Structure-based modeling of cysteine and serine disease variants of human proteome

Student: Dmitrii Podgalo
Supervisor: Petr Popov, Skoltech

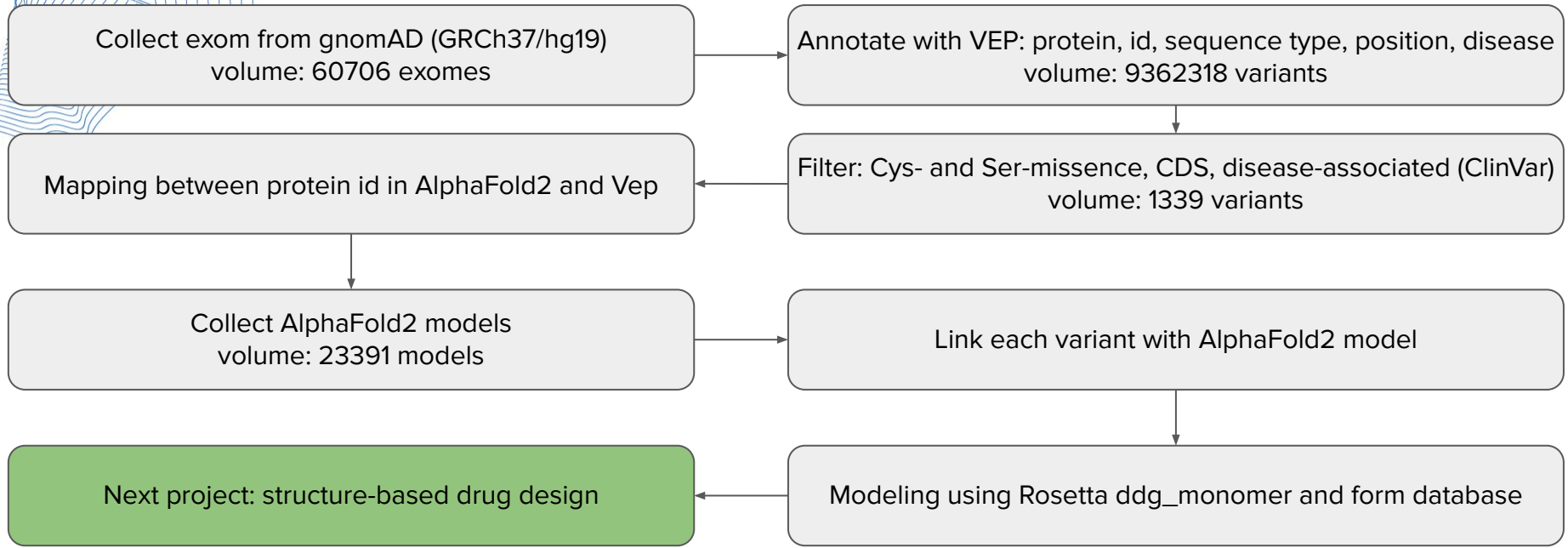
Introduction

There are many disease-associated mutations that endow pharmacological target (typically a protein), drug resistance, e.g. G12C amino acid substitution in oncogenic target KRAS. People carrying such mutations may need the development of personalized drugs, that take into account structural peculiarities of the mutated protein. One of the promising strategies is to develop covalent drugs that are specific for a given mutation.

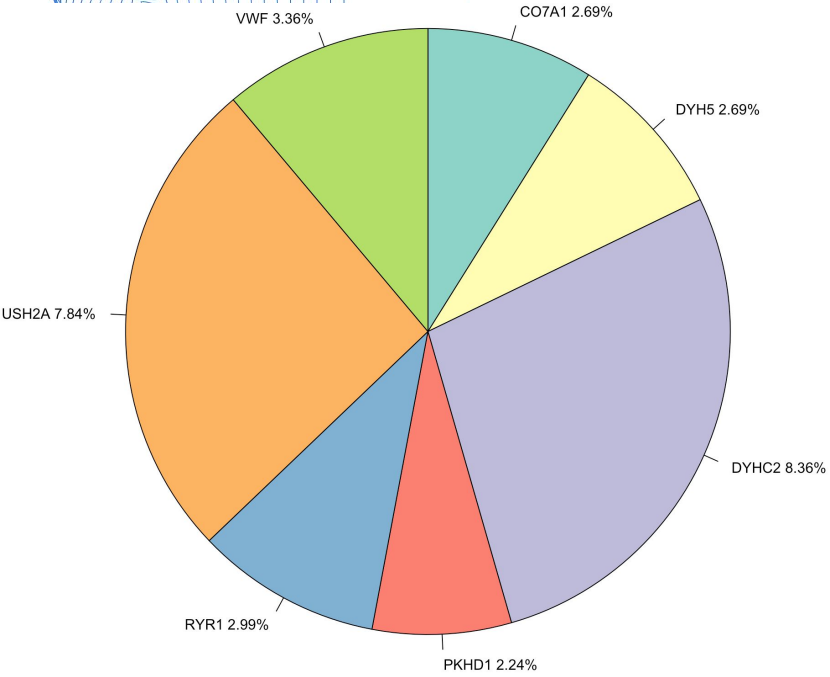
Aim

The goal of this project is to model structures of human proteins with disease-associated amino acid substitutions. Two types of amino acid substitutions are selected: X to Cysteine or X to Serine (X is any amino acid residue) – these residues are often used as the attachment points for covalent drugs.

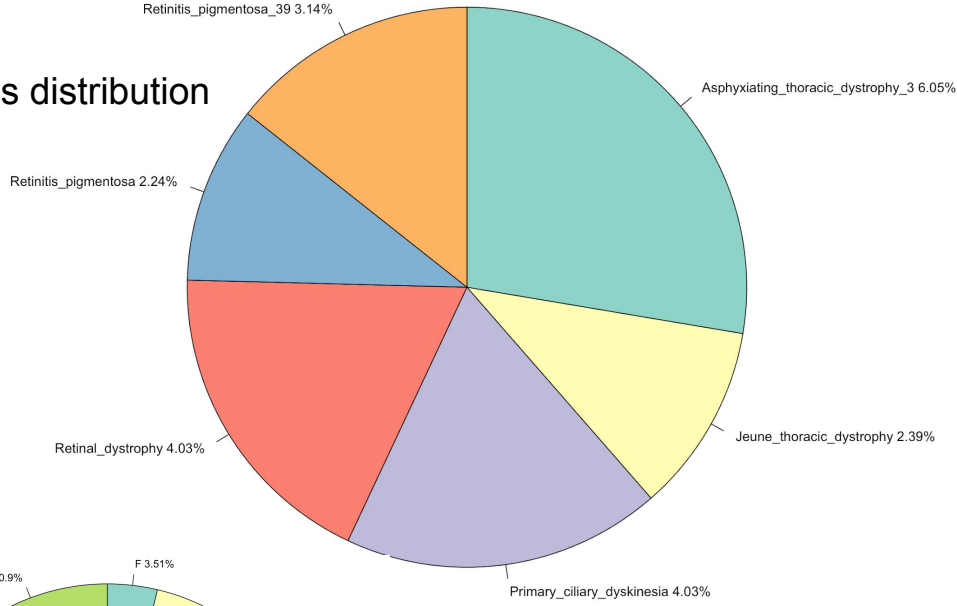
Workflow


Collect exom from gnomAD (GRCh37/hg19)
volume: 60706 exomes
Annotate with VEP: protein, id, sequence type, position, disease
volume: 9362318 variants
Filter: Cys- and Ser-missence, CDS, disease-associated (ClinVar)
volume: 1339 variants
Mapping between protein id in AlphaFold2 and Vep
Collect AlphaFold2 models
volume: 23391 models
Link each variant with AlphaFold2 model
Modeling using Rosetta ddg_monomer and form database
Next project: structure-based drug design

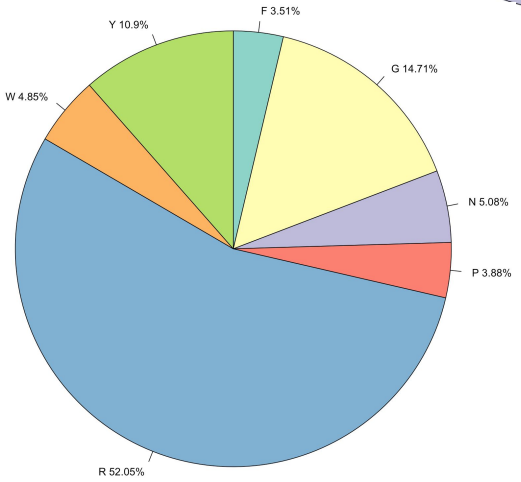
Statistics



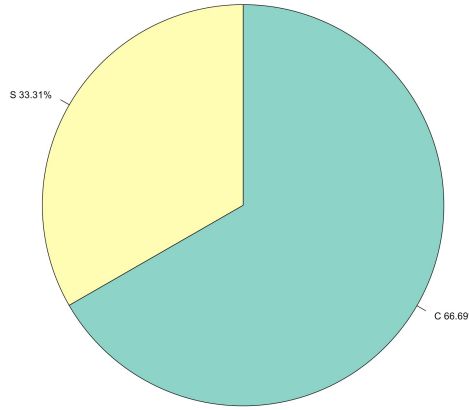
Proteins distribution



Diseases distribution



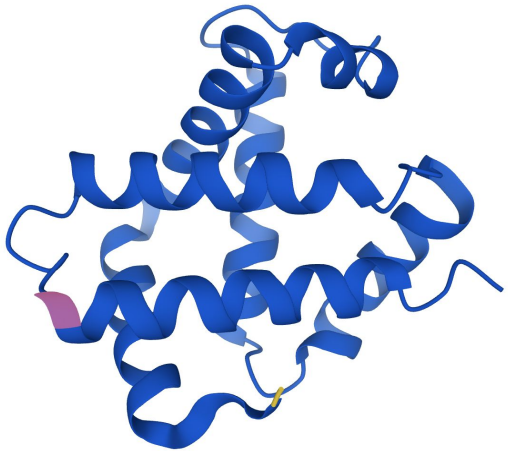
Normal amino acid distribution



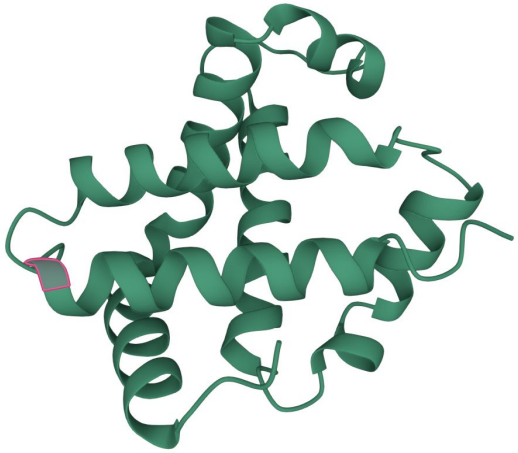
Mutant amino acid distribution

Example of Ser-variant

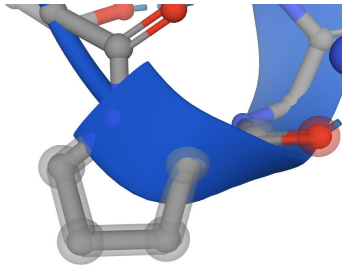
At 120 point of Hemoglobin α-chain proline replaced by serine and
this led to the disease

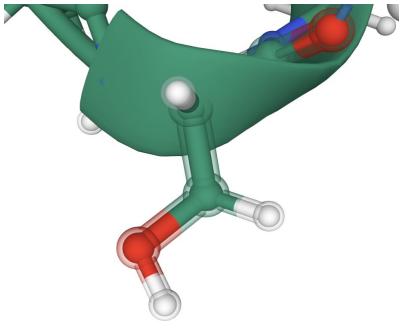


Hemoglobin α-chain normal



Hemoglobin α-chain mutated





Results

Collect 1339 variants associated with a disease. Got 1339 mutant 3D-models for each variants.

```
HFM1_HUMAN,A2PYH4,MLKSNDCLFSLENLFFEKPDEVENHPDNEKSLDWFLPPAPLISEIPDTQELEEELESHKLLGQEKRPKMLTSNLKITNEDTNYISLTQKFQFAFPSDKYEQDDLNLEGVGNNDLSHIAGKLTYASQKYKNHIGTEIAPEKSVPDDTKLVNFAEDKGESTSVFRKRLFKISDNIHGSAYSNDNELDS
HIGSVKIVQTEMNKGKSRNYSNSKQKFQYSANVFTANNAFSASEIGEGMFKAPSFSVAFQPHDIQEVTENGLGSLKAVTEIPAKFRSIFKEFPYFNYIQSKAFDDLLYTDRNFVICAPTGSGKTVVFELAITRLLMEVPLPWLNIKIVYMAPIKALCSQRFDDWKEKFGPIGLNCKELTGDTVMDDLFEIQHAHIIMTTPEKWD
SMTRKWRDNSLVQLVRLFLIDEVHIVKDENRGPTLEVVVSRMKTVQSVSQTLKNTSTAIPMRFVAVSATIPNAEDIAEWLSDGERPAVCLKMDESHRPVKLQKVVLGFPCSSNQTEFKFDLTLNYKIASVIQMYSDQKPTLVFCATRKGVQQAASVLVKDAKFIMTVEQKQRLQKYAYSVRDSKLRDILKDGAAYHHAGMELSD
RKVVEGAFTVGDLPVLFTTSTLAMGVNLPAHLVVIKSTMHYAGGLFEEYSETDILQMIGRAGRPQFDTTATAVIMTRLSTRDKYIQMLACRDTVESSLHRHLIEHLNAEIVLHTITDVNIAVEWIRSTLLYIRALKNPSHYGFASGLNKDGIEAKLQELCLKNLNDLSSLDLIKMDEGVNFKPTEAGRLMAWYYITFETVKKFY
TISGKETLSDLVTLIAGCKEFLDIQLRINEKKTLNTLNKDPNRITIRFPMEGRIKTREMKVNCLIQAQLGCIPIQDFALTQDTAKIFRHGSRITRWLSDFVAAQEKKFAVLLNSLILAKCFRCKLWENSLHVSKQLEKIGITLSNAIVNAGLTSFKKIEETDARELELILNRHPPFGTQIKETVMYLPKYELKVEQITRYSDTT
AEILVTVILRNFEQLQTKRTASDSHYVTLIIGDADNQVVYLHKITDSVLLKAGSWAKKIAVKRALKSEDLSINLISSEFVGLDIQQKLTVFYLEPKRFGNQITMQRKSETQISHSKHSDISTIAGPNKGTTASKKPGNRECNHLCKSKHTCGHDCCKIGVAQKSEIKESTISSYLSDLRNRNAVSSVPPVKRLKIQMNKSQSVD
LKEFGFTPKPSLPSISRSEYLNISELPIMEQWDQPEIYGKVRQEPSEYQDKEVLNVNFELGNEVWDDFDDENLEVTSFSTDTEKTKISGFGNTLSSSTRGSKLPLQESKSKFQREMSNSFVSSHEMSDISLSNSAMPKFSASSMTKLPQQAGNAVIVHFQERKPQNLSPEIEKQCFTFSEKNPNSSNYKKVDFFIRNSECKKEV
DFSMYHPDDEADEMKSLLGIFDGIF,884,I,S,Premature_ovarian_failure_9,AF-A2PYH4-F1-model_v2.pdb.gz
TM218_HUMAN,A2RU14,MAGTVLGVGAGVFILALLWVAVLLLCVLLSRASGAARFSVIFLFFGAVIITSVLLLFPRAGEFPAPEVEVKIVDDFFIGRYVLLAFLSAIFLGGLFLVLIHYVLEPIYAKPLHSY,80,R,C,Joubert_syndrome,AF-A2RU14-F1-model_v2.pdb.gz
S38A8_HUMAN,A6NNN8,MEGQTPGSRGLPEKPHPATAAATLSSMGAVFILMKSALGAGLLNFPWAFSKAGGVVPAFLVELVSLVFLISGLVILGYAAAVSGQATYQGVVRGLCGPAIGKLCEACFLLNLLMISVAFLRVIGDQLEKLCDSLLSGTPPAPQPWYADQRFTLPLLSVLVILPLSAPREIAFQKYTSILGTLAAC
YLALVITVQYYLWPQGLVRESHPSLSPASWTSVFSVFPTICFGFQCHEAAVSIYCSMRKRSLSHWALVSVLSLLACCLIYSLTGVYGFLTFGTEVSADVLMSYPGNDMVIIVARVLFAVSIVTVYPIVLFLGRSVMQDFWRRSCLGGWGPSALADPSGLWVRMPLTILWVTVTLAMALFMPDLSEIVSIIGGISSFFIFIFPGL
CLICAMGVEPIGPRVKCCLEVWGVVSVLVGTFIFGQSTAAAVWEMF,32,I,S,Foveal_hypoplasia,AF-A6NNN8-F1-model_v2.pdb.gz
```

In the future...

The obtained structural models will be used as the starting conformations for the structure-based drug design pipelines.

DmitriiPodgalo/POP