
Finding synonymous coding DNA sequences with maximum base pairing

Claudio Arbib¹, Andrea Manno², Fabrizio Rossi³, and Andrea D'ascenzo^{*†2}

¹Università degli Studi dell'Aquila [L'Aquila] (UNIVAQ.IT) – Via Giovanni Di Vincenzo 16/B, 67100 L'Aquila, Italy

²Università degli Studi dell'Aquila – Italy

³Department of Information Engineering, Computer Science and Mathematics University of L'Aquila – Università degli Studi dell'Aquila, Via Vetoio, I-67100 L'Aquila, Italy

Abstract

A Coding DNA Sequence (CDS) is a sequence of bases that, organized into triplets (codons), encodes for a protein. As different codons can encode the same amino acid, the same protein is encoded by a very large number of CDSs and a problem arises of finding optimal CDSs with respect to given indicators. Here the Base Pairing Number is considered, the largest number of hydrogen bonds the bases can form in a feasible folding. We developed an implicit enumeration algorithm based on dynamic programming to maximize base pairing, and tested it on random and human genome CDSs. The method performs an effective reduction of the search space and turns out to be more efficient than integer programming to solve the problem.

^{*}Speaker

[†]Corresponding author: andrea.dascenzo@graduate.univaq.it