Proc. Natl. Acad. Sci. USA
Vol. 94, pp. 12251–12254, November 1997

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 archaebacteria 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 recognizably 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.

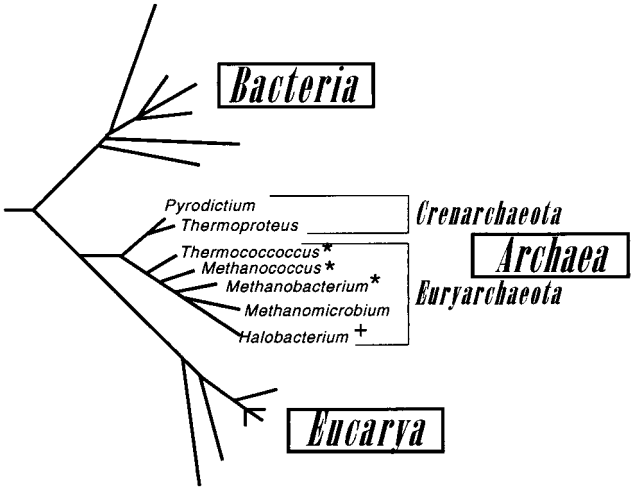

FIG. 1. The universal phylogenetic tree, as seen by Woese and collaborators (adapted from ref. 2), showing the three domains. The branching order and branch lengths are derived from sequence comparisons of rRNA. The branches pointing to different lineages of organisms are presented only for the domain *Archaea*. * Denotes archaeal 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).

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.

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

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

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.

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*A*). 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. 2*B*).

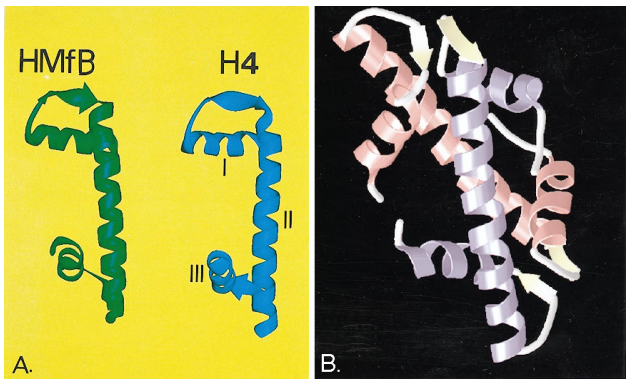

FIG. 2. (*A*) Ribbon representation of the three-dimensional fold of recombinant HMfB monomer (*Left*) (NMR structure, ref. 19), compared with that of the structured portion of chicken core histone H4 (*Right*) (crystal structure, ref. 20). Roman numerals denote helices I, II, and III, respectively. The similarity of structures is remarkable. (*B*) Ribbon model of a dimer of HMfB molecules, presenting the two monomers in the "handshake motif" typical of core histone heterodimers H2A-H2B and H3 and H4 (19, 20). Both images courtesy of M. Summers (University of Maryland, Baltimore).

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

seem to belong to the HMf family (15), although it does possess a limited but recognizable homology to eucaryal core histones (11); rather, it is perceived as a distant relative of the HU-family of bacterial HL proteins (15).

Studies involving *bona fide* members of the HMf family showed that these proteins, on binding to linear DNA molecules increased their electrophoretic mobility in agarose gels, suggesting some sort of DNA compaction. EM visualization of the complexes revealed NLS (e.g., ref. 16); the length of the constrained DNA was estimated to be around 90–150 bp. In addition, micrococcal nuclease (MNase) digestion produced a short ladder of fragments, multiples of ≈60 bp; a particle of ≈60 bp was the main product of more extensive digestion (21). The MNase digestion pattern was interpreted, by analogy with similar patterns of eucaryal chromatin, as resulting from preferential cleavage of naked DNA between NLS, with the DNA within the particles protected by virtue of its association with a tetramer of HMf. The MNase digestion results, however, raise several questions. Why is the new value of 60 bp so different from that previously estimated (90–150 bp; see above)? Why is the fragment so small? Does it reflect an internal cleavage at a specific exposed site within a larger particle, similarly to the situation observed with mononucleosomes reconstituted with $(H3–H4)_2$ tetramers only (22)? What does the observed ladder actually mean—that a small number of small particles of 60 bp are very closely packed with no linker DNA, or alternatively, that larger closely packed particles are also extensively internally cleaved? Finally, if we hypothesize that DNA is wrapped on the outside of a tetrameric HMf core, then the DNA will have to be severely bent, much more so than in the canonical eucaryal nucleosome (23). It is also not clear why fragments longer than 180 bp (''nucleosome trimers'') were never observed, even though the 200-fold molar excess of the protein over DNA (50-fold molar excess of histone tetramers) would certainly be expected to allow the formation of long ladders, especially on the pUC19 DNA.

Other published data may have a bearing on the role of HMf family in organization of archaeal DNA. The proteins preferentially bind to intrinsically curved DNA (24); however, when a set of cloned DNA fragments originating from MNase-resistant portions of reconstituted complexes were analyzed, no strong correlation between ''nucleosome'' positions and DNA curvature was noted (21). Another intriguing feature of the HMf-DNA complexes is that the DNA seems to be wrapped around the protein core in a right-handed superhelix (25), contrary to the sense of wrapping in eucaryal nucleosomes. This finding was recently explained by analogy with eucaryal nucleosomal particles, containing only the central $(H3–H4)_2$ tetramer. The DNA in such particles can apparently undergo a major, reversible switch from a left-handed superhelix to a right-handed one, and this switch is mediated by a similar switch in the topology of the tetramer itself involving a rotation of the two H3–H4 dimers about their H3–H3 interface (26). The relationship of the observation of Musgrave *et al.* (25) to the *in vivo* situation remains to be established.

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. salinarium* (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 cleavage were biochemically isolated and shown to contain monomers, dimers, and tetramers of HMf (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 HMf 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-HMf 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.

1. Woese, C. R. (1987) *Microbiol. Rev.* **51,** 221–271.
2. Olsen, G. J. & Woese, C. R. (1997) *Cell* **89,** 991–994.
3. Doolittle, R. F. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 2421–2423.
4. Woese, C. R., Kandler, O. & Wheelis, M. L. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 4576–4579.
5. Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., *et al.* (1996) *Science* **273,** 1058–1073.
6. Smith, D. R., Doucette-Stamm, L. A., DeLoughery, C., Lee, H., Dubois, J., *et al.* (1997) *J. Bacteriol.*, in press.
7. van Holde, K. E. (1988) *Chromatin* (Springer, New York).
8. Tsanev, R., Russev, G., Pashev, I. & Zlatanova, J. (1992) *Replication and Transcription of Chromatin* (CRC, Boca Raton, FL).
9. Pereira, S. L., Grayling, R. A., Lurz, R. & Reeve, J. N. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 12633–12637.
10. Searcy, D. G. & Stein, D. B. (1980) *Biochim. Biophys. Acta* **609,** 180–195.
11. Delange, R. J., Williams, L. C. & Searcy, D. G. (1981) *J. Biol. Chem.* **256,** 905–911.
12. Shioda, M., Sugimori, K., Shiroya, T. & Takayanagi, S. (1989) *J. Bacteriol.* **171,** 4514–4517.
13. Takayanagi, S., Morimura, S., Kusaoke, H., Yokoyama, Y., Kano, K. & Shioda, M. (1992) *J. Bacteriol.* **174,** 7207–7216.

14. Grayling, R. A., Sandman, K. & Reeve, J. N. (1996) *FEMS Microbiol. Rev.* **18,** 203–213.
15. Grayling, R. A., Sandman, K. & Reeve, J. N. (1996) *Adv. Protein Chem.* **48,** 437–467.
16. Sandman, K., Krzycki, J. A., Dobrinski, B., Lurz, R. & Reeve, J. N. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 5788–5791.
17. Arents, G. & Moudrianakis, E. N. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 11170–11174.
18. Baxevanis, A. D., Arents, G., Moudrianakis, E. N. & Landsman, D. (1995) *Nucleic Acids Res.* **14,** 2685–2691.
19. Starich, M. R., Sandman, K., Reeve, J. N. & Summers, M. F. (1996) *J. Mol. Biol.* **255,** 187–203.
20. Arents, G., Burlingame, R. W., Wang, B.-C., Love, W. E. & Moudrianakis, E. N. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 10148–10152.
21. Grayling, R. A., Bailey, K. A. & Reeve, J. N. (1997) *Extremophiles* **1,** 79–88.
22. Dong, F. & van Holde, K. E. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 10596–10600.
23. Morse, R. H. & Simpson, R. T. (1988) *Cell* **54,** 285–287.
24. Howard, M. T., Sandman, K., Reeve, J. N. & Griffith, J. D. (1992) *J. Bacteriol.* **174,** 7864–7867.
25. Musgrave, D. R., Sandman, K. M. & Reeve, J. N. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 10397–10401.
26. Hamiche, A., Carot, V., Alilat, M., De Lucia, F., O'Donohue, M.-F., Révet, B. & Prunell, A. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 7588–7593.
27. Reeve, J. N. (1992) *Annu. Rev. Microbiol.* **46,** 165–191.