 doi:10.20944/preprints202012.0089.v1

Web-based Tools for Computational Enzyme Design

Sérgio M. Marques[1,2], Joan Planas-Iglesias[1,2], Jiri Damborsky[1,2*]

[1] Loschmidt Laboratories, Department of Experimental Biology and RECETOX, Faculty of Science, Masaryk University, Kamenice 5/A13, 625 00 Brno, Czech Republic

[2] International Centre for Clinical Research, St. Anne's University Hospital Brno, Pekarska 53, 656 91 Brno, Czech Republic

*Corresponding author: Damborsky, Jiri (jiri@chemi.muni.cz)

Abstract

Enzymes are on high demand for very diverse biotechnological applications. However, natural biocatalysts often need to be engineered for fine-tuning their properties towards the end applications, such as the activity, selectivity, stability to temperature or co-solvents, and solubility. Computational methods are increasingly used in this task, providing predictions that narrow down the space of possible mutations significantly and can enormously reduce the experimental burden. Many computational tools are available as web-based platforms, making them accessible to non-expert users. These platforms are typically user-friendly, contain walk-throughs, and do not require deep expertise and installations. Here we describe some of the most recent outstanding web-tools for enzyme engineering and formulate future perspectives in this field.

Keywords

Web server; rational design; functionalizing enzymes; enzyme discovery; user-friendly tools.

 doi:10.20944/preprints202012.0089.v1

1. Introduction

Enzymes are the catalysts used by nature to perform the complex chemical reactions required to sustain life. They evolved over billions of years to achieve high efficiency and specificity required by each lifeform to survive and thrive in their environments. Biotechnology has emerged as a way of mankind to exploit such nature's creations, with numerous benefits over the classical chemical processes. In many cases, however, technological applications require particular properties beyond what is available in naturally occurring biomolecules, such as specific activity, selectivity, stability, solubility, etc. In such cases, they have to be reengineered on-demand [1].

The global protein-engineering market was evaluated in USD 1.9 billion in 2018, and it is projected to reach USD 3.9 billion in 2024 with the remarkable annual growth of 12.4% (CAGR) during this period. The rational protein design accounted for the largest technology segment of the market in 2018, while the biopharmaceutical companies accounted for the largest end-user [2]. This categorically demonstrates the growing importance of protein engineering.

Directed evolution methods have been extensively used to successfully improve natural biomolecules. However, they can be expensive and time-consuming. Therefore, the usage of computational methods for rational design is becoming more and more common. The predictive power of computational tools is gradually improving. Many of the existing tools are developing intuitive and user-friendly web-based platforms, which expand their usability to the broader community. Without the need for software installation or Unix command-line environment, these platforms are ideal for non-specialists.

In this review, we focus on web-based computational tools for enzyme engineering. We describe the recently developed web servers that we consider the most outstanding (Table 1), after a selection from a larger pool of tools (Supplementary Tables S1-S6). We have organized them by the focus: i) enzyme discovery, ii) protein solubility, iii) enzyme activity and specificity, iv) protein stability, v) protein dynamics, and vi) multipurpose. We omitted the tools specialized on the structure prediction or identification of protein-protein interactions to keep the focus on enzyme design.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 December 2020 doi:10.20944/preprints202012.0089.v1

Table 1. Selected web-based computational tools for enzyme engineering classified by the focus and published between 2018 and 2020. The full list of the tools is available in the Supplementary Tables S1-S6.

Web server	URL *Description*	Input[a]	Output[b]	Runtime	Ref.
	Enzyme discovery				
HEC-net	http://hecnet.cbrlab.org *Deep learning application to predict the enzymatic activity (EC number up to its fourth level) of a protein sequence.*	- Sequence	- EC class prediction, up to the fourth level	Minutes	[3]
Bio2Rxn	http://design.rxnfinder.org/bio2rxn *Consensus prediction of the enzymatic activity (EC number up to its fourth level) of a protein sequence from six different predictive methods*	- Sequence	- Reaction type - Reaction schema - EC class prediction, up to the fourth level - List of predictors agreeing with the predicted class	Hours	[4] [*]
GSP4PDB	https://structuralbio.utalca.cl/gsp4pdb *Search for deposited protein structures (PDB) compatible with the input graph-based structural pattern*	- Graph-based structural pattern	- List of compatible PDB structures	Seconds to minutes	[5]
LIBRA-WA	http://biochimica3.bio.uniroma3.it/LIBRAWA *Identification of protein function via recognition of binding pockets*	- PDB ID - PDB file - Ligand database - Ligand PDB ID - Search parameters	- List of PDB structures with compatible binding pocket. - Their ligands - Structural similarity score - Confidence	Minutes	[6]
EnzymeMiner	https://loschmidt.chemi.muni.cz/enzymeminer *To retrieve a list of protein sequences that potentially have the same enzymatic function than the input ones, considering their essential residue profiles.*	-Sequence(s) - Essential residues	- List of putative hits annotated with multiple scores - Similarity network view	Hours	[7] [**]
	Engineering protein solubility				
AGGRESCAN3D 2.O	http://biocomp.chem.uw.edu.pl/A3D2 *To predict the solubility of a protein from its three-dimensional structure, the effects of mutations on such structure, and the optimal substitutions for solubilization*	- PDB ID - PDB file - Mutations list -Non-mutable residues list - Preferences on optimization	- Aggregation profile - Per residue score - 3D view. - effect of mutations	Minutes	[8] [**]
SOLart	http://babylone.ulb.ac.be/SOLART *To predict the solubility of a protein from its three-dimensional structure*	- PDB ID - PDB file	- Predicted solubility value - Scores of the individual	Seconds	[9]

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 December 2020 doi:10.20944/preprints202012.0089.v1

			features used for prediction		
AggreRATE-pred	http://www.iitm.ac.in/bioinfo/aggrerate-pred *To predict the effects of mutations on solubility*	- Sequence - PDB ID - PDB file - Mutations list	- Predicted aggregation rate for each mutation	Minutes to hours	[10]
Solubility-Weighted Index	https://tisigner.com/sodope *To predict the solubility and flexibility of an input protein sequence, having the possibility of focusing only in a region of the sequence*	- Sequence	- Solubility probability - Flexibility score -hydropathy score - Hydrophobicity and flexibility - Solubility tags suggestion	Seconds	[11] [*]
SoluProt	https://loschmidt.chemi.muni.cz/soluprot *To predict the solubility of a protein specified by an input sequence*	-Sequence -List of sequences	- Solubility score	Seconds	[12] [**]
Engineering enzyme activity and selectivity					
FuncLib	http://FuncLib.weizmann.ac.il *To redesign an active site and create multiple-point designs. Based on conservation analysis and energy calculations*	- PDB ID - PDB file - Mutable residues - Essential residues - Ligands	- List of mutants ranked by ΔΔG -PDB structures of the best designs	Hours	[13] [**]
CaverDock (Caver Web)	https://loschmidt.chemi.muni.cz/caverweb *To calculate trajectory and interaction energy profiles of a ligand travelling through a protein tunnel. Available at the Caver Web interface*	- PBD ID - PDB file - Ligand file - Ligand drawing - Ligand smiles - ZINC ligand code	- Ligand trajectory as PDBQT file - Binding energy profiles - Energy barriers	Minutes to hours	[14]
DaReUS-Loop	http://bioserv.rpbs.univ-paris-diderot.fr/services/DaReUS-Loop *For modeling or remodeling loops in homology models and finding the best loops conformation*	- PBD ID - PDB file - Sequence	- PDB structures of the modeled proteins - Confidence scores	Minutes to hours	[15] [**]
nAPOLI	http://bioinfo.dcc.ufmg.br/napoli *To analyze protein-ligand interactions detecting important conserved interacting residues*	- PDB ID(s) - PDB file(s) - Interaction cutoffs	- Interactive view of contact residues - Interaction networks - Analysis charts and tables of contacts	Minutes	[16]
Engineering protein stability					

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 December 2020 doi:10.20944/preprints202012.0089.v1

FireProtASR	https://loschmidt.chemi.muni.cz/fireprotasr *To perform ASR and infer ancestral sequences and find more stable (and promiscuous) proteins*	- Sequence - Own multiple sequence alignment - Essential residues	- Multiple sequence alignment - Phylogenetic tree - Sequence analysis and visualization	Hours	[17] [*]
TKSA-MC	http://tksamc.df.ibilce.unesp.br *To find hotspots by optimizing the protein charge interactions. Calculates electrostatic free energy ΔG_{ele} of all polar/charged residues to identify destabilized ones*	- PDB ID - PDB file - pH value or range - Temperature	- Electrostatic energy for every ionizable residue - ΔG_{ele} vs. pH profile	Seconds	[18]
pStab	http://pbl.biotech.iitm.ac.in/pStab *To engineer protein stabilities through mutations involving charged residues. A statistical mechanical model is employed to predict the unfolding curves for the selected mutants as a function of temperature*	- PDB ID - PDB file - pH - Temperature	- Pairwise electrostatic interactions - Mutational hotspots - Stabilizing mutations and respective $\Delta\Delta G$ - Thermal unfolding curves - Local stability profiles	Minutes to hours	[19]
ProTSPoM	http://cosmos.iitkgp.ac.in/ProTSPoM *To estimate the thermodynamic stabilization $\Delta\Delta G$ upon single-point mutations. Based on machine learning*	- PDB file - Mutation	- $\Delta\Delta G$ value	Seconds	[20]
Yosshi	https://biokinet.belozersky.msu.ru/yosshi *To select hot-spots for introducing disulfide bonds which naturally occur in some proteins, based on multiple-sequence alignments*	- PDB file - Multiple sequence alignment	- Cysteine-mutation pairs - Disulfide frequencies - PyMOL session with hotspots - Structures of disulfide mutant	Seconds to minutes	[21] [*]
SSbondPre	http://liulab.csrc.ac.cn/ssbondpre *To predict disulfide bonds to enhance the protein structural stability based on machine learning and geometric restraints*	- PDB ID - PDB file	- Cysteine-mutation pairs - Energy and entropy change	Seconds	[22]
mCSM-membrane	http://biosig.unimelb.edu.au/mcsm_membrane *To predict the stability or pathogenic effects of mutations on membrane protein and the likelihood of them being disease-associated*	- PDB ID - PDB file - Uniprot ID - Sequence - Mutations	- $\Delta\Delta G$ value - PyMOL session with the contact maps	Seconds	[23]
DenseCPD	http://protein.org.cn/densecpd.html *To predict the probabilities of the 20 natural amino acids for each residue in a protein structure, considering the three-dimensional density*	- PDB ID - PDB file	- Probability scores for every mutation	Minutes	[24]

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 December 2020 doi:10.20944/preprints202012.0089.v1

	distribution of protein backbone. Uses machine learning				
Engineering protein dynamics					
DynaMut2	http://biosig.unimelb.edu.au/dynamut2 *To assess changes in stability and flexibility upon mutation.*	- PDB ID - PDB file - Mutations list	- ΔΔG value - 3D view (wild type or mutant) with predicted interactions of the mutated residue(s) and B-factor and hydrophobicity mapping - B-factor profile	Seconds to minutes	[25] [*]
CABS-flex 2.0	http://biocomp.chem.uw.edu.pl/CABSflex2 *To evaluate the flexibility of the input protein structure*	- PDB ID - PDB file - execution parameters	- Ensemble of structure files for all normal modes. - Contact (cross-correlations) map - Fluctuations plot	Minutes to hours	[26]
ProSNEx	http://prosnex-tool.com *To evaluate the flexibility of the input protein structure*	- PDB ID - PDB file - execution parameters	- 3D view mapping calculated properties on the structure - Network analysis - Cross-correlation maps	Minutes to hours	[27] [*]
AlloSigMA 2	http://allosigma.bii.a-star.edu.sg *To evaluate the allosteric effects of ligand binding or mutations*	- PDB ID - PDB file - mutation list - ligand-binding site	- Per residue ΔΔG value - Allostery modulation graph for the effects of ligand binding or mutation.	Minutes	[28]
LARMD	http://chemyang.ccnu.edu.cn/ccb/server/LARMD *To perform short-timed fully atomistic conventional (Int_mod) and steered (str_mod) molecular dynamic simulations, and normal mode analysis (nor_mod) for the study of ligand binding and unbinding.*	-PDB ID -PDB file - Ligand	- Ensemble of structure files representing the trajectory. - PCA analysis, conformation clusters and dynamic residue cross-correlations (int_mod)	Int_mod: minutes, Str_mod: hours	[29] [*]

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 December 2020 doi:10.20944/preprints202012.0089.v1

			-Hydrogen bond analysis and energy decomposition if protein-ligand is submitted (int_mod) - Tunnel and transport energy profile (str_mod) - Cross-correlations and residues fluctuations map (nor_mod)		
Multipurpose					
Caver Web	https://loschmidt.chemi.muni.cz/caverweb *To calculate tunnels in proteins with buried binding sites and analyze the ligand transport through those tunnels*	- PBD ID - PDB file - Ligand file, drawing, SMILES or ZINC code	- Enzyme cavities - Enzyme tunnels - Tunnel profiles - PyMOL session with tunnels - Tunnel residues - Ligand energy profiles - Energy barriers - Ligand trajectory	Pockets: seconds, Tunnels: seconds, Ligand transport : minutes to hours	[30] [**]
HotSpot Wizard 3.0	http://loschmidt.chemi.muni.cz/hotspotwizard *For automated identification of hotspots in semi-rational protein design to give improved protein stability, catalytic activity, substrate specificity and enantioselectivity*	- PDB ID - PDB file - Sequence	- Enzyme pockets - Enzyme tunnels - Multiple sequence alignment - Homology models - Correlated residues - Amino acid frequency - Mutational effect on function - Flexibility (B-factors) - Mutability scores -Sequence consensus	Analysis: hours, Mutations design: minutes to hours, Library design: seconds	[31] [*]

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 December 2020 doi:10.20944/preprints202012.0089.v1

			- ΔΔG of designed mutants - Library design for saturation mutagenesis		
ROSIE[c]	http://rosie.rosettacommons.org *Unified platform hosting several tools specialized in different aspects of molecular modeling and analysis of proteins and nucleic acids*	*Depends on the sub-tool*	*Depends on the sub-tool*	Minutes to days	[32] [*]
ProteinsPlus	https://proteins.plus *To analyze and predict multiple properties and features of proteins, their binding sites and interactions with ligands*	- PDB ID - PDB file - Ligand file - Keyword search	- Hydrogen prediction - Protein pockets and binding sites - 2D interaction diagrams - Ensemble compilation from PDB - Protein-protein interaction analysis - Metal coordination prediction - Protein-ligand affinity or activity - Placement of water molecules in the active site - Structure quality assessment	Seconds to minutes	[33] [*]
pPerturb	http://pbl.biotech.iitm.ac.in/pPerturb *To quantify the strength of an interaction network by employing perturbations (alanine mutations); can predict the extent of destabilization of proteins arising from side-chain truncations*	- PDB ID - PDB file - Target residue	- Perturbation profiles - Contact network plots - Allosteric hotspots - Thermal unfolding curves	Minutes	[34]

[a]Some of the listed items are mandatory and some are optional; [b]ΔΔG is the stabilization energy, and corresponds to the change in free energy upon each mutation from the wild-type or template; [c]ROSIE hosts many individual tools to be discussed individually.

2. Enzyme discovery

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 December 2020 doi:10.20944/preprints202012.0089.v1

A common strategy for getting a good catalyst for a given substrate is to find a new natural enzyme in the genomic databases. Interestingly, there exists a vast space allowing for the discovery of novel enzymes, since the proportion of protein sequences that have not yet been biochemically characterized is huge: only 1 in every 450 protein sequences present in the NCBI nr database [35] has a record potentially encompassing functional annotation in the manually curated UniProtKB/Swiss-Prot database [36]. Despite the availability of high throughput methods for biochemical characterization of large numbers of gene expression products [37], *in silico* approaches can conveniently reduce the time and costs of the process.

The task of discovering new enzymes can be tackled in different manners. A straightforward strategy is to predict the enzymatic activity of a protein from its sequence. HEC-Net [3] is a deep learning tool that exploits strategies based on sequence pattern recognition, sequence similarity and amino-acid biochemical properties to achieve prediction accuracy over 90% on the fourth level of the Enzyme Commission (EC) classification. Also exploiting deep learning, Bio2Rxn [4] [*] produces a consensus prediction based on six individual predictors. One of them is based on convolutional neural networks that are trained exclusively on EC-number annotated protein sequences. The other five are more traditional predictors based on sequence similarity, identification of sequence patterns and amino-acid biochemical properties. Bio2Rxn retains high precision values (over 90%) while increasing recall (close to 60%) when compared to similar tools.

A complementary strategy consists of identifying protein-ligand structural motifs from the protein structures in RCSB Protein Data Bank (PDB) [38]. This approach relies on the rationale that such binding interfaces are substrate-specific, which in turn is an essential fingerprint of the catalytic process. GSP4PDB [5] allows the user to design and define the so-called Graph-based Structural Patterns as the representations of the protein-ligand interface and then query for such patterns in the PDB, thus returning proteins that could potentially accommodate the ligand. LIBRA-WA[6] is a web-based application that exploits network theory to identify binding pockets in an input protein. Such identification is done upon comparison with two precompiled databases, a ligand-binding sites and the Catalytic Site Atlas [39].

A third strategy consists of finding the existing protein sequences that could carry on an enzymatic function. While the first approach relied on precisely predicting the enzymatic function of an input sequence, here the challenge is to comprehensively identify the maximum number of protein sequences able to perform a given catalytic function. EnzymeMiner [7] [**] accepts several proteins with known enzymatic function as input, infers their essential or catalytic residues, and exploits different tools for the assessment of sequence similarity to identify such potential catalysts (Figure 1). In contrast

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 December 2020 doi:10.20944/preprints202012.0089.v1

with the tools presented in the second strategy, EnzymeMiner does not rely on knowledge of 3D structure of proteins. The tool is fully automated, ranks sequences by their predicted solubility and provides annotations on source organism, extremophilicity, structure availability, etc., to guide the selection process.

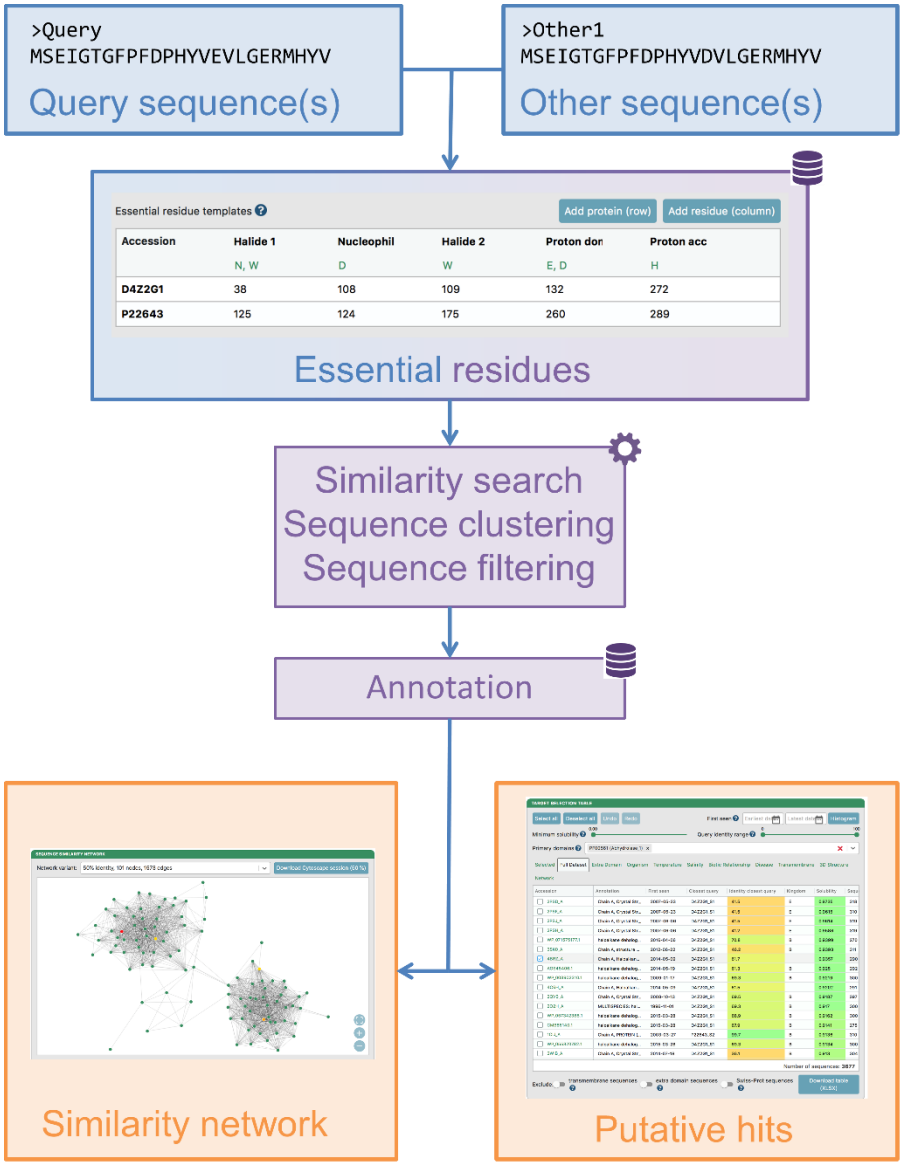

Figure 1. Illustration of the EnzymeMiner workflow [7] [**]. The web server accepts several sequences with the desired function. The user can also input 'other sequences', performing the desired function to help on the sequence filtering step. The server can retrieve catalytic residues from the Catalytic Site Atlas, or the user can define them (allowing for degenerated positions). The query proteins are used to search for homologs, and the obtained hits are subsequently clustered and filtered ensuring the presence of the defined essential residues. Multiple annotations are retrieved to enrich the information of the filtered list of hits. The final results are presented in two interactively integrated views: (i) Putative Hits allows for prioritization according to any of the retrieved annotations and (ii) the Similarity Network view presents the sequences clustered according to their sequence similarity.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 December 2020 doi:10.20944/preprints202012.0089.v1

3. Engineering protein solubility and aggregation

A recurring problem upon producing engineered variants of proteins is that they may suffer from diminished solubility or aggregation. Several approaches relying on sequence and structure properties provide solutions for solubility prediction and optimization [40,41].

Among the methods requiring the input of 3D structure, Aggrescan3D 2.0 [8] [**] is a well-established aggregation predictor that projects a pre-calculated intrinsic aggregation propensity scale to the query protein structure. Thus, the aggregation propensity values that are used to produce the final prediction are modulated by the specific structural context of the evaluated region or patch. The newest version improves the predictions by considering protein flexibility and stability and by providing suggestions of optimized solubility.

SOLart [9] relies on structure-derived statistical potentials to infer the query protein solubility. Differences in Gibbs free energy inferred from such statistical potentials - especially those considering backbone torsion angles, solvent accessibility and inter-residue distances - allow for accurate predictions of solubility when compared with experimental values, achieving a Pearson's correlation coefficient of 0.67 and 0.51 on independent validation set and modelled proteins, respectively.

AggreRATE-pred [10] integrates amino acid physicochemical and structural-based properties, and mutational and contact propensities in a multiple regression model to predict the effect of mutations on the aggregation rates. The chosen model to be applied depends on the protein length and the secondary structure type on where the mutation(s) occur. This strategy achieves a correlation in between experimental and predicted values of up to 0.82 and performs also well on modeled proteins. Interestingly, this approach does not rely on any structural information for short peptides (< 40 amino acids).

When the 3D structure of the protein to engineer is not available or obtaining a model becomes challenging, solubility can also be predicted from the protein sequence. Solubility-Weighted Index [11] [*] offers a pre-calculated compendium of per-residue flexibility propensities that were refined and optimized in a set of 12,216 target proteins from 196 different species that were expressed In *E. coli* using either a C- or N-terminal poly-histidine fusion tag. The strategy derives from the observation that, over almost 10,000 different studied protein properties, flexibility was the best predictor for solubility.

SoluProt [12] [*] is based on gradient boosting regression and provides solubility prediction from the protein sequence. The machine learning model has been developed using a manually curated TargetTrack database. Considering the amino-acid singlet and dimer content of the poly-peptidic chain, their physicochemical properties, membrane propensity and similarity to *E. coli* 3D proteome, this

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 December 2020 doi:10.20944/preprints202012.0089.v1

approach achieves an AUC of 0.60 on a newly compiled independent set. SoluProt is integrated in EnzymeMiner [7], providing an easy way to filter out unlikely soluble proteins in the process of novel enzyme discovery.

4. Engineering enzyme activity and specificity

Enzyme activity and selectivity are the key features normally targeted in enzyme engineering. Although activity and selectivity are very different properties, they can often be improved using similar computational approaches. Engineering the activity towards a substrate of interest is likely to result in the improvement of the selectivity towards this substrate. The most common strategy consists of introducing mutations in the active site and optimizing it towards the targeted substrate. Other approaches have also proven successful, namely the engineering of access tunnels, modification of the dynamic properties, editing recognition elements such as loops, or targeting allosteric sites.

Important computational tools for engineering enzyme function − among which is the gold-standard Rosetta toolbox [42] − have been reviewed [43,44]. Rosetta-based web tool FuncLib [13] [**] was specially designed to add multiple-point mutations to the binding site. Taking into account evolutionary information and energetically favorable substitutions, single-point mutations are combined and ranked by the predicted stabilization free energies (ΔΔG). The FuncLib workflow ensures that no deleterious mutations are introduced and it can account for potential epistatic effects resulting from combining multiple mutations.

CaverDock [14], integrated into the Caver Web [30] [**] (section 7), can be used for engineering enzyme activity and selectivity. This tool was designed to predict the trajectory and binding energy profile of (un)binding of a ligand travelling through the enzyme access tunnels using a constrained molecular docking algorithm. The user can run calculations for different ligands or for multiple enzyme variants, and assess which combinations provide the best energy profiles. This is especially useful when the limiting steps in the catalysis involve the substrate binding or the product release.

Enzyme specificity can also be modified by engineering loops, which represent the flexible elements that can modulate substrate recognition and binding specificity. DaReUS-Loop [15] [**] (re)models loops in homology models and it can search the databases for new loop conformations suitable to be introduced in the target structure. This can help users find new enzyme variants with diverse substrate specificities. nAPOLI [16] automatically identifies conserved protein-ligand interactions across a large data set, such as a list of PDB structures or any protein within a specified range of sequence identity. It compiles the type of interactions and networks formed to find hotspots

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 December 2020 doi:10.20944/preprints202012.0089.v1

within the binding sites or suggest mutations that can produce more favorable interactions with a specific substrate.

5. Engineering protein stability

Enzyme stability refers to the range of temperature, co-solvents, pH, and other general conditions in which enzymes can resist and remain active. For many biotechnological purposes, it is desirable that the enzymes survive longer time or harsher conditions beyond what the native variants normally would. One can push those boundaries by engineering their stability using: (i) energy calculations, (ii) phylogenetic analysis, (iii) machine learning, and (iv) combination of the previous ones. These strategies [45–48] and software tools [41] have been extensively reviewed.

Ancestral sequence reconstruction (ASR) is a strategy that is becoming increasingly used for protein stabilization. FireProt-ASR [17] [*] is the first fully automated platform for inferring the ancestral sequences by phylogenetic analysis. Based on a single protein sequence, the tool builds a dataset of homology sequences and performs a multiple sequence alignment to build a phylogenetic tree and reconstruct the ancestral nodes. The method can be used not only to improve thermostability, but also to expand the catalytic promiscuity and increase expressibility of enzymes.

Electrostatic interactions are crucial to protein folding and integrity. They also rule the effects of pH and ion concentration on protein stability. However, they are often underestimated or poorly predicted during enzyme engineering. TKSA-MC [18] and pStab [19] tools tackle this issue by assessing unfavorable electrostatic interactions and identifying charged hot-spot residues for mutagenesis.

A very different approach is protein stabilization by introduction of disulfide bonds. Yosshi [21] [*] and SSBondPre [22] are recent tools devoted to this strategy, the former using evolutionary analysis and the latter using machine learning. Most of the stability prediction methods have been developed for globular soluble proteins. mCSM-membrane [23] can predict the stability changes or the pathogenicity associated with mutations in membrane proteins.

6. Engineering protein dynamics

Proteins exist in dynamic, metastable conformational states, transitioning through an ensemble of possible local conformations. The motions resulting from such transitions can fundamentally influence the catalytic activity of an enzyme [49,50]. Thus, assessing and engineering enzyme dynamics

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 December 2020 doi:10.20944/preprints202012.0089.v1

may be crucial to achieve a desired activity output. It also has an impact on predicting protein solubility and stability.

DynaMut2 [25] [*] combines Normal Mode Analysis methods and graph-based signatures to investigate the effects of single- and multiple-point mutations on protein stability and dynamics. The server reports B-factors that characterize the predicted flexibility of the mutants and changes in the stability. Moreover, the server offers the possibility to independently run coarse-grained predictions on a structure using five different force fields.

CABS [51] is a coarse-grained force field accounting for side chain contacts, main-chain hydrogen bond networks, and local geometric preferences. It was validated against molecular dynamics and nuclear magnetic resonance ensembles, and is part of AGGRESCAN 3D [8]. Freshly re-implemented in a web server CABS-flex 2.0 [26], it allows for evaluating larger proteins with up to 2000 residues, to impose user-defined distance restraints, and offers an improved graphical output.

ProSNEx [27] [*] models inter-residue interaction networks from the input 3D coordinates of the protein to be studied. Such contacts are weighted according to dynamical cross-correlation maps either obtained from elastic network models or other normal mode applications, the graph theory based spectral clustering of side chains, or molecular dynamic simulations derived energies. These dynamics studies are enriched with subsequent network and sequence conservation analysis, and the results are presented in an easy-to-interpret graphic-intensive interface.

AlloSigMA 2 [28] studies allostery and is based on the implementation of a structure-based statistical mechanical model. The server allows for evaluating the allosteric free energy resulting from the perturbation of any residue in the input structure. It allows for testing the allosteric effects of introducing mutations and the impact of introducing a ligand into the studied system. An intuitive graphical user interface provides a rapid interpretation of the protein regions that changed their dynamics.

LARMD [29] [*] automates the execution of fully atomistic molecular dynamics simulations up to 4 ns long. Untrained users can opt for the suggested easy-to-set-up predefined conditions and more versed ones can fine-tune the execution parameters to their needs. The application is focused on deciphering the structural and dynamical effects of ligand binding, and to this end implements tunnel discovery tools such as CAVER 3.0 [52]. Furthermore, it offers a wide range of analyses on the obtained trajectories: (i) structural variability and fluctuation analyses, (ii) normal mode analysis, and (iii) trajectory clustering. The server provides a wide range of graphics and charts to ease the interpretation of the results.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 December 2020 doi:10.20944/preprints202012.0089.v1

7. Engineering multiple properties

Some protein engineering web-tools integrate multiple tasks in robust workflows. Caver Web [30] [**] can be used to identify molecular tunnels and channels in proteins with buried cavities and predict the transport of ligands through these tunnels (Figure 2). The workflow starts with the identification of the relevant pockets and computing the tunnels from the selected pocket to the surface using CAVER 3.0 [52]. The user then selects the tunnels and ligands for analysis of the transport using CaverDock (section 4). This integrated analysis allows identifying hotspots on the enzyme tunnels that can remove the barriers to the transport of the target substrates or products or increase their specificity, and thus improve the enzymatic function.

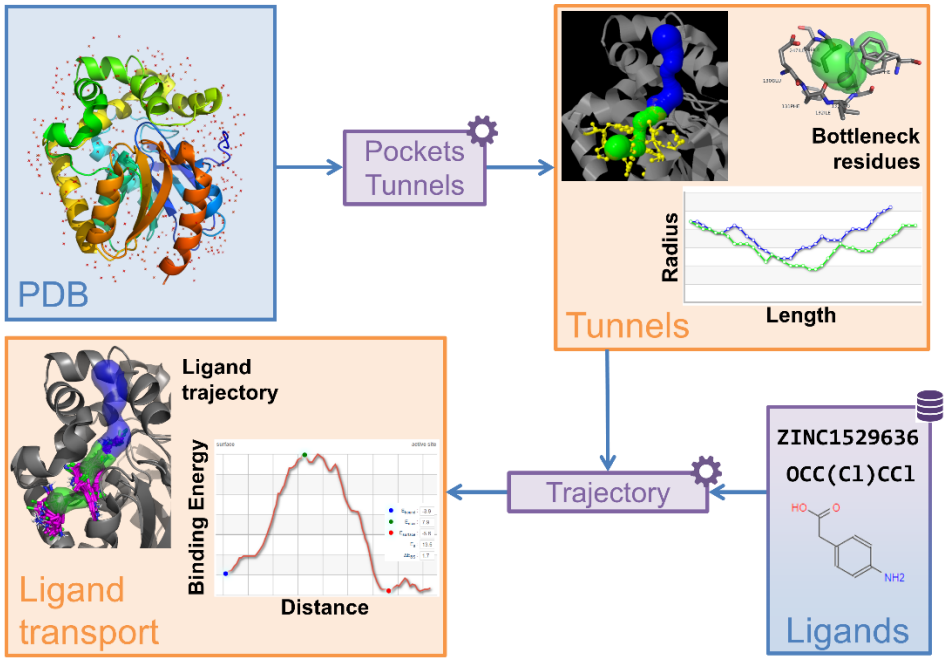

Figure 2. Illustration of the Caver Web workflow [30] [**]. The user enters a PDB file or PDB code. The pockets in the 3D structure are calculated and one of them is used as a starting point to calculate the tunnels to the surface. The identified tunnels can be analyzed for their properties, bottleneck residues and tunnel-lining residues. The user can enter one or multiple ligands as files, drawing, SMILES or ZINC codes, and calculate their trajectories through the selected tunnels. The user can analyze the binding energy profiles of the ligand, determine energy minima, maxima and energy barriers. The ligand trajectory and the list of bottleneck residues forming the energy barriers can be downloaded.

HotSpot Wizard 3 [31] [*] is a tool for identification of mutagenesis hotspots for improving stability, activity, and specificity, following a multi-stage automatic workflow (Figure 3). The tool sequentially calculates several parameters in order to identify: i) functional hotspots located in the active site pocket and/or access tunnels, ii) stability hotspots corresponding to flexible residues, (iii) stability

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 December 2020 doi:10.20944/preprints202012.0089.v1

hotspots based on back-to-consensus and (iv) correlated hotspots corresponding to co-evolving residue pairs. Recent updates have made possible the calculation of homology models from the protein sequence. The user can build smart libraries based on the amino acid frequencies, predict the stabilization energy of selected mutations, and even combine the interesting ones into multiple-point mutations.

ROSIE [32] [*] is probably the multipurpose web platform with the widest scope. Developed in 2013 and updated recently, it hosts under the same roof dozens of tools from the Rosetta family (currently 27) for modelling and designing proteins, nucleic acids, and other biopolymers. In this uniform and friendly environment, enzyme engineers can perform, for example, molecular docking (*Ligand Docking*), predict and design stability (*Sequence Tolerance*, *VIP*), and improve solubility (*Supercharge*). Other Rosetta-based tools useful for enzyme design, e.g. *RosettaDesign* and *Robetta* [42], are also available as web servers.

ProteinsPlus [33] [*] is a unified platform integrating multiple tasks of protein investigation, namely database exploration, structural quality assessment, conformational analysis, binding site analysis, 2D-interaction diagrams, pocket detection, etc. Although it is not devoted to enzyme engineering *per se*, it can provide comprehensive structural knowledge. pPerturb [34] aims primarily at assessing the importance of different residues to the stability by analyzing the effects of alanine mutations on the global number of contacts in the structure. The workflow is divided into perturbation profiles calculation (ΔQ), interaction networks, and the change in thermodynamic stability from truncating side chains. Overall, the tool can facilitate identifying residues that determine local stability and potential allosteric signal transduction paths.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 December 2020 doi:10.20944/preprints202012.0089.v1

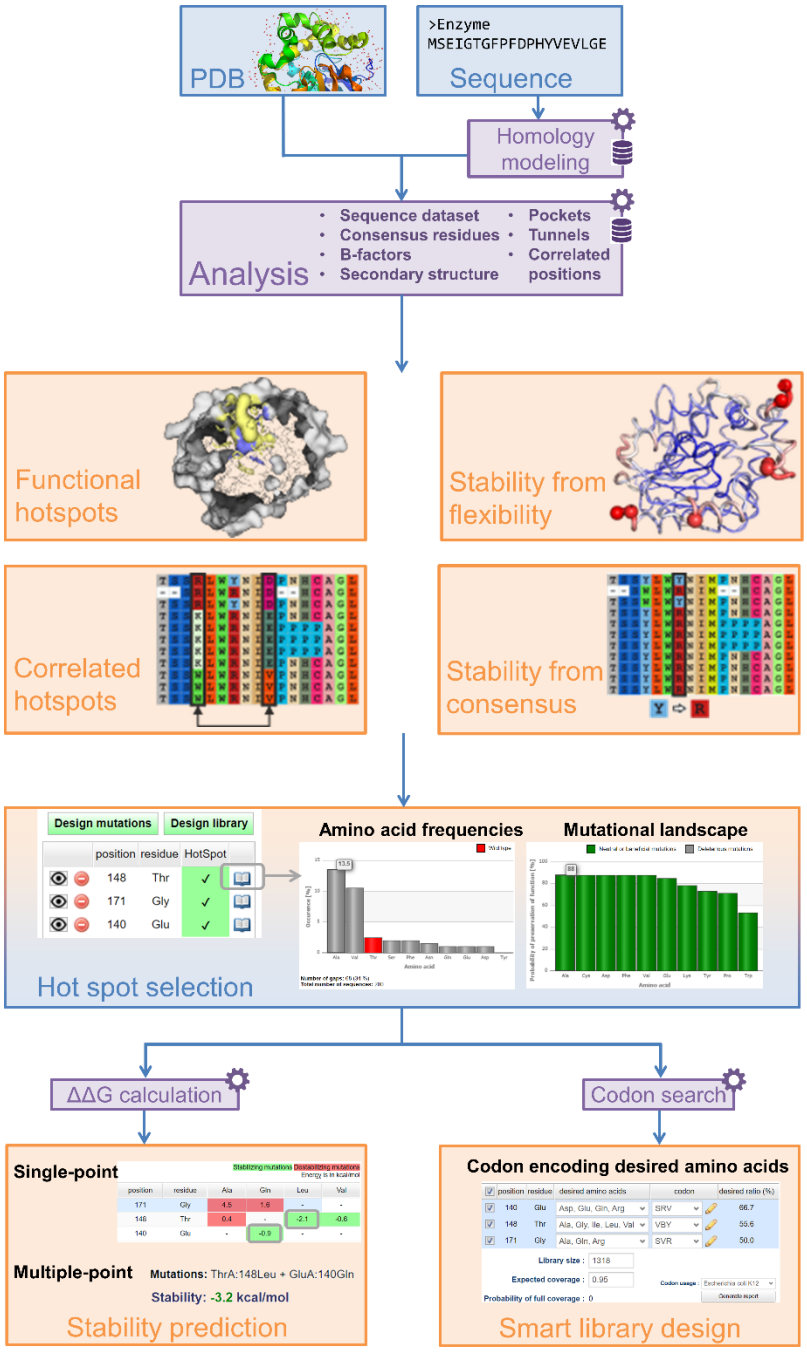

Figure 3. Illustration of the HotSpot Wizard 3 workflow [31] [*]. The user enters a PDB structure or a sequence that will be used to predict the structure by homology modelling. A sequence of different calculations are performed, leading to four types of hot-pot predictions: (i) functional hotspots (non-essential residues located on functional pockets or tunnels, ranked by mutability), (ii) correlated hot-spots (co-evolving pairs of residues, obtained from consensus and correlation analysis), (iii) stability from flexibility (hot-spots with higher B-factors), and (iv) stability from consensus (hot-spots recommended to be mutated to amino acids with higher frequency in the multiple sequence alignment). The user can select the hot-spots for mutagenesis based on the integrated overview of the suggested positions, such as mutability, secondary structure, amino acid frequency and mutational landscape. The user can predict the stabilization (ΔΔG) from all the selected single-point mutations on the selected hot-spots, and combine them into multiple-point mutations. The user can also calculate the optimal DNA codon content to build smart libraries for screening the selected positions with the desired set of amino acids.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 December 2020 doi:10.20944/preprints202012.0089.v1

8. Conclusions and perspectives

Here we reviewed the recently published web-based tools specialized in different aspects of enzyme engineering, which can be valuable resources to experimental scientists. The advantages of web-based tools are their immediate use without tedious installations, optimal settings already selected by the developers, regular updates and maintenance, and shared computational resources. We observe a boom of new methods and approaches, especially the rise of predictors based on machine learning, for which the quality of the experimental data used for training is of paramount importance. However, it is not always guaranteed by the available databases, which would highly benefit from stronger efforts of the community to supply high-quality, findable, annotated and curated data. These data will also provide essential input for machine learning as well as critical comparisons of newly developed tools. Modern high-throughput experimental technologies like fluorescent activated cell sorting, microfluidics, cell-free expression and deep mutational scanning will enable the collection of large and highly consistent data sets.

We observed a large number of tools devoted to enzyme discovery, although mainly focused on predicting the potential enzymatic activity of a protein sequence, but not for retrieving potential catalysts from a collection of orphan proteins. We also see a shift in the strategies for engineering activity and specificity, as many recent tools focus on non-active site elements, e.g. loops, tunnels, highly flexible and allosteric regions. In general, engineering catalytic activity, selectivity and protein solubility are insufficiently developed and more reliable tools are needed to provide practically useful predictions. With the constant increase of computational power, which allows more robust assessment of structural ensembles, we expect protein dynamics to become a more integral part of the next generation tools. We predict the same should happen with the design of catalytic activity using high-level methods, i.e. quantum mechanics or hybrid quantum mechanics/molecular dynamics, to be made accessible via web servers. We have witnessed a game-changing situation with the development of GPU cards and their use for computationally demanding tasks. We envisage another major breakthrough with gradually maturing quantum computing technologies.

Acknowledgements

The authors would like to express their thanks to the Czech Ministry of Education (CZ.02.1.01/0.0/0.0/16_026/0008451, CZ.02.1.01/0.0/0.0/17_043/0009632, LM2018121, LM2015047, LM2018140), the Technology Agency of the Czech Republic (TN01000013). This project has received funding from the European Union's Horizon 2020 research and Innovation programme under grant

agreement 814418 and 857560. The article reflects the author's view and the Agency is not responsible for any use that may be made of the information it contains.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest

** of outstanding interest

1. Arnold FH: **Innovation by evolution: bringing new chemistry to life (Nobel Lecture)**. *Angewandte Chemie International Edition* 2019, **58**:14420–14426.

2. *Protein Engineering Market - Global Forecast to 2024*. MarketsandMarkets; 2020.

3. Memon SA, Khan KA, Naveed H: **HECNet: a hierarchical approach to enzyme function classification using a siamese triplet network**. *Bioinformatics* 2020, doi:10.1093/bioinformatics/btaa536.

4. Zhang T, Tian Y, Yuan L, Chen F, Ren A, Hu Q-N: **Bio2Rxn: sequence-based enzymatic reaction predictions by a consensus strategy**. *Bioinformatics* 2020, **36**:3600–3601.

* Predicts the EC class (up to its fourth level) of the input sequence as per the consensus prediction of six different predictive methods, which combine deep learning, sequence pattern discovery and per residue biochemical properties.

5. Angles R, Arenas-Salinas M, García R, Reyes-Suarez JA, Pohl E: **GSP4PDB: a web tool to visualize, search and explore protein-ligand structural patterns**. *BMC Bioinformatics* 2020, **21**:85.

6. Toti D, Viet Hung L, Tortosa V, Brandi V, Polticelli F: **LIBRA-WA: a web application for ligand binding site detection and protein function recognition**. *Bioinformatics* 2018, **34**:878–880.

7. Hon J, Borko S, Stourac J, Prokop Z, Zendulka J, Bednar D, Martinek T, Damborsky J: **EnzymeMiner: automated mining of soluble enzymes with diverse structures, catalytic properties and stabilities**. *Nucleic Acids Res* 2020, **48**:W104–W109.

** Identifies proteins that potentially perform the inferred enzymatic function from input sequences, ensures essential amino-acids profile in the result hits, and allows to prioritize them by several properties, including solubility, extremophilicity, source organism or structure availability.

8. Kuriata A, Iglesias V, Pujols J, Kurcinski M, Kmiecik S, Ventura S: **Aggrescan3D (A3D) 2.0: prediction and engineering of protein solubility**. *Nucleic Acids Res* 2019, **47**:W300–W307.

** Estimates protein aggregation that in its newest version can be applied to significantly larger and multimeric proteins, simultaneous predicts changes in protein solubility and stability upon mutation. The tool provides a REST-ful service to incorporate calculations in automatic pipelines.

9. Hou Q, Kwasigroch JM, Rooman M, Pucci F: **SOLart: a structure-based method to predict protein solubility and aggregation**. *Bioinformatics* 2020, **36**:1445–1452.

10. Rawat P, Prabakaran R, Kumar S, Gromiha MM: **AggreRATE-Pred: a mathematical model for the prediction of change in aggregation rate upon point mutation**. *Bioinformatics* 2020, **36**:1439–1444.

11. Bhandari BK, Gardner PP, Lim CS: **Solubility-Weighted Index: fast and accurate prediction of protein solubility**. *Bioinformatics* 2020, doi:10.1093/bioinformatics/btaa578.

* Describes the study of a plethora of protein properties which lead to the observation that flexibility was the best predictor for solubility. Existing flexibility statistical pseudo-potentials were optimized and implemented as an easy-to-use web server.

12. Hon J, Marusiak M, Martinek T, Zendulka J, Bednar D, Damborsky J: **SoluProt: prediction of soluble expression from protein sequence**. *Bioinformatics* 2020, **Accepted**.

* Predicts expressibility and solubility of proteins in *Escherichia coli* using gradient boosted regression model developed using the manually curated TargetTrack database. The machine learning model is provided via the user-friendly web server and is also used for prioritization of hits in EnzymeMiner.

13. Khersonsky O, Lipsh R, Avizemer Z, Ashani Y, Goldsmith M, Leader H, Dym O, Rogotner S, Trudeau DL, Prilusky J, et al.: **Automated design of efficient and functionally diverse enzyme repertoires**. *Molecular Cell* 2018, **72**:178-186.e5.

** Predicts multiple-point mutations that can either modify the enzyme specificity or increase activity towards a given substrate. The single-point mutations adding energy improvements are combined and scored by Rosetta energy calculations after filtering out non-conserved mutations.

14. Vavra O, Filipovic J, Plhak J, Bednar D, Marques SM, Brezovsky J, Stourac J, Matyska L, Damborsky J: **CaverDock: a molecular docking-based tool to analyse ligand transport through protein tunnels and channels**. *Bioinformatics* 2019, **35**:4986–4993.

15. Karami Y, Rey J, Postic G, Murail S, Tufféry P, de Vries SJ: **DaReUS-Loop: a web server to model multiple loops in homology models**. *Nucleic Acids Res* 2019, **47**:W423–W428.

** Describes the development of the tool for the modeling of single or multiple loop regions in the context of homology models and assessment of their reliability. The alternative conformations are obtained from the Protein Data Bank.

16. Fassio AV, Santos LH, Silveira SA, Ferreira RS, de Melo-Minardi RC: **nAPOLI: a graph-based strategy to detect and visualize conserved protein-ligand interactions in large-scale**. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 2020, **17**:1317–1328.

17. Musil M, Khan R, Stourac J, Bednar D, Damborsky J: **FireProtASR: a web server for fully automated ancestral sequence reconstruction**. *Brief Bioinform* 2020, **Accepted**.

* Describes the first fully automated web server for ancestral sequence reconstruction, providing the results in a user-friendly interactive interface. No prior knowledge of the bioinformatics tools or the ancestral sequence reconstruction method is required.

18. Contessoto VG, Oliveira VM de, Fernandes BR, Slade GG, Leite VBP: **TKSA-MC: A web server for rational mutation through the optimization of protein charge interactions**. *Proteins: Structure, Function, and Bioinformatics* 2018, **86**:1184–1188.

19. Gopi S, Devanshu D, Krishna P, Naganathan AN: **pStab: prediction of stable mutants, unfolding curves, stability maps and protein electrostatic frustration**. *Bioinformatics* 2018, **34**:875–877.

20. Banerjee A, Mitra P: **Estimating the effect of single-point mutations on protein thermodynamic stability and analyzing the mutation landscape of the p53 protein**. *J Chem Inf Model* 2020, **60**:3315–3323.

21. Suplatov D, Timonina D, Sharapova Y, Švedas V: **Yosshi: a web-server for disulfide engineering by bioinformatic analysis of diverse protein families**. *Nucleic Acids Res* 2019, **47**:W308–W314.

* Designs disulfide bonds based on a systematic homology-driven analysis, facilitating a broader interpretation of disulfides. Allows adding disulfide bonds not only for structural stability but also as a mechanism to address functional diversity within a superfamily.

22. Gao X, Dong X, Li X, Liu Z, Liu H: **Prediction of disulfide bond engineering sites using a machine learning method**. *Scientific Reports* 2020, **10**:10330.

23. Pires DEV, Rodrigues CHM, Ascher DB: **mCSM-membrane: predicting the effects of mutations on transmembrane proteins**. *Nucleic Acids Res* 2020, **48**:W147–W153.

24. Qi Y, Zhang JZH: **DenseCPD: improving the accuracy of neural-network-based computational protein sequence design with DenseNet**. *J Chem Inf Model* 2020, **60**:1245–1252.

25. Rodrigues CH, Pires DE, Ascher DB: **DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability**. *Nucleic Acids Res* 2018, **46**:W350–W355.

* Combines two different methods of normal mode analysis for capturing the principal motions of the studied protein with graph-based signatures for representing the wild-type environment to accurately predict the effects of mutations on protein stability and dynamics.

26. Kuriata A, Gierut AM, Oleniecki T, Ciemny MP, Kolinski A, Kurcinski M, Kmiecik S: **CABS-flex 2.0: a web server for fast simulations of flexibility of protein structures**. *Nucleic Acids Res* 2018, **46**:W338–W343.

27. Aydınkal RM, Serçinoğlu O, Ozbek P: **ProSNEx: a web-based application for exploration and analysis of protein structures using network formalism**. *Nucleic Acids Res* 2019, **47**:W471–W476.

* Generates protein structure networks from an input structure and weights the network edges according to dynamic cross-correlation values, calculated from elastic network models or obtained from molecular dynamics. Provides a comprehensive analysis of the predicted motions of a protein.

28. Tan ZW, Guarnera E, Tee W-V, Berezovsky IN: **AlloSigMA 2: paving the way to designing allosteric effectors and to exploring allosteric effects of mutations**. *Nucleic Acids Res* 2020, **48**:W116–W124.

29. Yang J-F, Wang F, Chen Y-Z, Hao G-F, Yang G-F: **LARMD: integration of bioinformatic resources to profile ligand-driven protein dynamics with a case on the activation of estrogen receptor**. *Brief Bioinform* 2019, doi:10.1093/bib/bbz141.

* Automates the execution of short molecular dynamics simulations for analysis of ligand transport via tunnels calculated by Caver. Non-expert users can easily set up the simulations by accepting the default settings while more experienced ones can adjust calculation parameters.

30. Stourac J, Vavra O, Kokkonen P, Filipovic J, Pinto G, Brezovsky J, Damborsky J, Bednar D: **Caver Web 1.0: identification of tunnels and channels in proteins and analysis of ligand transport**. *Nucleic Acids Res* 2019, **47**:W414–W422.

** Integrates the calculation of protein cavities, molecular tunnels, and the trajectories of ligand moving through those tunnels. The result is a deep knowledge of the tunnel properties, tunnel residues, and energetic maxima and minima for the ligand transport.

31. Sumbalova L, Stourac J, Martinek T, Bednar D, Damborsky J: **HotSpot Wizard 3.0: web server for automated design of mutations and smart libraries based on sequence input information**. *Nucleic Acids Res* 2018, **46**:W356–W362.

* Combines structural and sequence information to identify mutagenesis hotspots, based on functional features, conservation analysis, residue correlations, etc. Predicts stabilizing mutations using B-factors and back-to-consensus. Enables rational design of point mutations using Rosetta calculations as well as construction of smart libraries for directed evolution.

32. Moretti R, Lyskov S, Das R, Meiler J, Gray JJ: **Web-accessible molecular modeling with Rosetta: The Rosetta Online Server that Includes Everyone (ROSIE)**. *Protein Science* 2018, **27**:259–268.

* Provides a common environment for hosting a large number of web-accessible Rosetta protocols for modeling and designing proteins and other biopolymers. This new implementation simplified the submission process and extended the number of protocols available.

33. Schöning-Stierand K, Diedrich K, Fährrolfes R, Flachsenberg F, Meyder A, Nittinger E, Steinegger R, Rarey M: **ProteinsPlus: interactive analysis of protein–ligand binding interfaces**. *Nucleic Acids Res* 2020, **48**:W48–W53.

* Supports the first steps of analyzing a protein structure, with special focus on the interaction with ligands, in a comparative manner. The new version integrates an extended range of modeling tools.

34. Gopi S, Devanshu D, Rajasekaran N, Anantakrishnan S, Naganathan AN: **pPerturb: a server for predicting long-distance energetic couplings and mutation-induced stability changes in proteins via perturbations**. *ACS Omega* 2020, **5**:1142–1146.

35. Sayers EW, Agarwala R, Bolton EE, Brister JR, Canese K, Clark K, Connor R, Fiorini N, Funk K, Hefferon T, et al.: **Database resources of the National Center for Biotechnology Information**. *Nucleic Acids Res* 2019, **47**:D23–D28.

36. UniProt Consortium: **UniProt: a worldwide hub of protein knowledge**. *Nucleic Acids Res* 2019, **47**:D506–D515.

37. Vanacek P, Sebestova E, Babkova P, Bidmanova S, Daniel L, Dvorak P, Stepankova V, Chaloupkova R, Brezovsky J, Prokop Z, et al.: **Exploration of enzyme diversity by integrating bioinformatics with expression analysis and biochemical characterization**. *ACS Catal* 2018, **8**:2402–2412.

38. Rose PW, Prlić A, Altunkaya A, Bi C, Bradley AR, Christie CH, Costanzo LD, Duarte JM, Dutta S, Feng Z, et al.: **The RCSB protein data bank: integrative view of protein, gene and 3D structural information**. *Nucleic Acids Res* 2017, **45**:D271–D281.

39. Furnham N, Holliday GL, de Beer TAP, Jacobsen JOB, Pearson WR, Thornton JM: **The Catalytic Site Atlas 2.0: cataloging catalytic sites and residues identified in enzymes**. *Nucleic Acids Res* 2014, **42**:D485-489.

40. Planas-Iglesias J, Marques SM, Pinto G, Musil M, Stourac J, Bednar D, Damborsky J: **Computational design of enzymes for biotechnological applications**. *Biotechnol Adv* 2020, **Under review**.

41. Musil M, Konegger H, Hon J, Bednar D, Damborsky J: **Computational design of stable and soluble biocatalysts**. *ACS Catal* 2019, **9**:1033–1054.

42. Bender BJ, Cisneros A, Duran AM, Finn JA, Fu D, Lokits AD, Mueller BK, Sangha AK, Sauer MF, Sevy AM, et al.: **Protocols for molecular modeling with Rosetta3 and RosettaScripts**. *Biochemistry* 2016, **55**:4748–4763.

43. Sequeiros-Borja CE, Surpeta B, Brezovsky J: **Recent advances in user-friendly computational tools to engineer protein function**. *Brief Bioinform* 2020, doi:10.1093/bib/bbaa150.

44. Weinstein J, Khersonsky O, Fleishman SJ: **Practically useful protein-design methods combining phylogenetic and atomistic calculations**. *Current Opinion in Structural Biology* 2020, **63**:58–64.

45. Goldenzweig A, Fleishman SJ: **Principles of protein stability and their application in computational design**. *Annu Rev Biochem* 2018, **87**:105–129.

46. Kazlauskas R: **Engineering more stable proteins**. *Chem Soc Rev* 2018, **47**:9026–9045.

 doi:10.20944/preprints202012.0089.v1

47. Marabotti A, Scafuri B, Facchiano A: **Predicting the stability of mutant proteins by computational approaches: an overview**. *Brief Bioinform* 2020, doi:10.1093/bib/bbaa074.

48. Mazurenko S: **Predicting protein stability and solubility changes upon mutations: data perspective**. *ChemCatChem* 2020, **12**:5590–5598.

49. Schafer JW, Schwartz SD: **Directed evolution's influence on rapid density fluctuations illustrates how protein dynamics can become coupled to chemistry**. *ACS Catal* 2020, **10**:8476–8484.

50. Campbell E, Kaltenbach M, Correy GJ, Carr PD, Porebski BT, Livingstone EK, Afriat-Jurnou L, Buckle AM, Weik M, Hollfelder F, et al.: **The role of protein dynamics in the evolution of new enzyme function**. *Nat Chem Biol* 2016, **12**:944–950.

51. Jamroz M, Orozco M, Kolinski A, Kmiecik S: **Consistent view of protein fluctuations from all-atom molecular dynamics and coarse-grained dynamics with knowledge-based force-field**. *J Chem Theory Comput* 2013, **9**:119–125.

52. Chovancová E, Pavelka A, Benes P, Strnad O, Brezovsky J, Kozlikova B, Gora A, Sustr V, Klvana M, Medek P, et al.: **CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures**. *PLoS Comput Biol* 2012, **8**:e1002708.