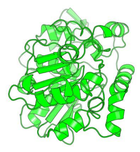
Open-access Mustguseal platform
for bioinformatic analysis in computational enzymology

Dmitry Suplatov[1*], Yana Sharapova[2], Daria Timonina[2], Elena Schmalhausen[1], Kateryna Fesko[5], Nina Popova[3], Vladimir Muronets[1], Vladimir Voevodin[3,4], Vytas Švedas[1,2]

[1] Belozersky Institute of Physicochemical Biology, [2] Faculty of Bioengineering and Bioinformatics, [3] Faculty of Computational Mathematics and Cybernetics, [4] Research Computing Center at Lomonosov Moscow State University, Moscow, Russia; [5] - Institute of Organic Chemistry, Graz University of Technology, Graz, Austria;
[*] d.a.suplatov@belozersky.msu.ru

Comparative analysis of homologous proteins in a functionally diverse superfamily is a valuable tool at studying structure-function relationship, but represents a methodological challenge. We have developed an open-access platform available at https://biokinet.belozersky.msu.ru/mustguseal consisting of free on-line methods to study the structure-function relationship in proteins, to select the most promising hot-spots for implementation of novel functions, improvement of stability and evolvability of useful proteins/enzymes, and to design their selective modulators.

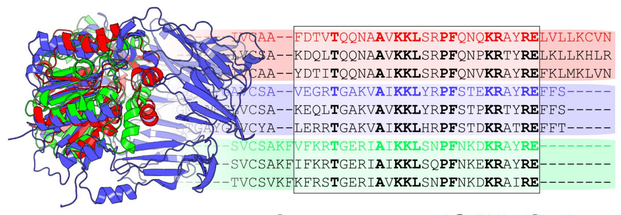
The key concept is to study the structure-function relationship of a particular protein by systematic bioinformatic analysis of the corresponding superfamily



Protein of interest

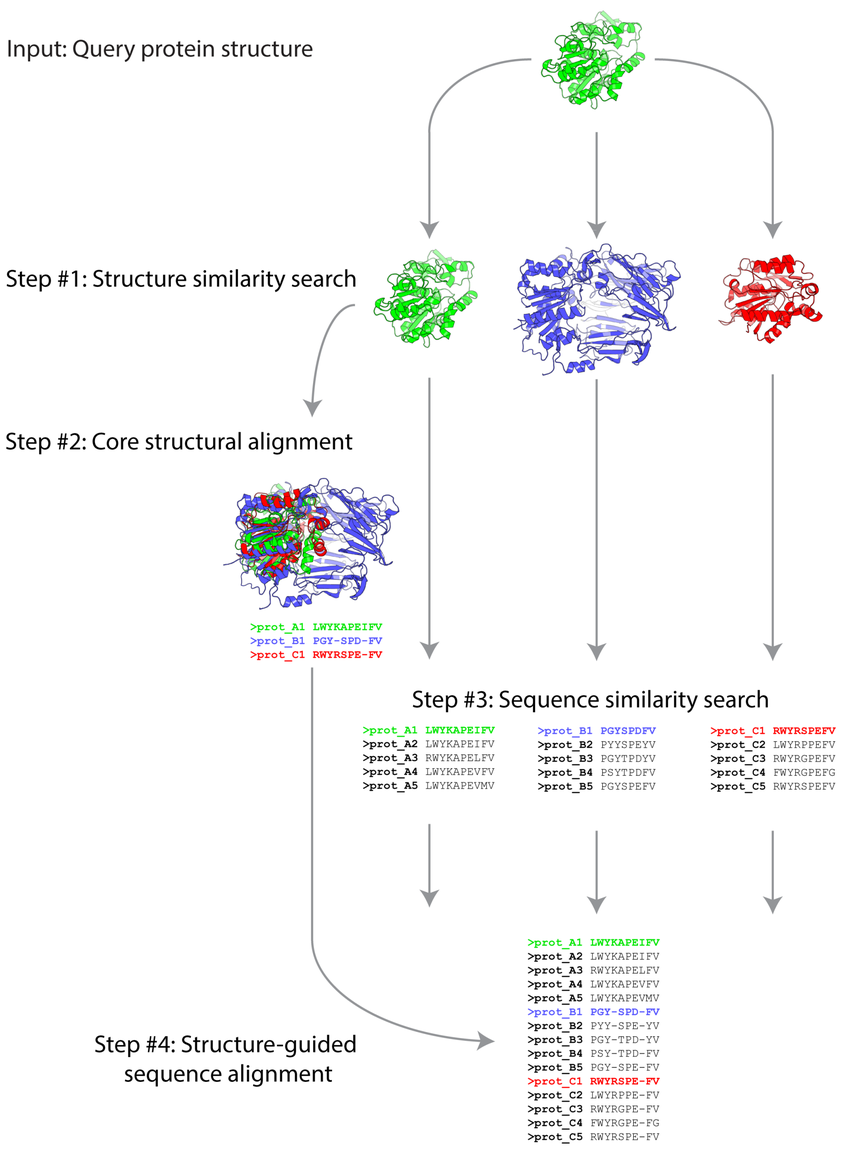
Collection and analysis of all the available sequence and structural data corresponding to the superfamily

Conclusions regarding the sequence/structure-function relationship in the particular protein



Corresponding superfamily

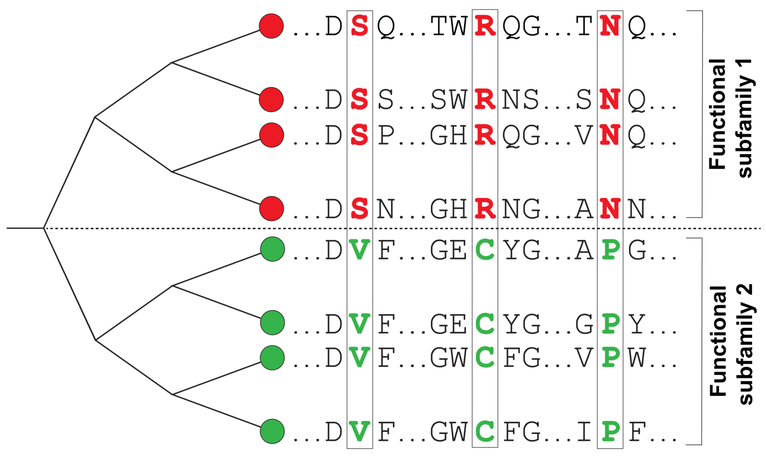
1. Construct multiple alignment of the protein superfamily



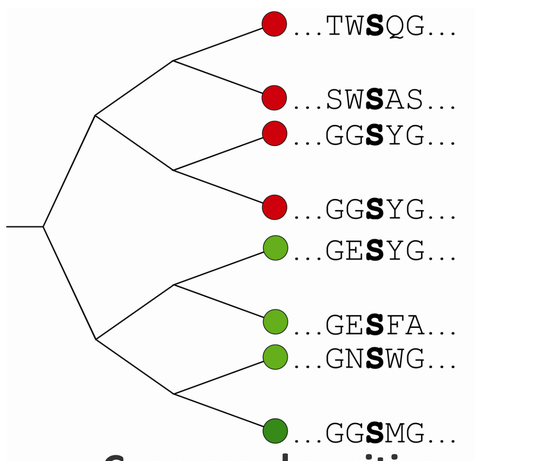
[*Bioinformatics*, 2018; 10.1093/bioinformatics/btx831]

- The key web-server **Mustguseal** (i.e., **Mu**ltiple **St**ructure-**Gu**ided **Se**quence **Al**ignment of Protein Families) can automatically collect from public databases and align thousands of sequences and structures of proteins within a superfamily to produce a large structure-guided sequence alignment;
- Four sister web-methods are available to further study the collected data;

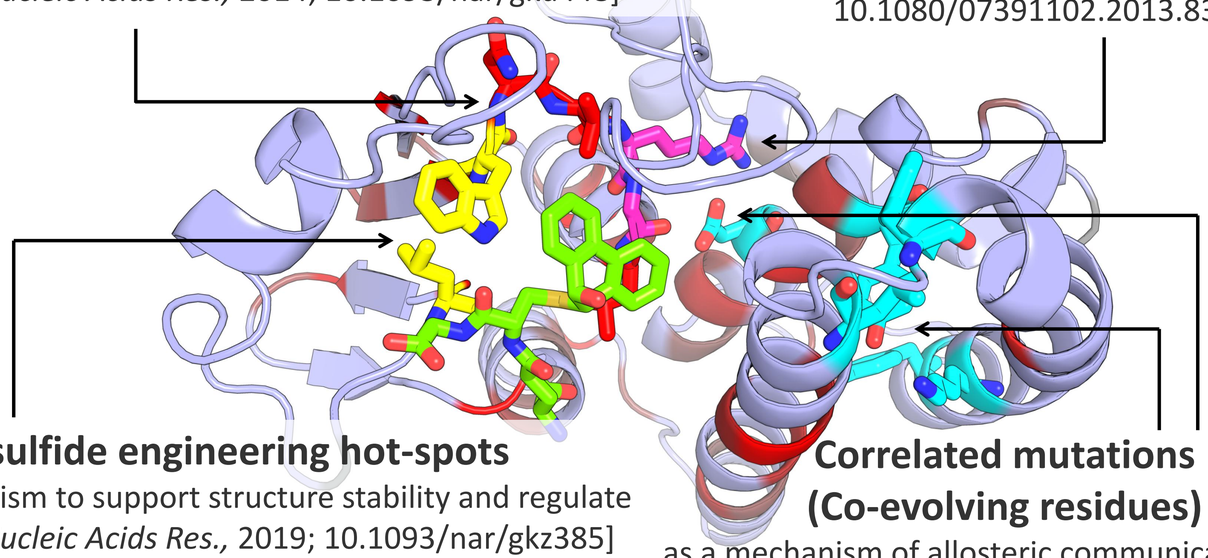
2. Annotate the protein of interest according to the bioinformatic analysis of the superfamily



Subfamily-Specific positions
as determinants of functional diversity and binding specificity [*Nucleic Acids Res.,* 2014; 10.1093/nar/gku448]

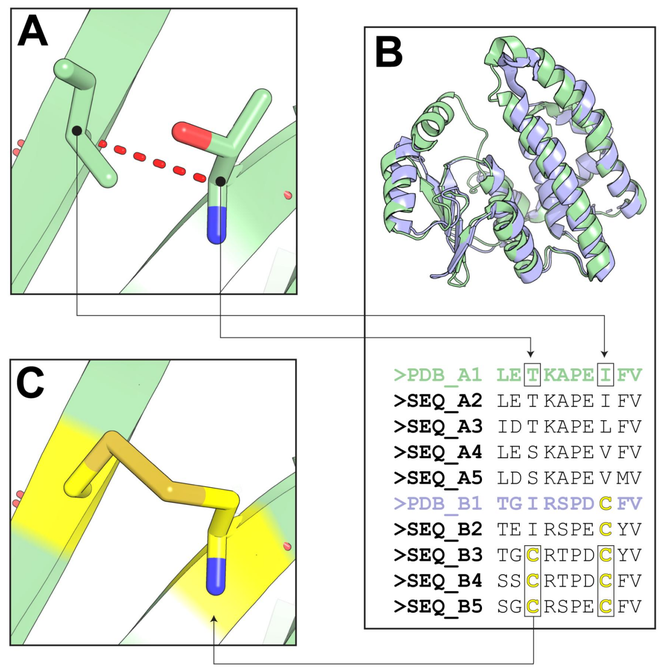


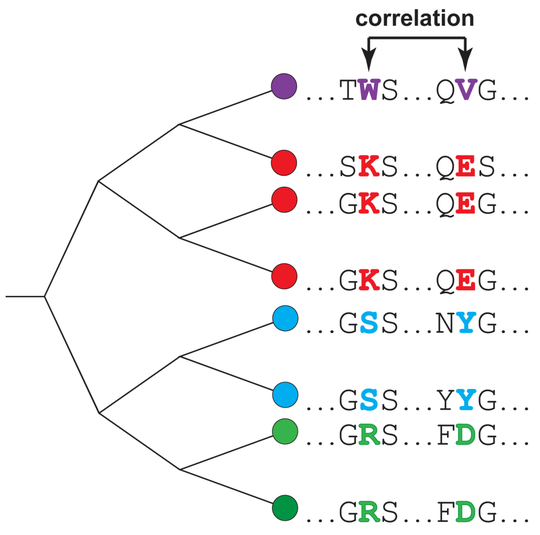
Conserved positions
as determinants of common functional properties [*J.Biomol.Struct.Dyn.*, 2014; 10.1080/07391102.2013.834514]



Disulfide engineering hot-spots
as a mechanism to support structure stability and regulate function [*Nucleic Acids Res.,* 2019; 10.1093/nar/gkz385]

Correlated mutations (Co-evolving residues)
as a mechanism of allosteric communication via a network of interacting residues, and a source of compensatory mutations for rational design [*J.Bioinform.Comput.Biol.,* 2018;10.1142/S021972001840005X]





3. Expert interpretation of the bioinformatic analysis followed by experimental evaluation

We have applied the developed methodology to study structure-functional relationship in various protein superfamilies, design improved enzymes and selective modulators of their activity:

- *FEBS J.*, 2018 (10.1111/febs.14486)
- *FEBS open bio*, 2018 (10.1002/2211-5463.12441)
- *Biochimie*, 2019 (10.1016/j.biochi.2018.12.017)
- *PLoS ONE*, 2014 (10.1371/journal.pone.0100643)
- *J.Biomol.Struct.Dyn.*, 2019 (10.1080/07391102.2018.1475260);
- Patent #RU2661151, 2018
- Patent #RU2564578, 2014
- *Biotechnology J.*, 2015 (10.1002/biot.201400150)
- *Protein Eng.Des.Sel.*, 2012 (10.1093/protein/gzs068)

This work was supported by the RFBR grant #18-29-13060. HPC computing resources at the MSU were supported by the project RFMEFI62117X0011.