

***Further Investigation into the Biosynthetic Pathways of the 20 Standard Amino Acids of the Genetic Code**

¹Onyilagha, J. C., ¹Trice, K., and ²Freeland, S.

¹ Department of Biology, University of Arkansas at Pine Bluff, 1200 North University Dr., Pine Bluff, AR 71601. E-mail: onyilaghaj@uapb.edu.

² Interdisciplinary Studies Program, University of Maryland Baltimore County, MD 21250.

Introduction: The origin of genetic coding is arguably the single most important event in the advent of life on Earth. The evolution of genetic coding connected the information storage and transmission potential of polymerized nucleotide sequences to the structural and catalytic properties of polymerized amino acid sequences. However, the events surrounding the emergence of the genetic code are full of mystery. From a pool of available molecules, life ended up using four nucleotides and twenty amino acids to encode and build its proteins. By the time of the Last Universal Common Ancestor (LUCA), the process of protein translation was largely fixed in the form of the standard genetic code. From this biochemical foundation, life has evolved such diversity that 21st century astrobiology is challenged to “*elucidate the biochemical capabilities that define the limits for cellular life*”.

In this present research effort, we set about to investigate whether a contemporary view of metabolic diversity supports the assertion that pathways of amino acid biosynthesis contain molecular fossils that connect “early” and “late” amino acids. The research is grounded on the proposition that genetic coding began with fewer than 20 amino acids. This “early” alphabet (comprising prebiotically plausible amino acids) was then augmented as metabolism evolved new possibilities, and incorporated them into genetic coding. Hence, the standard amino acid alphabet comprises a mixture of “early” versus “late” members: that is, some amino acids were available prebiotically and were therefore present from the start of genetic coding; others evolved later, as “inventions” of early metabolism. We also investigated whether metabolic pathways found in living organisms are indeed an accurate guide to ancient evolutionary events. The project goal was to provide additional insight into the emergence of a standard alphabet of 20 genetically encoded amino acids. Contemporary data from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database, and the MetaCyc database were used. Both are highly curated collections of metabolic pathway and enzyme data representing the full biodiversity of life on earth. Pathways to amino acid biosynthesis were analyzed. The enzymes catalyzing the pathways were sequenced and studied. Specifically, we studied the connection between the pathways

and enzymes of extant metabolism and the evolution of the standard amino acid alphabet.

Our findings are consistent with earlier reports that the 20 amino acids of the standard genetic code comprise of two different groups: “early” amino acids that were likely available at the origin for life through prebiotic syntheses, and “late” amino acids that are best understood as inventions of biology itself. However, the results question the underlying ideas of some of the theories surrounding the evolution or emergence of the 20 encoded amino acids of the standard genetic code, especially the “precursor-product” assertion of the “Co-Evolution theory”. Firstly, the theory asserted that each new amino acid (product) is synthesized from a pre-existing precursor. For example glutamine is synthesized from glutamic acid; tryptophan and cysteine are synthesized from serine. Secondly, that this relationship is reflected in the assignment of amino acids to codons within the genetic code. These assertions are inconsistent with our findings. For example, based on this theory, there is no precursor-product relationship that connects Glycine (Gly) and Threonine (Thr) amino acids. In other words, Thr cannot be synthesized from Gly and *vice-versa*. Nevertheless, our results show new pathways that facilitate the biosyntheses of Thr from Gly and/or *vice-versa*. Going by the co-evolution theory, the Gly→Thr and Thr→Gly biosyntheses are not possible. Consequently, we conclude that the whole idea about the origin of amino acids of the standard genetic code is far from being resolved and that there is need for a critical evaluation of the theories surrounding the emergence of the 20 standard amino acids.

* This research was sponsored by the NASA/MIRS program.