ON THE TRENDS IN PROTEIN MOLECULAR EVOLUTION: AMINO ACID COMPOSITION

ORLIN CH. IVANOV

Institute of Organic Chemistry with Centre of Phytochemistry, Bulgarian Academy of Sciences, Sofia 1040, Bulgaria

(Received 31 August, 1988)

Abstract. Data on the amino acid composition of proteins having various functions from organisms representing different evolutionary levels (83 superfamilies) are used in order to elucidate the trends in protein molecular evolution. The interconnections evolutionary rate (rate of mutation acceptance) – amino acid composition, and evolutionary level of the organism – amino acid composition (in case of proteins of the same or very similar function) are studied. The amino acid compositions of proteins performing jointly an evolutionarily old functions are also juxtaposed. The mean contemporary protein composition is used as a basis for comparison. The obtained results are evidence in favour of the existence of a trend for an increase of the special amino acids (Met, Ile, Gln, His, Lys, Asn, Phe, Tyr, Trp, Cys) at the expense of the usual ones (Thr, Pro, Ala, Ser, Arg, Gly, Leu, Val, Glu, Asp). The tests of statistical significance of the obtained results (comparison of the mean compositions of proteins from low evolutionary level organisms with that of all sequenced proteins; comparison of the mean contemporary protein composition with that obtained after simulation of the evolutionary process) confirm and universalize the observed trend. The above results direct the attention to the concept of a smaller number of amino acids in the ancient proteins and respectively simpler genetic code. A fluctuation around the initial primitive level is suggested to explain the conservatism of proteins of the same function in evolutionarily low level organisms. The observed trend could be applied for designing new proteins.

1. Introduction

The question of whether there exists a universal regularity (trend) in the changing amino acid composition of proteins in the course of time is of special importance for elucidating their evolution regarding both the past and future. Many authors consider that there should be a trend to the 5% level – an equal share of each protein amino acid. According to Dayhoff *et al.* (1978) there is a trend for increasing the share of the less mutable amino acids at the expense of the highly mutable ones. After computer simulation experiments the same authors established that, irrespective of the initial composition, there is an approach to the mean contemporary protein composition. Ivanov and Förtsch (1988) showed that the rate of this approach depends on the initial composition used.

We shall pose the question as follows. If one divides the 20 protein amino acids into two groups: usual (Thr, Pro, Ala, Ser, Arg, Gly, Leu, Val, Glu, Asp) and special (Met, Ile, Gln, His, Lys, Asn, Phe, Tyr, Trp, Cys), for which the grounds are given previously (Ivanov, 1989), is there a trend for increasing the share of special amino acids in the course of time?

Origins of Life and Evolution of the Biosphere **19**: 187–198, 1989.

2. Methods

The considerable number of known amino acid compositions provides a sufficiently large information base.

(A) Having in mind the small share of the special amino acids in conservative proteins (Ivanov, 1989) it is worthwhile to juxtapose amino acid composition with the evolutionary rate (V_e). The data on V_e from 53 cases (Dayhoff, 1978a*) were used. If there is a real trend for increasing the share of special amino acids in the course of time, this share should be small in the low V_e-proteins, and large in high V_e-proteins.

(B) Another promising approach is to compare the amino acid composition of proteins performing the same or very similar function from organisms representing different evolutionary levels.

(C) Studying the amino acid composition of a group of proteins performing together an evolutionarily old function is also a promising possibility.

(D) The cases of individual proteins performing an evolutionarily old function also deserve attention.

To check the existence of a trend for increasing the share of special amino acids in the course of time, two parameters were used: the percent share of Phe + Tyr + Trp and that of Cys. These parameters were chosen in view of the conservatism and functional value (see Volkenstein, 1982) of the above amino acids. In order to ignore the possibility that a transition special' → special'' amino acids would be interpreted as a transition usual → special amino acids, the percent share of the usual amino acids Thr + Pro + Ala + Ser + Gly + Glu + Asp was used as an additional third parameter. The data of Barker *et al.* (1984) on the mean contemporary protein composition were used as a basis for comparison: Phe + Tyr + Trp = 8.6%, Cys = 2.2%, Thr + Pro + Ala + Ser + Gly + Glu + Asp = 44.5%.

In order to check the statistical significance of the possible results the following tests were performed:

(1) comparison of the mean composition of the sequenced proteins from low evolutionary level organisms** (bacteria, viruses, yeasts, algae) – 67 sequences up to 1972 (Dayhoff, 1972) and 221 sequences up to 1978 (Dayhoff, 1978) with that of all proteins sequenced up to 1972 (108 families – Dayhoff and Hunt, 1972) and 1978 (314 families – Dayhoff *et al.*, 1978a);

(2) comparison of the mean compositions of several sets of proteins (sequenced up to 1972 – Dayhoff and Hunt, 1972; up to 1976 – 185 families, Dayhoff *et al.*,

* The citation refers to all values of V_e used.

** We do not insist that a human is higher than a bacterium. Having (most probably) a common evolutionary ancestor they have had an equal time to evolve. But, as a result human ('higher' organisms, respectively) have receded much farther from this ancestor than the bacterium ('lower' organisms, respectively) judging by fossil evidence (Schopf, 1972). This makes possible an extrapolation to the geological past.

1976; up to 1978 - Dayhoff *et al.*, 1978a; up to 1984 - 2898 proteins with 591717 amino acid residues, Barker *et al.*, 1984) with those obtained from them after simulation of the evolutionary process. This simulation was made by means of the mutation probability matrices (the probability of each amino acid to remain unchanged was used) of Dayhoff and co-workers for distances of 256 (Dayhoff *et al.*, 1972) and 250 accepted point mutations (PAMs) (Dayhoff *et al.*, 1978) based on 814 and 1572 PAMs respectively, accumulated from closely related proteins. In our opinion the comparison of mean compositions of the closely related proteins (the mutations among them being the base of the above matrices) before and after a simulation of evolutionary process is especially important. Actually, the normalized frequencies are used in the simulation. They are approximately proportional to the average composition of each group of closely related proteins multiplied by the number of mutation in the corresponding phylogenetic tree (see Dayhoff *et al.*, 1978).

3. Results and Discussion

(A) The results of juxtaposing amino acid composition - evolutionary rate (V_e) are presented in Table I. The cases of known V_e were divided in three groups depending on V_e-value. As seen, when V_e increases, the share of Phe + Tyr + Trp shows a trend to increase. But, in cases of especially high V_e ($V_e > 16$)* this trend disappears. This can be explained with the decrease of the value of the above amino acids: a high V_e can ensure a rapid adaptation to any requirements. As to the share of Cys, a clearly expressed trend to increase was observed. That gives some reason to consider the share of Cys as a measure of the time elapsed..

The results of the other approaches are presented in Table II.

(B) Cytochrome C is known as a highly conserved protein ($V_e = 2.2$). The share of Phe + Tyr + Trp for the lower evolutionary level (*Tetrahymena puriformis***) is smaller than for the higher one (horse, human). A similar picture is observed in the case of the conserved triosephosphate isomerase: the share of Phe + Tyr + Trp and that of Cys in vertebrates*** is larger than in bacteria. The PSTI-type inhibitors exhibit a similar picture. Thus, PSTI from dog and human demonstrate a larger share of Phe + Tyr + Trp and of Cys compared to the related subtilisin inhibitor of *Streptomyces albogriseolus*. Similar is the case of plastocyanins: the share of Phe + Tyr + Trp in green plant sequences is larger than that in the sequence of the blue green algae *Anabaena variabilis*. Another similar example: endolysin, Bacteriophage lambda (a peptide hydrolase) and lysozymes - bacteriophage and animal (glycosyl hydrolases) perform similar functions and are related distantly (Dayhoff *et al.*, 1976). There is a regular increase of the share of both Phe + Tyr

* Accepted point mutations per 100 residues per 100 mya.

** The sequences are taken from Barker *et al.* (1984).

*** In cases when no concrete organism is shown, mean values from a considerable number of sequences are used (see Dayhoff *et al.*, 1978a).

TABLE I

Comparison between the evolutionary rate (V_e) and percentage of the most conservative amino acids

V_e[a]	Number of cases	Cases where F+Y+W>8.6%[b]	Cases where C>2.2%[b]
0.0– 3.5	19	6	1
3.6–12.0	17	11	6
16.0–37.0	17	6	12

[a] Accepted point mutations per 100 residues per 100 mya. Data from Dayhoff (1978a) are used.

[b] The share (%) in the mean contemporary protein (Barker *et al.*, 1984).

+ Trp and Cys in the above succession. The situation with globins, arranged according to a decreasing conservatism, is to be noted: annelid globin → myoglobins (V_e = 12) → β-hemoglobin chains (V_e = 12). There is a regular increase of the value of both parameters used. There is evidence for the validity of Haeckel's biogenetic law on the molecular level (Acher *et al.*, 1977; Goodman *et al.*, 1982; Ivanov, 1987). On the other hand, the chronology of appearance of the individual hemoglobin chains in the course of embryonic development is known (Dickerson and Geis, 1983). It is connected with corresponding functional changes. Hence, ζ- and α-hemoglobin chains are older than ϵ and β-chains in both embryonic and evolutionary respects. Here, the inequalities /Phe + Tyr + Trp/$_{\zeta+\alpha}$ </Phe + Tyr + Trp/$_{\epsilon+\beta}$ (23 < 27 residues) and Cys$_{\zeta^4\alpha}$ < Cys$_{\epsilon^4\beta}$ (2<3) are to be noted.

(C) It is hard to challenge the ancientness of reproductive function. Because of that we paid attention to the protein taking part in its realization: ubiquitin, histones, protamines, LAC repressor protein, ribosomal proteins (*E. coli*). Except for ubiquitin (V_e = 0) and histones (V_e = 0.1 – 0.9) there are no data on the evolutionary rate of the above proteins. Having in mind the ancientness of the function and the conservatism of some of the participants, low values of both parameters were to be expected, which was observed (Table II). It is to be noted that the values of both of the parameters are lower for 50S ribosomal proteins compared to 30S ones (see also Ivanov *et al.*, 1984).

Another relatively ancient function is the locomotory one. The sequences of collagen, keratin, troponin, myosin, tropomyosin, actin, Ca-binding protein, Ca-dependent regulator protein, and flagellin from various sources are known. The evolutionary rate is known for troponin (1.5), collagen (1.7) and parvalbumin (7.0). Having in mind the considerations in the case of reproductive function, low values of both of the parameters were to be expected, which was also observed (Table II). Only for parvalbumin the share of Phe + Tyr + Trp is larger than in mean contemporary protein, which can be explained by the higher V_e (see Table I and the corresponding text). The very low values of both of the parameters for bacterial flagellin (*Bacillus subtilis* 168) are to be noted.

TABLE II

Comparison of the composition: proteins with the same, similar or cooperative functions from organisms representing different evolutionary levels

Protein	T+P+A+S+G+ E+D[44.5][a]	F+Y+W[8.6][a]	C[2.2][a]
Cytochrome C			
Tetrahymena puriformis	53.2	7.3	1.8
Horse	42.3	8.3	1.8
Human	41.3	8.3	1.8
Triosephosphate isomerase [5.3.1.1]			
Bacillus stearothermophilus	45.4	6.0	0.8
Vertebrate	47.0	7.3	1.9
Protease inhibitors (PSTI-type)			
Subtilisin inhibitor *Streptomyces albogriseolus*	57.5	6.2	3.5
Dog, submandibular gland	41.7	7.8	10.4
Human	41.1	7.1	10.7
Plastocyanin			
Anabaena variabilis	52.4	7.6	1.0
Plants	52.7	9.2	1.0
Globins			
Annelid	49.0	5.4	0.7
Myoglobins	42.7	7.3	0.0
Hemoglobin alpha chains	46.8	8.1	0.9
Hemoglobin beta-type chains	40.3	9.1	1.1
Reproductive function			
Chromosomal proteins			
Ubiquitin	40.6	4.1	0.0
Histones H2A, H2B, H3, H4	40.2-46.7	3.1-5.9	0.0-1.5
Protamines	26.4	0.8	0.0
LAC repressor (*E. coli*)	44.9	4.0	0.9
Ribosomal proteins (*E. coli*)			
50S ribosomal proteins	43.2	5.2	0.4
30S ribosomal proteins	42.7	5.6	0.6
Locomotory function			
Fibrous proteins			
Collagen	79.7	1.7	0.0
Keratin B2 IIIA, IIIB, feather	45.3-53.0	4.2-5.5	8.5-24.5
Contractile system proteins			
Troponin C, T, I	43.8-52.4	3.4-7.1	0.0-1.3
Myosin, alkali and DTNB light chains	45.7-49.9	5.8-8.9	0.5-1.2
Tropomyosin, alpha chain	50.0	2.5	0.4
Actin	47.3	8.6	1.3
Calcium-binding protein, intestinal	46.2	7.5	0.0
Calcium-dependent regulator protein	52.1	6.8	0.0
Parvalbumin	52.7	9.4	0.8
Flagellin (*Bacillus subtilis* 168)	46.4	2.0	0.0

Table II (continued)

Protein	T+P+A+S+G+ E+D[44.5][a]	F+Y+W[8.6][a]	C[2.2][a]
Peptide hydrolases			
Subtilisins [3.4.21.14]	55.0	6.3	0.0
Alpha-lytic protease Myxobacter 495 [3.4.21.12]	52.5	6.1	3.0
Bacterial proteases A and B *Streptomyces griseus*	58.3	8.4	2.2
Trypsin-like enzyme, *Streptomyces griseus* [3.4.21.4]	52.0	8.6	2.7
Chymotrypsinogens [3.4.21.1]	49.8	7.3	4.1
Trypsin [3.4.21.4]	44.2	7.6	5.4
Elastase [3.4.21.11]	41.7	8.7	3.3
Endolysin Bacteriophage lambda	42.7	8.3	0.6
Glycosyl hydrolases			
Bacteriophage lysozyme	39.1	8.5	1.2
Animal lysozyme related	40.3	9.7	6.2
Superoxide dismutase [1.15.1.1]	52.3	3.3	2.0
Penicillinase [3.5.2.6]	40.2-50.4	5.5-7.8	0.0
Tryptophan synthetase alpha chain [in 4.2.1.20]	47.5	7.0	1.1

[a] The share (%) in the mean contemporary protein (Barker *et al.*, 1984).

(D) The group of serine proteases is relatively well studied. The evolutionary relationships among α-lytic protease (Myxobacter 495), protease A, protease B, and the trypsin-like enzyme (the latter three from *Streptomyces griseus*), elastase, chymotrypsinogen, and trypsin is beyond question (Barker and Dayhoff, 1972; Young *et al.*, 1978). The ancientness of the protease function is also hardly to be challenged. The evolutionary rate is known for trypsin (5.9). Here also, as in C. cases (above), low values of special amino acids' share were expected. This was confirmed regarding the share of Phe + Tyr + Trp. As to the share of Cys, it is larger than that in mean contemporary protein. But, here also, the shares of both parameters, and especially that of Cys in bacterial proteases, are smaller than in vertebrates. The share of Cys is zero in subtilisins, for which an evolutionary relationship with the other serine proteases was established previously (Ivanov and Genov, 1987a, b). Noteworthy is that in both cases the same codon is used for the active Ser (Brenner, 1988).

Several proteins performing relatively old functions can also be listed: superoxide dismutase, penicillinase, tryptophan synthetase α-chain (participating in the biosynthesis of Trp). Here also, the values of both parameters are lower than those in the mean contemporary protein.

As to the additional third parameter (the percent share of Thr + Pro + Ala + Ser + Gly + Glu + Asp), in conservative proteins its values are considerable at the expense of those of the first two parameters (the percent share of Phe +

Tyr + Trp and that of Cys). In the vast majority of cases with the increase of complexity of the organism (progressive evolution) this additional parameter undergoes a change, which is opposite to that of the first two parameters. This is especially clear in cases of protein with the same function from organisms representing different evolutionary levels (Table II). Thus, we have serious grounds to reject the possibility for a considerable transition special' → " amino acids. Only one possibility remains: the evolutionary transition usual → special amino acids. Besides, due to the use of the additional third parameter (comprising most of the usual amino acids) it is clear that the above transition usual → special amino acids is valid for all special amino acids.

Most probably the transition usual → special amino acids ensures (due to the stereochemical peculiarities of the special amino acids) a better adaptation: improvement of the existing functions and creation of new ones. From the examples cited (Table II) it is clear that there is a correspondence between the increasing evolutionary level of the source-organism (progressive evolution) and the increased share of special amino acids.

3.1. The difficult questions

The existence of the above trend leads to some important and at the same time difficult questions. In a few words the situation is as follows. At the contemporary stage of evolution there is a difference in the amino acid composition of proteins with the same function. This difference depends on the evolutionary level of the organism: the usual amino acids predominate in the proteins from low evolutionary level organisms. The same type of amino acid composition is typical for the proteins performing conservative function with low evolutionary rate. It follows that the low evolutionary level organisms are more conservative than the higher level ones. Does that mean that the rate of mutation acceptance for proteins of the same function is different, depending on the degree of evolutionary development of the source-organism. But, if yes, it is to be expected the rate of acceptance of the mutations in low level organisms to be higher than in high level organisms because of the rapid change of generations. Hence, a paradox is formed: in the low evolutionary level organisms, where a high rate of mutation acceptance is to be expected, a conservatism is observed.

In our opinion, there is a possibility for solving the question: in the low evolutionary level organisms there exists a fluctuation with the same or even with a higher (because of the rapid change of generations) evolutionary rate around a primitive level, for which, having in mind the above considerations, a considerable share of C + G in the codons, leading to a corresponding share of the usual amino acids, is typical.

Naturally, such a possibility needs to be confirmed. We directed our attention to the introns. What is their role in contemporary proteins? This is not very clear at the moment. Still, the generalization of a number of data (Waring and Davies, 1984) showed a presence of conservatism and common secondary structure for a class of introns from low evolutionary level organisms. Besides, it proved that

there is a pairing between a part of intron and a part of exon. Most probably, introns must have played an important role in the past.

Let us consider some data on the introns in hemoglobin genes (hemoglobins are the best studied class of proteins up to now). According to Slightom *et al.* (1987) the introns accept mutations easier than the neighbouring exons from fetal globin genes in primates.

On the other hand Proudfoot *et al.* (1982) established that ζ-gene, the oldest both in embryonic and evolutionary respects, contains large introns consisting, in part, of simple repeat sequences of high C + G content.

Hence, there are at the same time considerable variability and, in spite of that, preserved conservatism: high C + G content (encoding the usual amino acids) and periodicity. Therefore, the data on introns, considered as a relic, are evidence in favour of the possibility for a fluctuation with a considerable evolutionary rate around a level with the following characteristics: considerable share of C + G and clearly expressed periodicity.

In the proteins from low evolutionary level organisms a high level of C + G and periodicity are also observed (e.g. ribosomal proteins from *E. coli*, Ivanov *et al.*, 1984), and a high evolutionary rate is to be expected. So, the situation is analogous to that of the introns. And this analogy is not formal: its validity is supported by the fact of pairing between intron and exon parts (Waring and Davies, 1984). Most probably, in the proteins of low evolutionary level organisms there is a fluctuation around a level of considerable C + G share and periodicity. This is in good accordance with the concept of adaptive evolution in prokaryotes (see Zavarzin, 1987), which ensures a readiness for a rapid adaptation to a possible change of the environment, but if there is not a serious change of above conditions there will not be an adequate, directed change of the organism (protein, respectively). Most probably, in the case of simply organized organisms, the extent of change of the environment requiring a directed answer from the organism is relatively high. The situation for the complex-balanced high organisms is completely different: the changes of the organism (protein, respectively) following (even insignificant) changes of the surroundings lead inevitably to a chain of connected, and because of that, directed changes – directed or progressive evolution.

As a result, there are reasons to consider the proteins from low evolutionary level organisms as 'living fossils', which provides a possibility for a simple interpretation of the results by comparing protein sequences.

3.2. Tests of statistical significance of the trend observed

The results of comparison of the mean protein compositions from low evolutionary level organisms (in the sense explained in Methods) with those of all sequenced proteins are given in Table III. The number of amino acid residues in the low organism proteins is considerable – 7760 up to 1972 and 30052 up to 1978. Besides the first set is included entirely into the second one.

The results of simulation of the mean compositions of different large sets of

TABLE III

Mean percent compositions (MC) of proteins from low evolutionary level organisms compared to those of all sequenced proteins[a]

Proteins	Amino acid	M	I	Q	H	K	N	F	Y	W	C	Usual	Special	Σ (F,Y,W,C)
Low level 1972[a]		1.8	5.0	3.7	1.4	7.1	5.5	3.3	3.2	1.1	2.2	65.7	34.3	9.8
MC 1972		1.6	4.6	3.6	2.2	7.0	4.4	3.5	3.4	1.2	3.4	65.1	34.9	11.5
Low level 1978[a]		1.9	4.9	3.6	1.9	6.7	4.5	3.4	3.3	1.1	1.9	66.8	33.2	9.7
MC 1978		1.7	4.5	3.9	2.0	6.6	4.3	3.6	3.4	1.3	2.9	65.8	34.2	11.2

[a] Data of Dayhoff (1972, 1978) are used. Data on MC are given in the text.

proteins – these of closely related proteins (the mutations among them are the base of the corresponding mutation probability matrices) and the larger sets of proteins sequenced up to 1972, 1976, 1978, 1984, are given in Table IV.

An inspection of these results (the right parts of Tables III and IV) reveals their considerable similarity. This makes it possible to be discussed together. In both cases, irrespective of the different approaches used and the different sets of sequences (each smaller of which is a part of the larger one) there is a confirmation of the trend in protein molecular evolution observed (see Tables I and II) – increasing the share of special amino acids on the account of the usual one. Noteworthy is the coincidence of a number of details. The largest increasing share exhibit Phe, Tyr, Trp, and Cys. Coincidences are also observed in the most cases of exception: increasing the share of the usual amino acids Gly and Leu and decreasing that of some special ones – Met, Ile, Asn. The analysis of the mutation probability matrices (bases on numerous mutations among closely related proteins) makes it possible to explain the increasing share of the above special amino acids. There is a discrepancy between the process of changing into other amino acids and the reverse process, which is due to the high mutability. Asn and Gln demonstrate a uniform and radial mutability. In case of Ile and Met the lower mutability is compensated by its directivity. Thus, in 20% of the cases the mutations of Met result in Leu (Dayhoff *et al.*, 1978). The analysis of the individual amino acid compositions from low level organisms shows especially large amplitude for the above mentioned amino acids.

Thus, Met, Ile, Asn and, to a certain extent Gln and Lys could be considered as intermediate links or, to be more precise, as decreasing reserves of the relatively rapid (accepted point mutation) evolution. Therefore, its rate should decrease in the course of time. Noteworthy is the constancy of the results (for each of the approaches) having in mind the differences in the compositions used.

Based on a considerable information, the tests of statistical significance, which are two more independent approaches, not only confirm but also universalize the trend in protein molecular evolution observed by us.

TABLE IV

Mean percent compositions (MC) and normalized frequencies of exposure to mutation (NF) compared to those after simulated evolution by using mutation probability matrices

Composition	Matrix used	Usual amino acids										Special amino acids										Usual	Special	Σ (F,Y,W,C)
		T	P	A	S	R	G	L	V	E	D	M	I	Q	H	K	N	F	Y	W	C			
NF 1972	256 PAMs (1972)	6.2	4.1	9.6	5.7	3.4	9.0	8.4	7.8	5.3	5.3	1.2	3.5	3.2	3.4	8.5	4.2	4.5	3.0	1.2	2.5	64.8	35.2	11.2
Result		3.5	3.8	7.0	2.1	4.6	12.1	14.8	6.9	2.7	3.3	0.2	1.6	1.2	3.5	9.3	1.5	7.2	6.4	3.0	5.3	60.8	39.2	21.9
MC 1972	256 PAMs (1972)	6.5	5.5	8.1	7.8	3.9	7.6	7.3	6.9	4.8[a]	4.8[a]	1.6	4.6	3.6[a]	2.2	7.0	4.4[a]	3.5	3.4	1.2	3.4	63.2[a]	34.9[a]	11.5
Result		3.9	5.4	6.1	3.0	5.5	10.7	13.4	6.4	2.6	3.1	0.3	2.2	1.4	2.4	8.0	1.7	5.9	7.5	3.1	7.5	60.1	40.0	24.0
MC 1976	250 PAMs (1978)	5.8	5.6	8.7	7.2	4.5	8.8	7.2	6.6	5.6[a]	5.5[a]	1.7	4.6	3.9[a]	2.1	6.6	4.2[a]	3.5	3.5	1.2	3.1	65.5[a]	34.4[a]	11.3
Result		3.3	5.8	5.9	3.8	4.0	12.4	12.8	5.9	3.5	3.2	0.5	2.4	2.0	1.6	8.3	1.3	5.8	5.7	3.4	8.4	60.6	39.4	23.3
NF 1978	250 PAMs (1978)	5.8	5.1	8.7	7.0	4.1	8.9	8.5	6.5	5.0	4.7	1.5	3.7	3.8	3.4	8.1	4.0	4.0	3.0	1.0	3.3	64.3	35.8	11.3
Result		3.2	5.2	5.7	3.6	3.5	12.2	14.7	5.6	3.0	2.6	0.5	1.9	1.9	2.6	9.9	1.2	6.5	4.7	2.8	8.7	59.3	40.7	22.7
MC 1978	250 PAMs (1978)	6.1	5.2	8.6	7.0	4.9	8.4	7.4	6.6	6.0	5.5	1.7	4.5	3.9	2.0	6.6	4.3	3.6	3.4	1.3	2.9	65.7	34.2	11.2
Result		3.5	5.4	5.9	3.7	4.4	11.9	13.2	5.9	3.8	3.2	0.5	2.4	2.0	1.6	8.3	1.4	6.0	5.5	3.7	7.9	60.9	39.3	23.1
MC 1984	250 PAMs (1978)	6.0	5.1	7.9	7.1	5.0	7.4	8.9	6.5	5.9	5.1	2.3	5.0	4.0	2.3	6.1	4.3	3.9	3.3	1.4	2.2	64.9	34.8	10.8
Result		3.5	5.4	5.4	3.7	4.5	10.5	16.0	5.8	3.7	3.0	0.7	2.6	2.1	1.8	7.7	1.4	6.6	5.4	4.1	6.0	61.5	38.4	22.1

[a] Values which would be larger if the amides had been determined in all proteins. The mutation probability matrices for the evolutionary distance of 256 PAMs (Dayhoff *et al.*, 1972) and 250 PAMs (Dayhoff *et al.*, 1978) are used. Data on MC are given in the text.

In our opinion the universality of the difference between the mean compositions of proteins from low evolutionary level organisms and that of the other proteins is especially important. This is an evidence in favour of our hypothesis of the fluctuation of composition of proteins from low evolutionary level organisms around a primitive level. Therefore, there are now more serious grounds to consider the information macromolecules from the above organisms as 'living fossils' and to use them as a base for extrapolation to the past. An extrapolation of the observed trend in reverse direction could provide important information concerning the design of new proteins.

4. Conclusions

The analysis of a considerable amount of information (protein sequences representing 83 superfamilies) by using several approaches: the dependence evolutionary rate – protein amino acid composition, the dependence evolutionary level of the source-organism – protein amino acid composition, studying amino acid composition of groups of proteins performing jointly an old function, as well as that of individual proteins with old functions, demonstrated that there is a trend for increasing the share of special amino acids at the expense of the usual ones in the course of evolution. The presence of such a trend makes it possible to throw a glance to the past as well as to the future. The look to the past (special amino acids → 0) leads again to the concept (Ivanov and Ivanov, 1980; Ivanov, 1989) of the small number of usual amino acids in the ancient proteins and the corresponding simpler genetic code.

The obtained results are a quantitative characteristic of the progressive evolution. They reflect the unidirectional character of the process – a result of the action of a number of factors. The thorough investigation of the individual factors of the chain protein biosynthesis → structure → function can shed an additional light on the details of the above process.

References

Acher, R., Chauvet, J. P., and Duguet, M.: 1977, *INSERM* **70**, 173.

Barker, W. C. and Dayhoff, M. O.: 1972, in M. O. Dayhoff (ed.), *Atlas of Protein Sequence and Structure*, Vol. 5, National Biomedical Research Foundation, Washington, D.C.,p. 101.

Barker, W. C., Hunt, L. T., Orcutt, B. C., George, D. G., Yeh, L. S., Chen, H. R., Blomquist, M. C., Johnson, G. C., Seibel-Ross, E. I., Hong, M. K., and Ledley, R. S.: 1984, *Protein Sequence Data Base of the Protein Identification Resource*, Release 3.0, November 15, 1984, National Biomedical Research Foundation, Washington, D.C.

Brenner, S.: 1988, *Nature* **334**, 528.

Dayhoff, M. O.: 1972, *Atlas of Protein Sequence and Structure*, Vol. 5, National Biomedical Research Foundation, Washington, D.C.

Dayhoff, M. O.: 1978, *Atlas of Protein Sequence and Structure*, Vol. 5, Suppl. 3, National Biomedical Research Foundation, Washington, D.C.

Dayhoff, M. O.: 1978a, in M. O. Dayhoff (ed.), *Atlas of Protein Sequence and Structure*, Vol. 5, Suppl. 3, National Biomedical Research Foundation, Washington, D.C., p. 1.

Dayhoff, M. O., Dayhoff, R. E., and Hunt, L. T.: 1976, in M. O. Dayhoff (ed.), *Atlas of Protein Sequence and Structure*, Vol. 5, Suppl. 2, National Biomedical Research Foundation, Washington, D.C., p. 301.

Dayhoff, M. O., Eck, R. V., and Park, C. M.: 1972, in M. O. Dayhoff (ed.), *Atlas of Protein Sequence and Structure*, Vol. 5, National Biomedical Research Foundation, Washington, D.C., p. 89.

Dayhoff, M. O. and Hunt, L. T.: 1972, in M. O. Dayhoff (ed.), *Atlas of Protein Sequences and Structure*, Vol. 5, National Biomedical Research Foundation, Washington, D.C., p. D-355.

Dayhoff, M. O., Hunt, L. T., and Hurst-Calderone, S.: 1978a, in M. O. Dayhoff (ed.), *Atlas of Protein Sequence and Structure*, Vol. 5, Suppl. 3, National Biomedical Research Foundation, Washington, D.C., p. 363.

Dayhoff, M. O., Schwartz, R. M., and Orcutt, B. C.: 1978, in M. O. Dayhoff (ed.), *Atlas of Protein Sequence and Structure*, Vol. 5, Suppl. 3, National Biomedical Research Foundation, Washington, D.C., p. 345.

Dickerson, R. E. and Geis, I.: 1983, *Hemoglobin: Structure, Function, Evolution, and Pathology*, The Benjamin/Cummings Publishing Company, Inc., Menlo Park, California.

Goodman, M., Romero-Herrera, A., E., Dene, H., Czelusniak, J., and Tashian, R. E.: 1982, in M. Goodman (ed.), *Macromolecular Sequences in Systematic and Evolutionary Biology*, Plenum Press, New York and London, p. 115.

Ivanov, O. Ch.: 1987, *Naturwissenschaften* **74**, 40.

Ivanov, O. Ch.: 1989, *Studia biophysica* (in press).

Ivanov, O. Ch. and Förtsch, B.: 1988, *Commun. Dept. Chem.* **21**, 27.

Ivanov, O. Ch. and Genov, N. Ch.: 1987a, *Common. Dept. Chem.* **20**, 244.

Ivanov, O. Ch. and Genov, N. Ch.: 1987b, *Commun. Dept. Chem.* **20**, 517.

Ivanov, O. Ch. and Ivanov, Ch. P.: 1980, *J. Mol. Evol.* **16**, 47.

Ivanov, O. Ch., Kenderov, P. S., and Revalski, J. P.: 1984, *Origins of Life* **14**, 557.

Proudfoot, N. J., Gil, A., and Maniatis, T.: 1982, *Cell* **31**, 553.

Schopf, J. W.: 1972, in C. Ponnamperuma (ed.), *Exobiology*, North-Holland Publishing Company, Amsterdam and London, p. 16.

Slightom, J. L., Theisen, T. W., Koop, B., and Goodman, M.: 1987, *J. Biol. Chem.* **262**, 7472.

Volkenstein, M. V.: 1982, Physics and Biology, Academic Press, Inc., London.

Waring, R. B. and Davies, R. W.: 1984, *Gene* **28**, 277.

Young, C. L., Barker, W. C., Tomaselli, C. M., and Dayhoff, M. O.: 1978, in M. O. Dayhoff (ed.), *Atlas of Protein Sequence and Structure*, Vol. 5, Suppl. 3, National Biomedical Research Foundation, Washington, D.C., p. 73.

Zavarzin, G. A.: 1987, in L. P. Tatarinov and A. P. Rasnitsin (eds.), *Evolution and Biocoenotic Crises*, Nauka, Moscow, p. 144 (in Russian).