University of Basel

Department
Biozentrum

SIB Swiss Institute of Bioinformatics

BIOZENTRUM
The Center for Molecular Life Sciences

Basel Computational Biology Seminar: 22830-01 Current Research in Bioinformatics I

Eugene V. Koonin

Computational Biology Branch, Division of Intramural Research
National Library of Medicine, National Institutes of Health
Bethesda, MD, USA

Emergence of globular protein folds from random amino acid sequences

The origin and evolution of protein folds are among the most challenging, long-standing problems in biology. Emphasizing the challenge, evolutionary reconstructions indicate that the main diversity of protein folds evolved very early in the history of life, even before the formation of the fully developed, modern-type translation system. Although many plausible scenarios of early protein evolution leading to fold nucleation have been proposed, realistic simulation of this process was not feasible because of the lack of efficient approaches for protein structure prediction, a situation that changed with the advent of powerful tools for fast and robust protein structure prediction, such as AlphaFold and ESMFold. I will present a recently developed computational approach for protein fold evolution simulation (PFES) with atomistic details that provide insights into the mechanisms of evolution of globular folds from random amino acid sequences*. PFES introduces random mutations in a population of protein sequences, evaluates the effect of mutations on protein structure, and selects a new set of proteins for further evolution. Repeating this process iteratively allows tracking the evolutionary trajectory of a changing protein fold that evolves under selective pressure for protein fold stability, interaction with other proteins, or other features shaping the fitness landscape. PFES was employed to show how globular protein folds could evolve from random amino acid sequences as monomers or in complexes with other proteins. The simulations reproduce the evolution of several simple folds of natural proteins as well as the evolution of a variety of distinct folds not known to exist in nature. Evolution of a stable fold from random sequences, on average, takes about one, and in some simulations, as few as 0.2 amino acid replacement per site, which is comparable to empirical data on protein evolution. Thus, the results of these computational experiments suggest that simple but stable protein folds can evolve relatively easily. These findings could shed light on the enigma of the rapid evolution of protein fold diversity at the earliest stages of life evolution. PFES tracks the complete evolutionary history from simulations that describes intermediate states at the sequence and structure levels and can be used to test a broad variety of hypotheses on protein fold evolution.

Date: Monday, September 29, 2025

Time: 16:15 h – 17:30h

Location: Biozentrum, 02.073

Contact: Aleksei Mironov (aleksei.mironov@unibas.ch)

* Sahakyan H, Babajanyan SG, Wolf YI, Koonin EV. In silico evolution of globular protein folds from random sequences. Proc Natl Acad Sci U S A. 2025 Jul 8;122(27):e2509015122.