

Mid Term Report

TITLE: EVOLUTION OF THE STANDARD GENETIC CODE

Course: Evolutionary Dynamics(CI 717)

Instructor: Supreet Saini

Name: Pulkit Ahuja

Roll Number: 150020056

INDEX

SNo	Title	Page No.
1	Introduction	3
2	Important Areas of Concern	4
3	Theories to explain Genetic Code	5
4	Important Definitions	7
5	The problem	8
6	Possible Approach	9
7	Conclusion	9
8	Reference	10

INTRODUCTION

Across all the organisms on Earth - plants, animals, microorganisms, etc. , the mapping of Codons to amino acids during the process of Translation remains identical. This is referred as Standard Genetic Code.

During the process of translation, mRNA is decoded in Ribosome in order to produce specific Amino acid sequences. In the process, the bases (4, Adenin, Guanine, Uracil, Cytosine) are read in form of triplets (codons) which translates to different Amino acids. The number of ways in which this mapping can be done can be thought of as a problem of putting $4^3 = 64$ distinct balls in 21 (20 amino acids + 1 end) boxes such that each box has at least one ball which comes out to be nearly 10^{84} .

Hence, the question arises is that out of this number of ways of mapping that is possible, why was a particular mapping got selected? In other words, What is so special with the Standard Genetic code?

		Second letter					
		U	C	A	G		
First letter	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA Stop UAG Stop	UGU } Cys UGC } UGA Stop UGG Trp	U C A G	Third letter
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	U C A G	
	A	AUU } AUC } Ile AUA } AUG Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U C A G	
	G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }	U C A G	

IMPORTANT AREAS OF CONCERN

Some of the important questions which needs to be answered are -

1. Why does block structure exist in Genetic Code?
2. What is the link between the second codon position and nature of the Amino acid?
3. What is the relationship between the second codon position and the class of aminoacyl t RNA synthetase?
4. Why is there a negative correlation between molecular weight of codon and number of codons allocated to it?
5. Why is there a positive correlation between number of synonymous codons for an amino acid and the frequency of the amino acid in proteins?
6. Why does Genetic code minimise the likelihood of mistranslation and point defects?

The first three question can be answered by Crick's theory of code extension which argues that primitive code was not triplet code but that originally the bases were read one at a time, then 2 at a time and only later evolved to the present triplet code. This would have lead to block structure of the genetic code. Also if we assume that the second codon position was first to be read this explains the greater influence of second codon position (with the help of stereochemical theory)

The other three questions don't have such a clear theory that can explain their existence. Hence these are some questions worth dwelling into.

THEORIES TO EXPLAIN GENETIC CODE

There are different theories given in order to explain the existence of Standard Genetic Code. Some of them are -

1. Stereochemical Theory

- Given by Gamow in 1954
- the codon assignments for particular amino acids are determined by a physicochemical affinity that exists between the amino acids and the cognate nucleotide triplets.
- Hence, there is a possibility that the specific structure of the code is not at all accidental but necessary and possibly unique
- Generally deemed improbable because of the failure of direct experiments to identify specific interaction between amino acids and cognate triplets
- weak and relatively nonspecific interactions between amino acids and their cognate triplets
- **“Escape triplet theory”** - the primordial code functioned through interactions between amino acids and cognate triplets that resided within amino acid binding RNA molecules

2. Frozen Accident Theory

- the code is universal because at present time, any change would be lethal
- This is because, in all organisms, the code determine the amino acid sequence of so many highly evolved process that any change to these would be highly disadvantageous
- Assumption that all life evolved from a single organism (LUCA)
- In its extreme form, the theory implies that the allocation of codons to amino acids at this point was entirely a matter of “chance”.

3. Crick's primitive code theory

- given by F.H.C. Crick(1968)
- Propose that at early step only a few amino acid were coded
- Initially the machinery only recognised one base, then slowly it started recognising 2 base codon and finally 3 base codon.
- More amino acids were selected in order to remove ‘cavity’ in the translation process

4. Coevolution Theory

- Structure of the standard code reflects the pathway of amino acid biosynthesis
- The code coevolved with the amino acid biosynthesis pathways ie during the code evolution, subset of codons for precursor amino acids have been reassigned to encode product amino acid
- Basic idea same as Crick's scenario of code extension
- Explicit identification

5. Adaptation Theory

- The structure of the genetic code was shaped under selective forces that made the code maximally robust
- Hence, it minimised the effect of errors on the structure and function of the synthesised proteins.
- So, the standard code evolved to minimise the effect of point mutations - **“lethal mutation” hypothesis**
- Most important pressure in the code's evolution was selection for minimisation of effect of the translational misreadings.
“translational error minimisation”
- Second position correlates best with amino acid properties is considered as evidence.

There are also possibilities of combination of two or more of these theories with varying factors to explain the solution. This may represent a more real solution as there is a good possibility that a combination of these factor were at work simultaneously during the evolution/selection.

Important Definition

1. **Codons** - Triplet of gene base pairs which during translation of RNA into protein translates to a particular amino acid.
2. **mRNA** - RNA formed by the process of transcription to copy message of DNA in order to be translated into portion
3. **Transcription** - The process of copying of message of DNA to RNA inside nucleus
4. **Translation** - The process of creation of specific protein through RNA with the help of Ribosome.
5. **Amino acids** - They are organic entities with $\text{NH}_2\text{-C-COOH}$ functional group and are the monomers for protein.
6. **Nucleic Bases** - nitrogen-containing biological compounds that form nucleosides. They are used to code DNA and RNA with information.
7. **Standard Genetic Code** - The mapping codon to Amino acid that is standard throughout all living organism
8. **LUCA(Last Universal Common Ancestor)** - It is the most recent population of organisms from which all organisms now living on Earth have a common descent.
9. **Fitness landscape** - Evolutionary landscape used to visualise the relationship between genotypes and reproductive success. It is assumed that every genotype has a well-defined replication rate called fitness which is represented as height of the landscape.
10. **Genetic Algorithm** - Genetic Algorithms are the heuristic search and optimisation techniques that mimic the process of natural evolution.
11. **Synonymous Codons(Same sense codon)** - Pair of Codons that translates to same Amino acid

THE PROBLEM

A lot of research work is done to answer the question of what makes the Genetic code special enough to be selected from other 10^{84} other options for mapping. As mentioned, there is a good possibility that the solution to the problem lies as a combination of 2 or more of these theories. So one question that one will like to answer is -

What combination of these theories can be closer to realistic picture?

Also, I would like to work on one of the six questions mentioned above-

Why is there a positive correlation between number of synonymous codons for an amino acid and the frequency of the amino acid in proteins?

As there is an elementary logic present i.e. the higher need of a particular amino acid in proteins may have led to more number of synonymous amino acid but the relation can be vice versa i.e due to the presence of more synonymous codons, the frequency of amino acid can be higher.

POSSIBLE APPROACH

Some plausible direction can be -

1. Coevolution theory with combination of error minimisation in order to account the robustness of standard code.
2. Combined scenario with selection for error minimisation in early stages and stereochemical affinities at later stages (which explains the unusual position for Arginine.)

For the second question, direction will be in the basic element of need of more of one amino acid which leads to the selection of particular codon or the presence of more synonymous codons that lead to particular amino acid. Though I will search for other possible approaches for this problem.

CONCLUSION

A lot of theories have been suggested to explain Standard genetic code, which serves as the basis of protein generation in all living organisms. In my project, I will try to analyse what combination of theories is best suited to explain the choice of the code and also on the reason for presence of synonymous codons in the Standard genetic code

REFERENCES

1. Balaji Kumar, Supreet Saini 2016 (<http://pubs.rsc.org/en/Content/ArticleHtml/2016/MB/c6mb00262e>)
2. E. V. Koonin and A. S. Novozhilov, IUBMB Life, 2009
3. F. H. Crick, J. Mol. Biol., 1968
4. NPTEL Genetic Algorithm (<https://nptel.ac.in/courses/105108081/37>)
5. Toshimichi Ikemura J. Nol. Biol. (1981) 151, 389-409