NIOPEK LAB

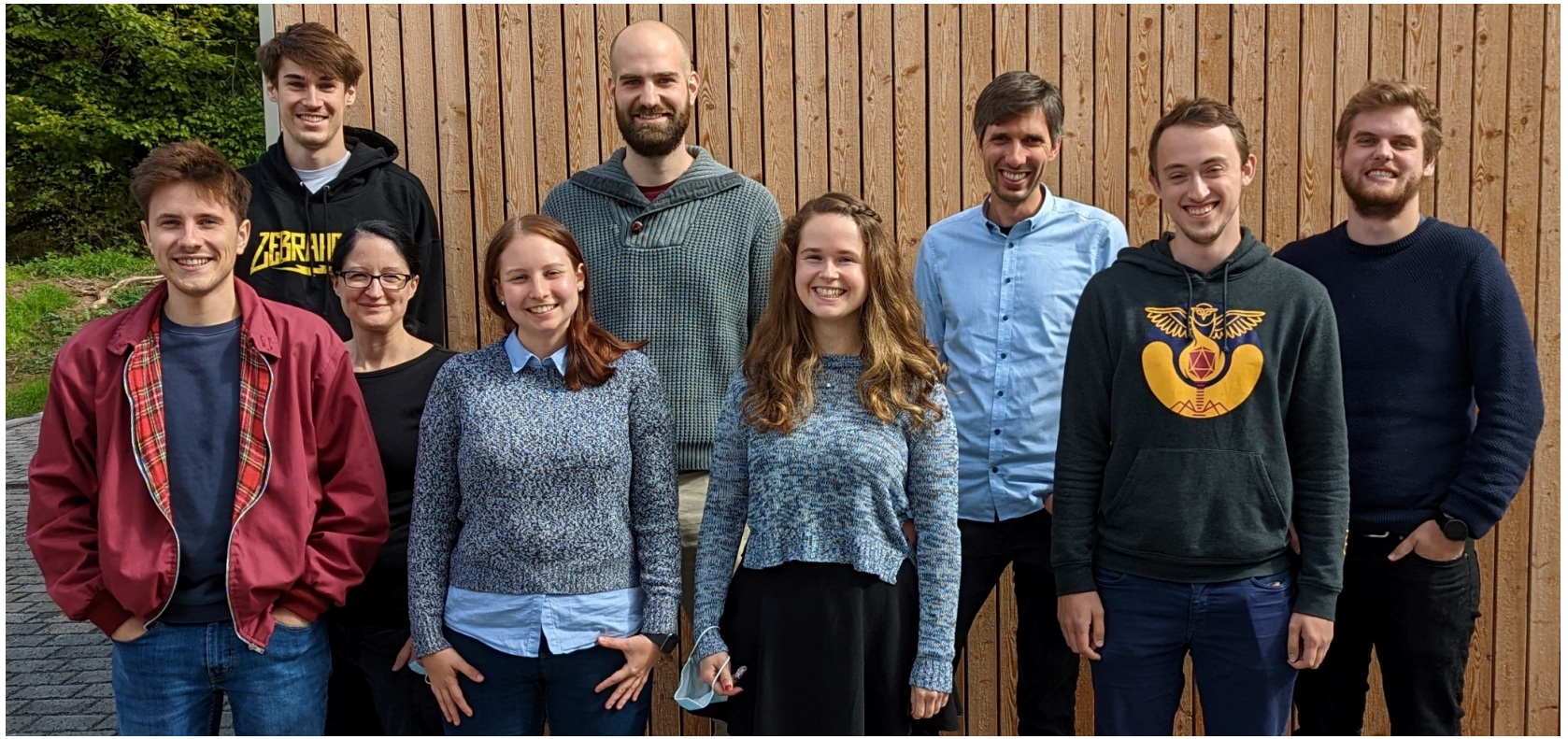
UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

Deep mutational scanning data analysis to reveal sequence-function relationships in Python

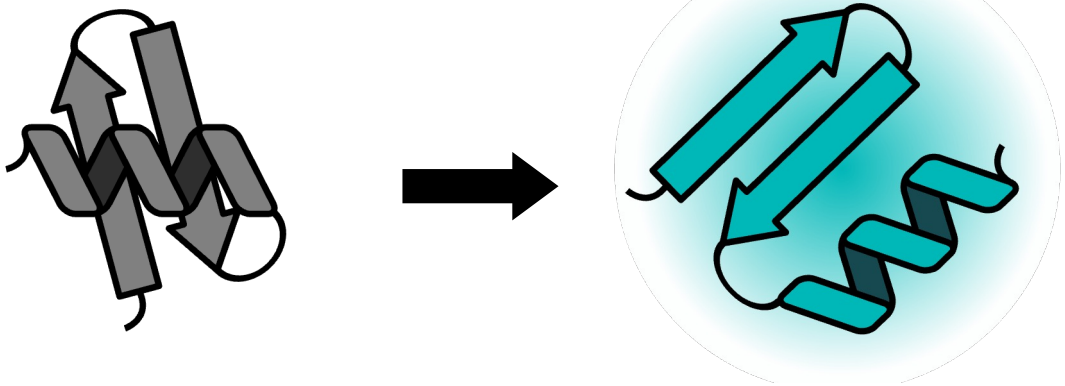
Supervisors:
Prof. Dominik Niopek
Jan Mathony

Tutor:
Benedict Wolf

Niopek lab @ IPMB, Heidelberg University

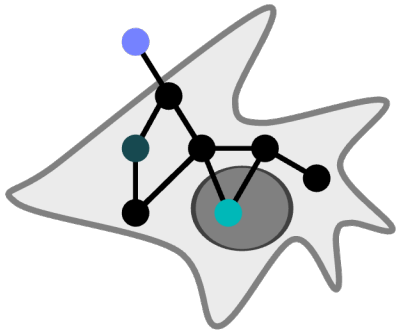


Protein engineering

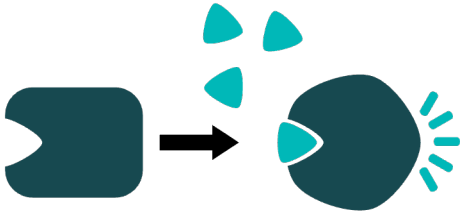


Wildtype protein

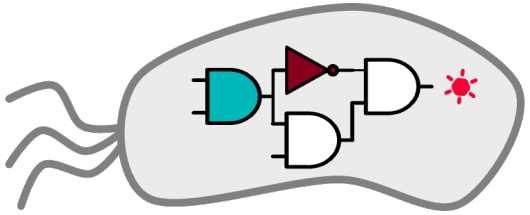
Desired function



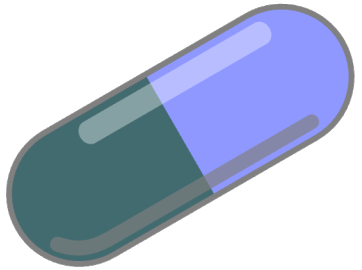
Basic research



Biosensor



Bioengineering



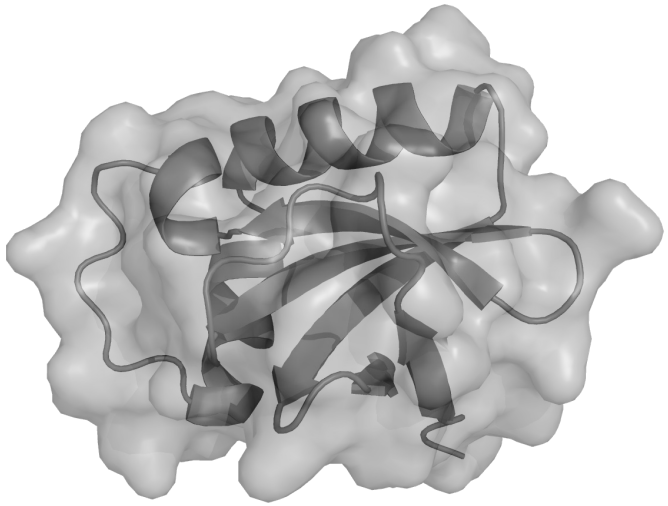
Therapeutics

From protein sequence to function

NIOPEK LAB

UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

```
>Protein
MANKTYKIGKNAGYDG
CGLCLAAISENEAIKV
KYLRDICPDYDGDDKA
EDWLRWGTDSRVKAAA
LEMEQYAYTSVGMASC
WEFVEL
```



Sequence

Structure

Function

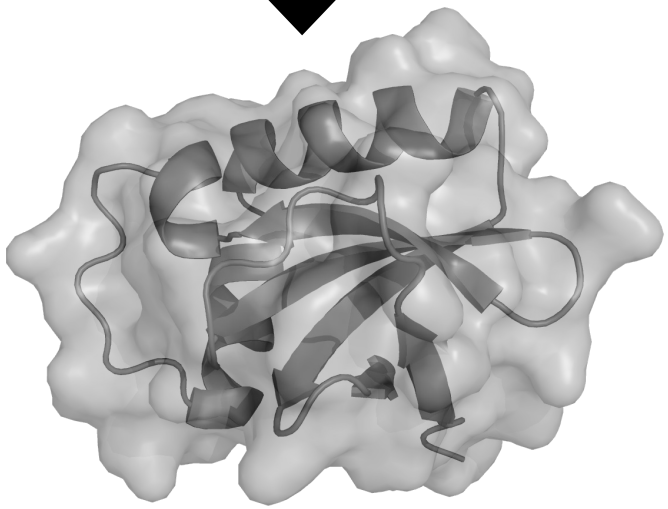
From protein sequence to function

NIOPEK LAB

UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

?

```
>Protein
MANKTYAIGKNAGYDG
CGLCLAAISENEAIKV
KYLRDICPDYDGDDKA
EDWLRWGTDSRVKAAA
LEMEQYAYTSVGMASC
WEFVEL
```

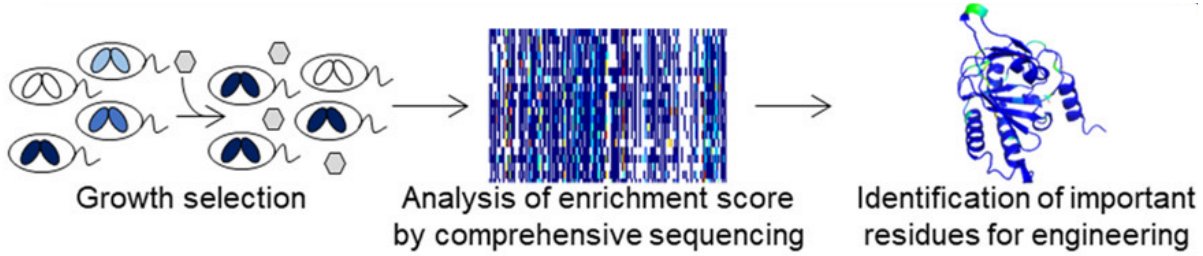


Sequence

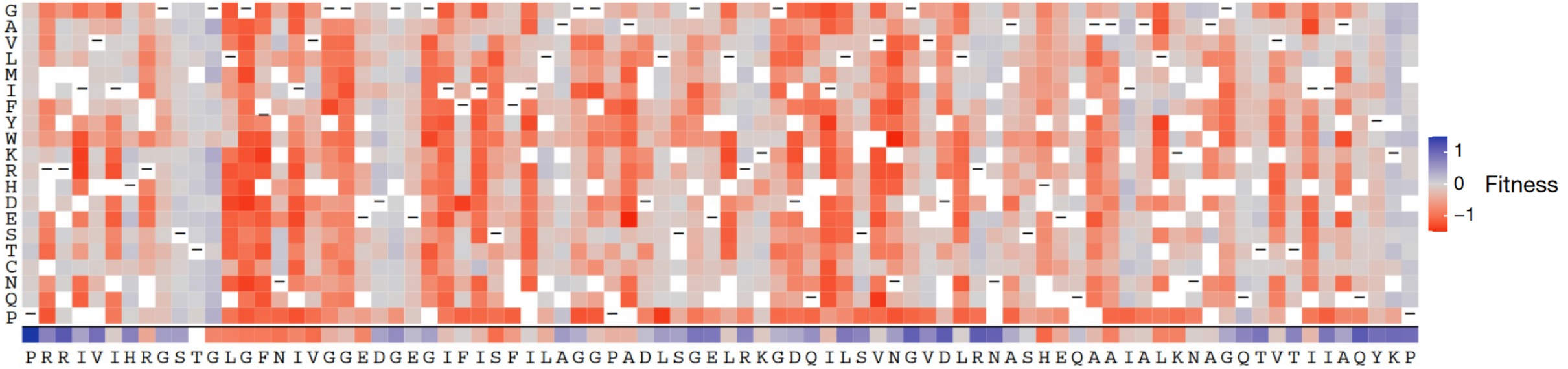
Structure

Function

Deep mutational scanning (DMS)

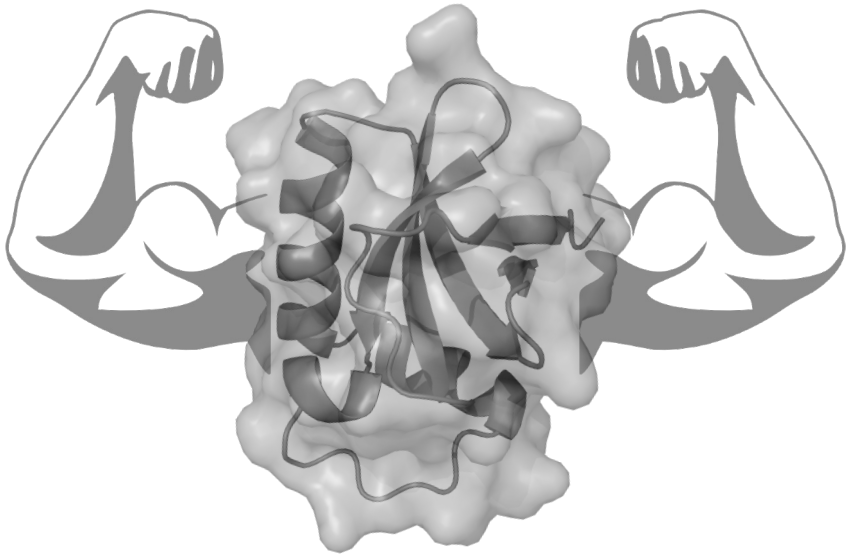


Ogawa et al., Acs Synth. Bio., 2022



Faure et al., Nature, 2022

ProteinGym



A **substitution benchmark** of **~1.5M missense variants** across **87 DMS datasets**
(Notin et al., Arxiv, 2022)

- Do you observe amino acid-specific substitution patterns?
- Does the mutation data relate secondary structure elements or functional sites?
- Do the observed trends correlate between proteins?

- Can you identify a subset of features that can help with the **prediction of mutation tolerance**?
- How do these compare to **evolutionary conservation**?