Fitness Function Analysis of Biological Genetic Codes using an Evolutionary Algorithm

William Seffens

Department of Biological Sciences
Center for Theoretical Study of Physical Systems
Clark Atlanta University
223 James P. Brawley Dr., SW, Atlanta, GA 30134

David Digby

Department of Biological Sciences
Clark Atlanta University
223 James P. Brawley Dr., SW, Atlanta, GA 30134

Abstract

An evolutionary algorithm model of early cells (protocells) has been studied to investigate the evolution of the universal and the several alternate genetic codes in biology. Up to one thousand protocells are initialized with randomly selected genetic codes, which are then mutated over many generations, with a continuous requirement to replicate a fixed genome consisting of one hundred genes. Representative real enzymes from biological protein databases were used for this gene set. Fidelity of replication was measured by reference to the real protein sequences using a standard table of amino acid similarities. An additional selection demands minimization of a cost function roughly proportional to the number of tRNA genes required to implement the protocell's genetic code. Various fitness functions can be evaluated in this evolutionary algorithm treatment of biological evolution of the genetic code. A function based on mRNA folding was evaluated to understand a biological bias observed in real gene sequences. Measured genetic drift was found to be dependent upon this mRNA folding fitness function for very small population sizes. The mRNA folding function is a symmetry operator acting upon the gene set by the genetic code. These results may have applicability to general problem solving in evolutionary computation.

1 RESULTS

Various fitness functions can be evaluated for biological relevance in an evolutionary algorithm treatment of evolution of the genetic codes (Digby and Seffens, 1999). Each genetic code is defined as a 64x20 matrix relating each one of the 64 codons to each one of the 20 amino acids. Every organism is evaluated for fidelity of replicating these real enzymes from its genome, using its own genetic code (with weighted random choices in case of degenerate assignments). The fidelity values are summed to yield a total fitness parameter for each of the organisms, and these scores are sorted to establish a fitness ranking of the organisms. Organisms with the lowest 50% of fitness values are "killed", and the organisms with the highest 50% fitness values are further evaluated. An additional test was devised to evaluate a bias found in a recent mRNA folding study (Seffens and Digby, 1999). The observed bias in mRNA folding was hypothesized to be due to the selection forces that drove the evolution of genetic codes.

To assess the importance of mRNA folding in the evolution of genetic codes, a perfect solution was presented to a very small population of protocells. Data shows the loss of fitness for very small populations over runs of 100 generations initialized with the biological genetic code. The populations range from 8 to 48 protocells. In all cases the long-stem mRNA selection losses the standard genetic code to a greater degree through genetic drift compared to short-stem selection. Populations greater than 100 organisms maintained the perfect solution for both long- and short-stem selection. These results support the hypothesis that there is a relationship between mRNA folding and the structure of the genetic code in biology. Current work is utilizing graph theory measures of the genetic code to track the large data produced by 1000 organisms over up to 50,000 generations.

Acknowledgments

This work was supported (or partially supported) by NIH grant GM08247 and NSF CREST Center for Theoretical Studies of Physical Systems (CTSPS) Cooperative Agreement #HRD-9632844.

References

D. Digby and W. Seffens (1999). *Evolutionary Algorithm Analysis of the Biological Genetic Codes*, Proceedings of 1999 Genetic and Evolutionary Computation Conference, Orlando, FL, p 1440.

W. Seffens and D.W. Digby (1999*). mRNAs Have Greater Calculated Folding Free Energies Than Shuffled Or Codon Choice Randomized Sequences*. Nuc. Acids Res.**27:**1578-84.