Engineering of protein structures
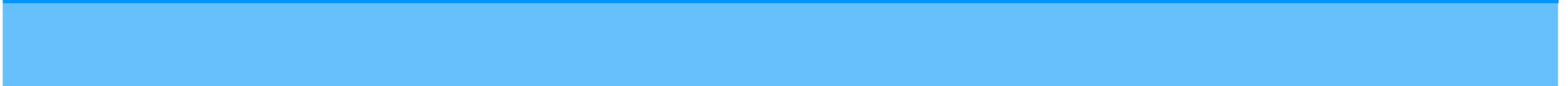
Outline

- Overview of mutations

- Databases of mutations

- Missense mutations

- Prediction of mutational effects

- Rational design of proteins

❑ **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
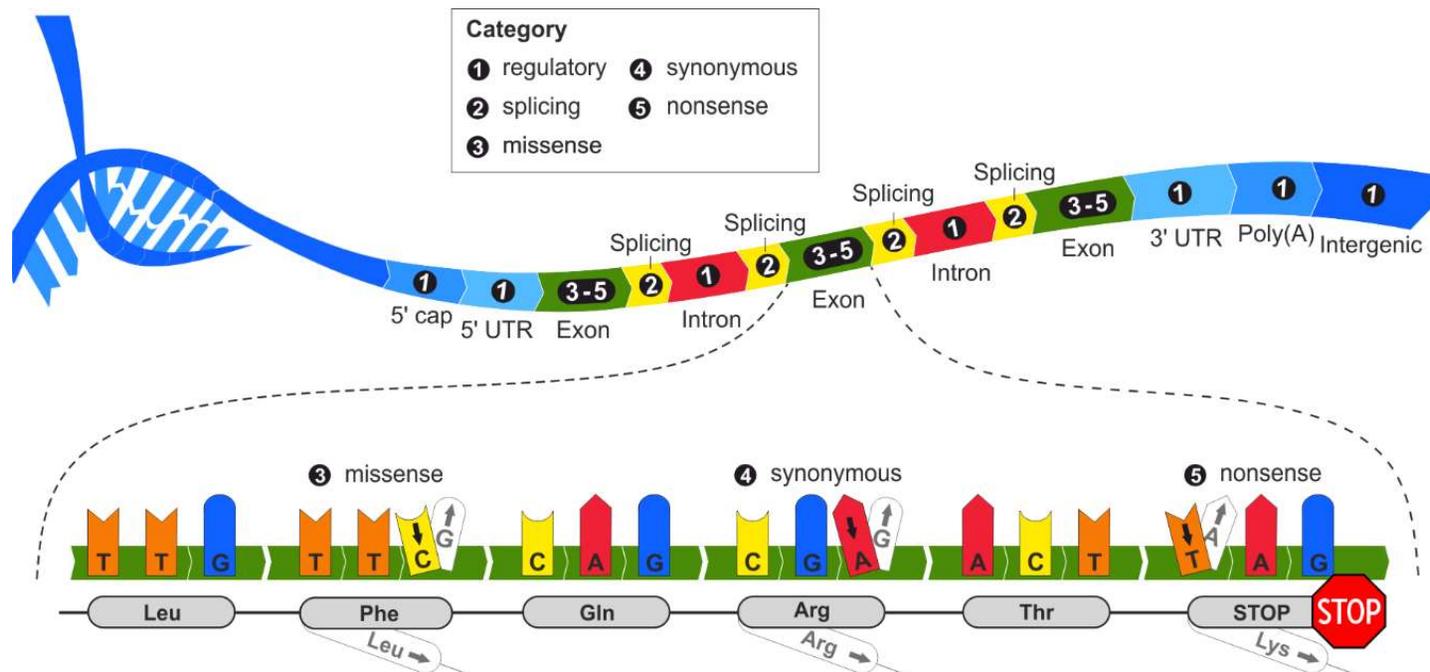
Overview of mutations

- ❑ **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

Central mutation databases

❑ **Online Mendelian Inheritance in Man – OMIM**

- ▪ http://omim.org/

- ▪ comprehensive database of human genes and genetic phenotypes

| About | Statistics ▾ | Downloads ▾ | Contact Us | MIMmatch | Donate ▾ | Help ▾ | ❓ |

| Search OMIM... | 🔍 | Options ▾ |
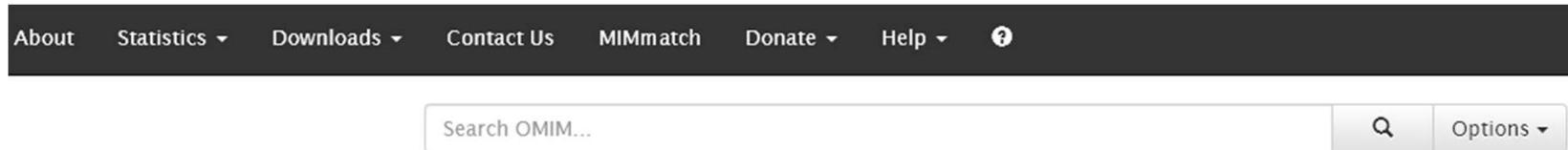
OMIM Entry Statistics

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

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

What is PMD ?

Last modified :*Mar 26, 2007*

Go to Japanese Page

- • **What is PMD ?** (see samples)

  Compliations 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 natur al or artificial, depending on the type of the mutation. For each entry the following ite ms 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 th e sequence. "CHANGE" indicates the position and kind of mutations, such as amino acid substitut ion, insertion and deletion, denoted with a specific notation. Any functional or structural features ("FUNCTION", "STRUCTURE", "STABILITY",etc) ob served 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**

| Gene / Description / OMIM	Database / URL	Curator / Institution
ATP-binding cassette, sub-family D (ALD), member 1 300371	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 ABO blood group (transferase A, alpha 1-3-N-acetylgalactosaminyltransferase; transferase B, alpha 1-3-galactosyltransferase) 110300	Blood Group Antigen Mutation Database http://www.ncbi.nlm.nih.gov/gv/mhc/xslcgi.cgi?cmd=bgmut/home	Olga O. Blumenfeld Department of Biochemistry, Santosh Patnaik, Department of Cell Biology, Albert Einstein College of Medicine New York, NY. U.S.A
ACAD8 acyl-CoA dehydrogenase family, member 8 604773	Innsbruck Metabolic Diseases Pages http://lovd.i-med.ac.at/home.php?select_db=ACAD8	Barbara Lanthaler, Stefanie Kalb and Martina Witsch-Baumgartner
ACADM acyl-CoA dehydrogenase, C-4 to C-12 straight chain 607008	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
ACADSB acyl-CoA dehydrogenase, short/branched chain 600301	Innsbruck Metabolic Diseases Pages http://lovd.i-med.ac.at/home.php?select_db=ACADSB	Barbara Lanthaler, Stefanie Kalb and Martina Witsch-Baumgartner
ACADVL acyl-CoA dehydrogenase, very long chain 609575	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
ACE2 angiotensin I converting enzyme (peptidyl-dipeptidase A) 2 300335	ACE2 database at LOVD http://www.LOVD.nl/ACE2	Johan T. den Dunnen Leiden Univ. Med Centre *(acting), Curator vacancy*
ACHE acetylcholinesterase (Yt blood group) 100740	Blood Group Antigen Mutation Database http://www.ncbi.nlm.nih.gov/gv/mhc/xslcgi.cgi?cmd=bgmut/home	Olga O. Blumenfeld Department of Biochemistry, Santosh Patnaik, Department of Cell Biology, Albert Einstein College of Medicine New York, NY. U.S.A
ACOT9 acyl-CoA thioesterase 9	ACOT9 database at LOVD http://www.LOVD.nl/ACOT9	Johan T. den Dunnen Leiden Univ. Med Centre *(acting), Curator vacancy*
ACSL4 acyl-CoA synthetase long-chain family member 4 300157	ACSL4 database at LOVD http://www.LOVD.nl/ACSL4	Johan T. den Dunnen Leiden Univ. Med Centre *(acting), Curator vacancy*
ACTA1 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
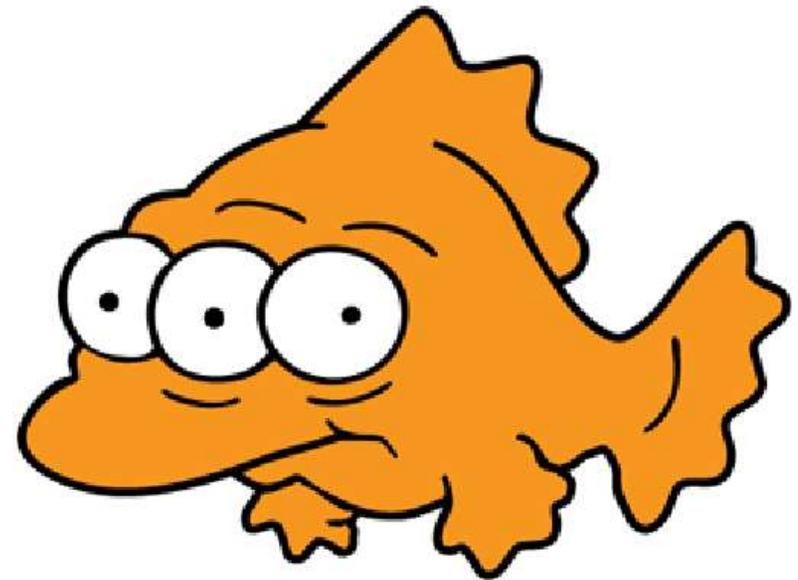
Mutations affecting structure

- **Molecular basis of mutations affecting folding & stability**

    - introduced clashes – common for small to large mutations in buried residues
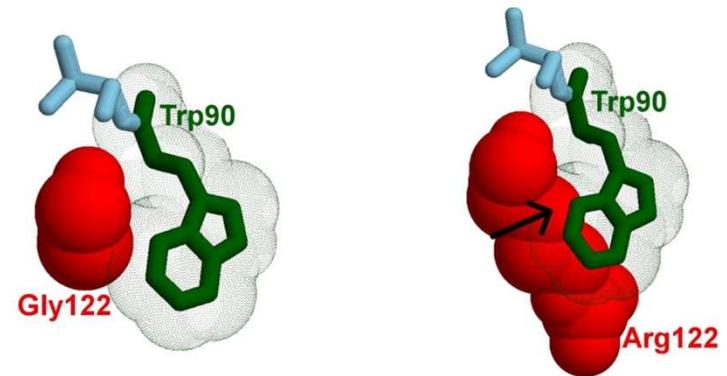
    - loss of interactions – most pronounced effects related to H-bonds, salt bridges and aromatic interactions
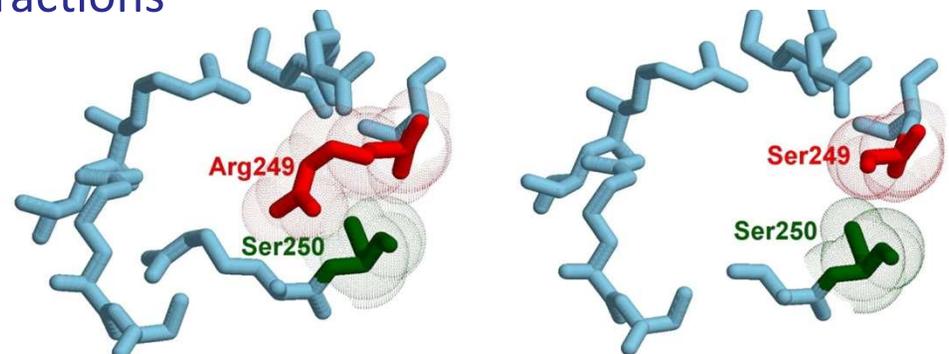
Mutations affecting structure

- **Molecular basis of mutations affecting folding & stability**

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

L75P   I77P   F78P   I88P   Y93P
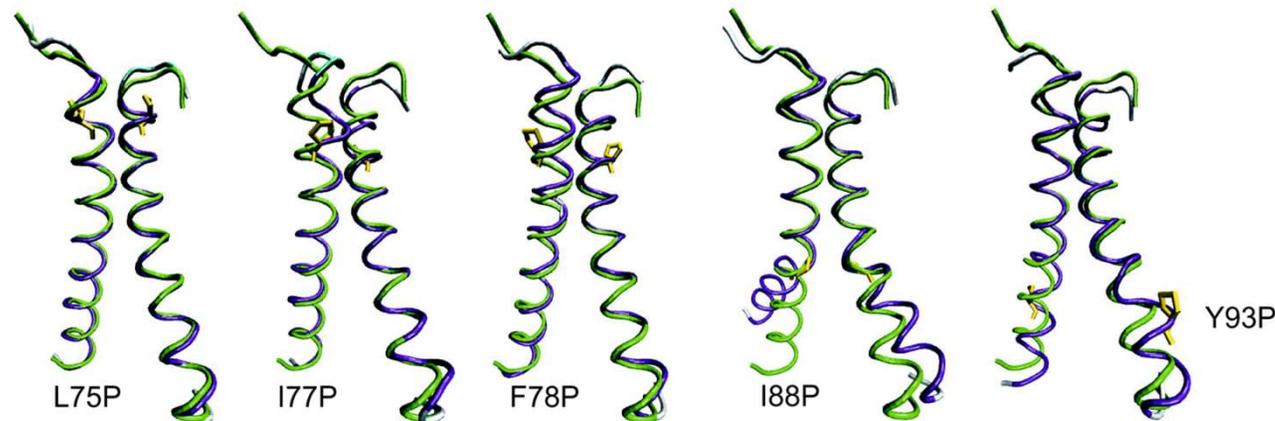
    - changes in charge/hydrophobicity

        - 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 affects 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

Ile84Val
3x lower K_i

Ile50Val
37 x lower K_i
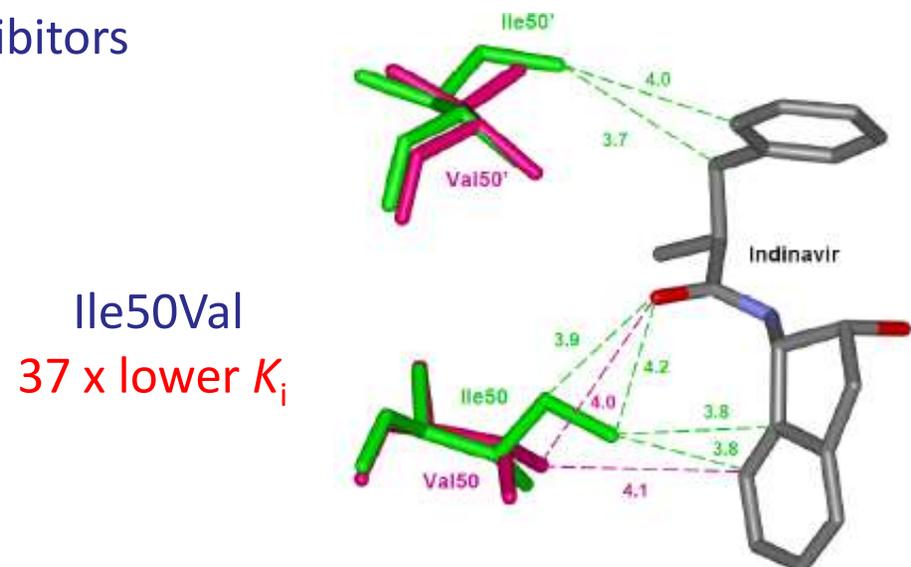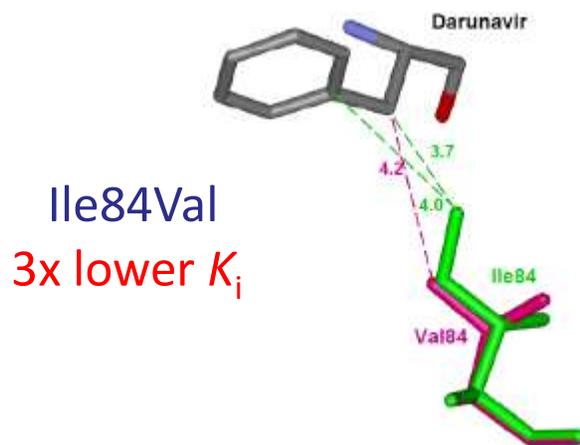
Mutations affecting function

- **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

LinB WT

LinB L177W
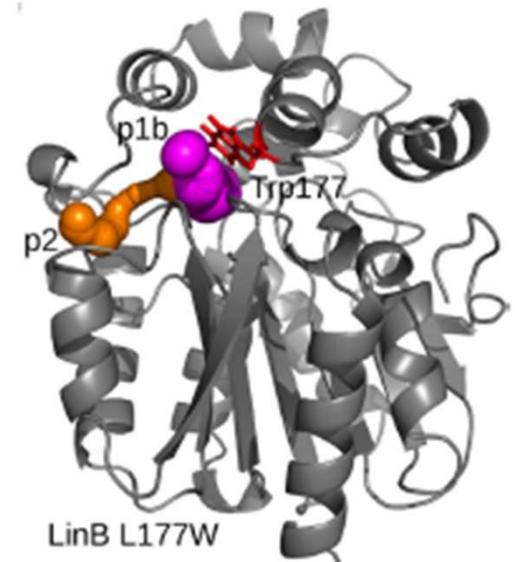
Mutations affecting function

❑ **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

Identification of mutable residues

Functional hot-spots
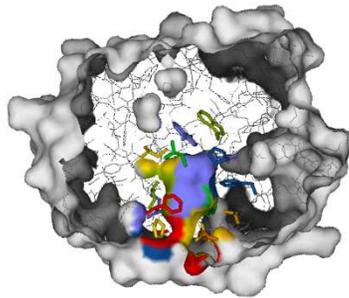
Stability hot-spots (flexibility)
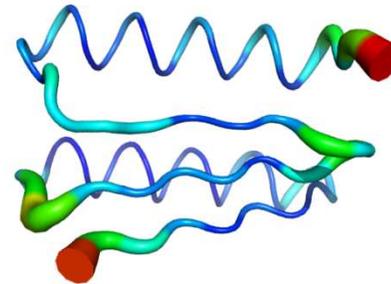
Stability hot-spots (evolution)

Correlated hot-spots

Identification of mutable residues

Identification of mutable residues

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 form 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; …
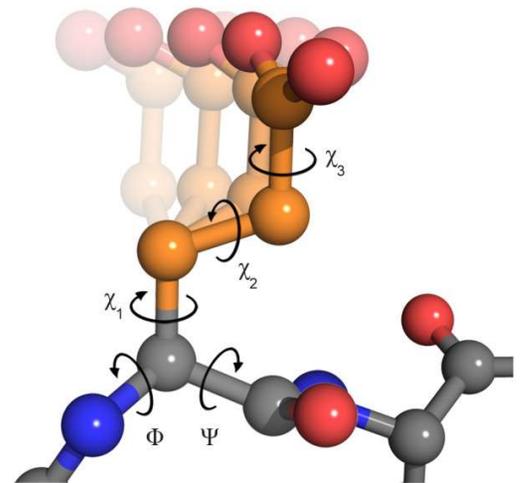
Prediction of effects on structure

- ❑ **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/
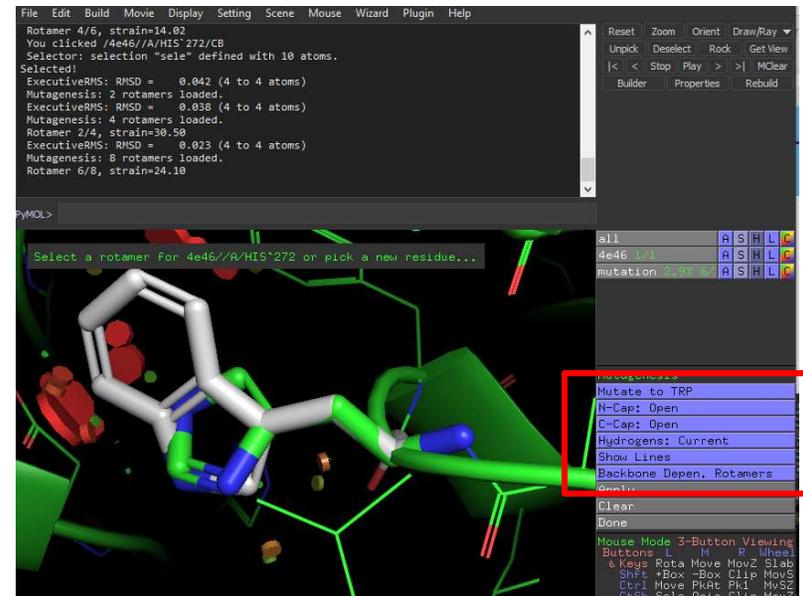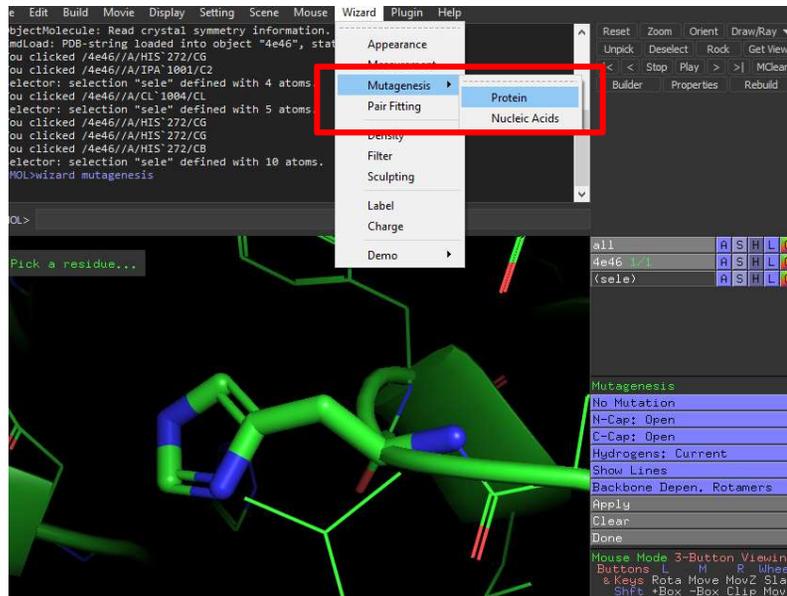
- stand alone, with plug-in to Yasara modeling tool

- 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**

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**   output

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

PREDICTSNP — Consensus classifiers for prediction of disease-related mutations

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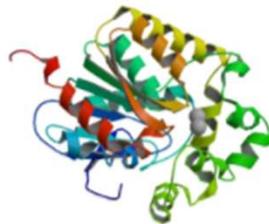
Consensus classifier for prediction of the effect of *nucleotide* substitutions.

Prediction of pathogenicity

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 binging (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
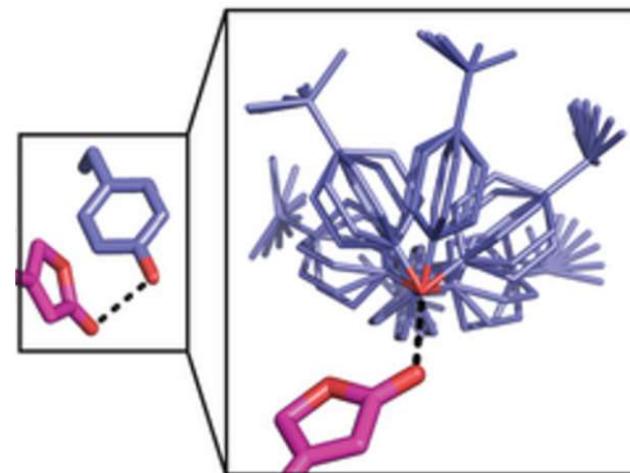
Improving ligand binding and activity

❑ **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
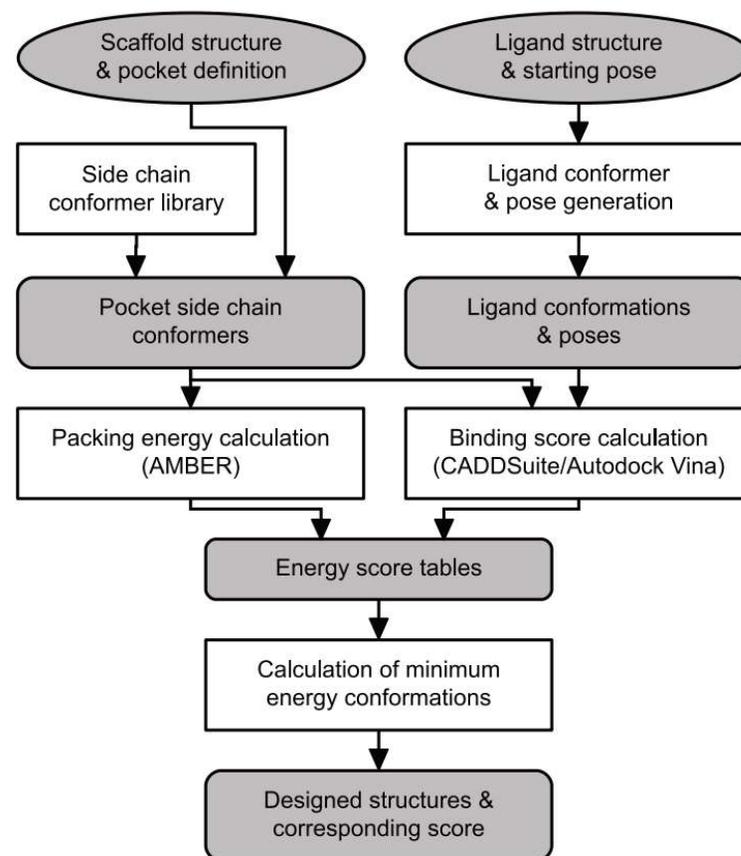
- utilizes evolution (conservation) and RosettaDesign (energy) to introduce multiple-point mutations to modify the properties of the binding site

- 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	3	
Maximal number of mutations per design	5	
Minimal PSSM threshold	-1 ⌄	
ΔΔG	5.5 ⌄	
Sequence space	143A	FY
	144A	P
	151A	FMY
	177A	LAGNST
	211A	ILMV
	247A	AGMSTVY
	248A	LIMV
Total number of designs in tolerated sequence space	3,313	

Reset Verify Proceed

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**

Improving protein stability

❑ **FireProt**

Mutations

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

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

Improving protein stability

❑ **FireProt^{ASR}**

- https://loschmidt.chemi.muni.cz/fireprotasr

- sequence-based stabilization: ancestral sequence reconstruction

- analysis of protein evolution and protein stabilization

Improving protein stability

❑ **PROSS**

- ▪ https://pross.weizmann.ac.il/step/pross-terms/

- ▪ Combination of mutations "allowed" by conservation analysis and

Rosetta calculations (energy)
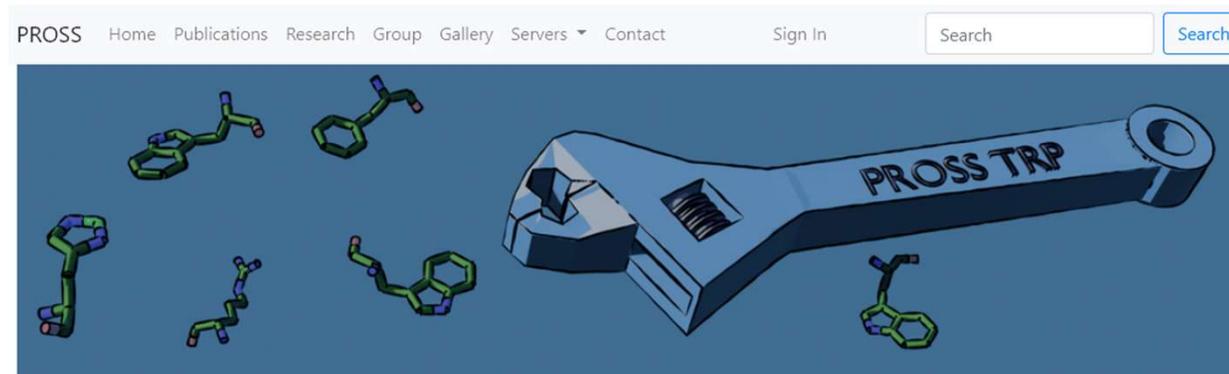
PROSS: the Protein Repair One-Stop Shop

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 (ΔΔG)

MVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEE-
IEGEGEGRPYEGTQTAKLKVTKGGPLP-
FAWDISPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWE
RVMNFEDGGVVTVTQDSSLQDGEFIYKVKL
RGTNFPSDGPVMQKKTMGWEASSERMYPEDG ALK-
GEIKQRLKLKDGGHYDAEVKTTYKAKVQLPGAYN-
VNIKLDITSHNEDYTIVEQYERAEGRHSTGGMDELYK

PVTVT

PVTVT   VVTVK
KVTVT   VVTPT
VRTVT   VVTVR
VETVT   VVTKT
VDTVT   VVEVT

THR   →   LYS
        ddG
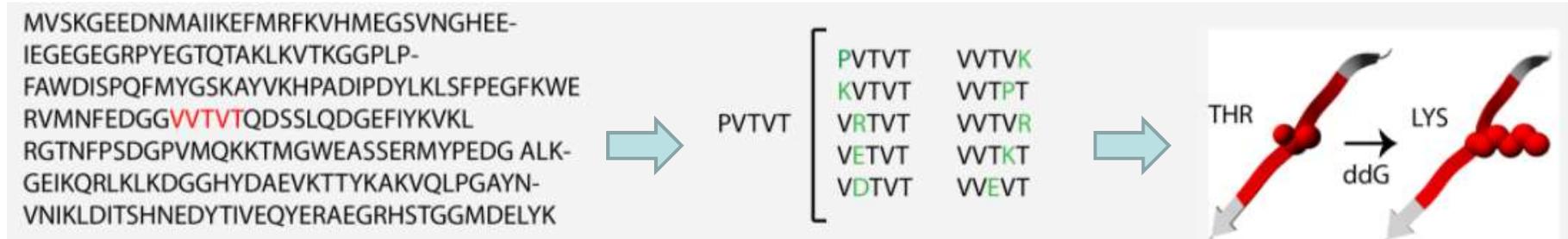
References I

- 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