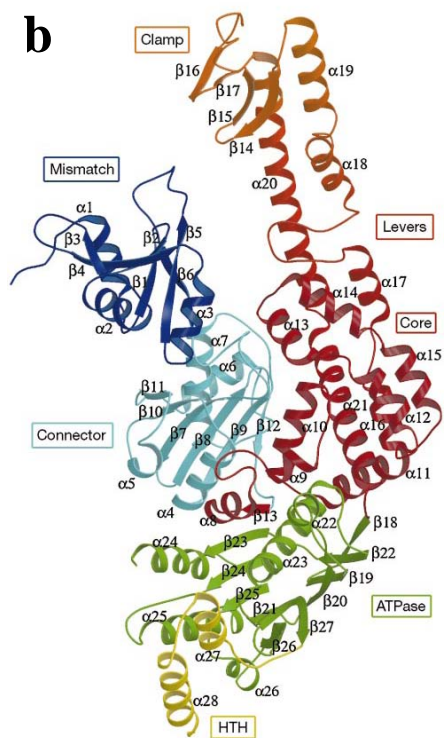


Fig 1.3 Crystal structure of E. coli MutS binding to G·T mismatch (reproduce from M. H. Lamers, et. al, *Nature*, 2000). a. Overview of MutS-DNA complex. The asymmetric unit contains a MutS dimer with G·T mismatched DNA running through the lower channel. DNA and ADP are colored red, the mismatch-binding monomer light green, and the second monomer blue. b. Mismatch-binding monomer colored by domain. The mismatch-binding domain is colored dark blue, the connector domain light blue, the core and levers from red to orange, the clamp domain orange, the ATPase domain green, and HTH (helix-turn-helix turn involved in dimer contacts) yellow. c. Mismatch specific binding motif, Phe-X-Glu. Phe 36 stakes with the mismatched thymine, which is rotated out into the minor groove by 3Å. Glu 38 forms hydrogen bond to the N3 of thymine 22 of the mismatch. d. Only one monomer has specific mismatch-binding contacts which are exclusive to its N-terminal mismatch-recognition domain. The equivalent domain in the second monomer has only loose DNA backbone contacts. The interaction between DNA and the mismatch binding domain expands the minor groove of DNA , kinking the DNA ~60° at G·T mismatch. Mismatch is colored yellow.

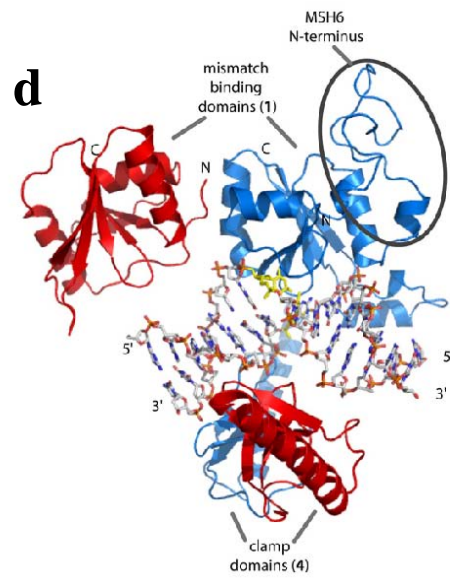
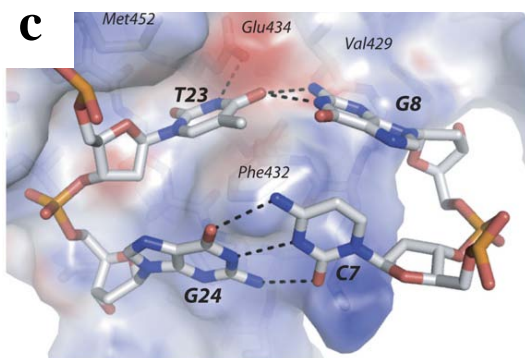
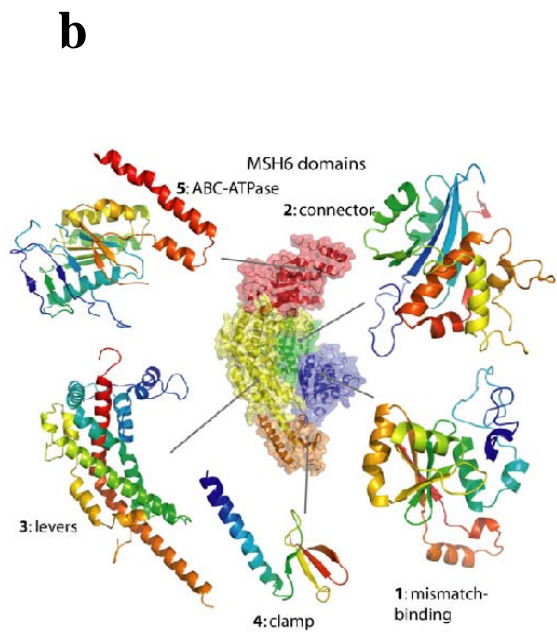
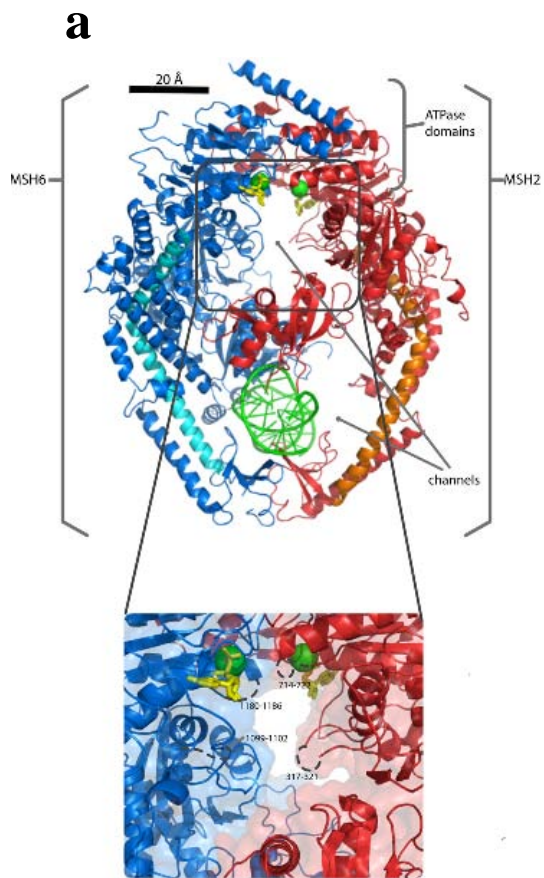
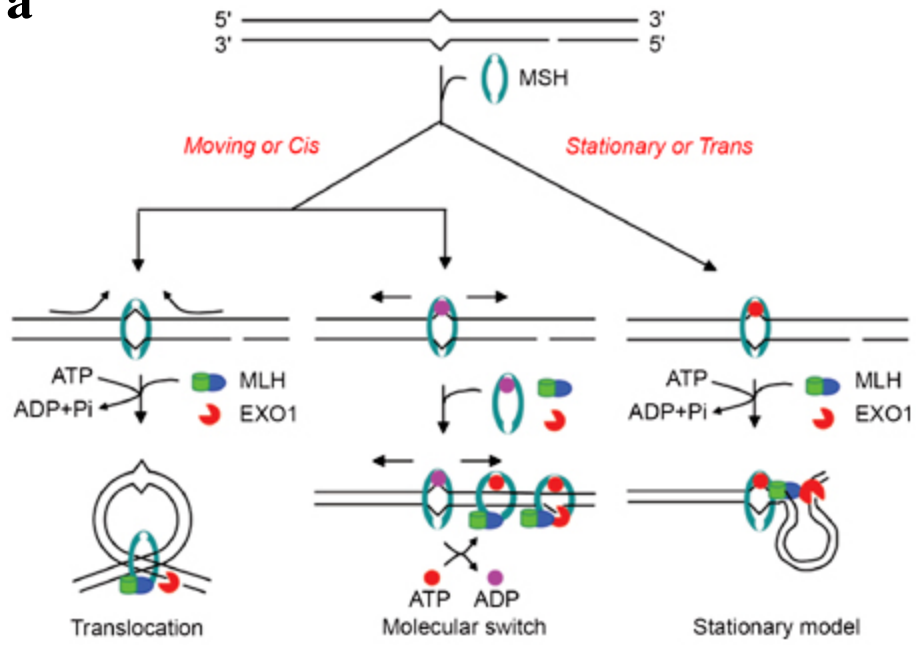


Figure 1.4 Crystal structure of Mut α binding with G·T mispair complex (reproduced from J.J. Warren, et.al, *Molecular Cell*, 2007). a. Ribbon diagram of the structure of MutS α -DNA complex. Msh6 is colored blue, Msh2 red, DNA green ribbon, and ADP green sphere. Long α helices connecting clamp and ATPase domains in Msh2 and Msh6 are colored orange and cyan respectively. The whole structure looks like the Greek letter “ Θ ”, pierced by two channels. The enlarged upper channel is shown below the structure. Disordered loops are shown as dashed lines with residue numbers. The DNA is bent by $\sim 45^\circ$, and bound in the lower channel of MutS α heteroduplex. b. Each protein can be divided into 5 domains, mismatch binding (blue), connector (yellow), levers (yellow), clamp (brown) and ABC-ATPase domain (red). c. Conserved mismatch specific binding motif, Phe-X-Glu motif. Glu434 of the conserved motif hydrogen bonds to the mispair thymine, which is sandwiched between Phe432 and Met459, and the backbone carbonyl of Val429 accepts a hydrogen bond from the mispaired guanine. d. Only the mismatch binding domain of Msh6 makes specific contacts with mismatch. The equivalent domain of Msh2 is rotated up away from DNA backbone and makes only one contact with DNA backbone. An additional ordered region at the N-terminal of Msh6 makes extended nonspecific binding with DNA.

a



b

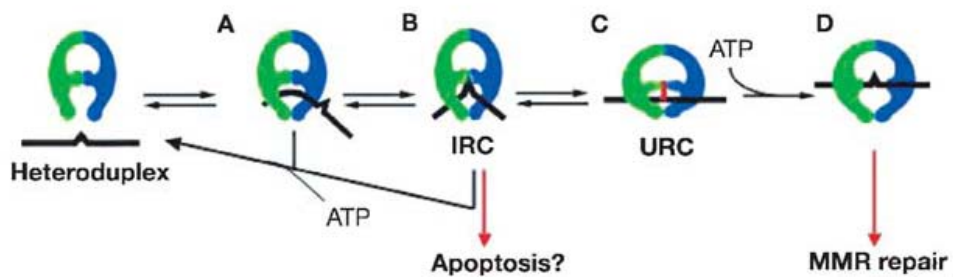
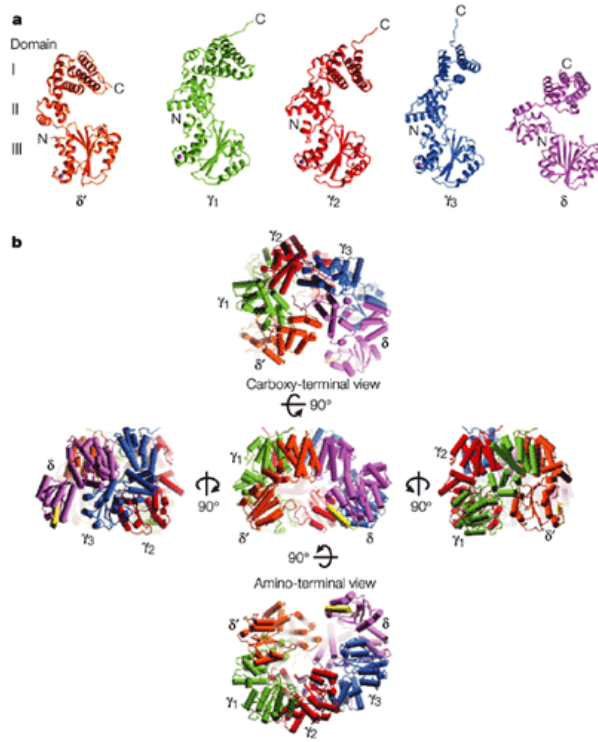


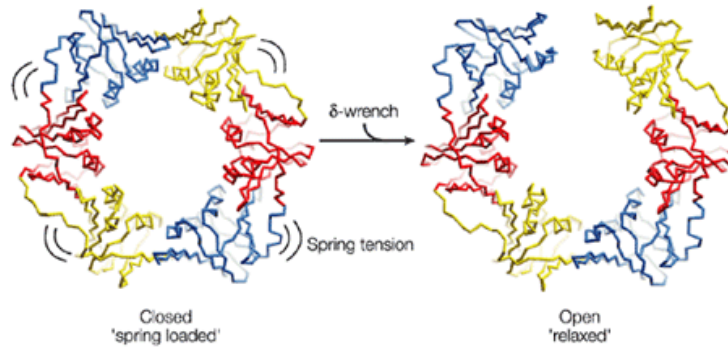
Figure 1.5. Models for signaling downstream MMR events following mismatch recognition and mismatch recognition mechanism of E. coli MutS. a. A schematic diagram for signaling between the mismatch and the strand discrimination signal is shown (reproduced from G. Li, *Cell Research*, 2008). Here, a 5' nick is the strand discrimination signal. Similar models apply for 3' nick-directed MMR. The "stationary" or "trans" model (right) emphasizes that MutS or its homologous (MSH) proteins remain bound at the mismatch. It is the protein-protein interactions that induce DNA bending or looping that brings the two distant sites together. The two DNA sites can cooperate in a "trans" configuration. In two "cis" or "moving" models, one called the "translocation" model (left) and the other called the "molecular switch" or "sliding clamp" model (middle), it is hypothesized that the MSH proteins bind to the mismatch and then move away from the site to search for the strand discrimination signal. The translocation model suggests that ATP hydrolysis drives bidirectional movement of the MSH proteins, resulting in the formation of an α -like loop. In the molecular switch model (center), binding of an MSH protein (in its ADP-bound state) to the mismatch triggers an ADP to ATP exchange that promotes bidirectional sliding of the protein away from the mismatch, thereby emptying the mismatch site for another incoming MSH protein. Mismatch excision begins when an MSH protein reaches the strand break. b. MutS binds to DNA nonspecifically and bends it in search of mismatch (A). Upon specific recognition of a mismatch, MutS undergoes a conformational change to an initial recognition complex (IRC) in which the DNA is kinked (B). MutS then undergoes further conformational

changes to the ultimate recognition complex (URC) in which the DNA is unbent (D), signaling the initiation of mismatch repair (D) (reproduced from H. Wang, *PNAS*, 2003)

a



b



c

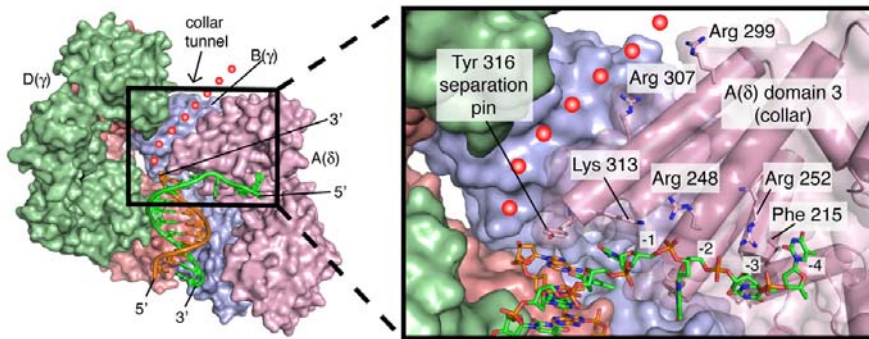
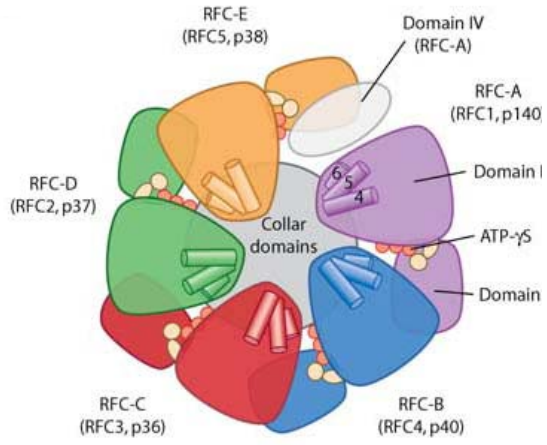
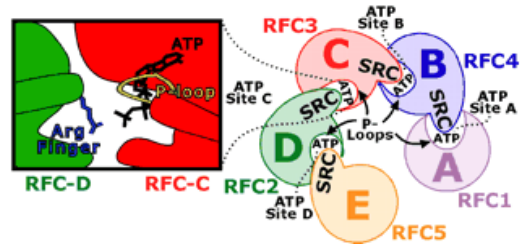
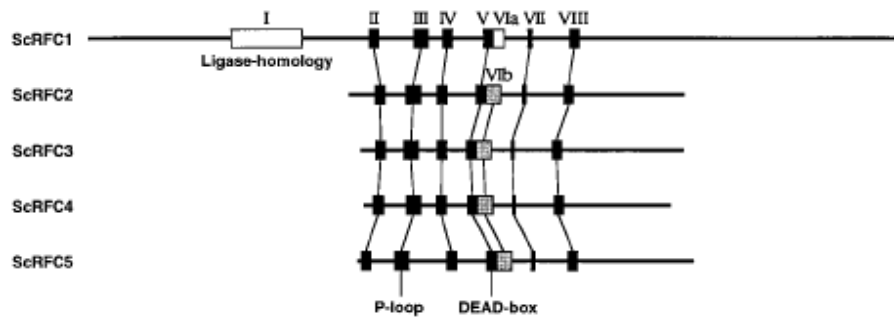
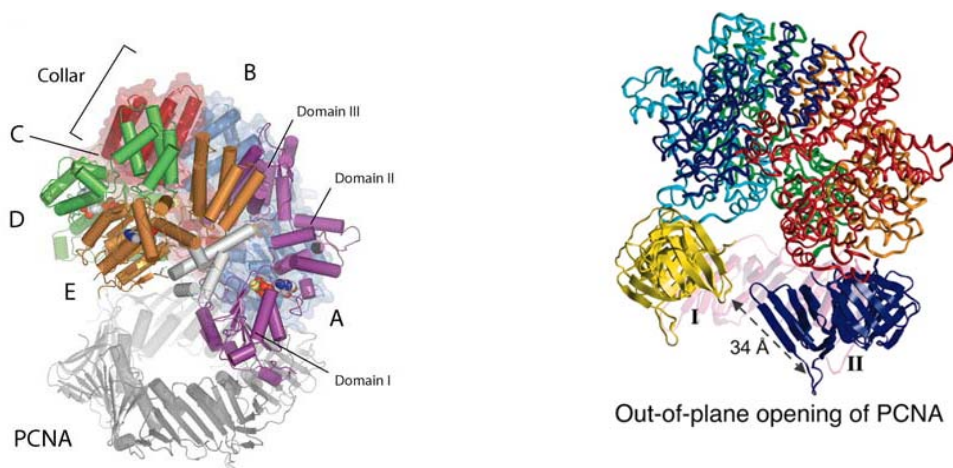
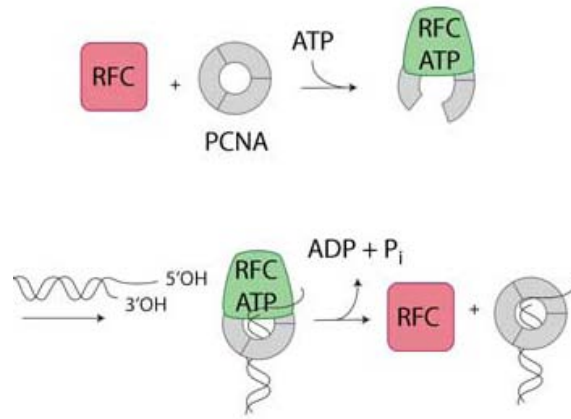


Figure 1.6 Crystal structure of E.coli γ -complex and schematic view of the mechanism of opening β -clamp. a. The individual subunits from the crystal structure of γ -complex (reproduced from M.J. Davey, J. Kuriyan, M. O'Donnell, *Nature Review*, 2002). The three domains (I, II and III) are indicated. The amino and carboxyl termini are marked (N and C respectively). The colour of each subunit corresponds to the colour of the subunit in the structure of γ -complex below. The crystal structure reveals that the five subunits are arranged in a circle, with δ and δ' bracketing the three γ subunits. b. The β clamp is composed of two crescent-shaped promoters (reproduced from M.J. Davey, J. Kuriyan, M. O'Donnell, *Nature Review*, 2002)., each consisting of three domains (each domain is indicated by a different color). In the dimer form (left) the promoters must adopt a bent shape to form the dimeric interfaces of the clamp, therefore placing the promoters under spring tension. When one interface of the clamp is disrupted by the δ wrench, the promoters relax (right), resulting in an opening at one interface of the clamp for DNA strand passage. c. . The structure of clamp loader with primer-template DNA is shown (refer to K. R. Simonetta, J. Kuriyan, M. O'Donnell, *Cell*, 2009)., with the δ' subunit removed to reveal a tunnel leading through the collar, indicated by red spheres. In the expanded view on the right, side chains presented by the collar domain of the δ subunit and that interact with DNA are shown. Two side chains that line the collar channel are also shown. The Tyr316 side chain stacks on the nucleotide base at the 3' end of the primer strand, resulting the termination of the primer strand and a sharp bend in the template strand as it exits the clamp loader chamber.

a**b****c****d**

e



f

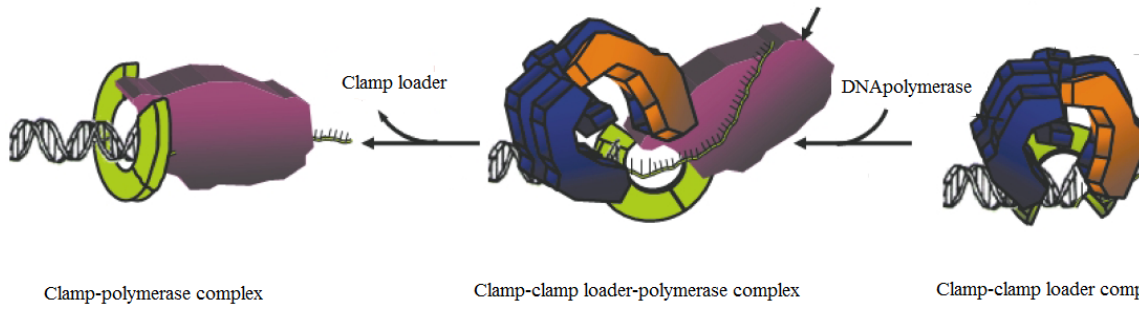
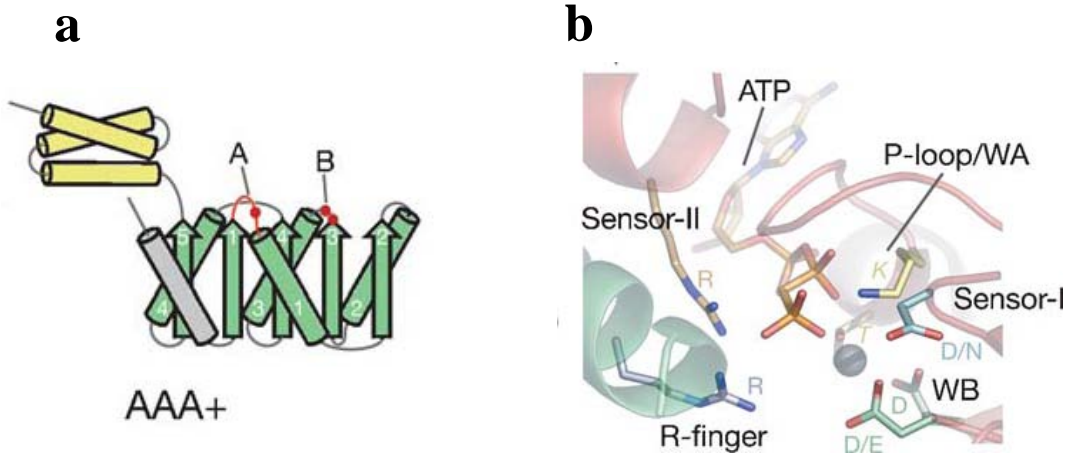


Figure 1.7 Crystal structure of RFC clamp loader complex and schematic of the clamp loading process. a. A schematic view of RFC complex viewed from the PCNA interacting face. The five subunits of the RFC complex are referred to as RFC-A, RFC-B, RFC-C, RFC-D, and RFC-E, respectively (reproduced from G.D. Bowman, M. O'Donnell, J. Kuriyan, *Nature*, 2004). b. Left: schematic diagram depicting ATPase site C from the crystal structure. Right: Schematic of the arrangement of ATP sites in the AAA+ modules of RFC heteropentamer. Each ATPase site is at a subunit interface. The neighboring subunit contains an arginine finger in a conserved SRC motif that interacts with the γ phosphate of the ATP bound to the adjacent subunit (reproduced from A. Jonson, J. Kuriyan, M. O'Donnell, *JBC*, 2006). c. The small subunits align to the middle part of the large subunit of RFC complex. There are eight conserved RFC boxes numbered consecutively from N-terminus to C-terminus. Box I is the DNA ligase homolog domain, and boxes II to VIII contain with them an ATP binding region (refer to G. Cullmann, B. Stillman, *Molecular and Cellular Biology*, 1995). d. A stereoview of RFC: PCNA complex. The N-terminal domain I and domain II form the AAA+ ATPase modules, assembling into a right-handed spiral. The C-terminal of each subunit packs together to form stable, cylindrical structure, referred as the 'collar'. Left is the complex of RFC-PCNA (close) (reproduced from G.D. Bowman, M. O'Donnell, J. Kuriyan, *Nature*, 2004); right is the model of RFC-PCNA (open) (Z. Zhaung, S. J. Benkovic, *PNAS*, 2006). e. The clamp loader cycle. ATP binding with RFC introduces RFC activation. Upon binding with PCNA, the activated RFC induces the conformational change in PCNA that opens one subunit interface. The RFC: PCNA complex

subsequently binds to primer-template DNA, and the recognition of the double strand/single strand junction stimulates ATP hydrolysis by clamp loader. This hydrolysis results in the dissociation of RFC from the clamp and DNA, leaving PCNA encircling DNA (reproduced from G.D. Bowman, M. O'Donnell, J. Kuriyan, *Nature*, 2004). f. A proposed structural representation of the polymerase competing with clamp loader at the same clamp face. A clamp-clamp loader-polymerase media complex may be formed during the process that polymerase takes place of clamp loader after it completes loading the clamp onto DNA. Clamp is colored green, clamp loader blue and orange, and polymerase purple. DNA is demonstrated as black helix (adapted from M. A. Trakselis, S.J. Bencovic, *JMB*, 2003).

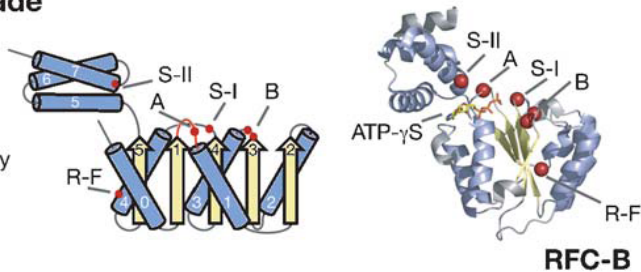


c

1 - Clamp loader clade

No unique features

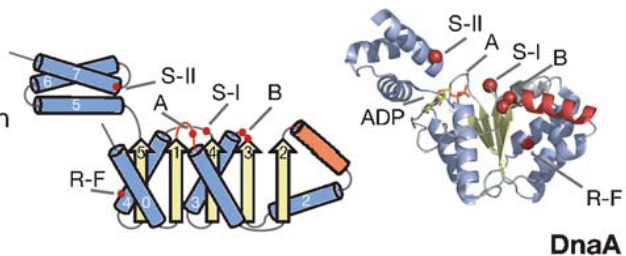
3 major families:
bacterial clamp loader family
RFC clamp loader family
WHIP proteins



2 - Initiator clade

Initiator specific helix
before $\alpha 2$ mediates
open filament formation

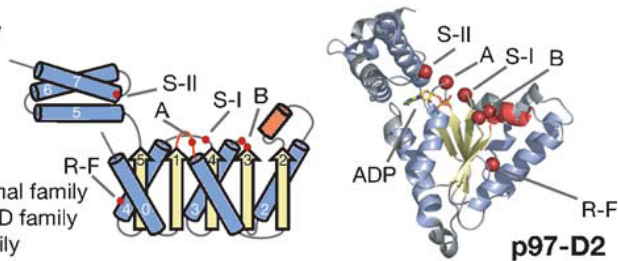
2 major families:
DnaA/DnaC/Hda family
Orc1/Cdc6 family



3 - Classic clade

Short substrate specificity
helix before $\alpha 2$ defines
translocation channel

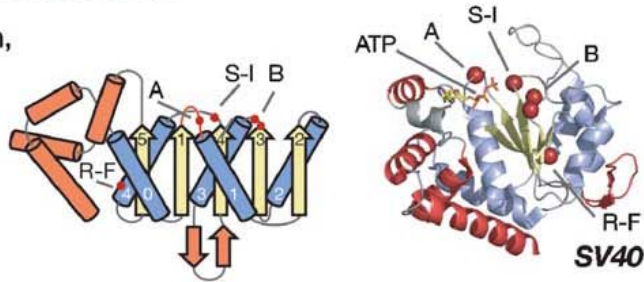
7 major families:
FtsH family Proteasomal family
Tip49 family ClpAB NTD family
Katanin family AFG1 family
Cdc48/NSF/Pex/Bcs family



4 - Superfamily III helicase clade

Lacks AAA+ lid domain,
unique helical bundle
composed of N- and
C-terminal elements.

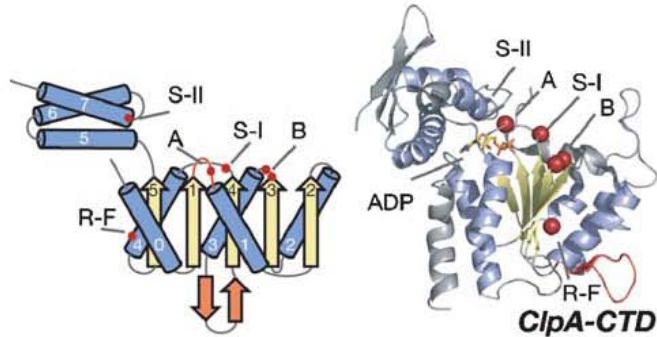
2 major families:
DNA virus family
RNA virus family



5 - HCLR clade

No features beyond
PS-I β -hairpin

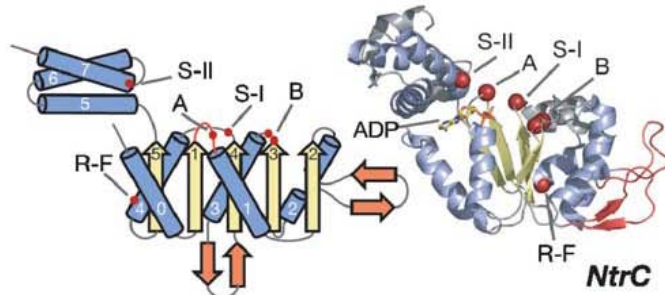
4 major families:
HslU/ClpX family
ClpAB-CTD family
Lon family
RuvB family



6 - H2-insert clade

β -hairpin insertion
in helix 2

2 major families:
NtrC family
McrB family



7 - PS-II insert clade

Helical insertion after
 $\alpha 5$ repositions the
Sensor-II motif

3 major families:
MCM family
MoxR family
Chelatase/YifB family
Dynein/Midasin family

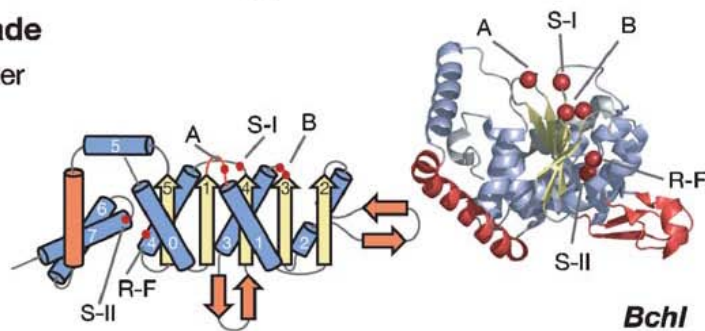


Figure 1.8 Schematic view of AAA+ ATPase family (Taken from J.P. Erzberger and J.M. Berger, *Annu. Rev. Biophys. Biomol. Struct.*, 2006). a. Topology diagram of AAA+ ATPase. The core ASCE fold is colored green, additional β strands colored grey and C-terminal helical bundle colored yellow. b. Detail of the active site of ATP-DnaA showing the position of nucleotide-interacting motifs and ATP. The coloring reflects subunit contribution. c. Basic AAA+ clades. The first three AAA+ clades show few structural changes relative to the basic ASCE fold. The last four AAA+ clades are Pre-sensor I insert superclade members share a common β -hairpin insertion but are also distinguished by additional clade-specific features. Basic AAA+ secondary structure elements (blue and yellow) as well as clade-specific structure feature (red) are depicted.

Reference List

- Acharya, S., Foster, P.L., Brooks, P., and Fishel, R. (2003). The coordinated functions of the *E. coli* MutS and MutL proteins in mismatch repair. *Mol. Cell* *12*, 233-246.
- Allen, D.J., Makhov, A., Grilley, M., Taylor, J., Thresher, R., Modrich, P., and Griffith, J.D. (1997). MutS mediates heteroduplex loop formation by a translocation mechanism. *EMBO J.* *16*, 4467-4476.
- Alley, S.C., Shier, V.K., Abel-Santos, E., Sexton, D.J., Soumillon, P., and Benkovic, S.J. (1999). Sliding clamp of the bacteriophage T4 polymerase has open and closed subunit interfaces in solution. *Biochemistry* *38*, 7696-7709.
- Antony, E., and Hingorani, M.M. (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*, 13115-13128.
- Antony, E., and Hingorani, M.M. (2003). Mismatch recognition-coupled stabilization of Msh2-Msh6 in an ATP-bound state at the initiation of DNA repair. *Biochemistry* *42*, 7682-7693.
- Ban, C., and Yang, W. (1998a). Crystal structure and ATPase activity of MutL: implications for DNA repair and mutagenesis. *Cell* *95*, 541-552.
- Ban, C., and Yang, W. (1998b). Structural basis for MutH activation in *E. coli* mismatch repair and relationship of MutH to restriction endonucleases. *EMBO J.* *17*, 1526-1534.
- Bellaoui, M., Chang, M., Ou, J., Xu, H., Boone, C., and Brown, G.W. (2003). Elg1 forms an alternative RFC complex important for DNA replication and genome integrity. *EMBO J.* *22*, 4304-4313.
- Bermudez, V.P., Maniwa, Y., Tappin, I., Ozato, K., Yokomori, K., and Hurwitz, J. (2003). The alternative Ctf18-Dcc1-Ctf8-replication factor C complex required for sister chromatid cohesion loads proliferating cell nuclear antigen onto DNA. *Proc. Natl. Acad. Sci. U. S. A.* *100*, 10237-10242.
- Bertram, J.G., Bloom, L.B., Hingorani, M.M., Beechem, J.M., O'Donnell, M., and Goodman, M.F. (2000). Molecular mechanism and energetics of clamp assembly in *Escherichia coli*. The role of ATP hydrolysis when gamma complex loads beta on DNA. *J. Biol. Chem.* *275*, 28413-28420.
- Bertram, J.G., Bloom, L.B., Turner, J., O'Donnell, M., Beechem, J.M., and Goodman, M.F. (1998). Pre-steady state analysis of the assembly of wild type and mutant circular

clamps of Escherichia coli DNA polymerase III onto DNA. *J. Biol. Chem.* 273, 24564-24574.

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

Bjornson, K.P., and Modrich, P. (2003). Differential and simultaneous adenosine di- and triphosphate binding by MutS. *J. Biol. Chem.* 278, 18557-18562.

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

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

Bloom, L.B. (2006). Dynamics of loading the Escherichia coli DNA polymerase processivity clamp. *Crit. Rev. Biochem. Mol. Biol.* 41, 179-208.

Bowman, G.D., O'Donnell, M., and Kuriyan, J. (2004). Structural analysis of a eukaryotic sliding DNA clamp-clamp loader complex. *Nature* 429, 724-730.

Cannavo, E., Marra, G., Sabates-Bellver, J., Menigatti, M., Lipkin, S.M., Fischer, F., Cejka, P., and Jiricny, J. (2005). Expression of the MutL homologue hMLH3 in human cells and its role in DNA mismatch repair. *Cancer Res.* 65, 10759-10766.

Chen, S., Levin, M.K., Sakato, M., Zhou, Y., and Hingorani, M.M. (2009). Mechanism of ATP-driven PCNA clamp loading by *S. cerevisiae* RFC. *J. Mol. Biol.* 388, 431-442.

Clark, A.B., Valle, F., Drotschmann, K., Gary, R.K., and Kunkel, T.A. (2000). Functional interaction of proliferating cell nuclear antigen with MSH2-MSH6 and MSH2-MSH3 complexes. *J. Biol. Chem.* 275, 36498-36501.

Cullmann, G., Fien, K., Kobayashi, R., and Stillman, B. (1995). Characterization of the five replication factor C genes of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 15, 4661-4671.

Cunningham, E.L., and Berger, J.M. (2005). Unraveling the early steps of prokaryotic replication. *Curr. Opin. Struct. Biol.* 15, 68-76.

- Dao, V., and Modrich, P. (1998). Mismatch-, MutS-, MutL-, and helicase II-dependent unwinding from the single-strand break of an incised heteroduplex. *J. Biol. Chem.* 273, 9202-9207.
- Datta, A., Schmeits, J.L., Amin, N.S., Lau, P.J., Myung, K., and Kolodner, R.D. (2000). Checkpoint-dependent activation of mutagenic repair in *Saccharomyces cerevisiae* pol3-01 mutants. *Mol. Cell* 6, 593-603.
- Drummond, J.T., Li, G.M., Longley, M.J., and Modrich, P. (1995). Isolation of an hMSH2-p160 heterodimer that restores DNA mismatch repair to tumor cells. *Science* 268, 1909-1912.
- Dzantiev, L., Constantin, N., Genschel, J., Iyer, R.R., Burgers, P.M., and Modrich, P. (2004). A defined human system that supports bidirectional mismatch-provoked excision. *Mol. Cell* 15, 31-41.
- Erzberger, J.P., and Berger, J.M. (2006). Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annu. Rev. Biophys. Biomol. Struct.* 35, 93-114.
- Erzberger, J.P., Mott, M.L., and Berger, J.M. (2006). Structural basis for ATP-dependent DnaA assembly and replication-origin remodeling. *Nat. Struct. Mol. Biol.* 13, 676-683.
- Fishel, R. (1998). Mismatch repair, molecular switches, and signal transduction. *Genes Dev.* 12, 2096-2101.
- Flores-Rozas, H., Clark, D., and Kolodner, R.D. (2000). Proliferating cell nuclear antigen and Msh2p-Msh6p interact to form an active mismatch recognition complex. *Nat. Genet.* 26, 375-378.
- Genschel, J., Bazemore, L.R., and Modrich, P. (2002). Human exonuclease I is required for 5' and 3' mismatch repair. *J. Biol. Chem.* 277, 13302-13311.
- Genschel, J., Littman, S.J., Drummond, J.T., and Modrich, P. (1998). Isolation of MutSbeta from human cells and comparison of the mismatch repair specificities of MutSbeta and MutSalpha. *J. Biol. Chem.* 273, 19895-19901.
- Genschel, J., and Modrich, P. (2003). Mechanism of 5'-directed excision in human mismatch repair. *Mol. Cell* 12, 1077-1086.
- Glover, B.P., and McHenry, C.S. (1998). The chi psi subunits of DNA polymerase III holoenzyme bind to single-stranded DNA-binding protein (SSB) and facilitate replication of an SSB-coated template. *J. Biol. Chem.* 273, 23476-23484.

Gomes, X.V., and Burgers, P.M. (2001). ATP utilization by yeast replication factor C. I. ATP-mediated interaction with DNA and with proliferating cell nuclear antigen. *J. Biol. Chem.* 276, 34768-34775.

Gomes, X.V., Gary, S.L., and Burgers, P.M. (2000). Overproduction in *Escherichia coli* and characterization of yeast replication factor C lacking the ligase homology domain. *J. Biol. Chem.* 275, 14541-14549.

Gomes, X.V., Schmidt, S.L., and Burgers, P.M. (2001). ATP utilization by yeast replication factor C. II. Multiple stepwise ATP binding events are required to load proliferating cell nuclear antigen onto primed DNA. *J. Biol. Chem.* 276, 34776-34783.

Gradia, S., Acharya, S., and Fishel, R. (2000). The role of mismatched nucleotides in activating the hMSH2-hMSH6 molecular switch. *J. Biol. Chem.* 275, 3922-3930.

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

Gu, L., Hong, Y., McCulloch, S., Watanabe, H., and Li, G.M. (1998). ATP-dependent interaction of human mismatch repair proteins and dual role of PCNA in mismatch repair. *Nucleic Acids Res.* 26, 1173-1178.

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

Guenther, B., Onrust, R., Sali, A., O'Donnell, M., and Kuriyan, J. (1997). Crystal structure of the delta' subunit of the clamp-loader complex of *E. coli* DNA polymerase III. *Cell* 91, 335-345.

Gulbis, J.M., Kelman, Z., Hurwitz, J., O'Donnell, M., and Kuriyan, J. (1996). Structure of the C-terminal region of p21(WAF1/CIP1) complexed with human PCNA. *Cell* 87, 297-306.

Guo, S., Presnell, S.R., Yuan, F., Zhang, Y., Gu, L., and Li, G.M. (2004). Differential requirement for proliferating cell nuclear antigen in 5' and 3' nick-directed excision in human mismatch repair. *J. Biol. Chem.* 279, 16912-16917.

Guo, S., Zhang, Y., Yuan, F., Gao, Y., Gu, L., Wong, I., and Li, G.M. (2006). Regulation of replication protein A functions in DNA mismatch repair by phosphorylation. *J. Biol. Chem.* 281, 21607-21616.

- Hanna, J.S., Kroll, E.S., Lundblad, V., and Spencer, F.A. (2001). *Saccharomyces cerevisiae* CTF18 and CTF4 are required for sister chromatid cohesion. *Mol. Cell. Biol.* *21*, 3144-3158.
- Hess, M.T., Gupta, R.D., and Kolodner, R.D. (2002). Dominant *Saccharomyces cerevisiae* msh6 mutations cause increased mispair binding and decreased dissociation from mispairs by Msh2-Msh6 in the presence of ATP. *J. Biol. Chem.* *277*, 25545-25553.
- Hingorani, M.M., and O'Donnell, M. (1998). ATP binding to the *Escherichia coli* clamp loader powers opening of the ring-shaped clamp of DNA polymerase III holoenzyme. *J. Biol. Chem.* *273*, 24550-24563.
- Iyer, L.M., Leipe, D.D., Koonin, E.V., and Aravind, L. (2004). Evolutionary history and higher order classification of AAA+ ATPases. *J. Struct. Biol.* *146*, 11-31.
- Iyer, R.R., Pluciennik, A., Burdett, V., and Modrich, P.L. (2006). DNA mismatch repair: functions and mechanisms. *Chem. Rev.* *106*, 302-323.
- Jeruzalmi, D., O'Donnell, M., and Kuriyan, J. (2001). Crystal structure of the processivity clamp loader gamma (gamma) complex of *E. coli* DNA polymerase III. *Cell* *106*, 429-441.
- Jeruzalmi, D., Yurieva, O., Zhao, Y., Young, M., Stewart, J., Hingorani, M., O'Donnell, M., and Kuriyan, J. (2001). Mechanism of processivity clamp opening by the delta subunit wrench of the clamp loader complex of *E. coli* DNA polymerase III. *Cell* *106*, 417-428.
- Jiang, J., Bai, L., Surtees, J.A., Gemici, Z., Wang, M.D., and Alani, E. (2005). Detection of high-affinity and sliding clamp modes for MSH2-MSH6 by single-molecule unzipping force analysis. *Mol. Cell* *20*, 771-781.
- Johnson, A., Yao, N.Y., Bowman, G.D., Kuriyan, J., and O'Donnell, M. (2006). The replication factor C clamp loader requires arginine finger sensors to drive DNA binding and proliferating cell nuclear antigen loading. *J. Biol. Chem.* *281*, 35531-35543.
- Joshi, A., Sen, S., and Rao, B.J. (2000). ATP-hydrolysis-dependent conformational switch modulates the stability of MutS-mismatch complexes. *Nucleic Acids Res.* *28*, 853-861.
- Junop, M.S., Obmolova, G., Rausch, K., Hsieh, P., and Yang, W. (2001). Composite active site of an ABC ATPase: MutS uses ATP to verify mismatch recognition and authorize DNA repair. *Mol. Cell* *7*, 1-12.

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

Kanellis, P., Agyei, R., and Durocher, D. (2003). Elg1 forms an alternative PCNA-interacting RFC complex required to maintain genome stability. *Curr. Biol.* 13, 1583-1595.

Kazmirski, S.L., Podobnik, M., Weitze, T.F., O'Donnell, M., and Kuriyan, J. (2004). Structural analysis of the inactive state of the Escherichia coli DNA polymerase clamp-loader complex. *Proc. Natl. Acad. Sci. U. S. A.* 101, 16750-16755.

Kazmirski, S.L., Zhao, Y., Bowman, G.D., O'donnell, M., and Kuriyan, J. (2005). Out-of-plane motions in open sliding clamps: molecular dynamics simulations of eukaryotic and archaeal proliferating cell nuclear antigen. *Proc. Natl. Acad. Sci. U. S. A.* 102, 13801-13806.

Kelman, Z., Yuzhakov, A., Andjelkovic, J., and O'Donnell, M. (1998). Devoted to the lagging strand-the subunit of DNA polymerase III holoenzyme contacts SSB to promote processive elongation and sliding clamp assembly. *EMBO J.* 17, 2436-2449.

Kenna, M.A., and Skibbens, R.V. (2003). Mechanical link between cohesion establishment and DNA replication: Ctf7p/Eco1p, a cohesion establishment factor, associates with three different replication factor C complexes. *Mol. Cell. Biol.* 23, 2999-3007.

Kondo, T., Matsumoto, K., and Sugimoto, K. (1999). Role of a complex containing Rad17, Mec3, and Ddc1 in the yeast DNA damage checkpoint pathway. *Mol. Cell. Biol.* 19, 1136-1143.

Kong, X.P., Onrust, R., O'Donnell, M., and Kuriyan, J. (1992). Three-dimensional structure of the beta subunit of E. coli DNA polymerase III holoenzyme: a sliding DNA clamp. *Cell* 69, 425-437.

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

Kramer, B., Kramer, W., Williamson, M.S., and Fogel, S. (1989). Heteroduplex DNA correction in Saccharomyces cerevisiae is mismatch specific and requires functional PMS genes. *Mol. Cell. Biol.* 9, 4432-4440.

- Krishna, T.S., Kong, X.P., Gary, S., Burgers, P.M., and Kuriyan, J. (1994). Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. *Cell* 79, 1233-1243.
- Kunkel, T.A., and Erie, D.A. (2005). DNA mismatch repair. *Annu. Rev. Biochem.* 74, 681-710.
- Lamers, M.H., Perrakis, A., Enzlin, J.H., Winterwerp, H.H., de Wind, N., and Sixma, T.K. (2000). The crystal structure of DNA mismatch repair protein MutS binding to a G x T mismatch. *Nature* 407, 711-717.
- Lau, P.J., and Kolodner, R.D. (2003). Transfer of the MSH2.MSH6 complex from proliferating cell nuclear antigen to mispaired bases in DNA. *J. Biol. Chem.* 278, 14-17.
- Lee, J.Y., Chang, J., Joseph, N., Ghirlando, R., Rao, D.N., and Yang, W. (2005). MutH complexed with hemi- and unmethylated DNAs: coupling base recognition and DNA cleavage. *Mol. Cell* 20, 155-166.
- Lee, J.Y., and Yang, W. (2006). UvrD helicase unwinds DNA one base pair at a time by a two-part power stroke. *Cell* 127, 1349-1360.
- Li, G.M. (2008). Mechanisms and functions of DNA mismatch repair. *Cell Res.* 18, 85-98.
- Li, G.M., and Modrich, P. (1995). Restoration of mismatch repair to nuclear extracts of H6 colorectal tumor cells by a heterodimer of human MutL homologs. *Proc. Natl. Acad. Sci. U. S. A.* 92, 1950-1954.
- Longley, M.J., Pierce, A.J., and Modrich, P. (1997). DNA polymerase delta is required for human mismatch repair in vitro. *J. Biol. Chem.* 272, 10917-10921.
- Majka, J., and Burgers, P.M. (2004). The PCNA-RFC families of DNA clamps and clamp loaders. *Prog. Nucleic Acid Res. Mol. Biol.* 78, 227-260.
- Majka, J., and Burgers, P.M. (2003). Yeast Rad17/Mec3/Ddc1: a sliding clamp for the DNA damage checkpoint. *Proc. Natl. Acad. Sci. U. S. A.* 100, 2249-2254.
- Maradeo, M.E., and Skibbens, R.V. (2009). The Elg1-RFC clamp-loading complex performs a role in sister chromatid cohesion. *PLoS ONE* 4, e4707.
- Martik, D., Baitinger, C., and Modrich, P. (2004). Differential specificities and simultaneous occupancy of human MutSalpha nucleotide binding sites. *J. Biol. Chem.* 279, 28402-28410.

- Mayer, M.L., Gygi, S.P., Aebersold, R., and Hieter, P. (2001). Identification of RFC(Ctf18p, Ctf8p, Dcc1p): an alternative RFC complex required for sister chromatid cohesion in *S. cerevisiae*. *Mol. Cell* 7, 959-970.
- Mazur, D.J., Mendillo, M.L., and Kolodner, R.D. (2006). Inhibition of Msh6 ATPase activity by mispaired DNA induces a Msh2(ATP)-Msh6(ATP) state capable of hydrolysis-independent movement along DNA. *Mol. Cell* 22, 39-49.
- Mendillo, M.L., Mazur, D.J., and Kolodner, R.D. (2005). Analysis of the interaction between the *Saccharomyces cerevisiae* MSH2-MSH6 and MLH1-PMS1 complexes with DNA using a reversible DNA end-blocking system. *J. Biol. Chem.* 280, 22245-22257.
- Messer, W. (2002). The bacterial replication initiator DnaA. DnaA and oriC, the bacterial mode to initiate DNA replication. *FEMS Microbiol. Rev.* 26, 355-374.
- Miyata, T., Suzuki, H., Oyama, T., Mayanagi, K., Ishino, Y., and Morikawa, K. (2005). Open clamp structure in the clamp-loading complex visualized by electron microscopic image analysis. *Proc. Natl. Acad. Sci. U. S. A.* 102, 13795-13800.
- Modrich, P. (2006). Mechanisms in eukaryotic mismatch repair. *J. Biol. Chem.* 281, 30305-30309.
- Modrich, P., and Lahue, R. (1996). Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annu. Rev. Biochem.* 65, 101-133.
- Moldovan, G.L., Pfander, B., and Jentsch, S. (2006). PCNA controls establishment of sister chromatid cohesion during S phase. *Mol. Cell* 23, 723-732.
- Mossi, R., Jonsson, Z.O., Allen, B.L., Hardin, S.H., and Hubscher, U. (1997). Replication factor C interacts with the C-terminal side of proliferating cell nuclear antigen. *J. Biol. Chem.* 272, 1769-1776.
- Naiki, T., Kondo, T., Nakada, D., Matsumoto, K., and Sugimoto, K. (2001). Chl12 (Ctf18) forms a novel replication factor C-related complex and functions redundantly with Rad24 in the DNA replication checkpoint pathway. *Mol. Cell. Biol.* 21, 5838-5845.
- Naktinis, V., Onrust, R., Fang, L., and O'Donnell, M. (1995). Assembly of a chromosomal replication machine: two DNA polymerases, a clamp loader, and sliding clamps in one holoenzyme particle. II. Intermediate complex between the clamp loader and its clamp. *J. Biol. Chem.* 270, 13358-13365.
- Naktinis, V., Turner, J., and O'Donnell, M. (1996). A molecular switch in a replication machine defined by an internal competition for protein rings. *Cell* 84, 137-145.

- Natrajan, G., Lamers, M.H., Enzlin, J.H., Winterwerp, H.H., Perrakis, A., and Sixma, T.K. (2003). Structures of Escherichia coli DNA mismatch repair enzyme MutS in complex with different mismatches: a common recognition mode for diverse substrates. *Nucleic Acids Res.* 31, 4814-4821.
- Obmolova, G., Ban, C., Hsieh, P., and Yang, W. (2000). Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. *Nature* 407, 703-710.
- Oku, T., Ikeda, S., Sasaki, H., Fukuda, K., Morioka, H., Ohtsuka, E., Yoshikawa, H., and Tsurimoto, T. (1998). Functional sites of human PCNA which interact with p21 (Cip1/Waf1), DNA polymerase delta and replication factor C. *Genes Cells* 3, 357-369.
- Olson, M.W., Dallmann, H.G., and McHenry, C.S. (1995). DnaX complex of Escherichia coli DNA polymerase III holoenzyme. The chi psi complex functions by increasing the affinity of tau and gamma for delta. *J. Biol. Chem.* 270, 29570-29577.
- Onrust, R., and O'Donnell, M. (1993). DNA polymerase III accessory proteins. II. Characterization of delta and delta'. *J. Biol. Chem.* 268, 11766-11772.
- Palombo, F., Gallinari, P., Iaccarino, I., Lettieri, T., Hughes, M., D'Arrigo, A., Truong, O., Hsuan, J.J., and Jiricny, J. (1995). GTBP, a 160-kilodalton protein essential for mismatch-binding activity in human cells. *Science* 268, 1912-1914.
- Palombo, F., Iaccarino, I., Nakajima, E., Ikejima, M., Shimada, T., and Jiricny, J. (1996). hMutSbeta, a heterodimer of hMSH2 and hMSH3, binds to insertion/deletion loops in DNA. *Curr. Biol.* 6, 1181-1184.
- Parnas, O., Zipin-Roitman, A., Mazor, Y., Liefshitz, B., Ben-Aroya, S., and Kupiec, M. (2009). The ELG1 clamp loader plays a role in sister chromatid cohesion. *PLoS ONE* 4, e5497.
- Peltomaki, P. (2005). Lynch syndrome genes. *Fam. Cancer.* 4, 227-232.
- Peltomaki, P. (2003). Role of DNA mismatch repair defects in the pathogenesis of human cancer. *J. Clin. Oncol.* 21, 1174-1179.
- Pietroni, P., and von Hippel, P.H. (2008). Multiple ATP binding is required to stabilize the "activated" (clamp open) clamp loader of the T4 DNA replication complex. *J. Biol. Chem.* 283, 28338-28353.

- Pietroni, P., Young, M.C., Latham, G.J., and von Hippel, P.H. (2001). Dissection of the ATP-driven reaction cycle of the bacteriophage T4 DNA replication processivity clamp loading system. *J. Mol. Biol.* *309*, 869-891.
- Podust, V.N., Tiwari, N., Stephan, S., and Fanning, E. (1998). Replication factor C disengages from proliferating cell nuclear antigen (PCNA) upon sliding clamp formation, and PCNA itself tethers DNA polymerase delta to DNA. *J. Biol. Chem.* *273*, 31992-31999.
- Ramilo, C., Gu, L., Guo, S., Zhang, X., Patrick, S.M., Turchi, J.J., and Li, G.M. (2002). Partial reconstitution of human DNA mismatch repair in vitro: characterization of the role of human replication protein A. *Mol. Cell. Biol.* *22*, 2037-2046.
- Raschle, M., Marra, G., Nystrom-Lahti, M., Schar, P., and Jiricny, J. (1999). Identification of hMutLbeta, a heterodimer of hMLH1 and hPMS1. *J. Biol. Chem.* *274*, 32368-32375.
- Rush, J., Lin, T.C., Quinones, M., Spicer, E.K., Douglas, I., Williams, K.R., and Konigsberg, W.H. (1989). The 44P subunit of the T4 DNA polymerase accessory protein complex catalyzes ATP hydrolysis. *J. Biol. Chem.* *264*, 10943-10953.
- Schmidt, S.L., Gomes, X.V., and Burgers, P.M. (2001). ATP utilization by yeast replication factor C. III. The ATP-binding domains of Rfc2, Rfc3, and Rfc4 are essential for DNA recognition and clamp loading. *J. Biol. Chem.* *276*, 34784-34791.
- Sexton, D.J., Kaboord, B.F., Berdis, A.J., Carver, T.E., and Benkovic, S.J. (1998). Dissecting the order of bacteriophage T4 DNA polymerase holoenzyme assembly. *Biochemistry* *37*, 7749-7756.
- Simonetta, K.R., Kazmirski, S.L., Goedken, E.R., Cantor, A.J., Kelch, B.A., McNally, R., Seyedin, S.N., Makino, D.L., O'Donnell, M., and Kuriyan, J. (2009). The mechanism of ATP-dependent primer-template recognition by a clamp loader complex. *Cell* *137*, 659-671.
- Skibbens, R.V., Corson, L.B., Koshland, D., and Hieter, P. (1999). Ctf7p is essential for sister chromatid cohesion and links mitotic chromosome structure to the DNA replication machinery. *Genes Dev.* *13*, 307-319.
- Smith, S., Hwang, J.Y., Banerjee, S., Majeed, A., Gupta, A., and Myung, K. (2004). Mutator genes for suppression of gross chromosomal rearrangements identified by a genome-wide screening in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.* *101*, 9039-9044.

- Snyder, A.K., Williams, C.R., Johnson, A., O'Donnell, M., and Bloom, L.B. (2004). Mechanism of loading the Escherichia coli DNA polymerase III sliding clamp: II. Uncoupling the beta and DNA binding activities of the gamma complex. *J. Biol. Chem.* 279, 4386-4393.
- Stukenberg, P.T., Studwell-Vaughan, P.S., and O'Donnell, M. (1991). Mechanism of the sliding beta-clamp of DNA polymerase III holoenzyme. *J. Biol. Chem.* 266, 11328-11334.
- Trakselis, M.A., Berdis, A.J., and Benkovic, S.J. (2003). Examination of the role of the clamp-loader and ATP hydrolysis in the formation of the bacteriophage T4 polymerase holoenzyme. *J. Mol. Biol.* 326, 435-451.
- Tran, H.T., Gordenin, D.A., and Resnick, M.A. (1999). The 3'-->5' exonucleases of DNA polymerases delta and epsilon and the 5'-->3' exonuclease Exo1 have major roles in postreplication mutation avoidance in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 19, 2000-2007.
- Tsurimoto, T., and Stillman, B. (1991). Replication factors required for SV40 DNA replication in vitro. I. DNA structure-specific recognition of a primer-template junction by eukaryotic DNA polymerases and their accessory proteins. *J. Biol. Chem.* 266, 1950-1960.
- Turner, J., Hingorani, M.M., Kelman, Z., and O'Donnell, M. (1999). The internal workings of a DNA polymerase clamp-loading machine. *EMBO J.* 18, 771-783.
- Uhlmann, F., Cai, J., Gibbs, E., O'Donnell, M., and Hurwitz, J. (1997). Deletion analysis of the large subunit p140 in human replication factor C reveals regions required for complex formation and replication activities. *J. Biol. Chem.* 272, 10058-10064.
- Umar, A., Buermeyer, A.B., Simon, J.A., Thomas, D.C., Clark, A.B., Liskay, R.M., and Kunkel, T.A. (1996). Requirement for PCNA in DNA mismatch repair at a step preceding DNA resynthesis. *Cell* 87, 65-73.
- Varlet, I., Canard, B., Brooks, P., Cerovic, G., and Radman, M. (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, 10156-10161.
- Wang, H., and Hays, J.B. (2002). Mismatch repair in human nuclear extracts. Time courses and ATP requirements for kinetically distinguishable steps leading to tightly controlled 5' to 3' and aphidicolin-sensitive 3' to 5' mispair-provoked excision. *J. Biol. Chem.* 277, 26143-26148.

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

Warbrick, E. (2000). The puzzle of PCNA's many partners. *Bioessays* *22*, 997-1006.

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

Williams, C.R., Snyder, A.K., Kuzmic, P., O'Donnell, M., and Bloom, L.B. (2004). Mechanism of loading the Escherichia coli DNA polymerase III sliding clamp: I. Two distinct activities for individual ATP sites in the gamma complex. *J. Biol. Chem.* *279*, 4376-4385.

Witte, G., Urbanke, C., and Curth, U. (2003). DNA polymerase III chi subunit ties single-stranded DNA binding protein to the bacterial replication machinery. *Nucleic Acids Res.* *31*, 4434-4440.

Yang, W., Junop, M.S., Ban, C., Obmolova, G., and Hsieh, P. (2000). DNA mismatch repair: from structure to mechanism. *Cold Spring Harb. Symp. Quant. Biol.* *65*, 225-232.

Yao, N., Coryell, L., Zhang, D., Georgescu, R.E., Finkelstein, J., Coman, M.M., Hingorani, M.M., and O'Donnell, M. (2003). Replication factor C clamp loader subunit arrangement within the circular pentamer and its attachment points to proliferating cell nuclear antigen. *J. Biol. Chem.* *278*, 50744-50753.

Yao, N., Turner, J., Kelman, Z., Stukenberg, P.T., Dean, F., Shechter, D., Pan, Z.Q., Hurwitz, J., and O'Donnell, M. (1996). Clamp loading, unloading and intrinsic stability of the PCNA, beta and gp45 sliding clamps of human, E. coli and T4 replicases. *Genes Cells* *1*, 101-113.

Yao, N.Y., Johnson, A., Bowman, G.D., Kuriyan, J., and O'Donnell, M. (2006). Mechanism of proliferating cell nuclear antigen clamp opening by replication factor C. *J. Biol. Chem.* *281*, 17528-17539.

Young, M.C., Weitzel, S.E., and von Hippel, P.H. (1996). The kinetic mechanism of formation of the bacteriophage T4 DNA polymerase sliding clamp. *J. Mol. Biol.* *264*, 440-452.

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

Zhuang, Z., Berdis, A.J., and Benkovic, S.J. (2006). An alternative clamp loading pathway via the T4 clamp loader gp44/62-DNA complex. *Biochemistry* *45*, 7976-7989.

Zhuang, Z., Yoder, B.L., Burgers, P.M., and Benkovic, S.J. (2006). The structure of a ring-opened proliferating cell nuclear antigen-replication factor C complex revealed by fluorescence energy transfer. *Proc. Natl. Acad. Sci. U. S. A.* *103*, 2546-2551.

CHAPTER TWO: AFM STUDY OF EUKARYOTIC RFC LOADING PCNA PROCESS

2.1 Introduction

Sliding clamp proteins increase DNA polymerase processivity from a few hundred nucleotides to several thousand nucleotides. However, the conserved closed circular structure of the sliding clamps requires that they must be loaded onto a DNA substrate by multi-protein complexes known as clamp loaders (Bloom, 2006; Davey, Jeruzalmi, et al, 2002; Stukenberg, Studwell-Vaughan, and O'Donnell, 1991).

Numerous studies of clamp loaders, including the γ -complex of *E.coli* bacteria, gp44/gp62 of T4 bacteriophage, and human and yeast RFC complexes have revealed detailed information about their structures and mechanisms of function. Clamp loaders from different organisms share some common features. (I) The peptide sequence and crystal structures of clamps are highly conserved from prokaryotes to eukaryotes (Bowman, O'Donnell, and Kuriyan, 2004; Cullmann, Fien, et al, 1995; Guenther, Onrust, et al, 1997; Jeruzalmi, O'Donnell, and Kuriyan, 2001; Yao, Turner, et al, 1996). All clamp loaders are composed of five subunits: δ , δ' and three copies of γ of γ -complex of *E.coli*, one gp62 and four copies of gp44 of gp44/gp62 of T4 clamp loader, and one large subunit(RFC-A) and four small subunits(RFC-B-C-D-E) of RFC. II. All three clamps are

AAA+ ATPases, with ATP binding sites lying between two subunits and an arginine finger sticking out from neighboring subunit to sense of ATP binding and stimulate ATP hydrolysis. III. Comparison the crystal structure of the γ -complex and RFC reveals that the N-terminals of all five subunits are arranged into a right-handed spiral and the C-termini of the five subunits form the C-terminal cylinder, termed “collar”, which is essential for oligomerization of the five subunits (Figure 1.6, 1.7). There is a DNA exit channel between the δ and δ' subunit of γ -complex and RFC-A and RFC-E subunit. IV. In regard to the clamp loading function of the clamp loaders, both the γ -complex and RFC recognize primer-template DNA and posit the DNA helix into the center of clamp. ATP binding is essential for binding PCNA and PCNA ring opening. ATP hydrolysis triggered by DNA substrate ejects RFC, leaving a PCNA-DNA complex after PCNA ring closing (Chen, Levin, et al, 2009; Gomes, and Burgers, 2001; Hingorani, and O'Donnell, 1998; Naktinis, Onrust, et al, 1995; Turner, Hingorani, et al, 1999). The clamp loading mechanism of γ -complex of *E.coli* has been well studied (Jeruzalmi, Yurieva, et al, 2001; Stewart, Hingorani, et al, 2001; Turner, Hingorani, et al, 1999). Crystal structure studies have led to a “wrench-motor-stator” model to explain how the five subunits of the γ -complex cooperate to open the β -clamp subunit interface. In this model, the δ subunit cracks open the β -dimer, even in the absence of γ and δ' , at one interface, and thus is termed the ‘wrench’. The three γ subunits are the only subunits of the clamp loader that bind ATP (Xiao, Naktinis, and O'Donnell, 1995), and therefore comprise the ‘motor’ that drives the clamp loading reaction. The relatively rigid protein δ' acts as a backboard, directing ATP-induced conformational changes of the other γ -complex subunits, and is

therefore termed the “stator”. ATP binding induces a conformational change in the γ complex, exposing δ for interaction with β and subsequent ring opening (Naktinis, Onrust, et al, 1995; Turner, Hingorani, et al, 1999). The δ subunit relaxes the spring tension of an α -helix at the interface of β clamp, thus disrupting the structure of the dimer interface (Jeruzalmi, Yurieva, et al, 2001). In the recently revealed structure of γ -complex with primer-template DNA (Simonetta, Kazmirski, et al, 2009), all the AAA+ residues of the three copies of γ subunit and δ' subunit form a right-handed spiral around the double-stranded portion of DNA, which is in a slightly distorted B-form conformation. Domain I of the δ subunit is disengaged from DNA. The collar domain of the δ subunit positions the side chain of Tyr 316 so that it stacks on the nucleotide base at the 3' end of the primer strand. This stacking results in a sharp bend in the template strand as it exits the clamp loader channel.

Although the γ clamp loader complex of *E. coli* has been well characterized, some features of clamp loading mechanism do not appear to be conserved from prokaryotes to eukaryotes system. In *E. coli*, the processivity clamp is termed as ‘ β clamp’, which is homodimer, however, both T4 bacteriophage clamp gp45 and eukaryotic PCNA clamp are homotrimers. In addition, unlike PCNA, which appears to have a tight trimer subunit interface (Yao, Turner, et al, 1996), gp45 clamp appears to have one subunit interface open in solution (Alley, Shier, et al, 1999). These differences between clamps in different species suggest that they may be loaded on DNA by different mechanisms. In fact, a few of studies have suggested that there are distinct mechanisms among these three species. In terms of the clamp loader structure, the γ -complex only has three active ATP-binding

sites, located at subunit interface $\delta'/\gamma 1$, $\gamma 1/\gamma 2$ and $\gamma 2/\gamma 3$, and two or three molecules of ATP can be bound (Jeruzalmi, O'Donnell, and Kuriyan, 2001; Kazmirski, Podobnik, et al, 2004; Williams, Snyder, et al, 2004). In contrast, the gp44/gp62 clamp loader of T4 bacteriophage and eukaryotic RFCs have four active ATP binding sites and RFC can bind five molecules of ATP in the presence of clamp and DNA (Bowman, O'Donnell, and Kuriyan, 2004; Chen, Levin, et al, 2009; Gomes, Schmidt, and Burgers, 2001; Gomes, and Burgers, 2001; Pietroni, Young, et al, 2001; Pietroni, and von Hippel, 2008; Trakselis, Berdis, and Benkovic, 2003). The fifth ATP is bound to RFC –E subunit which does not have a fully intact ATP binding site (Bowman, O'Donnell, and Kuriyan, 2004; Cai, Yao, et al, 1998; Cai, Yao, et al, 1998). So far, the function of the fifth ATP is unknown. Regarding the ATP utilization, three molecules of ATP of γ -complex are hydrolyzed upon DNA binding. The β -clamp alone does not stimulate ATP hydrolysis, but it enhances the hydrolysis rate stimulated by DNA (Bertram, Bloom, et al, 2000; Snyder, Williams, et al, 2004; Williams, Snyder, et al, 2004). Three different mechanisms for the role of ATP in clamp loading by the gp44/gp62 clamp loader uses have been proposed. All three mechanisms suggest different numbers of ATP bound and a different order of ATP hydrolysis stimulated by DNA and the gp45 clamp (Pietroni, Young, et al, 2001; Pietroni, and von Hippel, 2008; Sexton, Kaboord, et al, 1998; Trakselis, Berdis, and Benkovic, 2003; Zhuang, Berdis, and Benkovic, 2006). Recent research on RFC revealed that two of five RFC-bound ATPs are hydrolyzed in the presence of DNA and three of five RFC-bound ATPs hydrolyzed in the presence of DNA and PCNA (Chen, Levin, et al, 2009; Schmidt, Gomes, and Burgers, 2001).

In this study, we use single molecule technique AFM to investigate the details of the oligomerization of yeast RFC, and how the oligomerization is affected by nucleotide cofactors, PCNA, and DNA. Based on the intriguing observation that ATP causes the dissociation of RFC into small subunits and RFC subcomplex appears to remain with PCNA and DNA after ATP hydrolysis triggered by DNA binding, we hypothesize a putative mechanism of RFC clamp loading pathway. In addition, to better understand how RFC is involved into DNA mismatch repair pathway, we examine the binding properties of RFC to nicked DNA substrates. Intriguingly, we find that RFC forms filaments with nicked dsDNA in the absence of nucleotide cofactors and that RFC may possess helicase-like activity. Our results raise the possibility that RFC plays roles in other biological pathways.

2.2 Materials and Methods

Protein purification. Yeast RFC was a gift from Manju M. Hingorani Lab. PCNA was purified as described previously (Xue, Ratcliff, et al, 2002) .

DNA substrates for AFM study. To make pUC18Nick DNA substrate, the pUC18 plasmid is nicked by restriction enzyme Nt. BstNBI and purified by 1% agarose gel. 1077Nick DNA is made from pUC18Nick DNA substrate by cutting it using restriction enzymes, AlwNI and XmnI to make 1077 bp nicked DNA fragment. 1077Homo is created by cutting pUC18 plasmid using AlwNI and XmnI restriction enzymes and purified by 1% agarose gel.

AFM Imaging. Approximately 30nM RFC complex is incubated with ~5mM of different nucleotide cofactors for about 1 minute at room temperature in low salt binding buffer A (20mM Tris·HCl, pH 7.8, 50mM NaCl, 5mM BME, 5mM MgCl₂, 5% glycerol). The protein-DNA complex is formed by incubating ~5nM RFC (sometime with ~5nM PCNA trimer) with ~10nM DNA substrates in the absence or presence of ~5mM nucleotide cofactors for about 1 minute at room temperature in low salt binding buffer A or high salt binding buffer B (20mM Tris·HCl, pH 7.8, 100mM NaCl, 5mM BME, 5mM MaCl₂, 5% glycerol) in a total volume ~20μL. The reaction is deposited on freshly cleaved mica (Spruce Pina Mica Company) and quickly washed by deionized distilled water, blotted with filter paper and then dried under stream of nitrogen. All the images are collected at a scan rate ~3Hz in air with a Nanoscope IIIa (Digital Instrument) microscope in tapping mode. PiontprobeR tapping mode silicon probes (Molecular Imaging Corporation) with spring constants of ~50 N/m and resonance frequencies ~170 KHz are used for all imaging. The image resolution is 512 × 512 pixels.

2.3 Results

The oligomerization state of RFC depends on nucleotides

To examine the effects of nucleotide cofactors on the oligomerization state of RFC, we imaged RFC in the presence and absence of ATP, ADP and ATPγS. AFM is a powerful method for examining the oligomerization state of proteins (Ratcliff, and Erie, 2001; Xue, Ratcliff, et al, 2002; Yang, Wang, and Erie, 2003). It has been demonstrated

that the volume of proteins determined from AFM images depends linearly on molecular weight (AFM Volume =1.2MW-14.7).

Representative AFM images of RFC deposited in the presence of different nucleotide cofactors are shown in Figure 2.1. To quantitatively examine the oligomerization state of RFC, the volume of the proteins was determined as described previously (Ratcliff, and Erie, 2001; Xue, Ratcliff, et al, 2002; Yang, Wang, and Erie, 2003) and compared to the possible subcomplexes of RFC. RFC is composed of one large subunit with an approximate volume of 100 nm^3 and four small subunits with volumes of $\sim 40 \text{ nm}^3$. The predicated volumes of RFC and several RFC subcomplexes are shown in table 2.1.

To examine the effect of nucleotide on the oligomeric state of RFC, we deposited RFC with different nucleotide cofactors. Inspection of the histogram of the volume of RFC in the absence of nucleotide reveals two peaks at $\sim 30 \text{ nm}^3$, $\sim 250 \text{ nm}^3$. The peak at $\sim 250 \text{ nm}^3$ is consistent with RFC existing as an intact pentamer and the peak at $\sim 30 \text{ nm}^3$ the combination of the small subunits (Figure 2.2, a, Table 2.1). The ratio of the number of proteins in the peak at 30 nm^3 to that at 250 nm^3 is 1:5, which indicates $\sim 4\%$ of RFC is dissociated in the absence of nucleotide cofactor.

The volume distribution of RFC in the presence of ATP γ (Figure 2.2, c) is similar to that in the absence of nucleotides except that the second peak is shifted to higher volume (from $\sim 250 \text{ nm}^3$ to $\sim 330 \text{ nm}^3$). The peak ration is around 1:6, which indicated $\sim 3\%$ is dissociated in the presence of ATP γ S. It seems unlikely that RFC has gained more subunits; however, another source of the volume increase could be from a conformational

change of the RFC complex, because an AFM image is dialation between the tip and surface (Brar, Sacho, et al, 2008). Consequently, it is likely that this increase in volume is the result of a conformation change in RFC. To test this possibility, we measured the height of each RFC complex both in the absence and presence of ATP γ S. This statistical analysis (Figure 2.3) shows that the peak of the height distribution in the presence of ATP γ S relative to its absence is shifted to higher height by ~ 0.5 nm. This large change in height of proteins strongly suggests that ATP γ S induces a conformational change in RFC. A pre-steady kinetic study indicates that nucleotide binding induces a slow conformational change of RFC, which resulting a high binding affinity for PCNA. Here, we termed RFC in the presence of ATP γ S as “active” RFC and RFC without nucleotide binding as “inactive” RFC.

ATP causes dramatic dissociation of the RFC complex, depicted peaks at ~ 30 nm³ and ~ 90 nm³ (Figure 2.2, b).The peak ~ 90 nm³ is consisted of the large subunit of RFC and subcomplexes of two small subunits, such as RFC (D, E). The peak at ~ 30 nm³ is consistent with the size of the RFC small subunits. The ratio of these two peaks is about 2:1. In the situation where RFC dissociates into one large subunit and four small subunits, the peak ration is expected to be 4:1. If there are two small subunits and one larger subcomplex, the peak ratio is expected to be 1:1. The peak ratio of 2:1 suggests the peak at 90 nm³ is the mixture of RFC large subunit and subcomplex formed by two small subunits.

The volume histogram of RFC in the presence of ADP is broad with additional peaks appearing. The peaks at $\sim 30\text{nm}^3$ and $\sim 260\text{nm}^3$ are consistent with the peaks shown in the absence of nucleotide. The shoulder peak on the peak $\sim 260\text{nm}^3$ is Gaussian fit centered $\sim 100\text{nm}^3$, which seems like the recurrence peak as RFC with ATP. This result suggests that RFC is partially dissociated in the presence of ADP.

Four major volume peaks, 30nm^3 , 90nm^3 , 250nm^3 and 330nm^3 observed in this study represent five different conformations of RFC complex, small subunits, large subunit or subcomplex composed of two small subunits, inactive RFC and active RFC respectively.

RFC-PCNA complex in the presence of ATP or ATP γ S

We investigated how PCNA affects the oligomerization states of RFC by incubating RFC and PCNA together with ATP or ATP γ S. Representative AFM images in presence of ATP or ATP γ S are shown in Figure 2.4. Figure 2.5 shows the results of the volume analysis. In addition, we present the normalized sum of the volume distributions of PCNA and RFC with corresponding nucleotide cofactor. This sum distribution (pink) represents the null state in which no binding events happen between RFC and PCNA. In the presence of ATP, a nascent peak appears $\sim 400\text{nm}^3$, consistent with the size of a RFC-PCNA complex. In addition, compared with null state, there is a reduction in the peak at the volume of PCNA $\sim 90\text{nm}^3$, supporting the formation of RFC-PCNA complex. Intriguingly, a peak at $\sim 240\text{nm}^3$ appears. This peak is consistent with the size of RFC, or could possibly represent a subcomplex of RFC with PCNA. In the presence of ATP γ S,

instead of RFC-PCNA complex, we only find RFC subcomplexes with PCNA with a broad peak at $\sim 120\text{nm}^3$ to $\sim 180\text{nm}^3$, however we cannot discriminate which complexes the peak represents (Figure 2.5, b). It seems like $\text{ATP}\gamma\text{S}$ binding causes dissociation of the RFC-PCNA complex, although we cannot rule out the possibility that the dissociation happens during the deposition. Compared with the histogram of null state, only the large size of RFC $\sim 330\text{nm}^3$ (active RFC) appears to bind with PCNA, because the peak consistent with inactive RFC ($\sim 250\text{nm}^3$) is maintained. This result suggests that only active RFC efficiently interacts with PCNA in the presence of ATP or $\text{ATP}\gamma\text{S}$, but the interaction of RFC with PCNA in the presence of $\text{ATP}\gamma\text{S}$ appears to result in partial dissociation of RFC.

RFC oligomerization induced by binding with primer-template DNA with $\text{ATP}\gamma\text{S}$

During the DNA replication process, RFC specifically recognizes the DNA 3' primer-template DNA junction and loads PCNA with N-terminal facing to 3' end and the C-terminal tethers DNA polymerase to DNA substrate (Goedken, Kazmirski, et al, 2005; Hingorani, and Coman, 2002). In this study we inspect how primer-template affects the oligomerization states of RFC or RFC-PCNA complexes in the presence of $\text{ATP}\gamma\text{S}$. We use a 40/65 ptDNA substrate with a 40 base pair primer strand and 65 base pair template strand, leaving a 25 base 5' overhang (Figure 2.6). Representative AFM images of RFC-ptDNA complexes are shown in Figure 2.6 a. Most particles show a single RFC binding complex, some of them having a DNA tail sticking out of the complex (Figure 2.6 a, left). A small percentage of the particles contains a double RFC binding complex (Figure 2.6 a,

right). Volume calculations reveal that most single particles are consistent with full length of RFC, volume $\sim 330\text{nm}^3$, and most of doubly bound complexes contain one full length of RFC and a smaller size RFC subcomplex, volume $\sim 200\text{nm}^3$. Quantitative volume analysis shows the volume distribution of RFC-40/65 ptDNA in the presence of ATP γ S complex is double Gaussian centered $\sim 40\text{nm}^3$ and $\sim 330\text{nm}^3$ (Figure 2.6, b), however, the volume distribution of RFC-PCNA-40/65 ptDNA ternary complex is double Gaussian centered $\sim 40\text{nm}^3$ and $\sim 200\text{nm}^3$ (Figure 2.6,d). The size of 330nm^3 is consistent with the size of RFC with ATP γ S. The size of 200nm^3 is approximately the size of the particles occurring in the reaction of RFC incubation with PCNA in the presence of ATP γ S. Once again, it is most likely that the RFC-PCNA complex is not stable in the presence of ATP γ S, independent of in presence or absence of DNA substrate. The 200nm^3 particle is close to various RFC-PCNA subcomplexes, such as RFC-A-B-C, RFC-D-E-PCNA (Table 2.1). Determination of the composition of the complexes, however, requires further investigation.

RFC forms filaments with nicked dsDNA in the absence of nucleotide

Very little is known about the properties of RFC-nicked DNA complexes. To characterize the properties of RFC binding to nicked DNA, we imaged RFC with nicked and unnicked DNA in the presence and absence of nucleotide cofactors at low and high salt concentrations.

First, we used a 1077 bp DNA fragment with a single nick 477 bp from one end (Figure 2.3). In the absence of nucleotide cofactor and at low salt concentration (50mM

NaCl, Figure 2.8 a), we see several different classes of complexes of RFC with DNA (Examples shown in Table 2.2). Approximate 45% of the DNA fragments have no RFC bound (free DNA) and ~40% are fully coated with RFC. The average length of DNA covered by RFC is similar to that of free nicked dsDNA, indicating DNA maintains its natural conformation and it is not wrapped around RFC. In addition to fully coated DNA, we also see partially coated and more interestingly, partially and fully coated DNA branched structures (Table 2.1), where a RFC-coated side branch sticks out from the main DNA backbone and looks like letter “Y”. RFC appears to have promoted the unwinding of a short segment of the DNA at the nick. In addition, we found some fully coated RFC-DNA tracks, which are significantly shorter than full length DNA. We speculate that these complexes may be RFC-coated single-stranded DNA, because we know RFC can bind to single-stranded DNA (Hingorani, and Coman, 2002). Consistent with the idea that RFC binding can promote the melting of double stranded DNA containing a nick, we also observe two (or a few) RFCs bound with the DNA partially melted at the nick and similar to the double complexes seen with primer template DNA (Figure 2.6, a), one RFC bound at ss-ds junction and the other RFC bound to the single strand of DNA (Figure 2.8 d, Table 2.1). It seems like RFC is peeling the single strand DNA from the dsDNA. Statistical analysis shows except for 50% of DNA without RFC binding, fully coated RFC-DNA track contributes 38.8% of all RFC-DNA complexes. These results may explain the observation by others that the addition of RFC to primer templates DNA results in aggregation (Chen, Levin, et al, 2009). DNA with few particle binding is RFC binds to DNA, rather than DNA track and RFC peeing DNA complex.

To examine the dependency of the formation of the complexes, especially the filaments, on salt concentration, we used a high salt concentration buffer (100mM NaCl, Figure 2.8, b). Similar classes of RFC-DNA complexes are found. However, there is a sharp drop in the number of fully coated DNA fragments and an increase in DNAs with a few particles bound. There is also a small increase in the percentage of free DNA. These results suggest that high salt concentration interrupts the RFC-RFC interaction between RFC.

To assess the role of the nick in the formation of these different complexes, we also imaged RFC with unnicked DNA, 1077Unnick, which has the same length and sequence as 1077Nick except there is no nick (Figure 2.4, c). We see very few RFCs binding to unnicked DNA, which suggests that the filament formation is initiated at the nick.

RFC-DNA complexes formed in presence of ATP or ADP

We added ATP into the RFC- nicked DNA solution to examine how the nucleotide affects the RFC- DNA complex. Because we observed filament formation in the absence of ATP, we added all components in the order of RFC, ATP, and the DNA. In the presence of ATP, we see primarily very large RFC-DNA complexes (Figure 2.9 a). It appears that multiple RFCs aggregate with multiple 1077Nick DNAs and form “spider-like” complexes. It is obvious that the length of DNA is shorter than the length of 1077Nick DNA, indicating that DNA is wrapped up inside the complex. The size of the bulky clump in the center of the complex ranges from $\sim 300\text{nm}^3$ to 15000nm^3 , which is corresponds to 1-50 RFCs (not including the DNA volume). Only 2% of the complexes

show a single RFC bound. To determine if the formation of these large complexes is dependent on salt concentration, as is the formation of the filaments in the absence of ATP, we increased the NaCl concentration from 50mM to 100mM. As can be seen in the image in Figure 2.9 b, similar complexes are found as in the low salt concentration.

To assess the role of the nick in the formation of these large complexes, we replaced 1077Nick with 1077Unnick DNA as we did in the absence of nucleotide (Figure 2.9, c). RFC forms similar large complexes with unnicked DNA. However, many of these complexes have both ends of the DNA contained within the large RFC-DNA complex. These results suggest that DNA ends may be necessary for the formation of these large complexes.

Because the N-terminal domain of RFC has a DNA ligase motif, that appears to increase nonspecific binding of RFC to double-stranded DNA (Uhlmann, Cai, et al, 1997). We examined whether this domain was responsible for the formation of the large complexes. We deposited nicked DNA with the N-terminal-truncated RFC and ATP in low salt buffer (Figure 2.9, d). Smaller complexes are found in this situation. The average size of truncated RFC-DNA complex is $\sim 1800\text{nm}^3$, which is smaller than that of full length RFC-DNA complexes ($\sim 2300\text{nm}^3$), suggesting a fewer number of RFCs bind with DNA.

To determine if different nucleotide cofactor affect the formation of RFC-DNA complexes, we replaced ATP with ADP or ATP γ S. In the presence of ADP, we find complexes similar to those seen in the presence of ATP formed both in the case of

1077nick dsDNA and 1077 Unnick dsDNA (Figure 2.9, e,f). However, in the presence of ATP γ S, we do see single RFCs bound to DNA (Figure 2.10, a). The analysis of these results is discussed in the following section.

To examine if DNA ends are responsible for the formation of large complexes, we used a circular nicked DNA substrate, pUC18Nick in the presence of ATP and low salt (Figure 2.13, a). Volume analysis results show only RFC subcomplexes bind to DNA (Figure, 2.13, c). This result indicates that DNA ends are essential for the formation of large RFC-DNA complexes. The physiological function of any of such large RFC-DNA complexes remains unknown.

Bend angle and binding specificity analysis of RFC in the presence of ATP γ S

To determine if RFC has any binding specificity to the nick, we employed the 1077Nick DNA substrate which has a single nick positioned ~140 nm from one end and deposited it with RFC in the presence of ATP γ S. A representative image of the RFC-DNA complexes is shown in the Figure 2.10 a, left. Because we know the position of the nick (specific site) on the DNA, we can discriminate between RFC bound at nick (specific complexes) and RFC bound to unnicked sites (nonspecific complexes). To determine the RFC binding specificity, we measure the distance from each end of the DNA to the center of the RFC-DNA complex. The distribution of binding positions is broad, indicating that RFC exhibits weak binding specificity to the nick (Figure 2.10 b). We also deposited RFC with homoduplex DNA, 1077Unnick, in the presence of ATP γ S;

unfortunately, we did not see either RFC or RFC-DNA complexes on the surface (Figure 2.10 a, right).

A recent crystal structure of the γ -complex from *E.coli.* with primer-template DNA predicts that clamp loaders bend the template ssDNA as it exits from the complex (Simonetta, Kazmirski, et al, 2009). A similar conformation has been proposed for RFC bound to primer-template DNA; however nothing is known about the conformation of RFC bound to nicked or fully double-stranded DNA. We measured bend angles induced by RFC bound at nick and at homoduplex sites. For this analysis, we excluded data in which multiple RFCs are bound to the same DNA to avoid counting aberrant bend angles. The distributions of bend angles induced at the specific site and at nonspecific sites are double Gaussian with peaks centered at $\sim 50^\circ$ and $\sim 0^\circ$ (Figure 2.10 c), $\sim 36^\circ$ and $\sim 0^\circ$ (Figure 2.10 d), respectively. The larger bend angle induced by RFC at nick is probably due to the flexible structure of DNA at the nick. Both bend angle distributions are broad. The breadth of bend angle distribution is related with the flexibility of protein-DNA complex. The breadth of bend angle distribution is $\sim 40^\circ$ at non-specific sites and $\sim 20^\circ$ at the specific site, which suggests that RFC forms more flexible complexes with DNA at non-specific sites than at specific sites. It is unexpected that we find a peak with a 0° bend angle for RFC-DNA complexes. Non-specific binding produces more unbent complex than specific binding. In the crystal structure of the γ -complex with pt DNA, the DNA is bent as it exits from the crank formed by δ and δ' subunit. We speculate that the unbent complex may result from different binding conformations of RFC, or that the DNA exits from another channel, like C-collar channel formed by RFC collar domains. The

population of unbent complexes at specific site is significantly smaller than that of nonspecific sites. There is possibility that peak at 0° from the contamination of the specific complexes with non-specific one. Because the two DNA ends are undistinguishable, we cannot discriminate between RFC bound to the to 40% from the short end (specific binding) and RFC bound to the position 40% from the other end, and given the low specificity of binding, the 'specific' complex distribution will have significant contamination with nonspecific complexes.

To characterize the oligomerization states of RFC, we measured the volume of RFC complexes bound to the specific sites, nonspecific site, and the protein distributed on the background. The results indicate that full length RFC binds to the DNA at both specific sites and nonspecific sites (Figure 2.11, a, b). In contrast, only RFC subcomplexes are found on the background (Figure 2.11, c). Because ATP γ S alone does not cause the dissociation of RFC (Figure 2.1), this observation suggests that DNA is destabilizing the RFC-ATP γ S complex.

RFC forms five classes of complexes with nicked circular DNA in the absence of nucleotides

PUC18Nick is a circular DNA with a single nick. Nicked circular DNA is the typical substrates used in vital studies using RFC and PCNA. Although we cannot know the position of the nick with this substrate, it avoids complications from having DNA ends. From inspection of the AFM images (Figure 2.12a), we classified the complexes formed by RFC and nicked circular DNA in the absence of nucleotide into five categories,

which are shown in the Table 2.3. Most of the nicked DNAs have either one (52.1%) or two (22.8%) RFC's bound. In addition, 6.6% of RFC-DNA complexes have more than two RFCs bound. Similar to the complexes found in the presence of nicked linear dsDNA without nucleotide cofactor, we also find that 12% of RFC complexes show the DNA partially melted (Table 2.3) and 6.6% show a track of RFC-bound DNA. The difference between RFC-bound DNA track on nicked circular DNA and the RFC filament on nicked linear dsDNA (Figure 2.8, a) is that RFC appears to bring two dsDNA together with nicked circular DNA. Bend angle analysis for DNA with one or two RFCs bound shows similar bend angle distributions with peaks at 0° , 28° and 60° for single RFC binding (Figure 2.12, b) and at 0° , 22° and 52° for double RFC binding (Figure 2.12, c).

Bend angle and volume analysis of RFC on nicked circular DNA in the presence of ATP

To examine the conformation of RFC bound to the nicked circular DNA in the presence of ATP which is required to load PCNA, we incubated RFC and pUC18Nick DNA together with ATP for ~1 min, and deposited the reaction on the mica for imaging. Inspection of Figure 2.13 shows small RFC or RFC subcomplexes bound to the DNA. These images are different markedly from those seen with linear nicked DNA in the presence of ATP (Figure 2.9), which shows the formation of large complexes. These results indicate that the formation of the large complexes requires DNA ends. In addition, analysis of the bend angles reveals a single peak, centered $\sim 40^\circ$ (Figure 2.13, b), which is different from the bimodal distribution seen in the absence of ATP (Figure 2.10, c, d). Volume analysis of these complexes yields a single peak at $\sim 188\text{nm}^3$, which is too small

for full length RFC and is consistent with a RFC-A-B-C subcomplex (Table 2.1). This size is also close to the predicted AFM volume of RFC-B-C-D-E subcomplex; however, a previous study showed that the RFC-B-C-D-E subcomplex exhibits poor binding affinity to DNA substrates (Gomes, and Burgers, 2001). The distribution of volumes of the protein distributed on the background, is Gaussian centered at $\sim 94\text{nm}^3$, which is consisted with a RFC subcomplex with two small subunits and, or RFC large subunit. Because this size is complimentary to the size of RFC subcomplex bound with DNA, and because the RFC-D-E subcomplex is known to be stable to be in solution and can unload PCNA from DNA substrate by itself (Yao, Johnson, et al, 2006), we hypothesize this subcomplex is RFC-D-E subcomplex.

Bend angle and volume analysis of RFC-PCNA on nicked circular DNA substrate in the presence of ATP

It is surprising that we observe RFC subcomplex bound DNA after DNA-stimulated ATP hydrolysis, because it is generally believed that full length RFC is ejected from the DNA as a result of ATP hydrolysis. To determine if this RFC subcomplex is due to the absence of PCNA, we added PCNA into previous system. This experiment was done in high salt concentration buffer (Figure 2.14 a), because in low salt concentration buffer, multiple RFCs and DNA warp up together (Image is not showed). The distribution of AFM volumes of the protein bound to DNA is a double Gaussian centered $\sim 190\text{nm}^3$ and $\sim 290\text{nm}^3$ (Figure 2.14, b). The peak of $\sim 190\text{nm}^3$ is similar to the peak seen with RFC alone and DNA (Figure 2.13, c), which indicates a RFC subcomplex bound to

nicked circular DNA in the absence and presence of PCNA. The peak $\sim 290\text{nm}^3$ is consistent with a complex of RFC-A-B-C and PCNA (Table 2.1). RFC-PCNA-pUC18Nick DNA complex has a broader distribution of bend angle (Figure 2.14, c), compared with RFC-pUC18nick complex. It is probably caused by different binding conformations of RFC subcomplex and RFC+PCNA subcomplex.

2.4 Discussion

As the PCNA clamp loader, RFC plays an essential role in DNA replication. Currently, there are many questions that remain about the mechanism of RFC loading PCNA. These puzzles include how does RFC recognize and load PCNA onto different DNA substrates? What role does ATP play in different loading steps? How does the DNA stimulate ATP hydrolysis? Do all four ATP binding sites act in a cooperative pathway or independently? How do nucleotides and PCNA binding introduce RFC conformation changes? How does the ATP-induced conformation change in RFC affect the binding affinity to PCNA and DNA? What is the mechanism by which RFC is ejected from DNA substrate after PCNA loading?

In this study, we take advantage of the single molecule technique of AFM to investigate the conformational properties of RFC and its complexes with DNA and PCNA in the presence and absence of nucleotide cofactors. We demonstrate that ATP-dependent changes in the interactions between and within the clamp loader subunits promote conformational changes in the complex that alter interactions with the clamp and

DNA. Binding interactions with the clamp and DNA also promote changes in the clamp loader that alter interactions for the next step pathway.

ATP induces the dissociation of RFC into monomers and subcomplexes

Nucleotide cofactor binding to clamp loaders initiates the entire cycle of clamp loading (Bertram, Bloom, et al, 1998; Chen, Levin, et al, 2009; Gomes, and Burgers, 2001; Hingorani, and O'Donnell, 1998; Yao, Coryell, et al, 2003). In *E.coli*, the δ subunit is blocked from interaction with β -clamp by δ' ; however, ATP binding introduced a conformational change of γ complex, exposing the δ subunit for interaction with β -clamp (Jeruzalmi, Yurieva, et al, 2001; Turner, Hingorani, et al, 1999). In a pre-steady-state kinetic study of RFC, nucleotide binding appears to cause a slow conformational change of RFC before it can interact with PCNA (Chen, Levin, et al, 2009). Our results are consistent with this suggestion, in that ATP γ S causes an increase in both the height and volume of RFC (Figure 2.2, 2.3). The increase in height suggests that RFC undergoes a conformational change. ATP binding may drive a motion of RFC into a more spiral conformation, opening like a spring, resulting in increasing the height and volume of RFC. This 'more spiral' conformation may facilitate RFC opening of the closed PCNA ring and binding along the DNA helix.

Unlike ATP γ S, ATP induces the dissociation of RFC. This dissociation is probably caused by the weak ATPase activity of RFC. The dissociation product is classified into two sizes of subcomplex; one consistent with the size of RFC small subunits and the other consistent with the RFC large subunit or subcomplex of two small

subunits. Analyzing the peak ratio of these two sizes of protein, we found the peak of larger size protein probably represents the combination of large subunit and subcomplex of two small subunits. However, there is another possibility. Because only subcomplex of RFC-D-E, RFC-B-C, RFC-C-D, RFC-B and RFC-D are found to be soluble in solution (Yao, Coryell, et al, 2003), it is possible that the other insoluble subcomplexes or subunits, like RFC-A, have precipitated out of solution, which would affect the ratio of these two peaks representing different complexes.

The situation in the presence of ADP is a little more complicated. The size of most of RFCs in the presence of ADP is similar to that of RFC in the absence of nucleotide; however, some of the RFC is dissociated as there is a shoulder on the main peak with a volume $\sim 100\text{nm}^3$. This volume is similar to one of the peaks seen in the presence of ATP. The observations that RFC does not dissociate into subunits in the presence of ATP γ S, and that it is completely dissociated in the presence of ATP but only partially dissociated in the presence of ADP suggest that ATP hydrolysis and/ or the mixed occupancy of the ATPase site with both ATP and ADP leads to significant destabilization of the subunit interactions. These results suggest that interactions between nucleotide cofactor and arginine finger govern the stability of the subunit interactions. As discussed in Chapter 1, the ATPase site on RFC lie at the interface between the subunits, with the arginine finger of the ATPase site being contributed by the adjacent subunit. Perhaps in the absence of nucleotide, the five subunits of RFC are packed with the ATP binding sites and the arginine finger close to each other. ATP binding may open the ATPase site with the γ -phosphate of ATP interacting with the arginine finger (Johnson, Yao, et al, 2006). After

ATP hydrolysis, the absence of γ -phosphate in ADP would remove the connection between ATP binding site and arginine finger, which in turn may destabilize the connection between each subunit, resulting in the dissociation of RFC.

PCNA stabilizes RFC in the presence of ATP

It has been demonstrated that ATP or ATP γ S increases the binding affinity of RFC for PCNA (Chen, Levin, et al, 2009; Gomes, and Burgers, 2001). It has been proposed that the interaction between clamp and clamp loader induces a conformational change in clamp, resulting in the opening of clamp subunits at one interface (Chen, Levin, et al, 2009). To investigate how PCNA affects the conformation of RFC, we examine the oligomerization states of RFC-PCNA complex in the presence of ATP and ATP γ S.

Although ATP binding and hydrolysis result in the dissociation of RFC, PCNA appears to weaken the effects caused by the hydrolysis of ATP. Inspection of the volume data in Figure 2.5 reveals that PCNA dramatically stabilizes RFC in the presence of ATP. Specifically, in the absence of PCNA, ATP causes RFC to dissociation into monomers and small subcomplexes, whereas, in the presence of PCNA, a significant fraction of RFC exist as a full complex ($\sim 240\text{nm}^3$) and there is a peak consistent with RFC bound to PCNA ($\sim 400\text{nm}^3$). Studies of the ATPase activity of T4 clamp loaders have shown that the presence of clamp inhibits ATP hydrolysis by T4 clamp loader (Pietroni, and von Hippel, 2008). Based on our data, inhibition of hydrolysis is would be expected to stabilize RFC, and therefore, the RFC-PCNA complex. Notably, the presence of PCNA also appears to stabilize RFC alone (although it is possible that the peak $\sim 240\text{nm}^3$

represents an interaction of a subcomplex of RFC with PCNA). RFC may be stabilized by PCNA if it is cycling on and off and the re-association of PCNA is faster than the dissociation of RFC into subcomplexes.

Surprisingly, in the presence of ATP γ S, instead of full length RFC-PCNA complex, we only find subcomplexes of RFC with PCNA, indicated by the smaller size peaks from $\sim 160\text{nm}^3$ to $\sim 200\text{nm}^3$; that is, PCNA appears to promote the disassembly of RFC in the presence of ATP γ S (Figure 2.5 b). Even the addition of primer-template DNA into this does not inhibition the dissociation. We, however, could not identify the different subcomplexes that these peaks represent. This result suggests that RFC does not form stable complexes with PCNA in the presence of ATP γ S which in conflict with an EM study that visualized an RFC-PCNA complex in the presence of ATP γ S (Miyata, Suzuki, et al, 2005). However, in this study, the clamp-clamp loader complex was purified by gel filtration before the EM images were acquired. Consequently, this complex may not reflect the major population in the reaction. The PCNA-induced RFC dissociation in the presence of ATP γ S may be because ATP γ S binding induces RFC into a more spiral conformation, and this more spiral conformation of RFC, in turn, induces the PCNA docked under the RFC into an out-of-plan open conformation (Kazmirski, Zhao, et al, 2005). This RFC-PCNA (open) complex is a transient conformation during the clamp loading process. The out-of-plan conformation of PCNA does not appear to be the favorite conformation of PCNA. The motion of PCNA back to an in-plane conformation causes the loss of contact with RFC, which is indicated in the crystal structure of RFC-PCNA complex (Bowman, O'Donnell, and Kuriyan, 2004), in which PCNA is in an in-

plane conformation and does not contact the RFC-D and RFC-E subunits. If the interactions between PCNA and RFC subunits are more favorable than the interaction between RFC subunits, the interaction of RFC with PCNA may induce the disassembly of RFC, which is suggested by our results. This PCNA motion is probably combined with the PCNA-stimulated ATP hydrolysis because in the presence of ATP, we did not observe PCNA-induced RFC dissociation (Figure 2.5 a). Previous studies indicate that in the ATP bound the ATP site between RFC-D and RFC-E is specifically stimulated hydrolysis by PCNA ((Johnson, Yao, et al, 2006). It is possible that this PCNA-stimulated ATP hydrolysis reduces the binding affinity between PCNA and RFC subunits allowing PCNA to relax back to the in-plane conformation. Because ATP γ S inhibits this ATP turnover reaction which this relaxation of PCNA is inhibited in turn, may cause the dissociation of RFC. In addition, a fluorescence energy transfer study indicates that ATP and ATP γ S induces different conformations of RFC-PCNA complexes (Zhuang, Yoder, et al, 2006), with one of the PCNA interface being opened by RFC $\sim 34\text{\AA}$ in the presence of ATP and $\sim 27\text{\AA}$ in the presence of ATP γ S. These results suggest that there are different conformations of the RFC-PCNA complex in the presence of different nucleotides, or that different RFC-PCNA complexes are formed.

RFC recognition of nicked DNA substrate

RFC and PCNA are involved in DNA mismatch repair pathway. In a reconstituted bidirectional mismatch repair system, RFC and PCNA activate a latent endonuclease activity of MutL α in an ATP- and mismatch-dependent manner (Kadyrov, Dzantiev, et al,

2006). Incision by activated MutL α endonuclease occurs on both 3'- and 5'-heteroduplex strand. Exo I performs 5' to 3' excision from the MutL α incision site through and beyond the site of the mismatch (Genschel, Bazemore, and Modrich, 2002; Zhang, Yuan, et al, 2005). PCNA is also required for MMR initiation and resynthesis step (Gu, Hong, et al, 1998; Umar, Buermeier, et al, 1996). Although substantive research suggest that RFC specially recognizes primer-template junctions and position DNA in the center of PCNA to form a topological link between these two, little is known about how RFC recognizes nick DNA substrate, which is the typical DNA substrate in the reconstituted DNA repair system. The nick provides the strand discrimination signal in the MMR process. To characterize the conformation of RFC bound to the nicked DNA, we examined the binding of RFC to nicked dsDNA in the presence of ATP γ S. RFC exhibits weak binding specificity to the nick (Figure 2.10, b). It exhibits a bimodal population of bend angles: a bent population and an unbent population (Figure 2.10, c and d), at both nicked and unnicked sites. RFC bends DNA more at nick than at unnicked sites. In addition, more unbent RFC-DNA complexes are found at unnicked sites. These results suggest that RFC may have two different binding modes on the DNA and that the ease of bending nicked DNA (or primer-template DNA) may contribute to recognition.

We attempted to conduct a similar analysis with ATP instead of ATP γ S; however addition of ATP to linear DNA (nicked and unnicked) led to large 'spider-like' complexes containing multiple RFCs and multiple DNAs. The formation of these large complexes appears to be dependent on DNA ends, because such complexes were not formed in the presence of nicked circular DNA (Figure 2.13, a). The physiological role of

these complexes, if any, is unclear; however, it is possible that similar complexes could be involved in processes such as recombination or on double strand break repair.

Possible mechanism of PCNA loading

Although we cannot determine the position of the nick on nicked circular DNA, it is useful to analyze the size of complexes and the RFC induced conformational changes in the DNA. Interestingly, unlike in the presence of ATP γ S, the distribution of bend angles is a single Gaussian, with a peak at $\sim 40^\circ$. In addition, most of the DNAs have a only a single RFC bound. Perhaps these complexes represent RFC bound at the nick. This idea is consistent with the observation that more bent complexes are seen at the nick on linear DNA in the presence of ATP γ S. If true, this result would suggest that the ATP increases the specificity of RFC for a nick.

The most interesting observation comes from the volume analysis of RFC both bound to the DNA and free on the surface. The average volume of the complex bound to the DNA is $\sim 180\text{nm}^3$ and that of the complexes on the surface is $\sim 100\text{nm}^3$. Both of these volumes are too small to represent full length RFC and therefore, indicate that, in the absence of PCNA, ATP results in subcomplexes of RFC bound to the DNA. Because RFC bound to DNA in the presence of ATP γ S appears to contain all five subunits (based on volume analysis), these results suggest that ATP hydrolysis ejects some the RFC subunits, which is also consistent with our observation that ATP induces dissociation of RFC into smaller subunits in the absence of DNA (Figure 2.2). Based on the volume analysis result, we speculate that the small RFC subcomplex that remains on the DNA is

RFC-A-B-C subcomplex and that free RFC distributed on the surface is the RFC-D-E subcomplex (Figure 2.13, Table 2.1). We make this hypothesis based on the observation that RFC-D-E subunit is stable in solution (Yao, Coryell, et al, 2003) and that RFC-D-E can unload PCNA from a circular DNA substrates (Yao, Johnson, et al, 2006). We also examine the oligomerization of the complexes bound to the nicked DNA in the presence of both PCNA and ATP. Interestingly, we still find a population of complexes that are similar size ($\sim 190\text{nm}^3$) to the RFC subcomplex bound to DNA in the absence of PCNA. In addition, we also find larger complexes that are consisted with the size of RFC-A-B-C/PCNA complex ($\sim 290\text{nm}^3$) (Figure 2.14, Table 2.1). Taking all these results together, we make a hypothesis about the clamp loading process, depicted in the Figure 2.15. In this model, RFC undergoes conformational change upon ATP binding, which increases binding affinity to the PCNA clamp. Interaction with RFC results in PCNA ring opening (Chen, Levin, et al, 2009; Gomes, and Burgers, 2001) and RFC positions the DNA helix into the center of PCNA. DNA triggers RFC-bound ATP hydrolysis (Naktinis, Turner, and O'Donnell, 1996; Podust, Tiwari, et al, 1998), which results in the two subunits RFC-D-E being ejected from the PCNA-DNA complex. The RFC-A-B-C subcomplex maintains contact with PCNA and DNA, and topologically links these two. The idea that the subcomplex of RFC-D-E is ejected from DNA during the PCNA loading process is consistent with several other results. First, the interaction between RFC-D-E and PCNA appears to be primarily responsible for PCNA-ring opening (Yao, Johnson, et al, 2006). The ejection of RFC-D-E may cause PCNA-ring to close and circle around the DNA helix. Second, RFC-D-E is capable of unloading PCNA from DNA substrate (Yao,

Johnson, et al, 2006). The ejected RFC-D-E subcomplex could then attend unloading PCNA from synthesized okizaki segments. Regarding the function of RFC-A-B-C subcomplex, we speculate that this subcomplex could anchor PCNA at the junction of primer-template or nicked DNA. The N-terminal ligase homolog domain of RFC-A subunit, which is capable of non-specific to binding single-stranded and double-stranded DNA (Gomes, Gary, and Burgers, 2000; Hingorani, and Coman, 2002; Uhlmann, Cai, et al, 1997) and has a strong binding with PCNA (Bowman, O'Donnell, and Kuriyan, 2004; Yao, Johnson, et al, 2006), could stabilize this interaction. RFC-B and RFC-C also interact with DNA and PCNA to some extent. Consequently, RFC-A-B-C subcomplex may be capable of fixing PCNA at specific sites. Maintaining PCNA at a primer-template junction or a nick would give downstream proteins, such as Pol δ time to interact with PCNA at the specific site. The competition between PCNA-binding proteins may stimulate the release of the RFC subcomplex from the DNA substrate. In support of this idea, in the T4 clamp/clamp loader system, gp44/gp62 plays a role as a molecular chaperone (Sexton, Kaboord, et al, 1998; Trakselis, Berdis, and Benkovic, 2003). A complex of gp44/gp62·gp45·DNA·gp43 has been identified as an intermediate in the formation of gp45·DNA·gp43. The gp43, a DNA polymerase in T4 bacteriophage, interacts with the same PCNA face as the gp44/gp62 clamp loader.

RFC may have functions other than loading PCNA on DNA

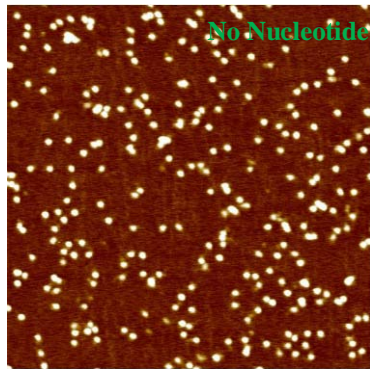
There is another surprising observation which is that RFC forms protein filaments with nicked dsDNA in the absence of nucleotide. RecA plays central in DNA

homologous recombination. The RecA protein binds strongly and in long clusters to ssDNA to form a nucleoprotein filament. The protein has more than one DNA binding site, and thus can hold a single strand and double strand together. This feature makes it possible to catalyze a DNA synapsis reaction between a DNA double helix and a homologous region of single stranded DNA. The reaction initiates the exchange of strands between two recombining DNA double helices (Guo, Zhang, et al, 2006). As a member of AAA+ ATPase family, RecA and RFC share the conserved ASCE core nucleotide-binding pocket. DnaA, another AAA+ ATPase protein, also forms protein filaments on DNA (Erzberger, Mott, and Berger, 2006; Erzberger, and Berger, 2006). Based on the structural similarity with other AAA+ ATPase proteins and our observations, it is possible RFC could possess other activities besides functioning as clamp loader. Another potential capability of RFC is that it could function as a DNA. We observed that in some RFC-DNA complexes, RFC splits DNA double strand from the nick and another RFC binds with the single stranded DNA (Figure 2.8, d). Considering many AAA+ ATPase proteins have helicase activity (Erzberger, and Berger, 2006), like DnaA, it would not be surprising if RFC had the helicase activity too, although more experiments will be required to assess this possibility. In support of this idea, the recently released E.coli γ -complex with primer-template DNA crystal structure mentioned that γ -complex has side chain Try316, which is reminiscent of the role of an aromatic side chain in the UvrD helicase that serves as a “separation pin” by splitting the path of DNA (Lee, and Yang, 2006; Simonetta, Kazmirski, et al, 2009). In addition, they found a second channel formed by C-terminal collar domain that could accommodate DNA, although

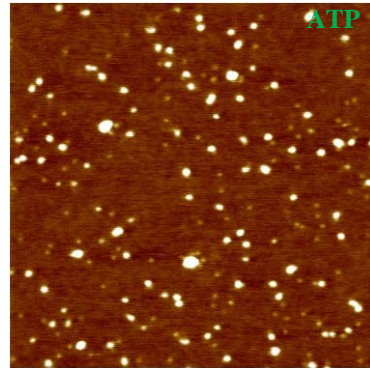
they did not present any direct evidence that DNA goes through this second channel. Based on the highly conserved sequence and structure between RFC and γ -complex (Cullmann, Fien, et al, 1995), RFC is likely to have the similar second channel and separation pin structure.

Besides the clamp loader RFC, there are other clamp loaders termed ‘alternative clamp loader’ that shared the same small subunits RFC-B-C-D-E but RFC-A is replaced by other subunits, such as Rad24, Ctf18 and Elg1, (Majka, and Burgers, 2004). These proteins function in the DNA damage checkpoint pathway and establishment of sister chromatid cohesion, which is essential for maintaining chromosome stability. Some of them function in the same pathways, such as Elg1-RFC and Ctf18- RFC which are both involved in the establishment of sister chromatid cohesion, although they have different functions (Maradeo, and Skibbens, 2009; Parnas, Zipin-Roitman, et al, 2009). All three alternative clamp loaders could interact with PCNA and PCNA is also found involved in their pathways (Moldovan, Pfander, and Jentsch, 2006). RFC shares a similar structure and clamp loader function with the alternative clamp loaders. Taken together, our observation that multiple RFCs and dsDNAs could form spider-like complex it suggest that RFC may be involved into these pathways too.

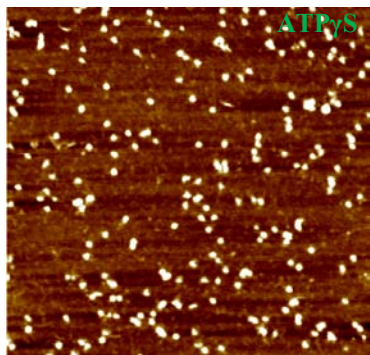
Briefly, RFC likely plays multiple roles in different pathways, like DNA recombination, DNA damage checkpoint, genome modification. As a complicated protein with multiple ATPase sites, it may have functions besides clamp loader function.



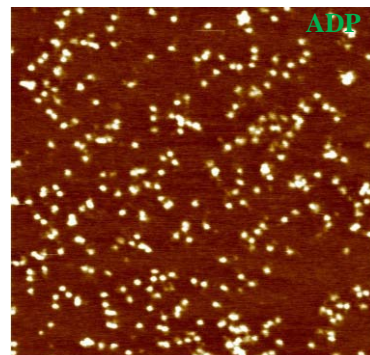
(a)



(b)



(c)



(d)

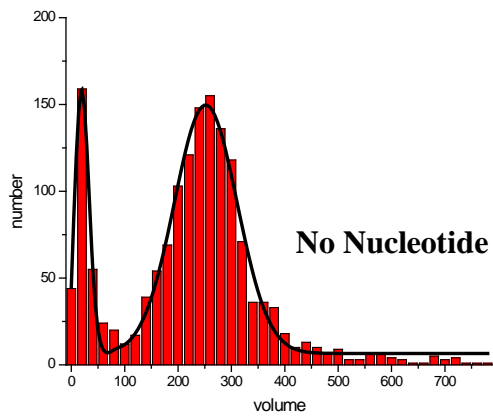
Figure 2.1 AFM images of RFC in the presence of different nucleotide cofactors. The image's size is 1 μ m \times 1 μ m. a, RFC in the absence of nucleotide. b, RFC in the presence of ATP. c, RFC in the presence of ATP γ S. d, RFC in the presence of ADP. All the experiments are performed in the same conditions. The concentration of RFC is 30 nM. The concentration of nucleotide is 5 mM. The protein and nucleotide are incubated in the buffer (Tris·HCl 20 mM (pH 7.4), NaCl, 50 mM, BME, 5mM, MgCl₂, 5mM, 5% Glycerol) for about 1 minute.

Table 2.1. Relationship of *S.cerevisiae* RFC' MW with AFM volume^a

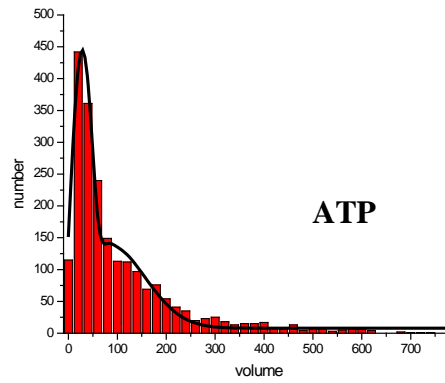
	MW(k Da)	AFM Volume (nm ³) ^b
RFC	248.8	283.9
RFC-A	94.9	99.2
RFC-E	39.9	33.2
RFC-A-B-C-D	208.9	236.0
RFC-B-C-D-E	153.9	170.0
RFC-A-B-C	169.2	188.3
RFC-D-E	79.6	80.8
PCNA	87.0	89.7
RFC+PCNA	335.8	388.3
RFC-A-B-C+PCNA	256.2	292.7
RFC-D-E+PCNA	166.6	185.2
RFC-E+PCNA	120.2	129.5

a. RFC assembled into right handed spiral in the order of RFC (A-B-C-D-E).

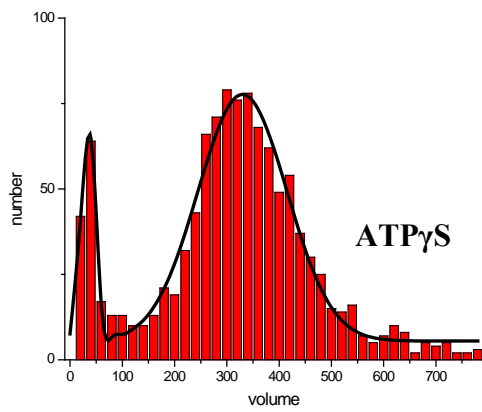
b. AFM volume is calculated based on the formula “AFM Volume=1.2×MW-14.7”



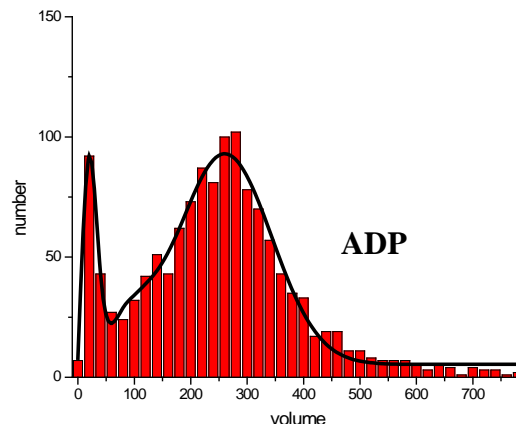
(a)



(b)



(c)



(d)

Figure 2.2 Histograms of RFC volume distributions with different nucleotide

cofactors. a, RFC is in the absence of nucleotide. The double Gaussian fit is centered at 21nm^3 , 258nm^3 , which is consistent with RFC small subunits and full length of RFC respectively. b, RFC is in presence of ATP. The double Gaussian fit is centered at 28nm^3 , 82nm^3 , which is consistent with RFC small subunits and the RFC large subunit or a dimer of RFC small subunits. c, RFC in presence of ATP γ S. The double Gaussian fit is centered at 34nm^3 and 330nm^3 , which is consistent with RFC small subunits and full length RFC. d, RFC in presence of ADP. The multiple Gaussian fit is centered at 28nm^3 , 90nm^3 , 260nm^3 , which is consistent with small RFC subunits, the RFC large subunit or a dimer of two RFC small subunit and full length RFC, respectively.

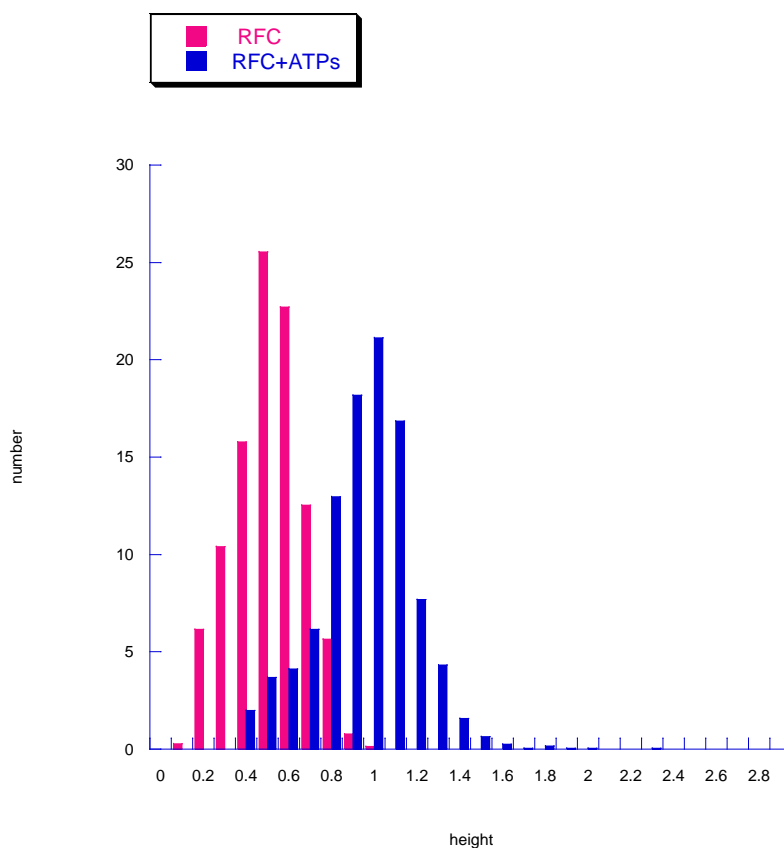
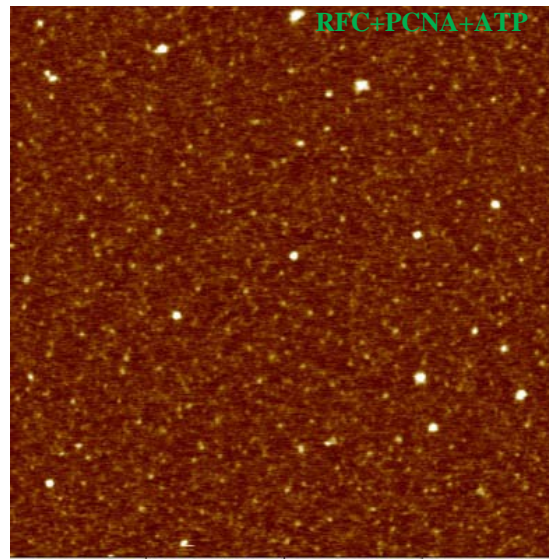
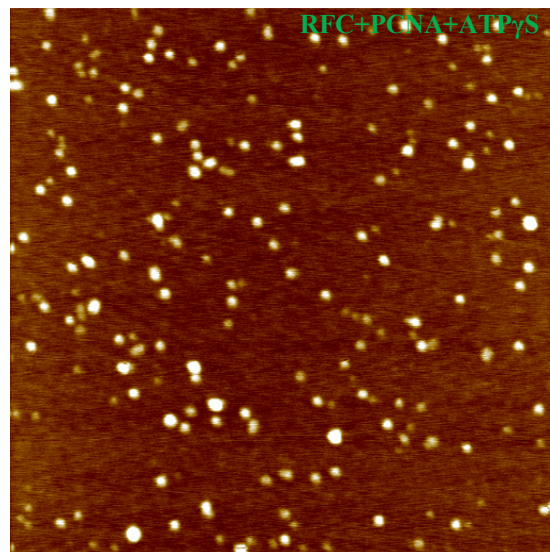


Figure 2.3 Nucleotide binding introduces RFC conformational change. The distribution of the height of RFC in the absence of nucleotide is fit by a single Gaussian centered ~ 0.5 nm (pink). The distribution of the height of RFC in the presence of ATP γ S is fit by a single Gaussian centered ~ 1 nm (blue). The AFM volume of RFC in the presence of ATP γ S is much larger than that of RFC alone. The volume accretion is due to the increase height of RFC upon interaction with ATP γ S. The center of the height histogram shifts to right by ~ 0.5 nm, comparing RFC alone (pink) with RFC \cdot ATP γ S (blue). The increase in height suggests that there is a conformational change of RFC

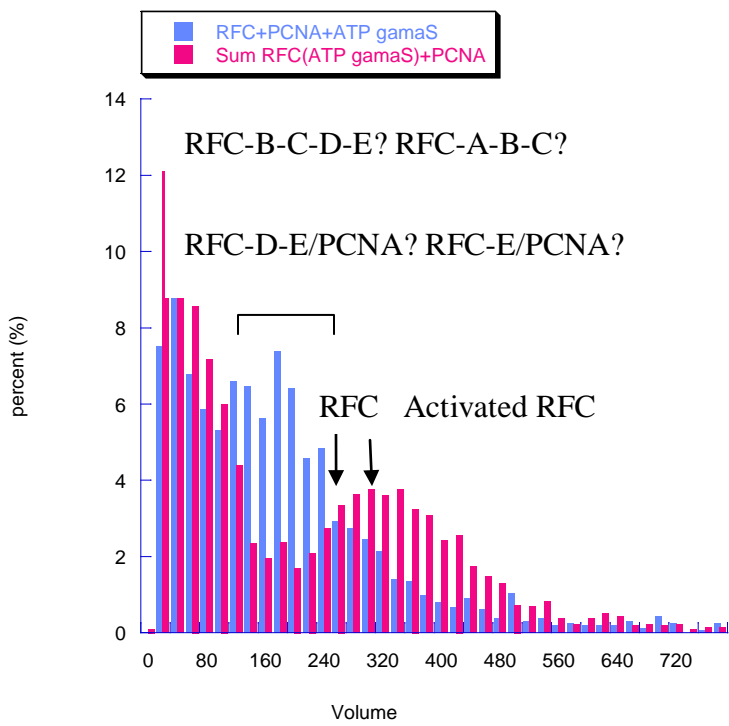
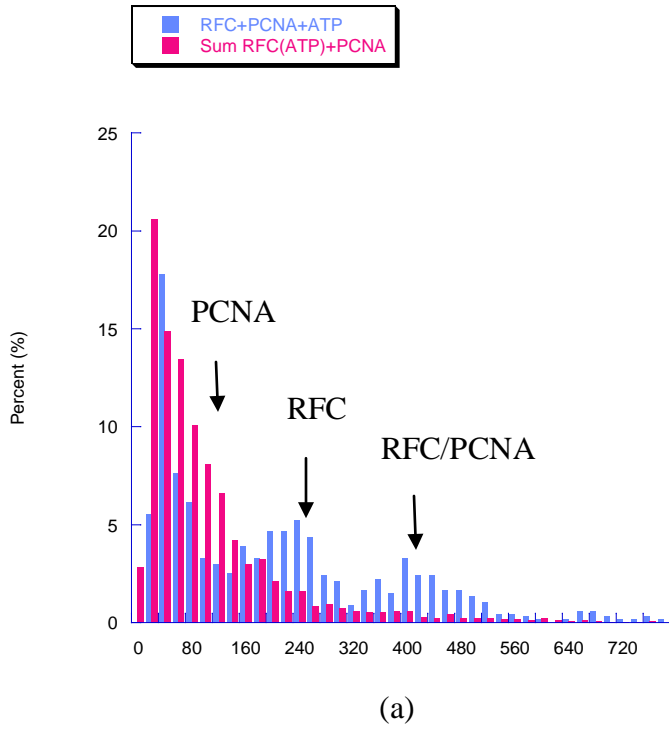


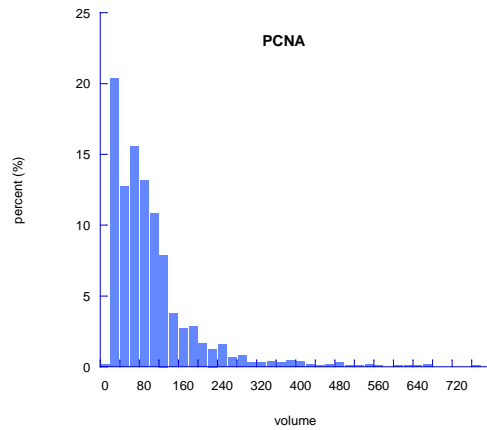
(a)



(b)

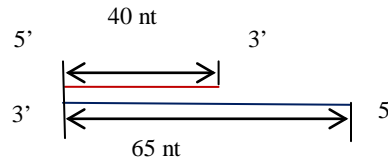
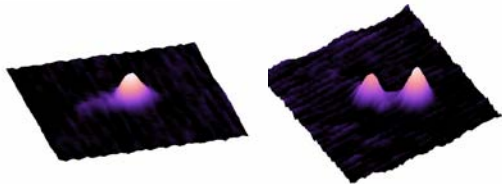
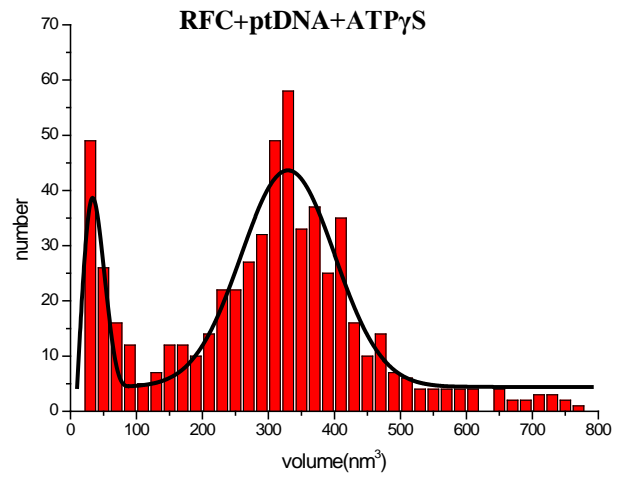
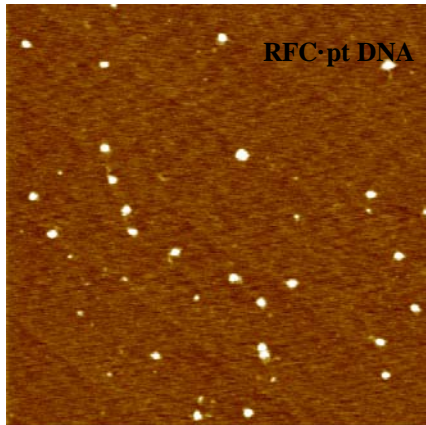
Figure 2.4 AFM images of RFC·PCNA complexes in the presence of ATP and ATP γ S respectively. The images are size 1 μ m \times 1 μ m. a.RFC (10nM) incubated with PCNA (10nM) in the presence of ATP (5mM). b. RFC (10nM) incubated with PCNA (10nM) in the presence of ATP γ S (5mM).





(c)

Figure 2.5 Histograms of volume of RFC-PCNA complexes in the presence of ATP and ATP γ S, and of PCNA. The concentration of PCNA and RFC is 10nM, and 10nM respectively. The y-axis represents the percentage of various particles distributed on the surface. The volume histogram of RFC-PCNA complex (blue) is presented with the sum of volume histogram of RFC with corresponding nucleotide and volume histogram of PCNA (pink) together. a. In the presence of ATP, interaction with PCNA reduces the dissociation of RFC, inferred by the peak around 240 nm³. The peak centered on 400 nm³ represents the RFC-PCNA complex. The decreasing percent of the PCNA peak, ~ 90 nm³, indicates the interaction between RFC and PCNA. b. In the presence of ATP γ S, instead of full length RFC-PCNA, a putative RFC subcomplex-PCNA is found, inferred by the peak around 180nm³. There are four different particles fit the peak here. It is hard to distinguish using AFM technique. Compared with sum volume distribution, the peak ~300nm³, which matches the size of activated RFC, is reduced. The peak ~240nm³, which matches the size of inactive RFC, remains the same percentage. c. Histogram of volumes of PCNA (~30nM as trimer). The peak at ~ 90nm³ represents PCNA trimer.



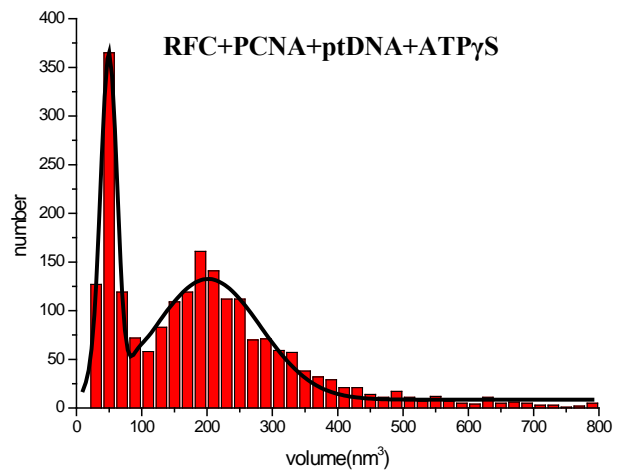
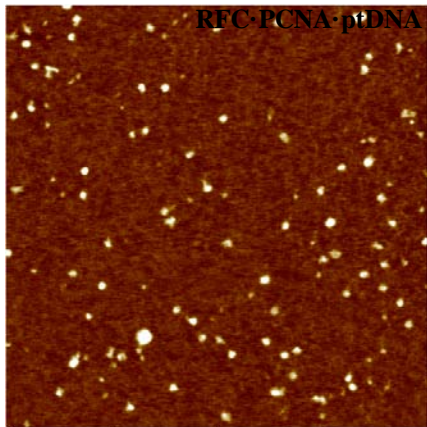
Single binding
Volume=291.9

Double binding
Volume=312.3, 207

40/65 pt DNA

(a)

(b)



(c)

(d)

Figure 2.6 AFM images and histograms of the volume distributions of RFC–primer template DNA complex with ATP γ S in the presence or absence of PCNA. a. Field view of RFC-40/65 ptDNA complex in the presence of ATP γ S. Representative surface plots of a single RFC binding and a double RFC binding complex are shown under the field view image. b. Histogram of volumes of RFC-40/65 ptDNA complexes in the presence of ATP γ S. The curves represents a double Gaussian fit with peaks are centered at 40nm³ and 330nm³, representing small subunits of RFC and full length RFC or a RFC-DNA complex, respectively. The schematic view of DNA substrate used in this study is shown under the histogram. The template strand and the primer strand contain 65 and 40 nucleotides respectively, resulting in a short 40/65 pt DNA substrate with a 5' overhang. c. Field view of RFC-PCNA40/65 ptDNA ternary complex in the presence of ATP γ S. d. Histogram of volumes of RFC-PCNA-40/65 ptDNA ternary complexes in the presence of ATP γ S. The double Gaussian fit peaks are centered on 50nm³ and 200nm³ respectively. The low volume peak represents PCNA and/or small subunits of RFC. The large size peak represent various subassemblies of the RFC-PCNA-pt DNA complex. The concentration of RFC, PCNA and ATP γ S is 5nM, 10nM and 5mM, respectively. The image size is 2 μ m \times 2 μ m.

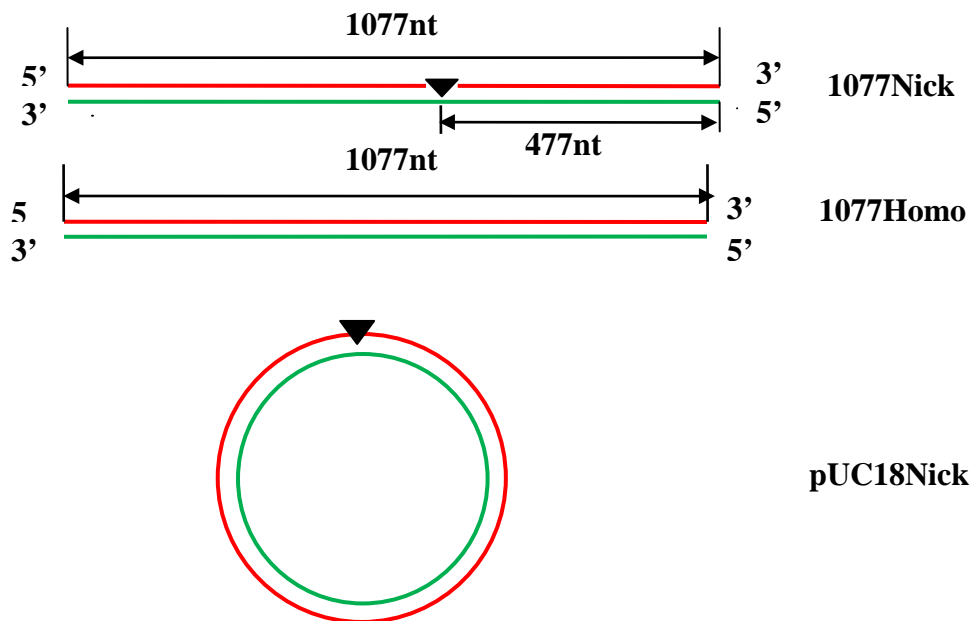
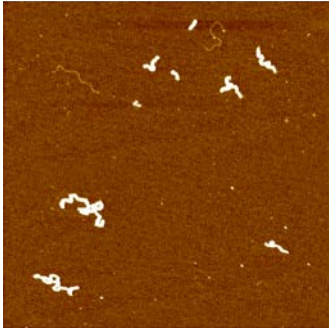


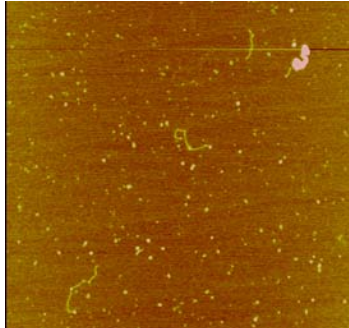
Figure 2.7 Schematic view of DNA substrates used in the AFM study. The DNA substrates are identified on the right by the length of the fragment and existence of the nick. The 1077 piece is made from XmnI and AlwNI digested pUC18 plasmid and amplified by PCR. The pUC18 plasmid is nicked by NbsI to break the 1077Nick DNA substrate before the cutting. The existence of the nick is checked by gel electrophoresis. pUC18Nick is made from NbsI digested the pUC18 plasmid and purified by 1% Agrose gel. The 1077Homo and 1077Nick have the same length and nucleotide sequence. 1077Nick has a nick at ~40% of the length from the 5' end of the bottom strand.

Nicked DNA /Low salt



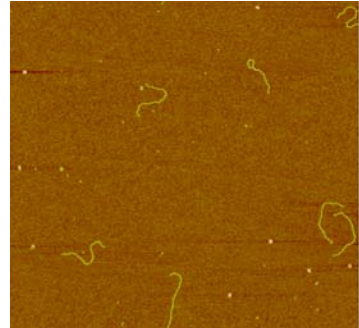
(a)

Nicked DNA /high salt

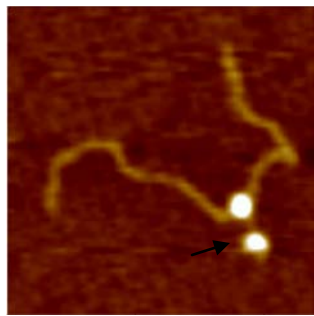


(b)

Unnicked DNA /Low salt



(c)



(d)

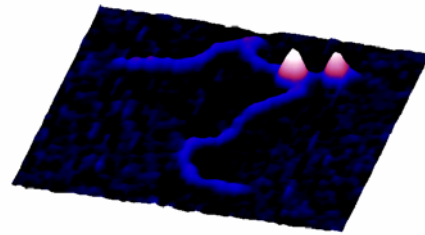
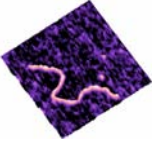

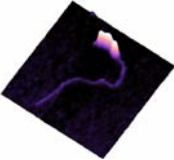

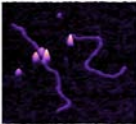
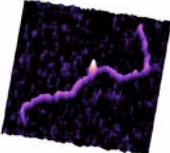
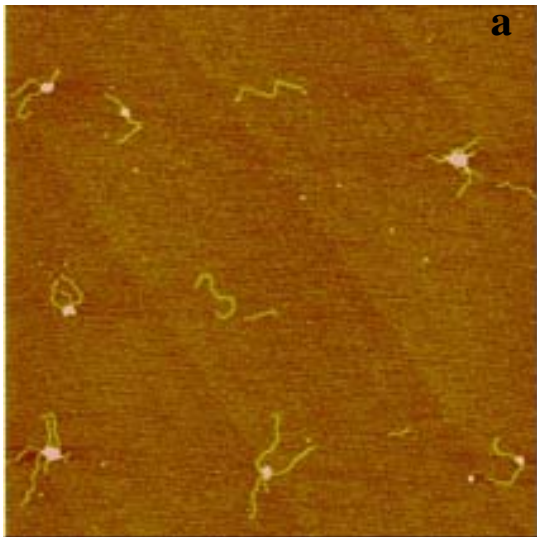


Figure 2.8 Representative AFM images of RFC complexes formed on the 1077Nick substrate in the absence of nucleotide. a, Filament of RFC on 1077 nicked dsDNA. The RFC concentration is 5nM. RFC and 1077 nicked dsDNA are incubated in low salt concentration buffer (Tris·HCl 20 mM (pH 7.4), NaCl, 50 mM, BME, 5mM, MgCl_2 , 5mM, 5% Glycerol) for ~1 minute. b, AFM image of RFC and nicked dsDNA filament in high salt concentration buffer (Tris·HCl 20 mM (pH 7.4), NaCl, 100 mM, BME, 5mM, MgCl_2 , 5mM, 5% Glycerol). The DNA and protein concentrations are the same as above. c, RFC incubated with 1077Homo in the absence of nucleotide. d, Example of a RFC nicked DNA complex in which the DNA is partially melted at the nick. Arrow points to the single strand. Left is topview image. Right is a surface image.

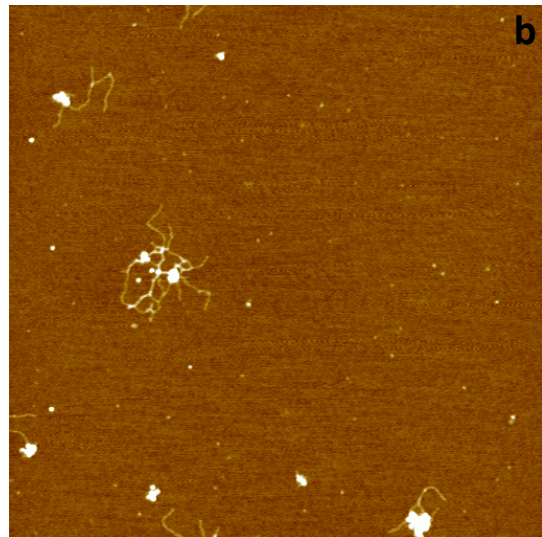
Table2.2 Statistical conformation analysis of RFC on 1077Nick DNA substrate in absence of nucleotide

		Low Salt (%)	High Salt (%)
Free DNA		44.6	58.1
Fully coated		38.8	8.8
Partially coated		5.8	2.5
With branch		2.5	0.6
RFC peeling dsDNA		4.1	2.5
With few particle bound		5.0	27.5
Total number of DNA		121	160

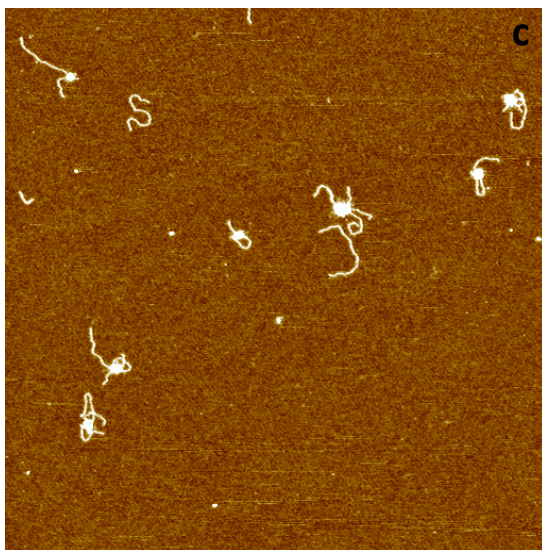
Nicked dsDNA +RFC + ATP/ Low salt



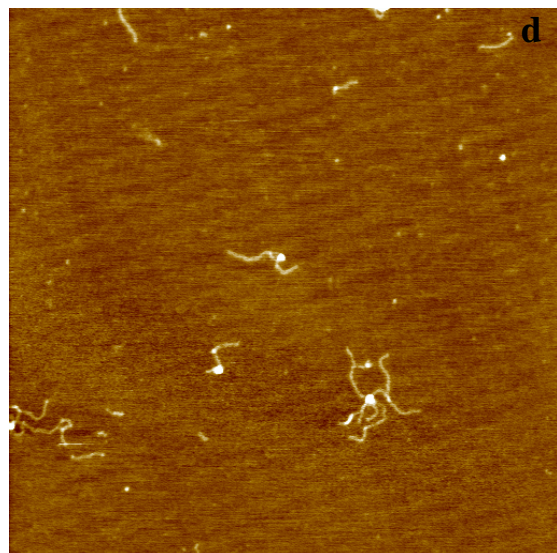
Nicked dsDNA +RFC + ATP/High salt



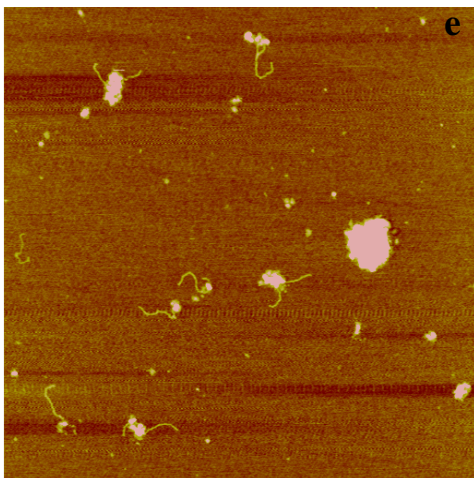
Homoduplex dsDNA+RFC + ATP/ Low salt



Nicked dsDNA + ATP+ Truncated RFC/ Low salt



Nicked dsDNA+RFC+ADP/Low salt



Homoduplex dsDNA dsDNA+RFC+ADP/Low salt

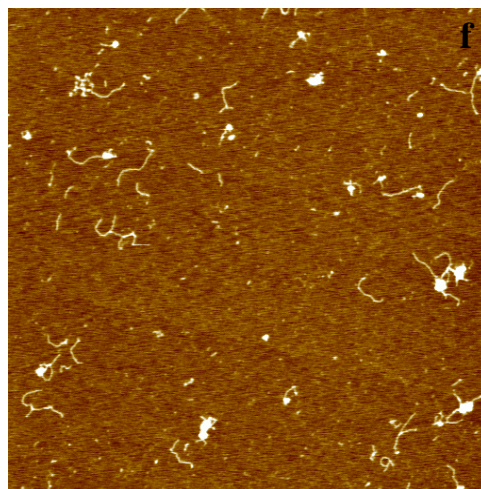
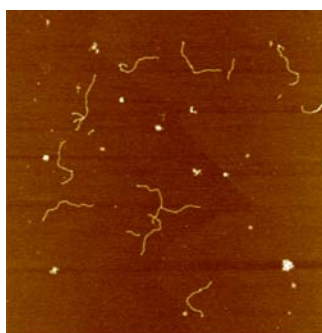
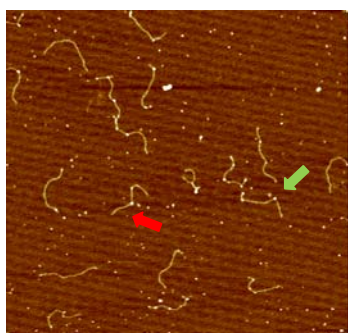
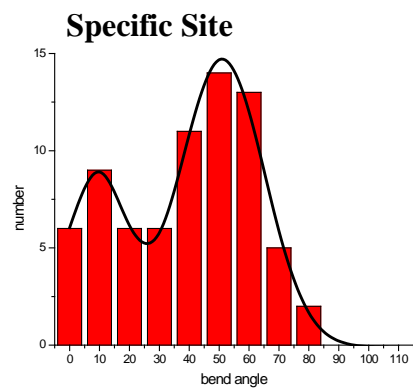


Figure 2.9 AFM images of RFC-DNA complexes in presence of ATP or ADP. a, RFC-1077Nick DNA complexes in low salt concentration buffer with ATP (NaCl, 50mM). The concentrations of RFC and DNA are 5nM and 10nM, respectively. The concentration of ATP is 5mM. The image size is $2\mu\text{m}\times 2\mu\text{m}$. b, RFC-1077Nick DNA complexes in high salt concentration buffer with ATP (NaCl, 100mM). c, RFC-1077Homo DNA complexes in the low salt concentration buffer with ATP (NaCl, 50mM). d Truncated RFC-1077Nick DNA complexes in low salt concentration buffer with ATP (NaCl, 50mM). Truncated RFC is mutated RFC without RFC1 ligase homology domain. e, RFC-1077Nick DNA complexes in the low salt concentration buffer with ADP (NaCl, 50nM). ADP is 5mM. f, RFC-1077Homo DNA complexes in the low salt concentration buffer with ADP (NaCl, 50nM). ADP is 5mM.

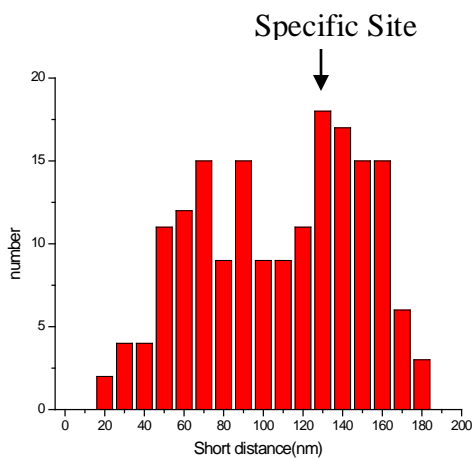
Nicked dsDNA+RFC+ATP γ S Homoduplex dsDNA dsDNA+RFC+ATP γ S



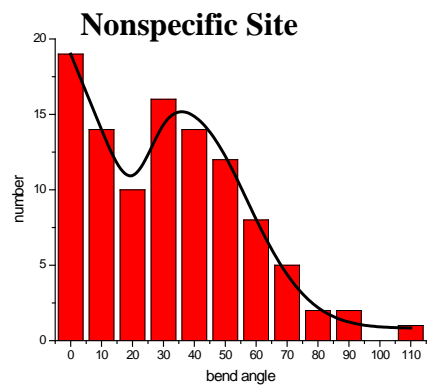
(a)



(c)

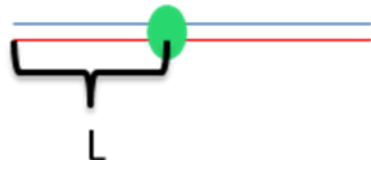


(b)

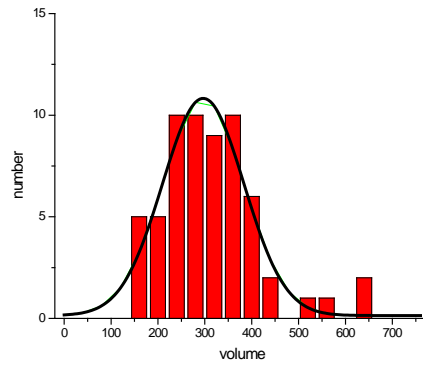


(d)

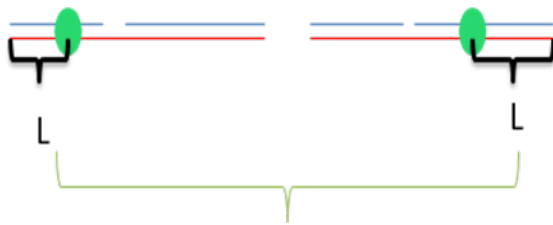
Figure 2.10 Representative AFM images and the bend angle and binding position distribution of RFC on nicked DNA in the presence of ATP γ S. a, Left: RFC-1077Nick DNA complexes in the presence of ATP γ S, Right: RFC-1077Homo DNA complex in the presence of ATP γ S. The concentration of RFC and DNA is 5nM and 10nM, respectively. The image size is 2 μ m \times 2 μ m. RFC and DNA are incubated in the high salt concentration buffer. The red arrow points out RFC bound at the nick. The green arrow points out nonspecific complexes. b, Histogram of the binding position of RFC on the 1077Nick DNA substrate. The contour DNA length is \sim 360 nm. The binding position is determined by measuring the distance from the center of RFC to the nearest DNA end. The arrow points to the RFC bound at the nick. The 1077Nick DNA substrate is 1077bp long and has a nick 477 nucleotides from the 3' end of the top strand. c, The distribution of bend angles induced by RFC bound at the nick with a double Gaussian fit with peaks at 0 $^\circ$ and 50 $^\circ$. d, The distribution of bend angles induced by nonspecific binding of RFC fit to a double Gaussian with peaks at 0 $^\circ$ and 36 $^\circ$.



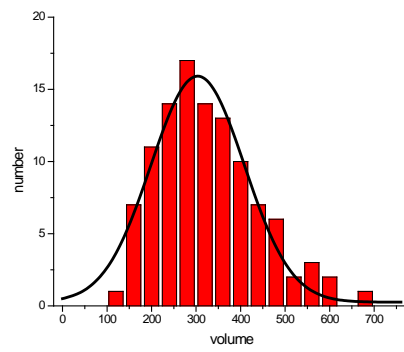
Specific Site



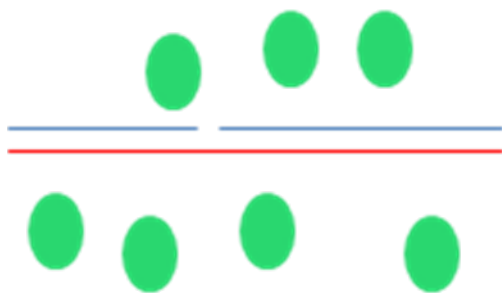
(a)



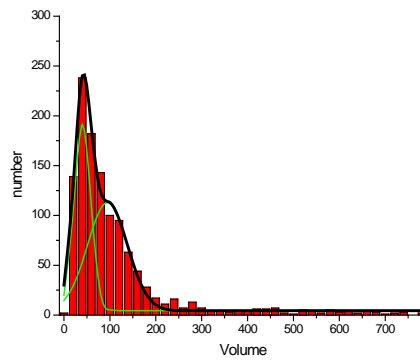
Nonspecific Site



(b)



Background



(c)

Figure 2.11 Schematic view of RFC binding to the 1077Nick dsDNA in the presence of ATP γ S and volume distribution of RFC bound at the nick (a), at nonspecific sites (b), and free on the surface (c). The distributions for RFC bound specifically (a) and nonspecifically (b) both fit to single Gaussian with peaks at $\sim 360\text{nm}^3$. The volume distribution of free RFC fits to a double Gaussian with peaks at 30nm^3 and 90nm^3 .

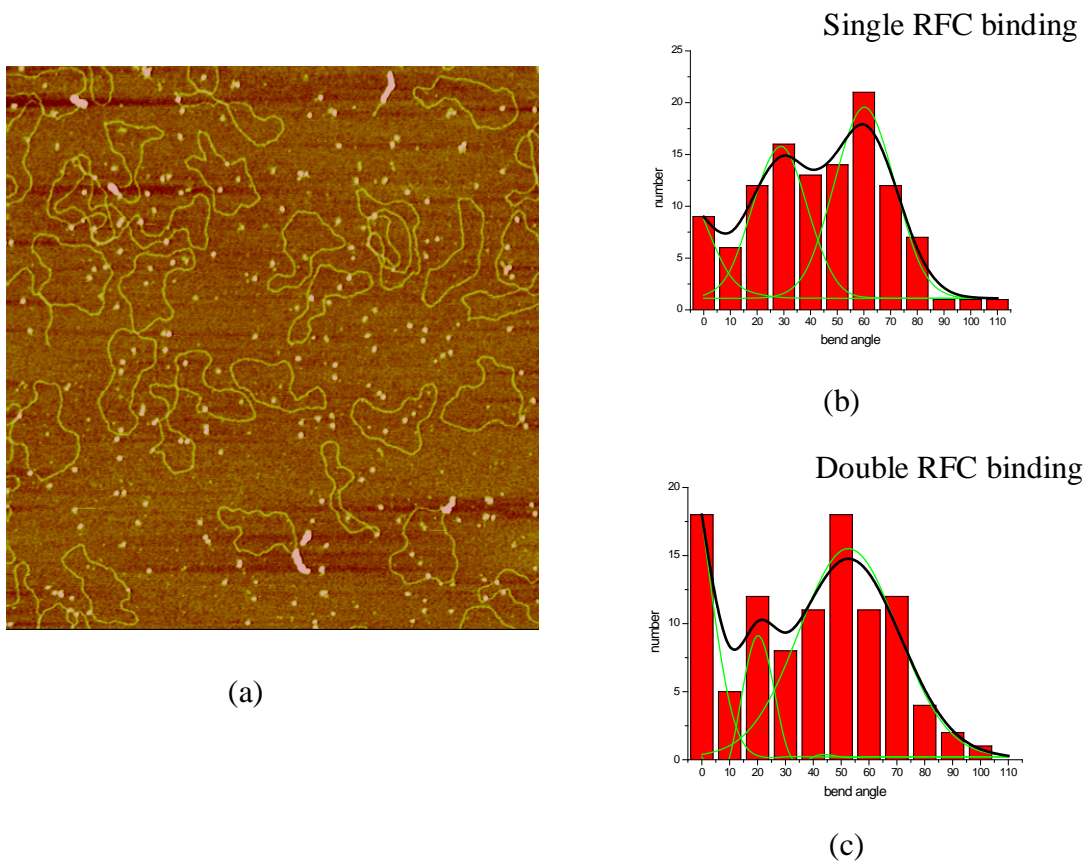
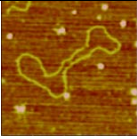
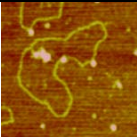
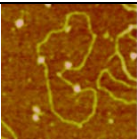
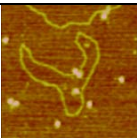
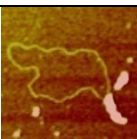
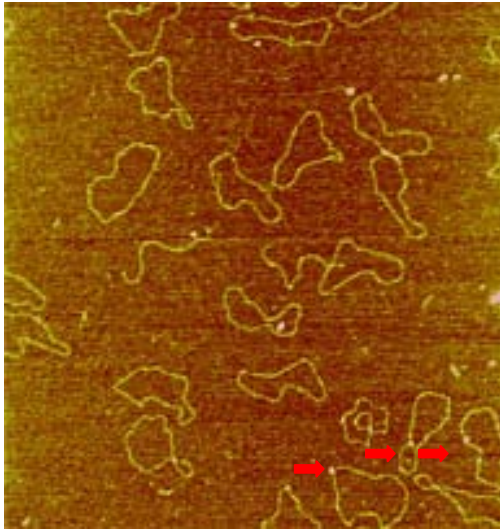


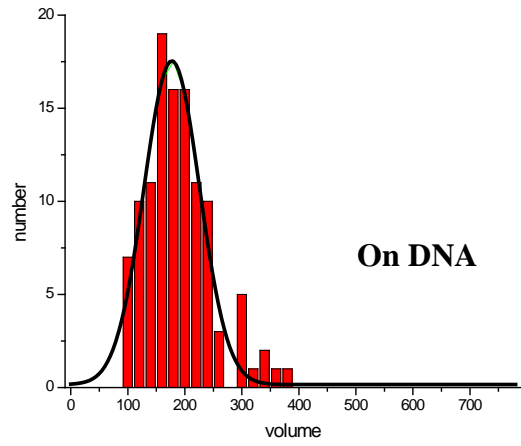
Figure 2.12: Representative AFM image and histograms of bend angle distribution of RFC on the pUC18Nick DNA substrate in the absence of nucleotide. a, RFC – pUC18Nick DNA complexes. The experiment is performed in high salt buffer. The concentrations of RFC and DNA are 5nM and 10nM, respectively. The size of image is $2\mu\text{m}\times 2\mu\text{m}$. b, Bend angle distribution for RFC singly bound is the curve are triple Gaussian fits with peaks at 0° , 28° , 60° . c, Bend angle distribution for double bound with triple RFC Gaussian fit with peaks at 0° , 22° , 52° .

Table 2.3 Statistical analysis of the conformation of RFC on the PUC18Nick DNA substrate in the absence of nucleotide

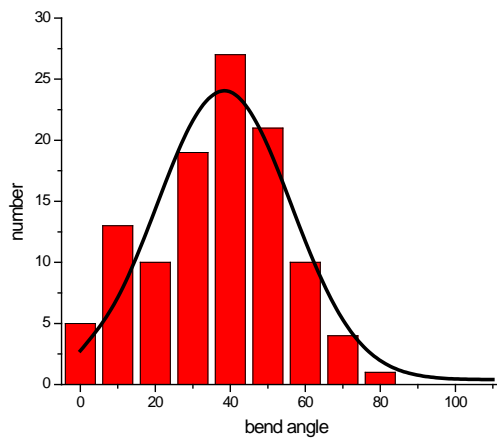
Single Binding		135	52.1%
Double Binding		59	22.8%
Multiple Binding		17	6.6%
Single Strand Complex		31	12%
RFC-DNA Track		17	6.6%



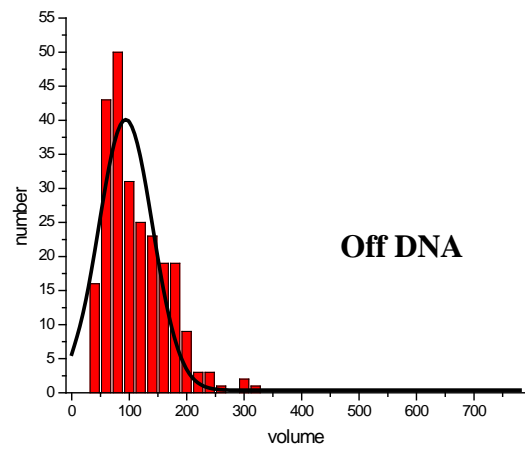
(a)



(c)

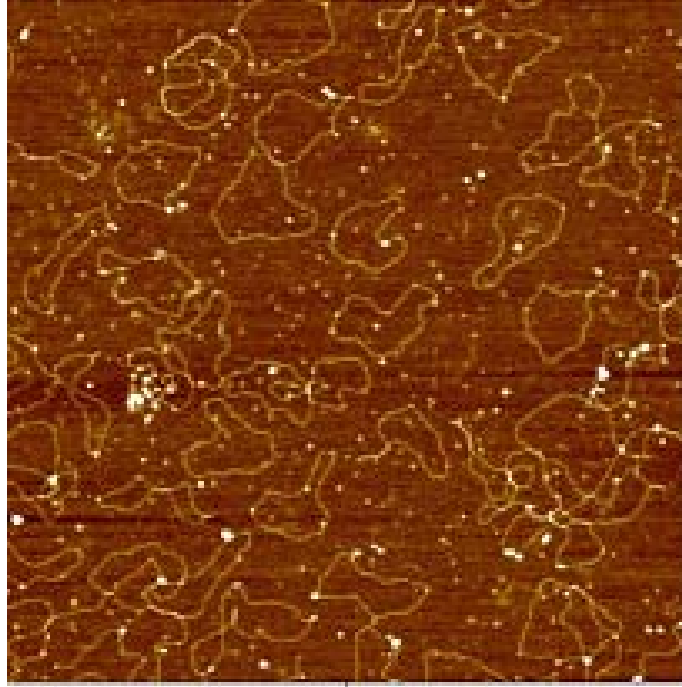


(b)

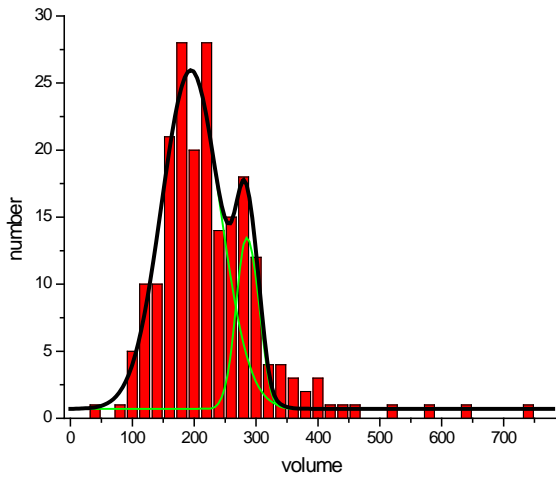


(d)

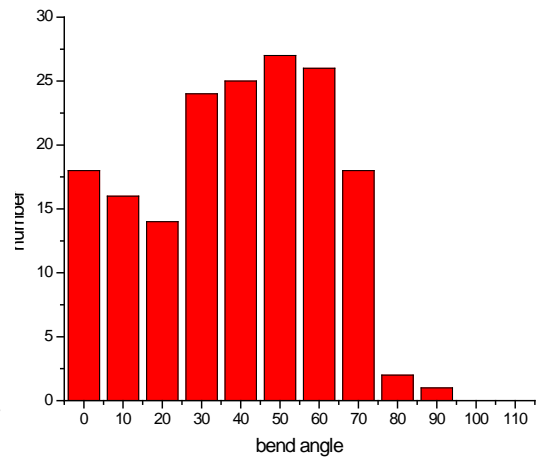
Figure 2.13 AFM image of RFC and pUC18Nick DNA deposited in the presence of ATP in high salt buffer and histograms of bend angles and volumes of the RFC - pUC18Nick DNA complexes. a, AFM image of RFC - pUC18Nick DNA complexes. RFC bound to DNA is pointed out by the red arrows. The size of image is $2\mu\text{m} \times 2\mu\text{m}$. The concentrations of RFC and DNA are 5nM and 10nM, respectively. b, Histogram of the bend angles of RFC on pUC18Nick DNA substrate. The bend angle distribution induced by RFC is Gaussian centered at $\sim 40^\circ$. c, Histogram of the volumes of RFC on pUC18Nick DNA substrate in the presence of ATP. The volume distribution of RFC on the DNA is Gaussian centered at $\sim 177\text{nm}^3$. d, Histogram of volumes of RFC distributed on the background in the presence of pUC18Nick DNA and ATP. The volume distribution of RFC on the background is Gaussian centered at $\sim 93\text{nm}^3$.



(a)



(b)



(c)

Figure 2.14 AFM image of RFC and PCNA with pUC18Nick DNA in the presence of ATP in high salt buffer. Histograms of the bend angles and the volumes of RFC – PCNA complexes on pUC18Nick DNA in the presence of ATP. a, AFM image of RFC – PCNA-pUC18Nick DNA complexes. The size of image is $2\mu\text{m}\times 2\mu\text{m}$. The concentrations of RFC, PCNA and DNA are 5nM, 5nM and 10nM, respectively. b, Histogram of the volumes of proteins bound to pUC18Nick DNA substrate. The volume distribution is fit by a double Gaussian centered at $\sim 194\text{nm}^3$ and 286nm^3 c, Histogram of the bend angles of RFC +PCNA on the pUC18Nick DNA substrate.

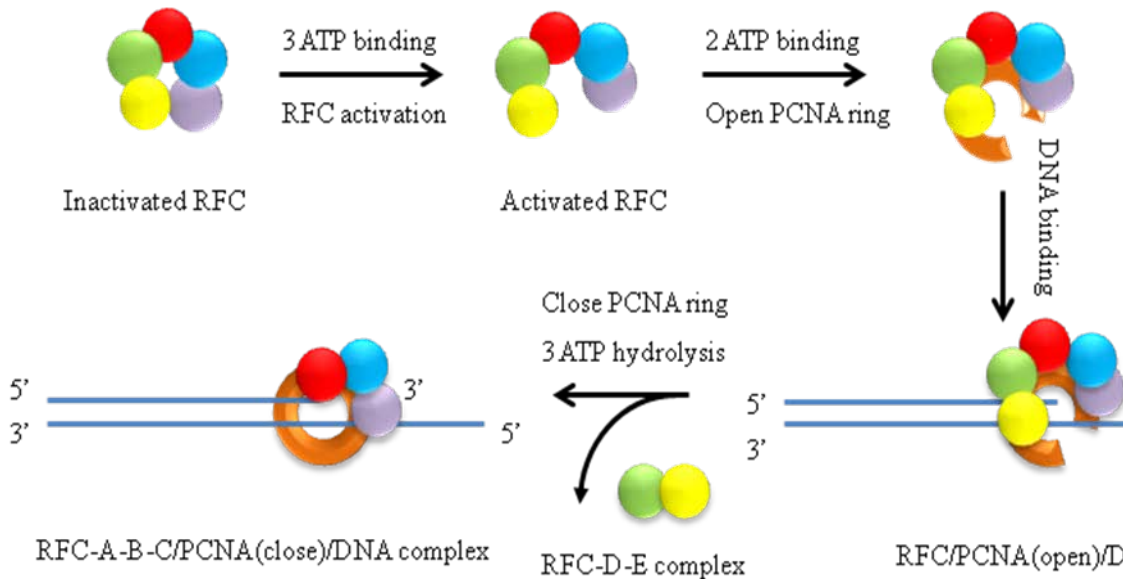


Figure 2.15 Schematic view of the proposed mechanism of RFC loading PCNA onto a primer-template DNA substrate. Combining previous research and our observations, we hypothesize a putative clamp loading mechanism. “Inactive” RFC is activated by binding with three ATPs. The “active” RFC is capable of binding with PCNA and another two ATPs. Interaction of PCNA with RFC introduces PCNA ring opening, and then RFC specifically loads PCNA at the primer-template junction. Three equivalent of ATP hydrolysis triggered by DNA binding ejects two subunits of RFC, RFC-D-E, from PCNA-DNA complex. RFC-A-B-C remains with PCNA at primer-template junction. Downstream PCNA-binding proteins, such as pol δ , promote the dissociation of the subcomplex from PCNA-DNA complex. RFC-D-E subcomplex may carry out the process of unloading PCNA from synthesized okazaki fragments.

Reference List

- Alley, S.C., Shier, V.K., Abel-Santos, E., Sexton, D.J., Soumilion, P., and Benkovic, S.J. (1999). Sliding clamp of the bacteriophage T4 polymerase has open and closed subunit interfaces in solution. *Biochemistry* 38, 7696-7709.
- Bertram, J.G., Bloom, L.B., Hingorani, M.M., Beechem, J.M., O'Donnell, M., and Goodman, M.F. (2000). Molecular mechanism and energetics of clamp assembly in *Escherichia coli*. The role of ATP hydrolysis when gamma complex loads beta on DNA. *J. Biol. Chem.* 275, 28413-28420.
- Bertram, J.G., Bloom, L.B., Turner, J., O'Donnell, M., Beechem, J.M., and Goodman, M.F. (1998). Pre-steady state analysis of the assembly of wild type and mutant circular clamps of *Escherichia coli* DNA polymerase III onto DNA. *J. Biol. Chem.* 273, 24564-24574.
- Bloom, L.B. (2006). Dynamics of loading the *Escherichia coli* DNA polymerase processivity clamp. *Crit. Rev. Biochem. Mol. Biol.* 41, 179-208.
- Bowman, G.D., O'Donnell, M., and Kuriyan, J. (2004). Structural analysis of a eukaryotic sliding DNA clamp-clamp loader complex. *Nature* 429, 724-730.
- Brar, S.S., Sacho, E.J., Tessmer, I., Croteau, D.L., Erie, D.A., and Diaz, M. (2008). Activation-induced deaminase, AID, is catalytically active as a monomer on single-stranded DNA. *DNA Repair (Amst)* 7, 77-87.
- Cai, J., Yao, N., Gibbs, E., Finkelstein, J., Phillips, B., O'Donnell, M., and Hurwitz, J. (1998). ATP hydrolysis catalyzed by human replication factor C requires participation of multiple subunits. *Proc. Natl. Acad. Sci. U. S. A.* 95, 11607-11612.
- Chen, S., Levin, M.K., Sakato, M., Zhou, Y., and Hingorani, M.M. (2009). Mechanism of ATP-driven PCNA clamp loading by *S. cerevisiae* RFC. *J. Mol. Biol.* 388, 431-442.
- Cullmann, G., Fien, K., Kobayashi, R., and Stillman, B. (1995). Characterization of the five replication factor C genes of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 15, 4661-4671.
- Davey, M.J., Jeruzalmi, D., Kuriyan, J., and O'Donnell, M. (2002). Motors and switches: AAA+ machines within the replisome. *Nat. Rev. Mol. Cell Biol.* 3, 826-835.
- Erzberger, J.P., and Berger, J.M. (2006). Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annu. Rev. Biophys. Biomol. Struct.* 35, 93-114.

Erzberger, J.P., Mott, M.L., and Berger, J.M. (2006). Structural basis for ATP-dependent DnaA assembly and replication-origin remodeling. *Nat. Struct. Mol. Biol.* *13*, 676-683.

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

Goedken, E.R., Kazmirski, S.L., Bowman, G.D., O'Donnell, M., and Kuriyan, J. (2005). Mapping the interaction of DNA with the Escherichia coli DNA polymerase clamp loader complex. *Nat. Struct. Mol. Biol.* *12*, 183-190.

Gomes, X.V., and Burgers, P.M. (2001). ATP utilization by yeast replication factor C. I. ATP-mediated interaction with DNA and with proliferating cell nuclear antigen. *J. Biol. Chem.* *276*, 34768-34775.

Gomes, X.V., Gary, S.L., and Burgers, P.M. (2000). Overproduction in Escherichia coli and characterization of yeast replication factor C lacking the ligase homology domain. *J. Biol. Chem.* *275*, 14541-14549.

Gomes, X.V., Schmidt, S.L., and Burgers, P.M. (2001). ATP utilization by yeast replication factor C. II. Multiple stepwise ATP binding events are required to load proliferating cell nuclear antigen onto primed DNA. *J. Biol. Chem.* *276*, 34776-34783.

Gu, L., Hong, Y., McCulloch, S., Watanabe, H., and Li, G.M. (1998). ATP-dependent interaction of human mismatch repair proteins and dual role of PCNA in mismatch repair. *Nucleic Acids Res.* *26*, 1173-1178.

Guenther, B., Onrust, R., Sali, A., O'Donnell, M., and Kuriyan, J. (1997). Crystal structure of the delta' subunit of the clamp-loader complex of E. coli DNA polymerase III. *Cell* *91*, 335-345.

Guo, S., Zhang, Y., Yuan, F., Gao, Y., Gu, L., Wong, I., and Li, G.M. (2006). Regulation of replication protein A functions in DNA mismatch repair by phosphorylation. *J. Biol. Chem.* *281*, 21607-21616.

Hingorani, M.M., and Coman, M.M. (2002). On the specificity of interaction between the Saccharomyces cerevisiae clamp loader replication factor C and primed DNA templates during DNA replication. *J. Biol. Chem.* *277*, 47213-47224.

Hingorani, M.M., and O'Donnell, M. (1998). ATP binding to the Escherichia coli clamp loader powers opening of the ring-shaped clamp of DNA polymerase III holoenzyme. *J. Biol. Chem.* *273*, 24550-24563.

Jeruzalmi, D., O'Donnell, M., and Kuriyan, J. (2001). Crystal structure of the processivity clamp loader gamma (gamma) complex of E. coli DNA polymerase III. *Cell* *106*, 429-441.

Jeruzalmi, D., Yurieva, O., Zhao, Y., Young, M., Stewart, J., Hingorani, M., O'Donnell, M., and Kuriyan, J. (2001). Mechanism of processivity clamp opening by the delta subunit wrench of the clamp loader complex of E. coli DNA polymerase III. *Cell* *106*, 417-428.

Johnson, A., Yao, N.Y., Bowman, G.D., Kuriyan, J., and O'Donnell, M. (2006). The replication factor C clamp loader requires arginine finger sensors to drive DNA binding and proliferating cell nuclear antigen loading. *J. Biol. Chem.* *281*, 35531-35543.

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

Kazmirski, S.L., Podobnik, M., Weitze, T.F., O'Donnell, M., and Kuriyan, J. (2004). Structural analysis of the inactive state of the Escherichia coli DNA polymerase clamp-loader complex. *Proc. Natl. Acad. Sci. U. S. A.* *101*, 16750-16755.

Kazmirski, S.L., Zhao, Y., Bowman, G.D., O'donnell, M., and Kuriyan, J. (2005). Out-of-plane motions in open sliding clamps: molecular dynamics simulations of eukaryotic and archaeal proliferating cell nuclear antigen. *Proc. Natl. Acad. Sci. U. S. A.* *102*, 13801-13806.

Lee, J.Y., and Yang, W. (2006). UvrD helicase unwinds DNA one base pair at a time by a two-part power stroke. *Cell* *127*, 1349-1360.

Majka, J., and Burgers, P.M. (2004). The PCNA-RFC families of DNA clamps and clamp loaders. *Prog. Nucleic Acid Res. Mol. Biol.* *78*, 227-260.

Maradeo, M.E., and Skibbens, R.V. (2009). The Elg1-RFC clamp-loading complex performs a role in sister chromatid cohesion. *PLoS ONE* *4*, e4707.

Miyata, T., Suzuki, H., Oyama, T., Mayanagi, K., Ishino, Y., and Morikawa, K. (2005). Open clamp structure in the clamp-loading complex visualized by electron microscopic image analysis. *Proc. Natl. Acad. Sci. U. S. A.* *102*, 13795-13800.

Moldovan, G.L., Pfander, B., and Jentsch, S. (2006). PCNA controls establishment of sister chromatid cohesion during S phase. *Mol. Cell* *23*, 723-732.

Naktinis, V., Onrust, R., Fang, L., and O'Donnell, M. (1995). Assembly of a chromosomal replication machine: two DNA polymerases, a clamp loader, and sliding

clamps in one holoenzyme particle. II. Intermediate complex between the clamp loader and its clamp. *J. Biol. Chem.* 270, 13358-13365.

Naktinis, V., Turner, J., and O'Donnell, M. (1996). A molecular switch in a replication machine defined by an internal competition for protein rings. *Cell* 84, 137-145.

Parnas, O., Zipin-Roitman, A., Mazor, Y., Liefshitz, B., Ben-Aroya, S., and Kupiec, M. (2009). The ELG1 clamp loader plays a role in sister chromatid cohesion. *PLoS ONE* 4, e5497.

Pietroni, P., and von Hippel, P.H. (2008). Multiple ATP binding is required to stabilize the "activated" (clamp open) clamp loader of the T4 DNA replication complex. *J. Biol. Chem.* 283, 28338-28353.

Pietroni, P., Young, M.C., Latham, G.J., and von Hippel, P.H. (2001). Dissection of the ATP-driven reaction cycle of the bacteriophage T4 DNA replication processivity clamp loading system. *J. Mol. Biol.* 309, 869-891.

Podust, V.N., Tiwari, N., Stephan, S., and Fanning, E. (1998). Replication factor C disengages from proliferating cell nuclear antigen (PCNA) upon sliding clamp formation, and PCNA itself tethers DNA polymerase delta to DNA. *J. Biol. Chem.* 273, 31992-31999.

Ratcliff, G.C., and Erie, D.A. (2001). A novel single-molecule study to determine protein--protein association constants. *J. Am. Chem. Soc.* 123, 5632-5635.

Schmidt, S.L., Gomes, X.V., and Burgers, P.M. (2001). ATP utilization by yeast replication factor C. III. The ATP-binding domains of Rfc2, Rfc3, and Rfc4 are essential for DNA recognition and clamp loading. *J. Biol. Chem.* 276, 34784-34791.

Sexton, D.J., Kaboord, B.F., Berdis, A.J., Carver, T.E., and Benkovic, S.J. (1998). Dissecting the order of bacteriophage T4 DNA polymerase holoenzyme assembly. *Biochemistry* 37, 7749-7756.

Simonetta, K.R., Kazmirski, S.L., Goedken, E.R., Cantor, A.J., Kelch, B.A., McNally, R., Seyedin, S.N., Makino, D.L., O'Donnell, M., and Kuriyan, J. (2009). The mechanism of ATP-dependent primer-template recognition by a clamp loader complex. *Cell* 137, 659-671.

Snyder, A.K., Williams, C.R., Johnson, A., O'Donnell, M., and Bloom, L.B. (2004). Mechanism of loading the Escherichia coli DNA polymerase III sliding clamp: II. Uncoupling the beta and DNA binding activities of the gamma complex. *J. Biol. Chem.* 279, 4386-4393.

Stewart, J., Hingorani, M.M., Kelman, Z., and O'Donnell, M. (2001). Mechanism of beta clamp opening by the delta subunit of Escherichia coli DNA polymerase III holoenzyme. *J. Biol. Chem.* 276, 19182-19189.

Stukenberg, P.T., Studwell-Vaughan, P.S., and O'Donnell, M. (1991). Mechanism of the sliding beta-clamp of DNA polymerase III holoenzyme. *J. Biol. Chem.* 266, 11328-11334.

Trakselis, M.A., Berdis, A.J., and Benkovic, S.J. (2003). Examination of the role of the clamp-loader and ATP hydrolysis in the formation of the bacteriophage T4 polymerase holoenzyme. *J. Mol. Biol.* 326, 435-451.

Turner, J., Hingorani, M.M., Kelman, Z., and O'Donnell, M. (1999). The internal workings of a DNA polymerase clamp-loading machine. *EMBO J.* 18, 771-783.

Uhlmann, F., Cai, J., Gibbs, E., O'Donnell, M., and Hurwitz, J. (1997). Deletion analysis of the large subunit p140 in human replication factor C reveals regions required for complex formation and replication activities. *J. Biol. Chem.* 272, 10058-10064.

Umar, A., Buermeier, A.B., Simon, J.A., Thomas, D.C., Clark, A.B., Liskay, R.M., and Kunkel, T.A. (1996). Requirement for PCNA in DNA mismatch repair at a step preceding DNA resynthesis. *Cell* 87, 65-73.

Williams, C.R., Snyder, A.K., Kuzmic, P., O'Donnell, M., and Bloom, L.B. (2004). Mechanism of loading the Escherichia coli DNA polymerase III sliding clamp: I. Two distinct activities for individual ATP sites in the gamma complex. *J. Biol. Chem.* 279, 4376-4385.

Xiao, H., Naktinis, V., and O'Donnell, M. (1995). Assembly of a chromosomal replication machine: two DNA polymerases, a clamp loader, and sliding clamps in one holoenzyme particle. IV. ATP-binding site mutants identify the clamp loader. *J. Biol. Chem.* 270, 13378-13383.

Xue, Y., Ratcliff, G.C., Wang, H., Davis-Searles, P.R., Gray, M.D., Erie, D.A., and Redinbo, M.R. (2002). A minimal exonuclease domain of WRN forms a hexamer on DNA and possesses both 3'- 5' exonuclease and 5'-protruding strand endonuclease activities. *Biochemistry* 41, 2901-2912.

Yang, Y., Wang, H., and Erie, D.A. (2003). Quantitative characterization of biomolecular assemblies and interactions using atomic force microscopy. *Methods* 29, 175-187.

Yao, N., Coryell, L., Zhang, D., Georgescu, R.E., Finkelstein, J., Coman, M.M., Hingorani, M.M., and O'Donnell, M. (2003). Replication factor C clamp loader subunit

arrangement within the circular pentamer and its attachment points to proliferating cell nuclear antigen. *J. Biol. Chem.* *278*, 50744-50753.

Yao, N., Turner, J., Kelman, Z., Stukenberg, P.T., Dean, F., Shechter, D., Pan, Z.Q., Hurwitz, J., and O'Donnell, M. (1996). Clamp loading, unloading and intrinsic stability of the PCNA, beta and gp45 sliding clamps of human, *E. coli* and T4 replicases. *Genes Cells* *1*, 101-113.

Yao, N.Y., Johnson, A., Bowman, G.D., Kuriyan, J., and O'Donnell, M. (2006). Mechanism of proliferating cell nuclear antigen clamp opening by replication factor C. *J. Biol. Chem.* *281*, 17528-17539.

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

Zhuang, Z., Berdis, A.J., and Benkovic, S.J. (2006). An alternative clamp loading pathway via the T4 clamp loader gp44/62-DNA complex. *Biochemistry* *45*, 7976-7989.

Zhuang, Z., Yoder, B.L., Burgers, P.M., and Benkovic, S.J. (2006). The structure of a ring-opened proliferating cell nuclear antigen-replication factor C complex revealed by fluorescence energy transfer. *Proc. Natl. Acad. Sci. U. S. A.* *103*, 2546-2551.

CHAPTER THREE: AFM STUDY OF MISMATCH RECOGNITION BY EUKARYOTIC MUTS HOMOLOGS

3.1 Introduction

Mismatched nucleotides arise from polymerase misincorporation errors, recombination between heteroallelic parental DNAs, and chemical or physical DNA damage. Postreplicative mismatch repair (MMR) corrects these errors before they lead to genomic instability. This repair pathway increases the fidelity of DNA replication by up to 1000 fold. Inactivation of the human mismatch repair system is the cause of hereditary nonpolyposis colon cancer (HNPCC) and has been implicated in the development of a subset of sporadic tumors that occur in a variety of tissues (Iyer, Pluciennik, et al, 2006; Kunkel, and Erie, 2005; Li, 2008; Modrich, and Lahue, 1996).

In *E.coli*, MutS, MutL and MutH promote strand-specific repair by taking advantage of the transiently unmethylated state of the newly synthesized strand. MutS and MutL function as dimers and have intrinsic ATPase activities that are essential for MMR. MMR is initiated by the binding of MutS to either a mismatch or short insertion/deletion loop (IDL). ATP binding introduces a conformational change of MutS and promotes its interaction with MutL. The MutS-MutL heterocomplex activates a latent endonuclease activity of MutH in an ATP-dependent manner, which incises the newly synthesized

strand at a nearest hemimethylated d (GATC) site. The strand specific nick provides entry for other proteins that complete the downstream steps of the pathway (Iyer, Pluciennik, et al, 2006; Kunkel, and Erie, 2005; Li, 2008; Varlet, Canard, et al, 1996).

In *E. coli*, base-base and insertion/deletion mismatches of up to 4 nt which arise during DNA replication are recognized by MutS. The MutS structure reveals a disc-shaped protein with two channels. The DNA is bound into the lower channel and kinked with an $\sim 60^\circ$ bend angle (Figure 1.3). The crystal structure of MutS reveals that the MutS homodimer binds to mismatch DNA in an asymmetric fashion. While each MutS subunit contains two DNA binding domains, I and IV, only domain I from one subunit directly interacts with mismatch mismatched base. Specific interactions between the mispaired base and the Phe-X-Glu mismatch recognition motif are identified. The mismatch base is rotated into the minor grooves by $\sim 3\text{\AA}$ and stacks with the phenylalanine. The glutamic acid forms a hydrogen bond to the same base. Most of the other contacts from the two subunits are to the backbone and are, therefore, DNA sequence nonspecific, as expected given the need to repair replication errors in a variety of different sequence contexts. These interactions narrow the major groove, and expand the minor groove where the phenylalanine interacts with the DNA mispair (Lamers, Perrakis, et al, 2000; Natrajan, Lamers, et al, 2003; Obmolova, Ban, et al, 2000).

In eukaryotes, two heterodimeric MutS homologs, Msh2Msh6 and Msh2Msh3 have been identified. These two complexes recognize different types of mismatches; Msh2Msh6 primarily acts in a pathway that repairs base-base and single

insertion/deletion mismatches, while Msh2Msh3 acts mainly to repair insertion/deletion loop mismatches 2 to 13 nt in size (Alani, 1996; Habraken, Sung, et al, 1996; Harrington, and Kolodner, 2007; Lee, Surtees, and Alani, 2007; Palombo, Iaccarino, et al, 1996). The Phe-X-Glu mismatch-recognition motif is conserved in Msh6 but is not in Msh2 or Msh3. In the recent crystal structures of Msh2Msh6 (the first 340 amino acids of Msh6 are truncated) bound with various mismatched DNA substrates, domain I of Msh6 contributes most of contacts with DNA, including non-specific contacts with the DNA backbone made by the N-terminal residues (360-398), and the specific contacts are made with the mispaired bases by the conserved Phe-X-Glu motif. By contrast, domain I of Msh2 makes only a few contacts with DNA and is dispensable for mismatch repair by Msh2Msh6. Msh2Msh6 bends G-T mismatched DNA in to an angle of $\sim 45^\circ$ (Warren, Pohlhaus, et al, 2007). Although the crystal structure of Msh2Msh3 is unknown, domain I of Msh3 appears to be important for IDL binding specificity and for suppressing nonspecific binding, while domain I of Msh2 contributes a nonspecific binding activity. Deletion of domain I of Msh2 results in the loss Msh2Msh3-mediated MMR and recombination function (Lee, Surtees, and Alani, 2007). The Phe-X-Glu motif is not conserved in Msh3, instead two lysines in yeast and a lysine and an arginine in human Msh3 are responsible for the binding of Msh2Msh3 to IDL DNA. Msh2Msh3 involves protein-DNA contacts that appear very different from those required for Msh2Msh6 mismatch binding.

It has been inferred that MutS homologs initially bind to DNA through three - dimension (3D) random collisions, and scan DNA for lesions or via 1D diffusion after

initial binding (Drummond, Li, et al, 1995). The question is how MutS homologs specifically recognize various DNA lesions and initiate repair pathways. It has been shown that affinities of MutS homologs are ~10 fold higher for mismatched than perfectly mismatched DNA (Gradia, Acharya, and Fishel, 2000; Schofield, Brownnewell, et al, 2001). Affinities vary with the composition of the mismatch and the local sequence context. The relatively low binding affinity difference between mismatched and matched DNA substrates may be caused, in part, by the small free energy differences between perfectly paired and mismatched base pairs (estimated to be in the range of 1-3 kcal/mol for internal pairs), which would correspond to only a 10-fold discrimination between the perfectly paired and mismatched bases (Echols, and Goodman, 1991; Loeb, and Kunkel, 1982). In addition, inspection the crystal structures of *E. coli* MutS with G-T, C-A, A-A, G-G, unpaired T heteroduplex, and Taq MutS with an unpaired T heteoduplex reveals a similar mismatch-recognition pathway for the different mismatches (Lamers, Perrakis, et al, 2000; Natrajan, Lamers, et al, 2003; Obmolova, Ban, et al, 2000). Human Msh2Msh6 bound with a G-T mispair, unpaired T, O6-methyl G-T mispair, and G-dU mispair heteroduplexes also exhibit very similar structures (Warren, Pohlhaus, et al, 2007).

Insight into the mechanism of mismatch-recognition and specificity comes from a recent study using AFM to directly visualize MutS bound to mismatch and to homoduplex DNA (Wang, Yang, et al, 2003). This study demonstrates that MutS adopts a single population of conformations, in which the DNA is bent, when bound to a homoduplex sites, but two populations of conformation, bent and unbent, when bound at the mismatch site. These results suggest the unique unbent complex bound at mismatched

sites is a signal of mispaired bases being recognized, and that the bent complex at mismatched sites may be an intermediate in the formation of the unbent complex.

In this study, we adopt the same AFM method to study the mechanism of mismatch-recognition and specificity of eukaryotic MutS homologs, yeast Msh2Msh3 and yeast Msh2Msh6. AFM is a single molecule technique which can resolve individual protein-DNA complexes. An ensemble of MutS-DNA complexes will provide snap-shots of the dynamic process during mismatch recognition. Furthermore, using AFM, the distribution of conformations within a complex population of molecules can be characterized (Bustamante, and Rivetti, 1996).

Our results demonstrate that yeast Msh2Msh3 and Msh2Msh6 bind different DNA substrates in a similar manner and exhibit low binding specificity to the mismatch. Intriguingly, unlike *E. coli* MutS or *Taq* MutS, two populations of conformations are observed both at mismatch sites and homoduplex sites. This bimodal distribution of conformations is independent on the presence or type of mismatch. Even though the structure and contacts with DNA of Msh-2Msh3 appears different from Msh2Msh6, our results suggest that these differences maybe not affect the protein-DNA binding modes. Taken together, a higher-grade mismatch-recognition mechanism is suggested for eukaryotic MutS homologs.

3.2 Materials and Methods

Msh2Msh3 and Msh2Msh6 proteins. Yeast Msh2Msh3 and Msh2Msh6 was provided from collaborator, Manju M. Hingorani Lab.

DNA substrates for AFM studies. The 782Homo, 783Tbulge and 982GT are purified as described previously (Wang, Yang, et al, 2003). The 1124Tbulge and 1125TTbulge are created by annealing a single-strand DNA (ssDNA) with a complementary DNA fragment, with the exception of one or two bases. A plasmid (pBluescript II) containing a phage origin of replication was chosen for the purpose to exacting single-stranded DNA. pBluescript contains the f1 phage origin of replication and can be used to produce ssDNA using a helper phage (VCSM13). ssDNA is isolated by precipitating in 2.5M NaCl/20% PEG8000 buffer and purified by phenol/chloroform extraction and ethanol precipitation. A point mutation is made in the XbaI site on pBluscript II plasmid to make complementary mismatched segment. The mutation knocks out the ability of XbaI to cut at the site which provides a method for screening for the mutation. A 6X concentration of ssDNA are annealed to 1X concentration of PCR fragments. To make a single T-bulge mismatch, a mutated PCR fragment with one T insertion is annealed to the WT ssDNA. To makde a double T- bulge mismatch a mutated PCR fragment with two T insertions is annealed to the WT ssDNA. The annealing products are cut by XbaI, BseyI and NaeI to make 1124 or 1125 DNA fragment and purified by agarose gel.

Atomic force microscopy. Protein-DNA complexes are formed by incubating ~5nM of yeast MutS α or MuS β with ~10nM of hetero- homoduplex DNA substrates for 1-5 minutes at room temperature in binding buffer (25mM Hepes, pH 8.1, 120mM NaCl, 0.1

mM EDTA, and 1.0 mM DTT, 5mM MgCl_2 , 10% glycerol) in a total volume of 20 μL . The reaction is deposited onto freshly cleaved mica (Spruce Pine Mica Company) at room temperature. The mica is quickly washed with HPLC grade water, blotted dry, and then dried under stream of nitrogen. The images are captured in air with the Nanoscope IIIa (Digital Instruments) microscope in tapping mode. Pointprobe^R tapping mode silicon probes (Molecular Imaging Corporation) with spring constants of ~ 50 N/m and resonance frequencies ~ 170 KHz are used for all imaging. Images are collected at a speed of 3-4 Hz and resolution of 512 \times 512 pixels.

Image analysis. DNA length and bend angles are measured using Nanoscope IIIa software. The angle, α , at the intersection of two extrapolated DNA arms extruding from MutS homologs is measured. The bend angle Θ , is $180 - \alpha$. The length of DNA is determined by dividing the contour length into connected straight segments and then measuring the length of each segment and summing all lengths up. The positions of the RFC binding sites on DNA templates are determined by measuring the length of the DNA from intersection of the two extrapolated DNA arms to each end. The binding position is defined as the ratio of the length of the shorter DNA tract divided by the total contour length. Classification of specific complexes was based on the consideration of protein size measured from AFM images and uncertainty in the measurement of DNA contour length. Only those proteins-DNA complexes with centers positioned within one standard deviation of the expected mismatch position are categorized as specific complexes. We do not end-label the DNA to unequivocally identify the DNA ends, which means that some nonspecific complexes will be counted as specific complexes, but

not *vice versa*. Volume analysis is performed as described previously (Erie, Yang, et al, 1994; Ratcliff, and Erie, 2001; Yang, Wang, and Erie, 2003). The program Kaleidagraph and OriginPro are used to generate statistical plots.

3.3 Results

AFM volume analysis of Msh2Msh3

A representative AFM image of Msh2Msh3 is shown in Figure 3.1 (a). Msh2Msh3 is composed of two different proteins with similar molecular weight: Msh2 of 107kDa and Msh3 of 117kDa. To examine the oligomerization states of Msh2Msh3 under the conditions of our experiment, we analyzed Msh2Msh3 in the absence of nucleotide using a volume analysis method described previously (Ratcliff, and Erie, 2001; Yang, Wang, and Erie, 2003). The protein volume measured by AFM is linearly related with the molecular weight of protein ($AFM\ Volume = 1.2\ (MW) - 14.7$). The peaks at $\sim 110\text{nm}^3$ and $\sim 290\text{nm}^3$ in the volume histogram (Figure 3.1, b) are consistent with the AFM volume of Msh2 or Msh3 and Msh2Msh3, respectively. The analysis reveals $\sim 60\%$ of Msh2Msh3 complexes exist as heterodimers. Given that the dissociation equilibrium ($Msh2Msh3 \leftrightarrow Msh2 + Msh3$) and the concentration of Msh2Msh3 in the reaction ($\sim 20\text{nM}$), we estimate the dissociation constant, K_d , to be $\sim 5.3\text{nM}$ under our experiment conditions.

Yeast Msh2Msh3 bound to various DNA substrates exhibits two conformations: bent and unbent despite of the difference between DNA substrates

In this study, we employed two different DNA substrates containing uniquely located mismatches (783Tbulge and 1125TTbulge; where the number indicates the length of the fragment, and the letters indicate the type of mismatch) and 782Homo homoduplex DNA (Figure 3.2). The 783Tbulge DNA has a single unpaired T positioned ~27% of the distance from one end, and 1125TTbulge DNA has two unpaired T ~33% from the end. 782Homo substrate is the same as 783Tbulge except that there is no mismatch. A previous AFM study indicates that a single base bulge does not introduced significant intrinsic bending of DNA compared with homoduplex DNA (Wang, Yang, et al, 2003). Msh2Msh3 acts mainly to repair insertion/deletion loops of 2 to 13 nucleotides in size and DNA flap structures predicted to form during genetic recombination, and it is also redundant with Msh2Msh6 in the repair of small IDLs (Habraken, Sung, et al, 1996; Harrington, and Kolodner, 2007). Therefore, in this study, we use 782Homo as Msh2Msh3 nonspecific recognition substrate, 783Tbulge as Msh2Msh3 redundant substrate with Msh2Msh6, and 1125TTbulge as Msh2-Msh3 specific recognition substrate.

-Yeast Msh2Msh3 binding specificity analysis-

Representative AFM images of yeast Msh2Msh3 deposited in the presence of 782Homo, 783Tbulge and 1125TTbulge DNA substrates are shown in Figure 3.3 (a), 3.4 (a), and 3.5 (a). To determine the binding positions of Msh2Msh3 on the DNA substrate, we measured the length of the DNA arms extruded from Msh2Msh3 complex. The binding position is defined as the ratio of the length of the shorter DNA arm divided by

the total contour length. The distributions of Msh2Msh3 on 782Homo, 783Tbulge DNA and 1125TTbulge DNA substrates are random (Figure 3.3 b, 3.4 b, and 3.5, b). These results suggest Msh2Msh3 has no significant binding preference to the single T-bulge site and double T-bulge site.

Because we know the positions of the mismatches on the DNA, we can discriminate Msh2Msh3 bound at a mismatch (specific complex) and Msh2Msh3 bound at a homoduplex site (nonspecific complex). In the presence of 783Tbulge DNA (Figure 3.4 b), we define Msh2Msh3 bound in the range of 24%-33% distance from the nearest DNA ends as specific complexes. In the presence of 1125TTbulge DNA (Figure 3.5, b), we define Msh2Msh3 complexes in the range of 30%-38% distance from the nearest DNA end as specific complexes. We observed ~ 31 of 102 Msh2Msh3 complexes bind specifically at T-bulge mismatch site, and ~44 of 145 Msh2Msh3 complexes bound at TT-bulge mismatch site (If Msh2Msh3 is bound at multiple position on the DNA, the complexes are included in binding position analysis, but not in the DNA bend angle analysis). These results suggest Msh2Msh3 exhibits weak binding specificity to both the single T-bulge site and the double T-bulge site.

-Yeast Msh2Msh3-introduced DNA bend angle analysis-

Msh2Msh3 introduces significant DNA bending both with specific and nonspecific DNA substrates. Figure 3.3(c) and 3.4(c) show the distribution of bend angles introduced by Msh2Msh3 bound to 782Homo and 783T-bulge DNA. A previous AFM study showed that the bend angle distribution of free DNA without protein binding is a half-Gaussian

distribution centered at 0° with 35° breadth (Wang, Yang, et al, 2003). Inspection of the distributions of Msh2Msh3 bend angles for both homoduplex and T-bulge DNA substrates reveal a broad distribution with a bend angle centered at $\sim 60^\circ$. Intriguingly, we also find a population of complexes that is not bent. The bimodal bend angle distributions are fit well by a double Gaussian distribution, but not single Gaussian distribution. Approximately 15% of unbent complexes and 85% of bent complexes are found in both distributions (The drop at $\sim 50^\circ$ in the bend angle distribution of 782Homo and 783Tbulge is the error of bins, which will disappear upon changing the bin size). Interestingly, analogous conformation distributions are also observed for Msh2Msh3 bound at mismatch and homoduplex sites on the 1125TTbulge DNA (Figure 3.5 (c) and (d)). A double Gaussian fit reveals two peaks centered at $\sim 0^\circ$ and $\sim 40^\circ$ for both Msh2Msh3 bound at mismatch and at homdupex site. About 30% of Msh2Msh3 specific complexes are unbent and 20% of the nonspecific complexes are unbent. The ratio of the populations of the unbent state to the bent state represents the equilibrium between these two states. The ratio ranges from 0.2 for Msh2Msh3 bound to homoduplex or T-bulge DNA substrate to 0.4 for Msh2Msh3 bound at TT-bulge mismatch, indicating that the unbent state is slightly less stable than the bent states.

In summary, these AFM data indicate that yeast Msh2Msh3 binds various DNA substrates in the same bimodal mode, imposing two conformations on the DNA, bent and unbent. These two Msh2Msh3-induced DNA conformations are independent on the presence of mismatch (homoduplex vs. heteroduplex), the type of mismatch (T-bulge vs. TT-bulge) and binding position (specific complex vs. nonspecific complex). In contrast

to *E. coli* MutS and *Taq* MutS (Wang, Yang, et al, 2003), Msh2Msh3 exhibits low binding specificity to mismatches. Taking all the results together, it appears Msh2Msh3 adopts a mismatch-recognition mechanism that is different from that of *E. coli* MutS and *Taq* MutS.

To examine if the bimodal binding mode is conserved in other eukaryotic species, we examined human Msh2Msh3 in the presence of 1125TTbulge DNA substrate (Figure 3.6). Human Msh2Msh3 appears to be distributed randomly on the TT-bulge DNA, suggesting low binding specificity. The distributions of DNA bend angles are double Gaussians with peaks at $\sim 0^\circ$ and $\sim 45^\circ$ for both specific and nonspecific complexes. About 20% of specific complexes and 25% of nonspecific complexes are unbent, suggesting an equilibrium constant of ~ 0.3 . A broader bend angle distribution of the bent states compared with the unbent states is seen for both specific and nonspecific complexes. These results suggest human Msh2Msh3 appears to recognize IDLs similarly to yeast Msh2Msh3.

Yeast Msh2Msh6 bound to various DNA substrates exhibits two conformations: bent and unbent despite of the difference between DNA substrates

In this study, we also employed two different DNA substrates containing uniquely located mismatches (982G/T and 1124Tbulge; where the number indicates the length of the fragment, and the letters indicate the type of mismatch) (Figure 3.2), and a 400bp homoduplex DNA substrate made from PCR fragments of pUC18. The 982GT has G/T

mismatch posited 42% from one end, and 1124Tbulge has an unpaired T 33% from one end.

-Yeast Msh2Msh6 binding specificity analysis-

Representative AFM images of yeast Msh2Msh6 bound to homoduplex, G/T mismatch and T-bulge DNA substrates are presented in Figure 3.7 (a), 3.8(a) and 3.9(a). Msh2Msh6 is uniformly distributed on the homoduplex DNA (nonspecific DNA) (Figure 3.7, b). In the presence of mismatch-recognition substrates (specific DNA), 982GT DNA and 1124Tbulge DNA, Msh2Msh6 exhibits low binding specificity to a T-bulge or G/T mismatch. About 100 of 253 Msh2Msh6 complexes specifically bind to G/T mismatch, and ~38 of 94 Msh2Msh6 complexes specifically bind to single T-bulge mismatch (Figure 3.8 b; Figure 3.9 b). Here, we defined Msh2Msh6 complexes bound to 982GT or 1124Tbulge DNA in the range of 40%-48%, 30%-38% respectively as specific complexes (The Msh2Msh6-DNA complexes with multiple Msh2Msh6 bound are counted in the binding specificity analysis, but not in the bend angle analysis).

-Yeast Msh2Msh6 bend angle analysis-

Similar to yeast Msh2Msh3, two populations of Msh2Msh6- DNA complexes, bent and unbent, are observed in the presence of homoduplex, G/T mismatch and T-bulge DNA substrates. In the presence of homoduplex DNA, the distribution of DNA bend angles is a double Gaussians distribution with peaks at $\sim 0^\circ$ and $\sim 40^\circ$ (Figure 3.7, c). The portion of unbent complex is $\sim 20\%$, indicating an equilibrium constant of ~ 0.25 . The

breadth of distribution for bent state is significantly broader than that of distribution of unbent state.

Intriguingly, Msh2Msh6 introduces similar distribution pattern of DNA bend angles at mismatch and homoduplex sites (Figure 3.8, c, d; Figure 3.9 c, d). The distribution are fit well to a double Gaussian distributions, with peaks centered $\sim 0^\circ$ and $\sim 50^\circ$ for both specific and nonspecific complexes. In addition, these distributions do not appear to depend on the types of mismatch, because the DNA bend angles introduced by Msh2Msh6 at G/T mismatch site is similar to the angles introduced by Msh2Msh6 bound at unpaired T-bulge site. The angle of 50° is close to the bend angle observed in the human Msh2Msh6-DNA complex crystal structure (Warren, Pohlhaus, et al, 2007). Statistically analysis indicates that $\sim 20\%$ of the Msh2Msh6 complexes bound at the G/T mismatch, at homoduplex sites, and at unpaired T-bulge are in an unbent conformation. These results appear to suggest the distribution between bent and unbent states of the Msh2Msh6-DNA complexes is independent on the properties of DNA substrates. The breadth of distribution of bent states is much broader than that of unbent states, indicating a large ensemble of bent Msh2Msh6-DNA conformations that may be in a dynamic equilibrium with a relatively rigid unbent Msh2Msh6-DNA complex.

3.4 Discussion

In this study, we employ the same AFM assays used for *E.coli* MutS to investigate the mechanism of mismatch recognition and specificity of eukaryotic MutS homologs, yeast Msh2Msh3 and yeast Msh2Msh6.

In the study of Msh2Msh3, we examined the binding specificity of Msh2Msh3 to three different DNA substrates, homoduplx, single T-bulge and double T-bulge DNA. Although it is suggested that Msh2Msh3 is redundant with Msh2Msh6 with the respect to the repair of small IDLs (Harrington, and Kolodner, 2007; Marsischky, and Kolodner, 1999), in our experiments, we do not observe any significant difference of binding affinity compared Msh2Msh3 bound to single T-bulge DNA substrate with homoduplex DNA. The distribution of Msh2Msh3 on the T-bulge DNA substrate is as random as that of Msh2Msh3 on homoduplex DNA. Msh2Msh3 also recognizes larger IDLs (Habracken, Sung, et al, 1996). However, compared with homoduplex DNA substrate, Msh2Msh3 does not show binding specificity to the specific site (double T –bulge), with only <30% of Msh2Msh3 bound at the specific site.

Msh2Msh6 functions in repair of base-base and small IDLs (Alani, 1996; Marsischky, and Kolodner, 1999). We examined the binding affinity of Msh2Msh6 in presence of these two kinds of mismatched DNAs: a G-T mismatched DNA and an unpaired T-bulge DNA. Although G-T mispair is the most efficiently repaired mismatch (Kramer, Kramer, and Fritz, 1984; Kramer, Kramer, et al, 1989), Msh2Msh6 does not exhibit high binding specificity to the G-T mismatch site, and only <40% of specific complexes are found. In addition, Msh2Msh6 binds to unpaired T-bulge DNA with similar binding specificity. Compared with Msh2Msh3 complexes, the relative higher binding specificity of Msh2Msh6 complex may be derived from the presence of Phe-X-Glu mismatch-recognition motif of Msh6 (Warren, Pohlhaus, et al, 2007). Although we cannot exclude the influence of DNA sequence context surrounding the mispair on the

binding specificity (Marsischky, and Kolodner, 1999), it appears that the efficiency of mismatch-recognition is not simply determined by the binding affinity to the mispaired bases of eukaryotic MutS homologs. In addition, the low binding specificity of MutS homologs is consistent with the bulk solution studies (Jiang, Bai, et al, 2005; Kijas, Studamire, and Alani, 2003; Marsischky, and Kolodner, 1999; Mendillo, Mazur, and Kolodner, 2005).

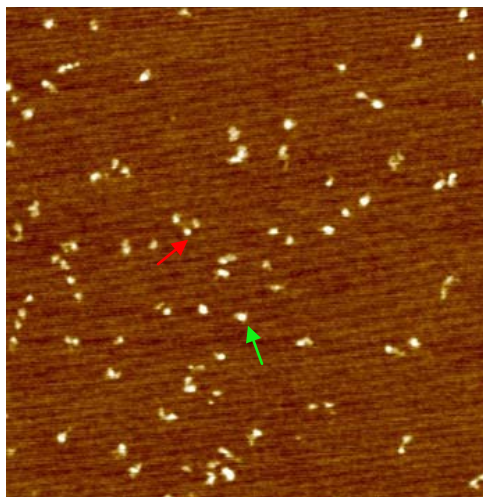
This AFM technique allows us to compare the conformations of MutS homolog-DNA complexes at homoduplex sites and at mismatches. In *E.coli*, the unbent state MutS-DNA complexes is unique for MutS bound at mispaired bases and is suggested to serve as the ultimate recognition complex (URC) in which the DNA is unbent with the mismatched base possibly being flipped out (Tessmer, Yang, et al, 2008; Wang, Yang, et al, 2003). It has been suggested that the repair efficiency will be determined in part, by the relative stability of the bent and unbent state. In regard to the eukaryotic MutS homologs, we find unbent complexes both at the mismatch and at homoduplex sites. It appears that eukaryotic MutS homologs may randomly check out any DNA nucleotide while diffusing along the DNA backbone. The unbent state with a base flipped out is suggested to be more stable at mismatch than at correctly paired base (O'Gara, Horton, et al, 1998). However, our results suggest there is no significant stability difference of unbent complex between specific (at mismatch) and non-specific complexes (at homoduplex site), because the portion of the population of unbent complexes bound at mismatch (~20%) is similar to that of unbent complexes bound at homoduplex sites (~20%). This result may suggest that, in eukaryotic MMR, there is a different mechanism

of recognition or that ATP binding by MutS homologs leads to higher specificity. However, in this study, MutS homologs exhibit very low binding specificity to the specific site, which means there is higher level of nonspecific complex contamination for the complexes bound at specific sites, complicating the analysis of conformation. It is possible that a similar binding mode for MutS homologs bound at specific site and at nonspecific site results from contamination of the nonspecific complexes in the analysis of the specific complexes.

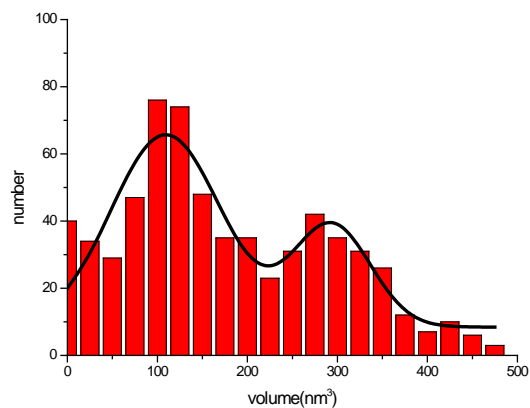
The breadths of distribution are related to the flexibility of the protein-DNA complex (Erie, Yang, et al, 1994; van Noort, Orsini, et al, 1999). The broad distribution of angles of bent complexes (For example, 10° vs. 40° for Msh2Msh3 bound to TT-bulge DNA at mismatch and at homoduplex site), represents an ensemble of a large number of potentially dynamic conformations of Msh2Msh3 or Msh2Msh6. The unbent conformation of the complexes may represent another dynamic conformation of the Msh2Msh3 or Msh2Msh6-DNA complexes, rather than a mismatch-recognition complex.

Msh2Msh6 appears to recognize different types of mismatch in a similar manner. The DNA bend angle distribution of Msh2Msh6 bound with G-T mismatched DNA is similar to that of Msh2Msh6 with unpaired T-bulge DNA, which is consistent with the crystal structure study of Msh2Msh6 Δ 340 with different substrates (Warren, Pohlhaus, et al, 2007). This result suggests that recognition of different DNA substrates is not dependent on the structure of Msh2Msh6 bound at mismatch.

In a conclusion, unlike prokaryotic MutS, mispaired bases appear to be recognized in a different manner for the eukaryotic MutS homologs. So far, the detail about the mechanism is unknown. However, it appears that the mismatch-recognition mechanism of eukaryotic MutS homologs is not simply dependent on the binding specificity of protein to the mismatch, the relative stability of unbent protein-DNA complex to bent protein-DNA complex or the conformation of eukaryotic MutS homolog bound with mismatch.



(a)



(b)

Figure 3.1 Representative AFM image of Msh2Msh3 and histogram of the volume distribution of Msh2msh3. a. AFM image of Msh2Msh3. The size of image is $1\mu\text{m}\times 1\mu\text{m}$. The Msh2Msh3 concentration is ~ 20 nM. The red arrow points to a Msh2Msh3 monomer. The green arrow points to a Msh2Msh3 dimer. b. The volume distribution of Msh2Msh3 is double Gaussian centered $\sim 109\text{nm}^3$ and $\sim 293\text{nm}^3$, representing monomers of Msh2 and Msh3 and Msh2Msh3 heterodimer, respectively.

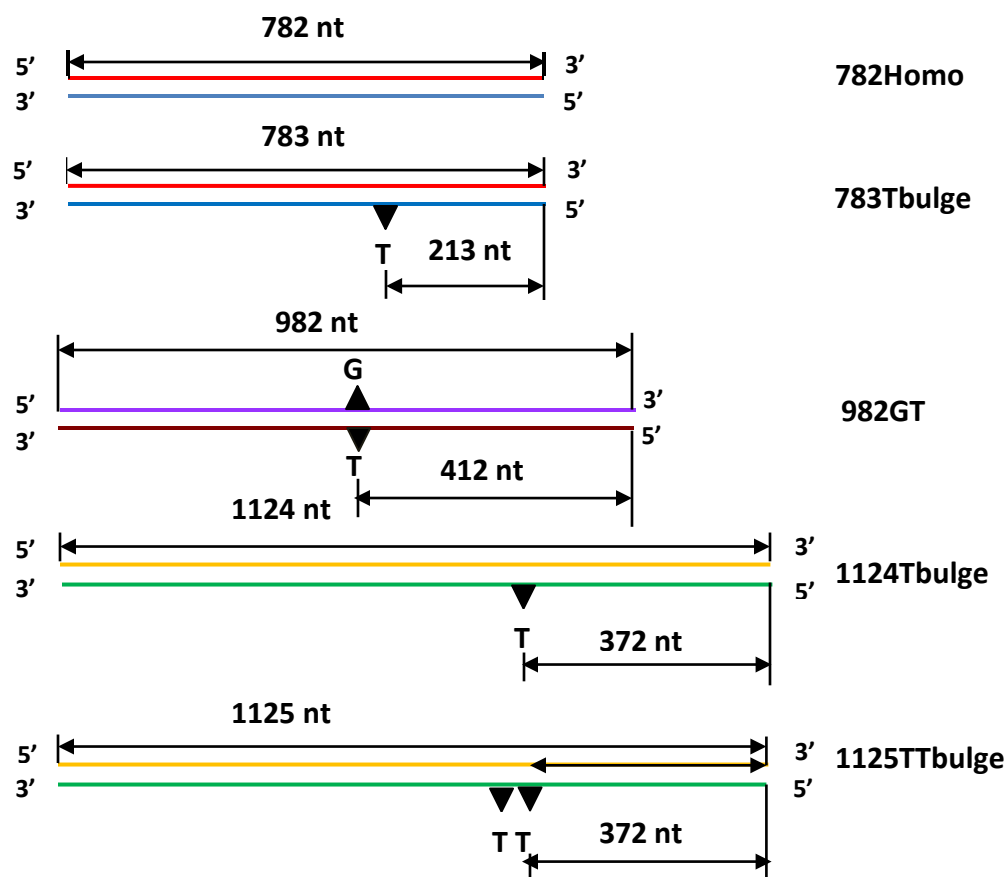
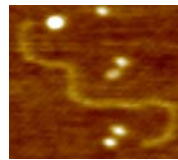
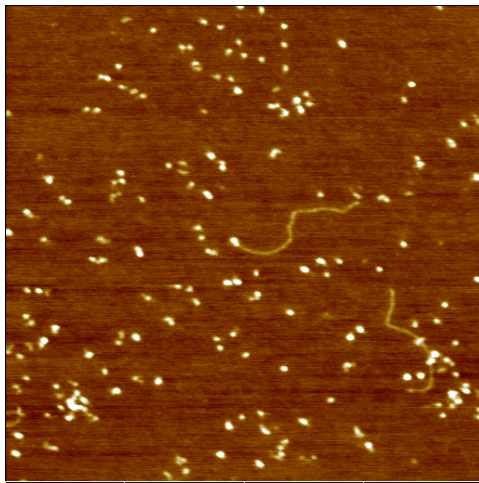
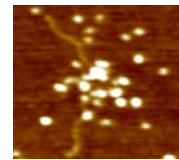


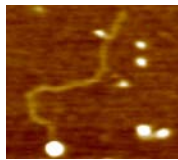
Figure 3.2: Schematic view of the DNA substrates. The substrates are identified on the right with the length of the fragment and the type of mismatch. The length of the fragments and the position of the mismatch in nucleotides from the nearest end are shown by the black arrows. 783T bulge heteroduplex DNA substrate is the same as 782Homo with the exception of having a 1T-bulge, 213 nucleotides from the 5' end of bottom strand. 1124Tbulge and 1125TTbulge are the same DNA substrates with 1 T-bulge and 2 T-bulge 372 nucleotides from the 5' end of bottom strand respectively. The 982GT substrate has a G/T mismatch 412 nucleotides from the 5' end of the bottom strand.



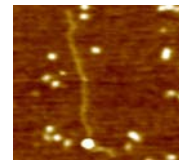
Free DNA: 32%



Cluster complex: 31%

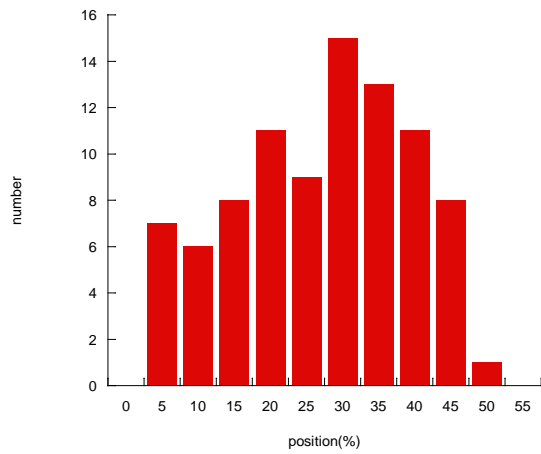


End binding: 18.5%

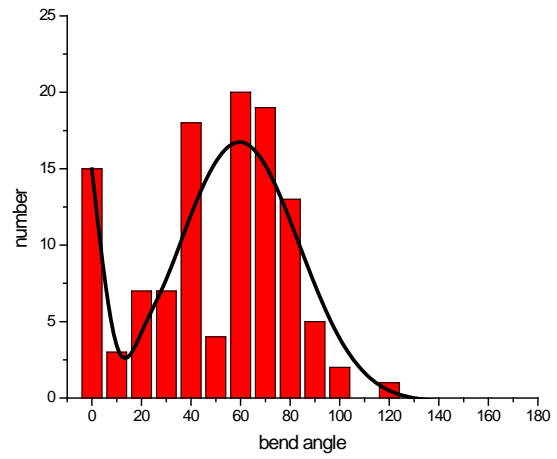


Single binding: 18.5%

(a)



(b)

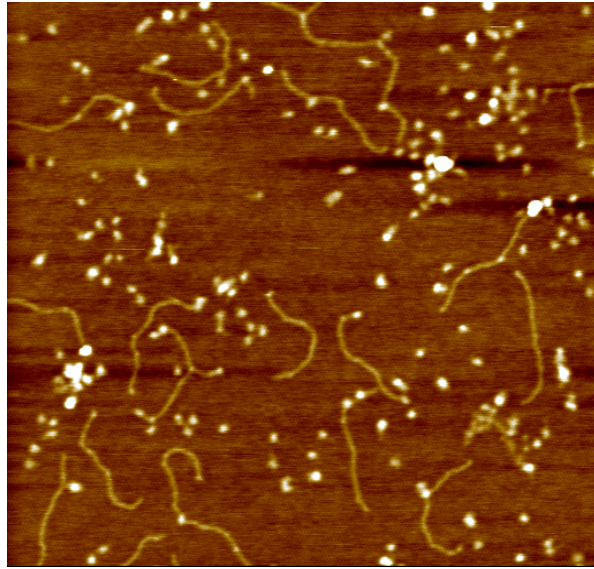


(c)

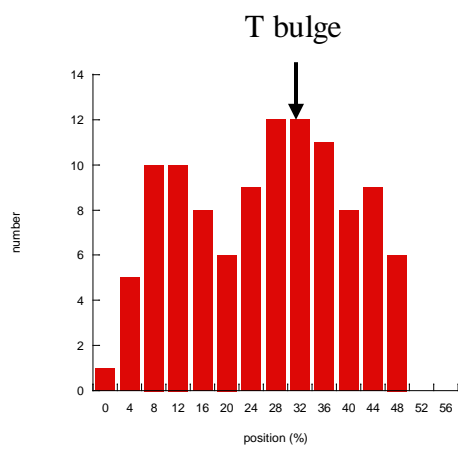
Figure 3.3 Representative AFM image of Msh2Msh3 in the presence homoduplex dsDNA substrate, the distribution of positions of Msh2Msh6 on homoduplex DNA and histogram of DNA bend angles induced by Msh2Msh3 binding.

a. Image of Msh2Msh3 -782Homo DNA complexes. The size of image is $2\mu\text{m}\times 2\mu\text{m}$. The concentrations of Msh2Msh3 and DNA are $\sim 5\text{nM}$ and $\sim 10\text{nM}$, respectively.

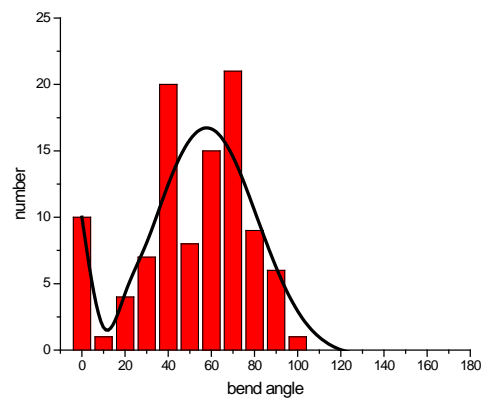
Representative “zoom in” images of four classes of Msh2Msh3-782Homo complex; free DNA (32%), cluster (31%), end binding (18.5%), single binding, (18.5%). b. The distribution of Msh2Msh3 on homoduplex DNA substrate, 782Homo is random. c. The DNA bend angle distribution induced by Msh2Msh3 is fit by a double Gaussian with peaks centered $\sim 0^\circ$ and $\sim 60^\circ$ respectively.



(a)

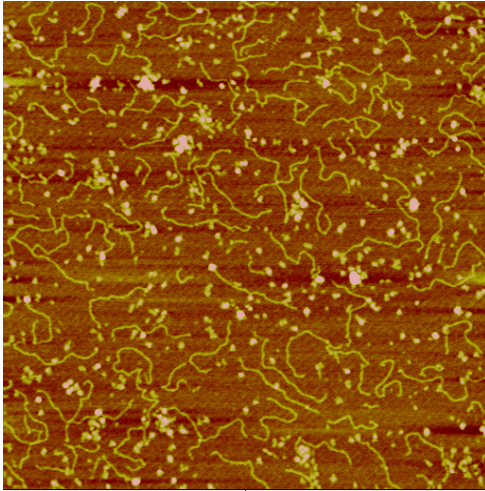


(b)

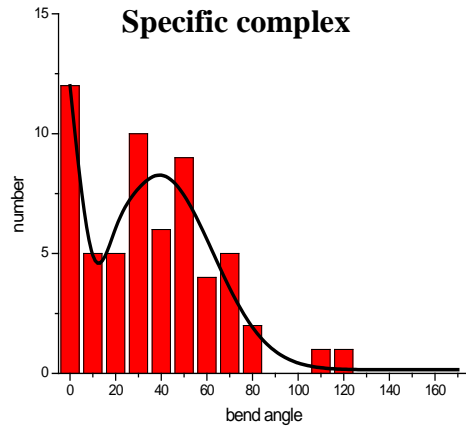


(c)

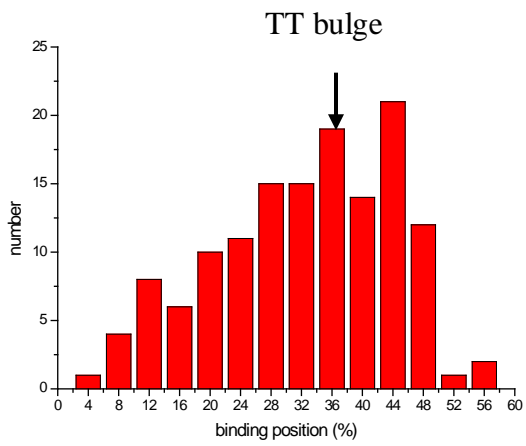
Figure 3.4 Representative AFM image of Msh2Msh3 in the presence of T-bulge DNA substrate, position distribution of Msh2Msh3 on T-bulge heteroduplex DNA, and histogram of DNA bend angle induced by Msh2Msh3 binding. a. Image of Msh2Msh3-783Tbulge DNA complexes. The size of image is $2\mu\text{m}\times 2\mu\text{m}$. The concentrations of protein and DNA are $\sim 5\text{nm}$ and $\sim 10\text{nM}$, respectively. b. The position distributions of Msh2Msh3 on single T-bulge DNA substrate and 783Tbulge DNA are random. c. The DNA bend angle distribution induced by Msh2Msh3 binding is a double Gaussian with peaks centered at $\sim 0^\circ$ and $\sim 60^\circ$.



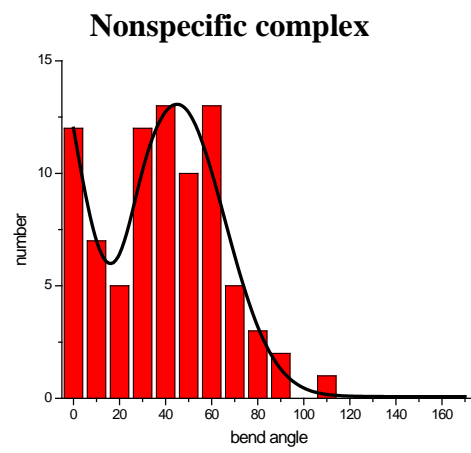
(a)



(c)

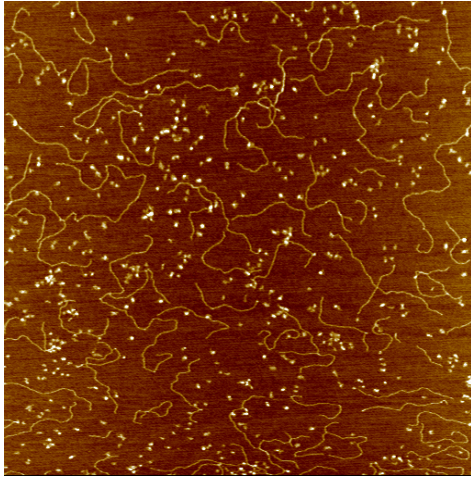


(b)



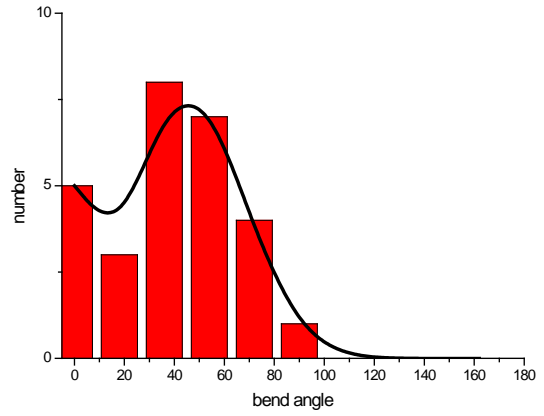
(d)

Figure 3.5 Representative AFM image of Msh2Msh3 in the presence of TT-bulge DNA substrate, position distribution of Msh2Msh3 on TT-bulge DNA and histogram of bend angle induced by Msh2Msh3 bound at double T-bulge sites and homoduplex sites on 1125TTbulge DNA. a, Msh2Msh3 - 1125TTBulge DNA complexes. The concentrations of Msh2Msh3 and 1125TTbulge DNA substrate are ~ 5nM and ~20 nM, respectively. The size of image is 2 μ m \times 2 μ m. b. The distribution of binding positions of Msh2Msh3 on 1125TT-bulge DNA substrate suggests Msh2Msh3 shows weak binding specificity to the mismatches. c. DNA bend angles of specific complex are shown for Msh2Msh3 bound at double T-bulge site on 1125TTbulge DNA substrates. The distribution of bend angles is fit by a double Gaussian with peaks at ~0° and ~40°. d. DNA bend angles of nonspecific complex are shown for Msh2Msh3 bound at homogenous sites on 1125TTbulge DNA substrate. The distribution of bend angles is fit by a double Gaussian with peaks at at ~0° and ~40°.

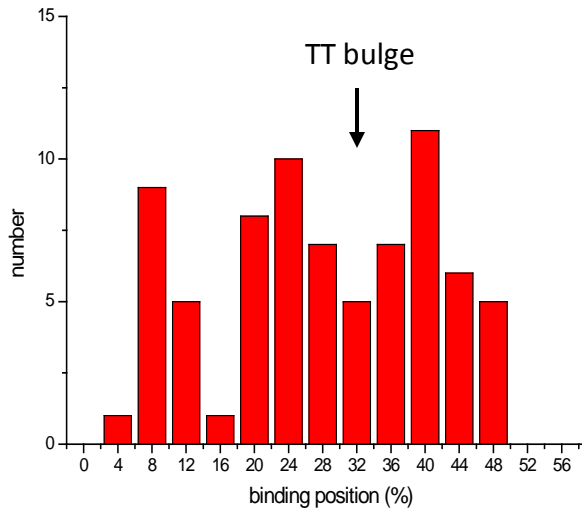


(a)

Specific complex

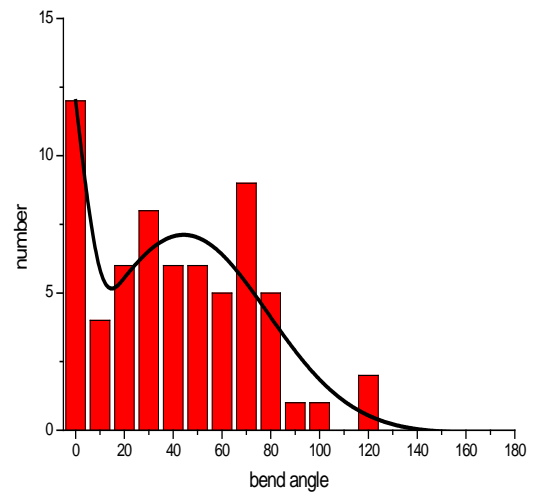


(c)



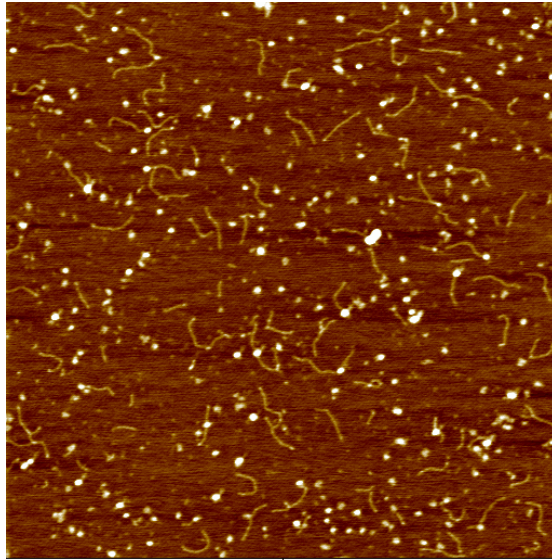
(b)

Nonspecific complex

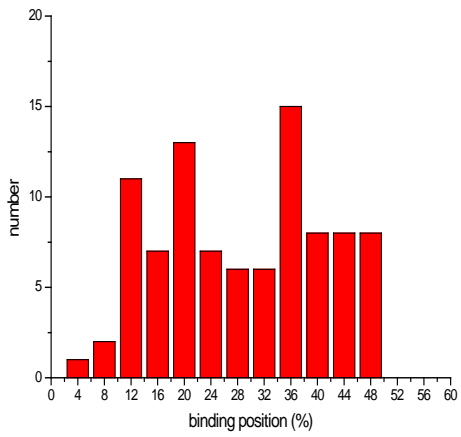


(d)

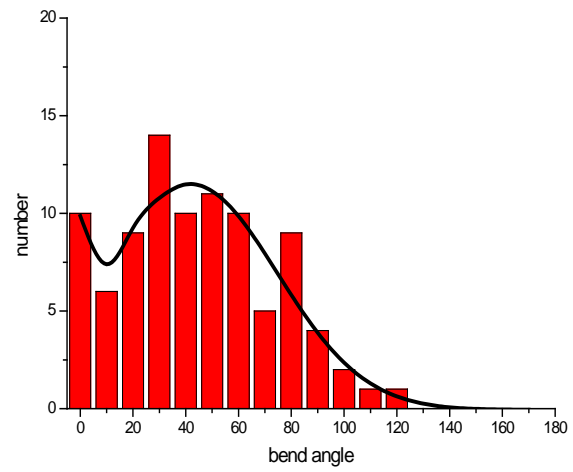
Figure 3.6 Representative AFM image of human Msh2Msh3 in the presence of TT-bulge DNA substrates, position distribution of human Msh2Msh3 on TT-bulge DNA and histogram of bend angle induced by human Msh2Msh3 bound at double T-bulge site and homoduplex site on 1125TTbulge DNA. a, hMsh2Msh3 - 1125TTBulge DNA complex. The concentrations of hMsh2Msh3 and 1125TTbulge DNA substrate are ~ 5nM and ~20 nM, respectively. The size of image is 2 μ m \times 2 μ m. b. The distribution of binding positions of hMsh2Msh3 on 1125TT-bulge DNA substrate is random. c. DNA bend angles of specific complex are shown for hMsh2Msh3 bound at double T-bulge site on 1125TTbulge DNA substrate. The distribution of bend angles is fit by a double Gaussian with peaks at ~0° and ~40°. d. DNA bend angles of nonspecific complex are shown for hMsh2Msh3 bound at homogenous sites on 1125TTbulge DNA substrate. The distribution of bend angles is fit by a double Gaussian with peaks at ~0° and ~40°.



(a)

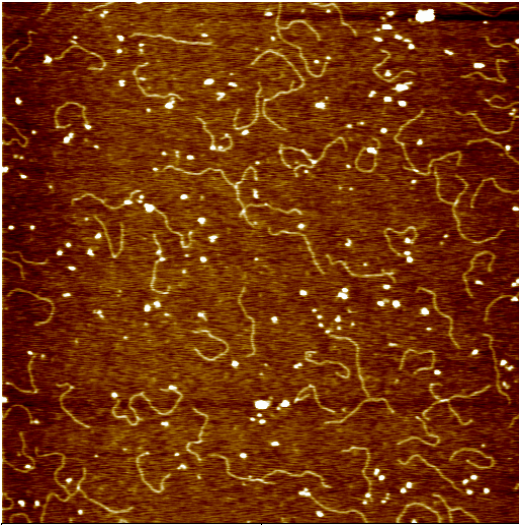


(b)

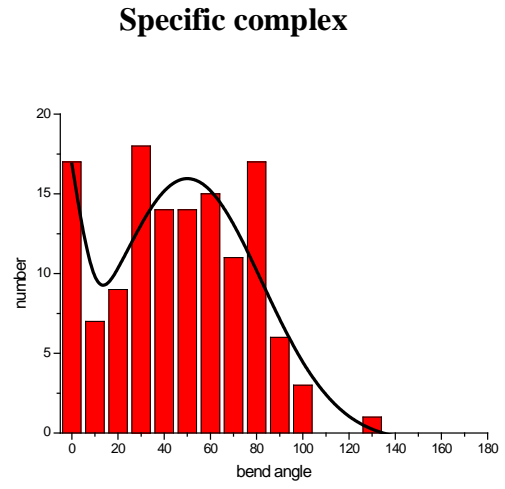


(c)

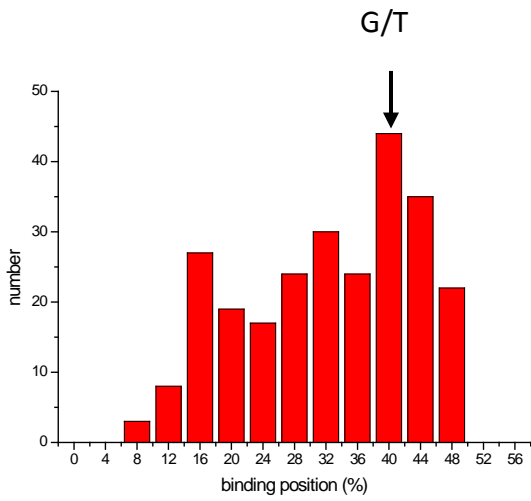
Figure 3.7 Representative AFM image of Msh2Msh6 in presence of homoduplex DNA, position distribution of Msh2Msh6 on homoduplex DNA and histogram of DNA bend angles introduced by Msh2Msh6 binding. a. Msh2Msh6-homoduplex DNA complex. The image is $2\mu\text{m}\times 2\mu\text{m}$. The concentrations of Msh2Msh6 and DNA are $\sim 5\text{nM}$ and $\sim 10\text{nM}$, respectively. The homoduplex DNA substrate is piece of pUC18 PCR fragment with 400 nucleotides length. b. The distribution of binding positions of Msh2Msh6 on homoduplex DNA is random. c. The distribution of bend angles induced by Msh2Msh6 binding is fit by a double Gaussian with peaks at $\sim 0^\circ$ and $\sim 40^\circ$.



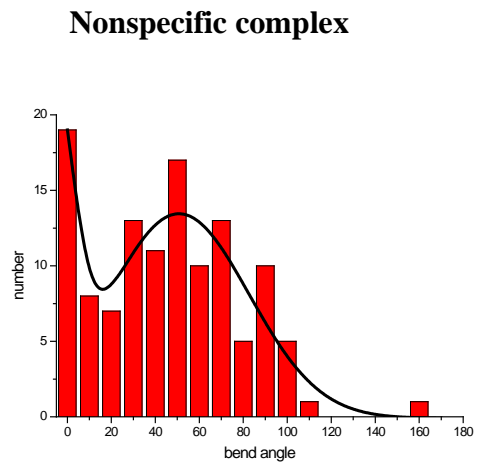
(a)



(c)

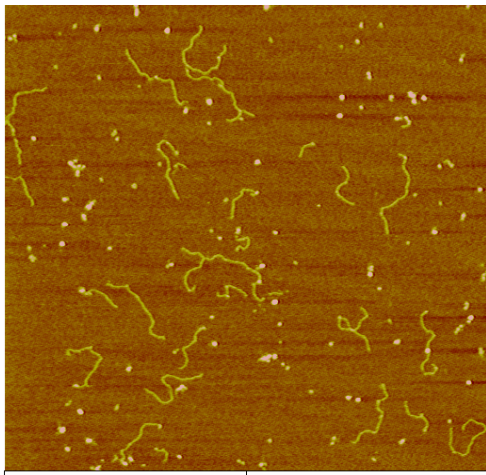


(b)

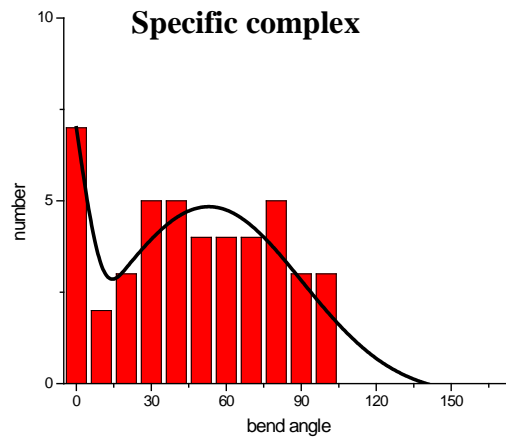


(d)

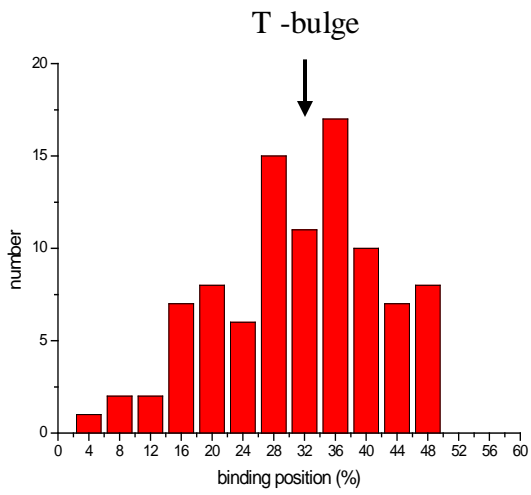
Figure 3.8 Representative AFM image of Msh2Msh6 in the presence of G/T mismatch DNA, position distribution of Msh2Msh6 on 982GT heteroduplex DNA and histogram of DNA bend angles introduced by Msh2Msh6 bound at G/T mismatch site and homogenous site on 982GT DNA substrate. a. Msh2Msh6 – 982GT DNA complex. The concentrations of Msh2Msh6 and 982GT DNA substrate are ~ 5nM and ~10 nM, respectively. The size of image is 2 μ m \times 2 μ m. b. The distribution of binding position of Msh2Msh6 on 982GT DNA substrate suggests Msh2Msh6 shows weak binding specificity to the G/T mismatch. c. DNA bend angles of specific complex are shown for Msh2Msh6 bound at G/T site on 982GT DNA substrate. The distribution of bend angles is fit by a double Gaussian with peaks at $\sim 0^\circ$ and $\sim 50^\circ$. d. DNA bend angles of nonspecific complex are shown for Msh 2Msh6 bound at homogenous sites on 982G/T DNA substrate. The distribution of DNA bend angles is fit by a double Gaussian with peaks at $\sim 0^\circ$ and $\sim 50^\circ$.



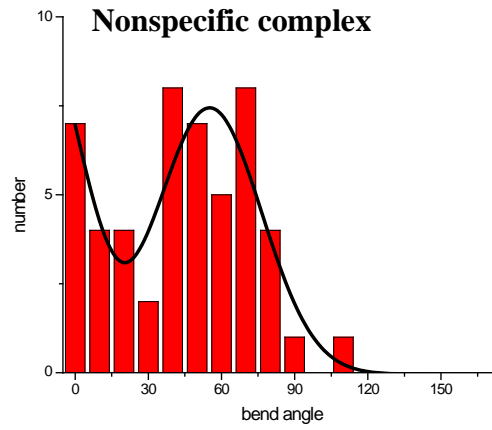
(a)



(c)



(b)



(d)

Figure 3.9 Representative AFM image of Msh2Msh6 in the presence of T-bulge DNA substrate, position distribution of Msh2Msh6 on T-bulge DNA and histogram of bend angle introduced by Msh2Msh6 bound at single T-bulge site and homoduplex site on 1124Tbulge DNA. a, Msh2Msh6 - 1124TBulge DNA complex. The concentrations of Msh2Msh6 and 1124Tbulge DNA substrate are ~ 5nM and ~10 nM, respectively. The size of image is 2 μ m \times 2 μ m. b. The distribution of binding position of Msh2Msh6 on 1124T-bulge DNA substrate suggests Msh2Msh6 shows weak binding specificity to the single T-bulge mismatch. c. DNA bend angles of specific complex are shown for Msh2Msh6 bound at single T-bulge site on 1124Tbulge DNA substrate. The distribution of bend angles is fit by a double Gaussian with peaks at ~0° and ~50°. d. DNA bend angles of nonspecific complex are shown for Msh 2Msh6 bound at homogenous sites on 1124Tbulge DNA substrate. The distribution of bend angles is fit by a double Gaussian with peaks at ~0° and ~50°.

Reference List

- Alani, E. (1996). The *Saccharomyces cerevisiae* Msh2 and Msh6 proteins form a complex that specifically binds to duplex oligonucleotides containing mismatched DNA base pairs. *Mol. Cell. Biol.* *16*, 5604-5615.
- Bustamante, C., and Rivetti, C. (1996). Visualizing protein-nucleic acid interactions on a large scale with the scanning force microscope. *Annu. Rev. Biophys. Biomol. Struct.* *25*, 395-429.
- Drummond, J.T., Li, G.M., Longley, M.J., and Modrich, P. (1995). Isolation of an hMSH2-p160 heterodimer that restores DNA mismatch repair to tumor cells. *Science* *268*, 1909-1912.
- Echols, H., and Goodman, M.F. (1991). Fidelity mechanisms in DNA replication. *Annu. Rev. Biochem.* *60*, 477-511.
- Erie, D.A., Yang, G., Schultz, H.C., and Bustamante, C. (1994). DNA bending by Cro protein in specific and nonspecific complexes: implications for protein site recognition and specificity. *Science* *266*, 1562-1566.
- Gradia, S., Acharya, S., and Fishel, R. (2000). The role of mismatched nucleotides in activating the hMSH2-hMSH6 molecular switch. *J. Biol. Chem.* *275*, 3922-3930.
- Habraken, Y., Sung, P., Prakash, L., and Prakash, S. (1996). Binding of insertion/deletion DNA mismatches by the heterodimer of yeast mismatch repair proteins MSH2 and MSH3. *Curr. Biol.* *6*, 1185-1187.
- Harrington, J.M., and Kolodner, R.D. (2007). *Saccharomyces cerevisiae* Msh2-Msh3 acts in repair of base-base mispairs. *Mol. Cell. Biol.* *27*, 6546-6554.
- Iyer, R.R., Pluciennik, A., Burdett, V., and Modrich, P.L. (2006). DNA mismatch repair: functions and mechanisms. *Chem. Rev.* *106*, 302-323.
- Jiang, J., Bai, L., Surtees, J.A., Gemici, Z., Wang, M.D., and Alani, E. (2005). Detection of high-affinity and sliding clamp modes for MSH2-MSH6 by single-molecule unzipping force analysis. *Mol. Cell* *20*, 771-781.
- Kijas, A.W., Studamire, B., and Alani, E. (2003). Msh2 separation of function mutations confer defects in the initiation steps of mismatch repair. *J. Mol. Biol.* *331*, 123-138.

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

Kramer, B., Kramer, W., Williamson, M.S., and Fogel, S. (1989). Heteroduplex DNA correction in *Saccharomyces cerevisiae* is mismatch specific and requires functional PMS genes. *Mol. Cell. Biol.* 9, 4432-4440.

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

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

Lee, S.D., Surtees, J.A., and Alani, E. (2007). *Saccharomyces cerevisiae* MSH2-MSH3 and MSH2-MSH6 complexes display distinct requirements for DNA binding domain I in mismatch recognition. *J. Mol. Biol.* 366, 53-66.

Li, G.M. (2008). Mechanisms and functions of DNA mismatch repair. *Cell Res.* 18, 85-98.

Loeb, L.A., and Kunkel, T.A. (1982). Fidelity of DNA synthesis. *Annu. Rev. Biochem.* 51, 429-457.

Marsischky, G.T., and Kolodner, R.D. (1999). Biochemical characterization of the interaction between the *Saccharomyces cerevisiae* MSH2-MSH6 complex and mispaired bases in DNA. *J. Biol. Chem.* 274, 26668-26682.

Mendillo, M.L., Mazur, D.J., and Kolodner, R.D. (2005). Analysis of the interaction between the *Saccharomyces cerevisiae* MSH2-MSH6 and MLH1-PMS1 complexes with DNA using a reversible DNA end-blocking system. *J. Biol. Chem.* 280, 22245-22257.

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

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

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

- O'Gara, M., Horton, J.R., Roberts, R.J., and Cheng, X. (1998). Structures of HhaI methyltransferase complexed with substrates containing mismatches at the target base. *Nat. Struct. Biol.* *5*, 872-877.
- Palombo, F., Iaccarino, I., Nakajima, E., Ikejima, M., Shimada, T., and Jiricny, J. (1996). hMutSbeta, a heterodimer of hMSH2 and hMSH3, binds to insertion/deletion loops in DNA. *Curr. Biol.* *6*, 1181-1184.
- Ratcliff, G.C., and Erie, D.A. (2001). A novel single-molecule study to determine protein--protein association constants. *J. Am. Chem. Soc.* *123*, 5632-5635.
- Schofield, M.J., Brownell, F.E., Nayak, S., Du, C., Kool, E.T., and Hsieh, P. (2001). The Phe-X-Glu DNA binding motif of MutS. The role of hydrogen bonding in mismatch recognition. *J. Biol. Chem.* *276*, 45505-45508.
- Tessmer, I., Yang, Y., Zhai, J., Du, C., Hsieh, P., Hingorani, M.M., and Erie, D.A. (2008). Mechanism of MutS searching for DNA mismatches and signaling repair. *J. Biol. Chem.* *283*, 36646-36654.
- van Noort, J., Orsini, F., Eker, A., Wyman, C., de Grooth, B., and Greve, J. (1999). DNA bending by photolyase in specific and non-specific complexes studied by atomic force microscopy. *Nucleic Acids Res.* *27*, 3875-3880.
- Varlet, I., Canard, B., Brooks, P., Cerovic, G., and Radman, M. (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*, 10156-10161.
- Wang, H., Yang, Y., Schofield, M.J., Du, C., Fridman, Y., Lee, S.D., Larson, E.D., Drummond, J.T., Alani, E., Hsieh, P., and Erie, D.A. (2003). DNA bending and unbending by MutS govern mismatch recognition and specificity. *Proc. Natl. Acad. Sci. U. S. A.* *100*, 14822-14827.
- Warren, J.J., Pohlhaus, T.J., Changela, A., Iyer, R.R., Modrich, P.L., and Beese, L.S. (2007). Structure of the human MutSalpha DNA lesion recognition complex. *Mol. Cell* *26*, 579-592.
- Yang, Y., Wang, H., and Erie, D.A. (2003). Quantitative characterization of biomolecular assemblies and interactions using atomic force microscopy. *Methods* *29*, 175-187.