

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



# Структурная биоинформатика: задачи и методы

Лекция 1, биоинформатика, 4 курс ФББ МГУ, осенний семестр  
Злобин А. С., [alexander.zlobin@fbb.msu.ru](mailto:alexander.zlobin@fbb.msu.ru)

# 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.

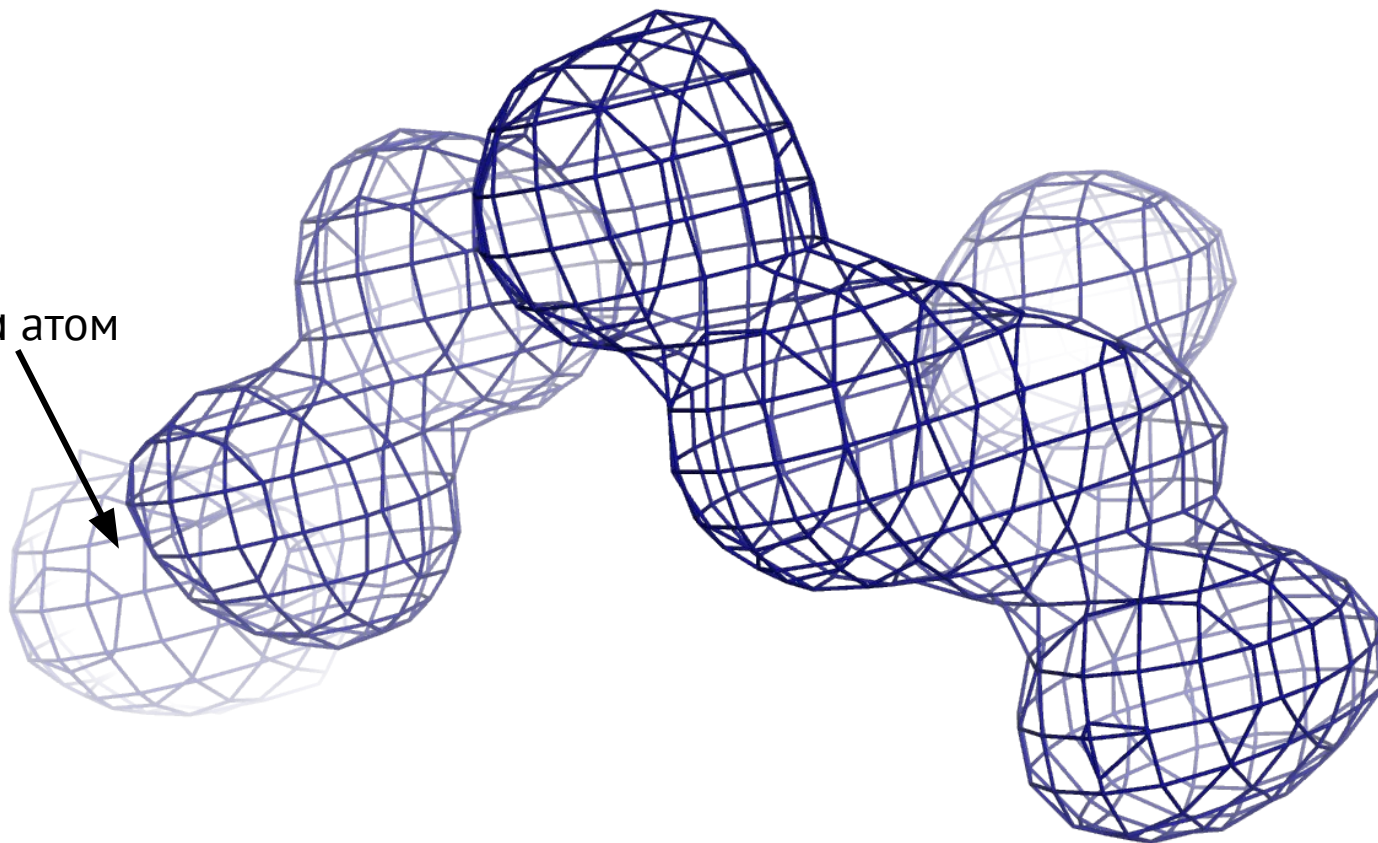
Однако кроме ограничений из эксперимента, у нас есть дополнительные ограничения из природы:

- 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

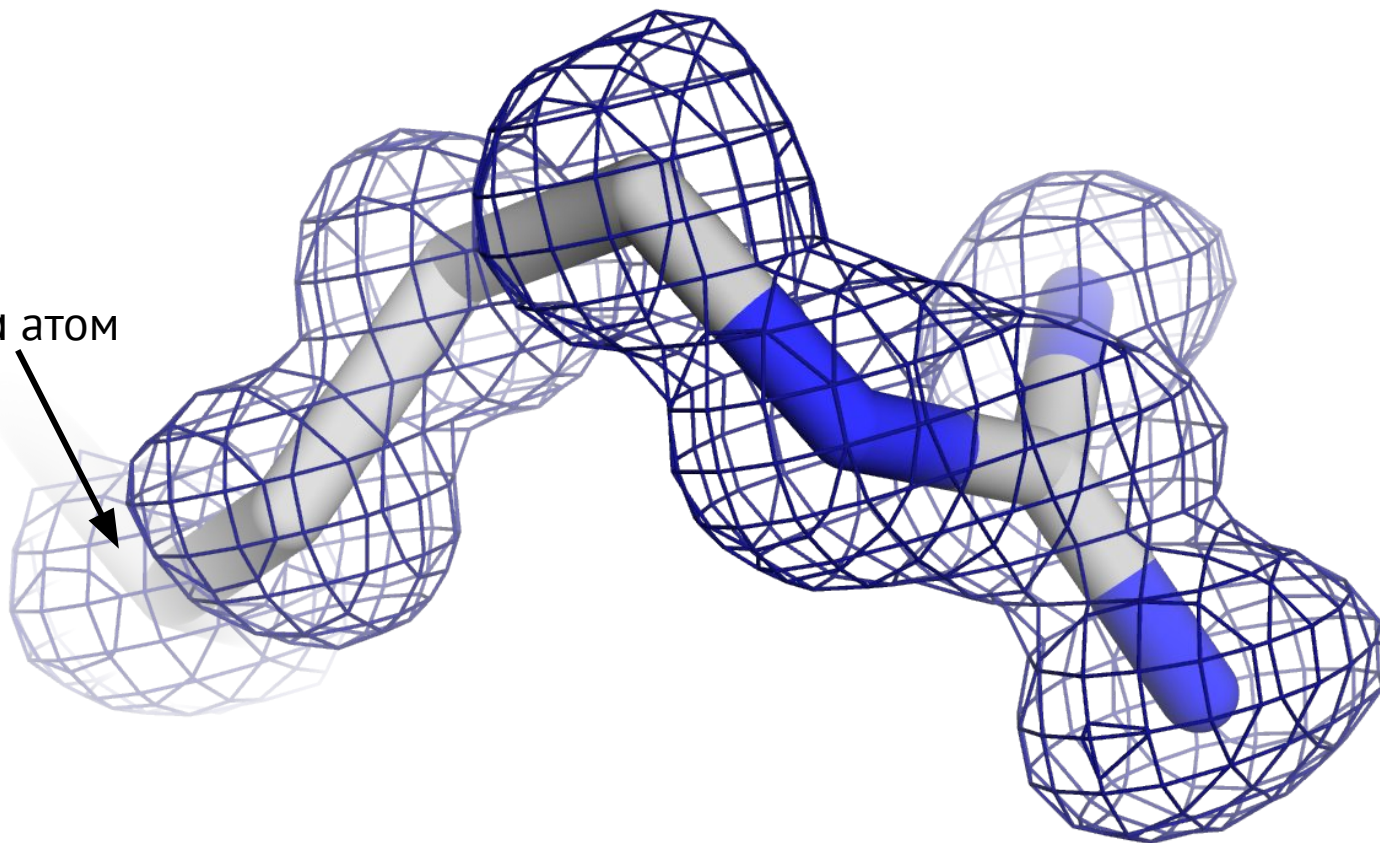
Это Ca атом



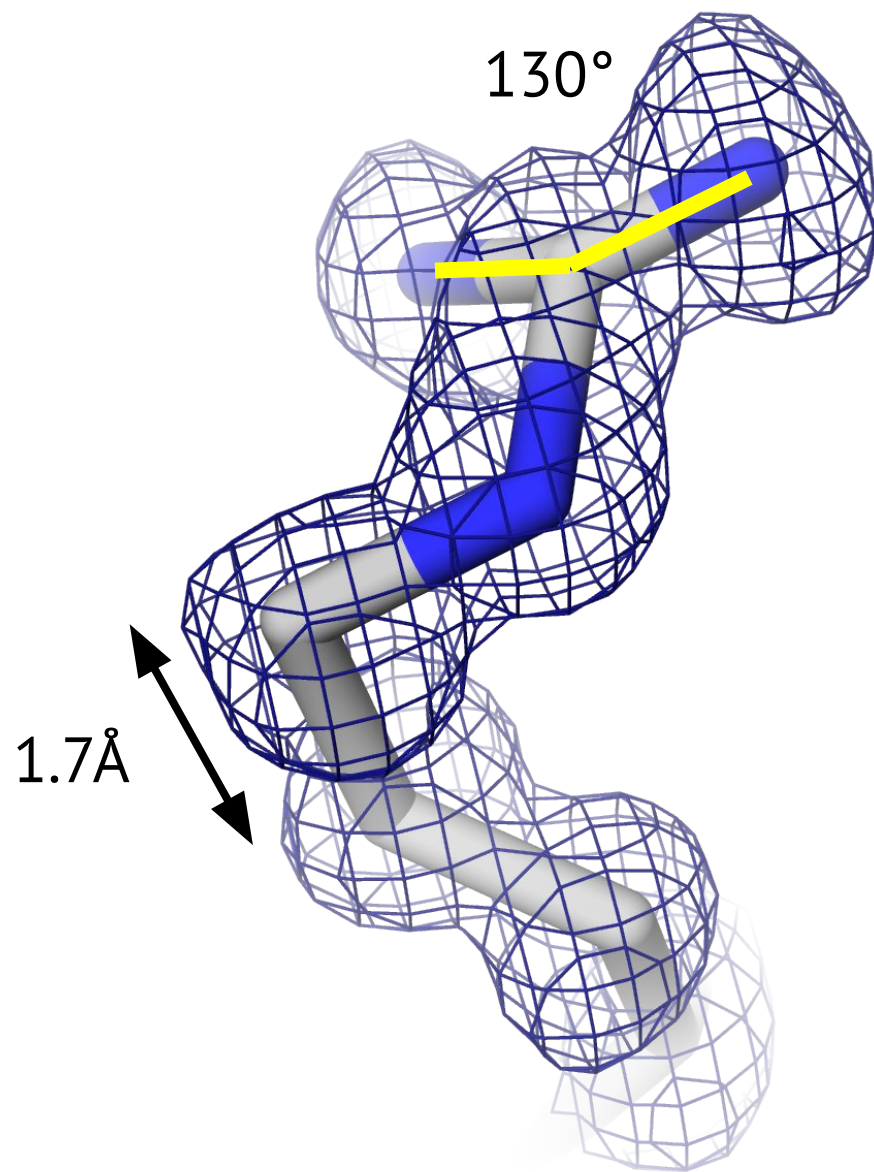
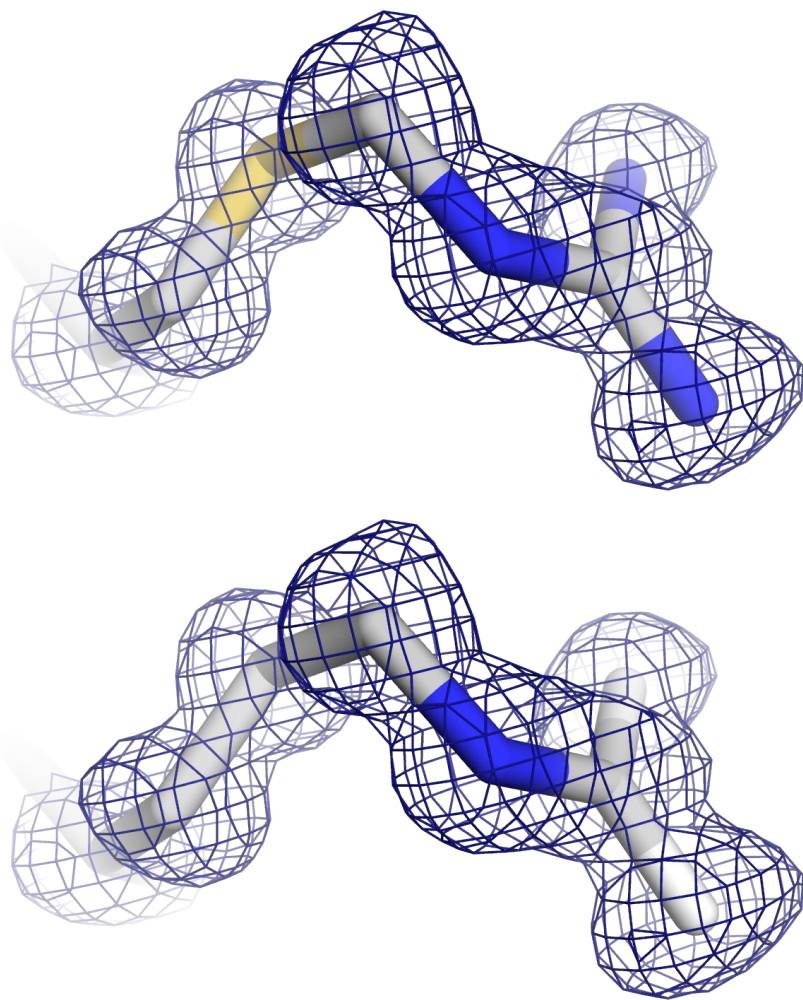
Данные РСА и КристоЭМ – области пространства с повышенной электронной плотностью. Боковой радикал какой аминокислоты хорошо соответствует этой плотности?

# РСА и КриоЭМ

Это Са атом

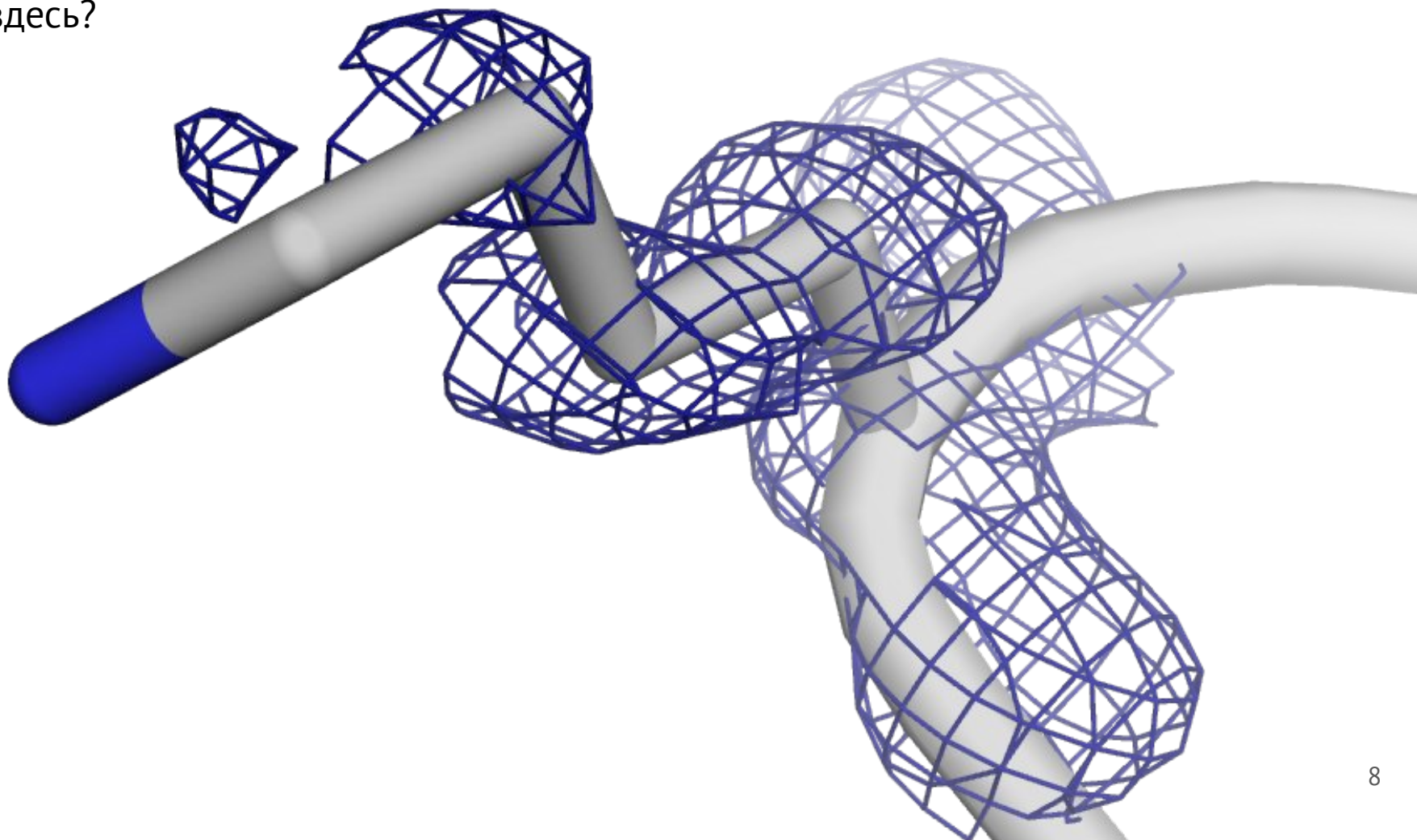


# Почему не так?



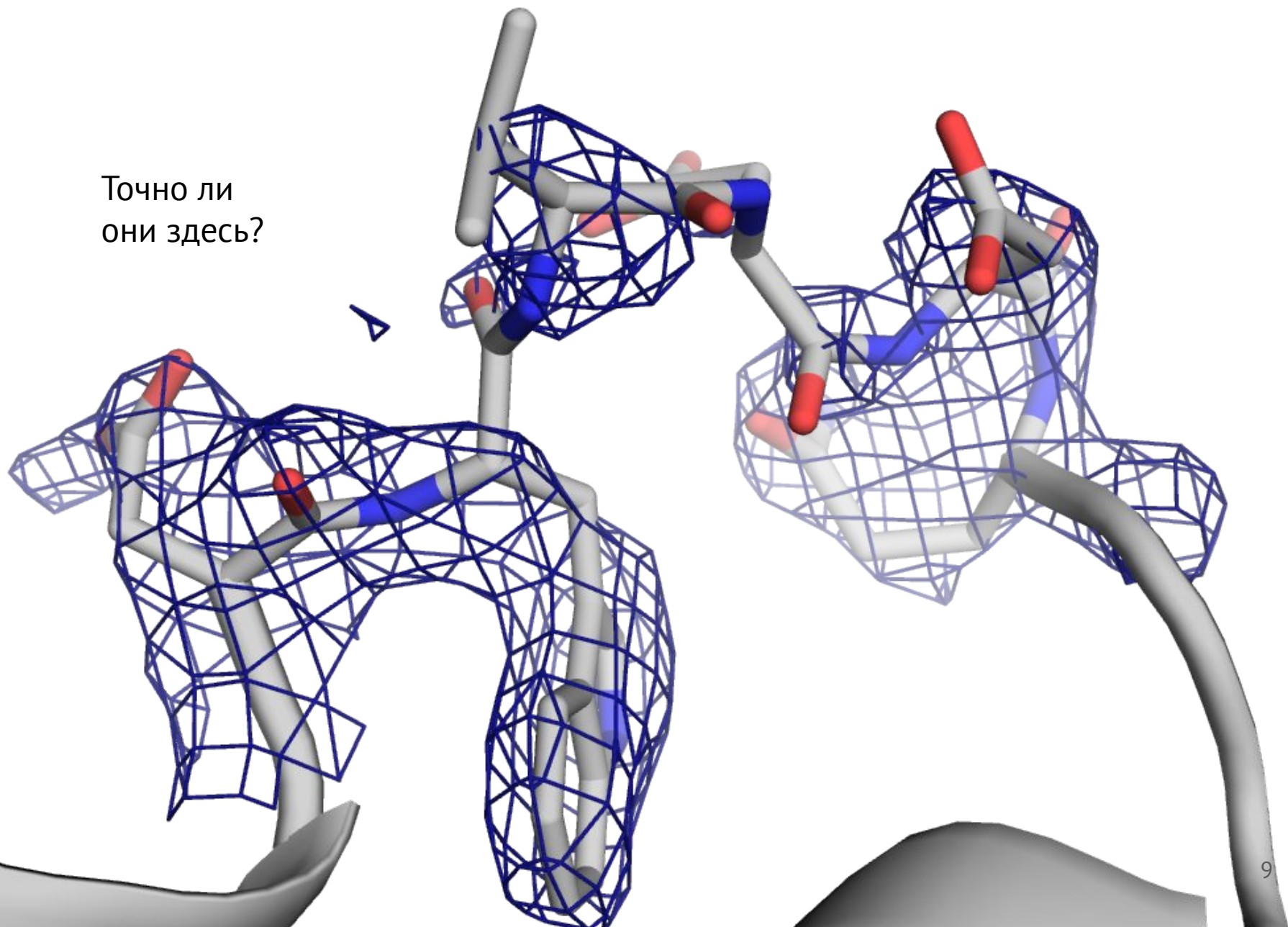
# Не всегда все так хорошо

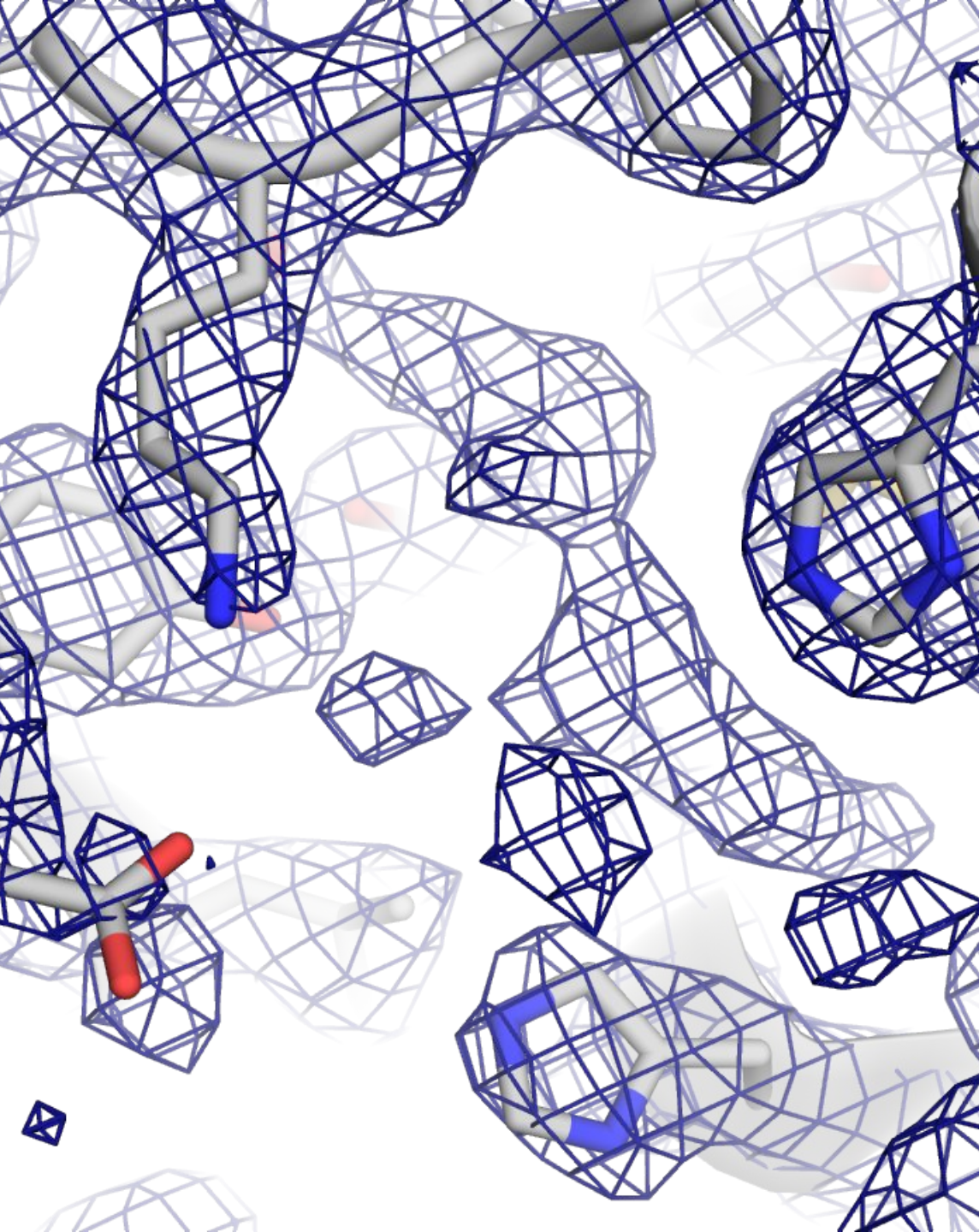
Точно ли  
он здесь?





Точно ли  
они здесь?





Что это?

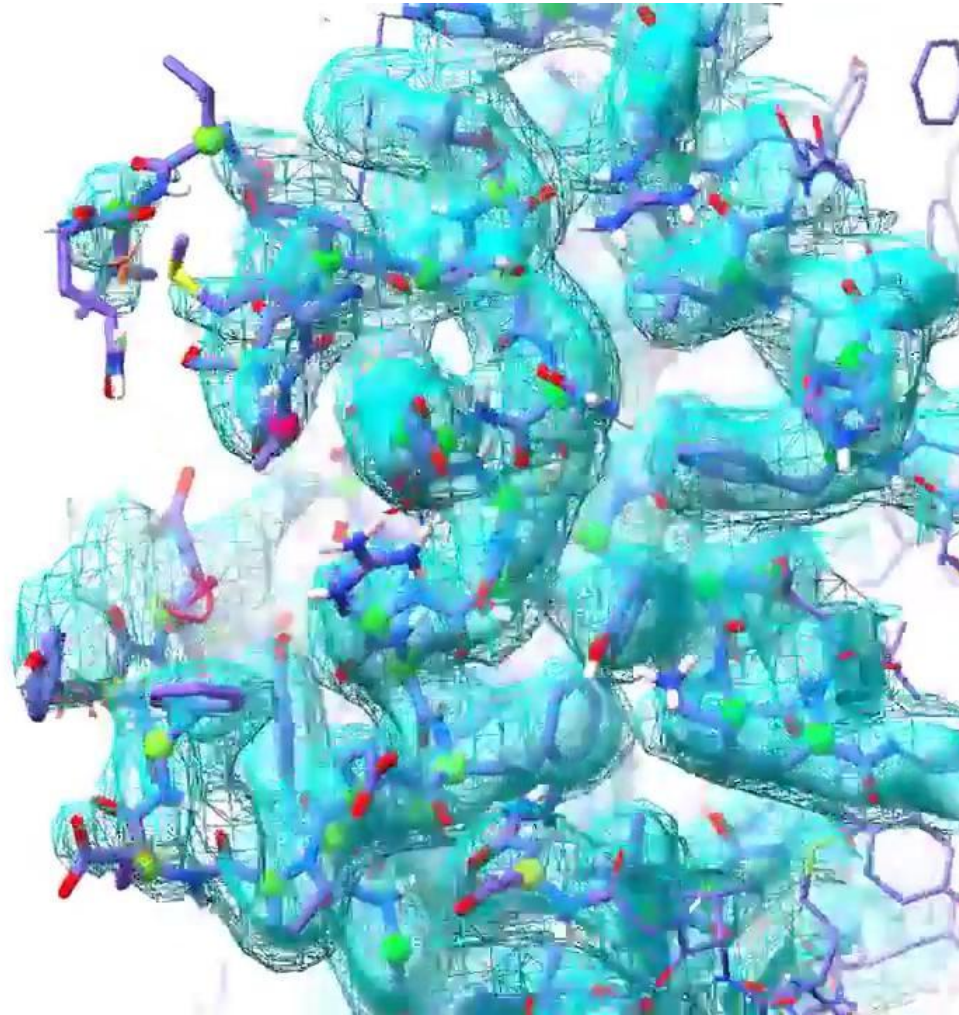
# Иногда все очень плохо



Tristan Croll  
@CrollTristan

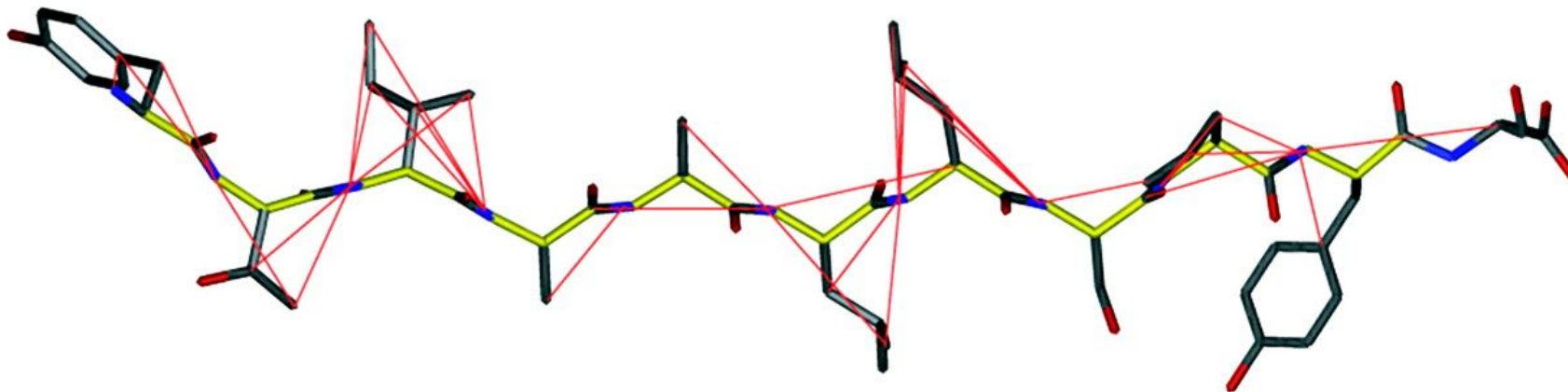


Oh, wow - looks like they're all out by \*nine\* residues (Pro 918 is where Pro 927 should be; Leu 907 is where Trp 916 should be). Would \*not\* have been easy to pick this in the original 6nur, but 7btf is both higher resolution and has slightly more "tail".

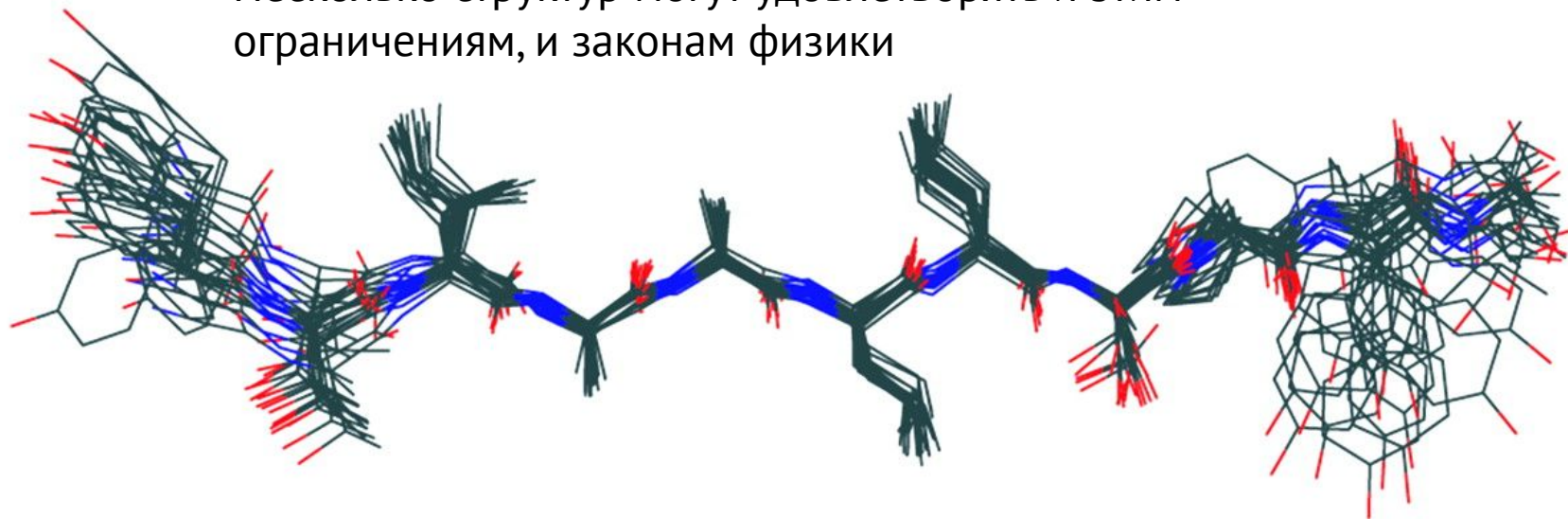


# ЯМР

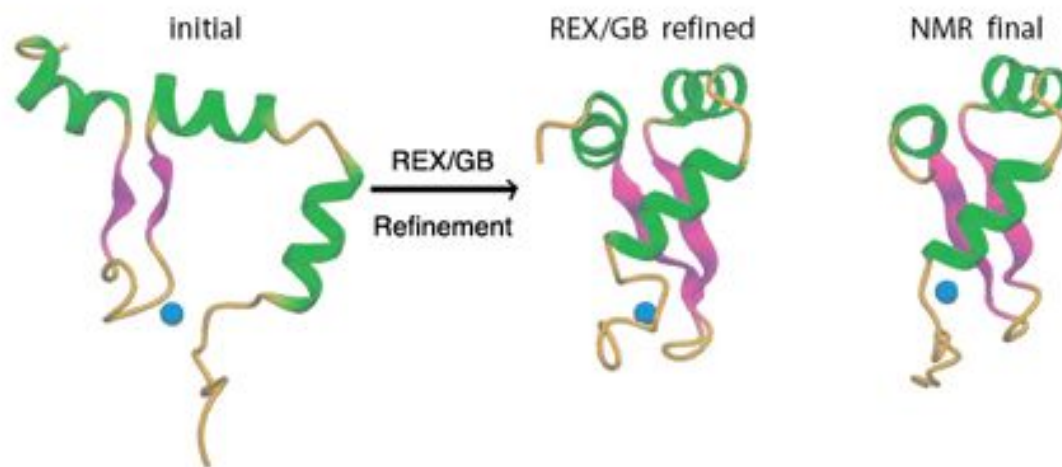
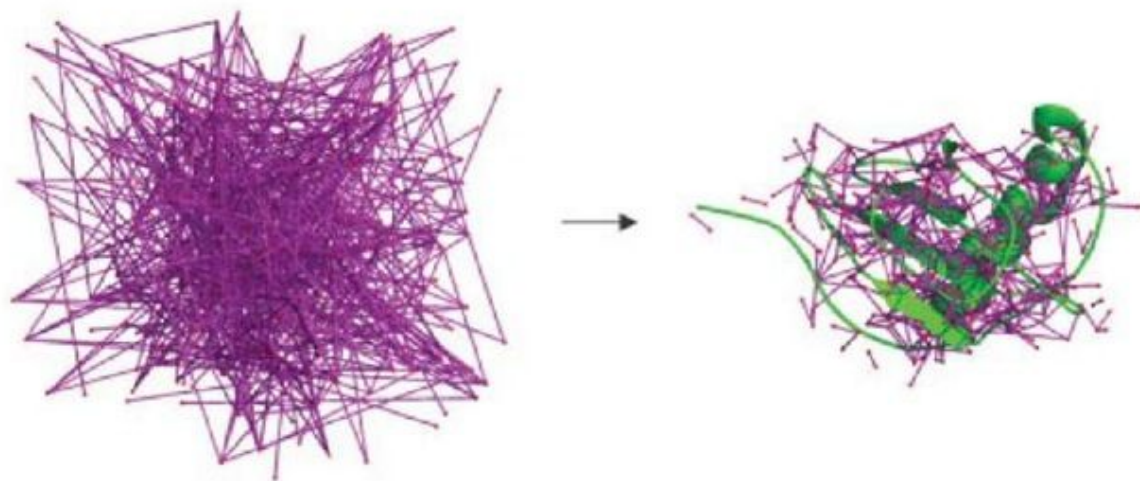
Данные ЯМР – набор ограничений на расстояния



Несколько структур могут удовлетворять и этим ограничениям, и законам физики



# ЯМР



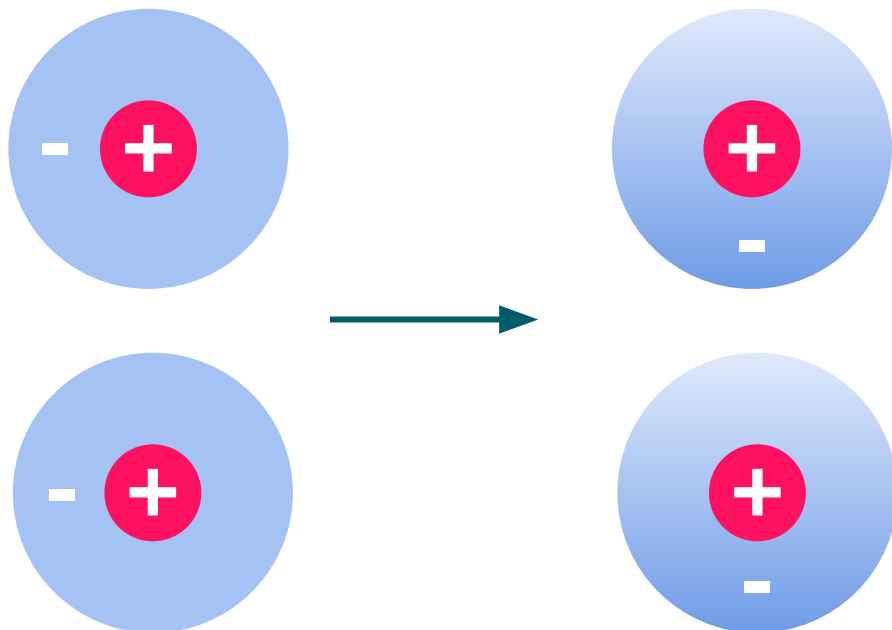
# Взаимодействия, определяющие структуру

*Можем ли мы верить полученной структуре? Почему эти структуры такие?*

*Какими структуры вообще могут быть, а какими не могут?*

Наведенный диполь и наведенный диполь	<b>Ne</b>	<b>Ne</b>	0.05-40 (кДж/моль)
Диполь и наведенный диполь	<b>H<sub>2</sub>O</b>	<b>O<sub>2</sub></b>	2-15
Диполь и диполь	<b>C=O</b>	<b>C=O</b>	5-25
Водородная связь	<b>H<sub>2</sub>O</b>	<b>H<sub>2</sub>O</b>	10-40
Ион и диполь	<b>Na<sup>+</sup></b>	<b>C=O</b>	40-600
Ион и ион	<b>R-NH<sub>3</sub><sup>+</sup></b>	<b><sup>-</sup>OOC-R</b>	400-4000
Ковалентная связь	<b>C-O</b>		150-1200

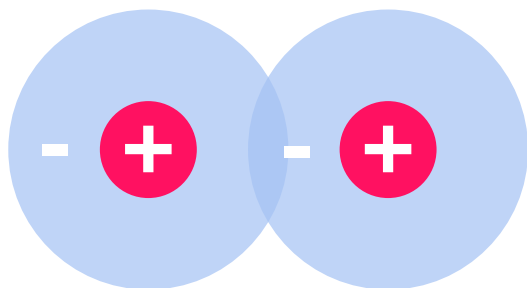
## Наведенные диполи – притяжение

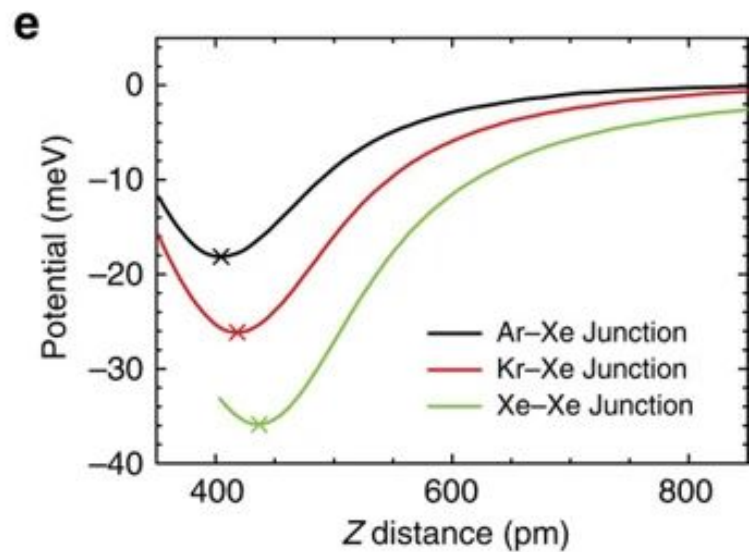
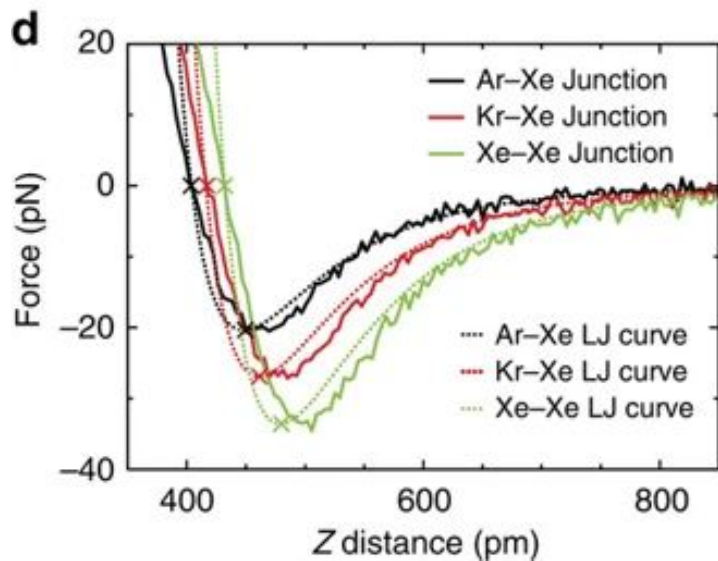
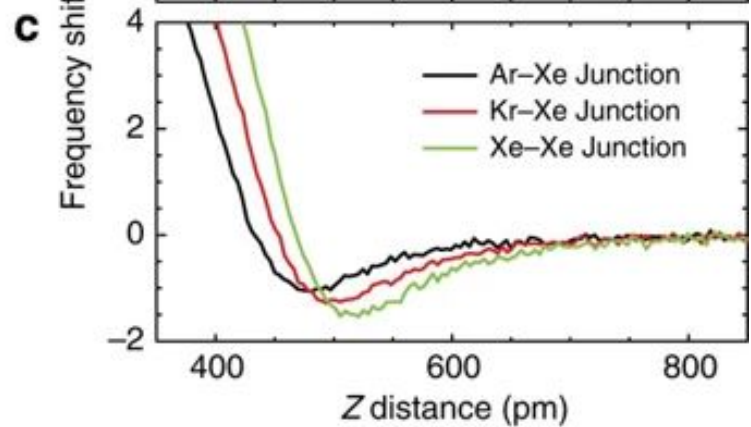
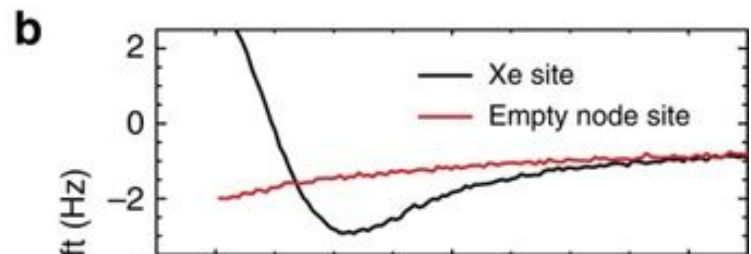
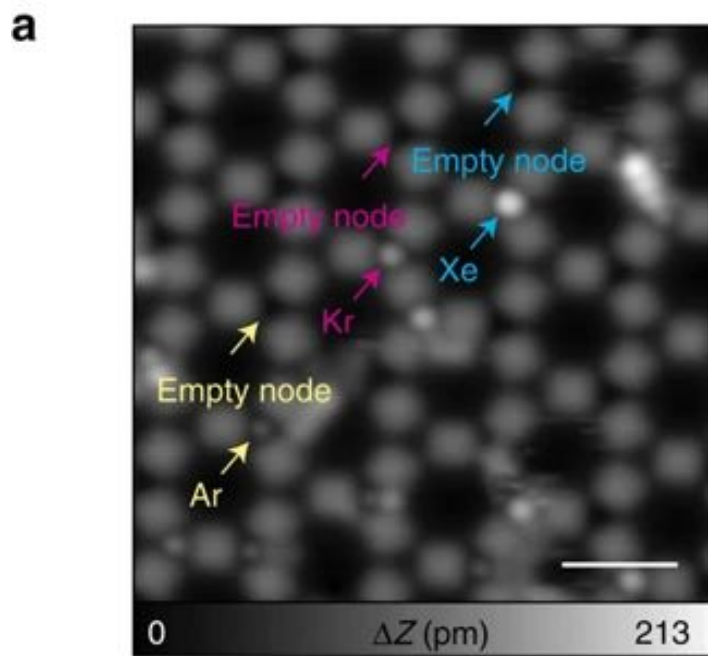


Дисперсионные взаимодействия  
(наведенный диполь и наведенный диполь)

В растворе часто можно пренебречь

## Электронные оболочки – отталкивание

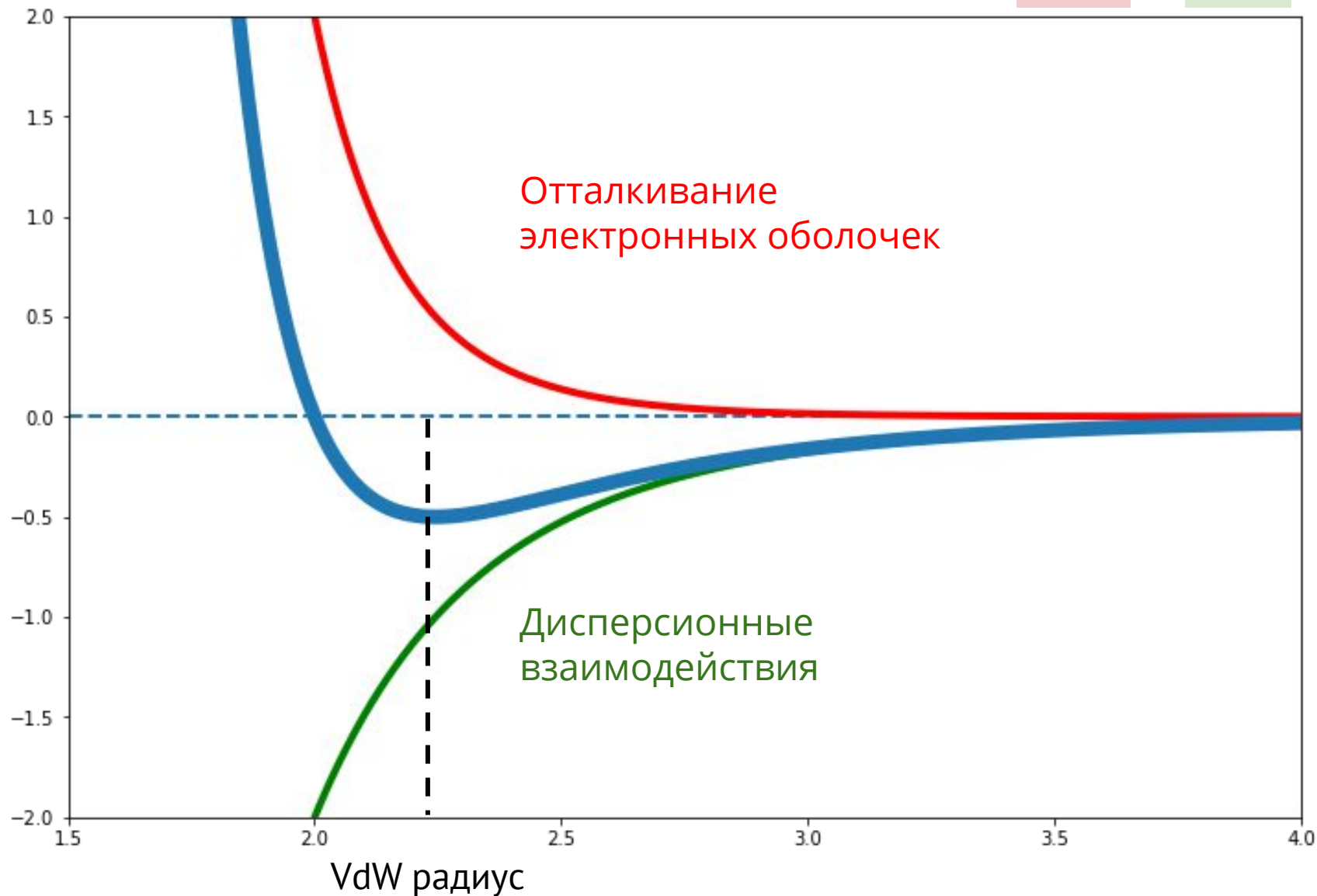






# Потенциал Леннарда-Джонса

$$V_{LJ} = 4\epsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right]$$



# VdW – следствия

- Перекрывания VdW радиусов в структурах – грубейшая ошибка
- Пустоты тоже вызывают вопросы. Даже в вакууме в отсутствие растворителя структура будет стремиться к компактности. Однако тут не все так однозначно. Вероятно, иногда отбор по каким-то причинам действует в сторону сохранения пустот (How Nothing Boosts Affinity: Hydrophobic Ligand Binding to the Virtually Vacated S1' Pocket of Thermolysin, JACS, 2017)

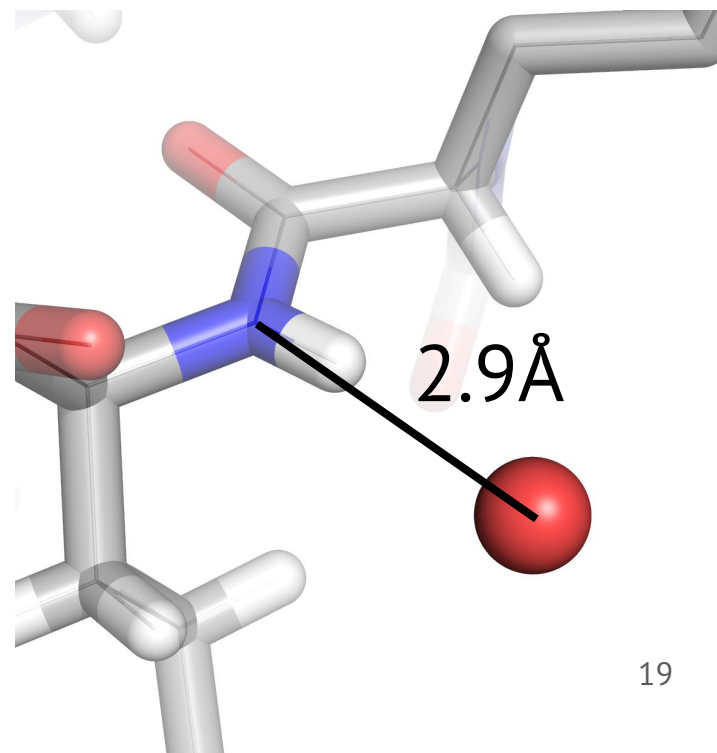
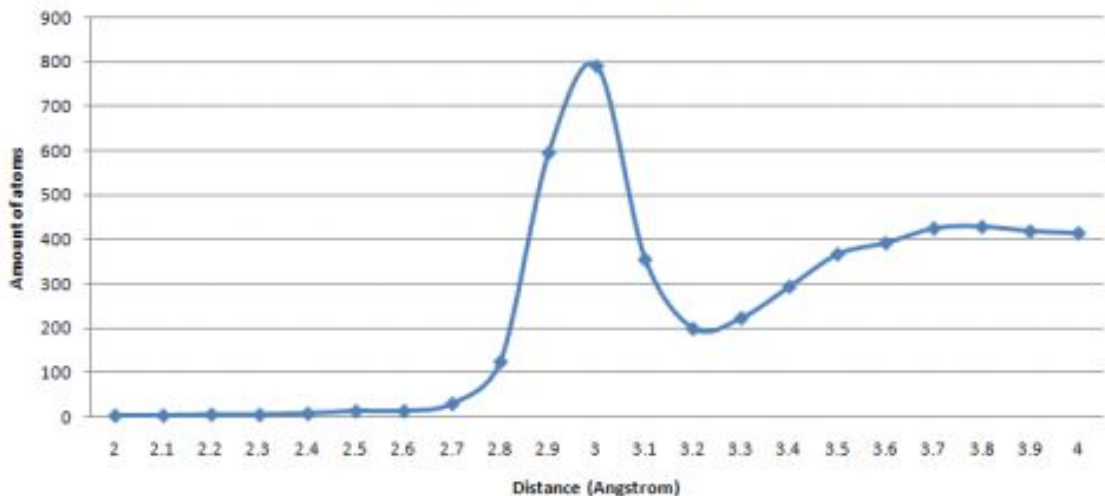
Author, year	R, Å								
	H	F	Cl	Br	I	O	S	N	C
Pauling, 1939	1.2	1.35	1.80	1.95	2.15	1.40	1.85	1.5	1.70
Bondi, 1964	1.20	1.47	1.75	1.85	1.98	1.52	1.80	1.55	1.70
Zefirov, 1974	1.16	1.40	1.90	1.97	2.14	1.29	1.84	1.50	1.71
Gavezzotti, 1983–1999	1.17	1.35	1.80	1.95	2.10	1.40	1.85	1.50	1.70
Batsanov, 1995			1.80	1.90	2.10	1.51	1.80		1.68
Wieberg, 1995		1.5	1.8	1.9	2.1	1.5	1.8	1.6	1.7
Rowland, 1996	1.10	1.46	1.76	1.87	2.03	1.58	1.81	1.64	1.77

# Водородная связь

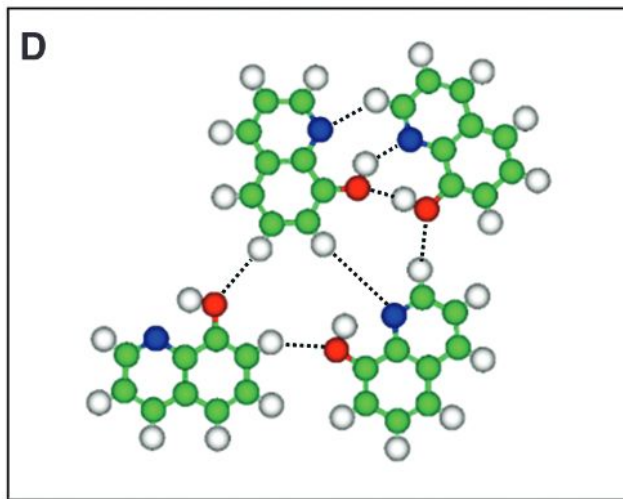
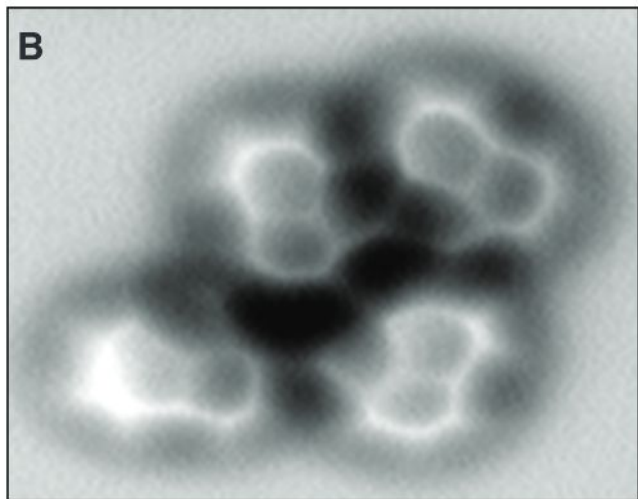
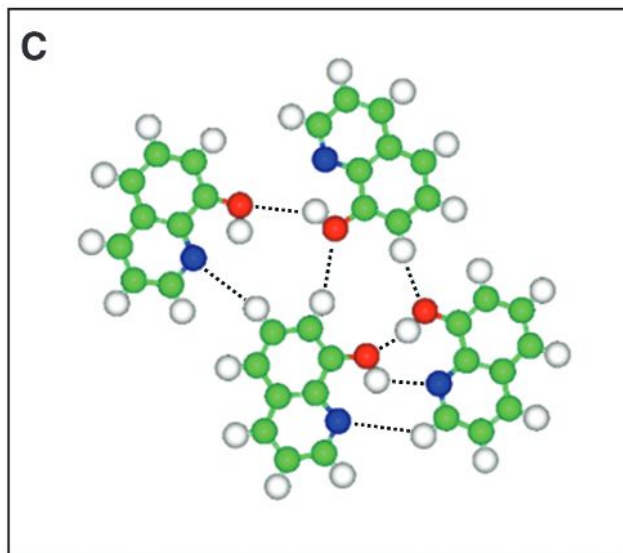
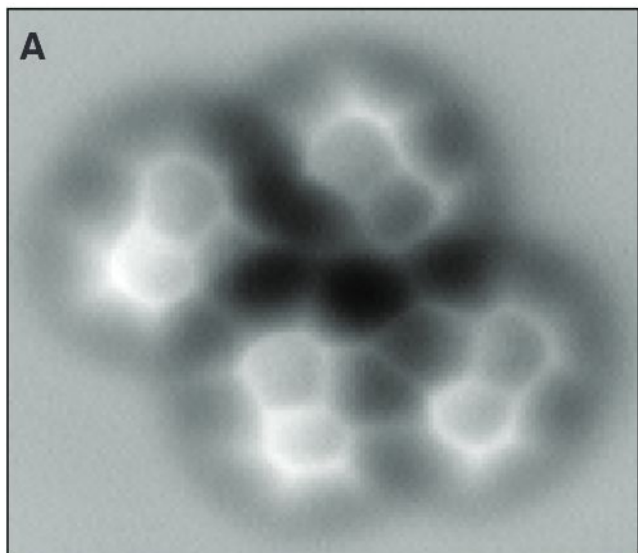
The hydrogen bond is **an attractive interaction** between a hydrogen atom from a molecule or a molecular fragment X–H in which X is more electronegative than H, and an atom or a group of atoms in the same or a different molecule, in which **there is evidence of bond formation**.

Один из не прямых источников экспериментальной информации – изучение кокристаллизованных молекул воды

Distribution of atoms amount by the distance  
HOH.O and backbone N  
At all 72436 atoms



# Водородная связь



Атомно-силовая  
микроскопия –  
прямой evidence  
of bond formation

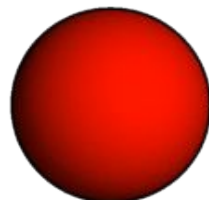
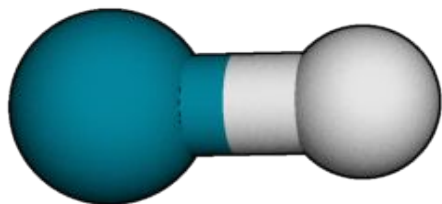
*Zhang et. al, Science,  
2013*

# Водородная связь

$\delta^-$

$\delta^+$

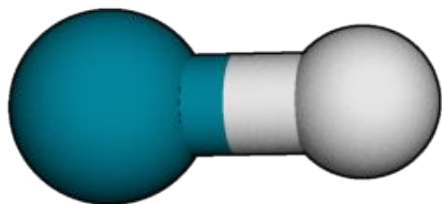
$\delta^-$



Электростатический  
компонент

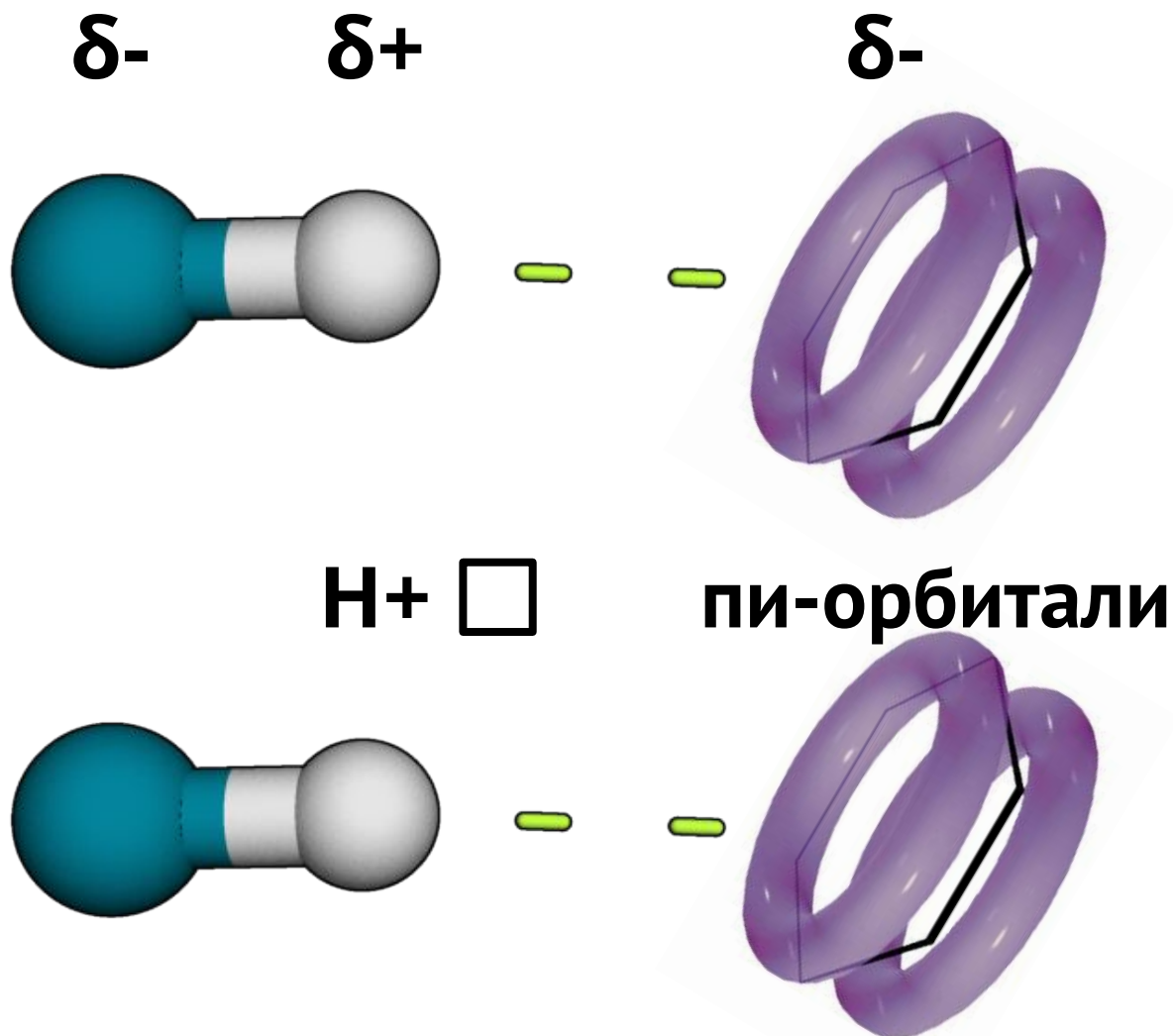
$H^+$   $\square$

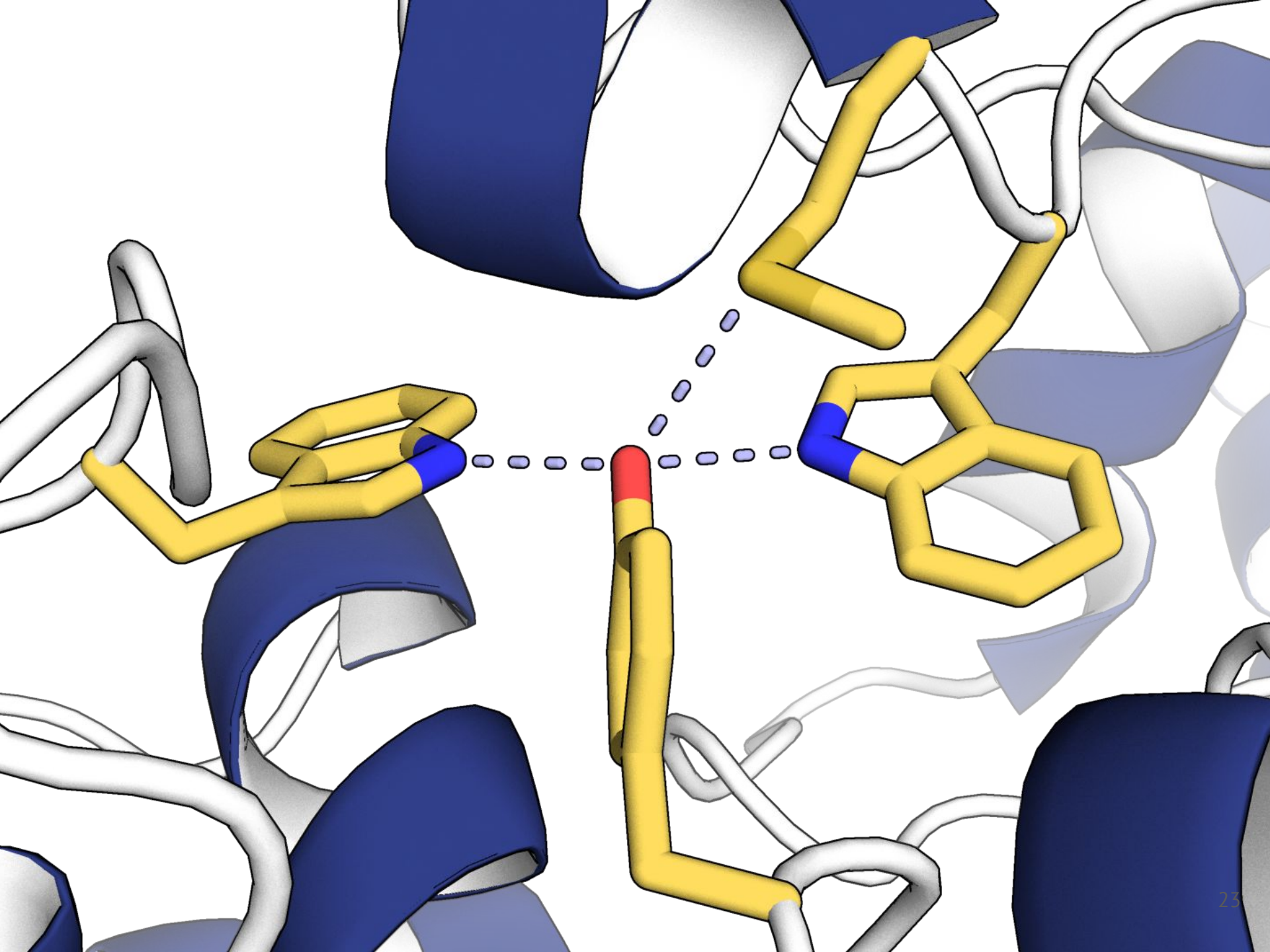
:

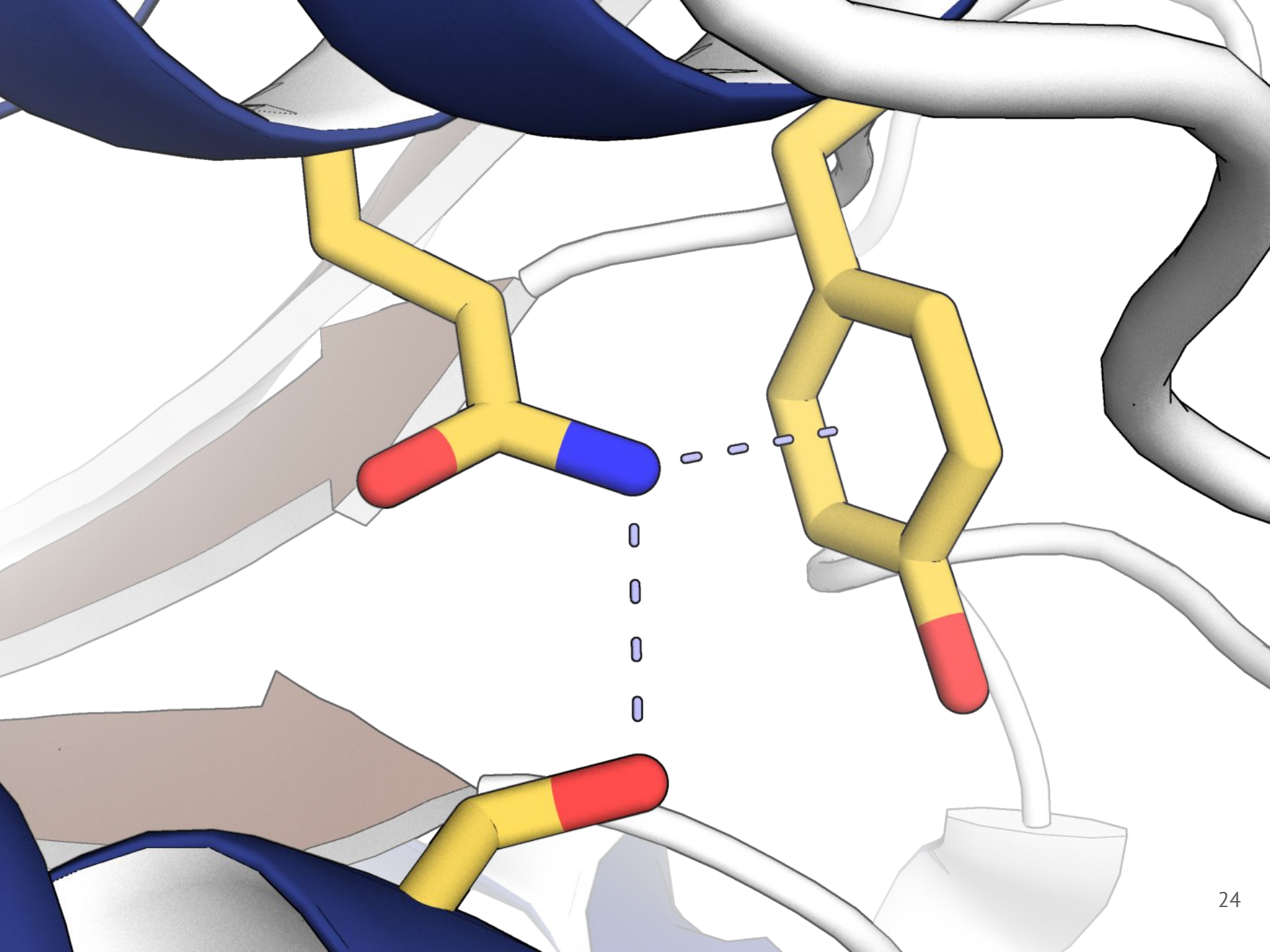


Донорно-акцепторный  
компонент

# Пи-водородная связь

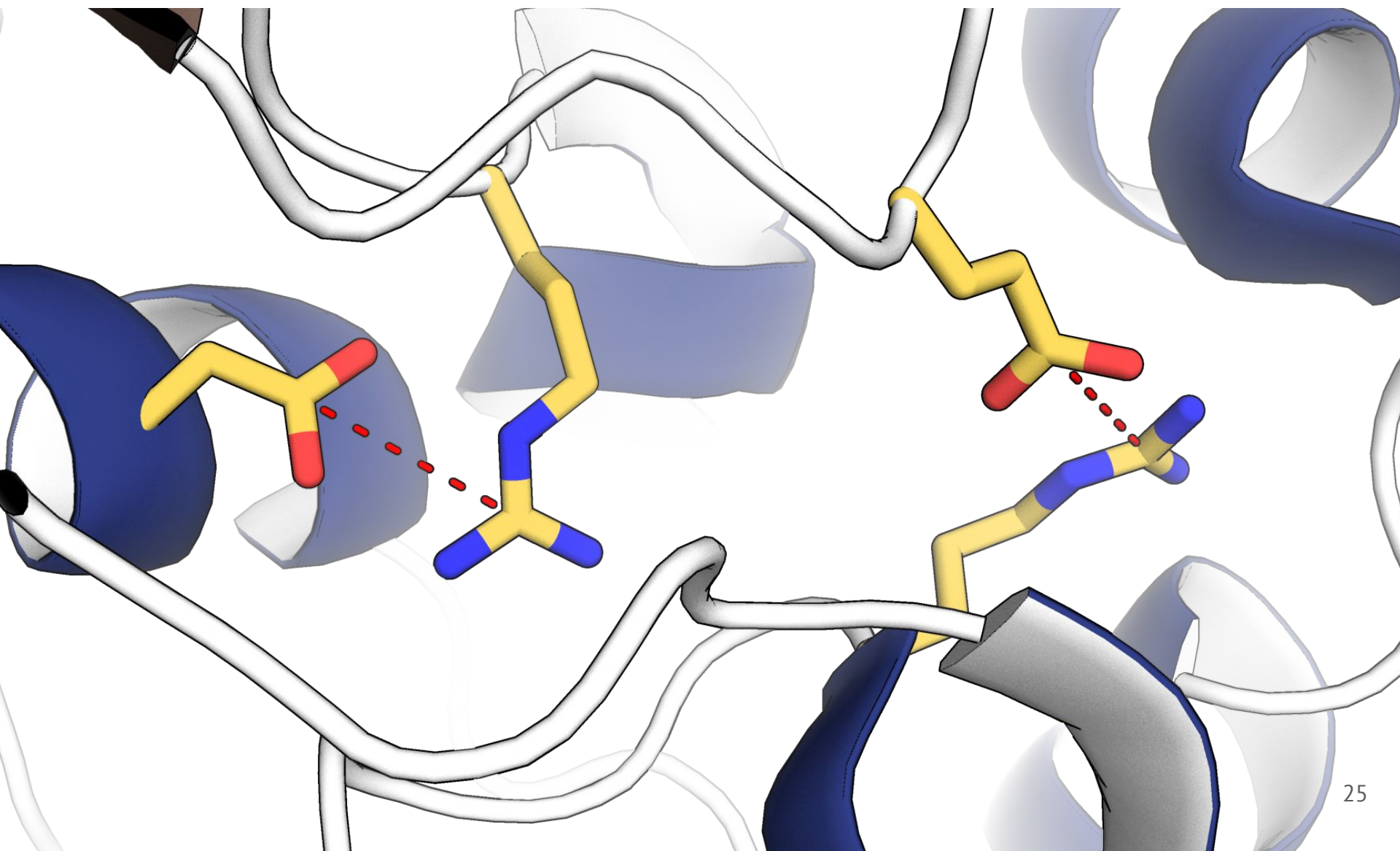




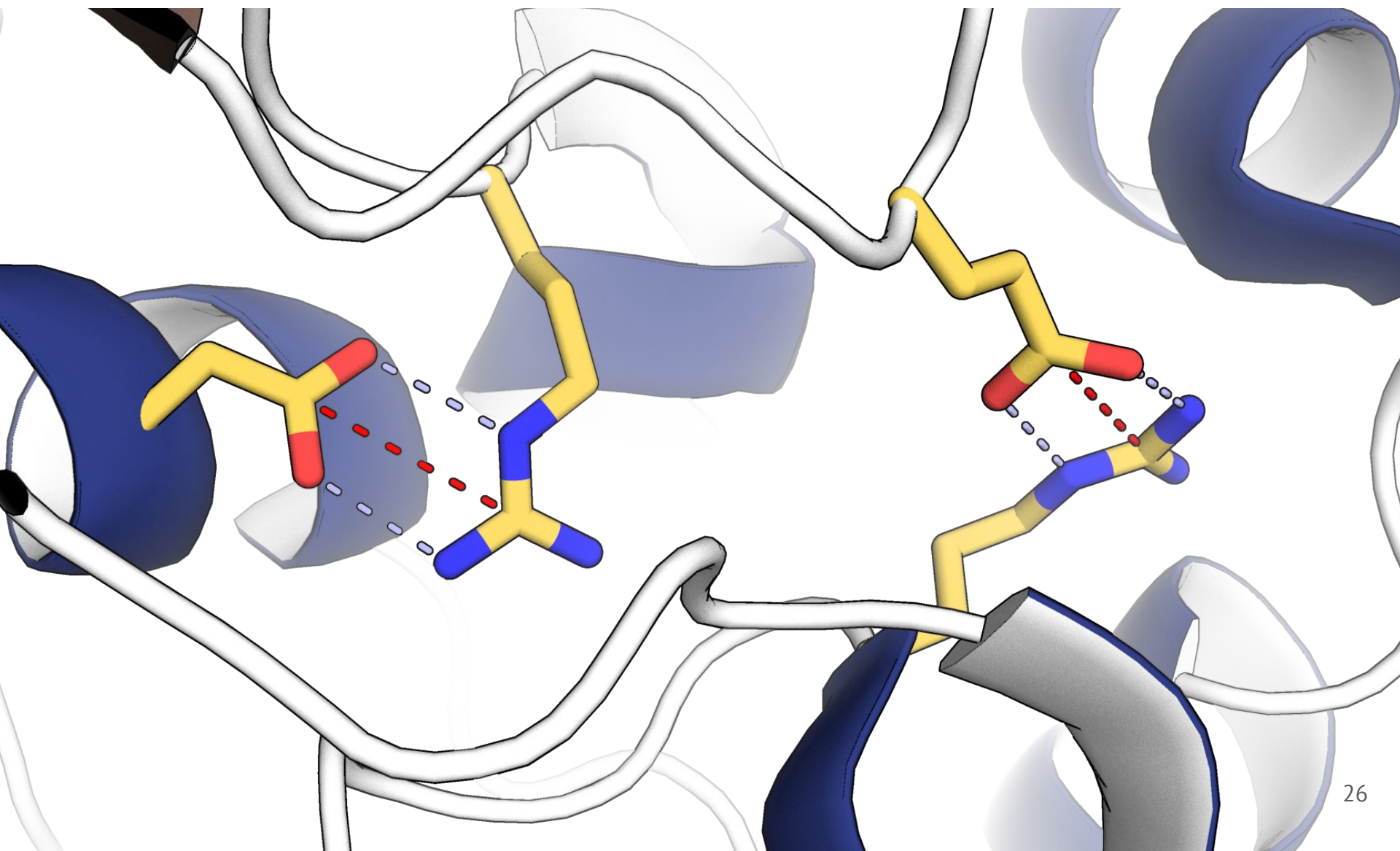


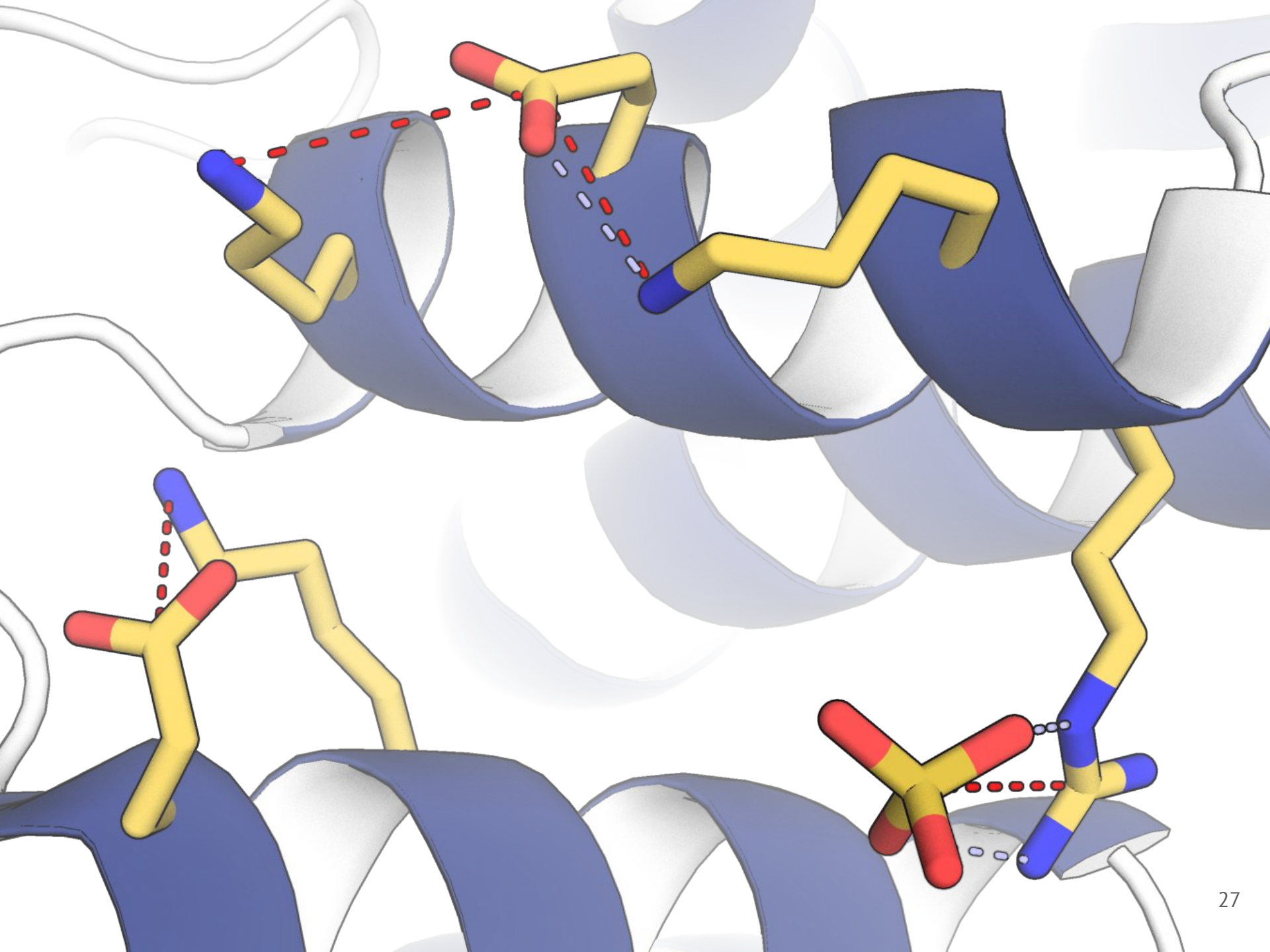


# Электростатика

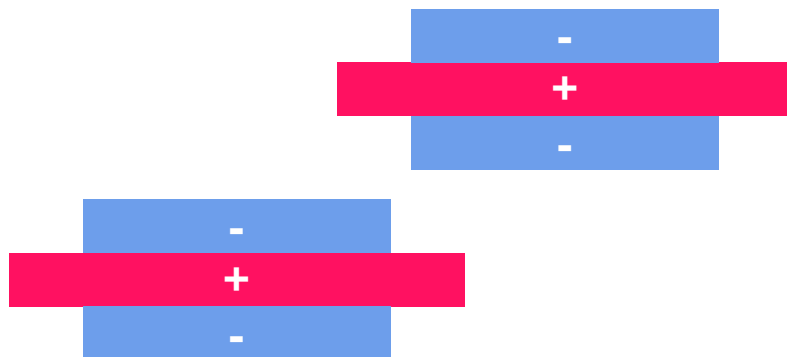


# Электростатика





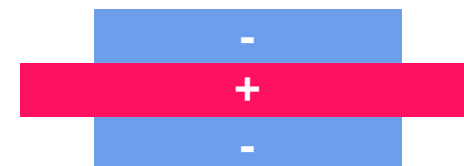
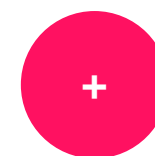
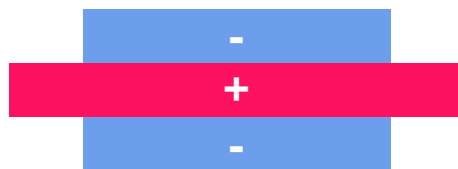
# Стэкинг



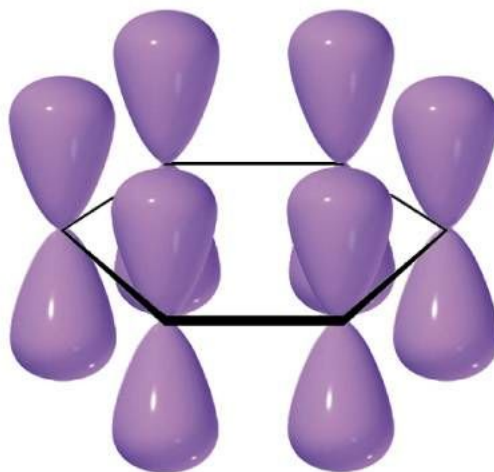
Параллельно сдвинутый



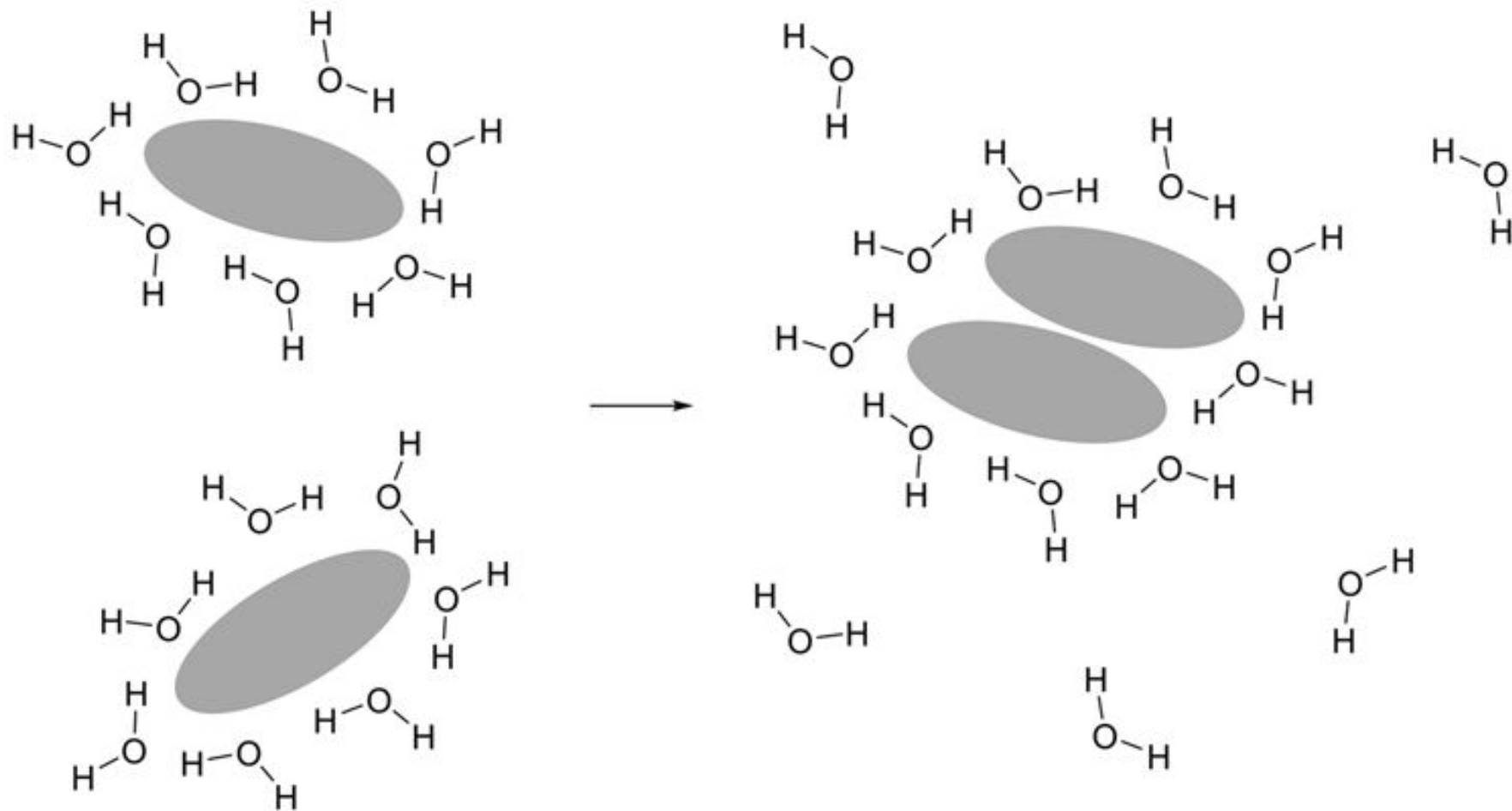
T-стекинг



$\pi$ -катионный



# Гидрофобный эффект



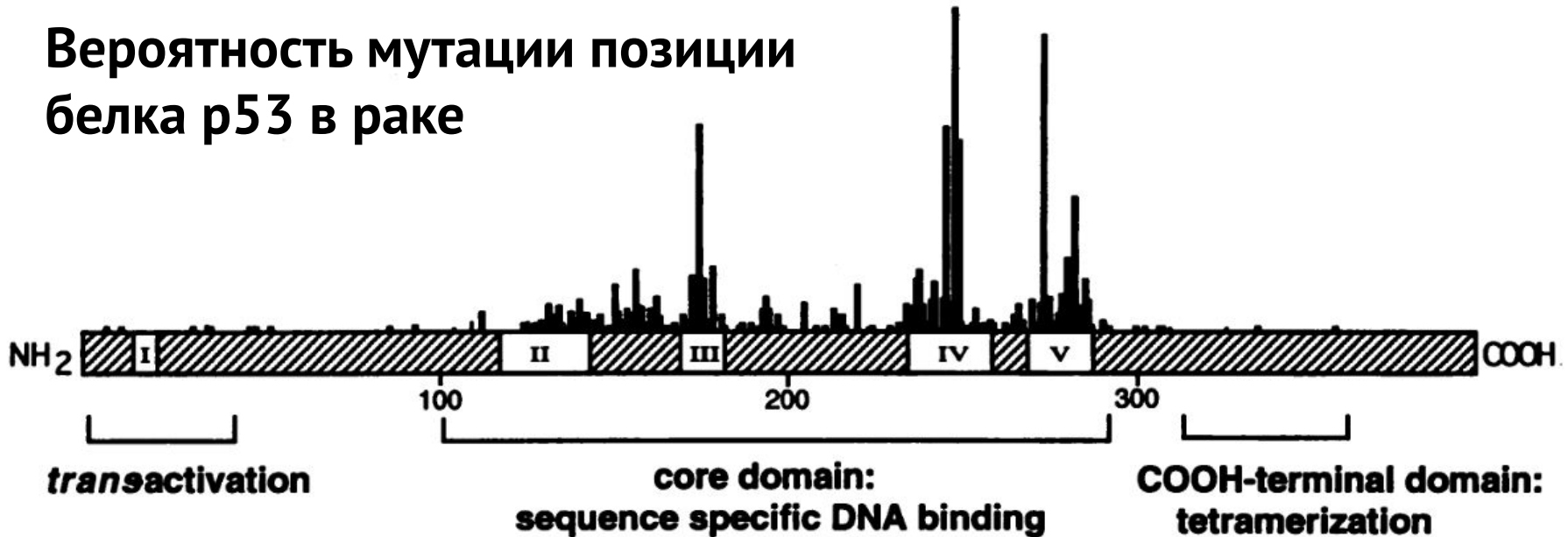
# Зачем мы получили структуру?

## Функция макромолекул “защита” в их структуре и динамике

Расположение атомов в пространстве – самая низкоуровневая реализация естественного отбора. Для связывания чего-либо, процессирования, возможности добавить ко всему этому регуляцию, в определенных местах должны быть определенные типы атомов. Это налагает ограничения на типы остатков и их позиции, сочетания типов в разных позициях.

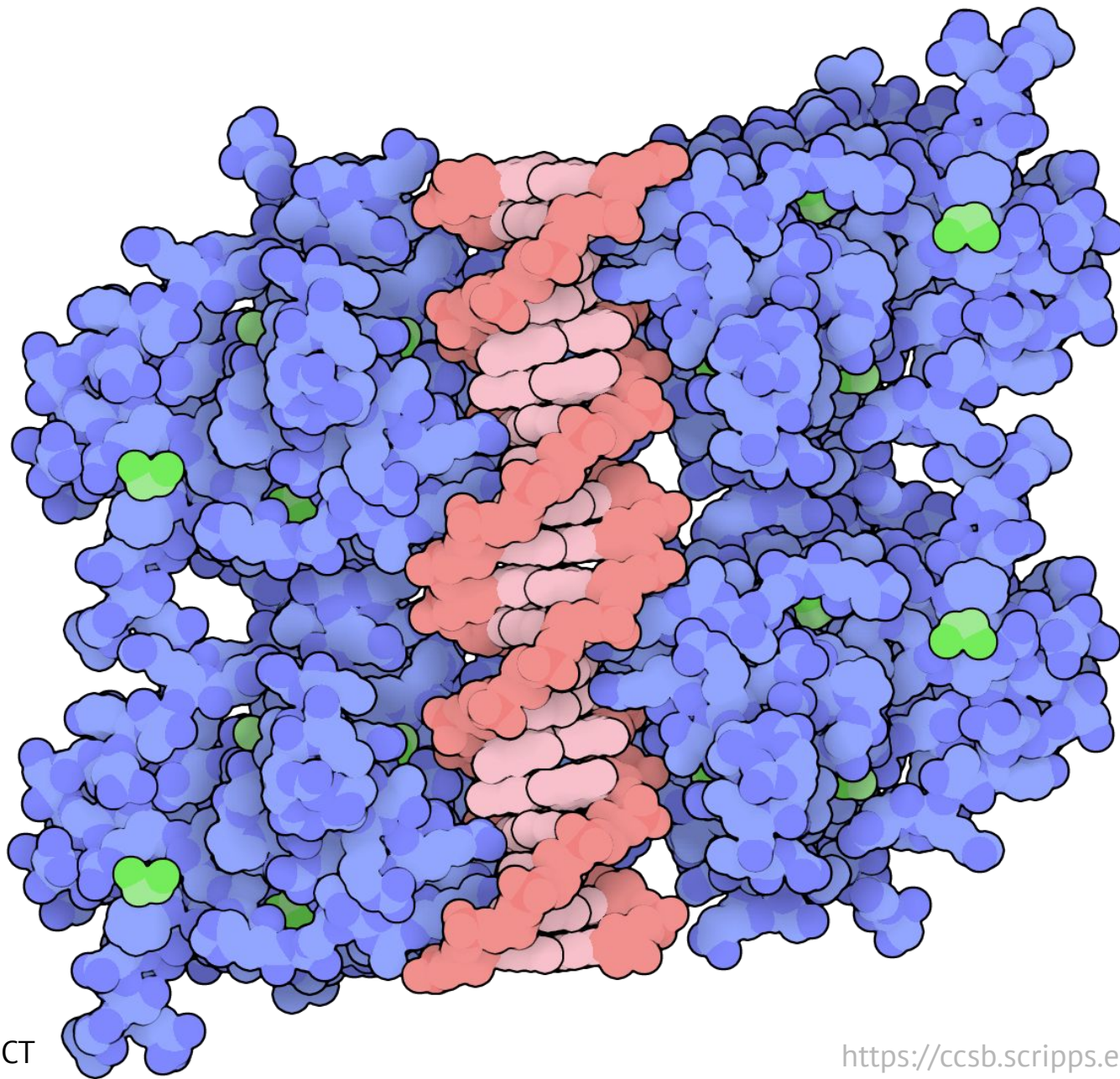
Задача структурной биоинформатики – разгадать этот паззл, как сверху-вниз (знаем последовательности, их эволюцию и функцию их транскриптов, хотим понять, почему вышло именно так), так и снизу-вверх (хотим подобрать последовательность под реализацию определенной функции).

# 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.

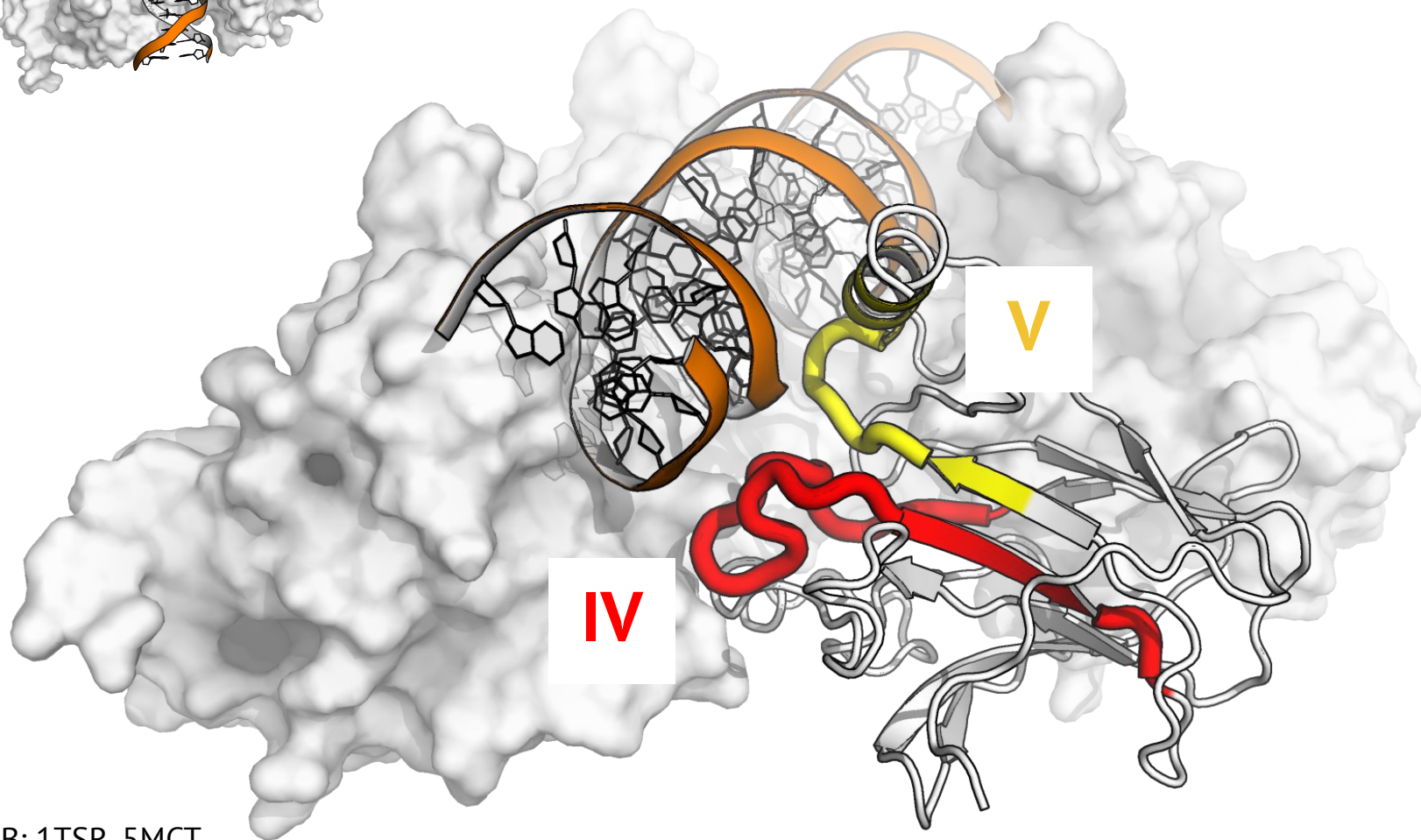
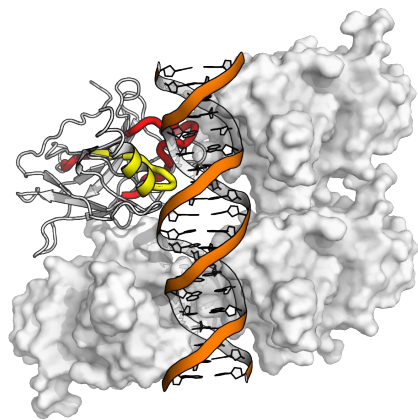
Структурная биоинформатика в данной ситуации должна ответить на вопрос **Почему именно эти участки последовательности консервативны?**  
**Какова их функция?**



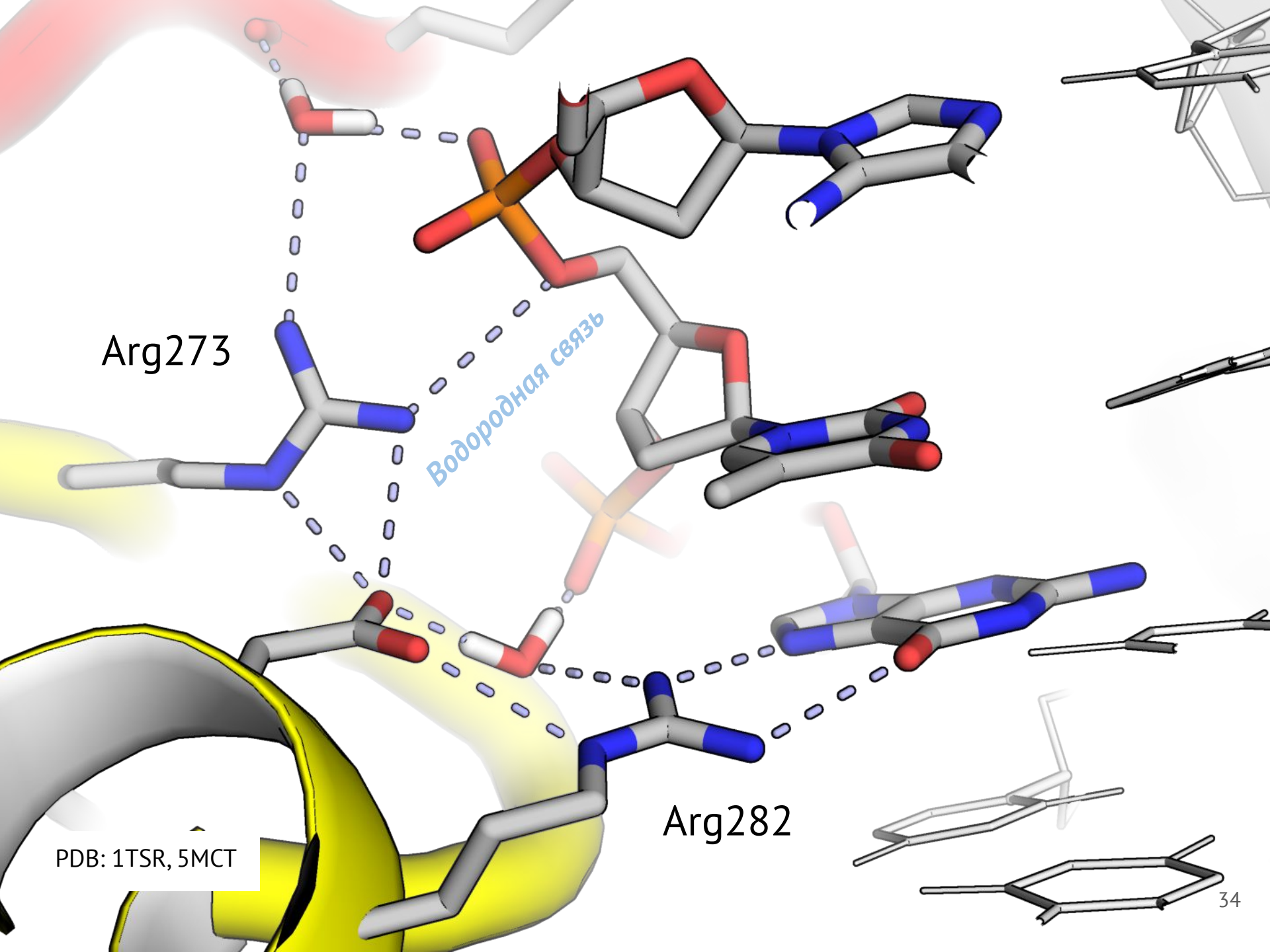
PDB: 1TSR, 5MCT

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





PDB: 1TSR, 5MCT

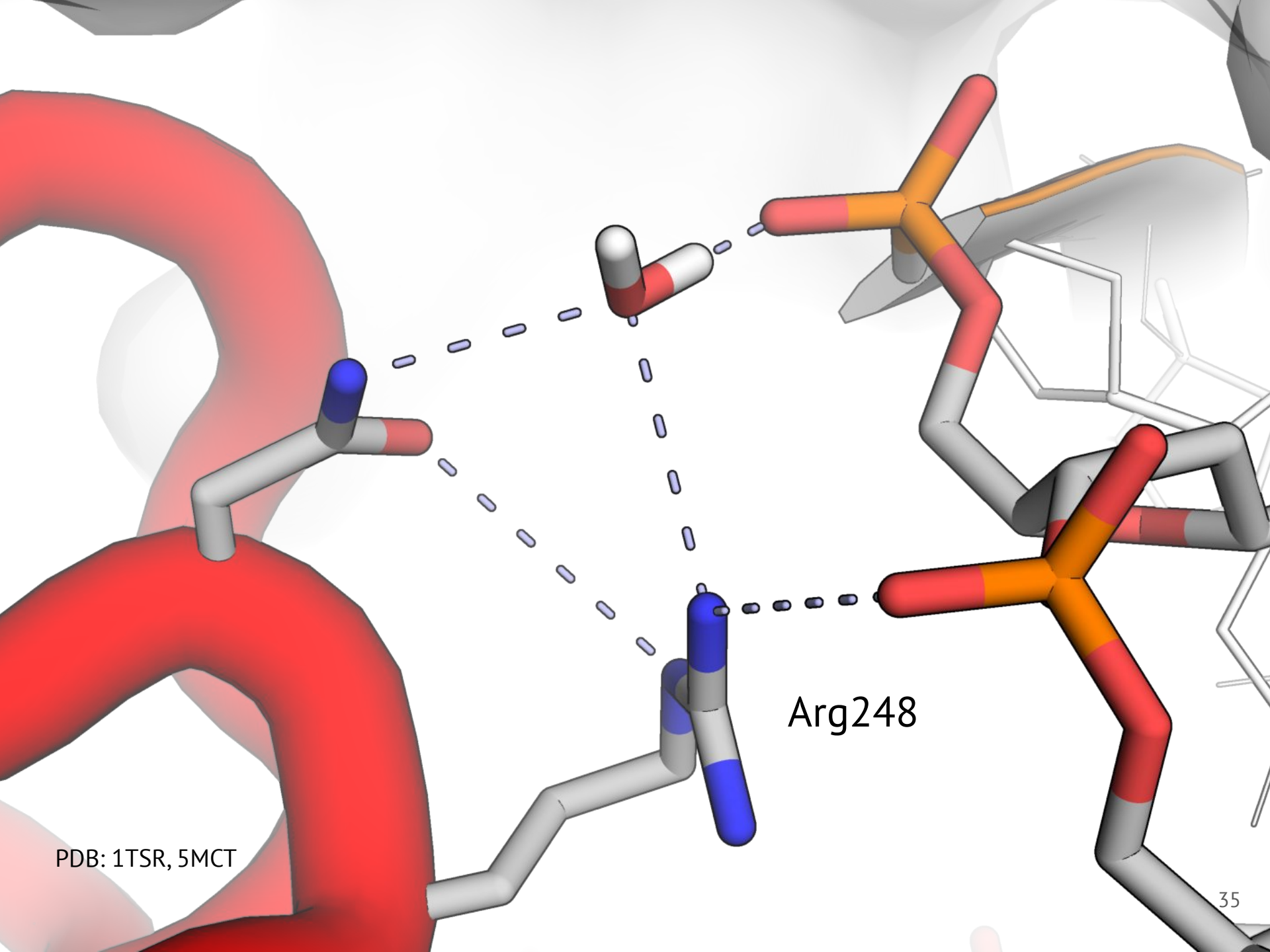


Arg273

Водородная связь

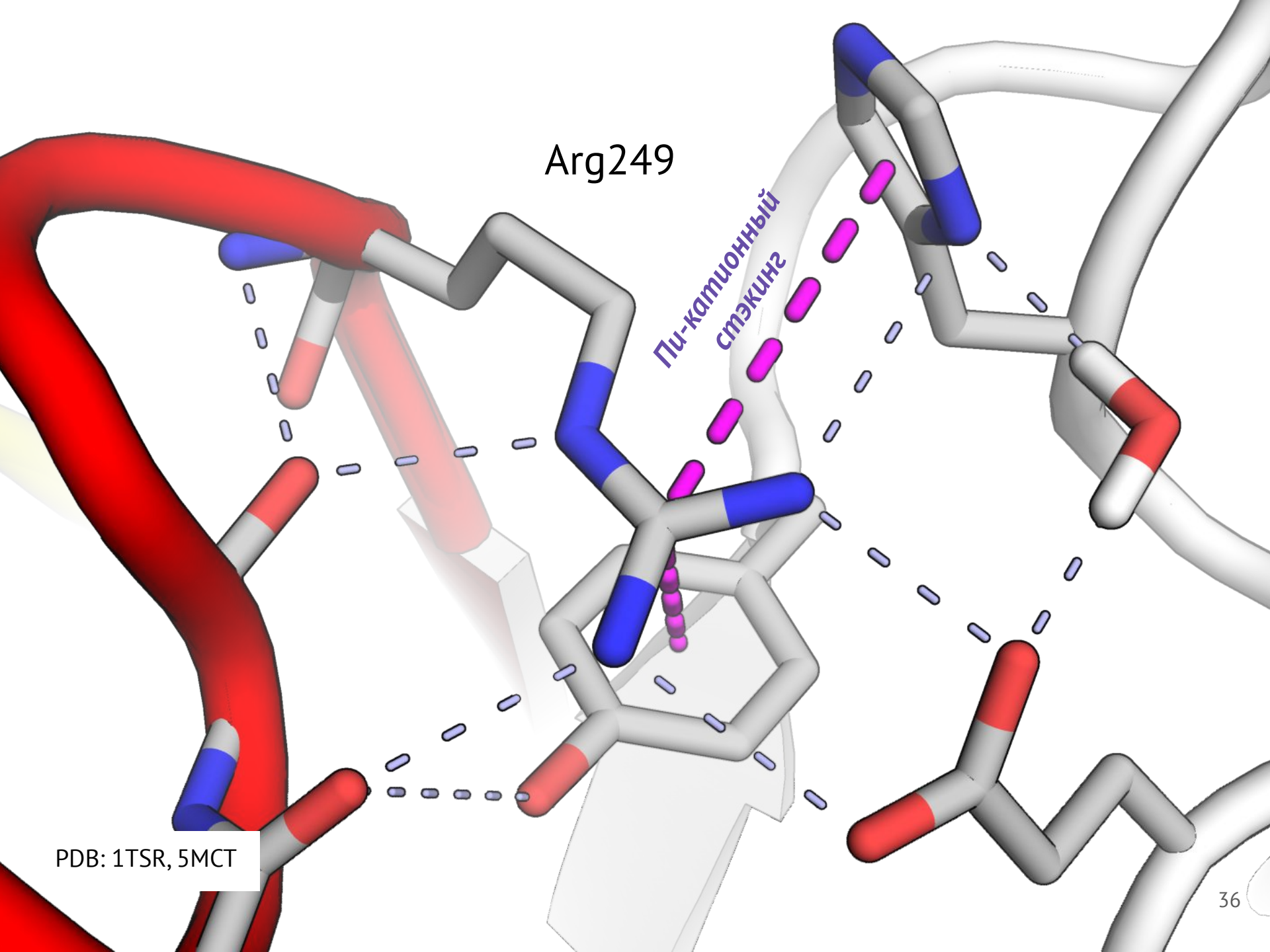
Arg282

PDB: 1TSR, 5MCT



PDB: 1TSR, 5MCT

Arg248



# Агрегация гемоглобина S

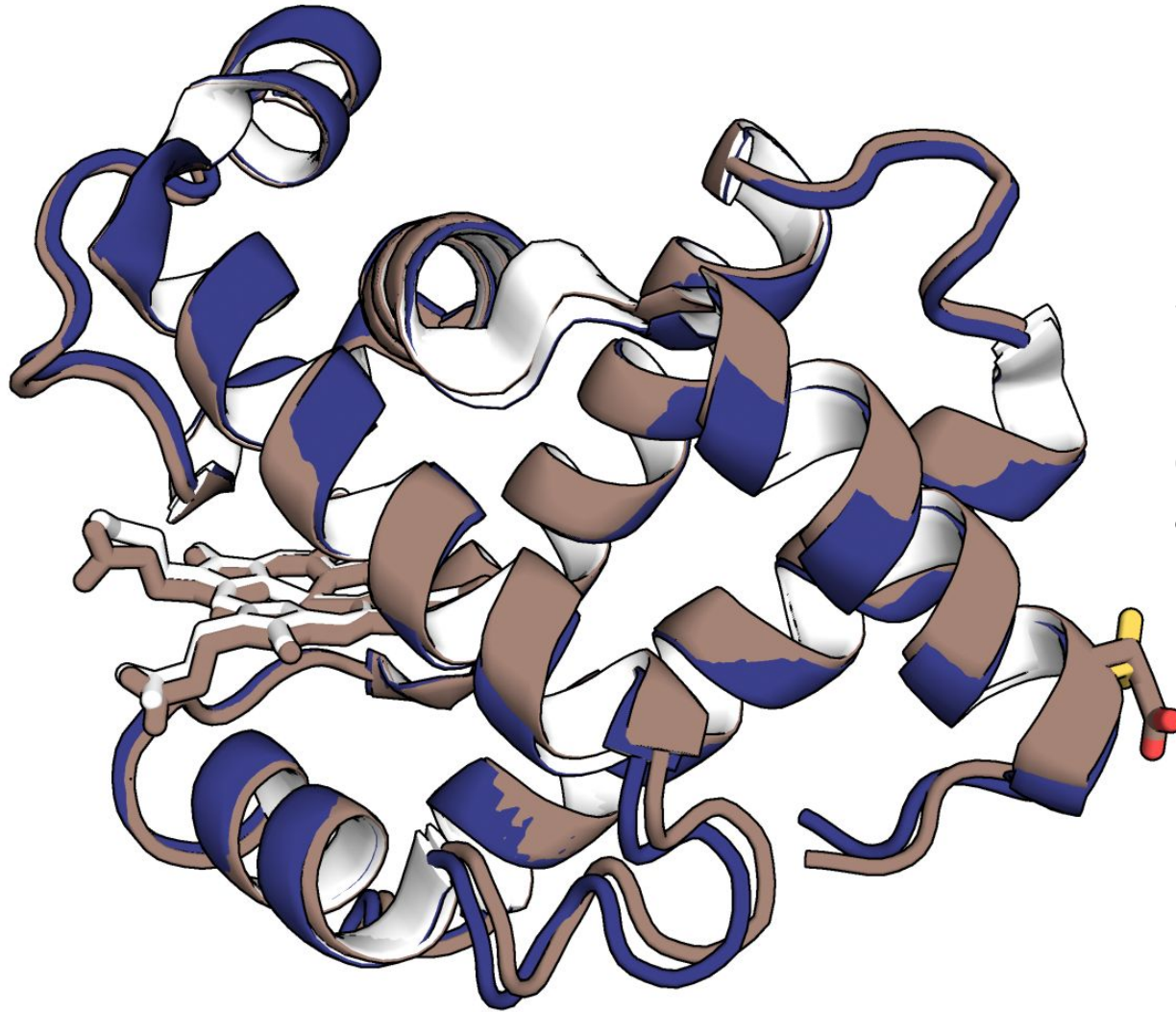
10	20	30	40	50
MVHLTPEEKS	AVTALWGKVN	VDEVGGEALG	RLLVVYPWTQ	RFFESFGDLS
60	70	80	90	100
TPDAVMGNPK	VKAHGKKVLG	AFSDGLAHLA	NLKGTFATLS	ELHCDKLHVD
110	120	130	140	
PENFRLLGNV	LVCVLAHHFG	KEFTPPVQAA	YQKVVAGVAN	ALAHKYH

10	20	30	40	50
MVHLTP <b>V</b> EKS	AVTALWGKVN	VDEVGGEALG	RLLVVYPWTQ	RFFESFGDLS
60	70	80	90	100
TPDAVMGNPK	VKAHGKKVLG	AFSDGLAHLA	NLKGTFATLS	ELHCDKLHVD
110	120	130	140	
PENFRLLGNV	LVCVLAHHFG	KEFTPPVQAA	YQKVVAGVAN	ALAHKYH

Замена E7V в бета-субъединице гемоглобина вызывает серповидноклеточную анемию. Как?

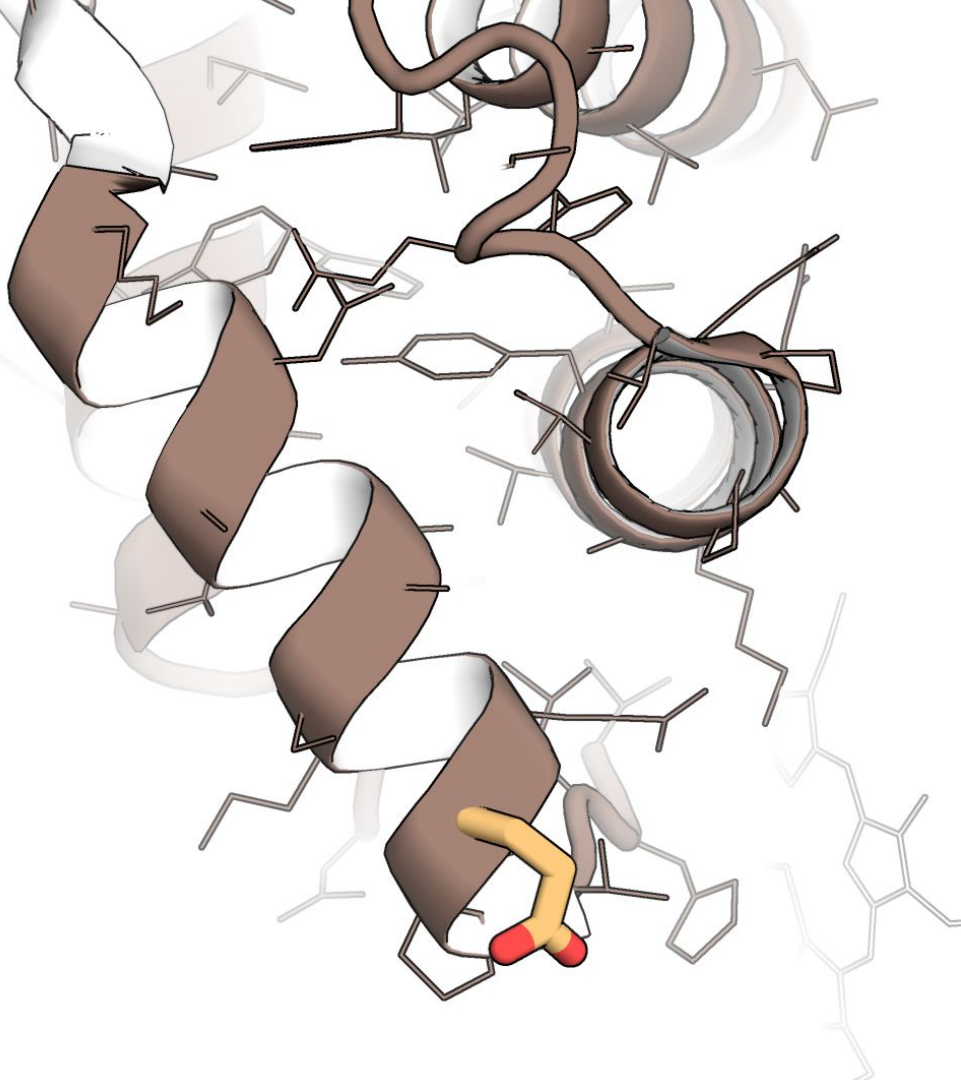
PDB: 2HBS, 4HHB

# Агрегация гемоглобина S

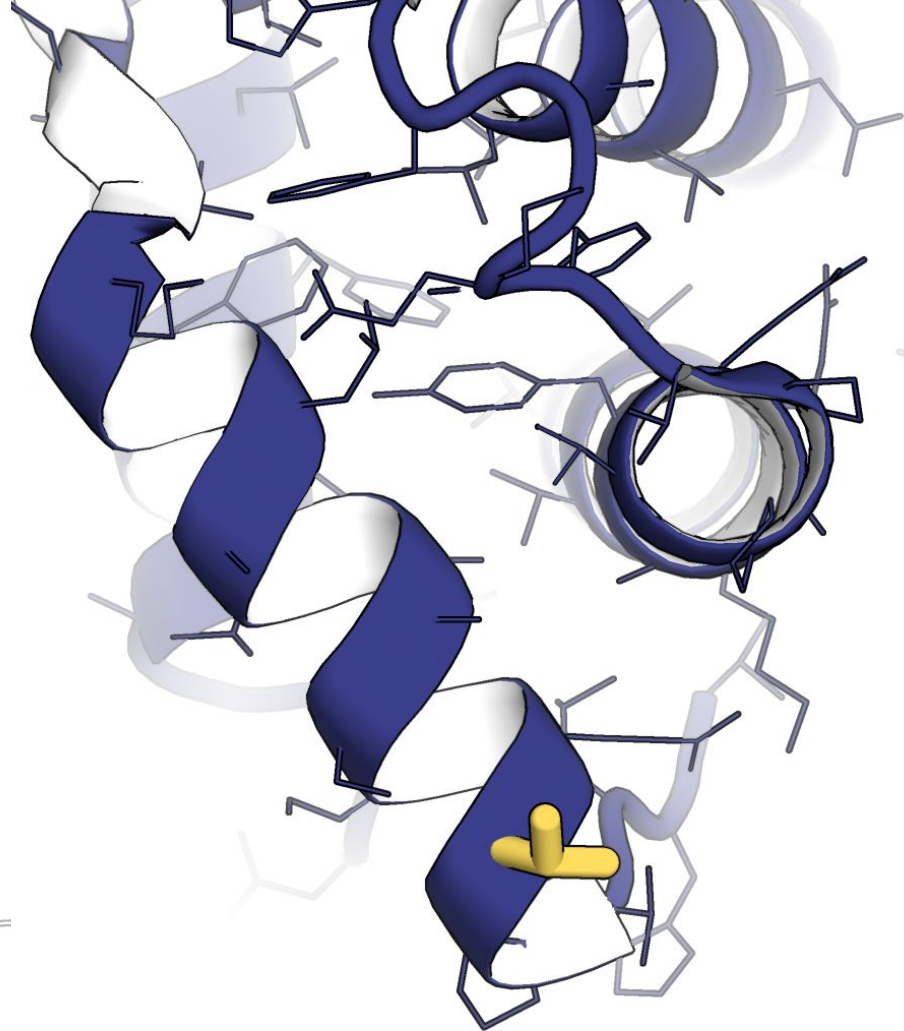


Замена E6V  
в бета-субъединице  
гемоглобина вызывает  
серповидноклеточную  
анемию. Как?

PDB: 2HBS, 4HHB

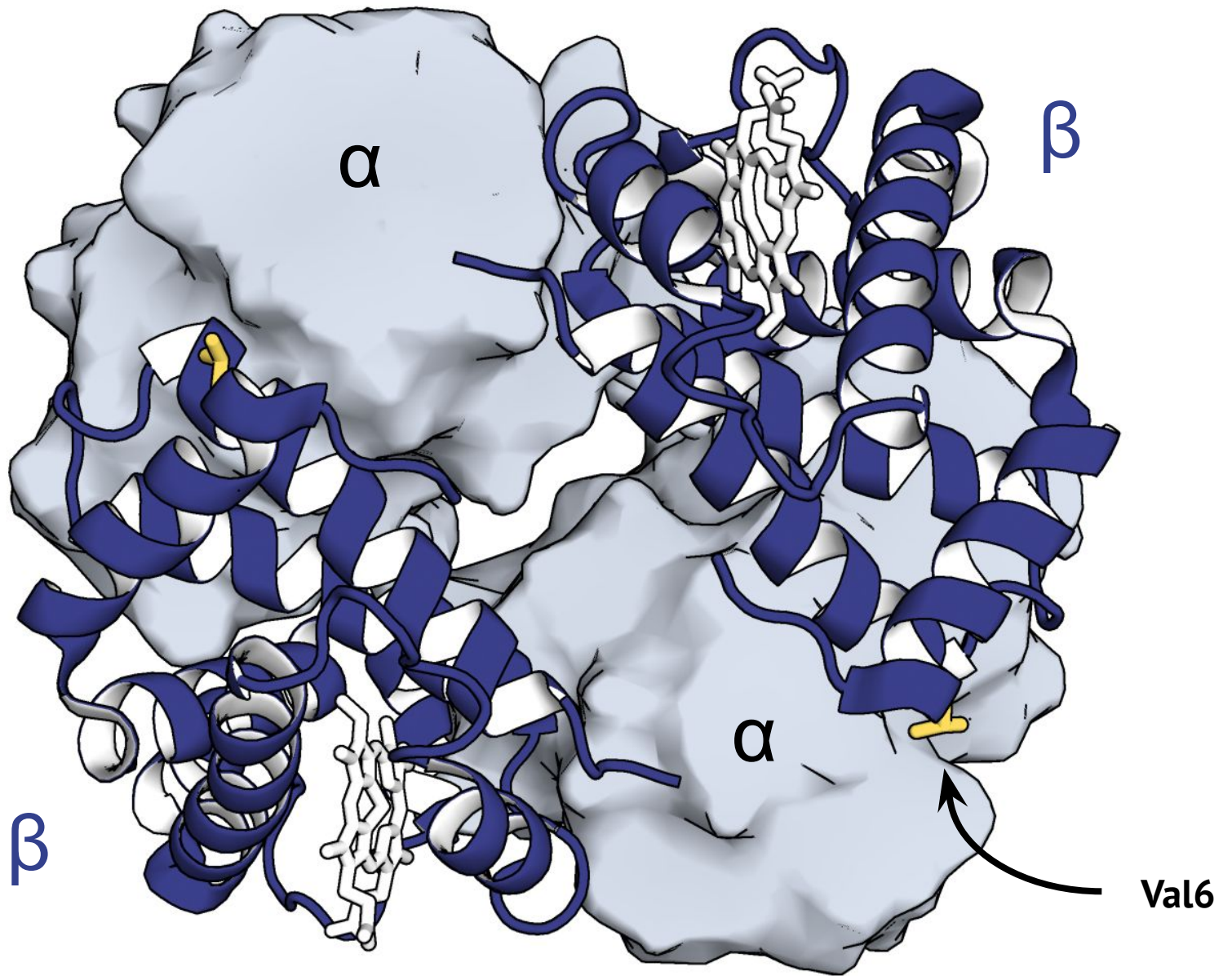


Дикий тип: на 6 позиции глутамат

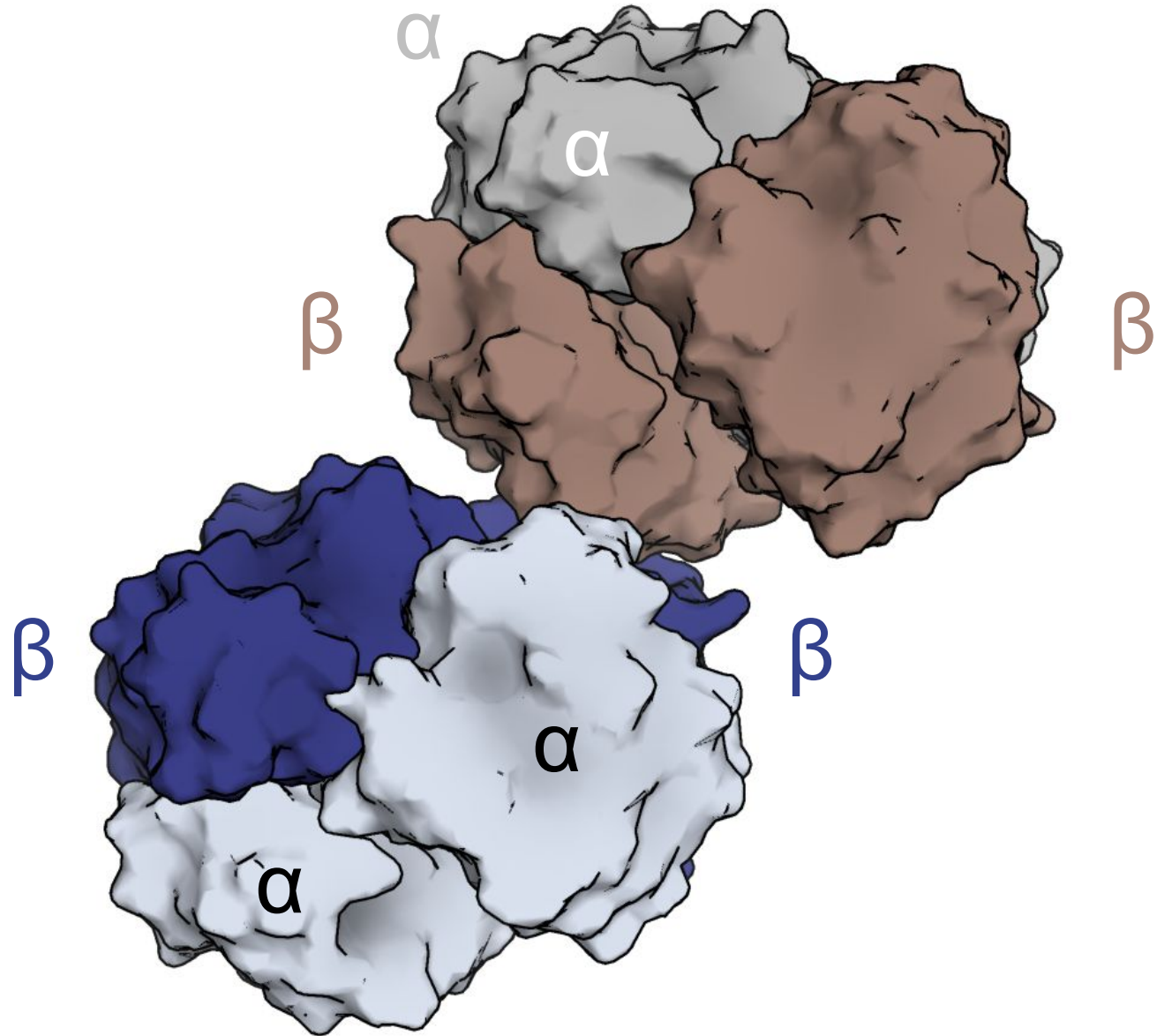


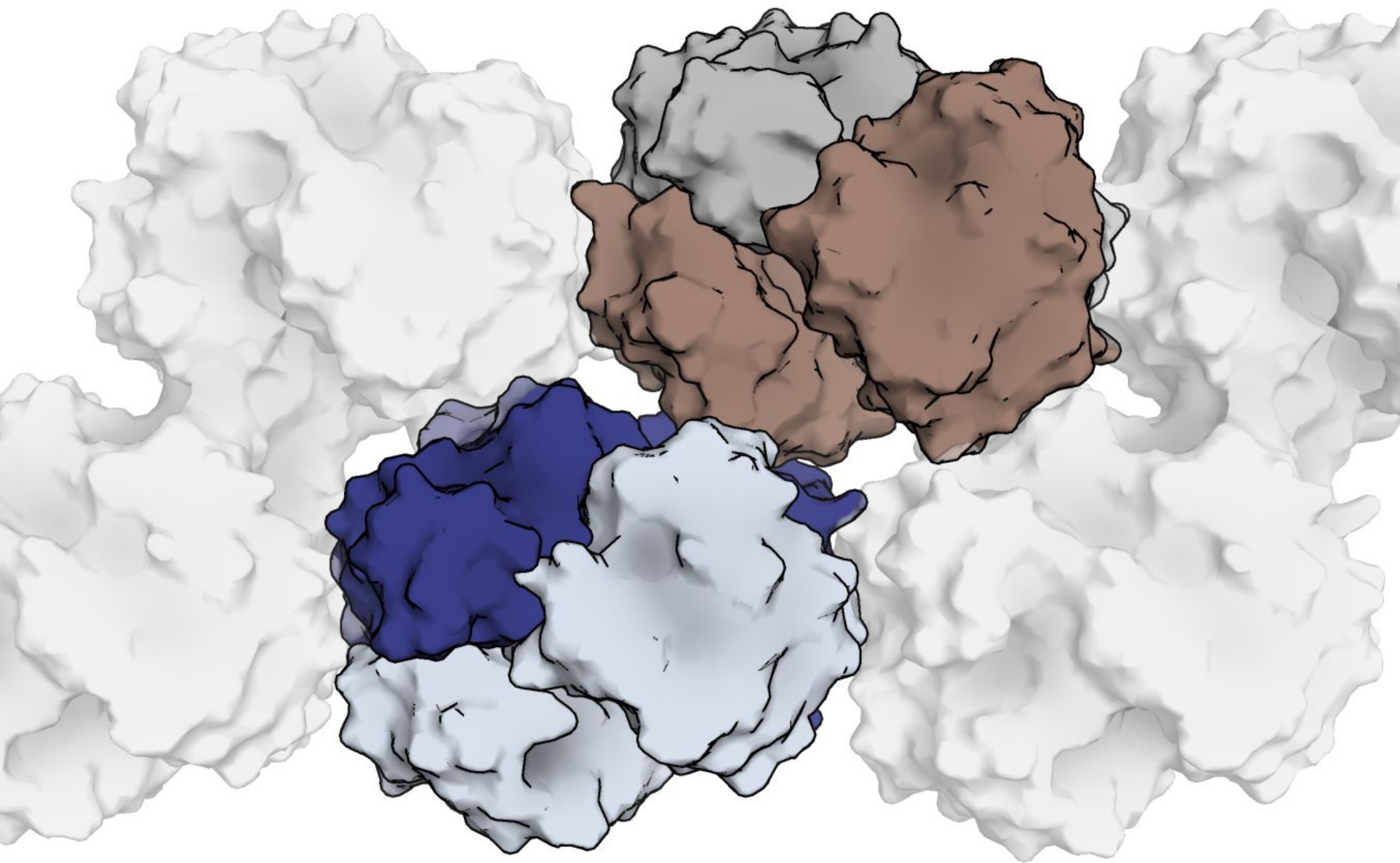
Мутант: на 6 позиции валин

Однако структуры в остальном идентичны →  
причина патологии не в мисфолдинге



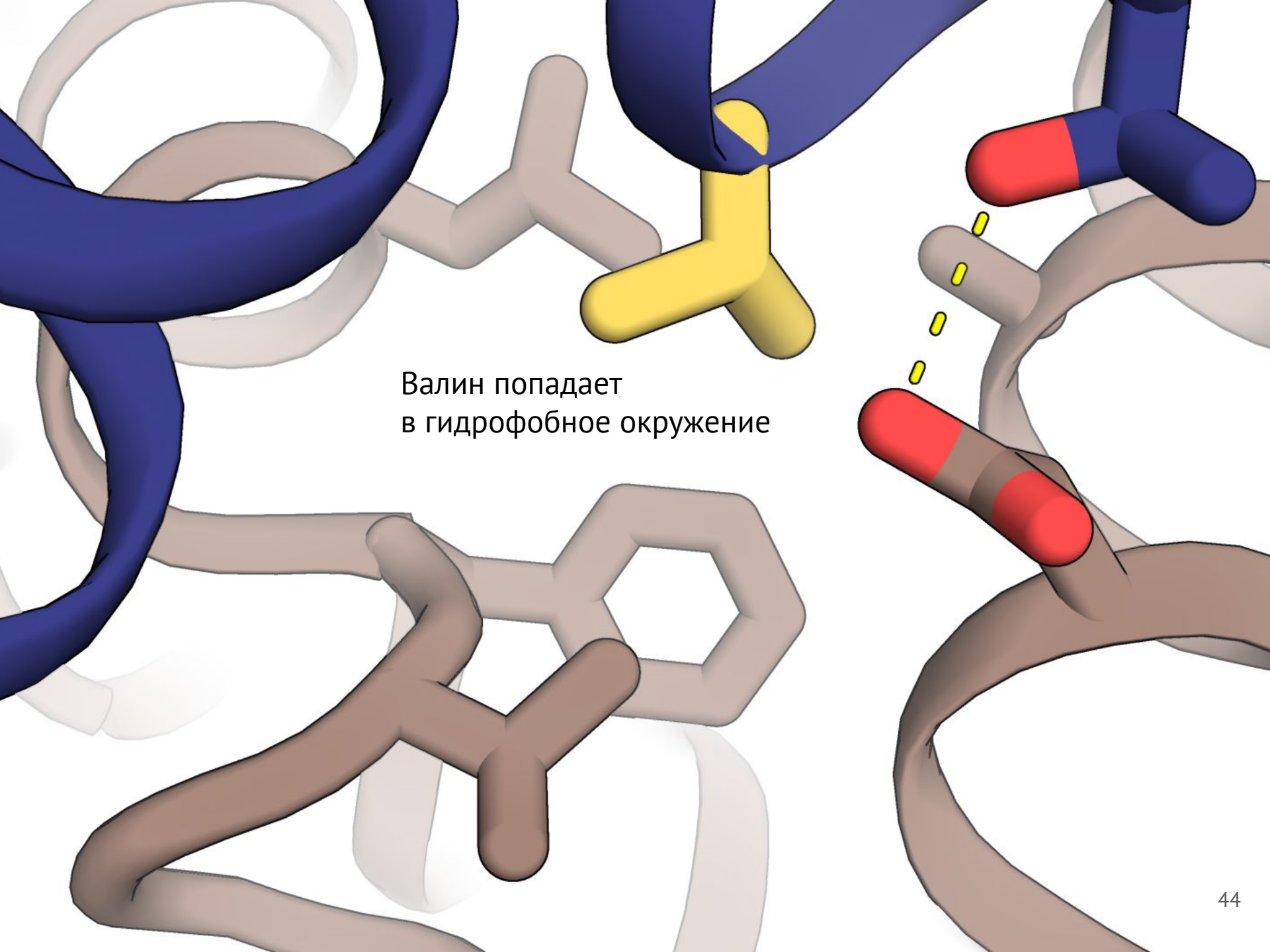






# Shape complementarity

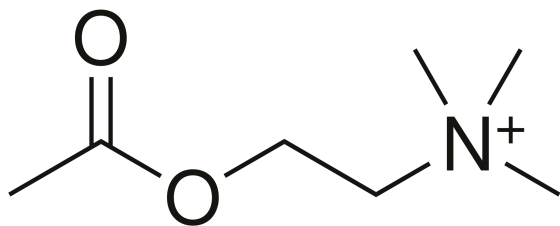
Один из ключевых терминов молекулярного распознавания  
Гидрофобный эффект + VdW



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

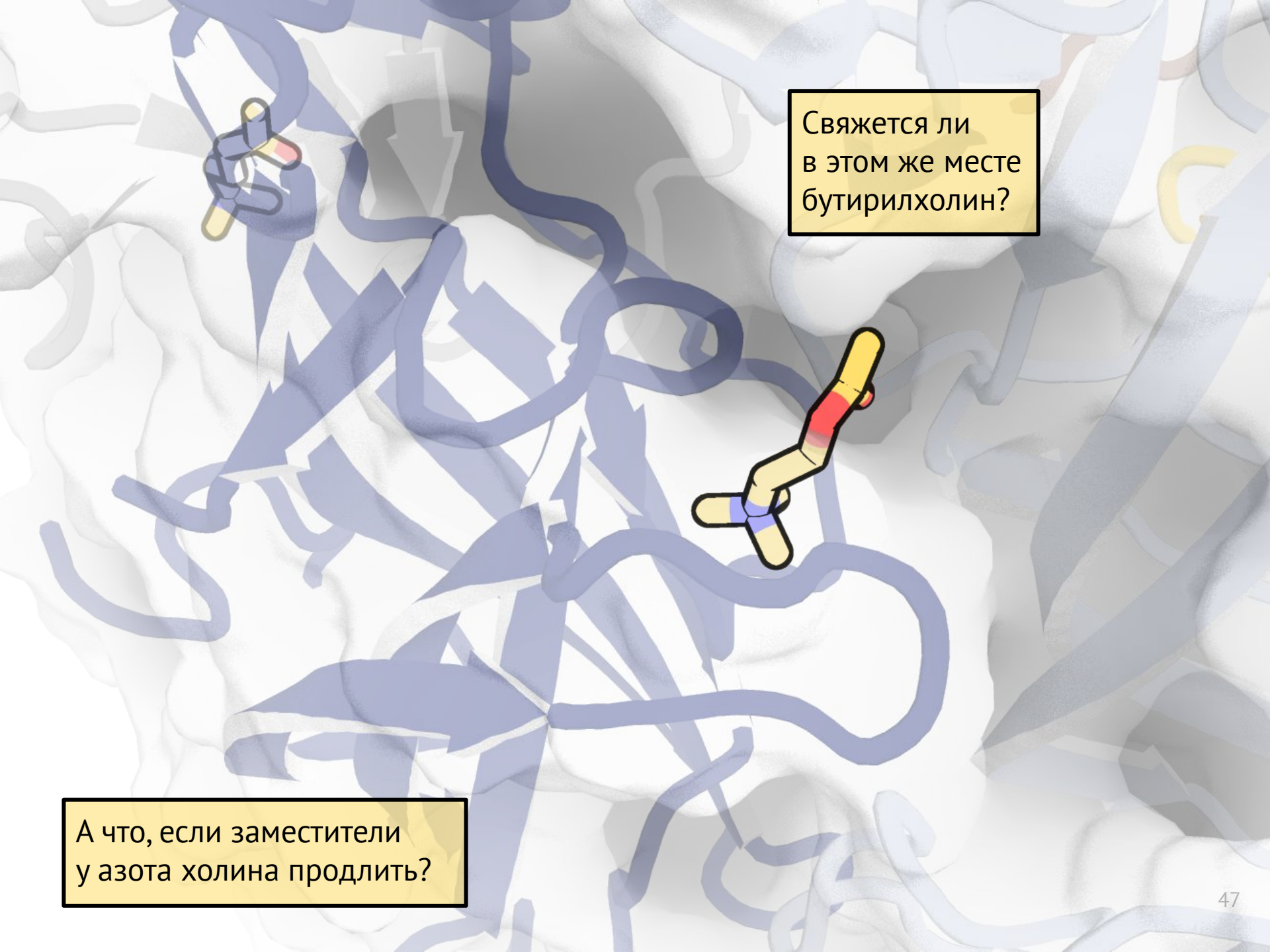


nAChR  
PDB 5KXI

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

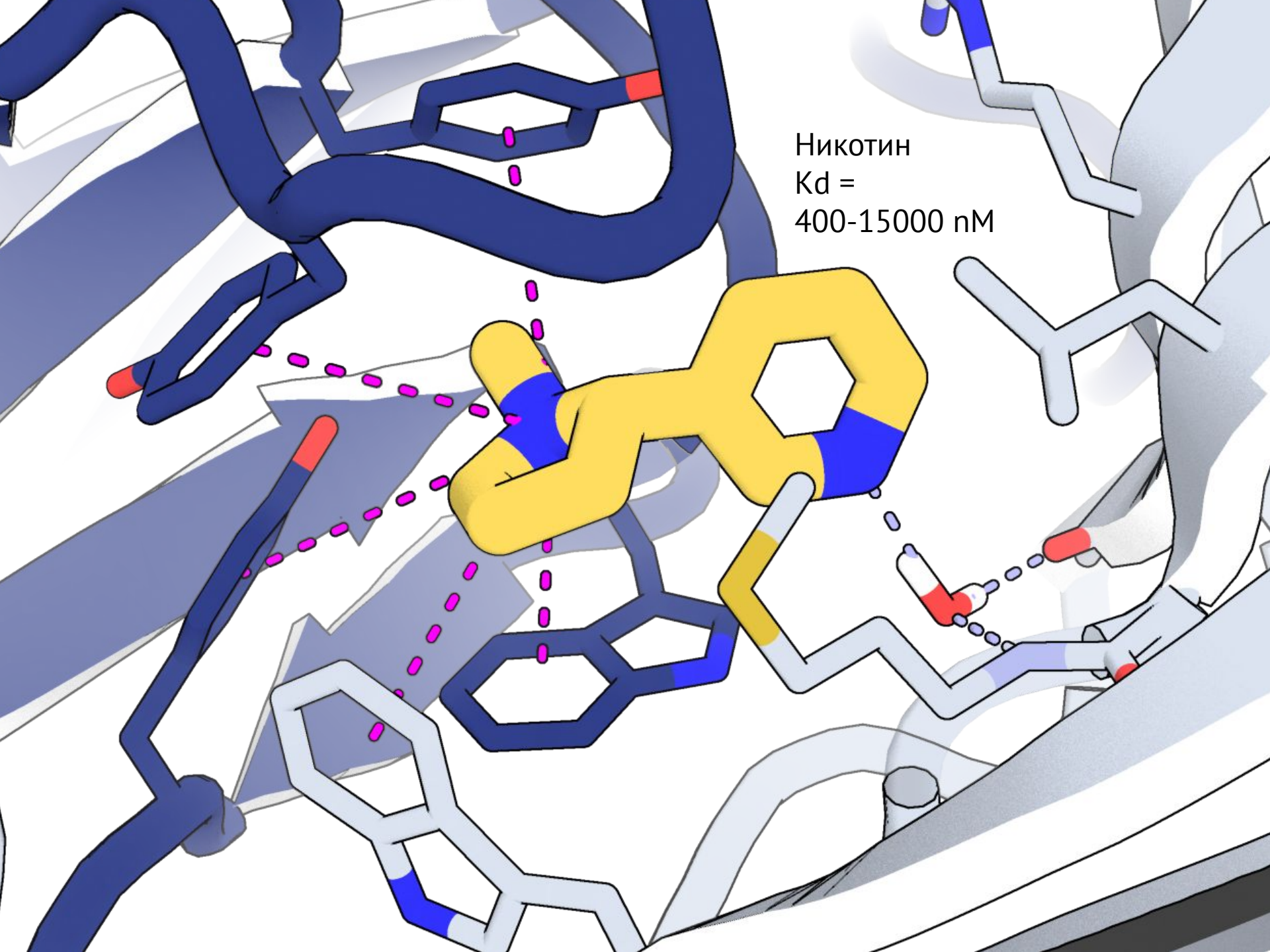
Гидрофобный  
карман,  
образованный  
аргинином  
и лейцином

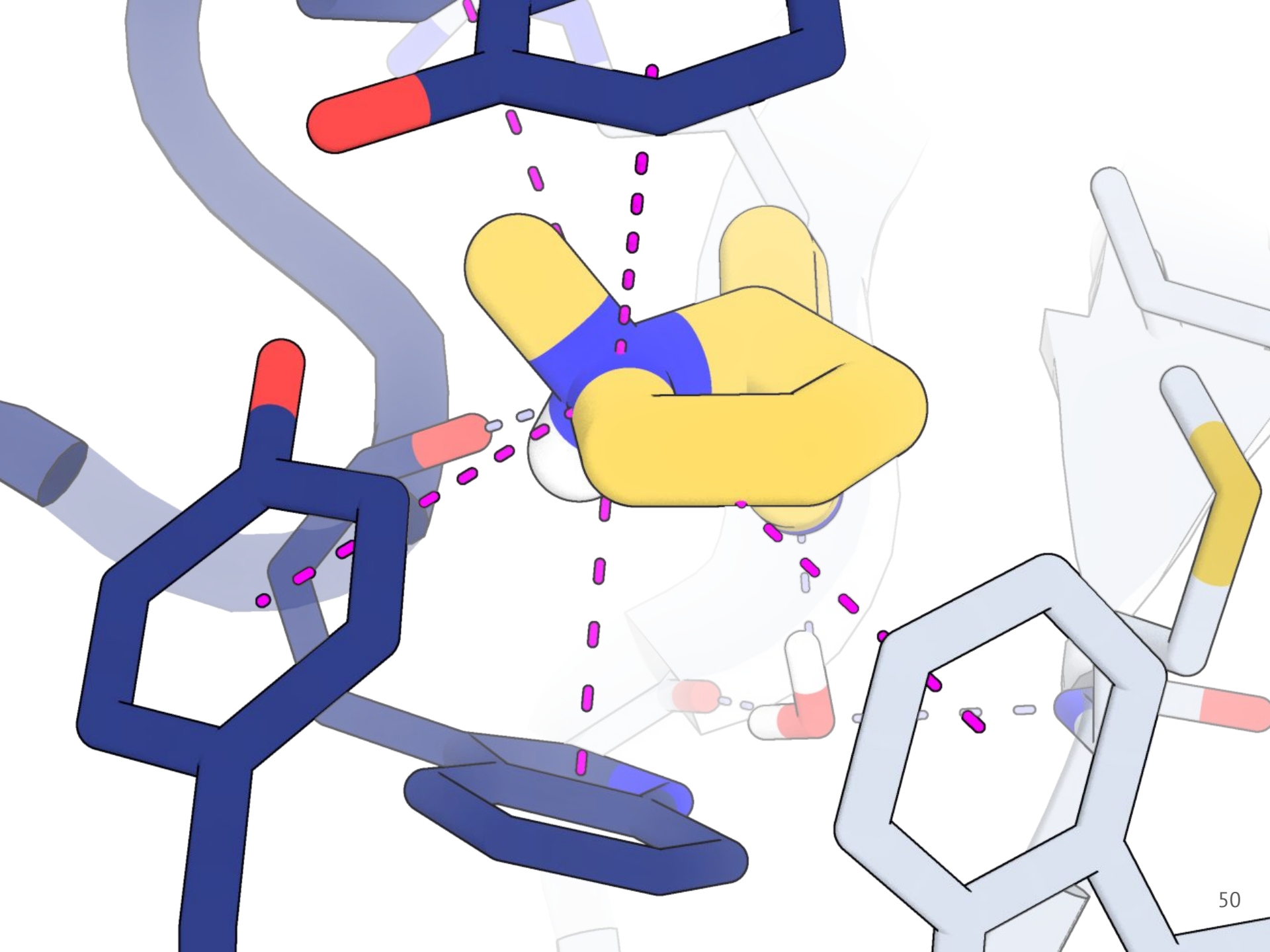
Пи-кат. +  
гидрофобика

Водяной  
МОСТИК

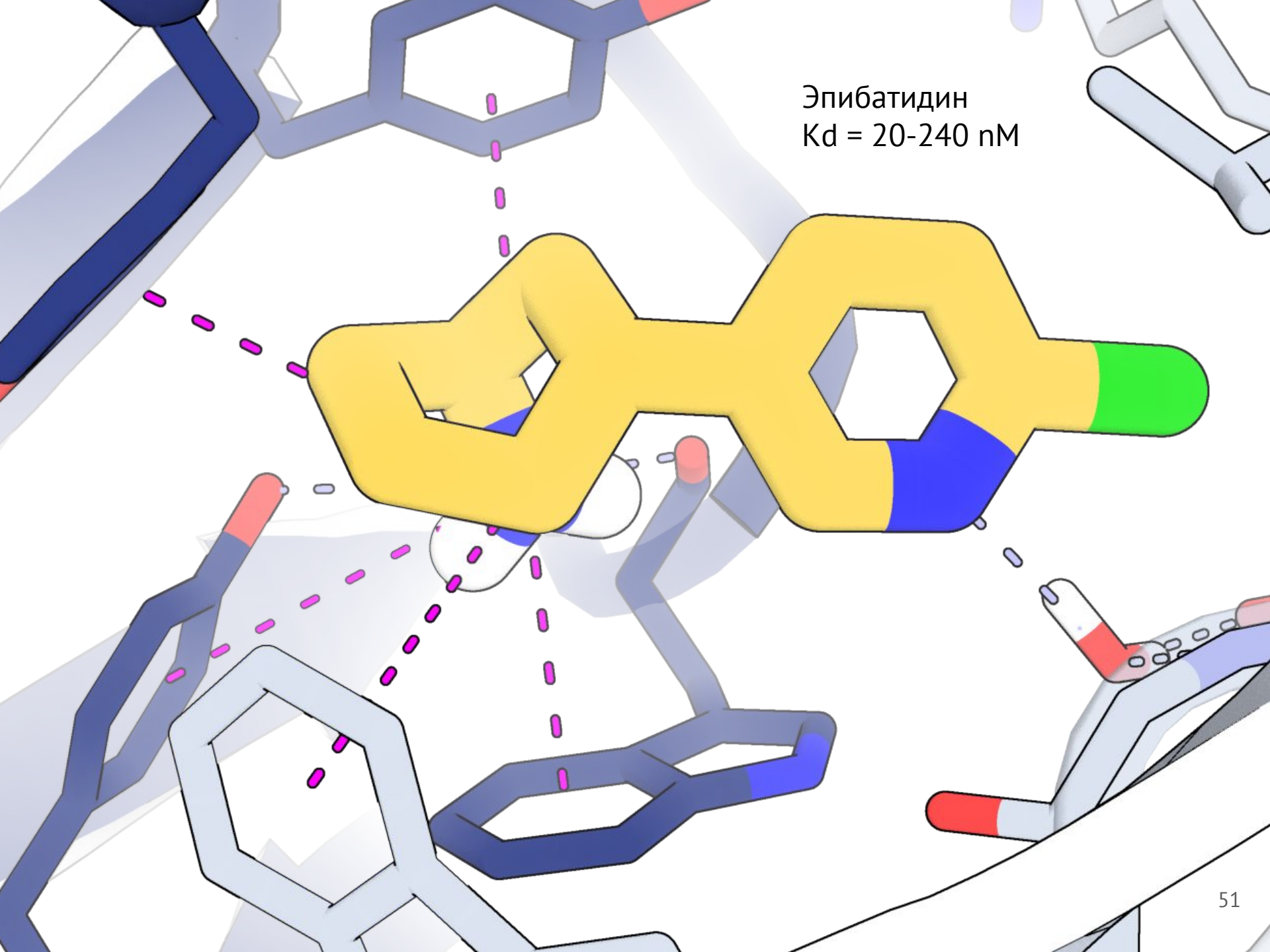


Никотин  
Kd =  
400-15000 nM

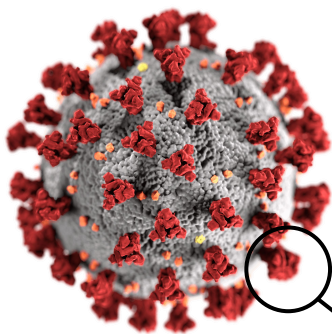




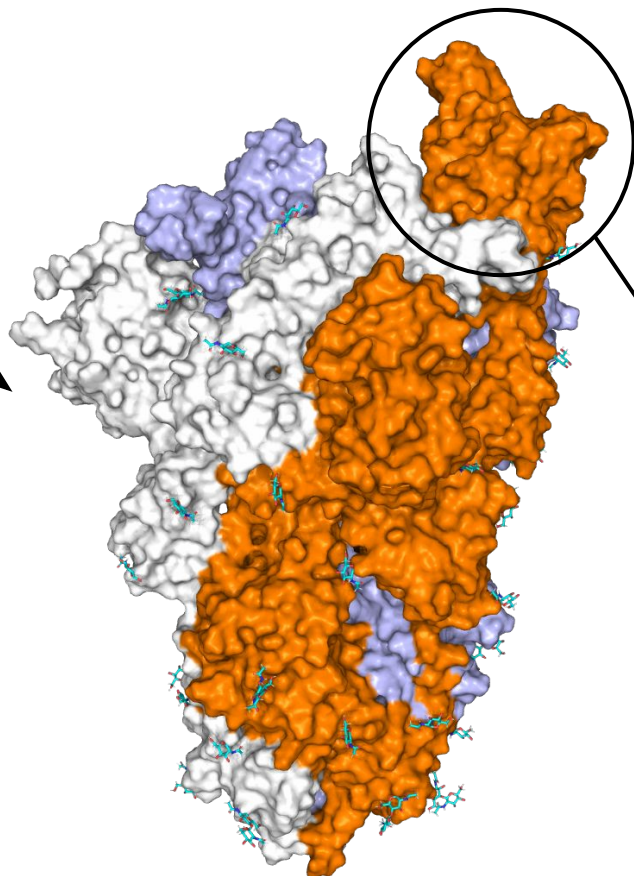
Эпibatидин  
Kd = 20-240 нМ



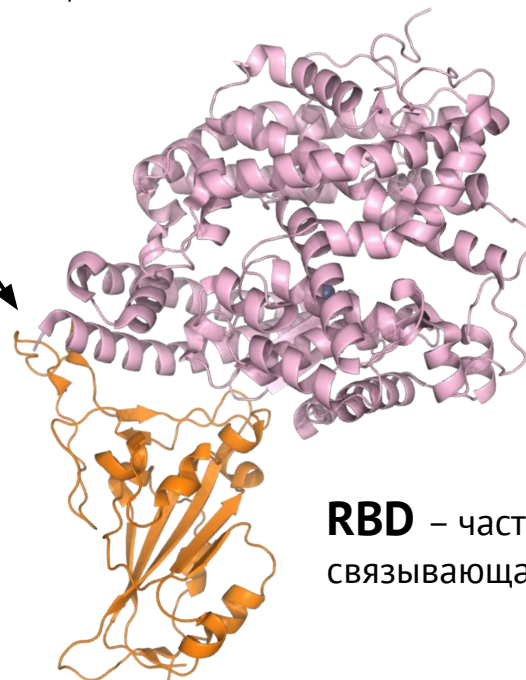
## Sars-Cov-2



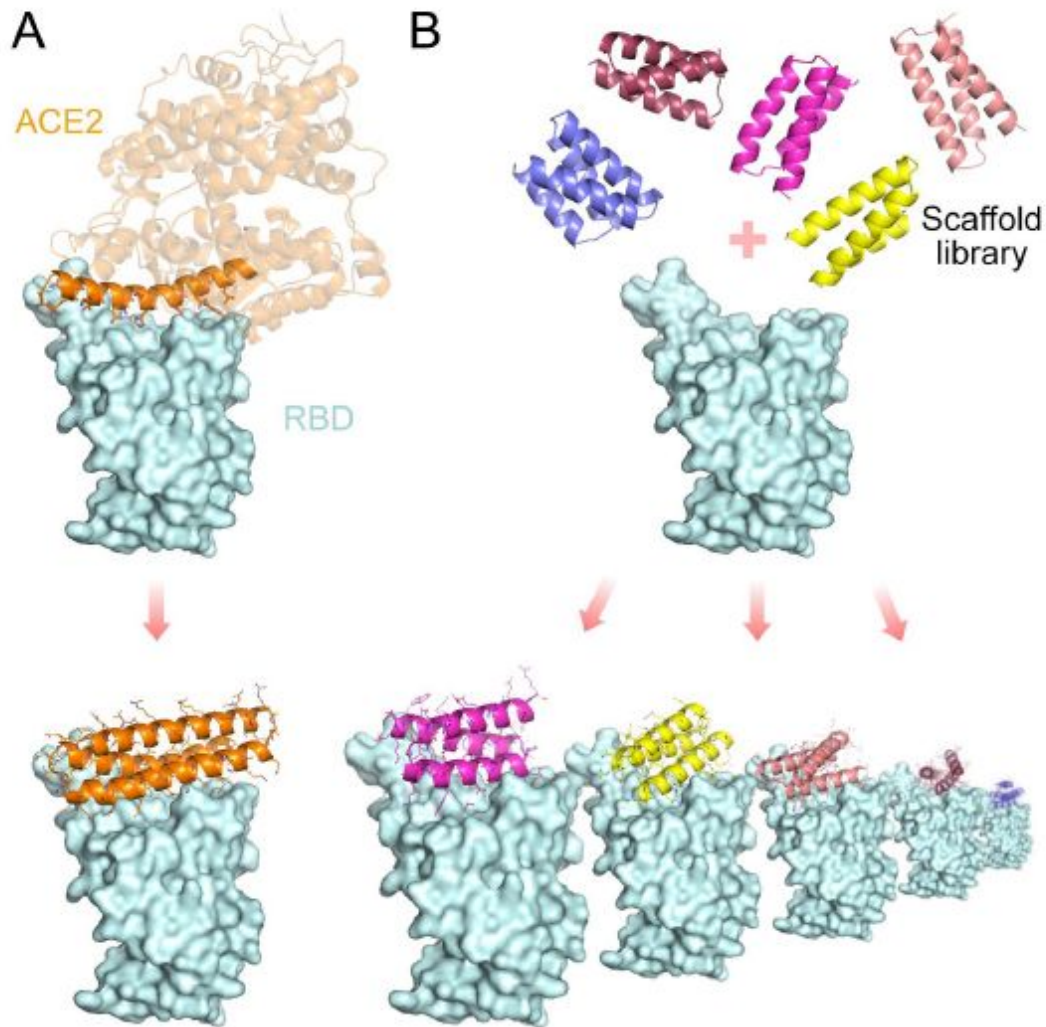
## Spike protein



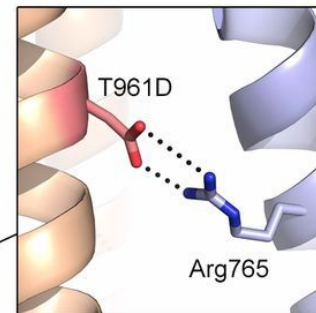
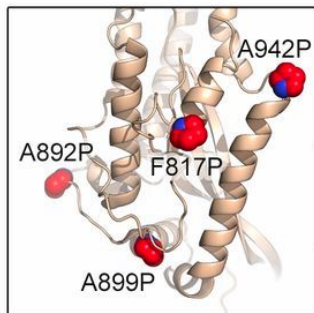
**ACE2** – наш рецептор, который вирус подло использует для своих целей



**RBD** – часть S-белка, связывающая ACE2

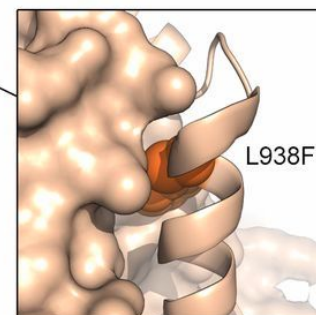
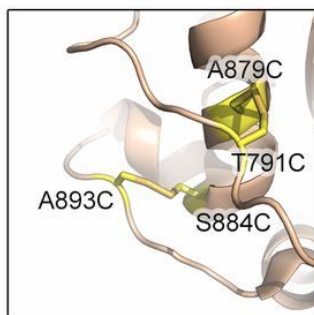


**Инженерия  
искусственных  
белков,  
связывающих RBD  
лучше, чем ACE2**

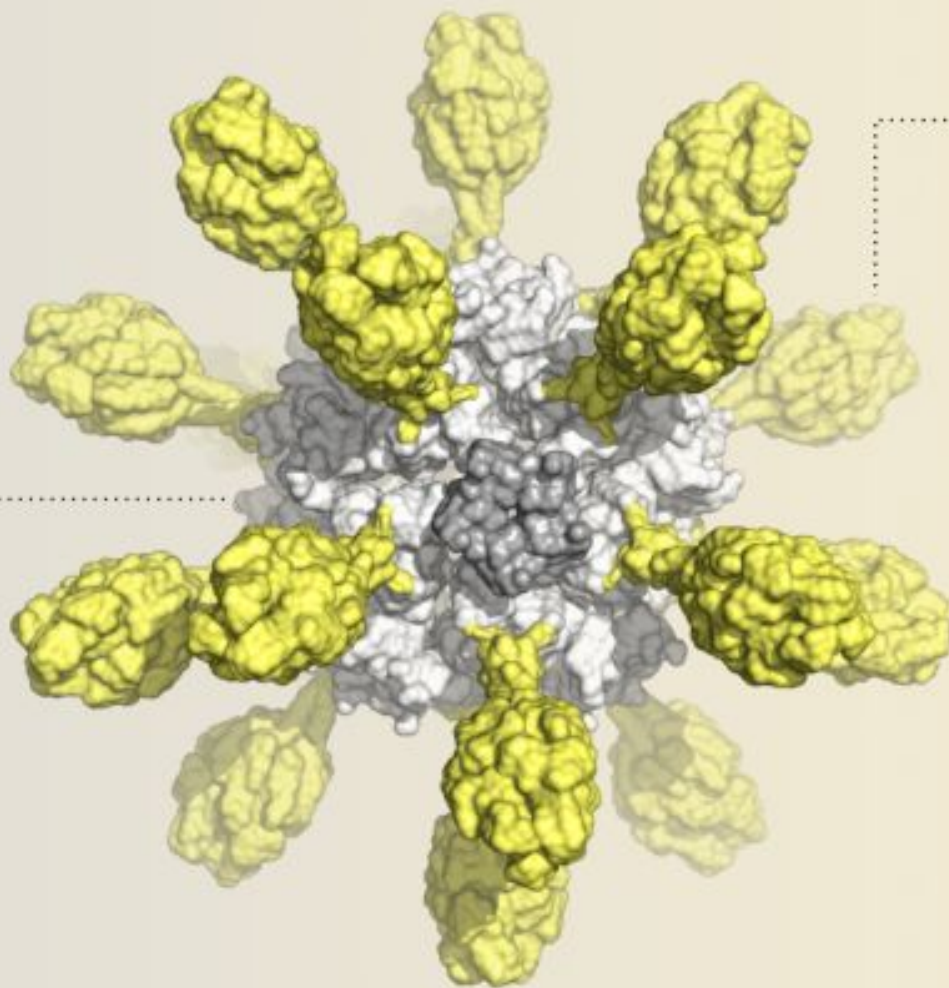


## Инженерия более стабильного S-белка

DOI: 10.1126/science.abd0826

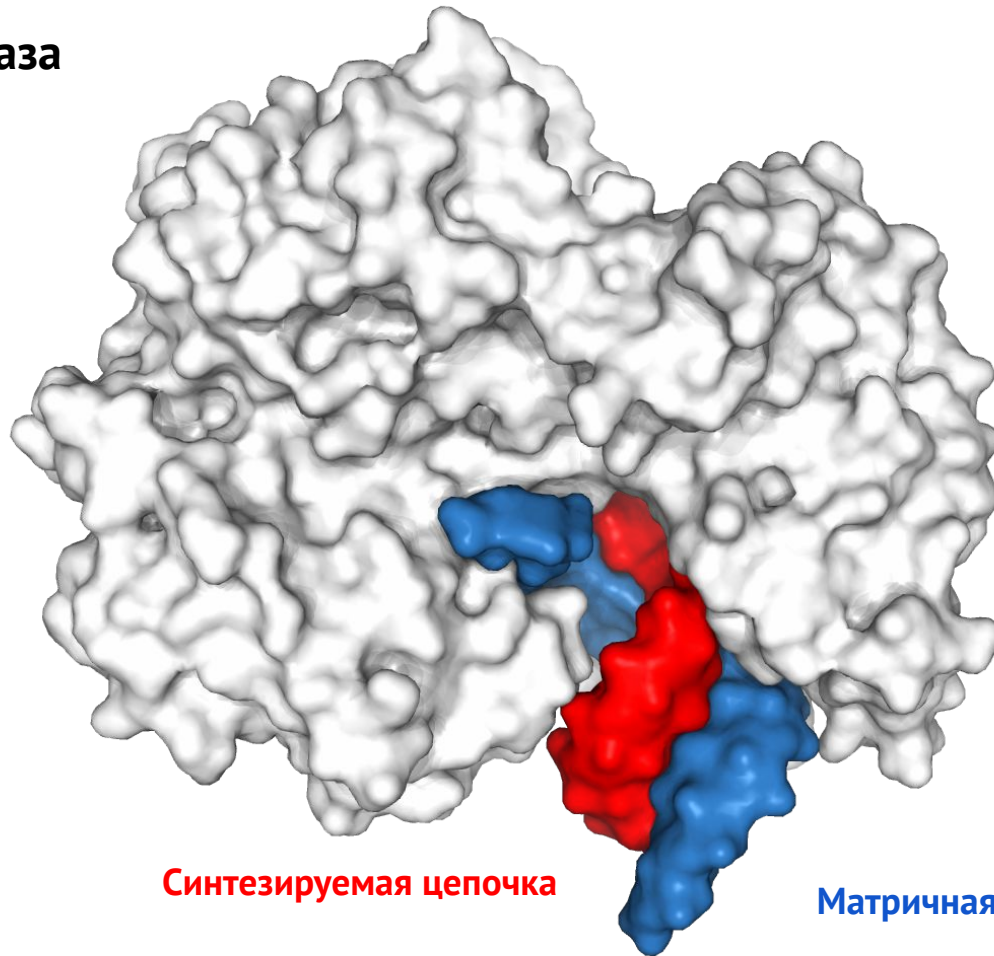


DESIGNED  
**NANOPARTICLE**



STABILIZED  
**ANTIGENS**

# РНК-полимераза

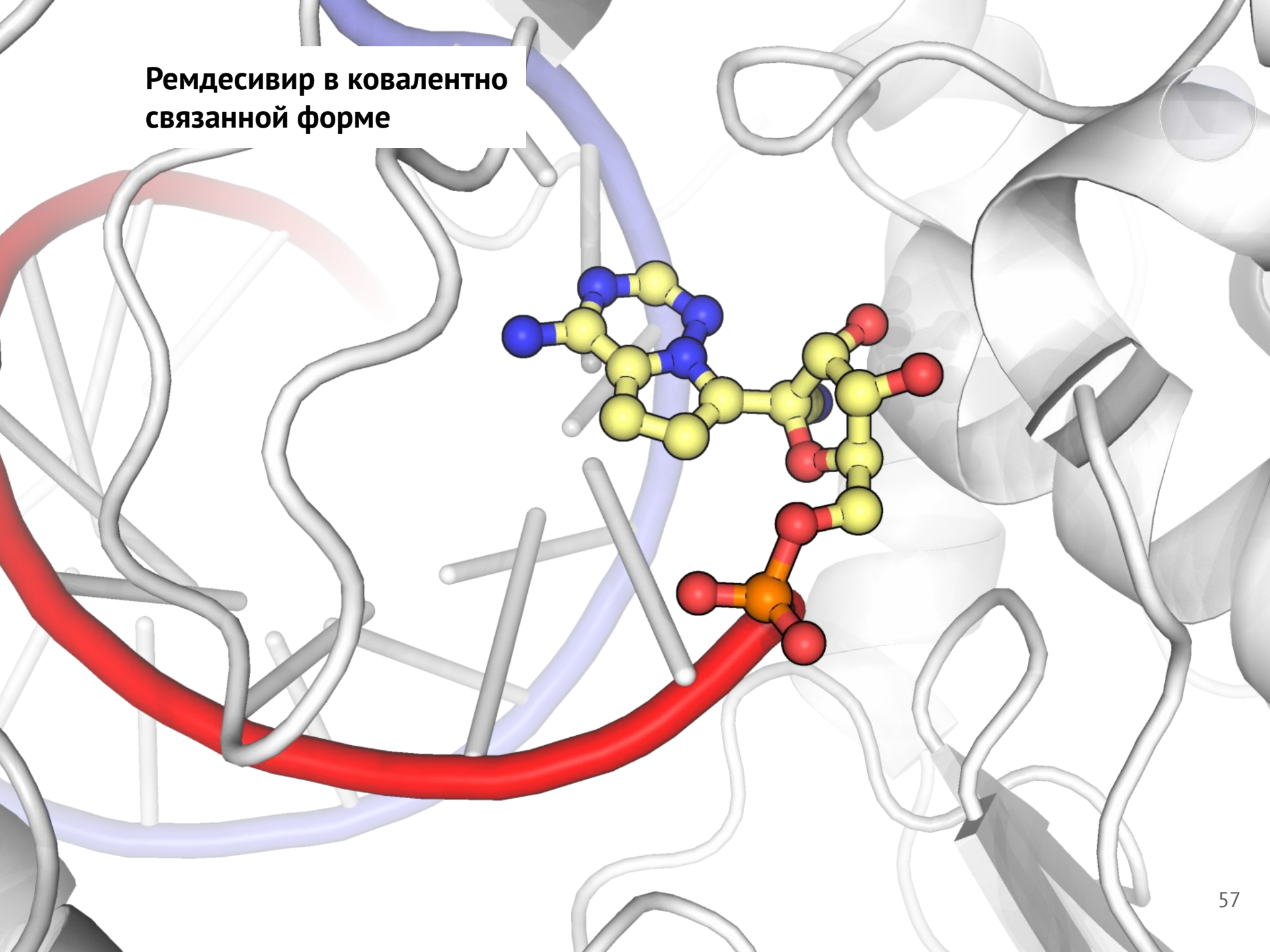


Синтезируемая цепочка

Матричная цепочка

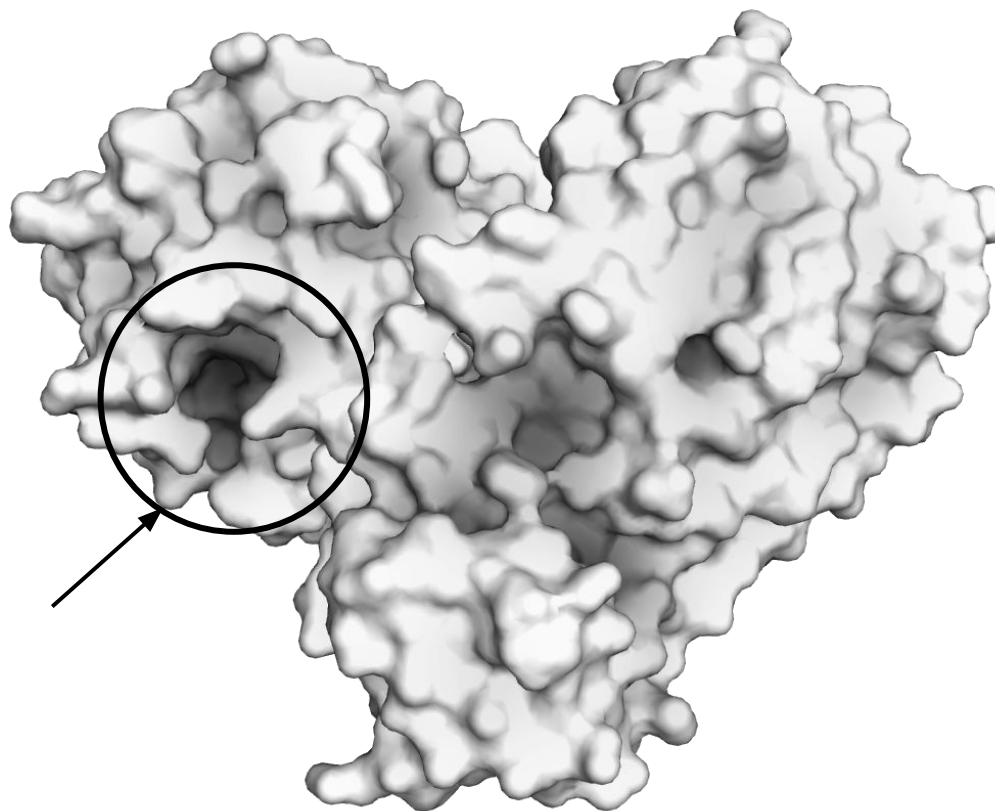


**Ремдесивир в ковалентно  
связанной форме**



## Главная протеаза коронавируса

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



**Что можно сделать?**





# Структурка и статьи: мультидисциплинарные работы

Без структурного рассмотрения невозможно ни одно высокоимпактное

- фармакологическое исследование: **механизмы действия** лекарств, механизмы потери чувствительности к лекарствам
- исследование эффекта мутации с **доказательством каузальности**
- исследование о **дизайне** модуляторов, белков и ферментов

nature



# Структурка и статьи: специализированные журналы

Nature Structural and Molecular Biology

Bioinformatics

Nature Chemical Biology

BMC Bioinformatics

Structure

Nucleic Acids Research

Molecular Cell

Proteins

Acta Crystallographica D, F

Current Opinion in Structural Biology

Frontiers in Molecular Biosciences

Journal of Structural Biology

# 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**