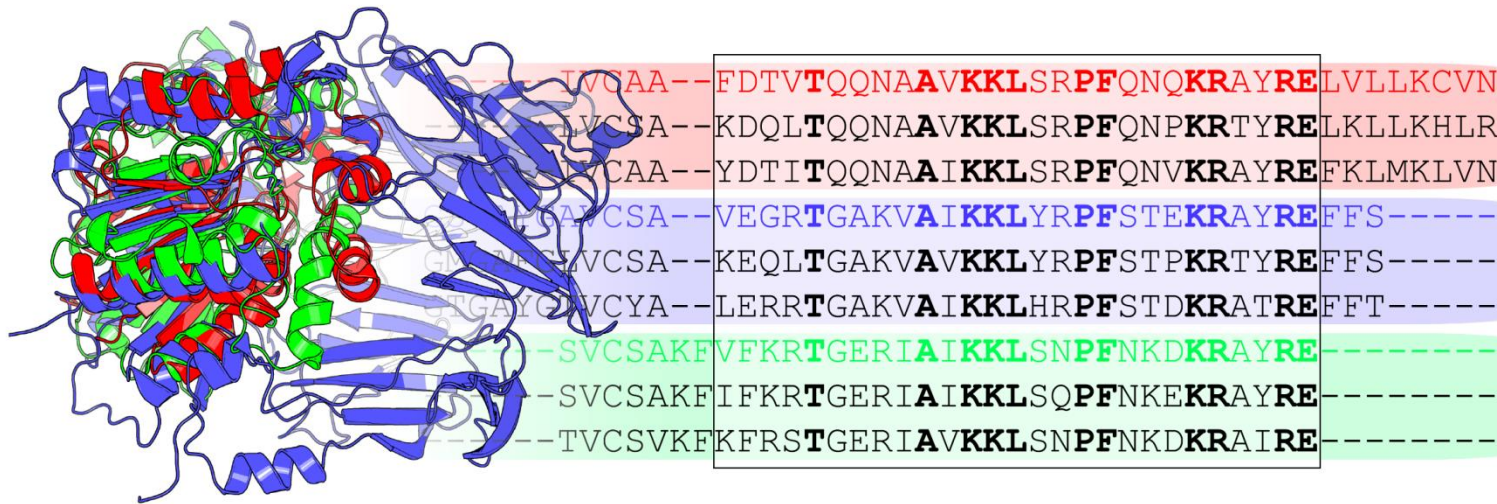


# Mustguseal

Open-access on-line platform for bioinformatic analysis in computational enzymology



<https://biokinet.belozersky.msu.ru/m-platform>

Introductory presentation

by Dmitry Suplatov

d.a.suplatov@belozersky.msu.ru

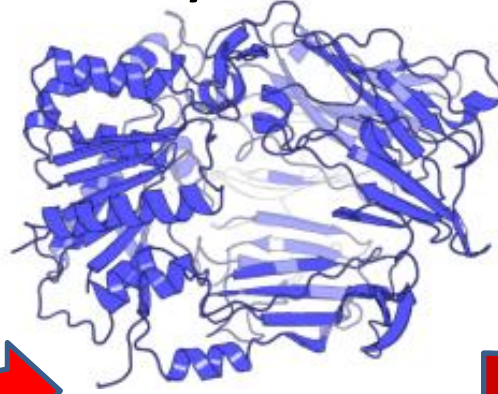
Lomonosov Moscow State University

Moscow, Russia

July 1<sup>st</sup>, 2020

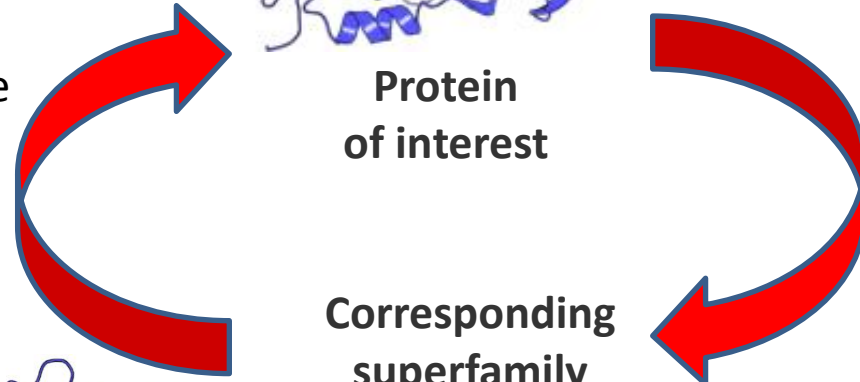
# The key concept

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

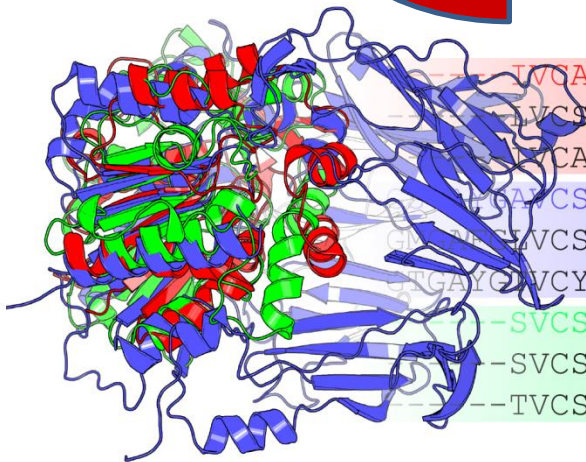


Protein of interest

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



Corresponding superfamily



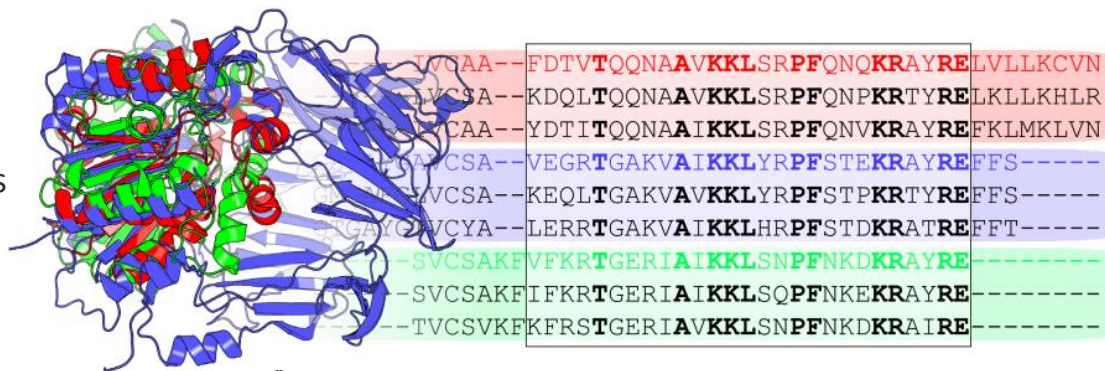
---IVGAA--	FDTV	TQQNA	AVKKLSR	PFQ	NQKRAYRE	LVLLKCVN
---IVCSA--	KDQL	TQQNA	AVKKLSR	PFQ	NPRTYRE	LKLLKHLR
---IVCAA--	YDTI	TQQNA	AIKKLSR	PFQ	NVKRAYRE	EKLMKLVN
---IVCSA--	VEGR	TGAKVA	IKKLYR	PFSTE	KRAYRE	FFS----
---IVCSA--	KEQL	TGAKVA	AVKKLYR	PFSTP	KRTYRE	FFS----
---IVCYA--	LERR	TGAKVA	IKKLR	PFSTD	KRATRE	FFT----
---SVCSAK	FVFKR	TGERIA	IKKLSN	PFNKD	KRAYRE	-----
---SVCSAK	FIFKR	TGERIA	IKKLSQ	PFNKE	KRAYRE	-----
---TVCSVK	FKFRS	TGERIA	AVKKLSN	PFNKD	KRAIRE	-----

# Open-access on-line platform for bioinformatic analysis in computational enzymology

## Mustguseal

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

*Bioinformatics*, 2018



## Zebra2

To identify and prioritize **subfamily-specific and conserved positions** as the determinants of functional diversity and key catalytic/structural residues, respectively

*Nucleic Acids Res.*, 2020

## pocketZebra

To identify and rank **binding sites** in proteins by functional significance and select particular positions in the structure that are important for selective binding of substrates and ligands

*Nucleic Acids Research*, 2014

## Yosshi

To systematically classify and study **disulfide bonds** in diverse protein families, and to assist at selecting hot-spots for disulfide engineering

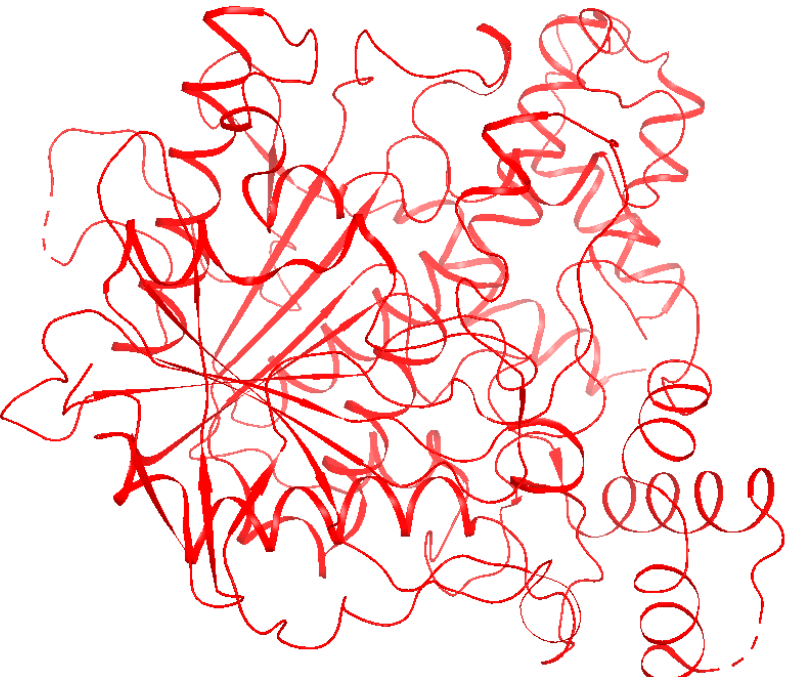
*Nucleic Acids Research*, 2019

## visualCMAT

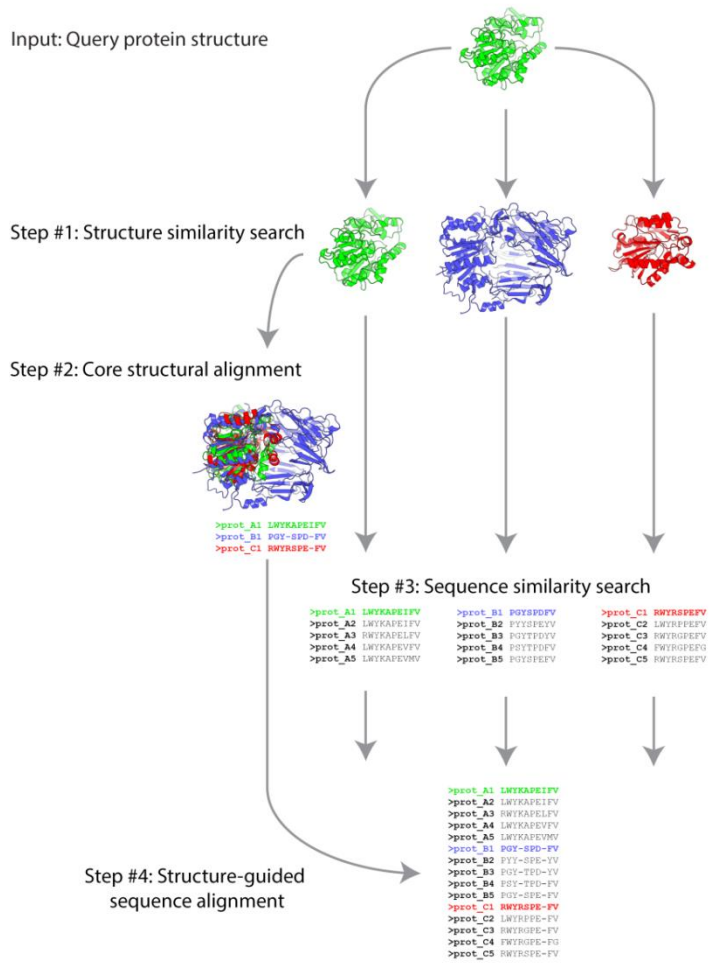
To predict and visualize **correlated mutations/co-evolving residues** in protein structures as a mechanism of allosteric communication, and a source of compensatory mutations for rational design

*J Bioinform Comput Biol.*, 2018

# Automatic construction of a large structure-guided sequence alignment of your protein family by the Mustguseal



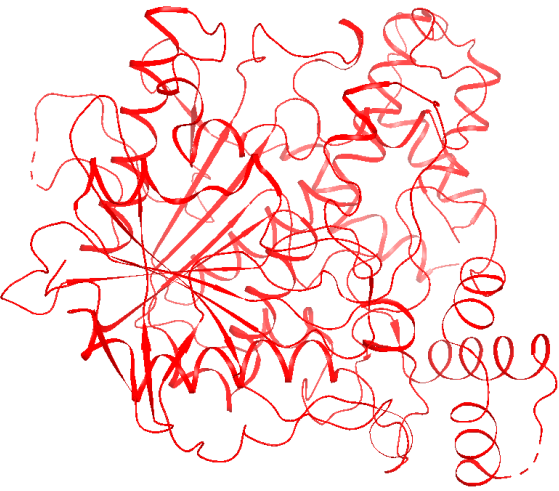
**The input:**  
PDB structure of human acetylcholinesterase



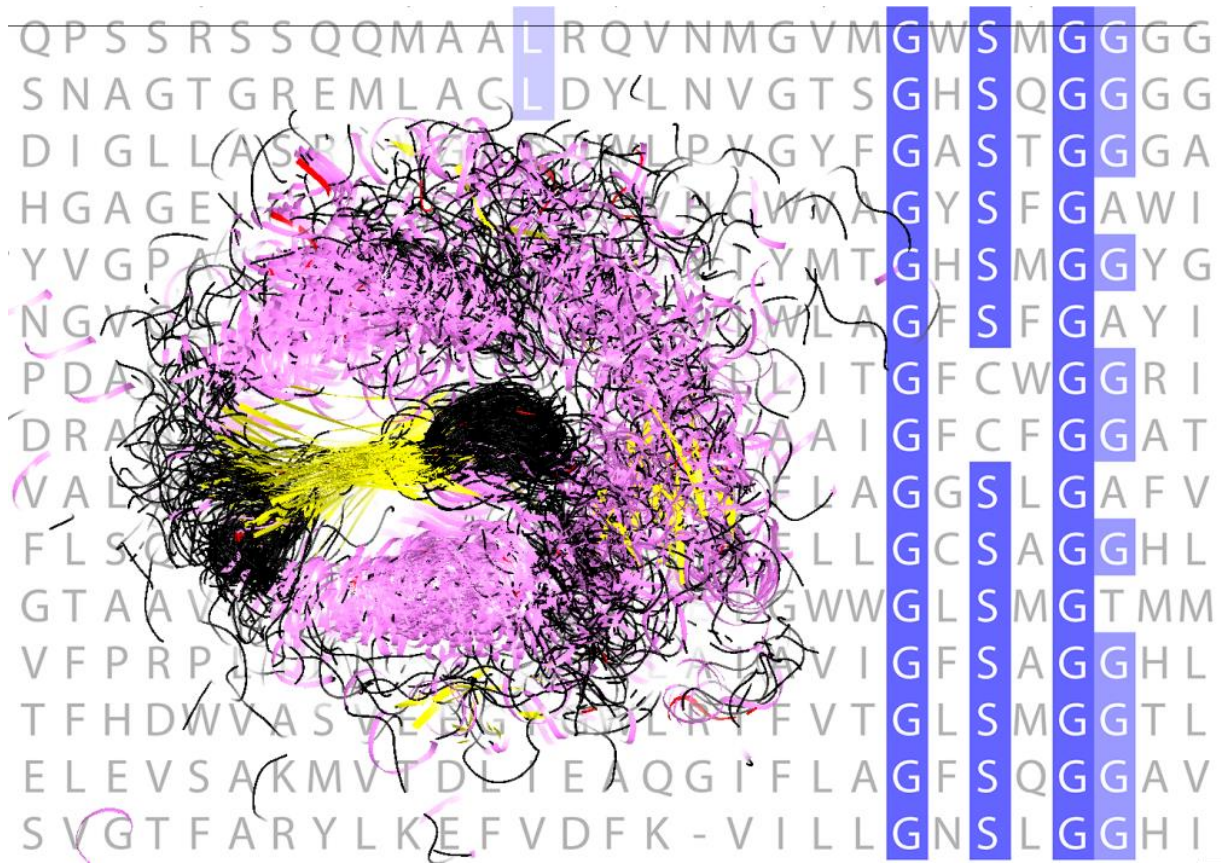
<https://biokinet.belozersky.msu.ru/mustguseal>

**The process:**  
Automatic collection and alignment of all the available protein sequences and structures from public databases

# Automatic construction of a large structure-guided sequence alignment of your protein family by the Mustguseal



**The input:**  
PDB structure of human acetylcholinesterase



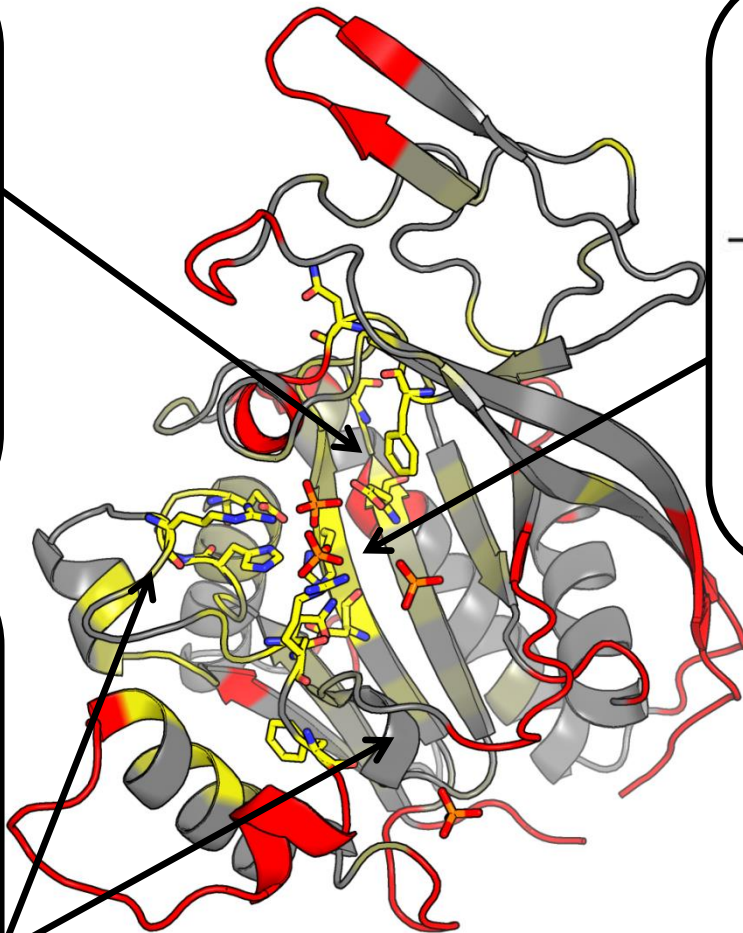
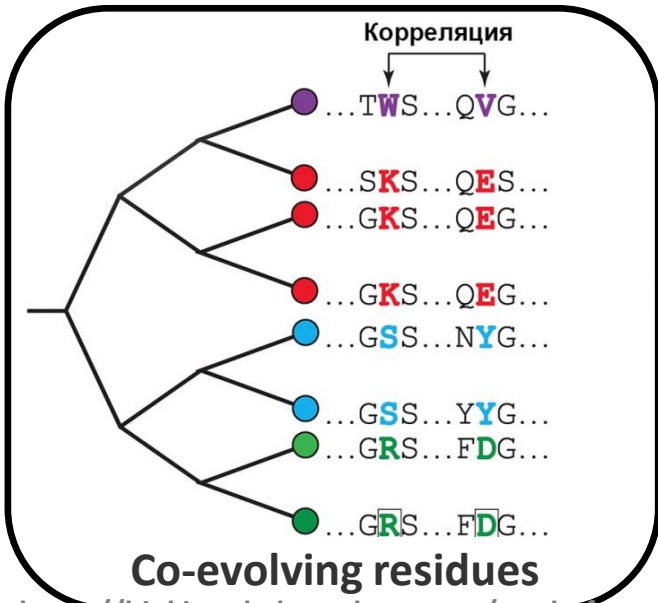
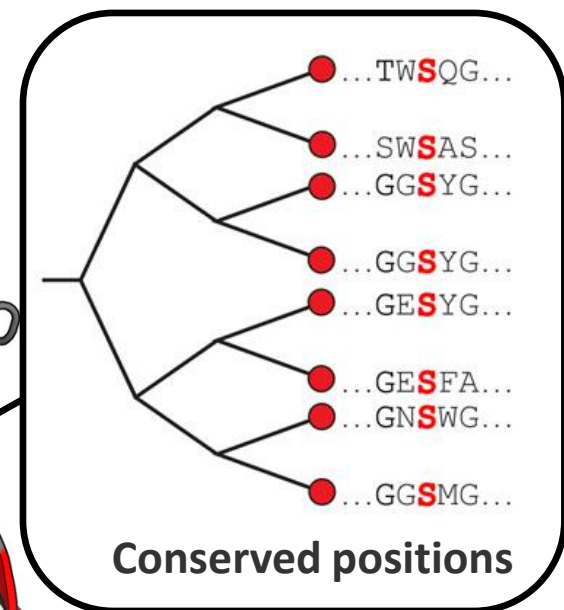
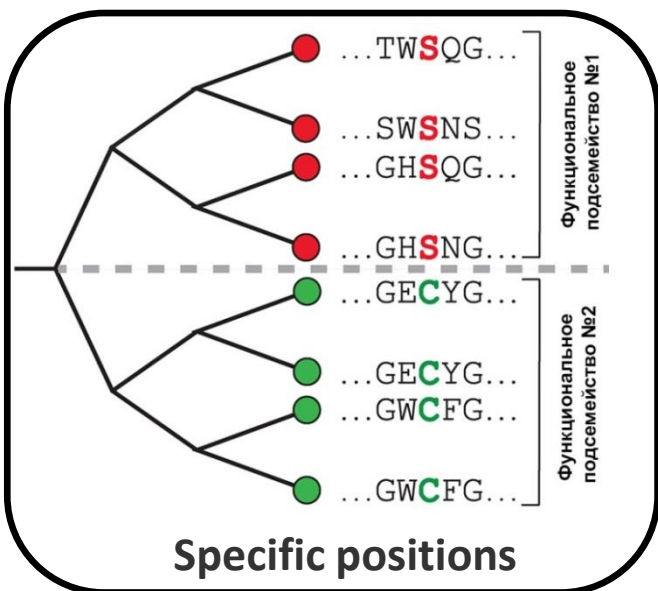
**The output:**  
Structure guided-sequence alignment of human acetylcholinesterase and its homologs from the  $\alpha/\beta$ -hydrolase superfamily

# Submit the final Mustguseal alignment for further analysis

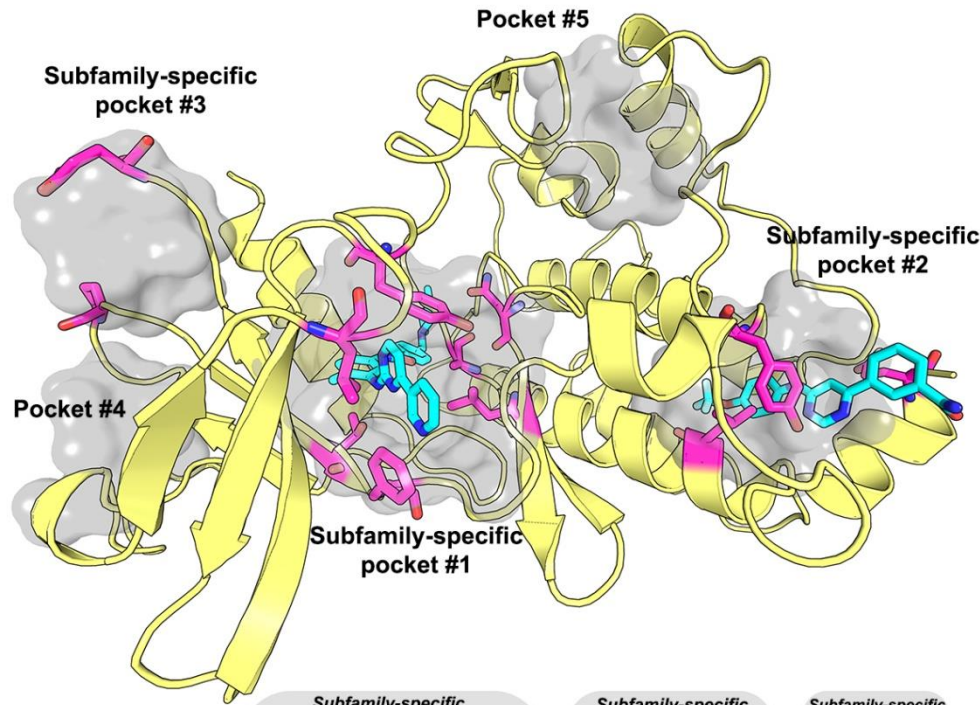
Submit the Final Mustguseal alignment	FINAL_A-ditxcb9jvnyuqz.fasta	12.0 MB	<pre>&gt;prot_A1 LMYKAEELFV &gt;prot_A2 LMYKAEELFV &gt;prot_A3 DMKAEELFV &gt;prot_A4 LMYKAEELFV &gt;prot_A5 LMYKAEELFV &gt;prot_B1 NGE-SDE-FV &gt;prot_B2 DYT-SDE-FV &gt;prot_B3 DGT-TEC-FV &gt;prot_B4 DGT-TEC-FV &gt;prot_B5 DGT-SDE-FV &gt;prot_C1 DMKAEELFV &gt;prot_C2 LMYKAEELFV &gt;prot_C3 DMKAEELFV &gt;prot_C4 DMKAEELFV &gt;prot_C5 DMKAEELFV</pre>	<a href="#">Submit to Zebra</a>
Submit the Final Mustguseal alignment and a PDB structure of the representative protein	FINAL_A-ditxcb9jvnyuqz.fasta	12.0 MB	<pre>&gt;prot_A1 LMYKAEELFV &gt;prot_A2 LMYKAEELFV &gt;prot_A3 DMKAEELFV &gt;prot_A4 LMYKAEELFV &gt;prot_A5 LMYKAEELFV &gt;prot_B1 NGE-SDE-FV &gt;prot_B2 DYT-SDE-FV &gt;prot_B3 DGT-TEC-FV &gt;prot_B4 DGT-TEC-FV &gt;prot_B5 DGT-SDE-FV &gt;prot_C1 DMKAEELFV &gt;prot_C2 LMYKAEELFV &gt;prot_C3 DMKAEELFV &gt;prot_C4 DMKAEELFV &gt;prot_C5 DMKAEELFV</pre>	<a href="#">Submit to Zebra</a> <a href="#">Submit to pocketZebra</a>
	0_1u8f_O.pdb	166 KB		<a href="#">Submit to visualCMAT</a>

- A new submission to Zebra, pocketZebra, visualCMAT, and Yosshi can be made directly from the Mustguseal Results page.

# Annotation of the protein of interest according to the bioinformatic analysis of the superfamily



# Identify and study the conserved and subfamily-specific positions



- The [Zebra](#) web-server can be used to identify and prioritize subfamily-specific and conserved positions in a functionally diverse superfamily and to select hot-spots for rational design of the query protein;

[Suplatov D., et al. \(2020\) Nucl. Acids Res.](#)

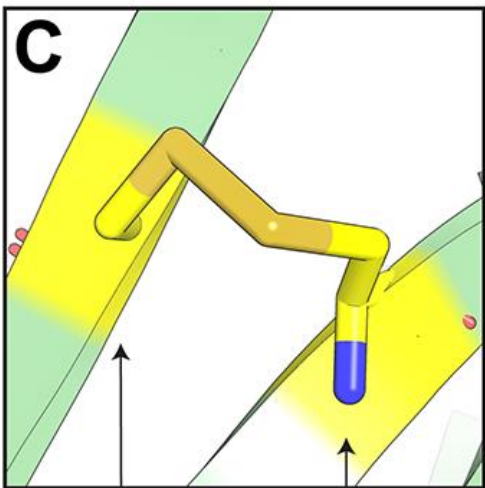
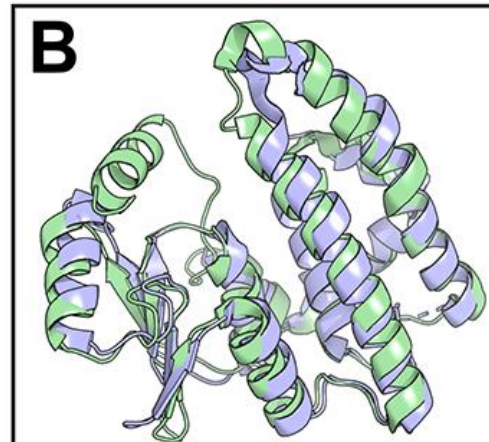
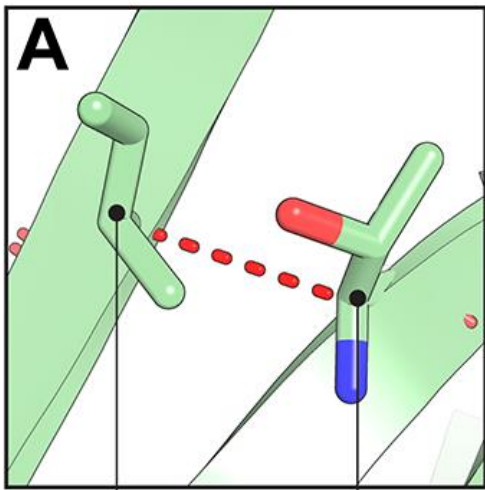
- The [pocketZebra](#) web-server can be used to identify and rank binding sites in proteins by functional significance and select particular positions in the structure that are important for selective binding of substrates/inhibitors/effectors;

[Suplatov D., et al. \(2014\) Nucl. Acids Res.](#)

	Subfamily-specific pocket #1				Subfamily-specific pocket #2				Subfamily-specific pocket #3											
Representative PDB	...	R	<b>LF</b>	...	<b>YL</b>	...	<b>TN</b>	Q	...	<b>HY</b>	...	<b>L</b>	<b>E</b>	M	...	<b>P</b>	...	<b>E</b>	K	...
Functional subfamily 1	...	R	LF	...	YL	...	TN	N	...	QY	...	L	E	L	...	LP	...	E	R	...
	...	K	LF	...	YL	...	TN	N	...	EY	...	L	E	V	...	VP	...	E	R	...
	...	N	LF	...	YL	...	TN	D	...	HY	...	L	E	V	...	VP	...	E	K	...
Functional subfamily 2	...	R	HL	...	LD	...	GV	Q	...	NA	...	F	Q	V	...	VA	...	K	K	...
	...	K	HL	...	LD	...	GV	E	...	QA	...	F	Q	L	...	LA	...	K	R	...
	...	Q	HL	...	LD	...	GV	N	...	HA	...	F	Q	M	...	VA	...	K	R	...
...	L	HL	...	LD	...	GV	V	...	NA	...	F	Q	M	...	VA	...	K	K	...	



# Systematically classify and study disulfide bonds in diverse protein families

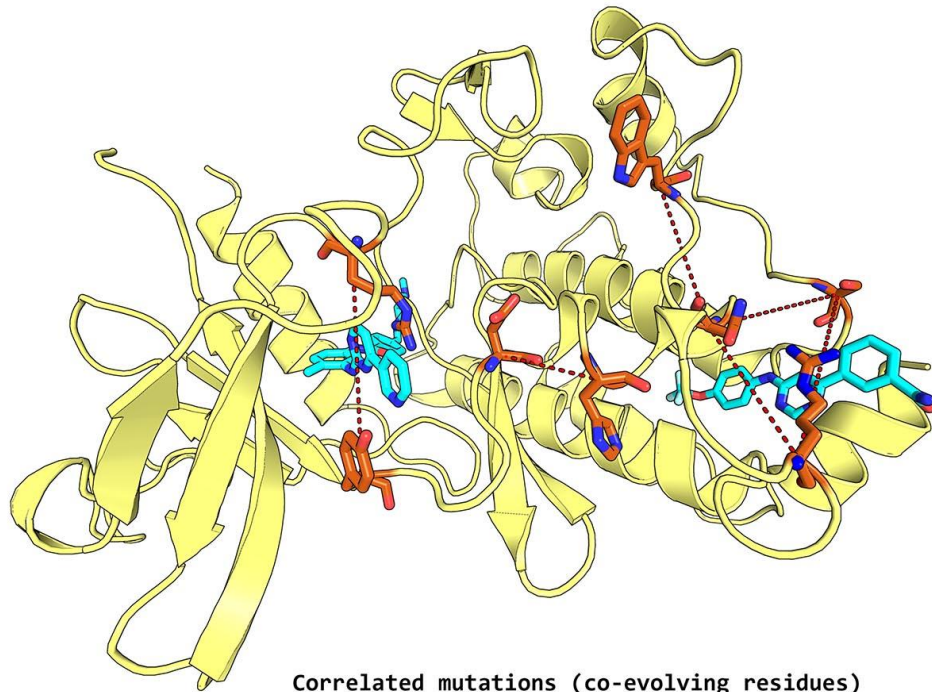


```

>PDB_A1 LE T KAPE I FV
>SEQ_A2 LE T KAPE I FV
>SEQ_A3 ID T KAPE L FV
>SEQ_A4 LE S KAPE V FV
>SEQ_A5 LD S KAPE V MV
>PDB_B1 TG I RSPD C FV
>SEQ_B2 TE I RSPE C YV
>SEQ_B3 TG C RTPD C YV
>SEQ_B4 SS C RTPD C FV
>SEQ_B5 SG C RSPE C FV
  
```

The [Yosshi](#) web-service can be used to systematically classify and study disulfide bonds in diverse protein families, and to assist at selecting hot-spots for disulfide engineering in the structure of your query protein. The "YOu web-server for S-S bond HarvestIng" is a new highly automated on-line tool for a systematic homology-driven analysis and engineering of disulfide bonds that can be easily used by a general biologist at a daily laboratory routine. The Yosshi facilitates a broader interpretation of disulfides not just as a factor of structural stability, but rather as a mechanism to implement diversity within a superfamily; [Suplatov D., et al. \(2019\) Nucl. Acids Res.](#)

# Predict and visualize the correlated mutations (co-evolving residues)



Representative PDB

Members of a protein superfamily

...	R	<b>D</b>	I	...	<b>H</b>	L	...	T	<b>W</b>	Q	...	<b>H</b>	<b>N</b>	...	<b>R</b>	...	<b>S</b>	M	...	<b>L</b>	<b>Y</b>	...	<b>R</b>	K	...
R	E	I	...	R	K	...	S	<b>K</b>	N	...	Q	...	<b>A</b>	...	<b>S</b>	...	<b>R</b>	L	...	<b>L</b>	<b>F</b>	...	<b>F</b>	R	...
...	K	E	I	...	R	L	...	T	<b>W</b>	N	...	E	<b>N</b>	...	<b>S</b>	...	<b>R</b>	V	...	<b>V</b>	<b>Y</b>	...	<b>R</b>	R	...
...	K	<b>D</b>	I	...	<b>H</b>	I	...	S	<b>K</b>	E	...	E	<b>A</b>	...	<b>S</b>	...	<b>R</b>	L	...	<b>L</b>	<b>R</b>	...	<b>F</b>	R	...
...	N	R	I	...	E	L	...	T	<b>W</b>	D	...	<b>H</b>	<b>N</b>	...	<b>T</b>	...	<b>R</b>	V	...	<b>V</b>	<b>R</b>	...	<b>Y</b>	K	...
...	R	<b>D</b>	I	...	<b>H</b>	V	...	S	V	Q	...	N	<b>A</b>	...	<b>E</b>	...	<b>K</b>	V	...	<b>V</b>	<b>F</b>	...	<b>R</b>	K	...
...	K	R	I	...	E	A	...	S	V	E	...	Q	<b>A</b>	...	<b>E</b>	...	<b>K</b>	L	...	<b>L</b>	<b>W</b>	...	<b>R</b>	R	...
...	Q	<b>D</b>	I	...	<b>K</b>	G	...	S	V	N	...	H	<b>W</b>	...	<b>R</b>	...	<b>D</b>	M	...	<b>V</b>	<b>W</b>	...	<b>R</b>	R	...
L	<b>D</b>	I	...	<b>K</b>	L	...	S	V	V	...	N	<b>W</b>	...	<b>R</b>	...	<b>D</b>	M	...	<b>V</b>	<b>R</b>	...	<b>W</b>	K	...	

<https://biokinet.belozersky.msu.ru/visualcmat>

- The [visualCMAT](#) web-service can be used to predict and visualize correlated mutations/co-evolving residues in protein structures. The visualCMAT can be used to understand the relationship between structure and function and identify co-evolution patterns in protein superfamilies, implemented at selecting hotspots and compensatory mutations for rational design and directed evolution experiments to produce novel enzymes with improved properties, and employed at studying the mechanism of selective ligand's binding and allosteric communication between topologically independent sites in protein structures; [Suplatov D., et al. \(2018\) J Bioinform Comput Biol.](#)

# Contacts

- Support and collaboration  
[d.a.suplatov@belozersky.msu.ru](mailto:d.a.suplatov@belozersky.msu.ru)  
[vytas@belozersky.msu.ru](mailto:vytas@belozersky.msu.ru)
- Press to ask your question on-line



**Ask for help**



Lomonosov Moscow State University, the Main building  
Dmitry Suplatov @ 2014 All Rights Reserved