DNA Replication Restart in Archaea
Roxane Lestini, Floriane Delpech, Hannu Myllykallio

To cite this version:
hal-01275161

HAL Id: hal-01275161
https://hal-polytechnique.archives-ouvertes.fr/hal-01275161
Submitted on 16 Feb 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Chapter 13
DNA Replication Restart in Archaea

Roxane Lestini, Floriane Delpech and Hannu Myllykallio

Additional information is available at the end of the chapter
http://dx.doi.org/10.5772/59994

1. Introduction

One fundamental challenge of cells is to accurately copy their genetic material for cell proliferation. This task is performed by core machineries considered conserved in all three domains of life: bacteria, archaea and eukaryotes [1].

For the vast majority of bacteria, the genome consists of one circular DNA molecule. Replication is initiated at a single replication origin from which two replication forks progress in the opposite direction. Replication termination takes place in the terminus region opposite the origin so that each replication fork has copied approximately one half of the genome. Studies of Escherichia coli mutants in key proteins for replication restart such as PriA strongly suggested that many replication forks encounter DNA damage or roadblocks leading to replisome inactivation under normal growth conditions. The reactivation of replication forks has been studied for several decades in bacteria. The picture that emerges is that bacterial proteins implicated in homologous recombination also play a key role in stabilizing and/or restoring blocked replication forks.

Unlike bacterial genomes, eukaryotic chromosomes contain numerous replication origins that can be used as backup origins to rescue arrested forks. Consequently, the importance of replication restart pathways in eukaryotes has long been ignored. However, recent studies have demonstrated that fork restart pathways operate also in eukaryotic cells and are important for cell viability under replication stress conditions. Eukaryotic replication restart pathways described also involve recombination proteins, as in bacteria. Thus it appears that general rules regarding replication restart and the key role of recombination proteins in these processes are conserved in bacteria, yeast and higher eukaryotes, but little is known in archaea, the third domain of life. This is of interest as archaea appear to be evolutionary hybrids between bacteria and eukaryotes.
Three main archaeal phyla are currently recognized: Crenarchaeota, Euryarchaeota [2] and Thaumarchaeota [3]. Similarly to most bacteria, archaeal genomes are also formed by a circular DNA molecule. However, unlike bacteria, some archaeal species have a single origin, whereas others have multiple origins per chromosome. Moreover, the ploidy of the genome in archaea varies considerably, with some species having one copy per cell whereas others have up to 25 copies of their genome in proliferating cells. As archaeal DNA replication consists both of evolutionary conserved as well as original features, understanding replication restart in these microorganisms will shed light on these fundamental but very complex pathways crucial to fulfill DNA replication. In this chapter we present recent advances on replication in archaea, followed by focused description of the Hef/XPF protein and its implication in replication restart in archaeal cells.

2. DNA replication origins in archaea

2.1. Multiple replication origins

Bacteria replicate their circular chromosome from a defined site called a replication origin. Two replication forks assemble at the replication origin and move in opposite directions. Each replication fork progresses at the same rate, and termination occurs at specific sites opposite the origin. Archaeal chromosomes are also circular, but whereas some archaea initiate replication from a single origin others replicate their chromosome from multiple replication origins, as observed for eukaryotic linear chromosomes (Table 1).

<table>
<thead>
<tr>
<th>Phylum</th>
<th>Organism</th>
<th>No. of replication origins</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Euryarchaeota</td>
<td><em>Pyrococcus abyssi</em></td>
<td>1</td>
<td>[4]</td>
</tr>
<tr>
<td></td>
<td><em>Haloferax volcanii</em></td>
<td>3</td>
<td>[5-7]</td>
</tr>
<tr>
<td></td>
<td><em>Haloferax mediterranei</em></td>
<td>2</td>
<td>[7]</td>
</tr>
<tr>
<td></td>
<td><em>Archaeglobus fulgidus</em></td>
<td>1</td>
<td>[8]</td>
</tr>
<tr>
<td></td>
<td><em>Halobacterium sp. NRC1</em></td>
<td>4</td>
<td>[9, 10]</td>
</tr>
<tr>
<td></td>
<td><em>Haloarcula hispanica</em></td>
<td>2</td>
<td>[11]</td>
</tr>
<tr>
<td></td>
<td><em>Methanothermobacter thermoautotrophicus</em></td>
<td>1</td>
<td>[12]</td>
</tr>
<tr>
<td>Crenarchaeota</td>
<td><em>Sulfolobus acidocaldarius</em></td>
<td>3</td>
<td>[13]</td>
</tr>
<tr>
<td></td>
<td><em>Sulfolobus solfataricus</em></td>
<td>3</td>
<td>[13]</td>
</tr>
<tr>
<td></td>
<td><em>Sulfolobus islandicus</em></td>
<td>3</td>
<td>[14]</td>
</tr>
<tr>
<td></td>
<td><em>Pyrobaculum calidifontis</em></td>
<td>4</td>
<td>[15]</td>
</tr>
<tr>
<td></td>
<td><em>Aeropyrum pernix</em></td>
<td>At least 2</td>
<td>[16]</td>
</tr>
<tr>
<td>Thaumarchaeota</td>
<td><em>Nitrospumilus maritimus</em></td>
<td>1</td>
<td>[7]</td>
</tr>
</tbody>
</table>

Table 1. Replication origins experimentally identified in archaeal chromosomes
Replication from a single replication origin was reported experimentally in the euryarchaea *Pyrococcus abyssi* [4] and in *Archaeoglobus fulgidus* [8]. But then several studies showed that various euryarchaea have multiple replication origins like the halophiles *Haloferax volcanii* [5-7] and *Haloarcula hispanica* [11]. Multiple replication origins have also been identified in *Sulfolobus solfataricus*, *Sulfolobus acidocaldarius* and *Sulfolobus islandicus* [13, 14] as well as in *Pyrobaculum calidifondis* [15] and *Aeropyrum pernix* [16] that belong to the crearchaeota phylum. Whether archaea from the recently discovered phylum thaumarchaeota have multiple origins remains unknown but a recent study in *Nitrosopumilus maritimus* identified a single replication origin in this organism that is conserved in the phylum, suggesting they have a single replication origin as also suggested by recent computational analysis [7, 17].

### 2.2. Archaeal replication initiator Orc1/Cdc6 proteins and origins recognition

How replication is regulated to allow a single circular DNA molecule to be replicated from uneven multiple origins is an ongoing question in archaea (Figure 1).

![Figure 1](image)

**Figure 1.** Uneven distribution of multiple replication origins in archaeal chromosomes. (A) in *Sulfolobus* species, (B) in *Haloarcula hispanica*, and (C) in *Haloferax volcanii*. Bubbles on chromosomes indicate replication origins, and arrows indicate bidirectional replication from each origin.

From each replication origin two replication forks are assembled and progress at the same rate so that termination of the replication is asynchronous. The origin region usually has a high content of adenine and thymine residues flanked by several conserved repeated motifs known as Origin Recognition Boxes (ORBs). In many cases archaeal replication origins are linked to replication genes [15, 18] and are located near genes coding initiator proteins. Despite the conservation of the replication origin-initiator structure, archaeal replication origins exhibit considerable diversity in terms of both ORB elements and their initiator genes [7, 11, 12]. Because replication origins can be dramatically diverse, it may facilitate differential usages by these microorganisms to adapt to various harsh environments.

All sequenced archaeal genomes encode proteins homologous to the eukaryotic initiator proteins Orc1 and Cdc6. Because the archaeal proteins are related both to the eukaryotic Orc1 subunit, involved in the replication origin recognition, and Cdc6, involved in the replicative helicase recruitment, they may combine both activities in a single polypeptide. Indeed, several
studies have shown that the archaeal replicative helicase MCM is recruited by Cdc6/Orc1 proteins at replication origins [14, 19, 20]. Archaeal Cdc6/Orc1 proteins also share mechanistic similarities with the bacterial initiator protein DnaA. In vitro studies on the binding of Cdc6/Orc1 proteins to ORBs in the *Methanothermobacter thermoautotrophicus* replication origin revealed that they bind cooperatively to the repeated sequences found in the vicinity of the oriC, as observed for the association of the bacterial initiator protein DnaA [12, 21].

How multiple replication origins are regulated by Cdc6/Orc1 proteins in archaeal cells is a complex question. The number of Orc1/Cdc6 proteins varies between species, and recent genetic studies attempting to delete *cdc6/orc1* genes revealed a complex regulation of replication, highlighting a specificity of initiator proteins at each origin.

For instance, four *cdc6/orc1* genes are found on *Halobacterium* NRC-1 chromosome that replicates from four distinct replication origins, but only two *cdc6/orc1* genes are located near a replication origin. Genetic studies of replication initiation at one of these origins showed that only the presence of the initiator protein associated was needed, revealing a specific binding of each replication origin by initiator proteins [10]. This is also the case in *Solfataricus* cells. Three *cdc6/orc1* genes are found in the chromosome. Out of the three replication origins, two were linked to a *cdc6/orc1* gene, and different subsets of the three Cdc6/Orc1 proteins recognized these replication origins [22]. The third origin was not recognized by Cdc6/Orc1 initiators. It was specifically recognized by the crenarchaeal-specific WhiP protein (for Winged-Helix initiator protein). WhiP proteins share sequence similarity with the essential eukaryal replication factor Cdt1 and display a domain organization reminiscent of bacterial plasmid initiator proteins. The conservation of WhiP-coding genes located near the replication origin in other crenarchaea suggested that this third replication origin was captured from extrachromosomal elements [16]. A similar situation is found in *Sulfobacterium islandicus*. *SisOriC-1* was bound by Orc1-1, *SisOriC-2* by Orc1-3 while no association of any Cdc6/Orc1 protein was observed at *SsiOriC-3* specifically recognized by *SsiWhiP* protein [14].

Moreover, additional role of initiator proteins independent of replication origins has recently been suggested by serial deletions of *cdc6/Orc1* genes in *S. Islandicus* and *H. hispanica*. In *S. Islandicus* none of the three *cdc6/Orc1* genes were essential for viability and all three possible double-mutants were viable. However, although one of the Cdc6/Orc1 proteins seemingly did not bind to any replication origin in vivo, the triple mutant could not be generated, further suggesting that the observed synthetic lethality may reflect additional role of replication initiator proteins [14]. Similarly, both replication origins in *H. hispanica* chromosome were shown to be controlled independently by specific *cdc6/orc1* genes. But while one of the replication origin could be deleted, the deletion of its associated *cdc6/orc1* gene lead to a severe growth defect, also suggesting a vital function of the protein outside replication initiation from its associated origin [23].

### 2.3. Are replication origins essential for viability in archaea?

The specific initiation sites, replication origins, on the chromosome of *H. hispanica* could be deleted separately but it was not possible to generate a mutant deleted for both origins at the same time. Attempt to delete also the replication origins of other replicons found in this
organism suggested that one active ori-cdc6 pairing on each replicon was essential for genome replication in *H. hispanica* [23]. But recent data obtained in the halophilic archaea *Haloferax volcanii* challenged the notion that replication origins are essential determinants of DNA replication. Indeed, Hawkins *et al.* revealed that not only cells were viable upon deletion of all known replication origins but they even grew faster than the corresponding wild-type cells [5].

How replication initiates in absence of replication origins? Because no evidence for activation of dormant origins has been found, authors favoured the hypothesis that replication initiation occurs randomly on the chromosome at recombination intermediates. Recombination-Dependent Replication (RDR) has first been observed in *E. coli* cells and extensively studied by various laboratories. From those studies it appears that replication fork inactivation occurs very frequently under normal growth conditions. Several replication restart pathways have evolved depending on the cause of arrest. They all share the common feature to involve recombination proteins such as the RecA recombinase and the PriA protein responsible for the loading of a replisome at recombination intermediates (Figure 2).

Indeed Hawkins *et al.* have shown that the archaeal RecA ortholog (RadA) is essential for viability in absence of replication origins.

But the deletion of radA alone impaired *H. volcanii* growth, highlighting that viability already relied on recombination [24]. Thus the essentiality of RadA in absence of replication origins may not reflect a direct need for recombination to start replication. Furthermore, in *E. coli* cells

---

**Figure 2.** Model for double-strand break repair and replication restart in *E. coli*. RecBCD (purple, green and blue egg-shaped) degrades double-stranded end until it encounters a Chi site (black region). A switch in RecBCD activity produces a 3’-single-stranded DNA on which RecA (yellow ball) proteins are loaded. Homology search and strand exchange forms a Holliday junction (HJ) adjacent to a D-loop. The Holliday junction is resolved by RuvABC. PriA then load the replisome on the D-loop at which replication restarts.
RDR is deleterious for growth and viability. It also seems to be the case in yeast [25] and higher eukaryotes in which replication defects are linked to genome rearrangements and diseases [26, 27]. In that context the better fitness of origins-deleted cells observed in *H. volcanii* is puzzling. Hawkins *et al.* argued that replicative helicases MCM were more efficiently recruited at recombination intermediates as they were not sequestered at replication origin(s). Whether MCM is a limiting factor for replication initiation at replication origins in *H. volcanii* cells is currently not known and would have to be investigated. They also argued that the polyploidy of *H. volcanii* genome (18 copies of the genome in exponential phase [28]) allows viability to rely on stochastic partitioning. This argument implies that all chromosome dimers generated by recombination events (including RDR) do not have to be resolved to provide viable daughter cells, and that proteins involved in resolution of recombination intermediates such as the Holliday junction resolvase Hjc are not essential for viability. This hypothesis is clearly worth of experimental testing.

An alternative explanation for RDR is activation of dormant origins randomly in cells so that no preferential origin emerged at the level of a cell population [29, 30]. In this scenario the essentiality of RadA could imply that randomly-initiated replication forks more often collapse and have to be restarted. This notion would be consistent with an organization of archaeal genes on the genome preventing collision of replication machinery with transcription machinery [18] and physical connections recently suggested between replication and transcription machineries [31].

In conclusion, this study by Hawkins *et al.* raised many interesting questions that need to be experimentally addressed to understand how *H. volcanii* genome is replicated in absence of replication origins. Future work should aim at unravelling the molecular mechanisms that allow archaeal cells lacking origins to be viable and to even show increased fitness. In that context, one interesting protein to focus on might be the helicase/nuclease Hef. Indeed it has recently been shown that Hef (i) is genetically linked to the HJ resolvase Hjc and (ii) is recruited at arrested replication forks in living *H. volcanii* cells [32, 33].

### 3. Archaeal Hef/XPF proteins from the XPF/MUS81/FANCM family

Proteins belonging to the XPF/MUS81/FANCM endonuclease family act on 3′-flap DNA structures that are formed during DNA repair or replication restart. They are found throughout eukarya and archaea but to date have not been identified in bacteria. Eukaryotes have several XPF/MUS81/FANCM family members that all share a conserved nuclease domain [34] whereas MUS81 proteins possess only an active nuclease domain. In XPF, an active nuclease domain is fused to a SF2-helicase domain that is degenerated and appears to be inactive [35]. By contrast, FANCM consists of a helicase:nuclease fusion in which the nuclease domain is degenerated [36, 37]. Other members can be found that have a degenerated nuclease and/or helicase domain. They assemble into heterodimeric complexes with MUS81, XPF or FANCM proteins to form distinct active complexes involved in DNA repair, meiotic recombination and replication restart [38] (Figure 3).
All archaea encode a protein of the XPF/MUS81/FANCM family of endonucleases. It exists in two forms. The long form, referred as Hef, consists of an N-terminal helicase fused to a C-terminal nuclease and is specific to the euryarchaea. The short form, referred as XPF, lacks the helicase domain and is specific to the crenarchaea and the thaumarchaea (Figure 3).

The long-formed Hef protein was first identified in *Pyrococcus furiosus* due to its activity on branched DNA structures, Hef meaning helicase-associated endonuclease fork-structure DNA [39]. Hef has the unique feature among XPF/MUS81/FANCM proteins of having both an active helicase domain and an active nuclease domain, allowing the identification of its human DNA replication restart in archaea.
ortholog FANCM protein also consisting of a helicase:nuclease fusion [37, 40]. What do we know about archaeal Hef/XPF function?

3.1. In vitro studies of crenarchaeal XPF proteins

The crystallographic structure of XPF from the crenarchaea *Aeropyrum pernix* was solved in presence and absence of double-stranded DNA [41, 42]. The protein has two domains, a N-terminal nuclease domain and an Helix-hairpin-Helix (HhH)\textsubscript{2} domain. *ApeXPF* formed homodimers. The interaction involved the two nuclease domains and the two (HhH)\textsubscript{2} domains from each monomer. However, only one monomer seemed catalytically active at a time when the homodimer was bound to DNA. The (HhH)\textsubscript{2} domain had a major role in interacting with DNA. This interaction triggered a domain movement coupling the (HhH)\textsubscript{2} domain to the nuclease domain to allow subsequent cleavage of the DNA substrate. The DNA was bent by around 90° upon interaction, suggesting that XPF binding causes distortion at double-strand/single-strand DNA junctions.

The nuclease activity of XPF from *Sulfulobus solfataricus* has been studied in more details. The replication factor PCNA (Proliferating Cell Nuclear Antigen) was required in vitro for nuclease activity of this “short” XPF [43]. In the cell, the trimeric PCNA ring encircles double-strand DNA (dsDNA) and firmly attaches the replicative polymerase to the template strand, enhancing its processivity. PCNA is a central protein as it also interacts with various proteins involved in replication and/or repair like Fen1. Interaction with PCNA often involves a conserved motif known as PCNA-Interacting Protein (PIP) motif conserved in XPF proteins. Indeed it was shown that *Sso* XPF interacts with PCNA through its conserved PIP motif [43, 44]. Intramolecular FRET experiments showed that the binding of *Sso* XPF to a 3′-flap indeed bent the DNA as observed in *Ape* XPF structure, but that the interaction with PCNA allowed *Sso* XPF to distort the DNA structure in a proper conformation for efficient cleavage [45, 46]. *Sso* XPF preferentially cleaved 3′-flap and processed them into gapped duplex products. It was also observed that *Sso* XPF can act on substrates containing a variety of DNA damages or modifications [47, 48].

3.2. In vitro characterization of euryarchaeal Hef proteins

As mentioned previously Hef was identified in *P. furiosus* due to its enzymatic activity on branched DNA structures [39]. In vitro experiments on *Pfu* Hef revealed a similar organization of the C-terminal region of archaeal XPF proteins, with a nuclease domain and a helix-hairpin-helix domain. Similarly, homodimers were observed with both the nuclease and the HhH domains forming domain-domain interfaces. Dimer formation appeared crucial for substrate recognition specificity [49]. A variety of branched DNA structures carrying single-strand DNA (ssDNA) portions, such as flapped and fork-structured DNAs, were recognized and cleaved by the C-terminal nuclease domain of *Pfu* Hef [50, 51]. The N-terminal domain of Hef displayed a structured-DNA specific helicase. Two conserved helicase motifs from Super-Family 2 (SF2) helicases were separated by a third domain that shares structure similarity with the “Thumb” domain of polymerases involved in dsDNA binding [51].
In vitro experiments suggested collaboration between the coupled helicase and the nuclease domains of PfuHef: the helicase domain binds and processes the fork-structured DNA, forming a four-way structure that is then cleaved by the endonuclease domain [52]. Yet replication fork restart can involve the formation of a four-way junction (Holliday junctions) from a three-way junction (fork-like structure). The four-way junction is then resolved by a Holliday junction resolvase. Thus euryarchaeal Hef could be involved in the resolution of stalled replication forks, as suggested for Mus81 complexes in eukaryotes. Indeed in both fission and budding yeast it seems that the primary function of Mus81 complexes is the restart of collapsed-replication forks by homologous recombination [53-57], a role that is functionally redundant with the helicase-nuclease Sgs1-Top3 and the Rqh1-Top3 complexes, respectively [53, 58]. The MUS81 complex is also found in humans [59], and promotes replication fork restart by homologous recombination [60-63].

More recently, Thermococcus kodakarensis Hef has been shown to interact with TkoPCNA1. The interaction with PCNA did not involve a canonical PIP motif but a disordered region of Hef between the helicase and nuclease domains. Interestingly, these long disordered regions connecting two catalytic domains are a common feature of euryarchaeal Hef and eukaryotic FANCM proteins [64].

These biochemical studies have indicated that both creanarchaeal XPF and euryarchaeal Hef proteins interact with PCNA and display biochemical activities consistent for being proteins involved in DNA repair and/or replication restart. Is this hypothesis supported by in vivo studies of Hef proteins?

3.3. What have we learned deleting hef gene in euryarchaea

The hef gene has been deleted in two different euryarchaea: in the hyperthermophile Thermococcus kodakarensis and in the halophile Haloferax volcanii. In both organisms, Hef was non-essential for cell viability under normal laboratory growth conditions.

T. kodakarensis cells deleted for hef showed increased sensitivity to a variety of DNA damaging agents [65], consistent with a role of Hef in the maintenance of genomic stability. The sensitivity to UV irradiation suggested that Hef was involved in the repair of UV lesions. Both helicase and nuclease domains of TkoHef were needed as the same phenotype was observed upon deletion of the entire gene and deletion of the helicase-coding region or the nuclease-coding region of the gene [65]. Nucleotide Excision Repair (NER) is the major pathway to repair DNA lesions after UV radiation. But the existence of an archaeal NER pathway has not been established yet. Most archaea have eukaryal-type NER genes, but most of eukaryal NER proteins have multiple cellular functions so the presence of several NER-like protein is not enough to conclude that a functional NER pathway can be found in archaea [66]. Interestingly the sensitivity to UV radiations of T. kodakarensis cells deleted for hef suggested that Hef was involved in Nucleotide Excision Repair (NER), as its human ortholog XPF-ERCC1 and its counterpart RAD1-RAD10 in the yeast Saccharomyces cerevisiae [67, 68]. And it also suggested that an active NER pathway exists in archaea. Clearly, additional experiments are now needed to better understand the role of TkoHef in NER and, more generally, to further dissect the pathway responsible for archaeal Nucleotide Excision Repair.
In contrast, the deletion of hef in H. volcanii neither affected sensitivity to various DNA damaging agents nor recombination frequency. We could only observed a slow-growth phenotype of the Δhef colonies when chronically exposed to mitomycin C (MMC) on plate [32]. Note that the direct comparison between these two studies on phenotypic analyses of Δhef strains is difficult as Table 2 illustrates major differences regarding experimental conditions used including drug concentrations and cell treatment methods.

<table>
<thead>
<tr>
<th></th>
<th>H. volcanii</th>
<th>T. kodakarensis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Concentration</td>
<td>Exposure</td>
</tr>
<tr>
<td>UV irradiation</td>
<td>Up to 150 J/m²</td>
<td>on plate</td>
</tr>
<tr>
<td>MMS</td>
<td>0.04%</td>
<td>1 hour in suspension</td>
</tr>
<tr>
<td>γ-rays</td>
<td>Up to 1000 Gy</td>
<td>on plate</td>
</tr>
<tr>
<td>Phleomycin</td>
<td>1 or 2 mg/ml</td>
<td>1 hour in suspension</td>
</tr>
<tr>
<td>Mitomycin C</td>
<td>0.02 µg/ml</td>
<td>On plate</td>
</tr>
</tbody>
</table>

Table 2. Methods used for exposure of H. volcanii and T. Kodakarensis Δhef cells to DNA damaging agents as reported in [32, 65, 69].

A possible explanation for these phenotypic differences is that NER proteins in Thermococcus and Haloferax species are very different. In fact H. volcanii also possesses bacterial-like NER proteins most probably acquired by lateral gene transfer [66], and it was shown that they were responsible for the repair of UV lesions [32].

To further investigate the role of Hef in H. volcanii, the observed lack of an obvious phenotype for Δhef cells prompted us to combine hef-deletion with other endonuclease or helicase deletions that may encode redundant functions with Hef. Among several combinations tested, we demonstrated that Hef was essential for viability in the absence of the Holliday junctions (HJs) resolvase Hjc. Holliday junctions are four-way branched DNA structures formed during homologous recombination strand exchange and recombination-dependent replication restart. HJs resolvases are found in bacteria, archaea and eukarya, although they are not evolutionary related. Hjc is conserved throughout archaea. The single deletion of hjc gene in H. volcanii cells (as well as in T. Kodakarensis cells) did not affect growth rate, DNA repair or recombination [32, 65]. Co-lethality of Hef and Hjc could be explained by redundant roles of Hef and Hjc as HJs resolvase. In this scenario Hef could use its helicase activity on arrested replication forks to process them into four-way DNA structures that can be resolved by its nuclease activity. This scenario was compatible with the in vitro studies described above.
Indeed, point mutations inactivating the helicase activity (HvoHef-K48A) or the nuclease activity (HvoHef-D679A) of HvoHef resulted in the same phenotype observed in the absence of the entire protein. This nicely demonstrated that both helicase and nuclease activities of Hef were required for fulfilling its role in the absence of the Hjc resolvase [32]. To test the hypothesis that Hef and Hjc were both acting as HJs resolvase, we deleted hef or hjc in a strain carrying a radA deletion. In absence of RadA recombinase, HJs are no longer formed by homologous recombination so that deleting HJs resolvases should not have any affect. Indeed cells deleted for both radA and hjc were phenotypically similar to cells only deleted for radA. This observation was consistent with a role of Hjc in the resolution of HJs formed by RadA-mediated strand exchange during homologous recombination. However, radA gene could not be deleted in hef-deleted cells [32], strongly reflecting that functional roles of Hjc and Hef were distinct. This observation also suggested that Hef was required for cell viability in absence of recombination. Which alternative pathways or additional functions could (i) depend on Hef, (ii) be essential for cell viability during normal growth condition, and (iii) implicate recombination proteins and/or Hef? Replication restart is one possible pathway. But how could we obtain more detailed information on functional roles of Hef if hef-deleted mutant strains hardly shows any phenotype or cannot be combined with other deletions? We decided to develop tools to allow dynamic localization of fluorescently-labelled Hef proteins in living Halofex volcanii cells.

### Table 3. Growth phenotypes of H. volcanii deletion mutants.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Growth phenotypes</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>Δhef</td>
<td>+++</td>
<td>[32]</td>
</tr>
<tr>
<td>Δhjc</td>
<td>++</td>
<td>[32]</td>
</tr>
<tr>
<td>ΔradA</td>
<td>+</td>
<td>[24]</td>
</tr>
<tr>
<td>Δhef Δhjc</td>
<td>-</td>
<td>[32]</td>
</tr>
<tr>
<td>ΔradA Δhjc</td>
<td>+</td>
<td>[32]</td>
</tr>
<tr>
<td>ΔradA Δhef</td>
<td>-</td>
<td>[32]</td>
</tr>
</tbody>
</table>

4. Dynamic localization of Hef proteins fused to the Green Fluorescent Protein (GFP) in living H. volcanii cells

The Green Fluorescent Protein (GFP) was originally isolated from the jellyfish *Aequora Victoria* [70]. It is encoded by a single polypeptide containing the chromophore. After translation of the protein, an autocatalytic process involving oxygen has to take place within the chromophore. Once active, the GFP has a major excitation peak at a wavelength of 395 nm and an emission peak at 509 nm. A deep understanding of the protein has enabled the development
of several GFP variants with modified spectral properties [71, 72]. All those FPs are now used in living cells/organisms to study protein localization, mobility, turnover, interactions, and much more [73]. Such approaches can reveal key features of proteins in vivo to complete our understanding of pathways, as illustrated for NER pathway in mammalian cells [74], but their use in archaeal cells has been rather limited until recently.

4.1. Fusion of the green fluorescent protein to the C-terminal end of Hef protein in H. volcanii

GFP has been used to investigate proteasome-dependant proteolysis and protein levels in H. volcanii [75, 76] as well as biofilm formation by the crenarchaea Sulfolobus solfataricus [77], two aerobic archaea. Because GFP’s variant can differ not only by their fluorescence properties but also by their maturation rate of the fluorophore, temperature and pH stability or oligomeric state for instance, one has to carefully choose the variants that optimally fit the lifestyle of the organism being studied. We have recently used GFP-fusions to investigate protein localization and behaviour in archaeal cells for the first time. These studies were performed using the halophile H. volcanii that has a relatively high intracellular salt concentration (around 2.5M in laboratory growth conditions) and an optimal growth temperature of 45°C. Expression of several GFP variants were previously tested in this species, demonstrating that the smRS-GFP could be used for further studies [75]. This variant has mutations increasing solubility (Phe99Ser, Met153Thr and Val63Ala) as well as a mutation in one of the three amino acids of the chromophore (Ser65Thr) that redshifts the absorption maximum to 488 nm without changing the emission properties of the protein [78, 79]. Based on this observation we fused the smRS-GFP to the C-terminal end of H. volcanii Hef. The resulting fusion protein was expressed under physiological expression levels and conditions from the native chromosomal locus of the hef gene [33].

Whether GFP-fused Hef proteins remained functional was then tested by comparing cells deleted for hef with cells expressing the hef::gfp allele. No growth delay was observed for hef::gfp cells on MMC plates, indicating functional complementation by Hef::GFP construct. In agreement with this notion, hjc could be deleted, although a growth defect was measured for hef::gfp Δhjc cells. Because we were interested in the localization of Hef in response to replication arrests, we exposed cells to aphidicolin (APD), an antibiotic that inhibits DNA synthesis in halophilic archaea [80], thus arresting replication forks. Exposing hef-deleted cells to increasing concentrations of APD decreased cell viability, showing that indeed Hef is involved in the genomic stability upon replication arrest. Such decrease in cell viability was not observed with cells expressing GFP-fused Hef.

4.2. Localising the fluorescence signal in H. volcanii living cells

We then observed the localisation of Hef::GFP proteins by fluorescence microscopy, comparing cells exposed to APD to non-treated control samples. Towards this goal, a drop of cells was spotted on an agarose slice placed on a glass slide. After allowing this drop to dry, the agarose pad was covered with a cover-slip for cell imaging studies using a wide-field microscope to visualize a large number of individual cells. Differential Interference Contrast (DIC) [also
known as Nomarski Interference Contrast (NIC), was first used to visualize the cells as it enhances the contrast in unstained, transparent samples. Then fluorescence imaging was performed (exciting at 474 nm and collecting at 527-554 nm). Note however that due to the small cell size (around 1 to 2 µm) and the use of the soft agarose, not all cells were in the same focal plane. In order not to lose any information, fluorescence images were acquired at different focal planes on the z-axis. Consecutive slices of cells in focus were then selected and used to perform a maximum intensity z-projection. At each pixel, the highest fluorescence signal was kept when comparing the selected images. This maximum intensity z-projection resulted into a two-dimensional picture where the maximal fluorescence signals from different focal planes were recorded (Figure 4).

![Figure 4](image)

**Figure 4.** Schematic representation of fluorescence signal analysed in cells. (A) Representation of a cell with fluorescence foci and the different focal planes used for imaging. (B) Representation of fluorescence signal in each focal plane. (C) Resulting image after projection of the maximum fluorescence signal at each pixel for the four focal planes.

Resulting images contained hundreds of cells that were analysed by quantitative image analyses using IMARIS software. Different imaging parameters were optimized to detect cells and fluorescence foci within cells using automatic thresholds to avoid user-bias. This approach allowed thousands of cells to be analysed in each condition tested, providing extremely high statistical power.

### 4.3. Hef::GFP molecules are recruited at arrested replication forks

Using such approach, we have shown that Hef::GFP proteins formed fluorescence foci even under normal growth condition, in the absence of any DNA damaging agents. The number of these foci was significantly increased from two to four foci per cell in response to aphidicolin exposure. We also observed that the number of foci per individual cell changed significantly.
While the majority of cells had one or two foci in normal growth conditions, a higher proportion of cells having more than two foci was observed upon APD exposure (Figure 5).

![Figure 5](image)

**Figure 5.** *In vivo* localization of GFP-labeled Hef in response to aphidicolin exposure. A total of 23760 spots within 13666 control cells and 15299 spots within 3721 APD-treated cells were analyzed. (A) Pictures of DIC and GFP signal of *hef::gfp* cells under control conditions and after exposure to 5 µg/ml aphidicolin. Bar equals 10 µm. (B) Average cell surface of *hef::gfp* cells in response to increasing concentrations of aphidicolin. (C) Mean number of GFP-Hef labeled fluorescence foci per cell in response to increasing concentrations of aphidicolin. (D) Relative frequency of number of foci per individual cell. All error bars represent SD. n > 3 experiments, *t* test are performed in comparison to control without aphidicolin. *** Significantly different, *p*<0.001. ** Significantly different, *p*<0.01. * Significantly different, *p*<0.05. From [33].

We have also observed that cell size was increased from 28 to 45 µm² in response to replication arrest (i.e. APD exposure). We have shown using other DNA damaging agents that increased cell size and number of foci were specific to APD treatment, suggesting that indeed *Hvo*Hef is recruited at arrested replication forks brought about by addition of aphidicolin.

### 4.4. Diffusing pattern of Hef::GFP molecule upon replication arrests

To investigate the diffusion of Hef::GFP molecules inside and outside fluorescence foci, we performed Fluorescence Recovery After Photobleaching (FRAP) and Number and Brightness (N&B) experiments. These experiments were performed using a confocal microscope on cells immobilized on a poly-D-lysine coated cover-slip.

In FRAP experiments a region of interest was photobleached in a cell. The speed of fluorescence recovery in that region was then measured, reflecting the diffusion of
Hef::GFP fluorescent molecules arriving from the non-photobleached region of the cell. In control cells (no aphidicolin), one major population of Hef::GFP diffusing molecules was observed. From the fit of the recovery curve we obtained the recovery constant, allowing then the apparent two-dimensional diffusion rate of Hef::GFP to be estimated at 0.8 to 2.3 µm² per second. This appeared markedly lower than expected for Hef dimer, as revealed by analytical ultracentrifugation experiments on purified Hvo
Hef further indicating that Hef has a peculiar elongated shape in solution. Several possibilities may explain this limited diffusion. In addition to the non-globular quaternary structure, physical constraints of the cytosol, possibly resulting from high DNA and salt concentration, and/or transient interactions with cellular components (DNA or proteins complexes) may explain this slow diffusion. But FRAP experiments performed on cells exposed to aphidicolin revealed an additional, even more slowly-diffusing population that was clearly induced by APD treatment (Figure 6).

Such changes in the diffusion pattern of Hef::GFP molecules upon replication arrests were also observed using N&B technique that measures fluctuation of fluorescence intensity in each analysed pixel [81]. These analyses were performed on one hundred images taken every 2 seconds, and cell regions including and excluding fluorescence foci were compared. Fluctuation of fluorescence intensity per pixel was then used to

Figure 6. Fluorescence Recovery After Photobleaching experiments to study the dynamic localization of GFP-labeled Hef molecules at fluorescence foci. (A) Images of a representative cell in response to aphidicolin treatment for FRAP analysis. FRAP regions are shown by white circles. Time after photobleaching in seconds. Bar equals 5 µm. (B) Fluorescence recovery curve averaged for 9 control cells. C) Fluorescence recovery curve averaged for 8 aphidicolin treated-cells. (D) Diffusion constants [Confidence interval at 95%] calculated for GFP-labeled Hef diffusing molecules. From [33].
determine the number of diffusing molecules and their brightness. This information can then be used to deduce changes in the oligomeric state of the fluorescent molecules. When N&B technique was applied to diffusing Hef::GFP molecules, changes in the oligomeric state (i.e. higher brightness) were observed upon APD treatment. This observation revealed oligomerization and/or co-localisation of several Hef::GFP molecules induced by APD exposure (Figure 7), and provided a feasible explanation for the slow diffusion in APD treated cells revealed by FRAP experiments.

Figure 7. Number & Brightness experiments to study the oligomeric state of GFP-labeled Hef molecules diffusing at fluorescence foci. (A) Images of representative cells for N&B analysis. Average intensity (A and C) and pseudo-coloured normalized brightness values (B and D) for representative control cells (A and B) and cells exposed to 5 µg/ml aphidicolin (C and D). Bar equals 5 µm. (B) Average number of Hef::GFP diffusing molecules per pixel. (C) Average brightness of Hef::GFP diffusing molecules per pixel. From [33].

Overall, the results obtained from FRAP and N&B experiments were consistent with the notion that Hef::GFP molecules are actively recruited at arrested replication forks. Whether the slow-diffusion pattern of Hef::GFP molecules reflects their recruitment directly on DNA and/or as part of protein complexes at arrested replication forks are questions that remain to be addressed. Interestingly, hjc deletion had effect neither on cell size nor on the number of foci per cell in normal growth condition as well as in response to APD treatment. These observations showed that recruitment of Hef to arrested replication forks was not increased in the absence of the alternative pathway involving Hjc (and RadA), suggesting that Hef is recruited at arrested replication forks even in the presence of the alternative HR-dependent pathway. We also noted that in eukarya recent studies have indicated that FANCM proteins can prevent homologous recombination [82-85]. This raised the possibility that HvoHef may prevent access of recombination proteins to arrested replication forks (Figure 8) [33]. Because both the helicase and nuclease activities of Hef are presumably needed for biological function, we also assumed that Hef is directly implicated in processing of arrested replication forks. These hypotheses will be addressed in the future experimental work.
Moreover, our work has also shown that replication forks arrest spontaneously occurred in *H. volcanii* cells that may contain a high number of replication forks. This is exemplified by the fact that a typical *H. volcanii* cell contains 18 genome copies and that each molecule can be replicated by up to 8 replication forks. Whether all genome copies are replicated simultaneously is not known and this striking question needs to be addressed in future work to better understand replication dynamic in archaea. But as we have observed up to 15 to 20 arrested replication forks in some individual cells, our imaging studies rather suggest that several copies of the genome may be replicated in one cell.

### 5. Concluding remarks

*In vitro* characterization of Hef/XPF proteins suggested a key role in genomic stability similarly to what was observed in eukaryote. Genetic experiments coupled to dynamic localization of Hef proteins fused to the Green Fluorescence Protein further revealed that Hef is recruited at arrested replication forks in *Haloferax volcanii* cells. Experimental approaches we have recently developed for halophiles provide a valuable tool for studying functional roles of Hef at arrested replication forks in living archaeal cells. Understanding how Hef is recruited at arrested replication forks and the processing taking place to allow replication restart is our next challenge. Whether XPF protein have similar role in crenarchaeal cells remains to be addressed. As genetic tools have been developed for crenarchaeal organisms, future studies might tackle this issue [86, 87].

In conclusion, as archaea possess hallmarks of both bacterial and eukaryotic replication systems we believe that continuation of studies underlined will shed light on the evolutionary
history of replication restart mechanisms and its complex machinery that we are just starting to unravel in eukaryotes and now archaia.

Acknowledgements

We thank A. Lestini for figures 1, 2, 3 and 4. Figures 5, 6, 7 and 8 were reproduced with permission from Lestini et al. [33] published by Oxford Journals in Nucleic Acides Research. We also want to acknowledge the Agence Nationale de la Recherche for funding (ARCREP and RETIDYNA grants).

Author details

Roxane Lestini*, Floriane Delpech and Hannu Myllykallio

*Address all correspondence to: roxane.lestini@polytechnique.edu

Laboratoire d’Optique et Biosciences, Ecole Polytechnique, CNRS UMR–INSERM U696, Palaiseau Cedex, France

References


[32] Lestini R, Duan Z, Allers T. 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)2010 Sep 4;9(9):994-1002.


[34] Enzlin JH, Scharer OD. The active site of the DNA repair endonuclease XPF-ERCC1 forms a highly conserved nuclease motif. EMBO J2002 Apr 15;21(8):2045-53.


