

Mapping Polypeptide Sequences as Protein Pipes in a 3D Sequence Space Based on a Graph Theory Representation of the Genetic Code

Carl Slater,
Mathematics Department, Morehouse College,

Fissehe Abebe,
Mathematics Department, Clark Atlanta University,
223 James Brawley Dr., S.W., Atlanta, GA 30134.

William Seffens,
Biology Department, Clark Atlanta University,
223 James Brawley Dr., S.W., Atlanta, GA 30134.
Corresponding author: Wseffens@cau.edu

May 18, 2007

We examine the bioinformatic significance of a graph theory representation of the genetic code that partitions the 20 amino acids into three families. The genetic code can be represented as a graph, with vertices labeled by the amino acids and lines representing complementary DNA bases arranged as three-letter codons paired with reverse-complement codons. The genetic graph is constructed by pairing codons with their reverse complement for all 20 amino acids. The codons, in turn, are grouped together according to the respective amino acids for which they code. Interestingly, a three-component graph is formed for the standard genetic code. Protein sequences were mapped into a 3D sequence space based on composition in the three families. Resulting images are described as 3D protein pipes or as a lattice

path, with the path exhibiting helices. The net helix rotation for a group of 634 proteins shows a bias that disappears when sequences are randomized. Helix rotation is also different when randomized partitions not based on the standard genetic code are used. Length of the helices is also longer for real peptides compared to randomized sequences. Together this suggests that the graph theory structure of the genetic code is biologically significant.