

Analyses from coding regions to protein structures: second codon position preferences and the organisation of the genetic code

M.L. Chiusano^{1,2}, F. Alvarez-Valin¹, M. Di Giulio¹, G. D'Onofrio¹, G. Ammirato², G. Colonna² and G. Bernardi¹

¹ Laboratorio di Evoluzione Molecolare, Stazione Zoologica Anton Dohrn, Villa Comunale, I-80121 Naples, Italy.

² Centro di Ricerca Interdipartimentale di Scienze Computazionali e Biotecnologiche, Seconda Università, via Costantinopoli 16, 80138 Naples, Italy.

In order to investigate on the relationships between the coding regions and the encoded proteins, we took advantage of a program (1) written in the C language, which aligns sequence and three-dimensional structure information. In particular, this software was applied to the analysis of protein secondary structures and the nucleotide composition in the corresponding mRNA regions. A comparative analysis of nucleotide frequencies in the three codon positions of coding sequence regions, corresponding to both predicted or experimentally determined secondary structures, revealed differences in the three codon positions, which become strongly evident for codon second positions. Indeed, nucleotide frequencies in the second codon position of coding regions and the secondary structures in the encoded proteins are different for helix, β -strand and aperiodic structures. These differences are related to the physico-chemical properties of protein secondary structures and suggest that the ge-

netic code had been organized to preserve secondary structures of proteins preventing deleterious amino acid substitutions that could modify the average physico-chemical requirements for an optimal structure (2).

1) M.L. Chiusano, G. Colonna (2000). From DNA to Protein: comparative analyses to investigate structural relationships. http://obelix.bio.uniroma2.it/www/abstr_2000.html

2) M.L. Chiusano, F. Alvarez-Valin, M. Di Giulio, G. D'Onofrio, G. Ammirato, G. Colonna and G. Bernardi (2000). Second codon position of genes and the secondary structures of proteins: relationships and implications for the origin of the genetic code. *Gene* 261, 63-69.