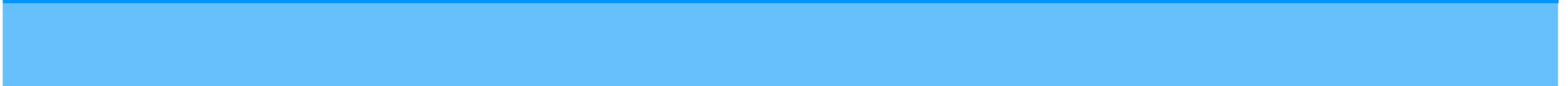


LOSCHMIDT
LABORATORIES



Engineering of protein structures



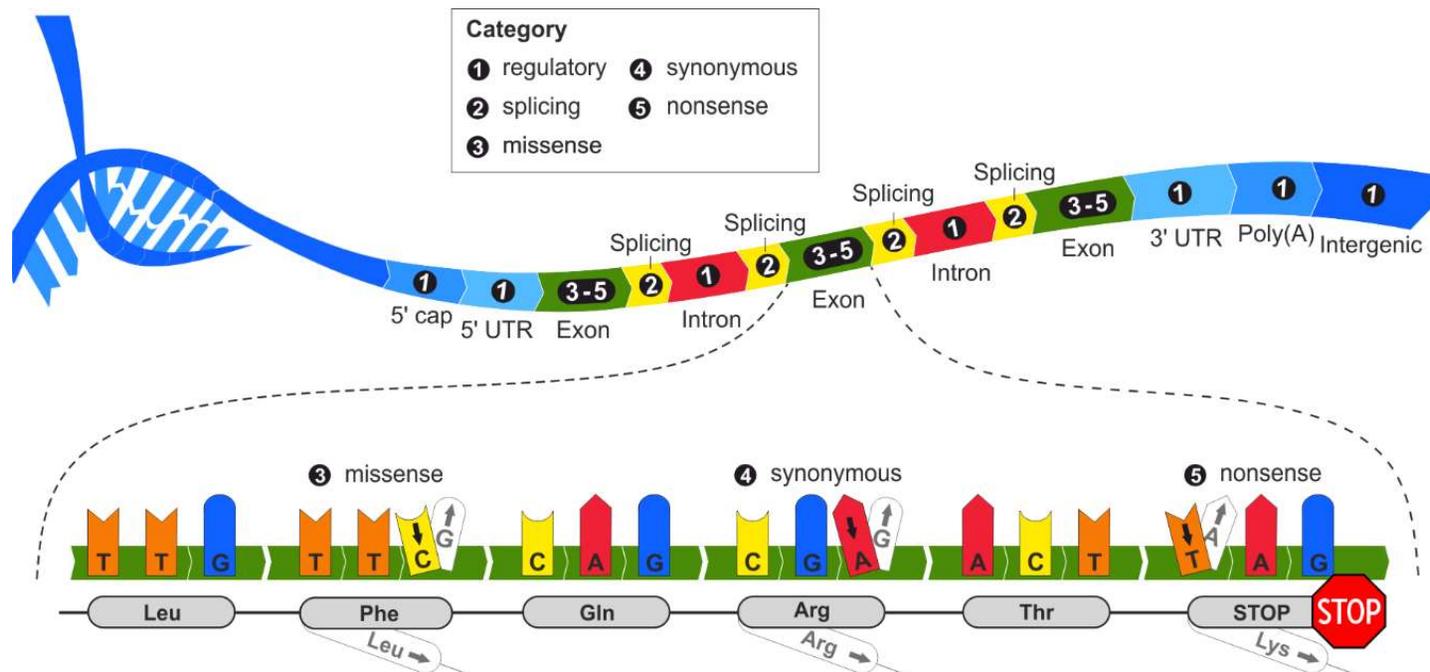
Outline

- ❑ Overview of mutations
- ❑ Databases of mutations
- ❑ Missense mutations
- ❑ Prediction of mutational effects
- ❑ Rational design of proteins

Overview of mutations

□ Location in the DNA

- non-coding region → affects protein expression (transcriptional regulation, mRNA stability, translation location, translation rates, etc.)
- **coding region → may affect protein sequence**



Overview of mutations



□ Types

- point mutation – a single nucleotide is changed in DNA or RNA
 - **substitutions**
 - **single nucleotide polymorphism (SNP)**
 - genetic variation; occurs in > 1 % of population
 - about 10,000,000 in the human genome
 - **insertions or deletions**
 - codons have triple nature (3 nucleotides → 1 amino acid)
 - potential for frameshift (change in the grouping of codons, resulting in a different translation)
 - can be very deleterious

Point mutations at protein level



□ Types of point mutations

- **silent** (synonymous SNP) – no effect on protein sequence

| | | | | |
|----------|-----|-----|-----|-------------------|
| | L | Q | T | ← protein seq. |
| normal: | ctg | cag | act | ← nucleotide seq. |
| | | * | | ← mutation |
| mutated: | ctg | caa | act | |
| | L | Q | T | |

- **missense** (non-synonymous SNP) – substitution of amino acid

| | | | | |
|----------|-----|-----|-----|-------------------|
| | L | Q | T | ← protein seq. |
| normal: | ctg | cag | act | ← nucleotide seq. |
| | | * | | ← mutation |
| mutated: | ctg | cgg | act | |
| | L | R | T | |

- **nonsense** – introduction of a stop codon -> protein truncation

| | | | | |
|----------|-----|-----|-----|-------------------|
| | L | Q | T | ← protein seq. |
| normal: | ctg | cag | act | ← nucleotide seq. |
| | | * | | ← mutation |
| mutated: | ctg | tag | act | |
| | L | *** | | |

Databases of mutations



- ❑ **Human Genome Variation Society**
 - <http://www.hgvs.org>
 - lists all the available databases of mutations
- ❑ **Central mutation databases (>20)**
 - substitutions in all genes
 - data mainly from literature
- ❑ **Locus-specific databases (about 700)**
 - substitutions in specific genes
 - typically manually annotated

Central mutation databases

❑ Database of Single Nucleotide Polymorphisms - dbSNP

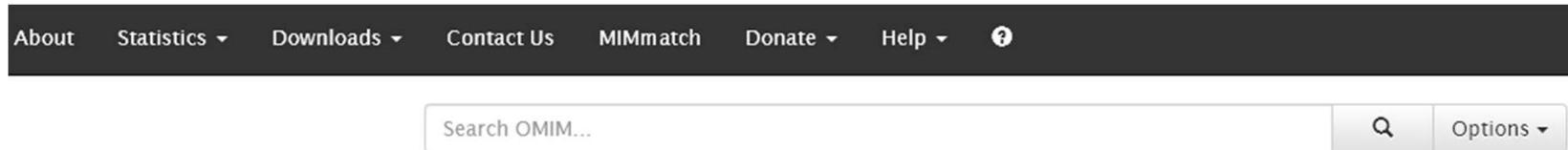
- <http://www.ncbi.nlm.nih.gov/SNP/>
- repository for both SNP and short deletion and insertion
- for human genome

The screenshot shows the NCBI dbSNP website. At the top, there is a navigation bar with links for PubMed, Nucleotide, Protein, Genome, Structure, PopSet, Taxonomy, OMIM, Books, and SNP. The main heading is "dbSNP Short Genetic Variations". Below this is a search bar with the text "Search for SNP on NCBI Reference Assembly". A search form is visible with "Search Entrez" and "SNP" entered, and a "Go" button. On the left side, there is a sidebar with a "Have a question about dbSNP? Try searching the SNP FAQ Archive!" section and a "Go" button. Below that is a "GENERAL HUMAN VARIATION" section with links for "Search, Annotate, Submit" and "Annotate and Submit Batch Data with Clinical Impact Attributes for Filtering Variation". Further down are sections for "SNP SUBMISSION DOCUMENTATION SEARCH" and "RELATED SITES". The main content area features an "ANNOUNCEMENT" box dated 09/20/2012 regarding the release of dbSNP Mouse_10090 and Cow_9913 data. Below the announcement is a "Search by IDs on All Assemblies" section with a note that "rs#" and "ss#" must be prefixed with "rs" or "ss", respectively. It includes a search form with "ID:" and "Reference cluster ID(rs#)" fields and "Search" and "Reset" buttons. The "Submission Information" section lists links for "By Submitter", "New Submitted Batches", "Method", "Population", and "Publication". The "Batch" section lists options for "Enter List" (with sub-links for NCBI Assay ID(ss), Reference SNP ID(rs), and Local SNP ID) and "Upload List" (with sub-links for NCBI Assay ID(ss), Reference SNP ID(rs), and Local SNP ID). A "Batch Query Help" link is located at the bottom right of the page.

Central mutation databases

❑ Online Mendelian Inheritance in Man – OMIM

- <http://omim.org/>
- comprehensive database of human genes and genetic phenotypes



OMIM Entry Statistics

Number of Entries in OMIM (Updated December 9th, 2020) :

| MIM Number Prefix | Autosomal | X Linked | Y Linked | Mitochondrial | Totals |
|---|-----------|----------|----------|---------------|--------|
| Gene description * | 15,554 | 744 | 51 | 37 | 16,386 |
| Gene and phenotype, combined + | 30 | 0 | 0 | 0 | 30 |
| Phenotype description, molecular basis known # | 5,565 | 349 | 5 | 33 | 5,952 |
| Phenotype description or locus, molecular basis unknown % | 1,414 | 115 | 4 | 0 | 1,533 |
| Other, mainly phenotypes with suspected mendelian basis | 1,660 | 103 | 3 | 0 | 1,766 |
| Totals | 24,223 | 1,311 | 63 | 70 | 25,667 |

Central mutation databases



❑ UniprotKB/Swiss-Prot

- <http://www.uniprot.org/>
- high-quality manually annotated protein entries with partial lists of known sequence variants

The screenshot shows the UniProt website interface. At the top, there are navigation tabs: Search, Blast, Align, Retrieve, and ID Mapping. Below these is a search bar with a dropdown menu set to 'Protein Knowledgebase (UniProtKB)' and a 'Query' input field. There are buttons for 'Search', 'Advanced Search >', and 'Clear'.

WELCOME

The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

What we provide

| | |
|-----------------|--|
| UniProtKB | Protein knowledgebase, consists of two sections: <ul style="list-style-type: none">★ Swiss-Prot, which is manually annotated and reviewed.★ TrEMBL, which is automatically annotated and is not reviewed. Includes complete and reference proteome sets. |
| UniRef | Sequence clusters, used to speed up sequence similarity searches. |
| UniParc | Sequence archive, used to keep track of sequences and their identifiers. |
| Supporting data | Literature citations, taxonomy, keywords, subcellular locations, cross-referenced databases and more. |

Getting started

- Text search
- Sequence similarity searches (BLAST)
- Sequence alignments
- Batch retrieval
- Database identifier mapping (ID Mapping)

NEWS

UniProt release 2012_10 - Oct 31, 2012
CIA: on your Genome service

- > Statistics for UniProtKB: Swiss-Prot - TrEMBL
- > Forthcoming changes
- > News archives

Follow @uniprot (456 followers)

SITE TOUR

Learn how to make best use of the tools and data on this site.

PROTEIN SPOTLIGHT

branching out October 2012

Humans are unique. Whichever way you look at it. We can talk. We can write. We can build skyscrapers, make art, design weapons and be a general nuisance to many other life forms. About 2.5 million years ago however, our ancestors could not...

Central mutation databases



❑ Protein Mutant Database - PMD

- <http://pmd.ddbj.nig.ac.jp/~pmd/pmd.html>
- compilation of literature data providing information on functional and/or structural influences of mutation at a specific position
- useful for protein engineering purposes



Last modified: Mar 26, 2007

[Go to Japanese Page](#)

• [What is PMD ? \(see samples\)](#)

Compilations of protein mutant data are valuable as a basis for protein engineering. They provide information on what kinds of functional and/or structural influences are brought about by amino acid mutation at a specific position of protein. The Protein Mutant Database (PMD) that we are constructing covers natural as well as artificial mutants, including random and site-directed ones, for all proteins except members of the globin and immunoglobulin families. The PMD is based on literature, not on proteins. That is, each entry in the database corresponds to one article which may describe one, several or a number of protein mutants.

The PMD is based on the literature, not on proteins. That is, each entry in the database corresponds to one article, which contains several or a number of protein mutants. Each database entry is identified by a serial number and is defined as either natural or artificial, depending on the type of the mutation. For each entry the following items are recorded: "JOURNAL", "TITLE", "CROSS-REFERENCE", "PROTEIN", "N-TERMINAL", "CHANGE", "FUNCTION", "STRUCTURE", "STABILITY", etc. "CROSS-REFERENCE" indicates the code names of the protein given in other databases such as Protein Identification Resources (2). "N-TERMINAL" shows the N-terminal sequence of five amino acids which may help to show the unambiguous numbering of the sequence. "CHANGE" indicates the position and kind of mutations, such as amino acid substitution, insertion and deletion, denoted with a specific notation. Any functional or structural features ("FUNCTION", "STRUCTURE", "STABILITY", etc) observed in the mutant are described immediately after "CHANGE". Relative differences in activity and/or stability, in comparison with the wild-type protein, are indicated with symbols [-,], [=], [+] or [+ +]. Complete loss of activity is denoted as [0].

• [Sample of PMD entry](#)

• [Detailed Description of PMD](#)

• [Statistics of the release Mar 26, 2007](#)

○ Number of entries: 45,239

Locus-specific databases

□ for information on gene-specific databases

| | | |
|---|---|---|
| ATP-binding cassette, sub-family D (ALD), member 1 300371 ABO | X-linked adrenoleukodystrophy Database http://www.x-ald.nl | Ronald R.J.A. Wanders Lab. of Genetic Metabolic Diseases Academic Medical Ctr. Amsterdam, The Netherlands. |
| ABO blood group (transferase A, alpha 1-3-N-acetylgalactosaminyltransferase; transferase B, alpha 1-3-galactosyltransferase) 110300 ACAD8 | Blood Group Antigen Mutation Database http://www.ncbi.nlm.nih.gov/gv/mhcxsc/cgi?cmd=bgmu/home | Olga O. Blumenfeld Department of Biochemistry, Santosh Patnaik, Department of Cell Biology, Albert Einstein College of Medicine New York, NY. U.S.A |
| acyl-CoA dehydrogenase family, member 8 604773 ACADM | Innsbruck Metabolic Diseases Pages http://lovd.i-med.ac.at/home.php?select_db=ACAD8 | Barbara Lanthaler, Stefanie Kalb and Martina Witsch-Baumgartner |
| acyl-CoA dehydrogenase, C-4 to C-12 straight chain 607008 ACADSB | CCHMC - Human Genetics Mutation Database https://research.cchmc.org/LOVD/home.php?select_db=ACADM | Ammar Husami, Brian Richardson, Edita Freeman, Kerry Shooner, Thedia Jacobs and Theru A. Sivakumaran |
| acyl-CoA dehydrogenase, short/branched chain 600301 ACADVL | Innsbruck Metabolic Diseases Pages http://lovd.i-med.ac.at/home.php?select_db=ACADSB | Barbara Lanthaler, Stefanie Kalb and Martina Witsch-Baumgartner |
| acyl-CoA dehydrogenase, very long chain 609575 ACE2 | CCHMC - Human Genetics Mutation Database https://research.cchmc.org/LOVD/home.php?select_db=ACADVL | Ammar Husami, Brian Richardson, Edita Freeman, Kerry Shooner, Thedia Jacobs and Theru A. Sivakumaran |
| angiotensin I converting enzyme (peptidyl-dipeptidase A) 2 300335 ACHE | ACE2 database at LOVD http://www.LOVD.nl/ACE2 | Johan T. den Dunnen Leiden Univ. Med Centre (<i>acting</i>), <i>Curator vacancy</i> |
| acetylcholinesterase (Yt blood group) 100740 ACOT9 | Blood Group Antigen Mutation Database http://www.ncbi.nlm.nih.gov/gv/mhcxsc/cgi?cmd=bgmu/home | Olga O. Blumenfeld Department of Biochemistry, Santosh Patnaik, Department of Cell Biology, Albert Einstein College of Medicine New York, NY. U.S.A |
| acyl-CoA thioesterase 9 ACSL4 | ACOT9 database at LOVD http://www.LOVD.nl/ACOT9 | Johan T. den Dunnen Leiden Univ. Med Centre (<i>acting</i>), <i>Curator vacancy</i> |
| acyl-CoA synthetase long-chain family member 4 300157 ACTA1 | ACSL4 database at LOVD http://www.LOVD.nl/ACSL4 | Johan T. den Dunnen Leiden Univ. Med Centre (<i>acting</i>), <i>Curator vacancy</i> |
| actin, alpha 1, skeletal muscle 102610 | Laing Laboratory Skeletal muscle alpha-actin (ACTA1) http://acta1.waimr.uwa.edu.au/home.php?select_db=ACTA1 | Nigel Laing and Kristen Nowak |

Missense mutations



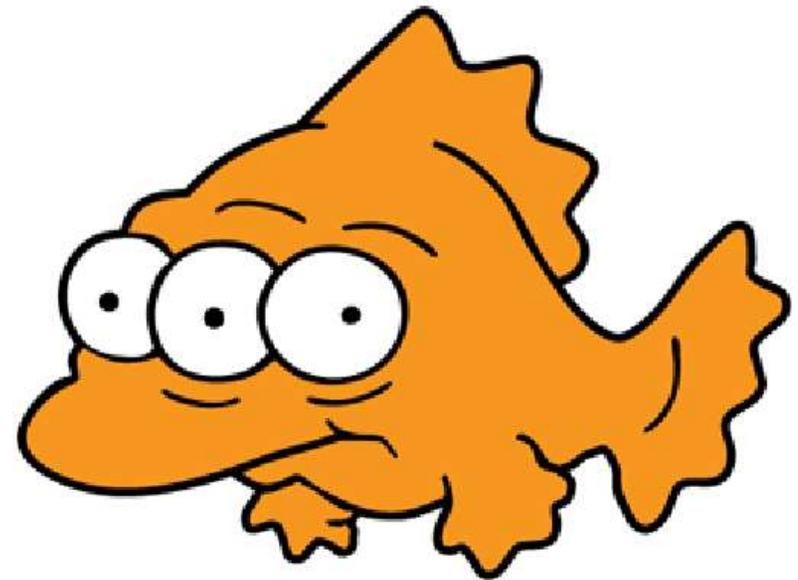
- ❑ Mutations affecting structure
 - stability & folding
 - aggregation

- ❑ Mutations affecting function
 - binding & catalysis
 - transport processes
 - protein dynamics
 - protein localization

Mutations affecting structure



- ❑ **Major pathogenic consequences of missense mutation**
 - compromised **folding** – the protein has modified folds or presents more unfolded states
 - decreased **stability** – the lifetime of the protein is decreased

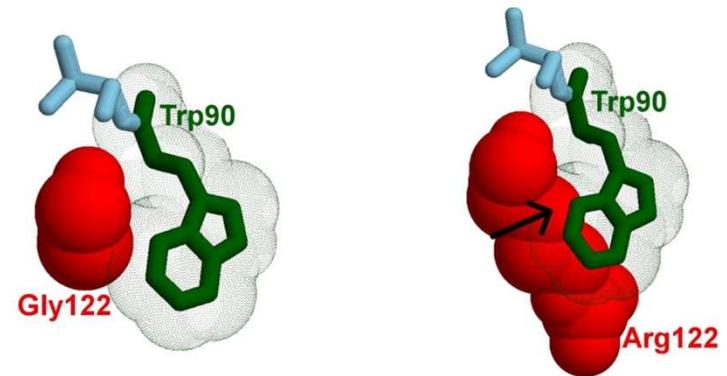


Mutations affecting structure

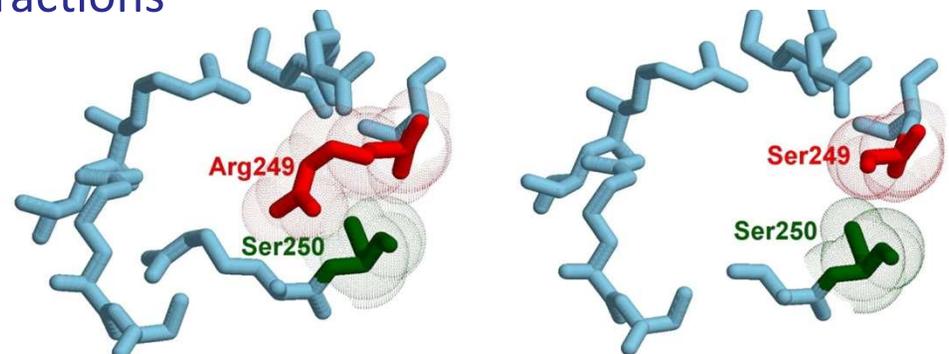


□ Molecular basis of mutations affecting folding & stability

- **introduced clashes** – common for small to large mutations in buried residues



- **loss of interactions** – most pronounced effects related to H-bonds, salt bridges and aromatic interactions

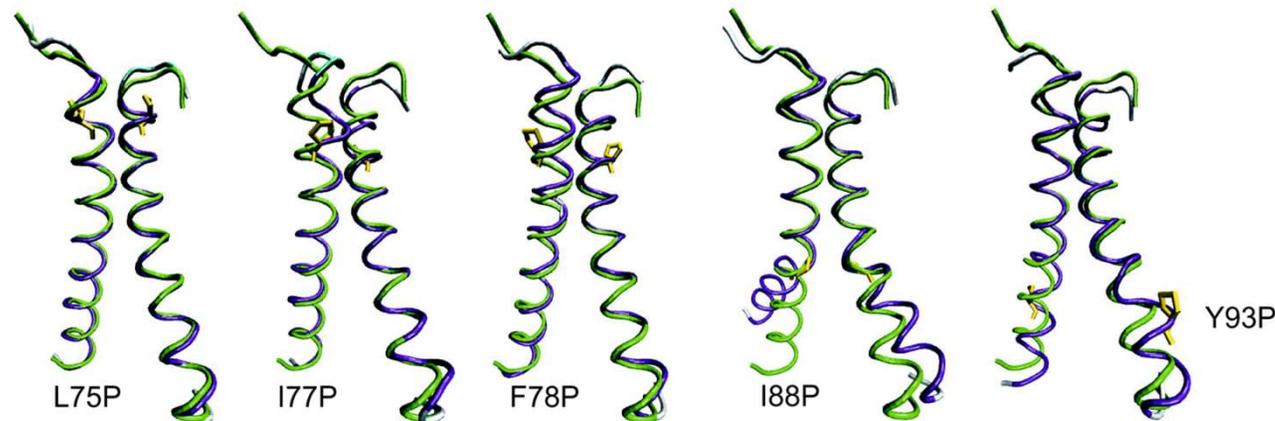


Mutations affecting structure



□ Molecular basis of mutations affecting folding & stability

- **altered conformation of protein backbone** – mutations concerning residues with specific backbone angles (glycine and proline)



- **changes in charge/hydrophobicity**
 - introducing hydrophilic/charged residue into the protein core
 - introducing hydrophobic residue onto the protein surface

Mutations affecting structure



- ❑ **Mutations can reduce solubility or increase aggregation**
 - alterations on the surface residues may affect the solubility (e.g. reduction of charge)
 - hydrophobic mutations can increase protein aggregation
 - aggregating proteins usually have high level of β -structures
- ❑ **Aggregation modulated by short specific sequences**
 - APR are sequences of 5-15 hydrophobic residues
 - they tend to stack and form amyloid fibrils (cross- β spines)
 - some mutations can increase the propensity to form such amyloid structures

Mutations affecting function

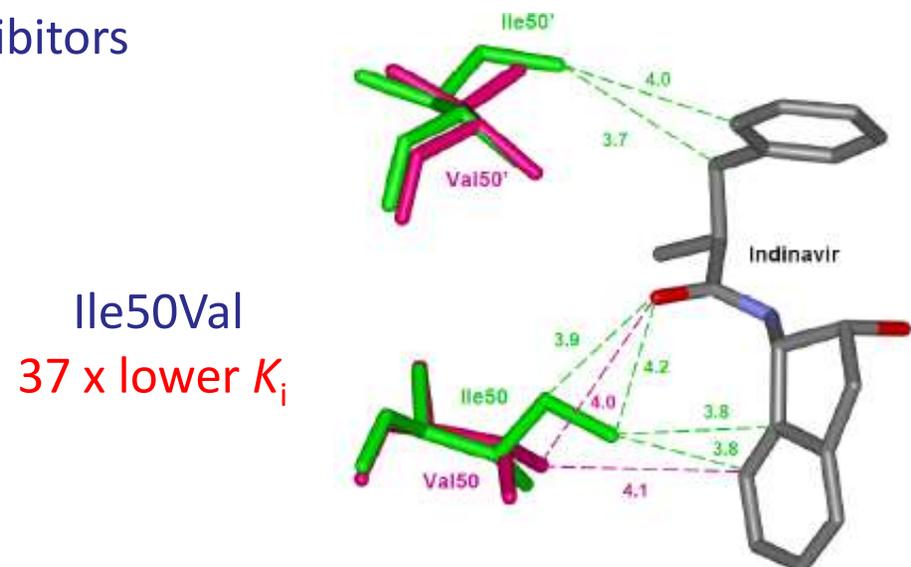
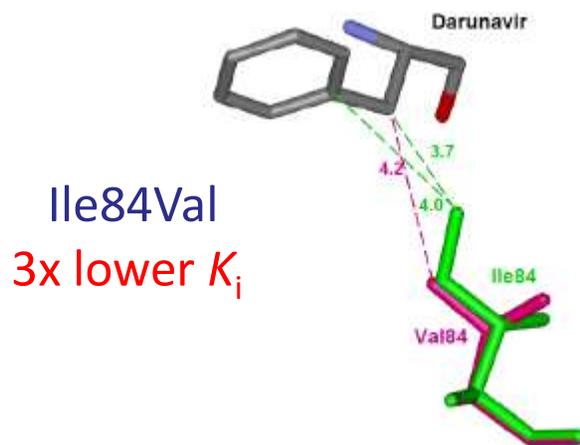


□ Effect on binding and catalysis

- binding and active sites are tuned to bind specific molecules and their transition states
- mutations can **improve or disrupt the binding and catalysis**

□ Example – drug-resistance of HIV-1 protease mutants

- loss of interactions with inhibitors



Mutations affecting function



❑ Effect on ligand transport

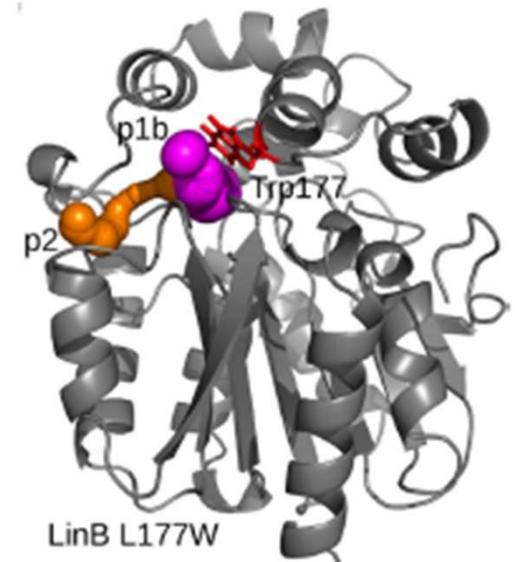
- pathways are adjusted to permit transport of specific molecules
- mutations can **speed-up or hinder** their transport or allow the transport of different molecules

❑ Example – transport in mutant of haloalkane dehalogenase



Leu177Trp in tunnel
p1a tunnel became closed

500x slower release of products



Mutations affecting function



□ Effect on protein dynamics

- dynamics enables proteins to adapt to their binding partners
- mutations can
 - make flexible regions more rigid (targeting hinge residues or very mobile ones) -> **reduced adaptability**
 - increase flexibility of rigid regions (targeting residues with many contacts on possibly mobile elements – i.e. loops)
-> **increased adaptability**
- these change may affect activity and specificity

Mutations affecting function



□ Effect on protein localization

- after translation protein must be translocated to the appropriate cellular compartment
- in many proteins, translocation is directed by short peptide sequences on the N-terminus acting as targeting signals
- mutations can **disrupt or alternate the signal** -> protein fails to be transported to the correct subcellular location
 - **missing protein** -> inactive reaction pathways or unregulated signaling cascades
 - **mislocalized protein** -> active in the wrong cellular compartment, causing harmful effects

Prediction of mutational effects



- ❑ Identification of mutable residues
- ❑ Prediction of the effects on structure
- ❑ Prediction of pathogenicity

Identification of mutable residues



- ❑ The effect of mutations on the protein can be directly assessed from the role of the modified residue

- ❑ **Mutation of evolutionary conserved residues**
 - residues important for protein function or stability tend to be **highly conserved** over evolution
 - mutation of highly conserved residues -> often lead to **destabilization or loss of function**
 - mutation of highly variable residues -> often **neutral**; can modulate function

Identification of mutable residues



❑ Mutations affecting stability & folding

- mutation of residues with many contacts or with favorable interaction energy -> often **destabilizing or compromise folding**
- mutation of residues in protein core -> **often destabilizing**
 - small residue to large -> **steric clashes**
 - large to small -> **loss of contacts** (creation of a void)
 - polar to non-polar -> **loss of H-bond**
 - neutral to charged -> introduction of **isolated charge**
- mutation of residues on protein surface (often neutral)
 - polar to hydrophobic -> **desolvation penalty** (destabilizing)
- mutation concerning proline or glycine -> **altered conformation**

Identification of mutable residues



❑ Mutations affecting function

- mutation of residues in binding or active sites -> **modified binding**
- mutation of residues in transport pathways -> **modified transport**
- mutation of hinge or mobile residues, residues on loops with many contacts -> **modified flexibility**
- mutation of residues directing protein localization -> **mislocalization of proteins**

Identification of mutable residues

- **Tools for annotating (identifying) the role of residues**
 - individual tools for specific analysis
 - evolutionary conservation – e.g. ConSurf, ...
 - residue contacts – e.g. Contact Map Web Viewer, ...
 - residue interactions – e.g. Protein Interaction Calculator, ...
 - accessible surface area – e.g. AsaView, Naccess, ...
 - binding sites – e.g. CASTp, metaPocket 2.0, meta-PPISP, ...
 - transport pathways – e.g. CAVER 3.0, POREWALKER, ...
 - protein dynamics – e.g. NMA, molecular dynamics
 - protein localization – e.g. SignalP, TargetP, Phobius, TMHMM, ...

Identification of mutable residues

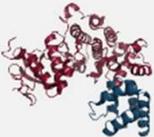
- ❑ **HotSpot Wizard – meta-server combining several tools**
 - <http://loschmidt.chemi.muni.cz/hotspotwizard/>
 - homology modelling, MSA, conservation, correlation, pockets and tunnels detection, docking, stability prediction, design of smart library

HOTSPOT WIZARD v3.1
Design of mutations and smart libraries in protein engineering

Submit new job Help Example Use cases Acknowledgement

Job ID: e.g. XXXXXX Find job

SELECT TYPE OF INPUT DATA

STRUCTURE 

SEQUENCE **IDDQD**
MSLGAKPF
GAAIAAFVRAM
VVLVVDHWGSLRGL

INPUT STRUCTURE Load example

Source: Enter PDB code
 Upload PDB file

PDB ID:

REFERENCE

Sumbalova, L., Stourac, J., Martinek, T., Bednar, D., Damborsky, J., 2018: HotSpot Wizard 3.0: Web Server for Automated Design of Mutations and Smart Libraries based on Sequence Input Information. *Nucleic Acids Research* 46 (W1): W356-W362.

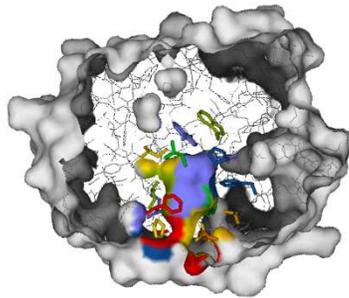
PubMed OPEN ACCESS

USER STATISTICS

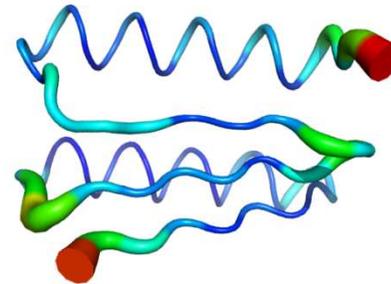
- Number of visitors: 58946
- Number of jobs: 32197

Identification of mutable residues

Functional hot-spots



Stability hot-spots (flexibility)



Stability hot-spots (evolution)

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| T | S | Y | L | W | Y | N | I | M | P | N | H | C | A | G | L | |
| - | - | S | W | L | W | R | N | I | M | - | - | H | C | A | G | L |
| T | S | Y | L | W | Y | N | I | M | P | N | H | C | A | G | L | |
| T | S | Y | L | W | R | N | I | M | P | N | H | C | A | G | L | |
| T | S | Y | L | W | R | N | I | M | P | P | P | P | A | G | L | |
| T | S | Y | L | W | R | N | I | M | P | P | P | P | A | G | L | |
| T | S | Y | L | W | R | N | I | M | P | N | H | C | A | G | L | |
| T | S | Y | L | W | R | N | I | M | P | N | H | C | A | G | L | |
| T | S | Y | L | W | R | N | I | M | P | N | H | C | A | G | L | |
| T | S | Y | L | W | R | N | I | M | P | N | H | C | A | G | L | |
| T | S | Y | L | W | R | N | I | M | P | N | H | C | A | G | L | |

Y ⇒ R

Correlated hot-spots

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| T | S | Y | L | W | Y | N | I | M | P | N | H | C | A | G | L | | | |
| - | - | S | W | L | W | R | N | I | M | D | - | - | H | C | A | G | L | |
| T | S | Y | L | W | Y | N | I | M | P | N | H | C | A | G | L | | | |
| T | S | Y | L | W | K | L | W | R | N | I | E | P | N | H | C | A | G | L |
| T | S | Y | L | W | K | L | W | R | N | I | E | P | P | P | P | A | G | L |
| T | S | Y | L | W | K | L | W | R | N | I | E | P | P | P | P | A | G | L |
| T | S | Y | L | W | K | L | W | R | N | I | E | P | N | H | C | A | G | L |
| T | S | Y | L | W | K | L | W | R | N | I | E | P | N | H | C | A | G | L |
| T | S | Y | L | W | W | L | W | R | N | I | V | P | N | H | C | A | G | L |
| T | S | Y | L | W | W | L | W | R | N | I | V | P | N | H | C | A | G | L |

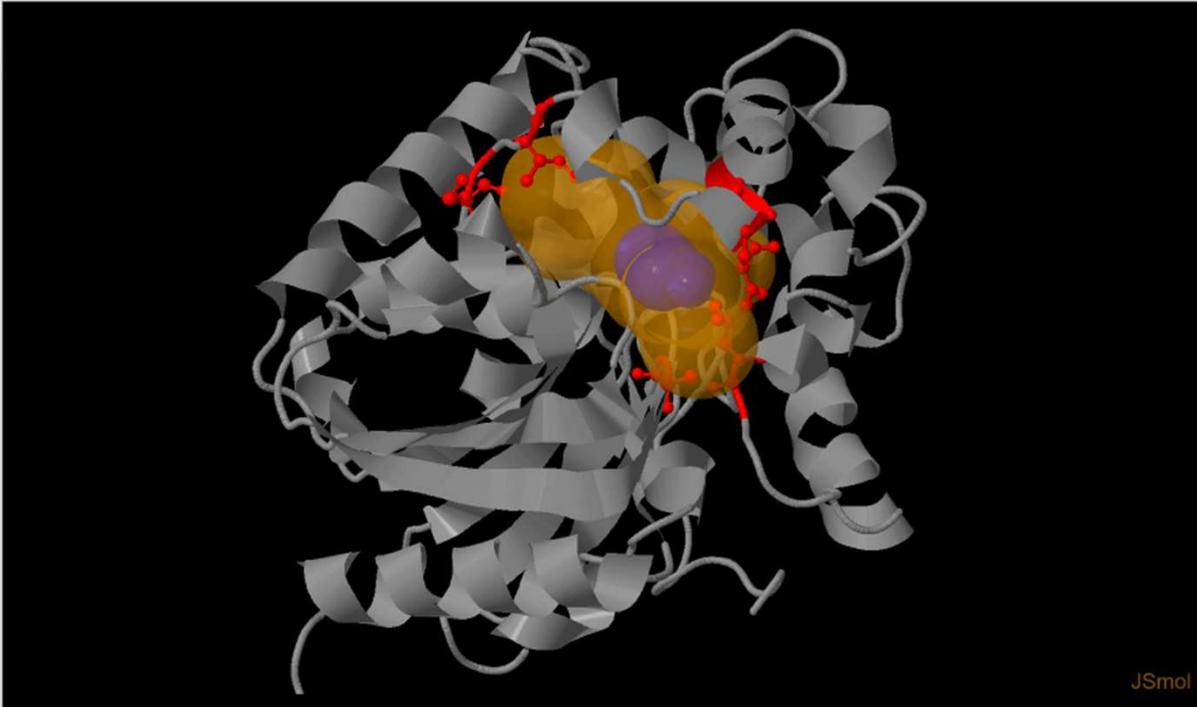
↑ ↑

Prediction of mutational effects - mutable residues

Identification of mutable residues

Functional hot spots of 1CV2

Viewer



Return to Results browser

Visualization settings

Structure visualization style:

Wireframe Cartoon

Sticks Trace

Balls & sticks Backbone

Balls

Hide all visualized residues

Save image

Reset view

Visualization quality:

1 8

Tunnels

| id | length (Å) | bottleneck radius (Å) |
|---------------------------------------|------------|-----------------------|
| Starting from pocket: 1 | | |
| <input checked="" type="checkbox"/> 1 | 7.7 | 1.5 |

Pockets

| id | chain(s) | relevance (%) | volume (Å ³) |
|---------------------------------------|----------|---------------|--------------------------|
| <input checked="" type="checkbox"/> 1 | A | 100 | 576 |
| <input type="checkbox"/> 2 | A | 82 | 883 |
| <input type="checkbox"/> 3 | A | 62 | 275 |
| <input type="checkbox"/> 4 | A | 28 | 753 |

Residue features

Exclude correlated positions
 Exclude catalytic pockets
 Exclude tunnels
 Exclude α -helices and β -sheets

Exclude buried residues
 Include residues with moderate mutability

| | chain | position | residue | mutable | non-essential | in tunnel | in catalytic pocket | HotSpot |
|-------------------------------------|-------|----------|---------|---------|---------------|-----------|---------------------|---------|
| Chain A | | | | | | | | |
| <input checked="" type="checkbox"/> | A | 146 | Gln | ✓ | ✓ | ✓ | ✓ | ✓ |
| <input checked="" type="checkbox"/> | A | 136 | Met | ✓ | ✓ | X | ✓ | ✓ |
| <input checked="" type="checkbox"/> | A | 147 | Asp | ✓ | ✓ | ✓ | ✓ | ✓ |
| <input checked="" type="checkbox"/> | A | 271 | Ala | ✓ | ✓ | ✓ | ✓ | ✓ |
| <input checked="" type="checkbox"/> | A | 138 | Ile | ✓ | ✓ | X | ✓ | ✓ |
| <input type="checkbox"/> | A | 247 | Ala | ✓ | ✓ | ✓ | ✓ | ✓ |
| <input type="checkbox"/> | A | 248 | Leu | ✓ | ✓ | ✓ | ✓ | ✓ |

Residues selected for mutagenesis

| | chain | position | residue | HotSpot |
|-------------------------------------|-------|----------|---------|---------|
| <input checked="" type="checkbox"/> | A | 146 | Gln | ✓ |
| <input checked="" type="checkbox"/> | A | 136 | Met | ✓ |

Identification of mutable residues

Library design

Standard SwiftLib

AA selection mode : Amino acid frequency Minimal frequency (%) : 5 Include wild-type Exclude wild-type

| chain | position | residue | desired amino acids | codon | desired ratio (%) | stop ratio (%) |
|-------|----------|---------|-----------------------------------|-------|-------------------|----------------|
| A | 136 | Met | Ala, Lys, Pro, Gln, Arg, Thr | VVR | 77.8 | 0.0 |
| A | 146 | Gln | Ala, Asp, Glu, Gly, Pro, Gln, Ser | BVV | 63.0 | 11.1 |
| A | 147 | Asp | Ala, Phe, Gly, Leu, Met, Thr, Val | DBS | 61.1 | 0.0 |

| codon | desired ratio (%) | stop ratio (%) | desired amino acids | encoded amino acids |
|-------|-------------------|----------------|---|---|
| DBS | 100.0 | 0.0 | Ala:2 Cys:1 Phe:1 Gly:2 Ile:1 Leu:1 Met:1 Arg:1 Ser:3 Thr:2 Val:2 Trp:1 | Ala:2 Cys:1 Phe:1 Gly:2 Ile:1 Leu:1 Met:1 Arg:1 Ser:3 Thr:2 Val:2 Trp:1 |
| DBK | 100.0 | 0.0 | Ala:2 Cys:1 Phe:1 Gly:2 Ile:1 Leu:1 Met:1 Arg:1 Ser:3 Thr:2 Val:2 Trp:1 | Ala:2 Cys:1 Phe:1 Gly:2 Ile:1 Leu:1 Met:1 Arg:1 Ser:3 Thr:2 Val:2 Trp:1 |
| DBB | 100.0 | 0.0 | Ala:3 Cys:2 Phe:2 Gly:3 Ile:2 Leu:1 Met:1 Arg:1 Ser:5 Thr:3 Val:3 Trp:1 | Ala:3 Cys:2 Phe:2 Gly:3 Ile:2 Leu:1 Met:1 Arg:1 Ser:5 Thr:3 Val:3 Trp:1 |
| DBN | 97.2 | 2.8 | Ala:4 Cys:2 Phe:2 Gly:4 Ile:3 Leu:2 Met:1 Arg:2 Ser:6 Thr:4 Val:4 Trp:1 | Ala:4 Cys:2 Phe:2 Gly:4 Ile:3 Leu:2 Met:1 Arg:2 Ser:6 Thr:4 Val:4 Trp:1 |
| DBV | 96.3 | 3.7 | Ala:3 Cys:1 Phe:1 Gly:3 Ile:2 Leu:2 Met:1 Arg:2 Ser:4 Thr:3 Val:3 Trp:1 | Ala:3 Cys:1 Phe:1 Gly:3 Ile:2 Leu:2 Met:1 Arg:2 Ser:4 Thr:3 Val:3 Trp:1 |
| DBD | 96.3 | 3.7 | Ala:3 Cys:1 Phe:1 Gly:3 Ile:2 Leu:2 Met:1 Arg:2 Ser:4 Thr:3 Val:3 Trp:1 | Ala:3 Cys:1 Phe:1 Gly:3 Ile:2 Leu:2 Met:1 Arg:2 Ser:4 Thr:3 Val:3 Trp:1 |
| NBS | 91.7 | 0.0 | Ala:2 Cys:1 Phe:1 Gly:2 Ile:1 Leu:3 Met:1 Arg:3 Ser:3 Thr:2 Val:2 Trp:1 | Ala:2 Cys:1 Phe:1 Gly:2 Ile:1 Leu:3 Met:1 Pro:2 Arg:3 Ser:3 Thr:2 Val:2 Trp:1 |
| NBK | 91.7 | 0.0 | Ala:2 Cys:1 Phe:1 Gly:2 Ile:1 Leu:3 Met:1 Arg:3 Ser:3 Thr:2 Val:2 Trp:1 | Ala:2 Cys:1 Phe:1 Gly:2 Ile:1 Leu:3 Met:1 Pro:2 Arg:3 Ser:3 Thr:2 Val:2 Trp:1 |
| NBB | 91.7 | 0.0 | Ala:3 Cys:2 Phe:2 Gly:3 Ile:2 Leu:4 Met:1 Arg:4 Ser:5 Thr:3 Val:3 Trp:1 | Ala:3 Cys:2 Phe:2 Gly:3 Ile:2 Leu:4 Met:1 Pro:3 Arg:4 Ser:5 Thr:3 Val:3 Trp:1 |
| NBN | 89.6 | 2.1 | Ala:4 Cys:2 Phe:2 Gly:4 Ile:3 Leu:6 Met:1 Arg:6 Ser:6 Thr:4 Val:4 Trp:1 | Ala:4 Cys:2 Phe:2 Gly:4 Ile:3 Leu:6 Met:1 Pro:4 Arg:6 Ser:6 Thr:4 Val:4 Trp:1 |

Library size : 7315

Expected coverage : 0.95

Probability of full coverage : 0

Codon usage : Escherichia coli K12

Generate report

Prediction of mutational effects - mutable residues

Prediction of effects on structure

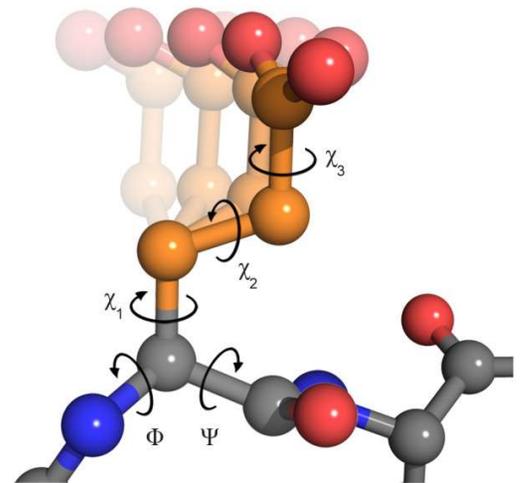


□ Prediction of mutant structures – general workflow

- mutated residue and its surroundings represented by rotamers from **rotamer library** (conformations derived from X-ray structures)
- the best set of rotamers selected **by Monte Carlo** approach
- optionally – **energy minimization, backbone flexibility**
- **comparing structures of mutant and native protein -> assessment of the mutational effect**

□ Available tools

- PyMOL; FOLDX
- WhatIF; RosettaBackrub; ...

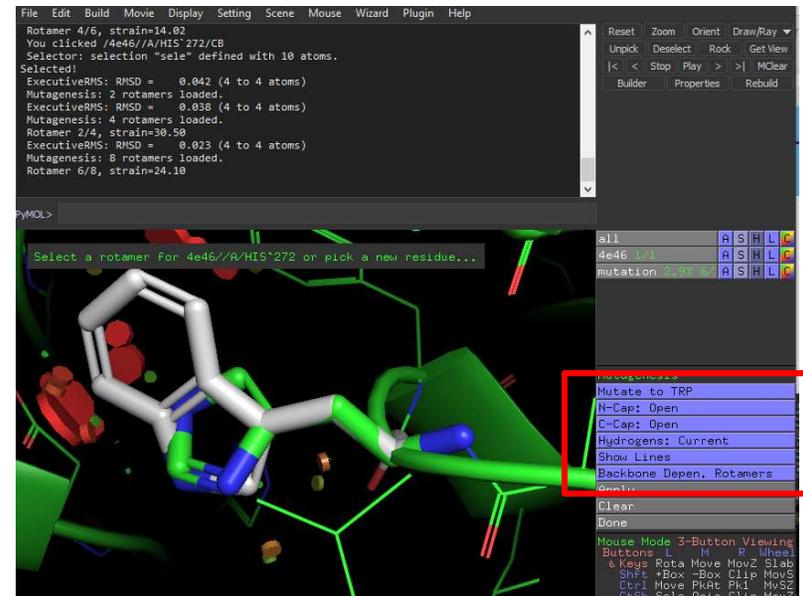
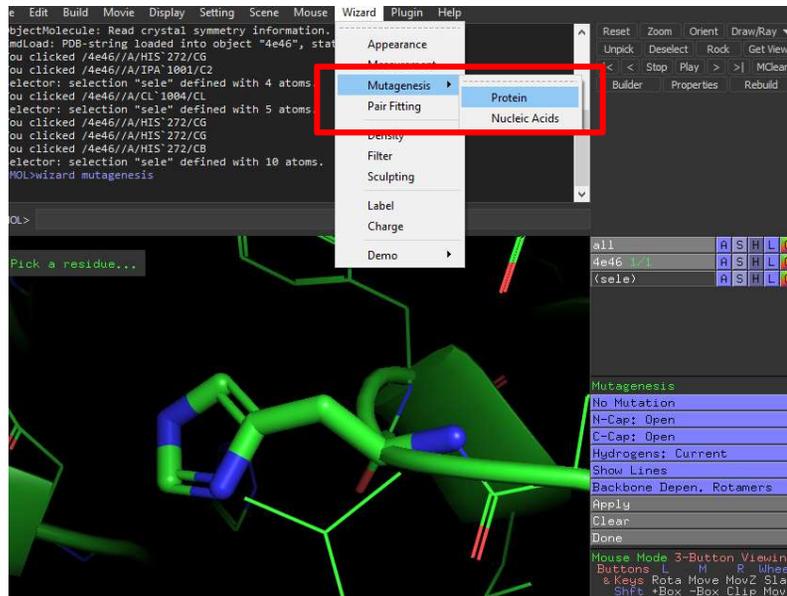


Prediction of effects on structure



PyMOL

- <https://pymol.org/>
- Mutagenesis module
- user can choose rotamers and visualize potential clashes; very fast
- fixed backbone; no mutational scoring



Prediction of effects on structure

□ FOLDX

- <http://foldxsuite.crg.eu/>
- **stand alone**, with plug-in to Yasara modeling tool
- **fast**
- **fixed backbone** conformation
- construction of **multiple mutants**
- empirical scoring function for calculation of **stability change** ($\Delta\Delta G$)

Prediction of effects on structure

□ FOLDX

The screenshot displays the FOLDX software interface. The main window shows a 3D ribbon representation of a protein structure. A menu is open, listing various analysis options. The 'FoldX' option is selected, and its sub-menu is visible, with 'Mutate residue' and 'Mutate multiple residues' highlighted by a red box. The interface includes a menu bar (File, Edit, Simulation, Analyze, View, Effects, Options, Window, Help), a left sidebar with 'ATOM PROPERTIES' and 'Bonds' sections, and a right sidebar with a 'SCENE CONTENT' table.

| Obj | Name | Visi | Acti | Atom |
|-----|------|------|------|------|
| 1 | 1crn | Yes | Yes | 1 |
| 2 | | No | No | |
| 3 | | No | No | |
| 4 | | No | No | |
| 5 | | No | No | |
| 6 | | No | No | |
| 7 | | No | No | |
| 8 | | No | No | |
| 9 | | No | No | |
| 10 | | No | No | |

Prediction of effects on structure

□ WHATIF

- <https://swift.cmbi.umcn.nl/servers/html/index.html>
- web server for multiple purpose including mutagenesis
- very fast
- fixed backbone conformation
- construction of single mutants only
- no scoring function

Prediction of effects on structure

□ RosettaBackrub

- <https://kortemmelab.ucsf.edu/backrub>
- **web server** primarily aimed to design protein backbone
- **slow**
- **fixed** or **flexible backbone** conformation (ensemble)
- construction of **multiple mutants**
- **general scoring function** for calculation of wild-type and mutant stability (user has to calculate the difference)
- **visualization** of possible mutants in Jmol

Prediction of effects on structure

□ RosettaBackrub

UCSF | kortemmelab | RosettaBackrub

[Other Services: [Alanine Scanning](#)]
[briza | Logout]

[Home] [Documentation] [Register]
[Submit] [Queue] [My Account]

Submit a new job

- Point Mutation
 [Smith and Kortemme, 2008]
 → One Mutation
 → Multiple Mutations
- Backrub Ensemble
 → Backrub Ensemble
 [Smith and Kortemme, 2008]
 → Backrub Ensemble Design
 [Friedland et al., 2008]
- Sequence Tolerance
 → Interface Sequence Tolerance
 [Humphris and Kortemme, 2008]
 → Generalized Protocol (Fold / Interface) Sequence Tolerance
 [Smith and Kortemme, 2010]
 [Smith and Kortemme, 2011]

Point Mutation

Select an amino acid

- ALA
- ARG
- ASN
- ASP
- CYS
- GLN
- GLU
- GLY
- HIS
- ILE
- LEU
- LYS
- MET
- PHE
- PRO
- SER
- THR
- TRP
- TYR

User Name:

Rosetta Version:

PDB:

General Settings

Job Name:

Number of structures:

Application Specific

| # | Chain ID | Residue ID | Select an amino acid |
|---|----------|------------|----------------------|
| 1 | A | 15 | Select an amino acid |

output

Total scores for the generated structures. Download files:

- total scores only
- detailed scores
- detailed scores (NTBL)
- detailed scores for residues (also in individual job files)
- detailed scores for residues (NTBL)

| Filename | Score |
|-------------------|--------|
| 1CQW_job.pdf | 23.942 |
| 1CQW_0001_job.pdf | 10.483 |
| 1CQW_0002_job.pdf | 10.892 |
| 1CQW_0003_job.pdf | 10.775 |
| 1CQW_0004_job.pdf | 9.246 |
| 1CQW_0005_job.pdf | 11.122 |
| 1CQW_0006_job.pdf | 11.204 |
| 1CQW_0007_job.pdf | 10.264 |
| 1CQW_0008_job.pdf | 10.507 |
| 1CQW_0009_job.pdf | 10.218 |
| 1CQW_0010_job.pdf | 10.545 |
| 1CQW_0011_job.pdf | 11.315 |
| 1CQW_0012_job.pdf | 11.220 |
| 1CQW_0013_job.pdf | 10.622 |
| 1CQW_0014_job.pdf | 10.979 |
| 1CQW_0015_job.pdf | 10.782 |
| 1CQW_0016_job.pdf | 11.406 |
| 1CQW_0017_job.pdf | 10.895 |
| 1CQW_0018_job.pdf | 11.406 |
| 1CQW_0019_job.pdf | 10.000 |
| 1CQW_0020_job.pdf | 11.399 |

Structural models for up to 10 of the best scoring structures. The query structure is shown in red, the mutated residue is shown as sticks representation.

Please wait, it may take a few moments to load the Cx face representation.

| Model | Prescribed |
|---|-------------------------------------|
| <input checked="" type="checkbox"/> 1CQW | <input type="checkbox"/> |
| <input checked="" type="checkbox"/> 1CQW_0004_job | <input checked="" type="checkbox"/> |
| <input checked="" type="checkbox"/> 1CQW_0019_job | <input checked="" type="checkbox"/> |
| <input checked="" type="checkbox"/> 1CQW_0009_job | <input checked="" type="checkbox"/> |
| <input checked="" type="checkbox"/> 1CQW_0007_job | <input checked="" type="checkbox"/> |
| <input checked="" type="checkbox"/> 1CQW_0001_job | <input checked="" type="checkbox"/> |
| <input checked="" type="checkbox"/> 1CQW_0002_job | <input checked="" type="checkbox"/> |
| <input checked="" type="checkbox"/> 1CQW_0017_job | <input checked="" type="checkbox"/> |
| <input checked="" type="checkbox"/> 1CQW_0013_job | <input checked="" type="checkbox"/> |
| <input checked="" type="checkbox"/> 1CQW_0003_job | <input checked="" type="checkbox"/> |
| <input checked="" type="checkbox"/> 1CQW_0015_job | <input checked="" type="checkbox"/> |

Prediction of pathogenicity



□ Prediction of impact of mutation on protein function

- tools employ **machine learning approaches**
- **trained on functional experimental data**
- predictions can be based on **sequence only**
- **qualitative results** – i.e. deleterious versus neutral
- primarily **intended for pathogenicity** prediction (leading to disease)

□ Available tools

- MutPred, SNAP, PhD-SNP, SIFT, MAPP ...
- **PredictSNP** – meta server combining many tools

Prediction of pathogenicity

- There are many more tools out there

| Method | Based on | Training set | Conservation analysis | Structural attributes | Annotations | Website |
|---------------|-------------------------|--------------------------------------|--|-------------------------------|-------------|---|
| MutPred | RF | HGMD, Swiss-Prot | SIFT, Pfam, PSI-BLAST | Predicted attributes | – | http://mutpred.mutdb.org/ |
| nsSNPAnalyzer | RF | Swiss-Prot | SIFT | Homologue mapping | – | http://snpanalyzer.uthsc.edu/ |
| Panther | Alignment scores | – | Panther library, HMMs | – | – | http://www.pantherdb.org/tools/csnpscoreForm.jsp |
| PhD-SNP | SVM | Swiss-Prot | Sequence environment, sequence profiles | – | – | http://gpcr2.biocomp.unibo.it/cgi/predictors/PhD-SNP/PhD-SNP.cgi |
| PolyPhen | Empirical rules | – | PSIC profiles | Homologue mapping/predictions | Swiss-Prot | http://genetics.bwh.harvard.edu/pph/ |
| PolyPhen2 | Bayesian classification | Swiss-Prot, neutral pseudo-mutations | PSIC profiles | Homologue mapping/predictions | Pfam domain | http://genetics.bwh.harvard.edu/pph2/ |
| SIFT | Alignment scores | – | MSAs | – | – | http://sift.jcvi.org/ |
| SNAP | NN | PMD, neutral pseudo-mutations | PSIC profiles, Pfam, PSI-BLAST | Predictions | – | http://rostlab.org/services/snap/ |
| SNPs&GO | SVM | Swiss-Prot | Sequence environment, sequence profiles, Panther | – | GO | http://snps-and-go.biocomp.unibo.it/snps-and-go/ |

Prediction of pathogenicity

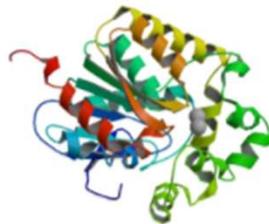
- ❑ **PredictSNP:** <http://loschmidt.chemi.muni.cz/predictsnp/>
 - ❑ Combines many tools for Protein or DNA assessment of SNPs



Consensus classifiers for prediction of disease-related mutations



Consensus classifier for prediction of the effect of *amino acid* substitutions.



Consensus classifier for prediction of the effect of *nucleotide* substitutions.



Prediction of pathogenicity



PREDICTSNP¹ Consensus classifier for prediction of disease related amino acid mutations



Home
Use cases

INPUT Load example

Insert protein sequence in FASTA format:

```
>HEA_HUGAN
MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKYFHEFDSHNSGSAQVTHGSD
KLVKALINVAIVGDMPHALSLSDLNHRIVDFVFNFLSHCLLVTLAAAEPAETTP
AIVDRLRFELASVSTVLTISKYR
```

Load

MUTATIONS Manual input

Select positions:

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|-----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|--|
| 1 | M | V | L | S | P | A | D | K | T | N | V | K | A | A | W | G | K | V | G | A | H | A | G | E | Y | G | A | E | A | L | E | R | M | F | L | S | F | P | T | T | |
| 41 | K | T | Y | F | P | H | F | D | L | S | H | G | S | A | Q | V | K | G | H | G | K | K | V | A | D | A | L | T | N | A | V | A | H | V | D | D | M | P | N | A | |
| 81 | L | S | A | L | S | D | L | H | A | H | K | L | R | V | D | P | V | N | F | K | L | L | S | H | C | L | L | V | T | L | A | A | H | L | P | A | E | F | T | P | |
| 121 | A | V | H | A | S | L | D | K | F | L | A | S | V | S | T | V | L | T | S | K | Y | R | | | | | | | | | | | | | | | | | | | |

| Pos | Wild-type | Mutations | Clear |
|-----|-----------|------------------|------------------------------------|
| 59 | H | Y - Tyr | ✖ |
| 80 | G | D - Asp, V - Val | ✖ |
| 83 | V | T - Thr | ✖ |
| 88 | T | V - Val | ✖ |
| 72 | A | E - Glu, V - Val | ✖ |

Clear all mutations

TOOLS FOR EVALUATION

| Tool name | Time demands | Expected accuracy |
|--|--------------|-------------------|
| <input checked="" type="checkbox"/> PredictSNP | 32 min | 73.4% |
| <input checked="" type="checkbox"/> MAPP | 10 min | 70.7% |
| <input checked="" type="checkbox"/> PhD-SNP | 32 min | 71.5% |
| <input checked="" type="checkbox"/> PolyPhen-1 | 15 min | 68.1% |
| <input checked="" type="checkbox"/> PolyPhen-2 | 15 min | 69.2% |
| <input checked="" type="checkbox"/> SIFT | 15 min | 70.3% |
| <input checked="" type="checkbox"/> SNAP | 30 min | 67.6% |

JOB CONTROL

Submit job

Job ID:

Find job

REFERENCE

Bendi, J., Stourac, J., Salanda, O., Paveka, A., Weben, E.D., Zందుకా, J., Brezovsky, J., Damborsky, J., 2014: PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. PLOS Computational Biology 10: e1003440.

USER STATISTICS

- Number of visitors: 32175
- Number of jobs: 25238

CONTACT

Loschmidt Laboratories

- predictsnp@sci.muni.cz
- <http://loschmidt.chemi.muni.cz>

OTHER TOOLS

RESOURCES

User guide

- Link: [PDF](#)

PredictSNP benchmark dataset

- 24,082 neutral / 19,800 deleterious
- Links: [XLS](#), [dataset statistics](#)

PMD testing dataset

- 1,248 neutral / 2,249 deleterious
- Links: [XLS](#), [dataset statistics](#)

MMP testing dataset

- 4,456 neutral / 7,539 deleterious
- Links: [XLS](#), [dataset statistics](#)

OVERFIT testing dataset

- 15,081 neutral / 17,695 deleterious
- Links: [XLS](#), [dataset statistics](#)

Rational design of proteins



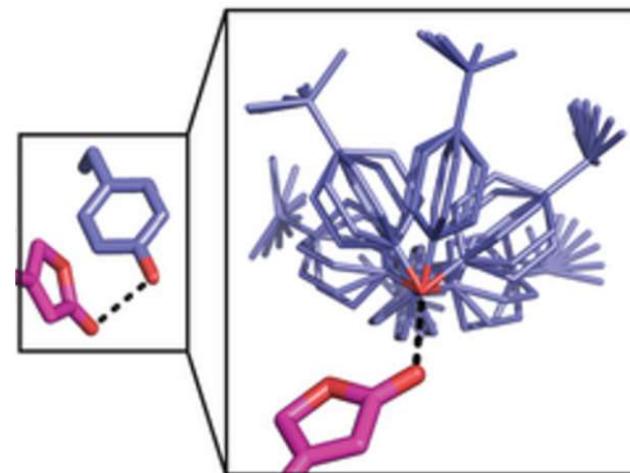
- We can use mutagenesis to rationally design proteins according to our needs (**protein engineering**)
- Properties that can be modified by mutagenesis
 - **Function**
 - Ligand binding (e.g., catalytic activity or substrate selectivity)
 - Macromolecular interface
 - **Stability**
 - **Solubility**

Improving ligand binding and activity



□ Rosetta

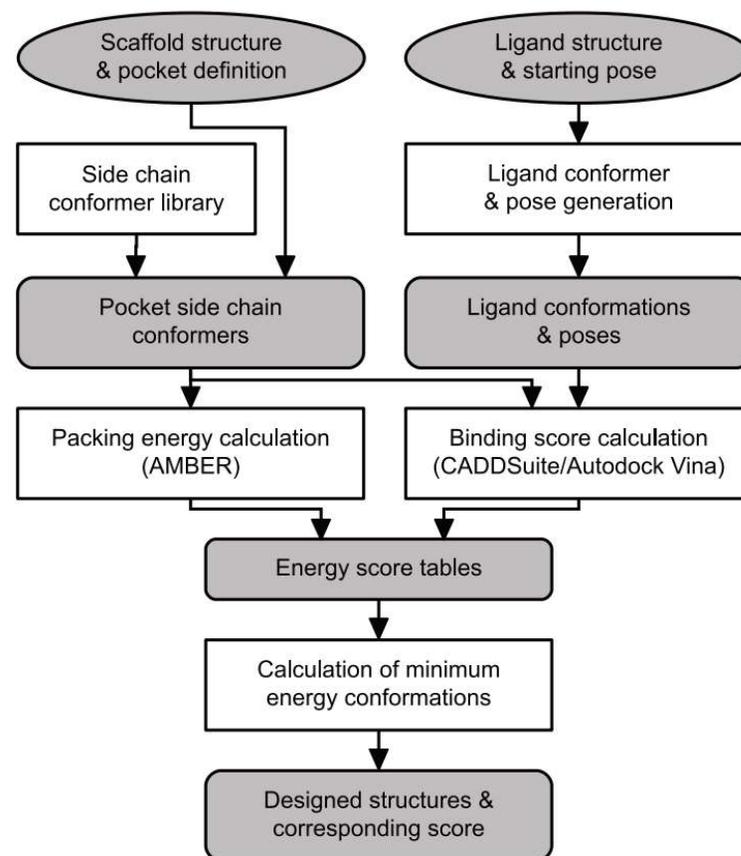
- <https://www.rosettacommons.org/>
- A large suite of many tools to model structures
- Predicts free energy changes upon mutations ($\Delta\Delta G$)
- Monte Carlo sampling (random search) to predict minimum-energy structure of mutants
- **RosettaDesign** webserver
 - <http://rosettadesign.med.unc.edu/>
 - helps design mutations to optimize the binding site and increase interactions with a ligand/substrate



Improving ligand binding and activity

□ PocketOptimizer

- <http://www.eb.tuebingen.mpg.de/birte-hoecker/algorithms-and-software/pocketoptimizer.html>
- Aimed at maximizing the affinity of a binding site towards a ligand
- Modular pipeline with different tools
- Docking, mutagenesis, scoring function
- Predicts global minimum-energy conformations among the designs



Improving ligand binding and activity

□ FuncLib

- <https://funclib.weizmann.ac.il/bin/steps>
- utilizes evolution (conservation) and RosettaDesign (energy) to **introduce multiple-point mutations** to modify the properties of the binding site
- it can be used to improve the binding affinity towards a ligand
- outputs up to 50 multiple-point mutants for protein synthesis

Improving ligand binding and activity

□ FuncLib

| Parameter | Value | | | | | | | | | | | | | | |
|---|--|------|----|------|---|------|-----|------|--------|------|------|------|---------|------|------|
| Minimal number of mutations per design | <input type="text" value="3"/> | | | | | | | | | | | | | | |
| Maximal number of mutations per design | <input type="text" value="5"/> | | | | | | | | | | | | | | |
| Minimal PSSM threshold | <input type="text" value="-1"/> | | | | | | | | | | | | | | |
| $\Delta\Delta G$ | <input type="text" value="5.5"/> | | | | | | | | | | | | | | |
| Sequence space | <table><tbody><tr><td>143A</td><td>FY</td></tr><tr><td>144A</td><td>P</td></tr><tr><td>151A</td><td>FMY</td></tr><tr><td>177A</td><td>LAGNST</td></tr><tr><td>211A</td><td>ILMV</td></tr><tr><td>247A</td><td>AGMSTVY</td></tr><tr><td>248A</td><td>LIMV</td></tr></tbody></table> | 143A | FY | 144A | P | 151A | FMY | 177A | LAGNST | 211A | ILMV | 247A | AGMSTVY | 248A | LIMV |
| 143A | FY | | | | | | | | | | | | | | |
| 144A | P | | | | | | | | | | | | | | |
| 151A | FMY | | | | | | | | | | | | | | |
| 177A | LAGNST | | | | | | | | | | | | | | |
| 211A | ILMV | | | | | | | | | | | | | | |
| 247A | AGMSTVY | | | | | | | | | | | | | | |
| 248A | LIMV | | | | | | | | | | | | | | |
| Total number of designs in tolerated sequence space | 3,313 | | | | | | | | | | | | | | |

Optimizing protein-protein interface

□ AffiLib

- <https://affilib.weizmann.ac.il/bin/steps>
- utilizes RosettaDesign (energy) and evolution (conservation) to **introduce mutations** and optimize macromolecular interface
- suggests mutations on interface positions for improvement of the binding affinity
- outputs up to 50 multiple-point mutants for protein synthesis

Optimizing protein-protein interface

□ mutation Cutoff Scanning Matrix (mCSM-PPI2)

- http://biosig.unimelb.edu.au/mcsm_ppi2/
- based on machine learning, evolutionary data and energy (FoldX)
- provides mutational $\Delta\Delta G$
- modes of calculations
 - **single mutation** – single point mutations on interface
 - **mutation list** – single mutations accordingly to a user
 - **systematic** – position saturation (all interface residues are mutated to all other 19 amino acids)
 - **alanine scanning** (all interface residues are mutated to alanine)

Improving protein stability



❑ Prediction of stability change upon mutation

- **structure** of mutant protein **may not be produced**
- tools often employ
 - **empirical scoring functions**
 - **machine learning approaches**

❑ Available tools

- FOLDX, RosettaBackrub
- PoPMuSiC
- ...
- Hybrid tools for protein stabilization

Improving protein stability

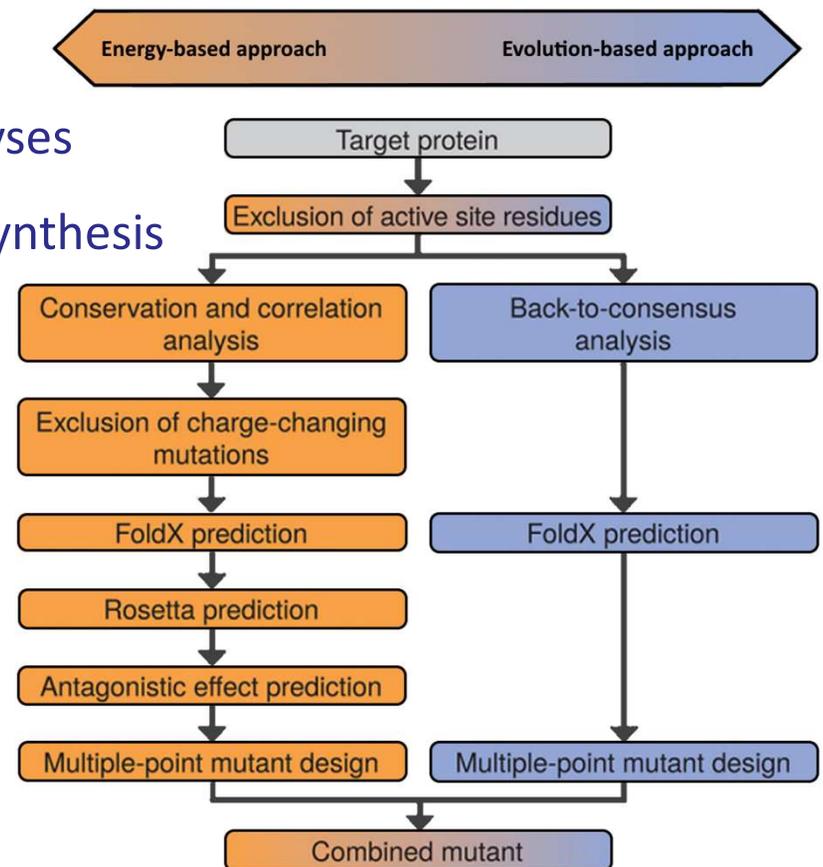


- **Prediction of Protein Mutant Stability Changes (PoPMuSiC)**
 - <http://dezyme.com/>
 - uses four statistical potentials weighted on the basis of the solvent accessibility of the mutated residue
 - commercial
 - three modes of calculations
 - **manual** – selected set of single point mutations
 - **mutations list** – evaluates list of mutations specified by the user
 - **systematic** – protein saturation (all mutations at all positions)

Improving protein stability

□ FireProt

- <https://loschmidt.chemi.muni.cz/fireprotweb>
- *In silico* analysis of all mutations
- Energy- and evolution-based analyses
- Multiple-point mutants for gene synthesis



Improving protein stability

□ FireProt

The screenshot displays the FireProt web interface. On the left is a 3D viewer showing a protein structure in grey ribbon representation with several residues highlighted in orange and blue. On the right is a control panel with three main sections: Visualization settings, FireProt protocol design, and Mutant designer.

Visualization settings:

- Structure visualization style: Wireframe, Cartoon (selected), Sticks, Trace, Balls & sticks, Backbone, Balls
- Buttons: Hide all visualized residues, Save image, Reset view
- Visualization quality: 1 (slider from 1 to 8)

FireProt protocol design:

- PDB ID: 4e46
- Length: 292
- Evolution mutant: -3.7 kcal/mol (6 mutations)
- Energy mutant: -20.85 kcal/mol (8 mutations)

Mutant designer:

- Buttons: Original selection, Save mutant, Download all designs (.zip)
- ENERGY MUTANT DDG: -20.85 KCAL/MOL

| | chain | position | ref | alt | foldx | rosetta |
|---|-------|----------|-----|-----|-------|---------|
| ⊖ | A | 11 | D | P | -1.39 | -1.89 |
| ⊖ | A | 33 | T | I | -1.31 | -1.94 |
| ⊖ | A | 145 | A | L | -2.77 | -1.71 |

Improving protein stability

□ FireProt

| Mutations | | | | | | | | | | |
|--------------------------|-------|---------------|-----|------------------|--------------------|----------------|---------|-----------------------|------------------|-------|
| Combined mutant | | Energy mutant | | Evolution mutant | | Wild-type | | | | |
| Mutation info | | | | | Energy information | | | Evolution information | | |
| visualize | chain | position | ref | alt | not conserved | not correlated | rosetta | mutable by majority | mutable by ratio | foldx |
| <input type="checkbox"/> | A | 11 | D | P | ✓ | ✓ | -1.89 | X | X | -1.39 |
| <input type="checkbox"/> | A | 20 | E | S | ✓ | ✓ | - | ✓ | ✓ | 0.08 |
| <input type="checkbox"/> | A | 33 | T | I | ✓ | ✓ | -1.94 | X | X | -1.31 |
| <input type="checkbox"/> | A | 119 | N | H | X | ✓ | - | ✓ | X | -1 |
| <input type="checkbox"/> | A | 145 | A | L | ✓ | ✓ | -1.71 | X | X | -2.77 |
| <input type="checkbox"/> | A | 148 | T | L | ✓ | ✓ | -2.15 | X | X | -1.84 |
| <input type="checkbox"/> | A | 155 | A | P | ✓ | ✓ | -0.85 | ✓ | ✓ | -1.1 |
| <input type="checkbox"/> | A | 164 | D | M | ✓ | ✓ | -1.85 | X | X | -1.18 |
| <input type="checkbox"/> | A | 176 | C | W | ✓ | ✓ | -6.69 | X | X | -1.76 |
| <input type="checkbox"/> | A | 187 | D | W | ✓ | ✓ | -2.81 | X | X | -1.1 |
| <input type="checkbox"/> | A | 198 | D | S | ✓ | ✓ | - | ✓ | X | -0.7 |
| <input type="checkbox"/> | A | 200 | E | R | ✓ | ✓ | - | ✓ | X | -0.4 |
| <input type="checkbox"/> | A | 217 | N | W | ✓ | ✓ | -1.76 | ✓ | ✓ | -1.38 |
| <input type="checkbox"/> | A | 285 | E | A | ✓ | ✓ | - | ✓ | X | -0.38 |

Improving protein stability

□ FireProt^{ASR}

- <https://loschmidt.chemi.muni.cz/fireprotasr>
- sequence-based stabilization: ancestral sequence reconstruction
- analysis of protein evolution and protein stabilization

FIREPROT^{ASR} v1.1 Fully automated ancestral sequence reconstruction

Submit new job Help Example Use cases Acknowledgement Job ID: e.g. xxxxxxx Find job

SELECT THE STARTING POINT

SEQUENCE USER DATA

STARTING FROM SEQUENCE Load example

Source : Enter own sequence Upload sequence file

Sequence :

Validate

REFERENCE

Musil, M., Khan, R., Beier, A., Stourac, J., Konegger, H., Damborsky, J., Bednar, D. 2020: FireProt-ASR: A Web Server for Fully Automated Ancestral Sequence Reconstruction. *Briefings in Bioinformatics*, 2020, bbaa337.

PubMed OPEN ACCESS

USER STATISTICS

- Number of visitors: 7286
- Number of jobs: 2109

CONTACT

Loschmidt Laboratories

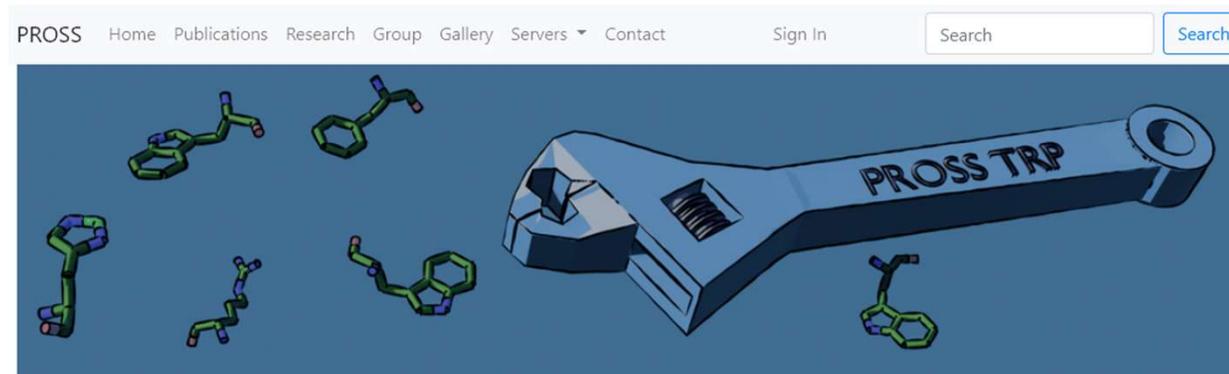
- fireprot@sci.muni.cz
- <https://loschmidt.chemi.muni.cz>

SUPPORTED BY

Improving protein stability

□ PROSS

- <https://pross.weizmann.ac.il/step/pross-terms/>
- Combination of mutations “allowed” by conservation analysis and Rosetta calculations (energy)



PROSS Home Publications Research Group Gallery Servers ▾ Contact Sign In Search Search

PROSS: the Protein Repair One-Stop Shop

Input

- 1 Pross Terms
- 2 Upload PDB?
- 3 Files
- 4 Structure info
- 5 Constraints
- 6 Upload MSA?
- 7 MSA file
- 8 PROSS MSA
- 9 Energy function

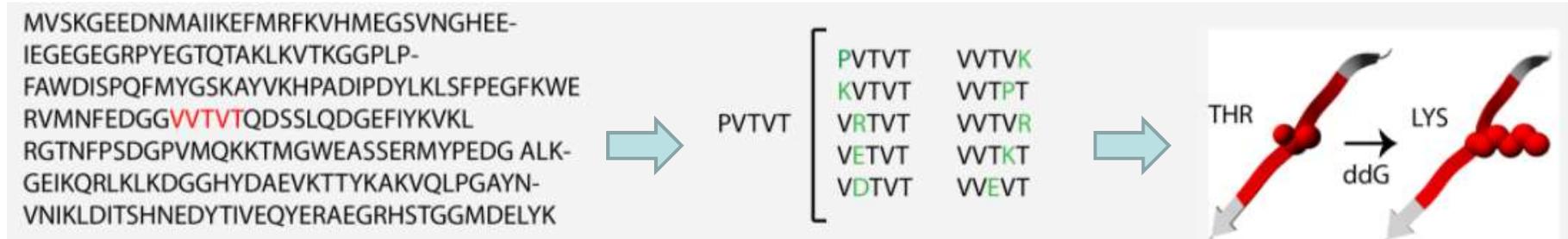
Email address*
me@example.com
Valid academic email address to send the reports to.

Improving protein solubility

❑ Aggrescan3D

❑ SolubiS

- <https://solubis.switchlab.org/>
- To identify stabilizing mutations that reduce the aggregation tendency of a protein
- 1) Identifies exposed APRs
- 2) Introduces “gatekeeper” residues (P, R, K, D and E) into APSs
- 3) Assesses the stability changes of mutations ($\Delta\Delta G$)



References I



- ❑ Ng, P. C. & Henikoff, S. (2006) Predicting the effects of amino acid substitutions on protein function. *Annual Review of Genomics and Human Genetics* **7**: 61-80.
- ❑ Thusberg, J. & Vihinen, M. (2009) Pathogenic or not? And if so, then how? Studying the effects of missense mutations using bioinformatics methods. *Human Mutation* **30**: 703-714.
- ❑ Potapov, V. *et al.* (2009) Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details. *Protein Engineering, Design & Selection* **22**: 553-560.

References II

- ❑ Khan, S. & Vihinen, M. (2010) Performance of protein stability predictors. *Human Mutation* **31**: 675-684.
- ❑ Bendl, J. *et al.* (2016) PredictSNP2: A Unified Platform for Accurately Evaluating SNP Effects by Exploiting the Different Characteristics of Variants in Distinct Genomic Regions. *PLOS Computational Biology* **12**: e1004962.
- ❑ Musil, M. *et al.* (2019) Computational Design of Stable and Soluble Biocatalysts. *ACS Catalysis* **9**: 1033–1054.
- ❑ Planas-Iglesias, J. *et al.* (2021) Computational design of enzymes for biotechnological applications. *Biotechnology Advances* **47**:107696