ABSTRACT

Title of Document: ARCHAEAL DNA REPLICATION PROTEINS: MEMBERS AND FUNCTIONS

Zhuo Li, Doctor of Philosophy, 2013

Directed By: Professor Zvi Kelman, Institute for Bioscience and Biotechnology Research

The mechanism of DNA replication in archaea, the third domain of life, has been studied for more than two decades using biochemical, structural and bioinformatic approaches. Historically, many of the proteins that participate in archaeal replication were identified via similarity to enzymes needed for DNA replication in bacteria and eukarya. This study uses a different approach to identify new factors that may be involved in replication.

Genetic tools developed for the thermophilic archaeon *Thermococcus kodakarensis* were used to identify new replication factors that could not be recognized through *in silico* methods. First, a network of proteins that may participate in replication was identified using *in vivo* tagging of known replication enzymes. Following affinity purification the proteins that co-purified with the tagged enzymes were identified using mass spectrometry. This study describes the identification of a number of new putative replication factors.

Next, the biochemical properties of two proteins identified in the screen were characterized. One, the product of gene TK1525, was identified via its interaction with the GINS complex. This protein was predicted to be an archaeal homologue of the bacterial RecJ nuclease. It was found that the protein is a processive, manganese-dependent, single strand DNA-specific exonuclease. The protein was designated GAN for GINS-associated nuclease. GAN forms a complex with GINS and also interacts with the archaeal-specific DNA polymerase D *in vivo*. Subsequent bioinformatic analysis suggested that GAN may be the archaeal homologue of the eukaryotic Cdc45 protein.

The second protein characterized is the product of TK0808. This protein was identified via its interactions with proliferating cell nuclear antigen (PCNA). The protein, upon binding to PCNA, inhibits PCNA-dependent activities. The protein was therefore designated TIP for Thermococcales inhibitor of PCNA. While most proteins that interact with PCNA do so via a PCNA-interacting peptide (PIP) motif that interacts with the inter domain connecting loop (IDCL) on PCNA, TIP neither contains the canonical PIP motif nor interacts with PCNA via the IDCL. These findings suggest a new mechanism for PCNA binding and suggest a new mechanism to regulate PCNA-dependent activities.

ARCHAEAL DNA REPLICATION PROTEINS: MEMBERS AND FUNCTIONS

By

Zhuo Li

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2013

Advisory Committee:
Professor Philip R. DeShong, Dean's representative
Professor Najib M. El-Sayed
Professor Jason D. Kahn
Professor Zvi Kelman, Chair
Professor Shuwei Li

Dedication

To

My Wife and Our Parents

Acknowledgements

I want to first say thanks to my supervisor Dr. Zvi Kelman. He is a strict scientist with great passion in research and humor in the lab. During my five years in his lab, I not only learned knowledge, but more importantly, experienced the systematic way to perform scientific research.

I want to say thanks to my dissertation advisory committee, Drs. Najib M. El-Sayed, Jason Kahn, and Shuwei Li for their suggestions and critical evaluations of the work throughout the duration of my graduate study.

I want to say thanks to Dr. John N. Reeve, who suggested that I contact Dr. Zvi Kelman while I was looking for an advisor for my Ph.D. study. I also want to say thanks to Dr. Thomas Santangelo, who provided key materials for this study. I also give my thanks to Drs. Jerard Hurwitz, James Edward, Jane Ladner, Richard Huang, Jeffrey Hudgens, and Amanda Altieri for their contributions towards my Ph.D. training.

I want to say thanks to the current and former lab members Cassiah, Charles, Daniel, Evan, Katie, Miao, Nozomi, Ravi. They provided great help and made the time in the lab enjoyable.

Last but not least, I want to say thanks to my family who supported my study. My dedicated wife, Shuhong spent her busy years while I was separately living far from home. Our parents traveled from our home country to help us with our little daughters.

Table of Contents

List of Tables

List of Figures

Abbreviations

AP	Affinity purification
bp	Base pair
BSA	Bovine serum albumin
Da	Dalton
dNTP	Deoxynucleotide triphosphate
dsDNA	Double-stranded DNA
EDTA	Ethylenediaminetetraacetic acid
Fen1	Flap endonuclease 1
GAN	GINS associated nuclease
GINS	Sld5, Psf1, Psf2 and Psf3 complex
IR	Inverted repeat
IPTG	Isopropyl-β-D-thiogalactopyranoside
MCM	Minichromosome maintenance
MS	Mass spectrometry
ORB	Origin recognition boxes
ORC	Origin recognition complex
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase Chain Reaction
PIP	PCNA-interacting protein motif
PolB	DNA polymerase B
PolD	DNA polymerase D

RFC	Replication factor C
RPA	Replication protein A
SDS	Sodium dodecyl sulfate
SDS-PAGE	SDS-Polyacrylamide gel electrophoresis
SSB	Single strand DNA binding protein
ssDNA	Single stranded DNA
TIP	*Thermococcales* inhibitor of PCNA

Chapter 1: *Thermococcus kodakarensis* DNA Replication

Introduction

DNA replication is required for the propagation and evolution of all life forms. The mechanism of DNA replication guarantees the duplication and transfer of genetic information during cell division. The process needs to be precise, and must occur once and only once per cell cycle.

The cell cycle of archaea is a mosaic of the bacterial and eukaryal systems, with some archaeal-specific components. For many years the study of archaeal DNA replication was hindered by a lack of genetic tools, so most of the information regarding the replication process was derived from biochemical and limited cellular studies. In the last several years, however, genetic tools have become available for several archaeal species. In this review we summarize the recent genetic and biochemical studies on the replication machinery of the thermophilic euryarchaeon *T. kodakarensis*.

T. kodakarensis is an anaerobic hyperthermophilic euryarchaea with optimal growth at 85°C that was isolated from a sofatara on the shore of Kodakara Island, Japan (1). It is an obligate heterotroph that grows on organic substrate, usually in the presence of elemental sulfur that is reduced to hydrogen sulfide. Its genome contains ~2.09 Mbp with about 2,300 annotated genes (2). It was found that the organism is naturally competent for DNA uptake and efficiently incorporates donor DNA into its genome by homologous recombination (3).

T. kodakarensis origin of replication

In all organisms, DNA replication starts at specific sites known as origins of replication. Origins contain AT-rich region(s) which are readily unwound, and specific inverted repeats that facilitate the binding of origin binding proteins (OBP). The binding of these proteins to the origin results in partial melting of the duplex DNA and formation of the initial replication bubble.

Although the origin of replication in *T. kodakarensis* is currently unknown, one can predict its location and sequence based on information gained from other well-characterized archaeal species [reviewed in (4,5)]. Most archaeal origins of replication are located upstream of the gene encoding the initiator protein, Cdc6 (6). The origins of replication from most archaeal species studied also contain origin recognition boxes (ORBs), inverted repeats that serve as binding sites for the Cdc6 protein (7,8). The inverted repeats from different species share some sequence similarities with a consensus referred to as the mini-ORB followed by a string of G nucleotides refer to as a "G-string" in some origins (9-11). In addition, most archaeal origin sequences are several hundred bases long and are located in intergenic regions of the chromosome (9,11,12). The DNA sequence upstream of the gene encoding Cdc6 in *T. kodakarensis* (TK1901) resembles that of other archaeal origins of replication. The sequence includes a long AT-rich region flanked by a number of inverted repeats resembling the consensus sequence of the mini-ORB; some repeats also contain G-strings. It is thus likely that this region is the origin of replication in *T. kodakarensis*. Supporting evidence comes from the observation that the putative origin is located in close proximity to other genes encoding for replication proteins

[e.g. DNA polymerase D (PolD)] as has been reported for other archaeal and bacterial origins.

However, several archaeal species contain multiple origins of replication, and in these cases some origins are not near the Cdc6 gene (13,14). It is thus possible that *T. kodakarensis* contain origins in addition to the putative origin upstream of the TK1901 gene.

Initiation of DNA replication

The mechanism by which DNA replication initiates in archaea is not yet clear. However, studies with a number of species suggest that the Cdc6 protein participates in the process and specifically binds to the ORB. This binding may facilitate the formation of the replication bubble (7,8) (Figure 1-1). Several studies also suggest a role for the Cdc6 protein in the assembly of the minichromosome maintenance (MCM) helicase at the origin [(15,16) and references therein]. The helicase is the first component of the replisome that is recruited to the origin of replication.

A gene encoding Cdc6 has been identified in the *T. kodakarensis* genome (TK1901) located upstream of the putative origin. The *T. kodakarensis* Cdc6 protein contains all the hallmarks of Cdc6 proteins from other species, suggesting it plays a similar, essential, role during the initiation process. Support for an important cellular role comes from the observation that the gene encoding Cdc6 cannot be deleted from the *T. kodakarensis* genome (17).

By *in vivo* tagging of the Cdc6 proteins it was found that the protein interacts with a number of known or putative replication factors including proliferating cell nuclear antigen (PCNA), replication factor C (RFC) and PolD (the individual proteins

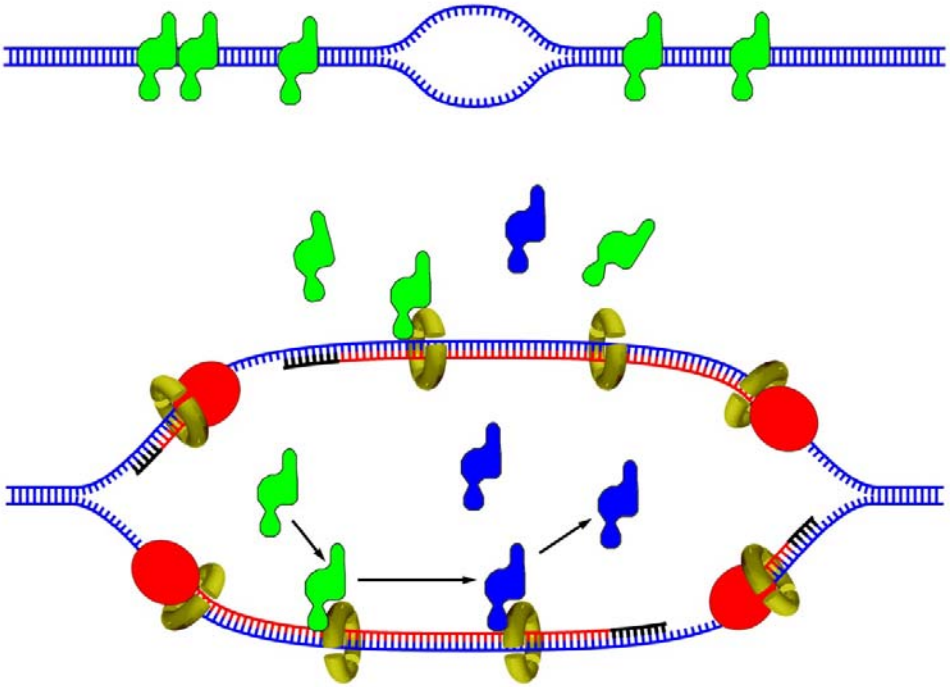

Figure 1-1: A hypothesized schematic representation of the *T. kodakarensis* replication initiation machinery.

Top: Cdc6 in the ATP-bound form (green) binds to the ORB and forms the initial replication bubble. Lower: A proposal for regulation of initiation. The replication machinery assembles at the origin and initiates bi-directional DNA synthesis [for simplicity, only the polymerase (red) and PCNA (dark yellow rings) are shown]. The moving replication fork removes the Cdc6 protein from the DNA. Cdc6 can re-bind to the newly replicated DNA, but interacts with PCNA rings left on the DNA after disengagement from the polymerase. The interaction between PCNA and the ATP-bound Cdc6 stimulates ATP hydrolysis by Cdc6 and results in an inactive enzyme (bright yellow). If this occurs, re-initiation of replication is prevented.

are discussed below). Cdc6 is the functional homolog of the *Escherichia coli* DnaA protein, which binds to the origin of replication to form the initial replication bubble and together with DnaC recruits the DnaB helicase to the DNA (18). It was also found that the DnaA protein interacts with the β-subunit of the DNA polymerase III

(PolIII) holoenzyme. This interaction plays a major role in ensuring that the *E. coli* chromosome will replicate only once per cell cycle via a mechanism called regulatory inactivation of DnaA (RIDA) (19). PCNA is the archaeal functional homolog of the β-subunit (20), and, as with many other PCNA-interacting proteins (21), the *T. kodakarensis* Cdc6 protein contains a putative PCNA interacting protein (PIP)-box. Thus, it is possible that archaea have a similar mechanism by which the interactions between Cdc6 and PCNA prevent re-initiation after origin firing (Figure 1-1).

T. kodakarensis replisome

The GINS complex

T. kodakarensis, like many other archaeal species, contains a tetrameric GINS (go-ichi-ni-san, or 5-1-2-3, from the subunits Sld5, Psf1, Psf2 and Psf3) complex assembled from two molecules each of GINS15 (TK0536p) and GINS23 (TK1619p), proteins most closely related to the eukaryal Sld5 and Psf1, and Psf2 and Psf3 proteins, respectively (22,23). As in many other archaeal species, the gene encoding GINS15 (TK0536) is in an operon with the gene encoding PCNA (TK0535) while the gene encoding GINS23 (TK1619) is in an operon with the gene encoding for MCM (TK1620) (23,24).

The three dimensional structure of the heterotetrameric *T. kodakarensis* GINS is similar overall to the human complex, although the contacts between the GINS15 and GINS23 subunits differ (25). The main structural difference between the two is the location of the C-terminal domain of the archaeal GINS15 subunit, which is located

about 30 Å away from the corresponding position of Psf1 subunits in the eukaryotic structures (25).

Although the function of the archaeal GINS is not yet clear it may be the functional homolog of the bacterial τ subunit. In bacteria, and presumably eukarya, the leading and lagging strand polymerases form a complex with the helicase and primase. In *E. coli* the τ subunit (DnaX) binds to the bacterial replicative DNA polymerase (PolIII), DNA helicase (DnaB) and primase (DnaG) at the replication fork and coordinates leading and lagging strand synthesis [reviewed in (26,27)]. Thus, the τ subunit can be thought of as the organizer of the replisome. The protein(s) that couples the polymerase, helicase and primase in archaea has not yet been identified but may be the archaeal GINS complex. This hypothesis is based in large part on *in vitro* and *in vivo* observations showing that the archaeal GINS complexes, including that of *T. kodakarensis*, interact with primase, MCM, the archaeal Cdc45 protein, PolD and PCNA (Figure 1-2) (28-31).

The MCM helicase

Following the formation of the replication bubble, the replisome is assembled at the bubble to form two replication forks and initiate bidirectional duplication of the chromosome. Based on studies in bacteria and eukarya the first replisomal enzyme associated with the replication bubble is the helicase, which unwinds the duplex DNA to provide the single-stranded (ss) DNA substrate for the polymerase. In archaea, the replicative helicase is the MCM complex [reviewed in (15,32,33)]. The biochemical and structural properties of MCM helicases from several archaea have been extensively studied. It was shown that MCM forms a hexameric ring that can bind

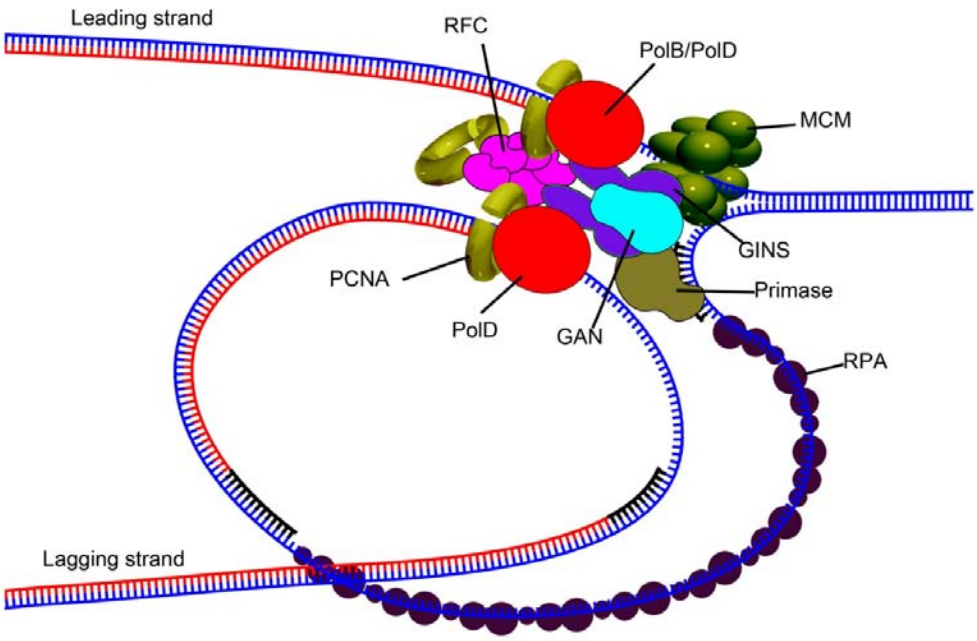

Figure 1-2: A hypothesized schematic representation of the *T. kodakarensis* replisome.

DNA Polymerase (PolB or PolD) is shown in red; MCM in dark green; RFC in pink; Primase in brown; RFC in magenta; GINS in purple; GAN in light blue; RPA in dark blue; and PCNA as dark yellow rings. The old strands of DNA are shown in blue and the newly replicated DNA in red. RNA primers are shown in black.

and translocate along ss and double-stranded (ds) DNA, has a 3'-5' directionality, requires ATP for DNA unwinding, unwinds DNA-RNA hybrids while translocating on the DNA strand, and can displace proteins from DNA [reviewed in (15,32,33)].

Most archaeal species contain a single gene encoding the MCM protein. *T. kodakarensis* is unusual by having three genes encoding MCM in its genome (TK0096, TK1361 and TK1620 encoding for MCM1, 2 and 3, respectively) (Table 1-1). It was found, however, that although all three genes are expressed and the proteins possess helicase activity, only one, MCM3, is essential for cell viability (34). Other observations also support the notion that MCM3 is the replicative helicase. The

Table 1-1. A comparison of the number of homologs of DNA replication proteins in *T. kodakarensis* to those in most other euryarchaeon and other kingdoms.

Protein (or functional unit)	*T. kodakarensis*	Euryarchaea	Crenarchaea	Korarchaeaota	Nanoarchaeota	Thaumarchaeota
Cdc6	1	1-2	1-3	2	1	1
MCM	3	1	1	1	1	1
RPA (1-3)	3	1-3	-	1	1	2
GINS (2)	1	1	1	1	1	1
Cdc45	1	1	1	1	-	-
DNA primase (2)	1	1	1	1	1	1
RFC (2)	1	1	1	1	1	1
PCNA	2	1	3	1	1	2
PolB	1	1	2-3	1	1	2
PolD (2)	1	1	-	1	1	2
Fen1	1	1	1	1	1	1
DNA ligase	1	1	1	1	1	2

When the active unit is a heterocomplex the number of subunits in the complex is shown in brackets. "-" indicates no homolog exists.

location of the MCM3 gene is in an operon with the gene encoding the GINS23 subunit (as is the case in other archaeal species) (23,24). In contrast, the genes encoding MCM1 and 2 are located in regions of the chromosome predicted to be of viral origin (2,35). It is thus possible that the proteins played a role in viral replication prior to integration into the *T. kodakarensis* genome. The MCM3 gene predicts a protein that looks more like a replicative helicase: the length of the MCM3 protein (682 residues) is similar to that of other archaeal MCM proteins, but MCM1 and 2 have long N-terminal extensions of 210 and 140 amino acids, respectively (34). These long N-terminal extensions in MCM1 and 2 proteins may have participated in viral replication processes. It is well established that helicases from many bacteriophages and eukaryotic viruses are fused to other proteins needed for replication. For examples, the bacteriophage T7 helicase is a part of a polypeptide that also includes the primase (36), and the Simian virus 40 (SV40) helicase, the large-T antigen (Tag), is a part of a protein that also contains the SV40 origin recognition protein (37). As only limited studies on archaeal viruses have been reported and only a handful have been sequenced, it is not clear if this hypothesis regarding the N-terminal extension is correct or even if MCM1 and 2 are of viral origin.

Replication protein A (RPA)

Following the assembly of the helicase around the DNA at the origin, MCM unwinds the duplex and enlarges the replication bubble. All euryarchaea studied contain homologs of the eukaryotic trimeric ssDNA binding protein (SSB), referred to as RPA. The proteins from different species, however, differ in sequence, subunit composition, and biochemical properties. While in some archaea RPA is a single

protein, in others it is a hetrotrimeric complex (as in eukarya). SSB plays a major role in DNA replication although it is not considered a part of the replisome. SSB protects ssDNA from nuclease degradation, melts secondary DNA structures that hinder polymerase movement, and regulates the hand-off of the primers from primase to the polymerase on the lagging strand (38,39).

T. kodakarensis contains three homologs of RPA (14, 31 and 41 kDa) that form heterotrimers like the eukaryotic complex (32). In contrast to bacteria and eukarya, but similar to other archaeal species, *T. kodakarensis* RPA inhibits DNA synthesis by DNA polymerase B (PolB). This may support the notion that PolB is not a replicative enzyme but is involved in other cellular processes such as repair and/or recombination.

Cdc45

The CMG (Cdc45, MCM, GINS) complex is the replicative helicase in eukarya (40). Archaeal homologs of the eukaryotic MCM and GINS proteins have been identified, but until recently no homolog of the Cdc45 protein was known. This has changed, however, as *in vivo* tagging of *T. kodakarensis* replication proteins identified a DNA nuclease that co-purified with the GINS complex [Chapter 2 and (29)]. This so-called GAN (GINS associated nuclease) [Chapter 3 and (28)] protein belongs to the RecJ family. Using computational analysis it was shown that the eukaryotic Cdc45 is an ortholog of RecJ and, by extension, the archaeal GAN (41). It was also shown that each archaeal species contains at least one RecJ homolog. It was suggested that the protein is the archaeal homolog of Cdc45. Thus, a complex of a

RecJ homolog, MCM and GINS (arCMG) may be homologous and functionally analogous to the eukaryotic CMG complex.

DNA synthesis

DNA primase

DNA polymerases require a 3'-hydroxyl primed template in order to elongate DNA chains and these primers are synthesized by DNA primases. In bacteria and eukarya, DNA primases initiate replication by synthesizing small RNA primers on both leading and lagging strands that are used by the replicative polymerases for subsequent elongation events. In bacteria, DNA primase consists of a single subunit, the DnaG protein (42). In eukaryotes, DNA primase is a part of a tetrameric complex, the DNA polymerase α (Polα)-primase. The primase part is a heterodimer containing a catalytic p48 subunit that associates tightly with a non-catalytic regulatory subunit (p58) that stabilizes and modulates the activity of the catalytic subunit (42). Following the synthesis of a short RNA primer by the primase part the Polα catalytic subunit, in association with the B regulatory subunit, elongates the primers using dNTP to form a RNA-DNA hybrid which serves as a primer for the replicative polymerases.

Archaeal genomes contain both a homolog of DnaG and a dimeric eukaryotic-type DNA primase with a catalytic and a regulatory subunit. However, studies have shown that the archaeal DnaG-like primase most likely participates in RNA degradation rather than DNA replication (29,43-45) and genetic studies suggest that

while the gene encoding DnaG is dispensable for cell growth (45), each of the eukaryotic-like primase subunits are essential (46).

In sharp contrast to the bacterial and eukaryal primases, the archaeal eukaryotic-like dimeric primases, including the *T. kodakarensis* enzyme, can initiate oligonucleotide chains *de novo* from either ribo- or deoxynucleotides [reviewed in (47)]. In fact, the *T. kodakarensis* primase was shown to have a lower Km for deoxynucleotides in comparison to ribonucleotides (48). It is thus possible that the archaeal primase incorporated the activity of both the primase and polymerase activities in the eukaryotic Polα-primase complex.

In vivo studies with several archaeal species, however, demonstrated that the primers on the lagging strand are made of ribonucleotides [for example see (49)]. The higher cellular levels of rNTPs in comparison to dNTPs may explain the apparent inconsistency between the *in vitro* and *in vivo* observation. Future studies will need to clarify these observations.

In addition to its ability to synthesize long rNTP and dNTP chains the *T. kodakarensis* primase was shown to have another, unexpected, activity. In the presence of dATP or dGTP and small molecules with -OH groups the enzyme generates dAMP adducts including dAMP-glycerol, dAMP-Tris and dAMP-Hepes (50). These products can be formed by the catalytic subunit alone or the dimeric complex in the presence or absence of ssDNA template. In addition to these derivatives, dNMP is also produced in the reaction, suggesting that H_2O can replace the other small molecule in this reaction (50). It is not yet clear if these, or similar, products play a role *in vivo*.

Polymerase accessory proteins

The replicative polymerase on its own has very low processivity. High processivity is achieved by a ring-shaped processivity factor, PCNA, that binds to the polymerase and tethers it to the DNA [reviewed in: (26,51-53)]. The clamp loader, RFC, assembles the clamp around the DNA in an ATP-dependent reaction [reviewed in: (52,54)]. In the genomes of most archaea, including *T. kodakarensis*, two homologs of RFC have been identified (RFCS and RFCL), forming a pentameric complex containing four subunits of RFCS associated with one subunit of RFCL.

While most euryarchaeota contain a single PCNA homolog that forms homotrimers, [reviewed in: (51)] the *T. kodakarensis* genome encodes two PCNA homologs (TK0535 and TK0582, encoding PCNA1 and 2, respectively) (Table 1-1). It was found that the two proteins form homotrimers in solution with similar three dimensional structures (55). However, although both PCNA1 and 2 stimulate the activity of PolB (55) and Fen1 (Chapter 4), only PCNA1 appears to be essential for cell viability (56). TK0582, encoding PCNA2, can be readily deleted from the chromosome, but repeated attempts to delete PCNA1 have failed.

Supporting the idea that PCNA1 is the replicative protein in *T. kodakarensis* is the observation that PCNA1 is in an operon with gene encoding the GINS15 protein, as has been found in other archaeal species (23,24). TK0582 is not adjacent to a GINS-encoding gene but is within a region of the *T. kodakarensis* genome thought to be of viral origin (2,35). Furthermore, although the overall three-dimensional structures of the PCNA1 and PCNA2 complexes are similar, the details of the

monomer-monomer interfaces are significantly different, with the PCNA1 interfaces most similar to those in other archaeal PCNA complexes (55).

DNA polymerase

Archaeal DNA polymerases have been purified and studied for many years, mainly for their use in PCR. Those studies concentrated mainly on members of the PolB family of DNA polymerases. This is the largest group of archaeal polymerases and all species sequenced contain at least one member of this group. Since the three replicative polymerases in eukarya, DNA polymerases (Pol) α, ε, and δ all belong to family B it was proposed that the members of this family also participate in chromosomal replication in archaea.

In addition to PolB, euryarchaeota species such as *T. kodakarensis* contain a second DNA polymerase, PolD. This dimeric enzyme contains a large and small subunit and is archaeal-specific. In many species, including *T. kodakarensis*, the genes encoding the two subunits of PolD are in an operon and located in a close proximity to the origin of replication. PolD may function at the replication fork as its small, non catalytic subunit shares amino acid sequence similarity with several of the small, non-catalytic subunits of the eukaryotic Polδ and Polε. In addition, it was shown that the large subunit of PolD shares sequence similarity with the catalytic subunit of the eukaryotic Polε and may be the functional homolog of Polε in archaea (57). Studies with halobacterium indicated that both PolB and PolD may be essential for cell viability (46). This may suggest that while in bacteria a single polymerase, PolIII, is dimerized at the replication fork to replicate the leading and lagging strands [reviewed in: (58)], archaea may have two different polymerases, PolB and PolD, at

the replication fork, similar to eukarya in which Polε replicates the leading strand while Polδ copies the lagging strand [reviewed in: (59)].

Okazaki fragment maturation

The RNA primers that initiate each Okazaki fragment on the lagging strand need to be removed, the gap in the DNA sealed, and the newly synthesized Okazaki fragment needs to be ligated to the previous fragment (Figure 1-3). It is thought that in archaea Okazaki fragment maturation is similar to eukarya: the polymerase, upon reaching the previous Okazaki fragment, displaces the primer and provides the substrate for flap endonuclease 1 (Fen1). Following Fen1 removal of the primer, the polymerase fills the gap and DNA ligase joins the two adjacent Okazaki fragments. It was shown in several archaeal species that both Fen1 and ligase are stimulated upon interaction with PCNA [for example see (60)]. *In vitro* studies showed that the *T. kodakarensis* PCNA stimulates the activity of Fen1 but no stimulation of DNA ligase activity could be detected (Chapter 4). Both Fen1 and DNA ligase, however, form complexes with PCNA *in vivo*. In contrast to other archaeal DNA ligases, the *T. kodakarensis* enzyme can utilize both ATP and NAD as cofactors for the ligation reaction although the activity with NAD was much lower in comparison to ATP (61).

What was learned from the study

DNA replication mechanism is conserved in all three domains of life. However, the proteins involved in the process are different. Although many replication peoteins could be identified via their similarities to enzymes in bacteria and eukary summarized in Table 1-2, some could not.

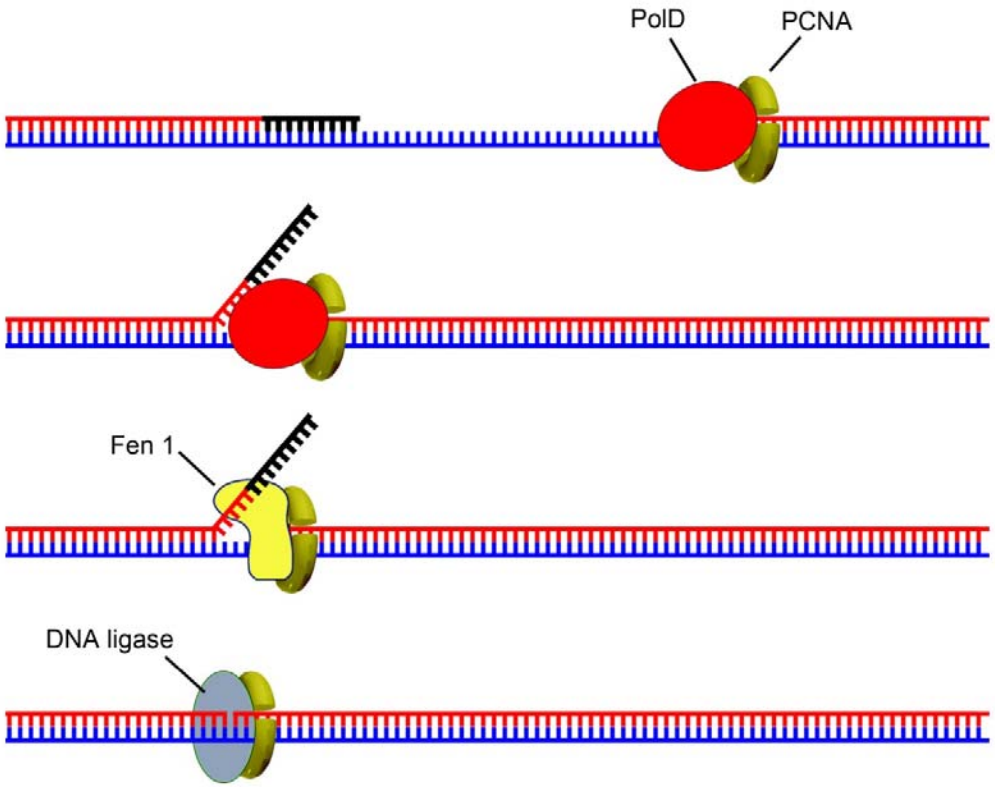

Figure 1-3: The steps in Okazaki fragment maturation on the lagging strand.

PolD is shown in red; PCNA as a dark yellow ring; FEN1 in light yellow; and DNA ligase in light gray. See text for details.

To identify new proteins that may participate in the replication process 19 *T. kodakarensis* strains, in which one know replication enzyme was tagged with a His_6-tag, were constractued. The proteins that form a complex, in vivo, with the tagged enzymes were identified, following afinity purification, using mass-spectrometry. This study identified severla new putative replicatoin enzymes (Chapter 2).

Two of the identified enzymes were bichemically characterized. One is the product of TK1252 which was identified via its interactions with the GINS complex.

The protein was shown to be a ssDNA specific exonuclease and therefore was designated GAN for GINS-associated nuclease (Chapter 3). The other is a small protein, the product of TK0808, that interacts with PCNA and upon binding inhibits the PCNA-dependent enzymes (Chapter 4). This is the first protein shown to inhibit the archaeal PCNA protein.

Table 1-2: A comparison of homologs of DNA replication proteins among three domains of life.

Function	Bacteria	Archaea	Eukaryote
Origin recognition	DnaA	ORC/CDC6	ORC1-6
Helicase loading	DnaC	Unknown	Cdc6
Helicase	DnaB	MCM	MCM2-7
Helicase enhancing	Unknown	GAN*	Cdc45
Replication scaffold	τ	GINS	GINS
Primase	DnaG	DnaG/Polα	Polα /Primase
DNA polymerase	PolIII core	PolB/PolD	Polε/Polδ
Replication clamp	β	PCNA	PCNA
Clamp regulation	Unknown	TIP*	P21
Clamp loader	γ complex	RFC	RFC
Single strand DNA binding	SSB	RPA	RPA

* Firstly identified in this study.

Chapter 2: Affinity Purification of an Archaeal DNA Replication Protein Network

Introduction

The replisome, the chromosomal DNA replication machinery, is composed of subcomplexes that separate the duplex DNA, prime and synthesize DNA, mature and ligate Okazaki fragments, and facilitate and stabilize events in replication (62). Figure 2-1 shows the model of DNA replication fork using *T. kodakarensis* proteins. Extensive biochemical and genetic research has led to the identification of conserved and domain-specific protein-protein and protein-DNA interactions that direct the assembly, and are required for the movement, functions, and stability of bacterial and eukaryotic replisomes. To date, archaeal replication has received far less experimental attention. Archaea are prokaryotes and, in common with Bacteria, most have a single circular chromosome (~0.8 to 8 Mbp) that is replicated bidirectionally from an origin of replication. With no nuclear membrane, the archaeal replisome is also assembled directly from proteins in the cytoplasm, but based on sequence conservation, most of the archaeal proteins predicted to participate in DNA replication are more closely related to eukaryotic than bacterial proteins [reviewed in (63)]. Only a few of these archaeal proteins have, however, been functionally characterized, and there are some obvious and intriguing absences of archaeal homologues of conserved bacterial and/or eukaryotic replisome proteins. For example, there are no known archaeal homologues of the *E. coli* τ subunit that couples the leading and lagging strand DNA

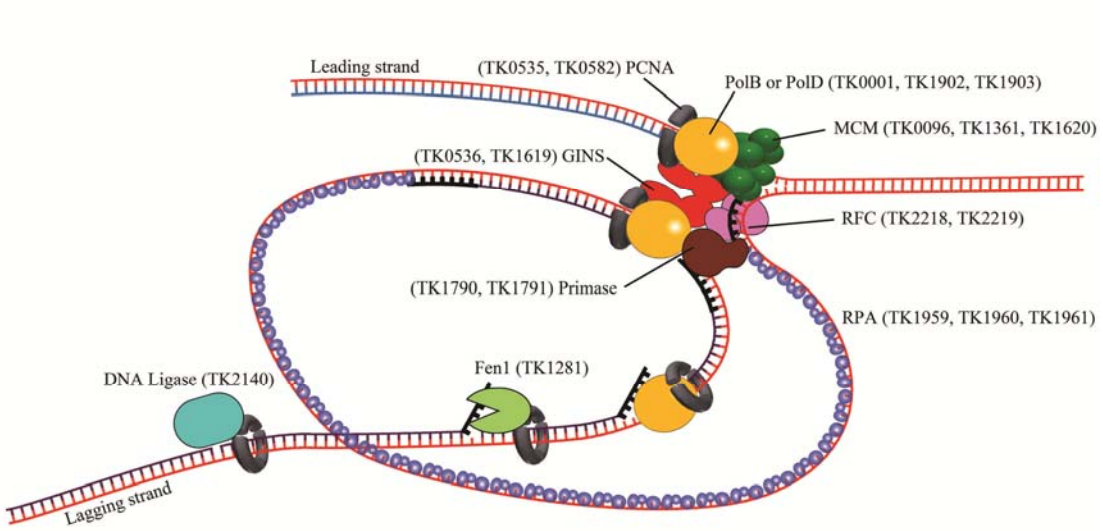

Figure 2-1: Components of the archaeal replisome.

The *T. kodakarensis* numerical gene designations are listed adjacent to the protein subcomplexes predicted to constitute the replisome (2). It remains uncertain if the replicative polymerase(s) is PolB (TK0001p) and/or PolD (TK1902p and TK1903p).

polymerases and helicase, or of Cdc45 and MCM10, proteins essential for eukaryotic chromosome replication. The functions of these proteins may therefore be carried out by unrelated archaeal proteins or by archaeal homologues with such divergent sequences that they are not readily identified by bioinformatics. Such divergence is exhibited by DNA replication processivity factors. Bacterial processivity factors (the β subunit of DNA PolIII) are homodimers of ~40-kDa subunits, whereas eukaryotic and euryarchaeal processivity factors PCNA are homotrimers of ~29-kDa subunits. Members of the order Crenarchaeota contain three different PCNA homologues that assemble into heterotrimers (64). The bacterial and eukaryotic/euryarchaeal PCNAs have only ~15% sequence identity but still form complexes with almost identical three-dimensional structures and retain the same functions (20). Here we report the

results of experiments that identify many of the proteins likely to participate in DNA replication in the euryarchaeon *T. kodakarensis*. With this experimentally documented database available, a firm foundation is established for focused research on individual archaeal replication components and for investigative exploitation of this simpler prokaryotic model for eukaryotic replication.

Obtaining the information reported was made possible by the recent development of genetic tools for *T. kodakarensis* KOD1, a heterotrophic hyperthermophile with a 2.09-Mbp genome that has ~2,300 annotated genes (2). *T. kodakarensis* is naturally competent for DNA uptake and incorporates added DNA into its chromosome by homologous recombination. By constructing DNA molecules with a target gene flanked by chromosomal sequences, the gene can be deleted, inactivated, or replaced with an allele that encodes a modified protein. For this project, we constructed 19 *T. kodakarensis* strains, each of which has a gene encoding a known or predicted replication protein replaced with the same gene with a hexahistidine-encoding sequence (His_6-tag) added in frame at either the 5' or the 3' terminus. As the modified genes were expressed from the wild-type loci, they were subject to the same regulation as the wild-type genes. The His_6-tagged proteins synthesized *in vivo* were isolated directly by Ni^{2+} affinity from clarified cell lysates, and the *T. kodakarensis* proteins that were co-isolated, as components of stable complexes assembled *in vivo*, were then identified by mass spectrometry. As reported and discussed, the identities of these proteins confirm some, but not all, of the predicted archaeal replisome interactions, reveal unpredicted associations, and provide experimental evidence for additional replication components.

Materials and Methods

Construction of transforming DNAs and transformation of *T. kodakarensis* KW128.

Genes encoding 19 known or putative replication proteins were amplified from *T. kodakarensis* genomic DNA (Table 2-1), and the His_6-encoding sequence (5' CATCATCATCATCATCAT 3') was added, in frame, to either the 3' or the 5' terminus by overlapping PCR (65). The amplified genes were cloned into pUMT2 (3) using restriction sites adjacent to *trpE* (TK0254) and flanked by ~2 kbp DNA molecules that were amplified from immediately upstream and downstream of the gene of interest. The DNA molecules and the organization of genes cloned into pUMT2 to generate the plasmids used to transform *T. kodakarensis* KW128 are illustrated in Figure 2-2. Plasmid preparations were isolated from *E. coli* DH5α cells and used directly to transform *T. kodakarensis* KW128 (*ΔpyrF ΔtrpE::pyrF*) as previously described (3,66). Transformants were selected by colony growth at 85°C on plates containing GELRITE-solidified minimal medium that lacked tryptophan. Cultures of representative transformants were grown to stationary phase in MA-YT medium (66) that contained 2 g/l sulfur. The cells were harvested, and genomic DNA was isolated. The presence of the desired chromosomal construction was confirmed by diagnostic PCR amplification and DNA sequencing as previously described (66). Homologous recombination within the flanking sequences directed integration of the transforming DNA into the *T. kodakarensis* chromosome. In each case, the wild-type gene of interest was replaced with *trpE* and the gene that encoded the His_6-tagged version of the replication protein.

Table 2-1: *T. kodakarensis* proteins His_6-tagged and used to isolate replisome components.

Protein	Gene number[1]	Known (or predicted) Function(s)
Cdc6	TK1901	Binds to origin of replication; participates in helicase assembly
DNA ligase	TK2140	Okazaki fragment maturation
DnaG-like	TK1410	Putative DNA primase
Fen1	TK1281	Okazaki fragment maturation
GINS1[2]	TK0536	Subunit of a complex that binds primase, helicase and polymerase
GINS2	TK1619	Subunit of a complex that binds primase, helicase and polymerase
MCM1[2]	TK0096	Putative replicative helicase
MCM2	TK1361	Putative replicative helicase
MCM3	TK1620	Putative replicative helicase
PCNA1[2]	TK0535	Increases polymerase processivity
PCNA2	TK0582	Increases polymerase processivity
PolB	TK0001	DNA polymerase B
PolD2	TK1903	Subunit of PolD
Primase[1]	TK1790	Subunit of DNA primase
RFC1	TK2218	Subunit of the processivity complex, assembles PCNA on DNA
RFC2	TK2219	Subunit of the processivity complex, assembles PCNA on DNA
RPA2[2]	TK1960	Subunit of the ssDNA binding protein
RPA3	TK1961	Subunit of the ssDNA binding protein
Function unknown[3]	TK1792	Protein encoded in an operon with DNA primase (TK1790p + TK1791p)

1. The numerical designation of the *T. kodakarensis* encoding gene.
2. Based on sequence homologies, *T. kodakarensis* has two GINS and PCNA proteins and three MCM and RPA proteins.
3. TK1792p, although annotated as a protein with unknown function, is in an operon with the two subunits of primase (TK1790p and TK1791p) and therefore was tagged in the study.

The His_6-encoding sequence was also added to genes that encode a subunit of RPA (RPA1; TK1959), the small subunit of euryarchaeal PolD-S (TK1902), and the small subunit of the dimeric primase (Pri-S; TK1791). Transformation with these constructs failed to generate viable *T. kodakarensis* transformants, suggesting that the His_6 extension resulted in defective enzymes.

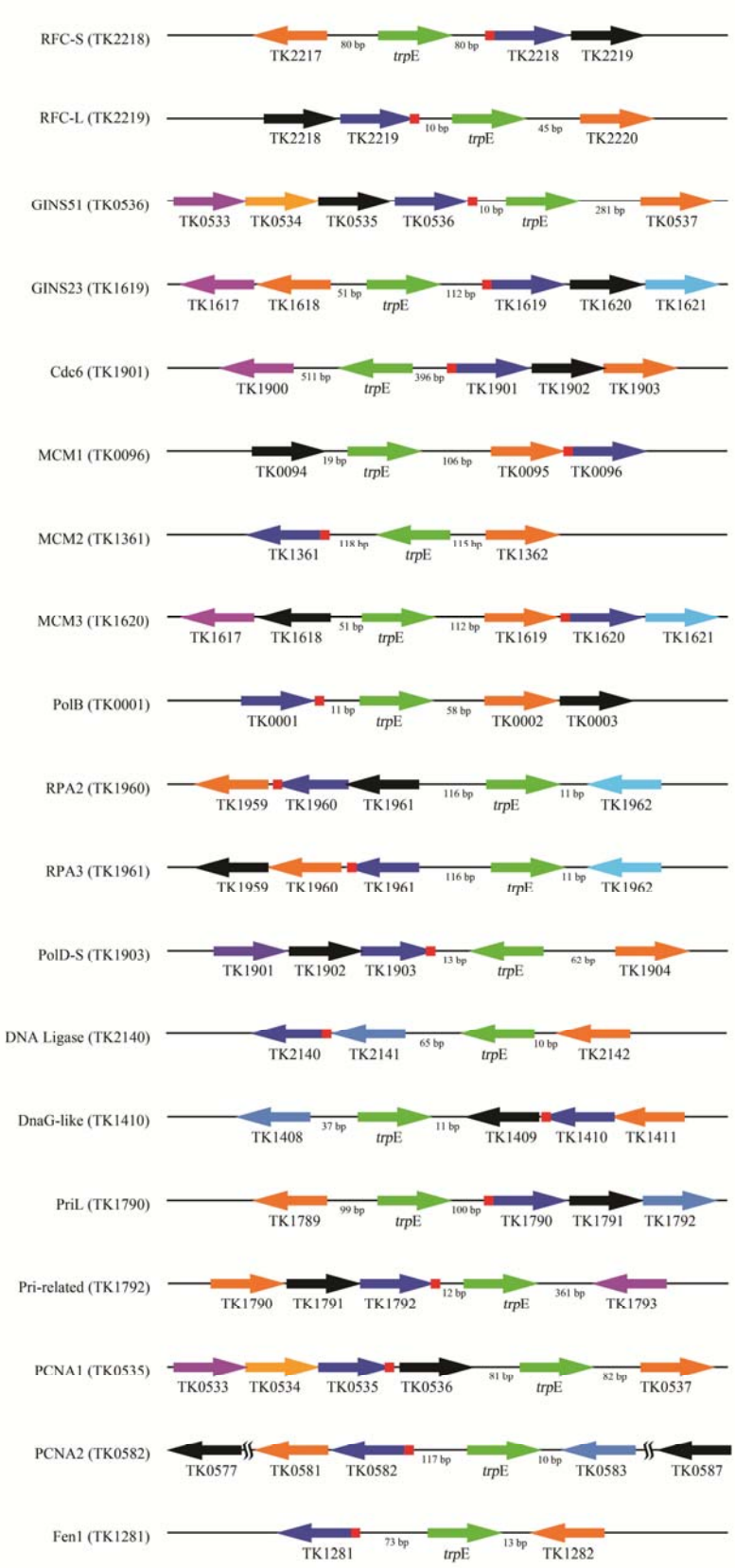
RFC-S (TK2218)
TK2217
80 bp
trpE
80 bp
TK2218
TK2219
RFC-L (TK2219)
TK2218
TK2219
10 bp
trpE
45 bp
TK2220
GINS51 (TK0536)
TK0533
TK0534
TK0535
TK0536
10 bp
trpE
281 bp
TK0537
GINS23 (TK1619)
TK1617
TK1618
51 bp
trpE
112 bp
TK1619
TK1620
TK1621
Cdc6 (TK1901)
TK1900
511 bp
trpE
396 bp
TK1901
TK1902
TK1903
MCM1 (TK0096)
TK0094
19 bp
trpE
106 bp
TK0095
TK0096
MCM2 (TK1361)
TK1361
118 bp
trpE
115 bp
TK1362
MCM3 (TK1620)
TK1617
TK1618
51 bp
trpE
112 bp
TK1619
TK1620
TK1621
PolB (TK0001)
TK0001
11 bp
trpE
58 bp
TK0002
TK0003
RPA2 (TK1960)
TK1959
TK1960
TK1961
116 bp
trpE
11 bp
TK1962
RPA3 (TK1961)
TK1959
TK1960
TK1961
116 bp
trpE
11 bp
TK1962
PolD-S (TK1903)
TK1901
TK1902
TK1903
13 bp
trpE
62 bp
TK1904
DNA Ligase (TK2140)
TK2140
TK2141
65 bp
trpE
10 bp
TK2142
DnaG-like (TK1410)
TK1408
37 bp
trpE
11 bp
TK1409
TK1410
TK1411
PriL (TK1790)
TK1789
99 bp
trpE
100 bp
TK1790
TK1791
TK1792
Pri-related (TK1792)
TK1790
TK1791
TK1792
12 bp
trpE
361 bp
TK1793
PCNA1 (TK0535)
TK0533
TK0534
TK0535
TK0536
81 bp
trpE
82 bp
TK0537
PCNA2 (TK0582)
TK0577
TK0581
TK0582
117 bp
trpE
10 bp
TK0583
TK0587
Fen1 (TK1281)
TK1281
73 bp
trpE
13 bp
TK1282

Figure 2-2: Genetic map of the regions with His_6-coding sequence.

DNA molecules constructed and cloned into pUMT2 to produce the plasmid DNAs used to transform *T. kodakarensis* KW128. The protein that was replaced with a His_6-tagged version and the gene that encodes it are shown to the left of each construct. In each construct, the target gene is shown as a blue arrow and the 5' - or 3' -terminal location of the His_6-encoding sequence is indicated by a red box. A constitutively expressed *trpE* (TK0254) gene (green arrow) that conferred tryptophan-independent growth and provided the positive selection used to isolate the desired *T. kodakarensis* KW128 transformant was incorporated into each construct.

Isolation of His_6-tagged proteins and complexes.

T. kodakarensis cells were harvested by centrifugation from 5-liter cultures grown to late exponential phase (optical density at 600 nm of ~0.8) (see Figure 2-3) at 80°C in MA-YT medium supplemented with 5 g sodium pyruvate/liter using a BioFlow 415 fermentor (New Brunswick Scientific). The cells were resuspended in 30 ml of buffer A [25 mM Tris-HCl (pH 8.0), 500 mM NaCl, 10 mM imidazole, 10% glycerol] and lysed by sonication. After centrifugation (10000g 15 min), the resulting clarified lysate was loaded onto a 1-ml HiTRAP chelating column (GE Healthcare) preequilibrated with $NiSO_4$. The column was washed with buffer A, and proteins were eluted using a linear imidazole gradient from buffer A to 67% buffer B [25 mM Tris-HCl (pH 8.0), 100 mM NaCl, 150 mM imidazole, 10% glycerol]. Fractions that contained the tagged protein were identified by Western blotting, pooled, and dialyzed against buffer C [25 mM Tris-HCl (pH 8.0), 500 mM NaCl, 0.5 mM EDTA, 2 mM dithiothreitol]. Thirty-microgram aliquots of the proteins present in solution were precipitated by adding trichloroacetic acid (TCA; 15% final concentration).

Identification of proteins by mass spectrometry.

The TCA-precipitated proteins were identified by multidimensional protein identification technology at the Ohio State University mass spectrometry facility (http://www.ccic.ohio-state.edu/MS/proteomics.htm) using the MASCOT search engine. A MASCOT score of >100 was considered meaningful. To obtain such a score, a minimum of two unique peptide fragments usually had to be identified from

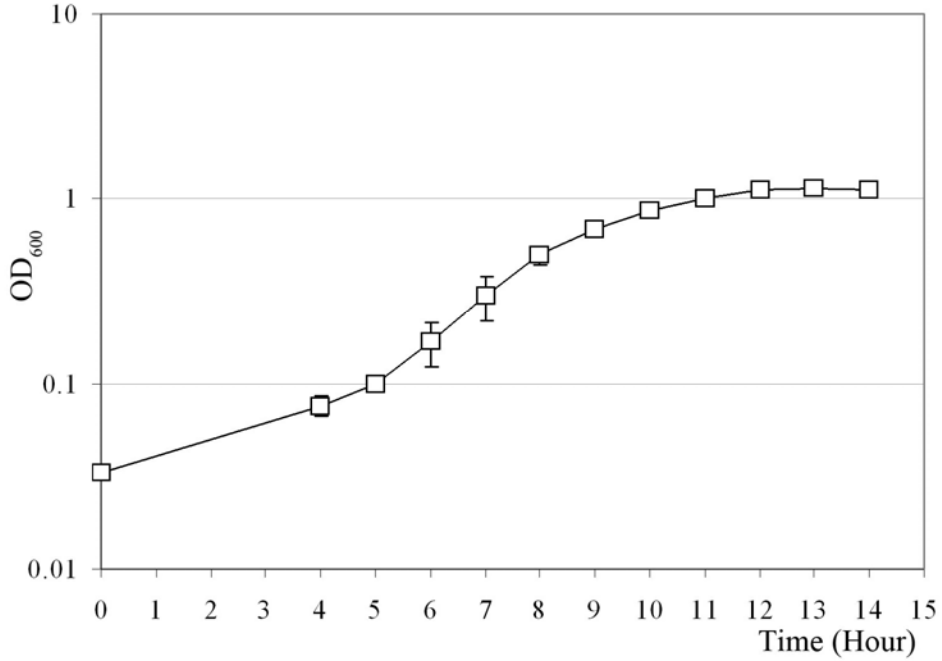

Figure 2-3: Growth curve of *T. kodakarensis* KW128.

The average values and standard deviations are shown for optical density at 600 nm (OD_{600}) measurements made during the growth of three independent cultures in MA-YT medium supplemented with 5 g sodium pyruvate/liter.

the same protein. Protein isolation and mass spectrometry analyses of lysates from two independent cultures of *T. kodakarensis* KW128 were also undertaken. From these controls, several *T. kodakarensis* proteins were identified that bound and eluted from the Ni^{2+} charged matrix in the absence of a His_6-tagged protein. All of the proteins identified in the experimental samples that had MASCOT scores of >100 and were not also present in the control samples are listed in Appendix 1.

Results

Overview

Lysates were generated from exponentially growing but not synchronized cell

populations and so contained complexes present at all stages of the replication cycle. The His_6-tagged proteins (Table 2-1) were all synthesized as soluble proteins and were present in readily detectable amounts in the clarified lysates. All of the putative protein-protein interactions detected, based on the co-isolation of a protein with a His_6-tagged protein, are documented in Appendix 1. The consistent interactions that remained, after the exclusion of proteins whose annotated functions argue strongly against a role in nucleic acid metabolic processes, are listed in Appendix 2 and illustrated as a network in Figure 2-4. Many of the interactions were confirmed by co-isolation of the same proteins when different interacting partners were His_6-tagged and used to isolate the complex. The results include both previously established and previously unknown interactions between documented, predicted, and previously unrecognized components of the archaeal replisome. In some cases, when two or three homologous proteins with very similar sequences were present and different homologues were His_6-tagged, the same proteins were co-isolated, consistent with functional redundancy. When this was not the case, the results argue for divergence of the homologues to the extent that different interactions are made, suggesting different functions.

Established archaeal replisome complexes.

As some replisome complexes were already well documented, the co-isolation of the proteins known to be components of these complexes validated and provided a measure of the sensitivity of the His_6-tag-dependent co-isolation technology. Some examples of these validating interactions are described individually below, and all are listed in Appendix 2 and documented in Figure 2-4.

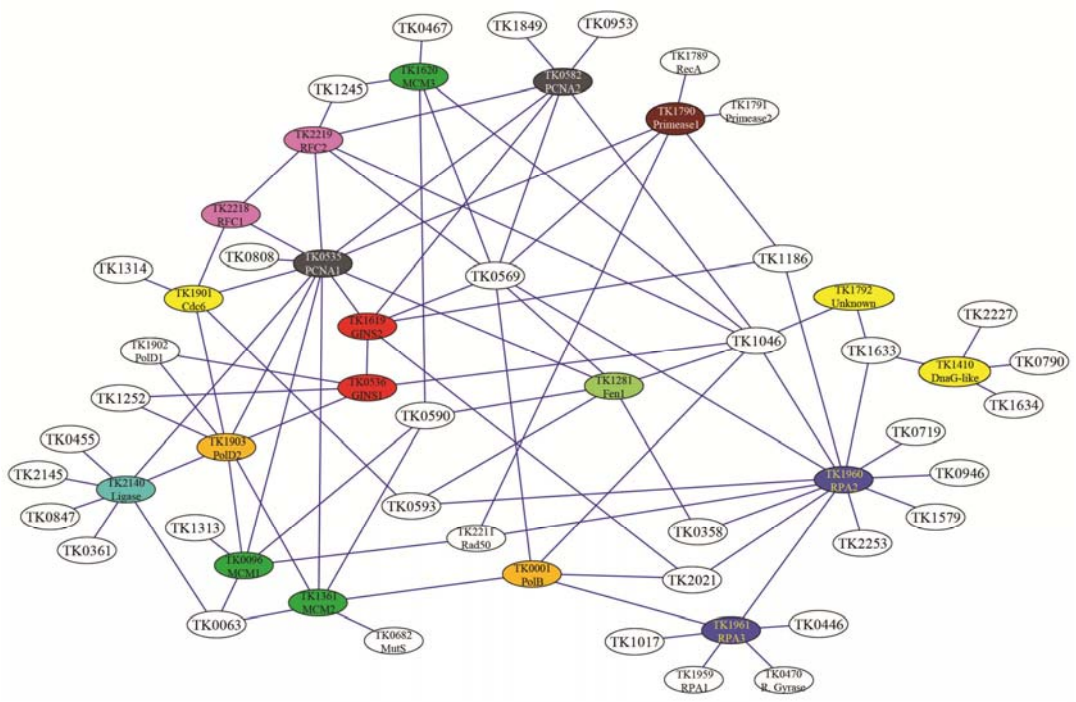

Figure 2-4: The interaction network documented for protein components of the *T. kodakarensis* replisome.

Proteins that were His_6-tagged and used to isolate interacting proteins are identified in colored ovals (the colors used are as in Figure 2-1). Proteins that were co-isolated with a His_6-tagged protein are identified in white ovals by the designations of the *T. kodakarensis* genes that encode them.

(i) Polypeptide subunits of PolD, primase, RFC, and the GINS complex. It is well established that the archaeal replisome components PolD, primase, RFC, and the GINS complex are each formed by the assembly of two different polypeptides (63,67,68). Consistent with this, the two polypeptides annotated as the subunits of these replisome proteins in *T. kodakarensis* were co-isolated with very high MASCOT scores (Appendix 2). Additional experiments with recombinant proteins also confirmed that, as predicted, the two primase, the two GINS, and the two RFC subunits assembled *in vitro* to form a heterodimer, a heterotetramer, and a heteropentamer, respectively (28,34,56,69).

(ii) RPA [single-stranded DNA (ssDNA)-binding protein] heterotrimers. *T. kodakarensis* has three genes (TK1959, TK1960, and TK1961) that encode homologues of the polypeptides that form the eukaryotic trimeric RPA complex (RPA1, RPA2, and RPA3, respectively). In *Pyrococcus furiosus*, three RPA homologues have also been identified and shown to form an active heteromeric complex (70). Consistent with this, *T. kodakarensis* RPA1 and RPA2 were co-isolated by Ni^{2+} binding of His_6-tagged RPA3 (Appendix 2).

(iii) RFC-PCNA complex formation. RFC-PCNA binding has been reported in all of the replication systems investigated (52). *T. kodakarensis* has two genes that encode PCNA homologues, PCNA1 and PCNA2 (TK0535p and TK0582p, respectively). Both the small (RFC-S; TK2218p) and large (RFC-L; TK2219p) subunits of RFC were co-isolated with His_6-tagged PCNA1, and PCNA2 was co-isolated with His_6-tagged RFC-L (Appendix 2). These co-isolation results are consistent with both PCNA homologues participating in DNA replication and with replication complexes assembled *in vivo* containing a mixture of PCNA1 and PCNA2. *In vitro* experiments have also confirmed functional interactions between RFC and both PCNA proteins (55,56,69).

(iv) PCNA-PolD-Fen1-ligase interactions. PCNA interactions with DNA polymerases increase their processivity (53). PCNA also binds and regulates the activity of a number of enzymes participating in Okazaki fragment maturation and postreplication processes [summarized in references (21,71)]. Consistent with these reports, PCNA1 was co-isolated in complexes with PolD-L (large subunit of euryarchaeon-specific PolD), Fen1, and DNA ligase (Appendix 2). There was no

evidence for a PCNA-PolB interaction when either PolB or the PCNA proteins were tagged. Such an interaction may not, however, be detectable in a soluble extract given that the bacterial and eukaryotic processivity factors (the β subunit and PCNA, respectively) must encircle the DNA to form a stable complex with the polymerase. PolD was also co-isolated with His_6-tagged DNA ligase, adding support to the hypothesis that DNA ligase is associated with the archaeal replication fork (72).

Novel interactions of archaeal replication proteins.

(i) PCNA interactions. Many proteins have been reported to interact with eukaryotic PCNA (21), but only a few of these have recognizable homologues in Archaea. Most of the proteins that bind to PCNA do so via a PIP box sequence (73,74). In addition to proteins expected to co-purify with PCNA (see above), Cdc6 (TK1901p), MCM1 (TK0096p), and MCM2 (TK1361p) were co-purified with His_6-tagged PCNA1, and these do contain PIP box-related sequences (QRAKEAFY in Cdc6p, QKPYENFW and QSKPGFY in MCM1p, and QERVIGFL in MCM2). Three additional proteins that have no known functions but also contain PIP box-related sequences were also routinely co-isolated in complexes with His_6-tagged PCNA2, namely, TK0569p (QPRSPFYP), TK0953p (QALAEWYA), and TK1046p (QGYRESFA). MCM and PCNA are both established replisome participants, but this is the first experimental evidence for their copresence within a stable complex and the presence of the PIP box sequence suggests a direct MCM-PCNA interaction. The possible roles of PCNA-Cdc6 interaction are discussed below. Homologues of TK0569p are present in Archaea and Bacteria, and homologues of TK1046p are present in all three domains (discussed below). Homologues of TK0953p are present

in a small number of archaeal and bacterial species and each appears to have an ATPase domain.

(ii) GINS interactions. In eukaryotes, the GINS complex is an assembly of four different polypeptides (designated Sld5, Psf1, Psf2, and Psf3) that interact with several replisome components, including MCM and the Polα-primase complex [references (23,75)]. The GINS complex plays a role in both the initiation and elongation phases of DNA replication. All archaeal genomes contain a single protein, designated GINS15, that has sequence similarity to Sld5 and Psf1. Some Archaea, including *T. kodakarensis*, also have a protein designated GINS23 that is related to Psf2 and Psf3 (22) and forms a tetrameric complex that contains two GINS15 and two GINS23 subunits (22). Both subunits of PolD were co-isolated using His_6-tagged GINS15, and PCNA1 and PCNA2 were both co-isolated with His_6-tagged GINS23, providing the first experimental evidence for a stable replisome association of the GINS complex with PolD and PCNA.

TK1252p, a protein co-isolated with His_6-tagged GINS15 (Appendix 2), is annotated as an ssDNA-specific exonuclease with some homology to bacterial RecJ. Intriguingly, a protein (SSO0295p) predicted to have a DNA-binding domain similar to that in RecJ, co-purified with the GINS complex from *Sulfolobus solfataricus* (31). RecJ plays a role in stalled replication fork activation in *E. coli* (76), suggesting that SSO0295p and TK1252p may similarly help in maintaining replication fork progression. SSO0295p and TK1252p are not, however, related proteins. These observations suggest that the eukaryotic GINS complex may also associate with an as-yet-unidentified nuclease.

(iii) Rad50 interactions. Eukaryotic Rad50 is part of a complex with Mre11 and Nbs1 that is required for double-strand DNA break repair [reviewed in reference (77)] and also plays a role during replication. This complex may help prevent replication fork-associated damage by serving as a scaffold that maintains the fork during replication pauses [for example, see reference (78)]. *T. kodakarensis* Rad50 (TK2211p) was co-isolated in complexes using His_6-tagged MCM1, primase, and RPA2 (Appendix 2). This is consistent with Rad50 also being present in the archaeal replisome and participating in a replication-related function in both eukaryotes and archaea. An interaction of eukaryotic Rad50 and RPA has also been reported (79), and based on the results obtained with *T. kodakarensis* Rad50, it seems reasonable to predict that eukaryotic Rad50 also interacts with helicase and primase.

(iv) MCM interactions. The MCM proteins are generally considered to function as replicative helicases (15,80), but the co-isolation of both Rad50 and MutS (TK0682p) with His_6-tagged MCM1 (Appendix 2) predicts that the MCM proteins may also participate in DNA repair.

(v) TK1046p interactions. Homologues of TK1046p are present in all three domains. The function(s) of this large protein (147.4 kDa) is unknown, although it does share some sequence similarity with nucleases and it is predicted to have an OB fold, a motif often used for nucleic acid recognition. TK1046p was co-isolated in complexes using His_6-tagged Fen1, GINS15, MCM3, PCNA2, PolB, RFC-L, RPA2, and TK1792p with very high MASCOT scores (Appendix 2). This large number of interactions with known replisome enzymes argues strongly that TK1046p is a component of the replication machinery. By extrapolation from the OB fold

prediction, TK1046p may be the first recognized example of a conserved nuclease that participates in DNA replication in all three domains.

Evidence against a putative archaeal replisome component.

TK1410p interacts with the exosome. TK1410p is predicted by sequence similarity to be related to the bacterial primase DnaG, and limited primase activity has been reported for a recombinant version of the TK1410p homologue from *S. solfataricus* (SSO0079p) (81). These observations suggested that this protein might be part of the replisome, but the complexes isolated using His_6-tagged TK1410p did not contain any known replisome proteins, but rather components of the exosome (TK1633p and TK1634p) were isolated. TK1410p was similarly not present in any complex isolated using a known His_6-tagged replication protein. Consistent with TK1410p being a part of the exosome, purified exosomes and exosome-containing membrane fractions from *S. solfataricus* also contain SSO0079p (43,44). Taken together, the results argue that TK1410p participates in exosome activity rather than in DNA replication.

Discussion

The interaction network (Figure 2-4) and the interactions listed in Appendix 1 and in Appendix 2 were documented using a systematic approach to isolate and identify all of the proteins that co-purified with known or predicted archaeal replisome components in *T. kodakarensis*. All of the *T. kodakarensis* strains that synthesized His_6-tagged proteins grew at the wild-type rate in all of the media tested, minimizing any concerns for the accumulation of aberrant structures or assembly into

nonnative complexes. To ensure the same regulation and expression levels, the genes encoding the His_6-tagged proteins were expressed from the native chromosomal locations using the wild-type gene expression signals. The results reported provide *in vivo* confirmation and validation of archaeal replication protein interactions previously documented *in vitro* and experimental evidence for several previously unrecognized replisome interactions that likely contribute to archaeal and potentially, by extrapolation, also to eukaryotic replication fork assembly, maintenance, and function.

What is the role of Cdc6-PCNA interaction?

The archaeal Cdc6 proteins bind to the origin of replication, where they are thought to direct the DNA strand separation needed for the initiation of DNA replication and also to recruit other components of the replisome to the origin of replication [summarized in (67,82)]. Thus, they are functional homologues of bacterial DnaA proteins. The complexes isolated from *T. kodakarensis* using His_6-tagged Cdc6 contained PCNA1, providing the first direct experimental support for an archaeal Cdc6-PCNA interaction, an observation that may be of major significance. A regulatory event known as the regulatory inactivation of DnaA (RIDA) ensures that the *E. coli* chromosome is replicated only once per cell cycle [reviewed in (83,84)]. RIDA stimulates the hydrolysis of the active replication initiator ATP-DnaA complex, resulting in inactive ADP-DnaA complexes. The β subunit of PolIII (the functional homologue of PCNA) and the homologous-to-DnaA (Hda) protein are required for this regulation [reviewed in references (83,84)]. The co-isolation of PCNA1 and Cdc6 is consistent with a mechanism similar to RIDA existing in Archaea. Hda belongs to

the AAA+ family of ATPases and has sequence similarity to the ATPase region of DnaA. As there is no identifiable archaeal Hda homologue, one of the proteins that co-isolated with His_6-tagged Cdc6 or PCNA proteins may embody the Hda function.

What is the function of the GINS complex?

In eukaryotes, Cdc45, MCM, and GINS form a tight complex (referred to as the CMG complex) that moves with the replication fork and is thought to function as the replicative helicase. The GINS complex also interacts with the Polα-primase complex, which is responsible for primer synthesis on the lagging strand [reference (23) and references therein]. To date, no archaeal homologue of Cdc45 has been identified but several proteins, and so potential candidates for Cdc45 functional homologues, co-purified with His_6-tagged GINS15 or GINS23, including TK0569p, TK1046p, and TK1186p, which also co-purified with His_6-tagged primase (Figure 2-4; Appendix 2). It has also been proposed that the GINS proteins maintain the integrity of the replisome by linking the replicative polymerase, primase, and helicases, but a direct interaction of GINS with DNA polymerase has not been documented. In Archaea, GINS was previously shown to interact with primase and MCM (31,79) but not with DNA polymerase. The results now reported (Figure 2-4; Appendix 2) confirm that both subunits of PolD form a complex with His_6-tagged GINS15 and both PCNA1 and PCNA2 interact with His_6-tagged GINS23. When added to the previously reported interactions, these results add substantial experimental support to the hypothesis that GINS functions as the center of the replisome, linking the polymerase, helicase, and primase components.

Are the archaeal replication proteins modified by small proteins?

In eukaryotes, the activities of PCNA and MCM are modulated by ubiquitination and sumoylation [reviewed in (85,86)]. A small protein similar in size to ubiquitin (~8 kDa; TK0808p) was consistently co-isolated with His_6-tagged PCNA1, and a second similarly sized protein (~8.5 kDa; TK0590p) was co-isolated in complexes using His_6-tagged Fen1, MCM1, MCM2, and MCM3 (Appendix 2). Currently, very little is known of protein modification in Archaea (87-89), but it seems possible that TK0590p and/or TK0808p could form protein conjugates that regulate archaeal replication as does ubiquitin and SUMO modification of replication proteins in eukaryotes. Some support for this notion is provided by the observation that PCNA in *Haloferax volcanii* is stabilized by proteosome disruption (90).

What are the roles of the three MCM proteins in *T. kodakarensis*?

MCM is a hexameric complex that assembles at the leading edge of the replication fork and unwinds the two DNA strands ahead of the replicative polymerase (15,32,33). In eukaryotes, MCM is a heterocomplex of six different polypeptides (MCM2 through MCM7). Most of the archaeal species studied in detail to date contain only one MCM polypeptide that assembles to form a homohexamer. Recently, some Archaea have been identified (35,91,92) with several MCM homologues that are thought to have resulted from gene duplication and/or lateral gene transfer from other Archaea (2,35,64,92).

T. kodakarensis has three genes (TK0096, TK1361, and TK1620) encoding MCM homologues, MCM1, MCM2, and MCM3, respectively, that could assemble to form three different MCM homohexamer complexes and/or many different MCM

heterohexamer complexes. The co-isolation results argue for the assembly of only homohexameric MCM complexes. MCM2 and MCM3 were not co-isolated with His_6-tagged MCM1, MCM1 and MCM3 were not co-isolated with His_6-tagged MCM2, and MCM1 and MCM2 were not co-isolated with His_6-tagged MCM3. The results obtained are consistent with both MCM1 and MCM2 being part of the replisome, and based on the similarity of their interactions, they may be functionally redundant. Ishino et al. and Pan et al. demonstrated that only MCM3, but not MCM1 or MCM2, is essential for the viability (34,93). The genes coding MCM1 and MCM2 located in an region of the *T. kodakarensis* genome where are believed to be of viral origin (2,35). In contrast, the results argue that MCM3 participates in complexes that differ from those formed by MCM1 and MCM2 (Figure 2-4; Appendix 2). Only proteins with unknown functions were co-isolated using His_6-tagged MCM3, and MCM3 was never co-isolated with a known His_6-tagged replication enzyme. MCM3 appears to be a member of the McmD group (92), one of the two groups of MCM proteins conserved within the order Methanococcales that overall contain four to eight MCM homologues. In eukaryotes, MCM homologues are thought also to participate in transcription, DNA repair, and chromatin remodeling and it seems possible that the archaeal McmD group of MCM proteins might similarly participate in one or more of these processes in Archaea, rather than in DNA replication.

Chapter 3: A Novel DNA Nuclease is Stimulated by Association with the GINS Complex

Introduction

Chromosomal DNA replication has many universally conserved features, but there are differences in the proteins and complexes that initiate and maintain DNA replication forks in Bacteria, Archaea and Eukarya. Many of the proteins required for bacterial and eukaryal replication have been isolated and characterized extensively, while most of the components of the archaeal replication machinery have only been putatively identified by bioinformatics. When archaeal proteins with sequences in common with bacterial and/or eukaryal replisome proteins have been investigated, the results obtained have generally confirmed their predicted replication functions (63,67,82). This *in silico* approach, however, does not readily identify archaeal-specific replisome proteins. To address this limitation, 20 *T. kodakarensis* strains were constructed, each of which synthesized an established archaeal replication protein with an amino- or carboxy-terminal hexahistidine extension (His_6-tag). These proteins were purified directly by nickel-affinity from *T. kodakarensis* cell lysates, and all proteins that were consistently co-isolated with each His_6-tagged replication protein were identified (Chapter 2). One of the co-purified proteins (encoded by TK1252) contains a conserved nuclease domain, providing a direct indication of its involvement in DNA metabolism. For this reason, we decided to select TK1252 encoded protein as the first candidate to perform further research. Here, we report the characterization of this novel archaeal nuclease that was present in the complexes

isolated by nickel-binding of His_6-tagged subunits of the *T. kodakarensis* GINS complex and the archaeal-specific PolD.

In Eukarya, the heterotetramer GINS complex associates with the mini-chromosome maintenance (MCM) proteins, Mcm2-7 and with Cdc45 to form the Cdc45, Mcm2-7, GINS (CMG) complex. This complex has a 3' → 5' DNA helicase activity and is thought to function as the replicative helicase (40,75,94). The GINS complex is required to establish and maintain replication forks (95-97) and also interacts with the Polα-primase complex that synthesizes primers on the lagging strand (23,75). With these features, the eukaryal GINS complex appears to be the functional homologue of the τ subunit (DnaX) of the *E. coli* replisome that binds to the bacterial replicative DNA polymerase (PolIII), DNA helicase (DnaB) and primase (DnaG) at the replication fork and coordinates leading and lagging strand syntheses (27). Sequence homologies predict that many Archaea, including *T. kodakarensis*, have a GINS complex assembled from two molecules each of GINS15 (TK0536p) and GINS23 (TK1619p), proteins most closely related to the eukaryal Psf1 and Sld5, and Psf2 and Psf3 proteins, respectively (22,23). Consistent with GINS being an archaeal replisome component, investigations of [$GINS15_2$-$GINS23_2$] complexes from several Archaea have documented interactions with the archaeal primase, MCM, PolD and PCNA (29-31,98).

In the *T. kodakarensis* genome annotation, the protein encoded by TK1252 is predicted to be a single-strand specific nuclease (2). The results reported here confirm that this protein does associate with the GINS complex, specifically with the GINS15 component, and demonstrate that it is a single-strand (ss) DNA-specific 5' → 3'

exonuclease. The exonuclease activity of this protein, designated GINS-associated nuclease (GAN), is stimulated by its interaction with GINS15. Possible roles for the GAN-GINS association during archaeal DNA replication are discussed.

Materials and Methods

Nuclease substrates.

[γ-^{32}P]ATP was purchased from Perkin Elmer. Unlabeled, Cy3- and Cy5-labeled deoxy- and ribo-oligonucleotides, with the sequences listed in Appendix 4, were obtained from the NIST/UMD nucleic acids synthesis facility. Double-strand (ds) DNA substrates were generated by annealing complementary oligonucleotides followed by PAGE purification, as previously described (99). To obtain linear and circular 200-mer substrates, 1.5 nmol of the 100-mer oligonucleotides A and B (Appendix 4) were phosphorylated by incubation with 40 U of T4 polynucleotide kinase for 1 h at 37°C. The phosphorylation reaction mixture for oligonucleotide B also contained 71 pmol of [α-^{32}P]ATP. To construct the linear substrate, 0.5 nmol of phosphorylated oligonucleotides A and B plus 2.5 nmol of the bridge oligonucleotide AB (Appendix 4) were mixed in 20 mM HEPES (pH 7.5), 150 mM NaCl, heated to 100°C and the mixture was then allowed to cool slowly to 22°C. This procedure was also used to generate the circular substrate, except that the reaction mixture also contained 2.5 nmol of the bridge oligonucleotide BA (Appendix 4). The reaction mixtures were placed at 16°C, 8000 U of T4 DNA ligase (Fermentas) were added and incubation continued for 14 h. The reaction products were separated by electrophoresis at 15 W for 75 min through 10% (w/v) polyacrylamide-8 M urea gels

run in TBE. The regions of the gel containing the desired 200-mer linear and circular ssDNAs were excised and the DNAs eluted from the gel into 0.5 M ammonium acetate, 10 mM magnesium acetate, 1 mM EDTA, ethanol precipitated and dissolved in 30 μl of TE (pH 8.0). The resulting solution was passed through a S300 mini-column filter (GE Healthcare).

Plasmid construction

For protein expression in *E. coli*, the genes encoding GAN (TK1252), GINS15 (TK0536) and GINS23 (TK1619) were PCR-amplified from *T. kodakarensis* genomic DNA using primers (listed in Appendix 5) that added an in-frame His_6-encoding sequence to the 3'-terminus of the amplified gene. The amplified DNAs were ligated with pET15b (TK1252) or pET21a (TK0536 and TK1619) linearized by digestion with the restriction enzyme listed in Appendix 5. A plasmid that directed the synthesis of GAN (D34A) was generated by site-specific mutagenesis from the plasmid that expressed TK1252 by using a QuikChange mutagenesis kit (Stratagene) using oligonucleotides with the sequences listed in Appendix 5.

To construct a *T. kodakarensis* strain that synthesized GAN-His_6 *in vivo*, TK1252 and DNA from immediately upstream and downstream of TK1252 were separately amplified from *T. kodakarensis* genomic DNA. An overlapping PCR was used to add an His_6-encoding sequence in-frame to the 5'-terminus of TK1252 (65). The three amplified DNAs were cloned into pUMT2 (3) adjacent to *trpE* (TK0254) to generate plasmid pZLE034 (Figure 3-1). In pZLE034, TK1252-His_6 is positioned between genomic sequences that are homologous to the DNA immediately upstream and downstream of TK1252 in the *T. kodakarensis* KW128 genome. An aliquot of

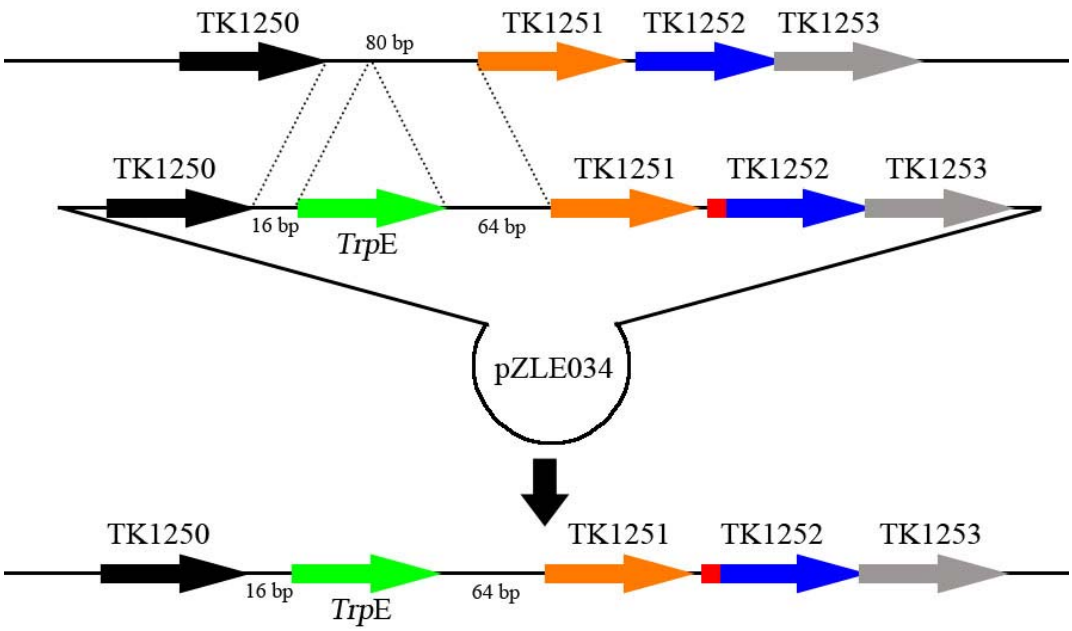

Figure 3-1: Plasmid pZLE034 used to transform *T. kodakarensis* to obtain the synthesis of the His_6-tagged GAN *in vivo*.

The structure of plasmid pZLE034 used as donor DNA to transform *T. kodakarensis* KW128 is shown above the genomic locus resulting from by homologous recombination in the regions flanking TK1252. The GAN encoding gene (TK1252) is shown as a blue arrow with the His_6-encoding sequence indicated by a red box. Transformation of *T. kodakarensis* KW128 with the constitutively expressed *trpE* (TK0254) gene (green arrow) confers tryptophan independent growth and was used as the positive selection to isolate the desired His_6-GAN expressing transformant.

pZLE034 DNA was used to transform *T. kodakarensis* KW128 (*ΔpyrF; ΔtrpE::pyrF*) as previously described (3,66) and transformants were selected by growth on plates lacking tryptophan. The desired replacement of TK1252 with the GAN-His_6 encoding gene was confirmed in a representative transformant, designated *T. kodakarensis* 34-5, by diagnostic PCR and sequencing (29).

Recombinant protein purification

The plasmids encoding GAN, GAN (D34A), GINS15 or GINS23 were

transformed into *E. coli* BL21 (DE3)-CodonPlus-RIL (Stratagene). Isopropyl-β-D-thiogalactopyranoside induction, expression at 16°C for 16 h and purification of the recombinant N-terminal His_6-tagged GAN and GAN (D34A), and C-terminal His_6-tagged GINS15 and GINS23 from *E. coli* cell lysates by Ni^{2+} affinity chromatography were carried out as previously described (100). Aliquots of the purified proteins were stored at -80°C.

Size exclusion chromatography

Aliquots of each experimental protein (100 μg) or protein mixture and Gel Filtration standards (Bio-Rad) were diluted in 200 μl of 25 mM Tris-HCl (pH 7.5), 100 mM NaCl, 10% (v/v) glycerol and loaded onto a Superdex-200 column (HR10/30; GE Healthcare) pre-equilibrated in the same buffer. Fractions (250 μl) were collected from the column at a flow rate of 0.5 ml/min. The proteins present in aliquots (80 μl) of each fraction were separated by electrophoresis through a 12% (w/v) polyacrylamide-SDS gel and stained with Coomassie brilliant blue (R250).

Nuclease assays

Unless otherwise noted in the figure legends, the nuclease assay reaction mixtures (20 μl) containing the DNA substrate, BSA (125 μg/ml), 25 mM Tris-HCl (pH 7.5), 2 mM $MnCl_2$ and GAN, were incubated at 70°C for 20 min. Nuclease digestion was stopped by adding 20 μl of 95% formamide, 0.1× TBE, 10 mM EDTA and incubation at 100°C for 2 min. The digestion products were visualized and quantified by phosphorimaging after electrophoretic separation through 20% (w/v) polyacrylamide-8 M urea gels run in TBE for 1.25 h at 15 W. For native gels nuclease

reactions were stopped by adding 5 μl of 50% glycerol, 20 mM EDTA. The digestion products were visualized and quantified by phosphorimaging after electrophoretic separation through 20% (w/v) polyacrylamide gels run in TBE for 2 h at 300 V.

Liquid chromatography-mass spectrometry

Aliquots (10 μM) of the DNA templates, 5'-AAAAAAGG and 5'-GGAAAAAA, were incubated in reaction mixtures (50 μl), with or without 20 pmol GAN, for 1 h at 70°C in a buffer containing 5 mM ammonium formate (pH 6.5), 2 mM $MnCl_2$. The products were subjected to liquid chromatography (LC)/mass spectrometry (MS) analyses using the negative ion mode with a Finnigan LTQ ion trap mass spectrometer (San Jose, CA, USA) equipped with nanospray ionization (NSI) interface coupled to an Agilent 1200 HPLC system (Palo Alto, CA, USA). The flow from the Aligent pump was split from 0.85 ml to 25 nl/min using a 75 μm internal diameter (ID) silica capillary as the flow splitter. Separations were performed using 50 μm ID silica capillary columns (Polymicro Technology, Phoenix, AZ, USA) with in-house made frit packed with 15 cm of 3 μm Atlantis T3 C18 aqueous reversed phase particles (Waters, Milford, MA, USA). The mobile phase A was 5 mM ammonium formate (pH 6.0) in water, and the mobile phase B was 5 mM ammonium formate in methanol. Analytes were eluted over 30 min using a 0-95% linear gradient of solvent B. The heated capillary was at 200°C. Fragmentation was activated by collision-induced dissociation of 35%. Selective reaction monitoring was implemented with the following transitions: dAMP: 330.1 to 195.1 m/z; dGMP: 346.1 to 195.1 m/z; dAMP-dAMP: 643.1 to 330.1 m/z; and dGMP-dGMP: 675.1 to 346.1 m/z. The instrument control, data acquisition, and data analysis were performed by

Xcalibur software (Thermo Electron Corporation, version 2.0.7 SP1).

Isolation and identification of His_6-tagged GAN and associated proteins from *T. kodakarensis*

T. kodakarensis 34-5 cultures (5 liter) were grown to late exponential phase (OD_{600} of 0.8) at 80°C in MA-YT medium supplemented with 5 g sodium pyruvate/l in a BioFlow 415 fermentor (New Brunswick Scientific). The cells were harvested by centrifugation, resuspended in 30 ml of buffer A [25 mM Tris-HCl (pH 8.0), 500 mM NaCl, 10 mM imidazole and 10% glycerol] and lysed by sonication. After centrifugation, the resulting clarified lysate was loaded onto a 1 ml HiTrap chelating column (GE Healthcare) pre-equilibrated with $NiSO_4$. The column was washed with buffer A and proteins were eluted using a linear imidazole gradient from buffer A to 67% buffer B [25 mM Tris-HCl (pH 8.0), 100 mM NaCl, 150 mM imidazole and 10% glycerol]. Fractions that contained the GAN protein were pooled and dialyzed against buffer C [25 mM Tris-HCl (pH 8.0), 500 mM NaCl, 0.5 mM EDTA, 2 mM DTT]. Aliquots (30 μg) of the protein were precipitated by adding trichloroacetic acid (TCA; 15% final concentration). The TCA-precipitated proteins were identified by multi-dimensional protein identification technology at the Ohio State University mass spectrometry facility (http://www.ccic.ohio-state.edu/MS/proteomics.htm) using the MASCOT search engine. The protein isolation and mass spectrometry analyses were also repeated twice using lysates from two independent cultures of *T. kodakarensis* KW128. These control experiments identified the *T. kodakarensis* proteins that bound and eluted from a Ni^{2+} charged matrix in the absence of a His_6-tagged protein. All proteins co-isolated with His_6-GAN by binding to Ni^{2+} matrix from *T. kodakarensis*

34-5 cell lysates that had high MASCOT scores, and that were not present in the control samples, are listed in Table 3-1.

Results

Purified GAN and GINS15 form complexes in solution.

GAN (TK1252p) was co-isolated with His_6-GINS15 (TK0536p) from *T. kodakarensis* cell lysates by His_6-GINS15 binding to a Ni^{2+} charged matrix followed by imidazole elution consistent with GAN forming a stable complex with GINS15 *in vivo* (Chapter 2). To determine if these proteins also interacted *in vitro*, recombinant GAN and GINS15 were mixed and the products examined by size exclusion chromatography. In the absence of GINS15, the GAN (52.9 kDa) elution profile was consistent with the presence of monomers and $(GAN)_2$ dimers (Figure 3-2A; elution peaks in fractions 53 and 59). GINS15 (21.5 kDa) alone eluted almost exclusively at a position consistent with a $(GINS15)_2$ dimer (Figure 3-2B, elution peak in fraction 56), as reported previously for GINS15 from *S. solfataricus* (31). When incubated together, GAN and GINS15 interacted to form several complexes that eluted in fractions consistent with the formation of complexes larger than $(GAN)_2$ and $(GINS15)_2$ dimers (Figure 3-2E, elution peaks in fractions 38 and 47). Incubation of GAN with GINS23 (19.2 kDa) did not result in the formation of larger complexes (Figure 3-2F). GAN bound the [$GINS15_2$-$GINS23_2$] complex (Figure 3-2G), suggesting that the GAN-GINS15 interactions did not disrupt the GINS complex (30,31).

Table 3-1. Proteins co-purified with GAN.

Gene #	Score	MW (Da)	Peptide matches	Percent coverage	Function
TK1903	1354	150190	54	19.5	PolD-L
TK1902	947	80848	43	28.4	PolD-S
TK1252	746	52858	26	37.1	GAN
TK1619	155	19154	6	14.6	GINS23
TK1637	105	29250	4	13.8	proteasome subunit alpha
TK1496	97	22992	3	12.9	30S ribosomal protein S2
TK0536	87	21583	2	14.9	GINS15
TK1748	81	125718	3	3.5	isoleucyl-tRNA synthetase
TK0847	79	8671	4	38.6	hypothetical protein TK0847
TK2217	79	44001	2	3.3	2-amino-3-ketobutyrate coenzyme A ligase
TK0593	73	45867	3	6.7	Unknown
TK2157	54	36667	3	6	Unknown
TK2106	50	46763	2	3.5	phosphopyruvate hydratase
TK1940	48	39764	3	11.2	small-conductance mechanosensitive channel
TK0171	48	41290	2	2.8	Unknown
TK0566	45	96249	2	1.6	DEAD/DEAH box RNA helicase
TK0714	43	74602	3	2.7	iron(II) transport protein B
TK2253	41	28743	13	3.2	Unknown
TK0470	40	198006	2	0.6	reverse gyrase
TK2276	39	23395	2	16.9	orotidine 5'-phosphate decarboxylase
TK1448	37	39518	2	2	5,10-methylenetetrahydrofolate reductase
TK2270	29	7456	2	30.2	Unknown
TK2255	28	48783	2	3	bifunctional phosphatase/dolichol-phosphate glucosyltransferase
TK0263	27	43206	2	3.5	3-phosphoshikimate 1-carboxyvinyltransferase

Proteins with at least two peptides matches are listed along with their molecular weight, MASCOT score, and the percentage of the amino acid sequence covered by the matching peptides. (See text for further details.)

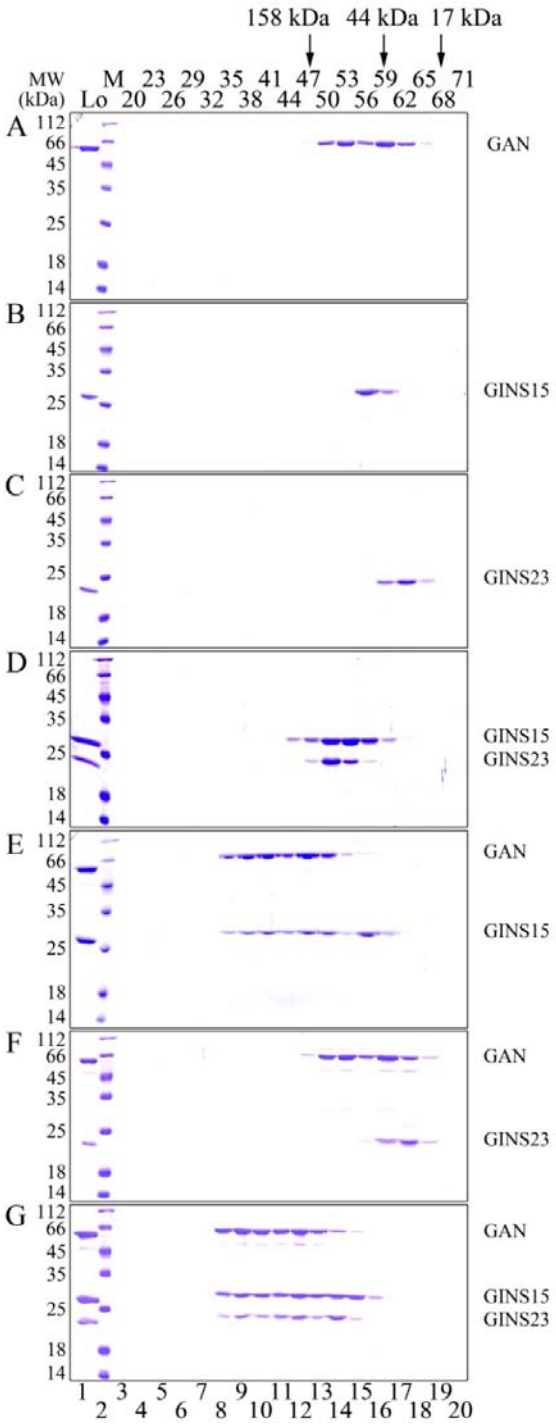

Figure 3-2: Gel-filtration assay indicates GAN interacts with GINS15.

A sample (100 μg) of each protein listed to the right of the corresponding panels (A through G) was subjected to Superdex-200 gel filtration analysis. Aliquots (80 μl) from each fraction were separated by electrophoresis through 12% polyacrylamide-

SDS gels and stained with Coomassie brilliant blue (R-250). The fractions in which γ-globulin (158 kDa), ovalbumin (44 kDa) and myoglobin (17 kDa) eluted are noted at the top of the figure.

GAN is a ssDNA nuclease.

Based on limited sequence similarities between GAN and *E. coli* RecJ (Figure 3-3), GAN was predicted to be a ssDNA nuclease (2). As shown in Figure 3-4A, this was confirmed. A Cy3-labeled single stranded deoxy-oligonucleotide (30-mer) was fully digested by GAN in the presence of Mn^{2+} (Figure 3-4A; lane 2) and limited activity was also observed with Mg^{2+} (Figure 3-4A; lane 8). There was no activity in the absence of metal ions, or with Zn^{2+}, Li^{+} or Ca^{2+} present. An aspartate, known to be important for RecJ activity (101), is conserved in the GAN-RecJ alignment (Figure 3-3, marked by an asterisk). This was replaced by alanine, and the GAN (D34A) variant had minimal nuclease activity with Mn^{2+} present (Figure 3-4A; lane 3) and no detectable activity with Mg^{2+} (Figure 3-4A; lane 9). An *E. coli* RecJ variant with the analogous alanine for aspartate replacement similarly retained a residual nuclease activity, but had 400-fold lower activity than wild-type RecJ (101).

As expected, digestion of the 30-mer ssDNA substrate was dependent on the GAN concentration (Figure 3-4B and C) and time of incubation at 70°C (Figure 3-4D and E) and, under all reaction conditions, the ssDNA substrate was digested almost exclusively to mononucleotides as determined by MS analysis (Figure 3-5). Consistent with GAN degrading the ssDNA molecules processively, there was no detectable accumulation of intermediate length oligonucleotides (Figure 3-4B and D). Providing further support for this conclusion, essentially the same pattern of digestion products was obtained when the 30-mer DNA was Cy3-labeled at the 3'-terminus or

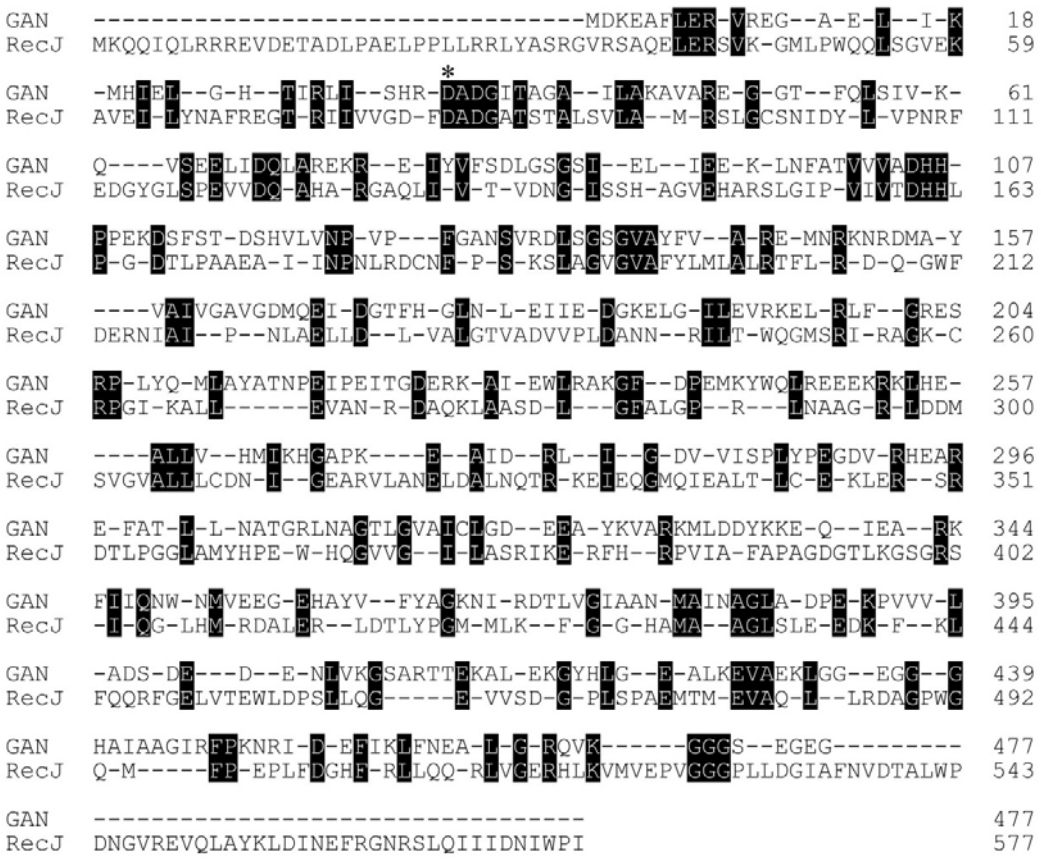

Figure 3-3: Alignment of the amino acid sequence of the *T. kodakarensis* GAN and *E. coli* RecJ.

T. kodakarensis GAN (TK1252: GI:57641187); *E. coli* RecJ (GI:16130794). The Asp residue that was replaced by Ala in the GAN (D34A) variant is marked by an asterisk.

Cy5-labeled at the 5'-terminus (Figure 3-4C and E). As illustrated in Figure 3-4E, the 30-mer ssDNA was degraded by GAN at a rate of 36.9 ± 13.3 nt/mol/min. Although the monomer-dimer equilibrium remains to be determined, both GAN monomers and dimers appeared to be active (Figure 3-6).

GAN acts as 5'-exonuclease on ssDNA.

GAN was incubated with the nucleic acids illustrated in Figure 3-7A to

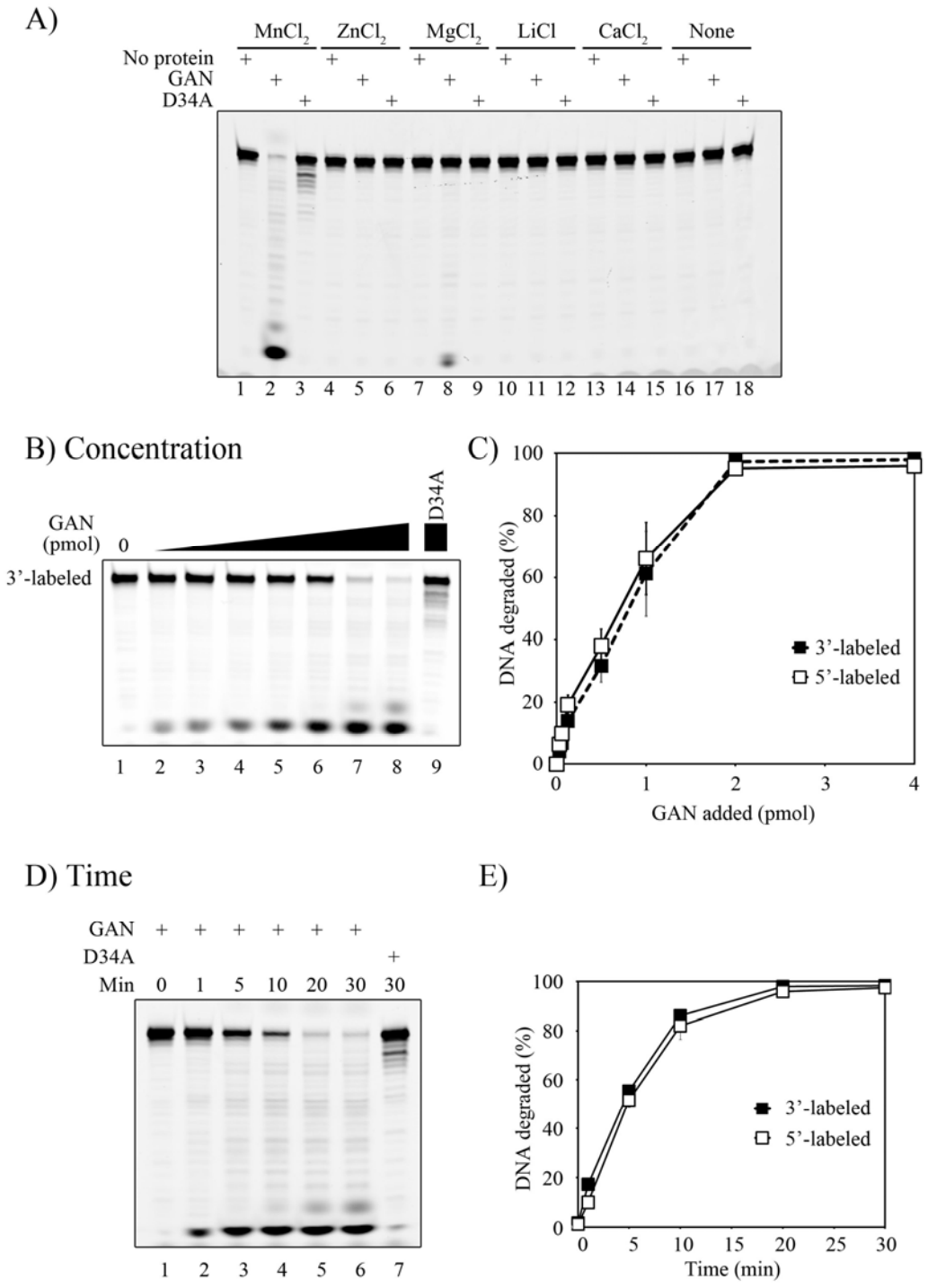

Figure 3-4: GAN is a Mn^{2+}-dependent exonuclease.

(A) Electrophoretic separation of the products of reaction mixtures (20 µl) that contained 7.5 pmol 3'-Cy3-labeled oligonucleotides (A4, Appendix 4), 4 pmol of GAN or GAN (D34A), 25 mM Tris-HCl (pH 7.5), 125 µg/ml BSA and 2 mM $MnCl_2$, $ZnCl_2$, $MgCl_2$, LiCl or $CaCl_2$, that were incubated at 70°C for 20 min. (B) Reaction

mixtures (20 µl) containing 7.5 pmol of 3'-Cy3 labeled oligonucleotides (A4, Appendix 4) in 25 mM Tris-HCl (pH 7.5), 2 mM $MnCl_2$, 125 µg BSA/ml and 0, 0.06, 0.125, 0.25, 0.5, 1, 2 and 4 pmol (lanes 1-8) of GAN protein. In lane 9, 4 pmol of GAN (D34A) was added in place of wild-type GAN. The reaction mixtures were incubated at 70°C for 20 min and the products were separated by electrophoresis, visualized and quantified by phosphorimaging. (C) Quantification of the A4 digestion products shown in (B), and of digestion products generated from 5'-Cy5 labeled oligonucleotide (A1, Appendix 4) under the same reaction conditions. The data shown are the averages, with standard deviations, from three independent experiments with each substrate. (D) Separation of the reaction products generated in reaction mixtures (20 µl) that contained 10 pmol of 3'-Cy3-labeled oligonucleotide (A4, Appendix 4) 1 pmol GAN or GAN (D34A), 2 mM $MnCl_2$ and 125 µg/ml BSA in 25 mM Tris-HCl (pH 7.5) incubated at 70°C for 0, 1, 5, 10, 20 and 30 min. (E) Quantification of the A4 digestion products shown in (D), and of digestion products generated from 5'-Cy5 labeled oligonucleotide (A1, Appendix 4) under the same reaction conditions. The data shown are the averages, with standard deviations, from three independent experiments with each substrate.

determine the direction of exonuclease digestion and substrate specificity. No nuclease activity was detected with linear dsDNA (substrate III; Figure 3-7A and B). This inability to degrade dsDNA provided an assay to determine the direction of GAN digestion of ssDNA. There was no digestion of a 3'-ssDNA extension from a dsDNA molecule (substrate II; Figure 3-7 and B) whereas a 5'-ssDNA extension was rapidly hydrolyzed (substrate I; Figure 3-7 and B). Digestion of the 5'-ssDNA extension was then followed by a much slower separation and degradation of the two strands of the dsDNA. In all experiments, DNA hydrolysis yielded mononucleotides (Figure 3-7A; lane 4; Figure 3-7C and D). Essentially the same results were obtained when the substrates were [^{32}P]-labeled rather than dye-labeled. Based on the results obtained, GAN is a 5' → 3' ssDNA-specific exonuclease, and given the almost simultaneous release of a label attached to either the 5'- or 3'-terminus of a ssDNA substrate (Figure 3-4B and D), the initial binding of GAN to the 5'-terminus is most likely the rate limiting step in the degradation of these substrates.

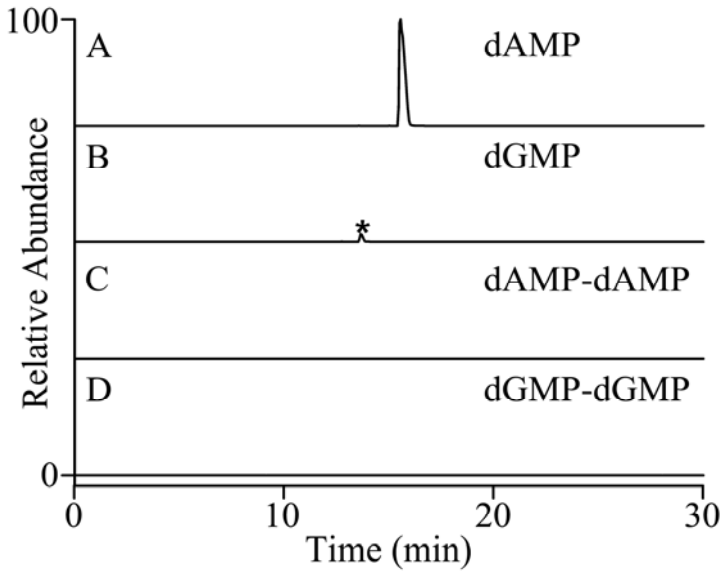

Figure 3-5: GAN nuclease produces mononucleotides.

Reconstructed ion chromatograms of the products of GAN digestion of 5'-AAAAAAGG identified by HPLC-MS/MS analysis. The products were evaluated specifically for the presence of (A) dAMP, (B) dGMP, (C) dAMP-dAMP and (D) dGMP-dGMP. The ratio of dAMP to dGMP was 9:1 and deviated from the expected 3:1 ratio due to differences in the response factor/sensitivities of dAMP and dGMP. Asterisk indicates peak detection with signal to noise ratio > 20. The absence of dinucleotides, dAMP-dAMP (C) and dGMP-dGMP (D), are indicated by the traces. MS/MS transitions are: dAMP (m/z 330.1 → 195.1), dGMP (m/z 346.1 → 195.1), dAMP-dAMP (m/z 643.1 → 330.1) and dGMP-dGMP (m/z 675.1 → 346.1). Essentially, the same results were obtained when the substrate was 5'-GGAAAAAA. (Experiment performed by Wei Yuan and James L. Edwards.)

To further confirm the requirements for nuclease activity, experiments were undertaken using linear and circular versions of a 200-mer ssDNA substrate (Figure 3-8A and B). When linear, this oligonucleotide was readily degraded by GAN but, when circular, remained intact. A small amount of linear 200-mer present in the circular 200-mer preparations was degraded by exposure to GAN (Figure 3-8A). As controls, the 200-mer substrates were incubated with a mixture of *E. coli* exonucleases I and III and, as expected, these well-characterized enzymes degraded

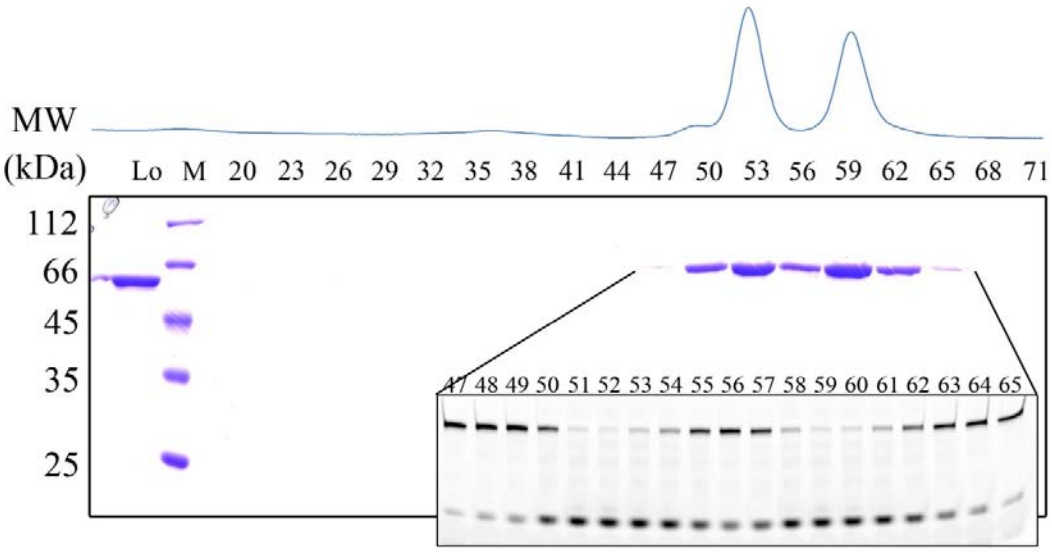

Figure 3-6: Monomers and dimers of GAN possess nuclease activity.

A GAN protein solution (100 μg) was subjected to Superdex-200 gel filtration analysis. The protein concentration in each eluted fraction was measured by ultraviolet absorption at 280 nm (top) and the proteins in aliquots (80 μl) from each fraction were separated by electrophoresis through 12% SDS-PAGE and stained with Coomassie Brilliant Blue (R-250). Aliquots (2 μl) of each indicated fraction (inset) were added to reaction mixtures that contained 7.5 pmol of 3'-Cy3 labeled oligonucleotides (A4, Appendix 4), 25 mM Tris-HCl (pH 7.5), 2 mM $MnCl_2$ and 125 μg/ml BSA. The reaction mixtures were incubated at 70°C for 10 min and the products generated were separated by electrophoresis and visualized by phosphorimaging.

only the linear substrate (Figure 3-8A). GAN appears to specifically degrade DNA as no degradation of RNA could be observed under all conditions tested (Figure 3-8C).

GINS15 stimulates the GAN nuclease activity.

To determine if GINS affected the activity of GAN, its nuclease activity was assayed in the presence or absence of GINS15, GINS23 or the [$GINS15_2$-$GINS23_2$] complex. The presence of GINS15 stimulated GAN nuclease activity (Figure 3-9; Figure 3-10A, lanes 1-5) but had no stimulatory effects on GAN (D34A) (Figure 3-

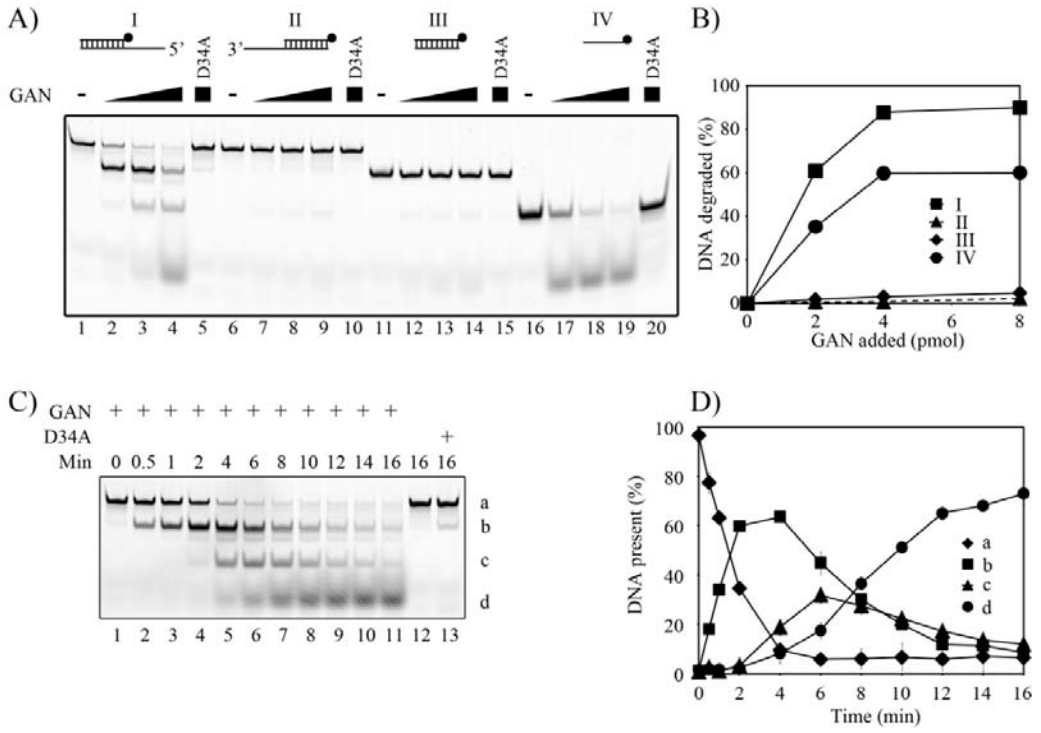

Figure 3-7: GAN is a 5' → 3' exonuclease.

(A) The structures of the DNA substrates, with the position of the Cy3-label indicated (dark circle), are illustrated above the electrophoretic separations of the GAN digestion products. Reaction mixtures (20 μl) that contained 7.5 pmol of Cy3-labeled substrate, 125 μg/ml BSA, 25 mM Tris-HCl (pH 7.5), 2 mM $MnCl_2$ and 0 (-), 2, 4 or 8 pmol of GAN or 8 pmol of GAN (D34A) were incubated for 15 min at 55°C. The reaction products were separated by electrophoresis through 20% native polyacrylamide gels, visualized and quantified by phosphorimaging. (B) Average values from three independent repetitions of the experiments shown in (A). (C) Native gel electrophoretic separation of the products of digestion of substrate I in reaction mixtures (20 μl) that contained 7.5 pmol of substrate I, 4 pmol of GAN (lanes 1-11) or GAN (D34A) (lane 13), 125 μg BSA/ml, 25 mM Tris-HCl (pH 7.5), 2 mM $MnCl_2$. The reaction mixtures were incubated at 55°C for 0, 0.5, 1, 2, 4, 6, 8, 10, 12, 14 or 16 min. (D) Average values, with standard deviations, from three independent repetitions of the experiments shown in C. The Cy3-labeled molecules indicated and quantified were a: substrate I (filled diamond); b: ds DNA product (filled square); c: ss oligonucleotides from substrate I (filled triangle); d: mononucleotides (filled circle).

10A; lanes 11 and 12) although GINS15 and GAN (D34A) were found to interact (Figure 3-11). GINS15 alone had no nuclease activity (Figure 3-10A; lane 6).

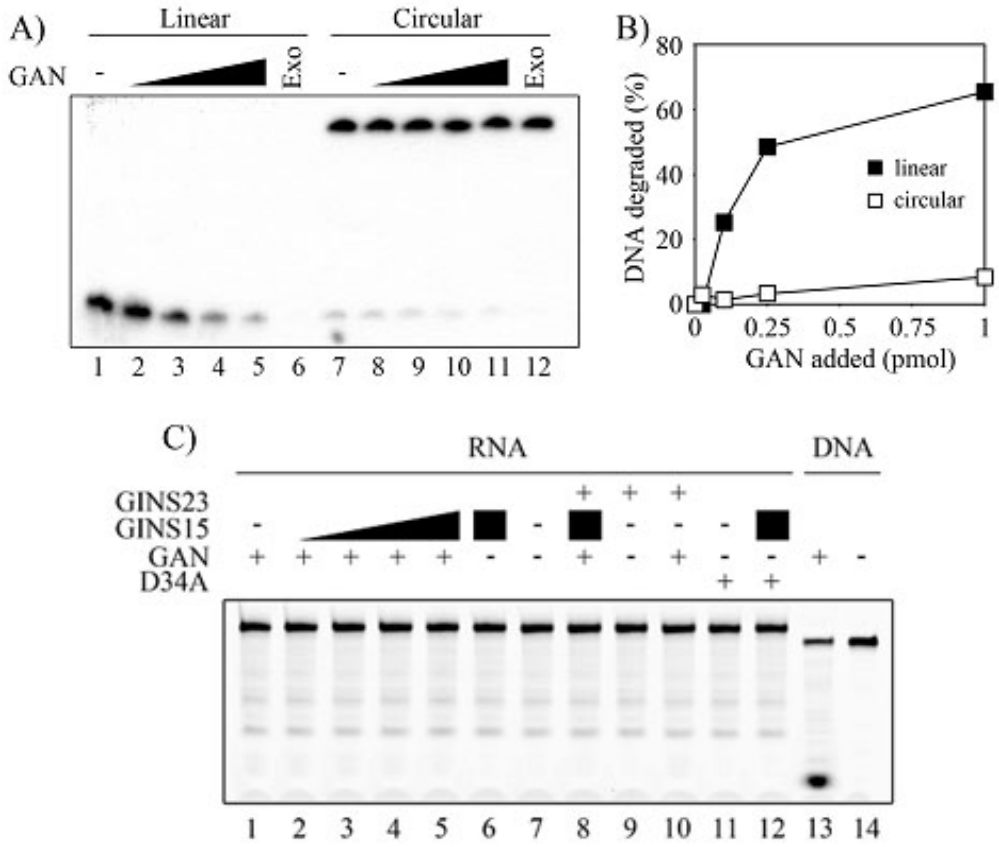

Figure 3-8: GAN is an ssDNA-specific exonuclease.

(A) Electrophoretic separation of the products of reaction mixtures (20 µl) that contained 0.75 pmol of [^{32}P]-labeled 200 nt linear or circular ssDNA, 25 mM Tris-HCl (pH 7.5), 2 mM $MnCl_2$ and 125 µg/ml BSA, and 0 (-), 0.025, 0.1, 0.25 and 1 pmol GAN incubated at 60°C for 10 min. The products of incubation of these substrates with a mixture of *E. coli* exonucleases I and III (10 and 50 units, respectively; Exo) for 30 min at 37°C were separated in lanes 6 and 12 (Experiment performed by Wiebke Chemnitz). (B) Quantification of the GAN digestion shown in (A). (C) Electrophoretic separation of the products of reaction mixtures (20 µl) incubated at 55°C for 10 min that contained 7.5 pmol of Cy3-labled RNA or DNA (A4R and A4, respectively, Appendix 4), 2 pmol of GAN or GAN (D34A), 125 µg/ml BSA, 25 mM Tris-HCl (pH 7.5), 2 mM $MnCl_2$ and 0.5, 1, 2 (lanes 2-4) or 4 (lanes 5, 6, 8, and 12) pmols of GINS15 or 4 pmols of GINS23 (lanes 8-10).

Incubation with GINS23 did not stimulate the GAN nuclease activity (Figure 3-10A; lane10) and the presence of [$GINS15_2$-$GINS23_2$] complexes had the same stimulatory

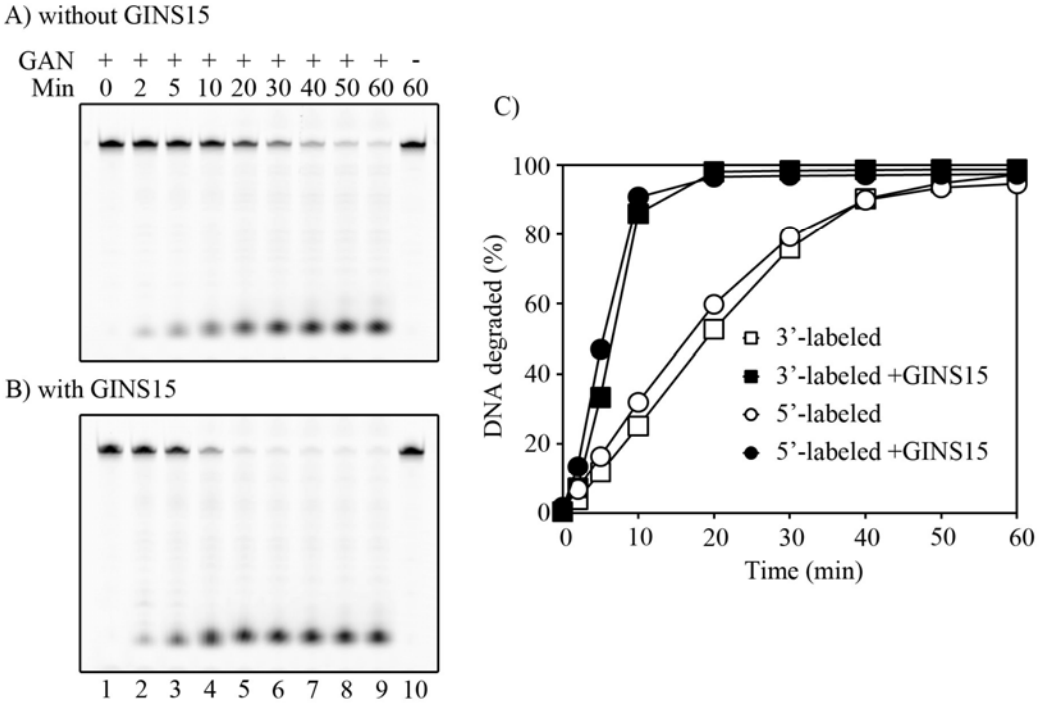

Figure 3-9: GINS15 stimulates GAN activity.

(A) and (B). Separation of the reaction products generated in reaction mixtures (20 µl) that contained 7.5 pmol of each 3'-Cy3 (A4, Appendix 4) and 5'-Cy5 (A1, Appendix 4) labeled oligonucleotides, 0.4 pmol GAN, 2 mM $MnCl_2$ and 125 µg/ml BSA in 25 mM Tris-HCl (pH 7.5) incubated at 70°C for 0, 2, 5, 10, 20, 30, 40, 50 and 60 min in the absence (A) or presence (B) of 0.8 pmol GINS15 protein. No GAN protein was added to lane 10. Only the gels with the 3'-Cy3 labels are shown. (C) Quantification of the digestion products shown in panels A and B. The data shown are the averages, with standard deviations, from three independent experiments with each substrate.

effect as the presence of GINS15 alone (Figure 3-10A; lanes 5 and 8). Incubation with GINS15 did not change the substrate specificity of GAN; with GINS15 present GAN still did not degrade dsDNA or RNA (Figure 3-8 C).

DNA polymerase D interacts with GAN *in vivo*.

GAN was identified as a protein that formed a stable complex*in vivo* with His_6-tagged GINS15 (Chapter 2). To determine if additional proteins formed complexes *in*

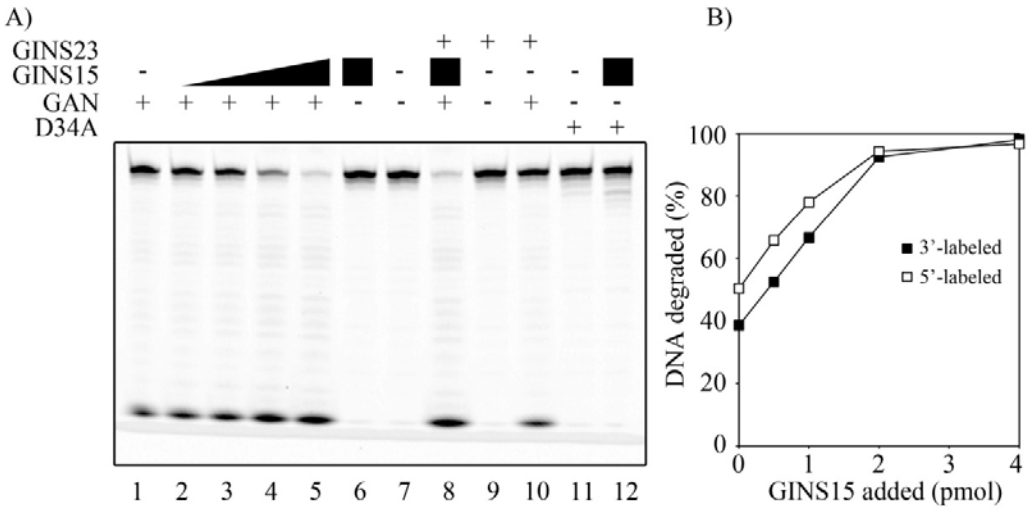

Figure 3-10: GAN nuclease activity is stimulated by GINS15.

(A) Electrophoretic separation of the products of digestion of a 3'-Cy3-labeled substrate (A4, Appendix 4) in reaction mixtures (20 µl) that contained 7.5 pmol of substrate, 125 µg BSA/ml, 25 mM Tris-HCl (pH 7.5), 2 mM $MnCl_2$ and, when indicated, 2 pmol GAN or GAN (D34A) in the presence of 0.5, 1, 2 (lanes 2-4) or 4 (lanes 5, 6, 8 and 12) pmol of GINS15, or 4 pmol GINS23 (lanes 8-10 and 12). The reaction mixtures were incubated at 70°C for 4 min. (B) Quantification of the A4 digestion products shown in (A), and of digestion products generated from 5'-Cy5-labeled oligonucleotides (A1, Appendix 4) under the same reaction conditions. The data shown are the averages from three independent experiments with each substrate.

vivo with GAN, a *T. kodakarensis* strain was constructed that synthesized His_6-tagged GAN (Figure 3-1) and proteins that co-purified with this tagged protein from cell lysates were identified by MS (Table 3-1). Both subunits of the GINS complex (GINS15 and GINS23) and both subunits of the euryarchaeal specific PolD were predominant among the co-purified proteins. This provides strong reciprocal evidence for the presence of a GAN-GINS complex *in vivo*, and is consistent with a larger replisome structure in which the GAN-GINS complex also interacts with PolD.

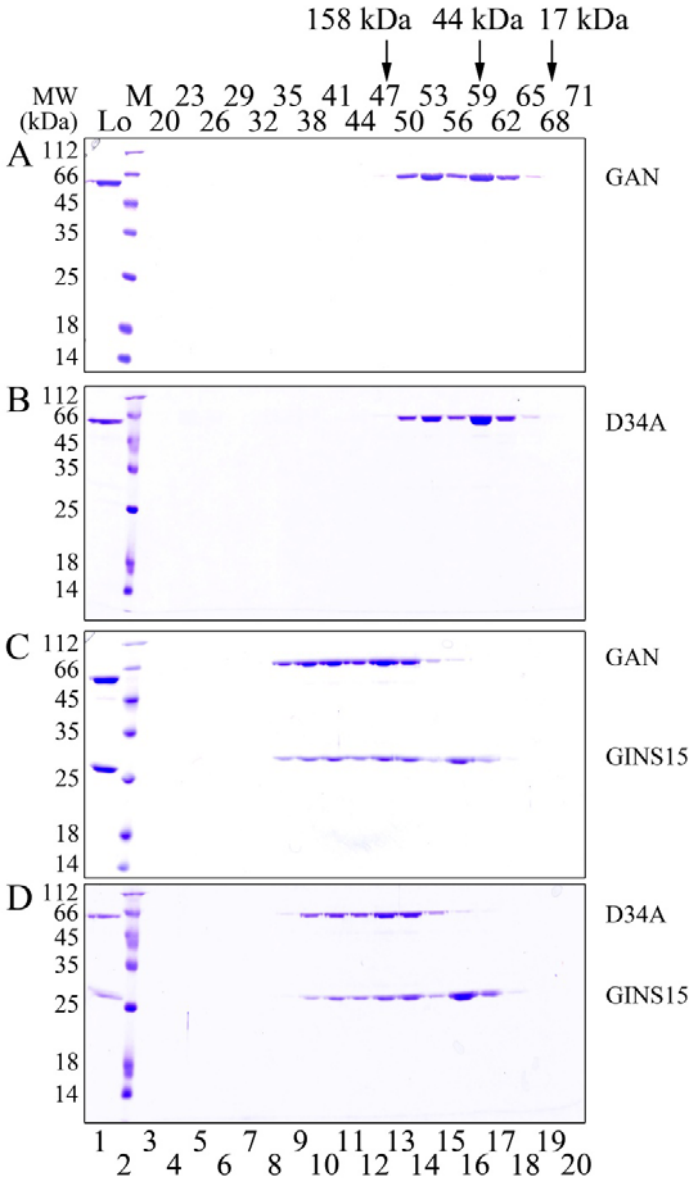

Figure 3-11: GAN (D34A) interacts with GINS15.

A sample (100 μg) of each protein listed to the right of the corresponding panels (A through D) was subjected to Superdex-200 gel filtration analysis. The proteins in aliquots (80 μl) from each fraction were separated by electrophoresis through 12% polyacrylamide-SDS gels and stained with Coomassie brilliant blue (R-250). Fractions in which γ-globulin (158 kDa), ovalbumin (44 kDa) and myoglobin (17 kDa) eluted are noted at the top of the figure. Panel A and C are the same gel shown in Figure 3-2A and 3-2E respectively.

Discussion

Genes encoding well-conserved GAN homologues are widely distributed in the Euryarchaea but are not present in Crenarchaea, Bacteria or in the Nanoarchaeum equitans genome (Figure 3-12). The most parsimonious interpretation is therefore that the GAN nuclease evolved in the archaeal domain, and specifically in the euryarchaeal lineage after separation from the crenarchaeal lineage. It seems likely that GAN plays a conserved role in euryarchaeal DNA metabolism, and given its robust 5' → 3' exonuclease activity, stimulated by association with GINS15, GAN could play a role directly in euryarchaeal genome replication. In Bacteria, the 5' exonuclease activity of DNA polymerase I (PolI) is required to remove the RNA strand. The strand displacement activity of eukaryal DNA polymerase δ (Polδ) can generate flap structures that contain the RNA primer (Figure 3-13B) and are recognized and cleaved by Fen1 (59,102). Dna2 helicase/nuclease may also participate in the resection of the flap structures (103). However, as Fen1 is not essential for viability (104,105), alternative mechanism(s) must exist to mature the eukaryal Okazaki fragments. In the Euryarchaea, which includes *T. kodakarensis*, PolD may synthesize the lagging strand (106). This polymerase does have strand displacement activity (106,107) which, on encountering an Okazaki fragment, could generate the 5' ssDNA required for archaeal Fen1 digestion (Figure 3-13C_I). Since Fen1 is also not essential for Euryrachaea viability (32) it is possible that the 5' ssDNA generated by the PolD displacement reaction can be removed by GAN (Figure 3-13C_II). While bacterial and eukaryal Okazaki fragments are initiated by short ribonucleotides, lagging strand synthesis in Archaea may involve only

Figure 3-12: Alignment of GAN in euryarchaeota.

An alignment of the amino acid sequences of the GAN protein from six euryarchaeota species. Highlighted colors represent residues with 100% identity in red, 80% identity in yellow and 60% identity in green. The accession numbers used are: *Methanosarcina acetivorans*, NP_615782.1; *Methanococcoides burtonii*,

YP_566162.1; *Methanocaldococcus jannaschii*, NP_247822.1; *T. kodakarensis*, YP_183665.1; *Pyrococcus abyssi*, YP_183665.1; *Methanothermobacter thermautotrophicus*, NP_276538.1.

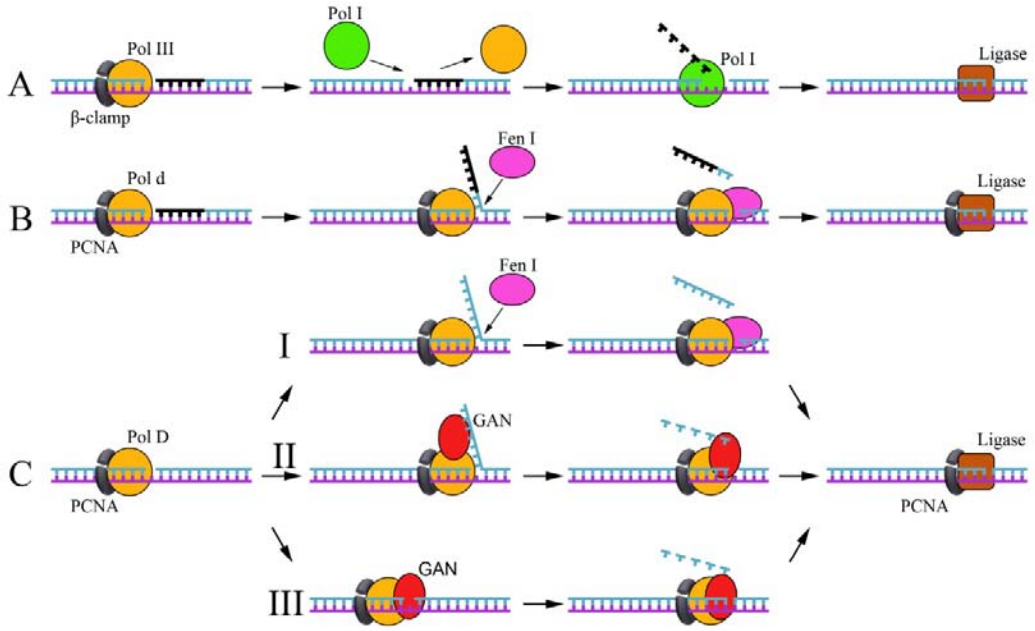

Figure 3-13: Proposed role for GAN during archaeal DNA replication.

Models for lagging strand DNA maturation in the (A) Bacteria, (B) Eukarya and (C) Archaea. See text for details.

deoxynucleotide primers (47). If this was the case, their degradation could becatalyzed by the 5' → 3' exonuclease activity of GAN. Here, we have established that a 5' ssDNA extension is required to initiate GAN activity *in vitro* but, *in vivo*, as a component of a replisome complex, GAN might have the ability to initiate DNA degradation from an unsealed nick in dsDNA. If this were the case, then GAN could provide a function equivalent to the 5' → 3' exonuclease activity of bacterial PolI (Figure 3-13C_III).

Chapter 4: Regulating PCNA activity by the binding of small protein.

Introduction

PCNA plays essential roles in DNA metabolic processes including replication, repair, recombination and cell cycle progression [for review see (21,108,109)]. PCNA is a ring-shaped trimeric protein that encircles duplex DNA (52,53) and tethers other enzymes to the DNA (21,110,111). To date, all activities described for the PCNA proteins require them to encircle the duplex; no biochemical function for PCNA off DNA has been reported. However, the PCNA proteins form stable rings that cannot assemble independently around the duplex. The RFC complex functions as the clamp loader and assembles the PCNA rings around the duplex [reviewed in (112)].

Most of the proteins that interact with PCNA do so via a PIP (PCNA-interacting peptide) motif (73,74). The PIP motif is a fairly weak consensus sequence of QXXhXXaa where "h" is a moderately hydrophobic amino acid (isoleucine, leucine or methionine) and "a" is an aromatic residue, followed by a non-conserved sequence containing basic amino acids. The PIP motif interacts with the loop that connects the two domains in each PCNA monomer [referred to as interdomain connecting loop (IDCL)] (113). Biochemical and structural analysis illustrated the requirement for an intact PIP motif for the interactions between replication enzymes and PCNA. Other, less common, PCNA-binding motifs have also been reported in several PCNA interacting enzymes [for example see (114)].

The β-subunit is the functional homologue of PCNA in bacteria forming a ring-

shaped structure similar to PCNA (20). Similar to PCNA, the β-subunit also interacts with a number of enzymes involved in nucleic acid metabolic processes many of which bind to the β-subunit via a consensus sequences, QL[S/D]LF and QLxLx[L/F], that resembles the PIP motif (115,116).

Genes encoding PCNA are present in all archaeal genomes. While the genomes of the euryarchaeota branch of archaea contain a single gene encoding for PCNA, the genome of *T. kodakarensis* contains two genes encoding for PCNA homologues (TK0535 and TK0582 encoding for PCNA1 and PCNA2, respectively) (56,69). It was found that both proteins form trimeric structures with characteristics similar to those of other archaeal and eukaryal PCNA proteins (51). Both proteins were also shown to support processive DNA synthesis by PolB and PolD (55,56,69).

Most, but not all, of the archaeal replication proteins were identified by their homology to the eukaryotic or bacterial counterparts using *in silico* approaches. Bioinformatic tools, however, cannot identify archaeal-specific replication proteins or proteins with similar function to the eukaryotic/bacterial counterparts but with diverse sequences. In the last few years, however, genetic tools have been used to expand the pool of putative archaeal replication factors. In one of these studies, several established replication proteins from the archaeon *T. kodakarensis* were tagged *in vivo* with an amino- or carboxy-terminal hexahistidine extension (His_6-tag) (29). These proteins were purified using nickel-affinity column from *T. kodakarensis* cell lysates, and the proteins that co-isolated were identified (29).

One of the proteins identified in the study was a small protein (~8.5 kDa, encoded by TK0808) that co-purified with PCNA1. Small PCNA binding proteins,

like P21, are important PCNA regulator in eukaryote. However, the PCNA regulator in archaea is still not known yet. So we hypothesize that this small protein may be a possible PCNA regulator in archaea. For this reason, we performed in detail study of TK0808. Bioinformatic analysis revealed that homologues proteins are present only in *Thermococcales* genomes. *In vitro* studies showed that the protein could bind both PCNA1 and PCNA2 resulting in the inhibition of PCNA dependent activities of DNA polymerase and Fen1 activities. Therefore the protein was designated TIP (*Thermococcales* inhibitor of PCNA). However, TIP does not contain a PIP motif or a sequence similar to those involved in proteins interactions with the bacterial β-subunit. In addition, using H/D exchange and site directed mutagenesis it was found that TIP does not interact with the IDCL on PCNA. These results suggest a new mechanism to bind and inhibit PCNA activity and may provide a new target for drug discovery.

Materials and Methods

Cloning and purification of recombinant proteins.

For protein expression in *E. coli* the genes encoding TIP (TK0808) and Fen1 (TK1281) were PCR-amplified from *T. kodakarensis* genomic DNA using the primers shown in Supplementary Table S1 which include an NdeI and SalI restriction enzymes in the forward and reverse primers, respectively. The amplified DNA was ligated with pET15b which include an in frame His_6-tag at the N-terminus (117). The construction of the expression vectors for *T. kodakarensis* PCNA1, PCNA2, RFC, PolB and *M. thermautotrophicus* PCNA were previously described (48,55,56,118).

The vectors to express the mutant forms of PCNA2 were generated using site-specific mutagenesis using a QuikChange kit (Agilent Technologies) and the plasmids that contain the wild-type proteins as templates. The oligonucleotides used are listed in Appendix 6.

The pET based vectors encoding for the different proteins were transformed into BL21 DE3 Rosetta cells (Life Technologies) and protein expressions were induced at 37°C by the addition of 0.5 mM IPTG and farther incubation for 3 h. The proteins were purified by absorption and elution from a Ni column as previously described for the PCNA proteins (55). Following purification the proteins were aliquoted and stored at -80°C.

Elongation assay of singly primed M13.

PolB catalyzed elongation of singly primed M13 ssDNA was carried out in reaction mixtures (20 μL) containing 40 mM Tris-HCl (pH 8.0), 250 mM NaCl, 1.5 mM DTT, 100 μg/ml BSA, 10 mM magnesium acetate, 2 mM ATP, 100 μM each of dCTP, dGTP, and dTTP, 20 μM [α-^{32}P]dATP (1.2×10^4 cpm/pmol), 10 fmol of singly primed M13mp18, 440 fmol of RFC and 100 fmol of PolB and PCNA1, PCNA2 and TIP proteins as indicated in the figure legend. Reaction mixtures were incubated for 20 min at 70°C. Following incubation, reactions were treated with 2 μl of stop solution (containing 0.1M EDTA, 5% SDS, 80 μg yeast tRNA, 20 μg proteinase K) and incubated for 20 min at 37°C. For quantitation, aliquots (4 μL) of reaction mixtures were removed, and DNA synthesis measured by adsorption to DE81 paper followed by liquid scintillation counting. The mixture was adjusted to 3 M ammonium acetate and 70% ethanol and the DNA collected by centrifugation.

Following *in vacuo* drying the pellets were dissolved in 0.05 M NaOH and 1 mM EDTA and separated by alkaline agarose electrophoresis.

Fen1 nuclease assay.

Three oligonucleotides were used to generate the substrate for Fen1 assay. H7 (5'-CTTCAATCGGCTCAGACCGAGCAGAATTCTATGTGTTTACCAAGCGCT G-3') was labeled at the 5'-end with ^{32}P using [γ-^{32}P]ATP and polynucleotide kinase. Following the labeling the DNA was hybridized to two additional oligonucleotides, H2 (5'-CAGCGCTTGGTAAACACATAGAATTCTGCTCGGTCTCTCGGCAGAT TCTAGAAATCGACGCTAGCAAGTGAC-3') and H5 (5'-GTCACTTGCTAGCGT CGATTTCTAGAATCTGCCGAG-3'). The substrate was purified from polyacrylamide gel as previously described (99).

Unless otherwise noted, Fen1 nuclease assays were performed in 20 μl reaction containing 20 fmol substrates, 25 mM Tris-HCl (pH 8.0), 10 mM $MgCl_2$, 125 μg/ml BSA, and proteins as indicated in the figure legends. Reactions were incubated at 60°C for 1 hr. Following incubation the reactions were stopped by adding 20 μl of 95% formamide, 10 mM EDTA and 0.1x TBE followed by incubation at 100°C for 2 min. The reaction products were separated on 20% polyacrylamide-8M urea gels in 1x TBE followed by visualization and quantification using phosphorimaging (GE Healthcare).

H/D exchange (HDX) mass spectrometry.

For H/D exchange (HDX) mass spectrometry analysis the protein stock solution was diluted in PBS buffer (20 mmol/l sodium phosphate, 500 mM NaCl, pH 7.6) to

prepare a 15 μmol/L final analytical concentration and equilibrated at 4°C for 2 hr. PCNA-TIP complex was prepared by mixing PCNA and TIP at a ratio of 1:1 (PCNA monomer : TIP monomer) and equilibrated at 4 °C for 16 hr.

HDX was conducted on an HDX PAL robot (LEAP Technologies, Carrboro, NC). Protein solution (5 μl) was diluted into 21 μl D_2O buffer (20 mmol/l sodium phosphate, 500 mmol/l NaCl, pD 7.6) at 3 °C. At selected times (30 sec, 5 min, 20 min, 40 min, and 60 min) the HDX was quenched by mixing with 40 μl of 3 mol/L urea, 1% TFA at 1°C. The quenched solution was injected into an on-line pepsin-digestion device for 3 min. The digested protein solution was trapped on a C18 guard column (1.0 mm dia., 5 μm, Grace Discovery Sciences). The peptide mixture was separated with a C18 analytical column (1.0 mm dia. x 5 cm length, 1.9 μm, Hypersil GOLD, Thermo Scientific) via a Dionex Ultimate 3000 UPLC with a 9.5 min gradient operated with a binary mixture of solvent A and B at 50 μl/min flow rate. The gradient settings were: 5% to 35% solvent B in 3 min, 35% to 70% solvent B in 5 min, 70% to 100% solvent B in 0.5 min, and isocratic flow at 100% solvent B for 0.5 min, then returned to 5% solvent B in 0.5 min. Solvent A was water containing 0.1% formic acid, and solvent B was 80% acetonitrile and 20% water containing 0.1% formic acid. All LC connection lines and valves were housed in the refrigerated compartment of the HDX PAL at 2°C. Peptides were analyzed on a Thermo LTQ Orbitrap Elite (Thermo Fisher, San Jose, CA). The instrument settings were: spray voltage, 3.7 kV; sheath gas flow rate, 25 (arbitrary units); capillary temperature, 270°C. Three replicates were obtained for each ion-exchange time point.

Peptide identification and HDX data processing.

Peptides of PCNA (PCNA1, PCNA2, PCNA2-A mutant, PCNA2-E mutant) and TIP were identified using tandem MS (MS/MS) on the Thermo LTQ Orbitrap Elite. One full mass spectral acquisition triggered six scans of MS/MS (precursor ion is activated by CID) whereby the most abundant precursor ions were sequenced. Peptide identification was achieved by submitting Thermo RAW files to both MASCOT (Matrix Science, Oxford, U.K.) and MassMatrix database search engines. The MASCOT settings were: enzyme, none; MS tolerance, 20 ppm; MS/MS tolerance, 0.6 Da; maximum number of missed cleavages, 3; peptide charge of 1+, 2+ and 3+. The MassMatrix settings were: enzyme, nonspecific; precursor ion tolerance, 10 ppm; product ion tolerance, 0.8 Da; minimum pp score, 5.0. All identifications of peptides were manually confirmed. From mass spectra obtained during HDX-MS experiments the centroid of each deuterated peptide envelope and the relative deuterium uptake by each peptide were calculated with HDX Workbench (119). Corrections for back exchange were made by considering the values of 80% deuterium content of the exchange buffer and an estimated 70% deuterium recovery. Paired t-tests were used in verifying the deuterium uptake differences, and a value of $p < 0.05$ is considered significant.

Size exclusion chromatography.

Aliquots of each experimental protein (100 μg) or protein mixture and Gel Filtration standards (Bio-Rad) were diluted in 200 μl of 25 mM Tris-HCl (pH 8.0), 200 mM NaCl, 0.5 mM EDTA and 10% (v/v) glycerol and loaded onto a Superdex-200 column (HR10/30; GE Healthcare) pre-equilibrated in the same buffer. Fractions

(250 µl) were collected from the column at a flow rate of 0.5 ml/min. The proteins present in aliquots (80 µl) of each fraction were separated by electrophoresis through a 15% (w/v) polyacrylamide-SDS gel and stained with Coomassie brilliant blue (R250).

T. kodakarensis strain construction and confirmation of genome structures.

T. kodakarensis strains were grown in artificial sea water (ASW) with 5 g/L each of yeast extract (Y) and tryptone (T), and 2 g/L sulfur (S°) at 85°C with the growth of cultures measured by an increase in optical density at 600 nm (OD_{600}) as previously described (120). Plasmids pOSU0808A, pOSU0808B, and pOSU0808D were created using standard molecular biology techniques as established and were maintained in *E. coli* (120). Transformation of *T. kodakarensis* cells were performed as previously described (121) and strain construction for markerless deletion of the TK0808 gene was performed as described in (120). Strains were constructed by transformation of *T. kodakarensis* TS559 (120) with transformants grow on media lacking agmatine for plasmid incorporation and then counter-selected on media containing 6-methylpurine to generate a strain containing the deletion of TK0808 with no selectable marker (120). Strain THH2 (ΔTK0808) was generated by deleting the first 189 bp of the TK0808 ORF; the final 6 bp overlap with TK0807 and were thus retained.

pTHH6 construction and confirmation.

Sequences encoding TIP were amplified via PCR with a 5' primer that contained the promoter for histone B from *M. thermautotrophicus* (PhmtB). The resultant

amplicon was cloned into pTS543 at a unique NotI restriction site using Clonetech In-Fusion® HD cloning.

Southern blot analysis.

The genomic organization of strain THH2 was confirmed via Southern blots of Acc65I and BglII digested genomic DNA. The Acc65I and BglII restriction fragments that hybridized to a digoxigenin (DIG)-labeled amplicon probe, PCR-generated from within TK0808 (probe B/C) and a flanking region (probe E/F) (Figure 4-1), were identified by using anti-DIG antibodies coupled to alkaline phosphatase as previously described (122).

T. kodakarensis growth curves.

T. kodakarensis strains were grown overnight in 5 ml ASW YT S° medium, and strains lacking pTHH6 were supplemented with 1 mM agmatine sulfate. Growth at 85°C of triplicate cultures was monitored at OD_{600} nm of 1:100 inoculated cultures.

Results

TIP interacts with PCNA1 and PCNA2 *in vitro*.

TIP protein, encoded by TK0808, was co-isolated with His_6-PCNA1 (TK0535) from *T. kodakarensis* cell lysates (20). To determine if the two proteins also interacted *in vitro*, recombinant TIP and PCNA1 were mixed and the products examined by size exclusion chromatography. As previously been reported (69,93) PCNA1 protein (29.1 kDa) elute as a trimeric complex (Figure 4-2A; elution peaks in fraction 55). TIP (9.8 kDa) alone eluted as a monomer (Figure 4-2C, elution peak in

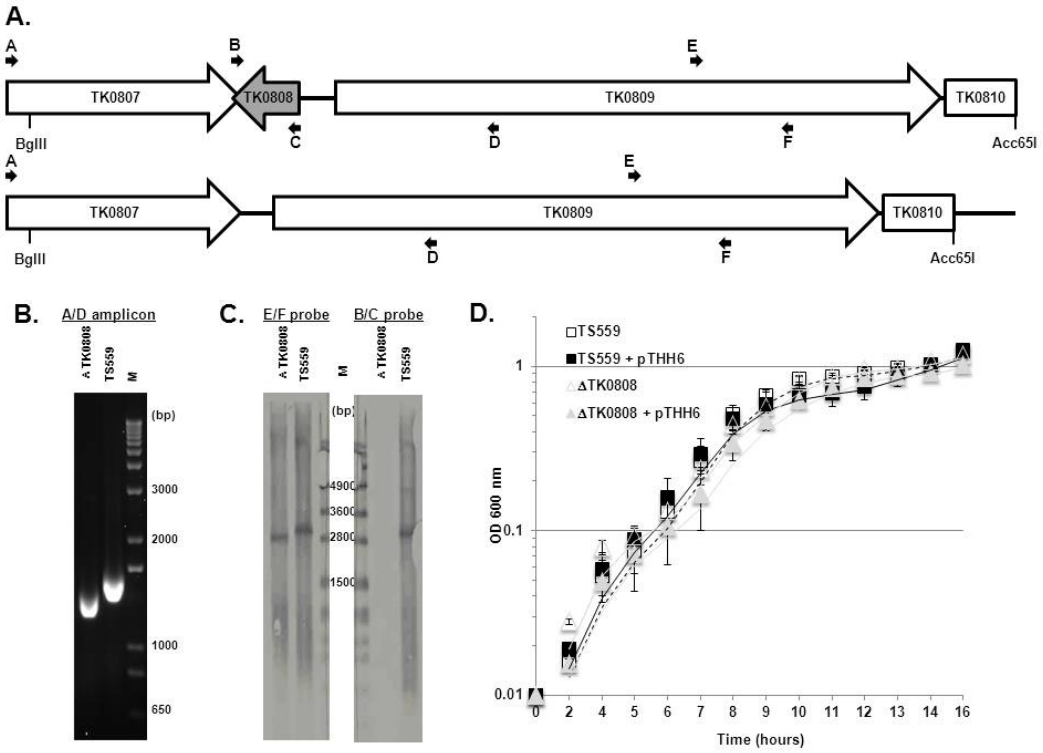

Figure 4-1: TIP is not essential for *T. kodakarensis* viability.

(A). Annotated sections of the *T. kodakarensis* strains TS559 (top) and ΔTK0808 (bottom) genomes, highlighting the locations of genes (open arrows) and oligonucleotide primers (black arrows; labeled A-F). The approximate locations of relevant Acc65I and BglII restriction sites are shown. (B). Ethidium bromide-stained, agarose gel electrophorectic separation of amplicons generated using primers A and D (panel A) from chromosomal DNA templates from strains TS559 and ΔTK0808 results in a ~190 bp smaller amplicon from *T. kodakarensis* strains lacking TK0808 than the parental strain, TS559. (C). Southern analysis confirms the genome structures of *T. kodakarensis* strains TS559 and ΔTK0808. Probes were generated with primers B/C (left) and E/F (right). M = DNA markers, in bp. (D). Growth curves of *T. kodakarensis* strains TS559 and ΔTK0808, alone and when containing plasmid pTHH6. The moving average (period = 2) is shown as a trend line for each curve. (Experiment performed by Thomas Santangelo)

fraction 67). When incubated together, TIP and PCNA1 interacted to form a complex as evident by the elution of TIP at earlier fractions (Figure 4-2D, elution peak in fraction 58) and the elution of PCNA1 in later fractions (Figure 4-2D, elution peak in

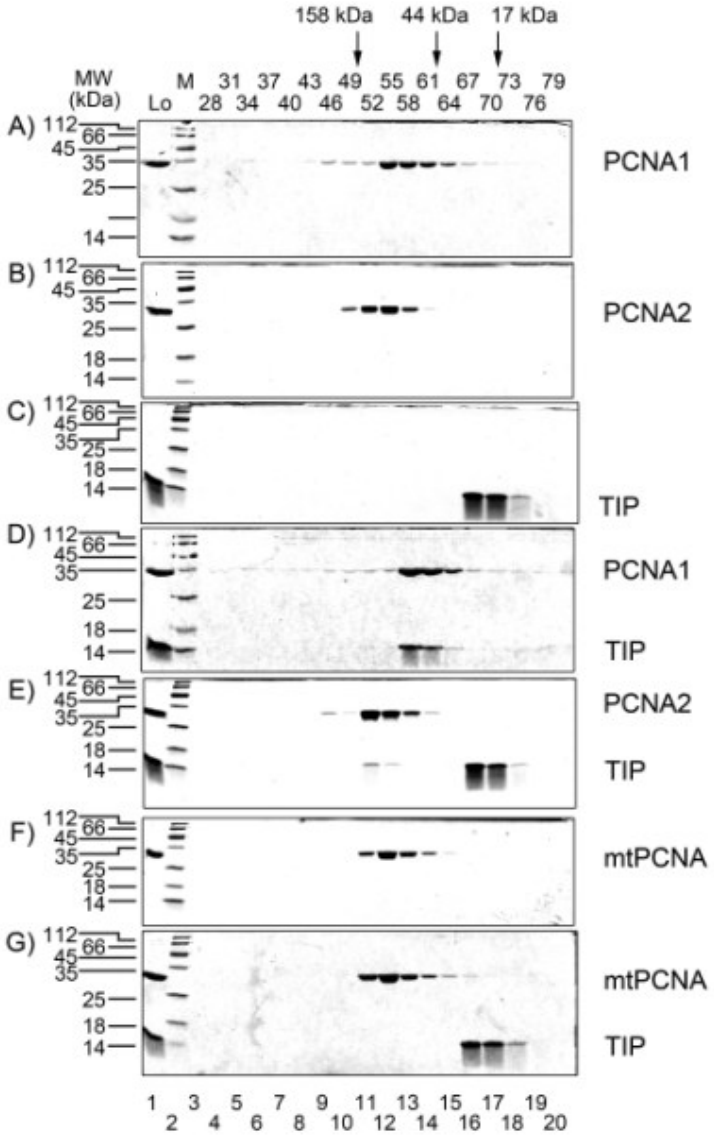

Figure 4-2: TIP interacts with PCNA1 and PCNA2.

One hundred micrograms of each protein were separated on Superdex-200 gel filtration column as described in "Material and Methods". Aliquots (80 µl) from the indicated fractions were separated on 15% SDS-PAGE and followed by Coomassie brilliant blue (R-250) staining. (A). PCNA1; (B). PCNA2; (C). TIP; (D), PCNA1 and TIP; (E), PCNA2 and TIP; (F), mtPCNA; (G), mtPCNA and TIP.

fraction 58). These results not only illustrate that the two proteins interacts but also that the binding of TIP to PCNA results in the dissociation of the trimeric ring.

As discussed above, *T. kodakarensis* genome encodes for two PCNA proteins (69,93) with similar overall structures (55). It was reported that PCNA2 present in low levels *in vivo* [<50 molecules per cell (69,93)] which may explain the inability to detect PCNA2-TIP interactions *in vivo*. Nevertheless, the ability of TIP to interact with PCNA2 was also been evaluated. As shown in Figure 4-2B, PCNA2 protein (29.3 kDa) elute as a trimeric complex (Figure 4-2B; elution peaks in fraction 55) as was previously reported (69,93). Similar to PCNA1, TIP also form a complex with PCNA2 as evident but the elution of TIP at earlier fractions than TIP alone (Figure 4-2E, elution peak in fraction 52).

Homologues of TIP have been identified only in the genomes of *Thermococcales*. This may suggest that only PCNA proteins from these organisms can bind TIP. Therefore, as a control, a PCNA homologue from a different euryarchaeon, *M. thermautotrophicus*, was used. As previously reported (118), the *M. thermautotrophicus* PCNA protein (30.4 kDa) elutes from the sizing column as a trimeric complex (Figure 4-2F; elution peaks in fraction 55). The presence of TIP did not affect this elution profile (Figure 4-2G) and the elution profile of TIP was not altered in the presence of *M. thermautotrophicus* PCNA (compare Figure 4-2G to 4-2C).

TIP inhibits PCNA-dependent PolB activity.

PCNA bind and stimulated the activity of many proteins involved in nucleic acid metabolic processes. The best understood function for PCNA is its role as the processivity factor for DNA polymerases. While the processivity of the replicative polymerase is low, it increased dramatically in the presence of PCNA (112). As TIP

binding to PCNA appears to affect PCNA oligomeric structure it may regulate PCNA functions. Therefore, the effect of TIP on PCNA stimulation of PolB activity was evaluated using a singly primed M13 template.

As was previously reported (55,56,69) the elongation of the ssM13 template by PolB was detected only in the presence of RFC and either PCNA1 (Figure 4-3 compare lanes 7-9 to lane 1) or PCNA2 (Figure 4-3 compare lanes 3 and 4 to lane 1). The presence of TIP, however, inhibited the stimulatory effect of both PCNA proteins (Figure 4-3 compare lanes 5 and 6 to lane 4; lanes 10-13 to lane 8 and lanes 14 and 15 to lane 7). TIP alone, without PCNA did not affect PolB activity (Figure 4-3 lane 2). It was previously showed that higher concentrations of PCNA1 in comparison to PCNA2 are needed to achieve similar stimulation of PolB (51,56). This is likely due to the lower stability of the PCNA1 rings (55). Therefore the concentrations of PCNA1 used in the experiments shown in Figure 4-3 are higher than those used for PCNA2.

The inhibition of PCNA effect on PolB by TIP can be due to its binding to PCNA or it may directly interact with PolB. The ability of TIP to bind PolB and the effect of TIP on PolB-PCNA interactions was evaluated using size exclusion chromatography. TIP (9.8 kDa) alone eluted as a monomer (Figure 4-4 panel C, elution peak in fraction67). PolB (89.6 kDa) alone also elute as monomer (Figure 4-4 panel D, elution peak in fraction 55). The elution profile does not change when TIP and PolB proteins are incubated together (Figure 4-4 compare panel I to panels C and D). PCNA1 (29.1 kDa) and PCNA2 (29.3 kDa) elute as a trimeric complex (Figure 4-4 panels A and B; elution peaks in fraction 55). When PCNA1 and PCNA2 incubated

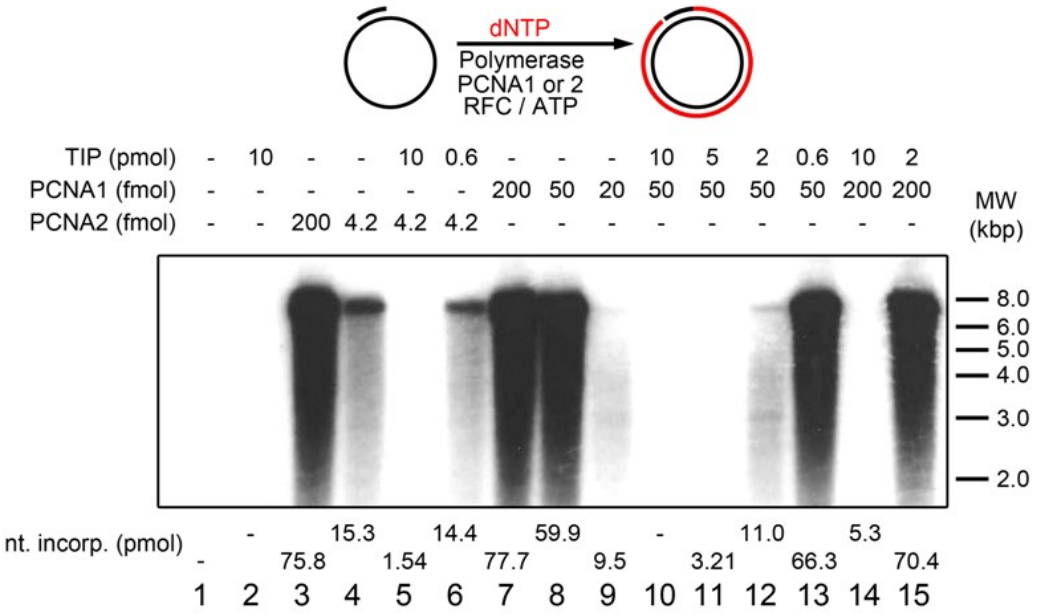

Figure 4-3: TIP inhibits PCNA stimulation of PolB.

Reaction mixtures (20 µl) were as described under "Materials and Methods" in the presence of 200 fmols (lanes 7, 14 and 15), 50 fmols (lanes 8 and 10-13), or 20 fmols (lane 9) of PCNA1 or 200 fmols (lane 3), or 4.2 fmols (lanes 4-6) of PCNA2 in the presence of 0.6 pmols (lanes 6 and 13), 2 pmols (lanes 12 and 15), 5 pmols (lane 11), or 10 pmols (lanes 2, 5, 10 and 14) of TIP. Reactions were incubated for 20 min at 70°C. An aliquot (4 µl) was used to measure DNA synthesis, and the remaining mixture was subjected to 1.1% alkaline-agarose gel electrophoresis. After drying, gels were autoradiographed for 15 min at -80°C and then developed. The assay is schematically shown at the top. (Experiment performed by Jerard Hurwitz)

with PolB they form a complex (Figure 4-4 panels E and G; elution peaks in fraction 49 and 49, respectively). When TIP is added to the complex, however, the interactions between PCNA1 and PCNA2 and PolB are severed (Figure 4-4 panels F and H). These results suggest that the inhibition of PCNA effect on PolB is by severing the interactions between PCNA and PolB likely due to the dissociation of the PCNA ring.

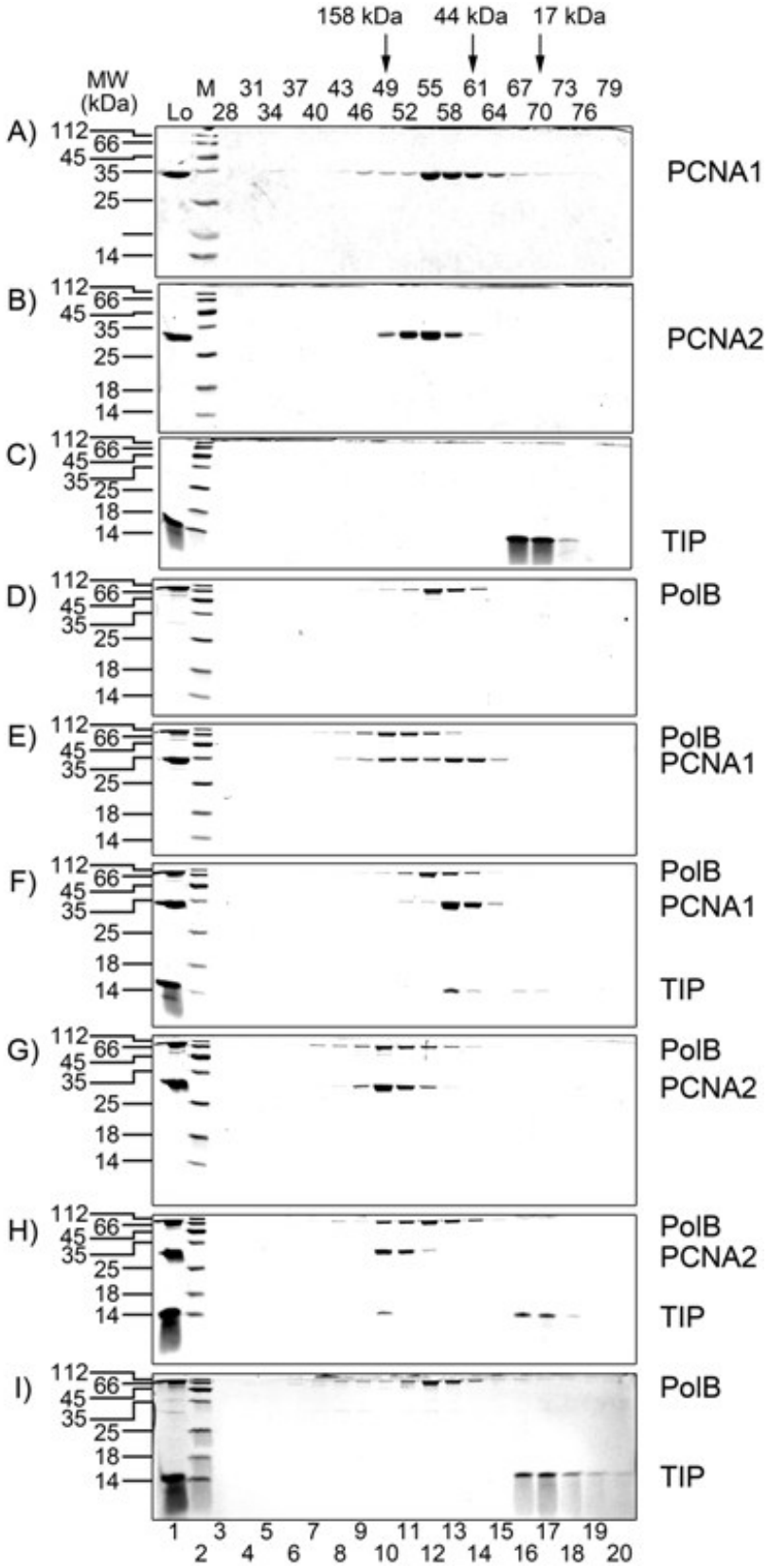
158 kDa
44 kDa
17 kDa
MW (kDa)
M 31 37 43 49 55 61 67 73 79
Lo 28 34 40 46 52 58 64 70 76
A)
B)
C)
D)
E)
F)
G)
H)
I)
112
66
45
35
25
18
14
PCNA1
PCNA2
TIP
PolB
PolB
PCNA1
PolB
PCNA1
TIP
PolB
PCNA2
PolB
PCNA2
TIP
PolB
TIP
1 3 5 7 9 11 13 15 17 19
2 4 6 8 10 12 14 16 18 20

Figure 4-4: TIP affects the interactions of PCNA1 and PCNA2 with PolB.

One hundred micrograms of each protein were separated on Superdex-200 gel filtration column as described in "Material and Methods". Aliquots (80 μl) from the indicated fractions were separated on 15% SDS-PAGE and followed by Coomassie brilliant blue (R-250) staining. (A). PCNA1; (B). PCNA2; (C). TIP; (D), PolB; (E), PCNA1 and PolB; (F), PCNA1, PolB and TIP; (G), PCNA2 and PolB; (H), PCNA2, PolB and TIP; (I), PolB and TIP.

TIP inhibits PCNA-dependent Fen1 activity.

In a number of organisms, it was found that PCNA also stimulate the activity of Fen1, a structure specific nuclease involved in Okazaki fragment maturation (123). The enzyme removes the flap structure that contains the RNA primer prior to gap filling by the polymerase and subsequent ligation of the two adjacent Okazaki fragments (124). Although PCNA effect on Fen1 activity was shown for other organisms it has not yet been reported for *T. kodakarensis*. Therefore, the effect of PCNA1 and PCNA2 on Fen1 activity was evaluated using a Fen1 assay (schematically shown in Figure 4-5A). As shown in Figure 4-5B and C while Fen1 alone has very weak nuclease activity (Figure 4-5B lanes 1 and 7) both PCNA1 (Figure 4-5B lanes 2-5) or PCNA2 (Figure 4-5B lanes 8-11) stimulate the activity of Fen1 in a concentration dependent manner. As expected, PCNA1 or PCNA2 do not possess nuclease activity on their own (Figure 4-5B lanes 6 and 12).

Next the effect of TIP on Fen1 stimulation by both PCNA proteins was evaluated. As shown in Figure 4-5D and E, Fen1 activity was inhibited in the presence of TIP in a concentration dependent manner (Figure 4-5D compare lanes 3-7 to lane 2 and lanes 10-14 to lane 9, also see panel E). Although both PCNA stimulate Fen1 activity to similar extent (Figure 4-5C) TIP is more efficient in inhibiting PCNA2 than

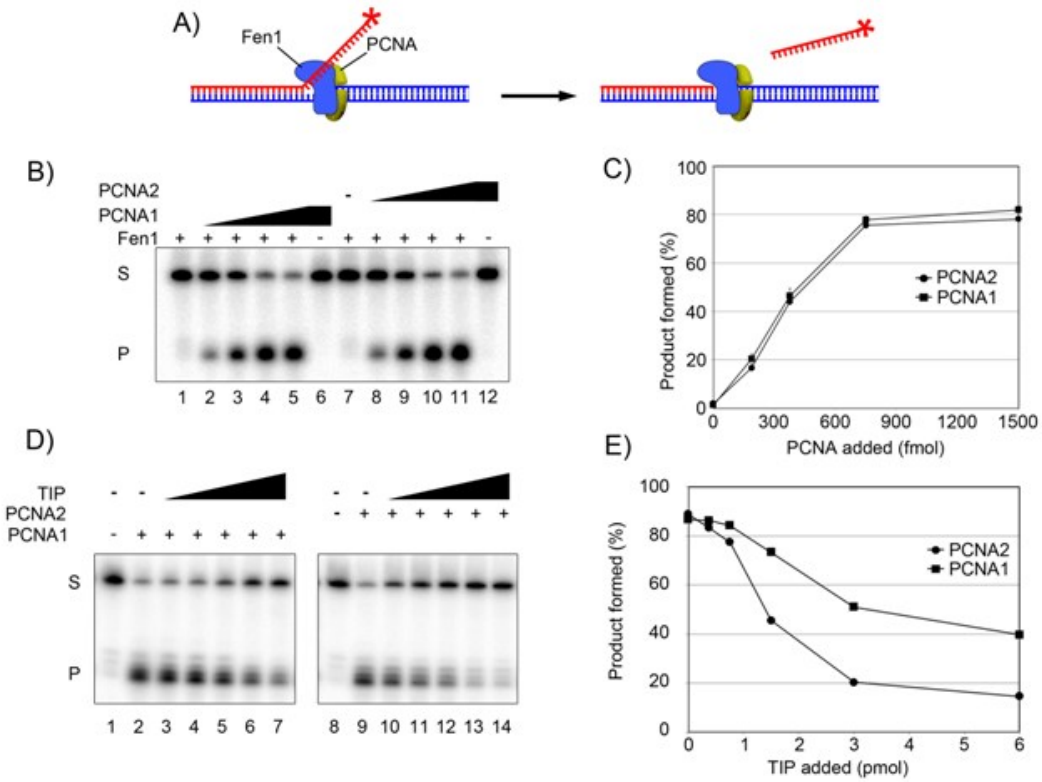

Figure 4-5: TIP inhibits PCNA stimulation of Fen1 activity.

(A). A schematic illustration of the assay. (B) and (C). The effect of PCNA on Fen1 activity was measured as described in "Material and Methods" in a reaction mixture (20 µl) that contained 20 fmols substrates, 250 fmols Fen1 (lanes 1-5 and 7-11), 188 (lane 2), 375 (lane 3), 750 (lane 4), 1500 (lanes 5 and 6) fmols of PCNA1, or 188 (lane 8), 375 (lane 9), 750 (lane 10), 1500 (lanes 11 and 12) fmols of PCNA2. The reactions mixtures were incubated at 60°C for 60 min and the products were separated on 20% (w/v) polyacrylamide-8M urea gels, visualized and quantified by phosphorimaging. A representative gel is shown in panel (B) and the averages, with standard deviations, from three independent experiments are shown in panel (C). (D) and (E). The effect of TIP on PCNA stimulation of Fen1 activity was measured as described in "Material and Methods" in a reaction mixture (20 µl) that contained 20 fmols substrates, 250 fmols Fen1, 1500 fmols PCNA1 (lanes 2-7), 1500 fmols PCNA2 (lanes 9-14) in presence of 375 (lanes 3 and 10), 750 (lanes 4 and 11), 1500 (lanes 5 and 12), 3000 (lanes 6 and 13), 6000 (lanes 7 and 14) fmols of TIP. The reactions mixtures were incubated at 60°C for 60 min and the products were separated on 20% (w/v) polyacrylamide-8M urea gels, visualized and quantified by phosphorimaging. Representative gel is shown in panel (D) and the averages, with standard deviations, from three independent experiments are shown in panel (E).

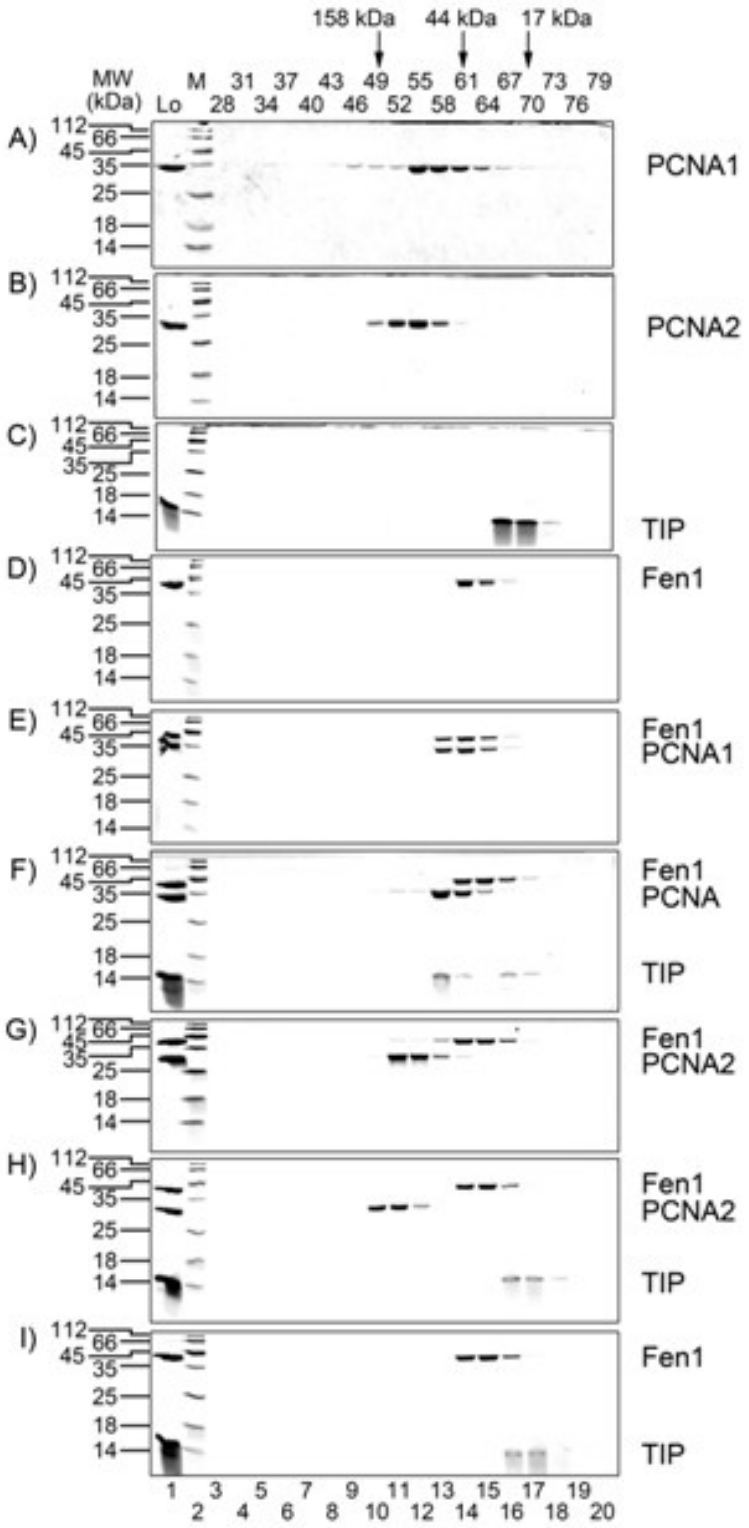
158 kDa
44 kDa
17 kDa
MW (kDa)
M 31 37 43 49 55 61 67 73 79
Lo 28 34 40 46 52 58 64 70 76
A)
B)
C)
D)
E)
F)
G)
H)
I)
112
66
45
35
25
18
14
PCNA1
PCNA2
TIP
Fen1
Fen1
PCNA1
Fen1
PCNA
TIP
Fen1
PCNA2
Fen1
PCNA2
TIP
Fen1
TIP
1 3 5 7 9 11 13 15 17 19
2 4 6 8 10 12 14 16 18 20

Figure 4-6: TIP affects the interactions of PCNA1 and PCNA2 with Fen1.

One hundred micrograms of each protein were separated on Superdex-200 gel filtration column as described in "Material and Methods". Aliquots (80 µl) from the indicated fractions were separated on 15% SDS-PAGE and followed by Coomassie brilliant blue (R-250) staining. (A). PCNA1; (B). PCNA2; (C). TIP; (D), Fen1; (E), PCNA1 and Fen1; (F), PCNA1, Fen1 and TIP; (G), PCNA2 and Fen1; (H), PCNA2, Fen1 and TIP; (I), Fen1 and TIP.

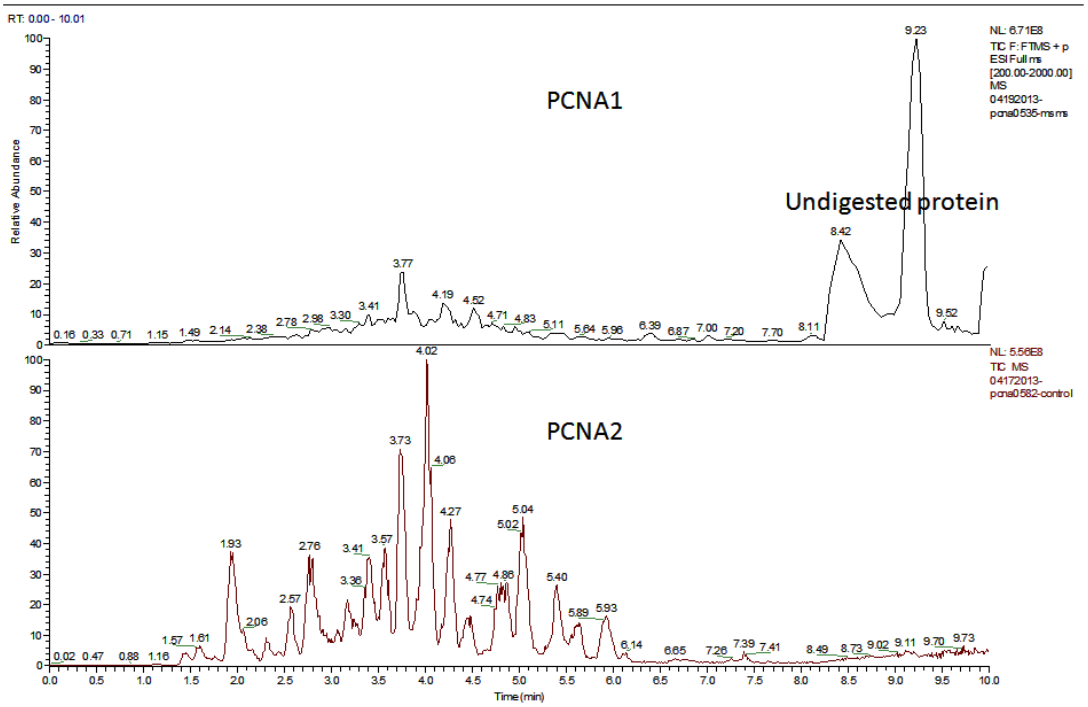

Figure 4-7: Total ion chromatogram of peptides obtained from PCNA.

PCNA1 (top) showed poor digestion efficiency and produced mainly the undigested protein. PCNA2 had good digestion efficiency. (Experiments were performed by Richard Huang)

PCNA1 (Figure 4-5E). Nevertheless, as with the effect of TIP on PolB activity, the binding of TIP to PCNA inhibits Fen1 activity. Similar to PolB, the interactions between Fen1 and PCNA were affected in present of TIP (Figure 4-6).

Protein dynamics of PCNA and TIP.

While the structure of PCNA is known that of TIP is not. One approach to

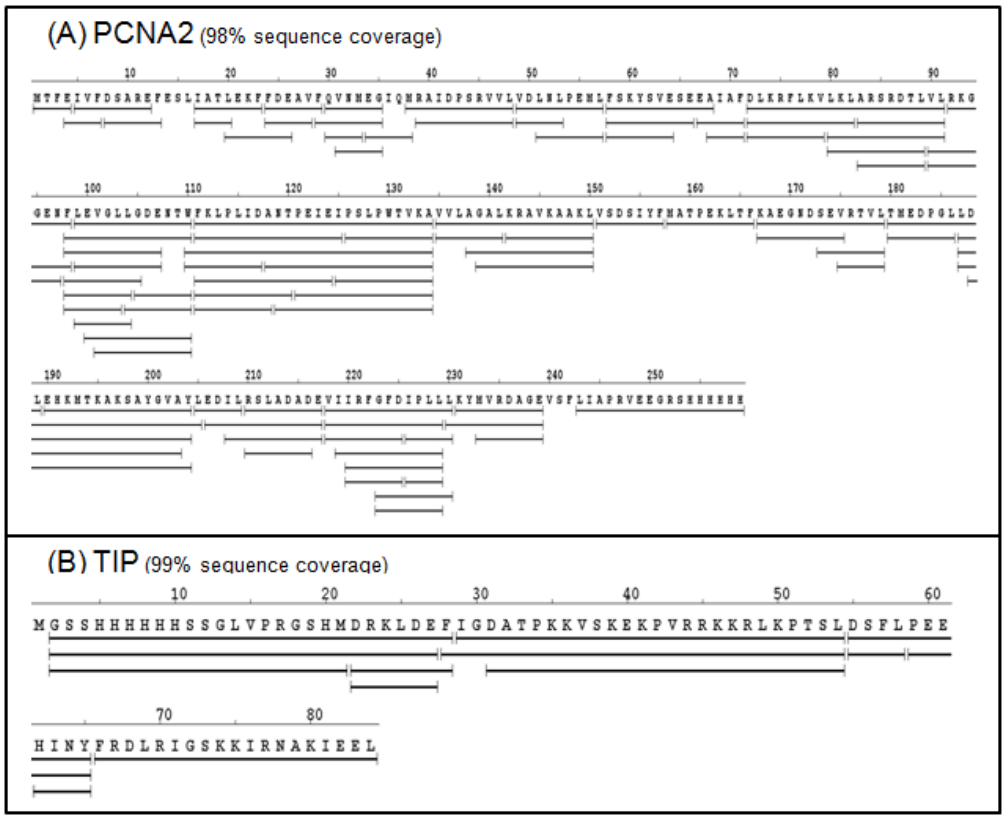

Figure 4-8: Peptic peptides of native PCNA2 and TIP.

An on-line digestion device provides 98% sequence coverage of PCNA2 (A) and 99% sequence coverage of TIP (B) (125). (Experiments were performed by Richard Huang)

understand the structural properties of these two partners is by studying their conformational dynamics in solution. Measurements of peptide-level HDX kinetics have the potential to reveal protein dynamics in solution with a spatial resolution of 6-10 residues (126). To attain such resolution, proteolysis of the subject proteins during the HDX-MS measurements must efficiently produce a distribution of peptides of overlapping sequence and leave negligible parent protein. Immobilized pepsin digests PCNA2 efficiently, but it failed to adequately digest PCNA1 for study (Figure 4-7); hence, our HDX-MS study focuses on the PCNA2 system. Proteolytic digestion of

native PCNA2 and TIP with pepsin yields HDX information for 98% of the PCNA2 backbone and 99% of the TIP backbone (Figure 4-8).

In solution, PCNA2 is known to form trimers (55,56,69). When mapped onto the crystal structure of PCNA2 (PDB: 3LX2) (55), HDX-MS kinetic data show that the interior-helical regions near the trimer interface are structurally stable (stronger hydrogen bonds) (Figure 4-9A), as they exchange an extremely low amount of deuterium over durations ranging from seconds to hours. On the other hand, the outer helices and loop regions, which are close to the trimer interface, are relatively more flexible, as their amide sites exchange to attain substantial fractions of deuterium on a time scale of minutes (Figure 4-9A). The exterior, long-loop regions show milder dynamics in which their amide sites exchange to acquire deuterium on the time scale of hours. These HDX-MS results are in accord with the trimeric quaternary structure of PCNA2 (Figure 4-9A).

The higher order structure of TIP is unknown. As shown by the HDX-MS exchange rate heat map of TIP (Figure 4-9B), amide groups of TIP, especially in the C-terminal region, rapidly exchange to contain over 80% deuterium within the first 30 sec after dilution in D_2O. Such fast conformational dynamics are in accord with the view that TIP is mostly unstructured.

Binding interfaces between PCNA2 and TIP.

HDX-MS has the potential to reveal binding interfaces in the protein complex byobserving changes in deuterium uptake rates among backbone amides in response to the formation of the protein complex (127-129). Pairwise comparisons of apo-PCNA2 (PCNA2 trimer alone) and holo-PCNA2 (PCNA2-TIP complex) and apo-TIP

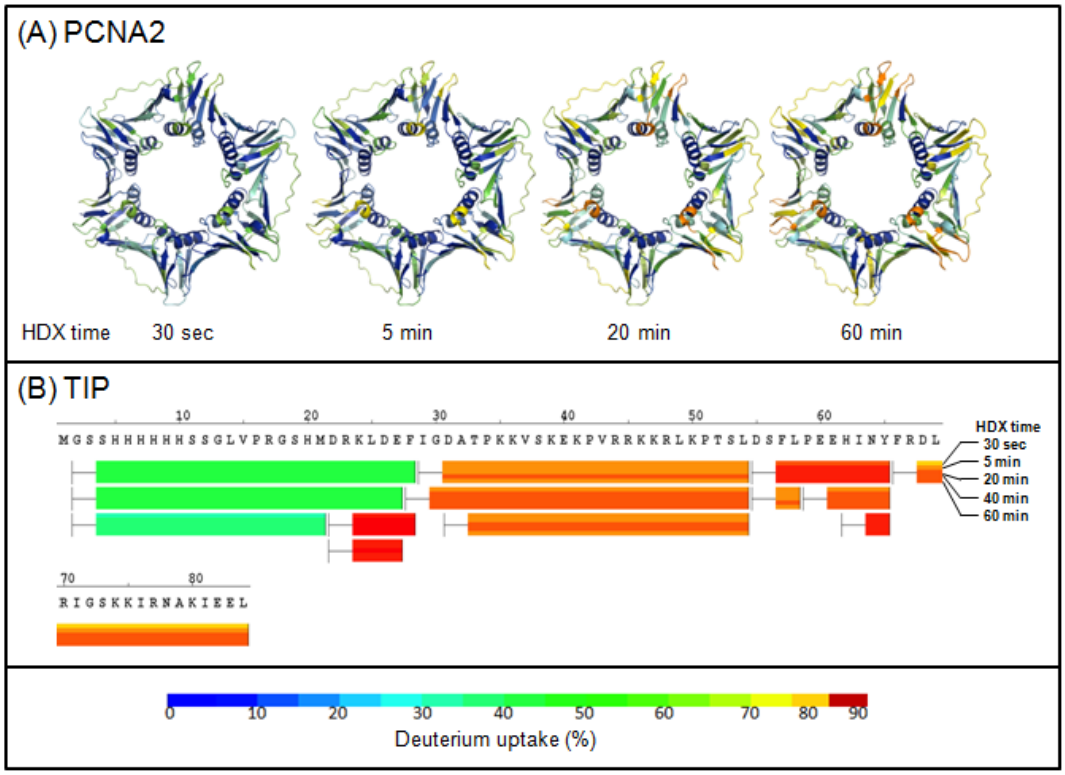

Figure 4-9: Protein dynamics of PCNA2 and TIP.

The HDX kinetics results of native PCNA2 are plotted onto the crystal structure (PDB: 3LX2) (55) (A). The HDX kinetics results of TIP are plotted according to the peptide sequence (125). The extent of relative deuterium uptake from low to high is presented with color ranging from blue to red. (Experiments were performed by Richard Huang)

and holo-TIP provide information regarding the interfaces located on PCNA2 and TIP that stabilize the holo-PCNA2 complex.

The HDX-MS results for the PCNA2-TIP complex (PCNA monomer: TIP monomer = 1:1) show that peptide regions 187-204 and 243-259 of native PCNA2 (Figures 4-10A and 4-11), located outside of the ring structure, exchange significantly less deuterium ($p \leqslant 0.03$, where any value of $p < 0.05$ is considered significant) upon

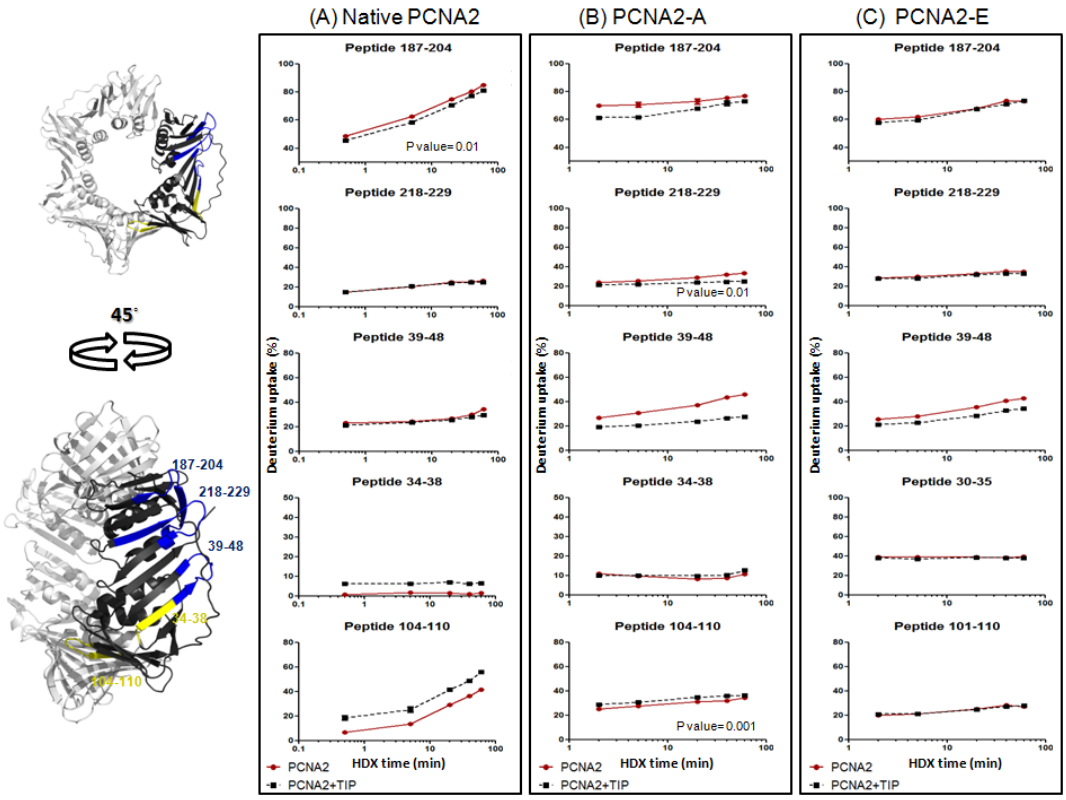

Figure 4-10: Mutation in PCNA2 affects its interactions with TIP.

HDX differences of apo-PCNA2 (red curves) vs. holo-PCNA2 (black curves). Regions of native PCNA2 that show significant differences in D uptake upon TIP binding are shown in (A) and plotted onto the crystal structure (PDB: 3LX2) (left). Regions show decrease in D uptake upon TIP binding are colored blue, whereas regions show increase in D uptake upon TIP binding are colored yellow. The HDX kinetics of the same regions of PCNA2-A mutant and PCNA2-E mutant are plotted in (B) and (C), respectively. (Experiments were performed by Richard Huang)

TIP binding. The diminished exchange rate exhibited by peptide region 187-204 in native PCNA2 suggests that these amides are involved in PCNA2-TIP binding. We regard the diminished D-uptake rates of peptide region 243-259 somewhat more cautiously because these comprise mostly artificial histidine residues; thus, the observed decrease in solvent accessibility at this region may be caused by non-specific interactions. Interestingly, peptide regions 34-38 and 104-110 of native

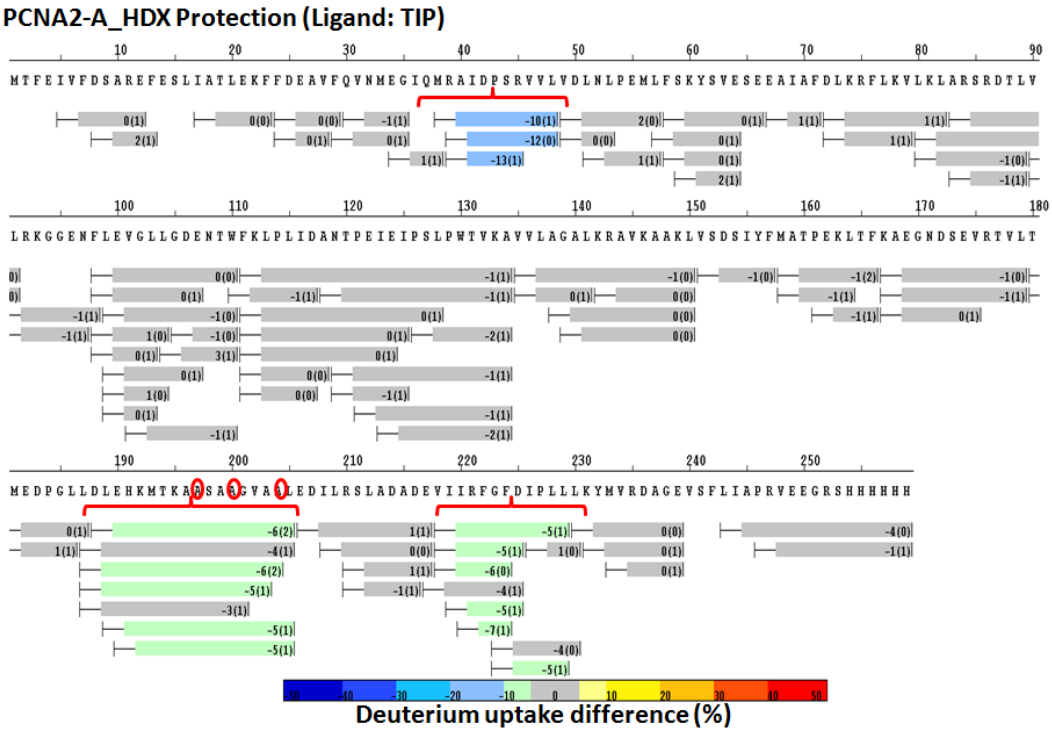

Figure 4-11: Differential HDX map of PCNA2-A mutant.

Differential HDX map of PCNA2-A mutant in which the regions having negative values of deuterium uptake difference (apo-PCNA2-A mutant vs. holo-PCNA2-A mutant) suggests the protein conformation changes toward a more protected conformation upon the binding of TIP. (Experiments were performed by Richard Huang)

PCNA2 exchange more deuterium upon TIP binding, indicating diminished protection factors. Since these amides are located at the trimer interfaces, the diminished protection factors suggest that binding with TIP facilitates the dissociation of PCNA2 trimer.

Mutational analysis of PCNA2-TIP interactions.

The region of 187-204 of PCNA2 was selected for site directed mutagenesis to

examine the role of this region in regulating the PCNA2-TIP interactions. Three residues in this region K197, Y200, and Y204 were mutated to Ala (Figure 4-12B, PCNA2-A) or Glu (Figure 4-12B, PCNA2-E). First, the effect of the mutation on the regulation by PCNA by TIP was evaluated using the Fen1 activity assay (Figure 4-12A). While the stimulatory effect of wild-type and PCNA2-A on Fen1 activity was inhibited by TIP to similar extent (Figure 4-12C and D), PCNA2-E inhibition was substantially reduced (Figure 4-12C and D). Although TIP affects each mutant differently, the mutant proteins on their own stimulate Fen1 activity to similar extent (Figure 4-12D).

The effect of the mutations on the TIP-PCNA interactions was then been evaluated using HDX-MS. PCNA2-A mutant showed significant decrease of deuterium uptake not only in region 187-204, but also in regions 39-48 and 218-229 (Figures 4-12B, 4-13). Interestingly, these regions are composed of short loops and are geometrically close to each other (blue regions in Figure 4-12), suggesting potential binding interfaces. The HDX-MS results for the PCNA2-A mutant suggest that these mutations strengthen the PCNA2-TIP interactions. This observation may support the stronger inhibitory effect of TIP on PCNA2-A stimulation of Fen1 at low TIP concentrations (Figure 4-12D). Surprisingly, the PCNA2-E mutant showed significant decrease of deuterium uptake only in region 39-48 (Figures 4-10C and 4-14). The absence of a deuterium uptake difference in region 187-204 indicates the importance of this region in modulating the PCNA2-TIP interactions.

Turning to an examination of the differential deuterium uptake of apo- and holo-TIP, we note that peptide regions 22-28 and 55-58 of holo-TIP (Figures 4-15A and 4-

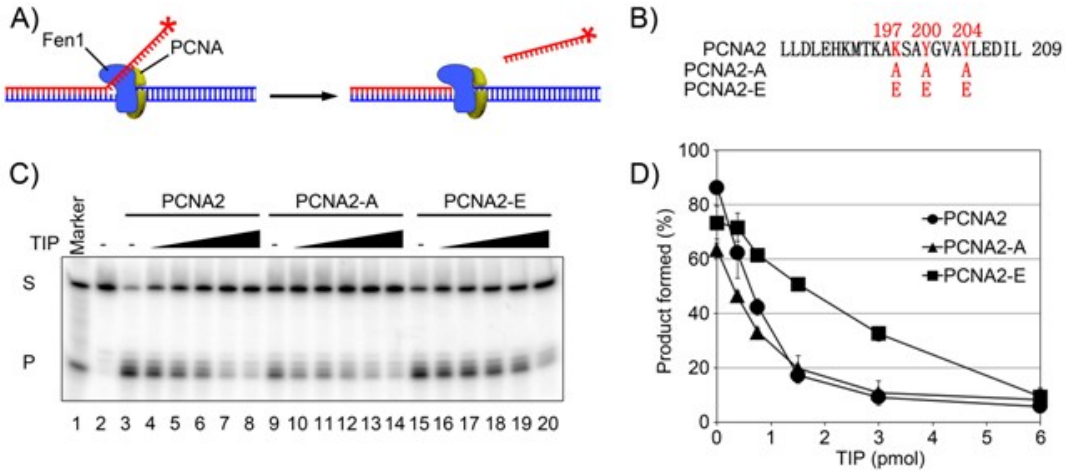

Figure 4-12: Effect of TIP on mutant PCNA stimulation of Fen 1 activity.

(A). A schematic illustration of the assay. (B). The residue mutated in PCNA2-A and PCNA2-E. C. The effect of TIP on PCNA wild-type and mutant proteins stimulation of Fen1 activity was measured as described in "Material and Methods" in a reaction mixture (20 μl) that contained 20 fmols substrates, 250 fmols Fen1, 1500 fmols PCNA2 (lanes 3-8), PCNA2-A (lanes 9-14) PCNA2-E (lanes15-20), in the presence of 375 (lanes 4, 10 and 16), 750 (lanes 5, 11 and 17), 1500 (lanes 6, 12 and 18), 3000 (lanes 7, 13 and 19), 6000 (lanes 8, 14 and 20) fmols of TIP. Oligonucleotides of 49 bases and 13 bases were separated in lane 1 and marked as "S" and "P", respectively. The reactions mixtures were incubated at 60°C for 60 min and the products were separated on 10% (w/v) polyacrylamide-8M urea gels, visualized and quantified by phosphorimaging. A representative gel is shown in panel (C) and the averages, with standard deviations, from three independent experiments are shown in panel (D).

16A) show slow deuterium uptake rates in the native PCNA2-TIP complex. These diminished D-uptake rates suggest that TIP engages with PCNA2 through interactions with peptide region 187-204 of native PCNA2. Moreover, the HDX kinetics data show that the deuterium uptake of peptide regions 22-28 and 55-58 of TIP in the presence of native PCNA2 eventually reaches the same level as that observed in the apo form (TIP alone), evidencing the flexible nature of TIP conformation. Not surprisingly, our results for the PCNA2-A mutant-TIP complex exhibit a similar trend of HDX kinetics in regions 22-28 and 55-58 of TIP (Figures 4-15B and 4-16B) as

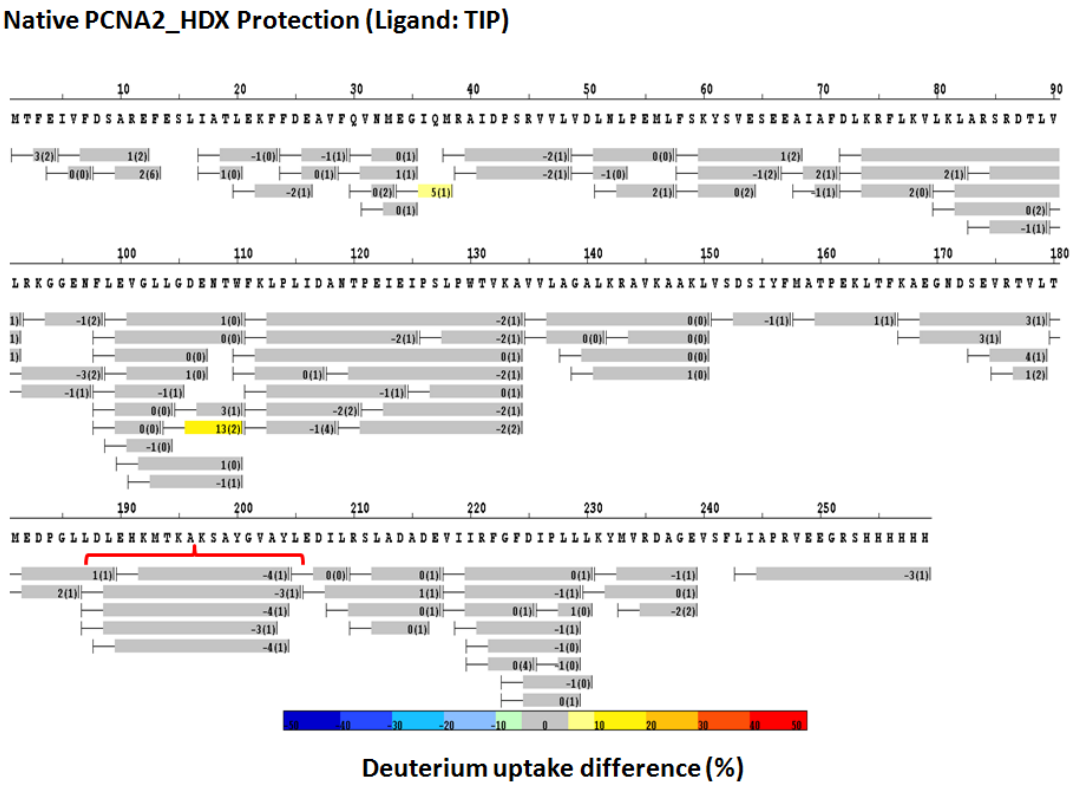

Figure 4-13: Differential HDX map of wild-type PCNA2.

Differential HDX map of wild-type PCNA2 in which the regions having negative values of deuterium uptake difference (apo-PCNA2 vs. holo-PCNA2) suggests the protein conformation changes toward a more protected conformation upon the binding of TIP. (Experiments were performed by Richard Huang)

compared to the native PCNA2-TIP complex (Figure 4-15A). This similarity is in accord with the formation of the stable PCNA2-A mutant-TIP complex. In contrast, apo-TIP and holo-TIP for the PCNA2-E mutant-TIP complex exhibit no significant difference in deuterium uptake behaviors (Figures 4-15C and 4-16C), indicating that the PCNA2-E mutant interaction with TIP is weak. In summary, our HDX analyses of PCNA2-TIP complexes show the importance of region 187-204 in regulating the PCNA2-TIP interactions.

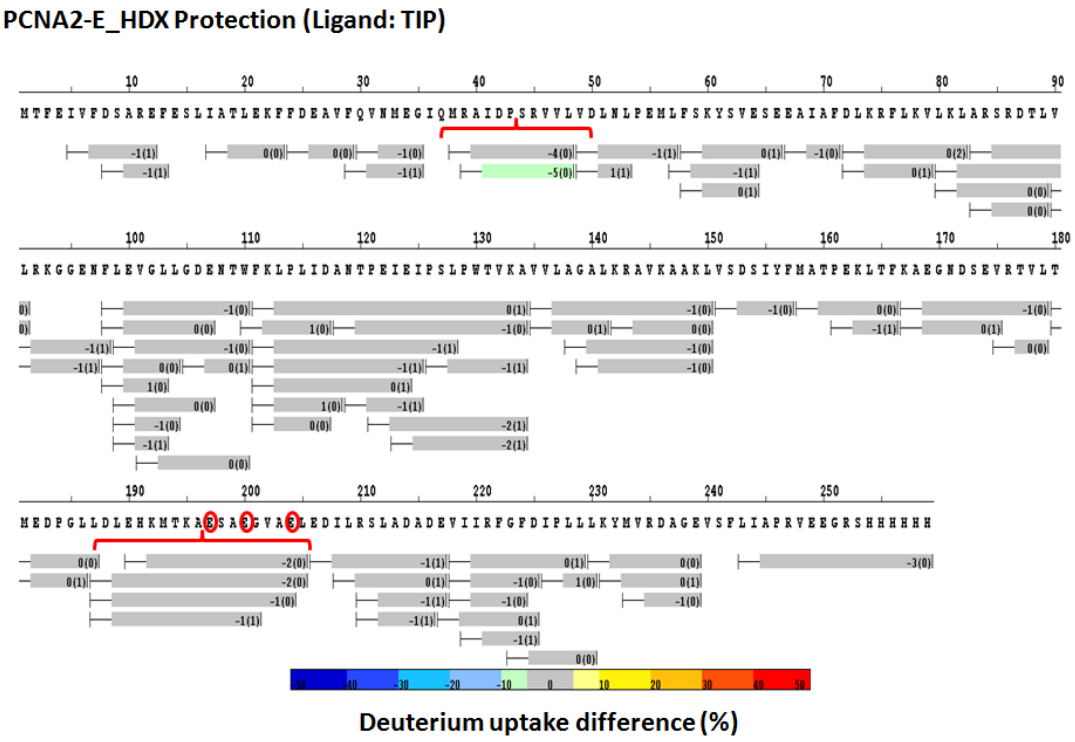

Figure 4-14: Differential HDX map of PCNA2-E mutant.

Differential HDX map of PCNA2-E mutant in which the regions having negative values of deuterium uptake difference (apo-PCNA2-E mutant vs. holo-PCNA2-E mutant) suggest the protein conformation change toward a more protected conformation upon the binding of TIP. (Experiments were performed by Richard Huang)

TIP is not essential for *T. kodakarensis* viability.

PCNA is an essential protein and plays essential roles in many nucleic acid metabolic processes. Therefore, TIP, which regulates PCNA activity, may also be essential for *T. kodakarensis* viability. Therefore a *T. kodakarensis* strain in which the TIP encoding genes, TK0808 deleted. *T. kodakarensis* THH2 (ΔTK0808) was readily obtained with the genome structure confirmed by diagnostic PCR (Figure 4-1B) and Southern blots (Figure 4-1C). The growth of the deleted strain was indistinguishable

from the wild- type strain (Figure 4-1D) suggesting that under normal growth conditions, TIP is not essential for *T. kodakarensis* viability.

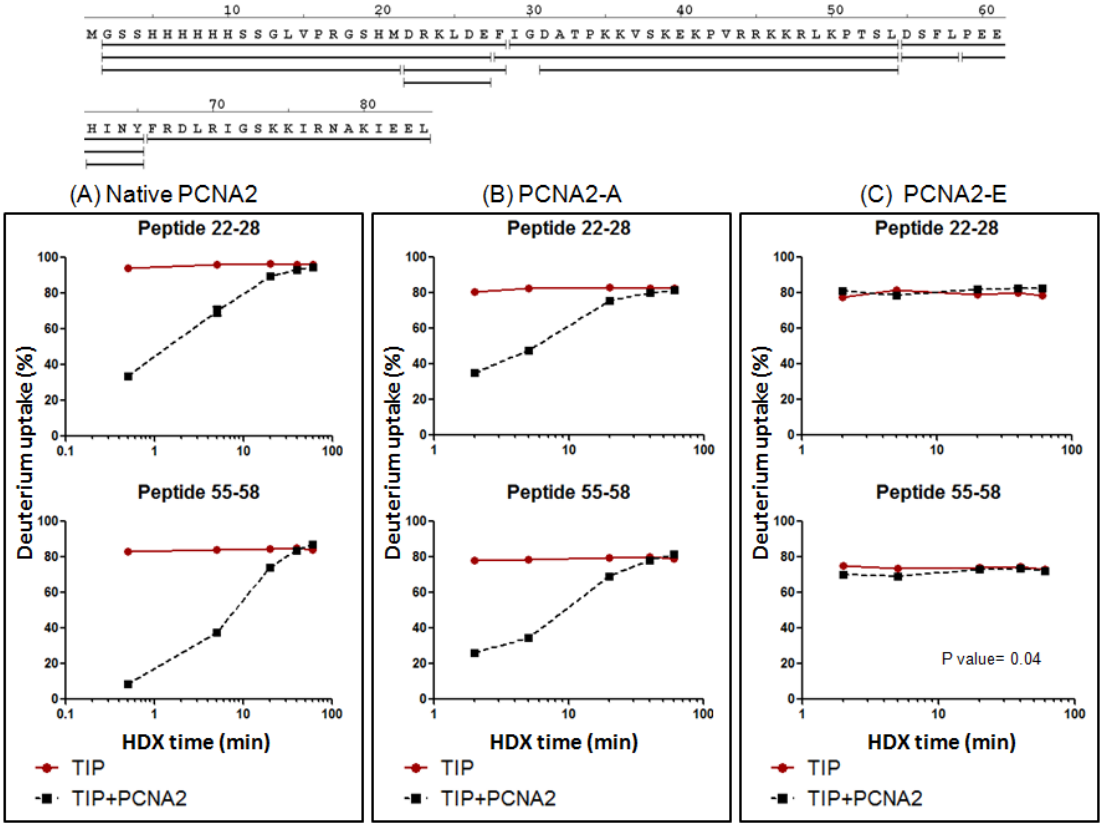

Figure 4-15: Mutation in PCNA2 affects its interactions with TIP.

HDX differences of apo-TIP (red curves) vs. holo-TIP (black curves). Regions of TIP that show significant differences in D uptake upon native PCNA2 binding are shown in (A). The HDX kinetics of the same regions of TIP in the presence of PCNA2-A mutant and PCNA2-E mutant are plotted in (B) and (C), respectively. (Experiments were performed by Richard Huang)

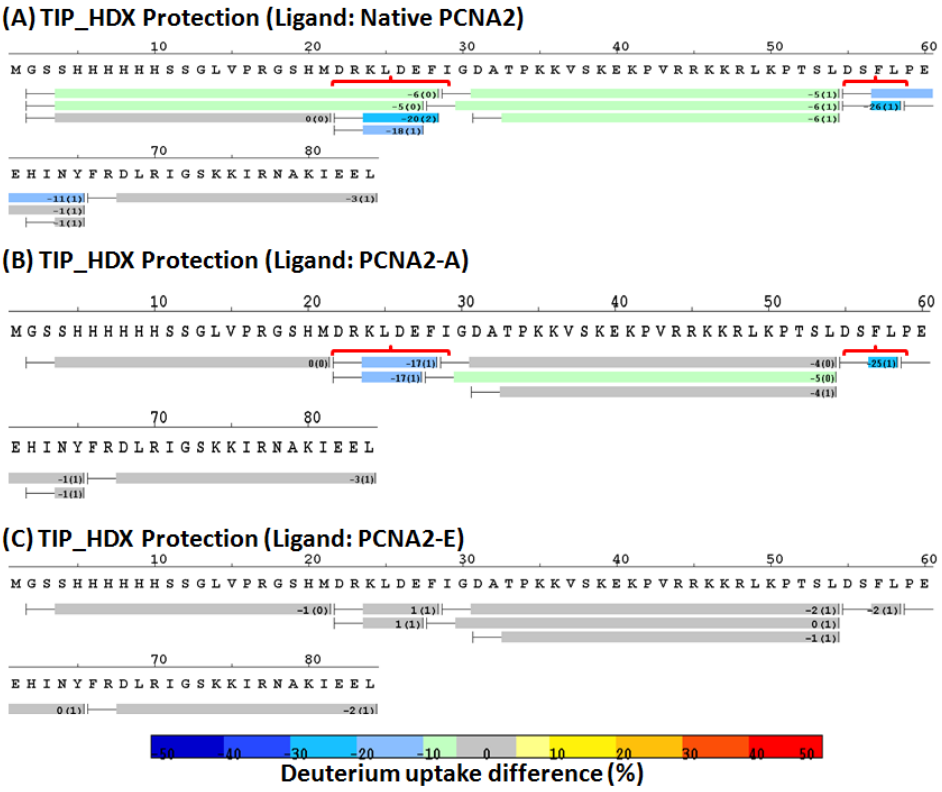

Figure 4-16: Differential HDX map of TIP in present of PCNAs.

Differential HDX map of TIP in which the regions having negative values of deuterium uptake difference (apo-TIP vs. holo-TIP) suggest the protein conformation change toward a more protected conformation upon the binding of native PCNA2 (A), PCNA2-A mutant (B), and PCNA2-E mutant (C). (Experiments were performed by Richard Huang)

Discussion

PCNA plays an essential role in many cellular processes, and thus the protein requires tight spatial and temporal regulation. In eukarya, a number of mechanisms have been shown to regulate PCNA functions including post-translational modification such as phosphorylation, acetylation, ubiquitination and sumoylation [for reviews see: (109,130)]. Although PCNA modification has not yet been reported in archaea, the presence of protein kinases and small-molecule modifiers in archaea

[reviewed in (131,132)] may suggest that similar mechanisms to regulate PCNA may also exist in this domain. Binding of a small protein is another mechanism for regulation of the eukaryotic PCNA. For example, the cell cycle regulator p21 binds to the IDCL and prevents PIP-containing client proteins from associating with PCNA.

The results presented here identified a possible new mechanism by which PCNA is regulated. Binding of TIP to PCNA does not involve a PIP-motif or the IDCL, yet it results in the inactivation of PCNA. However, TIP may regulate PCNA by two different mechanisms. The binding of TIP to PCNA1 appears to result in ring dissociation while the binding to PCNA2 does not (Figure 4-2). However, the activity of both PCNAs is inhibited. In one case it may be the dissociation of the ring while in the other TIP binding may prevent other protein interactions with the client enzyme. However, it was suggested that *in vivo* only PCNA1 may be required for cell viability (56,69). It is thus possible that, *in vivo*, ring dissociation in the main mechanism by which TIP regulate PCNA functions.

Although homologues of TIP have been identified only in *Thermococcales*, other archaeal species may also contain small proteins that bind PCNA and may have similar effects on the integrity of the PCNA ring.

Although PCNA is expressed in all cell types and has a long half-life, the protein is also overexpressed in tumor cells. It was suggested that PCNA could be a valuable target for cancer therapy. A number of small molecules that inhibit PCNA have been developed, most of which target the IDCL and thus block PCNA interactions with client enzymes [for examples see (133)]. The structures of PCNAs from all organisms are very similar (52), and therefore it is possible that the data presented here will

provide a new general mechanism for PCNA inhibition. One could design small molecule that mimic TIP, bind PCNA in the same region, and thus dissociate the trimeric ring. In the future, when the structure of PCNA-TIP complex is determined other interaction region may also be identified.

The bacterial β-subunit plays an essential role in DNA replication and it is the structural and functional homologue of PCNA (20). It was suggested that inhibition of the β-subunit function either by preventing it from interaction with other proteins or by other means may be a good antimicrobial drug [see (134) for a detail discussion]. The overall structures of PCNA and the β-subunit are very similar although β-subunit is a dimer and PCNA is a timer and there is no amino acid similarity between the two proteins (20). It may be possible to design a small molecule that will bind to similar region on the β-subunit as TIP binds on PCNA. Such binding may have similar effect resulting in the dissociation of the β-subunit ring.

Chapter 5: Concluding Remarks

Summary of results

The mechanisms of DNA replication are conserved through all life forms. The proteins involved in this process have similar functional property in the three domains of life. By analyzing genome sequences, archaea was shown to have similar DNA replication machinery to eukaryotes, which are evolutionarily distinct from that of bacteria. At the same time, the archaeal replication machinery involved fewer proteins and complexes in comparison the eukayotic process.. For this reason, archaeal DNA replication machinery is considered a simplified form of the eukaryotic machinery. *In vitro* studies also showed that archaeal replication enzymes have similar functions to their eukaryotic homologues. Based on bioinformatics approaches, a subset of eukaryote DNA replication protein homologues in archaea were identified. However, some DNA replication proteins that are essential in eukaryotes and/or bacteria are missing in archaea, which in turn possess a number of their own unique replication proteins. This raises the possibility that archaea might have more specific DNA replication proteins that have not yet been identified by bioinformatics approach.

The studies presented in this thesis aimed in identifying new proteins participating in archaeal DNA replication. To achieve this goal the *T. kodakarensis* was used as a model organism. In Chapter 2, AP followed by proteomic analysis was used to identify an archaeal DNA replication protein interaction network. In this network, previously documented and predicted interactions were confirmed, which

provides experimental evidence for unrecognized interactions between proteins with known and unknown functions, through which expanded the knowledge for archaeal replication research.

In Chapter 3 and 4, I continued to verify the protein interactions involved. GAN (TK1252)-GINS was selected as the bait protein GINS is an essential DNA replication protein in eukaryote. The GAN protein, by sequence analysis, contains a conserved DNA nuclease domain, which is a strong indication of participation in DNA metabolism. For the above reason, the GAN-GINS interaction was verified by *in vitro* experiments, which is described in Chapter 3. GAN is predicted to be a homologue of bacterial RecJ-like nuclease. A protein encoded by SSO0295 in *S. solfataricus* co-purified with the GINS complex from this crenarchaeon (31) and, in common with GAN, SSO0295p also has some limited sequence similarity to the DNA-binding domain of *E. coli* RecJ (31). In eukaryotes, Cdc45, MCM and GINS form a complex (CMG) and move along with the replication fork. All components of the CMG are essential for the replication of eukaryotic cells and hence their viability. Archaea possesses both MCM and GINS but no Cdc45 homologue has been identified based on sequence similarity. Due to the essentiality of Cdc45, we predict that it should exist in archaea and that the GAN is the leading candidate for the archaeal version of Cdc45. After the isolation and characterization of GAN, colleagues using bioinformatics technology reported the similarity between GAN and Cdc45 (41), which strongly supported that the GAN is a functional homologue of the eukaryotic Cdc45. To confirm the linkage between GAN and Cdc45, more effort

including structural studies on Cdc45 and GAN are necessary and are currently being pursued.

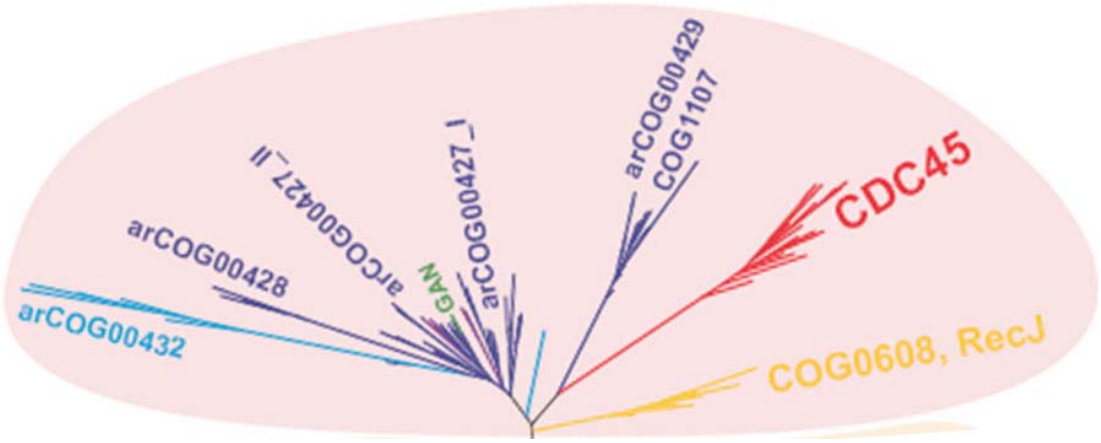

Figure 5-1: Phylogeny of the DHH superfamily.
Archaeal GAN and eukaryotic Cdc45 are homologue based on secondary structure and HHpred predication. This picture was adapted from (41).

To further characterize this network in Chapter 4, I investigated TIP (TK0808), a PCNA-binding protein from the interaction network. As a platform for a variety of DNA replication and repair proteins, PCNA is regulated by a series of proteins. However, the PCNA regulation protein in archaea has not yet been identified. In contrast to most other PCNA binding proteins, which interact with PCNA through conserved sequences (including PIP or KA motifs), the TIP does not have any known identified motifs. Our data suggest that TIP may act as a DNA replication negative regulator by disrupting the association between PCNA and its partners. A combination of H/D exchange and size exclusion chromatography showed that TIP binds to a novel interface that is distinct from classical IDCL region on PCNA.

Outlook for future studies

It is necessary to pay further attention to other new proteins in this network. Two of the candidates were encoded by TK1313 and TK0358, which were pulled down by MCM (TK0096) and RFC, respectively. A recent comprehensive whole-genome study showed the homologues of hypothetical genes TK1313 and TK0358 were essential to the viability of *Methanococcus maripaludis* (135). Although the detailed functions of these genes have not been confirmed, the association of these proteins with known replication proteins strongly argues that they are involved in DNA replication. Guided by this network, more efforts are needed to explore the new members participating in DNA replication.

It will be critical to determine the function of GAN's nuclease at the replication fork. It has been previously observed that the Fen1 protein, which is responsible for removing the 5' flap on the pre-matured Okazaki fragments, is not essential for viability (104). The recent progress from our collaborator indicates that the gene coding for GAN is not required for cell viability. This allows for speculation that GAN and Fen1 might have complementary function in the maturation of Okazaki fragments. One way to validate this possibility is to build a strain with a switch, which can control the expression of at least one of the GAN and Fen1 genes. Although inducible promoters are commonly used in bacteria and yeast the developing of similar switches in archaea is lagging behind. When available, a more detailed study on the roles of Fen1 and GAN could be evaluated.

Further study of the function and binding mechanism of TIP to PCNAs will also be necessary. The newly identified interface on PCNA has the potential to become a

new target for PCNA inhibitors, which have been used as drug targeting due to its up-regulation in cancer cells. In eukaryotes, the product of oncogene p53 activates the expression of p21, which is a peptide serving as a PCNA inhibitor. This may suggest that archaea has a similar mechanism to regulate PCNA through a small protein. TIP, although sharing no primary sequence identity with p21, might be a good candidate to be an archaeal version of p21. To explore the proteins interacting with TIP is also an important way to reveal its function(s). The AP-MS/MS strategy used in Chapter 2 should be interesting to be applied to expand the network involving TIP.

DNA replication is important. The process involves the participation and coordination of a multitude of proteins. However, the current understandings of DNA replication in eukaryotes are still limited, partly due to the complexity of the eukaryotic DNA replication machinery. Archaea provides a similar and simple platform to study this process. The information provided in this study will be able to provide clues for studying DNA replication in eukaryotes. The strategy used in this study can be applied to other archaea species in order to address DNA replication processes. The isolation of protein complex upstream of MS is neither limited to affinity purification used in this study. Alternative separation methods like ultracentrifugation and immunoprecipitation are also expected to isolate the protein complexes. In addition, AP-MS/MS can be called upon to identify new proteins in other research fields outside DNA replication.

This work mainly focuses on discovery and verification of new proteins and interactions involved in DNA replication using *T. kodakarensis* as a model system. Using genetic and biophysical methods, unknown proteins were shown to be involved

in DNA replication. These new proteins are strongly hypothesized to be involved in DNA replication. *In vitro* experiments were also used to verify the new identified interactions. Followed scientists are expected to continue expanding the study of this network. Plasmids for expressing fish proteins are listed in Appendix 7 and waiting for the new explorers.

List of Appendices

Appendices

Appendix 1: Protein-protein interactions identified in the study.

Gene #	Score	MW(Da)	Peptide matches	Percent coverage	Function
TK1901 (Cdc6) was tagged					
TK1901	1259	47777	83	55	Cdc6
TK0038	448	31444	16	29	flagellin
TK2218	336	37170	13	13	RFC-S
TK0535	313	28222	15	33	PCNA1
TK2257	287	20549	12	40	deoxycytidylate deaminase
TK0563	281	18180	19	40	6-pyruvoyl-tetrahydropterin synthase-related protein
TK1990	269	44387	14	23	cysteine desulfurase
TK0040	238	27426	9	24	flagellin
TK2106	237	46763	4	14	phosphopyruvate hydratase
TK1314	233	50095	9	26	ATPase
TK1903	186	150190	10	2	PolD-S
TK2100	165	36201	7	12	thioredoxin reductase
TK1415	162	10781	3	29	50S ribosomal protein L12P
TK1696	156	11435	7	47	30S ribosomal protein S24e
TK2217	152	44001	5	7	2-amino-3-ketobutyrate coenzyme A ligase
TK0657	140	47268	4	10	ABC transporter periplasmic component
TK0309	122	82109	5	6	elongation factor EF-2
TK1981	118	34865	5	8	2-ketoisovalerate ferredoxin oxidoreductase subunit beta
TK1276	115	17263	5	24	30S ribosomal protein S19e
TK1966	100	32950	2	9	D-3-phosphoglycerate dehydrogenase
TK0593	100	45867	2	6	Protein with unknown function
TK2140 (DNA ligase) was tagged					
TK2140	3453	64042	774	82	DNA ligase
TK1903	291	150190	13	4	PolD-S
TK1009	250	96657	13	9	putative 5-methylcytosine restriction system, GTPase subunit
TK0563	229	18180	15	40	6-pyruvoyl-tetrahydropterin synthase-related protein
TK0063	196	21240	12	28	nucleotidyltransferase/DNA-binding domain-containing protein
TK2217	175	44001	6	9	2-amino-3-ketobutyrate coenzyme A ligase
TK0361	146	16513	1	25	Protein with unknown function
TK1481	141	48923	7	13	NADH:polysulfide oxidoreductase
TK0455	131	37834	4	12	Protein with unknown function

TK2303	128	59178	4	7	chaperonin beta subunit
TK2145	126	69683	6	3	Protein with unknown function
TK0847	124	8671	4	59	Protein with unknown function
TK1505	120	21184	4	13	30S ribosomal protein S4
TK1022	117	30374	3	9	D-aminopeptidase
TK0759	105	49864	6	7	asparaginyl-tRNA synthetase
TK2100	105	36201	6	9	thioredoxin reductase
TK0535	103	28222	4	13	PCNA1

TK1410 (DnaG-like) was tagged

TK0790	589	16212	35	82	Protein with unknown function
TK1633	352	29757	15	46	exosome complex RNA-binding protein Rrp42
TK2100	278	36201	13	22	thioredoxin reductase
TK2278	268	42443	13	19	myo-inositol-1-phosphate synthase
TK2227	235	54037	11	12	RNA-binding protein FAU-1
TK1634	202	27668	10	23	exosome complex exonuclease Rrp41
TK1047	194	30376	8	19	translation initiation factor IF-2B subunit alpha
TK1481	186	48923	13	22	NADH:polysulfide oxidoreductase
TK2303	132	59178	6	10	chaperonin beta subunit
TK2074	121	39247	6	7	putative glutamate synthase subunit beta
TK1331	121	8750	5	46	Lrp/AsnC family transcriptional regulator
TK0309	114	82109	3	4	elongation factor EF-2
TK1254	104	23007	5	16	30S ribosomal protein S3Ae

TK1281 (Fen1) was tagged

TK1281	2608	38786	581	82	Fen1
TK0535	1967	28222	361	89	PCNA1
TK0038	336	31444	11	21	flagellin
TK0643	305	13183	10	35	prefoldin subunit beta
TK0563	304	18180	21	36	6-pyruvoyl-tetrahydropterin synthase-related protein
TK1005	252	16265	16	27	prefoldin subunit alpha
TK2303	249	59178	6	10	chaperonin beta subunit
TK2106	229	46763	7	11	phosphopyruvate hydratase
TK0040	204	27426	10	15	flagellin
TK0569	200	38673	7	11	Protein with unknown function
TK1046	199	147354	5	3	Protein with unknown function
TK1532	196	13355	8	40	30S ribosomal protein S17P
TK0590	192	8502	10	63	Protein with unknown function
TK1481	185	48923	5	7	NADH:polysulfide oxidoreductase
TK0358	184	53699	5	6	Protein with unknown function TK0358
TK2074	183	39247	10	10	putative glutamate synthase subunit beta
TK1622	183	57145	4	7	methylmalonyl-CoA decarboxylase, alpha subunit

TK1254	182	23007	7	18	30S ribosomal protein S3Ae
TK0330	178	14679	3	20	methylmalonyl-CoA epimerase
TK0296	169	34405	4	9	quinolinate synthetase
TK1500	164	15318	6	25	30S ribosomal protein S9P
TK0309	153	82109	4	5	elongation factor EF-2
TK0593	153	45867	5	13	Protein with unknown function
TK0180	145	41161	3	7	acetyl-CoA acetyltransferase
TK2049	138	72164	5	7	acylamino acid-releasing enzyme
TK1984	132	36379	2	12	pyruvate ferredoxin oxidoreductase subunit beta
TK2035	131	45164	3	8	glycine cleavage system aminomethyltransferase T
TK2100	130	36201	5	7	thioredoxin reductase
TK0657	124	47268	3	7	ABC transporter periplasmic component
TK2278	120	42443	2	4	myo-inositol-1-phosphate synthase
TK1542	113	39022	3	7	50S ribosomal protein L3P
TK1083	111	127604	4	3	DNA-directed RNA polymerase subunit beta
TK0814	107	47062	2	4	type A flavoprotein
TK1840	106	58957	2	5	cobalt-activating carboxypeptidase
TK1521	101	26487	4	13	30S ribosomal protein S5P

TK0536 (GINS51) was tagged

TK1903	895	150190	41	15	PolD-S
TK1252	858	52858	25	32	ssDNA-specific exonuclease
TK1046	659	147354	22	10	Protein with unknown function
TK1902	488	80848	14	15	PolD-L
TK0643	393	13183	18	52	prefoldin subunit beta
TK0536	364	21583	14	44	GINS51
TK1005	353	16265	11	55	prefoldin subunit alpha
TK2278	269	42443	8	18	myo-inositol-1-phosphate synthase
TK1619	191	19154	7	26	GINS23
TK1696	151	11435	3	47	30S ribosomal protein S24e
TK1521	140	26487	5	11	30S ribosomal protein S5P
TK1505	118	21184	4	13	30S ribosomal protein S4
TK1529	105	27753	4	7	30S ribosomal protein S4e
TK0348	103	29743	3	13	membrane protease subunit stomatin/prohibitin-like protein

TK1619 (GINS23) was tagged

TK1619	1565	19154	141	74	GINS23
TK0582	1119	28429	53	82	PCNA2
TK1186	451	43957	15	26	Protein with unknown function
TK2129	359	23756	15	25	triosephosphate isomerase
TK2303	267	59178	7	15	chaperonin beta subunit
TK2100	264	36201	11	20	thioredoxin reductase
TK0535	175	28222	5	17	PCNA1

TK1984	143	36379	7	11	pyruvate ferredoxin oxidoreductase subunit beta
TK1981	134	34865	4	16	2-ketoisovalerate ferredoxin oxidoreductase subunit beta
TK1618	119	18769	5	18	calcineurin superfamily metallophosphoesterase
TK0569	115	38673	2	9	Protein with unknown function
TK0664	115	24671	3	8	hypoxanthine/guanine phosphoribosyltransferase
TK2021	107	32017	4	6	ParA/MinD family ATPase
TK2074	103	39247	6	7	putative glutamate synthase subunit beta
TK1284	100	18461	6	20	intracellular protease I

TK0096 (MCM1) was tagged

TK0096	676	103285	30	18	MCM1
TK0563	395	18180	29	48	6-pyruvoyl-tetrahydropterin synthase-related protein
TK1313	195	23503	12	20	Protein with unknown function
TK0063	194	21240	8	36	nucleotidyltransferase/DNA-binding domain-containing protein
TK1903	174	150190	11	3	PolD-S
TK0535	137	28222	5	14	PCNA1
TK2211	109	102779	6	4	chromosome segregation protein
TK0590	106	8502	3	21	Protein with unknown function
TK1990	102	44387	5	8	cysteine desulfurase

TK1361 (MCM2) was tagged

TK2021	109				
TK1622	108	57145	4	4	methylmalonyl-CoA decarboxylase, alpha subunit
TK0664	108	24671	2	8	hypoxanthine/guanine phosphoribosyltransferase
TK2074	107	39247	7	7	putative glutamate synthase subunit beta
TK1506	106	16981	3	20	30S ribosomal protein S13P
TK0570	106	53782	2	4	ABC-type iron(III) transport system, periplasmic component
TK0765	105	37269	3	7	glyceraldehyde-3-phosphate dehydrogenase
TK1840	103	58957	4	4	cobalt-activating carboxypeptidase

TK1903(PolD-S) was tagged

TK1903	427	150190	21	8	PolD-S
TK1902	197	80848	9	6	PolD-L
TK1005	181	16265	6	20	prefoldin subunit alpha
TK0643	152	13183	6	27	prefoldin subunit beta
TK0535	123	28222	4	12	PCNA1
TK1252	102	52858	4	5	ssDNA-specific exonuclease

TK1790(Pri-L)

was tagged

TK1791	442	40443	26	28	Pri-S
TK2100	342	36201	18	26	thioredoxin reductase
TK2303	315	59178	14	18	chaperonin beta subunit
TK1481	264	48923	12	31	NADH:polysulfide oxidoreductase
TK1790	187	46893	9	15	Pri-L
TK1186	185	43957	7	24	Protein with unknown function
TK2278	176	42443	10	20	myo-inositol-1-phosphate synthase
TK1966	172	32950	12	18	D-3-phosphoglycerate dehydrogenase
TK0569	154	38673	5	12	Protein with unknown function
TK0309	153	82109	5	8	elongation factor EF-2
TK1254	139	23007	8	20	30S ribosomal protein S3Ae
TK1415	135	10781	4	29	50S ribosomal protein L12P
TK2211	132	102779	10	5	chromosome segregation protein
TK0535	131	28222	5	10	PCNA1
TK0563	122	18180	9	21	6-pyruvoyl-tetrahydropterin synthase-related protein
TK2217	121	44001	4	7	2-amino-3-ketobutyrate coenzyme A ligase
TK2052	119	8624	6	39	Lrp/AsnC family transcriptional regulator
TK1496	113	22992	4	11	30S ribosomal protein S2
TK1974	112	13096	6	38	carboxymuconolactone decarboxylase-related protein
TK1789	109	28203	4	15	ATPase

TK1792(Primase related protein) was tagged

TK1005	280	16265	13	34	prefoldin subunit alpha
TK1046	279	147354	12	5	Protein with unknown function
TK1481	203	48923	7	19	NADH:polysulfide oxidoreductase
TK2074	148	39247	5	11	putative glutamate synthase subunit beta
TK1633	133	29757	3	17	exosome complex RNA-binding protein Rrp42
TK1990	125	44387	8	12	cysteine desulfurase
TK1521	121	26487	3	9	30S ribosomal protein S5P
TK2100	112	36201	2	9	thioredoxin reductase

TK2218(RFC-S) was tagged

TK2218	968	37170	58	26	RFC-S
TK2219	388	57239	24	15	RFC-L
TK2100	174	36201	8	13	thioredoxin reductase
TK0330	136	14679	4	19	methylmalonyl-CoA epimerase
TK0814	131	47062	3	6	type A flavoprotein
TK1505	119	21184	5	13	30S ribosomal protein S4

TK2219(RFC-L) was tagged

TK2218	1412	37170	149	29	RFC-S

TK0582	885	28429	45	60	PCNA2
TK2219	841	57239	62	32	RFC-L
TK0643	264	13183	7	28	prefoldin subunit beta
TK2138	198	20232	6	38	orotate phosphoribosyltransferase
TK1496	170	22992	3	18	30S ribosomal protein S2
TK1974	139	13096	3	38	carboxymuconolactone decarboxylase-related protein
TK0569	131	38673	2	9	Protein with unknown function
TK0038	118	31444	8	8	flagellin
TK1245	113	15009	5	15	Protein with unknown function
TK1505	106	21184	6	13	30S ribosomal protein S4
TK1046	105	147354	3	1	Protein with unknown function

TK1960(RPA2) was tagged

TK1046	1196	147354	43	17	Protein with unknown function
TK0795	879	79742	29	34	putative 5-methylcytosine restriction system, GTPase subunit
TK1521	552	26487	23	48	30S ribosomal protein S5P
TK1005	516	16265	20	55	prefoldin subunit alpha
TK0657	416	47268	11	19	ABC transporter periplasmic component
TK0643	405	13183	14	44	prefoldin subunit beta
TK2303	396	59178	13	15	chaperonin beta subunit
TK0038	392	31444	16	29	flagellin
TK0183	384	25316	13	37	fibrillarin
TK0358	378	53699	10	17	Protein with unknown function
TK0309	375	82109	11	11	elongation factor EF-2
TK1276	369	17263	15	56	30S ribosomal protein S19e
TK1254	343	23007	20	30	30S ribosomal protein S3Ae
TK1532	333	13355	23	43	30S ribosomal protein S17P
TK1496	316	22992	16	26	30S ribosomal protein S2
TK1542	312	39022	9	20	50S ribosomal protein L3P
TK1536	301	23340	12	35	30S ribosomal protein S3P
TK1529	287	27753	9	33	30S ribosomal protein S4e
TK0040	252	27426	11	16	flagellin
TK0563	246	18180	13	20	6-pyruvoyl-tetrahydropterin synthase-related protein
TK1481	229	48923	10	17	NADH:polysulfide oxidoreductase
TK1099	209	7283	8	62	30S ribosomal protein S27e
TK2074	207	39247	6	18	putative glutamate synthase subunit beta
TK1622	206	57145	5	10	methylmalonyl-CoA decarboxylase, alpha subunit
TK1539	204	26082	6	13	50S ribosomal protein L2P
TK2100	200	36201	7	12	thioredoxin reductase
TK1696	179	11435	12	47	30S ribosomal protein S24e
TK0184	174	47201	5	10	snoRNP component, Nop56p/58p-like protein
TK0180	174	41161	2	9	acetyl-CoA acetyltransferase

TK1541	166	28695	5	10	50S ribosomal protein L4P
TK0330	166	14679	4	19	methylmalonyl-CoA epimerase
TK2106	164	46763	3	7	phosphopyruvate hydratase
TK1500	160	15318	9	27	30S ribosomal protein S9P
TK0348	158	29743	8	18	membrane protease subunit stomatin/prohibitin-like protein
TK0423	158	70055	5	7	prolyl endopeptidase
TK2278	155	42443	5	8	myo-inositol-1-phosphate synthase
TK2021	152	32017	4	14	ParA/MinD family ATPase
TK1078	149	16439	7	40	30S ribosomal protein S12P
TK1633	148	29757	5	14	exosome complex RNA-binding protein Rrp42
TK1981	143	34865	5	10	2-ketoisovalerate ferredoxin oxidoreductase subunit beta
TK1417	142	24053	4	13	50S ribosomal protein L1P
TK1505	140	21184	9	23	30S ribosomal protein S4
TK1984	139	36379	3	10	pyruvate ferredoxin oxidoreductase subunit beta
TK1083	138	127604	6	3	DNA-directed RNA polymerase subunit beta
TK0296	138	34405	2	7	quinolinate synthetase
TK1637	136	29250	6	12	proteasome subunit alpha
TK0471	136	30863	5	12	transcription regulator
TK0814	136	47062	4	10	type A flavoprotein
TK0946	133	15865	4	19	Protein with unknown function
TK2035	131	45164	2	5	glycine cleavage system aminomethyltransferase T
TK0765	130	37269	2	6	glyceraldehyde-3-phosphate dehydrogenase
TK1538	130	15411	8	11	30S ribosomal protein S19P
TK0569	130	38673	5	8	Protein with unknown function TK0569
TK1502	128	13764	4	20	50S ribosomal protein L18e
TK1112	127	6556	6	35	transcription elongation factor NusA
TK0044	124	54137	6	8	flagella-related protein D, internal insertion
TK2290	123	49681	5	8	ribulose bisophosphate carboxylase
TK2253	121	28743	9	20	Protein with unknown function
TK1880	118	51019	3	8	succinyl-CoA synthetase (NDP forming), large subunit
TK1506	110	16981	5	22	30S ribosomal protein S13P
TK2246	110	32678	4	6	L-asparaginase
TK0759	110	49864	4	6	asparaginyl-tRNA synthetase
TK1186	110	43957	4	11	Protein with unknown function
TK0719	109	39062	4	8	ABC-type molybdate transport system, ATPase component
TK0664	109	24671	2	8	hypoxanthine/guanine phosphoribosyltransferase
TK1579	107	33947	1	5	ABC-type multidrug transport system, ATPase component
TK1596	107	11840	3	18	V-type ATP synthase subunit H

TK2091	107	48428	3	5	membrane bound hydrogenase, NiFe-hydrogenase large subunit 2
TK0593	105	45867	3	12	Protein with unknown function
TK1960	103	14356	5	13	RPA2
TK1548	103	42933	2	6	serine--glyoxylate aminotransferase, class V
TK2211	100	102779	21	3	Rad50
TK1961(RPA3) was tagged					
TK1959	1472	31240	267	65	RPA1
TK0470	932	141310	67	14	reverse gyrase
TK1960	734	14356	105	86	RPA2
TK1188	468	46297	44	29	sugar-phosphate nucleotydyltransferase
TK0563	168	18180	9	20	6-pyruvoyl-tetrahydropterin synthase-related protein
TK1017	133	135911	10	3	chromosome segregation ATPase
TK0001	117	90030	17	3	PolB
TK2074	112	39247	8	12	putative glutamate synthase subunit beta
TK1083	112	127604	6	3	DNA-directed RNA polymerase subunit beta
TK0446	110	27210	4	9	Protein with unknown function
TK0296	109	34405	3	6	quinolinate synthetase

Proteins with MASCOT scores higher than 100 identified in the eluate containing His-tagged proteins are listed along with the proteins molecular weight, number of peptides matched, and the percentage of its amino acid sequence covered by the matching peptides. None of the listed proteins was detected in equivalent column fractions prepared from the untagged KW128 strain. (See text for further details.)

Appendix 2: A subset of the protein-protein interactions identified in the study.

Gene #	Score	MW (Da)	Peptide Matches	Percent coverage	Function
Cdc6	TK1901				
TK1901	1259	47777	83	55	Cdc6
TK2218	336	37170	13	13	RFC-S
TK0535	313	28222	15	33	PCNA1
TK1314	233	50095	9	26	ATPase
TK1903	186	150190	10	2	PolD-S
TK0593	100	45867	2	6	Protein with unknown function
DNA ligase	TK2140				
TK2140	3453	64042	774	82	DNA ligase
TK1903	291	150190	13	4	PolD-S
TK0063	196	21240	12	28	nucleotidyltransferase/DNA-binding domain-containing protein
TK0361	146	16513	1	25	Protein with unknown function
TK0455	131	37834	4	12	Protein with unknown function
TK2145	126	69683	6	3	Protein with unknown function
TK0847	124	8671	4	59	Protein with unknown function
TK0535	103	28222	4	13	PCNA1
DnaG-like	TK1410				
TK0790	589	16212	35	82	Protein with unknown function
TK1633	352	29757	15	46	exosome complex RNA-binding protein Rrp42
TK2227	235	54037	11	12	RNA-binding protein FAU-1
TK1634	202	27668	10	23	exosome complex exonuclease Rrp41
Fen1	TK1281				
TK1281	2608	38786	581	82	Fen1
TK0535	1967	28222	361	89	PCNA1
TK0569	200	38673	7	11	Protein with unknown function
TK1046	199	147354	5	3	Protein with unknown function
TK0590	192	8502	10	63	Protein with unknown function
TK0358	184	53699	5	6	Protein with unknown function
TK0593	153	45867	5	13	Protein with unknown function

GINS51	TK0536				
TK1903	895	150190	41	15	PolD-S
TK1252	858	52858	25	32	ssDNA-specific exonuclease
TK1046	659	147354	22	10	Protein with unknown function
TK1902	488	80848	14	15	PolD-L
TK0536	364	21583	14	44	GINS51
TK1619	191	19154	7	26	GINS23
GINS23	TK1619				
TK1619	1565	19154	141	74	GINS23
TK0582	1119	28429	53	82	PCNA2
TK1186	451	43957	15	26	Protein with unknown function
TK0535	175	28222	5	17	PCNA1
TK0569	115	38673	2	9	Protein with unknown function
TK2021	107	32017	4	6	ParA/MinD family ATPase
MCM1	TK0096				
TK0096	676	103285	30	18	MCM1
TK1313	195	23503	12	20	Protein with unknown function
TK0063	194	21240	8	36	nucleotidyltransferase/DNA-binding domain-containing protein
TK1903	174	150190	11	3	PolD-S
TK0535	137	28222	5	14	PCNA1
TK2211	109	102779	6	4	Rad50
TK0590	106	8502	3	21	Protein with unknown function
MCM2	TK1361				
TK0535	925	28222	46	66	PCNA1
TK1903	358	150190	21	6	PolD-S
TK0063	230	21240	9	32	nucleotidyltransferase/DNA-binding domain-containing protein
TK0590	152	8502	11	51	Protein with unknown function
TK0682	150	65448	11	12	MutS-like DNA mismatch repair ATPase
TK0001	100	90030	17	2	PolB
MCM3	TK1620				
TK1245	167	15009	8	23	Protein with unknown function
TK0590	164	8502	7	53	Protein with unknown function
TK0569	146	38673	5	7	Protein with unknown function

TK0467	124	118109	4	3	Protein with unknown function
TK1046	100	147354	6	2	Protein with unknown function

PCNA1	TK0535

TK0535	2098	28222	828	91	PCNA1
TK2218	265	37170	13	10	RFC-S
TK0808	259	7606	15	48	Protein with unknown function
TK2219	105	57239	5	6	RFC-L

PCNA2	TK0582

TK0582	1723	28429	166	84	PCNA2
TK1046	754	147354	28	11	Protein with unknown function
TK0569	173	38673	6	15	Protein with unknown function
TK0535	132	28222	5	18	PCNA1
TK0953	130	67743	5	7	ATPase
TK1849	110	19094	2	11	Protein with unknown function

PolB	TK0001

TK0001	4027	90030	556	35	PolB
TK1046	576	147354	25	10	Protein with unknown function
TK0569	210	38673	9	15	Protein with unknown function
TK2021	109	32017	4	8	ParA/MinD family ATPase

PolD-S	TK1903

TK1903	427	150190	21	8	PolD-S
TK1902	197	80848	9	6	PolD-L
TK0535	123	28222	4	12	PCNA1
TK1252	102	52858	4	5	ssDNA-specific exonuclease

Pri-L	TK1790

TK1791	442	40443	26	28	Pri-S
TK1790	187	46893	9	15	Pri-L
TK1186	185	43957	7	24	Protein with unknown function
TK0569	154	38673	5	12	Protein with unknown function
TK2211	132	102779	10	5	Rad50
TK0535	131	28222	5	10	PCNA1
TK1789	109	28203	4	15	ATPase

Unknown	TK1792

TK1046	279	147354	12	5	Protein with unknown function
TK1633	133	29757	3	17	exosome complex RNA-binding protein Rrp42
RFC-S	TK2218				
TK2218	968	37170	58	26	RFC-S
TK2219	388	57239	24	15	RFC-L
RFC-L	TK2219				
TK2218	1412	37170	149	29	RFC-S
TK0582	885	28429	45	60	PCNA2
TK2219	841	57239	62	32	RFC-L
TK1245	113	15009	5	15	Protein with unknown function
TK0569	131	38673	2	9	Protein with unknown function
TK1046	105	147354	3	1	Protein with unknown function
RPA2	TK1960				
TK1046	1196	147354	43	17	Protein with unknown function
TK0358	378	53699	10	17	Protein with unknown function
TK2021	152	32017	4	14	ParA/MinD family ATPase
TK1633	148	29757	5	14	exosome complex RNA-binding protein Rrp42
TK0946	133	15865	4	19	Protein with unknown function
TK2253	121	28743	9	20	Protein with unknown function
TK0719	109	39062	4	8	ABC-type molybdate transport system, ATPase component
TK1579	107	33947	1	5	ABC-type multidrug transport system, ATPase component
TK0593	105	45867	3	12	Protein with unknown function
TK1960	103	14356	5	13	replication factor A complex, RPA14 subunit
TK2211	100	102779	21	3	Rad50
RPA3	TK1961				
TK1959	1472	31240	267	65	RPA1
TK0470	932	141310	67	14	reverse gyrase
TK1960	734	14356	105	86	RPA2
TK1017	133	135911	10	3	chromosome segregation ATPase
TK0001	117	90030	17	3	PolB
TK0446	110	27210	4	9	Protein with unknown function

Proteins with MASCOT scores higher than 100 identified in the eluate containing His-tagged proteins are listed along with the proteins molecular weight, number of peptides matched, and the percentage of its amino acid sequence covered by the matching peptides. None of the listed proteins was detected in equivalent column fractions prepared from the untagged KW128 strain.

Appendix 3: Primers used to generate the constructs for the study in Chapter 2.

Genes to tag	Name	Sequences	Res. Sites
TK1619 (GINS2)	Zhuo_0001	5'-AGGAAGCTTAGTACCTCGAAAGGACTTCG-3'	Hind III
	Zhuo_0002	5'-ATTCTGCAGTCACGATAACCCTATAAGCC-3'	Pst I
	Zhuo_0003	5'-AGCTCTAGAACTGCCCTGACAGAATGATC-3'	Xba I
	Zhuo_0006	5'-TTAGAATTCCCTCAAGAATCTTCTTGAATATG-3'	EcoR I
	Zhuo_0306	5'-GAAAGGTGCAAATCATGCATCATCATCATCATCATT TCACGGGTAAAGC-3'	
	Zhuo_0307	5'-GCTTTACCCGTGAAATGATGATGATGATGATGCATGA TTTGCACCTTTC-3'	
TK0001 (PolB)	Zhuo_0007	5'-CAGGCATGCGAACTACGTTGAGATACCGAAG-3'	Sph I
	Zhuo_0008	5'-TGACTGCAGAAATGGAAAGGTCAATGATGATGATGAT GATGAGTTCCCTTCGGCTTCAGCCAAG-3'	Pst I
	Zhuo_0009	5'-CCATCTAGAGTTTTCCAGCGGATAACCCT-3'	Xba I
	Zhuo_0010	5'-CTCGGTACCCCGAGAACCTCATCGTAGCGCG-3'	Kpn I
TK2219 (RFC2)	Zhuo_0011	5'-ATGGCATGCGGTAAGTGATCCCTATGACG-3'	Sph I
	Zhuo_0012	5'-AAACTCGAGGCAATACAATTCAATGATGATGATGAT GATGCTTCTTGAGGAAGTCGAACAGC-3'	
	Zhuo_0012m	5'-AAAGCATGCGCAATACAATTCAAT-3'	SphI
	Zhuo_0013	5'-AAATCTAGAGAAAAATATAAAAACCCACT-3'	Xba I
	Zhuo_0014	5'-AAAGAATTCCGAGCATATCCCTCGCCAGCT-3'	EcoR I
TK0536 (GINS1)	Zhuo_0015	5'-GCATAAGCATGCGCGCAAT-3'	Sph I
	Zhuo_0016	5'-TCTCTGCAGAGGACTTACTTTAATGATGATGATGATGA TGGAGGAATATCCTTACTCTTC-3'	Pst I
	Zhuo_0017	5'-CAGGGATCCTTTAAGTTGCTCCAGTTTTTATTC-3'	BamH I
	Zhuo_0018	5'-TTTAAGGAATTCCGATCCGGATGGGCTGTTC-3'	EcoR I
TK2218 (RFC1)	Zhuo_0019	5'-TATGCATGCAGTGAGCAGCGCCACTAGGCA-3'	Sph I
	Zhuo_0020	5'-GGGCTGCAGTTGTCTTCTCGTGGAACTGCA-3'	Pst I
	Zhuo_0021	5'-CCCTCTAGATGGTTGGTTGGGTACAGCGTCCCAGTGA GACGGCAAAAGCCTTAAACTCCGGGGGCTCATAAAGG TTTAG-3'	Xba I
	Zhuo_0022	5'-CTCATAAAGGTTTAGGTGAAAATCCATGCATCATCATC ATCATCATAT GTCCGAGGAAGTGAAGGAAGTTAAAATTC-3'	
	Zhuo_0023	5'-GGAGAATTCGGTCGTAGATGACGAAGTCG-3'	EcoR I
TK0535 (PCNA1)	Zhuo_0015	5'-GCATAAGCATGCGCGCAAT-3'	Sph I
	Zhuo_0017	5'-CAGGGATCCTTTAAGTTGCTCCAGTTTTTATTC-3'	BamH I
	Zhuo_0018	5'-TTTAAGGAATTCCGATCCGGATGGGCTGTTC-3'	EcoR I
	Zhuo_0034	5'-AGAGTTGAGGGAGGTCAATGATGATGATGATGATGG CTCCTCAACGCGCGGAGCGA-3'	
	Zhuo_0035	5'-TCGCTCCGCGCGTTGAGGAGCATCATCATCATCATCA	

		TTGACCTCCCTCAACTCT-3'	
	Zhuo_0036	5'-ATATCTGCAGAGGACTTACTTTAGAGGAATATC-3'	Pst I
TK1960 (RPA2)	Zhuo_0037	5'-AAGAGCATGCTCAGAGCGGCTCGTAGAAGC-3'	Sph I
	Zhuo_0040	5'-GTGTCTGCAGGCCCTATAGTTTTCCCATTTC-3'	Pst I
	Zhuo_0041	5'-GTGTTCTAGACTAGAAAAACTTCAATCGGAC-3'	Xba I
	Zhuo_0042	5'-TATGGAATTCGAGGGGATTCATGCCACCAC-3'	EcoR I
	Zhuo_0043m	5'-GAGTTCTTCGGAGGTGGTGAGCATCATCATCATCATC ATGAATGAAGAAGCGCCTACCAG-3'	
	Zhuo_0044m	5'-CTGGTAGGCGCTTCTTCATTCATGATGATGATGATGAT GCTCACCACCTCCGAAGAACTC-3'	
TK1961 (RPA3)	Zhuo_0037	5'-AAGAGCATGCTCAGAGCGGCTCGTAGAAGC-3'	Sph I
	Zhuo_0040	5'-GTGTCTGCAGGCCCTATAGTTTTCCCATTTC-3'	Pst I
	Zhuo_0041	5'-GTGTTCTAGACTAGAAAAACTTCAATCGGAC-3'	Xba I
	Zhuo_0042	5'-TATGGAATTCGAGGGGATTCATGCCACCAC-3'	EcoR I
	Zhuo_0045	5'-GAAGGAGTACGGCCTTGAGCATCATCATCATCATCAT TGAGGTGGTTGAAATGGAGG-3'	
	Zhuo_0046	5'-CCTCCATTTCAACCACCTCAATGATGATGATGATGAT GCTCAAGGCCGTACTCCTTC-3'	
TK1901 (Cdc6)	EJ5	5'-CCCGGTACCTTGATGTGGGTAAAACATGGGGAGG-3'	Kpn I
	EJ6	5'-CCCGGATCCGAAAACACGAACAATGATGGACAAA AGGG-3'	BamH I
	Zhuo_0047	5'-GGGCTGCAGATGTATATCCCTGTTCATTCCTC-3'	Pst I
	Zhuo_0050	5'-CCCCTGCAGACCCATTCTCACTCCCGTTTTCCACTG GAAGGG-3'	Pst I
	Zhuo_0051	5'-CGAATATTGAACTGAGGTAGTCATCGTCGTGGTG GTGGTGGTGGTGCATCGTCCCACTCTCCACTTG-3'	
	Zhuo_0052	5'-CAAGTGGAGAGTGGGACGATGCACCACCACCACCA CCACGACGATGACTACCTCAGTTCAATATTCG-3'	
TK1903 (PolD2)	Zhuo_0053	5'-CTCGGTACCAAGAGGGCACTCAGAGAAGC-3'	Kpn I
	Zhuo_0054m	5'-GGGTCTAGACAAGAGGAGAATGTTAGTGGTGGTGG TGGTGGTGCGAGCCGAAGAACTCGTCGAGGCTTATGC CTTTCCGCTTC-3'	Xba I
	Zhuo_0054m1	5'-AGGCTTATGCCTTTCCGCTTC-3'	
	Zhuo_0054m2	5'-GGGTCTAGACAAGAGGAGAATGTTAGTGG-3'	
	Zhuo_0055	5'-GGGCTGCAGGTTTCCTTTTCGTGTGTTCG-3'	Pst I
	Zhuo_0056	5'-GGGCTGCAGTAACTTCCGCAAGAATGTAC-3'	Pst I
TK1281 (Fen1)	Zhuo_0062	5'-CTGCGCATGCAGGCTTCTGAAGGTCCCGAGAAC-3'	Sph I
	Zhuo_0063	5'-CTTCTTTTTCTCATCTTTTCCGGTGGTGAGAATGCATC ATCATCATCATCATGGAGTCCAGATAGGTGAGCTG-3'	
	Zhuo_0063m1	5'-ACCTATCTGGACTCCATGATGATGATGATGATGCAT TCTCACCACCGGA-3'	
	Zhuo_0063m2	5'-TCCGGTGGTGAGAATGCATCATCATCATCATCATGG AGTCCAGATAGGT-3'	
	Zhuo_0064	5'-AAAACTGCAGGGAGATACTGTACAACTGGCAGGCT GAGGCAAAGCCTAAATACTTCTTTTTCTCATCTTTTCC GG-3'	Pst I

	Zhuo_0064m	5'-AAAACTGCAGGGAGATACTGTACAACTGGC-3'	Pst I
	Zhuo_0065	5'-GGGTCTAGAAAAGTGTAGAGAGTCACCCA-3'	Xba I
	Zhuo_0066	5'-CCTTTGAATTCGGCAGGTCGAGGTATGGAAC-3'	EcoR I
TK1790 (Primease1)	Zhuo_0080:	5'-CTCGGCATGCTTCCAGATTGATATGTATCG-3'	Sph I
	Zhuo_0081:	5'-TCTGCTGCAGTCTGAACATCCAAAAGTTTTATATTG-3'	Pst I
	Zhuo_0082:	5'-ATGCTCTAGACACCATAGGGAATTTTATTGTCTTC ACAGTACATTATTGTGCACCTCCCAAAAAGGAAACC GTTTTATTGTTGATTATG-3'	Xba I
	Zhuo_0082m1:	5'-ATGCTCTAGACACCATAGG-3'	Xba I
	Zhuo_0083:	5'-CCGTTTTATTGTTGATTATGTCACCCATATTAGGTGGGA GTATGCATCATCATCATCATCTCGACCCCTTTGGAAA AAGAGCG-3'	
	Zhuo_0083m1:	5'-CGACCCCTTTGGAAAAAGAGCGGAGAGC-3'	
	Zhuo_0083m2:	5'- GCTCTCCGCTCTTTTTCCAAAGGGGTCG-3'	
	Zhuo_0084:	5'-TCTGGATCCCTCCGTTCAATGCTCTCCAC-3'	BamH I
TK1792 (Primease Related)	Zhuo_0089:	5'-GTGCGCATGCCCAATCTGCCTCGAAGATGC-3'	Sph I
	Zhuo_0090:	5'-ATGCCTGCAGGGCCCCCTCGCCTCAATGATGATGA TGATGATGTCTTCTCAGAGCATCCA-3'	Pst I
	Zhuo_0092:	5'-GTTTGGATCCGAGGACGTTCGTTGTAGCAG-3'	BamH I
	Zhuo_0301	5'-ATGCTCTAGATGAGGCGAGGGGGCCGTAGCCCG CCTTC-3'	Xba I
TK0096 (MCM1)	EJ14m	5'-GTCAAAATTGATGTCCTCACTGTGGTGGTGGTGGT GGTGCATCCTTCATCCCTCTATTTGCGCC-3'	
	EJ15m	5'-GGCGCAAATAGAGGGATGAAGGATGCACCACCAC CACCACCACAGTGAGGACATCAATTTTGAC-3'	
	Zhuo_0093:	5'-GCTGCATGCGAGTTAGTGGACAAATCAC-3'	Sph I
	Zhuo_0094:	5'-ATGCCTGCAGCATCATTGTGAAGAAGTGATCAC-3'	Pst I
	Zhuo_0095:	5'-ATGCTCTAGATATAAGTAAGCGCTGTTCTGTAC-3'	Xba I
	Zhuo_0096:	5'-CCCGAATTCTTCACGCCATTGCTCTTCTC-3'	EcoR I
TK1410 (DnaG like)	Zhuo_0098:	5'-GTTGCATGCCTCCGCATTGAGGACGTCAG-3'	Sph I
	Zhuo_0099:	5'-GAAAATTAGAAAAGGAATCACTCATGATGATGATG ATGATGAGCGAAAGTGATGATCTTG-3'	
	Zhuo_0100:	5'-CAAGATCATCACTTTCGCTCATCATCATCATCATCA TGAGTGATTCCTTTTCTAATTTTC-3'	
	Zhuo_0101:	5'-GTACCTGCAGAAAGTTTAGTCCTAGTGTTC-3'	Pst I
	Zhuo_0102:	5'-ATGCTCTAGAATAACGTCCCCACCTCTTCT-3'	Xba I
	Zhuo_0103:	5'-CTCTGGATCCATGCTCCTGTCCGCGGAAGAT-3'	BamH I
TK2140 (Ligase)	Zhuo_0114	5'-ACTCGCATGCAGACCCTCGAGTGGGTTGTC-3'	Sph I
	Zhuo_0115	5'-TTTCTGCAGGTTGAAGTTCACTCCTCAACG-3'	Pst I
	Zhuo_0116	5'-CCCTCTAGACTTCTTTCCTTTTATATCCTG-3'	Xba I
	Zhuo_0117	5'-TCAGAGTAGCGCATATCGCTATGATGATGATGATG ATGCATTCTCACCACCGGATAA-3'	
	Zhuo_0118	5'-TTATCCGGTGGTGAGAATGCATCATCATCATCATCA TAGCGATATGCGCTACTCTGA-3'	

	Zhuo_0119	5'-GCGAGGTACCGCTTTATCTTCTCGCTCTCC-3'	Kpn I
TK1620 (MCM3)	Zhuo_0148	5'-TACCGGCATGCGGAAATGGCGAAGATGTACT-3'	Sph I
	Zhuo_0149	5'-ATCTTCTGCAGGCCATTTTTGATCATTCTG-3'	Pst I
	Zhuo_0150	5'-ATCTTCTCTAGAAGCCAATCTCGGTTCATTCAG-3'	Xba I
	Zhuo_0151	5'-ATCATCTCTTCCCTGTCATGATGATGATGATGATGCATCAGGCATCACCGA-3'	
	Zhuo_0152	5'-TCGGTGATGCCTGATGCATCATCATCATCATCATGACAGGGAAGAGATGAT-3'	
	Zhuo_0153	5'-TAGCCGAATTCGATGAACTCTTCCCGCTCGTGTA-3'	EcoR I
TK0582 (PCNA2)	Zhuo_0031	5'-CGCGCTGCAGGTTTGTTTTCGTGAAGATAGACTAC-3'	Pst I
	Zhuo_0033m:	5'-GCTTAGGTACCATACTTTGTCGTGAGAGGTA-3'	Kpn I
	Zhuo_0154	5'-CATTCAGCATGCCTCTCAGTC-3'	Sph I
	Zhuo_0155	5'-GAGGAGGTGATGGCATGCATCATCATCATCATCATACTTTTGAGATTGTGT-3'	
	Zhuo_0156	5'-ACACAATCTCAAAAGTATGATGATGATGATGATGCATGCCATCACCTCCTC-3'	
	Zhuo_0242	5'-CCGTTCTAGATGCTGGCCGGCACTACATCAG-3'	Xba I
TK1361 (MCM2)	Zhuo_0158	5'-CTCCAGAATTCCCGCTCATAGACCGGAAGA-3'	EcoR I
	Zhuo_0159	5'-GAGAGAGAGAGGGTGAGATGCATCATCATCATCATCATTTAACAAAAGTTACAGACG-3'	
	Zhuo_0160	5'-CGTCTGTAACTTTTGTTAAATGATGATGATGATGATGCATCTCACCCTCTCTCTCTC-3'	
	Zhuo_0161	5'-ATGCTCTAGACTTCGGGATCCTGCTCAGG-3'	Xba I
	Zhuo_0162	5'-ATGCCTGCAGGAGATCCCTATCATGCTCC-3'	Pst I
	Zhuo_0163	5'-CTTTTGCATGCAGAGTATCTCCGCTCGCAG-3'	Sph I

Underlined sequences indicate the restriction sites.

Appendix 4. Oligonucleotides and substrates used for biochemistry study in Chapter 3

Name	Sequences	Observation
A1	5'-GGGGCGAGTCCAGGTCAGGACCTTGCGGGG-3'	Cy5 labeled at 5' end
A2	5'-CCCCGCAAGGTCCTGACCTGGACTCGCCCC-3'	
A3	5'-AATAATCAATACTATAAGTACATATCAATACCCCGCAAGGTC CTGACCTGGACTCGCCCC-3'	
A4	5'-GGGGCGAGTCCAGGTCAGGACCTTGCGGGG-3'	Cy3 labeled at 3' end
A4R	5'-GGGGCGAGTCCAGGTCAGGACCTTGCGGGG-3'	RNA, Cy3 labeled at 3' end
A5	5'-CCCCGCAAGGTCCTGACCTGGACTCGCCCCA ATAATCAATACTATAAGTACATATCAATA-3'	
F1	A4+A3	5' overhang
F2	A4+A5	3' overhang
F3	A4+A2	Blunt
A	5'-GCGGGCAACAGCAACCGGAGCAGCAGACGAAGGAAACCAAGG AGGCGAAACCAGGGCCCAAGCAGCAGGAAAGGAGCACCAGGAGC AAAAGCAGGCACAG-3'	
B	5'-CAACAGGGAACGAGCACGGGCGCGGCAGCCACACGAGCCAGG AGCGACGACCGCCAGCGCGCAAACGAACGCCGAGCGCCGAAGGC ACGGAAAGCCAGCA-3'	
Bridge AB	5'-GTTCCCTGTTGCTGTGCCTGC-3'	
Bridge BA	5'-CTGTTGCCCGCTGCTGGCTTTC-3'	

Appendix 5: Oligonucleotides used for cloning in Chapter 3

Name	Sequences	Res. Sites
Zhuo_0232	5'-ATATCATATGGATAAGGAGGCTTTTTTGGAGCG-3'	NdeI
Zhuo_0233	5'-GCATGTCGACTCAACCCTCGCCTTCACTTCCAC-3'	SalI
Zhuo_0314	5'-CATCTCCCACAGGGCTGCTGACGGCATCA-3'	
Zhuo_0315	5'-TGATGCCGTCAGCAGCCCTGTGGGAGATG-3'	
Zhuo_0323	5'-TGAGGAAGCTTGTTGTGTAAACACTGACCTTCG-3'	HindIII
Zhuo_0324	5'-ATGCCGCATGCCGGAATGAGTTTAAAAAGG-3'	SphI
Zhuo_0325	5'-GATCATCTAGACAACGCCGATGCGAAGGGGCGGCT-3'	XbaI
Zhuo_0326	5'-AAGCCTCCTTATCATGATGATGATGATGATGCACGCGGCATCAC-3'	
Zhuo_0327	5'-GTGATGCCGCGTGCATCATCATCATCATCATGATAAGGAGGCTT-3'	
Zhuo_0328	5'-GCATGCGGATCCTCTCGTTGTCCACCTGGATTG-3'	BamHI
Miao_029	5'-CCGCATATGGATATAGTGAAGCTCAGGGAAC-3'	NdeI
Miao_030	5'-CCGGGATCCTTAGTGATGGTGATGGTGATGGAGGAATATCCTTACTCTTCGTGC-3'	BamHI
Miao_031	5'-CCGCATATGTTCACGGGTAAAGCCCTC-3'	NdeI
Miao_032	5'-CCGGGATCCTCAGTGATGGTGATGGTGATGGGCATCACCGAGCCACTCGTTTC-3'	BamHI

Underlined sequences indicate the restriction sites.

Appendix 6: Oligonucleotides used for cloning in Chapter 4

Name	Used to generate	Sequences
Zhuo_0494	PCNA1-A	5'-GACCGCCAGTGCCGCCGGCATAGCCGCCCTCAGCGACA TGG-3'
Zhuo_0495	PCNA1-A	5'-CCATGTCGCTGAGGGCGGCTATGCCGGCGGCACTGGCGG TC-3'
Zhuo_0496	PCNA1-E	5'-GTCGAGGAAGAGACCGAAAGTGCCGAAGGCATAGAAGAG CTCAGCGACATGGTC-3'
Zhuo_0497	PCNA1-E	5'-GACCATGTCGCTGAGCTCTTCTATGCCTTCGGCACTTTCGG TCTCTTCCTCGAC-3'
Zhuo_0498	PCNA2-A	5'-AGATGACAAAGGCGGCCAGTGCGGCCGGGGTTGCAGCC CTC-3'
Zhuo_0499	PCNA2-A	5'-GAGGGCTGCAACCCCGGCCGCACTGGCCGCCTTTGTCA TCT-3'
Zhuo_0500	PCNA2-E	5'-AGATGACAAAGGCGGAAAGTGCGGAAGGGGTTGCAGAA CTCGAGGACATC-3'
Zhuo_0501	PCNA2-E	5'-GATGTCCTCGAGTTCTGCAACCCCTTCCGCACTTTCCGCCT TTGTCATCT-3'
Zhuo_0395	Fen1	5'-ATGCCATATGGGAGTCCAGATAGGTGAGC-3'
Zhuo_0396	Fen1	5'-AATTGTCGACCTATCGACCGAACCAGCTCTC-3'
Zhuo_0277	TIP	5'-ATGCCATATGGACAGGAAGCTCGACGAG-3'
Zhuo_0278	TIP	5'-GTACGTCGACTTATAACTCCTCGATTTTCGCG-3'

Underlined sequences indicate the restriction sites.

Appendix 7: Cloned plasmids for expressing proteins identified in the interaction network.

Plasmid name	Gene	Current Statues
pZF003	TK0643	Soluble protein expressed
pZF005	TK1005	Soluble protein expressed
pZF008	TK1252	Studied in Chapter 3
pZF012	TK0569	Soluble protein expressed
pZF013	TK0808	Studied in Chapter 4
pZF016	TK2106	Soluble protein expressed
pZF019	TK1245	Soluble protein expressed
pZF021	TK1633	Soluble protein expressed
pZF022	TK1634	Soluble protein expressed
pZF023	TK0063	Soluble protein expressed
pZF024	TK0593	Soluble protein expressed
pZF026	TK1789	Soluble protein expressed
pZF027	TK2021	Soluble protein expressed
pZF028	TK2101	Soluble protein expressed
pZF029	TK0790	Soluble protein expressed
pZF030	TK0953	Soluble protein expressed
pZF031	TK1281	Studied in Chapter 4
pZF032	TK0679	Soluble protein expressed
pZF033	TK0801	Soluble protein expressed
pZF034	TK2250	Soluble protein expressed
pZF035	TK2140	Soluble protein expressed

Bibliography

1. Atomi, H., Fukui, T., Kanai, T., Morikawa, M. and Imanaka, T. (2004) Description of *Thermococcus kodakaraensis* sp. nov., a well studied hyperthermophilic archaeon previously reported as *Pyrococcus* sp. KOD1. *Archaea*, **1**, 263-267.

2. Fukui, T., Atomi, H., Kanai, T., Matsumi, R., Fujiwara, S. and Imanaka, T. (2005) Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1 and comparison with *Pyrococcus* genomes. *Genome Res*, **15**, 352-363.

3. Sato, T., Fukui, T., Atomi, H. and Imanaka, T. (2003) Targeted gene disruption by homologous recombination in the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. *J Bacteriol*, **185**, 210-220.

4. Robinson, N.P. and Bell, S.D. (2005) Origins of DNA replication in the three domains of life. *FEBS J*, **272**, 3757-3766.

5. Kelman, L.M. and Kelman, Z. (2004) Multiple origins of replication in archaea. *Trends Microbiol*, **12**, 399-401.

6. Kelman, Z. (2000) The replication origin of archaea is finally revealed. *Trends Biochem Sci*, **25**, 521-523.

7. Gaudier, M., Schuwirth, B.S., Westcott, S.L. and Wigley, D.B. (2007) Structural basis of DNA replication origin recognition by an ORC protein. *Science*, **317**, 1213-1216.

8. Dueber, E.L., Corn, J.E., Bell, S.D. and Berger, J.M. (2007) Replication origin recognition and deformation by a heterodimeric archaeal Orc1 complex. *Science*, **317**, 1210-1213.

9. Norais, C., Hawkins, M., Hartman, A.L., Eisen, J.A., Myllykallio, H. and Allers, T. (2007) Genetic and physical mapping of DNA replication origins in *Haloferax volcanii*. *PLoS Genet*, **3**, e77.

10. Capaldi, S.A. and Berger, J.M. (2004) Biochemical characterization of Cdc6/Orc1 binding to the replication origin of the euryarchaeon *Methanothermobacter thermoautotrophicus*. *Nucleic Acids Res*, **32**, 4821-4832.

11. Robinson, N.P., Dionne, I., Lundgren, M., Marsh, V.L., Bernander, R. and Bell, S.D. (2004) Identification of two origins of replication in the single chromosome of the archaeon *Sulfolobus solfataricus*. *Cell*, **116**, 25-38.

12. Myllykallio, H., Lopez, P., Lopez-Garcia, P., Heilig, R., Saurin, W., Zivanovic, Y., Philippe, H. and Forterre, P. (2000) Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon. *Science*, **288**, 2212-2215.

13. Robinson, N.P. and Bell, S.D. (2007) Extrachromosomal element capture and the evolution of multiple replication origins in archaeal chromosomes. *Proc Natl Acad Sci U S A*, **104**, 5806-5811.

14. Lundgren, M., Andersson, A., Chen, L., Nilsson, P. and Bernander, R. (2004) Three replication origins in *Sulfolobus* species: synchronous initiation of chromosome replication and asynchronous termination. *Proc Natl Acad Sci U S A*, **101**, 7046-7051.

15. Sakakibara, N., Kelman, L.M. and Kelman, Z. (2009) Unwinding the structure and function of the archaeal MCM helicase. *Mol Microbiol*, **72**, 286-296.

16. Sakakibara, N., Kelman, L.M. and Kelman, Z. (2009) How is the archaeal MCM helicase assembled at the origin? Possible mechanisms. *Biochem Soc Trans*, **37**, 7-11.

17. Li, Z., Kelman, L.M. and Kelman, Z. (2013) *Thermococcus kodakarensis* DNA replication. *Biochem Soc Trans*, **41**, 332-338.

18. Kaguni, J.M. (2006) DnaA: controlling the initiation of bacterial DNA replication and more. *Annu Rev Microbiol*, **60**, 351-375.

19. Leonard, A.C. and Grimwade, J.E. (2011) Regulation of DnaA assembly and activity: taking directions from the genome. *Annu Rev Microbiol*, **65**, 19-35.

20. Kelman, Z. and O'Donnell, M. (1995) Structural and functional similarities of prokaryotic and eukaryotic DNA polymerase sliding clamps. *Nucleic Acids Res*, **23**, 3613-3620.

21. Vivona, J.B. and Kelman, Z. (2003) The diverse spectrum of sliding clamp interacting proteins. *FEBS Lett*, **546**, 167-172.

22. Swiatek, A. and Macneill, S.A. (2010) The archaeo-eukaryotic GINS proteins and the archaeal primase catalytic subunit PriS share a common domain. *Biol Direct*, **5**, 17.

23. MacNeill, S.A. (2010) Structure and function of the GINS complex, a key component of the eukaryotic replisome. *Biochem J*, **425**, 489-500.

24. Berthon, J., Cortez, D. and Forterre, P. (2008) Genomic context analysis in Archaea suggests previously unrecognized links between DNA replication and translation. *Genome Biol*, **9**, R71.

25. Oyama, T., Ishino, S., Fujino, S., Ogino, H., Shirai, T., Mayanagi, K., Saito, M., Nagasawa, N., Ishino, Y. and Morikawa, K. (2011) Architectures of archaeal GINS complexes, essential DNA replication initiation factors. *BMC Biol*, **9**, 28.

26. Langston, L.D., Indiani, C. and O'Donnell, M. (2009) Whither the replisome: emerging perspectives on the dynamic nature of the DNA replication machinery. *Cell Cycle*, **8**, 2686-2691.

27. Pomerantz, R.T. and O'Donnell, M. (2007) Replisome mechanics: insights into a twin DNA polymerase machine. *Trends Microbiol*, **15**, 156-164.

28. Li, Z., Pan, M., Santangelo, T.J., Chemnitz, W., Yuan, W., Edwards, J.L., Hurwitz, J., Reeve, J.N. and Kelman, Z. (2011) A novel DNA nuclease is stimulated by association with the GINS complex. *Nucleic Acids Res*, **39**, 6114-6123.

29. Li, Z., Santangelo, T.J., Cubonova, L., Reeve, J.N. and Kelman, Z. (2010) Affinity purification of an archaeal DNA replication protein network. *MBio*, **1**, e00221-00210.

30. Yoshimochi, T., Fujikane, R., Kawanami, M., Matsunaga, F. and Ishino, Y. (2008) The GINS complex from *Pyrococcus furiosus* stimulates the MCM helicase activity. *J Biol Chem*, **283**, 1601-1609.

31. Marinsek, N., Barry, E.R., Makarova, K.S., Dionne, I., Koonin, E.V. and Bell, S.D. (2006) GINS, a central nexus in the archaeal DNA replication fork. *EMBO Rep*, **7**, 539-545.

32. Brewster, A.S. and Chen, X.S. (2010) Insights into the MCM functional mechanism: lessons learned from the archaeal MCM complex. *Crit Rev Biochem Mol Biol*, **45**, 243-256.

33. Costa, A. and Onesti, S. (2009) Structural biology of MCM helicases. *Crit Rev Biochem Mol Biol*, **44**, 326-342.

34. Pan, M., Santangelo, T.J., Li, Z., Reeve, J.N. and Kelman, Z. (2011) *Thermococcus kodakarensis* encodes three MCM homologs but only one is essential. *Nucleic Acids Res*, **39**, 9671-9680.

35. Krupovic, M., Gribaldo, S., Bamford, D.H. and Forterre, P. (2010) The evolutionary history of archaeal MCM helicases: a case study of vertical evolution combined with hitchhiking of mobile genetic elements. *Mol Biol Evol*, **27**, 2716-2732.

36. Mendelman, L.V., Notarnicola, S.M. and Richardson, C.C. (1992) Roles of bacteriophage T7 gene 4 proteins in providing primase and helicase functions *in vivo*. *Proc Natl Acad Sci U S A*, **89**, 10638-10642.

37. Fanning, E. and Knippers, R. (1992) Structure and function of simian virus 40 large tumor antigen. *Annu Rev Biochem*, **61**, 55-85.

38. Lohman, T.M. and Ferrari, M.E. (1994) *Escherichia coli* single-stranded DNA-binding protein: multiple DNA-binding modes and cooperativities. *Annu Rev Biochem*, **63**, 527-570.

39. Wold, M.S. (1997) Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annu Rev Biochem*, **66**, 61-92.

40. 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. *Proc Natl Acad Sci U S A*, **103**, 10236-10241.

41. Makarova, K.S., Koonin, E.V. and Kelman, Z. (2012) The CMG (CDC45/RecJ, MCM, GINS) complex is a conserved component of the DNA replication system in all archaea and eukaryotes. *Biol Direct*, **7**, 7.

42. Frick, D.N. and Richardson, C.C. (2001) DNA primases. *Annu Rev Biochem*, **70**, 39-80.

43. Evguenieva-Hackenberg, E., Walter, P., Hochleitner, E., Lottspeich, F. and Klug, G. (2003) An exosome-like complex in *Sulfolobus solfataricus*. *EMBO Rep*, **4**, 889-893.

44. Roppelt, V., Hobel, C.F., Albers, S.V., Lassek, C., Schwarz, H., Klug, G. and Evguenieva-Hackenberg, E. (2010) The archaeal exosome localizes to the membrane. *FEBS Lett*, **584**, 2791-2795.

45. Lao-Sirieix, S.H. and Bell, S.D. (2004) The heterodimeric primase of the hyperthermophilic archaeon *Sulfolobus solfataricus* possesses DNA and RNA primase, polymerase and 3'-terminal nucleotidyl transferase activities. *J Mol Biol*, **344**, 1251-1263.

46. Berquist, B.R., DasSarma, P. and DasSarma, S. (2007) Essential and non-essential DNA replication genes in the model halophilic Archaeon, *Halobacterium* sp. NRC-1. *BMC Genet*, **8**, 31.

47. Lao-Sirieix, S.H., Pellegrini, L. and Bell, S.D. (2005) The promiscuous primase. *Trends Genet*, **21**, 568-572.

48. Chemnitz Galal, W., Pan, M., Kelman, Z. and Hurwitz, J. (2012) Characterization of DNA primase complex isolated from the archaeon, *Thermococcus kodakaraensis*. *J Biol Chem*, **287**, 16209-16219.

49. Matsunaga, F., Norais, C., Forterre, P. and Myllykallio, H. (2003) Identification of short 'eukaryotic' Okazaki fragments synthesized from a prokaryotic replication origin. *EMBO Rep*, **4**, 154-158.

50. Chemnitz Galal, W., Pan, M., Giulian, G., Yuan, W., Li, S., Edwards, J.L., Marino, J.P., Kelman, Z. and Hurwitz, J. (2012) Formation of dAMP-glycerol and dAMP-Tris derivatives by *Thermococcus kodakaraensis* DNA primase. *J Biol Chem*, **287**, 16220-16229.

51. Pan, M., Kelman, L.M. and Kelman, Z. (2011) The archaeal PCNA proteins. *Biochem Soc Trans*, **39**, 20-24.

52. 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.

53. Jeruzalmi, D., O'Donnell, M. and Kuriyan, J. (2002) Clamp loaders and sliding clamps. *Curr Opin Struct Biol*, **12**, 217-224.

54. Kelch, B.A., Makino, D.L., O'Donnell, M. and Kuriyan, J. (2012) Clamp loader ATPases and the evolution of DNA replication machinery. *BMC Biol*, **10**, 34.

55. Ladner, J.E., Pan, M., Hurwitz, J. and Kelman, Z. (2011) Crystal structures of two active proliferating cell nuclear antigens (PCNAs) encoded by *Thermococcus kodakaraensis*. *Proc Natl Acad Sci U S A*, **108**, 2711-2716.

56. Pan, M., Santangelo, T.J., Cubonova, L., Li, Z., Metangmo, H., Ladner, J., Hurwitz, J., Reeve, J.N. and Kelman, Z. (2013) *Thermococcus kodakarensis* has two functional PCNA homologs but only one is required for viability. *Extremophiles*, **17**, 453-461.

57. Tahirov, T.H., Makarova, K.S., Rogozin, I.B., Pavlov, Y.I. and Koonin, E.V. (2009) Evolution of DNA polymerases: an inactivated polymerase-exonuclease module in Pol epsilon and a chimeric origin of eukaryotic polymerases from two classes of archaeal ancestors. *Biol Direct*, **4**, 11.

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

59. Burgers, P.M. (2009) Polymerase dynamics at the eukaryotic DNA replication fork. *J Biol Chem*, **284**, 4041-4045.

60. Beattie, T.R. and Bell, S.D. (2012) Coordination of multiple enzyme activities by a single PCNA in archaeal Okazaki fragment maturation. *EMBO J*, **31**, 1556-1567.

61. Nakatani, M., Ezaki, S., Atomi, H. and Imanaka, T. (2000) A DNA ligase from a hyperthermophilic archaeon with unique cofactor specificity. *J Bacteriol*, **182**, 6424-6433.

62. Kornberg, A. and Baker, T.A. (1992) *DNA replication*. 2nd ed. W.H. Freeman, New York.

63. Grabowski, B. and Kelman, Z. (2003) Archeal DNA replication: eukaryal proteins in a bacterial context. *Annu Rev Microbiol*, **57**, 487-516.

64. Chia, N., Cann, I. and Olsen, G.J. (2010) Evolution of DNA replication protein complexes in eukaryotes and Archaea. *PLoS One*, **5**, e10866.

65. Kasiviswanathan, R., Shin, J.H., Melamud, E. and Kelman, Z. (2004) Biochemical characterization of the *Methanothermobacter thermautotrophicus* minichromosome maintenance (MCM) helicase N-terminal domains. *J Biol Chem*, **279**, 28358-28366.

66. Santangelo, T.J., Cubonova, L., James, C.L. and Reeve, J.N. (2007) TFB1 or TFB2 is sufficient for *Thermococcus kodakaraensis* viability and for basal transcription *in vitro*. *J Mol Biol*, **367**, 344-357.

67. Barry, E.R. and Bell, S.D. (2006) DNA replication in the archaea. *Microbiol Mol Biol Rev*, **70**, 876-887.

68. Duggin, I.G. and Bell, S.D. (2006) The chromosome replication machinery of the archaeon *Sulfolobus solfataricus*. *J Biol Chem*, **281**, 15029-15032.

69. Kuba, Y., Ishino, S., Yamagami, T., Tokuhara, M., Kanai, T., Fujikane, R., Daiyasu, H., Atomi, H. and Ishino, Y. (2012) Comparative analyses of the two proliferating cell nuclear antigens from the hyperthermophilic archaeon, *Thermococcus kodakarensis*. *Genes Cells*, **17**, 923-937.

70. Komori, K. and Ishino, Y. (2001) Replication protein A in *Pyrococcus furiosus* is involved in homologous DNA recombination. *J Biol Chem*, **276**, 25654-25660.

71. Lopez de Saro, F.J. (2009) Regulation of interactions with sliding clamps during DNA replication and repair. *Curr Genomics*, **10**, 206-215.

72. Dionne, I., Nookala, R.K., Jackson, S.P., Doherty, A.J. and Bell, S.D. (2003) A heterotrimeric PCNA in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Mol Cell*, **11**, 275-282.

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

74. Warbrick, E., Heatherington, W., Lane, D.P. and Glover, D.M. (1998) PCNA binding proteins in *Drosophila melanogaster* : the analysis of a conserved PCNA binding domain. *Nucleic Acids Res*, **26**, 3925-3932.

75. Labib, K. and Gambus, A. (2007) A key role for the GINS complex at DNA replication forks. *Trends Cell Biol*, **17**, 271-278.

76. Courcelle, J., Donaldson, J.R., Chow, K.H. and Courcelle, C.T. (2003) DNA damage-induced replication fork regression and processing in *Escherichia coli*. *Science*, **299**, 1064-1067.

77. Borde, V. and Cobb, J. (2009) Double functions for the Mre11 complex during DNA double-strand break repair and replication. *Int J Biochem Cell Biol*, **41**, 1249-1253.

78. Mirzoeva, O.K. and Petrini, J.H. (2003) DNA replication-dependent nuclear dynamics of the Mre11 complex. *Mol Cancer Res*, **1**, 207-218.

79. Robison, J.G., Elliott, J., Dixon, K. and Oakley, G.G. (2004) Replication protein A and the Mre11.Rad50.Nbs1 complex co-localize and interact at sites of stalled replication forks. *J Biol Chem*, **279**, 34802-34810.

80. Bochman, M.L. and Schwacha, A. (2009) The Mcm complex: unwinding the mechanism of a replicative helicase. *Microbiol Mol Biol Rev*, **73**, 652-683.

81. Zuo, Z., Rodgers, C.J., Mikheikin, A.L. and Trakselis, M.A. (2010) Characterization of a functional DnaG-type primase in archaea: implications for a dual-primase system. *J Mol Biol*, **397**, 664-676.

82. Kelman, L.M. and Kelman, Z. (2003) Archaea: an archetype for replication initiation studies? *Mol Microbiol*, **48**, 605-615.

83. Leonard, A.C. and Grimwade, J.E. (2009) Initiating chromosome replication in *E. coli*: it makes sense to recycle. *Genes Dev*, **23**, 1145-1150.

84. Katayama, T. and Sekimizu, K. (1999) Inactivation of *Escherichia coli* DnaA protein by DNA polymerase III and negative regulations for initiation of chromosomal replication. *Biochimie*, **81**, 835-840.

85. Ulrich, H.D. (2009) Regulating post-translational modifications of the eukaryotic replication clamp PCNA. *DNA Repair (Amst)*, **8**, 461-469.

86. Forsburg, S.L. (2004) Eukaryotic MCM proteins: beyond replication initiation. *Microbiol Mol Biol Rev*, **68**, 109-131.

87. Makarova, K.S. and Koonin, E.V. (2010) Archaeal ubiquitin-like proteins: functional versatility and putative ancestral involvement in tRNA modification revealed by comparative genomic analysis. *Archaea*, **2010**, 710303.

88. Bienkowska, J.R., Hartman, H. and Smith, T.F. (2003) A search method for homologs of small proteins. Ubiquitin-like proteins in prokaryotic cells? *Protein Eng*, **16**, 897-904.

89. Humbard, M.A., Miranda, H.V., Lim, J.M., Krause, D.J., Pritz, J.R., Zhou, G., Chen, S., Wells, L. and Maupin-Furlow, J.A. (2010) Ubiquitin-like small archaeal modifier proteins (SAMPs) in *Haloferax volcanii*. *Nature*, **463**, 54-60.

90. Kirkland, P.A. and Maupin-Furlow, J.A. (2009) Stabilization of an archaeal DNA-sliding clamp protein, PCNA, by proteasome-activating nucleotidase gene knockout in *Haloferax volcanii*. *FEMS Microbiol Lett*, **294**, 32-36.

91. Walters, A.D. and Chong, J.P. (2009) *Methanococcus maripaludis*: an archaeon with multiple functional MCM proteins? *Biochem Soc Trans*, **37**, 1-6.

92. Walters, A.D. and Chong, J.P. (2010) An archaeal order with multiple minichromosome maintenance genes. *Microbiology*, **156**, 1405-1414.

93. Ishino, S., Fujino, S., Tomita, H., Ogino, H., Takao, K., Daiyasu, H., Kanai, T., Atomi, H. and Ishino, Y. (2011) Biochemical and genetical analyses of the three mcm genes from the hyperthermophilic archaeon, *Thermococcus kodakarensis*. *Genes Cells*, **16**, 1176-1189.

94. Gambus, A., Jones, R.C., Sanchez-Diaz, A., Kanemaki, M., van Deursen, F., Edmondson, R.D. and Labib, K. (2006) GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. *Nat Cell Biol*, **8**, 358-366.

95. Kubota, Y., Takase, Y., Komori, Y., Hashimoto, Y., Arata, T., Kamimura, Y., Araki, H. and Takisawa, H. (2003) A novel ring-like complex of *Xenopus* proteins essential for the initiation of DNA replication. *Genes Dev*, **17**, 1141-1152.

96. Takayama, Y., Kamimura, Y., Okawa, M., Muramatsu, S., Sugino, A. and Araki, H. (2003) GINS, a novel multiprotein complex required for chromosomal DNA replication in budding yeast. *Genes Dev*, **17**, 1153-1165.

97. Yang, X., Gregan, J., Lindner, K., Young, H. and Kearsey, S.E. (2005) Nuclear distribution and chromatin association of DNA polymerase alpha-primase is affected by TEV protease cleavage of Cdc23 (Mcm10) in fission yeast. *BMC Mol Biol*, **6**, 13.

98. Chang, Y.P., Wang, G., Bermudez, V., Hurwitz, J. and Chen, X.S. (2007) Crystal structure of the GINS complex and functional insights into its role in DNA replication. *Proc Natl Acad Sci U S A*, **104**, 12685-12690.

99. Shin, J.H., Jiang, Y., Grabowski, B., Hurwitz, J. and Kelman, Z. (2003) Substrate requirements for duplex DNA translocation by the eukaryal and archaeal minichromosome maintenance helicases. *J Biol Chem*, **278**, 49053-49062.

100. Kasiviswanathan, R., Shin, J.H. and Kelman, Z. (2005) Interactions between the archaeal Cdc6 and MCM proteins modulate their biochemical properties. *Nucleic Acids Res*, **33**, 4940-4950.

101. Sutera, V.A., Jr., Han, E.S., Rajman, L.A. and Lovett, S.T. (1999) Mutational analysis of the RecJ exonuclease of *Escherichia coli*: identification of phosphoesterase motifs. *J Bacteriol*, **181**, 6098-6102.

102. Hubscher, U. and Seo, Y.S. (2001) Replication of the lagging strand: a concert of at least 23 polypeptides. *Mol Cells*, **12**, 149-157.

103. Bae, S.H. and Seo, Y.S. (2000) Characterization of the enzymatic properties of the yeast dna2 Helicase/endonuclease suggests a new model for Okazaki fragment processing. *J Biol Chem*, **275**, 38022-38031.

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

105. Reagan, M.S., Pittenger, C., Siede, W. and Friedberg, E.C. (1995) Characterization of a mutant strain of Saccharomyces cerevisiae with a deletion of the *RAD27* gene, a structural homolog of the *RAD2* nucleotide excision repair gene. *J Bacteriol*, **177**, 364-371.

106. Henneke, G., Flament, D., Hubscher, U., Querellou, J. and Raffin, J.P. (2005) The hyperthermophilic euryarchaeota *Pyrococcus abyssi* likely requires the two DNA polymerases D and B for DNA replication. *J Mol Biol*, **350**, 53-64.

107. Rouillon, C., Henneke, G., Flament, D., Querellou, J. and Raffin, J.P. (2007) DNA polymerase switching on homotrimeric PCNA at the replication fork of the euryarchaea *Pyrococcus abyssi*. *J Mol Biol*, **369**, 343-355.

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

109. Stoimenov, I. and Helleday, T. (2009) PCNA on the crossroad of cancer. *Biochem Soc Trans*, **37**, 605-613.

110. Kelman, Z. and Hurwitz, J. (1998) Protein-PCNA interactions: a DNA-scanning mechanism? *Trends Biochem Sci*, **23**, 236-238.

111. Winter, J.A. and Bunting, K.A. (2012) Rings in the extreme: PCNA interactions and adaptations in the archaea. *Archaea*, **2012**, 951010.

112. Yao, N.Y. and O'Donnell, M. (2012) The RFC Clamp Loader: Structure and Function. *Subcell Biochem*, **62**, 259-279.

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

114. Gilljam, K.M., Feyzi, E., Aas, P.A., Sousa, M.M., Muller, R., Vagbo, C.B., Catterall, T.C., Liabakk, N.B., Slupphaug, G., Drablos, F. *et al.* (2009) Identification of a novel, widespread, and functionally important PCNA-binding motif. *J Cell Biol*, **186**, 645-654.

115. 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. *Proc Natl Acad Sci U S A*, **98**, 11627-11632.

116. Kurz, M., Dalrymple, B., Wijffels, G. and Kongsuwan, K. (2004) Interaction of the sliding clamp beta-subunit and Hda, a DnaA-related protein. *J Bacteriol*, **186**, 3508-3515.

117. Shen, Y., Musti, K., Hiramoto, M., Kikuchi, H., Kawarabayashi, Y. and Matsui, I. (2001) Invariant Asp-1122 and Asp-1124 are essential residues for polymerization catalysis of family D DNA polymerase from *Pyrococcus horikoshii*. *J Biol Chem*, **276**, 27376-27383.

118. Kelman, Z. and Hurwitz, J. (2000) A unique organization of the protein subunits of the DNA polymerase clamp loader in the archaeon *Methanobacterium thermoautotrophicum* deltaH. *J Biol Chem*, **275**, 7327-7336.

119. Pascal, B.D., Willis, S., Lauer, J.L., Landgraf, R.R., West, G.M., Marciano, D., Novick, S., Goswami, D., Chalmers, M.J. and Griffin, P.R. (2012) HDX workbench: software for the analysis of H/D exchange MS data. *J Am Soc Mass Spectrom*, **23**, 1512-1521.

120. Hileman, T.H. and Santangelo, T.J. (2012) Genetics Techniques for *Thermococcus kodakarensis*. *Front Microbiol*, **3**, 195.

121. Sato, T., Fukui, T., Atomi, H. and Imanaka, T. (2005) Improved and versatile transformation system allowing multiple genetic manipulations of the hyperthermophilic archaeon *Thermococcus kodakaraensis*. *Appl Environ Microbiol*, **71**, 3889-3899.

122. Cubonovaa, L., Katano, M., Kanai, T., Atomi, H., Reeve, J.N. and Santangelo, T.J. (2012) An archaeal histone is required for transformation of *Thermococcus kodakarensis*. *J Bacteriol*, **194**, 6864-6874.

123. Balakrishnan, L. and Bambara, R.A. (2013) Okazaki fragment metabolism. *Cold Spring Harb Perspect Biol*, **5**, a010173.

124. Balakrishnan, L. and Bambara, R.A. (2013) Flap Endonuclease 1. *Annu Rev Biochem*, **82**, 119-138.

125. Mayne, L., Kan, Z.Y., Chetty, P.S., Ricciuti, A., Walters, B.T. and Englander, S.W. (2011) Many overlapping peptides for protein hydrogen exchange

experiments by the fragment separation-mass spectrometry method. *J Am Soc Mass Spectrom*, **22**, 1898-1905.

126. Rand, K.D., Zehl, M., Jensen, O.N. and Jorgensen, T.J. (2009) Protein hydrogen exchange measured at single-residue resolution by electron transfer dissociation mass spectrometry. *Anal Chem*, **81**, 5577-5584.

127. Huang, R.Y., Garai, K., Frieden, C. and Gross, M.L. (2011) Hydrogen/deuterium exchange and electron-transfer dissociation mass spectrometry determine the interface and dynamics of apolipoprotein E oligomerization. *Biochemistry*, **50**, 9273-9282.

128. Huang, R.Y., Wen, J., Blankenship, R.E. and Gross, M.L. (2012) Hydrogen-deuterium exchange mass spectrometry reveals the interaction of Fenna-Matthews-Olson protein and chlorosome CsmA protein. *Biochemistry*, **51**, 187-193.

129. Chalmers, M.J., Busby, S.A., Pascal, B.D., West, G.M. and Griffin, P.R. (2011) Differential hydrogen/deuterium exchange mass spectrometry analysis of protein-ligand interactions. *Expert Rev Proteomics*, **8**, 43-59.

130. Jackson, S.P. and Durocher, D. (2013) Regulation of DNA damage responses by ubiquitin and SUMO. *Mol Cell*, **49**, 795-807.

131. Maupin-Furlow, J.A. (2013) Ubiquitin-like proteins and their roles in archaea. *Trends Microbiol*, **21**, 31-38.

132. Kennelly, P.J. (2003) Archaeal protein kinases and protein phosphatases: insights from genomics and biochemistry. *Biochem J*, **370**, 373-389.

133. Punchihewa, C., Inoue, A., Hishiki, A., Fujikawa, Y., Connelly, M., Evison, B., Shao, Y., Heath, R., Kuraoka, I., Rodrigues, P. *et al.* (2012) Identification of small molecule proliferating cell nuclear antigen (PCNA) inhibitor that disrupts interactions with PIP-box proteins and inhibits DNA replication. *J Biol Chem*, **287**, 14289-14300.

134. Georgescu, R.E., Yurieva, O., Kim, S.S., Kuriyan, J., Kong, X.P. and O'Donnell, M. (2008) Structure of a small-molecule inhibitor of a DNA polymerase sliding clamp. *Proc Natl Acad Sci U S A*, **105**, 11116-11121.

135. Sarmiento, F., Mrazek, J. and Whitman, W.B. (2013) Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon *Methanococcus maripaludis*. *Proc Natl Acad Sci U S A*, **110**, 4726-4731.