

# Human Gene Mutation Database (HGMD) – databáze mutací lidských genů v kódujících oblastech.

(vychází s primárních dat, manuální a automatické prohledávání, více jak 250 časopisů je skenováno)



Obsahuje 120 000 různých mutací

4400 genů – záznamy všech dědičných mutací způsobující onemocnění

30 000 referencí

The screenshot shows the HGMD website interface. At the top, there's a navigation bar with links like 'Statistics', 'New genes', 'What's new', 'Background', 'Publications', 'Contact us', 'Register', and 'Log in'. Below that is a search bar with a 'Gene symbol' dropdown and a 'Go!' button. The main content area features the HGMD logo and a large heading 'Úvod do HGMD®'. The text below the heading is in Czech and describes the database's scope, including germline mutations and references to scientific literature. It also mentions the availability of HGMD Professional data.

- Quickly access detailed reports for human inherited disease mutations without tedious and time consuming literature searches

- Easily verify whether an observed mutation is novel or has been previously described

- Understand the mutational spectrum of a particular gene or disease

- Analyze / Identify candidate genes for disease linkage and predisposition

Human Gene Mutation Database (HGMD) – databáze mutací lidských genů  
 (vychází s primárních dat, manuální a automatické prohledávání, více jak  
 250 časopisů je skenováno)

Gene Symbol	Chromosomal location	Gene name	cDNA sequence	Extended cDNA	Mutation viewer
PAH <small>(Aliases: available to <a href="#">subscribers</a>)</small>	12q22-q24.2	Phenylalanine hydroxylase <small>(Aliases: available to <a href="#">subscribers</a>)</small>	<input type="text" value="NM_000277.1"/>	Not available	<b>BIOBASE</b> <small>Feature available to subscribers</small>

Mutation type	Number of mutations	Mutation data by type ( <a href="#">register</a> or <a href="#">log in</a> )
Missense/nonsense	401	<input type="button" value="Get mutations"/>
Splicing	79	<input type="button" value="Get mutations"/>
Regulatory	1	<input type="button" value="Get mutations"/>
Small deletions	65	<input type="button" value="Get mutations"/>
Small insertions	8	<input type="button" value="Get mutations"/>
Small indels	6	<input type="button" value="Get mutations"/>
Gross deletions	27	<input type="button" value="Get mutations"/>
Gross insertions/duplications	3	<input type="button" value="Get mutations"/>
Complex rearrangements	1	<input type="button" value="Get mutations"/>
Repeat variations	0	No mutations
Get all mutations by type	<b>BIOBASE</b> <small>Feature available to subscribers</small>	
<b>Public total</b> ( <a href="#">HGMD Professional 2012.2</a> total)	<b>591</b> (694)	

Více než 60% všech mutací tvoří missense mutace u monogenních onemocnění

- funkční mutace (10%)
- strukturní mutace (80%)
- Jiné efekty (10%)

Online Mendelian Inheritance in Man (**OMIM**) ... při NCBI  
Katalogizuje všechny známé dědičné choroby

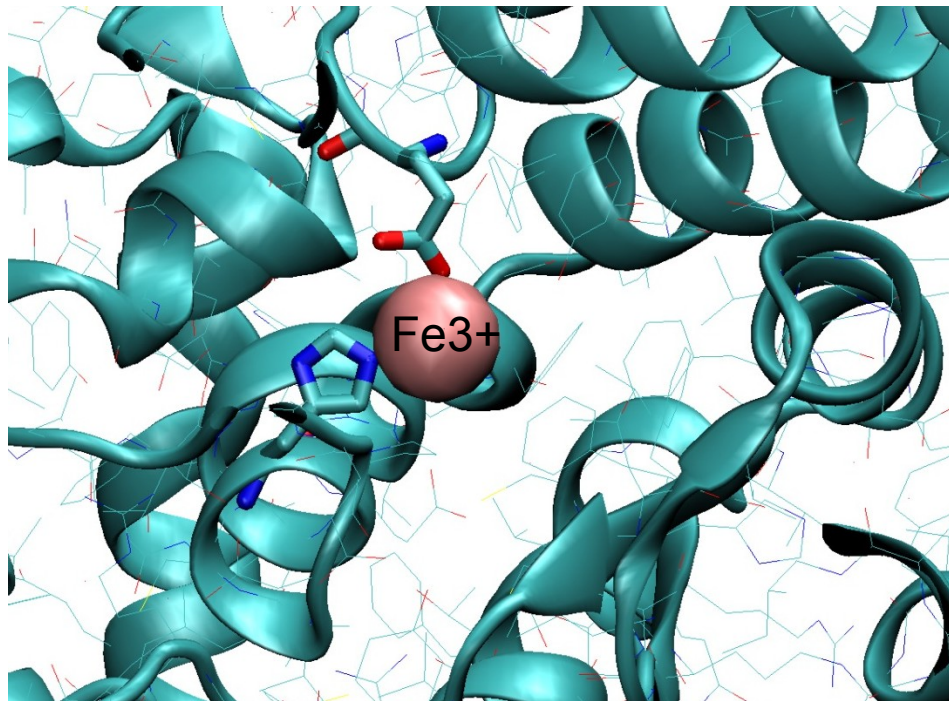
The Single Nucleotide Polymorphism database (**dbSNP**) ... při NCBI  
Katalogizuje variace v genech (výskyt více jak 1% v populaci).

**POZOR: Tato databáze ale uvádí i kauzální varianty!**

funkční mutace



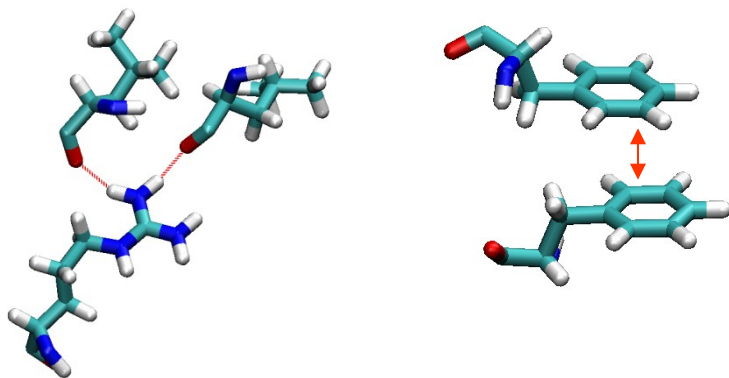
(např. vazbu ligandu/kofaktoru, příp. vazba specifických iontů součást konformace aktivního místa)



Obr. PAH

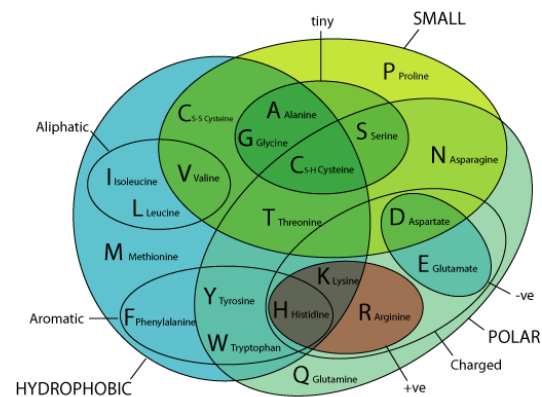
strukturní mutace ➔

Vytvářejí síť kontaktů s okolními aminokyselinami  
*H-bonding, stacking, salt bridges*



narušení struktury proteinu vede k destabilizaci proteinu,  
unfoldingu příp. agregaci

změna fyzikálně-chemických parametrů:  
*změna náboje,*  
*Polarity*  
*velikost aminokyseliny*



Ovšem ne vždy substituce aminokyselin je škodlivá!

Jak rozlišit mezi

single nucleotide polymorphisms (SNPs) v kódující oblasti, která je funkčně neutrální

A kauzální mutací?

„amino acids in the cores of proteins are relatively conserved compared with those in the solvent accessible regions

catalytic amino acids responsible for enzymatic reaction are also well conserved throughout evolution

Mutations tend to be accepted where evolutionary pressure is relatively relaxed and where they can remain in the population without selective disadvantage (or advantage)“

disease-related SNPs occur less in  $\alpha$ -helices but more frequently in  $\beta$ -strands than neutral nsSNPs

We find that cancer somatic mutations and disease-related variants occur more frequently at amino acids making hydrogen bonds from side chains than neutral polymorphisms.

Further, nsSNPs occur less frequently at the solvent inaccessible region of proteins, whereas disease-related mutations occur much more frequently than the average

***Sungsam Gong, and Tom L. Blundell Plos one 2010***

## Přístupy studia efektů missense mutací:

- Funkční analýzy (exprese proteinu, stabilita, specifické charakteristiky)  
časově a finančně náročné
- Automatické *in silico* programy (SIFT, PolyPhen, SNPs3D, FODLX..)  
rychlé a spolehlivost průměrně 70 %
- Strukturní analýza pomocí MD a bioinformatiky  
časově a finančně středně náročné, pohled do mechanismu účinky  
je to ale stále predikce



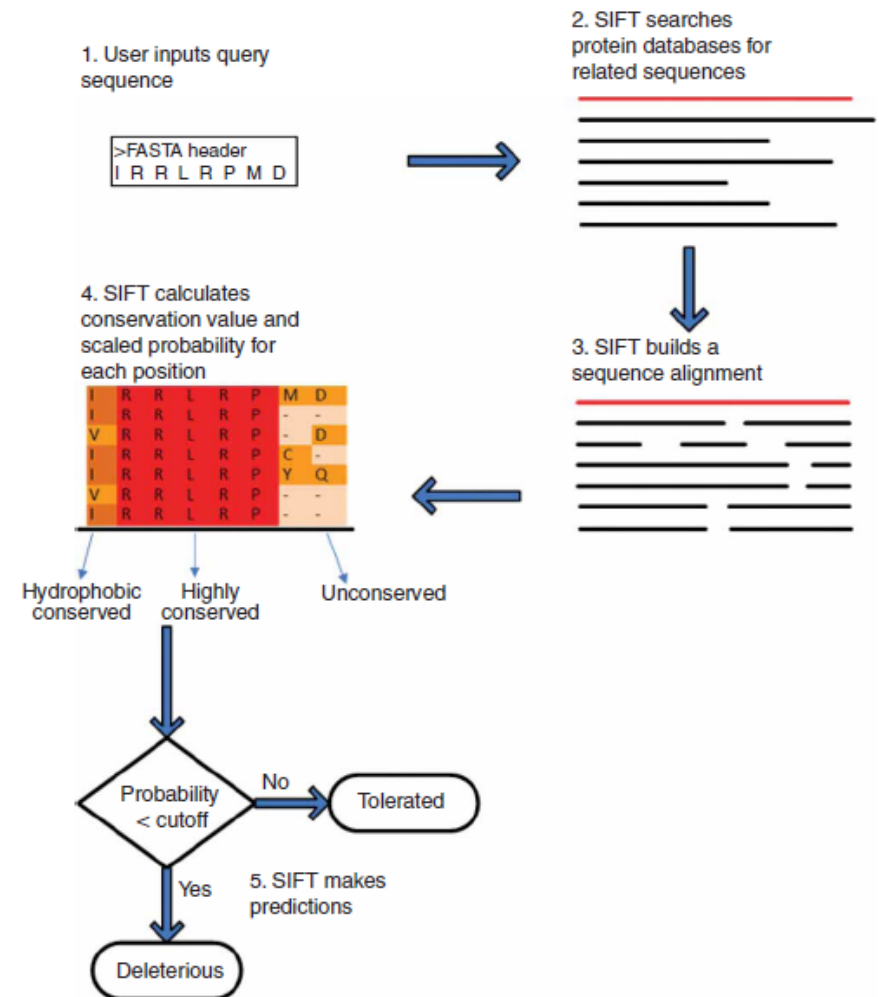
# Automatické analýzy missense mutací pomocí bioinformatických nástrojů

**SIFT** - Sorting Tolerant From Intolerant  
(<http://sift.jcvi.org>)

SIFT assumes that important positions in a protein sequence have been conserved throughout evolution and therefore substitutions at these positions may affect protein function.

Using sequence homology, SIFT predicts the effects of all possible substitutions at each position in the protein sequence, 5–20 min.

Úspěšnost predikce pro skupinu kauzálních mutací 69%.  
Falešně pozitivní 19 %.



Amino acid substitution models such as PAM and BLOSUM describe the degree of substitutions as log-odd ratio values where the positive scores suggest commonly occurring and preferred substitutions, whereas the negative scores imply very rare substitutions which are disfavoured in nature. Those substitution tables were widely used to assess and predict the effects of nsSNPs

## Substituční matice pro aminokyseliny

- AA s podobnými chemickými vlastnostmi (např: náboj, velikost) mají vyšší skóre
  - Acids & Amides DENQ (Asp, Glu, Asn, Gln)
  - Basic HKR (His, Lys, Arg)
  - Aromatic FYW (Phe, Tyr, Trp)
  - Hydrophilic ACGPST (Ala, Cys, Gly, Pro, Ser, Thr)
  - Hydrophobic LMV (Ile, Leu, Met, Val)

	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W	
C	9																				C
S	-1	4																			S
T	-1	1	5																		T
P	-3	-1	-1	7																	P
A	0	1	0	-1	4																A
G	-3	0	-2	-2	0	6															G
N	-3	1	0	-2	-2	0	6														N
D	-3	0	-1	-1	-2	-1	1	6													D
E	-4	0	-1	-1	-1	-2	0	2	5												E
Q	-3	0	-1	-1	-1	-2	0	0	2	5											Q
H	-3	-1	-2	-2	-2	-2	1	-1	0	0	8										H
R	-3	-1	-1	-2	-1	-2	0	-2	0	1	0	5									R
K	-3	0	-1	-1	-1	-2	0	-1	1	1	-1	2	5								K
M	-1	-1	-1	-2	-1	-3	-2	-3	-2	0	-2	-1	-1	5							M
I	-1	-2	-1	-3	-1	-4	-3	-3	-3	-3	-3	-3	-3	1	4						I
L	-1	-2	-1	-3	-1	-4	-3	-4	-3	-2	-3	-2	-2	2	2	4					L
V	-1	-2	0	-2	0	-3	-3	-3	-2	-2	-3	-3	-2	1	3	1	4				V
F	-2	-2	-2	-4	-2	-3	-3	-3	-3	-3	-1	-3	-3	0	0	0	-1	6			F
Y	-2	-2	-2	-3	-2	-3	-2	-3	-2	-1	2	-2	-2	-1	-1	-1	-1	3	7		Y
W	-2	-3	-2	-4	-3	-2	-4	-4	-3	-2	-2	-3	-3	-1	-3	-2	-3	1	2	11	W

# Automatické analýzy missense mutací pomocí bioinformatických nástrojů

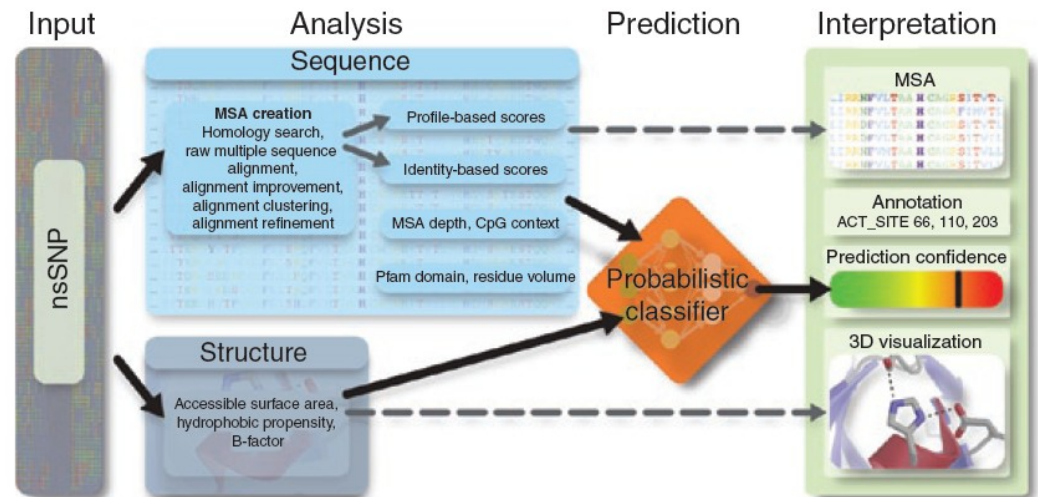
**Polyphen -2** (<http://genetics.bwh.harvard.edu/pph/>)

Uses sequence conservation, structure and SWISS-PROT annotation



Úspěšnost predikce 92% a 73%  
Pro HumDiv a HumVar datasets

Falešně pozitivní ~ 20 %.



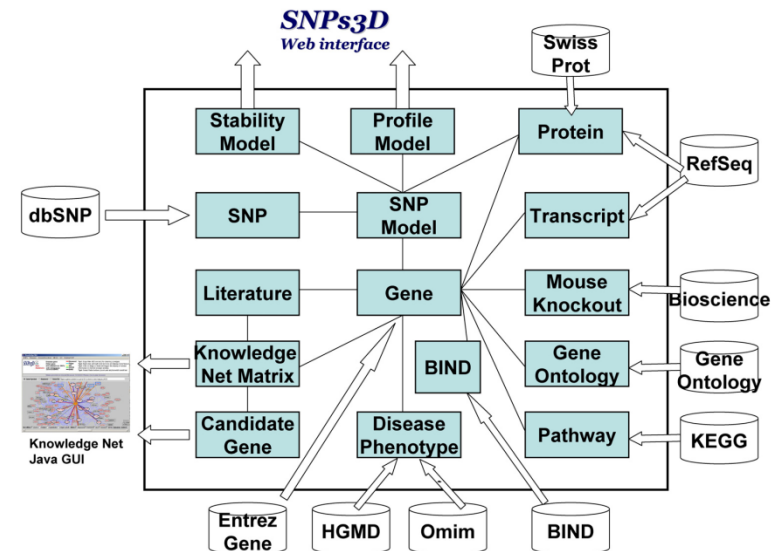
# Automatické analýzy missense mutací pomocí bioinformatických nástrojů

**SNPs3D** (<http://www.snps3d.org/>)

Uses the SupportVector Machine method (analyze data and recognize patterns) together with 15 parameters (structure based) or 5 parameters (alignment based).

*E.g. Loss of polar–polar interactions, Loss of hydrophobic interactions, Loss of a salt-bridge, Buried charge, Over-packing, Cavity formation...*

These parameters (when positive) indicate decrease of protein stability.



Úspěšnost predikce 74%  
Falešně pozitivní 15 %.

## Automatické analýzy missense mutací pomocí bioinformatických nástrojů

**FOLDX** (<http://foldx.embl.de/>)

Změna celkové volné energie

$$\Delta\Delta G = \Delta G_{mut} - \Delta G_{wt}$$

$\Delta G_{mut}$  - free energy difference between the folded and unfolded states

Empirical force field that was developed for the rapid evaluation of the effect of mutations on the stability, folding and dynamics of proteins and nucleic acids

$$\begin{aligned}\Delta G = & a \cdot \Delta G_{vdw} + b \cdot \Delta G_{solvH} + c \cdot \Delta G_{solvP} + d \cdot \Delta G_{wb} \\ & + e \cdot \Delta G_{hbond} + f \cdot \Delta G_{el} + g \cdot \Delta G_{kon} + h \cdot T\Delta S_{mc} \\ & + k \cdot T\Delta S_{sc} + l \cdot \Delta G_{clash}.\end{aligned}$$

(a . . . l) are relative weights of the different energy terms used for the free energy calculation

Výpočty na lokálním PC ne přes web rozhraní nyní.

Lze měnit počáteční nastavení: T(K), c(M)...

## Automatické analýzy missense mutací pomocí bioinformatických nástrojů

**FOLDX** (<http://foldx.embl.de/>)

<i>FoldX component</i>	<i>Energy (kcal.mol-1)</i>
Backbone Hbond	-48.30
Sidechain Hbond	-18.40
Van der Waals	-62.33
Electrostatics	-2.79
Solvation Polar	84.34
Solvation Hydrophobic	-76.71
Van de Waals clashes	0.00
entropy side chain	34.93
entropy main chain	81.94
sloop_entropy	0.00
mloop_entropy	0.00
cis_bond	0.00
torsional clash	0.40
backbone clash	1.20
helix dipole	0.00
water bridge	0.00
disulfide 0.00	
electrostatic kon	0.00
partial covalent bonds	0.00
Total	-5.72

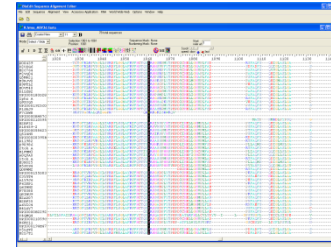
$$\Delta\Delta G(\text{change}) = \Delta G(\text{MT}) - \Delta G(\text{WT})$$

$\Delta\Delta G(\text{change}) > 1 \text{ kcal/mol}$ : the mutation is destabilizing

$\Delta\Delta G(\text{change}) < -1 \text{ kcal/mol}$ : the mutation is stabilizing

Úspěšnost predikce 60%

# Vlastní analýza missense mutací pomocí MD a bioinformatiky



## Hodnocené znaky:

Specifické kontakty postraními řetězci AA

Výskyt AA v aktivním místě

Zanořenost AA v proteinu

- změna objemu AA

- změna náboje

- změna polarity

Konzervovanost AA > 50%

Přítomnost helix/turn breakers

Mutační pravděpodobnost (Blosum62)

**Ukazatele kauzality**

Blundell:

Clanek v plos one

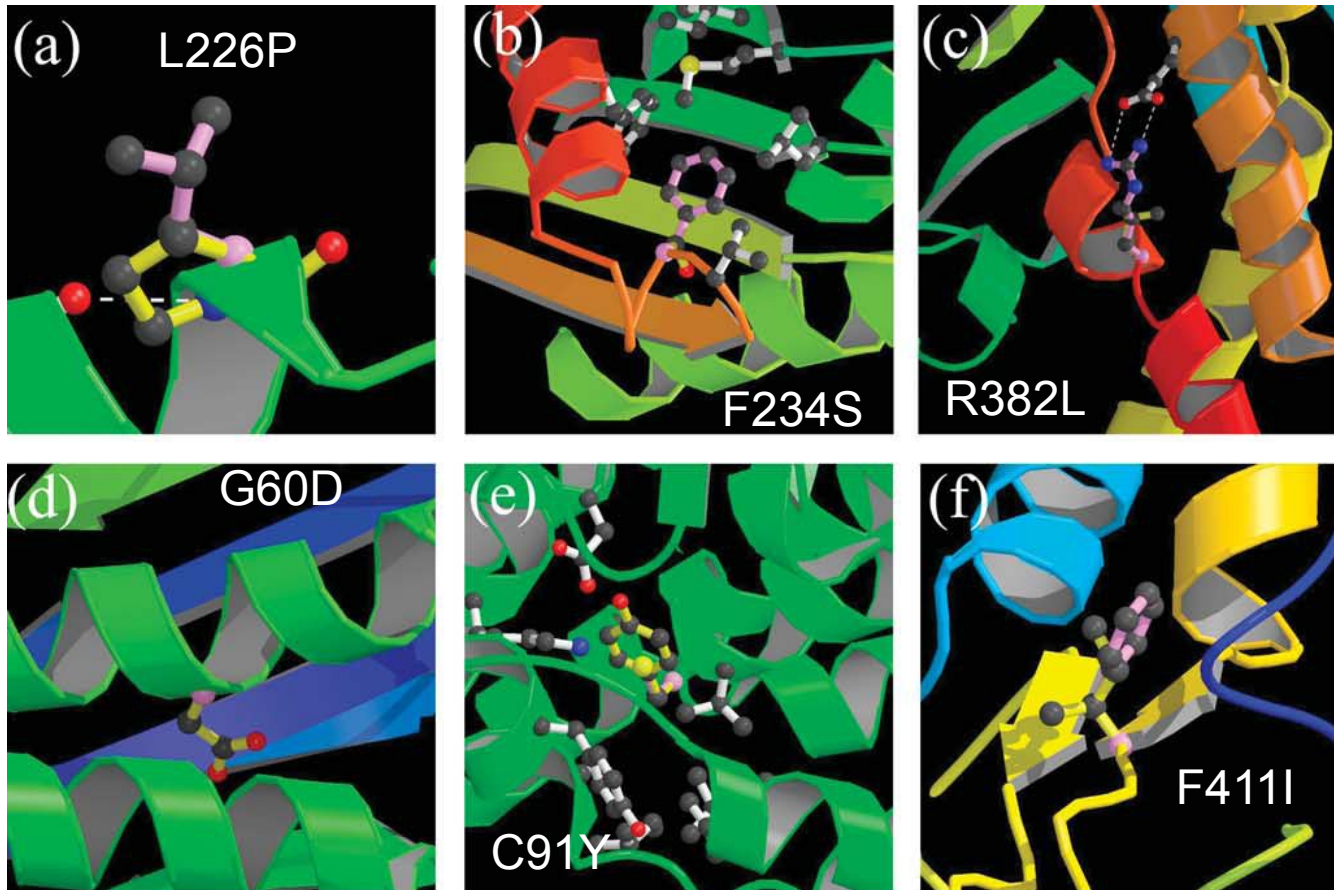
Kauzální mutace uvnitř proteinu, polymorfismy spíš napovrchu

Kauzální mutace: wt mutace častěji vytváří H-bond kontakty



Examples of disease caused by structure destabilizing factors:  
In silico analysis of mutations

Causal mutations:



bonds of wild-type side-chains are shown purple,  
and bonds of the mutant side-chains are yellow.