

Анализ трехмерных структур



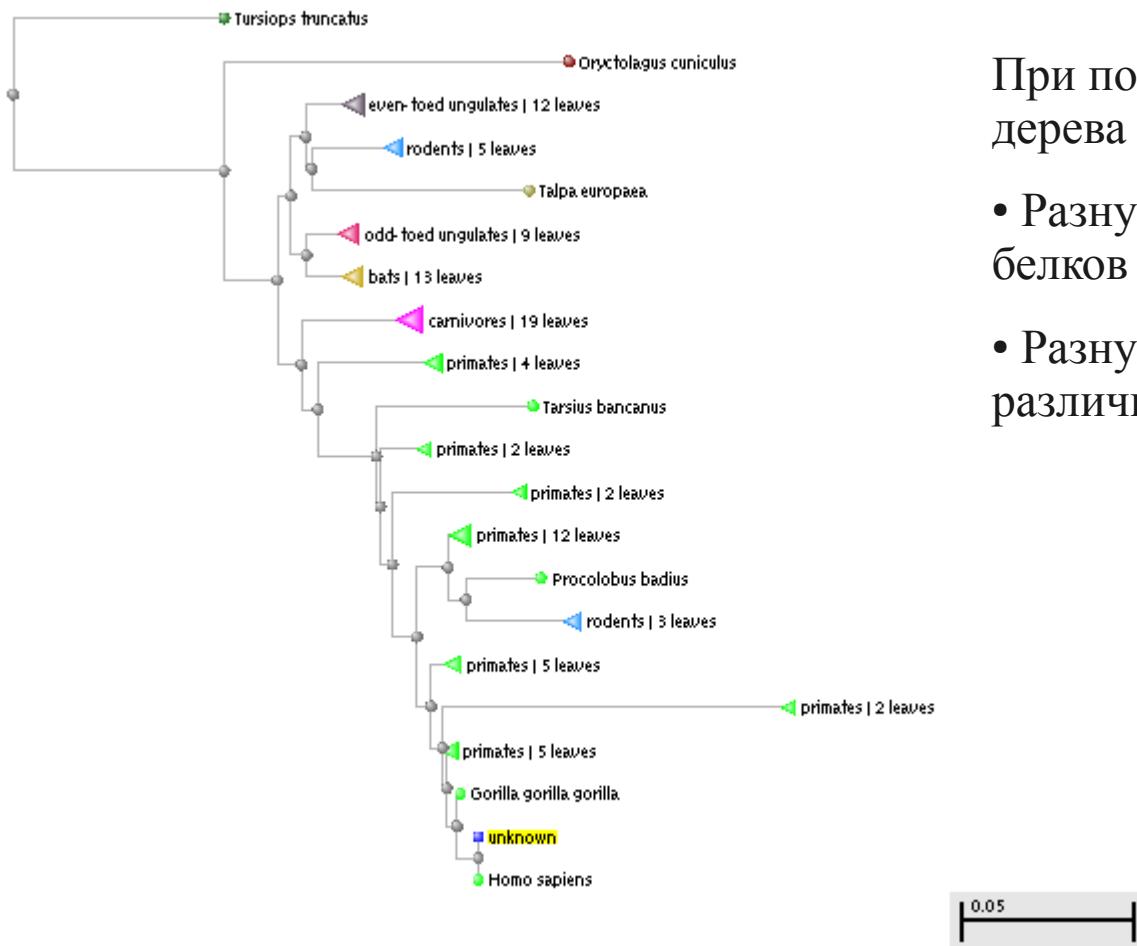
каф. биофизики
биологического ф-та МГУ

Зленко Д.В.

ЭВОЛЮЦИЯ

Ключевым, очень сильным и мало доказанным положением любой теории, претендующей на установление родства между группами организмов является **линейная связь между родством и сходством**.

Принимая это, на основе сходства последовательностей можно попытаться определить степень родства организмов и как давно они разошлись.



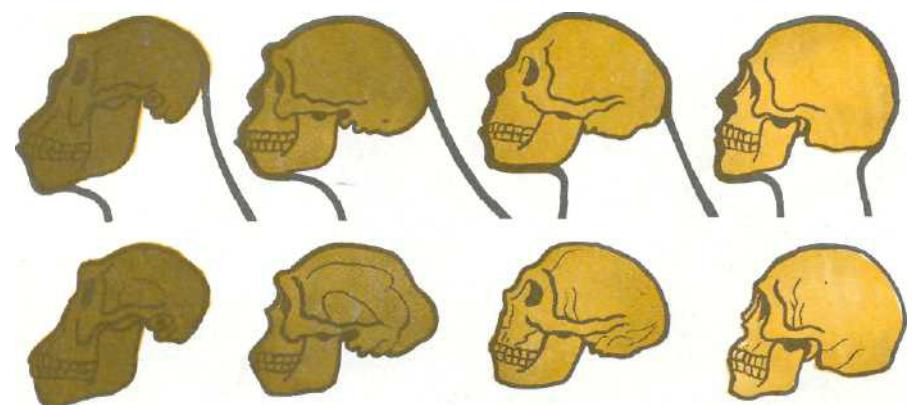
При построении бинарного филогенетического дерева необходимо учитывать дополнительно:

- Разную скорость эволюции различных групп белков
- Разную скорость эволюции в пределах различных таксономических групп

Однонуклеотидные полиморфизмы

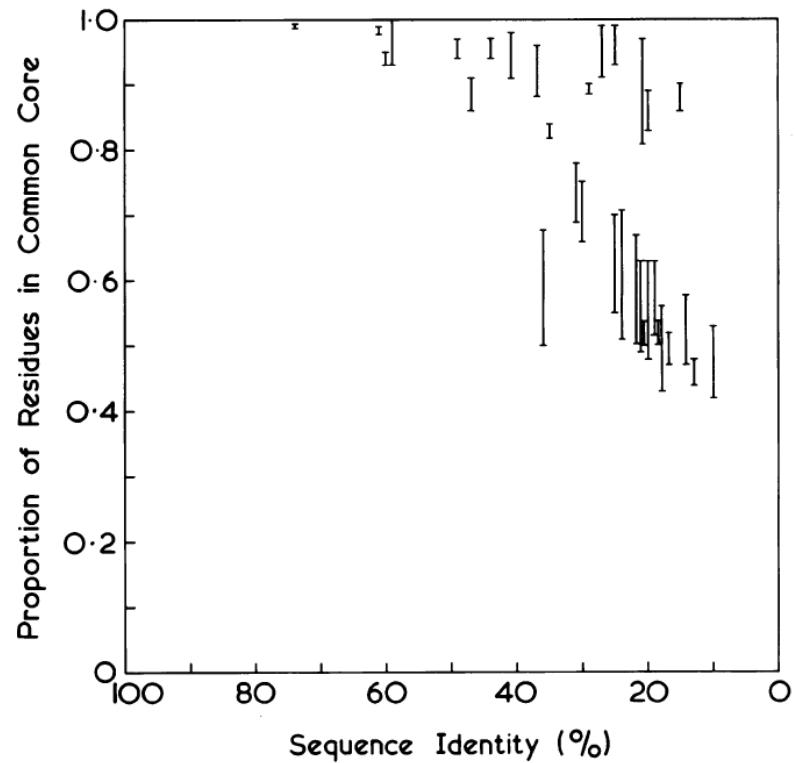
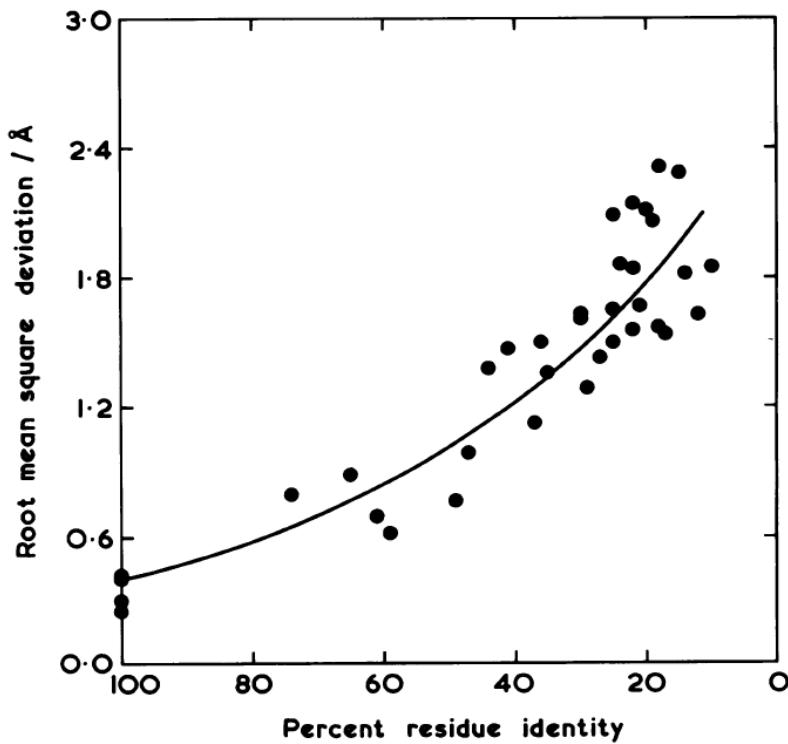
SNP - это точечные однонуклеотидные замены, которые характеризуют изменчивость от особи к особи.

- в среднем встречаются раз в 500 - 2000 п.о.
- можно подразделить на значащие, и синонимичные.
- анализ синонимичных замен позволяет определить время расхождения двух групп особей.
- анализ значащих замен позволяет выявить области, подверженные давлению отбора.



Эволюция белков

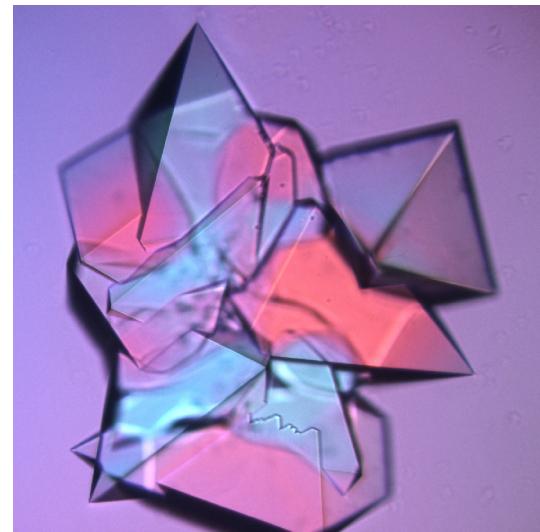
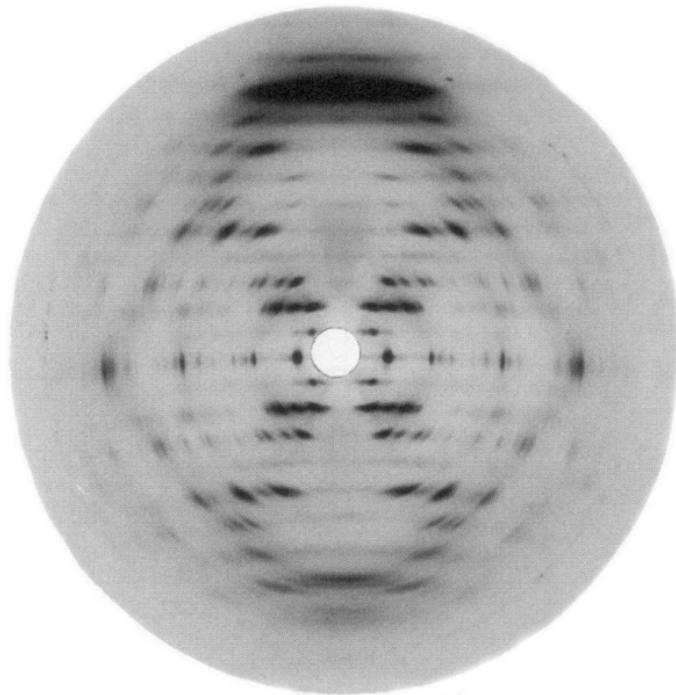
- Наибольшей вариабельностью обладают расположенные на периферии фрагменты последовательности
 - Схема укладки цепи консервативнее собственно первичной последовательности
 - Мутации, приводящие к изменению во взаимном расположении элементов вторичной структуры встречаются чаще, чем мутации меняющие эти элементы.



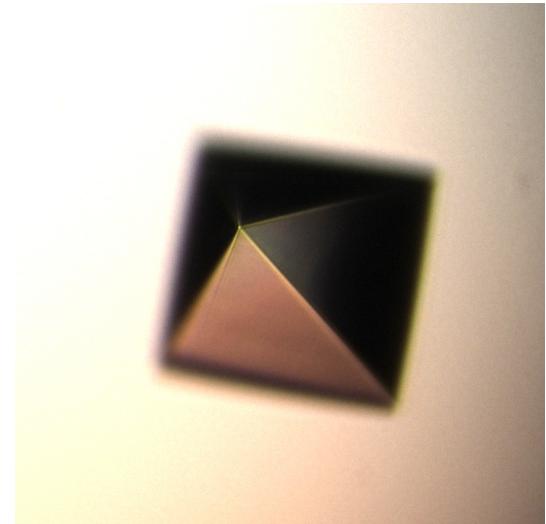
Chothia C. and Lesk A.M. "The relation between the divergence of sequence and structure in proteins". 1986. EMBO J. 5(4): 823–826.

Рентгеноструктурный анализ

- В основе метода лежит явление дифракции рентгеновского излучения на электронах атомов кристалла.
- Рентгеновское излучение - тормозное излучение электронов с длинами волн в диапазоне от 0.1 до 100 \AA .
- Дифракция есть явление преломления света на границе двух сред.
- Кристалл можно представить в виде набора дифракционных решеток.



Вильгельм Рентген
(1845 - 1923)

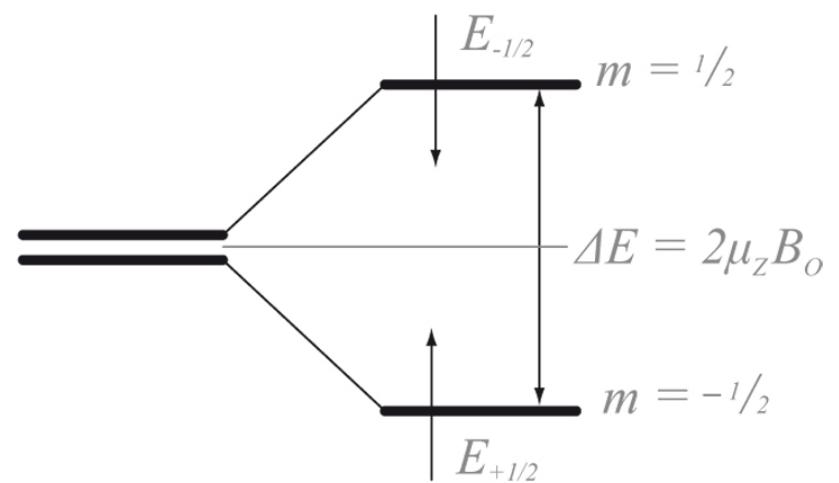
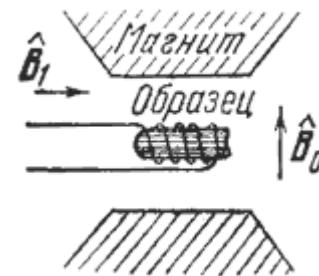


Ядерный Магнитный Резонанс

- В основе метода лежит явление поглощения электромагнитного излучения за счет переориентации ядерных спинов во внешнем магнитном поле.
- Спин ядра имеет разную энергию в зависимости от ориентации относительно внешнего поля.
- Переориентация спинов возможна с поглощением или излучением кванта света.
- Каждое ядро находится не только во внешнем поле, но и в поле соседних ядер.



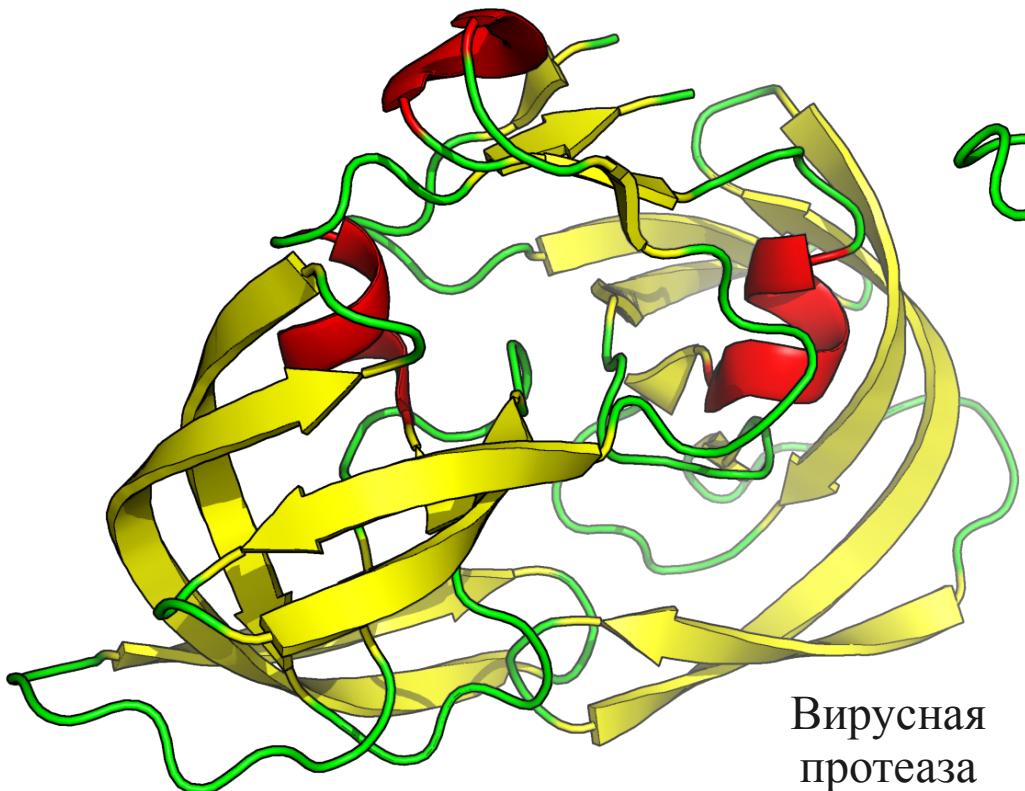
Айзек Раби
(1898 - 1988)



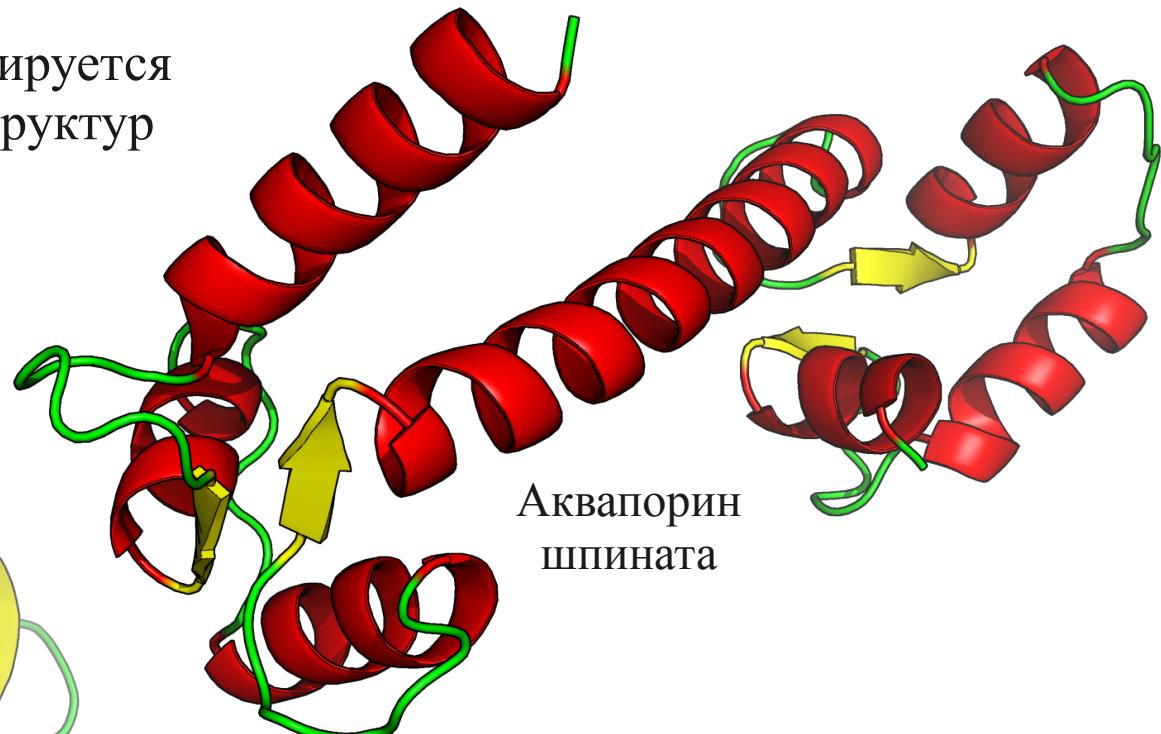
PDB - Protein Data Bank



- Объем базы данных PDB составляет без малого 100 тыс последовательностей
- Меньший объём успешно компенсируется трёхмерностью задачи сравнения структур друг с другом



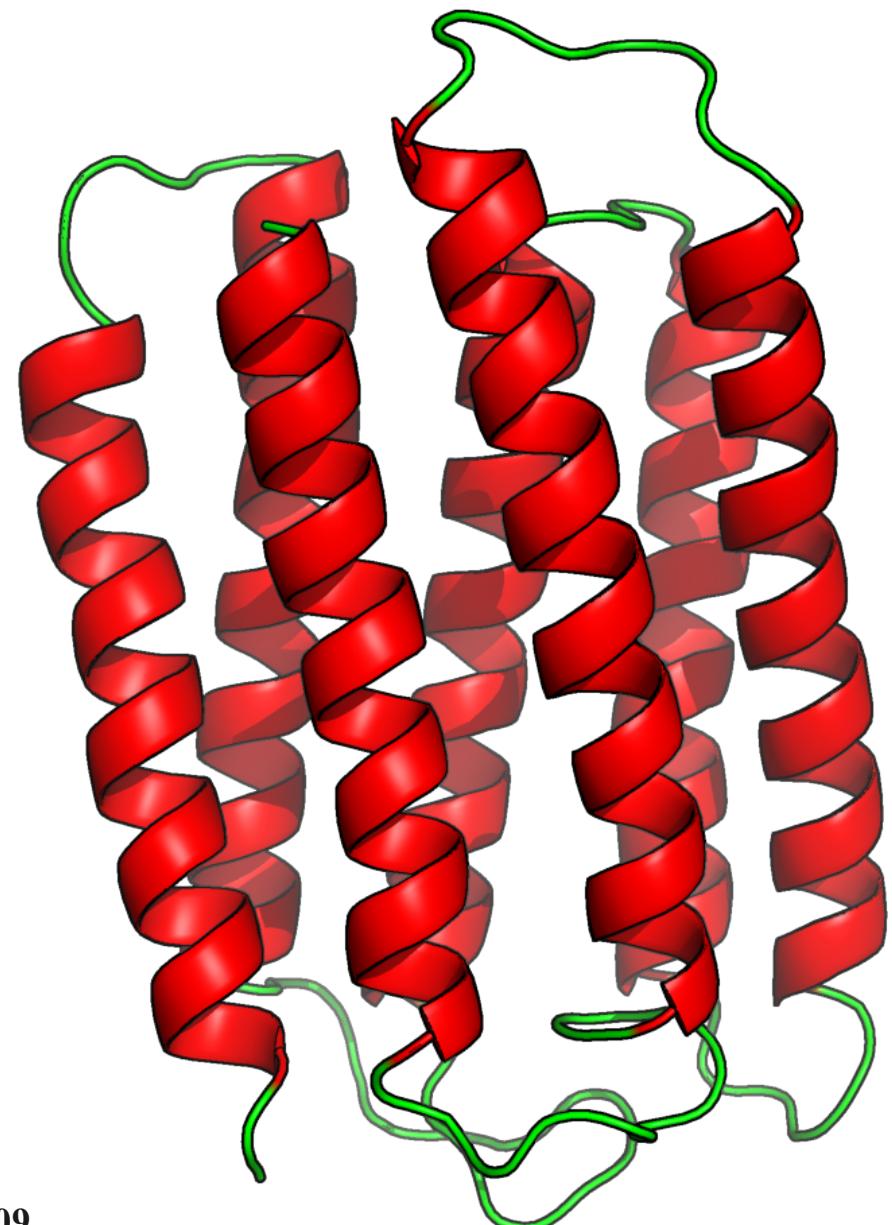
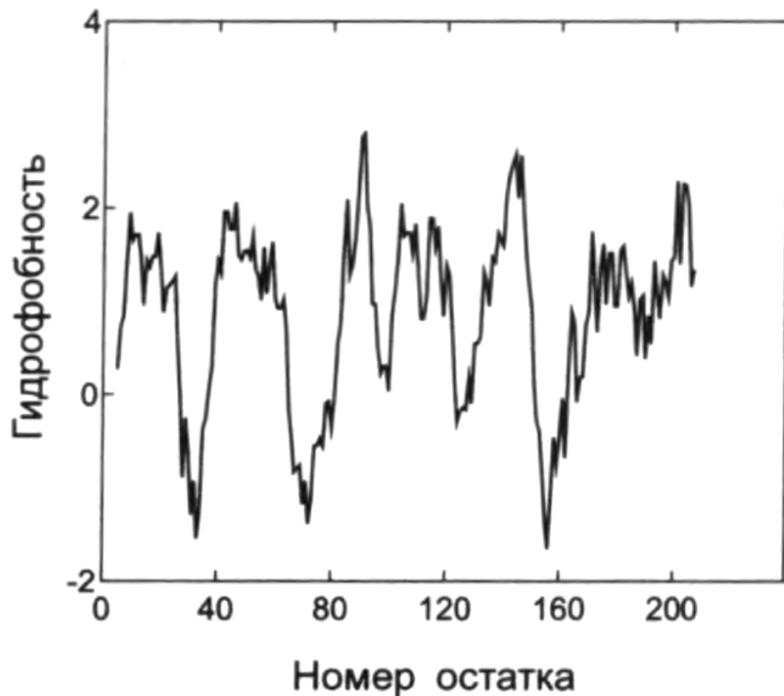
Вирусная
протеаза



Аквапорин
шпината

Профили гидрофобности

- Гидрофобность аминокислотных остатков коррелирует со свойствами их окружения
- Трансмембранные α -спирали сложены гидрофобными остатками
- Многие α -спирали имеют гидрофобную и гидрофильную поверхности.



Леск А. «Введение в биоинформатику». БИНОМ. Москва. 2009.

Krystek S.R., Metzler W.J., Novotny J. "Hydrophobicity Profiles for Protein Sequence Analysis. Curr Protoc Protein Sci". 2001. doi: 10.1002/0471140864.ps0202s00.

Выравнивание не показательно

Парное выравнивание леггемоглобина люпина и миоглобина кашалота не показательно (сходство 16.1%), как и для бычьего химотрипсина и субтилизина *Bacillus licheniformis* (7.0%):

Alignment

🖨️ Learn how to print this alignment in color

1	MNNSKIIISKVLLSLSLFTVGASAFVIQDELMQKHNHAKAEVSAAEIKKHEEKWNKYYGVNA	CGVPA	5 P00766 CTRA_BOVIN
1	*** *		60 P09331 ETA_STAAU
6	TQPVLSGLSPIVNGEEAVPGSWFWQSILQ-DITGIFFCGGSLINENVVVTAAHCGVITS-		63 P00766 CTRA_BOVIN
61	ENLFLKEELFSKV--DEKDRCYTYNTIGNVFVKGTSATGVLIGINTVLTNPRIAKPANG		117 P09331 ETA_STAAU
64	--DVIYAG--EDPSSSEKTIQKLIAKVKNSKYNLSITINNDITIYLVIST--AIS		114 P00766 CTRA_BOVIN
118	DPSIIVSRSPINTDDNNTETYGEYEVEIQLQ---PFGAGVLDALIRPKEDQNGVSI		173 P09331 ETA_STAAU
115	SQTISAVCLPSASDDFAAGTTCVTTGWGLTRYTNANTPDLQQASLPLLNSNTNKKYWG		174 P00766 CTRA_BOVIN
174	DKLISPAK-IGTSNDLFDGDKELELIGYFED---HKVNQHPRSEIEUTTISPGRLRYYG		227 P09331 ETA_STAAU
175	KIKDAMICAGASGVSSCMGDGGGPLVCKTKNGAVTLVGVISVGSSTCSTSTPGVYARVTAI		234 P00766 CTRA_BOVIN
228	TVPQ-----N-SSE-IFNSNG--ELVGVIHSSEYSHLDPEHQINY-GVGI		267 P09331 ETA_STAAU
235	VNVVQQTAAAN-- 245 P00766 CTRA_BOVIN		
268	QNYKPIINEKNE 280 P09331 ETA_STAAU		

Annotation

- Beta strand
 - Mutagenesis
 - Domain
 - Propeptide
 - Disulfide bond
 - Signal peptide
 - Chain
 - Active site
 - Turn
 - Helix
- Amino acid properties
- Similarity
 - Hydrophobic
 - Negative
 - Positive
 - Aliphatic
 - Tiny
 - Aromatic
 - Charged
 - Small
 - Polar
 - Big
 - Serine Threonine

- Обе пары белков выполняют аналогичные функции

- Тем не менее пространственные структуры первой пары очень похожи, а для второй не имеют ничего общего

Alignment

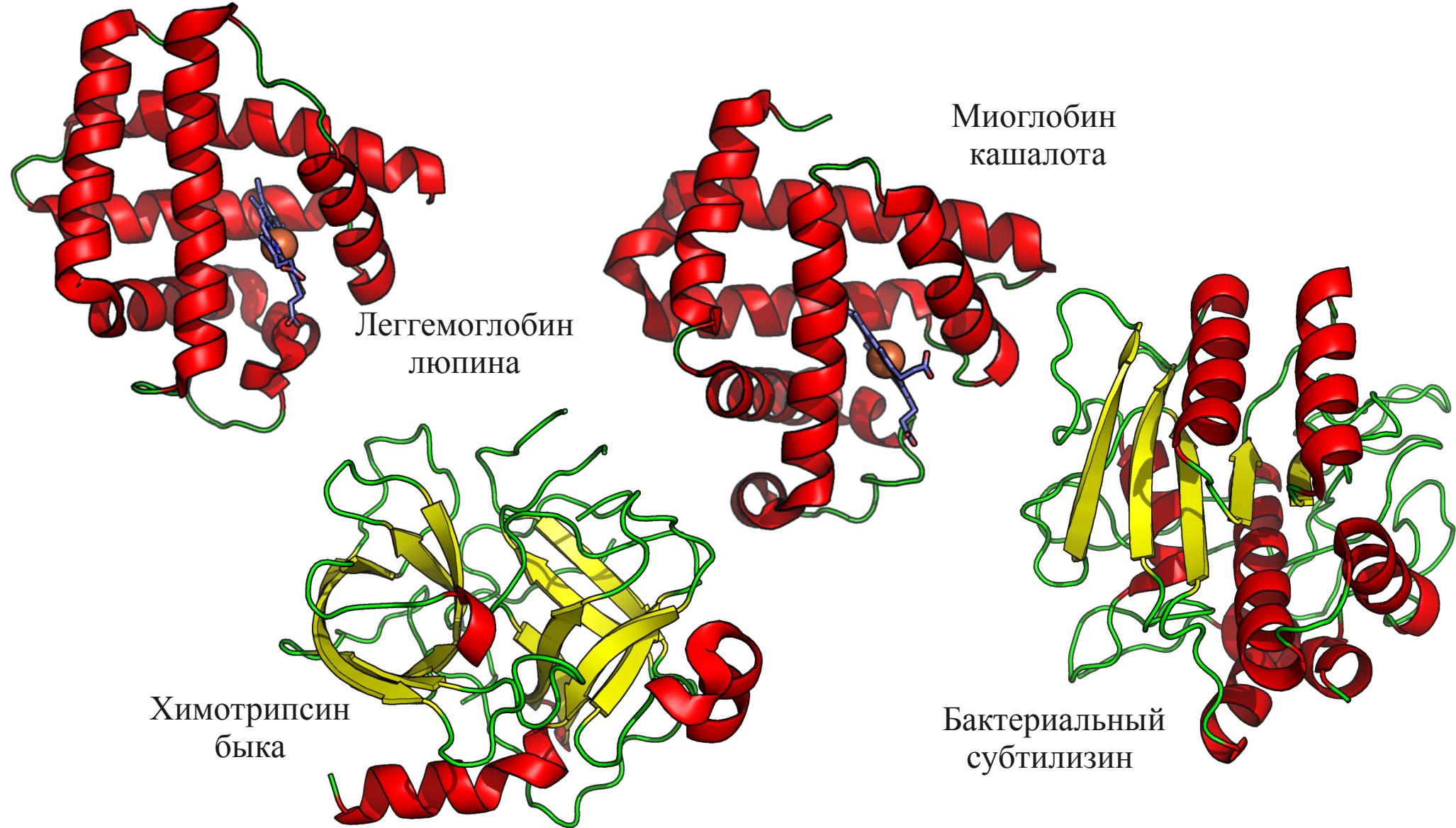
🖨️ Learn how to print this alignment in color

1	CGVPAIQPVLSGLSRIVNGEEAVPGSWFWQSILQ--TGFHFCG-C-SLTINEN	50 P00766 CTRA_BOVIN
1	-----MMRKSPNLGMLTAFMLVFTMAFSASASAQPAKNEKO	39 P00780 SUBT_BACLI
51	WVVTAAHGVITSVVVAGEFDQSSSEK-TQKLIAKVK-FNSKYNLSITINNDITLKL	108 P00766 CTRA_BOVIN
40	YIVG-FKSGVKTKSVKNDIIESGGKVKDQFFIINAKLKEALKEVKNDPDVAYE-	97 P00780 SUBT_BACLI
109	STAASESQTVSAVCLPSASDDFAAGTTCVTTGWGLTRYTNANTPDLQQASLPLLNS-T	166 P00766 CTRA_BOVIN
98	EDHVHAQALQTVPYGI-----PLIKADKV	121 P00780 SUBT_BACLI
167	NCKKYNOTKTDAMICAGASGVSSCMGDGG-----PLVCKK--NG	205 P00766 CTRA_BOVIN
122	QAGFKGANVKVAVLDTGIAASHPDLNVGGASFVAGEAYNTDGNQHGTIVACTVAALDN	181 P00780 SUBT_BACLI
206	ATLVG---IVSNGSSTCSTSTPGVYARVTAALVNVQQTAAAN-----	245 P00766 CTRA_BOVIN
182	TTGVLGWAPSISLYAVKVLNNSGSGTYSGINSGIENATTNGMDVINMSLGPGSCTAMKQ	241 P00780 SUBT_BACLI
246	AVDNAYARGVVVVAAGNSGSSGNTNTIGYPAKYDSVIAVGAVDNSNRASFSSVGAEL	245 P00766 CTRA_BOVIN
242		301 P00780 SUBT_BACLI
246	VMAPGAGVYSTYPTSTYATLNGTSMASPHVAGAAALILSKHPNLSASQVRNPLSSTAYI	245 P00766 CTRA_BOVIN
302		361 P00780 SUBT_BACLI
246		245 P00766 CTRA_BOVIN
362	GSSFYYOKGLINVEAAQ	379 P00780 SUBT_BACLI

Annotation

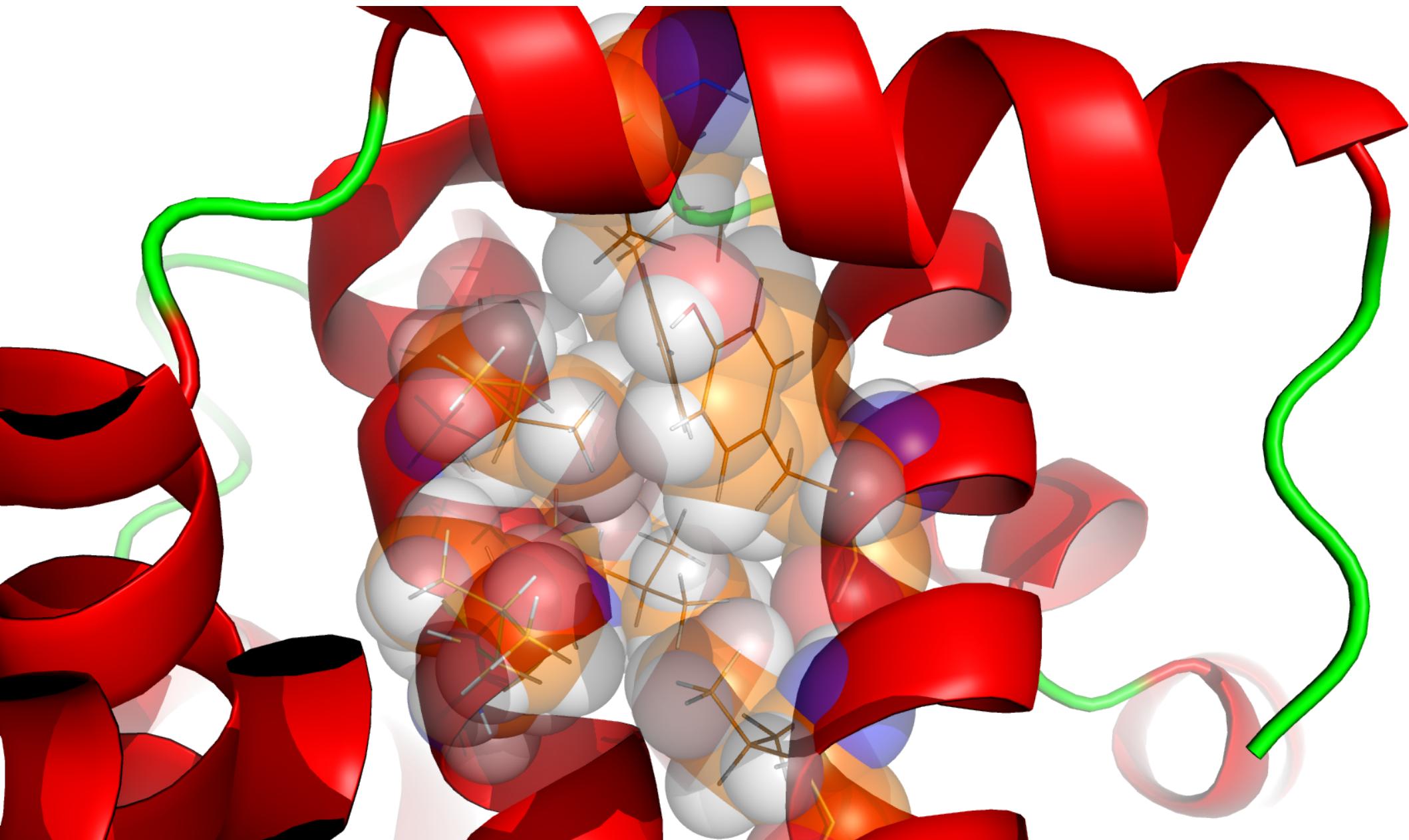
- Chain
 - Beta strand
 - Signal peptide
 - Disulfide bond
 - Propeptide
 - Helix
 - Domain
 - Metal binding
 - Active site
 - Sequence conflict
 - Turn
- Amino acid properties
- Similarity
 - Hydrophobic
 - Negative
 - Positive
 - Aliphatic
 - Tiny
 - Aromatic
 - Charged
 - Small
 - Polar
 - Big
 - Serine Threonine

Структурное сходство/различие очевидно



Трехмерная структура белка много консервативнее первичной последовательности.

Контакты между остатками



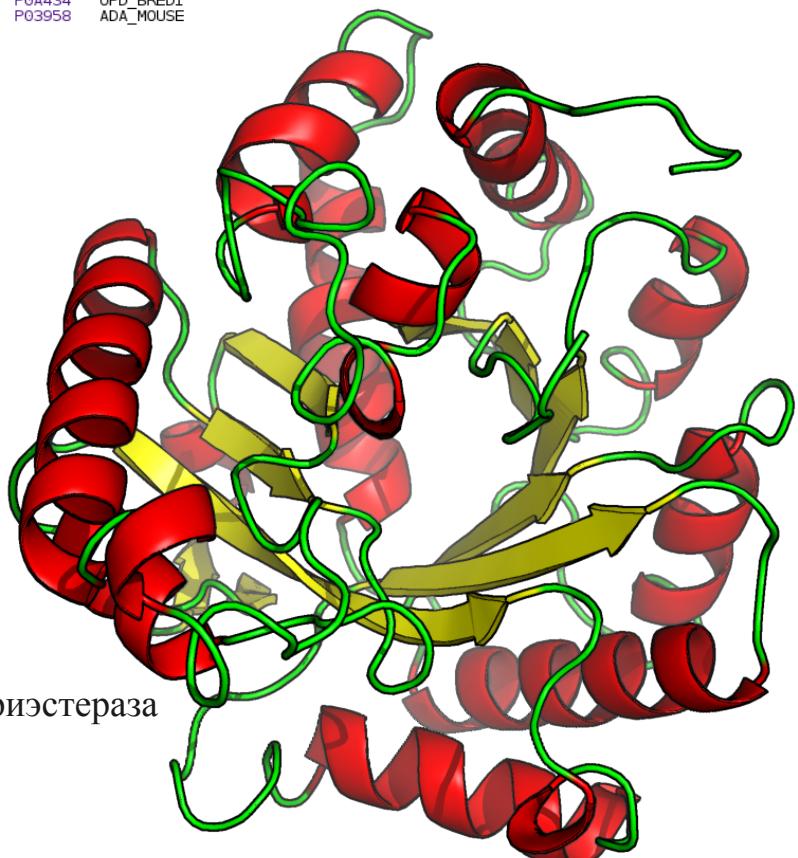
Наиболее консервативной чертой трехмерной структуры белка оказываются взаимные контакты между аминокислотами.

Дальние гомологи

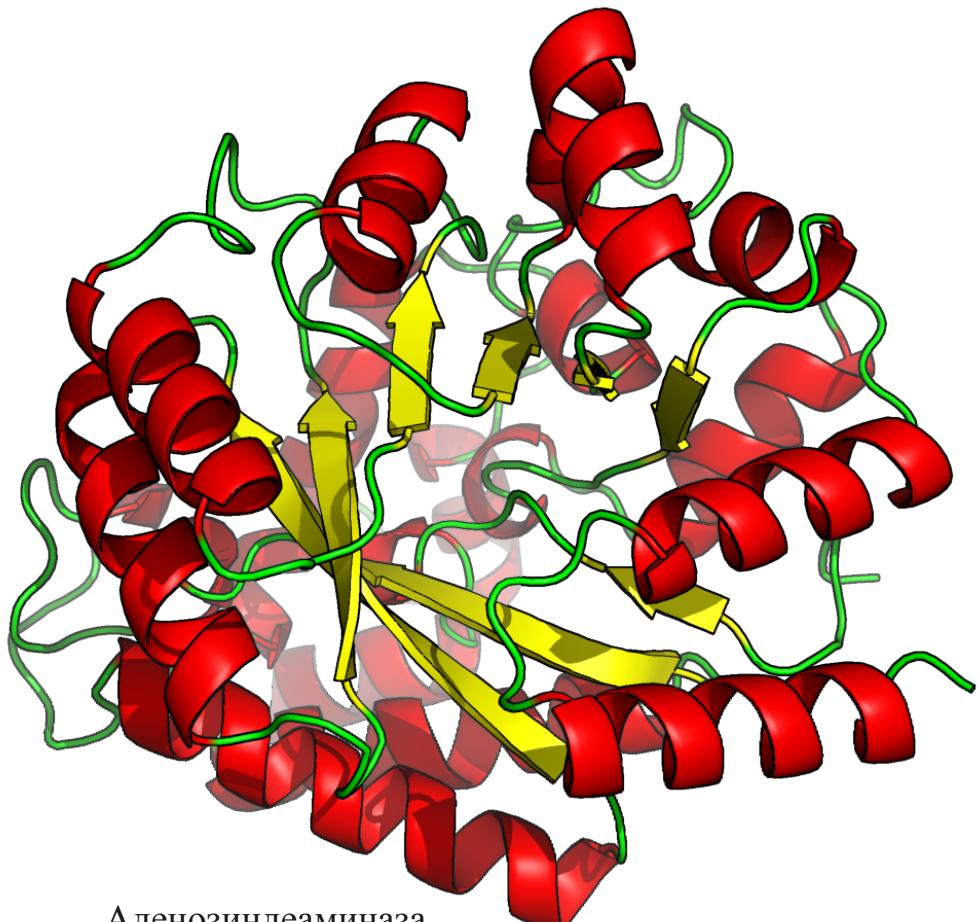
Alignment

Learn how to print this alignment in color

1	-----MOTRRVV-----LKSAAAAGT-----LLGGLAGCAGVAGSITGCDRINTVRGPITISE	48	P0A434	OPD_BREDI
1	MATQPAFNPKVELHVHLDGAIAKPEILYFGKKRGIALPA-----DTVEELRNIGMD-	53	P03958	ADA_MOUSE
49	AGFTLTTHEICGSAGFLRAW-----PEFFGSRKALAEKAVRGLRRARAAGVTRIVDVSTF	104	P0A434	OPD_BREDI
54	KPLSLPGFLAKFDYMPVIAGCREAIKIAYEFEVEMAKEGVVY-----V	98	P03958	ADA_MOUSE
105	DIGRDVSLLAEVSAADVHIVAATGLWEDPPLSNRLRSWEEITQFFLREIQYQIEEDTGIR	164	P0A434	OPD_BREDI
99	EVRYSPHLANS-K-----VDPMPPNQTEGDNTF-----DDVVDLVNQGLQEGEOQAFGK	147	P03958	ADA_MOUSE
165	AGIIKVATTGKATPPQELVLIKAAAISLATGVPVTTHTIASQPDGEQQAAIFSEQLSPS	224	P0A434	OPD_BREDI
148	VPSILCCMRH-QPSMSLEVLCKKYNQKTVVAM-----DLAG-----DETLIESG	191	P03958	ADA_MOUSE
225	RVCIGHSHDDTDD-----LSYTALALARGLIGLIDHIPHSATIGLEDNASASA	270	P0A434	OPD_BREDI
192	SLEPG-VEAYEGAVNGIHRTHVACEGVSPPEVVRPEAVDIKTERVGH-GYHTEDEALYN-	250	P03958	ADA_MOUSE
271	LLGIRSNQTRALLIKALIDQGYMKQILVSDWLF-----GFSSYTNIMDVW-----DRVNPF	322	P0A434	OPD_BREDI
251	R-LLKENMHPFEVCPWNSSYLTGAWDKPTTHAVWRKKN-----	285	P03958	ADA_MOUSE
323	DGMIFIPLRVIPF-----LREKGVPQETLAGITVTNP-----RFLSPTLRAS-----	365	P0A434	OPD_BREDI
286	DKANYSLNDDPLFKSTLDTYQMTKDDMGFTEEEKFPLNNNAKSSPEEKKELLE	345	P03958	ADA_MOUSE
366	-----RILYREYQ-----	365	P0A434	OPD_BREDI
346	-----RILYREYQ-----	352	P03958	ADA_MOUSE



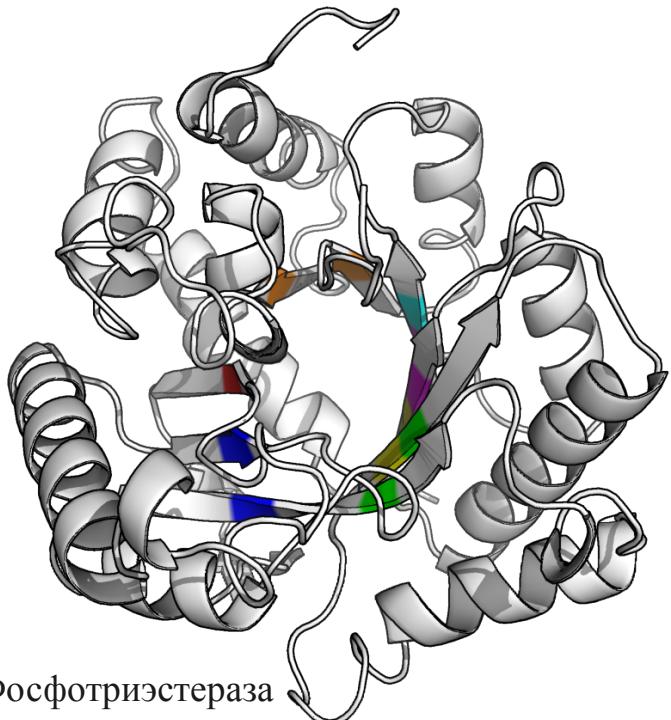
Фосфотриэстераза



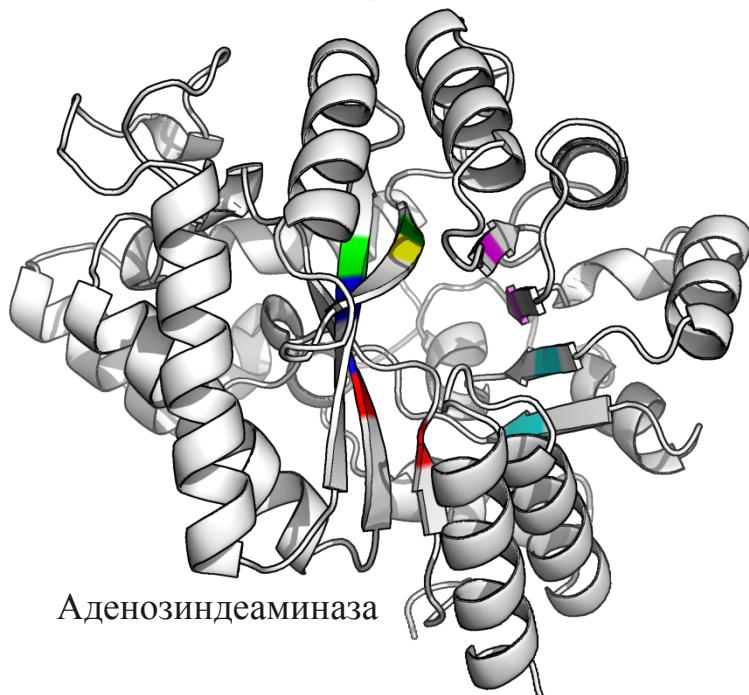
Аденозиндеаминаза

- Уровень гомологии аденоzinдеаминазы мыши и фосфотриэстеразы (органиофосфат гидролазы) бактерии *Pseudomonas diminuta* составляет 13.6%, что не позволяло установить гомологию.
- Сравнение пространственных структур показало наличие в них общего домена TIM-бочонка (в честь ТриозофосфатИзоМеразы).

Матрицы контактов

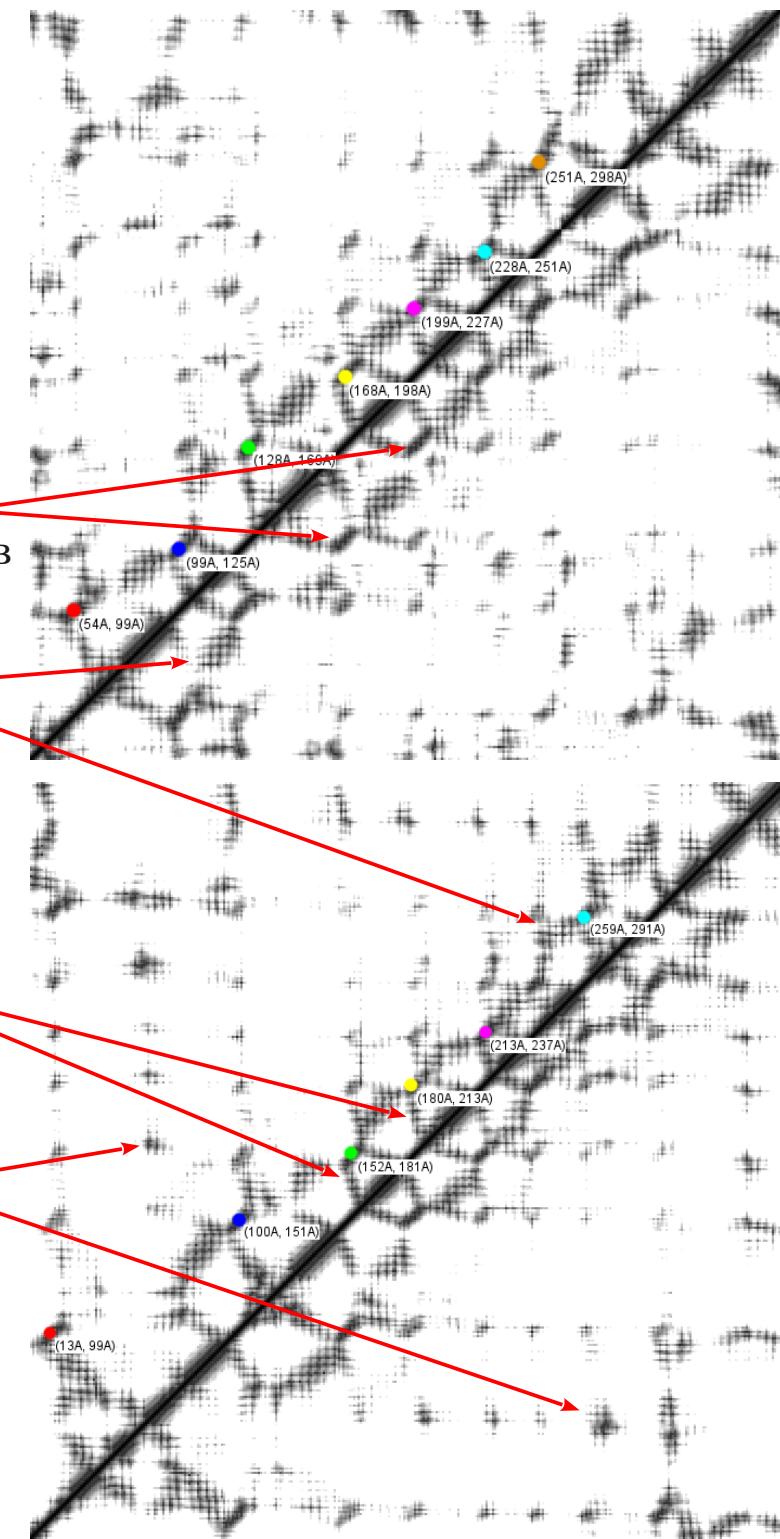


Фосфотриэстераза



Аденозиндеаминаза

- Матрицы контактов аденозиндеаминазы мыши и бактериальной фосфотриэстеразы очень похожи
- Узкие штрихи, идущие вдоль диагонали, соответствуют участкам параллельных β -слоёв
- Широкие штрихи, идущие вдоль диагонали, соответствуют участкам параллельных α -спиралей
- Широкие штрихи, идущие под углами, соответствуют областям контактов между антипараллельными α -спиралями и β -слоями
- Изолированные области сближения соответствуют точечным контактам между петлями и элементами вторичной структуры

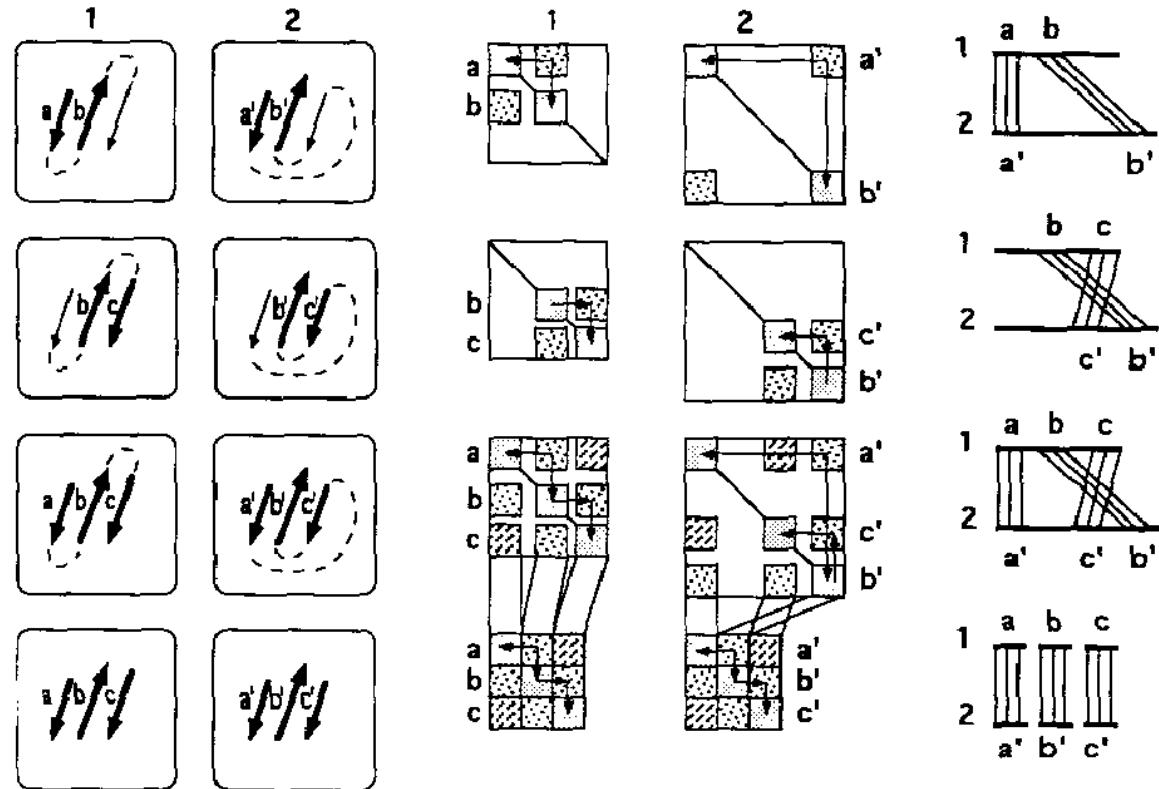


DALI — Distance matrix ALignment

- Последовательности разбивают на 6-ти аминокислотные фрагменты
- Для каждой пары фрагментов рассчитывают оценочную функцию S

$$S = \sum_{i,j}^N \theta^r - |d_{ij}^A - d_{ij}^B|$$

- Если S ниже порогового, фрагменты считаются выравненными и программа пытается их расширить
- Итогом формирования списка выравненных фрагментов является затравка, основа консервативной части структуры пары белков (фолд)
- Затравка расширяется наименее невыгодным способом с использованием оценочной функции S



3D

2D

1D

Перекрестный поиск

Цель: найти структуры, соответствующие данной последовательности, и наоборот, последовательности, соответствующие данной структуре.

Эта задача известна как задача распознавания **фолда**, то есть общей схемы третичной структуры белка.

1. 3D-профили.

- Для каждой аминокислоты анализируют окружение в базе структур или ее части
- Рассчитывают вероятности для каждой аминокислоты находиться в том или ином окружении
- Используя одномерный профиль свойств окружения для некоторого фолда можно рассчитать вероятность такой укладки тестируемой последовательности

R. Liithy, J.U. Bowie, D. Eisenberg. "Assessment of protein models with three-dimensional profiles". **1992**. Nature. 356, 83-85.

2. Протягивание.

- Тестируемую последовательность выравнивают против всех последовательностей, для которых имеется 3D структура.
- Тестируемую последовательность укладывают согласно выравниванию
- Качество полученной укладки контролируют, анализируя взаимные расстояния между различными аминокислотами.

Связь последовательности и структуры

В 1969 году С. Левинталь подсчитал, что полипептид, состоящий из 100 а.к. может пребывать в 10^{143} конформаций.

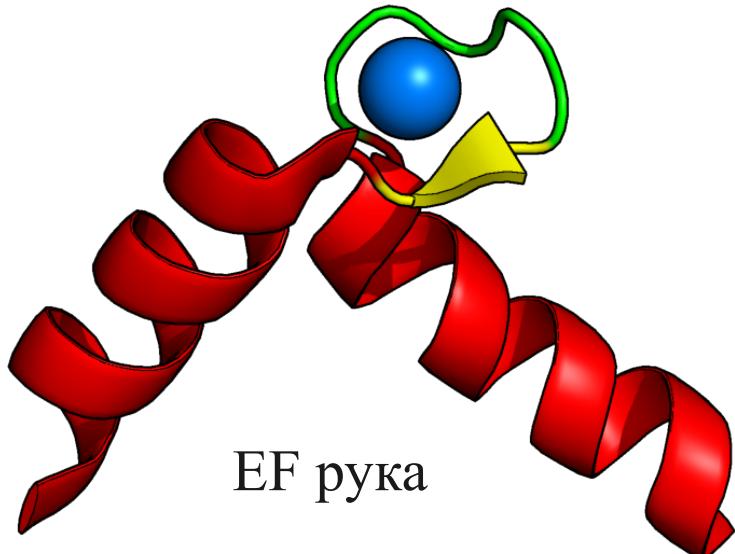
Задача предсказания трехмерной структуры белка *ab initio* не решена, однако:

- Элементы вторичной структуры могут быть предсказаны
- Вариантов взаимного расположения этих элементов уже значительно меньше.
- Помимо элементов вторичной структуры существуют более сложно организованные блоки белковой структуры.

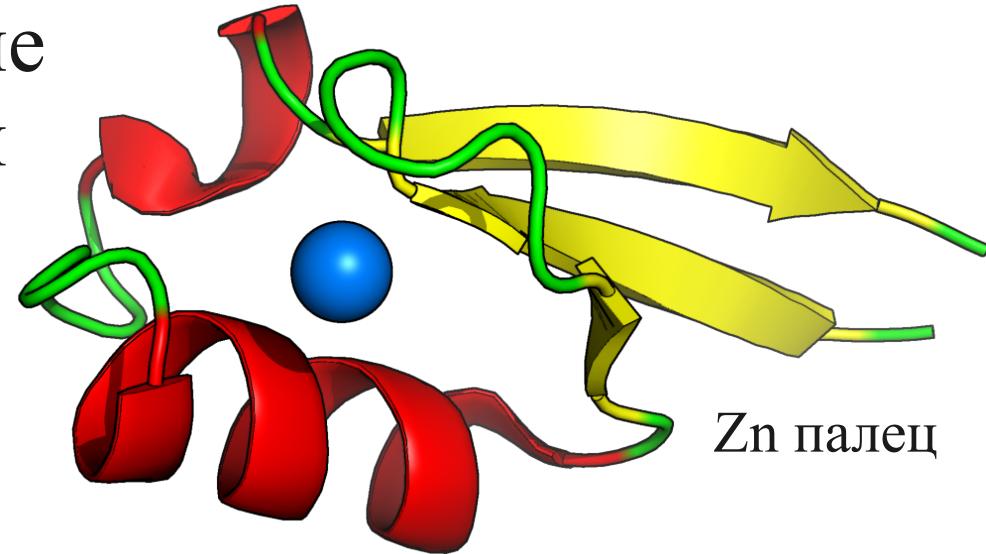


Сайрус Левинталь (1922 – 1990)

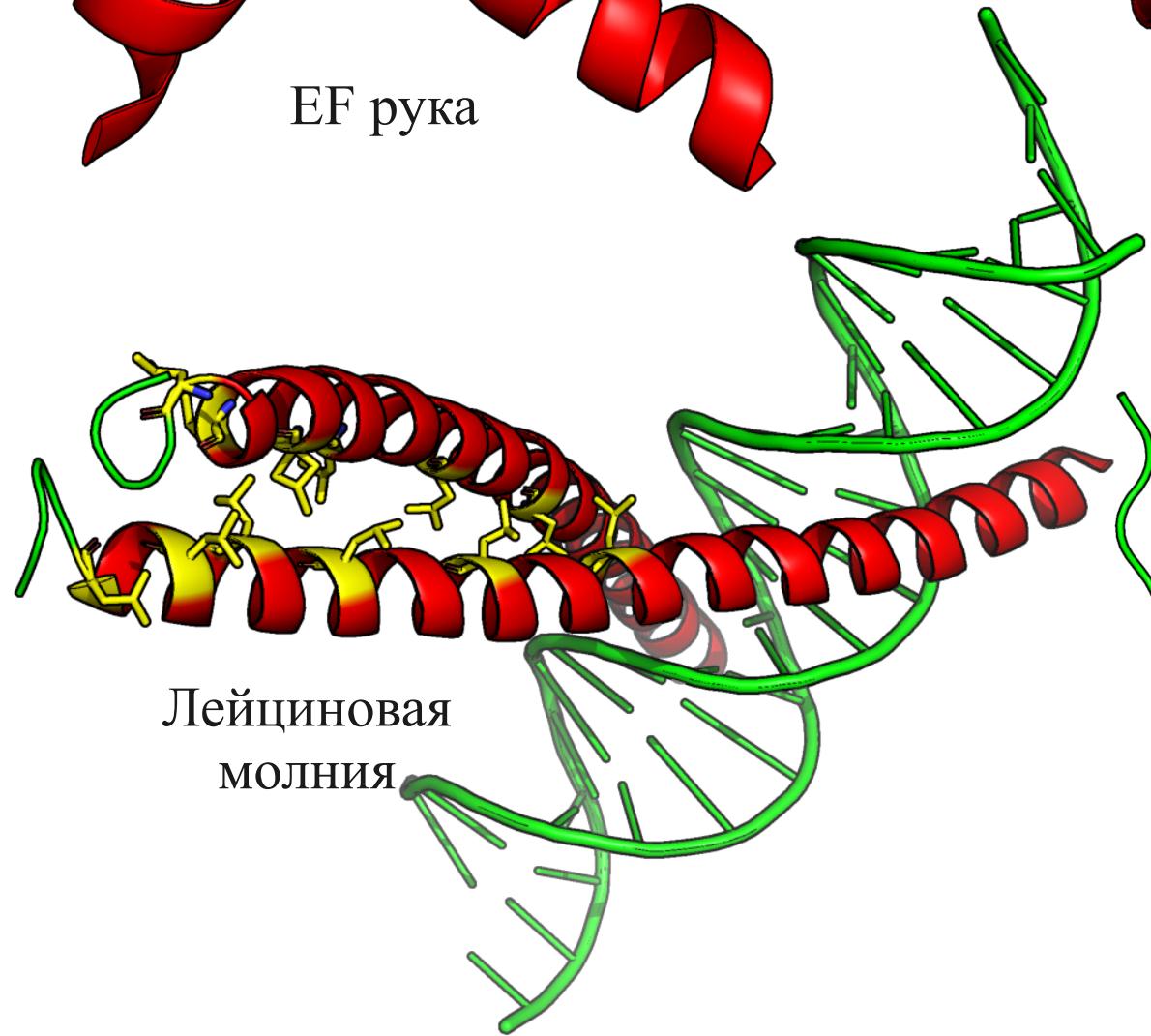
Структурные МОТИВЫ



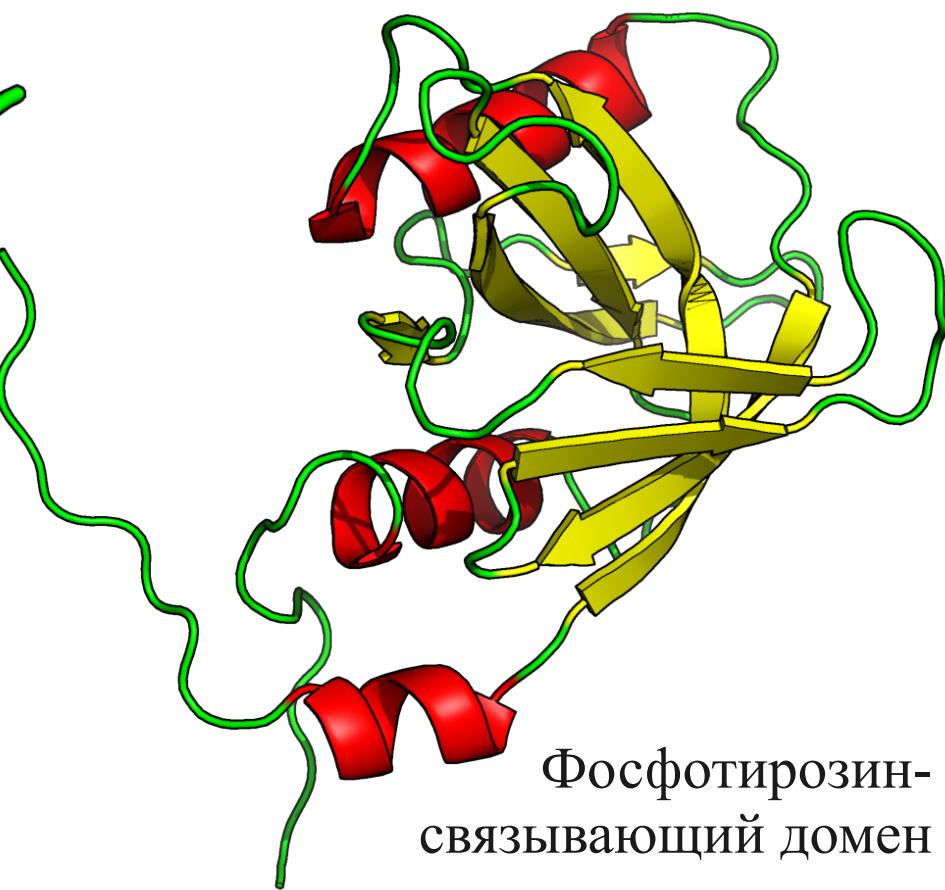
EF рука



Zn палец



Лейциновая
молния



Фосфотирозин-
связывающий домен

Предсказание вторичной структуры

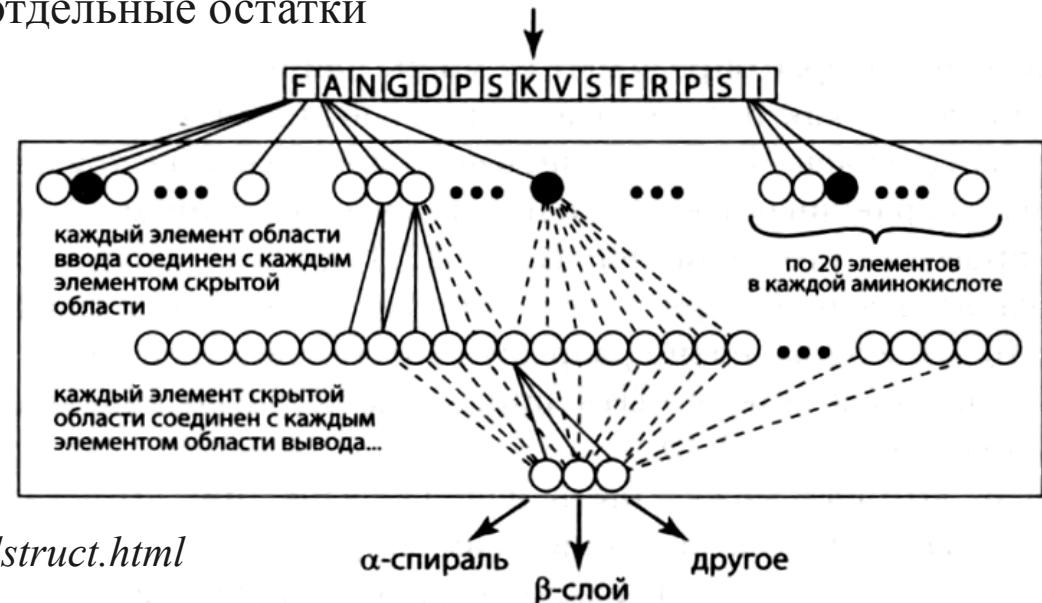
1. Метод Чоу-Фасмана

- Аминокислоты разделяют по типам вторичной структуры
- 4 из 6-ти α -спиральных остатка свидетельствуют об образовании спирали, а 3 из 5-ти β -слойных - образование β -слоя
- Точность метода не превышает 60%

Chou P.Y., Fasman G.D. "Prediction of protein conformation". Biochemistry. 1974. 13(2): 222–245.

2. Метод GOR и нейронные сети

- Используют условные вероятности давать те или иные типы вторичной структуры.
- Анализируют широкую «рамку», а не отдельные остатки
- Точность составляет ок. 90%

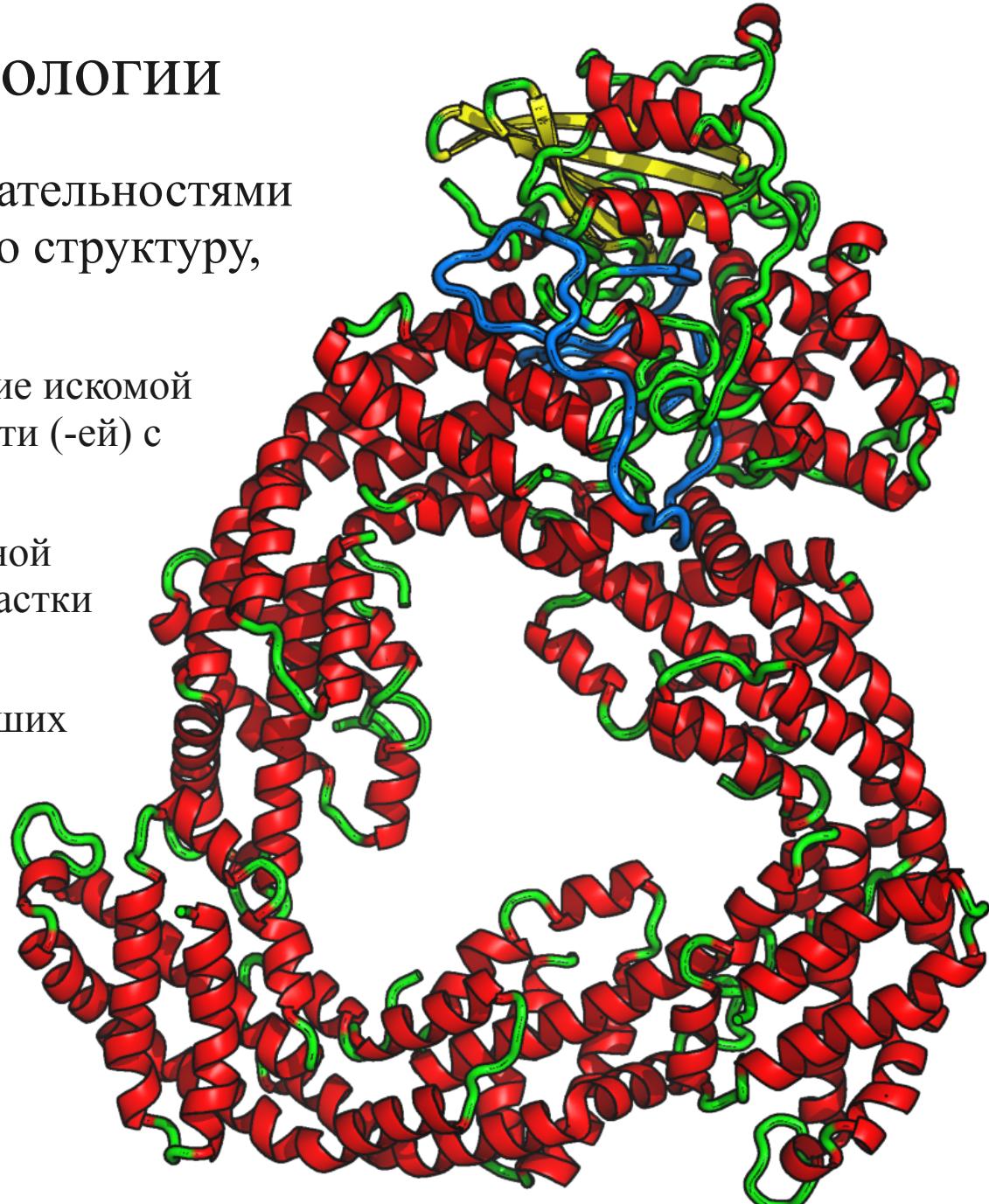


http://www.geneinfinity.org/sp/sp_proteinsecondstruct.html

Моделирование по гомологии

Если белки со схожими последовательностями имеют схожую пространственную структуру, тогда:

- Построим множественное выравнивание искомой последовательности и последовательности (-ей) с известной структурой
- Наложим совпадающие участки основной цепи искомой последовательности на участки цепи известной структуры
- Заменим боковые радикалы мутировавших остатков, сохраняя по возможности их конформации
- Устраним наложения атомов и другие шероховатости
- Оптимизируем геометрию системы, минимизировав ее потенциальную энергию



*MODELLER (<http://salilab.org/modeller/>),
SWISS_MODEL (<http://swissmodel.expasy.org/>)*

Спасибо за внимание!!!