Mini review

Archaeal histones: dynamic and versatile genome architects

Bram Henneman and Remus T. Dame *

Leiden Institute of Chemistry, Gorlaeus Laboratories, Leiden University, Leiden, The Netherlands

* Correspondence: E-mail: rtdame@chem.leidenuniv.nl.

Abstract: Genome organization and compaction in Archaea involves different chromatin proteins, among which homologues of eukaryotic histones. Archaeal histones are considered the ancestors of their eukaryotic counterparts, which is reflected in the way they position along the genome and wrap DNA. Evolution from the archaeal modes of action to the prototypical eukaryotic nucleosome may be attributed to altered histone-histone interactions and DNA sequence determinants cooperating to yield stable multimeric structures. The identification of a new candidate phylum, proposed to be a missing link between archaea and eukaryotes, Lokiarchaeota, may be instrumental in addressing this hypothesis.

Keywords: archaeal chromatin; genome architecture; nucleoid; Archaea; Lokiarchaeota; histone; HMf; HTk; Alba

1. Introduction

Organisms from the domain Archaea—one of the three domains of life—share similarities with both bacteria and eukaryotes [1]. While these organisms have a cell shape, mode of replication and some metabolic routes that are similar to those in bacteria, archaea are genetically closer related to eukaryotes. This is emphasized by the recently proposed two-domain theory of life [2]. Generally, archaeal genomes consist of a single circular chromosome, as in bacteria; some archaeal species, notably members of the phylum Euryarchaeota express histones, which are similar to histones H3 and H4 found in eukaryotic cells in terms of quaternary structure [3]. Histones are the prototypical organizers of genomic DNA in eukaryotes, wrapping DNA around a generally octameric histone core into nucleosomes, imposing a first level of organization within chromosomes [3]. Archaeal nucleosomes are generally thought to be composed of two histone dimers, which, unlike their
eukaryotic counterparts, lack a flexible tail at the amino-terminus that is subject to post-translational modification \[4,5\]. In addition, archaea express other small abundant proteins, which are believed to be involved in genome architecture, such as Alba \[6\]. Histone proteins are absent in bacteria, which instead express a number of small abundant proteins, which are believed to be involved in genome architecture, as well as regulation of gene expression \[7–9\].

Recently, a new candidate archaeal phylum, Lokiarchaeota, was proposed based on metagenomic sequencing of microbes in deep marine sediments. The genomes of Lokiarchaeota contain genes that previously had been found exclusively in eukaryotes \[10\]. This includes genes coding for actin homologues, small GTPases and components of the ESCRT-complex, involved in membrane budding and vesicle formation. Therefore, these newly discovered organisms may be important evolutionary links between archaea and eukaryotes. Here, we describe the identification of histones and other chromatin proteins in the genomes of Lokiarchaeota, and discuss them in relation to chromatin proteins described in other archaea and in eukaryotes. Specifically, we discuss the apparent differences in mechanism by which archaeal histones affect genome architecture and the role of histones in transcription regulation, with emphasis on recent developments in this field.

2. Structure and Alignment

Histones are the only archaeal chromatin proteins that resemble the eukaryotic histones H3 and H4 in structure and sequence \[11\]. These proteins occur in every archaeal phylum, although histone-coding genes are not conserved throughout all orders of the phylum Crenarchaeota \[12\]. The best-studied histones are HMfA and HMfB, two histones from the euryarchaeon \textit{Methanothermus fervidus}, which are closely related in terms of sequence \[13\]. These histones occur in solution as homo- or heterodimers, but have been shown to bind to specific DNA sequences as tetramers, possibly resembling the (H3-H4)\textsubscript{2} tetrasomes formed as intermediates during histone octamer assembly in eukaryotes \[11,14\]. Archaeal histone tetramers were shown already in early studies to protect ~60 bp of DNA from micrococcal nuclease (MN) digestion \[15\] and to assemble at specific high-affinity sequences \[16,17\]. Conversely, MN digestion of chromatin from the euryarchaeal species \textit{Thermococcus kodakarensis} yields digests with fragments 60 bp in size, representing a tetrameric histone, but also fragments of 30 bp and multiples of 30 bp up to ~500 bp. This observation indicates that ‘nucleosomes’ formed by histone HTkB are not of defined size and suggests that dimers, tetramers and larger multimers are functionally relevant (either in chromatin organization or in relation to transcription regulation) \[18\]. Although in our model we envision interactions between adjacent nucleosomes and archaeal histones lack N-terminal tails as found in eukaryotic histones, it is analogous to the interactions found to exist between eukaryotic nucleosomes, mediated by an acidic patch formed by H2A and H2B and the N-terminal tail of H4 of a neighboring nucleosome \[19,20\]. Building further upon a model proposed by Maruyama \textit{et al.}, we envision a model for multimerization of histone proteins, in which histone dimers can be added and removed from both ends to assemble into a DNA-coiling multimer that consists of dimers that each cover 30 bp of DNA (figure 1C). Studies on \textit{Methanothermobacter thermautotrophicus} chromatin are in accordance with this observation, which suggests that the basic functional unit of archaeal histones might not be a tetramer, as generally assumed, but a dimer \[18,21\]. The observed differences between species might be due to differences in genomic sequence and/or protein sequence, altering dimer-dimer interactions.
Figure 1. Model for multimerization of histone proteins in *Thermococcus kodakarensis* and *Methanothermobacter thermautotrophicus*. Blue: histone dimers; red: DNA duplex. A. DNA is bent by a histone dimer, which covers 30 bp of DNA. B. A histone tetramer wraps the DNA in a left- or right-handed manner and is able to switch between these configurations, for example in response to changes in salt concentrations [22]. C. Histone dimers can further associate to form a larger multimer. Every additional dimer wraps another 30 bp of DNA. At both extremities of the multimer, dimers can be added or removed. The right-handed wrapping depicted here is arbitrary, since also this large histone-multimer-DNA complex might accommodate both left-handed and a right-handed wrapping.

Although sequence homology between histones is sometimes limited, the histone fold (i.e. typical structural elements and their configuration) is well conserved between species [23]. Histones consist of three α-helices separated by short β-loops (figure 2A). Residues P4, R10, R19, E33, K53 and T54, located in the N-terminal α-helix and the C-terminal β-loop, were, based on crystal structure and mutagenesis studies, predicted to be mainly responsible for interacting with the DNA backbone. Dimer-dimer interaction, tetramerization, has been proposed to be mediated by residues L46, H49, A50, D59 and L62 in the C-terminal and middle α-helix and the C-terminal β-loop [22,24]. Euryarchaeal histones have a high degree of sequence identity and are very similar in length, although some have an extended N-terminal loop and/or helix, or an extended C-terminus of ~30 amino acids (not shown in the alignment). It has been proposed that these unconventional histone proteins are modulators of DNA binding [23,25]. Also, some histones from halophilic archaea consist of two histone domains fused together, whereas other histones assemble into dimers.
following synthesis [26]. In these naturally occurring histone-histone fusion proteins, the histone fold

A.

Figure 2. Alignment of histone protein sequences from several eukaryotic and archaeal species and Alba from several archaeal species. Secondary structure is indicated at the top of the figure; dark gray bars represent α-helices, light gray bars represent β-sheets. Arrows indicate amino acid residues of which the function is described in the text. A. Alignment of histone proteins. Numbers and secondary structure are based on HMfB from \textit{M. fervidus}. Red and bold: residues that are shared between most lokiarchaeal histone proteins and other archaea but not with eukaryotic histones; green and bold: residues that are shared between at least one eukaryotic histone and at least two lokiarchaeal histone protein but not with histone proteins from other archaea; blue and bold: residues from Lokiarchaeal histone proteins of known function which are different in all other organisms in this alignment. H3 and H4: from \textit{Drosophila melanogaster}; HMfA and HMfB: \textit{Methanothermus fervidus}; HTkA and HTkB: \textit{Thermococcus kodakarensis}; HstA: \textit{Haloferax volcanii} (N-terminal and C-terminal histone domain shown separately); HLK-A-E: Lokiarchaeota (GenBank accession numbers KKK41688.1, KKK44894.1, KKK40642.1, and KKK45508.1, respectively), possibly from different species. B. Alignment of Alba proteins. Numbers and secondary structure are based on Alba1 from \textit{Sulfolobus solfataricus}. Blue indicates residues which are poorly conserved among Alba proteins (divergent residues in Lokiarchaeota in bold). Sso: \textit{Sulfolobus solfataricus}; Tko: \textit{Thermococcus kodakarensis}; Mfe: \textit{Methanothermus fervidus}; Mma: \textit{Methanococcus maripaludis}; Loki Alba1, Alba2: Lokiarchaeota (GenBank accession numbers KKK43110.1, KKK44501.1, and KKK43470.1, respectively), possibly from different species.
at the C-terminus is more similar to the histone fold of other euryarchaeal histones than the one at the
N-terminus [23,27]. HTkB and HMtA2, which bind along the genome as a dimer, tetramer or larger
multimer in *T. kodakarensis* and *M. thermautotrophicus*, respectively, contain all residues identified
as responsible for tetramerization in HMfB from *M. fervidus*. Therefore the difference in minimal
size of the functional histone unit cannot be attributed to residues known to be involved in
tetramerization.

We scrutinized the published metagenome sequence of Lokiarchaeota [10]. We identified five
histone-coding sequences in the genomes of this novel phylum. These histones are also partly similar
in sequence to eukaryotic histones (figure 2A). Interestingly, some residues are shared between
Lokiarchaeota and eukaryotes, but not with Euryarchaeota. These residues include V25 and to a
lesser extent Q14, which are of unknown function, but occupy positions that in other histones are
positively charged (indicated in green). Conversely, Lokiarchaeota share some residues with histone
proteins from Euryarchaeota but not with eukaryotes, of which P7, V20 and A36 are most obvious
(red). The hybrid nature of the lokiarchaeal histone proteins might be a reflection of Lokiarchaeota
being an intermediate between Archaea and Eukaryotes. Furthermore, histones from Lokiarchaeota
lack some but not all of the residues that in Euryarchaeota were predicted to be responsible for
interaction with DNA and tetramerization, most notably P4, R19 and T54 in all five Lokiarchaeal
histones identified here (blue), but also E33, A50, D59 and A62 in individual cases. This suggests
that DNA-binding and tetramerization in histones from Lokiarchaeota occur differently than in other
histones.

Further, we identified three more chromatin proteins in Lokiarchaeota, which are members of
the Alba protein family. Alba proteins, which have been shown to bind in trans between two DNA
duplexes as well as in cis along a single DNA duplex, have been found in all archaeal phyla. Some
organisms, such as *Sulfolobus solfataricus*, express two Alba variants, Alba1 and Alba2, with
distinct architectural properties, attributed to a phenylalanine at position 60 in Alba1, which is
crucial to (cooperative) binding in trans and in cis [6,28]. Based on this definition, the genomes of
Lokiarchaeota encode two Alba1 homologues and one Alba2 homologue (figure 2B). The two Alba1
homologues, which differ only at positions 53, 54, 58 and 59, are highly similar to the Alba(1)
proteins of *T. kodakarensis*, *M. fervidus*, *Methanococcus maripaludis* and *S. solfataricus*, but
similarity between Alba2 from Lokiarchaeota and Alba2 from *S. solfataricus* is much smaller. The
Alba homologues from Lokiarchaeota lack K16, the residue that is present in most Alba homologues
and can be (de-)acetylated to modulate DNA binding [29]. Instead, both Alba1 homologues contain a
lysine on positions 17 and 18, which might also be targets of similar post-translational modification.
Besides histones and Alba, we were not able to identify any other known chromatin proteins in this
phylum.

3. Binding Motif

Archaeal histones have a preference for binding GC-rich sequences with alternating (G/C)_{2/3}
and (A/T)_{2/3} motifs, which are separated by half a helical turn. This compresses the minor and major
groove on one side of the helix with A:T facing towards the histones and G:C facing outwards,
causing the DNA to bend. This mechanism is very similar to the one found in eukaryotes. Studies on
HMfB from *M. fervidus* reveal that binding motifs are more complex: AT and GC are not equal to
TA and CG, respectively, in terms of facilitating archaean nucleosome assembly [17]. Also, when
analyzing substitution patterns in the DNA of nucleosomes in sister lineages of *Haloferax volcanii*, it was found that there is a preference for G:C near the dyad, whereas changes towards A:T were more common further away from the dyad, near the ends of the nucleosome DNA [30,31]. These observations are again very similar to observations reported for human nucleosome positioning sequences.

In eukaryotes, a histone wraps 147 bp of DNA 1.65 times around its core, by means of which the DNA is compacted and supercoils are constrained [32]. In archaea, tetrameric histones have been reported to wrap DNA without making a full turn, which results in a horseshoe-like conformation [33]. However, interactions with the DNA may determine the extent of wrapping, which means DNA sequence motifs may affect the exact number of turns that is made by the DNA.

Furthermore, HMf proteins from *M. fervidus* can wrap DNA around its tetramer core, but also are able to bend DNA as a dimer *in vitro* [34]. Wrapping can be considered an advanced form of bending, rather than a separate mechanism, and may be an evolutionary consequence of acquiring dimers at adjacent high-affinity sites. *In vivo* bending and wrapping by histones are expected to occur in parallel. Not only wrapping by histones, but also loop formation by trans-acting elements such as Alba contributes to DNA organization and compaction. Loop formation is an important organizing principle among prokaryotes; it may stabilize topologically isolated domains in which supercoiling is preserved, and in which a subset of genes can be co-regulated [7,35,36].

### 4. Histones as a Transcription Factor

Analogous to many chromatin proteins throughout the domains of life, histone proteins likely also play roles as transcription factors, possibly functionally modulated by tetramer composition, supercoiling and other chromatin proteins [12,23,37,38]. Archaeal histones are able to repress or activate transcription, dependent on genomic context and growth phase [39]. The genomes of histone-expressing archaea harbor one to six histone genes, encoding histones that are slightly different from each other. Knocking out a subset of histone genes does not result in significant changes in growth rate, but strains devoid of any histone gene exhibit severely hampered growth in most species [39–42]. Histone proteins can form both homo- and heterodimers, which means that an organism that expresses two histone proteins can form six different histone tetramers. It was shown *in vivo* and *in vitro* that binding patterns are different for homodimers of HtkA compared to HtkB in *T. kodakarensis* [43]. Studies on HMfA and HMfB from *M. fervidus* showed that HMfA is prevalent during exponential growth, while HMfB is the main histone protein in the stationary phase. Similar observations have been reported for other archaeal species [39,44]. Also, HMfB has different DNA-binding affinity and supercoiling activities compared to HMfA [45]. Combined, these data suggest that the composition of the nucleosome modulates gene regulation, probably on a local level.

Although *in vivo* evidence lacks to date, it has been shown that the nucleosome slows down transcription elongation by RNA polymerase *in vitro*, thereby acting as a repressor of transcription [21]. Analysis of the position of nucleosomes on *T. kodakarensis* genomic DNA *in vivo* and *in vitro* showed that very active and vital genes, such as the ribosomal DNA operon, are not occupied by nucleosomes, whereas adjacent genes do contain nucleosomes [43]. Also, nucleosome depletion was observed at intergenic regions, which harbor the promoter and transcription termination elements. Nucleosome-depleted transcriptional start sites did contain nucleosomes up- and downstream of their location on the genome [30]. In addition, binding of histones is regulated in
an indirect way by the interplay between histones and the chromatin protein TrmBL2, which is affected by supercoiling. HTmB from *T. kodakarensis* competes with the highly abundant protein TrmBL2, which can form stiff filaments on the DNA in a non-specific manner at low KCl concentrations (<300 mM), and in a sequence-specific manner at high KCl concentrations (>300 mM) *in vitro* [38,46]. By forming filamentous structures along DNA, TrmBL2 antagonizes DNA packaging and transcription repression by histones; it should however be noted that these filamentous structures may in turn cause transcription repression. TrmBL2 is only stable on relaxed DNA; over- or underwinding will reduce DNA-binding interaction of TrmBL2 but does not affect binding interaction between histones and DNA. Archaeal histones can easily adapt to changes in supercoiling because accommodate both right- and left-handed wrapping configurations as a tetramer, a property that is shared with (H3-H4)_2 tetramers [22,38,47,48]. However, octameric eukaryotic histones solely wrap DNA in a left-handed manner. Archaeal histone multimers may adopt the wrapping configuration of the tetramer-DNA complex at the moment it binds other dimers to form a larger multimer, and may not be able to change handedness after multimerization (figure 1B, 1C). The wrapping configuration of archaeal histones has been shown to be influenced by ionic strength: HMfA and HMfB wrap the DNA in a right-handed manner at low salt concentrations, corresponding to positive supercoiling, whereas negative supercoiling is generated at high salt concentrations. [22,48]. In eukaryotes, nucleosomes are thought to restrain supercoiling and to respond structurally to changes in DNA topology, which may create more favorable conditions for binding of transcription factors and RNA polymerase [49,50]. A similar interplay between DNA topology and genome structure/accessibility may also occur within archaeal chromatin. Since TrmBL2, unlike histones, loses affinity for supercoiled DNA, introduction of supercoiling by topoisomerases could remove TrmBL2 from the DNA, allowing histones to wrap and thereby silence a gene or operon. Furthermore, Alba could introduce topologically isolated domains in which supercoiling can be locally regulated, thereby regulating transcription of a relatively small selection of genes. This suggests that histones are part of a system, which influences transcription and which can act both globally and locally.

5. Conclusion

Archaeal histones appear as dynamic and versatile organizers of the genome due to different binding stoichiometries, histone composition and DNA sequence determinants. These properties may give clues to the evolutionary path from dimeric histones, simply bending DNA analogous to other architectural DNA bending proteins, to tetrameric and larger multimeric structures, such as eukaryotic histone octamers. We propose that clustering of DNA-bending histone dimers, driven by dimer-dimer interactions and/or evolutionary clustering of sequence determinants enhancing cellular fitness, underlies formation of the ‘archaeal’ tetrameric nucleosome. Due to further clustering and recruitment of H2A and H2B proteins this ‘simple’ nucleosome evolved into the prototypical eukaryotic nucleosome. The genomes of Lokiarchaeota, as well as other archaeal phyla yet to be identified, potential missing links between archaea and eukaryotes, could be instrumental in investigating this hypothesis.
Acknowledgments

Research on the topic of this review is supported by grants from the Netherlands Organization for Scientific Research [VIDI 864.08.001], High Tech Systems & Materials NanoNextNL program 8B, the FOM Foundation for Fundamental Research on Matter program ‘Crowd management: The physics of genome processing in complex environments’ and the Human Frontier Science Program [RGP0014/2014].

Conflict of Interest

The authors declare no conflict of interest in this research.

References


