

Bachelor thesis project

Analysis of Protein Repeats

Evolution of Cryptic Oligo-peptide Repeats

Background: Oligo-peptide repeats are short (20-50 residues) stretches of highly conserved sequence found in multiple copies within a single peptide chain. About 25-35% of all proteins contain oligo-peptide repeats although many of these have gone unnoticed, even in proteins with known structures. We have identified thousands of clusters of proteins that contain these cryptic oligo-peptide repeats.

Objectives: The student will focus on one of these clusters, studying, analyzing, and categorizing one of these repeat clusters. The repeat will be formally defined and its evolutionary relationship to other proteins will be determined.

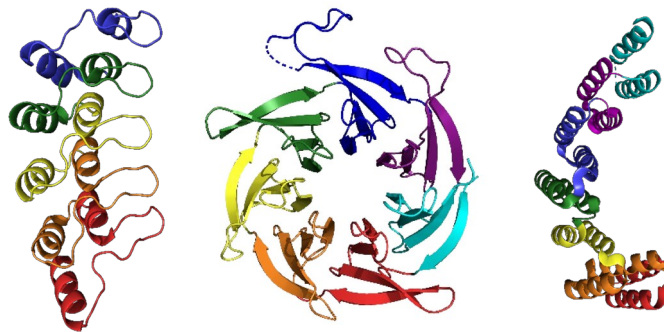
Requirements:

- Adequate completed coursework and interest in biochemistry
- Adequate completed coursework and interest in math and statistics
- Programming skills in one or more languages (*e.g.* R, Python, C++, FORTRAN, etc.)

Methods:

- Creation of multiple sequence alignments
- Statistical analysis of protein sequence features

Figure 1: Examples of oligo-peptide repeat proteins. Each repeat unit is colored individually for easy recognition.



Supervision: Matthew Merski, IEB 22, [email: mmerski@uni-muenster.de](mailto:mmerski@uni-muenster.de)
Molecular Evolution and Bioinformatics group (<http://bornberglab.org/>)

Selected Literature:

- 1) Schuler, A., Bornberg-Bauer, E. "Evolution of Protein Domain Repeats in Metazoa" *Mol. Biol. Evol.* **33**, 3170 (2016)
- 2) Kajava, A.V. "Tandem repeats in proteins: from sequence to structure" *J. Struct. Biol.* **179**, 279 (2012)
- 3) Merski, M. et al. "Self-analysis of repeat proteins reveals evolutionary conserved patterns" *BMC Bioinformatics* **21**, 179 (2020)