

# Lower Limit to the Size of the Primeval Amino Acid Alphabet

Ariel Fernández

Institute for Biophysical Dynamics, Department of Computer Science, Ryerson Hall,  
The University of Chicago, Chicago, Illinois 60637.

Permanent address: Indiana University School of Informatics, Center for Computational  
Biology and Bioinformatics, 714 N. Senate Ave., Indianapolis, IN 46202. Fax: 31 72 78 92 17.  
E-mail: ariel@uchicago.edu

Z. Naturforsch. **59c**, 151–152 (2004); received June 16/July 17, 2003

Here I systematically examine the information complexity of all primary sequences of natural proteins deposited in the Swiss-Prot database. The sequence complexity is assessed by determining the frequency of occurrence of each amino acid type on sequence windows of fixed length, calculating the Shannon entropy of the window and then averaging over all windows covering the sequence. The minimum value in information content obtained from the present-day record imposes a lower limit in the number of letters that a primeval amino acid alphabet must have had.

*Key words:* Genetic Code, Amino Acid Alphabet, Translation