

Prediction of the effect of single aminoacid protein variants using deep mutational scanning data

University of Bologna — Master Thesis in Bioinformatics

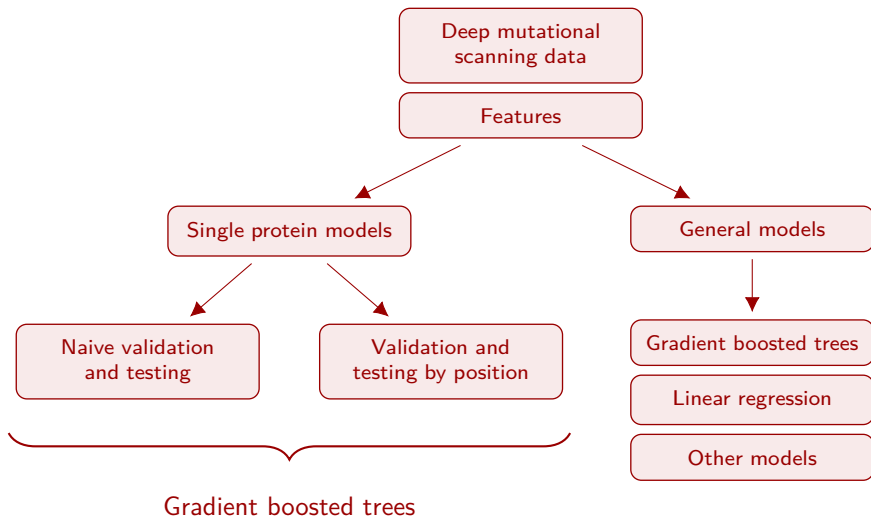
Pierotti Saul

Internal Advisor: Prof. Pietro Di Lena

External Advisor: Prof. Arne Elofsson (Stockholm University)

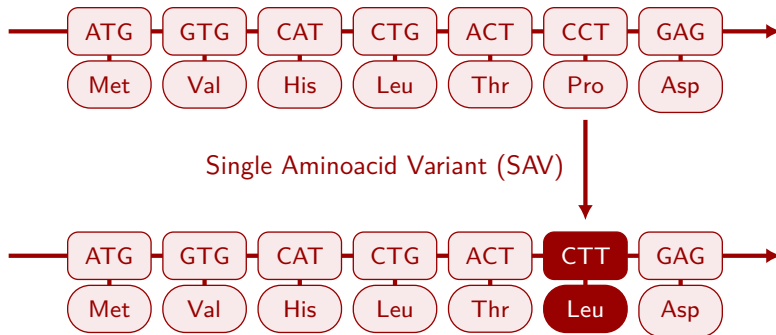
July 19, 2021

Structure of the project

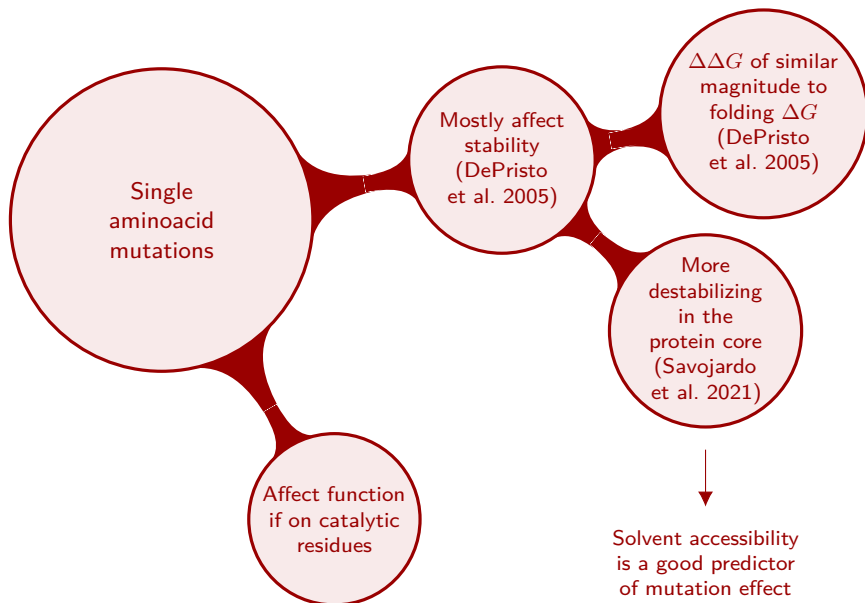


Single aminoacid variants

In this work I focused exclusively on point missense mutations. Nonsense mutations, indels, and mutations in non-coding regions were not considered.

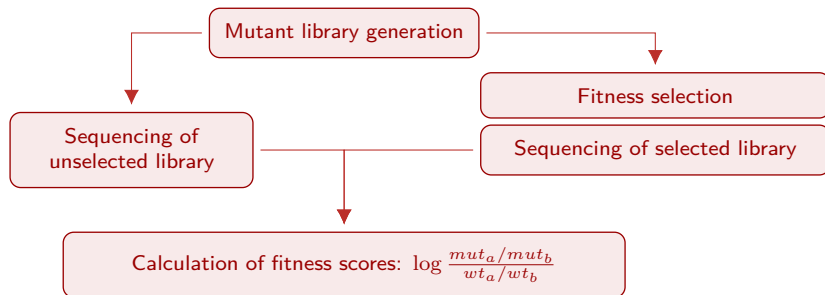


Effect of single aminoacid mutations in proteins

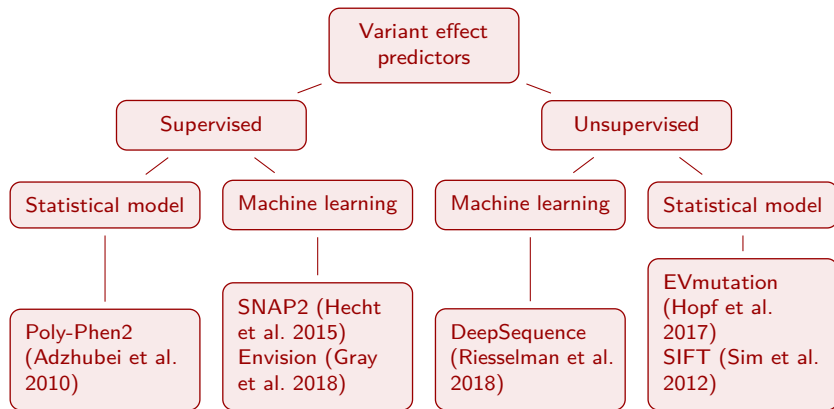


Deep mutational scanning

High-throughput technique for obtaining fitness information on a large number of mutations.



Variant effect predictors



Envision, EVmutation, and DeepSequence provided quantitative predictions. Envision was trained on deep mutational scanning data while the others are either unsupervised or trained on SNP annotations.

Training data

I used the training dataset of Envision (Gray et al. 2018), composed of nine independent experiments on eight different proteins. The distribution of fitness scores is bimodal and very variable across datasets.

