The eukaryotic lineage arose from bacterial and archaeal cells that underwent a symbiotic merger. At the origin of the eukaryote lineage, the bacterial partner contributed genes, metabolic energy, and the building blocks of the endomembrane system. What did the archaeal partner donate that made the eukaryotic experiment a success? The archaeal partner provided the potential for complex information processing. Archaeal histones were crucial in that regard by providing the basic functional unit with which eukaryotes organize DNA into nucleosomes, exert epigenetic control of gene expression, transcribe genes with CCAAT-box promoters, and a manifest cell cycle with condensed chromosomes. While mitochondrial energy lifted energetic constraints on eukaryotic protein production, histone-based chromatin organization paved the path to eukaryotic genome complexity, a critical hurdle en route to the evolution of complex cells.

Eukaryotes from Prokaryotes: Symbiotic Contributions

The origin of eukaryotes remains one of evolution’s more pressing unresolved questions; however, it is beginning to yield some of its secrets. Over the past several decades, perspectives on eukaryogenesis have undergone a significant transformation [1]. Mitochondria are now thought to have been present in the eukaryote common ancestor [2,3], and the broad outlines of eukaryote lineage origin have increasingly come to include the concept of symbiosis [4–7]. Both the host lineage and the mitochondrial endosymbiont contributed to the eukaryote origin.

The mitochondrial contribution to the origin of eukaryotes encompassed energy [8], genes [9–11], and lipids [12]. A hallmark of eukaryotic cells is their endomembrane system, which comprises the endoplasmic reticulum (ER), the nucleus, and vesicular membrane traffic. Although eukaryotes are typically seen as being descendant from archaea [13], they possess bacterial lipids. Prokaryotes synthesize lipids at the plasma membrane, eukaryotes synthesize lipids in the ER and in mitochondria [14]. Even the origin of the eukaryotic endomembrane system, which has been a longstanding puzzle in cell evolution, now connects to mitochondria [14–16]. Mitochondria secrete single-membrane-bounded vesicles in the cytosol called mitochondrial-derived vesicles (MDVs) [14–16]. They are homologous to the outer-membrane vesicles (OMVs) of Gram-negative bacteria [17]. OMVs produced by the mitochondrial endosymbiont at the eukaryote origin are a promising candidate for the biochemical and physical source of the endomembrane system. MDVs produced by the ancestral mitochondrion in the cytosol of an archaeal host generate a membrane topology identical to that of the ER [15]. They also provide an ancestrally outward-directed vectoral membrane flux to the host’s plasma membrane and thus a simple mechanism by which the archaeal lipids of the host could be replaced by the bacterial lipids of the mitochondrial endosymbiont. With energy, genes, and membranes coming from mitochondria (Figure 1), what did the archaeal host contribute to the eukaryotic lineage?
Since their discovery, archaea have been implicated as relatives of the host lineage at the origin of eukaryogenesis and mitochondria\(^{18,19}\). There is now much discussion about the possibility that complex, phagocytosing archaea might be yet discovered in nature, based on metagenomic data\(^{20–22}\). In that view, the contribution of archaea is unequivocal: the archael host brought preformed complexity to the eukaryotic lineage. Indeed, the possibility that some archaea or archael relatives might possess a primitive cytoskeleton has been discussed for decades\(^{23–26}\).
The prokaryotic precursors of actin and tubulin are present in many archaea as well as in bacteria [27,28]. So why did prokaryotes never utilize that starting material to evolve eukaryotic cellular complexity? From the bioenergetic standpoint, the simplest answer is that mitochondrial power was required for the evolution of eukaryotic complexity [8,29,30] and that prokaryotes lack true complexity because they lack mitochondria.

However, some researchers doubt that mitochondria had significance for eukaryote origin, arguing that the presence of mitochondria in the eukaryote common ancestor is pure coincidence, with phagotrophy (eating other cells as a form of obtaining carbon and energy) being the key innovation that allowed eukaryotes to become complex [31,32]. The insurmountable problem with that view, however, is that phagotrophy only provides physiological benefit to cells that already possess mitochondria (reviewed in [33]), which is very likely the reason why no phagocytosing prokaryotes have ever been found.

Doubts of a different nature are coming to bear upon the issue, however. That is, some researchers now doubt that the metagenomic data currently being interpreted as indicating complex archaea really reflect the existence of complex archaea [14,30,33–38]. We know that eukaryotes are complex because we can observe their complexity both under the microscope and in everyday life. The complex archaea about which much is being written [20–22] have yet to be seen. Until direct evidence comes forth for archaea with eukaryotic complexity, it remains possible, if not probable, that the archaean contribution to the symbiotic merger was something other than preformed complexity.

For the sake of reasoning, let us assume for a moment that doubts about the existence of complex archaea [14,30,33–38] are justified. We would then ask the following question. What was the archaean trait that made the archaean contribution to the symbiosis a success? We have known for decades that archaea contributed the information-processing system, or informational genes, to the eukaryotic lineage [9,10,39,40]. The genes in the eukaryotic lineage that have remained most conserved in function relative to archaean counterparts are involved in information processing: the RNA polymerase [4,41], the ribosome [42,43], and aminoacyl tRNA synthetases [44]. But the machinery of information processing itself does not set eukaryotes apart from prokaryotes. The information-processing machineries of eukaryotes and archaea are largely the same. Rather it is the amount of information that eukaryotes maintain, process, and inherit that sets them apart from prokaryotes.

**Eukaryotic DNA Content**

Although the increase in metabolic capacity of eukaryotic cells greatly exceeds that of their prokaryotic progenitors, the increase in DNA content of eukaryotic cells over their prokaryotic progenitors is even greater. A slight historical diversion into DNA content may be helpful in appreciating a critical capacity afforded by an increased amount of DNA. When the study of molecular biology commenced, bacteria and their viruses were the primary topics of investigation. The DNA content of these organisms seemed rational, the more complex the organism the larger the amount of DNA [45]. As eukaryotic cells came under scrutiny, this rational system seemed to break down, by and large eukaryotic cells had a great deal more DNA than expected [46]. Organisms in the same species had identical amounts of DNA, but very similar species often had vastly different amounts of DNA [47,48]. This phenomenon was so prevalent that it became known as the ‘C value paradox’, C value being the haploid amount of DNA associated with a species [49]. Eukaryotic cells have an inordinately large amount of DNA relative to prokaryotic cells, and the differences in DNA content across eukaryotes was not correlated with increased complexity, sophistication, or even gene number [45,50].
The differences between prokaryotic and eukaryotic genome size are accounted for primarily by mobile DNA because it can create new copies of itself, but without a positive contribution to the phenotype of the organism [51]. Mobile DNA is also called ‘selfish DNA’ [52] and makes up a substantial portion of the eukaryotic genome, well over 90% in Homo sapiens [53]. The presence of selfish DNA in the genomes of eukaryotes clearly indicates that the constraint on the amount of DNA in prokaryotic cells is not a limitation in eukaryotic cells, which can exhibit a super abundance of DNA. The ability to accommodate a vast amount of DNA produced by mobile DNA may have led to the origin of introns in the eukaryotic lineage and selective pressures that required the evolution of a nucleus in the eukaryote common ancestor [54]. The necessity of intron excision likely precipitated the decoupling of transcription from translation in the eukaryote ancestor, a major feature distinguishing eukaryotes from their prokaryotic predecessors.

The amount of DNA found in eukaryotes is so vastly greater than in prokaryotes that there is a minimal overlap of the distributions (Figure 2). Eukaryotes have DNA contents with a range spanning over 200,000-fold [48], while the range of expressed genes (proteomes) is much more limited, probably spanning about 50-fold. Early in the development of molecular biology, the replication of DNA was viewed as a major expense in the cell’s energy economy and thus the expansion of DNA in eukaryotes was thought to be very expensive. A more detailed analysis of cellular energy expenditures indicates that the synthesis of proteins is far and away the major consumer of cellular energy, about 75% of the cell’s energy budget, while DNA synthesis is relatively inexpensive, about 3% of the energy budget [8, 55]. The limitation on prokaryotic genome size is related to the ability to manipulate and control the expression of large amounts of DNA, rather than the expense of synthesizing the DNA. Why are eukaryotic cells capable of manipulating vastly larger amounts of DNA than their prokaryotic progenitors? The answer would appear to lie in the organization of the DNA into nucleosomes.

Archaeal Histones

It would be an oversimplification to view the DNA in a cell independently of proteins bound to the DNA. All organisms have proteins bound to their DNA in order to prevent collapse of DNA at high

Figure 2. The Ranges of Genome Sizes for Various Groups. The ranges of genome sizes are displayed on a logarithmic scale. The vertical dimension indicates, but is not strictly proportional to, the number of taxa. The specific shape of the distributions is not the main point here, the range is accurately portrayed. (A) The comparative ranges of genome sizes of archaea, bacteria, and eukarya. (B) The ranges of genome sizes for various groups of eukaryotes. The range of genome sizes for plants is somewhat greater than for animals, but this difference is not evident on a logarithmic scale (www.genomesize.com) [50].
In bacteria there are a number of DNA-binding proteins that fulfill this role [57]. Archaea in particular, have several proteins that fulfill this role, including proteins with the histone-fold, which are widespread among the euryarchaea and nanoarchaea [58,59]. In these archaea, DNA binds to histone-fold proteins, which form homodimers and appear to produce extended fibers, as recent crystal structures for archaeal histones reveal [60]. The diversity of archaeal histones is not fully charted [61] but neither eukaryotic type nucleosomes nor eukaryotic type histone modifications have been found in archaea so far. Nuclear eukaryotic DNA, in contrast, is organized into nucleosomes, slightly less than 150 base pairs of DNA wrapped around an octamer histone core consisting of two H2A/H2B dimers and a H3/H4 tetramer [62]. The filament nature of archaeal histone polymerization contrasts to the ‘beads on a string’ nature of eukaryotic histones (Figure 3).

Histones make up a major portion of a nucleosome, ~ 55% of the mass, and their special orientation creates the nucleosome core. The nucleosomes are connected to each other by short ‘linker’ segments of DNA and comprise the primary structure of chromatin [62,63]. Organization of DNA into nucleosomes leads to a compaction of 30- to 40-fold [64].

Recent advances in chromatin imaging reveal 5 nm and 24–30 nm fibers of chromatin organization in living cells [65], underscoring the central role of nucleosomes in the primary organization of eukaryotic DNA. The nucleosome organization appears to permit a high degree of flexibility and removes a major constraint on manipulating large amounts of DNA [65,66]. Compaction and flexibility permit large amounts of DNA to be manipulated and the cellular genome size is free to expand, almost without limit [64] (Figure 3).

**Histones Control Gene Expression**

A dramatic increase in the DNA content of cells requires not only improved mechanisms for replication and maintenance but also enhanced control of gene expression [67,68]. The evolution and development of a nucleosome-based chromatin structure in eukaryotes addresses both of these requirements. Initially, nucleosomes were viewed as static structural elements facilitating the manipulation of the DNA; the role of histones in organizing eukaryotic DNA into nucleosomes provides a basis for enhanced control of gene expression. Having a greatly expanded DNA content dramatically increases the necessity for control of genetic expression in eukaryotes well beyond the prokaryotic mechanisms for regulation of gene expression. Within the nucleosome structure, particularly in the polypeptide extensions – the N terminal and C terminal ‘tails’ – beyond the core histone-fold protein, histone modification provides a powerful element for the control of eukaryotic gene expression [67–72].

Central to the role of histones as a mechanism for the control of gene expression are histone acetyltransferases (HATs) [70,71]. Histone modification plays the central role in the massively expanded epigenetic control of gene expression in eukaryotes relative to prokaryotes. The ‘histone barcode’ concept has it that specific histone modification, or combinations thereof, can affect distinct downstream cellular events by altering the structure of chromatin or by generating a binding platform for effector proteins [68,70,73]. The role of covalent modification of histones in the regulation of gene expression has become a central theme in eukaryote gene regulation [74].

The central portion of the eukaryotic histones, dominated by the histone-fold, is the core of the nucleosome; however, extensions of this core – particularly at the amino terminus – provide the substrate for modifications that control gene expression [75]. Post-translational acetylation of histone termini, particularly of lysine residues, as well as phosphorylation and methylation, dramatically affects their charge and function. Histone modification affords the elaborate control of gene
Figure 3. Histones in Archaea and Eukaryotes. (A) The crystal structure of DNA bound to archaeal histones was recently determined to 4 Å for Methanothermus fervidus [60]. The schematic representation of the structure shown here indicates DNA wrapped four times around the central protein superhelix of nine archaeal histone dimers. The archaeal structure forms a fiber. (B) Eukaryotic histones form the familiar nucleosome structure that is the basic structural and functional unit of eukaryotic chromatin. The red surfaces indicate the N terminal (top) and C terminal extensions of eukaryotic histones. These extensions contain the main sites of histone modification that modulate chromatin condensation/decondensation and that govern gene expression. These functional extensions are crucial to eukaryotic nucleosome organization, but are lacking in archaeal histones [60,61]. In contrast to archaea, eukaryotes have a bona fide cyclin-dependent cell cycle in which metaphase chromosomes condense for chromosome segregation and cell division (lower panel). Regulation of the cell cycle and decisions to exit into G0, arrest, or autophagy involve histone modifications and are governed by nutrient sensing (Box 1). Note that, in eukaryotes with a syncytial (coenocytic) lifestyle, chromosome division and cell division are not coupled [54], in contrast to archaea and bacteria. AMPK, AMP-activated protein kinase; CDKs, cyclin-dependent kinases; TOR, target of rapamycin.
expression required by the vastly expanded cellular DNA complement of eukaryotes (Figure 3, Box 1). This epigenetic control can persist over many cell generations, allowing the differentiation of stable gene-expression patterns in various cell types [76]. Such gene control is essential in multicellular organisms in which different portions of the genome are expressed in different tissues. This type of control permits ‘silencing’ of large portions of the genome, making the expansion of eukaryotic genomes by ‘selfish DNA’ feasible [77]. These developmental patterns of gene expression are fundamental to the evolution of multicellular organisms, and a hallmark of eukaryotes.

**Energy and Information: A Good Recipe for Complexity**

The origin of eukaryotic histones traces to archaea. Histones are required for genomic complexity and the eukaryotic cell cycle, both of which are integral to eukaryotic cellular complexity (Figure 3). Dennis Searcy was investigating histone-like proteins in the archaea even before the archaea were recognized as a domain [78]. John Reeve and colleagues found bona fide histones in methanogens (euryarchaeotes) [58–60] and other archaea [79], emphasizing over decades the evolutionary significance of archaeal histones with regard to the eukaryote origin. It is possible that the role of histones in the process of eukaryote origin has been underappreciated. Perhaps histones were key.

The mitochondrion contributed genes, energy, and membranes to the origin of the eukaryote lineage. What did the archaean partner contribute? That is, what archaean trait allows eukaryotic cells...
to accumulate, inherit, and express the large complex genomes that underpin eukaryote complexity? The answer might be simple: with the energetic configuration that mitochondria bestowed upon eukaryotes, the histone-dependent organization of DNA into chromatin allowed eukaryotes to take the first hurdle towards true complexity by regulating and managing large amounts of DNA [66–68]. The archaeal partner brings proteins with the histone-fold, from which eukaryotic histones, nucleosomes, and the substrate for epigenetic control of gene expression evolved [58–60]. The multiple-origin nature of archaeal DNA replication is also essential in the propagation of large amounts of DNA, which was necessary to replicate chromosomes of eukaryotic size [80,81].

Unlike the metabolic contribution of the bacterial partner, which is fully developed at the time of the symbiotic merger, the archaeal partner’s contribution is only the ‘potential’ for manipulating large amounts of DNA in the form of proteins with the histone-fold. These proteins evolve into the eukaryotic histones, which are at the heart of the eukaryotic information management [70, 75]. Stated another way, the union of a bacterium and an archaeon set the stage for the development of the eukaryotic lineage whereby the archaeal host compartment underwent a great deal more evolutionary modification than the mitochondrion did – the archaeal host was transformed from within, via gene transfer from the endosymbiont.

**Histones, Decisions, and Signaling in the Cell Cycle**

Energy powers the evolutionary process because energy powers life – without energy, no life, much less evolution. There are very interesting and important connections between histone modification and the energetic status of the eukaryotic cell. The essence of histone acetylation is that acetylation preferentially takes place at lysine residues, reducing their positive charge and hence their affinity to DNA, opening up chromatin for transcription. Mitochondrial proteins tend to undergo nonenzymatic (spontaneous) acylation by acetyl-CoA, an energy-rich thioester intermediate of core energy metabolism [82]. Might such a simple mechanism have been the seed of eukaryotic gene regulation through histone acetylation (see Outstanding Questions)? At eukaryote origin, sensing the nutritional status of the cell in the thioester currency of acetyl-CoA levels (histone acetylation) could have transferred information about the physiological state of the cell directly to chromosome activity in a meaningful manner. Acetylation of histones H3 and H4 promotes chromatin decondensation [83], removal of histone acetylation signals low acetyl-CoA levels (low nutrients) and leads to condensed inactive chromosomes. In an unrefined initial process, high nutrient levels would evoke decondensed active chromosomes, while low nutrient levels would lead to condensed, transcriptionally inactive chromatin.

Today, histone modification is highly regulated and tightly linked to (i) the energetic status of the cell [84], (ii) chromatin condensation [85], and (iii) decisions about the eukaryotic cell cycle [86]. Eukaryotes sense the energetic state of the cell with the help of the sensor proteins AMP-activated protein kinase (AMPK) and target of rapamycin (TOR) [87–89], and they sense the nutritional status of the cell (carbon, energy, and amino acid availability) via TOR and GCN2 [90]. By dry weight, nonphotosynthetic cells are about 50% protein and 10% nitrogen [91]. The three main energy and nutrient signaling pathways of eukaryotes – AMPK, TOR, and GCN2 – are conserved across eukaryotic groups [90]. All three master regulators transmit signals to chromatin via histone modification [88,89] (Box 1). By sensing the nutrient state of the cell they connect mitochondrial energy supply – and mitochondrial quality-control via autophagy [92,93] – with the eukaryotic cell cycle. The cell cycle is the basic and indivisible quantum unit of eukaryotic cell survival [94] (Figure 3). Of course, many eukaryotes have reduced the bioenergetic functions of their mitochondria or have transferred ATP synthesis to the cytosol [5,32], but in evolutionary terms the cell cycle was built upon archaeal histones in a cell that had mitochondria.
Even the CCAAT Box Traces Back to Histones

Archaeal histones were even instrumental in the early evolution of basic eukaryotic transcriptional regulation. How so? The CCAAT-box is found in 30% of eukaryotic promoters [95]. It binds the CCAAT-box binding complex, CBC, which promotes the recruitment of RNA polymerase II to eukaryotic DNA for transcription. Because it has undergone many duplications and functional specialization for regulation of various genes, CBC has several names in the literature, the most common being nuclear factor Y, or NF-Y [96]. The core subunits of CBC are evolutionarily derived from eukaryotic and archaeal histones [97,98]. The structural basis for CBC binding was resolved for the protein from Aspergillus nidulans [99]. Aspergillus CBC has the crystal structure of a histone H2A/H2B heterodimer [99]. It bends DNA, it can interact with the histone pair H3/H4, and it possesses a specific structural motif, the Nα helix of HapB, which binds directly to the CCAAT box and is strictly conserved throughout all eukaryotes [99]. Though lacking in archaea, the CCAAT box is ubiquitous among eukaryotic genomes, hence it was present in the eukaryotic ancestor. Its binding protein, CBC, is derived from archaean histones, it is modified in the same manner as histones are [100], it is recruited to promoters specific to the cell cycle [101], it has retained the DNA-bending structure of histones [99], and it helps to activate about one in three eukaryotic genes [95]. Clearly, archaean histones were important for the establishment of gene regulation that led to chromatin condensation, the establishment of a cell cycle (Figure 3), transcriptional response to nutrient availability, and complexity in eukaryotes that involves CCAAT box-dependent gene regulation. Taken together, that is a substantial contribution from archaean histones.

Concluding Remarks

At eukaryotic origin, the bacterial partner contributed genes, energy, and membranes to the union, resulting in the eukaryotic lineage, while the host contributed information-processing ability.
A case can be made that the major contribution of the archaeal partner to the union was the provision of proteins with the histone-fold that evolved into bona fide eukaryotic histones, which became the core of both eukaryotic nucleosomes [59] and the cell cycle (Figure 3).

The compaction of eukaryotic DNA into nucleosomes allowed the nascent eukaryotic lineage to store and manage vastly more information than was possible in prokaryotes, and thus to express a dramatically larger number of different proteins than were expressed in prokaryotes. At the same time, the multiple replication origins germane to archaeal DNA [80,81] allowed eukaryotic chromosomes to expand to the size range now characteristic of eukaryotic cells. Through acetylation and methylation, histones also provided a simple mechanism of information transfer about the nutrient and energetic state of the nascent eukaryotic cell into chromatin condensation and activity.

With the combination of increased energy available per gene, provided by multiple mitochondria, and virtually unlimited genetic capacity, provided by eukaryotic chromatin structure, the eukaryotic lineage embarked on the exploration of evolutionary innovations that came to include meiosis and the cell cycle, multicellular organisms with different cell types, and complex cellular interactions within an organism. Increased genetic capacity allowed for the elucidation of complex developmental programs, many underpinned by CCAAT box-dependent gene regulation, as well as innovation in protein structure. The characteristics we associate with the eukaryotic lineage are the direct result of the increased energy provided by the bacterial partner (mitochondria) coupled with an increased capacity to handle DNA afforded by organization of eukaryotic DNA into nucleosomes contributed by the histones from the archaeal partner. This combination of traits endowed the first eukaryote with evolutionary options, particularly in the realm of cell structures and multicellularity, unavailable to any prokaryotic lineage. The cell cycle is the cornerstone of eukaryotic cell division and growth. At its heart reside histones, both in a regulatory and a structural role, giving them a special place in the origin and evolution of both the eukaryotic cell and the eukaryotic cell cycle.

**Acknowledgments**

W.F.M. thanks the European Research Council (Grant 666053) and the Volkswagenstiftung, Germany (Grant 93 046) for financial support. We thank Sriram Garg, Sven Gould, and Michael Grunstein for discussions and comments on the paper, and Verena Zimorski for help in preparing the manuscript.

**References**