Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Theoretical Biology 235 (2005) 318–325
Journal of Theoretical Biology
www.elsevier.com/locate/yjtbi

On the optimality of the genetic code, with the consideration of coevolution theory by comparison of prominent cost measure matrices

Hani Goodarzi[a,*], Hamed Shateri Najafabadi[a], Kasra Hassani[a], Hamed Ahmadi Nejad[b], Noorossadat Torabi[a]

[a]*Department of Biotechnology, Faculty of Science, University of Tehran, Enghelab Ave., Tehran, Iran*
[b]*Department of Computer Engineering, Sharif University of Technology, Azadi Ave., Tehran, Iran*

Received 10 December 2004; received in revised form 20 January 2005; accepted 24 January 2005
Available online 13 March 2005
Communicated by Clas Blomberg

Abstract

Statistical and biochemical studies have revealed non-random patterns in codon assignments. The canonical genetic code is known to be highly efficient in minimizing the effects of mistranslation errors and point mutations, since it is known that when an amino acid is converted to another due to error, the biochemical properties of the resulted amino acid are usually very similar to those of the original one. In this study, using altered forms of the fitness functions used in the prior studies, we have optimized the parameters involved in the calculation of the error minimizing property of the genetic code so that the genetic code outscores the random codes as much as possible. This work also compares two prominent matrices, the Mutation Matrix and Point Accepted Mutations 74–100 (PAM_{74-100}). It has been resulted that the hypothetical properties of the coevolution theory of the genetic code are already considered in PAM_{74-100}, giving more evidence on the existence of bias towards the genetic code in this matrix. Furthermore, our results indicate that PAM_{74-100} is biased towards the single base mistranslation occurrences in second codon position as well as the frequency of amino acids. Thus PAM_{74-100} is not a suitable substitution matrix for the studies conducted on the evolution of the genetic code.

Keywords: Coevolution theory; Fitness function; Genetic code; Load minimization; Optimality

1. Introduction

The canonical genetic code was long thought to be a "frozen accident" (Crick, 1968). However, later studies revealed the existence of some slight changes in the genetic code, which prompted a search for an applicable theory. Several hypotheses have been presented to explain the evolution of the genetic code to its present form (Crick, 1968; Dillon, 1973; Wong, 1975; Woese, 1965; Pele, 1965; Goldberg and Wittes, 1966; Woese et al., 1966; Szathmary, 1991; Szathmary and Zintzaras, 1992; Goldman, 1993; Jukes, 1997; Judson and Haydon, 1999; Houen, 1999; Ronneberg et al., 2000; Freeland et al., 2000; Ardell and Sella, 2001; Di Giulio, 2000; Freeland, 2002; Sella and Ardell, 2002; Di Giulio, 2003; Goodarzi et al., 2004). One of the latest scenarios, designated "code-metabolism coevolution", postulates that the earliest genetic code used a small number of prebiotically synthesized amino acids, and subsequently expanded to its present form by incorporating novel derivatives of these primordial amino acids as biosynthetic pathways evolved (Wong, 1975, 1976, 1981; Wong and Bronskill, 1979; Di Giulio, 2001a). However, the statistical significance of this theory was debated by Ronneberg et al. (2000). On the other hand, statistical

*Corresponding author. Tel.: +98 21 8058210; fax: +98 21 8040284.
E-mail address: hani.goodarzi@gmail.com (H. Goodarzi).

0022-5193/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2005.01.012

studies have supported the theory that the genetic code has evolved so as to minimize the consequences of errors during transcription and translation (Woese, 1965; Sonneborn, 1965; Haig and Hurst, 1991; Freeland and Hurst, 1998; Ardell, 1998; Knight et al., 1999 Freeland et al., 2000; Goodarzi et al., 2004). In order to test this hypothesis, attempts have been made to compute the optimality of the genetic code by quantifying the cost of single-base changes reflected in protein synthesis (Alff-Steinberger, 1969; Haig and Hurst, 1991; Ardell, 1998; Freeland and Hurst, 1998; Gilis et al., 2001).

More recently, Haig and Hurst (1991), and Freeland and Hurst (1998), improved their approach by comparing the canonical genetic code with randomly generated ones in order to measure the efficiency of the code in limiting the consequences of errors during transcription and translation. Haig and Hurst (1991) proposed a fitness function, designated φ, for estimating the efficiency of a code in load minimization and computed the fraction of random codes which scored better than the canonical genetic code (i.e. had smaller values of φ). Testing several fitness functions based on different physicochemical parameters, they found that when using differences in polarity between amino acids (Woese, 1965; Di Giulio, 1989) as a cost measure, single base changes in the natural code resulted a very small average load. The nearer the score of the canonical genetic code becomes to the minimum of φ, the higher load minimization it indicates, where φ is defined as:

$$\varphi^{Haig\ and\ Hurst} = \sum_{c,c'} [h(a(c)) - h(a(c'))]^2. \quad (1)$$

In the function presented, $a(c)$ is the amino acid coded by codon c and $h(a)$ is the hydropathy index of amino acid a. Calculating φ scores for thousands of randomly generated codes, Haig and Hurst (1991) found the fraction of random codes that outscored the natural genetic code to be in the order of 10^{-4}. Later, consideration of transition/transversion biases and different probabilities of mistranslation for the three codon positions, led to the proposition of a new fitness function which modeled the probability of translational errors more accurately (Freeland and Hurst, 1998). With the improved fitness function, Freeland and Hurst (1998), decreased the fraction of random genetic codes that are better than the natural one, from 10^{-4} to 10^{-6}.

Gilis et al. (2001), highlighted the importance of another parameter in the optimization of genetic code, namely the frequency at which different amino acids occur in proteins. Although this frequency differs from protein to protein, a prevailing pattern can be recognized in general (King and Jukes, 1969; Gilis et al., 2001; Brooks et al., 2002). A high correlation has been revealed between the number of synonymous codons and the frequencies in which their relative amino acids occur, emphasizing the significance of this parameter in the optimization of the genetic code. In addition, they brought further improvement to φ by using quantities other than polarity to measure the costs of different amino acid substitutions in protein conformation and stability. They devised a cost function designated "Mutation Matrix" by evaluating, in silico, the changes in folding free energy caused by all possible point mutations in a set of protein structures. Mutation Matrix is alleged to be unbiased towards the genetic code, based on the fact that it compares the impact of mutations on the accurate folding of a protein, which makes it a suitable cost function to use when studying genetic code (Gilis et al., 2001). The derivation is based on a dataset of 141 protein structures, determined by X-ray crystallography and studied by Wintjens et al. (1996). Each residue in each of these 141 proteins is mutated into the 19 non-wild-type amino acids and for each of these mutations the change in folding free energy is evaluated using the procedure detailed by Gilis et al. (2001). For the sake of comparison, besides their Mutation Matrix, Gilis et al. (2001) also made use of Point Accepted Mutations 74–100 (PAM_{74-100}) scoring matrix. The Point Accepted Mutations (PAM) matrices are a family of matrices derived from amino acid substitution frequencies observed from within naturally occurring homologous proteins (Benner et al., 1994). Although PAM_{74-100} is the chosen matrix for many studies (Ardell, 1998; Freeland et al., 2000; Gilis et al., 2001; Goodarzi et al., 2004), its validity has been argued by some researchers (Di Giulio, 2001b).

Gilis et al. (2001), conducted the classic experiment of generating random genetic codes, while applying their new fitness function (stop codons are excluded):

$$\varphi^{faa} = \sum_{c=1}^{64} \frac{p(a(c))}{n(a(c))} \sum_{c'=1}^{64} p(c'|c)\, g(a(c), a(c')). \quad (2)$$

In the above equation, $p(a(c))$ returns the relative frequency of the amino acid $a(c)$ coded by codon c, $n(a(c))$ is an integer standing for the number of synonymous codons that amino acid $a(c)$ possesses, and $g(a(c), a(c'))$ is a cost measure function which illustrates the deleterious effect of the amino acid substitution resulted from the misinterpretation of codon c as c'. Throughout this text, the functions $p(a)$ and $n(a)$ are used as well, returning the relative frequency and the number of synonyms of an amino acid, respectively. The corresponding results revealed a lower f, the fraction of random codes which scored better than the canonical genetic code. f was calculated to be 2×10^{-9} when applying Mutation Matrix or PAM_{74-100}.

In a recent work by Goodarzi et al. (2004) a new group of parameters came into consideration to measure the efficiency of the genetic code in load minimization, including a constant designated K, in order to consider

the nonsense mutations. K is defined so that

$$g(a, term) = -K, \tag{3}$$

where $g(a, term)$ stands for the cost of misinterpretation of any codon as a stop codon.

Furthermore, Goodarzi et al. (2004) incorporated a novel function named $f(c)$ into φ^{faa}, resulting in a new fitness function named φ^{HH}. $f(c)$ is declared as

$$f(c) = 1 - \int_0^{|y|} \frac{1}{\sqrt{2\pi}} e^{-(t^2/2)}\, dt,$$
$$y = \frac{1}{S} \cdot \frac{\big(n(a(c))/(n(total) - n(term))\big) - p(a(c))}{\sigma(a(c))} \tag{4}$$

where $n(total)$ and $n(term)$ are the total number of codons and the number of termination codons, respectively. Note that in all the randomly generated codes, $n(total)$ equals 64 and $n(term)$ returns 3. The value designated S was defined as a constant value, fixed at 100 (for further discussion see Goodarzi et al., 2004). As a result, φ^{HH} was declared as

$$\varphi^{HH} = \sum_{c=1}^{64} \frac{p(a(c))}{n(a(c))} f(c) \sum_{c'=1}^{64} p(c'|c)\, g(a(c), a(c')). \tag{5}$$

Goodarzi et al. (2004) found that when applying φ^{HH}, the probability of finding a code better than the canonical genetic code decreases compared to φ^{faa}.

According to "code-metabolism coevolution" theory, codon assignments coevolved with invention of biosynthetic pathways for new amino acids, which means that product amino acids being synthesized from precursor ones usurped codons assigned to their precursors (Wong, 1975; Amirnovin, 1997; Ronneberg et al., 2000). The theory postulates that biosynthetically related amino acids in the genetic code should be positioned near each other in terms of mutation. The 12 biochemically related precursor–product amino acid pairs introduced by Ronneberg et al. (2000) are shown in Table 1. All pairs are located in the same biochemical branch and all conversions are energetically favorable.

Goodarzi et al. (2004) dictated extra weight to these pairs by altering the scoring matrix using the following equation:

$$g(a, a') = g(a, a') + Inc|g(a, a')|. \tag{6}$$

A cost measure is added by a known fraction of itself, if the pair a/a' is in the set of pairs shown in Table 1, thus favoring the two amino acids of a pair being positioned near each other.

However, it has been stated by Ronneberg et al. (2000) that the four last pairs of these 12 precursor–product pairs can be omitted (Table 2).

Table 1
Precursor–product pairs, indicated by Ronneberg et al. (2000)

Precursor	Product	Precursor	Product
Ser	Trp	Asp	Asn
Ser	Cys	Asp	Met
Phe	Tyr	Asp	Lys
Thr	Ile	Glu	Pro
Gln	His	Asp	Arg
Glu	Gln	Asp	Thr

Table 2
Precursor–product pairs, indicated by Ronneberg et al. (2000), an altered form of the 12 pairs according to the assumptions presented by Ronneberg et al. (2000)

Precursor	Product	Precursor	Product
Ser	Trp	Gln	His
Ser	Cys	Glu	Gln
Phe	Tyr	Asp	Asn
Thr	Ile	Asp	Met

These pairs are a result of removing the last four pairs in Table 1.

2. Methods

In the work presented, some altered forms of fitness functions, stated previously by Gilis et al. (2001) and Goodarzi et al. (2004), are used to measure the efficiency of the canonical genetic code in reducing the effects of errors during transcription and translation.

z-value scoring method was used to measure the optimality of the canonical genetic code in accordance to the random codes (Goodarzi et al., 2004):

$$z = \frac{\varphi_{cgc} - \mu}{\sigma}, \tag{7}$$

where φ_{cgc} is the fitness score of the canonical genetic code, μ is the average score of the randomly generated codes, and σ is the standard deviation of distribution of scores obtained from the random codes.

Throughout this text, optimizing a parameter means to find a value for that parameter which results in a maximum z-value.

Rules for generating random genetic codes (Freeland and Hurst, 1998):

1. The "codon space" is divided into 21 non-overlapping sets of codons observed in the canonical code, each set specifying an amino acid in the natural genetic code (one set consists of stop codons).
2. Each alternative code is obtained by randomly assigning each of the 20 amino acids to one of these

sets. All three stop codons remain invariant in position for all alternative codes.

2.1. Determining the mistranslation related parameters by a new method

No known tRNA anticodon base modification, natural or engineered, can discriminate third-base pyrimidines within any codon, despite the large number of non-standard codes now recognized and diverse experimental manipulation of coding components (Ronneberg et al., 2000). So, codon XYU is not distinguishable from XYC and vice versa. Thus, XYU can mutate to any codon which XYC is able to mutate to and XYC can mutate to any codon which XYU is able to mutate to. For example, UUC can be mutated to UGU, in the same way as UUU does.

With these considerations, a new fitness function, named φ_{ac}^{faa}, was defined on the same basis as φ^{faa}, with the difference that double-base mutations were allowed as long as one of them was a transition in a pyrimidine located in the third position:

$$\varphi_{ac}^{faa} = \sum_{c=1}^{64} \frac{p(a(c))}{n(a(c))} \sum_{c'=1}^{64} p(c'|c)\, g(a(c), a(c')). \quad (8)$$

φ_{ac}^{faa} was used for optimizing the transition/transversion weighting in the first and the second codon position. Both Mutation Matrix and PAM_{74-100} were used as a reference to the values of $g(a(c), a(c'))$.

2.2. Inclusion of nonsense mistranslations

In order to include the nonsense mutations, K was defined the same as stated by Goodarzi et al. (2004) and presented in Eq. (3), using φ_{ac}^{faa} and transition / transversion weightings resulted in the previous step. K was optimized for both Mutation Matrix and PAM_{74-100}. The same mutation biases among different codon positions were used as stated by Freeland and Hurst (1998).

2.3. Optimizing the value of S

Hereafter, φ_{ac}^{HH} was defined, based on φ^{HH} declared by Goodarzi et al. (2004), allowing double-base mutations if causing a transition in a pyrimidine in the third codon position:

$$\varphi_{ac}^{HH} = \sum_{c=1}^{64} \frac{p(a(c))}{n(a(c))} f(c) \sum_{c'=1}^{64} p(c'|c)\, g(a(c), a(c')). \quad (9)$$

Using φ_{ac}^{HH} with Mutation Matrix and PAM_{74-100}, the constant S, presented in Eq. (4), was optimized. The same transition/transversion weightings and K, resulted from φ_{ac}^{faa}, were used in calculation of φ_{ac}^{HH}.

2.4. Applying code-metabolism coevolution theory

Finally, the value of the constant *Inc*, presented in qEq. (6), was also optimized in accordance to both the 12 pairs of biochemically related amino acid pairs introduced by Ronneberg et al. (2000), which are presented in Table 1, and the eight pairs presented in Table 2.

3. Results

3.1. Determination of transition/transversion weightings

Z-values obtained by applying φ_{ac}^{faa} and different transition/transversion weightings in the first codon position are plotted in Figs. 1 and 2, referring to Mutation Matrix and PAM_{74-100}, respectively. Using Mutation Matrix resulted in an optimum transition/transversion weighting of 1.7 for the first codon position. This value was determined to be 1.5 for PAM_{74-100}.

The same analysis was done regarding to the second codon position (Figs. 3 and 4). Hereby, no maximum value was resulted by tracking z-values obtained from transition/transversion weightings in the range of 0–25, neither for Mutation Matrix nor for PAM_{74-100}. So we conventionally chose the second base transition/transversion weighting to be the value that resulted in 99% of the extrapolated maximum z-value. These values are 1194.63 and 7.6 regarding Mutation Matrix and PAM_{74-100}, respectively. Based on the new transition/transversion weightings, values of $p(c'|c)$ have been established and are presented in Table 3.

3.2. Optimizing the cost of nonsense mistranslations

New values of $p(c'|c)$ were used to optimize the constant K (Eq. (3)). As shown in Figs. 5 and 6, the

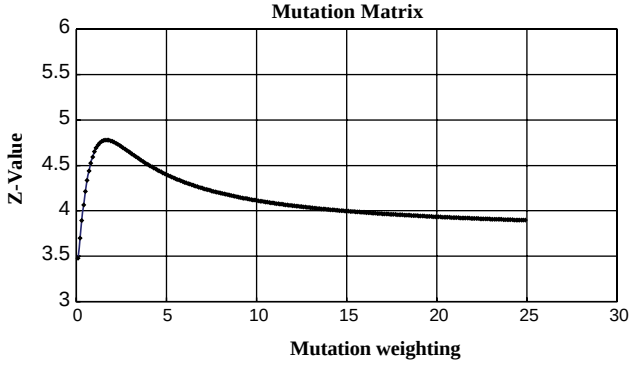

Fig. 1. Z-value versus transition/transversion weighting in the first codon position using Mutation Matrix.

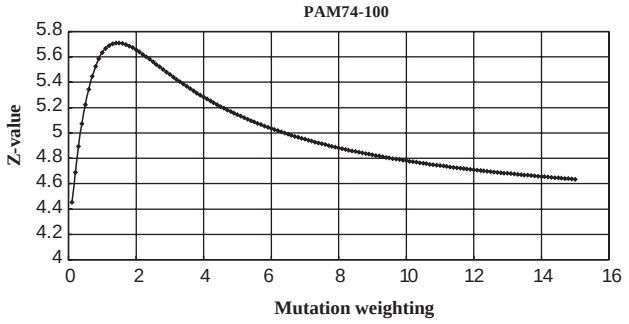

Fig. 2. Z-value versus transition/transversion weighting in the first codon position using PAM_{74-100}.

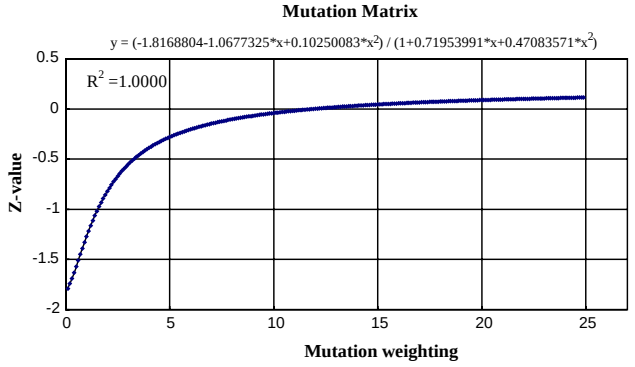

Fig. 3. Z-value versus transition/transversion weighting in the second codon position using Mutation Matrix.

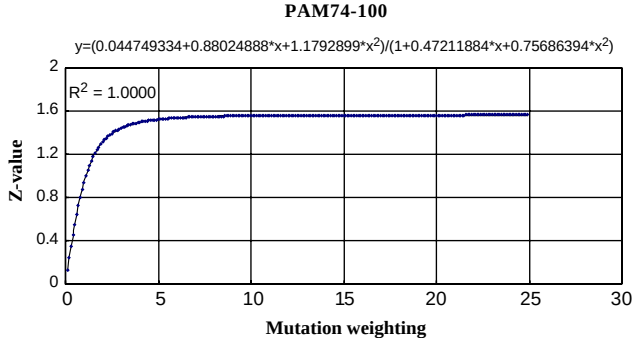

Fig. 4. Z-value versus transition/transversion weighting in the second codon position using PAM_{74-100}.

values of 4.5 and 5 resulted in maximum z-values for Mutation Matrix and PAM_{74-100}, respectively.

3.3. Optimizing the constant S

Using φ_{ac}^{HH} while applying new values of $p(c'|c)$ and the constant K, optimization of the constant S (Eq. (4)) resulted in the values of 9 and 3 for Mutation Matrix and PAM_{74-100}, respectively (Figs. 7 and 8).

Table 3
New values for $p(c'|c)$ in accordance to load minimization of transition/transversion mutation scores

Mutation type and position	$p(c'\|c)$ for PAM_{74-100}	$p(c'\|c)$ for mutation matrix
First position transition	$1/N$	$1/N$
First position transversion	$0.667/N$	$0.588/N$
Second position transition	$0.5/N$	$0.5/N$
Second position transverion	$0.066/N$	$0.000419/N$

N is the normalization factor selected so that $\sum_{c'} p(c'|c) = 1$.

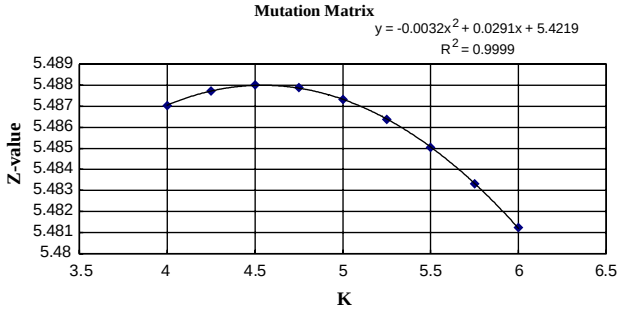

Fig. 5. Z-value versus constant K using Mutation Matrix.

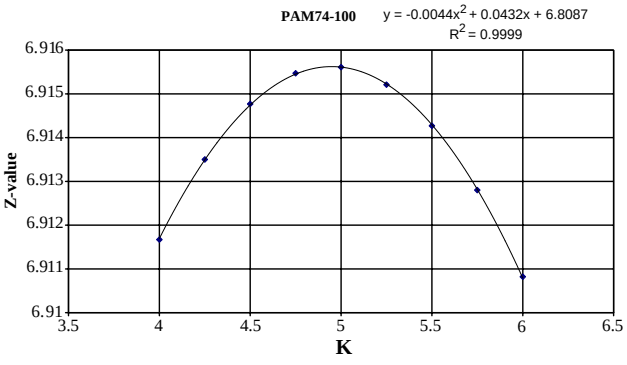

Fig. 6. Z-value versus constant K using PAM_{74-100}.

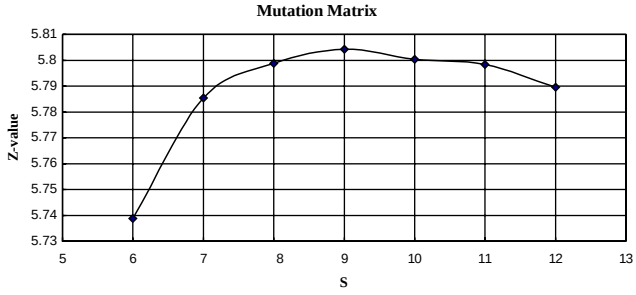

Fig. 7. Z-value versus constant S presented in Eq. (3), using Mutation Matrix.

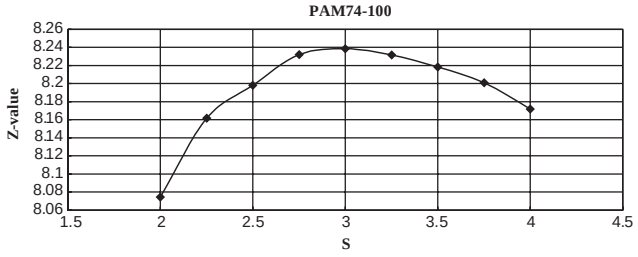

Fig. 8. Z-value versus constant S presented in Eq. (3), using PAM_{74-100}.

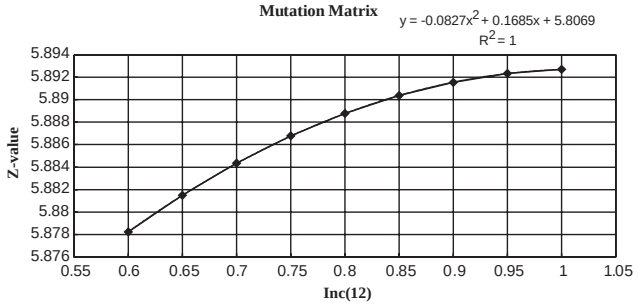

Fig. 9. Z-value versus constant *Inc* for the precursor–product pairs presented in Table 1, using Mutation Matrix.

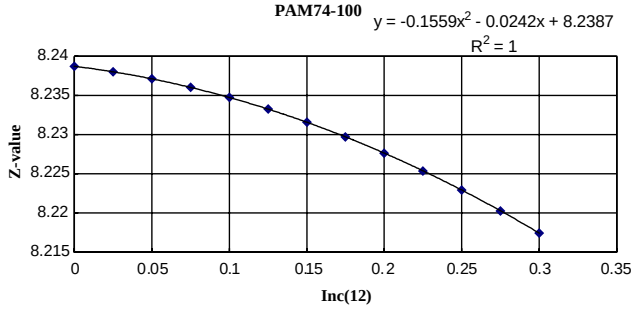

Fig. 10. Z-value versus constant *Inc* for the precursor–product pairs presented in Table 1, using PAM_{74-100}.

3.4. Code-metabolism coevolution theory

In the final step, the optimization of the constant *Inc* was analysed, first according to the classic point of view (including 12 precursor–product pairs, Table 1) and then by omitting four pairs and using the remaining eight (Table 2). The results are plotted in Figs. 9–12 for Mutation Matrix and PAM_{74-100}. As the graphs indicate, no optimized value can be concluded from the analysis of the 12 amino acid pairs as the plots are strictly incremental or decremental. Hence, when observing the results of the eight amino acid pair analysis, the graph regarding to PAM_{74-100} possesses a maximum in $Inc = 0.075$ and thus, an optimized value of *Inc* constant

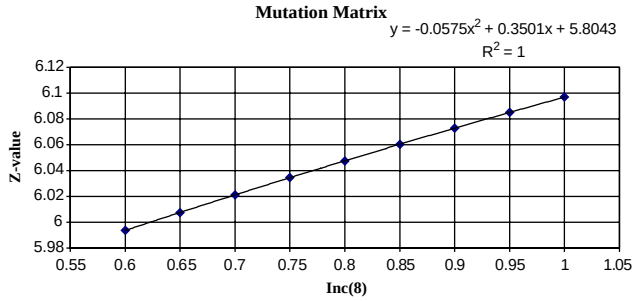

Fig. 11. Z-value versus constant *Inc* for the precursor–product pairs presented in Table 2, using Mutation Matrix.

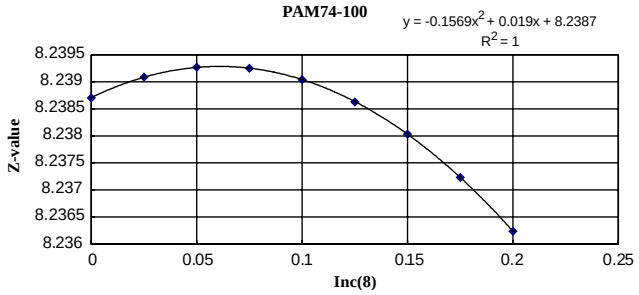

Fig. 12. Z-value versus constant *Inc* for the precursor–product pairs presented in Table 2, using PAM_{74-100}.

can be concluded. In contrast, the graph regarding to the Mutation Matrix is strictly incremental and possesses no maximum value.

4. Discussion and conclusion

Our results indicated at least three reasons that PAM_{74-100} is not a suitable measure for studies conducted on the optimality of the genetic code:

4.1. The optimal probability of transversion occurrence in the second codon position

As shown in Table 3, optimization of the transition/transversion weighting in the first codon position results in similar values for Mutation Matrix and PAM_{74-100}. However, in the case of the second codon position, when Mutation Matrix is used, the genetic code shows the highest optimality actually when no transversion occurs. This is while using PAM_{74-100}, optimization of the transition/transversion weighting for the second position of codon results in a value about five times of the transition/transversion weighting for the first codon position, causing the probability of transversion in the second codon position to be 0.1 of that in the first codon position. Applying Mutation Matrix, φ_{ac}^{faa} does not favor the occurrence of transversional mistranslations in

the second codon position even at a very low probability, whereas PAM_{74-100} admits a rather considerable probability.

4.2. PAM_{74-100} reflects the frequencies of amino acids

Goodarzi et al. (2004) introduced S to be as conservative as possible. Higher values of S causes the function $f(c)$ to be less effective in φ_{ac}^{HH}. PAM_{74-100} results in the minimum probability of finding a better code than the canonical genetic code when S is chosen to be a small value, compared to the optimized value of S for Mutation Matrix. It implies that in the case of PAM_{74-100}, when the relative frequency of an amino acid deviates from the linear relationship proposed by King and Jukes (1969) as indicated by $f(c)$, the corresponding $g(a(c), a(c'))$ returns a higher value, resulting in a higher φ_{ac}^{HH}. It can be taken into interpretation that PAM_{74-100} is correlated with the frequencies of amino acids, in addition to the correlation between PAM matrices and the genetic code as mentioned previously by Di Giulio (2001b).

4.3. Weightings due to code-metabolism coevolution are already indicated in PAM_{74-100}

As shown in Fig. 10, the incorporation of the constant *Inc* in accordance with the 12 pairs of amino acids presented in Table 1, when using PAM_{74-100}, makes the z-value to decrease. Also the optimization of the constant *Inc* in accordance to the eight pairs presented in Table 2 results in a value of 0.05 with respect to PAM_{74-100}. The incorporation of this value to PAM_{74-100} changes it only very slightly. It can be taken to indicate that, as some researchers suggested, although PAM_{74-100} has been acquired of highly diverged homologous proteins, it is still affected by the function and the evolutionary history of the genetic code, so that the biosynthetic relations of amino acids are already considered in PAM_{74-100}. These results also support the coevolution theory by finding its footsteps in PAM_{74-100}, as a matrix that reflects the properties of the genetic code.

It should be highlighted that in all steps of this work, the genetic code possessed significantly higher z-values when PAM_{74-100} was used as the cost measure of amino acid substitution, compared with z-values obtained when Mutation Matrix was used. This is at least partly due to the correlation of PAM_{74-100} with the genetic code. One other possible explanation is that still there are other parameters (similar to *Inc*) that are already considered in PAM_{74-100}. In other words, there are even more hidden biases behind the values reflected in PAM_{74-100}.

Acknowledgements

Authors are grateful to Elahe Elahi for her useful comments and advices.

References

Alff-Steinberger, 1969. The genetic code and error transmission. Proc. Natl Acad. Sci. USA 64, 584–591.
Amirnovin, R., 1997. An analysis of the metabolic theory of the origin of the genetic code. J. Mol. Evol. 44, 473–476.
Ardell, D.H., 1998. On error-minimization in a sequential origin of the standard genetic code. J. Mol. Evol. 47, 1–13.
Ardell, D.H., Sella, G., 2001. On the evolution of redundancy in genetic codes. J. Mol. Evol. 53, 269–281.
Benner, S.A., Cohen, M.A., Gonnet, G.H., 1994. Amino acid substitution during functionally constrained divergent evolution of protein sequences. Protein Eng. 7 (11), 1323–1332.
Brooks, D.J., Fresco, J.R., Lesk, A.M., Singh, M., 2002. Evolution of amino acid frequencies in proteins over deep time: inferred order of introduction of amino acids into the genetic code. Mol. Biol. Evol. 19, 1645–1655.
Crick, F.H., 1968. The origin of genetic code. J. Mol. Biol. 38, 367–379.
Di Giulio, M., 1989. The extension reached by the minimization of the polarity distances during the evolution of the genetic code. J. Mol. Evol. 29, 288–293.
Di Giulio, M., 2000. The origin of the genetic code. Trends Biochem. Sci. 25, 44–47.
Di Giulio, M., 2001a. A blind empiricism against the coevolution theory of the origin of the genetic code. J. Mol. Evol. 53, 724–732.
Di Giulio, M., 2001b. The origin of the genetic code cannot be studied using measurements based on the PAM matrix because this matrix reflects the code itself, making any such analyses tautologous. J. Theor. Biol. 208, 141–144.
Di Giulio, M., 2003. The early phases of genetic code origin: conjectures on the evolution of coded catalysis. Origins Life Evol. Biosphere 33 (4), 479–489.
Dillon, L.S., 1973. The origins of the genetic code. Bat. Rev. 39, 301–345.
Freeland, S.J., 2002. The genetic code: an adaptation for adapting? J. Genetic Program. Evolv. Mach. 3 (2), 113–127.
Freeland, S.J., Hurst, L.D., 1998. The genetic code is one in a million. J. Mol. Evol. 47, 238–248.
Freeland, S.J., Knight, R.D., Landweber, L.F., Hurst, L.D., 2000. Early fixation of an optimal genetic code. Mol. Biol. Evol. 17, 511–518.
Gilis, D., Massar, S., Cerf, N.J., Rooman, M., 2001. Optimality of the genetic code with respect to protein stability and amino acid frequencies. Genome Biol. 2 (11), 49.1–49.12.
Goodarzi, H., Nejad, H.A., Torabi, N., 2004. On the optimality of the genetic code with the consideration of termination codons. BioSystems 57 (1–3), 63–73.
Goldberg, A.L., Wittes, R.E., 1966. Genetic code: aspects of organization. Science 153, 420–424.
Goldman, N., 1993. Further results on error minimization in the genetic code. J. Mol. Evol. 37, 662–664.
Haig, D., Hurst, L.D., 1991. A quantitative measure of error minimization on the genetic code. J. Mol. Evol. 33, 412–417.
Houen, G., 1999. Evolution of the genetic code: the nonsense, antisense, and antinonsense codes make no sense. Biosystems 54, 39–46.
Judson, O.P., Haydon, D., 1999. The genetic code: what is it good for? J. Mol. Evol. 49, 539–550.

Jukes, T.H., 1997. Neutral changes and modifications of the genetic code. Theor. Popul. Biol. 49, 143–145.

King, J.L., Jukes, T.H., 1969. Non-Darwinian evolution. Science 164, 788–798.

Knight, R.D., Freeland, S.J., Landweber, L.F., 1999. Selection, history, and chemistry: the three faces of the genetic code. Trends Biochem. Sci. 24, 241–247.

Pele, S.R., 1965. Correlation between coding-triplets and amino acids. Nature 207, 597–599.

Ronneberg, T.A., Landweber, L.F., Freeland, S.J., 2000. Testing a biosynthetic theory of the genetic code: fact or artifact? Proc. Natl Acad. Sci. USA 97, 13690–13695.

Sella, G., Ardell, D.H., 2002. The impact of message mutation on the fitness of a genetic code. J. Mol. Evol. 54 (5), 638–651.

Sonneborn, T.M., 1965. Degeneracy of the genetic code: extent, nature, and genetic implications. In: Evolving Genes and Proteins. Academic Press, New York, pp. 97–166.

Szathmary, E., 1991. Codon swapping as a possible evolutionary mechanism. J. Mol. Evol. 32, 178–182.

Szathmary, E., Zintzaras, E., 1992. A statistical test of hypotheses on the Organization and origin of the genetic code. J. Mol. Evol. 35, 185–189.

Wintjens, R.T., Rooman, M.J., Wodak, S.J., 1996. Automatic classification of $\alpha\alpha$-turn motifs in proteins. J. Mol. Biol. 255, 235–253.

Woese, C.R., 1965. On the evolution of the genetic code. Proc. Natl Acad. Sci. USA 54, 1546–1552.

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

Wong, J.T., 1975. A co-evolution theory of the genetic code. Proc. Natl Acad. Sci. USA 72, 1909–1912.

Wong, J.T., 1976. The evolution of a universal genetic code. Proc. Natl. Acad. Sci. USA 73, 2336–2340.

Wong, J.T., 1981. Coevolution of genetic code and amino acid biosynthesis. Trends Biochem. Sci. 6, 33–36.

Wong, J.T., Bronskill, P.M., 1979. Inadequacy of pre-biotic synthesis as the origin of proteinaceous amino acids. J. Mol. Evol. 13, 115–125.