# Discovering the general architecture of protein families with OverProt

Midlik Adam1,2, Hutařová Vařeková Ivana1,2,3, Hutař Jan1,2, Chareshneu Aliaksei1,2, Svobodová Radka1,2,   
Berka Karel4

1 CEITEC - Central European Institute of Technology, Masaryk University, Brno, midlik@mail.muni.cz

2 National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Brno

3 Faculty of Informatics, Masaryk University, Brno

4 Department of Physical Chemistry, Faculty of Science, Palacký University, Olomouc

Every protein family has a set of characteristic secondary structure elements (SSEs) – helices and β-strands. Their number, order, and spatial arrangement is relatively consistent throughout the whole family; thus they define the general architecture of the family and provide a good guide for orientation within the structures. However, there are always some variations within the family, and a single structure is not enough to represent the general architecture of the whole family. Therefore we developed OverProt, which gathers the secondary structure information from all family members and creates the secondary structure consensus. This consensus shows the general architecture of the family as well as its variation, and thus provides a useful insight into the family (just as the sequence logo does for a family of sequences). OverProt server (https://overprot.ncbr.muni.cz/) provides precomputed consensus for all CATH superfamilies plus user-defined computations, visualized by an interactive viewer, which shows the SSE type, length, frequency of occurrence, spatial variability, and β-connectivity. OverProt is also utilized in the visualization tool 2DProts [1], which has been integrated into the CATH database.

[1] Hutařová Vařeková I, Hutař J, Midlik A, Horský V, Hladká E, Svobodová R, Berka K (2021) 2DProts: database of family-wide protein secondary structure diagrams. Bioinformatics, 37, 4599–4601.