

What we wanted to do

Insights

Improvement vs. Deterioration

Hotspots & Patterns

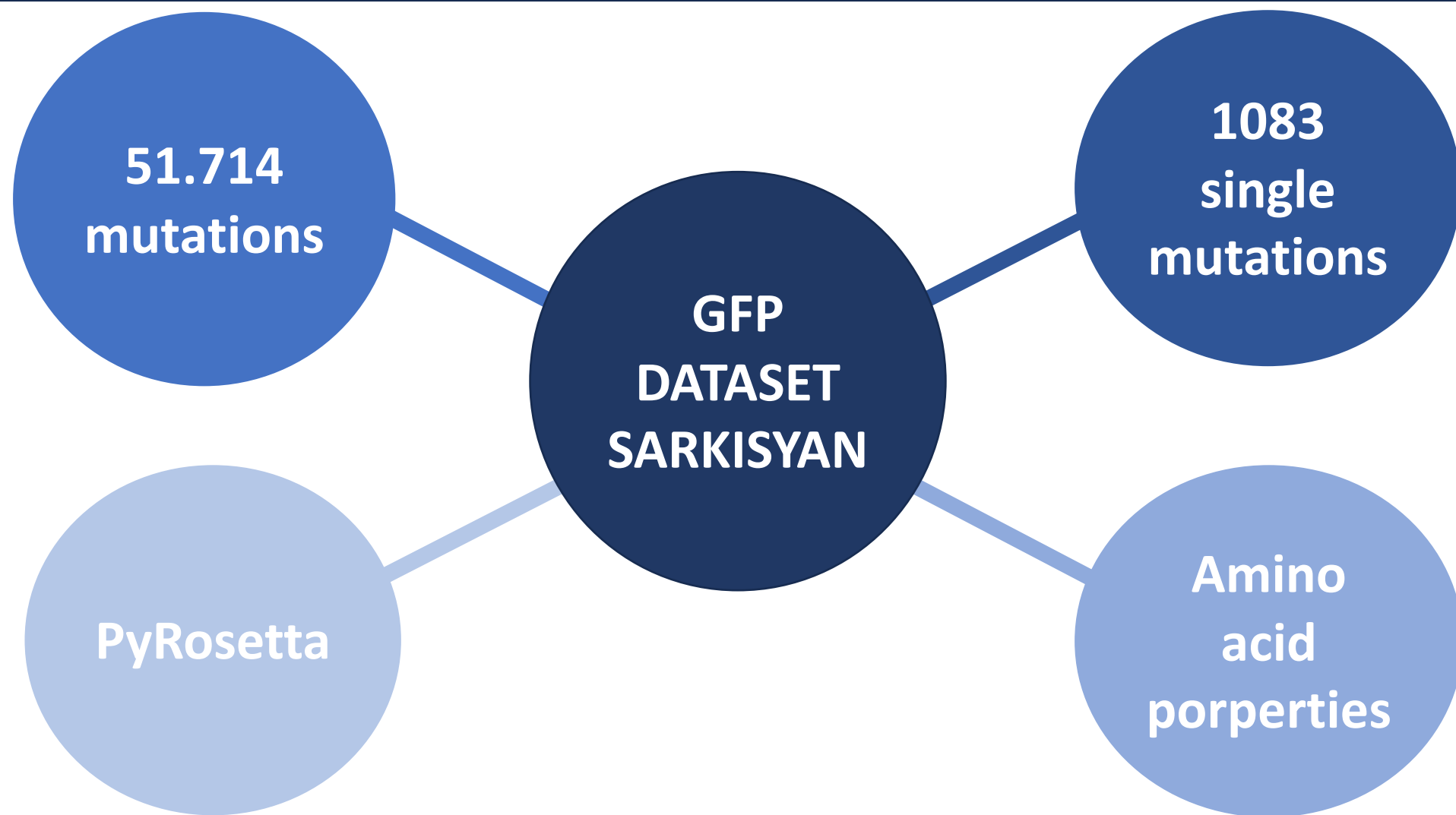
Further Analysis

How to improve GFP's tolerance towards mutations

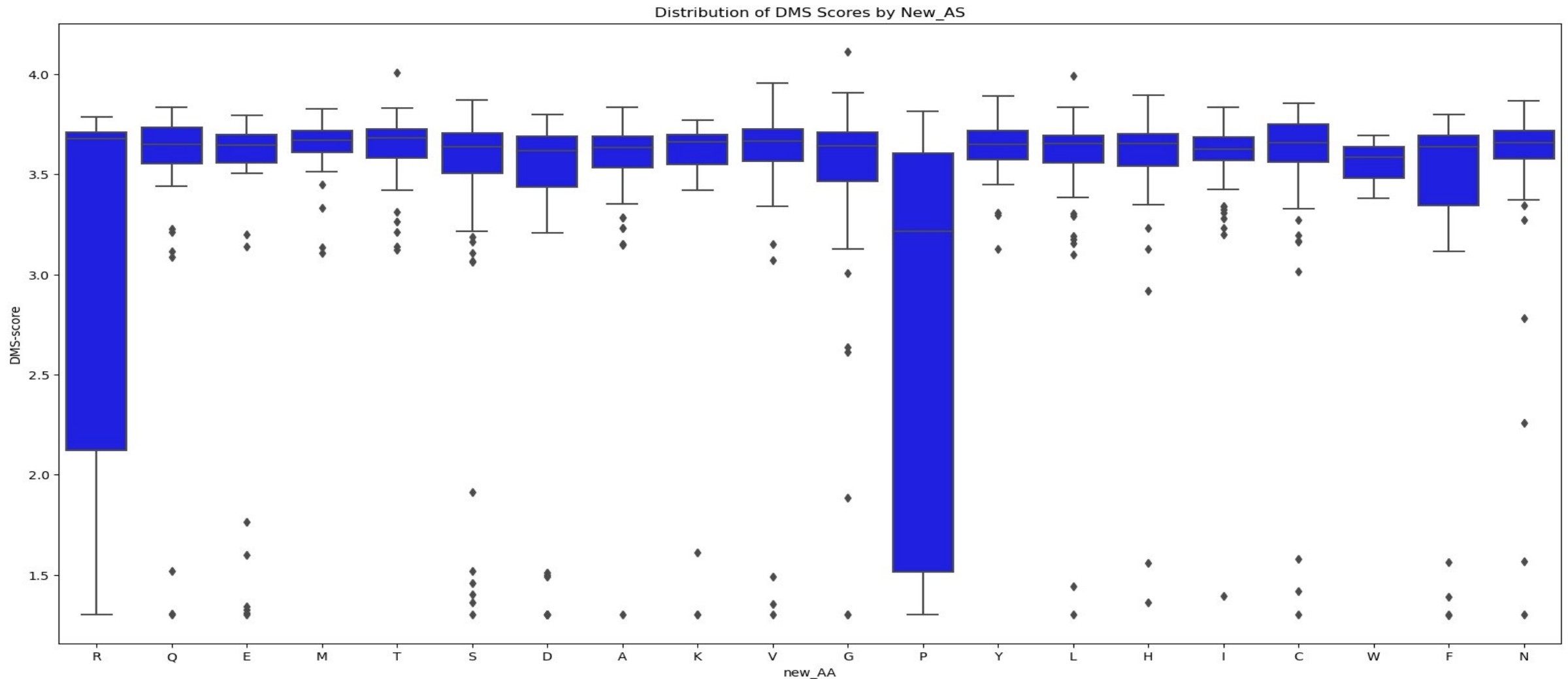
Data analysis group project 2023

A presentation by Roman Kurley, Angela Ma, Lisa Duttenhöfer, Rebecca Ress

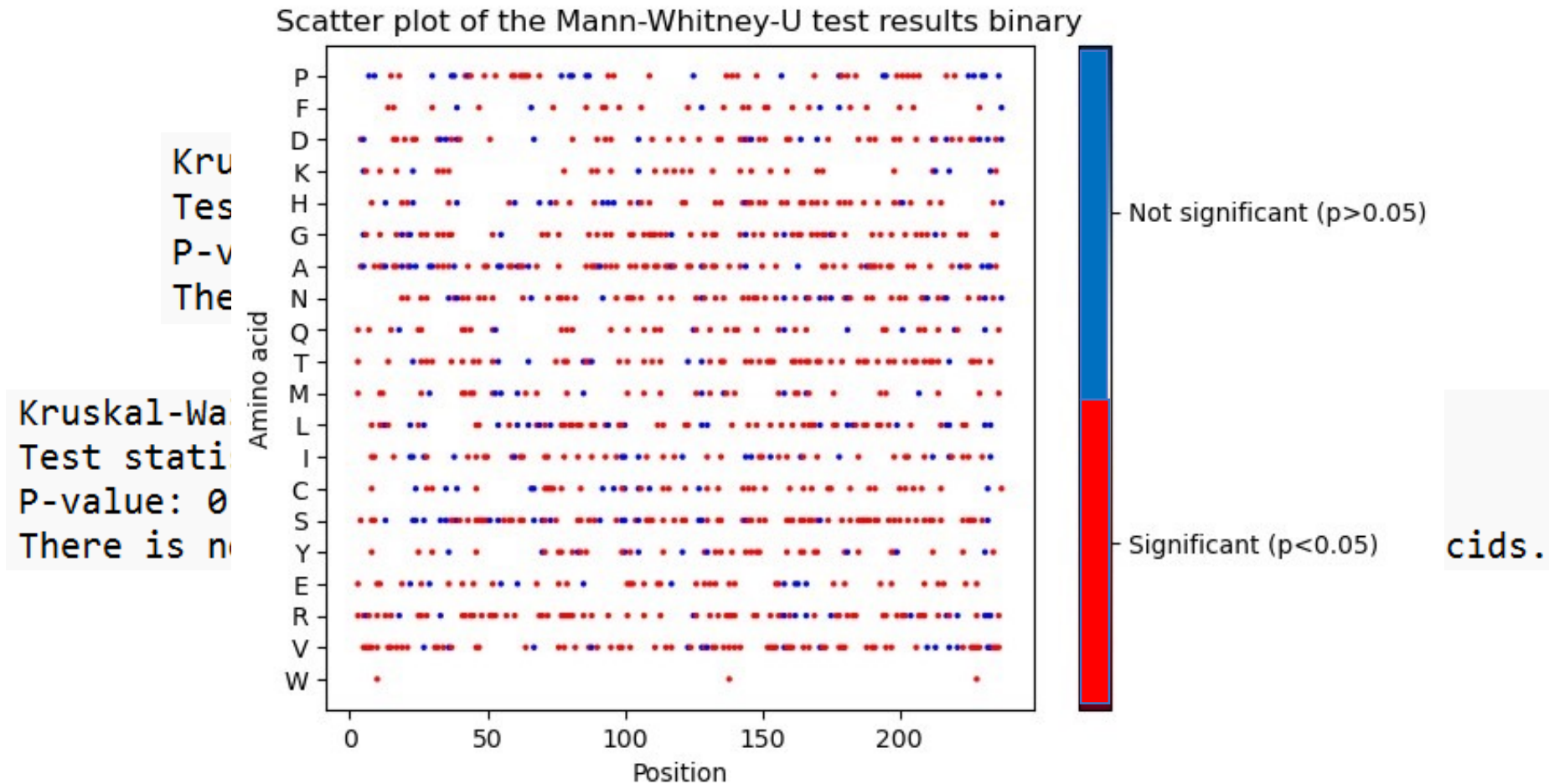
Our dataset



Does the position or new amino acid have a bigger influence on dms-scores?



Does the position or new amino acid have a bigger influence on dms-scores?



Amino acid properties

Unmutated sequence: ERTIFFKDDGNYKTRA**EV**KFEGDTLVNRIELKG

AEVRFEG

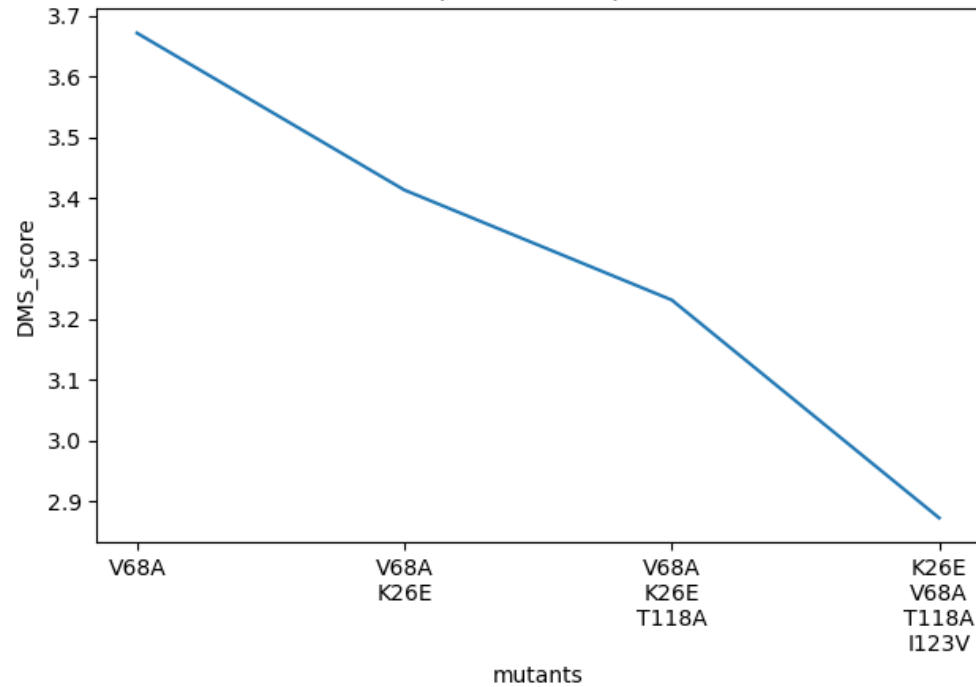
AEVRFEG

Neighbourhood

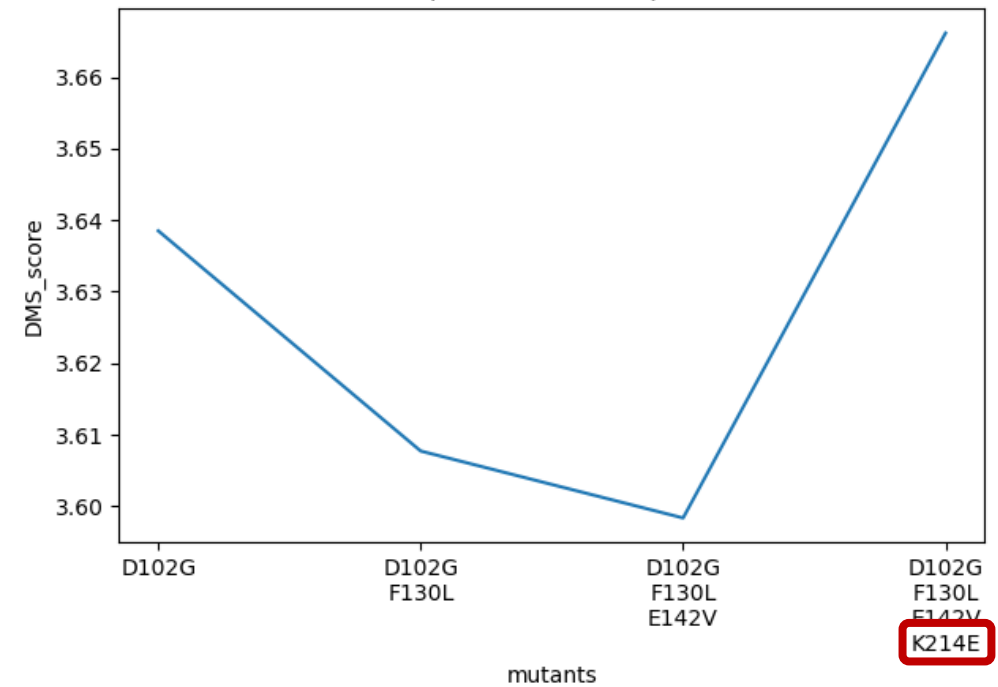
	Mutation	Neighbourhood	Molecular Weight	Residue Weight	pKa1	pKb2	pI4	H	VSC	P1	P2	SASA	NCISC
972	K113R	AEVRFEG	914.97	788.87	15.38	66.42	40.61	-0.64	443.5	63.3	1.069	12.219	0.337986
973	K113R-unmut	AEVKFEG	886.96	760.86	15.39	66.33	39.59	0.39	438.5	64.1	0.997	11.917	0.312107

Sequential Mutants

Example without Epistasis



Example for Positive Epistasis



Epistasis



mutations can have different effects on their own than when they occur together

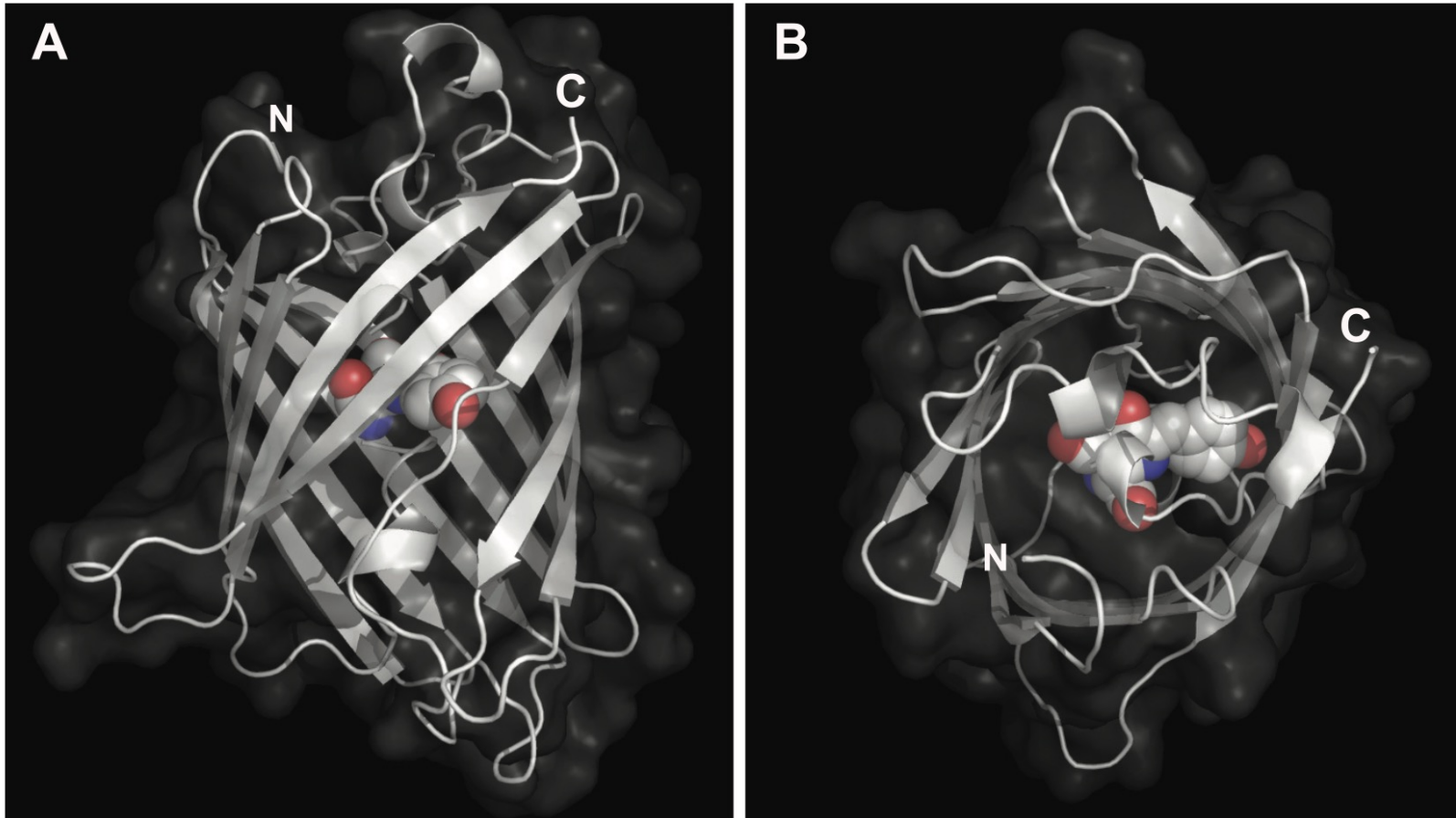


masking effect



in our case: interaction of amino acid residues within the protein

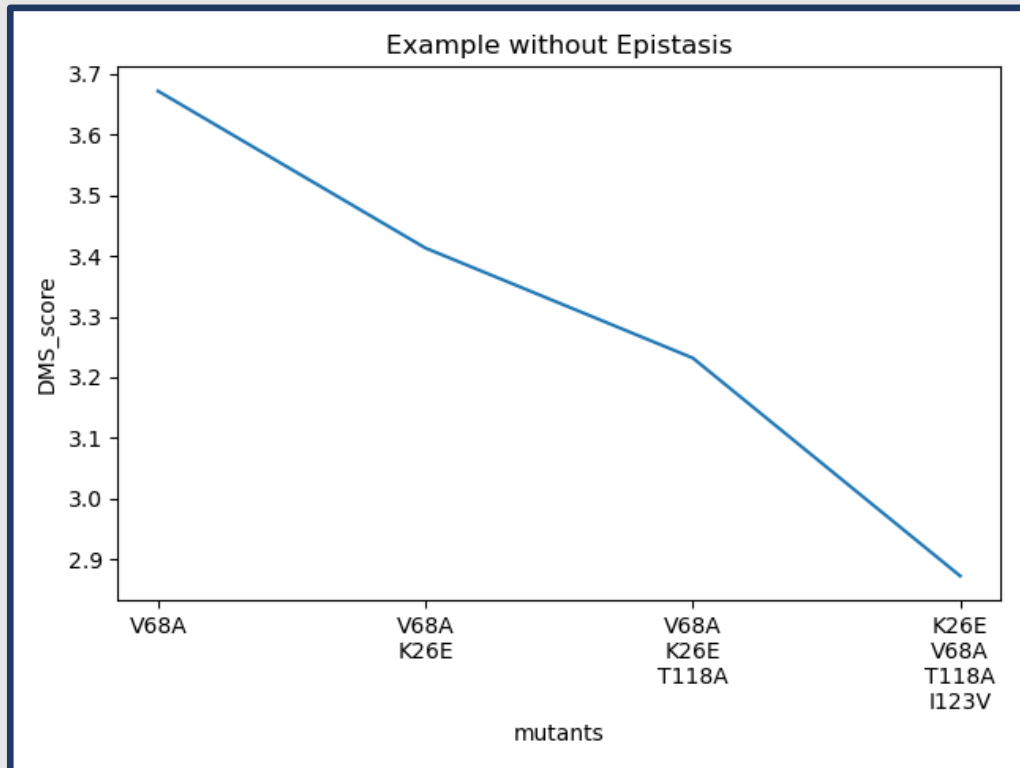
Structural Analysis



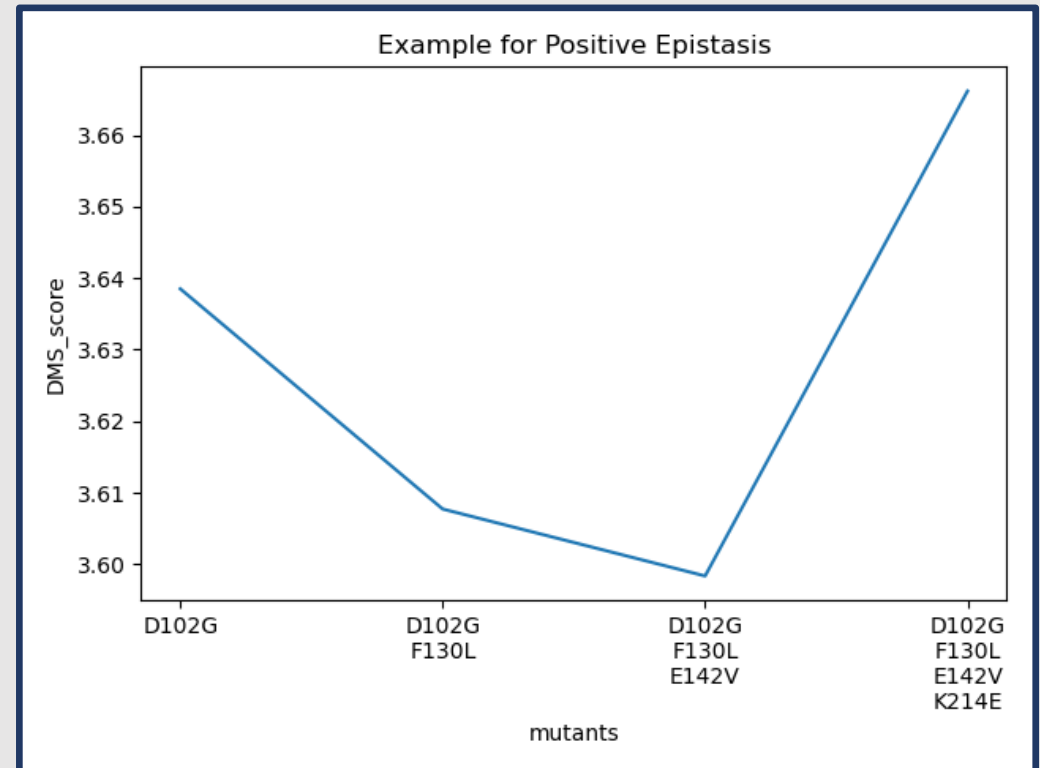
→ epistatic effects are stronger for buried amino acid residues

Analysis of stabilizing mutations

WITHOUT EPISTASIS

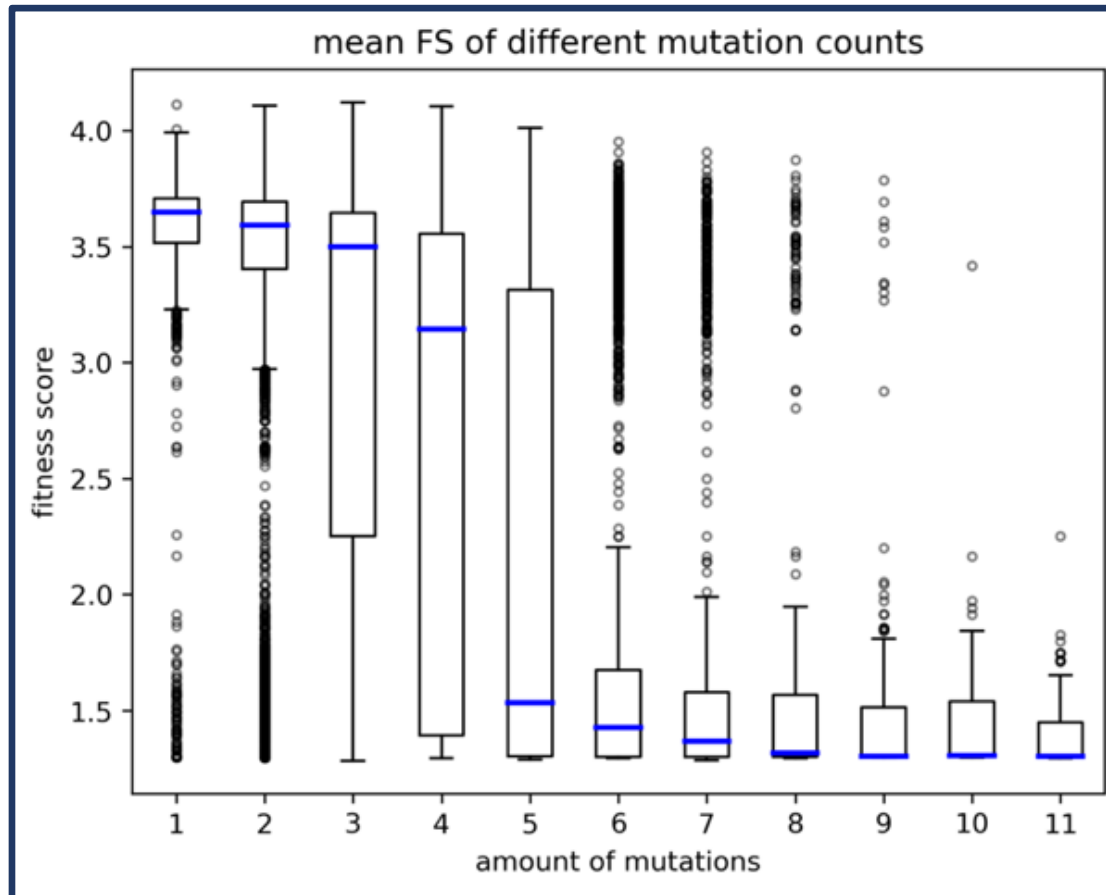


POSITIVE EPISTASIS



➔ Find mutations that have a positive epistatic effect on mutants

The impact of the amount of mutations



- Big impact of mutation count on protein fitness
- Mann - Whitney – U Test:
→ from seven mutations on protein is mostly non-functional
- Involve in further analysis

Ranking results

ranking3_score

V163A 388.828780
I171V 257.867360
S175G 219.571456
K113R 216.440643
I167V 212.858116
D117G 188.616871
K156R 187.935465
N144D 180.941018
K214E 180.634375
D129G 173.797917
K158R 172.884196
N105S 169.893861
I167T 168.779007
I123V 164.639468
T38A 164.520162

Formel
1

ranking5_score

V163A 5006189.732314
I171V 3284110.661276
S175G 3094377.775957
I167V 2866717.401939
K113R 2701526.013427
K214E 2389711.380513
K156R 2353862.372659
T38A 2150345.797784
D129G 2146870.162859
N144D 2143126.811354
D117G 2134821.905162
N105S 2095489.897094
I167T 2058640.353017
K158R 2051455.254767
I123V 2021080.178848

Formel 2

ranking5_score_weighted1

V163A 4430848.344827
I171V 4293815.100407
K214E 3983177.620595
F99L 3863113.192576
S175G 3802096.004937
T97A 3376546.402760
K158R 3332382.935368
K113R 3325591.957366
T43A 3237913.089066
N144D 3214758.885099
F223S 3133546.974888
K79R 3102521.890041
I167V 3059591.012304
M153V 3039109.402483
N121S 2976874.671938

Formel 3

How to define stabilizing mutations

1. Wie viel besser sind die fscores der Mutanten die die Mut beinhalten im Gegensatz zu den anderen?

→ **Differenz** = $\text{mean_aller_fscores_MIT} - \text{mean_aller_fscores_OHNE}$

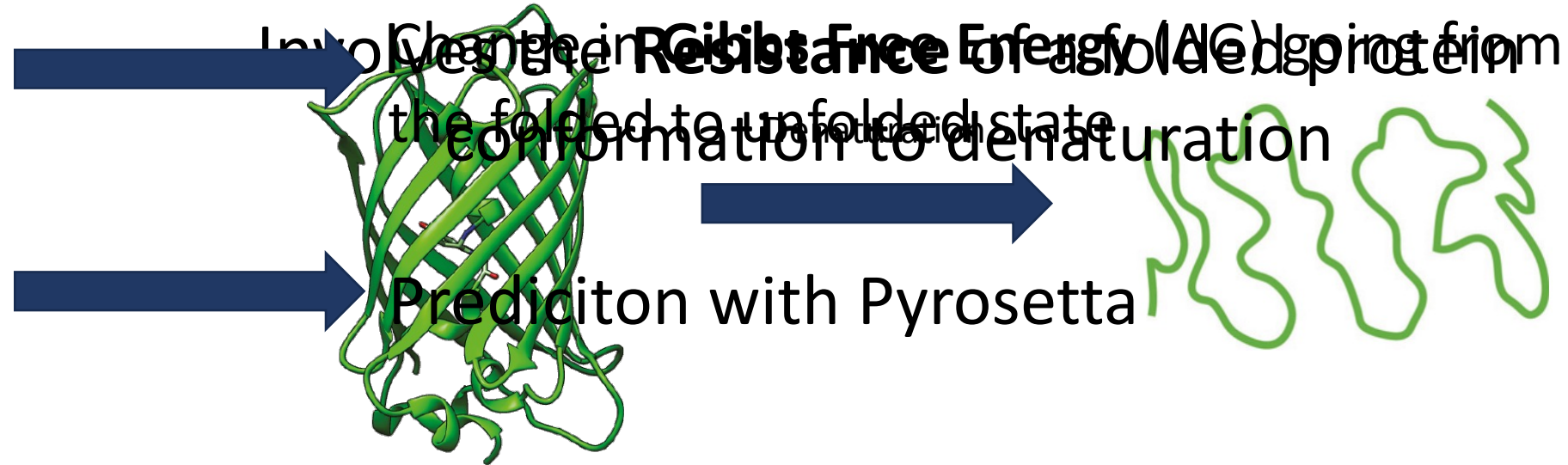
→ Wenn stabilisierend (positiver Effekt) → $\text{MIT} > \text{OHNE} \rightarrow \text{Differenz} > 0$

2. Sehr unterschiedlich wie oft eine Mut im Datenset vorkommt

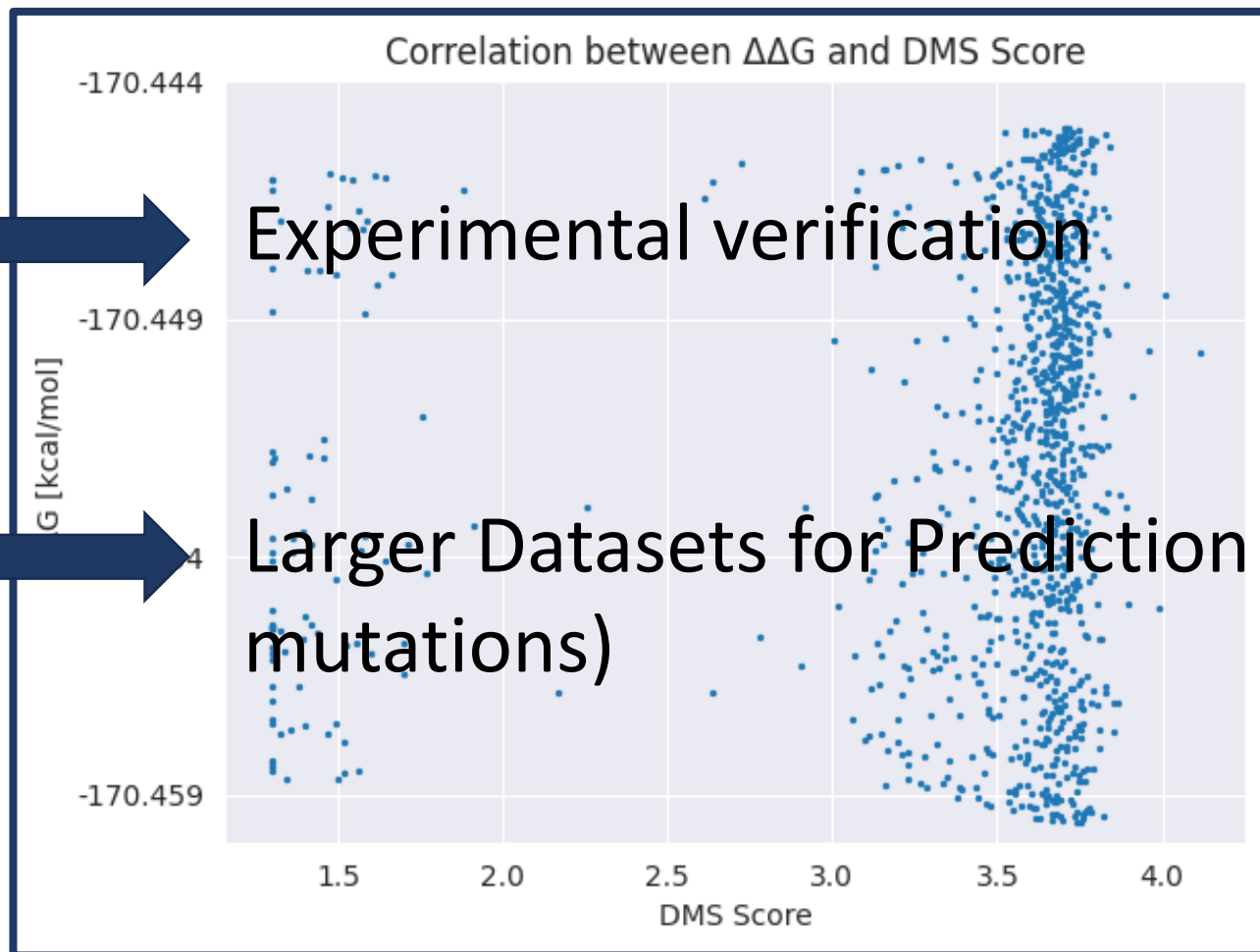
→ Einbeziehen (**Anzahl der Vorkommnisse** einer Mut im Datenset)

Protein Stability

Thermodynamic stability



Relationship between $\Delta\Delta G$ and DMS Score



CONCLUSION:

- PyRosetta used as Prediction Model
- $\Delta\Delta G$ = Difference between ΔG of the WT and Mutants
- No correlation between $\Delta\Delta G$ and DMS Score

Thanks for listening!