Spring 5-10-2014

Modeling Assemblies and Interactions at the Replication Fork: Sliding Clamps and Clamp Interacting Enzymes

Xiaojun Xu
Georgia State University

MODELING ASSEMBLIES AND INTERACTIONS AT THE REPLICATION FORK: SLIDING CLAMPS AND CLAMP INTERACTING ENZYMES

by

XIAOJUN XU

Under the Direction of Dr. Ivaylo Ivanov

ABSTRACT

DNA sliding clamps are structurally conservative toroid-shape proteins that encircle and slide along DNA, serving as scaffold for other functional enzymes to act on DNA and ensuring the replication proccessivity, thereby, of fundamental biological significance across domains of life. Mechanistic details and related functional implications concerning clamp opening, interaction between clamp and clamp-interactive proteins, post-translational modification of sliding clamp remain largely elusive due to technical difficulties in single molecule level manipulations and structural studies on large biological complex.

Toward the end of providing a unified molecular-level description on clamp loading that would account for all available experimental observations, we calculated the interface binding energy and depicted the residue pair contributions using MM/PBSA and MM/GBSA calculations

to compare the different interfaces of sliding clamps, and dissolved the uncertainty in comparative stability of different sliding clamp interfaces. The possible interface breaking pathways were investigated by sampling the opening state of interfaces using SMD simulations.

Functioning as a polymerase accessory factor, sliding clamp associates with the dual enzymatic functional polymerase B (PolB) as DNA replication occurs. The massive conformational switch of PolB between replicating and editing mode is recognized for its functional significance but little is understood in the context of the PCNA/PolB/DNA complex. We integrated the structural informations from individual and binary crystal structures, as well as low-resolution structural information and other functional assay results, to build the complex atomistic models in both modes and refine them through atomistic simulations. The transition process was probed using TMD and ENM to reveal the structural characteristics and determinants of the transition. Sliding clamp is also a master coordinator of cellular responses to DNA damage. Efforts with the same methodology were made on human PCNA/FEN1/DNA ternary complex to investigate the reversible associations of FEN1 to PCNA and the conformational switching leading to exchange of repair intermediates.

In the third thrust, we modeled the ubiquitin-modified and SUMO-modified PCNA using protein-protein docking and atomic simulation. Alongside with the SAXS data, our results revealed the structural basis for the distinct functional outcomes upon different posttranslational modification of PCNA.

INDEX WORDS: DNA replication, DNA repair, Sliding clamp, Proliferating cell nuclear antigen (PCNA), Polymerase B (PolB), Flap endonuclease-1 (FEN1), MM/PBSA, MM/GBSA, Molecular dynamics (MD), Steered molecule dynamics (SMD), Targeted molecule dynamics (TMD), Small angle X-ray scattering (SAXS)

MODELING ASSEMBLIES AND INTERACTIONS AT THE REPLICATION FORK: SLIDING CLAMPS AND CLAMP INTERACTING ENZYMES

by

XIAOJUN XU

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2014

MODELING ASSEMBLIES AND INTERACTIONS AT THE REPLICATION FORK: SLIDING CLAMPS AND CLAMP INTERACTING ENZYMES

by

XIAOJUN XU

Committee Chair: Ivaylo Ivanov

Committee: Donald Hamelberg

Ivaylo Ivanov

Jenny Yang

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

May 2014

ACKNOWLEDGEMENTS

First and foremost I would like to thank my advisor Dr. Ivaylo Ivanov for his guidance in my graduate study and developing a professional career. He has been a great inspiration and icon to me in terms of both his scientific expertise and personality. I've learned and been giving far beyond what I expected. It's really a fortune that a graduate student could have this extraordinary level of support from his/her advisor, and I am and will be always deeply grateful for his effort, patience, and tolerance in guiding me through the years of training.

I am also greatly indebted to my previous advisor Dr. Jenny Yang. I would not have been in here if she did grant me the opportunity to work with her in the first place. She has being tremendously generous and supportive to me, especially during the difficult times. There would not be such a way leading me toward the Ph.D. degree if not for her.

Also, I would like to express my gratitude to Dr. Hamelberg for serving in my dissertation committee, and for his valuable advises and ideas on my dissertation research and career development.

I am indebted to my mentor, Dr. Carlo Guardiani for his meticulous guidance and directions when I firstly came into the lab. Even it was only for a half year, I treasure the experience of working with him greatly since it paved a solid path for me to come the long way through.

Special thanks to my colleges, Dr. Chunli Yan and Dr. Buddhadev Maiti; in particular, Chunli and I work closely on many projects and she has never been even slightly hesitating to share what she knew. Many of the projects would not have been possibly completed without her guidance and contribution. Dr. Maiti taught me and helped me to figure out lots of technical difficulties for numerous times. I can never thank them enough. I also want to thank Bradley Kossmann for his essential inspections and insightful suggestion on my manuscripts, and Kath-

leen Carter for the stimulating discussions we had on our projects and day-to-day cheerful conversations.

The six years in Atlanta would have been colorless without the companionship and constant support from Alice, who is with the most caring and noble heart I've ever met. It is with her I shared the miles on the road, ups and downs, laughs and tears, and was able to get over with all the uncertainties and frustrations. Not for a single moment shall the best time of life be forgotten.

Thanks go to many others with whom we crossed our lives here and there, sharing wonderful moments and stories. Kendra Hubbard—my talkative non-kin American sister, always difficult to hang her up over the phone; Shuo Wang—one of the very few girls I know with a fine sense of humor, could very well have become a standalone comedienne if not being a chemist, contributed greatly in eliminating my poor working style of procrastination; Jina Jingjuan Qiao—the walking Encyclopedia of "what should I do", came from my own home city; William Liangwei Li—Jina's husband, the notorious Mr. Mean, never been outwitted by anyone over argumentations; Jeff Jiafeng Geng and Dan Cui couple—my favorite beer companion; Bob Wohlhueter—a erudite scholar who really enjoys research and life as a whole, he is a role-model for me; Chris Loper—my best badminton teacher and partner; Stan Singleton—my Bible study teacher in the first year after I came in US. The friendships we shared kept me not feeling alone over the long time of being far away from home. I'm truly grateful for all of them.

I acknowledge that many chapters of this dissertation have or will be appeared in print elsewhere. Chapter 3 is adapted from Nucleic Acids Research (2013) 41(16): 7804-7814; I thank the coauthors of collaborators, Dr. Lin Wang, Dr. Ravindra Kumar, Dr. Tony Liu, Dr. Tae-Hee Lee, and Dr. Stephen Benkovic. This research was supported by the National Science Foundation (NSF-CAREER MCB-1149521 to I.I.), start-up funds from Georgia State Univ. to I.I.

Computational resources were provided in part by allocations from the NSF XSEDE program (CHE110042) and the National Energy Research Scientific Computing Center supported by the DOE Office of Science (contract DE-AC02-05CH11231).

Chapter 4 appears in Nucleic Acids Research (2013) 41(22): 10020-10031; the dissertation author was the primary author of the paper. This work was supported by a National Science Foundation CAREER award [MCB-1149521 to I.I.] and Georgia State University start-up funds to [I.I.]. Computational resources were provided in part by a National Science Foundation XSEDE allocation [CHE110042] and through an allocation (BIP007) at NERSC supported by the U.S. Department of Energy Office of Science [contract DE-AC02-05CH11231]. Funding for open access charge: National Science Foundation CAREER award [MCB-1149521 to I.I.]

Chapter 5 is a completed manuscript about to be submitted; the dissertation author is the primary author of the paper. This work was supported by a National Science Foundation CAREER award [MCB-1149521 to I.I.], Georgia State University start-up funds to [I.I.]. Computational resources were provided in part by a National Science Foundation XSEDE allocation [CHE110042] and through an allocation (BIP007) at NERSC supported by the U.S. Department of Energy Office of Science [contract DE-AC02-05CH11231].

Chapter 6 is adapted from PNAS (2012) 109(22): 8528-8533; I thank the coauthors of our collaborator, Dr. Jordi Querol-Audi, Dr. Susan Tustakawa, Dr. Miaw-Sheue Tsai, Dr. John Tainer, Dr. Priscilla Cooper, and Dr. Eva Nogales. The authors thank Dr. Gabriel Lander and Patricia Grob for advice during EM data collection and processing. Computational resources were provided in part by an NSF Teragrid allocation (CHE110042) and through an allocation (BIP007) at the National Energy Research Scientific Computing Center (NERSC) supported by the DOE. Work on the project is supported by Georgia State University (to I.I.), Cleon C. Ar-

rington research initiation grant (to I.I.), P01 CA092584 (NCI to J.A.T.), R01 CA081967 (NCI to J.A.T). E.N. is a Howard Hughes Medical Institute Investigator.

Chapter 7 is a completed manuscript has been submitted to Plos Biology (under peer review as the time of uploading this disseration). I thank the coauthors of our collaborator, Dr. Susan Tsutakawa, Dr. Chirstopher Weinacht, Dr. Bret Freudenthal, Dr. Zhihao Zhuang, Dr. Todd Washington, and Dr. John Tainer. This work was supported by a NSF CAREER Grant MCB-1149521 (to I.I.), Georgia State University start-up funds (to I.I.), P01 CA092584 (NCI to J.A.T.), R01 CA081967 (NCI to J.A.T), NSF Grant MCB-0953764 (to Z.Z.) and R01 GM108027 (to M.T.W.). Computational resources were provided in part by a National Science Foundation XSEDE allocation [CHE110042] and through an allocation at National Energy Research Scientific Computing Center (NERSC) supported by the U.S. Department of Energy Office of Science [contract DE-AC02-05CH11231]. This SAXS data was collected at BL12.3.1 at the Advanced Light Source (ALS), supported by the Integrated Diffraction Analysis Technologies (IDAT) program (DOE/BER), by DOE contract DE-AC02-05CH11231, and by NIH MINOS (R01GM105404).

I thank the Molecular Basis Diseases fellowship at Georgia State University for the financial support for my graduate study.

Above all, I want to thank my Mom Liping Xia, my Dad Jingpei Xu, and my sister Xiaoling Xu for their incomparable love to me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. SLIDING CLAMP IN DNA REPLICATION AND REPAIR

1.1 Replisome

Genome replication is a vital event in the life cycle of all organisms. During the chromosome replication phase, the physical integrity of DNA is preserved through a precisely controlled semiconservative mechanism, which is carried out by a macromolecular assembly that constitutes numerous proteins, the replisome (1). During DNA replication, the replisome has to couple the unwinding of parental double-stranded (ds) DNA with nascent DNA synthesis on both strands. Two dsDNA duplexes with the exact same sequence as the parental dsDNA are produced via replication, while in real life it is not as straightforward as it appears to be since the replisome also needs to integrate DNA damage response signals into the progression, arresting, and resume of replication fork progression (1,2).

The replisome's various functionalities arise from its constituent modules, common to all domains of life: the helicase, single strand binding protein (SSB), primase, DNA polymerase, proofreading 3'-5' exonucleases, primer removal endonuclease, ligase, sliding clamp and the clamp loader. These modules facilitate the loading of toroidal sliding clamps onto primed DNA, guaranteeing high processivity in DNA replication (1,3,4). Followed by the assembly of the replisome at sequence-specific positions along DNA, known as origin sequences, duplex DNA is separated into leading and lagging template strands by a helicase. The SSBs then bind to the two separated DNA strands to stabilize them and prevent annealing. To initiate DNA synthesis, a short strand of RNA primer must be synthesized by primase for either leading or lagging-strand. The replicase is then being able to start synthesizing the nascent DNA with the template and primer on each strand. Since replicases proceed in a unidirectional manner during replication, and because both strands of the nascent DNA synthesis occur in the 5' to 3' direction, the anti-

parallel nature of dsDNA dictates that the leading strand be extended in the same direction as replication fork movement, while the lagging strand has to be synthesized in the opposite direction. As a result, the lagging-strand template accumulates single-strand (ss) DNA as a collection of Okazaki fragments (5). It's worth noting that the sliding clamp has to be loaded onto each template strand by a clamp loader prior to the initiation of replication to ensure processsivity (6). Even though replication is carried out by a high fidelity replicase, it is still subject to an intrinsic low error rate due to the misincorporation of incorrect deoxynucleoside triphosphate (dNTP) into the growing chain. Most of these errors can be detected and corrected by the intrinsic 3' to 5' exonuclease activity of the replicase. The primer for nascent strand DNA strand needs to be removed by organism-specific enzymes such as DNA polymerase I, RNAse H, flap endonuclease 1 (FEN1), etc. In the last stage, the ligase must connect the Okazaki fragments on the lagging strand to complete the replication (1-3).

Despite substantial conservation of the core components and functional mechanisms of cellular replisomes, significant differences in the complexity, inter-component interactions and mechanistic regulation have evolved between different domains of life (3). A comparison of these differences between a selection of different organisms is shown in Table 1.1.

1.1.1 Prokaryotic replisome

As with many other *in vivo* studies, examining the replisome is easiest in prokaryotes. For example, the *E. coli* replisome is comprised of the replicase DNA polymerase III (Pol III), sliding clamp (β-clamp), the clamp loader (γ-complex) etc. (Table 1). The homohexameric DnaBencircles one strand of the parental DNA, utilizing the energy from ATP hydrolysis to separate the duplex DNA into two daughter strands by translocating along the strand in the 5' to 3' direction within its DNA-binding pore. SSB proteins are then recruited to each strand to remove

thermodynamically favorable secondary structural formations that may impede replication and prevent proper annealing. The Pol III complexes for both leading- and lagging-strand are tethered to a β-clamp (also complexed with DnaB), synthesize complementary DNA strands with template and primer, which is created by DnaG, an RNA primase.

Table 1.1 The replicase components from different domains of life

Component	*E.coli*	Eukaryotes	Archaea	T4 Phage
Polymerase	Pol III Core ($\alpha\varepsilon\theta$)	Pol $\alpha/\delta/\varepsilon$	Pol δ	gp43
Sliding clamp	β-clamp	PCNA	PCNA	gp45
Clamp loader	γ-complex	RFC	RFC	gp44/62
SSB	SSB	RPA	RPA	gp32
Primase	DnaG	Pol α-primase	Primase	gp61
Helicase	DnaB	MCM2-7/CMG	MCM	gp41

With the help of accessory proteins, Pol III becomes very efficiency and attains the highest specificity of all *E. coli* DNA polymerases. Each Pol III core consists of 3 subunits: α, ε, and θ (7,8). The α subunit is encoded by dnaE gene, possesses DNA polymerase activity, is capable of incorporating 8 nucleotides/second (ntd/s), on the order of magnitude of Pol III core activity (20 ntd/s) (9). The ε subunit is a proofreading 3′-5′ exonuclease, encoded by dnaQ. While the ε

subunit is essential to maintain the processivity of Pol III , the small θ subunit (holE) bears no substantial recognized function but a slight stimulation on the activity of ε (10). Pol III operates with high fidelity, misinorporating one incorrect dNTP in about every 10^6 to 10^7 bases (11). The ε subunit of the polymerase removes incorrect dNTPs to augment fidelity. The dimeric, ring-shaped β-clamp, tethered to Pol III, topologically encircles template DNA, preventing the dissociation of the polymerase from the DNA and furthering proccessivity. With energy from ATP hydrolysis, the clamp loader opens and then reseals the β-clamp subunit-subunit interface around DNA (3).

Pol III synthesizes lagging-strand Okazaki fragments in a series of 1–2 kilo-base (kb) pieces by traveling in the opposite direction of replication fork progression (12,13). The C-terminal extension of the clamp loader τ subunits' simultaneous binding of Pol III and DnaB keeps the replisome from falling apart while creating a DNA loop between the lagging strand β-clamp and the helicase. When Okazaki fragment synthesis is nearly complete, the β-clamp dissociates from Pol III and DNA, and the loop region relaxes (14,15). The Pol III complex and β-clamp then recycles on the upstream RNA primer for the new Okazaki fragment synthesis. After Okazaki fragment synthesis, DNA Pol I has to replace the RNA primer with DNA when an Okazaki fragment is completed. The gaps between Okazaki fragments are fixed by ligase in the end (3).

1.1.2 Eukaryotic replisome

The replisome in eukaryotic cells contains modules of similar functionalities to those found in *E. coli* replisome. The homohexameric helicase MCM (Minichromosome Maintenance, Mcm2-7) contains six distinct subunits; its prokaryotic analog being DnaB in *E. coli*. MCM encircles the template leading-strand and translocates in the direction of 3' to 5' to disrupt the du-

plex DNA. Instead of Pol III operating on both strands for DNA synthesis as in *E. coli*, Pol ε and Pol δ act as two separate polymerases on the leading and lagging-strand, respectively (16). The eukaryotic, homotrimeric sliding clamp Proliferating Cell Nuclear Antigen (PCNA) is the homotrimer sliding clamp in eukaryote, adopts the ring-shaped structure. The Pol α-primase contains four subunits that include a primase and DNA polymerase, capable of synthesizing the RNA-DNA hybrid primer rather than the single-stranded RNA fragment. The heterotimeric Replication Protein A (RPA) is the *E. coli* SSB functional counterpart in eukaryotes, removing secondary structure impediments and stabilize the single-stranded template DNA.

Some other proteins such as Cdc45, Sld2, Sld3, Dpb11, GINS complex (Go, Ichi, Nii, and San), etc., are also involved in the initiation and progression of replication without known prokaryotic counterparts. The exact functions of most of these participants remain largely unclear (17), while it has been proved that the functioning processes of many of these proteins are precisely regulated by posttranslational modification, such as phosphorylation, ubiquitination and sumoylation. These modifications are thought to be critical modulators in cell cycle control and DNA-damage responses (18). The synthesis of the lagging-strand follows a similar mechanism as that in *E. coli*, with the eukaryotic Okazaki fragments being only 100–200 bases, much shorter than the 1–2 kb in *E. coli* (19). The RNA primers are hydrolyzed by either DNA2 nuclease or FEN1. The gaps between Okazaki fragments are then fixed by DNA ligase I.

1.2 Sliding Clamp

Sliding clamps tether polymerases to DNA templates, drastically increasing both the length of DNA being synthesized per association event and the rate of DNA synthesis. Although sliding clamps are largely viewed as accessory or scaffold proteins, they also play critical roles in

cell cycle control and DNA repair checkpoints by coordinating numerous proteins involved in signaling pathways.

1.2.1 Structural features of sliding clamps

Sliding clamps from different domains of life do not share much sequence identity; however, they all adopt the toroid shape structure that could encircle the dsDNA. Bacterial clamps are ring-shaped homodimers, such as the β-clamp in *E. coli* (20), PCNA eukaryotic clamp (21,22), gp45 clamps from T4 bacteriophage (23) and most archaeal PCNA clamps are homotrimers (24-26). On contrary, the Rad9-Hus1-Rad1 checkpoint clamp and the clamp from the archaebacteria Sulfolobus solfataricus are heterotrimeric complexes (27,28). Regardless of the subunit composition difference and the little sequence homology over sliding clamps from different species, the overall structures of them are substantially similar. Each clamp consists of six domains held together through non-covalent interactions between adjacent domains. Interdomain connector (IDC) loops, positioned on the outer surface of the ring structure serve as linkers between domains. Each clamp subunit contains three domains in the case of dimeric clamps such as β-clamp, and contains two domains in trimeric clamps such as PCNA. In the clamp loading process, the subunit-subunit interfaces formed by the non-covalent interaction between neighboring subunits need to be disrupted before template DNA can be threaded in.

Subunits within a clamp are arranged in a head-to-tail fashion so that each subunit adopts the same directionality around the rotational axis of symmetry. In addition, this arrangement creates two surfaces with distinct structural features on the surface of the planar ring. It's been shown that all proteins known to interact with a sliding clamp associate with the same clamp surface that has the IDC loops on it, binding the IDC loops through a clamp-binding motif (which features a cluster of hydrophobic residues) in both prokaryotes and eukaryotes (29,30). Sequence

alignments and binding assays suggest a consensus binding sequence for PCNA has been proposed to be Qxx(I/L/M)xxF(F/Y) for PCNA (31,32) and QLxLF (33) or QxxL(x)F (34) for *E. coli* β-clamp based on sequence alignments and binding assays. These motifs are capable of binding to a hydrophobic patch on the surface of IDC loops of the clamps (29,30). Each clamp subunit contains one such binding pocket, so there are multiple binding pockets for either PCNA or *E. coli* β-clamp. However, steric hindrances prevent the clamp from binding to multiple interactive partners. For example, clamp loader dissociation is prerequisite to polymerase-sliding clamp interaction.

PCNA is known to interact with a variety of protein partners involved in DNA replication and repair, translesion DNA synthesis, chromatin structure remodeling, cell cycle control and a variety of metabolic pathways by coordinating the functions of various proteins (35-38). The structural basis for the broad array of PCNA-interacting proteins is a weak consensus sequence binding motif. PCNA's protein association versatility is exemplified by specific RFC clamp loader binding. It's been known that an alternative RFC clamp loaders in which one subunit is replaced by another protein that specializes in different function; the alternative clamp loader is capable of load the checkpoint clamp (9-1-1) instead of PCNA at specific target sites (39,40).

Interactions between the DNA backbone and the side chains lining the PCNA pore are primarily non-specific electrostatic interactions between phosphorylate group and positively charged residue side chains, providing an interface conducive to sliding at physiological ionic strength.

1.2.2 Clamp loading process

An opening and resealing process of a clamp is required for loading it onto DNA. This process is facilitated by a clamp-loader, which is a multi-protein clamp-loading ATPase. In eu-

karyotic cells, the replication factor C (RFC) is the clamp loader, consisting of five different homologous proteins that are in turn homologous to the γ and δ' subunits of *E. coli* γ-complex, the prokaryotic analog of RFC (29,41). All subunits of RFC are members of the AAA+ protein family, and they are arranged in a circle (42,43). Common sequence motifs among these clamp-loader AAA+ proteins have been termed RFC boxes.

In *E. coli*, the β-clamp enables the distributive DNA polymerase III to become highly processive, remaining bound with DNA template through the extensive reactions of incorporating dNTPs during DNA replication. The γ-complex binds to the β-clamp, triggering ring opening, which then leads to DNA association. ATP binding is necessary for triggering the clamp opening and DNA template positioning, and the subsequent ATP hydrolysis reseals the clamp (44). This loading process occurs only once per round of replication on the leading strand, while multiple loading events are necessary on the lagging strand due to the nature of the Okazaki fragments. Finally, the clamp dissociates from the replication machinery upon reaching the 5' end of the pre-existing daughter DNA.

In eukaryotic cells, RFC loads PCNA onto a 3′ primer/template junction and then dissociates upon ATP binding, freeing PCNA for association with Pol δ (3). The pentameric arrangement of the RFC subunits was originally proposed based on sequence and subunit interaction similarities to *E. coli* γ-complex, and was later proven by structural studies (29,45). The five RFC subunits in *S. cerevisiae* are referred to as RFC1 through RFC5. RFC1 resembles the δ subunit of *E. coli* γ-complex in the conserved clamp-interacting residues and its position in the complex. RFC2–4 share characteristics of the γ subunit of *E. coli* γ-complex in forming a trimeric ATPase subassembly. The positioning of RFC5 in RFC complex resembles that of the δ' subunit in *E. coli* γ-complex. RFC5 contains an SRC motif (a conserved serine-arginine-cysteine motif

present in all known clamp loaders) but lacks the consensus phosphate-binding loop (P-loop). The ATP binding sites of RFC or γ-complex ($\gamma_3\delta\delta'$) are located at subunit-subunit interfaces, while RFC contains four competent ATPase sites as opposed to three since RFC1 also binds ATP whereas δ does.

The PCNA loading process is highly sequential. RFC binds two ATP molecules before associating with PCNA. RFC then binds a third ATP before threading the primer/template DNA through the PCNA pore. The DNA binding triggers ATP hydrolysis followed closely by PCNA resealing and RFC dissociation (Figure 1.1) (46). Mutagenesis studies of the residues involved in ATP binding have illustrated that all four consensus ATP binding sites are critical for RFC activity (47,48). P-loop region mutants retain the ability to bind PCNA but are incapable of binding DNA (47). A number of crystal structures of RFC-PCNA complexes have provided detailed information on how clamp loaders interact with clamps, greatly enhancing our understanding of clamp loading mechanisms (29,49).

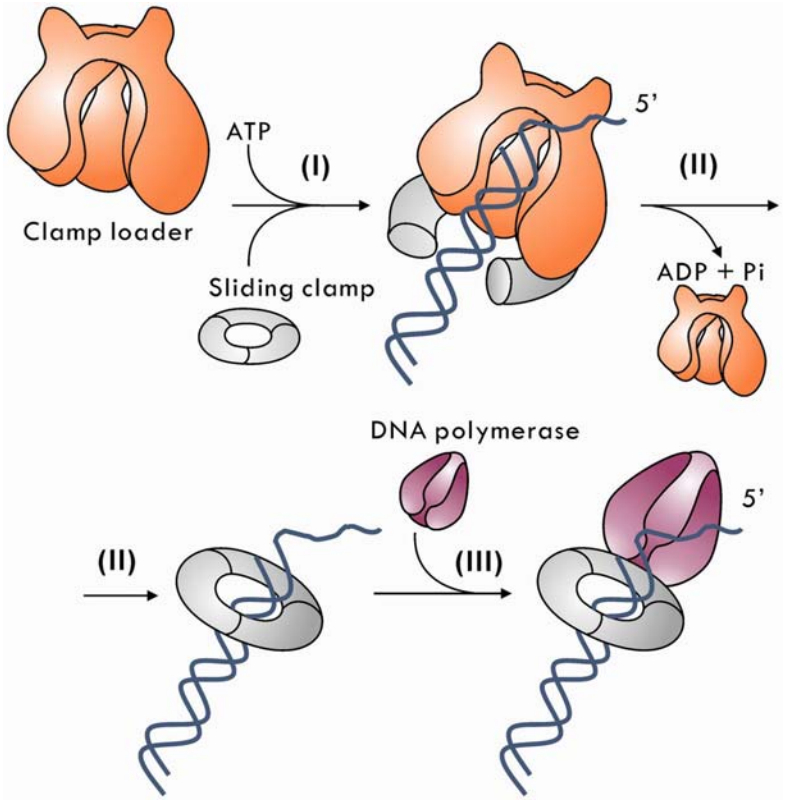

Figure 1.1 Clamp-loading mechanism and subsequent exchange of the clamp loader with DNA polymerase on the sliding clamp.

In the presence of ATP, the clamp loader binds and opens the ring-shaped DNA sliding clamp for its loading on the DNA. The clamp loader then hydrolyzes ATP before its dissociation from the clamp–DNA complex. In the final step, DNA polymerase binds to the clamp to begin DNA replication.

CHAPTER 2. METHODS

2.1 Molecular Dynamics

Since the first application of Molecular Dynamics (MD) simulation to protein in 1977 (50), MD simulation has been vastly developed and become an established powerful tool in studying biological macromolecular systems, probing the structural, dynamic, kinetics, thermodynamic, and other time dependent features (51-53). The successful applications of MD simulations have covered many studies on bimolecular stability, conformational and allosteric properties, dynamic properties, molecular recognition and the properties of complexes, ion and small molecule transport, protein lipid interaction etc. (54). With MD simulations, detailed dynamic atomistic level information can be obtained to facilitate the interpretation of experimental data to gain access to information not directly accessible or inaccessible by experiments. Therefore, it is a valuable tool in advancing our knowledge of many biomolecular systems.

The motions of the particles in a MD simulation obey Newton's laws of motions so the electronic degrees of freedom are not explicitly handled in a MD system. Therefore, MD simulations are only applicable in describing the bond and angle distortions instead of any events that involve bond breaking/formation, isomerizations, or excited states (55). In general, a MD simulation involves the following stages:

1. Obtaining the initial coordinates and assigning the velocities to the atoms in the MD system. Available structural information of biomolecules is provided through x-ray crystallography or NMR studies and can be retrieved from the PDB databank. Initial velocities of the atoms are usually assigned by Maxwell-Boltzmann distribution.

2. Computing the instantaneous forces exerted on the particles in the system and numerically integrating the Newton's equation of motion using small finite timestep to evolve the sys-

tem along a specified timescale. This is conducted iteratively to accumulate the ensemble that to be analyzed.

3. Analyzing the resulting ensemble (often refer to trajectory) to retrieve the desired time-dependent properties mentioned above.

2.1.1 Amber Force Field

The intra- and inter- molecular forces experienced by the atoms are computed with an empirical force field in a MD simulation. The empirical force field, also known as the potential energy function, allows for the calculation of the potential energy (V) as a function of the coordinates of the system in a 3-dimensional space (R). The potential energy function is an additive function consists of several energy terms describing bonded and non-bonded interactions between atoms.

$$V(R)_{total} = V(R)_{intramolecular} + V(R)_{intermolecular} \qquad (2.1)$$

The bonded interactions terms include bonds, bond angels and dihedral angles; the non-bonded interactions contain Coulombic and Lennard-Jones interactions. The bond lengths and bond angels are both considered independent, and the magnitude of atomic partial charges are fixed so that the potential energy function is purely additive. In a carefully designed force field with all the finely tuned parameters primarily obtained by experimental and/or quantum mechanical studies, the harmonic approximations for each energy term are sufficiently accurate for biomolecular simulations due to the fact that the simulations of biological systems are typically carried out at physiological temperatures where the bond lengths and angles only varies very little about the equilibrium values. Many force fields have been specifically developed for biomolecule simulations, including CHARMM22 (56-59), AMBER force field (60) as the most prevail-

ing ones. In particular, the AMBER force field is used in this dissertation, with the following functional form:

$$U = \sum_{bonds} K_r\left(r - r_{eq}\right)^2 + \sum_{angles} K_\theta\left(\theta - \theta_{eq}\right)^2 + \sum_{dihedrals} K_\phi\left[1 + \cos\left(n\phi - \delta\right)\right]$$
$$+ \sum_{impropers} K_\varphi\left(\varphi - \varphi_{eq}\right)^2 + \sum_{electrostatic} \frac{q_i q_j}{\varepsilon_l r_{ij}} + \sum_{vdw}\left[\varepsilon_{ij}\left(\frac{R_{ij}^{\min}}{r_{ij}}\right)^{12} - 2\left[\frac{R_{ij}^{\min}}{r_{ij}}\right]^6\right] \quad (2.2)$$

where the first four terms are the intramolecular terms, and the last two terms are the intermolecular terms. K_r, K_θ, K_ϕ and K_φ are the force constants of bond length, bond angle, dihedral angle and improper angle energy term, respectively. Accordingly, r, θ, ϕ, φ are the instantaneous values of the bond length, bond angel, dihedral angel and improper angle. The subscript "eq" stands for the equilibrium values for each type of interaction. ε_l is the effective dialectic constant. q_i and q_j represent the partial charge of atom i and j, respectively. In the dihedral term, n is the multiplicity, ϕ is the instantaneous torsion angle, and δ is the phase shift. The first intermolecular term is the Lennard-Jones term, where R_{ij}^{min} is the minimum interaction radius (dependent on the interacting atoms and determined by experiment), ε_{ij} is the Lennard-Jones well depth, $(\frac{1}{r_{ij}})^6$ describes the attractive interactions, and $(\frac{1}{r_{ij}})^{12}$ describes the repulsive interactions. All atoms including hydrogens in AMBER force field are treated explicitly. The parameter sets are optimized from experimental measurements, such as vibrational spectra, torsional energy surfaces, experimental heats and free energies of vaporization, solvation and sublimation, dipole moments and molecular volume etc., supplemented with *ab initio* calculations, so that they are largely empirical. AMBER force field for the most part is sufficient to model general biomolecules with its comprehensive parameter sets including proteins, nucleic acid, li-

pid, and carbohydrates with consistency between these sets (60). However, it is likely that we need extra or modified parameters for certain particular purpose or when the regular sets lead into problems such as when we need to use the PARMBSC0 in our simulations to improve the DNA performance in my dissertation studies (61).

2.1.2 Integration of Newton's equation

Once the initial structure of the system is obtained and the initial velocities of all the atoms have been assigned, the evolution of the system can be determined by integrating Newton's second law of motion:

$$F = ma \tag{2.3}$$

where F is the force exert of an atom, m is the atom mass, and a is the acceleration. Alternatively, the force can be described in the form of the gradient of the potential energy function:

$$F = -\nabla U\left(R\right) = -\frac{dU}{dR} \tag{2.4}$$

As the acceleration can be expressed as the second order derivative of position with respect to time, we can now relates the derivative of the potential energy with respect to coordinates to the changes in position as a function of time:

$$-\frac{dU}{dR} = m\frac{d^2R}{dt^2} \tag{2.5}$$

The MD simulation system evolves on the basis of the equation 2.5, the Newtonian equation. Numerical integration is done with finite difference methods using small time interval, which is typically referred to "timestep", Δt, in MD simulations. Once the acceleration is obtained, we can derive the new positions and velocities at time $t + \Delta t$ using the positions and velocities at time t. By iterating this step, we can keep updating the system till a specified time point. Since the highest frequency motions in a MD system is the stretching vibrational motions of the heavy

atom-hydrogen bond, which occurs on a ~ 10 fs timescale, the Δt is then empirically chosen to be ~ 1/10 of this scale (1 fs) to accurately describe the fastest degrees of freedom of the system (62). To improve computational efficiency, algorithms including the SHAKE were developed to enable using timestep of 2 fs (63,64). By applying SHAKE algorithm, all the covalent hydrogen bonds in the system are constrained in order to remove the highest frequency motions, this procedure is validated by the fact the these highest frequency motions are mostly of the least interest comparing with other low frequency motions.

One of the representative integrators that have been implemented to conduct the integration is the Verlet algorithm (65), in which two Taylor expansion of the atom position as a function of time were utilized to obtain the updated atom position:

$$r_{n+1} = r_n + v_n \Delta t + \frac{1}{2}\left(\frac{F_n}{m}\right)\Delta t^2 + \mathrm{O}(\Delta t^3) \qquad (2.6)$$

$$r_{n-1} = r_n - v_n \Delta t + \frac{1}{2}\left(\frac{F_n}{m}\right)\Delta t^2 - \mathrm{O}(\Delta t^3) \qquad (2.7)$$

where r_n is the position at step n, r_{n+1} is the position at next step. $\mathrm{O}(\Delta t^n)$ is the term of order n or higher as a function of Δt. Adding equation 2.6 and 2.7 gives the Verlet algorithm for updating the atom positions:

$$r_{n+1} = 2r_n - r_{n-1} + \left(\frac{F_n}{m}\right)\Delta t^2 + \mathrm{O}(\Delta t^4) \qquad (2.8)$$

With equation 2.8, the r_{n+1} can be obtained by using the current position r_n, the previous position r_{n-1}, and the current force F_n acting on the atom. This calculation is done for all the atoms in the system for each timestep over the simulation. Note that the velocity of each atom is not explicitly involved in the algorithm, but it can be easily computed by subtracting equation 2.8 from 2.7:

$$v_n = \frac{r_{n+1} - r_{n-1}}{2\Delta t} + \mathrm{O}(\Delta t^3) \qquad (2.9)$$

A different version of the integrator is the leapfrog algorithm (66), in which a modification was made to reduce the errors over the velocity propagation when using Verlet algorithm. Instead of determine the velocity at the end point of a timestep, it was evaluated at the midpoint, therefore, the algorithm becomes:

$$v_{n+1/2} = v_{n-1/2} + \left(\frac{F_n}{m}\right)\Delta t \qquad (2.10)$$

$$r_{n+1} = r_n + v_{n+1/2}\Delta t \qquad (2.11)$$

where $v_{n\pm1/2}$ is the velocity at the midpoint of a timestep, $t \pm (\frac{1}{2})\Delta t$. With leapfrog algorithm, we firstly also calculate the current force F_n. The force and the previous midpoint velocity were then used to compute the next midpoint velocity with equation 2.10. Finally, the next step position can be obtained by using equation 2.11.

2.1.3 Non-bonded interactions

For an explicit solvent biomolecular MD simulation system of thousands of atoms, the calculation of non-bonded terms turns to be quite computationally expensive since they include many more interactions than the bonded interactions for each atom, and this is even more problematic for the electrostatic terms. The number of bonded terms in a given force field is proportional to the number of atoms in the system, N, while the non-bonded terms exist between every pair of atoms so the number increases on the order of N^2. The attractive and repulsive forces of the Lennard-Jones potential decay rapidly as the distance increases: at distance of 2.5 σ, it decreases to only ~ 1% of the value at σ (the force between the two interacting particles is zero at distance of σ). However, the electrostatic energy does not decay as rapidly as the Lennard-Jones potential does; it is proportional to the reciprocal of the distance between two point charges. A

variety of algorithms have been developed to reduce the computational cost on calculating the nonbond terms.

2.1.3.1 Cutoffs

Cutoff methods are commonly applied when computing the Lennard-Jones interaction energies. These methods calculate the interaction energies between all pairs of atoms within a specified threshold, which is known as the cutoff distance. The cutoff value implemented for biomolecular systems is generally set to 10-12 Å (67).

An abrupt truncation at the cutoff distance introduces discontinuities in the potential energy function and would cause instability in simulation if not fixed. A few approaches have been developed to circumvent this problem. One is to use a shifted potential function, in which a constant term is subtracted from the original potential at all values so that the potential at cutoff value is of zero and the continuity of the function is maintained. Another way is to use a polynomial switching function $S(r)$ that modifies the non-bonded energy smoothly over a buffer window of typically 1-2 Å before the cutoff point so that the potential energy gradually approaches to zero over this region and remains the same values before the buffer region (68).

2.1.3.2 Long-range electrostatic interaction

Cutoff methods used to be applied to treat electrostatics interactions as well to reduce the computational cost. However, artifacts were found in the simulation that involves substantial localized charges, such as the distortion of DNA and instability of α-helices when using cutoffs for electrostatic interactions (69). As a matter of fact, the long-range electrostatic interactions are important in maintaining protein structural integrity. Therefore, more accurate way of treating the electrostatic interactions becomes necessary. The most popular way to handle the long-range

interaction is the Ewald summation (70), which was developed by Paul Peter Ewald in the 1920s to compute the interaction energy in crystals.

In the Ewald method, a particle interacts with all other particles in the unit cell and with all of their images in the periodic system. The electrostatic contribution to the potential energy for all pairs of charges in the central cell is given by:

$$V = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \qquad (2.12)$$

where N is the number of atoms in the system, q_i, q_i are the partial charges on atom i and j respectively, ε_0 is the dielectric constant, r_{ij} is the minimum distance between atoms i and j, and $1/2$ is to avoid double-counting. The charge-charge contribution between the atoms in the central cell and all the atoms in the images of the periodic system is given by:

$$V = \frac{1}{2}\sum_{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j}{4\pi\varepsilon_0 \left|r_{ij}+n\right|} \qquad (2.13)$$

where L is the length of the central unit box, n is a vector whose components are integral multiples of L ($n = n_x L, n_y L, n_z L$). The Ewald method breaks the summation into two series that both of them converge much faster than the right-hand-side of equation 2.14, which converges very slowly. This conversion is based on the following mathematical foundation:

$$\frac{1}{r} = \frac{f(r)}{r} + \frac{1-f(r)}{r} \qquad (2.14)$$

The choice of the $f(r)$ has to consider the rapid variation of electrostatic interactions when r is small and the slow decay of electrostatic interactions when r is large. The first term is for real space and is short-ranged, and the second term is for reciprocal space and is a long-ranged term (this term can be Fourier-transformed). The Ewald sum uses the complementary error function $erfc(r)$ as $f(r)$.

$$erfc(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} \exp(-t^2) dt \qquad (2.15)$$

In the Ewald sum, each charge is considered to be surrounded by a neutralizing charge distribution which is of equal magnitude but of opposite sign. This is commonly a Gaussian distribution of the form:

$$\rho_i(r) = \frac{q_i \alpha^3}{\pi^{(3/2)}} \exp(-\alpha^2 r^2) \qquad (2.16)$$

The point charges summation is now the summation of the interactions between charges and the neutralizing Gaussian distributions. The real space part of Ewald sum is of the following form:

$$V = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} {\sum_{|n|=0}}' \frac{q_i q_j}{4\pi\varepsilon_0} \frac{erfc(\alpha |r_{ij} + n|)}{|r_{ij} + n|} \qquad (2.17)$$

The summation with the prime indicates that the series does not include the interaction $i = j$ at $n = 0$. The new summation with the error function converges rapidly and the value is considered negligible beyond some cutoff distance. The rate of convergence depends on α (the width of the Gaussian distribution), and the wider it is, the faster the sum converges.

A background charge distribution is then added to counteract the neutralizing Gaussian distribution. This summation is conducted in the reciprocal space sum and is a Fourier series:

$$V = \frac{1}{2} \sum_{k \neq 1} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\pi L^3} \frac{q_i q_j}{4\pi\varepsilon_0} \frac{4\pi^2}{k^2} \exp\left(-\frac{k^2}{4\alpha^2}\right) \cos(k \cdot r_{ij}) \qquad (2.18)$$

where k are reciprocal vectors. The reciprocal space sum also converges much more quickly than the original point charges sum. The only problem is that the reciprocal space sum converges more quickly for a small α rather than a large α that is favorable for the real space sum. In addition, the number of terms included in the reciprocal sum increases as α does, result-

ing in higher computational cost. A value of α of $5/L$ and 100-200 reciprocal vectors k have been shown to provide acceptable results. The real space sum of the neutralizing Gaussians includes the interaction of each Gaussian function with itself. Therefore, the following term must be subtracted:

$$V = -\frac{\alpha}{\sqrt{\pi}} \sum_{k=1}^{N} \frac{q_k^2}{4\pi\varepsilon_0} \tag{2.19}$$

Another correction term is also required if the simulation is conducted in vacuum, which has the relative permittivity of 1:

$$V = -\frac{2\pi}{3L^3} \left| \sum_{i=1}^{N} \frac{q_i}{4\pi\varepsilon_0} r_i \right|^2 \tag{2.20}$$

In the end, the full Ewald summation is given by:

$$V = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left\{ \begin{array}{l} \sum_{|n|=0}^{\infty}{}' \frac{q_i q_j}{4\pi\varepsilon_0} \frac{erfc(\alpha |r_{ij}+n|)}{|r_{ij}+n|} \\ + \sum_{k \neq 0} \frac{1}{\pi L^3} \frac{q_i q_j}{4\pi\varepsilon_0} \frac{4\pi^2}{k^2} \exp\left(-\frac{k^2}{4\alpha^2}\right) \cos(k \cdot r_{ij}) \\ - \frac{\alpha}{\sqrt{\pi}} \sum_{k=1}^{N} \frac{q_k^2}{4\pi\varepsilon_0} + \frac{2\pi}{3L^3} \left| \sum_{i=1}^{N} \frac{q_i}{4\pi\varepsilon_0} r_k \right|^2 \end{array} \right\} \tag{2.21}$$

In general, Ewald summation uses the combination of real and reciprocal space sums of interaction energies to evaluate long-range electrostatic energies. It is computationally expensive to implement as it scales to N^2); however, methods have been devised to cut back on computational cost (71). By optimizing α, the balance between the sums will reduce the Ewald sum to the complexity of $N^{3/2}$, which is still quite expensive for large biomolecular systems. The particle mesh Ewald sum (PME) developed by Darden and coworkers reduces the Ewald sum to $Nlog(N)$ (72). In the PME method, the trigonometric function values of the Fourier series present in the reciprocal space sum are evaluated by a smooth approximation of the potential over a

grid, or mesh. The smoothing function is known as the Euler spline, which expresses the value of the trigonometric function at the actual charge coordinates in terms of the charge value at neighboring grid points. The resulting sums over the grid points are then evaluated efficiently by a fast Fourier transform. The PME method is implemented in the NAMD simulation package (73,74), which is used intensively throughout this thesis. It's worth mentioning that there are still internal limitations in the treatment of electrostatics due to the basic assumption of "fixed-charge" model used in most of the current force fields, in which each atom is assigned by a single value for the atomic charge that is not influenced by the local electrostatic environment. For a more practical matter, electronic polarization in current force field is accounted for in an average way to produce reasonable agreement with experimental measurements. Advances such as "fluctuating-charge" (75,76) and dipole model (77) to take polarization into account are underway.

2.1.4 Solvent model and boundary conditions

Since the stability and functions of biomolecules are sensitive to the environment, a reasonable representation of the solvent environment of these biomolecule models in simulations becomes crucial for characterizing the properties of the solute. As a result, biomolecular simulations have to be conducted in certain solvent environment in accordance to the reality.

With implicit solvent model, the macromolecule interacts only with itself, but the electrostatic interactions are modified to account for the solvent. All solvent effects are contained in the dielectric constant ε (78,79). For example, $\varepsilon = 1$ for vacuum, $\varepsilon = 1{\sim}20$ for internal region of protein, $\varepsilon = 78{\sim}80$ for water solvent.

With explicit solvent model, the macromolecule is surrounded by solvent molecules (water, ions), with which the macromolecule interacts. Specific nonbond interactions are calculated explicitly using the force field potential energy functions. Over the years, the explicit solvent

models have been made possible by current understanding on it and the growing computing resources (80). A particular example is the TIP3P (81) water model, which is a rigid water model with each atom treated explicitly. This water model is used to study all the protein or protein/DNA complex models in the thesis.

To recover the macroscopic properties from a MD simulation with limited number of solute and solvent molecules, we have to consider using the boundary conditions. A conventional way to simulate an infinite system is to use the periodic boundary condition, in which the solute and solvent of the simulated system are confined in a unit cell that has an infinite number of images in all three dimensions (55). Only the molecules in the unit cell are considered explicitly during the simulation. However, if an atom in a unit cell leaves the cell and enters an image unit, an image atom would enter to replace it to maintain the number of atoms in the system. Practically, a minimum image convention is usually applied so forces are only calculated with the closest periodic unit cell that the computational cost would be limited. Different shapes of the periodic cell are available, including the cubic cell, rectangular cell, and spherical cell. The choice of the shape of the periodic cell is dependent on the underlying geometry of the simulation system.

2.1.5 Temperature control

The total energy is conservative in conventional MD simulations. Therefore, the time averages from MD simulation are equivalent to ensemble averages computed from the NVE ensemble (constant number of particles, constant volume, and constant total energy; also known as microcanonical ensemble) given the simulation is long enough. The flexibility of MD is greatly enhanced by having techniques extend MD simulation to the NVT (constant number of particles, constant volume, and constant temperature) or NPT (constant number of particles, constant pressure, and constant temperature) ensembles as well. To control the temperature in MD simulation,

there are essentially three ways: scaling velocities, adding stochastic forces and/or velocities, and using “extended Lagrangian” formalisms (68).

2.1.5.1 Velocity Scaling: Isokinetics and the Berendsen Thermostat

Thy temperature of the canonical ensemble (NVT) is related to the time average of the kinetic energy, which is given by:

$$\frac{3}{2}Nk_BT = \frac{1}{2}\sum_{i=1}^{N} m_i v_i^2 \qquad (2.22)$$

In simple velocity scaling, the velocities are chosen randomly from a range (for example, [-0.5, 0.5]) and could be then rescaled to result in a desired temperature with equation 2.22. It could be converted to a dynamic scheme with which the velocities are continuously scaled so that the total kinetic energy is conserved with the isokinetic thermostat. Let the temperature immediately after a velocity update be T_i, and the instaneous temperature be T. If the velocities were scaled by a constant λ, we can calculate the associated temperature change as:

$$\Delta T = \frac{1}{2}\sum_{i=1}^{N}\frac{2}{3}\frac{m_i\left(\lambda v_i\right)^2}{Nk_B} - \frac{1}{2}\sum_{i=1}^{N}\frac{2}{3}\frac{m_i v^2}{Nk_B} = \left(\lambda^2 - 1\right)T \qquad (2.23)$$

where the scaling factor $\lambda = \sqrt{T / T_i}$. The isokinetic thermostat is not applicable in in the canonical ensemble simulation, but is good for the initialization or heating-up stage of a MD simulation. This velocity rescaling can be performed at each or every few steps.

Another popular velocity scaling thermostat is that of Berendsen (82), in which the system is coupled to an external heat bath that can be fixed at the desired temperature. The bath provides or absorbs heat from the simulation system. The velocity is such scaled that temperature change rate is proportional to the temperature difference between the bath and the system.

$$\frac{dT_i}{dt} = \frac{1}{\tau}\left(T_{bath} - T_i\right) \qquad (2.24)$$

where τ is a constant called the "rise time" of the thermostat. It describes the strength of the coupling of the system to a hypothetical heat bath. The larger the τ is, the weaker the coupling. The change in temperature between successive timesteps is:

$$\Delta T = \frac{\Delta t}{\tau}\left(T_{bath} - T_i\right) \tag{2.25}$$

The velocity scaling factor is then:

$$\lambda = \left[1 + \frac{\Delta t}{\tau}\left(\frac{T_{bath}}{T_i} - 1\right)\right]^{\frac{1}{2}} \tag{2.26}$$

where T_{bath} can be set to the desired temperature, Δt is the integration time step, a larger τ means it takes longer time to reach to a prescribed temperature. Regardless of the simple implementation and efficiency, velocity scaling thermostats are not commonly used in production stages of MD simulation since they do not strictly conform to the canonical ensemble (83).

2.1.5.2 Stochastic NVT Thermostats

The Andersen thermostat (84) introduces a stochastic element to the temperature by having random collisions of molecules with an imaginary heat bath at the desired temperature. In the single-particle approach, a random particle is chosen and its velocity is drawn from the Maxwell-Boltzmann distribution at the desired temperature at each step:

$$p(v_{ix}) = \left(\frac{m_i}{2\pi k_B T}\right)^{1/2} \exp\left[-\frac{1}{2}\frac{m_i v_{ix}^2}{k_B T}\right] \tag{2.27}$$

where $p(v_{ix})$ is the probability that an atom i of mass m_i has velocity v_{ix} in the x direction at a temperature T , k_B is the Boltzmann constant. This applies to each component of the particle's velocity. Alternatively, a massive collision can be used in which each velocity component of

every particle is reassigned simultaneously. The time between collisions, or the number of collisions in some (short) time interval is decided randomly, with the following Poisson distribution:

$$p(t) = ce^{-ct} \tag{2.28}$$

where c is the stochastic collision frequency. The system evolves at constant energy between collisions. In practice, if c is too high the temperature control algorithm dominates and the system does not fluctuates in kinetic energy; if the collision rate is too low the system does not sample from a canonical distribution of energies. An optimal choice has been found:

$$c = \frac{2a\kappa V^{1/3}}{3k_B N} = \frac{2a\kappa}{3k_B \rho^{1/3} N^{2/3}} \tag{2.29}$$

where a is a dimensionless constant, κ is the thermal conductivity, V is the volume, k_B is the Boltzmann constant, and ρ is the number density of particles, $\rho = N/V$. Andersen thermostat has been shown to generate correct NVT probabilities in the limit of infinitely long trajectory averaged over many heat bath collisions. Nevertheless, it should be noted that the random collisions decorrelate the velocities of particles faster than the NVE dynamics. In other words, the particles would lose memory of their velocities from some previous time. As a result, Andersen thermostat does not preserve true molecular kinetics.

Another scheme is the Langevin thermostat (85), in which the particles experience a random force and a constant friction. The average magnitude of the random forces and the friction are related to obey the "fluctuation-dissipation theorem" (86), thereby guarantees the NVT statistics. The Newtonian's equation becomes the Langevin's equation as:

$$m\ddot{r}_i = -\nabla_i U - m\Gamma\dot{r}_i + W_i(t) \tag{2.30}$$

where U is the particle interaction potential, Γ is a friction coefficient, and the dot is a time derivative ($\dot{r}$ is the velocity and $\ddot{r}$ is the acceleration). The random force $W_i(t)$ is determined by a

Gaussian distribution to add kinetic energy; the variance of $W_i(t)$ is a function of the prescribed temperature and timestep. In this way, the random force is balanced with the frictional force to maintain the temperature.

2.1.5.3 Nosé-Hoover thermostat

In the Nosé-Hoover thermostat (87,88), a heat bath is considered as an integral part of the system. An artificial coordinate $\tilde{s}$ (identical to the real system), associated "mass" Q ($Q > 0$), and velocity $\dot{\tilde{s}}$ are assigned to the heat bath. The magnitude of Q determines the coupling strength between the bath and the real system and hence, the temperature fluctuation. $\tilde{s}$ is a time-scaling parameter, which stretches the timescale in the extended system by the factor of $\tilde{s}$. The real and virtual variables are related as following:

$$d\tilde{t} = \tilde{s}dt \tag{2.31}$$

$$\tilde{r} = r \tag{2.32}$$

$$\dot{\tilde{r}} = \dot{r} / \tilde{s} \tag{2.33}$$

$$\tilde{s} = s \tag{2.34}$$

$$\dot{\tilde{s}} = \dot{s} / \tilde{s} \tag{2.35}$$

For the extended system, the Lagrangian is:

$$L = \sum_i \frac{m_i}{2}\tilde{s}^2\dot{\tilde{r}}^2 - U(\tilde{r}) + \frac{1}{2}Q\dot{\tilde{s}}^2 - gk_B T_0 \ln\tilde{s} \tag{2.32}$$

The first term of the right hand side is the kinetic energy, and the second is potential energy of the real system. The rest terms are the kinetic energy the potential of $\tilde{s}$. The potential is chosen to ensure producing a canonical ensemble. With Nosé-Hoover formalism, $g = N_{df}$ (number of degree of freedom) for real-time sampling, while with Nosé-formalism, $g = N_{df} + 1$

for virtual-time sampling. The equations of motion is therefore given by the following two equations:

$$\ddot{\tilde{r}} = \frac{\tilde{F}_i}{m_i \tilde{s}^2} - \frac{2\dot{\tilde{s}}\dot{\tilde{r}}_i}{\tilde{s}} \tag{2.34}$$

$$\ddot{\tilde{s}} = \frac{1}{Q\tilde{s}}\left(\sum_i m_i \tilde{s}^2 \dot{\tilde{r}}_i^2 - g k_B T_0 \right) \tag{2.35}$$

In the extended system $(\tilde{r}, \tilde{p}, \tilde{t})$, a microcanonical ensemble is sampled by these equations while the real system energy oscillates. Heat transfers between the real system and the heat bath occurs when $\tilde{s}$ fluctuates, thereby, regulates the temperature of the real system. It can be shown that a canonical ensemble for the real system can be sampled by the equations. The Nosé equations of motion can be reformulated using real system variables with the following transformations:

$$\dot{s} = \tilde{s}\dot{\tilde{s}} \tag{2.36}$$

$$\ddot{s} = \tilde{s}^2 \ddot{\tilde{s}} + \tilde{s}\dot{\tilde{s}}^2 \tag{2.37}$$

$$\dot{r} = \tilde{s}\dot{\tilde{r}} \tag{2.38}$$

$$\ddot{r} = \tilde{s}^2 \ddot{\tilde{r}} + \tilde{s}\dot{\tilde{r}}^2 \tag{2.39}$$

Given $\gamma = \dot{s} / s$, the Lagrangian equations of motion can then be re-written with:

$$\ddot{r}_i = \frac{F_i}{m_i} - \gamma r_i \tag{2.40}$$

$$\dot{\gamma} = \frac{-k_B N_{df}}{Q} T(t) \left(\frac{g}{N_{df}} \frac{T_0}{T(t)} - 1 \right) \tag{2.41}$$

Q determines the coupling strength and thereby the temperature control. If Q is too large (loose coupling), the canonical distribution can only be recovered with very long simulation; if it is too

small (tight coupling), high-frequency temperature oscillation would be introduced. The Nosé equations of motion can be expressed with a more intuitive choice for the coupling strength:

$$\dot{\gamma} = -\frac{1}{\tau_{NH}}\left(\frac{g}{N_{df}}\frac{T_0}{T(t)} - 1\right) \qquad (2.42)$$

with the effective relaxation time being:

$$\tau_{NH}^2 = \frac{Q}{N_{df} k_B T_0} \qquad (2.43)$$

By computing the oscillation frequency for small deviations $\delta\tilde{s}$ from the average $\langle\tilde{s}\rangle$, the relaxation time can be approximated.

2.1.6 Pressure control

In an isothermal-isobaric ensemble, the volume is a dynamic variable that fluctuates in the simulated system. The pressure can be calculated using the Clausius virial theorem:

$$\mathrm{P} = \frac{2}{3V}\left(E_{kin} - \Xi\right) \qquad (2.44)$$

where V is the box volume, E_{kin} is the kinetic energy, and Ξ is the inner virial for additive pairwise interactions, the explicit form of it is:

$$\Xi = \frac{1}{2}\sum_{i<j} r_{ij} \cdot f\left(r_{ij}\right) \qquad (2.45)$$

where $f\left(r_{ij}\right)$ is the force between particles *i* and *j* at a distance r_{ij}. While the pressure in a periodic system a second order tensor instead of a scalar, it can still be expresses as a scalar in a isotropic system:

$$P = Tr(\mathrm{P})/3 \qquad (2.46)$$

As a common procedure for many types of barostats, the pressure can be controlled through adjusting the inner virial through scaling the distances of inter particle. Barostats with constant box shape (isotropic barostats) includes Berendsen barostats (82), Andersen barostats (84), Nosé-Hoover barostats (87,88), etc. Only two isotropic barostats will be described in the following.

2.1.6.1 Berendsen barostat

With Berendsen barostat, the simulation system is weakly coupled to an external "pressure bath". The rate of change of pressure is given by:

$$\left(\frac{dP(t)}{dt}\right)_{bath} = \frac{P_{bath} - P(t)}{\tau_p} \qquad (2.47)$$

where τ_p is the time constant for the coupling, P_{bath} is the pressure of the "bath", and $P(t)$ is the instantaneous pressure at time t. A proportional coordinate scaling factor α is introduced to the equations of motion:

$$\dot{x} = v + \alpha x \qquad (2.48)$$

So the corresponding volume change is:

$$\dot{V} = 3\alpha V \qquad (2.49)$$

Since the pressure change can be related to the isothermal compressibility β:

$$\frac{dP(t)}{dt} = -\frac{1}{\beta V}\frac{dV}{dt} = -\frac{3\alpha}{\beta} \qquad (2.50)$$

we then have α evaluated to:

$$\alpha = -\frac{\beta\left(P_{bath} - P(t)\right)}{3\tau_p} \qquad (2.51)$$

Substituting equation 2.47 into 2.44 gives us the final equation of motion in Berendsen barostat:

$$\dot{x} = v - \frac{\beta\left(P_{bath} - P(t)\right)}{3\tau_p} x \qquad (2.52)$$

2.1.6.2 Andersen barostat

In Andersen barostat, the coordinates r_i is replaced by scaled coordinates as:

$$\rho_i = r_i / V^{1/3} \qquad (2.53)$$

The new Lagrangian is given by:

$$L\left(\rho^N, \dot{\rho}^N, Q, \dot{Q}\right) = \frac{1}{2} Q^{2/3} \sum_{i=1}^{N} m_i \dot{\rho}_i^2 - \sum_{i<j=1}^{N} U\left(Q^{1/3} \rho_{ij}\right) + \frac{1}{2} M \dot{Q}^2 - p_0 Q \qquad (2.54)$$

If we consider Q as the volume V, the first two terms of the right hand side are the original Lagrangian. The third and fourth terms are the kinetic and potential energy of Q respectively. The simulation system can now be considered as is in a container and can be compressed by a piston (of mass M) with an external pressure p_0 acting it, with Q (whose value is the volume V) being the piston coordinate. With equations of motion for the scaled system derived and translated back to the real system, the final equations of motion are given (detailed derivations are omitted in here):

$$\frac{dr_i}{dt} = \frac{\mathrm{P}_i}{m_i} + \frac{1}{3} r_i \frac{d \ln V}{dt} \qquad (2.55)$$

$$\frac{d\mathrm{P}_i}{dt} = -\sum \hat{r}_{ij} U'\left(r_{ij}\right) - \frac{1}{3} r_i \frac{d \ln V}{dt} \qquad (2.56)$$

$$\frac{M d^2 V}{dt^2} = p_0 + \left(\frac{2}{3} \sum \frac{\mathrm{P}_i^2}{2 m_i} - \frac{1}{3} \sum r_{ij} U'\left(r_{ij}\right) \right) \Big/ V \qquad (2.57)$$

2.2 Advanced Sampling Methods in MD

2.2.1 Steered Molecular Dynamics

Steered molecular dynamics (SMD) simulation introduces a biased force to steer the systems along a particular degree of freedom. The biased force can be time-dependent or position-dependent. SMD allows for the sampling the rare events that might only happen over a large timescale while keeping manageable computational expense. In many respect, it serves as a computational counterpart to the experimental techniques such as optical tweezers (89), atomic force microscopy (AFM) (90), biomembrane force probes (91), and dynamic force spectroscopy (92), in which an external mechanical manipulations to molecules were usually applied.

SMD would be equivalent to umbrella sampling when the force applied was sufficiently small since the induced conformational change in each timestep will be minimum. However, it is often more interesting to violate this limit in cases where major structural changes and also correspondingly deviations from equilibrium are experienced. Examples of such cases include ligand dissociation and protein unfolding, etc. Free-energy differences can be calculated from the exponential averages of irreversible works from multiple SMD simulations, as revealed by the Jarzynski's equality equation (93,94).

$$e^{-\beta \Delta G} = \left\langle e^{-\beta W} \right\rangle \qquad (2.58)$$

where ΔG is the free energy change of the process, and W is the irreversible work. It can be seen that the average of exponential term of Jarzynski's equality is dominated by trajectories producing small work values. These trajectories are usually very few in the simulations, so the sufficient sampling can be a problem (93,94).

2.2.2 Targeted Molecular Dynamics

A related technique is targeted molecular dynamics (TMD) in which a bias force dependent on the structural difference between the current conformation and a target conformation is applied to drive the evolution of the specified molecule toward the target conformation. The desired conformational change would occur on a much shorter timescale than the biological timescale. The bias force is established on the root mean square deviation (RMSD) of the current conformation with respect to the target conformation (computed over the specified atoms). In a TMD simulation, the RMSD constraint is initialized as the RMSD between the initial and the target conformation in the beginning, and is decreased by some small value at each timestep over the simulation. The instant biased force is along the direction from the current to the target positions, and the magnitude of it is given by:

$$U_{TMD} = \frac{1}{2}\frac{k}{N}\left[RMSD(t) - RMSD^{*}(t) \right]^{2} \qquad (2.60)$$

where $RMSD(t)$ is the instantaneous RMSD of the current from the target conformation; and $RMSD^{*}(t)$ evolves linearly from the initial RMSD at the first TMD step to the final RMSD at the last step. The force constant k is scaled down by the number of targeted atoms, N. Since the biased force alters the energy barriers, the resulting TMD pathway is not guaranteed to be the minimum energy pathway. However, TMD is efficient in the sense that it produces all-atom transition pathways that reveal important qualitative characteristics of the transition process.

2.3 Integrative Modeling

To understand the mechanistic of a biological assembly's function, an atomistic model of it would be great value as it provides the detailed structural information of each component and a route to dissect the interaction between them. Structures of atomic level resolution are usually obtained through X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy,

however, the flexibility of biological assemblies often poses technical difficulties and renders them to be ineffective in determining the structures of large biological assemblies. Tremendous amount of critical cellular activities are actually carried by multi-biomolecular assemblies instead of a single protein component. The unveiled structures of these assemblies have impeded the process of exploring the mechanistic of their functions. Over the last decade, this situation has been alleviated by the advancement of cryo-electron microscopy (cryo-EM). Cryo-EM usually generates models of 8-20 Å resolution, which is substantially lower than a common X-ray crystal structure. On the other hand, it does not require as high concentration as X-ray crystallography does, and it has the advantage of generating the in-solution structure information, which is generally of more relevance to the in vitro condition.

Apart from the aforementioned techniques, structural information in terms of the shape of biological molecule or assembly can be obtained through Electron microscopy (EM), small angel X-ray scattering (SAXS) etc.; interaction information are accessible through various biochemical and biophysical avenues, such as chemical footprinting, cross-linking, FRET, MS, proteomics studies, and so on. The large body of these heterogeneous information, however, are mostly dispersed in literatures, lacking a high level of uniformity. Therefore, an integrative modeling approach which aiming at combing these experimental data and translate them into a uniform structural representation would be ideal in advancing our understanding on these assemblies, hence, the relevant biological functions. Such an approach would greatly reduce the time cost on conventional modeling and improve the homogeneity of the results. The generated model would also be of more reference value to the experimentalists as it could not only consolidate/disprove the experimental data, but also provide new hypothesis for further experimental tests.

To start the integrative modeling, all available data from different areas of study including experimental, physical, bioinformatics, and statistical have to be collected. Upon a proper choice of the resolution on which the system of interest will be defined in the model, the applicable data that was collected would have to be translated into spatial restraints on the entire or part of the system. For example, a residue-residue contact can be incorporated by applying a harmonic function on the distance between these two residues, and a cryo-EM density map can be used to generate a 3D-grid based function to bias the system to evolve toward it. To sample these restraint functions as a whole, various methods could then be applied, such as molecular dynamics, Monte Carlo, Brownian dynamics, docking, etc. In the end, an ensemble of models is generated for analysis and determination of the final model. Recent successes in implementing integrative modeling include a variety of systems, utilizing experimental data from X-ray, NMR, cryo-EM and SAXS. These applications have brought forth to prosperity of integrative modeling and many innovative findings in biomolecular assemblies. Karca *et al*. gave a comprehensive review on how different types of experimental data could be translated into one of the four categories of restraints such as binding sites, distance, orientation, and shape on a high level abstraction level (95). In this dissertation, applications on cryo-EM and SAXS based integrative modeling using atomistic MD simulation will be exemplified by chapter 5, chapter 6, and chapter 7.

2.4 MM/PBSA and MM/GBSA

The Molecular Mechanics/Poisson-Boltzmann Surface Area (MM/PBSA) and the Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) models are used to calculate binding free energies for macromolecules by combining molecular mechanics and continuum solvation models (96). In the MM/PBSA or MM/GBSA model, the energy of a molecule, G,

comprises the MM energy in gas phase and the solvation free energy (equation 2.61). The gas phase MM energy can be calculated by equation 3.

$$G = G_{gas} + \Delta G_{solv} \tag{2.61}$$

$$G_{gas} = H_{gas} - TS_{MM} \approx E_{gas} - TS_{MM} \tag{2.62}$$

A thermodynamic cycle in Figure 2.1 has to be utilized to calculate the binding free energy of a ligand (L) binding to a receptor (R).

$$\begin{array}{ccc} L_{aqueous} + R_{aqueous} & \xrightarrow{\Delta G_{bind}} & LR_{aqueous} \\ \downarrow -\Delta G^{L}_{solv} \quad \downarrow -\Delta G^{R}_{solv} & & \downarrow -\Delta G^{LR}_{solv} \\ L_{gas} \quad + \quad R_{gas} & \xrightarrow{\Delta G_{gas}} & LR_{gas} \end{array}$$

Figure 2.1. Thermodynamic cycle

$$\Delta G_{bind} = G^{LR} - (G^{L} + G^{R}) = \Delta G_{gas} + \Delta G_{solv} \tag{2.63}$$

The binding free energy, ΔG_{bind}, comprises the gas phase MM energy ΔG_{gas} and the solvation free energy ΔG_{sol}. Given the assumption that the binding process does not change the intra-molecular energies significantly, ΔE_{int} can be safely neglected. Hence, ΔG_{gas} can be calculated by equations 2.64 and 2.65:

$$\Delta G_{gas} = \Delta H_{gas} - T\Delta S_{MM} \approx \Delta E_{gas} - T\Delta S_{MM} \tag{2.64}$$

$$\Delta E_{gas} = \Delta E_{\mathrm{int}ernal} + \Delta E_{elec} + \Delta E_{vdw} \tag{2.65}$$

The solvation free energy ΔG_{solv} can be computed with equations 2.66 and 2.77 using the thermodynamic cycle. Each term on the right hand side of equation 2.66 comprises the polarization energy (electrostatic interaction energy), $\Delta G_{PB/GB}$, and the non-polar energy (non-

electrostatic energy), ΔG_{SA}. The polarization energy can be computed with either the PB or GB model, while the non-polar energy is estimated with solvent accessible surface area (SAS).

$$\Delta G_{solv} = \Delta G_{solv}^{LR} - (\Delta G_{solv}^{L} + \Delta G_{solv}^{R}) \tag{2.66}$$

$$\Delta G_{PBSA/GBSA} = \Delta G_{PB/GB} + \Delta G_{SA} \tag{2.67}$$

The polarization energy by PB is the difference in the work of charging a molecule in the gas phase with a dielectric constant of unity and in solution with a dielectric constant of ε (78~80 for water environment). The work of creating the charge distribution in a dielectric media is called the reaction field energy $G_{reaction}$, which can be determined with equation 2.68:

$$G_{reaction} = -\frac{1}{2}\int \rho(r)\phi(r)dr \tag{2.68}$$

where $\emptyset(r)$ is the electrostatic potential can be calculated by numerically solving the Poisson-Boltzmann equation:

$$\nabla.\varepsilon(r)\nabla\phi(r) - \varepsilon(r)\lambda(r)\kappa^2 \frac{k_B T}{q}\sinh\left[\frac{q\phi(r)}{k_B T}\right] = -4\pi\rho(r) \tag{2.69}$$

Here $\rho(r)$ is the free charge density; q is the electronic charge; k_B is the Boltzmann constant; κ^2 is a simple switching function, which is zero in regions inaccessible to the electrolyte and one otherwise; and $\lambda(r)$, the Debye length, is a function of ionic strength of the electrolyte solution. For a conducting sphere with a charge spreading out uniformly on the surface, the Poisson-Boltzmann equation has an analytical solution and the equation used to calculate the polarization energy of this ideal system is so-called the Born equation. The GB model is an extension of the Born model for a molecule with an arbitrary shape. The polarization energy of the GB model is calculated with equation 2.70:

$$G_{GB} = -\frac{1}{2}\left(1 - \frac{1}{\varepsilon}\right)\sum_{i,j}^{atoms} q_i q_j \gamma_{ij} \tag{2.70}$$

where γ_{ij} has a unit of inverse length and a well-recognized functional form of is as the following:

$$\gamma_{ij} = \left(r_{ij} + \alpha_i \alpha_j e^{-r_{ij}/\alpha_i \alpha_j} \right)^{-\frac{1}{2}} \quad (2.71)$$

r_{ij} is the inter-atomic distance; α is effective Born radius. The non-polar energy in both models is the leftover excluding the polarization energy from the solvation free energy. It is comprised of the free energy required to form the solute cavity in a solvent that mainly accounts for the entropy penalty associated with the reorganization of solvent molecules around a solute, and the van der Waals interaction between the solute and solvent. The non-polar energy G_{SA} is simply estimated by multiplying a constant, which is known as the surface tension, to the solvent accessible surface area (A) of the solute plus an intercept b:

$$G_{SA} = \gamma A + b \quad (2.72)$$

In MM/PBSA or MM/GBSA, each free energy component is subjected to intrinsic substantial uncertainty so that the accumulated error could end up being intolerable if the calculation protocol is poor designed. Nevertheless, comparing to the free energy perturbation (FEP) and Thermodynamic Integration (TI) (97) methods that require a sampling of the entire transformation path, MM/PBSA or MM/GBSA only evaluates the end-points of the system and therefore, is much more computationally efficient (96).

2.5 Trajectory Analysis

2.5.1 RMSD

Root Mean Square Deviation (RMSD) is a quantitative measurement of the structural difference between two conformations of a molecular structure. To calculate the RMSD, the two

conformations of the structure have to be overlaid so as to eliminate the translational and rotational deviations. The RMSD can then be calculated by:

$$RMSD = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(r_n^i - r_n^j\right)^2} \qquad (2.73)$$

where N is the number of atoms included in the measurement, n is the index of the atom, and r is the coordinates of the n-th atom, i and j refer to the i-th and j-th conformation of molecule respectively. The summation in equation 2.73 can include any set of atoms of the structure and is usually in the unit of Å. With biomolecular simulations system, the RMSD analysis often includes only the heavy atoms (non-hydrogen atoms). In the case of proteins and DNA, the analysis usually only includes the backbone or alpha carbons (C_α) of the protein and phosphate groups of the DNA backbone since they are usually sufficient to describe the RMSD difference.

2.5.2 Principal Components Analysis

Principal components analysis (PCA) is generally a data analysis method that identifies patterns in multi-dimensional data and highlights both similarities and differences from different samples. In PCA, the original data set is transformed to a new coordinate system, in which the greatest variance exists when the original data was projected onto the first principal component (PC), and the second greatest variance does on the second PC, and so on and so forth. PCA has been shown to alleviate the difficulty in grasping the largest-amplitude modes (slowest dynamic motions), which are often of the most interest to us, from biomolecular simulation systems (98).

Consider an m-dimensional trajectory data set consisting n snapshots from a MD simulation. This trajectory data set can be expressed as a matrix X of m columns and n rows. The data set is multidimensional, and the whole point of PCA is to reduce dimensionality in order to examine the relationships of the different conformations from the simulations. To perform PCA,

the mean of each dimension of the data set is firstly subtracted from each dimension following by constructing a covariance matrix C as showing equation 2.74. The covariance between any pair of variables (dimensions) indicates the magnitude of one variable of varying from the mean with respect to the other.

$$\mathrm{cov}(x,y) = \frac{\sum_{i=1}^{m}(x_i - \bar{x})(y_i - \bar{y})}{(m-1)} \qquad (2.74)$$

With the multi-dimensional MD trajectory data set, there are multiple covariance can be calculated using different pairs of variables. As a result, C would become a m by m matrix for a m dimensional trajectory data, with each element being the covariance between one pair of variables. Noticing that this covariance matrix is symmetric since $cov(x,y) = cov(y,x)$, and the elements on its diagonal are the covariance between one variable and itself ($cov(x,y)$), which is the variances of it. The eigenvectors of the covariance matrix are the PCs we are interested in. By diagonalizing the covariance matrix, the eigenvectors and corresponding eigenvalues will be obtained. We could then represent the original trajectory data in terms of these orthogonal eigenvectors instead of the original Cartesian coordinates. The variance magnitude in the direction of one particular eigenvector is indicated by the corresponding eigenvalue.

After the eigenvectors and eigenvalues for the covariance matrix are found, they are sorted from high to low according to eigenvalues, with the first PC being the eigenvector the with highest eigenvalues, indicating it contributes the most to the variance of the trajectory data set. It is usually the first few PCs that interest us since they account for the most part of variance (the percentage of the contribution of one PC can be calculated by dividing the associated eigenvalue by sum of all eigenvalues). Projecting the original data set onto the chosen PCs would allow for

reducing the dimensionality and retaining the large-scale motions of the of the trajectory data set. The transformation can be done by using equation 2.75:

$$X_r = E^T \times X^T \qquad (2.75)$$

where X_r is the transformed data matrix with reduced dimensionality given not all the eigenvectors were chosen, E^T is the transposed matrix of the chosen eigenvectors, and X^T is the transpose of the original data matrix with the means subtracted for each dimension. The original trajectory data set is now represented in terms of the eigenvectors of choice. It's usually of interest to plot the first up to 3 PCs against one another to visualize the projections of the data onto them so that the large-scale motions can be identified by locating the paths between clusters in the plot.

2.5.3 Normal Mode Analysis

Strictly speaking, Normal Mode Analysis (NMA) a simulation technique to probe the large-scale motion of the biomolecules instead of being an analysis tool for MD trajectory. A standard NMA requires a set of coordinates, a force field describing the interactions between the atoms. The process of a standard NMA includes minimization of the potential energy function, constructing the Hessian matrix (the second derivative matrix of the potential energy function with respect to the mass-weighted atomic coordinates), and the diagonalization of the Hessian matrix (the mathematical treatment is the same as it was with the covariance matrix in PCA). It is computationally expensive due the high dimensionality of the biomolecular systems. Therefore, a simplified model called Elastic Network Model (ENM) was introduced into protein study (99) and become popular due to its simplicity. In ENM, the atoms are connected by a network of elastic springs. With ENM, the energy minimization is not needed given all the elastic springs are at the equilibrium distances, and the number of atoms included in the model is greatly reduced to

the number of residues (or even less). Even with a great deal of simplification comparing to the standard NMA, a respectable degree of correspondence has been shown between the two methods (99).

CHAPTER 3. PROBING DNA CLAMP MECHANICS WITH SINGLE-MOLECULE FORCE SPECTROSCOPY

3.1 Abstract

Detailed mechanisms of DNA clamps in prokaryotic and eukaryotic systems were investigated by probing their mechanics using single-molecule force spectroscopy. Specifically, the mechanical forces required for the *E. coli* and yeast clamp opening were measured at the single-molecule level by optical tweezers. With the aid of steered molecular dynamics simulations, these forces were translated into subunit interface binding energies associated with the clamp opening processes. In combination with additional computational studies we identified the contact networks that contribute the most to the interface stability of the yeast and *E. coli* clamps. These studies provide a vivid picture of the mechanical and energy landscape for clamp opening and reveal how the prokaryotic and eukaryotic clamps function through different pathways.

3.2 Introduction

DNA sliding clamps play a pivotal role in DNA replication and repair and, hence, cell reproduction and survival. DNA clamps have been investigated extensively. Crystallographic studies indicate that DNA clamps retain a well-conserved structure in all three domains of life. Further structural investigations of clamp loader complexes with their cognate clamps (6,29,100-102) have shown more detailed interactions between sliding clamps and clamp loaders. The yeast clamp loader RFC (replication factor C), a pentameric AAA+ ATPase (29,103-105), was observed in complex with its corresponding clamp in a close conformation (29). In the absence of ATP, RFC has a weak interaction (*K*d ~20 nM) with PCN. However, in the presence of ATP, RFC affinity to PCNA was significantly increased (*K*d ~1 nM) (106). In the crystal structure RFC assumes a spiral arrangement with three of the five RFC subunits (RFC-A, RFC-B and

RFC-C) in contact with the surface of the clamp. Yet, an archaeal complex examined by electron microscopy revealed an open clamp in complex with its clamp loader and DNA, suggestive of an intermediate stage during clamp loading (100). Though they do not reveal the dynamics or the underlying energy landscape, these structures have provided vital information to facilitate detailed mechanistic studies both in the literature and in our current research.

A general mechanism suggests that DNA clamps serve as a mobile platform for DNA polymerases to slide freely on the DNA thereby increasing the processivity of these polymerases. More detailed ensemble kinetic studies reveal that clamps are opened by clamp loaders in the presence of either ATP or ATPγS, suggesting that ATP hydrolysis is not required for the clamp opening. Instead ATP hydrolysis is implicated in the clamp ring closure or clamp loader dissociation from the DNA (106,107). Steady-state examination of the events involving ATP binding by RFC suggests that the yeast clamp loader binds two ATP molecules before binding the sliding clamp PCNA (proliferating cell nuclear antigen) (22,37). Subsequently, the RFC/PCNA complex bound one more ATP molecule followed by the binding of DNA and an additional ATP molecule (108). Hydrolysis of an unknown number of ATP molecules led to the loading of the clamp on DNA and clamp loader dissociation from the clamp-DNA complex. Based on pre-steady state kinetic studies, Hingorani et al. proposed a model for the ATP binding and hydrolysis along the clamp loading pathway (107,109). In their model the clamp loader binds three molecules of ATP and initiates a slow conformational change that enables it to bind and open the clamp as well as to bind two additional ATPs. The clamp loader thus locks the clamp in an active loading conformation. After the hydrolysis of three ATP molecules, the clamp closes around the DNA.

Additional research featured a combination of single-molecule and ensemble techniques to examine the kinetics of the PCNA loading–unloading pathways (110). A fluorescence reso-

nance energy transfer (FRET) reporter system was constructed at the interface of the clamp to follow its opening and loading on DNA and provided crucial information about the timescale and the extent of interfacial strand separation during clamp opening. Using primarily FRET signals from a fluorescently labeled forked-DNA and PCNA, clamp loading has been shown to be stochastic and is successful only after several failed attempts (111).

To better understand the mechanics of clamp opening, theoretical studies have been carried out to map the underlying free energy landscape. Computational investigation (112) of the yeast PCNA/RFC complex has suggested that RFC, instead of destabilizing the closed state of PCNA, functions by selective stabilization of the open clamp conformation. The new interface formed between RFC and PCNA compensates for the binding energy lost upon PCNA subunit interface disruption. Furthermore, the barrier for the transition of the closed state to the open state was consistent with clamp opening occurring on a millisecond timescale.

Previous studies have suggested that the mechanism of yeast clamp opening by its cognate clamp loader is very similar to that of *E. coli* clamp opening (6). Their very similarity has attracted us to examine them as paradigms for prokaryotic and eukaryotic systems. In the current research we have measured the forces required to open an interface of either the yeast or the *E. coli* clamp with optical tweezers. We have found that, remarkably similar in terms of overall structures and functions as they are, the eukaryotic clamp (yeast PCNA) differs considerably from its prokaryotic counterpart (*E. coli* clamp) in detailed enzymatic actions. Combined with the steered molecular dynamics (SMD) simulations, the study provides detailed mechanistic information of clamp opening for both yeast and *E. coli* clamps.

3.3 Experimental Procedures

Construction of single-chain clamps- The *S. cerevisiae* PCNA's native cysteines (C22/C30/C62/C81) were mutated to serines. A single-chain PCNA was cloned into the pET28a vector preceded by an N-terminal 6-His tag as described by MacNally *et al*. (113). The C-terminal glutamate residues of the first and second PCNA subunits and the N-terminal methionine residues of the second and third subunits in the chain were all deleted. The linker sequence between the first and second subunits was GSNSQSNGSGA and the linker sequence between the second and third subunits was GSNSQASNSGA. A cysteine was introduced by site-directed mutagenesis at positions K107 or D122 in the first subunit and F185 or D223 in the third subunits respectively. Similarly, a single-chain β-clamp was cloned in the pET28 vector and a cysteine was introduced at E64 or G102 in the first subunit and at E314 or E298 in second subunit.

Biotin and digoxygenin DNA handles- The yeast PCNA gene was amplified from the pET22b PCNA plasmid using either 5′-Biotin-TTATTCTTCGTCATTAAATTTAGG-3′ or 5′-Dig-TTATTCTTCGTCATTAAATTTAGG-3′and5′-AmATGTTAGAAGCAAAATTTGAAG-3′. The 777bp PCR fragment was purified by a Qiagen PCR cleanup kit and quantified by NanoDrop™ spectroscopy. The 777bp PCR fragment containing a biotin on one end and an amine group on the other end was reacted with a water-soluble amine-to-sulfhydryl linker (sulfo-SMCC, Thermo Scientific) at room temperature in buffer-A (20 mM sodium phosphate, 150mM sodium chloride, pH 8.3). After 30 min of reaction the cross-linked DNA was purified quickly with a Qiagen PCR cleanup kit and eluted with buffer-B (100 mM sodium phosphate, 150mM sodium chloride, pH 7.2). The 777bp PCR fragment containing a digoxigenin on one end and an amine group on the other was made and purified similarly. The DNA handles were made for immediate use.

Coupling of DNA handles to clamps- We used sulfhydryl-maleimide chemistry to link maleimide activated DNA linkers to the PCNA cysteines. Single-chain PCNA was incubated at room temperature for 10 min in buffer-B containing a hundred molar excess of tris-(2-carboxethyl) phosphine (TCEP). The biotin and digoxygenin DNA handles mixed in 1:1 ratio before being added to the TCEP treated PCNA and the reaction was incubated at room temperature for 30 min and then at 4 °C overnight. The DNA handles were similarly attached to the *E. coli* single-chain β-clamp.

Tethering of PCNA on a quartz slide- The surface of a quartz slide was functionalized with biotin as reported elsewhere (114). Then the slide surface was passivated with liposome (ref.) followed by an extensive buffer wash. A solution of approximately 50nM streptavidin in Buffer B was applied on the slide surface and incubated for 15 min. and then washed away with buffer B. A complex of PCNA or β-clamp with DNA handles was first incubated for 30 min. with anti-digoxigenin-coated beads (2 micron) before being applied onto the functionalized quartz surface for incubation for over 1 hr to get tethered to the slide surface via biotin-avidin interactions.

Force measurements by optical tweezers- A homebuilt optical tweezers setup was used to measure the PCNA rupture force (114). The quartz slide with the tethered PCNA was mounted on a computer-controlled, custom-built, x-y translation piezo stage. Anti-digoxigenin-coated beads were captured by the laser trap and tracked one at a time. While the stage was moving at a designated speed (100 $nm \cdot s^{-1}$) to rupture the clamp interface, a video was recorded through a 100X1.49 NA oil immersion objective by charged coupled device (CCD) camera. Traces for bead displacements from the center of the optical trap were extracted by processing the recorded videos. The displacements were converted to the actual force applied by the optical tweezers on the clamps.

Molecular Dynamics Simulation- The crystal structure of *E coli* β-clamp and *Saccharomyces cerevisiae* PCNA were obtained from the Protein Data Bank (PDB ID: 2pol and 1plq, respectively). Molecular dynamics simulations were performed using intact trimeric clamps. Subsequent molecular mechanics Poisson−Boltzmann surface area (MM/PBSA) calculations were carried out on reduced models generated from the above MD trajectories, which contained a single subunit interface. The truncated-clamp model retained the N-terminal domain from the first subunit (residues 1 to 241) and the C-terminal domain (residues 122 to 366) of the second subunit. For the yeast clamp we simply removed one of the equivalent subunits leaving a single interface between the other two remaining subunits. The xLeap module of AMBER 9 (115) was used to add hydrogen atoms to the models. All ionizable side chains were assigned to their protonation states at pH 7.0 using the WHATIF server. Each system was then solvated with TIP3P (116) water leaving a minimum distance of 10.0 Å from the protein surface to the edge of the simulation box. Counter ions were added to achieve charge neutralization and additional 100 mM NaCl concentration was introduced to mimic physiological conditions.

The systems were then minimized for 5000 steps with fixed backbone atoms followed by 5000 steps of minimization with harmonic restraints (k= 15 kcal mol^{-1} $Å^{-2}$) to remove unfavorable contacts. All systems were then gradually heated to 300 K over 50 ps in the NVT ensemble while keeping the protein backbone restrained. The equilibration was then continued for another 1.8 ns in the NPT ensemble and the harmonic restraints were gradually released in 5 stages.

Production runs were carried out in the isothermal isobaric ensemble (1 atm and 300 K) for 20 ns for all the systems. Long-range electrostatic interactions were evaluated with the smooth particle mesh Ewald (SPME) algorithm (117). For the short-range non-bonded interactions we employed a cutoff of 10 Å with a switching function at 8.5 Å. The integration time step was 2 fs

and the bonds between hydrogen and heavy atoms were fixed to eliminate the most rapid oscillatory motions in the system. The r-RESPA multiple time step method (118) was adopted with a 2 fs time step for bonded, 2 fs for short-range non-bonded interactions and 4 fs for long-range electrostatic interactions. All simulations were performed using the NAMD 2.7 code (73) with the AMBER Parm99SB parameter set (119) containing the force field for nucleic acids and proteins, on Hopper II, a Cray XE6 system at the National Energy Research Scientific Computing Center. Data were analyzed using the PTRAJ utility in AMBER (115) and custom VMD (120) TCL scripts.

Steered Molecular Dynamics- For the steered molecular dynamics runs we selected 5 conformations ~2 ns apart along the MD production trajectories of the yeast and β-clamp systems. These statistically independent, uncorrelated configurations were used to initiate constant velocity steered molecular dynamics (SMD) simulations (93,94,121-124). External forces in SMD were applied on the following pairs of residues: K107-R185 and D122-D223 for yeast PCNA; E64-E314 and G102-E298 for the β-clamp. In each case we selected the center of mass (COM) distance between the residue pair as our SMD pulling coordinate. The two residues were pulled away from each other by applying a harmonic restraint to their COMs ($k = 25$ kcal mol^{-1} $Å^{-2}$) and moving the restraint with constant velocity of 3 Å ns^{-1} along the line connecting the COMs of the two residues. Configurations from the SMD pulling trajectories were sampled for analysis at an interval of 10 ps. To analyze the disruption of contacts at the subunit interfaces we computed the number of hydrogen bonds as a function of pulling distance. For this hydrogen bond analysis a H-bond distance cutoff of 3.2 Å and angle cutoff of 30° were assumed. For each hydrogen bond detected during the pulling trajectory the time of last occurrence was used to identify the contact break-up. The interface opening mechanisms were then mapped onto the force-extension profile.

Binding Free Energy Calculation- The molecular mechanics Poisson−Boltzmann surface area (MM/PBSA) method in AMBER 9.0 (115) was applied to calculate relative binding energies for the subunit interfaces of the β-clamp and yeast PCNA. The MM/PBSA calculations were carried out using the last 10 ns from the free MD trajectory for each clamp system. In total, 2500 frames were used for averaging. The free energy of binding can be calculated for each snapshot from the following expression:

$$\Delta G_b = \Delta E_{MM} + \Delta G_{sol} - T\Delta S \qquad (3.1)$$

where ΔE_{MM} is the gas-phase molecular mechanics binding energy, contributed by van der Waals and electrostatic interactions; ΔG_{sol} is the change in solvation free energy upon binding, comprised of electrostatic and nonpolar interactions; ΔS is the gas-phase entropy change upon binding. The electrostatic solvation energy is determined using the finite difference Poisson−Boltzmann (PB) method. In the PB calculation, a 0.5 Å grid size was used, and the dielectric constants of protein and water were set to 1.0 and 78.0, respectively. The nonpolar contribution to the solvation free energy was determined from the solvent-accessible surface-area according to the equation:

$$\Delta G_{nonpolar} = \gamma A + b \qquad (3.2)$$

where A is the solvent-accessible surface area, and the solvation parameters γ and b are 0.0072 kcal $mol^{-1}Å^{-2}$ and 0 kcal mol^{-1}, respectively. The probe radius of the solvent was set to 1.4 Å. The surface area A was calculated using MolSurf in AMBER 9. The optimized set of atomic radii in AMBER 9 was used, and the atomic charges of the protein were taken from the ff99SB force field. It is important to note that the entropy contribution was not included and, therefore, the computed binding energies should be considered only relative to one another.

The total binding energies were further decomposed into individual contributions from residues forming the subunit interfaces of yeast and *E. coli* clamps. The energy contribution from individual residue was calculated by summing its interaction energy over all residues in the clamp system. The same dynamics trajectories utilized in the MM-PBSA calculations were used for energy decomposition. The contribution to the binding energy from each residue includes three terms: van der Waals contribution (ΔE_{vdw}), electrostatic contribution (ΔE_{ele}), and solvation contribution (ΔG_{GBSA}). PTRAJ module of AMBER TOOLS 12 and VMD (120) were used for the analysis of trajectories and structural visualization.

3.4 Results and Discussion

Construction of single-chain clamps- Yeast PCNA is a homotrimer whereas the *E. coli* β-clamp, a homodimer. However, both retain a ring-shaped structure with inner diameters sufficient to accommodate dsDNA. To measure the mechanical force required for clamp opening using optical tweezers we constructed a single-chain clamp, which possesses only one subunit interface that can be opened, thereby avoiding potential complications involved in the measurements. PCNA is composed of three monomers (A, B and C) aligned in a head-to-tail fashion. In brief, a single-chain clamp was created by incorporating 10 amino acid linkers between the subunits A-B and B-C leaving a single open interface between subunits A-C. In addition we introduced a cysteine on subunits A and C by site-directed mutagenesis. These cysteines were used to attach DNA handles via sulfhydryl-maleimide chemistry. Previously MacNally et al. have reported that the single-chain PCNA generated by this method is structurally identical to wild-type PCNA while retraining full activity. The enhanced ATPase activity of RFC in the presence of the clamp served as an assay for clamp functionality (46). We have measured the ATPase activity of the clamp loader in the presence of the single-chain clamp using an enzyme-coupled reaction in

which the production of ADP was coupled to the depletion of NADH by pyruvate kinase and lactate dehydrogenase (125) and found that the ATPase activity of the clamp loader was stimulated by the single-chain clamp to the same extent as the wild-type clamp. Similarly a single-chain *E. coli* β-clamp was constructed that likewise was fully active in the ATPase assay.

Rupture force measurement- Yeast PCNA is a homotrimer with three identical interfaces. To study the biophysical changes and nature of forces operating in DNA clamp loading and unloading, we measured the mechanical force required for the PCNA opening using optical tweezers. In the measurements one end of the single-chain PCNA attached to a DNA linker was tethered to a quartz slide surface via a biotin-avidin interaction and the other end, attached to a DNA linker capped with an anti-digoxigenin coated bead (Figure 3.1).

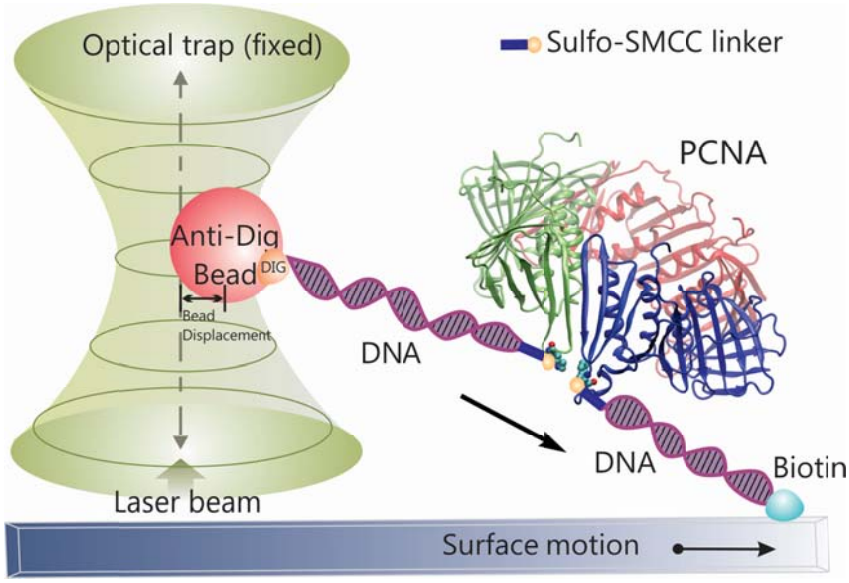

Figure 3.1 Schematic representation of clamp rupture force measusrment with optical tweezers.

A typical rupture event of PCNA was depicted in Figure 3.2. Multiple traces for the rupture events were collected and histograms were constructed (Figure 3.3A and 3.3B) for determin-

ing the average bead displacements. To convert bead displacements in arbitrary units to the rupture force, we multiplied the observed bead displacement by the trap stiffness, 0.15 pN·unit^{-1}. We also found that the mechanical force required for the PCNA interface opening was dependent on the position of attachment. When yeast PCNA was pulled at residues K107 and F185, we observed that 1.7 ± 0.6 pN was sufficient for the PCNA ring opening. In contrast, when pulling PCNA at residues D122 and D223, a substantially higher rupture force, 11 ± 1 pN, was required. A control experiment using only a 1-kb DNA linker with an anti-dig bead on one end and biotin on the other was performed to confirm that the observed rupture forces are due to the clamps (Figure 3.4). In these control experiments, no rupture events were observed within the operative regime of the optical trap.

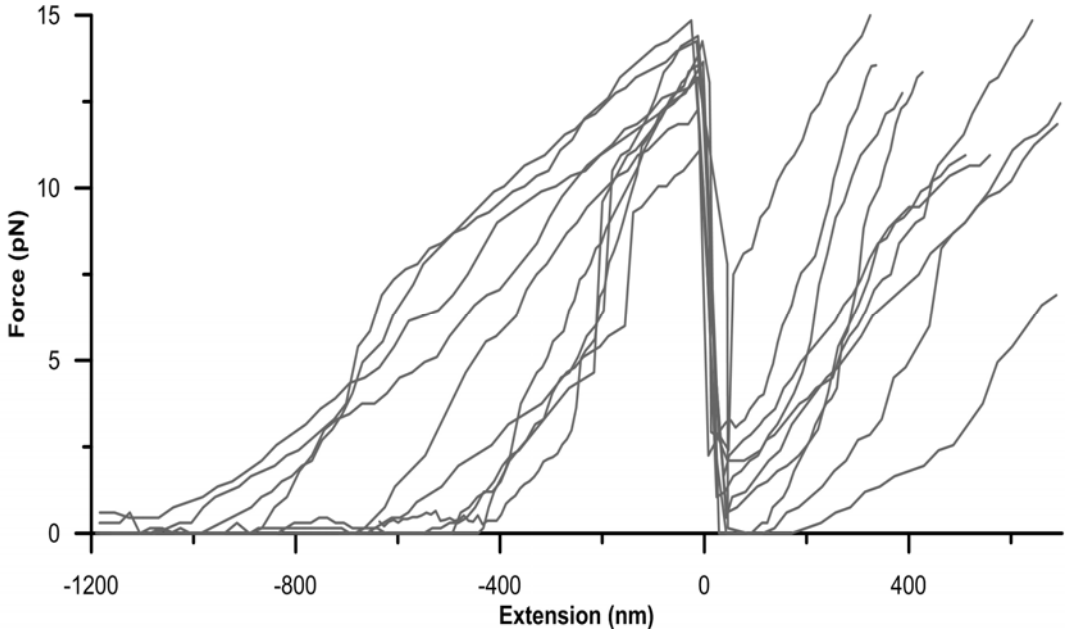

Figure 3.2 Ten force-extension curve examples

The DNA-clamp complex was pulled at a constant speed (100 nm/s). The applied force was calculated by multiplying the bead displacement by the trap stiffness (0.15 pN·nm^{-1}). The extension is calculated by subtracting the bead displacement from the stage displacement and post-synchronized to be zero at the moment of rupture.

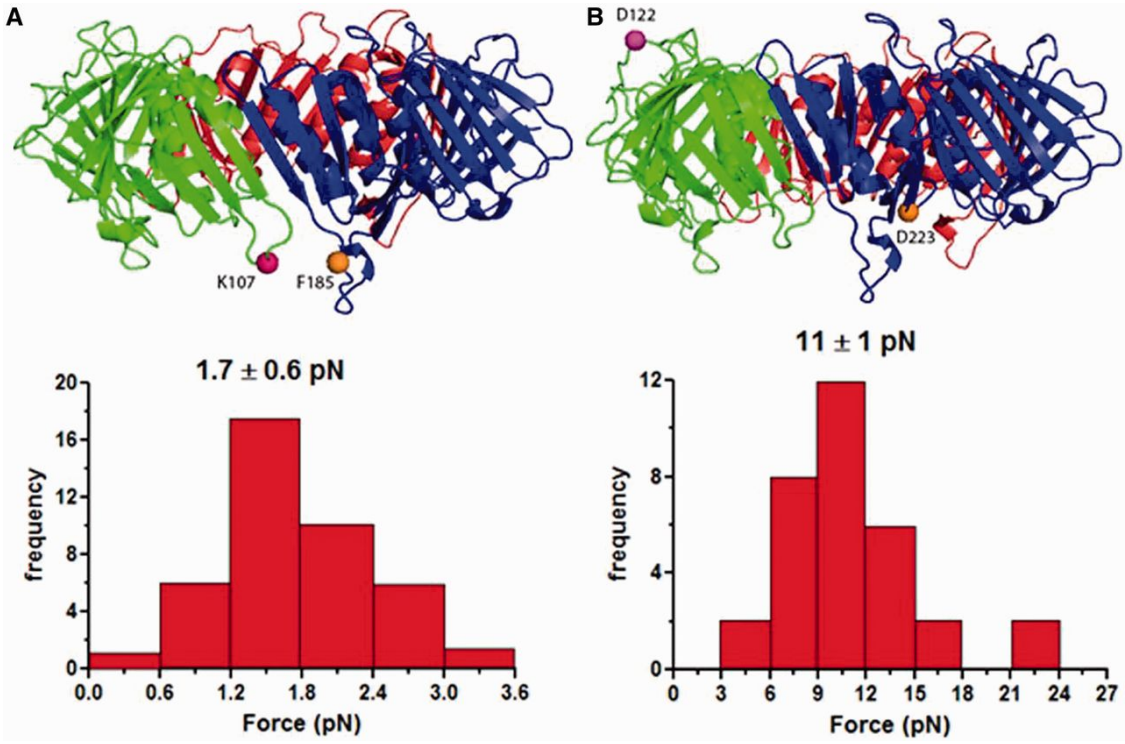

Figure 3.3 Histograms of PCNA interface rupture force

A) PCNA pulled at residues K107 and F185. The interface was ruptured at 1.7 ± 0.6 pN. B) PCNA pulled at residues 122 and 223. The interface was ruptured at 11 ± 1 pN.

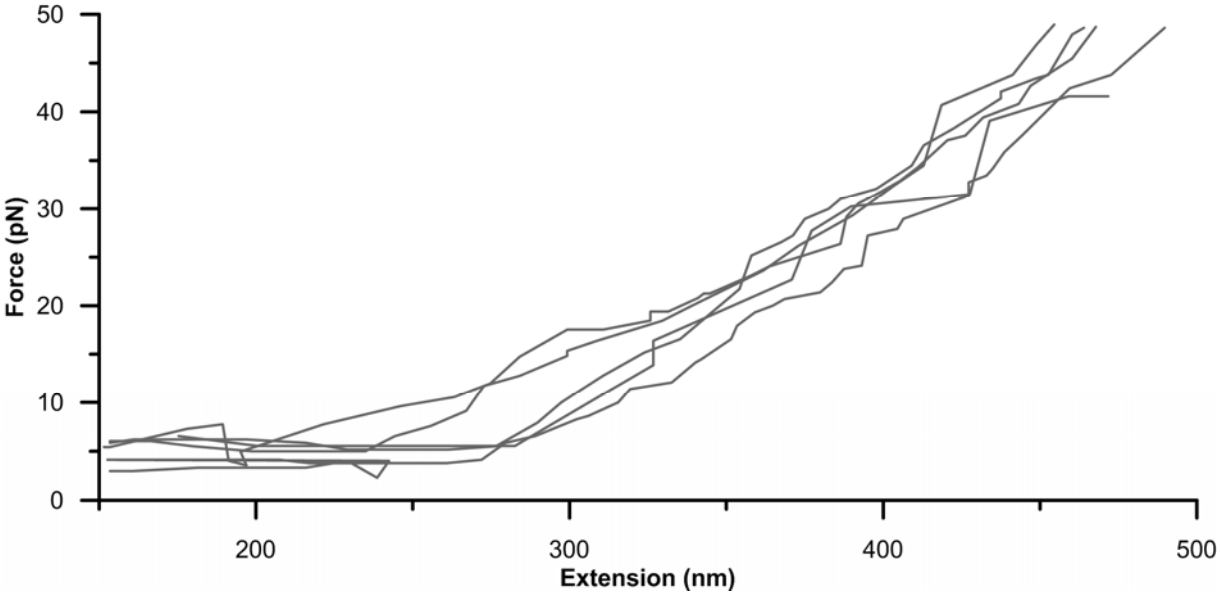

Figure 3.4 Typical force extension curves

1kb DNA linker with an anti-dig bead on one end and biotin on the other. No rupture events were observed within the operative regime of the optical trap (0 - 40 pN).

In a similar fashion we performed experiments with the *E. coli* β-clamp and constructed histograms (Figure 3.5A and 3.5B). In the case of the *E. coli* clamp we found that clamp opening force was also dependent on the position of attachment. When β-clamp was pulled at residues E64 and E314, we observed that 5.8 ± 0.5 pN was sufficient for its opening whereas pulling at residues G102 and E298 resulted in a higher rupture force of 12 ± 0.3 pN. By comparing the minimum force required for yeast PCNA opening vs. *E. coli* β-clamp opening, we found that the *E. coli* β-clamp requires at least three times higher opening force than the yeast PCNA.

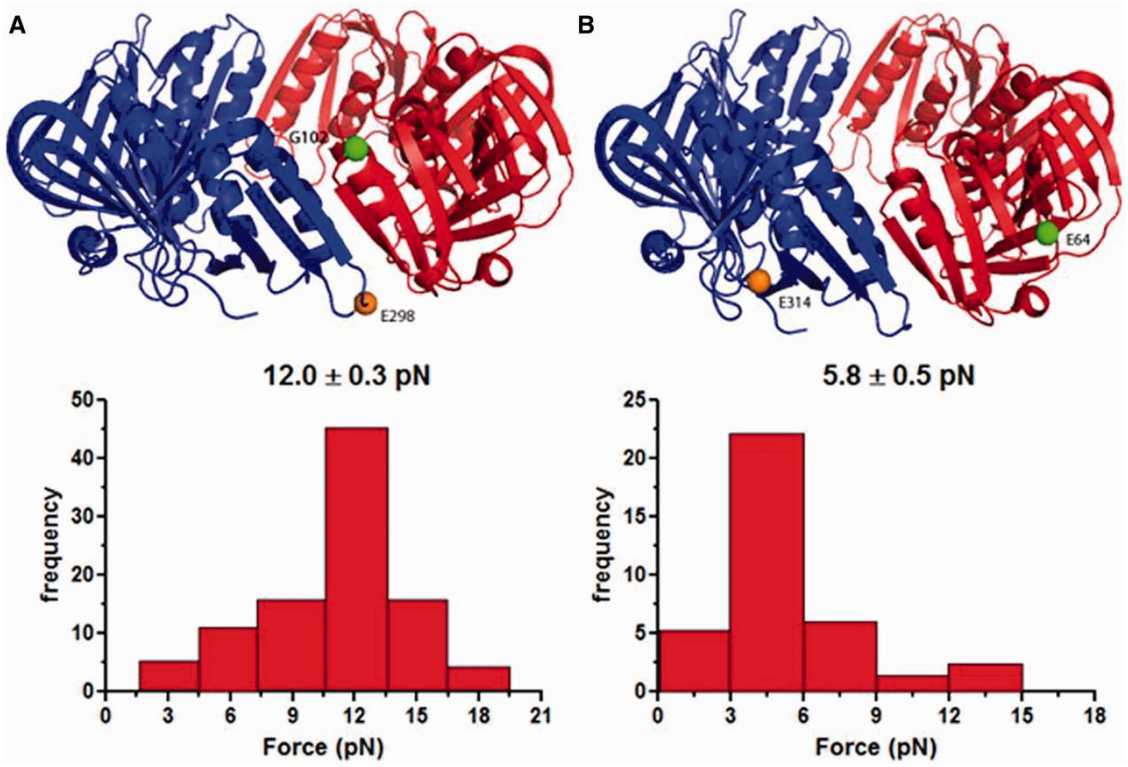

Figure 3.5 Histograms of the β-clamp interface rupture force

A) β-clamp pulled at residues E64 and E314. The interface was ruptured at 5.8 ± 0.5 pN. B) β-clamp pulled at residues G102 and E298.The interface was ruptured at 12 ± 0.3 pN.

The data show that in both systems, the mechanical force required for clamp opening was dependent on the position of attachment. This can be explained by inspecting the locations of the attachments. Smaller rupture forces were observed when the pulling force was applied directly near the subunit–subunit interface (PCNA: K107-F185, Figure 3.2A; β-clamp: E64-E314, Figure 3.3A). Larger rupture force was required when the pulling force was applied away from the subunit–subunit interfaces (Figure 3.3B and 3.5A), which might be due to disruptions of other interactions outside of the subunit interface. Thus, the smaller force values are more accurate assessments of the forces necessary to disrupt the interactions in the clamp subunit interfaces.

Interface interaction energies - To analyze the energetics of the clamp interfaces, various components of the interaction energy of the subunit-subunit complexes were evaluated using MM/PBSA analysis (Table 3.1). We note that unlike the case of MM/PBSA applied to small ligands, it is difficult to accurately assess the entropic contribution to binding. Calculating the entropic component from normal mode analysis would typically lead to large standard deviations for the binding free energies. Therefore, to quantitatively compare the interfaces we have opted to focus only on the enthalpic contribution. Thus, the values in Table 3.1 represent binding energies rather than free energies. As expected, due to lack of entropy-enthalpy compensation, the absolute binding energies are overestimated compared to the corresponding experimental free energies. Nonetheless, previous applications of MM/PBSA to protein interfaces have shown the usefulness of comparing the relative stabilities of the binding epitopes and identifying binding hotspots based on energy decomposition(126).

Table 3.1 Free Energy Analysis (kcal mol-1) for the formation of clamp subunit interfaces

Contribution	PCNA	*β*-clamp
ΔE_{ele}	2138.80 ± 58.24	-486.51 ± 36.39
ΔE_{vdw}	-73.95 ± 4.37	-60.99 ± 5.93
$\Delta G_{nonpolar}$	-11.28 ± 0.45	-11.19 ± 0.52
ΔG_{polar} (PB)	-2076.70 ± 59.39	463.11 ± 32.38
ΔG_{sol} (PB)[a]	-2087.98 ± 59.22	451.92 ± 32.17
ΔG_{ele} (PB)[b]	62.09 ± 8.32	-23.40 ±10.45
$\Delta\Delta G_b$ (PB)	-23.14 ± 7.64	-95.58 ± 8.25

[a]The polar/nonpolar ($\Delta G_{sol} = \Delta G_{polar} + \Delta G_{nonpolar}$) contribution. [b]The electrostatic ($\Delta G_{ele} = \Delta E_{ele} + \Delta G_{polar}$) contributions.

The MM/PBSA analysis for our complexes leads to substantial negative binding energies for both the *E. coli* β-clamp and *S. cerevisiae* PCNA. This is consistent with the experimental evidence suggesting both clamps remain closed in solution (127). Furthermore, the binding energy ΔG_b for the *E. coli* β-clamp interface was found to be four times larger than the corresponding energy for the yeast PCNA interface (Table 3.1). The difference is significant (even when we take into account the neglect of the entopic effects) and it also correlates very well with the outcome from the optical trap rupture force measurements.

More importantly, the factors contributing to interface stability are entirely different for the *E. coli* β-clamp versus *S. cerevisiae* PCNA. To evaluate these factors we compared the individual components of the binding energies in Table 3.1. The intermolecular (gas-phase) van der Waals interactions are favorable and contribute almost equally to the stability of both interfaces, while the gas-phase electrostatic interactions are favorable for the *E. coli* β-clamp interface and unfavorable for the yeast PCNA interface. The nonpolar solvation energy ($\Delta G_{nonpolar}$) corresponds to the burial of solvent-accessible surface area (SASA) and contributes favorably to the stability of both interfaces. By contrast, the electrostatic solvation energy (ΔG_{polar}) disfavors binding for the *E. coli* β-clamp interface whereas it favors greatly binding for the yeast PCNA interface. Thus, there is a compensation effect between the gas-phase electrostatics and the electrostatic solvation energy, which act in opposite directions for each clamp interface. For the yeast clamp the total electrostatic energy (sum of unfavorable gas-phase and favorable solvation energies) was found to disfavor binding. By contrast, the β-clamp interface is characterized by unfavorable electrostatic solvation energy and favorable gas-phase electrostatic contribution,which in the balance produce total electrostatic energy that still favors binding. This outcome can be rationalized by the fact that the opposing sides of the yeast PCNA subunit interface are both nega-

tively charged. By contrast the β-clamp interface is composed of an anti-parallel beta sheet with the positive and negative residues sequestered on each strand and forming salt bridges across the interface. Such salt bridge interactions, found in the context of largely hydrophobic environment, contribute to the far greater overall stability of the *E coli* clamp when compared with the yeast clamp. As expected, MM/PBSA overestimates the absolute value of binding free energy due to the omission of entropy in the calculation.

The total binding energy at each subunit interface was further decomposed into individual residue contributions as shown in Figure 3.6. Residues contributing more than ±1.5 kcal/mol toward the stability of each interface are labeled explicitly (Figure 3.6A and 3.6C). The residues contributions are also mapped onto the structure of the two subunit interfaces and color-coded from positive (blue) to negative (red) in panels 3.6B and 3.6D.

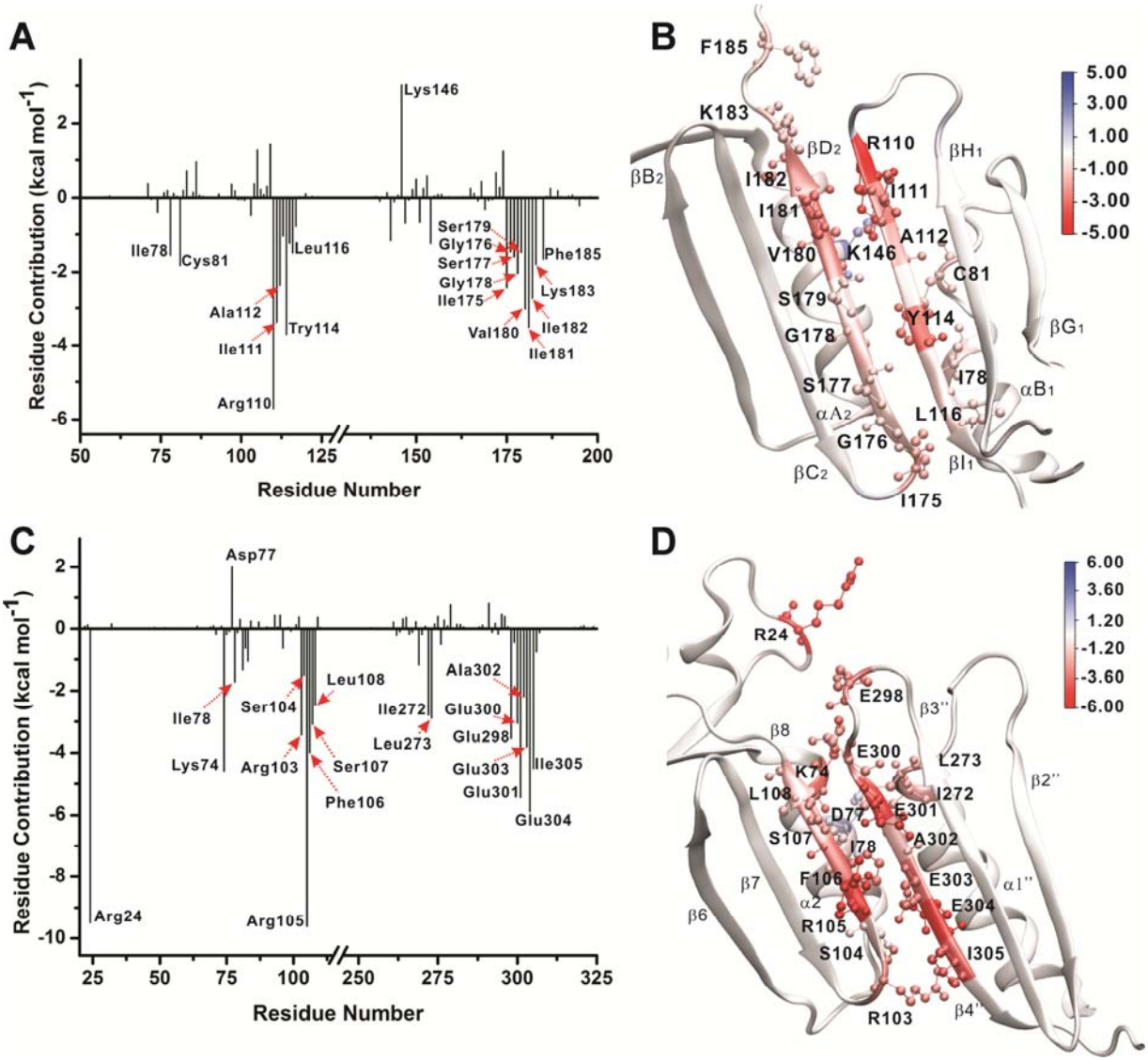

Figure 3.6 Binding energy contributions from the interface residues

A-B) yeast PCNA and C-D) *E. coli* β-clamp. Individual residue contributions are indicated on the graphs and residues contributing more than ±1.5 kcal·mol^{-1} are labeled explicitly. The binding energy contributions are also mapped onto the structures of the two interfaces and are colored from red (negative) to blue (positive).

For the yeast clamp the dominant contribution to the binding epitope comes from residues positioned along the interfacial antiparallel β−strands (Ile175 to Lys183 on strand βD2, and Phe185on the loop region right above strand βD2; and Arg110 to Ala112, Tyr114, Leu116 on βI1). Thus, the subunit interface is largely hydrophobic and, in this hydrophobic context, stability is primarily determined by the presence of 7 main chain hydrogen bonds. Notably, for the yeast PCNA the residues with greatest contributions to stability are not evenly distributed along the length of the beta strands giving rise to the possibility of asymmetric disruption of the interface upon opening. In such a scenario, the interface would preferentially unzip from one end of the β-sheet wherein the residue contacts are least stable.

The α-helices at the back of the antiparallel β-sheet, facing the central hole of PCNA (αB1 and αA2), were found to contribute to a far lesser extent to the interface stability. Favorable contributions arise from residues Ile78, Cys81 on helix αB1, and Leu151, Leu154, on the opposing αA2 helix. Furthermore Lys146 on αA2 helix was found to destabilize the interface by 3.02 $kcal \cdot mol^{-1}$ due to unfavorable electrostatic interaction with the Arg110 on the opposing βI1 strand.

The β-clamp interface is organized differently with electrostatic interactions playing a much more prominent role. Five glutamate residues, Glu298, Glu300, Glu301, Glu303, and Glu304 are all positioned along the interfacial β-strand β4’’. Arg103, Arg105 from the β8 strand and Lys74 from the α2 helix are all segregated across the interface from the negative residues. This positioning contributes greatly to the total binding energy (Figure 3.6B). Meanwhile, hydrophobic residues such as Ile78 on α2, Leu108 on β8, Ile272 and Leu273 from α1’’ greatly stabilize the interface through hydrophobic interactions. Residues Phe106, Ser107 and Leu108 on β8, Ala302 and Ile305 on β4’’ also contribute to interface stability by forming a hydrophobic

core and additionally through backbone hydrogen bonds across the β8-β4'' sheet. Additionally, Arg24 on a distant loop connecting α1 and β2 stabilizes the interface, mostly likely through electrostatic interaction with the negatively charged Glu298 on the β4'' strand.

Interface rupture mechanisms- Sliding clamp opening is an activated process. Activated rare events are not commonly accessible to straightforward MD simulations. To enforce separation of the two interfacial beta strands of *S. cerevisiae* PCNA and the β-clamp we relied on steered molecular dynamics: a method that involves the use of a harmonic constraint moving at a constant velocity to "steer away" the centers of mass of groups of atoms and, thus, promote enhanced sampling. SMD was used to force open the clamp subunit interfaces by applying external forces to steer the centers of mass of the selected residue pairs along the direction connecting the residue COMs. A force constant k of 25 kcal·mol^{-1}·Å^{-2} was sufficient to open the interfaces of either PCNA or the β-clamp. The steering velocity v was selected to be 3.0 Å/ns for all 5 different starting conformations for each clamp. Additional SMD with reduced steering velocity of 1.0 Å/ns was later carried out on the starting conformations which resulted in the lowest work profile from each system to consolidate the 3.0 Å/ns pulling results (data not shown). SMD simulations probe the energy barriers for interface disruption in a way analogous to single molecule optical tweezers experiments. However, the timescales involved in computational pulling runs are much shorter than the corresponding experimental timescales. As a result, the computed force-extension profiles include a sizable contribution from non-equilibrium effects, which is not present in the corresponding experimental profiles. Therefore, the rupture forces probed by the SMD simulations are not directly comparable to the experimental pulling data. Nonetheless, mechanistic details regarding subunit interface disruption under external forces can be analyzed and the relative magnitude of the forces between different systems can be compared. Residue

contacts along the interface break in stages leading to a characteristic seesaw pattern (several peaks and shoulders) along the profile. Integrating the force/extension plot yields a work vs. distance profile.

Five runs of the yeast clamp (Lys107-Phe185 pair) all opened up completely after approximately 40 ns simulation, while only one β-clamp (G102-Glu298) completely opened after 50 ns. The other four β-clamp systems eventually open but undergo contact breaking and reformation between residues not involved in the original interface. Force-extension profiles from the run that opened the β-clamp without complications and from the run that yielded the lowest work profile for the yeast clamp are shown in Figure 3.7. The force applied by the virtual spring on the pulling residue accrues upon the extension until a threshold is reached, indicated by the peaks in the plots, results in the breaking of the interaction between the subunits and subsequent force decrease. The different stability of the yeast and *E. coli* clamp interfaces is evident from the fact that the yeast clamp requires a smaller maximum force to open up compared to the β-clamp. The major peak on the β-clamp profile is 15.52 kcal·mol^{-1}·Å^{-1} - approximately 3.3 times larger than the corresponding peak of the yeast clamp profile (4.74 kcal·mol^{-1}·Å^{-1}). This finding is in good correspondence with the outcome of the MM/PBSA analysis and also consistent with the experimental pulling data (Figure 3.3 and 3.5).

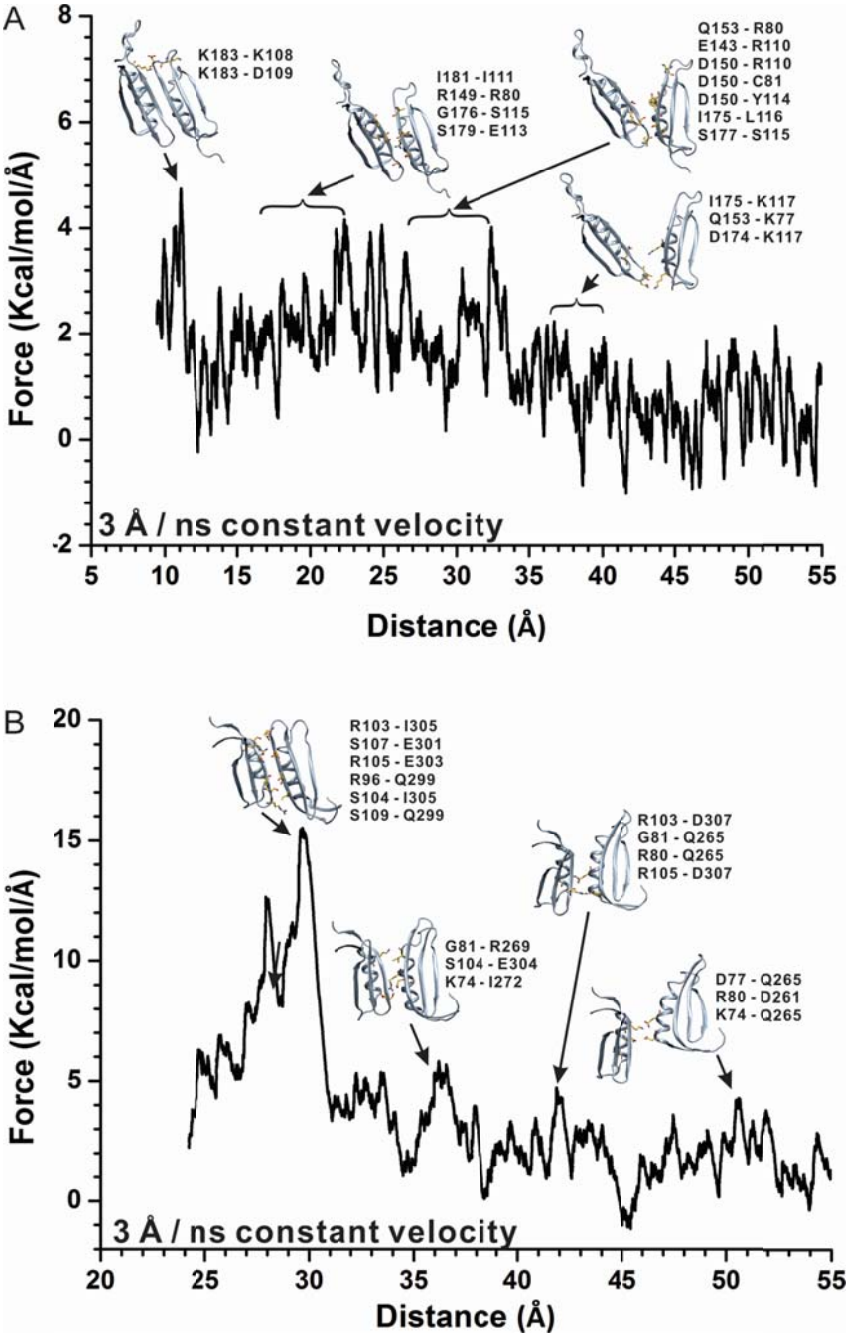

Figure 3.7 Interface disruption mechanism as determined from the SMD force-extension profiles

A) yeast PCNA and B) *E. coli* β-clamp. Representative snapshots of the subunit interfaces at different stages of separation are shown. Arrows indicate the correspondence to distinct

peaks in the force-extension profiles. The force is representative for the running average of every 3 Å to smooth the plots.

The structural differences at the subunit interfaces in yeast PCNA and *E. coli* β-clamp lead to observed distinct opening mechanisms in the SMD trajectories. Four distinct groups of residue contacts get disrupted consecutively along the yeast PCNA pulling trajectory with corresponding peaks readily identified in the force extension profile (Figure 3.7A). Specifically, yeast PCNA was found to open by gradual unzipping of the interfacial anti-parallel beta sheet βD2-βI1. The strand separation is initiated from the end of the interface holding the residue pair Lys107-Phe185. Interestingly, this end is presumed to be more stable based on the MM-PBSA binding energy decomposition analysis. This indicates that the exact position where the external forces are applied matters for the observed directionality of opening. This finding is also consistent with the results from the optical tweezers experiments, which showed that the maximum rupture force is dependent on the positions of the residue pair being pulled.

Unlike the yeast PCNA, β-clamp interface is stabilized the most by residues in the middle of the antiparallel beta sheet β8-β4''. Its strength is derived from the presence of a hydrophobic core as well as main chain hydrogen bonds and salt bridge interactions. Therefore the SMD pulling run reveals a much more abrupt opening whereby a large number of contacts are broken cooperatively (as indicated by a prominent initial peak in the force extension profile in Figure 3.7B). The remaining few contacts in the middle of the interface correspond mainly to charged contacts between positive and negative residues on the opposing sides of the interface. These contacts are the last to be cleared simply due to the long side chains of Arg residues that allow substantial extension before breaking of the contact.

Energy requirements for clamp opening- A significant amount of mechanical energy is required for clamp opening. Among all possible energies the minimum ones are most meaningful. To roughly estimate the minimum energies required, we combined the clamp-opening distances derived from the steered molecular dynamics (SMD) simulations and the forces measured by optical tweezers. SMD simulations afforded the pulling distances until the clamps were fully opened. For yeast PCNA the clamp-opening distance was estimated to be 35Å and for *E. coli* β-clamp, 31Å. Experimentally the clamp-opening distance for yeast PCNA was determined by FRET to be about 34Å (128), a distance similar to the one computed by SMD. That result confirmed the validity of our SMD simulations. The mechanical forces required to open the clamps were provided by our optical tweezers experiments. The minimum forces measured 1.7 ± 0.6 pN at residues K107 and F185 for yeast PCNA and 5.8 ± 0.5 pN at residues E64 and E314 for *E. coli* β-clamp. Generally, as the pulling distance is rather small, we can legitimately assume that the force (*f*) linearly increases with the pulling distance (*d*) and thereby, approximating the integral of the pulling force over the distance as the area of a triangle, we can roughly calculate the energy (*E*) by the following equation:

$$E = \frac{1}{2} \cdot f \cdot d \qquad (3.3)$$

Using the equation above, we have estimated that the minimum energy required for clamp-opening is around 0.43kcal·mol^{-1} for the yeast PCNA and 1.29kcal·mol^{-1} for the *E. coli* β-clamp.

The overall clamp loading process is driven by ATP hydrolysis. So naturally we expect that most of the energy released by ATPs should be used for clamp opening. Yet, based on the number of hydrolyzed ATPs estimated from ensemble kinetic studies (107,109), there exist huge disparities between the energy generated by ATP hydrolysis and that required to open the clamp. Consequently, quite contrary to our expectations, only a small fraction of the energy by ATP hy-

drolysis is used to open clamps. Thus, the opened clamp together with its clamp loader forms a low energy and relatively stable complex instead of a short-lived high energy one. The kinetic advantage of the complex for clamp loading is well illustrated by our single-molecule kinetic studies by FRET (111), in which several attempts are required before the clamps are fully loaded on DNA. During clamp loading, a long-lived complex will have more opportunity to survive failed attempts and hence enhance the overall efficiency of clamp loading. In the loading process the energy by ATP hydrolysis is mostly likely to be dissipated after the clamp is successfully loaded on DNA. Consequently, ATP hydrolysis may be ultimately associated with dissociation of the clamp loader in addition to closure of the clamp (129,130).

3.5 Conclusion

The studies conducted by optical tweezers probed the mechanics of clamp loading and revealed the mechanical forces and energies required for opening the subunit interfaces of the two clamps. And the differences in clamp opening were investigated structurally and mechanistically by MD or SMD simulations. Specifically, we quantify these differences in terms of specific hydrophobic and polar contacts from MM/PBSA analysis as well as delineate the detailed mechanism of disruption of the interfaces. We observe distinct opening mechanisms – gradual unzipping in the case of yeast PCNA versus abrupt cooperative disruption for the *E. coli* β-clamp. These differences are dictated by the distinct nature of the contacts formed at the subunit interfaces. In comparison with yeast PCNA, at the β-clamp interface we observe a far greater number of strong hydrophobic interactions in addition to favorable electrostatics and strong specific salt-bridge interactions. All of them result in a large overall difference in interface stabilities and thus in the mechanical forces required for clamp opening in the yeast and *E. coli* clamps.

CHAPTER 4. OPENING PATHWAYS OF THE DNA CLAMPS PROLIFERATING CELL NUCLEAR ANTIGEN AND RAD9-RAD1-HUS1

4.1 Abstract

Proliferating cell nuclear antigen (PCNA) and the checkpoint clamp Rad9-Rad1-Hus1 (9-1-1) topologically encircle DNA and act as mobile platforms in the recruitment of proteins involved in DNA damage response and cell cycle regulation. To fulfill these vital cellular functions both clamps need to be opened and loaded onto DNA by a clamp loader complex – a process, which involves disruption of the DNA clamp's subunit interfaces. Herein, we compare the relative stabilities of the interfaces using the molecular mechanics Poisson−Boltzmann solvent accessible surface (MM/PBSA) method. We identify the Rad9-Rad1 interface as the weakest and, therefore, most likely to open during clamp loading. We also delineate the dominant interface disruption pathways under external forces in multiple–trajectory steered molecular dynamics runs. We show that, similar to the case of protein folding, clamp opening may not proceed through a single interface breakdown mechanism. Instead, we identify an ensemble of opening pathways, some more prevalent than others, characterized by specific groups of contacts that differentially stabilize the regions of the interface and determine the spatial and temporal patterns of breakdown. In 9-1-1 the Rad9-Rad1 and Rad9-Hus1 interfaces share the same dominant unzipping pathway whereas the Hus1-Rad1 interface is disrupted concertedly with no preferred directionality.

4.2 Introduction

Genome duplication and maintenance are essential for all life and the dynamic molecular machinery responsible for these vital biological functions is the replisome (131). Proliferating Cell Nuclear Antigen (PCNA) (22,38,132) and a related checkpoint protein, Rad9-Rad1-Hus1

(9-1-1) (133) are DNA clamps, which act as platforms for the assembly of core replisomal components on DNA (35). In this capacity, DNA clamps are essential in cellular activities ranging from DNA replication, repair of DNA damage, chromatin structure maintenance, chromosome segregation, cell-cycle progression and apoptosis (35,38,134-139). PCNA is a recognized master coordinator of multiple pathways controlling replication and DNA damage response. The clamp has a toroidal shape that wraps around DNA and topologically links the replicating DNA polymerase to its substrate. During lagging strand DNA synthesis, PCNA organizes three core replication proteins - DNA polymerase, Flap endonuclease 1 (FEN1), and DNA ligase I (Lig1) (36,140). Similarly, the checkpoint clamp 9-1-1 is a crucial constituent of complexes responsible for checkpoint signaling and base excision repair (BER). While PCNA features three identical subunits, 9-1-1 (28,141,142) is heterotrimeric (Figure 1) reflecting the distinct roles of PCNA and 9-1-1 in DNA processing (143). Unlike PCNA, 9-1-1 does not associate with replicative polymerases but recruits checkpoint effector kinases to sites of DNA damage. 9-1-1 also stabilizes stalled replication forks (136,144,145) and stimulates BER enzymes (including FEN1, and Lig1), thus linking BER activities to checkpoint coordination (146-149).

PCNA and 9-1-1 are composed of three subunits forming closed rings encircling DNA. Each subunit includes two wedge-shaped domains connected by a long inter-domain connector loop (IDCL) (22,132). Most partner proteins associate with the clamps through the IDCL using a consensus sequence called PIP-Box. The clamp subunits delimit an inner cavity whose walls are rich in positively charged residues needed for contacting DNA. The fact that PCNA and 9-1-1 form stable trimers implies an activated mechanism for loading onto DNA (6,108,150,151). A clamp-loader (AAA+ ATPase) opens and subsequently re-closes the clamps around DNA (105). Specifically, PCNA is loaded onto primer-template DNA by the pentameric replication factor C

(RFC) complex (29,103,105,108,109,152). 9-1-1 is loaded by a variant clamp loader wherein the large Rfc1 subunit has been replaced by Rad17 (135,153) The clamp-loading mechanism is conserved in the three domains of life and involves opening of the pre-formed rings along a single subunit interface. Since all three 9-1-1 interfaces are distinct, naturally, a question arises whether one interface is preferentially opened during clamp loading. Furthermore, since the protein machinery responsible for clamp loading is essentially the same for PCNA and 9-1-1 (besides the replacement of Rfc1 by Rad17), this prompts two important questions: (i) does subunit interface stability play a role in determining which interface is opened; and (ii) is there an intrinsic difference in the energy required to open and load PCNA versus 9-1-1?

For the heterotrimeric 9-1-1 clamp, the question of which interface opens to allow a pre-formed ring to be loaded onto DNA junctions is still unresolved. There have been two competing proposals. First, Dore' at el. argued that the Rad9–Rad1 interface serves as an opening gate during clamp loading. In their 9-1-1 structure (28) (PDB id: 3G65) the Rad9–Rad1 interface was found to be most polar and had the smallest buried surface area (1100 $Å^2$). By contrast, Sohn et al. (141) independently solved the 9-1-1 structure (PDB id: 3A1J) and argued that the Hus1–Rad1 interface was most likely to open due to its close structural resemblance to the human PCNA interface. Finally, Jiang et al. (142) published a third 9-1-1 crystal structure (PDB id: 3GGR) and proposed association between Rad1 and human Rad17 in the complex of 9-1-1 with the clamp loader. Due to the chiral arrangement of the 9-1-1 subunits (Rad9, Rad1, and Hus1 in anticlockwise order viewed from the top) the Rad9 subunit would then be located beneath the Rfc5 subunit of the clamp loader. This would position the Rad9-Rad1 interface at the gap between Rfc5 and Rad17. Such positioning also implied the Rad9-Rad1 interface could be the gateway to open the checkpoint clamp.

Despite the availability of three independently determined 9-1-1 structures, this conundrum cannot be easily resolved for several reasons: (i) buried surface area (BSA) is an imperfect indicator of interface stability; (ii) specific residue contacts at the interfaces differ among the three available structures; (iii) there are unresolved residues close to the interfaces in two of the three experimental structures (28,141,142). In this contribution, we analyze the DNA clamp interfaces to determine the interface serving as the opening gateway in 9-1-1. Specifically, we apply the MM/PBSA (MM/GBSA) method (96,154) in the AMBER package (115,155) to compute binding energies for the subunit interfaces in PCNA and 9-1-1, providing an unbiased comparison of relative interface stabilities. Furthermore, pairwise decomposition of the MM/GBSA binding energies allowed us to delineate the significant interactions contributing to the stability of each interface. Finally, by analyzing series of multiple-trajectory steered molecular dynamics runs (pulling runs), we characterized the mechanical properties of the DNA clamp interfaces and their breakdown mechanisms under external forces.

4.3 Material and Methods

We adopted a three-step computational strategy to evaluate the clamp interfaces. First, the molecular mechanics Poisson−Boltzmann (Generalized Born) solvent accessible surface area (MM/PBSA or MM/GBSA) method was applied to calculate binding energies and compare the relative stabilities of the four subunit interfaces of human PCNA and 9-1-1. Second, the computed binding energies were decomposed to reveal aggregate per residue contributions to interface stabilization. We also carried out pairwise MM/GBSA decomposition to construct a matrix of pair interaction energies for residues forming each interface. Third, we performed multiple-trajectory steered molecular dynamics simulations to monitor the pattern of breakdown for all

energetically significant residue contacts identified in the MM/GBSA pairwise decomposition. Description for each of the three stages in our modeling protocol is given below.

Systems Setup and Molecular Dynamics- The crystal structures of PCNA (22) (PDB ID: 1VYM) and Rad9-Rad1-Hus1 (28,141,142) (PDB ID: 3A1J, 3GGR, 3G65) were obtained from the Protein Data Bank. Molecular dynamics simulations prior to MM/PBSA analysis were performed using the intact trimeric clamps. Unresolved residues in the 3G65 and 3A1J structures were fixed based on the 3GGR structure. For the subsequent MM-PBSA calculations we removed one clamp subunit to ensure only one interface (between the remaining two subunits) was evaluated. To limit computational expense, we carried out all SMD runs on reduced models, wherein we retained only the N-terminal domain of the first subunit and the C-terminal domain of the neighboring subunit. This approach is justified by the outcome of the MM/GBSA analysis, which showed no substantial contributions to the binding energies from residues outside the two adjacent domains defining each interface. The XLeap module of AMBER 9 (115,155) was used to add hydrogen atoms. All ionizable side chains were assigned to their ionization states at pH 7.0 using the WHATIF server (156). Each system was solvated with TIP3P water molecules (81) leaving a minimum distance of 10.0 Å from the protein surface to the edge of the simulation box. Counter ions were added to achieve charge neutralization and additional 100 mM NaCl concentration was introduced to mimic physiological conditions.

The systems were minimized for 5000 steps with fixed backbone atoms followed by 5000 steps of minimization with harmonic restraints on the protein backbone atoms (k=25 kcal/mol) to remove unfavorable contacts. All systems were then gradually heated to 300 K over 200 ps in the NVT ensemble while keeping the protein backbone constrained. The equilibration was continued for another 1.0 ns in the NPT ensemble and the harmonic restraints were gradually released.

Production runs were carried out in the isothermal isobaric ensemble (1 atm and 300 K) for 20 ns for the intact clamp systems (for MM/PBSA). Long-range electrostatic interactions were evaluated with the smooth particle mesh Ewald (SPME) algorithm (117). For the short-range non-bonded interactions we employed a cut off of 10 Å with a switching function at 8.5 Å. The integration time step was 2 fs and the bonds between hydrogen and heavy atoms were fixed to eliminate the most rapid oscillatory motions. The r-RESPA multiple time step method (118) was adopted with a 2 fs time step for bonded, 2 fs for short-range non-bonded interactions and 4 fs for long-range electrostatic interactions. All simulations were performed using the NAMD 2.7 code (73,74) with the AMBER Parm99SB parameter set (119) containing the force field for nucleic acids and proteins. Data were analyzed using the PTRAJ utility in AMBER (115) and custom VMD TCL scripts (157).

Binding Energy Calculation- To calculate binding energies for all four clamp interfaces we employed the molecular mechanics Poisson−Boltzmann solvent accessible surface area (MM/PBSA) method (96,154). For the purposes of MM/PBSA analysis we sampled frames at 4ps intervals from the last 10 ns of the MD trajectories (the first 10 ns were discarded as equilibration). In total 2500 frames were used for averaging. Free energy of binding can be computed as follows (same as equation 3.1):

$$\Delta G_b = \Delta E_{MM} + \Delta G_{sol} - T\Delta S \qquad (4.1)$$

where ΔE_{MM} represents gas-phase molecular mechanics binding energy (van der Waals and electrostatics); ΔG_{sol} is the change in solvation free energy and ΔS is the gas-phase entropy change upon binding. The electrostatic solvation energy contribution to ΔG_{sol} can be estimated using the finite difference Poisson−Boltzmann (PB) method. A grid size of 0.5 Å was used in the PB calculation, and the dielectric constants for protein and water were set to 1.0 and 78.0, re-

spectively. The nonpolar contribution to ΔG_{sol} was estimated by the solvent-accessible surface area according to equation (same as equation 4.2):

$$\Delta G_{nonpolar} = \gamma A + b \qquad (4.2)$$

The solvation parameters γ and b are set to be 0.0072 kcal mol^{-1} $Å^{-2}$ and 0 kcal mol^{-1}, respectively. The surface area A was calculated using MolSurf in AMBER 9 (115). The probe radius of the solvent was set to 1.4 Å. The optimized set of atomic radii in AMBER 9 was used, and the atomic charges of the protein were taken from the ff99SB force field. It is important to note that the entropy contribution was not included and, therefore, the computed values for the binding energies should be considered only relative to one another. Computed relative binding energies represent an appropriate measure of interface stability under the assumption that the gas-phase entropy terms are similar for the PCNA and 9-1-1 interfaces. The similar extent and secondary structure of the four clamp interfaces justifies such an approach.

Finally, we carried out two types of binding energy decomposition calculations. First aggregate contributions of individual residues toward interface stability were evaluated (1-dimensional decomposition). Second, a pairwise (2-dimensional) decomposition was carried out to pinpoint residue pairs contributing significantly (above -1.0 kcal/mol for PCNA and Hus1-Rad1; -2.0 kcal/mol for Rad9-Hus1 and Rad9-Rad1) toward the interface binding energy. To reduce the computational expense of the pairwise decomposition, we employed model truncation analogous to the one used in SMD. The MM/GBSA decomposition follows an established method to estimate the contributions of residues from each of the two subunits to the total binding energy by means of component analysis (158). PTRAJ module of AMBER TOOLS 12 and VMD (157) were used for the analysis of trajectories and structural visualization.

Steered Molecular Dynamics- To examine the mechanical properties of the subunit interfaces in PCNA and 9-1-1 and establish interface disruption mechanisms under external forces we relied on the steered molecular dynamics (SMD) method (122). To initiate the SMD runs we selected snapshots from preliminary 10 ns unbiased simulations of PCNA and 9-1-1 (3A1J) at equal intervals along the trajectories. These statistically uncorrelated configurations were used to initiate constant velocity steered molecular dynamics (SMD). In each interface we have applied constraints to the -strand of the N-terminal domain while applying a harmonic force k = 15 kcal mol^{-1} $Å^{-2}$ to the center of mass of the other strand. The pulling direction was along the line connecting the centers of mass of the two domains. The constant pulling velocity was 1 Å ns^{-1} and the pulling run had a total duration of 12 ns so as to achieve an opening of 12 Å of the interface. During the pulling trajectories 1200 structures were sampled at intervals of 10 ps for analysis. In total, the SMD simulation time exceeded 1.4 s necessitating the use of extensive supercomputing resources. Thus, the project required over 1 million CPU hours at two supercomputing facilities (NICS Kraken and NERSC Hopper II supercomputers).

Analysis of the SMD runs was carried out by following the time evolution of the contacts identified as significant contributors to binding in the pairwise (2-D) MM/GBSA decomposition (158). The contacts were classified as hydrogen bonds, salt-bridges, hydrophobic interactions and suitable cutoffs applied to determine the presence of the contact in the trajectory frames. For each contact detected during the pulling trajectory the time of last occurrence (allowing for the possibility of contact breaking and reforming) was used to identify the breakdown time. The interface opening mechanism could then be followed on a contact map wherein the color-coding scheme signified groups of contacts cleaved within the same time interval.

4.4 Results and Discussion

Rad9-Rad1 interface requires least amount of energy to open – To understand the structural and energetic basis for interface stability in PCNA and 9-1-1, we analyzed all four subunit interfaces (PCNA-PCNA, Rad9-Rad1, Rad9-Hus1 and Hus1-Rad1) through MD simulations along with the application of the molecular mechanics Poisson−Boltzmann solvent accessible surface area (MM/PBSA) method (96,154). We employ complete and fully relaxed models of the trimeric sliding clamps (PDB id: 3GGR, 3A1J and 3G65) (28,141,142) embedded in an aqueous solvent environment. Subunit interface binding energies were averaged over 2500 independent conformations. Various components of the subunit interaction energy ΔG_b were evaluated for all three available 9-1-1 X-ray structures (Table 4.1, 4.2 and 4.3). When MM/PBSA is applied to extended interfaces it is challenging to accurately assess the gas-phase entropic contribution. Therefore, quantitative comparison of the interfaces involved only the enthalpic contribution and the values in Tables 4.1, 4.2 and 4.3 represent binding energies rather than free energies. As expected, due to lack of entropy-enthalpy compensation, the absolute binding energies are overestimated. Nonetheless, previous applications of MM/PBSA to protein interfaces have shown the usefulness of comparing the relative stabilities of binding epitopes and identifying binding hotspots based on energy decomposition (126,159).

Table 4.1 Binding energy analysis (kcal mol^{-1}) for the DNA clamp interfaces in PCNA and 9-1-1 (3GGR model)

Contribution	3GGR			
	PCNA/PCNA	Rad9/Hus1	Hus1/Rad1	Rad9/Rad1
ΔE_{ele}	866.04 (46.68)	-479.45 (88.28)	-38.87 (37.04)	-584.25 (53.97)
ΔE_{vdw}	-80.65 (5.38)	-98.14 (7.77)	-92.49 (6.27)	-80.12 (6.12)
$\Delta G_{nonpolar}$	-11.68 (0.42)	-16.58 (0.70)	-12.62 (0.55)	-12.84 (0.46)
ΔG_{polar}	-827.27 (46.05)	502.33 (84.81)	69.78 (34.50)	618.16 (50.82)
ΔG_{sol}[a]	-838.95 (45.83)	485.75 (84.34)	57.17 (34.42)	605.31 (50.74)
ΔG_{ele}[b]	38.77 (10.49)	22.88 (10.65)	30.91 (7.73)	33.90 (10.94)
ΔG_b	**-53.56** (8.91)	**-91.84** (9.08)	**-74.19** (8.16)	**-59.06** (8.68)
ΔG_b Ratio	**1.00**	**1.71**	**1.39**	**1.10**
BSA (Å²)	1555	2117	1629	1645

[a]Polar/nonpolar ($\Delta G_{sol}=\Delta G_{polar}+\Delta G_{nonpolar}$) contributions to ΔG_b. [b]Electrostatic ($\Delta G_{ele}=\Delta E_{ele}+\Delta G_{polar}$) contributions to ΔG_b. Calculation of ΔG_b does not explicitly consider entropy contributions. Standard deviations are shown in parentheses. Averaged buried surface areas (BSA) for the interfaces are given units of Å^2. These apply to Table 4.2 and 4.3 as well.

Table 4.2 Binding energy analysis (kcal mol^{-1}) for the DNA clamp interfaces in PCNA and 9-1-1 (3A1J model)

Contribution	3A1J			
	PCNA/PCNA	Rad9/Hus1	Hus1/Rad1	Rad9/Rad1
ΔE_{ele}	866.04 (46.68)	-411.36 (63.18)	-89.79 (35.88)	-694.78 (44.58)
ΔE_{vdw}	-80.65 (5.38)	-89.24 (8.55)	-84.54 (5.68)	-76.44 (5.51)
$\Delta G_{nonpolar}$	-11.68 (0.42)	-14.18 (0.96)	-11.46 (0.59)	-13.03 (0.56)
ΔG_{polar}	-827.27 (46.05)	437.34 (60.61)	109.80 (34.61)	724.07 (42.57)
ΔG_{sol}[a]	-838.95 (45.83)	423.16 (59.93)	98.34 (34.45)	711.04 (42.40)
ΔG_{ele}[b]	38.77 (10.49)	25.98 (13.44)	20.01 (7.49)	29.29 (11.87)
ΔG_b	**-53.56** (8.91)	**-77.44** (11.37)	**-75.99** (6.34)	**-60.18** (11.29)
ΔG_b Ratio	**1.00**	**1.45**	**1.42**	**1.12**
BSA (Å²)	1555	1836	1475	1639

Table 4.3 Binding energy analysis (kcal mol^{-1}) for the DNA clamp interfaces in PCNA and 9-1-1 (3G65 model)

Contribution	3G65			
	PCNA/PCNA	Rad9/Hus1	Hus1/Rad1	Rad9/Rad1
ΔE_{ele}	866.04 (46.68)	-583.33 (49.33)	-97.19 (30.16)	-753.18 (54.20)
ΔE_{vdw}	-80.65 (5.38)	-91.96 (8.42)	-85.71 (4.85)	-80.15 (7.87)
$\Delta G_{nonpolar}$	-11.68 (0.42)	-15.48 (0.98)	-12.21 (0.59)	-12.76 (0.88)
ΔG_{polar}	-827.27 (46.05)	590.96 (46.84)	115.53 (29.59)	784.04 (54.66)
ΔG_{sol}[a]	-838.95 (45.83)	575.48 (46.56)	103.32 (29.32)	771.28 (54.10)
ΔG_{ele}[b]	38.77 (10.49)	7.62 (12.65)	18.33 (7.91)	30.86 (12.69)
ΔG_b	**-53.56** (8.91)	**-99.82** (10.87)	**-79.59** (6.78)	**-62.05** (11.58)
ΔG_b Ratio	**1.00**	**1.86**	**1.49**	**1.16**
BSA (Å²)	1555	2037	1556	1621

From the data in Table 4.1 (also Table 4.2 and 4.3) we immediately conclude that the 9-1-1 interfaces differ substantially in stability. For the most complete structure (3GGR), the computed binding energies ΔG_b establish the following order of stability: Rad9-Hus1>Hus1-Rad1>Rad9-Rad1≈PCNA-PCNA (ratios of ΔG_b of 1.71 : 1.39 : 1.10 : 1.0). This ordering is maintained regardless of which x-ray structure was used in constructing the model (*e.g.* for 3G65 the ΔG_b ratios are 1.86 : 1.49 : 1.16 : 1.0). The only outlier value is the Rad9-Hus1 binding energy in the 3A1J model. Missing residues at this interface were added by homology modeling (using 3GGR as template) likely leading to a marginally lower ΔG_b. Comparison among the independently set-up 3G65, 3A1J and 3GGR models reveal that the binding energy ratio ΔG_b (Rad9-Rad1)/ΔG_b (PCNA) is remarkably consistent irrespective of the initial structural model (ratio of 1.16, 1.12 and 1.10, respectively). The Rad9-Rad1 interface is the weakest and, thus, a clear choice to be the gateway for opening the checkpoint clamp. Furthermore, the binding energy of

Rad9-Rad1 is practically indistinguishable from the binding energy of the PCNA-PCNA interface. Therefore, the energetic requirement for the Rad17-RFC2–5 complex to open 9-1-1 is essentially same as the requirement for RFC to open the PCNA ring. From an evolutionary standpoint this outcome implies 9-1-1 and PCNA may have been subject to evolutionary pressure to optimize the overall stability of one subunit interface (Rad9-Rad1 or PCNA-PCNA) to match the requirements of the clamp loading machinery.

Despite the almost perfect match in overall binding energies, the stability of Rad9-Rad1 and PCNA-PCNA interfaces is achieved by very different means. The PCNA interface is largely flat and hydrophobic and does not display electrostatic complementarity between the opposing binding epitopes (Figure 4.1C). By contrast, the Rad9-Rad1 interface features more polar contacts and a shorter hydrophobic patch between the two antiparallel β-strands S9-S13 (141) (Figure 4.3B). The binding surfaces from the Rad9 and Rad1 subunits display high electrostatic complementarity (Figure 4.1F) as reflected in the computed gas-phase electrostatic energy. ΔE_{ele} is favorable for the Rad9-Rad1 interface but unfavorable for the PCNA interface (Table 1). By contrast, the electrostatic solvation energy (ΔG_{polar}) disfavors Rad9-Rad1 binding whereas it favors PCNA interface formation. Rad9-Hus1 and Hus1-Rad1 also display favorable ΔE_{ele} due to electrostatic complementarity between the 9-1-1 subunits (Figure 4.1D-F) – a feature, which is absent in the PCNA interface.

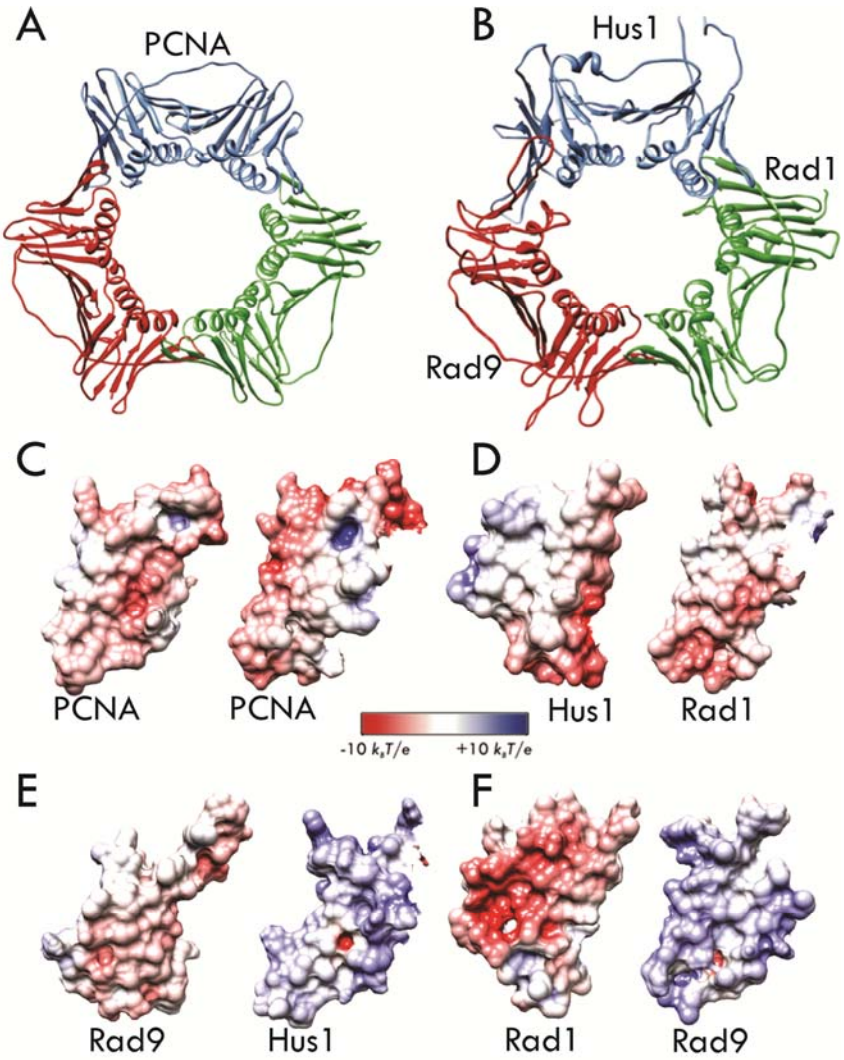

Figure 4.1 Common toroidal architecture of the DNA clamps PCNA and 9-1-1

A) Structure of PCNA with the equivalent subunits shown in red, blue and green; B) Structure of the checkpoint clamp with Rad9 shown in red, Rad1 in green and Hus1 in blue; C) Surface electrostatics of the PCNA interface; D) Surface electrostatics of the Hsu1-Rad1 interface; E) Surface electrostatics of the Rad9-Hus1 interface; F) Surface electrostatics of the Rad9-Rad1 interface. Electrostatic potential is mapped onto the molecular surface for each interface and color-coded from red (negative) to blue (positive).

Finally, it's notable that the computed BSA values for the interfaces averaged over the simulation trajectories (Table 4.1, 4.2 and 4.3) do not correlate well with the computed binding energies. Using only $\Delta G_{nonpolar}$ (or alternatively BSA) to evaluate the 9-1-1 interfaces would suggest Hus1-Rad1 to be marginally less stable than Rad9-Rad1, inverting the order of stability established through the total binding energies ΔG_b. However, interface formation is driven by both polar (ΔG_{polar} and ΔE_{ele}) and non-polar interactions ($\Delta G_{nonpolar}$ and ΔE_{vdw}), which in the balance determine ΔG_b. Since BSA accounts only for the $\Delta G_{nonpolar}$ contribution, it is a poor gauge of relative interface stability.

Structural determinants of PCNA and 9-1-1 interface stability – To delineate the structural features that affect 9-1-1 interface stabilities, we decomposed the total binding energies ΔG_b into aggregate per residue contributions (1-D decomposition) (158). All residues contributing substantially toward interface stabilization/destabilization above a ±1.5 kcal mol^{-1} threshold are shown in Figure 4.2 and Figure 4.3. Residue contributions are also color mapped onto the structure of the respective interfaces. First, we note that all interfaces in 9-1-1 and PCNA are stabilized entirely by local interactions with no above-threshold contributions arising from residues outside the adjacent N-terminal and C-terminal domains. Despite dissimilar sequences, the 9-1-1 and PCNA interfaces share common structural features (22,28). At the core of each clamp interface is an antiparallel β-sheet (strands S9 and S13), forming part of the outer shell of the trimeric PCNA or 9-1-1 ring. This β-sheet is reinforced to varying degrees by two adjacent α-helices (labeled H2 and H3) facing toward the central cavity of each clamp. Such an arrangement of secondary structure elements imparts structural stability to the clamps. Residue contacts in the loop regions above and below the central β-sheet contribute with varying extent to the overall binding energy of each interface. The planar ring architecture results in the clamps being two-sided: the

front (top) face exposes the IDC loops and is thus responsible for binding most replication factors (notably DNA polymerases and the clamp loader). The back (bottom) side of PCNA is characterized by three loops denoted as P-loops (or guide loops). The 9-1-1 interfaces likewise have a front and back side and in the analyses that follow we have oriented all interfaces so that the IDCL-presenting face is denoted as "top" and the back face is "bottom". Classification of the opening pathways follows the same convention.

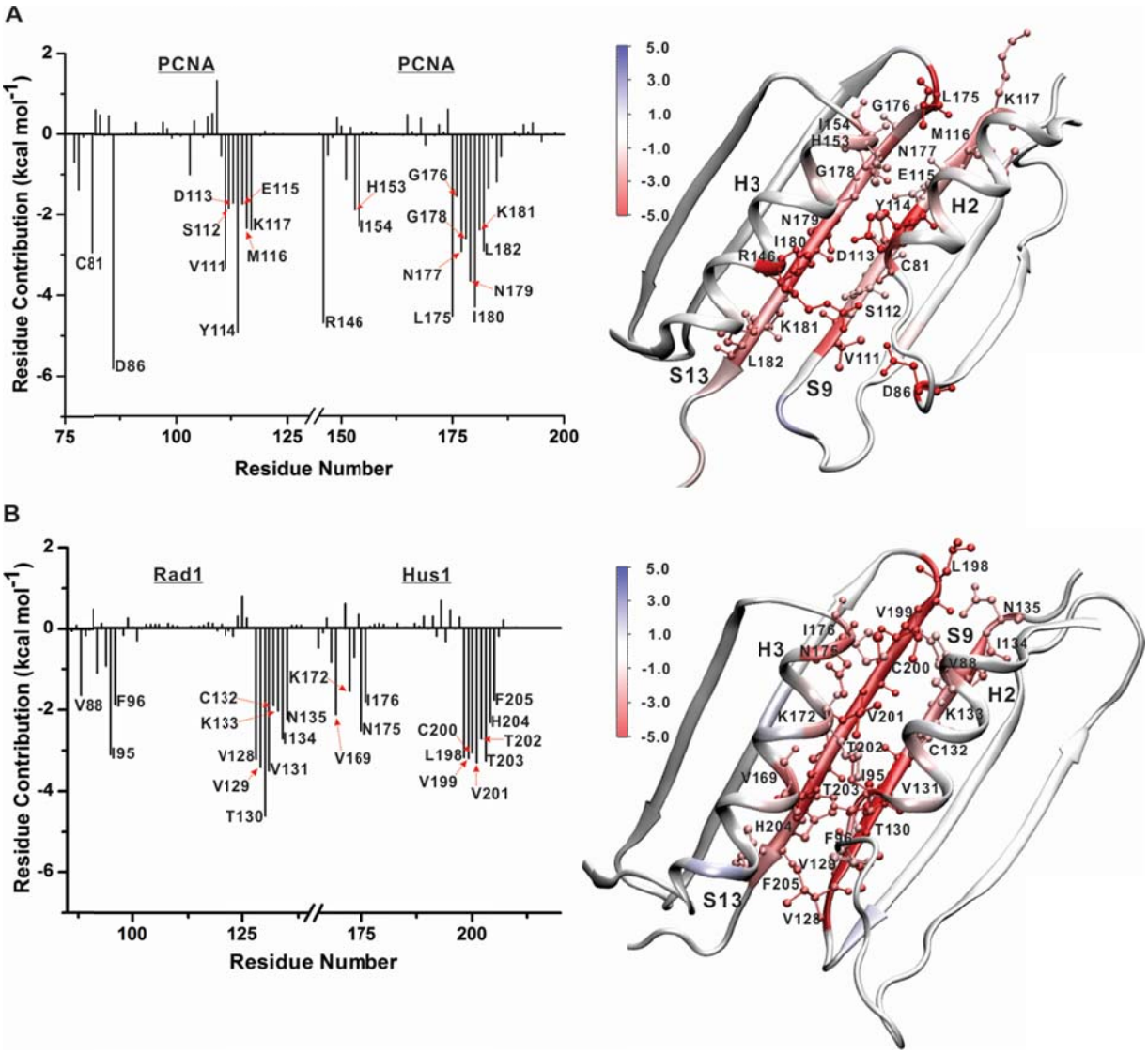

Figure 4.2 Origins of interface stability for the PCNA and Hus1-Rad1 interfaces from 1-D MM/GBSA decomposition analysis

A) Aggregate binding energies for the individual residues of the PCNA interface (left) and residue contributions mapped onto the structure (right); B) Aggregate binding energies for the individual residues of the Hus1-Rad1 interface (left) and residue contributions mapped onto the structure (right). Only residues contributing above a ±1.5 kcal mol^{-1} threshold in G_b are represented.

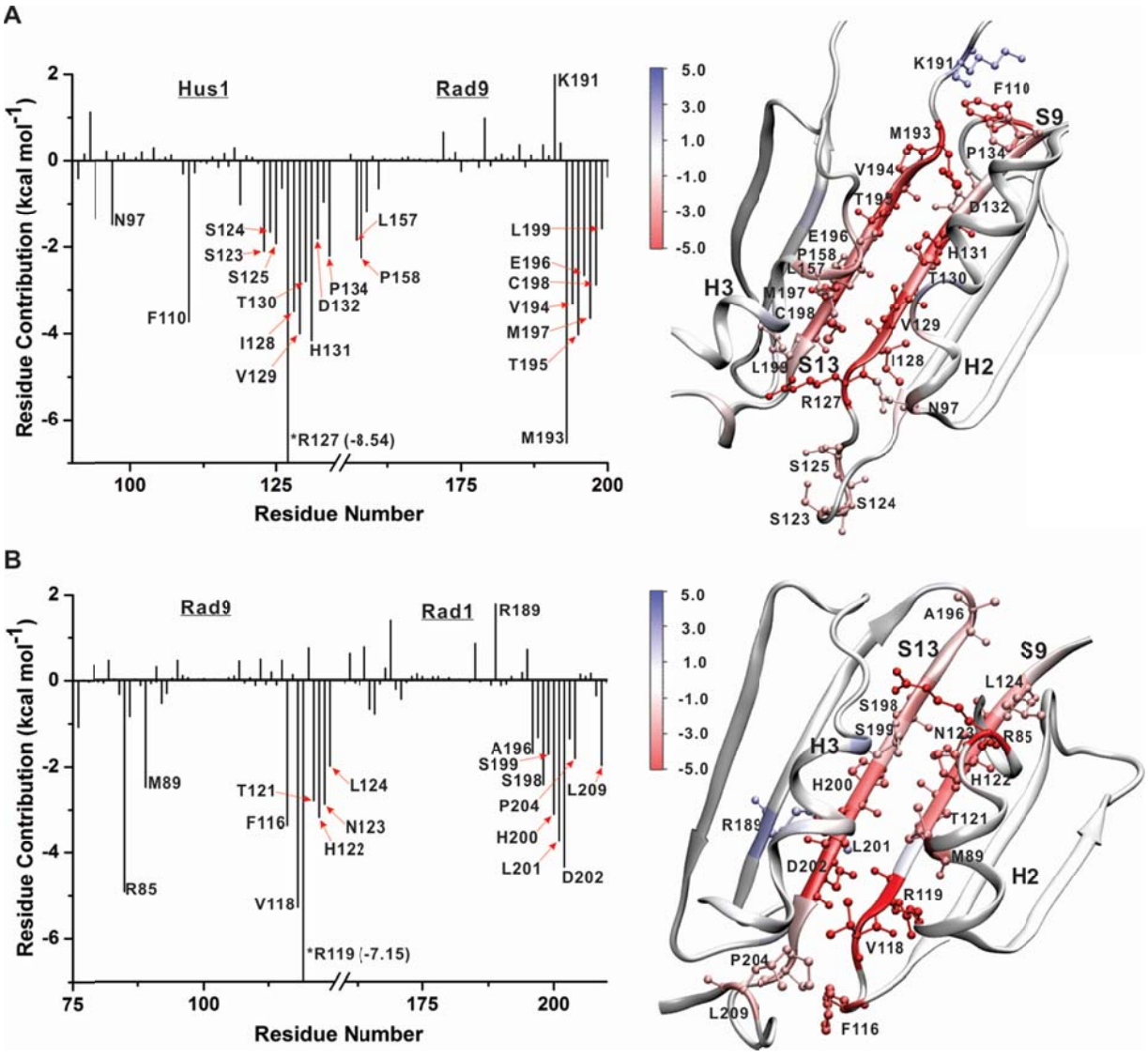

Figure 4.3 Origins of interface stability for the Rad9-Hus1 and Rad9-Rad1 interfaces from 1-D MM/GBSA decomposition analysis

A) Aggregate binding energies for the individual residues of the Rad9-Hus1 interface (left) and residue contributions mapped onto the structure (right); B) Aggregate binding energies for the individual residues of the Rad9-Rad1 interface (left) and residue contributions mapped onto the structure (right). Only residues contributing above a ±1.5 kcal mol^{-1} threshold in G_b are represented.

In Figure 4.2 we first compare the interfaces PCNA-PCNA and Hus1-Rad1, which were suggested by Sohn et al. (141) to be structurally most similar. Indeed, both interfaces are flat and hydrophobic, which is clearly reflected in the presence of two prominent largely symmetric peaks in the binding energy graphs. The peaks correspond to the two antiparallel β-strands at the core of the PCNA or Hus1-Rad1 interface and together account for most of the computed ΔG_b. The roughly symmetric peaks suggest that there is no preferred directionality to open these two interfaces from either the top or the bottom side. For PCNA the central β-sheet (V111-K117 and L175-L182) is the primary determinant of interface stability with 15 residues having above-threshold contributions. The interface is stabilized by the presence of 7 stable backbone hydrogen bonds in a largely hydrophobic sequence context. The residues in the middle of the sheet contribute more significantly to ΔG_b than the flanking residues at the edge of the sheet. There are very few residues outside the β-sheet with above-threshold contributions. Interactions between the adjacent helices H2 and H3 are considerably less substantial, except for a single salt bridge between residues D86 and R146 at the back of the interface. Together these two residues add -10.56 kcal/mol to the total binding energy. Three additional residues C81, H153 and I154 from helices H2 and H3 also contribute -2.95, -1.89 and -2.30 kcal/mol, respectively.

While the Hus1-Rad1 interface superficially resembles the PCNA interface (16 contributing residues in the core β-sheet; V128-N135 and L198-F205), the average computed binding energy per residue within the β-sheet is higher (-2.91 kcal/mol) for Hus1-Rad1 compared to PCNA-PCNA (-2.65 kcal/mol). Additionally, there are significantly more hydrophobic residues from the adjacent H2 and H3 helices contributing above the threshold level (V88, I95, F96, V169, N175, I176). Such differences in the extent of the hydrophobic contacts justify the observed order of stability with the PCNA-PCNA interface being easier to open than Hus1-Rad1.

In Figure 4.3 we compare the two more polar interfaces - Rad9-Hus1 and Rad9-Rad1. Once again interactions within the central β-sheets play the most prominent role (R127-P134 and M193-L199 for Rad9-Hus1; V118-L124 and A196-P204 for Rad9-Rad1;) resulting in two primary peaks in the 1-D MM/GBSA decomposition graphs. In contrast to PCNA-PCNA and Hus1-Rad1 the peaks in Figure 3 are largely asymmetric, featuring a broader distribution of binding energies. Stronger contacts are clustered at the bottom of the two interfaces. These stronger contacts are likely to break late during the interface rupture induced by the clamp loader, while weak contacts would break first. Therefore, the decomposition of ΔG_b (either 1-D or pairwise 2-D) gives us a clear prediction of the preferred direction for opening the subunit interfaces under external forces. The observed asymmetry of the peaks for Rad9-Rad1 and Rad9-Hus1 is indicative of preferred direction to open both interfaces from the top side. This trend is especially well pronounced for Rad9-Rad1, which is also the most polar of the three 9-1-1 interfaces. However, there are also substantial differences between Rad9-Rad1 and Rad9-Hus1 in terms of the number and type of residues contributing to ΔG_b. Rad9-Rad1 is clearly the weaker interface due to fewer contacts within the core β-sheet and only two above-threshold contributions from outside the β-sheet: residues M89 and R85 from helix H2. By contrast, the Rad9-Hus1 interface features not

only exceptionally strong contributions within the core β-sheet (per residue average of -3.51 kcal/mol) but also many contributing residues outside the core of the interface - N97 from helix H2, L157 and P158 from helix H3, F110 and S123-S125 from loop regions of the interface. Thus, Rad9-Hus1 appears to be the most substantial of all 9-1-1 subunit interfaces.

Clamp opening pathways under external force – To address the mechanisms and pathways for clamp opening we monitored subunit interface disruption under external forces. This was accomplished by multiple-trajectory steered molecular dynamics (SMD) runs using truncated PCNA and 9-1-1 models. Truncation to include only the N-terminal and C-terminal domains from each interface is fully justified by the MM/GBSA results, which showed the clamp subunits to be held together entirely by local interactions. The external forces were applied to pull apart the β-strand at the core of each interface. Visual classification of the SMD trajectories/opening pathways (Table 4.4) was performed prior to detailed contact analysis. Our SMD simulations showed that similar to the case of protein folding, the clamp opening for both PCNA and 9-1-1 does not occur through a single uniform breakdown mechanism. Instead, the results are consistent with the existence of an ensemble of parallel clamp-opening pathways, some more prevalent than others. These pathways are characterized by specific groups of contacts that differentially stabilize the regions of the interface and, thus, determine the spatial and temporal patterns of breakdown. The stochastic nature of interface opening is consistent with recent work of Benkovic *et al*. (111), which pointed out that clamp loading proceeds by an inefficient and largely heuristic search. Instead of a single, highly confined pathway the process involves multiple attempts and many intermediates (possibly also off-pathway states). Thus, in addition to its primary function, the clamp loader has to recycle incorrectly loaded clamps.

Table 4.4 Classification of the DNA clamp opening pathways from SMD

Number of pathway classified as “top-down”, “bottom-up” or “concerted” for each clamp subunit interface are shown.

Pathway Classification	Subunit Interface			
	PCNA/PCNA	Rad9/Hus1	Hus1/Rad1	Rad9-Rad1
Top-down	12	16	9	17
Bottom-up	10	2	9	5
Concerted	8	4	1	0

Despite the stochastic nature of clamp opening, the dominant pathways exhibit certain common features. Interface break-up almost invariably involves unzipping of the central β-sheet (strands S9-S13) followed by separation of the more loosely connected H2 and H3 helices and the loop regions of the interface. The unzipping of the β-sheet can occur in the top-down, bottom-up direction or by cooperative collapse of the contacts between the S9 and S13 strands. Therefore, in Table 4.4 we classify the opening pathways as “top-down”, “bottom-up” or “concerted”. We note that the classification is based on the overall direction of opening for the entire interface (not just the β-sheet collapse). In a small number of SMD runs (less than 10%), the initial opening was followed by partial re-formation of the contacts at the interface. These runs were excluded from the pathway classification in Table 4.4.

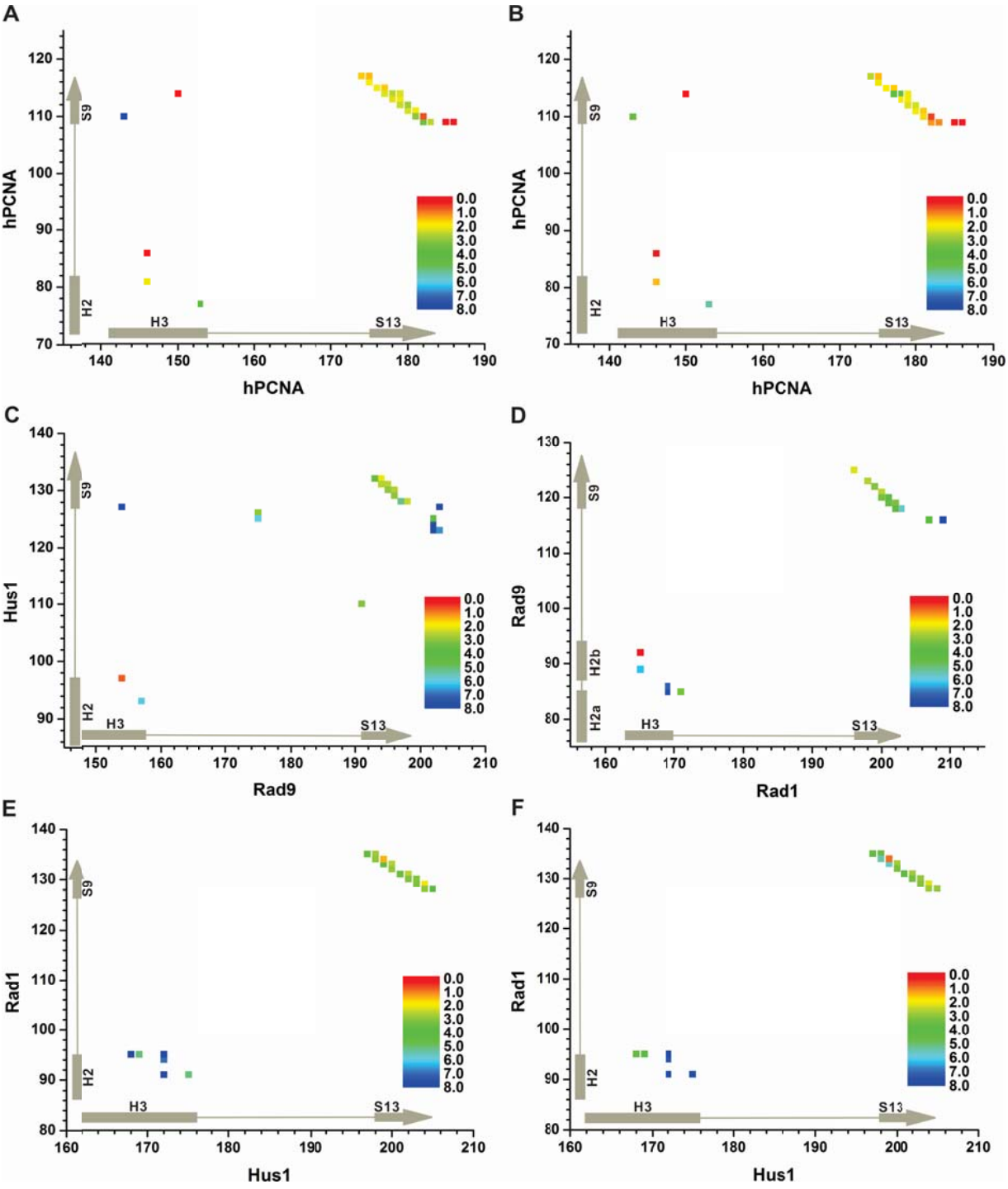

Figure 4.4 Dominant pathways for DNA clamp opening from contact analysis of SMD simulations

Time evolution of significant contacts for A) the PCNA interface in the “top-down” pathway; B) the PCNA interface in the “bottom-up” pathway; C) the Rad9-Hus1 interface in the “top-down” pathway; D) the Rad9-Rad1 interface in the “top-down” pathway; E) the Hus1-Rad1 interface in the “top-down” pathway; F) the Hus1-Rad1 interface in the “bottom-up” pathway. Averaged break-up times of above-threshold contacts from the pairwise MM/GBSA decomposition are color-mapped on the panels from red (0ns) to blue (8ns).

Under external forces the four clamp interfaces display distinct rupture mechanisms. For PCNA-PCNA and Hus1-Rad1 clamp opening in the “top-down” or “bottom-up” direction is equally probable. Indeed for PCNA all three pathways, including the concerted pathway, are equally represented. For Hus1-Rad1 there is a higher tendency for the opening process to start from the ends of the interface. However, once initiated, the unzipping is completed within a narrow time interval, which is indicative of almost cooperative disruption of the backbone hydrogen bonds at the interface core (Figure 4.4E and 4.4F). By contrast the more polar Rad9-Rad1 and Rad9-Hus1 interfaces open preferentially in the “top-down” direction. This outcome is fully consistent with our prediction from the one-dimensional and pairwise MM/GBSA decompositions (Figure 4.2, Figure 4.3 and Figure 4.5). The origin of preferred directionality lies in the uneven distribution of strong polar or charged contacts along these interfaces. Moreover, polar contacts found outside the central β-sheet exert more significant bias on the opening direction. The reason for this outcome is that these contacts involve specific side-chain interactions, while contacts within the core β-sheet occur mostly through the backbone and are thus non-specific. Preferred directionality is lost when the core of the interface is primarily hydrophobic and held together by backbone hydrogen bonds (as is the case for PCNA-PCNA and Hus1-Rad1).

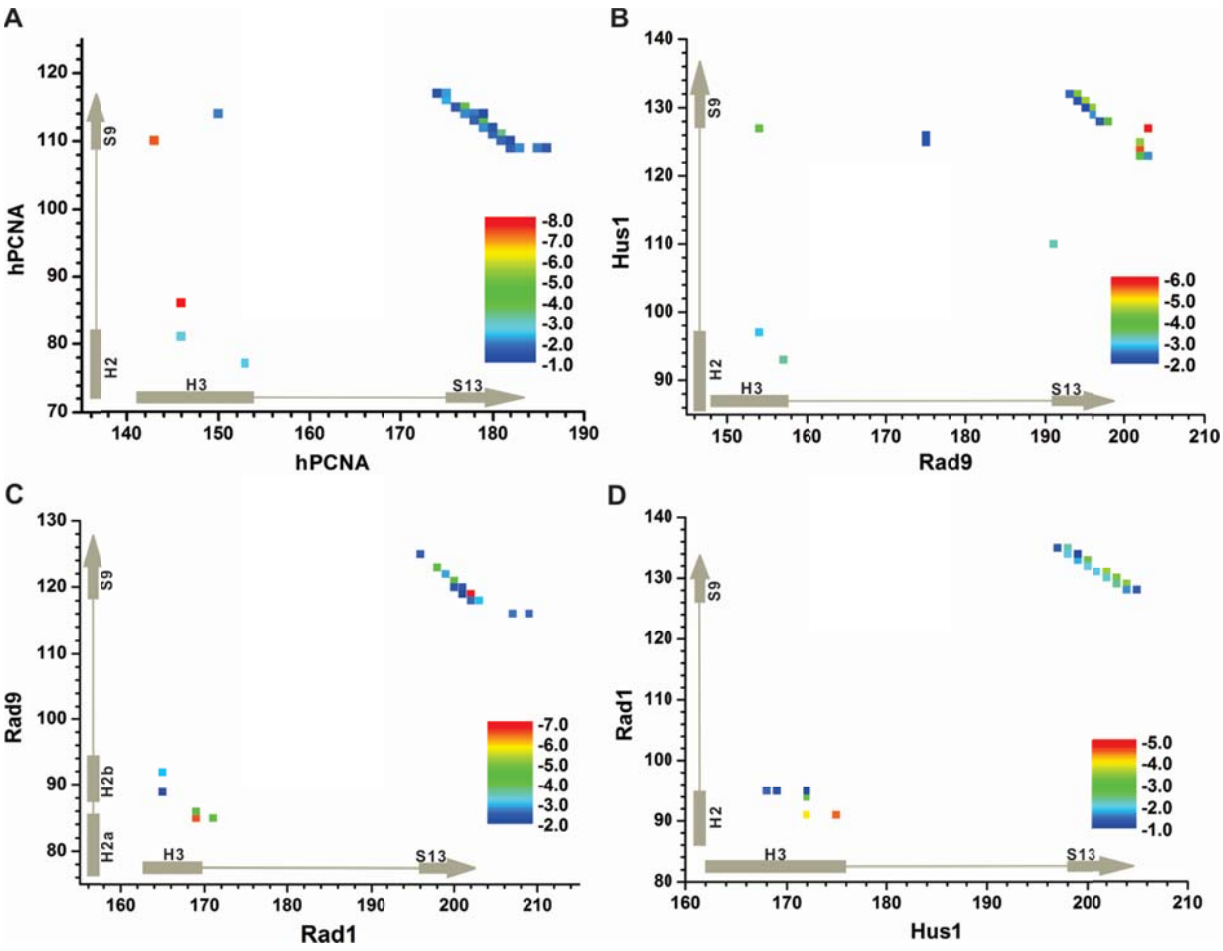

Figure 4.5 Significant (above-threshold) contacts from pairwise MM/GBSA binding energy decomposition values

A) the PCNA interface; B) Rad9-Hus1 interface; C) the Rad9-Rad1 and D) the Hus1-Rad1 interface. Binding energies of residue pairs (in kcal/mol) are color-mapped on the panels from red to blue.

Next, we carried out detailed analysis of the interface rupture mechanisms in PCNA and 9-1-1. To this end we subdivided all pulling runs by type of opening pathway (Table 4.4) and analyzed the pattern of contact disruption as the interfaces were pulled apart. Only residue con-

tacts contributing above-threshold in the pairwise MM/GBSA decomposition (Figure 4.5) were considered and their time evolution monitored throughout the pulling trajectory. To follow the contact dynamics, we first classified the MM/GBSA residue pairs by type of interaction. For hydrogen bond contacts (backbone, side-chain H-bonds and direct salt bridges) we employed heavy atom distance cutoff of 3.3 Å and an angle cutoff of 50º. Hydrophobic contacts were identified by considering whether heavy atoms from the first hydrophobic residue were in van der Waals contact with any heavy atom from the second hydrophobic residue. Polar residues not involved in hydrogen bonding were treated by applying the same distance criterion used for hydrophobic contacts. The time of last occurrence (within the specified cutoffs) of each MM/GBSA pair was recorded and these times were averaged over the trajectories belonging to the same pathway. Naturally, identical contacts collapse at different times in different trajectories. However, we observed that within each pathway the rupture events for specific contacts occurred within narrow time intervals, justifying the use of averaged break-up times in defining the opening mechanism.

Results from this analysis are shown in Figure 4.4 and reveal the characteristics of the dominant pathways for subunit interface disruption in PCNA and 9-1-1. Description of the contacts includes average break-up times in nanoseconds and interaction energies ΔG_b from the pairwise MM/GBSA decomposition (noted in parenthesis after each contact). For PCNA (Figure 4.4A and 4.4B) the opening mechanism is dominated by the disruption of the two antiparallel β-strands in the outer shell of the clamp. There is no clear distinction between the "top-down" and "bottom-up" pathways in terms of timing of contact disruption. Indeed the majority of contacts between the *β*-strands disappear in a relatively narrow time interval from 1.5 ns to 4 ns regardless of whether strand unzipping is initiated from the top or bottom side.

For the 9-1-1 interfaces, it generally took more time to initiate S9-S13 strand separation. This is illustrated by the shift in color toward the green/blue end of the scale as compared to PCNA in overall agreement with our ΔG_b results. Pulling the Hus1-Rad1 interface did not result in a single dominant pathway (Figure 4.4E and 4.4F) consistent with the flattened binding energy profile in the 1-D or pairwise 2-D MM/GBSA decompositions. Rupture of the core β-sheet in Hus1-Rad1 occurred with no obvious directionality and within an even narrower time range compared to PCNA: 2.9 ns – 3.6 ns for the top-down pathway, and 2.9 ns – 4.0 ns for the bot-tom-up pathway. The pair Val199–Lys133 (-1.76 kcal/mol) lasted longer than any other β-sheet contact (5.8 ns) due to the long flexible side chain of Lys133 reaching across the interface. The strongest contacts from the H2, H3 α-helices N175-D91 (-4.54 kcal/mol), K172-D91 (-4.05 kcal/mol) and K172-S94 (-2.80 kcal/mol) lasted beyond the complete separation of the β-strands in both pathways (rupture times of 8.2, 8.1 and 7.5 ns, respectively). Therefore helix separation on the inner side of the interface occurred after the rupture of the core β-strands. We conclude that just like in PCNA, the Hus1-Rad1 interface opens by cooperative, non-directional unzipping of the core β-strands dominated by disruption of hydrophobic contacts.

The top-down pathway was the dominant mechanism for opening the Rad9-Hus1 inter-face (Figure 4.4C). Disruption of the core β-strands for this interface occurred in a much broader time range (2.0 ns – 7.0 ns) compared to PCNA, indicating the presence of persistent contacts whose break-up requires much larger external forces. This notion is consistent with Rad9-Hus1 being the most stable of all 9-1-1 interfaces. The strongest interaction identified in the pairwise decomposition was between charged residues D203–R127 (-10.8 kcal/mol) and was positioned at the extreme back side of the interface. Not surprisingly, this was the last contact to break at 10.6 ns. Other robust contacts included E154–R127 (-4.12 kcal/mol), E202–S125 (-4.45 kcal/mol),

E202–Ser124 (-5.62 kcal/mol), D203–S123 (-2.66 kcal/mol) and L157–K93 (-3.38 kcal/mol), which all resided in the bottom region of the interface and cooperatively resisted separation until average break-up times well above 6.5 ns. All of these pairs included polar or charged residues, corroborating the importance of charged interactions in maintaining interface stability and establishing a predominant opening direction.

The opening mechanism of the Rad9-Rad1 interface (Figure 4.4D) revealed an even more pronounced tendency for directional opening as compared to Rad9-Hus1. Rupture of the interface was initiated from the top side, which incidentally is also the side binding the clamp loader. Thus it is attractive to speculate that the Rad9-Rad1 interface is weakest precisely at the point of application of mechanical force by the clamp loader. A group of resilient pairwise interactions at the bottom of the interface is responsible for the observed "top-down" directionality of opening. These interactions feature both charged and hydrophobic residue contacts: R119-D202 (-16.50 kcal/mol), R85-E169 (-6.55 kcal/mol), F116-L209 (-2.54 kcal/mol), F116-S207 (-2.66 kcal/mol), and V118-Y203 (-2.98 kcal/mol). The strongest contact R119–D202 resides at the lower end of the β-sheet and along with V118-Y203 did not break until a very late stage (5.8 ns) of interface separation. Phe116–Leu209 was typically the last contact to rupture (11.7 ns) in most of the SMD trajectories. Despite the overall top-down rupture pattern, there were a couple of persistent contacts at the top of Rad9-Rad1. These involved the salt bridge R85–E169 (-6.55 kcal/mol) and the S86–E169 pair, both located at the top of the H2, H3 α-helices. Maintaining these two contacts until the latter stages of interface separation (7.8 ns) was made possible by the long side chains of Arg85 and E169, which could reach across the partially opened S9-S13 β-strands. Once again, the results highlight the importance of salt bridges in determining the interface opening mechanism and for the overall architecture of the DNA clamps.

4.5 Concluding remarks

The toroidal architecture adopted by DNA clamps is a perfect example of how molecular shape follows function. Despite low sequence similarity, clamps from bacteriophage to archaea and humans all form trimeric rings encircling DNA. The function of these protein rings is to serve as mobile scaffolds in the coordinated assembly of the DNA replication and repair machinery. In this case, function requires a subtle balance between the ability to form stable trimers and the necessity to open to be loaded onto DNA. All clamps employ a common loading mechanism, wherein a single subunit interface is ruptured by the clamp loader to allow threading of primer-template DNA and followed by closure of the ring.

Why have two distinct clamps, PCNA and 9-1-1, evolved in humans? One may speculate that the purpose of the heterotrimeric 9-1-1 is to ensure selective recruitment of different partners to different subunits on the same ring-shaped platform. This cannot be accomplished by PCNA, except in a stochastic manner, because the PCNA subunits are equivalent. Importantly, the different subunits of 9-1-1 not only differentiate the clamp from PCNA but also open up the possibility that affinity differences among the clamp interfaces may have functional significance. Specifically, it has been proposed that the Rad9-Rad1 interface of the checkpoint clamp is the weakest and, therefore, serves as the opening gate for loading 9-1-1 onto chromatin. However, this proposal remains controversial with others positing Hus1–Rad1 is the opening gate based on structural resemblance to the PCNA interface. Through detailed subunit interface analysis, we identified Rad9-Rad1 as having the lowest affinity among all 9-1-1 interfaces. This finding makes Rad9-Rad1 a clear candidate to be the 9-1-1 opening gate. The Hus1-Rad1 interface was similar to PCNA in overall structure and hydrophobicity and opened up by an analogous rupture mechanism. However, the total binding energy ΔG_b of Hus1-Rad1 was higher than the energy of

Rad9-Rad1. Furthermore, the affinities of Rad9-Rad1 and PCNA-PCNA were very closely matched, despite pronounced dissimilarities in interface contacts and polar character. Thus, matching affinities may reflect the common energetic requirements of the clamp loading machinery to open PCNA and Rad9-Rad1.

Using external SMD forces as a surrogate of the clamp loader, we also identified the dominant pathways for interface disruption in PCNA and 9-1-1. In agreement with recent experimental work, we demonstrate the stochastic nature of clamp opening and observe an ensemble of opening pathways. We find that the overall directionality and cooperativity of the dominant pathways are distinct among the four clamp interfaces. Groups of contacts (especially charged and polar contacts) differentially stabilize the top and bottom regions of the Rad9-Rad1 and Rad9-Hus1 interfaces and determine the spatial and temporal patterns of breakdown. In Rad9-Rad1 and Rad9-Hus1 the dominant pathway involves unzipping of secondary structure elements in the "top-down" direction. By contrast the predominantly hydrophobic Hus1-Rad1 and PCNA interfaces are disrupted concertedly with no preferred directionality. Collectively, the results provide a framework for future experiments to examine the DNA clamp opening mechanisms. Specifically, single molecule micromanipulation experiments (with optical or magnetic tweezers) along with mutational substitution of residue pairs at the clamp interfaces could be used to probe the computationally identified pathways and establish the barriers and energy requirements for clamp opening.

CHAPTER 5. SWITCHING FROM POLYMERASE TO EXONUCLEASE ACTIVITY IN THE DNA POLYMERASE POLB INVOLVES LARGE-SCALE CONFORMATIONAL REARRANGEMENT OF POLB ON PCNA

5.1 Abstract

Replicative DNA polymerases (Pols) frequently possess two distinct DNA processing activities – DNA synthesis (polymerization activity) or proofreading (3'-5' exonuclease activity). The polymerase (pol) and exonuclease (exo) reactions are performed alternately and are spatially separated in different protein domains. Thus, the growing DNA primer terminus has to undergo dynamic conformational switching between two distinct functional sites on the polymerase. Furthermore, the pol-to-exo mode transition must occur in the context of the replisome, wherein the polymerase is physically associated to its processivity factor PCNA. The mechanism of this conformational switching had remained obscure, largely due to the dynamic nature of the ternary Pol/PCNA/DNA assemblies. Here, we present computational models of the ternary assemblies for an archaeal polymerase, PolB. We combine all available structural information for the binary complexes, low-resolution negative stain EM data on the exo-mode PolB complex and refine atomistic models for the ternary PolB/PCNA/DNA complexes using extensive molecular dynamics simulations. In addition to the canonical PIP-box/IDCL interface of PolB with PCNA, contact analysis of the simulation trajectories revealed secondary binding interfaces, distinct between the pol and exo states. Using targeted molecular dynamics (TMD) we also explored the conformational transition from pol to exo-mode. In the TMD simulations we identified a flexible hinge region in the palm domain of PolB, critical for conformational switching. With the thumb domain anchored onto the PCNA surface, the neighboring palm domain executed a simple rotational motion around the hinge, bringing PolB down to form an additional interface with the PCNA

surface. A helix from PolB containing a patch of Arg residues was involved in the binding, locking the complex in the exo mode conformation. Together, these results provide a structural view of how the pol and exo states of PolB are coordinated through PCNA to achieve efficient proofreading.

5.2 Introduction

DNA polymerases catalyze the template-dependent incorporation of deoxynucleoside triphosphates (dNTPs) to extend the 3' end of a DNA primer strand. In addition to this polymerization activity many polymerases also possess 3'-5' exonuclease activity (proofreading), which allows the removal of misincorporated nucleotides from the nascent DNA (160,161). The fact that polymerization and proofreading are sequestered into spatially separated active sites within the polymerase suggests a switching mechanism must exist to transfer the primer terminus from one site to the other. The archaeal family-B DNA polymerases (PolB) share a common architecture comprised of five protein domains - fingers, palm, thumb, N-terminal, and exonuclease domain (24-26,162-164). The distance between the polymerase and exonuclease domains exceeds 20 Å in all available structures, requiring the substrate DNA to transition between the polymerase and exonuclease sites to ensure replication fidelity (165). This process is accompanied by substantial conformational changes of both the polymerase and DNA as evident in the crystal structures of PolB in replication and editing mode (163,164,166,167). While structures of PolB/DNA complexes in both pol and exo mode exist, details of the switching mechanism between the two modes have remained elusive. Moreover, during replication PolB functions as part of the replisome and must accomplish conformational switching while bound to multiple proteins, including proliferating cell nuclear antigen (PCNA, sliding clamp). PolB's association with PCNA is essential for full PolB activities (35,168). The trimeric PCNA protein encircles duplex

DNA as a closed ring and serves as a mobile platform for DNA polymerase to act on DNA, thereby, ensuring processive replication (38,168,169). PCNA-interacting proteins commonly include within their flexible tail regions a conserved sequence termed PCNA-interacting protein motif (PIP-box). This motif associates with a hydrophobic patch beneath the PCNA inter-domain connector loop (IDCL). The PIP-box/IDCL interaction confers the primary affinity for binding to the clamp (32,38,168). Consistent with this notion, disruption of the PolB PIP-box interaction results in substantial loss of enzymatic activity in both pol and exo mode. Since different PCNA partner proteins feature similar PIP motifs (which localize to the same sites on the clamp surface), the PIP-box interactions alone achieve only simple tethering to the clamp and are insufficient to ensure active orientation for core replication proteins (e.g. Pol). Emerging evidence has pointed to the critical role of secondary interfaces with the PCNA subunits to attain functional complexes. By locking the Pol enzyme in an optimal orientation these contacts regulate access to the substrate. Such secondary contacts have been previously identified for the ternary PolB/PCNA/DNA complexes (170,171) from low-resolution EM structural models. Since the contacts appear mutually exclusive in pol and exo modes, they have been proposed to modulate the conformational transition from polymerization to proofreading (171). In this contribution we provide atomically detailed models to more fully describe these secondary interfaces for the archaeal polymerase PolB from Pyrococcus furiosus (Pfu). We also delineate the large-scale conformational dynamics of the PolB/PCNA/DNA ternary complexes underlying the switch from polymerase to exonuclease function.

5.3 Results and Discussion

Structural models for the PolB/PCNA/DNA assemblies in Different Mode - Multinanosecond MD simulations were carried out to explore the conformational and structural dynamics of

the ternary Pfu PCNA/PolB/DNA complex in both pol-mode and exo-mode. The initial model of pol-mode was constructed based on PCNA/PolB and PolB/DNA binary structures from the Protein Data Bank (accession codes 3A2F and 4AIL) (164,170). The exo-mode was obtained by fitting the homology structure of Pfu PolB/DNA based on crystal structure of Pyrococcus abyssi polymerase (accession code 4FLV) (163) and the PCNA from PCNA/PolB binary structure into the corresponding ternary complex cryo-EM density map (EMDB accession code emd-5220) (171) using MDFF (172,173). The DNA in the active sites in each mode was preserved as what they were in the corresponding crystal structure; an extra dsDNA was added to each mode to extend the DNA so that it could pass through the PCNA ring.

The final refined model for either mode is the closest structure to the centroid of the major cluster generated by pairwise rmsd clustering analysis with corresponding MD trajectory. The sliding clamp in either model shows no structural discrepancies with the original crystal structure and no steric hindrance between proteins or DNA (Figure 5.1A, 5.1C). In our pol-mode model, PolB is anchored on the interface of PIP-IDCL with the core of the enzyme leaning toward the central hole of the PCNA ring, adopting an overall tilted position that partially obscures the central cavity of PCNA ring looking from the side of the IDCLs (cartoon representation, Figure 5.1A). This configuration is in line with the "tethered model" that was developed from the RB69 DNA polymerase/PCNA complex, in which the polymerase also fully covers the central hole of the clamp in either pol-mode or exo-mode (166,167). A resting ''locked-down'' configuration was observed in the *E.coli* translesion DNA polymerase (pol IV)/β-clamp complex, where the polymerase attached to the β clamp on the rim, taking no coverage over the central cavity (174). The polymerase in the latest "stand-by model" developed from the Pfu PCNA/PolB structure adopted an intermediate state between the active tethered state and the resting locked-down state

(170). In fact, the tethered model is the only one that allows for the substrate dsDNA to pass through the PCNA ring without steric hindrance, as evidenced by the initial pol-mode model (Figure 5.2B). Substantial kinking of the DNA duplex would have to be induced to accommodate the DNA in either the stand-by or locked-down models without adjusting the relative position of polymerase with respect to the sliding clamp, therefore, these models are likely not in an active mode. The functioning state of the complex with substrate DNA, on the other hand, would have to have the polymerase to associate with the DNA substrate in a more natural relaxed manner to gain the proccessivity, which was exactly what was observed in our MD simulation.

The single-particle EM has been utilized to investigate the Pfu PCNA/PolB/DNA assembly structure in exo-mode and produced a EM map, in which the PolB adopted a fully coverage over the PCNA central cavity as it is in the tethered model, forming a more extensive interface with the PCNA surface comparing with that in the pol-mode, primarily accomplished through the formation of the new interactions between the palm domain of PolB and the PCNA2 subunit surface (Figure 5.1C) (171). Functional studies have revealed a direct interaction between the arginine patch (R379R380L381R382 in PolB palm domain) and the E171 on the PCNA surface in the exo-mode. To integrate these crucial information into the atomistic structure, we applied MDFF to generate the initial exo-mode ternary complex structure, which is shown in cartoon in contrast to the original cryo-EM density envelope in gray surface at a conservative density threshold of 2.25 (Figure 5.3B). Upon the MDFF on PolB part, a substantial agreement in terms of the overall structure can be seen in the comparison. More importantly, the arginine patch on the palm domain of PolB was driven into the proper density region, in the vicinity of the PCNA E171 sidechain, with both R379 and R382 having a good chance to interact with E171 through electrostatic interaction (in this orientation, R380 was not likely to interact with E171). The final

structure generated by the subsequent MD simulation starting from the initial MDFF model does not drift away from the initial to any substantial extends. PolB maintained its fully coverage on the central cavity of PCNA and the extensive interface with the PCNA surface over the MD simulation, indicating a stable state of the MDFF-fitted model (Figure 5.1C).

For each mode, the flexibility of the complex was investigated by computing the B-factors for each residue, over the relevant trajectory, with the first 100 ns discarded as equilibration stage. Upon comparison of the B-factors for both modes, it is clear that the exo-mode complex adopts a more rigid structure than the pol-mode (Figure 5.1B, 5.1D). This disparity is also reflected in the rmsd profiles for both modes, wherein the pol-mode converges more slowly and shows a higher level of fluctuation overall (Figure 5.4). This is a functionally important detail of the PolB transition dynamic. In polymerase mode, PolB must move quickly and be more accommodating to the dynamics inherent in DNA, while in edit mode the complex may need to pause temporarily for proofreading and would need a more rigid structure to do so.

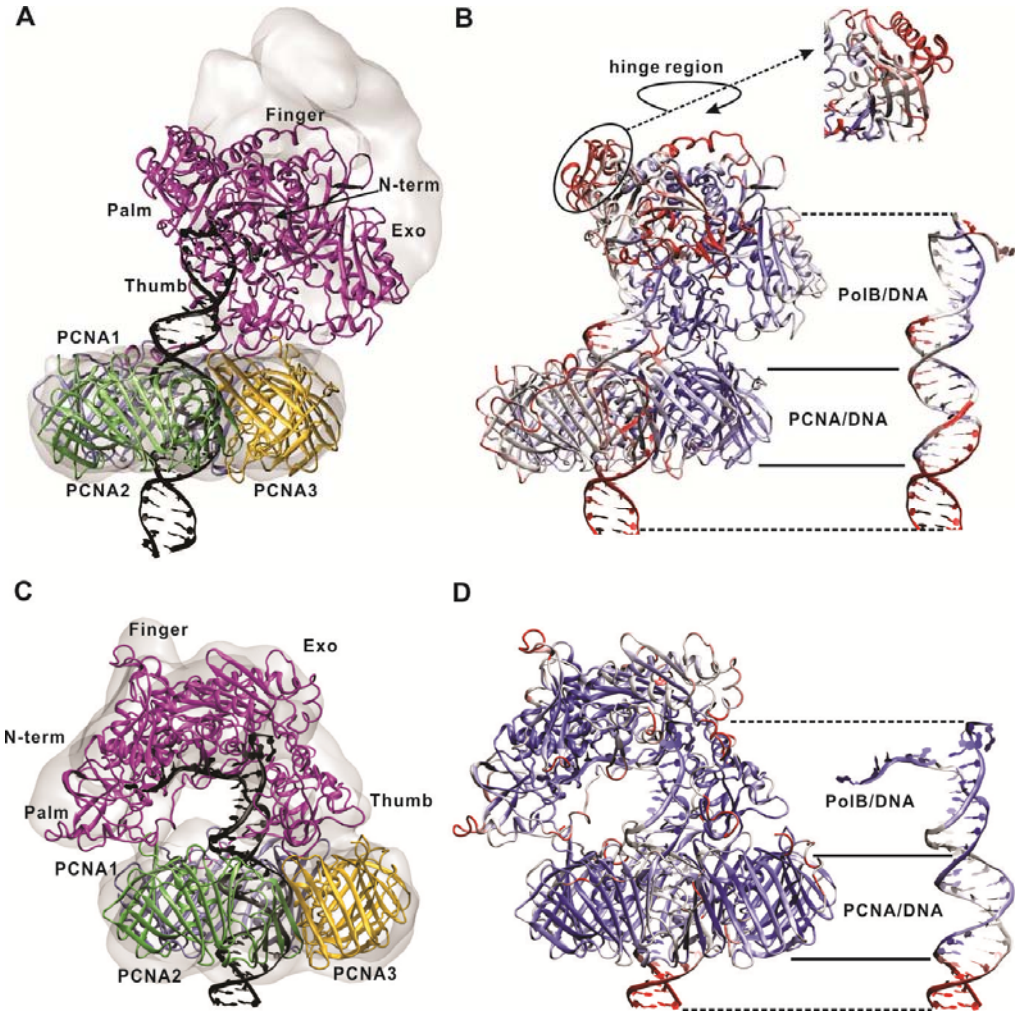

Figure 5.1 The closest structure to the centroid of the dominant cluster for the ternary Pfu PCNA/PolB/DNA complex from MD trajectory clustering analysis

A) pol-mode and C) exo-mode; complexes are shown in cartoon representations. The PCNA subunits are shown in different color: PCNA1 in ice blue, PCNA2 in lime, PCNA3 in gold, PolB in magenta, and DNA in black. The gray surface is the PCNA/PolB crystal structure (3A2F) in pol-mode, and the conformation obtained after the MDFF fitting in exo-mode, i.e., the starting conformation of the free MD for exo-mode; C and D) Computed B-factors mapped onto the ternary structures in each mode. Coloring corresponds to computed B-factor values from high (red) to low (blue). (Insets) Highlight of the B-factor difference for DNA parts.

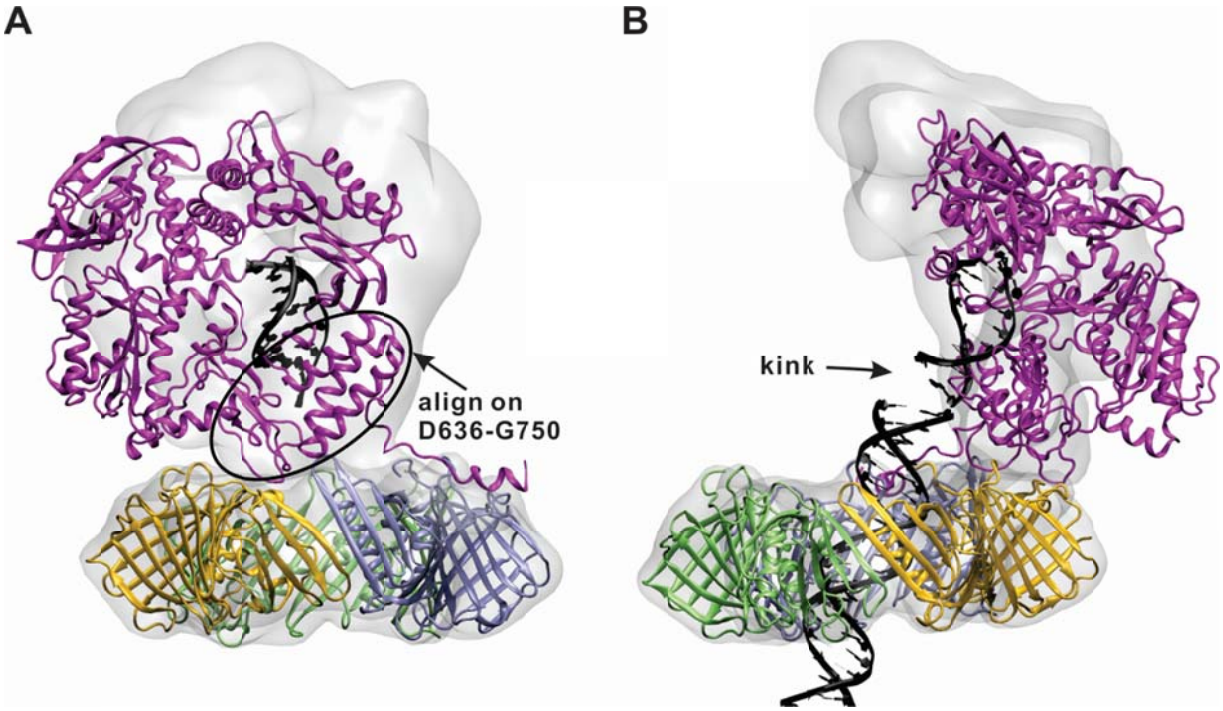

Figure 5.2 Initial setup of pol-mode with the dsDNA going through PCNA ring bearing a substantial kink.

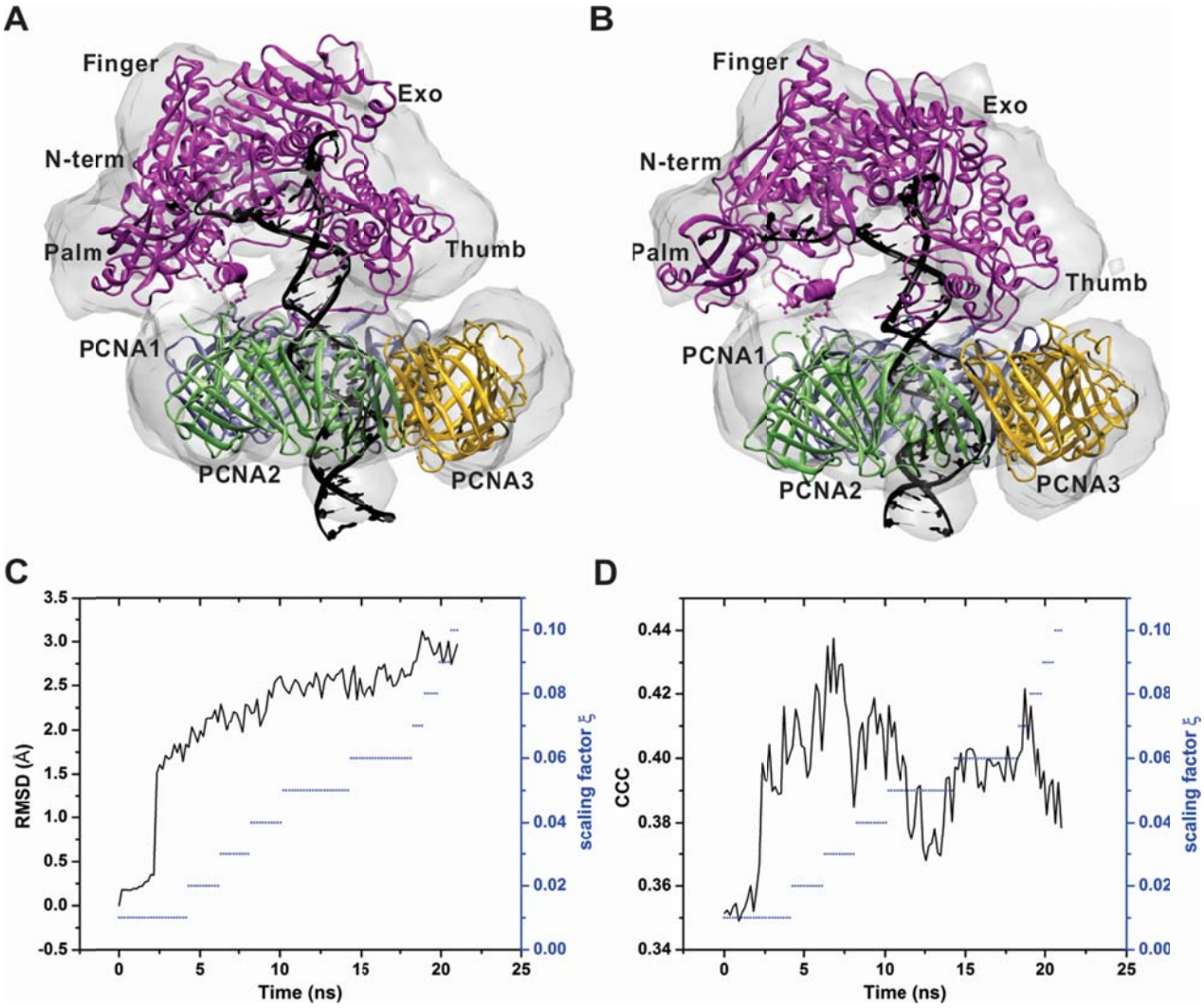

Figure 5.3 MDFF of exo-mode PCNA/PolB/DNA complex

A) Cartoon representation of the complex before fitted into the density map; B) Cartoon representation of the complex after fitted into the density map; C) RMSD of PolB over the MDFF simulation; D) CCC profile of PolB over the MDFF simulation.

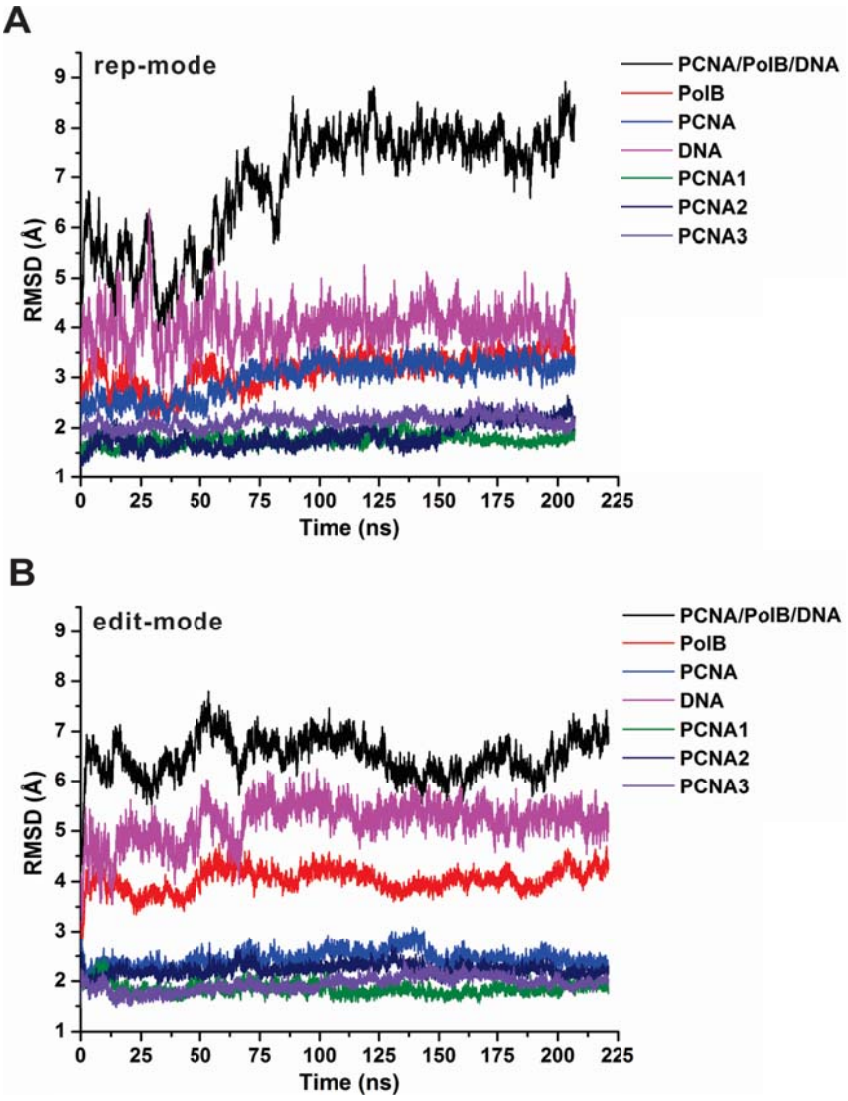

Figure 5.4 RMSD profiles of MD simulations

A) RMSD evolution of pol-mode assembly and each component from it; B) RMSD evolution of exo-mode assembly and each component from it.

DNA Is More Tightly Bound in the Editing Mode than in the Replicating Mode Complex - protein-DNA interactions in the pol- and exo- modes were found to be largely different (Figure 5.5). To explore the dynamic interaction of the DNA to PCNA or PolB, we sampled all the pos-

sible interactions between the sidechain of positively charged residues and oxygen atoms in the DNA backbone phosphate groups. The association of DNA with PCNA in either case tends to be asymmetric in terms of the number of the residues involved in PCNA is not evenly distributed among its three subunits. PCNA2 does not participate much in PCNA/DNA interaction in pol-mode (with the exception of residue K11) (Figure 5.5A, 5.5C), although it interacts strongly with DNA in the exo-mode (Figure 5.5B, 5.5C). This can be understood by envisioning the pol-mode to exo-mode transition, which would naturally force the DNA from one side of the PCNA pore to the other. The DNA bending showing by the DNA axis calculated with the Curvers+ program (175) reflects the relative positioning of DNA with respect to PCNA (Figure 5.5C). In addition, slight geometric parameter discrepancies of the DNA can be observed in the exo-mode: the spanning difference between major and minor grooves become trivial due to the compressing of the duplex part right above the PCNA ring induced by the transition.

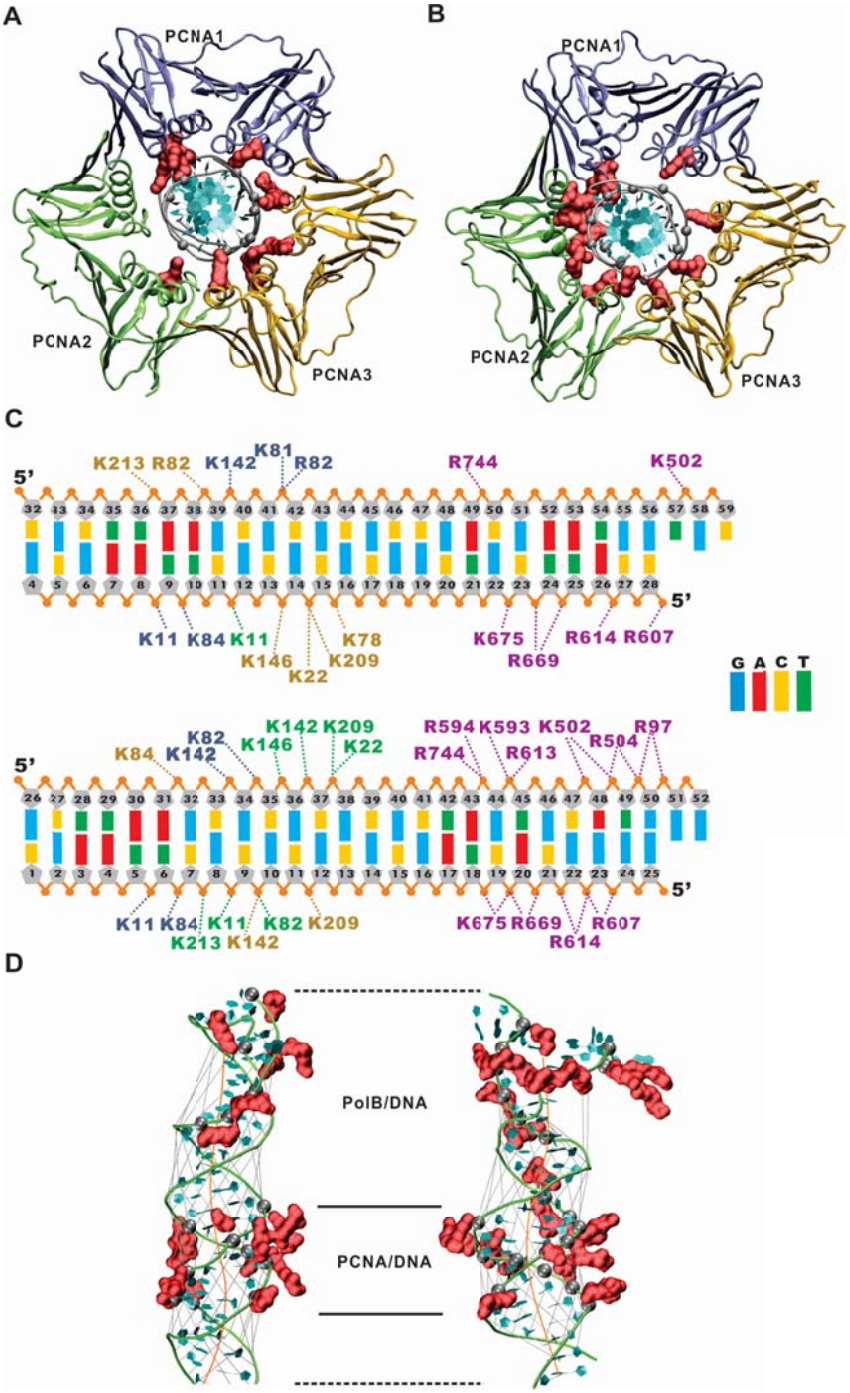

Figure 5.5 Pfu PCNA/PolB/DNA ternary complex in different modes associate the substrate DNA duplex in distinct binding modes

A and B) Cartoon representations of PCNA binds to dsDNA in pol-mode and exo-mode respectively. The PCNA subunits are colored differently: PCNA1 in ice blue, PCNA2 in lime,

PCNA3 in gold. The dsDNA phosphodiester groups and the side-chain heavy atoms of basic residues on the inner surface of PCNA involved in persistent contacts (observed in more than 50% of the frames in the MD trajectories) are shown in gray spheres and red surfaces respectively; C) All persistent DNA-protein contacts are listed explicitly for rep pol-mode (upper panel) and exo-mode (lower-panel) respectively. The residue label texts are colored so they are in consistence with corresponding subunit color; D) Averaged structure of the DNA duplex from pol-mode (left), and exo-mode (right) simulation. The DNA is shown in ribbons representation. The DNA axis and the widths of major or minor grooves were computed with the program Curves+ and are shown in orange and gray lines respectively.

Moreover, it is intriguing to notice that the PolB-DNA association in the exo-mode is stronger that the pol-mode with respect to the number of persistent contacts of PolB-DNA (7 vs. 18) (Figure 5.5D). The greater number of the contacts in exo-mode makes it possible that the DNA duplex is consistently tracked after it leaves the PCNA cavity (Figure 5.5D). Alternately, the two discrete contact groups evident in the pol-mode yield a discontinuity in the protein-DNA association (Figure 5.5C). The difference in the coupling strength is likely rooted in the functional differences between the two modes: a less extensive coupling might favor the fast process of the replication machinery while a stronger coupling is preferable for the complex to pause and fix the errors. Therefore, a correlation can be drawn between the numbers of protein-DNA contacts, the rigidity of the complex and the function PolB is performing (editing or replication).

The PolB Orientation in either Mode is determined by Specific Secondary Interactions with PCNA - It has been noted that the primary PIP-IDCL interactions between functional enzymes and sliding clamps do not confer much beyond an anchor point. The secondary contacts

that are off the primary sites are actually the ones determining the functional orientation of the enzymes, which has been seen in many cases including the FEN1/PCNA, Ligase/PCNA and RNA pol/PCNA systems (143,176). Our computed model also revealed many other direct interaction sites between PolB and PCNA surface other than the canonical PIP-IDCL interaction in either mode. We used geometric criteria to determine the interactions (H-bond, salt-bridge, hydrophobic interaction) throughout the production phase of the MD trajectories, considering interactions that occur < 50% of the trajectory to be transient. Therefore, all contacts shown are considered stable and persistent for each model (Figure 5.6). The long C-terminal tail of PolB makes stable contacts with the IDCL on PCNA1 subunit through a hydrophobic cluster, generally involving the same group of hydrophobic residues in both modes (Figure 5.6B, 5.6D). While there are only two H-bonds between PolB and PCNA1, both near the hydrophobic pocket in the pol-mode (Figure 5.6A), a stronger H-bond and salt-bridge network contributes substantially to the relatively lockedup orientation of the C-terminal tail of the exo-mode PolB (Figure 5.6C). In the pol-mode, the major interface other than the PIP-IDCL site is the formed by electrostatic complementarity between the thumb domain of PolB and the PCNA (Figure 5.7A). Four persistent salt-bridges and one H-bond are formed on this interface, such as K686-D42 (PCNA3), K724-E119 (PCNA3), K725-D117 (PCNA3), R706-E171 (PCNA1), and R706-E26 (PCNA3) (Figure 5.6A), whereas the R706-E171 was retrieved with 90.8% occupancy over the trajectory and was previously identified to be critical to the polymerase activity in the complex (170). In addition, a hydrophobic cluster of 3 residues further strengthens the stability of this interface (Figure 5.6B). Although the PIP-IDCL interaction is invariant, the salt-bridge network and the hydrophobic cluster are largely disappeared in the exo-mode, leaving a much smaller hydrophobic pocket, two salt-bridges and one H-bond, such as the K690-E249 (PCNA3), R706-D25 (PCNA3), and R706-

S21 (PCNA3), on the slightly rearranged interface between thumb domain of PolB and PCNA (Figure 5.6C, 5.6D, 5.7B). The difference of the PolB-PCNA interface in exo-mode from pol-mode is mostly introduced by the formation of the interaction between the palm domain of PolB and PCNA2. Notably, the salt-bridge formed by R382 from the arginine patch in the palm domain and E171 from PCNA2 is consistent with the previously reported importance of the arginine patch in maintaining the full exonuclease activity of PolB in ternary complex (171). The electrostatic surface potential profiles of the protein components show that the E171 of PCNA2 fits nicely into the groove of the arginine patch on PolB (Figure 5.7B). Interestingly, the R379 that was found to be critical did not reveal itself as contributing to any substantial level to this interface in our simulation. Considering the heterogeneity of the ternary complex EM images, and the relatively minor degradation of exonuclease activity (20%) of the ternary complex with complete charge reversion mutant R379E, it might have overshot the contribution of a single residue to the stability of this interface by highlighting R379 only. Without R380 or R382 point mutant, it's hard to elucidate the importance of a singular arginine within the arginine patch. Regardless of the discrepancy, it is more reasonable to accredit the entire arginine patch rather than any single arginine given the multiple mutations greatly decreased the exonuclease activity (171).

Given the persistency of these secondary interactions, it's highly possibly that they play important roles in maintaining the proper distinct fully functional orientation of the PolB in different mode. While the direct available experimental evidence only support one single pair interaction for each mode, it's very likely that the single pair interaction is in cooperation with other interaction network to lock a complex of this size in a particular orientation. In addition, the residues involved in the identified contacts mentioned above are largely conservative cross many species as revealed by the multi-sequence alignment (Figure S5.1), therefore, it is worthwhile to

conduct further mutagenesis studies on the other residues participated in forming the stable contacts in the simulations.

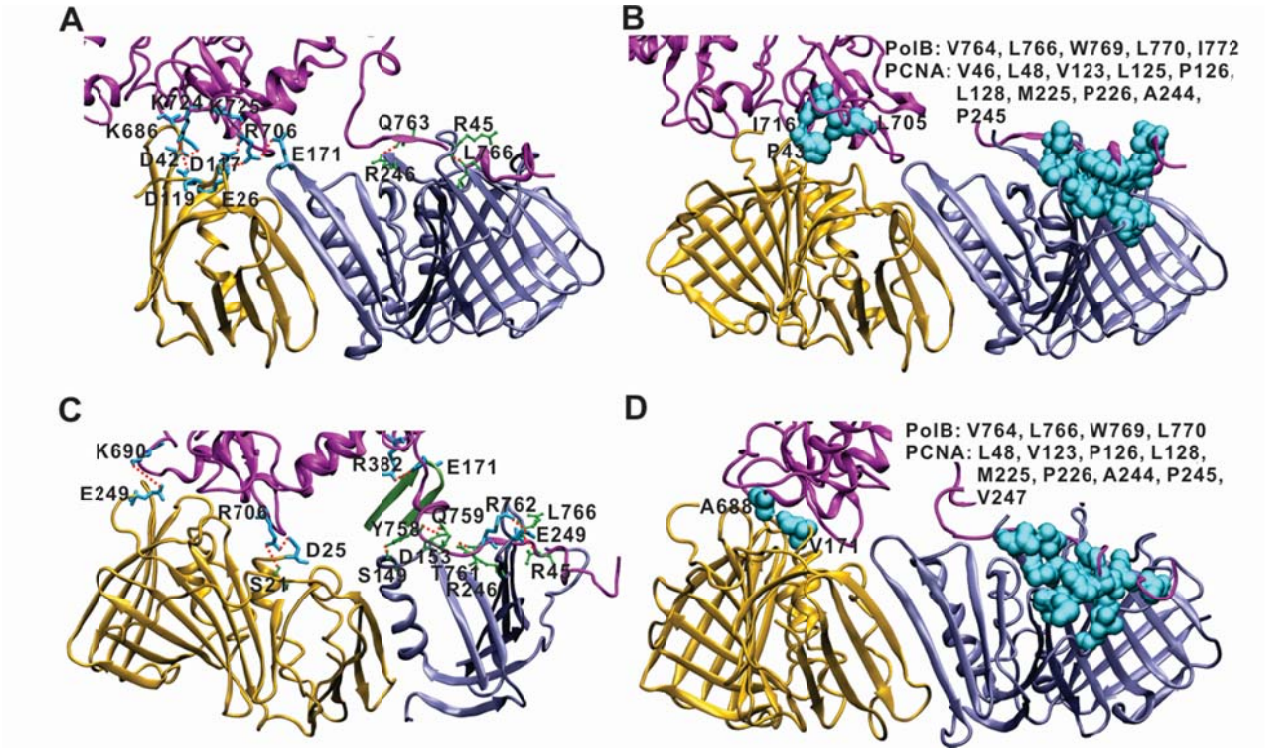

Figure 5.6 Distinct secondary contacts between PolB and PCNA determines the productive orientation of PolB with respect to PCNA in different mode

A and C) H-bonds and salt-bridges formed between PolB and PCNA are shown as red dash lines, with the residues involved in H-bonds colored in green (stick and ball representation) and residues involved in salt-bridges in cyan (bond representation); B and D) Hydrophobic clusters formed by PolB and PCNA in pol-mode and exo-mode. The hydrophobic residues involved are colored in cyan (ball representation).

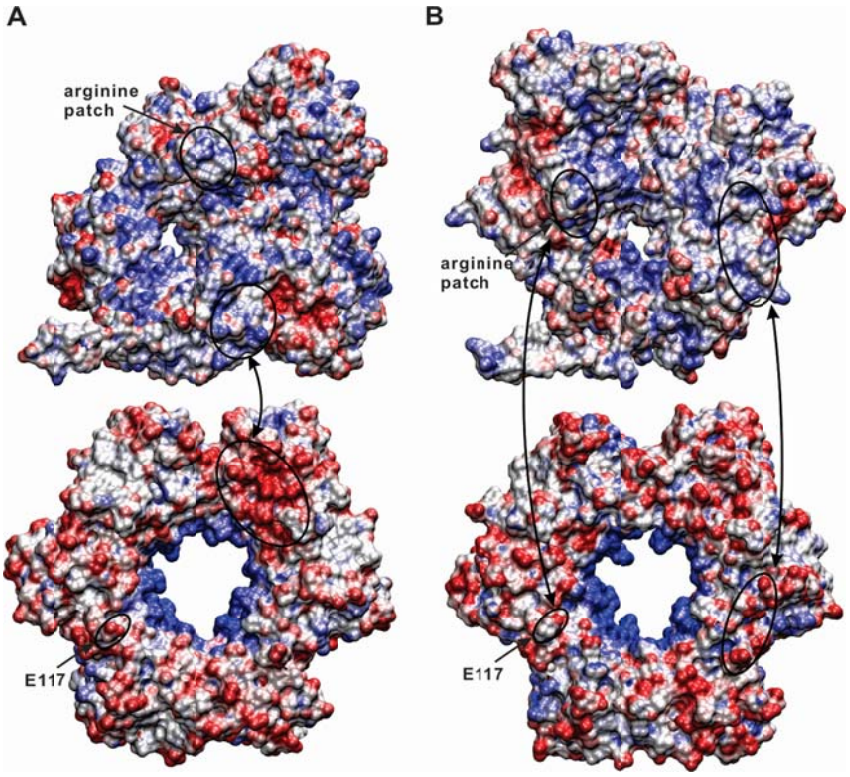

Figure 5.7 Surface electrostatic potential of the protein components in PCNA in the tenary complexes

A) Electrostatic complementarity between interacting regions on PCNA and PolB from pol-mode assembly; B) Electrostatic complementarity between interacting regions on PCNA and PolB from exo-mode assembly.

Transition from Replicating to Editing Mode Majorly Involves Internal Pivoting Motion of PolB - PCA has been used as a versatile tool to analyze the low-frequency motions of macromolecule dynamics as it reduces the dimensionality of the MD trajectory. The first few principal components are of special attention since they capture the important dynamic features of the system, which are often the slow large-amplitude motions. In our PCA results, the first two eigenvalues accounts for 48% and 39% for the pol-mode and exo-mode MD trajectory respectively.

To highlight the relative motions observed in PolB, we aligned the trajectory (projected onto corresponding principal component) based on PCNA and color-coded the most mobile regions of PolB by physical time. Motion in the first component of pol-mode reveals a horizontal twisting motion of PolB with respect to the PCNA ring planar (Figure 5.8A), and the second component motion indicates a vertical swing motion of the PolB toward the PCNA ring (Figure 5.8B). A slanted twisting motion, which can be viewed as the combination of the horizontal twisting and vertical swing is revealed by the first principal component motion of the exo-mode (Figure 5.8C). Also present is the vertical swing captured by the second component of the exo-mode (Figure 5.8D). The magnitude of motions from the exo-mode is uniformly smaller than that from the pol-mode due to the overall structural flexibility difference of them (Figure 5.1B, 5.1D; Figure 5.4).

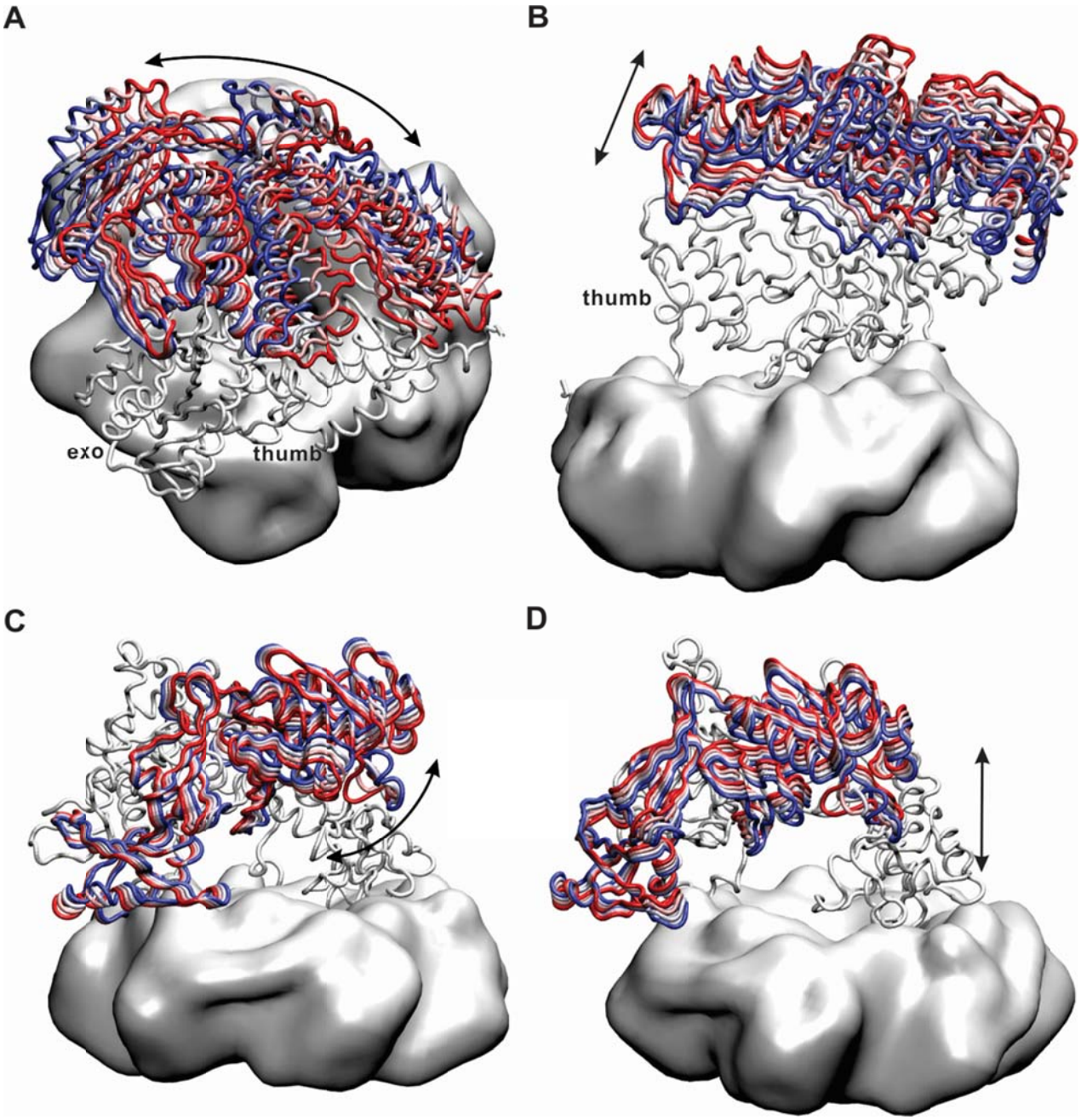

Figure 5.8 Large-scale motions of PolB with respect to PCNA revealed by PCA

A) The horizontal twisting motion is captured as the most significant global motion by projecting the pol-mode MD trajectory onto its first principal component; B) The vertical swing motion is captured as the second significant global motion by projecting the pol-mode MD tra-

jectory onto its second principal component; C) A slanted twisting motion is captured as the most significant global motion by projecting the exo-mode MD trajectory onto its first principal component; D) The vertical swing motion is captured as the second significant global motion by projecting the exo-mode MD trajectory onto its second principal component. The structures were all aligned on PCNA to highlight the relative motion of PolB with respect to PCNA, and the progression of the transition is color coded on PolB (excluding the less mobile thumb domain) from red to blue (or from blue to red).

Similar to the first few principal components from PCA, the low-frequency modes from normal mode analysis (NMA) have been shown related to large-scale motion of macromolecules (177,178). To explore the global motions from a different perspective we conducted the NMA for both ternary complex models using the PATH-ENM server (179). In the pol-mode, the slanted twisting motion corresponding to the lowest frequency normal mode (Figure 5.9A) and the vertical swing motion to the second lowest frequency normal mode (Figure 5.9B) are consistent with the principal components from MD (Figure 5.8A, 5.8B). As for the exo-mode, the horizontal twisting and vertical swing in the two lowest frequencies normal modes for exo-mode (Figure 5.9C, 5.9D) indicate a considerable correlation with corresponding PCA results (Figure 5.8C, 5.8D). This consistency is in line with the observation that the low-frequency normal modes and the first few principal components obtained from MD simulations have shown clear similarity (180,181).

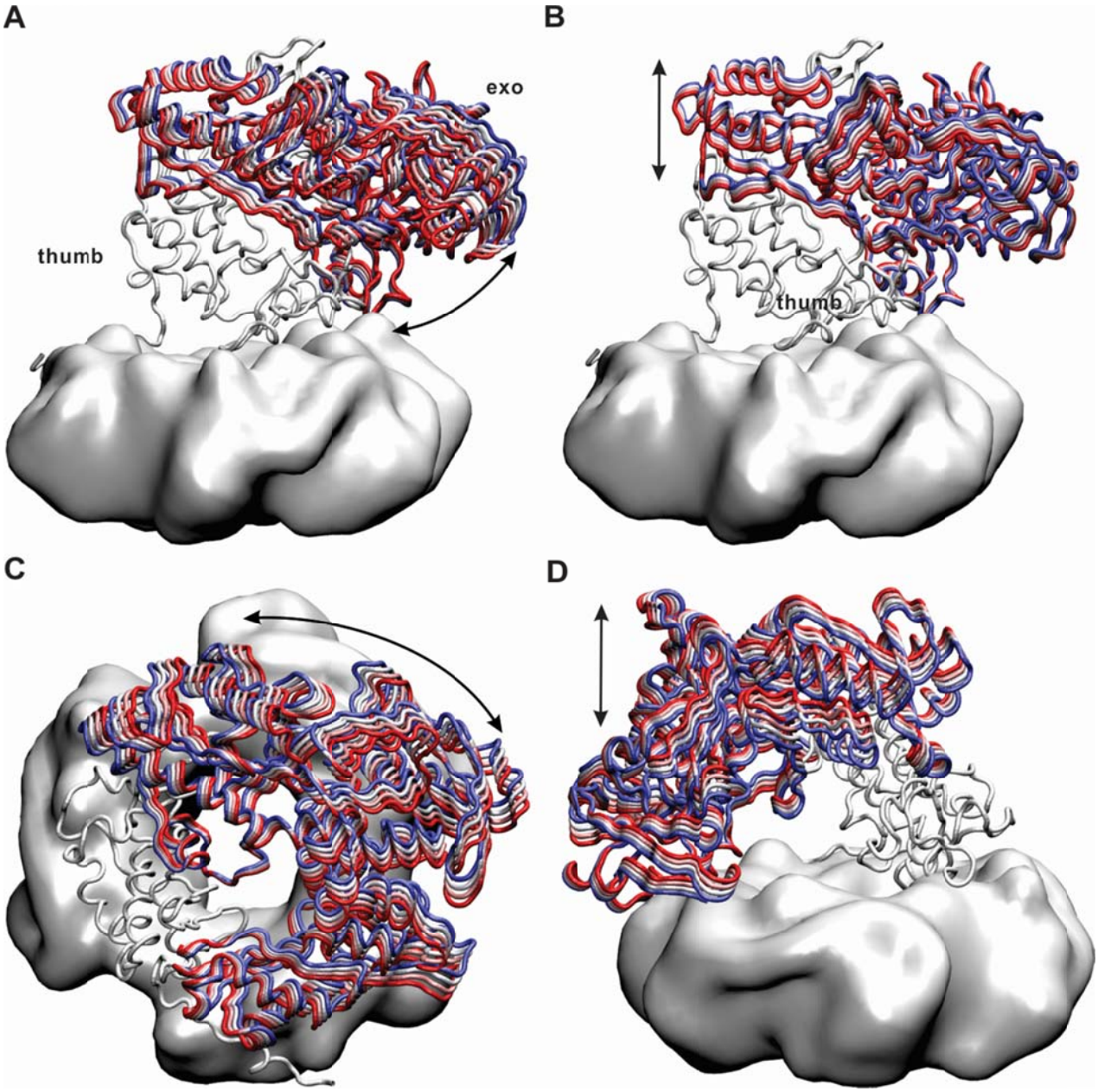

Figure 5.9 Lowest-frequency motions of PolB with respect to PCNA revealed by NMA

A) The lowest frequency mode in pol-mode model is a slanted twisting motion; B) The second lowest frequency in pol-mode model mode is the vertical swing motion; C) The lowest frequency mode in exo-mode model is the horizontal twisting motion; D) The second lowest frequency mode in exo-mode model is the vertical swing motion. The structures were all aligned on

PCNA to highlight the relative motion of PolB with respect to PCNA, and the progression of the transition is color coded on PolB (excluding the less mobile thumb domain) from red to blue (or from blue to red).

The transition between pol-mode and exo-mode of polymerase happens on the millisecond timescale (182). Therefore, it's difficult to sample the transitioning dynamics with free MD with a system of the ternary complex size (~240000 atoms). To overcome the timescale limitation, the TMD method was used to accelerate the transition, wherein a rmsd based biased potential was introduced to drive the instant state of the system toward the targeted final state. Although TMD does not guarantee a minimum free-energy pathway and might produce unphysical intermediate states when was conducted using large force constant, it's rather computationally efficient in exploring qualitative characteristics of large-scale conformational changes with proper consideration given to the force constant applied (183-185). In this effort, TMD was used to probe the dynamics of the Pfu PCNA/PolB/DNA over the transition from pol-mode to exo-mode. A profile of the rsmd on all the backbone atoms of the ternary complex over the TMD simulation indicates a smooth transition without jumps or discontinuities (Figure 5.10A). Upon a visual inspection on the TMD trajectory, it becomes noticeable that besides the compression and separation of the DNA duplex (rmsd variation is 5.66 ± 2.84 Å, Figure 5.10B), the internal motion is mostly contributed by the conformational change in PolB, which is evident by the rmsd profiles of different components computed following alignment to the initial conformation of each of them (Figure 5.10B). The PCNA has a rmsd of 1.99 ± 0.98 Å, and the rmsd of PCNA with the thumb domain of PolB is slightly larger as of 3.28 ± 1.61 Å. This is consistence with the observation that the thumb domain maintains its relative positioning with respect to PCNA throughout

the transition. As a result, it is the motions of the N-terminal, exonuclease and palm domain all together with respect to the thumb domain that reflects the internal motion of PolB. The relatively quasi-rigid, concerted motion of these three domains is also reflected by its small rmsd variation of 2.06 ± 1.13 Å over the transition. Therefore, the overall internal motion of the ternary complex can be captured by the relative motion of the three domains with respect to the other protein components, which is a simple pivoting of them on the hinge region locates at the interface of thumb and palm domain (Figure 5.11A). This pivoting motion can be decomposed into two components: one is the clockwise twisting in the plane parallel to PCNA ring viewing from above PolB; the other is the vertical swing toward the PCNA ring until the arginine patch in the palm domain is in contact with PCNA surface (Figure 5.11A, 5.11C). The substrate DNA close to the polymerase active site does not separate until late stage of the TMD simulation (~16 ns of the 21 ns simulation), and the 3' end template DNA continuously tracks with the long loop region of the palm domain (Figure 5.11D). On the thumb domain end, the contact of it with PCNA surface rearranges to accommodate the final orientation of PolB in the exo-mode (Fig. 5.11B). This is also reflected by the distinct group of contacts shown on this interface in either mode (Fig. 3).

It can be easily identified that horizontal twisting and vertical swing components of the simple twisting observed in the TMD transition (Figure 5.11A) perfectly correlates with the first and second principal components of the MD trajectory from either rep- or exo-mode complex (Figure 5.8A). A fairly similar strong correlation pattern holds when we compare the two components of the twisting motion to the low frequency motions revealed by NMA. In another word, the pivoting motion observed in TMD can be reconstituted with the first two principal components from PCA or the lowest frequency vibrations of the complex. Therefore, this simple pivot-

ing motion is likely to be a fundamental mechanical characteristic of the transition process given both PCA and NMA results are in support with it.

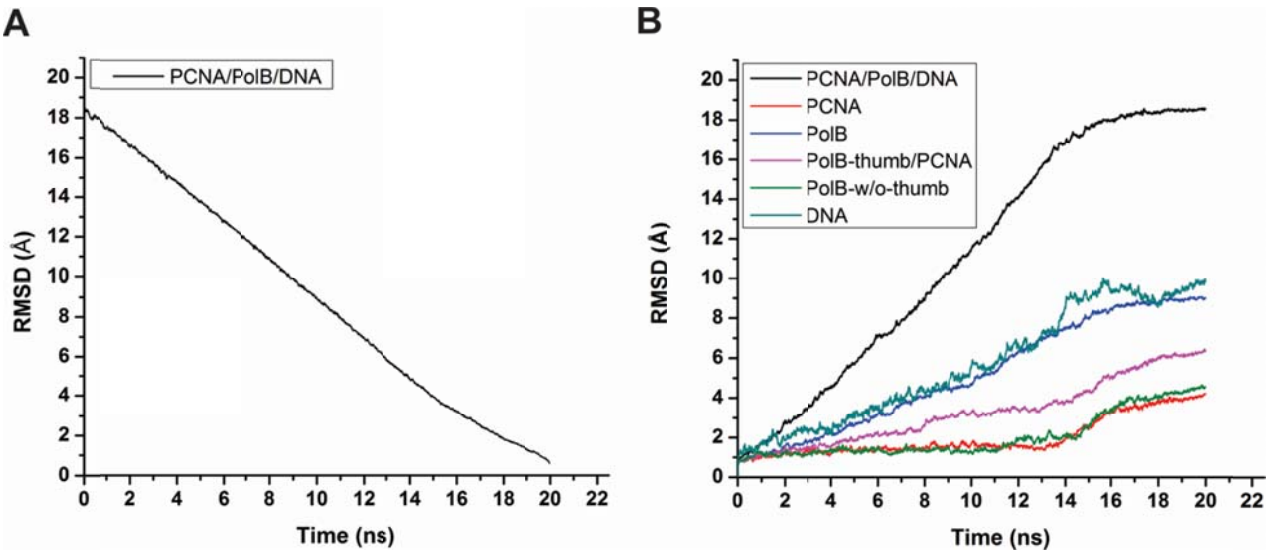

Figure 5.10 RMSD profiles of TMD simulation

A) backbone RMSD of PCNA/PolB/DNA aligning to the target structure (exo-mode); B) backbone RMSD of different components aligning to the corresponding component in the initial structure (pol-mode).

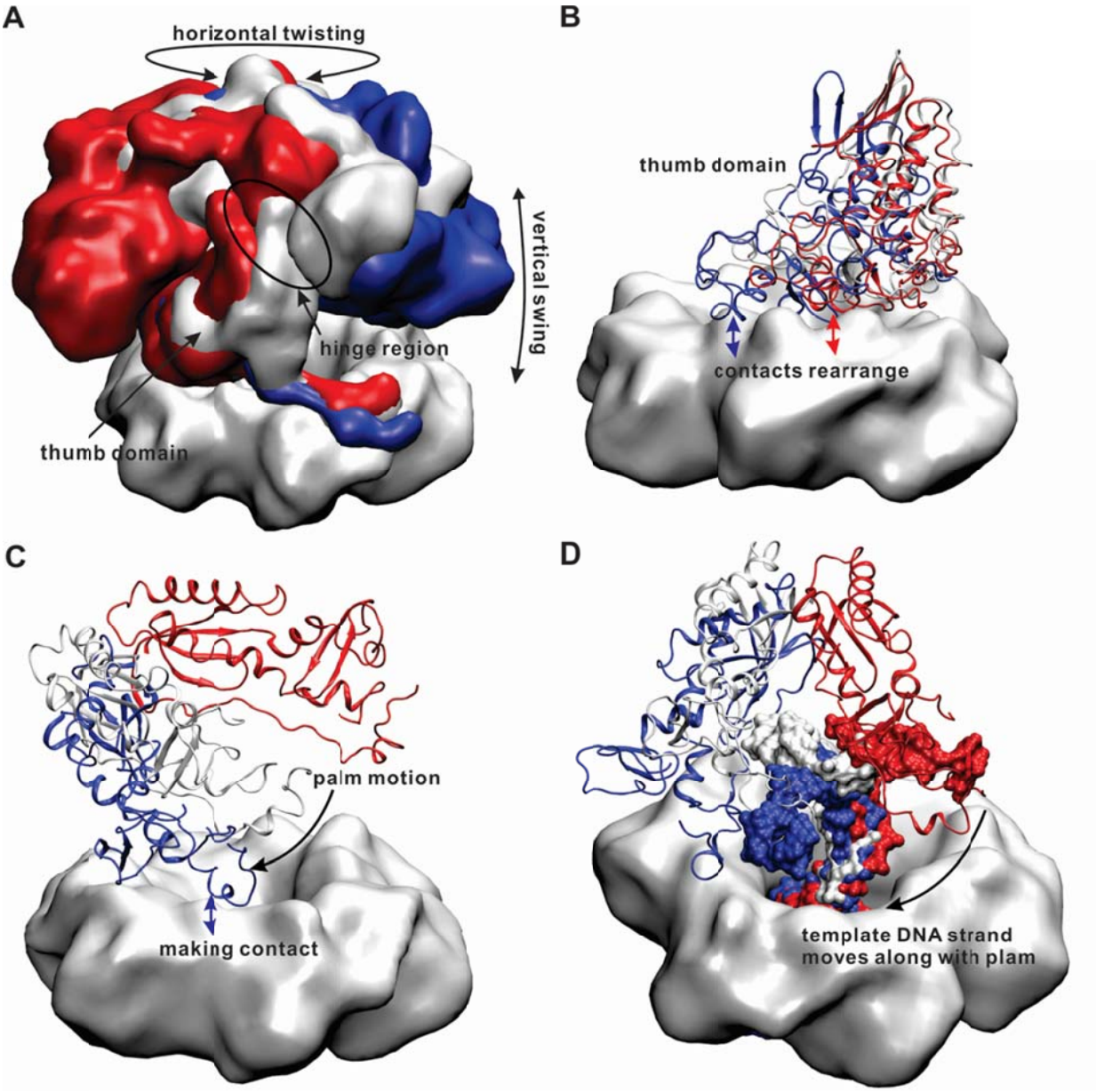

Figure 5.11 Collective motion of PolB with respect to PCNA captured in TMD with the trajectory aligned on pol-mode PCNA

A) Surface representation of the complex without DNA showing; The pol-mode to exo-mode transition is captured as a simple pivoting of the PolB on the hinge region (the interface of thumb and palm domains) with the thumb domain and PCNA remains largely still. The pivoting composes the horizontal twisting and vertical swing (indicated by double-arrow arcs); B) Polb

thumb domain (cartoon representation) detaches and reattaches to the PCNA surface over the transition, with double arrow arcs showing the different contact sites; C) View of the transition from the palm domain side with palm domain showing in cartoon representation and other domains of PolB omitted. The overall motion is reflected by the swing of the palm from distant with PCNA2 to interacting with its surface; D) The single stranded part of the template DNA strand (surface representation) tracks along the palm domain (cartoon representation) as it swings. The progression of the transition is color coded from red to blue in all 4 panels.

5.4 Conclusion

The motion in the TMD can be recapitulated with a rather simple pivoting motion of PolB on the interface between the thumb and palm domain, herein referred to as the hinge region. Projection of the free MD trajectories onto the first and second principal components obtained by principal component analysis (PCA) revealed the large scale conformational change in either mode is consistent with the simple pivoting that was captured in TMD. In addition, the lowest and second lowest frequency normal modes generated by the elastic network model (ENM) (179) reproduce the pivoting motion obtained by TMD, correlating well with the PCA results. All together, these suggest a fundamental dynamic feature of the transition between pol-mode and exo-mode of DNA polymerases.

The transition between the pol-mode and exo-mode is important for polymerases possess exonuclease activity to function properly. With distinct structural features of PolB in different mode being appreciated by the crystal structures and cryo-EM studies, still very little is known about the switching mechanism. The results from our computational approach have provided insights into the Pfu PCNA/PolB/DNA ternary complex by obtaining the thorough functional related atomistic level interaction information between different components in the complex in-

cluding the protein-protein and protein-DNA interactions. Specific interactions in different mode modulate the complex in its corresponding functional orientation and probably determine the free energy profile and hence the transition rate between the modes. In short, the most detailed and complete picture of the PolB transition between pol-mode and exo-mode to date was captured in the context of ternary complex. The pivoting motion of PolB on its hinge region was found to be fundamental to this transition, therefore, shed lights on the switching mechanism of DNA polymerases, and serve as a springboard for future experimental lines of evidence to further illustrate the biological significance of this unique process.

5.5 Materials and Methods

Model construction and MD simulation - The replicating ternary complex was constructed using two structures: the PCNA/PolB complex (PDB ID code 3A2F) (170) and the PolB/DNA structure (PDB ID code 4AIL) (164). In the PCNA/PolB binary structure, PolB adopts an upright position of with respect to the PCNA ring (Fig. 1A). An initial overlay of the PolB/DNA structure onto the PCNA/PolB structure based on an alignment over all PolB backbone atoms reveals a discrepancy in the conformation of PolB, whereas PolB adopts an overall more compact conformation in the PolB/DNA structure than in the PCNA/PolB structure due to it interacts with the substrate DNA. Therefore, an overlay across entire PolB would be sub-optimal in maintaining the original contacts formed by the PolB and PCNA in the PCNA/PolB binary structure. To circumvent this problem, the two binary structures were superimposed based on the part of the thumb domain only (D636 to G750) so that the thumb domain in PolB/DNA structure would have a better chance to form contact with the PCNA from the PCNA/PolB structure (Figure 5.2A). With the initial relative positions of the PolB and PCNA been determined, an upstream dsDNA (Figure 5.12) extension would only be allowed to pass through the PCNA ring with a

substantial kink. To relax the DNA duplex in the ternary complex, the system was equilibrated with the well-defined α-helices and β-sheets protein constrained for 4 ns before replacing the corresponding parts with the original crystal structures to maintain the crystal contacts. In addition, a short TMD based on the RMSD of the sidechains of a small group of residues (R706, K724, K725 from PolB, E171 from PCNA1, D96, L99 from PCNA3, the 3 subunits of PCNA are numbered counterclockwise from the top view of PolB side as from PCNA1 to PCNA3 in this study) involved in the aforementioned thumb-PCNA contacts in PCNA/PolB crystal structure were then applied to maximally restore the contacts before the whole system was refined through the ~200 ns free MD.

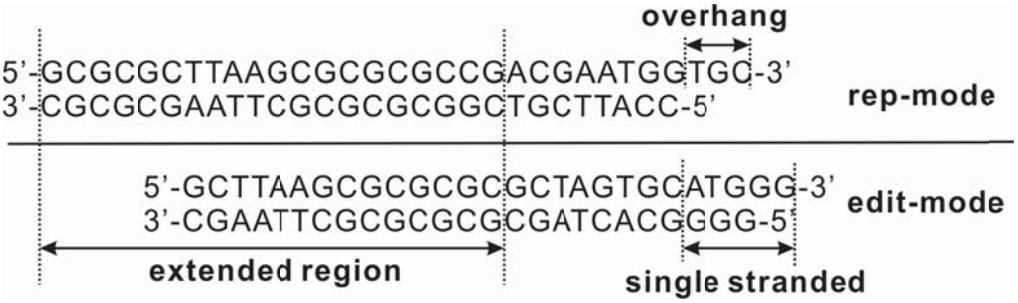

Figure 5.12 DNA Sequences in the complexes

Whereas the pol-mode Pfu PolB/DNA structure was available, homology modeling based on the exo-mode *Pyrococcus abyssi* PolB/DNA structure (PDB ID code 4FLV, sharing 83.1% sequence identity with Pfu PolB) (163) was applied to obtain the exo-mode Pfu PolB/DNA structure. The homology model of PolB/DNA and PCNA from PCNA/PolB crystal structure were then separately fitted into the same experimental EM density map with rigid-body fitting. Similarly, a strand of dsDNA was introduced to extend the substrate DNA through PCNA ring, while

in this case, without any substantial clashes with the PCNA ring. 21 ns MDFF was ensued to fit the PolB into the corresponding cryo-EM map density in order to recover the extensive PCNA/PolB interface in this mode. In MDFF, external biased forces proportional to the EM density gradient are applied to drive the atoms into the high-density regions of the EM density map. The DNA substrate was excluded from the MDFF fitting procedure due to the incomplete density of DNA. The PCNA was also not included since the PCNA interface would have to be ruptured to fit into the density map of PCNA part. After the MDFF, the system was also refined by a ~200 ns free MD.

Production runs were carried out in the NPT ensemble (1 atm and 300 K) for ~200 ns for the each complex. Long-range electrostatic interactions were evaluated with the smooth-particle mesh Ewald algorithm (186). The short-range nonbonded interactions were calculated used a cutoff of 10 Å with a switching function at 8.5 Å. The integration time step was 2 fs and the SHAKE algorithm was applied to fix the bonds between the hydrogens and the heavy atoms. The r-RESPA multiple time-step method (118) was utilized with a 2 fs time step for bonded and short-range nonbonded interactions, and 4 fs for long-range electrostatic interactions. All simulations were performed using the NAMD 2.9 code (73,187) with the AMBER Parm99SB parameter set (188) containing the basic force field for nucleic acids and proteins, as well as the refined parameters for backbone dihedrals for protein (SB) and nucleic acids dihedrals (BSC0) on Edison, a Cray XE6 system at the National Energy Research Scientific Computing Center. Data were analyzed using the PTRAJ utility in AMBER and custom VMD TCL scripts. DNA structural parameters were analyzed with the program CURVES + (175).

Targeted Molecular Dynamics - To accelerate the transition from pol-mode to exo-mode of the ternary complex, a TMD simulation was performed in the NPT ensemble by applying the following time-dependent, harmonic energy bias to all the backbone atoms in the system.

$$U_{TMD} = \frac{1}{2} Nk(rmsd - \rho(t))^2 \qquad (5.1)$$

where N is the number of atoms subjected to the bias, k is the harmonic force constant, and RMSD is the root-mean-square deviation between a configuration at time point t and the target configuration. When the reference RMSD value $\rho(t)$ at time t is monotonically decreasing, the initial structure is gradually driven toward the target structure. Different k vaues of 200, 400, 600 and 800 kcal mol^{-1} Å^{-2} were tested with 2 ns for each to determine an optimal balance between accuracy and efficiency. The 600 kcal mol^{-1} Å^{-2} was then chosen for the 20 ns production run by the standard of minimal value, which did not compromise any secondary structures in the model. The initial structure of the TMD was chosen to be the closest conformation to the centroid of the dominant cluster from the second half of pol-mode MD simulation trajectory. To eliminate the difference in the DNA length, the upstream DNA in the pol-mode was cut by 3 base pairs to match that in the exo-mode; and the system was then equilibrated with a 4 ns free MD before the TMD was introduced. Similarly, the target structure is the closest conformation to the centroid of the dominant cluster from the second half of the exo-mode MD simulation trajectory.

Molecular Dynamics Flexible Fitting - The initial complex of the exo-mode was obtained by flexible fitting into the EM map using the molecular dynamics flexible fitting (MDFF) method (172,173). A preliminary rigid body fitting was performed using Chimera, a program from University of California San Francisco. MDFF involves a MD simulation in which external forces proportional to the EM density gradient are applied, biasing the motion of the protein atoms

toward the high density regions of the map (PCNA and DNA was excluded from the MDFF fitting procedure). To prevent secondary structural distortions, bond, angle and dihedral angle restraints were imposed on the well-defined secondary structure regions defined by DSSP (189). The MDFF was applied in several stages with progressively higher values of the MDFF force scaling factor ξ, which was chosen to vary from 0.01 to 0.1. The convergence of fitting was then assessed by monitoring the RMSD and the cross-correlation coefficient (CCC) of PolB between the atomic model and the original EM density map (Figure 5.3).

CHAPTER 6. REPAIR COMPLEXES OF FEN1 ENDONUCLEASE, DNA, AND RAD9-HUS1-RAD1 ARE DISDINGUISED FROM THEIR PCNA COUNTERPARTS BY FUNCTIONALLY IMPORTANT STABILITY

6.1 Abstract

Processivity clamps such as proliferating cell nuclear antigen (PCNA) and the checkpoint sliding clamp Rad9/Rad1/Hus1 (9-1-1) act as versatile scaffolds in the coordinated recruitment of proteins involved in DNA replication, cell-cycle control and DNA repair. Association and handoff of DNA-editing enzymes such as flap endonuclease 1 (FEN1) with sliding clamps are key processes in biology, which are incompletely understood from a mechanistic point of view. We have used an integrative computational and experimental approach to define the assemblies of FEN1 with double-flap DNA substrates and either PCNA or the checkpoint sliding clamp 9-1-1. Fully atomistic models of these two ternary complexes were developed and refined through extensive MD simulations to expose their conformational dynamics. Clustering analysis revealed the most dominant conformations accessible to the complexes. The cluster centroids were subsequently used in conjunction with single particle electron microscopy (EM) data to obtain a 3-dimensional EM reconstruction of the human 9-1-1/FEN1/DNA assembly at 18 Å resolution. Comparing the structures of the complexes revealed key differences in the orientation and interactions of FEN1 and double-flap DNA with the two clamps that are consistent with their respective functions in providing inherent flexibility for lagging strand DNA replication or inherent stability for DNA repair.

6.2 Introduction

Flap endonuclease 1 (FEN1) belongs to a class of essential nucleases (the FEN1 5′ nuclease superfamily) present in all domains of life (190). FEN1 catalyzes the endonucleolytic cleav-

age of bifurcated DNA or RNA structures known as 5′ flaps. 5′ flaps are generated during lagging strand DNA synthesis or during long-patch base excision repair (LP-BER). The FEN1 substrates are in fact double flap DNA (dfDNA) with DNA on the opposite side of the 5′ flap forming a single nucleotide 3′ flap when bound to the enzyme (191,192). By removing the 5' ssDNA or RNA flap from such substrates FEN1 produces a single nicked product that could be sealed by the subsequent action of a DNA ligase. Consistent with its crucial role in DNA replication and repair, FEN1 is highly expressed in all proliferative tissues and its activity is key for the maintenance of genomic integrity (193). FEN1 has been identified as a cancer susceptibility gene and mutations in it have been linked to a number of genetic diseases, such as myotonic dystrophy, Huntington's disease, several ataxias, fragile X syndrome and cancer (134,194-197).

The FEN1 nuclease activity can be stimulated by its association with processivity clamps such as proliferating cell nuclear antigen (PCNA), which encircle DNA at sites of replication and repair (198-200). PCNA is a recognized master coordinator of cellular responses to DNA damage and interacts with numerous DNA repair and cell cycle control proteins. In this capacity, PCNA serves not only as a mobile platform for the attachment of these proteins to DNA but, importantly, plays an active role in the recruitment and release of these crucial participants at the replication fork(36,38). The dominant mode of interaction for many of these factors is through attachment to the interdomain connector loop (IDCL) of PCNA and the PCNA C-terminus (191,200,201). The trimeric PCNA ring can provide at most three binding sites for replication and repair factors. The crystal structure of the human FEN1 with PCNA indeed revealed three FEN1 enzymes bound to the sliding clamp in different orientations(201). Additionally, a biochemical study of the *Sulfolobus solfataricus* proteins supported the idea that distinct protein partners such as DNA polymerase, FEN1, and DNA ligase could simultaneously associate with

PCNA(101). The competition among proteins to simultaneously bind to the surface of PCNA as well as to their common DNA substrate has led to the notion of conformational switching and handoffs of repair intermediates(101,202). These are key processes in PCNA biology, which are incompletely understood from a mechanistic perspective.

In addition to PCNA FEN1 is known to associate with the alternative checkpoint clamp Rad9-Rad1-Hus1 (9-1-1 complex). While PCNA is comprised of three identical subunits, 9-1-1 is an example of a heterotrimeric sliding clamp (141,142,203). This fact reflects the different protein partners the two clamps engage and the distinct roles these complexes play in coordinating DNA processing. In contrast to PCNA, the 9-1-1 complex is thought to serve as a recruitment platform to bring checkpoint effector kinases to sites of DNA damage, thus activating checkpoint control, and also functions to stabilize stalled replication forks that have encountered DNA lesions(136,144,145). It has also been demonstrated that 9-1-1 interacts with and stimulates enzymes involved in base excision repair (BER), such as NEIL1, MYH, TDG, FEN1, and DNA Ligase I, thus potentially linking BER activities to checkpoint coordination(146-149).

In view of their crucial involvement in replication and LP-BER, a detailed structural comparison of the ternary PCNA/FEN1/DNA and 9-1-1/FEN1/DNA complexes would be of great value. While structural snapshots are available for the individual components of such assemblies (PCNA (pdb id: 1VYM), human 9-1-1 (3GGR)) (Figure S1) and, for some binary complexes (FEN1/DNA (3Q8L) and FEN1/PCNA (1UL1)(22,141,142,192,201,203), the larger ternary assemblies present extreme challenges to molecular crystallography (MX). We have adopted a strategy that integrates high and low resolutions structural data from different sources through computational methods to characterize the structures of such complexes. Here we report models for the ternary PCNA/FEN1/dfDNA and 9-1-1/FEN1/dfDNA assemblies, which were

constructed by combining all available high resolution MX data for the individual components and subassembles. The models were refined by multi-nanosecond atomistic MD simulations. Single particle electron microscopy of negatively stained samples indicated that the structure defined by the 9-1-1/FEN1/dfDNA model exists in solution. Subsequently, the computational model was integrated with the EM data resulting in a 3D reconstruction for the ternary assembly determined at 18 Å resolution. Finally, we present a detailed comparative analysis of the two ternary complexes highlighting the differences in interactions and in the modes of association for FEN1 and dfDNA with PCNA and with the checkpoint clamp.

6.3 Results and Discussion

Overall structure of the ternary FEN1 complexes- To shed light on the conformations and structural dynamics of the ternary complexes we relied on extensive multi-nanosecond molecular dynamics simulations. Integrating the available high-resolution structural data, we were able to refine credible models for the ternary complexes of FEN1/PCNA/DNA and FEN1/9-1-1/DNA, which exhibited no structural discrepancies or steric hindrance between the sliding clamps, DNA and FEN1 (details provided in Supplementary Methods). The models indicated that FEN1 occupies an overall upright position on the polymerase binding face of PCNA or on the corresponding face of the checkpoint clamp. In both models the enzyme was facing the central cavity of the PCNA or 9-1-1 ring. The final models were selected after pairwise RMSD clustering analysis of the MD trajectories (204,205) and the structure closest to the centroid of the most populated cluster was chosen as representative for each ternary complex (Figures 6.1A and 6.1B). Time evolution of the two systems over 100 ns of dynamics was conveniently monitored in terms of RMSD values for each ternary assembly and its constituent parts (Figure 6.2). Our data revealed that the most adaptable component in both complexes was the double-flap DNA substrate, which under-

went the largest displacement in average heavy atom RMSD (7.14 ± 0.76 Å (s.d.) for the PCNA and 6.11 ± 0.54 Å (s.d.) for the 9-1-1 complex, respectively). By contrast the core domain of FEN1 and the subunits of PCNA and the 9-1-1 clamp displayed minimal internal displacements (1.90±0.14 Å (s.d.) for the FEN1 core, 2.31±0.18 Å (s.d) for PCNA and 3.54±0.24 Å (s.d.) for 9-1-1) The most significant internal motions were confined to surface loops of the clamps (*e.g.* P-loop of PCNA, IDC loops of Hus1 and Rad9), the extreme C-terminal ends of PCNA and a few flexible outer helices in FEN1 (α13, α4 and α5, α10, α11 and α12; detailed view of these structural elements is presented in Figure 6.2). Importantly, although motions bringing FEN1 closer to the PCNA central cavity were observed in the simulation (Figure 6.1A and Figure 6.3), the overall displacement along this mode was rather subtle. It was reflected in the observed RMSD for the complexes (3.58±0.49 Å (s.d.) for the PCNA complex and 6.94±0.38 Å (s.d) for the 9-1-1 complex excluding DNA). The finding that the DNA substrate is the most adaptable component in both assemblies is further corroborated by the data presented in Figure 6.1B-E. The figure depicts the computed B-factors mapped onto the structures of the two ternary complexes. Not surprisingly, the entire DNA substrate was found to be highly flexible. The one exception is the central portion of the upstream DNA duplex, wherein the DNA backbone was found to form extensive contacts with the positive inner surface of the sliding clamps. Notably, the B-factors for this middle region of the upstream duplex are lower in the case of the 9-1-1 complex indicating tighter association between 9-1-1 and dsDNA as compared to the PCNA complex. Other mobile regions included the H2TH DNA binding motif (helices α10, α11 and α12) and the "helical arch" region (helices α4 and α5), which caps the 5' flap and the active site of FEN1.

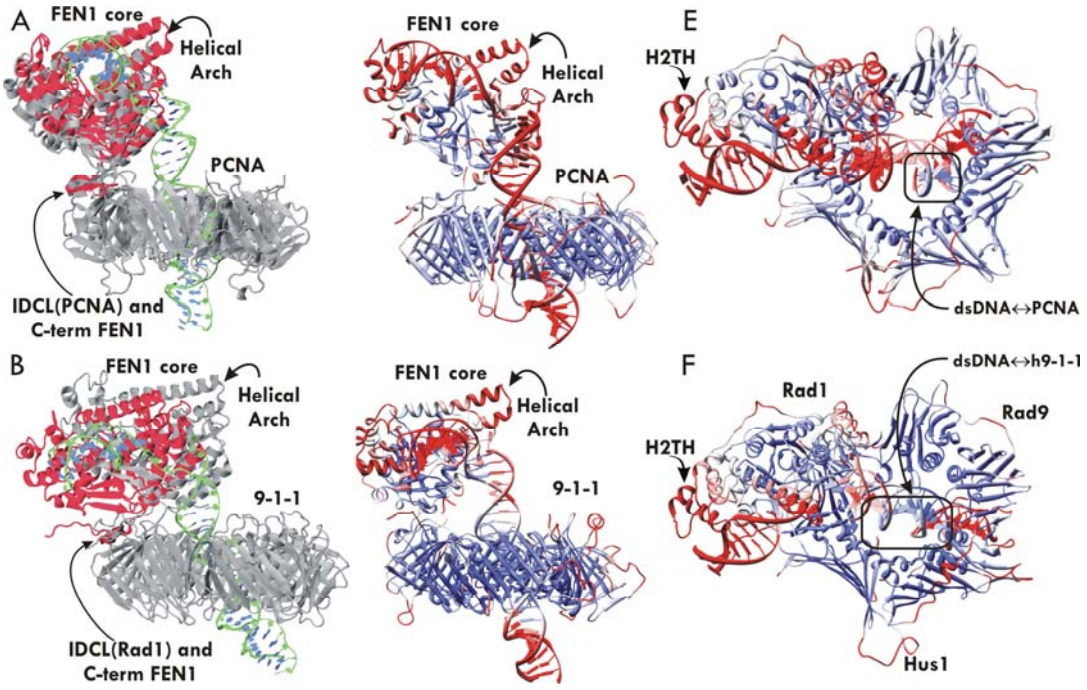

Figure 6.1 Representative structures for the FEN1/PCNA/dfDNA and FEN1/9-1-1/dfDNA complexes from integrative computational modeling

A) the PCNA complex overlayed onto the original DNA-free x-ray structure (1UL1, chain Y); B) the 9-1-1 complex overlayed onto a model based on the 1UL1 structure wherein PCNA has been replaced by 9-1-1 (retaining the relative FEN1 orientation); C-F) Computed B-factors mapped onto the PCNA and 9-1-1 models. The coloring is by B-factor values ranging from blue to red. For both complexes lateral and anterior views are presented and important structural elements are labelled accordingly.

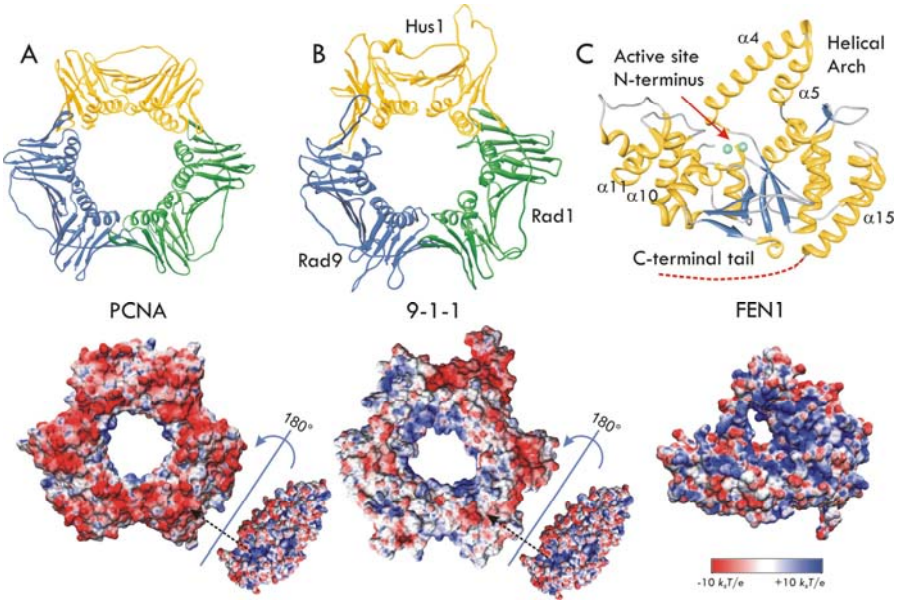

Figure 6.2 Structures and electrostatic potential maps

A) PCNA B) the checkpoint clamp 9-1-1 and C) FEN1. The structures of PCNA and 9-1-1 are colored by individual subunits; FEN1 is colored by secondary structure. The electrostatic potential is mapped onto the molecular surface and colored from red (negative) to blue (positive potential). The surface of FEN1 facing the clamps is characterized by basic patches, which electrostatically complement an acidic patch on the top side of the PCNA ring. In 9-1-1, only the Rad1 subunit possesses an electrostatic pattern complementary to the bottom surface of FEN1.

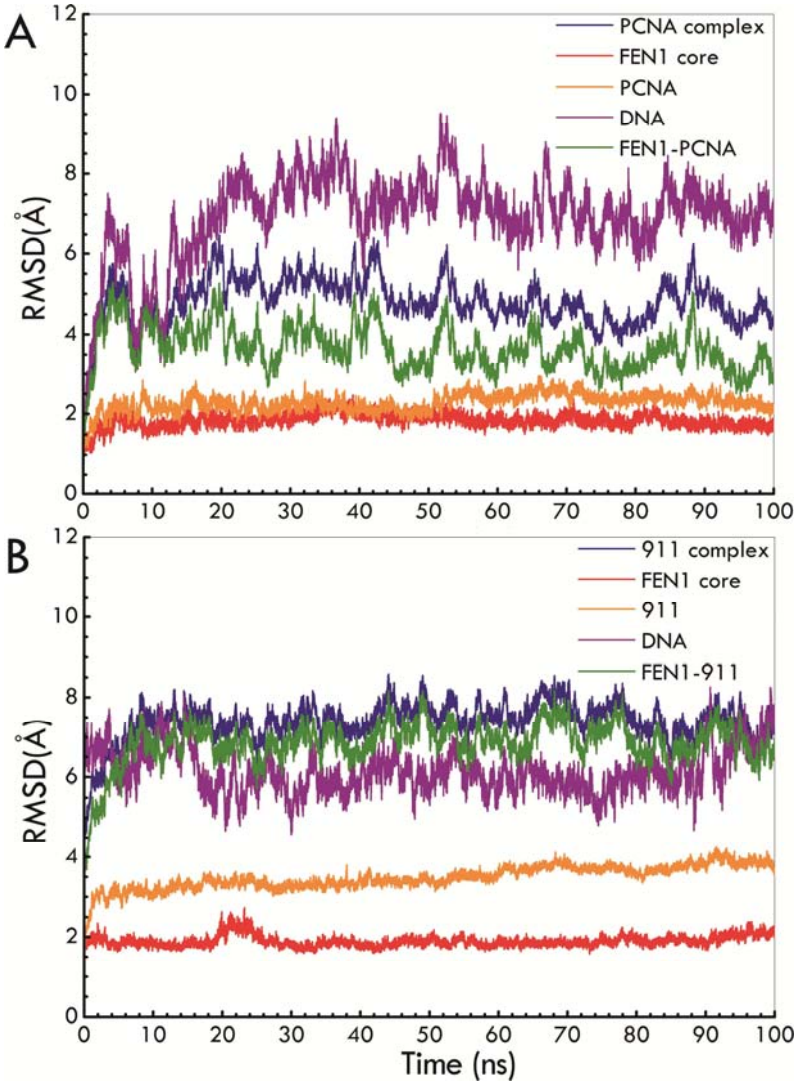

Figure 6.3 Time evolution of the RMSD values for the ternary assemblies and their constituent parts

A) RMSD values for the FEN1/dfDNA/PCNA complex (blue), FEN1 core excluding the C-terminal tail (red), PCNA (yellow), dfDNA(purple) and FEN1-PCNA(green); B) RMSD values for the FEN1/dfDNA/9-1-1 complex (blue), FEN1 core excluding the C-terminal tail (red), 9-1-1 (yellow), dfDNA(purple) and FEN1/9-1-1(green). In both cases the reference structure reflects the FEN1 orientation observed in the x-ray structure of Sakurai et al (1UL1, Chain Y).

Structural determinants of flexibility for the ternary assemblies-The modest change in the orientation of FEN1 from the original crystallographic position could be explained by two intrinsic features of the complexes that permit structural adaptation: (i) the dfDNA substrate is kinked(192,202) by ~100˚and features a moderately flexible single stranded region; and (ii) a pronounced tilt is observed between the axis of the upstream DNA duplex and the plane of the PCNA or 9-1-1 ring (Figure 6.4). The dsDNA fragment goes through the plane of the sliding clamp ring at a sharp tilt angle: 17.34±2.87˚(s.d) for PCNA and 27.19±1.96˚(s.d) for 9-1-1. Interactions between the dsDNA phosphodiester backbone and basic residues lining the inner surface of the clamps are responsible for the observed binding mode in the two ternary complexes and will be discussed later in greater detail. Our findings carry strong parallels to previous computational work (206) on the association of PCNA with dsDNA, which arrived at a similar markedly asymmetric model for the PCNA/dsDNA assembly. Conventional models for PCNA encircling DNA entail that the clamp is perpendicular to the DNA axis, rendering all three PCNA binding sites equivalent. In contrast, recent evidence has collectively suggested that sliding clamps bind DNA asymmetrically. The breaking of the threefold symmetry of the clamp upon DNA binding offers an ideal mechanism for handoff of protein partners. Experimentally, the x-ray structure of the bacterial β-clamp-DNA complex (207) revealed dsDNA passing through the clamp at a sharp angle of 22°. More recently, EM analysis of DNA ligase-PCNA-DNA and Pol-β/PCNA/DNA assemblies showed that dsDNA was tilted by 16° and 13°, respectively (208,209). By contrast, in the replication factor C/PCNA/DNA ternary assembly the DNA axis was found to be nearly perpendicular to the PCNA ring (100). This versatile mode of association between the sliding clamp and DNA appears to be a universal feature observed in bacterial, archaeal and eu-

karyotic clamps and may explain how the sliding clamp is able to accommodate distinct protein partners and correctly position them to complete their functions.

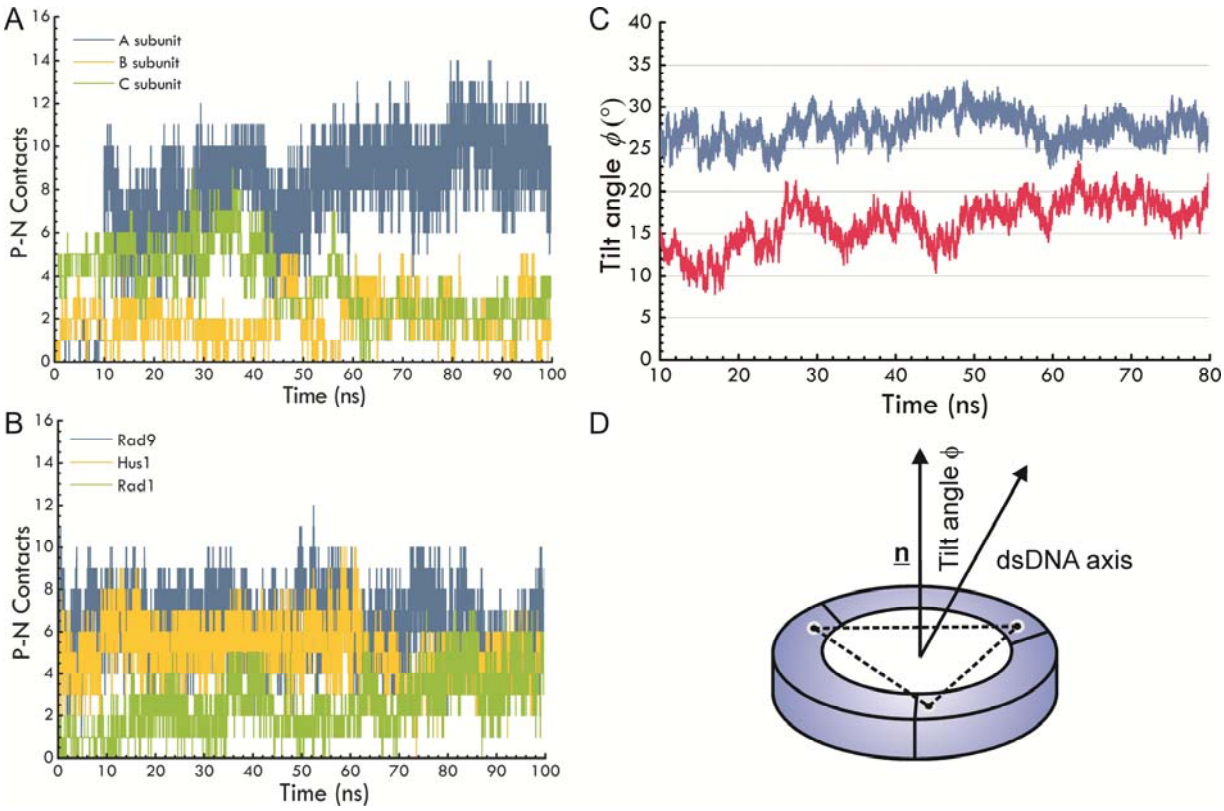

Figure 6.4 Time evolution of the contacts between the side chain N atoms of basic residues positioned within 4.5 Å of DNA backbone P atoms

Data for the PCNA complex is shown in panel A and for the 9-1-1 complex in panel B. In both panels the contacts are broken down by individual subunits and colored in blue (PCNA-A and Rad9), yellow (PCNA-B and Hus1) or green (PCNA-C and Rad1) Panel C shows the time evolution of the tilt angle between the DNA axis and the ring-plane of the sliding clamp for the PCNA (red) and 9-1-1 (blue) complex, respectively. Panel D schematically illustrates how the tilt angle is defined.

EM analysis and MDFF flexible fitting of the 9-1-1/FEN1 binary and ternary complexes- Hybrid methods for structural analysis constitute an emergent area, which could shed light on biological complexes with high degree of structural plasticity. The FEN1 ternary assembles are examples of this type of systems, wherein the inherent flexibility has so far precluded crystallographic analysis. As a direct visualization technique, single particle electron microscopy (EM) has the potential to investigate flexible assemblies (210) (211). Given the importance ascribed to the dfDNA substrate in determining the "upright" position of FEN1 in the models, the next step in our analysis involved comparison of a binary (DNA-free) and a ternary (substrate-bound) complex using single particle EM.

Specifically, binary complex formation for FEN1 and 9-1-1 was assessed by incubating both proteins at 30°C and subsequently running size-exclusion chromatography. Analysis of the corresponding fractions by SDS-PAGE showed 9-1-1 and FEN1 eluting at different retention times, thus indicating weak interaction under our experimental conditions (Figure 6.5). To increase the stability of the complex the sample was cross-linked prior to the chromatography step. Fractions containing the 9-1-1/FEN1 complex were then visualized by EM and data collected as described in Materials and Methods. In parallel a 3-dimensional EM reconstruction of the 9-1-1 checkpoint clamp was generated for comparison purposes (Figure 6.6).

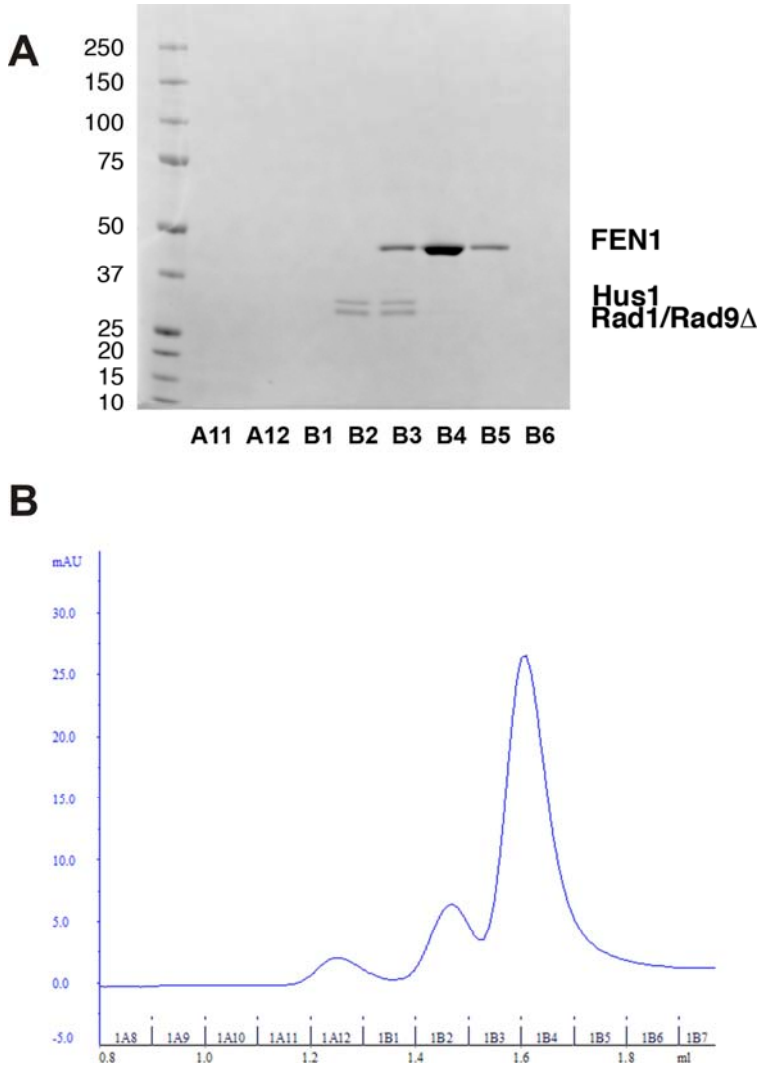

Figure 6.5 FEN1 and 9-1-1 do not form a stable complex under the experimental conditions tested

h9-1-1 in the presence of 3-fold excess of FEN1 was incubated at 30°C for 15min and then applied to a Superdex 200 size-exclusion column. Visualization of the corresponding fractions (labeled from A11 to B6) by SDS-PAGE shows that proteins elute at different retention times. Molecular weight markers are indicated on the left. FEN1/dfDNA/9Δ-1-1 sample purification through size-exclusion chromatography for EM studies. Peaks corresponding to the ternary

complex and DNA are indicated. Fraction numbers and lane numbers of the SDS-PAGE gel are equivalent (first lane: molecular weight markers).

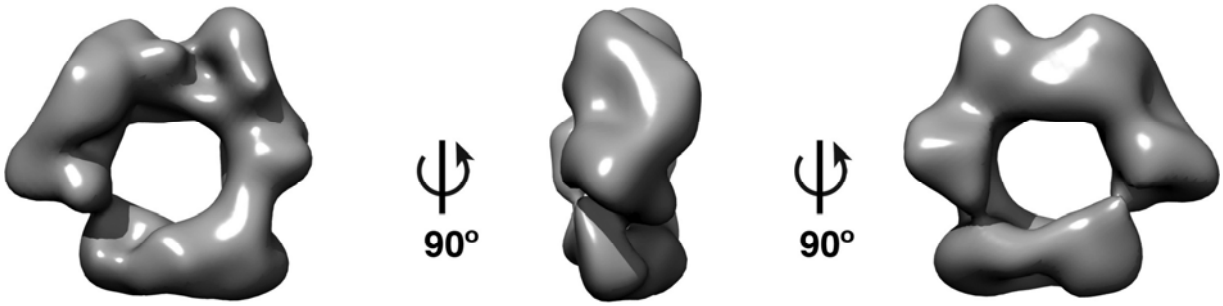

Figure 6.6 Three-dimensional reconstruction of the full-length h9-1-1 complex expressed in insect cells

Front, lateral and back views are shown in solid grey.

Figure 6.7A shows reference-free 2D class averages for the binary 9-1-1/FEN1 complex. Individual images could be classified as either six-lobed rings with an additional density in the periphery of the ring or as compact two-lobed rods with an extra density region to one side. Comparing the images to the 3D reconstruction of 9-1-1, the six-lobed rings and two-lobed rods were respectively assigned as top and side views of the checkpoint clamp. The extra density adjacent to the 9-1-1 ring was highly variable in its orientation and corresponded in all cases to a single FEN1 enzyme bound to the clamp. Our finding corroborates the idea that only one of the 9-1-1 subunits is competent to associate with FEN1. Importantly, the side views clearly showed how FEN could swing by at least 90 degrees from a position in-line with the plane of the ring to an upright orientation facing the central hole of 9-1-1. This is a particularly interesting observation in light of the multiple FEN1 orientations found by Sakurai et al (201) in the binary PCNA complex. In their structure chain X corresponded to the in-line (“sideways”) position while chains Y and Z corresponded to the “upright” position. Our results show that the binary FEN1/9-

1-1 complex displays the same range of conformational variability as the binary FEN1/PCNA complex, pointing to the functional significance of flexibility in these two systems. In fact, the FEN1 positioning in the Sakurai structure could, in principle, be limited by specific crystal contacts and steric constrains within the crystal. By contrast, our EM results are likely to reflect the conformational range of the binary complexes in solution. Thus, we propose that both the 9-1-1/FEN1 and PCNA/FEN1 binary assemblies are flexibly tethered. The functional role of such flexibility could be to allow initial FEN1 binding in the "sideways" orientation. In this position FEN1 could stay associated and move along DNA without affecting the function of proteins bound to the front face of the clamp. Upon encountering the correct DNA substrate, FEN1 could swing into the catalytically competent "upright" orientation and engage in handoffs with proteins simultaneously attached to the other two clamp subunits.

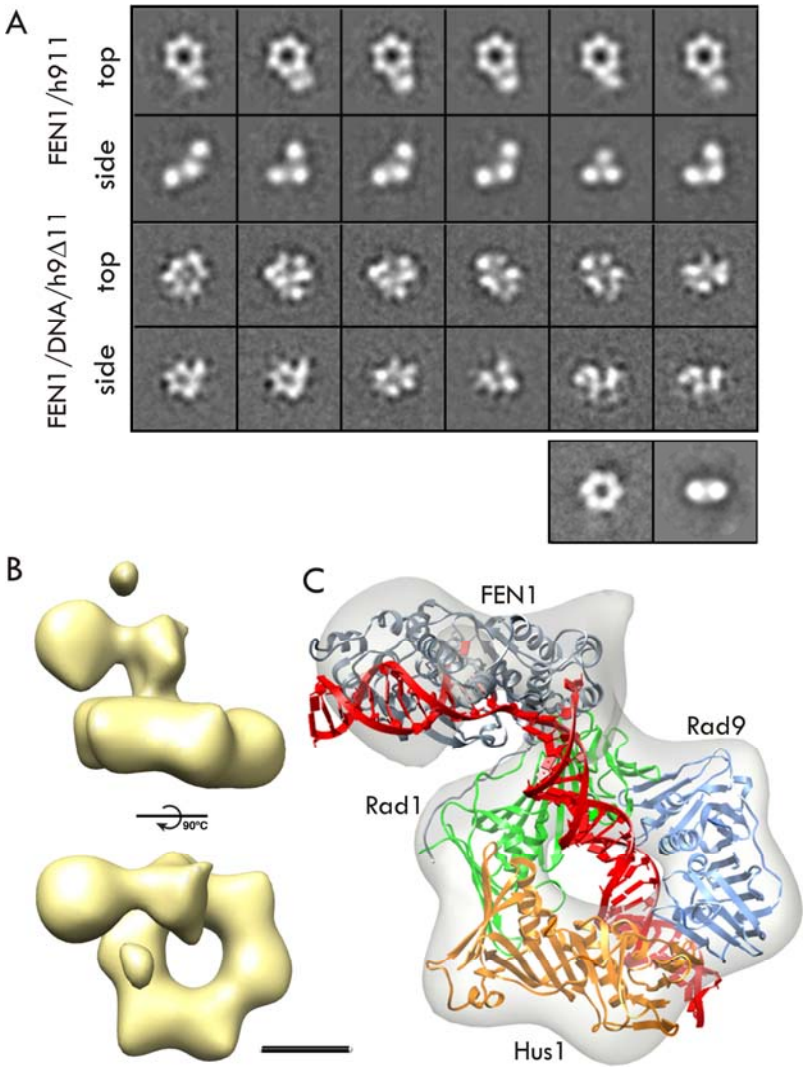

Figure 6.7 FEN1 interacts flexibly with Rad1 and adopts a fixed position in the presence of the DNA substrate

A) Six representative reference-free 2D class averages (both top and side views) for the FEN1/h911 binary complex are compared to those corresponding to the FEN1/DNA/h9Δ11 ternary complex. Top and side views of the h911 complex are shown in the lower panel for reference. B) Side and top views of the FEN1/DNA/h9Δ-1-1 3D reconstruction (scale bar 40 Å). C) MDFF flexible fitting of the FEN1/9-1-1/dfDNA complex into the 3D map of the ternary com-

plex with individual components colored as follows: FEN1 in grey, Rad9 - blue, Rad1-green and Hus1 – yellow and dfDNA in red.

Next, we focused on the structural characterization of the ternary FEN1/9-1-1/DNA complex. The selection of DNA substrate followed the work of Tsutakawa et al (192) except that 10 additional bases were introduced at the 3' end. A band shift assay was performed in the presence of either full-length 9-1-1 or a truncation construct lacking the extra C-terminal tail of Rad9 (h9Δ-1-1). No band shift was observed with the full-length h9-1-1 (Figure 6.8, upper left panel) suggesting that in the absence of the check-point clamp loader the Rad9 C-terminal tail may interfere with DNA binding. However, the appearance of an extra band was observed when FEN1 and DNA where incubated with the truncated version of h9-1-1 clearly indicating the formation of a FEN1/h9Δ-1-1/DNA ternary complex (Figure 6.8, upper right panel). After cross-linking and purification of the ternary complex by size-exclusion chromatography (Figure 6.8, lower panel) samples were visualized by negative stain EM. The small size of the complex (140 kDa) precluded us from carrying out cryo-EM studies.

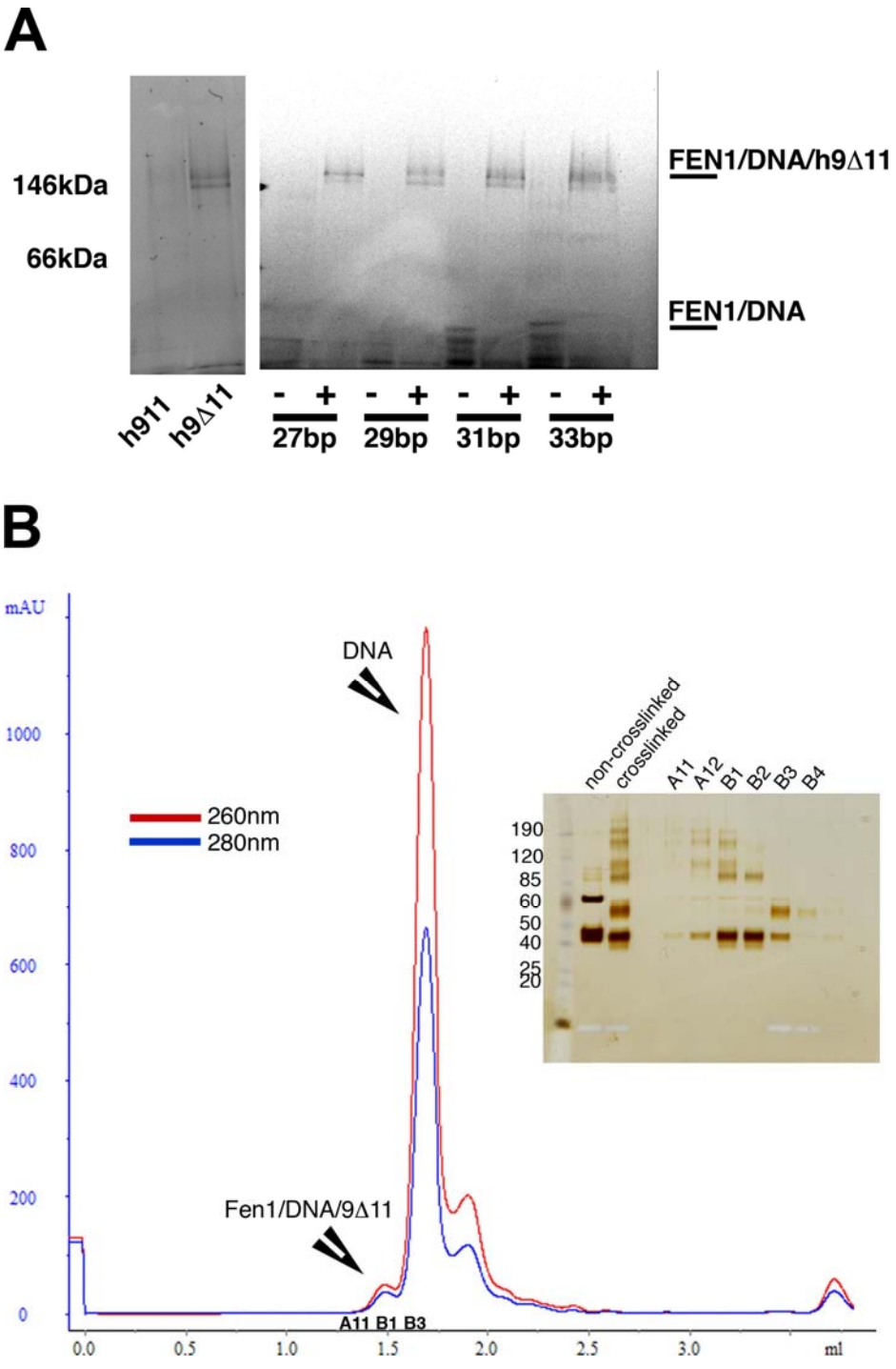

Figure 6.8 Band shif assay on FEN1/DNA/h9Δ-1-1 complex

A) FEN1/DNA/h9Δ-1-1 forms a stable complex as judged by band shift assays. Four different DNA substrates varying in length were incubated with FEN1 in the presence or absence of

h9Δ-1-1 (right panel; + and – indicate the presence or absence of h9Δ-1-1). The presence of a band around 140kDa shows the formation of a ternary complex. No band shift was observed when full-length 9-1-1 was used (left panel, first lane); B) FEN1/dfDNA/9Δ11 sample purification through size-exclusion chromatography for EM studies. Peaks corresponding to the ternary complex and DNA are indicated. Fraction numbers and lane numbers of the SDS-PAGE gel are equivalent (first lane: molecular weight markers)

Figure 6.7A (middle panel) shows representative reference-free 2D class averages of the ternary FEN1/9-1-1/DNA complex. It is evident from the images that in the presence of DNA the extra density lobe above the 9-1-1 ring adopts a well-defined fixed position. Thus, the class averages strongly indicated that addition of dfDNA locks FEN1 into the "upright" position facing the central cavity of the 9-1-1 clamp. This finding is fully consistent with the simulation results for FEN1/9-1-1/DNA, which similarly displayed only this upright conformation.

In order to produce a 3D reconstruction from our EM data we took advantage of the FEN1/9-1-1/DNA model generated from our simulations. The atomistic model was used to create a three-dimensional density map that was low-pass filtered to 60 Å resolution. Computationally generated projections of this low-resolution model qualitatively resembled the EM reference-free class averages (Figure 6.9). Thus, we decided to use this filtered structure as an initial model to assign relative orientations to the different experimental views of the complex. We performed our three-dimensional refinement using iterative projection matching in EMAN2 (212,213) reaching a final 3D reconstruction of the FEN1-DNA-h9Δ11 complex at an estimated resolution of 18 Å (Figure 6.7B). The 3D map contains two principal features - a hexagonal non-symmetrical ring with an external diameter of ~85 Å and a central channel of ~30 Å correspond-

ing to h9Δ11; and a two-lobed extra density ~70 Å in length, corresponding to FEN1. This density was connected to a single lobe of the 9-1-1 ring. The atomic model was positioned by eye into the density as a rigid body and docking was further refined using the fitting tool of the Chimera program. Although the presence of DNA has a clear effect in the relative positioning of Fen1 with respect to the h911 ring, the reconstruction does not show clear density for DNA. However, this was not very surprising, given the well-documented difficulty in visualizing DNA by negative staining.

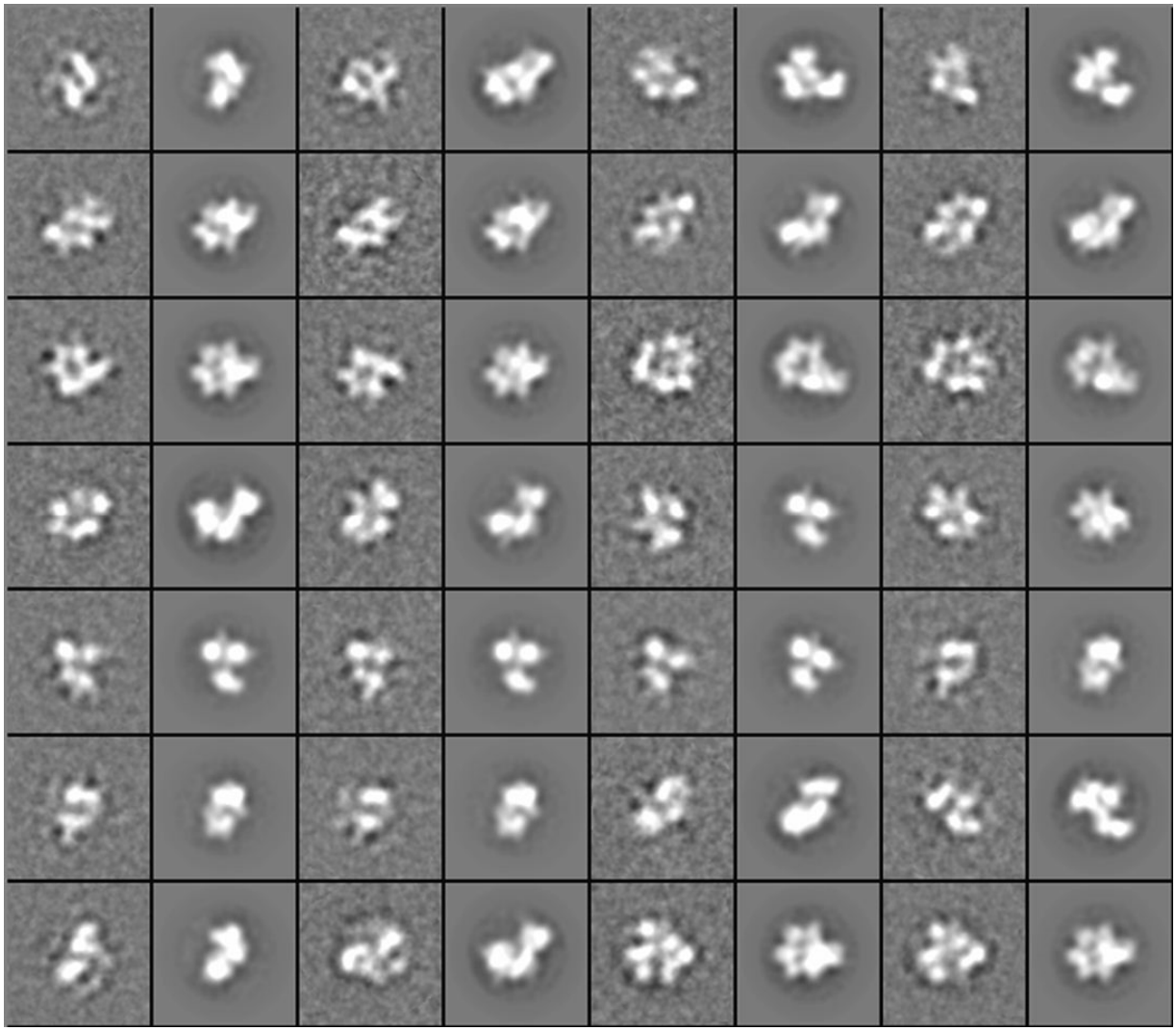

Figure 6.9 2D class averages of the FEN1/DNA/9d11 complex

Panel of reference-free 2D class averages of the FEN1/DNA/9Δ-1-1 complex (odd columns) compared to the forward projections of the model filtered at 60 Å used for projection matching (even columns).

As a final stage of our atomistic model generation and refinement we employed flexible fitting into the experimental EM map using the MDFF method (172,173) (Figure 6.7C). MDFF involves a molecular dynamics simulation wherein external forces proportional to the EM density gradient are applied to bias the atoms of the model into the high-density regions of the EM map. The fitting was carried out in stages as described in the SI Methods section. After the refinement and minimization the FEN1/9-1-1/DNA atomic model exhibited excellent agreement with the EM experiment with fewer than 300 atoms found outside the EM density envelope (at a conservative density threshold of 3.6). All flexible elements including the surface loops of 9-1-1 and the helical arch of FEN1 were consistent with the EM map. The main observation we draw from the fitted atomistic model is that despite the overall upright position for FEN1 on 9-1-1 there is a lateral shift of the FEN1 core (also seen in the simulation) so that the dominant conformation of FEN1 is more "lopsided" compared to the FEN1/PCNA/DNA model and the FEN1 core ends up in closer proximity to the surface of the checkpoint clamp. Thus, a closer comparison of the two ternary complexes reveals significant differences in the mode of association, which correlate with their distinct functions in DNA replication and repair and will be elaborated in the following sections.

The dsDNA interactions of 9-1-1 and PCNA in the ternary complexes are distinct- Our experimental results established that FEN1 is flexibly tethered to 9-1-1 and that DNA binding locks the FEN1 enzyme into an upright conformation in the 9-1-1 complex. How could the dou-

ble flap DNA affect the overall conformation of the complex? In the simulations, we observe specific interactions in the upstream DNA duplex with the inner surface of the sliding clamps. PCNA(22,37) and 9-1-1 (141,142,203) are each composed of three subunits that come together and delimit a ring-shaped inner cavity capable of encircling duplex DNA. The walls of this inner cavity are rich in basic residues (Figure 6.2), largely conserved from yeast to human and important for making contacts with the DNA backbone. We argue that these interactions are responsible for the correct positioning of FEN1 and allow the enzyme to assume an upright position on the two sliding clamps.

Despite similarity in overall architecture, PCNA and 9-1-1 present distinct inner surfaces to dsDNA that are characterized by different shapes, electrostatics, hydrophobic and polar patches and secondary structure elements. The dynamic association of the upstream DNA duplex to the inner cavity of PCNA and 9-1-1 was exhaustively sampled in the simulations, which allowed us to delineate and compare the interactions of dsDNA with the two clamps. Surprisingly, the mode of association with dfDNA was strikingly different for PCNA versus 9-1-1 (Figure 6.10). For the PCNA ternary complex a highly asymmetric mode of association was confirmed, with only one of the PCNA subunits forming the majority of stable contacts with the DNA backbone (K20, K217, R149, R146 and K14). The second subunit was contributing three stable contacts (K217, K80, K20) and the third subunit made no persistent contacts with DNA at all. Furthermore, the Arg and Lys residues making salt bridge interactions with the DNA phosphate groups were found to be evenly distributed between the two strands of the DNA duplex. By contrast, in the 9-1-1 complex the majority of such contacts track along the template strand as it passes through the 9-1-1 central hole. The extended interface formed between the template strand and 9-1-1 contains 10 persistent contacts. Three additional contacts (K229 from Rad9, K25 from Hus1

and K92 from Rad1) are made with the opposite strand and are positioned roughly one helical turn apart. Consistent with the previously noted differences in the computed B-factors, 9-1-1 displays significantly tighter association with DNA, forming 13 stable contacts with the DNA backbone (compared with 9 contacts in the PCNA complex).

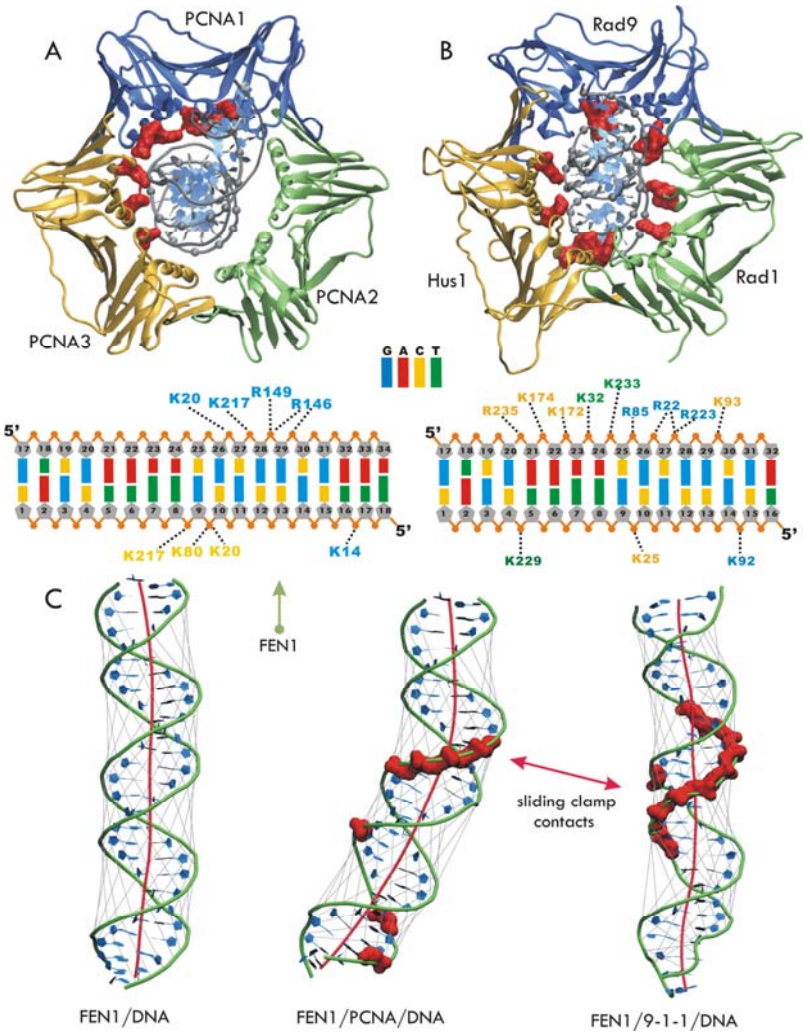

Figure 6.10 In the ternary complexes PCNA and 9-1-1 engage the upstream DNA duplex in two distinct binding modes

A-B) Cartoon representations of PCNA and 9-1-1 binding to dsDNA, colored in blue for Rad9 and PCNA1, yellow for Hus1 and PCNA3 and green for Rad1 and PCNA2. The dsDNA

phosphodiester groups (grey spheres) and basic residues on the inner surface of PCNA and 9-1-1 (red surfaces) are shown. Below we explicitly list all persistent DNA contacts (observed in more than 50% of the frames in the MD trajectories) for the PCNA and the 9-1-1 complex, respectively. C) Averaged structures for the upstream DNA duplex from FEN1/dfDNA(left), FEN1/dfDNA/PCNA(middle) and FEN1/dfDNA/9-1-1(right) simulations. The DNA is shown in ribbons representation. The DNA axis was computed with the program Curves+ and shown in red. The gray lines are representative of the widths of the major and minor grooves of DNA.

Unlike the asymmetric interface in PCNA, the set of contacts with 9-1-1 are evenly distributed among the three subunits (4 contacts with Hus1, 4 with Rad1 and 3 with Rad9) and are also more persistent in the MD trajectories compared with the PCNA complex (Figure 6.4). The distinct distribution of basic residue contacts in the two complexes has consequences for the overall positioning of the dsDNA substrate (Figure 6.4C). It also results in observed differences in the local structure of the upstream DNA duplex as it passes through the sliding clamp ring. In Figure 3c, we compare results from three simulations: (i) the FEN1/dfDNA complex in the absence of PCNA or 9-1-1; (ii) the FEN1/dfDNA/PCNA ternary complex and (iii) the FEN1/dfDNA/9-1-1 ternary complex. In each case, the structure of the DNA substrate was averaged over 80 ns (excluding the first 20 ns as equilibration) and the axis of the upstream duplex was computed with the program Curves+ (214). In the absence of either clamp, the dsDNA axis is essentially straight and DNA duplex conforms closely to the expected canonical B-form geometry. By contrast, the asymmetric mode of association observed for the FEN1/dfDNA/PCNA complex results in a moderate bending of the DNA axis by ~20 degrees. Consistent with the more symmetric contacts distribution, the 9-1-1 complex exhibits no such tendency for DNA

bending. While the evenly spaced basic residues lead to a straight axis, the DNA is not entirely free from structural distortion – we observe contraction of the minor groove at the level of the 9-1-1 interface accompanied by major groove widening above and below the interface.

Collectively, our findings point to a different mode of association for the double flap substrate when bound within the PCNA versus the 9-1-1 complex. We argue that such differences may allow the 9-1-1 complex to function specifically in the sensing and repair of damaged DNA and, thus, carry important implications for the way 9-1-1 operates in base excision repair. Surprisingly, a recent study, which carried out a large-scale comparative analysis of processivity clamps from a variety of organisms, showed that the electrostatic properties of the human checkpoint protein Hus1 and to a lesser degree Rad9, but not of Rad1, were more similar to non-ring viral processivity factors than to the eukaryotic PCNA (215). These viral processivity factors (*e.g.* UL42, UL44) possess an increased positive charge on the DNA-binding face compared to the ring-forming clamps (PCNA, PolIII subunit). They are also known to bind to the DNA directly rather than rely on a topological connection to DNA. Our results are consistent with this view that elevated positive potentials on the inner surface of 9-1-1 (specifically for Hus1 and Rad9; Figure 6.2) lead to tighter association of the checkpoint clamp with DNA, which is likely related to the dual role of the checkpoint clamp as a recruiting platform in base excision repair and as a damage sensor in checkpoint control.

Specific contacts with the clamp surface modulate the FEN1 orientation in the complexes- While the interactions of dfDNA with PCNA or the 9-1-1 clamp are crucial in establishing the "upright" orientation of FEN1 in the ternary complexes, a multitude of additional direct interactions between FEN1 and the surface of the two clamps appear to be significant. These interactions are schematically presented in Figure 6.11 and classified by type as hydrogen bonding,

side chain hydrophobic or charged contacts. Notably, in our analysis we have taken care to average along the simulation trajectories and, based on a geometric criterion for each interaction type, to exclude contacts occurring only transiently. Thus, all contacts shown are stable and persistent above a characteristic threshold level. Importantly, these contacts at the interface between FEN1 and PCNA or 9-1-1 appear important in fine-tuning the orientation of the FEN1 core. Specifically, we note that the long C-terminal tail of FEN1 makes extensive contacts with Rad1 or PCNA along the IDC loop region. Biochemical analyses have highlighted the role of the FEN1 C-terminal tail, which is required for stimulation of FEN1 activity by either PCNA or the checkpoint clamp. Intriguingly, the exact C-terminal residues responsible for stimulation by the two clamps appear to be distinct (216). The residues within the PIP-box are critical for stimulation by PCNA. By contrast, residues at the extreme C-terminal end (last 21 residues) are important for stimulation by 9-1-1. In our two models, the FEN1/PCNA interface features more hydrophobic contacts, displaying two distinct hydrophobic pockets ((i) P234, V233, A208, I128, L47, V45 and (ii) M119, L118, A96, M68, A67 and L66). Residues from these pockets interact directly with the FEN1 PIP-box (βA–αA motif). In contrast, the Rad1 surface appears less hydrophobic in this region, lacking both hydrophobic pockets (Figure 6.11). Suggestively, the closer proximity of the tilted FEN1 catalytic core to Rad1 might allow the FEN1 C terminus to come back up to interact with the FEN1 catalytic core in the 9-1-1 complex, providing a rationale for the perplexing data that the extreme C terminus is required for 9-1-1 stimulation but not for stimulation by PCNA.

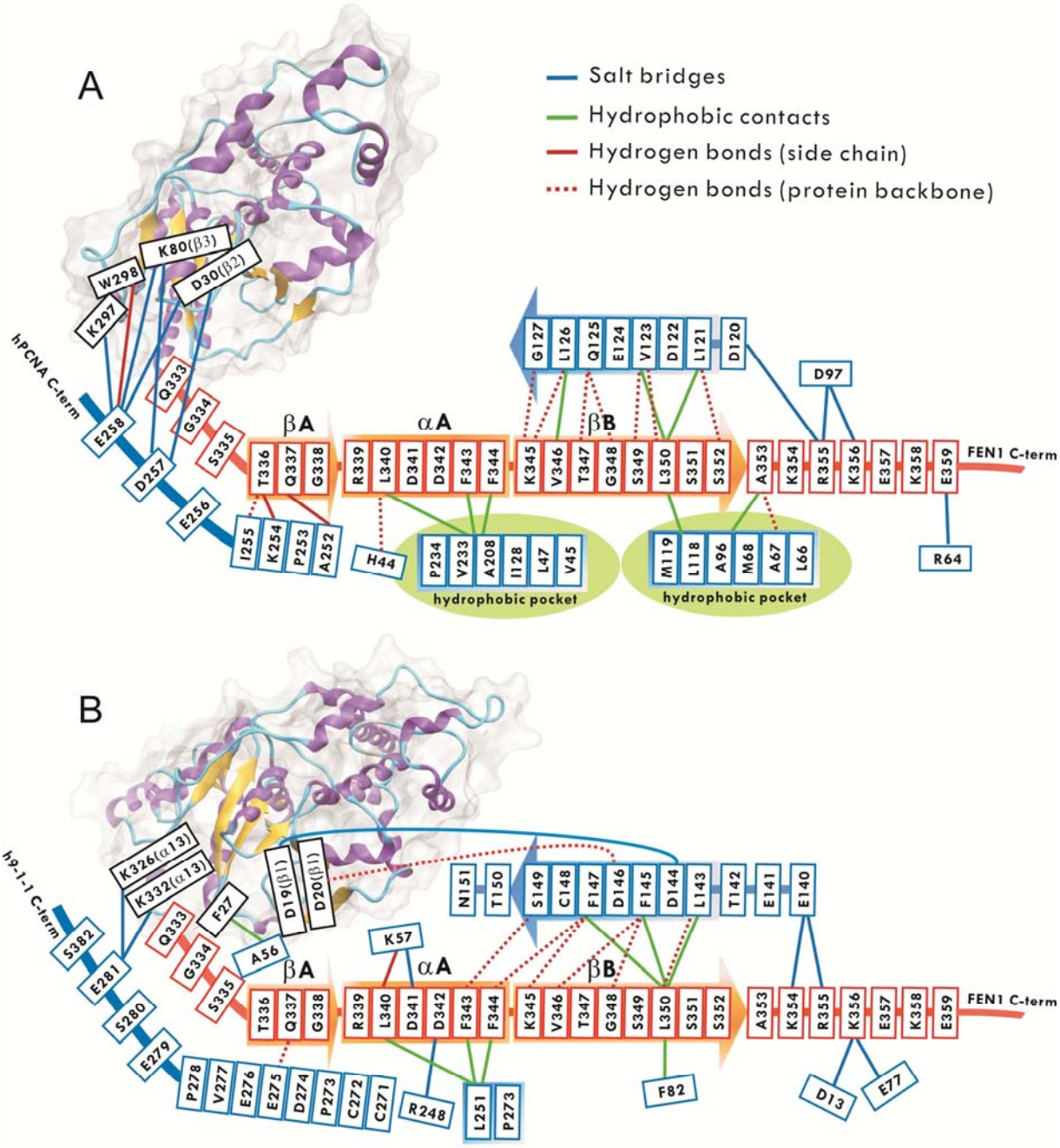

Figure 6.11 Schematic representations of the FEN1/PCNA and FEN1/9-1-1 interfaces reveal significant differences in the way FEN1 engages the two clamps

Panel A) shows FEN1/PCNA interface contacts; panel B) shows FEN1/9-1-1 interface contacts. Persistent residue contacts (more than 50% for H-bonds and salt bridges; more that 90% for hydrophobic contacts) are shown as lines and colored according to type – salt bridg-

es(blue), hydrogen bonds (red; side chain – lines; backbone – dotted lines), side chain hydrophobic contacts (green). Secondary structure elements are shown for the FEN1 C-terminal tail in orange and sliding clamp (PCNA/9-1-1) in blue. The core of FEN1 is shown in ribbon representation and colored by secondary structure. Hydrophobic pockets on the PCNA surface are highlighted in green.

6.4 Concluding remarks

The role of sliding clamps in increasing processivity for DNA-editing enzymes has been long appreciated. Whether these clamps simply act as tethers or play a more active mechanistic role in DNA replication and repair remains controversial. Our combined EM and computational results provide insights into the ternary complexes PCNA/FEN1/DNA and 9-1-1/FEN1/DNA (Figure 6.12). and imply an active role for the clamps. The observed FEN1 interactions in the models are consistent with distinct functions for the complexes. Our models show the DNA substrate is functionally bound by FEN1 with the upstream duplex passing through the inner hole of the ringed clamps. Due to segmental flexibility, in the DNA-free complexes FEN1 is able to occupy multiple positions. In the ternary complexes the DNA substrate locks FEN1 into an upright position producing a single incision-capable state. The DNA passes through the rings at a tilt and the tilt angle is larger by 9° in the 9-1-1 complex. The computational analyses also suggest that 9-1-1, which has more positive inner surface than PCNA, has almost 50% more charged interactions with the DNA. It is interesting to note that the more symmetrical PCNA ring shows a biased DNA interaction toward one side of the ring, whereas the 9-1-1 interactions involve all three subunits. The distinct DNA interactions are consistent with different roles for the two clamps in DNA replication and repair. During replication, PCNA must be mobile and slide on DNA along with replicative polymerases. However, the checkpoint clamp 9-1-1 acts as a tempo-

rary scaffold during repair of a section of DNA. Such functional differences arguably dictate corresponding differences in the interactions and the mode of association of the clamps in their respective replication/repair complexes. The tethered flexibility observed in the binary complexes of PCNA/FEN1 and 9-1-1/FEN1 would allow FEN1 to swing from an out-of-the-way position on the outside of the ring to the upright functional position. This flexibility is consistent with the rapid interchange between PCNA-bound polymerase and FEN1 during Okazaki fragment maturation.

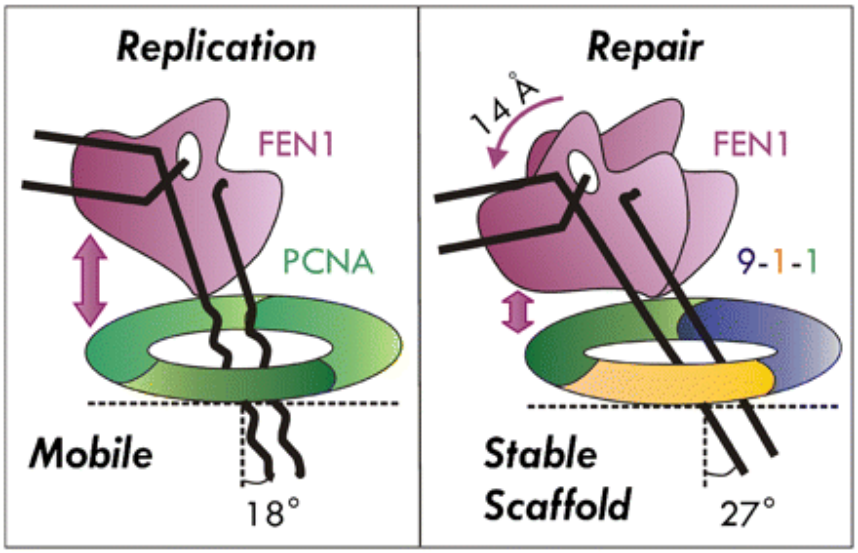

Figure 6.12 Functionally important differences between the PCNA/FEN1/DNA and 9-1-1/FEN1/DNA complexes.

PCNA and 9-1-1 serve different functions and, as a consequence, display different modes of association to FEN1 and DNA. The DNA and FEN1 are more tilted in 9Δ-1-1 complex than in the corresponding PCNA complex. The 9Δ-1-1 ring shows more interactions with the upstream duplex and stabilizes DNA passing through the ring. Such differences at the structural level may translate into functional differences in the ways these complexes engage at the replication fork. PCNA/DNA interactions are linked to processive and mobile sliding, whereas the tighter 9-1-1 interactions would be consistent with a scaffolding role for 9-1-1. Materials and Methods

Sample preparation, EM data collection and processing- FEN1-h911 binary complex was formed by incubating h911 with a 3 molar excess of FEN1 at RT for 15min in 20mM Hepes pH=7.6, 80mM KCl. Complex was cross-linked and purified by size-exclusion chromatography. FEN1-DNA-h9Δ11 ternary complex was cross-linked and further processed as for the binary complex. 4 µl of the purified complexes were loaded onto continuous carbon grids and complexes were stained either with 2% PTA (FEN1-h911 binary complex) or 2%UAc (FEN1/DNA/-h9Δ11 ternary complex and h911 alone). After cross-linking and purification of the complexes the samples were visualized by negative stain EM. Samplers were visualized using the Leginon data collection software (217) on a Fei Tecnai F20 microscope. Data was collected using a 80000X magnification (1.5Å/pixel) in low dose mode ($20e^-/Å^2$) on a Gatan 4×4K pixel CCD camera (15µm pixel size). All the processing of two-dimensional data was performed using programs and utilities contained within the Appion processing environment (25).

Computational models and protocols- Initial models were constructed based on structures from the Protein Data Bank (accession codes 1UL1 and 3Q8L) (192,201). The FEN1-DNA complex was superimposed onto the FEN1-PCNA structure by optimally aligning FEN1 to the Y chain in the 1UL1 complex and the 3' end of the dfDNA was extended so that it could pass through the PCNA/9-1-1 ring. Hydrogen atoms, counterions (Na^+), an additional 100 mM NaCl concentration and TIP3P solvent(218) we introduced. The systems were then minimized and equilibrated. Production runs were carried out in the NPT ensemble (1 atm and 300 K) for 120 ns for the FEN1/PCNA/DNA and 100 ns for the FEN1/9-1-1/DNA complex. The simulations employed SPME electrostatics(117), 10 Å cutoff for short-range non-bonded interactions and 2 fs integration time step. All the simulations were performed using the NAMD 2.7 code(219) with

the AMBER Parm99SB parameter set(119) with modified nucleic acid parameters (BSC0) (61) on Hopper II, a Cray XE6 system at the National Energy Research Scientific Computing Center.

CHAPTER 7. STRUCTURALLY DISTINCT COMPLEXES OF UBIQUITIN- AND SUMO-MODIFIED PCNA LEAD TO DISTINCT DNA DAMAGE RESPONSE PATHWAYS

7.1 Abstract

Proliferating cell nuclear antigen (PCNA) is a pivotal replication protein, which also controls cellular responses to DNA damage. Posttranslational modification of PCNA by SUMO and ubiquitin plays a substantial role in modulating these responses. How the modifiers alter PCNA-dependent DNA repair and damage tolerance pathways (e.g. translesion syntheisis or homologous recombination) is largely unknown. Here we employed a hybrid computational protocol to model ubiquitin-modified and SUMO-modified PCNA. We combined the resulting models with small angle X-ray scattering (SAXS) data to assess the structural differences among three modified PCNA complexes: $PCNA_{K164}$-Ub, $PCNA_{K107}$-Ub and $PCNA_{K164}$-SUMO. Our results argue against allosteric models for the role of the modifier. We show that SUMO and ubiquitin have distinct modes of association to PCNA. Ubiquitin binding is characterized by segmental flexibility with multiple docked Ub positions on PCNA. By contrast, SUMO associates by simple tethering and adopts extended flexible conformations. Such dissimilarities could be traced to the anticorrelated surface electrostatic potentials of the two modifiers. Thus, our integrative modeling approach reveals a structural basis for the distinct functional outcomes upon PCNA ubiquitination or SUMOylation.

7.2 Introduction

Dynamic assembly, coordinated access to DNA, and conformational switching are critical aspects of DNA replication and repair, essential processes upon which life depends. In both processes, the sliding clamp proliferating cell nuclear antigen (PCNA) (22,38,132) acts as a mas-

ter coordinator of multiple pathways controlling replication and DNA damage responses (DDR). PCNA's toroidal shape allows it to topologically encircle DNA while also binding to core replisomal constituents, numerous repair and cell cycle control proteins (35,36,38,134). In this capacity, the sliding clamp acts as a platform for the assembly of the replication machinery on DNA. How PCNA achieves coordination of multiple replication and repair activities is incompletely understood. Partner proteins bind to PCNA predominantly through a consensus sequence termed PIP box motif. The PIP-box is considered the primary element conferring affinity for partner proteins. However, different enzymes feature similar PIP motifs, which in turn localize to the same sites on the clamp surface. Therefore, it is unclear how PIP-box interactions alone could mediate the mechanisms wherein PCNA differentially coordinates activities of multiple partners. Furthermore, the PIP-box motifs are commonly found within flexible tail regions of the proteins. Therefore, direct PIP-box interactions achieve only simple tethering to the clamp. If not the PIP-boxes then what are the controlling elements responsible for partner coordination? Clues have emerged from recent studies, suggesting that secondary contacts with the PCNA subunits and the tilting of the DNA duplex going through the ring-shaped clamp play an important role (143,206). By locking each enzyme in an optimal orientation these contacts regulate access to the substrate. They may also temporarily block PCNA attachment of other partners to ensure correct ordering of the DNA-editing transactions in replication and repair.

Emerging evidence suggests that posttranslational modifications (PTMs) of PCNA also play a critical role in coordinating DNA damage responses to suppress genome instability. This added level of regulation is realized through the interplay of PTMs and partner protein recruitment to PCNA. Specifically, reversible covalent attachment of ubiquitin or ubiquitin-like (UBL) proteins (220-226) (e.g. SUMO) selects among alternative damage processing pathways such as

homologous recombination or translesion synthesis (TLS). The structural basis for such selection has so far remained elusive. Ubiquitin and SUMO both attach to their PCNA target through a glycine residue from the end of their flexible carboxy termini. This glycine forms an isopeptide linkage with a specific lysine side chain on the target. Despite sharing a common β-GRASP fold, SUMO and ubiquitin elicit distinct functional outcomes, and characteristically have been portrayed as having antagonistic roles (220). However, recent studies (227) have significantly attenuated this picture by showing the two modifiers could act in concert to affect cellular fates.

UV irradiation or exposure to DNA damaging agents commonly results in lesions that cause replication fork stalling. The cellular response in this case is mono-ubiquitination of PCNA at a conserved lysine K164, which in turn triggers translesion synthesis (TLS) – an essential damage tolerance pathway (224,227). In TLS the replicative polymerase is transiently exchanged with a specialized TLS polymerase capable of progressing past lesions in the template strand (228). Ubiquitin attachment provides an additional binding surface for the TLS polymerase, which usually possesses an ubiquitin-binding motif (222). By contrast, PCNA SUMOylation at positions K164 or K127 leads to suppression of homologous recombination through the recruitment of the antirecombinogenic helicase Srs2 (229). Srs2 interacts with PCNA through its carboxy-terminal domain, which harbors tandem receptor motifs (230,231) (a SUMO interaction motif (SIM) and a non-canonical PIP-box). Both motifs are required for specific recognition of SUMO–PCNA. The function of Srs2 is to inhibit the formation of Rad51 filaments (232,233), thereby assisting the ubiquitin-dependent TLS pathway. Thus, PCNA ubiquitination and SUMOylation offer a striking example of crosstalk between pathways controlled by distinct PTMs (221,227). While attachment of a single ubiquitin is the dominant DDR response in mammalian cells, yeast PCNA can also undergo polyubiquitination through the non-proteasomal

K63 linkage (224). Polyubiquitination of PCNA at K164 channels DDR to yet another pathway - error-free damage bypass by template switching. Finally, in yeast PCNA mono-ubiquitination at position K107 helps cells to overcome defects in DNA ligation (234). How could PTMs introduced at different positions on PCNA result in such vastly divergent functional outcomes? Furthermore, how do posttranslational modifications of PCNA facilitate recruitment of subsequent effector proteins in these pathways?

To answer these questions we modeled PCNA covalently modified by ubiquitin and SUMO at different attachment positions using a multiscale computational protocol. The models were used in conjunction with solution X-ray scattering data (SAXS) (235-238) to assess the structural differences among three PTM-PCNA complexes ($PCNA_{K164}$-Ub, $PCNA_{K107}$-Ub and $PCNA_{K164}$-SUMO). Here we show that SUMO and Ub have distinct modes of binding to PCNA and that the position of ubiquitin attachment, 107 versus 164, alters conformation. Ubiquitin bound to PCNA can dynamically adopt multiple discrete conformations and has segmental flexibility. By contrast, SUMO associates to the clamp by simple tethering, resulting in extended flexible conformations.

7.3 Results

Distinct architectures of modified PCNA complexes uncovered by SAXS data- We conducted small angle X-ray scattering (SAXS) experiments to probe the overall architecture and flexibility of $PCNA_{K107}$-Ub and $PCNA_{K164}$-SUMO complexes in solution. These experiments involved constructs comprised of Saccharomyces cerevisiae PCNA with Ub/SUMO moieties exclusively attached at positions K107 and K164, respectively. Covalent attachment in the PCNA-Ub complex was achieved by chemical cross-linking with a K107C PCNA mutant (239). The PCNA-SUMO complex was formed by split-fusion (231,240). The split-fusion and chemical li-

gation methods yield linkages close to the enzymatically produced isopeptide bond and offer a simple and specific approach to introduce Ub or SUMO modifications without the complexity of multiple enzymatic steps. Previous studies have validated these strategies by showing that chemically and enzymatically ubiquitinated PCNA have identical biochemical properties (239). Furthermore, previous solution structural results (241) on $PCNA_{K164}$-Ub species were found to be consistent, irrespective of the exact linkage (split-fusion or chemically cross-linked). For $PCNA_{K107}$-Ub and $PCNA_{K164}$-SUMO, SDS-PAGE analysis showed that the Ub and SUMO moieties were present in 1:1 ratio relative to the PCNA subunit in both constructs. Therefore, each PCNA clamp harbors three ubiquitin or SUMO moieties corresponding to the three equivalent points of attachment on PCNA.

SAXS data were collected for $PCNA_{K107}$-Ub and $PCNA_{K164}$-SUMO complexes and compared to the previously obtained scattering profile of $PCNA_{K164}$-Ub and simulated profiles from available crystal structures (split-fusion $PCNA_{K164}$-Ub (240) and $PCNA_{K164}$-SUMO (230); PDB ID 3L10 and 3V60, respectively). All three experimental scattering profiles (Figure 7.1) are distinct, revealing that the three PTM complexes adopt conformations with different levels of compactness in solution. $PCNA_{K107}$-Ub is slightly more compact than $PCNA_{K164}$-Ub. SUMOylated PCNA presents an I(q) profile (Figure 7.1A) with similar overall shape to the SAXS profile of the 3V60 crystal structure, but considerably more extended (Figure 7.1B). Fitting to the SAXS data, the crystal structures 3L10 and 3V60 produced high χ_{free} values consistent with significant discrepancies between the conformations adopted in the crystal structures and the solution phase ensembles. Guinier plots (Figure 7.1C) for $PCNA_{K107}$-Ub and $PCNA_{K164}$-SUMO were linear, indicating that the samples were not aggregated. The radius of gyration, Rg, is accurately defined by SAXS regardless of sample concentration and contrast and therefore provides an important

constraint for possible solution structures (236). The Rg values (Table 7.1) based on the Guinier analyses were 37.2 and 35.6 Å for $PCNA_{K107}$-Ub and $PCNA_{K164}$-Ub, respectively. These values are comparable to the 36.05 Å Rg calculated from the 3L10 crystal structure. The $PCNA_{K164}$-SUMO complex features a significantly larger Rg of 46.8 Å suggesting that SUMO may occupy more extended positions on PCNA compared to ubiquitin.

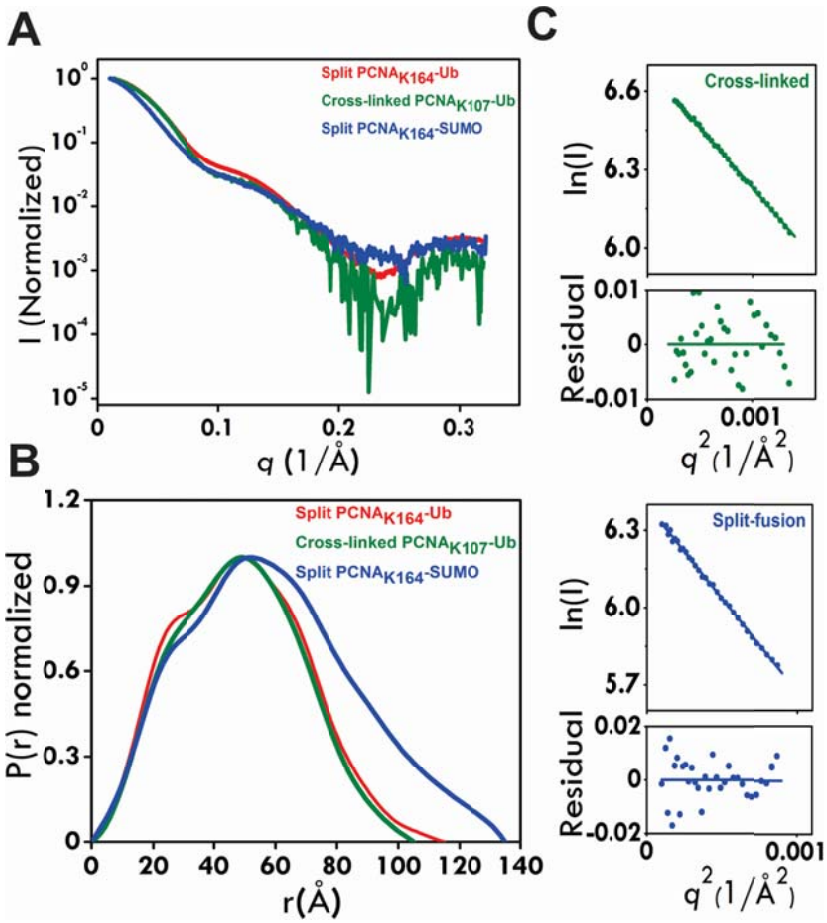

Figure 7.1 Distinct architectures of PCNA-Ub and PCNA-SUMO complexes from SAXS analysis

A) SAXS profiles of PCNAK164-SUMO (blue), PCNAK107-Ub (green) and PCNAK164-Ub (red); B) P(r) functions for PCNAK164-SUMO (blue), PCNAK107-Ub (green)

and PCNAK164-Ub (red); C) Guinier analyses of SAXS data for PCNAK107-Ub (green) and PCNAK164-SUMO (blue) showing relative linearity of the samples in Guinier region, indicating lack of aggregation.

Table 7.1 Models identified by the minimal ensemble search (MES)

Model	PCNA-Ub (K107)[a]	PCNA-SUMO (K164)[a]
1[b]	43%	50%
2[b]	26%	22%
3[b]	13%	20%
4	6%	8%
5	6%	
6	6%	

[a]Contributed by each model; [b]Models were highlighted in Figure 3

Ubiquitin- and SUMO-PCNA conformations from conjugated protein docking- To further explore the substantial differences between ubiquitin and SUMO positioning in the modified PCNA complexes, we constructed computational models of ubiquitin (SUMO) ligated to PCNA. We used a recently developed protocol for conjugated protein docking in Rosetta 3.4. The method samples the conformational space available to ligated proteins (242-244) using a standard Rosetta Metropolis Monte Carlo random sampling tool (245). The protocol involved docking a single Ub (SUMO) moiety to PCNA conjugated at the K107 or K164 position, respectively. In total, 4,970 PCNA-Ub models and 7,525 PCNA-SUMO models were produced, of which 4,791 $PCNA_{K107}$-Ub models and 4,449 $PCNA_{K164}$-SUMO models passed the score and buried surface area filter. The short linker between the modifier and the PCNA attachment point limited accessible conformations. This allowed us to locally saturate the PCNA surface with decoys (i.e.

Ub/SUMO docking poses) and achieve extensive sampling. Scoring results from Rosetta are shown in Figure 7.2. For $PCNA_{K107}$-Ub, the all-atom score of the docking decoys is plotted as a function of the Cα root-mean-square deviation (RMSD) from the lowest scoring PCNA-Ub structure. For $PCNA_{K164}$-SUMO, the X-ray structure 3V60 (230) was used as a reference to compute the Cα RMSD. Figure 7.2 revealed a clear distribution pattern for the decoys, which clustered into bands around specific RMSD values. For PCNA-Ub the lowest scoring decoys formed score "funnels" indicating several preferred position for Ub to reside on the surface of PCNA. The majority of $PCNA_{K107}$-Ub decoys had Rosetta energy scores ranging from −1360 to −1100. For $PCNA_{K164}$-SUMO most scores were found in a much narrower range between −1030 to −920. Collectively, the higher Rosetta scores, the narrow score range and the absence of low-scoring outliers for the PCNA-SUMO complex implied significantly weaker association between PCNA and SUMO as compared to ubiquitin. Next, we selected the lowest-scoring structurally distinct models from the Rosetta output (Figure 2). These docking positions (one chosen from each score funnel) were refined using all-atom explicit solvent molecular dynamics (MD) with the program NAMD (73,74). The MD simulation and analysis allowed us to eliminate conformations where the Ub or SUMO moiety interacted primarily through polar contacts and consequently departed from the PCNA surface during MD refinement. Nine unique docked positions for each system were selected to build triplet structures of modified PCNA. These structures after the MD refinements (sorted by Rosetta score) are presented in Figure 7.3.

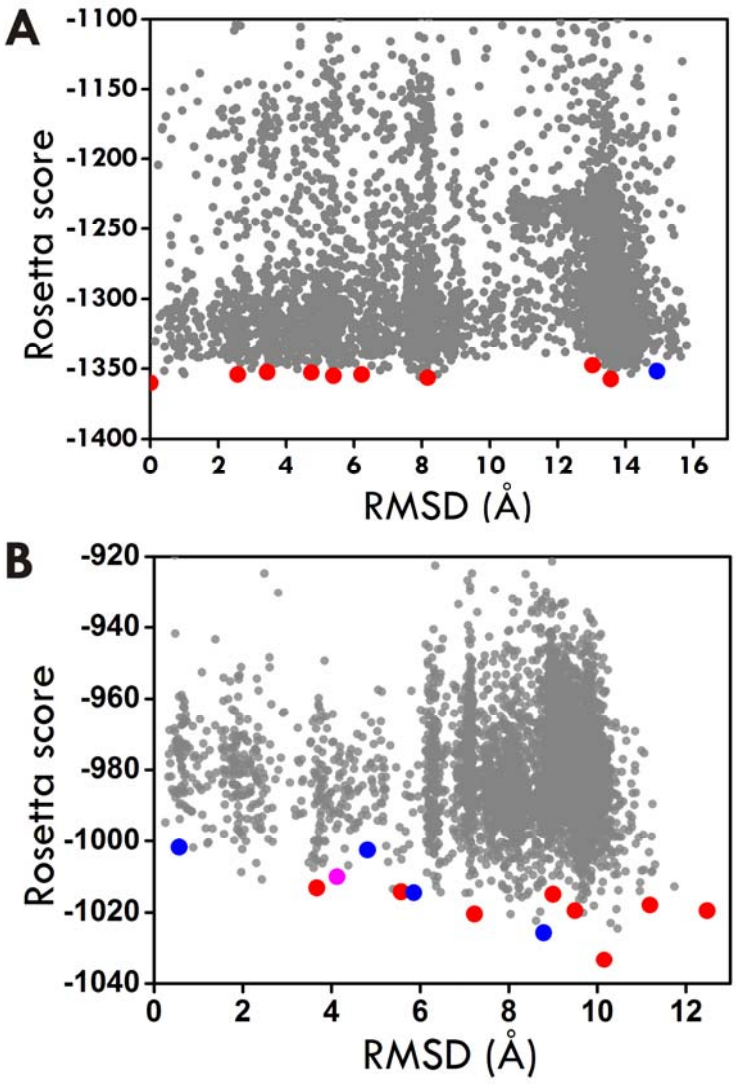

Figure 7.2 Rosetta score versus RMSD plots for ubiquitinated and SUMOylated PCNA

A) Decoys from PCNAK107-Ub docking are shown in gray. Lowest-scoring structurally distinct models (selected for building triplets) are shown red. One model (blue dot) departed from its binding position during MD and was not considered for MES analysis; B) Decoys from PCNAK164-SUMO docking are shown in gray. Lowest-scoring models (selected for building triplets) are shown red. Four models (blue dots) departed during MD and were not considered for

MES analysis. One partially flexible position (purple) was subsequently included in MES analysis.

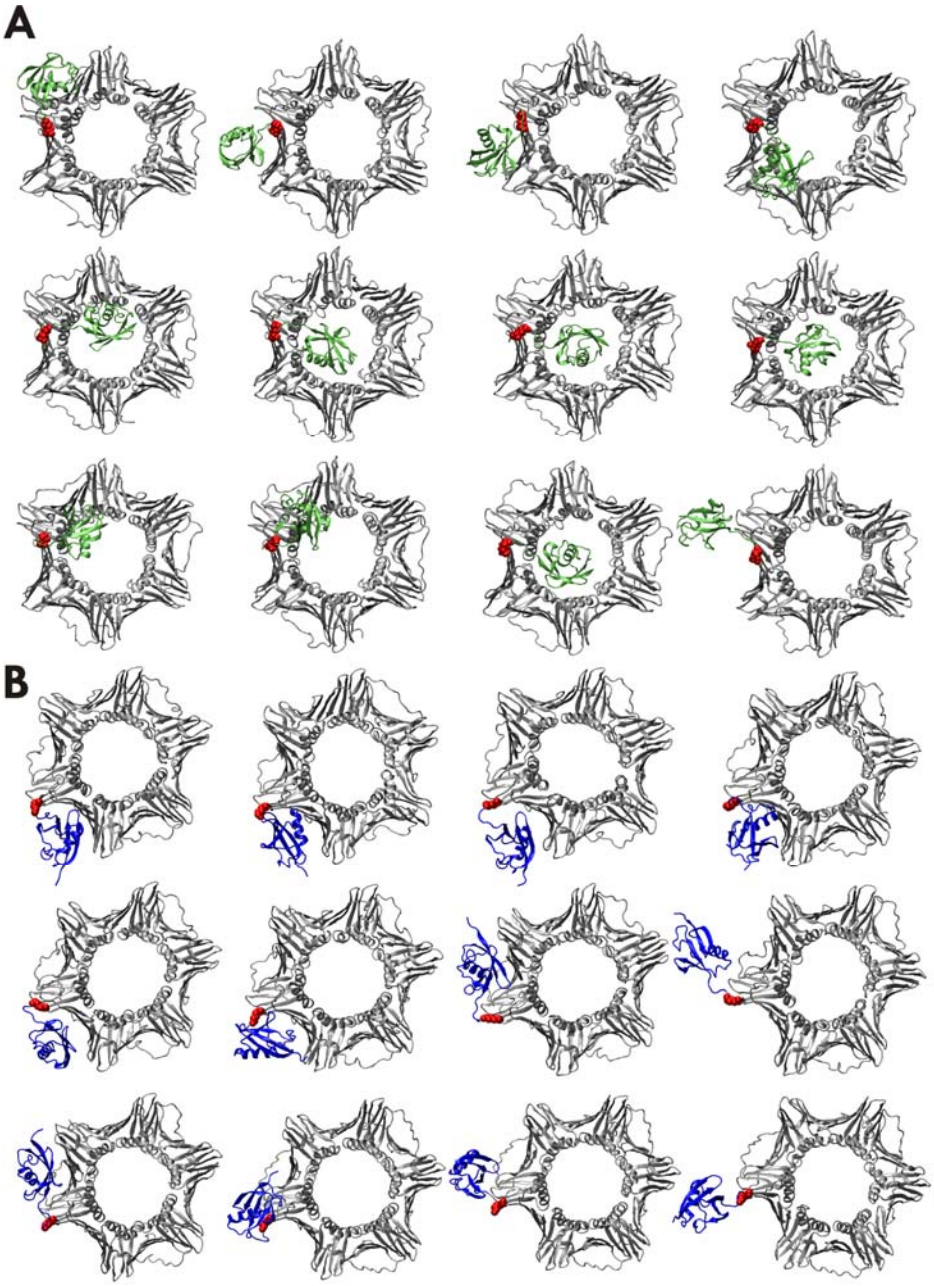

Figure 7.3 Rosetta binding positions used in constructing PCNA triplet structures for MES analysis

A) PCNAK107-Ub and B) PCNAK164-SUMO structures. PCNA, Ub and SUMO are shown in cartoon representation and colored in gray, green and blue, respectively. The K107 and K164 residues of PCNA are depicted as a red spheres.

Optimal positions of Ubiquitin/SUMO on PCNA from Minimal Ensemble Search- To determine which positions were occurring in solution, we generated models with three ubiquitin or three SUMO moieties linked to homotrimeric PCNA for comparison to the experimental SAXS data. Twelve positions for $PCNA_{K107}$-Ub (including 3 detached flexible Ub positions identified by averaging from the MD ensemble) and twelve positions for $PCNA_{K164}$-SUMO (including the 3V60 X-ray structure and 3 detached flexible SUMO positions) were used to build triplet structures for the modified complexes. All possible combinations of positions were generated with three Ub or SUMO moieties per PCNA homotrimer. Thus, we produced a final set of 568 PCNA-Ub triplet models and 1728 PCNA-SUMO triplet models.

Structural information on dynamic assemblies (such as modified PCNA in solution) can be determined by SAXS (235,241,246,247). Fourier transform of SAXS data yields a distribution of electron pair distances P(r). Since the P(r) function encodes all interatomic distance information (even from low-populated flexible conformations) it allows SAXS to sensitively discriminate among structural models. Thus, structural interpretation is predicated on the generation and filtering of an ensemble of atomistic models. To implement this strategy we computed theoretical SAXS profiles for all triplet models of the modified complexes using the program FoXS and fitted these profiles to the experimental scattering data. The goodness-of-fit χ values provide a measure to assess the agreement between the experimental scattering data and each of the structural models. We recalculated the identified ensembles using χ_{free} (248), to give a noise-reduced

assessment of the fit. Computed χ values for $PCNA_{K107}$-Ub and $PCNA_{K164}$-SUMO are plotted in Figure 7.4A, B as a function of Cα RMSD for each conformation. The single best fit to the experimental data was χ=1.80 (χ_{free}=2.32) for $PCNA_{K107}$-Ub and χ=2.55 (χ_{free}=2.97) for $PCNA_{K164}$-SUMO, respectively. However, in a flexible system, multiple conformations will contribute to the SAXS profile. Therefore, a minimal ensemble search (MES) (249) was employed to identify a small subset of conformations that optimally represent the scattering data. In the case of $PCNA_{K107}$-Ub an excellent agreement to the experimental curve was obtained (χ=1.60; χ_{free}=2.10) (Figure 7.4C) by a set of three distinct conformers with ensemble contributions of 45%, 47% and 8% (Figure 7.4G). In these conformers the Ub occupied primarily docked positions, closely associating with the PCNA surface. In the case of $PCNA_{K164}$-SUMO, however, the ensemble was dominated by two flexible, extended conformations (45% and 31% occupancy) and a third flexible but more compact conformation (24% occupancy) (Figure 7.4H). MES notably improved the goodness-of-fit to the PCNA-SUMO SAXS data (χ=2.17, χ_{free}=2.41, Figure 7.4D). Conversely, large discrepancies between the X-ray structure 3V60 and the solution SAXS profile were observed (χ=6.70, χ_{free}=7.65, Figure 7.4B) indicating that the conformation found in the crystal is not prevalent in the solution phase ensemble.

In all models selected by MES we calculated the occupancy of docked versus extended flexible positions for $PCNA_{K107}$-Ub and $PCNA_{K164}$-SUMO. In the PCNA-Ub complex, the ubiquitin was found in the docked position (in the groove above the subunit interface of PCNA) with occupancy of 50%, with 34% occupancy in the back of PCNA along the P-loop and 16% occupancy in the central cavity of PCNA. The possibility for Ub to interact with the helices of the central cavity of PCNA distinguishes the $PCNA_{K107}$-Ub from the $PCNA_{K164}$-Ub complex. The reason is that K107 is positioned on the back face of the clamp close to the subunit interface.

This makes it possible for Ub through its flexible linker to extend further toward the central cavity of the clamp. Ub position blocking the central cavity of PCNA is not expected to be functionally relevant in the situation when the clamp is loaded onto DNA. However, it may reflect the small difference in compactness between the $PCNA_{K164}$-Ub and $PCNA_{K107}$-Ub complexes seen by SAXS. In the case of PCNA-SUMO, extended positions were included in the MES calculation. As shown in Figure 7.4D, MES analysis identifies SUMO with 84% occupancy in two extended flexible positions. Interestingly, MES did not pick any positions docked to the PCNA surface despite inclusion of the 3V60 crystallographic position in the MES optimization. This outcome indicates that SUMO is largely flexible in solution. The docked positions observed by crystallography (230,231) (3PGE and 3V60) are probably stabilized by lattice contacts.

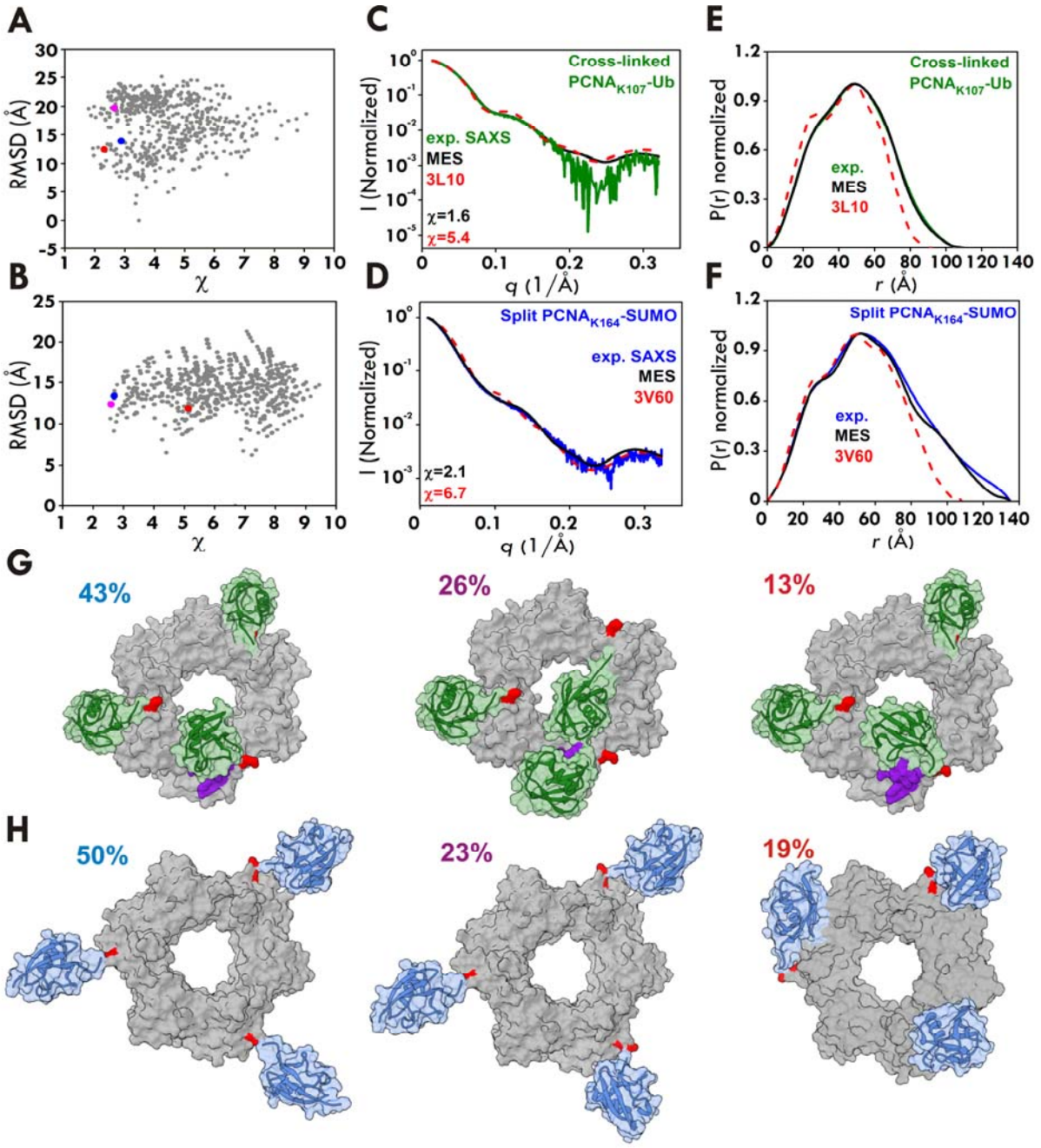

Figure 7.4 Ub primarily adopts docked positions in PCNAK107-Ub while SUMO occupies extended positions in PCNAK164-SUMO

A minimal ensemble search (MES) produces the best fit to the experimental SAXS data for PCNAK107-Ub and PCNAK164-SUMO. A) χ values for the triplet PCNAK107-Ub structures plotted against RMSD. Conformations selected by MES are highlighted in blue, magenta and

red, respectively; B) χ values for the triplet PCNAK164-SUMO structures plotted against RMSD. Conformations selected by MES are highlighted in blue, magenta and red, respectively; C) Overlaid experimental SAXS profile for PCNAK107-Ub (green), computed profile for 3L10 X-ray structure (red dotted line) and computed profile from MES model (black); D) Overlaid experimental SAXS profile for PCNAK164-SUMO (blue), computed profile for 3V60 X-ray structure (red dotted line) and computed profile from MES model (back); E) P(r) functions for PCNAK107-Ub (green), 3L10 X-ray structure (red dotted line) and MES model (black); F) P(r) functions for PCNAK164-SUMO (blue), 3V60 X-ray structure (red dotted line) and MES model (black); G) The three most populated structures from MES analysis of PCNAK107-Ub. (H) The three most populated structures from MES analysis of PCNAK164-SUMO. The P-loop is shown in purple; the K107 and K164 attachment points are depicted in red. PCNA, Ub and SUMO are shown in gray, green and blue, respectively. The MES occupancies for the three conformations are labeled in blue, magenta and red, respectively.

Detailed interactions in the MES identified models- To identify specific structural features that might drive the closer association of Ub to PCNA, we inspected the three most populated MES triplet models. One distinct position selected by MES identified Ub positioned in a groove of the PCNA ring directly above the PCNA subunit interface (Figure 7.4G). The surface of the ubiquitin that interacted with PCNA was the hydrophobic surface centered on Ile44 and Val70 known to associate with a variety of other proteins (250,251). In addition to hydrophobic contacts, we observed several electrostatic and hydrogen bonding interactions: notably, salt bridges $Lys48_{Ub}$-$Glu113_{PCNA}$, $Arg42_{Ub}$-$Glu104_{PCNA}$ and $Glu51_{Ub}$-$Lys168_{PCNA}$. The MES results also imply that a substantial fraction of ubiquitin is found on the back face of the ringed clamp and the

C-terminus linker of Ub closely interacts with the P-loop of PCNA. This is consistent with the split-fusion PCNA-Ub crystal structure, which identified Ub in two positions close to the P-loop. We further verified that Ub positioned along the P-loop did not obstruct duplex DNA passing through the PCNA ring (Figure 7.5). Since PCNA was not loaded onto DNA, ubiquitin in $PCNA_{K107}$-Ub could also be situated in the central cavity of PCNA. Our MES results further suggest that covalently attached SUMO makes few contacts to PCNA and adopts extended flexible positions away from the tips of the PCNA subunits. The crystal structure of $PCNA_{K164}$-SUMO revealed that the interface with the modifier comprises loop regions of PCNA and SUMO, which involve acidic residues from both sides. Furthermore, the outer surface of PCNA is overwhelmingly negatively charged. We verified that these interfacial contacts from the crystal structure could not be maintained in an MD simulation. The reason becomes apparent upon inspection of the surface electrostatic potentials for Ub, SUMO and PCNA (Figure 7.6). For all three dominant docked conformations of $PCNA_{K107}$-Ub we observe electrostatic complementarity at the Ub/PCNA interface. By contrast, there is clear repulsion for the $PCNA_{K164}$-SUMO crystal structure where both surfaces are negatively charged. Thus, it is likely that the docked conformation identified in the X-ray structure is stabilized by the crystal environment and may not be highly populated in solution.

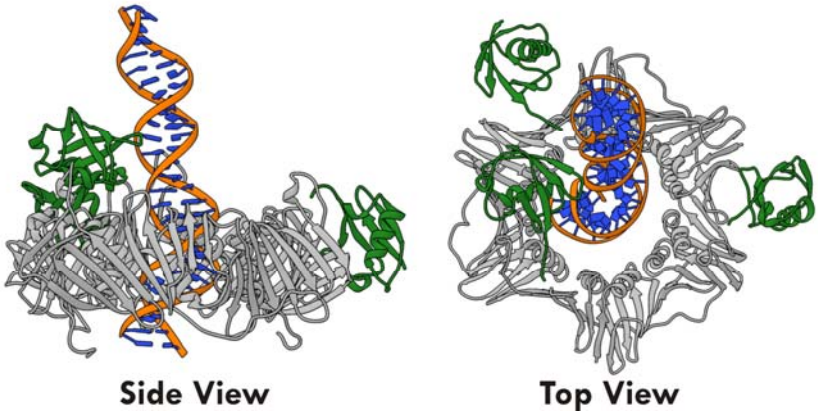

Figure 7.5 Model of PCNAK107-Ub with duplex DNA threaded through the clamp

The most populated structure from MES analysis was used to represent PCNAK107-Ub. The DNA was positioned by overlaying PCNA to the E. coli beta-clamp-DNA structure (PDB ID 3BEP). The complex was then relaxed in a 40ns MD simulation. The structure shown is from the end of the trajectory and demonstrates that Ub at the back of PCNA does not obstruct the passage of DNA through the PCNA ring.

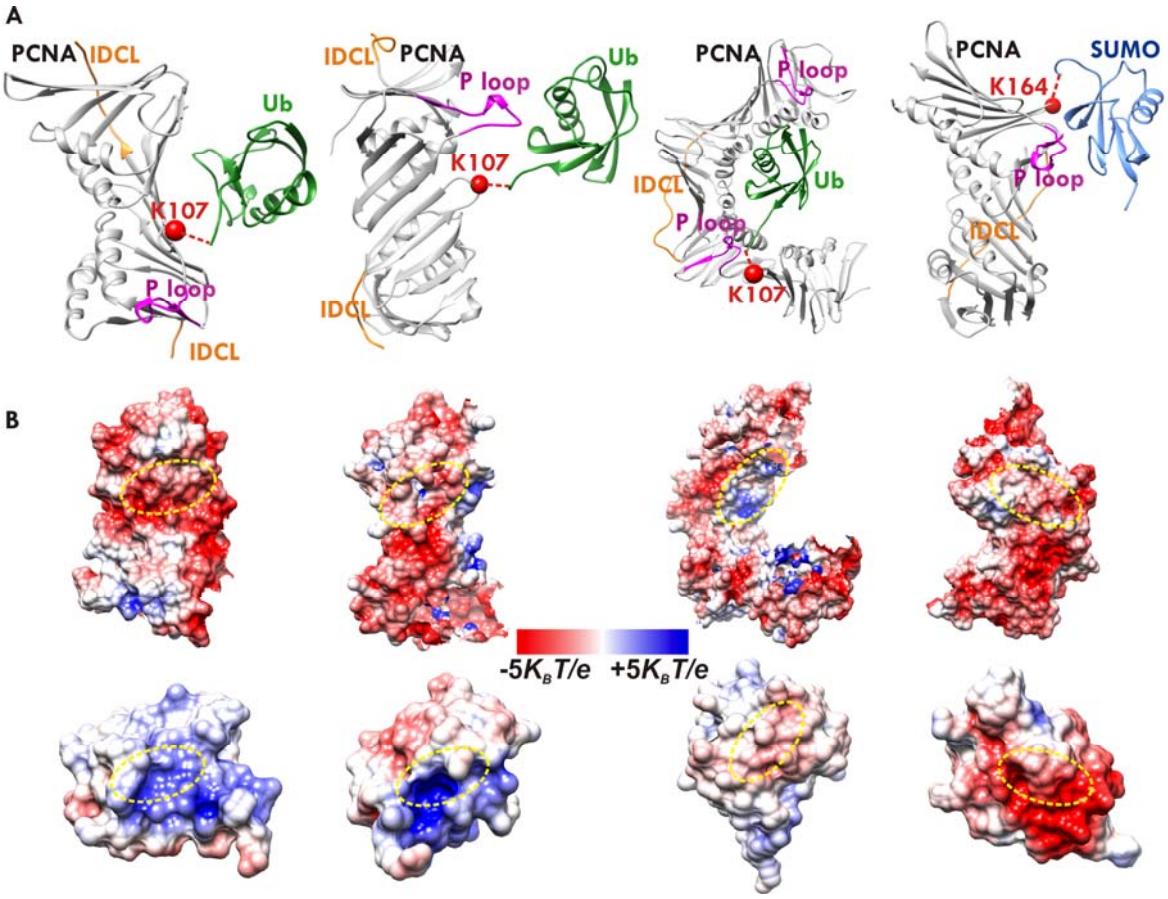

Figure 7.6 Structural differences in PCNAK107-Ub and PCNAK164-SUMO complexes from the anticorrelated electrostatic potential of Ub and SUMO

A) Ub docked onto PCNA in the most populated MES positions (subunit interface groove, P-loop and central cavity positions) and SUMO position from the 3V60 crystal structure. Ub, SUMO and PCNA are colored green, blue and grey, respectively. The IDCL and P-loops on PCNA are shown in orange and magenta, respectively. The attachment positions, K107 and K164 residues, are displayed as red balls; B) Electrostatic potential surfaces corresponding to the bound positions of Ub and SUMO on PCNA (groove, P-loop and central cavity, from left to right for PCNAK1o7-Ub and 3V60 structure for PCNAK164-SUMO). Charges were assigned using PDB2PQR server and electrostatic potentials were calculated using APBS. 100 mM NaCl con-

centration was introduced to screen the electrostatics mimicking physiological conditions. The potential varies from −5KBT/e to +5KBT/e and is depicted from red to blue, respectively.

7.4 Discussion

PCNA monoubiquitination and SUMOylation are reversible dynamic modifications that orchestrate cellular events in response to DNA damage. How Ub and SUMO associate with PCNA and how this association may facilitate subsequent partner interactions has not been fully established. Three general models can be envisioned: (i) simple tethering; (ii) structured interface formation or (iii) allosteric response. Under the simple tethering model, the only function of the modifier (ubiquitin or SUMO) is to provide affinity to the protein partner carrying ubiquitin- or SUMO-interacting domain. Under the structured interface model, the modifier participates with PCNA in the assembly of higher order complexes with downstream effector proteins. In this case, affinity to PCNA augments the affinity for the ubiquitin- or SUMO-interacting partner. In the allosteric model, Ub or SUMO conjugation induces conformational change in PCNA, which in turn facilitates recruitment of downstream effectors in the respective pathways. This allosteric hypothesis is exemplified by the case of thymine DNA glycosylase (TGD) wherein product release by the enzyme involves large conformational change assisted by SUMO modification (252,253).

Our results argue strongly against an allosteric model for the role of PCNA ubiquitination or SUMOylation in effecting distinct functional responses to DNA damage. Multiple trajectory MD simulations fail to observe any substantial changes in the conformation of PCNA in the modified complexes. In the ubiquitinated complexes the modifier appears to be stabilized against the PCNA surface and is able to adopt several discrete positions relative to the PCNA ring. From MES the most populated position in $PCNA_{K107}$-Ub was the sideways position above the subunit

interface of the PCNA ring. A similar orientation was found as predominant also for the $PCNA_{K164}$-Ub complex. Ub can also reside at the back of the PCNA ring close to the P-loop in both the K107 and K164 ubiquitinated complexes. Finally, for K107 a small population of Ub can interact with the helices of the central hole of PCNA. To describe this situation, we use the concept of segmental flexibility and define it as the stabilization of "functionally-relevant" positions in otherwise flexible systems. Both $PCNA_{K107}$-Ub and $PCNA_{K164}$-Ub exhibit such segmental flexibility and feature primarily docked positions for the Ub modifier. Thus, for the Ub modified PCNA complexes our results support the structured interface mechanism for partner protein recruitment. Furthermore, segmental flexibility is likely a common characteristic of eukaryotic ubiquitin regulatory systems.

By contrast, the $PCNA_{K164}$-SUMO complex adopted extended flexible conformations wherein SUMO was attached to K164 but was not otherwise interacting with the surface of PCNA. Therefore, for SUMOylated PCNA recruitment the results support the simple tethering model, consistent with the model of the SRS2 interaction with SUMOylated PCNA that adopted two positions (230). The structurally distinct modes of association of Ub and SUMO to the sliding clamp rationalize the ability of the modifiers to promote distinct functional responses even when the PTMs occur at the exact same position on PCNA. Notably, the inherently dynamic PTM positioning may not be easily construed from static crystal structures available for PCNA-SUMO or PCNA-Ub suggesting combined computational-experimental approaches may be critical to accurate structural analyses.

In this respect, our work on modified PCNA underscores the power of combining hybrid computational models with experimental SAXS data to gain structural understanding of protein complexes in solution (254). Segmental flexibility is often necessary for the function of protein

machines (143,255-257). In crystals, conformations can be blocked or formed by crystal contacts (230,231) or influenced by crystallization conditions, thus, restricting our knowledge of intrinsically flexible systems to the few available X-ray structures. Therefore, generation of conformational ensembles for the complexes by hybrid methods (starting from high-resolution structures of their individual components) and validating the models by SAXS reveals a more accurate and complete representation of the diverse conformational states present in solution. In this context, the results afford insights into how PCNA post-translational modifications by Ub and SUMO provide both the necessary specificity and flexibility to regulate recruitment and coordinated actions of effector proteins to promote genomic stability.

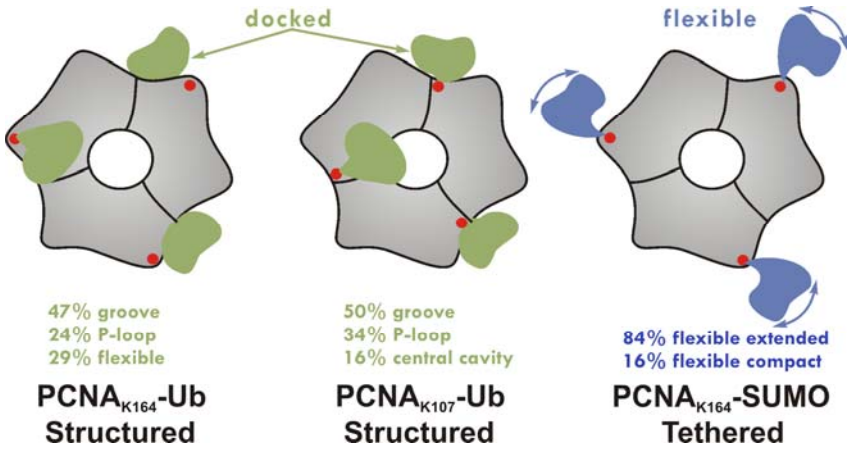

Figure 7.7 Biological implications of the different mode of association of ubiquitin and SUMO to PCNA

The PCNA trimer is shown in gray; the Ub modifier in green; SUMO in blue; the attachment positions are indicated by a red dots; curved arrows indicate flexible attachment. Approximate occupancy (%) of the identified distinct positions of the modifiers on PCNA a given below each model.

7.5 Materials and Methods

SAXS analysis of $PCNA_{K164}$-Sumo and $PCNA_{K107}$-Ub- SUMOylated and cross-linked yeast PCNA was purified as before (231,239,258,259). SAXS data were collected at the SIBYLS 12.3.1 beamline at the Advanced Light Source, LBNL (238,260,261). Scattering measurements were performed on 20 l samples at 15 °C loaded into a helium-purged sample chamber, 1.5 m from the Mar165 detector. Prior to data collection, modified PCNA were purified by size exclusion chromatography on a 24 mL Superose6 column equilibrated in 20 mM Tris pH 7.5, 150 mM NaCl, 5% glycerol. Data were collected on both the original gel filtration fractions and samples concentrated ~2x-8x from individual fractions. Fractions prior to the void volume and concentrator eluates were used for buffer subtraction. Sequential exposures (0.5, 0.5, 5, and 0.5 s for $PCNA_{K164}$-SUMO and 0.5, 0,5, 2, 5, 0.5 s for $PCNA_{K107}$-Ub) were taken at 12 keV. Although scattering of the split-fusion $PCNA_{K164}$-SUMO showed radiation-induced aggregation, the scattering of the cross-linked $PCNA_{K107}$-Ub showed a decrease in slope, indicative that the molecules in solution were becoming smaller. It is likely due to irradiation breaking the disulfide bond. The first and second 0.5 second exposures overlaid, so the first exposure was assumed to have only minimal damage. The best data, based on signal-to-noise and Guinier, were collected on split-fusion $PCNA_{K164}$-Sumo (2.7 mg/ml, 0.5 sec exposure) and cross-linked $PCNA_{K107}$-Ub (4.1 mg/ml, 0.5 sec exposure). Data was analyzed using PRIMUS. The Porod exponent was determined from a linear regression analysis (I vs Q) of the top of the first peak in the Porod-Debye plot (q^4*I(q) vs q^4) of the scattering data, implemented in SCÅTTER, available at beamline 12.3.1(236,261). X^2_{free} of selected models was calculated using Scatter (248).

Computational models and protocols- To model the $PCNA_{K107}$-Ub and $PCNA_{K164}$-SUMO complexes, we employed a chemically conjugated docking protocol written as a part of the Ro-

setta 3.4 suite (242-244). We used the standard Rosetta scoring function − score12 to rank and select the top-scoring models (245,262). The protocol was designed to search the conformational space available to ubiquitin chemically conjugated via an isopeptide bond and samples rotations about torsional angles in the vicinity of the isopeptide bond. To initiate the sampling for $PCNA_{K107}$-Ub we used the 1UBQ structure for ubiquitin (263) and 1PLQ structure for yeast PCNA (22). For the $PCNA_{K164}$-SUMO complex, we used the structure of SUMOylated PCNA (PDB ID: 3V60) (230). The initial PDB structures were minimized with the Rosetta relax protocol while employing all-heavy-atom constraints prior to conjugated docking. For the isopeptide linker, protocol UBQ_Gp_LYX-cterm was used. Torsions allowed to change included: the chi angles of Lys107 or Lys164 of PCNA, the isopeptide bond and both phi and psi for the Gly76, Gly75 and Arg74 of ubiquitin (Gly98, Gly97 and Ile96 of SUMO). Sampling was performed with a standard Rosetta Metropolis Monte Carlo search protocol (245). Results were automatically filtered according to the solvent accessible surface area (SASA) buried at the protein interface (> 500 $Å^2$) and total score of the docked complex (< 0). In total the sampling produced 4,791 decoys for PCNA-Ub and 4,499 decoys for PCNA-SUMO using 20,000 Monte Carlo cycles per trajectory. All Rosetta docking calculations were performed on the Stampede supercomputer at the Texas Advanced Computing Center (TACC). To identify low-scoring outliers we relied on Rosetta score vs. RMSD plots. In this case RMSD is defined in reference to an arbitrary structure and is merely a way to spread the points out and make groups of points that stand out clearly from the cloud. Selected low-scoring decoys were used for further MD refinement.

Model refinement- Outliers with best scores from Rosetta docking were refined using all-atom explicit solvent molecular dynamics. In setting up for MD hydrogen atoms, counterions (Na^+) and TIP3P solvent (264) were introduced using the XLeap module in the AMBER 9

(115,265,266). Additionally, 100 mM NaCl concentration was introduced to mimic physiological conditions. The systems were then minimized for 5000 steps with backbone atoms fixed followed by 5000 steps of minimization with harmonic restraints to remove unfavorable contacts. The systems were then gradually brought up to 300 K and run for 50 ps in the NVT ensemble while keeping the protein backbone restrained. The equilibration was continued for another 2 ns in the NPT ensemble and the harmonic restraints were gradually released. The 60 ns production simulations were performed in the NPT ensemble (1 atm and 300 K) without constraints. A short-range cutoff of 10 Å was used for the short-range non-bonded interactions with a switching function at 8.5 Å. The long-range electrostatic interactions were treated with a smooth particle mesh Ewald method (117). An integration time step of 2 fs was used under a multiple time stepping scheme. The r-RESPA multiple timestep method (118) was adopted with a 2 fs time step for bonded, 2 fs for short-range non-bonded interactions and 4 fs for long-range electrostatic interactions. Bonds between hydrogen atoms and heavy atoms of the protein were constrained with SHAKE. All simulations were performed using the NAMD 2.8 code (73,74) using the AMBER Parm99SB force field. Models, which departed substantially from the initial Rosetta docking position during MD were eliminated from further consideration.

Minimal Ensemble Search- Twelve positions for $PCNA_{K107}$-Ub (including 3 detached flexible Ub positions identified by averaging from the MD ensemble) and twelve positions for $PCNA_{K164}$-SUMO (including the 3V60 X-ray structure and 3 detached flexible SUMO positions) were used to build triplet structures for the modified complexes. All possible combinations of positions were generated with three Ub or SUMO moieties per PCNA homotrimer. Thus, we produced a final set of 568 PCNA-Ub triplet models and 1728 PCNA-SUMO triplet models. Computation of scattering profiles from the models and comparison to the experimental data

used the FoXS code. FoXS (247) is a method for computing a theoretical scattering profile of a structure and fitting to the experimental profile. For a given structure the SAXS profile is computed based on the Debye formula:

$$I(q) = \sum_{i=1}^{N} \sum_{j=1}^{N} f_i(q) f_j(q) \frac{\sin(qd_{ij})}{qd_{ij}} \quad (7.1)$$

where the intensity, I(q), is a function of the momentum transfer, $q = \frac{4\pi \sin\theta}{\lambda}$; 2θ is the scattering angle and λ is the wavelength of the incident X-ray beam; $f_i(q)$ is the form factor of an atom i, d_{ij} is the distance between atoms i and j, and N is the number of atoms in the system. The SAXS profile computed from the structure is fitted to the experimental SAXS profile by minimizing χ:

$$\chi = \sqrt{\frac{1}{M} \sum_{i=1}^{M} \left(\frac{I_{exp}(q_i) - cI(q_i)}{\sigma(q_i)}\right)^2} \quad (7.2)$$

where $I_{exp}(q_i)$ and $I(q_i)$ are the experimental and computed profiles, $\sigma(q_i)$ is the error of the experimental profile, M is the number of points in the profile, and c is the scaling factor.

The coexistence of different conformations that contribute to the experimental scattering curve has to be taken into account by considering the flexibility of different Ub or SUMO positions. An algorithm developed by Pelikan *et al.* (249) was used to search for the minimal ensemble (MES) of conformations from the pool of all Rosetta generated triplet models. The multi-conformational scattering I(q) from such a minimal ensemble is computed by averaging the individual scattering patterns from the conformers. The final model was fitted to the experimental curve I(q) minimizing χ between the experimental and calculated multi-conformational curve according to equation 7.2.

REFERENCES

1. Kornberg, A., and Baker, T. A. (1992) *DNA replication*, 2nd ed., W.H. Freeman, New York
2. Lodish, H. F., Matsudaira, P. T., Kaiser, C., and Krieger, M. (2004) *Molecular cell biology*, 5th ed., W.H. Freeman and Company, New York
3. Johnson, A., and O'Donnell, M. (2005) Cellular DNA replicases: Components and dynamics at the replication fork. *Annual Review of Biochemistry* **74**, 283-315
4. McHenry, C. S. (2011) DNA Replicases from a Bacterial Perspective. *Annu Rev Biochem* **80**, 403-436
5. Sinha, N. K., Morris, C. F., and Alberts, B. M. (1980) Efficient in vitro replication of double-stranded DNA templates by a purified T4 bacteriophage replication system. *The Journal of biological chemistry* **255**, 4290-4293
6. Indiani, C., and O'Donnell, M. (2006) The replication clamp-loading machine at work in the three domains of life. *Nat Rev Mol Cell Biol* **7**, 751-761
7. McHenry, C. S., and Crow, W. (1979) DNA polymerase III of Escherichia coli. Purification and identification of subunits. *The Journal of biological chemistry* **254**, 1748-1753
8. Studwell-Vaughan, P. S., and O'Donnell, M. (1993) DNA polymerase III accessory proteins. V. Theta encoded by holE. *The Journal of biological chemistry* **268**, 11785-11791
9. Maki, H., and Kornberg, A. (1985) The polymerase subunit of DNA polymerase III of Escherichia coli. II. Purification of the alpha subunit, devoid of nuclease activities. *The Journal of biological chemistry* **260**, 12987-12992
10. Taft-Benz, S. A., and Schaaper, R. M. (2004) The theta subunit of Escherichia coli DNA polymerase III: a role in stabilizing the epsilon proofreading subunit. *J Bacteriol* **186**, 2774-2780
11. Bloom, L. B., Chen, X., Fygenson, D. K., Turner, J., O'Donnell, M., and Goodman, M. F. (1997) Fidelity of Escherichia coli DNA polymerase III holoenzyme. The effects of beta, gamma complex

processivity proteins and epsilon proofreading exonuclease on nucleotide misincorporation efficiencies. *The Journal of biological chemistry* **272**, 27919-27930

12. Hamdan, S. M., and Richardson, C. C. (2009) Motors, Switches, and Contacts in the Replisome. *Annual Review of Biochemistry* **78**, 205-243
13. Perumal, S. K., Yue, H. J., Hu, Z. X., Spiering, M. M., and Benkovic, S. J. (2010) Single-molecule studies of DNA replisome function. *Bba-Proteins Proteom* **1804**, 1094-1112
14. Lia, G., Michel, B., and Allemand, J. F. (2012) Polymerase Exchange During Okazaki Fragment Synthesis Observed in Living Cells. *Science* **335**, 328-331
15. Lia, G., Michel, B., and Allemand, J. F. (2012) Polymerase exchange during Okazaki fragment synthesis observed in living cells. *Science* **335**, 328-331
16. Kunkel, T. A., and Burgers, P. M. (2008) Dividing the workload at a eukaryotic replication fork. *Trends Cell Biol* **18**, 521-527
17. Moyer, S. E., Lewis, P. W., and Botchan, M. R. (2006) Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 10236-10241
18. Kurth, I., and O'Donnell, M. (2013) New insights into replisome fluidity during chromosome replication. *Trends in biochemical sciences* **38**, 195-203
19. DePamphilis, M. L., and Wassarman, P. M. (1980) Replication of eukaryotic chromosomes: a close-up of the replication fork. *Annu Rev Biochem* **49**, 627-666
20. 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
21. Kontopidis, G., Wu, S. Y., Zheleva, D. I., Taylor, P., McInnes, C., Lane, D. P., Fischer, P. M., and Walkinshaw, M. D. (2005) Structural and biochemical studies of human proliferating cell nuclear

antigen complexes provide a rationale for cyclin association and inhibitor design. *Proc Natl Acad Sci U S A* **102**, 1871-1876

22. 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

23. Moarefi, I., Jeruzalmi, D., Turner, J., O'Donnell, M., and Kuriyan, J. (2000) Crystal structure of the DNA polymerase processivity factor of T4 bacteriophage. *Journal of molecular biology* **296**, 1215-1223

24. Hopfner, K. P., Eichinger, A., Engh, R. A., Laue, F., Ankenbauer, W., Huber, R., and Angerer, B. (1999) Crystal structure of a thermostable type B DNA polymerase from Thermococcus gorgonarius. *Proc Natl Acad Sci U S A* **96**, 3600-3605

25. Hashimoto, H., Nishioka, M., Fujiwara, S., Takagi, M., Imanaka, T., Inoue, T., and Kai, Y. (2001) Crystal structure of DNA polymerase from hyperthermophilic archaeon Pyrococcus kodakaraensis KOD1. *Journal of molecular biology* **306**, 469-477

26. Kim, S. W., Kim, D. U., Kim, J. K., Kang, L. W., and Cho, H. S. (2008) Crystal structure of Pfu, the high fidelity DNA polymerase from Pyrococcus furiosus. *International journal of biological macromolecules* **42**, 356-361

27. Williams, G. J., Johnson, K., Rudolf, J., McMahon, S. A., Carter, L., Oke, M., Liu, H., Taylor, G. L., White, M. F., and Naismith, J. H. (2006) Structure of the heterotrimeric PCNA from Sulfolobus solfataricus. *Acta crystallographica. Section F, Structural biology and crystallization communications* **62**, 944-948

28. Dore, A. S., Kilkenny, M. L., Rzechorzek, N. J., and Pearl, L. H. (2009) Crystal structure of the rad9-rad1-hus1 DNA damage checkpoint complex--implications for clamp loading and regulation. *Mol Cell* **34**, 735-745

29. 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

30. 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

31. Warbrick, E. (1998) PCNA binding through a conserved motif. *Bioessays* **20**, 195-199

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

33. Dalrymple, B. P., Kongsuwan, K., Wijffels, G., Dixon, N. E., and Jennings, P. A. (2001) A universal protein-protein interaction motif in the eubacterial DNA replication and repair systems. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 11627-11632

34. Lopez de Saro, F. J., Georgescu, R. E., Goodman, M. F., and O'Donnell, M. (2003) Competitive processivity-clamp usage by DNA polymerases during DNA replication and repair. *The EMBO journal* **22**, 6408-6418

35. Jonsson, Z. O., and Hubscher, U. (1997) Proliferating cell nuclear antigen: more than a clamp for DNA polymerases. *Bioessays* **19**, 967-975

36. Maga, G., and Hubscher, U. (2003) Proliferating cell nuclear antigen (PCNA): a dancer with many partners. *J Cell Sci* **116**, 3051-3060

37. Kelman, Z. (1997) PCNA: structure, functions and interactions. *PCNA Oncogene* **14**, 629-640

38. Moldovan, G. L., Pfander, B., and Jentsch, S. (2007) PCNA, the maestro of the replication fork. *Cell* **129**, 665-679

39. Khair, L., Chang, Y. T., Subramanian, L., Russell, P., and Nakamura, T. M. (2010) Roles of the checkpoint sensor clamp Rad9-Rad1-Hus1 (911)-complex and the clamp loaders Rad17-RFC and Ctf18-RFC in Schizosaccharomyces pombe telomere maintenance. *Cell Cycle* **9**, 2237-2248

40. Shiomi, Y., Shinozaki, A., Nakada, D., Sugimoto, K., Usukura, J., Obuse, C., and Tsurimoto, T. (2002) Clamp and clamp loader structures of the human checkpoint protein complexes, Rad9-1-1 and Rad17-RFC. *Genes to cells : devoted to molecular & cellular mechanisms* **7**, 861-868

41. 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

42. Neuwald, A. F., Aravind, L., Spouge, J. L., and Koonin, E. V. (1999) AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome research* **9**, 27-43

43. Erzberger, J. P., and Berger, J. M. (2006) Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annual review of biophysics and biomolecular structure* **35**, 93-114

44. Hanson, P. I., and Whiteheart, S. W. (2005) AAA+ proteins: Have engine, will work. *Nat Rev Mol Cell Bio* **6**, 519-529

45. O'Donnell, M., Jeruzalmi, D., and Kuriyan, J. (2001) Clamp loader structure predicts the architecture of DNA polymerase III holoenzyme and RFC. *Current biology : CB* **11**, R935-946

46. 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. *The Journal of biological chemistry* **276**, 34776-34783

47. 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. *The Journal of biological chemistry* **276**, 34784-34791

48. 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

49. Kelch, B. A., Makino, D. L., O'Donnell, M., and Kuriyan, J. (2011) How a DNA Polymerase Clamp Loader Opens a Sliding Clamp. *Science* **334**, 1675-1680

50. McCammon, J. A., Gelin, B. R., and Karplus, M. (1977) Dynamics of folded proteins. *Nature* **267**, 585-590

51. Cheatham, T. E., 3rd, and Kollman, P. A. (2000) Molecular dynamics simulation of nucleic acids. *Annual review of physical chemistry* **51**, 435-471

52. Karplus, M., and McCammon, J. A. (2002) Molecular dynamics simulations of biomolecules. *Nature structural biology* **9**, 646-652

53. Moraitakis, G., Purkiss, A. G., and Goodfellow, J. M. (2003) Simulated dynamics and biological macromolecules. *Rep Prog Phys* **66**, 383-406

54. Adcock, S. A., and McCammon, J. A. (2006) Molecular dynamics: survey of methods for simulating the activity of proteins. *Chemical reviews* **106**, 1589-1615

55. Leach, A. R. (2001) *Molecular modelling : principles and applications*, 2nd ed., Prentice Hall, Harlow, England ; New York

56. MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B* **102**, 3586-3616

57. Mackerell, A. D., Wiorkiewiczkuczera, J., and Karplus, M. (1995) An All-Atom Empirical Energy Function for the Simulation of Nucleic-Acids. *Journal of the American Chemical Society* **117**, 11946-11975

58. Feller, S. E., and MacKerell, A. D. (2000) An improved empirical potential energy function for molecular simulations of phospholipids. *Journal of Physical Chemistry B* **104**, 7510-7515

59. Kuttel, M., Brady, J. W., and Naidoo, K. J. (2002) Carbohydrate solution simulations: Producing a force field with experimentally consistent primary alcohol rotational frequencies and populations. *J Comput Chem* **23**, 1236-1243

60. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1996) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules (vol 117, pg 5179, 1995). *Journal of the American Chemical Society* **118**, 2309-2309

61. Pérez, A., Marchán, I., Svozil, D., Sponer, J., Cheatham, T. E., Laughton, C. A., and Orozco, M. (2007) Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of ±/³ Conformers. *Biophysical journal* **92**, 3817-3829

62. Schlick, T., Barth, E., and Mandziuk, M. (1997) Biomolecular dynamics at long timesteps: Bridging the timescale gap between simulation and experimentation. *Annual review of biophysics and biomolecular structure* **26**, 181-222

63. Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **23**, 327-341

64. van Gunsteren, W. F., and Berendsen, H. J. C. (1977) Algorithms for macromolecular dynamics and constraint dynamics. *Molecular Physics* **34**, 1311-1327

65. Verlet, L. (1967) Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review* **159**, 98-103

66. Hockney, R. W. (1970) Potential calculation and some applications. *Methods Comput. Phys.*

67. Steinbach, P. J., and Brooks, B. R. (1994) New Spherical-Cutoff Methods for Long-Range Forces in Macromolecular Simulation. *J Comput Chem* **15**, 667-683

68. Frenkel, D., and Smit, B. (2002) *Understanding molecular simulation : from algorithms to applications*, 2nd ed., Academic, San Diego, Calif. ; London

69. Brooks, C. L., Pettitt, B. M., and Karplus, M. (1985) Structural and energetic effects of truncating long ranged interactions in ionic and polar fluids. *Journal of Chemical Physics* **83**, 5897

70. Ewald, P. P. (1921) Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen der Physik* **369**, 253-287

71. Sagui, C., and Darden, T. A. (1999) Molecular dynamics simulations of biomolecules: long-range electrostatic effects. *Annual review of biophysics and biomolecular structure* **28**, 155-179

72. Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **98**, 10089-10092

73. Kale, L., Skeel, R., Bhandarkar, M., Brunner, R., Gursoy, A., Krawetz, N., Phillips, J., Shinozaki, A., Varadarajan, K., and Schulten, K. (1999) NAMD2: Greater scalability for parallel molecular dynamics. *Journal of Computational Physics* **151**, 283-312

74. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, and Schulten K. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* **26**

75. Patel, S., and Brooks, C. L. (2004) CHARMM fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations. *J Comput Chem* **25**, 1-16

76. Patel, S., Mackerell, A. D., Jr., and Brooks, C. L., 3rd. (2004) CHARMM fluctuating charge force field for proteins: II protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. *J Comput Chem* **25**, 1504-1514

77. Dang, L. X. (1998) Importance of Polarization Effects in Modeling the Hydrogen Bond in Water Using Classical Molecular Dynamics Techniques. *The Journal of Physical Chemistry B* **102**, 620-624

78. Cramer, C. J., and Truhlar, D. G. (1999) Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chemical reviews* **99**, 2161-2200

79. Simonson, T. (2001) Macromolecular electrostatics: continuum models and their growing pains. *Curr Opin Struct Biol* **11**, 243-252

80. Levy, R. M., and Gallicchio, E. (1998) Computer simulations with explicit solvent: recent progress in the thermodynamic decomposition of free energies and in modeling electrostatic effects. *Annual review of physical chemistry* **49**, 531-567

81. Jorgensen, W. L., Chandrasekhar, J., Madura1, J. D., Impey, W. L., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water *Journal of Chemical Physics* **79**, 10

82. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **81**, 3684-3690

83. Hunenberger, P. (2005) Thermostat algorithms for molecular dynamics simulations. *Adv Polym Sci* **173**, 105-147

84. Andersen, H. C. (1980) Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics* **72**, 2384-2393

85. Grest, G. S., and Kremer, K. (1986) Molecular dynamics simulation for polymers in the presence of a heat bath. *Physical review. A* **33**, 3628-3631

86. Felderhof, B. U. (1978) On the derivation of the fluctuation-dissipation theorem. *Journal of Physics A: Mathematical and General* **11**, 921

87. Nosé, S. (1984) A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics* **81**, 511-519

88. Hoover, W. G. (1985) Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A* **31**, 1695-1697

89. Ashkin, A., and Dziedzic, J. M. (1987) Optical trapping and manipulation of viruses and bacteria. *Science* **235**, 1517-1520

90. Binnig, G., Quate, C. F., and Gerber, C. (1986) Atomic Force Microscope. *Physical Review Letters* **56**, 930-933

91. Evans, E., Ritchie, K., and Merkel, R. (1995) Sensitive Force Technique to Probe Molecular Adhesion and Structural Linkages at Biological Interfaces. *Biophysical journal* **68**, 2580-2587

92. Evans, E. (2001) Probing the relation between force - Lifetime - and chemistry in single molecular bonds. *Annual review of biophysics and biomolecular structure* **30**, 105-128

93. Park, S., Khalili-Araghi, F., Tajkhorshid, E., and Schulten, K. (2003) Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *Journal of Chemical Physics* **119**, 3559-3566

94. Park, S., and Schulten, K. (2004) Calculating potentials of mean force from steered molecular dynamics simulations. *Journal of Chemical Physics* **120**, 5946-5961

95. Karaca, E., and Bonvin, A. M. (2013) Advances in integrative modeling of biomolecular complexes. *Methods* **59**, 372-381

96. Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D. A., and Cheatham, T. E., 3rd. (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Accounts of chemical research* **33**, 889-897

97. Kollman, P. (1993) Free-Energy Calculations - Applications to Chemical and Biochemical Phenomena. *Chemical reviews* **93**, 2395-2417

98. David, C., and Jacobs, D. (2014) Principal Component Analysis: A Method for Determining the Essential Dynamics of Proteins. in *Protein Dynamics* (Livesay, D. R. ed.), Humana Press. pp 193-226

99. Tirion, M. M. (1996) Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Phys Rev Lett* **77**, 1905-1908

100. 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. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 13795-13800

101. Dionne, I., Brown, N. J., Woodgate, R., and Bell, S. D. (2008) On the mechanism of loading the PCNA sliding clamp by RFC. *Molecular Microbiology* **68**, 216-222

102. Seybert, A., Singleton, M. R., Cook, N., Hall, D. R., and Wigley, D. B. (2006) Communication between subunits within an archaeal clamp-loader complex. *Embo J.* **25**, 2209-2218

103. Bowman, G. D., Goedken, E. R., Kazmirski, S. L., O'Donnell, M., and Kuriyan, J. (2005) DNA polymerase clamp loaders and DNA recognition. *FEBS Letters* **579**, 863-867

104. Duderstadt, K. E., and Berger, J. M. (2008) AAA plus ATPases in the initiation of DNA replication. *Critical Reviews in Biochemistry and Molecular Biology* **43**, 163-187

105. O'Donnell, M., and Kuriyan, J. (2006) Clamp loaders and replication initiation. *Current Opinion in Structural Biology* **16**, 35-41

106. 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. *The Journal of biological chemistry* **276**, 34768-34775

107. 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. *Journal of molecular biology* **388**, 431-442

108. Majka, J., and Burgers, P. M. (2004) The PCNA-RFC families of DNA clamps and clamp loaders. *Progress in nucleic acid research and molecular biology* **78**, 227-260

109. Sakato, M., Zhou, Y., and Hingorani, M. M. (2012) ATP binding and hydrolysis-driven rate-determining events in the RFC-catalyzed PCNA clamp loading reaction. *Journal of molecular biology* **416**, 176-191

110. Zhuang, Z. H., Yoder, B. L., Burgers, P. M. J., and Benkovic, S. J. (2006) The structure of a ring-opened proliferating cell nuclear antigen-replication factor C complex revealed by fluorescence energy transfer. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 2546-2551

111. Kumar, R., Nashine, V. C., Mishra, P. P., Benkovic, S. J., and Lee, T. H. (2010) Stepwise loading of yeast clamp revealed by ensemble and single-molecule studies. *Proc Natl Acad Sci U S A* **107**, 19736-19741

112. Tainer, J. A., McCammon, J. A., and Ivanov, I. (2010) Recognition of the ring-opened state of proliferating cell nuclear antigen by replication factor C promotes eukaryotic clamp-loading. *Journal of the American Chemical Society* **132**, 7372-7378

113. McNally, R., Bowman, G. D., Goedken, E. R., O'Donnell, M., and Kuriyan, J. (2010) Analysis of the role of PCNA-DNA contacts during clamp loading. *BMC structural biology* **10**, 3

114. Liu, C. Y., Qureshi, M. T., and Lee, T. H. (2011) Interaction strengths between the ribosome and tRNA at various steps of translocation. *Biophysical journal* **100**, 2201-2208

115. Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) The Amber biomolecular simulation programs. *J Comput Chem* **26**, 1668-1688

116. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**

117. Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995) A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577-8593

118. Tuckerman, M., Berne, B. J., and Martyna, G. J. (1992) Reversible Multiple Time Scale Molecular-Dynamics. *Journal of Chemical Physics* **97**, 1990-2001

119. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins* **65**, 712-725

120. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics. *J Mol Graphics* **14**, 33-&

121. Isralewitz B, Baudry J, Gullingsrud J, Kosztin D, and Schulten K. (2001) Steered molecular dynamics investigations of protein function. *J Mol Graph Model.* **19**, 13-25.

122. Isralewitz, B., Gao, M., and Schulten, K. (2001) Steered molecular dynamics and mechanical functions of proteins. *Current Opinion in Structural Biology* **11**, 224-230

123. Izrailev S, Stepaniants S, Isralewitz B, Kosztin D, Lu H, Molnar F, Wriggers W, and Schulten K. (1998) Steered molecular dynamics. in *Lecture Notes in Computational Science and Engineering.* (Deuflhard P, Hermans J, Leimkuhler B, Mark AE, Reich S, and Skeel RD eds.), Springer-Verlag, Berlin. pp 39-65.

124. Nome, R. A., Zhao, J. M., Hoff, W. D., and Scherer, N. F. (2007) Axis-dependent anisotropy in protein unfolding from integrated nonequilibrium single-molecule experiments, analysis, and simulation. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 20799-20804

125. Barker, S. C., Kassel, D. B., Weigl, D., Huang, X., Luther, M. A., and Knight, W. B. (1995) Characterization of pp60c-src tyrosine kinase activities using a continuous assay: autoactivation of the enzyme is an intermolecular autophosphorylation process. *Biochemistry* **34**, 14843-14851

126. Wong, S., Amaro, R. E., and McCammon, J. A. (2009) MM-PBSA Captures Key Role of Intercalating Water Molecules at a Protein-Protein Interface. *Journal of chemical theory and computation* **5**, 422-429

127. 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 to cells : devoted to molecular & cellular mechanisms* **1**, 101-113

128. Zhuang, Z., Yoder, B. L., Burgers, P. M. J., and Benkovic, S. J. (2006) The structure of a ring-opened proliferating cell nuclear antigen–replication factor C complex revealed by fluorescence energy transfer. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 2546-2551

129. Anderson, S. G., Thompson, J. A., Paschall, C. O., O'Donnell, M., and Bloom, L. B. (2009) Temporal Correlation of DNA Binding, ATP Hydrolysis, and Clamp Release in the Clamp Loading Reaction Catalyzed by the Escherichia coli γ complex. *Biochemistry* **48**, 8516-8527

130. Hayner, J. N., and Bloom, L. B. (2013) The β Sliding Clamp Closes around DNA prior to Release by the Escherichia coli Clamp Loader γ Complex. *Journal of Biological Chemistry* **288**, 1162-1170

131. Benkovic, S. J., Valentine, A. M., and Salinas, F. (2001) Replisome-mediated DNA replication. *Annual Review of Biochemistry* **70**, 181-208

132. Kelman, Z. (1997) PCNA: structure, functions and interactions. *Oncogene* **14**, 629-640

133. Eichinger, C. S., and Jentsch, S. (2011) 9-1-1: PCNA's specialized cousin. *Trends in biochemical sciences* **36**, 563-568

134. Stolimenov, I., and Helleday, T. (2009) PCNA on the crossroad of cancer. *Biochemical Society Transactions* **37**, 605-613

135. Ellison, V., and Stillman, B. (2003) Biochemical characterization of DNA damage checkpoint complexes: Clamp loader and clamp complexes with specificity for 5 ' recessed DNA. *Plos Biol* **1**, 231-243

136. Parrilla-Castellar, E. R., Arlander, S. J. H., and Kamitz, L. (2004) Dial 9-1-1 for DNA damage: the RaO-Hus1-Rad1 (9-1-1) clamp complex. *DNA Repair* **3**, 1009-1014

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

138. Delacroix, S., Wagner, J. M., Kobayashi, M., Yamamoto, K., and Karnitz, L. M. (2007) The Rad9-Hus1-Rad1 (9-1-1) clamp activates checkpoint signaling via TopBP1. *Genes & development* **21**, 1472-1477

139. Komatsu, K., Miyashita, T., Hang, H., Hopkins, K. M., Zheng, W., Cuddeback, S., Yamada, M., Lieberman, H. B., and Wang, H. G. (2000) Human homologue of S. pombe Rad9 interacts with BCL-2/BCL-xL and promotes apoptosis. *Nature cell biology* **2**, 1-6

140. Friedrich-Heineken, E., Toueille, M., Tannler, B., Burki, C., Ferrari, E., Hottiger, M. O., and Hubscher, U. (2005) The two DNA clamps Rad9/Rad1/Hus1 complex and proliferating cell nuclear antigen differentially regulate flap endonuclease 1 activity. *Journal of molecular biology* **353**, 980-989

141. Sohn, S. Y., and Cho, Y. (2009) Crystal Structure of the Human Rad9-Hus1-Rad1 Clamp. *Journal of molecular biology* **390**, 490-502

142. Xu, M., Bai, L., Gong, Y., Xie, W., Hang, H. Y., and Jiang, T. (2009) Structure and Functional Implications of the Human Rad9-Hus1-Rad1 Cell Cycle Checkpoint Complex. *Journal of Biological Chemistry* **284**, 20457-20461

143. Querol-Audi, J., Yan, C., Xu, X., Tsutakawa, S. E., Tsai, M. S., Tainer, J. A., Cooper, P. K., Nogales, E., and Ivanov, I. (2012) Repair complexes of FEN1 endonuclease, DNA, and Rad9-Hus1-Rad1 are distinguished from their PCNA counterparts by functionally important stability. *Proc Natl Acad Sci U S A* **109**, 8528-8533

144. Navadgi-Patil, V. M., and Burgers, P. M. (2011) Cell-cycle-specific activators of the Mec1/ATR checkpoint kinase. *Biochemical Society Transactions* **39**, 600-605

145. Lieberman, H. B., Bernstock, J. D., Broustas, C. G., Hopkins, K. M., Leloup, C., and Zhu, A. P. (2011) The role of RAD9 in tumorigenesis. *J Mol Cell Biol* **3**, 39-43

146. Balakrishnan, L., Brandt, P. D., Lindsey-Boltz, L. A., Sancar, A., and Bambara, R. A. (2009) Long patch base excision repair proceeds via coordinated stimulation of the multienzyme DNA repair complex. *Journal of Biological Chemistry* **284**, 15158-15172

147. Gembka, A., Toueille, M., Smirnova, E., Poltz, R., Ferrari, E., Villani, G., and Hubscher, U. (2007) The checkpoint clamp, Rad9-Rad1-Hus1 complex, preferentially stimulates the activity of apurinic/apyrimidinic endonuclease 1 and DNA polymerase beta in long patch base excision repair. *Nucleic Acids Research* **35**, 2596-2608

148. Guan, X., Madabushi, A., Chang, D. Y., Fitzgerald, M. E., Shi, G., Drohat, A. C., and Lu, A. L. (2007) The human checkpoint sensor Rad9-Rad1-Hus1 interacts with and stimulates DNA repair enzyme TDG glycosylase. *Nucleic Acids Research* **35**, 6207-6218

149. Helt, C. E., Wang, W., Keng, P. C., and Bambara, R. A. (2005) Evidence that DNA damage detection machinery participates in DNA repair. *Cell Cycle* **4**, 529-532

150. Jeruzalmi, D., O'Donnell, M., and Kuriyan, J. (2002) Clamp loaders and sliding clamps. *Current Opinion in Structural Biology* **12**, 217-224

151. Tainer, J. A., McCammon, J. A., and Ivanov, I. (2010) Recognition of the ring-opened state of proliferating cell nuclear antigen by replication factor C promotes eukaryotic clamp-loading. *J Am Chem Soc* **132**, 7372-7378

152. Chen, S. Y., Levin, M. K., Sakato, M., Zhou, Y. Y., and Hingorani, M. M. (2009) Mechanism of ATP-Driven PCNA Clamp Loading by S. cerevisiae RFC. *Journal of molecular biology* **388**, 431-442

153. Bermudez, V. P., Lindsey-Boltz, L. A., Cesare, A. J., Maniwa, Y., Griffith, J. D., Hurwitz, J., and Sancar, A. (2003) Loading of the human 9-1-1 checkpoint complex onto DNA by the checkpoint clamp loader hRad17-replication factor C complex in vitro. *Proc Natl Acad Sci U S A* **100**, 1633-1638

154. Brice, A. R., and Dominy, B. N. (2011) Analyzing the robustness of the MM/PBSA free energy calculation method: application to DNA conformational transitions. *J Comput Chem* **32**, 1431-1440

155. Case DA, Darden TA, Cheatham TE III, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Wang B, Pearlman DA, Crowley M, Brozell S, Tsui V, Gohlke H, Mongan J, Hornak V, Cui G, Beroza P, Schafmeister C, Caldwell JW, Ross WS, and Kollman PA. (2004) AMBER8.

156. Vriend, G. (1990) WHAT IF: a molecular modeling and drug design program. *J Mol Graph* **8**, 52-56

157. Humphrey W, Dalke A, and Schulten K. (1996) VMD: visual molecular dynamics. *J Mol Graph Model* **14**, 33-38

158. Gohlke, H., Kiel, C., and Case, D. A. (2003) Insights into Protein–Protein Binding by Binding Free Energy Calculation and Free Energy Decomposition for the Ras–Raf and Ras–RalGDS Complexes. *Journal of molecular biology* **330**, 891-913

159. Wang, L., Xu, X., Kumar, R., Maiti, B., Liu, C. T., Ivanov, I., Lee, T. H., and Benkovic, S. J. (2013) Probing DNA clamps with single-molecule force spectroscopy. *Nucleic Acids Res*

160. Benkovic, S. J., Valentine, A. M., and Salinas, F. (2001) Replisome-mediated DNA replication. *Annual review of biochemistry* **70**, 181-208

161. Joyce, C. M., and Benkovic, S. J. (2004) DNA polymerase fidelity: kinetics, structure, and checkpoints. *Biochemistry* **43**, 14317-14324

162. Zhao, Y., Jeruzalmi, D., Moarefi, I., Leighton, L., Lasken, R., and Kuriyan, J. (1999) Crystal structure of an archaebacterial DNA polymerase. *Structure* **7**, 1189-1199

163. Gouge, J., Ralec, C., Henneke, G., and Delarue, M. (2012) Molecular recognition of canonical and deaminated bases by P. abyssi family B DNA polymerase. *Journal of molecular biology* **423**, 315-336

164. Wynne, S. A., Pinheiro, V. B., Holliger, P., and Leslie, A. G. (2013) Structures of an apo and a binary complex of an evolved archeal B family DNA polymerase capable of synthesising highly cy-dye labelled DNA. *PloS one* **8**, e70892

165. Kunkel, T. A. (2004) DNA replication fidelity. *The Journal of biological chemistry* **279**, 16895-16898

166. Shamoo, Y., and Steitz, T. A. (1999) Building a replisome from interacting pieces: sliding clamp complexed to a peptide from DNA polymerase and a polymerase editing complex. *Cell* **99**, 155-166

167. Franklin, M. C., Wang, J., and Steitz, T. A. (2001) Structure of the replicating complex of a pol alpha family DNA polymerase. *Cell* **105**, 657-667

168. Maga, G., and Hubscher, U. (2003) Proliferating cell nuclear antigen (PCNA): a dancer with many partners. *J Cell Sci* **116**, 3051-3060

169. Kelman, Z. (1997) PCNA: structure, functions and interactions. *Oncogene* **14**, 629-640

170. Nishida, H., Mayanagi, K., Kiyonari, S., Sato, Y., Oyama, T., Ishino, Y., and Morikawa, K. (2009) Structural determinant for switching between the polymerase and exonuclease modes in the PCNA-replicative DNA polymerase complex. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 20693-20698

171. Mayanagi, K., Kiyonari, S., Nishida, H., Saito, M., Kohda, D., Ishino, Y., Shirai, T., and Morikawa, K. (2011) Architecture of the DNA polymerase B-proliferating cell nuclear antigen (PCNA)-DNA ternary complex. *Proc Natl Acad Sci U S A* **108**, 1845-1849

172. Trabuco, L. G., Villa, E., Mitra, K., Frank, J., and Schulten, K. (2008) Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* **16**, 673-683

173. Trabuco, L. G., Villa, E., Schreiner, E., Harrison, C. B., and Schulten, K. (2009) Molecular dynamics flexible fitting: A practical guide to combine cryo-electron microscopy and X-ray crystallography. *Methods* **49**, 174-180

174. Tori, K., Kimizu, M., Ishino, S., and Ishino, Y. (2007) DNA polymerases BI and D from the hyperthermophilic archaeon Pyrococcus furiosus both bind to proliferating cell nuclear antigen with their C-terminal PIP-box motifs. *J Bacteriol* **189**, 5652-5657

175. Lavery, R., Moakher, M., Maddocks, J. H., Petkeviciute, D., and Zakrzewska, K. (2009) Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res* **37**, 5917-5929

176. Mayanagi, K., Kiyonari, S., Saito, M., Shirai, T., Ishino, Y., and Morikawa, K. (2009) Mechanism of replication machinery assembly as revealed by the DNA ligase-PCNA-DNA complex architecture. *Proc Natl Acad Sci U S A* **106**, 4647-4652

177. Brooks, B., and Karplus, M. (1985) Normal modes for specific motions of macromolecules: application to the hinge-bending mode of lysozyme. *Proc Natl Acad Sci U S A* **82**, 4995-4999

178. Thomas, A., Hinsen, K., Field, M. J., and Perahia, D. (1999) Tertiary and quaternary conformational changes in aspartate transcarbamylase: a normal mode study. *Proteins* **34**, 96-112

179. Zheng, W. J., and Doniach, S. (2003) A comparative study of motor-protein motions by using a simple elastic-network model. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 13253-13258

180. Hayward, S., Kitao, A., and Berendsen, H. J. (1997) Model-free methods of analyzing domain motions in proteins from simulation: a comparison of normal mode analysis and molecular dynamics simulation of lysozyme. *Proteins* **27**, 425-437

181. Amadei, A., Linssen, A. B., and Berendsen, H. J. (1993) Essential dynamics of proteins. *Proteins* **17**, 412-425

182. Patel, S. S., Wong, I., and Johnson, K. A. (1991) Pre-steady-state kinetic analysis of processive DNA replication including complete characterization of an exonuclease-deficient mutant. *Biochemistry* **30**, 511-525

183. Marco, E., Foucaud, M., Langer, I., Escrieut, C., Tikhonova, I. G., and Fourmy, D. (2007) Mechanism of activation of a G protein-coupled receptor, the human cholecystokinin-2 receptor. *The Journal of biological chemistry* **282**, 28779-28790

184. Lee, H. S., Robinson, R. C., Joo, C. H., Lee, H., Kim, Y. K., and Choe, H. (2006) Targeted molecular dynamics simulation studies of calcium binding and conformational change in the C-terminal half of gelsolin. *Biochem Bioph Res Co* **342**, 702-709

185. Cheng, X. L., Wang, H. L., Grant, B., Sine, S. M., and McCammon, J. A. (2006) Targeted molecular dynamics study of C-loop closure and channel gating in nicotinic receptors. *Plos Comput Biol* **2**, 1173-1184

186. Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995) A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **103**, 8577-8593

187. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781-1802

188. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **65**, 712-725

189. Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577-2637

190. Grasby, J. A., Finger, L. D., Tsutakawa, S. E., Atack, J. M., and Tainer, J. A. (2011) Unpairing and gating: sequence-independent substrate recognition by FEN superfamily nucleases. *Trends in biochemical sciences*

191. Chapados, B. R., Hosfield, D. J., Han, S., Qiu, J. Z., Yelent, B., Shen, B. H., and Tainer, J. A. (2004) Structural basis for FEN-1 substrate specificity and PCNA-mediated activation in DNA replication and repair. *Cell* **116**, 39-50

192. Tsutakawa, S. E., Classen, S., Chapados, B. R., Arvai, A. S., Finger, L. D., Guenther, G., Tomlinson, C. G., Thompson, P., Sarker, A. H., Shen, B. H., Cooper, P. K., Grasby, J. A., and Tainer, J. A. (2011) Human Flap Endonuclease Structures, DNA Double-Base Flipping, and a Unified Understanding of the FEN1 Superfamily. *Cell* **145**, 198-211

193. Zheng, L., Jia, J., Finger, L. D., Guo, Z. G., Zer, C., and Shen, B. H. (2011) Functional regulation of FEN1 nuclease and its link to cancer. *Nucleic Acids Research* **39**, 781-794

194. Freudenreich, C. H., Kantrow, S. M., and Zakian, V. A. (1998) Expansion and length-dependent fragility of CTG repeats in yeast. *Science* **279**, 853-856

195. Schweitzer, J. K., and Livingston, D. M. (1998) Expansions of CAG repeat tracts are frequent in a yeast mutant defective in Okazaki fragment maturation. *Hum Mol Genet* **7**, 69-74

196. Kucherlapati, M., Yan, K., Kuraguchi, M., Zhao, J., Lia, M., Heyer, J., Kane, M. F., Fan, K. H., Russell, R., Brown, A. M. C., Kneitz, B., Edelmann, W., Kolodner, R. D., Lipkin, M., and Kucherlapati, R. (2002) Haploinsufficiency of Flap endonuclease (Fen1) leads to rapid tumor progression. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 9924-9929

197. Henneke, G., Friedrich-Heineken, E., and Hubscher, U. (2003) Flap endonuclease 1: a novel tumour suppresser protein. *Trends in biochemical sciences* **28**, 384-390

198. Jonsson, Z. O., Hindges, R., and Hubscher, U. (1998) Regulation of DNA replication and repair proteins through interaction with the front side of proliferating cell nuclear antigen. *Embo J.* **17**, 2412-2425

199. Li, X. Y., Li, J., Harrington, J., Lieber, M. R., and Burgers, P. M. J. (1995) Lagging-Strand DNA-Synthesis at the Eukaryotic Replication Fork Involves Binding and Stimulation of Fen-1 by Proliferating Cell Nuclear Antigen. *Journal of Biological Chemistry* **270**, 22109-22112

200. Gomes, X. V., and Burgers, P. M. J. (2000) Two modes of FEN1 binding to PCNA regulated by DNA. *Embo J.* **19**, 3811-3821

201. Sakurai, S., Kitano, K., Yamaguchi, H., Hamada, K., Okada, K., Fukuda, K., Uchida, M., Ohtsuka, E., Morioka, H., and Hakoshima, T. (2005) Structural basis for recruitment of human flap endonuclease 1 to PCNA. *Embo J.* **24**, 683-693

202. Chapados, B. R., Hosfield, D. J., Han, S., Qiu, J., Yelent, B., Shen, B., and Tainer, J. A. (2004) Structural basis for FEN-1 substrate specificity and PCNA-mediated activation in DNA replication and repair. *Cell* **116**, 39-50

203. Dore, A. S., Kilkenny, M. L., Rzechorzek, N. J., and Pearl, L. H. (2009) Crystal Structure of the Rad9-Rad1-Hus1 DNA Damage Checkpoint Complex - Implications for Clamp Loading and Regulation. *Molecular cell* **34**, 735-745

204. Daura, X., Jaun, B., Seebach, D., van Gunsteren, W. F., and Mark, A. E. (1998) Reversible peptide folding in solution by molecular dynamics simulation. *Journal of molecular biology* **280**, 925-932

205. Lindahl, E., Hess, B., and van der Spoel, D. (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling* **7**, 306-317

206. Ivanov, I., Chapados, B. R., McCammon, J. A., and Tainer, J. A. (2006) Proliferating cell nuclear antigen loaded onto double-stranded DNA: dynamics, minor groove interactions and functional implications. *Nucleic Acids Research* **34**, 6023-6033

207. Georgescu, R. E., Kim, S. S., Yurieva, O., Kuriyan, J., Kong, X. P., and O'Donnell, M. (2008) Structure of a sliding clamp on DNA. *Cell* **132**, 43-54

208. Mayanagi, K., Kiyonari, S., Nishida, H., Saito, M., Kohda, D., Ishino, Y., Shirai, T., and Morikawa, K. (2011) Architecture of the DNA polymerase B-proliferating cell nuclear antigen (PCNA)-DNA ternary complex. *P Natl Acad Sci USA* **108**, 1845-1849

209. Mayanagia, K., Kiyonari, S., Saito, M., Shirai, T., Ishino, Y., and Morikawa, K. (2009) Mechanism of replication machinery assembly as revealed by the DNA ligase-PCNA-DNA complex architecture. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 4647-4652

210. Orlova, E. V., and Saibil, H. R. (2010) Methods for Three-Dimensional Reconstruction of Heterogeneous Assemblies. *Method Enzymol* **482**, 321-341

211. Zhang, W., Kirnmel, M., Spahn, C. M. T., and Penczek, P. A. (2008) Heterogeneity of Large Macromolecular Complexes Revealed by 3D Cryo-EM Variance Analysis. *Structure* **16**, 1770-1776

212. Ludtke, S. J., Baldwin, P. R., and Chiu, W. (1999) EMAN: semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol* **128**, 82-97

213. Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I., and Ludtke, S. J. (2007) EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol* **157**, 38-46

214. Lavery, R., Moakher, M., Maddocks, J. H., Petkeviciute, D., and Zakrzewska, K. (2009) Conformational analysis of nucleic acids revisited: Curves. *Nucleic Acids Research* **37**, 5917-5929

215. Kazlauskas, D., and Venclovas, C. (2011) Computational analysis of DNA replicases in double-stranded DNA viruses: relationship with the genome size. *Nucleic Acids Research* **39**, 8291-8305

216. Friedrich-Heineken, F., Toueille, M., Tannler, B., Burki, C., Ferrari, E., Hottiger, M. O., and Hubscher, U. (2005) The two DNA clamps Rad9/Rad1/Hus1 complex and proliferating cell nuclear antigen differentially regulate flap endonuclease 1 activity. *Journal of molecular biology* **353**, 980-989

217. Suloway, C., Pulokas, J., Fellmann, D., Cheng, A., Guerra, F., Quispe, J., Stagg, S., Potter, C. S., and Carragher, B. (2005) Automated molecular microscopy: the new Leginon system. *J Struct Biol* **151**, 41-60

218. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926-935

219. (!!! INVALID CITATION !!!).

220. Perry, J. J. P., Tainer, J. A., and Boddy, M. N. (2008) A SIM-ultaneous role for SUMO and ubiquitin. *Trends in biochemical sciences* **33**, 201-208

221. Ulrich, H. D. (2005) Mutual interactions between the SUMO and ubiquitin systems: a plea of no contest. *Trends Cell Biol* **15**, 525-532

222. Bienko, M., Green, C. M., Crosetto, N., Rudolf, F., Zapart, G., Coull, B., Kannouche, P., Wider, G., Peter, M., Lehmann, A. R., Hofmann, K., and Dikic, I. (2005) Ubiquitin-binding domains in Y-family polymerases regulate translesion synthesis. *Science* **310**, 1821-1824

223. Hicke, L. (2001) Protein regulation by monoubiquitin. *Nat Rev Mol Cell Biol* **2**, 195-201

224. Hoege, C., Pfander, B., Moldovan, G. L., Pyrowolakis, G., and Jentsch, S. (2002) RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature* **419**, 135-141

225. Ulrich, H. D., and Walden, H. (2010) Ubiquitin signalling in DNA replication and repair. *Nat Rev Mol Cell Biol* **11**, 479-489

226. Ulrich, H. D. (2012) Ubiquitin and SUMO in DNA repair at a glance. *Journal of Cell Science* **125**, 249-254

227. Stelter, P., and Ulrich, H. D. (2003) Control of spontaneous and damage-induced mutagenesis by SUMO and ubiquitin conjugation. *Nature* **425**, 188-191

228. Yang, K., Weinacht, C. P., and Zhuang, Z. (2013) Regulatory Role of Ubiquitin in Eukaryotic DNA Translesion Synthesis. *Biochemistry*

229. Pfander, B., Moldovan, G. L., Sacher, M., Hoege, C., and Jentsch, S. (2005) SUMO-modified PCNA recruits Srs2 to prevent recombination during S phase. *Nature* **436**, 428-433

230. Armstrong, A. A., Mohideen, F., and Lima, C. D. (2012) Recognition of SUMO-modified PCNA requires tandem receptor motifs in Srs2. *Nature* **483**, 59-63

231. Freudenthal, B. D., Brogie, J. E., Gakhar, L., Kondratick, C. M., and Washington, M. T. (2011) Crystal Structure of SUMO-Modified Proliferating Cell Nuclear Antigen. *Journal of molecular biology* **406**, 9-17

232. Krejci, L., Van Komen, S., Li, Y., Villemain, J., Reddy, M. S., Klein, H., Ellenberger, T., and Sung, P. (2003) DNA helicase Srs2 disrupts the Rad51 presynaptic filament. *Nature* **423**, 305-309

233. Veaute, X., Jeusset, J., Soustelle, C., Kowalczykowski, S. C., Le Cam, E., and Fabre, F. (2003) The Srs2 helicase prevents recombination by disrupting Rad51 nucleoprotein filaments. *Nature* **423**, 309-312

234. Das-Bradoo, S., Nguyen, H. D., Wood, J. L., Ricke, R. M., Haworth, J. C., and Bielinsky, A. K. (2010) Defects in DNA ligase I trigger PCNA ubiquitylation at Lys 107. *Nature cell biology* **12**, 74-79; sup pp 71-20

235. Putnam, C. D., Hammel, M., Hura, G. L., and Tainer, J. A. (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev Biophys* **40**, 191-285

236. Rambo, R. P., and Tainer, J. A. (2011) Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers* **95**, 559-571

237. Rambo, R. P., and Tainer, J. A. (2010) Bridging the solution divide: comprehensive structural analyses of dynamic RNA, DNA, and protein assemblies by small-angle X-ray scattering. *Current Opinion in Structural Biology* **20**, 128-137

238. Hura, G. L., Menon, A. L., Hammel, M., Rambo, R. P., Poole, F. L., 2nd, Tsutakawa, S. E., Jenney, F. E., Jr., Classen, S., Frankel, K. A., Hopkins, R. C., Yang, S. J., Scott, J. W., Dillard, B. D., Adams, M. W., and Tainer, J. A. (2009) Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat Methods* **6**, 606-612

239. Chen, J. J., Ai, Y. X., Wang, J. L., Haracska, L., and Zhuang, Z. H. (2010) Chemically ubiquitylated PCNA as a probe for eukaryotic translesion DNA synthesis. *Nat Chem Biol* **6**, 270-272

240. Freudenthal, B. D., Gakhar, L., Ramaswamy, S., and Washington, M. T. (2010) Structure of monoubiquitinated PCNA and implications for translesion synthesis and DNA polymerase exchange. *Nat Struct Mol Biol* **17**, 479-U123

241. Tsutakawa, S. E., Van Wynsberghe, A. W., Freudenthal, B. D., Weinacht, C. P., Gakhar, L., Washington, M. T., Zhuang, Z., Tainer, J. A., and Ivanov, I. (2011) Solution X-ray scattering combined with computational modeling reveals multiple conformations of covalently bound ubiquitin on PCNA. *Proc Natl Acad Sci U S A* **108**, 17672-17677

242. Leaver-Fay, A., Tyka, M., Lewis, S. M., Lange, O. F., Thompson, J., Jacak, R., Kaufman, K., Renfrew, P. D., Smith, C. A., Sheffler, W., Davis, I. W., Cooper, S., Treuille, A., Mandell, D. J., Richter, F., Ban, Y. E., Fleishman, S. J., Corn, J. E., Kim, D. E., Lyskov, S., Berrondo, M., Mentzer, S., Popovic, Z., Havranek, J. J., Karanicolas, J., Das, R., Meiler, J., Kortemme, T., Gray, J. J., Kuhlman, B., Baker, D., and Bradley, P. (2011) ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods in enzymology* **487**, 545-574

243. Baker, R., Lewis, S. M., Sasaki, A. T., Wilkerson, E. M., Locasale, J. W., Cantley, L. C., Kuhlman, B., Dohlman, H. G., and Campbell, S. L. (2013) Site-specific monoubiquitination activates Ras by impeding GTPase-activating protein function. *Nat Struct Mol Biol* **20**, 46-52

244. Saha, A., Lewis, S., Kleiger, G., Kuhlman, B., and Deshaies, R. J. (2011) Essential role for ubiquitin-ubiquitin-conjugating enzyme interaction in ubiquitin discharge from Cdc34 to substrate. *Molecular cell* **42**, 75-83

245. Rohl, C. A., Strauss, C. E., Misura, K. M., and Baker, D. (2004) Protein structure prediction using Rosetta. *Methods in enzymology* **383**, 66-93

246. Hammel, M. (2012) Validation of macromolecular flexibility in solution by small-angle X-ray scattering (SAXS). *European biophysics journal : EBJ* **41**, 789-799

247. Schneidman-Duhovny, D., Hammel, M., and Sali, A. (2010) FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res* **38**, W540-544

248. Rambo, R. P., and Tainer, J. A. (2013) Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* **496**, 477-481

249. Pelikan, M., Hura, G. L., and Hammel, M. (2009) Structure and flexibility within proteins as identified through small angle X-ray scattering. *Gen Physiol Biophys* **28**, 174-189

250. Hurley, J. H., Lee, S., and Prag, G. (2006) Ubiquitin-binding domains. *The Biochemical journal* **399**, 361-372

251. Bomar, M. G., Pai, M. T., Tzeng, S. R., Li, S. S. C., and Zhou, P. (2007) Structure of the ubiquitin-binding zinc finger domain of human DNA Y-polymerase eta. *Embo Rep* **8**, 247-251

252. Hardeland, U., Steinacher, R., Jiricny, J., and Schar, P. (2002) Modification of the human thymine-DNA glycosylase by ubiquitin-like proteins facilitates enzymatic turnover. *The EMBO journal* **21**, 1456-1464

253. Steinacher, R., and Schar, P. (2005) Functionality of human thymine DNA glycosylase requires SUMO-regulated changes in protein conformation. *Current biology : CB* **15**, 616-623

254. Rambo, R. P., and Tainer, J. A. (2013) Super-resolution in solution X-ray scattering and its applications to structural systems biology. *Annual review of biophysics* **42**, 415-441

255. Duda, D. M., Borg, L. A., Scott, D. C., Hunt, H. W., Hammel, M., and Schulman, B. A. (2008) Structural insights into NEDD8 activation of cullin-RING ligases: conformational control of conjugation. *Cell* **134**, 995-1006

256. Williams, G. J., Williams, R. S., Williams, J. S., Moncalian, G., Arvai, A. S., Limbo, O., Guenther, G., SilDas, S., Hammel, M., Russell, P., and Tainer, J. A. (2011) ABC ATPase signature helices in Rad50 link nucleotide state to Mre11 interface for DNA repair. *Nat Struct Mol Biol* **18**, 423-431

257. Pascal, J. M., Tsodikov, O. V., Hura, G. L., Song, W., Cotner, E. A., Classen, S., Tomkinson, A. E., Tainer, J. A., and Ellenberger, T. (2006) A flexible interface between DNA ligase and PCNA supports conformational switching and efficient ligation of DNA. *Molecular cell* **24**, 279-291

258. Freudenthal, B. D., Gakhar, L., Ramaswamy, S., and Washington, M. T. (2010) Structure of monoubiquitinated PCNA and implications for translesion synthesis and DNA polymerase exchange. *Nature Structural & Molecular Biology* **17**, 479-U123

259. Chen, J., Ai, Y., Wang, J., Haracska, L., and Zhuang, Z. (2010) Chemically ubiquitylated PCNA as a probe for eukaryotic translesion DNA synthesis. *Nat Chem Biol* **6**, 270-272

260. Classen, S., Rodic, I., Holton, J., Hura, G. L., Hammel, M., and Tainer, J. A. (2010) Software for the high-throughput collection of SAXS data using an enhanced Blu-Ice/DCS control system. *J Synchrotron Radiat* **17**, 774-781

261. Classen, S., Hura, G. L., Holton, J. M., Rambo, R. P., Rodic, I., McGuire, P. J., Dyer, K., Hammel, M., Meigs, G., Frankel, K. A., and Tainer, J. A. (2013) Implementation and performance of SIBYLS: a dual endstation small-angle X-ray scattering and macromolecular crystallography beamline at the Advanced Light Source. *J Appl Crystallogr* **46**, 1-13

262. Kuhlman, B., and Baker, D. (2000) Native protein sequences are close to optimal for their structures. *Proc Natl Acad Sci U S A* **97**, 10383-10388

263. Vijay-Kumar, S., Bugg, C. E., and Cook, W. J. (1987) Structure of ubiquitin refined at 1.8 A resolution. *Journal of molecular biology* **194**, 531-544

264. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79**, 926-935

265. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *Journal of the American Chemical Society* **117**, 5179-5197

266. Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G. M., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., Caldwell, J., Wang, J. M., and Kollman, P. (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J Comput Chem* **24**, 1999-2012