July 16-21, 2017 at UC San Diego, CA, USA

Information Analysis is used to Determine the Identity Elements of the Operational tRNA Code

Gabriel S. Zamudio[1] and Marco V. José*[1]

[1]Theoretical Biology Group, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, Ciudad de México, C.O. 04510, Mexico

* e-mail: marcojose@biomedicas.unam.mx

Introduction: The correct implementation of the genetic code involves a complex machinery of biological pathways. At the core of such a process lies the transfer RNA (tRNA), a key molecule responsible of the translation process. To sustain such convoluted process, the tRNA holds two genetic codes. The anticodon code designed to read the messenger RNA, and a second genetic code known as the operational code [1], which is commonly mapped [2] to the acceptor stem of the tRNA. This second genetic code directs the correct identification of each tRNA by its cognate aminoacyl-tRNA synthetase (aaRS) by stereochemical means. The problem of deciphering this recognition code and identifying the sites along the tRNA structure involved is known as "the identity problem". In this work, the information theory is used to analyze $\sim 53{,}000$ sequences of tRNA genes of the 20 amino acids to determine the specific sites of the tRNA structure participating in the positive recognition of their respective aaRSs. The variation of information $VI(X,Y)$, is a measure of information distance and it is given by: $VI(X,Y) = H(X) + H(Y) - 2I(X,Y)$, where $H(X)$ is the Shannon's entropy and $I(X,Y)$ is mutual information. A list of sets of sites is provided for each tRNA isoacceptor. According to this measure, the sites which are nearby to the anticodon, bridge the information of the anticodon to the rest of the structure, thereby these sites are the ones involved in the second genetic code. The list provides for each tRNA a set of sites that were modified with complete synchronicity among all samples, i.e., these sets comprise the identity elements for each tRNA associated to each amino acid. Previously undetected identity elements are reported.

[1] C. De Duve (1988). The second genetic code *Nature*, 333: 117–118. [2] Y.-M. Hou, P. Schimmel (1988). A simple structural feature is a major determinant of the identity of a transfer RNA *Nature*, 333: 140–145.