Available online at www.sciencedirect.com
SCIENCE DIRECT
Gene 346 (2005) 7–12
GENE
SECTION
EVOLUTIONARY GENOMICS
www.elsevier.com/locate/gene

The ocean abysses witnessed the origin of the genetic code

Massimo Di Giulio*

Institute of Genetics and Biophysics 'Adriano Buzzati Traverso', CNR, Via G. Marconi 10, 80125 Naples, Napoli, Italy

Received 20 June 2004; accepted 26 July 2004
Available online 31 December 2004

In memory of Franco

Abstract

The comparison of proteins from a non-barophilous and a barophilous organism makes it possible to define the barophily ranks of amino acids. The correlation of these ranks with the number of codons attributed to amino acids in the genetic code, together with another straightforward argument based on an optimisation percentage of a barophily index (BI) (easily defined by barophily ranks) which can be associated to the genetic code table, suggest that the genetic code originated under high hydrostatic pressure. Moreover, as the BI value can be calculated for the sequence of any protein, it also makes it possible to define the BI for the genetic code if the number of codons attributed to the amino acids in the code is assumed to be the frequency with which the amino acids appeared in ancestral proteins. Finally, sampling the BI variable between many non-barophile organisms and from many proteins of a single non-barophile organism leads to the conclusion that the BI value of the genetic code is not typical of these organisms. Whereas, since the genetic code BI value is statistically higher than that of these non-barophile organisms, it supports the hypothesis that genetic code structuring took place under high hydrostatic pressure.

Keywords: Hydrostatic pressure; Amino acid barophily ranks; Protein barophily index; Hydrothermal vents; Prebiotic environment; Origin of life

1. Introduction: the search for the environment in which life originated

Many difficulties are encountered when attempting to falsify the theories suggested to explain the origin of life. This stems from the absolute lack of any reference point regarding, for instance, the type of chemical structure that first triggered the process that was to become the origin of life proper. Other obstacles lie, for example, in the consideration that the set of chemical reactions characterising the autotrophic theory (Wachtershauser, 1988, 1990) were said to have played an important role in enriching the primordial soup (Bada and Lazcano, 2002), as instead is envisaged by the heterotrophic theory of the origin of life (Lazcano and Miller, 1996). This would clearly diminish the possibilities for testing these two theories because, if abiotic synthesis reactions typical of the autotrophic theory may have contributed to enriching the primordial soup (Bada and Lazcano, 2002), then experimental evidence of the plausibility of some synthesis reactions predicted by the autotrophic theory would not actually lead to discriminating between the two theories. Another singular difficulty, which cannot nevertheless be excluded, derives from the consideration that on a planet like the Earth, the different theories made to explain the origin of life might be 'simultaneously' true. In other words, it is conceivable that life can originate in a number of different ways and hence the various theories may not be wrong a priori but only a posteriori, in that only one of these theories might have ultimately led to the establishment of life as we now know it. For example, it has been shown that even if life on this planet had originated independently no less than 10 times, it is more than likely, and as the sole effect of chance, that only 1 of the 10 forms would ultimately have survived (Raup and Valentine, 1983). Clearly life in these 10 hypothetical independent origins

Abbreviations: BI, barophily index; *LUCA*, the last universal common ancestor; PAI, hydrostatic pressure asymmetry index.

* Tel.: +39 081 7257313; fax: +39 081 5936123.
E-mail address: digiulio@iigb.na.cnr.it.

0378-1119/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2004.07.045

may have originated differently according to the predictions of some of the different theories. Therefore, testing the theories suggested for explaining the origin of life is extremely problematic.

One way to remove many of the difficulties encountered in falsifying the theories put forward to explain the origin of life seems to me to lie in trying to identify the environment in which life originated. Clearly, once the environment in which life originated has been identified, at least some of the theories suggested as an explanation of this origin would disappear because this environment would turn out to be incompatible with some of these theories. Moreover, the attempt to identify the environment in which life originated should preferably be based on biology as, unlike physics and chemistry, it includes the history of the living world and could preserve the 'memory' of the environment in which the origin of life took place.

A first attempt in this direction was made by Galtier et al. (1999) who, by exploiting the correlation between the optimal growth temperatures of various organisms and the G+C content of the ribosomal RNA and reconstructing the rRNA sequence of the last universal common ancestor (LUCA), were able to claim that the LUCA was a mesophile. These authors (Galtier et al., 1999) were thus able to make inferences on the environment in which the earliest organisms lived and in which life presumably originated. Di Giulio (2000, 2001, 2003a,b) extended the logic introduced by Galtier et al. (1999) to include proteins, which made it possible to define that the temperature at which the LUCA lived and at which the genetic code originated seems indeed to be the one at which hyperthermophile organisms live.

These methods (Di Giulio, 2000, 2001, 2003a,b) can be conceptually applied to any physical or chemical variable referring to the environment in which the earliest organisms lived and, therefore, could allow inferences to be made on the place where life originated. In the present paper, one of these methods (Di Giulio, 2000) is applied to hydrostatic pressure in the hope of clarifying whether life originated on the coastline, in lagoons or in the depths of the oceans.

2. Materials and methods

The proteins used in the comparison between *Pyrococcus furiosus* and *Pyrococcus abyssi* were taken from the Kyoto database at the site www.genome.ad.jp/kegg/. All the other proteins used in the analysis were taken from the NBCI by means of BLASTP (Altschul et al., 1997).

The alignments between orthologous proteins were constructed using CLUSTALX (Thimpson et al., 1997) with the default parameters. Only the highly conserved regions between amino acid sites, which were also highly conserved, were used in the analysis, whereas regions containing gaps or which were badly aligned were all eliminated.

3. Results and discussion

3.1. The construction of a barophily index

I have recently compared 141 orthologous proteins from *P. furiosus* and *P. abyssi* (Di Giulio, 2005). The aim of this comparison was to identify the amino acid substitution pattern between a non-barophile organism (*P. furiosus*: isolated at a depth of 0.5 m) and a barophile or barotolerant organism (*P. abyssi*: isolated at a depth of 2000 m). Some of these results are reported in Table 1. It was thus possible to identify which amino acids (Arg, Ser, Val, Asp, and Gly) are preferentially and significantly used in the barophile organism and which (Tyr, Gln, Thr, Ile, Pro, Lys and Asn) are preferentially used in the non-barophile organism (Table 1) (Di Giulio, 2005). This identification also makes it possible, on the basis of probability values, to attribute a barophily rank naturally to all the amino acids and similarly to what has already been done for temperature (Di Giulio, 2000). This yields a range from the most barophilic, arginine with a rank of 20, to the least barophilic, tyrosine with a rank of 1 (Table 1). Once again

Table 1
Results of the total amino acid substitutions involving the single amino acids and deriving from the comparison of proteins from *P. furiosus* and *P. abyssi* (Di Giulio, 2005)

Substitution direction: non-Barophilic-AA→Barophilic-AA		χ^2	*P*	Barophily ranks
AAs→R=831	R→AAs=633	26.50	<0.0001	20
AAs→S=628	S→AAs=476	20.66	<0.0001	19
AAs→V=910	V→AAs=775	10.66	0.0011	18
AAs→D=513	D→AAs=432	6.78	0.0092	17
AAs→G=295	G→AAs=238	5.88	0.015	16
AAs→L=621	L→AAs=577	1.54	0.21	11.5
AAs→H=106	H→AAs=103	0.02	0.89	11.5
AAs→F=209	F→AAs=212	0.00	1.00	11.5
AAs→M=204	M→AAs=215	0.24	0.62	11.5
AAs→E=936	E→AAs=959	0.26	0.61	11.5
AAs→A=449	A→AAs=481	1.04	0.31	11.5
AAs→C=10	C→AAs=15	–	0.21	11.5
AAs→W=24	W→AAs=33	–	0.14	11.5
AAs→N=399	N→AAs=449	2.84	0.092	7
AAs→K=1031	K→AAs=1135	4.90	0.027	6
AAs→P=111	P→AAs=151	5.80	0.016	5
AAs→I=819	I→AAs=961	11.16	0.00083	4
AAs→T=301	T→AAs=395	12.42	0.00042	3
AAs→Q=173	Q→AAs=255	15.32	<0.0001	2
AAs→Y=130	Y→AAs=205	16.34	<0.0001	1

The direction of the substitution is, as indicated, non-barophilic amino acid (AA)→barophilic amino acid. For instance, AAs→R=831 means that in the comparison 831 amino acid substitutions were observed, which from the amino acids (AAs) of the non-barophile organisms 'transformed' into the same number of arginines (R) of the barophile organism (Di Giulio, 2005). The χ^2 with one degree of freedom is calculated on the basis of the expected frequencies in the ratio 50:50 (Di Giulio, 2005); whereas, for two amino acids (Cys=C and Trp=W) the probability was calculated using the binomial formula (Di Giulio, 2005). The last column reports the barophily ranks: the values of 11.5 units are simply the mean of the ranks of the respective amino acids having insignificant probability values. See text for further information.

in analogy with what has already been achieved for temperature (Di Giulio, 2000), a barophily index (BI) can be defined which can be associated to the sequence of any protein:

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

where R_j are the barophily ranks (Table 1) of the 20 amino acids and N is the length of the protein considered, i.e. the number of amino acids of which it is made up.

3.2. Hydrostatic pressure is linked to the organisation of the genetic code

I have recently shown that the same physicochemical properties of amino acids (polarity and 'size' (molecular weight)) that were important in the origin of the genetic code seem to have been important in explaining the barophily of amino acids (Di Giulio, 2005). Moreover, I have found that the hydrostatic pressure asymmetry index (PAI), which is a measure of the barophilicity of amino acids, positively and significantly correlates with the number of codons attributed to amino acids in the genetic code (Di Giulio, 2005). All this seems to indicate that hydrostatic pressure was important in organising the genetic code.

Barophily ranks (Table 1) obviously correlate with PAI values (r=+0.750, F=23.18, df=19, P=10^{-4}). However, barophily ranks correlate only marginally with the number of codons attributed to amino acids in the genetic code (r=+0.398, F=3.39, df=19, P=0.082). This correlation becomes significant (r=+0.623, F=6.36, df=11, P=0.030) only if we eliminate the eight amino acids that do not seem important in defining barophily (Table 1). Nevertheless, it is only the partial correlation (Table 2 and its legend) between the number of codons, barophily ranks (Table 1) and amino acid polarity (Woese et al., 1966) variables that clearly points out the close relationship between the number of codons attributed to the amino acids in the genetic code and the barophily ranks (Table 2). If the influence of amino acid polarity values is removed from the correlation between the number of codons and the barophily ranks (as what happens in the partial correlation) then the significance of this correlation increases with respect to the one reported above (Table 2 and its legend). For instance, in the multiple regression of barophily ranks (Table 1) versus the number of codons attributed to amino acids in the genetic code and their polarity values (Woese et al., 1966), we obtain a highly significant regression coefficient relative to the number of codons (β=+3.759, t=+3.651, df=11, P=0.0053). (An equivalent result is obtained by making, in the multiple regression, the number of codons variable a dependent and not independent variable.) This all seems to support the conclusion that the number of codons attributed to amino acids in the genetic code was at least partly determined by barophily ranks and, therefore, by hydrostatic pressure (Di Giulio, 2005). However, the following simple argument seems to make the latter claim even more likely.

If the number of codons attributed to amino acids in the genetic code was guided by natural selection (Hasegawa and Miyata, 1980; Di Giulio, 1989, 2000, 2005; Taylor and Coates, 1989; Dufton, 1997), then it seems sensible to consider the number of codons as reflecting the frequency with which amino acids appeared in ancestral proteins. This is because if the number of codons was a variable guided by natural selection, then it should have made the number of codons attributed to amino acids in the genetic code 'equivalent' to their frequency in ancestral proteins, i.e. optimised. This therefore justifies the barophily index (BI), if associated to the genetic code, on the basis of the number of codons codifying for the various amino acids in the code. More explicitly, the number of codons attributed to amino acids in the genetic code would represent the frequency with which amino acids appeared in ancestral proteins. For example, arginine which is codified by six codons in the genetic code would have had a frequency of 6/61 in ancestral proteins. This makes it possible to associate a mean protein, and hence its BI value, to the genetic code (BI_{code}). This value is BI_{code}=11.639. We can also define a mean BI (BI_{mean}), i.e. of a completely randomised genetic code in attributing the number of codons to amino acids, such as the one in which all the amino acids are used in proteins with the same frequency (1/20). The latter value is BI_{mean}=10.500. Finally, we can define a maximum BI (BI_{max}) as the one associating the highest barophily rank values (Table 1) with the highest number of codon values. For instance, arginine with a rank of 20 (Table 1) is associated with a number of codons equal to 6, and likewise for serine with a rank of 19 and valine with a rank of 18, whereas aspartic acid with a rank of 17 would be given a plurality of 4, and so on to tyrosine, with a rank of 1 (Table 1), which would be attributed with a plurality of 1. The BI_{max} defined in this way is 12.500. Finally, the following percentage can

Table 2
Partial correlation coefficient matrix between the variables: (1) number of codons attributed to the amino acids in the genetic code, (2) barophily ranks (Table 1) and (3) amino acid polarity (Woese et al., 1966)

	Codon number	Woese polarity	Barophily ranks
Codon number	+1.000	−0.622	+0.773
Woese polarity		+1.000	+0.638
Barophily ranks			+1.000

The partial correlation was carried out only on 12 observations, and therefore excluding the eight amino acids with a barophily rank equal to 11.5 (Table 1), i.e. the amino acids having neither a barophilic nor a non-barophilic character (Table 1). The reported coefficients are significant at least at the level of significance of 5%.

be calculated to indicate the shift from the mean code which the genetic code underwent on the basis of the number of codons attributed to amino acids: $[(BI_{code}-BI_{mean})/(BI_{max}-BI_{mean})]\times 100$. This percentage is 57.0%. This seems to indicate a considerable distancing from the randomised mean code and, hence, an optimisation towards the code with BI_{max}. Indeed, in a Mann–Whitney test (Balaam, 1972) the sample of barophily ranks (Table 1) with the plurality imposed by the genetic code (in other words the ones used to calculate the BI_{code}) comes from the same population that yielded the sample associating the barophily ranks with the number of codons in the genetic code so as to obtain the maximum BI value (BI_{max}) (n_1=61, n_2=61, U=1643, U'=2078, Z=−1.141 (corrected for ties), P=0.13). This indicates that the two samples are not significantly different. Moreover, this optimisation percentage which exceeds 50% seems to indicate this if we consider that the same calculation performed with the thermophily index (TI) (Di Giulio, 2000) yields a percentage of only 7.5% (TI_{mean}=10.500, TI_{code}=10.684, TI_{max}=12.943) even if the genetic code nevertheless seems to have been structured at high temperature (Di Giulio, 2000). Therefore, perhaps this should also be true for hydrostatic pressure, if not more so.

Overall, the observations reported in this section together with those presented in another work (Di Giulio, 2005) seem to support the hypothesis that the origin of the genetic code took place under high hydrostatic pressure. However, it would be preferable to have more direct evidence of this, which is the aim of the following section.

3.3. The genetic code did not originate in a non-barophile 'organism' and was therefore structured under high hydrostatic pressure

We may wonder whether the value of the BI_{code}=11.639 is a typical value of non-barophile or barophile organisms. The answer to this question clearly lies in sampling the BI variable for these organisms. Unfortunately, it must be stated at once that there is no sufficiently high number of sequences from barophile organisms for this sampling to be performed. However, what we can adequately do is to sample the BI variable of non-barophile organisms. This has been done in Table 3 in which a total of 799 proteins distributed over 23 different orthologous proteins were used to calculate, from the 23 mean BI values (Table 3), the mean of means of BI values, which turned out to be 11.273 with a standard deviation of 0.238. A t-test (Balaam, 1972) indicates that the mean of means (11.273) cannot be considered as extracted from a populating having a mean of $\mu=BI_{code}$=11.639 (Di Giulio, 2000). We obtain a highly significant value of t=−7.375 ($t=(11.273-11.639)/(0.238/(23)^{1/2})$, df=22, $P<<10^{-3}$) which refutes the null hypothesis of equality between the two means. More correctly, as a result of the interpretation given to the statistical test, the test should have been

Table 3
Sampling of the barophily index variable (BI)

Proteins	Alignment length amino acid number	Mean BI	Standard deviation	Number of sequences
Carbamoyl-phosphate synthetase	501	11.137	0.268	67
Signal recognition particle 54 kDa	356	11.291	0.437	53
Alanyl-tRNA synthetase	288	11.336	0.404	55
Acetolactate synthase	412	11.188	0.414	23
Acetylornithine aminotransferase	226	11.289	0.408	19
Inosine-5′-monophosphate dehydrogenase	338	11.642	0.289	45
Beta subunit of tryptophan synthase	354	11.285	0.373	24
S-adenosyl-L-homocysteine hydrolase	343	11.257	0.302	31
Cell division protein (ftsZ)	296	11.545	0.372	26
Glutamate-1-semialdehyde aminotransferase	321	11.310	0.379	24
Glutamine-fructose-6-phosphate transferase	339	11.126	0.270	28
Phosphoribosylamine-glycine ligase	243	11.643	0.455	29
Anthranilate synthetase alpha-subunit	268	11.668	0.444	24
Threonine synthase	179	11.049	0.470	21
Aspartate aminotransferase	302	10.685	0.411	19
Succinyl-CoA synthetase beta subunit	288	11.223	0.374	27
Adenylosuccinate synthase	240	11.538	0.204	26
Acetyl-CoA synthetase	257	11.009	0.291	23
Valyl-tRNA synthetase	434	10.987	0.387	48
Glutamate dehydrogenase (GDH)	293	11.263	0.382	47
CTP synthetase	356	11.249	0.255	40
Argininosuccinate lyase	305	11.470	0.504	53
Glucosamine-fructose-6-phosphate aminotransferase	318	11.091	0.348	47

Proteins from the three domains of life (Bacteria, Archaea and Eukarya) were aligned as reported in Materials and methods. The BI values for each individual protein sequence were then calculated. The mean of these values (mean BI) and their relative standard deviation were then calculated. Elsewhere, the meaning of the table is self-explanatory.

conducted under the null hypothesis of BI_{code} being greater than the mean of means of BI for non-barophiles. Obviously, the results reported above are equivalent to those obtainable under the latter null hypothesis.[1]

The conclusion is that the genetic code did not originate in a non-barophilous organism because the BI_{code}=11.639 is statistically different from the grand average (BI=11.273) of the BI mean values for the sequences of the specific proteins of non-barophiles. Moreover, the statistical test also implies that the BI_{code} is significantly higher than that of the grand average of the BI values and, consequently, entails that the origin of genetic code organisation took place at high hydrostatic pressure. This is because, if the BI_{code} does not belong to non-barophile organisms, it must belong to barophile organisms if we admit, as seems to be the case, that hydrostatic pressure has a significant effect on the amino acid composition of proteins (Di Giulio, 2005).

As already mentioned, the lack of sequences of proteins from a sufficiently high number of barophile organisms does not allow us to see whether the BI_{code} value is typical of barophile organisms. Nevertheless, another check can be performed, not by sampling the BI variable within numerous barophilous organisms, but in many proteins from a single barophile organism and from a non-barophile, i.e. in a strictly controlled comparison having a relative and not an absolute significance. I therefore aligned and compared a random sample of 124 orthologous proteins from *P. furiosus* and *P. abyssi*, from which the BI values were calculated. I then calculated the mean of the BI values for these 124 proteins, which turned out to be 10.951 for *P. furiosus* and 11.139 for *P. abyssi*. Both values are different from the BI_{code}=11.639 (for *P. furiosus*: $t=(10.951-11.639)/(0.367/(124)^{1/2})=-20.88$, $df=123$, $P<<10^{-3}$; for *P. abyssi*: $t=(11.139-11.639)/(0.374/(124)^{1/2})=-14.89$, $df=123$, $P<<10^{-3}$). Therefore, the BI_{code}=11.639 is statistically different and higher than the mean value BI=10.951 from *P. furiosus*. This supports the origin of genetic code structuring under high hydrostatic pressure. The other result is less clear, as it also sees the mean value BI=11.139 of proteins from *P. abyssi* as being different and lower than BI_{code}=11.639. Clearly only when we have the possibility to conduct an extended inter- and intraspecific comparison of numerous proteins from barophilous organisms will we be able to understand this observation. For the time being, we can say that this can be expected in that an organism like *P. abyssi*, for reasons related to its particular evolution, might have a value of the mean BI=11.139 for its proteins which are not only lower than BI_{code} but also close to that of the mean of non-barophile proteins (BI=11.273) because these BI values in actual fact refer to a different sampling of the BI variable, respectively intra- and interspecies. (A more extensive sampling of the BI variable is therefore necessary, but this will only be possible after accumulating sequences from other barophile organisms.) But what is clearly important for the conclusion supported in the present paper is that the proteins from *P. furiosus* possess, as already mentioned, a mean value of BI=10.951 which is statistically different and lower than that of BI_{code}=11.639, thus adding support to the hypothesis of code structuring under high hydrostatic pressure. This is because the value BI=10.951 might actually have been statistically equal to, or even higher than, BI_{code}=11.639, which is not the case.[2]

4. Implications and prospects

The main conclusion of this paper is that it adds greater likelihood to the claim (Di Giulio, 2005) that the origin of genetic code organisation took place under high hydrostatic pressure. I think that this has two main implications. The first regards the theories put forward to explain the origin of life. A genetic code structured under high hydrostatic pressure may imply (Di Giulio, 2000) that the origin of life itself took place under high hydrostatic pressure. This favours theories such as that of Wachtershauser (1988, 1990) which see life as having originated under high hydrostatic pressure (Wachtershauser, 1992) at the expense of theories such as the heterotrophic one (Lazcano and Miller, 1996) which do not seem to be based on such an assumption. The second implication concerns the tree of life. If the main conclusion of this paper is true, then we should be able to observe that the deepest branches of the tree of life (those closest to the LUCA node) will be occupied by barophile organisms. This might turn out to be true with the furthering of our knowledge on barophile biology. Moreover, phylogenetic analysis places *Methanopyrus kandleri*, a baro-tolerant which has been isolated at a depth of 2000 m, close to the root of the archaeal tree (Burggraf et al., 1991; Rivera and Lake, 1996; Nolling et al., 1996) and, in particular, Xue et al. (2003) propose a rooting of the tree of life close to *M. kandleri*. All this is therefore in perfect agreement with the contents of the paper, namely that barophily is an ancestral trait. Whereas, the main prospect for the analysis herein

[1] More generally, if we consider that the mean of means of the BI value (11.273) is also a BI value, then it would seem justifiable to use, as standard deviation in the test, the one obtained by calculating the mean of means of the standard deviations weighted with their respective degrees of freedom, which turned out to be 0.366. However, even using the latter value as the standard deviation in the mean difference test, we obtain a highly significant value of t ($t=-4.796$, $df=22$, $P<10^{-3}$).

[2] Obviously the value of the BI means from the two compared organisms (*P. furiosus* and *P. abyssi*) are statistically different from one another ($t=(11.139-10.951)\times(124)^{1/2}/(0.370/(2)^{1/2})=+4.001$, $df=+\infty$, $P<10^{-3}$). Moreover, it must be pointed out that this analysis did not make use of sequences from, for example, *Methanococcus jannaschii*, which is another barophilous organism because there are no species of ideal methanococci that can be used in a strictly controlled comparison such as the one carried out here between *P. furiosus* and *P. abyssi*.

referred is that these methods (Di Giulio, 2000, 2001, 2003a,b, 2005) might also be used to investigate the chemical environment in which the last universal common ancestor lived.

References

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. Nucleic Acids Res. 25, 3389–3402.

Bada, J.L., Lazcano, A., 2002. Some like hot, but not the first biomolecules. Science 296, 1982–1983.

Balaam, L.N., 1972. Fundamentals of Biometry. George Allen & Unwin, London.

Burggraf, S., Stetter, K.O., Rouviere, P., Woese, C.R., 1991. *Methanopyrus kandleri*: an archaeal methanogen unrelated to all other methanogens. Syst. Appl. Microbiol. 14, 346–351.

Di Giulio, M., 1989. Some aspects of the organization and evolution of the genetic code. J. Mol. Evol. 29, 191–201.

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 universal ancestor was a thermophile or a hyperthermophile. Gene 281, 11–17.

Di Giulio, M., 2003a. The universal ancestor was a thermophile or a hyperthermophile: tests and further evidence. J. Theor. Biol. 221, 425–436.

Di Giulio, M., 2003b. The universal ancestor and the ancestor of bacteria were hyperthermophiles. J. Mol. Evol. 57, 721–730.

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

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

Galtier, N., Tourasse, N., Gouy, M., 1999. A nonhyperthermophilic common ancestor to exant life forms. Science 283, 220–221.

Hasegawa, M., Miyata, T., 1980. On the antisymmetry of the amino acid code table. Orig. Life 10, 265–270.

Lazcano, A., Miller, S.L., 1996. The origin and early evolution of life: prebiotic chemistry, the pre-RNA world, and time. Cell 85, 793–798.

Nolling, J., Elfner, A., Palmer, J.R., Steigerwald, V.J., Pihl, T.D., Lake, J.A., Reeve, J.N., 1996. Phylogeny of *Methanopyrus kandleri* based on methyl coenzyme M reductase operons. Int. J. Syst. Bacteriol. 46, 1170–1173.

Raup, D.M., Valentine, J.W., 1983. Multiple origins of life. Proc. Natl. Acad. Sci. U. S. A. 80, 2981–2984.

Rivera, M.C., Lake, J.A., 1996. The phylogeny of *Methanopyrus kandleri*. Int. J. Syst. Bacteriol. 46, 348–351.

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

Thimpson, 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. Nucleic Acids Res. 25, 4876–4882.

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

Wachtershauser, G., 1990. The case for the chemo-autotrophic origin of life in an iron-sulphur world. Orig. Life Evol. Biosph. 20, 173–176.

Wachtershauser, G., 1992. Groundworks for an evolutionary biochemistry: the iron-sulphur world. Prog. Biophys. Mol. Biol. 58, 85–201.

Woese, C.R., Dugre, D.H., Dugre, S.A., Kondo, M., Saxinger, W.C., 1966. On the fundamental nature and evolution of the genetic code. Cold Spring Harbor Symp. Quant. Biol. 31, 723–736.

Xue, H., Tong, K.L., Marck, C., Grosjean, H., Wong, J.T., 2003. Transfer RNA paralogs: evidence for genetic code-amino acid biosynthesis coevolution and archaeal root of life. Gene 310, 59–66.