Archaeal DNA polymerases: new frontiers in DNA replication and repair
Christopher D. O. Cooper 1*

1Department of Biological and Geographical Sciences, School of Applied Sciences, University of Huddersfield, Queensgate, Huddersfield, West Yorkshire, HD1 3DH, United Kingdom.

*Corresponding author:
c.d.cooper@hud.ac.uk
Tel: +44 (0)1484 471736

Running title
Archaeal DNA polymerases

Keywords
Archaea; DNA polymerase; exonuclease; DNA replication; DNA repair; replisome; holoenzyme; polymerase chain reaction; sequencing; biotechnology; polB1; Dpo1; PolB; PolD; PCNA, RFC.

Abstract
Archaeal DNA polymerases have long been studied due to their superior properties for DNA amplification in the Polymerase Chain Reaction and DNA sequencing technologies. However a full comprehension of their functions, recruitment and regulation as part of the replisome during genome replication and DNA repair lags behind well-established bacterial and eukaryotic model systems. The archaea are evolutionarily very broad, but a number of studies in the major model systems of both Crenarchaeota and Euryarchaeota are starting to yield significant increases in understanding of the functions of DNA polymerases in the respective phyla. Recent advances in biochemical approaches and in archaeal genetic models allowing knockout and epitope tagging have led to significant increases in our understanding, including DNA polymerase roles in Okazaki fragment maturation on the lagging strand, towards reconstitution of the replisome itself. Furthermore, poorly characterised DNA polymerase paralogues are finding roles in DNA repair and CRISPR immunity. This review attempts to provide a current update on the roles of archaeal DNA polymerases in both DNA replication and repair, addressing significant questions that remain for this field.
Introduction

All forms of life face a requirement to replicate their genomes in an accurate and timely manner [1]. Furthermore, DNA may be damaged following both endogenous and exogenous insults, with mutations potentially resulting if the lesions are not corrected by specific repair pathways, prior to DNA replication. Many archaebacteria inhabit physical or chemical extremes, so could be expected to experience significant levels of exogenously-induced damage [2], therefore archaebacteria are presumed to encode resilient DNA repair mechanisms [3]. Organisms hence require DNA polymerases, not only to synthesise daughter copies of genomic DNA during replication, but to also fill across gaps arising from lesion removal during DNA repair.

Since the discovery of Pol I from Escherichia coli in the 1950’s [4], although much research has been undertaken into DNA polymerases and their roles in replication and repair, a significant force driving their study has been their biotechnological exploitation as reagents for DNA sequencing and DNA amplification in the Polymerase Chain Reaction (PCR). DNA polymerases from thermophilic archaebacteria in particular have been well-studied (e.g. Pfu from Pyrococcus furiosus), generally being equipped with a 3’-5’ exonuclease activity allowing them to ‘proofread’ DNA synthesis, ensuring accurate DNA amplification [5]. Significant advances have also been made in engineering archaebacterial polymerases for biotechnology, including fusion to DNA binding domains to enhance DNA synthesis processivity [6], and mutagenesis to allow incorporation of modified [7] or even artificial nucleotides [8]. Such mutagenesis approaches have been key to next-generation sequencing approaches [5,9], for instance, specific variants of Thermostococcus sp. 9°N PoIB DNA polymerase are able to incorporate bulky reversible dye-terminators required for sequencing-by-synthesis approaches, such as Illumina.

DNA polymerases are classified into 6 families (A, B, C, D, X and Y) and comprise a conserved nucleotidyltransferase catalytic mechanism, with strongly conserved motifs for both polymerase and exonuclease activities [10,11]. These DNA polymerase families share a core structure, shaped like a cupped right hand that encompasses the DNA primer-template [12,13]. This comprises several sub-domains important for catalysis (thumb, fingers, palm), with some appended specialised domains, such as the uracil binding domain in some Family B polymerases [14]. Family B polymerases are found across all cellular life and viruses, and at least one Family B polymerase (PoIB) is present in all archaebacteria (Figure 1) [15,16]. Biochemical analysis and genome sequencing however soon indicated the presence of Y family polymerases in some Crenarchaeota [17], and a non-canonical DNA polymerase (Family D, PoLD) in the Euryarchaeota (Figure 1) [18]. Unusually, PoLD is a heterodimer comprising a smaller DP1 3’-5’ exonuclease subunit and the DP2 DNA polymerase subunit [18], and has also been successfully applied to PCR [19]. The evolutionary relationship of PoLD to other DNA polymerases is unclear, although DP1 is thought to be ancestral to euarkyotic Family B polymerase structural subunits [15], but the DP2 sequence is very different to the canonical DNA polymerases. However, the recent structural determination of PoLD [20] demonstrates that the DP2 subunit has a double-psi β-barrel fold conserved in multi-subunit RNA polymerases, potentially bridging DNA replication and transcription, rather than originating from a highly-divergent Family B polymerase [15,20].
Archaea also encode primases, DNA-dependent RNA polymerases recruited to origins to initiate DNA replication by de novo synthesis of RNA primers [21]. Archaeal primases are similar to eukaryotic primases, originally thought to comprise a heterodimer of catalytic PriS and regulatory PriL subunits (Figure 1) [22]. However, recent studies suggest a third subunit PriX is important for primase activity, albeit phylogenetically restricted to some Crenarchaeota (Figure 1) [23]. Surprisingly, archaeal primases can catalyse not only RNA but also DNA synthesis, able to replicate up to several kilobases of product [22,24] and so maybe classed as non-canonical DNA polymerases. Hence, primases are thought to sequentially hand-off between RNA and DNA synthesis [25], prior to transferring to a DNA polymerase. Putative bacterial-like DnaG primases are also found in archaeal genomes [26,27], although they appear to be associated with RNA degradation rather than replication [27,28].

Archaeal DNA polymerase phylogenetic relationships are complex, as many Crenarchaeota contain two additional Family B members (polB2/Dpo2 and polB3/Dpo3) resulting from gene duplications, in addition to the polB1 (Dpo1) presumed major replicase (Figure 1) [29,30]. Furthermore, PolB from Euryarchaeota are more closely related to polB3 [15], suggesting polB1 and polB2 were derived from an ancestral polB3 following the Euryarchaeota/Crenarchaeota split. Moreover, although TACK superphylum members (Thaumarchaeota, Aigarchaeota, Crenarchaeota, Korarchaeota) [31] encode polB1/Dpo1, they all apart from Crenarchaeota also contain the euryarchaeotal-like PolD [16]. Such confusing evolutionary histories within the archaean are likely to result from accelerated rates of evolution in some clades, alongside significant horizontal gene transfer [15].

Although significant research has uncovered the core components of the archaean DNA replication and repair pathways, and how they relate to and regulate DNA polymerases, many questions remain. Current advances in archaean genetic manipulation and biochemical techniques allowing reconstitution and study of at least parts of replisomes are making great strides in elucidating replication and repair mechanisms. This review intends to set out the current status of DNA polymerase research in archaean DNA replication and repair, and suggests potential future directions for the field.

**Archaeal DNA polymerases at the replication fork**

DNA replication is initiated at replication origins following recruitment of specific proteins [32] such as Orc1/Cdc6 [33] and MCM [34]. DNA polymerases are subsequently recruited and interact with additional enzymes and structural proteins to form the replicase holoenzyme, the multi-protein complex responsible for the semi-discontinuous synthesis of both leading and lagging strands of DNA. Although well characterised in bacteria and eukaryotes [35], little is known of the proteins that support the archaean DNA polymerase holoenzyme. New advances in biochemistry, structural biology, and archaean genetics in both the Eury- and Crenarchaeota are helping to fill in many of the missing gaps in our knowledge [36,37].

**Euryarchaeotal replication: current status**

The Euryarchaeota is a diverse archaean phylum encompassing halophiles, methanogens and hyperthermophiles, with multiple model species being used to study replication [36]. At first
glance studying multiple systems may seem redundant, but this is useful considering the evolutionary distance and genetic and phenotypic diversity exhibited by this phylum. Although methanogen and halophile systems have been available for some time, tractable genetic systems have only recently become available for Thermococcus kodakarensis [38], of particular importance as much prior biochemical analysis of euryarchaeal replication focused on related Thermococcus and Pyrococcus species.

Analysis of the euryarchaeal replication fork suggests a coordinated action of PolD and PolB centred around the homotrimeric PCNA ‘sliding clamp’ or processivity factor (Figure 2A) [39]. PolB interacts with a wide range of replication fork components in addition to PCNA, including RPA and primase [40]. PolD from both Pyrococcus abyssi and T. kodakarensis associates with the GINS complex [27,40], and TkoPolD also binds to DNA ligase [27]. Genetic knockout analysis reveals TkoPolB is not essential for viability but provides resistance to UV irradiation, but the two TkoPolD subunits are absolutely required in the Euryarchaeota and hence, that PolD is the replicative polymerase [41]. Similarly, PolD but not PolB was essential for the distantly related methanogen, Methanococcus maripalidus [42]. However, both the chromosomally-encoded PolB1 and PolD were essential in Halobacterium sp. NRC-1, whilst the plasmid-encoded PolB2 was dispensable [43].

This conflicting genetic data suggests either that PolD is the sole replicative DNA polymerase (i.e. the DNA polymerase responsible for synthesising the bulk of both leading and lagging strands in the genome), or that PolB in fact could be a replicative/leading strand polymerase, but functionally redundant with PolD. However proofreading-deficient TkoPolB mutants do not exhibit increased mutation rates in vivo, arguing against this [41]. This also suggests an additional role for PolB in DNA repair, or in another as yet undetermined but essential process in the Halobacteriales. Recent pre-steady state kinetic data support a role for PolD as a replicative DNA polymerase, as Thermococcus sp. 9’N PolD follows a similar polymerisation scheme to other DNA polymerases, albeit with a very slow polymerisation rate and higher error rate than other typical replicative polymerases [44]. However as the authors of this study point out, some PolD kinetic parameters and fidelity could be dependent on additional replisome components in vivo, and furthermore, it is unclear if the PolD 3’-5’ exonuclease preference of Mn^{2+} over Mg^{2+} influences PolD fidelity in vivo.

P. abyssi PCNA (PabPCNA) is observed to bind PabPolB, requiring a C-terminal PIP (PCNA-interacting protein) box motif [45], and both N- and C-terminal PIP boxes for PabPolD [46]. A PCNA interaction was not observed however for T. kodakarensis PolB (TkoPolB) from affinity pulldowns [27], although this may reflect a similar requirement for primed DNA to potentiate such physical interactions as with PabPCNA [47]. Both PabPolD and PabPolB can extend DNA primers, with PabPolD requiring PabPCNA for efficient DNA synthesis. PabPolD conversely exhibits DNA strand displacement activity in the absence of PabPCNA, but PabPolB requires PabPCNA for this activity [47,48], with strand displacement activity inhibited for both polymerases in P. furiosus by the HPfA1 histone protein [49]. Only PabPolD can extend RNA primers, with RNA strand displacement only occurring for PabPCNA-stimulated PabPolD [47], suggesting a role for PolD in initial extension of primed templates on both leading and lagging strands. Furthermore PabPolB is observed to displace PabPolD from DNA but as PabPolB is inhibited by downstream RNA primers [50], this data suggests displacement is likely to occur only on the leading strand. This suggested PabPolD
initially extends the leading strand RNA primer, prior to switching to PabPolB/PabPCNA for processive synthesis, with PabPolD/PabPCNA involved in lagging strand Okazaki fragment primer extension and maturation.

Recent multiplex capillary electrophoresis analysis with *Thermococcus* sp. 9°N proteins however indicates a requirement for PolB in Okazaki fragment maturation (Figure 2B), as PolD stops 4 nucleotides before downstream Okazaki fragments, even in the presence of PCNA and other replication factors [51]. PolB then fills in this gap, with its strand displacement activity creating a flap structure subsequently processed by FEN1 and DNA ligase, supporting the observation of PabPolB displacing PabPolD already discussed [50]. This begs the question of how PolB can be dispensible for growth following the knockout studies already described [41,42]. It is feasible that the residual full Okazaki processing observed for *Thermococcus* sp. 9°N in the absence of PolB (~5%), could result from the weak strand displacement activity of PolD (also seen for PabPolD [47]), and this could potentially be sufficient to sustain this in PolB deletion mutants *in vivo* [51].

**Crenarchaeotal replication: current status**

Although an evolutionarily broad phylum, the Crenarchaeota are typically thermoacidophiles, with perhaps the most well characterised being the *Sulfolobus* genus [52]. *Sulfolobus solfataricus* is the most biochemically characterised crenarchaeote in terms of replication, but *S. acidocaldarius* [53,54] and *S. islandicus* [55] have offered more tractable genetic systems. *Sulfolobus* present a particularly excellent model system for studying DNA archaeal replication, as apart from extremely powerful genetics, their synchronisable cell cycle is the most understood in archaea [56] and their proteins are generally readily expressed in bacterial systems [57]. Furthermore, significant advances have been made in imaging sub-cellular localisation of replication components and nascent DNA using the incorporation of fluorescent nucleoside analogues [58].

The Family B polymerase member Dpo1 has been the major focus for crenarchaeotal DNA polymerases since its first isolation from *S. acidocaldarius* [59]. It was proposed as the crenarchaeal replicase following its association with other core replication factors such as the heterotrimeric SsoPCNA (Figure 2C) [60]. Furthermore, SsoRFC physically interacts with SsoDpo1 and stimulates both DNA polymerase and 3′-5′ exonuclease activities [61] by facilitating SsoDpo1 recruitment to DNA. Orc1/Cdc6 origin initiators also interact with SsoDpo1, stimulating SsoDpo1 DNA binding but inhibiting polymerase activity, with SsoCdc6-1/2 inhibiting 3′-5′ exonuclease activity [62]. Moreover, only Dpo1 is required for replication by the rudivirus SIRV2 following knockout studies during *S. islandicus* infection [63], suggesting viral subversion of the Dpo1 cellular replicase.

The lagging strand maturation ability of *S. solfataricus* (Figure 2D) was demonstrated from the observations that PCNA1 binds FEN1 endonuclease, PCNA2 binds Dpo1 and PCNA3 binds Lig1 ligase, although PCNA3 also weakly binds FEN1 and Dpo1 [60]. An *in vitro* Okazaki fragment maturation system was subsequently reconstituted, comprising the PCNA1-2-3 heterotrimer, Dpo1, FEN1 and Lig1 alone [64]. This study demonstrated PCNA-stimulated lagging strand DNA synthesis and RNA primer strand displacement by SsoDpo1, creating a flap structure which could then be cleaved by FEN1, followed by covalent ligation by Lig1 of the upstream and downstream Okazaki fragments [64]. Use of a PCNA1-2-3 fused protein
confirmed the ‘molecular toolbelt’ model, where Dpo1, FEN1 and Lig1 are simultaneously coordinated around a single PCNA heterotrimer molecule, with different PCNA subunits driving assembly of multiprotein complexes [64], in contrast to the homotrimeric PCNAs of the Euryarchaeota, which more closely resemble eukaryotic PCNA.

Electron microscopy studies confirm this simultaneous engagement of PCNA by the three separate proteins on separate subunits, all positioned on the front face of PCNA [65], supporting previous P. furiosus PCNA-polymerase EM studies [66]. The polymerase contacts PCNA extensively in two regions [65-67], with the PfuPolB C-terminal PIP box directing PCNA interaction [67]. As with the euryarchaeotal PolB polymerases, a C-terminal PIP box was found in SsoDpo1 that was essential for the interaction with SsoPCNA2 [68]. However, this is in addition to the originally identified N-terminal PIP box shown to be important for PCNA2/3 binding [60], suggesting a situation instead similar to the euryarchaeotal PolD family, requiring both N- and C-terminal PIP boxes for PCNA binding [46]. Small-angle X-ray scattering studies support binding of the SsoDpo1 C-terminal PIP box to PCNA2 in the holoenzyme complex [68], similar to that seen for PfuPolB [67].

Surprisingly, the Dpo1-PCNA1-2-3 holoenzyme supports only distributive rather than processive DNA synthesis, with the SsoDpo1 C-terminal PIP box essential for continual recruitment and exchange of Dpo1 with PCNA2 in the holoenzyme [68]. DNA polymerase exchange is also essential to preventing blockage of the replication fork by DNA lesions in bacterial [69] and eukaryotic systems [70]. Translesion (TLS) polymerases are exchanged with a replicative polymerase from the holoenzyme, allowing lesion bypass followed by switching back to a replicative polymerase. As described, Crenarchaeota encode a Y family TLS polymerase (Dpo4/PolY), with SsoDpo4 and SsoDpo1 physically interacting [71]. In addition to the independent Dpo1-PCNA2 and Dpo4-PCNA1 interactions [72], an additional hydrophobic region of Dpo1 behind the palm subdomain contacts Dpo4, allowing recruitment and stabilisation of an the Dpo1-Dpo4-PCNA1-2-3 holoenzyme and thereby increasing DNA synthesis processivity [73]. Direct contacts between the bacterial Pol III replicative and Pol IV TLS polymerases appear important for polymerase switching [73] with the Dpo1-Dpo4 interface appearing conserved [74], suggesting archaea too exhibit DNA polymerase switching from replicative to translesion synthesis.

DNA polymerase dynamics also appear concentration and temperature dependent, as SsoDpo1 trimerises in the presence of DNA in vitro which increases both polymerase kinetic rate and processivity (Figure 2Cvii) [75]. Such oligomerisation is not unusual in the absence of accessory factors, as some bacterial [76] and eukaryotic DNA polymerases [77] have been noted to have stoichiometries greater than one. SsoDpo1 oligomerisation is suggested to also assist in polymerase/exonuclease switching during proofreading [75]. SsoDpo4 itself dimerises at higher and physiological temperatures [78], however, oligomeric Dpo1 also binds undamaged DNA more strongly than Dpo4 at physiological temperatures, suggesting thermodynamic regulation of replisome assembly. Hence, local polymerase concentration and thermodynamic-mediated oligomerisation could work in tandem with PCNA-mediated replisome assembly for both Dpo1 and Dpo4 switching determining replisome composition over normal or damaged templates [73], however, Dpo1/Dpo4 oligomerisation requires validation in vivo.
Beyond PCNA acting as a nexus at the replication fork where it coordinates binding of polymerases Dpo1 and Dpo4, proteins involved with Okazaki fragment maturation and some DNA repair enzymes [79], little else is known of the composition of the crenarchaeotyal replisome, or if structural components exist analogous to the tau subunit that brings three Pol III units together in the bacterial replisome [35]. Recent studies however using epitope-tagged proteins in *S. acidocaldarius* have elucidated two small proteins that interact with *Sulfolobus* Dpo1, so Dpo1 should be considered as a stable heterotrimer [80]. These Dpo1 (polB1)-binding proteins (PBP1, PBP2) significantly increase the thermostability of SsoDpo1. PBP1 negatively influences Dpo1 strand displacement activity *in vitro* and encourages Dpo1 release on encountering downstream Okazaki fragments, thereby reducing inefficient DNA re-synthesis. PBP2 enhances DNA synthesis and moderates PBP1 influence on Dpo1 distributive synthesis. Hence crenarchaeotyal replicative DNA polymerases may be analogous to the replicative eukaryotic polymerases Polδ and Polε, requiring small accessory proteins as part of their overall structure [81,82]. However as PBP1 and PBP2 are absent from the Thermoproteales (Figure 1) [80], it is feasible that this is not a universal observation for the Crenarchaeota, or that functionally analogous proteins to PBP1 and PBP2 are yet to be found in these species. Effects of PBP1 or PBP2 on the propensity for Dpo1 oligomerisation are yet to be demonstrated however.

**Archaeal DNA polymerases in DNA repair – unanswered questions**

Archaea require DNA polymerases for most repair pathways, likely filling in gaps generated following lesion removal [3]. Replicative DNA polymerases however may encounter lesions prior to their repair, potentially stalling due to their active sites being unable to accommodate bulky lesions [83]. Such discrimination is well studied for archaean Family B polymerases, with the loop region of the fingers determining accurate genome replicating or error-prone TLS functionality [84]. Replicative polymerases may instead exchange with (often Y-family) TLS polymerases on encountering lesions, as TLS polymerases are better suited for dealing with lesions due to their less spatially-constrained active sites, resulting often in error-free lesion bypass [85].

Family Y DNA polymerases are found in all bacteria and eukaryotes [86], but exhibit a varying phyletic distribution in archaea, presumably following significant lateral gene transfer and deletion events (Figure 1). *S. solfataricus* Dpo4 [87] is an extremely well-characterised and amenable Y-family model DNA polymerase for biochemical, kinetic and structural studies [88]. SsoDpo4 and its *S. acidocaldarius* orthologue Dbh [89] are demonstrated to be recruited to lesions *in vivo* to avoid their incorrect replication [90]. Although SsoDpo4 is exchanged with Dpo1 on PCNA as discussed, little is known of the process of Dpo4 recruitment or if a mechanism similar to recruitment via PCNA mono-ubiquitlylation in eukaryotes is possible in archaea [90]. It is also unclear how lesion bypass could occur in those archaea lacking a Family Y polymerases. Recent reports however demonstrate that euryarchaeotyal primases are able to bypass different lesions to rescue stalled replication forks [91]. It is unknown if primases are specifically recruited for this purpose, or remain in close association with DNA polymerases throughout the replication cycle [40]. Furthermore, homologous recombination or fork regression mechanisms could also potentially indirectly mediate polymerase lesion bypass [92].
Crenarchaeotal Dpo2 and Dpo3 – mysterious repair polymerases?

Although Dpo1 and Dpo3 are ubiquitous in the Crenarchaeota, phylogenetic analysis suggests Dpo2 presence is widespread, but absent in some Crenarchaeota [15]. Despite initial expectations that Dpo2 and Dpo3 would be inactive from lacking certain otherwise conserved catalytic residues [93], similar to Dpo1 and Dpo4, both Dpo2 and Dpo3 were expressed at the RNA and protein level in S. solfataricus [57]. Moreover, they exhibit DNA polymerase and 3′-5′ exonuclease activities [57,94,95], indicating they are bona fide proteins and not pseudogenes. In contrast to SsoDpo1 and SsoDpo4 however, SsoDpo2 and SsoDpo3 bind DNA weakly with limited thermostability [57], although other accounts suggest a higher thermostability for Dpo3 [94]. SsoRFC and SsoPCNA together increased primer extension with SsoDpo1/SsoDpo4, but only weakly for SsoDpo2/SsoDpo3 with no obvious PIP boxes for PCNA interaction observed for the latter two [57]. Further, addition of SsoSSB strongly enhanced primer extension with SsoDpo1 and SsoDpo4 but conversely strongly inhibited SsoDpo2 and SsoDpo3 [57]. This accumulated evidence suggests a major role as replicative DNA polymerases is unlikely for Dpo2 and Dpo3.

As with SsoDpo4, SsoDpo2/Dpo3 however both exhibit significant lesion bypass capabilities and are able to bypass 8-oxoG and hypoxanthine lesions [57]. SsoDpo2 can also bypass uracil lesions and interestingly, SsoDpo3 but not SsoDpo2 can bypass bulky cyclobutane thymine dimer lesions [57]. This is surprising as SsoDpo2 is upregulated by UV damage [96,97], suggesting an unclear function for SsoDpo2 in processing UV-induced lesions. Dpo2 was also observed to be upregulated in S. islandicus by the Orc1-2 DNA damage response regulator following exposure to NQO, a DNA adduct forming compound [98]. Furthermore, Dpo2 and Dpo3 may also be involved in BER, as both can synthesise across gaps during in vitro P. aerophilum BER [99]. Hence, significant evidence points to an involvement for Dpo2 in DNA repair.

Recent reports suggest a further role for Dpo2 in de novo spacer acquisition during CRISPR-mediated immunity. Both Dpo2 and Dpo4 were upregulated in S. islandicus, following spacer acquisition resulting from either infection with the STSV2 monocaudavirus [100], or following overexpression of the Csa3a CRISPR transcriptional regulator [101]. Spacer acquisition is dependent on DNA replication in S. islandicus [102], suggesting an active involvement of Dpo2 in this process. The homologous recombination-related NurA and HerA genes were also upregulated under the same conditions [100,101], potentially implicating Dpo2 in double-strand break repair (at least in the context of CRISPR), potentially similar to that seen for Pol I in E. coli spacer acquisition [103].

Hence, it appears that Dpo2 and Dpo3 are functionally related, but are unlikely to play a role in replicative DNA synthesis from their non-optimal characteristics. They may instead act as specialised repair polymerases, most likely switching with Dpo1 on encountering specific lesions, potentially being redundant with Dpo4 TLS polymerase. Future genetic analyses on Dpo2/3 should shed further light on their specific roles in genome integrity.

Roles of archaeal DNA polymerases in uracil recognition

Potentially mutagenic deaminated bases in DNA are a prevalent risk for thermophilic archaea in particular [2]. Both euryarchaeotal [104] and crenarchaeotal [105] replicative
DNA polymerases are found to stall at uracil and hypoxanthine bases [104], although this is not observed for those from bacteria or eukaryotes [106]. This is most prominent for PolB polymerases, stopping at 4 bases prior to the deaminated base where it is tightly bound in an N-terminal pocket [14,107]. Euryarchaeotal PolD polymerases are also inhibited by uracil [108] and more weakly by hypoxanthine [109], potentially by a different mechanism from PolB. Instead, PolD polymerases slow down DNA synthesis rather than stopping at a defined site upstream of the deaminated base, and inhibition may also be mediated by a deaminated base in trans on the non-replicated strand [108,109].

DNA polymerase stalling at these lesions is thought to allow recruitment of BER proteins, preventing the lesion being copied and fixed into the genome as a mutation. Indeed, crosstalk between stalled PolB and PolD polymerases and BER components is observed, although this counter-intuitively inhibits uracil-DNA glycosylase, EndoV and EndoQ [110,111], most likely resulting from the tight binding of uracil preventing access by BER enzymes [111]. This shutdown of BER by sequestering lesions could potentially allow replication fork regression/reversal to form a ‘chicken foot’ structure [92], thereby restoring the deaminated base across from its correct cognate base. This would allow accurate BER repair potentially using PolB to synthesis across the gap [109]. As described previously however, the situation may be even more complex as euryarchaeotal primases were found to bypass not only oxidative and UV damage-induced lesions, but also can replicate past uracil lesions. This may potentially rescue PolB-stalled replication forks [91], suggesting primase-replicase cross-talk may function to ensure replication proceeds.

Conclusions and remaining questions

For many years, research on archaeal DNA polymerases focused on their biochemical mechanisms and biotechnological exploitation [5]. Recent advances however in biochemical techniques and genetics have yielded deeper insights into the biology of archaeal DNA polymerases and associated proteins in genome replication and repair [36,37]. In particular, establishment of in vitro systems reconstituting Okazaki fragment on the lagging strand [51,64] and in situ epitope tagging to elucidate additional replisome components [80]. A number of outstanding questions still remain however, in particular, the role of Dpo1 oligomerisation and the regulation of the Dpo4 polymerase switch in Crenarchaeota, and the roles of Dpo2/3 in DNA repair and CRISPR. Although only briefly referred to here, the functional relationship of DNA polymerases with primases certainly deserves greater attention, as although archaeal primases comprise an active area of research [25,112], the switch from primases to polymerases during the replication cycle and the relationship to lesion bypass is still unclear [91].

A further enduring question is whether archaeal DNA polymerases are post-translationally modified to regulate their activities and interactions. Both bacterial [113] and eukaryotic [114] DNA polymerases may be modified, and other archaeal genome integrity factors are observed to undergo regulation by post-translational modification. Notably, in Sulfolobus lysine methylation enhances MCM helicase activity [115], Mre11-Rad50 methylation follows DNA damage [116], and the chromatin protein Alba’s DNA binding affinity is regulated by acetylation [117]. E1/E2/E3 ubiquitin-like modification pathways have been discovered in
other archaeal species and potentially associated with protein turnover [118,119]. Hence, it would be interesting to assess if any such modifications play similar roles in archaea, as shown for ubiquitin-mediated polymerase switching at the eukaryotic replication fork [90].

This review has focused on archaeal DNA polymerases in a cellular context, however, many archaeal mobile DNA elements and viruses encode poorly characterised DNA polymerases. Mobilisation through horizontal gene transfer has been proposed to drive DNA polymerase evolution in diverse archaeal phyla and eukaryotes [15,120]. Although the ruddivirus SirV2 subverts the \textit{S. solfataricus} Dpo1 polymerase for viral replication, viruses encoding DNA polymerases are seen across the archaeal spectrum \textit{e.g.} infecting halophiles [121,122], thermophiles [123] and methanogens [124]. Such polymerases are not necessarily thought to be active as missing key catalytic motifs, but as has been seen previously for Dpo2/Dpo3 [93], activity may still be present [57] and their presence suggests a role in evolutionary fitness or infection. DNA polymerases found in other genetic elements include the AEP (archaeo-eukaryotic primase) family found in archaeal plasmids [125] and the arCOG04926 Family B polymerases contained within CRISPR-associated archaeal Casposons [15,126]. Characterisation of such novel non-cellular DNA polymerases will shed light not only on archaeal virus/plasmid replication and DNA polymerase evolution, but could potentially further drive the development of novel biotechnological reagents. The future is certainly bright for archaeal DNA polymerase research.

\textbf{Summary}

\begin{itemize}
  \item Advances in biochemical analysis, structural biology and model system genetics have been pivotal in understanding the functional roles of archaeal DNA polymerases.
  \item Knowledge of the archaeal replisome protein complement is improving, but more work is needed to further understand DNA polymerase recruitment and regulation.
  \item Roles of euryarchaeotal PolB/PolD are becoming understood, but their relative function in leading/lagging strand synthesis is still not completely clear.
  \item Full \textit{in vivo} roles of crenarchaeotal DNA polymerase oligomerisation are unknown.
  \item Functions of Dpo2 and Dpo3 in crenarchaeota need to be fully established.
\end{itemize}

\textbf{Abbreviations}


\textbf{Acknowledgements}

The author is grateful for funding from the University of Huddersfield.

\textbf{Competing interests}

The author declares no competing interests associated with this study.
Figure legends

Figure 1. Phyletic distribution of archaeal DNA polymerases
Distribution of DNA polymerases and associated structural/regulatory subunits in representative archaeal genomes. Filled circles represent gene presence with colour reflecting the respective text for species. Letters in the first column reflect phylum (K, Korarchaeota; C, Crenarchaeota; A, Aigarchaeota; T, Thaumarchaeota; N, Nanoarchaeota; E, Euryarchaeota). Species text colouring reflects taxonomic order, apart from for the Thaumarchaeota. Numbers beside filled circles represent number of paralogue copies found in genomes. Filled circles positioned between two columns represent fused genes. Phyletic data retrieved from Raymann et al. [16], Makarova et al. [15], Yan et al. (PBP1/2) [80] and Liu et al. (PriX) [23]. Dpo4/poiY distributions were determined by the author using BLASTP homology searches [127], querying S. solfataricus P2 Dpo4 (AAK42588.1) against the non-redundant NCBI database. BLAST hits were compared against the CDD database [128] to confirm identity as Family Y polymerases.

Figure 2. Current models of archaeal replication fork enzymology
Key denotes proteins, X denotes DNA lesions and transparency/dotted outlines represent unconfirmed, tentative or speculative observations. Double-headed dotted arrows represent potentially reversible/dynamic binding interactions. Proteins are not to scale. A. Euryarchaeota leading strand replication. (i) RFC and homotrimeric PCNA are recruited to primed template DNA at replication fork, similar to 2Ci. (ii) PolD binds PCNA. (iii) PolD synthesises DNA from RNA primed template. PolB may potentially be used for continued DNA synthesis from DNA primer, but is redundant in some species. (iv) On encountering DNA lesions (X), DNA primase (PriL/PriS) may bypass the lesion, before DNA polymerase continues elongation. B. Euryarchaeota lagging strand replication. (i) PolD binds to PCNA and elongates from an RNA primer in an Okazaki fragment, but stops 4 nucleotides before the downstream Okazaki fragment. (ii) PolB displaces PolD, synthesising the remaining DNA and displacing the RNA primer. (iii) FEN1 cleaves the displaced RNA/DNA ‘flap’ and PolB synthesises across the gap. (iv) DNA ligase covalently seals the nick. C. Crenarchaeota leading strand replication. (i) and (ii) Heterotrimeric PCNA is loaded onto primed DNA by RFC, requiring ATP hydrolysis. (iii) Dpo1 (heterotrimeric, with PBP1/PBP2 subunits) is recruited to PCNA (PCNA2). (iv) Dpo1 distributively synthesises DNA but RFC can stimulate Dpo1, although it is not clear if RFC rebinds. If the fork encounters lesions (X), Dpo4 is recruited to PCNA1 and can act potentially as a dimer. (v) Dpo4 may bypass the lesion, but the roles of additional translesion Dpo2 and Dpo3 are not fully established. (vi) Dpo1 continues to synthesise once lesion is bypassed, but it is not clear if Dpo4 leaves the holoenzyme or has a more stable association. (vii) Dpo1 may trimerise, increasing activity and processivity (red arrow), although it is not yet established in vivo or if PBP1/PBP2 or PCNA are present on Dpo1 trimerisation. D. Crenarchaeota lagging strand replication. (i) Heterotrimeric Dpo1 binds PCNA2, FEN1 endonuclease to PCNA1 and LIG1 DNA ligase to PCNA3 in a ‘molecular toolbelt’. (ii) Dpo1 synthesises from an RNA primer and displaces the downstream RNA primer, creating an RNA/DNA ‘flap’. (iii) FEN1 cleaves the ‘flap’ and once Dpo1 has synthesised to the downstream DNA, LIG1 covalently seals the nick.


Bailey, M.F., Van der Schans, E.J. and Millar, D.P. (2007) Dimerization of the Klenow fragment of Escherichia coli DNA polymerase I is linked to its mode of DNA binding. Biochemistry. 46, 8085-8099


Processing of DNA lesions by archaean DNA polymerases from Sulfolobus solfataricus.
Nucleic Acids Res. 31, 4024-4030

Uracil recognition by replicative DNA polymerases is limited to the archaea, not occurring with bacteria and eukarya.
Nucleic Acids Res. 36, 705-711

Structural basis for uracil recognition by archaeal family B DNA polymerases.
Nucleic Acids Res. 38, 705-711

Novel inhibition of archaeal family-D DNA polymerase by uracil.
Nucleic Acids Res. 41, 4207-4218

Archaeal DNA Polymerase D: A Zinc-Binding Protein Inhibited by Hypoxanthine and Uracil.
J Mol Biol. 428, 2805-2813

Interplay between DNA polymerase and proliferating cell nuclear antigen switches off base excision repair of uracil and hypoxanthine during replication in archaea.
J Mol Biol. 383, 762-771

Archaeal DNA Polymerase-B as a DNA Template Guardian: Links between Polymerases and Base/Alternative Excision Repair Enzymes in Handling the Deaminated Bases Uracil and Hypoxanthine.
Archaea. 2016, 1510938

Primer synthesis by a eukaryotic-like archaeal primase is independent of its Fe-S cluster.
Nat Commun. 8, 1718

Posttranslational modification of the umuD-encoded subunit of Escherichia coli DNA polymerase V regulates its interactions with the beta processivity clamp.
Proc Natl Acad Sci U S A. 99, 5307-5312

Arginine methylation regulates DNA polymerase beta.
Mol Cell. 22, 51-62

The Helicase Activity of Hyperthermophilic Archaeal MCM is Enhanced at High Temperatures by Lysine Methylation.
Front Microbiol. 6, 1247

Post-translational methylations of the archaeal Mre11:Rad50 complex throughout the DNA damage response.
Mol Microbiol. 100, 362-378

The interaction of Alba, a conserved archaeal chromatin protein, with Sir2 and its regulation by acetylation.
Science. 296, 148-151

Functional reconstruction of a eukaryotic-like E1/E2/(RING) E3 ubiquitylation cascade from an uncultured archaeon.
Nat Commun. 8, 1120

Involvement of a eukaryotic-like ubiquitin-related modifier in the proteasome pathway of the archaeon Sulfolobus acidocaldarius.
Nat Commun. 6, 8163

Evolution of Eukaryotic DNA Polymerases via Interaction Between Cells and Large DNA Viruses.
J Mol Evol. 81, 24-33

122 Bath, C., Cukalac, T., Porter, K. and Dyall-Smith, M.L. (2006) His1 and His2 are distantly related, spindle-shaped haloviruses belonging to the novel virus group, Salterprovirus. Virology. 350, 228-239


<table>
<thead>
<tr>
<th>Phylum</th>
<th>Order</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td></td>
<td>Candidatus Korarchaeum cryptofilum</td>
</tr>
<tr>
<td>C</td>
<td>Sulfolobales</td>
<td>Sulfolobus solfataricus</td>
</tr>
<tr>
<td>C</td>
<td>Sulfolobales</td>
<td>Sulfolobus islandicus</td>
</tr>
<tr>
<td>C</td>
<td>Sulfolobales</td>
<td>Acidiamonas hospitalis</td>
</tr>
<tr>
<td>C</td>
<td>Sulfolobales</td>
<td>Metallosphaera sedula</td>
</tr>
<tr>
<td>C</td>
<td>Sulfolobales</td>
<td>Sulfolobus tokodaii</td>
</tr>
<tr>
<td>C</td>
<td>Desulfurococcales</td>
<td>Ignicoccus hospitalis</td>
</tr>
<tr>
<td>C</td>
<td>Desulfurococcales</td>
<td>Hyperthermus butylicus</td>
</tr>
<tr>
<td>C</td>
<td>Desulfurococcales</td>
<td>Pyrolobus fumari</td>
</tr>
<tr>
<td>C</td>
<td>Desulfurococcales</td>
<td>Staphylothermus marinus</td>
</tr>
<tr>
<td>C</td>
<td>Desulfurococcales</td>
<td>Thermus marinus</td>
</tr>
<tr>
<td>C</td>
<td>Desulfurococcales</td>
<td>Aeropyrum pernix</td>
</tr>
<tr>
<td>C</td>
<td>Acidilobales</td>
<td>Acidilobus saccharovorans</td>
</tr>
<tr>
<td>C</td>
<td>Thermoproteales</td>
<td>Thermophilum pendens</td>
</tr>
<tr>
<td>C</td>
<td>Thermoproteales</td>
<td>Thermosphaera arsenaticum</td>
</tr>
<tr>
<td>C</td>
<td>Thermoproteales</td>
<td>Thermoactinemonas aerophilum</td>
</tr>
<tr>
<td>C</td>
<td>Thermoproteales</td>
<td>Thermoproteus tenax</td>
</tr>
<tr>
<td>C</td>
<td>Thermoproteales</td>
<td>Caldivigia maquilagensis</td>
</tr>
<tr>
<td>C</td>
<td>Thermoproteales</td>
<td>Vulcansilvaetra distributa</td>
</tr>
<tr>
<td>A</td>
<td></td>
<td>Candidatus Caldarchaeum subterraneum</td>
</tr>
<tr>
<td>T</td>
<td>Nitrososphaerales</td>
<td>Candidatus Nitrososphaera gargentis</td>
</tr>
<tr>
<td>T</td>
<td>Cenarchaeales</td>
<td>Cenarchaeum symbiosum</td>
</tr>
<tr>
<td>T</td>
<td>Nitrosopumilales</td>
<td>Nitrosopumilus maritimus</td>
</tr>
<tr>
<td>N</td>
<td>Nanarchaeales</td>
<td>Nanarchaeum equitans</td>
</tr>
<tr>
<td>E</td>
<td>DHVE2 group</td>
<td>Acidilobus profundum boonei</td>
</tr>
<tr>
<td>E</td>
<td>Thermoplasmales</td>
<td>Thermoplasma acidophilum</td>
</tr>
<tr>
<td>E</td>
<td>Thermoplasmales</td>
<td>Thermoplasma volcanium</td>
</tr>
<tr>
<td>E</td>
<td>Thermoplasmales</td>
<td>Ferroplasma acidarmanus</td>
</tr>
<tr>
<td>E</td>
<td>Thermoplasmales</td>
<td>Picrophicus torridus</td>
</tr>
<tr>
<td>E</td>
<td>Archaeoglobales</td>
<td>Archaeoglobus fulgidus</td>
</tr>
<tr>
<td>E</td>
<td>Archaeoglobales</td>
<td>Archaeoglobus profundus</td>
</tr>
<tr>
<td>E</td>
<td>Archaeoglobales</td>
<td>Ferroglobus placidus</td>
</tr>
<tr>
<td>E</td>
<td>Halobacteriales</td>
<td>Halobacterium volcanii</td>
</tr>
<tr>
<td>E</td>
<td>Halobacteriales</td>
<td>Halobacterium borinquense</td>
</tr>
<tr>
<td>E</td>
<td>Halobacteriales</td>
<td>Halobacterium walsbyi</td>
</tr>
<tr>
<td>E</td>
<td>Halobacteriales</td>
<td>Halobacterium salinarum</td>
</tr>
<tr>
<td>E</td>
<td>Halobacteriales</td>
<td>Halobacterium steigeri</td>
</tr>
<tr>
<td>E</td>
<td>Halobacteriales</td>
<td>Halobacterium ruber</td>
</tr>
<tr>
<td>E</td>
<td>Halobacteriales</td>
<td>Natrallia magadi</td>
</tr>
<tr>
<td>E</td>
<td>Halobacteriales</td>
<td>Haloterrigena turkenica</td>
</tr>
<tr>
<td>E</td>
<td>Halobacteriales</td>
<td>Natriniema pullirubrum</td>
</tr>
<tr>
<td>E</td>
<td>Halobacteriales</td>
<td>Halopigier xantheniensis</td>
</tr>
<tr>
<td>E</td>
<td>Halobacteriales</td>
<td>Natronoebacterium gregory</td>
</tr>
<tr>
<td>E</td>
<td>Halobacteriales</td>
<td>Natrononasomas pharaonis</td>
</tr>
<tr>
<td>E</td>
<td>Halobacteriales</td>
<td>Halohabibus tiamatea</td>
</tr>
<tr>
<td>E</td>
<td>Halobacteriales</td>
<td>Halobacterium utahensis</td>
</tr>
<tr>
<td>E</td>
<td>Halobacteriales</td>
<td>Halobacterium mukohataei</td>
</tr>
<tr>
<td>E</td>
<td>Halobacteriales</td>
<td>Halobacterium hispanicana</td>
</tr>
<tr>
<td>E</td>
<td>Halobacteriales</td>
<td>Halobacterium marismortai</td>
</tr>
<tr>
<td>E</td>
<td>Methanomicrobiales</td>
<td>Methanococcus polymorphum labreanum</td>
</tr>
<tr>
<td>E</td>
<td>Methanomicrobiales</td>
<td>Methanococcus bergensis</td>
</tr>
<tr>
<td>E</td>
<td>Methanomicrobiales</td>
<td>Methanoplanus petrolearius</td>
</tr>
<tr>
<td>E</td>
<td>Methanomicrobiales</td>
<td>Methanosphaera palustris</td>
</tr>
<tr>
<td>E</td>
<td>Methanomicrobiales</td>
<td>Methanoregula boonei</td>
</tr>
<tr>
<td>E</td>
<td>Methanomicrobiales</td>
<td>Methanospirillum hungatei</td>
</tr>
<tr>
<td>E</td>
<td>Methanocellales</td>
<td>Methanocella paludicola</td>
</tr>
<tr>
<td>E</td>
<td>Methanosarcinales</td>
<td>Methanoseta harundinacea</td>
</tr>
<tr>
<td>E</td>
<td>Methanosarcinales</td>
<td>Methanoseta concilii</td>
</tr>
<tr>
<td>E</td>
<td>Methanosarcinales</td>
<td>Methanoseta thermophilosa</td>
</tr>
<tr>
<td>E</td>
<td>Methanosarcinales</td>
<td>Methanosarcina barcki</td>
</tr>
<tr>
<td>E</td>
<td>Methanosarcinales</td>
<td>Methanosarcina acetivorans</td>
</tr>
<tr>
<td>E</td>
<td>Methanosarcinales</td>
<td>Methanosarcina mazi</td>
</tr>
<tr>
<td>E</td>
<td>Methanosarcinales</td>
<td>Methanococcoides barnanti</td>
</tr>
<tr>
<td>E</td>
<td>Methanococcales</td>
<td>Methanocaldococcus inferni</td>
</tr>
<tr>
<td>E</td>
<td>Methanococcales</td>
<td>Methanocaldococcus jannaschi</td>
</tr>
<tr>
<td>E</td>
<td>Methanococcales</td>
<td>Methanothermococcus okinawensis</td>
</tr>
<tr>
<td>E</td>
<td>Methanococcales</td>
<td>Methanococcus voitae</td>
</tr>
<tr>
<td>E</td>
<td>Methanococcules</td>
<td>Methanococcus maripapudis</td>
</tr>
<tr>
<td>E</td>
<td>Methanopyrales</td>
<td>Methanopyrus kandleri</td>
</tr>
<tr>
<td>E</td>
<td>Methanobacteriales</td>
<td>Methanothermus fervidus</td>
</tr>
<tr>
<td>E</td>
<td>Methanobacteriales</td>
<td>Methanosphaera stadtmanae</td>
</tr>
<tr>
<td>E</td>
<td>Methanobacteriales</td>
<td>Methanobrevisibacter smithii</td>
</tr>
<tr>
<td>E</td>
<td>Methanobacteriales</td>
<td>Methanothermococcus thermotrophicus</td>
</tr>
<tr>
<td>E</td>
<td>Thermococcales</td>
<td>Pyrococcus furiosus</td>
</tr>
<tr>
<td>E</td>
<td>Thermococcales</td>
<td>Pyrococcus abyssi</td>
</tr>
<tr>
<td>E</td>
<td>Thermococcales</td>
<td>Pyrococcus horikoshii</td>
</tr>
<tr>
<td>E</td>
<td>Thermococcales</td>
<td>Thermococcus carnarius</td>
</tr>
<tr>
<td>E</td>
<td>Thermococcales</td>
<td>Thermococcus kodakarensis</td>
</tr>
<tr>
<td>E</td>
<td>Thermococcales</td>
<td>Thermococcus barophilus</td>
</tr>
<tr>
<td>E</td>
<td>Thermococcales</td>
<td>Thermococcus litoralis</td>
</tr>
</tbody>
</table>