DNA Repair in the Archaea – an emerging picture

Malcolm F White[1] and Thorsten Allers[2]

[1] Biomedical Sciences Research Complex, School of Biology, University of St Andrews, Fife, KY16 9ST.

[2]School of Life Sciences, University of Nottingham, Nottingham NG2 2UH.

For correspondence: email mfw2@st-andrews.ac.uk or Thorsten.Allers@nottingham.ac.uk

Abstract

There has long been a fascination in the DNA Repair pathways of archaea, for two main reasons. Firstly, many archaea inhabit extreme environments where the rate of physical damage to DNA is accelerated. These archaea might reasonably be expected to have particularly robust or novel DNA repair pathways to cope with this. Secondly, the archaea have long been understood to be a lineage distinct from the bacteria, and to share a close relationship with the eukarya, particularly in their information processing systems. Recent discoveries suggest the eukarya arose from within the archaeal domain, and in particular from lineages related to the TACK superphylum and *Lokiarchaea*. Thus, archaeal DNA repair proteins and pathways can represent a useful model system. This review focuses on recent advances in our understanding of archaeal DNA repair processes including Base Excision Repair (BER), Nucleotide Excision Repair (NER), Mismatch Repair (MMR) and Double Strand Break Repair (DSBR). These advances are discussed in the context of the emerging picture of the evolution and relationship of the three domains of life.

Introduction

Although double-stranded DNA is a stable, chemically inert molecule, damage to DNA is largely unavoidable, and can have serious consequences for a cell, including mutation and death. While some level of mutation is acceptable, and indeed constitutes the raw material for evolution, high mutational load is incompatible with life. Efficient repair of DNA damage is therefore essential for all forms of life. The Archaea are no exception, and indeed they often inhabit challenging environments and are thus exposed to extremes of temperature, salinity, pressure or pH. Archaea would thus be expected to have particularly robust DNA repair pathways, and they do, but we don't yet understand them very well. As has been noted in previous reviews of the topic, there are many

enigmas in the field of archaeal DNA repair (Grogan, 1998, White, 2003, Rouillon & White, 2011, Grogan, 2015). Some of these are gradually being resolved whilst others remain stubbornly opaque. In this review, we focus on recent research that illuminates aspects of the four universal DNA repair pathways: Base Excision Repair (BER), Nucleotide Excision Repair (NER), Mismatch Repair (MMR) and Homologous Recombination / Double Strand Break Repair (HR/DSBR) (Figure 1).

The last few years have seen rapid advances in several areas. Genomics has given us vast new datasets and unveiled a diverse array of new archaeal species that are shaking our view of the tree of life (Adam *et al.*, 2017, Spang *et al.*, 2017). Genetic systems are being developed for key model organisms such as the Halophiles, Methanogens, *Sulfolobales* and *Thermococcales* that allow the increasingly-sophisticated study of archaeal gene function (Farkas *et al.*, 2013). Biochemical and structural studies are revealing mechanistic detail on individual DNA repair proteins and pathways. Used in combination, these approaches can lead to swift and significant advances in understanding. A good example is the discovery of a non-canonical Mismatch Repair pathway, based on the EndoMS nuclease, by the Ishino lab (Ishino *et al.*, 2016). This advance, described in detail below, has the potential to answer one of the major outstanding questions of the archaeal DNA Repair field.

This is a field in transition. Much of the early work on DNA replication and repair in the archaea arose from a desire to study simpler model systems of eukaryal (ultimately, human) processes. This approach led to many notable successes. However, as the need for model systems has faded, there is a growing realisation that the archaea are not a niche player in the biosphere but rather a major, significant component that deserves study in their own right. Their cellular and molecular biology is often distinct from those of the bacteria and eukarya, and this is certainly true for their DNA repair pathways.

DNA repair and the origin of the eukarya

Although still not universally agreed, the recent discovery of new archaeal lineages known collectively as the “ASGARD” archaea, which includes the species *Lokiarchaeota* and *Thorarchaeota*, have caused a reassessment of the relationship between the archaeal and eukaryal domains (reviewed in (Eme *et al.*, 2017). The large number of gene families previously thought to be specific to the eukarya that are found in ASGARD genomes has led to the suggestion that Eukarya arose from an archaeal species related to the ASGARD archaea. Other experts however disagree with this

interpretation of the data (Da Cunha *et al.*, 2017). What can the distribution of DNA repair genes across the archaea add to this hot topic (Figure 2)? If we take the example of the XPF nuclease, it comes in two "flavours" in archaea. The short version consists only of a nuclease domain, which interacts with PCNA, and is found only in the TACK superphylum (Rouillon & White, 2011). The long version has a nuclease fused to a helicase domain matching eukaryal XPF. This is present predominantly in the euryarchaea, but also in the ASGARD archaea. Similarly, a eukaryal-type Replication Protein A (RPA, a single stranded DNA binding protein) is present in most archaea with the exception of the crenarchaea and *Thermoplasma*, which have a short version (Rouillon & White, 2011). Focussing on the two examples of ASGARD archaea in Figure 2, it is apparent that *Lokiarchaea* and *Thorarchaea* have the complement of eukaryal-type repair proteins one would expect for an ancestor of the eukarya. This includes copies of the bacterial-type mismatch repair proteins MutS and MutL, which are also present throughout the eukaryal lineage. Intriguingly, the ASGARD archaea have also picked up the bacterial UvrABC NER system. Overall, the distribution pattern of DNA repair genes in the archaea, and the ASGARD lineage in particular, is consistent with the hypothesis that the latter gave rise to the eukaryal domain of life.

Mismatch Repair (MMR)

The canonical MutL-MutS pathway

Mismatch Repair (MMR) is the process by which bases incorporated in error by the DNA replication machinery are detected and corrected. The MutL-MutS MMR pathway first characterised in *E.coli* is present in most bacteria (with the notable exception of the actinobacteria) and in the eukarya, but is the exception rather than the rule in the archaea (Kelman & White, 2005). Most archaea lack plausible MutS and MutL homologues, and those that have them tend to be temperature mesophiles such as halophiles and methanogens that most likely captured these genes by lateral gene transfer from bacteria (Figure 2). The mode of inheritance of a bacterial-type MMR pathway from bacteria to the eukarya is a matter of conjecture. One possibility is that endosymbiotic event that led to the evolution of the mitochondrion from an alpha-proteobacterium allowed the bacterial genes for MMR to become established in the early eukaryal genome. An alternative possibility is that the eukarya inherited the bacterial MMR machinery via their archaeal lineage. It is notable that the ASGARD archaea including *Lokiarchaeum* and *Thorarchaeum*, which have been proposed as the most closely related extant archaea to the progenitor of the eukarya (Eme *et al.*, 2017), possess clear MutS and MutL homologues.

The emerging role of EndoMS

The lack of canonical MMR in most archaea is not reflected in high mutation rates (Grogan, 2004), and deletion of MutS-MutL in *Halobacterium salinarum* did not give rise to a hypermutation phenotype (Busch & DiRuggiero, 2010). These observations suggest that alternative pathways exist to detect and remove mismatches post DNA replication.

To search for this pathway, Ishino and colleagues devised a functional screen for enzymes capable of cleaving DNA mismatches in *Pyrococcus furiosus* (Ishino *et al.*, 2016). This resulted in the identification of an enzyme, which was named EndoMS for endonuclease mismatch-specific, capable of cleaving a range of mismatched DNAs by the introduction of staggered cleavages in both strands of the DNA, leaving 5 nt 5'-overhangs (Ishino *et al.*, 2016). EndoMS had originally been identified in the Millikallio lab and named NucS, based on its activity against single-stranded DNA (Ren *et al.*, 2009). The structure of NucS revealed a dimeric, two-domain organisation, and the enzyme was shown to form a physical interaction with the sliding clamp PCNA (Proliferating Cell Nuclear Antigen) (Ren *et al.*, 2009). As the enzyme has a much higher specificity for mismatches than for branched or ssDNA, the nomenclature "EndoMS" will be used henceforth. The recent DNA:protein co-crystal structure reveals that EndoMS wraps around mismatched DNA substrates, flipping out two bases and cleaving the DNA backbone in a manner reminiscent of type II restriction enzymes (Nakae *et al.*, 2016)(Figure 3). The enzyme is active against G-T, G-G, T-T, T-C and A-G mismatches, but not against C-C, A-C or A-A mismatches *in vitro* (Ishino *et al.*, 2016), which is consistent with higher binding affinities for substrates with a mismatched G or T (Nakae *et al.*, 2016).

EndoMS has a complex distribution in the archaea (Figure 2), with examples in the halophiles, various thermophiles from the crenarchaeal and euryarchaeal phyla, and Thorarchaeum from the ASGARD phylum. EndoMS is also present in some bacterial genomes, particularly the phylum Actinobacteria where MutS-MutL is generally absent. A screen for mutation avoidance genes showed that deletion of the gene encoding EndoMS in *Mycobacterium smegmatis* resulted in a hypermutation phenotype, increasing background mutation rate by about 100-fold (Castaneda-Garcia *et al.*, 2017). The higher rates of mutation were due to elevated levels of transitions (A:T to G:C or G:C to A:T), which is a hallmark of a MMR defect, and similar effects were observed when EndoMS was deleted in *Streptomyces coelicolor*. Mycobacterial EndoMS has no nuclease activity

when presented with mismatched DNA substrates *in vitro*, suggesting that further components in this non-canonical MMR pathway remain to be identified (Castaneda-Garcia *et al.*, 2017).

Taken together, the studies in archaea and bacteria make a compelling case that EndoMS participates in a MMR pathway. However, many important aspects of this pathway remain to be elucidated. The generation of double strand breaks by *P. furiosus* EndoMS is suggestive of an MMR process that functions via homologous recombination / DSBR (Ishino *et al.*, 2016). This has the advantage that there is no need to identify nascent DNA strands to pinpoint the mismatched base, as both will be resected during DSBR. The observation that EndoMS is sometimes found in an operon with the RadA recombinase lends further support to this hypothesis (Ren *et al.*, 2009). However, generation of a double strand break each time a mismatch is detected seems a risky strategy, unless homologous recombination is very efficient. This is probably the case in many of the euryarchaea, which are highly polyploid. It is much less obvious for the crenarchaea, which have a eukaryal-like cell cycle with monoploid and diploid stages (Lundgren & Bernander, 2007). Clearly, dissection and reconstitution of the pathway using genetic and biochemical techniques is a pressing priority. The interaction of archaeal EndoMS with the sliding clamp PCNA may provide a means to locate EndoMS at the replication fork to interrogate newly synthesised DNA, and could give the opportunity for co-location of a variety of DNA manipulation enzymes on the PCNA toolbelt (Beattie & Bell, 2011). In this regard, it will be interesting to see whether the bacterial EndoMS protein requires an interaction with the bacterial sliding clamp for activity.

Nucleotide Excision Repair (NER)

NER is a pathway that removes bulky, helix-distorting lesions such as photoproducts from DNA (Figure 1). Because it does not rely on direct detection of the lesion, but rather the resultant structural perturbation, it can repair many different types of DNA damage. The NER pathway in bacteria is catalysed by the UvrABC proteins, where UvrA is involved in damage recognition, UvrB is the helicase that opens the dsDNA and UvrC the nuclease that cuts on both sides of the lesion. In eukarya, an analogous and more complex pathway exists, which involves damage recognition by XPC-hr23b, DNA opening by Transcription factor IIH (TFIIH), subsequent binding of the XPA and RPA proteins, resulting in recruitment of the nucleases XPF-ERCC1 and XPG to cut on either side of the lesion. Archaea have a diverse and, frankly, confusing range of NER proteins encoded in their genomes (Figure 2), (Rouillon & White, 2011). In archaea that have co-opted the bacterial NER genes

encoding UvrABC, the bacterial system seems to be dominant for NER. For example, the NER patch repair size of 10-11 bp for *Methanothermobacter thermautotrophicum* is consistent with UvrABC function (Ogrunc *et al.*, 1998). Likewise, deletion of the genes for UvrA, UvrB or UvrC in *Halobacterium* NRC-1 resulted in a severe UV sensitivity despite the fact that this organism also has homologues of the eukaryal-type NER proteins XPF, XPB and XPD (Crowley *et al.*, 2006). Furthermore, there are no recognisable orthologues of the damage recognition proteins XPC and XPA in archaea. The SSB protein, which can melt damaged DNA specifically (Cubeddu & White, 2005) and can bind quickly and cooperatively on ssDNA (Morten *et al.*, 2015) could conceivably carry out this function.

Since most archaea have at least some eukaryal type NER genes, the question of their function is pertinent. Genetic studies of the putative archaeal NER pathway have been limited. Deletion of the XPD and XPB genes in *Thermococcus kodakaraensis* resulted in only very mild repair phenotypes (Fujikane *et al.*, 2010). In contrast, deletion of the XPF homologue Hef in this organism resulted in a marked sensitivity to the crosslinking agent mitomycin C (MMC), methylmethanesulfonate (MMS) and gamma radiation, suggesting an important role for Hef in multiple repair pathways including crosslink repair and replication restart (Fujikane *et al.*, 2010). This is consistent with the known roles of the eukaryal XPF and Mus81 proteins, which share a common ancestor with Hef (Rouillon & White, 2011). Both the helicase and nuclease activities of Hef were shown to be important, suggesting that Hef needs to unwind and cleave DNA during repair (Fujikane *et al.*, 2010). In the crenarchaeon *Sulfolobus islandicus*, deletion of the XPD, XPB and Bax1 genes has been reported with no resulting phenotype (She *et al.*, 2009). Although these results should be viewed as preliminary until published in more detail in a peer reviewed journal, they are consistent with the work in *T. kodakaraensis.* Overall then, genetic studies have shown that putative NER proteins are not essential, but have not progressed our understanding of the archaeal NER pathway very far. This has led Grogan to speculate that there is no NER pathway per se in archaea lacking UvrABC – raising the possibility that bulky NER-type lesions, which would represent a barrier to the replication fork, are removed by pathways that restart stalled forks (Grogan, 2015). Although this is an interesting hypothesis, it does beg the question: why do most archaea have XPB and XPD genes? After all, they must be doing *something*.

XPD helicase

Although we still have a rather limited understanding of archaeal NER, study of the XPD and XPB helicases has none-the-less been quite revealing. XPD is a 5' to 3' helicase with an essential iron-sulfur cluster (Rudolf *et al.*, 2006). In eukarya, XPD exists in the ten-subunit transcription factor TFIIH, along with the XPB helicase. TFIIH is involved in both NER, where DNA around a lesion is unwound, and transcription initiation, where RNA polymerase II promoters are unwound. XPD is essential for DNA unwinding in NER, but its activity is not required in transcription (Kuper *et al.*, 2014). Until recently, TFIIH was difficult to study at a structural level and the archaeal XPD, which is a monomer, was thus an attractive model system. Three groups independently reported the structure of archaeal XPD (Fan *et al.*, 2008, Liu *et al.*, 2008, Wolski *et al.*, 2008), revealing a four-domain organisation with two motor domains, an Arch and FeS domain (Figure 4). The mutations that cause the genetic condition xeroderma pigmentosum in humans, which arises from defective NER, could be mapped onto the archaeal XPD structures. The residues targeted by mutation are highly conserved, and cluster in areas involved in the catalytic mechanism of the archaeal enzyme – a striking example of conservation of function spanning the archaeal and eukaryal domains (Liu *et al.*, 2008).

In eukaryal NER, XPD has been shown to "proofread" for the presence of a DNA lesion in the translocated strand as a mechanism to increase the specificity of the NER reaction (Mathieu *et al.*, 2013). A lesion recognition pocket, close to the FeS cluster and immediately adjacent to the pore through which XPD pulls ssDNA, was identified. Two amino acids, Tyr-192 and Arg-196, were identified as an important part of this pocket, and mutations at these positions reduced DNA repair in a eukaryal system (Mathieu *et al.*, 2013). The authors went on to make the same changes in XPD from the archaeon *Ferroplasma acidophilum* (FacXPD), which correspond to residues Tyr-171 and Lys-175. This enzyme had been shown previously to stall at CPD lesions on the translocated strand (Mathieu *et al.*, 2010). They found that mutation of these residues did indeed abrogate the ability of FacXPD to stall at a CPD lesion, although helicase activity was unaffected (Mathieu *et al.*, 2013). However, XPD from *S. acidocaldarius* is not stalled by CPD or extrahelical fluorescein adducts in model substrates (Rudolf *et al.*, 2010). This may point to differences in the functions of XPD in the eury- and crenarchaea. Both SacXPD and FacXPD display only modest (~2-fold) increases in binding affinity for damaged versus undamaged DNA (Rudolf *et al.*, 2010, Ghoneim & Spies, 2014), suggesting that damage recognition, when it occurs, could be subtle. This picture is further complicated by the observation from Atomic Force Microscopy studies that TacXPD binds to

extrahelical fluorescein lesions in the translocated strand, but CPDs in the displaced strand (Buechner *et al.*, 2014). Furthermore, single molecule studies of FacXPD revealed the unexpected ability of the enzyme to bypass a bound single-strand DNA binding protein without either protein dissociating from the nucleic acid – a phenomenon that is still not fully understood (Honda *et al.*, 2009). Clearly, further work in this area would be desirable to improve our understanding of damage recognition by the XPD helicase.

Further studies of archaeal XPDs have revealed mechanistic insights into DNA binding and associated conformational changes. The Kisker lab succeeded in co-crystallising TacXPD with a short piece of ssDNA, demonstrating unequivocally the polarity of unwinding by the enzyme (Kuper *et al.*, 2012). The DNA was bound by motor domain 2, and the authors predicted that, since XPD can unwind bubble structures (Rudolf *et al.*, 2010) and eukaryal NER functions on DNA without ends, full engagement with DNA would require the opening of the interface between the Arch and FeS domains to allow DNA passage. This hypothesis was confirmed by the observation of transient opening of the interface in a single molecule study by the Spies lab (Ghoneim & Spies, 2014) and recently nailed down by a study which covalently closed the interface with a crosslinker (Constantinescu-Aruxandei *et al.*, 2016). Constantinescu and co-workers demonstrated that TacXPD can still bind DNA with high affinity when the interface between the Arch and FeS domain is covalently closed, but cannot function as a helicase. They proposed a two stage binding mechanism for XPD, with ssDNA initially bound tightly by motor domain 2, followed by transient opening of the Arch domain to allow passage through the central pore (Constantinescu-Aruxandei *et al.*, 2016). This mechanism is likely to hold true for eukaryal XPD in the context of TFIIH.

XPB helicase (or not?)

XPB has historically been considered to be a 3’ to 5’ DNA helicase, however the evidence supporting this assignment is rather thin. Helicase activity was ascribed to XPB from *Archaeoglobus fulgidus* (Fan *et al.*, 2006), but was not detected in either XPB protein from *S. solfataricus* (Richards *et al.*, 2008). The structure of AfuXPB revealed an unusual conformation, with the motor domains rotated away from the canonical structure by 170 °. The structure revealed two accessory domains, which were named the Damage Recognition domain (DRD) and Thumb (Fan *et al.*, 2006). The White lab reported that XPB is often found in an operon with a protein they named Bax1, and that the two proteins from a 1:1 complex (Richards *et al.*, 2008). Subsequently, Bax1 was shown to be a nuclease (Roth *et al.*, 2009), and a detailed study revealed that XPB and Bax1 function in concert to extend

bubble structures and cleave DNA (Rouillon & White, 2010). The Thumb domain was shown to be essential for DNA unwinding by XPB, and the DRD was shown to be essential for the function of the XPB-Bax1 complex, as no unwinding or nuclease activity was observed when it was deleted (Rouillon & White, 2010). In the past few years, evidence from studies of eukaryal TFIIH has accumulated that supports a role for XPB as a dsDNA translocase rather than a helicase. In this model, XPB binds dsDNA and catalyses opening of a DNA bubble downstream of the binding site in an ATP-dependent reaction (He *et al.*, 2016). Recent cryo-EM studies of the structural biology of transcription initiation appear to place this model beyond doubt (Schilbach *et al.*, 2017), at least for transcription and most likely for NER too. The work on archaeal XPB is largely consistent with a function as a dsDNA translocase rather than a helicase. The XPB-Bax1 complex could thus function as a stripped-down version of the eukaryal NER apparatus by binding at the site of helix-destabilising lesions, opening a bubble through XPB's ATP-dependent translocase activity and cleavage at the lesion by Bax1. Such a mechanism is still largely speculative however, requiring further study.

Transcription coupled repair (TCR)

TCR differs from Global Genome Repair (GGR, described above) in being initiated by stalling of RNA polymerase on the transcribed strand of genes. A coupling factor (Mfd in bacteria; CS-B/RAD26 in eukarya) is then recruited to the stalled complex and in turn recruits the NER machinery to repair the damage. This alternative NER pathway typically has faster kinetics than GGR, meaning that DNA lesions in transcribed strands are repaired more quickly than those in non-transcribed ones. RNA polymerase from the archaeon *Thermococcus kodakarensis* has been shown to stall when encountering a variety of DNA lesions in template strands during transcription, suggesting that stalled RNA polymerase molecules are a common sensor for DNA damage in all domains of life (Gehring & Santangelo, 2017). Accelerated TCR has been observed in the halophiles and shown to be dependent on UvrA in *H. salinarum* (Stantial *et al.*, 2016). This suggests a mechanism similar to that in bacteria, although there is no clear Mfd orthologue in in archaea. On the other hand, two independent studies have demonstrated that TCR is not faster than GGR in *S. solfataricus* (Dorazi *et al.*, 2007, Romano *et al.*, 2007). A comparison of the rates of repair of transcribed and non-transcribed strands in *S. solfataricus*, *E. coli* and *S. cerevisiae* suggests that the archaeon has a significantly faster rate of GGR, which may explain the apparent lack of accelerated TCR (Dorazi *et al.*, 2007). At any rate, the identification of coupling factors in archaea that link stalled RNA polymerase to the NER pathways is an important area for further study.

DNA transfer systems

Two independent studies of the transcriptional response to UV radiation in the *Sulfolobales* highlighted the upregulation of an operon comprised of 5 genes of unknown function (Fröls *et al.*, 2007, Götz *et al.*, 2007). UV treatment was also observed to result in significant and reversible cell aggregation in *S. acidocaldarius*. Subsequent analysis revealed that the operon encoded genes specific for a type IV pilus structure, leading the renaming of the operon as the *ups* operon (for UV-inducible pili operon of *Sulfolobus*) (Fröls *et al.*, 2008). These findings led to the hypothesis that the Ups system represents a machinery for the exchange of DNA following DNA damage (Fröls *et al.*, 2008). Pili were shown to mediate species-specific aggregates and to support large increases in the rate of homologous recombination, providing a survival advantage in following DNA damage (Ajon *et al.*, 2011). Downstream of the *ups* operon in *S. acidocaldarius* are four conserved genes predicted to encode a ParB-like nuclease, a glycosyl transferase, an EndoIII-like nuclease and a helicase. Deletion of these genes did not abrogate UV-induced aggregation, but did result in a significant decrease in survival following UV irradiation, suggesting that this operon likely plays a role in DNA repair that is related in some way to the UV-inducible DNA transfer pathway (van Wolferen *et al.*, 2015). A further twist to the story came with the identification of the Ced (Crenarchaeal system for exchange of DNA) system for DNA import. The UV-inducible *ced* genes encode CedA – an integral membrane protein, and CedB - a membrane bound hexameric DNA translocase related to HerA (van Wolferen *et al.*, 2016). CedA and CedB are thought to assemble to form a machine for the import of DNA following Ups-mediated cell aggregation, thus enhancing recombination and DNA repair. This fascinating system seems to be unique to the crenarchaea – no other examples of a DNA import (rather than export) machinery is currently known in the prokaryotes (van Wolferen *et al.*, 2016).

Base Excision Repair and Alternative Excision Repair

Damage to individual bases, caused for example by hydrolytic deamination, oxidation or methylation, is the most common and unavoidable type of DNA damage. Therefore, it is perhaps not surprising that the DNA repair pathway responsible for detection and correction of these lesions, Base Excision Repair (BER) is ubiquitous and fundamentally conserved across all domains of life. The canonical BER pathway is initiated when a glycosylase specific for a particular damaged base detects the lesion, usually by base flipping, and cleaves the glycosidic bond, generating an abasic (AP) site. The AP site in DNA is detected by AP Endonuclease, which cleaves the

phosphodiester backbone on the 5' side of the lesion, allowing DNA polymerase to initiate repair synthesis. Depending on circumstances, BER is completed by flap displacement and subsequent removal by the Fen1 nuclease (long patch repair), or by removal of the abasic nucleotide by RP lyase (short patch repair) – with both pathways resulting in nicked DNA that can be ligated (reviewed in (Grasso & Tell, 2014)). The Alternative Excision Repair (AER) pathway is imitated by an endonuclease (rather than a glycosylase), which nicks the DNA backbone next to a DNA lesion (reviewed in (Yasui, 2013)).

Since rates of hydrolytic deamination increase with growth temperature, this type of damage is particularly problematic for thermophiles and hyperthermophiles. Deamination of uracil, guanine and adenine, which give rise to uracil, xanthine and hypoxanthine respectively, are a particular problem as they have the potential to result in altered base pairing and hence mutation if not repaired quickly. Endonuclease V (EndoV) is a nuclease found in all domains of life that cuts at the 3' side of hypoxanthine residues in DNA, initiating the AER pathway Many archaeal genomes, including the majority of the thermophiles, possess a gene encoding EndoV (Kiyonari *et al.*, 2014). Biochemical studies that the EndoV enzyme from *A. fulgidus* and *P. furiosus* has the canonical specificity for inosine (Liu *et al.*, 2000); (Kiyonari *et al.*, 2014). In contrast, EndoV from *Ferroplasma acidarmanus* has a broader specificity for deaminated bases (Kanugula *et al.*, 2005). Recently, a second nuclease, Endonuclease Q (EndoQ) has been identified in *P. furiosus* which cleaves the DNA backbone on the 5' side of deaminated bases or abasic sites (Shiraishi *et al.*, 2015). In contrast to EndoV, the EndoQ enzyme has a narrow distribution in the archaea (Shiraishi *et al.*, 2015). EndoQ forms a physical and functional interaction with the sliding clamp PCNA (Shiraishi *et al.*, 2016), which may help direct the enzyme to the replication fork, increase the efficiency of the enzyme and allow coordinated repair with Fen1, DNA polymerase and DNA ligase, which are all PCNA-interacting enzymes (Figure 5). PCNA is an important partner for many other DNA repair enzymes, including AP Endonuclease (Kiyonari *et al.*, 2009), Uracil DNA glycosylase (Kiyonari *et al.*, 2008), the XPF nuclease (Roberts *et al.*, 2003) and the more recently characterised Nre protein (Giroux & MacNeill, 2016), which has a clear though as yet undefined role in DNA repair. In the future, we can expect that further BER enzymes, particularly nucleases, will be discovered in the distinct archaeal lineages. Orthologues of EndoQ outwith the *Thermococcales*, for example, seem very likely to exist but are as yet unidentified.

Double-strand Break Repair

Pathways of DSB repair

Double-stranded breaks (DSBs) are arguably the most lethal form of DNA damage that cells can incur. DSBs have the potential to block essential processes such as transcription, DNA replication, and cell division. Because both strands of the DNA duplex are broken, the inaccurate repair of DSBs can lead not just to mutations but also to genome rearrangements. The most accurate form of DSB repair, which largely avoids such collateral damage, is homologous recombination (HR). However, this is a complex and energetically-demanding process and for this reason, simpler but less accurate pathways of DSB repair operate alongside HR (Figure 6).

Non-homologous end-joining (NHEJ) is initiated by binding of the Ku protein complex, which acts as a scaffold to recruit nucleases, polymerases and ligases that process and repair the DSB (Figure 5). It is a rapid and versatile pathway of repair, which can accommodate DNA ends with a variety of lesions that would otherwise be refractory to ligation. Although it is error-prone, NHEJ is commonly used in eukaryotic cells, in particular higher eukaryotes that are quiescent in the G1 phase of the cell cycle, because it does not depend the presence of a homologous DNA duplex. However, NHEJ in archaea is rare because it requires the Ku protein and this is found in only a small number of species. In fact, a complete NHEJ complex, comprising Ku, polymerase, phosphoesterase and ligase, has only been found in *Methanocella paludicola* (Bartlett *et al.*, 2013). Crystal structures of these archaeal enzymes have demonstrated a conservation with the bacterial NHEJ counterparts (Bartlett *et al.*, 2016).

Microhomology-mediated end-joining (MMEJ) is a primitive method of DSB repair that does not require the Ku complex. Instead, DSBs are resected by exonucleases to expose short single-stranded tracts of homology that anneal with each other. Trimming of the resulting flaps is followed by DNA synthesis and ligation (Figure 6); like NHEJ, this method of DSB repair can result in deletions. MMEJ has been observed in *Haloferax volcanii* (Delmas *et al.*, 2009, Stachler *et al.*, 2017) and *S. islandicus* (Zhang & Whitaker, 2018), but the enzymatic basis is currently unknown.

Homologous recombination (HR) is the only error-free method of DSB repair, because it uses a second copy of DNA as a template (Figure 6). As suggested by its name, HR requires the intact template to be homologous to the broken DNA duplex, but genetic studies in *S. acidocaldarius* have

found that archaea might utilise shorter tracts of homology than bacteria or eukaryotes (Grogan & Stengel, 2008, Rockwood *et al.*, 2013). There are three steps to HR. *(1) Pre-synapsis*. The DSB is resected by exonucleases to generate 3' single-stranded DNA tails that are bound by the RecA-family recombinase, which in archaea is RadA. *(2) Synapsis*. The nucleoprotein filament formed by RadA engages in a homology search with an intact duplex, whereupon it catalyses strand exchange to form a displacement loop (D-loop); the 3' end in the D-loop is used to prime DNA synthesis. *(3) Post-synapsis.* At this point the invading strand may be displaced by a helicase, and the newly-synthesised section of DNA will allow it to reanneal with the other end of the DSB. This method of HR results exclusively in non-crossovers. Alternatively, capture of the second DSB end by the D-loop will result in the formation of a double Holliday junction structure. This is resolved by structure-specific endonucleases to yield either crossover or non-crossover products, depending on the orientation of the cuts.

HR is the best-studied pathway of DSB repair in archaea (White, 2011). In addition to its primary role in DNA repair (Fujikane *et al.*, 2010, Liang *et al.*, 2013, Zhang *et al.*, 2013), HR is used to promote genetic diversity following DNA transfer between *Sulfolobus* species (van Wolferen *et al.*, 2013, van Wolferen *et al.*, 2015, van Wolferen *et al.*, 2016) and between *Haloferax* species (Naor *et al.*, 2012, Naor *et al.*, 2016). HR is also used to restart DNA replication at stalled forks, which arise at DNA damage or protein roadblocks. This ability to initiate DNA replication using the invading 3' end of a D-loop is harnessed in strains of *H. volcanii* and *Thermococcus kodakarensis* that are deleted for replication origins. In origin-less mutants, HR is essential because it is used constitutively to initiate all DNA replication (Hawkins *et al.*, 2013, Gehring *et al.*, 2017).

HR pre-synapsis – Mre11-Rad50 and NurA-HerA

DSBs must be processed by exonucleases to generate the 3' single-stranded DNA tails that form nucleoprotein filaments with RadA. In *Escherichia coli*, this resection is carried out by RecBCD helicase/exonuclease. In eukaryotes, the Mre11 and Rad50 proteins form a complex that initiates resection by limited 3' to 5' degradation, followed by extensive resection by 5' to 3' exonucleases. Mre11 and Rad50 are conserved in archaea and structural studies have shown that they form a complex with DNA binding, unwinding and resection activities (Deshpande *et al.*, 2014, Sung *et al.*, 2014, Liu *et al.*, 2016). In *S. acidocaldarius*, the Mre11-Rad50 complex undergoes post-translational methylation in response to γ-irradiation (Kish *et al.*, 2016), and in *H. volcanii* the Mre11-Rad50

complex act in both the repair of DSBs and the compaction of the nucleoid after DNA damage (Delmas *et al.*, 2009, Delmas *et al.*, 2013).

In many archaeal species, the genes for Mre11 and Rad50 are found in an operon with those for the hexameric HerA helicase and the NurA nuclease, and the NurA-HerA complex has recently been the subject of much exciting research. Structural studies have revealed that NurA forms a toroidal dimer with a narrow central channel that can accommodate the two strands of an unwound duplex (Blackwood *et al.*, 2012, Byrne *et al.*, 2014). In complex with a HerA hexamer, the NurA dimer generates a continuous channel, indicating that HerA-driven translocation propels the DNA duplex through the NurA nuclease ring, where it is unwound and degraded (Figure 7) (Rzechorzek *et al.*, 2014, Ahdash *et al.*, 2017). The nuclease activity of NurA is modulated by HerA, and was found to be essential for cell viability in *S. islandicus* (De Falco *et al.*, 2015, Huang *et al.*, 2015). Bacterial homologues of NurA-HerA have been identified in *Deinococcus radiodurans*, and play a role in HR (Cheng *et al.*, 2015, Cheng *et al.*, 2015).

HR synapsis – SSB, RadA and its paralogues

The 3' single-stranded tail is bound by the RecA-family recombinase RadA (Morrical, 2015). RadA polymerisation is driven by the insertion of an invariant phenylalanine in the N-terminal domain into a binding pocket of an adjacent monomer (Figure 8). The DNA in this nucleoprotein filament is stretched ~1.5x in length, which facilitates the search for homologous sequences and the strand exchange process (Figure 6). To form the nucleoprotein filament, RadA must first displace single-stranded DNA binding protein (SSB), a ubiquitous protein with an oligonucleotide-binding (OB) fold, a twisted β-barrel with a binding site that accommodates four nucleotides of ssDNA (Lin *et al.*, 2008). The SSBs found in Euryarchaea are similar to the heterotrimeric eukaryotic replication protein A (RPA), which forms a heterotrimer, whereas the SSBs in Crenarchaea are more akin to the homotetrameric bacterial SSB; both the euryarchaeal RPA and crenarchaeal SSBs show a greater variety of architectures than their eukaryotic or bacterial counterparts. The *S. solfataricus* SSB has been shown to interact with RadA and inhibit its single-stranded DNA-dependent ATPase activity (Rolfsmeier & Haseltine, 2010). In order to stimulate strand exchange and overcome inhibition by SSB, the Rad54 protein of *S. solfataricus* can interact with RadA and remodel the topology of the homologous duplex DNA (Haseltine & Kowalczykowski, 2009).

The role of displacing SSB from single-stranded DNA and loading RadA more commonly falls to RadA paralogues (Lin *et al.*, 2006) and in this capacity, they are known as recombination mediators. RadB is found only in Euryarchaea, it interacts with RadA (Patoli *et al.*, 2017) and functions as a recombination mediator in *H. volcanii*, where it has been proposed to induce a conformational change in RadA and thereby promote its polymerisation on DNA (Wardell *et al.*, 2017). Similarly in *S. solfataricus*, the RadA paralogue SsoRal1 enhances RadA binding of single-stranded DNA and stabilises the nucleoprotein filament (Graham *et al.*, 2013). By contrast, the *S. solfataricus* paralogue Sso2452 and the *Sulfolobus tokodaii* paralogue stRadC2 have been found to inhibit strand exchange and D-loop formation by RadA (McRobbie *et al.*, 2009, Wang *et al.*, 2012). An *in vivo* study of two RadA paralogues in *S. islandicus*, RadC1 and RadC2, has shown that both are involved in DNA repair but the effect on HR has yet to be determined (Liang *et al.*, 2013).

HR post-synapsis – Hel308, Hef and Hjc

Once a D-loop is formed it can be used to prime DNA synthesis; the nascent 3' end may then be unwound to reanneal with the other side of the DSB. This is known as synthesis-dependent strand annealing (SDSA) and yields only non-crossover products. In archaea, the enzyme responsible for unwinding the invading strand is likely to be Hel308, a Ski2-family helicase found in archaea and metazoans but not in bacteria or yeast (Woodman & Bolt, 2009). Hel308 is essential for cell viability in *S. tokodaii* (Hong *et al.*, 2012, Song *et al.*, 2016) but not in *H. volcanii* (TA, unpublished). It interacts with RPA (Woodman *et al.*, 2011) and structural studies have shown that when Hel308 is bound to a 3' single-strand tailed partial duplex (Figure 9), the helicase domains encircle single-stranded DNA in a "ratchet" for directional translocation (Richards *et al.*, 2008). It has recently been found that DNA binding and unwinding by Hel308 requires a distinctive winged helix domain (Northall *et al.*, 2017). Taken together, these studies suggest that Hel308 controls HR at the D-loop step and assists in the restart of stalled DNA replication forks (Northall *et al.*, 2016).

Instead of being unwound, the D-loop may capture the second end of the DSB and thereby form a four-way Holliday junction structure. An enzyme that most likely mediates this transition in Euryarchaea is Hef (Lestini *et al.*, 2015). A member of the XPF/MUS81 family of structure-specific endonucleases, Hef comprises two distinct domains: an N-terminal domain of the DEAH helicase family and a C-terminal domain of the XPF endonuclease family, it acts on nicked, flapped and forked DNA (Komori *et al.*, 2004). Hef forms specific localisation foci *in vivo* in response to replication fork

arrest (Lestini *et al.*, 2013), and has been shown to interact with several DNA repair and replication proteins, including RecJ-like exonucleases and the PCNA sliding clamp of the DNA replication apparatus (Ishino *et al.*, 2014, Rohleder *et al.*, 2016, Nagata *et al.*, 2017). In *H. volcanii*, Hef is essential for cell viability when the Holliday junction resolvase Hjc is absent, and both the helicase and nuclease activities of Hef are indispensable (Lestini *et al.*, 2010). It has been proposed that Hef and Hjc provide alternative means to restart stalled DNA replication forks by processing Holliday junctions.

In contrast to Hef, Hjc has only nuclease activity and is specific for four-way DNA structures (Komori *et al.*, 1999). Enzymes of this class are known as Holliday junction resolvases and are capable of generating crossover products (Figure 6). A second Holliday junction resolvase, Hje, is present in Sulfolobales and a genetic study of Hjc and Hje in *S. islandicus* found that while deletion of either *hje* or *hjc* had no effect on cell viability, deletion of both *hje* and *hjc* is lethal (Huang *et al.*, 2015). This parallels the redundancy between Hjc and Hef in *H. volcanii* (Lestini *et al.*, 2010). Hjc has been observed to interact with many DNA repair proteins such as the RadA paralogue RadC2 (Wang *et al.*, 2012), the Hel308 helicase (Hong *et al.*, 2012) and a novel ATPase from *S. islandicus* termed SisPINA (Zhai *et al.*, 2017); the latter forms hexameric rings, similar to the bacterial Holliday junction migration helicase RuvB. Another novel protein that has been reported to bind to Holliday junctions is the phMutS5 mismatch repair enzyme from *Pyrococcus horikoshii* (Ohshita *et al.*, 2017), but unlike eukaryotic MutS homologues that act in HR, phMutS5 showed no nuclease activity on branched DNA.

Applications of DSB repair

HR is not only an error-free method of DSB repair but also a cornerstone of archaeal genetics (Leigh *et al.*, 2011, Farkas *et al.*, 2013). The ability to target a specific gene for deletion or mutation, using plasmid constructs with flanking regions of homology, relies on HR (Figure 10). Refinements of these methods have enabled the high-throughput generation and screening of targeted mutants in *Pyrococcus furiosus* (Farkas *et al.*, 2012), *S. islandicus* (Zhang *et al.*, 2013) and *H. volcanii* (Kiljunen *et al.*, 2014); the latter is notable for using a transposon insertion library to carry out saturation mutagenesis, which facilitates the identification of non-essential genes in any specific pathway.

Other pathways of DSB repair have been harnessed in genetic manipulation and genome engineering. MMEJ has recently been used for a high-throughput method for targeted gene inactivation in *S. islandicus*, in one case the minimal size of micro-homology for marker replacement was as few as 10 bp (Zhang & Whitaker, 2018). In *Methanosarcina acetivorans*, a system of CRISPR-Cas9-mediated genome engineering has been developed and it was found that co-expression the NHEJ machinery from *M. paludicola* allowed efficient genome editing without the need for a repair template (Nayak & Metcalf, 2017).

The enzymes involved in DSB repair have also found applications *in vitro*. For example, the thermostable RadA recombinase from *Pyrococcus woesei* enhances the specificity of simplex and multiplex PCR assays (Stefanska *et al.*, 2016). Similarly, the Hel308 helicase from *Thermococcus gammatolerans* has found a new lease of life as a motor protein for nanopore sequencing. Owing to its ability to unwind duplex DNA and ratchet the single stranded DNA through the nanopore in a step-wise manner, Hel308 significantly improves the accuracy of single-molecule sequencing (Craig *et al.*, 2015, Derrington *et al.*, 2015, Craig *et al.*, 2017).

Concluding Remarks

Research into DNA repair in the archaea has flourished since the turn of the millennium, driven largely by the availability of genome sequences. However, the emerging picture fits with neither of the preconceptions that were held twenty years ago. Archaea are neither “odd” bacteria, a view held by detractors of the third domain of life, nor are they “mini-eukaryotes” as proposed by those who believed they would serve as simplified models for human cells. Instead, archaea have proved to be every bit as unique and diverse as bacteria and eukaryotes, and the archaeal systems for DNA repair reflect this distinctive status. Genomic surveys have revealed a patchwork of bacterial and eukaryotic repair enzymes, alongside proteins that are unique to archaea, but laboratory studies have shown that these enzymes do not necessarily behave in the same way as their bacterial or eukaryotic counterparts. Nevertheless, one aspect of the field has not changed in twenty years – archaea and their systems for DNA repair continue to serve as a window into our evolutionary past.

Acknowledgements

MFW is the recipient of a Royal Society Wolfson Research Merit Award. Work in TA’s laboratory is funded by the Biotechnology and Biological Sciences Research Council (BBSRC).

References

Adam PS, Borrel G, Brochier-Armanet C & Gribaldo S (2017) The growing tree of Archaea: new perspectives on their diversity, evolution and ecology. *ISME J* **11**: 2407-2425.

Ahdash Z, Lau AM, Byrne RT, Lammens K, Stuetzer A, Urlaub H, Booth PJ, Reading E, Hopfner KP & Politis A (2017) Mechanistic insight into the assembly of the HerA-NurA helicase-nuclease DNA end resection complex. *Nucleic Acids Res* **45**: 12025-12038.

Ajon M, Frols S, van Wolferen M, Stoecker K, Teichmann D, Driessen AJ, Grogan DW, Albers SV & Schleper C (2011) UV-inducible DNA exchange in hyperthermophilic archaea mediated by type IV pili. *Mol Microbiol* **82**: 807-817.

Bartlett EJ, Brissett NC & Doherty AJ (2013) Ribonucleolytic resection is required for repair of strand displaced nonhomologous end-joining intermediates. *Proc Natl Acad Sci U S A* **110**: E1984-1991.

Bartlett EJ, Brissett NC, Plocinski P, Carlberg T & Doherty AJ (2016) Molecular basis for DNA strand displacement by NHEJ repair polymerases. *Nucleic Acids Res* **44**: 2173-2186.

Beattie TR & Bell SD (2011) Molecular machines in archaeal DNA replication. *Curr Opin Chem Biol* **15**: 614-619.

Blackwood JK, Rzechorzek NJ, Abrams AS, Maman JD, Pellegrini L & Robinson NP (2012) Structural and functional insights into DNA-end processing by the archaeal HerA helicase-NurA nuclease complex. *Nucleic Acids Res* **40**: 3183-3196.

Buechner CN, Heil K, Michels G, Carell T, Kisker C & Tessmer I (2014) Strand-specific recognition of DNA damages by XPD provides insights into nucleotide excision repair substrate versatility. *J Biol Chem* **289**: 3613-3624.

Busch CR & DiRuggiero J (2010) MutS and MutL are dispensable for maintenance of the genomic mutation rate in the halophilic archaeon *Halobacterium salinarum* NRC-1. *PLoS One* **5**: e9045.

Byrne RT, Schuller JM, Unverdorben P, Forster F & Hopfner KP (2014) Molecular architecture of the HerA-NurA DNA double-strand break resection complex. *FEBS Lett* **588**: 4637-4644.

Castaneda-Garcia A, Prieto AI, Rodriguez-Beltran J*, et al.* (2017) A non-canonical mismatch repair pathway in prokaryotes. *Nat Commun* **8**: 14246.

Cheng K, Chen X, Xu G, Wang L, Xu H, Yang S, Zhao Y & Hua Y (2015) Biochemical and Functional Characterization of the NurA-HerA Complex from *Deinococcus radiodurans*. *J Bacteriol* **197**: 2048-2061.

Cheng K, Zhao Y, Chen X, Li T, Wang L, Xu H, Tian B & Hua Y (2015) A Novel C-Terminal Domain of RecJ is Critical for Interaction with HerA in *Deinococcus radiodurans*. *Front Microbiol* **6**: 1302.

Constantinescu-Aruxandei D, Petrovic-Stojanovska B, Penedo JC, White MF & Naismith JH (2016) Mechanism of DNA loading by the DNA repair helicase XPD. *Nucleic Acids Res* **44**: 2806-2815.

Craig JM, Laszlo AH, Derrington IM, Ross BC, Brinkerhoff H, Nova IC, Doering K, Tickman BI, Svet MT & Gundlach JH (2015) Direct Detection of Unnatural DNA Nucleotides dNaM and d5SICS using the MspA Nanopore. *PLoS One* **10**: e0143253.

Craig JM, Laszlo AH, Brinkerhoff H, Derrington IM, Noakes MT, Nova IC, Tickman BI, Doering K, de Leeuw NF & Gundlach JH (2017) Revealing dynamics of helicase translocation on single-stranded DNA using high-resolution nanopore tweezers. *Proc Natl Acad Sci U S A* **114**: 11932-11937.

Crowley DJ, Boubriak I, Berquist BR, Clark M, Richard E, Sullivan L, Dassarma S & McCready S (2006) The uvrA, uvrB and uvrC genes are required for repair of ultraviolet light induced DNA photoproducts in *Halobacterium* sp. NRC-1. *Saline sys* **2**: 11.

Cubeddu L & White MF (2005) DNA damage detection by an archaeal single-stranded DNA-binding protein. *J Mol Biol* **353**: 507-516.

Da Cunha V, Gaia M, Gadelle D, Nasir A & Forterre P (2017) Lokiarchaea are close relatives of Euryarchaeota, not bridging the gap between prokaryotes and eukaryotes. *PLoS Genet* **13**: e1006810.

De Falco M, Catalano F, Rossi M, Ciaramella M & De Felice M (2015) NurA is endowed with endo- and exonuclease activities that are modulated by HerA: new insight into their role in DNA-end processing. *PLoS One* **10**: e0142345.

Delmas S, Duggin IG & Allers T (2013) DNA damage induces nucleoid compaction via the Mre11-Rad50 complex in the archaeon *Haloferax volcanii*. *Mol Microbiol* **87**: 168-179.

Delmas S, Shunburne L, Ngo HP & Allers T (2009) Mre11-Rad50 promotes rapid repair of DNA damage in the polyploid archaeon *Haloferax volcanii* by restraining homologous recombination. *PLoS Genet* **5**: e1000552.

Derrington IM, Craig JM, Stava E*, et al.* (2015) Subangstrom single-molecule measurements of motor proteins using a nanopore. *Nat Biotechnol* **33**: 1073-1075.

Deshpande RA, Williams GJ, Limbo O*, et al.* (2014) ATP-driven Rad50 conformations regulate DNA tethering, end resection, and ATM checkpoint signaling. *EMBO J* **33**: 482-500.

Dorazi R, Gotz D, Munro S, Bernander R & White MF (2007) Equal rates of repair of DNA photoproducts in transcribed and non-transcribed strands in *Sulfolobus solfataricus*. *Mol Microbiol* **63**: 521-529.

Eme L, Spang A, Lombard J, Stairs CW & Ettema TJG (2017) Archaea and the origin of eukaryotes. *Nat Rev Microbiol* **10:** 711-723.

Fan L, Arvai AS, Cooper PK, Iwai S, Hanaoka F & Tainer JA (2006) Conserved XPB core structure and motifs for DNA unwinding: Implications for pathway selection of transcription or excision repair. *Mol Cell* **22**: 27-37.

Fan L, Fuss JO, Cheng QJ, Arvai AS, Hammel M, Roberts VA, Cooper PK & Tainer JA (2008) XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations. *Cell* **133**: 789-800.

Farkas J, Stirrett K, Lipscomb GL, Nixon W, Scott RA, Adams MW & Westpheling J (2012) Recombinogenic properties of *Pyrococcus furiosus* strain COM1 enable rapid selection of targeted mutants. *Appl Environ Microbiol* **78**: 4669-4676.

Farkas JA, Picking JW & Santangelo TJ (2013) Genetic techniques for the archaea. *Annu Rev Genet* **47**: 539-561.

Fröls S, Gordon PM, Panlilio MA, Duggin IG, Bell SD, Sensen CW & Schleper C (2007) Response of the hyperthermophilic archaeon *Sulfolobus solfataricus* to UV damage. *J Bacteriol* **189**: 8708-8718.

Fröls S, Ajon M, Wagner M, Teichmann D, Zolghadr B, Folea M, Boekema EJ, Driessen AJ, Schleper C & Albers SV (2008) UV-inducible cellular aggregation of the hyperthermophilic archaeon *Sulfolobus solfataricus* is mediated by pili formation. *Mol Microbiol* **70**: 938-952.

Fujikane R, Ishino S, Ishino Y & Forterre P (2010) Genetic analysis of DNA repair in the hyperthermophilic archaeon, *Thermococcus kodakaraensis*. *Genes Genet Syst* **85**: 243-257.

Gehring AM & Santangelo TJ (2017) Archaeal RNA polymerase arrests transcription at DNA lesions. *Transcription* **8**: 288-296.

Gehring AM, Astling DP, Matsumi R, Burkhart BW, Kelman Z, Reeve JN, Jones KL & Santangelo TJ (2017) Genome Replication in *Thermococcus kodakarensis* Independent of Cdc6 and an Origin of Replication. *Front Microbiol* **8**: 2084.

Ghoneim M & Spies M (2014) Direct correlation of DNA binding and single protein domain motion via dual illumination fluorescence microscopy. *Nano Lett* **14**: 5920-5931.

Giroux X & MacNeill SA (2016) A novel archaeal DNA repair factor that acts with the UvrABC system to repair mitomycin C-induced DNA damage in a PCNA-dependent manner. *Mol Microbiol* **99**: 1-14.

Götz D, Paytubi S, Munro S, Lundgren M, Bernander R & White MF (2007) Responses of hyperthermophilic crenarchaea to UV irradiation. *Genome Biol* **8**: R220.

Graham WJt, Rolfsmeier ML & Haseltine CA (2013) An archaeal RadA paralog influences presynaptic filament formation. *DNA Repair (Amst)* **12**: 403-413.

Grasso S & Tell G (2014) Base excision repair in Archaea: back to the future in DNA repair. *DNA Repair (Amst)* **21**: 148-157.

Grogan DW (1998) Hyperthermophiles and the problem of DNA instability. *Mol Microbiol* **28**: 1043-1049.

Grogan DW (2004) Stability and repair of DNA in hyperthermophilic Archaea. *Curr Issues Mol Biol* **6**: 137-144.

Grogan DW (2015) Understanding DNA Repair in Hyperthermophilic Archaea: Persistent Gaps and Other Reasons to Focus on the Fork. *Archaea* **2015**: 942605.

Grogan DW & Stengel KR (2008) Recombination of synthetic oligonucleotides with prokaryotic chromosomes: substrate requirements of the *Escherichia coli*/lambdaRed and *Sulfolobus acidocaldarius* recombination systems. *Mol Microbiol* **69**: 1255-1265.

Haseltine CA & Kowalczykowski SC (2009) An archaeal Rad54 protein remodels DNA and stimulates DNA strand exchange by RadA. *Nucleic Acids Res* **37**: 2757-2770.

Hawkins M, Malla S, Blythe MJ, Nieduszynski CA & Allers T (2013) Accelerated growth in the absence of DNA replication origins. *Nature* **503**: 544-547.

He Y, Yan CL, Fang J, Inouye C, Tjian R, Ivanov I & Nogales E (2016) Near-atomic resolution visualization of human transcription promoter opening. *Nature* **533**: 359-365.

Honda M, Park J, Pugh RA, Ha T & Spies M (2009) Single-molecule analysis reveals differential effect of ssDNA-binding proteins on DNA translocation by XPD helicase. *Mol Cell* **35**: 694-703.

Hong Y, Chu M, Li Y, Ni J, Sheng D, Hou G, She Q & Shen Y (2012) Dissection of the functional domains of an archaeal Holliday junction helicase. *DNA Repair (Amst)* **11**: 102-111.

Huang Q, Liu L, Liu J, Ni J, She Q & Shen Y (2015) Efficient 5'-3' DNA end resection by HerA and NurA is essential for cell viability in the crenarchaeon *Sulfolobus islandicus*. *BMC Mol Biol* **16**: 2.

Ishino S, Nishi Y, Oda S, Uemori T, Sagara T, Takatsu N, Yamagami T, Shirai T & Ishino Y (2016) Identification of a mismatch-specific endonuclease in hyperthermophilic Archaea. *Nucleic Acids Res* **44**: 2977-2986.

Ishino S, Yamagami T, Kitamura M*, 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.

Kanugula S, Pauly GT, Moschel RC & Pegg AE (2005) A bifunctional DNA repair protein from *Ferroplasma acidarmanus* exhibits O6-alkylguanine-DNA alkyltransferase and endonuclease V activities. *Proc Natl Acad Sci U S A* **102**: 3617-3622.

Kelman Z & White MF (2005) Archaeal DNA replication and repair. *Curr Opin Microbiol* **8**: 669-676.

Kiljunen S, Pajunen MI, Dilks K, Storf S, Pohlschroder M & Savilahti H (2014) Generation of comprehensive transposon insertion mutant library for the model archaeon, *Haloferax volcanii*, and its use for gene discovery. *BMC Biol* **12**: 103.

Kish A, Gaillard JC, Armengaud J & Elie C (2016) Post-translational methylations of the archaeal Mre11:Rad50 complex throughout the DNA damage response. *Mol Microbiol* **100**: 362-378.

Kiyonari S, Uchimura M, Shirai T & Ishino Y (2008) Physical and functional interactions between uracil-DNA glycosylase and proliferating cell nuclear antigen from the euryarchaeon *Pyrococcus furiosus*. *J Biol Chem* **283**: 24185-24193.

Kiyonari S, Egashira Y, Ishino S & Ishino Y (2014) Biochemical characterization of endonuclease V from the hyperthermophilic archaeon, *Pyrococcus furiosus*. *J Biochem* **155**: 325-333.

Kiyonari S, Tahara S, Shirai T, Iwai S, Ishino S & Ishino Y (2009) Biochemical properties and base excision repair complex formation of apurinic/apyrimidinic endonuclease from *Pyrococcus furiosus*. *Nucleic Acids Res* **37**: 6439-6453.

Komori K, Sakae S, Shinagawa H, Morikawa K & Ishino Y (1999) A Holliday junction resolvase from *Pyrococcus furiosus*: functional similarity to *Escherichia coli* RuvC provides evidence for conserved mechanism of homologous recombination in Bacteria, Eukarya, and Archaea. *Proc Natl Acad Sci U S A* **96**: 8873-8878.

Komori K, Hidaka M, Horiuchi T, Fujikane R, Shinagawa H & Ishino Y (2004) Cooperation of the N-terminal Helicase and C-terminal endonuclease activities of Archaeal Hef protein in processing stalled replication forks. *J Biol Chem* **279**: 53175-53185.

Kuper J, Wolski SC, Michels G & Kisker C (2012) Functional and structural studies of the nucleotide excision repair helicase XPD suggest a polarity for DNA translocation. *EMBO J* **31**: 494-502.

Kuper J, Braun C, Elias A, Michels G, Sauer F, Schmitt DR, Poterszman A, Egly JM & Kisker C (2014) In TFIIH, XPD helicase is exclusively devoted to DNA repair. *PLoS Biol* **12**: e1001954.

Leigh JA, Albers SV, Atomi H & Allers T (2011) Model organisms for genetics in the domain Archaea: methanogens, halophiles, Thermococcales and Sulfolobales. *FEMS Microbiology Reviews* **35**: 577-608.

Lestini R, Duan Z & 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.

Lestini R, Delpech F & Myllykallio H (2015) DNA replication restart and cellular dynamics of Hef helicase/nuclease protein in *Haloferax volcanii*. *Biochimie* **118**: 254-263.

Lestini R, Laptenok SP, Kuhn J, Hink MA, Schanne-Klein MC, Liebl U & Myllykallio H (2013) Intracellular dynamics of archaeal FANCM homologue Hef in response to halted DNA replication. *Nucleic Acids Res* **41**: 10358-10370.

Liang PJ, Han WY, Huang QH, Li YZ, Ni JF, She QX & Shen YL (2013) Knockouts of RecA-like proteins RadC1 and RadC2 have distinct responses to DNA damage agents in *Sulfolobus islandicus*. *J Genet Genom* **40**: 533-542.

Lin Y, Lin LJ, Sriratana P, Coleman K, Ha T, Spies M & Cann IK (2008) Engineering of functional replication protein a homologs based on insights into the evolution of oligonucleotide/oligosaccharide-binding folds. *J Bacteriol* **190**: 5766-5780.

Lin Z, Kong H, Nei M & Ma H (2006) Origins and evolution of the recA/RAD51 gene family: evidence for ancient gene duplication and endosymbiotic gene transfer. *Proc Natl Acad Sci U S A* **103**: 10328-10333.

Liu H, Rudolf J, Johnson KA, McMahon SA, Oke M, Carter L, McRobbie AM, Brown SE, Naismith JH & White MF (2008) Structure of the DNA repair helicase XPD. *Cell* **133**: 801-812.

Liu J, He B, Qing H & Kow YW (2000) A deoxyinosine specific endonuclease from hyperthermophile, Archaeoglobus fulgidus: a homolog of *Escherichia coli* endonuclease V. *Mutat Res* **461**: 169-177.

Liu Y, Sung S, Kim Y*, et al.* (2016) ATP-dependent DNA binding, unwinding, and resection by the Mre11/Rad50 complex. *EMBO J* **35**: 743-758.

Lundgren M & Bernander R (2007) Genome-wide transcription map of an archaeal cell cycle. *Proc Natl Acad Sci U S A* **104**: 2939-2944.

Mathieu N, Kaczmarek N & Naegeli H (2010) Strand- and site-specific DNA lesion demarcation by the xeroderma pigmentosum group D helicase. *Proc Natl Acad Sci U S A* **107**: 17545-17550.

Mathieu N, Kaczmarek N, Ruthemann P, Luch A & Naegeli H (2013) DNA quality control by a lesion sensor pocket of the xeroderma pigmentosum group D helicase subunit of TFIIH. *Curr Biol* **23**: 204-212.

McRobbie AM, Carter LG, Kerou M, Liu H, McMahon SA, Johnson KA, Oke M, Naismith JH & White MF (2009) Structural and functional characterisation of a conserved archaeal RadA paralog with antirecombinase activity. *J Mol Biol* **389**: 661-673.

Morrical SW (2015) DNA-pairing and annealing processes in homologous recombination and homology-directed repair. *Cold Spring Harb Perspect Biol* **7**: a016444.

Morten MJ, Peregrina JR, Figueira-Gonzalez M, Ackermann K, Bode BE, White MF & Penedo JC (2015) Binding dynamics of a monomeric SSB protein to DNA: a single-molecule multi-process approach. *Nucleic Acids Res* **43**: 10907-10924.

Nagata M, Ishino S, Yamagami T, Simons JR, Kanai T, Atomi H & Ishino Y (2017) Possible function of the second RecJ-like protein in stalled replication fork repair by interacting with Hef. *Sci Rep* **7**: 16949.

Nakae S, Hijikata A, Tsuji T, Yonezawa K, Kouyama KI, Mayanagi K, Ishino S, Ishino Y & Shirai T (2016) Structure of the EndoMS-DNA complex as mismatch restriction endonuclease. *Structure* **24**: 1960-1971.

Naor A, Lapierre P, Mevarech M, Papke RT & Gophna U (2012) Low species barriers in halophilic archaea and the formation of recombinant hybrids. *Curr Biol* **22**: 1444-1448.

Naor A, Altman-Price N, Soucy SM, Green AG, Mitiagin Y, Turgeman-Grott I, Davidovich N, Gogarten JP & Gophna U (2016) Impact of a homing intein on recombination frequency and organismal fitness. *Proc Natl Acad Sci U S A* **113**: E4654-4661.

Nayak DD & Metcalf WW (2017) Cas9-mediated genome editing in the methanogenic archaeon *Methanosarcina acetivorans*. *Proc Natl Acad Sci U S A* **114**: 2976-2981.

Northall SJ, Ivancic-Bace I, Soultanas P & Bolt EL (2016) Remodeling and Control of Homologous Recombination by DNA Helicases and Translocases that Target Recombinases and Synapsis. *Genes (Basel)* **7**(8): E52.

Northall SJ, Buckley R, Jones N, Penedo JC, Soultanas P & Bolt EL (2017) DNA binding and unwinding by Hel308 helicase requires dual functions of a winged helix domain. *DNA Repair (Amst)* **57**: 125-132.

Ogrunc M, Becker DF, Ragsdale SW & Sancar A (1998) Nucleotide excision repair in the third kingdom. *J Bacteriol* **180**: 5796-5798.

Ohshita K, Fukui K, Sato M, Morisawa T, Hakumai Y, Morono Y, Inagaki F, Yano T, Ashiuchi M & Wakamatsu T (2017) Archaeal MutS5 tightly binds to Holliday junction similarly to eukaryotic MutSgamma. *FEBS J* **284**: 3470-3483.

Patoli BB, Winter JA, Patoli AA, Delahay RM & Bunting KA (2017) Co-expression and purification of the RadA recombinase with the RadB paralog from *Haloferax volcanii* yields heteromeric ring-like structures. *Microbiol* **163**: 1802-1811.

Ren B, Kuhn J, Meslet-Cladiere L, Briffotaux J, Norais C, Lavigne R, Flament D, Ladenstein R & Myllykallio H (2009) Structure and function of a novel endonuclease acting on branched DNA substrates. *EMBO J* **28**: 2479-2489.

Richards JD, Cubeddu L, Roberts J, Liu H & White MF (2008) The archaeal XPB protein is a ssDNA-dependent ATPase with a novel partner. *J Mol Biol* **376**: 634-644.

Richards JD, Johnson KA, Liu H, McRobbie AM, McMahon S, Oke M, Carter L, Naismith JH & White MF (2008) Structure of the DNA repair helicase hel308 reveals DNA binding and autoinhibitory domains. *J Biol Chem* **283**: 5118-5126.

Roberts JA, Bell SD & White MF (2003) An archaeal XPF repair endonuclease dependent on a heterotrimeric PCNA. *Mol Micro* **48**: 361-371.

Rockwood J, Mao D & Grogan DW (2013) Homologous recombination in the archaeon *Sulfolobus acidocaldarius*: effects of DNA substrates and mechanistic implications. *Microbiol* **159**: 1888-1899.

Rohleder F, Huang J, Xue Y, Kuper J, Round A, Seidman M, Wang W & Kisker C (2016) FANCM interacts with PCNA to promote replication traverse of DNA interstrand crosslinks. *Nucleic Acids Res* **44**: 3219-3232.

Rolfsmeier ML & Haseltine CA (2010) The single-stranded DNA binding protein of *Sulfolobus solfataricus* acts in the presynaptic step of homologous recombination. *J Mol Biol* **397**: 31-45.

Romano V, Napoli A, Salerno V, Valenti A, Rossi M & Ciaramella M (2007) Lack of strand-specific repair of UV-induced DNA lesions in three genes of the archaeon *Sulfolobus solfataricus*. *J Mol Biol* **365**: 921-929.

Roth HM, Tessmer I, Van Houten B & Kisker C (2009) Bax1 is a novel endonuclease: implications for archaeal nucleotide excision repair. *J Biol Chem* **284**: 32272-32278.

Rouillon C & White MF (2010) The XBP-Bax1 helicase-nuclease complex unwinds and cleaves DNA: implications for eukaryal and archaeal nucleotide excision repair. *J Biol Chem* **285**: 11013-11022.

Rouillon C & White MF (2011) The evolution and mechanisms of nucleotide excision repair proteins. *Res Microbiol* **162**: 19-26.

Rudolf J, Rouillon C, Schwarz-Linek U & White MF (2010) The helicase XPD unwinds bubble structures and is not stalled by DNA lesions removed by the nucleotide excision repair pathway. *Nucleic Acids Res* **38**: 931-941.

Rudolf J, Makrantoni V, Ingledew WJ, Stark MJ & White MF (2006) The DNA repair helicases XPD and FancJ have essential iron-sulfur domains. *Mol Cell* **23**: 801-808.

Rzechorzek NJ, Blackwood JK, Bray SM, Maman JD, Pellegrini L & Robinson NP (2014) Structure of the hexameric HerA ATPase reveals a mechanism of translocation-coupled DNA-end processing in archaea. *Nat Commun* **5**: 5506.

Schilbach S, Hantsche M, Tegunov D, Dienemann C, Wigge C, Urlaub H & Cramer P (2017) Structures of transcription pre-initiation complex with TFIIH and Mediator. *Nature* **551**: 204-209.

She Q, Zhang C, Deng L, Peng N, Chen Z & Liang YX (2009) Genetic analyses in the hyperthermophilic archaeon *Sulfolobus islandicus*. *Biochem Soc Trans* **37**: 92-96.

Shiraishi M, Ishino S, Yamagami T, Egashira Y, Kiyonari S & Ishino Y (2015) A novel endonuclease that may be responsible for damaged DNA base repair in *Pyrococcus furiosus*. *Nucleic Acids Res* **43**: 2853-2863.

Shiraishi M, Ishino S, Yoshida K, Yamagami T, Cann I & Ishino Y (2016) PCNA is involved in the EndoQ-mediated DNA repair process in Thermococcales. *Sci Rep* **6**: 25532.

Song X, Ni J & Shen Y (2016) Structure-Based genetic analysis of Hel308a in the hyperthermophilic archaeon *Sulfolobus islandicus*. *J Genet Genomics* **43**: 405-413.

Spang A, Caceres EF & Ettema TJG (2017) Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. *Science* **357**: 6351.

Stachler AE, Turgeman-Grott I, Shtifman-Segal E, Allers T, Marchfelder A & Gophna U (2017) High tolerance to self-targeting of the genome by the endogenous CRISPR-Cas system in an archaeon. *Nucleic Acids Res* **45**: 5208-5216.

Stantial N, Dumpe J, Pietrosimone K, Baltazar F & Crowley DJ (2016) Transcription-coupled repair of UV damage in the halophilic archaea. *DNA Repair (Amst)* **41**: 63-68.

Stefanska A, Gaffke L, Kaczorowska AK, Plotka M, Dabrowski S & Kaczorowski T (2016) Highly thermostable RadA protein from the archaeon *Pyrococcus woesei* enhances specificity of simplex and multiplex PCR assays. *J Appl Genet* **57**: 239-249.

Sung S, Li F, Park YB, Kim JS, Kim AK, Song OK, Kim J, Che J, Lee SE & Cho Y (2014) DNA end recognition by the Mre11 nuclease dimer: insights into resection and repair of damaged DNA. *EMBO J* **33**: 2422-2435.

van Wolferen M, Ma X & Albers SV (2015) DNA Processing Proteins Involved in the UV-Induced Stress Response of Sulfolobales. *J Bacteriol* **197**: 2941-2951.

van Wolferen M, Ajon M, Driessen AJ & Albers SV (2013) Molecular analysis of the UV-inducible pili operon from *Sulfolobus acidocaldarius*. *Microbiol open* **2**: 928-937.

van Wolferen M, Wagner A, van der Does C & Albers SV (2016) The archaeal Ced system imports DNA. *Proc Natl Acad Sci U S A* **113**: 2496-2501.

Wang L, Sheng D, Han W, Huang B, Zhu S, Ni J, Li J & Shen Y (2012) *Sulfolobus tokodaii* RadA paralog, stRadC2, is involved in DNA recombination via interaction with RadA and Hjc. *Sci China Life Sci* **55**: 261-267.

Wardell K, Haldenby S, Jones N, Liddell S, Ngo GHP & Allers T (2017) RadB acts in homologous recombination in the archaeon *Haloferax volcanii*, consistent with a role as recombination mediator. *DNA Repair (Amst)* **55**: 7-16.

White MF (2003) Archaeal DNA repair: paradigms and puzzles. *Biochem Soc Trans* **31**: 690-693.

White MF (2011) Homologous recombination in the archaea: the means justify the ends. *Biochem Soc Trans* **39**: 15-19.

Wolski SC, Kuper J, Hanzelmann P, Truglio JJ, Croteau DL, Van Houten B & Kisker C (2008) Crystal structure of the FeS cluster-containing nucleotide excision repair helicase XPD. *PLoS Biol* **6**: e149.

Woodman IL & Bolt EL (2009) Molecular biology of Hel308 helicase in archaea. *Biochem Soc Trans* **37**: 74-78.

Woodman IL, Brammer K & Bolt EL (2011) Physical interaction between archaeal DNA repair helicase Hel308 and Replication Protein A (RPA). *DNA Repair (Amst)* **10**: 306-313.

Yasui A (2013) Alternative excision repair pathways. *Cold Spring Harb Perspect Biol* **5**(6) a012617.

Zhai B, DuPrez K, Doukov TI, Li H, Huang M, Shang G, Ni J, Gu L, Shen Y & Fan L (2017) Structure and Function of a Novel ATPase that Interacts with Holliday Junction Resolvase Hjc and Promotes Branch Migration. *J Mol Biol* **429**: 1009-1029.

Zhang C & Whitaker RJ (2018) Microhomology-Mediated High-Throughput Gene Inactivation Strategy for the Hyperthermophilic Crenarchaeon *Sulfolobus islandicus*. *Appl Environ Microbiol* **84**.

Zhang C, Tian B, Li S, Ao X, Dalgaard K, Gokce S, Liang Y & She Q (2013) Genetic manipulation in *Sulfolobus islandicus* and functional analysis of DNA repair genes. *Biochem Soc Trans* **41**: 405-410.

Figure Legends

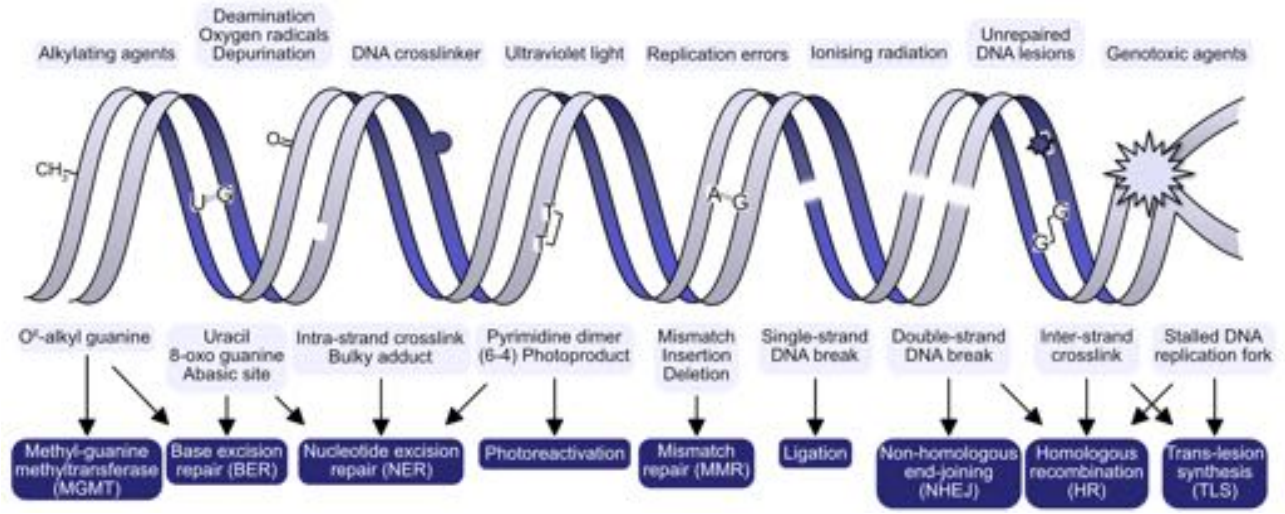

Figure 1. Schematic of DNA damage causes, consequences and repair pathways. Further details are found in the main text.

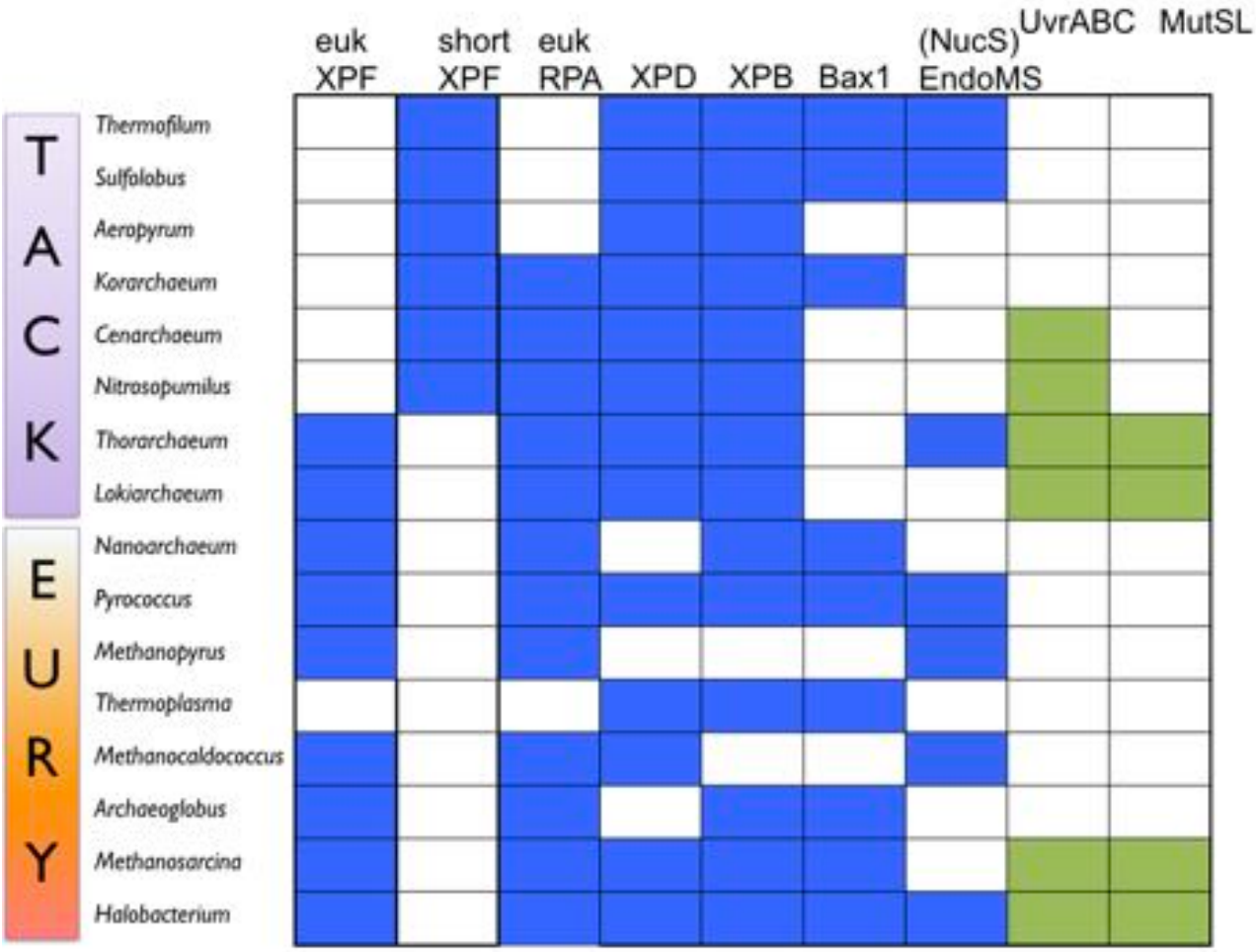

Figure 2. Distribution of DNA repair genes in the archaea. Genus names on the left are organised as members of the TACK superphylum and Euryarchaea. For each genus, a shaded box indicates the presence of the relevant gene. Bacterial genes probably acquired by lateral gene transfer are shown in green, others in blue. Accession numbers are shown in table S1.

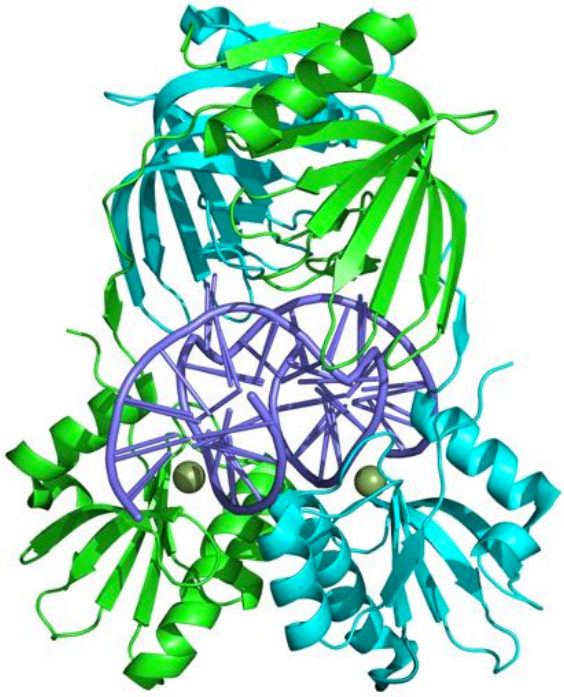

Figure 3. Structure of the EndoMS dimer bound to DNA (Nakae *et al.*, 2016). EndoMS subunits are shown in cyan and green, with the N-terminal dimerization domain at the top and the C-terminal nuclease domains at the bottom. The two catalytic sites are indicated by the green spheres that denote the active site Magnesium ions. The DNA duplex (blue) is distorted by EndoMS binding and two bases are flipped out.

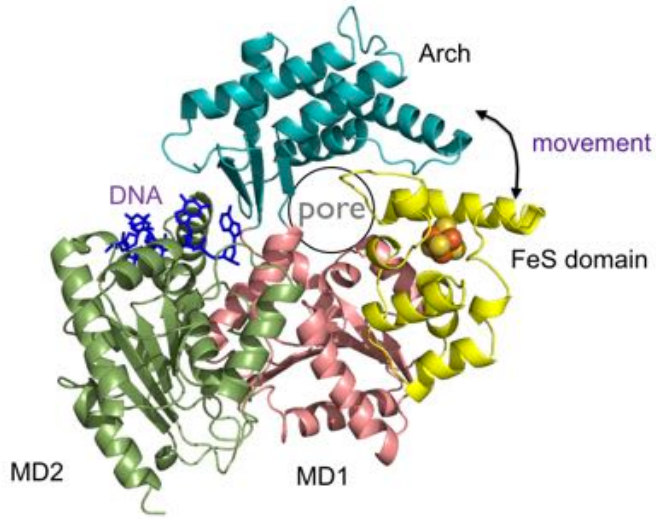

Figure 4. Structure of XPD from *T. acidophilum* (Constantinescu-Aruxandei *et al.*, 2016). Motor domain 1 (MD1) is pink, Motor domain 2 (MD2) green, the FeS domain yellow and the Arch domain teal. The covalently bound 5 nt of DNA is shown in blue. The interface between the Arch and FeS domains that must open is indicated, and the central pore through which DNA must pass is labelled.

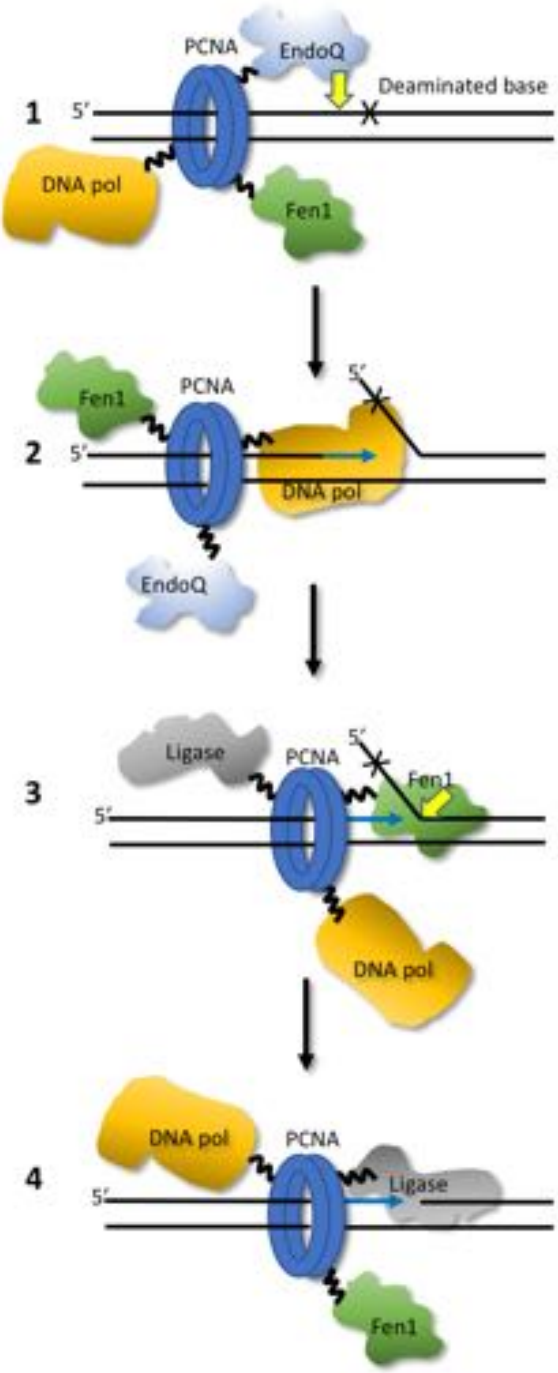

Figure 5. EndoQ pathway for Alternative Excision Repair of deaminated DNA. 1. EndoQ detects deaminated base, cleaving DNA backbone on 5' side. 2. DNA polymerase extends the 3' end of DNA, displacing a DNA flap including the lesion. 3. Fen1 removes the 5' flap, leaving nicked DNA that is ligated by DNA ligase (4). The process may be coordinated by PCNA, which interacts with each of the enzymes. Similar pathways may pertain for other glycosylases and DNA repair nucleases that interact with PCNA. It is not yet clear whether this "molecular toolbelt" view of PCNA reflects reality, as protein partners will associate and dissociate in dynamic equilibrium.

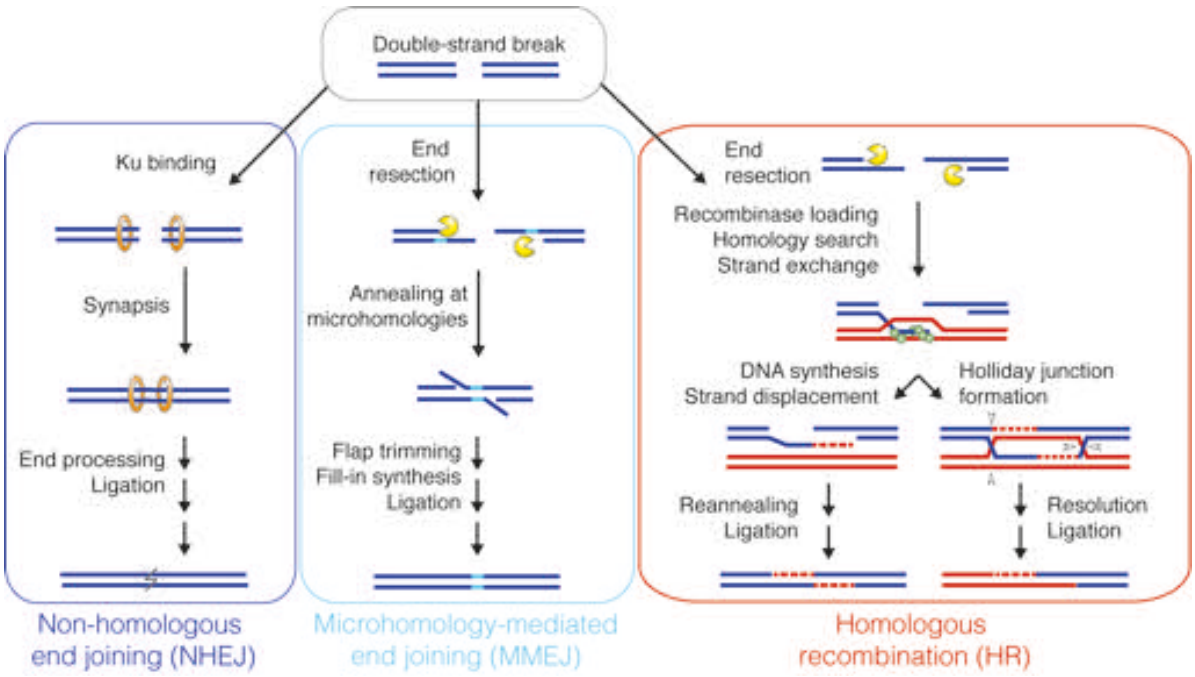

Figure 6. Pathways of DSB repair. Double-strand DNA breaks are repaired by non-homologous end-joining (NHEJ), microhomology-mediated end-joining (MMEJ) or homologous recombination (HR).

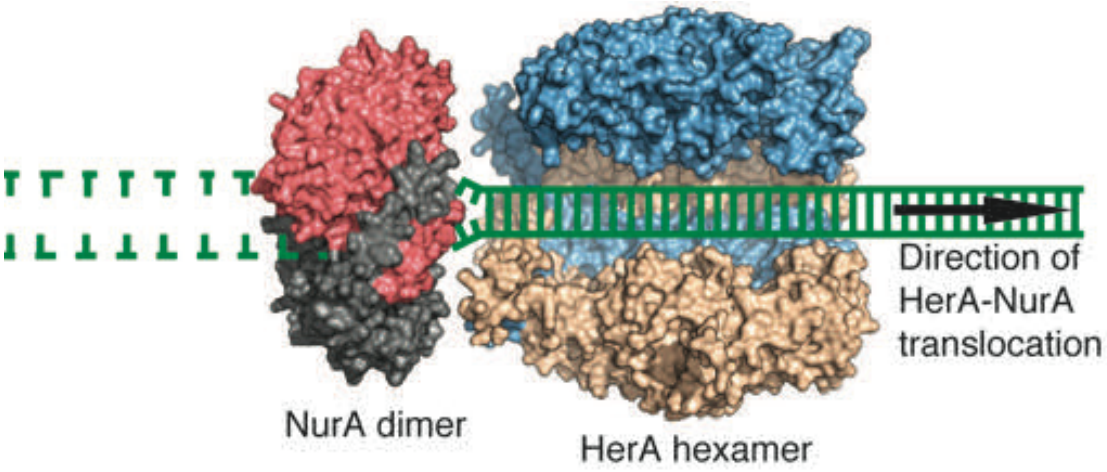

Figure 7. NurA-HerA complex. Model for how HerA and NurA might process DNA ends. Double-stranded DNA is channelled through HerA helicase and unwound by the ploughshare motif in NurA. Both DNA strands are degraded by the NurA nuclease. From (Rzechorzek *et al.*, 2014).

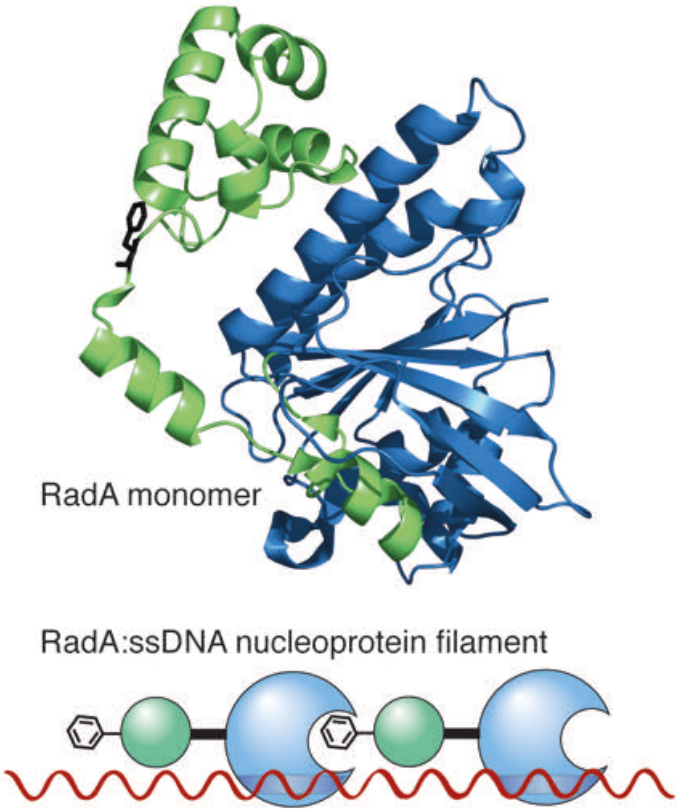

Figure 8. RadA recombinase. Rad:DNA nucleoprotein filament formation by insertion of phenylalanine into binding pocket of an adjacent RadA monomer. From (Wardell *et al.*, 2017).

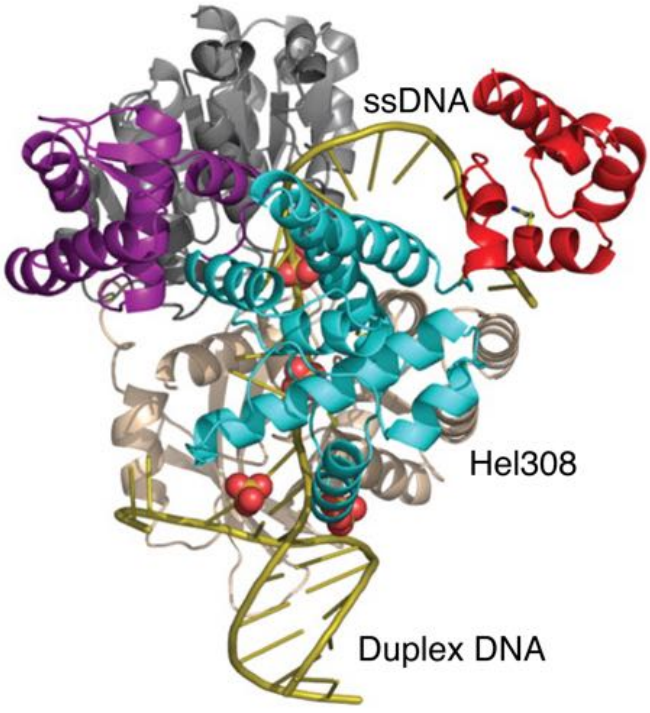

Figure 9. Hel308 helicase. DNA duplex is unwound into single-strands by Hel308. From (Richards *et al.*, 2008).

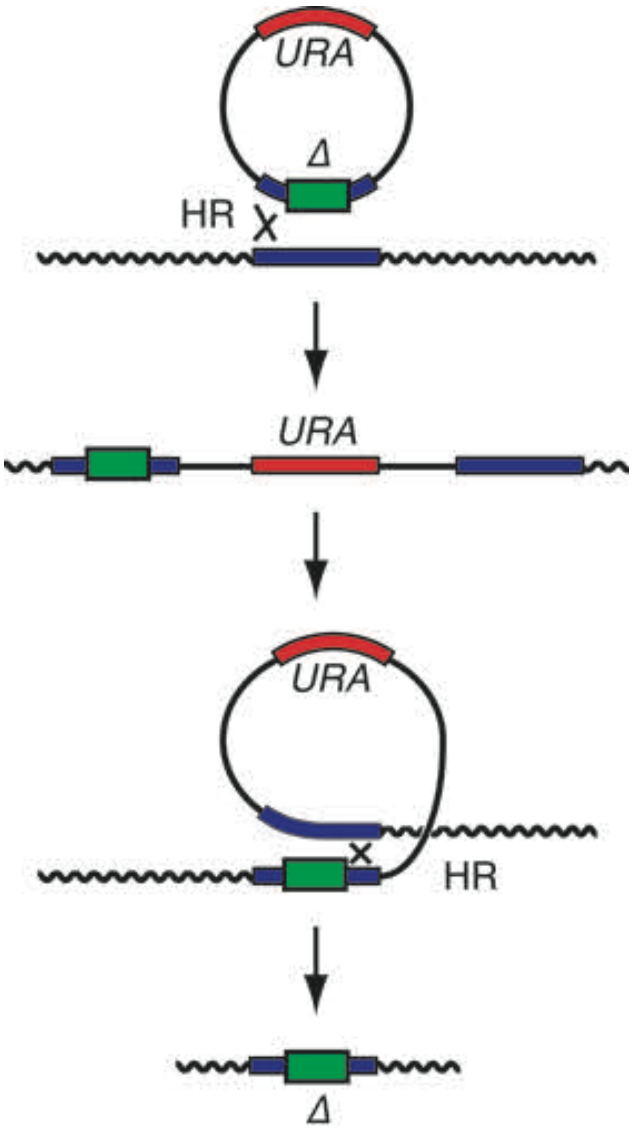

Figure 10. Typical strategy for gene deletion in archaea using HR. A plasmid with flanking homology is used to delete and replace a target gene with a selectable marker (Δ). A second marker for uracil biosynthesis (URA) is used for selection and counter-selection (using 5-FOA) of cells that have undergone HR as indicated.