Commentary

Archaeal chromatin: Virtual or real?

Jordanka Zlatanova*

Department of Biochemistry and Biophysics, Oregon State University, Corvallis, OR 97331-7305; and Institute of Genetics, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria

As we find out more about the archaeabacteria our sense of their strangeness increases, but its explanation lies in a shared ancestry, not in individually evolved idiosyncrasies.

Carl Woese (1)

The more one reads about Archaea, the more one sees a whole new world, revealing itself. What are these mysterious organisms, how did they come into existence, and why did it take us so long to even recognize their distinctness among living creatures? Where does the watershed lie between these microbes and the rest of the living world? What about the dichotomous division between prokaryotes and eukaryotes? The answers to many of these questions are found in the enlightening accounts of Woese (e.g., refs. 1 and 2). Indeed, from the viewpoint of a systematist, animals and plants are rich in complex morphological detail, which can serve as the basis for their systematics; however, not much phylogenetic information can be derived from the simple morphologies and physiologies of the bacteria. Only the revolutionary capability to sequence nucleic acids turned bacterial systematics into a tangible enterprise.

Molecular comparisons, originally of rRNA sequences, followed by analysis of several protein families (discussed in ref. 3) prompted a revolutionary change in our view of the evolutionary relationship among living organisms. The three-domain concept was forwarded (4): life on Earth comprises three domains, Bacteria, Archaea, and Eucarya, each of which contains several kingdoms (Fig. 1). The proposed universal phylogenetic tree places the Archaea in a clearly distinct realm of organisms, differing in fundamental ways from the Bacteria and Eucarya. Although from a cytological point of view Archaea are prokaryotic (they lack nuclei, cytoskeleton, and organelles, see Table 1), at the molecular level they represent a complex mosaic of features of either prokaryotic or eukaryotic nature, as well as their own unique features. This mosaic is itself partitioned into two aspects: metabolism and information processing. Molecular analyses, culminating in the sequencing of the complete genomes of Methanococcus jannaschii (5) and Methanobacterium thermoautotrophicum strain ΔH (6), provide a clear general view of this structural dichotomy. Although the metabolic facets of these organisms are closely related to those seen in Bacteria, the molecules, and hence probably the mechanisms, involved in information processing seem to be recognizable more similar to those in Eucarya (for description of some molecular features of Archaea, in comparison with Bacteria and Eucarya, see Table 1).

If Archaea and Eucarya have a common evolutionary ancestor, it must have had some of the molecular features of the information-processing apparatus seen in a modified form in both domains. If this is the case, one may ask how the genome is organized. Eucarya package and regulate (in part) the activity of their genomes by organizing them into chromatin, a nucleoprotein fiber built of nucleosomal particles separated by linker DNA (7, 8). Do Archaea have a structural counterpart of chromatin? It is this important question that has been experimentally addressed in the paper by Pereira et al. (9), published in this issue.

The Archaea "chromatin" field developed along two lines of independent research: the description of histone-like proteins and their interaction with DNA in vitro, and electron microscopy (EM) visualization of nucleoprotein fibers spread out of cells. A small, basic, histone-like protein, HTa, that can organize DNA into nucleosome-like structures (NLS) was identified in Thermoplasma acidophilum (10, 11). The same work presented EM images of native nucleoprotein spreads that contained globular particles 5–6 nm in diameter along the DNA fibers (see below). High-quality, fairly convincing "chromatin" fiber images from Halobacterium salinarium were reported later (12, 13) (unfortunately, no relevant biochemical data are available on this species). Finally, many studies, mainly from Reeve's laboratory, have been reported on the isolation, sequences, and DNA-binding properties of a family of HL proteins now known as the HMf family of archaeal histones (for reviews see refs. 14 and 15).

The Histone Proteins of Archaea

The histone family in Archaea, named after its first described members, the two closely related HL proteins from Methanothermus fervidus (16), now contains 18 members (refs. 5, 14,

*To whom reprint requests should be addressed at: Oregon State University. e-mail: zlatanoj@ucs.orst.edu.

© 1997 by The National Academy of Sciences 0027-8424/97/9412251-4$2.00/0
PNAS is available online at http://www.pnas.org.

FIG. 1. The universal phylogenetic tree, as seen by Woese and collaborators (adapted from ref. 2), showing the three domains. The branching order and branch lengths are derived from sequence comparisons of rRNA. The branches pointing to different lineages of organisms are presented only for the domain Archaea. + Denotes archaean groups known to contain proteins homologous to eucaryal core histones (note that none have been so far reported for kingdom Crenarchaeota). * Denotes groups in which chromatin-like organization has been directly visualized by electron microscopy (EM).
and 15; J. N. Reeve, personal communication). They are strongly conserved in sequence among themselves; moreover, and of particular importance, they have strong sequence similarities with the folded regions of eucaryal core histones. In fact, each consensus sequence for a eucaryal core histone is more similar to the HMf histone sequences than to the other core histone sequences (15), putting the two groups of proteins into a distinct group, separate from all other DNA-binding proteins.

The resemblance of the two histone groups is further substantiated from secondary structure predictions (15, 17, 18) and analysis (19). Both protein classes possess the “histone fold,” originally described by Arents et al. (20) in all four core histones. The fold consists of three α-helices interconnected by loop/β-strand segments (Fig. 2). The histone fold allows the histone monomers to dimerize by the antiparallel pairing of their long central α-helices, with the formation of the so-called handshake motif. NMR solution studies of recombinant HMfB (19) demonstrated that the histone fold in the core histones is superimposable with that in HMfB; moreover, two HMfB monomers interact with each other to form the handshake motif characteristic of core histone heterodimers (Fig. 2B).

HMf proteins exist as dimers in solution, as demonstrated by gel filtration and chemical cross-linking (15). Both homo- and heterodimers can be formed in vitro, and mixtures of the three possible dimers, (HMfA)₂, (HMfB)₂, and HMfA–HMfB, have been directly isolated from cells. Because the accumulation of the two monomers is dependent on growth phase, and they have somewhat different DNA-binding features, it has been suggested that the cell combines the relative abundance of different dimers with their different binding properties to differentially regulate the structure of the M. fervidus chromosome during growth (15). Finally, and presumably of physiological relevance, when cross-linking is performed on DNA-bound proteins, tetramers can be formed, in addition to dimers.

### Formation of NLS in Vitro

The first NLS formed by the interaction of an archaeal HL protein (HTa) with naked DNA were reported back in 1980 (10). The relationship of these particles to those studied later in Reeve’s laboratory is unclear partly because HTa does not

---

**Table 1. Major cytological and molecular features of the domain Archaea**

<table>
<thead>
<tr>
<th>Property</th>
<th>Bacteria</th>
<th>Archaea</th>
<th>Eucarya</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cytological features</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nucleus</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Cytoskeleton</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Organelles (mitochondria, chloroplasts, Golgi apparatus)</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Molecular features</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA topology</td>
<td>Negatively supercoiled</td>
<td>Relaxed or positively supercoiled (in hyperthermophilic Archaea that contain reverse gyrase)</td>
<td>Negatively supercoiled</td>
</tr>
<tr>
<td>Promoter structure</td>
<td>Two conserved boxes at −10 (TATAAT) and −35 (TTGACA) from transcription start site</td>
<td>TATA box and/or initiator element</td>
<td>TATA box and/or initiator element</td>
</tr>
<tr>
<td>RNA polymerase</td>
<td>One type; relatively simple subunit composition; binds directly to promoter (can be footprinted)</td>
<td>One type; complex subunit structure (subunit pattern, genes, and serological properties similar to eucaryal RNA polymerase II); can be footprinted, but still requires basal transcription factors for promoter recognition</td>
<td>Three types; complex subunit compositions; cannot be footprinted; require basal transcription factors for promoter recognition/binding</td>
</tr>
<tr>
<td>Basal transcription factors</td>
<td>No</td>
<td>TBP, TFIIH, and TFIIH homologs of eucaryal RNA polymerase II-associated factors described thus far</td>
<td>TBP, TAFs, TFIIA, TFIIIB, TFIIE, TFIIH, TFIIH required for RNA polymerase II initiation; P-TEFb, TFIIIS, TFIIIF, elongin, and ELL required for elongation</td>
</tr>
<tr>
<td>Poly(A) tails in RNA</td>
<td>Short</td>
<td>Short (avg. 12 bases in length)</td>
<td>Long</td>
</tr>
<tr>
<td>Chromatin</td>
<td>No</td>
<td>?</td>
<td>Yes</td>
</tr>
</tbody>
</table>

For further features and references see the series of minireviews on Archaea published in the June 27, 1997, issue of Cell, and refs. 5 and 27.
Evidence for Particulate Deoxyribonucleoprotein (DNP) Structures in Situ

As mentioned above, EM images of spreads of DNA fibers revealed NLS in T. acidophilum (10) and H. salinarum (12, 13). The work of Pereira et al. (9) adds M. thermoautotrophicum to this list. In addition, it directly demonstrates by immuno-EM that the bead-like structures contain HMT. Again, as in the in vitro experiments described above, MNase digestion of formaldehyde-fixed protoplasts produces a short ladder of multiples of 60 bp, going up to a barely distinguishable trimer. Qualitatively similar results were obtained earlier on T. acidophilum nucleoprotein, with the formation of a stable 40-bp fragment and its dimer (10). Additional fragments of unknown nature can also be seen in the pattern, especially at positions between those of the monomer and dimer bands. Importantly, the monomeric DNP complexes obtained by MNase digestion were biochemically isolated and shown to contain monomers, dimers, and tetramers of HMT (these experiments were performed with M. fervidus).

Two other important observations are reported in ref. 9. First, it was estimated, from measurements of the amount of HMT in the cell, that there is enough of the protein (one tetramer per 67 bp of DNA) to compact the entire genome. How much of the genome is actually packed in chromatin remains to be established. Second, DNA from cross-linked complexes purified by immunoprecipitation with anti-HMT antibodies was probed for the presence of several genes. In every case, with one exception, the specific gene sequences were present in the complexes, albeit at considerably different levels. Understanding the implications of these results would require additional studies. At present the results suggest that differential organization of different genes in chromatin may constitute one level of their transcriptional regulation, as is the case for eucaryal chromatin.

Missing Pieces of the Puzzle

From the above, it seems clear that archaeal DNA may be organized in an archaeal version of eucaryal chromatin. The data, although appearing at an ever-increasing pace, only set the stage for much more research. What is the actual structure of the archaeal particles? How long is the DNA constrained by the protein core, and what is the sense of the DNA superhelix? How are the two partners, DNA and proteins, interacting? How are the particles interconnected to form a fiber and what is its structure? How does the chromatin fiber interact with the complex replication and transcriptional machineries to allow (regulated) processing of the genetic information? The answers to all these intriguing questions will require the concerted efforts of researchers from many different disciplines. The picture that will emerge as a result of this effort will undoubtedly contribute to our understanding of this fascinating domain of life, and of life in general.

I thank Dr. John Reeve for last-minute hectic discussions and for communicating unpublished data and data in press. I also acknowledge Dr. M. Summers for providing Fig. 2, Dr. S. H. Leuba for useful discussions, and Dr. K. E. van Holde for critical reading.