

# The origin of the eukaryotic cell

Radhey S. Gupta and G. Brian Golding

Molecular sequence data are beginning to provide important insights into the evolutionary origin of eukaryotic cells. Global phylogenies of numerous protein sequences indicate that the eukaryotic cell nucleus is a chimera, which has received major contributions from both a Gram-negative eubacterium and an archaeobacterium. Recent studies also indicate that the formation of the nuclear envelope and the endoplasmic reticulum was accompanied by duplication of genes for the molecular chaperone proteins (e.g. hsp70, hsp90), which facilitate protein transport across membranes. Based on these observations, it is suggested that the ancestral eukaryotic cell arose by a unique endosymbiotic event involving engulfment of an eocyte archaeobacterium by a Gram-negative eubacterial host.

**THE ORIGIN OF** the eukaryotic cell constitutes one of the major unsolved problems in biological sciences. Eukaryotes, which comprise all true multicellular organisms, as well as a broad range of single-cell organisms, are distinguished from prokaryotes by the presence of several internal structures. These include the nucleus, cytomembrane, cytoskeleton and a variety of organelles. However, the hallmark feature of all eukaryotes is the presence of a nucleus (i.e. separation of DNA and cytoplasm by a nuclear envelope) and any organism lacking this property is by definition a prokaryote<sup>1,2</sup>. The structural complexity of the eukaryotic cells, in conjunction with the fossil evidence, which suggests that the prokaryotes might have predated eukaryotes by as much as 1.0–2.0 billion years<sup>3</sup>, has led to the notion that eukaryotic cells have evolved from more simple prokaryotic organisms with which they share numerous common (or related) molecules<sup>1,2,4–6</sup>. Indeed, studies done during the past 10–15 years have established that a number of eukaryotic cell organelles (namely mitochondria and chloroplasts) bear a close evolutionary relationship to specific groups of prokaryotes (namely  $\alpha$ -proteobacteria and cyanobacteria, respectively)<sup>7</sup>. These organelles are believed to have evolved from prokaryotic ancestors through endosymbiotic capture by eukaryotic hosts lacking such

structures<sup>7,8</sup>. However, the origin of the nucleus, which presumably marked the beginning of the eukaryotic cell, remains an unresolved issue at the present time.

To explain the origin of the eukaryotic cells a number of theories have been put forward (see Box 1). Several of these theories are conjectural in nature, whereas others are based on limited

experimental evidence. However, as sequence data has become more plentiful, the reconstruction of a more accurate picture of the evolutionary history of the eukaryotic cells has been permitted.

## Current model for the origin of the eukaryotic cell

Woese and colleagues were among the first to make use of sequence data to understand molecular systematics<sup>10</sup>. They promoted the use of ribosomal RNA sequences to understand the evolutionary relationship between all life forms. Among their many discoveries, an important finding was the recognition that, based on ribosomal RNA (rRNA) sequences, prokaryotes formed two distinct phylogenetic groups, originally termed archaeobacteria and eubacteria. In terms of genetic distance, these two groups were as different from each other as either was from the eukaryotes<sup>10</sup>. Subsequent work on a few protein sequences that were examined (namely EF-Tu/1a, EF-G/2, RNA polymerase and F- and V-type ATPases) also supported the phylogenetic distinctness of archaeobacterial and eubacterial groups<sup>13,14,19</sup>. Initially, only an unrooted tree was possible with these three groups of organisms (namely archaeobacteria, eubacteria and eukaryotes)<sup>10</sup>. The

### Box 1. Earlier theories for the origin of eukaryotic cells

Symbiotic origin of eukaryotic cells proposed.	Mereschkowsky <sup>9</sup>
Endosymbiotic origin of mitochondria, chloroplasts and eukaryotic motility apparatus (namely cytoskeleton) from specific groups of bacteria proposed. Endosymbiotic origin of nucleus not favored.	Margulis <sup>1,8</sup>
Independent origin of archaeobacteria, eubacteria and eukaryotes from a universal ancestor (progenote) proposed.	Woese <sup>4,10</sup>
The suggestion that the nucleus could have evolved by engulfment of an archaeobacterial species was made.	Lake <i>et al.</i> <sup>11,12</sup>
Evolution of eukaryotic cells from a transient intermediate between archaeobacteria and Gram-positive bacteria proposed. Suggestion concerning origin of various eukaryotic cell structural features made. Opposed to endosymbiotic origin of nucleus.	Cavalier-Smith <sup>5</sup>
Direct evolution of eukaryotic nuclear genome from an archaeobacterial ancestor, which is the currently favored view, proposed.	Iwabe <i>et al.</i> <sup>13</sup> Gogarten <i>et al.</i> <sup>14</sup> , Woese <i>et al.</i> <sup>15</sup>
The eocyte group of archaeobacteria was indicated as the closest relative of eukaryotes.	Rivera and Lake <sup>16</sup>
Evolution of eukaryotic cell by engulfment of a prokaryotic organism (archaeobacterium) <sup>18</sup> by a proto-eukaryotic lineage that contained RNA-based metabolism proposed.	Hartman <sup>17</sup> , Sogin <sup>18</sup>
The possibility that the eukaryotic nuclear genome could be a chimera formed by primary fusion of an archaeobacterium and a bacteria suggested.	Zillig <sup>6</sup>

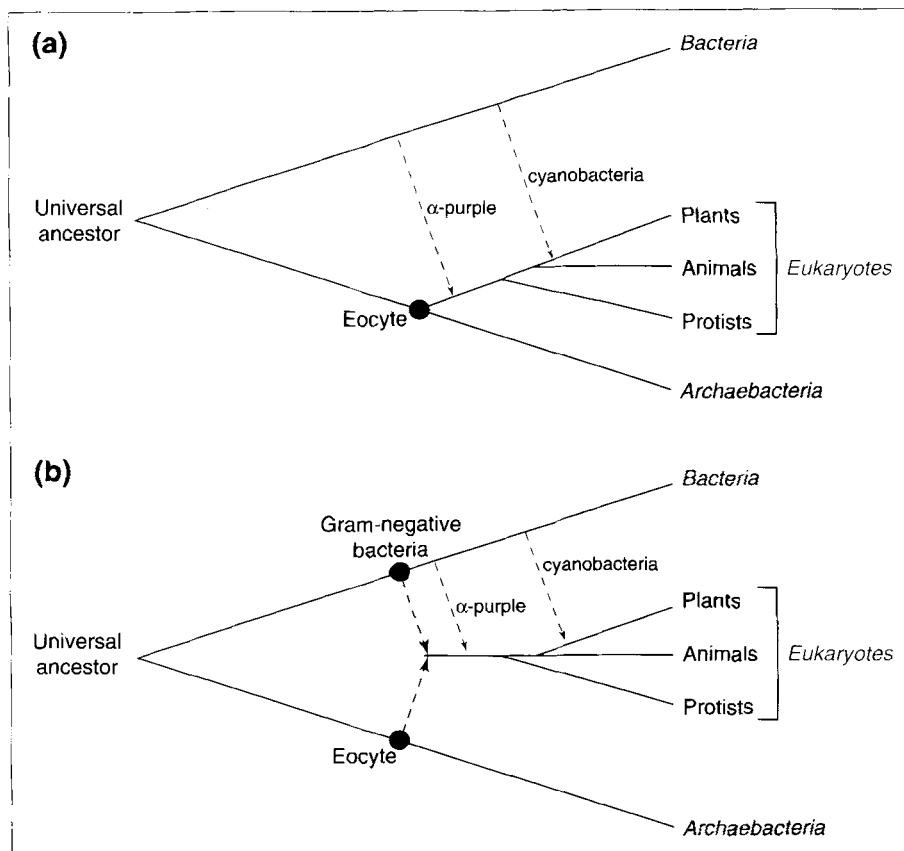
R. S. Gupta is at the Department of Biochemistry; and G. B. Golding is at the Department of Biology, McMaster University, Hamilton, Ontario, Canada L8N 3Z5.

root of this tree was determined using duplicated gene sequences such as those of EF-Tu/G and ATPase V/F. A critical assumption here is that the gene duplication event must have occurred before the divergence of the three groups of organisms. If this assumption is correct then one set of duplicated sequences could be used to root the other set. Using this approach, Iwabe *et al.*<sup>13</sup> and Gogarten *et al.*<sup>14</sup> placed the root of the tree between archaeobacteria and eubacteria. In both these studies, the eukaryotes branched off from the lineage leading to archaeobacteria (Fig. 1). In 1990, Woese *et al.*<sup>15</sup> adopted this rooting and elevated the three groups from kingdoms to 'domains of life' and renamed them Archaea, Bacteria and Eucarya<sup>15</sup>. The above rooting (referred to as the archaeobacterial model in Fig. 1): (1) implies that the eukaryotic organisms are the direct descendants of an archaeobacterial ancestor; (2) has become widely accepted; and (3) represents the prevalent view regarding the origin of the eukaryotic nuclear genome.

#### Inadequacy of the present model to explain global protein phylogenies

The above view of the origin of the eukaryotic nuclear genome is based on a limited number of gene sequences. However, as sequence data on more proteins have become available, this model has not always been supported, and sequence features and phylogenies based on several proteins give a very different picture of the history of these three groups of organisms<sup>20–27</sup>. Protein sequences have several advantages over rRNA for phylogenetic studies. In proteins, each character has 20 states rather than only four different states for nucleotides. This provides more information per site and reduces the likelihood of convergent or parallel substitutions, or mere chance alignment of certain sequence regions. In addition, organism-specific differences in G/C content can cause biased phylogenetic reconstructions. However, owing to degeneracy of the genetic code, protein sequences are less affected by differences in G/C content than are nucleotide sequences<sup>28</sup>. Lastly, by contrast to the rRNA, where sequence alignment is dependent on the secondary structure of the molecule, no such assumption is necessary for the alignment of protein sequences.

Several recent studies by independent laboratories have found proteins that do not support the traditional



**Figure 1**

Evolutionary origin of eukaryotic cells. **(a)** The traditional archaeobacterial model and **(b)** the newly proposed chimeric model. According to the archaeobacterial model, the eukaryotic nuclear genome is directly derived from an archaeobacterial ancestor by means of normal evolutionary mechanisms. By contrast, the chimeric model proposes that the ancestral eukaryotic cell arose as the result of a unique fusion event between a Gram-negative eubacterium and an 'eocyte' archaeobacterium, and that both these parents have made major contributions to its nuclear genome. Both models indicate that mitochondria and chloroplasts originated as a result of later endosymbiotic events from specific groups of eubacteria.

branching pattern shown in Fig. 1 (Refs 20–27). The 70 kDa heat shock proteins (hsp70/dnaK), which constitute the most conserved proteins found in all species, provide the clearest example of such phylogenies<sup>23–25</sup>. In eukaryotic cells, unique hsp70 homologs are present in the cytosol, endoplasmic reticulum (ER), mitochondria and chloroplasts<sup>24,29,30</sup>. However, the various forms of hsp70 are readily distinguished from each other based on their characteristic signature sequences (Fig. 2). From a number of such signature sequences, the mitochondrial and chloroplast forms are seen to be clearly prokaryotic-like and are distinct from eukaryotic hsp70s from the ER and cytosol<sup>24,25,30</sup>. The cytosolic and the ER-resident forms (BiP) of hsp70s form paralogous gene families that arose by gene duplication very early in the evolution of eukaryotic cells<sup>25,30</sup>. However, the most distinctive feature of the hsp70 family of proteins is a relatively conserved insert of 23–27 amino acids, which is present in

the same position in all hsp70s from Gram-negative eubacteria\* and eukaryotic species, but absent in all of the homologs from archaea and Gram-positive eubacteria (Fig. 2; Refs 23–25, 30). Such evidence points to a specific relationship between Gram-negative bacteria\* and eucarya on one hand, and archaea and Gram-positive bacteria on the other. This inference is strongly supported by the detailed phylogeny based on hsp70 sequences<sup>23–25</sup>.

The hsp70 protein family is by no means unique in its indication of a different branching order. Trees based on glutamine synthetase (*glnA*)<sup>21,26</sup> and glutamate dehydrogenase (*gdh*)<sup>22</sup> also suggest similar relationships. All of these proteins are well-conserved molecules that retain excellent homology among the major domains of life. Except for the branching of eucarya with the

\*Throughout the text, the terms eubacteria and bacteria are used interchangeably to refer to prokaryotes that are not archaeobacteria.

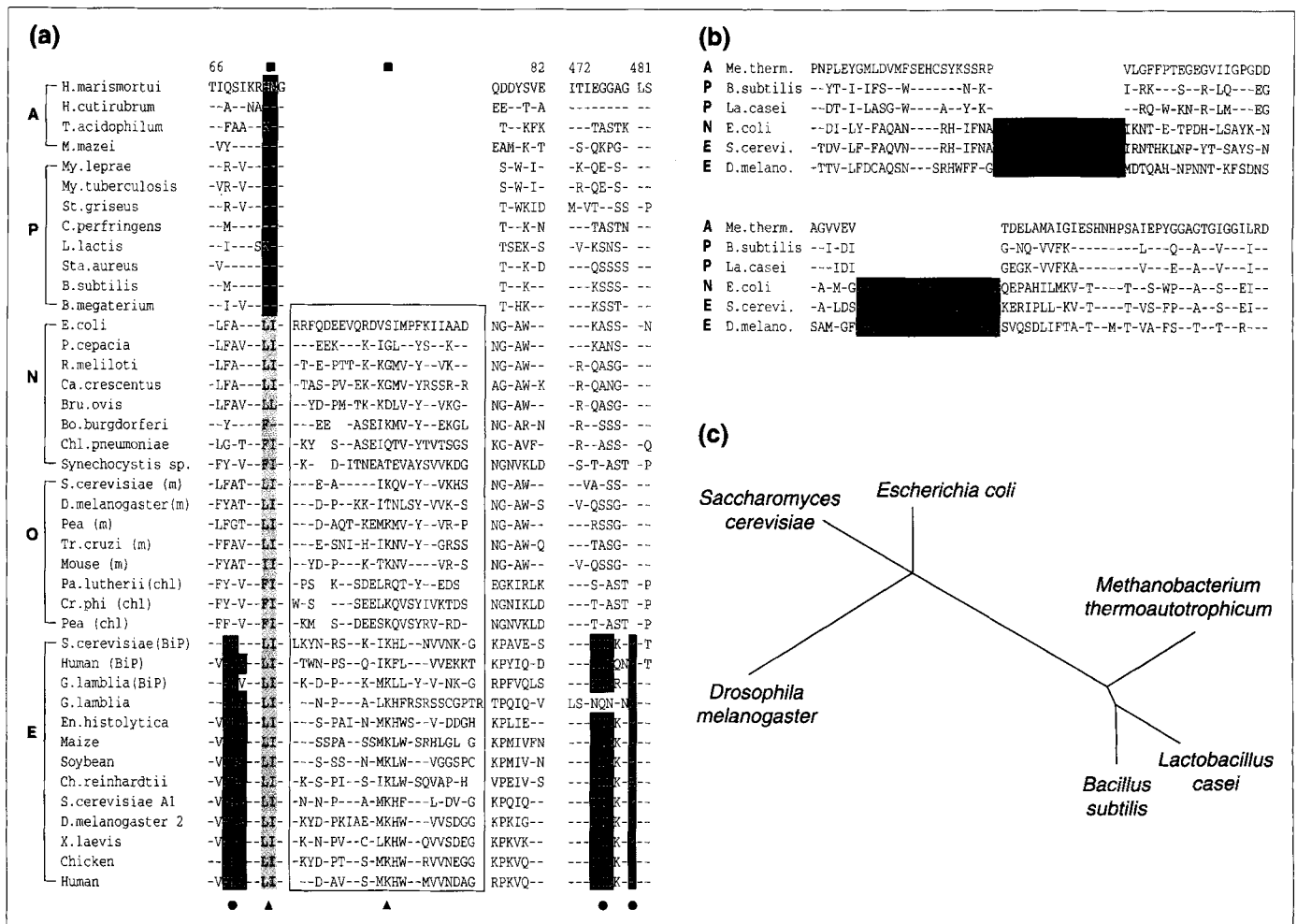


Figure 2

**(a)** Sequence signatures in the hsp70 protein family showing the relatedness of eukaryotic cytosolic homologs to the Gram-negative eubacteria. The letters A, E, N, P and O refer to sequences from archaeobacteria (A), eukaryotes (cytosolic and endoplasmic reticulum; E), Gram-negative bacteria (N), Gram-positive bacteria (P) and eukaryotic organellar sequences [namely mitochondria (m) and chloroplasts (chl); O], respectively. The numbers on the top refer to the position in the *Halobacterium marismortui* hsp70 sequence. The dashes (–) indicate identity with the amino acid in the top line. Symbols above or below the shaded regions refer to different types of sequence signatures: ■, sequence signatures common to only archaeobacteria and Gram-positive bacteria; ●, sequence features present only in eukaryotic cytosolic and ER (BiP) homologs; and ▲, shared sequence features between Gram-negative eubacteria and eukaryotic homologs. Not all signature sequences of the above kinds are shown. The boxed region shows the large insert present in only Gram-negative eubacteria and eukaryotic homologs. Sequences from only representative species are shown, however, homologs from all other species examined also contain the indicated signature sequences. Abbreviations used in the species names are as follows: B, *Bacillus*; Bo, *Borellia*; Bru, *Brucella*; C, *Clostridium*; Ca, *Caulobacter*; Ch, *Chlamydomonas*; Chl, *Chlamydia*; Cr, *Cryptomonas*; D, *Drosophila*; E, *Escherichia*; En, *Entamoeba*; G, *Giardia*; H, *Halobacterium*; L, *Lactococcus*; M, *Methanobacteria*; My, *Mycobacterium*; P, *Pseudomonas*; Pa, *Pavlova*; R, *Rhizobium*; S, *Saccharomyces*; St, *Streptomyces*; Sta, *Staphylococcus*; T, *Thermoplasma*; Tr, *Trypanosoma*; X, *Xenopus*. The notations (mit), (chl) and (BiP) in parenthesis denote mitochondrial, chloroplast and ER-resident forms of hsp70, respectively. The database accession numbers of these sequences are described in earlier work<sup>24</sup>. **(b)** Alignment of sequences of phosphoribosylformylglycinamide synthase (FGARAT) from various species. The species names and GenBank accession numbers of the sequences are as follows: *Escherichia coli* (M19501), *Saccharomyces cerevisiae* (U14566), *Drosophila melanogaster* (U00683), *Methanobacterium thermoautotrophicum* (Me. therm.) (M59245), *Bacillus subtilis* (J02732), *Lactobacillus (La.) casei* (M85265). The shaded regions show the shared insertions in Gram-negative bacteria and eukaryotes. **(c)** A distance tree based on the FGARAT sequences.

Gram-negative bacteria, the phylogenetic trees based on these protein sequences are otherwise conventional and indicate relationships between organisms that would have been expected (in the sense that plants cluster together, vertebrates cluster together and mitochondria and chloroplast homologs are specifically related to  $\alpha$ -proteobacteria and cyanobacteria, respectively)<sup>22–26,30,31</sup>.

Another example of a protein sequence that demonstrates a close relationship

between the eucarya and the Gram-negative bacteria is the enzyme phosphoribosylformylglycinamide synthase (FGARAT). From a partial alignment of FGARAT sequences presented in Fig. 2, it is readily seen that a number of conserved indels (i.e. either insertions or deletions) are uniquely shared by Gram-negative bacteria and eukaryotic homologs on one hand, and Gram-positive bacteria and archaeobacteria on the other. A specific relationship between

these groups is also strongly supported by phylogenetic analysis of the sequence data (Fig. 2).

To investigate this problem further, we have collected all proteins in the databases for which sequences were available at the time from Gram-positive bacteria (P), Gram-negative bacteria (N), archaea (A) and eucarya (E)<sup>27</sup>. A total of 24 proteins were found that fulfilled these criteria. After alignment of the sequences, they were evaluated to

determine which of the three possible monophyletic arrangements they support [namely the archaeobacterial tree (P,N)(A,E), the tree supported by hsp70, *gdh* and *glnA* (P,A)(N,E), or the third possible topology (P,E)(N,A)]<sup>27</sup>. The analysis was performed using both the maximum likelihood and the maximum parsimony methods<sup>27</sup>. The results are summarized in Table I. In addition to hsp70, glutamine synthetase, glutamate dehydrogenase and FGARAT proteins, a number of other proteins significantly preferred a phylogeny that clusters the eucarya with Gram-negative bacteria and the archaea with Gram-positive bacteria. In fact, the protein-based trees are almost evenly split between this topology and the presently favored topology that groups archaea with eucarya. The proteins for which no significant preference was observed do not support or refute either phylogeny. However, it is significant to note that none of the proteins supported the third phylogeny (P,E)(N,A), which groups the Gram-positive bacteria with the eucarya. This result provides strong evidence that the methods are not generating false significance scores, because the sequences are too distantly related or no longer contain traces of their ancestry. If this were the case, an equal number of proteins would be expected to support each of the three alternative phylogenies.

**Orthology–paralogy and horizontal gene transfer**

How can the two different kinds of phylogenies inferred from these protein sequences be reconciled? Because some of the proteins suggest the same topology as the rRNA sequences, should we prefer that particular inference and ignore the others? It is important to exclude two possibilities that could lead to incorrect phylogenies before considering other models to explain these results. First, all sequences used to reconstruct phylogenies must be orthologous. If they are not then an incorrect phylogeny will be inferred. However, for many of the proteins that fail to cluster the archaea and eucarya (e.g. hsp70, glutamine synthetase and glutamate dehydrogenase), multiple forms are well known and have been incorporated into their phylogenies<sup>21–26</sup>. In fact, the distinct clustering of various hsp70 isoforms into their phylogenetic trees gives further confidence that the phylogeny inferred from these proteins is reliable. For example, the mitochondrial

forms of hsp70 consistently tree within the  $\alpha$ -proteobacteria, and the chloroplast forms within the cyanobacteria<sup>23,24,30,31</sup>. The cytosolic and ER forms of hsp70, which are found in all eukaryotic species, including the amitochondrial protist *Giardia lamblia*<sup>25</sup>, contain distinctive signature sequences that are not present in any organellar homologs (see Fig. 2; Refs 24, 25). Thus, it is highly unlikely that the unusual branching pattern for this protein might be a consequence of comparison between non-orthologous sequences.

Second, it has been suggested that some of the unusual topologies could be because of horizontal gene transfers<sup>20–22</sup>, but there are several lines of evidence that show this is not the cause. For the hsp70 sequence data, this possibility has been ruled out<sup>24,25</sup>. Furthermore, if horizontal transfer is invoked to explain the unusual phylogeny of approximately half of all proteins, then why must it occur only between eucarya and Gram-negative bacteria on the one hand, and archaea and Gram-positive bacteria on the other? All of the gene/protein sequences give: (1) a similar relationship between various eukaryotic species; and (2) they also indicate a similar origin of mitochondria and chloroplasts from specific groups of bacteria, thus arguing strongly against extensive gene transfers.

**Chimeric model for the origin of the eukaryotic nuclear genome**

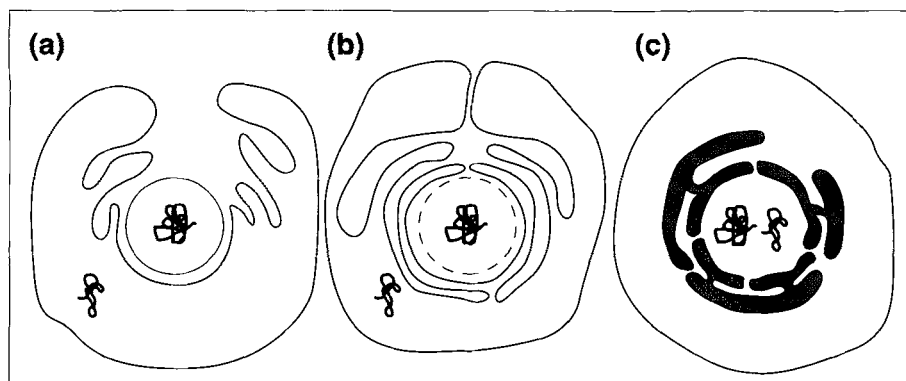
If the observed results are not anomalous, then to explain them it is necessary to postulate that the eukaryotic nuclear genome received major contributions from both an archaeobacterium as well as a Gram-negative eubacterium. Or in other words, it is a chimera formed by the fusion of these groups of parental organisms. Based on Rivera and Lake's results with the EF-1a and EF-2 sequences<sup>16</sup>, the archaeobacterial partner is likely to be a thermoacidophilic 'eocyte' group of archaeobacterium. The nature of the Gram-negative fusion partner is not clear at present, but in the hsp70 phylogeny, a closer relationship of eukaryotic cytoplasmic homologs to the cyanobacterial species is often observed<sup>23,24</sup>. Following the fusion event, the two sets of genes/proteins that were present in the resulting pro-eukaryote cell had then to be resolved into a single functioning cell. By choosing a mixture or set of genes from each of the two parental organisms, a eukaryotic cell evolved that inherited

Table I. Global phylogeny of protein sequences <sup>a</sup>
<b>Eukaryote–Gram-negative clades</b> Aspartate aminotransferase Hsp70 Ferredoxin Glutamate dehydrogenase Glutamine synthetase Phosphoribosylformylglycinamide synthase Pyrroline-5-carboxylate reductase
<b>Eukaryote–Archaeobacterial clades</b> Arginino-succinate synthase ATP synthase $\alpha$ -chain ATP synthase $\beta$ -chain Dihydrofolate reductase Elongation factor 1 Elongation factor 2 Indole-3-glycerol phosphate synthase Phosphoglycerate kinase DNA-directed RNA polymerase
<b>Equivocal clades</b> $\beta$ -galactosidase Carbamoyl phosphate synthase Deoxyribopyrimidine photolyase Glyceraldehyde 3-phosphate dehydrogenase Superoxide dismutase Anthranilate synthase components I and II Anthranilate phosphoribosyltransferase N-(5'-phosphoribosyl) anthranilate isomerase
<b>Eukaryote–Gram-positive clade</b> None

<sup>a</sup>A summary of the phylogenetic analysis based on different protein sequences<sup>27</sup>. For the proteins listed under the first two categories above, the observed phylogenetic relationships in all cases is statistically significant. The proteins listed under the equivocal clades did not yield statistically significant relationships<sup>27</sup>.

the different characteristics of the two parents. This in turn explains the many cellular and biochemical features that are shared by eukaryotes with the archaea on the one hand and with the eubacteria on the other<sup>4–6</sup>.

During the evolution of the eukaryotic genome from the chimeric ancestor, it is expected that the genes/proteins that were already functioning as a unit within cells (e.g. translation and perhaps also transcription apparatus) were taken as a whole from one of the parents. Indeed, the sequence data on various translational/transcriptional related genes (e.g. rRNA, ribosomal proteins, EF-G, EF-Tu, RNA polymerase, TATA-element-binding protein, etc.)<sup>32</sup> suggest that they originated from the archaeobacterial ancestor. In view of this, it is probable that the translational/transcriptional-related gene sequences do not provide independent estimates of the eukaryotic phylogeny. Hence, the



**Figure 3**

Origin of the eukaryotic cell nucleus and endomembrane system by the chimeric model. **(a)** The key event in the origin of the eukaryotic cell is postulated to be the engulfment of an 'eocyte' archaeobacterium by a Gram-negative eubacterium that presumably lacked a cell wall. **(b)** As the membrane of the host surrounded the guest species, its own membrane (containing ether-linked lipids) became redundant and was lost. **(c)** Eventual separation of the internalized membrane from the plasma membrane led to the formation of the nuclear envelope and the endoplasmic reticulum. The formation of these new compartments was preceded or accompanied by duplication of the genes for the chaperone proteins (e.g. hsp70, hsp90, etc.). The transfer of the host genome to the newly formed nucleus, and an assortment of the genes from the two parents, led to the formation of the ancestral eukaryotic cell.

interpretation of the rooting of the universal tree based on such sequences (namely EF-Tu/G) is questionable<sup>24,27,33</sup>.

#### Origin of nucleus and endoplasmic reticulum

Important insight into the origin of the eukaryotic cell nucleus is provided by the proteins that are found in the ER. As the ER is contiguous with the nuclear envelope as well as the cytomembrane in eukaryotic cells, its evolution most probably took place in concert with the nucleus<sup>5,12,25</sup>. For a number of proteins (namely hsp70 and hsp90), which function as molecular chaperones in the transport of other 'passenger proteins' across membranes, distinct homologs are found in the ER and cytosolic compartments<sup>25,34</sup>. Recently, the genes for both ER and cytosolic hsp70s have been cloned from *G. lamblia*, which is one of the most primitive eukaryotic species presently known<sup>25</sup>. The cloning of a gene encoding an ER hsp70 homolog from *G. lamblia* provides strong evidence that the ER originated very early in the eukaryotic cell history. Both the ER and cytosolic hsp70s from all eukaryotic species share a number of sequence features with the Gram-negative bacteria (including the 23–27 amino acid insert in the amino-terminal quadrant), but they also contain numerous unique shared sequence signatures that are not present in any prokaryotic or organellar homologs (see Fig. 2; Ref. 25). Phylogenetic analyses of ER and cytosolic hsp70s from various eukaryotic species indicate that they form distinct subfamilies,

which evolved by a gene duplication event that took place very early in the evolution of eukaryotic cells<sup>25</sup>. A similar inference has been reached for the ER and cytosolic forms of the hsp90 family of proteins<sup>34</sup>. These observations strongly indicate that the formation of ER (and through inference, nuclear envelope) was accompanied by duplication of the genes for proteins that facilitate membrane transport of other proteins.

These observations and the chimeric nature of the eukaryotic nuclear genome discussed above could be explained if the origin of eukaryotic cells involved an endosymbiotic event between a Gram-negative eubacterial host and an archaeobacterial 'eocyte' species<sup>12,24,25</sup>. In one possible scenario, shown in Fig. 3, a Gram-negative eubacterium lacking a cell wall developed a symbiotic relationship with an archaeobacterium. The host eubacterium, either naturally or as a result of this association, developed numerous membrane infolds. Over a period of time the membrane of the host organism completely surrounded the guest species, which became the nucleus and some of the membrane infolds developed into the ER. (The membrane of the guest species, which became redundant under these conditions was eventually lost.) The formation of the nuclear envelope, or ER, involved the creation of a new compartment in the cell that had to communicate (i.e. import/export proteins and other molecules) with the rest of the cell. Therefore, formation of this compartment was either accompanied

or preceded by duplication of the genes for the chaperone proteins (e.g. hsp70, hsp90, etc.), which are essential for this purpose<sup>29,30</sup>. Subsequently, the genome of the host organism was transferred to the newly formed nucleus, probably by a similar mechanism as that responsible for the nuclear transfer of genes from the later endosymbionts, mitochondria and chloroplasts<sup>7</sup>.

#### Concluding remarks

The model proposed here can explain many of the distinctive genetic, biochemical and structural characteristics of the eukaryotic cells, including the ester-linked nature of their membrane lipids<sup>1–6,12</sup>. The origin of some of the other distinctive features of eukaryotic cells (e.g. cytoskeleton) remains unclear at present, however, structural components for these could also be derived from one of the prokaryotic parents<sup>35–37</sup>. By contrast to the presently accepted view, this model proposes that the ancestral eukaryotic cell evolved by a very different mechanism than that responsible for the diversity seen within the prokaryotic world or subsequently within the eukaryotic world. Unlike the prokaryotic species, which are homogenomic, all of the eukaryotic species are indicated to be heterogenomic<sup>38</sup>. Furthermore, our observation that all of the eukaryotic, cytosolic and ER hsp70s contain a number of shared sequence signatures, which are not found in any prokaryotic or organellar homologs, strongly suggests that the major evolutionary transition from the prokaryotes to the eukaryotes was the result of a single, unique endosymbiotic fusion event<sup>25,39</sup>. Thus, the model proposes that, by contrast to the 'pre-karyotic' nature of the prokaryotes, all of the eukaryotic species are 'post-karyotes', where karyosis refers to the original endosymbiotic event. Lastly the protein-based phylogenies also raise important questions concerning the relationship between archaeobacterial and eubacterial species, some of which have been discussed previously<sup>23,24,27,40</sup>.

#### Acknowledgements

We thank J. Cechetto for assistance with the artwork. Work from R. S. G.'s laboratory is supported by a research grant from the Medical Research Council of Canada. G. B. G. is a fellow of the Canadian Institute for Advanced Research and is supported by a Natural Sciences and Engineering Research Council of Canada grant.

## References

- 1 Margulis, L. (1970) *Origin of Eukaryotic Cells*, Yale University Press
- 2 Stanier, R. Y., Dondroff, M. and Adelberg, E. A. (1990) *The Microbial World* (3rd edn), Prentice Hall
- 3 Knoll, A. H. (1992) *Science* 256, 622–627
- 4 Woese, C. R. (1981) *Sci. Am.* 244, 98–122
- 5 Cavalier-Smith, T. (1987) *Ann. NY Acad. Sci.* 503, 17–54
- 6 Zillig, W. (1991) *Curr. Opin. Genet. Dev.* 1, 544–551
- 7 Gray, M. W. (1992) *Int. Rev. Cytol.* 141, 233–357
- 8 Margulis, L. (1993) *Symbiosis in Cell Evolution* (2nd edn), W.H. Freeman & Co.
- 9 Mereschowsky, C. (1905) *Biol. Zbl.* 25, 595–604
- 10 Woese, C. R. (1987) *Microbiol. Rev.* 51, 221–271
- 11 Lake, J. A., Henderson, E., Clark, M. W. and Matheson, A. T. (1982) *Proc. Natl Acad. Sci. USA* 79, 5948–5952
- 12 Lake, J. A. and Rivera, M. C. (1994) *Proc. Natl Acad. Sci. USA* 91, 2880–2881
- 13 Iwabe, N. et al. (1989) *Proc. Natl Acad. Sci. USA* 86, 9355–9359
- 14 Gogarten, J. P. et al. (1989) *Proc. Natl Acad. Sci. USA* 86, 6661–6665
- 15 Woese, C. R., Kandler, O. and Wheelis, M. L. (1990) *Proc. Natl Acad. Sci. USA* 87, 4576–4579
- 16 Rivera, M. C. and Lake, J. A. (1992) *Science* 257, 74–76
- 17 Hartman, H. (1984) *Spec. Sci. Technol.* 7, 77–83
- 18 Sogin, M. L. (1991) *Curr. Opin. Genet. Dev.* 1, 457–463
- 19 Puhler, G. et al. (1989) *Proc. Natl Acad. Sci. USA* 86, 4569–4573
- 20 Smith, M. W., Feng, D. F. and Doolittle, R. F. (1992) *Trends Biochem. Sci.* 17, 489–493
- 21 Tiboni, O., Cammarano, P. and Sanangelantoni, A. M. (1993) *J. Bacteriol.* 175, 2961–2969
- 22 Benachenhou-Lahfa, N., Forterre, P. and Labedan, B. (1993) *J. Mol. Evol.* 36, 335–346
- 23 Gupta, R. S. and Golding, G. B. (1993) *J. Mol. Evol.* 37, 573–582
- 24 Gupta, R. S. and Singh, B. (1994) *Curr. Biol.* 4, 1104–1114
- 25 Gupta, R. S., Aitken, K., Falah, M. and Singh, B. (1994) *Proc. Natl Acad. Sci. USA* 91, 2895–2899
- 26 Brown, J., Masuchi, Y., Robb, F. T. and Doolittle, W. F. (1994) *J. Mol. Evol.* 38, 566–576
- 27 Golding, G. B. and Gupta, R. S. (1995) *Mol. Biol. Evol.* 12, 1–6
- 28 Steel, M. A., Lockhart, P. J. and Penny, D. (1993) *Nature* 364, 440–442
- 29 Morimoto, T. I., Tissieres, A. and Georgopoulos, C. (eds), (1994) *The Biology of Heat Shock Proteins and Molecular Chaperones*, Cold Spring Harbor Laboratory Press
- 30 Boorstein, W. R., Zeigelhoffer, T. and Craig, E. A. (1994) *J. Mol. Evol.* 38, 1–17
- 31 Falah, M. and Gupta, R. S. (1994) *J. Bacteriol.* 176, 7748–7753
- 32 Klank, H.P. and Doolittle, R. F. (1994) *Curr. Biol.* 4, 920–922
- 33 Doolittle, R. F. (1995) *Proc. Natl Acad. Sci. USA* 92, 2421–2423
- 34 Gupta, R. S. (1995) *Mol. Biol. Evol.* 12, 1063–1073
- 35 Bermudes, D., Hinkle, G. and Margulis, L. (1994) *Microbiol. Rev.* 58, 387–400
- 36 Erickson, H. (1995) *Cell* 80, 367–370
- 37 Gupta, R. S. and Soltys, B. J. *Biochem. Mol. Biol. Int.* (in press)
- 38 Margulis, L. (1996) *Proc. Natl Acad. Sci. USA* 93, 1071–1076
- 39 Szathmari, E. and Maynard-Smith, J. (1995) *Nature* 374, 227–232
- 40 Gupta, R. S. (1995) *Mol. Microbiol.* 15, 1–11

## Type VI membrane proteins?

It is well established that integral membrane proteins can be of a variety of different types. Under the Singer classification<sup>1</sup>, type I and II membrane proteins have one transmembrane polypeptide domain, but differ in their orientation. Type III membrane proteins contain multiple transmembrane regions and type IV proteins are made up of multimers of subunits that, when assembled together, make a transmembrane channel. Proteins attached to the lipid bilayer only by covalently bound lipids are classified as type V (see Fig. 1).

Recently, some membrane proteins have emerged that contain two different modes of membrane attachment on the same protein molecule. These proteins are anchored by one or more transmembrane-spanning domains and also by a glycosyl-phosphatidylinositol (GPI) anchor. This novel group of membrane-anchored proteins have been designated as type VI (Ref. 2; and Fig. 1).

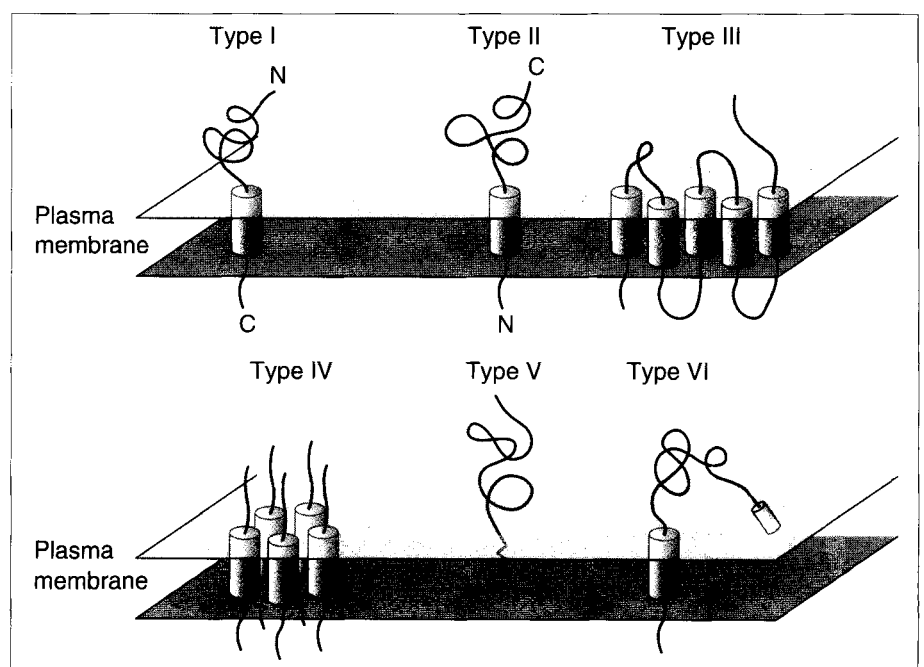
A construct containing an uncleaved amino-terminal signal peptide and a signal for GPI-anchor attachment was shown to produce a protein anchored by both transmembrane and GPI-anchors<sup>3</sup>. This demonstrated that it was possible for the cellular machinery to synthesize such proteins. Around the same time, reports of natural type VI proteins began to emerge. One such protein was ponticulin, a cell-surface protein that forms the major high affinity link between the plasma membrane and the actin cortical network<sup>4</sup>. Cloning of the ponticulin gene from *Dictyostelium discoideum* suggested that the hydrophobic domain encoded at the 5' end of the gene was part of a GPI-anchor-

attachment signal sequence<sup>2</sup>. Upon closer examination, ponticulin was shown to incorporate both [<sup>3</sup>H]-palmitic acid and [<sup>3</sup>H]-ethanolamine, compounds found in GPI-anchors. Enzymes that can usually cleave GPI-anchored proteins, for example phosphatidylinositol-specific phospholipase C (PI-PLC), could not release ponticulin from purified plasma membranes. Furthermore, an antibody recognizing native ponticulin was shown to bind only to the cytoplasmic-exposed side of the protein, indicating the presence of a cytosolic domain(s). It is quite likely that a human homologue of ponticulin exists, as experiments using anti-ponticulin

antibodies have identified a 17 kDa protein in polymorphonuclear leucocytes<sup>5</sup>.

Another protein that has been suggested to be anchored in this manner is Sm23 (Ref. 6). Sm23 is a cell-surface protein found on the human gut parasite, *Schistosoma mansoni*. The presence of a GPI anchor has been demonstrated on this protein<sup>7</sup>. Experiments utilizing [<sup>3</sup>H]-ethanolamine incorporation and detergents have implicated the presence of both a GPI anchor and transmembrane-spanning domains on Sm23 (Ref. 6).

Research into this dual mechanism of membrane attachment is still in its infancy and poses many questions. Some proteins



**Figure 1**

Types of membrane-anchored protein. The lipid bound to type V proteins is represented as a zigzag and the glycosyl-phosphatidylinositol (GPI) anchor of type VI proteins is shown as a yellow cylinder.