Amino acid evolution: an alternative hypothesis

Peter Andras[a], Alina Andras[b], Csaba D. Andras[c,d]
[a]School of Computing Science,
University of Newcastle, Newcastle upon Tyne, NE1 7RU, UK
[b]Institute for Ageing and Health,
University of Newcastle, Newcastle upon Tyne, NE4 6BE, UK
[c]Department of Chemical Engineering,
Budapest University of Technology and Economics, Budapest, Hungary
[d]Department of Technology and Natural Sciences,
Sapientia – Hungarian University of Transylvania, Miercurea Ciuc, Romania

The understanding of the evolution of amino acid usage in proteins of living organisms is a fundamental issue in the context of theories about the origins of life. The commonly accepted view is that life started with a few amino acids and newer ones were added to the amino acid library of organisms during evolution. Here we propose an alternative hypothesis, suggesting that life might have started with an initial expansion of amino acids followed by gradual decrease of the number of amino acids used in proteins.

It has been recently reported that the replacement frequencies of amino acids are asymmetrical [1], e.g., during the independent evolution of human and chimpanzee genomes the replacement of leucin codons by phenylalanine codons happened almost twice as many times than the reverse replacement. This phenomenon can be interpreted according to the hypothesis that early amino acids (i.e., amino acids produced in origins-of-life experiments [2] or found in meteorites [3]) are replaced by newer amino acids, which emerged as living systems evolved [4]. The data presented by Jordan et al. [1] show that indeed the amino acids which loose in terms of frequency are those which can be considered early amino acids, and those which gain are the other amino acids.

We propose an alternative hypothesis for the interpretation of asymmetric amino acid replacement frequencies. We hypothesise that in the early stages of life there were many amino acids composing proteins of early organisms and during evolution, due to optimisation pressures, the set of amino acids used in proteins was reduced to the currently known 22 amino acids [1,5], which participate in protein formation in extant organisms. In our view the asymmetrical replacement frequencies reflect the effects of this optimisation process.

Plants are able to synthesize more than 200 amino acids that are not directly incorporated into proteins during the translation of mRNA molecules. The existence of natural tRNA molecules with anticodons of length 4 and possibly more (e.g., stop-supressor tRNAs, +1/-1 frameshift tRNAs) indicate that the coding potential of tRNA-s is much larger than the number of genetically encoded amino acids. The discovery in the 70s of the rare protein forming amino acid seleno-cysteine [5], and the recent discovery of the genetically encoded pyrrolysine in archaea [5] show that it is possible that some non-early amino acids might be on the track of disappearance from the list of genetically encoded amino acids.

The cornerstone of our hypothesis is the assumption that simplification of the amino acid lexicon may lead to more efficient living systems incorporating more complex and more functional proteins. This assumption is based on the finding that codon usage is more restricted in more complex organisms [6]. Our assumption is also supported by recent findings that the number of genes of complex organisms is much smaller than their expected number, and that the complex behaviour of these

organisms is achieved by complex regulatory combination of a relatively small number of genomic components [7,8] (i.e., genes and regulatory segments of the DNA).

According to our hypothesis life started with an initial expansion of amino acids and early organisms built their proteins using a wide range of amino acids encoded by codons of length k, with $k \geq 3$. During evolution the optimisation pressures selected organisms using shorter codons and fewer amino acids resulting in the currently dominant set of 20 amino acids and length 3 codons. Remnants of early living systems are the rarely used genetically encoded amino acids (seleno-cysteine and pyrrolysine), tRNA molecules with anticodons of length 4 or more, and viruses that produce frameshift and stop-supressor tRNAs [9]. Our hypothesis also provides an explanation for the observed code simplification of mitochondria, implying that these highly specialised organelles evolved further the simplification of the tRNA lexicon resulting in their reduced set of tRNAs and increased number of unused codons [9].

Our interpretation suggests that asymmetric replacement frequencies do not indicate the replacement of early amino acids by newer ones, but instead indicate the optimisation of the genetic code and the set of encoded amino acids towards a functionally more complex and more efficient form of life based on fewer amino acids and simpler, less ambiguous encoding of these amino acids.

1. Jordan, I.K., Kondrashov, F.A., Adzhubei, I.A., Wolf, Y.I., Koonin, E.V., Kondrashov, A.S. & Sunayev, S. *Nature*, **433**, 633-638 (2005).
2. Miller, S.L. & Orgel, L.E. *The Origins of Life on the Earth*. Englewood Cliffs, NJ, Prentice Hall (1974).
3. Pizzarello, S. *Origins of Life and Evolution of the Biosphere*, **34**, 25-34 (2004).
4. Brooks, D.J., Fresco, R.F., Lesk, A.M., & Singh, M. *Molecular Biology and Evolution*, **19**, 1645-1655 (2002).
5. Cobucci-Ponzano, B., Rossi, M., & Moracci, M. *Molecular Microbiology*, **55**, 339-348 (2005).
6. Duret, L. & Mouchiroud, D. *Proceedings of the National Academy of Sciences*, **96**, 4482-4487 (1999).
7. Andras, P. & Andras, C.D. *Medical Hypotheses*, **64**, 678-688 (2005).
8. Mattick, J.S. & Gagen, M.J. *Molecular Biology and Evolution*, **18**, 1611-1630 (2001).
9. Söll, D. & RajBhandary U.L. (eds.) *tRNA Structure, Biosynthesis and Function*. (ASM Press, Washington, DC, 1995).