

Дизайн карманов связывания

Дизайн белков, НТУ Сириус

Головин А.В.^{1 2} Беляева Ю.Д.²

¹ НТУ Сириус, ЦИИиИТ

² МГУ им. М.В.Ломоносова, Факультет Биоинженерии и
Биоинформатики

Сириус, 2021

Способы дизайна карманов связывания

Дизайн

“re-purposing”

de novo

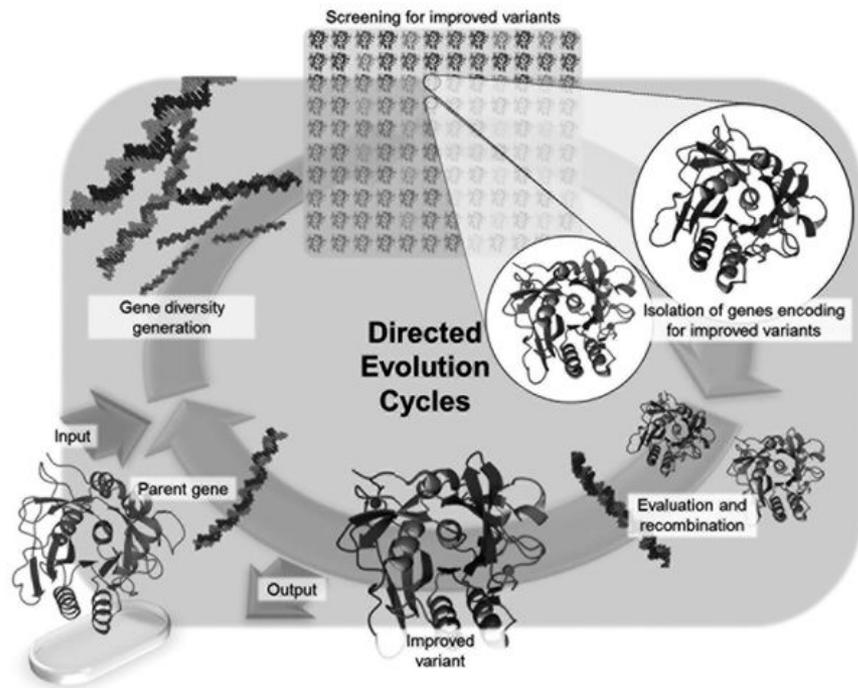
Направленная
эволюция

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.



Общая идея метода направленной эволюции

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

Вычислительное моделирование

Процесс дизайна белкового байндера посредством перепрофилирования состоит из нескольких **этапов**:

- 1) составление и курирование библиотеки белковых скаффолдов
- 2) помещение лиганда в карман связывания
- 3) дизайн последовательности
- 4) докинг лигандов + MD
- 5) 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

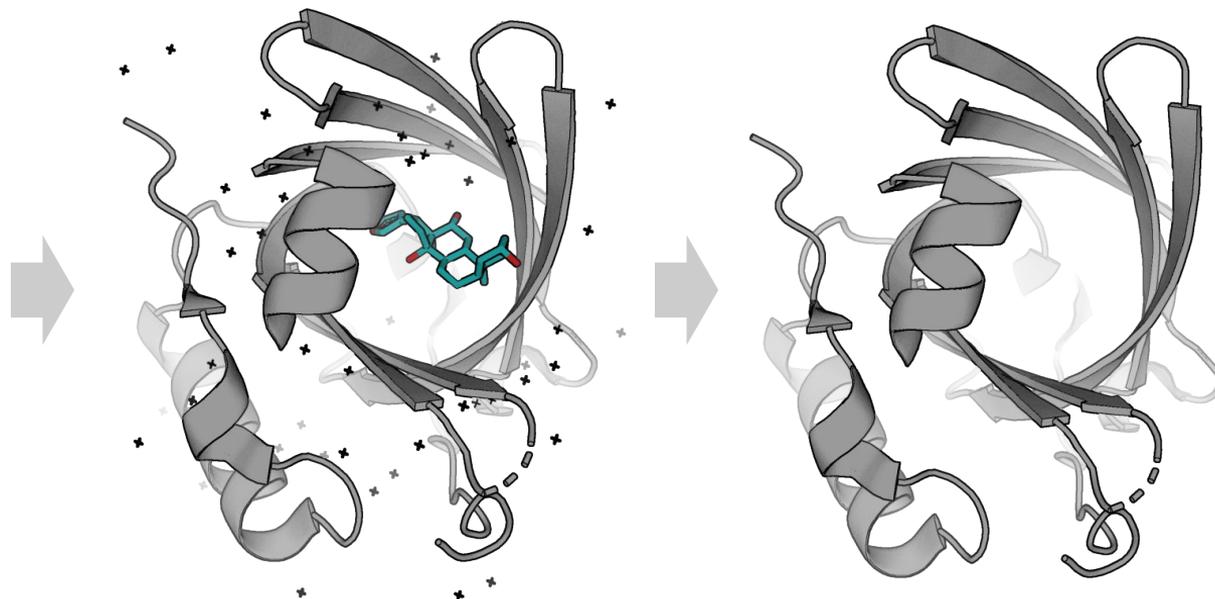
Method: X-RAY DIFFRACTION

Resolution: 1.90 Å

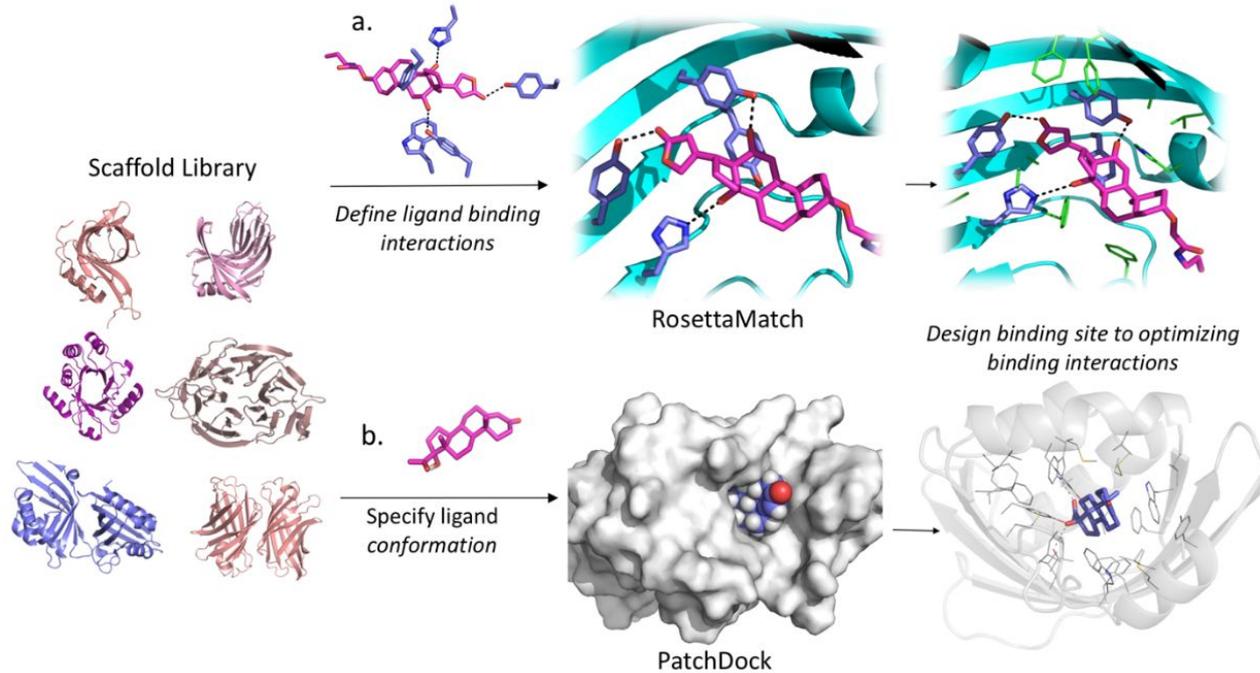
R-Value Free: 0.248

R-Value Work: 0.185

R-Value Observed: 0.188



Помещение лиганда в карман связывания. RosettaMatch



Вычислительное моделирование белковых байндеров
("re-purposing")

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

Дизайн последовательности, оптимизация ротамеров

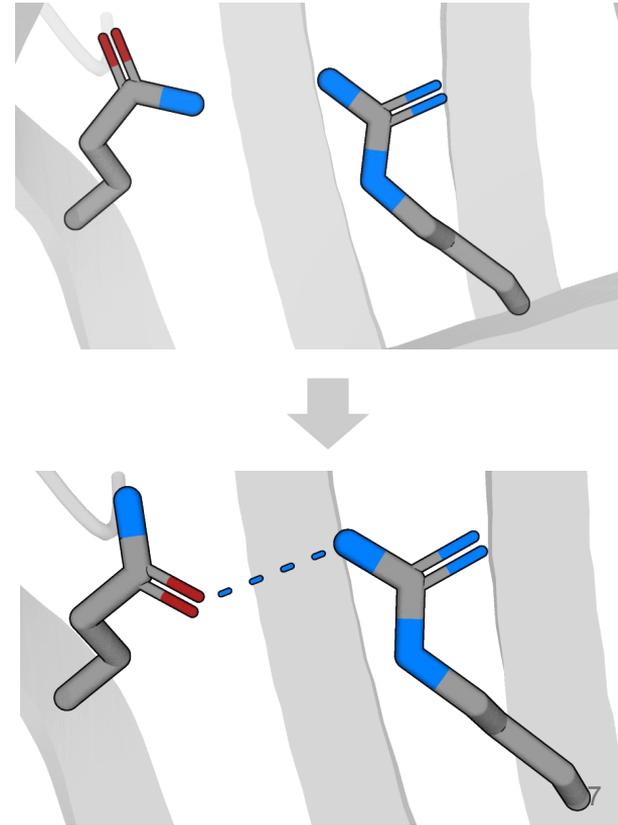
Используется алгоритм стохастического моделирования отжига (метод Моенте-Карло, имплементированный в Rosetta). Дизайн проводится при фиксированном белковом остове.

Принцип работы в общем виде:

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

$$p = \min(1; \exp(-\frac{\Delta E}{k_B T}))$$

- 5) повторяем



Вычислительное моделирование

Процесс дизайна белкового байндера посредством перепрофилирования состоит из нескольких **этапов**:

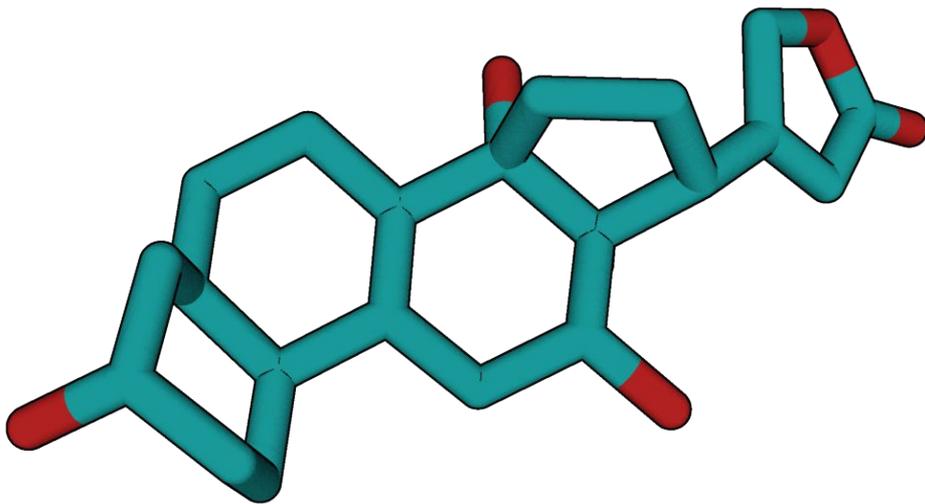
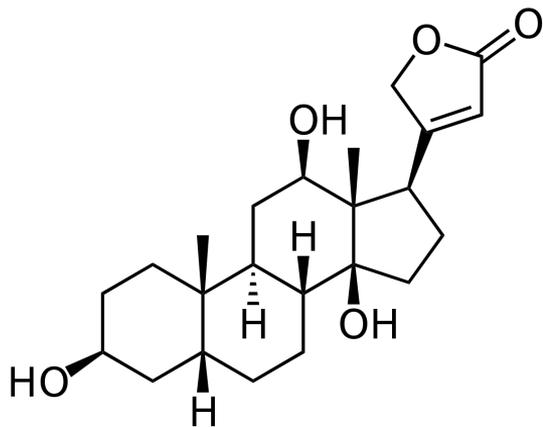
- 1) составление и курирование библиотеки белковых скаффолдов
- 2) помещение лиганда в карман связывания
- 3) дизайн последовательности
- 4) докинг лигандов + MD
- 5) экспериментальная валидация результатов

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

Примеры дизайнов.

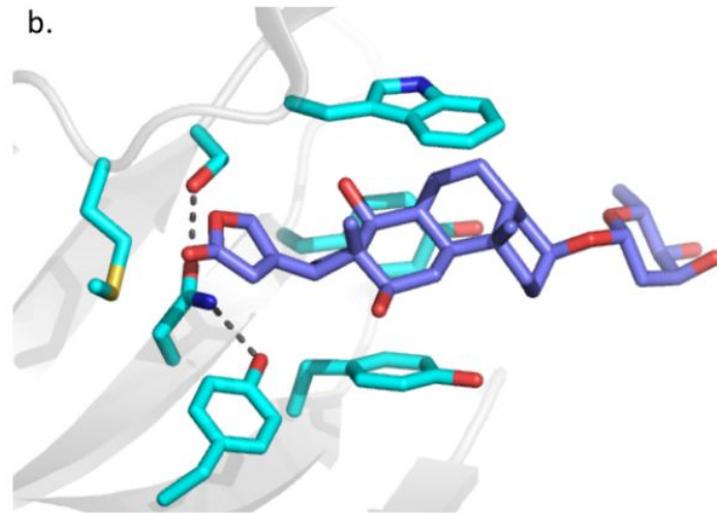
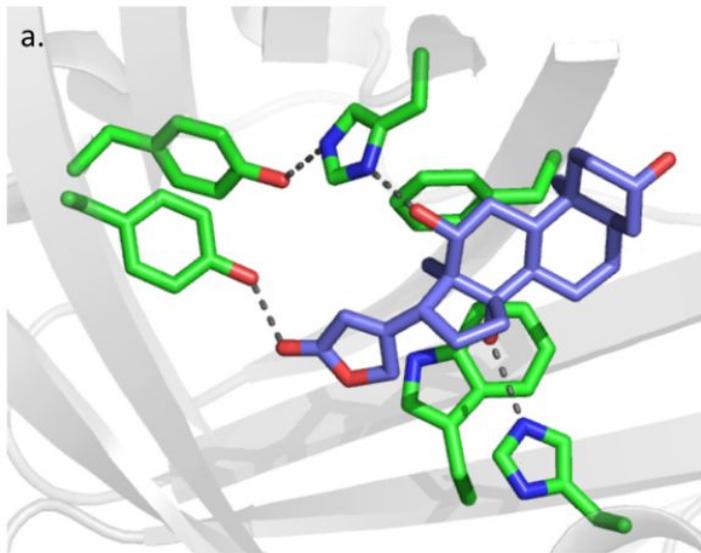
Дизайн байндера дигоксигенина (DIG)

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



Примеры дизайнов.

Дизайн байндера дигоксигенина (DIG)



(a) Сайт связывания в DIG-связывающем липокаллине ($K_d \sim 30.2$ нМ, PDB: 1LKE)

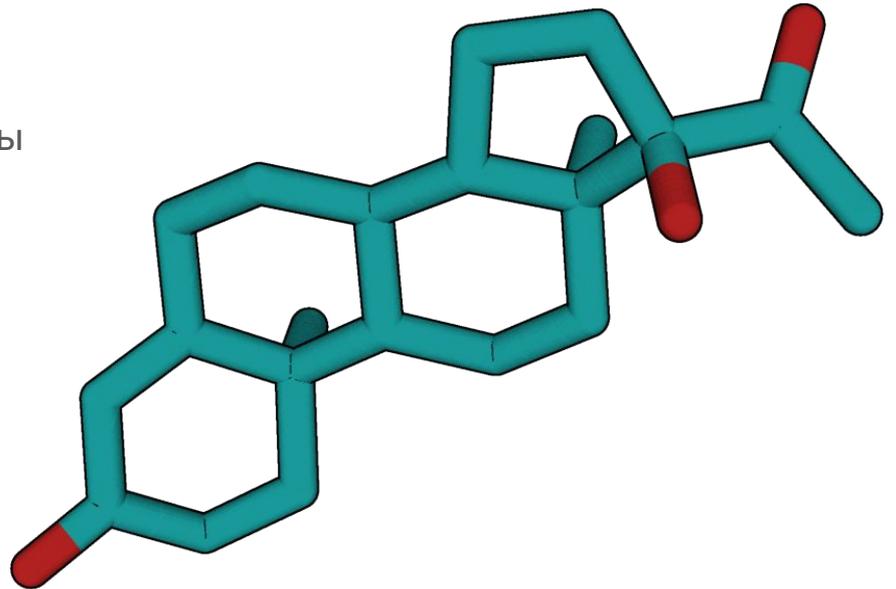
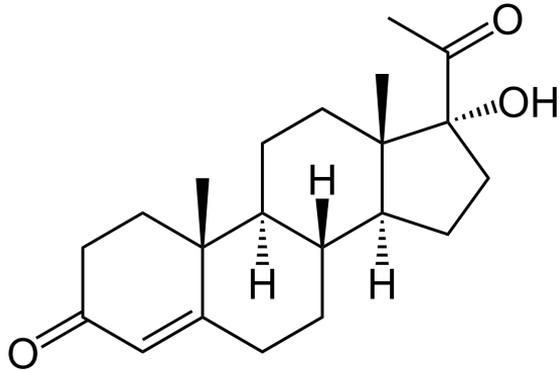
(b) Сайт связывания анти-DIG антитела ($K_d \sim 0.1$ нМ, PDB: 1IGJ)

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

Примеры дизайнов.

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”

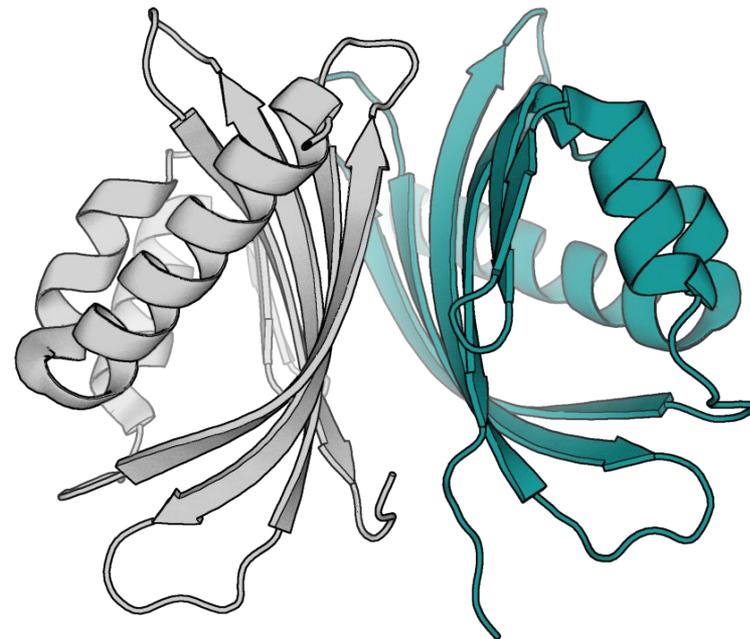


Примеры дизайнов.

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



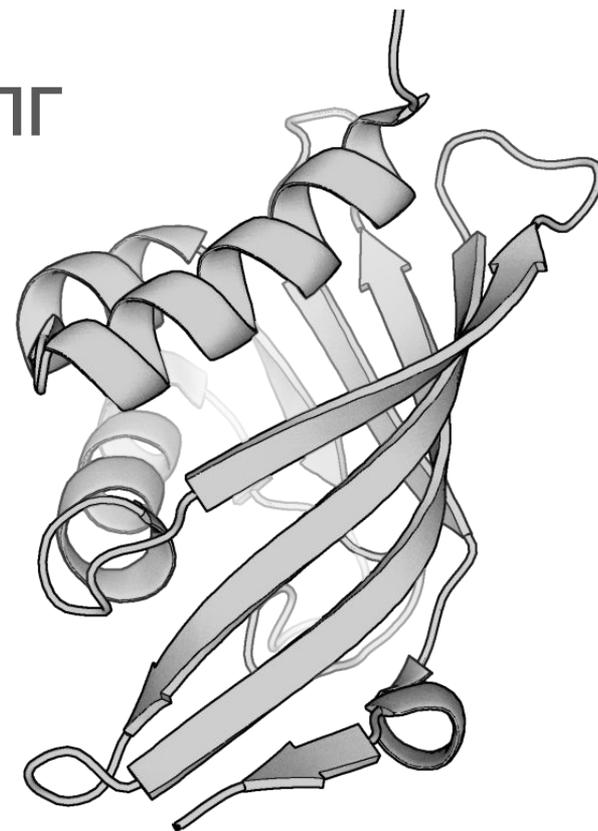
NTF2 белок
PDB ID: [1JB4](#)

Примеры дизайнов.

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*

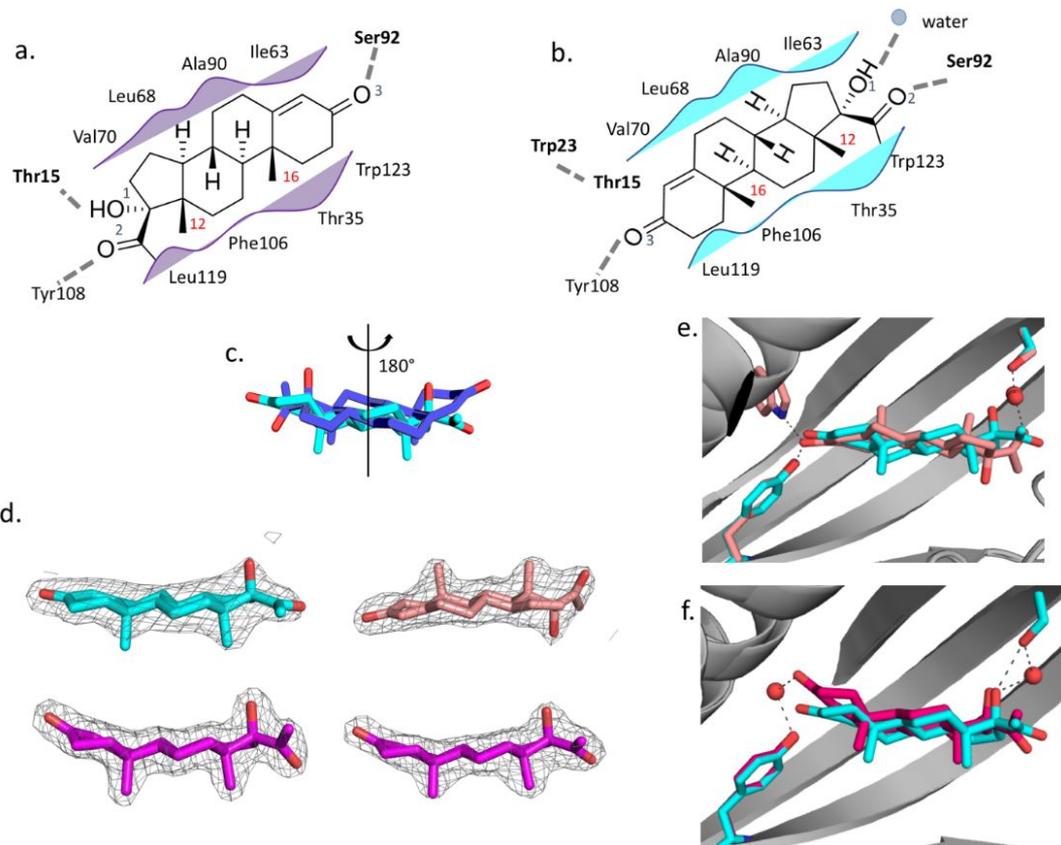
По результатам первой экспериментальной проверки (метод дрожжевого дисплея), этот дизайн не смог связать 17-ОПГ



Белок с неизвестной функцией (RV0760) из *Mycobacterium tuberculosis*
PDB ID: [2A15](#)

Примеры дизайнов. 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



Дизайн *de novo*

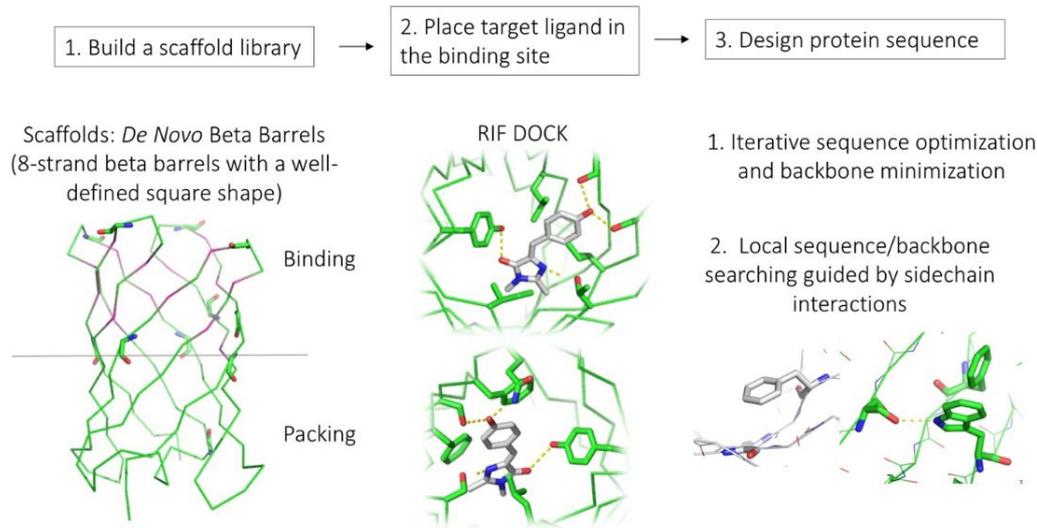
В отличие от дизайна с перепрофилированием белковой функции, описанного ранее, *de novo* дизайн в качестве входных данных не требует существования предопределенного остова (из природного белка)

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

- 1) создание библиотеки белковых скаффолдов, собранных *de novo* из коротких пептидных фрагментов; протокол лежит в открытом доступе
- 2) помещение интересующей нас малой молекулы в заданный сайт связывания посредством алгоритма RIF (Rotamer-Interaction-Field)
- 3) итеративная оптимизация последовательности и минимизация остова (Rozetta)
- 4) оптимизация взаиморасположения лиганда и окружающих его боковых цепей аминокислот (увеличиваем количество взаимодействий)
- 5) 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”

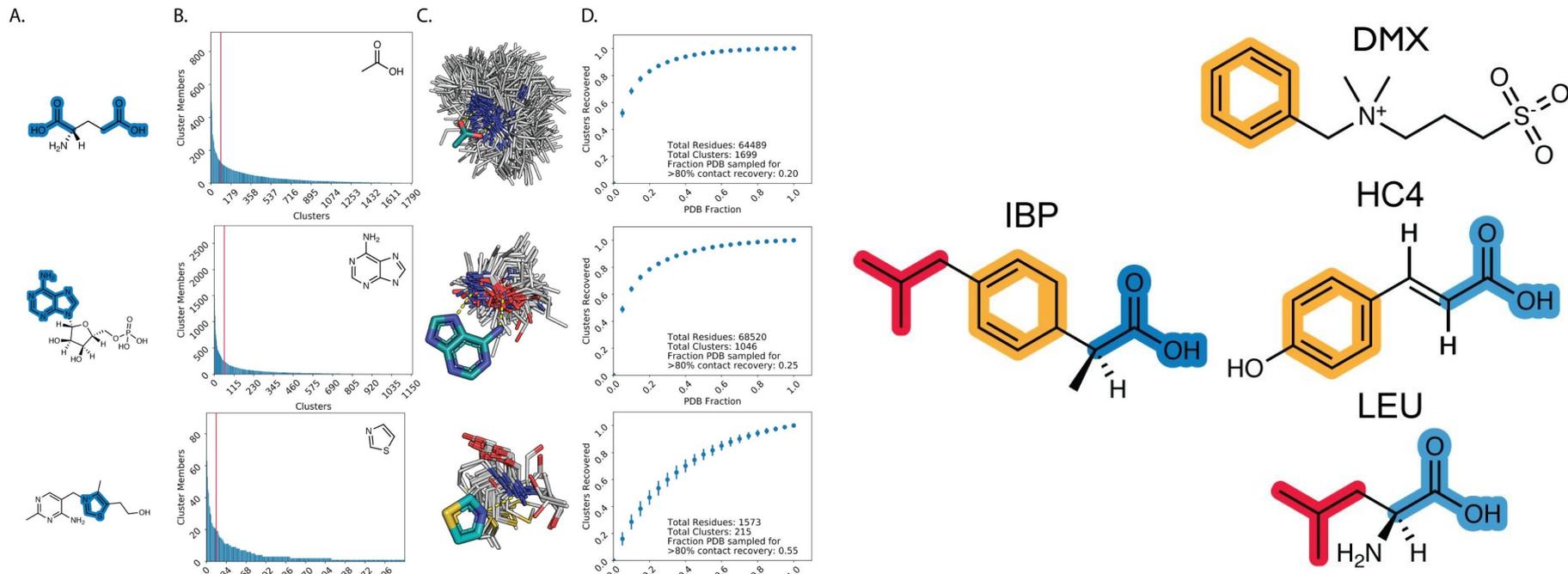
Источник: <https://doi.org/10.1371/journal.pcbi.1008178>

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



Декомпозиция лигандов на фрагменты

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

Пример алгоритма дизайна

Перед началом дизайна кармана связывания, проводилась предварительная фильтрация найденных контактов по значениям функции энергии (all-atom energy function, Rosetta):

$$\Delta E_{total} = \sum_i w_i E_i(\Theta_i, aa_i)$$

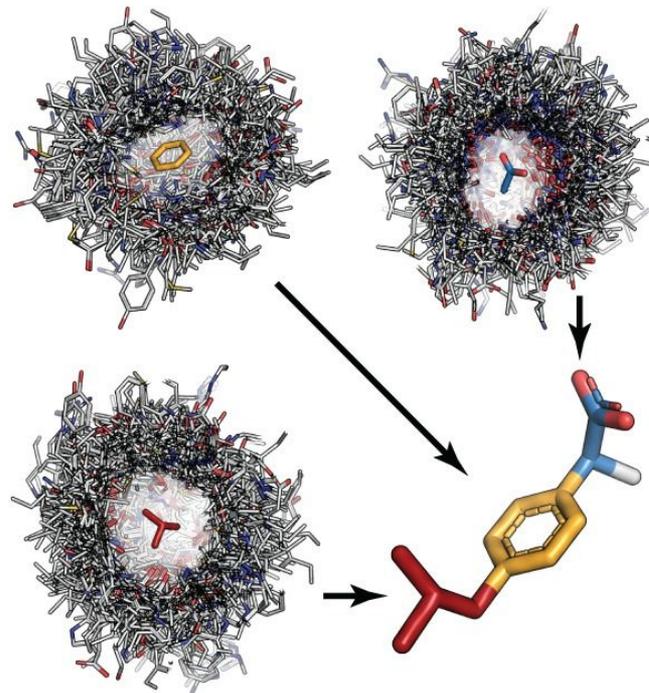
Используемые авторами термы энергии:

fa_rep	энергия отталкивания между двумя атомами, принадлежащих разным остаткам
fa_atr	энергия притяжения между двумя атомами, принадлежащих разным остаткам,
fa_elec	энергия взаимодействия между двумя несвязанными заряженными атомами,
hbond_sc	энергия водородных связей, существующих между двумя боковыми цепями
hbond_bb_sc	энергия водородных связей между боковой цепью и остовом
fa_sol	энергия сольватации

Пример алгоритма дизайна

Для ансамбля найденных контактов каждого из выделенных фрагментов проводилась кластеризация с целью выделения хорошо представленных контактов “фрагмент-белок”. Отбирались наиболее представленные кластеры

Все кластеризованные контакты “белок-фрагмент” позиционировались относительно лиганда - так был получен ансамбль выровненных контактов лиганда с белком



Пример алгоритма дизайна

К каждому конформеру лиганда-мишени был применен **алгоритм Монте-Карло** - производилась сборка дискретных боковых цепей аминокислот из ансамбля контактов в композитный сайт связывания

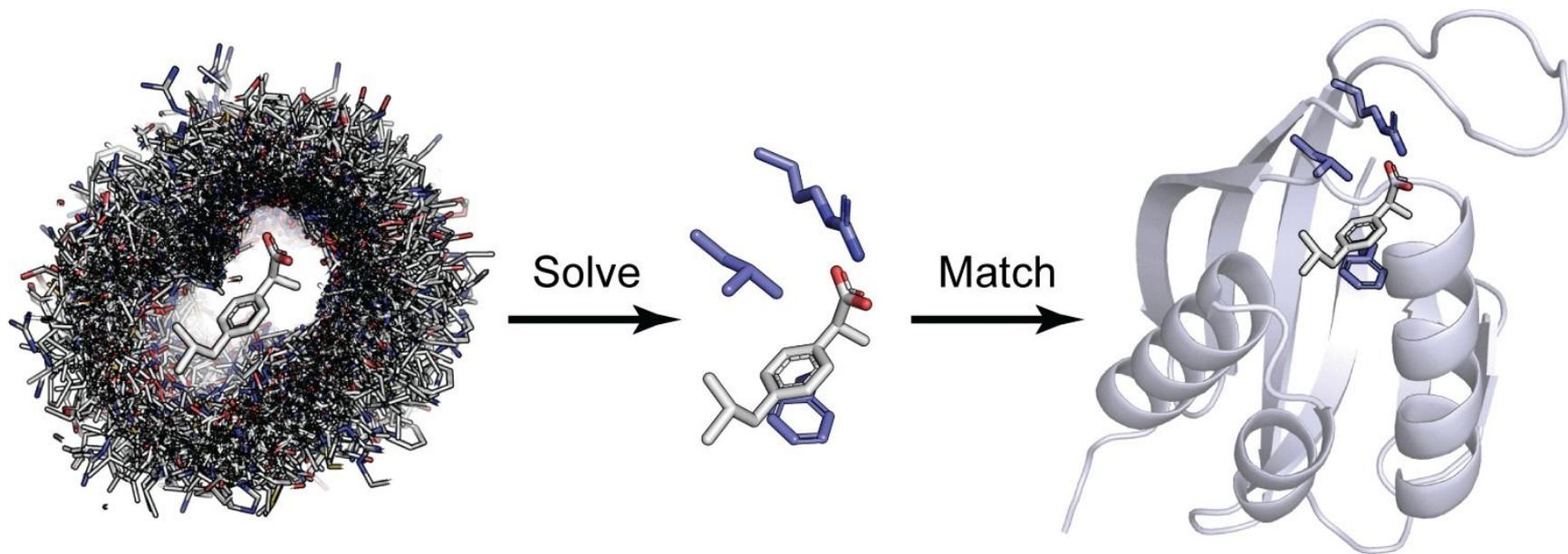
Сайт связывания лиганда инициализировался тремя случайными остатками из ансамбля контактов

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)

Пример алгоритма дизайна



Пример алгоритма дизайна

“Combinatorial methods for small-molecule placement in computational enzyme design”

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

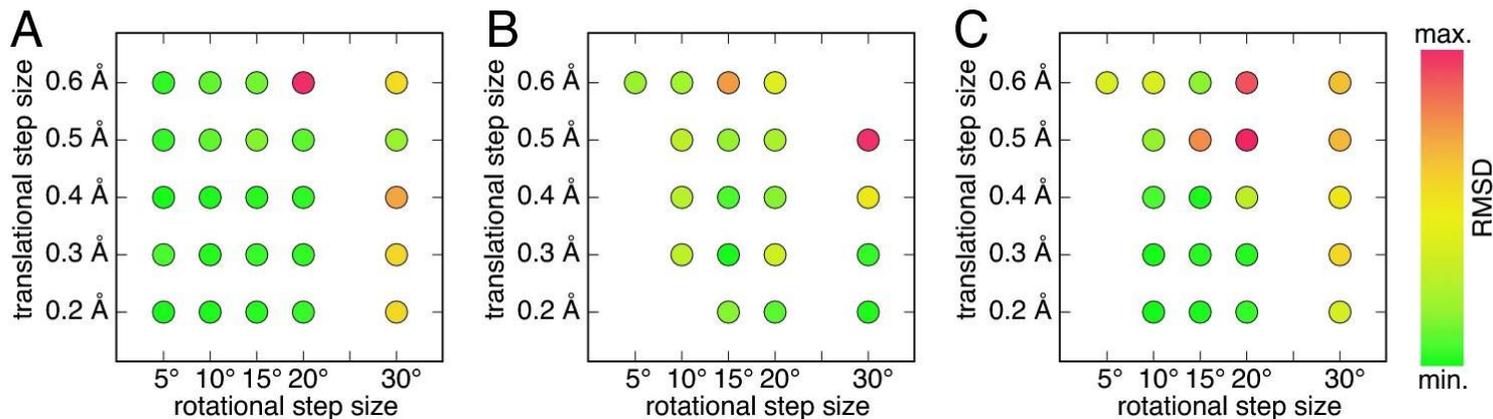
Алгоритм дизайна:

- 1) Создание большого набора координат лиганда
- 2) Редукция сгенерированного набора координат субстрата
- 3) Расчет попарных взаимодействий лиганд-белок
- 4) Опциональная поправка значений энергии → 'улучшение' контактов между аминокислотами белка и лигандом, важных для реализации белком каталитических функций/функций байндера
- 5) Определение оптимальной по энергии аминокислотной последовательности белка (сиквенса)

Создание большого набора координат лиганда

Набор координат лиганда генерировался одним из двух способов:

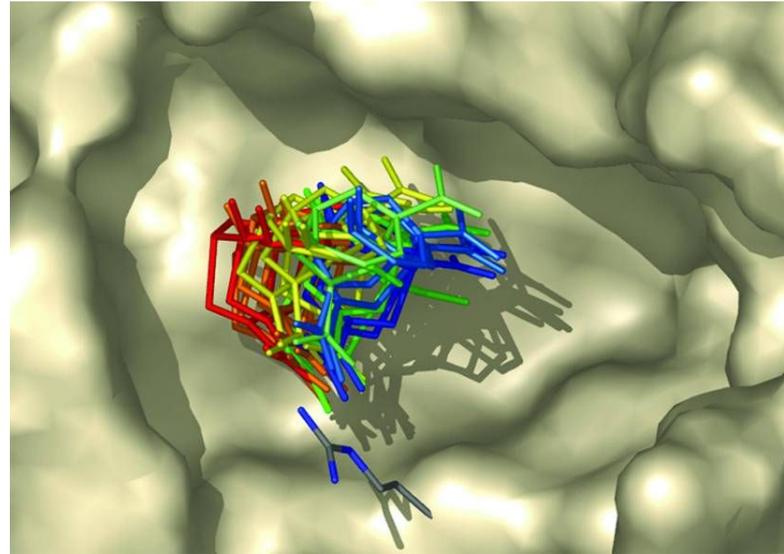
- 1) простое преобразование исходных координат атомов в виде линейного смещения и вращения
- 2) метод таргетного размещения



Создание большого набора координат лиганда

Набор координат лиганда генерировался одним из двух способов:

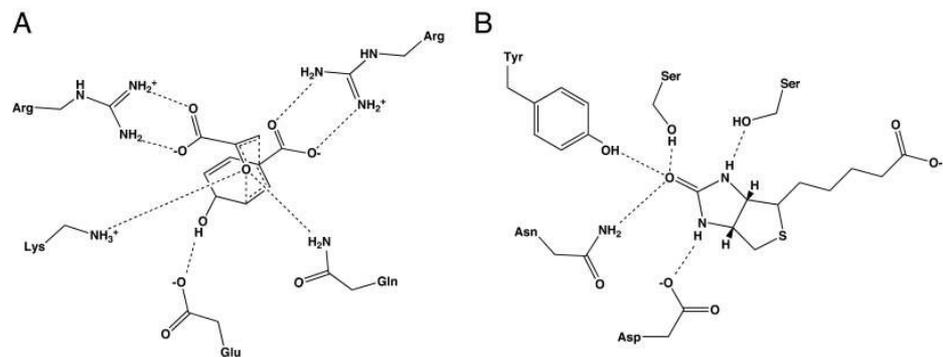
- 1) простое преобразование исходных координат атомов в виде линейного смещения и вращения
- 2) метод таргетного размещения



Источник: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1636520/>

Редукция сгенерированного набора координат субстрата. Расчет энергии взаимодействий

- $10^6 - 10^9$ вариантов координат лиганда
→ 20 000 вариантов
- для фильтрации конформаций производилась **проверка на наличие клэшей** между лигандом и белковым остовом (потенциал твердых сфер)
- дополнительный раунд фильтрации - **удовлетворение конформации лиганда геометрическим параметрам**, определенных пользователем (характеристика контактов боковых цепей аминокислот с малой молекулой)



Источник: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1636520/>

Определение белкового сиквенса

- поиск последовательностей, соответствующих оптимальной по энергии структуре белка производился с помощью методов FAST/HERO
- происходила оптимизация последовательности всей белковой цепи, за исключением позиций, 'защищенных' пользователем от внесения мутаций
- результат - минимальная по энергии белковая последовательность, реализующая максимальное количество определенных пользователем контактов аминокислота - лиганд

Список полезных источников

- 1) Dou J. Exploring the Molecular Design of Ligand Binding Sites by Computational Protein Design : дис. – 2017.
- 2) Lucas J. E., Kortemme T. New computational protein design methods for de novo small molecule binding sites //PLoS computational biology. – 2020. – Т. 16. – №. 10. – С. e1008178.
- 3) Alford R. F. et al. The Rosetta all-atom energy function for macromolecular modeling and design //Journal of chemical theory and computation. – 2017. – Т. 13. – №. 6. – С. 3031-3048.
- 4) Nerattini F. et al. Design of Protein–Protein Binding Sites Suggests a Rationale for Naturally Occurring Contact Areas //Journal of chemical theory and computation. – 2018. – Т. 15. – №. 2. – С. 1383-1392.
- 5) Tinberg, C. E., & Khare, S. D. (2016). Computational Design of Ligand Binding Proteins. Computational Protein Design, 363–373.
- 6) Lassila J. K. et al. Combinatorial methods for small-molecule placement in computational enzyme design //Proceedings of the National Academy of Sciences. – 2006. – Т. 103. – №. 45. – С. 16710-16715.