Consensus temporal order of amino acids and evolution of the triplet code

E.N. Trifonov*

Department of Structural Biology, The Weizmann Institute of Science, Rehovot 76100, Israel

Abstract

Forty different single-factor criteria and multi-factor hypotheses about chronological order of appearance of amino acids in the early evolution are summarized in consensus ranking. All available knowledge and thoughts about origin and evolution of the genetic code are thus combined in a single list where the amino acids are ranked chronologically. Due to consensus nature of the chronology it has several important properties not visible in individual rankings by any of the initial criteria. Nine amino acids of the Miller's imitation of primordial environment are all ranked as topmost (G, A, V, D, E, P, S, L, T). This result does not change even after several criteria related to Miller's data are excluded from calculations. The consensus order of appearance of the 20 amino acids on the evolutionary scene also reveals a unique and strikingly simple chronological organization of 64 codons, that could not be figured out from individual criteria: New codons appear in descending order of their thermostability, as complementary pairs, with the complements recruited sequentially from the codon repertoires of the earlier or simultaneously appearing amino acids. These three rules (Thermostability, Complementarity and Processivity) hold strictly as well as leading position of the earliest amino acids according to Miller. The consensus chronology of amino acids, **G/A**, **V/D**, **P**, **S**, **E/L**, **T**, R, N, K, Q, I, C, H, F, M, Y, W, and the derived temporal order for codons may serve, thus, as a justified working model of choice for further studies on the origin and evolution of the genetic code. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Abiotic; Chronology of amino acids; Chronology of codons; Origin of life; Origin of code

1. Introduction

Three years ago Thomas Bettecken and myself (Trifonov and Bettecken, 1997) developed an idea that, perhaps, some repeating sequences frequently expanding in disease, in particular, GCT triplets, may have enjoyed their expandability as an advantage in the very early days of evolution of nucleic acids. The very first codons, therefore, could have been the GCU triplet and its point change derivatives. The list of the earliest amino acids could be reconstructed from the yields of amino acids in the imitation experiments of Miller (1953), giving priority to chemically simplest ones, and those which are served today by more ancient tRNA synthetases of class II. Spectacularly, resulting six amino acids, ala, asp, gly, pro, ser and thr, are, indeed, encoded today by the GCU derivative triplets. The success of the fusion of these two independent 'predictions', on the earliest codons and the earliest amino acids (Trifonov and

Bettecken, 1997), suggests that, perhaps, one would be also able to make a more detailed reconstruction of the chronology of amino acids and of respective codons. For that purpose one could recruit many more criteria of the evolutionary age of amino acids, along with three criteria mentioned above. In this paper 40 different estimates are analyzed both collected from literature and suggested anew. Results of the reconstruction, unexpectedly informative, are presented below.

2. Results

2.1. Criteria

The list of collected criteria is shown in Table 1, in no special order. A brief comment and reference to each of the criteria follow. The numbers for the criteria below correspond to the listing in the table.

2.1.1. Various multi-factor theories

N9. According to Jukes (1973) initially only ten amino acids were present in the code, each one invol-

---

Table 1
Forty criteria for amino-acid chronology

1.	Simplicity (number of non-hydrogen atoms)
2.	Involvement with more ancient synthetases of class II
3.	Yield in the Miller's experiments
4.	Amino-acid composition of extant proteins
5.	Chemical inertness
6.	Stability of codon-anticodon interactions
7.	Molecular clock sequence analysis of synthetases
8.	Stability of ('older') assignments in the table of the code
9.	Jukes' theory of the origin of the code
10.	Co-evolution theory of Wong
11.	GCU-based theory of Trifonov and Bettecken
12.	RRY hypothesis of Crick
13.	RNY hypothesis, Eigen and Schuster
14.	Hypothesis of Hartman
15.	Hypothesis of Ferreira
16.	Prebiotic physicochemical code of Altshtein-Efimov
17.	Early copolymerization code of Nelsestuen
18.	Composition of proteinoids of Fox
19.	Co-evolution theory of Dillon
20.	Yield in imitation experiments of Fox and Windsor
21.	Yield in experiments of Harada and Fox, high temperatures.
22.	Yield in shock wave experiments of Bar-Nun
23.	Co-evolution theory of Wächtershäuser
24.	Remnants of primordial code in tRNA (Möller and Janssen)
25.	Evolutionary distances between isoacceptor tRNAs
26.	Hypothesis of O. Ivanov
27.	Match scores of BLOSUM matrix
28.	A/U start, Jimenez-Sanchez
29.	N-fixing amino acids first, Davis
30.	GNN codons first, Taylor and Coates
31.	Algebraic model of Hornos and Hornos
32.	Composition of translated Urgen
33.	Murchison meteorite
34.	Minimal graph complexity, amino acids
35.	Minimal graph complexity, amino-acid residues
36.	Hypothesis of Jimenez-Montano
37.	'Size/complexity' score, Dufton
38.	Minimal alphabet for folding
39.	DNA stability
40.	RNA duplex stability

ving four to eight codons. Later the excessive codons had been reassigned to additional amino acids.

N10. Co-evolution theory of Wong (1981, 1988) is based on the observation that biosynthetically related similar amino acids share similar codons rather than similar physicochemical properties. The amino acids of the Miller's mix are considered the earliest.

N11. GCU-based theory (Trifonov and Bettecken, 1997).

N12. RRY theory of Crick et al. (1976).

N13. RNY theory of Eigen, Schuster and Winkler-Oswatitsch (Eigen and Schuster, 1978; Eigen and Winkler-Oswatitsch, 1981a,b; Eigen et al., 1981).

N14. According to Hartman (1975a, 1975b, 1978, 1995) the code started from a singlet code and eventually developed to a doublet and triplet code, starting with G and C.

N15. This speculation is based on gradual transition of enzymatic functions from RNA-like oligomers to randomly synthesized peptides (Ferreira and Coutinho, 1993).

N16. The code could start as selective interactions of amino acids and nucleotides in progenes – mixed anhydrides of amino acids and trinucleotides (Altshtein and Efimov, 1988).

N17. This is based on the hypothesis on copolymerization of amino acids and bases, and hydrogen bonding between respective aminoacyl nucleotides (Nelsestuen, 1978).

N19. Biochemical co-evolution model based on metabolism of simple sulfur bacteria in the genus *Thioploca* (Dillon, 1978).

N23. Possible early synthesis of amino acids in sulfur-iron surface systems in volcanic vents (Wächtershäuser, 1988).

N28. Analysis of the metabolism of pyrimidine biosynthesis leads to an idea that the code began in RNA world with the letters A and U (Jimenez-Sanchez, 1995).

N29. This suggested chronology is based on the N-fixing amino acids (D, E, N, Q) as an initial set. These amino acids are thought to have been ambiguously translated from polyA in mineral surface reactions. Further additions to the code are assumed to follow path lengths of amino acid biosynthesis, starting with reductive citrate cycle (Davis, 1999).

N30. The order is based on the key positions in the synthetic pathways of the amino acids, on the presence in the Miller's mixture, and on the preference to amino acids encoded by GNN triplets (Taylor and Coates, 1989).

N31. Order suggested by group-theoretical analysis of the symmetries in the genetic code (Hornos and Hornos, 1993).

N36. Model based on the group theory, thermodynamics of codon-anticodon interactions and on complexity of amino acids (Jimenez-Montano, 1999).

2.1.2. Yields in experiments imitating primordial conditions

It is assumed that in the primordial conditions those abiotic amino acids which had higher concentrations in the environment were also first to be incorporated in the early code. The very first experiment of this kind indicated detectable amounts of glycine (Löb, 1913; see also Yockey et al., 1997).

N3. Experiment of Miller with electric discharge in imitated conditions of early Earth atmosphere (Miller, 1953; Weber and Miller, 1981).

N18. Composition of proteinoids in the experiments by Fox and Waehneldt (1968).

N20. Experiments of Fox and Windsor (1970) at moderate temperatures.

N21. Experiments of Harada and Fox (1964) at high temperatures.

N22. Shock wave experiments (Bar-Nun et al., 1970).

2.1.3. Criteria based on complexity of the amino acids

One would expect that more complex amino acids would appear later in the evolution of the code. There are several rather different algorithms to evaluate complexity of amino acids and their side residues.

N1. Complexity, estimated by simple counts of non-hydrogen atoms in the side residues (Trifonov and Bettecken, 1997).

N34. Minimal graph complexity of amino acids and

N35. Minimal graph complexity of amino-acid residues (Papentin, 1982). In these criteria traditional chemical presentations of the amino acids are transformed in linear graphs, and complexity of the graphs then calculated.

N37. 'Size/complexity' score (Dufton, 1997) is calculated that takes into account both complexity of chemical structure and bulk physico-chemical characteristics of the residues (size, weight, charge, etc.).

2.1.4. Criteria based on amino-acid composition of various ensembles of proteins

N4. Present-day proteins. Their composition is taken from (Arques and Michel, 1996), for prokaryotes and eukaryotes, and rank averaged. One may expect that the latest amino acids would be largely underrepresented in the composition of modern proteins. Thus the ranking by composition would also reflect amino-acid chronology.

N26. Functionally most ancient proteins may contain more of earliest amino acids (Ivanov, 1989).

N32. The composition of the protein encoded by the earliest mRNA (master tRNA or Urgen of Eigen and Winkler-Oswatitsch, 1981a,b) should be dominated by the earliest amino acids.

2.1.5. Criteria based on thermostability of early nucleic acids

N6. Codon-anticodon interactions. More stable interactions would be more favorable in the early, presumably higher temperature conditions, and in early simple complexes with no additional stabilizing interactions. The melting enthalpies of the dinucleotide stacks (Xia et al., 1998) corresponding to the first and second codon positions are taken as measure of their thermostability.

N39. Stability of early DNA as carrier of genetic memory may be of importance. The amino acids with higher stability of the respective complementary pairs of triplets in DNA would be expected then to be earlier ones. The stability of the triplet pairs is calculated as sum of respective melting enthalpies for the dinucleotide stack components (Gotoh, 1983).

N40. Stability of the early duplex RNA genes. Respective enthalpies for the triplet pairs are calculated as above, by using values for RNA dinucleotides (Xia et al., 1998). See Table 2.

2.1.6. Criteria that involve amino-acyl-tRNA synthetases

N2. Involvement with more ancient synthetases of class II (Eriani et al., 1990). Since the class II synthetases are believed to be more ancient, than class I enzymes (Hartman, 1995), respective ten amino acids served by the class II enzymes would be expected to have been available earlier than others.

N7. Two amino-acyl-tRNA synthetases, for tyr and trp, are found to be sequence-wise youngest of all, keeping inter-species similarities higher than intra-species similarities between the synthetases (Ribas de Pouplana et al., 1996).

Table 2
Thermostability of the codons (complementary pairs, kcal/M)

A	GCC	28.3	K	AAG	17.3	R	AGG	23.9
	GCG	25.5		AAA	13.6		AGA	22.9
	GCU	25.4	L	CUC	22.9	S	UCC	25.8
	GCA	25.3		CUG	20.9		UCG	23.1
C	UGC	25.3		CUA	18.2		UCU	22.9
	UGU	21.8		CUU	17.3		UCA	22.9
D	GAC	23.8	L	UUG	17.3	S	AGC	25.4
	GAU	21.8		UUA	14.5		AGU	21.9
E	GAG	22.9	M	AUG	19.8	T	ACC	24.8
	GAA	19.3	N	AAC	18.2		ACG	22.0
F	UUC	19.3		AAU	16.3		ACU	21.9
	UUU	13.6	P	CCC	26.8		ACA	21.8
G	GGC	28.3		CCG	24.0	V	GUC	23.8
	GGG	26.8		CCU	23.9		GUG	21.8
	GGA	25.8		CCA	23.8		GUA	19.1
	GGU	24.8	Q	CAG	20.9		GUU	18.2
H	CAC	21.8		CAA	17.3	W	UGG	23.8
	CAU	19.8	R	CGC	25.5	Y	UAC	19.1
I	AUC	21.8		CGG	24.0		UAU	17.1
	AUA	17.1		CGA	23.1			
	AUU	16.3		CGU	22.0			

2.1.7. Other single-factor criteria

N5. Chemical inertness would be important in the early stages of life when the amino acids were not yet protected by sophisticated cellular homeostases. The amino acids are divided in three groups – inert, moderately reactive and highly reactive (Trifonov, 1999).

N8. Many of assignments in the table of the triplet code are unstable, that is in some species a given codon may serve a different amino acid. Such instabilities may indicate which of the amino acids are acquired more recently. The ranking of the instabilities is evaluated from data in (Osawa et al., 1992).

N24. Remnants of primitive code at positions 3 to 5 in tRNA sequences (Möller and Janssen, 1990).

N25. The order may be estimated from evolutionary distances between isoacceptor tRNAs (Chaley et al., 1999).

N27. Currently used matrices for amino acid substitution, e.g. BLOSUM62 (Henikoff and Henikoff, 1992) may reflect not only similarity of properties of the interchangeable amino acids, but the time of appearance of the amino acid as well. In particular, C, H and W may have high diagonal values in the matrices rather because these are young amino acids that had less time for the replacements.

N33. Amino acids delivered by meteorites may reflect

prebiotic conditions. The extraterrestrial origin of life may be assumed (Kvenvolden et al., 1971).

N38. Minimal and, presumably, early alphabet sufficient for rapid folding of small β-sheet protein (Riddle et al., 1997).

2.2. Consensus chronology of amino acids

2.2.1. Averaging raw data

In Fig. 1 the rankings of the amino acids by their order of appearance in evolution of the triplet code are indicated for all 40 above criteria of their chronology. Many of them do not provide detailed ranking in which case the amino acids are grouped under the same rank for all group members. For example, by the criterion 8 the amino acids A, D, E, F, G, H and P are all equally earliest (seven amino acids of the same average rank 4), then follows V (rank 8), I (rank 9), K, N, S, Y (four amino acids of the same rank 11.5), and so on. For every amino acid the corresponding 40 rank values are averaged. The lowest average rank value would correspond to the amino acid that is most frequently on the left – the earliest one. In Table 3 in the column 'Raw data' the aver-

```
        1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20

 (1. G  A  -CS-   -  PTV  -  -  -DILMN-  -  -  EKQ  -  H  -FR-   Y  W)
 1* G  A  -CS-   P  V  T  M  L  K  -DI-  N  E  Q  H  F  R  Y  W
 2. -AG-  -  -  -DFHKNPST-  -  -  -  CEILMQRVWY  -  -  -
 (3. A  G  D  V  L  E  I  S  P  T  M  K  -  -  -CFHNQRWY-  -  -)
 3* G  A  D  E  S  -PV-  I  L  T  F  Y  K  C  M  R  -NQ-  H  W
 4. L  A  G  S  -VE-  -IT-  K  D  R  P  N  Q  F  Y  -HM-  -CW-
 5. -  -  AFGILPV  -  -  -  NQST  -  -  -  -  CDEHKMRWY  -  -  -
 (6. A  -GP-  -  DES  -  -TV-  R  L  -  CHQW  -  -IM-  Y  -  FKN -)
 6* A  G  P  S  D  R  E  -TW-  -CV-  -HL-  Q  M  I  Y  N  F  K
 7. -  -  -  -  -  -  ACDEFGHIKLMNPQRSTV  -  -  -  -  -  -  -WY-
 8. -  -  ADEFGHP  -  -  V  I  -  KNSY  -  -  MTW-  L  R  -CQ-
 9. -  -  -  ADEGHLPQRV  -  -  -  -  -  -CFIKNSTY-  -  -MW-
(10. -  -ADEGS-  -  V  -PT-  -IL-  F  C  Y  -KR-  -NQ-  H  -MW-)
 11. A  -  -  DGPSTV  -  -  -  CFHIKLMNQRWY  -  -  -  -  -
 11* A  G  -  DSTV  -  P  -IN-  E  -  -  -  CFHKLMQRWY  -  -  -
 12. -  DGNS  -  -  -  -  -  ACEFHIKLMPQRTVWY  -  -  -  -  -
(13. -AG-  -  -  DINSTV  -  -  -  CEFHKLMPQRWY  -  -  -  -)
 14. G  P  A  R  -  -  DENQST  -  -  -HK-  C  -  -FILVY-  -  -MW-
 15. -  -  FGKLNP  -  -  -  -  CDEHQRSTVW  -  -  -  -  AIMY  -
 16. -  -  -  ADEGKRSTV  -  -  -  -  -CFHILMNPQWY-  -  -
 17. -  -  -  -  DEFHIKLMSTVY  -  -  -  -  -  -ACGNPQRW-  -  -
 18. A  E  V  -GK-  M  L  C  Y  -NQ-  I  -DF-  R  H  P  W  T  S
 19. G  A  D  V  E  Q  -  HLPR  -  N  T  -IS-  -KM-  F  -CY-  W
(20. G  I  -AP-  S  E  D  F  L  V  -  -  CHKMNQRTWY  -  -  -)
(21. G  A  E  D  L  -PV-  S  I  T  -FY-  -  -  -CHKMNQRW-  -  -)
 22. G  A  V  L  -  -  -  -  CDEFHIKMNPQRSTWY  -  -  -  -
 23. -DE-  -  -  -ACGNPQST-  -  -  -  ILMV  -  -  -  FHKRWY  -  -
 24. -  ADGV  -  -  -  -  CEFHIKLMNPQRSTWY  -  -  -  -  -
 25. Q  H  P  -LS-  G  C  W  R  V  -DE-  A  Y  T  -IM-  F  -KN-
 26. -  -  -  ADEGLPRSTV  -  -  -  -  -  -  CFHIKMNQWY  -  -  -
 27. -  -AILSV-  -  -  -  EKMQRT  -  -  -  DFGN  -  -PY-  H  C  W
 28. -  -  FIKLMNY  -  -  -  -  CDEHQRSTVW  -  -  -  -  AGP  -
 29. -  DENQ  -  -  -  APSV  -  -CG-  T  -  ILM  -  R  K  -FY-  H  W
 30. -  -ADEGV-  -  -  -  -  -  -  CFHIKLMNPQRSTWY  -  -  -  -
 31. -  -CDFSV-  -  -  -  -EKLRY-  -  -HP-  -  -  -AGIMNQTW-  -  -
 32. V  -  AGP  -  -  -  ENRT  -  -  LQS  -  -  -  CDFHIKMYW  -  -  -
 33. -AG-  -  DEPV  -  -  -  -  -CFHIKLMNPQRSTWY  -  -  -  -
 34. G  A  D  P  -CS-  N  E  V  K  Q  T  L  M  I  R  H  F  Y  W
(35. G  A  -CS-  P  V  K  M  T  L  -DI-  N  E  Q  H  F  R  Y  W)
 36. -  ADGV  -  -  LPR  -  -  CIKQST  -  -  -  -  EFHMNWY  -  -
 37. G  A  V  -IL-  S  T  K  P  D  N  E  Q  F  R  Y  C  H  M  W
 38. -  -AGEIK-  -  -  -  -  -  CDFHLMNPQRSTVWY  -  -  -  -
 39. A  G  S  R  C  T  D  V  P  E  W  -HN-  F  L  I  Y  M  -KQ-
(40. G  A  P  W  -RS-  C  D  T  E  H  V  -LM-  Q  I  Y  N  F  K)
```

Fig. 1. Amino-acid chronology ranking by individual criteria. Criteria N1*, N3*, N6* and N11* are combined criteria (see Section 2.2.2.1). Respective excluded criteria are enclosed in parentheses. When the amino acids are grouped together under the same rank, dashes are placed in unoccupied rank positions, to assist reading the ranks from the figure.

Table 3
Consensus chronology of amino acids

	Raw data			Filtered data		Miller	
						+	−
G	4.4 ± 0.7	1	G	2.9 ± 0.3	1	G	A
A	4.9 ± 0.8	2	A	2.9 ± 0.3	2	A	G
V	6.9 ± 0.6	3	V	6.6 ± 0.6	3	V	V
D	7.2 ± 0.7	4	D	7.0 ± 0.7	4	D	D
S	7.9 ± 0.7	5	E	7.2 ± 0.6	5	E	E
E	8.2 ± 0.7	6	P	7.5 ± 0.6	6	P	P
P	8.3 ± 0.7	7	S	7.7 ± 0.7	7	S	S
L	9.4 ± 0.7	8	L	9.5 ± 0.7	8	L	L
T	10.1 ± 0.6	9	T	9.8 ± 0.6	9	T	T
I	11.2 ± 0.7	10	R	11.5 ± 0.7	10		
N	11.8 ± 0.7	11	N	12.2 ± 0.7	11		
R	12.0 ± 0.7	12	K	12.3 ± 0.5	12		
K	12.0 ± 0.7	13	Q	13.0 ± 0.4	13		
Q	12.4 ± 0.7	14	I	13.0 ± 0.5	14	I	I
C	12.4 ± 0.7	15	C	14.3 ± 0.6	15		
F	13.0 ± 0.7	16	H	14.9 ± 0.5	16		
H	13.3 ± 0.6	17	F	15.1 ± 0.4	17		
M	14.0 ± 0.6	18	M	15.4 ± 0.4	18		
Y	14.7 ± 0.5	19	Y	15.6 ± 0.4	19		
W	15.8 ± 0.6	20	W	16.7 ± 0.5	20		

age ranking values are shown together with calculated errors for the averages. The amino acids are sorted here according to the rank averages. According to the raw data analysis gly and ala are the first amino acids to appear followed by val and asp. Tyr and trp appear last. Although the raw data result needs some refinement (filtering), most of the features of the raw chronology stay unchanged after the treatment.

2.2.2. Filtering

Two steps of filtering are applied: combining of related criteria, and elimination of the noise background.

2.2.2.1. Related criteria. Some criteria are related by the concept they are based on, but are sufficiently different. While other criteria may be unrelated by the underlying idea, and yet the respective rankings are very similar. A quantitative estimate of the relatedness of various criteria is given by correlation coefficients for pair-wise comparisons between the rankings. The Spearman correlation coefficients are calculated, based on normalized total misfits between compared rankings. Table 4 displays the 39 × 40 triangular matrix for all pair-wise correlations between the 40 rankings. Positive correlations significantly exceed the negative ones which is an encouraging result indicating that large proportion of the criteria are relevant, indeed, and the respective rankings, apparently, convey some common truth. One could discard those criteria that show overall negative correlation with others. They may, however, give correct order if not ranks for some amino acids. Besides, the non-orthodox views are especially valuable for the analysis, since other, 'positive'

criteria may be biased in favor of dominating speculations or results. This is why all criteria are taken into the calculation. On the other hand, there could be some criteria that correlate strongly due to their relatedness *a priori*, being based on similar ideas. That is, essentially, they offer largely the same rankings. Such data should be combined (averaged), in order to avoid the obvious bias. Of 22 strongly correlated (correlation coefficient >0.5) pairs 16 pairs are found to be apparently independent. For example, rankings N24 and N30 (based on tRNA sequences, and on the domination of GNN codons, respectively) show highest correlation (0.81). Six pairs, on the other hand, are clearly related. They make four groups: N1 and N35 – these are criteria of complexity of amino acids; N3, N10, N20 and N21 – all based on or closely related to the Miller's imitation experiments; N6 and N40 – criteria of thermostability; and N11 and N13 – related theories based on GCU and RNY triplets, respectively. These criteria are, thus, combined in four averaged rankings (N1*, N3*, N6* and N11* in the Fig. 1). This reduction, from 40 to 34 rankings, makes the first step of filtering of the raw data.

2.2.2.2. Reduction of the noise level. The rank estimates for each of 20 amino acids provided by the remaining 34 rankings, if all irrelevant, would be scattered randomly over a whole range from rank 1 to rank 20, with the density 1.74 per rank unit (note that there are fractional ranks, but not smaller than 1.0). In reality the suggested estimates for individual amino acids rather form clusters (data not shown). For example, 20 of 34 rank estimates for alanine concentrate between the ranks 1 and 3. Most of remaining estimates, far away from the average, can be considered as noise. To eliminate this noise component the low density (<1.0 per rank unit) flanks of rank distributions for individual amino acids are discarded. New, corrected averages are calculated. These are presented in the column 'Filtered data' of the Table 3.

2.2.2.3. Robustness of the derived chronology of amino acids. In Table 5 the raw data chronology, filtered chronology (two steps separately) as well as earlier estimates, on smaller sets of the criteria, are presented. The boldface amino acids correspond to those with estimated ranks identical or one rank off the final (last column) chronology. Already by only three criteria (Trifonov and Bettecken, 1997) ten ranks of 20 are practically the same as by extensive 40 criteria analysis with two-step filtering. The earlier chronologies become closer to the final one with the increase in the number of criteria through 7 (Trifonov, 1999a), 25 (Trifonov, 1999b), 28 (Trifonov, 2000) and 40 criteria (this work). Remarkably, the last column shows almost the same order as in the raw data for the 40 criteria, with only three amino acids having ranks more than one unit off (arg, ile, and ser). Data for one and two steps of filtering are within accuracy of one rank unit almost identical. These comparisons, thus, demonstrate

Table 4
Correlation coefficients for all pairwise comparisons of the rankings suggested by 40 criteria

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	–	0.09	0.30	0.24	0.14	0.17	0.22	−0.17	−0.08	0.29	0.31	0.04	0.35	−0.04	−0.29	0	−0.33	−0.09	0.10	0.15
2	–	–	0.05	−0.02	0.09	0.06	0.04	0.14	0	0.04	0.30	0.16	0.36	0.38	0.04	0.18	−0.08	−0.39	−0.03	0.20
3	–	–	–	0.46	0.21	0.27	0.07	0.12	−0.04	0.47	0.43	0.04	0.32	0	−0.33	0.34	0.09	−0.01	0.31	0.40
4	–	–	–	–	0.16	0.18	0.12	−0.06	0.11	0.42	0.28	−0.03	0.28	0.05	−0.19	0.33	−0.08	−0.04	0.22	0.23
5	–	–	–	–	–	−0.05	0.09	0.06	−0.07	0.07	0.14	−0.14	0.21	−0.04	0.05	−0.16	−0.26	−0.16	0.03	0.34
6	–	–	–	–	–	–	0	−0.06	0.14	0.33	0.51	0.10	0.18	0.21	−0.14	0.32	−0.35	−0.33	0.30	0.29
7	–	–	–	–	–	–	–	−0.06	0.04	0.07	0	0.16	0	0.16	0.16	0.01	0.02	−0.04	0.18	0.03
8	–	–	–	–	–	–	–	–	−0.10	0.13	0.18	−0.12	0.01	0	−0.21	0	−0.08	−0.26	0.09	0.28
9	–	–	–	–	–	–	–	–	–	0.17	0.25	−0.04	0.10	0.16	−0.18	0.19	−0.16	−0.29	0.29	−0.08
10	–	–	–	–	–	–	–	–	–	–	0.46	0.13	0.18	−0.12	−0.31	0.25	−0.07	−0.11	0.18	0.41
11	–	–	–	–	–	–	–	–	–	–	–	0.20	**0.58**	0.18	−0.22	**0.56**	−0.15	−0.11	0.21	0.38
12	–	–	–	–	–	–	–	–	–	–	–	–	0.34	0	0.12	0.10	−0.20	−0.24	−0.02	0.06
13	–	–	–	–	–	–	–	–	–	–	–	–	–	0.04	−0.28	0.21	−0.12	−0.17	0.07	0.28
14	–	–	–	–	–	–	–	–	–	–	–	–	–	–	−0.02	0.21	−0.44	−0.40	0.25	0.03
15	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	−0.26	−0.36	−0.38	−0.22	−0.16
16	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	−0.02	−0.07	0.25	0.14
17	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	−0.20	−0.39	−0.06
18	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	−0.09	−0.21
19	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.09
20	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
21	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
22	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
23	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
24	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
25	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
26	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
27	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
28	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
29	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
30	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
31	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
32	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
33	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
34	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
35	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
36	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
37	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
38	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
39	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–

Table 4 (*continued*)

	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
1	0.18	0.18	0.28	0.14	−0.26	0.17	−0.02	−0.63	0.16	0.04	−0.36	0.21	0.16	0.47	**0.79**	0.22	0.36	−0.10	0.20	0.12
2	0.09	0.03	0.02	0.12	−0.38	0.12	−0.38	−0.43	−0.14	0.03	−0.31	0.06	0.16	0.19	0.09	−0.03	0.16	0.03	0.11	0.03
3	**0.67**	0.30	0.13	0.32	−0.36	0.44	0.17	−0.57	0.13	0.36	−0.18	0.17	0.39	0.26	0.26	0.42	0.45	0.12	0.02	−0.01
4	0.36	0.27	−0.06	0.14	−0.33	0.38	0.45	−0.49	−0.05	0.20	−0.23	0.21	0.17	0.12	0.23	0.24	**0.63**	**0.19**	0.06	−0.14
5	0.28	0.25	−0.08	0.08	−0.27	0.15	0.10	−0.25	−0.09	0	−0.42	0.38	0.09	−0.02	0.05	0.22	0.38	0	−0.23	−0.27
6	0.30	0.16	0.17	0.28	0.16	**0.52**	−0.19	−0.74	0.10	0.29	−0.21	0.36	0.42	0.28	0.10	0.28	0.04	0.12	0.41	**0.53**
7	−0.03	0.16	0.12	0.16	−0.12	0.04	0.17	−0.11	0.11	0.09	−0.07	0.06	0.04	0.18	0.18	0.13	0.12	0.09	−0.03	−0.06
8	0.29	0	−0.14	0.17	−0.39	−0.01	−0.52	−0.52	−0.17	0.26	−0.22	−0.13	0.34	−0.01	−0.20	−0.15	−0.13	0.09	−0.12	−0.27
9	0.10	−0.04	0.10	−0.04	−0.05	0.48	−0.15	−0.57	0.14	0.01	−0.22	0.26	0.08	0.02	−0.23	0.06	0.06	−0.29	0	−0.15
10	**0.64**	0.13	0.20	0.26	−0.32	0.40	0	−0.68	0.22	0.38	−0.04	0.12	0.36	0.22	0.20	0.28	0.35	0.10	0.17	−0.03
11	0.36	0.25	0.21	0.40	−0.15	**0.59**	−0.04	−0.67	0.18	0.45	−0.13	0.34	**0.63**	0.36	0.23	0.22	0.26	0.11	0.36	0.25
12	0.02	0.40	0.16	**0.60**	−0.13	0.02	−0.26	−0.02	0.15	0.42	−0.06	−0.07	0.28	0.20	−0.04	0.01	−0.03	0.24	0.07	−0.01
13	0.26	0.29	0.10	0.40	−0.43	0.28	−0.03	−0.35	0.08	0.34	−0.38	0.18	0.36	0.22	0.16	0.13	0.39	0.19	0.22	0.02
14	−0.02	−0.03	0.04	0.04	−0.14	0.20	−0.15	−0.67	−0.07	0.01	−0.39	0.37	0.14	0.24	−0.03	0.13	0.07	0.01	0.16	0.24
15	−0.27	0.09	−0.30	−0.06	−0.07	−0.16	−0.45	0.13	−0.29	−0.16	−0.16	−0.05	−0.10	−0.08	−0.21	−0.11	−0.11	−0.03	−0.23	−0.21
16	0.17	0.10	0.03	0.25	−0.32	**0.62**	0.08	−0.54	0	0.36	−0.02	0.23	0.32	0.11	0.05	0.17	0.14	0.20	0.28	0.12
17	0.05	−0.20	−0.40	0.20	−0.54	−0.08	0.01	0.13	−0.43	−0.17	0.28	−0.47	−0.30	−0.47	−0.26	−0.39	−0.12	−0.17	−0.50	−0.58
18	−0.06	0.23	−0.18	0.11	−0.57	−0.23	0.01	−0.22	−0.26	0.21	−0.30	−0.10	0.10	−0.09	−0.04	−0.10	−0.02	0.20	−0.43	−0.54
19	0.27	0.25	0.13	0.32	−0.15	0.29	−0.01	−0.72	0.24	0.38	−0.27	0.32	0.42	0.27	0.01	0.31	0.18	0.02	0.01	−0.01
20	**0.51**	0.13	0.02	0.14	−0.16	0.40	−0.05	−0.46	−0.01	0.21	−0.16	0.15	0.37	0.20	0.11	0.09	0.27	0.23	0.05	0.09
21	–	0.27	0.14	0.25	−0.26	0.46	0	−0.57	0.12	0.35	−0.14	0.17	0.43	0.17	0.12	0.31	0.35	0.12	0.05	0
22	–	–	−0.16	**0.79**	−0.10	0.16	0.12	−0.02	−0.13	**0.62**	−0.16	0.19	**0.51**	0.07	0.18	0.29	0.35	0.43	0.08	0.01
23	–	–	–	0.04	−0.18	0.18	−0.25	−0.67	**0.64**	0.13	−0.42	0.14	0.12	0.45	0.11	0.05	−0.06	−0.07	0.02	0.05
24	–	–	–	–	−0.24	0.16	−0.05	−0.09	0.07	**0.81**	−0.06	0.15	**0.64**	0.22	0.11	0.37	0.20	0.43	0.17	0.06
25	–	–	–	–	–	0.01	−0.46	−0.54	−0.20	−0.30	−0.22	−0.15	−0.20	−0.19	−0.28	−0.13	−0.37	−0.39	−0.06	0.21
26	–	–	–	–	–	–	0.05	−0.65	0.14	0.25	−0.03	0.47	0.36	0.15	0.09	0.36	0.23	−0.05	0.31	0.25
27	–	–	–	–	–	–	–	−0.22	−0.15	−0.05	−0.24	0.03	−0.16	−0.16	0.06	0.03	0.37	0.04	−0.28	−0.41
28	–	–	–	–	–	–	–	–	−0.65	−0.33	−0.23	−0.65	−0.47	−0.67	−0.56	−0.61	−0.39	0.01	−0.82	−0.83
29	–	–	–	–	–	–	–	–	–	0.19	−0.37	0.10	0.18	**0.63**	0.04	0.08	−0.04	−0.04	−0.13	−0.12
30	–	–	–	–	–	–	–	–	–	–	−0.06	0.19	**0.77**	0.24	0.01	0.24	0.12	**0.60**	**0.12**	0.02
31	–	–	–	–	–	–	–	–	–	–	–	−0.37	0.17	−0.22	−0.27	−0.13	−0.33	−0.26	−0.14	−0.33
32	–	–	–	–	–	–	–	–	–	–	–	–	0.27	0.19	0.04	0.29	0.20	−0.01	0.13	0.10
33	–	–	–	–	–	–	–	–	–	–	–	–	–	0.34	0.14	0.26	0.15	0.43	0.18	0.18
34	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.45	0.27	0.20	0	0.15	0.22
35	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.22	0.32	−0.01	0.10	0.10
36	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.30	−0.06	0.07	0.10
37	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.08	0.03	−0.16
38	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	−0.10	−0.10
39	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.48

Table 5
Stability of the ranking

	Number of criteria (simple averaging)					Filtered	
	3	7	25	28	40	One step	Two steps
1.	**G**	**A**	**G**	**G**	**G**	**G**	**G**
2.	**A**	**G**	**A**	**A**	**A**	**A**	**A**
3.	S	S	**D**	**V**	**V**	**V**	**V**
4.	**D**	P	**V**	**D**	**D**	**D**	**D**
5.	**P**	V	**P**	S	S	S	**E**
6.	T	T	**S**	**P**	**E**	**E**	**P**
7.	V	**L**	**E**	**E**	**P**	**P**	**S**
8.	**L**	D	**L**	**L**	**L**	**L**	**L**
9.	I	I	**T**	**T**	**T**	**T**	**T**
10.	K	E	I	I	I	**N**	**R**
11.	**N**	**N**	**N**	**N**	**N**	**R**	**N**
12.	E	F	F	**R**	**R**	**K**	**K**
13.	C	**K**	H	F	**K**	**I**	**Q**
14.	M	**R**	K	K	**Q**	**Q**	**I**
15.	**H**	Q	R	Q	**C**	**H**	**C**
16.	**F**	**C**	Q	**H**	F	**C**	**H**
17.	Q	**H**	C	C	**H**	F	**F**
18.	R	**M**	**M**	**M**	**M**	**M**	**M**
19.	**Y**	**W**	**Y**	**Y**	**Y**	**Y**	**Y**
20.	**W**	**Y**	**W**	**W**	**W**	**W**	**W**

that the derived chronology is robust, and neither addition of new criteria, nor possible additional filtering steps would result in any appreciable changes.

2.2.2.4. Relation to experiments by S. Miller. The most important reading from the derived consensus chronology of amino acids is that the consensus places at least nine of ten amino acids detected in the initial experiment of Miller (1953) to the top of the list. This can not be explained by possible domination of criteria related to Miller's data, since this is taken care of by combining the Miller-related criteria in just one, No. 3* (to replace criteria No. 3, 10, 20 and 21). Even if the combined ranking No. 3* is excluded from calculations, the result stays the same, with only little change of order within the top nine (see Table 3). Since uncertainties in calculation of the averaged ranks are rather high, the isoleucine and arginine are within respective error bars indistinguishable and, thus, T in fact may well be immediately followed by I which would place all ten amino acids of Miller to the top.

This observation, first, demonstrates that the significance of the Miller's results goes well beyond mere demonstration of possibility of abiotic synthesis of amino acids. The amino acids of the Miller's mix appear also to be the very first ones to be accommodated in the evolving triplet code. On the other hand, this observation makes the derived consensus chronology highly relevant apparently reflecting, indeed, the order of events in early evolution of the code. This becomes even more evident when the reconstruction of the chronology of codons is attempted.

2.3. Chronology of codons

2.3.1. Initial reconstruction, 20 codon pairs

As originally suggested by Eigen and Schuster (1978), glycine and alanine could have been the very first amino acids, being the highest yield components of the imitated primordial mixture of Miller (1953). They also speculated that, perhaps, the first codons for glycine and alanine had been GGC and GCC, respectively, making a very stable complementary combination. This is, indeed, the highest stability pair of the whole list of 32 combinations. Table 2 lists all the codons and stabilities of the respective complementary triplet pairs, expressed in melting enthalpies. The initial enthalpy values for individual base-pair stacks used for the derivation of the data in the Table are taken from the latest estimates for double-stranded RNA (Xia et al., 1998). Encouragingly, the next two amino acids in the chronological list are valine and aspartic acid for which, similarly, a complementary pair of codons can be chosen from the known repertoire of the codons for these amino acids: GUC and GAC. Both are the most stable triplets in the repertoires. (Here and below stability of triplets means thermostability of respective complementary pairs). It, thus, appears that the original suggestions by Eigen and Schuster, on the primacy of the thermostability, and on acquisition of new codons in form of complementary pairs of the codons, could have been rules applicable for the development of the entire code. One of corollaries of the codon complementarity rule, as also speculated in the cited work, would be possible complementarity of respective ancient tRNAs. This, indeed, has been confirmed by sequence analysis of a large ensemble of tRNA sequences (Rodin et al., 1993, 1996).

To check whether the rules of thermostability and complementarity are valid also for the rest of the consensus chronology, in Fig. 2 the most stable codons (bold face) for all 20 amino acids are displayed in one diagram with their respective amino acids. There is uncertainty in assignment of temporal order for amino acids encoded by two sets of triplets – serine, leucine and arginine. One of each pair could appear after the first one at any later moment in the chronology. For an initial guess they are put here together, as contemporaries. As Fig. 2 demonstrates, sequential acquisition of all amino acids according to the consensus chronology perfectly obeys (with only two negotiable exceptions) the thermostability and complementarity rules. For every strong triplet on the diagonal there is complementary one that belongs to the codon repertoire of one of the amino acids acquired earlier. This is reflected in the conspicuous above-diagonal positioning of almost all the complementary wobble triplets. First four amino acids discussed above are followed by the glutamic acid that is one of the mentioned apparent exceptions. Its stable triplet GAG has a complement CUC, the most stable triplet for leucine, that appears nominally only after proline and serine. However, considering the accuracy of the ranking
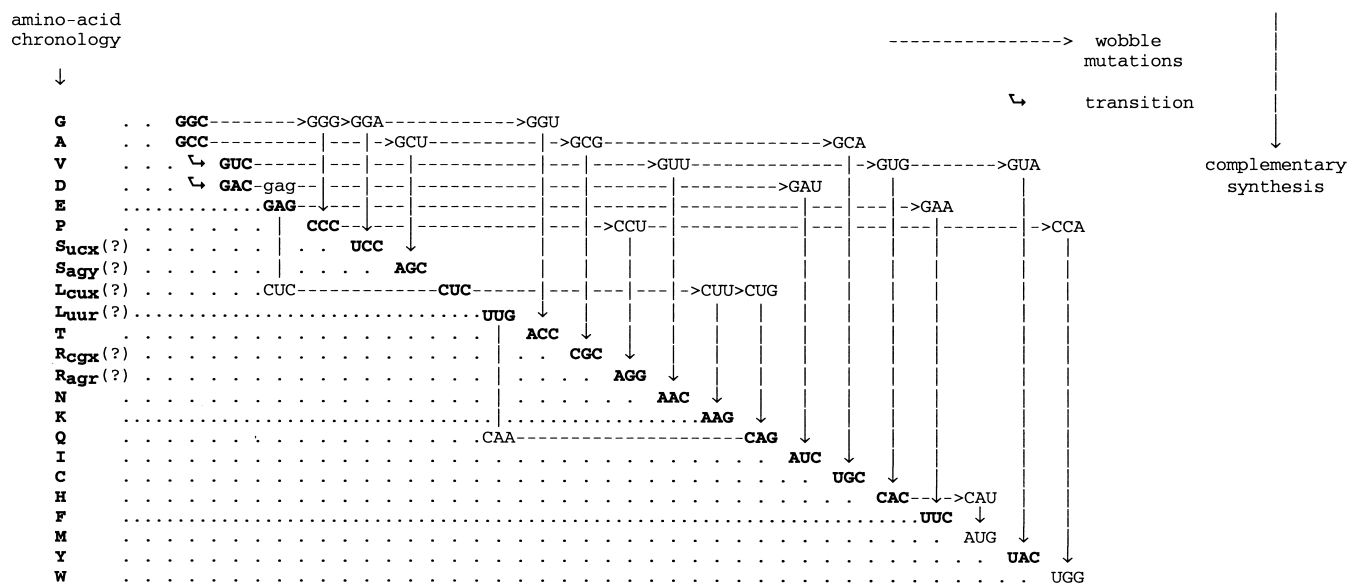
```
amino-acid
chronology                                        ---------------->  wobble
                                                                     mutations
   ↓
                                                         ↳    transition
                                                                                        ↓
  G    . .  GGC-------->GGG>GGA------------>GGU                                     complementary
  A    . .  GCC----------|---|->GCU----------|->GCG-------------------->GCA            synthesis
  V    . . . ↳  GUC------|---|--|------------|---|------>GUU--------------|->GUG-------->GUA
  D    . . . ↳  GAC-gag--|---|--|------------|---|-------|--------->GAU    |
  E    ............GAG--|---|--|------------|---|-------|------------|---|---|->GAA      |
  P    . . . . . . |  CCC--↓--|------------|---|->CCU--|------------|---|---|------|->CCA
 S_ucx(?) . . . . . |  . . UCC  ↓          |   |      |            |   |   |      |   |
 S_agy(?) . . . . . |  . . . .  AGC        |   |      |            |   |   |      |   |
 L_cux(?) . . . . . .CUC------------CUC----|---|---|->CUU>CUG      |   |   |      |   |
 L_uur(?).................................UUG  ↓   |   |          |   |   |      |   |
  T    . . . . . . . . . . . . . . |  ACC   ↓   |   |             |   |   |      |   |
 R_cgx(?) . . . . . . . . . . . .   |  . . CGC  ↓   |             |   |   |      |   |
 R_agr(?) . . . . . . . . . . . .   |  . . . AGG     |             |   |   |      |   |
  N    . . . . . . . . . . . . .    |  . . . .  AAC  ↓             |   |   |      |   |
  K    ...................................|..........AAG  ↓        |   |   |      |   |
  Q    . . . . . . . . . . . . . . . .CAA---------------->CAG  ↓    |   |   |      |   |
  I    . . . . . . . . . . . . . . . . . . . . . . .  AUC  ↓        |   |   |
  C    . . . . . . . . . . . . . . . . . . . . . . .  UGC  ↓        |   |
  H    . . . . . . . . . . . . . . . . . . . . . . .  CAC--↓->CAU   |   |
  F    .............................................UUC  ↓          |
  M    . . . . . . . . . . . . . . . . . . . . . . . . .  AUG  ↓    |
  Y    . . . . . . . . . . . . . . . . . . . . . . . . .  UAC  ↓
  W    . . . . . . . . . . . . . . . . . . . . . . . . . . .  UGG
```

Fig. 2. Reconstructed chronology of 20 codon pairs. The triplets corresponding to the most stable ones in the respective repertoires are shown boldface.

(see Table 3) the order E after P and S is equally acceptable, in which case E and L become next to each other, thus, putting the GAG and CUC triplets together, as the two rules require. The stablest codon CCC for proline is complementary to GGG for glycine, that came earlier. That is, the wobble version GGG of the glycine codon GGC appears simultaneously with its complementary counterpart CCC. Next in the chronology is serine which is encoded by two different sets of codons, UCX and AGY. The triplet UCC of the first set is stablest and it is complementary to GGA of glycine that is already present. The AGC triplet of the second set obeys the rules as well, being complementary to GCU, for alanine that is also already present in the developing chronology. As to the leucine encoded by UUG (this is second apparent exception), to satisfy the rules it only has to be put somewhere after glutamine, to make the complementary triplet CAA readily available. All subsequent stable codons appear simultaneously with complementary codons derived by wobble mutations from the above-diagonal triplets for the amino acids acquired at earlier steps of the chronology. Thus, suggestion of Eigen and Schuster is solidly confirmed being fully applicable not only to ala and gly, but to all amino acids of the consensus chronology. The third strict rule, the rule of processivity, can be formulated as well: new codons appear as derivatives of chronologically earlier codons. Majority of them are wobble mutations of the earlier codons, and their complements. Only in one case (from G/A to V/D) the new codons are generated, apparently, by transitions in the second codon positions or one transition (any of two possible ones) and one complementary copying step. Importantly, the processivity is due to the very special order of amino acids offered by the consensus chronology of amino acids. For example, putting W or R few ranks before P, F before E, or K before L, etc. – would destroy the triangular pattern.

Within the emerging picture of development of the code the starters GGC and GCC play unique role. Their origin is beyond the scheme and is more pertinent to the problem of the origin of life.

2.3.2. All 32 codon pairs

The thermostability rule holds that in every repertoire of codons for a given amino acid the most stable triplet is engaged first. Although there is also general decline of the thermostability towards the end of the codon chronology it is not strictly monotonous. For example, codon ACC for threonine is more stable than GUC for valine though valine comes earlier. There are, obviously, additional selection pressures other than just thermostability. This particular case of non-monotony has a simple explanation. All 64 triplets can be divided in two families, one having G or C in the central position, and another – A or U in the center. There is no way to derive A/U central triplets from G/C central triplets by a wobble mutation. The GUC/GAC (Val/Asp) pair is the most stable in the A/U central family. It is, apparently, formed by transition mutation(s) from GGC/GCC (Gly/Ala) pair. These both pairs occupy corner position, thus, supplying wobble and complementary versions corresponding to all remaining amino acids.

Considering an *a priori* importance of the thermostability one would also expect that within any given codon repertoire not only the first codon is chosen according to its stability, but second and third codons as well. Fig. 2 does not show it, perhaps, because the locations for second codon sets of serine, leucine and arginine are arbitrarily chosen here as immediately next to the first sets. They should be placed at some later position, like it is already suggested by apparent misplacement of the UUG triplet for leucine to begin with. Indeed, in the succession GCC, GCU, GCG, GCA for alanine (Fig. 2) the GCU codon complementary

Fig. 3. Reconstructed chronology of 32 codon pairs.

to AGC triplet for serine is less stable than GCG codon. By placing S_{agy} after R_{cgx} in the amino acid chronology one gets the monotonously descending order in stability of sequentially acquired alanine codons: GCC, GCG, GCU, GCA (see Table 2). Thus, by adjustment of the flexible positions for the double sets of serine, leucine and arginine one may attempt to organize the codon chronology to follow the extended thermostability rule: not only the first stablest, but second and third as well, consecutively. An additional flexibility is permitted by uncertainties of the average ranking values (Table 3). For example, D, E, P and S are indistinguishable in this respect, as well as the amino acids in the groups (N, K), (Q, I), and (H, F, M, Y). In the complete reconstruction of the codon chronology shown in Fig. 3 only one such allowed change in the amino-acid order is made, as already mentioned above: the (E, P, S) sequence is changed to (P, S, E). As a result, the derived complete codon chronology now strictly follows the rules of complementarity and processivity, and with only two violations – the extended rule of thermostability.

The reader is invited to trace in detail the order of engagement of the codons in every repertoire (line) of Fig. 3 and compare it with the thermostability orders given in Table 2.

The exceptions are the GUU codon for valine and CUU codon for leucine. According to their thermostability they should be at the end of the lines, after GUA and CUA, respectively. One possible explanation of the exceptions is that, probably, initial codon sets for valine and leucine were doublets GUY and CUY, rather than quartets, and the additional doublets, GUR and CUR were acquired at some later stage. In the present-day codon table the sets in form of ABY doublets are as common as quartets: UUY for phe, UAY for tyr, UGY for cys, CAY for his, AAY for arg, AGY for ser and GAY for asp.

With this comment the chronology of the codons displayed in Fig. 3 becomes fully consistent with the rules of extended thermostability, complementarity and processivity, while the amino-acid chronology (on the left) stays consistent with the calculated order in Table 3, within respective error bars (for E, P and S).

There are still many uncertainties in fine details of the chronologies that may or may not be clarified by further studies. Few additional criteria may emerge. However, merely due to large number of criteria contributing to the consensus picture, the new data could only change the rank averages for amino acids by fractions of unit, keeping the temporal orders essentially unchanged.

3. Discussion

The unique merit of the derived chronologies is that they are not built on any *a priori* assumptions but rather combine in as objective as possible way all ideas and suggestions regarding the order of acquisition of the amino acids during the evolution of the code. Four major and in many ways most natural discoveries are made:

1. The first amino acids to have been incorporated in early code were of abiotic origin, namely those which were obtained in classical imitation experiments by S. Miller.
2. In the development of the triplet code a major role was played by thermostability of codon-anticodon interactions.
3. New codons appeared in complementary pairs.
4. New codons were simple derivatives of chronologically earlier ones.

None of 40 individual criteria of chronology of amino acids alone would suggest these four fundamental conclusions. Since both amino-acid and codon chronologies derived as above reveal in the previously unexplored way the new basic knowledge, they, apparently, better represent true chronology of events than any earlier suggestions. It will be hard to challenge as fundamental and as obvious features as the four above, which are quite likely, thus, to stay.

The author has to admit that the spectacular order that came out of the disorder of very different, noisy and frequently rather questionable criteria is a great surprise. One could only hope, that most of the major factors that influenced the development of the code, are already present in the list of 40 criteria, that the noise level in the rankings by the individual criteria is, in average, not too high, and that the evolution of the code was not geared to a single dominating factor. Apparently, these expectations are met, indeed.

The robustness of the derived consensus ranking of amino acids is additionally demonstrated by the fact that even if the chronology of amino acids based on raw data (no filtering) is considered, the four rules strictly hold as well (data not shown), despite some differences in the succession of the amino acids (see Table 3). This also says that some uncertainty in the details of the chronologies still remains, even within the rather strict limits of the four rules. Further refinement of the orders will be, perhaps, possible after analysis of reconstructed ancient sequences of proteins and respective mRNAs.

An attempt to build the chronological orders on the basis of the four rules only, without taking into consideration all the numerous criteria, allows to outline the limits of the remaining uncertainties. It turns out that the first nine amino acids make a rather unique order: G/A > P > S > T, and G/A > V/D > E/L (sign > means here 'earlier'), with no other uncertainties. The remaining amino acids can be arranged in many possible combinations, all obeying the rules. The restrictions may be expressed in the following inequalities: $R_{cgx} > S_{agy} > C$, $N > H > Y$, $R_{agr} > W$, $K > Q > L_{uur}$, and $H > M$. These inequalities are valid for many alternative temporal orders. For example, to keep with the order GCC > GCG > GCU > GCA, S_{agy} may be placed anywhere after

R(CGC) but before C(UGC), see Fig. 3. For the time being such uncertain parts of the chronologies may be arranged by satisfying best the overall descend of the thermostability, as it is done in Fig. 3. Accordingly, the best guess on the amino acid chronology, consistent with the consensus and adjusted as allowed to fit the four rules is:**G/A**, **V/D**, **P**, **S**, **E/L**, **T**, R, N, K, Q, I, C, H, F, M, Y, W.

The thermostability rule is readily interpretable. By using the enthalpy values for deoxyribonucleotides instead of ribonucleotides one gets consistent picture as well (data not shown), due to largely the same order of descending thermostabilities of the DNA triplets. The question then is stability of what was so crucial in the evolution of the code. It could be stability of protein-coding DNA or RNA duplexes, or stability of codon-anticodon interactions in the early translation apparatus. One could imagine also codon-anticodon-like interactions between respective tRNAs, perhaps in prototype amino-acyl-tRNA synthetase complexes. RNA triplet interactions of the latter two suggestions seem to be better choice, being more dependent on the complementary triplet pair stabilities, then the duplexes. During the development of the code each time when the new codon is offered by the wobble mutation, other, next in the row mutation(s) of the same repertoire is, probably, present as well. The one that makes the most stable complementary pair is, apparently, selected.

Interestingly, both initiation codons and stop codons appear at the end of the derived codon chronology. It may correspond to the point of transition between progenote stage and first branchings in the basic tree of life all branches of which possess the initiation and termination codons. These codons could also appear at earlier stages. For example, initiation codon may have been introduced to exclude second strand of the early 'mRNA duplex' from translation. The termination could have been introduced to prevent further extensions of the protein chains by end-to-end fusion of their respective genes.

Intermediate stages of the development of the codon table leave many triplets unassigned. Perhaps, every appearance of such unassigned triplets would be lethal at these stages (if not at the end of the message). On the other hand, there could have been some precursor amino acids to be served by such triplets, e. g., in accordance with Wong's co-evolution theory (Wong, 1981, 1988). The reconstruction described in this work did not require such assumption. This does not exclude, however, some important undisclosed detours in the codon chronology that may be discovered in future studies.

Acknowledgements

References

Altshtein, A.D., Efimov, A.V., 1988. Physicochemical basis of origin of the genetic code: stereochemical analysis of interaction of amino acids and nucleotides on the basis of the progene hypothesis. Mol. Biol. 22, 1133–1149.

Arques, D.G., Michel, C.J., 1996. A complementary circular code in the protein coding genes. J. Theor. biol. 182, 45–58.

Bar-Nun, A., Bar-Nun, N., Bauer, S.H., Sagan, C., 1970. Shock synthesis of amino acids in simulated primitive environments. Science 168, 470–473.

Chaley, M.B., Korotkov, E.V., Phoenix, D.A., 1999. Relationships among isoacceptor tRNAs seem to support the co-evolution theory of the origin of the genetic code. J. Mol. Evol. 48, 168–177.

Crick, F.H.C., Brenner, S., Klug, A., Pieczenik, C., 1976. A speculation on the origin of protein synthesis. Origins Life 7, 389–397.

Davis, B.K., 1999. Evolution of the genetic code. Prog. Biophys. Mol. Biol. 72, 157–243.

Dillon, L.S., 1978. The Genetic Mechanism and the Origin of Life. Plenum Press, New York.

Dufton, M.J., 1997. Genetic code synonym quotas and amino acid complexity: cutting the cost of proteins? J. Theor. Biol. 187, 165–173.

Eigen, M., Schuster, P., 1978. The hypercycle. A principle of natural self-organization. Part C: The realistic hypercycle. Naturwissenschaften 65, 341–369.

Eigen, M., Winkler-Oswatitsch, R., 1981a. Transfer-RNA: The early adaptor. Naturwissenschaften 68, 217–228.

Eigen, M., Winkler-Oswatitsch, R., 1981b. Transfer-RNA, an early gene? Naturwissenschaften 68, 282–292.

Eigen, M., Gardiner, W., Schuster, P., Winkler-Oswatitsch, R., 1981. The origin of genetic information. Sci. Am. 244, 88–118.

Eriani, G., Delarue, M., Poch, O., Gangloff, J., Moras, D., 1990. Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. Nature 347, 203–206.

Ferreira, R., Coutinho, K.R., 1993. Simulation studies of self-replicating oligoribotides, with a proposal for the transition to a peptide-assisted stage. J. Theor. Biol. 164, 291–305.

Fox, S.W., Waehneldt, T.V., 1968. The thermal synthesis of neutral and basic proteinoids. Bioch. Bioph. Acta 160, 246–249.

Fox, S.W., Windsor, C.R., 1970. Synthesis of amino acids by heating of formaldehyde and ammonia. Science 170, 984–986.

Gotoh, O., 1983. Prediction of melting profiles and local helix stability for sequenced DNA. Adv. Biophys. 16, 1–52.

Harada, K., Fox, S.W., 1964. Thermal synthesis of natural amino acids from a postulated primitive terrestrial atmosphere. Nature 201, 335–336.

Hartman, H., 1975a. Speculations on the evolution of the genetic code. Origins of Life 6, 423–427.

Hartman, H., 1975b. Speculations on the origin and evolution of metabolism. J. Mol. Evol. 4, 359–370.

Hartman, H., 1978. Speculations on the evolution of the genetic code II. Origins of Life 9, 133–136.

Hartman, H., 1995. Speculations on the origin of the genetic code. J. Mol. Evol. 40, 541–544.

Henikoff, S., Henikoff, J.G., 1992. Amino acid substitution matrices for protein blocks. Proc. Natl. Acad. Sci. 89, 10915–10919.

Hornos, J.E.M., Hornos, Y.M.M., 1993. Algebraic model for the evolution of the genetic code. Phys. Rev. Lett. 71, 4401–4404.

Ivanov, O.C., 1989. On the possible evolution of genetic code: the composition of primitive proteins. Stud. Biophys. 134, 201–214.

Jimenez-Montano, M.A., 1999. Protein evolution drives the evolution of the genetic code and vice versa. BioSystems 54, 47–64.

Jimenez-Sanchez, A., 1995. On the origin and evolution of the genetic code. J. Mol. Evol. 41, 712–716.

Jukes, T.H., 1973. Possibilities for the evolution of the genetic code from a preceding form. Nature 246, 22–26.

Kvenvolden, K.A., Lawless, J.G., Ponnamperuma, C., 1971. Nonprotein amino acids in the Murchison meteorite. Proc. Natl. Acad. Sci. USA 68, 486–490.

Löb, W., 1913. Über das Verhalten des Formamids unter der Wirkung der stillen Entladung: Ein Betrag zur Frage der Stickstoff-Assimilation. Ber. Dtsch. Chem. Ges. 46, 684–697.

Miller, S.L., 1953. Production of amino acids under possible primitive earth conditions. Science 117, 528–529.

Möller, W., Janssen, G.M.C., 1990. Transfer RNAs for primordial amino acids contain remnants of a primitive code at position 3 to 5. Biochimie 72, 361–368.

Nelsestuen, G.L., 1978. Amino-acid directed nucleic acid synthesis. A possible mechanism in the origin of life. J. Mol. Evol. 11, 109–120.

Osawa, S., Jukes, T.S., Watanabe, K., Muto, A., 1992. Recent evidence for evolution of the genetic code. Microbiol. Rev. 56, 229–264.

Papentin, F., 1982. On order and complexity II. Application to chemical and biochemical structures. J. Theor. Biol. 95, 225–245.

Ribas de Pouplana, L., Frugier, M., Quinn, C.L., Schimmel, P., 1996. Evidence that two present-day components needed for the genetic code appeared after nucleated cells separated from eubacteria. Proc. Natl. Acad. Sci. USA 93, 166–170.

Riddle, D.S., Santiago, J.V., Bray-Hall, S.T., Doshi, N., Grantcharova, V.P., Yi, Q., Baker, D., 1997. Functional rapidly folding proteins from simplified amino acid sequences. Nature Struct. Biol. 4, 805–809.

Rodin, S., Ohno, S., Rodin, A., 1993. Transfer RNAs with complementary anticodons: could they reflect early evolution of discriminative genetic code adapters? Proc. Natl. Acad. Sci. USA 90, 4723–4727.

Rodin, S., Rodin, A., Ohno, S., 1996. The presence of codon-anticodon pairs in the acceptor stem of tRNAs. Proc. Natl. Acad. Sci. USA 93, 4537–4542.

Taylor, F.J.R., Coates, D., 1989. The code within the codons. BioSystems 22, 177–187.

Trifonov, E.N., 1999a. Elucidating sequence codes: three codes for evolution. Ann. NY Acad. Sci. 870, 330–338.

Trifonov, E.N., 1999b. Glycine clock: Eubacteria first Archaea next, Protoctista, Fungi, Planta and Animalia at last. Gene Therapy Mol. Biol. 4, 313–322.

Trifonov, E.N., 2000. Leap into life's beginnings tracking the chronology of amino acids. Sci. Spectra 20, 62–71.

Trifonov, E.N., Bettecken, T., 1997. Sequence fossils, triplet expansion, and reconstruction of earliest codons. Gene 205, 1–6.

Weber, A.L., Miller, S.L., 1981. Reasons for the occurrence of the twenty coded protein amino acids. J. Mol. Evol. 17, 273–284.

Wong, J.T.-F., 1981. Co-evolution of genetic code and amino acid biosynthesis. Trends Bioch. Sci. 6, 33–36.

Wong, J.T.-F., 1988. Evolution of the genetic code. Microbiol. Sci. 5, 174–181.

Wächtershäuser, G., 1988. Before enzymes and templates: theory of surface metabolism. Microbiol. Rev. 52, 452–484.

Xia, T., SantaLucia, J., Burkard, M.E., Kierzek, R., Schroeder, S.J., Jiao, X., Cox, C., Turner, D.H., 1998. Thermodinamical parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. Biochemistry 37, 14719–14735.

Yockey, H.P., Löb, Walther, Stanley, L., 1997. Miller and prebiotic 'building blocks' in the silent electrical discharge. Persp. Biol. Med. 41, 125–131.