

Лекция 2. Межмолекулярные взаимодействия белок-лиганд.

Курс: Машинное обучение в структурной биологии

Головин А.В.¹

¹МГУ им М.В. Ломоносова, Факультет Биоинженерии и Биоинформатики

Москва, 2021

» Взаимодействия

- * Межмолекулярные взаимодействия формируют материю
- * Взаимодействия в живом ключевой процесс
- * Взаимодействия основаны на слабых силах
- * Функция многих белков включает взаимодействие с малыми молекулами



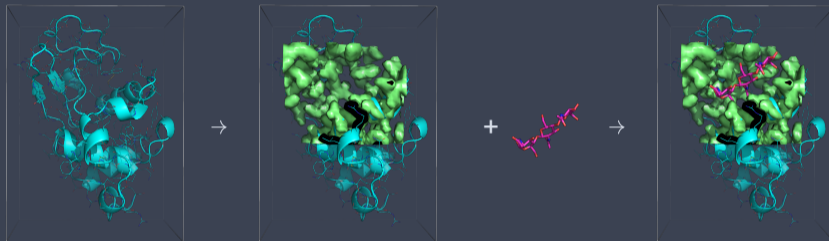
» Белок-лиганд

- * Поверхность белка значительно больше поверхности малых молекул
- * Поверхность белка может меняться с его конформационной динамикой
- * Взаимодействие с лигандом зачастую основано на комплементарности поверхностей



» Докинг белок-лиганд

Метод поиска способа связывания лиганда с белком



В результате мы можем узнать положение лиганда в комплексе с белком и оценить константу связывания

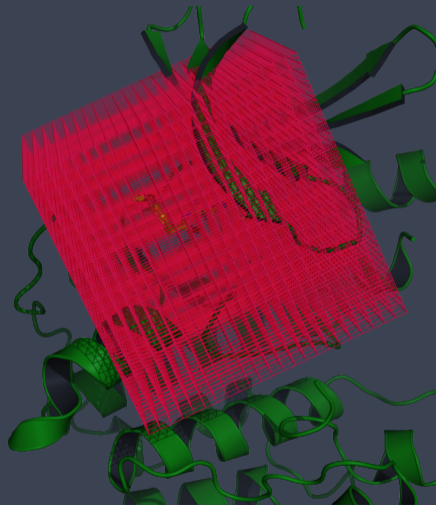


» Положение в сайте связывания

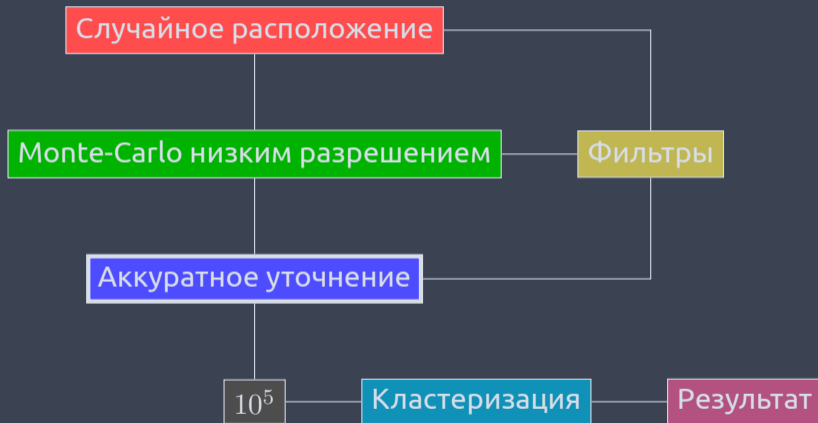
- * Сайт связывания — место связывания лиганда
- * Геометрия связывания — место связывания, ориентация и конформация лиганда



» Представление места связывания



» Поиск позиции и конформации лиганда

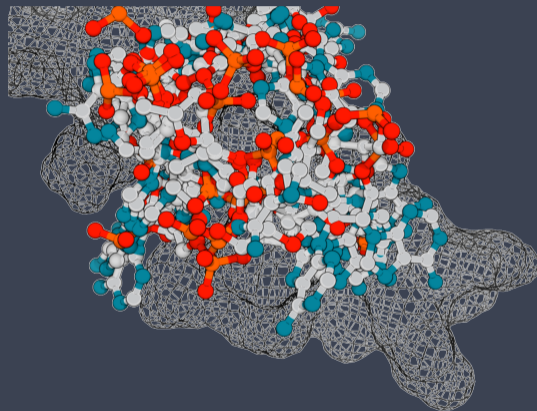
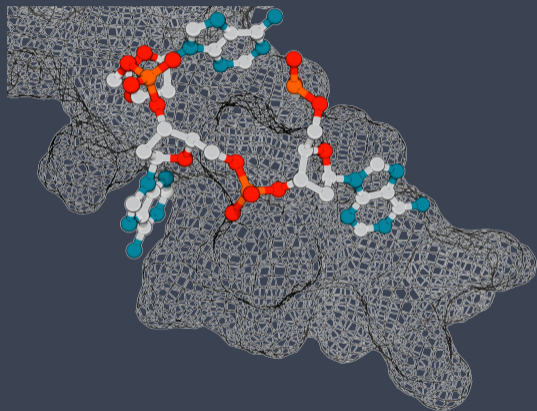


» Генетические алгоритмы

- * Задаём оптимизируемый объект как “ген(ы)”
- * Задаём функцию, которая вычисляет вероятность выживания в зависимости от имеющихся генов
- * Позволяем особям давать потомство, позволяем мутации, делеции, вставки, кроссинговеры, в зависимости от задачи
- * Вводим смертность особей



» Представление места связывания



» Использование докинга

Основные цели докинга:

- * Виртуальный поиск лигандов
- * Определение геометрии связывания лиганда

Если мы знаем, как связывается лиганд, то:

- * Мы можем узнать, какие части важны для связывания
- * Можно предложить изменения для улучшения константы связывания
- * Можем избежать ошибок



» Два основных компонента программ для докинга

- * Алгоритм поиска
 - * Установление места связывания
 - * Установление геометрии связывания
- * Алгоритм расчёта константы связывания областей с низкой энергией.



» Реализация

Сегодня существует много программ для докинга

- * AutoDock, DOCK, e-Hits, FlexX, FRED, Glide, GOLD, LigandFit, QXP, Surflex-Dock...и т.д.
- * разные алгоритмы оценки аффинности и разные алгоритмы поиска
- * Важно не путать лиганд-белок докинг и белок-белок докинг



» Практические аспекты

- * Часто PDB-структура содержит молекулы воды, почти всегда их надо убрать.
- * Надо добавлять протоны к структуре; His?
- * Часто в PDB неточно определена ориентация некоторых групп, что сказывается на паттерне водородных связей.
- * Протонирование лиганда и его таутомерные формы.



» Rigid|Flexible докинг

- * Rigid: лиганд не имеет внутренних степеней свободы, т.е. вращение вокруг связей запрещено.
- * Flexible: предполагает учёт вращения вокруг связей лиганда.
- * Часто белок рассматривается как жёсткое тело



» Как можно учесть подвижность рецептора?

- * Autodock Vina позволяет задать остатки рецептора, для которых разрешено вращение. Приводит к пересчёту сеток и сильно увеличивает время счёта
- * Длинная молекулярная динамика и нахождение самых часто встречающихся конформаций



» Фрагментарное построение лиганда

- * Сканирование по базам данных удобно, так как позволяет тут же проверить молекулу.
- * Сканирование не предполагает создание лиганда *de novo*.
- * Суть фрагментарного построения лиганда состоит в поиске мест в активном центре белка, где связываются небольшие фрагменты молекул
- * Соединяя фрагменты связями при сохранении места связывания, добиваются высоких констант.



» Фрагментарное построение лиганда



Leach, 1996



» Фрагментарное построение лиганда, реализация

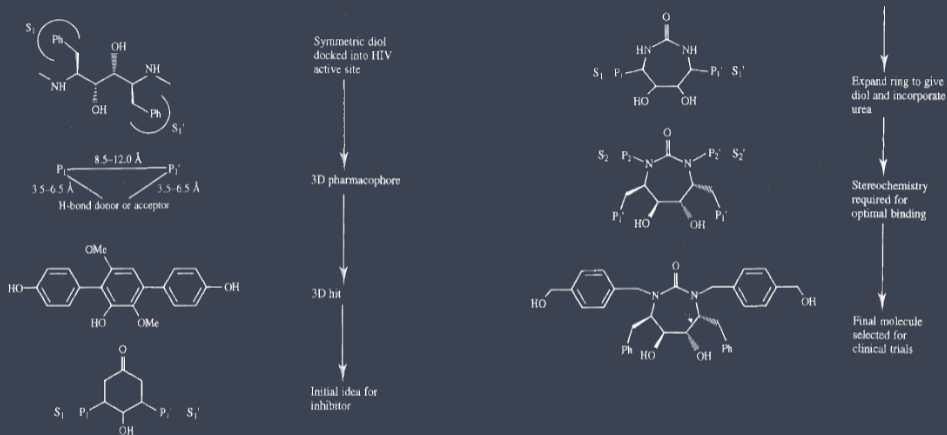
- * GRID аналог докинга.
- * MCSS сайт наполняется фрагментами и с помощью EM вычленяется место где фрагмент наиболее предпочтителен. Взаимодействия между фрагментами не учитываются.
- * LUDI использует информацию из банка PDB для задания фрагментов образующих водородные связи и т.д.
- * Необязательно всё моделировать, можно использовать PCA и ЯМР для определения места связывания фрагмента.



» Реализация ковалентного связывания фрагментов

- * Если у Вас есть два и более фрагмента, то можно искать способ их соединения по базам данных.
- * Реализовано в CAVEAT.
- * Можно строить автоматически строить скелеты. Главный критерий это сохранение взаимного положения фрагментов.
- * Переход от скелета к молекуле сложен, так как надо реализовать возможность синтеза молекулы.





» Основные направления применения ML

- * Перевзвешивание (Rescoring)
- * Подготовка библиотеки из химического разнообразия
- * Поиск сайтов связывания, выявление правил взаимодействия
- * Генерация новых соединений с высокой аффинностью



» Перевзвешивание

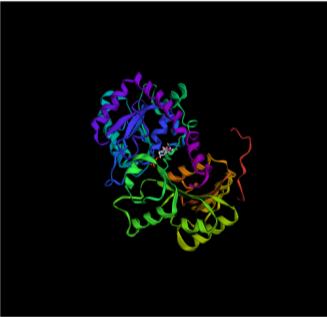
- * Докинг достаточно производительный метод, можно получить много данных о положении лигандов в белке
- * Классические оценочные функции это парные взаимодействия
- * Результат геометрии связывания лиганда можно оценить с помощью оценочной функции из ML



» Данные, PDBbind

PDBbind Current version: 2020
Total entries: 23,496

HOME BROWSE DATA LIGAND SEQUENCE DOWNLOAD APPLICATION CASF



Entry Information

PDB ID	3oka
Complex Type	Protein-Ligand
PDBbind Subset	general set
Protein Name	GDP-mannose-dependent alpha-(1-6)-phosphatidylinositol monomannoside mannosyltransferase:
Ligand Name	GDD
EC Number	E.C. 2.4.1.57
Resolution	2.2(Å)
Affinity (Kd/Ki/IC50)	Kd=19uM
Release Year	2010
Protein/NA Sequence	Check fasta file
Primary Reference	(2010) J. Biol. Chem. Vol. 285: pp. 37741-37752

Ligand Properties

Formula	C ₁₀ H ₁₈ N ₅ O ₁₁ P ₂
Molecular Weight	446.224

Display Options: Structure:

19,443 Записей

» Представление

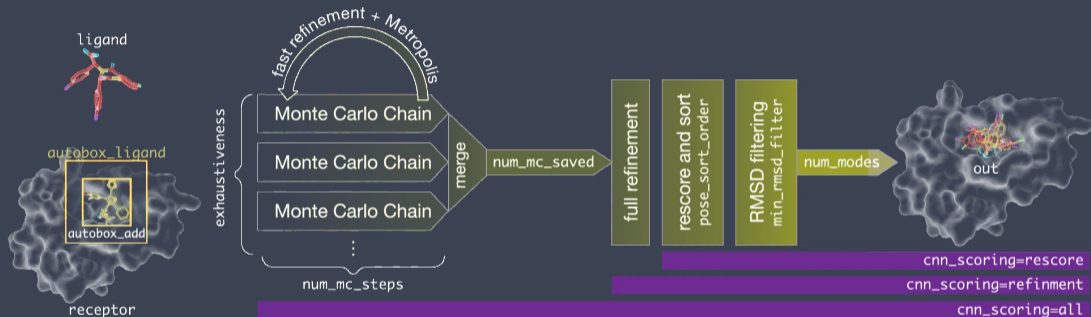
- * CNN исходно оптимизированы на изображения
- * Сайт связывания представляется как 3D сетка
- * Атомам присваиваются типы, 30 у лигандов, 16 у белков
- * Атомы представляются как распределение плотности

$$A(d, r) = \begin{cases} e^{-\frac{2d^2}{r^2}} & 0 \leq d < r \\ \frac{4}{e^2 r^2} d^2 - \frac{12}{e^2 r} d + \frac{9}{e^2} & r \leq d < 1.5r \\ 0 & d \geq 1.5r \end{cases}$$

10.1021/acs.jcim.6b00740



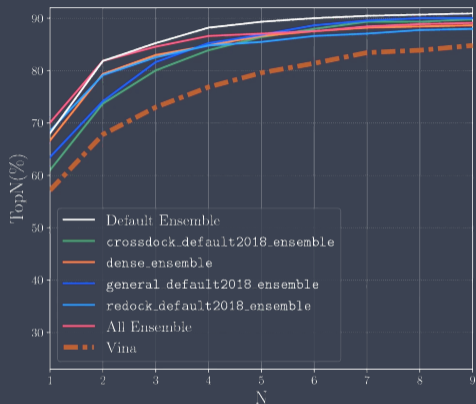
» GNINA



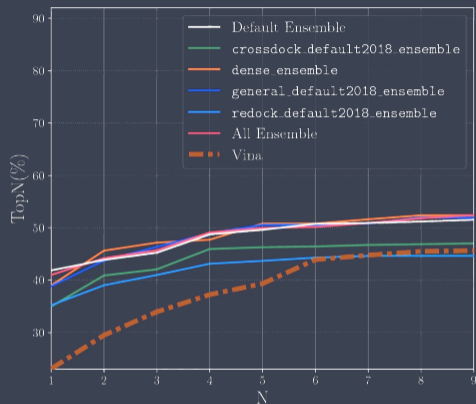
10.1186/s13321-021-00522-2



» GNINA



(a) Redocking Ensembles



(b) Cross-docking Ensembles

TopN is the percentage of targets ranked above or at N with a RMSD less than 2 Å ,
10.1186/s13321-021-00522-2

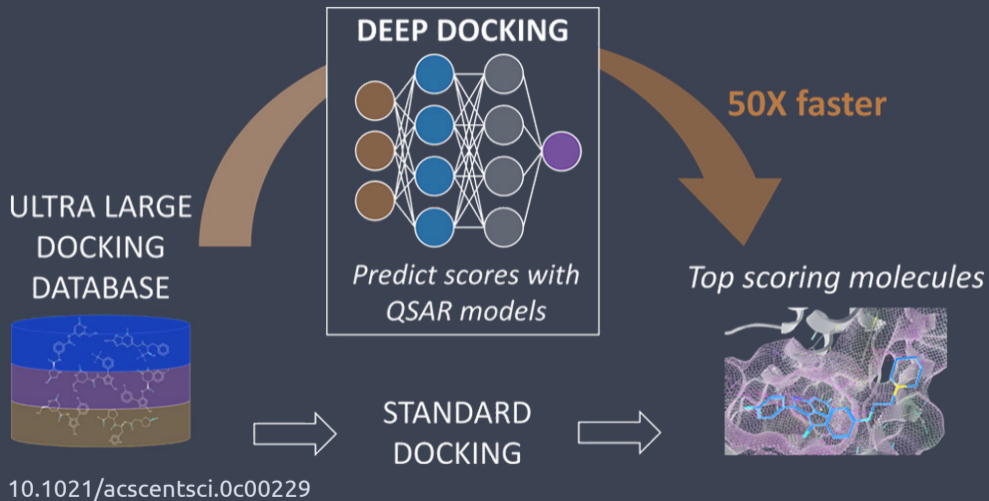


» Профилирование библиотек

- * Химическое разнообразие сравнимо с 10^{23}
- * Даже первичная генерация молекул требует гигантских ресурсов
- * Профилирование разнообразия под конкретную задачу



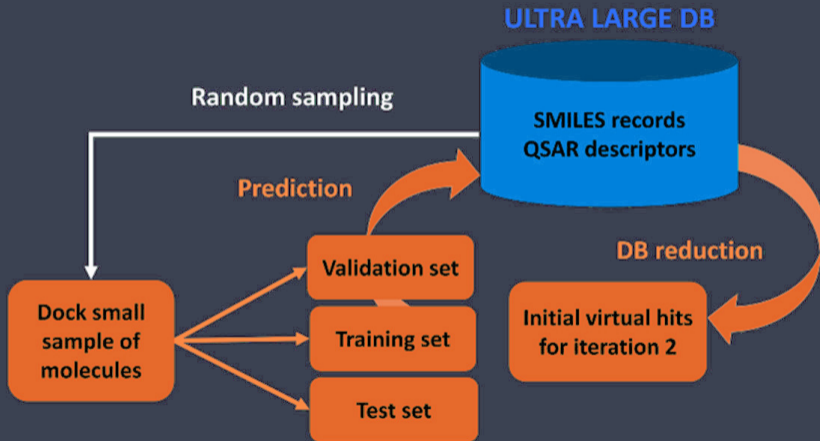
» Deep Docking, шаг 1



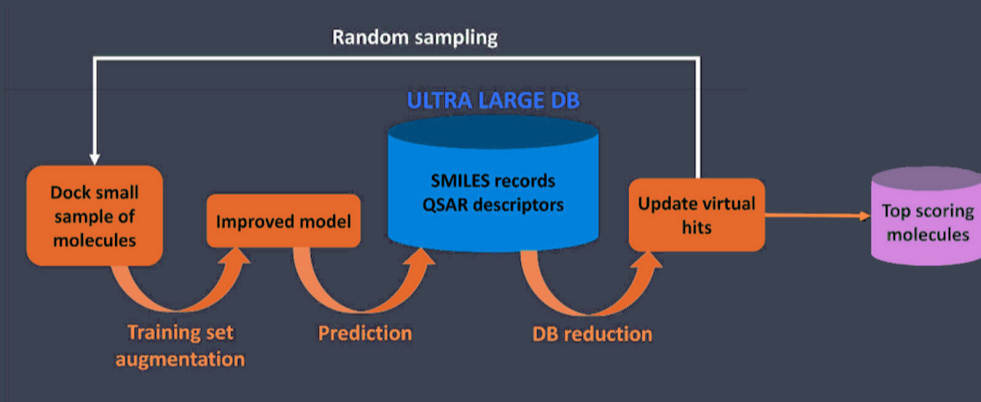
10.1021/acscentsci.0c00229



» Deep Docking, шаг 1



» Deep Docking, шаг 2



» Данные, представление, архитектура

- * Отбор из базы данных, $10^3 - 10^6$
- * Используются молекулярные QSAR дескрипторы, точнее Morgan fingerprints
- * "feed-forward" DNN сети в библиотеке Keras



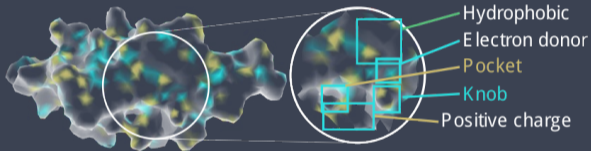
» Выявление сайтов связывания

- * Поверхность белка определяет взаимодействие
- * В поверхности важна не только "геодезия" но и типы атомов
- * Составление "fingerprints"
- * Обучение, выявление потенциальных сайтов и поиск подходящих отпечатков



» MASIF, общая идея

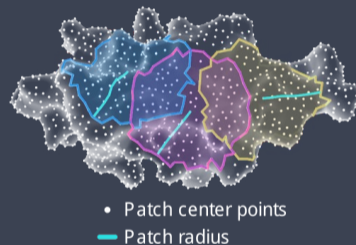
Protein molecular surface



Interaction fingerprint



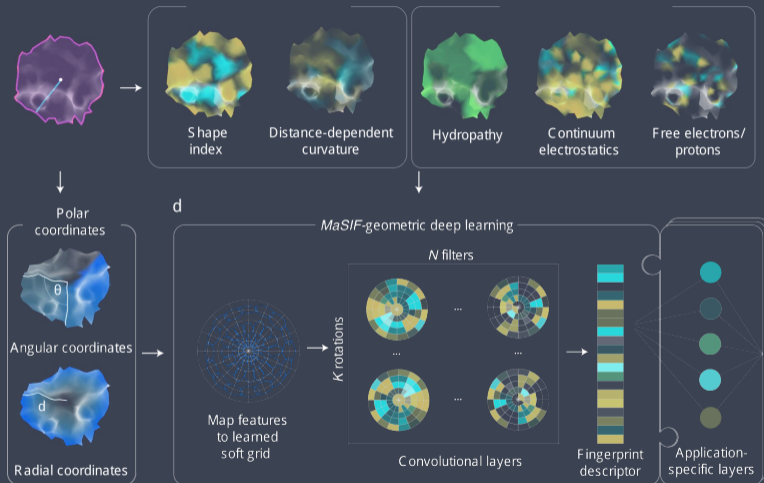
Approach, systematic extraction of patches



10.1038/s41592-019-0666-6



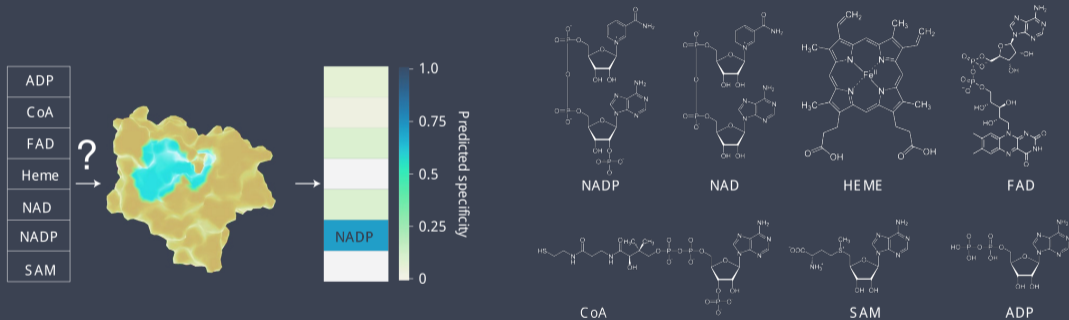
» MASIF, реализация



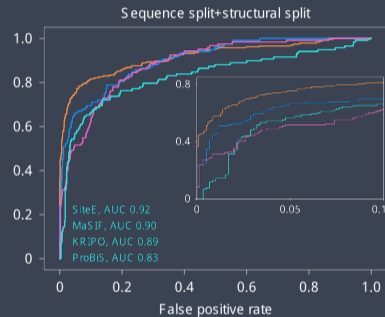
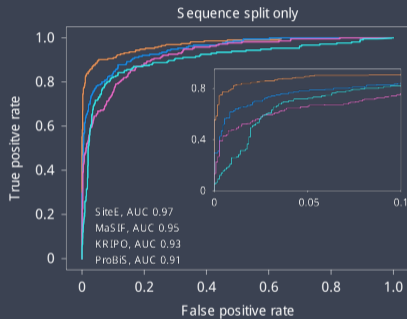
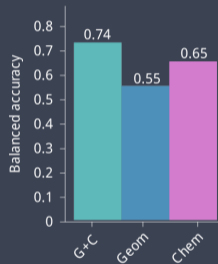
Данные: PDBbind, SAbDab



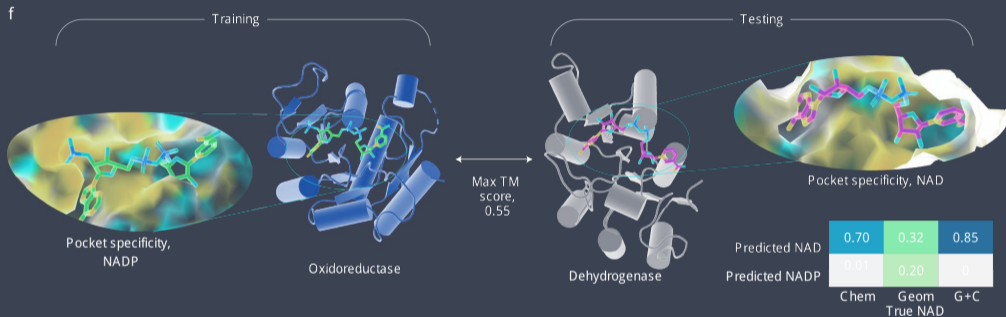
» Результат



» Аккуратность



» Пример



<https://github.com/lpdi-epfl/masif>

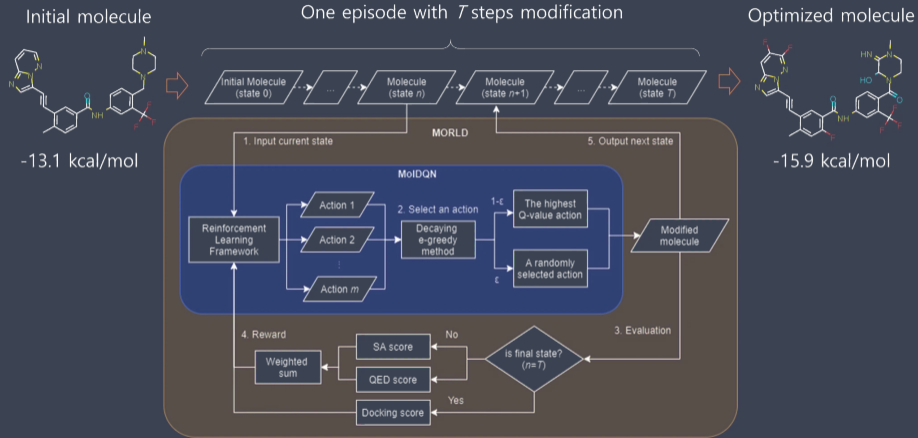


» Оптимизация соединений

- * Наличие соединения в сайте связывания
- * Обучение происходит в процессе работы
- * Вознаграждение и штрафы приходят из докинга



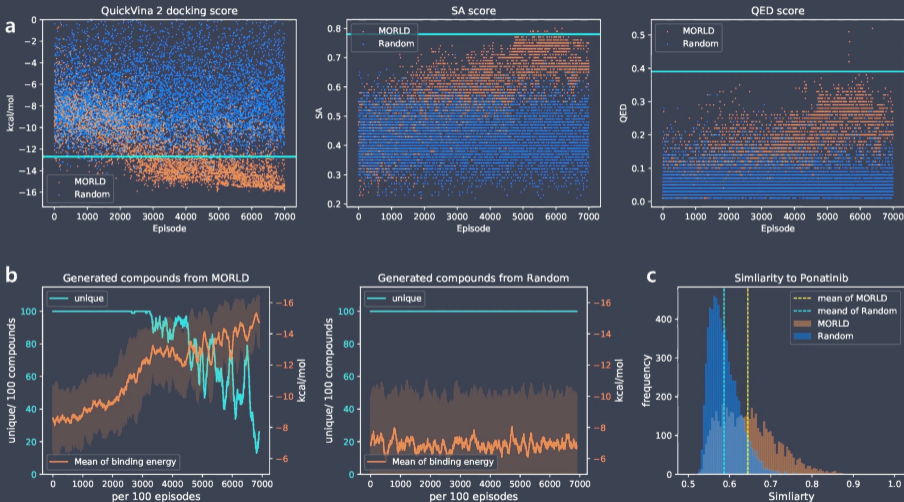
» MORD



10.1038/s41598-020-78537-2



» MORLD, результаты



» GENTRL

