The archaeal RecJ-like proteins: nucleases and ex-nucleases with diverse roles in replication and repair

Stuart A. MacNeill

School of Biology
University of St Andrews
North Haugh
St Andrews
Fife KY16 9ST
UK.

Correspondence: Stuart MacNeill (stuart.macneill@st-and.ac.uk)

Running title: Archaeal RecJ-like proteins

Abstract

RecJ proteins belong to the DHH superfamily of phosphoesterases that has members in all three domains of life. In bacteria, the archetypal RecJ is a 5'→3' ssDNA exonuclease that functions in homologous recombination, base excision repair and mismatch repair, while in eukaryotes, the RecJ-like protein Cdc45 (which has lost its nuclease activity) is a key component of the CMG (Cdc45-MCM-GINS) complex, the replicative DNA helicase that unwinds double-stranded DNA at the replication fork. In archaea, database searching identifies genes encoding one or more RecJ family proteins in almost all sequenced genomes. Biochemical analysis has confirmed that some but not all of these proteins are components of archaeal CMG complexes and has revealed a surprising diversity in mode of action and substrate preference. In addition to this, some archaea encode catalytically inactive RecJ-like proteins, and others a mix of active and inactive proteins, with the inactive proteins being confined to structural roles only. Here, I summarise current knowledge of the structure and function of the archaeal RecJ-like proteins, focusing on similarities and differences between proteins from different archaeal species, between proteins within species, and between the archaeal proteins and their bacterial and eukaryotic relatives. Models for RecJ-like function are described and key areas for further study highlighted.

Summary

- The bacterial RecJ nuclease functions in homologous recombination, base excision repair and mismatch repair, while the eukaryotic RecJ-like protein Cdc45 is a key component of the replicative helicase, the CMG complex
- The archaea encode multiple RecJ-like proteins with presumed roles in DNA replication and repair and the maintenance of genome stability
- Structural analysis of the archaeal RecJ-like proteins reveals a high degree of similarity to bacterial RecJ, as well as features characteristic of eukaryotic Cdc45

- Biochemical analysis of the archaeal RecJ-like proteins reveals remarkable differences in substrate preference and mode of action
- Both active and inactive archaeal RecJ-like proteins (nucleases and ex-nucleases) have been implicated as CMG complex components, although the function of the CMG-associated nuclease activity remains unclear
- Attempts to further understand the cellular functions of the archaeal RecJ-like proteins are ongoing

Introduction

High-fidelity DNA replication and efficient DNA repair are fundamental for life on Earth. Understanding the biology of DNA replication and repair across the broad sweep of archaeal diversity presents a significant challenge. While core components of the chromosomal DNA replication machinery are generally well conserved across archaeal evolution, non-core components are much more divergent (1). Likewise, DNA repair biology varies enormously across the archaeal phyla, with proteins apparently important for repair processes in one phyla being completely absent in others (2).

The bacterial RecJ protein and its eukaryotic homologue Cdc45 play conserved, essential roles in DNA repair and DNA replication, respectively. RecJ is a 5'→3' ssDNA exonuclease (3) that acts in various DNA repair pathways, including mismatch repair, base excision repair and homologous recombination. The RecJ protein family forms part of the DHH superfamily of phosphoesterases, named for a characteristic amino acid signature within the catalytic domain of these proteins (4). A comprehensive phylogenetic analysis of DHH proteins across evolution identifies three major clades within this superfamily: the Ppx1 clade, the COG2404 clade and the RecJ clade (5). The Ppx1 clade contains proteins from all three domains of life, including bacterial inorganic pyrophosphatases such as Ppx1 and the eukaryotic Prune protein, while the COG2404 clade is largely made up of archaeal proteins but also includes bacterial nanoRNAse proteins. The RecJ clade includes bacterial RecJ, eukaryotic Cdc45 and the various archaeal RecJ-like proteins that are the subject of this review (5,6). Eukaryotic Cdc45 is an essential component of the CMG complex, the replicative DNA helicase that unwinds double-stranded DNA at the

replication forks (7). The CMG comprises Cdc45, the hexameric MCM helicase and the tetrameric GINS complex (thus the name CMG, for Cdc45–MCM–GINS). Unlike bacterial RecJ, Cdc45 is not active as a nuclease: key catalytic residues within the DHH domain have been lost during evolution and no nuclease activity can be detected (5,6). Instead, Cdc45 appears to play a structural role in the CMG only (8). Phylogenetically, the archaeal RecJ-like proteins can be grouped into several clusters, some of which possess all the sequence motifs characteristic of active RecJ nucleases and others that do not (5). Some of the RecJ-like proteins, both nucleases and ex-nucleases, have been shown to be components of archaeal CMG complexes, while for others, novel roles are being sought in replication and repair. Here I summarise current knowledge of the structure and function of the archaeal RecJ-like proteins, highlighting similarities and differences between proteins from different archaeal species, between proteins within species, and between the archaeal proteins and their bacterial and eukaryotic relatives. Finally, I highlight a number of outstanding questions, the answers to which will be vital for gaining a fuller understanding of the functions of the RecJ-like proteins in the archaeal kingdom.

RecJ-like proteins in the euryarchaea

Thermococcus kodakarensis

T. kodakarensis, a member of the Thermococcales, encodes two RecJ-like proteins: TK1252 and TK0155, known as TkoGAN and TkoHAN respectively (Table 1). The TkoGAN protein is perhaps the best characterised RecJ-like protein in the archaea, with information available on activity (9,10), structure (11) and *in vivo* function (9,12). TkoGAN was initially identified on the basis of its interaction with Tko GINS proteins, thus the name GAN, for GINS-associated nuclease (10,13). Figure 1 shows a schematic representation of the domain structure of the protein, with bacterial RecJ and eukaryotic Cdc45 for comparison. The TkoGAN protein comprises DHH and DHHA1 domains, the former with an integral CID (CMG-interacting domain, see below), separated by a linker region.

Purified recombinant TkoGAN displays processive 5’→3’ exonuclease activity on ssDNA substrates but similar to bacterial RecJ, has no activity on either dsDNA or

RNA (10). The three-dimensional structure of TkoGAN has been solved by X-ray crystallography, as the apo-protein and in complex with the C-terminal B-domain of the GINS subunit TkoGINS51 (11). The structure comprises the N-terminal DHH domain (Figure 2A, blue) and the C-terminal DHHA1 (orange) connected by a long α-helical linker (green). The CID (red) emerges from the DHH domain. The active site of the enzyme is located in the cleft between the DHH and DHHA1 domains, with key residues for substrate binding and catalysis being located in both. The spacing of the DHH and DHHA1 domains in the TkoGAN structure is thought to represent an open conformation (discussed further below).

Comparison of the TkoGAN structure with bacterial RecJ and eukaryotic Cdc45 proteins (Figures 2C, 2D) reveals striking similarity, underlining the evolutionary relatedness of RecJ family proteins from the three kingdoms. This is particularly evident when comparing the TkoGAN and human Cdc45 structures, as bacterial RecJ proteins possess additional C-terminal domains that are not found in either the archaeal or human proteins (Figure 2D). Co-crystallisation of the TkoGINS51 B-domain (Figure 2E, purple) shows the B-domain binding to the DHH domain of TkoGAN, consistent with earlier biochemical analysis (10). Taken together, these results are consistent with TkoGAN being a component of a CMG complex in *T. kodakarensis* and this appears to the case, as the three putative CMG components can interact *in vitro* and can be co-purified in pairwise combinations from *T. kodakarensis* cell extracts (9).

Genetic analysis has shown that TkoGAN is non-essential for growth and division of *T. kodakarensis* (9,12), in sharp contrast to the eukaryotic Cdc45 protein which is essential for DNA replication and cell cycle progression. In addition to its non-essentiality, the fact that TkoGAN is potentially active as a nuclease in the CMG also sets it apart from the inactive eukaryotic Cdc45 protein. It is not at all clear what function the 5’→3’ ssDNA-specific exonuclease would perform at the replication fork and it has even been argued that having an active nuclease in such close proximity to the replication fork could present an impediment to efficient fork progression. One model for TkoGAN function has it acting during Okazaki fragment processing by removing 5’ flap structures generated by strand displacement synthesis by DNA polymerase D (Pol D). It is clear from biochemical studies that TkoGAN interacts with Pol D (10,13) and the available genetic evidence (specifically, the inability to generate *T. kodakarensis* strains lacking both GAN and either flap endonuclease 1

(Fen1) or RNAseHII, both of which likely function in Okazaki fragment processing) is consistent with a role for TkoGAN in this process (12). Note, however, that interactions between RecJ proteins and Pol D have not been reported in other archaeal organisms as yet and so models for TkoGAN function may not translate more widely. Further analysis will be required to address this issue.

The second RecJ-like protein encoded by *T. kodakarensis* is known as HAN for Hef-associated nuclease (Table 1). TkoHAN (TK0155) was first identified in a yeast two-hybrid screen for proteins that interacted with the intrinsically disordered region (IDR) that separates the N-terminal helicase and C-terminal nuclease domains of the TkoHef protein (14). Hef (helicase-associated endonuclease for fork-structured DNA) has been shown to play a vital role in processing of stalled replication forks in the euryarchaea. With both *T. kodakarensis* (15) and *Haloferax volcanii* (16,17), cells lacking Hef are highly sensitive to the DNA damaging agent mitomycin C which causes formation of interstrand crosslinks (ICLs) that result in replication fork stalling. The human Hef homologue FANCM has a similar role and mutations in the *FANCM* gene result in the rare genetic disease Fanconi anaemia (18,19). Amongst the archaea, HAN proteins are found across the euryarchaeal and in some Asgard archaea (see below) but are largely absent from the crenarchaea.

The archaeal HAN proteins have a distinct domain structure that sets them apart from typical RecJ homologues like TkoGAN. The N-terminal part of the protein comprises four structural motifs: two putative zinc finger motifs (absent in the haloarchaeal HAN proteins) followed by two OB folds (Figure 1). The RecJ-like region including DHH and DHHA1 domains is C-terminal.

Biochemical analysis of TkoHAN (20) has shown the protein to have robust 3'→5' exonuclease activity on both ssDNA and RNA substrates, in contrast to TkoGAN which possesses 5'→3' activity and on ssDNA only. With TkoHAN, cleavage stops ~ 10 nucleotides from the 5' end of the substrate, suggesting that the 5' region is bound by the protein and inaccessible to the active site. The putative zinc fingers and OB folds offer possible means of nucleic acid binding but this is yet to be investigated. Incubating synthetic replication fork structures simultaneously with TkoHAN and TkoHef reveals both stimulatory and inhibitory effects: TkoHef stimulates TkoHAN activity, while TkoHAN inhibits TkoHef nuclease activity (20). Genetic analysis shows that cells lacking TkoHAN are viable and somewhat

sensitive to MMC, consistent with the protein functioning in the repair of interstrand crosslinks.

Pyrococcus furiosus

The hyperthermophilic euryarchaeon *Pyrococcus furiosus*, like *T. kodakarensis* a member of the Thermococcales, encodes two RecJ-like proteins: PfuRecJ (PF2055) and a HAN orthologue (PF0399) that has yet to be characterised. PfuRecJ possesses intact DHH nuclease motifs and like TkoGAN, displays 5'→3' exonuclease activity on ssDNA (21). However, PfuRecJ also exhibits 3'→5' exonuclease activity on ssRNA substrates and on double-stranded RNA/DNA hybrid substrates with recessed RNA 3' ends. The latter activity, which is stimulated by RPA binding to the exposed single-stranded DNA region, is highest when the 3' end of the RNA strand is mismatched to its partner DNA. This *in vitro* situation mirrors that seen *in vivo* at the replication fork when primase adds the wrong ribonucleotide to a growing Okazaki fragment primer, raising the possibility that PfuRecJ could act to proofread 3' mismatched primers to facilitate their extension by PfuPolB. PfuPolB appears unable to extend 3'-mismatched primers, implying that such an activity must exist to allow efficient replication to take place *in vivo* (21).
The structure of recombinant PfuRecJ protein has been solved (22), revealing an overall three-dimensional structure made up of DHH, DHHA1 and CID domains that is very similar to that of TkoGAN (11). One difference between the PfuRecJ and TkoGAN structures is in the orientation of the long α-helical linker that connects the DHH and DHHA1 domains, which results in these domains being located closer to one another in PfuRecJ than in TkoGAN (Figure 2B). It is speculated that formation of this closed conformation is necessary for PfuRecJ activity. Like TkoGAN, PfuRecJ interacts with GINS via the DHH domain of PfuRecJ binding to the B-domain of the GINS51 subunit of PfuGINS (11), suggesting that PfuRecJ is also likely to assemble into a CMG. Whether both the 5'→3' and 3'→5' exonuclease activities of PfuRecJ are active in the CMG remains to be seen.

Thermoplasma acidophylum

T. acidophylum, a thermoacidophilic eukaryarcheon of the order Thermoplasmatales, encodes a pair of RecJ-like proteins, designated TaRecJ1 and TaRecJ2 (see Table 1), both of which are members of the RecJ clade that includes TkoGAN (23). *T. acidophylum* does not appear to encode a HAN homologue.

Biochemical analysis of recombinant proteins has confirmed that both TaRecJ1 and TaRecJ2 are exonucleases, but with different substrate preferences and different polarities: TaRecJ1 possesses 5'→3' ssDNA-specific exonuclease activity (like TkoGAN) while TaRecJ2 has 3'→5' exonuclease activity on RNA and ssDNA, with RNA being the preferred substrate (23). The similarities in polarity and substrate preference between TaRecJ1 and TkoGAN (10) would suggest TaRecJ1 as a component in the *T. acidophylum* CMG but remarkably, this is not the case: it is TaRecJ2 that forms a stable complex with TaGINS. Two molecules of TaRecJ2 bind to the homotetrameric TaGINS complex in an interaction for which the C-terminal B-domain of the TaGINS protein is both necessary and sufficient. TaRecJ2 does not associate with TaMCM but the TaRecJ2-TaGINS complex does, forming an intact CMG that can also be detected in *T. acidophylum* cell extracts. No interaction is detected between TaRecJ1 and either TaGINS or TaMCM, either *in vitro* or *in vivo* (23). The results obtained with TkoGAN and TacRecJ2 highlight a key conundrum when one comes to consider the roles of the the archaeal RecJ-like proteins as CMG complex components: with their different substrate preferences and polarities, the TkoGAN and TacRecJ2 proteins cannot be performing the same function in their respective CMGs. In short, not all archaeal CMGs are the same.

Methanocaldococcus jannaschii

The thermophilic methogen *Methanocaldococcus jannaschii*, a member of the Methanococcales, encodes three RecJ-like proteins: two members of the RecJ clade that includes TkoGAN (MJ0977 and MJ0831) and a TkoHAN homologue (MJ1198) that is largely unstudied, although the solution structure of part of the N-terminal region of this protein was determined by NMR as part of a structural genomics project (PDB ID: 2K52). Investigation of the function of these enzymes began with the demonstration that expression of either MJ0977 or MJ0831 (but not MJ1198) in

E. coli was able to partially rescue the UV sensitivity of a *recJ* mutant strain, implying at least partial conservation of function across evolution (24).
The MJ0977 and MJ0831 proteins have subsequently been designated MjaRecJ1 and MjaRecJ2, respectively, and characterised in detail biochemically (Table 1) (25). MjaRecJ1, which is the more closely related to TkoGAN and TacRecJ1, possesses 5'→3' exonuclease activity on both ssDNA and ssRNA substrates but with a preference for DNA over RNA, while MjaRecJ2 acts with 3'→5' polarity on both ssDNA and ssRNA but with a preference for RNA over DNA, broadly mirroring what is seen with the TacRecJ1 and TacRecJ2 proteins. However, unlike the situation in *T. acidopylum*, where TacRecJ2 has been shown to bind stably to GINS, neither MjaRecJ1 and MjaRecJ2 appears capable of forming a stable complex with *M. jannaschii* GINS, raising the possibility that interaction could require the presence of third partner (*M. jannaschii* MCM being the obvious candidate), or be transient in nature and thus escape detection using standard methods, or that neither protein is a component of the *M. jannaschii* replicative helicase.

Haloferax volcanii

The genetically tractable halophilic euryarchaeon *Haloferax volcanii* is an important model for studying archaeal DNA replication and repair (26). A number of components of the *H. volcanii* replication machinery have already been characterised genetically, including origin binding proteins (27), the MCM helicase (28), the sliding clamp PCNA (29,30) and both ATP- and NAD-dependent DNA ligases (31). *H. volcanii* encodes four RecJ-like proteins designated RecJ1-RecJ4, although only RecJ1 and RecJ3 have retained intact catalytic domains (see following section for discussion). The ex-nuclease HfxRecJ2 co-purifies from cell extracts with the HfxGINS protein (unpublished results from the author's group) and is presumably a component of the *H. volcanii* CMG, whereas the nuclease HfxRecJ3 is the HAN orthologue. None of the four proteins is essential for viability although some of the deletion strains exhibit mild sensitivity to DNA damaging agents.

RecJ-like proteins in the crenarchaea

The studies described above amply illustrate the diversity of RecJ family protein function in the euryarchaeal, but these organisms constitute only a single branch of the archaeal tree, albeit one that has been particularly extensively studied. In the crenarchaea, and more widely in the TACK superphylum (32) that includes the Thaumarchaeota, Aigarchaeota, Crenarchaeota and Korarchaeota phyla as well as number of newly discovered phyla such as the Bathyarchaeota (33), there is even greater diversity, but to date the only RecJ-like protein from these organisms to be characterised comes from the crenarchaeal Sulfolobales. This protein (originally designated RecJdbh, now called simply Cdc45) was in fact the first archaeal RecJ-like protein to be proposed as a component of the presumptive archaeal CMG (34), several years before the phylogenetic link between RecJ and Cdc45 was established (5,6).

The *Sulfolobus solfataricus* Cdc45 protein was identified through its interaction with GINS (34) but its similarity to archaeal RecJ-like proteins such as TkoGAN is low. The reason for this is that it has a highly degenerate DHH domain that lacks key catalytic residues found in other RecJ homologues. As a result, the SsoCdc45 protein is not a nuclease. Disruption of the catalytic domain is not confined to *S. solfataricus* or the Sulfolobales either, as other crenarchaeal RecJ-like proteins share this property (5). Beyond the Sulfolobales, in the Thaumarchaeota or Bathyarchaeota for example, more typical GAN-like RecJ proteins are widespread (data not shown).

Despite the lack of catalytic activity, analysis of the Sulfolobales Cdc45 protein in *S. acidocaldarius* (Sac) and *S. islandicus* (Sis) has provided important insights into its properties (35). The *S. islandicus* CMG complex has been reconstituted *in vitro* using purified SisCdc45, GINS and MCM proteins. These experiments revealed that Cdc45-GINS stimulates MCM helicase activity. By combining cell cycle synchronisation and chromatin immunoprecipitation techniques, the SacCdc45 protein has been shown to be enriched at replication origins in late G1/early S-phase but that this enrichment is lost in favour of enrichment at origin-distal loci in mid-S-phase and G2 (35), observations that are consistent with the Cdc45 protein moving with replication forks as part of the CMG.

RecJ-like proteins in the Asgard superphylum

Very recently, a new superphylum of uncultivated archaea, the Asgard archaea, was identified (36,37). Phylogenetic analysis of metagenomes from members of this group suggests that these organisms are the closest extant relatives of eukaryotes and that eukaryotic life may have arisen from within this lineage or from an as yet undiscovered sister lineage . The genomes of Asgard archaea encode numerous proteins that were once considered to be specific to eukaryotes (so-called eukaryotic signature proteins or ESPs) such as actin homologues and Ras-superfamily small GTPases (38,39). Investigating the Asgard genomes could offer insights into the evolution of the archaeal RecJ protein towards eukaryotic Cdc45.
Although none of the assembled Asgard metagenomes is complete, database searching identifies several RecJ-like proteins encoded by representatives of this superphylum (Table 2). The metagenomes of all four recognised Asgard phyla – the Thorarchaeota, Odinarchaeota, Heimdallarchaeota and Lokiarchaeota – encode a protein with approximately 30% amino sequence identity to TkoGAN and all the sequence motifs characteristic of an active RecJ-like enzyme (see Supplementary information Figure S1A for sequence alignment). In addition, homologues of TkoHAN can be found encoded by four Heimdallarchaeota metagenomes. These proteins are 25-35% identical to TkoHAN, have the characteristic HAN domain structure (including N-terminal zinc finger and OB fold domains) and the conserved amino acids residues required for nuclease activity (Figure S1B). No HAN homologues are apparent in the Thorarchaeota, Odinarchaeota or Lokiarchaeota, however, implying that the gene has been lost during the evolution of these phyla (Table 2).

Outstanding questions

Analysis of the activities, structure and genetics of the archaeal RecJ-like proteins has produced intriguing insights into the differing substrate preferences and modes of action of this diverse group of proteins. However, a number of questions remain to be answered before we can fully understand the functions of these proteins. Perhaps most importantly, are the RecJ-like exonucleases that function as part of the CMG (TkoGAN and TacRecJ2 for example) active at replication forks, and if so, on what substrates? And if these proteins are not active as part of the CMG, how is their

activity suppressed? Exonuclease activity is clearly not required for archaeal CMG function, at least not in the Sulfolobales, so an archaeal CMG containing an inhibited rather than a permanently inactivated RecJ-like exonuclease would be entirely possible. In which case, could the RecJ-like proteins have different roles (catalytic versus structural) depending on whether they are free in solution or part of a CMG, and if so, how can we tease these roles apart? Neither of the models put forward for TkoGAN in Okazaki fragment processing (12) and PfuRecJ in proofreading 3' mismatched primers (22) seems entirely consistent with the exonuclease active site being located in the advancing CMG unwinding DNA at the replication fork. Could these represent non-CMG roles for these enzymes? In addition, for the RecJ-like proteins such as TaRecJ1 that do not appear to be CMG components, how can we identify their cellular functions in systems that are not currently genetically tractable? At present, few archaeal systems offer this option, and none outside the well-studied Euryarchaeota and Crenarchaeota. In conclusion, it is clear that the RecJ-like proteins have the potential to be important players in DNA replication and repair in the archaea and present an exciting area for further investigation.

Abbreviations

CID (CMG interacting domain); CMG (Cdc45-MCM-GINS); GAN (GINS-associated nuclease); GINS (go-ichi-ni-san); HAN (Hef-associated nuclease); Hef (helicase-associated endonuclease for fork-structured DNA); PolD (DNA polymerase D).

Funding

Research on archaeal genome stability in the MacNeill lab was most recently funded by the USAF Office of Scientific Research [award number FA9550-10-1-0421].

Acknowledgements

I am grateful to my colleagues in St Andrews and elsewhere for helpful discussions on this topic.

Competing interests

The Author declares that there are no competing interests associated with this manuscript.

References

1. Kelman, L.M. and Kelman, Z. (2014) Archaeal DNA replication. *Annu Rev Genet*, **48**, 71-97.
2. White, M.F. and Allers, T. (2018) DNA repair in the Archaea - an emerging picture. *FEMS Microbiol Rev*., doi: 10.1093/femsre/fuy020.
3. Lovett, S.T. and Kolodner, R.D. (1989) Identification and purification of a single-stranded-DNA-specific exonuclease encoded by the *recJ* gene of *Escherichia coli*. *Proc Natl Acad Sci U S A*, **86**, 2627-2631.
4. Aravind, L. and Koonin, E.V. (1998) A novel family of predicted phosphoesterases includes *Drosophila* prune protein and bacterial RecJ exonuclease. *Trends Biochem Sci*, **23**, 17-19.

5. 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.
6. Sanchez-Pulido, L. and Ponting, C.P. (2011) Cdc45: the missing RecJ ortholog in eukaryotes? *Bioinformatics (Oxford, England)*, **27**, 1885-1888.
7. Burgers, P.M. and Kunkel, T.A. (2017) Eukaryotic DNA Replication Fork. *Annu Rev Biochem*., **86**, 417-438.
8. Pellegrini, L. (2017) Structural insights into Cdc45 function: was there a nuclease at the heart of the ancestral replisome? *Biophys Chem*, **225**, 10-14.
9. Nagata, M., Ishino, S., Yamagami, T., Ogino, H., Simons, J.R., Kanai, T., Atomi, H. and Ishino, Y. (2017) The Cdc45/RecJ-like protein forms a complex with GINS and MCM, and is important for DNA replication in *Thermococcus kodakarensis*. *Nucleic Acids Res*, **45**, 10693-10705.
10. 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.
11. Oyama, T., Ishino, S., Shirai, T., Yamagami, T., Nagata, M., Ogino, H., Kusunoki, M. and Ishino, Y. (2016) Atomic structure of an archaeal GAN suggests its dual roles as an exonuclease in DNA repair and a CMG component in DNA replication. *Nucleic Acids Res*, **44**, 9505-9517.
12. Burkhart, B.W., Cubonova, L., Heider, M.R., Kelman, Z., Reeve, J.N. and Santangelo, T.J. (2017) The GAN exonuclease or the flap endonuclease Fen1 and RNase HII are necessary for viability of *Thermococcus kodakarensis*. *J Bacteriol*, **199**, doi: 10.1128/JB.00141-17.
13. Li, Z., Santangelo, T.J., Cubonova, L., Reeve, J.N. and Kelman, Z. (2010) Affinity purification of an archaeal DNA replication protein network. *MBio*, doi: 10.1128/mBio.00221-10.
14. Ishino, S., Yamagami, T., Kitamura, M., Kodera, N., Mori, T., Sugiyama, S., Ando, T., Goda, N., Tenno, T., Hiroaki, H. *et al.* (2014) Multiple interactions of the intrinsically disordered region between the helicase and nuclease domains of the archaeal Hef protein. *J Biol Chem*, **289**, 21627-21639.

15. Fujikane, R., Ishino, S., Ishino, Y. and Forterre, P. (2010) Genetic analysis of DNA repair in the hyperthermophilic archaeon, *Thermococcus kodakaraensis*. *Genes & Genet Syst*, **85**, 243-257.
16. Lestini, R., Laptenok, S.P., Kuhn, J., Hink, M.A., Schanne-Klein, M.C., Liebl, U. and Myllykallio, H. (2013) Intracellular dynamics of archaeal FANCM homologue Hef in response to halted DNA replication. *Nucleic Acids Res*, **41**, 10358-10370.
17. Lestini, R., Duan, Z. and Allers, T. (2010) The archaeal XPF/Mus81/FANCM homolog Hef and the Holliday junction resolvase Hjc define alternative pathways that are essential for cell viability in *Haloferax volcanii*. *DNA Repair (Amst)*, **9**, 994-1002.
18. Mosedale, G., Niedzwiedz, W., Alpi, A., Perrina, F., Pereira-Leal, J.B., Johnson, M., Langevin, F., Pace, P. and Patel, K.J. (2005) The vertebrate Hef ortholog is a component of the Fanconi anemia tumor-suppressor pathway. *Nat Struct Mol Biol*, **12**, 763-771.
19. Meetei, A.R., Medhurst, A.L., Ling, C., Xue, Y., Singh, T.R., Bier, P., Steltenpool, J., Stone, S., Dokal, I., Mathew, C.G. *et al.* (2005) A human ortholog of archaeal DNA repair protein Hef is defective in Fanconi anemia complementation group M. *Nat Genet*, **37**, 958-963.
20. Nagata, M., Ishino, S., Yamagami, T., Simons, J.R., Kanai, T., Atomi, H. and Ishino, Y. (2017) Possible function of the second RecJ-like protein in stalled replication fork repair by interacting with Hef. *Sci Rep*, **7**, 16949.
21. Yuan, H., Liu, X.P., Han, Z., Allers, T., Hou, J.L. and Liu, J.H. (2013) RecJ-like protein from *Pyrococcus furiosus* has 3'-5' exonuclease activity on RNA: implications for proofreading of 3'-mismatched RNA primers in DNA replication. *Nucleic Acids Res*, **41**, 5817-5826.
22. Li, M.J., Yi, G.S., Yu, F., Zhou, H., Chen, J.N., Xu, C.Y., Wang, F.P., Xiao, X., He, J.H. and Liu, X.P. (2017) The crystal structure of *Pyrococcus furiosus* RecJ implicates it as an ancestor of eukaryotic Cdc45. *Nucleic Acids Res*, **45**, 12551–12564.
23. Ogino, H., Ishino, S., Kohda, D. and Ishino, Y. (2017) The RecJ2 protein in the thermophilic archaeon *Thermoplasma acidophilum* is a 3'-5' exonuclease that associates with a DNA replication complex. *J Biol Chem*, **292**, 7921-7931.

24. Rajman, L.A. and Lovett, S.T. (2000) A thermostable single-strand DNase from *Methanococcus jannaschii* related to the RecJ recombination and repair exonuclease from *Escherichia coli*. *J Bacteriol*, **182**, 607-612.
25. Yi, G.S., Song, Y., Wang, W.W., Chen, J.N., Deng, W., Cao, W., Wang, F.P., Xiao, X. and Liu, X.P. (2017) Two archaeal RecJ nucleases from *Methanocaldococcus jannaschii* show reverse hydrolysis polarity: implication to their unique function in archaea. *Genes (Basel)*, doi: 10.3390/genes8090211.
26. Giroux, X. and MacNeill, S.A. (2015) Molecular genetic methods to study DNA replication protein function in *Haloferax volcanii*, a model archaeal organism. *Methods Mol Biol*, **1300**, 187-218.
27. 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.
28. Kristensen, T.P., Maria Cherian, R., Gray, F.C. and MacNeill, S.A. (2014) The haloarchaeal MCM proteins: bioinformatic analysis and targeted mutagenesis of the β7-β8 and β9-β10 hairpin loops and conserved zinc binding domain cysteines. *Frontiers in microbiology*, **5**, 123.
29. Winter, J.A., Christofi, P., Morroll, S. and Bunting, K.A. (2009) The crystal structure of Haloferax volcanii proliferating cell nuclear antigen reveals unique surface charge characteristics due to halophilic adaptation. *BMC Struct Biol*, **9**, 55.
30. Morgunova, E., Gray, F.C., MacNeill, S.A. and Ladenstein, R. (2009) Structural insights into the adaptation of proliferating cell nuclear antigen (PCNA) from *Haloferax volcanii* to a high-salt environment. *Acta Cryst D*, **65**, 1081-1088.
31. Zhao, A., Gray, F.C. and MacNeill, S.A. (2006) ATP- and NAD^+-dependent DNA ligases share an essential function in the halophilic archaeon *Haloferax volcanii*. *Mol Microbiol*, **59**, 743-752.
32. Guy, L. and Ettema, T.J. (2011) The archaeal 'TACK' superphylum and the origin of eukaryotes. *Trends Microbiol*, **19**, 580-587.
33. Castelle, C.J. and Banfield, J.F. (2018) Major new microbial groups expand diversity and alter our understanding of the Tree of Life. *Cell*, **172**, 1181-1197.

34. 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.
35. Xu, Y., Gristwood, T., Hodgson, B., Trinidad, J.C., Albers, S.V. and Bell, S.D. (2016) Archaeal orthologs of Cdc45 and GINS form a stable complex that stimulates the helicase activity of MCM. *Proc Natl Acad Sci U S A*, **113**, 13390-13395.
36. Zaremba-Niedzwiedzka, K., Caceres, E.F., Saw, J.H., Backstrom, D., Juzokaite, L., Vancaester, E., Seitz, K.W., Anantharaman, K., Starnawski, P., Kjeldsen, K.U. *et al.* (2017) Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature*, **541**, 353-358.
37. Spang, A., Saw, J.H., Jorgensen, S.L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A.E., van Eijk, R., Schleper, C., Guy, L. and Ettema, T.J. (2015) Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature*, **521**, 173-179.
38. Spang, A., Eme, L., Saw, J.H., Caceres, E.F., Zaremba-Niedzwiedzka, K., Lombard, J., Guy, L. and Ettema, T.J.G. (2018) Asgard archaea are the closest prokaryotic relatives of eukaryotes. *PLoS Genet*, **14**, e1007080.
39. Eme, L., Spang, A., Lombard, J., Stairs, C.W. and Ettema, T.J.G. (2018) Archaea and the origin of eukaryotes. *Nat Rev Microbiol*, **16**, 120.
40. Zimmermann, L., Stephens, A., Nam, S.Z., Rau, D., Kubler, J., Lozajic, M., Gabler, F., Soding, J., Lupas, A.N. and Alva, V. (2017) A completely reimplemented MPI Bioinformatics Toolkit with a new HHpred server at its Core. *J Mol Biol*. 430, 2237-2243.

Table 1: Archaeal RecJ-like proteins

Protein name	Properties	References
TkoGAN (TK1252)	Non-essential 5'→3' ssDNA-specific exonuclease. Component of Tko CMG complex. Structure solved (PDB ID: 5GHT).	(9-13)
TkoHAN (TK0155)	Non-essential 3'→5' exonuclease acting on both ssDNA and RNA substrates. Protein has tandem N-terminal zinc finger and OB fold motifs, and interacts with Hef nuclease-helicase.	(14,20)
TacRecJ1 (TA_RS02725)	5'→3' ssDNA-specific exonuclease	(23)
TacRecJ2 (TA_RS05865)	3'→5' exonuclease with preference for RNA over ssDNA. Component of Tac CMG complex.	(23)
PfuRecJ (PF2055)	Possesses both 5'→3' ssDNA exonuclease and 3'→5' RNA exonuclease activities. Structure solved (PDB ID: 5X4H).	(21,22)
SsoCdc45 (SSO0295) and related Sulfolobales orthologues	Key catalytic residues absent so lacks nuclease activity. Structural component of Sulfolobales CMG complex.	(34,35)
MjaRecJ1 (MJ0977)	5'→3' ssDNA-specific exonuclease.	(24,25)
MjaRecJ2 (MJ0831)	3'→5' exonuclease with preference for ssRNA	(24,25)

Species abbreviations: Tko (*Thermococcus kodakarensis*), Tac (*Thermoplasma acidophilum*), Pfu (*Pyrococcus furiosus*), Sso (*Sulfolobus solfataricus*), Sac (*S. acidocaldarius*), Sis (*S. islandicus*), Mja (*Methanocaldococcus jannaschii*). Note that the database sequence for TacRecJ2 (TA_RS05865) is incorrect – the corrected sequence can be found in ref. (23).

Table 2: RecJ-like proteins in Asgard archaea

Phylum	GAN-like protein	HAN-like protein
Lokiarchaeota	✓	-
Thorarchaeota	✓	-
Odinarchaeota	✓	-
Heimdallarchaeota	✓	✓

Asgard RecJ-like proteins were identified by BLAST (blastp) searching of Asgard group (taxid: 1935183) proteins in the NCBI non-redundant protein sequences database (nr) using default parameters with TkoGAN, TkoHAN, PfuRecJ, TacRecJ1, TacRecJ2 and SsoCdc45 as query sequences.

Figure legends

Figure 1: Domain structure of RecJ-like proteins in the three kingdoms of life. Schematic representation of the domain structures of RecJ family proteins from bacteria (*E. coli*, EcRecJ), archaea (*T. kodakarensis*, TkoGAN and TkoHAN) and eukaryotes (human Cdc45). Note that the Cdc45 lacks conserved catalytic residues in the DHH and DHHA1 domains and is not an active nuclease. Individual protein domains are coloured as follows: DHH, dark blue; DHHA1, orange; CID, red; the linker region between DHH and DHHA1 domains, dark green; OB fold domain (DrRecJ only), cyan; tandem zinc finger domain in TkoHAN, dark red. See text for details and references.

Figure 2: Structures of RecJ family proteins. A. TkoGAN nuclease (PDB ID: 5GHT). **B.** PfuRecJ (5X4H). **C.** Human Cdc45 (5DGO). **D.** DrRecJ (5F55). **E.** Structure of TkoGAN bound to the B-domain of the TkoGINS51 protein (5GHS). The GINS51 B-domain binds to the DHH domain of TkoGAN. Individual protein domains are coloured as follows: DHH, dark blue; DHHA1, orange; CID, red; the linker region between DHH and DHHA1 domains, dark green; OB fold domain (DrRecJ only), cyan; C-terminal domain (CTD, DrRecJ only), light green; GINS51 B-domain, magenta. See text for details and references. Figures were prepared using PyMol (The PyMOL Molecular Graphics System, Version 2.0.7 Schrödinger, LLC).

Figure 1

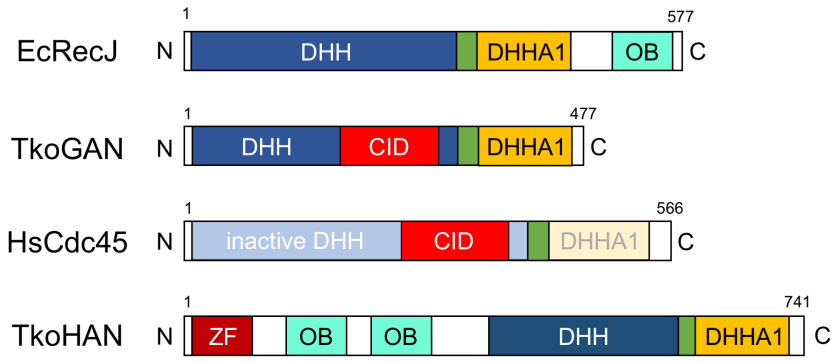
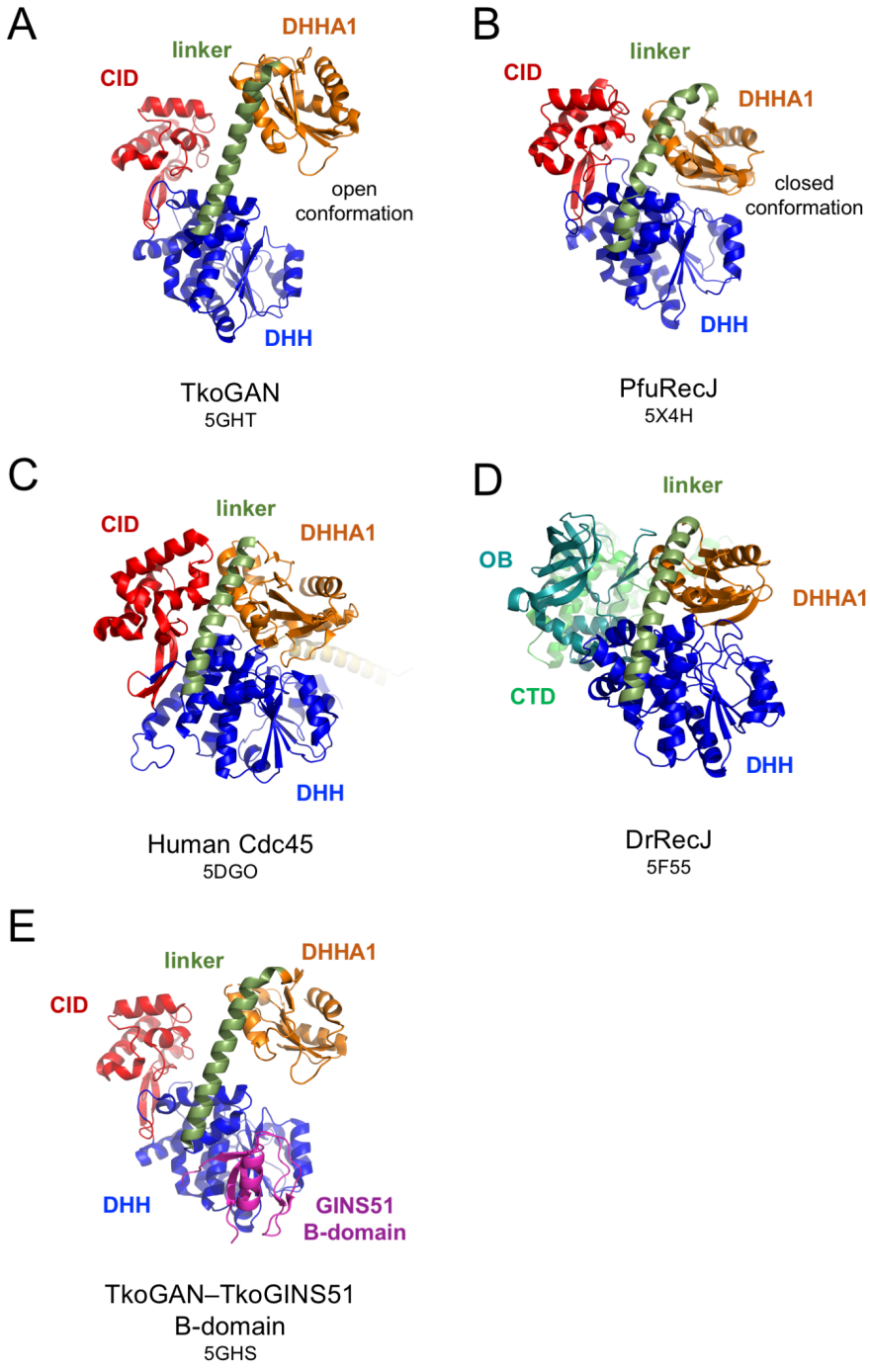

EcRecJ
N
1
577
DHH
DHHA1
OB
C
TkoGAN
N
1
477
DHH
CID
DHHA1
C
HsCdc45
N
1
566
inactive DHH
CID
DHHA1
C
TkoHAN
N
1
741
ZF
OB
OB
DHH
DHHA1
C

Figure 2


A
DHHA1
linker
CID
open
conformation
DHH
TkoGAN
5GHT
B
linker
CID
DHHA1
closed
conformation
DHH
PfuRecJ
5X4H
C
CID
linker
DHHA1
DHH
Human Cdc45
5DGO
D
linker
OB
DHHA1
CTD
DHH
DrRecJ
5F55
E
linker
DHHA1
CID
DHH
GINS51
B-domain
TkoGAN–TkoGINS51
B-domain
5GHS