Yosshi: the bioinformatic approach to protein disulfide engineering

Dmitry Suplatov[1*], Daria Timonina[2], Yana Sharapova[2], Vytas Švedas[1,2]

[1] Belozersky Institute of Physicochemical Biology and [2] Faculty of Bioengineering and Bioinformatics at Lomonosov Moscow State University, Moscow, Russia

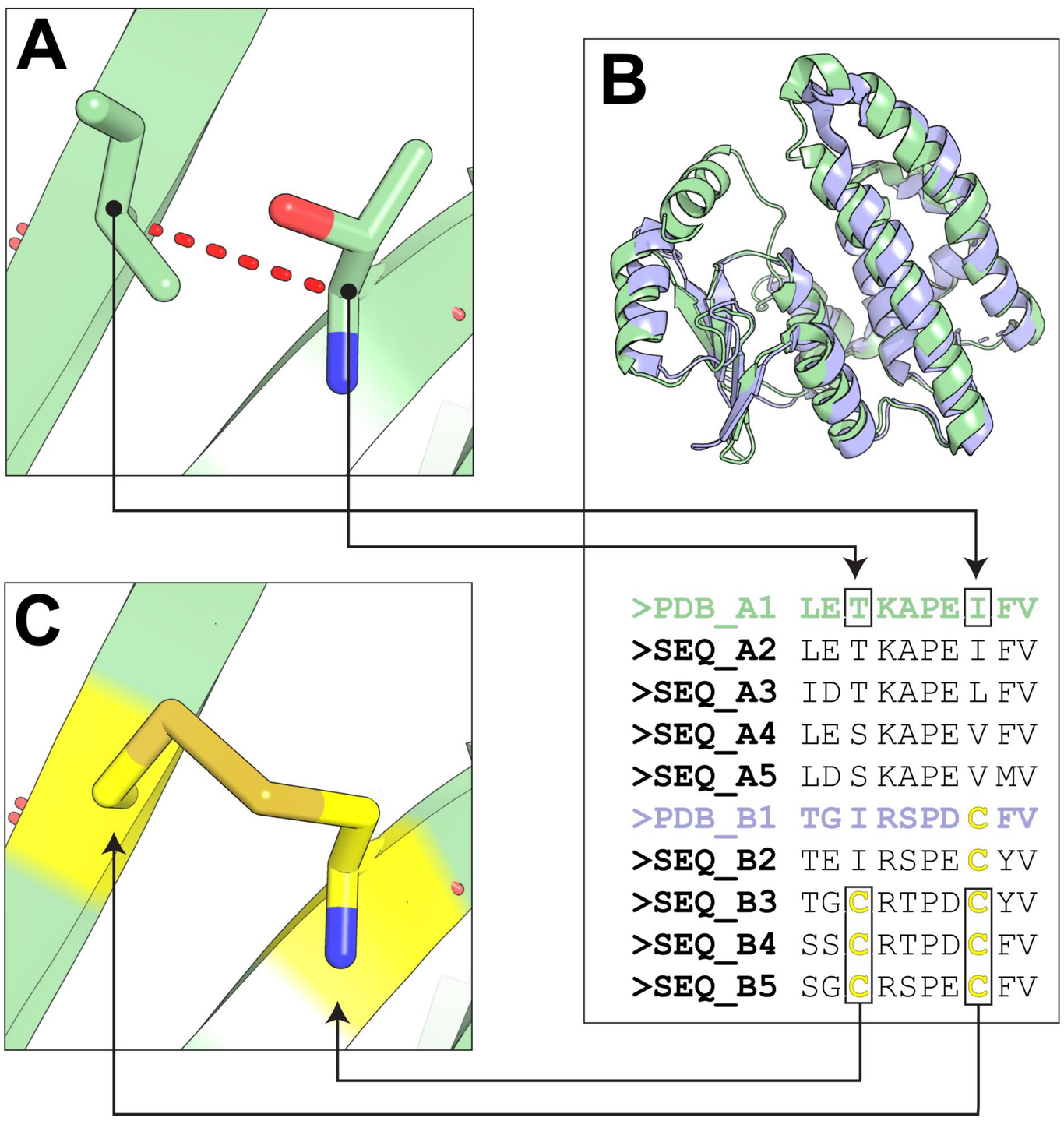
[*] d.a.suplatov@belozersky.msu.ru

You are 90 minutes away from a comprehensive evolution-inspired guide to disulfide engineering of your protein

Disulfide bonds are covalent crosslinks between the thiol groups of two cysteine residues that play a significant role in protein stability, function or regulation but are poorly conserved among evolutionarily related proteins. This provides an opportunity to study the roles that disulfide bonds play within a common structural fold of a superfamily by comparative analysis of homologs, as well as to introduce S-S bridges which naturally occur in some proteins into the structures of their homologs to improve stability or modulate function. We have developed the Yosshi – "Your web-server for S-S bond harvesting" in protein superfamilies – a new highly automated on-line tool for a systematic homology-driven analysis and engineering of disulfide bonds available at https://biokinet.belozersky.msu.ru/yosshi. The Yosshi facilitates a broader interpretation of disulfides not just as a factor of structural stability, but rather as a more universal, still insufficiently explored evolutionary instrument to implement functional diversity within a common structural fold of a superfamily.

I. Algorithm outline

The key novelty of the Yosshi is the use of bioinformatic analysis to search for pairs of cysteine residues in sequences of homologs followed by the 3D-motif analysis to evaluate whether introduction of the selected cysteines at corresponding positions in the user-submitted query protein can result in a formation of a disulfide bond

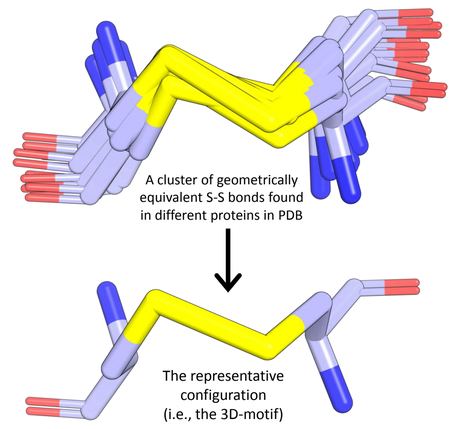


II. 3D-motif analysis

Assessing protein structure flexibility at disulfide engineering

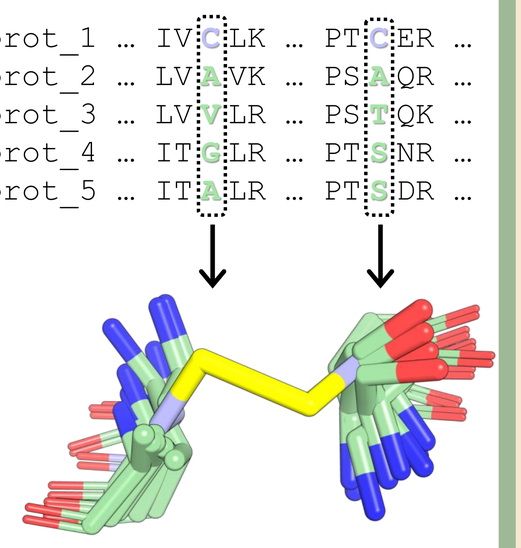
1. Library of 3D-motifs of S-S bonds

- Clustering of S-S bonds with equivalent geometry from different protein structures;
- Selection of a representative configuration in each cluster, i.e., a 3D-motif;
- **273 3D-motifs of S-S bonds were identified in the PDB** database (≤2.5Å)



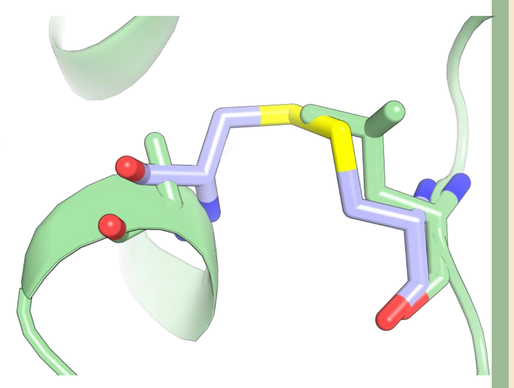
2. "Flexible" statistical model

Comparison of the S-S bond in a 3D-motif (blue) with its non-bonded equivalents in structures of homologous proteins (green): RMSD of the backbone atoms between the S-S bond and its non-bonded equivalents in homologs is calculated to establish the limits of flexibility of a pair of amino acid residues capable of a disulfide bond formation upon mutation to cysteines



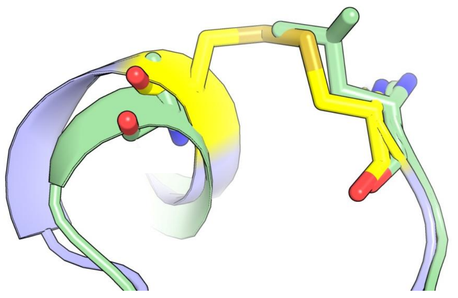
3. 3D-motif analysis of the query protein

A pair of positions in the query protein is confirmed as a promising site for S-S bond formation **if the backbone RMSD with at least one 3D-motif is within X=0.70Å**, what corresponds to a p-value of $P(x > X)=0.05$ of a normal distribution with $\mu=0.39$Å and $\sigma=0.19$Å.



4. Construction of molecular models of mutants

The respective **pair of positions in the query protein structure is replaced by the most similar 3D-motif** (i.e. selected by the lowest RMSD of the backbone) and this initial three-dimensional model of the mutant is further geometry-optimized



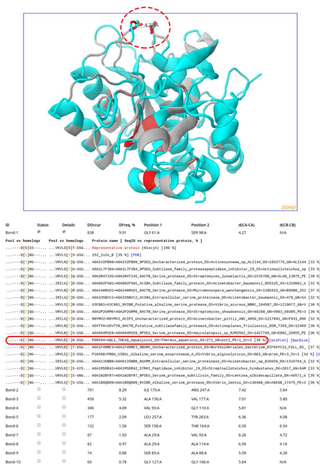
III. Output

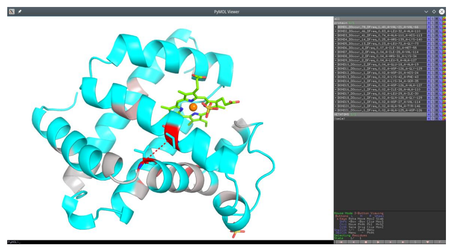
Yosshi provides a detailed homology-based annotation of disulfides within a common structural fold of the superfamily

Scan for details

The first output is a **list of pairs of positions** in the structure of the query protein that can form a disulfide bond assuming both residues are mutated to cysteines, or are already occupied by cysteines that can form a crosslink;

The second output is a **list of homologs** of the query protein, which contain cysteines in equivalent positions for each pair of such residues.





The results are delivered as a content-rich **PyMol session file** or can be studied **online** using the HTML5-based interactive analysis tools

IV. Publications

1. D.Suplatov, D.Timonina, Y.Sharapova, V.Švedas (2019). Yosshi: a web-server for disulfide engineering by bioinformatic analysis of diverse protein families, *Nucleic Acids Res.*, 47(W1), 308–14.
2. D.Suplatov, K.Kopylov, N.Popova, V.Voevodin, V.Švedas (2018). Mustguseal: a server for multiple structure-guided sequence alignment of protein families, *Bioinformatics*, 34(9):1583-85.

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