# Protein family based 2D Diagrams of Secondary Structure Elements

Hutařová Vařeková I.1,2,3, Hutař J. 1,2, Midlik A. 1,2, Navrátilová V. 4, Svobodová Vařeková R. 1,2, Berka K.4

*1 CEITEC - Central European Institute of Technology, Masaryk University Brno, Kamenice 5, 625 00 Brno-Bohunice, Czech Republic*

*2 National Centre for Biomolecular Research, Faculty of Science, Kamenice 5, 625 00 Brno-Bohunice, Czech Republic*

*3 Faculty of Informatics, Botanická 68a, 602 00 Brno-Bohunice, Czech Republic*

*4 Department of Physical Chemistry, Regional Centre of Advanced Technologies and Materials, Faculty of Science, Palacký University, 17. listopadu 1192/12, 771 46 Olomouc, Czech Republic*

Secondary structure elements (SSEs) such as α-helices and β-sheets are important part of protein structure. Their positions and distances in protein are often characteristic for proteins within whole protein family. Unfortunately, current methods focused on 2D visualization of SSEs (e.g., PROMOTIF [1], Pro-origami [2]) are based on one input protein only. Moreover, they usually do not consider information about real distances of SSEs and therefore SSEs that are close to each other in 3D might be visualized far from each other in 2D. As a result, even when two proteins from the same family differ only slightly in 3D, their 2D SSE diagrams can be totally different.

For this reason, we focused on development of a methodology for 2D SSE diagrams generation, which is based on structures representing whole protein families. In our approach we use three criteria: The first is to minimize the error of SSEs projection from 3D to 2D. Then we concentrate to highlight the similarities of protein families in each protein diagram. Specifically, we use a “skeleton” concept: For each protein family, we find the most firm and stable SSEs. Their position changes minimally in whole family. This “skeleton” get static position in 2D SSE diagrams for proteins from this protein family. The third criterion was to keep the differences between protein 3D structures and transfer them to 2D SSE diagram (classified e.g. by RMSD number). We tested our approach on all protein families described in CATH and this way we showed its applicability.

[1] Hutchinson, E. G., Thornton, J. M. (1996). PROMOTIF—a program to identify and analyze structural motifs in proteins. Protein Science, 5(2), 212-220.

[2] Stivala, A., Wybrow, M., Wirth, A., Whisstock, J. C., & Stuckey, P. J. (2011). Automatic generation of protein structure cartoons with Pro-origami. Bioinformatics, 27(23), 3315-3316.