

# Protein alphabet reduction based on frequency dictionaries

Misha Kudryashev

Institute of Computational Modeling, Akademgorodok, Krasnoyarsk, Russia;

and University of Heidelberg, Germany

With A. Gorban and T. Popova

## ABSTRACT

**Motivation:** What are proteins made from, as the working parts of the living cells protein machines? To answer this question, we need a technology to disassemble proteins onto elementary functional details and to prepare lumped description of such details. This lumped description might have a multiple material realization (in amino acids). Our hypothesis is that informational approach to this problem is possible. We propose a way of hierarchical classification that makes the primary structure of protein maximally non-random. The first steps of the suggested research program are realized: the method and the analysis of optimal informational protein binary alphabet. The general method is used to answer several specific questions, for example:

- Is there a syntactic difference between Globular and Membrane proteins?
- Are proteins random sequences of amino acids (a long discussion)?

For these questions, the answers are as follows:

- There exists significant syntactic difference between Globular and Membrane proteins, and this difference is described;
- Amino acid sequences in proteins are definitely not random.

---

\* To whom correspondence should be addressed.