

# Combination of geometrical learning and language models for bioinformatics

Выполнил: Клыпа Р.С.

Научный руководитель: Грудинин С.В.

МФТИ

14 декабря 2022 г.

- Предсказывание вторичной и третичной структуры РНК.
- Предсказывание динамики РНК.
- Предсказывания взаимодействий РНК-протеин и РНК-РНК.

## Предсказывают 3D структуры протеинов

- Evolutionary-scale prediction of atomic level protein structure with a language model. Zeming Lin, Halil Akin et al.  
<https://doi.org/10.1101/2022.07.20.500902>

## Предсказывают функции протеинов

- Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. Alexander Rives, Joshua Meier, et al.  
<https://doi.org/10.1073/pnas.2016239118>

## Предсказывают функции участков DNK

- DNA language models are powerful zero-shot predictors of non-coding variant effects. Gonzalo Benegas, Sanjit Singh Batra, Yun S. Song.  
<https://doi.org/10.1101/2022.08.22.504706>

# Deep learning for collective variables

## Предсказывают координаты реакции

- Learning molecular dynamics with simple language model built upon long short-term memory neural network. Sun-Ting Tsai, En-Jui Kuo and Pratyush Tiwary. <https://doi.org/10.1038/s41467-020-18959-8>

## Определяют набор наиболее информативных координат для симуляции

- Past-future information bottleneck for sampling molecular reaction coordinate simultaneously with thermodynamics and kinetics. Yihang Wang, Joao Marcelo Lamim Ribeiro and Pratyush Tiwary. <https://doi.org/10.1038/s41467-019-11405-4>

## Предсказывают связанные структуры лиганд и протеинов

- Gabriele Corso, Hannes Stärk, Bowen Jing, Regina Barzilay, Tommi Jaakkola. DiffDock: Diffusion Steps, Twists, and Turns for Molecular Docking. arXiv:2210.01776
- Zhuoran Qiao, Weili Nie, Arash Vahdat, Thomas F. Miller III, Anima Anandkumar. Dynamic-Backbone Protein-Ligand Structure Prediction with Multiscale Generative Diffusion Models. arXiv:2209.15171

## Решают motif-scaffolding проблему

- Brian L. Trippe, Jason Yim, Doug Tischer, David Baker, Tamara Broderick, Regina Barzilay, Tommi Jaakkola. Diffusion probabilistic modeling of protein backbones in 3D for the motif-scaffolding problem. arXiv:2206.04119

- Использование диффузионных сетей для анализа взаимодействий РНК.
- Акцент на молекулярной динамике.
- Мультимодельная система.