Archaeal DNA repair: paradigms and puzzles

M.F. White

Centre for Biomolecular Science, University of St. Andrews, North Haugh, St. Andrews, Fife KY16 9ST, U.K.

Abstract

It is now generally accepted that the Archaea share many similarities in their information-processing pathways with eukarya. Archaeal and eukaryal DNA replication and transcriptional machineries show particularly striking similarities, and the archaeal processes have been used extensively as simpler models of the much more complex eukaryal ones. Archaeal DNA-repair pathways are not yet well characterized, and their relationship with repair pathways in bacteria and eukarya are still open to question. There are also strong distinctions between the major subdivisions crenarchaea and euryarchaea within the archaeal domain. This review highlights some of these similarities and differences using specific examples arising from our studies of the double-stranded and single-stranded DNA-binding proteins and the repair endonuclease XPF in the crenarchaeote Sulfolobus solfataricus.

Introduction

The idea that cellular life can be classified into the three fundamental domains of bacteria, Archaea and eukarya was first proposed by Fox and Woese in 1977 [1] on the basis of rRNA sequence analysis (Figure 1). Although initially greeted with a degree of scepticism, this theory has been vindicated by subsequent genome sequencing and bioinformatics comparisons. Woese's seminal ideas, coupled with the explosion in genomics, have resulted in a surge in interest in the molecular biology of the Archaea that shows no sign of abating. One area where the Archaea have proven to be particularly interesting concerns their informational pathways (DNA replication, transcription, translation etc.). The similarities between the archaeal and eukaryal DNA-replication and -transcriptional apparatus are particularly striking, and highly distinct from the equivalent proteins in bacteria [2,3]. This has raised a slew of questions about the nature of the last universal cellular ancestor (LUCA), and in particular about the adoption of DNA as the predominant genetic material and the evolution of DNA-replication machineries [4,5].

Our main area of interest lies in the characterization of DNA-binding and -repair proteins in the Archaea; an area where there is a much less clear-cut picture of the relationships between the three domains [6]. This short review will concentrate on the complex picture that emerges from bioinformatic and biochemical studies of these proteins, and will emphasize the differences apparent between the two major subdivisions of the Archaea: the euryarchae and crenarchae.

Archaeal chromatin proteins

All organisms use double-strand DNA-binding proteins to compact and package their genetic material. Eukarya, with their large, complex genomes, have a particular problem in DNA compaction and a parallel problem in ensuring and controlling the accessibility of DNA. The basic unit of DNA compaction in eukarya is of course the nucleosome, an octamer of histone proteins wrapping 146 bp of double-stranded DNA (dsDNA). Histone tails are targets for a variety of covalent modifications that regulate DNA accessibility by altering higher-order chromatin structure [7,8]. Bacteria and Archaea, with much smaller genomes, are thought to have much simpler methods for DNA compaction that lack the sophistication and levels of control seen in eukarya. Nevertheless, the euryarchaeal branch of the Archaea (Figure 1) have long been known to possess true histones that form tetrameric nucleosomes, wrapping about 70 bp of DNA [9,10]. Whereas the relationship between the eukaryal and euryarchaeal histones is clear, the latter show no evidence for covalent modification, and furthermore histones are absent from some euryarchae and all known crenarchaeae [11] (Table 1).

Recently, another dsDNA-binding protein has been identified in Archaea. Alba (‘acetylation lowers binding affinity’) is a dimeric protein that is found in many Archaea, in particular those growing at elevated temperatures (Table 1). Its distribution overlaps that of the archaeal histones and is not confined to either archael subdomain [12]. Alba binds dsDNA, affording protection against nucleases, but does not appear to exert any significant compaction [13,14]. Notably, the Alba protein is subject to covalent modification by acetylation of a single lysine residue. Acetylated Alba has a lowered affinity for dsDNA, and acetylation is reversed by the NAD+-dependent deacetylase Sir2 [15]. This is therefore the first hint that Archaea may have rather more complex mechanisms for control of chromatin structure and therefore gene expression than previously thought. As the subject of

Key words: Archaea, chromatin, DNA repair, nucleotide excision repair, single-stranded DNA-binding protein (SSB).

Abbreviations used: NER, nucleotide excision repair; dsDNA, single-stranded DNA; SSB, single-stranded DNA-binding protein; RPA, replication protein A; LUCA, last universal cellular ancestor; dsDNA, double-stranded DNA; Alba, acetylation lowers binding affinity.

©2003 Biochemical Society

1 e-mail mfw2@st-and.ac.uk

Table 1. Key features of DNA-binding and DNA repair proteins in Archaea

<table>
<thead>
<tr>
<th>Protein</th>
<th>Subdomain</th>
<th>Temperature</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alba</td>
<td>Crenarchaeota</td>
<td>90°C</td>
<td>DNA binding and repair</td>
</tr>
<tr>
<td>Sir2</td>
<td>Euryarchaeota</td>
<td>37°C</td>
<td>DNA repair</td>
</tr>
<tr>
<td>NER</td>
<td>All</td>
<td>Variable</td>
<td>DNA repair</td>
</tr>
</tbody>
</table>

Figure 1. Phylogenetic tree of archael domains.

[Diagram not shown]
archaeal chromatin has been recently reviewed extensively [16], it will not be discussed further here.

Archaeal single-stranded DNA (ssDNA)-binding proteins (SSBs)

DNA is double stranded by default under most conditions. However, many cellular processes require the separation of DNA strands to form ssDNA. ssDNA is more susceptible to DNA damage, and lacks a partner strand to act as a template for repair. Accordingly, ssDNA is shepherded by SSBs that control its formation, protect it and allow reannealing to occur in a controlled way. SSBs can also melt dsDNA, detect DNA damage and recruit repair proteins [17]. The importance of SSB function is underlined by the observations that, firstly, SSBs are present in all known cellular organisms and many viruses and, secondly, that the basic unit of ssDNA binding, the OB fold, is conserved in every life form and must have been one of the basic types of protein domain present in LUCA.

Bacteria utilize a homotetrameric SSB, with each subunit consisting of a single OB fold for ssDNA binding plus a flexible C-terminal tail for protein–protein interactions [18]. In contrast, the eukarya have a heterotrimeric SSB known as replication protein A (RPA) with six OB folds, four that bind ssDNA and two that mediate subunit interactions [19,20]. The third OB fold in the large (RPA70) domain is interrupted by a zinc-binding domain. What about the Archaea? It is becoming a theme of this review that this question tends to result in rather complicated answers, and this is no exception. The euryarchaea appear to utilize an RPA-like SSB; that is to say a polypeptide or polypeptides with multiple OB folds and the characteristic zinc domain [21,22]. There seems to be a fair degree of variation, with some euryarchaeal SSBs appearing heterotrimeric, others heterodimeric.

The crenarchaeal SSB is another surprise: on initial analysis it resembles the bacterial SSB very closely, with a single OB fold followed by a flexible C-terminal tail [23]. The tail is not required for DNA binding and is known to interact with other proteins [23] and D.J. Richard and M.F. White, unpublished work). However, more careful analysis of the OB-fold domains shows that the eukaryal and archaeal

![Figure 1](image.png)

**Figure 1** | The tree of life, based on analysis of rRNA sequences, as proposed by Fox and Woese

LUCA, last universal cellular ancestor.

<table>
<thead>
<tr>
<th>Species</th>
<th>Short XPF</th>
<th>Long XPF</th>
<th>PCNA</th>
<th>RPA</th>
<th>SSB</th>
<th>Histones</th>
<th>Alba</th>
<th>UvrABC</th>
<th>Optimum temperature (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Aeropyrum pernix</em></td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>95</td>
</tr>
<tr>
<td><em>Pyrobaculum aerophilum</em></td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td><em>Sulfolobus solfataricus</em></td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>80</td>
</tr>
<tr>
<td><em>Sulfolobus tokodaii</em></td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>80</td>
</tr>
<tr>
<td><em>Thermoplasmophilum</em></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>60</td>
</tr>
<tr>
<td><em>Thermoplasma volcanii</em></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>60</td>
</tr>
<tr>
<td><em>Methanopyrus kandleri</em></td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td><em>Pyrococcus furiosus</em></td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td><em>Pyrococcus horikoshii</em></td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td><em>Pyrococcus abyssi</em></td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td><em>Methanococcus jannaschii</em></td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>85</td>
</tr>
<tr>
<td><em>Archeoglobus fulgidus</em></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>85</td>
</tr>
<tr>
<td><em>Methanothermobacter thermotrophicus</em></td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>70</td>
</tr>
<tr>
<td><em>Methanosarcina acetivorans</em></td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>&lt;40</td>
</tr>
<tr>
<td><em>Methanosarcina mazei</em></td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>&lt;40</td>
</tr>
<tr>
<td><em>Halobacterium</em></td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>&lt;40</td>
</tr>
</tbody>
</table>

©2003 Biochemical Society
sequences share a higher level of sequence identity, and this has recently been confirmed by the solution of the crystal structure of *Sulfolobus solfataricus* SSB [24]. Although this supports the closer relationship of Archaea and eukarya inferred from the tree in Figure 1, it raises the question of how the crenarchaeal protein has ended up with these chimaeric features: a eukarya-like OB fold coupled to a bacteria-like tail. A further question remains unanswered: although it was stated above that all organisms use SSBs, it is very difficult to detect a plausible example in the genome of the crenarchaeote *Pyrobaculum aerophilum* (Table 1; M.F. White, unpublished work). Assuming *Pyrobaculum* does code for a SSB, it may be even more diverged than those of its crenarchaeal cousins.

**Nucleotide excision repair (NER) in the Archaea**

Many types of bulky lesion, including photoproducts induced by UV irradiation, are repaired by the NER pathway. NER requires the initial detection of damage, unravelling of the two strands with a helicase activity, and concerted strand scissions introduced on either side of the lesion, allowing the removal of a patch of ssDNA containing the damage so that the undamaged strand can act as a template for repair synthesis [25]. In bacteria, this task is accomplished in typically sparing fashion by the UvrABC endonuclease complex, with a diagnostic patch size of approx. 11 nucleotides [26]. In contrast, eukarya take the usual approach of using a ‘sledgehammer to crack a nut’, with up to 30 proteins acting in concert to detect and remove a larger patch of 24–30 nt [27]. And the Archaea? Well, as you may have guessed, the situation is a bit complicated in Archaea. Initial assays of archaeal NER yielded a patch size of about 10 nt, and on that basis a bacteria-like repair pathway was predicted [28]. Indeed the archaeon studied, *Methanothermobacter thermautotrophicus*, does possess a full set of very clear UvrABC orthologues. Suspiciously clear, in fact, and rather too close to have been separated by the required billions of years of evolution. In fact, only a select group of Archaea (mesophilic methanogens and halophiles) have UvrABC orthologues, and this is likely to represent a case of relatively recent lateral gene transfer from bacteria to Archaea.

The Archaea that don’t have UvrABC (and indeed some of those that do) do have detectable orthologues of the nucleases XPF and XPG, and the helicases XPB and XPD, that are implicated in eukaryal NER. On this basis, we can tentatively conclude that the original NER pathway in the Archaea was eukaryal in character, and has been supplanted in some species by the bacterial version. However, there is so far little experimental evidence for NER in the Archaea, and furthermore there is a puzzling lack of any obvious homologue of the eukaryal damage-recognition proteins XPA and XPC. It is of interest to note that the recently published genome sequence of *Plasmodium falciparum* lacks homologues of XPA and XPC, which may indicate that these are an invention of more complex eukarya [29].

**The repair nuclease XPF**

The endonuclease XPF introduces a strand scission on the 5′ side of a DNA lesion during eukaryal NER. Human XPF has a structural specificity for 3′-flap structures [30], and forms a heterodimer with ERCC1 [31], which in turn interacts with the eukaryal single-stranded-DNA-binding protein RPA [32]. The C-terminal third of XPF encodes the nuclease domain of the protein, and may constitute a novel nuclease motif [33]. The N-terminal two-thirds resembles a helicase domain, but appears degenerate and is probably not an active helicase [34]. The euryarchaia have a close homologue of eukaryal XPF, with a C-terminal nuclease domain and an N-terminal domain that may be a functional helicase [34,35]. In contrast, the crenarchaea encode a protein that appears to have only the C-terminal nuclease domain. We have recently characterized this protein, and shown that it is indeed a functional 3′-flap endonuclease [36]. However, its activity is dependent on the sliding clamp PCNA (a heterotrimer in *S. solfataricus* [37]), with which it interacts via a characteristic C-terminal peptide motif [38]. This is reminiscent of the requirement for PCNA of the 5′-flap endonuclease XPG [39]. The sequence and structural specificity of *Sulfolobus* XPF–PCNA is strikingly similar to that of human XPF–ERCC1, suggesting a fundamental conservation of function [36].

I propose a scheme for the evolution for XPF (Figure 2), starting from a core nuclease domain like that present in crenarchaeota, dependent on PCNA for DNA targeting. Subsequent fusion with a helicase, possibly one already involved in NER, produced the euryarchaeal enzyme...
and dispensed with the requirement for PCNA. The helicase subsequently degenerated in eukarya, presumably because its cellular function was taken over by another protein. At some point in eukarya, duplication and divergence of the nuclease domain of XPF resulted in the ERCC1 protein, which provides a bridging function in the complex eukaryal NER machinery. A similar process probably resulted in the evolution of the Mus81 protein, which is an XPF homologue that is specialized for the processing of stalled replication forks [40,41].

Relationship of euryarchaeae, crenarchaeae and eukarya

There is still vigorous debate about the precise relationship of the archaenal and eukaryal lineages, and due to the paucity of information and the scrambling effects of lateral gene transfer we can never hope to know for sure exactly how life evolved into three domains. Restricting ourselves to the proteins described in this review, and summarized in Table 1, we can make the broad generalization that the euryarchae appear more similar to the eukarya than do the crenarchaeae. This is based on the observations that euryarchaeae have histones (mostly), a full-length XPF homologue (mostly), a single homotrimeric processivity factor and an RPA-like SSB. There are, no doubt, numerous counter-examples, and it is also true that it is almost impossible to make any broad statement about archaenal proteins; there will always be exceptions, such as the missing SSB of *Py. aerophilum*. It is also clear that the *Thermoplasma* spp. have been engaged in extensive gene-swapping with crenarchaeae, and have probably traded in their histone proteins for the bacterial chromatin protein HU [16]. What is certain is that future studies of archaenal DNA-repair proteins will continue to yield surprises and insights of intrinsic interest and relevance to human DNA repair.

M.F.W. is a Royal Society University Research Fellow. Thanks to the Biotechnology and Biological Sciences Research Council and Wellcome Trust for grant support.

References


Received 18 January 2003