Amino Acids (1999) 17: 207–226

Amino
Acids

Evolutionary changes reflected by the cellular amino acid composition

K. Sorimachi

Department of Microbiology, Dokkyo University School of Medicine, Mibu, Tochigi, Japan

Accepted July 22, 1998

Summary. Comparison of the amino acid composition of cell-proteins using 17 amino acids has been used to investigate the biological evolution of organisms such as bacteria, blue-green alga, green alga, fungi, slime mold, protozoa and vertebrates. The degree of difference in the amino acid ratios between any two groups reflects the degree of divergency in biological evolution. The amino acid composition of the Gram-negative bacteria (*Escherichia coli*, *Klebsiella*, *Proteus*, and *Vibrio alginolyticus*) was identical. However, the amino acid composition of *Staphylococcus aureus* and *Bacillus subtilis*, which are Gram-positive bacteria, differed from each other and from the Gram-negative bacteria. The amino acid composition of the blue-green alga (Cyanobacterium, *Chroococcidiopsis*) was quite similar to that of *E. coli*. A marked difference in the amino acid composition was observed between *E. coli* and green alga (*Chlorella*), and significant differences were observed between *E. coli* and other organisms, such as fungi, protozoa (*Tetrahymena*), slime mold (*Dictyostelium discoideum*) and vertebrates. In conclusion, the change in cellular amino acid composition reflects the divergence which has occurred during biological evolution, whereas a basic pattern of amino acid composition is maintained in spite of a long period of evolutional divergence among the various organisms. Thus, it is proposed that the primitive life forms established at the end of prebiotic evolution had a similar amino acid composition.

Keywords: Amino acid – Evolution – Procaryotic cell – Eucaryotic cell

Introduction

Cellular proteins are synthesized based on the genetic information encoded by DNA, and biological evolution is a result of changes in the nucleotide sequences of the structural genes. Biological evolution has been investigated using the changes in the amino acid sequence of cytochrome C (Dayhoff and

Schwartz, 1981; Dayhoff et al., 1972) and other proteins (McLaughlin and Dayhoff, 1972), and in the nucleotide sequence of ribosomal RNA (Sogin et al., 1986). These studies accurately showed that small divergencies had occurred in the evolutionary tree, although enormous labour is required for this kind of investigation.

Living organisms provide us with most of the data available to study biological evolution. In addition, the discovery of fossils in sedimentary rocks provide us with morphological information to aid our understanding of biological evolution. The palaeontological study of fossil microorganisms in Precambrian rocks (Barghoon and Schope, 1966; Peat and Lloyd, 1974; MacGregor et al., 1974; Nagy and Zumberge, 1976) has shown that microorganisms are related to a more ancient form of life. Thus, evidence for biological evolution has been infered by pure scientific research.

On the other hand, we have only limited information concerning prebiotic evolution, because we do not have a suitable experimental system to generate meaningful data. At present, it is impossible to establish an experimental system taking into account the real time factor. However, Miller originally showed that electrical discharges in a millieu of NH_3, CH_4, H_2 and H_2O, which he believed to have existed on primitive Earth, induced the formation of amino acids such as glycine, alanine and asparatic acid (Miller, 1955; Miller and Urey, 1959). In addition, these amino acids have been identified in meteorites (Kvenvolden et al., 1970; Wolman et al., 1972). Thus, it is possible that these amino acids might have been formed initially on primitive Earth. It has also been shown that amino acid polymerization without enzymes can occur in the presence of clay (Lahav et al., 1978). Although these experimental results suggest that the formation of amino acids would be the first step in prebiotic evolution, additional events are required to proceed from this point to the evolution of microorganisms. The establishment of rules to characterize primitive life which may be tested on living organisms may help us to understand prebiotic evolution. It is reasonable to assume that certain rules are conserved and applicable to both prebiotic evolution and biological evolution. Of course, the establishment of codon formation, which was necessary for the appearance of life, must have been the decisive event at the boundary of prebiotic and biological evolution.

Occasionally, data based on the amino acid sequence or the corresponding nucleotide sequence of a protein are so complex that the overall trend might not be obvious, whereas a macro-investigation may reveal a trend which may otherwise be hidden. Based on this hypothesis, we have previously investigated whether the cellular amino acid composition reflects biological evolution (Okayasu et al., 1997), as the amino acid composition of proteins represents the nucleotide composition used for coding the protein structures, and the cellular amino acid composition represents the average value of all the proteins within a cell. In addition, the amino acid composition analysis of whole cells is much simpler to perform than analysis of the amino acid sequence or the corresponding nucleotide sequence of a protein. In our previous study (Okayasu et al., 1997), we found that the amino acid composition differed in mammalian and bacterial cells, but were almost identical within

mammalian cells. The present study has been designed to investigate the relationship between the changes in amino acid composition and biological evolution.

Materials and methods

Cell culture

Rat hepatoma cells (M) (Katsuta and Takaoka, 1968), human urinary bladder carcinoma cells (HUB-15) (Kakuya et al., 1983), and golden fish scale cells (Akimoto, unpublished) were cultured according to the method described previously (Okayasu et al., 1997). Green alga (*Chlorella*) was also cultured using DM-160 medium, which is used for mammalian cell culture (Katsuta and Takaoka, 1976). Bacterial cells (*Escherichia coli*, *Klebsiella*, *Vibrio alginolyticus*, *Proteus*, *Staphylococcus aureus* and *Bacillus subtilis*) were cultured on 1.5% agar containing bouillon. The bacterial cells were fixed with 70% ethanol in plastic micro-centrifuge tubes for more than 3 days and washed twice with 70% ethanol to remove amino acids originating from the medium. Fungi (*Trichophyton mentagrophytes*, *Microsporum canis*, *Microsporum gypseum*, *Fonsecae pedrosoi*, *Candida albicans*, *Cryptococcus neoformans*, *Aspergillus fumigatus*, *Aspergillus terreus*, *Rhizopus orizae*, and *Alternaria alterinata*) were kindly supplied by Prof. Kazuko Nishimura and Prof. Makoto Miyaji of the Research Center for Pathogenic Fungi and Microbial Toxicossis, Chiba University. The fungi were cultured on potato dextrose agar or glucose Trypto-Soya broth yeast extract agar for 1–2 weeks, and fixed according to the procedure described above. Other cells were kindly supplied as follow: Slime mold (*Dictyostelium discoideum*); Prof. Saburo Uchiyama of Dokkyo University, Tetrahymena; Prof. Koei Hamana of Gunma University, Blue-green alga (Cyanobacterium, *Chroococci-diopsis*) (Hayashi et al., 1995); Prof. Sunao Yamazaki of Tokyo University.

Amino acid analysis

Cells were homogenized at least twice with a Physcotron (a microblender consisting of metal cutters) NITI-ON I Rikagaku Seisakusho (Tokyo, Japan) at 20,000 rpm for 20 sec in water. One aliquot (50 μl) was used for the amino acid analysis. Cell homogenates dried *in vacuo* were hydrolyzed in 6 N HCl at 105°C for 24 h. The hydrolyzate was applied to a HITACHI L-8500A amino acid analyzer (Okayasu et al., 1997). As glutamine was converted to glutamic acid during acidic hydrolysis, the glutamic acid values presented in this study represent the sum of both amino acids. Similarly, as asparagine was converted to asparatic acid, the asparatic acid values represent the sum of asparagine and aspartic acid.

Arginine is essential for sustained growth of various cell lines. Although arginine is synthesized from citrulline in most cultured cells, ornithine, the immediate precursor of citrulline, can not substitute for arginine in culture media (Niwa et al., 1980; Yamamoto and Niwa, 1993). However, rat hepatoma R-Y121B cells have ornithine carbamoyltransferase activity and are able to grow in culture medium containing ornithine instead of arginine (Niwa et al., 1980). The cellular amino acid composition was identical for liver cells and R-Y121B cells cultured in a modified EMEM in which arginine was replaced with ornithine (Okayasu et al., 1997). In addition, heparin inhibits collagen fiber formation in rat hepatoma cells, however, the cellular amino acid composition, with the exception of hydroxyproline was identical for heparin-treated cells and heparin-untreated cells (Akimoto et al., 1997). We also showed that the cellular amino acid composition was almost identical for rat primary hepatocytes, liver cell lines and transformed cells (Okayasu et al., 1997). These results show that the present method cannot detect changes in the cellular amino acid composition due to cell transformation, environmental differences or drug effects. However, the present method can detect the great changes in amino acid composition which result from biological evolution.

Results

Amino acid analysis of procaryotic cells

In our previous experiments, we found that the amino acid composition was almost identical for cell homogenates and ethanol precipitated cell homogenates (Okayasu et al., 1997). It appears that the contribution of free amino acids to the cellular amino acid composition is small. Therefore, we used cell homogenates without protein precipitation in the present study.

Although the biochemical characteristics of Gram-negative bacteria, *E. coli*, *V. alginolyticus*, *Proteus* and *Klebsiela* differ, their amino acid ratios were almost identical, as shown in Table 1. The standard deviations were less than 1%, with the exception of methionine and valine which slightly exceeded this limit. When these data were expressed graphically, the pattern for each strain showed a good correction, as shown in Fig. 1. These results indicate that the present analytical method has an excellent level of accuracy and reproducibility. Thus, the HITACHI L8500A amino acid analyzer was used throughout the present study.

S. aureus and *B. subtilis*, which belong to the Gram-positive bacteria, were also examined. The amino acid compositions of these bacteria differed from *E. coli*, as shown in Table 1 and Figs. 2A and 2B.

The procaryote blue-green alga (Cyanobacterium) was examined. The amino acid composition is shown in Table 1 and Fig. 2C. A small difference in amino acid composition was observed between Cynobacterium and *E. coli*. The concentrations of glycine, alanine and arginine were greater than those of *E. coli*, whereas lysine was present at a lower level.

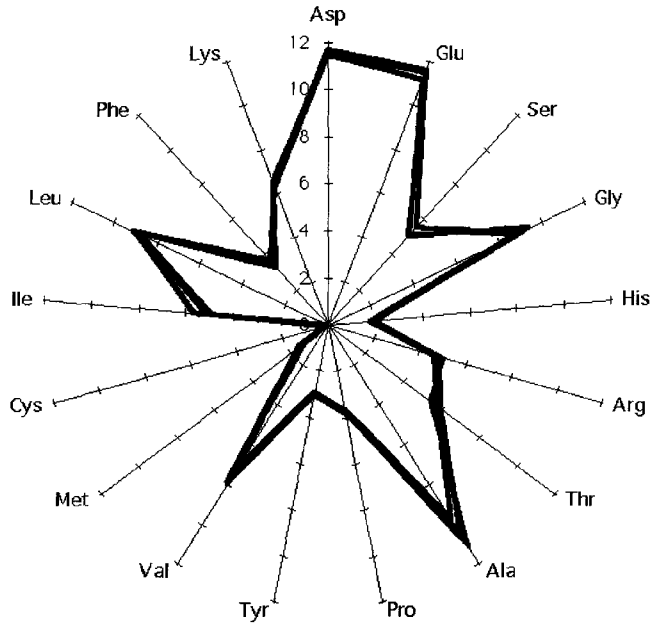

Fig. 1. Radar graphs of the amino acid compositions of Gram-negative bacteria. The graph was drawn using the mean values of *E. coli*, *Klebsiella*, *V. alginolyticus* and *Proteus*, and each graph was overlaid

Amino acid analysis of eucaryotic cells

Protozoa (*Tetrahymena*), *Chlorella* and slime mold (*Dictyostelium discodium*) were examined as representative primitive eucaryotes. Their amino acid compositions are shown in Table 2. The amino acid composition of *Tetrahymena* differed from that of *E. coli*, as shown in Fig. 2D. The concentration of alanine

Table 1. Amino acid compositions of bacteria and Cyanobacteria

A.A	*E. coli*	*V. alg.*	*Prot.*	*Kleb.*	*S. aur.*	*B. sub.*	Alga
Asp	11.46 ± 0.46	11.40 ± 0.14	11.61 ± 0.44	11.68 ± 0.08	9.23 ± 0.18	9.97 ± 0.44	10.32 ± 0.06
Glu	11.12 ± 0.19	11.49 ± 0.12	11.44 ± 0.15	11.58 ± 0.13	11.73 ± 0.17	11.35 ± 0.22	11.64 ± 0.07
Ser	5.16 ± 0.08	5.05 ± 0.04	5.55 ± 0.05	5.57 ± 0.03	5.25 ± 0.07	4.89 ± 0.03	5.32 ± 0.30
Gly	8.96 ± 0.24	9.39 ± 0.15	8.68 ± 0.23	9.13 ± 0.13	17.71 ± 0.58	8.01 ± 0.30	10.44 ± 0.09
His	1.90 ± 0.02	1.78 ± 0.01	1.77 ± 0.03	1.87 ± 0.05	1.48 ± 0.07	1.68 ± 0.04	1.18 ± 0.05
Arg	4.91 ± 0.09	4.88 ± 0.04	4.62 ± 0.02	5.02 ± 0.06	2.98 ± 0.06	3.61 ± 0.07	5.69 ± 0.11
Thr	5.96 ± 0.49	5.57 ± 0.47	5.76 ± 0.45	5.33 ± 0.42	4.98 ± 0.33	5.39 ± 0.15	6.24 ± 0.64
Ala	10.54 ± 0.52	10.88 ± 0.12	9.76 ± 0.40	11.09 ± 0.10	11.78 ± 0.23	11.23 ± 0.46	12.16 ± 0.05
Pro	3.84 ± 0.12	3.75 ± 0.11	3.70 ± 0.09	3.78 ± 0.11	2.59 ± 0.09	3.19 ± 0.16	4.17 ± 0.12
Tyr	3.01 ± 0.08	2.93 ± 0.05	3.03 ± 0.01	2.89 ± 0.05	2.38 ± 0.08	2.76 ± 0.10	3.36 ± 0.19
Val	7.97 ± 1.14	7.38 ± 0.31	7.97 ± 0.06	7.35 ± 0.51	6.46 ± 1.89	10.83 ± 1.96	6.60 ± 0.65
Met	1.12 ± 0.13	1.41 ± 1.20	1.49 ± 1.24	1.37 ± 1.19	0.46 ± 0.80	1.86 ± 0.25	1.05 ± 1.06
Cys	0.15 ± 0.13	0.05 ± 0.02	0.10 ± 0.01	0.04 ± 0.04	0.06 ± 0.05	0.19 ± 0.03	0.09 ± 0.09
Ile	5.07 ± 0.14	5.02 ± 0.16	5.66 ± 0.13	4.82 ± 0.15	4.78 ± 0.16	5.21 ± 0.22	5.26 ± 0.20
Leu	8.96 ± 0.32	9.10 ± 0.15	9.01 ± 0.22	8.86 ± 0.15	6.11 ± 0.17	7.34 ± 0.22	9.17 ± 0.10
Phe	3.62 ± 0.07	3.48 ± 0.07	3.52 ± 0.05	3.24 ± 0.08	2.96 ± 0.10	3.40 ± 0.11	3.34 ± 0.08
Lys	6.24 ± 0.16	6.42 ± 0.11	6.33 ± 0.15	6.38 ± 0.18	9.07 ± 0.31	7.10 ± 0.30	3.97 ± 0.04

The value is expressed as the percentage of total amino acids and is the mean ± S.D. of 3 or 4 independent experiments. Number of experiments: *E. coli*; 3, *V. alg.*; 3, *Prot.*, 3, *Kleb.*; 3, *S. aur.*; 4, *B. sub.*; 3, Alga; 4.

Table 2. Cellular amino acid compositions

Amino acid	Slime mold	*Tetrahymena*	*Chlorella*
Asp	11.36 ± 0.30	11.70 ± 0.25	9.73 ± 0.38
Glu	10.83 ± 0.22	12.62 ± 0.23	9.97 ± 0.47
Ser	8.08 ± 0.40	5.79 ± 0.39	4.48 ± 0.70
Gly	10.43 ± 0.16	7.85 ± 0.31	8.79 ± 0.35
His	1.72 ± 0.09	1.87 ± 0.07	3.50 ± 0.08
Arg	4.00 ± 0.11	4.15 ± 0.10	8.77 ± 0.48
Thr	6.97 ± 0.61	5.85 ± 0.56	4.80 ± 0.66
Ala	6.78 ± 0.12	7.28 ± 0.18	11.38 ± 1.21
Pro	4.88 ± 0.26	3.73 ± 0.33	4.60 ± 0.37
Tyr	3.02 ± 0.06	3.72 ± 0.26	2.29 ± 0.44
Val	7.03 ± 1.34	6.40 ± 1.54	5.82 ± 1.20
Met	0.47 ± 0.72	0.88 ± 0.97	0.65 ± 0.63
Cys	0.01 ± 0.02	0.13 ± 0.13	0.09 ± 0.09
Ile	6.22 ± 0.13	6.14 ± 0.33	3.16 ± 0.13
Leu	7.40 ± 0.12	8.64 ± 0.27	7.07 ± 0.14
Phe	3.75 ± 0.05	4.51 ± 0.18	2.86 ± 0.07
Lys	7.05 ± 0.22	8.73 ± 0.27	12.03 ± 0.30

See the legend to Table 1. Number of experiments: slime mold; 3, *Tetrahymena*; 6, *Chlorella*; 6.

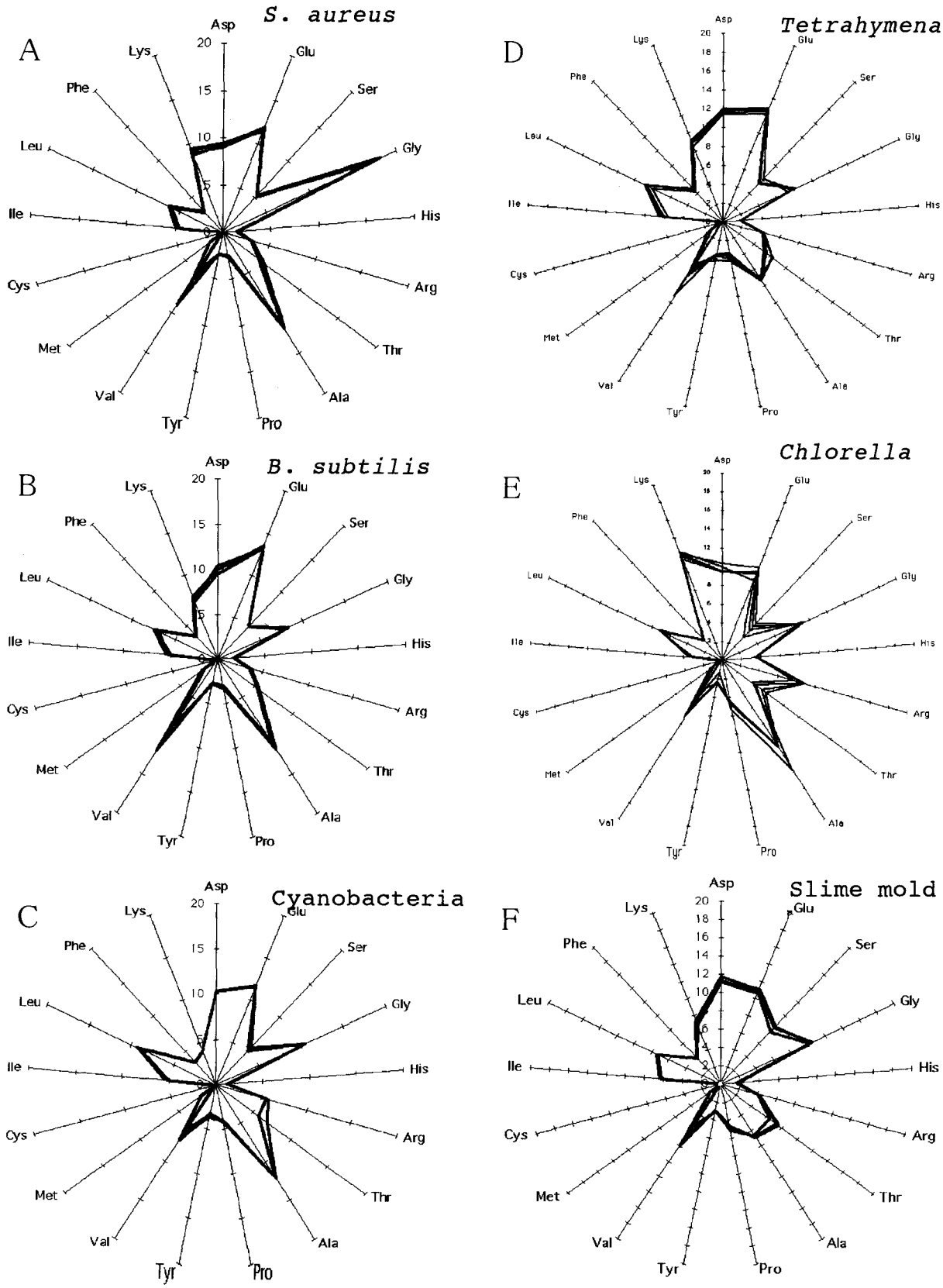

Fig. 2. Radar graphs of the amino acid compositions of various cells. The graph was drawn using each experimental value. **A** *S. aureus*, **B** *B. subtillis*, **C** Cyanobacterium, **D** *Tetrahymena*, **E** *Chlorella*, **F** slime mold

in *Tetrahymena* was less than that of *E. coli*, whereas lysine and glutamic acid were present at greater levels.

The amino acid composition of *Chlorella* proteins differed markedly from that of *E. coli*, as shown in Fig. 2E. In Chlorella, the concentrations of alanine, arginine and lysine were increased compared to those of *E. coli*, whereas levels of asparatic acid, glutamic acid, serine, leucine and isoleucine were significantly reduced.

In slime mold, the concentration of alanine was reduced compared with that of *E. coli*, whereas the concentration of serine had increased, as shown in Fig. 2F.

Fungi are eucaryotic and are often multicellular organisms. Although the fungi used in this study basically consist of a hypha and a spore, the samples used in the present study were not separated into their components. The amino acid compositions of the fungi are shown in Table 3 and are expressed graphically in Fig. 3A. The patterns of amino acid composition differ slightly between species (Fig. 3A), but *Trichophyton mentagrophytes, Microsporum canis*, and *Microsporum gypseum*, which belong to the dermatophytes, have similar amino acid compositions (Fig. 3B).

Amino acid analysis of mammalian cells

In our previous study (Okayasu et al., 1997), a variety of mammalian cells were examined to compare their amino acid compositions. It was concluded that there was no significant difference in amino acid composition, although some small differences were observed. As described above, the present method has a excellent level of accuracy and reproducibility, however, the amino acid compositions of two mammalian cell lines and one fish cell line (golden fish) were analyzed to substantiate the previous conclusion. The amino acid compositions of rat liver cells (M), human urinary bladder carcinoma cells (HUB-15), and golden fish scale cells are shown in Table 4 and Figs. 4A, 4B and 4C, respectively. The amino acid compositions of the three vertebrate cell lines were quite similar. However, in our previous study (Okayasu et al., 1997), the differences shown here were less clear due to experimental errors. The concentration of asparatic acid was lower in the veterbrates than in *E. coli*, but the concentrations of glutamic acid, serine and lysine were much higher in the mammalian cell lines.

Discussion

Prebiotic evolution

In *S. aureus*, the concentrations of glycine and alanine were much higher than in other cells examined in the present study. This may be due to the composition of the *S. aureus* cell wall. The peptidoglycan which constitutes the *S. aureus* cell wall contains a large amounnt of D- and L-alanine in the peptide chain, and glycine in the bridge between the peptide chains.

Table 3. Amino acid compositions of fungi

A.A.	No. 1	No. 2	No. 3	No. 4	No. 5	No. 6	No. 7	No. 8	No. 9	No. 10
Asp	11.10 ± 0.15	10.42 ± 0.28	11.16 ± 0.17	9.72 ± 0.24	10.07 ± 1.06	11.06 ± 0.08	9.69	9.89	10.75	12.13
Glu	11.06 ± 0.08	10.82 ± 0.37	10.54 ± 0.40	10.61 ± 0.33	11.74 ± 1.87	11.23 ± 0.22	9.81	10.98	11.04	11.18
Ser	7.59 ± 0.22	7.35 ± 0.23	7.23 ± 0.13	7.57 ± 0.14	7.69 ± 0.18	7.07 ± 0.04	7.33	7.04	7.14	6.96
Gly	8.68 ± 0.02	10.24 ± 0.44	11.57 ± 0.32	9.89 ± 0.16	9.43 ± 1.11	9.69 ± 0.14	9.22	9.87	9.58	9.24
His	1.88 ± 0.01	2.02 ± 0.02	2.72 ± 0.08	2.04 ± 0.05	1.99 ± 0.03	2.02 ± 0.02	1.97	1.95	1.84	2.76
Arg	3.66 ± 0.05	4.54 ± 0.14	4.23 ± 0.09	4.79 ± 0.18	4.47 ± 0.57	4.99 ± 0.05	4.39	4.61	5.03	4.69
Thr	7.48 ± 0.36	6.27 ± 0.28	6.45 ± 0.21	6.20 ± 0.20	6.49 ± 0.88	6.28 ± 0.21	8.36	6.37	6.48	5.78
Ala	8.83 ± 0.06	9.20 ± 0.24	9.93 ± 0.09	9.89 ± 0.09	8.16 ± 1.17	9.05 ± 0.06	9.40	9.74	9.16	8.31
Pro	4.71 ± 0.20	6.59 ± 0.34	6.95 ± 0.54	5.38 ± 0.26	6.43 ± 0.69	5.68 ± 0.14	6.33	5.96	5.36	4.19
Tyr	3.28 ± 0.03	2.85 ± 0.08	2.58 ± 0.08	3.02 ± 0.19	3.08 ± 0.14	2.73 ± 0.11	2.95	2.87	2.84	3.32
Val	5.90 ± 0.87	6.27 ± 0.73	5.36 ± 0.57	6.82 ± 1.31	5.06 ± 2.46	5.74 ± 0.72	7.82	5.70	5.28	5.68
Met	0.54 ± 0.45	1.00 ± 1.73	0.30 ± 0.52	0.33 ± 0.40	0.58 ± 0.79	0.76 ± 0.48	0.85	0.89	1.08	1.72
Cys	0.01 ± 0.02	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.08 ± 0.13	0.05	0.00	0.36	0.00
Ile	5.56 ± 0.06	4.98 ± 0.18	4.34 ± 0.15	5.01 ± 0.16	5.01 ± 0.22	4.89 ± 0.14	4.61	4.41	4.87	4.90
Leu	8.03 ± 0.05	7.97 ± 0.16	7.63 ± 0.11	8.59 ± 0.16	9.56 ± 0.38	8.35 ± 0.48	7.78	9.40	8.50	7.82
Phe	3.91 ± 0.08	3.47 ± 0.18	3.27 ± 0.07	3.76 ± 0.17	4.44 ± 0.31	3.54 ± 0.18	3.75	3.52	3.70	3.55
Lys	7.77 ± 0.13	6.01 ± 0.25	5.75 ± 0.15	6.38 ± 0.19	5.78 ± 0.24	6.84 ± 0.42	5.69	6.80	6.99	7.77

See the legends to Table 1. The name of species is shown and the number of experiments is shown in brackets: 1; *C. albicans* (3), No. 2; *A. fumigatus* (3), No. 3; *A. terreus* (3), No. 4; *C. neoformance* (3), No. 5; *M. canis* (3), No. 6; *M. gypseum* (3), No. 7; *F. pedrosoi* (2), No. 8; *A. alterinata* (2), No. 9; *T. mentagrophytes* (2), No. 10; *R. orizae* (1).

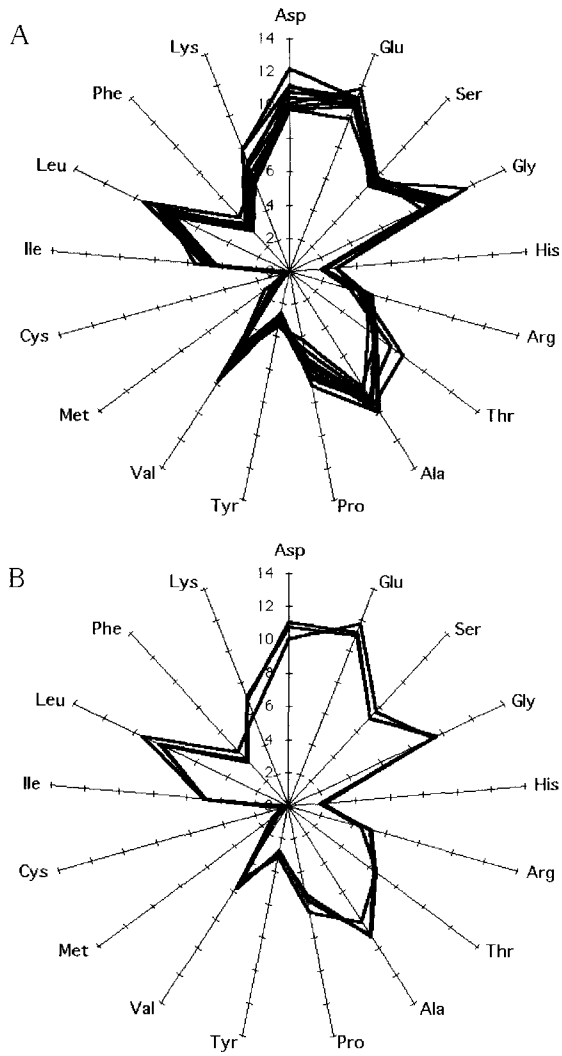

Fig. 3. Radar graphs of the amino acid compositions of fungi. The graph was drawn using the mean values of 10 species of fungi (**A**), and the dermatophytes(**B**)

Electric discharges induced the formation of glycine, alanine and asparatic acid in a gas designed to simulate the atmosphere of the primitive Earth (Miller, 1955; Miller and Urey, 1959). In addition, these amino acids have been identified in meteorites (Kvenvolden et al., 1970; Wolman et al., 1972). The monomer composition of a pre-organism constituted from the polymer compounds formed by chemical reactions may reflect the monomer concentrations present on primitive Earth, because polymerization was also chemically induced. If *S. aureus*, which may be an ancient form of life, uses amino acids which were abundant on primitive Earth to constitute its body, the relative concentrations of glycine, alanine and aspartic acid might be higher compared to the other amino acids.

Table 4. Amino acid composition of vertebrate cells

Amino acid	Fish	Rat	Human
Asp	9.53 ± 0.13	8.77 ± 0.57	9.61 ± 0.02
Glu	14.41 ± 0.18	12.35 ± 0.25	12.60 ± 0.06
Ser	6.03 ± 0.41	6.59 ± 0.19	6.89 ± 0.05
Gly	9.29 ± 0.09	11.16 ± 0.51	8.99 ± 0.02
His	2.03 ± 0.06	1.85 ± 0.03	2.02 ± 0.03
Arg	4.76 ± 0.05	5.62 ± 0.26	5.87 ± 0.05
Thr	5.10 ± 0.61	5.71 ± 0.38	5.89 ± 0.04
Ala	9.28 ± 0.09	9.02 ± 0.17	8.20 ± 0.02
Pro	4.27 ± 0.15	5.27 ± 0.11	5.02 ± 0.05
Tyr	2.90 ± 0.14	2.58 ± 0.29	3.08 ± 0.05
Val	6.55 ± 0.33	5.86 ± 0.10	6.20 ± 0.15
Met	1.45 ± 1.19	0.48 ± 0.69	0.12 ± 0.02
Cys	0.34 ± 0.20	0.13 ± 0.23	0.07 ± 0.03
Ile	4.62 ± 0.17	4.46 ± 0.11	4.57 ± 0.08
Leu	8.42 ± 0.05	9.04 ± 0.43	9.05 ± 0.06
Phe	3.44 ± 0.04	3.68 ± 0.17	3.68 ± 0.01
Lys	7.58 ± 0.16	7.42 ± 0.07	8.14 ± 0.10

See legends to Table 1. The name of the cells used and the number of experiments: golden fish scale cells (5), rat; M cells (3), human; HUB-15 cells (3).

Codon

It is considered that life appeared on Earth via a pre-organism formed by prebiotic evolution. During this period of prebiotic evolution, both polypeptides and polynucleotides were probably formed independently by chemical reactions with reaction rates determined by the monomer concentration. However, it is not clear whether protein formation preceded codon formation, or whether codon formation preceded protein formation, although the establishment of the codon was necessary for the appearance of life. There are many reports regarding codon frequency (Grosjean and Fiers, 1982; Henzen et al., 1981; Wain-Hobson et al., 1981; Sharp et al., 1988; Alff-Steinberger, 1987), and it is clear that the probability of codon usage differs not only in a single gene but also in different genes. Since these results were obtained from the statistical analyses of both protein amino acid sequences and nucleic acid sequences, we could not determine whether codon formation preceded protein formation or vice versa. If codon formation preceded protein formation during the establishment of life, the probability of amino acid usage will depend on the codon numbers involved in the chemical reaction. This is possible, because the cellular amino acid composition represents the variety of proteins contained within the whole cell, and the amino acid composition reflects codon usage, as shown in Fig. 5. On the other hand, the amino acid composition of the various cells (Figs. 1–4) differ from the amino acid composition determined from the codon data. Thus, codon forma-

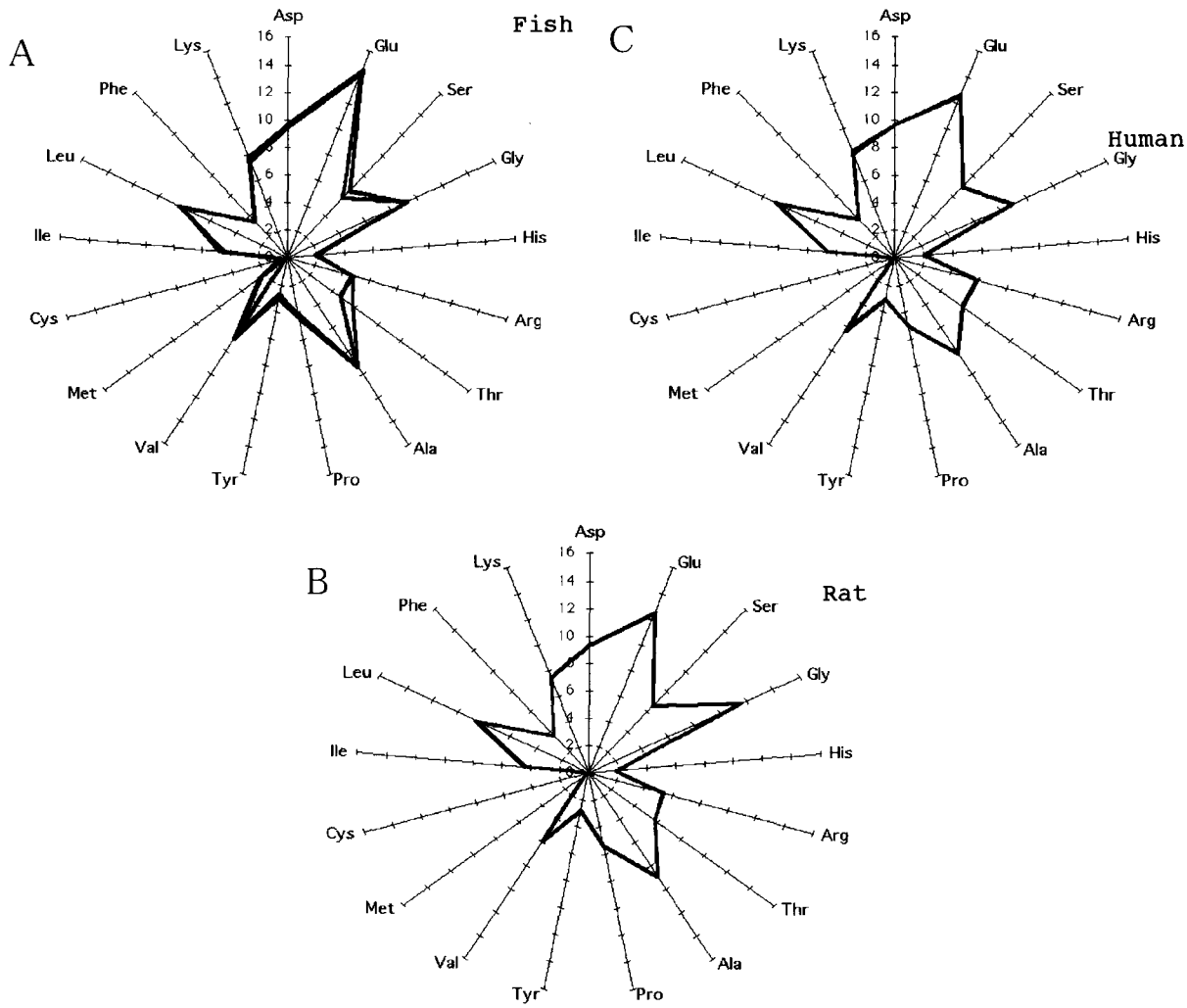

Fig. 4. Radar graphs of the amino acid compositions of vertebrates. The graph was drawn using each experimental value. **A** Golden fish scale cells, **B** rat M cells, **C** human HUB-15 cells

tion might not have preceded protein formation when life appeared on the Earth.

After the appearance of primitive life, protein synthesis was carried out in a hereditable manner according to the DNA sequences which had been determined in chemical reactions. In this primitive type of life form, the early proteins were basically synthesized without the linkage of function, however, the probability of amino acid usage was regulated by the amino acid composition available on Earth rather than the codon numbers. Based on this hypothesis, genome formation might have been strongly influenced by the selective force of cooperation between protein amino acid composition and function, and the cellular amino acid composition which resulted from the primitive amino acid composition is reflected even now.

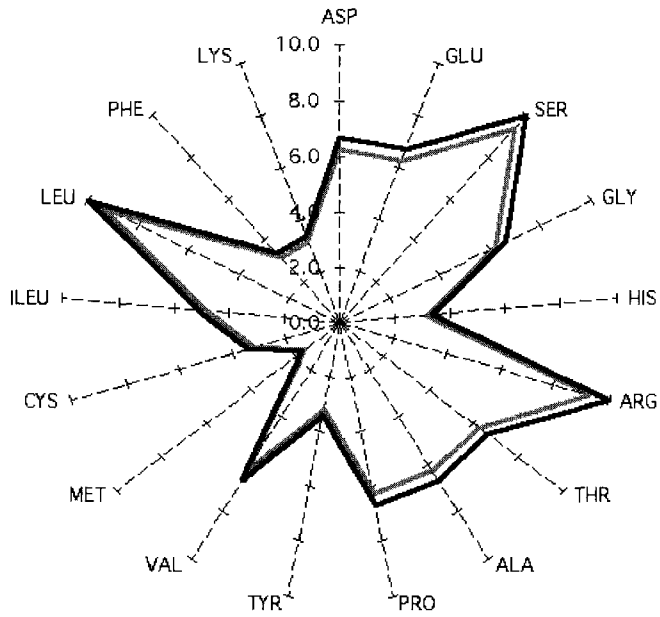

Fig. 5. Amino acid composition based on codon number. The number for each amino acid was divided by the total codon number, 64, (*black*), however tryptophan and stop codons were excluded from the total codon number (*gray*)

UV light

On the primitive Earth, strong ultraviolet irradiation induced the decomposition of tyrosine, phenylalanine and tryptophan, which are sensitive to near ultraviolet light. This resulted in low concentrations of these amino acids during polypeptide formation, and thus resulting in low concentrations of these amino acids in the polypeptides formed. Therefore, the low concentrations of tyrosine and phenylalanine (Figs. 1–4) in the whole cell amino acid composition might reflect their low concentrations on the primitive Earth. In addition, since tryptophan is decomposed during the acidic hydrolysis used in the present study, we are not able to comment on the frequency of this amino acid. However, the percentage of tryptophan calculated from codon frequency was 1.0% of total codon frequency, based on 35 complete human genes (Alff-Steinberger, 1987). This also suggests that the concentration of tryptophan was extremely low on the primitive Earth.

Polarity

On the other hand, when the amino acid compositions of the various cells are expressed by radar charts, nonpolar amino acids such as glycine, alanine, valine, leucine and isoleucine, produced peaks in all samples (Figs. 1–4). (The order of amino acid presentation is based on their elution order by the high performance liquid chromatography used in our previous study) (Okayasu et al., 1997). Polypeptides consisting of a large amount of nonpolar amino

acids may be better suited to assembly by hydrophobic force in the primitive ocean in which prebiotic evolution proceeded and where the concentration of organic compounds may have been extremely low. Thus, polypeptides abundant in nonpolar amino acids could contribute to the formation of a preorganism capable of collecting materials present at low concentration, and may have contributed to the appearance of life. However, charged amino acids such as glutamic acid, asparatic acid, lysine, histidine, and arginine are important hydrophilic amino acids possessing, basic or acidic characters, because the chemical reaction might then be carried out in solution. In addition, uncharged polar amino acids such as asparagine, glutamine, threonine, and tyrosine are required to make certain functional polypeptides. Therefore, increased concentrations of lysine and arginine are found in *Chlorella*, while the concentrations of valine, isoleucine and leucine are low (Fig. 6). During the evolutionary process which produced *Tetrahymena* and the vertebrates, the concentrations of lysine and glutamic acid increased, while that of alanine decreased compared to *E. coli.*

Biological evolution

Biological evolution has proceded three-dimensinally in the evolutional tree. If the analogy of a tree is used, it is possible to assume that the comparison of the 17 amino acids may be linked to the three-dimensional spreading of the branches, some going in the same direction but in different planes. If the amino acid composition patterns of samples compared with *E. coli* are similar, the evolution of the two species has progressed in the same evolutionary direction (Fig. 6A). The amino acid composition patterns of *S. aureus* and *B. subtillis* differed from each other (Figs. 2A and 2B), and their amino acid patterns were also different from *E. coli*, as shown in Fig. 6A. However, if the differences in the concentrations of glycine and valine compared with *E. coli* are excluded, differences in the other amino acid concentrations are similar for these two species. This may be due to the fact that both *S. aureus* and *B. subtilis* belong to the Gram-positive bacteria. Of course, the amino acid patterns of *S. aureus* and *B. subtilis* differed from those of other samples, as shown in Fig. 6A.

Cyanobacterium differed from cells such as protozoa, slime mold, *Chlorella*, fungi and vertebrates. The fact that the concentration of lysine was much lower than in *E. coli* was observed only in Cyanobacterium (Fig. 6A). This shows that Cyanobacterium differs from the other species used in the present study. In addition, the fact that the amino acid composition of blue-green alga is similar to that of *E. coli* means that this procaryote has not diverged from *E. coli* to any great degree in evolutionary terms.

The amino acid pattern of *Chlorella* is markedly different from the other cells examined in the present study (Fig. 6A). Thus, we consider that *Chlorella* evolved independently.

Comparison of the amino acid pattern of *Tetrahymena* with *E. coli* gives results similar to those of slime mold, fungi and vertebrates such as gold fish,

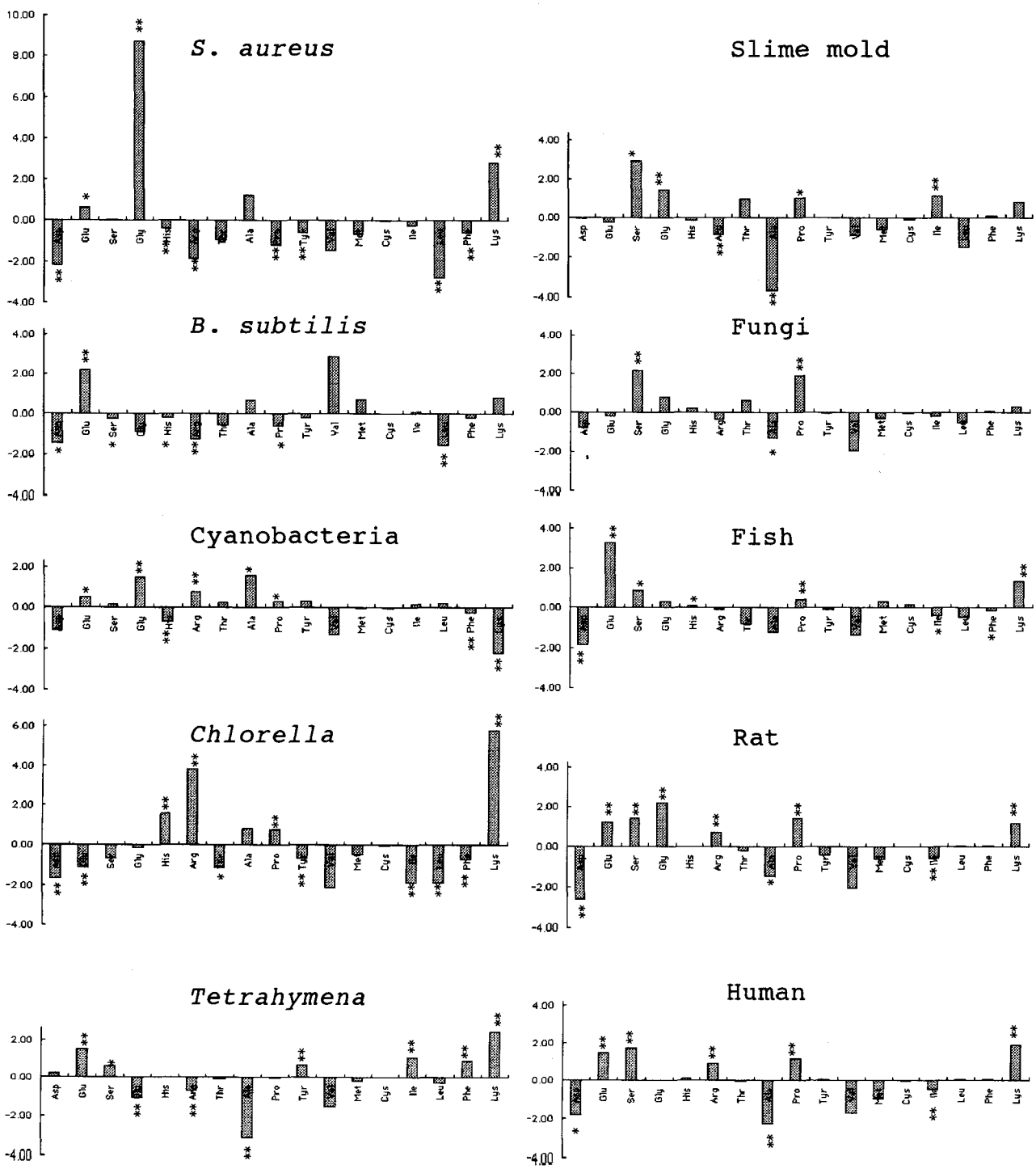

Fig. 6. Amino acid compositions of sample cells compared with *E. coli* (**A**), and changes in cellular Ala, Val, Ile, Leu, Pro or Ser concentration in biological evolution (**B**). The mean values of *E. coli* were substracted from those of the sample cells. Statistical differences between *E. coli* and other samples were evaluated using Student's *t*-test (* $P < 0.01$, ** $P < 0.005$)

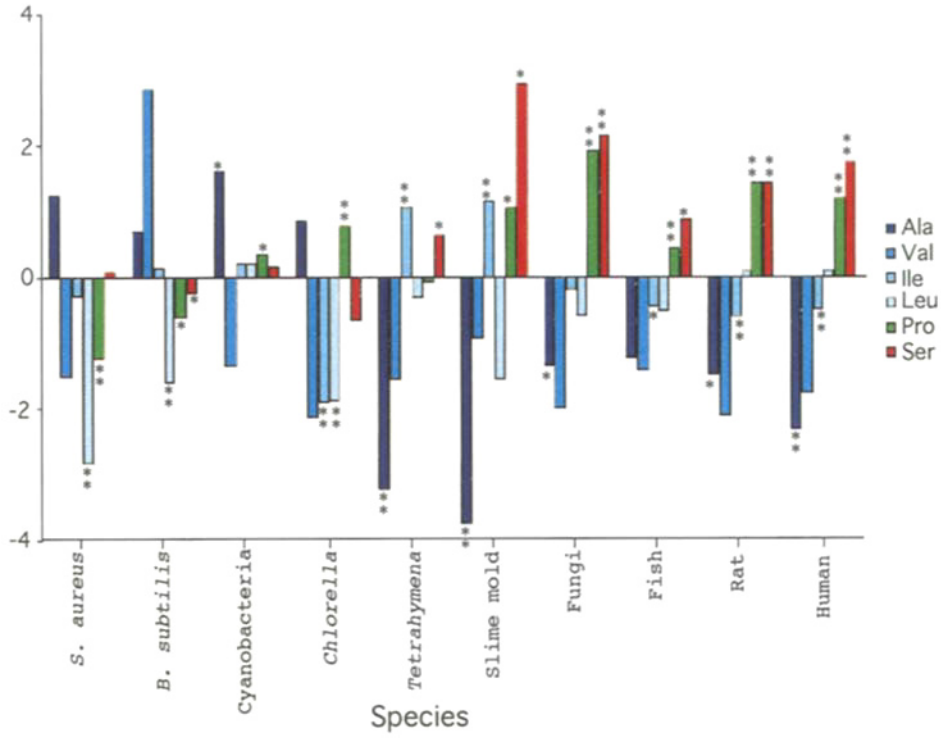

Fig. 6B

rat and human cells, as shown in Fig. 6A. These results based on the total-cellular amino acid composition suggest that they progressed in the same direction for a long period, and then separated into different directions. These conclusions are consistent with data obtained from cytochrome C (Dayhoff and Schwartz, 1981; Dayhoff et al., 1972) and r-RNA sequence analyses (Sogin et al., 1986).

Six amino acids (Ala, Val, Ile, Leu, Pro and Ser) are considered to be important in biological evolution based on data shown in Fig. 6B. In general, the concentrations of all or some of nonpolar amino acids such as alanine, valine, leucine and isoleucine decreased as biological evolution progressed, while the concentrations of proline, which is also an nonpolar amino acid, and serine increased (Fig. 6B). The concentration of proline is increased in many organisms which are more evolved than *E. coli*, and it is slightly increased in Cyanobacteria. On the other hand, proline concentration has increased in the Gram-positive bacteria and is almost constant in *Tetrahymena* (Fig. 6B). The increase in proline concentration eventually enabled the formation of collagen, which is an important component constituting extracellular matrices in multicellular organisms. Serine is present in the active site of many enzymes, or proteins in which chemical modifications take place. Thus, the increase in serine concentration suggests that serine has contributed to the formation of functional polypeptides during biological evolution, while decreases in alanine and valine concentration occurred. On the other hand, an inverse

relationship is observed in the Gram-positive bacteria, which are less evolved than *E. coli*. Cyanobacteria, *Chlorella* and *Tetrahymena* differ from the former two groups, although slime mold is close to fungi and the vertebrates. Thus, all organisms examined are classified into the following four groups and they have evolved in the following order; Gram-positive bacteria, Gram-negative bacteria, primitive eucaryotic organisms (Cyanobacteria, *Chlorella* and *Tetrahymena*), multicellular eucaryotic organisms, although in evolutional terms, Cyanobacteria and *Tetrahymena* are close to *E. coli* and to slime mold, respectively. It is possible to conclude that changes in the cellular amino acid composition reflect biological evolution.

Comparison of cellular amino acid composition with composition based on protein sequences

In the present and previous studies, we assumed that the cellular amino acid composition closely represents the amino acid composition of the total proteins contained within the whole cell. To validate this assumption, the amino acid compositions obtained from the reported amino acid sequences of certain proteins have been used to make the radar graphs shown in Fig. 7. The simple combination of only 4 *E. coli* proteins such as lactose operon repressor (Dayhoff, 1976), L-asparaginase (Maita et al., 1974), β-galactosidase (Dayhoff, 1978; Fowler and Zabin, 1977) and L-arabinose binding protein (Hogg and Hermodson, 1977) (only these four proteins have been listed as *E. coli* proteins in The Biochemical Data Book I, Japan Biochemistry Society, ed., Tokyo Kagaku Dojin, Tokyo Japan, 1979) showed a similar amino acid composition pattern to that obtained for *E. coli* (Fig. 7A).

In human cells, when the data obtained from 35 complete human genes (Alff-Steinberger, 1987) was used to construct the radar graph, the amino acid composition was similar to that of human HUB-15 cells, as shown in Fig. 7B. Thus, the amino acid composition of whole cells is a close representation of the amino acid composition of total cellular proteins.

Histones are important proteins which form chromatin structures in eucaryotic cells. Therefore, it is interesting to examine the amino acid composition found in the histones, and look for similarity with the cellular amino acid composition. Four calf thymus histones H2A (Dayhoff, 1972a), H2B (Dayhoff, 1972a), H3 (Dayhoff, 1973b) and H4 (Dayhoff, 1973b) (only these four histones have been listed as histones in The Biochemical Data Book I, 1979) were examined. Using the average of the 4 histone amino acid compositions, a radar graph very similar to that of *Chlorella*, with the exception of asparatic acid concentration, was constructed, as shown in Fig. 7C. In addition, recent studies have shown that the amino acid sequences of the *Drosophila* heat shock proteins resemble that of human α-crystallin (Ingolia and Craig, 1982), and that a α-crystallin has molecular chaperone activity as an additional function (Horwitz, 1992; Muchowski et al., 1997). Thus, it seems that existing proteins might have been used to establish primitive life forms, unlike the newly synthesized proteins which are produced for certain functions.

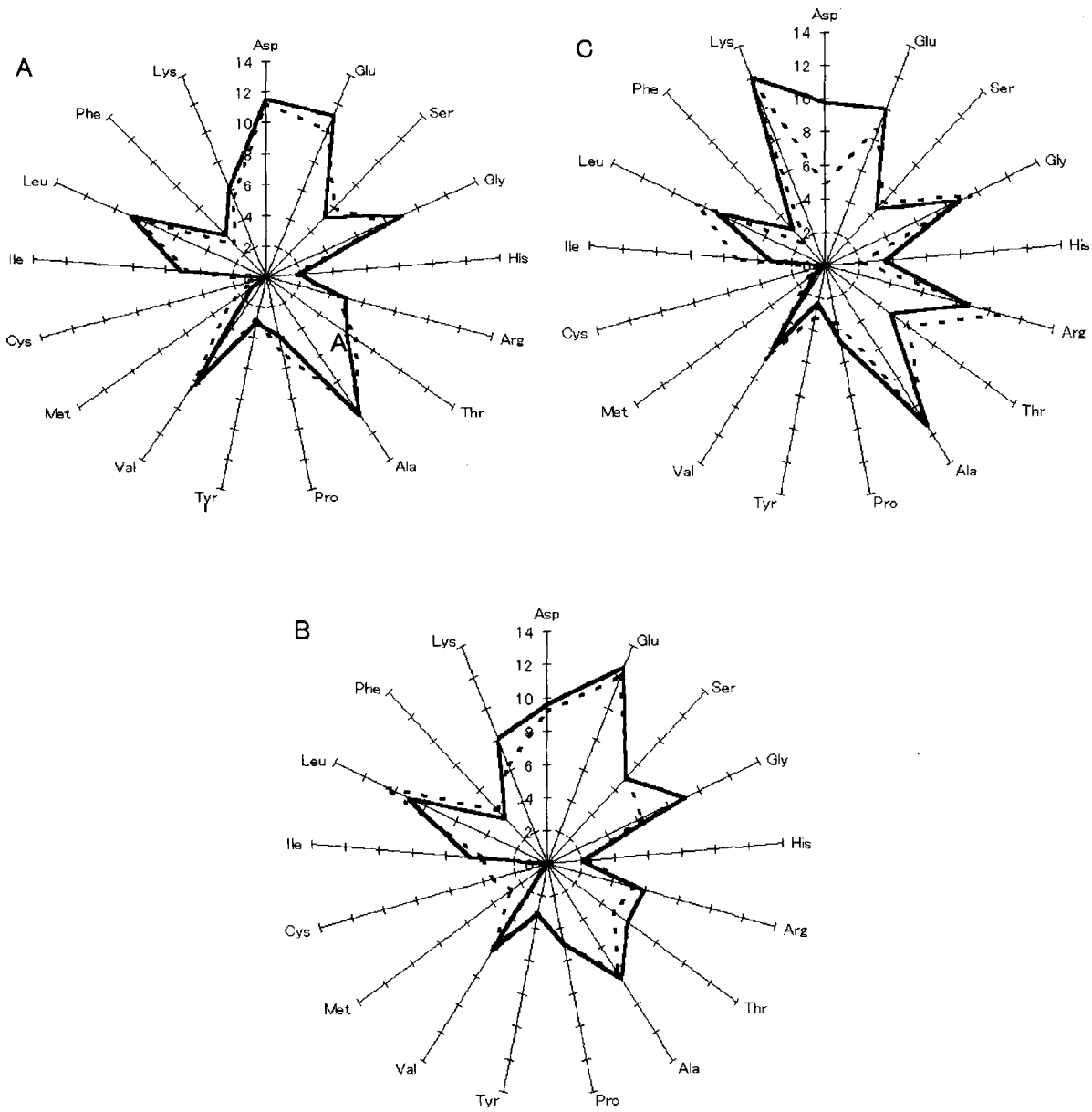

Fig. 7. Comparison of radar graphs of the cellular amino acid composition and protein amino acid compositions. **A** *E. coli*, **B** human HUB-15 cells, **C** *Chlorella* and calf thymus histones. The dotted lines represent the average amino acid compositions due to proteins or codons

Primitive life

The patterns of cellular amino acid composition differed among the different species, with the exception of the Gram-negative bacteria (Figs. 1–4). However, if these patterns were overlaid on each other, their shapes are basically similar, as shown in Fig. 8, even though the genome numbers differ among the various species. It was reported that the amino acid compositions of alanine-

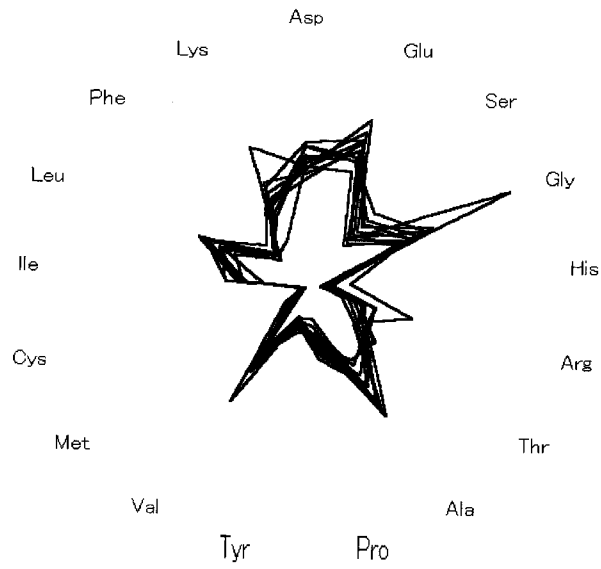

Fig. 8. Radar graphs of the amino acid compositions of all cells used in the present study

rich acidic ribosomal proteins are similar in procaryotes and eucaryotes (Visentin and Yaguchi, 1979). During prebiotic evolution, the monomer compositions of polymers might reflect their free monomer concentrations which were initially used in chemical reactions. Therefore, the present data suggests that the amino acid composition on the primitive Earth may be represented by the amino acid composition of the original life form which appeared there, and which in turn developed from the polypeptides formed during the prebiotic evolution period. Various microorganisms have been observed as fossils in Precambrian sedimentary rocks (Barghoon and Schope, 1966; Peat and Lloyd, 1974; Macgregor et al., 1974; Nagy and Zumberge, 1976). This fact shows that microorganisms are close in evolutionary terms to a more ancient form of life. It is reasonable to assume that the original pattern of amino acid composition seen in primitive life forms might have been preserved up to the present. Thus, the cellular amino acid composition, which is mainly based on the cellular proteins, of the various cells examined may reflect the progress, not only of biological but also of prebiotic evolution.

Acknowledgements

The author is greatly indebted to Dr. Masashi Ikeda, Dr. Teiji Okayasu and Dr. Kazumi Akimoto of this University for amino acid analysis, data analysis and vertebrate cell culture, respectively. Thanks are due to Prof. Hiroto Naora of the Australian National University for helpful discussion, and to Dr. Elizabeth Anne Kamei of Gunma University for revising this manuscript.

References

Akimoto K, Ikeda M, Sorimachi K (1997) Inhibitory effect of heparin on collagen fiber formation in hepatic cells in culture. Cell Struc Funct 22: 533–538

Alff-Steinberger C (1987) Codon usage in *Homo sapiens*: evidence for a coding pattern on the non-coding strand and evolutionary implications of dinucleotide discrimination. J Theor Biol 124: 89–95

Barghoon ES, Schope JW (1966) Microorganisms three billion years old from the Precambrian of South Africa. Science 152: 758–763

Dayhoff MO (1972a) Atlas of protein sequence and structure, vol 5. National Biomedical Foundation, Washington D.C., p D-279

Dayhoff MO (1972b) Atlas of protein sequence and structure, vol 5. National Biomedical Foundation, Washington D.C., p D-280

Dayhoff MO (1973a) Atlas of protein sequence and structure, vol 5. National Biomedical Foundation, Washington D.C., p S-68 [Suppl 1]

Dayhoff MO (1973b) Atlas of protein sequence and structure, vol 5. National Biomedical Foundation, Washington D.C., p S-69 [Suppl 1]

Dayhoff MO (1976) Atlas of protein sequence and structure, vol 5. National Biomedical Foundation, Washington D.C., p 229 [Suppl 2]

Dayhoff MO (1978) Atlas of protein sequence and structure, vol 5. National Biomedical Foundation, Washington D.C., p 123 [Suppl 3]

Dayhoff MO, Park CM, McLaughlin PJ (1972) Bilding a phylogenetic tree: cytochrome C. In: Atlas of protein sequence and structure, vol 5. National Biomedical Foundation, Washington D.C., pp 7–16

Dayhoff MO, Schwartz RM (1981) Evidence on the origin of eukaryotic mitochondria from protein and nucleic acid sequences. Ann N Y Acad Sciences, pp 92–104

Fowler AV, Zabin I (1977) The amino acid sequence of β-galactosidase of *Escherichia coli*. Proc Natl Acad Sci USA 74: 1507–1510

Grosjean H, Fiers W (1982) Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in effiently expressed genes. Gene 18: 199–209

Hayashi N, Ishida T, Peerapornpisal Y, Igarashi Y, Kodama T (1995) Effect of carbon dioxide concentration on the growth and Rubis CO activity of a thermophilic cyanobacteria, *Chroococcidiopsis*. sp. strain TS-821. J Ferment Bioeng 80: 507–509

Henzen D, Chevallier A, Garel JP (1981) Differential usage of iso- accepting tRNASer species in silk glands of Bombyx mori. Nature 290: 267–269

Hogg WW, Hermodson MA (1977) Amino acid sequence of the L-arabinose-binding protein from *Esherichia coli* B/r. J Biol Chem 252: 5135–5141

Horwitz J (1992) α-Crystallin can function as a molecular chaperone. Proc Natl Acad Sci USA 89: 10449–10453

Ingolia TD, Craig EA (1982) Four small *Drosophila* heat shock proteins are related to each other and to mammalian α-crystallin. Proc Natl Acad Sci USA 79: 2360–2364

Kakuya H, Yamada T, Yokokawa M, Ueda T (1983) Establishment of cell strains from human urothelial carcinoma and their morphological characterization. In Vitro 19: 591–599

Katsuta H, Takaoka T (1968) Cytobiological transformation of normal rat liver cells by treatment with 4-dimethylamino azobenzene after Nagisa culture. In: Katsuta H (ed) Cancer cells in culture. University of Tokyo Press, Tokyo, University Park Press, Baltimore and State College, Pennsylvania, pp 321–334

Katsuta H, Takaoka T (1976) Improved synthetic media suitable for tissue culture of various mammalian cells. In: Prtescott DM (ed) Methods in cell biology, vol 14. Academic Press, New York San Francisco London, pp 145–159

Kvenvolden K, Lawless J, Pering K, Peterson E, Flores J, Ponnamperumsa C, Kaplan IR, Moore C (1970) Evidence for extraterrestrial amino acids and hydrocarbons in the Murchison meteorite. Nature 228: 923–926

Lahav N, White D, Chang S (1978) Peptide formation in the prebiotic era: thermal condensation of glycine in fluctuating clay environments. Science 201: 67–69

MacGregor IM, Truswell JF, Eriksson KA (1974) Filamentous algae from the 2,300 m.y. old Transvaal Dolomite. Nature 247: 538–540

Maita T, Morokuma K, Matsuda G (1974) Amino acid sequence of L-asparaginase from *Escherichia coli*. J Biochem 76: 1351–1354

McLaughlin PJ, Dayhoff MO (1972) Evolution of species and proteins: a time scale. In: Atlas of protein sequence and structure, vol 5. National Biomedical Foundation, Washington D.C., pp 47–52

Miller SL (1955) Production of some organic compounds under possible primitive Earth conditions. J Am Chem Soc 77: 2351–2361

Miller SL, Urey HC (1959) Organic compound synthesis on the primitive Earth. Science 130: 245–251

Muchowski PJ, Bassuk JA, Lubsen NH, Clark JJ (1997) Human αB-crystallin. Small heat shock protein and molecular chaperone. J Biol Chem 272: 2578–2582

Nagy LA, Zumberge JE (1976) Fossil microorganisms from the approximately 2,800 to 2,500 million-year-old Bulawayan stromatolite: application of ultramicrochemical analyses. Proc Natl Acad Sci USA 73: 2973–2976

Niwa A, Yamamoto K, Sorimachi K, Yasumura Y (1980) Continuous culture of Reuber hepatoma cells in serum-free, arginine-, gultamine- and tyrosine-deprived chemically defined medium. In Vitro 16: 987–993

Okayasu T, Ikeda M, Akimoto K, Sorimachi K (1997) The amino acid composition of mammalian and bacterial cells. Amino Acids 13: 379–391

Peat CJ, Lloyd BJ (1974) Direct observation of rock macerates. Nature 251: 294–295

Sharp PM, Cowe E, Higgins DG, Shields DC, Wolfe KH, Wright F (1988) Codon usage patterns in *Escherichia coli*, *Bacillus subtillis*, *Saccharomyces pombe*, *Drosophila melanogaster* and *Homo sapiens*; a review of the considerable within-species diversity. Nucleic Acids Res 16: 8207–8211

Sogin ML, Elwood HJ, Gunderson JH (1986) Evolutionary diversity of eukaryotic small subunit rRNA genes. Proc Natl Acad Sci USA 83: 1383–1387

Visentin LP, Yaguchi M (1979) Structural homologies in alanine-rich acidic ribosomal proteins from procaryotes and eucaryotes. Can J Biochem 57: 719–726

Wain-Hobson S, Nussinov R, Brown RJ, Sussman JL (1981) Preferential codon usage in genes. Gene 13: 355–364

Wolman Y, Haverland W, Miller SL (1972) Nonprotein amino acids from spark discharges and their comparison with the Murchison meteorite amino acids. Proc Natl Acad Sci USA 69: 809–811

Yamamoto K, Niwa A (1993) Amino acid and vitamin requirements in mammalian cultured cells (Review Article). Amino Acids 5: 1–16

Author's address: Dr. Kenji Sorimachi, Department of Microbiology, Dokkyo University School of Medicine, Mibu, Tochigi 321-0293, Japan.

Received June 13, 1998

Verleger: Springer-Verlag KG, Sachsenplatz 4–6, A-1201 Wien. – Herausgeber: Prof. Dr. Gert Lubec, Universitäts-Kinderklinik, Währinger Gürtel 18–20, A-1090 Wien. – Redaktion: Sachsenplatz 4–6, A-1201 Wien. – Hersteller: Satz- und Umbruch: Vogel Medien GmbH, A-2102 Bisamberg; Offsetdruck: Novographic, Ing. W. Schmid, Maurer-Lange-Gasse 64, A-1238 Wien. – Verlagsort: Wien – Herstellungsort: Wien. – Printed in Austria.