**Protein fitness landscape paper list**

**（一）基准测试与数据集**

1. Dallago, C. *et al.* FLIP: Benchmark tasks in fitness landscape inference for proteins. *bioRxiv*, 2021.2011. 2009.467890 (2021).
2. Groth, P. M., Michael, R., Salomon, J., Tian, P. & Boomsma, W. Flop: Tasks for fitness landscapes of protein wildtypes. *bioRxiv*, 2023.2006. 2021.545880 (2023).
3. Thomas, N., Agarwala, A., Belanger, D., Song, Y. S. & Colwell, L. J. Tuned Fitness Landscapes for Benchmarking Model-Guided Protein Design. *bioRxiv*, 2022.2010. 2028.514293 (2022).
4. Notin, P. *et al.* Proteingym: Large-scale benchmarks for protein fitness prediction and design. *Advances in Neural Information Processing Systems* **36** (2024).
5. Sun, H. *et al.* Accelerating protein engineering with fitness landscape modeling and reinforcement learning. *bioRxiv*, 2023.2011. 2016.565910 (2023).

【注：应归为（六）更合适，为平衡数量放到这里了】

**（二）****深度生成模型与变分方法**

1. Riesselman, A. J., Ingraham, J. B. & Marks, D. S. Deep generative models of genetic variation capture the effects of mutations. *Nature methods* **15**, 816-822 (2018).
2. Ding, X., Zou, Z. & Brooks III, C. L. Deciphering protein evolution and fitness landscapes with latent space models. *Nature communications* **10**, 5644 (2019).
3. McGee, F. *et al.* The generative capacity of probabilistic protein sequence models. *Nature communications* **12**, 6302 (2021).
4. Ngo, N. K., Tran, T. V., Duy Nguyen, V. T. & Hy, T. S. Latent-based Directed Evolution accelerated by Gradient Ascent for Protein Sequence Design. *bioRxiv*, 2024.2004. 2013.589381 (2024).
5. Widatalla, T., Rafailov, R. & Hie, B. Aligning protein generative models with experimental fitness via Direct Preference Optimization. *bioRxiv*, 2024.2005. 2020.595026 (2024).

**（三）传统深度学习与机器学习方法**

1. Gelman, S., Fahlberg, S. A., Heinzelman, P., Romero, P. A. & Gitter, A. Neural networks to learn protein sequence–function relationships from deep mutational scanning data. *Proceedings of the National Academy of Sciences* **118**, e2104878118 (2021).
2. Kulikova, A. V., Diaz, D. J., Loy, J. M., Ellington, A. D. & Wilke, C. O. Learning the local landscape of protein structures with convolutional neural networks. *Journal of Biological Physics* **47**, 435-454 (2021).
3. Zhang, Y., Gao, Z., Tan, C. & Li, S. Z. Efficiently predicting protein stability changes upon single-point mutation with large language models. *arXiv preprint arXiv:2312.04019* (2023).
4. Blaabjerg, L. M. *et al.* Rapid protein stability prediction using deep learning representations. *Elife* **12**, e82593 (2023).
5. Sun, J., Zhu, T., Cui, Y. & Wu, B. Structure-based self-supervised learning enables ultrafast protein stability prediction upon mutation. *The Innovation* **6** (2025).

**（四）语言模型**

1. Notin, P. *et al.* Tranception: Protein fitness prediction with autoregressive transformers. *International Conference on Machine Learning.* 16990-17017 (PMLR).
2. Gelman, S. *et al.* Biophysics-based protein language models for protein engineering. *bioRxiv* (2024).
3. Zhou, Z. *et al.* Enhancing efficiency of protein language models with minimal wet-lab data through few-shot learning. *Nature Communications* **15**, 5566 (2024).
4. Jiang, F. *et al.* A general temperature-guided language model to design proteins of enhanced stability and activity. *Science Advances* **10**, eadr2641 (2024).
5. Tran, T. V. & Hy, T. S. Protein design by directed evolution guided by large language models. *IEEE Transactions on Evolutionary Computation* (2024).

**（五）突变效应预测**

1. Mansoor, S., Baek, M., Juergens, D., Watson, J. L. & Baker, D. Zero‐shot mutation effect prediction on protein stability and function using RoseTTAFold. *Protein Science* **32**, e4780 (2023).
2. Mansoor, S., Baek, M., Juergens, D., Watson, J. L. & Baker, D. Accurate mutation effect prediction using rosettafold. *BioRxiv*, 2022.2011. 2004.515218 (2022).
3. Gurusinghe, S. N., Wu, Y., DeGrado, W. & Shifman, J. M. ProBASS–a language model with sequence and structural features for predicting the effect of mutations on binding affinity. *bioRxiv* (2024).
4. Lan, T. *et al.* Generating mutants of monotone affinity towards stronger protein complexes through adversarial learning. *Nature Machine Intelligence* **6**, 315-325 (2024).
5. Cheng, P. *et al.* Zero-shot prediction of mutation effects with multimodal deep representation learning guides protein engineering. *Cell Research* **34**, 630-647 (2024).

**（六）综合优化与生物学实验**

1. Yadav, A. J., Kumar, S., Maurya, S., Bhagat, K. & Padhi, A. K. Interface design of SARS-CoV-2 symmetrical nsp7 dimer and machine learning-guided nsp7 sequence prediction reveals physicochemical properties and hotspots for nsp7 stability, adaptation, and therapeutic design. *Physical Chemistry Chemical Physics* **26**, 14046-14061 (2024).
2. Hie, B. L. *et al.* Efficient evolution of human antibodies from general protein language models. *Nature Biotechnology* **42**, 275-283 (2024).
3. Ding, K. *et al.* Machine learning-guided co-optimization of fitness and diversity facilitates combinatorial library design in enzyme engineering. *Nature Communications* **15**, 6392 (2024).
4. Hsu, C., Nisonoff, H., Fannjiang, C. & Listgarten, J. Learning protein fitness models from evolutionary and assay-labeled data. *Nature biotechnology* **40**, 1114-1122 (2022).
5. Ma, Z. *et al.* EvoAI enables extreme compression and reconstruction of the protein sequence space. *Nature Methods*, 1-11 (2024).