

Факультет биоинженерии и биоинформатики,
Московский государственный университет имени М. В. Ломоносова



Подвижность, регуляция, дизайн

Лекция 7, биоинформатика, 4 курс ФББ МГУ, осенний семестр
Злобин А. С., alexander.zlobin@fbb.msu.ru

Подвижность – масштабы

Термические колебания связей и углов	Фемтосекунды
Акт химической реакции	Фемтосекунды
Переключение водородных связей в воде	Пикосекунды
Изменение конформаций боковых радикалов	Пикосекунды
Конформационные перестройки подвижных петель	Наносекунды
Процесс связывания субстрата/диссоциации продукта	Наносекунды
Конформационные перестройки доменов	Микросекунды-миллисекунды
Фолдинг	Микросекунды-миллисекунды
Мат. ожидание времени до химической реакции	Миллисекунды-часы

Подвижность – как изучать?

1. Спекулировать на значениях В-факторов
2. Получить из данных ЯМР релаксации
3. Экстраполировать из статичных структур (морфинг)
4. Изучить главные колебательные степени свободы
5. Промоделировать методами молекулярной динамики

Морфинг

Популярный способ визуализации в статьях не-моделистов

Не несет почти никакой количественной информации, только визуализация

NMA

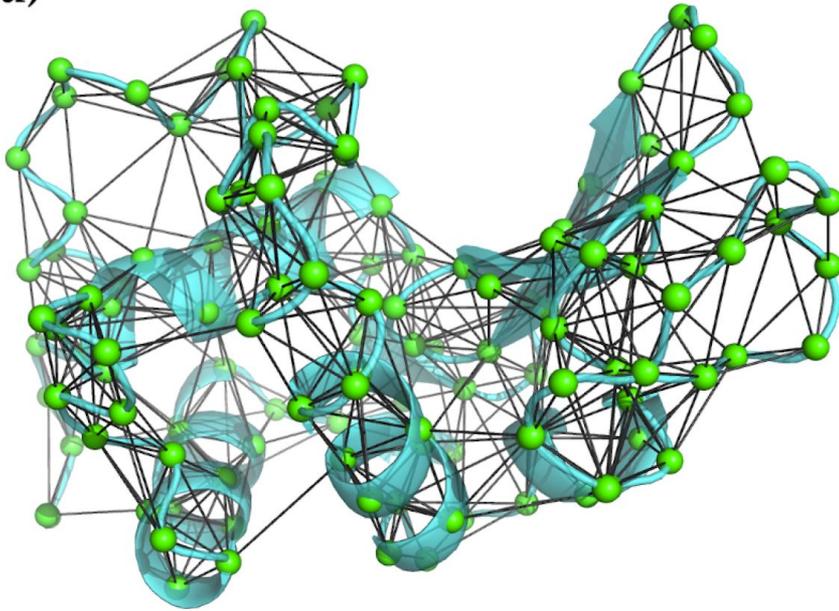
Анализ главных колебаний (Normal Mode Analysis)

Вычисление сил, действующих на каждый атом со стороны всех остальных

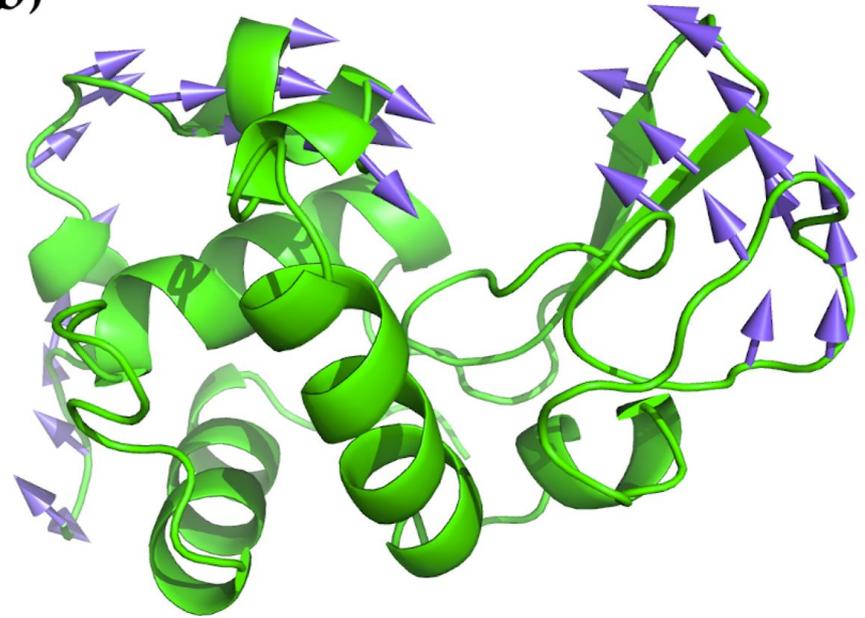
Выделение основных колебательных степеней свободы

Самые медленные степени свободы отвечают самым вероятным направлениям развития системы в динамике

(a)



(b)



А как мы можем рассчитать силы?

Нужно знать заряды на каждом атоме, жесткость каждой связи, угла, торсиона, силу ван-дер-Ваальсовых взаимодействий для каждой пары атомов. Вид функции, задающей каждый тип взаимодействия, и нужные эмпирические параметры вместе называются **силовым полем**.

- Параметры подбираются для воспроизведения макроскопических свойств (сохранение структуры, радиуса гирации, плотности воды, температуры кипения, координационного числа металла...)
- Функциональная форма подбирается для нужного уровня баланса скорости и точности счета

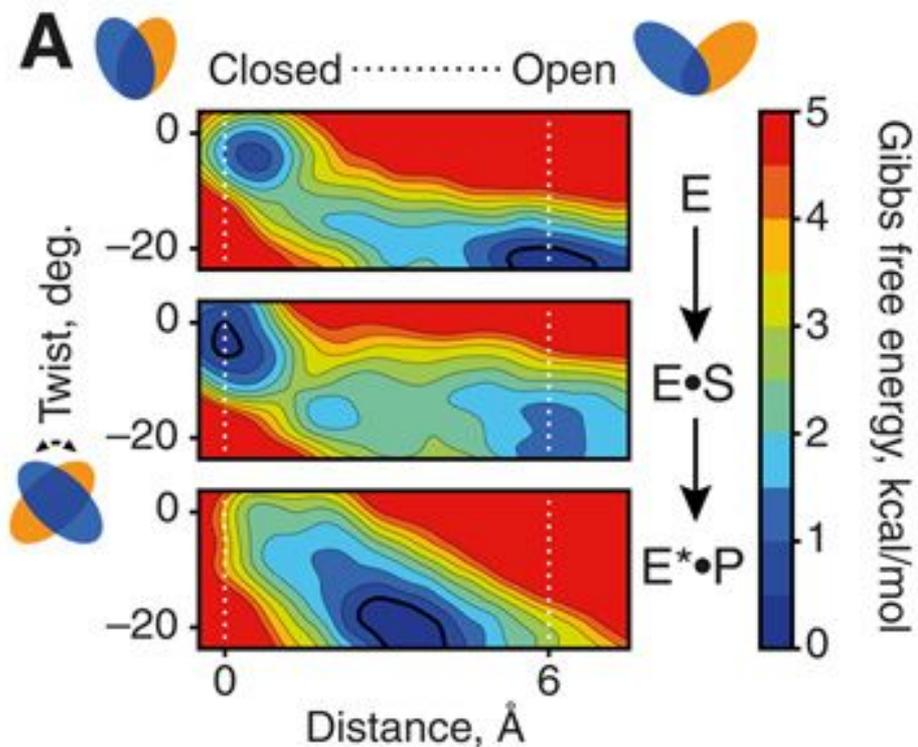
Регуляция

При работе с любой белковой структурой в реальности мы на самом деле имеем дело с ансамблем ее конформаций.

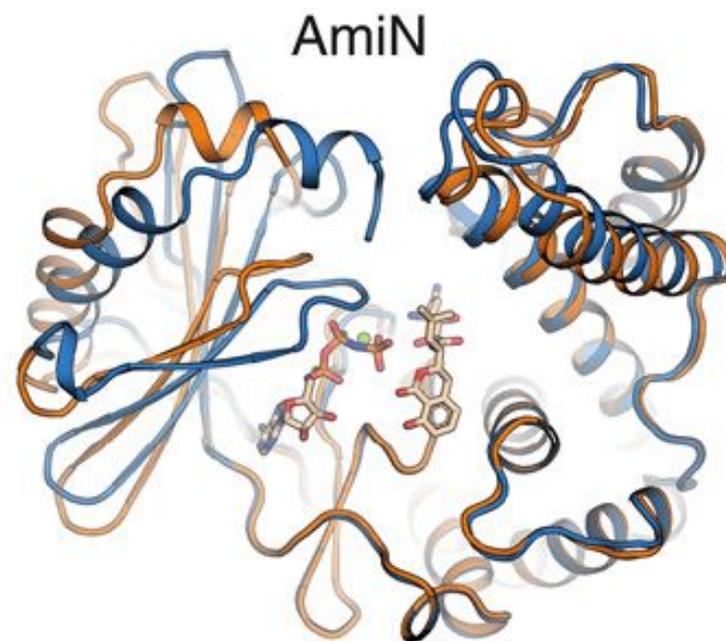
Внешние стимулы могут сдвигать равновесие в сторону тех или иных конформаций:

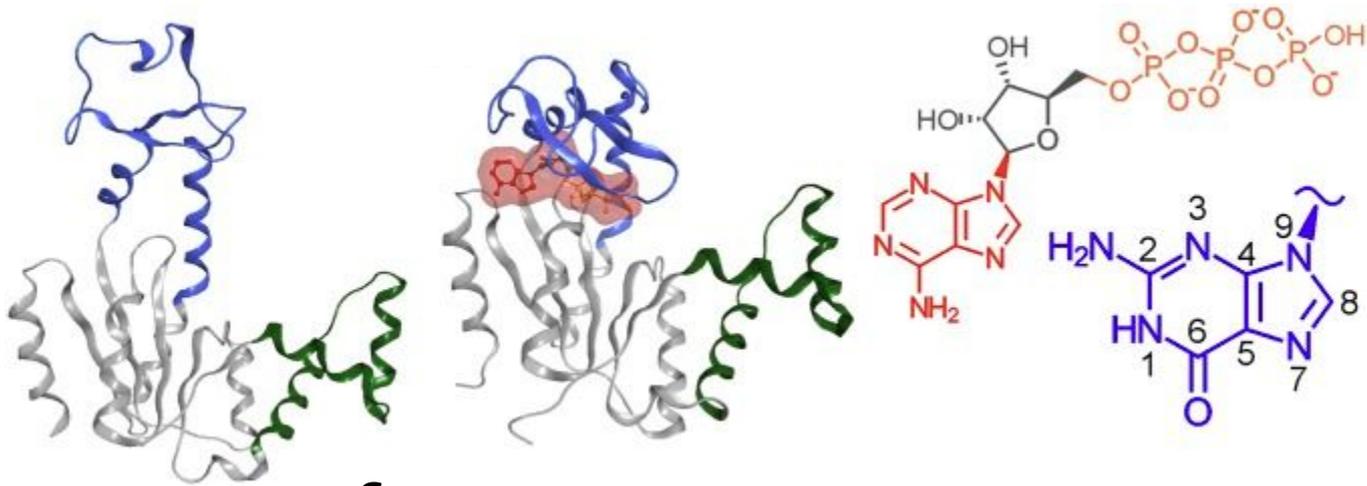
- Связывание нужного кофактора
- Связывание подходящего субстрата
- Фосфорилирование
- Протеолиз
- pH
- Связывание аллостерических модуляторов

Адаптация к правильному субстрату

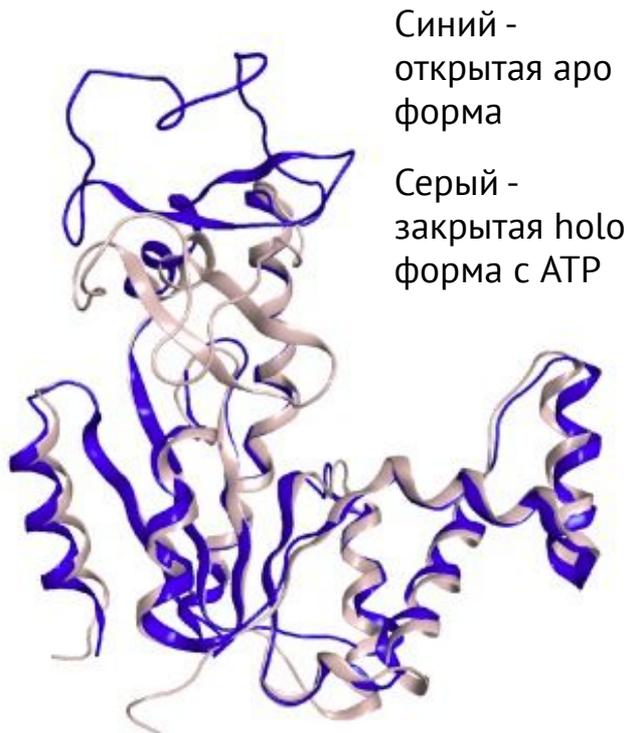


E = AmiN(D202)•ATP•Mg²⁺
E* = AmiN(D202-H)•ADP•Mg²⁺



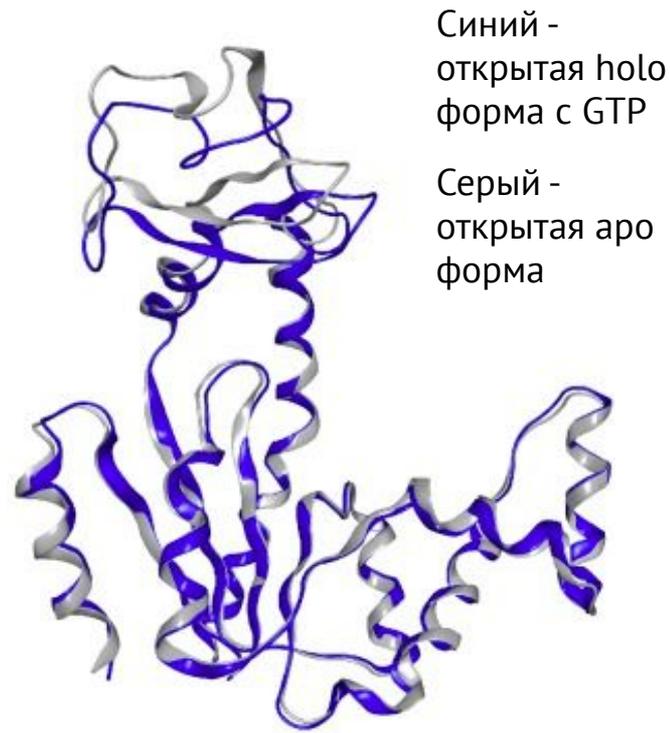


Селективность аденилат киназы



Синий -
открытая apo
форма

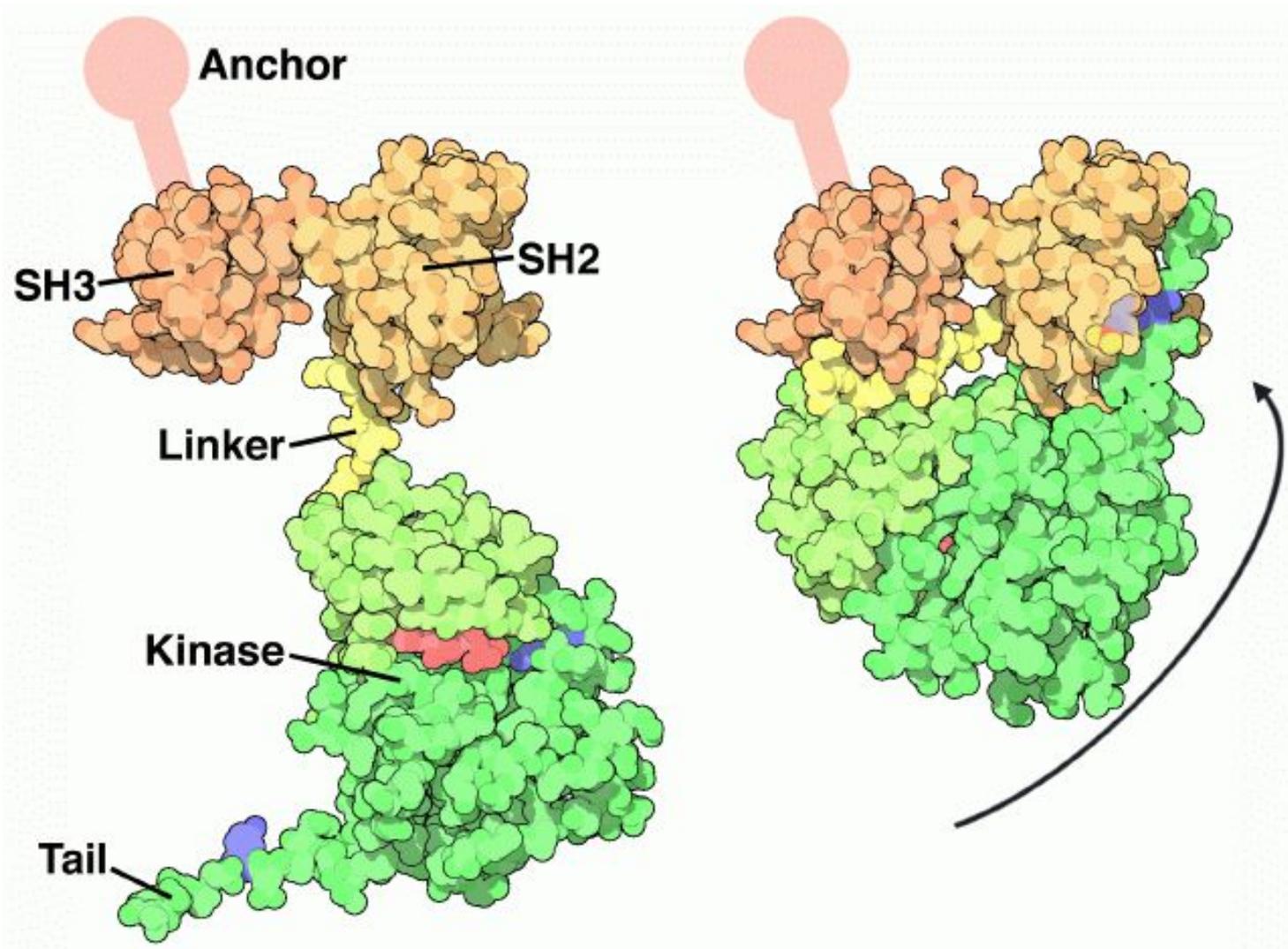
Серый -
закрытая holo
форма с ATP

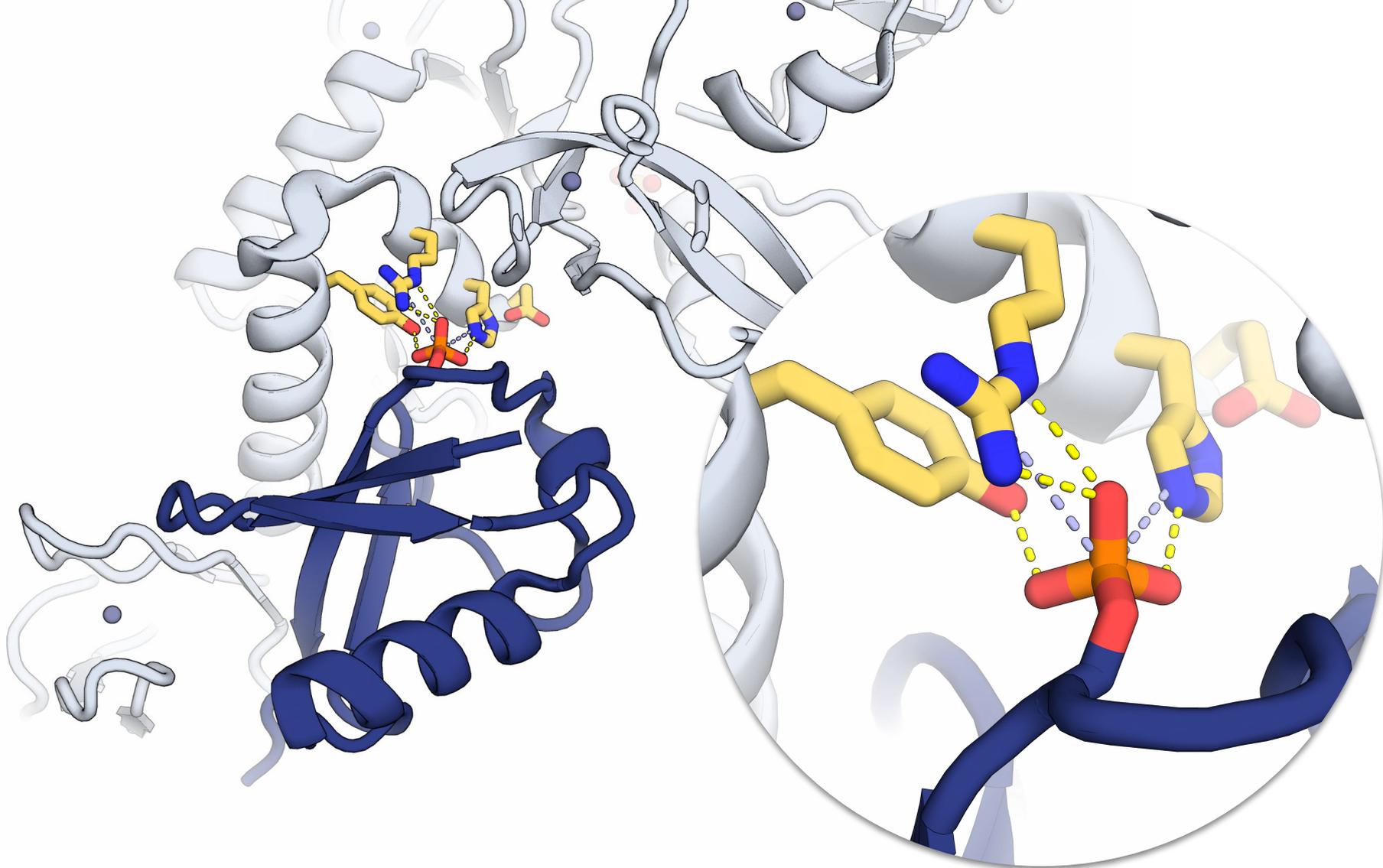


Синий -
открытая holo
форма с GTP

Серый -
открытая apo
форма

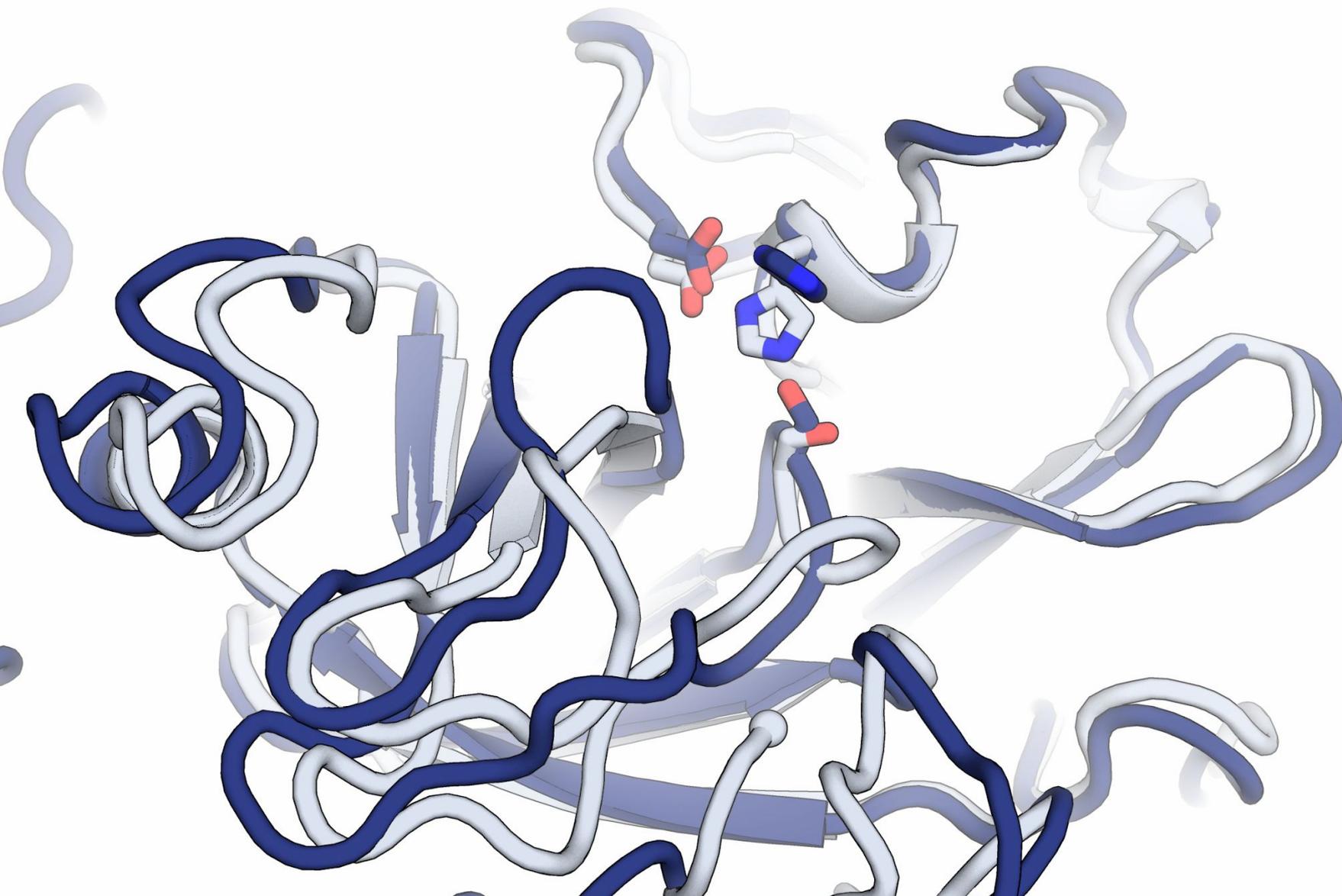
Адаптация к ковалентной модификации

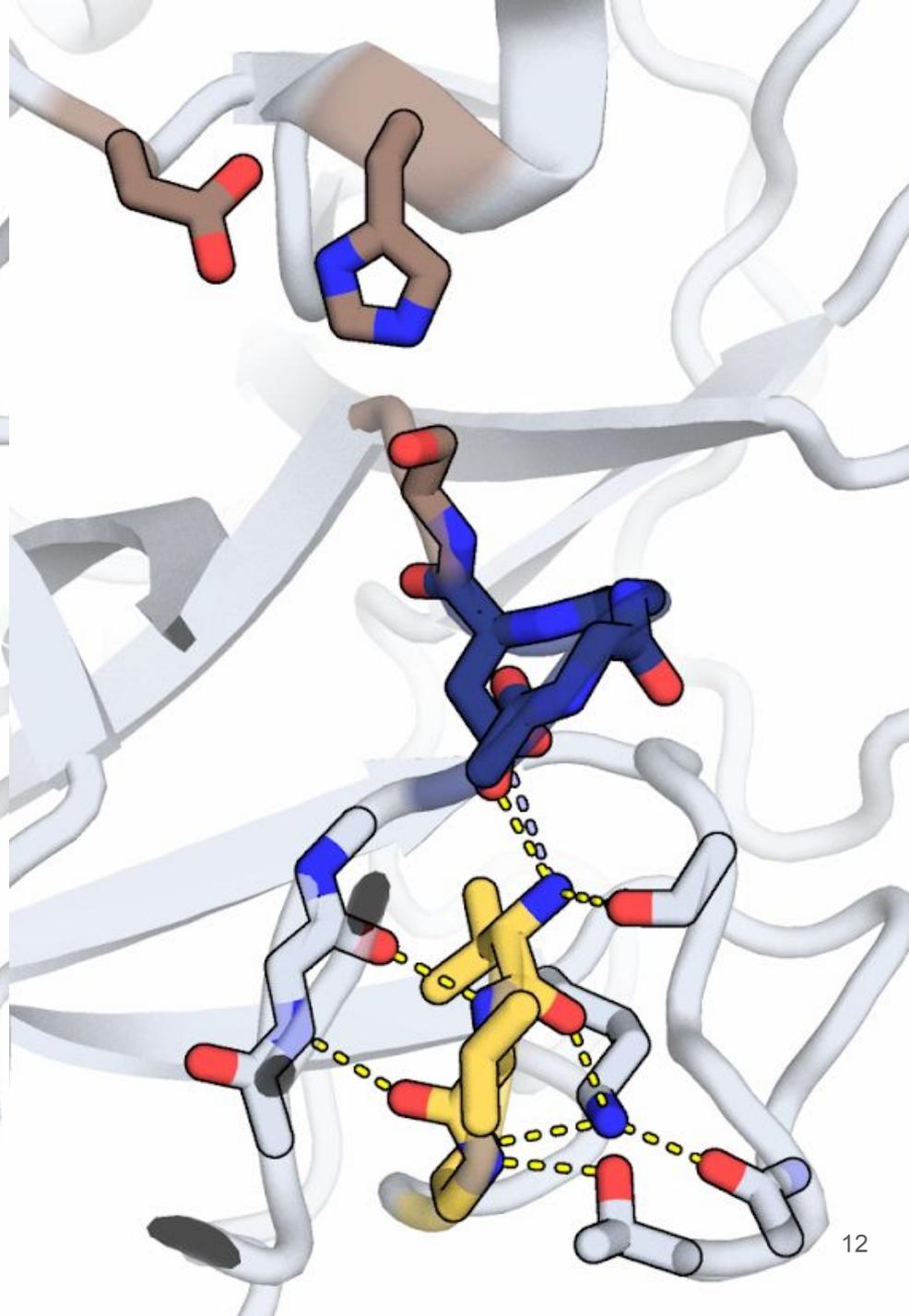
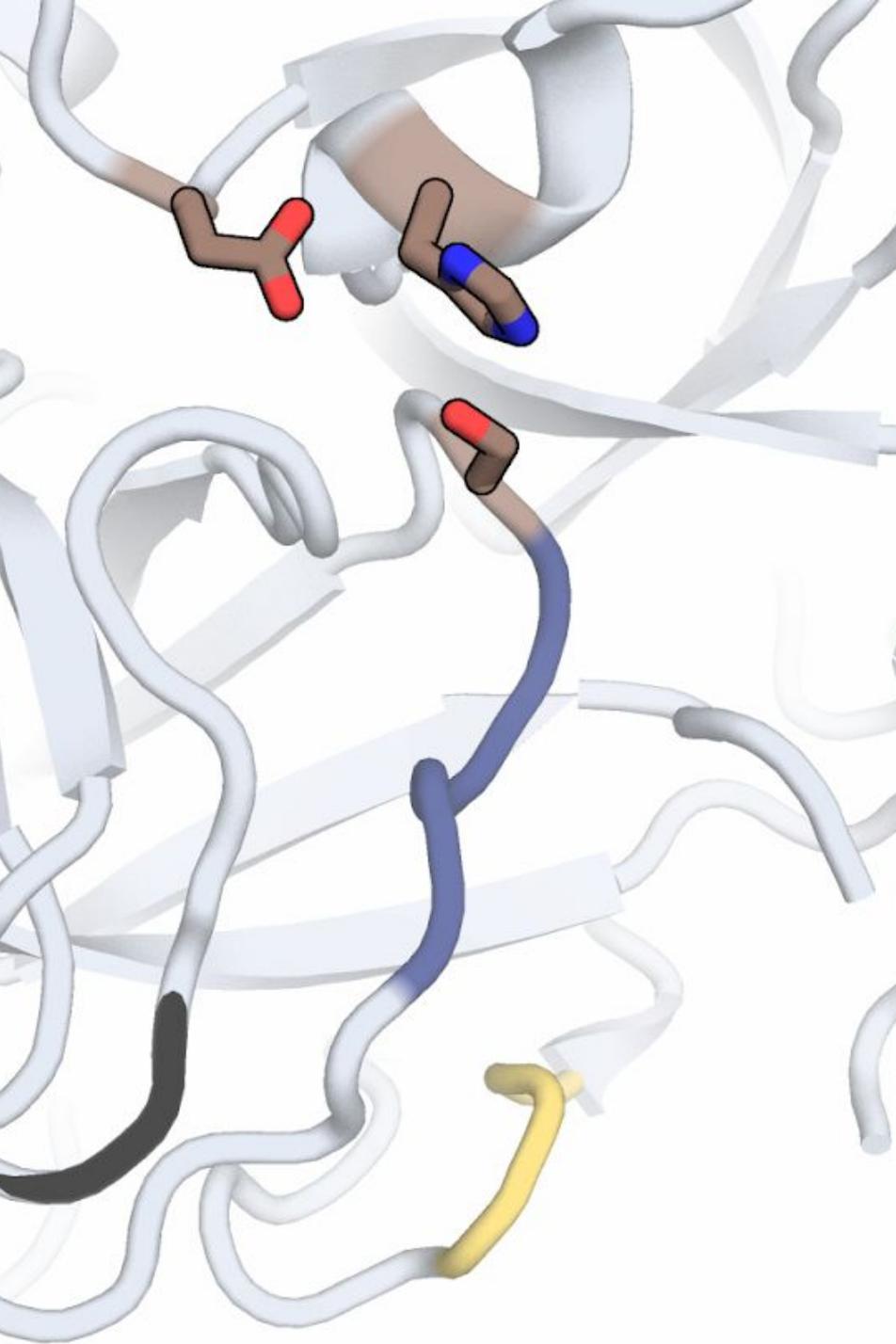




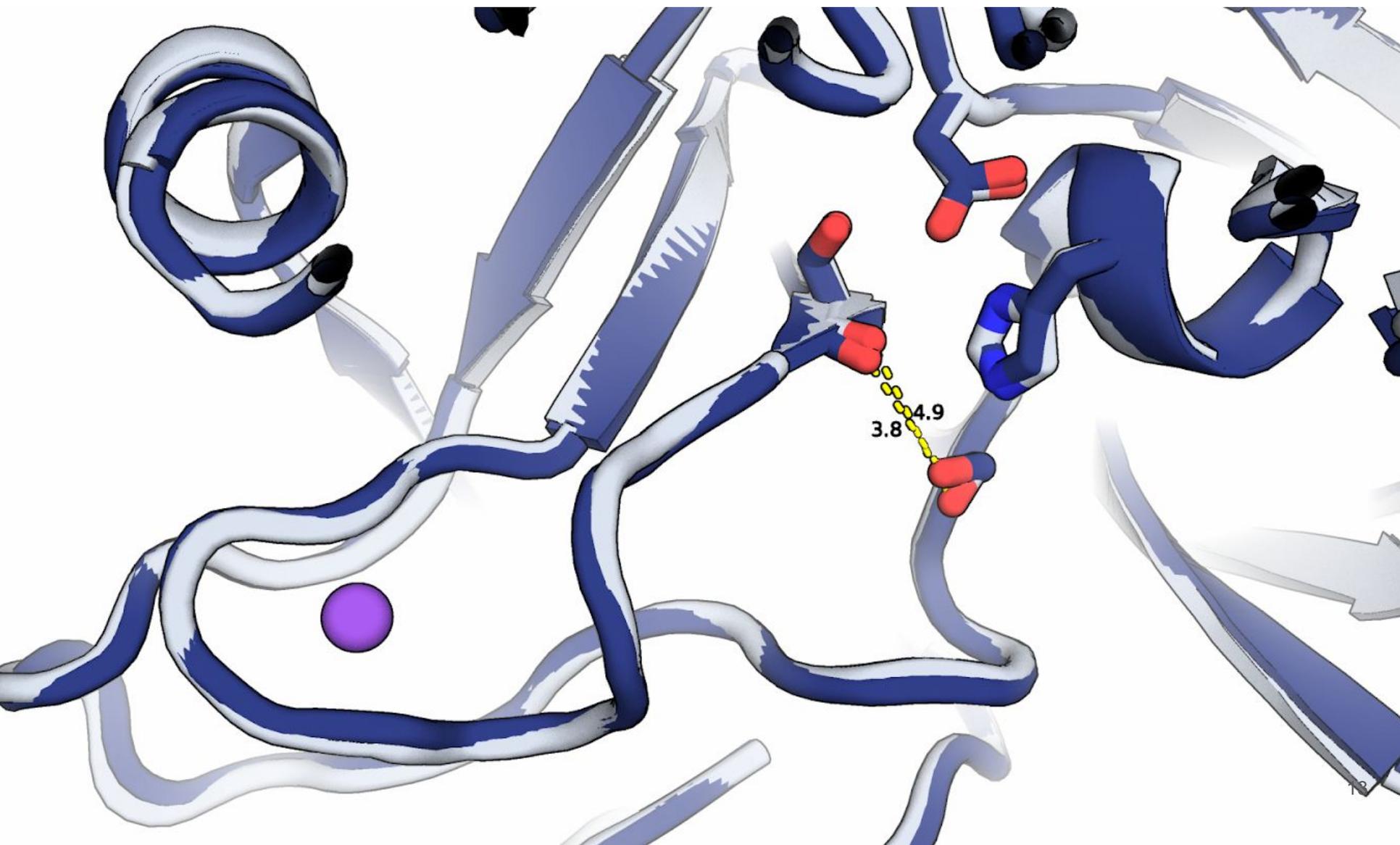
Паркин (убиквитин лигаза, серая) узнает фосфорилированный убиквитин (синий), а нефосфорилированный связывает гораздо хуже

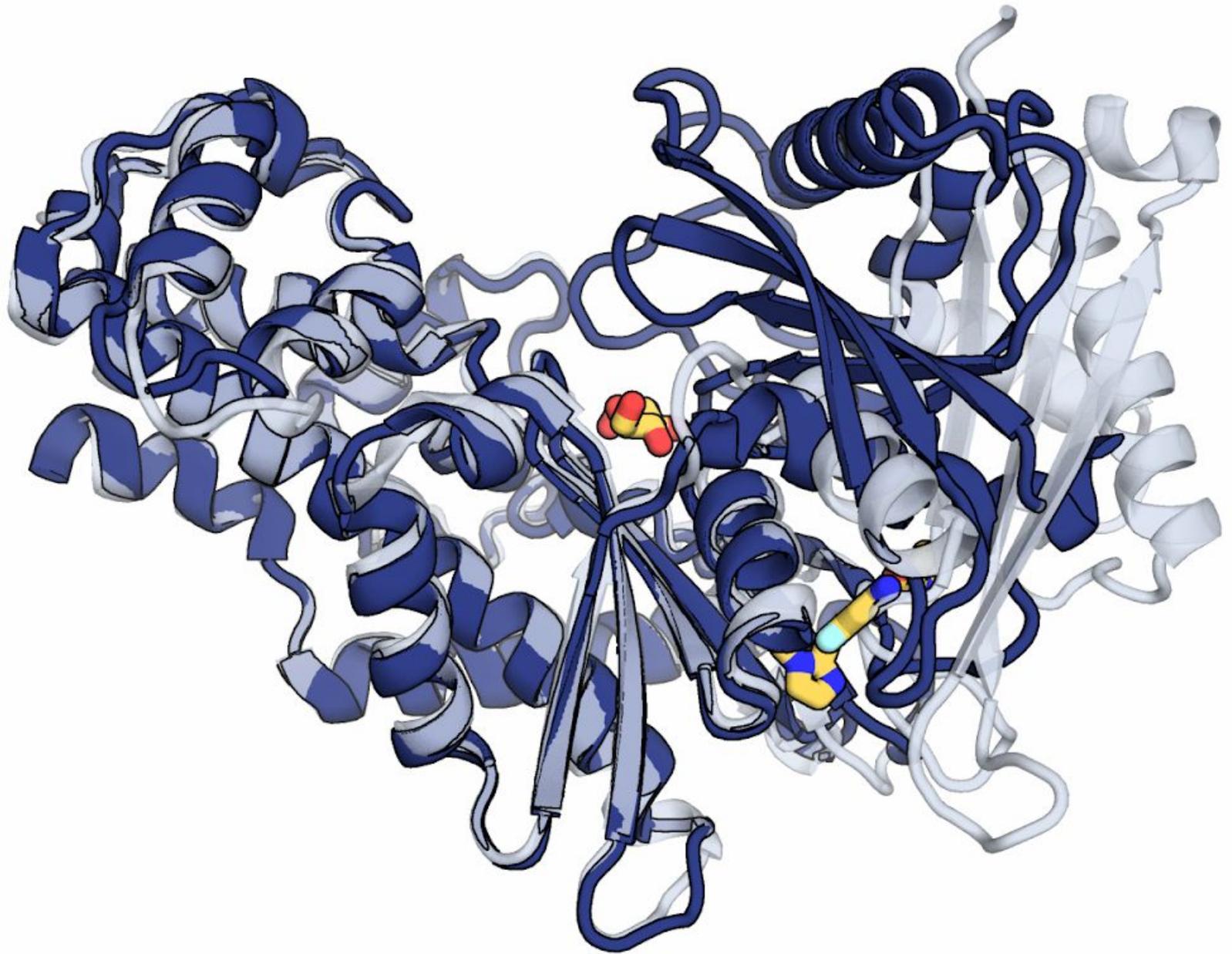
Адаптация к разрезанию

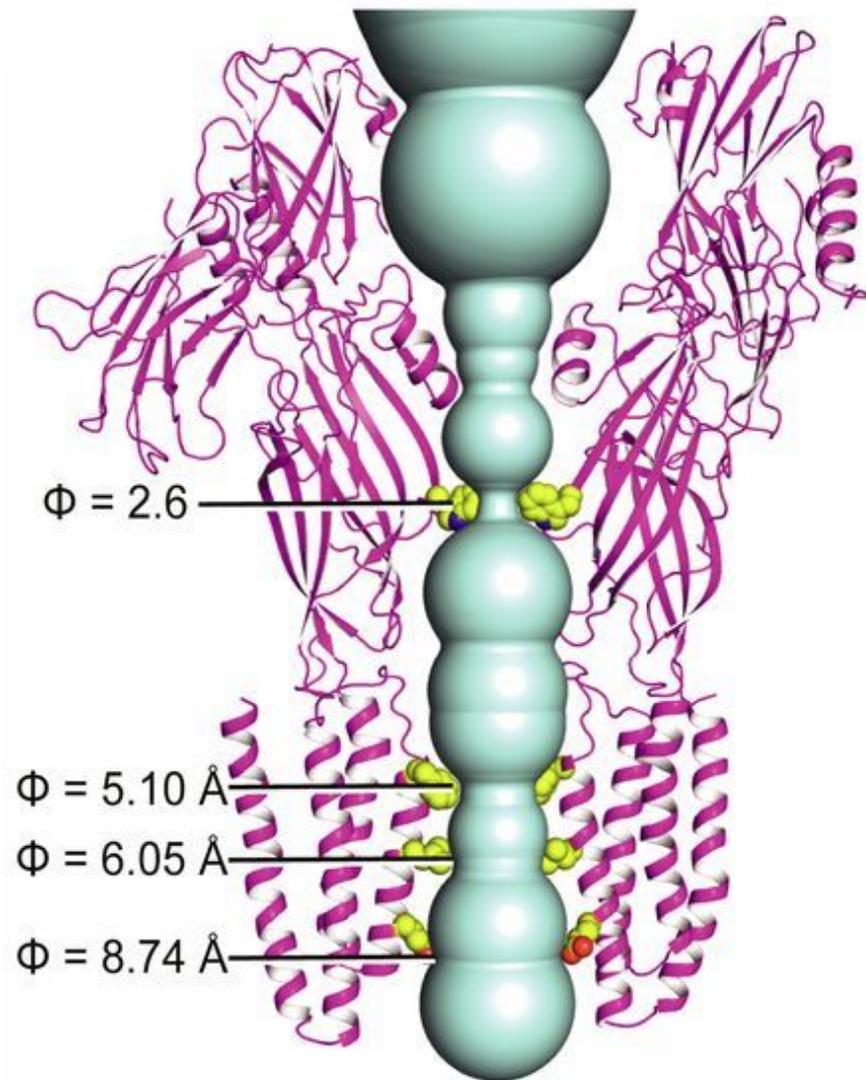
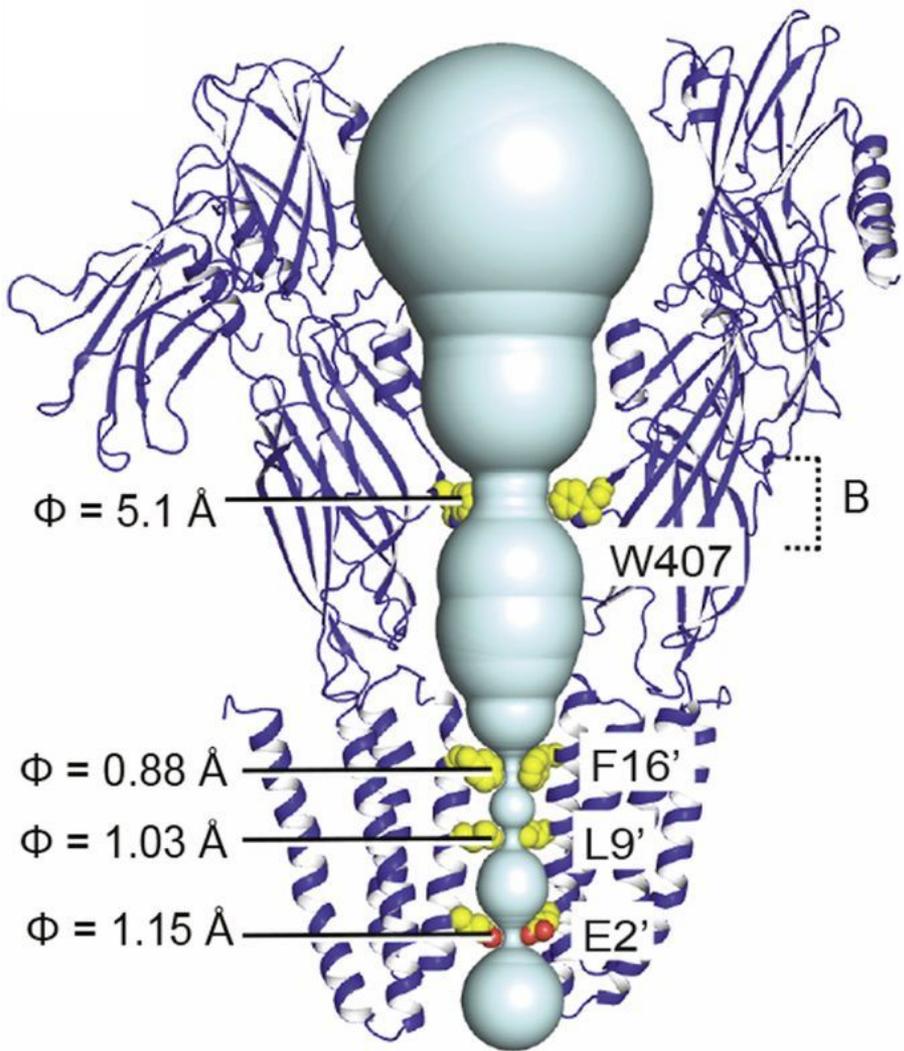




Аллостерические модуляторы

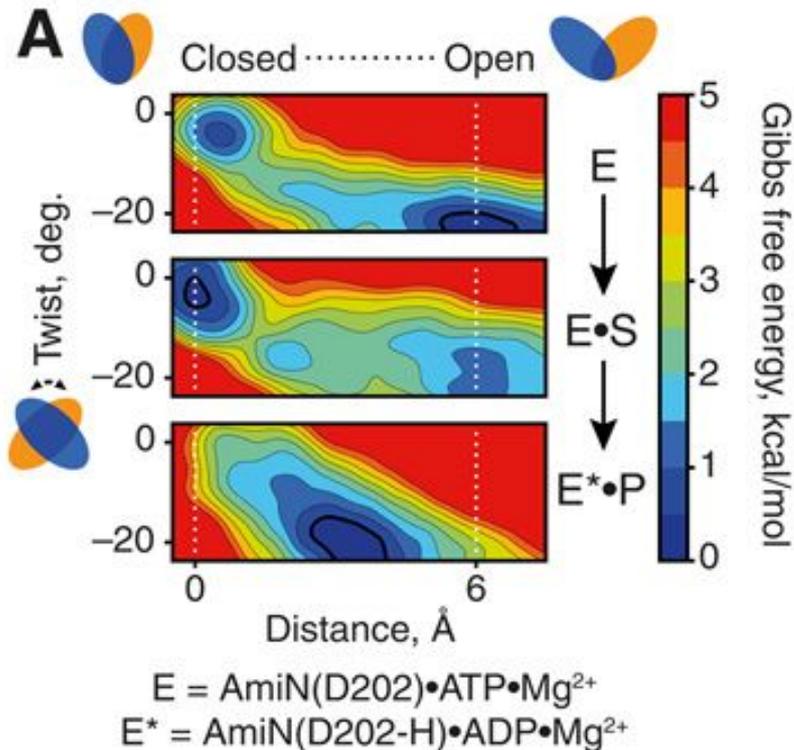






Аллостерические модуляторы

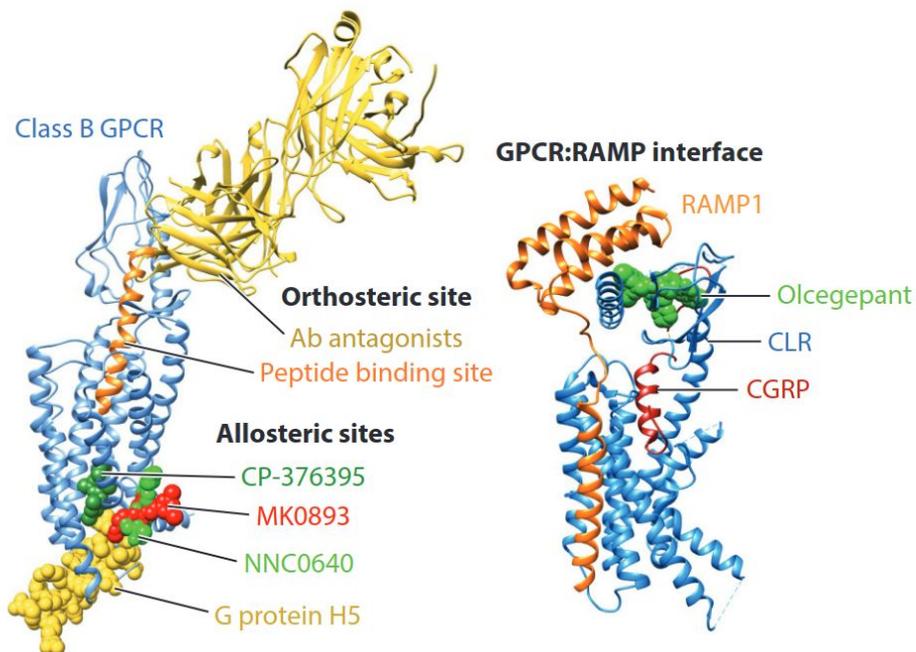
Не всегда эффект очевиден из кристаллографических структур, так как связывание модулятора влияет на динамические свойства и ансамбль конформаций, но кристаллизация возможна только из какой-то одной конформации.



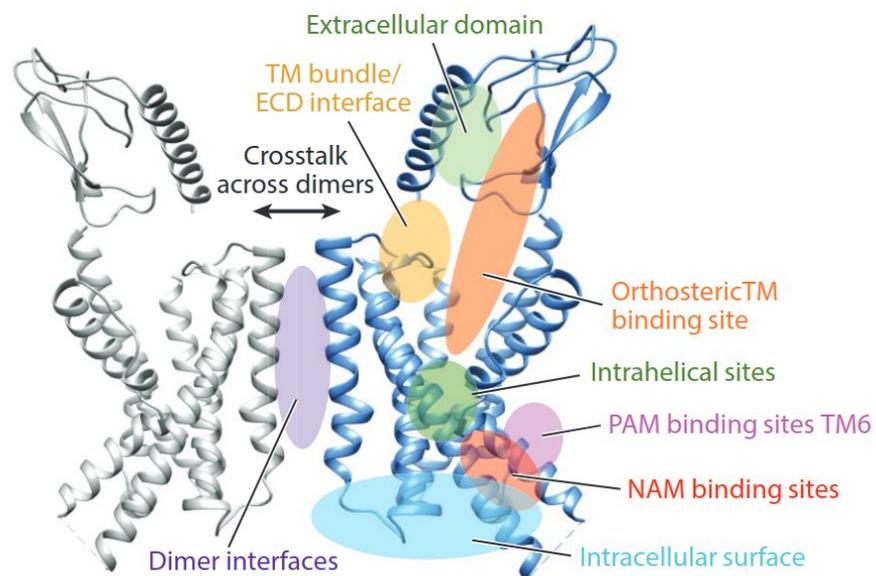
Также модуляция путем сдвига населенности состояний возможна на любом из этапов вдоль реакции.

Поиск аллостерических сайтов

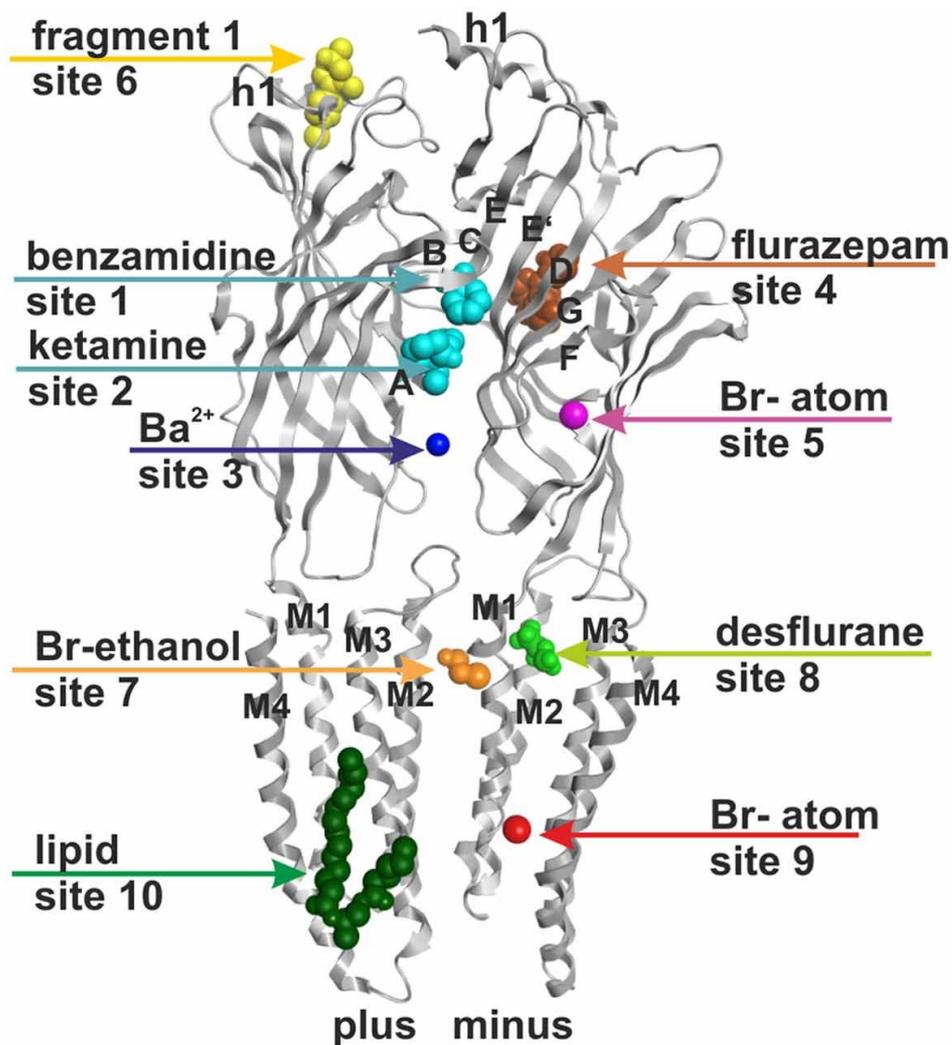
a Class B GPCR binding sites confirmed in structures



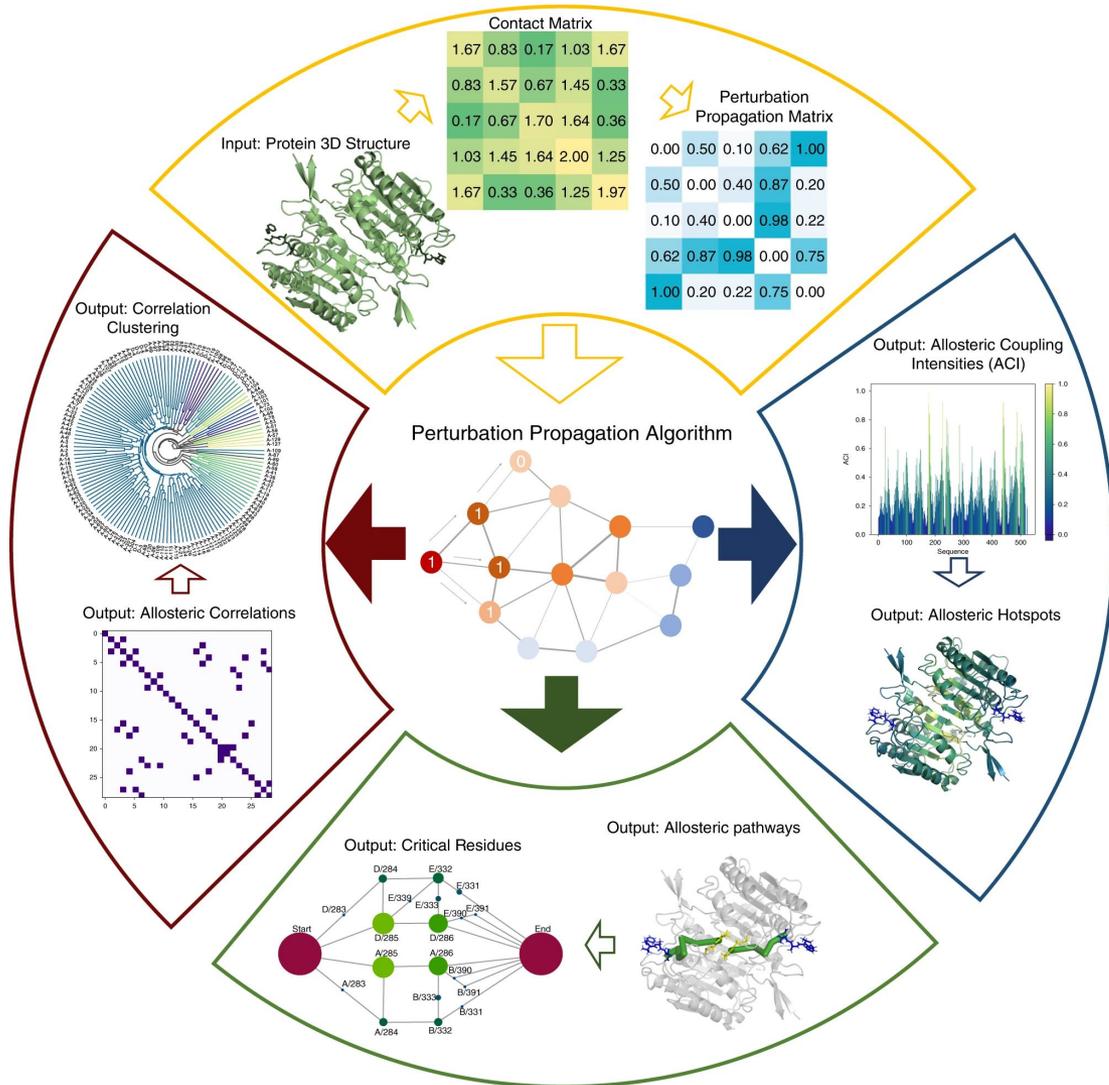
b Potential/predicted class B GPCR allosteric binding sites



Поиск аллостерических сайтов



А как это все работает-то?

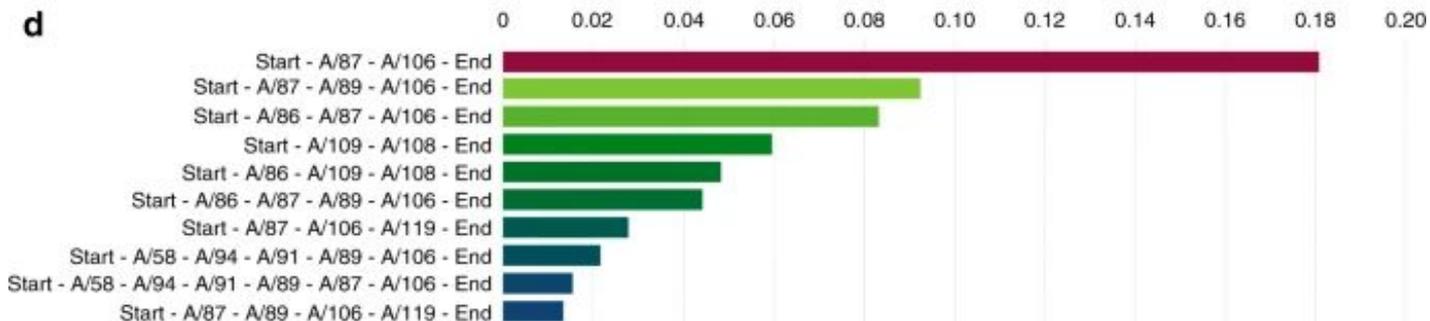
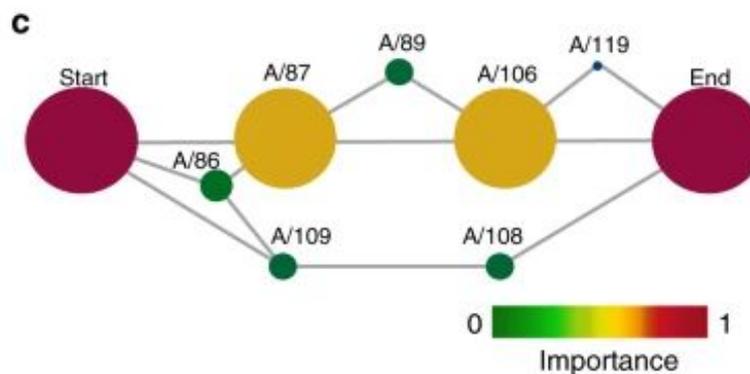
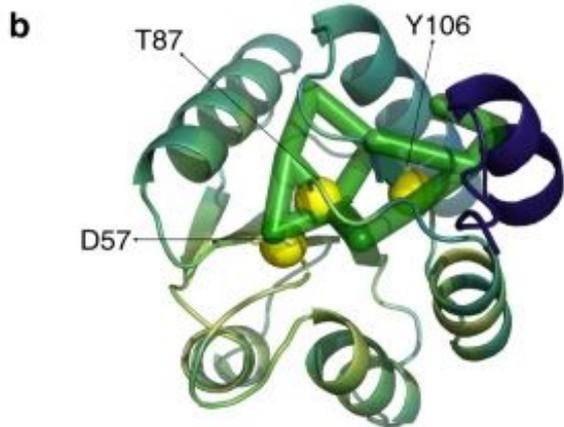


Движения ортостерического и аллостерического сайтов скоррелированы

Чтобы это достигалось, должны быть механизмы передачи информации через структуру белка

Предполагают наличие путей передачи информации через структурные возмущения. Их можно попытаться выделить на основании скоррелированности движений остатков во времени или анализа статичных структур

А как это все работает-то?



Сервер [OHM](#) работает на статической карте контактов по одному PDB

Перерыв

Задачи моделирования

Предсказание структуры, в которую свернется последовательность

Получение информации о подвижности и конформационных ансамблях

Получение информации о механизмах процессов и их энергетических параметрах

Задачи дизайна

Подбор последовательности для реализации заданной функции, описанной через информацию о структуре

Предсказание структуры

Моделирование по гомологии

Если есть структура у родственного белка, взять ее как темплейт и адаптировать

Трединг

Если нет структур родственных белков, попытаться “примерить” последовательность на набор различных фолдов, лучший использовать как темплейт

Knowledge-based фолдинг

Использование какой-либо информации из PDB, но не темплейтов целиком

Аб-инишио фолдинг (из первых принципов)

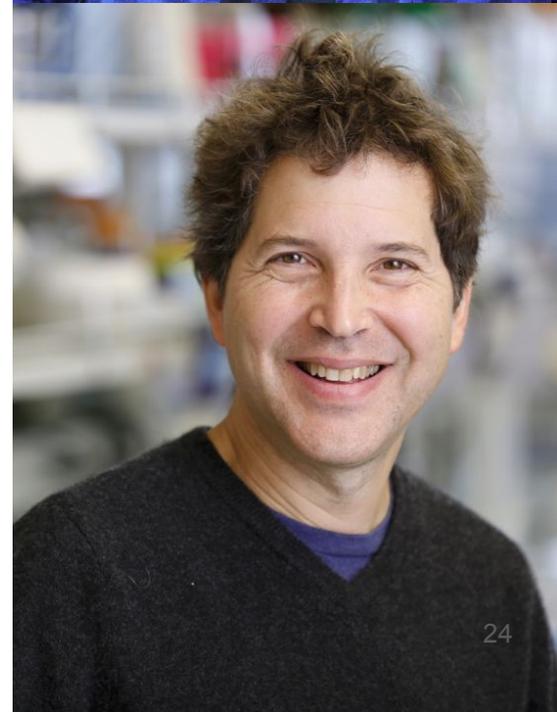
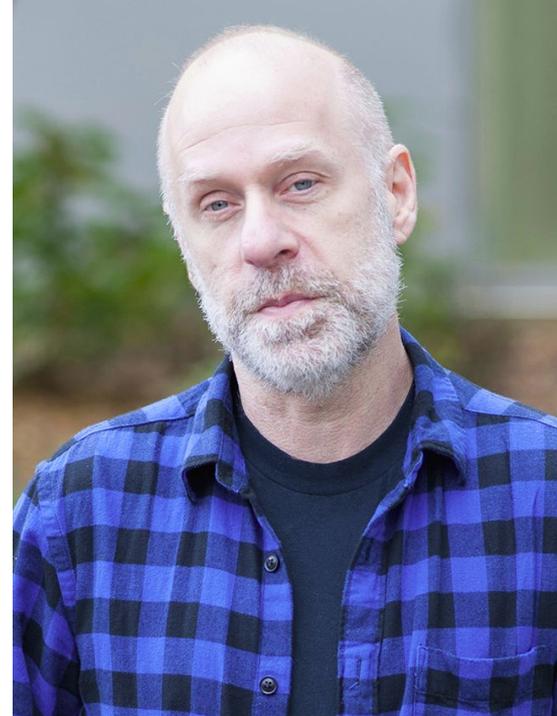
Не использовать темплейт или какие-либо статистические данные

Rosetta

Фреймворк для написания протоколов по предсказанию структуры (исходно) и дизайну (впоследствии) белков

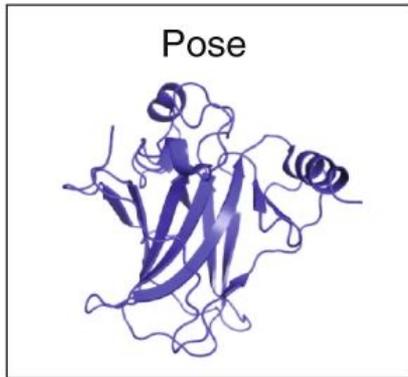
Основные составляющие:

- Библиотека ротамеров
- Библиотека остовных геометрий
- Процедура генерации конформации
- Скоринг-функция для оценки благоприятности конформации
- Процедура принятия/отбраковывания конформации

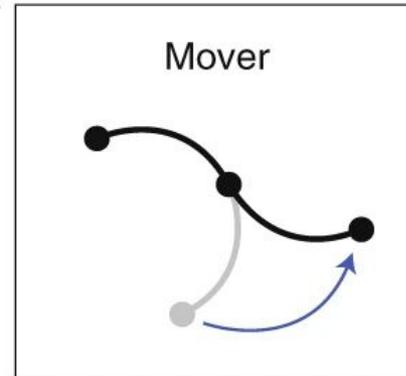


a

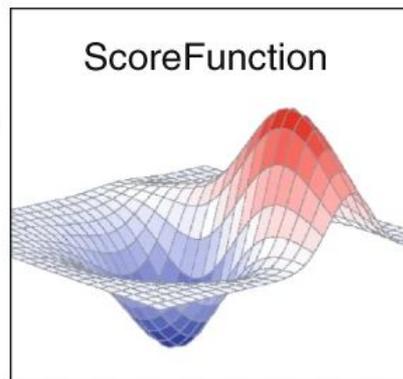
Select residues using
ResidueSelectors



Define side-chain identity and
organization by TaskOperations

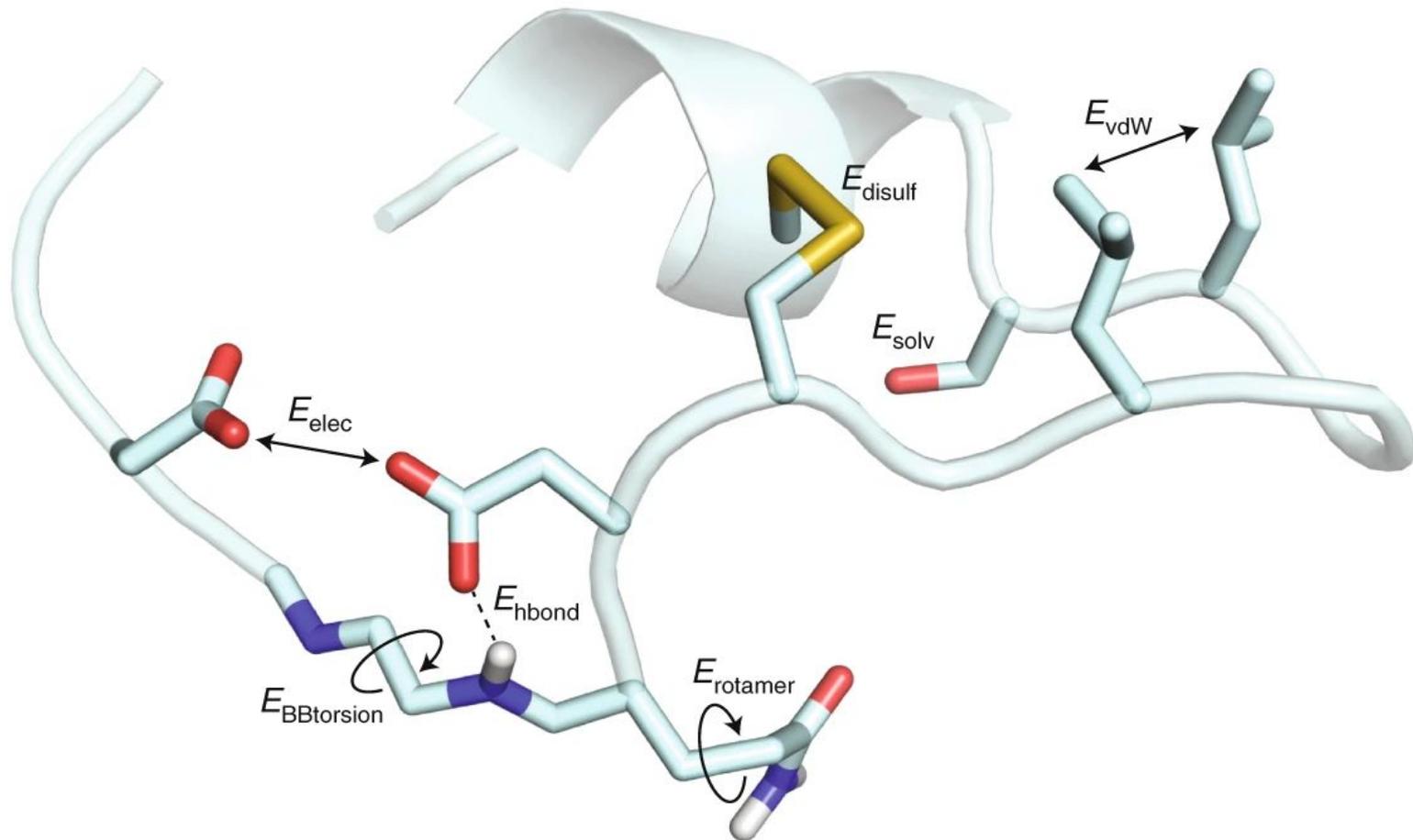


ScoreFunction
evaluates
whether new
conformation
is accepted



New conformation
is created during a move

- E_{vdW} Lennard–Jones for attractive or repulsive interaction
- E_{hbond} Hydrogen bonding allows buried polar atoms
- E_{elec} Electrostatic interaction between charges
- E_{disulf} Disulfide bonds between cysteines

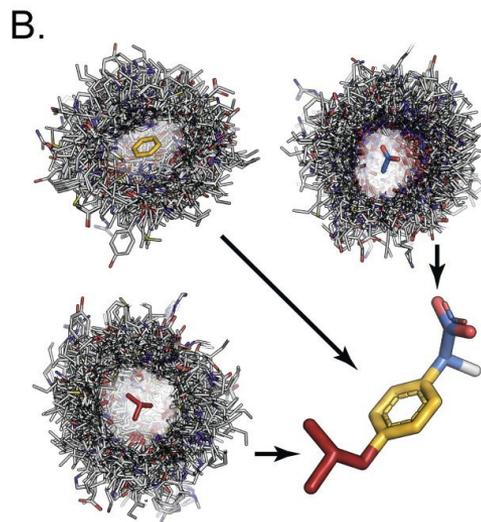
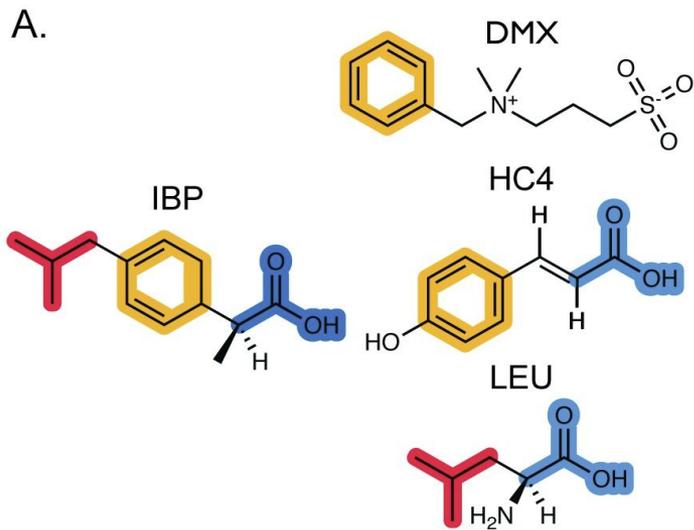


- E_{solv} Implicit solvation model penalizes buried polar atoms
- $E_{\text{BBtortion}}$ Backbone torsion preferences from main-chain potential
- E_{rotamer} Side-chain torsion angles from rotamer library
- E_{ref} Unfolded state reference energy for design
- E_{vdW} Lennard–Jones for attractive or repulsive interaction
- E_{hbond} Hydrogen bonding allows buried polar atoms
- E_{elec} Electrostatic interaction between charges
- E_{disulf} Disulfide bonds between cysteines

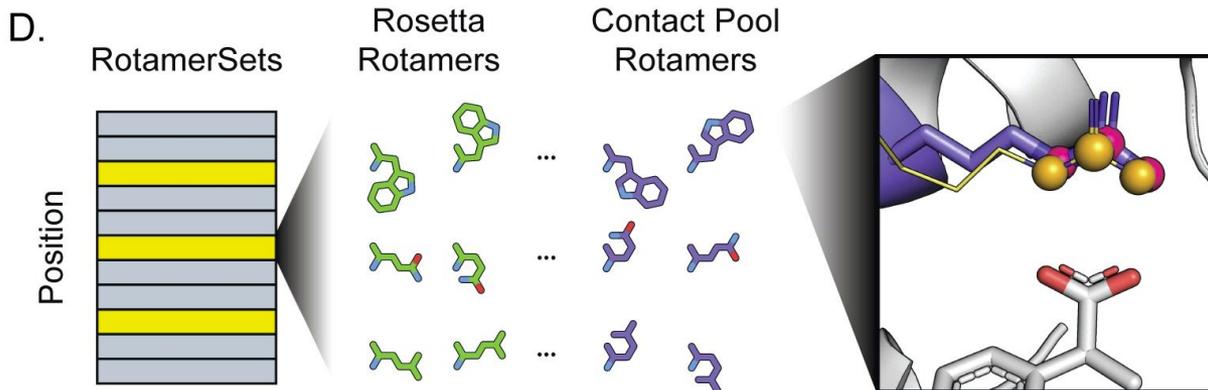
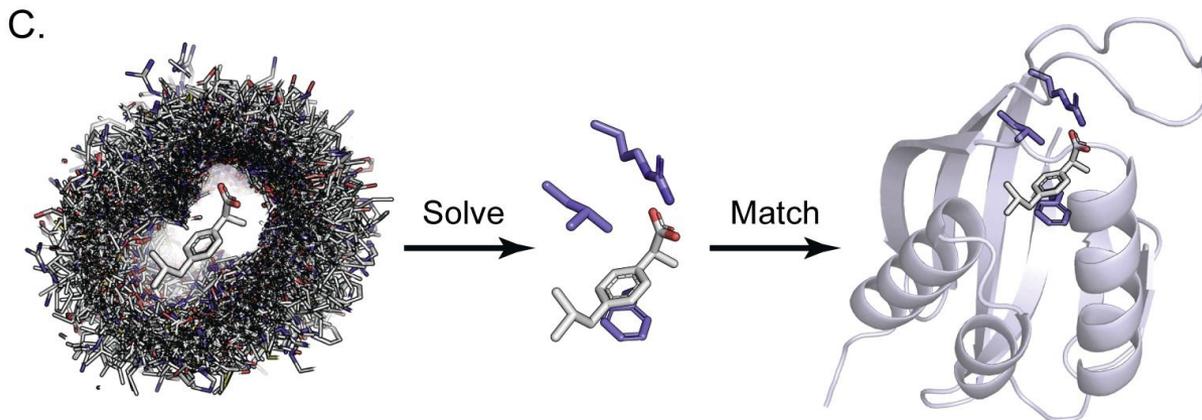
Rosetta: дизайн

Подбор набора субоптимальных последовательностей под заданную структуру

Чаще всего структура описывается через положение остова. Самый примитивный протокол работает с неизменяемым положением остова. Более продвинутые могут его случайным образом шевелить или полностью перестраивать.

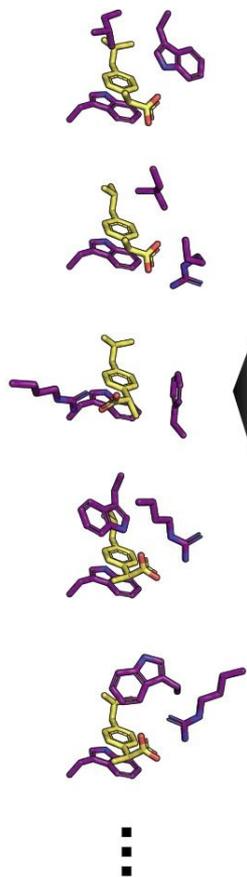


Дизайн связывания малых молекул

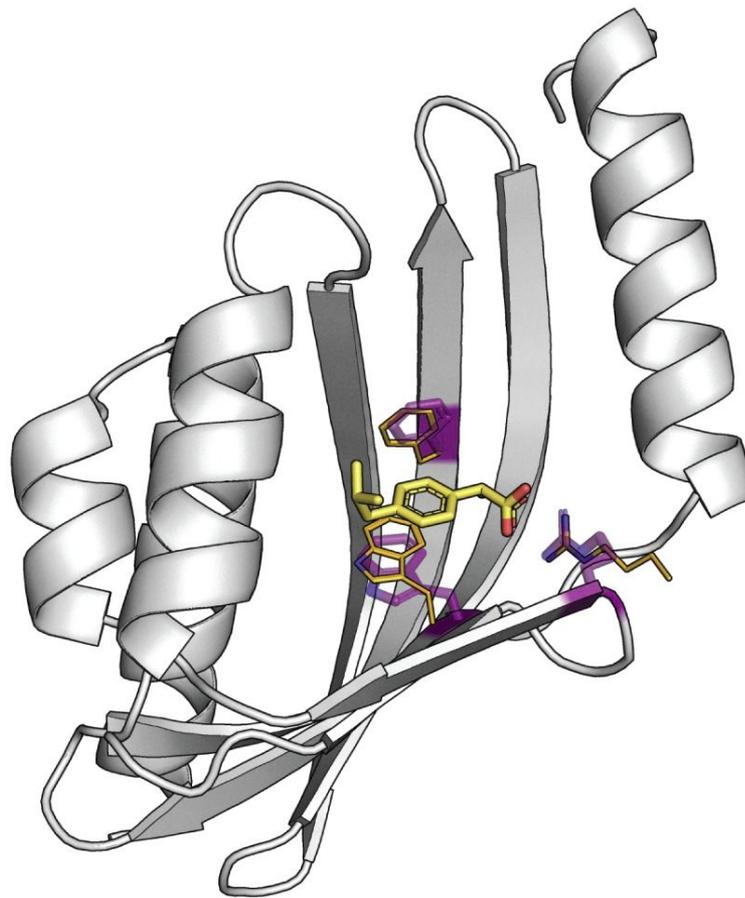


Дизайн связывания малых молекул

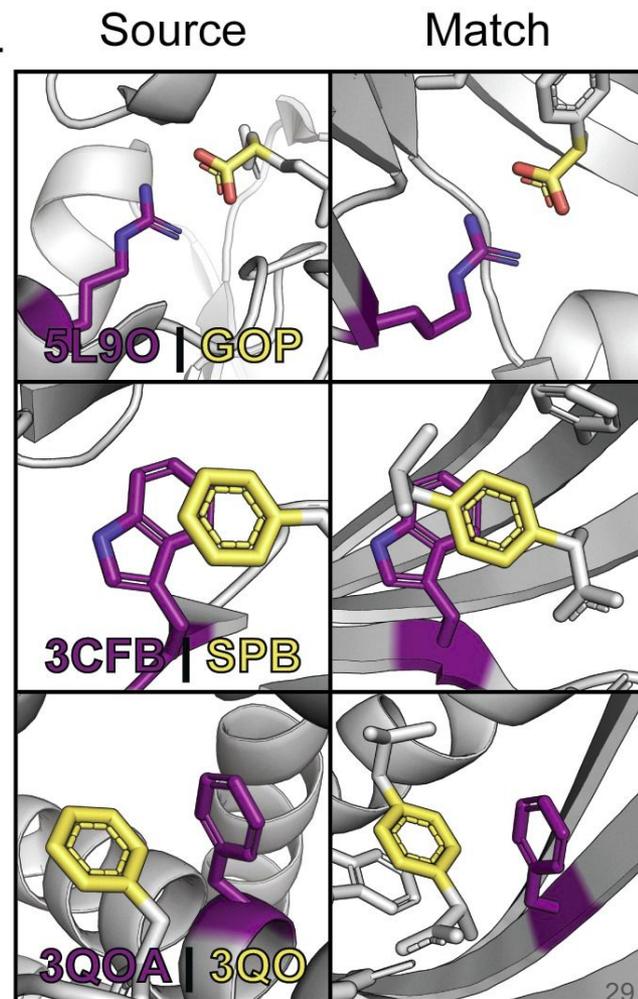
A.



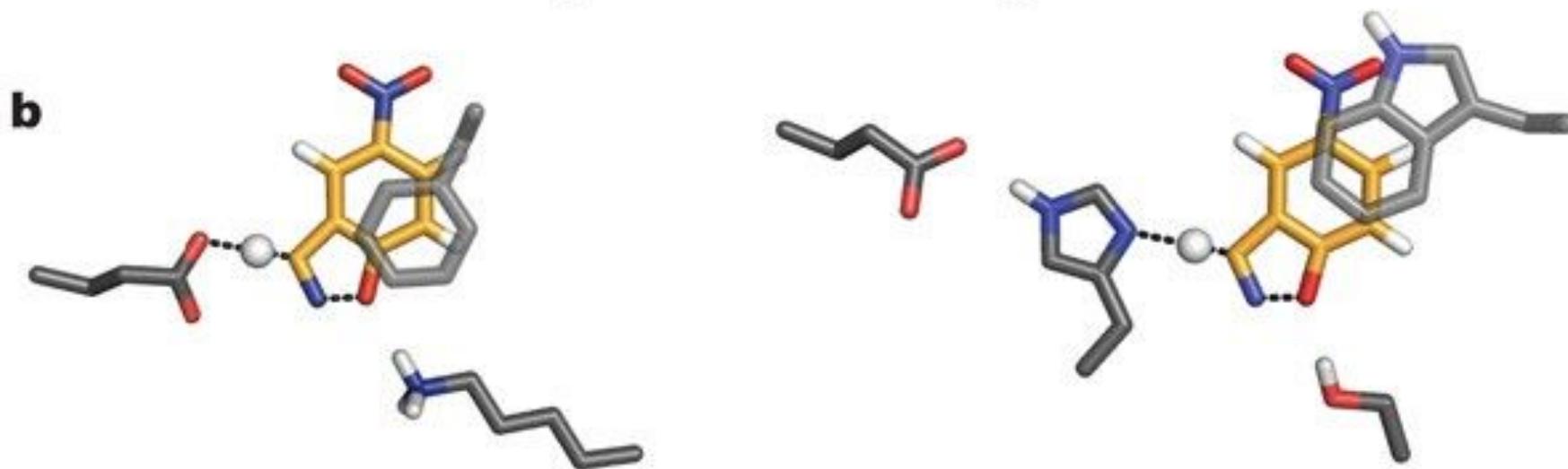
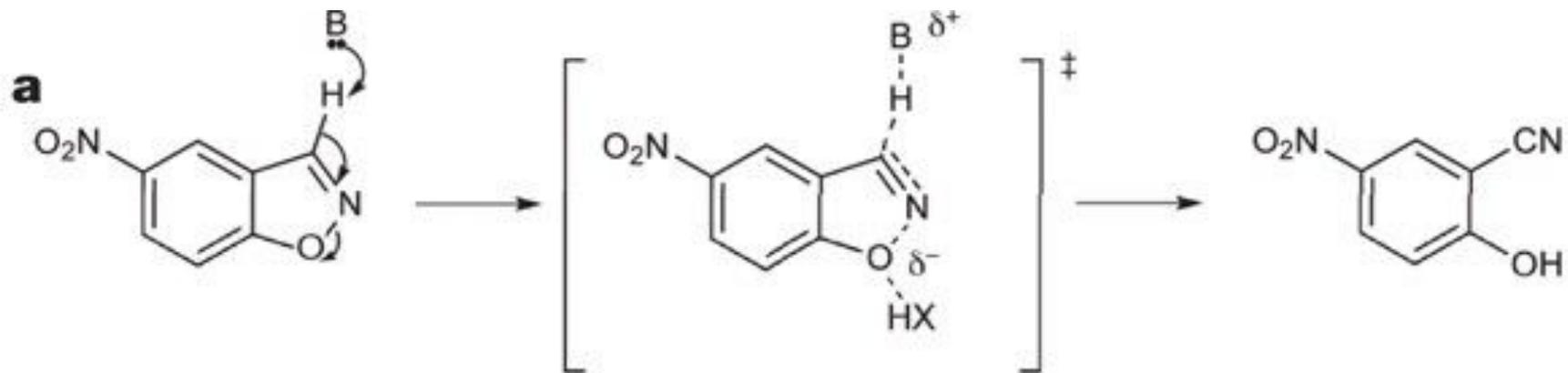
B.



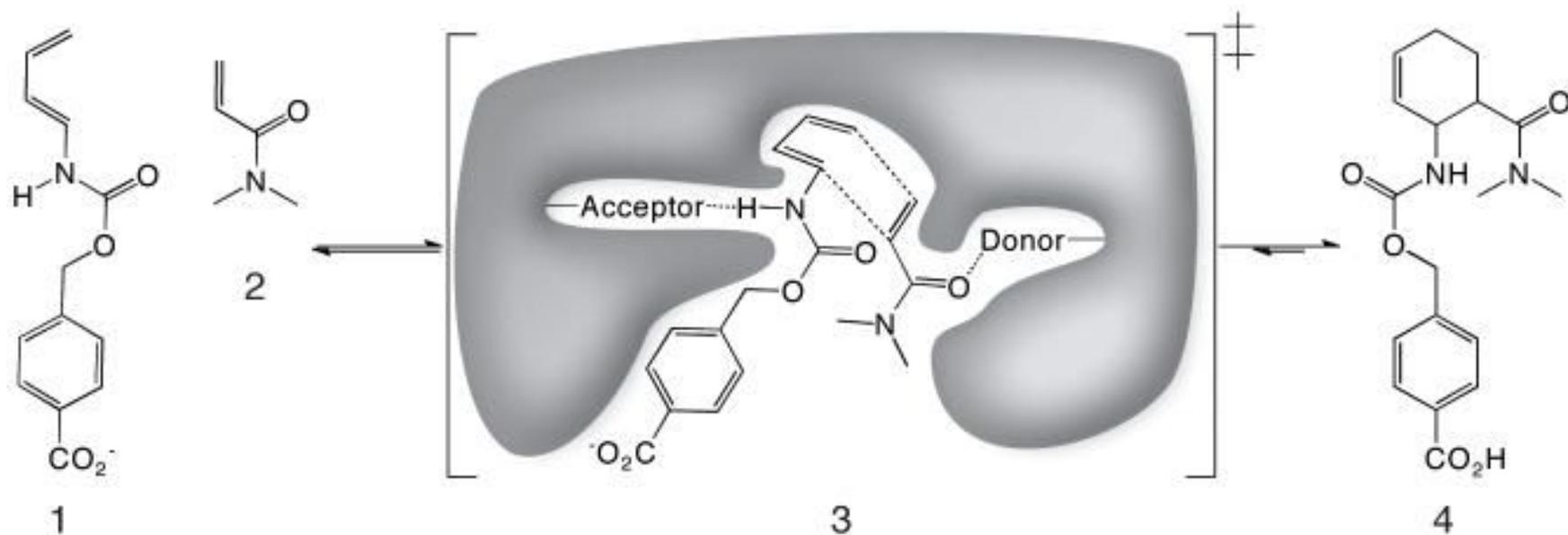
C.



Дизайн ферментов: стабилизация TS



Дизайн ферментов: стабилизация TS



Дизайн ферментов: стабилизация TS

Удалось сделать несколько де-ново ферментов (для реакций, не имеющих своего фермента в природе):

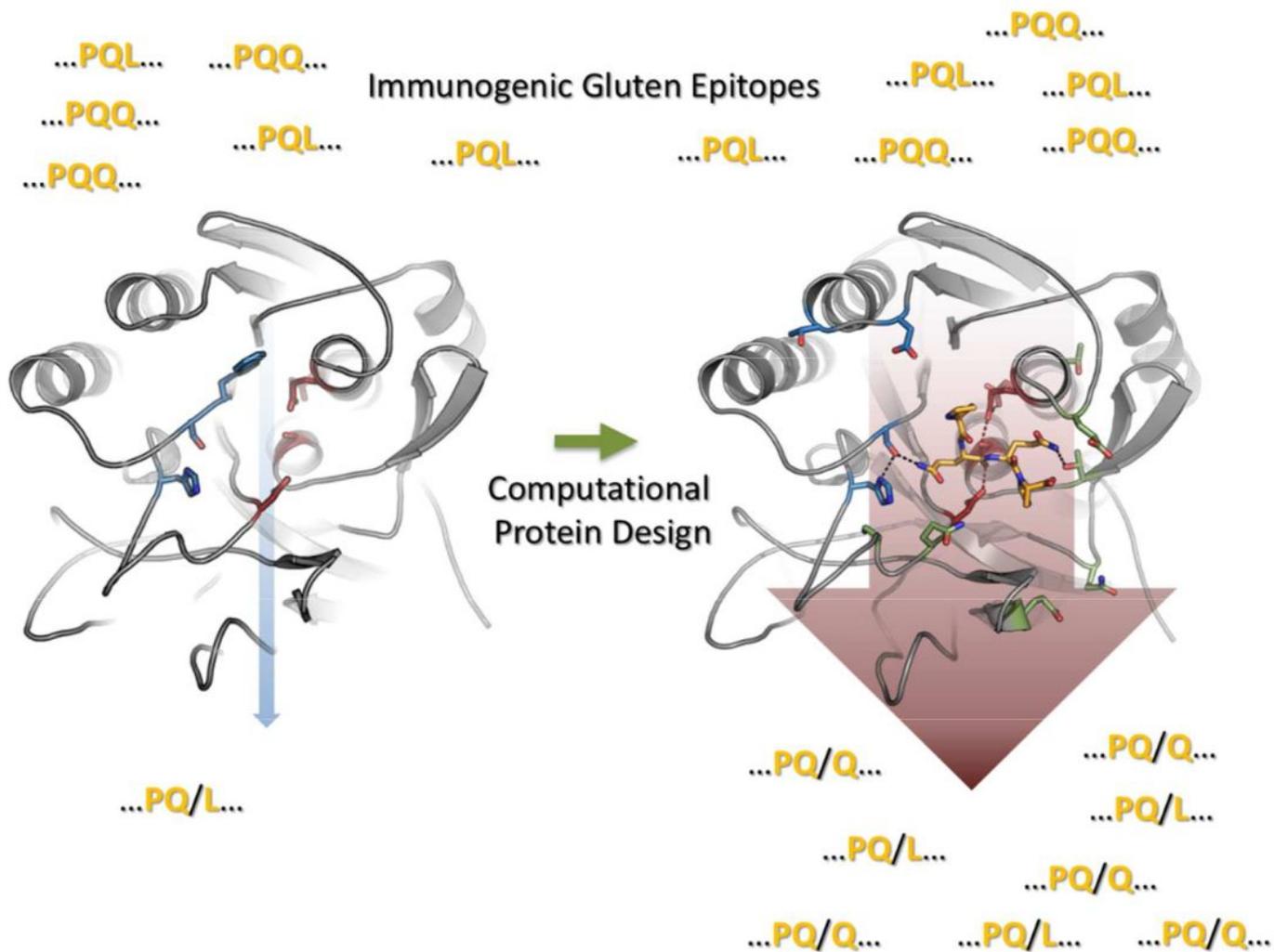
- 1) Только 1 стадия
- 2) Любое число > 0 = успех
- 3) Структура TS получена из КМ моделирования

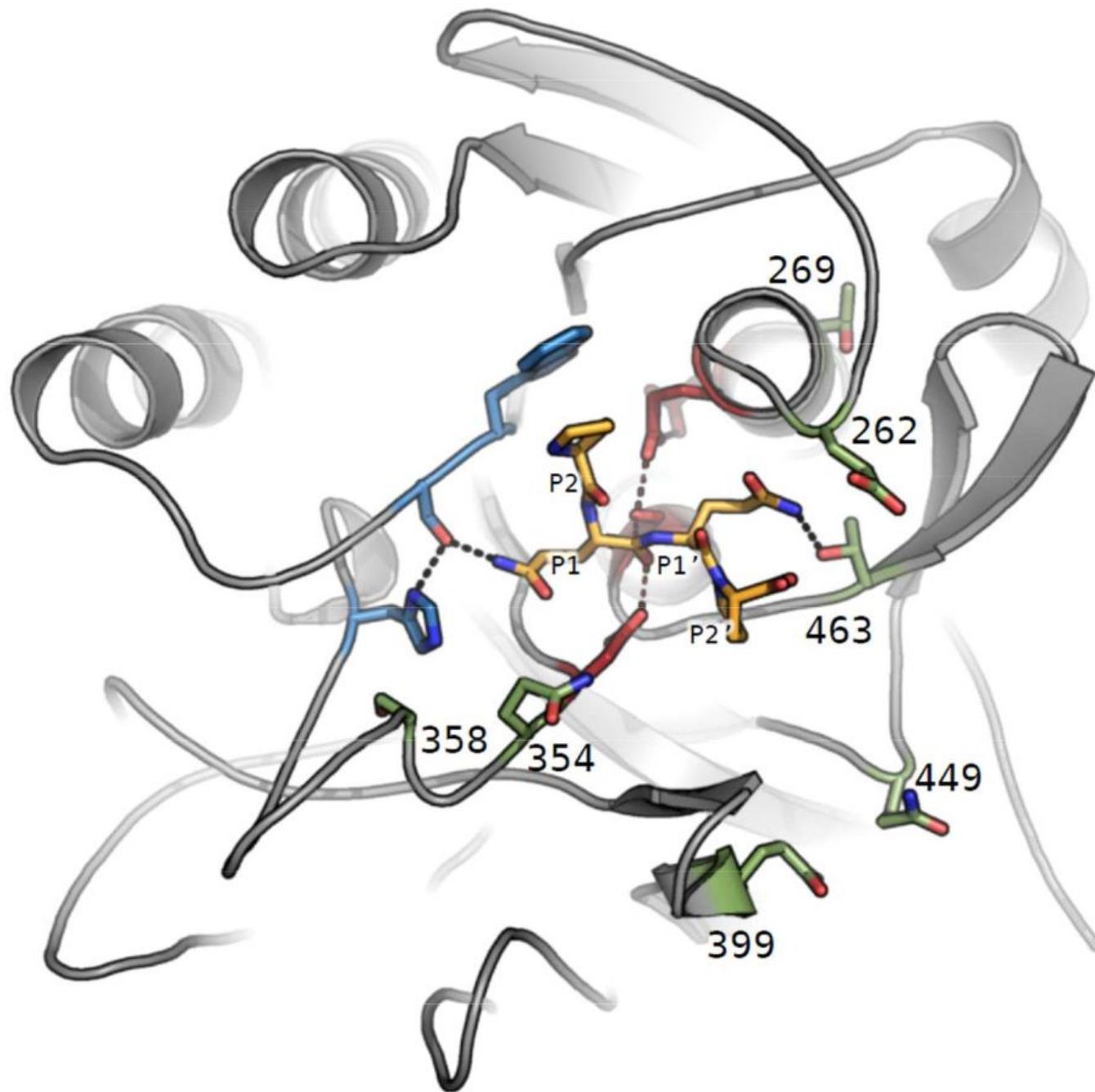
Не удастся улучшить уже имеющиеся ферменты

- 1) Ошибка в вычислении сора выше, чем ожидаемый результат
- 2) Имеющиеся ферменты это уже итог долгой эволюции – простые решения (например, единичные замены) исчерпаны. Требуется масштабный **редизайн**
- 3) Очень просто сломать

Не известны работы по успешному де-ново дизайну ферментов для сложных реакций (больше 1 стадии, наличие кофакторов/коферментов)

Оптимизация связывания





Collecting sequences

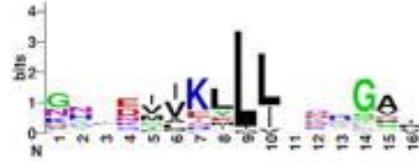
```

KYPQLIQELLQGSNFEA-
KNLQPKKLLKYGEDVTE-
PINDVLRVLLMHGVDVTE-
GHVEVEXHLLQGNAPLTK-
NTERAKYLLDNGADPTL-
WHZEVKMLLTIGADVTE-
DNEEPVSPVQVSSAVDE-
DNARDVLLVHLLGSDIET-

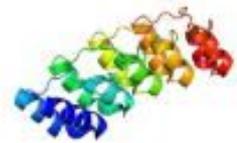
```



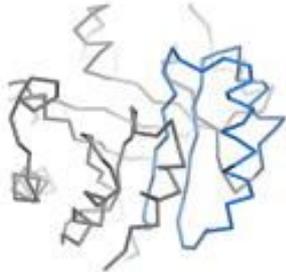
Calculating consensus



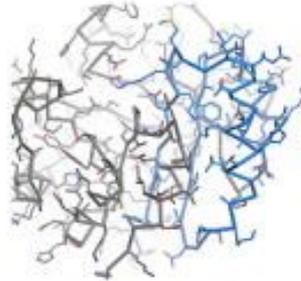
Consensus sequence



Backbone generation



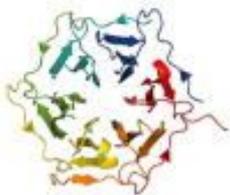
Side-chain placement



Repacking and minimization



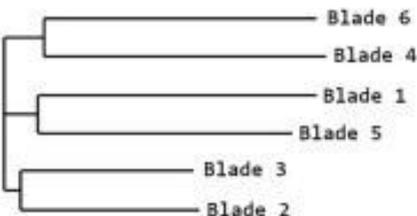
Pseudo-symmetric template



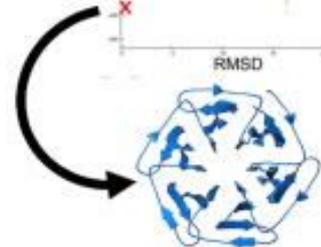
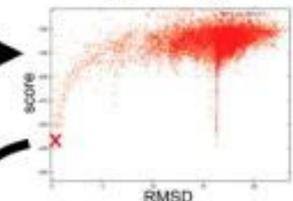
Symmetrical backbone



Ancestral sequence reconstruction



Sequence mapping

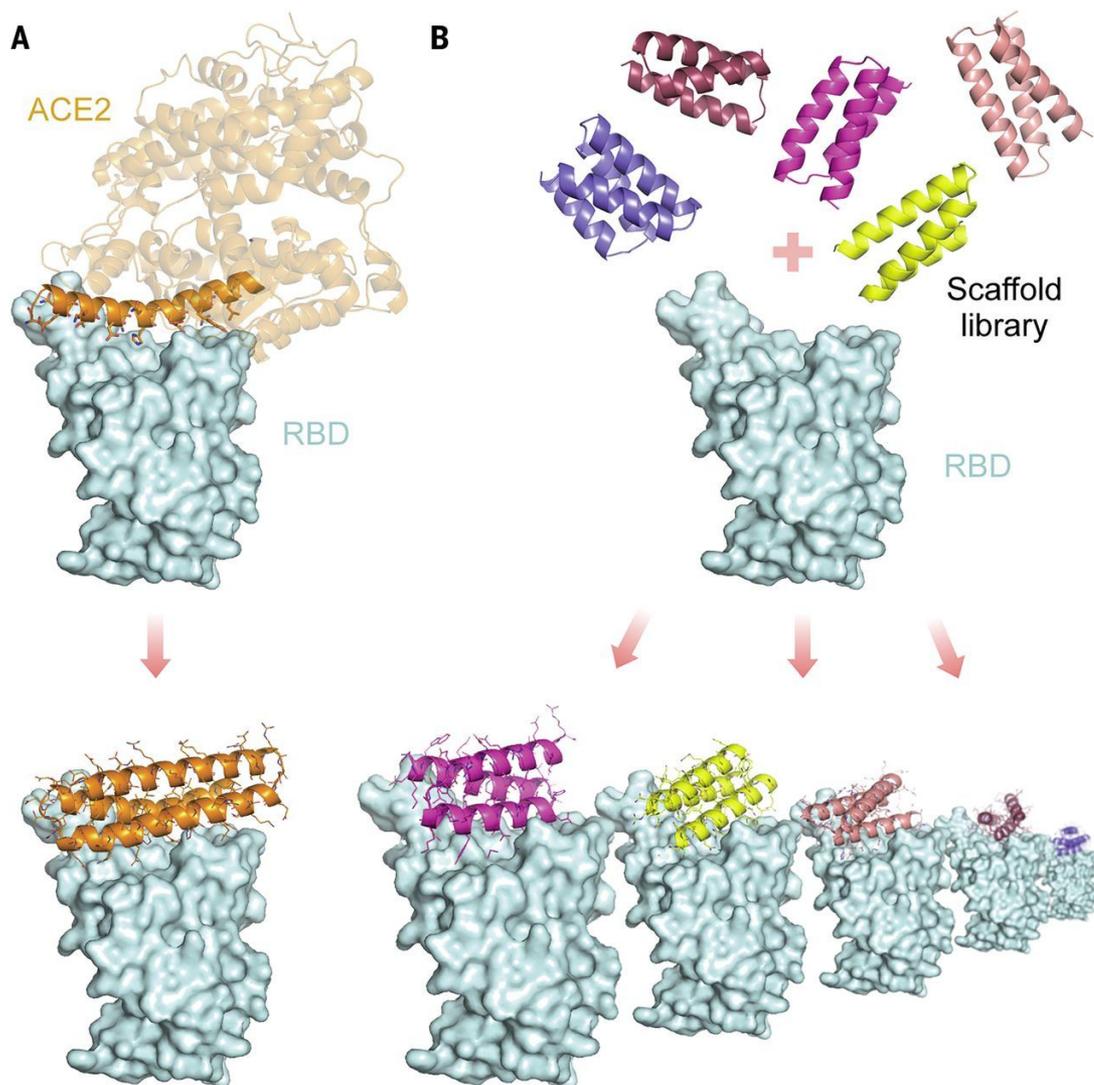


```

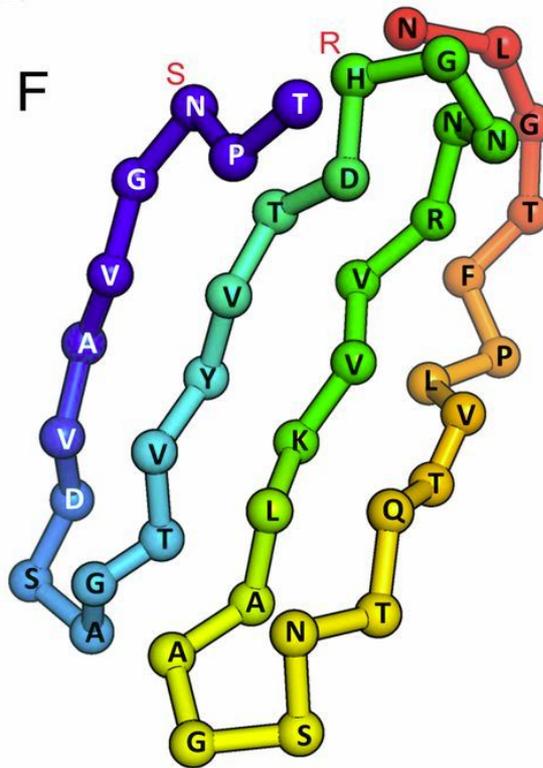
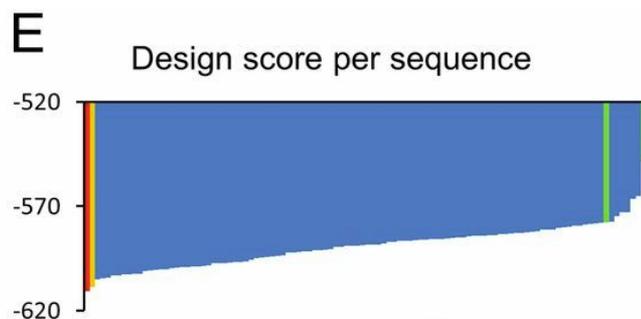
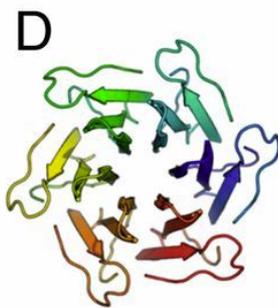
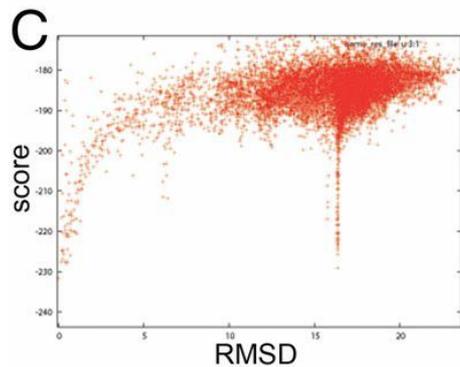
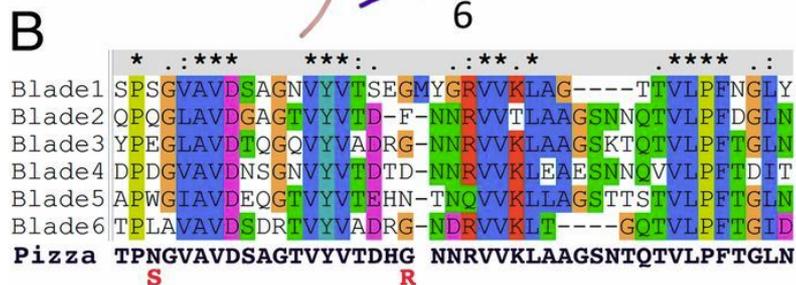
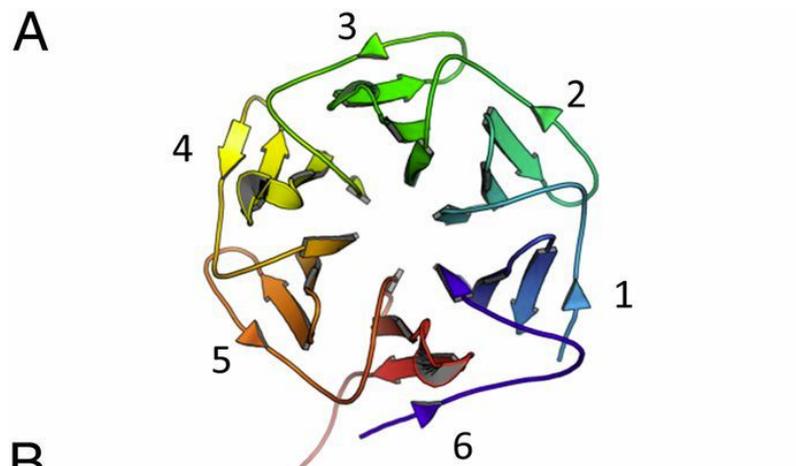
Blade 1 SPSGVAVDSAGNYYVTSEGMVGRVVKLATGSTGTVLVPFNGLY
Blade 6 TPLAVAVDSDRTVYVADRG-NDRVVKLT----GQTVLPFTGID
Blade 4 DPDGVAVDNSGNYYVTDTD-NNRVVKLEAESNNQVVLFFDIT
Blade 5 APWGIAVDEAGTVYYVTEHN-TNQVVKLLAGSTTSTVLPFTGLN
Blade 2 QPQGLAVDAGTVYYVTDFF--NNRVVTLAAGSNQTVLPFDGLN
Blade 3 YPEGLAVDTQGAIVYVADRG-NNRVVKLAAGSKTQTVLPFTGLN

```

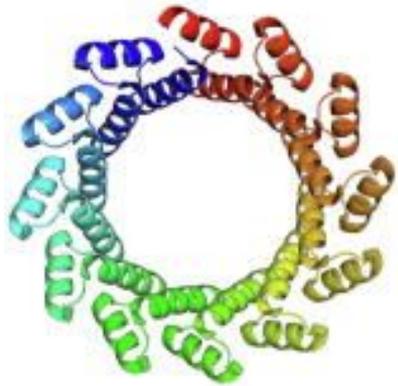
Backbone-based de novo дизайн



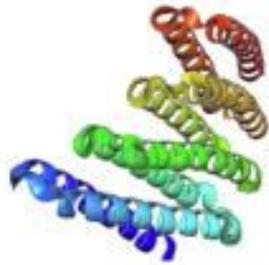
Оптимизация связывания комплекса



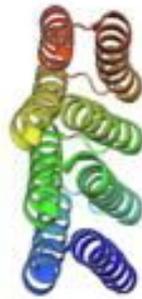
α -toroid



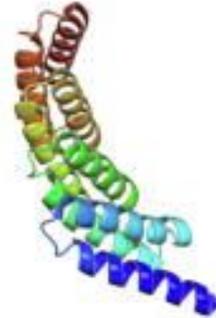
DHR64



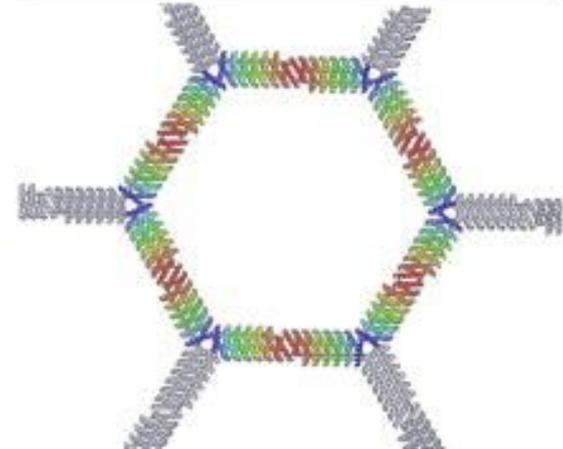
DHR10



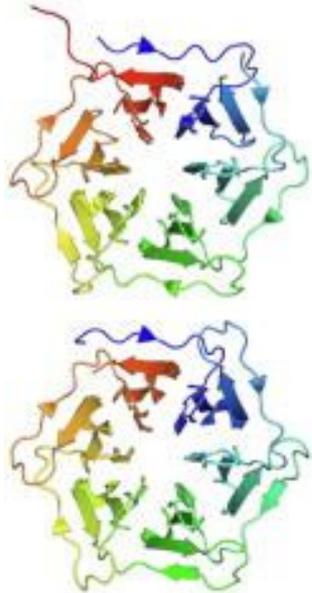
DHR53



Honey-comb lattice

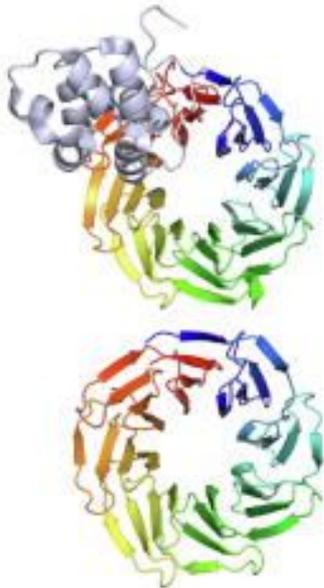


Pizza



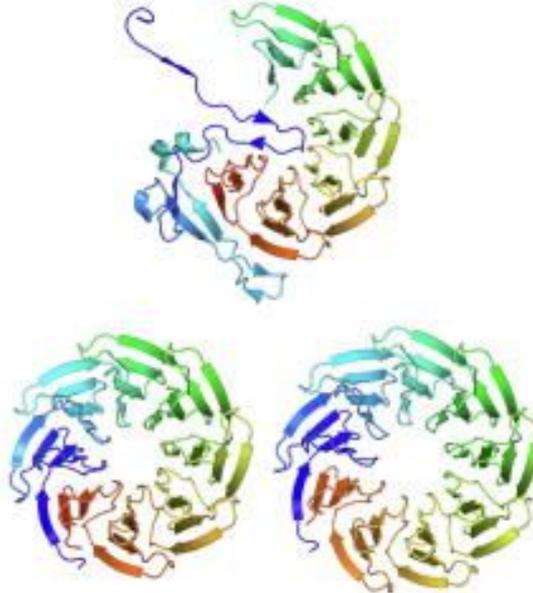
6

Ika



8

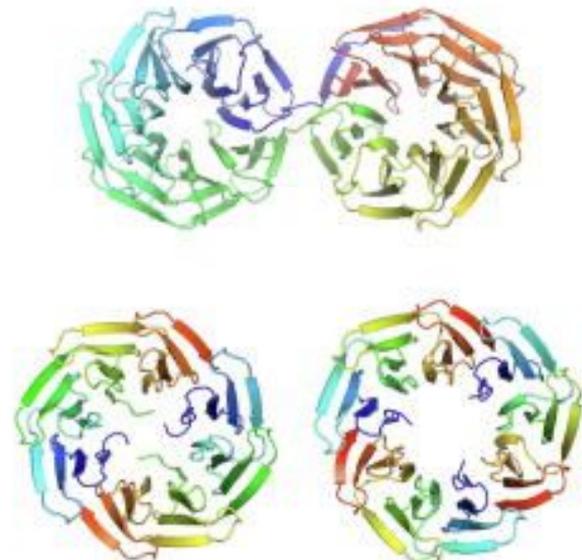
Cake



8

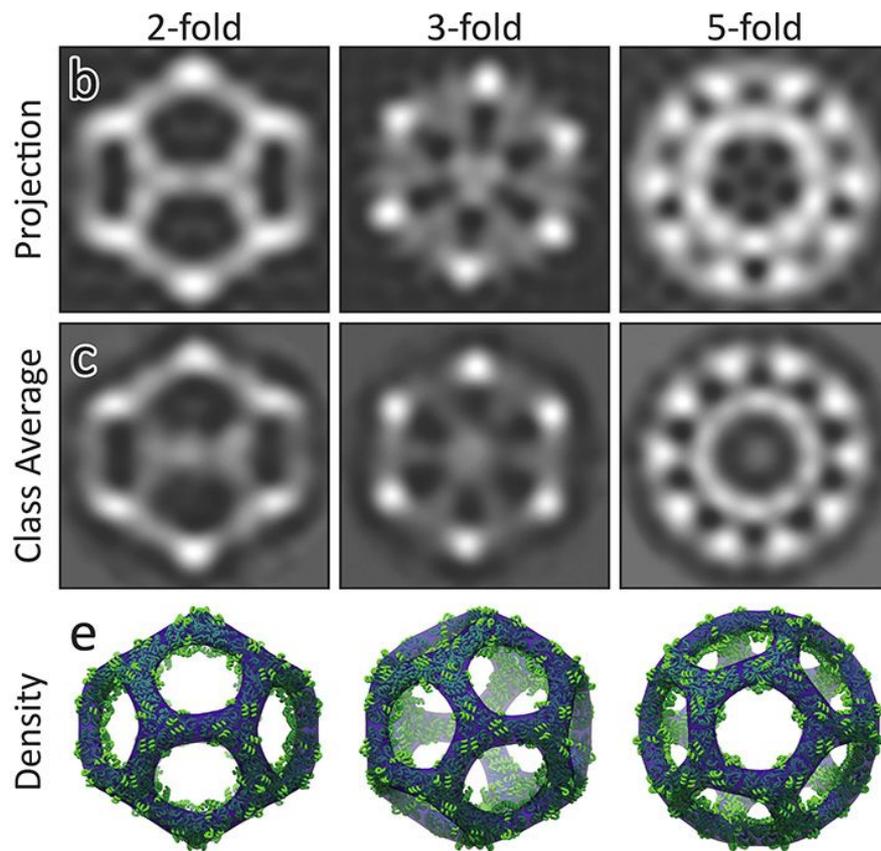
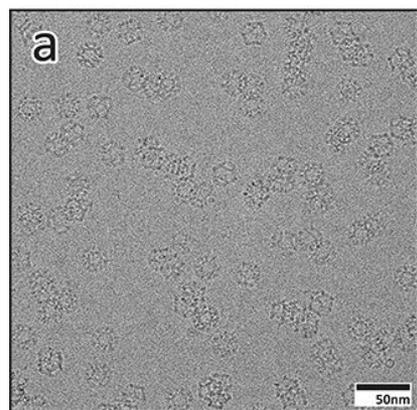
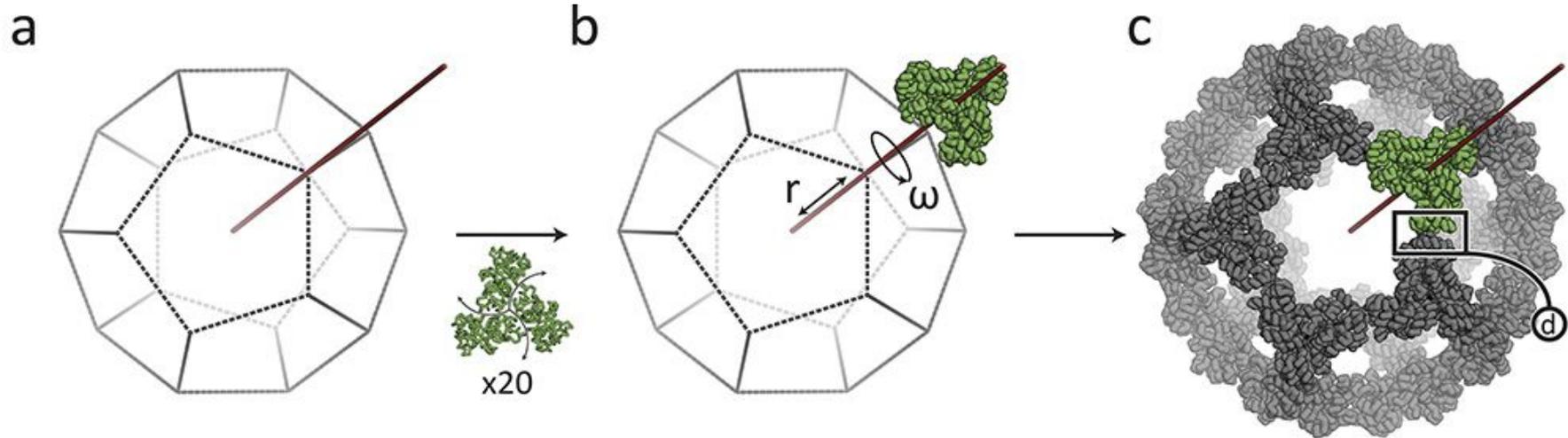
9

WRAP



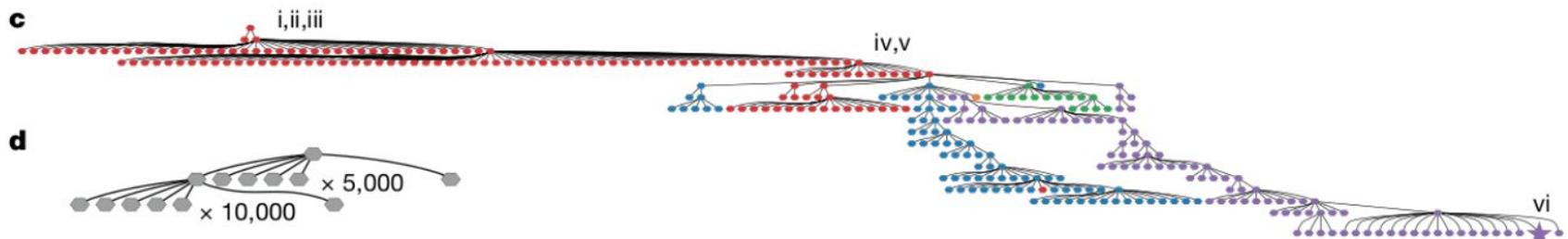
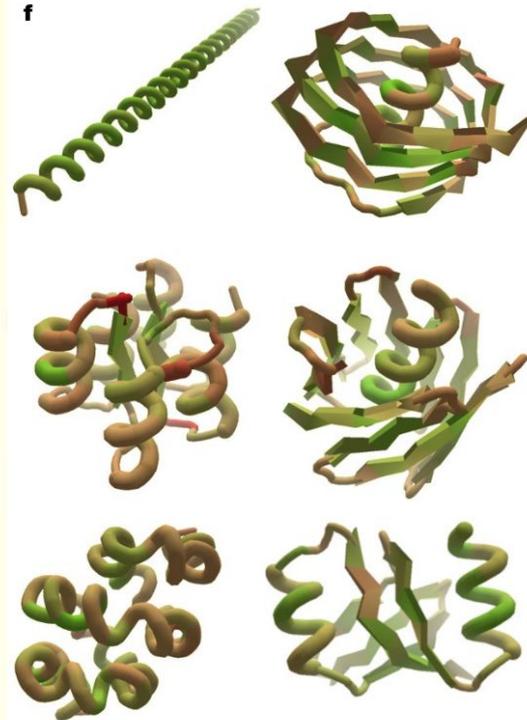
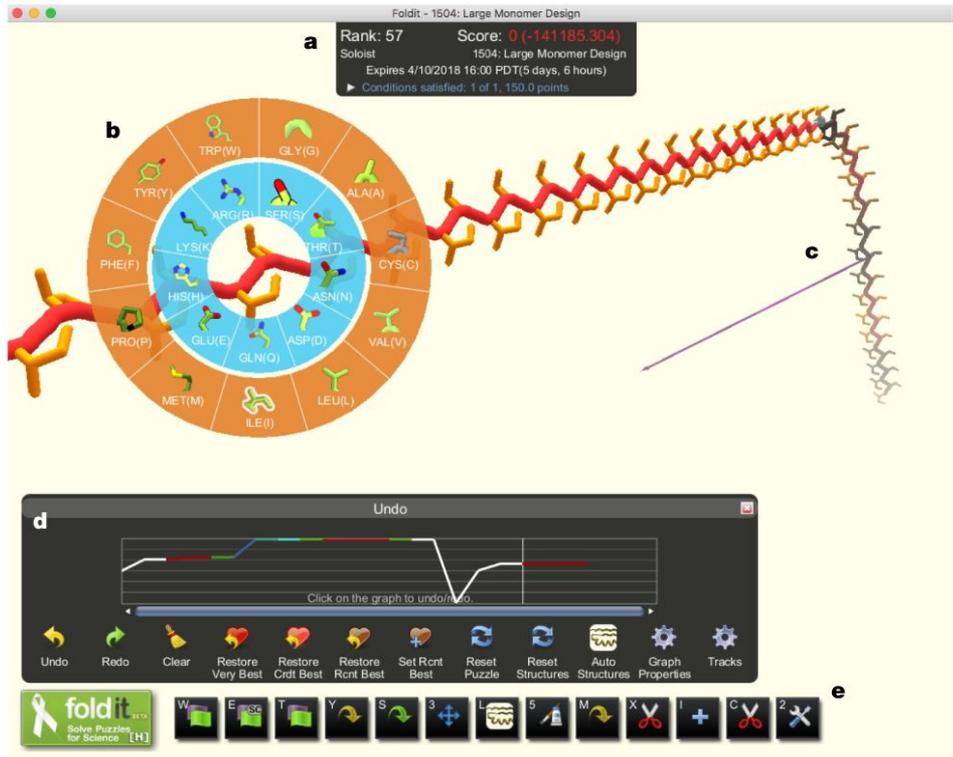
8

9



FoldIt!

Brian Koepnick, Jeff Flatten, Tamir Husain, Alex Ford, Daniel-Adriano Silva, Matthew J. Bick, Aaron Bauer, Gaohua Liu, Yojiro Ishida, Alexander Boykov, Roger D. Estep, Susan Kleinfelter, Toke Nørgård-Solano, Linda Wei, Foldit Players, Gaetano T. Montelione, Frank DiMaio, Zoran Popović, Firas Khatib, Seth Cooper and David Baker. **De novo protein design by citizen scientists** *Nature* (2019).



FoldIt! Education

Josh A. Miller, Firas Khatib, Seth Cooper, Scott Horowitz. **Introducing FoldIt Education Mode** *Nature Structural & Molecular Biology* (2020).

The screenshot displays the FoldIt! Education interface. At the top left, a button labeled "Pull Mode" is visible. In the top center, a dark box shows the "Score: 7717 of 8550". At the top right, a button labeled "Wiggle!" is present. The main area features a 3D protein structure with various colored segments (blue, orange, green, brown). A green tooltip box is overlaid on the left side of the protein, containing a graph with a wavy line and two red arrows pointing to a local minimum. Below the graph, the text reads: "WIGGLE uses a method called 'minimization' to make very small changes that reduce the energy of your protein (increasing your score)." At the bottom of the tooltip are "Back" and "Next" buttons. At the very bottom of the screen, a navigation bar includes "Actions", "Undo", "Modes", "View", and "Menu" buttons.

FoldIt! Education

Structure Mode

Score: 7105 of 8000

Idealizing Structure Angles

Ramachandran Map

This is the Ramachandran Map, or Rama Map. It shows the secondary structure of the protein in terms of the two different angles that make up the protein backbone, phi (x-axis) and psi (y-axis).

Back Next

ARG 7

ARGININE

The screenshot displays the FoldIt! Education interface. At the top left, a 'Structure Mode' button is visible. In the top center, a score of '7105 of 8000' is shown. At the top right, there is a button for 'Idealizing Structure Angles'. The main area features a 3D protein structure on the left, a 'Ramachandran Map' window on the right, and a green tooltip in the center. The tooltip explains that the Ramachandran Map shows the secondary structure of the protein in terms of phi (x-axis) and psi (y-axis) angles. The Ramachandran Map window shows a plot with a red region in the bottom-left, a green region in the top-right, and a blue region in the top-left. A point labeled 'ARG 7' is located in the top-right region, and 'ARGININE' is labeled at the bottom right of the plot. A small 3D model of a protein backbone is shown in the top right corner of the Ramachandran Map window. At the bottom of the interface, there is a navigation bar with buttons for 'Actions', 'Undo', 'Modes', 'View', and 'Menu'.

FoldIt! Standalone

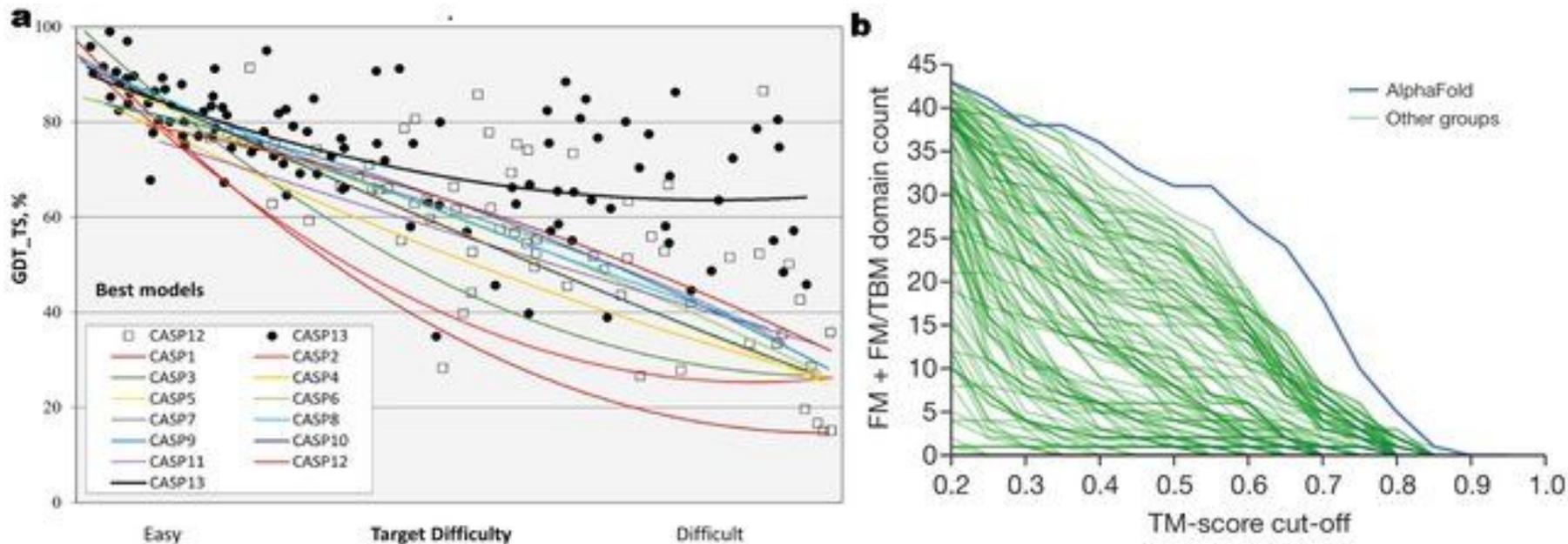
Robert Kleffner, Jeff Flatten, Andrew Leaver-Fay, David Baker, Justin B. Siegel, Firas Khatib and Seth Cooper. **FoldIt Standalone: a video game-derived protein structure manipulation interface using Rosetta.** *Bioinformatics* (2017).

CASP

Community Wide Experiment on the **Critical Assessment** of Techniques for Protein Structure Prediction

- The **High Accuracy Modeling** category will include domains where the majority of submitted models are of sufficient accuracy for detailed analysis. This category replaces the previous Template Based Modeling category.
- The **Topology** category (formerly Free Modeling) will assess domains where submitted models are of relatively low accuracy.
- The **Contact and Distance Prediction** category will assess the ability of methods to predict contacts and inter-residue distances.
- The **Refinement** category will analyze success in refining models beyond the accuracy obtained in the initial submissions. For each target, one of the best initial models will be selected, and reissued as the starting structure for refinement.
- The **Assembly** category will assess how well current methods can determine domain-domain, subunit-subunit, and protein-protein interactions. As in CASPs 11-13, we hope to work closely with CAPRI in this category.
- The **Accuracy Estimation** category will assess the ability to provide useful accuracy estimates for the overall accuracy of models and at the domain and residue level.
- The **Data Assisted** category will assess how much the accuracy of models is improved by the addition of sparse data. Targets for which such data are available will be re-released after initial data independent models have been collected, together with the available data. Data types are expected to include crosslinking data and SAXS.
- The **Biological Relevance** category will assess models on the basis of how well they provide answers to biological questions. Target providers will be asked to say what questions prompted the determination of the experimental structure. The usefulness of the models in answering those questions will be compared with the that of the experimental structures.

CASP

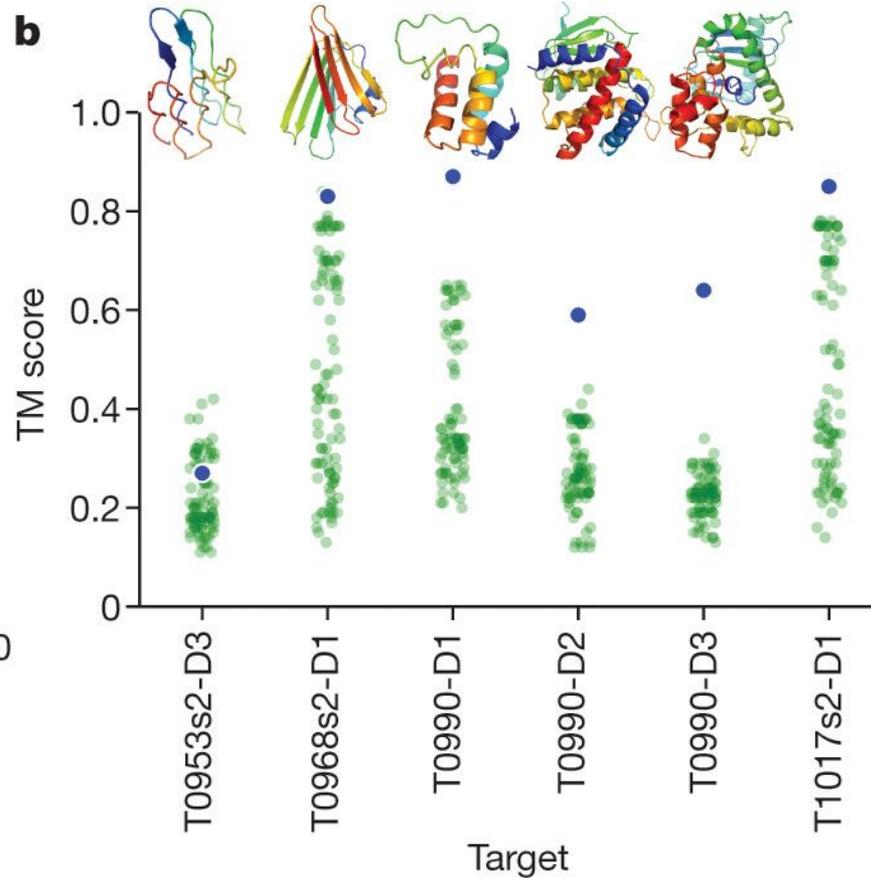
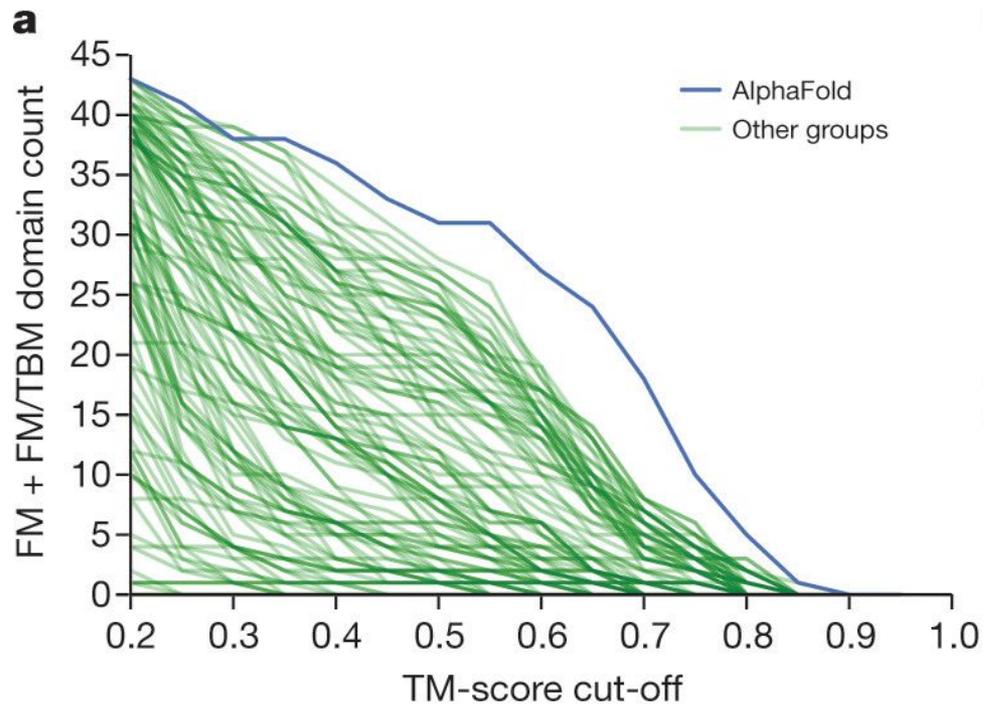


Соревнование, проводится раз в 2 года

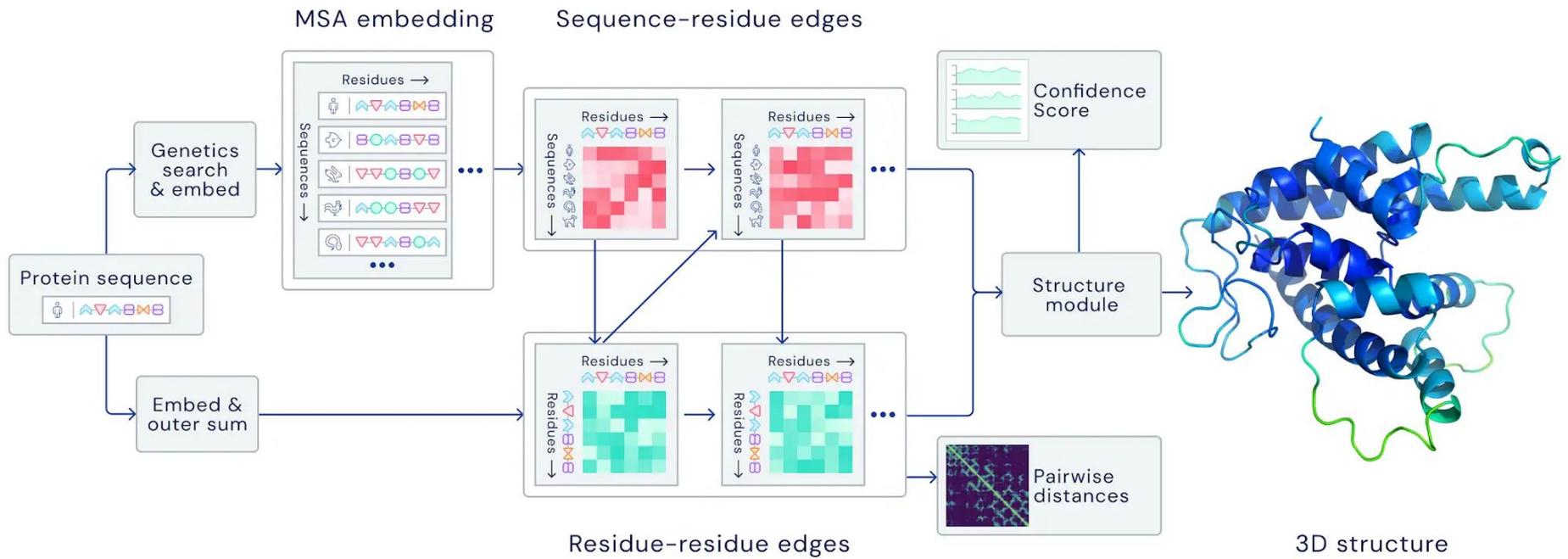
Авторы структур придерживаются их публикации и отправляют как задачи на CASP

Группы соревнуются в предсказании структуры по последовательности

AlphaFold



AlphaFold2



AlphaFold и (не)революция

Не получаем никакой информации о процессе фолдинга

Что делать с метаморфами?

Что делать с белками, принимающими стабильную укладку только при связывании с партнером?

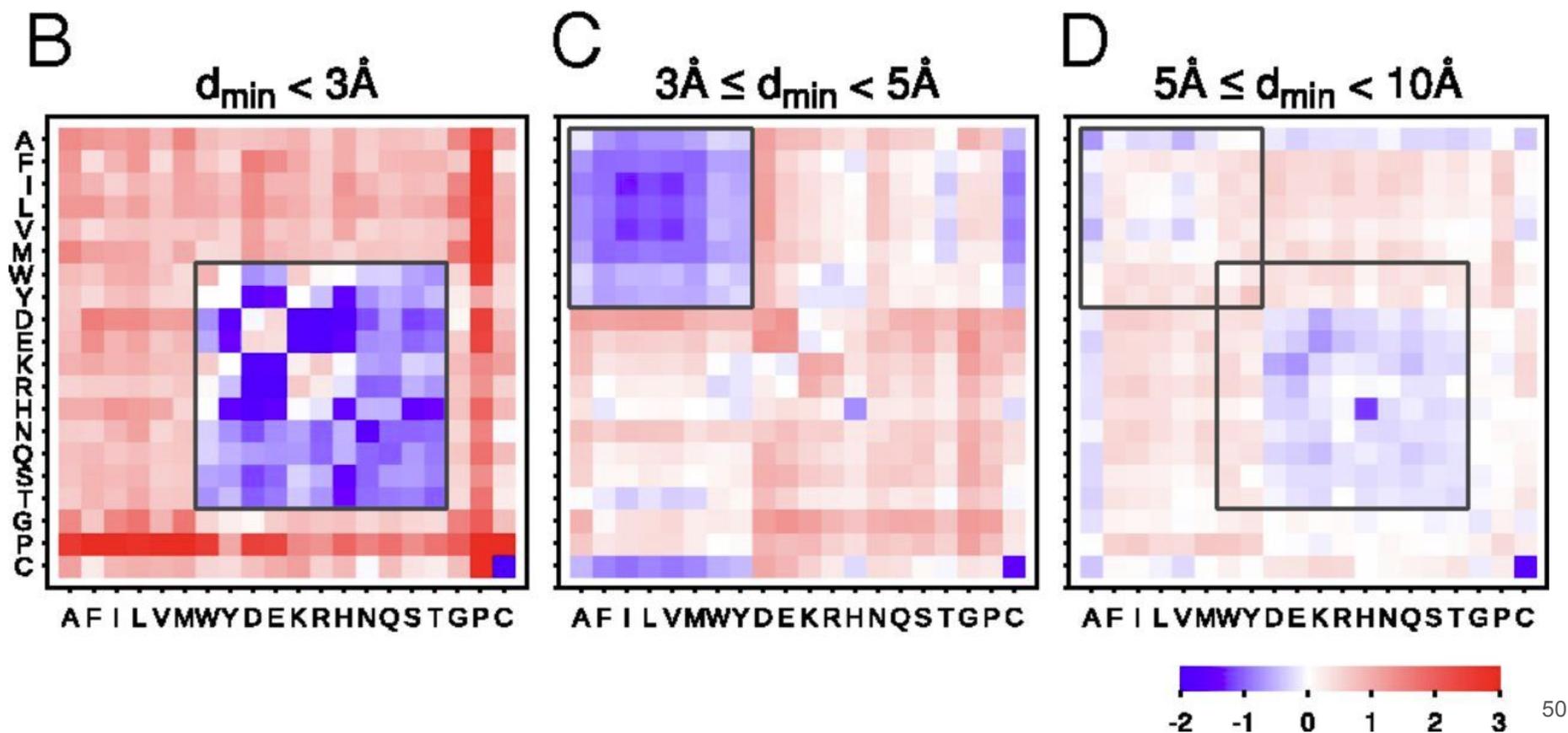
Что делать с коферментами и кофакторами?

Что делать с подвижностью?

Биас в сторону уже изученных белков

Почему помогает коэволюция?

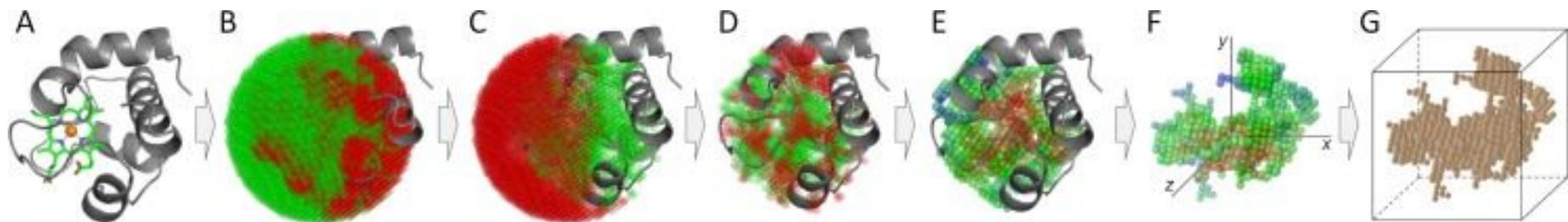
91% коэволюционирующих остатков непосредственно контактируют хотя бы в одном представителе семейства.



ML в структурной биологии

Как можно использовать структуры для ML?

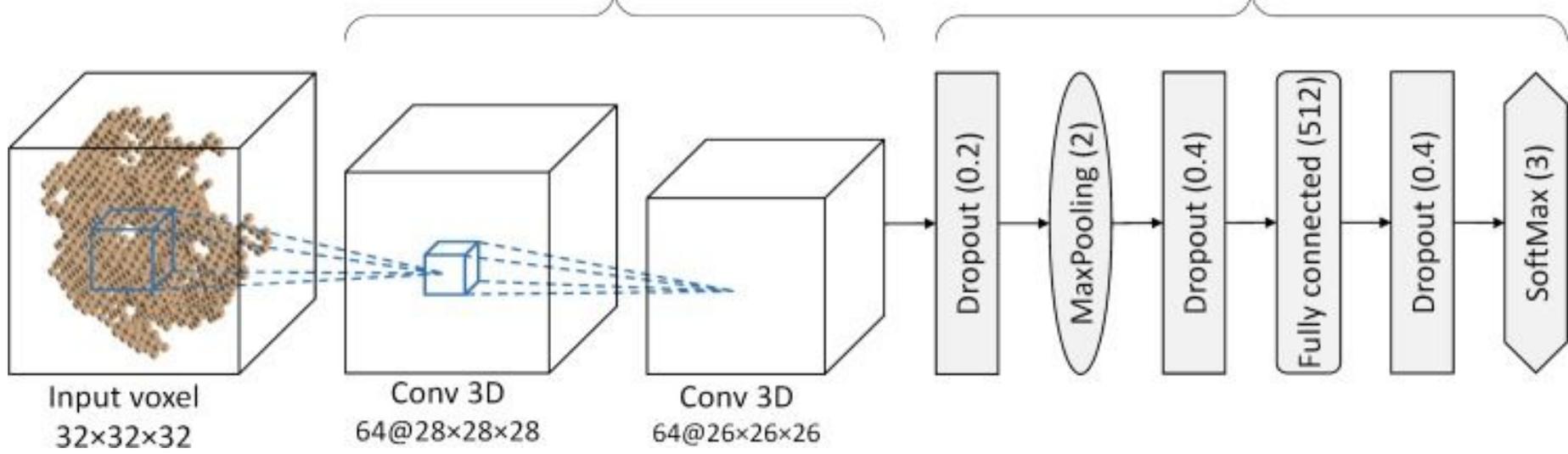
- Сопоставить фичи позициям в последовательности (1D дескриптор)
- Сопоставить фичи парам позиций: матрица расстояний (2D дескриптор)
- Картировать фичи на поверхность (2D дескриптор)
- Работать со структурой как таковой (вокселизация: 3D дескриптор)



A

B

C



Input voxel
32×32×32

Conv 3D
64@28×28×28

Conv 3D
64@26×26×26

Dropout (0.2)

MaxPooling (2)

Dropout (0.4)

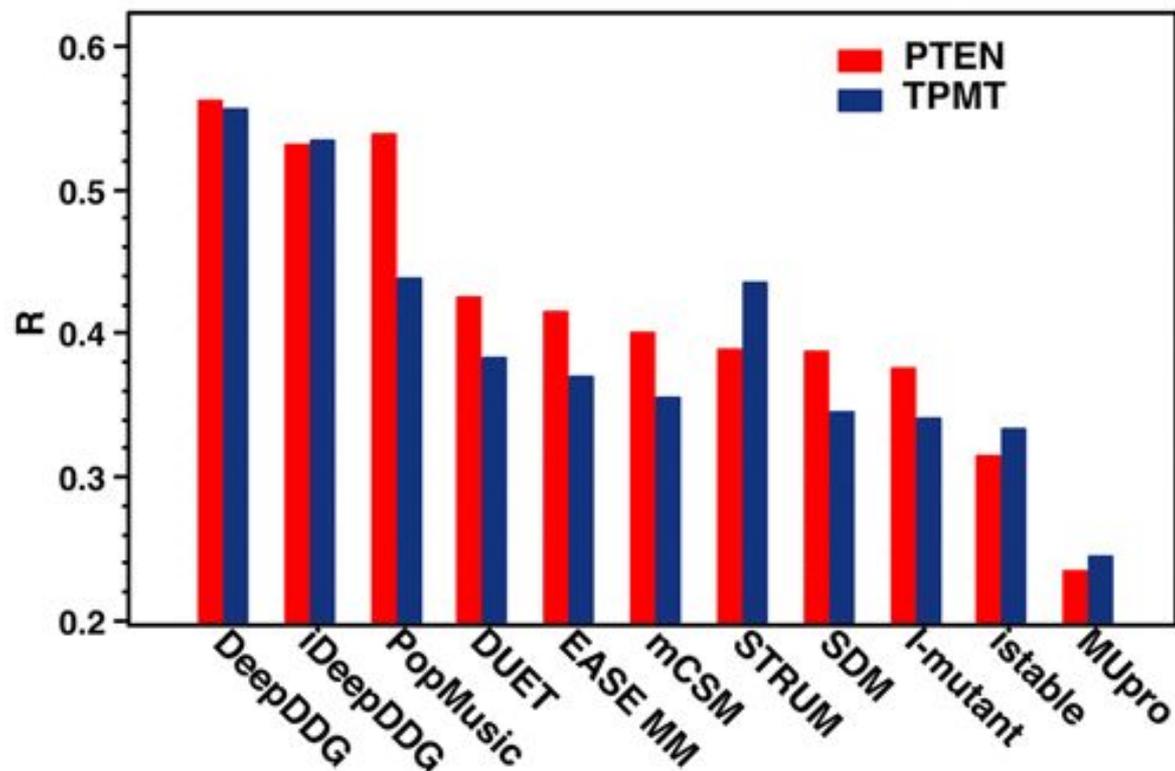
Fully connected (512)

Dropout (0.4)

SoftMax (3)

Задачи ML: ddG

Предсказание изменения в энергии фолдинга при внесении замены. Зачем: инструмент для стабилизации промышленных ферментов. Проблема: выборки.



Задачи: Gene Ontology

