A further thought of genetic coding of RNA on the basis of theological biology

Wan-Chung Hu
Department of Clinical pathology, Taipei Tzu Chi Hospital, New Taipei City, Taiwan

Abstract
Genetic codons are a three-letter code to be related to specific amino acids. Since there are four nucleotides including adenine, guanine, cytosine, and uracil, the permutation for the three-letter codons will lead to 4x4x4=64 possible codes for amino acids. But, essential amino acids are only twenty kinds plus start and end codons. Thus, the third position of the genetic codon is called wobble code. The first two codes can lead to 4x4=16 permutations compared to 20 kinds of essential amino acids. Thus, the first two codes are more important to decide the amino acids in evolution. And, hydrophilic amino acids are related to adenosine in codons and hydrophobic amino acids are related to uracil in codons. Both have reasons in theological biology.

Main text
The central dogma of biology is DNA transcribed to RNA and RNA translated to protein. The genetic codes can decide the sequences of amino acids composing proteins. This is decided by the genetic codons on mRNA which reflect the DNA sequence. The genetic codon is a three-letter code consisting of three nucleotides. Because we have four kinds of nucleotides including adenine (A), cytosine (C), guanine (G), and uracil (U), we can have 4x4x4=64 permutations to link to each type of amino acid. However, we only use twenty kinds of essential amino acids to build the protein. Including start and termination code, we will only need 22 codons for translating to proteins. The third position is called wobble code, and the specific amino acid is mainly decided by the first two codes of genetic codon. But, in special situation, the third position of genetic code is important[1]. We will discuss this later.

According to previous studies and observations, the codons can be divided into two groups: hydrophilic XAX codons and hydrophobic XUX codons. And, the neutral or biphasic codons are XCX or XGX codons[2,3]. There is no satisfied explanation for this phenomenon yet. I think this should be due to different solubility of tRNA carrying hydrophobic or hydrophilic amino acids. There is an anti-codon on tRNA. If the genetic codon on mRNA is XAX, then the anti-codon on tRNA is YUY. If the genetic codon on mRNA is XUX, then the anti-codon on tRNA is YAY. X and Y are complimentary nucleotides. Adenine is hydrophobic and uracil is hydrophilic. Thus,

tRNAs with YUY are more hydrophilic and tend to carry hydrophilic amino acids. tRNAs with YAY are more hydrophobic and tend to carry hydrophobic amino acids. Besides, the hydrophobicity of guanine and cytosine is in-between, so the tRNAs with YCY or YGY anti-codons tend to carry neutral or biphasic amino acids. Because the anti-codon is also a three-letter code, the central position of this three-letter code seems to affect the hydrophobicity of tRNAs most importantly. That is why the second letter is more important to decide the hydrophobicity of tRNAs as well as their carried amino acids. The first letter of anti-codon is less important than the second letter of anti-codon to affect the hydrophobicity of tRNAs with amino acids. The third letter of anti-codon is least important since it is a wobble nucleotide. There is another interesting observation. AUG is the initiation codon used. The first two letter is A and U which include both hydrophilic and hydrophobic. The stop codon UGA is most frequently used in human genome. It is a permutation of the initiation codon AUG. In addition, UAA is the most frequently and evolutionally used codon in high level expression genes and house-keeping genes and AT rich genes. UAA is the most frequently used stop codon in bacteria and mammalian mitochondria. The first two letters are UA which also include hydrophobic and hydrophilic. But, the sequence is just the opposite of the first two letters, AU, of the initiation codon. The efficiency of the three termination codon is UAA>UAG>UGA. UAA is the stop codon for low GC content genes and UAG as well as UGA are stop codons for high GC content genes[4]. There is another interesting observation. Tyrosine genetic codon has first two letters reading UA, and tyrosine is neutral semi-polar amino acid. It has both effects of hydrophobic and hydrophilic from U and A. The first two letters of isoleucine is AU. However, isoleucine is a hydrophobic amino acid. This time, the second position of nucleotide plays a key role in determining the hydrophobicity of the amino acid. The initiation codon is AUG for methionine. Methionine tends to be hydrophobic. It also suggests that second position U is more important than first position A[5,6].

We can apply the above rules. The codons can be divided into two groups: hydrophilic XAX codons and hydrophobic XUX codons. The central letter is purine A or pyrimidine U can decide the hydrophobicity. And, the neutral or biphasic codons are XCX or XGX codons. The hydrophobic amino acids include leucine (UU/CU), isoleucine (AU), methionine (AU), phenylalanine (UU), and valine (GU). The hydrophilic amino acids include asparagine (AA), aspartic acid (GA), glutamic acid (GA), glutamine (CA), lysine (AA), arginine (AG/CG), and histidine (CA). The neutral semi-polar amino acids include serine (UC), proline (CC), threonine (AC), alanine (GC), glycine (GG), tryptophan (UG), and cysteine (UG). We see the only exception of

the above rules is arginine (AG/CG). And, the first two letters in the hydrophilic amino acid group are duplicated. Here, the third letter is important to decide the specific amino acid in the hydrophilic group. It is worth to note that the third position can be either purine (A/G) or pyrimidine (C/U). The third position is C/U pyrimidine for aspartic acid and is A/G purine for glutamic acid. The third position is C/U pyrimidine for asparagine and is A/G purine for lysine. The third position is C/U pyrimidine for histidine and is A/G purine for glutamine. This principle can aid to distinguish acid and basic amino acids in the hydrophilic group. The third position A/G suggests that this hydrophilic amino acid has its H or OH group longer distance from the central carbon atom compared to that of hydrophilic amino acid with third letter C/U. The codes GA, CA, and AA are all used for hydrophilic amino acids with second letter A. Thus, AG is used for the last hydrophilic amino acid: arginine with its first position A to promote its hydrophilic characteristics. This explanation help to comply with the above principle[7]. Last observation is about the aromatic rings. Amino acids with aromatic rings include phenylalanine (UU), tyrosine (UA), and tryptophan (UG). They all have the first letter U in their codon. These aromatic ring related amino acids tend to be hydrophobic. However, aromatic rings are more complicated to be synthesized in evolution. No more second position U is provided for these three aromatic amino acids. Thus, the first letter U is used for these three aromatic amino acids. In their third position, there can be either purine (A/G) or pyrimidine (C/U). Compared to amino acids with aliphatic side chains or with hydroxylic (OH) groups that have four choices (A/G/C/U) in the third position in their codons, amino acids with aromatic rings tend to evolve later and is with less choices[2,8-10].

References

1. Jukes, T.H. Genetic code 1990. Outlook. *Experientia* **1990**, *46*, 1149-1157, doi:10.1007/bf01936925.
2. Dien, V.T.; Morris, S.E.; Karadeema, R.J.; Romesberg, F.E. Expansion of the genetic code via expansion of the genetic alphabet. *Current opinion in chemical biology* **2018**, *46*, 196-202, doi:10.1016/j.cbpa.2018.08.009.
3. Saier, M.H., Jr. Understanding the Genetic Code. *Journal of bacteriology* **2019**, *201*, doi:10.1128/jb.00091-19.

4. Belin, D.; Puigbò, P. Why Is the UAG (Amber) Stop Codon Almost Absent in Highly Expressed Bacterial Genes? *Life (Basel, Switzerland)* **2022**, *12*, doi:10.3390/life12030431.
5. Štambuk, N.; Konjevoda, P. Determining amino acid scores of the genetic code table: Complementarity, structure, function and evolution. *Bio Systems* **2020**, *187*, 104026, doi:10.1016/j.biosystems.2019.104026.
6. Taylor, F.J.; Coates, D. The code within the codons. *Bio Systems* **1989**, *22*, 177-187, doi:10.1016/0303-2647(89)90059-2.
7. Rapino, F.; Zhou, Z.; Roncero Sanchez, A.M.; Joiret, M.; Seca, C.; El Hachem, N.; Valenti, G.; Latini, S.; Shostak, K.; Geris, L.; et al. Wobble tRNA modification and hydrophilic amino acid patterns dictate protein fate. *Nature communications* **2021**, *12*, 2170, doi:10.1038/s41467-021-22254-5.
8. Barik, S. The Uniqueness of Tryptophan in Biology: Properties, Metabolism, Interactions and Localization in Proteins. *International journal of molecular sciences* **2020**, *21*, doi:10.3390/ijms21228776.
9. Davis, B.K. Evolution of the genetic code. *Progress in biophysics and molecular biology* **1999**, *72*, 157-243, doi:10.1016/s0079-6107(99)00006-1.
10. Rusch, S.L.; Kendall, D.A. Signal sequences containing multiple aromatic residues. *Journal of molecular biology* **1992**, *224*, 77-85, doi:10.1016/0022-2836(92)90577-7.