



НАУЧНАЯ КОНФЕРЕНЦИЯ

Суперкомпьютерные дни в России

Моделирование структуры подвижных участков белка с помощью GPU-ускоренной метадинамики с коллективными переменными на основе вариационных автоэнкодеров

К.Е. Копылов^{*}, Е.М. Кирилин^{*‡}, В.К. Швядас^{*§}

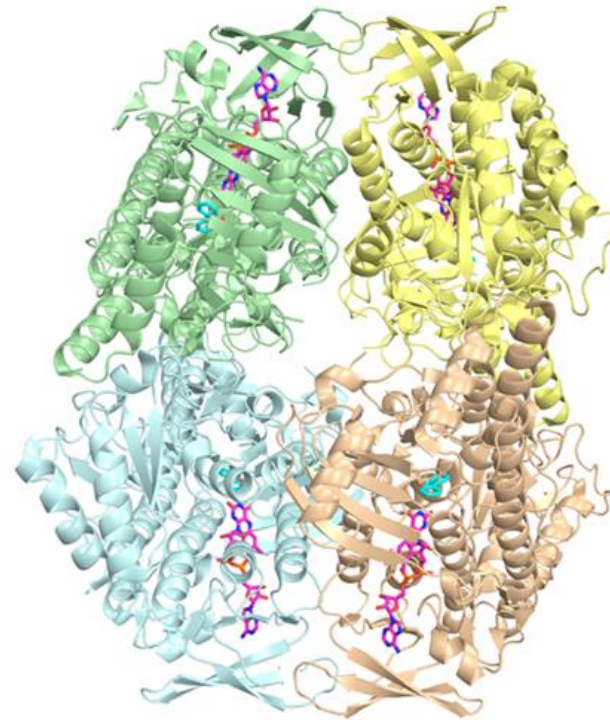
НИВЦ^{}, НИИ ФХБ имени А.Н. Белозерского[‡] и
факультет биоинженерии и биоинформатики[§]*

МГУ имени М.В. Ломоносова, Москва, Россия

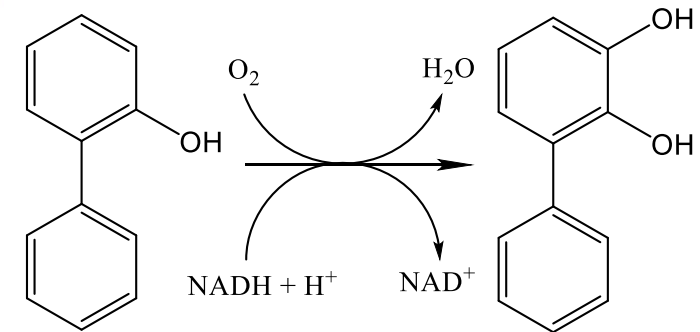


2-гидроксибифенил-3-монооксигеназа

- FAD и NADH-зависимый фермент
- Перспективный катализатор для фармацевтической промышленности
- **Отсутствует полноатомная структура активного центра - не разрешены петли Tyr256-Ile266 и Arg228-Val236**



Структура тетрамера 2-гидроксибифенил-3-монооксигеназы

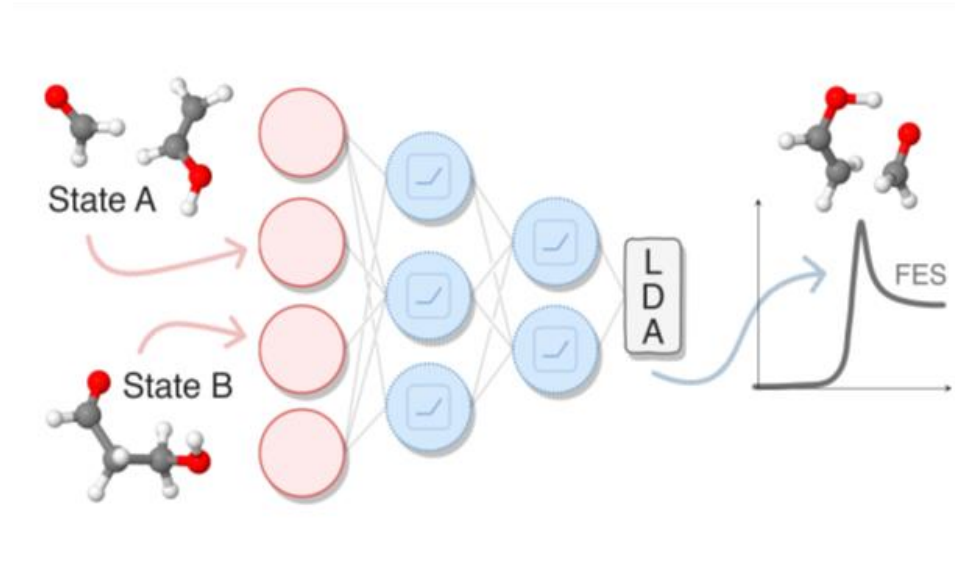


Реакция, катализируемая 2-гидроксибифенил-3-монооксигеназой



Метадинамика с нейросетевой CV

- Вариант молекулярной динамики с расширенным сэмплированием
- Исследование поверхности потенциальной энергии (ППЭ) в координатах «коллективных переменных» (collective variables, CV), в том числе дифференцируемых сложных функций, таких как нейронные сети
- В пространстве выходов добавляется энергия в виде гауссианов – действует на исходное пространство с помощью обратного распространения

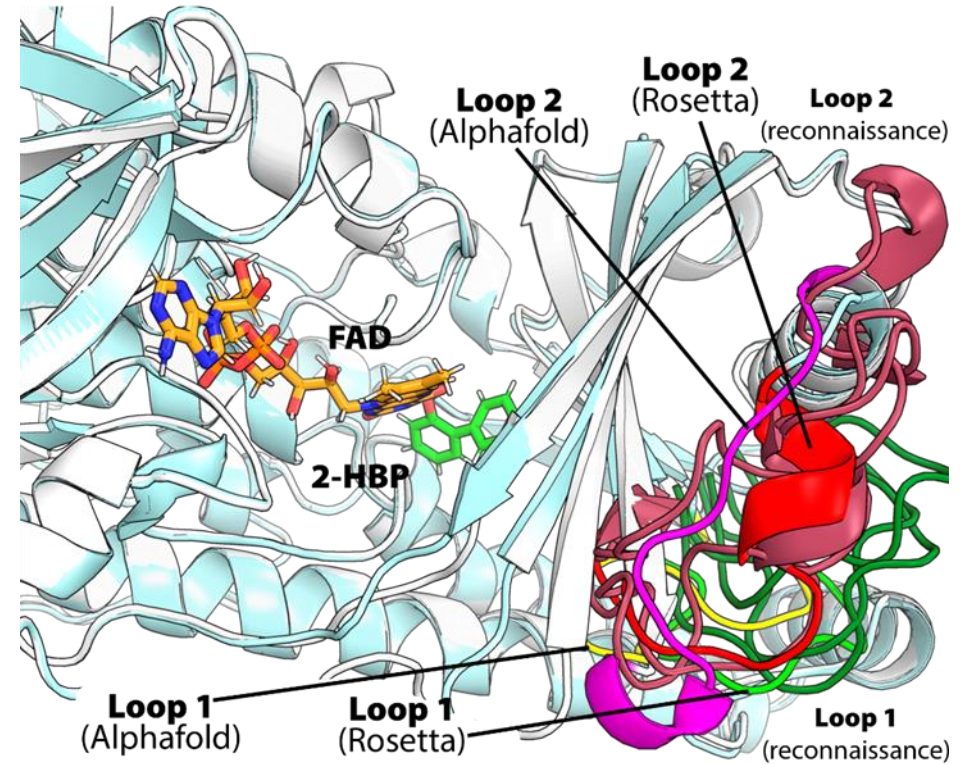
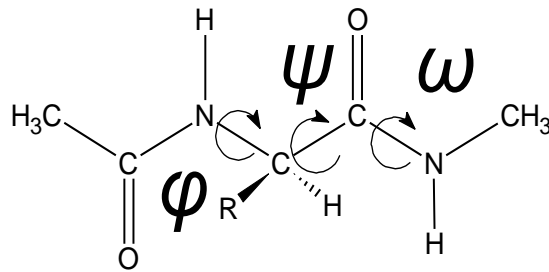
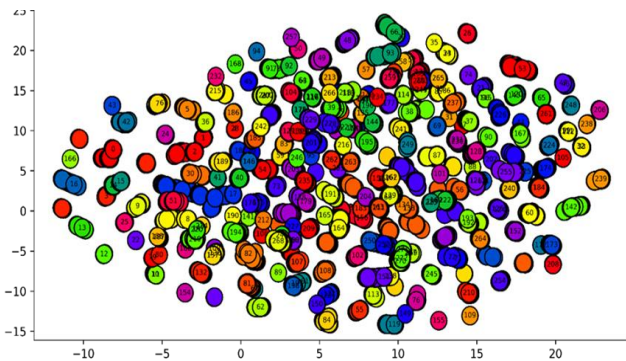


Bonati L., Rizzi V., Parrinello M. Data-driven collective variables for enhanced sampling // The journal of physical chemistry letters. ACS Publications, 2020. Vol. 11, № 8. P. 2998–3004.



Датасет

- Всего ~30 000 структур (AlphaFold, Rosetta, разведывательная метадинамика)
- 24 остатка, для каждого двугранные углы φ и ψ (48 всего)



Визуализация датасета двугранных углов участков HbrA с помощью UMAP с раскраской по HDBSCAN

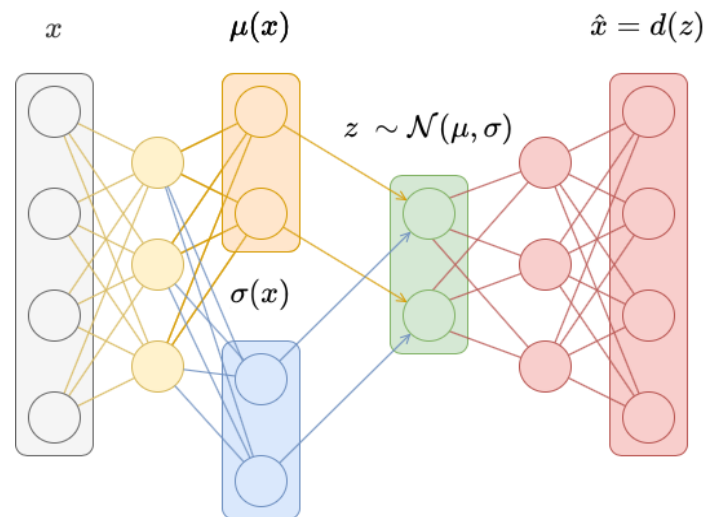
Положение подвижных участков в разных структурах датасета



Нейросетевая модель

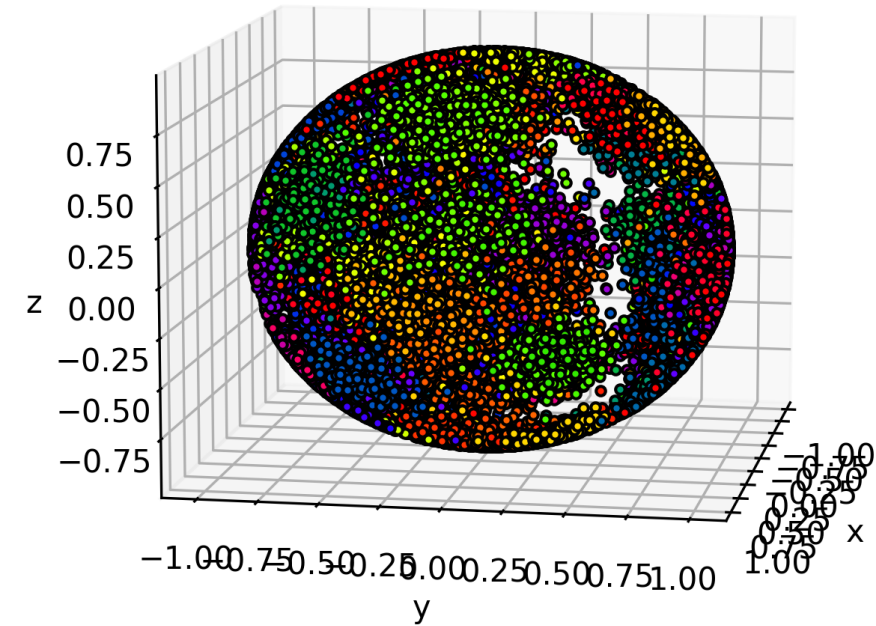
- Описано применение **вариационных автоэнкодеров** для понижения размерности третичных структур белков
- С учетом периодичности данных выбран **гиперсферический** вариационный автоэнкодер (распределение фон Мизеса-Фишера на многомерной сфере)
- Размерность входов и выходов - 48 двугранных углов
- Размерность латентного пространства – 3
- **Loss = cosine distance + 0,5 KL divergence**

Важный гиперпараметр



Структура слоёв вариационного автоэнкодера

[<https://avandekleut.github.io/vae/>]

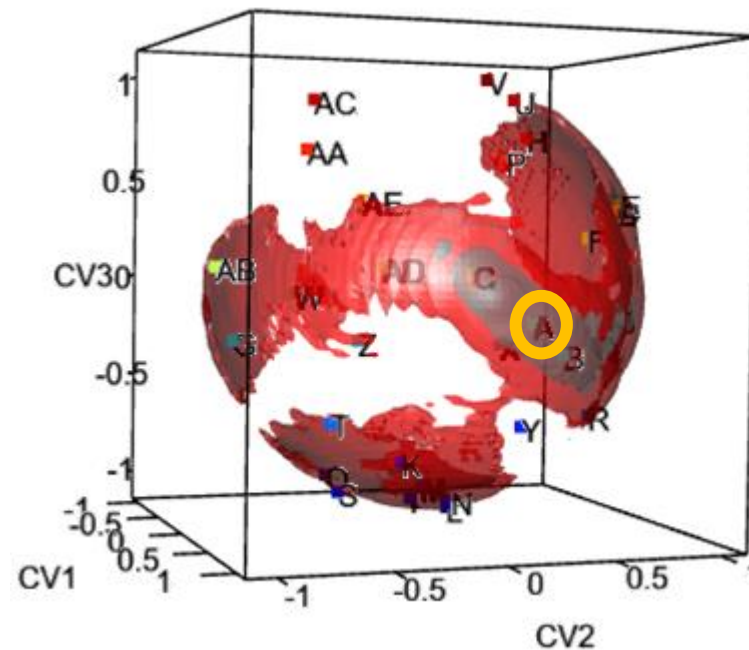


Отображение датасета структур NbprA на трёхмерное латентное пространство с помощью гиперсферического VAE

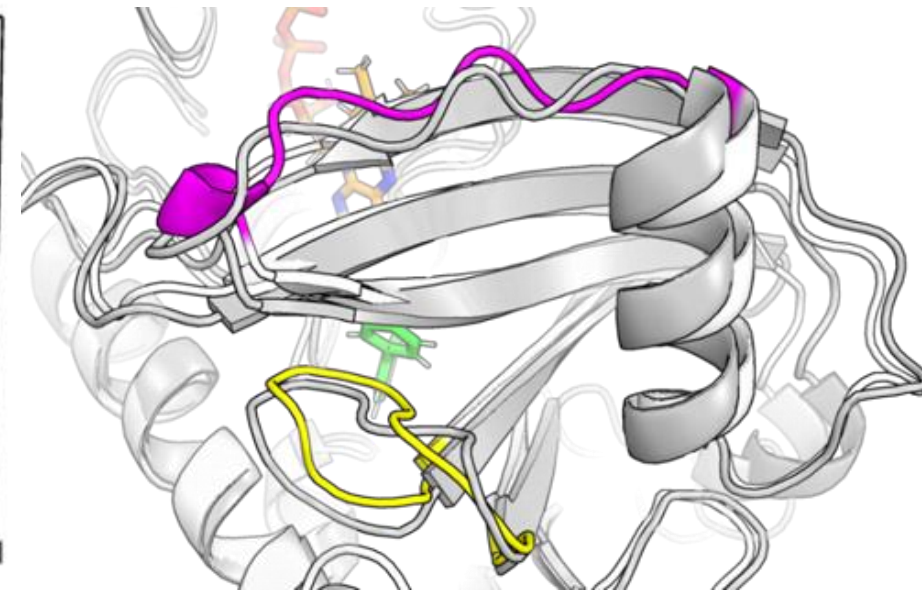


Результаты

- Метадинамика с гиперсферическим вариационным автоэнкодером показала минимум потенциальной энергии, близкий к структурам, предсказанным AlphaFold
- Разница со следующим классом минимумов (из молекулярной динамики) - порядка 15 ккал/моль



Изоповерхность потенциальной энергии метадинамики с гиперсферическим ВАЭ на уровнях -140 (голубой) и -110 кДж/моль (красный).

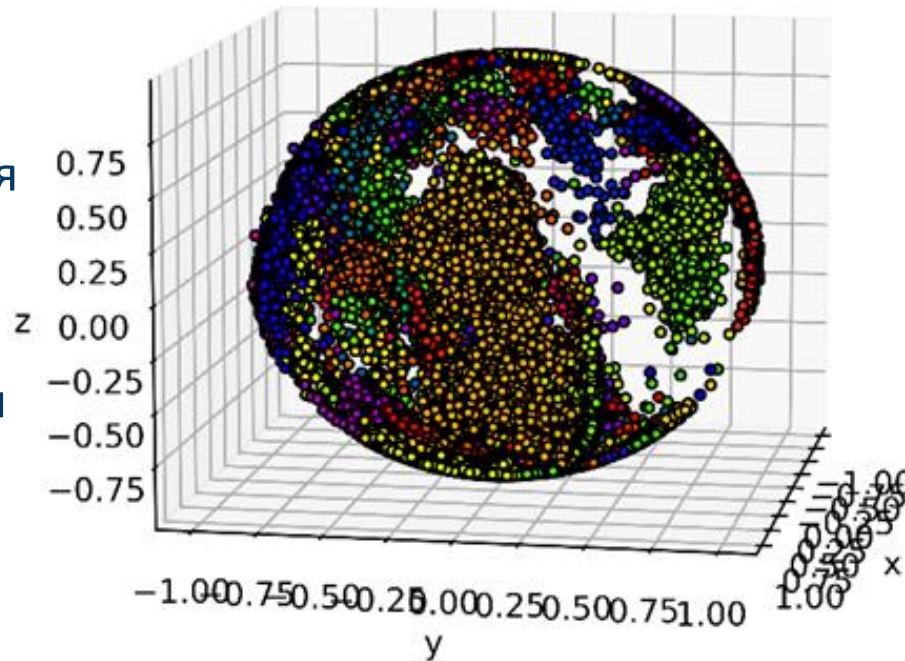


Серый - оптимальная структура (минимум "A")
Цветной - предсказание AlphaFold

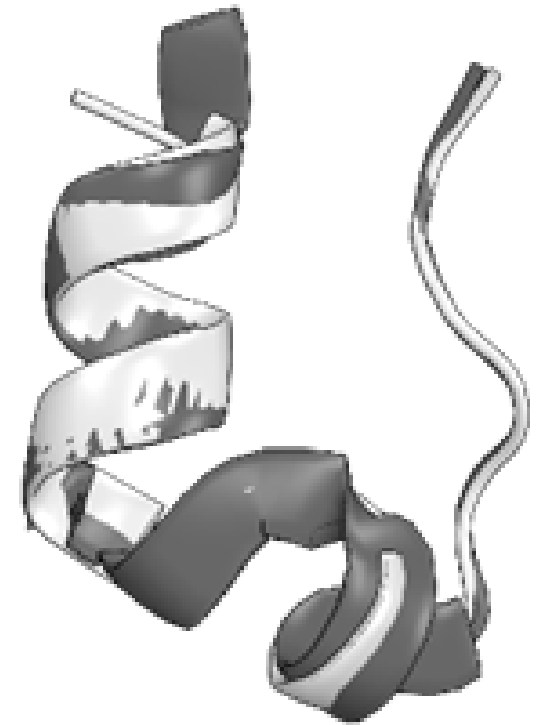


Валидация

- Белок Trp-Cage - распространённая модель для молекулярно-динамического фолдинга
- 20 остатков, для каждого углы ϕ и ψ
⇒ 38 входов VAE
- Минимум на поверхности потенциальной энергии соответствует структуре, сходной с литературными данными и структурами ЯМР из банка данных PDB



Отображение датасета Trp-Cage с помощью гиперсферического VAE



Белый - оптимальная структура из метадинамики (минимум "А")
Серый - структура ЯМР из PDB 1L2Y