

LOSCHMIDT
LABORATORIES



Engineering of protein structures

Outline

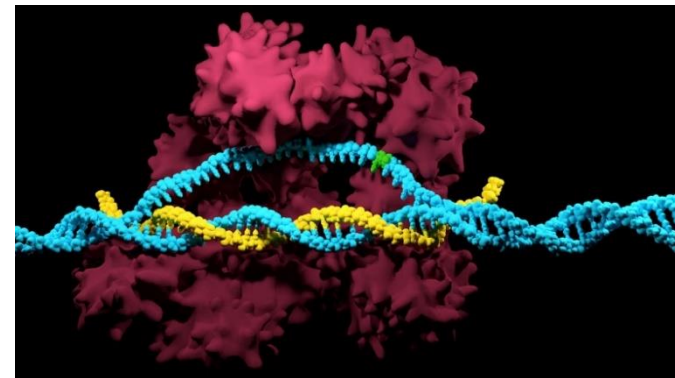
- ❑ Overview of mutations
- ❑ Databases of mutations
- ❑ Missense mutations
- ❑ Prediction of mutational effects
- ❑ Rational design of proteins

Overview of mutations

❑ Mutations in DNA or RNA may occur

- Errors in DNA replication during cell division
- Exposure to mutagens (physical or chemical agents)
- Viral infections
- ...Or scientist intervention 😊

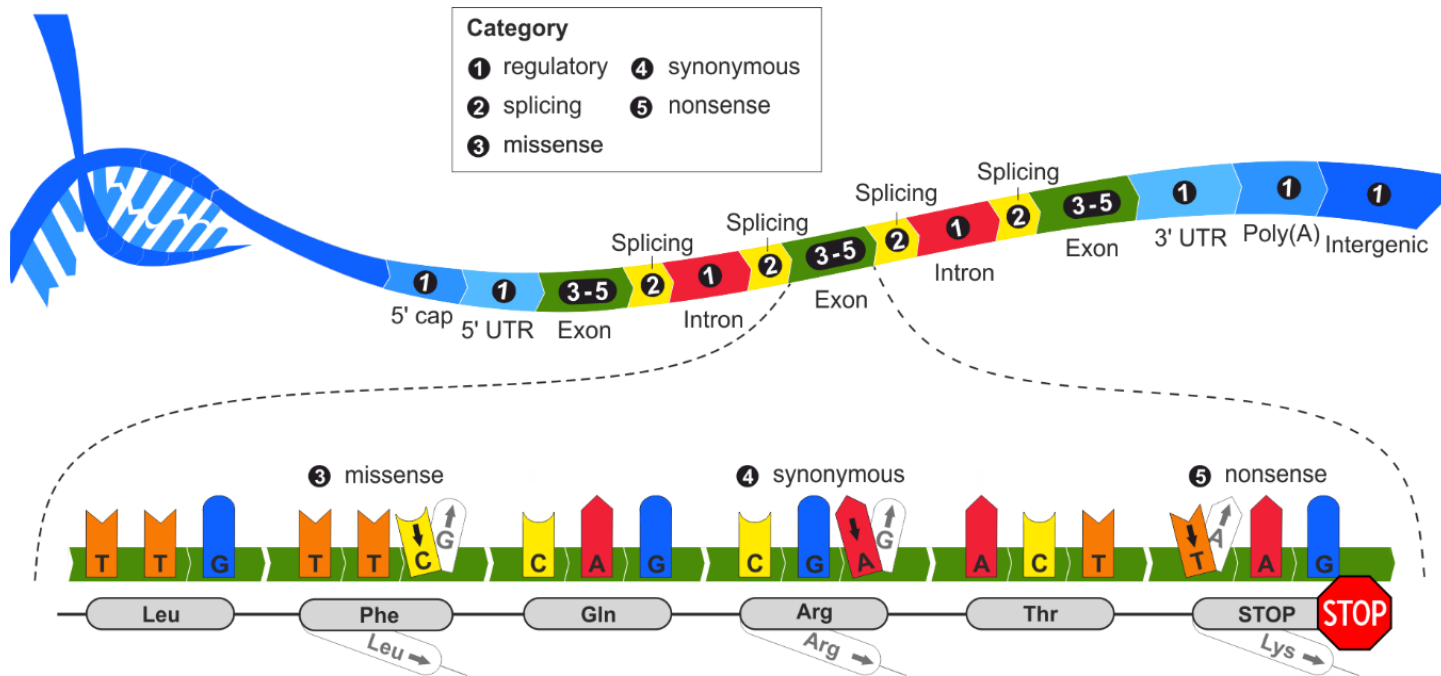
❑ Can be harmful or not



Overview of mutations

□ Location in the DNA

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



Overview of mutations



□ Types

- **Point mutations** – a single nucleotide is changed in DNA (or RNA)
 - **Substitutions**
 - **Single nucleotide polymorphism (SNP – pronounced “snip”)**
 - 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
- **Other types** (duplications, translocations, inversions, etc.)

Point mutations at protein level



□ Types of point mutations

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

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

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

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

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

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

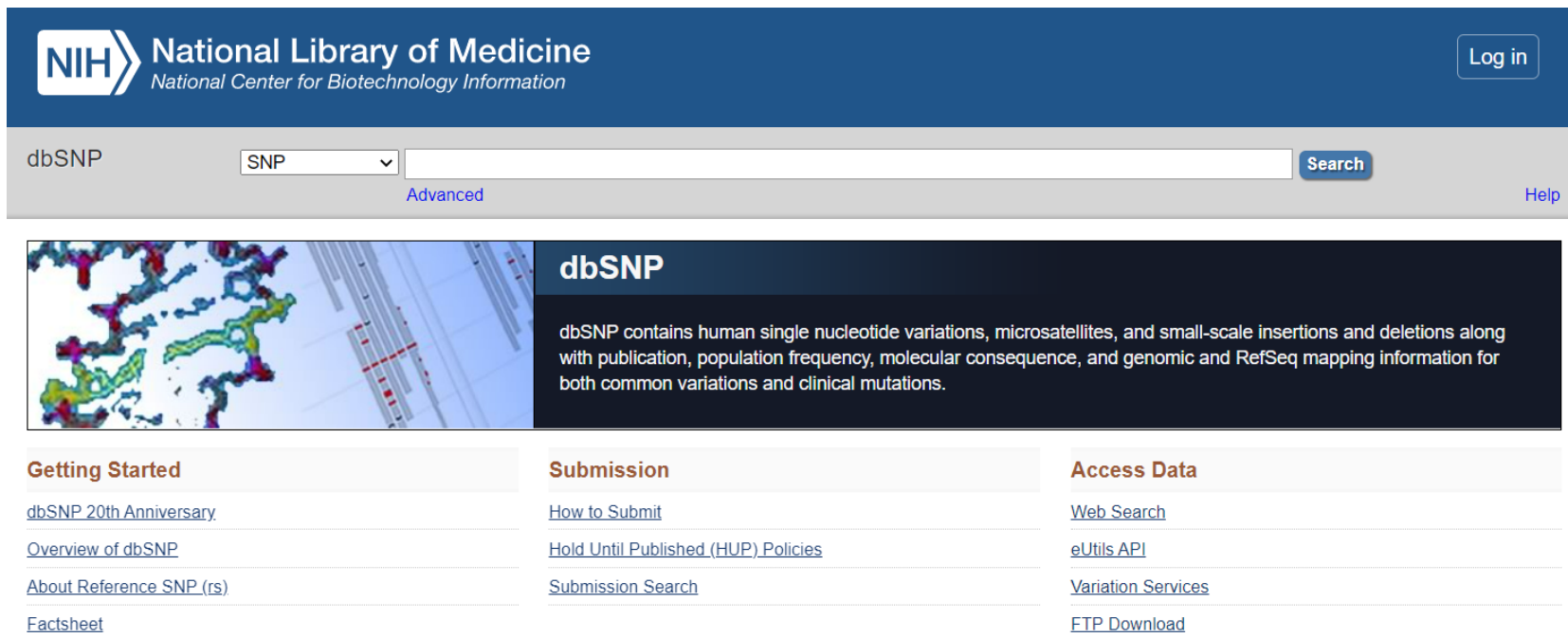


- ❑ **Human Genome Variation Society**
 - <http://www.hgvs.org>
 - Lists all the available databases of human mutations
- ❑ **Central mutation databases (>20)**
 - Substitutions in all genes
 - Variability in protein sequences
 - Data mainly from literature
- ❑ **Locus-specific databases (about 700)**
 - 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



NIH National Library of Medicine
National Center for Biotechnology Information

Log in

dbSNP SNP Search

Advanced Help

dbSNP

dbSNP contains human single nucleotide variations, microsatellites, and small-scale insertions and deletions along with publication, population frequency, molecular consequence, and genomic and RefSeq mapping information for both common variations and clinical mutations.

Getting Started

- [dbSNP 20th Anniversary](#)
- [Overview of dbSNP](#)
- [About Reference SNP \(rs\)](#)
- [Factsheet](#)

Submission

- [How to Submit](#)
- [Hold Until Published \(HUP\) Policies](#)
- [Submission Search](#)

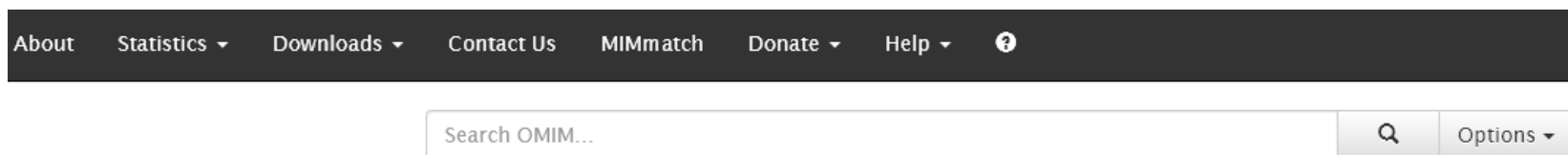
Access Data

- [Web Search](#)
- [eUtils API](#)
- [Variation Services](#)
- [FTP Download](#)

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

❑ Human Gene Mutation Database - HGMD

- <http://www.hgmd.cf.ac.uk/ac/index.php>
- Comprehensive collection of mutations in nuclear genes that underlie or are associated with human inherited disease

The Human Gene Mutation Database (HGMD®) represents an attempt to collate all known (published) gene lesions responsible for human inherited disease and is maintained in Cardiff by D.N. Cooper, E.V. Ball, P.D. Stenson, A.D. Phillips, K. Evans, S. Heywood, M.J. Hayden, M.M. Chapman, M.E. Mort, L. Azevedo and D.S. Millar.

Get HGMD Professional Please note that this latest up-to-date public version of our database is freely available only to registered users from academic institutions/non-profit organisations. All commercial users are required to purchase a license from QIAGEN®, our commercial partner. A license to HGMD Professional is available to both commercial and academic non-profit users wishing to access the most up-to-date version of the database (visit QIAGEN® to request a free trial of HGMD Professional). Read more about how HGMD is funded. You may not copy, store or re-distribute HGMD data without express written permission (i) from the curators or (ii) via your license agreement. Copyright © Cardiff University 2017. All rights reserved.

[Register for Public Version](#)

| Table: | Description: | Public entries: <small>This site Academic non-profit users only</small> | Total entries: <small>HGMD Professional 2019.4</small> |
|---|--|--|---|
| Mutation totals (as of 2020-12-10) | | 189186 | 275716 |
| Gene symbol | The gene description, gene symbol (as recommended by the HUGO Nomenclature Committee) and chromosomal location is recorded for each gene. In cases where a gene symbol has not yet been made official, a provisional symbol has been adopted which is denoted by lower-case letters. | 7677 | 10902 |
| cDNA sequence | cDNA reference sequences are provided, numbered by codon. | 7729 | 11079 |
| Genomic coordinates | Genomic (chromosomal) coordinates have been calculated for missense/nonsense, splicing, regulatory, small deletions, small insertions and small indels. | 0 | 250578 |
| HGVS nomenclature | Standard HGVS nomenclature has been obtained for missense/nonsense, splicing, regulatory, small deletions, small insertions and small indels. | 0 | 250862 |
| Missense/nonsense | Single base-pair substitutions in coding regions are presented in terms of a triplet change with an additional flanking base included if the mutated base lies in either the first or third position in the triplet. | 106004 | 159705 |
| Splicing | Mutations with consequences for mRNA splicing are presented in brief with information specifying the relative position of the lesion with respect to a numbered intron donor or acceptor splice site. Positions given as positive integers refer to a 3' (downstream) location, negative integers refer to a 5' (upstream) location. | 17183 | 23868 |
| Regulatory | Substitutions causing regulatory abnormalities are logged in with thirty nucleotides flanking the site of the mutation on both sides. The location of the mutation relative to the transcriptional initiation site, initiator codon, polyadenylation site or termination codon is given. | 3544 | 4575 |
| Small deletions | Micro-deletions (20 bp or less) are presented in terms of the deleted bases in lower case plus, in upper case, 10 bp DNA sequence flanking both sides of the lesion. The numbered codon is preceded in the given sequence by the caret character (^). | 28155 | 39822 |
| Small insertions | Micro-insertions (20 bp or less) are presented in terms of the inserted bases in lower case plus, in upper case, 10 bp DNA sequence flanking both sides of the lesion. The numbered codon is preceded in the given sequence by the caret character (^). | 11745 | 16881 |
| Small indels | Micro-indels (20 bp or less) are presented in terms of the deleted/inserted bases in lower case plus, in upper case, 10 bp DNA sequence flanking both sides of the lesion. The numbered codon is preceded in the given sequence by the caret character (^). | 2679 | 3652 |
| Gross deletions | Information regarding the nature and location of each lesion is logged in narrative form because of the extremely variable quality of the original data reported. | 14186 | 19491 |
| Gross insertions | Information regarding the nature and location of each lesion is logged in narrative form because of the extremely variable quality of the original data reported. | 3445 | 4945 |
| Complex rearrangements | Information regarding the nature and location of each lesion is logged in narrative form because of the extremely variable quality of the original data reported. | 1747 | 2231 |
| Repeat variations | Information regarding the nature and location of each lesion is logged in narrative form because of the extremely variable quality of the original data reported. | 498 | 546 |

9,438,337 queries successfully served since 2007.

Central mutation databases



□ UniProtKB/Swiss-Prot

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

UniProtKB 251,702,059 results

| Entry | Entry Name | Protein Names | Gene Names | Organism | Length |
|-------------------------------------|-------------|---|--------------------------------------|----------------------|----------|
| <input type="checkbox"/> A0A0C5B5G6 | MOTSC_HUMAN | Mitochondrial-derived peptide MOTS-c[...] | MT-RNR1 | Homo sapiens (Human) | 16 AA |
| <input type="checkbox"/> A0A1B0GTW7 | CIROP_HUMAN | Ciliated left-right organizer metalloproteinase[...] | CIROP, LMLN2 | Homo sapiens (Human) | 788 AA |
| <input type="checkbox"/> A0JNW5 | BLT3B_HUMAN | Bridge-like lipid transfer protein family member 3B [...] | BLTP3B, KIAA0701, SHIP164, UHRF1BP1L | Homo sapiens (Human) | 1,464 AA |
| <input type="checkbox"/> A0JP26 | POTB3_HUMAN | POTE ankyrin domain family member B3 | POTEB3 | Homo sapiens (Human) | 581 AA |
| <input type="checkbox"/> A0PK11 | CLRN2_HUMAN | Clarin-2 | CLRN2 | Homo | 232 AA |

Locus-specific databases

☐ For information on gene-specific databases

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

Missense mutations



- ❑ What are they?...
- ❑ How can they affect proteins?



Missense mutations



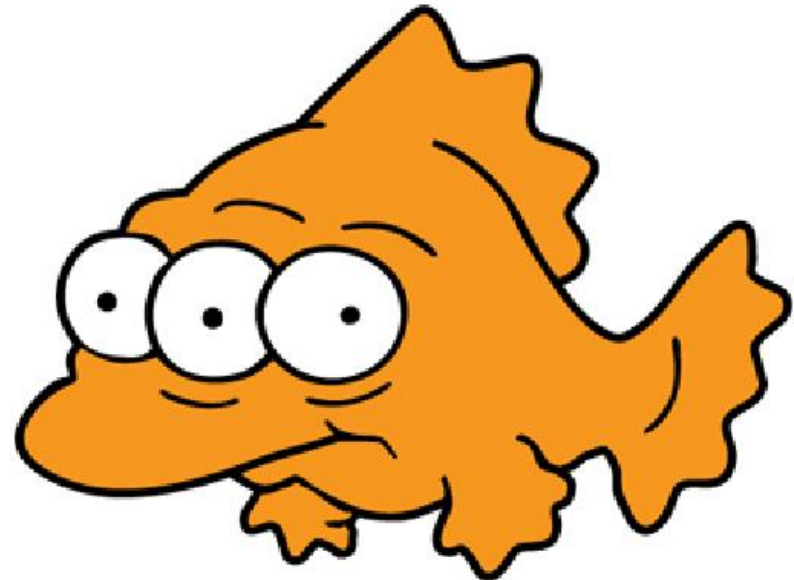
- ❑ Mutations affecting structure
 - Stability & folding
 - Aggregation
- ❑ Mutations affecting function
 - Binding & catalysis
 - Transport processes
 - Protein dynamics
 - Protein localization

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
- Increased **aggregation**

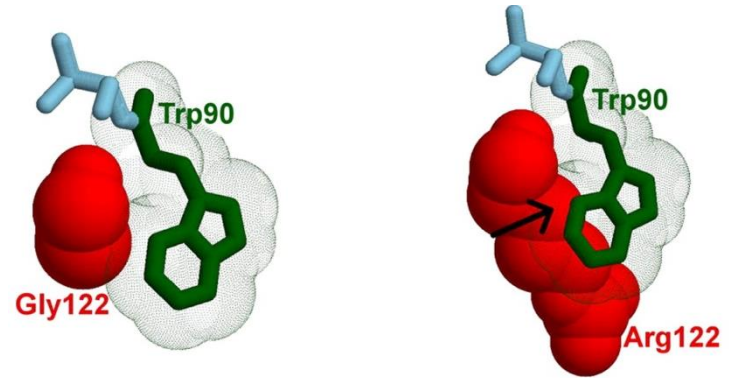


Mutations affecting structure

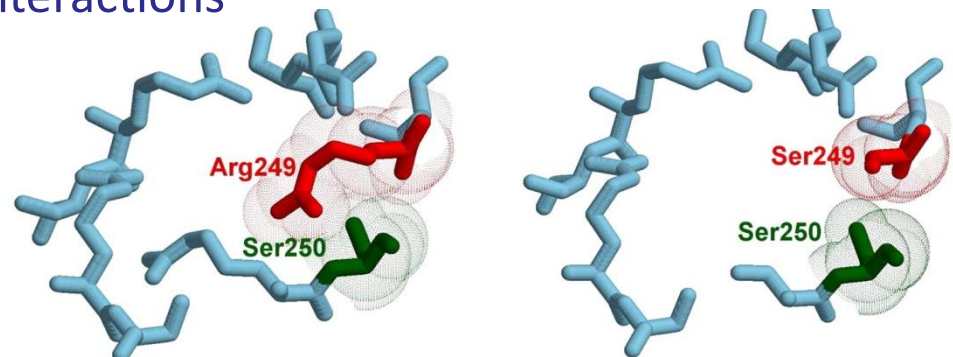


□ Molecular basis of mutations affecting folding & stability

- **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

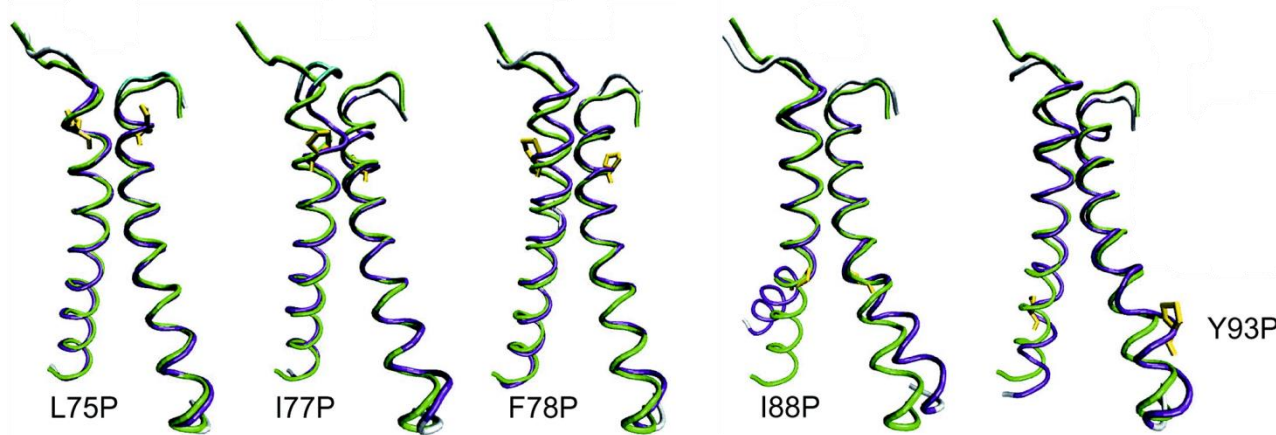


Mutations affecting structure



□ Molecular basis of mutations affecting folding & stability

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



NOTE:

- Glycine – the most flexible amino acid
- Proline – the most rigid

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

Mutations affecting structure

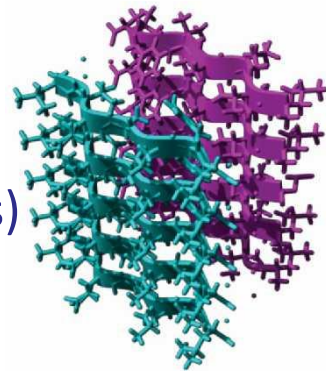


❑ Mutations can reduce solubility or increase aggregation

- Alterations on the surface residues may affect the solubility (ex: reduction of **charge**)
- **Hydrophobic** mutations can increase protein aggregation
- Aggregating proteins usually have high level of β -structures

❑ Aggregation modulated by short specific sequences

- **Aggregation-prone regions (APRs)** 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

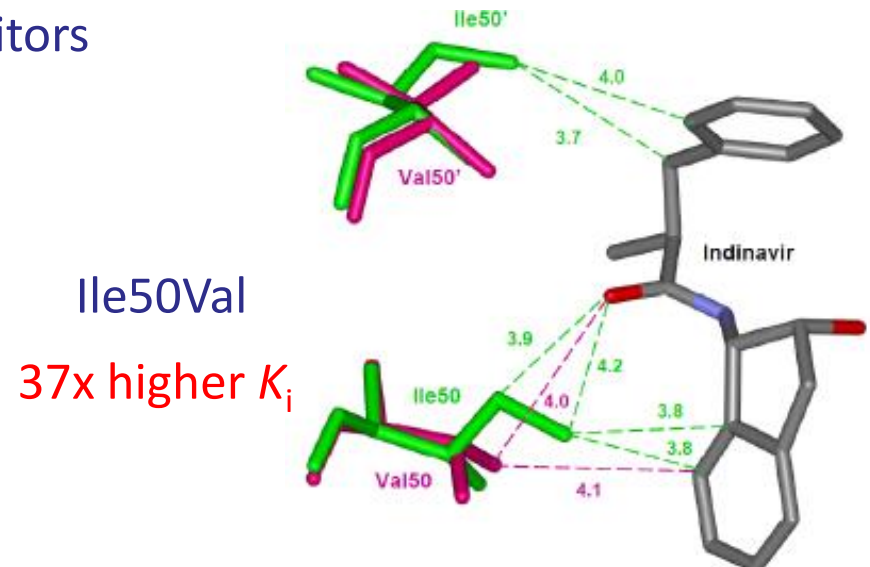
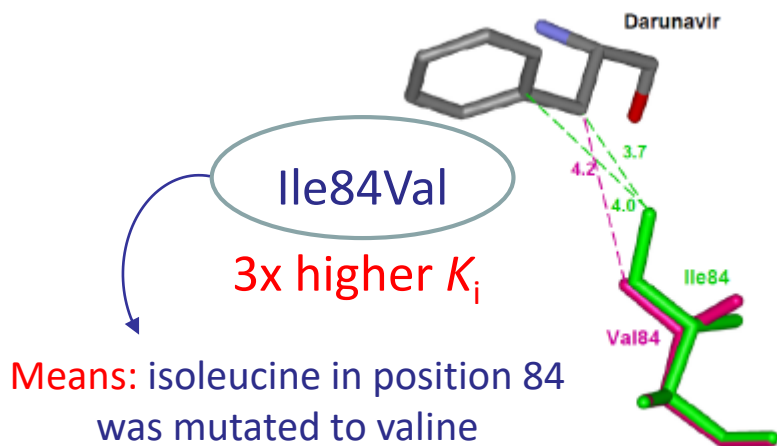


□ Effect on binding and catalysis

- Binding sites are tuned to bind specific molecules and stabilize transition states
- Mutations can **improve** or **disrupt the binding** and **catalysis**

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

- Loss of interactions with inhibitors

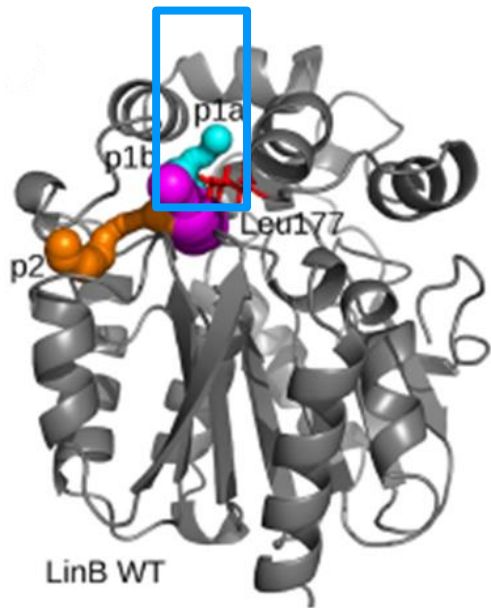


Mutations affecting function



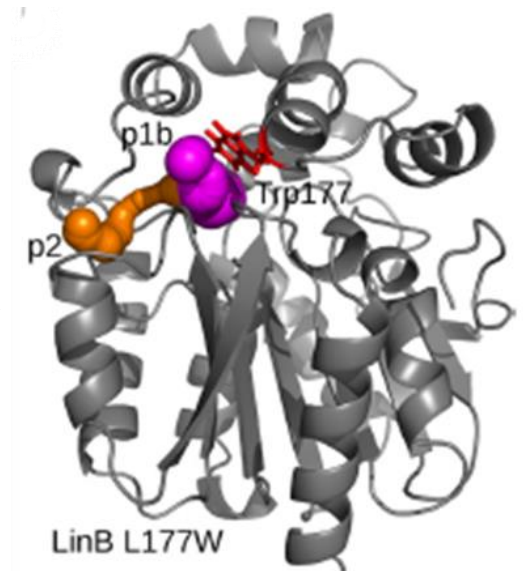
□ Effect on ligand transport

- Pathways are adjusted to permit transport of specific molecules
- Mutations can **speed-up or disrupt** their transport or allow the transport of different molecules



Leu177Trp => **tunnel**
becomes almost *closed*

release of products 500x slower



Mutations affecting function



□ Effect on protein dynamics

- Dynamics enables proteins to adapt to their binding partners and interchanging between conformations
- Mutations can:
 - Make regions more rigid (targeting hinge or very mobile regions, ex.: loops) -> **reduced adaptability**
 - Increase flexibility of rigid regions (targeting residues with many contacts in mobile elements) -> **increased adaptability**
- These change **may affect activity, specificity** or even **recognition**

Mutations affecting function



□ Effect on protein localization

- After translation, the protein must be translocated to the appropriate cellular compartment
- Translocation can be regulated by short sequences (Signal Peptides) on the **N-terminus**, by Translocation Complexes, Chaperones, etc.
- Mutations can **disrupt** or alter the **signal**, or **complex formation** -> 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



- ❑ What is it?



Identification of mutable residues



- The effect of mutations on the protein can be predicted directly 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**



□ 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 involving proline or glycine -> **altered conformation**



□ Mutations affecting function

- Mutation of residues in binding or active sites -> **modify binding or catalysis**
- Mutation of residues in transport pathways -> **modify transport**
- Mutation of hinge or mobile residues, residues on loops with many contacts -> **modify 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 – ex.: ConSurf, ...
 - Residue contacts – ex: Contact Map Web Viewer, ...
 - Residue interactions – ex: Protein Interaction Calculator, ...
 - Accessible surface area – ex: AsaView, Naccess, ...
 - Binding sites – ex: CASTp, metaPocket 2.0, meta-PPISP, ...
 - Transport pathways – ex: CAVER 3.0, POREWALKER, ...
 - Protein dynamics – ex: NMA, molecular dynamics, ...
 - Protein localization – ex: SignalP, TargetP, Phobius, TMHMM, ...

Identification of mutable residues

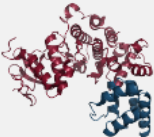
- ❑ **HotSpot Wizard – meta-server combining several tools**
 - <http://loschmidt.chemi.muni.cz/hotspotwizard/>
 - Homology modelling, MSA, conservation, correlation, pockets and tunnels detection, docking, stability prediction, design of smart library

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
VVLVVHDWGSALRGL

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.

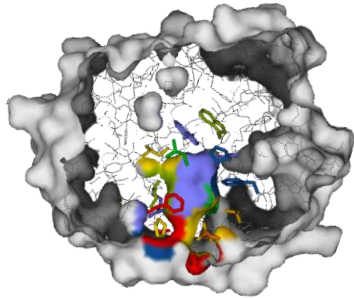
PubMed OPEN ACCESS

USER STATISTICS

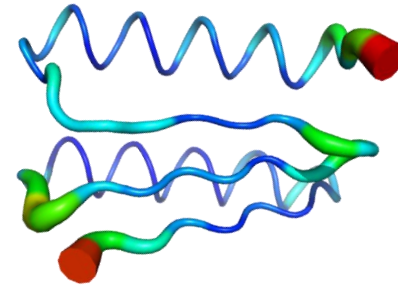
- Number of visitors: 58946
- Number of jobs: 32197

Identification of mutable residues

Functional hot-spots



Stability hot-spots (flexibility)



Stability hot-spots (evolution)

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

Y ⇒ R

Correlated hot-spots

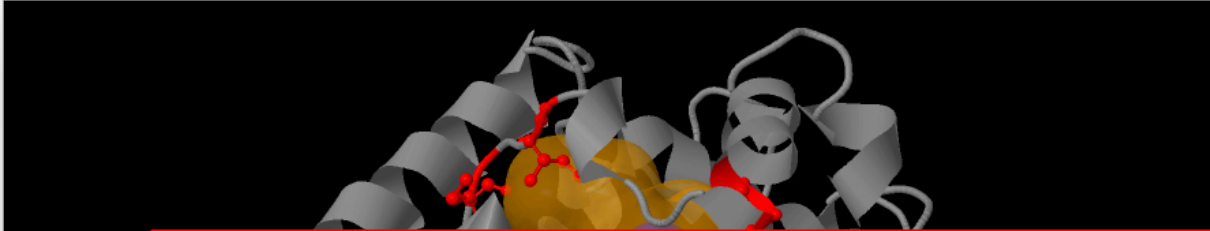
| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| T | S | S | R | L | W | Y | N | I | D | P | N | H | C | A | G | L |
| - | - | S | R | L | W | R | N | I | D | - | - | H | C | A | G | L |
| T | S | S | R | L | W | R | N | I | D | P | N | H | C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

↑ ↑

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

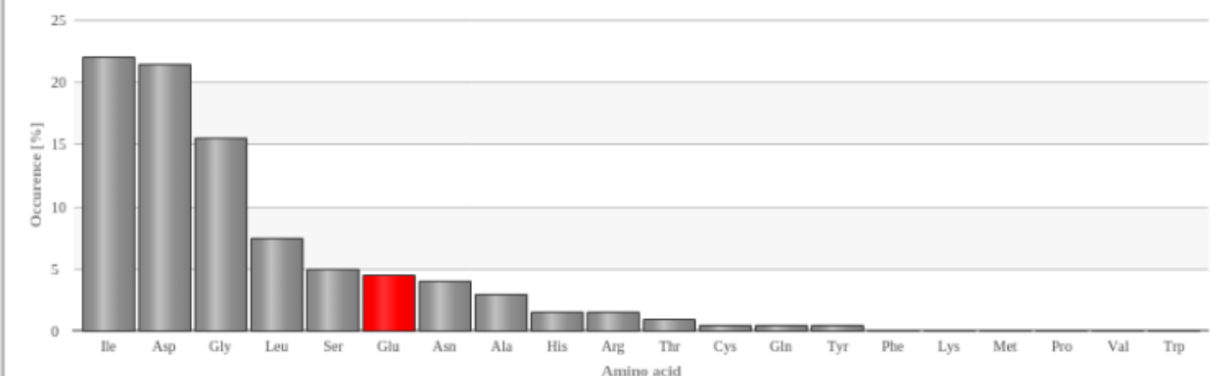
Hide all visualized residues

Save image

Reset view

Details for residue A:78 Glu

Overview Tunnels & pockets **Amino acid frequencies** Mutational landscape Correlated positions



Number of gaps: 23 (11.5%)
Total number of sequences: 200

Residue features

Exclude correlated residues

Exclude buried residues Include residues with moderate mutability

| | chain | position | residue | mutable | non-essential | in tunnel | in catalytic pocket | HotSpot |
|-------------------------------------|-------|----------|---------|---------|---------------|-----------|---------------------|---------|
| <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 | ✓ | ✓ | ✓ | ✓ | ✓ |

mutagenesis

Design mutations

Design library

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

Visualization settings (continued)

bottleneck radius (Å): 1.5

| relevance (%) | volume (Å ³) |
|---------------|--------------------------|
| 100 | 576 |
| 82 | 883 |
| 62 | 275 |
| 28 | 753 |

Identification of mutable residues

Design mutations

Single Point Multiple Point Results summary

Stabilizing mutations Destabilizing mutations
Energy is in kcal/mol

| chain | position | residue | Ala | Arg | Asn | Asp | Cys | Gln | Glu | Gly | His | Ile | Lys |
|-------|----------|---------|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| A | 249 | Thr | 0.4 | - | - | - | - | - | - | - | - | - | - |
| A | 145 | Glu | -2.1 | - | - | - | - | - | - | - | - | - | - |
| A | 138 | Ile | 7.6 | - | - | - | - | - | - | - | - | - | - |
| A | 248 | Leu | 6.2 | - | - | - | - | - | - | - | - | - | - |
| A | 173 | Val | 5.1 | - | - | - | - | - | - | - | - | - | - |
| A | 177 | Leu | 4.4 | - | - | - | - | - | - | - | - | - | - |
| A | 146 | Gln | -0.4 | - | - | - | - | - | - | - | - | - | - |
| A | 253 | Met | 6.7 | - | - | - | - | - | - | - | - | - | - |
| A | 147 | Asp | -3.5 | - | - | - | - | - | - | - | - | - | - |
| A | 136 | Met | 4.3 | - | - | - | - | - | - | - | - | - | - |

Export table to CSV

Evaluate multiple point stability

Codon usage : Escherichia coli K12

Generate report

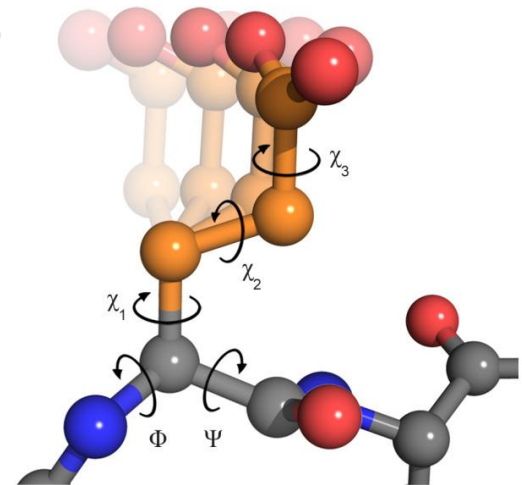
Prediction of effects on structure



- **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** ($\Delta\Delta G = \Delta G^{\text{Mut}} - \Delta G^{\text{Native}}$)

- **Available tools**

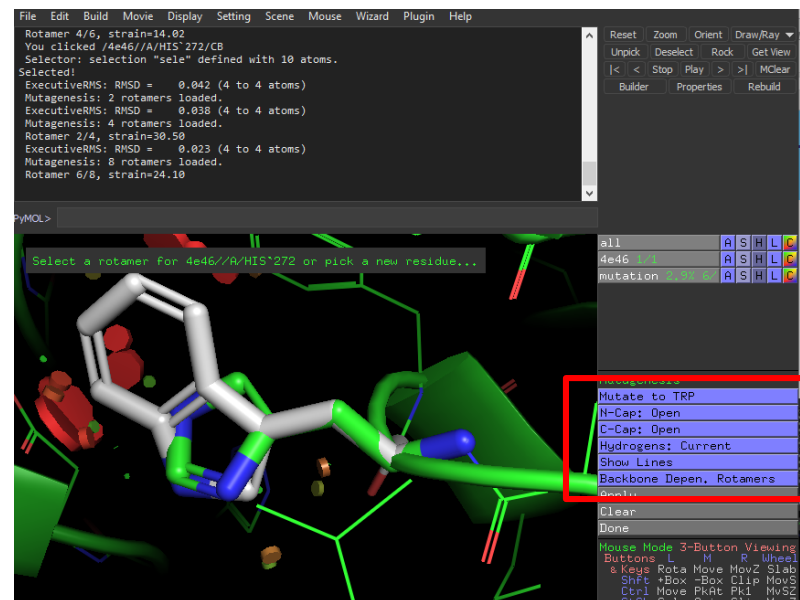
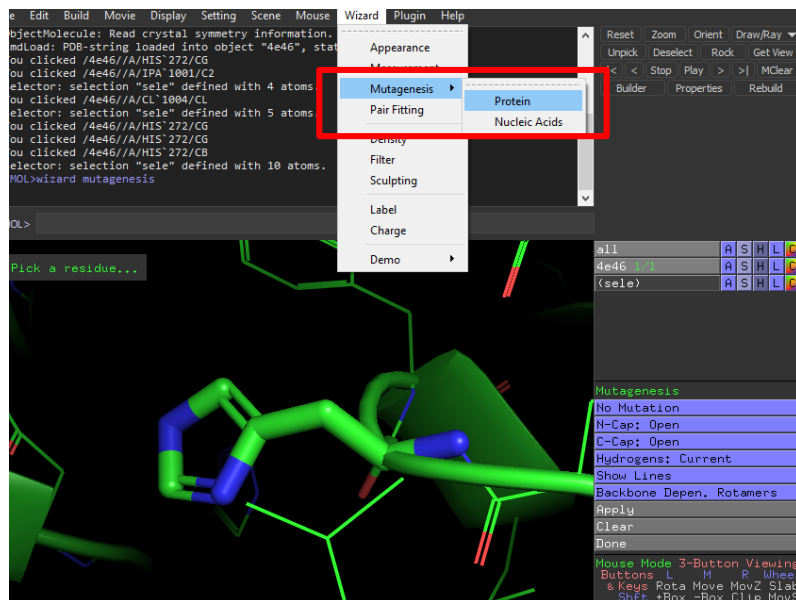
- Geometric: PyMOL; WhatIF
- Energy-based: FOLDX, Rosetta-ddG
- Homology: Swiss Model, MODELLER, etc.



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

□ 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, or stabilizing Proline mutations
- No scoring function

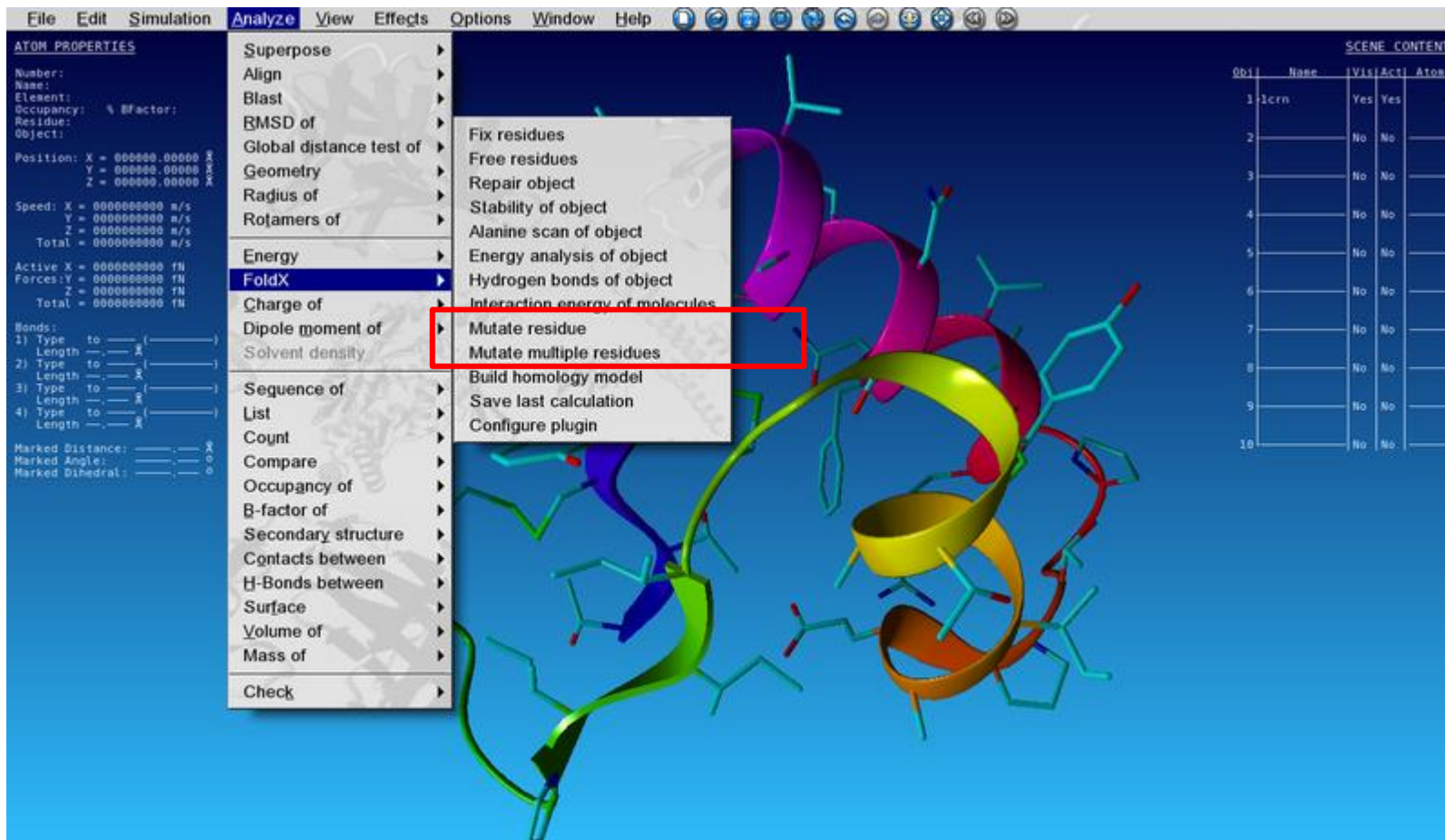
Prediction of effects on structure

□ FOLDX

- <http://foldxsuite.crg.eu/>
- **Stand alone**, with plug-in to Yasara modeling tool
- **Fast** (minutes)
- **Fixed backbone** conformation
- Construction of **single** or **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 a sub-menu is visible, highlighting 'Mutate residue' and 'Mutate multiple residues' with 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 | Vis | Act | Atom |
|-----|------|-----|-----|------|
| 1 | icrn | 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

□ Rosetta-ddG

- Under <https://www.rosettacommons.org/>
- **Stand alone** with bash and python scripts available
- **Slow** (hours-days)
- **Fixed** or **flexible backbone** conformation
- Construction of **single** or **multiple mutants**
- **Empirical force field** for calculating structure and stability of wild-type and mutant
- Construction of **PDB** and prediction of **stability change** ($\Delta\Delta G$)



- **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

❑ PredictSNP:

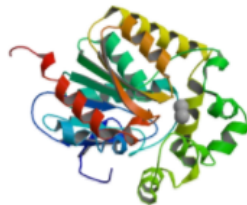
- ❑ <http://loschmidt.chemi.muni.cz/predictsnp/>
- ❑ Combines many tools for Protein or DNA assessment of SNPs



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:

```
>HSA_HUMAN
MVLSPADKTNVAAAGKVGAHAGEYGAEALERFLEFFTKRYEYHFDLHNSGAQVKGHG
KRVADALINAVRVDGDFWALSGLDRAHRLRVDFVNFLLSHCLLVTLAHLPAEFTF
AVHSLDKFLASVSTVLTSTYR
```

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 | ✖ |

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 | 67.8% |

JOB CONTROL

Job ID:

REFERENCE

Bendl, J., Stourac, J., Salanda, O., Pavelka, A., Weben, E.D., Zendluka, J., Brezovsky, J., Damborsky, J., 2014: PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. *F10S 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

- 4,450 neutral / 7,539 deleterious
- Links: [XLS](#), [dataset statistics](#)

OVERFIT testing dataset

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

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://roslab.org/services/snap/ |
| SNPs&GO | SVM | Swiss-Prot | Sequence environment, sequence profiles, Panther | – | GO | http://snps-and-go.biocomp.unibo.it/snps-and-go/ |



- ❑ **Protein engineering**: sometimes we can use mutagenesis to rationally design proteins according to our needs
- ❑ Properties that can be modified by mutagenesis
 - **Such as?...**





- ❑ **Protein engineering**: sometimes we can use mutagenesis to rationally design proteins according to our needs
- ❑ Properties that can be modified by mutagenesis
 - **Stability**
 - **Function**
 - Binding site (catalytic activity or substrate specificity)
 - Macromolecular interface
 - Molecular tunnels/channels
 - **Solubility**



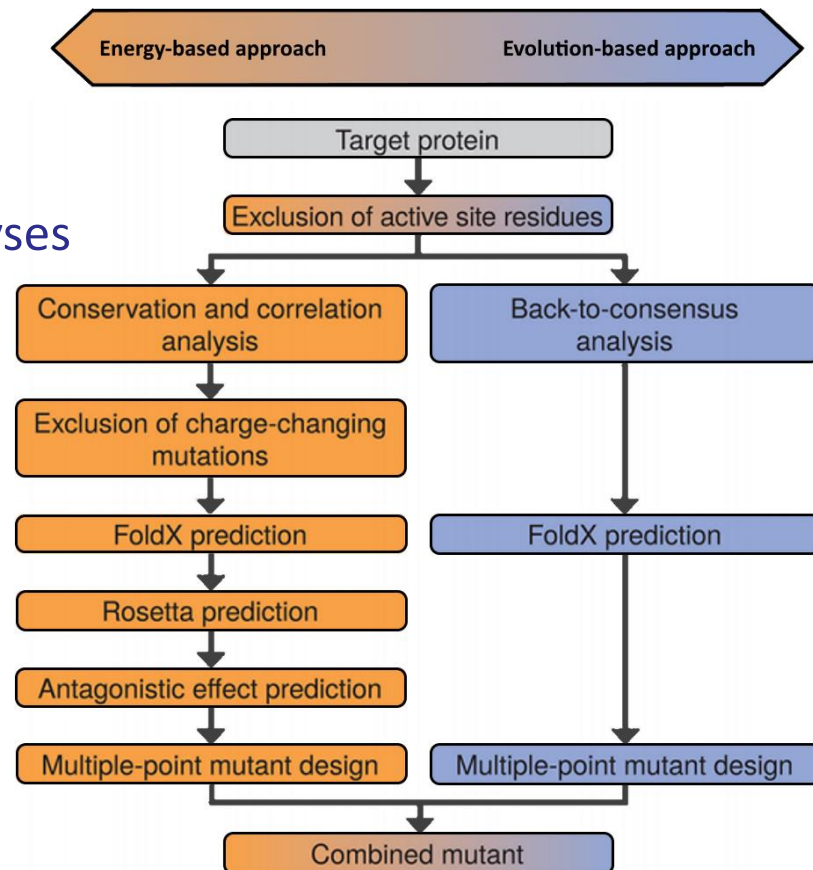
- ❑ **Prediction of stability change upon mutation**
 - Structure of mutant protein may not be produced
 - Tools often employ
 - Empirical scoring functions
 - Evolutionary conservation analysis (ex: back-to-consensus)
 - Machine learning approaches

- ❑ Available tools
 - Energy-based: Rosetta-ddG, FOLDX ✓
 - Evolution-based: FireProt^{ASR}
 - Hybrid approaches: FireProt, PROSS

Rational design: stability

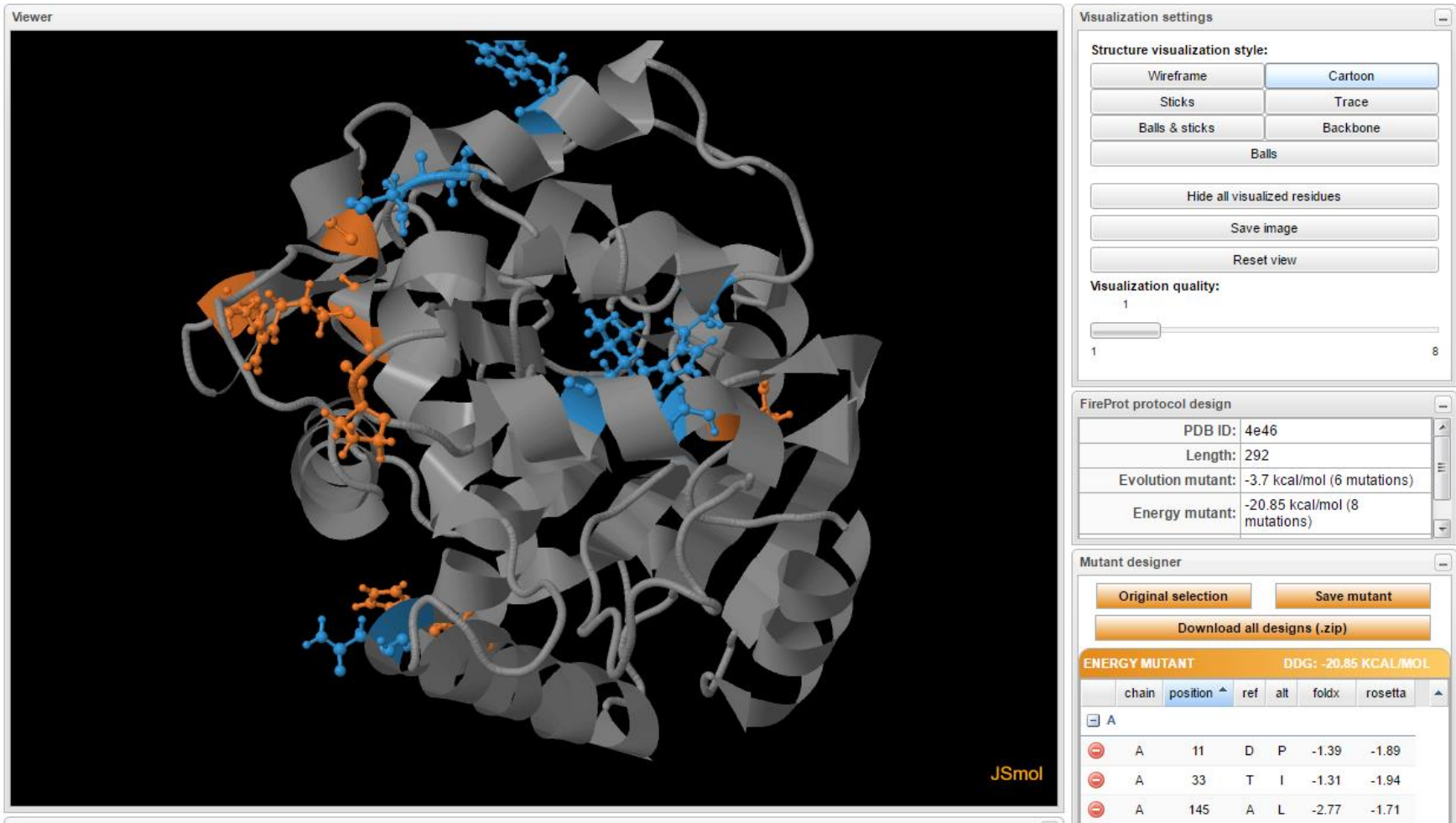
□ FireProt

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



Rational design: stability

□ FireProt



The screenshot displays the FireProt software interface. The main window shows a 3D visualization of a protein structure (PDB ID: 4e46) in a cartoon style, with some residues highlighted in orange and blue. The interface includes several panels:

- Viewer:** The main 3D visualization area.
- Visualization settings:** Controls for structure visualization style (Wireframe, Cartoon, Sticks, Trace, Balls & sticks, Backbone, Balls), buttons for "Hide all visualized residues", "Save image", and "Reset view", and a "Visualization quality" slider (set to 1).
- FireProt protocol design:** A table showing design parameters.
- Mutant designer:** Buttons for "Original selection", "Save mutant", and "Download all designs (.zip)", along with a table of energy mutants.

| 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 |

Rational design: stability

□ FireProt

Mutations

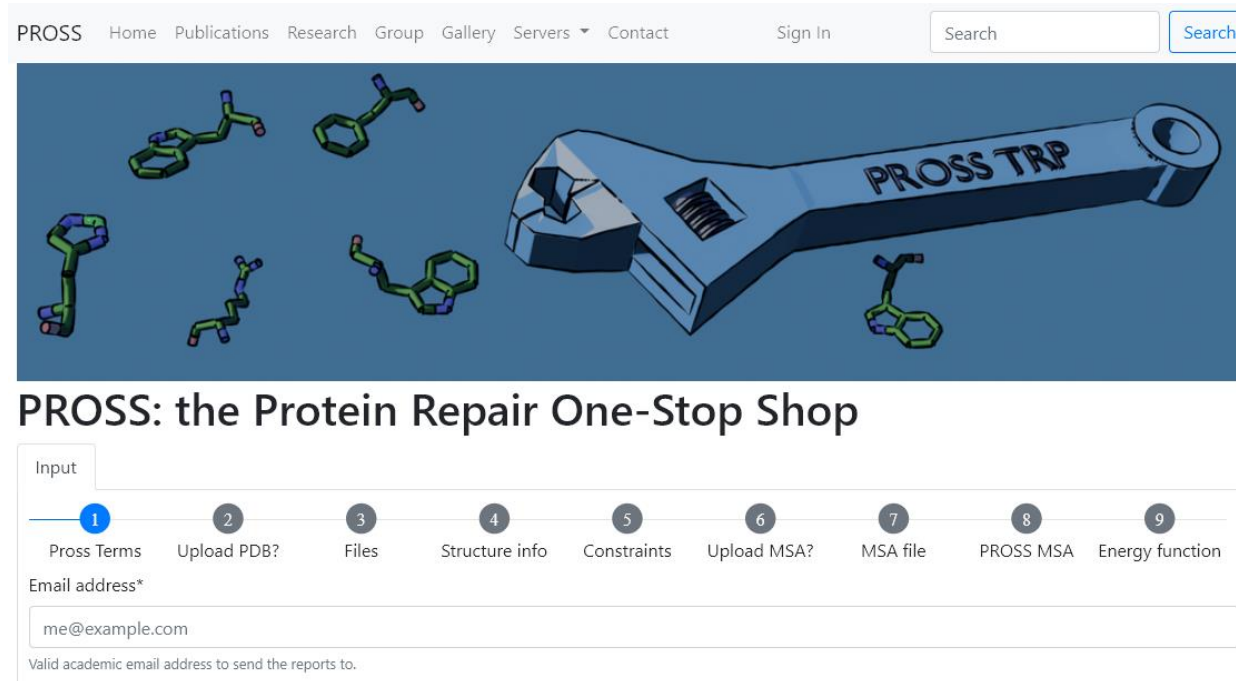
Combined mutant | Energy mutant | Evolution mutant | Wild-type

| 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 |

Rational design: 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

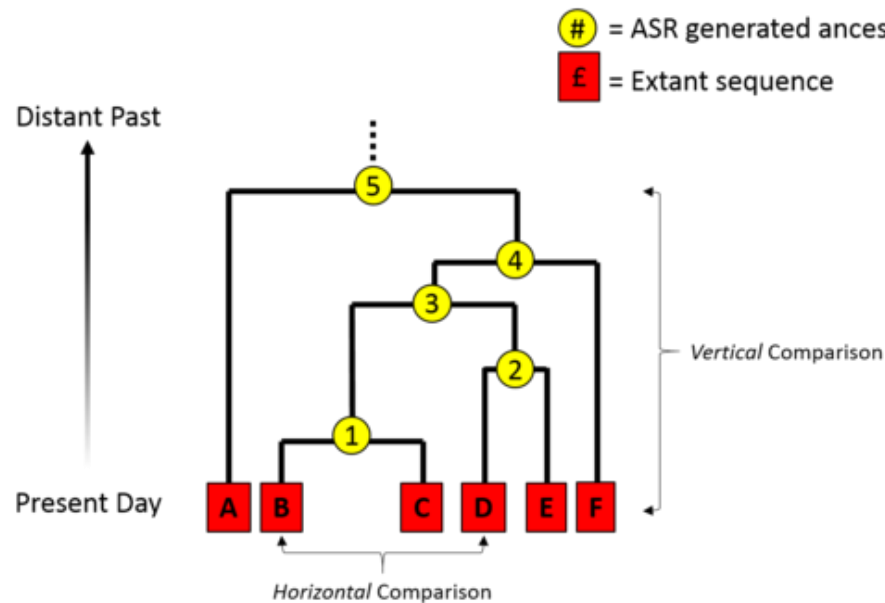
- 1 Pross Terms
- 2 Upload PDB?
- 3 Files
- 4 Structure info
- 5 Constraints
- 6 Upload MSA?
- 7 MSA file
- 8 PROSS MSA
- 9 Energy function

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

Rational design: stability

□ FireProt^{ASR}

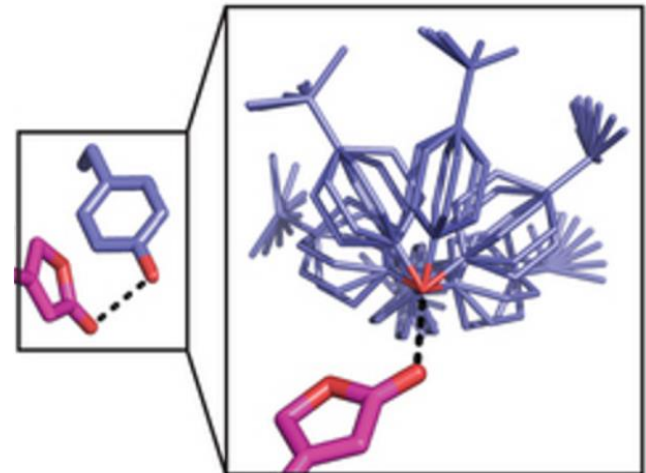
- <https://loschmidt.chemi.muni.cz/fireprotasr>
- Ancestral sequence reconstruction (ASR)
- Automated ancestral inference & phylogenetic tree
- Useful to find stable ancestral enzymes





□ RosettaDesign

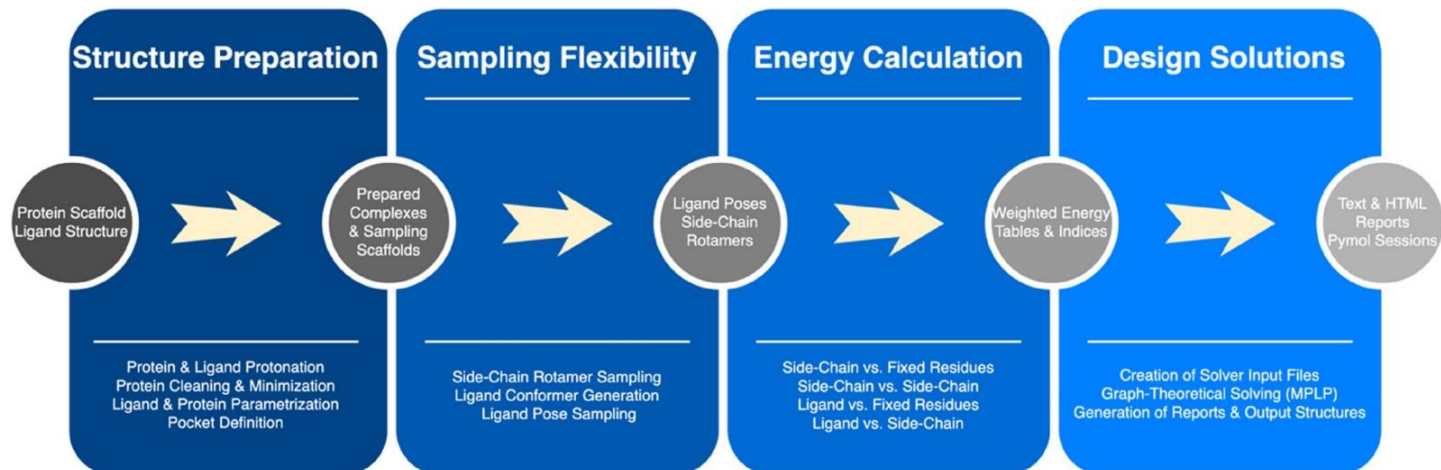
- <http://rosettadesign.med.unc.edu/>
- Monte Carlo sampling (random search) to predict minimum-energy structure of mutants
- Predicts free energy changes upon mutations ($\Delta\Delta G$)
- Helps design mutations to **optimize the binding site** and increase interactions with a ligand/substrate



Rational design: function

□ PocketOptimizer

- <https://github.com/Hoecker-Lab/pocketoptimizer/>
- Aimed at maximizing the affinity of a binding site towards a ligand
- Modular pipeline with different tools
 - Flexibility, docking, mutagenesis, energy calculation
 - Predicts global minimum-energy designs



Rational design: function

□ FuncLib

- <https://funclib.weizmann.ac.il>
- To redesign and/or optimize **binding site**
- Utilizes evolution (conservation) and Rosetta calculations (energy) to **introduce multiple-point mutations** to modify the properties of the binding site
- Can be used to improve the binding affinity towards a ligand
- Outputs up to 50 multiple-point mutants for protein synthesis

Rational design: function

□ 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 | | | | | | | | | | | | | | |

Rational design: function

□ AffiLib

- <https://affilib.weizmann.ac.il>
- To optimize **protein-protein interface**
- Utilizes evolution (conservation) and Rosetta (energy) to **introduce mutations** and optimize macromolecular interface
- Suggests mutations on the interface residues to improve the binding affinity
- Outputs up to 50 multiple-point mutants for protein synthesis

□ Mutation Cutoff Scanning Matrix (mCSM-PPI2)

- http://biosig.unimelb.edu.au/mcsm_ppi2/
- To optimize **protein-protein interface**
- 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
 - **Alanine scanning** (all interface residues are mutated to alanine)
 - **Systematic** – position saturation (all interface residues are mutated to all other 19 amino acids)

Rational design: solubility



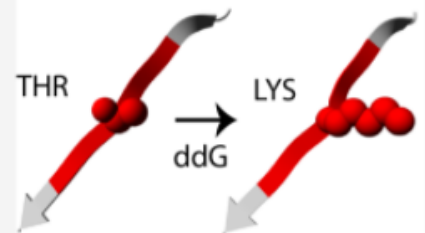
- ❑ **Aggrescan3D; SoluProt** (see lecture 6 - Analysis of protein structures)
- ❑ **SolubiS**
 - <https://solubis.switchlab.org/>
 - To identify stabilizing mutations that reduce the aggregation tendency of a protein
 - 1) Identifies **exposed APRs**
 - 2) Introduces **“gatekeeper”** residues (P, R, K, D and E) into APRs
 - 3) Assesses the stability changes of mutations ($\Delta\Delta G$)

MVSKGEEDNMAIIKEFMRFKVVHMEGSVNGHEE-
IEGEGEGRPYEGTQTAKLKVTKGGPLP-
FAWDISPQFMYGSKAYVKHPADIPDYLLKLSFPEGFKWE
RVMNFDGQVWTVTQDSSLQDGEFIYKVKL
RGTNFPDGPVMQKKTMGWEASSERMYPEDG ALK-
GEIKQRLKLDGGHYDAEVKTTYKAKVQLPGAYN-
VNIKLDITSHNEDYTIVEQYERAEGRHSTGGMDELYK



PVTVT

| | |
|-------|------|
| PVTVT | WTVK |
| KVTVT | WTPT |
| VRTVT | WTVR |
| VETVT | WTKT |
| VDTVT | WEVT |



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