Helix Engineering: Combining the Power of 3DM with AI to Disrupt Protein Engineering
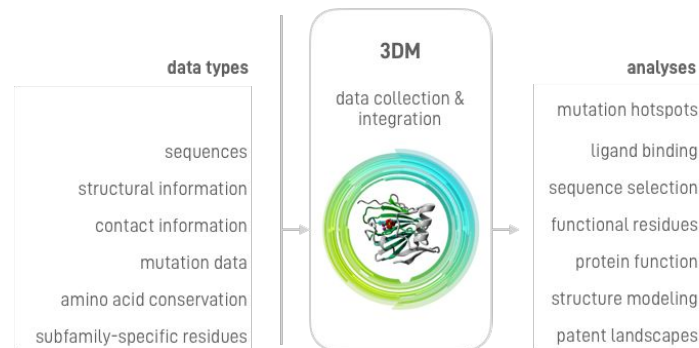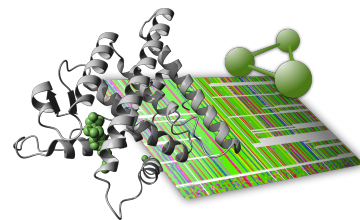
Bio-Prodict

Bio-Prodict's mission is "Catalyzing Protein Research".

Bio-Prodict collaborates with
- 7 out of top 10 pharmaceutical companies
- biotech companies
- academic & research institutes

3DM
- a protein data and analysis platform
- better understand the role of amino acids
- design better experiments
- widely used in the protein engineering field

data types	3DM	analyses
	data collection & integration	
sequences		mutation hotspots
structural information		ligand binding
contact information		sequence selection
mutation data		functional residues
amino acid conservation		protein function
subfamily-specific residues		structure modeling
		patent landscapes

3DM collects many types of data for all sequences of complete superfamilies:

- Mutation data from literature, patents, mutation databases (we have biggest literature mutation collection in the world)
- Protein-protein and ligand contact data.
- Evolutionaly alignment fingerprints (e.g. conservation, correlated mutations, aa distributions, etc)
- Disease information, chemical reaction data, SNP data

3DM contains many tools for the analysis of all this data:

- Tools for visualization of data in any protein structure
- Tools for easy analysis of the alignment.
- Cornet, for the analysis of correlated mutations
- All data and tools are connected in 3DM via the alignment.
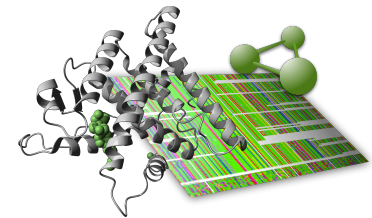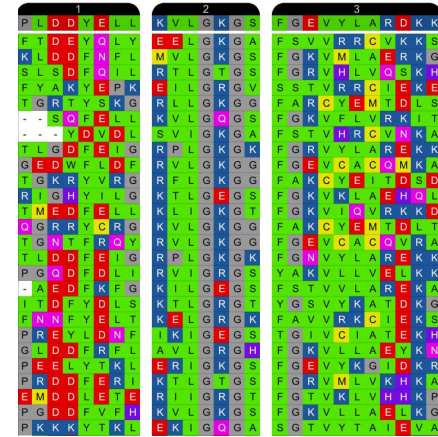
- The next step? Combining this data with Machine Learning
- Built on 3DM features, using data from >33.000 3DM systems for ==100% human variant coverage.==
- Helix Pathogenicity Predictor
  - Top of the line clinical predictions
  - High standards for performance assessment
  - Intuitive reports, web-interface, and APIs available
- Outperforms all other pathogenicity predictors in novel datasets for published cancer research and datasets with real novel clinical data (e.g. BRCA1, CHEK2, PALB2).
- **Predicts if a variant is pathogenic or not with ==95% accuracy== on the complete human exome.**
- **Applicable to any organism (e.g. for strain engineering)**
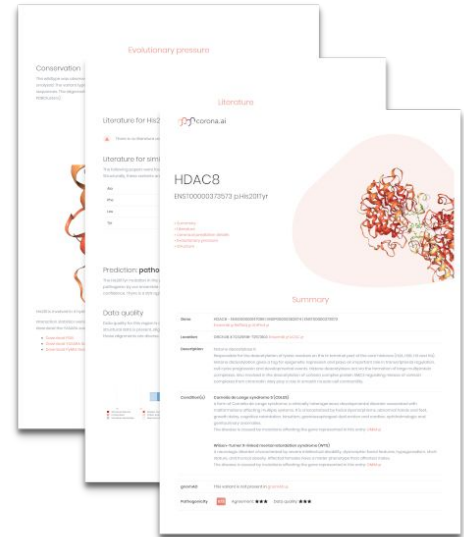
Bio-Prodict 3DM + AI

- Our next frontier: **Helix Engineering**

- Use Helix Pathogenicity strategy to enable AI based Protein Engineering
  - 56 number of tests have been performed with public data and several pilots are currently being run using customer data.
  - Low number of mutations needed (50 to 100) for the prediction of a high quality second round of evolution.
  - Can be applied to different protein features, such as activity, stability, co-factor binding, and many others.

- **Helix Engineering is a generic solution that boosts Protein Engineering with 3DM and AI insights**
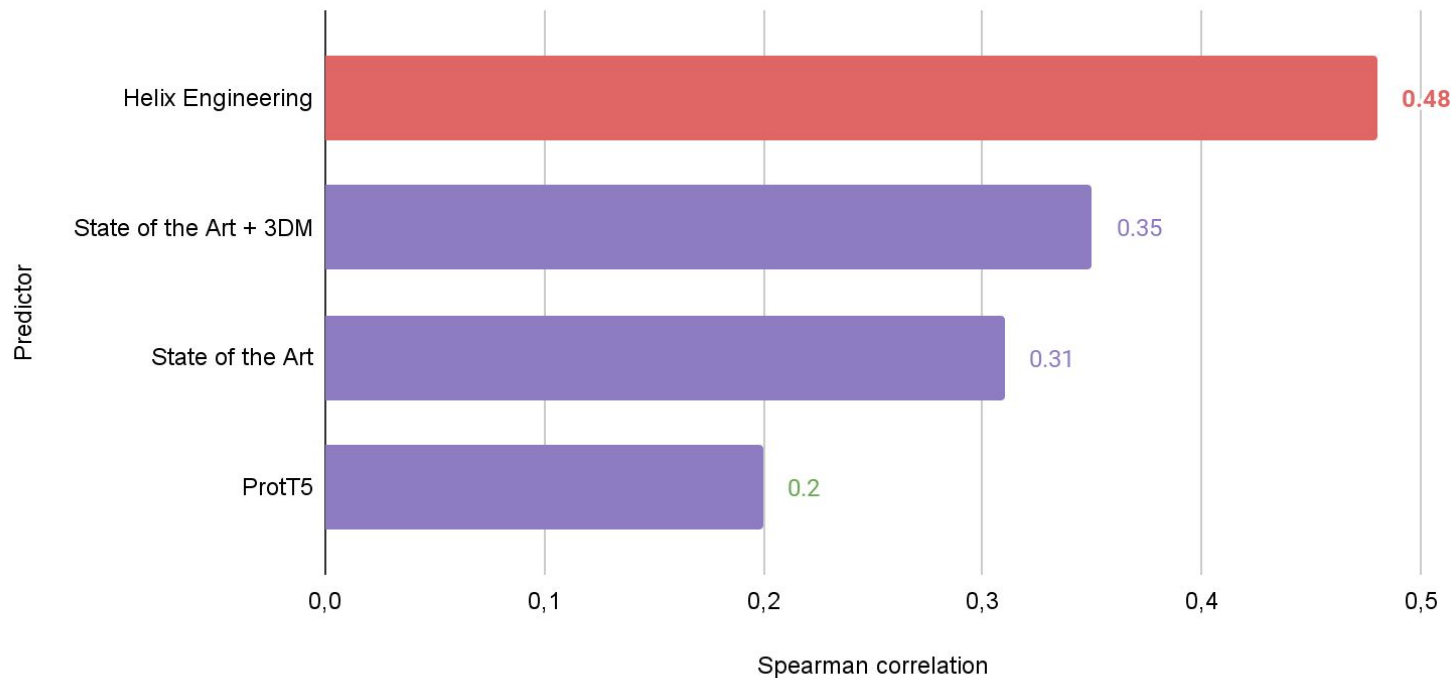
Prediction performance

Measure the number of hits (better than wild-type) per experiment on average, compared to random selection in 56 test sets.
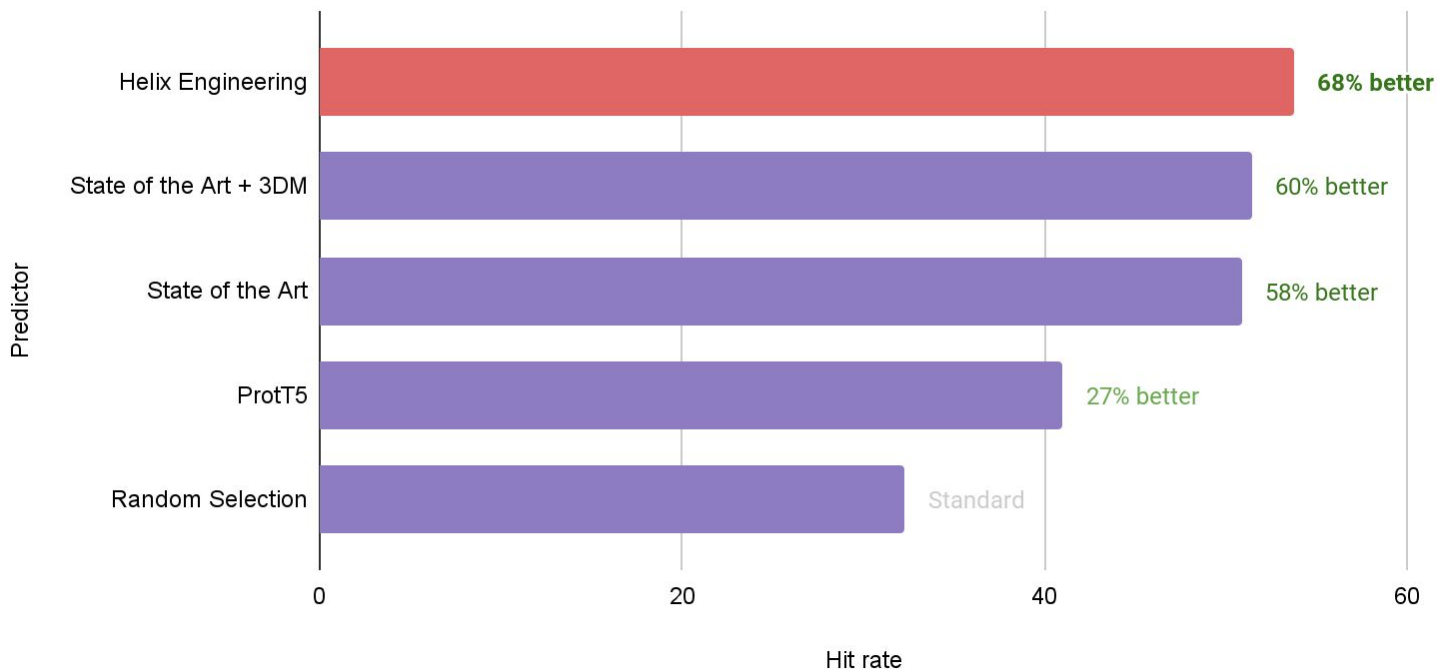
Prediction performance

Spearman Correlation (ranking performance) with true fitness of all variants in 56 test sets

Conclusion

- Using modern machine learning techniques, we can increase fitness prediction performance by **>60%** over legacy techniques, like random selection or one-hot encoding.

- By intelligently selecting training variants using a generic 3DM based search strategy, we are able to increase this up to **100%**

- In a collaborative approach we combine domain expertise to get the most out of each evolutionary run using a minimum of mutations to measure.

- We offer a full-service package for protein engineering with low turnaround times and low overhead

- The result: **Faster and better protein engineering results**

Q&A

More information:
www.bio-prodict.nl
helixlabs.ai

Whitepaper: https://arxiv.org/abs/2104.01033

bio-prodict
protein predictions

Helix Labs