International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor 2024: 7.101

Reassessing the Notion of a Second Genetic Code, A Historical Review and a Path Toward Clearer Genomic Classification

Bijan B Mohanty

E-Mail: bijanstudy[at]gmail.com

Abstract: ***In this review article, the author describes the circumstances behind the origin of the term "Second Genetic Code". its expected features, and the hype generated by the so-called discovery of the code, fueled by media reports as far back as in 1988, only to be proved untrue subsequently. However, every time a sequence code was found in DNA or protein, the same was touted as the Second Genetic Code, and media publicity buzz followed. It almost became a joke that there were four to five Second Genetic Codes doing the rounds. None of these so-called codes either resembled or had any link, either structural or functional, with the Standard Genetic Code. The author charts out those chronological developments and attempts to suggest a fresh classification of genomic codes to avoid the prevalent confusion and put the codes in a more defined framework. The above confusing scenario also prompts one to ponder objectively as to what is exactly meant by the Second Genetic Code and whether such a code exists at all. The same was recently addressed in a quantitative manner, which raises a fresh hope of discovering the second genetic code.***

Keywords: Genetic Code, Standard Genetic Code, Second Genetic Code, acceptor stem code, genomic code sequence

1.Introduction

After the discovery of DNA structure in 1953, focus shifted to unraveling the puzzle of transfer of chemical information stored in the string of nucleotide bases of DNA leading to the assembly of string of amino acids called proteins. Many legendary brains of that time slogged over the issue and exciting progress was made. George Gamow, a physicist who strayed into molecular biology, first computed that the units of chemical information had to be a set of minimum 3 bases out of the 4 designated ones, i.e. Adenine (A), Guanine (G), Cytosine (C) and Thymine (T) / Uracil (U). Calculation showed that there were 64 such sets. Protein chains are made from the pool of 20 specific amino acids (AA) and hence he predicted that many of these units would be degenerate with regard to an AA. Crick, in 1962, coined the term CODON for these triplet units [1]. Those were interesting times because, in parallel, exciting strides were being made in information science and technology pioneered by stalwarts like John Von Neumann, Alan Turing, and Shannon. CODE was a favourite word of information scientists and CODON meant code inherited which reflected the innate link between biology and information science at most rudimentary level. In 1962, Marshall Nirenberg and Matthei linked the codon **uuu** to amino acid phenylalanine through a series of experiments known as Poly-U experiments. In next five years, all 64 codons were linked to the 20 AA along with stop signal.

In the meantime, messenger RNA was discovered by two teams, one having Gross and Watson [2] and the other having Jacob and Sydney Brenner [3], in 1961. In 1958, tRNA was discovered by Zamecnik & Hoagland [4]. Both tRNA and mRNA were key players in the process of transcription and translation of genetic information resulting in assembly of amino acid chains. Of particular importance was the tRNA molecule which acted as an adaptor molecule as predicted by Crick in his adaptor hypothesis [5].

2.Developments

Why Search for Second Genetic Code?

Typical functioning of the tRNA molecule lay at the centre of frantic search for the Second Genetic Code. Individual amino acids are charged to the respective tRNAs at position A76 of the 3'-end of the acceptor stem by respective synthetase enzymes called aaRS. By respective tRNA, we mean a tRNA molecule which contains in its anticodon loop, an anticodon corresponding to the same AA which is charged to the same tRNA at acceptor stem. tRNA has a L shaped structure with anticodon loop at one extremity of L where identity of AA is embedded and aminoacylation site at the other extremity where vetting of the AA identity of anticodon loop is done. The distance between the two extremities is about 76 angstrom units which is the farthest distance between any two positions of L. Biologists could not identify any vetting mechanism which would transmit the AA identity information from anticodon loop to acceptor stem site of aminoacylation.

While research was going on to establish a link between anticodon loop and acceptor stem, a famous molecular biologist Christian de Duve appeared on the scene. His article "The Second Genetic Code" published in journal "Nature" in May 1988 [6] created a sensation. In the article, de Duve predicted about the existence of a Second Genetic Code in either synthetase molecules or in acceptor stem of tRNA which independently carried the identity of the same AA which was already identified by the anticodons. Duve did not elaborate how two different sites in the same RNA identified the same AA but had clearly hinted at a possible chemical similarity between the two which made this identification synchronized. It was also suggested that the Second Genetic Code was established prior to the Standard Genetic Code (SGC) and amino acids probably exhibited degeneracy with respect to a codon instead of the reverse situation as in case of SGC. This was a remarkable article

for it precisely predicted the defining features of the possible Second Genetic Code. It stated, *inter alia*, that the codons might be or might not be triplets but would be some multiplets and location would be more likely in aaRS. In the same issue of "Nature", another article was published by Schimmel et al [7] which identified a particular base pair which recognized AA Alanine. We reproduce the abstract.

"Analysis of a series of mutants of an Escherichia coli alanine transfer RNA shows that substitution of a single G-U base pair in the acceptor helix eliminates aminoacylation with alanine in vivo and in vitro. Introduction of that base pair into the analogous position of a cysteine and a phenylalanine transfer RNA confers upon each the ability to be amino acylated with alanine. Thus, as little as a single base pair can direct an amino acid to a specific transfer RNA".

The news was so sensational, New York Times posted a news item [8] under the headlines "Second Genetic Code Deciphered, Solving a Protein Synthesis Puzzle" on 13th May, 1988.

In 1995, Schimmel again reported about species-specific aminoacylation of minihelices switched by a single nucleotide. But, this time, he talked about an Operational RNA Code because a single nucleotide driven recognition process did not qualify to be termed as a Genetic Code. Subsequently, quite a few cases were reported wherein either a single base, or a pair of bases (either contiguous or not) served as recognition elements for a particular AA. Such base conglomerations were located mostly in acceptor stem and variable stem. All such AA identifiers were categorized as Operational RNA codes. Thus, the term Operational RNA Code gradually substituted the term Second Genetic Code.

But still the excitement of deciphering the second genetic code persisted for right reasons and papers were published where the title either incorporated the term Second Genetic Code or the scientific press releases announced the cracking of the second genetic code. Over a period, 4 to 5 second genetic codes appeared, creating confusing and funny situation prompting biophysicist Edward F Trifonov **[16]** to remark as under-

'Bewildered reader, naturally, would say "I'm done with seconds, can I have a third?" '

We put below a tentative list of 5 number of second genetic codes and then shall discuss whether they qualify to be called second genetic codes-

Table 1: Articles on Second genetic code

Serial	Dt	Article title	content	Press Reference
1	12/5/ 1998	A simple structural feature is a major determinant of the identity of a transfer RNA [7]	Recognition of tRNA by aaRS enzymes	New York Times [8] Science [8A]
2	10/8/ 2001	Translating the Histone Code [9]	Post translational modification of chromatin that impinges on histone amino-terminals, regulating access to underlying DNA	NewScientist [10]
3	19/7/ 2006	A genomic code For nucleosome positioning [11]	To construct a Eukaryotic nucleosome-DNA interaction model which suggests that genomes encode an intrinsic nucleosome organisation that explain about half of the nucleosome positioning.	New York Times [12]
4	1/3/ 2008	Cracking the second genetic code [13]	Interaction specificities between protein and DNA. Binding specificities of 300 mouse DNA-binding proteins established	
5	6/5/ 2010	Deciphering the splicing code [14]	Set of rules governing gene splicing	Nature [15]

Before one examines the above sequence codes, it is essential to understand and appreciate the threshold characteristics of a code to qualify as a candidate for second genetic code. For this, one must go back to the origin of the necessity followed by quest for the same and de Duve's conceptualization of the would-be code. The Author outlines his own basic expectation of features of the second genetic code based on above considerations –

(1) Its role should be identical to that of Standard Genetic Code. It should consist of codons/anticodons which shall encode the 20 amino acids.
(2) The codon length may be triplets or doublets or quadruplets or any multiplets.
(3) The entire codon pool should fully map the entire amino acid pool of 20 AA. There should not remain any unallocated codons or amino acids
(4) The Standard Genetic Code is a 4-alphabet code. The number of alphabets participating in the proposed second genetic code may be four or less than that but the number should have a logical basis.
(5) All the information units which are codons /anticodons in the instant case should have fixed code length each. Variable code lengths are not permitted. The basic premises behind this assumption is that variable code length shall destroy most of the symmetry manifested in the SGC **[18]**
(6) The second genetic code should be amenable to digital framework in terms of binary formulation just like SGC [**17**].

With above yardsticks in place, none of the sequence codes in **Table-1** qualify to be categorized as second Genetic codes.

Before examining the status of all the codes mentioned in **Table-1**, the Author proposes a fresh classification criterion for codes which are not genetic code variants.

The codes as mentioned in **Table-1** arise out of the interaction specificities involving two broad entities i.e. proteins and nucleic acids (RNA, DNA) which may be as under

(a) **Protein-nucleic acid interaction**-we call them operational hybrid codes
(b) **Protein-protein interaction** - we call them as operational protein codes.
(c) **Nucleic acid-nucleic acid interaction**-we call them as operational genomic codes.

In each category, we subdivide the codes into 3 sub categories based on the functional aspect of interaction. These are-

(a) **Switch codes**: The codes which result in activating (ON) or deactivating (OFF) a particular function.
(b) **Regulatory Codes**: The codes which regulate the speed or frequency of a particular task thereby affecting the degree of effectiveness.
(c) **Recognition Codes**: The codes which recognise another code or sequence, contiguous or discrete and the recognition is followed by an action other than switch action or regulatory action. Gene splicing may be put in this category.

We incorporate the word 'operational' in the same logic as described for above structurally classified codes.

Hence, each code can be one of the following 9 variants—

(ad), (ae), (af), (bd), (be), (bf), (cd), (ce), (cf).

Armed with above classification methodology, we tabulate the category of each code of Table-1.

Table 2: Classification of supposedly second genetic codes)

Serial No.	Article Title	Category of Code
1	A simple structural feature is a major determinant of the identity of a transferRNA [7]	(af) Recognition followed by aminoacylation. Known as Operational RNA codes
2	Translating the Histone Code [9]	(ae)
3	A genomic code For nucleosome positioning [11]	(af) Recognition followed by nucleosome organization
4	Cracking the second genetic code [13]	(ae)
5	Deciphering the splicing code [14]	(af)

The word operational is used to indicate that the underlying logic of the code structure is not yet deciphered. Upon deciphering, the word may be dropped.

These codes are the result of mostly post translational chemical modifications of DNA, RNA and associated proteins which act as epigenetic markers.

The Search for elusive Second Genetic Code:

In the meantime, while the search continues, the author succeeded in computationally arriving at the maximum number of possible genetic codes and the possible features of the second genetic code **[19]** which shall assist in discovering the same.

3.Conclusion

The review article traces brief chronological history of Genetic Code since discovery of the Standard Genetic Code, the compulsions which led to the search for the second genetic code, discovery of various types of codes and passing them as the second genetic code with accompanying confusion. Author reviews those codes which were touted as Second Genetic Codes and expresses his opinion and suggestion to tide over the confusion.

Author's Contribution

Author is an independent researcher and is not affiliated to any organization. He has solely contributed to the research funding.

Conflict of Interest: Author declares no conflict of interest.

References

[1] H. Chant Renne, "Information in Biology", Nature, 197, 27-30, 1963
[2] Gros F., Hiatt H., Gilbert W., Kurland C. G., Risebrough R. W. and Watson J. D., "Unstable Ribonucleic Acid Revealed by Pulse Labelling of Escherichia Coli, " *Nature*, 190, 581-585 (1961).
[3] Brenner S., Jacob F. and Meselson M, "*An Unstable Intermediate Carrying Information from Genes to Ribosomes for Protein Synthesis*, " *Nature*, 190, 576-81 (1961).
[4] HOAGLAND MB, STEPHENSON ML, SCOTT JF, HECHT LI, ZAMECNIK PC. J A soluble ribonucleic acid intermediate in protein synthesis. Biol Chem.1958 Mar; 231 (1): 241-57.
[5] Crick's Adaptor Hypothesis, Wikipedia
[6] Christian de Duve". The Second Genetic Code", Nature Volume 333, 12 May 1988
[7] Y M Hou & Paul Schimmel, "A simple structural feature is a major determinant of the identity of a transfer RNA", Nature Volume 333, 12 May 1988
[8] Jina Colata, "Second Genetic Code deciphered, Solving a Protein Synthesis Puzzle ", New York Times, 13, May, 1988
[8A] M M Waldrop', The structure of the "second genetic code" ', Science.1989 Dec 1; 246 (4934)
[9] Thomas Jenuwein and C David Allis, "Translating the Histone Code ", Science Vol.293, Issue 5532, pp 1074-1080, 10 Aug 2001
[10] Emma Young, "Packaging proteins may be second genetic code ", NewScientist, 9 Aug 2001
[11] Eran Segal et al, "A genomic code for nucleosome positioning ", Nature, 442, 772-778, 19 July 2006

[12] Nicholas Wade, "Scientists say They've found a code beyond Genetics in DNA ", New York Times, 442, 25 July 2006
[13] Tim Hughes, "Cracking the second genetic code", FASEB Journal, Volume 22, Issue S1, 01 March 2008
[14] Joseph Barash et al, "Deciphering the splicing code ", Nature, 465, 53-59, 6 May 2010
[15] J. Ramon Tejedor & Juan Val Caral, "Breaking the second genetic code ", Nature, 465, 45-46, 5 May 2010
[16] Edward F Trifonov, "Thirty Years of Multiple Sequence Codes ", Genomics, Proteomics Bioinformatics, 2011 Apr; 9 (1-2): 1-6
[17] Bijan Bihari Mohanty, "Are Binary Codes assigned to bases in codons a matter of Choice?"
[18] Bijan Bihari Mohanty, "Unique Binary representation of Bases & symmetry In the Standard Genetic Code ".
[19] Bijan Bihari Mohanty, "Does a Second Genetic Code really exist?"