

Single-stranded DNA binding proteins in the Archaea

Najwa Taib ^{1,2}, Simonetta Gribaldo ¹ and Stuart A. MacNeill ³

¹Unit Evolutionary Biology of the Microbial Cell

Department of Microbiology

Institut Pasteur

25-28 Rue du Dr Roux

75015 Paris

France

²Hub Bioinformatics and Biostatistics

Department of Computational Biology

Institut Pasteur

25-28 Rue du Dr Roux

75015 Paris

France

³Biomedical Sciences Research Complex

School of Biology

University of St Andrews

North Haugh

St Andrews

KY16 9ST

UK

Corresponding author:

Dr Stuart A. MacNeill

E-mail: stuart.macneill@st-andrews.ac.uk

Abstract

Single-stranded (ss) DNA binding proteins are found in all three domains of life where they play vital roles in nearly all aspects of DNA metabolism by binding to and stabilising exposed ssDNA and acting as platforms onto which DNA processing activities can assemble. The ssDNA binding factors SSB and RPA are extremely well conserved across bacteria and eukaryotes, respectively, and comprise one or more OB fold ssDNA binding domains. In the third domain of life, the Archaea, multiple types of ssDNA binding protein are found with a variety of domain architectures and subunit compositions, with OB fold ssDNA binding domains being a characteristic of most, but not all. This chapter summarises current knowledge of the distribution, structure and biological function of the archaeal ssDNA binding factors, highlighting key features shared between clades and those that distinguish the proteins of different clades from one another. The likely cellular functions of the proteins are discussed and gaps in current knowledge identified.

Running head: Archaeal RPA/SSB proteins

1. Introduction

Single-stranded DNA binding proteins play essential roles in almost all aspects of DNA metabolism in all three domains of life: bacteria, eukaryotes and archaea (1). In bacteria, the major single-stranded DNA binding protein is SSB, whereas in eukaryotes, this role is performed by the trimeric RPA (replication protein A) complex. SSB and RPA are characterised by the presence of one or more OB (oligosaccharide-oligonucleotide binding) fold domains (2). OB fold domains span 75-150 amino acids and consist of a five-stranded β -sheet that is coiled to form a closed β -barrel structure, often capped by an α -helix. The heterogeneity in length reflects the presence of variable loop regions generally located between the β -strands. Single-stranded DNA binding by the SSB and RPA OB folds is not sequence-specific, reflecting the need for the SSB/RPA proteins to engage, transiently, with single-stranded DNA in a wide variety of functional contexts (2). In bacteria, the archetypal *Escherichia coli* SSB is a 255 amino acid protein that comprises a single N-terminal OB fold domain followed by a relatively short, flexible C-terminal tail (3) that mediates protein-protein interactions (**Figure 1**). Individual bacterial SSB proteins of this type (i.e. with a single OB fold domain per protein) assemble into homotetrameric complexes. In contrast to the situation in *E. coli*, in certain other bacterial lineages (*Deinococcus* and *Thermus* being the best studied) SSB proteins with a tandem pair of OB fold domains are found (4). The second OB fold domain is followed by a flexible C-terminal tail related in sequence to that of *E. coli* SSB. These proteins assemble to form homodimers that, like *E. coli* SSB, contain four OB fold domains in total (5). In eukaryotes, replication protein (RPA) is the major single-stranded DNA binding factor (6,7). RPA is a heterotrimeric complex made up of the RPA1, RPA2 and RPA3

proteins (also known as RPA70, RPA32 and RPA14, respectively). **Figure 2A** shows the domain organisation of the well-studied RPA from the budding yeast *Saccharomyces cerevisiae*. RPA1 contains four OB fold domains (designated DBD-A, DBD-B, DBD-C and DBD-F, where DBD is short for DNA binding domain), RPA2 contains one OB fold (DBD-D) and RPA3 one OB fold also (DBD-E). DNA binding domains DBD-A, DBD-B, DBD-C and DBD-D are primarily responsible for ssDNA binding by RPA. DBD-E plays a structural role at the heart of the RPA complex while DBD-F functions as a site of protein-protein interaction. RPA2 also possesses an extended N-terminal domain that is the target for regulatory phosphorylation and a C-terminal winged helix (wH) domain, both of which are also involved in protein-protein interactions (8). A number of partial RPA structures have been solved, both with and without bound ssDNA, such as the *Ustilago maydis* DBD-A/B/C/D/E structure shown in **Figure 2B** (9).

RPA is essential for eukaryotic chromosome replication and plays important roles in a variety of DNA repair processes (6,7). In addition to RPA, variant RPA-like complexes are found in a number of lineages (10-12), including mammals and plants, and the related CST (Cdc13-Stn1-Ten1) complex plays an important role in telomere maintenance (13). Single OB fold SSBs (hSSB1 and hSSB2) have also been identified and characterised in mammals and shown to have roles in DNA damage repair (14-16).

This chapter provides an overview of the current state of knowledge of the biology of single-stranded DNA binding proteins in the Archaea. Archaeal organisms make up the third domain of life on Earth, are ubiquitous in nature, make up an estimated 20% of the planet's biomass, and play important roles in biosphere and atmosphere. OB fold-containing RPA- and/or SSB-like ssDNA binding factors have been identified in a number of well-studied archaeal species, with the RPA-like proteins

displaying a range of domain architectures with varying numbers of OB folds (**Figures 3, 4**). Biochemical characterisation of these proteins has allowed conclusions to be drawn regarding their ssDNA binding properties *in vitro*, while genetic analysis has inferred vital roles for the proteins in archaeal DNA replication and repair *in vivo*. Here, we present an up-to-date phylogenetic analysis of the distribution of RPA/SSB proteins across the archaeal domain and provide a comprehensive review of current knowledge of the properties of these proteins. Taken together, the data provides an excellent foundation for further studies of the biology of ssDNA binding in the archaea.

2. Phylogenetic distribution of archaeal SSB/RPA proteins

Figure 3 illustrates the phylogenetic distribution of OB fold-containing SSB/RPA proteins across the Archaea (17) and provides a number of important insights into single-stranded DNA binding capacity in different lineages. Individual SSB/RPA proteins can be classified into four groups (shown in schematic form in **Figure 3A**): RPA41 and RPA32 proteins, readily recognisable by the presence of characteristic C-terminal zinc finger OB fold and winged helix-turn-helix (wH) domains, respectively; single OB fold proteins, including both RPA14 and SSB proteins; and a broad class of multiple OB fold RPA proteins containing neither a zinc finger nor a wH domain.

The archaeal RPA41 proteins are structurally and evolutionarily related to eukaryotic RPA1 and are defined by the presence of one or several OB folds and a C-terminal C4 or C3H zinc finger motif. The first archaeal RPA to be characterised, from *Methanocaldococcus jannaschii* (18), was a member of this group and was originally reported to comprise four tandem OB fold domains followed by a C4 zinc

finger motif (18,19). More recent analysis has shown that the zinc finger is in fact embedded in a fifth OB fold that is related to the DBD-C OB fold in eukaryotic RPA1 (SM, unpublished). The presence of a DBD-C-like OB fold is a defining characteristic of all archaeal RPA41 proteins. To date, RPA41 proteins have been found to be encoded by almost all archaeal lineages, with exceptions being the Micrarchaeota, the closely related Nanoarchaeota and Parvarchaeota, and almost all lineages within the TACK superphylum (originally named for the Thaumarchaeota, Aigarchaeota, Crenarchaeota and Korarchaeota) (**Figure 3B**).

The archaeal RPA32 proteins are structurally and evolutionarily related to eukaryotic RPA2 and are defined by a structure that comprises one or two OB folds followed by the winged helix-turn-helix (wH) domain. Very few of these proteins have been characterised biochemically and in those that have, the presence of the C-terminal wH domain and the similarity to RPA2 went entirely unnoticed (20,21).

Like the RPA41 proteins, RPA32 proteins are widespread across the archaea, although it is noticeable that a small number of lineages that encode RPA41 do not encode RPA32 and that an even smaller number of lineages (specifically, the Nanoarchaeota and Parvarchaeota) appear to encode an RPA32 but not an RPA41 (**Figure 3B**). These observations suggest that the RPA41 and RPA32 proteins are each capable of functioning independently as single-stranded DNA binding proteins, even if they are likely to form dimeric RPA41-RPA32 complexes in most species (discussed further below).

Single OB-fold proteins are also widespread and fall into at least two groups (**Figure 3**). A number of species of the Thermococcales have been shown to encode trimeric RPA complexes that are very similar in organisation to eukaryotic RPA (see section 3.4 below). In these cases, the RPA41, RPA14 and RPA32 proteins (akin to eukaryotic RPA1, RPA3 and RPA2, respectively) are encoded by adjacent genes and co-

transcribed (22,23). This genetic organisation is seen in the Theionarchaea and in some species of Methanococcales (section 3.1.1), suggesting that these organisms also encode a trimeric RPA (**Figure 3**). However, in addition to this, many archaeal lineages carry genes that encode single OB fold proteins that are not linked genetically to those encoding RPA41 and RPA32 (and where the genes encoding RPA41 and RPA32, if both are present, are not necessarily linked to one another). In the absence of biochemical evidence, it is not possible to be determined with certainty if these single OB-fold proteins are the RPA14 component of a trimeric RPA, or if they act independently as monomeric or possibly homomultimeric SSBs. In the Crenarchaeota, which lack both RPA41 and RPA32, the single OB fold SSB protein has been extensively characterised (see section 4). SSB proteins of this type appear to be a shared feature of the TACK archaea (with exception of the Korarchaeota) that lack RPA41 and RPA32.

The final type of archaeal RPA proteins possess multiple OB-folds and can be distinguished from the RPA41 and RPA32 proteins by the absence of a C-terminal DBD-C-like zinc finger-containing OB fold or a winged helix (wH) domain, respectively. These proteins are scattered across a range of species (**Figure 3**), with no obvious pattern in relation to the presence or absence of the other three types of RPA/SSB (RPA41, RPA32, or single OB fold proteins).

The following sections summarise current knowledge of the biochemical properties and biological function of studied archaeal RPA/SSB proteins. With one exception (see section 3.1.1), we refer to these proteins by their original given names (for example MthRPA or HvoRpap1) rather than renaming them to take into account their new-found relatedness to RPA41 or RPA32.

3. Euryarchaeal RPA proteins

3.1 RPA proteins in class I methanogens

The Euryarchaeota represent a broad range of archaeal species with highly diverse characteristics, including methanogens, thermophiles, and halophiles (**Figure 3A**). The class I methanogens form the Methanomada superclass (24). These organisms are obligate anaerobes. RPA proteins from three different class I methanogens, representing three different taxonomic groupings, have been characterised biochemically: *Methanocaldococcus jannaschii* (formerly *Methanococcus jannaschii*), a member of the order Methanococcales and the first archaeal organism to have its genome completely sequenced (25), *Methanothermobacter thermoautotrophicus* (formerly *Methanobacterium thermoautotrophicum*), a member of the Methanobacteriales, and *Methanopyrus kandleri*, a member of the Methanopyrales. Although broadly similar in structure, these RPA proteins display <30% pairwise sequence identity and differ significantly in size and in the number of OB folds that each possesses (**Figure 4**).

3.1.1 *Methanocaldococcus jannaschii*

M. jannaschii encodes a single putative single-stranded DNA binding protein, originally designated MjaSSB (18) but referred to here as MjaRPA to avoid confusion with single OB fold proteins such as *E. coli* SSB or the crenarchaeal SSB proteins described below (section 4). MjaRPA is 645 amino acids in length and was initially thought to comprise four tandem OB fold domains followed by a C4 zinc finger motif (18,19) (**Figure 4A**). Subsequent bioinformatic analysis (SM, unpublished) identified a fifth OB fold that includes the zinc finger motif and which is related to the DBD-C OB fold in eukaryotic RPA1 (shaded light grey in **Figure 4A**), allowing

MjaRPA to be classified as an RPA41 protein. *M. jannaschii* also encodes an RPA32 homologue (17), and it seems reasonable to suggest that, *in vivo*, MthRPA will form a heterodimeric RPA complex with this protein. As yet, however, the RPA32 protein has not been characterised, either in isolation or with MjaRPA.

Biochemical characterisation of recombinant MjaRPA demonstrated that the protein is able to bind ssDNA with an affinity similar to human RPA and with a binding site of 15-20 nucleotides. No further biochemical or functional analysis of the *M.*

jannaschii MjaRPA protein has been reported but the structure of the second of the five MjaRPA OB fold domains (labelled OB2 in **Figure 4**, amino acids 172-270) was determined by X-ray crystallography in the course of a structural genomics project (**Figure 5A, Table 1**). In addition to this, the structure of the second OB fold of the closely related (44% identical) RPA protein from *Methanococcus maripaludis* (MmpRPA1) has been determined by both X-ray crystallography and NMR (**Table 1**). Evidence obtained from saturation transposon insertion mutagenesis of *M. maripaludis* (26) suggests that the MmpRPA1 protein is essential for cell viability, most likely suggesting a key role for the protein in chromosome replication.

In addition to MmpRPA1, *M. maripaludis* appears also to encode a single-stranded DNA binding factor that is akin to the trimeric RPAs found in *Pyrococcus furiosus* and *Thermococcus kodakarensis* (discussed in detail in section 3.4 below, see **Figure 4**) and in eukaryotic cells (**Figure 2**). Three adjacent genes encode the RPA41, RPA14 and RPA32 type proteins MmpRPA3 (Mmp1022), Mmp0121 and Mmp0120, respectively. MmpRPA3 is a 302 amino acid protein containing two OB folds, the second of which is of the DBD-C type with integral C4 zinc finger, while Mmp0121 and Mmp0120 are short proteins (105 and 128 amino acids, respectively) comprising a single OB fold only. Notably, Mmp120 appears to lack the wH domain found in other RPA32

proteins (**Figures 2, 4**). Unlike MmpRPA1, MmpRPA3 is apparently not required for cell viability (26). Equivalent genes are not found in *M. jannaschii*.

3.1.2 *Methanothermobacter thermautotrophicus*

The MthRPA protein is an RPA41 type protein, 792 amino acids in length (i.e. ~150 amino acids longer than MjaRPA) and comprising six tandem OB fold domains, with the sixth (shaded grey in **Figure 4A**) containing an integral C4 zinc finger like eukaryotic DBD-C (18,27) (SM, unpublished results). The structure of the first of the OB fold domains (amino acids 62-167) has been solved by NMR (see **Figure 5B, Table 1**). Only limited information is available regarding MthRPA function, but unlike *M. jannaschii*, *M. thermautotrophicus* does not appear to encode an RPA32 type protein. *In vitro*, the MthRPA protein, which is able to bind ~15-20 nucleotides of ssDNA, interacts directly with the *M. thermautotrophicus* family B DNA polymerase MthPolB to specifically inhibit DNA polymerase activity without affecting 3'→5' proofreading exonuclease activity (27). MthPolB activity is not inhibited by *E. coli* SSB and other DNA polymerases (*Taq* DNA polymerase, *Pfu* DNA polymerase) are not inhibited by MthRPA, highlighting the specificity of the MthRPA-MthPolB interaction (27). MthRPA also interacts with the DNA repair helicase MthHel308 via an arginine residue located in a conserved sequence motif at the C-terminus of the latter (28). Binding of MthRPA does not appear to affect MthHel308 unwinding activity, leading to the suggestion that MthRPA may have a role in targeting MthHel308 to its substrates.

3.1.3 *Methanopyrus kandleri*

The third RPA to be characterised from a class I methanogen is that of *M. kandleri*, a hyperthermophilic organism that can grow at temperatures $> 100^{\circ}\text{C}$ (29). MkaRPA is a 432 amino acid RPA41 type protein comprising three OB folds, with the most C-terminal being of the eukaryotic RPA1 DBD-C type and containing an integral zinc finger (see **Figure 4**) (SM, unpublished). MkaRPA has been shown to bind single-stranded DNA with high affinity in EMSAs, with FRET analysis suggesting two different binding modes (compacting then stretching) as the protein to DNA concentration increases. Gel filtration analysis of recombinant MkaRPA indicates that the protein is a trimer in solution. Similar to the situation with MthRPA, full-length MkaRPA (but not a C-terminally truncated protein that removes the DBD-C-like OB fold) inhibits the primer extension activity of the *M. kandleri* family B DNA polymerase MkaPolBI *in vitro* (29). As in the case of *M. jannaschii*, *M. kandleri* encodes an RPA32 type protein (17) that may interact with MkaRPA to form a dimeric complex *in vivo* but which has not been studied.

3.2 RPA proteins in the class II methanogens

The class II methanogens form the Methanomicrobia class and are phylogenetically distinct from the class I methanogens such as *M. jannaschii* and *M. thermautotrophicus* (24). Indeed, the class II methanogens are more closely related to the non-methanogenic Halobacteria (which, despite their name are not bacteria, but archaea), something that is borne out by consideration of their RPA proteins. To date, with the exception of a single structural study, investigation of the properties of these proteins has been confined to a single representative species, *Methanosarcina acetivorans* (29-33).

3.2.1 *Methanosarcina acetivorans*

M. acetivorans encodes three RPA homologues designated MacRPA1, MacRPA2 and MacRPA3 (29-31). MacRPA1 is a 484 amino acid protein containing four tandem OB fold domains, but no zinc finger motif (**Figure 4**), and is therefore related to the *Haloferax volcanii* Rpa2 protein and *Halobacterium salinarum* Rfa1 proteins discussed below, with the absence of the zinc finger being a shared feature (see section 3.3). Interestingly, the RPA1 protein of the closely related species *Methanococcoides burtonii* (MbuRPA1) contains only three, not four, OB fold domains (33). Multiple sequence alignments point to the protein having arisen due to deletion in sequences encoding the second and third OB folds of a MacRPA1-like ancestral RPA with four OB folds.

Biochemical analysis suggests that recombinant MacRPA1 can exist as both a dimer and a tetramer in solution (30). Deletion of either one or two OB folds from the N-terminal or C-terminal end of the MacRPA1 protein reduces, to varying degrees, but does not abolish, ssDNA binding in EMSA and/or fluorescence polarisation assays, while removal of a third OB fold (leaving only the most N-terminal or most C-terminal fold, OB1 and OB4 in **Figure 4**, respectively) abolishes ssDNA binding altogether. There is clearly a degree of redundancy among the OB folds in these *in vitro* assays, something that is borne out by *in vivo* assays of HvoRpa2 function in *H. volcanii* (discussed in section 3.3.1 below).

In vitro, MacRPA1 is capable of inhibiting the activity of the *M. acetivorans* flap endonuclease (Fen1) homologue MacFEN1 to cleave 20-nucleotide 5' flap structures but cannot inhibit cleavage of 5-nucleotide flaps (32). The ability to inhibit MacFEN1 activity is not shared by truncated MacRPA1 proteins lacking more than just the N- or C-terminal OB fold, or by MacRPA2 or MacRPA3.

MacRPA2 and MacRPA3 are RPA41 type proteins, 417 and 450 amino acids in length and closely related to one another. Each protein comprises an N-terminal tandem pair of OB fold domains followed by a C-terminal DBD-C type OB fold with an integral C3H zinc finger (**Figure 4**) (SM, unpublished). There is no structural information for either of the MacRPA2 or MacRPA3 proteins, but the structure of the first OB fold domain of the *M. mazei* RPA2 protein (MmaRPA2), which is 96% identical to MacRPA2, has been determined by NMR (**Figure 3C, Table 1**).

Recombinant MacRPA2 and MacRPA3 have been reported to exist as dimers in solution, to contain bound zinc, as would be expected, and to specifically bind ssDNA over dsDNA, with MacRPA3's affinity for ssDNA being greater than that of MacRPA2 (30). Both proteins, as well as MacRPA1, are capable of stimulating primer extension by *M. acetivorans* DNA polymerase B (MthPolB). Mutating individual cysteines in the MacRPA3 DBD-C zinc finger to alanine reduces single-stranded DNA binding and the ability of the proteins to stimulate MacPolBI activity *in vitro* but also causes significant structural changes (31), while removing the entire DBD-C region does not greatly alter ssDNA binding (29,30).

Although there is abundant published information regarding the biochemical properties of recombinant MacRPA2 and MacRPA3 proteins in isolation, *M. acetivorans* also encodes unstudied RPA32 proteins (17) and it is highly likely that MacRPA2 and MacRPA3 form heterodimeric RPA41-RPA32 complexes *in vivo*. As a result, as with *M. jannaschii* and *M. kandleri*, discussed above, the earlier biochemical data obtained with recombinant MacRPA2 and MacRPA3 in isolation should be treated with a degree of caution when considering the *in vivo* behaviour of these proteins.

3.3 RPA proteins in the Halobacteriales

The Halobacteria are relatively closely related to Methanosarcinales such as *M. acetivorans*, *M. burtonii* and *M. mazei* (**Figure 3**) and comprise a range of euryarchaeal organisms that inhabit hypersaline environments such as salt lakes and solar salterns and which have found an important place as model systems in archaeal research due to ease with which they can be grown in the lab and their tractability to molecular genetic analysis (34).

3.3.1 *Haloferax volcanii*

3.3.1.1 Rpa1-Rpap1 and Rpa3-Rpap3 complexes

There is substantial similarity between the RPA proteins encoded by the Halobacteriales and the Methanosarcinales. The *H. volcanii* Rpa1 and Rpa3 proteins (hereafter referred to as HvoRpa1 and HvoRpa3, also known as RpaA1 and RpaB1) are RPA41 type proteins and homologues of *M. acetivorans* RPA2 (MacRPA2) and RPA3 (MacRPA3), respectively. HvoRpa1 is a 427 amino acid protein that comprises three putative OB folds, with the most C-terminal being the eukaryotic DBD-C type with integral C3H zinc finger, while at 311 amino acids, HvoRpa3 is somewhat shorter, comprising only two OB folds, the second being the DBD-C type with C3H zinc finger (SM, unpublished) (**Figure 4**). At the chromosome level, the *rpa1* gene overlaps with *rpap1*, which encodes a 623 amino acid RPA32 protein containing an N-terminal OB fold and a C-terminal winged helix (wH) domain, while *rpa3* is located adjacent to *rpap3*, which encodes a 196 amino acid RPA32 protein with an N-terminal OB fold and a C-terminal wH domain (17). Consistent with their classification as RPA41 and RPA32 proteins, HvoRpa1 and HvoRpap1 can be co-

purified from native extracts prepared from overexpressing cells, as can HvoRpa3 and HvoRpap3 (21). There is no evidence for the existence of an RPA14 protein associated with these RPA41-RPA32 complexes, as is the case in *P. furiosus* and *T. kodakarensis* (see section 3.4 below), and the current assumption is that the *H. volcanii* RPAs are heterodimers. No structural information is available for any part of any of the four proteins that make up the two heterodimers, but HvoRpa3 – in isolation, i.e. without the corresponding RPA32 protein HvoRpap3 – has been purified in recombinant form and shown to be able to bind DNA in high salt conditions mimicking the internal salt concentrations of the *H. volcanii* cell (35). The N-terminal OB fold alone has reduced affinity for ssDNA compared to the full-length HvoRpa3 protein, whereas the C-terminal DBD-C type OB fold did not bind ssDNA under the conditions used.

Reverse genetic analysis has allowed initial insight into the cellular functions of the HvoRpa1-HvoRpap1 and HvoRpa3-HvoRpap3 complexes (20,21). The *rpa1-rpap1* and *rpa3-rpap3* operons can be deleted from the chromosome individually, but not at the same time, indicating that HvoRpa1-HvoRpap1 and HvoRpa3-HvoRpap3 share an essential function in the cell. Consistent with this, repression of *rpa3-rpap3* expression in $\Delta rpa1-rpap1$ cells leads to a significant growth retardation (20). Cells lacking the HvoRpa3-HvoRpap3 complex ($\Delta rpa3-rpap3$) display increased sensitivity to UV light and to the DNA damaging agent mitomycin C, while $\Delta rpa1-rpap1$ cells lacking HvoRpa1-HvoRpap1 retain full repair capacity (21). UV exposure causes formation of cyclobutane pyrimidine dimers (CPDs) and pyrimidine 6-4 photoproducts (6-4PPs) in DNA, whereas MMC produces three types of MMC damage: monoadducts, and intra- and inter-strand cross-links. The observed sensitivity of $\Delta rpa3-rpap3$ cells is therefore indicative of the HvoRpa3-HvoRpap3 complex playing a role in ssDNA binding during the repair of these DNA lesions, even if the molecular mechanisms of UV and

MMC damage repair in *H. volcanii* are not fully understood, while the shared essentially of the HvoRpa1-HvoRpap1 and HvoRpa3-HvoRpap3 complexes most likely implies a role in chromosomal DNA replication.

3.3.1.2 Rpa2

The *H. volcanii* Rpa2 protein (HvoRpa2, also known as RpaC) protein is an orthologue of *M. acetivorans* RPA1 (MacRPA1) (**Figure 4**). HvoRpa2 is a 483 amino acid protein and like MacRPA1, comprises a short, conserved N-terminal domain, followed by four OB fold domains, the most C-terminal of which was not immediately apparent at the time of initial publication, but no zinc finger (20). No structural information is available for any part of HvoRpa2.

Using reverse genetic methodology, HvoRpa2 has been shown to be essential for cell viability: the *rpa2* gene cannot be deleted from the chromosome and down-regulation of *rpa2* expression in otherwise wild-type cells results in significant growth retardation (20). Interestingly, this slow growth phenotype can be partially rescued by increased expression of *rpa3-rpap3*, though not *rpa1-rpap1*, indicating that HvoRpa3-HvoRpap3 can partly substitute for Rpa2 *in vivo*. Overexpression of HvoRpa2 leads to enhanced resistance to DNA damage caused by exposure to UV light, 4NQO (4-nitroquinoline 1-oxide), the alkylating agent methyl methanesulfonate (MMS) and the antibiotic phleomycin, which causes DNA strand breaks.

The Rpa2 protein has been used a platform for *in vivo* analysis of the role of individual OB fold domains, with various truncation and deletion derivatives of the *Halogeometricum borinquense* Rpa2 protein (HgmRpa2) being tested for their ability to rescue the slow growth defect resulting from repression of *H. volcanii rpa2*

expression (20). The HgmRpa2 protein is 75% identical to HvoRpa2 and is able to rescue to near wild-type levels. Deletion of the C-terminal OB fold (OB4 in HvoRpa2 in Figure 3) did not affect the ability of the HgmRpa2 protein to rescue the growth defect in the absence of exogenous DNA damage, but did result in increased sensitivity to UV irradiation and MMS exposure. Deletion of OB folds OB1 and OB2 led to somewhat reduced ability to rescue the growth defect, while deletion of OB3 significantly impaired rescue ability. Some rescue was seen when OB1 and OB2 were deleted together, but none when OB1 and OB3, or OB2 and OB3, were deleted.

Taken together, these results indicate a particularly important role for OB3 in Rpa2 function, assisted by OB1, OB2 and OB4. (20). It remains to be seen how these genetic observations correlate with the actual ssDNA binding dynamics.

To gain further insights into the biological role of the essential HvoRpa2 protein, *H. volcanii* cells have been engineered to express a GFP-Rpa2 fusion protein from the native chromosomal locus, with the *gfp:rpa2* open reading frame replacing the endogenous *rpa2* gene (36). Microscopic examination of either exponentially growing or stationary phase cells reveals that the GFP-Rpa2 protein is found at a small number of discrete foci, even in the absence of exogenous DNA damage, with the fact that slightly more foci are observed in exponentially growing cells leading to the suggestion that at least some of these may correspond to activate DNA replication forks (36). UV treatment of cells leads to a marked reduction in the number of GFP-Rpa2 foci, with single large foci being observed in cells treated with higher UV doses. In contrast, treatment of cells with the DNA polymerase inhibitor aphidicolin leads to a doubling in the number of GFP-Rpa2 foci, while treatment with the DNA damaging agent phleomycin leads to an approximate five-fold increase. It is assumed that the patterns observed with different damaging agents reflect the amount of ssDNA available to be bound by HvoRpa2 at each dose (36).

3.3.2 *Halobacterium salinarum*

RPA function has also been investigated in closely related organism *Halobacterium salinarum* (also known as *Halobacterium sp.* NRC-1). *H. salinarum* encodes homologues of HvoRpa1, HvoRpap1, HvoRpa3 and HvoRpap3, designated HsaRfa2, HsaRfa7, HsaRfa3 and HsaRfa8, respectively, as well as an HvoRpa2 homologue, HsaRpa1. The proteins are very similar in size and composition across the species, with the exception of the RPA32 protein HsaRfa7 which is ~160 amino acids shorter in its central section (between the OB fold and the wH domain) than its *H. volcanii* counterpart HvoRpap1 (Figure 3). Biochemical analysis shows that both HsaRfa3 and HsaRfa8 can be purified on the basis of one or other of the proteins, or both, having affinity for ssDNA, but falls short of proving that the two proteins interact with one another, as would be expected by analogy with HvoRpa3 and HvoRpap3 in *H. volcanii* or RPA41-RPA32 more generally (37).

A number of studies have shown that the *rfa3-rfa8* operon is upregulated in response to DNA damage caused by ionising radiation (IR) and UV exposure and that this results in a concomitant increase in Rfa3 and Rfa8 protein levels (38-40). Upregulation is also seen in *H. salinarum* strains selected for their enhanced resistance to IR (37,39,41) and has subsequently been shown to be sufficient in itself for increased IR resistance (37). Gene deletion analysis (42) has further underlined the importance of the Rfa3 and Rfa8 proteins in *H. salinarum*, as strains deleted for either *rpa3* or *rpa8* are sensitive to both IR and UV exposure, with the UV sensitivity mirroring what is seen with $\Delta rpa3$ -*rpa8* strains in *H. volcanii* (where IR sensitivity has not been tested) (21).

In contrast to the situation in *H. volcanii*, deletion of the *H. salinarum rfa1* gene, encoding the homologue of the essential HvoRpa2 protein, is viable, although slow

growing and severely sensitive to ionising radiation, UV and mitomycin C, while deletion of *rfa2*, but not *rfa7*, appears to be lethal, despite the products of these genes likely forming a dimeric complex. This raises the possibility that Rfa2 may have a function distinct from that of the presumed Rfa2-Rfa7 complex. In support of this notion, *rfa2* expression is markedly upregulated in response to UV and MMC treatment whereas *rfa7* expression is not (42). Further analysis will be required to confirm whether this is the case.

3.4 RPA proteins in the Thermococcales

RPA proteins have been characterised from two well-studied representatives of the Thermococcales, *P. furiosus* (22) and *T. kodakarensis* (23). Both organisms encode RPA41, RPA14 and RPA32 proteins that form a heterotrimeric RPA complex, similar to eukaryotic RPA.

In *P. furiosus*, the RPA41 protein is 360 amino acids in length and contains two OB folds, the most C-terminal of which is of the DBD-C type with an integral C3H zinc finger (**Figure 4**) (SM, unpublished). The RPA32 protein is 273 amino acids in length and contains an N-terminal OB fold and a C-terminal wH domain, while the RPA14 protein is 122 amino acids in length and contains a single OB fold (**Figure 4**). The *T. kodakarensis* proteins have identical domain organisation and share 60-70% sequence identity with the *P. furiosus* archetypes. In both species, the three proteins are encoded from a single operon, with the genes arranged in the order *rpa41-rpa14-rpa32* (22,23).

Biochemically, PfuRPA has been shown to bind ssDNA with high affinity and specificity, to stimulate RadA-mediated strand-exchange *in vitro*, and to co-immunoprecipitate from *P. furiosus* cell extracts with PfuRadA, mimicking the RPA-Rad51 interaction seen in eukaryotes (22). Weaker interactions with other

recombination and replication proteins, such as the Holliday junction resolvase Hjc, replication factor C and the DNA polymerase PolD, were also seen. In the related species *P. abyssi*, RPA was also reported to interact on ssDNA with PolD and primase, and to interact with and stimulate RNA polymerase (43).

TkoRPA has also been shown to bind ssDNA with high affinity and specificity, and to be able to relieve DNA polymerase pausing in *in vitro* reactions by resolving DNA template secondary structure while at the same time reducing DNA polymerase processivity (23). TkoRPA is also reported to be able to interact in solution with both *T. kodakarensis* DNA polymerases TkoPolB and TkoPolD. Like PfuRPA and PabRPA, TkoRPA may also interact with primase, TkoRadA, TkoRad50 and reverse gyrase (23,44).

3.5 RPA proteins in the Thermoplasmatales

Ferroplasma acidarmanus is a representative of the Thermoplasmatales, a clade of acidophilic euryarchaea. *F. acidarmanus* was originally reported to encode two OB fold containing ssDNA binding proteins, designated FacRPA1 and FacRPA2 (29). FacRPA1 is a 369 amino acid RPA41 protein possessing three OB folds, the third of which is of the DBD-C type with integral zinc finger, while FacRPA2 resembles a bacterial SSB protein, with a single OB fold and short C-terminal tail (discussed further in section 4 below) (**Figure 4**). As with several of the species discussed above, *F. acidarmanus* also encodes an unstudied RPA32 protein, with a single OB fold and C-terminal wH domain, encoded by the gene downstream of that encoding FacRPA1, suggesting the presence of a heterodimeric RPA complex (SM, unpublished). Biochemical analysis has shown that both FacRPA1 (in isolation, i.e. in the absence of its cognate RPA32) and FacRPA2 are able to bind ssDNA with high

affinity and specificity. FacRPA1, but not FacRPA2, is also able to stimulate the *in vitro* activity of the *M. acetivorans* PolB polymerase to overcome pausing on DNA templates (29) and both proteins, but particularly FacRPA2, have been shown to be able to stimulate efficient unwinding of forked DNA substrates by the *F. acidarmanus* XPD helicase (45).

4. Crenarchaeal SSB proteins

The Crenarchaeota are a constituent phylum of the TACK superphylum (46) (**Figure 3A**). Perhaps the best characterised crenarchaeal organisms belong to the Sulfolobales, an order that includes the genus *Sulfolobus*. Various *Sulfolobus* species have been used as models for diverse aspects of archaeal biology, but *S. solfataricus* in particular has found a role as a key model organism (34).

4.1 *Sulfolobus solfataricus*

The *S. solfataricus* SSB protein (SsoSSB) was the first crenarchaeal single-stranded DNA binding protein to be biochemically characterised (47,48) and also the first to have its three-dimensional structure determined (49,50). SsoSSB was initially identified in *S. solfataricus* cell extracts on the basis of its ability to bind to ssDNA and is an abundant, 148 amino acid protein comprising an N-terminal OB fold domain (residues 1-114) and a flexible C-terminal tail rich in glycine, arginine and glutamate residues (48). Overall, this simple structure resembles that of the bacterial SSB proteins. Indeed, despite having very limited primary sequence similarity to *E.coli* SSB, the SsoSSB protein is able to support growth of a temperature-sensitive

E. coli ssb-1 strain at its restrictive temperature, and is capable of stimulating DNA strand exchange by *E. coli* RecA (47).

The oligomeric state of SsoSSB has been the subject of much discussion, with both monomeric (50,51) and tetrameric (47,52) forms being proposed and no straightforward way to reconcile the differences in the published observations. It is agreed however that SsoSSB binds ~5 nucleotides of ssDNA per monomer and with high selectivity for ssDNA over dsDNA. Interestingly, recent results indicate that in addition to binding ssDNA, SsoSSB is also able to bind with high affinity to RNA (53), although the biological significance of this remains to be elucidated.

The structure of the N-terminal OB fold domain of SsoSSB has been solved by X-ray crystallography (49,50) and in complex with ssDNA by NMR (**Figure 6, Table 1**) (51,54). Structural comparisons with *E. coli* SSB and eukaryotic RPA point to the SsoSSB OB fold being more similar to the latter, specifically to the DBD-B OB fold (**Figure 2**) (50). Binding relies upon base stacking involving three conserved aromatic residues Trp56, Trp75 and Phe79, and is unidirectional, with each SsoSSB protein binding ssDNA with the same polarity with respect to the 5' and 3' ends of the DNA (51,54). A detailed biophysical analysis of SsoSSB-ssDNA binding has been reported, again highlighting similarities in the dynamics of binding by SsoSSB and DBD-B, rather than *E. coli* SSB (55).

In common with other SSBs, SsoSSB is able to destabilise double-stranded DNA, a property that is enhanced by the presence in the DNA of single mismatches or more complex lesions such as CPDs (cyclobutene pyrimidine dimers), the result of UV damage to double-stranded DNA (56). These observations suggest a potential role for SsoSSB in recognising DNA lesions *in vivo*, melting the damaged DNA and acting as a platform for recruitment of necessary repair factors. Several proteins have been identified as interacting with SsoSSB-coated ssDNA including the XPB helicase

Rad50, topoisomerase 1, RNA polymerase (RNAP) and reverse gyrase, although some of these proteins may be interacting with the ssDNA rather than with SsoSSB (56). Both RNA polymerase and reverse gyrase had previously been shown to interact directly with SsoSSB. Interaction with RNAP stimulates transcription (57), while interaction with reverse gyrase stimulate all the steps in *in vitro* reverse gyrase activity assays: DNA binding, cleavage, strand passage and ligation (58).

Given the undoubted importance of ssDNA binding proteins in all three domains of life, it would be reasonable to expect that the crenarchaeal SSB would be essential for cell viability. The essentiality of SSB has been examined in two closely related *Sulfolobus* species, *S. islandicus* and *S. acidocaldarius*, but with sharply contradictory results. *S. islandicus* *ssb* gene deletions are absent from genome-wide transposon insertion libraries and the gene cannot be deleted in three different *S. islandicus* strains (59). These results strongly suggest that SisSSB is essential for cell viability, as would be expected. In sharp contrast however, recent evidence suggests that the *S. acidocaldarius* SSB protein (SacSSB) is not required for cell viability as the *ssb* gene can apparently be deleted from the chromosome and viable Δ *ssb* progeny obtained (60). Cells deleted for *ssb* grow indistinguishably from wild-type at temperatures ranging from 60-80°C, but display cold sensitivity at 50-55°C, heightened sensitivity to novobiocin at 55°C, and sensitivity to heat shock at 90°C. If confirmed, the *S. acidocaldarius* observations raise very significant questions about the cellular functions of the SacSSB protein and how cells are able to cope in its absence.

4.2 Displacement of SSB in the Thermoproteales

The Thermoproteales are a clade of the Crenarchaeota that includes well-studied archaeal organisms such as *Thermoproteus tenax* and *Pyrobaculum aerophilum*.

Remarkably, unlike all other members of the Crenarchaeota, the majority of

sequenced Thermoproteales genomes do not encode a recognisable crenarchaeal-like SSB or euryarchaeal-like RPA. Instead, these organisms encode a distinct ssDNA-binding protein known as ThermoDBP, exemplified by the *T. tenax* protein referred to here as TteThermoDBP (61). The gene encoding the 196 amino acid ThermoDBP protein appears to have replaced the ancestral Thermoproteales SSB gene in an example of non-orthologous gene displacement. First identified in a screen for ssDNA binding proteins in *T. tenax* cell extracts, recombinant TteThermoDBP has been shown to bind strongly to ssDNA, weakly to ssRNA and not at all to duplex DNA, consistent with it acting as an SSB *in vivo* (61).

The structure of the N-terminal ssDNA binding domain (amino acids 10-148) of TteThermoDBP has been solved by X-ray crystallography, revealing a compact globular domain comprising four α -helices and a four-stranded antiparallel β -sheet (**Figure 7A**, Table 1). A putative DNA binding cleft has also been identified, consisting of a solvent-exposed cleft lined along its length by hydrophobic residues and with positively charged residues at its outer edge (61), and ideally structured for nucleobase and backbone interactions, respectively. Absent from the crystal structure is the C-terminal region of the protein which is helical in nature, possibly forms a basic leucine zipper structure, and is responsible for the dimerization of the recombinant protein (61).

In addition to ThermoDBP, two groups of ThermoDBP-related proteins (ThermoDBP-RPs) have been identified on the basis of sequence similarity to the N-terminal ssDNA binding domain of TteThermoDBP (61). Representatives of the first group, termed ThermoDBP-RP1s, are found in diverse crenarchaeal species including *S. solfataricus* and euryarchaeal species such as *P. furiosus* and *T. kodakarensis*. Representatives of the second group, termed ThermoDBP-RP2s, are

found in crenarchaea such as *Aeropyrum pernix* and euryarchaea such as *Archaeoglobus profundus*.

Biochemical analysis of purified recombinant ThermoDBP-RP1 and ThermoDBP-RP2 proteins has shown that *A. pernix* ThermoDBP-RP2 binds with high affinity to short (21 nt) mixed sequence or homo-pyrimidine ssDNAs, while not binding, or binding very weakly, to homo-purine ssDNA, dsDNA, or single- or double-stranded RNAs of a similar length (62). *P. furiosus* ThermoDBP-RP1 failed to bind to any of these short substrates, but bound weakly to a longer (45 nt) ssRNA. While earlier results had suggested that *S. solfataricus* ThermoDBP-RP1 protein was associated with the box C/D small RNAs and the 30S ribosome subunit, *in vitro* interaction of *P. furiosus* ThermoDBP-1 with box C/D small RNAs and to ribosomes was weak at best, leaving the cellular role of this protein unclear (62).

The three-dimensional structures of the PfuThermoDBP-RP1 and ApeThermoDBP-RP2 protein have been determined by X-ray crystallography, in both cases revealing an N-terminal domain structure similar to that of *T. tenax* ThermoDBP (61) (**Figure 7A**). In the case of PfuThermoDBP-RP1, the N-terminal domain is followed by amphipathic α -helix. PfuThermoDBP-RP1 is a tetramer in solution: in the crystal structure, four of these helices are seen to come together to form an anti-parallel four-helix bundle with a pair of N-terminal domains located at either end of this central rod-like structure (**Figure 7B**) (62).

The structure of the *A. pernix* ThermoDBP-RP2 is also tetrameric, comprising a dimer of intertwined dimers (**Figure 7C**). Each monomer contains a ThermoDBP-like N-terminal domain, but also a globular C-terminal domain comprising a five-stranded antiparallel β -sheet flanked by two α -helices, connected to the N-terminal domain by a bent linker helix. As with PfuThermoDBP-RP1, ApeThermoDBP-RP2 is a tetramer in solution. The structure of the ApeThermoDBP-RP2 protein bound to ssDNA has

also been determined. Binding occurs within a central tunnel-like structure that spans multiple subunits and results in significant distortion of the DNA backbone without any change to protein conformation (**Figure 7C**). The internal dimensions of the tunnel and the required distortion of the ssDNA backbone are not compatible with dsDNA binding, nor with binding to homo-purine ssDNA (see above). Superimposition of the PfuThermoDBP-RP1 and ApeThermoDBP-RP2 structures reveals that the potential DNA binding surface is conserved but that in PfuThermoDBP-RP1, this is occluded by the C-terminal helical tail of another monomer, providing an explanation for the lack of ssDNA binding observed *in vitro* and at the same time suggesting a simple mechanism for activation of PfuThermoDBP-RP1 ssDNA binding by re-orientation of the C-terminal helical tail (62).

Despite this detailed structural knowledge, the cellular functions of the ThermoDBP-RPs, and how these relate to the function of the archetypal ThermoDBP (presumed to be a stand-in for the absent SSB), remain unclear. It is not known if the ThermoDBP-RPs are essential for cell viability, for example, whether they play non-essential roles in cellular DNA transactions, or whether they act in concert with other cellular factors. Further work is clearly required to answer these questions to get a full understanding of ThermoDBP and ThermoDBP-RP functions.

5. Summary

Since the discovery of the first archaeal RPA-like proteins in 1998 (19), considerable progress has been made in identifying and characterising ssDNA binding proteins with diverse structural characteristics in a broad range of archaeal organisms. Biochemical analysis, primarily of proteins expressed and purified in recombinant

form, has identified core ssDNA binding properties, whilst molecular genetic analysis, most notably in the halophilic archaea, has led to significant new insights into biological function. In the future, technological advances, such as the application of super-resolution microscopy to archaeal systems (36), will allow a more detailed analysis of RPA/SSB behaviour *in vivo*, ultimately leading to a comprehensive understanding of the cellular roles of these proteins across the third domain of life.

References

1. Marceau, A.H. (2012) Functions of single-strand DNA-binding proteins in DNA replication, recombination, and repair. *Methods Mol Biol*, **922**, 1-21.
2. Theobald, D.L., Mitton-Fry, R.M. and Wuttke, D.S. (2003) Nucleic acid recognition by OB-fold proteins. *Annu Rev Biophys Biomol Struct*, **32**, 115-133.
3. Bianco, P.R. (2017) The tale of SSB. *Prog Biophys Mol Biol*, **127**, 111-118.
4. Eggington, J.M., Haruta, N., Wood, E.A. and Cox, M.M. (2004) The single-stranded DNA-binding protein of *Deinococcus radiodurans*. *BMC Microbiol*, **4**, 2.
5. Bernstein, D.A., Eggington, J.M., Killoran, M.P., Misic, A.M., Cox, M.M. and Keck, J.L. (2004) Crystal structure of the *Deinococcus radiodurans* single-stranded DNA-binding protein suggests a mechanism for coping with DNA damage. *Proc Natl Acad Sci U S A*, **101**, 8575-8580.
6. Iftode, C., Daniely, Y. and Borowiec, J.A. (1999) Replication protein A (RPA): the eukaryotic SSB. *Crit Rev Biochem Mol Biol*, **34**, 141-180.
7. Wold, M.S. (1997) Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annu Rev Biochem*, **66**, 61-92.

8. Prakash, A. and Borgstahl, G.E. (2012) The structure and function of replication protein A in DNA replication. *Subcell Biochem*, **62**, 171-196.
9. Fan, J. and Pavletich, N.P. (2012) Structure and conformational change of a replication protein A heterotrimer bound to ssDNA. *Genes Dev*, **26**, 2337-2347.
10. Kemp, M.G., Mason, A.C., Carreira, A., Reardon, J.T., Haring, S.J., Borgstahl, G.E., Kowalczykowski, S.C., Sancar, A. and Wold, M.S. (2010) An alternative form of replication protein A expressed in normal human tissues supports DNA repair. *J Biol Chem*, **285**, 4788-4797.
11. Haring, S.J., Humphreys, T.D. and Wold, M.S. (2010) A naturally occurring human RPA subunit homolog does not support DNA replication or cell-cycle progression. *Nucleic Acids Res*, **38**, 846-858.
12. Mason, A.C., Haring, S.J., Pryor, J.M., Staloch, C.A., Gan, T.F. and Wold, M.S. (2009) An alternative form of replication protein A prevents viral replication in vitro. *J Biol Chem*, **284**, 5324-5331.
13. Flynn, R.L. and Zou, L. (2010) Oligonucleotide/oligosaccharide-binding fold proteins: a growing family of genome guardians. *Crit Rev Biochem Mol Biol*, **45**, 266-275.
14. Li, Y., Bolderson, E., Kumar, R., Muniandy, P.A., Xue, Y., Richard, D.J., Seidman, M., Pandita, T.K., Khanna, K.K. and Wang, W. (2009) HSSB1 and hSSB2 form similar multiprotein complexes that participate in DNA damage response. *J Biol Chem*, **284**, 23525-23531.
15. Richard, D.J., Bolderson, E., Cubeddu, L., Wadsworth, R.I., Savage, K., Sharma, G.G., Nicolette, M.L., Tsvetanov, S., McIlwraith, M.J., Pandita, R.K. *et al.* (2008) Single-stranded DNA-binding protein hSSB1 is critical for genomic stability. *Nature*, **453**, 677-681.

16. Wu, Y., Lu, J. and Kang, T. (2016) Human single-stranded DNA binding proteins: guardians of genome stability. *Acta Biochim Biophys Sin (Shanghai)*, **48**, 671-677.
17. Taib, N. and Gribaldo, S. (2020). SSB/RPA distribution in the Archaea - dataset. <https://data.mendeley.com/datasets/gc4xm4jn5x/draft?a=cdfb530e-9551-45e2-a180-d3e94ea82074>
18. Kelly, T.J., Simancek, P. and Brush, G.S. (1998) Identification and characterization of a single-stranded DNA-binding protein from the archaeon *Methanococcus jannaschii*. *Proc Natl Acad Sci U S A*, **95**, 14634-14639.
19. Chédin, F., Seitz, E.M. and Kowalczykowski, S.C. (1998) Novel homologs of replication protein A in archaea: implications for the evolution of ssDNA-binding proteins. *Trends Biochem Sci*, **23**, 273-277.
20. Skowyra, A. and MacNeill, S.A. (2012) Identification of essential and non-essential single-stranded DNA-binding proteins in a model archaeal organism. *Nucleic Acids Res*, **40**, 1077-1090.
21. Stroud, A., Liddell, S. and Allers, T. (2012) Genetic and Biochemical Identification of a novel single-stranded DNA-binding complex in *Haloferax volcanii*. *Front Microbiol*, **3**, 224.
22. Komori, K. and Ishino, Y. (2001) Replication protein A in *Pyrococcus furiosus* is involved in homologous DNA recombination. *J Biol Chem*, **276**, 25654-25660.
23. Nagata, M., Ishino, S., Yamagami, T. and Ishino, Y. (2019) Replication protein A complex in *Thermococcus kodakarensis* interacts with DNA polymerases and helps their effective strand synthesis. *Biosci Biotechnol Biochem*, **83**, 695-704.
24. Adam, P.S., Borrel, G., Brochier-Armanet, C. and Gribaldo, S. (2017) The growing tree of Archaea: new perspectives on their diversity, evolution and ecology. *ISME J*, **11**, 2407-2425.

25. Bult, C.J., White, O., Olsen, G.J., Zhou, L., Fleischmann, R.D., Sutton, G.G., Blake, J.A., FitzGerald, L.M., Clayton, R.A., Gocayne, J.D. *et al.* (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science*, **273**, 1058-1073.
26. Sarmiento, F., Mrazek, J. and Whitman, W.B. (2013) Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon *Methanococcus maripaludis*. *Proc Natl Acad Sci U S A*, **110**, 4726-4731.
27. Kelman, Z., Pietrokovski, S. and Hurwitz, J. (1999) Isolation and characterization of a split B-type DNA polymerase from the archaeon *Methanobacterium thermoautotrophicum* Δ H. *J Biol Chem*, **274**, 28751-28761.
28. Woodman, I.L., Brammer, K. and Bolt, E.L. (2011) Physical interaction between archaeal DNA repair helicase Hel308 and Replication Protein A (RPA). *DNA Repair (Amst)*, **10**, 306-313.
29. Robbins, J.B., McKinney, M.C., Guzman, C.E., Sriratana, B., Fitz-Gibbon, S., Ha, T. and Cann, I.K. (2005) The Euryarchaeota, Nature's medium for engineering of single-stranded DNA-binding proteins. *J Biol Chem*, **280**, 15325-15339.
30. Robbins, J.B., Murphy, M.C., White, B.A., Mackie, R.I., Ha, T. and Cann, I.K. (2004) Functional analysis of multiple single-stranded DNA-binding proteins from *Methanosarcina acetivorans* and their effects on DNA synthesis by DNA polymerase BI. *J Biol Chem*, **279**, 6315-6326.
31. Lin, Y., Robbins, J.B., Nyannor, E.K., Chen, Y.H. and Cann, I.K. (2005) A CCCH zinc finger conserved in a replication protein A homolog found in diverse Euryarchaeotes. *J Bacteriol*, **187**, 7881-7889.

32. Lin, Y., Guzman, C.E., McKinney, M.C., Nair, S.K., Ha, T. and Cann, I.K. (2006) *Methanosarcina acetivorans* flap endonuclease 1 activity is inhibited by a cognate single-stranded-DNA-binding protein. *J Bacteriol*, **188**, 6153-6167.
33. Lin, Y., Lin, L.J., Sriratana, P., Coleman, K., Ha, T., Spies, M. and Cann, I.K. (2008) Engineering of functional replication protein a homologs based on insights into the evolution of oligonucleotide / oligosaccharide-binding folds. *J Bacteriol*, **190**, 5766-5780.
34. Farkas, J.A., Picking, J.W. and Santangelo, T.J. (2013) Genetic techniques for the archaea. *Annu Rev Genet*, **47**, 539-561.
35. Winter, J.A., Patoli, B. and Bunting, K.A. (2012) DNA binding in high salt: analysing the salt dependence of replication protein A3 from the halophile *Haloferax volcanii*. *Archaea*, **2012**, 719092.
36. Delpech, F., Collien, Y., Mahou, P., Beaurepaire, E., Myllykallio, H. and Lestini, R. (2018) Snapshots of archaeal DNA replication and repair in living cells using super-resolution imaging. *Nucleic Acids Res*, **46**, 10757-10770.
37. Karan, R., DasSarma, P., Balcer-Kubiczek, E., Weng, R.R., Liao, C.C., Goodlett, D.R., Ng, W.V. and Dassarma, S. (2014) Bioengineering radioresistance by overproduction of RPA, a mammalian-type single-stranded DNA-binding protein, in a halophilic archaeon. *Appl Microbiol Biotechnol*, **98**, 1737-1747.
38. McCready, S., Muller, J.A., Boubriak, I., Berquist, B.R., Ng, W.L. and Dassarma, S. (2005) UV irradiation induces homologous recombination genes in the model archaeon, *Halobacterium sp.* NRC-1. *Saline Systems*, **1**, 3.
39. Webb, K.M., Yu, J., Robinson, C.K., Noboru, T., Lee, Y.C. and DiRuggiero, J. (2013) Effects of intracellular Mn on the radiation resistance of the halophilic archaeon *Halobacterium salinarum*. *Extremophiles*, **17**, 485-497.

40. Whitehead, K., Kish, A., Pan, M., Kaur, A., Reiss, D.J., King, N., Hohmann, L., DiRuggiero, J. and Baliga, N.S. (2006) An integrated systems approach for understanding cellular responses to gamma radiation. *Mol Syst Biol*, **2**, 47.
41. DeVeaux, L.C., Muller, J.A., Smith, J., Petrisko, J., Wells, D.P. and DasSarma, S. (2007) Extremely radiation-resistant mutants of a halophilic archaeon with increased single-stranded DNA-binding protein (RPA) gene expression. *Radiat Res*, **168**, 507-514.
42. Evans, J.J., Gygli, P.E., McCaskill, J. and DeVeaux, L.C. (2018) Divergent Roles of RPA Homologs of the model archaeon *Halobacterium salinarum* in survival of DNA damage. *Genes (Basel)*, **9**, 223.
43. Pluchon, P.F., Fouqueau, T., Creze, C., Laurent, S., Briffotiaux, J., Hogrel, G., Palud, A., Henneke, G., Godfroy, A., Hausner, W. *et al.* (2013) An extended network of genomic maintenance in the archaeon *Pyrococcus abyssi* highlights unexpected associations between eucaryotic homologs. *PLoS One*, **8**, e79707.
44. Li, Z., Santangelo, T.J., Cubonova, L., Reeve, J.N. and Kelman, Z. (2010) Affinity purification of an archaeal DNA replication protein network. *MBio*, **1**.
45. Pugh, R.A., Lin, Y., Eller, C., Leesley, H., Cann, I.K. and Spies, M. (2008) *Ferroplasma acidarmanus* RPA2 facilitates efficient unwinding of forked DNA substrates by monomers of FacXPD helicase. *J Mol Biol*, **383**, 982-998.
46. Guy, L. and Ettema, T.J. (2011) The archaeal 'TACK' superphylum and the origin of eukaryotes. *Trends Microbiol*, **19**, 580-587.
47. Haseltine, C.A. and Kowalczykowski, S.C. (2002) A distinctive single-strand DNA-binding protein from the archaeon *Sulfolobus solfataricus*. *Mol Microbiol*, **43**, 1505-1515.

48. Wadsworth, R.I. and White, M.F. (2001) Identification and properties of the crenarchaeal single-stranded DNA binding protein from *Sulfolobus solfataricus*. *Nucleic Acids Res*, **29**, 914-920.
49. Kerr, I.D., Wadsworth, R.I., Blankenfeldt, W., Staines, A.G., White, M.F. and Naismith, J.H. (2001) Overexpression, purification, crystallization and data collection of a single-stranded DNA-binding protein from *Sulfolobus solfataricus*. *Acta Crystallogr D Biol Crystallogr*, **57**, 1290-1292.
50. Kerr, I.D., Wadsworth, R.I., Cubeddu, L., Blankenfeldt, W., Naismith, J.H. and White, M.F. (2003) Insights into ssDNA recognition by the OB fold from a structural and thermodynamic study of *Sulfolobus* SSB protein. *EMBO J*, **22**, 2561-2570.
51. Gamsjaeger, R., Kariawasam, R., Gimenez, A.X., Touma, C., McIlwain, E., Bernardo, R.E., Shepherd, N.E., Ataide, S.F., Dong, Q., Richard, D.J. *et al.* (2015) The structural basis of DNA binding by the single-stranded DNA-binding protein from *Sulfolobus solfataricus*. *Biochem J*, **465**, 337-346.
52. Rolfsmeier, M.L. and Haseltine, C.A. (2010) The single-stranded DNA binding protein of *Sulfolobus solfataricus* acts in the presynaptic step of homologous recombination. *J Mol Biol*, **397**, 31-45.
53. Morten, M.J., Gamsjaeger, R., Cubeddu, L., Kariawasam, R., Peregrina, J., Penedo, J.C. and White, M.F. (2017) High-affinity RNA binding by a hyperthermophilic single-stranded DNA-binding protein. *Extremophiles*, **21**, 369-379.
54. Gamsjaeger, R., Kariawasam, R., Touma, C., Kwan, A.H., White, M.F. and Cubeddu, L. (2014) Backbone and side-chain (1)H, (1)(3)C and (1)(5)N resonance assignments of the OB domain of the single stranded DNA binding

- protein from *Sulfolobus solfataricus* and chemical shift mapping of the DNA-binding interface. *Biomol NMR Assign*, **8**, 243-246.
55. Morten, M.J., Peregrina, J.R., Figueira-Gonzalez, M., Ackermann, K., Bode, B.E., White, M.F. and Penedo, J.C. (2015) Binding dynamics of a monomeric SSB protein to DNA: a single-molecule multi-process approach. *Nucleic Acids Res*, **43**, 10907-10924.
 56. Cubeddu, L. and White, M.F. (2005) DNA damage detection by an archaeal single-stranded DNA-binding protein. *J Mol Biol*, **353**, 507-516.
 57. Richard, D.J., Bell, S.D. and White, M.F. (2004) Physical and functional interaction of the archaeal single-stranded DNA-binding protein SSB with RNA polymerase. *Nucleic Acids Res*, **32**, 1065-1074.
 58. Napoli, A., Valenti, A., Salerno, V., Nadal, M., Garnier, F., Rossi, M. and Ciaramella, M. (2005) Functional interaction of reverse gyrase with single-strand binding protein of the archaeon *Sulfolobus*. *Nucleic Acids Res*, **33**, 564-576.
 59. Zhang, C., Phillips, A.P.R., Wipfler, R.L., Olsen, G.J. and Whitaker, R.J. (2018) The essential genome of the crenarchaeal model *Sulfolobus islandicus*. *Nat Commun*, **9**, 4908.
 60. Suzuki, S. and Kurosawa, N. (2019) Robust growth of archaeal cells lacking a canonical single-stranded DNA-binding protein. *FEMS Microbiol Lett*, **366**.
 61. Paytubi, S., McMahon, S.A., Graham, S., Liu, H., Botting, C.H., Makarova, K.S., Koonin, E.V., Naismith, J.H. and White, M.F. (2012) Displacement of the canonical single-stranded DNA-binding protein in the Thermoproteales. *Proc Natl Acad Sci U S A*, **109**, E398-405.

62. Ghalei, H., von Moeller, H., Eppers, D., Sohmen, D., Wilson, D.N., Loll, B. and Wahl, M.C. (2014) Entrapment of DNA in an intersubunit tunnel system of a single-stranded DNA-binding protein. *Nucleic Acids Res*, **42**, 6698-6708.
63. Raghunathan, S., Kozlov, A.G., Lohman, T.M. and Waksman, G. (2000) Structure of the DNA binding domain of E. coli SSB bound to ssDNA. *Nat Struct Biol*, **7**, 648-652.
64. Zimmermann, L., Stephens, A., Nam, S.Z., Rau, D., Kubler, J., Lozajic, M., Gabler, F., Soding, J., Lupas, A.N. and Alva, V. (2017) A Completely Reimplemented MPI Bioinformatics Toolkit with a new HHpred server at its core. *J Mol Biol*, **430**, 2237-2243.

Table 1: Archaeal SSB/RPA protein structures				
Protein	Species	Method	PDB code	References
MjaRPA [OB2 only]	<i>M. jannaschii</i>	●	3DM3	PDB
MthRPA [OB1 only]	<i>M. thermautotrophicus</i>	○	2K50	PDB
MmpRPA1 [OB1 only]	<i>M. maripaludis</i>	● ○	3E0E 2K5V	PDB
MmaRPA2 [OB1 only]	<i>M. mazei</i>	○	2KEN 2KBN	PDB
TacRPA1 [OB2 only]	<i>T. acidophilum</i>	○	2K75	PDB
ThermoDBP	<i>T. tenax</i>	●	3TEK	(61)
Thermo-DBP- RP1	<i>P. furiosus</i>	●	4PSL, 4PSM	(62)
Thermo-DBP- RP2	<i>A. pernix</i>	●	4PSN, 4PSO (with bound DNA)	(62)
SsoSSB	<i>S. solfataricius</i>	○	2MNA (with bound DNA)	(51,54)
SsoSSB	<i>S. solfataricius</i>	●	1O7I	(49,50)
Methods: X-ray crystallography (filled circles), NMR (open circles)				

Figure legends

Figure 1: Bacterial SSB. **A.** Domain organisation of the archetypal *E. coli* SSB protein. **B.** Crystal structure of the SSB tetramer bound to single-stranded DNA (PDB: 1EYG) (63). The structure was obtained using SSB protein lacking the unstructured C-terminal tail (amino acids 1-116 of 178).

Figure 2: Eukaryotic RPA. **A.** Domain organisation of budding yeast *S. cerevisiae* RPA proteins RPA1 (RPA70), RPA2 (RPA32) and RPA3 (RPA14) with OB fold motifs labelled DBD-A through DBD-F and RPA2 C-terminal winged helix domain labelled wH. **B.** Structure of the fungus *U. maydis* RPA1 DBD-A, DBD-B and DBD-C, RPA2 DBD-D and RPA3 DBD-E bound to single-stranded DNA (PDB: 4GNX) (9).

Figure 3: Domain structures and phylogenetic distribution of archaeal SSB/RPA proteins. **A.** Schematic showing representative structures of various archaeal SSB/RPA proteins. The numbers of OB folds in the RPA41, RPA32 and multiple OB-fold RPA proteins vary from lineage to lineage. **B.** Taxonomic distribution and diversity of SSB/RPA across the archaeal domain. HMM profiles were built for each component using reference sequences and searched against a local data bank of 258 archaeal genomes using HMMSEARCH. The hits were manually checked using domain composition and structure prediction. The presence/absence in each phylum is indicated using a colour gradient according to the number of occurrences within the group. White is complete absence and dark red is presence in all the genomes. The full dataset can be found as ref. (17).

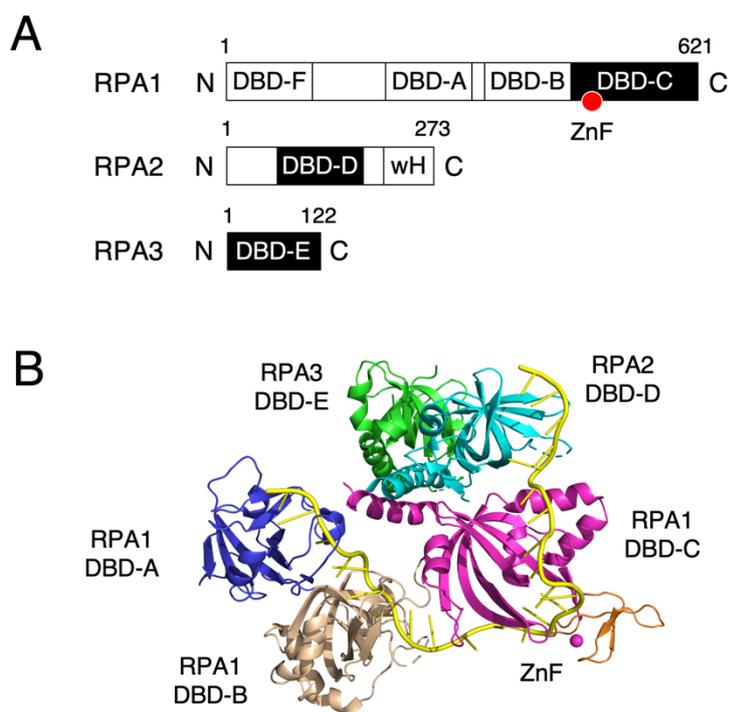
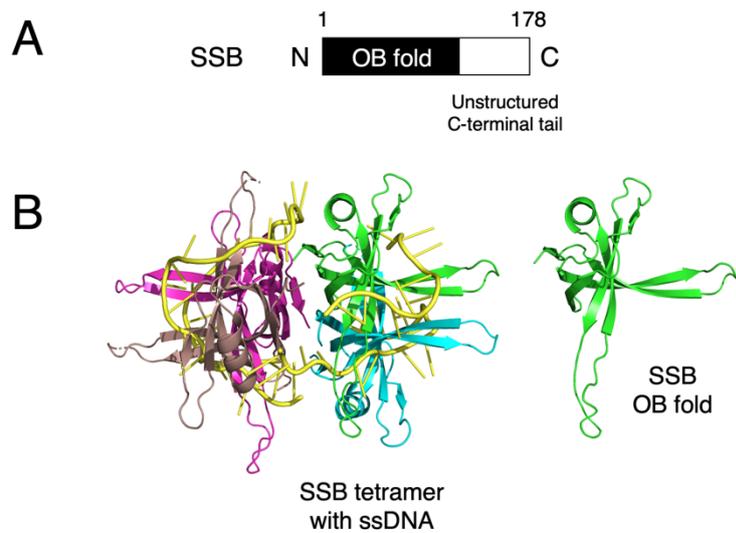
Figure 4: Archaeal RPA proteins. Domain organisation of representative RPA proteins from various archaeal species. Structural motifs (OB fold and wH domains) shaded in light grey were not identified as such in the original publications describing these proteins but were subsequently identified with high confidence (except for the low confidence prediction of the WH domain of HvoRpap3) using HHpred (64). PDB codes (green) are given for OB folds whose structure has been determined. PDB codes followed by an asterisk indicate the OB fold structure is from a closely related orthologue (see **Table 1**): these OB folds are shaded in dark grey. Proteins with UniProt KB codes: MjaRPA (Q58559), MthRPA (O27438), Mka (Q8TVF1), MacRPA1 (Q8TH77), MacRPA2 (Q8TLL7), MacRPA3 (Q8TT49), HvoRpa1 (D4GXL0), HvoRpa2 (D4GS55), HvoRpa3 (D4GZS1), HvoRpap1 (D4G XK9), HvoRpap3 (D4GZS0), PfuRPA41 (Q977W9), PfuRPA32 (Q977W7), PfuRPA14 (Q977W8), FacRPA1 (S0ATL9) and FacRPA2 (S0ANV3).

Figure 5: Structures of OB folds from archaeal RPA proteins. PDB codes: MjaRPA OB2 (3DM3), MthRPA OB1 (2K50), MmaRPA2 OB2 (2KEN) and TacRPA1 OB2 (2K75). TacRPA1 and MmaRPA2 are highly related to FacRPA1 and MacRPA2, respectively (see **Figure 3**).

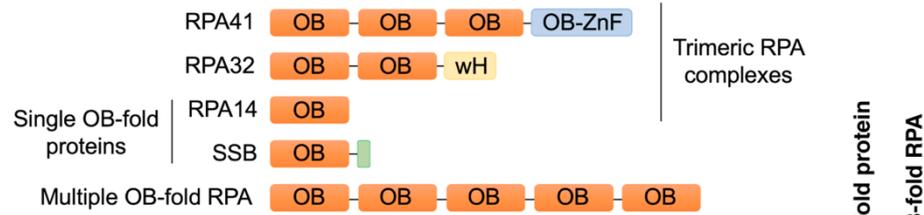
Figure 6: Solution structure of the SsoSSB bound to single-stranded DNA. The structure corresponds to amino acids 1-117 of *S. solfataricus* SSB (PDB: 2MNA) bound to single-stranded DNA (shown in red) (51,54).

Figure 7: Structure of ThermoDBP and ThermoDBP-related proteins. A. Structures of the N-terminal domain (amino acid 10-148) of *T. tenax* ThermoDBP (PDB: 3TEK) (61) and conserved N-terminal domains of *P. furiosus* ThermoDBP-RP1 (PDB: 4PSL)

and *A. pernix* ThermoDBP-RP2 (PDB: 4PSO) (62). **B. and C.** Structures of *P. furiosus* ThermoDBP-RP1 and *A. pernix* ThermoDBP-RP2 tetramers. In each case, individual monomers are shown in different colours, with the conserved N-terminal domain of one monomer coloured as in part A. Single-stranded DNA bound to ApeThermoDBP-DP2 is coloured in dark blue.



A



B

