



Научно-технологический
университет

Сириус

Структурная биоинформатика | Лекция 1

Вводная лекция. Задачи и методы

Александр Злобин

Структурная биоинформатика

В узком смысле это набор вычислительных методологий для работы с информацией о структурах макромолекул

Но в данном курсе нас также будут волновать:

- Как мы получили эту информацию? Можем ли мы ей верить?
- Почему эти структуры такие? Какими структуры вообще могут быть, а какими не могут?
- Зачем мы вообще получали структуру? Как связана структура и функция?
- Какие типичные задачи встают перед исследователем?
Можем ли мы спроектировать структуру?
- Насколько релевантна информация, полученная из статичной структуры?
- Как эволюционируют структуры?

Как мы получили информацию о структуре?

Физико-химические методы:

- Рентгеноструктурный анализ (РСА)
- Криоэлектронная микроскопия (КриоЭМ)
- Ядерный магнитный резонанс (ЯМР)
- Малоугловое рассеяние рентгеновских лучей (SAXS)

■ Ни один метод не дает точно описать структуру

Каждый метод способен дать набор ограничений на то, как структура должна выглядеть:

- РСА и КриоЭМ – как в пространстве локализована электронная плотность
- ЯМР – на каких расстояниях находятся друг от друга химические группы
- SAXS – какого размера и формы частица

Превращение набора ограничений в набор координат атомов – сложная вычислительная задача.

Как мы получили информацию о структуре?

Превращение набора ограничений в набор координат атомов – сложная вычислительная задача.

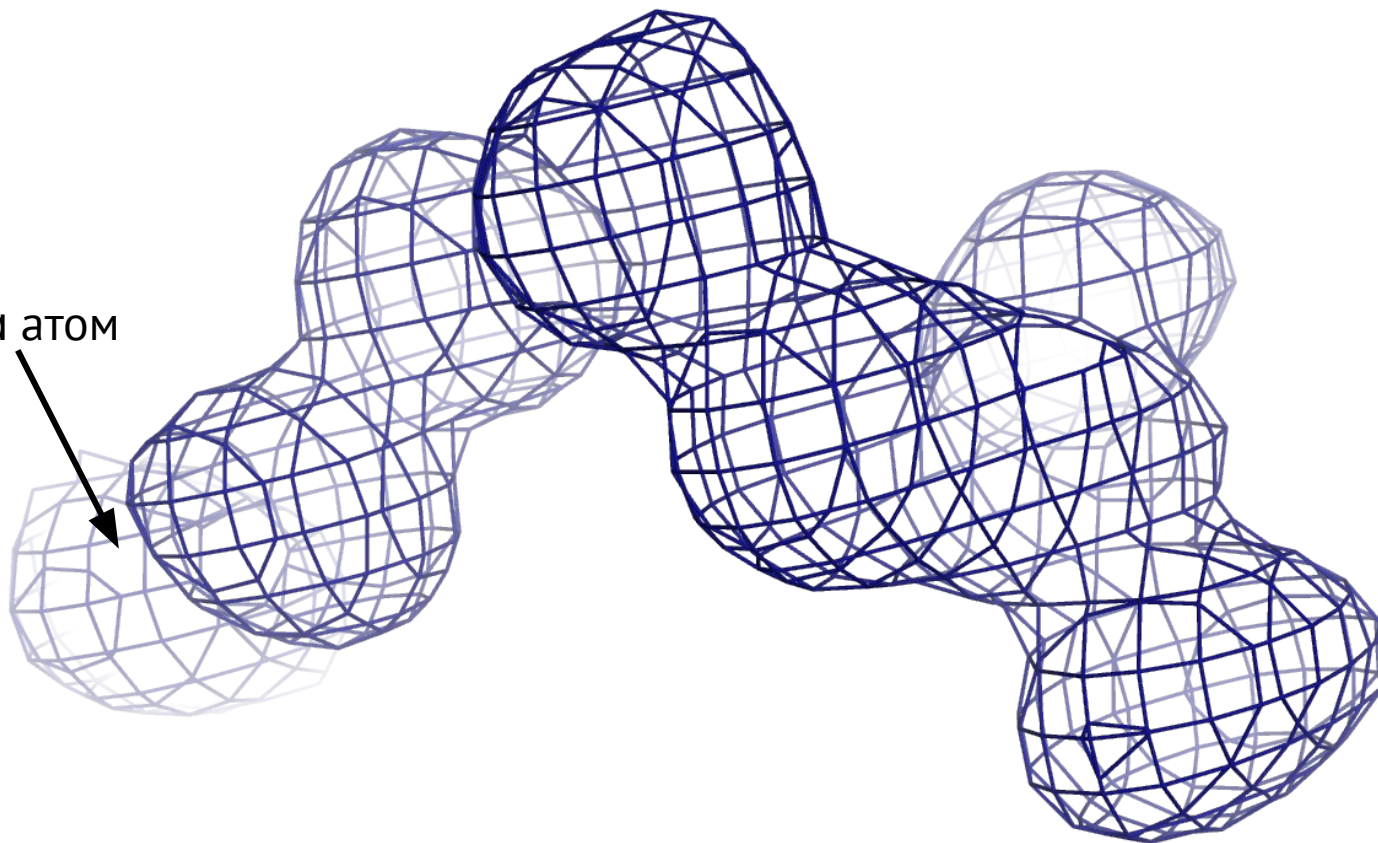
Однако кроме ограничений из эксперимента, у нас есть дополнительные ограничения из природы:

- Число электронов у атомов элементов
- Длины связей
- Углы между связями
- Допустимые значения торсионных углов
- Параметры нековалентных взаимодействий

Все это неизменно для любой структуры любой макромолекулы

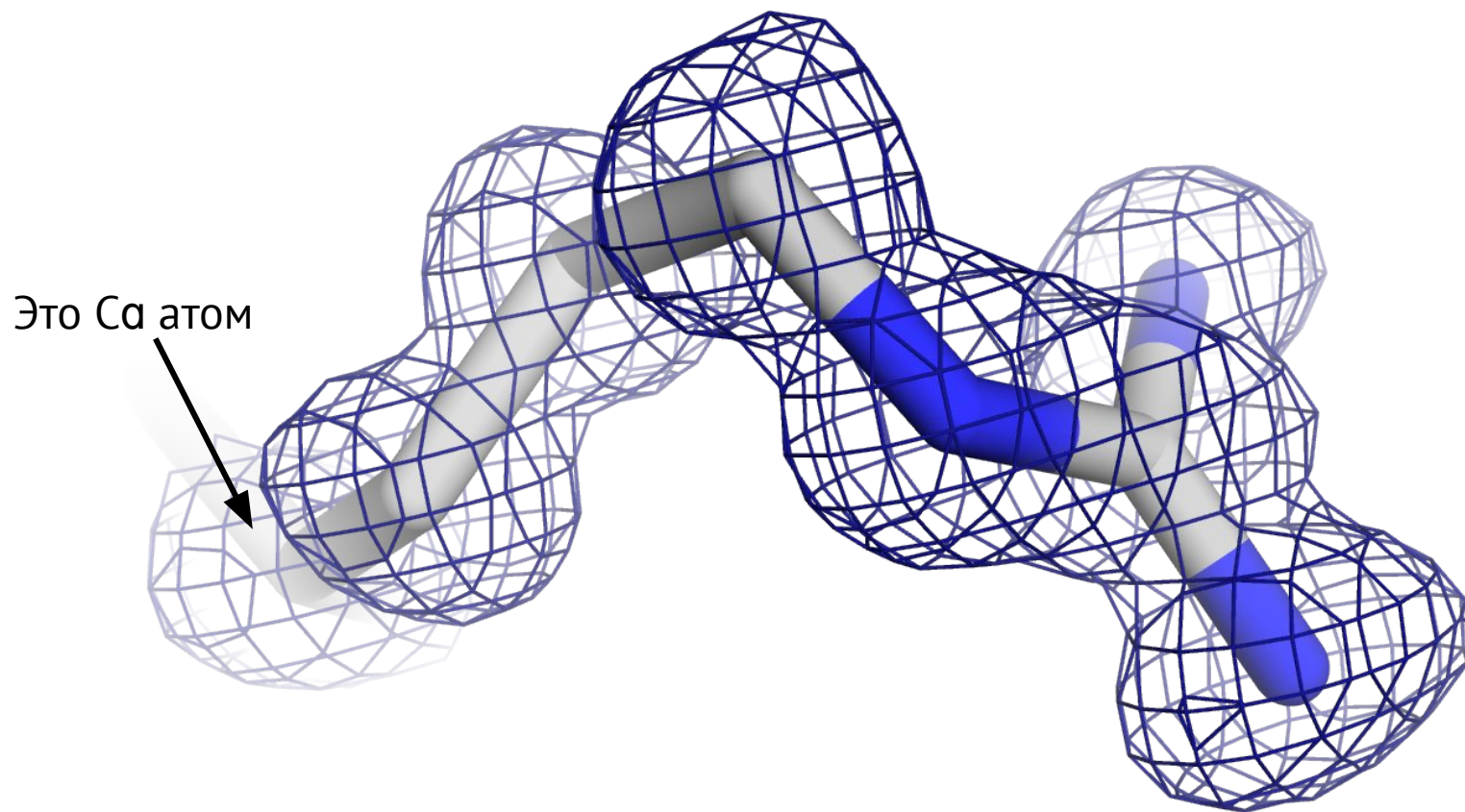
РСА и КриоЭМ

Это Ca атом

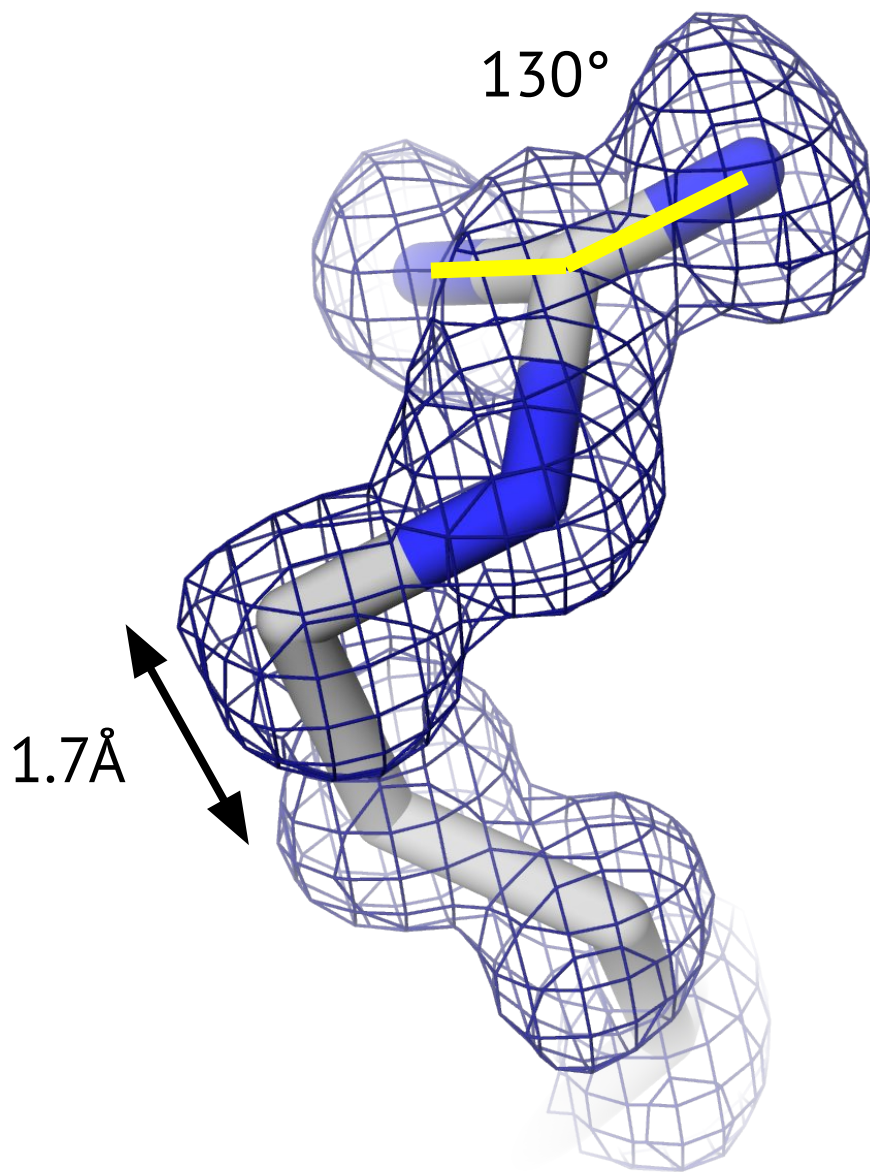
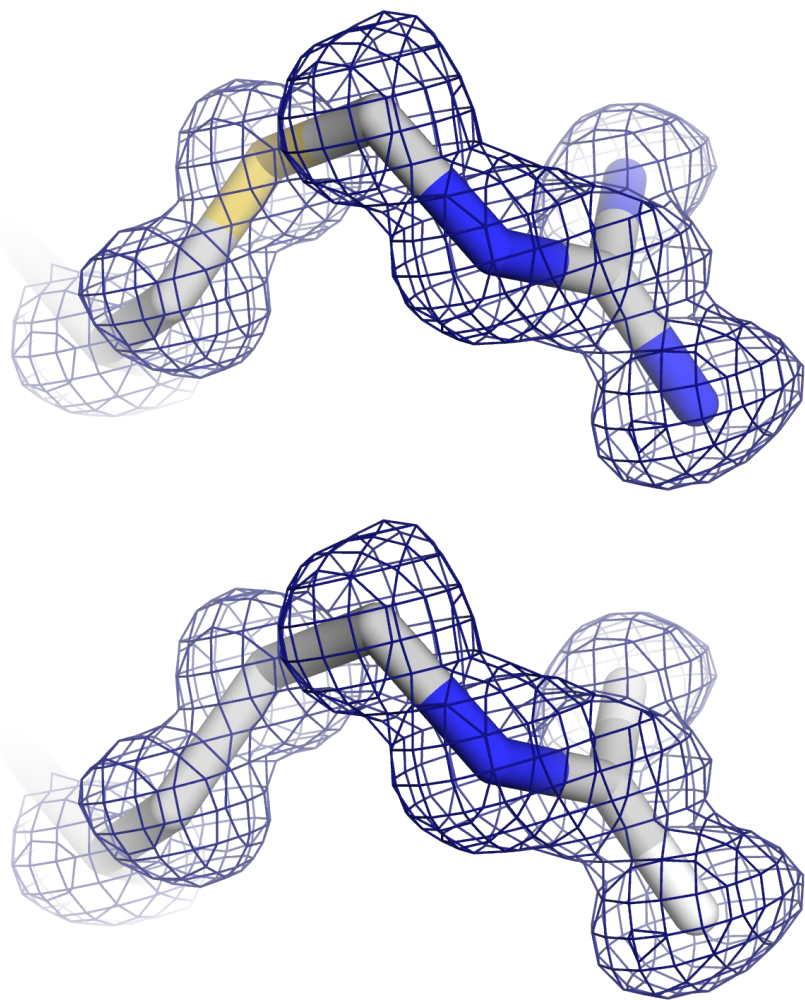


Данные РСА и КриоЭМ – области пространства с повышенной электронной плотностью. Боковой радикал какой аминокислоты хорошо соответствует этой плотности?

РСА и КриоЭМ

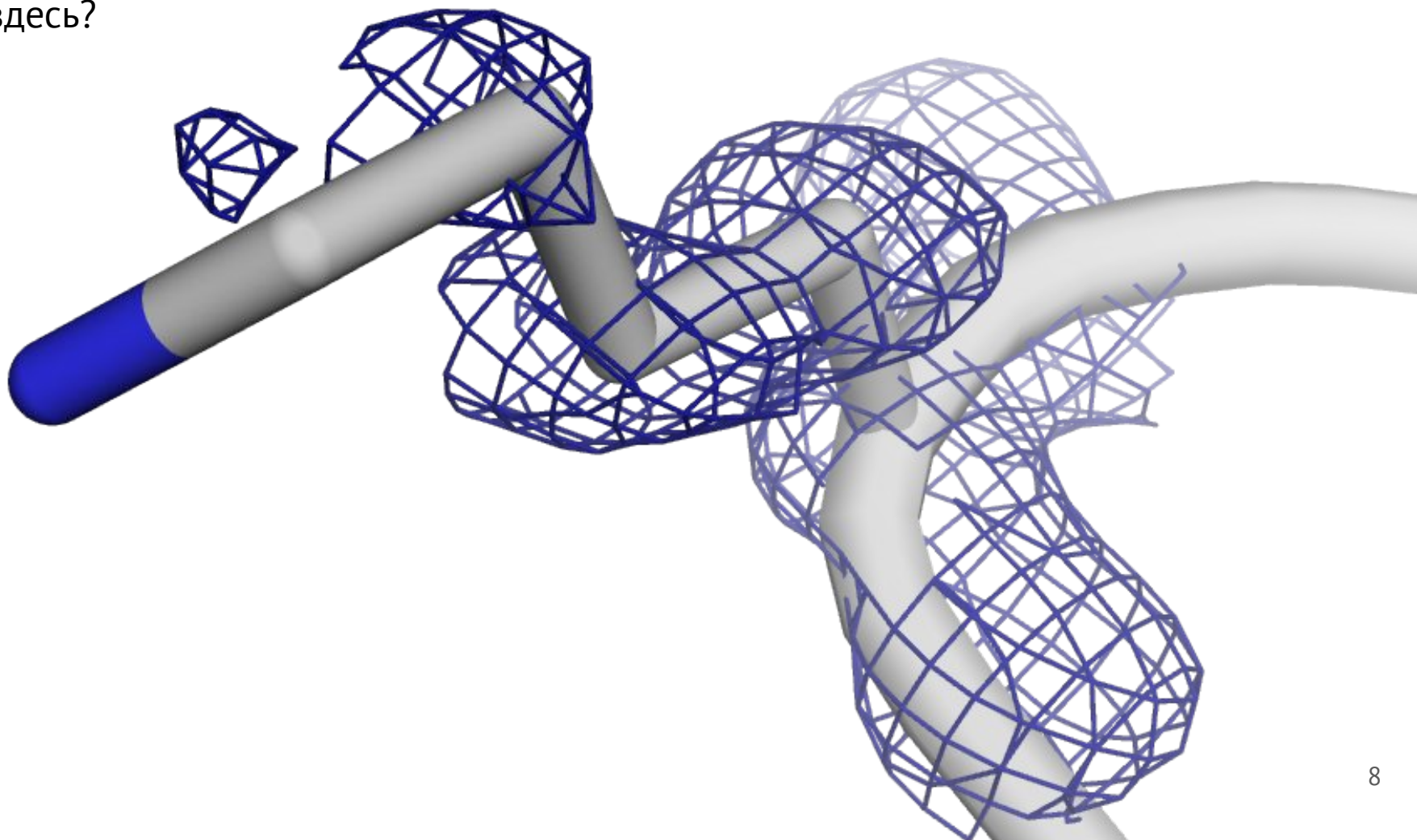


Почему не так?

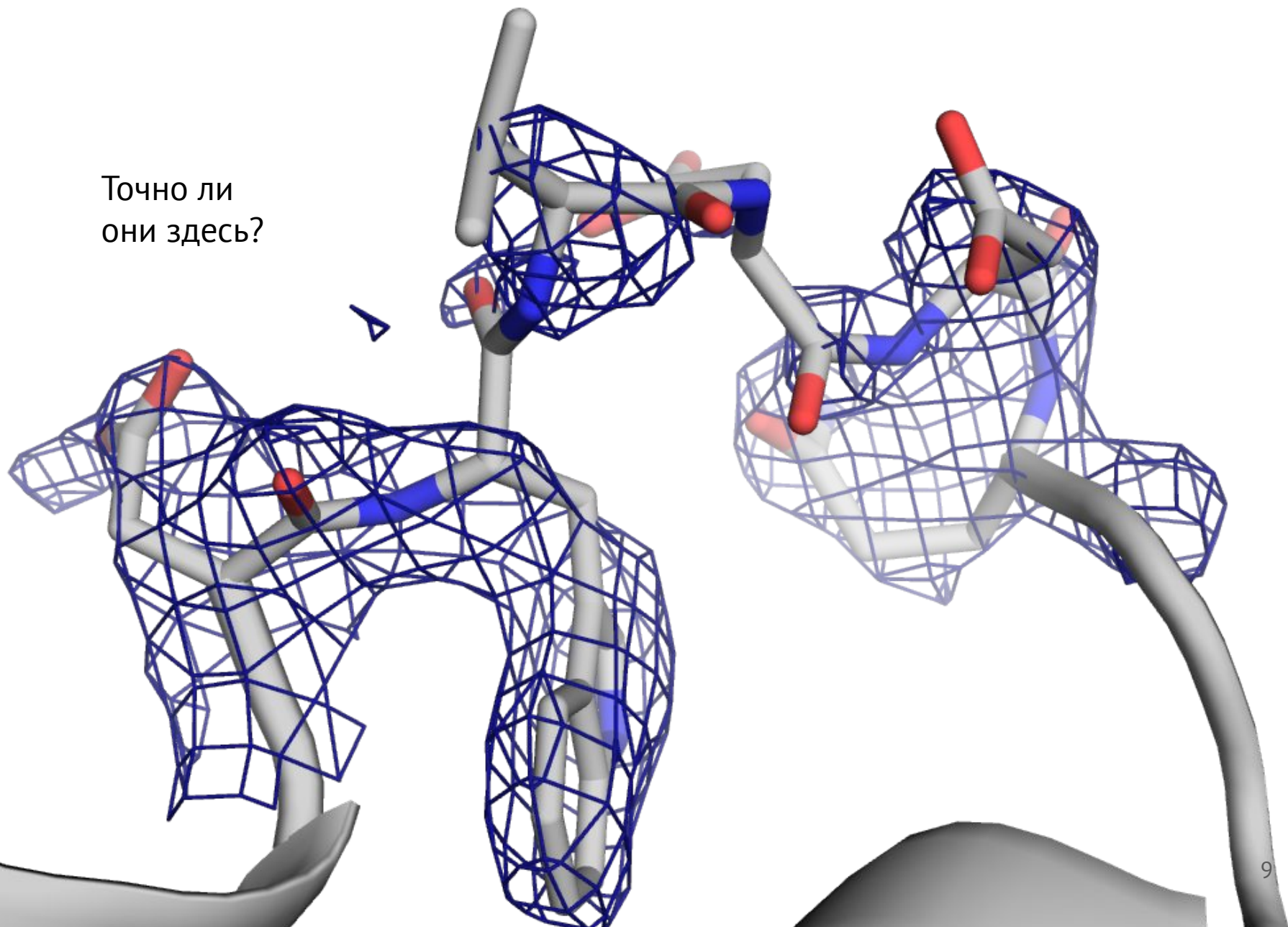


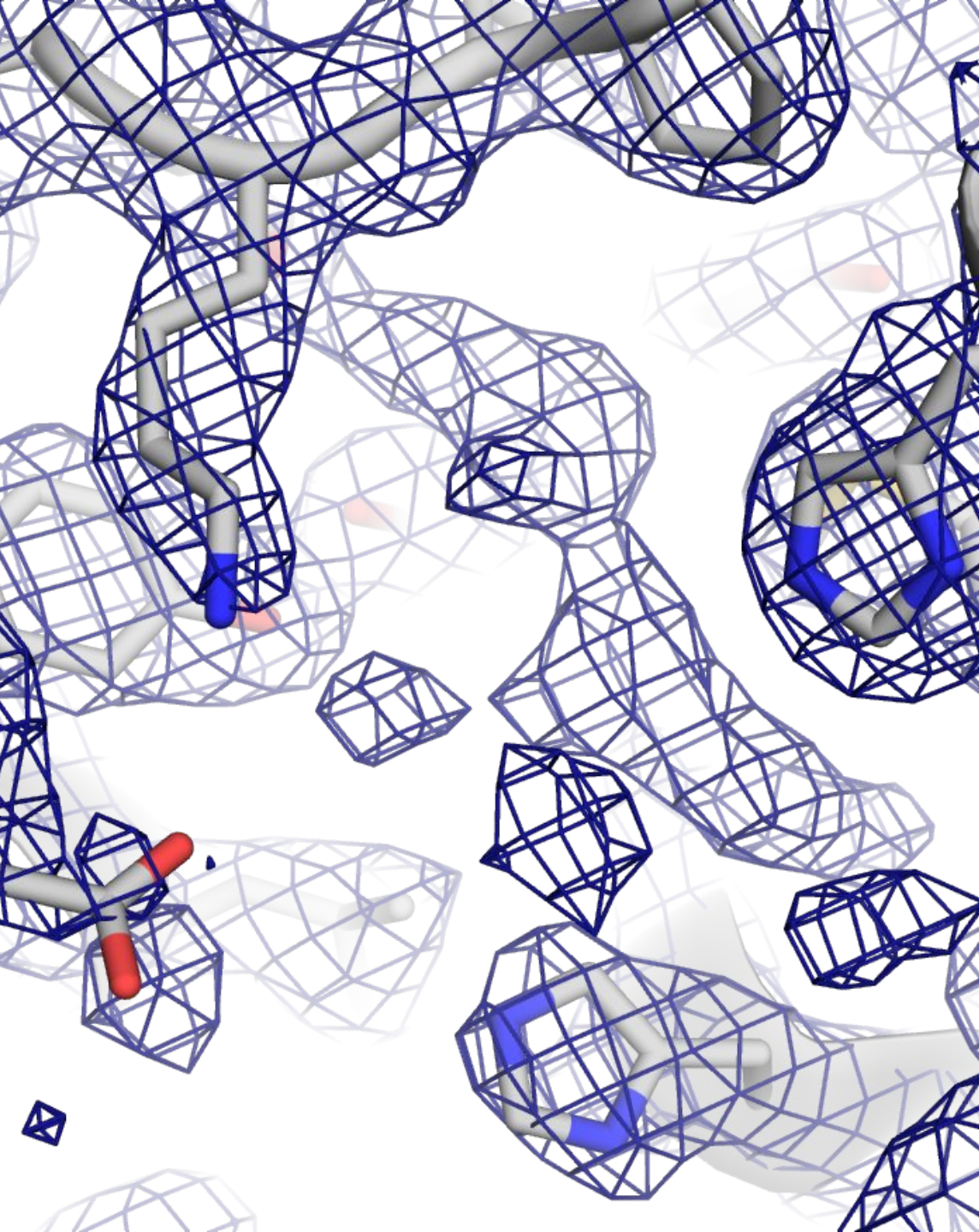
Не всегда все так хорошо

Точно ли
он здесь?



Точно ли
они здесь?





Что это?

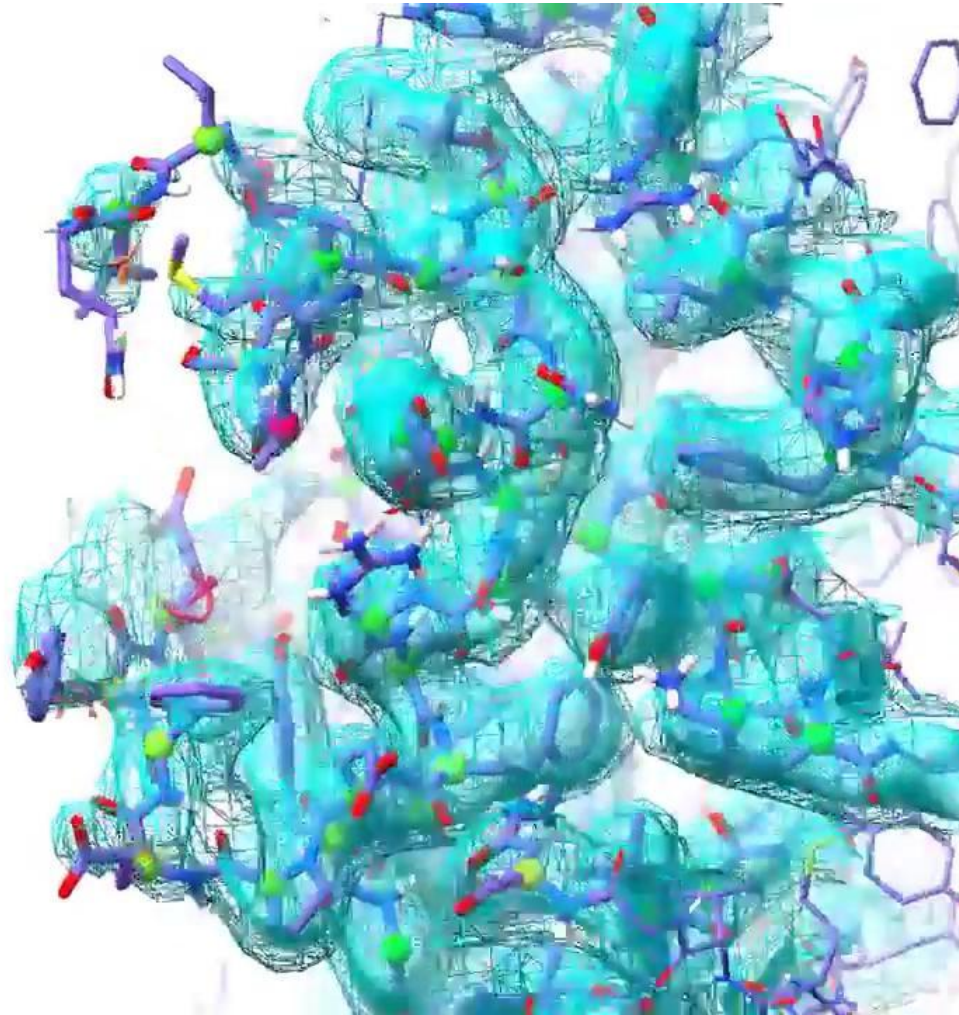
Иногда все очень плохо



Tristan Croll
@CrollTristan

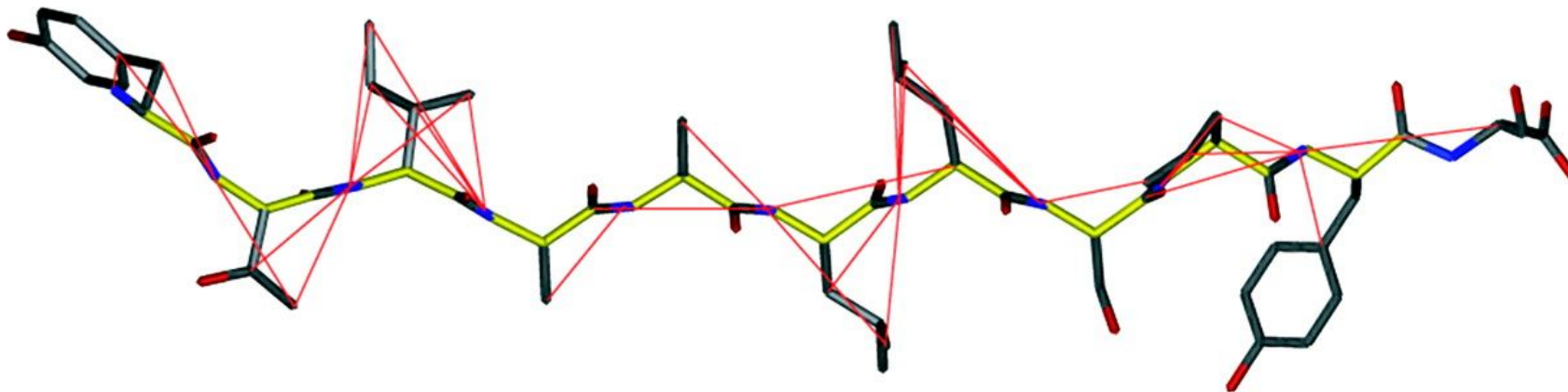


Oh, wow - looks like they're all out by *nine* residues (Pro 918 is where Pro 927 should be; Leu 907 is where Trp 916 should be). Would *not* have been easy to pick this in the original 6nur, but 7btf is both higher resolution and has slightly more "tail".

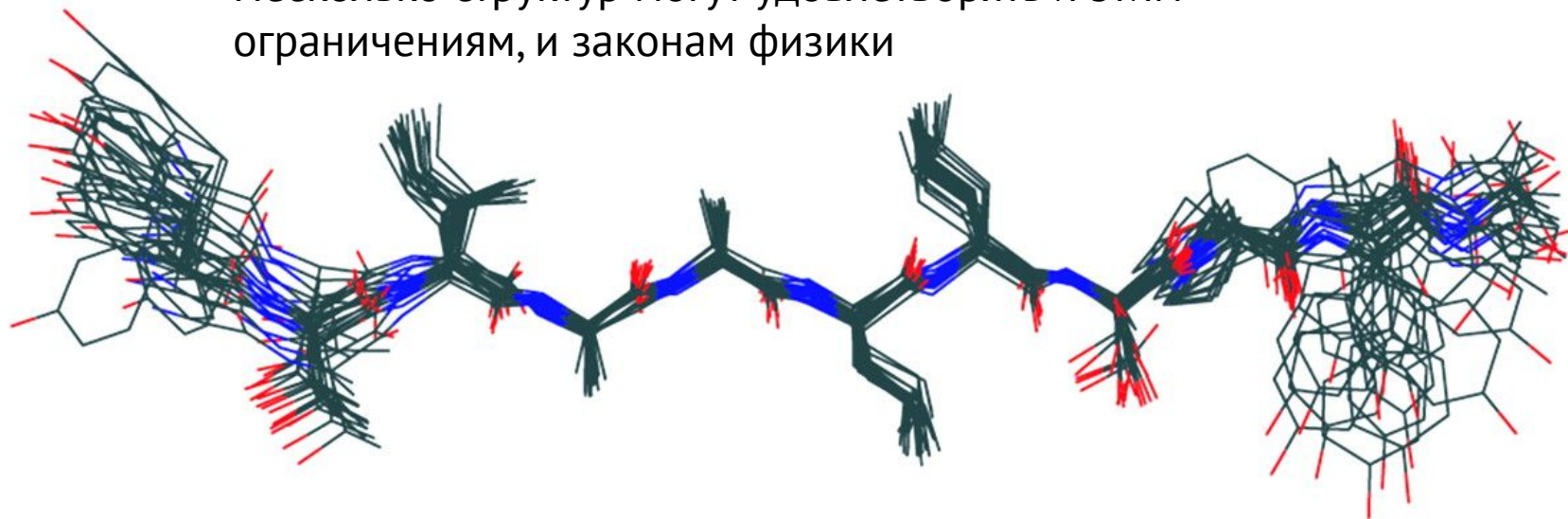


ЯМР

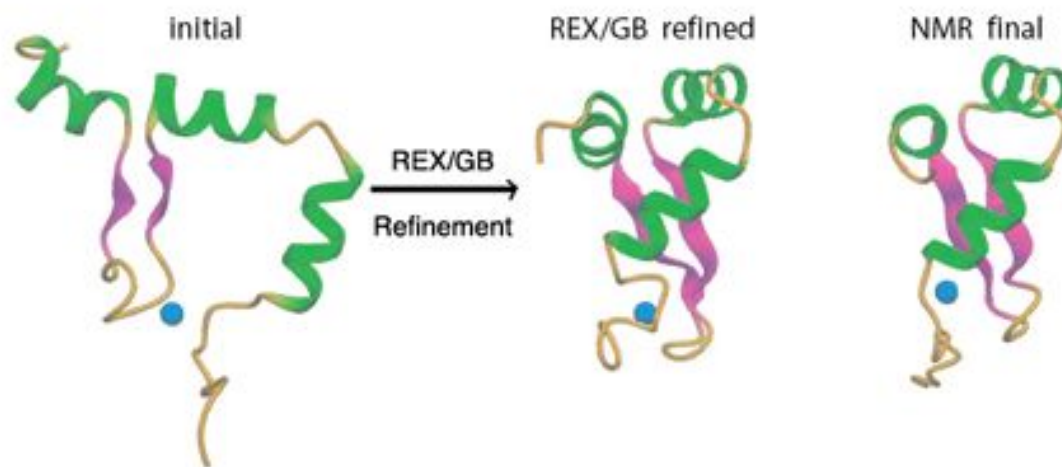
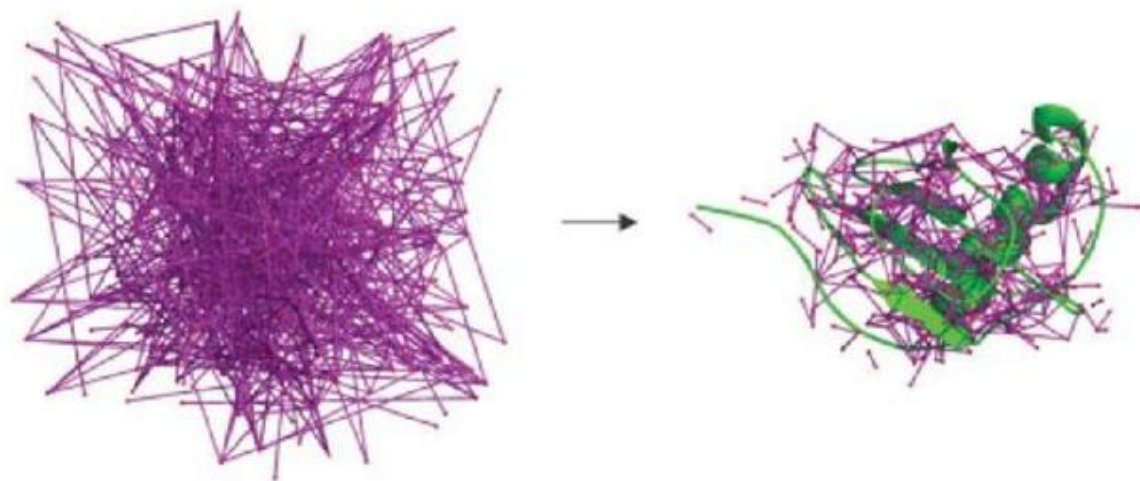
Данные ЯМР – набор ограничений на расстояния



Несколько структур могут удовлетворять и этим ограничениям, и законам физики



ЯМР

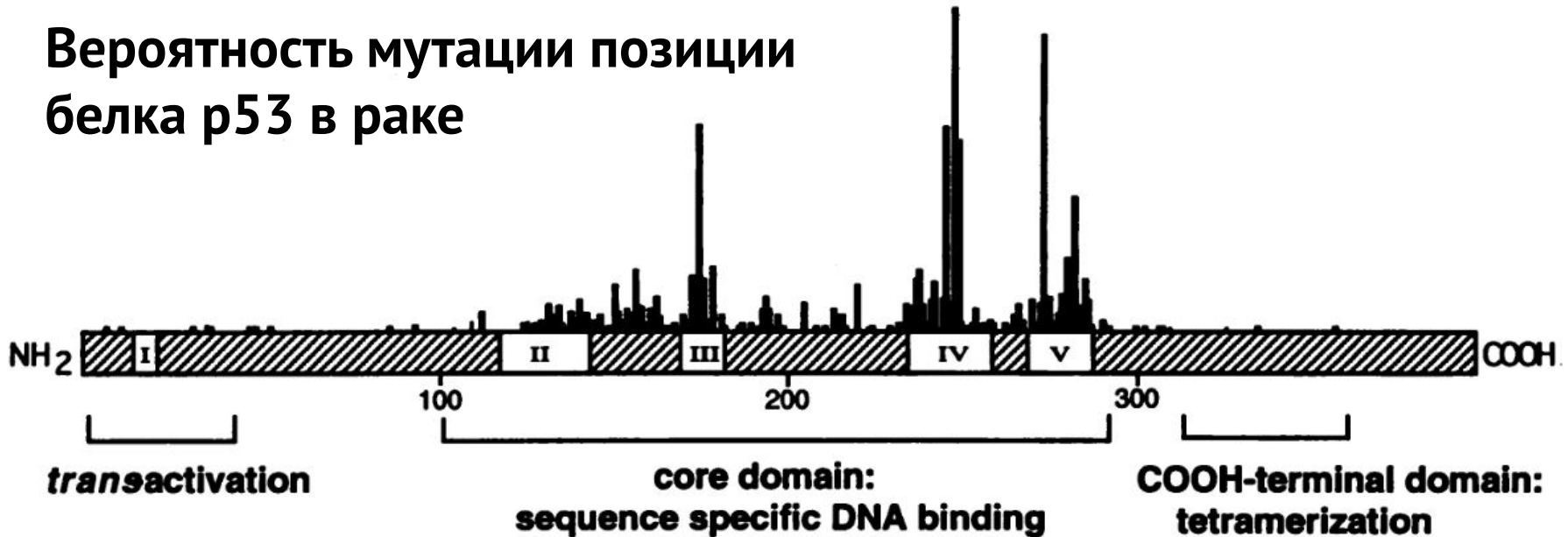


Зачем мы получили информацию о структуре?

Функция макромолекул “защита” в их структуре и динамике и влияет на последовательность через эволюционный отбор

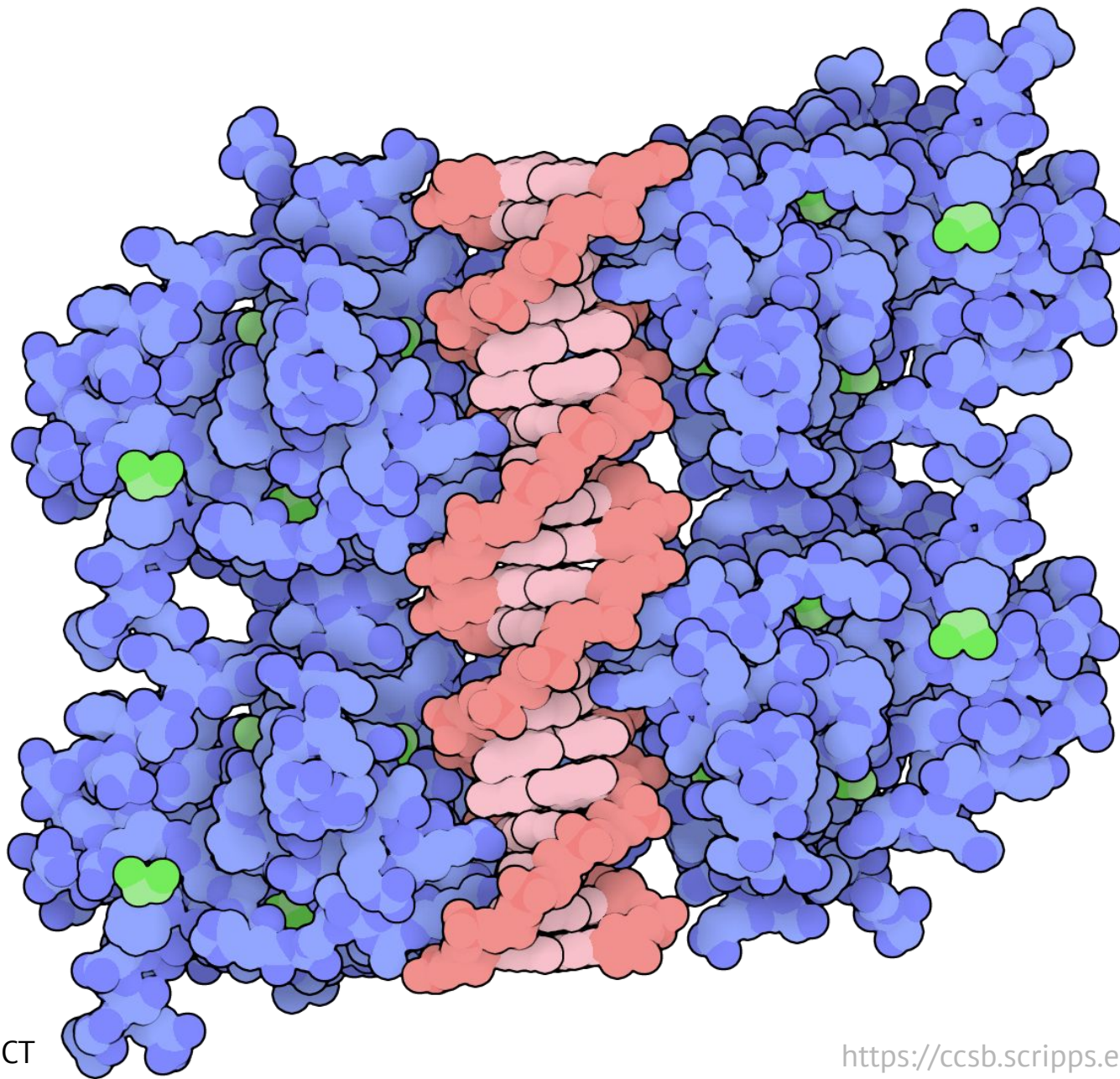
- Рассмотрение на уровне структуры и динамики отвечает на вопросы
- Как из ограниченного алфавита собрать разные функции? Какие функции можно ожидать от биомолекул, а какие нет?
 - Почему отбор действует на сохранение/избегание данных букв в данных позициях?
 - Как достигается избирательность в молекулярном узнавании – субстратов, эффекторов, регуляторных последовательностей ДНК

Вероятность мутации позиции белка p53 в раке



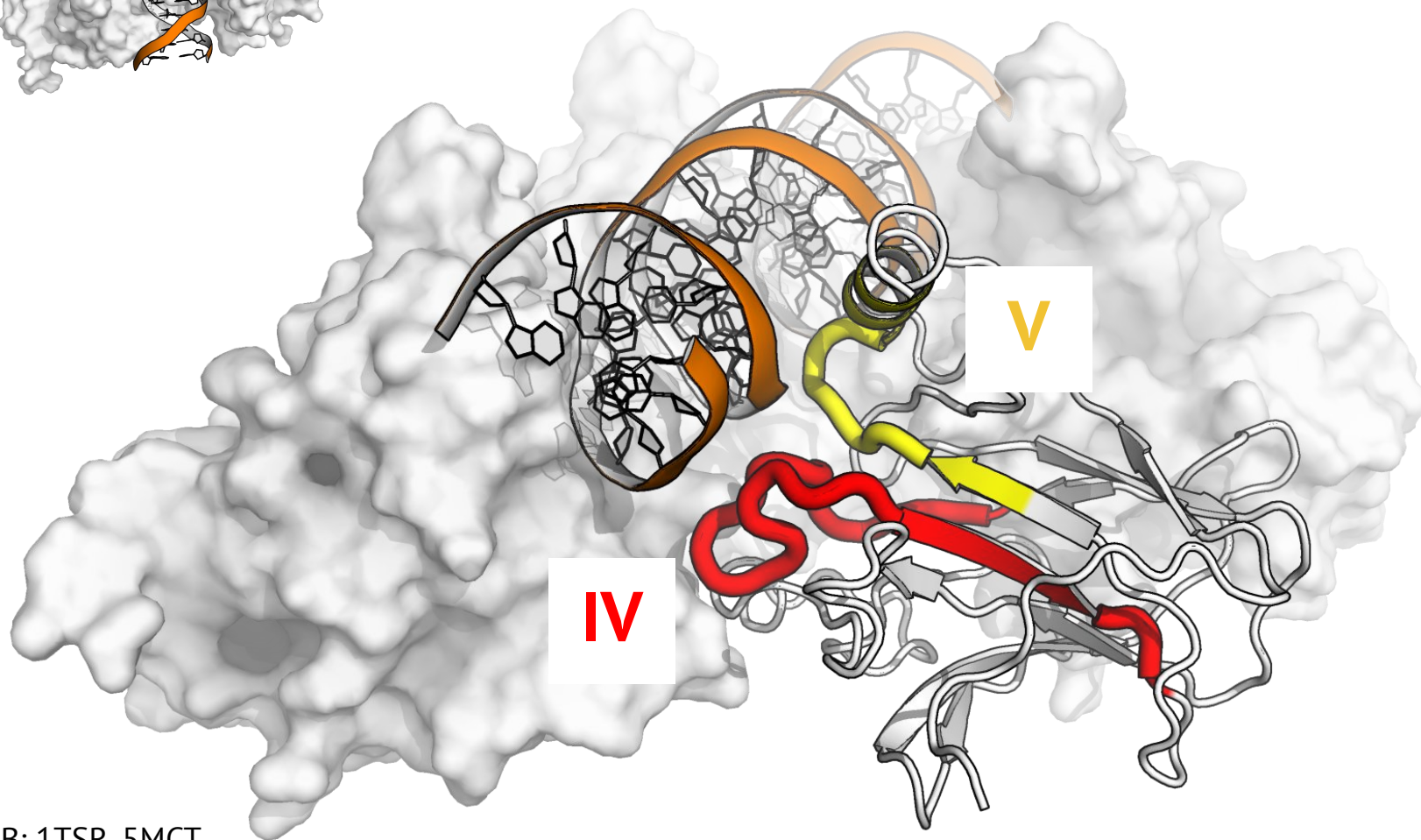
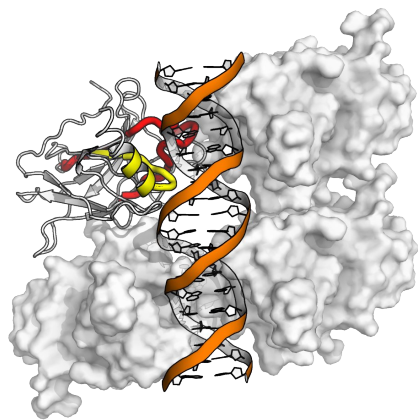
Римские цифры показывают участки с высокой консервативностью последовательности. Очевидно – если произошла мутация в консервативном регионе, то функция белка вероятно повреждена и это привело к канцерогенезу.

Структурная биоинформатика в данной ситуации должна ответить на вопрос **Почему именно эти участки последовательности консервативны?**
Какова их функция?

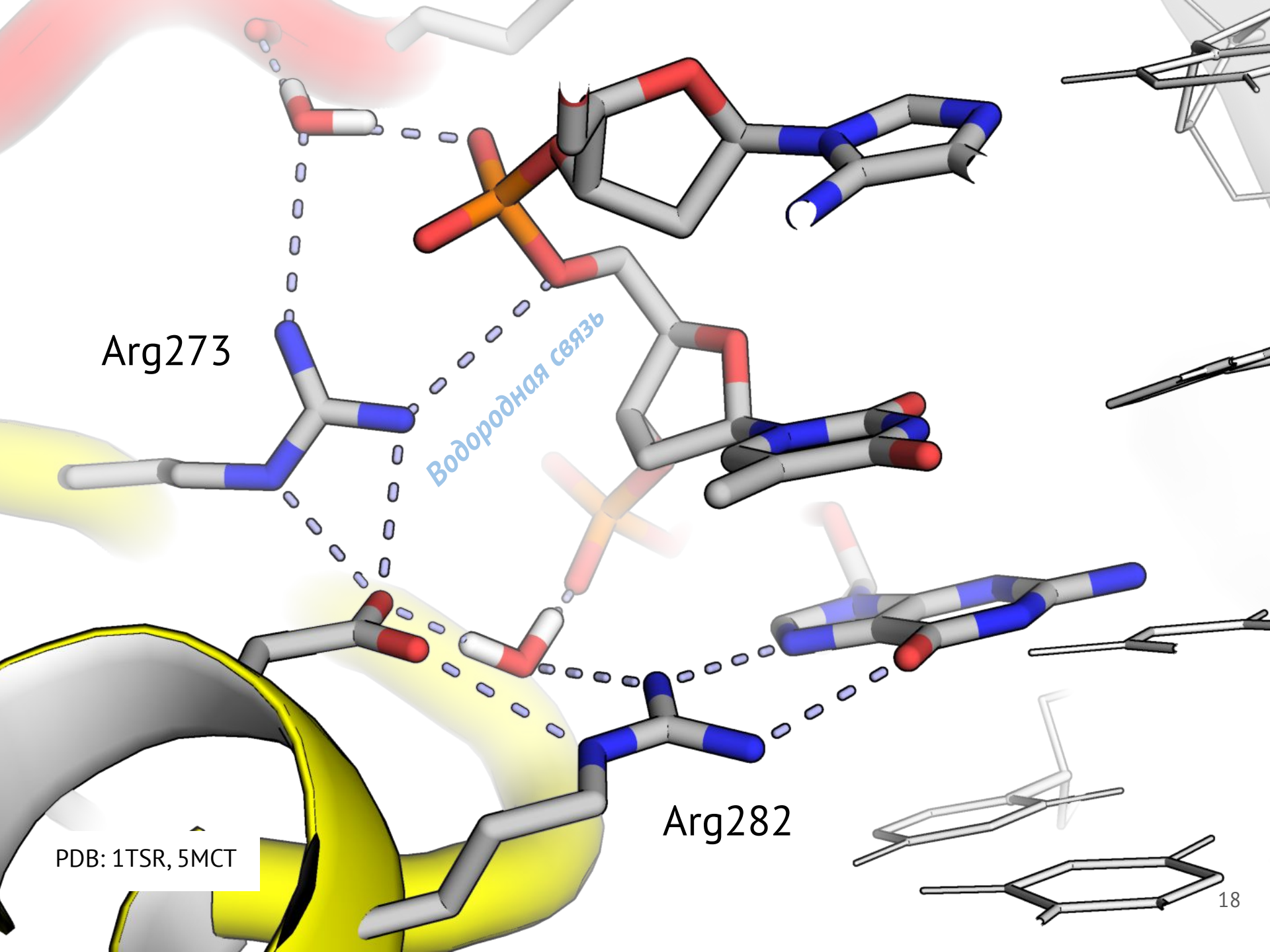


PDB: 1TSR, 5MCT

<https://ccsb.scripps.edu/illustrate/>



PDB: 1TSR, 5MCT

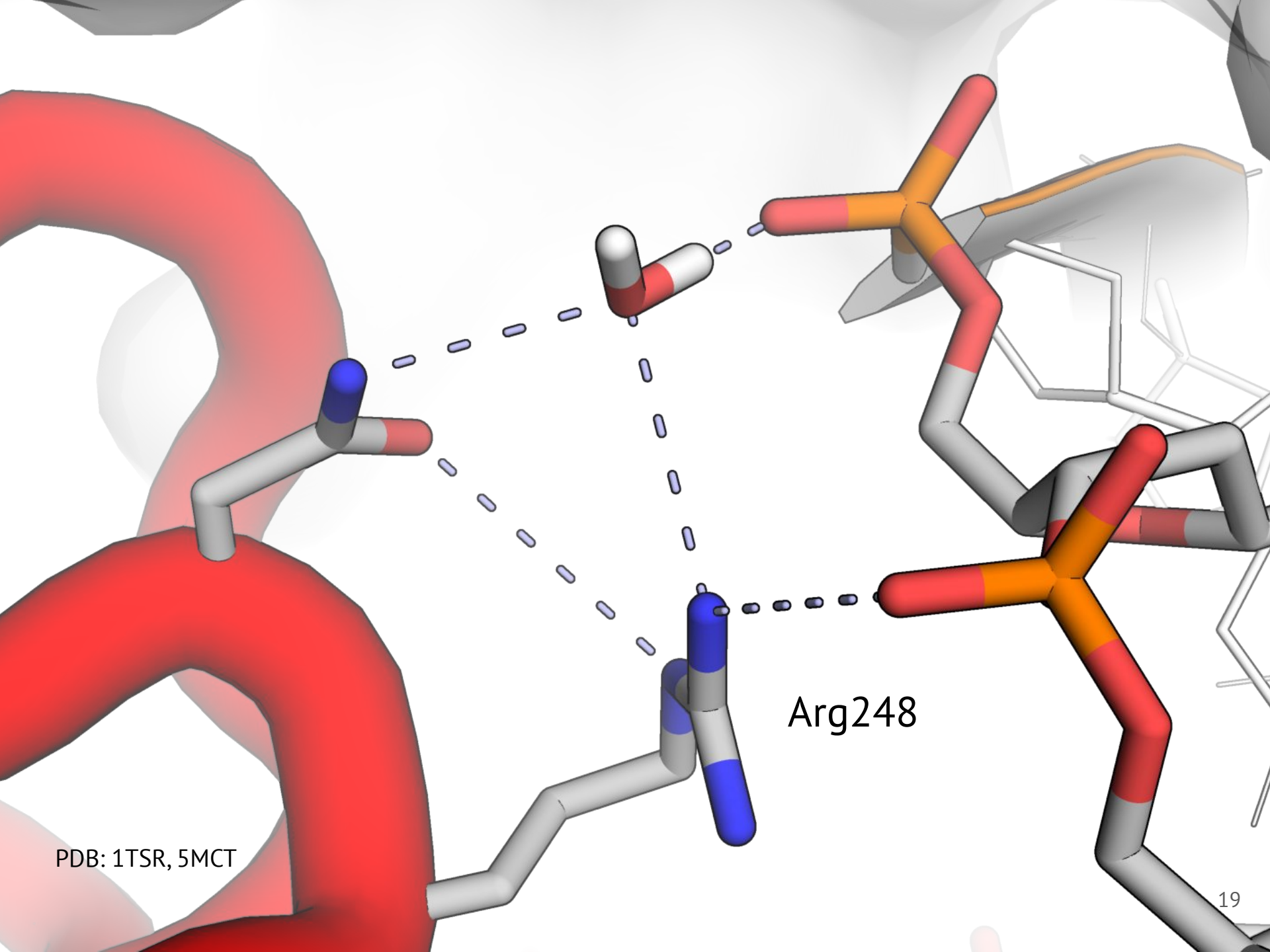


Arg273

Водородная связь

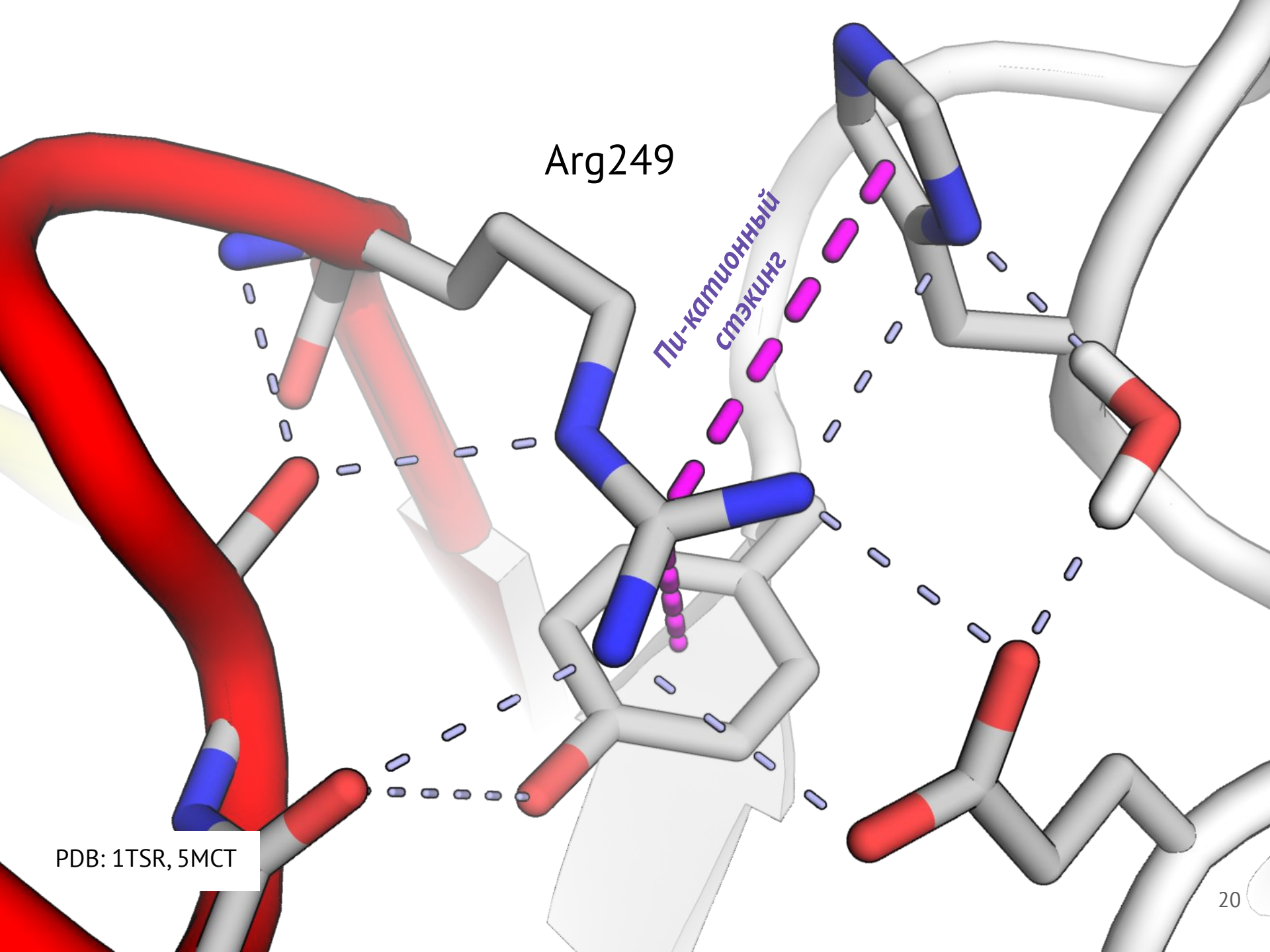
Arg282

PDB: 1TSR, 5MCT



PDB: 1TSR, 5MCT

Arg248

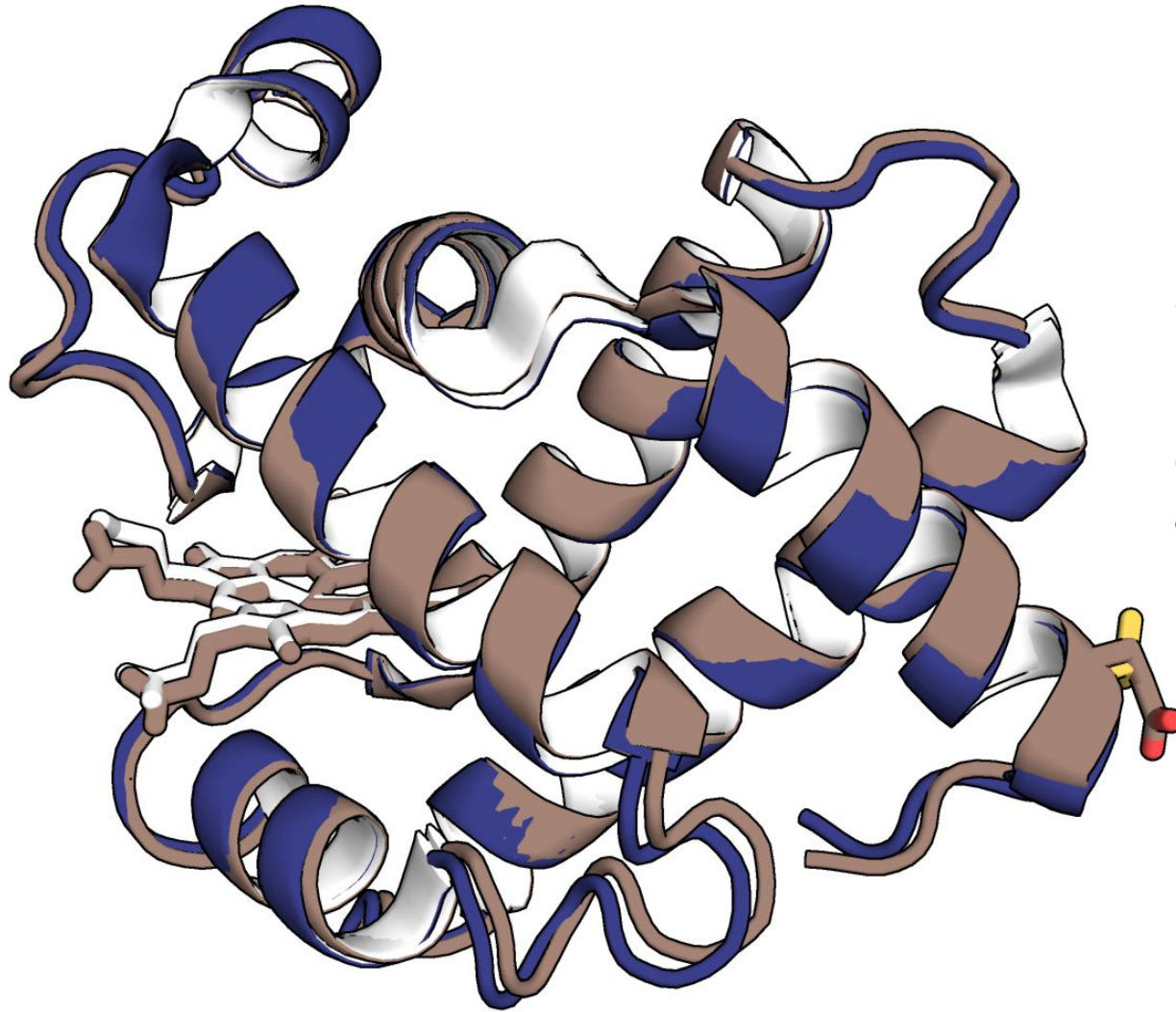


Arg249

Пи-катионный
стэкинг

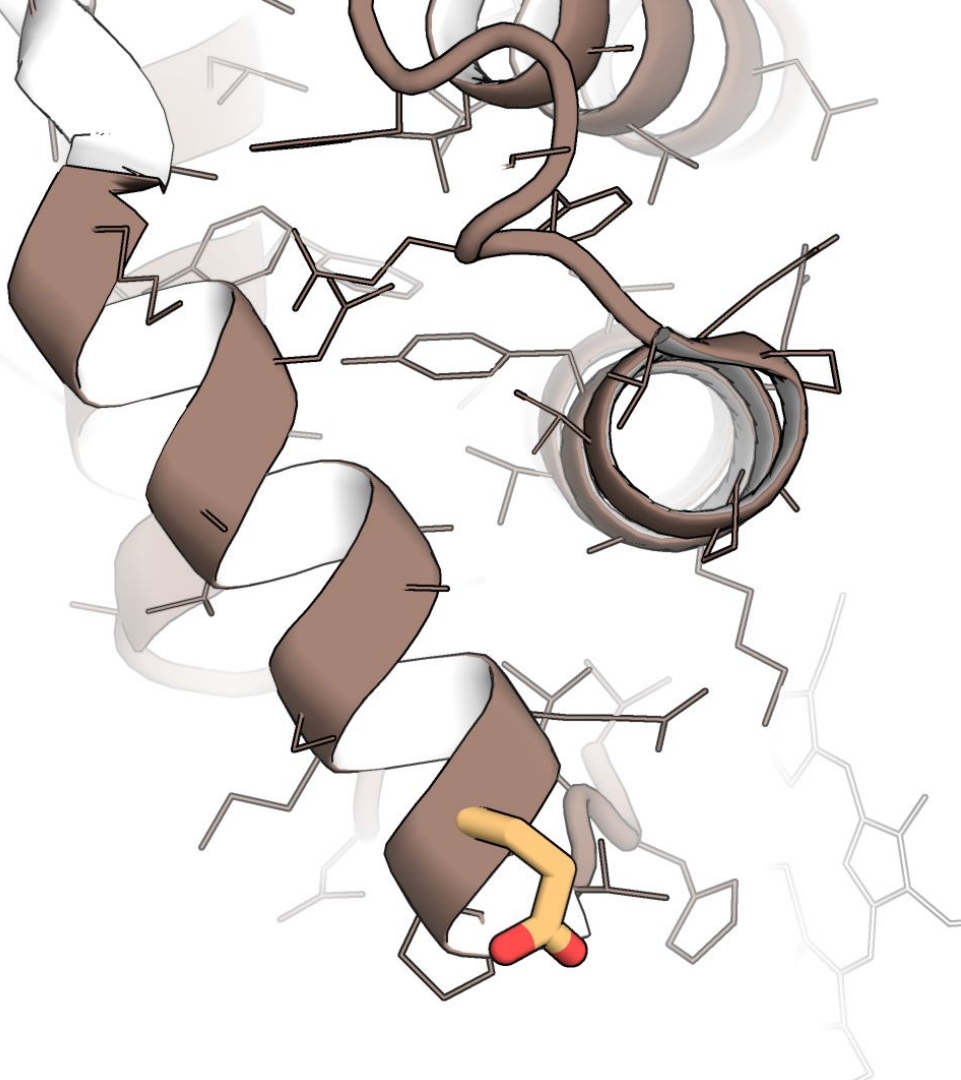
PDB: 1TSR, 5MCT

Агрегация гемоглобина S

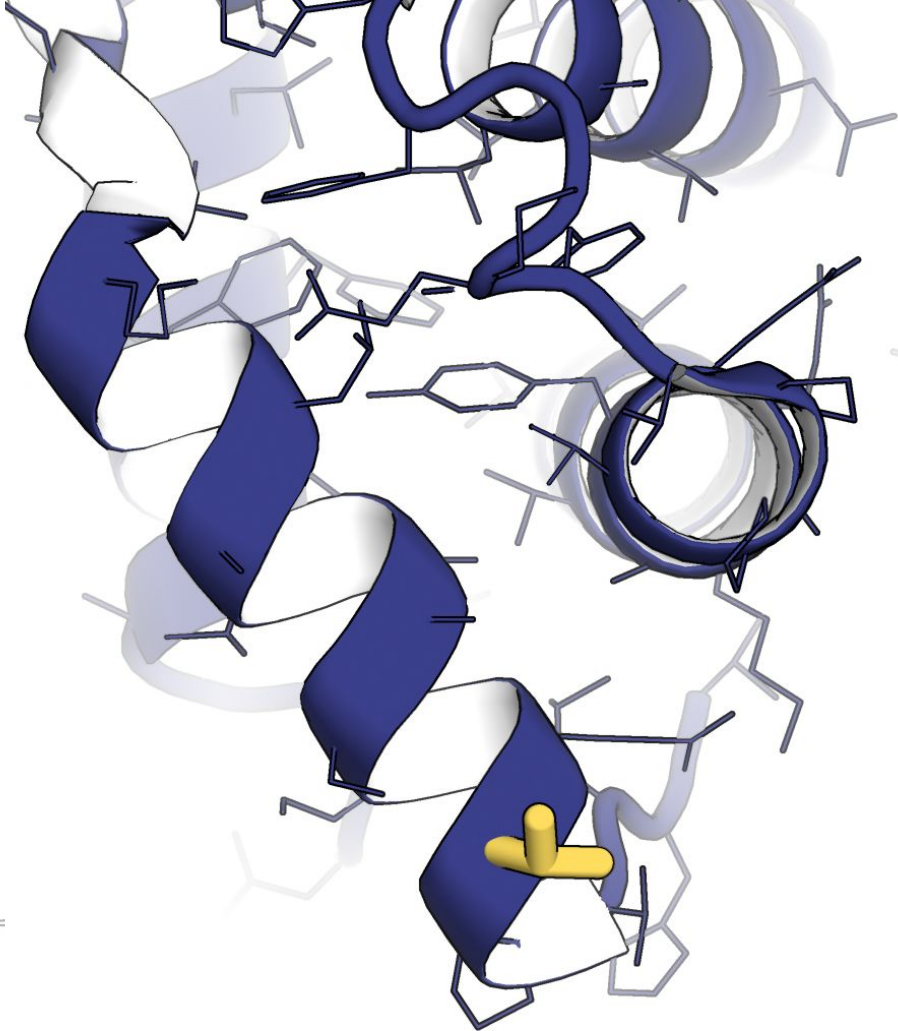


Замена E6V
в бета-субъединице
гемоглобина вызывает
серповидноклеточную
анемию. Как?

PDB: 2HBS, 4HHB

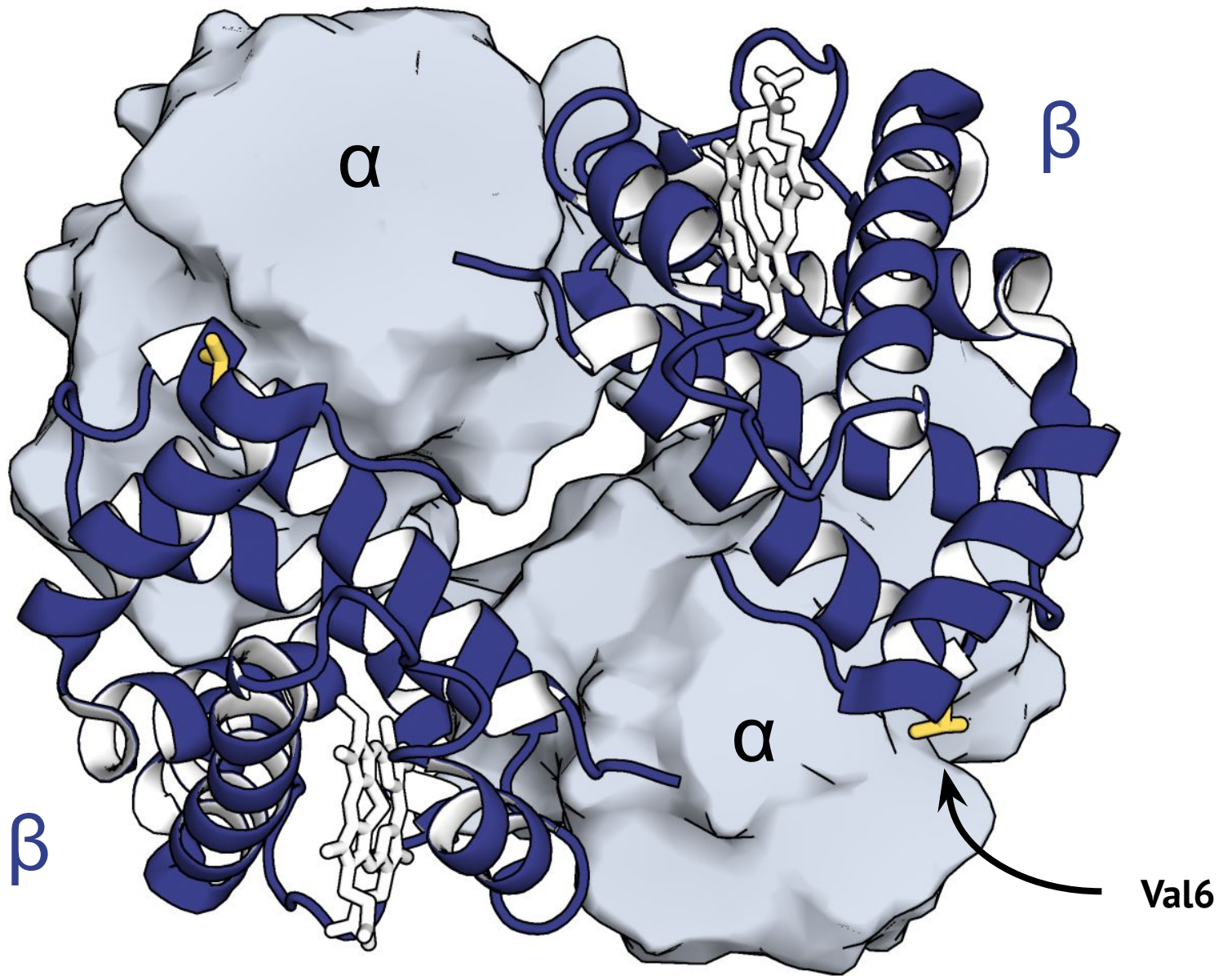


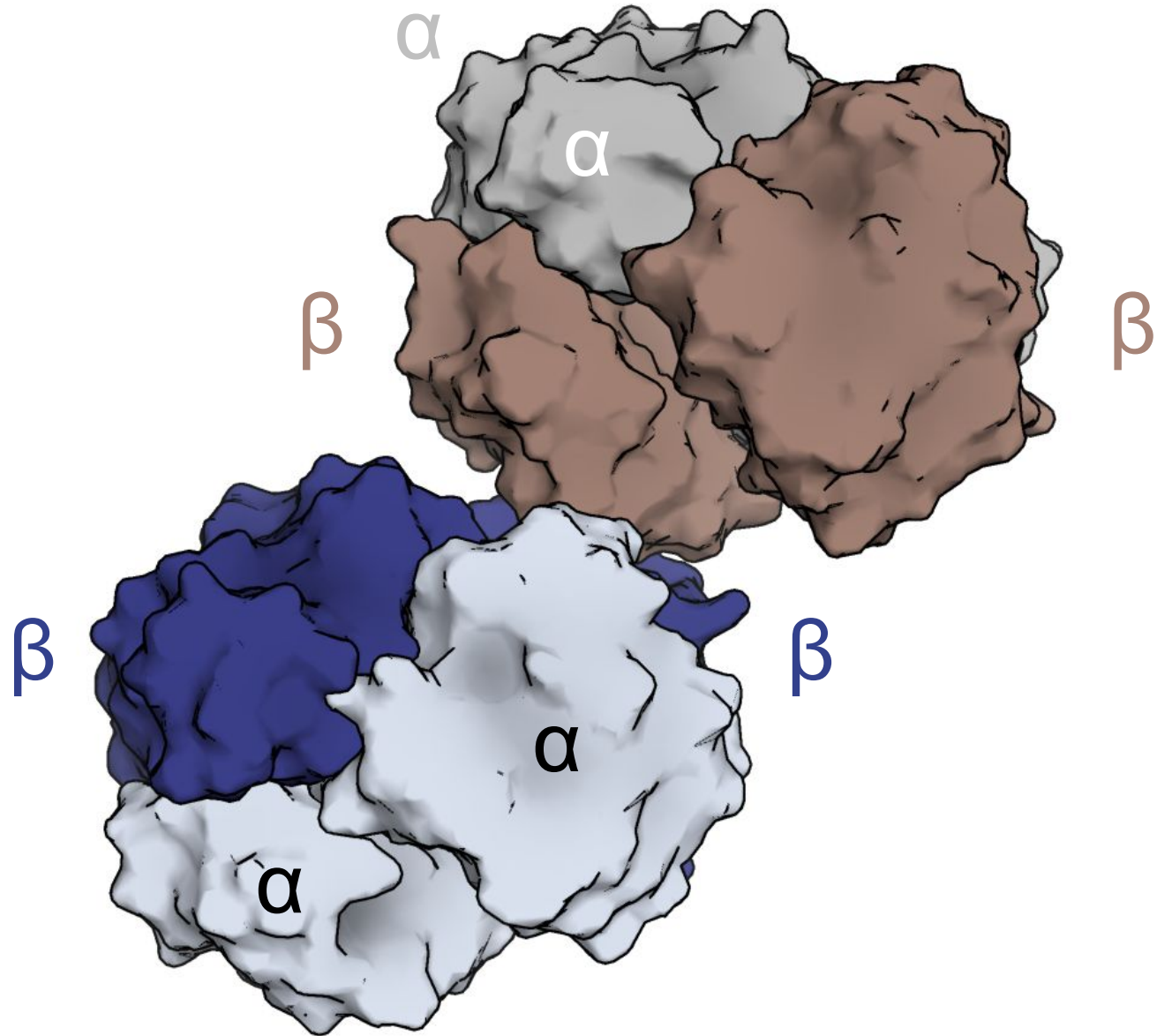
Дикий тип: на 6 позиции глутамат

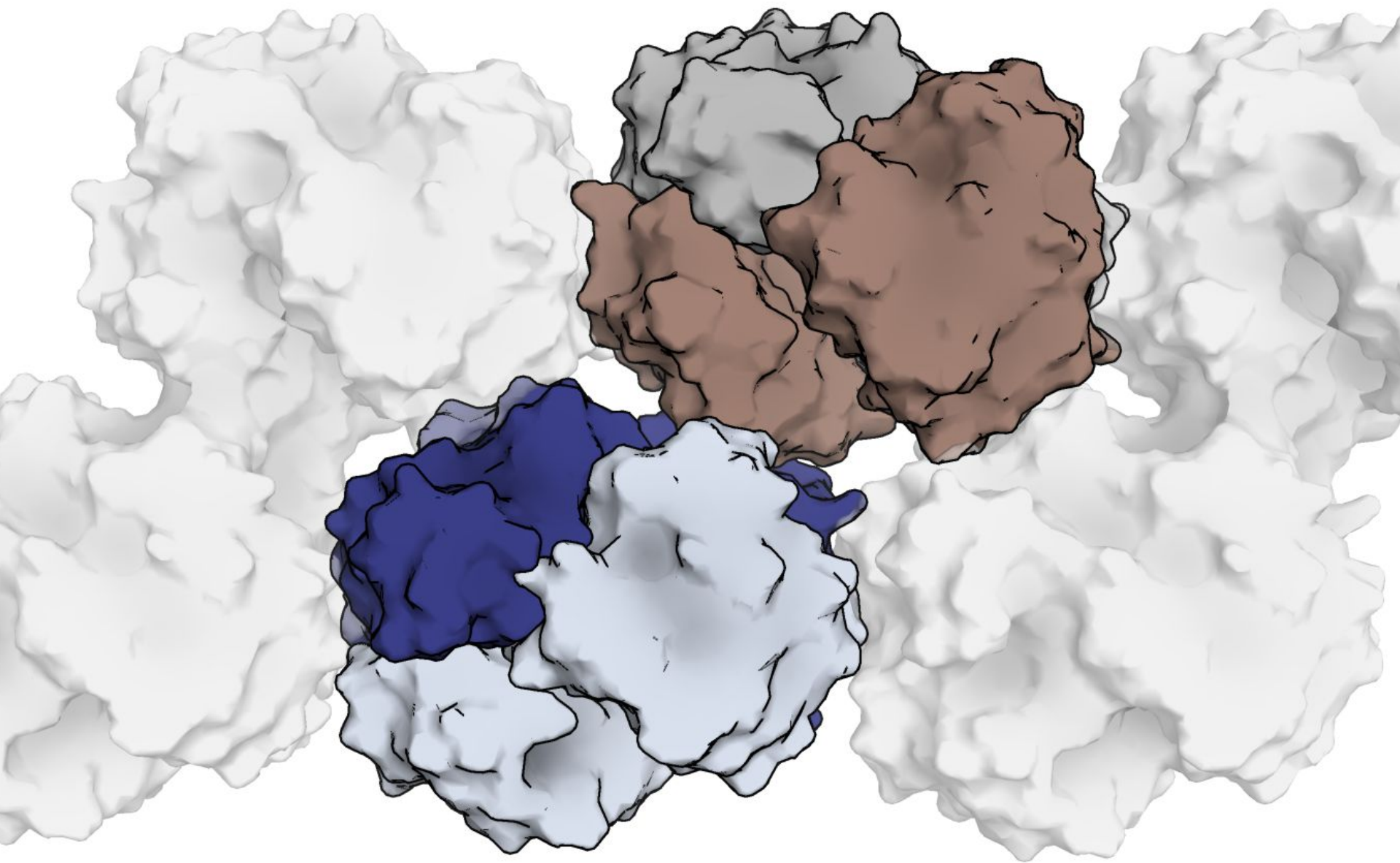


Мутант: на 6 позиции валин

Однако структуры в остальном идентичны →
причина патологии не в мисфолдинге

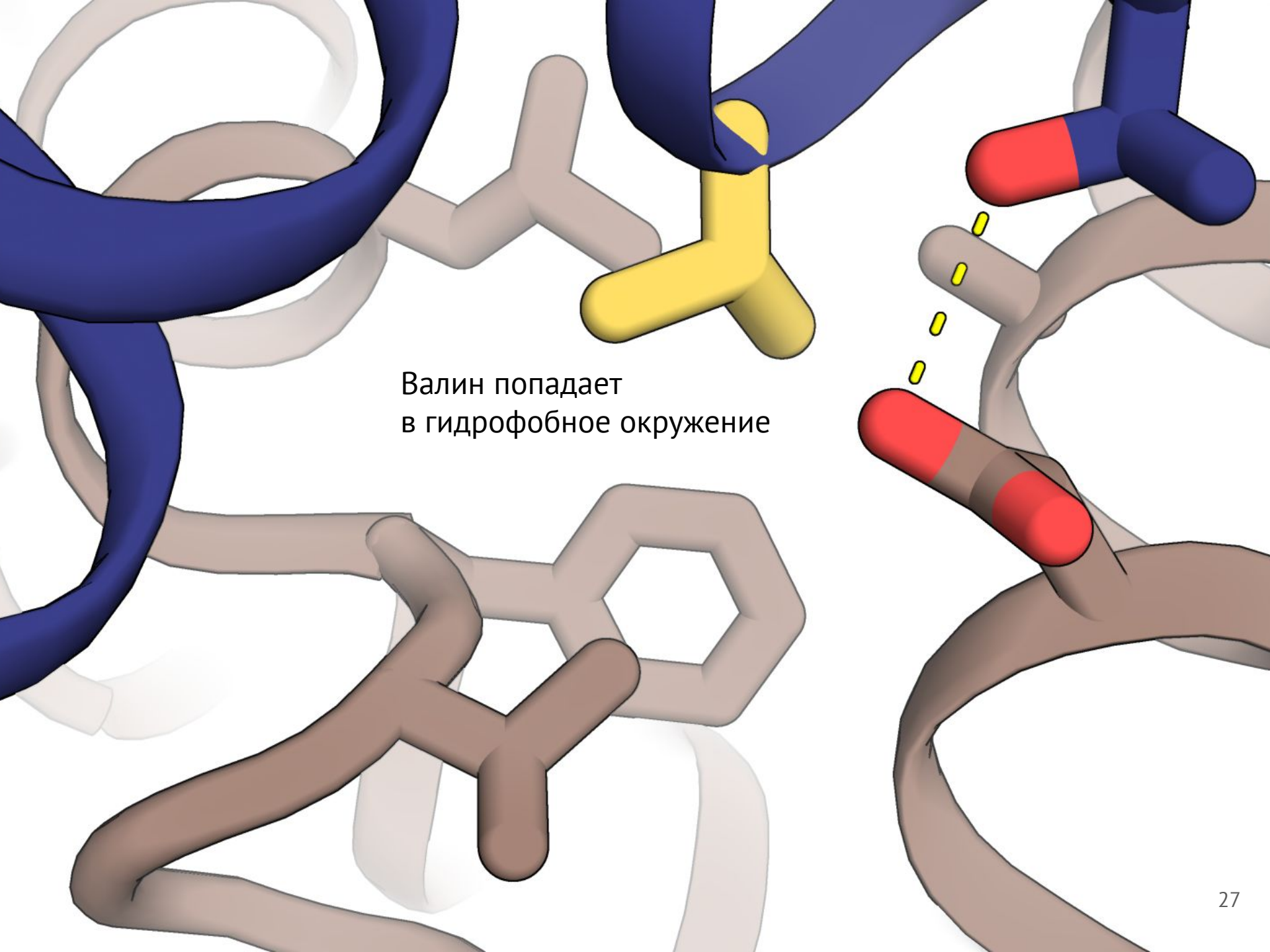






Shape complementarity

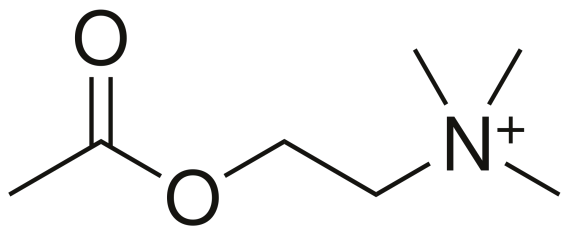
Один из ключевых терминов молекулярного распознавания
Гидрофобный эффект + VdW



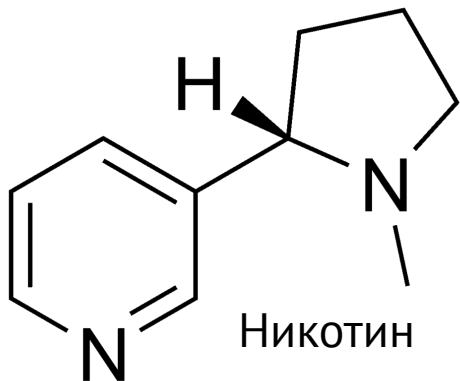
Валин попадает
в гидрофобное окружение

Неспецифичность

Арсенал нековалентных взаимодействий ограничен, что приводит к ограниченной специфичности рецепторов и создает возможность эксплоитов не-нативными лигандами.



Ацетилхолин



Никотин

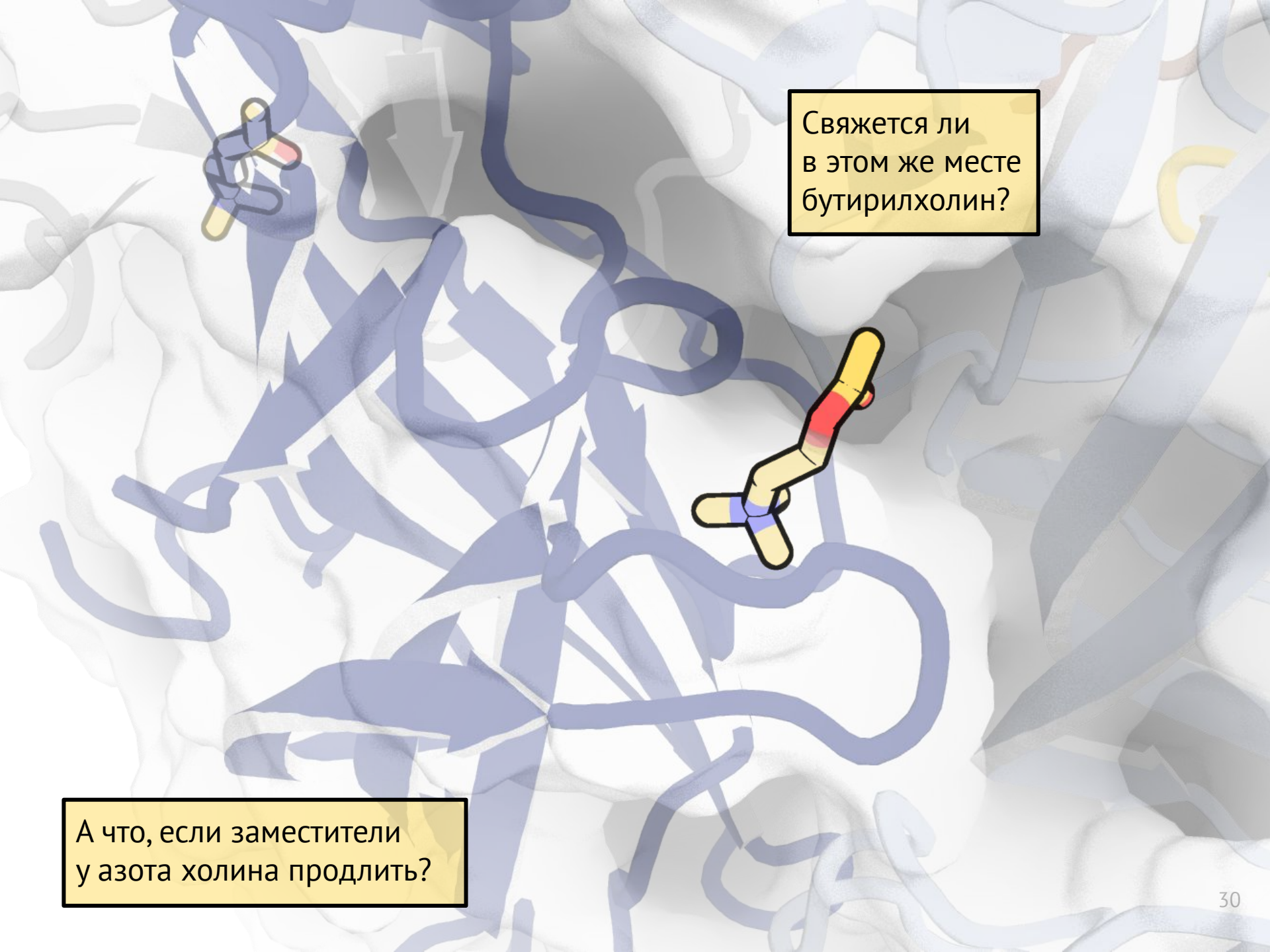


nAChR
PDB 5KXI

Ацетилхолин связывается
в 5 идентичных полостях
между субъединицами

Число связавшихся молекул
коррелирует с величиной
сигнала





Свяжется ли
в этом же месте
бутирилхолин?

А что, если заместители
у азота холина
продлить?

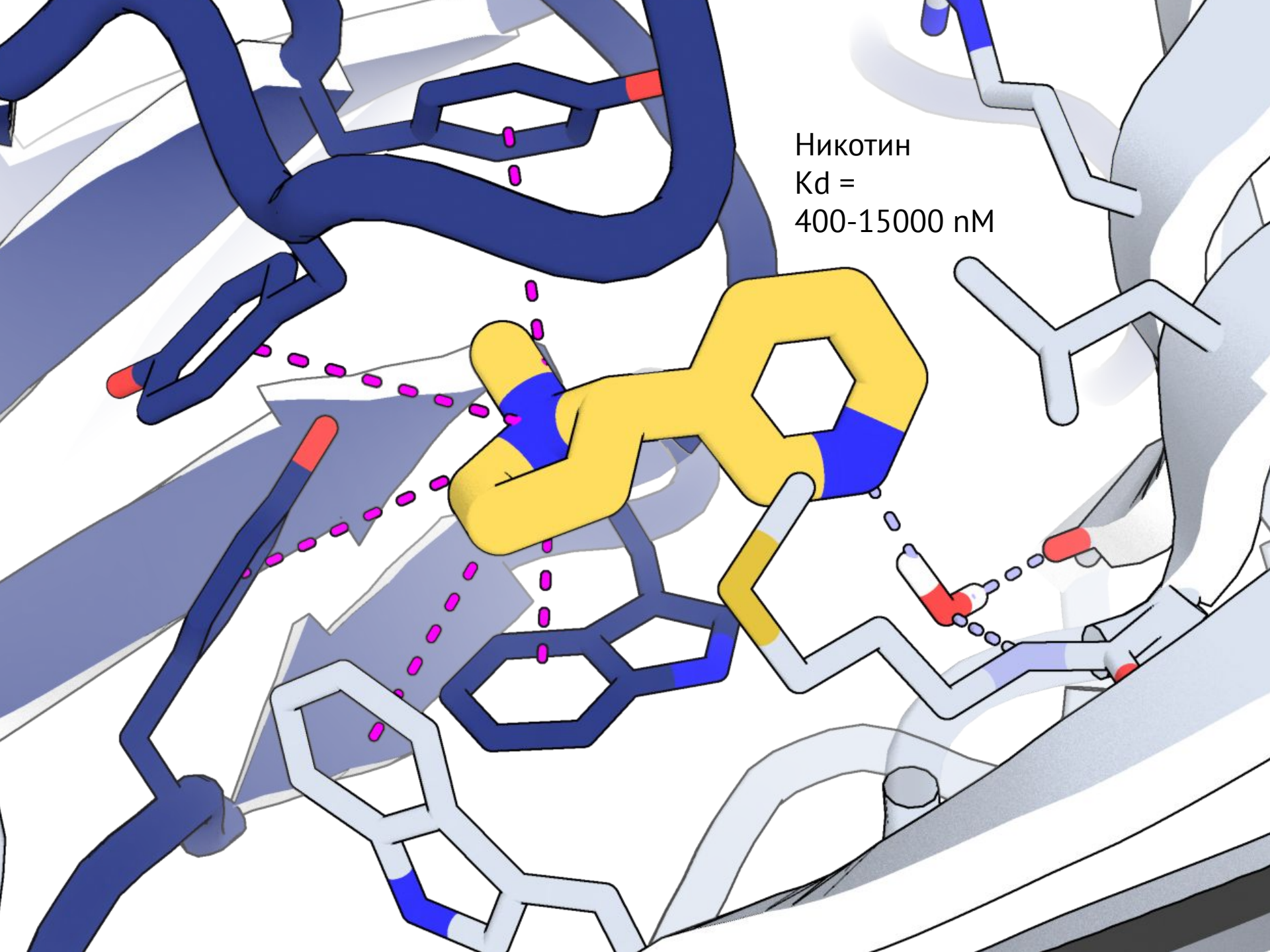
$K_d =$
4000-180000 nM

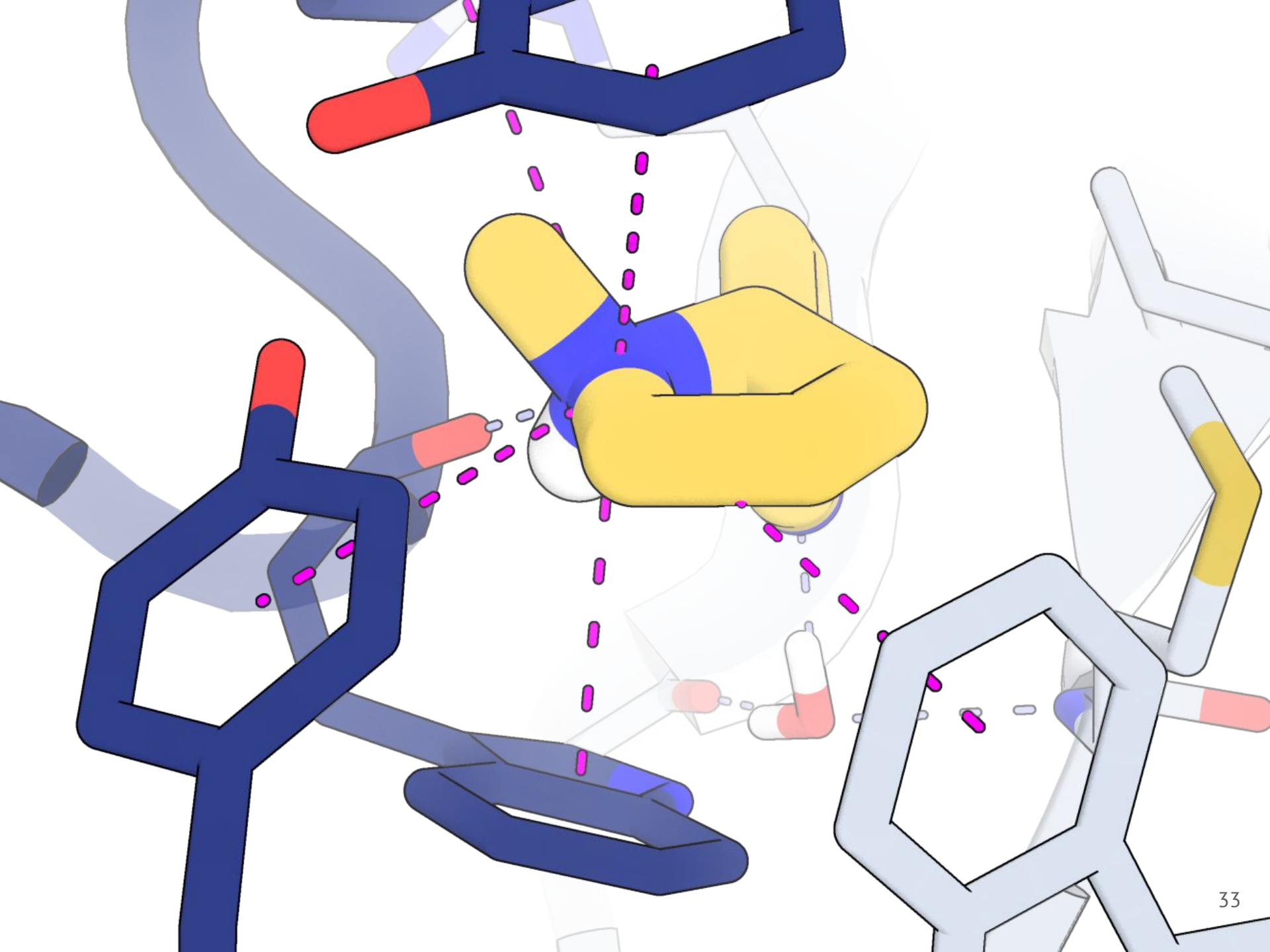
Гидрофобный
карман,
образованный
аргинином
и лейцином

Пи-кат. +
гидрофобика

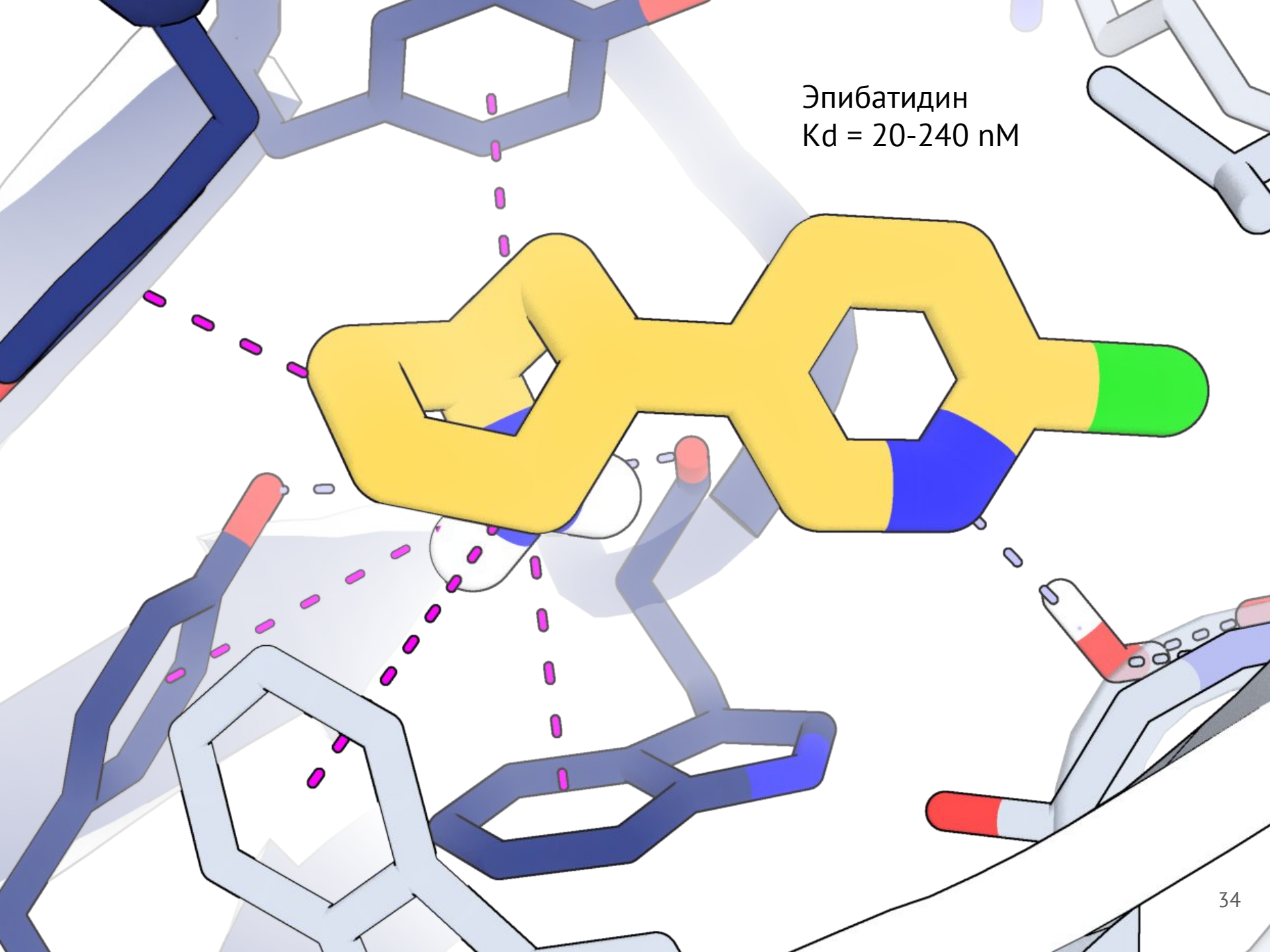
Водяной
МОСТИК

Никотин
Kd =
400-15000 nM





Эпibatидин
Kd = 20-240 нМ



Трехмерность и образность – баг или фича?

Особенность в работе со структурами заключается в том, что они находятся в 3Д пространстве и содержат в себе явные паттерны. Обе эти особенности естественным образом ловятся человеческим глазом, однако крайне трудны в формализации для компьютера.

Любые вычислительные методы для 3Д структур проигрывают человеку в точности

Любые вычислительные методы лишь помогают человеку в генерации гипотез, а не заменяют его

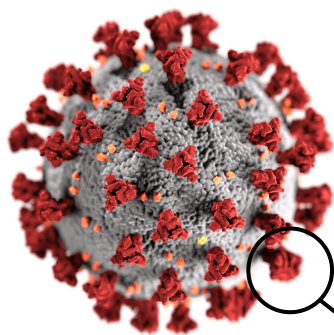
При работе с единичными белками ручной анализ возможен – и поэтому **необходим**

При работе с динамикой белка или при сравнении больших выборок ручной анализ невозможен – чисто инструментальный становится **допустим**

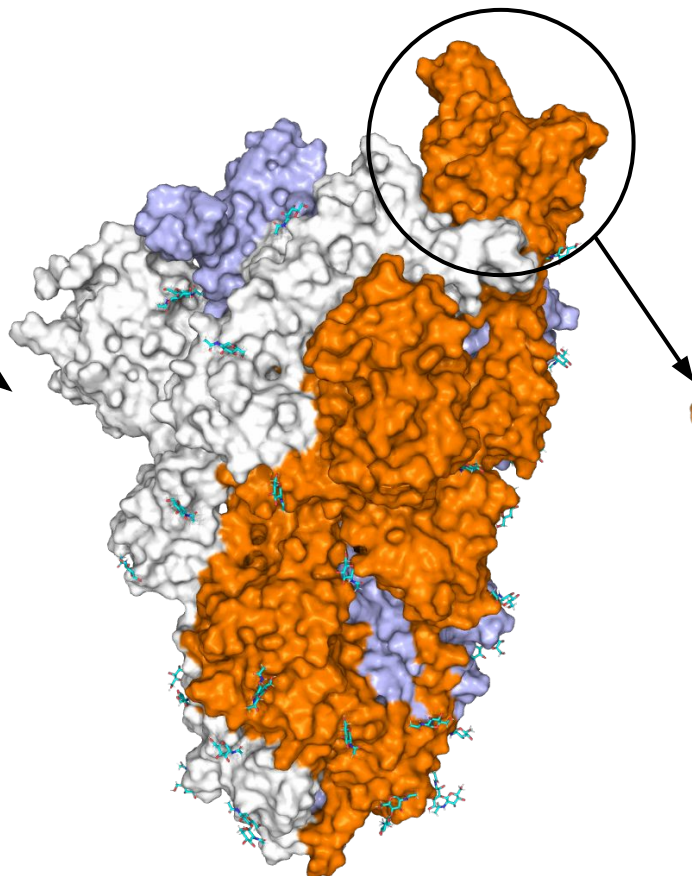
Задачи

- Объяснить эффект мутации
 - Объяснить эффект низкомолекулярного модулятора на белок
 - Объяснить поведение молекулы при изменении условий
 - Объяснить механизм работы канала/рецептора/фермента/...
-
- Спроектировать низкомолекулярные модуляторы
 - Изменить свойства канала/рецептора/фермента/...
 - Спроектировать новый канал/рецептор/фермент/...

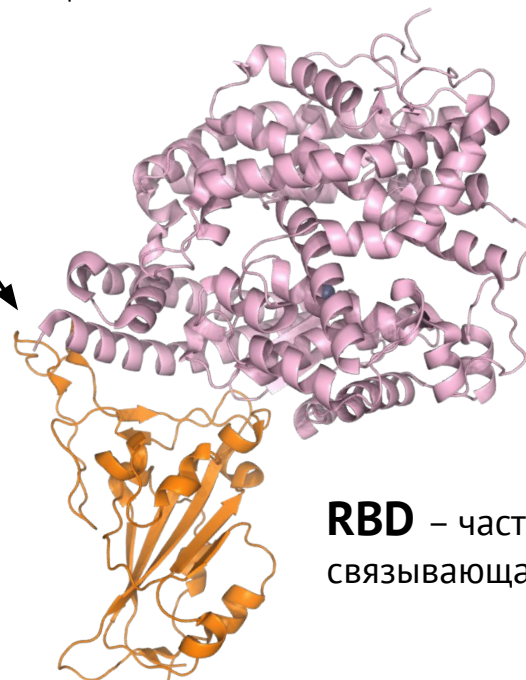
Sars-Cov-2



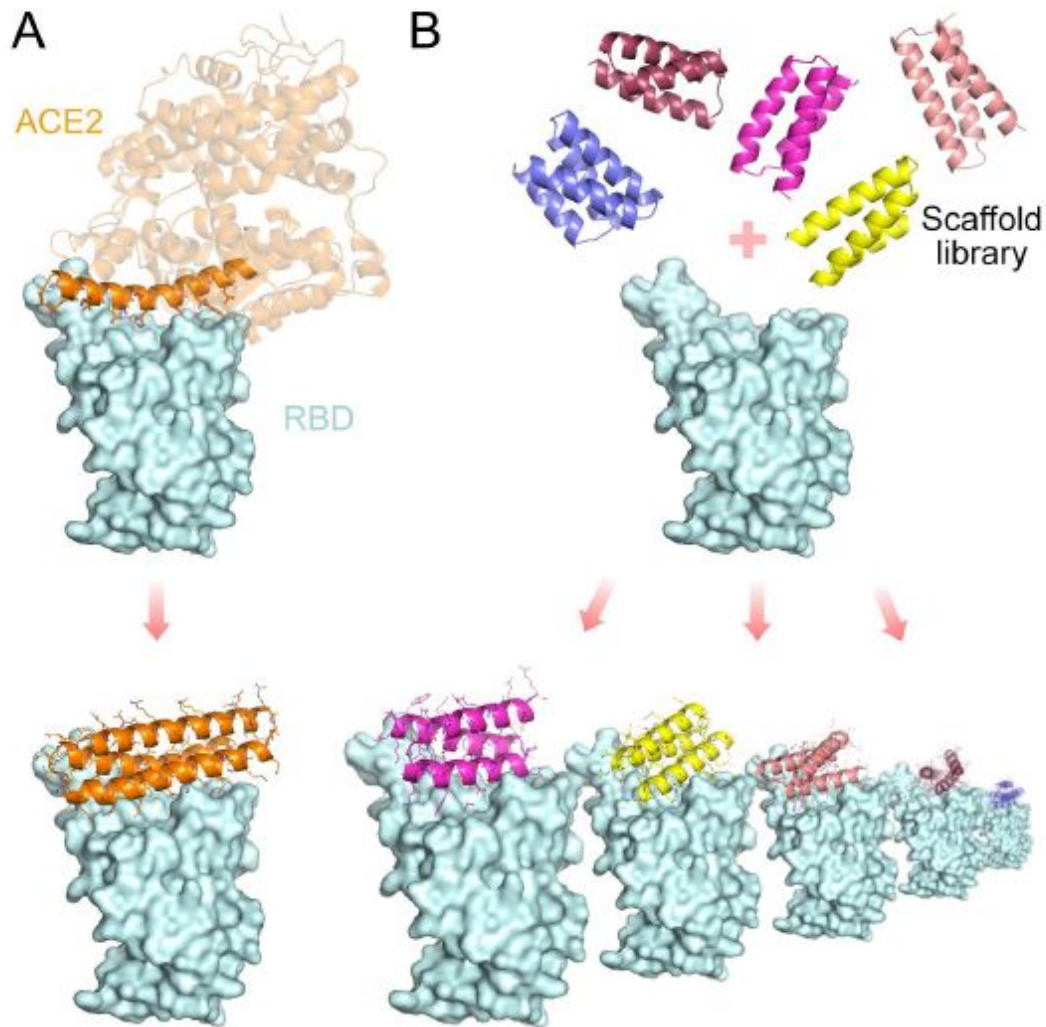
Spike protein



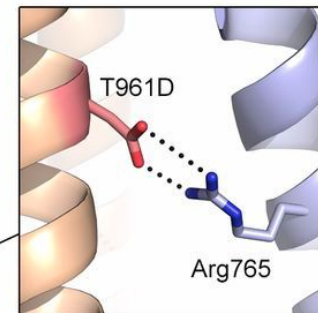
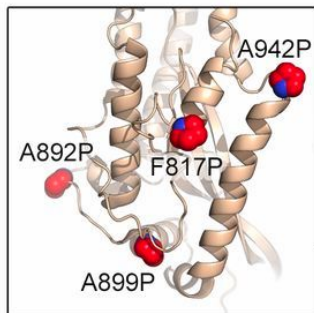
ACE2 – наш рецептор, который вирус подло использует для своих целей



RBD – часть S-белка, связывающая ACE2

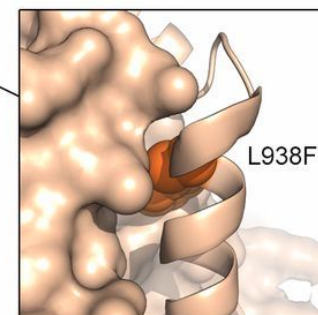
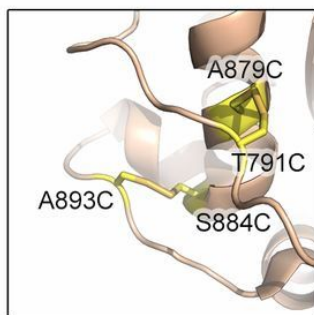


**Инженерия
искусственных
белков,
связывающих RBD
лучше, чем ACE2**

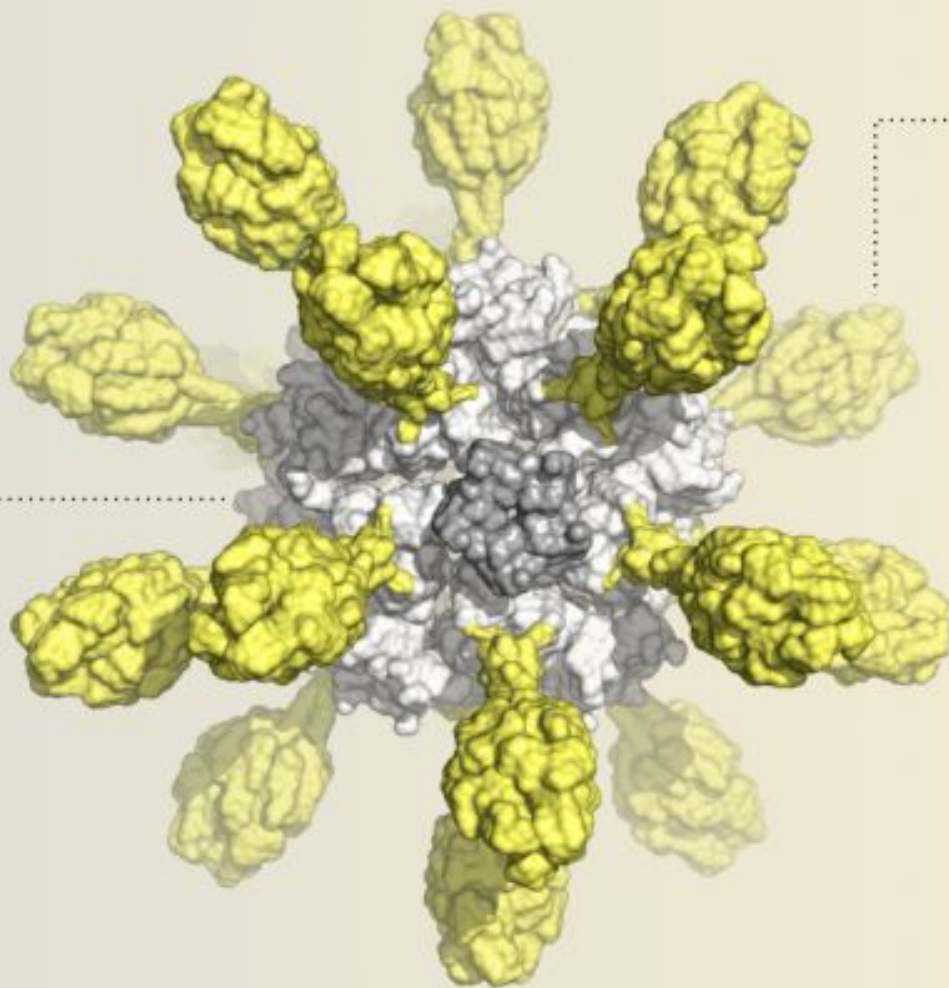


Инженерия более стабильного S-белка

DOI: 10.1126/science.abd0826

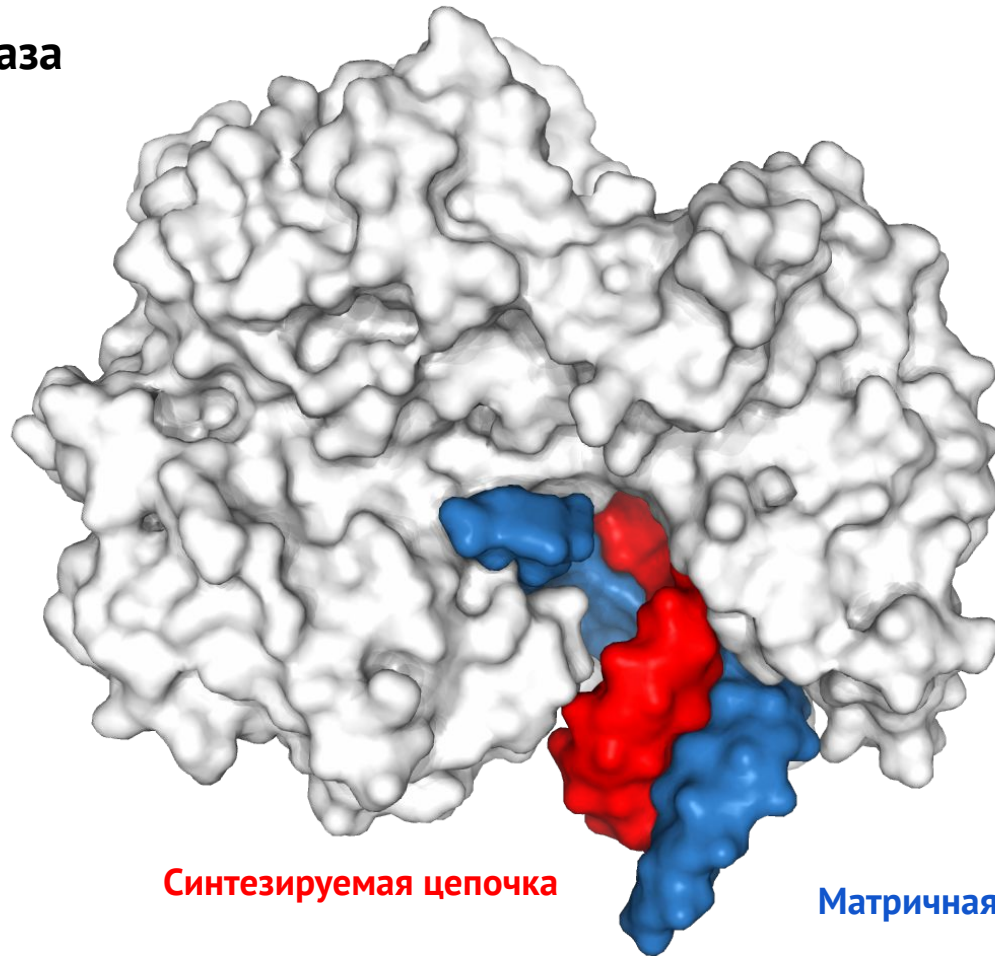


DESIGNED
NANOPARTICLE



STABILIZED
ANTIGENS

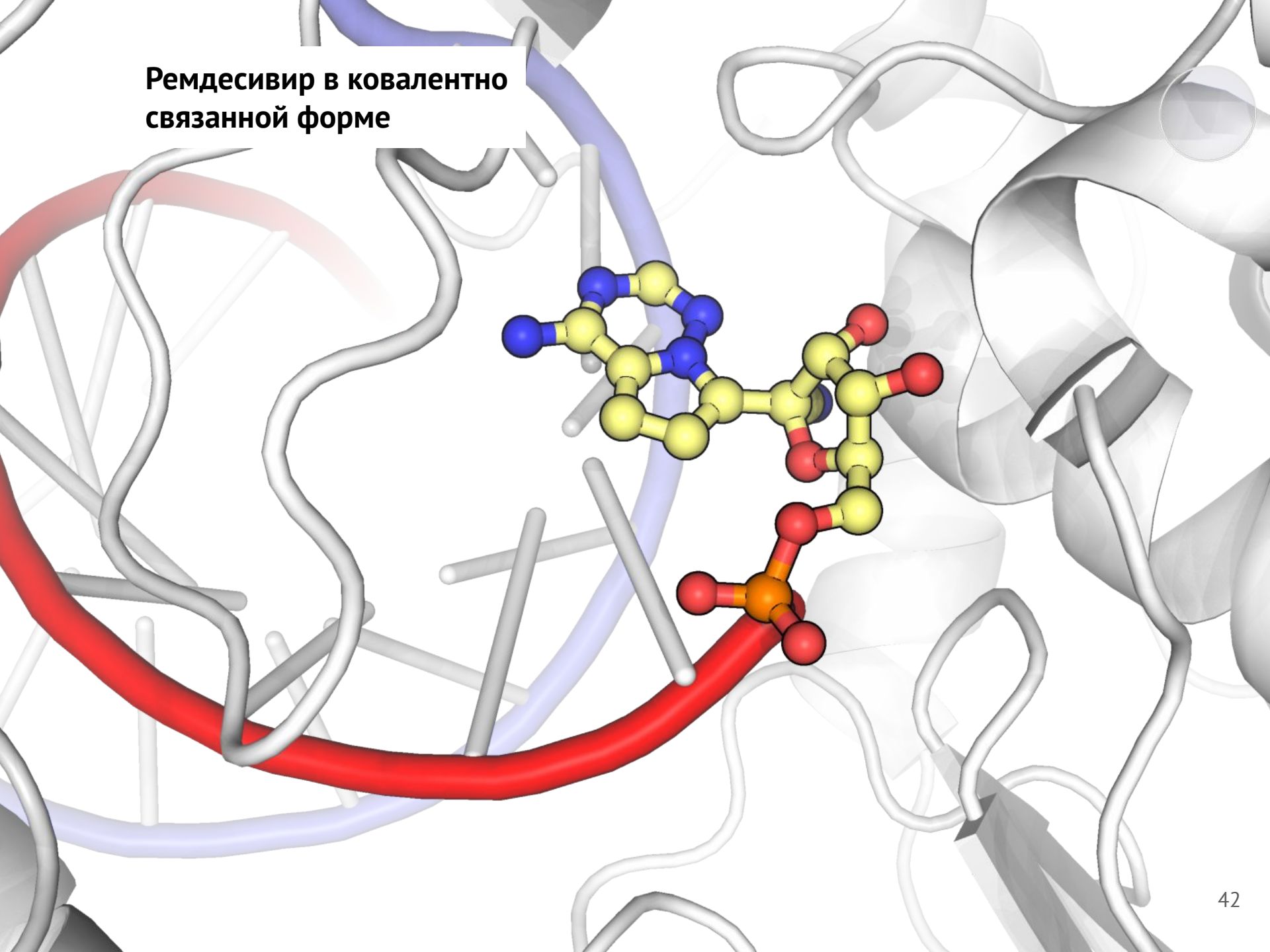
РНК-полимераза



Синтезируемая цепочка

Матричная цепочка

**Ремдесивир в ковалентно
связанной форме**



Главная протеаза коронавируса

Тут связывается белок,
который должен быть
разрезан этой
протеазой для
прохождения вируса по
его жизненному циклу



Что можно сделать?