Available online at www.sciencedirect.com
ScienceDirect
Journal of Theoretical Biology 245 (2007) 169–174
Journal of Theoretical Biology
www.elsevier.com/locate/yjtbi

# The evolution of the genetic code took place in an anaerobic environment

Marco Archetti[a], Massimo Di Giulio[b,*]

[a]Department of Zoology, Oxford University, South Parks Road, OX1 3PS Oxford, UK
[b]Institute of Genetics and Biophysics 'Adriano Buzzati Traverso', CNR, Via P. Castellino, 111, 80131 Naples, Napoli, Italy

Received 10 July 2006; received in revised form 21 September 2006; accepted 22 September 2006
Available online 24 September 2006

**Abstract**

We have compared orthologous proteins from an aerobic organism, *Cytophaga hutchinsonii*, and from an obligate anaerobe, *Bacteroides thetaiotaomicron*. This comparison allows us to define the oxyphobic ranks of amino acids, i.e. a scale of the relative sensitivity to oxygen of the amino acid residues. The oxyphobic index (OI), which can be simply obtained from the amino acids' oxyphobic ranks, can be associated to any protein and therefore to the genetic code, if the number of synonymous codons attributed to the amino acids in the code is assumed to be the frequency with which the amino acids appeared in ancestral proteins. Sampling of the OI variable from the proteins of obligate anaerobes and aerobes has established that the OI value of the genetic code is not significantly different from the mean OI value of anaerobe proteins, while it is different from that of aerobe proteins. This observation would seem to suggest that the terminal phases of the evolution of genetic code organization took place in an anaerobic environment. This result is discussed in the framework of hypotheses suggested to explain the timing of the evolutionary appearance of the aerobic metabolism.
© 2006 Elsevier Ltd. All rights reserved.

*Keywords:* Origin of life; LUCA; Prebiotic environment; Respiration-early hypothesis; Amino acid substitution pattern; Oxygen sensitivity of amino acid residues; Oxyphobic index

## 1. Introduction

It is widely accepted that the evolution of aerobic metabolism was possible only after the oxygen released by photosynthesis became abundant in the atmosphere; a viewpoint which is for instance typically taken for granted in text books (Skelton, 1993; Alberts et al., 1994). Nevertheless, Kasting (1993) suggests a model for the evolution of the primitive atmosphere that envisages a first stage in which the atmosphere is reducing but has oases on the ocean surface with a relatively high concentration of oxygen, deriving from the photodissociation of water vapour (see also Klein, 1992; Holland, 2006). Therefore, the terminal phases of the origin of genetic code organization might have taken place in the presence of oxygen. Indeed, Castresana et al. (1994) suggest a scenario for the evolution of aerobic metabolism, which envisages that respiration was already present at the time of the last universal common ancestor (LUCA). On the basis of phylogenetic analysis, Castresana et al. (1994) propose that uroxidase might have allowed the LUCA to live in the oxygen oases suggested by Kasting (1993) where a rudimentary aerobic metabolism might therefore have been able to evolve. Thus, the respiration-early hypothesis (Castresana and Saraste, 1995) predicts that aerobic respiration might have evolved in the LUCA before large quantities of oxygen were released into the atmosphere by photosynthetic organisms, thanks to these very oxygen oases.

On the other hand, the terminal phases of the origin of genetic code organization would seem to coincide with, or at least overlap, those of the evolution of the first universal common ancestor (FUCA), i.e. the early phases of the evolution of the LUCA (Di Giulio, 2001). More explicitly, it is possible that not only did the terminal phases of genetic code origin coincide with the evolutionary stage of the FUCA, but also that the entire evolution of the LUCA

*Corresponding author. Tel.: +39 081 6132369; fax: +39 081 6132706.
*E-mail address:* digiulio@igb.cnr.it (M. Di Giulio).

0022-5193/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2006.09.025

might have been characterized by the evolution of genetic code structuring (Di Giulio, 2001). Therefore, it can be reasonably assumed that at least some phases of genetic code structuring coincided with some phases of the LUCA's evolution.

A number of ideas and methods have been introduced in order to investigate the physical environment in which the genetic code was structured (Di Giulio, 2000, 2005a–c). In the present paper, these ideas are extended in order to investigate the presence/absence of oxygen in the physical environment in which the genetic code evolved. While it is clear that such a study might only be a partial test of the respiration-early hypothesis (see Section 4), it might nevertheless enhance our knowledge of the physical environment in which the terminal phases of the origin of genetic code organization took place. This is why we have decided to carry out this study.

## 2. Materials and methods

The proteins from *Cytophaga hutchinsonii*, and *Bacteroides thetaiotaomicron* were taken from the Kyoto database on the web site: www.genome.jp/kegg/. All other proteins, from the three domains, were taken from the NBCI database using BLASTP (Altschul et al., 1997).

The species of aerobes and obligate anaerobes were identified by consulting Jacobs and Gerstein (1960) and Staley et al. (1984) and the web site: www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD = File&DB = genomeprj.

The two or more proteins were aligned using CLUSTALX (Thompson et al., 1997). Only the highly conserved regions were used in the analysis, while those containing gaps or which were poorly conserved were removed from these alignments.

For each amino acid (Table 2) or for each amino acid pair (Table 3), the significant deviation from the expected 50:50 ratio was determined by calculating the precise binomial probability.

The oxyphobic index (OI) which can be associated to any protein is defined by means of

$$\mathrm{OI} = \sum_{j=1}^{N} R_j/N,$$

where $R_j$ is the oxyphobic rank (Table 2) of the $j$th amino acid and $N$ is the protein length in amino acids.

Where not specified, the methods and ideas referred in equivalent analyses (Haney et al., 1999; McDonald et al., 1999; Di Giulio, 2000, 2005a–c) hold.

## 3. Results and discussion

### 3.1. Protein adaptive patterns in the presence/absence of oxygen

In order to obtain information regarding the pattern of amino acid substitution between an aerobic organism and an anaerobe, we have compared proteins from an aerobe, *C. hutchinsonii*, and from an obligate anaerobe,

Table 1
This matrix shows the amino acid substitution pattern between an aerobic organism, *Cytophaga hutchinsonii*, and an obligate anaerobe, *Bacteroides thetaiotaomicron*

| *Cytophaga hutchinsonii* (aerobic) | *Bacteroides thetaiotaomicron* (anaerobic) | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | A | C | D | E | F | G | H | I | K | L | M | N | P | Q | R | S | T | V | W | Y | |
| A | 2106 | 52 | 76 | 139 | 13 | 117 | 12 | 33 | 104 | 31 | 27 | 63 | 76 | 53 | 44 | 237 | 92 | 121 | 0 | 20 | 3416 |
| C | 26 | 257 | 0 | 3 | 4 | 4 | 1 | 9 | 2 | 7 | 4 | 4 | 0 | 1 | 3 | 22 | 11 | 21 | 0 | 2 | 381 |
| D | 55 | 1 | 1712 | 351 | 3 | 43 | 18 | 1 | 51 | 19 | 9 | 146 | 14 | 39 | 21 | 56 | 34 | 12 | 0 | 6 | 2591 |
| E | 118 | 5 | 316 | 2114 | 6 | 42 | 13 | 24 | 161 | 34 | 20 | 86 | 31 | 128 | 35 | 66 | 67 | 39 | 0 | 15 | 3320 |
| F | 17 | 4 | 4 | 8 | 1314 | 4 | 16 | 59 | 9 | 108 | 31 | 13 | 3 | 7 | 8 | 7 | 16 | 23 | 13 | 199 | 1863 |
| G | 113 | 6 | 52 | 41 | 4 | 2728 | 12 | 7 | 52 | 7 | 3 | 64 | 7 | 16 | 14 | 71 | 20 | 7 | 0 | 5 | 3229 |
| H | 13 | 5 | 19 | 19 | 16 | 8 | 556 | 9 | 29 | 15 | 1 | 45 | 9 | 20 | 21 | 25 | 11 | 11 | 2 | 49 | 883 |
| I | 58 | 5 | 9 | 35 | 47 | 2 | 7 | 2085 | 41 | 341 | 104 | 15 | 8 | 19 | 18 | 19 | 62 | 552 | 3 | 23 | 3453 |
| K | 112 | 4 | 84 | 191 | 10 | 54 | 41 | 40 | 2035 | 49 | 24 | 108 | 41 | 106 | 293 | 65 | 93 | 44 | 3 | 13 | 3410 |
| L | 68 | 11 | 15 | 39 | 95 | 14 | 11 | 359 | 57 | 2743 | 191 | 20 | 10 | 46 | 24 | 26 | 49 | 190 | 6 | 43 | 4017 |
| M | 17 | 6 | 6 | 18 | 19 | 4 | 0 | 65 | 22 | 154 | 669 | 7 | 3 | 26 | 9 | 11 | 27 | 50 | 4 | 13 | 1130 |
| N | 48 | 8 | 158 | 83 | 9 | 49 | 43 | 12 | 100 | 25 | 10 | 1119 | 13 | 46 | 39 | 74 | 63 | 21 | 0 | 15 | 1935 |
| P | 57 | 4 | 25 | 38 | 2 | 15 | 4 | 5 | 23 | 11 | 4 | 8 | 1380 | 22 | 12 | 28 | 23 | 13 | 1 | 2 | 1677 |
| Q | 51 | 4 | 37 | 142 | 6 | 22 | 28 | 15 | 103 | 38 | 33 | 50 | 10 | 780 | 63 | 41 | 41 | 28 | 2 | 10 | 1504 |
| R | 23 | 5 | 15 | 33 | 10 | 10 | 21 | 7 | 222 | 16 | 15 | 40 | 8 | 53 | 1554 | 20 | 25 | 15 | 4 | 9 | 2105 |
| S | 268 | 24 | 51 | 88 | 9 | 83 | 18 | 14 | 82 | 22 | 12 | 127 | 41 | 44 | 33 | 1222 | 188 | 38 | 2 | 14 | 2380 |
| T | 121 | 20 | 61 | 98 | 11 | 16 | 8 | 65 | 89 | 51 | 26 | 68 | 26 | 31 | 36 | 162 | 1402 | 95 | 1 | 19 | 2406 |
| V | 115 | 30 | 17 | 48 | 38 | 6 | 5 | 468 | 38 | 169 | 59 | 20 | 14 | 14 | 13 | 36 | 90 | 1924 | 2 | 20 | 3126 |
| W | 4 | 1 | 2 | 3 | 8 | 0 | 1 | 2 | 2 | 4 | 1 | 2 | 0 | 0 | 5 | 0 | 0 | 1 | 276 | 16 | 328 |
| Y | 17 | 8 | 3 | 12 | 147 | 6 | 36 | 25 | 16 | 25 | 9 | 12 | 2 | 8 | 13 | 9 | 8 | 14 | 15 | 1062 | 1447 |
| | 3407 | 460 | 2662 | 3503 | 1771 | 3227 | 851 | 3304 | 3238 | 3869 | 1252 | 2017 | 1696 | 1459 | 2258 | 2197 | 2322 | 3219 | 334 | 1555 | |

See text for further information.

*B. thetaiotaomicron*. These two organisms (i) are both Bacteroides, (ii) have an equivalent C + G content, and (iii) have equivalent optimal growth temperatures as they are mesophiles (Staley et al., 1984). It is therefore expected that the amino acid substitution pattern must be affected primarily by the presence/absence of oxygen because this is the most significant difference between these two organisms (McDonald et al., 1999; Di Giulio, 2005a).

We have therefore compared 93 pairs of orthologous proteins from *C. hutchinsonii* and from *B. thetaiotaomicron* for a total of 44,601 amino acids (Table 1). This sample seems to be fairly representative of the amino acid substitution pattern because it presents a total of 15,563 variable amino acid positions, with a mean identity percentage in these proteins of 65.1% (Table 1).

Table 2 shows how the total amino acid substitutions regarding a single amino acid are distributed between the two compared organisms. Table 3 reports the statistically significant deviations from the expected 50:50 ratio of the single amino acid substitutions observed in the sample of all the amino acid substitutions (Table 1). Equivalent analyses have already been conducted in a similar way for other variables (Haney et al., 1999; McDonald et al., 1999; Di Giulio, 2000, 2005a, c).

Table 2
The total amino acid substitution values for the single amino acids, i.e. the sum of all the amino acid substitution values towards (→X) or from (X→) a certain X amino acid (Table 1)

| Total values of substitutions | | | | P | Rank |
|---|---|---|---|---|---|
| →C | 203 | C→ | 124 | $1.5 \times 10^{-5}$ | 20 |
| →R | 704 | R→ | 551 | $1.8 \times 10^{-5}$ | 19 |
| →M | 583 | M→ | 461 | $1.8 \times 10^{-4}$ | 18 |
| →Y | 493 | Y→ | 385 | $3.0 \times 10^{-4}$ | 17 |
| →E | 1389 | E→ | 1206 | $3.6 \times 10^{-4}$ | 16 |
| →N | 898 | N→ | 816 | 0.050 | 15 |
| →V | 1295 | V→ | 1202 | 0.066 | 10 |
| →D | 950 | D→ | 879 | 0.10 | 10 |
| →P | 316 | P→ | 297 | 0.47 | 10 |
| →W | 58 | W→ | 52 | 0.63 | 10 |
| →G | 499 | G→ | 501 | 0.97 | 10 |
| →A | 1301 | A→ | 1310 | 0.87 | 10 |
| →Q | 679 | Q→ | 724 | 0.24 | 10 |
| →H | 295 | H→ | 327 | 0.21 | 10 |
| →T | 920 | T→ | 1004 | 0.059 | 10 |
| →F | 457 | F→ | 549 | $4.1 \times 10^{-3}$ | 5 |
| →I | 1219 | I→ | 1368 | $3.6 \times 10^{-3}$ | 4 |
| →L | 1126 | L→ | 1274 | $2.7 \times 10^{-3}$ | 3 |
| →K | 1203 | K→ | 1375 | $7.5 \times 10^{-4}$ | 2 |
| →S | 975 | S→ | 1158 | $8.2 \times 10^{-5}$ | 1 |

The direction of substitution indicated by the arrow is: aerobic amino acid → anaerobic amino acid. *P* indicates the probability calculated using binomial distribution of the deviation from the expected 50:50 ratio for each amino acid. The ranks are simply the oxyphobic ranks of amino acids, defined on the basis of *P*. For instance, the sum of amino acid substitutions leading to C (Cys) in the anaerobic organism significantly exceeds that of the aerobic organism and, having a smaller *P*, it is assigned the highest rank (20 units). Whereas, the sum of the amino acid substitutions leading to S (Ser) of the anaerobic organism is significantly lower than that of the aerobic organism, and having a smaller *P*, it is assigned the lowest rank (1 unit), and so on. The rank of 10 units is the mean rank of amino acids having a non-significant *P*.

Table 3
All the statistically significant deviations from the expected 50:50 ratio of the single amino acid pairs

| | | | | P |
|---|---|---|---|---|
| NS | 127 | SN | 74 | $2.3 \times 10^{-4}$ |
| AL | 68 | LA | 31 | $2.6 \times 10^{-4}$ |
| KR | 222 | RK | 293 | $2.0 \times 10^{-3}$ |
| IM | 65 | MI | 104 | $3.3 \times 10^{-3}$ |
| AC | 26 | CA | 52 | $4.3 \times 10^{-3}$ |
| DK | 84 | KD | 51 | $5.7 \times 10^{-3}$ |
| FY | 147 | YF | 199 | $6.0 \times 10^{-3}$ |
| DT | 61 | TD | 34 | $7.3 \times 10^{-3}$ |
| IV | 468 | VI | 552 | $9.3 \times 10^{-3}$ |
| AI | 58 | IA | 33 | $1.1 \times 10^{-2}$ |
| AR | 23 | RA | 44 | $1.4 \times 10^{-2}$ |
| ET | 98 | TE | 67 | $1.9 \times 10^{-2}$ |
| DI | 9 | ID | 1 | $2.1 \times 10^{-2}$ |
| KP | 23 | PK | 41 | $3.3 \times 10^{-2}$ |
| LY | 25 | YL | 43 | $3.8 \times 10^{-2}$ |
| IR | 7 | RI | 18 | $4.3 \times 10^{-2}$ |
| QV | 14 | VQ | 28 | $4.4 \times 10^{-2}$ |

The first amino acid refers to the amino acid in the anaerobic organism, the second amino acid to that in the aerobe. For instance, the substitution NS = 127 (Table 1) indicates that 127 times N (Asn) has been substituted by S (Ser) passing from the anaerobic to the aerobic organism. The probability *P* was calculated using binomial distribution.

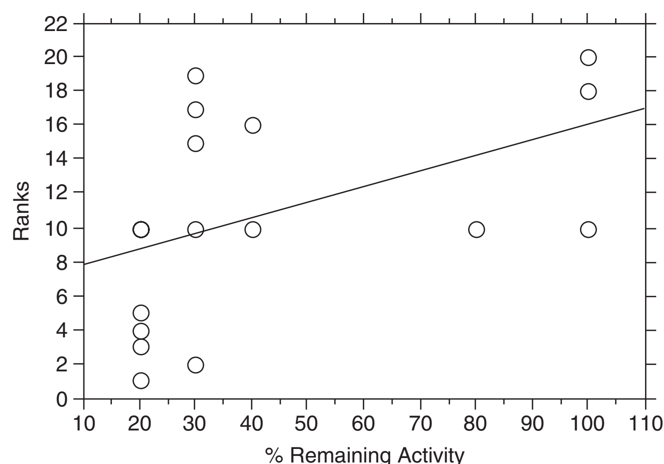


Fig. 1. The correlation between the values of the amino acids' oxyphobic ranks (Table 2) and the values of the percentages of the remaining activity of lysozyme in the presence of $CCl_3O_2$ and amino acids (Willson, 1985, Table IV, p. 62).

### 3.2. *The amino acid substitution pattern seems to be correlated to the scale measuring the amino acids' oxygen sensitivity*

At least for two amino acids, Cys and Met, which are particularly oxidation-sensitive because they contain sulphur (Berlett and Stadtman, 1997), the total amino acid

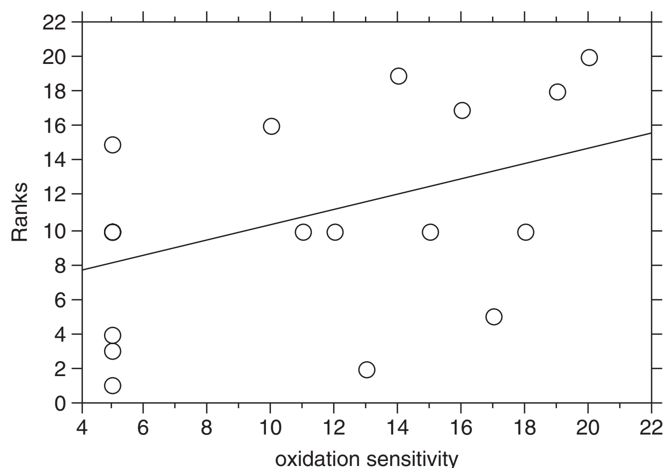


Fig. 2. The correlation between the values of the amino acids' oxyphobic ranks (Table 2) and the values of the ranks (oxidation sensitivity) attributed to the amino acids in Table 1 of Berlett and Stadtman (1997, p. 20315) in the order in which they appear in that table. For instance, Cys, which is the first amino acid in that table (Berlett and Stadtman, 1997, Table 1, p. 20315) has been attributed the value of 20 units. Met, which is the second amino acid in the table, has been attributed the value of 19 units, and so on. The amino acids not appearing in Bertlett and Stadtman's Table 1 have been attributed with a mean rank of 5 units.

substitutions regarding the single amino acids seem to reflect this sensitivity. Indeed, the ranks of these two amino acids are among the highest observed values (Table 2).

In order to ascertain whether or not the comparison, and therefore also the amino acid substitution pattern (Tables 1–3), reflects the amino acids' oxidation sensitivity, we have performed two correlations (Figs. 1 and 2 and their legends). The first is between the oxyphobic ranks of amino acids (Table 2) reflecting the amino acids' presumed oxygen sensitivity and the inactivation of lysozyme by $CCl_3O_2$ in the presence of amino acids used as antioxidants, i.e. the percentage of remaining activity of lysozyme (Willson, 1985). Fig. 1 reports this correlation, which is statistically significant ($r = +0.476$, $F = 5.27$, df = 19, $p = 0.034$). The second correlation is between the oxyphobic ranks of amino acids (Table 2) and the ranks of amino acids reflecting their oxidation sensitivity (Berlett and Stadtman, 1997; see legend of Fig. 2). Fig. 2 reports this second correlation, which is also statistically significant ($r = +0.448$, $F = 4.52$, df = 19, $p = 0.048$).

These two correlations, although only marginally significant and not independent of each other given the correlation existing between the percentage of the

Table 4
The data referring to the sampling of the values of the oxyphobic index variable (OI) of proteins from aerobic and obligate anaerobic organisms

| | Aln | Aerobic | | | Anaerobic | | |
|---|---|---|---|---|---|---|---|
| | | *n* | Mean | Std.dev. | *n* | Mean | Std.dev. |
| methionyl-tRNA synthetase | 332 | 60 | 9.5156 | 0.24 | 30 | 9.5481 | 0.31 |
| glutamate dehydrogenase | 240 | 38 | 9.5376 | 0.44 | 16 | 9.6891 | 0.36 |
| leucyl-tRNA synthetase | 583 | 45 | 9.4658 | 0.28 | 21 | 9.7120 | 0.32 |
| threonyl-tRNA synthetase | 504 | 53 | 9.4588 | 0.33 | 19 | 9.7007 | 0.27 |
| tryptophan synthetase | 324 | 39 | 9.7725 | 0.35 | 10 | 9.8192 | 0.24 |
| adenylosuccinate synthetase | 138 | 54 | 9.5407 | 0.38 | 24 | 9.7720 | 0.30 |
| CTP synthetase | 332 | 58 | 9.4383 | 0.24 | 26 | 9.5390 | 0.24 |
| ribosomal protein | 209 | 22 | 9.7154 | 0.32 | 20 | 9.7791 | 0.38 |
| thiamine biosynthesis protein | 246 | 23 | 9.4092 | 0.54 | 25 | 9.5003 | 0.63 |
| pyruvate dehydrogenase | 384 | 26 | 9.4682 | 0.34 | 9 | 9.6588 | 0.37 |
| DNA gyrase subunit A | 662 | 16 | 9.5520 | 0.17 | 30 | 9.5693 | 0.27 |
| prolyl-tRNA synthetase | 180 | 61 | 9.5679 | 0.47 | 28 | 10.0331 | 0.29 |
| arginyl-tRNA synthetase | 197 | 37 | 9.1882 | 0.55 | 28 | 9.3168 | 0.37 |
| tyrosyl-tRNA synthetase | 242 | 42 | 9.3421 | 0.40 | 15 | 9.5046 | 0.35 |
| isoleucyl-tRNA synthetase | 366 | 47 | 9.6639 | 0.31 | 18 | 9.7279 | 0.29 |
| valyl-tRNA synthetase | 409 | 46 | 9.7022 | 0.25 | 29 | 9.8190 | 0.24 |
| asparaginyl-tRNA synthetase | 246 | 40 | 9.8044 | 0.41 | 20 | 10.1711 | 0.24 |
| ferredoxin reductase | 207 | 45 | 9.4288 | 0.36 | 31 | 9.8763 | 0.29 |
| cysteine synthase | 264 | 35 | 9.0123 | 0.40 | 20 | 9.2873 | 0.31 |
| methionine adenosyltransferase | 292 | 31 | 9.1702 | 0.23 | 6 | 9.7401 | 0.14 |
| fumarate reductase | 195 | 41 | 10.3335 | 0.31 | 10 | 10.4426 | 0.24 |
| gamma-glutamyl phosphate reductase | 292 | 55 | 9.2125 | 0.49 | 17 | 9.3335 | 0.27 |
| arginine biosynthesis protein | 248 | 50 | 9.2663 | 0.42 | 30 | 9.4270 | 0.38 |
| selenophosphate synthetase | 188 | 16 | 9.0439 | 0.39 | 17 | 9.1232 | 0.35 |
| homocysteine *S*-methyltransferase | 164 | 46 | 9.5274 | 0.39 | 13 | 9.7227 | 0.59 |
| Histidyl-tRNA synthetase | 136 | 34 | 9.3546 | 0.56 | 31 | 9.6014 | 0.53 |
| argininosuccinate synthase | 308 | 41 | 9.5413 | 0.31 | 30 | 9.6012 | 0.28 |

Aln indicates the length of the alignment in amino acids; *n* indicates the number of species used by aerobic and obligate anaerobic organisms with the relative value of the mean and standard deviation.

remaining activity of lysozyme and the amino acids' oxidation sensitivity ($r = +0.684$, $F = 15.79$, $df = 19$, $p = 9.0 \times 10^{-4}$), nevertheless seem to clearly show that the oxyphobic ranks of amino acids (Table 2) reflecting the amino acid substitution pattern truly reflect the oxygen sensitivity of amino acid residues, i.e. they measure what they are supposed to measure.

### 3.3. The genetic code originated in an anaerobe

The mean protein defined on the basis of amino acid composition imposed by the number of synonymous codons in the genetic code has an OI value ($OI_{code}$) of 9.639 (see Section 2). In calculating this value, Trp for instance, which is specified by a single codon in the genetic code, is assumed to have a frequency in ancestral proteins of 1/61, while Arg with 6 codons in the genetic code must have had a frequency of 6/61, and so on (for a justification of this assumption, see Di Giulio, 2000). We can therefore test whether the value of the $OI_{code} = 9.639$ is typical of the proteins of aerobic organisms or of anaerobes. In order to answer this question we must evidently sample the OI variable in these two groups of organisms. Table 4 reports the main information regarding the sampling of the OI variable for 27 different orthologous proteins from aerobic and obligate anaerobic organisms. During this sampling, a further 41 orthologous proteins displayed a mean OI value ($OI_{mean}$) higher in aerobes than in anaerobes, and these proteins were all eliminated from the analysis. The mean value of the OI for the 27 proteins from the aerobes sample is $OI_{mean} = 9.483$, while that of the obligate anaerobes is $OI_{mean} = 9.667$ (Table 4). And while the $OI_{mean} = 9.483$ of aerobe proteins is significantly different from the $OI_{code} = 9.639$ ($t = -3.047$, $df = 26$, $0.001 < p < 0.01$), the mean of the anaerobe proteins, i.e. the value $OI_{mean} = 9.667$, is not different from $OI_{code} = 9.639$ ($t = +0.531$, $df = 26$, $p = 0.60$) (Balaam, 1972; Di Giulio, 2000, 2005b, c). This might indicate that the genetic code might have evolved in an anaerobic 'organism', given that the $OI_{code}$ value is not significantly different from the $OI_{mean}$ of proteins from anaerobes but it is different from that of aerobes.

## 4. Conclusions

The main result of the analysis here conducted seems to establish that the terminal phases of genetic code evolution must have taken place in an anaerobic environment. Although this result may seem to state the obvious because the primordial atmosphere is commonly assumed not to have had large quantities of oxygen (Klein, 1992; Schopf, 1993; Holland, 2006), the presence of oxygen oases on the surface of the primitive ocean (Klein, 1992; Kasting, 1993; Holland, 2006) makes it reasonable to suggest that the LUCA might have been an aerobe or, at least, an 'organism' possessing a rudimentary aerobic metabolism (Castresana et al., 1994; Castresana and Saraste, 1995). The observations made in the present paper do not substantiate this respiration-early hypothesis (Castresana and Saraste, 1995) because the mean protein which can be associated to the genetic code seems to be more the typical one of anaerobes than that of the aerobes. However, the test performed here could refer to a different evolutionary period than the one referred to by the respiration-early hypothesis and, therefore, further analyses would evidently be necessary to falsify this intriguing suggestion. Moreover, high concentrations of oxygen are not necessarily needed to enable aerobic respiration, as can be seen in some cytochrome oxidases of certain organisms which work at very low oxygen pressures (Castresana and Saraste, 1995). Therefore, if these were also the conditions in the LUCA, then the test here performed might not be suitable because it might not be sensitive to these low oxygen pressures. On the other hand, if this is not the case, we must conclude that it is more likely that the genetic code evolved in an anaerobic environment.

## References


Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P., 1994. Molecular Biology of the Cell, third ed. Graland Publishing.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl. Acid. Res. 25, 3389–3402.

Balaam, L.N., 1972. Fundamentals of Biometry. George Allen & Unwin, London, pp. 120–142.

Berlett, B.S., Stadtman, E.R., 1997. Protein oxidation in aging, disease, and oxidative stress. J. Biol. Chem. 272, 20313–20316.

Castresana, J., Saraste, M., 1995. Evolution of energetic metabolism: the respiration-early hypothesis. Trends Biochem. Sci. 20, 443–448.

Castresana, J., Lubenn, M., Saraste, M., Higgins, D.G., 1994. Evolution of cytochrome oxidase, an enzyme older than atmospheric oxygen. EMBO J. 13, 2516–2525.

Di Giulio, M., 2000. The late stage of genetic code structuring took place at a high temperature. Gene 261, 189–195.

Di Giulio, M., 2001. The non-universality of the genetic code: the universal ancestor was a progenote. J. Theor. Biol. 209, 345–349.

Di Giulio, M., 2005a. A comparison of protein from *Pyrococcus furiosus* and *Pyrococcus abyssi*: bariphily in the physicochemical properties of amino acids and in the genetic code. Gene 346, 1–6.

Di Giulio, M., 2005b. The ocean abysses witnessed the origin of the genetic code. Gene 346, 7–12.

Di Giulio, M., 2005c. Structuring of the genetic code took place at acidic pH. J. Theor. Biol. 237, 219–226.

Haney, P.J., Badger, J.H., Buldak, G.L., Reich, C.I., Woese, C.R., Olsen, G.J., 1999. Thermal adaptation analyzed by comparison of proteins sequences from mesophilic and extremely thermophilic *Methanococcus* species. Proc. Natl Acad. Sci. USA 96, 3578–3583.

Holland, H.D., 2006. The oxygenation of the atmosphere and oceans. Philos. Trans. R. Soc. London B Biol. Sci. 361, 903–915.

Jacobs, M.J., Gerstein, M.J., 1960. Handbook of Microbiology. D. Van Nostrand Company, Inc., London.

Kasting, J.F., 1993. Earth's early atmosphere. Science 259, 920–926.

Klein, C., 1992. In: Schopf, W., Klein, C. (Eds.), The Proterozoic Biosphere. Cambridge University Press, Cambridge, pp. 137–174.

McDonald, J.H., Grasso, A.M., Rejto, L.K., 1999. Patterns of temperature adaptation in proteins from *Methanococcus* and *Bacillus*. Mol. Biol. Evol. 16, 1785–1790.

Schopf, J.W., 1993. Microfossils of the early Archean apex chert: new evidence of the antiquity of life. Science 260, 640–646.

Skelton, P., 1993. Evolution. A Biological and Paleontological Approach. Addison-Wesley, The Open University, Reading, MA.

Staley, J.T., Bryant, M.P., Plennig, N., Holt, J.G., 1984. In: Hensyl, W.R. (Ed.), Bergey's Manual of Systematic, vol. 3. Lippincut Williams & Wilkins, Philadelphia.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucl. Acids Res. 25, 4876–4882.

Willson, R.L., 1985. Organic peroxy free radicals as ultimate agents in oxygen toxicity. In: Helmut, S. (Ed.), Oxidative Stress. Academic Press, New York, pp. 41–72.