**AMINO ACID FREQUENCY**

**Introduction**

Genetic information contained in [mRNA](http://www.tiem.utk.edu/~gross/bioed/webmodules/glossary.htm#mRNA) is in the form of [codons](http://www.tiem.utk.edu/~gross/bioed/webmodules/glossary.htm#codons), sequences of three [nucleotides](http://www.tiem.utk.edu/~gross/bioed/webmodules/glossary.htm#nucleotide), which are translated into [amino acids](http://www.tiem.utk.edu/~gross/bioed/webmodules/glossary.htm%20#aminoacids) which then combine to form proteins. At certain sites in a [protein's structure](http://www.tiem.utk.edu/~gross/bioed/webmodules/glossary.htm#proteinstructure), amino acid composition is not critical. Yet certain amino acids occur at such sites up to six times more often than other amino acids. In the 1960's, molecular biologists sought to determine if amino acid composition was a reflection of the [genetic code](http://www.tiem.utk.edu/~gross/bioed/webmodules/glossary.htm#geneticcode) or if certain amino acids were naturally selected as optimal.

**Importance**

The amino acid composition at non-critical sites in protein structure may or may not contribute to the overall quality of a protein. If certain amino acids are optimal for protein structure, [natural selection](http://www.tiem.utk.edu/~gross/bioed/webmodules/glossary.htm#naturalselection) should have acted over evolutionary time to increase the frequency of these amino acids. An alternate hypothesis is that amino acid composition at a site is merely a random permutation of the genetic code. We can use probabilities to answer questions about amino acid evolution.

**Question**

Are frequencies of particular amino acids simply a consequence of random [permutations](http://www.tiem.utk.edu/~gross/bioed/webmodules/glossary.htm#permutation) of the genetic code or instead a product of natural selection?

**Method**

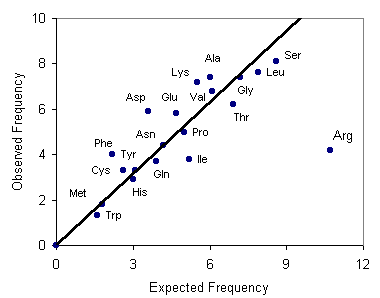
If a particular amino acid is in some way adaptive, then it should occur more frequently than expected by chance. This can easily be tested by calculating the [expected frequencies](http://www.tiem.utk.edu/~gross/bioed/webmodules/glossary.htm#expectedfrequency) of amino acids and comparing to [observed](http://www.tiem.utk.edu/~gross/bioed/webmodules/glossary.htm#observedfrequency). The codons and observed frequencies of particular amino acids are given in the table.

|  |  |  |
| --- | --- | --- |
| Amino Acids | Codons | Observed Frequency in Vertebrates |
| Alanine | GCU, GCA, GCC, GCG | 7.4 % |
| Arginine | CGU, CGA, CGC,CGG, AGA, AGG | 4.2 % |
| Asparagine | AAU, AAC | 4.4 % |
| Aspartic Acid | GAU, GAC | 5.9 % |
| Cysteine | UGU, UGC | 3.3 % |
| Glutamic Acid | GAA, GAG | 5.8 % |
| Glutamine | CAA, CAG | 3.7 % |
| Glycine | GGU, GGA, GGC, GGG | 7.4 % |
| Histidine | CAU, CAC | 2.9 % |
| Isoleucine | AUU, AUA, AUC | 3.8 % |
| Leucine | CUU, CUA, CUC, CUG, UUA, UUG | 7.6 % |
| Lysine | AAA, AAG | 7.2 % |
| Methionine | AUG | 1.8 % |
| Phenylalanine | UUU, UUC | 4.0 % |
| Proline | CCU, CCA, CCC, CCG | 5.0 % |
| Serine | UCU, UCA, UCC, UCG, AGU, AGC | 8.1 % |
| Threonine | ACU, ACA, ACC, ACG | 6.2 % |
| Tryptophan | UGG | 1.3 % |
| Tyrosine | UAU, UAC | 3.3 % |
| Valine | GUU, GUA, GUC, GUG | 6.8 % |
| Stop Codons | UAA, UAG, UGA | --- |

The frequencies of DNA bases in nature are 22.0% uracil, 30.3% adenine, 21.7% cytosine, and 26.1% guanine. The expected frequency of a particular codon can then be calculated by multiplying the frequencies of each DNA base comprising the codon. The expected frequency of the amino acid can then be calculated by adding the frequencies of each codon that codes for that amino acid.

As an example, the RNA codons for tyrosine are UAU and UAC, so the random expectation for its frequency is (0.220)(0.303)(0.220) + (0.220)(0.303)(0.217) = 0.0292. Since 3 of the 64 codons are nonsense or stop codons, this frequency for each amino acid is multiplied by a correction factor of 1.057.

By plotting the expected frequency against the observed frequency, we can see if some amino acids are occurring more or less often than expected by chance. If the observed and expected frequencies are close to equal, we would expect a line with a [slope](http://www.tiem.utk.edu/~gross/bioed/webmodules/glossary.htm#slopeofaline) = 1.



**Interpretation**

Excluding the outlier, arginine, the correlation between observed and expected frequencies was highly significant (r = 0.89). Only arginine (Arg) frequency seems to be a product of natural selection acting on one or more of the arginine codons.

**Conclusion**

Average amino acid composition passively reflects random permutations of the genetic code. In other words, instead of particular amino acids being optimal, the most important thing determining the frequency of amino acids in a protein is simply the number of codons coding for it.

**Additional Questions**

1. Calculate the expected frequencies for some of the amino acids.

2. Look at the outlier arginine. Does it occur more or less often than expected? Can you think of some hypotheses for why this might be?

**Sources**

Dyer, K. F. 1971. The quiet revolution: A new synthesis of biological knowledge. Journal of Biological Education **5**:15-24 King, J. L. and T. H. Jukes. 1969. Non-Darwinian evolution. Science **164**:788-798

**About this Resource**

This material was originally distributed as part of the “Alternative Routes to Quantitative Literacy for the Life Sciences[[1]](#endnote-1)” - National Science Foundation Award DUE-9752339 to the University of Tennessee, Knoxville for August 1, 1998 - July 31, 2000. Principal Investigator: Louis J. Gross. Co-Principal Investigators: Beth C. Mullin and Susan E. Riechert.

This material is now being revised as part of the “Resources for Improving Quantitative Skills in Community College Biology[[2]](#endnote-2)” project. As part of that project is also aligned with the OpenStax Biology Textbook[[3]](#endnote-3).

It is published using the QUBES Open Education Resources publishing platform[[4]](#endnote-4).

1. http://www.tiem.utk.edu/~gross/bioed/ [↑](#endnote-ref-1)
2. https://qubeshub.org/community/groups/quantbioatcc/ [↑](#endnote-ref-2)
3. https://openstax.org/details/books/biology-2e [↑](#endnote-ref-3)
4. https://qubeshub.org/qubesresources/publications/### [↑](#endnote-ref-4)