PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

The Origins and Evolution of Eukaryotic Proteins

Russell F. Doolittle

Phil. Trans. R. Soc. Lond. B 1995 **349**, 235-240
doi: 10.1098/rstb.1995.0107

The origins and evolution of eukaryotic proteins

RUSSELL F. DOOLITTLE

Center for Molecular Genetics, University of California, San Diego La Jolla, California 92093–0634, U.S.A.

SUMMARY

The common ancestry of eukaryotes, archaebacteria and eubacteria is well demonstrated by amino acid sequence comparisons of numerous proteins that are common to all three groups. On the other hand, there are a few proteins, like ubiquitin, that are common to eukaryotes and archaebacteria and which have yet to be observed in eubacteria. Some proteins appear to be wholly restricted to eukaryotes; this is especially true of cytoskeletal proteins. Recently, actin has been found by crystallography to be homologous with an ATP-binding domain found in a heat shock protein and several other proteins common to all three urkingdoms. This observation is puzzling on several counts. Most cytoskeletal proteins like actin and tubulin are very slow changing and must have been so for a very long time. How is it, then, that no sequence resemblance can be discerned with their alledged prokaryotic antecedents? The question is addressed by considering two bacterial fts proteins which appear to be related to actin, on the one hand, and tubulin, on the other. One answer may be that the rate of change of these proteins changed dramatically at a key point in their history. Another possibility is that eukaryotes are much older than some of their other proteins indicate.

1. INTRODUCTION

Until 1977, the universe of cellular organisms was divided simply into those that had nuclei and those that did not. At that juncture, however, a comparison of ribosomal RNA sequences revealed that there were really three realms, the bacteria falling into two distinct groups (Woese & Fox 1977*b*). The three 'urkingdoms' were: (i) the previously recognized eukaryotes; (ii) ordinary bacteria (now to be called 'eubacteria'); and (iii) a new assemblage of bacteria which were called the 'archaebacteria'. The evolutionary relation of the three groups remains enigmatic and controversial. The controversy has swirled around two aspects: first, are the archaebacteria ancient (as their name implies), or were they derived from preexisting eubacteria? Second, are the eukaryotes the result of a very early divergence that occurred shortly after life began, or are they relatively modern and derived from an archaebacterial ancestor?

Ribosomal RNA sequences were unable to provide answers these questions, the three groups being more or less equidistant as measured by differences in the RNA sequences, and the hope was that a thorough analysis of protein sequences common to all three groups would solve the riddle. Alas, the more data that have been accumulated, the more perplexing the matter seems to become. Thus some proteins from eubacteria are clearly much more akin to the eukaryotic equivalents than they are to archaebacterial ones (see figure 1*a*), but some others put the archaebacteria nearer the eukaryotes (see figure 1*b*). Other proteins cluster the two bacterial groups together (see figure 1*c*), and yet others are so mismatched, it is claimed, that the only explanation seems to be that eukaryotes are the product of a fusion of a eubacterium and an archaebacterium (see figure 1*d*).

In this brief article I describe what I consider to be the major dilemma about the origin of eukaryotes and those proteins that most set them apart from the other two groups. In particular, I will attempt to illustrate the problem by considering the origin of two cytoskeletal proteins: actin and tubulin. These proteins are noteworthy for their ability to polymerize, reversibly, into filaments or microtubules in association with the hydrolysis of ATP or GTP, respectively. As such, they control most of the mechanics of cellular movement

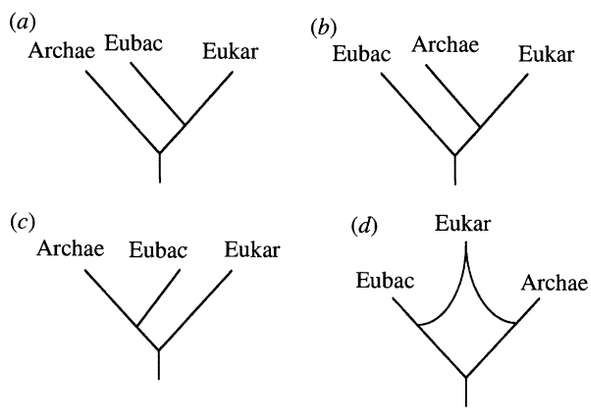

Figure 1. Diverse phylogenies for the three major biological groupings as suggested by sequence comparisons involving different proteins. (*a*) phosphglycerate kinase, gap dehydrogenase, (*b*) elongation fractors, HMG-CoA reductase, (*c*) glutamine synthetase, citrate synthetase, (*d*) heat shock 70 family.

Phil. Trans. R. Soc. Lond. B (1995) **349**, 235–240
Printed in Great Britain

including cell division, phagocytosis, general motility and other dynamic phenomena. To appreciate their present circumstances and their possible origins, it will be necessary for us to review some general principles of protein evolution. Before that, however, let us examine in a little more detail some of the scenarios that have been suggested for the evolution of eukaryotes themselves.

2. THE ORIGIN OF EUKARYOTIC CELLS

Early notions about these complex cells presumed that they merely evolved from simpler cells of the bacterial type. Even with the wide acceptance of the endosymbiotic hypothesis, whereby it was proposed that the principal organelles of the eukaryotic cell were derived from endosymbiotic bacteria, it was still presumed that the host was itself a large bacterium ('prokaryote'). This notion was challenged by Stanier (1970), and later by Woese & Fox (1977*a*), among others, all of whom underscored the fundamental difference in the cytosolic makeup of 'prokaryotes' and 'eukaryotes.' In particular, bacterial cells maintain their integrity largely by way of cell walls, whereas eukaryotic cells depend on an internal cytoskeleton. It was this consideration that led Woese & Fox (1977*a*) to propose that prokaryotes (bacteria) and eukaryotes diverged from a common ancestor at a very early stage in the history of life, presumably more than 3.5 billion years ago, inasmuch as there are microfossils of that age that resemble modern bacteria. They dubbed this primitive common ancestor the 'progenote.'

In contrast, Cavalier-Smith (1975, 1987) has made a strong argument favouring the progressive evolution of the three cellular types, beginning with eubacteria, the cell walls of various lineages differentiating into the chemically distinguishable gram-positive and gram-negative. It was the loss of the ability to make a muramic acid-containing cell wall, he argues, that gave rise to the characteristic membranous coverings of archaebacteria, and the further development of the eukaryotic cytoskeleton that allowed the complete shedding of an outer cell wall (Cavalier-Smith 1987). Like Stanier (1970) and Woese & Fox (1977*a*), he underscores that it is the eukaryotic cytoskeleton that allows phagocytosis, which as an incidental consequence permitted the existence of the endosymbionts which were to become mitochondria and chloroplasts, and perhaps some other organelles as well.

In the interim since these two widely different scenarios were posed, a considerable amount of macromolecular sequence data has been accumulated that have a bearing on the argument. Ribosomal RNA sequences continue to affirm that there are three primary kingdoms, called by most the Eubacteria, the Archaebacteria and the Eukarotes, but by some, the Archaea, Bacteria and Eukarya (Woese *et al.* 1990). As noted above, however, it has not been possible to prioritize the appearance of the three groups with the ribosomal RNA data. In contrast, protein sequences permit the use of a rooting strategy based on duplicated

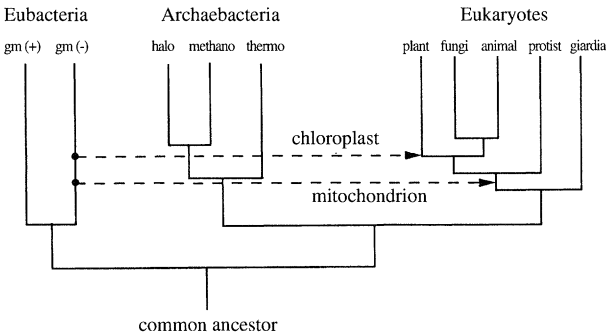

Figure 2. 'Tree of Life' showing archaebacteria sharing common ancestor with eukaryotes more recently than eukaryotes with eubacteria.

genes (Iwabe *et al* 1989), and the bulk of that evidence (see for example, Brown & Doolittle 1995) appears to favor the stepwise scheme proposed by Cavalier-Smith (see figure 2). Moreover, eukaryotes and archaebacteria share a number of advanced features, including proteasomes (Wenzel & Baumeister 1993) and the protein ubiquitin (Wolf *et al.* 1993), that are absent from eubacteria.

Nonetheless, there are a number of anomalies, and some investigators have come to regard the eukaryote – mitochondria and chloroplasts aside – as a chimera of a eubacterium and an archaebacterium (Pühler *et al.* 1989; Gupta & Golding 1993). In particular, some of these proposals point to the presence of the nucleus itself as evidence of an early endosymbiotic event that allowed the mixing of the two kinds of genome (Lake & Rivera 1994). It must be emphasised, however, that there is no precedent for a bacterium ever being able to maintain an endosymbiont. So far as is known, only a cell with a cytoskeleton has the machinery for effecting such a relation.

The internal cytoskeleton is composed of an array of proteins unique to eukaryotes, the most abundant of which are actin, myosin, tubulin, cytokeratins and their relatives. If we could trace the origins of these proteins, we should be able to unearth the origins of the eukaryotic cell itself and perhaps decide between the most extreme views about its origin.

3. EVOLUTIONARY CHANGE IN PROTEINS

Although many different factors can influence the rate of change in an amino acid sequence, it is generally observed that two diverging protein sequences change in a more or less stochastic manner consistent with an exponential decay (see figure 3). With the passage of time, a point will eventually be reached where the amount of sequence resemblance will not be sufficient to offer assurance of common ancestry. It is well known, however, that the three-dimensional aspect of related proteins is preserved long after all sequence resemblance has been eroded, and comparisons of *X*-ray and nuclear magnetic resonance (NMR) structures are able to reveal relations that are not apparent by simple sequence comparison. There is

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

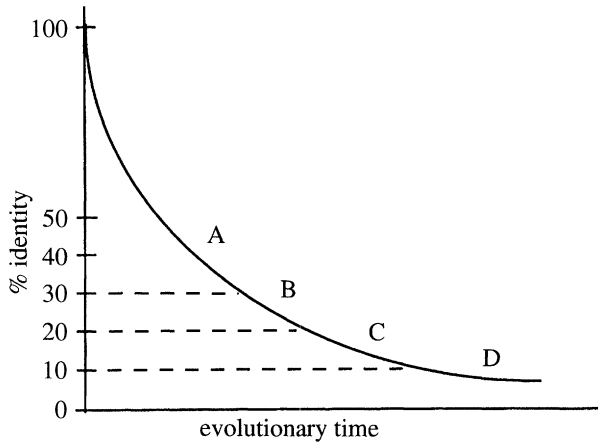

Figure 3. Sequence divergence from a common ancestor occurs as a negative exponential. If two (sufficiently long) sequences are more than 30% identical (A), simple visual inspection is enough to warrant a judgment about common ancestry. As the resemblance decays (B–D), increasingly sophisticated means are required to validate the relationship (from Doolittle 1995). A: simple sequence comparison; B: weighted scales, statistics, C: multiple alignments, profile analysis, D: 3-D structure comparison.

an intermediate stage between simple observation and complete obliteration of sequence information wherein a relation can be discerned by various computer-based pattern analyses (see figure 3). The cut-off points for these various stages are not hard and fast, but as a rule of thumb, any two sequences that are more than 30% identical can safely be presumed to have diverged from a common ancestor. If they have diverged much more than that, multiple sequence comparisons and/or other statistical considerations may have to be enjoined.

Different proteins change at different but characteristic rates (see figure 4). The factors that determine a particular protein's rate of change may be genetic, structural or functional. The fastest changing proteins in eukaryotic cells still maintain obvious sequence resemblance over the course of hundreds of millions of years, and the slowest changing ones may maintain virtually identical sequences up to as much as a billion years. Most metabolic enzymes are changing at a rate such that their eukaryotic and eubacterial versions are readily recognizable, the average percent identity between the two groups being more than 30% (Doolittle 1993). On the other hand, the eukaryotic cytoskeletal proteins are the slowest changing proteins known, their rate of change being less than 10% that of the average mainstream metabolic enzyme (see figure 4). One reason that they change so slowly may have to do with the large number of macromolecular interactions they must preserve. Thus not only must they maintain a structural compatibility that allows reversible polymerization, they need to be recognized also by a host of other gene products. In the case of actin, for example, these may include myosin, profilin, gelsolin, fragmin, villin, and severin, *inter alia*. As such the various interactants are genetically interlocked, change in any one having a potential consequence for the rest.

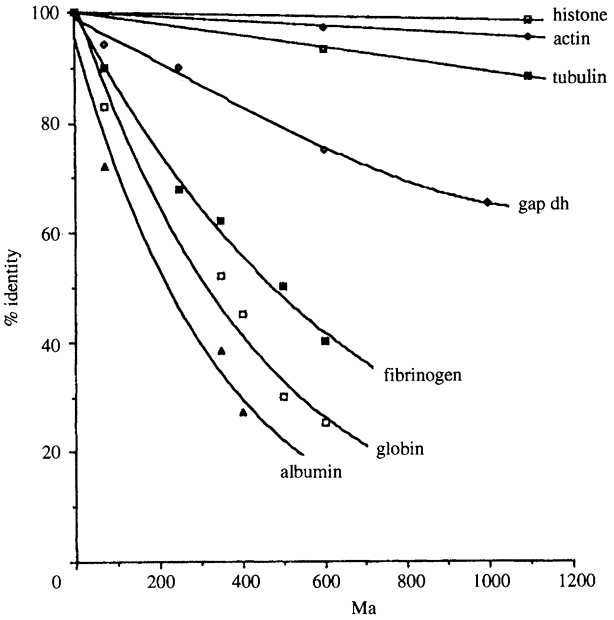

Figure 4. Different proteins change at different but characteristic rates. Note the very slow rate of change of histone, actin and tubulin relative to a mainstream enzyme like glyceraldehyde-3-phosphate dehydrogenase (gapdh) (from Doolittle 1992).

4. THE ORIGINS OF ACTIN AND TUBULIN

As noted above, actin and tubulin are at the heart of eukaryotic cell dynamics. They are among the slowest changing proteins known (see figure 4), and, until recently there was not a hint of similar proteins among the bacteria. Then, in 1991, the X-ray structure of an actin was determined, and comparisons with previously reported three-dimensional structures for other proteins revealed several unexpected relations (Flaherty *et al.* 1991). Foremost was an obvious resemblance to the ATP-binding domain of heat shock (HS-70) proteins. There was also an apparent relation to the ATP-binding portion of hexokinase. The structural similarities were strong enough that common ancestry seems certain.

Other investigators (Bork *et al.* 1992), using a pattern based on the structural alignment of actins, HS-70 proteins and hexokinase, were able to extend the network of related proteins, including most notably a protein involved in bacterial cell division called ftsA (filamenting, temperature sensitive).

Meanwhile, cell biologists had characterized forms of actin that differed significantly in sequence from previously reported actins (Lees-Miller *et al.* 1992). These actins were associated with microtubules or centrosomes and have come to be known as centractins (Clark & Meyer 1992, 1993). An alignment of portions of sequences from these two forms of actin and two bacterial ftsA proteins is presented in figure 5*a*; a phylogenetic tree based on this alignment is shown in figure 6*a*.

It is the functional connection of centractin with microtubules and tubulin that I want to emphasise here. Tubulin, like actin, falls into several families, denoted alpha, beta and gamma. The form associated with the centrosome is called gamma tubulin (Zheng *et*

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

(*a*)

```
ATHU        ..LVCDNGSGLVK AGFAGDDAPRAVFPSI VGR PRHQGVMVGMGQKDSYVGDE AQSKRGILTLKYPIEHGII..
ATAX        ..LVVDNGSGMCK AGFAGDDAPRAVFPSI VGR PRHVSVMAGMGQKDAYVGDE AQSKRGILTLKYPIEHGIV..
ATBY        ..LVIDNGSGMCK AGFAGDDAPRAVFPSI VGR PRHQGIMVGMGQKDSYVGDE AQSKRGILTLRYPIEHGIV..

CTHU        ..VVIDNGSGVIK AGFAGDQIPKYCFPNY VGR PKHVRVMAGALEGDIFIGPK AEEHRGLLSIRYPMEHGIV..
CTPN        ..ICIDNGSGVIK AGFAGEDQPKSFFPSY VGR PKHLKIMAGAIEGDIFIGNK AQELRGLLKIKYPIEHGIV..
CTNC        ..IVLDNGSGTIR AGFAGDDVPKCHFPSF VGR PKHLRVLAGALEGEVFIGQK AASELRGLLKIRYPLEHGIV..

FTBS        ..VSLDLGTSNTK  VIVGEMTGDS LNIIGVGNVPSEGLKKGSIVDIDETVHSIRKAFDQAERMVGFPLRKAIV..
FTEC        ..LVVGLEIGTAKVAALVGEVLPDGMVNIIGVGSCPSRGMDKGGVNDLESVVKCVQRAIDQAELMAD  CQISSV..
```

(*b*)

```
TUAH        ..HYTIGKEIIDLVLDRIRKLADQCTGLQGFLVFHSFGGGTGSGFTSLLM        ERLSVDYGK KSKLEFSIYP APQVST
TUAN        ..HYTIGKEIVDLCLDRIRKLADNCTGLQGFLVFSSVGGGTGSGLGALLL        ERLSVDYGK KSKLGFTVYP SPQVAT
TUBG        ..HYTEGAELVDAVLDVVRKRSEACDCLQGFQICHSLGGGTGAGMGTLLI        AKIREEYPD RMMCTFSVVP SPKVSD
TUBH        ..HYTEGAELVDAVLDVVRKEAESCDCLQGFQLTHSLGGGTGSGMGTLLI        SKMREEFPD RIMNTFSVVP SPKVSD
TUGH        ..GFSQGEKIHEDIFDIIDREADGSDSLEGFVLCHSIAGGTGSGLGSYLL        ERLNDRYPK KLVQTYSVFPNQDEMSD
TUGY        ..GYSHAERIFEDIMDMIDREADGSDSLEGFSLLHSIAGGTGSGLGSFLL        ERLNDRYPK KIIQTYSVFPNSQSVSD

FTEC        ..NPEVGRNAADEDRDALRAALEGADMV  F IAAGMGGGTGTG AAPVV        AEVAKDLG    ILTVAVVTKPFNFEG
FTWO        ..LPDVGKGAAEESIDEIMEHIKDSHML  F ITAGMGGGTGTGAAPVIAKAAREARAAVKDRAPKEKKILTVGVVTKPFGFEG
FTRM        ..QPEVGRAAAEECIDEIIDHLQGTHMC  F VTAGMGGGTGTGAAPIVA            QAARNKGILTVGVVTKPFHFEG

TUAH        AVVEPYNSILTTHTTLEHSDCAFMVDNEAIYDICRRNLDIERPTYTNLNRLIGQIVSSITASLRFDGALNVDLTEFQTNLVP..
TUAN        AVVEPYNSVLSTHALLEHTDVAVMLDNEAIYDICRRSLDIQRPTYTNLNRLIAQVISSLTASLRFDGALNVDVTEFQTNLVP..
TUBG        TVVEPYNATLSVHQLVEHADEVFCIDNEALYDICFRTLKLTCPTYGDLNHLVSLVMSGCTSCLRFPGQLNADLRKLAVNLIP..
TUBH        TVVEPYNATLSVHQLVENTDETYCIDNEALYDICFRTLKLTTPTYGDLNHLVSATMSGTVTCLRFPGQLNADLRKLAVNMVP..
TUGH        VVVQPYNSLLTLKRLTQNADCLVVLDNTALNRIATDRLHIQNPSFSQINQLVSTIMSASTTTLRYPGYMNNDLIGLIASLIP..
TUGY        VVVQPYNSLLALKRLTLNADSVVVLDNAALAHIAADRLHTQNPTFHQQNQLVSTVMSASTTTLRYPGYMNNDLVSIIASLIP..

FTEC        KKRMAF AEQGITELSKHVNSLITIPNDKLLKVLGRGISL LDAFGAANDVLKGAVQGIAELITRPGLMNVDFADVRTVMSE..
FTWO        VRRMPI AELGLEELQKYVDTLIVIPNQNLFRIANEKTTF SDAFKLADNVLHIGIRGVTDLMVMPGLINLDFADIETVMSE..
FTRM        GRRMRI ADQGISDLQKSVDTLIVIPNQNLFRIANDKTTF ADAFAMADQVLYSGVACITDLMVKEGLINLDFADVRSVMRE..
```

Figure 5. (*a*) Alignment of portions of various actin sequences (eukaryotic) and corresponding parts of two bacterial ftsA proteins. Residues where the two groups have significant similarity are emboldened. Codes: AT and CT = actin and centractin, respectively. FT = ftsA protein. Species: HU, human; AX, *Entamoeba*; BY, bakers yeast; NC, *Neurospora crassa*; PN, *Pneumocystis*; BS, *Bacillus subtilis*; EC, *Escherichia coli*. (*b*) Alignment of portions of various tubulins (eukaryotic) and corresponding parts of three bacterial ftsZ proteins. Codes: TUA, TUB and TUG = alpha, beta and gamma tubulins, respectively. FT = ftsZ proteins. Species: H, human; N, *Neurospora*; G, *Giardia*; Y, yeast; EC, *E. coli*; WO, *Wolbachia*; RM, *Rhizobium melioti*.

al. 1991). Recently, a motif common to all the tubulins was recognized in a bacterial protein called ftsZ (de Boer *et al.* 1992; RayChaudhuri & Park 1992). This protein is a member of the same apparatus involved in bacterial cell division that was mentioned above. Indeed it is a GTP-binding protein (a GTPase) and lies adjacent to ftsA in the bacterial genome (Vinella *et al.* 1993); this cannot be a coincidence. An alignment of portions of various members of the tubulin family and three ftsZ sequences is presented in figure 5*b* and a phylogenetic tree based on that alignment shown in figure 6*b*.

5. THE PARADOX

Consider the facts. Two of the slowest changing proteins in eukaryotes have functional counterparts in eubacterial systems. In the one case, common ancestry is clear from three-dimensional considerations. In the other, a common ancestor seems very likely on the basis of multiple sequence comparison (see figure 5*b*). But in both cases the sequence resemblance between the bacterial and eukaryotic proteins is very low. How is it that the sequences for the much faster changing metabolic enzymes are easily recognized in bacteria and eukaryotes, but these otherwise slowly changing sequences are not?

We must reflect on the degree of rate change we are considering here. Eukaryotic actins have changed at a rate of about 10% per billion years. As noted above, the course of divergent change follows an exponential decay rule, and the 50% difference between actins and centractins, if linearly extrapolated, indicates a duplication that clearly precedes the divergence of eukaryotes and bacteria (or for that matter, the origin of the solar system). The ftsA protein, which has both structural and functional resemblances to actins, has less than 20% sequence identity. If the divergence between these families took place about two billion years ago, which is an estimated date for the divergence of eukaryotes and eubacteria (Doolittle *et al.* 1989), then the minimum rate of change over the next billion years that could account for the observed degree of resemblance would have to be about 50 times greater than the rate over the next billion years, and considerably greater than was occurring for most enzyme proteins during the same period.

The characteristic rate of change for a family of proteins can change, of course, as is evident from casual inspection of the phylogenies in figure 6. For example,

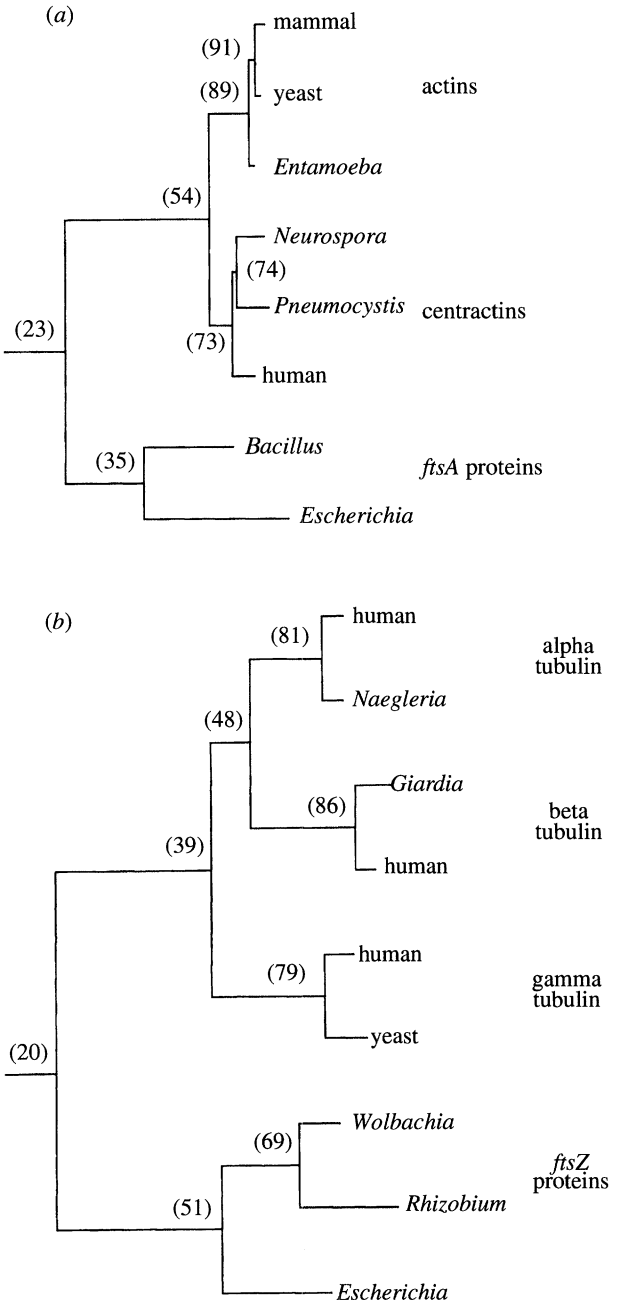

Figure 6. (*a*) Phylogenetic tree of actins, centractins and ftsA proteins based on alignment in figure 5*a*. The numbers in parentheses indicate the (average) percent identity for the taxa joined by that node. (b) Phylogenetic tree of tubulins and ftsZ proteins based on alignment shown in figure 5*b*.

centractins are changing at least twice as fast as actins. This is a far cry, however, from the drastic amount of change required to account for the differences between members of the eukaryotic actin family and related proteins in eubacteria.

Are there alternative or supplementary explanations? One possibility is that the divergence of the eukaryotic actins and tubulins and their eubacterial cousins occurred longer ago than was the case for those other proteins in the cell that have easily recognized counterparts in both groups. Indeed, the original suggestion by Woese & Fox (1977*b*), later amplified by Hartman (1984) and Sogin (1991), envisioned an RNA-based 'urkaryote' that was capable of making cytoskeletal proteins, if not much else. The rate of protein sequence change in an RNA-based system would have been enormously greater than occurs in DNA-based systems. At some later point, the genes for other proteins, including what are now considered mainstream metabolism enzymes, were acquired by the phagocytosis of a DNA-based bacterium. Sogin (1991) also made the point that if the endosymbiont were an archaebacterium, then the close kinship of so many eukaryotic and archaebacterial proteins would be explained. The argument is not without merit. I am also moved by the fact that Hartman (1975) had long ago pointed out that the centriole was a primitive RNA-dependent structure critical to the evolution of the eukaryotic cell.

There is one important observation that stands in the way of this explanation of events. The 'urkaryote' would have needed to employ a system of coding and protein manufacture that was fundamentally the same as would be passed on to the other early diverging lineages. Sogin (1991) specifically lists the aminoacyl tRNA synthetases as a part of the armamentarium of the primitive host-to-be. The aminoacyl tRNA synthetases, however, are among the many enzymes whose sequences are readily recognized between eukaryotes and eubacteria (Nagel & Doolittle 1991). Either the 'urkaryote' was still making proteins in a wholly RNA-based context (no aminoacyl tRNA synthetases), or we are stymied again by cytoskeletal proteins experiencing an inordinate rate of change.

6. CONCLUSIONS AND PERSPECTIVES

In spite of the identification of some related proteins in bacteria, the invention and evolution of actin and tubulin remain mysterious. Their very slow rates of change over the course of the last billion years stands in marked contrast to the vast amounts of sequence change that would have had to occur in the period immediately following their divergence from the common ancestor that also led to the functionally similar fts proteins, or, in the case of actin, to HS-70 and other ATP-binding proteins. One possibility is that these divergences occurred before the invention of those replication features that slowed protein evolution in general. The same can not be true of many other proteins, however, obviously similar sequences existing in eukaryotes and eubacteria. It may be that early suggestions of an RNA-based progenote capable of making cytoskeletal proteins are correct, and that this progenote acquired its genes for most other proteins from an early endosymbiont. The alternative is an abrupt change in the rate of sequence evolution for actins and tubulins and their relatives that is of the order of ten- to a hundredfold.

I thank Karen Anderson for generous help in the preparation of this manuscript and Glen Cho for assistance in preparing some of the figures and Paul Sharp for pointing out a mistaken organism assignment in an EMBL Data base entry. My laboratory is supported by the U.S. National Institutes of Health.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

REFERENCES

Bork, P, Sander, C. & Valencia, A. 1992 An ATPase domain common to prokaryotic cell cycle proteins, sugar kinases, actin, and hsp70 heat shock proteins. *Proc. natn Acad. Sci. U.S.A.* **89**, 7290–7294.

Brown, J. R. & Doolittle, W. F. 1995 The root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications. Proc. natn Acad. Sci. U.S.A. **92**, 2441–2445.

Cavalier-Smith, T. 1975 The origin of nuclei and of eukaryotic cells. *Nature, Lond.* **256**, 463–468.

Cavalier-Smith, T. 1987 The origin of eukaryote and archaebacterial cells. Ann. N.Y. Acad. Sci. **503**, 17–54.

Clark, S. W. & Meyer, D. I. 1992 Centractin is an actin homologue associated with the centrosome. *Nature, Lond.* **359**, 246–250.

Clark, S. W. & Meyer, D. I. 1993 Long lost cousins of actin. *Curr. Biol.* **3**, 54–55.

de Boer, P., Crossley, R. & Rothfield, L. 1992 The essential bacterial cell-division protein FtsZ is a GTPase. *Nature, Lond.* **359**, 254–256.

Doolittle, R. F. 1992 Reconstructing history with amino acid sequences. *Protein Sci.* **1**, 191–200.

Doolittle, R. F. 1993 Amino acid sequence comparison as an aid to determining evolutionary origins. In *Methods in protein sequence analysis* (ed. K. Imahori & F. Sakiyama), pp. 241–246. New York: Plenum Press.

Doolittle, R. F. 1995 The multiplicity of domains in proteins. *A. Rev. Biochem.* **64**, 287–314.

Doolittle, R. F., K. L. Anderson & D.-F. Feng 1989 Estimating the Prokaryote–Eukaryote Divergence Time From Protein Sequences. In *The hierarchy of life. Molecules and morphology in phylogenetic analysis* (ed. B. Fernholm, K. Bremer & H. Jörnvall), pp. 73–85. Amsterdam: Elsevier.

Flaherty, K. M., McKay, D. B., Kabsch, W. & Holmes, K. C. 1991 Similarity of the three-dimensional structures of actin and the ATPase fragment of a 70 kDa heat shock cognate protein. *Proc. natn Acad. Sci. U.S.A.* **88**, 5041–5045.

Gupta, R. S. & Golding, G. B. 1993 Evolution of HSP70 gene and its implications regarding relationships between archaebacteria, eubacteria, and eukaryotes. *J. molec. Evol.* **37**, 573–582.

Hartman, H. 1975 The centriole and the cell. *J. theor. Biol.* **51**, 501–509.

Hartman, H. 1984 The Origin of the eukaryotic cell. Spec. Sci. Technol. **7**, 77–81.

Iwabe, N., Kuma, K.-I., Hasegawa, M., Osawa, S. & Miyata, T. 1989 Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc. natn Acad. Sci. U.S.A.* **86**, 9355–9359.

Lake, J. A. & Rivera, M. C. 1994 Was the nucleus the first endosymbiont? *Proc. natn Acad. Sci. U.S.A.* **91**, 2880–2881.

Lees-Miller, J. P., Helfman, D. M. & Schroer, T. A. 1992 A vertebrate actin-related protein is a component of a multisubunit complex involved in microtubule-based vesicle motility. *Nature, Lond.* **359**, 244–245.

Nagel, G. & Doolittle, R. F. 1991 Evolution and relatedness in two aminoacyl-tRNA synthetase families. *Proc. natn Acad. Sci. U.S.A.* **88**, 8121–8125.

Pühler, G. *et al.* 1989 Archaebacterial DNA-dependent RNA polymerases testify to the evolution of the eukaryotic nuclear genome. *Proc. natn Acad. Sci. U.S.A.* **86**, 4569–4573.

RayChaudhuri, D. & Park, J. T. 1992 *Escherichia coli* cell-division gene *ftsZ* encodes a novel GTP-binding protein. *Nature, Lond.* **359**, 251–254.

Sogin, M. L. 1991 Early evolution and the origin of eukaryotes. *Curr. Opin. Genet. Dev.* **1**, 457–463.

Stanier, R. Y. 1970 Some aspects of the biology of cells and their possible evolutionary significance. *Symp. Soc. gen. Microbiol.* **20**, 1–38.

Vinella, D., Bouloc, P. & D'Ari, R. 1993 GTPase enters the ring. *Curr. Biol.* **3**, 65–66.

Wenzel, T. & Baumeister, W. 1993 *Thermoplasma acidophilum* proteasomes degrade partially unfolded and ubiquitin-associated proteins. *FEBS* **326**, 215–218.

Woese, C. R. & Fox, G. E. 1977*a* The concept of cellular evolution. *J. molec. Evol.* **10**, 1–6.

Woese, C. R. & Fox, G. E. 1977*b* Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. natn Acad. Sci. U.S.A.* **74**, 5088–5090.

Woese, C. R., Kandler, O. & Wheelis, M. L. 1990 Towards a natural system of organisms: proposal for the domains archaea, bacteria, and eucarya. *Proc. natn Acad. Sci. U.S.A.* **87**, 4576–4579.

Wolf, S., Lottspeich, F. & Baumeister, W. 1993 Ubiquitin found in the archaebacterium *Thermoplasma acidophilum*. *FEBS* **326**, 42–44.

Zheng, Y., Jung, M. K. & Oakley, B. R. 1991 γ-Tubulin is present in drosophila melanogaster and homo sapiens and is associated with the centrosome. *Cell* **65**, 817–823.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES