

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



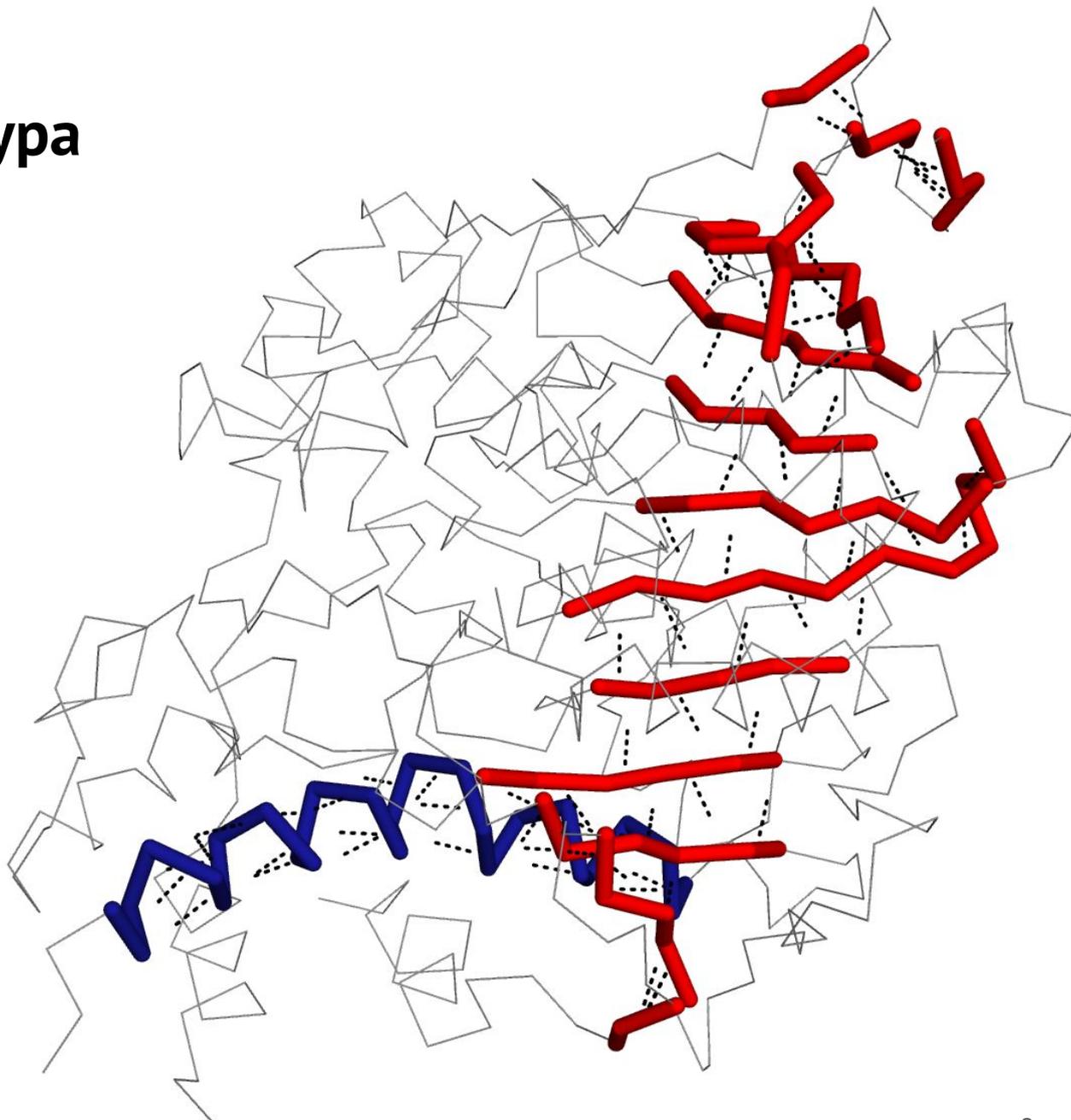
Структурные паттерны. Структурные домены

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

Вторичная структура

Вторичная структура -
паттерн во взаимном
расположении атомов
остова

Наслаивается на паттерн
меж-остовных
водородных связей
(определяется им?)

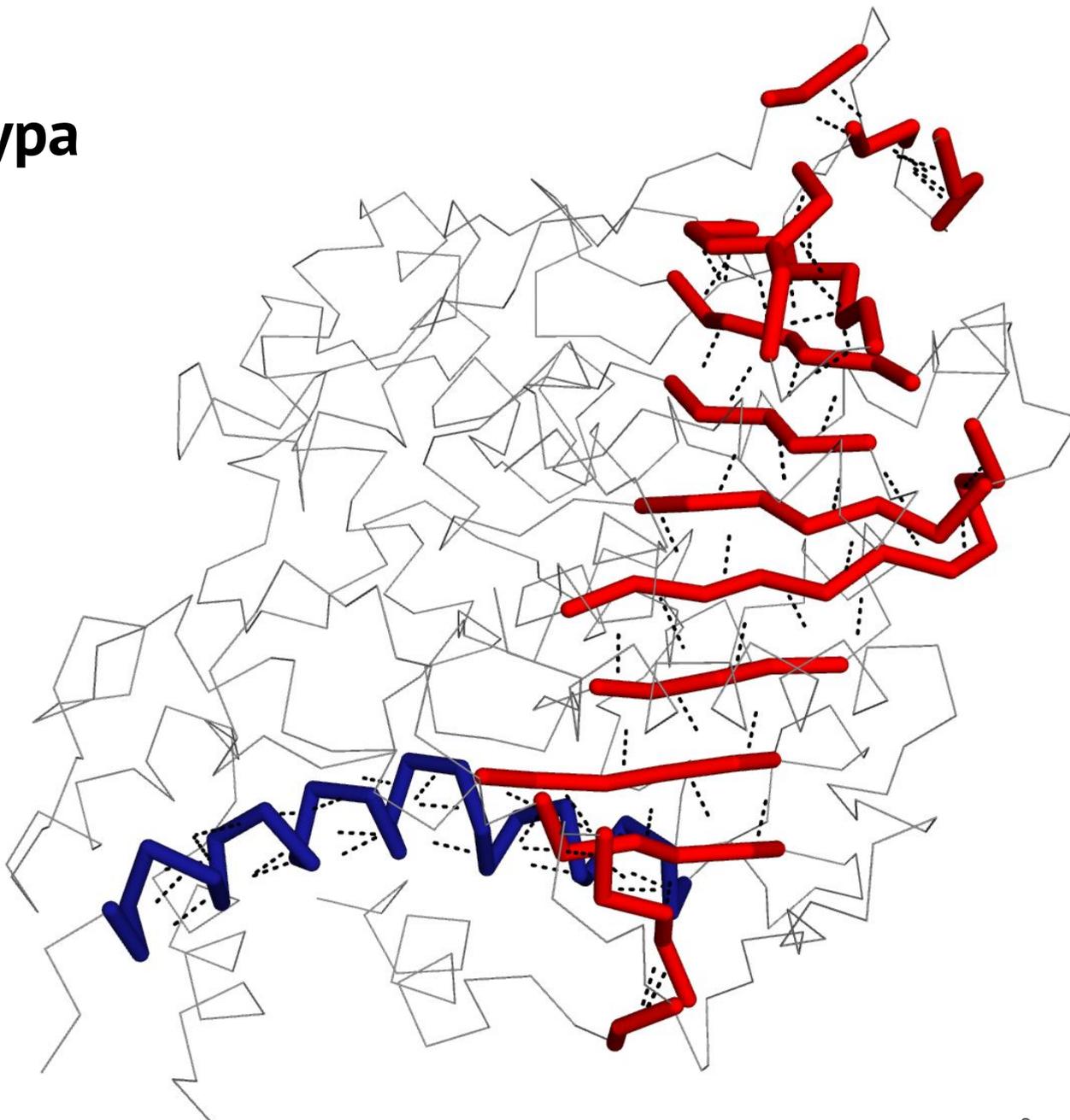


Вторичная структура

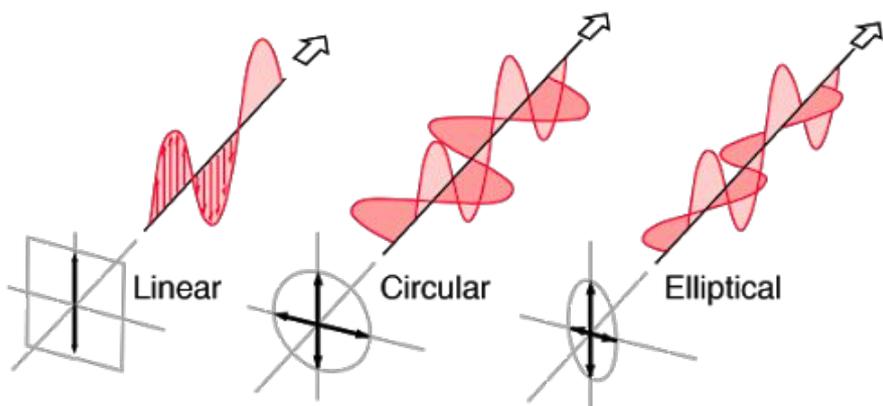
Вторичная структура -
паттерн во взаимном
расположении атомов
остова

Наслаивается на паттерн
меж-остовных
водородных связей
(определяется им?)

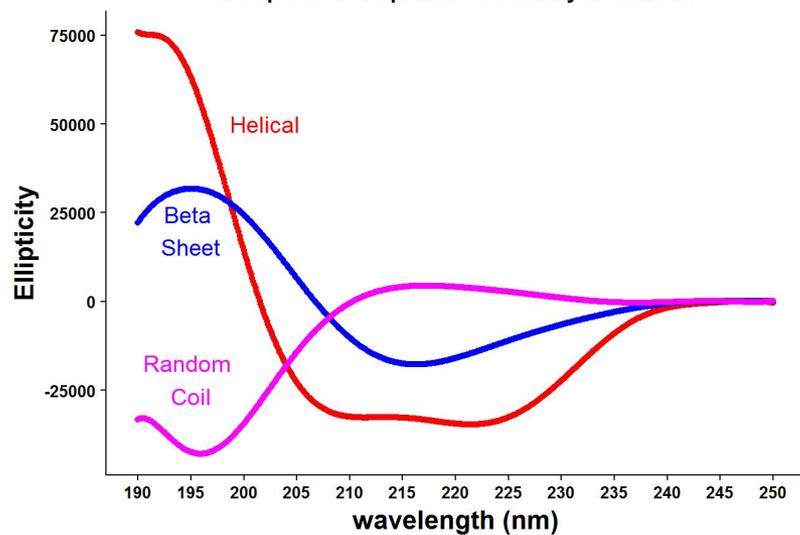
**Существует ли
“вторичная структура”
вне нашей головы?**



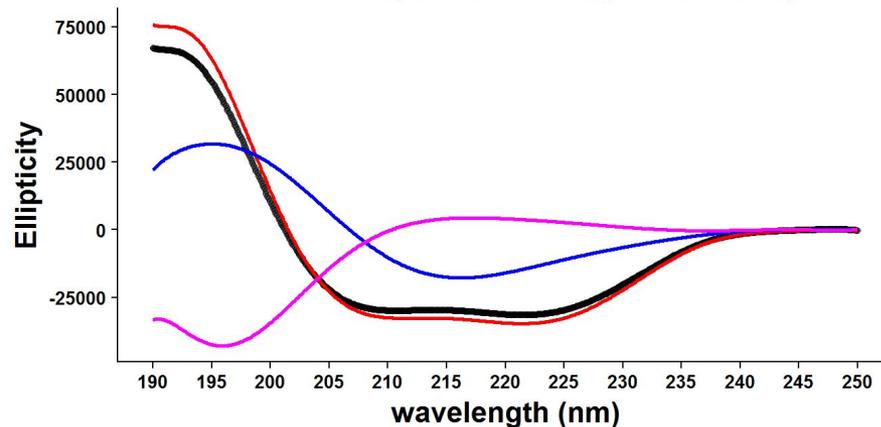
Спектроскопия кругового дихроизма



CD spectrum of protein secondary structures

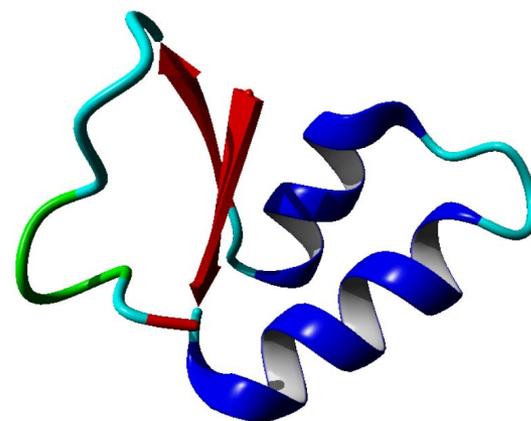
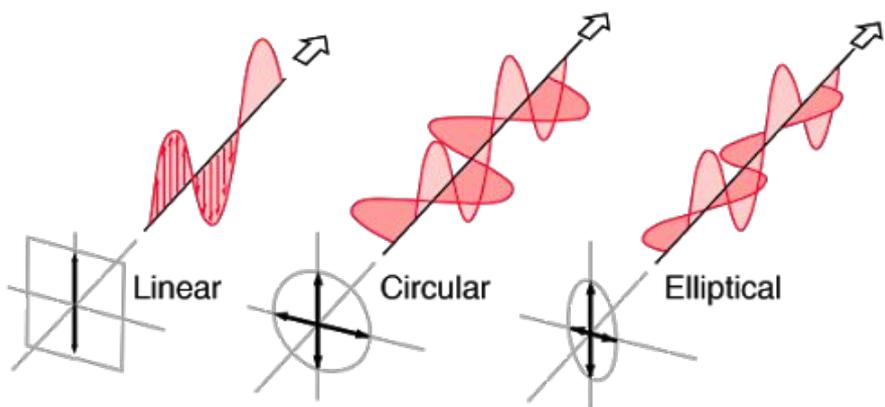


Predicted CD spectrum of BST2 (PDB ID: 3MQB)

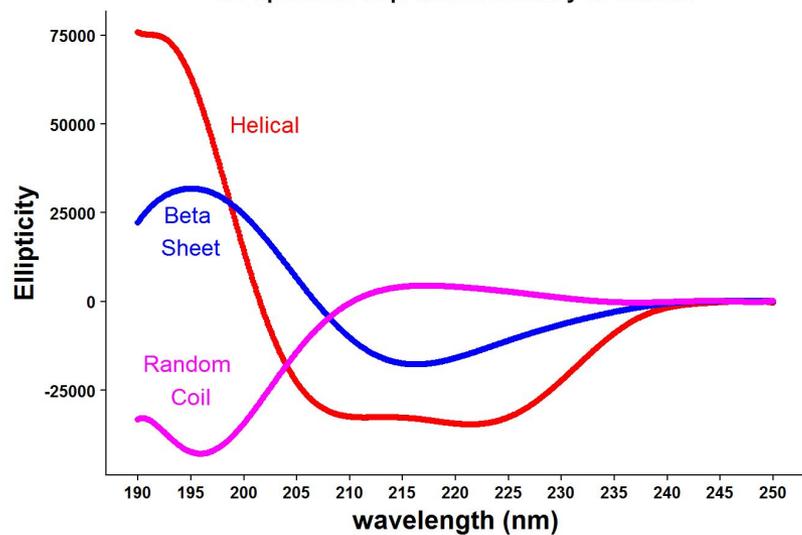


92% helix, 8% random coil

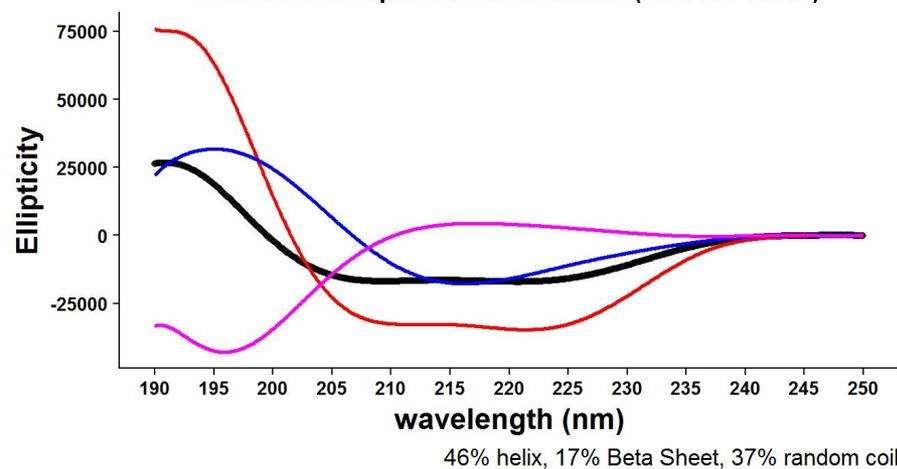
Спектроскопия кругового дихроизма



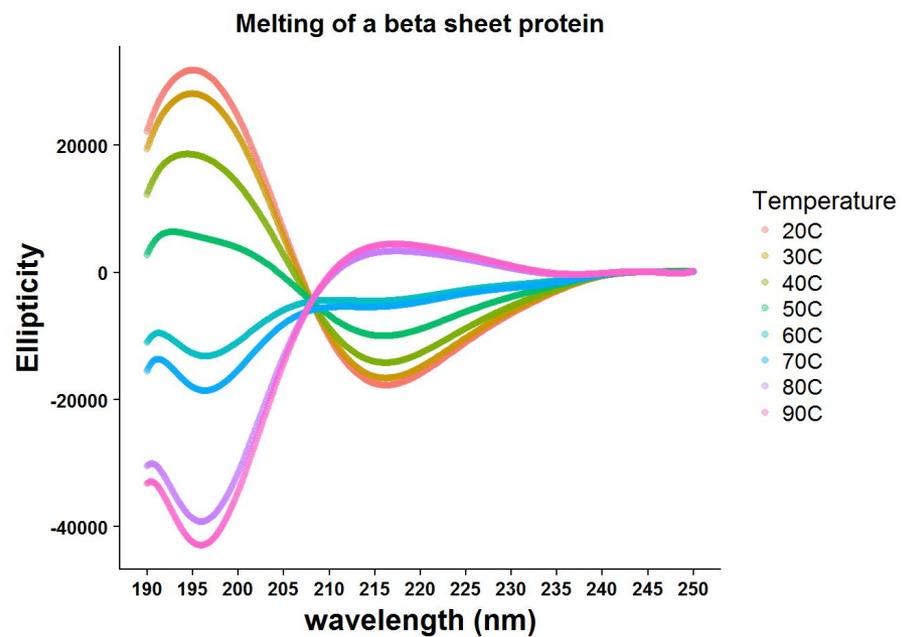
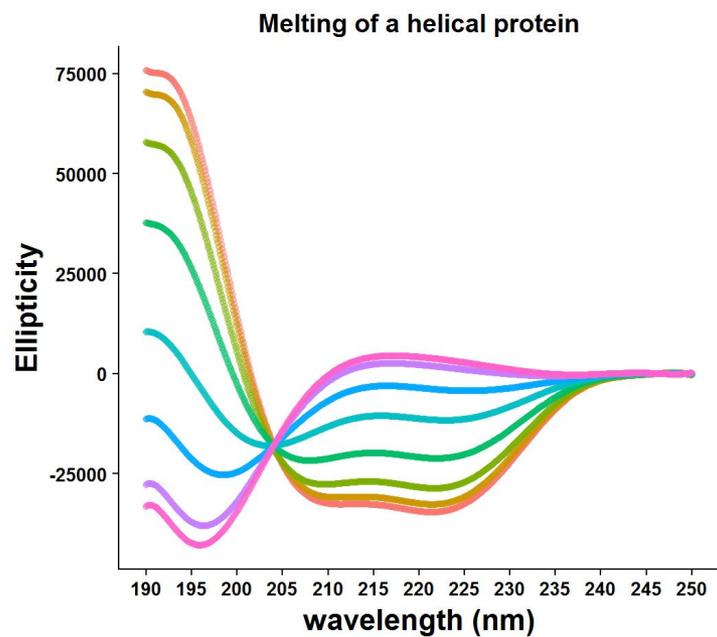
CD spectrum of protein secondary structures



Predicted CD spectrum of Crambin (PDB ID: 1CRN)



Спектроскопия кругового дихроизма

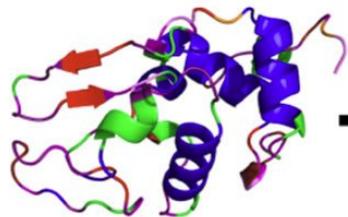


По структуре можно предсказать CD спектр

A

Secondary Structure Composition:

α -Helical (HEL): 34.1%
 β -strand (BS): 9.5%
 β -turns (BT): 4.7%
 3-Helical (3HT): 18.3%
 Irregular β (IRB): 1.6%
 LH-turns (LHT): 2.4%
 Other turns (OTT): 13.5%
 Unclassified (UC): 15.9%

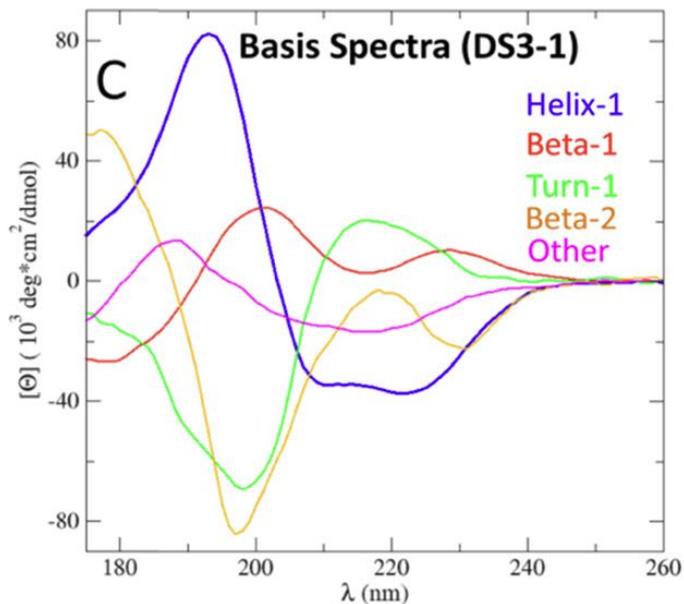


B

Assignment Matrix:

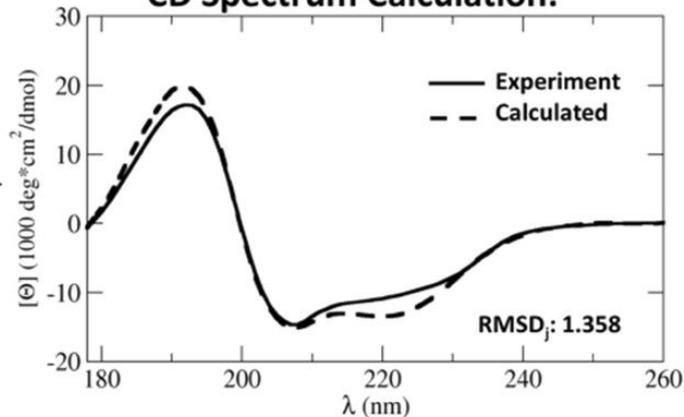
Assignment (α_{ki})		Helix-1	Beta-1	Turn-1	Beta-2	Other
DS3-1	k/i	1	2	3	4	5
HEL	1	1	0	0	0	0
BS	2	0	1	0	0	0
BT	3	0	1	0	0	0
3HT	4	0	0	1	0	0
IRB	5	0	0	0	1	0
LHT	6	0	0	0	0	1
OTT	7	0	0	0	0	1
UC	8	0	0	0	0	1

coefficients (C_{ji}): 0.341 0.142 0.183 0.016 0.318



D

CD Spectrum Calculation:



Вторичная структура и “вторичная структура”

- С т.з. эксперимента не важна бинарная принадлежность остатка к определенному виду вторичной структуры, а важен целиком состав белка. При этом сигнал зависит от типа и представленности вторичной структуры и не зависит от конкретной реализации (последовательности аминокислот)
- С т.з. структурной биологии мы не всегда однозначно можем сказать, к какому типу вторичной структуры принадлежит данный остаток – он может находиться где-то “посередине”
- С т.з. биоинформатики, однако, задача разметки вторичной структуры ставится как задача присвоения одной категориальной метки каждой позиции белка

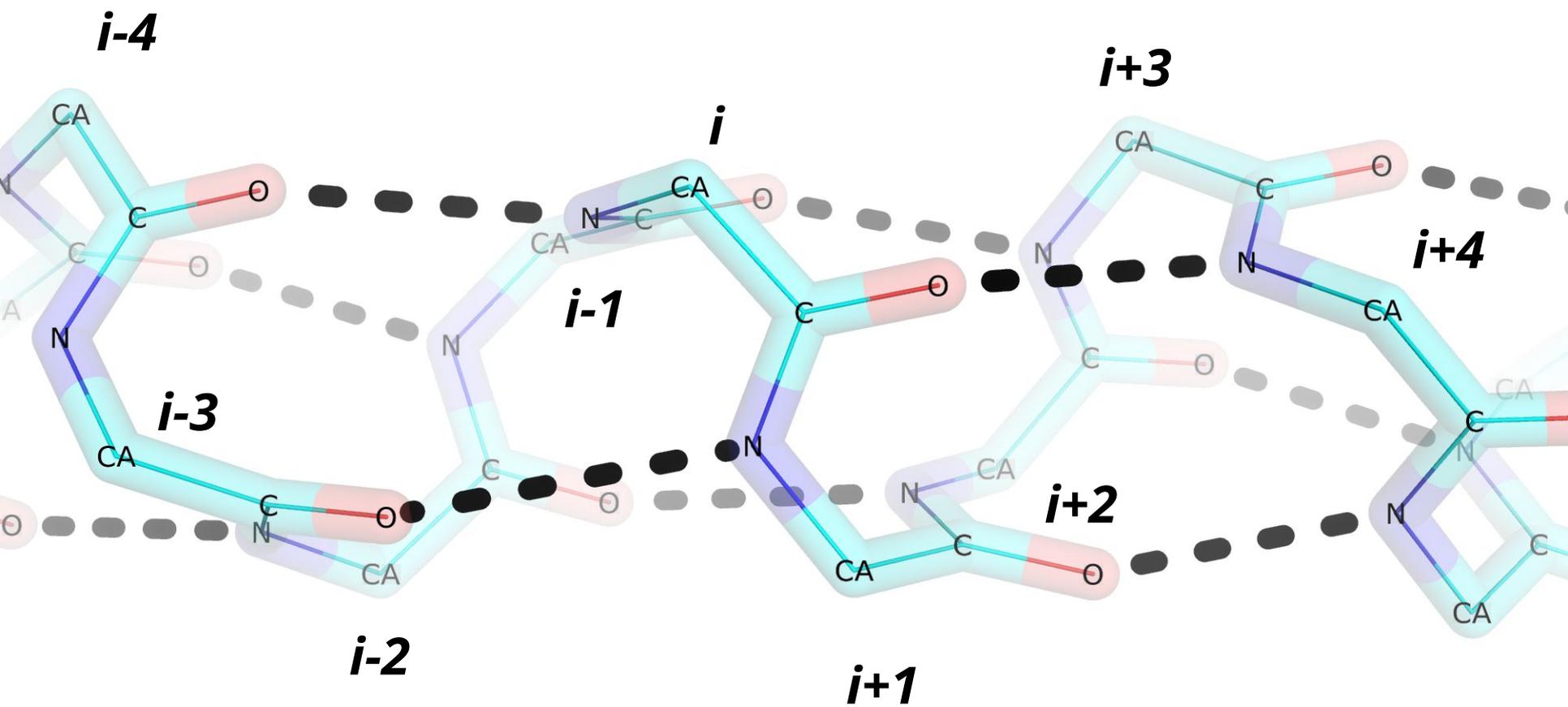
На основании чего мы можем попытаться выделить эти типы структур?

А на основании чего вообще выделяют типы?

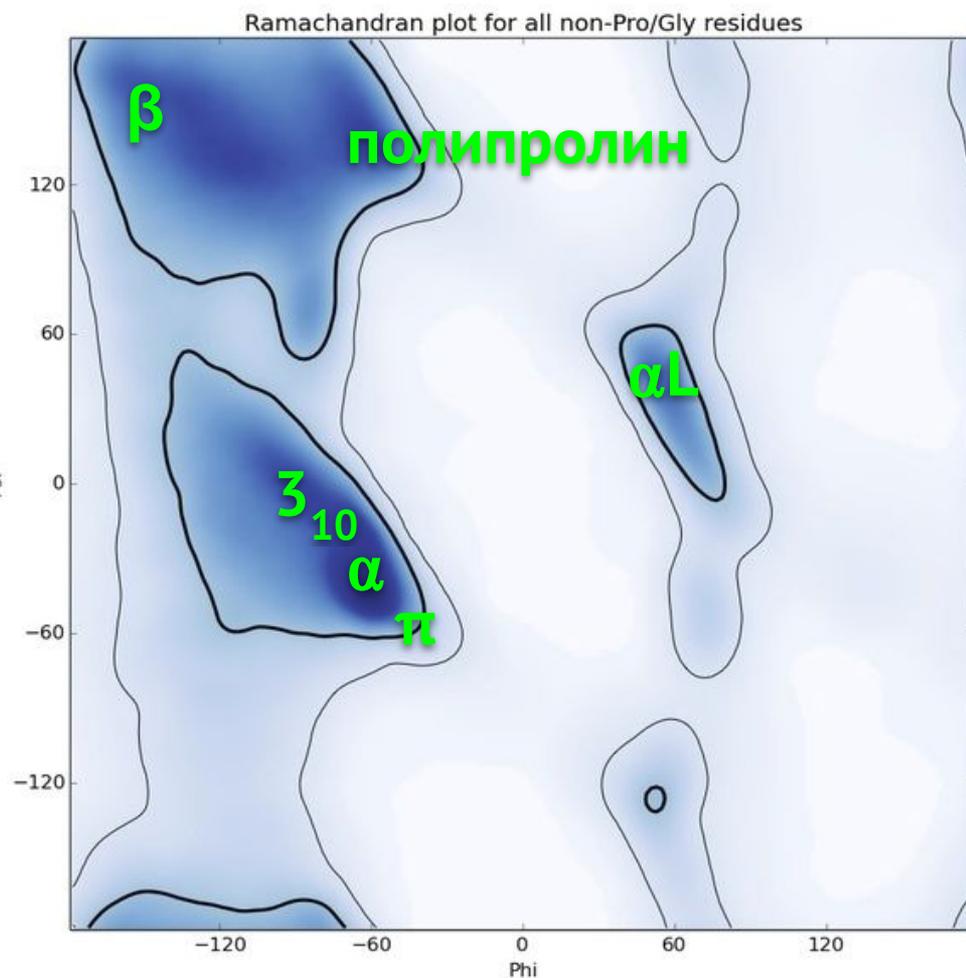
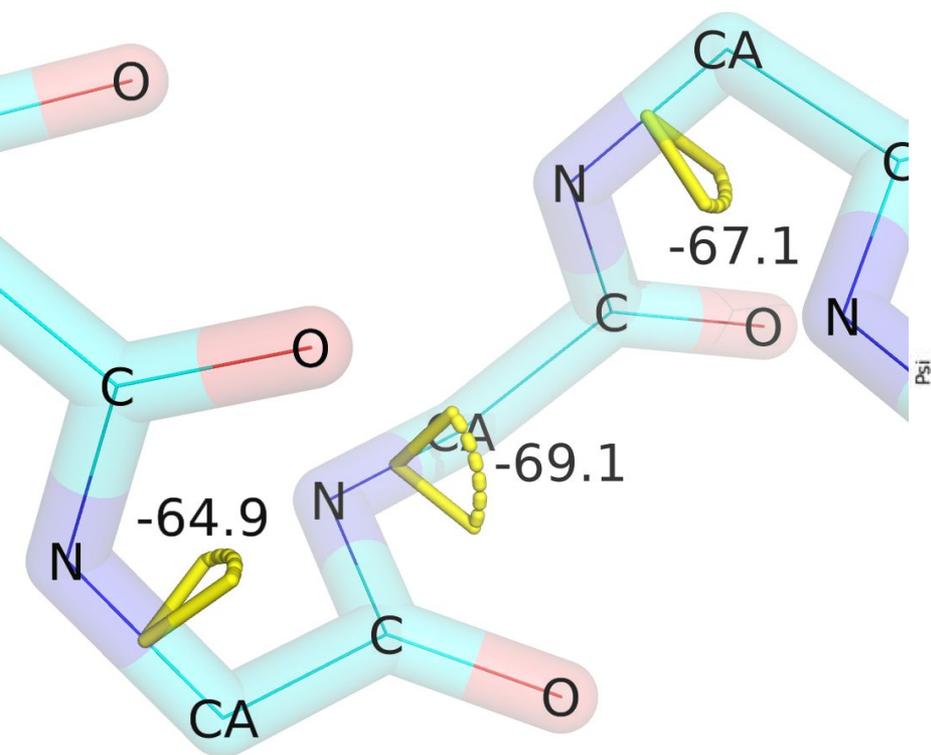
Перефразируя: введя какие дескрипторы мы можем увидеть, что вся остовная структура белка распадается на отдельные классы возможных остовных геометрий?

- Паттерн остов-остовных водородных связей
- Паттерн торсионных углов остова (ϕ , Ψ)
- Паттерн положения следующего остатка относительно положения предыдущих

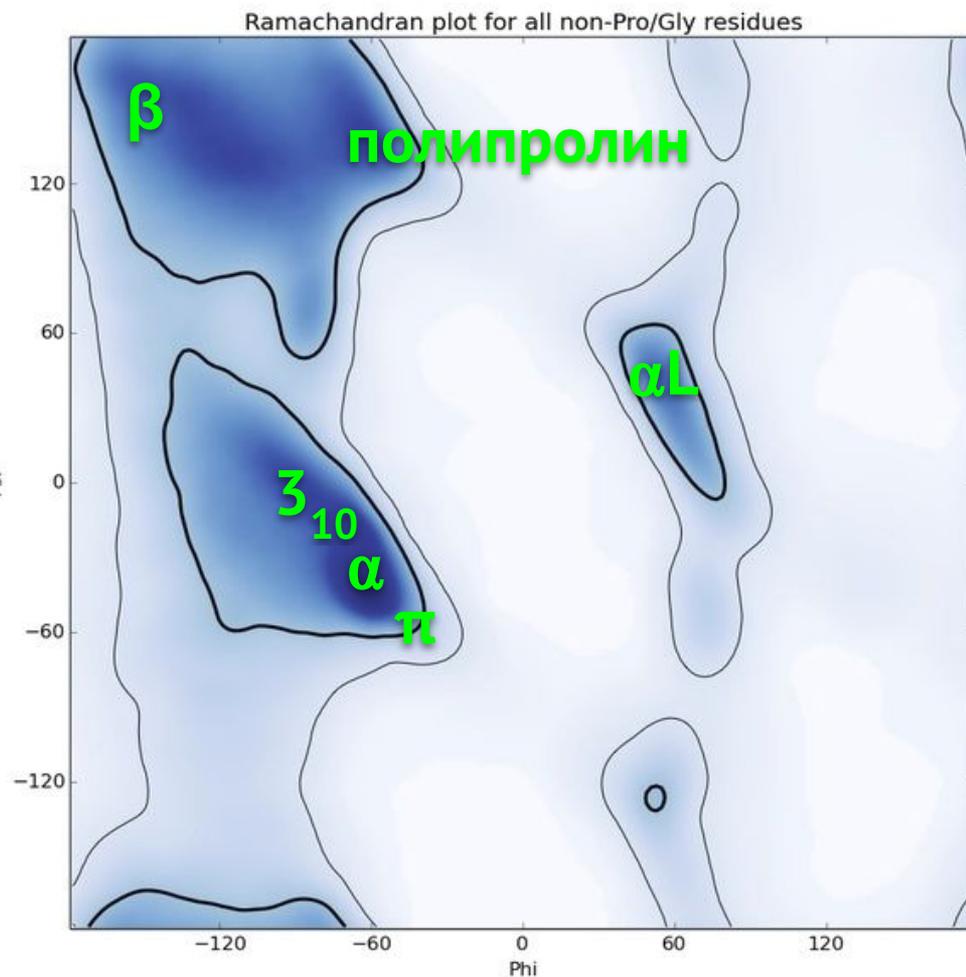
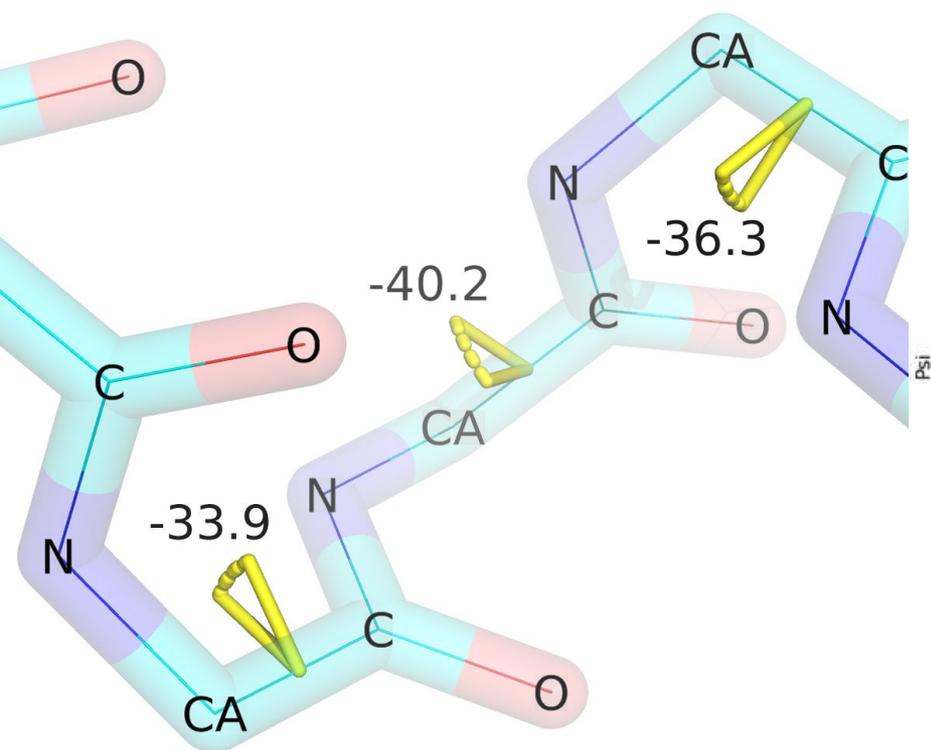
α -спираль и водородные связи



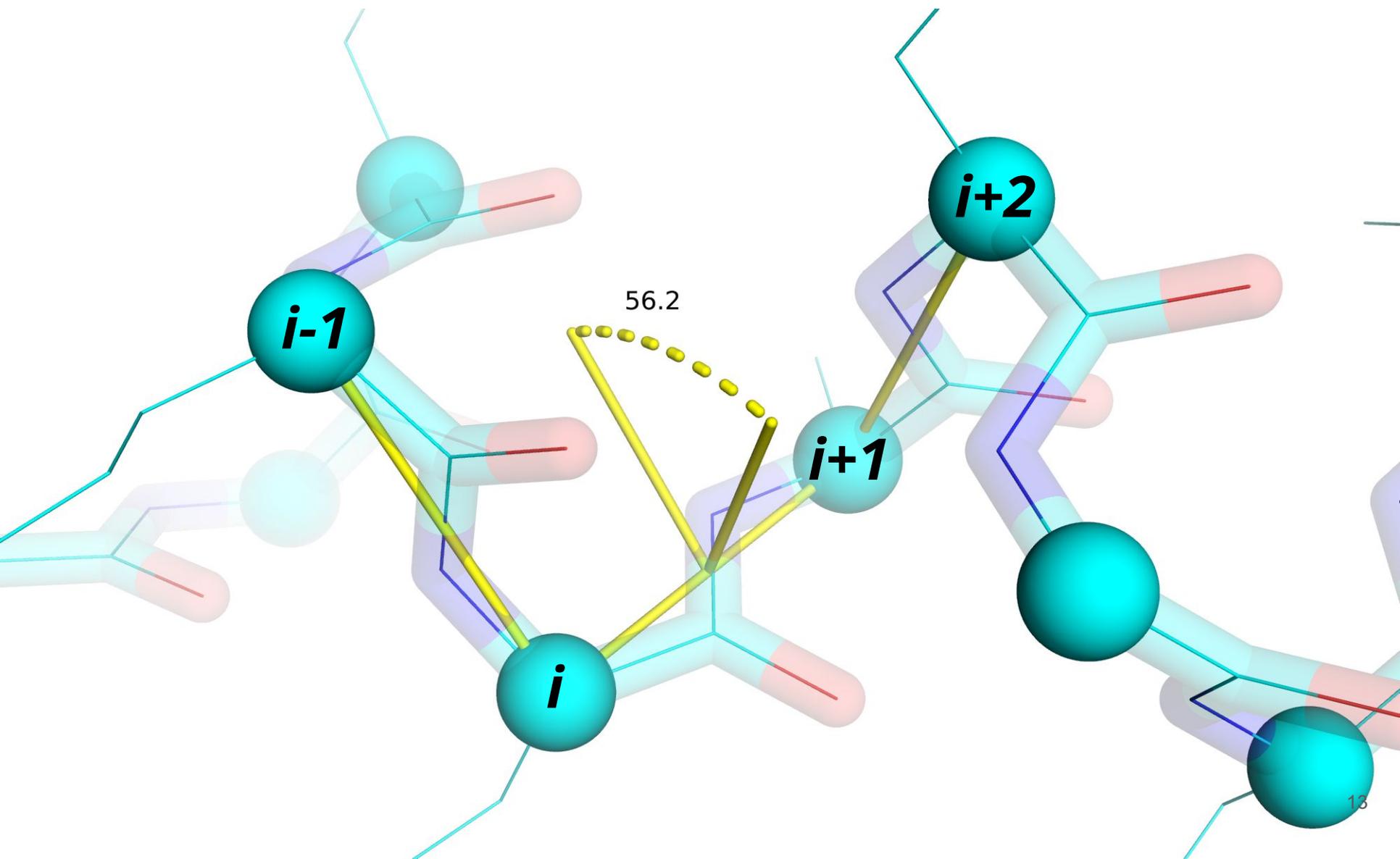
α -спираль и торсионные углы: ϕ



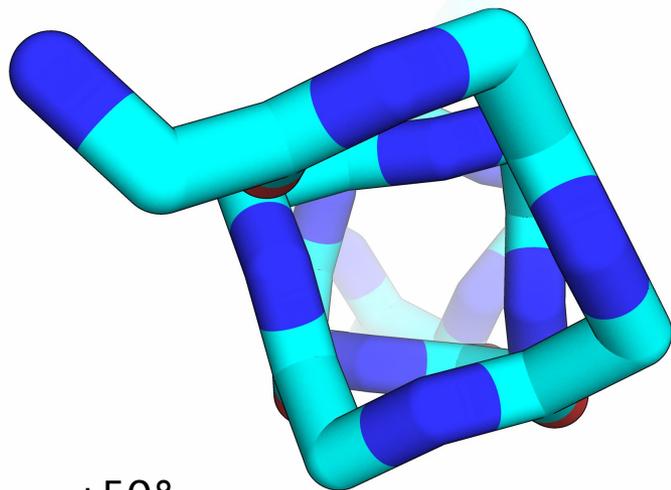
α -спираль и торсионные углы: Ψ



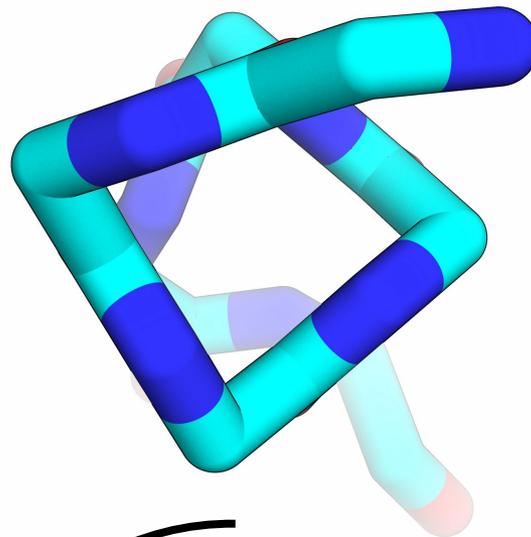
α -спираль и закрученность



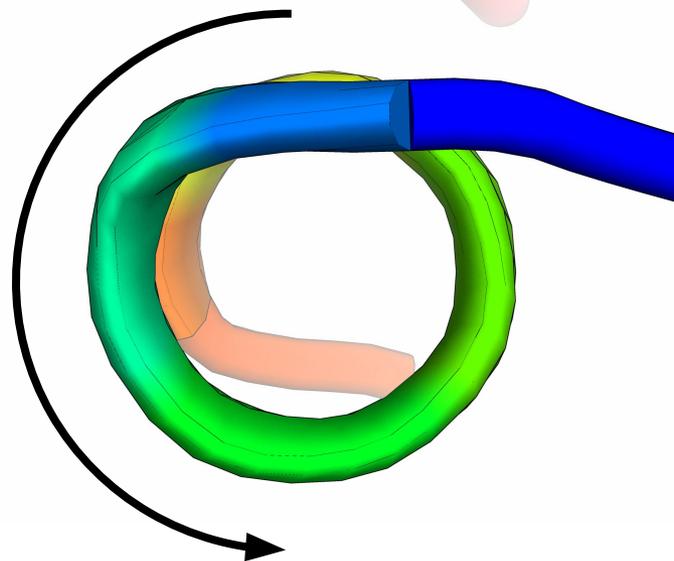
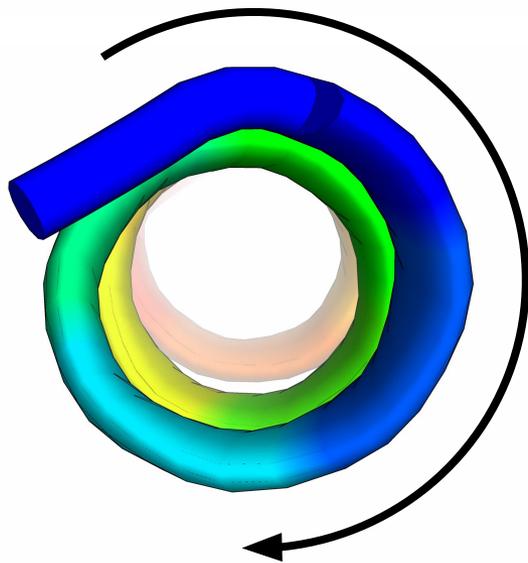
α -спираль и закрученность



+50°

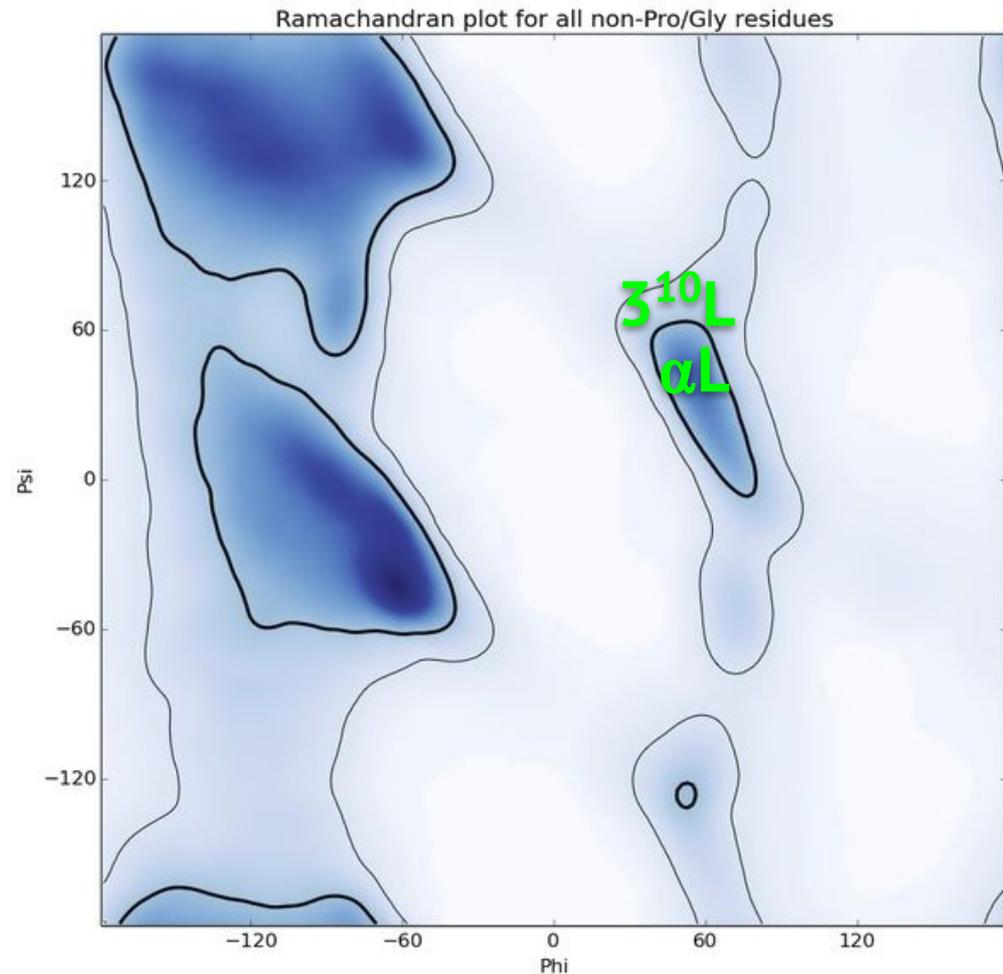


-50°



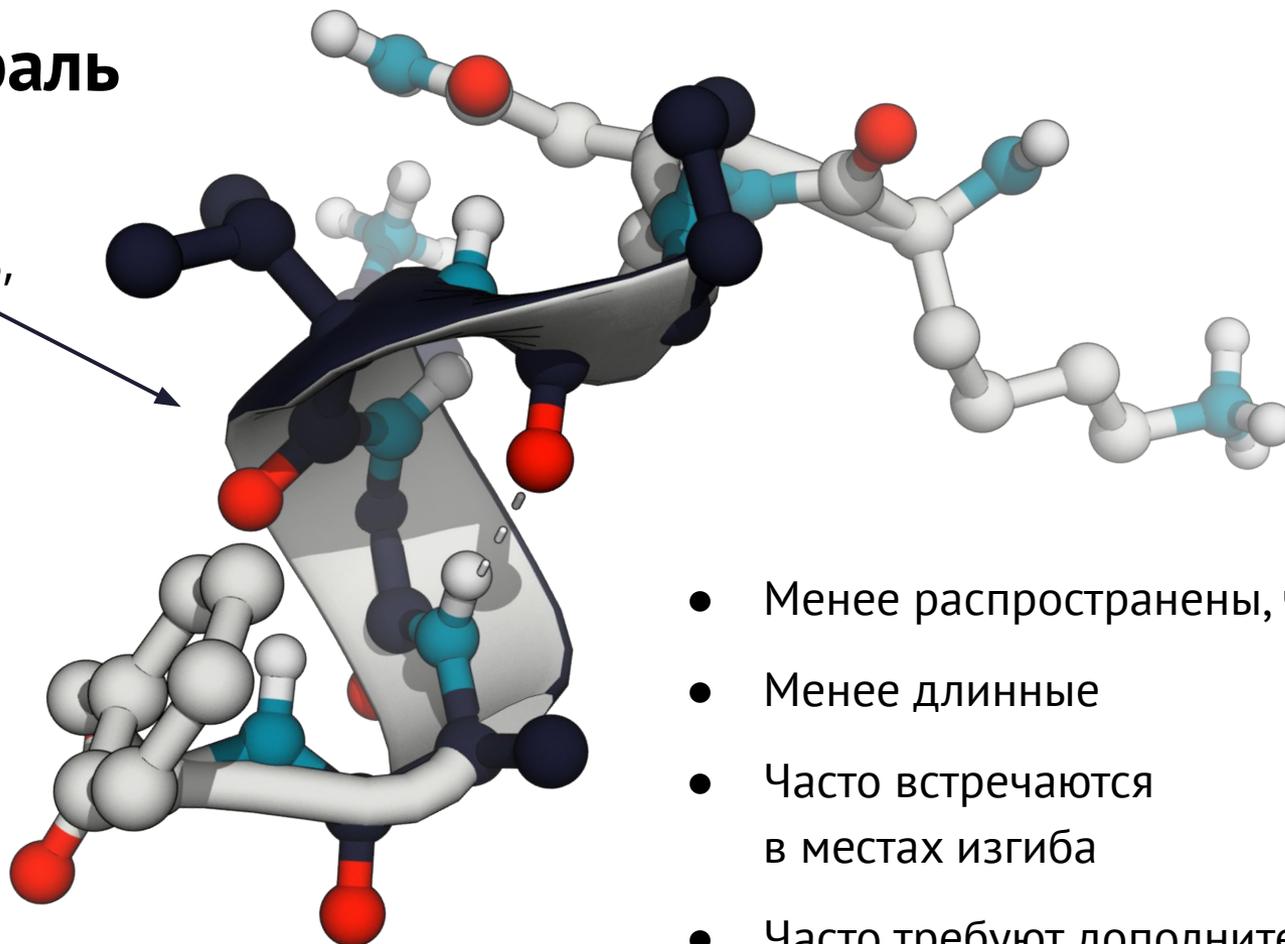
Левозакрученная спираль

- Значительно менее распространены, чем правозакрученные
- Если есть, то короткие
- Если есть, то не случайно – ожидайте рядом функционально значимые остатки



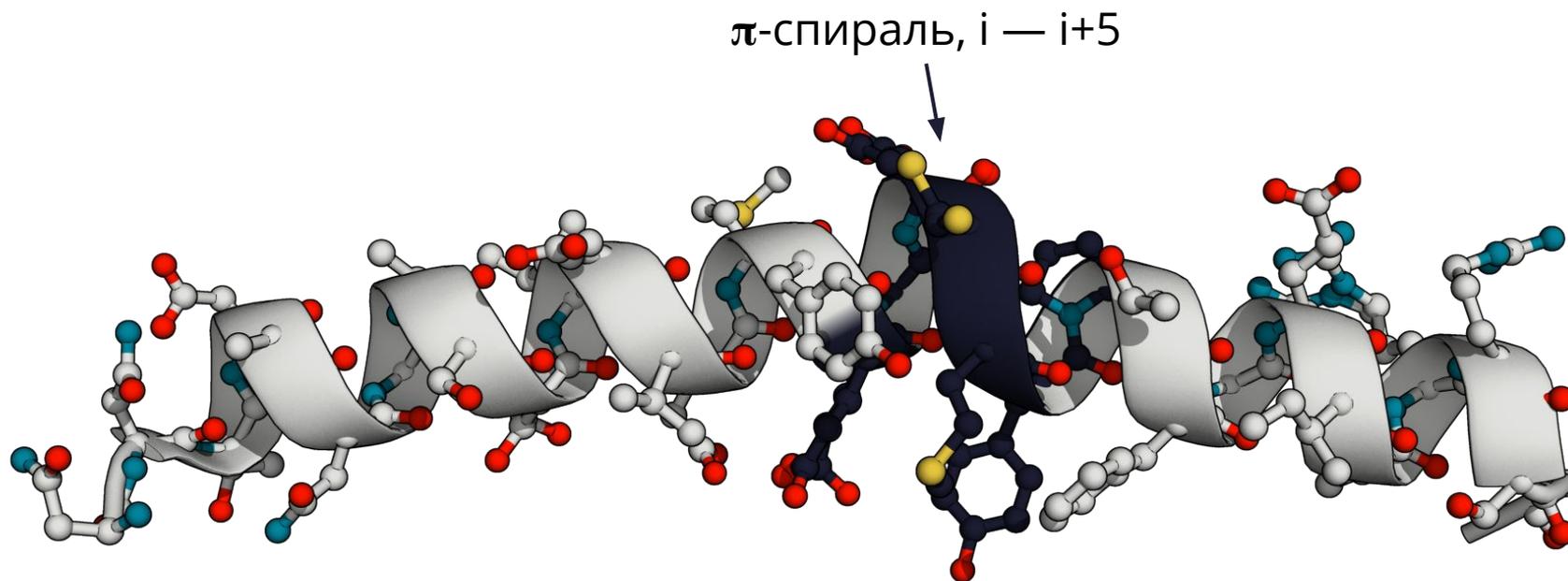
3_{10} -спираль

3_{10} -спираль,
 $i - i+3$



- Менее распространены, чем α
- Менее длинные
- Часто встречаются в местах изгиба
- Часто требуют дополнительной поддержки на конце со стороны сайдчейнов

π -спираль

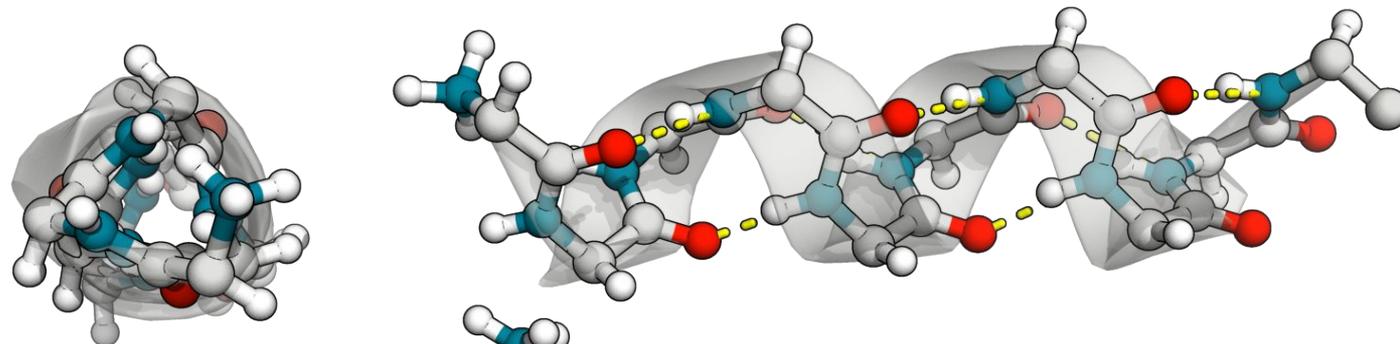


Еще более редкие, чем 3_{10} -спирали

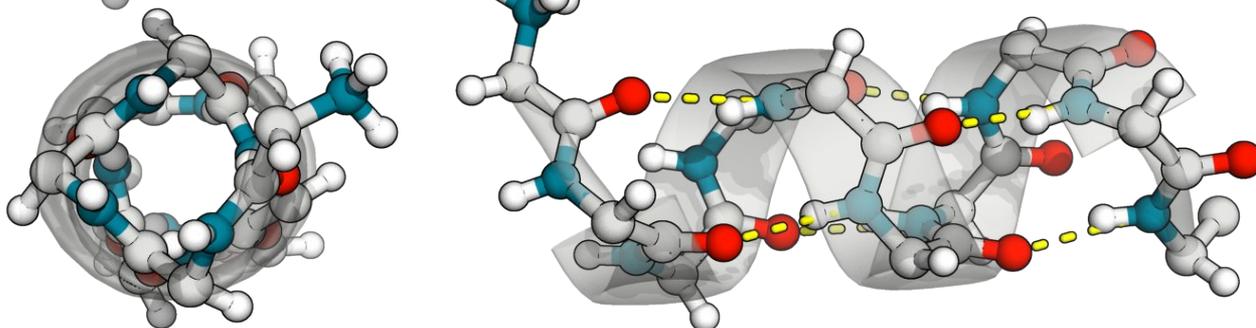
Изгибают альфа-спираль

Все спирали

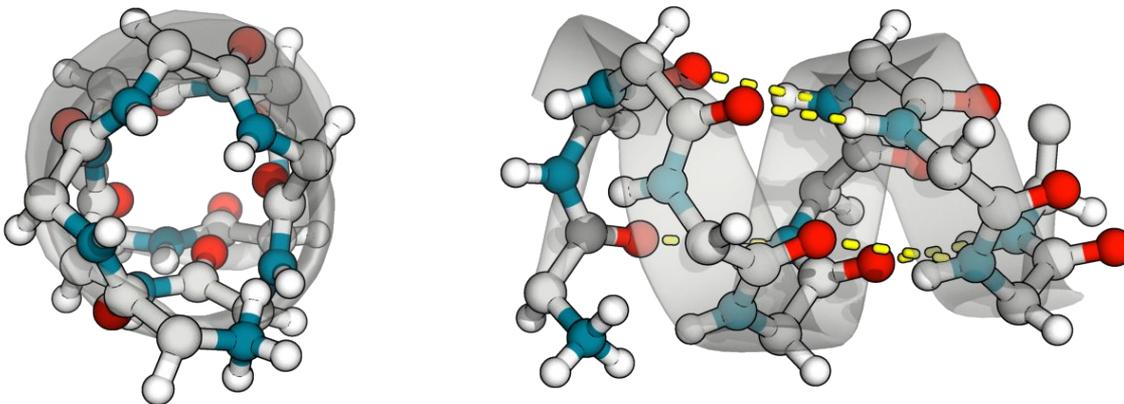
3_{10}



$4_{13'} \alpha$



$5_{16'} \pi$



Или даже так

› [Biopolymers](#). 2013 Nov;100(6):621-36. doi: 10.1002/bip.22267.

Multiple, consecutive, fully-extended 2.0₅-helix peptide conformation

[Cristina Peggion](#)¹, [Alessandro Moretto](#), [Fernando Formaggio](#), [Marco Crisma](#), [Claudio Toniolo](#)

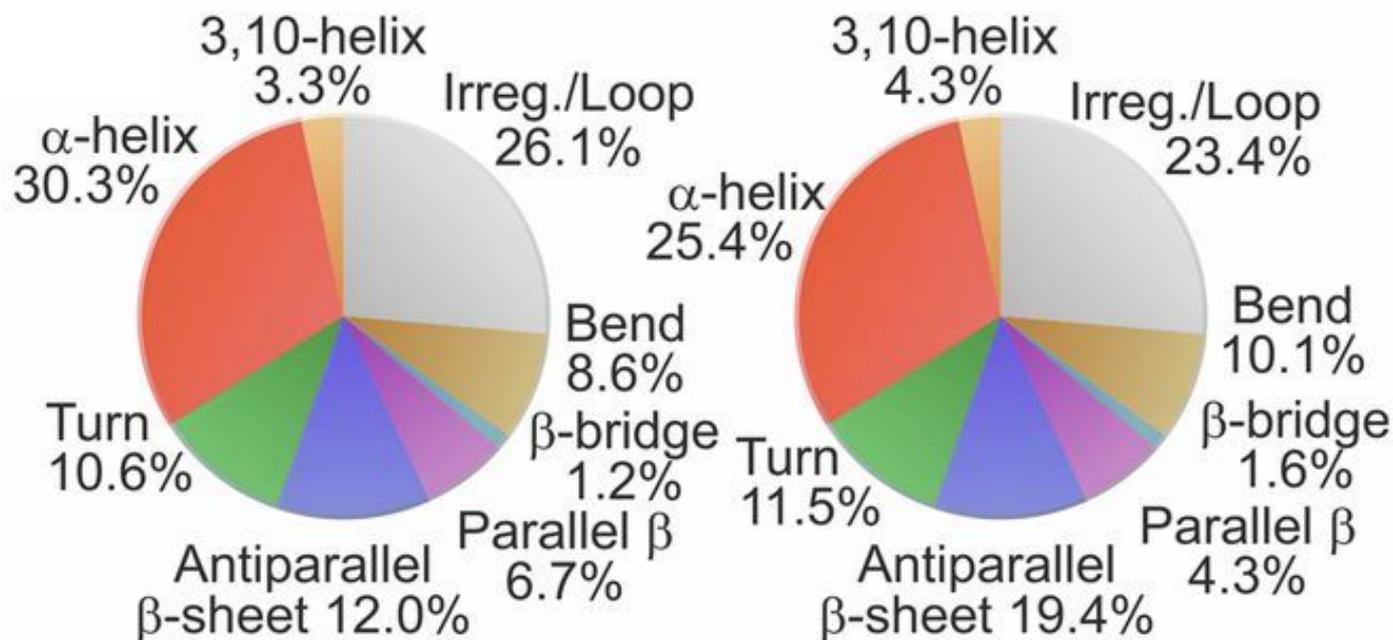
Affiliations + expand

PMID: 23893391 DOI: [10.1002/bip.22267](#)

Abstract

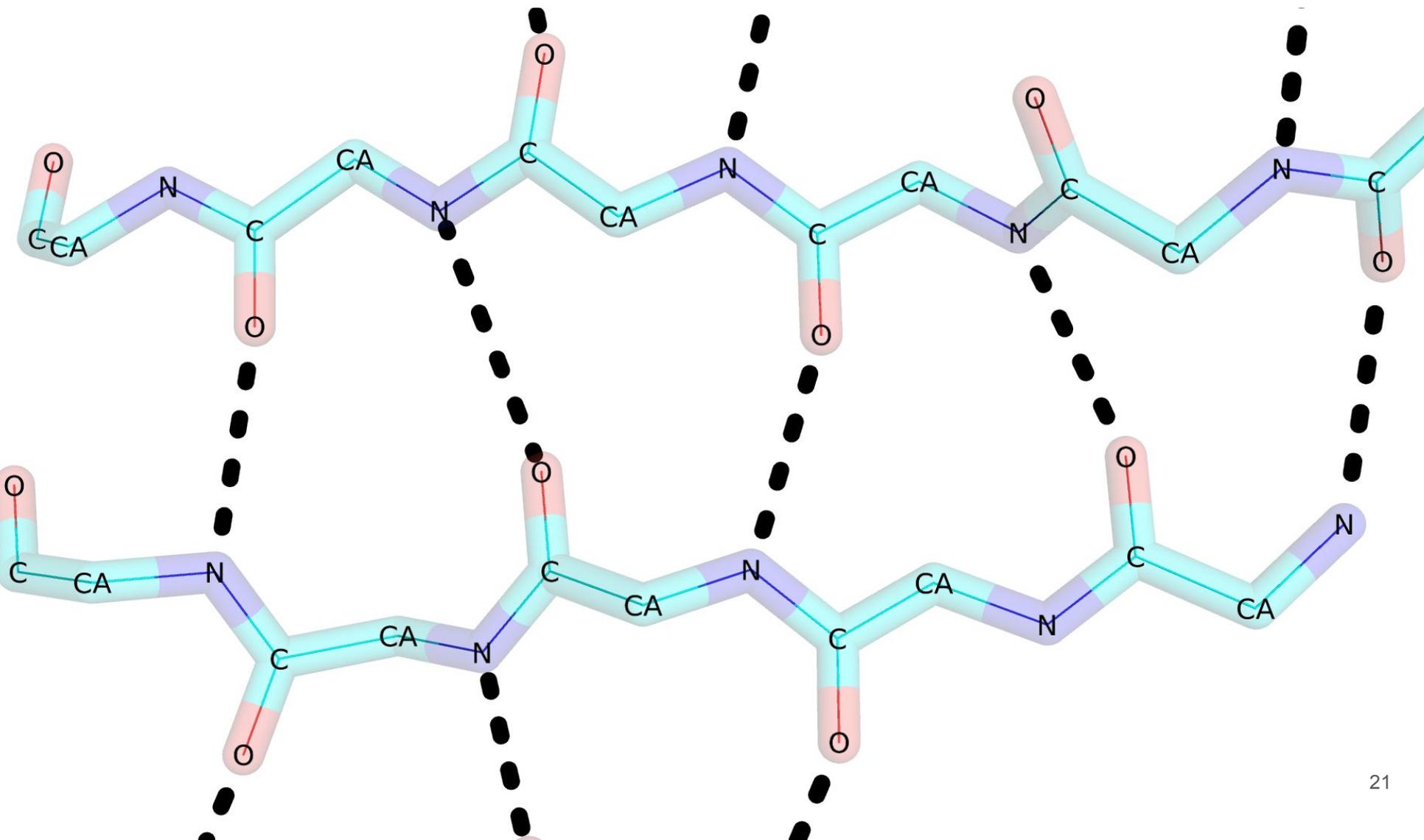
The peptide 2.0(5)-helix does exist. It has been experimentally authenticated both in the crystalline state (by X-ray diffraction) and in solution (by several spectroscopic techniques). It is the most

Но не стоит об этом сильно заморачиваться

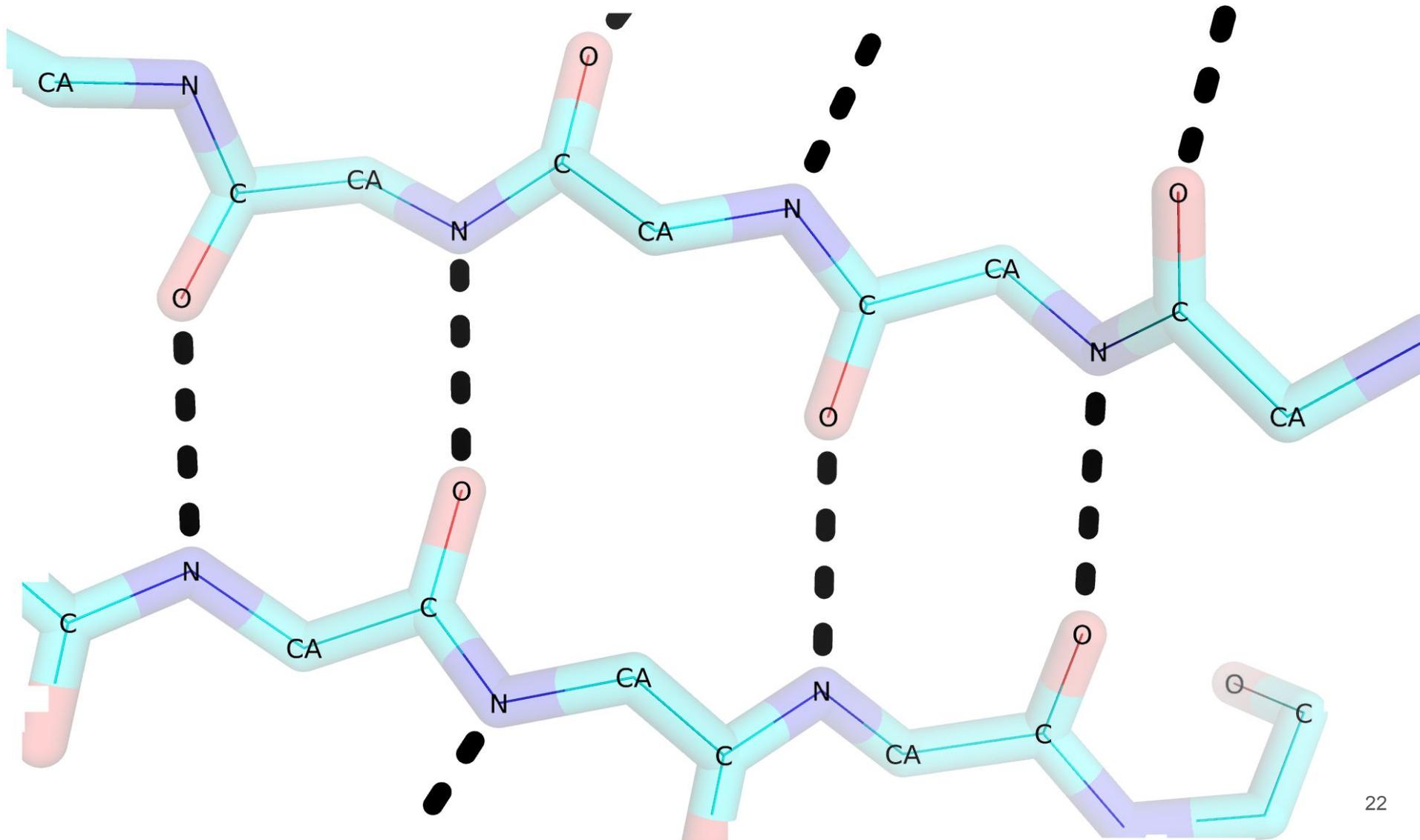


PDB	Class	SP175
4	Architecture	4
40	Topology	19
1313		53

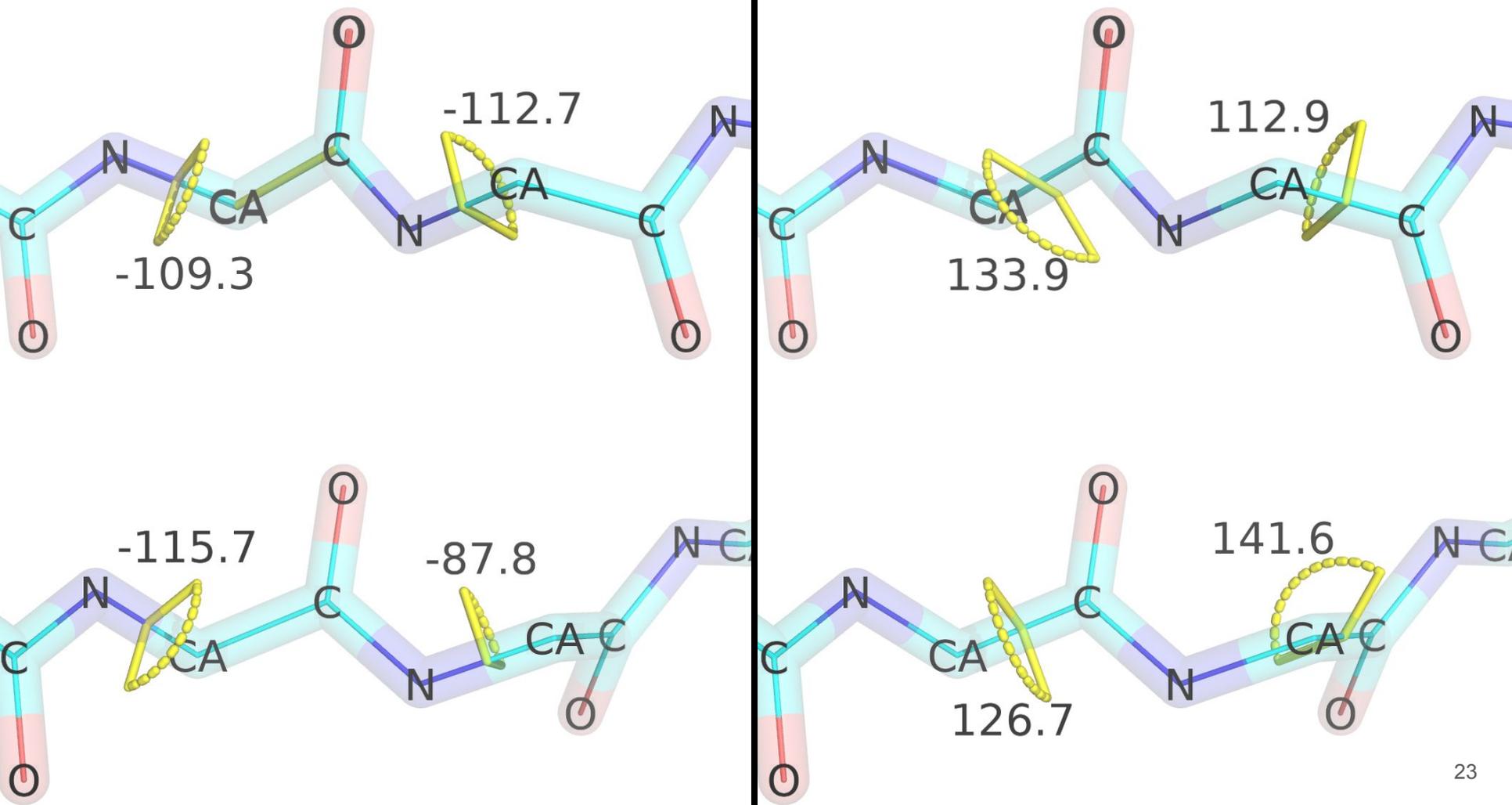
Параллельный β -лист



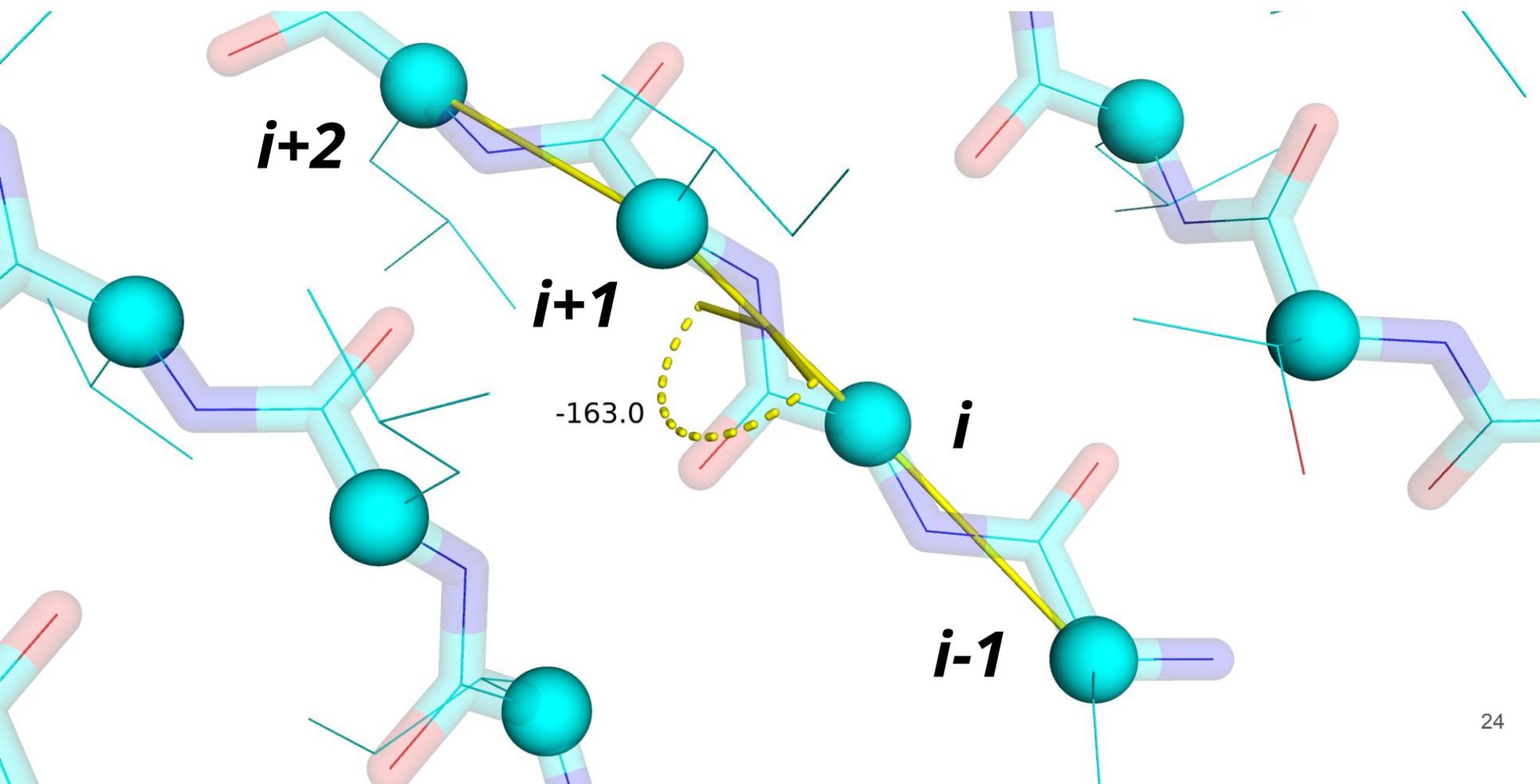
Антипараллельный β -лист



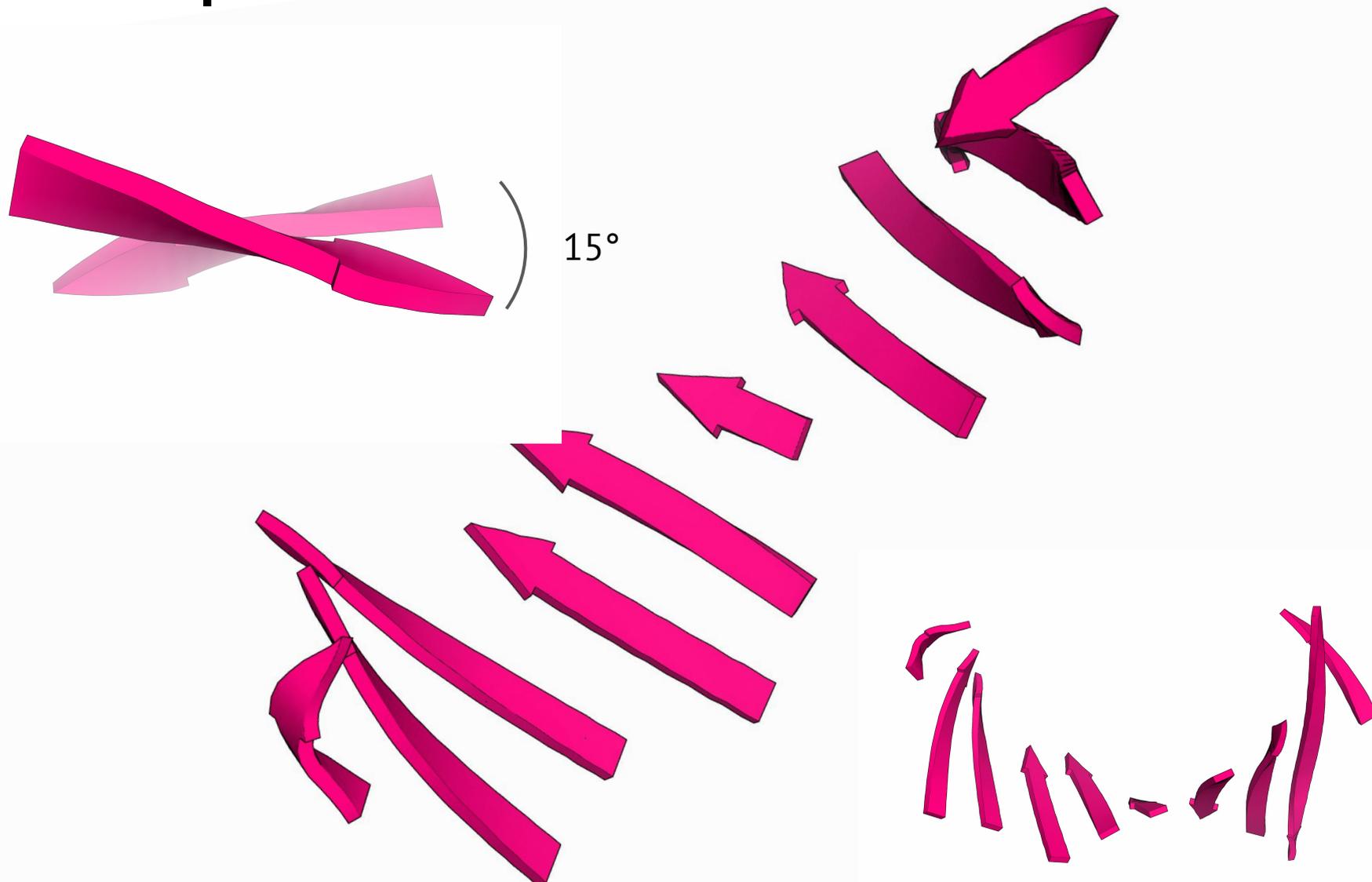
β-лист и торсионные углы

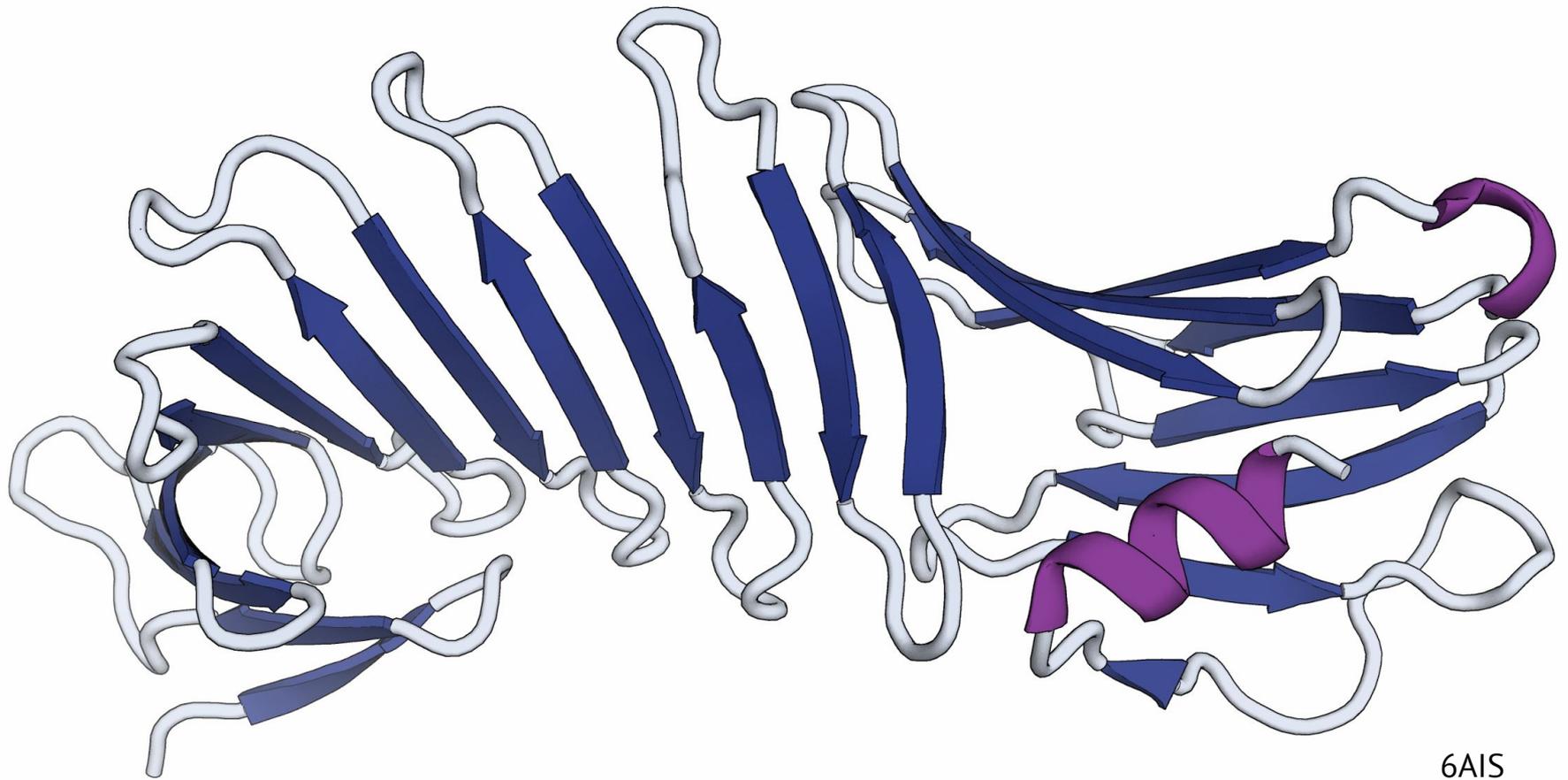


β -лист и закрученность



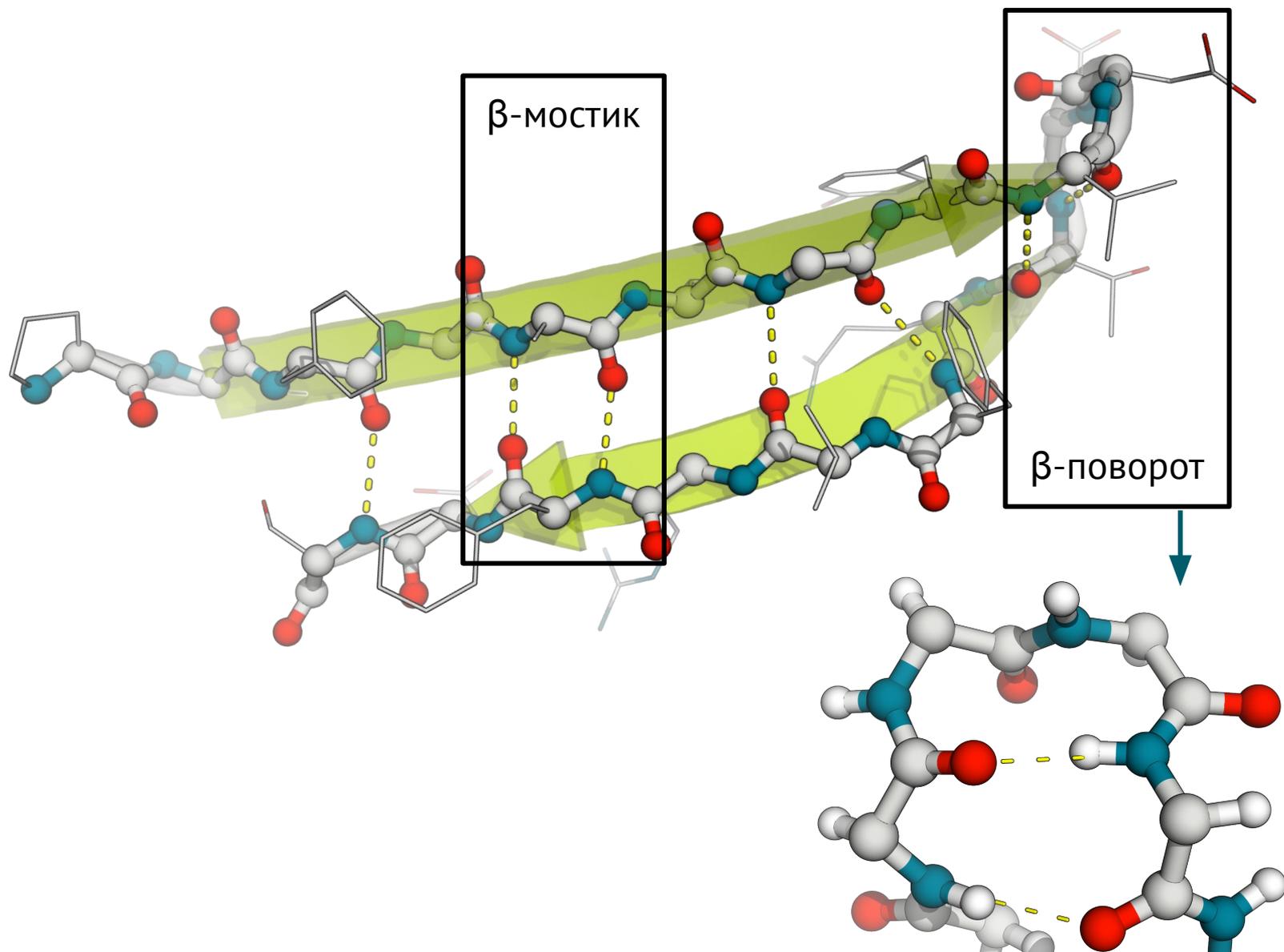
Изгибы β -листов





6AIS

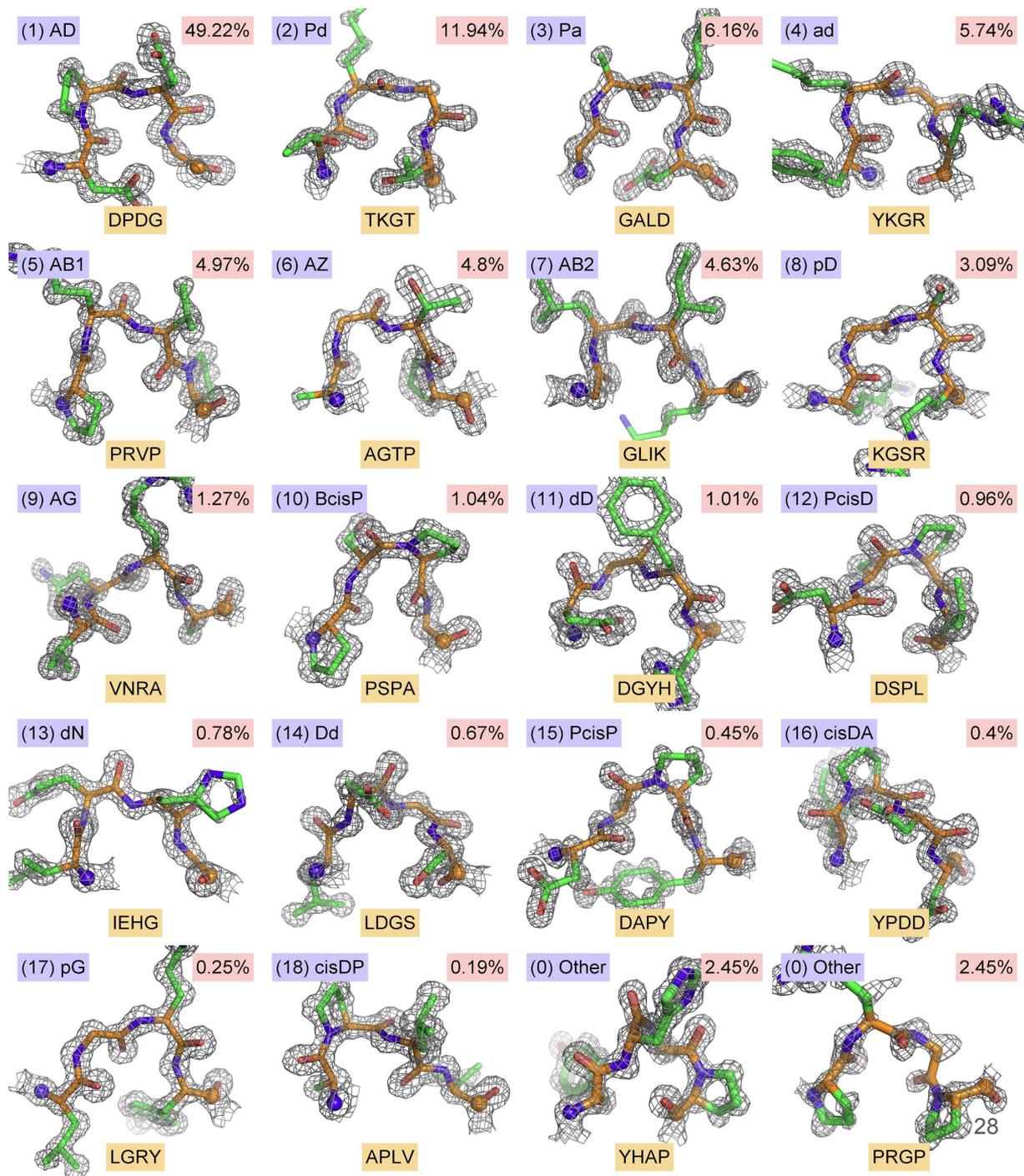
β-лист



β -кошмар (повороты)

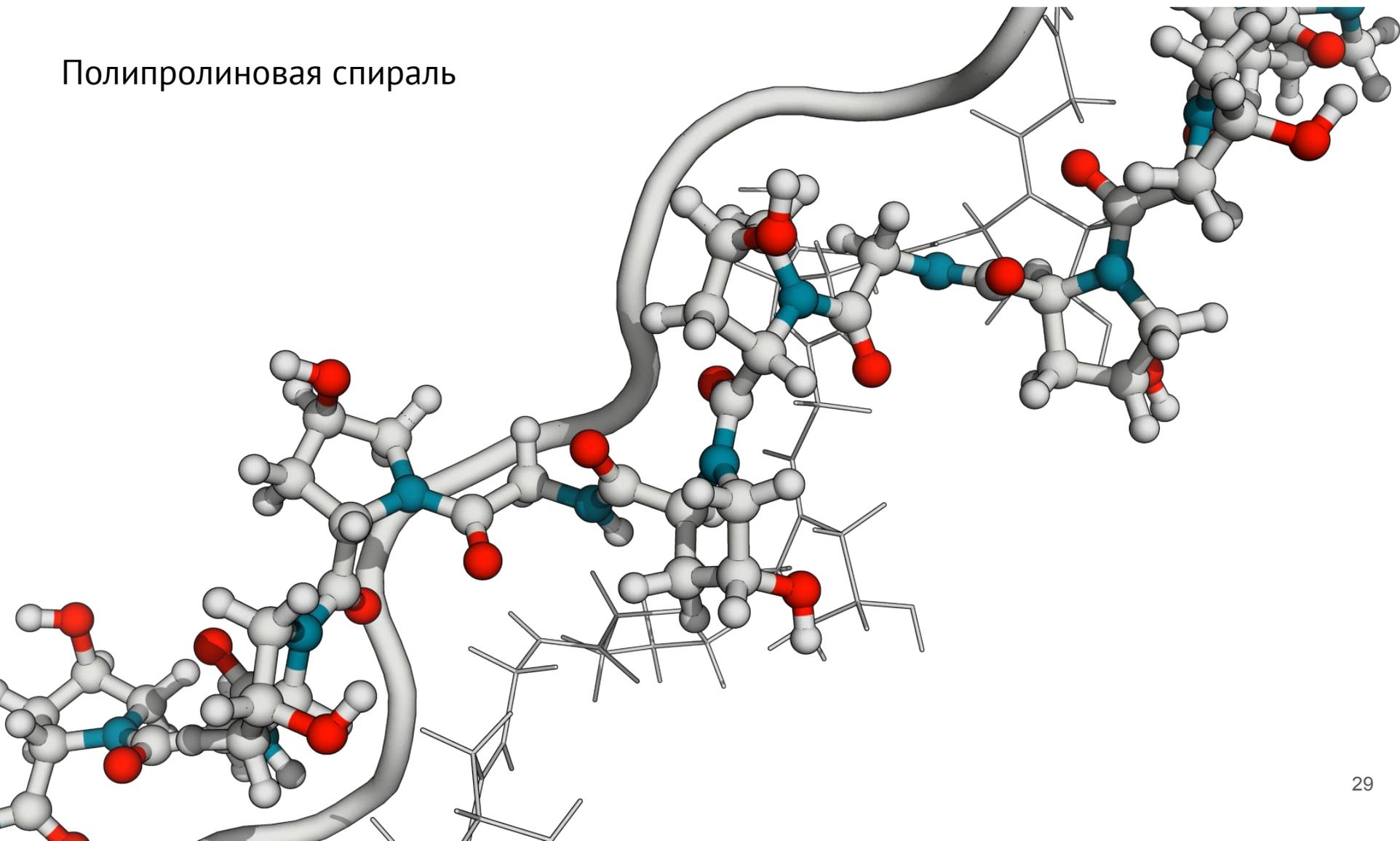
Одна из свежих
классификаций выделяет
аж 18 типов

Shapovalov M, Vucetic S, Dunbrack
RL Jr (2019) **A new clustering and
nomenclature for beta turns
derived from high-resolution
protein structures.** *PLoS Comput
Biol*



Вторичная структура без водородных связей

Полипролиновая спираль



В моем белке бета-лист. И что?

- Жесткость листа позволяет ему служить передатчиком механического сигнала на большие расстояния

**nature
communications**

Article | Published: 11 June 2014

Correlated motions are a fundamental property of β -sheets

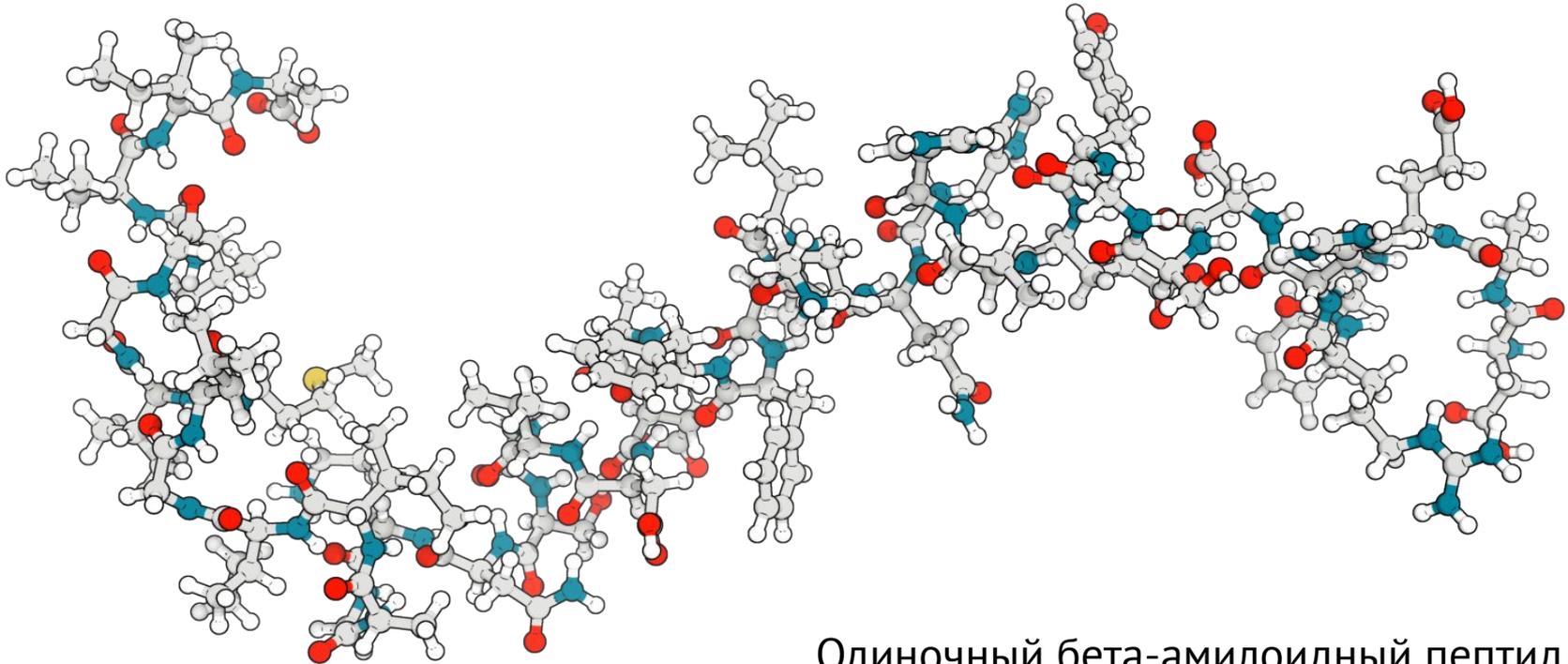
R. Bryn Fenwick, Laura Orellana, Santi Esteban-Martín, Modesto Orozco  & Xavier Salvatella 

Nature Communications 5, Article number: 4070 (2014) | [Cite this article](#)

- Бета-листовые структуры склонны к агрегации

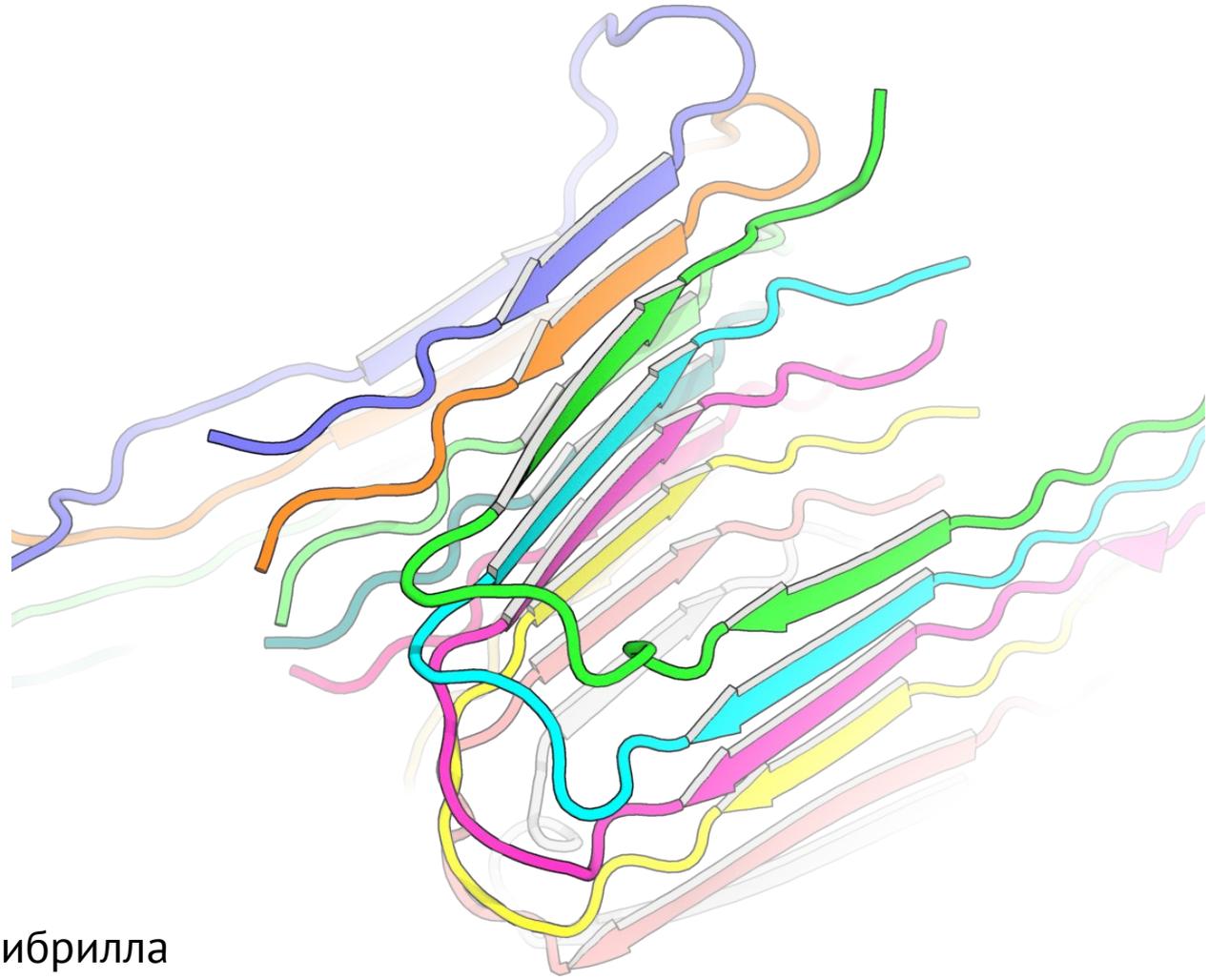
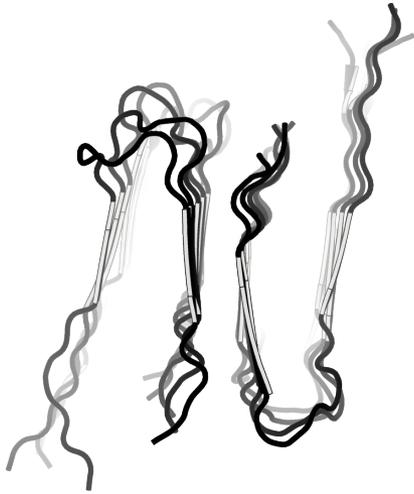


Амилоиды



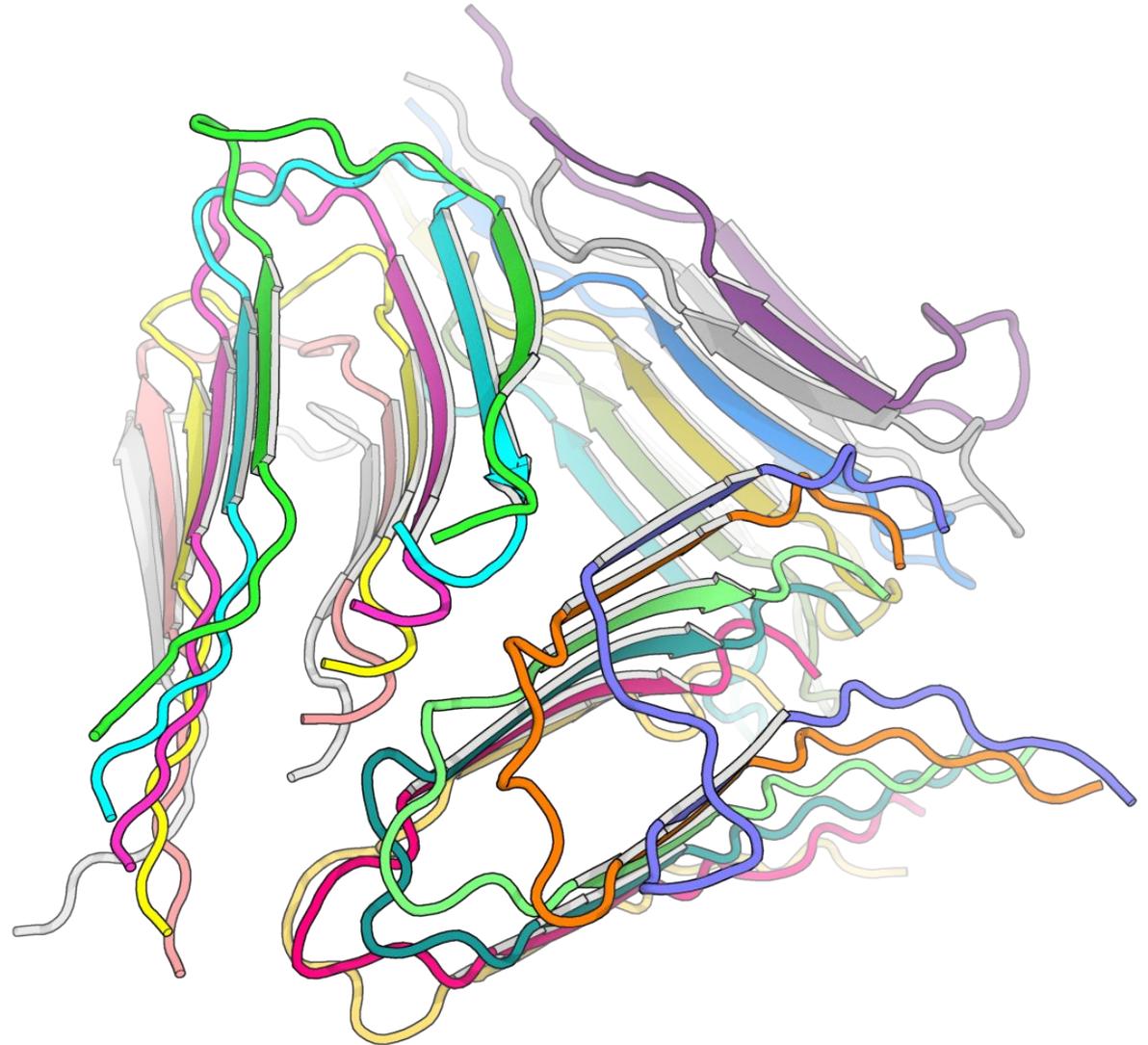
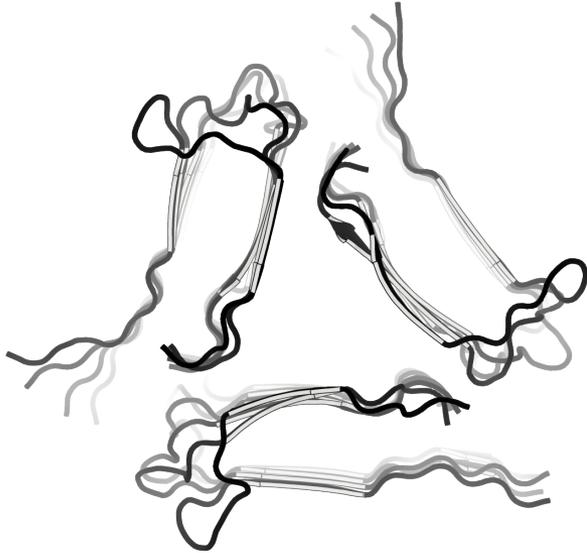
Одиночный бета-амилоидный пептид

Амилоиды



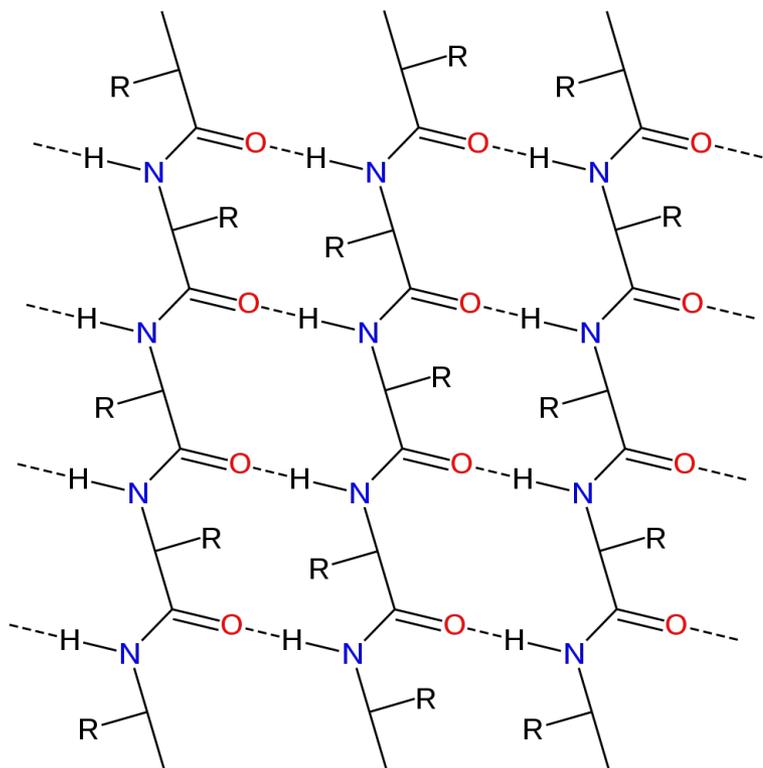
2-fold бета-амилоидная фибрилла

Амилоиды



3-fold бета-амилоидная
фибрилла

Забытая вторичная структура – альфа-лист



npj | Biofilms and Microbiomes

Article | [Open Access](#) | Published: 03 July 2017

Designed α -sheet peptides suppress amyloid formation in *Staphylococcus aureus* biofilms

Alissa Bleem, Robyn Francisco, James D. Bryers & Valerie Daggett [✉](#)

npj Biofilms and Microbiomes 3, Article number: 16 (2017) | [Cite this article](#)

539 Accesses | 8 Citations | 16 Altmetric | [Metrics](#)

α -Sheet secondary structure in amyloid β -peptide drives aggregation and toxicity in Alzheimer's disease



Dylan Shea, Cheng-Chieh Hsu, Timothy M. Bi, Natasha Paranjapye, Matthew Carter Childers, Joshua Cochran, Colson P. Tomberlin, Libo Wang, Daniel Paris, Jeffrey Zonderman, Gabriele Varani, Christopher D. Link, Mike Mullan, and Valerie Daggett

PNAS April 30, 2019 116 (18) 8895-8900; first published April 19, 2019 <https://doi.org/10.1073/pnas.1820585116>

Edited by Angela M. Gronenborn, University of Pittsburgh School of Medicine, Pittsburgh, PA, and approved March 26, 2019 (received for review December 4, 2018)

В моем белке альфа-спираль. И что?

- Спираль - диполь. На N-конце частичный заряд до +0.9. Концы спиралей имеют большую плотность водородных связей. Какой в этом смысл? Конец спирали + примыкающая петля часто являются местом связывания анионных субстратов

nature

Article | Published: 08 June 1978

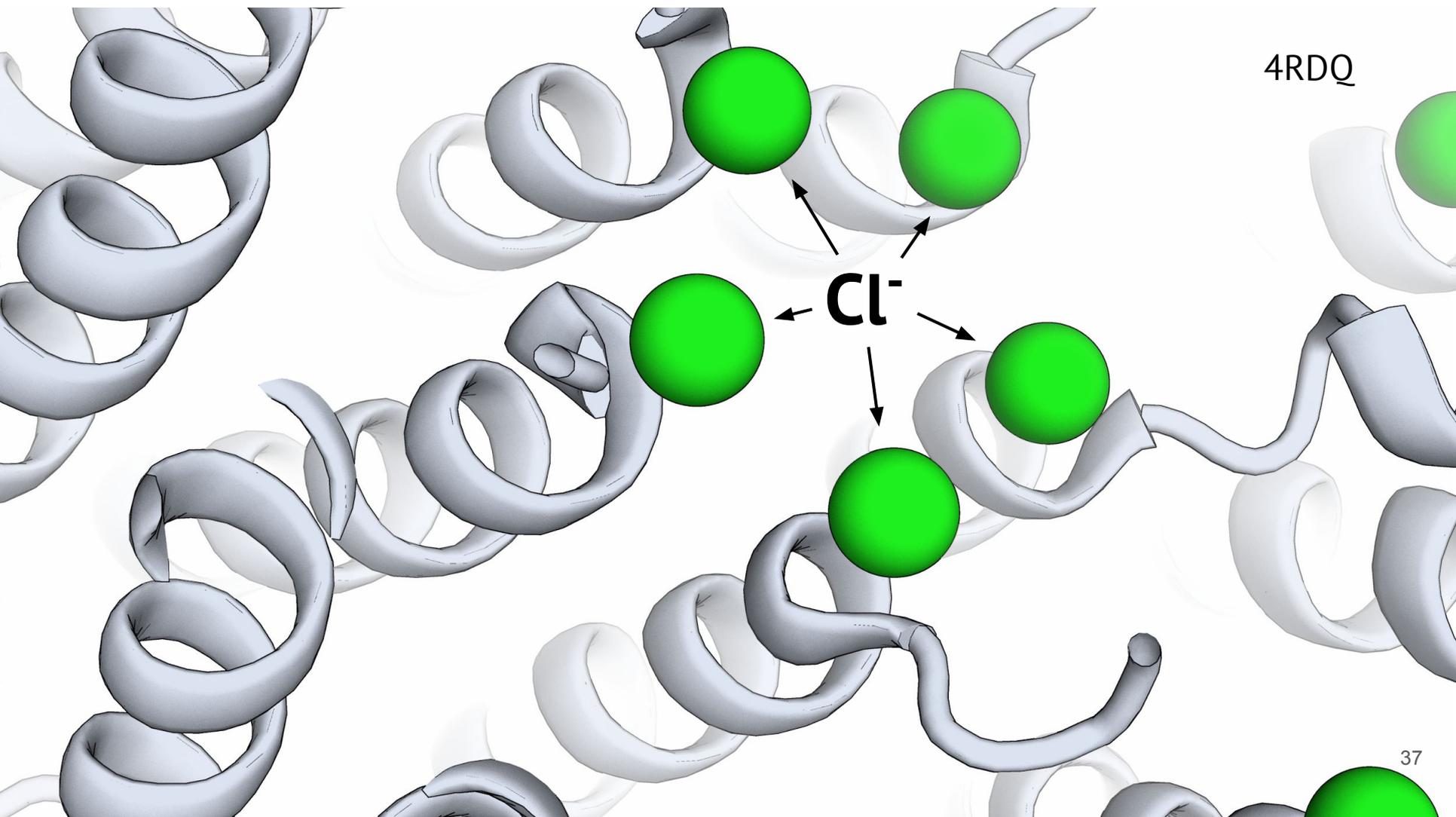
The α -helix dipole and the properties of proteins

W. G. J. Hol, P. T. van Duijnen & H. J. C. Berendsen

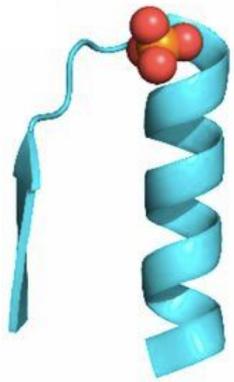
Nature 273, 443–446(1978) | [Cite this article](#)

- Дипольный момент спирали может сдвигать pK_a остатков. Играет роль в активации (депротонировании) цистеина в тиоредоксине, тирозин фосфатазах и не только

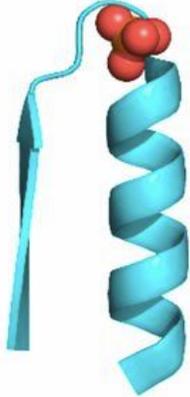
Какой конец альфа-спирали удерживает хлор?



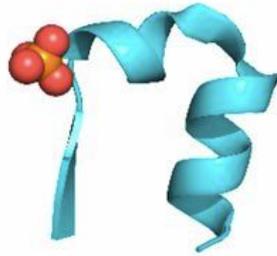
Альфа-спираль – старейший байндер фосфатов



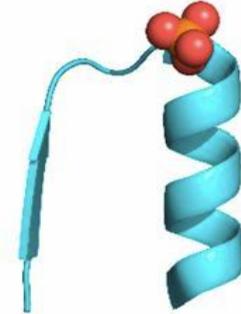
HUP domain



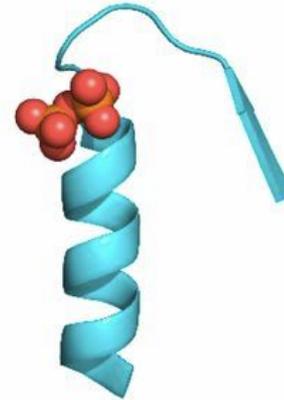
Flavodoxin



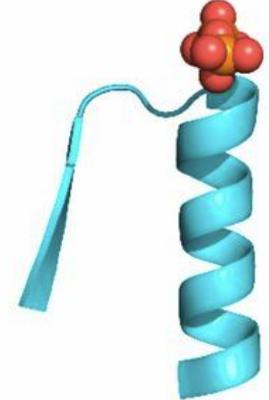
TIM β/α -barrel



Rossmann
with crossover



P-loop domains



Rossmann

Short and simple sequences favored the emergence of N-helix phospho-ligand binding sites in the first enzymes. Liam M. Longo, Dušan Petrović, Shina Caroline Lynn Kamerlin, and Dan S. Tawfik
PNAS, 2020

Simple yet functional phosphate-loop proteins. Maria Luisa Romero Romero, Fan Yang, Yu-Ru Lin, Agnes Toth-Petroczy, Igor N. Berezovsky, Alexander Goncarenco, Wen Yang, Alon Wellner, Fanindra Kumar-Deshmukh, Michal Sharon, David Baker, Gabriele Varani, and Dan S. Tawfik. *PNAS*, 2018

α -спираль прямая и жесткая



Но бывают и длиннее, например, из 98 остатков:

Wolny, M., Batchelor, M., Bartlett, G. et al. **Characterization of long and stable de novo single alpha-helix domains provides novel insight into their stability.** *Sci Rep*, 2017

Описаны единичные спирали вплоть до 200 остатков.

DSSP

Definition of the **S**econdary **S**tructure of **P**roteins

Kabsch,W., Sander,C.

Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, no.12 (1983)

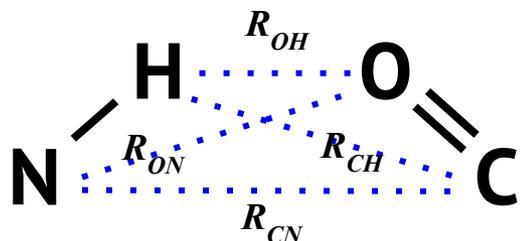
Вход алгоритма: PDB-файл

Выход алгоритма: текстовый файл с информацией о каждом остатке

Этапы алгоритма:

- Нахождение водородных связей между остовными N и O.
- Поиск паттернов водородных связей – “зародышей” элементов вторичной структуры
- Объединение паттернов в элементы вторичной структуры.

Алгоритм DSSP, шаг 1



Положение водородов на азотах остова восстанавливается однозначно

$$E = 0.084 \cdot \left(\frac{1}{R_{ON}} + \frac{1}{R_{CH}} - \frac{1}{R_{OH}} - \frac{1}{R_{CN}} \right) \cdot 322 \text{ kcal/mol}$$

$$q_1 \cdot q_2, \text{ где}$$

$$q_1 = -0.42$$

$$q_2 = 0.20$$

q_1 заряд на O
 q_2 заряд на H

“We choose a generous cutoff to allow for bifurcated H bonds and errors in coordinates”

То, что имеет $E < 0.5$ ккал/моль, то и есть водородная связь. При «идеальном» расположении атомов получается примерно -3 ккал/моль.

Алгоритм DSSP, элементарные паттерны

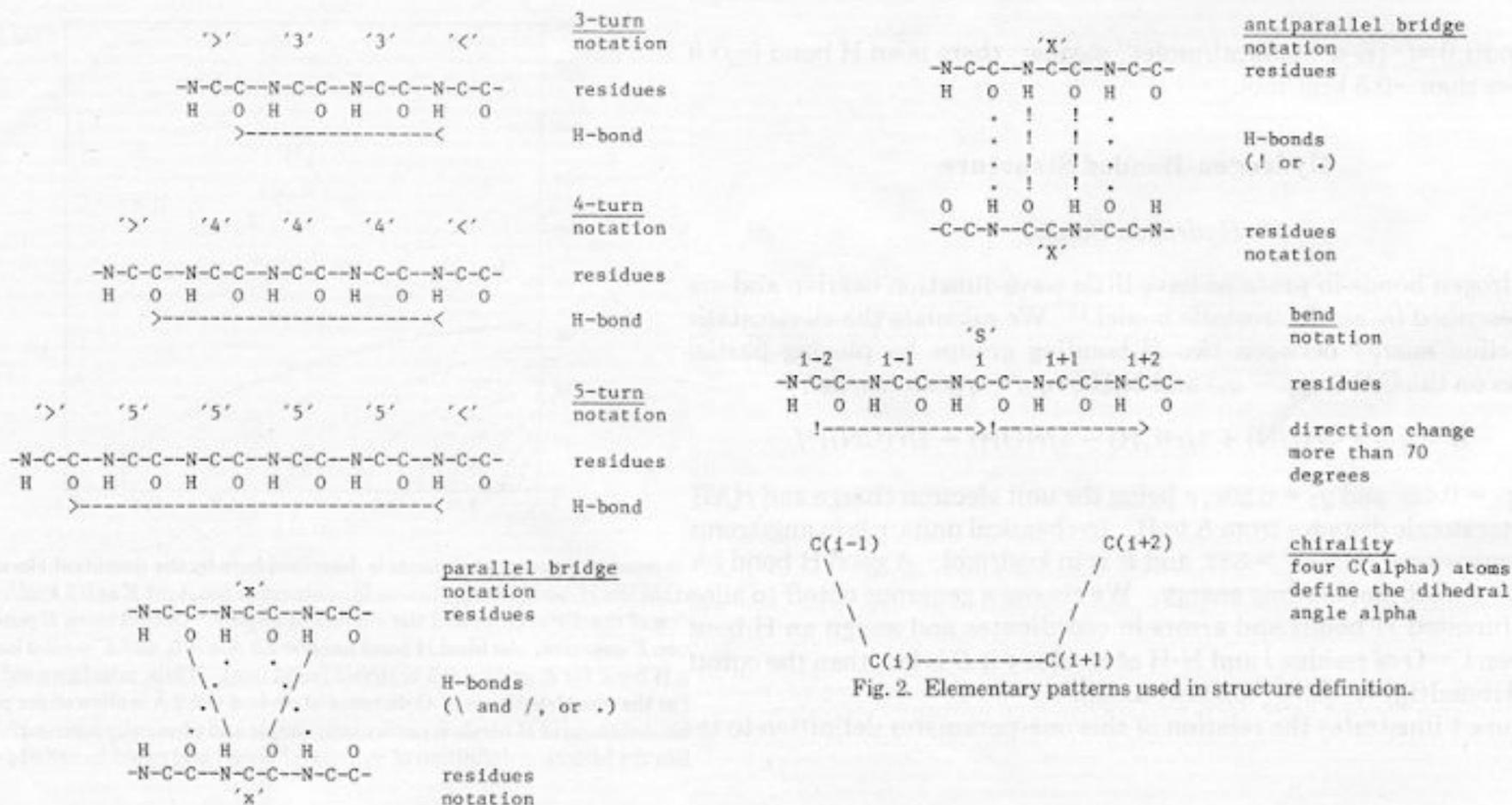
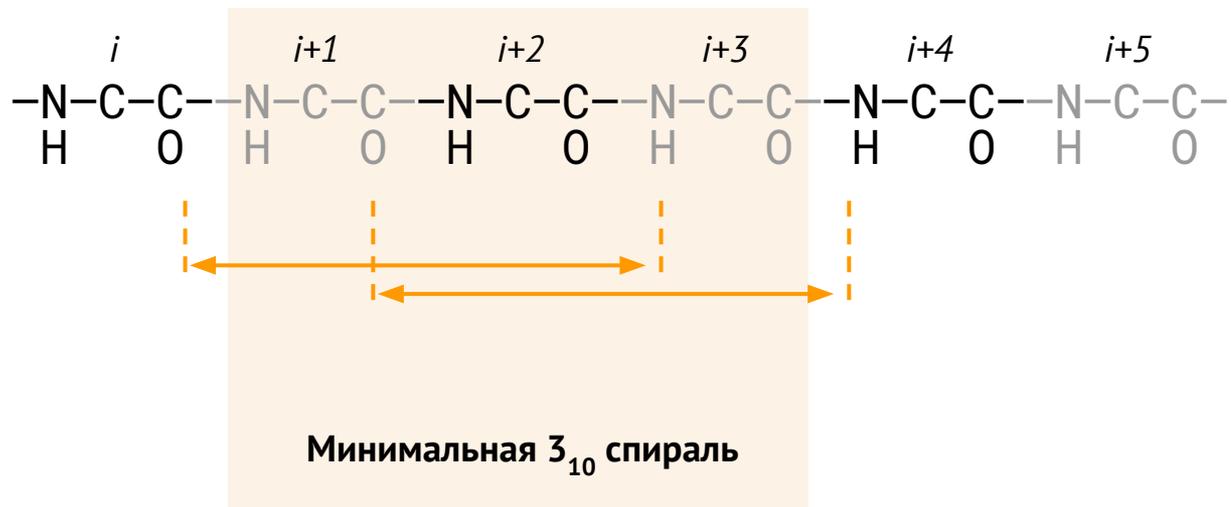
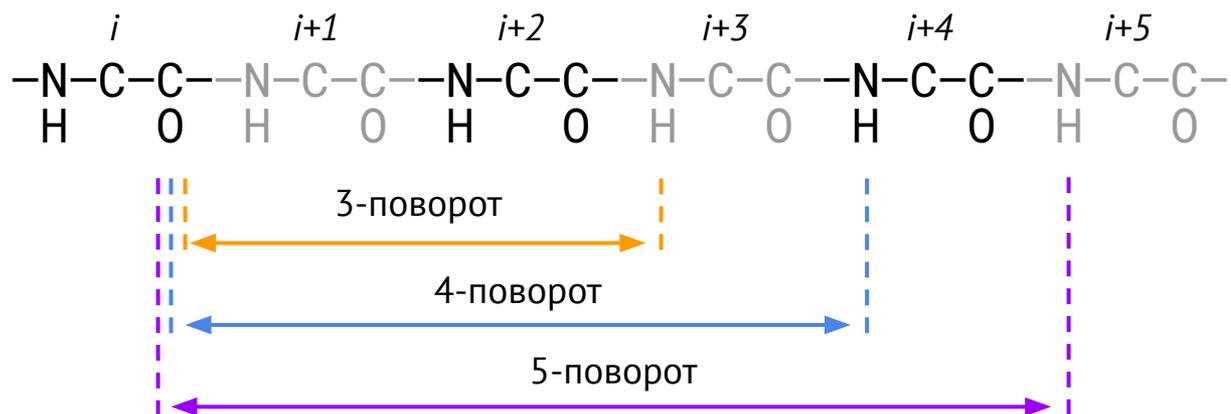


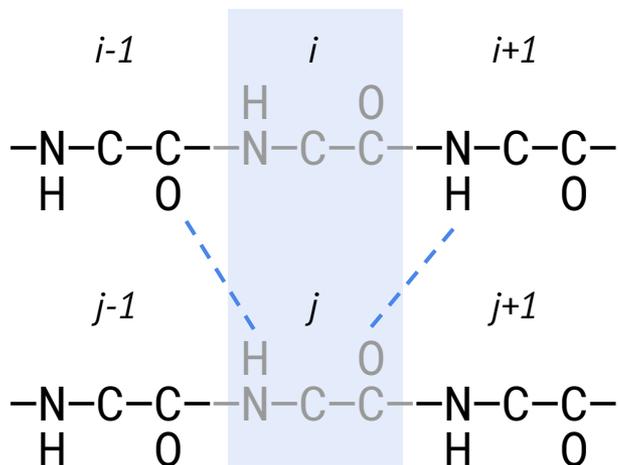
Fig. 2. Elementary patterns used in structure definition.

n-поворот и элементарная спираль

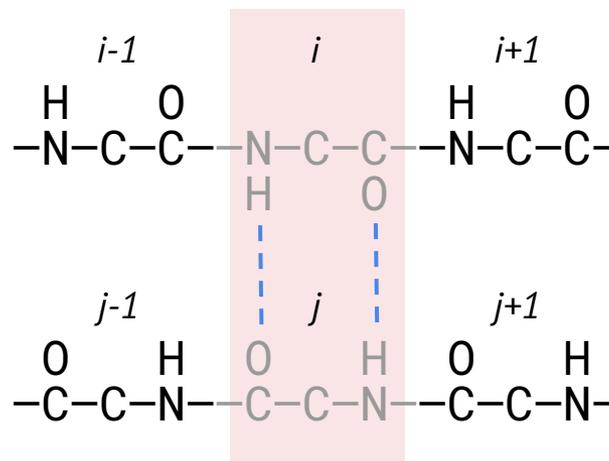


Терминальным остаткам метка вторичной структуры не присваивается

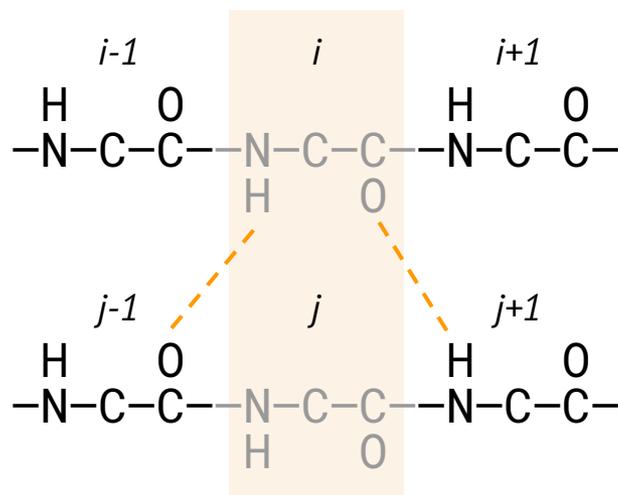
Мостики и лестницы



$i-1.O \cdots j.N$ и $j.O \cdots i+1.N$



$j.O \cdots i.N$ и $i.O \cdots j.N$



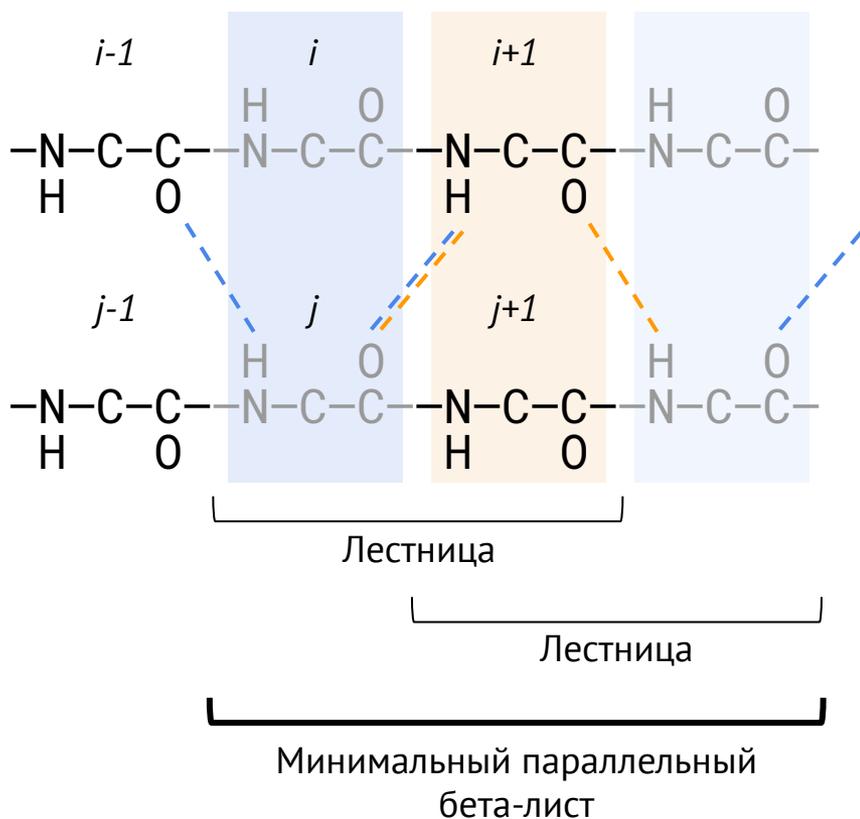
$j-1.O \cdots i.N$ и $i.O \cdots j+1.N$

Лестница = один или более мостиков **одного типа**, идущих подряд (по одному тяжу не более 1 остатка между лестницами, по другому тяжу не более 4)

Две лестницы связаны, если имеют общий остаток

Бета-лист: максимальный связный набор лестниц

Мостики и лестницы



Лестница = один или более мостиков **одного типа**, идущих подряд (по одному тяжу не более 1 остатка между лестницами, по другому не более 4)

Две лестницы связаны, если имеют общий остаток

Бета-лист: максимальный связный набор лестниц

Мостики и лестницы

Лестница = один или более мостиков **одного типа**, идущих подряд

Две лестницы связаны, если имеют общий остаток

Бета-лист: максимальный связный набор лестниц

Нерегулярности: допускаются выпячивания, по одному тяжу не более 1 остатка между лестницами, по другому тяжу не более 4. Это решает проблему с разметкой в качестве бета-листа двух параллельных бета-тяжей.

Дополнительные разметки

Излом (bend): изменение направления более чем на 70°

Хиральность. Остатку i присписывается знак (“+” или “-”) торсионного угла, образованного C α -атомами $i-1$, i , $i+1$, $i+2$

Обозначения

H = alpha helix

B = residue in isolated beta-bridge

E = extended strand, participates in beta ladder

G = 3-helix (3/10 helix)

I = 5-helix (pi helix)

T = hydrogen bonded turn

S = bend

Участки 3-, 4- или 5-поворотов, не составляющие спираль достаточной длины, обозначаются T

Приоритет при детектировании нескольких ситуаций у одного остатка: сверху вниз (от H к S)

Если не детектировано ничего из перечисленного, в соответствующей колонке ставится пробел (хотя некоторые программы обработки заменяют его на букву C, от слова “coil”).

В исходном варианте DSSP массово недоразмечал пи-спирали. Ревизии программы перестроили приоритет, присвоив наибольший как раз пи-спиралям.

DSSP: вариации на тему

SABLE (2016): контроль над бифуркациями, возможность донора никому не донировать

DSSPcont (2003): вместо однозначного присвоения типа – сопоставление остатку правдоподобия быть в составе этого типа. Интегрирование результатов по 10 запускам DSSP с разными порогами на образование водородной связи, от -1.0 ккал/моль до -0.2 ккал/моль

STRIDE

Frishman,D., Argos,P. **Knowledge-based protein secondary structure assignment.**
Proteins 23 , no, 4 (1995)

Этапы алгоритма:

- Расчет энергии водородной связи для пар остовных атомов N и O.
Способ расчета отличен от используемого в DSSP
- Для каждого остатка: расчет вероятности того, что остаток с данными значениями торсионных углов ϕ и ψ принадлежит альфа-спирали (бета-тяжу)
- Детекция «зародышей» элементов вторичной структуры
- Объединение зародышей в элементы

Алгоритм основан на подходе “knowledge-based”, то есть использовании статистики экспертных аннотаций.

STRIDE и выход из плоскости

$$E_{hb} = E_r \cdot E_t \cdot E_p$$

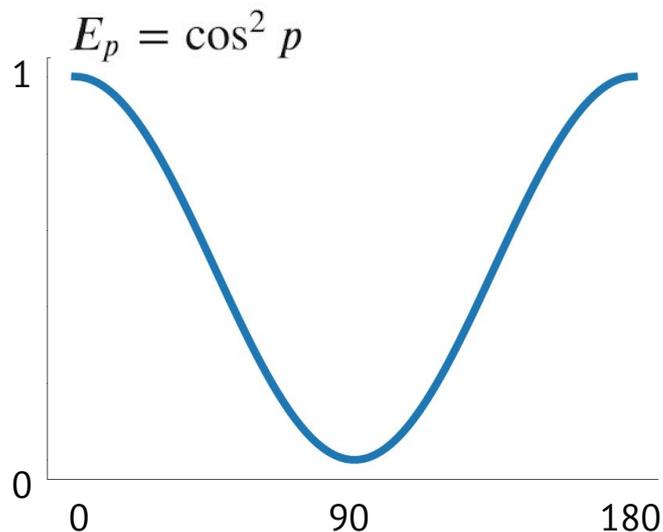
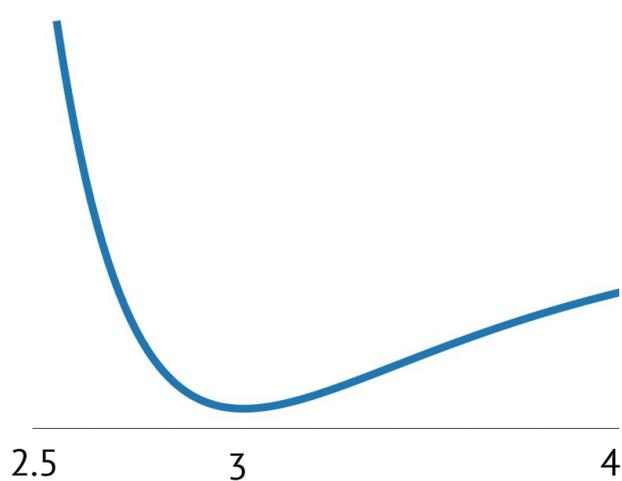
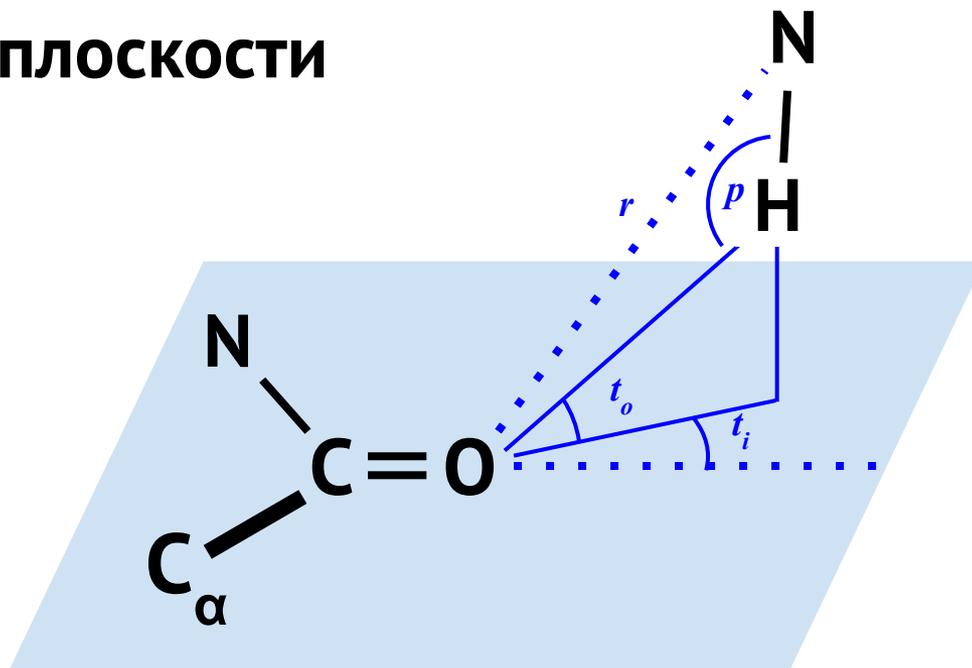
$$E_r = \frac{C}{r^8} - \frac{D}{r^6}$$

$$C = -3E_m r_m^8 \quad \text{kcal} \cdot \text{\AA}^8 / \text{mol}$$

$$D = -4E_m r_m^6 \quad \text{kcal} \cdot \text{\AA}^6 / \text{mol}$$

$$E_m = -2.8 \quad \text{kcal/mol}$$

$$r_m = 3.0 \quad \text{\AA}$$

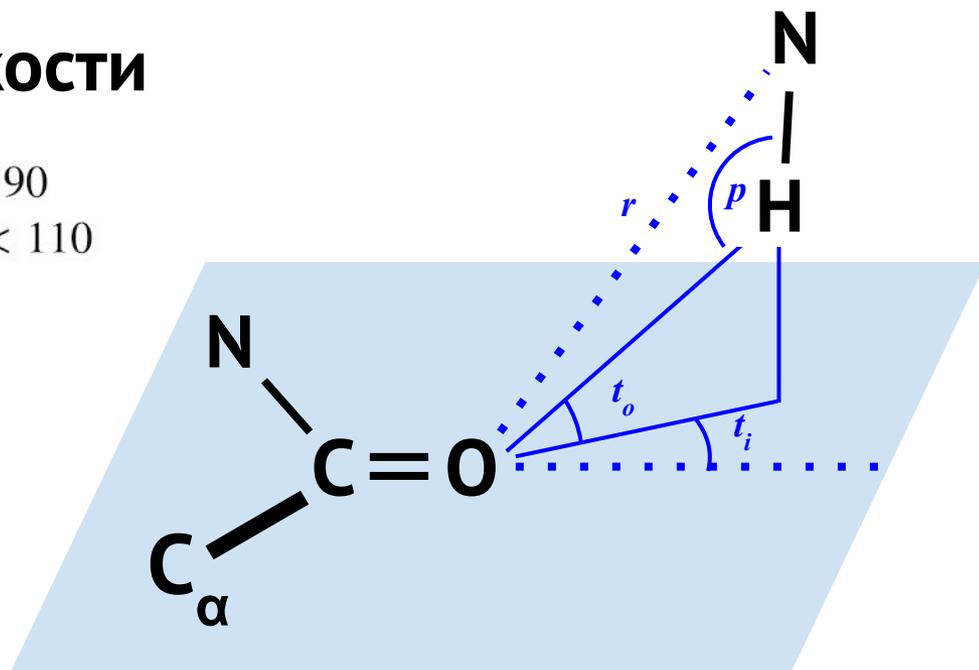
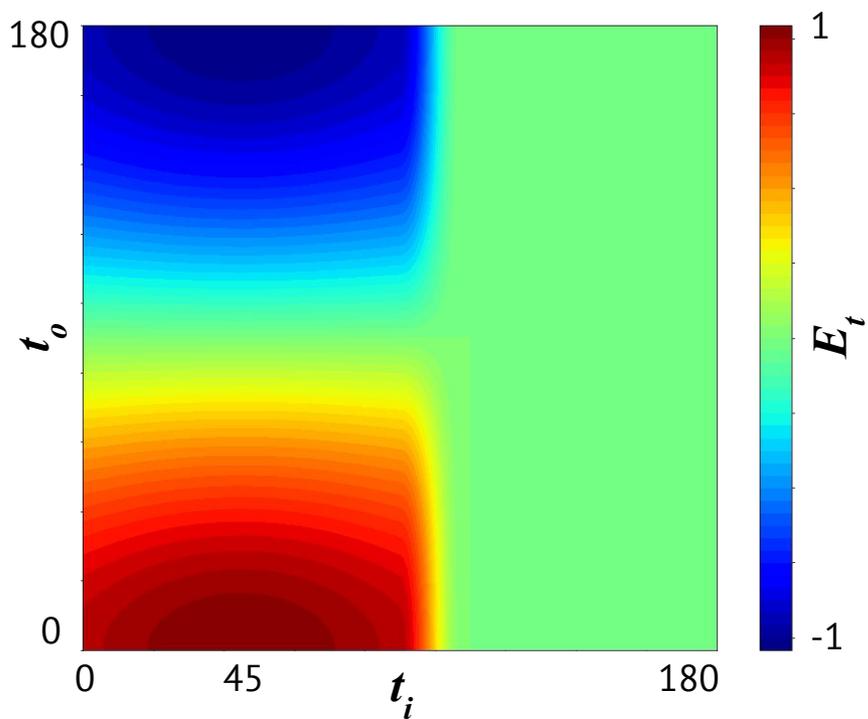


STRIDE И ВЫХОД ИЗ ПЛОСКОСТИ

$$E_t = \begin{cases} (0.9 + 0.1 \sin 2t_i) \cos t_o, & 0 < t_i < 90 \\ K_1(K_2 - \cos^2 t_i)^3 \cos t_o, & 90 < t_i < 110 \\ 0, & t_i > 110 \end{cases}$$

$$K_1 = 0.9 / \cos^6(110^\circ)$$

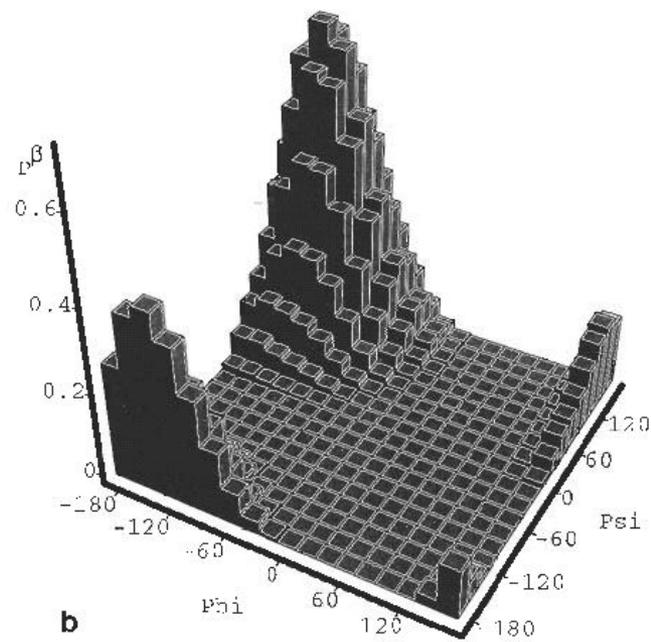
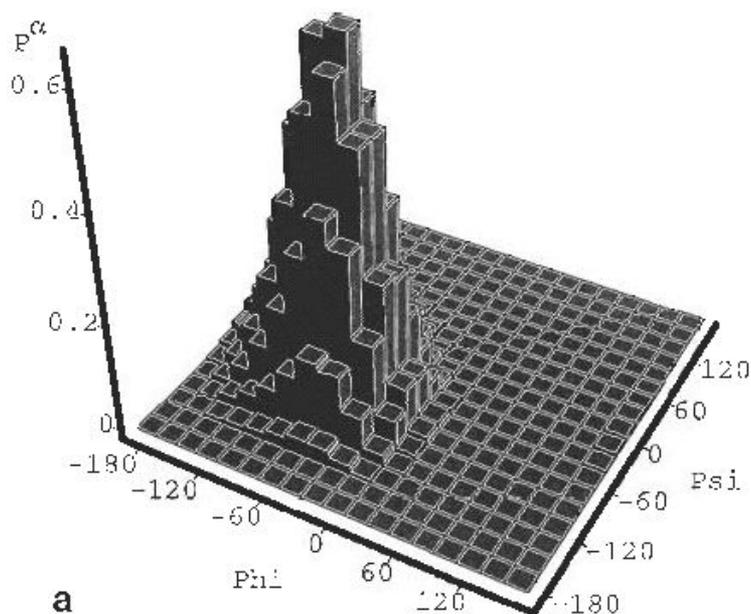
$$K_2 = \cos^2(110^\circ)$$



Собрана статистика по специально составленной выборке из 226 структур (“материал обучения” для уточнения параметров алгоритма). Принадлежность остатка альфа-спирали (бета-тяжу) определялась по аннотации PDB файла.

Карта Рамачандрана разбита на квадратики $20^\circ \times 20^\circ$, построена гистограмма (отдельно для бета- и альфа-).

Удалены точки из областей, заведомо запрещенных для альфа-спиралей (бета-тяжей) – это вероятные ошибки и артефакты.



Выделение зародышей

Критерий для пары $(i, i+4)$, чтобы считаться зародышем альфа-спирали:

$$E_{\text{hb}}^{i, i+4} \cdot \left(1 + W_1^\alpha + W_2^\alpha \cdot \frac{P_i^\alpha + P_{i+4}^\alpha}{2} \right) < T_1^\alpha$$

Минимальная альфа-спираль – два идущих подряд зародыша $(i, i+4)$, $(i+1, i+5)$; спиралью всегда считаются остатки $i+1, i+2, i+3, i+4$; а остатки i и $i+5$ – при выполнении условий $P_i^\alpha > T_2^\alpha$ и $P_{i+5}^\alpha > T_3^\alpha$, соответственно.

Константы $W_1^\alpha, W_2^\alpha, T_1^\alpha, T_2^\alpha, T_3^\alpha$ подобраны так, чтобы на материале обучения получить наилучшее совпадение с аннотированными в PDB альфа-спиралями.

Для бета-листов процедура аналогична.

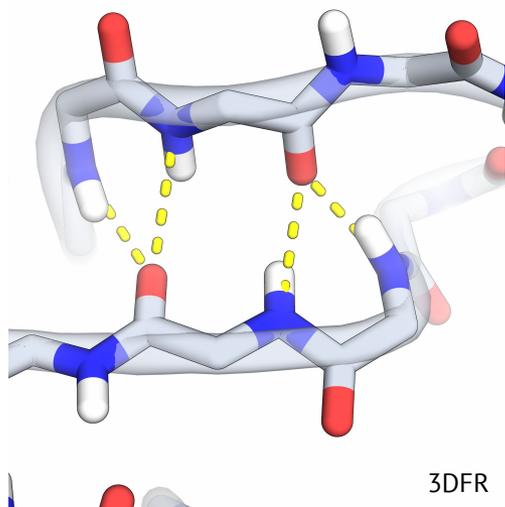
Другие

- DEFINE (1988)
- P-CURVE (1989)
- P-SEA (1997)
- XTLSSTR (1999)
- SECSTR (2002)
- VoTAP (2004)
- KAKSI (2005)
- PROSIGN (2008)
- SABA (2011)
- DISICL (2014)
- PCASSO (2014)
- SACF (2016)
- SABLE(2016)

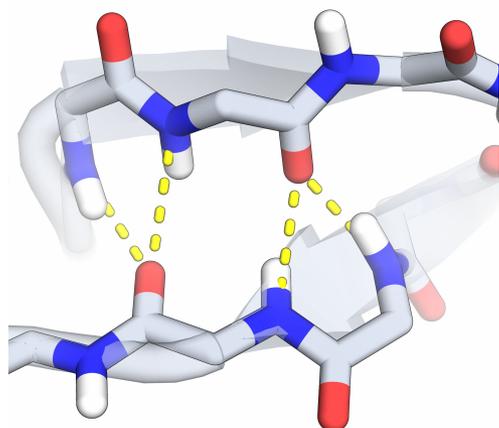
Многие из них используют только информацию о положении C α -атомов.
Наиболее популярным и де-факто стандартом по-прежнему остаётся DSSP

Алгоритм vs люди

PDB



DSSP



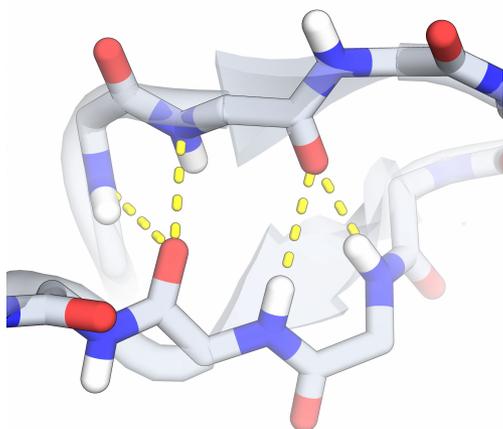
Вторичная структура, отображаемая PyMol по умолчанию, основывается на разметке, представленной в PDB файле авторами структуры:

```
HELIX 1 HB LEU A 23 THR A 34 1 12
HELIX 2 HC GLY A 42 PHE A 49 1 8
HELIX 3 HE ASP A 78 HIS A 89 1 12
HELIX 4 HF GLY A 99 LYS A 107 1 9
SHEET 1 S1 8 VAL A 74 VAL A 76 0
SHEET 2 S1 8 GLU A 56 THR A 63 1 0 ASN A 59 N VAL A 74
SHEET 3 S1 8 GLY A 36 GLY A 42 1 0 LYS A 37 N THR A 58
SHEET 4 S1 8 GLU A 93 GLY A 98 1 0 GLU A 93 N ILE A 38
SHEET 5 S1 8 THR A 1 GLN A 7 1 N ALA A 2 0 LEU A 94
SHEET 6 S1 8 ASP A 111 ALA A 119 1 0 THR A 112 N PHE A 3
SHEET 7 S1 8 THR A 152 LYS A 161 -1 N LYS A 160 0 ASP A 111
SHEET 8 S1 8 ASP A 135 VAL A 144 -1 N THR A 137 0 GLN A 159
```

Разными программами и разными людьми эта разметка может быть сделана по-разному. Люди менее восприимчивы к небольшим отклонениям в координатах.

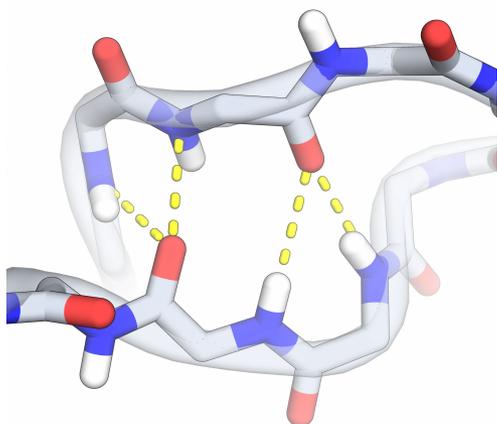
Алгоритм vs люди

PDB



2L28

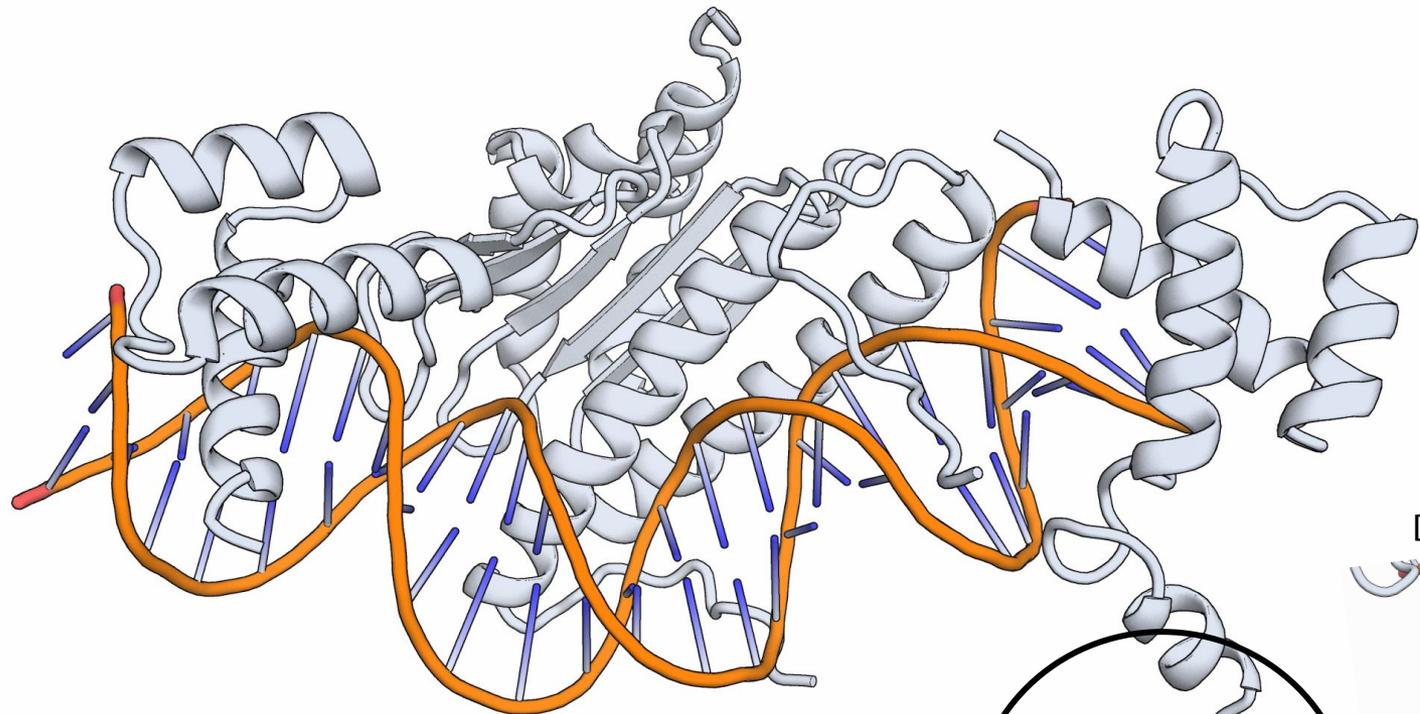
DSSP



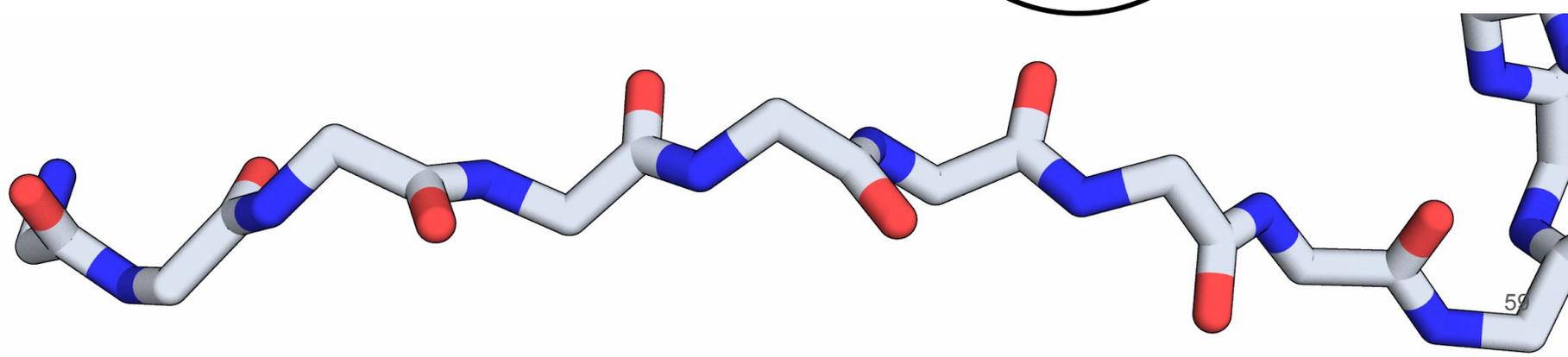
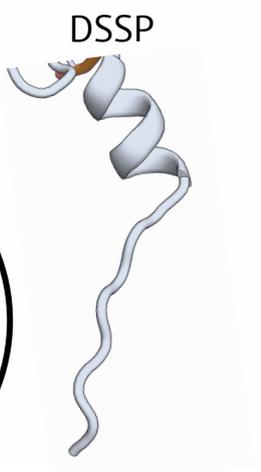
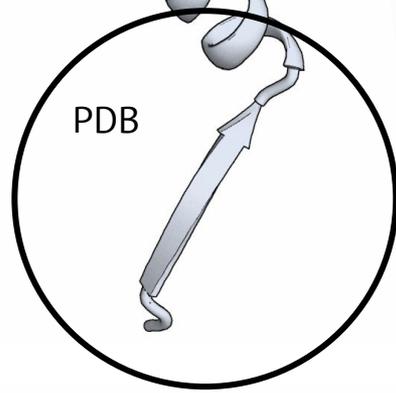
Вторичная структура, отображаемая PyMol по умолчанию, основывается на разметке, представленной в PDB файле авторами структуры:

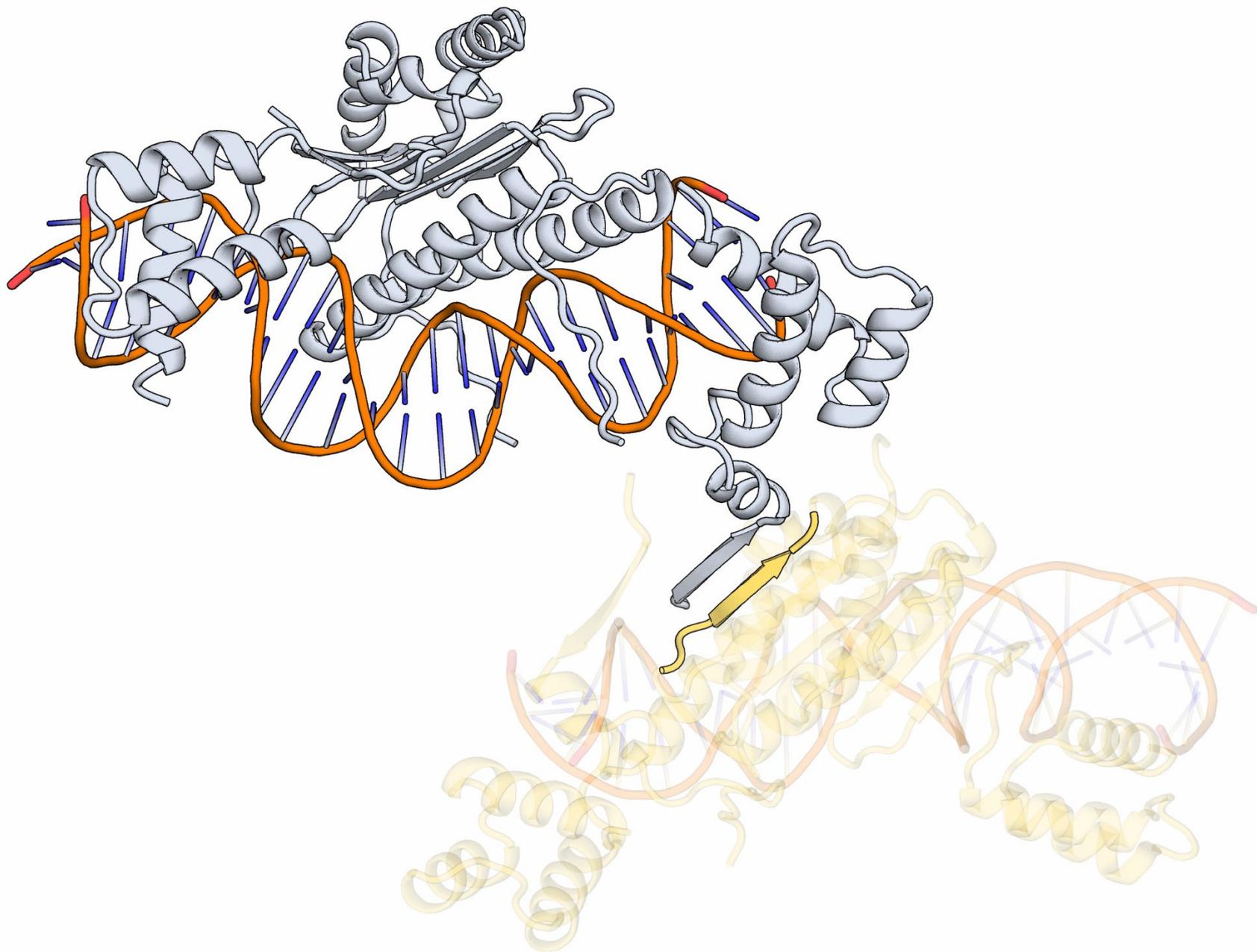
```
HELIX 1 HB LEU A 23 THR A 34 1 12
HELIX 2 HC GLY A 42 PHE A 49 1 8
HELIX 3 HE ASP A 78 HIS A 89 1 12
HELIX 4 HF GLY A 99 LYS A 107 1 9
SHEET 1 S1 8 VAL A 74 VAL A 76 0
SHEET 2 S1 8 GLU A 56 THR A 63 1 0 ASN A 59 N VAL A 74
SHEET 3 S1 8 GLY A 36 GLY A 42 1 0 LYS A 37 N THR A 58
SHEET 4 S1 8 GLU A 93 GLY A 98 1 0 GLU A 93 N ILE A 38
SHEET 5 S1 8 THR A 1 GLN A 7 1 N ALA A 2 0 LEU A 94
SHEET 6 S1 8 ASP A 111 ALA A 119 1 0 THR A 112 N PHE A 3
SHEET 7 S1 8 THR A 152 LYS A 161 -1 N LYS A 160 0 ASP A 111
SHEET 8 S1 8 ASP A 135 VAL A 144 -1 N THR A 137 0 GLN A 159
```

Разными программами и разными людьми эта разметка может быть сделана по-разному. Люди менее восприимчивы к небольшим отклонениям в координатах.

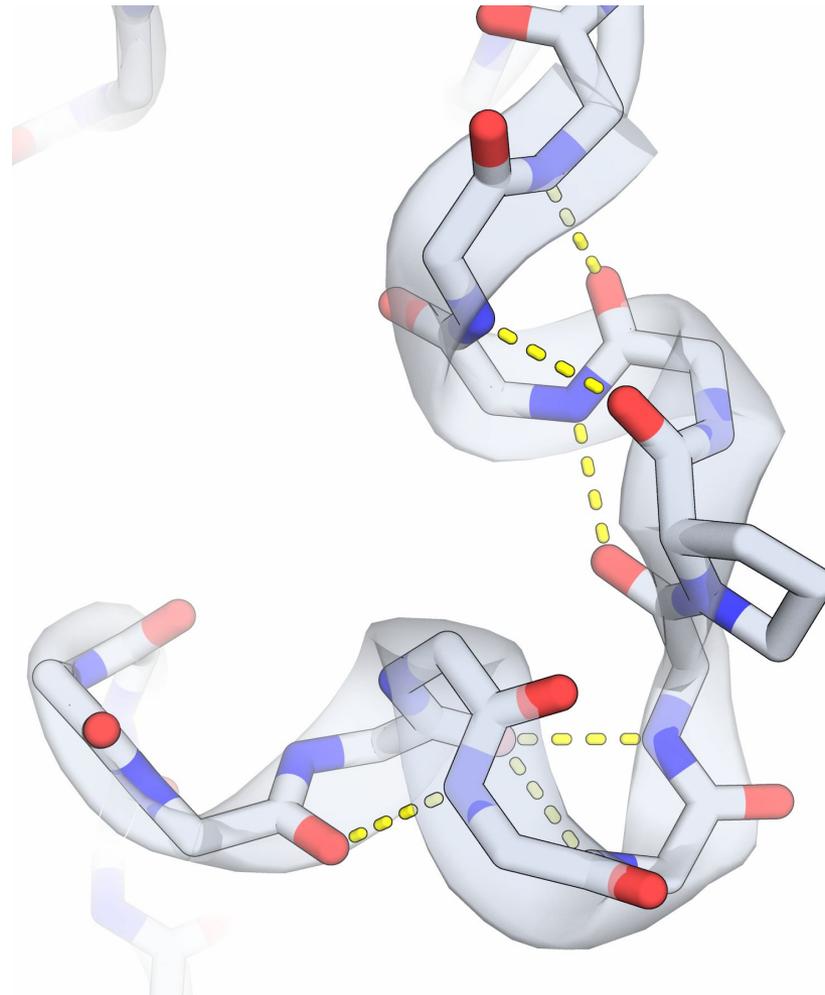
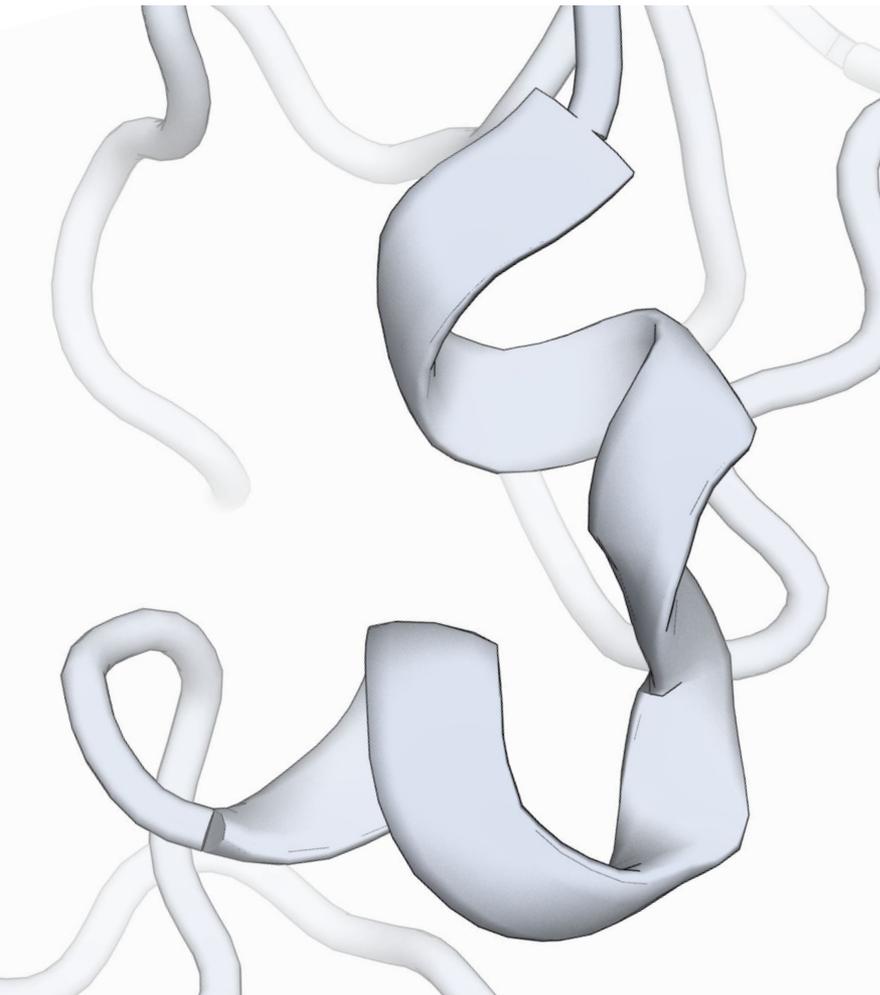


PDB 1MNM

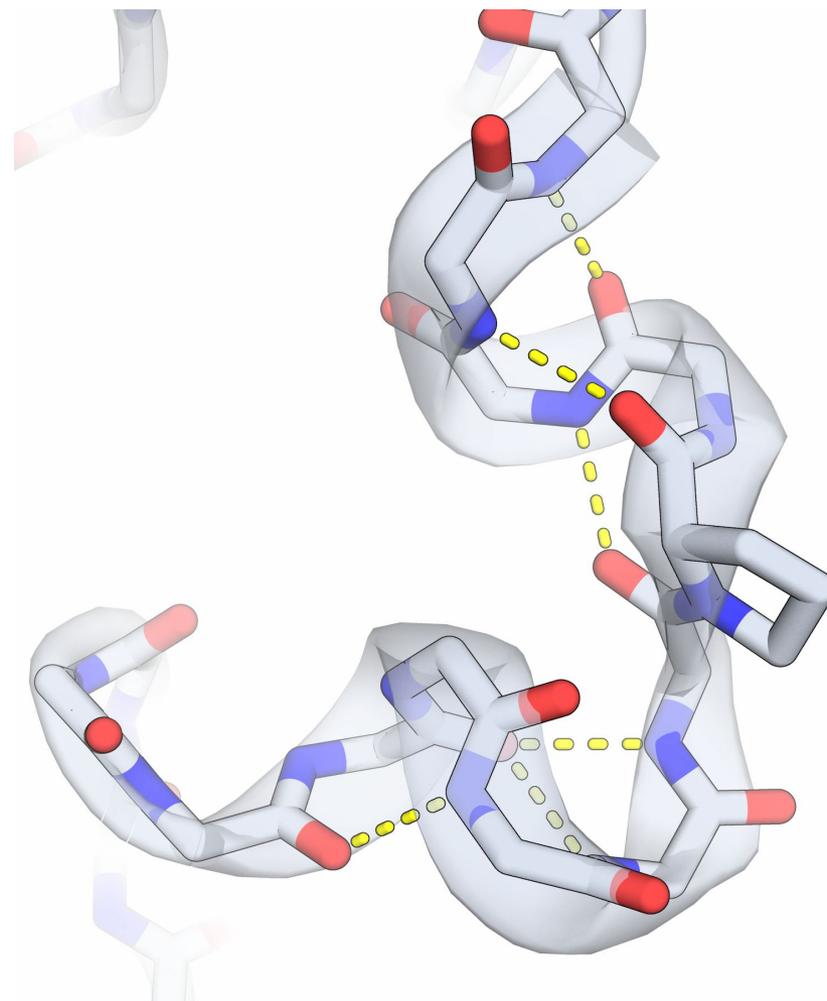
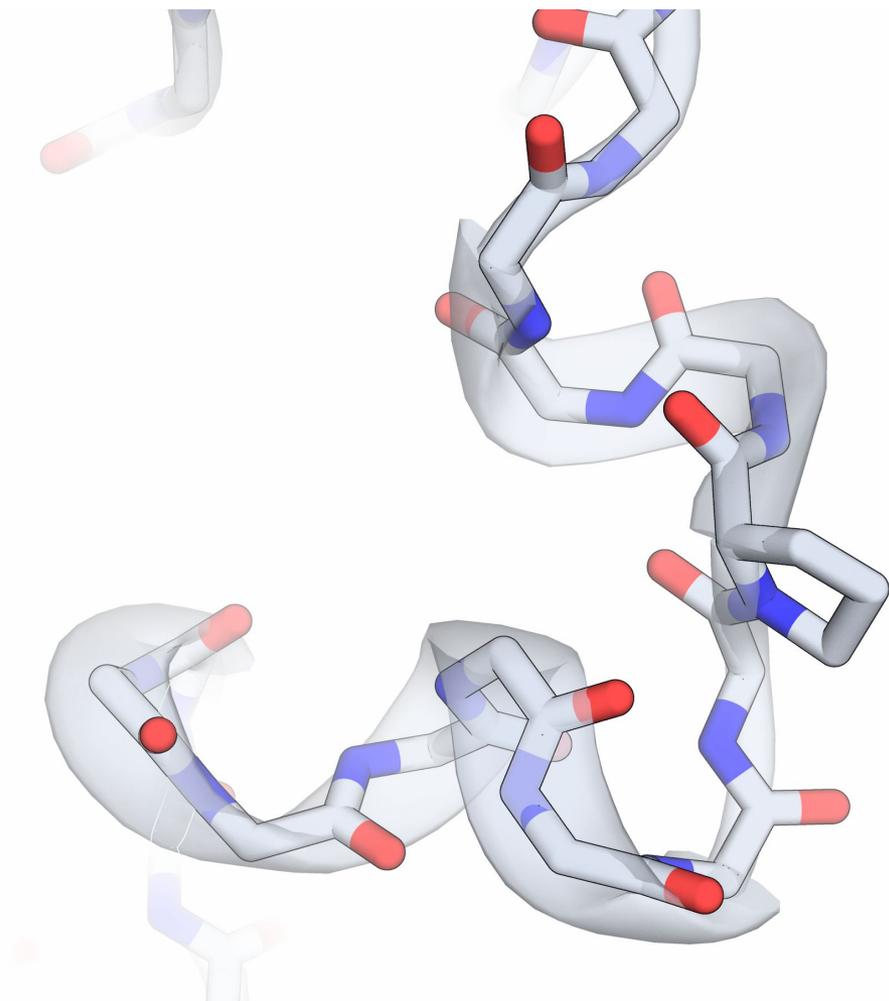


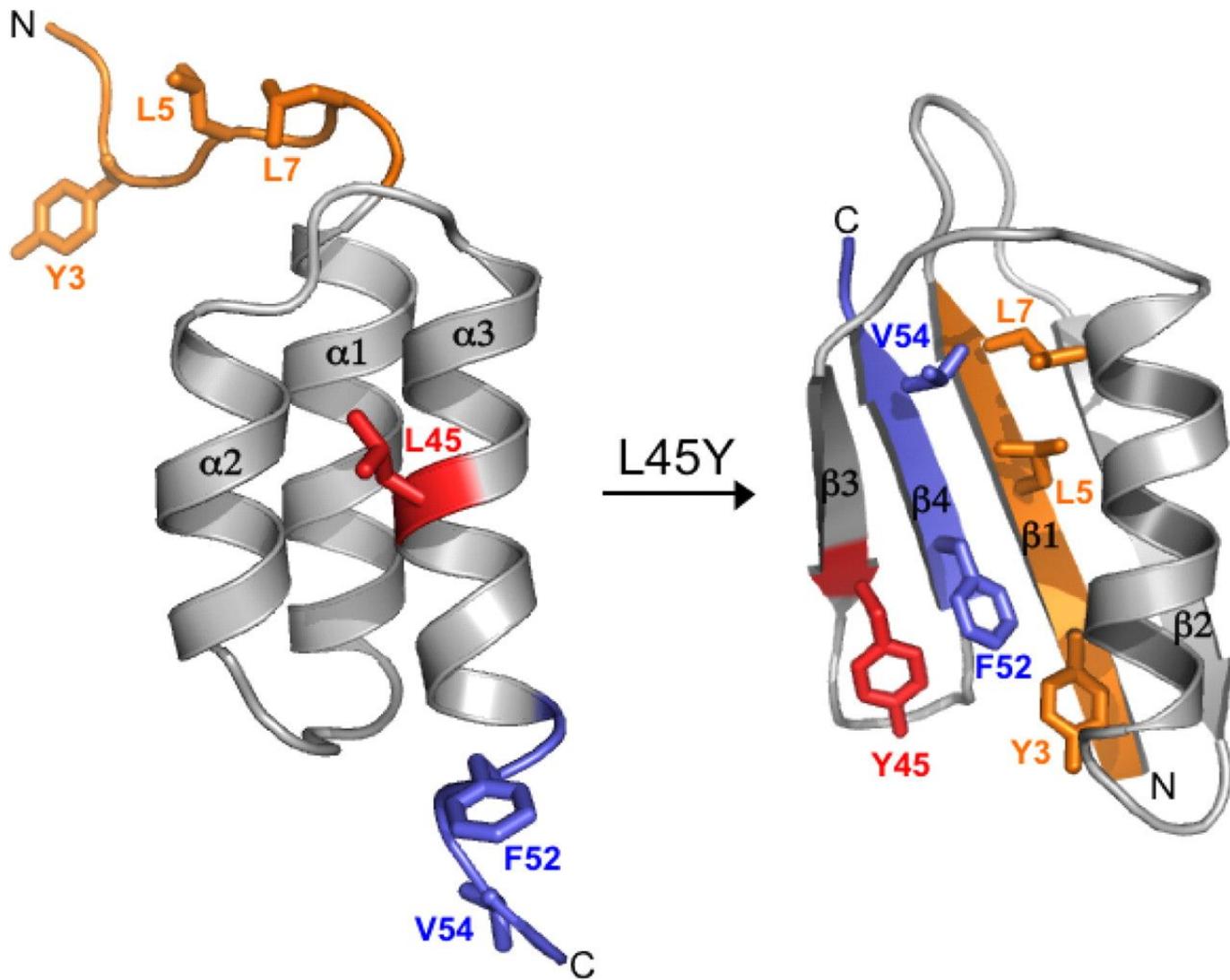


3K34



3K34



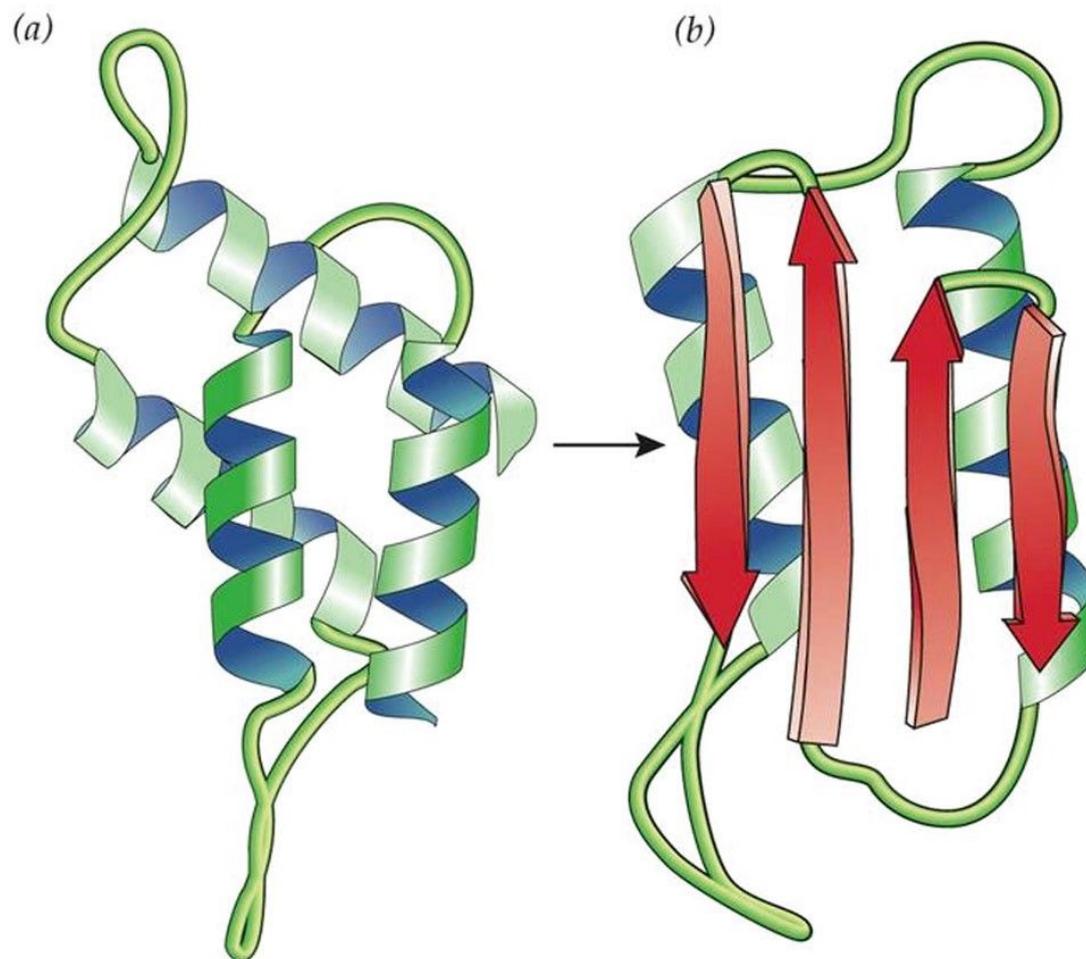


A minimal sequence code for switching protein structure and function

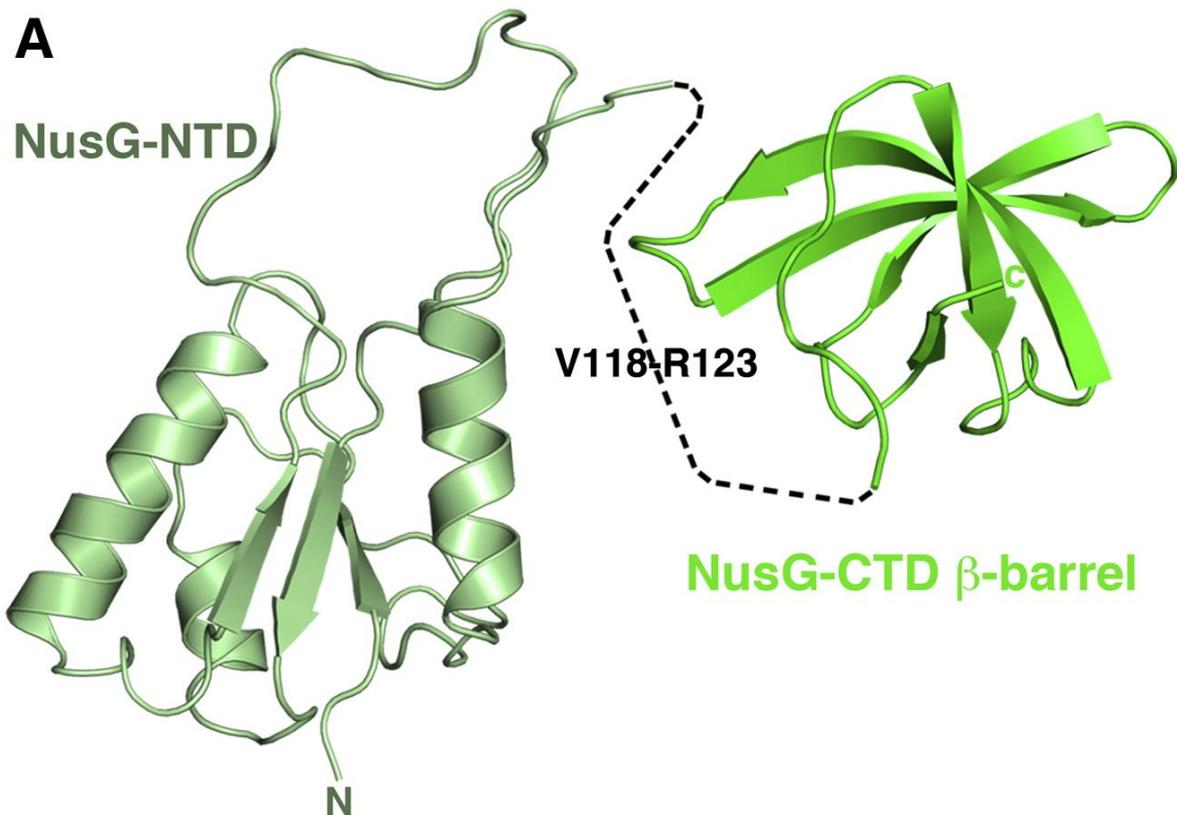
Patrick A. Alexander, Yanan He, Yihong Chen, John Orban, and Philip N. Bryan

PNAS December 15, 2009

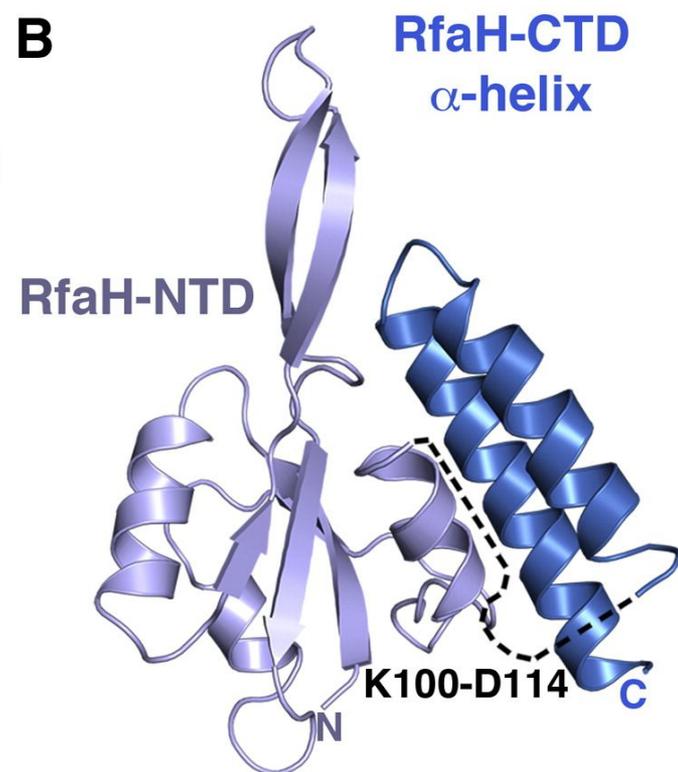
Прионы



Метаморфы



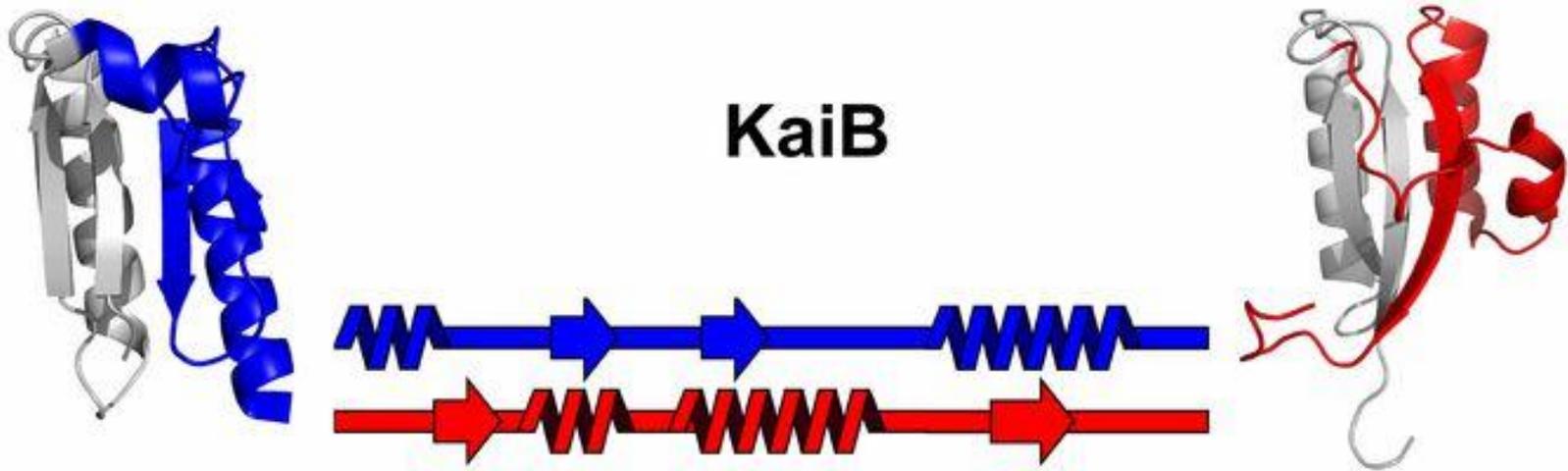
NusG is in an open form in solution and has an all β -fold in its CTD



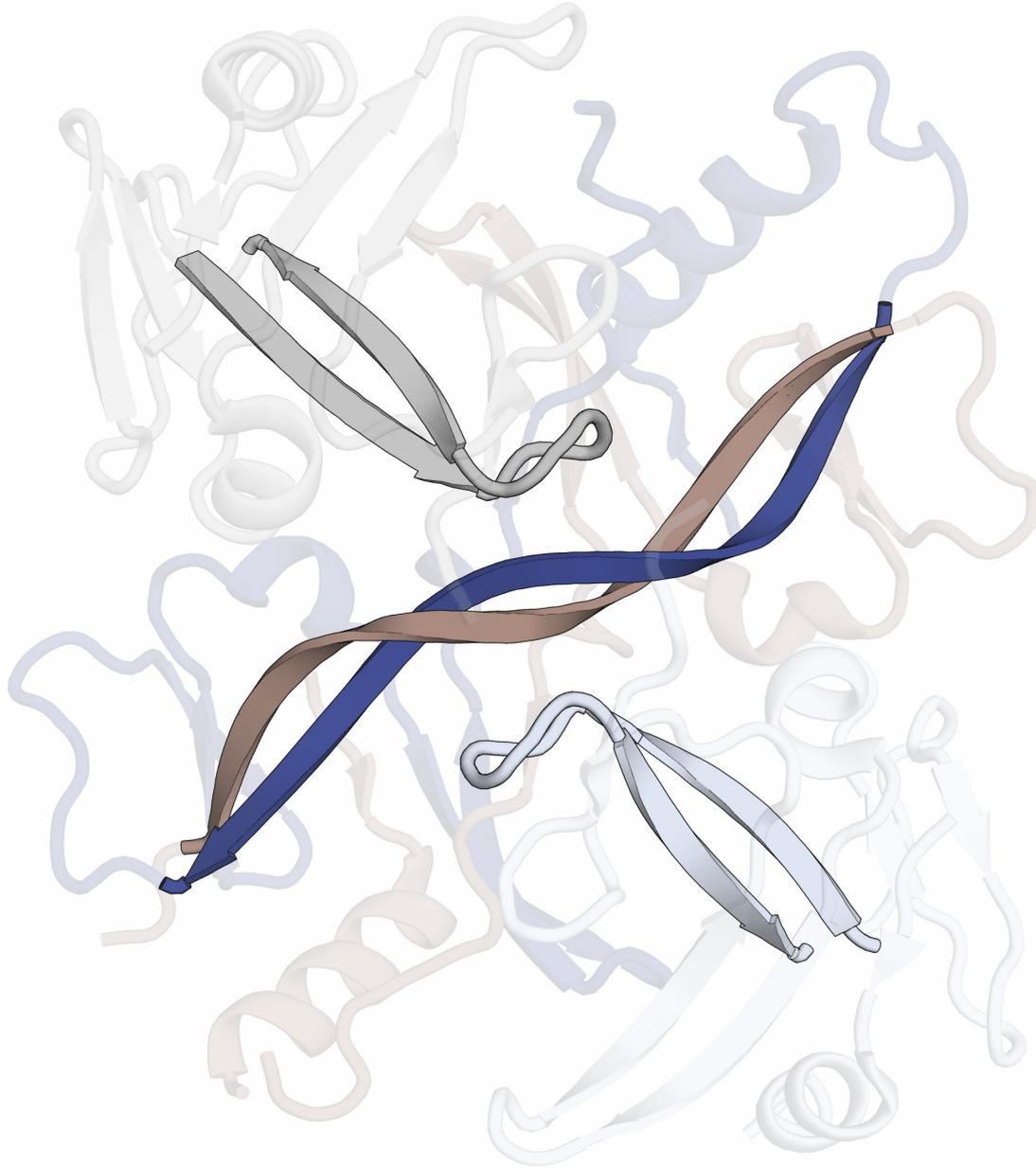
RfaH is in a closed state in the protein crystal and the CTD is in an all α -helical fold

Метаморфы

A HETERO-OLIGOMERS



Динамика спонтанного перехода между формами KaiB задает циркадные ритмы в цианобактериях



Надвторичная структура

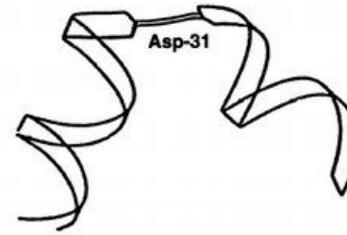
- Комбинации последовательно идущих элементов вторичной структуры, пребывающие в определенном взаимном расположении (т. е. выделение чисто структурное, без функционального или эволюционного подтекста)
- Промежуточный уровень абстракции между вторичной и третичной структурой
- Строительные блоки и затравки фолдинга. Иногда используется синоним “единица фолдинга” (folding unit)

Выделение структурного паттерна как элемента надвторичной структуры не имеет определенных критериев и довольно произвольно. Зачастую большие элементы надвторичной структуры могут также считаться типами третичной структуры (укладки).

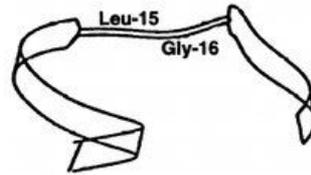
(a) H-b-H (2cyp)



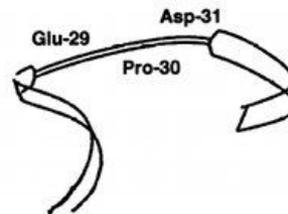
(b) H-t-H (1eca)



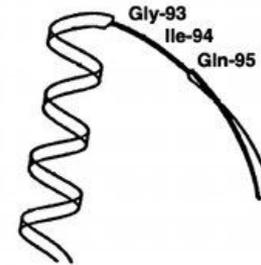
(c) H-bb-H (2utg)



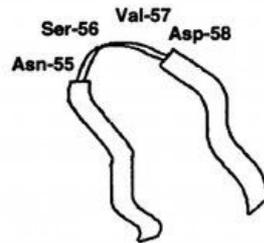
(d) H-lbb-H (2utg)



(e) H-lba-E (4pfk)



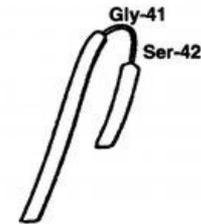
(f) E-aaal-E (2cna)



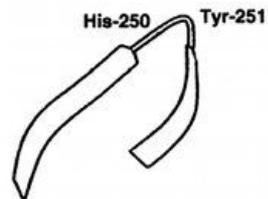
(g) E-aa-E (3enl)



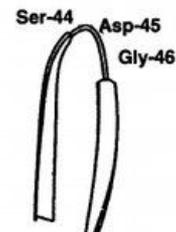
(h) E-ea-E (1cdl)



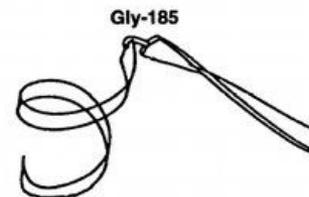
(i) E-ll-E (6tmn)

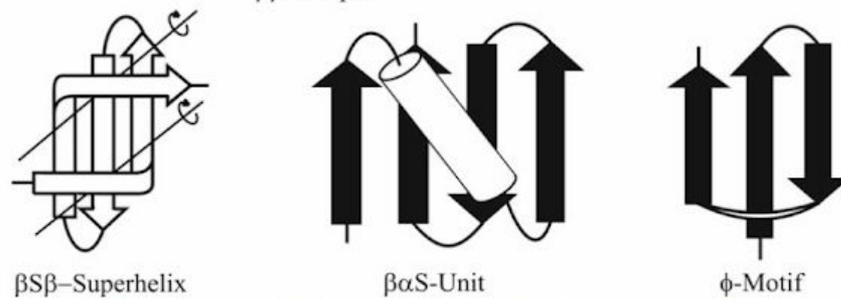
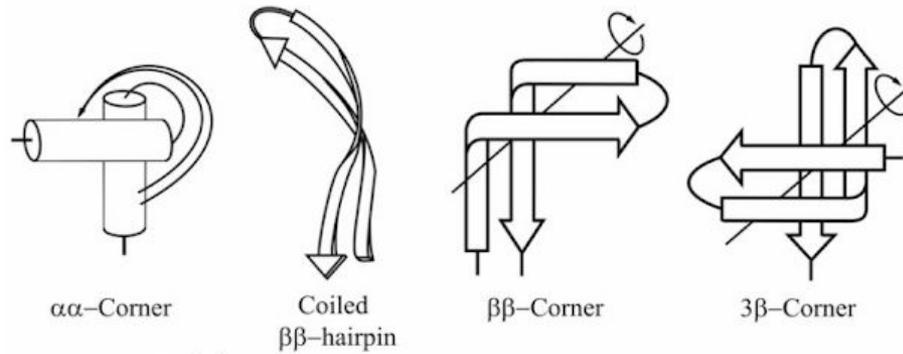
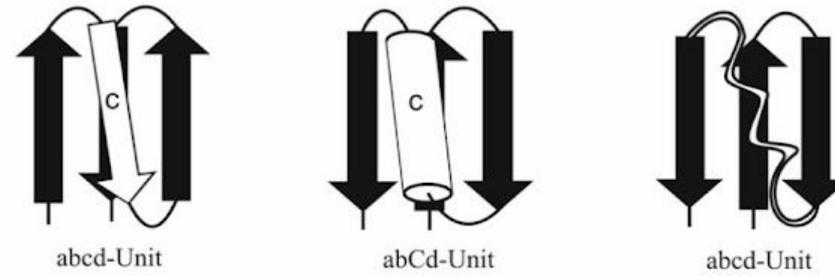
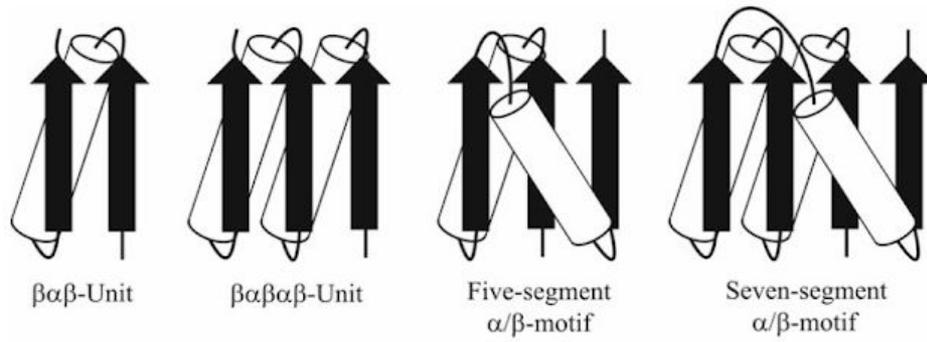


(j) E-aal-E (1il8)



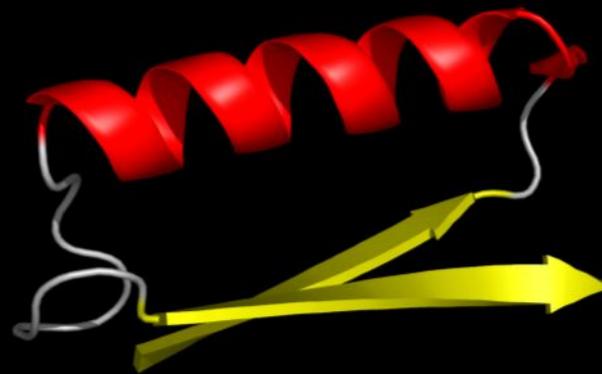
(k) H-l-E (2tsl)







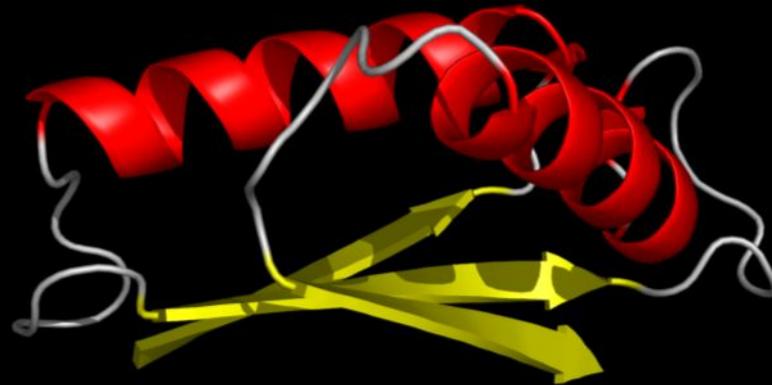
β-шпилька



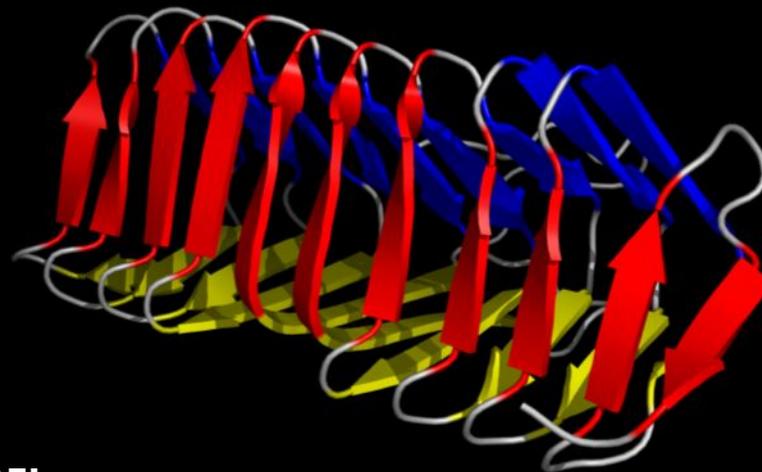
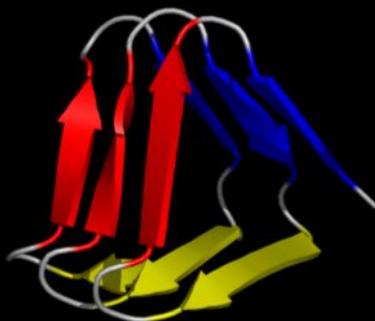
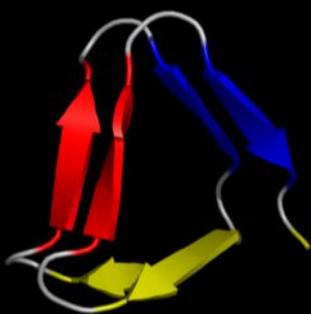
βαβ-спираль



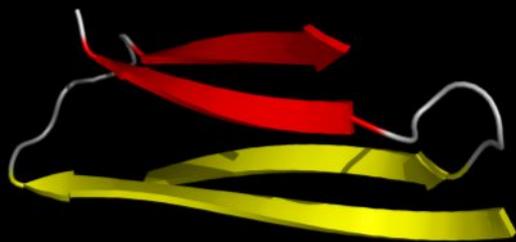
β-меандр



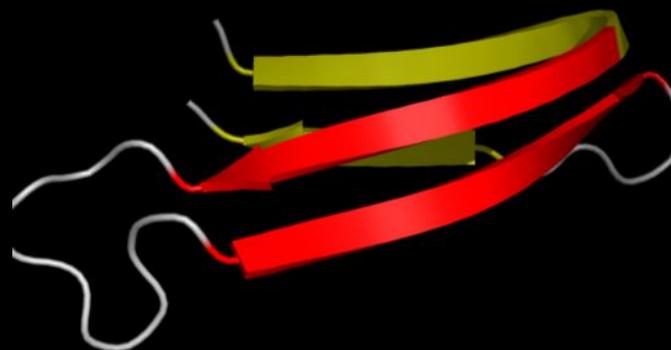
βαβ-мотив



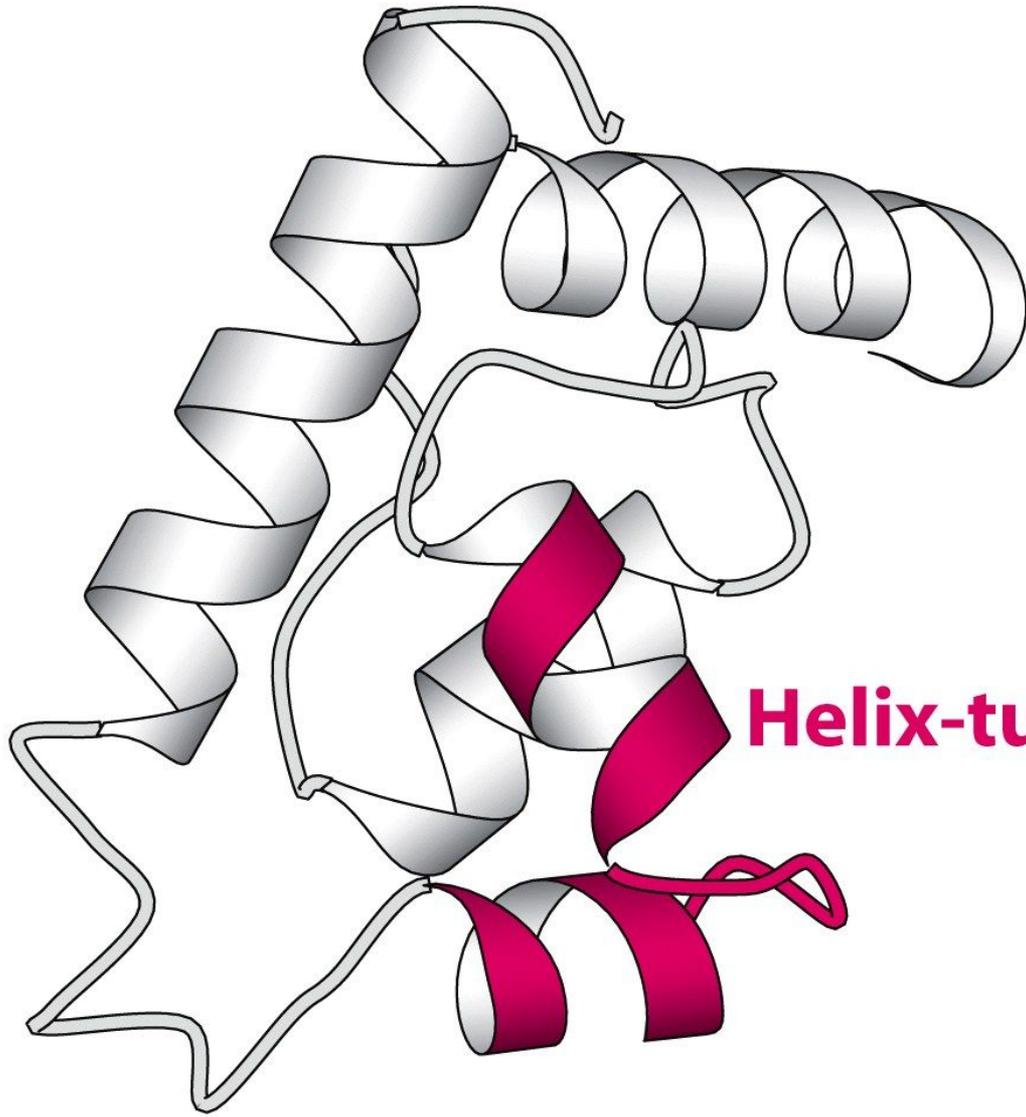
β-спираль



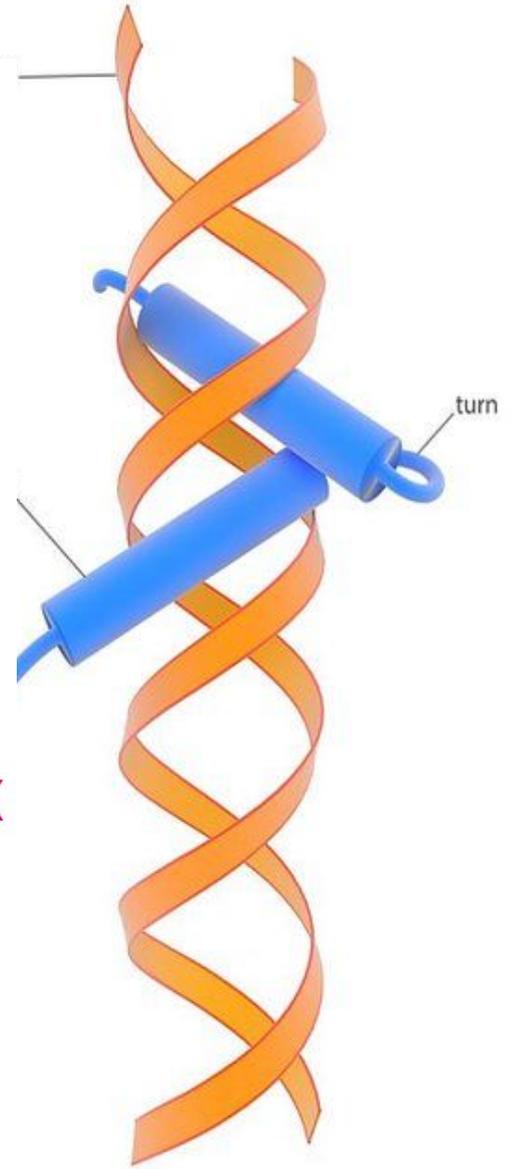
Интерлок



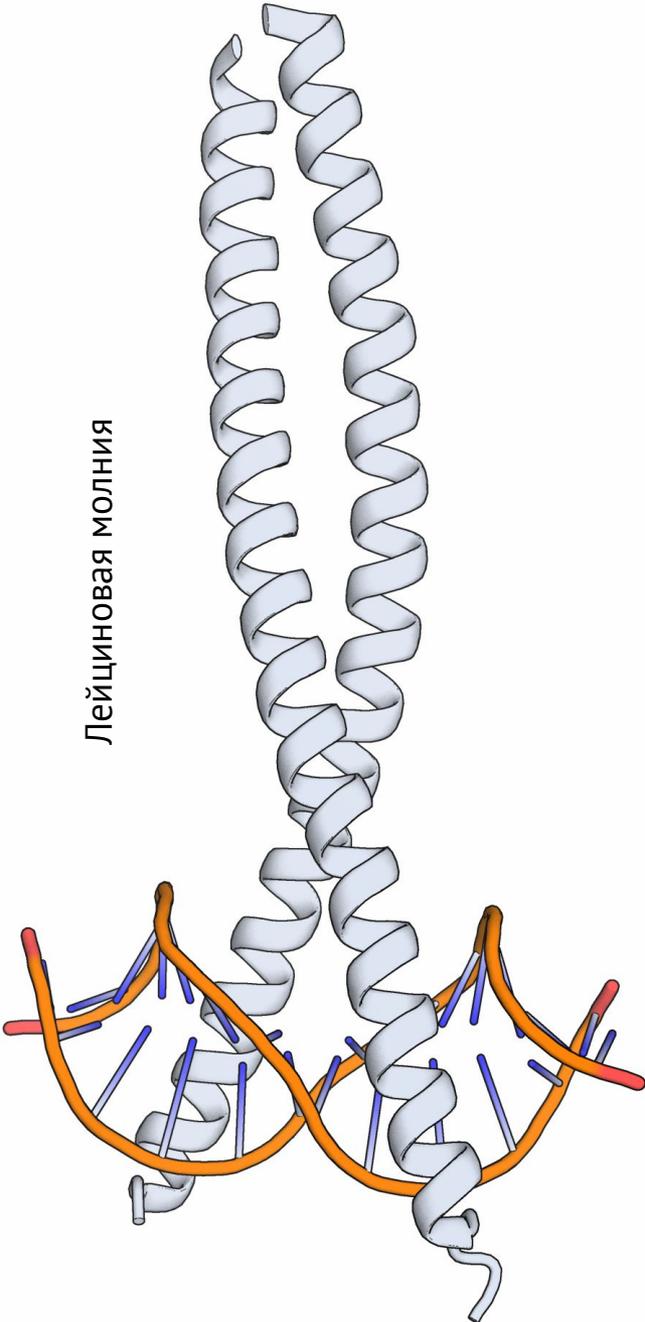
Греческий ключ



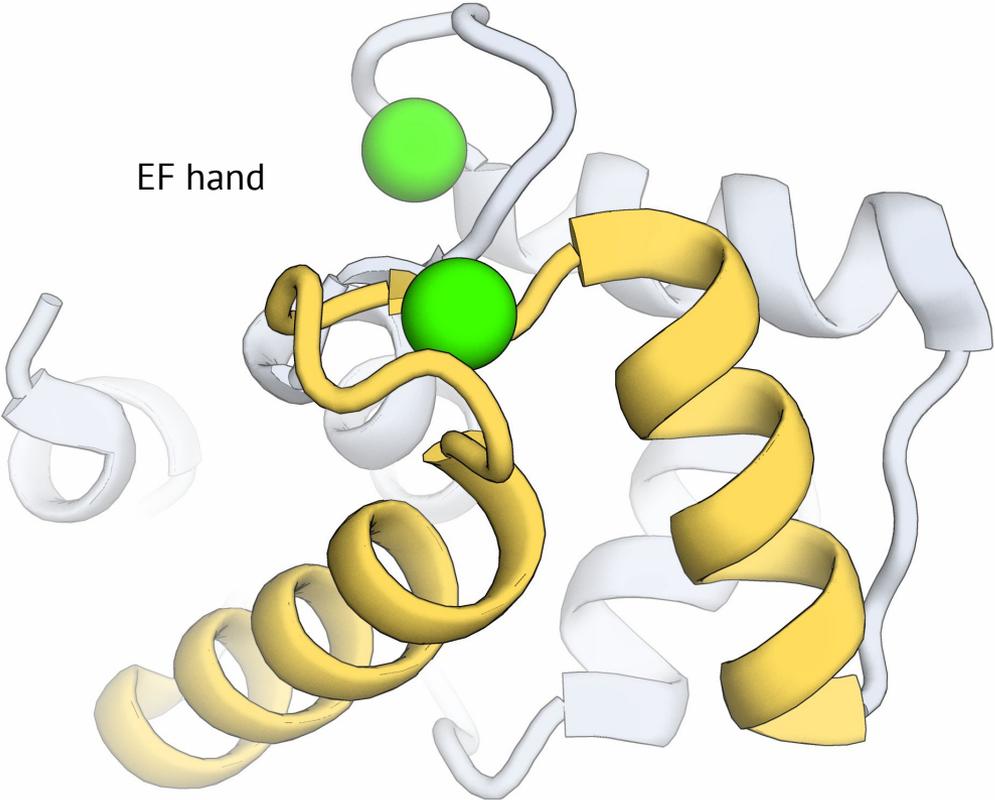
Helix-turn-helix



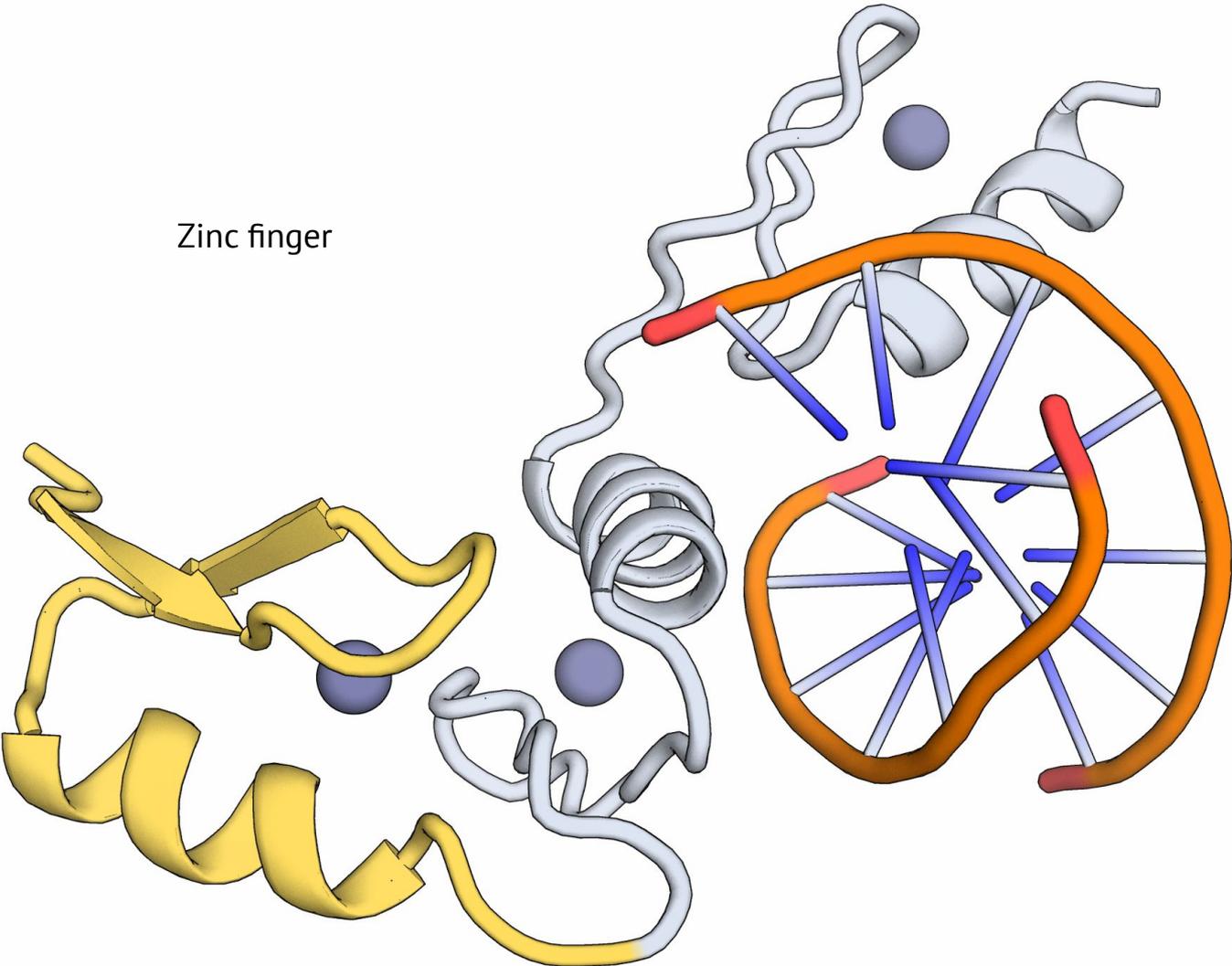
Лейциновая молния



EF hand



Zinc finger



Укладки

- Функция определяется на уровне структуры
- Структуры консервативнее последовательностей
- Хотим узнать функцию нашего белка и знаем его структуру – сначала стоит посмотреть на уже аннотированные похожие структуры
- Подспорье в изучении структурно-функциональных взаимоотношений – базы данных расклассифицированных укладок структурных доменов

Но что такое структурный домен?

Домены

Что такое домен?

Три определения:

- По функции (функциональный домен)
- По сравнению последовательностей (эволюционный домен)
- По структуре (структурный домен)

Функциональный домен (биохимия/биоинженерия)

Минимальная часть полипептидной цепи, которая:

- сохраняет (in vitro) как минимум одну из активностей полного белка
- как правило, может автономно свернуться в правильную, нативную структуру

Genetic definition of a protein-splicing domain: Functional mini-inteins support structure predictions and a model for intein evolution

Victoria Derbyshire, David W. Wood, Wei Wu, John T. Dansereau, Jacob Z. Dalgaard, and Marlene Belfort

PNAS October 14, 1997 94 (21) 11466-11471; <https://doi.org/10.1073/pnas.94.21.11466>

Рекомбиназа А из *Mycobacterium tuberculosis* (790 а.о.) содержит интеин (440 а.о.), белок, обладающий способностью автономно вырезаться из полипептидной цепи белка-предшественника (явление белкового сплайсинга). Это – первая активность интеина.

There are 83 sequences with the following architecture: DnaB, DnaB_C, Intein_splicing, LAGLIDADG_3

[DNAB_MYCTU](#) [*Mycobacterium tuberculosis* (strain ATCC 25618 / H37Rv)] Replicative DNA helicase
EC=3.6.4.12 EC=3.1.-.- (874 residues)

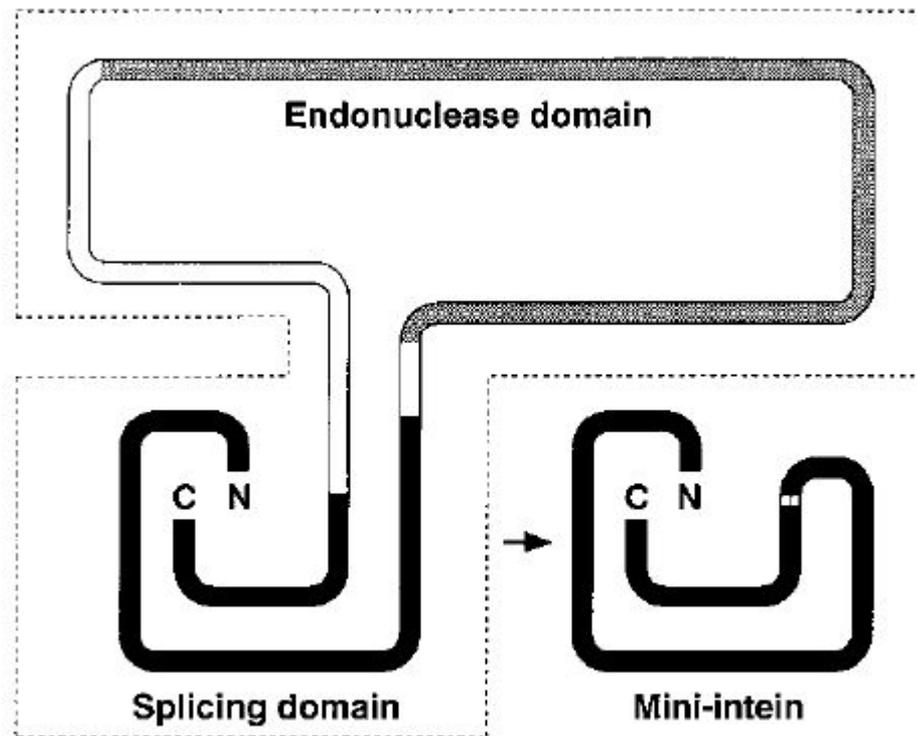


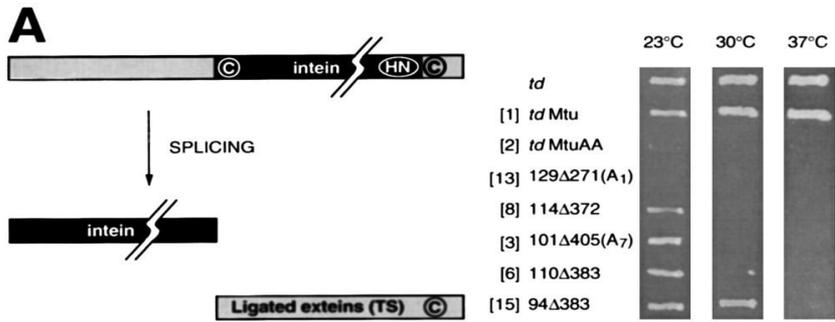
[Show](#) all sequences with this architecture.

Этот интеин обладает также эндонуклеазной активностью (вторая активность).

По сходству последовательности этого белка с последовательностями других, более изученных интеинов, в т.ч. интеина с расшифрованной РСА структурой (1VDE), была высказана гипотеза о том, что за две разные активности отвечают разные домены.

При этом за белковый сплайсинг отвечает домен, который составлен из N-концевого и C-концевого участков полипептидной цепи





Для проверки гипотезы авторы создали 21 конструктор генов интеина, в которых удалены разные внутренние участки полипептидной цепи.

B

	Intein size (aa)	TS phenotype	MIC splicing
1	440	+5	+5
2	440	-	-
3	137	+2	0*
4	147	+1	+1
5	159	+1	+1
6	168	+2	+3
7	171	+3	+2
8	183	+1	+2
9	192	+1	+1
10	165	+1	+1*
11	187	+1	+1
12	198	+1	+1
13	299	-	+1
14	154	+3	+2
15	152	+3/+4	+2
16	149	-	-
17	149	+2	+1
18	170	+1	+1
19	160	-	-
20	142	+2	+1
21	137	+1	+1

A B C D E H FG

endonuclease/spacer

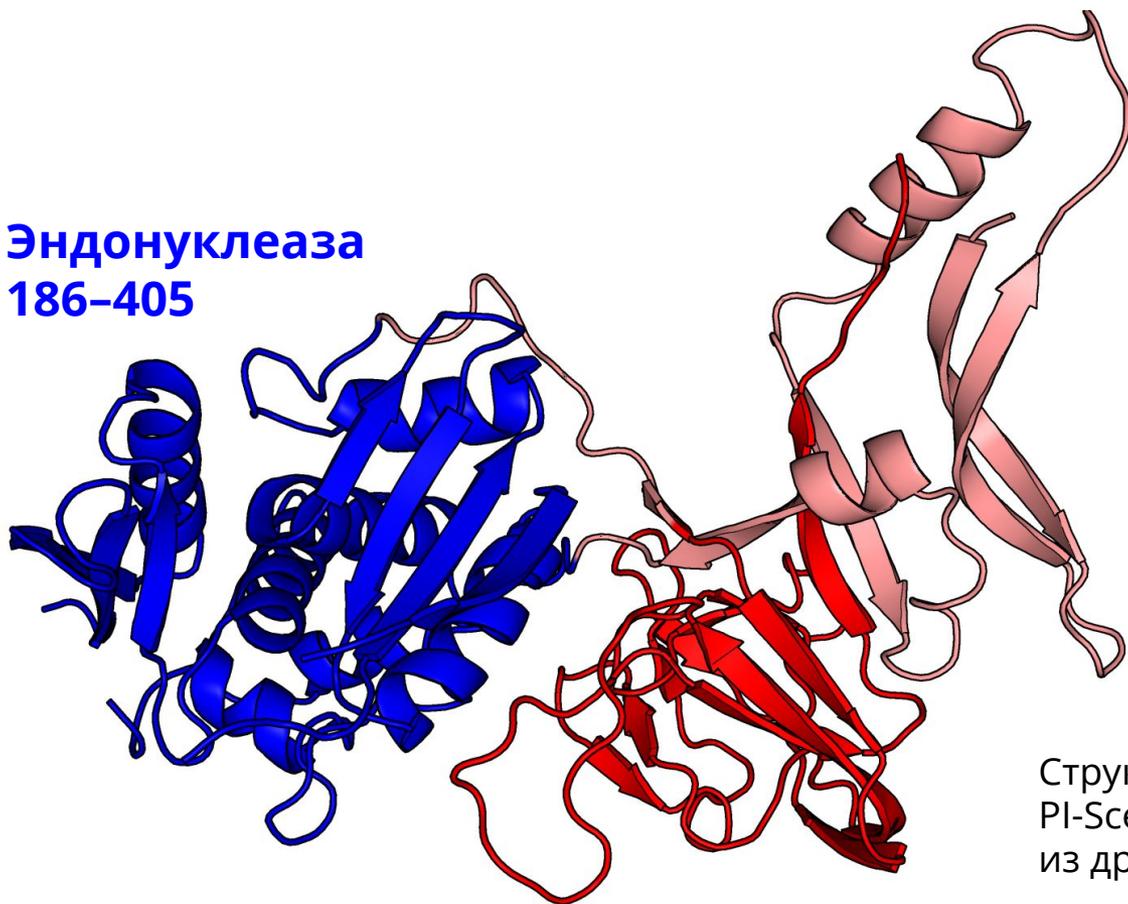
Конструкторы были встроены в ген другого белка (тимидилатсинтазы, TS) и экспрессировались в *E.coli*

Активность проверялась по наличию нативного белка TS (без вставки интеина)

Результат: белковый сплайсинг сохранялся в тех случаях, когда удаленный участок не затрагивал первые 96 и последние 35 а.о.

Вывод авторов: функциональный домен автономного белкового сплайсинга состоит из остатков 1–96 и 406–440 (всего 131 из полных 440)

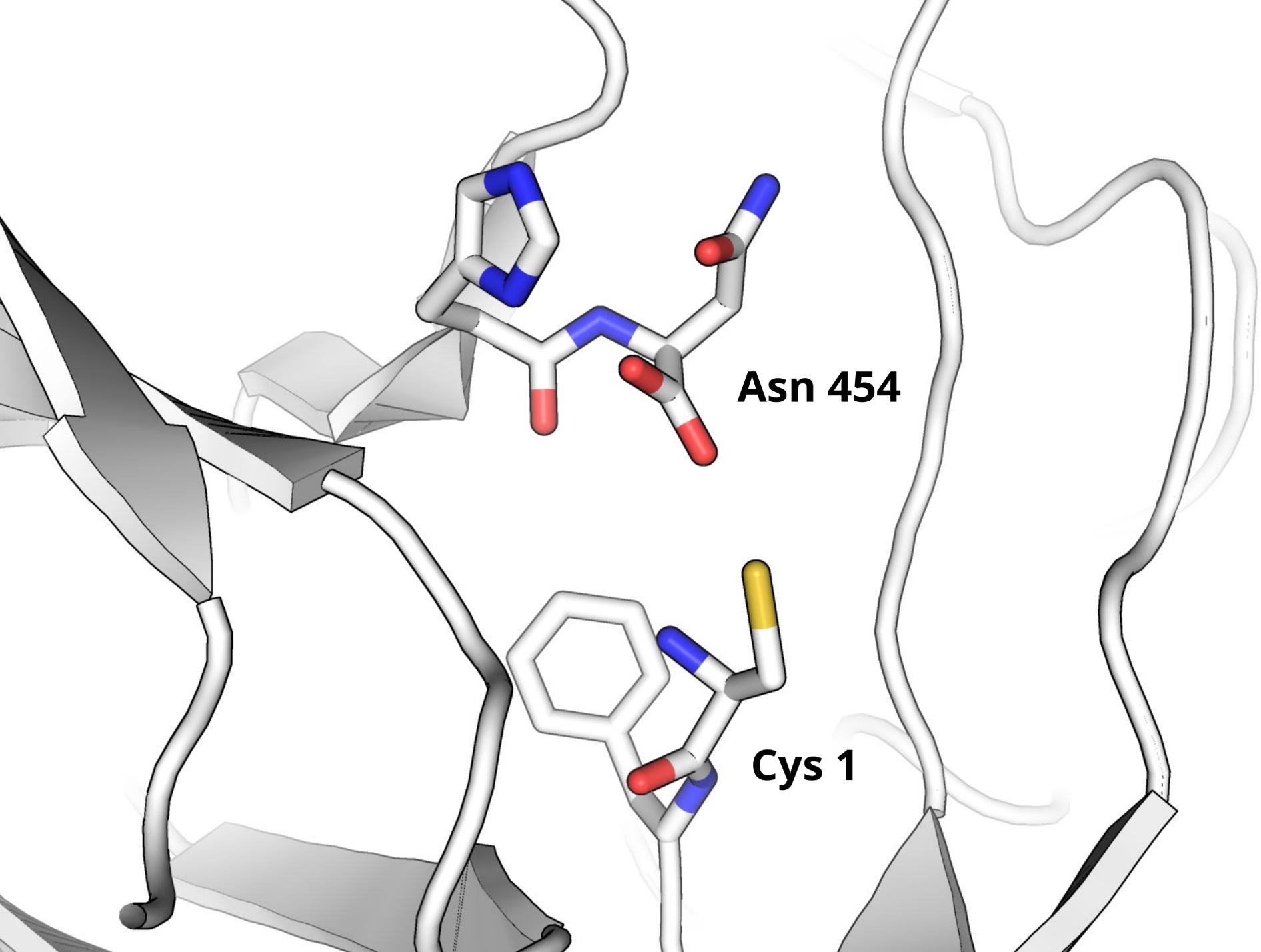
Эндонуклеаза
186–405



Интеин
1–181,
406–454

Минимально
функциональный
интеин
1–96,
419–454

Структура гомологичного белка
PI-SceI – хоминг эндонуклеазы
из дрожжей (PDB код 1VDE)



Последовательность интеина консервативна

Об этом свидетельствуют доменные архитектуры трех белков из разных грибов, описанные в Pfam



Доменная структура полноразмерного белка PI-Scel



Доменная структура белка VMA1



Доменная структура белка TFP1 (аннотирован по сходству)

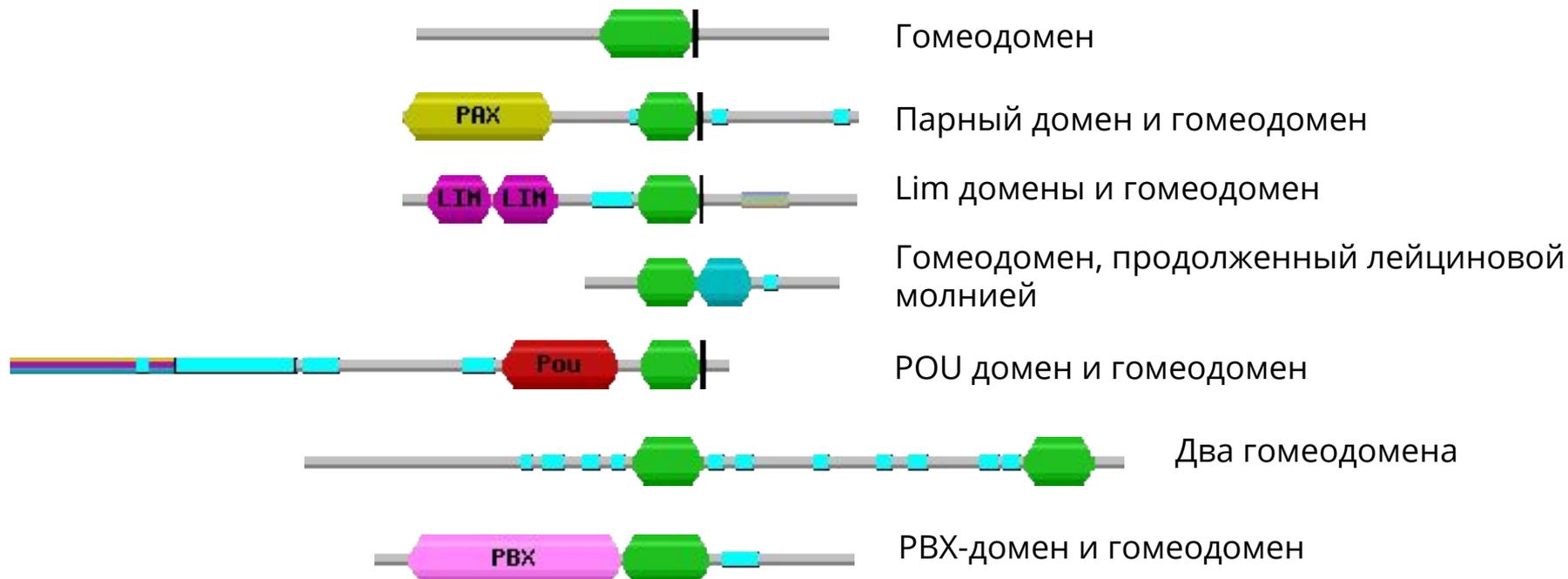
Эволюционный домен (биоинформатика: последовательности)

Достаточно длинный участок полипептидной цепи, который:

- эволюционно консервативен — существуют достоверно сходные участки в других белках
- Обычно замечен в перемешивании доменов (domain shuffling), то есть имеются примеры белков, где есть достоверно сходные с ним участки, но есть также несходные между собой (но эволюционно консервативные) участки

Гомеодомены активно перемешивались в эволюции

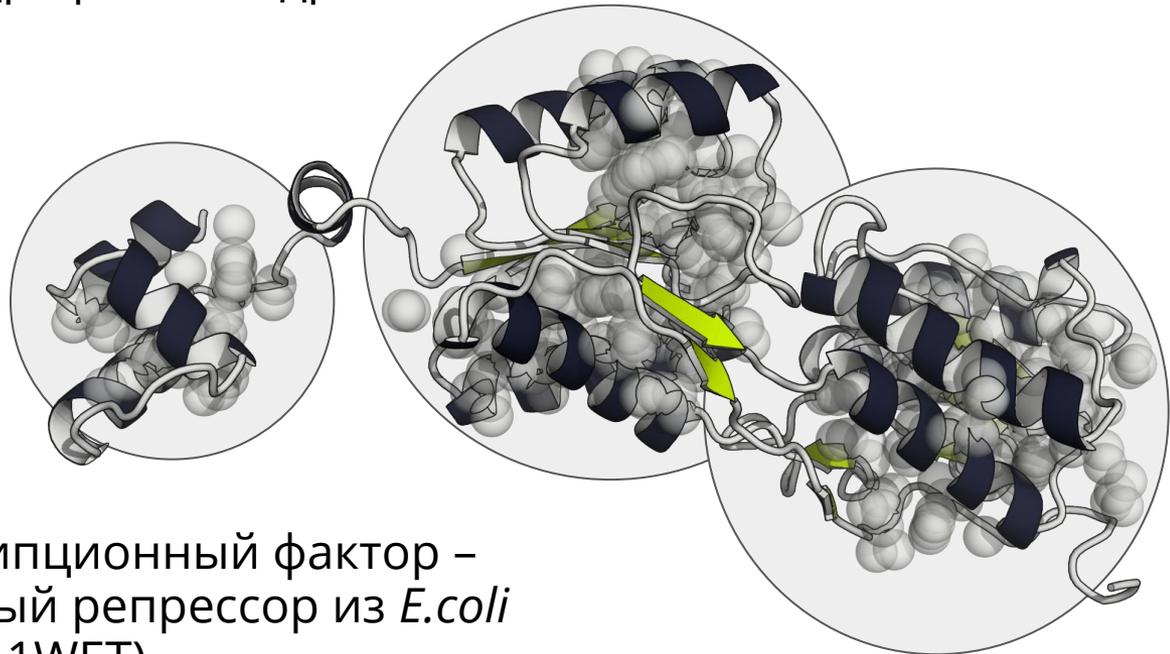
Об этом можно судить по 609 ! (2018 год) различным доменным архитектурам гомеобелков (Pfam)



Структурный домен (биоинформатика 3D структур)

Обособленная в пространстве часть белка,
его структурная единица, имеющая:

- сравнительно мало контактов с другими частями белка
- собственное гидрофобное ядро



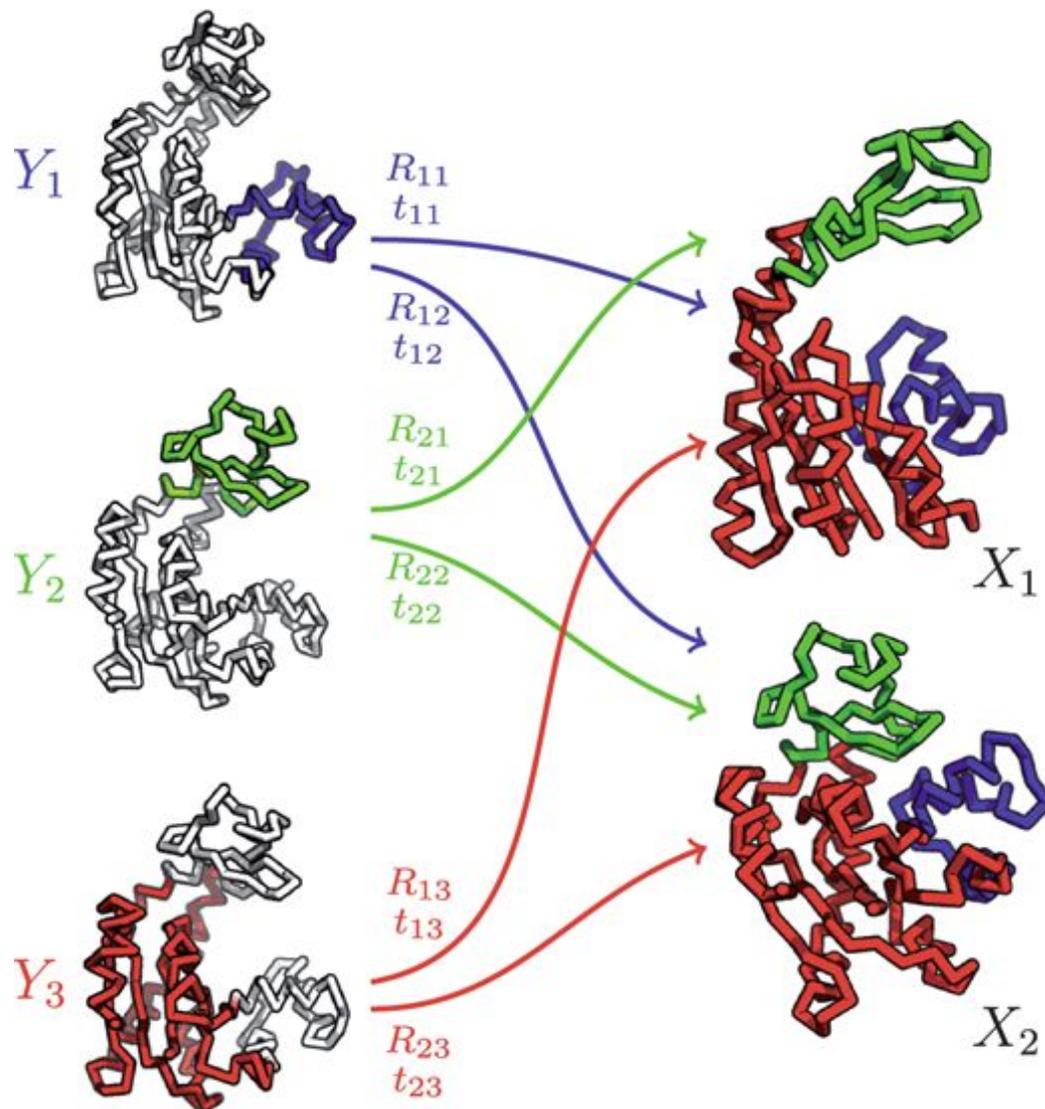
Транскрипционный фактор –
пуриновый репрессор из *E.coli*
(PDB код 1WET)

Структурный домен (биофизика)

Часть молекулы
белка, движущаяся
как твердое тело

**A probabilistic model for
detecting rigid domains
in protein structures**

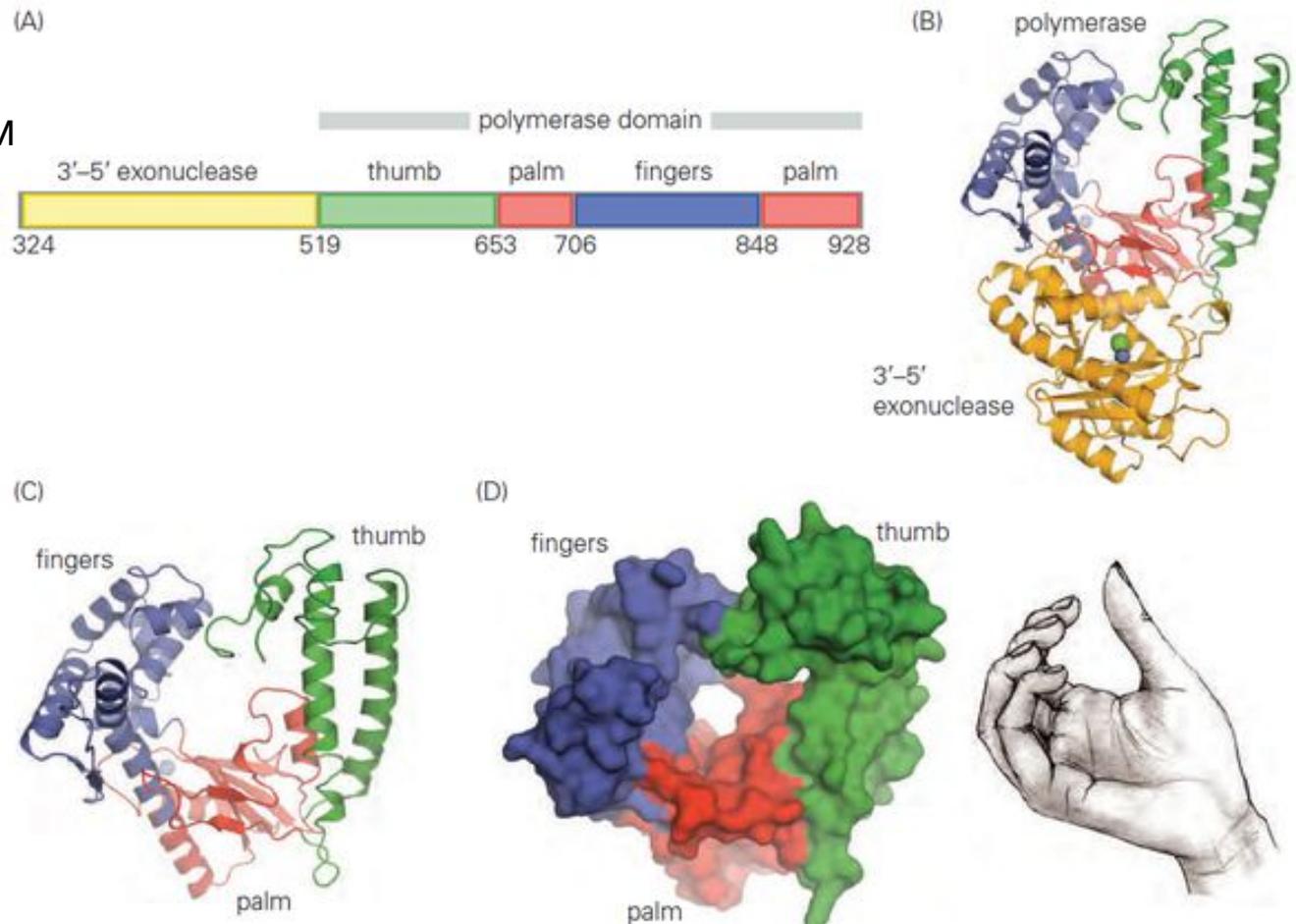
Thach Nguyen, Michael Habeck
Bioinformatics, Volume 32, Issue 17, 1
September 2016



Домен белка XXX (жизнь)

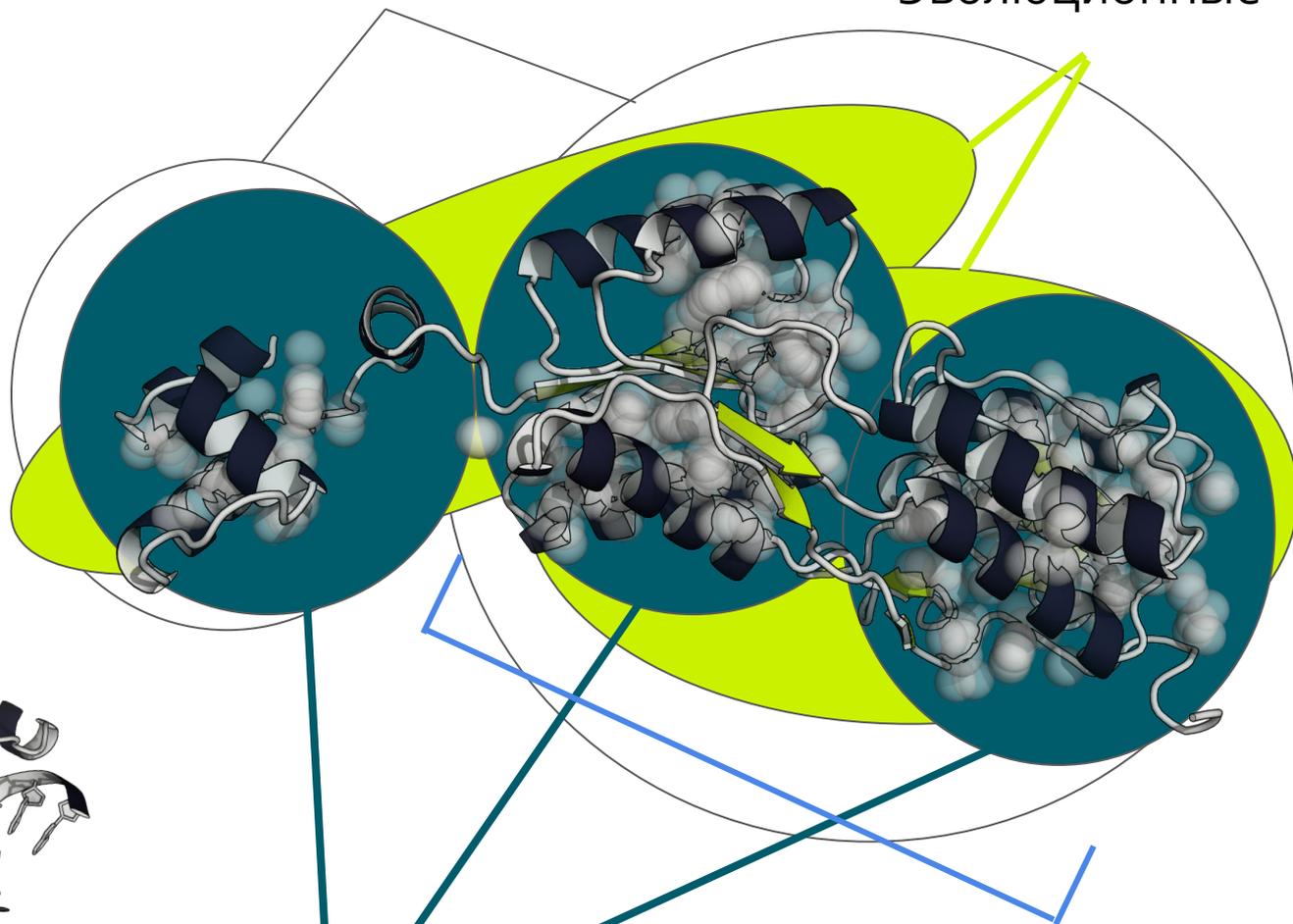
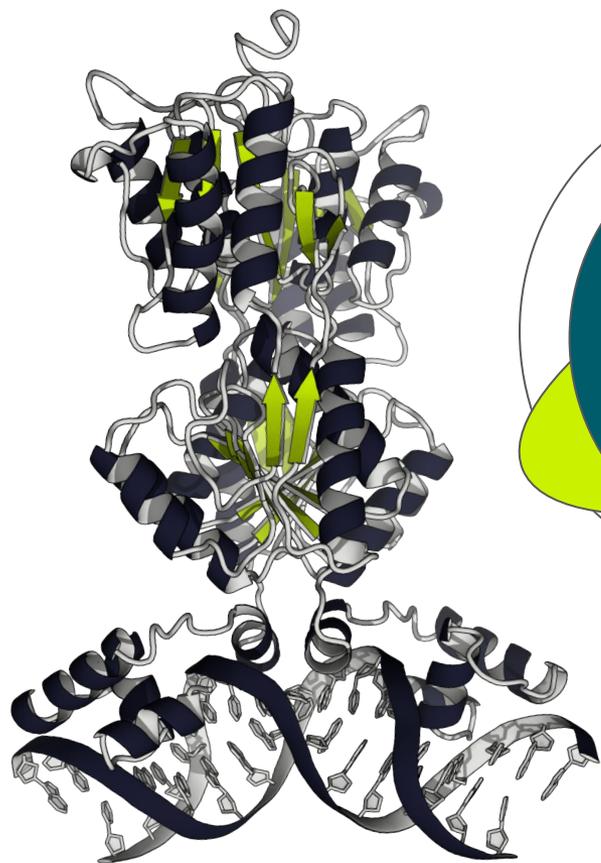
Часть белка, названная доменом

- Субъективизм
- Образность
- Традиция



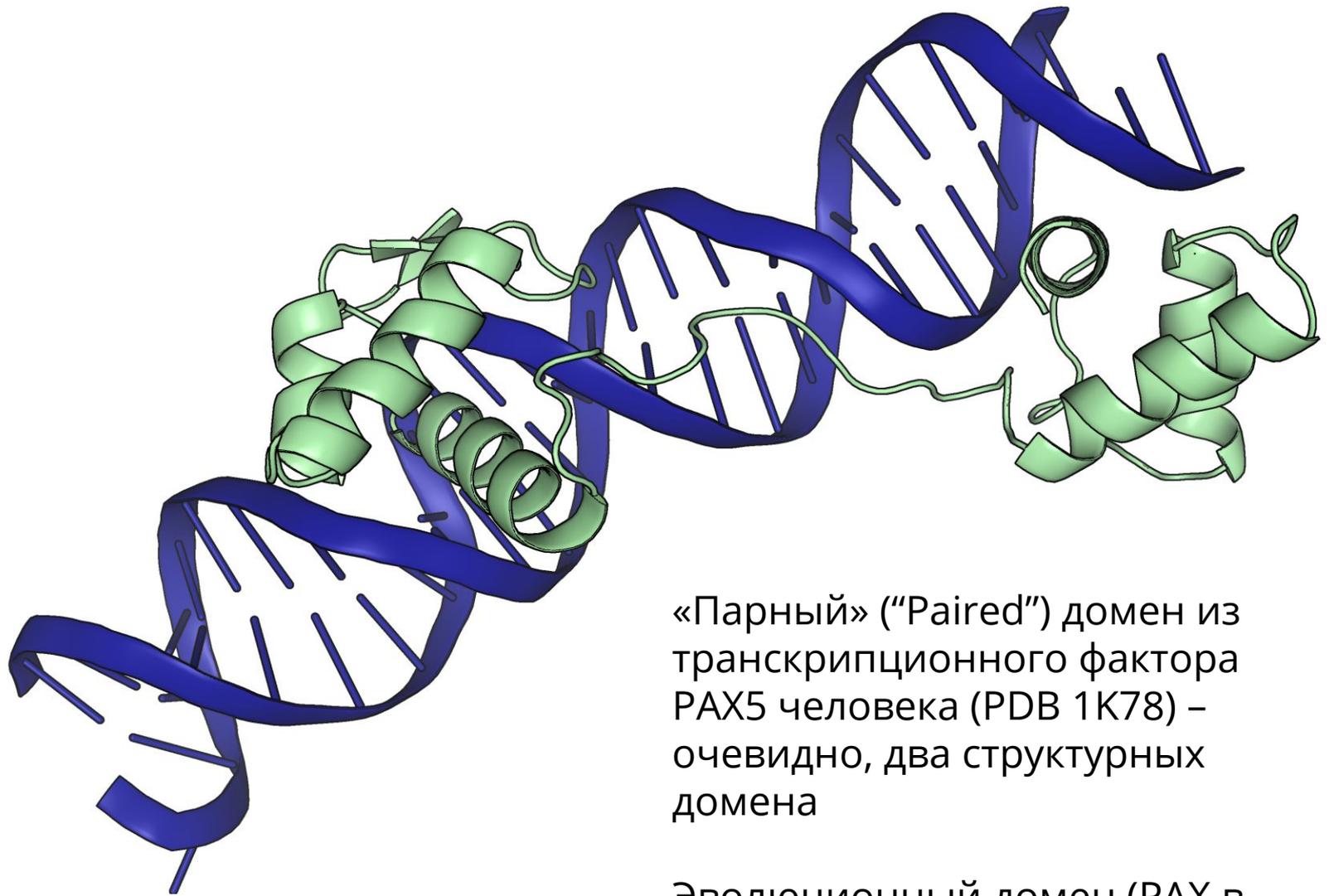
Функциональные

Эволюционные



Структурные

Классифицируется
(БД SCOP) как один
структурный домен



«Парный» ("Paired") домен из транскрипционного фактора PAX5 человека (PDB 1K78) – очевидно, два структурных домена

Эволюционный домен (PAX в Pfam) включает оба структурных домена (126 а.о.)

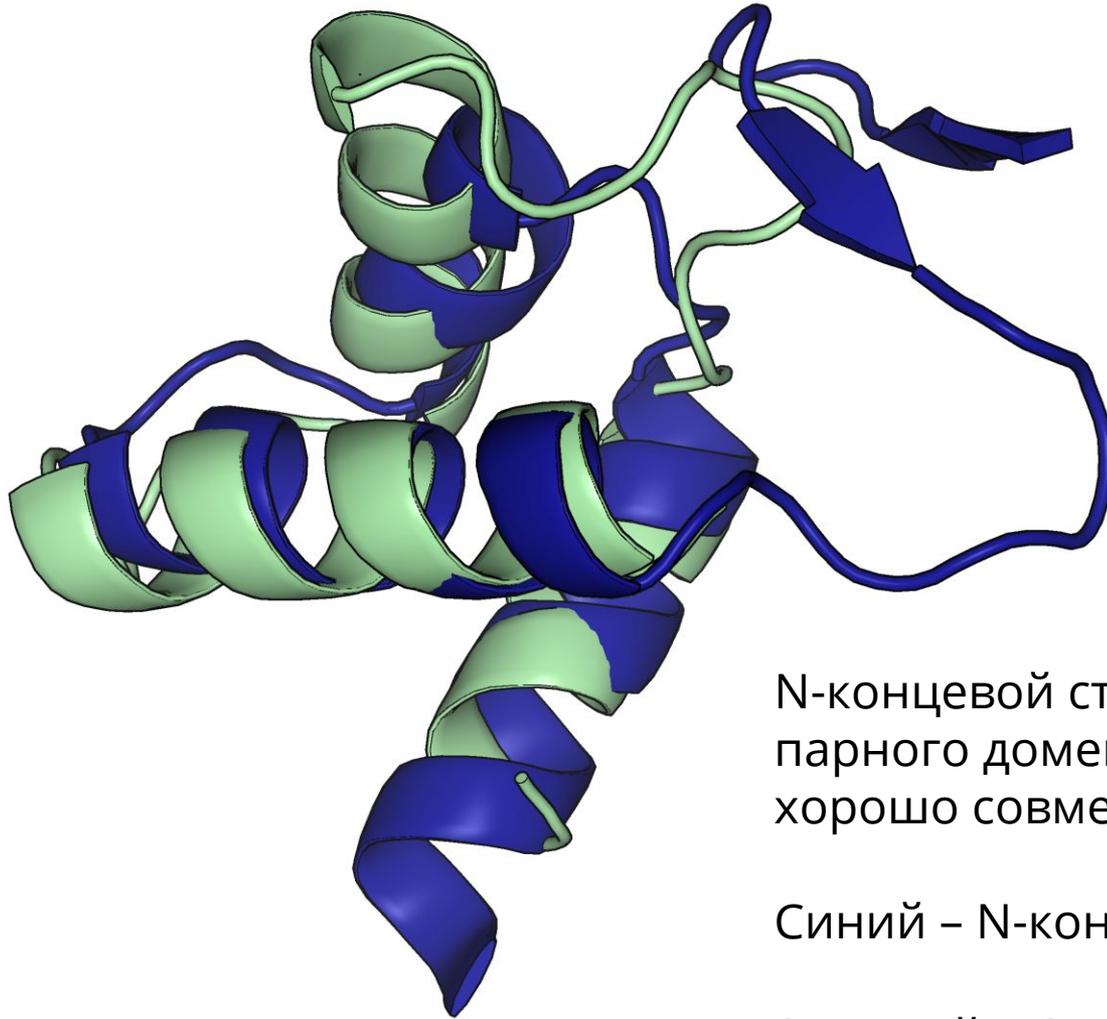
Последовательности PAX/prd доменов консервативны по всей длине

```

          *           20           *           40           *           60
PAX6_BRARE : SHSGVNQLGGVVFVNGRPLPDSTRQKIVELAHSGARPCDISRIHQVSNCGCVSKILGRYYETGSIRP : 65
PAX2_HUMAN : GHGGVNQLGGVVFVNGRPLPDVVRQRIVELAHQGVRPCDISRQLRVSHGCVSKILGRYYETGSIKP : 65
PAX1_MOUSE : TYGEVNQLGGVVFVNGRPLPNAIRLRIVELAQLGIRPCDISRQLRVSHGCVSKILARYNETGSILP : 65
PAX3_HUMAN : GQGRVNQLGGVVFINGRPLPNHIRHKIVEMAHHGIRPCVISRQLRVSHGCVSKILCRYQETGSIRP : 65
GSBN_DROME : GQGRVNQLGGVVFINGRPLPNHIRLKI VEMAASGVRPCVISRQLRVSHGCVSKILNRYQETGSIRP : 65
GSB_DROME/ : GQGRVNQLGGVVFINGRPLPNHIRRQIVEMAAAGVRPCVISRQLRVSHGCVSKILNRFQETGSIRP : 65
          g VNQLGGVVF6NGRPLP1 R IVE6A G RPC ISRqLrVShGCVSKIL R5 ETGSI P

          *           80           *           100           *           120
PAX6_BRARE : RAIGGSKPR-VATPEVVGKIAQYKRECPSIFAWWEIRDRLISEGVCTNDNIPSVSSINRVLR : 125
PAX2_HUMAN : GVIGGSKPK-VATEKVVVDKIAEYKRQNPTMFAWEIRDRLIAEGICDNDTVPSVSSINRIIR : 125
PAX1_MOUSE : GAIGGSKPR-VTTPNVVKHIRDYKQGDPIFAWEIRDRLIADGVCDKYNVPSVSSISRILR : 125
PAX3_HUMAN : GAIGGSKPKQVTPPDVEKKIEEYKRENPFGMFSWEIRDKLLKDAVCDRNTVPSVSSISRILR : 126
GSBN_DROME : GVIGGSKPK-VTSPEIETRIDELRKENPSIFSWEIREKLIKEGFAD---PPSTSSISRLLR : 122
GSB_DROME/ : GVIGGSKPR-VATPDIESRIEELKQSQPGIFSWEIRAKLIEAGVCDKQNAPSVSSISRLLR : 125
          g IGGSKP4 V 3P 6 I 4 P 6F WEIR 4L6 g cd PSvSSI R66R

```



N-концевой структурный домен
парного домена
хорошо совмещается с C-концевым

Синий – N-концевой

Зеленый – C-концевой

Но достоверного сходства
последовательностей не наблюдается