

Лекция 6. Дизайн карманов связывания

Курс: Методы машинного обучения в дизайне белков

Головин А.В. ¹

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

Москва, 2023

» Способы дизайна карманов связывания

- * De novo
- * “re-purposing”
 - * Направленная эволюция
 - * Вычислительное моделирование



» Направленная эволюция

- * Направленная эволюция - это общий термин для описания набора методов молекулярной биологии, которые, представляют собой искусственный процесс мутирования и отбора.
- * Эти методы подразумевают собой случайное (или направленное) внесение замен на генетическом уровне с последующим отбором белков, обладающих нужными функциями.

Источник: <http://dx.doi.org/10.4067/S0716-97602013000400011>



» Вычислительное моделирование

Процесс дизайна белкового байндера посредством перепрофилирования состоит из нескольких этапов:

- * составление и курирование библиотеки белковых скаффолдов
- * помещение лиганда в карман связывания
- * дизайн последовательности
- * докинг лигандов + MD
- * экспериментальная валидация результатов

Приведенная последовательность действий является только примером организации вычислительного эксперимента, нежели строгим протоколом



» Составление и курирование библиотеки белковых скаффолдов

RCSB **PDB**
PROTEIN DATA BANK

Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 1.90 Å

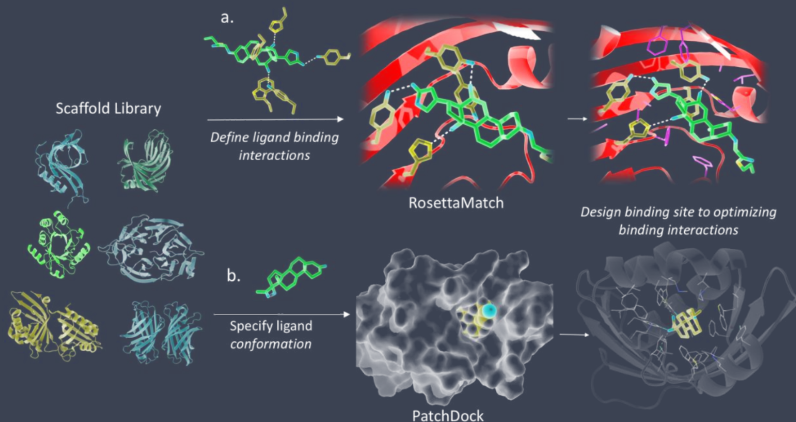
R-Value Free: 0.248

R-Value Work: 0.185

R-Value Observed: 0.188



» Помещение лиганда в карман связывания. RosettaMatch



Источник: <http://hdl.handle.net/1773/39951>



» Дизайн последовательности, оптимизация ротамеров

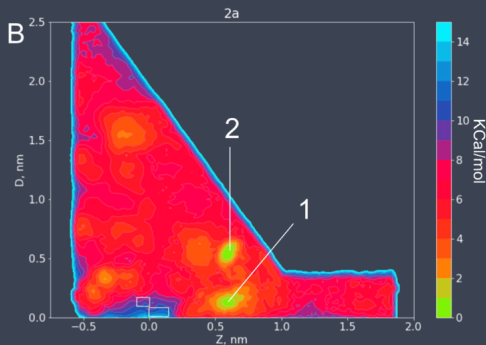
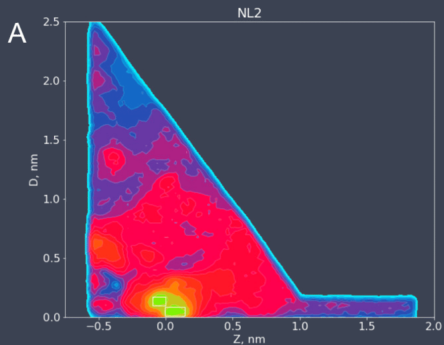
Используется алгоритм стохастического моделирования отжига (метод Моенте-Карло, имплементированный в Rosetta). Дизайн проводится при фиксированном белковом остоле. Принцип работы в общем виде:

- * имеем начальный вид аминокислотного состава последовательности и ротамеров в окрестности кармана связывания
- * расчет энергии
- * производится замена случайно выбранного остатка (точечная мутация)
- * расчет энергии, если $E_{new} < E_{old}$ то новый ротамер принимается, иначе ротамер принимается с некоторой вероятностью (к примеру, по критерию Метрополиса):
- * повторяем



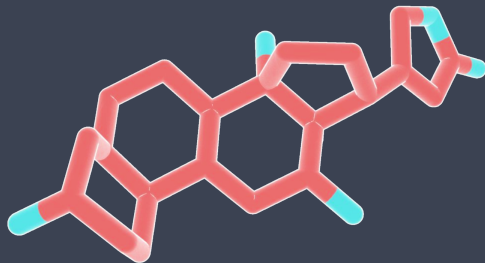
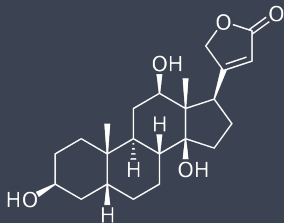
» Валидация

- * Докинг можно использовать только с критическим анализом, из-за неподвижного остова белка
- * Методы расширенной выборки в молекулярной динамике: метадинамика, awh ...

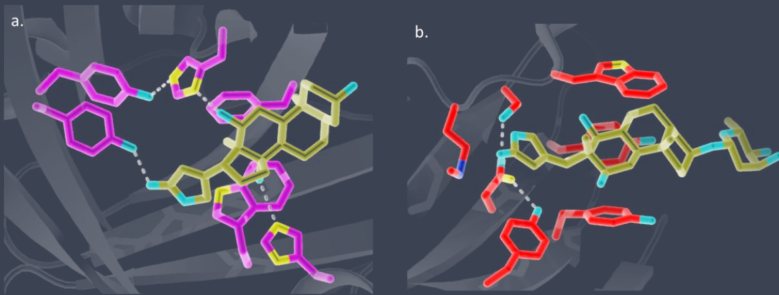


» Примеры дизайнов. Дизайн байндера дигоксигенина (DIG)

Дигоксигенин (DIG) - агликон дигоксина, сердечного гликозида, который используется для лечения некоторых сердечно-сосудистых заболеваний



» Примеры дизайнов. Дизайн байндера дигоксигенина (DIG)



(а) Сайт связывания в DIG-связывающем липокаллине (K_d 30.2 нМ, PDB: 1LKE)

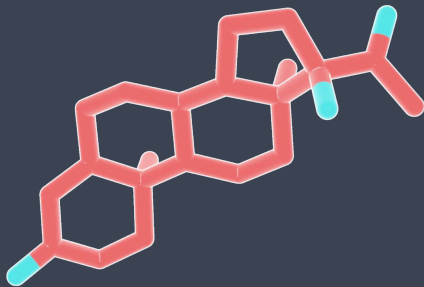
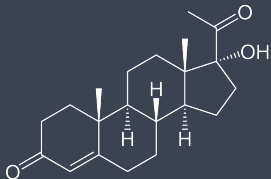
(б) Сайт связывания анти-DIG антитела (K_d 0.1 нМ, PDB: 1IGJ)

Источник: <http://hdl.handle.net/1773/39951>



» Дизайн белков, связывающих 17-ОПГ

17-ОПГ - 17-гидроксипрогестерон - гормон, производимый жёлтым телом яичников и корой надпочечников. Является биомаркером группы аутосомально-рецессивных заболеваний из группы "врожденная гиперплазия надпочечников"



» Дизайн белков, связывающих 17-ОПГ

Для создания рабочей библиотеки скаффолдов использовались NTF2-подобные белки из базы данных PDB

В связи с тем, что 17-ОПГ имеет небольшое количество полярных групп, авторы несколько экспериментировали с вычислительными подходами и первым делом определяли комплементарность молекулярных форм лиганд - белок (PathDock), а после искали водородные связи с боковыми цепями аминокислот



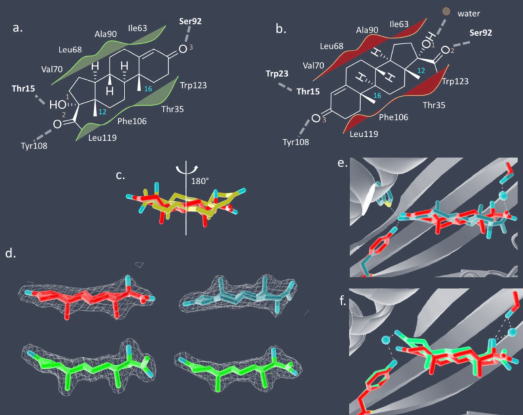
» Дизайн белков, связывающих 17-ОПГ

- * Внимание авторов привлек результат одного из дизайнов, белковой основой для которого послужил белок с неизвестной функцией (RV0760) из *Mycobacterium tuberculosis*
- * По результатам первой экспериментальной проверки (метод дрожжевого дисплея), этот дизайн не смог связать 17-ОПГ



» Дизайн белков, связывающих 17-ОПГ

(a) взаимодействия в модели дизайна
 (b) взаимодействия в полученной структуре
 (c) структура 17-ОПГ в модели дизайна показана фиолетовым, она сопоставлена со структурой 17-ОПГ из кристалла (цепь А, циан)
 (d) структуры 17-ОПГ из кристалла - цепь А (циан), цепи В и D (маджента), цепь С (розовый)
 (e, f) лиганда в контексте белка Неудачный дизайн байндера 17-ОПГ



» Дизайн de novo

В отличие от дизайна с перепрофилированием белковой функции, описанного ранее, de novo дизайн в качестве входных данных не требует существования predetermined остова (из природного белка)
Сложности подхода:

- * сайты связывания малых молекул в однодоменных белках зачастую требуют наличия достаточно глубоких полостей, что опосредует нестабильность фолда
- * для дизайна кармана связывания малых молекул требуется высокая точность определения нужной конформации остова и расположения боковых цепей
- * задача построения стабильного белкового скаффолда все еще остается нерешенной



» Пример протокола de novo дизайна белков

- * создание библиотеки белковых скаффолдов, собранных de novo из коротких пептидных фрагментов;
- * протокол лежит в открытом доступе помещению интересующей нас малой молекулы в заданный сайт связывания посредством алгоритма RIF (Rotamer-Interaction-Field)
- * итеративная оптимизация последовательности и минимизация остова (Rosetta)
- * оптимизация взаиморасположения лиганда и окружающих его боковых цепей аминокислот (увеличиваем количество взаимодействий)
- * экспериментальная проверка

Пример стратегии de novo дизайна белка-байндера, источник:
<http://hdl.handle.net/1773/39951>



» Пример алгоритма дизайна

“New computational protein design methods for de novo small molecule binding sites”

Источник: <https://doi.org/10.1371/journal.pcbi.1008178>

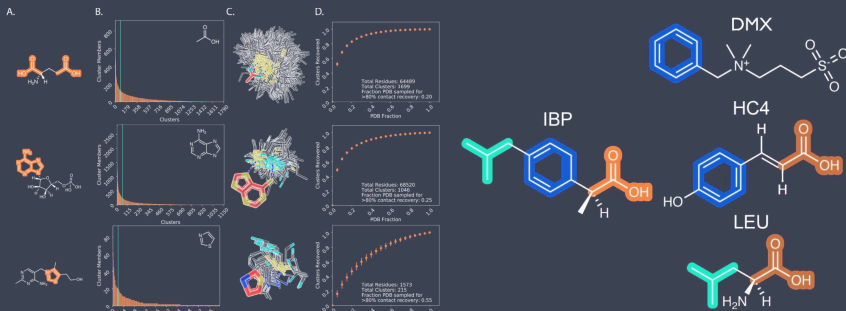
Авторы: James E. Lucas, Tanja Kortemme

Некоторые особенности метода:

- * декомпозиция лиганда на подструктуры/фрагменты, для которых в PDB можно найти большое количество примеров взаимодействий. “Фрагмент” определяем как субструктуру молекулы-лиганда, представляющую собой самостоятельную химическую группу и состоящую как минимум из трех атомов;
- * делается предположение, что найденные контакты фрагмент - белок формируют пул возможных контактов целого лиганда с белком;



» Пример алгоритма дизайна. Декомпозиция лиганда



» Детали реализации

- * К каждому конформеру лиганда-мишени был применен алгоритм Монте-Карло - производилась сборка дискретных боковых цепей аминокислот из ансамбля контактов в композитный сайт связывания
- * Сайт связывания лиганда инициализировался тремя случайными остатками из ансамбля контактов
- * Во время каждого шага алгоритма, боковая цепь случайного аминокислотного остатка в сайте связывания заменялась случайно выбранной боковой цепью из ансамбля контактов
- * Вычислялся скор “нового” сайта связывания (энергия), шаг алгоритма принимался на основании критерия Метрополиса; цель - минимизация функции энергии
- * Сайты связывания, характеризующиеся лучшими значениями энергии связывания релаксировались с использованием алгоритма FastRelax (Rosetta)



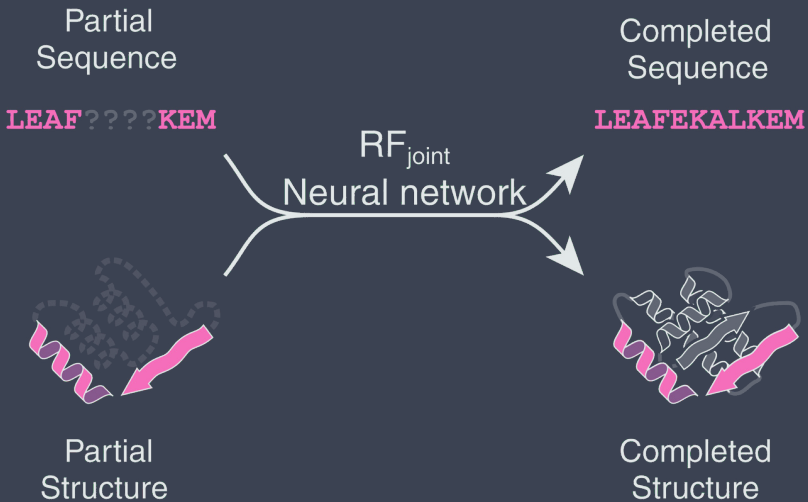
» Inpainting



- * Широкий спектр проблем проектирования структуры белков можно аналогичным образом сформулировать как проблемы восстановления недостающей информации

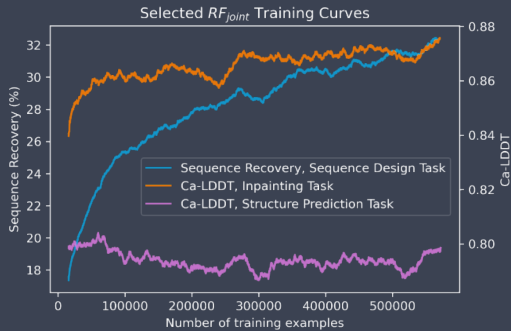
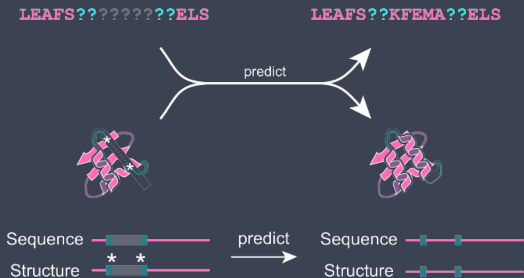


» Inpainting



» RFjoint

Joint Sequence & Structure recovery

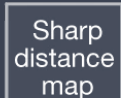
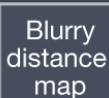


» "network hallucination"

Structure prediction

Amino acid
sequenceStructure-
prediction network3D
structure

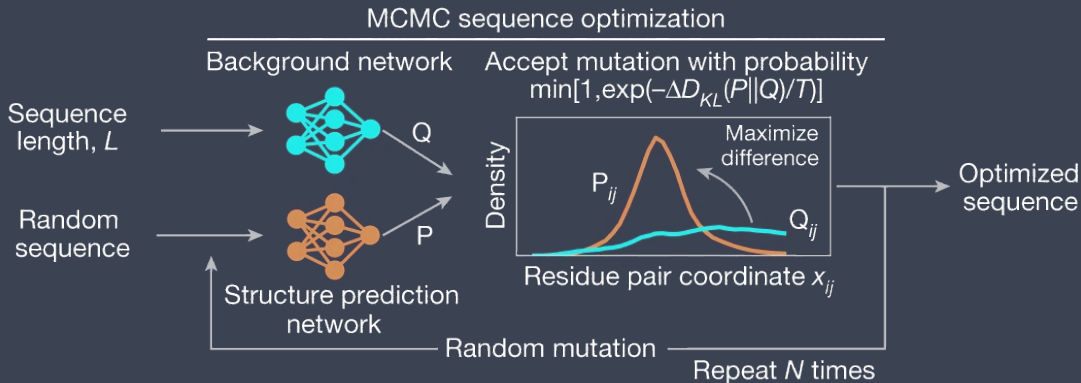
Protein design by network hallucination

Random
amino acid
sequence

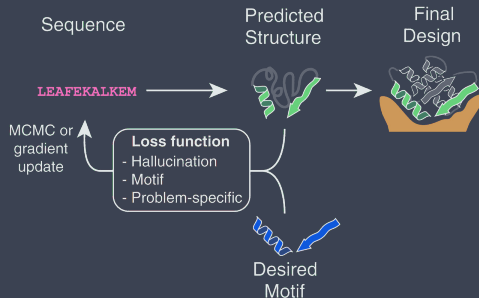
MCMC sequence optimization

3D
structure

» "network hallucination"



» Ограниченные галлюцинации

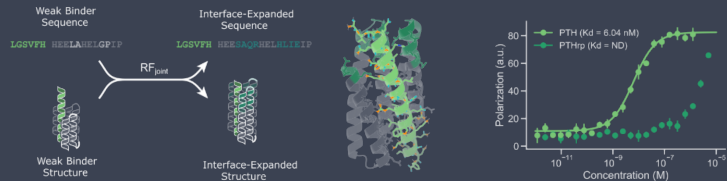


- * Сложная функцию потери, которая сочетает в себе часть из галлюцинаций с частью реконструкции мотива.
- * Подход с ограниченными галлюцинациями требует больших вычислительных ресурсов, поскольку для каждого шага градиентного спуска во время оптимизации последовательности требуется прямой и обратный проход через сеть.

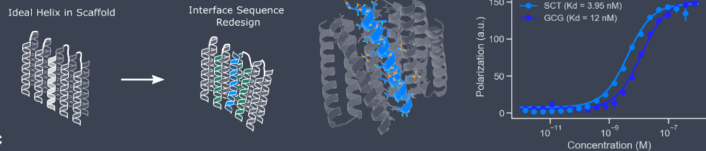


» Стратегии применения ML моделей для дизайна карманов

a



b

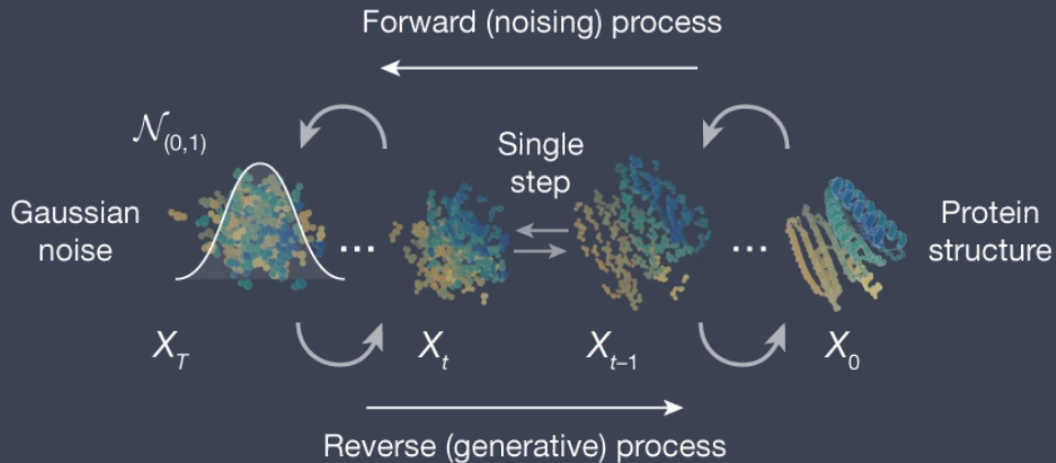


c



» Диффузия

Diffusion model



» Диффузия

RoseTTAFold

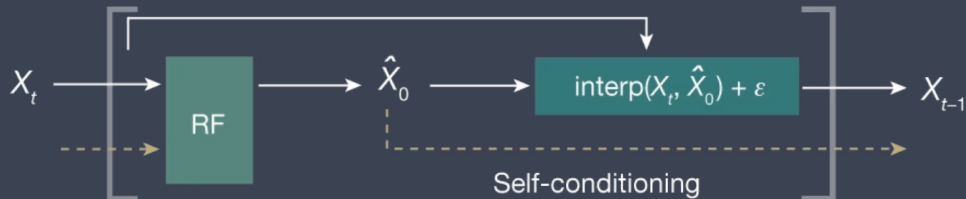


RFdiffusion



» Диффузия

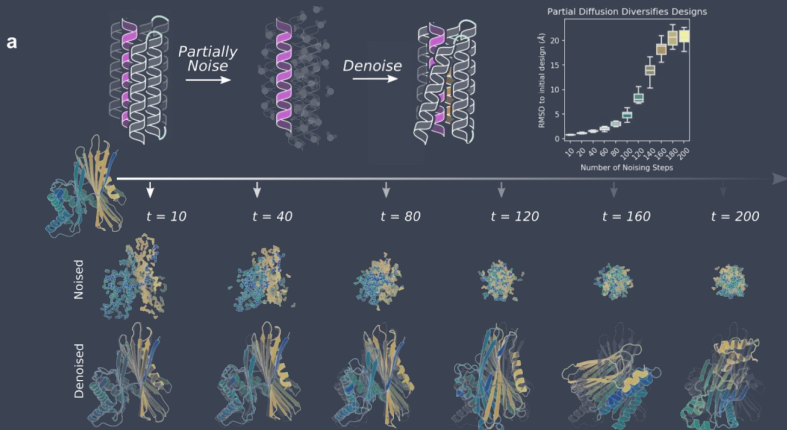
Single RFdiffusion step



На каждом шаге t модель берет $X(t+1)$, а затем прогнозирует обновленную структуру $X(0)$. Следующий ввод координат $X(t-1)$ в модель генерируется путем интерполяции с шумом в направлении $X(0)$.

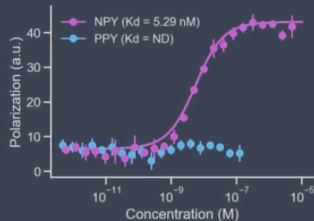
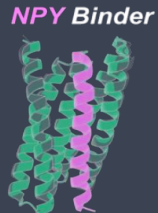
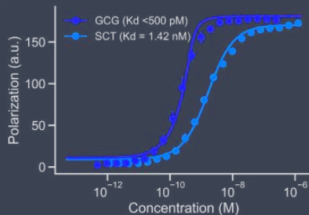
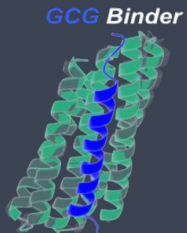


» Partial Noise



» Partial Noise, результаты

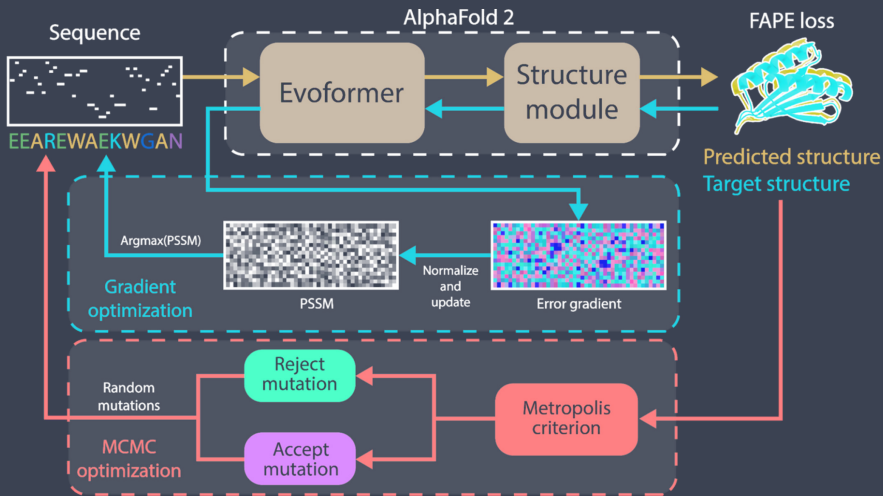
C



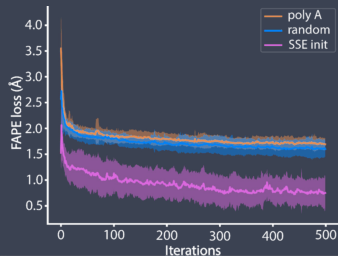
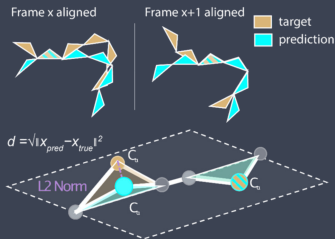
Высокая степень совпадения с AF2



» Дизайн по варианту укладки с AlphaFold



» Дизайн по варианту укладки с AlphaFold

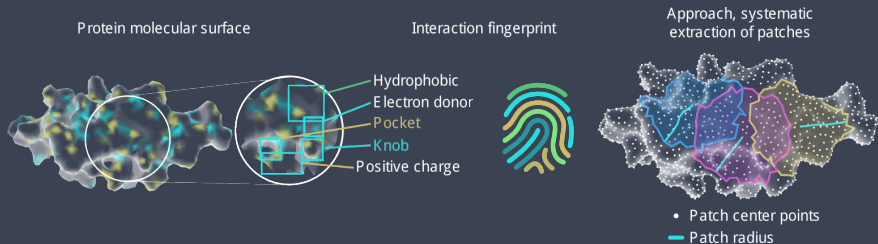


» Частичный дизайн с AlphaFold

- * Частичные галлюцинации
- * Дизайн при заданном ходе остова
- * "binder hallucination"
- * Custom loss



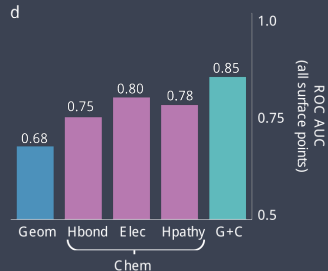
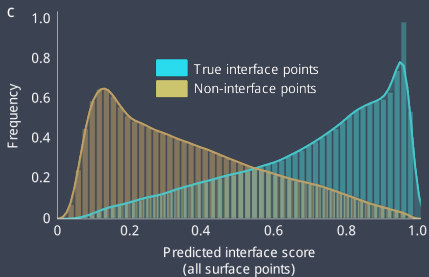
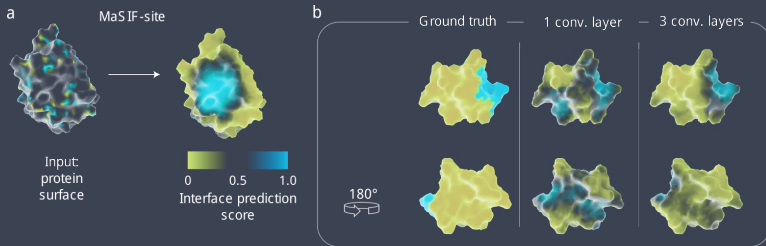
» MASIF



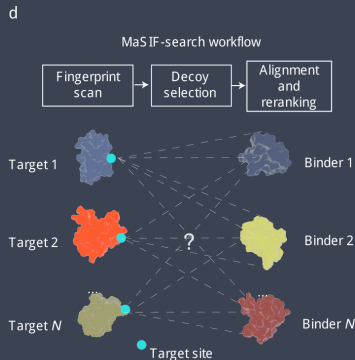
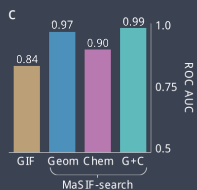
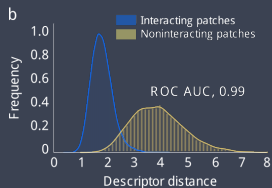
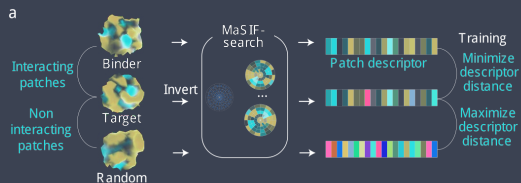
MaSIF-site: классификатор, на входе поверхность белка, на выходе прогнозируемая оценка для каждой вершины поверхности на вероятность участия в PPI



» MASIF, предсказание участков



» Ultrafast scanning with MASIF



MaSIF-search inverts the numerical features of one protein partner (multiplied by -1), with the exception of hydrophathy.



» Masif-seed

