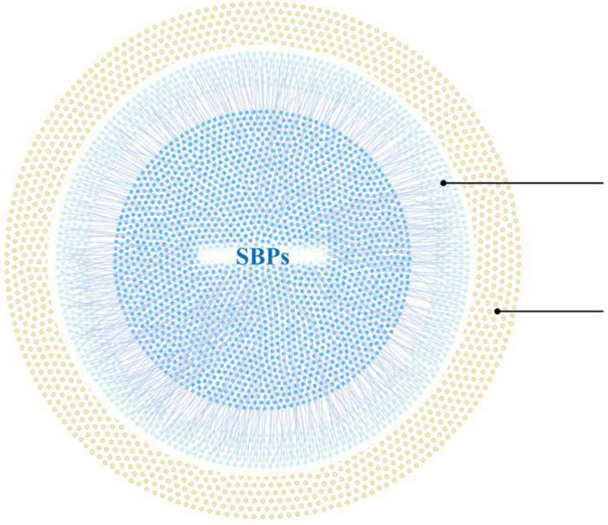
Expanding the sequence spaces of synthetic binding protein using deep learning-based framework ProteinMPNN

Yanlin LI, Wantong JIAO, Ruihan LIU, Xuejin DENG, Feng ZHU, Weiwei XUE

Frontiers of Computer Science, DOI: 10.1007/s11704-024-31060-3

Problems & Ideas

- Problems of conventional protein design approaches:
  - Directed evolution explores only the sequence space regions around natural proteins.
  - Site-directed mutagenesis is highly dependent on the physiological properties and structure of parental protein.
  - Protein de novo design software based on physical methods relies on energy functions to predict the stability and folding behavior.
- Ideas: Expanding the sequence spaces of synthetic binding protein using advance deep learning-based framework ProteinMPNN.

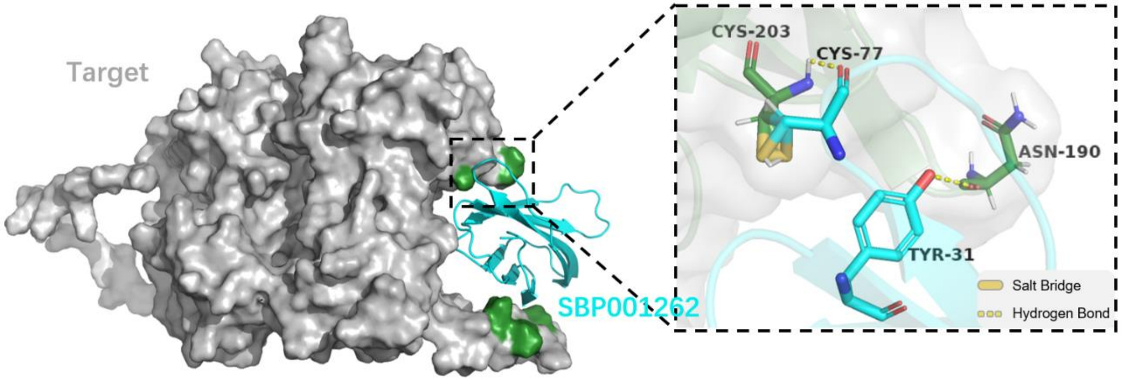


ProteinMPNN explores the space of synthetic proteins for the first time

ProteinMPNN can iteratively explore the synthetic protein space an unlimited number of times

Main Contributions

- Contributions:
  - Comprehensive bioinformatics analysis suggested that the new protein sequences generated by ProteinMPNN have improved solubility and stability compared with the original SBPs;
  - It is found that the sequences generated based on monomer structure are better than that of complex structure in terms of solubility and stability. In contrast, the sequences designed based on the complex structure have better performance in calculated binding energy;
  - Eight scaffolds with significantly improved solubility and stability were screened, including Neocarzinostatin based binder, Diabody, CI2 based binder, scFv, Repebody, Fab, Affilin, and Evibody;
  - Synthetic proteins designed by ProteinMPNN show superior performance compared to classical protein engineering methods.



Analysis of the interaction interface between SBP001262 and its target protein.