Affinity Purification of an Archaeal DNA Replication Protein Network

Zhuo Li,[a] Thomas J. Santangelo,[b] L'ubomíra Čuboňová,[b] John N. Reeve,[b] and Zvi Kelman[a]

Institute for Bioscience and Biotechnology Research, Rockville, Maryland, USA, and Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, Maryland, USA,[a] and Department of Microbiology, Ohio State University, Columbus, Ohio, USA[b]

Z.L. and T.J.S. contributed equally to this article.

ABSTRACT Nineteen *Thermococcus kodakarensis* strains have been constructed, each of which synthesizes a different His_6-tagged protein known or predicted to be a component of the archaeal DNA replication machinery. Using the His_6-tagged proteins, stable complexes assembled *in vivo* have been isolated directly from clarified cell lysates and the *T. kodakarensis* proteins present have been identified by mass spectrometry. Based on the results obtained, a network of interactions among the archaeal replication proteins has been established that confirms previously documented and predicted interactions, provides experimental evidence for previously unrecognized interactions between proteins with known functions and with unknown functions, and establishes a firm experimental foundation for archaeal replication research. The proteins identified and their participation in archaeal DNA replication are discussed and related to their bacterial and eukaryotic counterparts.

IMPORTANCE DNA replication is a central and essential event in all cell cycles. Historically, the biological world was divided into prokaryotes and eukaryotes, based on the absence or presence of a nuclear membrane, and many components of the DNA replication machinery have been identified and characterized as conserved or nonconserved in prokaryotic versus eukaryotic organisms. However, it is now known that there are two evolutionarily distinct prokaryotic domains, *Bacteria* and *Archaea*, and to date, most prokaryotic replication research has investigated bacterial replication. Here, we have taken advantage of recently developed genetic techniques to isolate and identify many proteins likely to be components of the archaeal DNA replication machinery. The results confirm and extend predictions from genome sequencing that the archaeal replication system is less complex but more closely related to a eukaryotic than to a bacterial replication system.

Address correspondence to John N. Reeve, reeve.2@osu.edu.

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 (Fig. 1) (1). 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 reference 2). 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 *Escherichia coli* τ subunit that couples the leading and lagging strand DNA 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 polymerase III [PolIII]) are homodimers of ~40-kDa subunits, whereas eukaryotic and euryarchaeal processivity factors (proliferating cell nuclear antigen [PCNA]) are homotrimers of ~29-kDa subunits. Members of the order *Crenarchaeota* contain three different PCNA homologues that assemble into heterotrimers (3). 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 (4). Here we report the results of experiments that identify many of the proteins likely to participate in DNA replication in the euryarchaeon *Thermococcus 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.
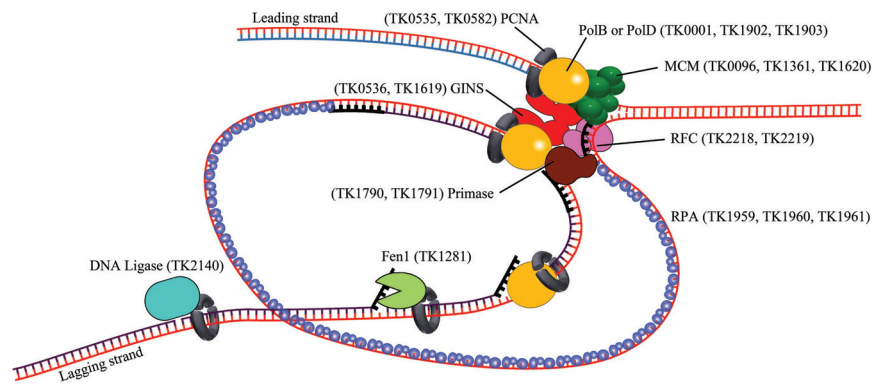
FIG 1 Components of the archaeal replisome. The *T. kodakarensis* numerical gene designations are listed adjacent to the protein subcomplexes predicted to constitute the replisome (5). It remains uncertain if the replicative polymerase(s) is PolB (TK0001p) and/or PolD (TK1902p and TK1903p).

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 (5). *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 coisolated, 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.

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 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 coisolation of a protein with a His$_6$-tagged protein, are documented in Table S1 in the supplemental material. 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 Table 2 and illustrated as a network in Fig. 2. Many of the interactions were confirmed by coisolation 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 coisolated, 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.

TABLE 1 *T. kodakarensis* proteins His$_6$ tagged and used to isolate replisome components

Protein	Gene[a]	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
GINS15[b]	TK0536	Subunit of a complex that binds primase, helicase, and polymerase
GINS23	TK1619	Subunit of a complex that binds primase, helicase, and polymerase
MCM1[b]	TK0096	Putative replicative helicase
MCM2	TK1361	Putative replicative helicase
MCM3	TK1620	Putative replicative helicase
PCNA1[b]	TK0535	Increases polymerase processivity
PCNA2	TK0582	Increases polymerase processivity
PolB	TK0001	DNA polymerase B
PolD-L	TK1903	Large subunit of DNA polymerase D
Pri-L	TK1790	Subunit of DNA primase
RFC-S	TK2218	Small subunit of the processivity complex, assembles PCNA on DNA
RFC-L	TK2219	Large subunit of the processivity complex, assembles PCNA on DNA
RPA2[b]	TK1960	Subunit of ssDNA-binding protein
RPA3	TK1961	Subunit of ssDNA-binding protein
TK1792p[c]	TK1792	Unknown function but encoded in an operon with DNA primase subunits (TK1790p + TK1791p)

[a] The designation of the *T. kodakarensis* gene encoding the protein (5).

[b] Based on sequence homologies, *T. kodakarensis* has two GINS and PCNA proteins and three MCM and RPA proteins.

[c] TK1792p, although annotated as a protein with unknown function, is encoded by an operon that also encodes the two subunits of primase (TK1790p and TK1791p) and therefore was tagged in this study.

TABLE 2 A subset of the protein-protein interactions identified in this study[a]

Protein, gene	MASCOT score	Protein molecular mass (Da)	No. of peptide matches	% Coverage	Annotated function (5)
Cdc6, TK1901					
TK1901	1,259	47,777	83	55	Cdc6
TK2218	336	37,170	13	13	RFC-S
TK0535	313	28,222	15	33	PCNA1
TK1314	233	50,095	9	26	ATPase
TK1903	186	150,190	10	2	PolD-L
TK0593	100	45,867	2	6	Unknown
DNA ligase, TK2140					
TK2140	3,453	64,042	774	82	DNA ligase
TK1903	291	150,190	13	4	PolD-L
TK0063	196	21,240	12	28	Nucleotidyltransferase/DNA-binding-domain-containing protein
TK0361	146	16,513	1	25	Unknown
TK0455	131	37,834	4	12	Unknown
TK2145	126	69,683	6	3	Unknown
TK0847	124	8,671	4	59	Unknown
TK0535	103	28,222	4	13	PCNA1
DnaG-like, TK1410					
TK0790	589	16,212	35	82	Unknown
TK1633	352	29,757	15	46	Exosome complex RNA-binding protein Rrp42
TK2227	235	54,037	11	12	RNA-binding protein FAU-1
TK1634	202	27,668	10	23	Exosome complex exonuclease Rrp41
Fen1, TK1281					
TK1281	2,608	38,786	581	82	Fen1
TK0535	1,967	28,222	361	89	PCNA1
TK0569	200	38,673	7	11	Unknown
TK1046	199	147,354	5	3	Unknown
TK0590	192	8,502	10	63	Unknown
TK0358	184	53,699	5	6	Unknown
TK0593	153	45,867	5	13	Unknown
GINS15, TK0536					
TK1903	895	150,190	41	15	PolD-L
TK1252	858	52,858	25	32	ssDNA-specific exonuclease
TK1046	659	147,354	22	10	Unknown
TK1902	488	80,848	14	15	PolD-S
TK0536	364	21,583	14	44	GINS15
TK1619	191	19,154	7	26	GINS23
GINS23, TK1619					
TK1619	1,565	19,154	141	74	GINS23
TK0582	1,119	28,429	53	82	PCNA2
TK1186	451	43,957	15	26	Unknown
TK0535	175	28,222	5	17	PCNA1
TK0569	115	38,673	2	9	Unknown
TK2021	107	32,017	4	6	ParA/MinD family ATPase
MCM1, TK0096					
TK0096	676	103,285	30	18	MCM1
TK1313	195	23,503	12	20	Unknown
TK0063	194	21,240	8	36	Nucleotidyltransferase/DNA-binding-domain-containing protein
TK1903	174	150,190	11	3	PolD-L
TK0535	137	28,222	5	14	PCNA1
TK2211	109	102,779	6	4	Rad50
TK0590	106	8,502	3	21	Unknown
MCM2, TK1361					
TK0535	925	28,222	46	66	PCNA1
TK1903	358	150,190	21	6	PolD-L
TK0063	230	21,240	9	32	Nucleotidyltransferase/DNA-binding-domain-containing protein
TK0590	152	8,502	11	51	Unknown
TK0682	150	65,448	11	12	MutS-like DNA mismatch repair ATPase
TK0001	100	90,030	17	2	PolB
MCM3, TK1620					
TK1245	167	15,009	8	23	Unknown
TK0590	164	8,502	7	53	Unknown
TK0569	146	38,673	5	7	Unknown
TK0467	124	118,109	4	3	Unknown
TK1046	100	147,354	6	2	Unknown

TABLE 2 Continued

Protein, gene	MASCOT score	Protein molecular mass (Da)	No. of peptide matches	% Coverage	Annotated function (5)
PCNA1, TK0535					
TK0535	2,098	28,222	828	91	PCNA1
TK2218	265	37,170	13	10	RFC-S
TK0808	259	7,606	15	48	Unknown
TK2219	105	57,239	5	6	RFC-L
PCNA2, TK0582					
TK0582	1,723	28,429	166	84	PCNA2
TK1046	754	147,354	28	11	Unknown
TK0569	173	38,673	6	15	Unknown
TK0535	132	28,222	5	18	PCNA1
TK0953	130	67,743	5	7	ATPase
TK1849	110	19,094	2	11	Unknown
PolB, TK0001					
TK0001	4,027	90,030	556	35	PolB
TK1046	576	147,354	25	10	Unknown
TK0569	210	38,673	9	15	Unknown
TK2021	109	32,017	4	8	ParA/MinD family ATPase
PolD-L, TK1903					
TK1903	427	150,190	21	8	PolD-L
TK1902	197	80,848	9	6	PolD-S
TK0535	123	28,222	4	12	PCNA1
TK1252	102	52,858	4	5	ssDNA-specific exonuclease
Pri-L, TK1790					
TK1791	442	40,443	26	28	Pri-S
TK1790	187	46,893	9	15	Pri-L
TK1186	185	43,957	7	24	Unknown
TK0569	154	38,673	5	12	Unknown
TK2211	132	102,779	10	5	Rad50
TK0535	131	28,222	5	10	PCNA1
TK1789	109	28,203	4	15	ATPase
RFC-S, TK2218					
TK2218	968	37,170	58	26	RFC-S
TK2219	388	57,239	24	15	RFC-L
RFC-L, TK2219					
TK2218	1,412	37,170	149	29	RFC-S
TK0582	885	28,429	45	60	PCNA2
TK2219	841	57,239	62	32	RFC-L
TK1245	113	15,009	5	15	Unknown
TK0569	131	38,673	2	9	Unknown
TK1046	105	147,354	3	1	Unknown
RPA2, TK1960					
TK1046	1,196	147,354	43	17	Unknown
TK0358	378	53,699	10	17	Unknown
TK2021	152	32,017	4	14	ParA/MinD family ATPase
TK1633	148	29,757	5	14	Exosome complex RNA-binding protein Rrp42
TK0946	133	15,865	4	19	Unknown
TK2253	121	28,743	9	20	Unknown
TK0719	109	39,062	4	8	ABC-type molybdate transport system, ATPase component
TK1579	107	33,947	1	5	ABC-type multidrug transport system, ATPase component
TK0593	105	45,867	3	12	Unknown
TK1960	103	14,356	5	13	Replication factor A complex, RPA14 subunit
TK2211	100	102,779	21	3	Rad50
RPA3, TK1961					
TK1959	1,472	31,240	267	65	RPA1
TK0470	932	141,310	67	14	Reverse gyrase
TK1960	734	14,356	105	86	RPA2
TK1017	133	135,911	10	3	Chromosome segregation ATPase
TK0001	117	90,030	17	3	PolB
TK0446	110	27,210	4	9	Unknown
Unknown, TK1792					
TK1046	279	147,354	12	5	Unknown
TK1633	133	29,757	3	17	Exosome complex RNA-binding protein Rrp42

[a] These interactions are shown in Fig. 3. Proteins with MASCOT scores of >100 identified in the eluate containing His$_6$-tagged proteins are listed along with each protein's molecular mass, the number of peptides matched, and the percentage of its amino acid sequence covered by the matching peptides. None of the proteins listed was detected in equivalent column fractions prepared from the untagged *T. kodakarensis* KW128. See text for further details.
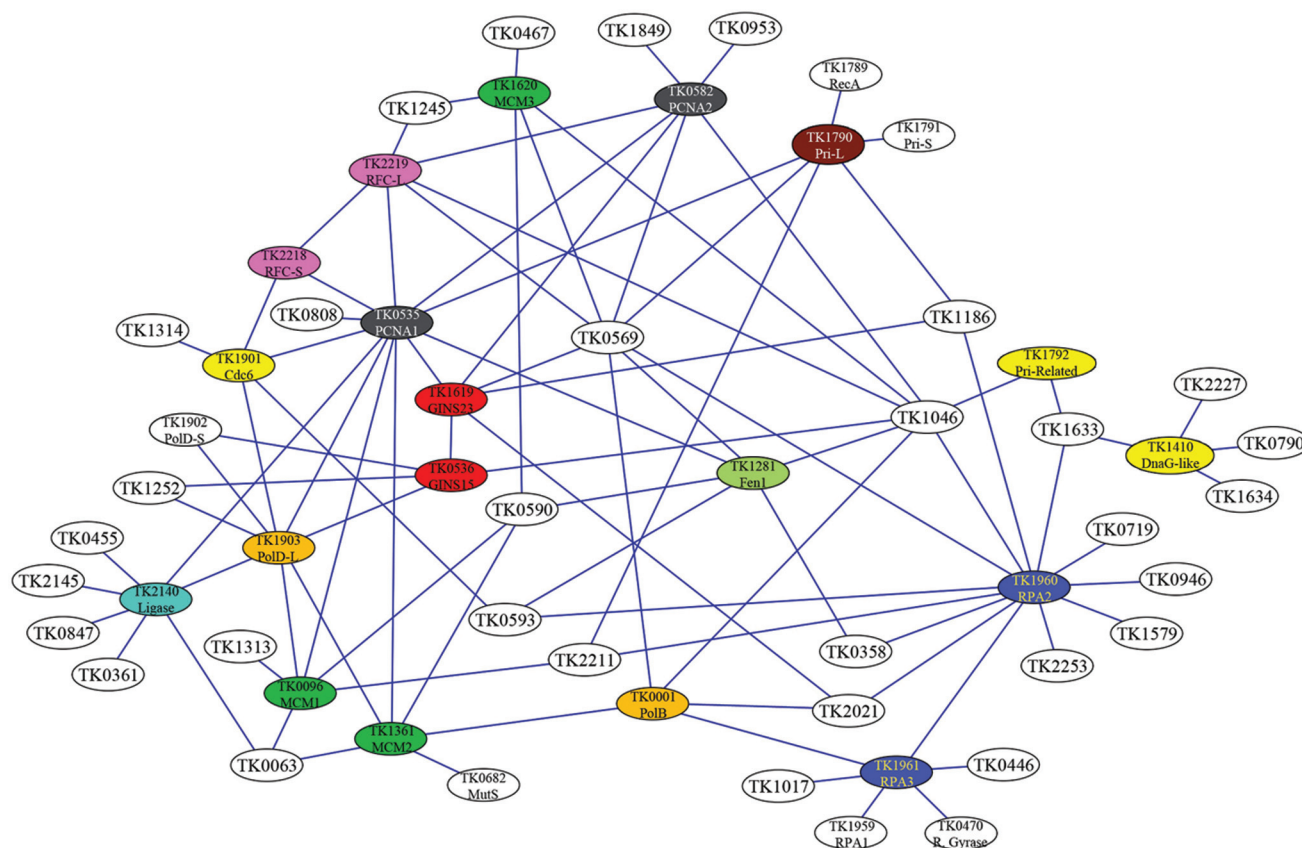
FIG 2 The interaction network documented for protein components of the *T. kodakarensis* replisome. Proteins that were His₆ tagged and used to isolate interacting proteins are identified in colored ovals (the colors used are as in Fig. 1). Proteins that were coisolated with a His₆-tagged protein are identified in white ovals by the designations of the *T. kodakarensis* genes that encode them.

Established archaeal replisome complexes. As some replisome complexes were already well documented, the coisolation of the proteins known to be components of these complexes validated and provided a measure of the sensitivity of the His₆ tag-dependent coisolation technology. Some examples of these validating interactions are described individually below, and all are listed in Table 2 and documented in Fig. 2.

(i) Polypeptide subunits of DNA polymerase D (PolD), primase, replication factor C (RFC), and the GINS complex. It is well established that the archaeal replisome components PolD, primase, RFC, and the GINS complex (from the Japanese go-ichi-ni-san, meaning 5-1-2-3, after the four related subunits of the eukaryotic complex, Sld5, Psf1, Psf2, and Psf3) are each formed by the assembly of two different polypeptides (2, 6, 7). Consistent with this, the two polypeptides annotated as the subunits of these replisome proteins in *T. kodakarensis* were coisolated with very high MASCOT scores (Table 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 (data not shown).

(ii) Replication protein A (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 (8). Consistent with this, *T. kodakarensis* RPA1 and RPA2 were coisolated by Ni²⁺ binding of His₆-tagged RPA3 (Table 2).

(iii) RFC-PCNA complex formation. RFC-PCNA binding has been reported in all of the replication systems investigated (9). *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 coisolated with His₆-tagged PCNA1, and PCNA2 was coisolated with His₆-tagged RFC-L (Table 2). These coisolation 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 (J. Hurwitz; personal communication).

(iv) PCNA-PolD-Fen1-ligase interactions. PCNA interactions with DNA polymerases increase their processivity (10). PCNA also binds and regulates the activity of a number of enzymes participating in Okazaki fragment maturation and postreplication processes (summarized in references 11 and 12). Consistent with these reports, PCNA1 was coisolated in complexes with PolD-L (large subunit of euryarchaeon-specific PolD), Fen1, and DNA ligase (Table 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 coisolated with His$_6$-tagged DNA ligase, adding support to the hypothesis that DNA ligase is associated with the archaeal replication fork (13).

Novel interactions of archaeal replication proteins. (i) PCNA interactions. Many proteins have been reported to interact with eukaryotic PCNA (11), 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 (14, 15). In addition to proteins expected to copurify with PCNA (see above), Cdc6 (TK1901p), MCM1 (TK0096p), and MCM2 (TK1361p) were copurified 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 coisolated 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 Pol α-primase complex (summarized in references 16 and 17). 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 (18) and forms a tetrameric complex that contains two GINS15 and two GINS23 subunits (18). Both subunits of PolD were coisolated using His$_6$-tagged GINS15, and PCNA1 and PCNA2 were both coisolated 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 coisolated with His$_6$-tagged GINS15 (Table 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, copurified with the GINS complex from *Sulfolobus solfataricus* (19). RecJ plays a role in stalled replication fork activation in *E. coli* (20), 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 21) 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 22). *T. kodakarensis* Rad50 (TK2211p) was coisolated in complexes using His$_6$-tagged MCM1, primase, and RPA2 (Table 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 (23), 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 (24, 25), but the coisolation of both Rad50 and MutS (TK0682p) with His$_6$-tagged MCM1 (Table 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 coisolated in complexes using His$_6$-tagged Fen1, GINS15, MCM3, PCNA2, PolB, RFC-L, RPA2, and TK1792p with very high MASCOT scores (Table 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: (i) 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) (26). 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 (27, 28). Taken together, the results argue that TK1410p participates in exosome activity rather than in DNA replication.

DISCUSSION

The interaction network (Fig. 2) and the interactions listed in Table 2 and in Table S1 in the supplemental material were documented using a systematic approach to isolate and identify all of the proteins that copurified 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 references 6 and 29). 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 references 30 and 31). 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 30 and 31). The coisolation 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 coisolated 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 Pol α-primase complex, which is responsible for primer synthesis on the lagging strand (reference 17 and references therein). To date, no archaeal homologue of Cdc45 has been identified but several proteins, and so potential candidates for Cdc45 functional homologues, copurified with His$_6$-tagged GINS15 or GINS23, including TK0569p, TK1046p, and TK1186p, which also copurified with His$_6$-tagged primase (Fig. 2; Table 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 (19, 32) but not with DNA polymerase. The results now reported (Fig. 2; Table 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 references 33 and 34). A small protein similar in size to ubiquitin (~8 kDa; TK0808p) was consistently coisolated with His$_6$-tagged PCNA1, and a second similarly sized protein (~8.5 kDa; TK0590p) was coisolated in complexes using His$_6$-tagged Fen1, MCM1, MCM2, and MCM3 (Table 2). Currently, very little is known of protein modification in *Archaea* (35–37), but it seems possible that TK0590p and/or TK0808p could form protein conjugates that regulate archaeal replication as does ubiquitin and SUMO modi-

fication 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 (38).

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 (24, 39, 40). 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 (41–43) with several MCM homologues that are thought to have resulted from gene duplication and/or lateral gene transfer from other *Archaea* (3, 5, 42, 43).

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 coisolation results argue for the assembly of only homohexameric MCM complexes. MCM2 and MCM3 were not coisolated with His$_6$-tagged MCM1, MCM1 and MCM3 were not coisolated with His$_6$-tagged MCM2, and MCM1 and MCM2 were not coisolated 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. In contrast, the results argue that MCM3 participates in complexes that differ from those formed by MCM1 and MCM2 (Fig. 2; Table 2). Only proteins with unknown functions were coisolated using His$_6$-tagged MCM3, and MCM3 was never coisolated with a known His$_6$-tagged replication enzyme. MCM3 appears to be a member of the McmD group (42), 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.

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 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 (44). Full details of the primers used are available upon request. The amplified genes were cloned into pUMT2 (45) using restriction enzymes 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 Fig. 3. Plasmid preparations were isolated from *E. coli* DH5α cells and used directly to transform *T. kodakarensis* KW128 ($\Delta pyrF$ $\Delta trpE::pyrF$) as previously described (45, 46). 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 (46) that contained 2 g S/liter. 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 (46). 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
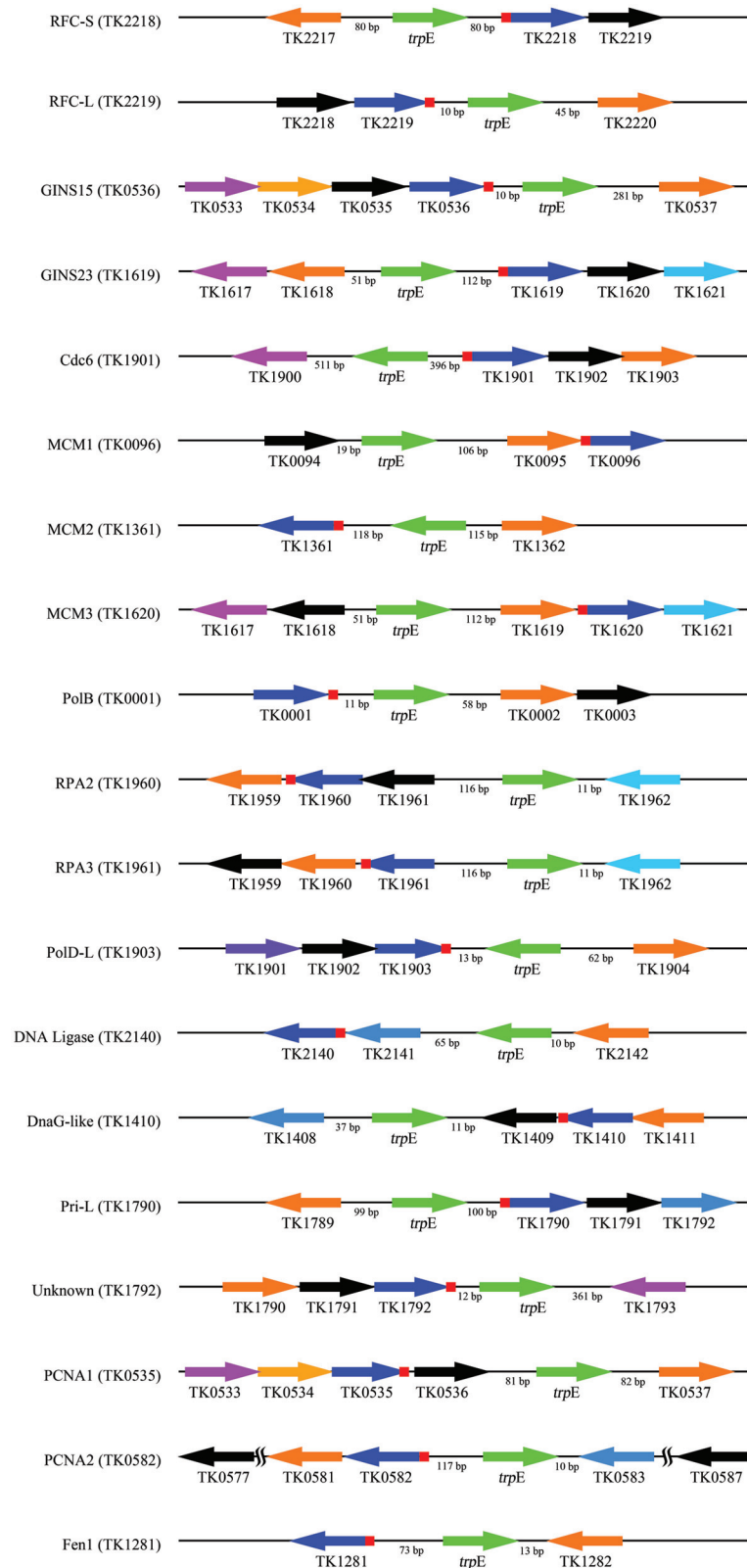
FIG 3 DNA molecules constructed and cloned into pUMT2 to produce the plasmid DNAs used to transform *T. kodakarensis* KW128. The protein that was replaced with a His₆-tagged version and the gene that encodes it are shown to the left of each construct. In each construct, the target gene is shown as a blue arrow and the 5′- or 3′-terminal location of the His₆-encoding sequence is indicated by a red box. A constitutively expressed *trp*E (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.

trpE and the gene that encoded the His$_6$-tagged version of the replication protein.

The His$_6$-encoding sequence was also added to genes that encode a subunit of RPA (RPA1; TK1959), the small subunit of euryarchaeal DNA polymerase D (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.

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 Fig. S1 in the supplemental material) 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], 500 mM NaCl, 10 mM imidazole, 10% glycerol) and lysed by sonication. After centrifugation, 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], 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], 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 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 Table S1 in the supplemental material.

ACKNOWLEDGMENTS

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.00221-10/-/DCSupplemental.

Figure S1, PDF file, 0.057 MB.

Table S1, PDF file, 0.038 MB.

REFERENCES

1. **Kornberg, A., and T. A. Baker.** 1992. DNA replication, 2nd ed. W. H. Freeman, New York, NY.
2. **Grabowski, B., and Z. Kelman.** 2003. Archaeal DNA replication: eukaryal proteins in a bacterial context. Annu. Rev. Microbiol. **57:**487–516.
3. **Chia, N., L. Cann, and G. J. Olsen.** 2010. Evolution of DNA replication protein complexes in eukaryotes and Archaea. PLoS One **5:**e10866.
4. **Kelman, Z., and M. O'Donnell.** 1995. Structural and functional similarities of prokaryotic and eukaryotic DNA polymerase sliding clamps. Nucleic Acids Res. **23:**3613–3620.
5. **Fukui, T., H. Atomi, T. Kanai, R. Matsumi, S. Fujiwara, and T. Imanaka.** 2005. Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1 and comparison with *Pyrococcus* genomes. Genome Res. **15:**352–363.
6. **Barry, E. R., and S. D. Bell.** 2006. DNA replication in the Archaea. Microbiol. Mol. Biol. Rev. **70:**876–887.
7. **Duggin, I. G., and S. D. Bell.** 2006. The chromosome replication machinery of the archaeon *Sulfolobus solfataricus*. J. Biol. Chem. **281:**15029–15032.
8. **Komori, K., and Y. Ishino.** 2001. Replication protein A in *Pyrococcus furiosus* is involved in homologous DNA recombination. J. Biol. Chem. **276:**25654–25660.
9. **Indiani, C., and M. O'Donnell.** 2006. The replication clamp-loading machine at work in the three domains of life. Nat. Rev. Mol. Cell Biol. **7:**751–761.
10. **Jeruzalmi, D., M. O'Donnell, and J. Kuriyan.** 2002. Clamp loaders and sliding clamps. Curr. Opin. Struct. Biol. **12:**217–224.
11. **Vivona, J. B., and Z. Kelman.** 2003. The diverse spectrum of sliding clamp interacting proteins. FEBS Lett. **546:**167–172.
12. **López de Saro, F. J.** 2009. Regulation of interactions with sliding clamps during DNA replication and repair. Curr. Genomics **10:**206–215.
13. **Dionne, I., R. Nookala, S. P. Jackson, A. J. Doherty, and S. D. Bell.** 2003. A heterotrimeric PCNA in the hyperthermophilic archaeaon *Sulfolobus solfataricus*. Mol. Cell **11:**275–282.
14. **Warbrick, E.** 1998. PCNA binding through a conserved motif. Bioessays **20:**195–199.
15. **Warbrick, E., W. Heatherington, D. P. Lane, and D. M. Glover.** 1998. PCNA binding proteins in *Drosophila melanogaster*: the analysis of a conserved PCNA binding domain. Nucleic Acids Res. **26:**3925–3932.
16. **Labib, K., and A. Gambus.** 2007. A key role for the GINS complex at DNA replication forks. Trends Cell Biol. **17:**271–278.
17. **MacNeill, S. A.** 2010. Structure and function of the GINS complex, a key component of the eukaryotic replisome. Biochem. J. **425:**489–500.
18. **Swiatek, A., and S. A. MacNeill.** 2010. The archaeo-eukaryotic GINS proteins and the archaeal primase catalytic subunit PriS share a common domain. Biol. Direct **5:**17.
19. **Marinsek, N., E. R. Barry, K. S. Makarova, I. Dionne, E. V. Koonin, and S. D. Bell.** 2006. GINS, a central nexus in the archaeal DNA replication fork. EMBO Rep. **7:**539–545.
20. **Courcelle, J., J. R. Donaldson, K. H. Chow, and C. T. Courcelle.** 2003. DNA damage-induced replication fork regression and processing in *Escherichia coli*. Science **299:**1064–1067.
21. **Borde, V., and J. Cobb.** 2009. Double functions for the Mre11 complex during DNA double-strand break repair and replication. Int. J. Biochem. Cell Biol. **41:**1249–1253.
22. **Mirzoeva, O. K., and J. H. Petrini.** 2003. DNA replication-dependent nuclear dynamics of the Mre11 complex. Mol. Cancer Res. **1:**207–218.
23. **Robison, J. G., J. Elliott, K. Dixon, and G. G. Oakley.** 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.
24. **Sakakibara, N., L. M. Kelman, and Z. Kelman.** 2009. Unwinding the structure and function of the archaeal MCM helicase. Mol. Microbiol. **72:**286–296.
25. **Bochman, M. L., and A. Schwacha.** 2009. The mcm complex: unwinding the mechanism of a replicative helicase. Microbiol. Mol. Biol. Rev. **73:**652–683.
26. **Zuo, Z., C. J. Rodgers, A. L. Mikheikin, and M. A. Trakselis.** 2010. Characterization of a functional DnaG-type primase in archaea: implications for a dual-primase system. J. Mol. Biol. **397:**664–676.
27. **Evguenieva-Hackenberg, E., P. Walter, E. Hochleitner, F. Lottspeich, and G. Klug.** 2003. An exosome-like complex in *Sulfolobus solfataricus*. EMBO Rep. **4:**889–893.
28. **Roppelt, V., C. F. Hobel, S. V. Albers, C. Lassek, H. Schwarz, G. Klug, and E. Evguenieva-Hackenberg.** 2010. The archaeal exosome localizes to the membrane. FEBS Lett. **584:**2791–2795.
29. **Kelman, L. M., and Z. Kelman.** 2003. Archaea: an archetype for replication initiation studies? Mol. Microbiol. **48:**605–615.
30. **Leonard, A. C., and J. E. Grimwade.** 2009. Initiating chromosome replication in *E. coli*: it makes sense to recycle. Genes Dev. **23:**1145–1150.
31. **Katayama, T., and K. Sekimizu.** 1999. Inactivation of *Escherichia coli* DnaA protein by DNA polymerase III and negative regulations for initiation of chromosomal replication. Biochimie **81:**835–840.
32. **Yoshimochi, T., R. Fujikane, M. Kawanami, F. Matsunaga, and Y. Ishino.** 2008. The GINS complex from *Pyrococcus furiosus* stimulates the MCM helicase activity. J. Biol. Chem. **283:**1601–1609.
33. **Ulrich, H. D.** 2009. Regulating post-translational modifications of the eukaryotic replication clamp PCNA. DNA Repair **8:**461–469.

34. **Forsburg, S. L.** 2004. Eukaryotic MCM proteins: beyond replication initiation. Microbiol. Mol. Biol. Rev. **68**:109–131.

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

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

37. **Makarova, K. S., and E. V. Koonin.** 2010. Archaeal ubiquitin-like proteins: functional versatility and putative ancestral involvement in tRNA modification revealed by comparative genomic analysis. Archaea doi:10.1155/2010/710303.

38. **Kirkland, P. A., and J. A. Maupin-Furlow.** 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.

39. **Costa, A., and S. Onesti.** 2009. Structural biology of MCM helicases. Crit. Rev. Biochem. Mol. Biol. **44**:326–342.

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

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

42. **Walters, A. D., and J. P. Chong.** 2010. An archaeal order with multiple minichromosome maintenance genes. Microbiology **156**:1405–1414.

43. **Krupovic, M., S. Gribaldo, D. H. Bamford, and P. Forterre.** 25 June 2010. The evolutionary history of archaeal MCM helicases: a case study of vertical evolution combined with hitch-hiking of mobile genetic elements. Mol. Biol. Evol. [Epub ahead of print.]

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

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

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