Structure-based modeling of Cys– and Ser–disease variants of human proteome

BIOINFORMATICS INSTITUTE

Student:
Dmitrii Podgalo

Supervisor:
Petr Popov, Skoltech

Aim: to model structures of human proteins with disease-associated amino acid substitution

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

↓

Collect AlphaFold2 models
volume: 23391 models

↓

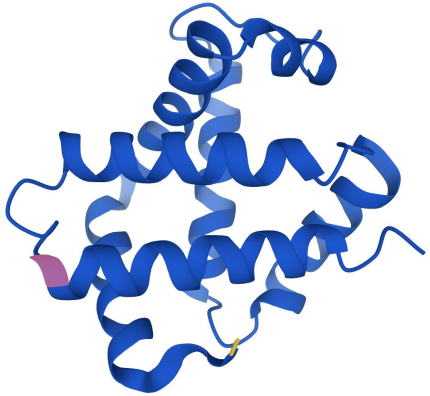
Link each variant with AlphaFold2 model

↓

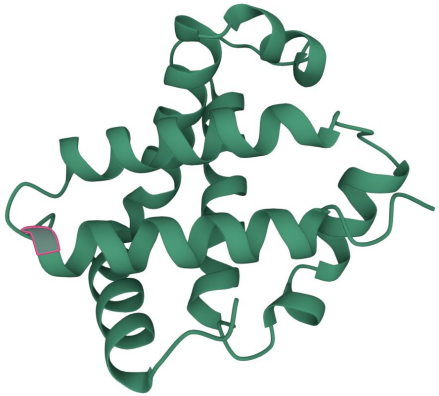
Modeling using Rosetta ddg_monomer and form database

↓

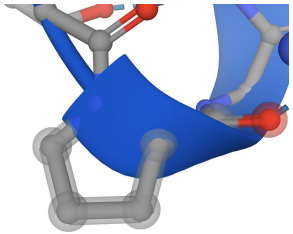
Next project: structure-based drug design

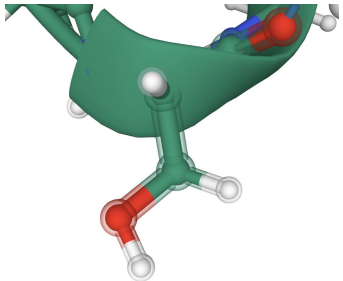


Hemoglobin α-chain
normal



Hemoglobin α-chain
mutated





Proline ⇨ Serine at 120 point ⇨ thalassemia