(This is a sample cover image for this issue. The actual cover is not yet available at this time.)

Journal of Theoretical Biology 304 (2012) 183–187

Contents lists available at SciVerse ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Why has nature invented three stop codons of DNA and only one start codon?

Michal Křížek [a,*], Pavel Křížek [b]

[a] Institute of Mathematics, Academy of Sciences, Žitná 25, CZ–115 67 Prague 1, Czech Republic
[b] Institute of Cellular Biology and Pathology, First Faculty of Medicine, Charles University in Prague, Albertov 4, CZ–128 00 Prague 2, Czech Republic

ARTICLE INFO

Article history:
Received 30 August 2011
Received in revised form
19 March 2012
Accepted 20 March 2012
Dedicated to Prof. Václav Pačes on his 70th birthday.
Available online 30 March 2012

Keywords:
DNA
RNA
Stop codon
Synchronization shift
Drosophila genome

ABSTRACT

We examine the standard genetic code with three stop codons. Assuming that the synchronization period of length 3 in DNA or RNA is violated during the transcription or translation processes, the probability of reading a frameshifted stop codon is higher than if the code would have only one stop codon. Consequently, the synthesis of RNA or proteins will soon terminate. In this way, cells do not produce undesirable proteins and essentially save energy. This hypothesis is tested on the AT-rich Drosophila genome, where the detection of frameshifted stop codons is even higher than the theoretical value. Using the binomial theorem, we establish the probability of reading a frameshifted stop codon within n steps. Since the genetic code is largely redundant, there is still space for some hidden secondary functions of this code. In particular, because stop codons do not contain cytosine, random C → U and C → T mutations in the third position of codons increase the number of hidden frameshifted stops and simultaneously the same amino acids are coded. This evolutionary advantage is demonstrated on the genomes of several simple species, e.g. *Escherichia coli*.

1. Briefly from the history of genetic code and stop codon analysis

When the structure of DNA was discovered by Watson and Crick (1953) the first attempts to establish the genetic code started. George Gamow got the idea that combinatorics and number theory could be used to explain the rich variability of species and functioning of genes. He realized that 20 amino acids constituting the proteins cannot be coded by pairs of nucleotides, since there exists only $16=4\times4$ different nucleotide pairs from the alphabet {A, C, G, T}, where A stands for adenine, C cytosine, G guanine, and T is thymine. To avoid this drawback, he proposed a special overlapping code (Gamow, 1954), where only one pair of nucleotides was considered, but a particular amino acid was assigned after reading the first nucleotide from the next pair (e.g. AC G A…). It is obvious that such a partial overlapping code would prescribe very restrictive conditions on the arrangement of particular amino acids. Later, it was found that this idea is wrong (see Watson, 1999).

Francis Crick also tried to establish the genetic code. His ingenuity can be illustrated on the so-called *comma-free codes*. Crick correctly assumed that each amino acid is coded by a triple of nucleotides called *codons* (i.e., there are $64=4\times4\times4$ different triplets). He knew that these triplets are by no means separated, i.e., it is not clear from where the genetic information is to be read. Crick and his coauthors (Crick et al., 1957) developed the following code:

They assumed that triplets AAA, CCC, GGG, and TTT do not code anything and that the remaining 60 triplets are divided into 20 equivalence classes. Two triplets were supposed to be equivalent, if there exists a circular permutation which transforms one triplet to another. For example AAC, ACA, and CAA are three equivalent triplets and there is no other triplet that would be equivalent to them. In this way they got a one-to-one mapping between 20 amino acids and 20 equivalence classes. Moreover, their code had a certain advantage, namely, the sequence

…AAC AAC AAC GAC GAC TAC…

would code the same protein as the sequence

…AA CAA CAA CGA CGA CTA C…,

which arises by a cyclic permutation of each triplet and the order of nucleotides is the same in both sequences. In this special and artificial case, it does not matter from which nucleotide of a given triplet the genetic information is read (therefore, the comma-free code).

However, later Nirenberg and Matthaei (1961) discovered that the amino acid phenylalanine is coded by the triplet TTT. Since this triplet was in Crick's code excluded, the comma-free code

* Corresponding author. Tel.: +420 222 090 712.
E-mail addresses: krizek@cesnet.cz, krizek@math.cas.cz (M. Křížek), pkriz@lf1.cuni.cz (P. Křížek).

0022-5193/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jtbi.2012.03.026

Table 1
The standard RNA genetic code.

First nucleotide	Second nucleotide				Third nucleotide
	U	*C*	*A*	*G*	
U	*UUU* phenylalanine	*UCU* serine	*UAU* tyrosine	*UGU* cysteine	*U*
	UUC	*UCC*	*UAC*	*UGC*	*C*
	UUA leucine	*UCA*	*UAA* **stop**	*UGA* **stop**	*A*
	UUG	*UCG*	*UAG* **stop**	*UGG* tryptophan	*G*
C	*CUU* leucine	*CCU* proline	*CAU* histidine	*CGU* arginine	*U*
	CUC	*CCC*	*CAC*	*CGC*	*C*
	CUA	*CCA*	*CAA* glutamine	*CGA*	*A*
	CUG	*CCG*	*CAG*	*CGG*	*G*
A	*AUU* isoleucine	*ACU* threonine	*AAU* asparagine	*AGU* serine	*U*
	AUC	*ACC*	*AAC*	*AGC*	*C*
	AUA	*ACA*	*AAA* lysine	*AGA* arginine	*A*
	AUG methionine	*ACG*	*AAG*	*AGG*	*G*
G	*GUU* valine	*GCU* alanine	*GAU* aspartic acid	*GGU* glycine	*U*
	GUC	*GCC*	*GAC*	*GGC*	*C*
	GUA	*GCA*	*GAA* glutamic acid	*GGA*	*A*
	GUG	*GCG*	*GAG*	*GGG*	*G*

was also found to be incorrect. Five years later Marshall Nirenberg established the final shape of the standard RNA genetic code, where thymine T is replaced by uracil U (Alberts et al., 1997; Păun et al., 2006).

This code is thoroughly investigated until now. An extensive survey of many of its properties is presented in Crick (1968). By Woese (1965) codons that differ by one letter usually encode the same amino acids or chemically related ones, i.e., single point mutations result in minimal effects on the translated protein (Itzkovitz and Alon, 2007). There are some exceptions. For instance, the single point mutation UGU → UGG corresponds to two quite different amino acids: cysteine and tryptophan.

Table 1 contains three special codons UAA, UAG, and UGA that normally function as stop signals. The discussion on the optimal number of stop codons has a long history dating back to the paper by Clarke and Miller (1982) on frameshift mutations of *E. coli* genes.

The above RNA genetic code from Table 1 is not universal. Many small deviations are surveyed in Santos et al. (2004). For instance, several prokaryota (e.g. the *Entomoplasmatales* and *Mycoplasmatales*) utilize an alternative genetic code in which UGA is not a stop codon (Bove, 1993). In some archaea, the UGA triplet encodes the 21st amino acid (selenocysteine). Furthermore, as reported in Zhang and Gladyshev (2007) in a symbiotic deltaproteobacterium of the gutless worm, *Olavius algarvensis*, UAG encodes the 22nd amino acid (pyrrolysine). Therefore, it is probable that very ancient primitive life could use only one stop codon for protein encoding, later two, and then three stop codons. Even four stop codons can be found in mitochondrial genomic data of vertebrates or *Thraustochytrium*. For genetic codes with different stop codon assignments, see Seligmann and Pollock (2004) and Singh and Pardasani (2009). In this note we discuss why it is advantageous from the evolutionary point of view to have more stop signals and only one start signal.

The 'ambush' hypothesis suggests that frameshifted stop codons (termed also out-of-frame stop codons or off frame stop codons) are sometimes selected for, see Seligmann and Pollock (2003) for the first appearance of this hypothesis. In other words, the functional significance of an increased frameshifted stop codons frequency can be explained by the *ambush hypothesis*. Frameshifted stops could reduce the metabolic cost of accidental frameshifts, and a positive correlation between the usage of codons and the number of ways codons can be part of hidden stops is expected, see Tse et al. (2010).

Some genomic data are in agreement with this hypothesis suggesting biotechnological applications (Seligmann and Pollock, 2004). Frameshifted stops indeed decrease the production of undesirable proteins (Seligmann, 2007). According to Itzkovitz and Alon (2007), genetic codes are designed so that they maximize the number of frameshifted stop codons and simultaneously they can carry some abundant additional information. As shown in Seligmann (2010a), there also exists a positive correlation between developmental stability and frameshifted stops. Moreover, translational frameshifted errors play important roles in "regulating ribosomal activity after frameshifts", see Seligmann (2011, p. 272).

Genetic codes possessing fewer stops usually have shorter genomes (Johnson, 2011). But this phenomenon seems to be compensated by the usage of alternative genetic codes in situations where suppressor tRNAs (see Seligmann, 2010b, 2011) are expressed.

The ambush hypothesis was further developed in Singh and Pardasani (2009) and Seligmann (2010a). A substantial number of genomes display a significant correlation between shifted stop codons and codon usage frequencies. Nevertheless, some evaluated evidences against the ambush hypothesis have been analyzed in Singh and Pardasani (2009).

2. The function of frameshifted stop codons

Consider a sequence of bases

...AGCGUUACCAU...

Which sequence of amino acids does it define? According to Table 1, it may be:

$$\begin{aligned}&\ldots + \text{serine} + \text{valine} + \text{threonine} + (\text{AU}) \text{ or}\\ &\quad (\text{A}) + \text{alanine} + \text{leucine} + \text{proline} + (\text{U}) \text{ or}\\ &\quad (\text{AG}) + \text{arginine} + \text{tyrosine} + \text{histidine} + \ldots,\end{aligned} \tag{1}$$

which are diametrically opposed triplets of amino acids. Which of these three sequences should be synthesized?

The standard genetic code has one start (initiation) codon AUG which establishes the beginning of synthesis and at the same time it codes the amino acid methionine (see Table 1). Downstream of the AUG there is usually a long sequence of hundreds or thousands of codons defining the protein. This sequence does not contain any stop codon UAA, UAG, or UGA except for the last one.

Since codons are in no way separated, any synchronization shift during transcription or translation by $\pm n$ bases, where n is not divisible by three, produces a wrong sequence of triplets

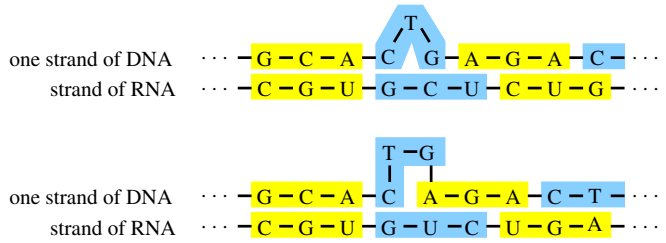

Fig. 1. A schematic illustration of a frameshift by one or two bases during transcription.

(see Fig. 1). Therefore, it seems very advantageous that nature invented three stop codons in the standard genetic code. When a synchronization shift appears, the synthesis of RNA or protein is soon stopped by one of the three *artificial* stop codons shifted one nucleotide to the left or right.

Example 1. Consider the following gene that codes one polypeptide strand of human hemoglobin. For clarity, we separate all codons by spaces (even though no separation marks are on the real DNA double helix). We observe that the first (start) codon ATG corresponds to methionine and the last triplet TAA is the stop codon. Further triplets are denoted by small letters as they do not belong to this gene.

ATG GTG CAC CTG ACT CCT GTG GAG AAG TCT GCC GTT

ACT GCC CTG TGG GGC AAG GTG AAC GTG GAT GAA GTT

GGT GGT GAG GCC CTG GGC AGG CTG CTG GTG GTC TAC

CCT TGG ACC CAG AGG TTC TTT GAG TCC TTT GGG GAT

CTG TCC ACT CCT GAT GCA GTT ATG GGC AAC CCT AAG

GTG AAG GCT CAT GGC AAG AAA GTG CTC GGT GCC TTT

AGT GAT GGC CTG GCT CAC CTG GAC AAC CTC AAG GGC

ACC TTT GCC ACA CTG AGT GAG CTG CAC TGT GAC AAG

CTG CAC GTG GAT CCT GAG AAC TTC AGG CTC CTG GGC

AAC GTG CTG GTC TGT GTG CTG GCC CAT CAC TTT GGC

AAA GAA TTC ACC CCA CCA GTG CAG GCT GCC TAT CAG

AAA GTG GTG GCT GGT GTG GCT AAT GCC CTG GCC CAC

AAG TAT CAC TAA gct cgc ttt ctt gct gtc caa ttt

cta tta aag...

Notice that the start triplet (shifted to the left) with space A TG does not appear in this genomic sequence, whereas another start triplet (shifted to the right) with space AT G (printed overlined) appears five times. If the genetic information is read from this wrong triplet AT G, where T is replaced by U in RNA, then the synthesis of proteins is relatively quickly stopped by one of the three frameshifted stop triplets TA A , TA G , or TG A (for clarity they are underlined). The synthesis is also stopped when any other undesirable shift of the reading frame appears. In this sophisticated manner, cells do not produce undesirable proteins (cf. (1)) and substantially save energy. Note that the above gene even contains 11 frameshifted stop triplets T AA, T AG, and T GA by one nucleotide to the left.

Notice that the cyclic permutation of the start codon ATG yields the stop codon TGA, i.e., the start codon shifted one nucleotide to the right leads to an immediate stop provided the next base is A. The fact that there exists only one start codon AUG in the standard genetic code (see Table 1) has also a certain evolutionary advantage, since the number of positions, from where the genetic information is read, is minimal. If there were two or more start codons in a genetic code, then the probability of reading frameshifted start codons would be larger, which would lead to a larger production of dysfunctional proteins than for one start codon.

Finally, let us emphasize that ATG is not the only universal start codon. For instance, ATA (corresponding to the purine bases mutation G → A in the third position) stands for the start in some mitochondrial genetic codes. Also ATT may rarely serve as a special start codon, see Faure et al. (2011).

3. Frequency of frameshifted stop codons in the Drosophila genome

We examine first theoretically and then practically the situation, when the genetic information is read shifted by one nucleotide to the left or right. For simplicity we assume that shifted triplets are distributed randomly with the same probability. Then the probability that we read one of the three stop codons in one step is

$$p=\frac{3}{64}=4.6875\%. \tag{2}$$

Hence, the probability that these three codons will stop reading of shifted information just at the nth step is $p(1-p)^{n-1}$. Taking into account that

$$1=(p+(1-p))^n,$$

we get by the binomial theorem (Rektorys, 1994) that the probability $P(n)$ of reading at least one stop codon within n steps is given by

$$P(n)=p^n+\binom{n}{1}p^{n-1}(1-p)+\cdots+\binom{n}{n-1}p(1-p)^{n-1}=1-(1-p)^n. \tag{3}$$

From this we find that the probability of reading at least one shifted stop codon is more than 50% after 15 steps, since $1-(1-p)^{15}>0.5$. After 96 steps this probability will be even higher than 99%.

For some archaebacteria, UGA is not a stop codon (cf. Table 1), but it codes selenocysteine, and there are only two stop codons (i.e. $p=1/32$). In this case the probability of reading at least one stop codon is more than 50% after 22 steps and after 100 steps this probability will be almost 96% due to formula (3).

If there were only one stop codon (i.e., $p=1/64$), then by (3) the probability that we read this codon is more than 50% after 45 steps and after 100 steps this probability will be less than 80%. For instance, the mitochondrial code of the flatworm has only one stop codon UAG, see Seligmann and Pollock (2004). These facts illustrate (see Fig. 2) why three stop codons are so advantageous.

Test 1. The above theoretical data were compared with real data from GenBank1. A single nucleodite polymorphism from the Drosophila Heterochromatin Genome Project available in GenBank2 was taken into account, but appeared to be negligible. The genomic sequence containing only genes of the third chromosome was shifted one nucleotide to the left. In this case, the mean density of frameshifted stop codons was about 5.31 per 100 codons (cf. formula (2)). From Fig. 2 we observe that the probability of reading at least one shifted stop codon is more than 50% after 13 steps. After 62 steps this probability will be higher than 99%. Therefore, the probability observed in the genome of reading one of the frameshifted stop codons is even higher than the theoretical probability given by (3).

The main reason of this remarkable phenomenon is that the frequency of all bases A and T in this particular instance is higher than that of C and G (see Table 2). Therefore, the probability of reading one of the frameshifted stop codons TAA, TAG, and TGA can be higher than $p=3/64$.

Another reason is that genes do not contain these three codons except for the last term and thus, the theoretical probability of their appearance in shifted positions (± 1) is slightly higher than the probability $p=3/64$ corresponding to a random distribution of codons. Furthermore, we observe that the stop codons TAA, TAG, and TGA do not overlap when frame is shifted (± 1).

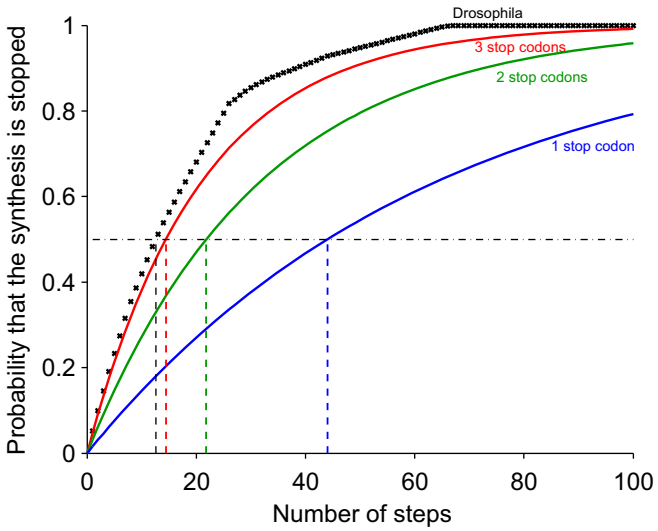

Fig. 2. Comparison of the theoretical probability that reading of a frameshifted genetic information is terminated after n steps with the probability based on real DNA data. The theoretical probabilities are depicted by solid lines. Cross markers correspond to the third chromosome of Drosophila shifted by one nucleotide to the left.

Test 2. According to Crick (1968, p. 369) any two codons differing only in C–U in the third position cannot be distinguished by the translation apparatus, as they code the same amino acid (see Table 1). Since none of the three stop codons in the standard genetic code contains cytosine, it is clear that random mutations from C to U (or C to T) increase, in general, the number of frameshifted stop codons.

Note that thymine and uracil have almost the same chemical structure. The only difference is that the methyl group CH_3 in thymine (see the left upper part of Fig. 3) is replaced by hydrogen in uracil. Since cytosine has a similar structure as thymine (see Fig. 3) and uracil, it is well known that C → T and C → U mutations of these pyrimidine bases are not rare.

Table 2 shows frequencies of particular bases for the same input data as in previous Test 1. Notice that the ratio between the total number of T's and C's is $\#T:\#C=1.403$ for all bases, whereas this ratio $\#T:\#C=1.783$ is higher when only the third bases are counted. This test illustrates that C → U and C → T mutations in the third position of codons became advantageous during the evolution of life—supporting the ambush hypothesis.

Test 3. The mitochondrial genome of *Drosophila melanogaster* is available in GenBank3. From Table 3 we observe that the frequency of all (A + T)'s is about 3.2 times larger than that of (C + G)'s. However, the frequency of (A + T) in the third position is at least 16 times larger than that of (C + G) in the third position. Similar results have been obtained also for *Drosophila acutilabella*, *Drosophila cardini*, etc., from the same GenBank3. These tests again illustrate that C → U and C → T mutations in the third position of codons are profitable (compare the last two rows in Table 3), since they decrease energy costs during frameshifted protein synthesis.

Denoting by N an arbitrary nucleotide, we find that N → A mutations in the third position could also be advantageous (see Table 3) because of the frameshifted stops UA A and UA G.

Table 2
Frequencies of bases in the Drosophila genome.

Base	A(%)	C(%)	G(%)	T(%)
All bases	31.8	21.7	16.0	30.5
Third bases	29.5	19.8	15.4	35.3

Table 3
Frequencies of bases in the mitochondrial *Drosophila melanogaster* genome.

Base	A(%)	C(%)	G(%)	T(%)
1st bases	30.9	11.3	20.2	37.6
2nd bases	20.1	20.6	13.7	45.6
3rd bases	45.5	**3.5**	2.2	**48.8**
All bases	32.2	**11.8**	12.0	**44.0**

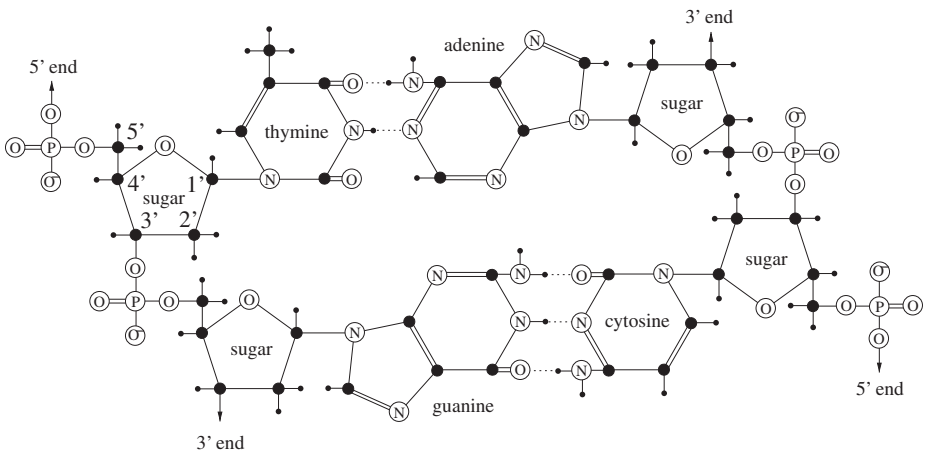

Fig. 3. Chemical structure of DNA. Hydrogen bonds shown as dotted lines.

Table 4
Frequencies of selected bases in the third position and in total in genomes of several simple species.

Genome	C in 3rd pos. (%)	C in all bases (%)	T in 3rd pos. (%)	T in all bases (%)
Escherichia coli	26.1	24.2	26.2	23.9
Plastid *E. coli*	25.6	23.5	27.7	22.6
Mycoplasma agalactiae	9.8	13.3	38.2	28.5
Acholeplasma laidlawii	11.9	15.2	38.7	29.2
Hydra vulgaris	11.2	16.5	39.9	28.5
mt *Hydra vulgaris*	3.4	10.2	54.2	44.9

Test 4. In Table 4 we observe similar phenomena to those that appear in Table 3 for some other simple species whose genomes are also available in GenBank3. Namely, the percentage of T's in the third position is higher than in all positions (see the last two columns of Table 4). This is advantageous from the evolutionary point of view, since there are more hidden stops, in general, and again supports our working hypothesis.

The degeneracy of genetic codes thus enables one to gain new capability that may essentially save energy expenses of cells. Recently Seligmann (2012) has also found exceptional roles of bases appearing in the 3rd codon positions.

4. Conclusions

Assigning A=00, C=01, G=10, U=11, we observe that genetic information can be rewritten in the binary digital system, which was in fact first discovered by nature 3.5 Gyr ago, i.e., much earlier than by man. Moreover, the genetic code during its long-term evolution gained many useful properties (Crick, 1968). Since it is largely redundant, there is still a space for some hidden secondary functions of this code as we have shown in Tests 1–4.

In particular, we found that in an AT-rich DNA sequence the probability of detection of frameshifted stop codons in nuclear and mitochondrial genomic data is even higher than the theoretical value. Frameshifted stop codons thus increase average rate of protein production, since long dysfunctional proteins after ribosomal slippages are not produced.

Our tests from Section 3 support the ambush hypothesis for the genome of Drosophila and several standard simple organisms. This hypothesis implies that hidden stop codons prevent frameshifted gene reading, and thus, the number of hidden stops reflects the complete genome. However, this topic deserves further exploration for other genomic data from GenBank4, GenBank5, and GenBank6, etc.

Acknowledgments

The authors thank Guy Hagen, Lucie Kárná, František Katrnoška, Christian Lanctôt, Roman Pipek, Hervé Seligmann, Lawrence Somer, Jarmila Štruncová, and the reviewers for insightful and encouraging comments, and spirited discussions. The paper was supported by Grant no. IAA 100190803 of the Grant Agency of the Academy of Sciences of the Czech Republic, RVO 67985840, and by Grant no. P205/12/P392 of the Grant Agency of the Czech Republic.

References

Alberts, B., et al., 1997. Essential Cell Biology. Garland Publishing, New York.
Bove, J.M., 1993. Molecular features of mollicutes. Clin. Infect. Dis. 17, S10–S31.
Clarke, C.H., Miller, P.G.G., 1982. Consequences of frameshift mutations in the trp A, trp B and lac I genes of *Escherichia coli* and in *Salmonella typhimurium*. J. Theor. Biol. 96, 367–379.
Crick, F.H.C., 1968. The origin of the genetic code. J. Mol. Biol. 38, 367–379.
Crick, F.H.C., Griffith, J.S., Orgel, L.E., 1957. Codes without commas. Proc. Natl. Acad. Sci. USA 43, 416–421.
Faure, E., Delaye, L., Tribolo, S., Levasseur, A., Seligmann, H., Barthélémy, R.-M., 2011. Probable presence of an ubiquitous cryptic mitochondrial gene on the antisense strand of the cytochrome oxidase I gene. Biol. Direct 6, 56.
Gamow, G., 1954. Possible relation between deoxyribonucleic acid and the protein structures. Nature 173, 318.
GenBank1. ⟨http://flybase.org⟩.
GenBank2. ⟨http://www.fruitfly.org/SNP/3-Exelixis.html⟩.
GenBank3. ⟨http://www.kazusa.or.jp/codon⟩.
GenBank4. ⟨http://www.ncbi.nlm.nih.gov⟩.
GenBank5. ⟨http://www.bioinf.man.ac.uk/ogre⟩.
GenBank6. ⟨http://www.ensembl.org/index.html⟩.
Itzkovitz, S., Alon, U., 2007. The genetic code is nearly optimal for allowing additional information within protein-coding sequences. Genome Res. 17, 405–412.
Johnson, L.J., et al., 2011. Stops making sense: translational trade-offs and stop codon reassignment. BMC Evol. Biol. 11, 227.
Nirenberg, M.W., Matthaei, J.H., 1961. The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. Proc. Natl. Acad. Sci. USA 47, 1588–1602.
Păun, G., Rozenberg, G., Salomaa, A., 2006. DNA Computing. New Computing Paradigms. Springer, Berlin.
Rektorys, K., 1994. Survey of Applicable Mathematics. Kluwer, Dordrecht.
Santos, M.A., Moura, G., Massey, S.E., Tuite, M.F., 2004. Driving change: the evolution of alternative genetic codes. Trends Genet. 20, 95–102.
Seligmann, H., 2007. Cost minimization of ribosomal frameshift. J. Theor. Biol. 249, 162–167.
Seligmann, H., 2010a. The ambush hypothesis at the whole-organism level: off frame hidden'stops in vertebrate mitochondrial genes increase developmental stability. Comput. Biol. Chem. 34, 80–85.
Seligmann, H., 2010b. Avoidance of antisense, antiterminator tRNA anticodons in vertebrate mitochondria. Biosystems 101, 42–50.
Seligmann, H., 2011. Two genetic codes, one genome: frameshifted primate mitochondrial genes code for additional proteins in presence of antisense antitermination tRNAs. Biosystems 105, 271–285.
Seligmann, H., 2012. Coding constraints modulate chemically spontaneous mutational replication gradients in mitochondrial genomes. Curr. Genomics 13, 37–54.
Seligmann, H., Pollock, D.D., 2003. The ambush hypothesis: hidden stop codons prevent off-frame gene reading. Midsouth Comput. Biol. and Bioinf. Soc., Abstract 36.
Seligmann, H., Pollock, D.D., 2004. The ambush hypothesis: hidden stop codons prevent off-frame gene reading. DNA Cell. Biol. 23, 701–705.
Singh, T.R., Pardasani, K.R., 2009. Ambush hypothesis revisited: evidences for phylogenetic trends. Comput. Biol. Chem. 33, 239–244.
Tse, H., Cai, J.J., Tsoi, H.-W., Lam, E., Yuen, K.-Y., 2010. Natural selection retains overrepresented out-of-frame stop codons against frameshift peptids in prokaryotes. BMC Genomics 11, 491.
Watson, J.D., 1999. Genes, Girls and Gamow. Alfred A. Knopf, New York.
Watson, J.D., Crick, F.H.C., 1953. Genetic implications of the structure of deoxyribonucleic acid. Nature 171, 964–969.
Woese, C.R., 1965. Order in the genetic code. Proc. Natl. Acad. Sci. 54, 71–75.
Zhang, Y., Gladyshev, V.N., 2007. High content of proteins containing 21st and 22nd amino acids, selenocysteine and pyrrolysine, in a symbiotic deltaproteobacterium of gutless worm *Olavius algarvensis*. Nucleic Acids Res. 35, 4952–4963.