ВВЕДЕНИЕ В БИОИНФОРМАТИКУ

Лекция №19

Структурная биоинформатика

Новоселецкий Валерий Николаевич
к.ф.-м.н., доц. каф. биоинженерии
valery.novoseletsky@yandex.ru

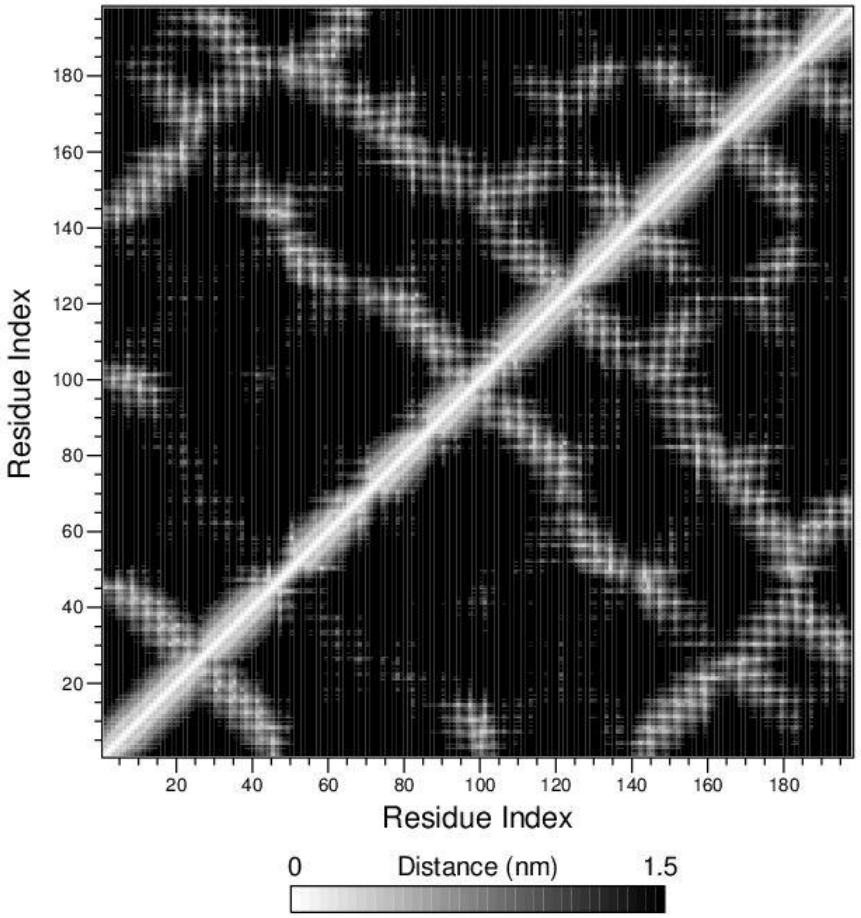
Сайт курса http://intbio.org/bioinf2020-2021

Сравнение структур

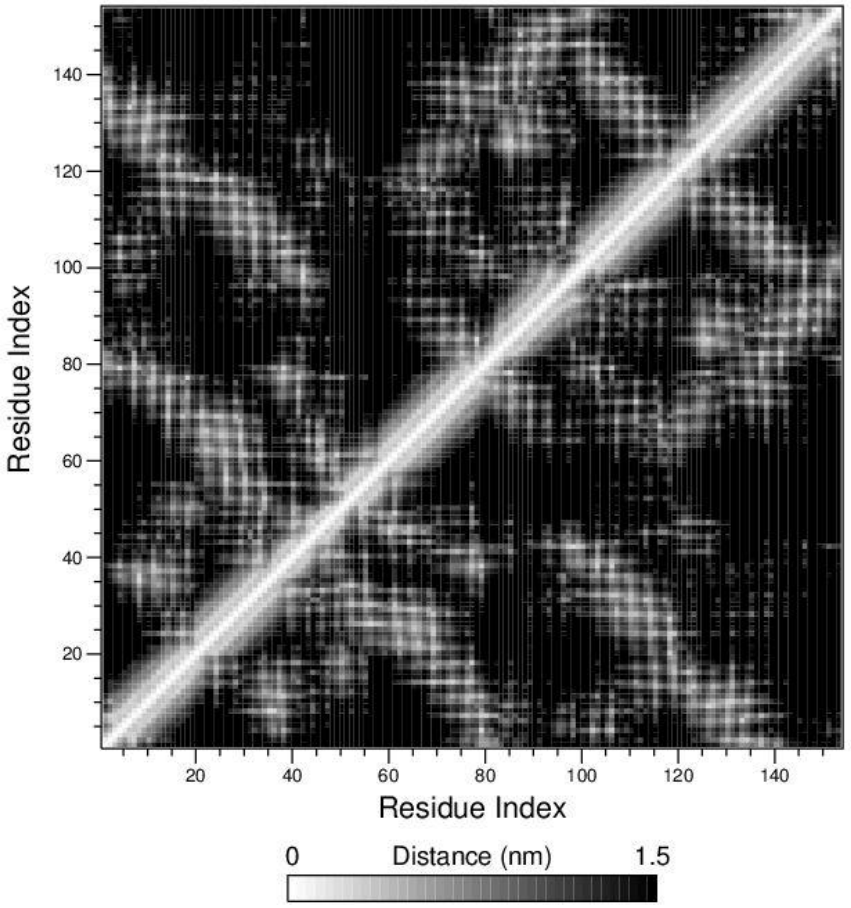
Если структуры схожи, то сохраняются паттерны контактов между остатками

$\Rightarrow$ анализируя матрицы расстояний (Ca-Ca), можно распознать схожие структуры

colicin



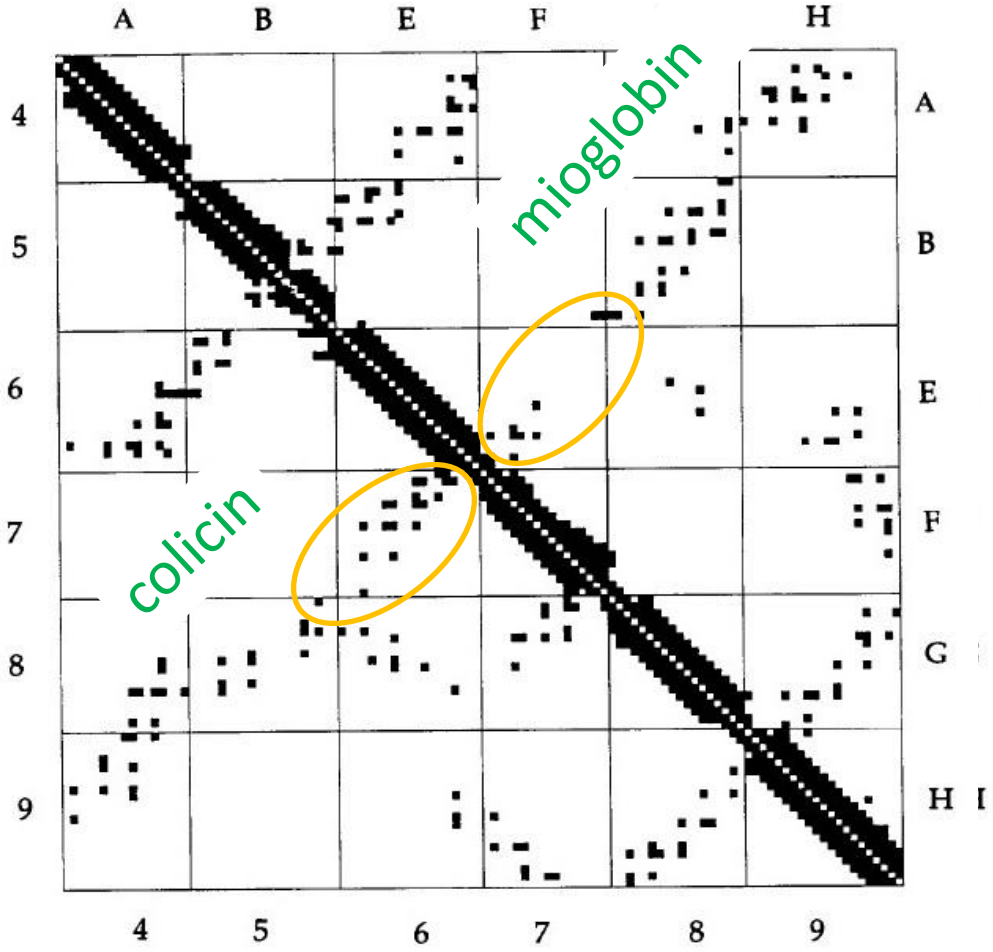
mioglobin



Сравнение структур

Выравнивание матриц расстояний – программа **DALI** (**D**istance-matrix **ALI**ngment)

(Holm & Sander, 1993)

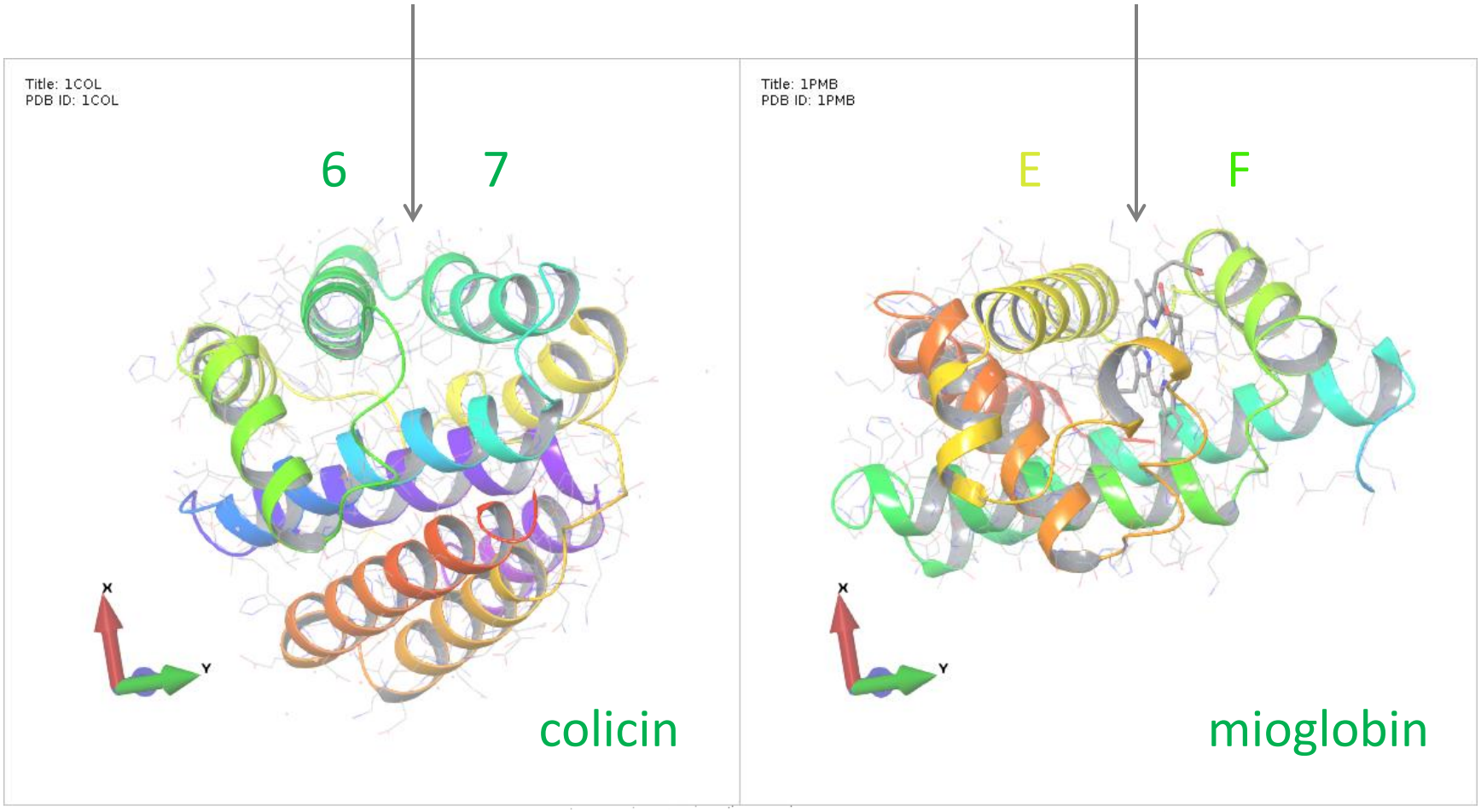


Особенности: свободная топология обнаруженных сходств в структуре, в т.ч. и «обратных» фрагментов.

Сравнение структур

Есть контакт между спиралями

Гем; нет контакта между спиралями



Сравнение структур

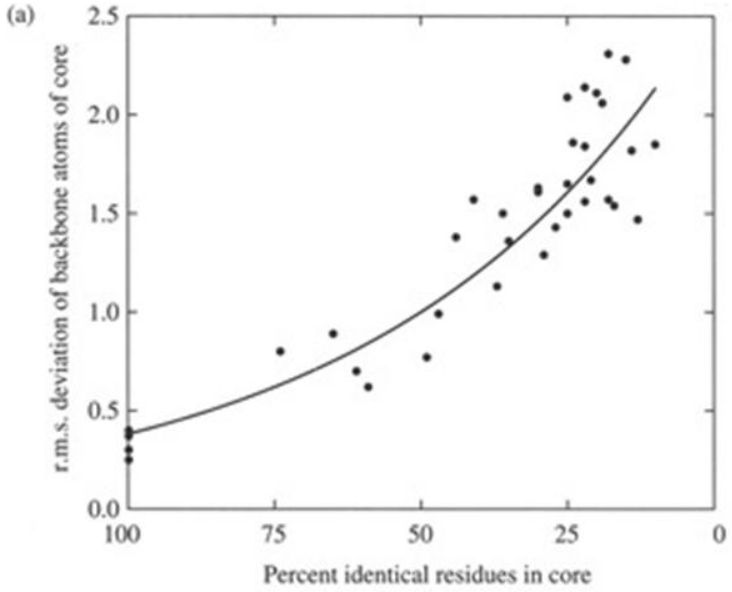
Structural alignment:

```
1COL SSA    5 CCCHHHHHHHHHHHHHHHHHHHHHHCHHHHHHHHHHHHHHHHCCCCCCCCCHHHHHHHHHH
1PMB SSA      ------------------------------------------------------------
1COL        5 AKDERELLEKTSELIAGMGDKIGEHLGDKYKAIAKDIADNIKNFQGKTIRSFDDAMASLN
1PMB          ------------------------------------------------------------

1COL SSA   65 HHHHCCCCCCCHHHHHHHHHHHHHH--CHHHHHHH-HHH--H-CCCCC-C-C--------
1PMB SSA    1 -------C-CHHHHHHHHHHHHHHHHHCHHHHHHHHHHHHHHHCHHHHHHHCCCCCCCCCH
1COL       65 KITANPAMKINKADRDALVNAWKHV--DAQDMANK-LGN--L-SKAFK-V-A--------
1PMB        1 -------G-LSDGEWQLVLNVWGKVEADVAGHGQEVLIRLFKGHPETLEKFDKFKHLKSE
                     :   :   ::* * :*  *      : *

1COL SSA  119 -------CHHHHHHHHHHHHHHHHHH-CCHH-HHHHHHHHHHHHCCCH-HHHHHHHHHHH
1PMB SSA   53 HHHHHHCHHHHHHHHHHHHHHHHHHCCCCHHHHHHHHHHHHHH-HC-CCC-HHH-HHHHH
1COL      119 -------DVVMKVEKVREKSIEGYET-GNWG-PLMLEVESWVLSGIAS-SVALGIFSATL
1PMB       53 DEMKASEDLKKHGNTVLTALGGILKKKGHHEAELTPLAQSHAT-KH-KIP-VKY-LEFIS
                    *:  :   *        :  *     *     *

1COL SSA  179 HHHHHHH-------CCCHHHHHHHHHHHHHHHHHH--CH-----HHHHHHHHHCCC
1PMB SSA  109 HHHHHHHHHHCCCCC-CHHHHHHHHHHHHHHHHHHHHHHHHCCCC-----------
1COL      179 GAYALSL-------GVPAIAVGIAGILLAAVVGALI--DD-----KFADALNNEIIR
1PMB      109 EAIIQVLQSKHPGDF-GADAQGAMSKALELFRNDMAAKYKELGFQG-----------
               *    *          * * *     *  :   :    :
```

RMSD: 3.827 Angstrom

Эволюция белковых структур

Вариации, встречающиеся в семействах гомологичных белков с одинаковой функцией, показывают, как структура приспосабливается к изменениям в последовательности: **структура устойчива к мутациям**.



Свободно могут мутировать участки на поверхности белка, не влияющие на функцию. В частности, внешние петли легко адаптируются к изменению количества остатков, в то время как мутации, изменяющие число внутренних остатков, приводят к изменению взаимной ориентации спиралей и листов, но не их конформации.

Эволюция белковых структур

Миоглобин кашалота (1mbn, 1969)

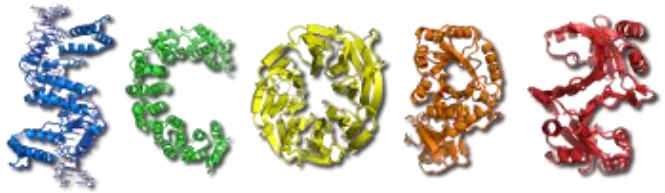
Леггемоглобин люпина (1gdj, 1995) (ИК РАН)

Эволюция белковых структур

миоглобин кашалота и леггемоглобин люпина

Идентичность 18%*, но родство подтверждается сходством структуры и функции

Классификация структур белков. SCOP

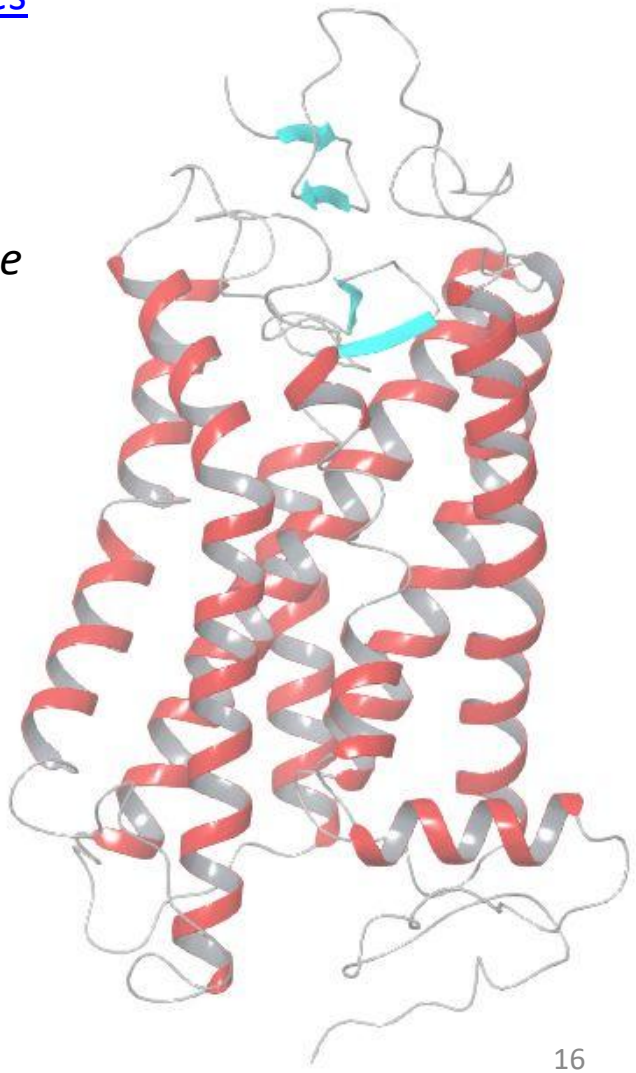


SCOP (**S**tructural **c**lassification **o**f **p**roteins) (Murzin et al., 1994 - 2009) – организация структур для отображения их эволюционного происхождения и структурного сходства. Основные уровни организации:

- **Классы**: α, β, α+β, α/β и разнообразные «малые белки», зачастую имеющие слабую вторичную структуру.
- **Фолды**: надсемейства, имеющие общую топологию укладки (одинаковые элементы вторичной структуры с одинаковым чередованием и одинаковым расположением, по крайней мере в «ядре»), наличие эволюционного предка маловероятно;
- **Надсемейства**: вероятно эволюционно близкие белки с низкой идентичностью, но функции и структуры которых позволяют предположить наличие общего предка (например, актин, АТФ-азный домен HSP и гексакиназы);
- **Семейства**: очевидно эволюционно близкие белки с идентичностью остатков, как правило, не менее 30% (глобины – 15%);

Классификация структур белков. SCOP

- **Root:** scop
- **Class:** Membrane and cell surface proteins and peptides
  Does not include proteins in the immune system
- **Fold:** Family A G protein-coupled receptor-like
  core: up-and-down bundle of seven transmembrane helices tilted 20 degrees with respect to the plane of the membrane
- **Superfamily:** Family A G protein-coupled receptor-like
- **Family:** Rhodopsin-like
  Individual TM segments have a number of kinks and distortions



http://scop2.mrc-lmb.cam.ac.uk/

Классификация структур белков. CATH

CATH (Orengo et al., 1997) – полуавтоматическая иерархическая классификация белковых доменов. Основные уровни организации:

Class – эквивалентно уровню «класс» в SCOP

Architecture – эквивалентно уровню «фолд» в SCOP

Topology – нечеткий уровень, объединяющий фолды с характерными особенностями

Homologous superfamily – эквивалентно уровню «надсемейство» в SCOP

CATH Superfamily 1.20.1070.10

Rhopdopsin 7-helix transmembrane proteins

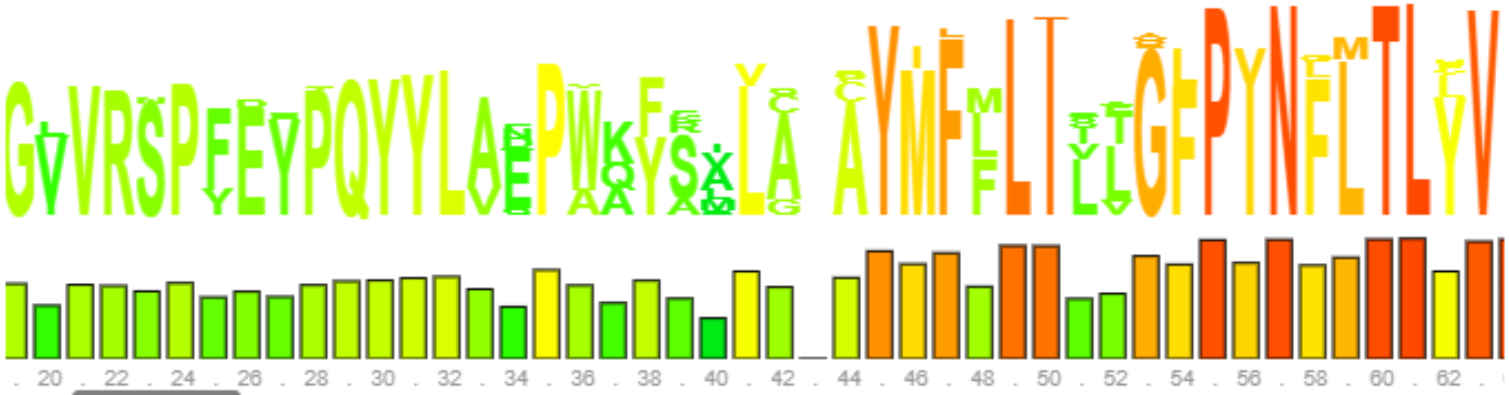
Хорошо заметно «ядро»

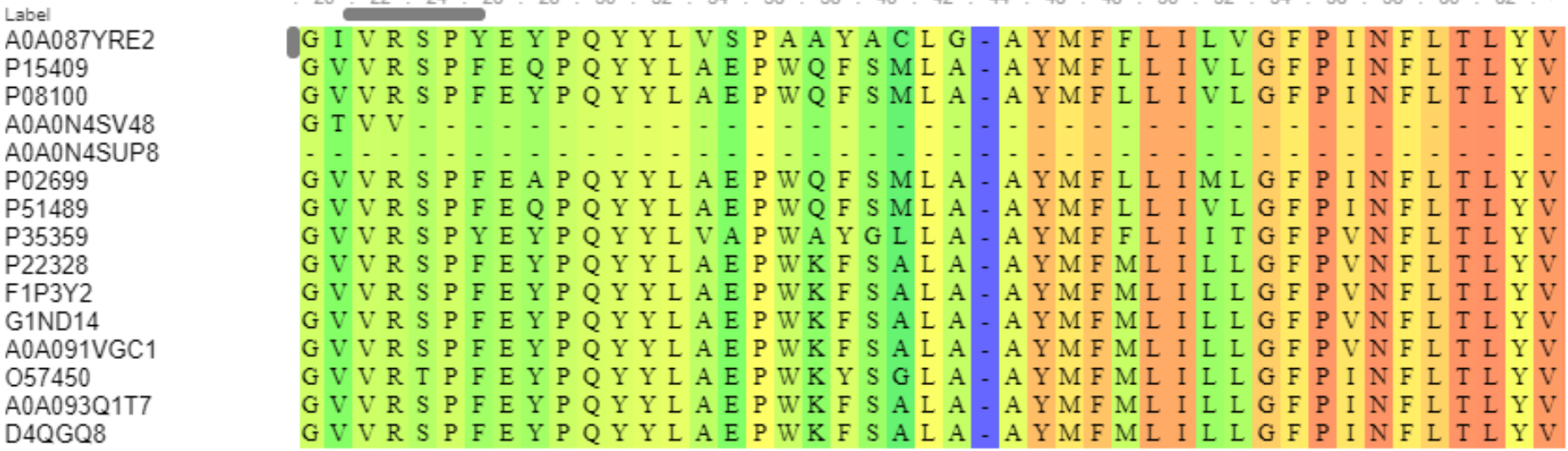
http://www.cathdb.info

Классификация структур белков. CATH

CATH Superfamily 1.20.1070.10

Rhodopsin 7-helix transmembrane proteins

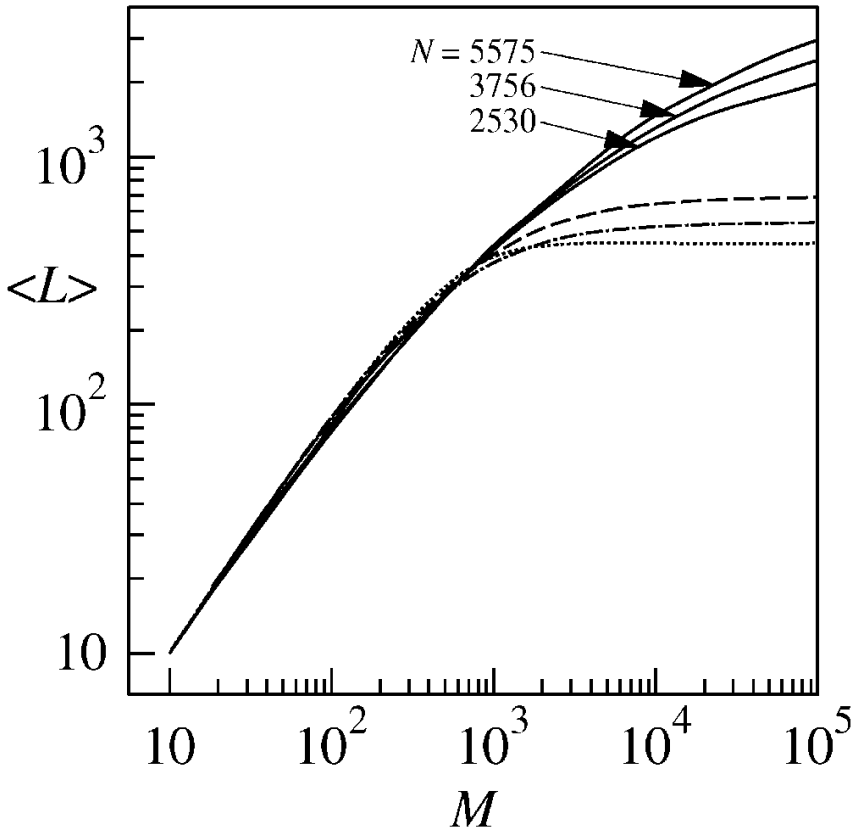




Классификация структур белков

Estimating the total number of protein folds (1999):

"Our results suggest that there are approximately 4,000 possible folds...



There has been interest in generating a comprehensive set of all protein folds.

… according to the most optimized model, although we have only observed 375 out of the 3,756 possible folds, this set still includes the structures of 70% of all protein families, even if there were an infinite number of such families. A catalog of only 930 folds would encompass approximately 90% such families."

The expected total number of observed folds, <*L*>, computed using Eq. (8) as a function of the number of protein families of known structure, *M*.

Структурная геномика

Цель: получение максимального количества разнообразных типов укладки (фолдов) белковых структур.

Методы: **РСА и спектроскопия ЯМР.**

Сроки: 2000 - 2015

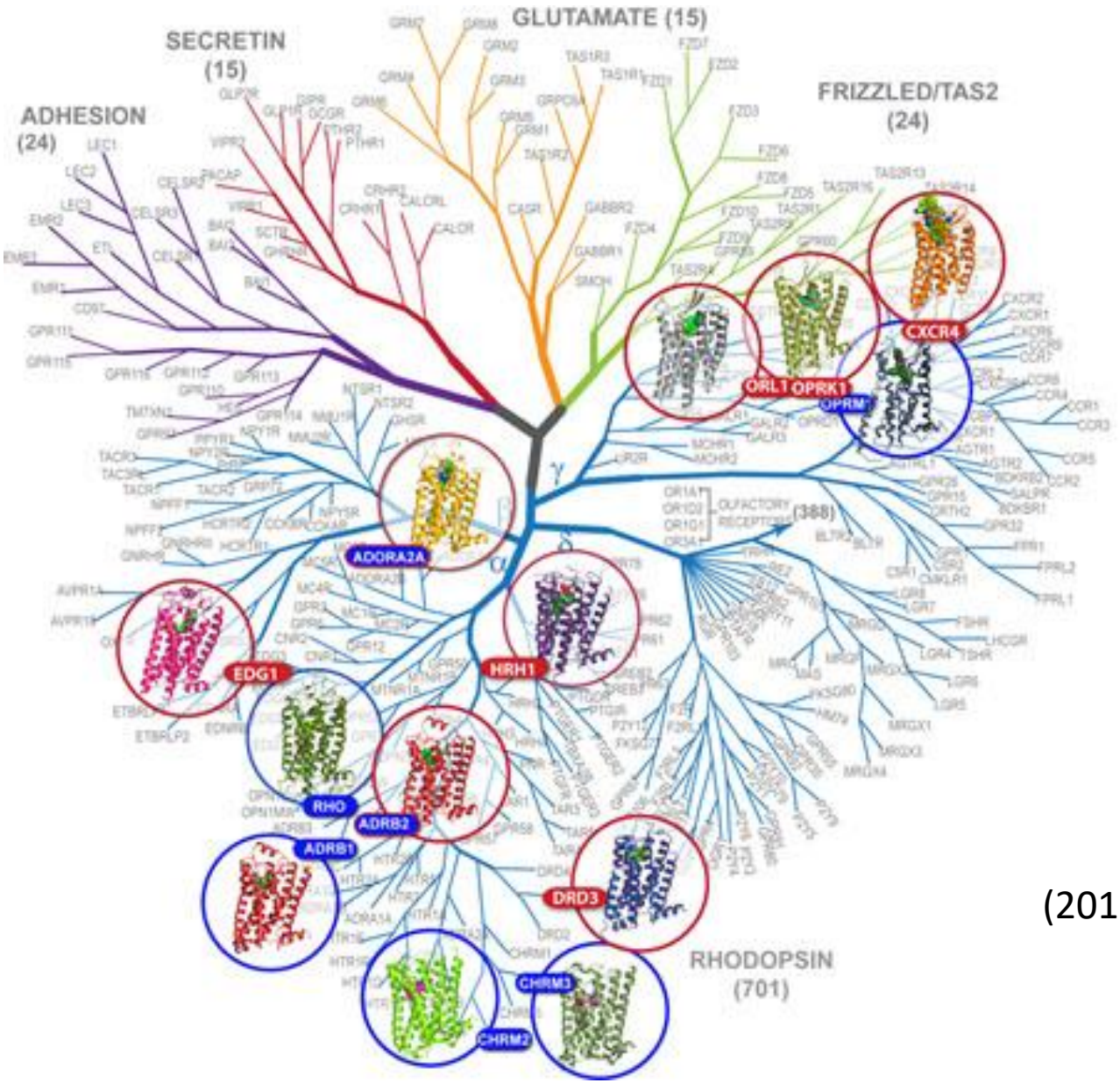
Правила отбора мишеней:

1) **Новизна**, т.е. в идеале последовательность не должна иметь сходства с белками с уже известной структурой;
2) **Актуальность**, т.е. наличие перспективы практического использования полученной структуры, а не только академический интерес;
3) **Удобство в работе**, т.е желательно, чтобы белки были растворимыми, имели повышенное содержание метионина (для решения фазовой проблемы в РСА) и т.п.

- поиск «под фонарем».

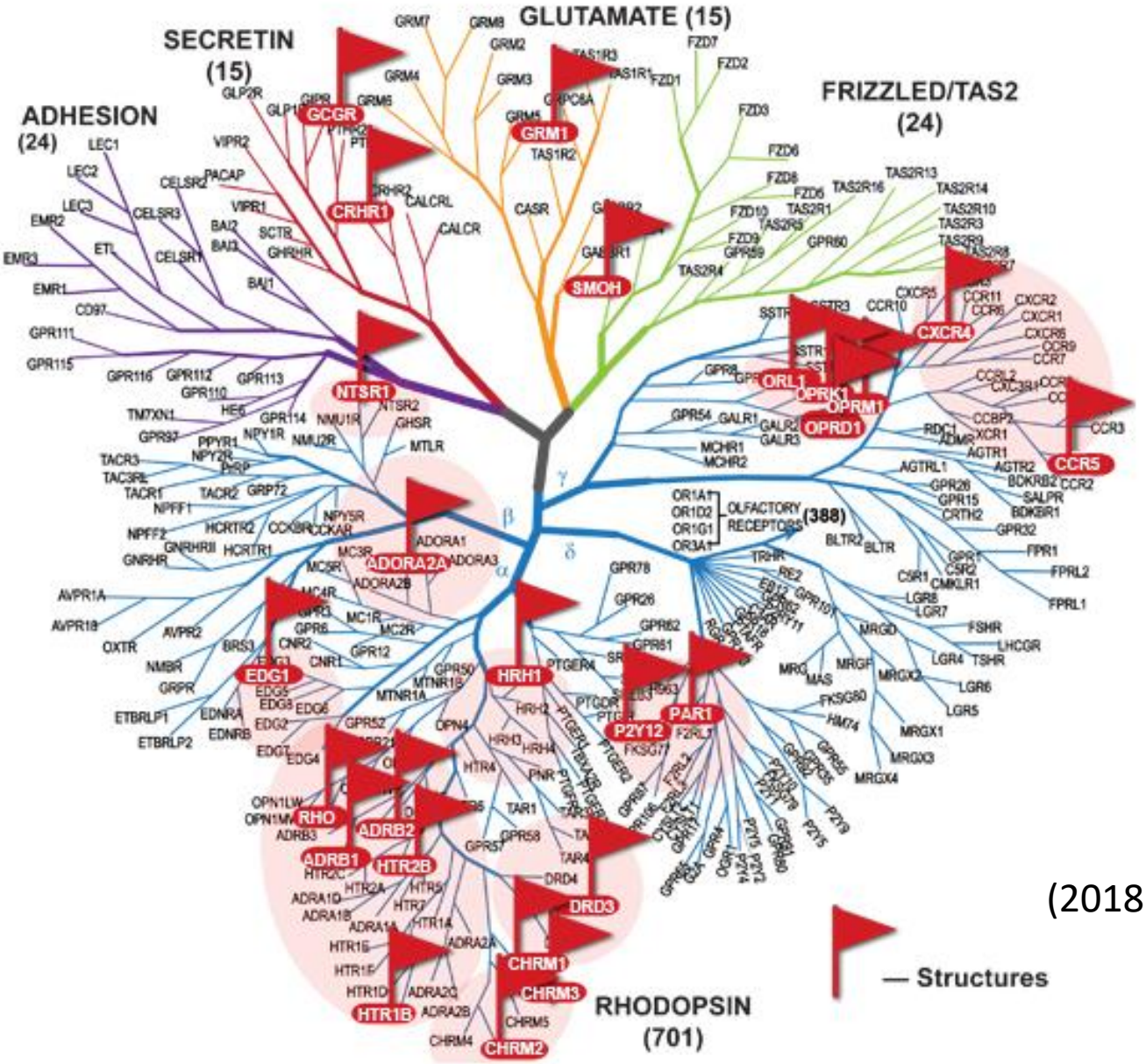
Функции расшифрованных структур зачастую были неизвестны и становились предметом для отдельных исследований

Структурная геномика. GPCR



(2012)

Структурная геномика. GPCR



(2018)

Поиск белков со схожей структурой

Protein Data Bank in Europe

Bringing Structure to Biology

PDBeFold

Share Feedback

Structure Alignment Results.

Query: pdb entry 1u19

CRYSTAL STRUCTURE OF BOVINE RHODOPSIN AT 2.2 ANGSTROMS RESOLUTION

Examined 1 entry, (1 chain). **Displaying Matches 1-2 of 2.**

Back to query Sort by Q-score

##	Scoring			RMSD	N_{align}	N_g	$\%_{seq}$	Query			Target (PDB entry)				
	Q	P	Z					Ch	N_{res}	$\%_{sse}$	Match	$\%_{sse}$	N_{res}	×	Title
1	0.23	1.4	7.1	2.20	235	11	22	A	351	55	2rh1:A	27	443	☑	HIGH RESOLUTION CRYSTAL STRUCTURE OF HUMAN B2-ADRENERGIC G PROTEIN- COUPLED RECEPTOR.
2	0.23	1.3	7.1	2.23	236	9	21	B	351	73	2rh1:A	36	443	☑	HIGH RESOLUTION CRYSTAL STRUCTURE OF HUMAN B2-ADRENERGIC G PROTEIN- COUPLED RECEPTOR.

Examined 1 entry, (1 chain). **Displaying Matches 1-2 of 2.**

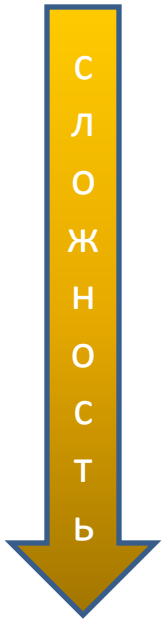
Back to query Sort by Q-score

Предсказание структуры белков

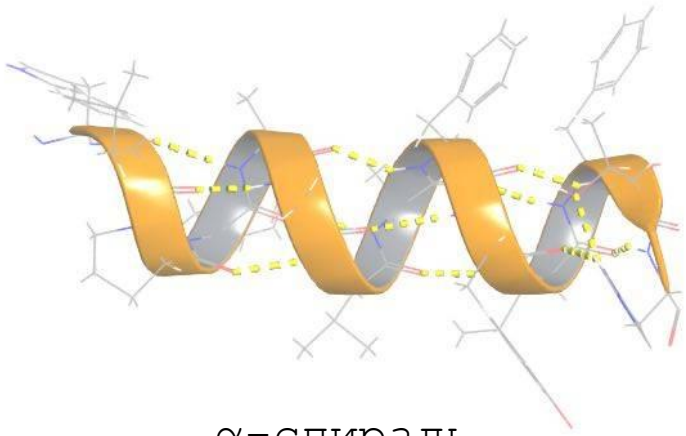
Сворачивание белка в уникальную конформацию наводит на мысль об алгоритме формирования структуры белка по его последовательности, но доказательством полноты и правильности нашего понимания могла бы стать его **реализация в виде компьютерной программы**...

Методы предсказания структуры по последовательности:

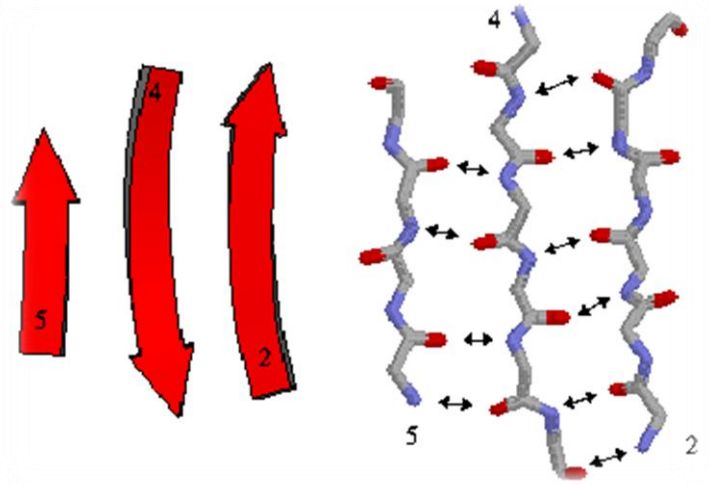
- Предсказание вторичной структуры;
- Распознавание топологии;
- Моделирование по гомологии;
- Распознавание типов укладки (по известной библиотеке фолдов);
- Априорное предсказание новых типов укладки.



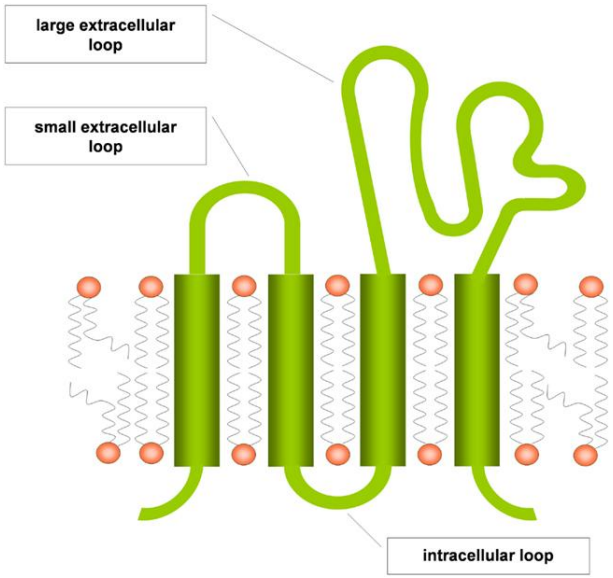
Типы вторичной структуры белков



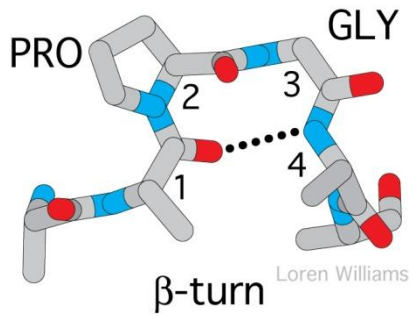
α-спираль



β-лист, состоящий из β-тяжей



петля



поворот

Предсказание вторичной структуры

В настоящее время предсказание может быть выполнено с относительно высокой точностью - около 80% элементов вторичной структуры выявляется правильно.

```
                      10        20        30        40        50
AA sequence    ALVEDPPLKVSEGGLIREGYDPDLDALRAAHREGVAYFLELEERERERTG
Prediction     HH----------EEE------HHHHHHHHHH-HHHHHHHHHHHHHHH
Experiment     -E----------E-----HHHHHHHHHHHHHHHHHHHHHHHHHHHH-

                      60        70        80        90        100
AA sequence    IPTLKVGYMAVFGYYLEVTRPYYERVPKEYRPVQTLKDRQRYTLPEMKEK
Prediction     --EEEEEEEEEEEEEEE----------EEEEEEEE--EEEE-HHHHHH
Experiment     ----EEEEE---EEEEEEEHHHHHH-----EEEEE---EEEEE-HHHHHH

                      110       120
AA sequence    EREVYRLEALIRRREEEVFLEVRERAKRQ
Prediction     HHHHHHHHHHHHHHHHHHHHHHHHHHHH
Experiment     HHHHHHHHHHHHHHHHHHHHHHHHHHH--
```

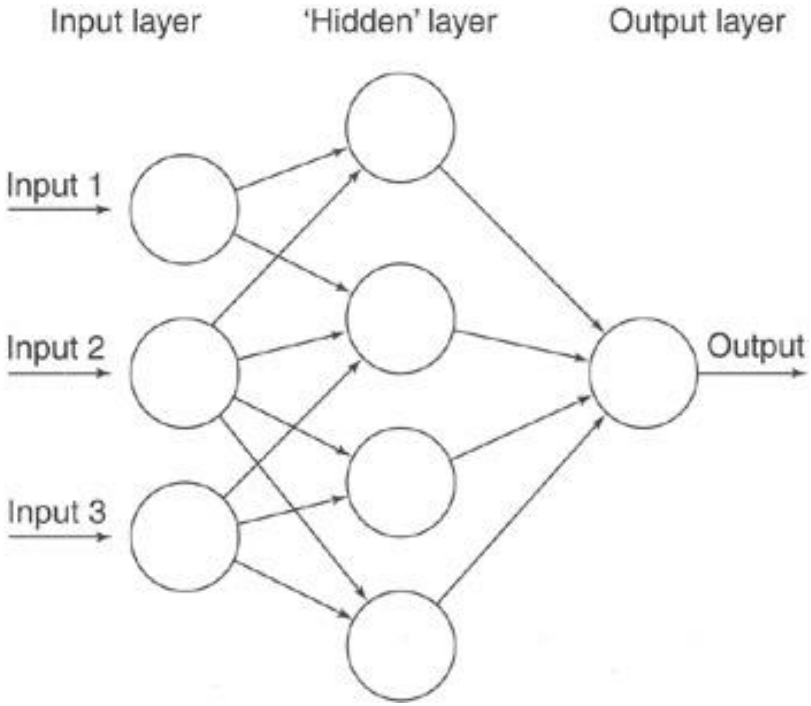
Наиболее мощные методы предсказания вторичной структуры основаны на нейронных сетях.

Нейронные сети

Искусственные нейронные сети (ИНС) — математические модели, а также их программные или аппаратные реализации, построенные по принципу организации и функционирования сетей нервных клеток живого организма.

В вычислительной схеме одиночный нейрон является вершиной графа с несколькими входящими ребрами и одним исходящим.

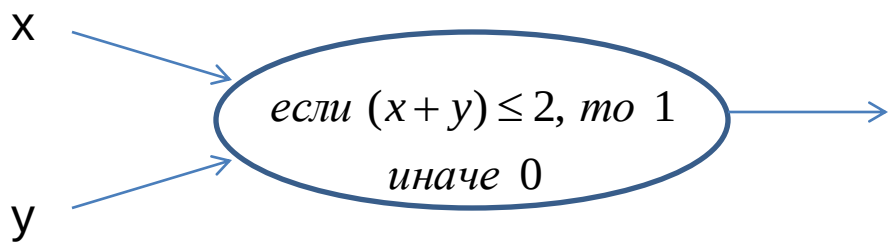
Для формирования сети необходимо соединить выходы одних нейронов со входами других.

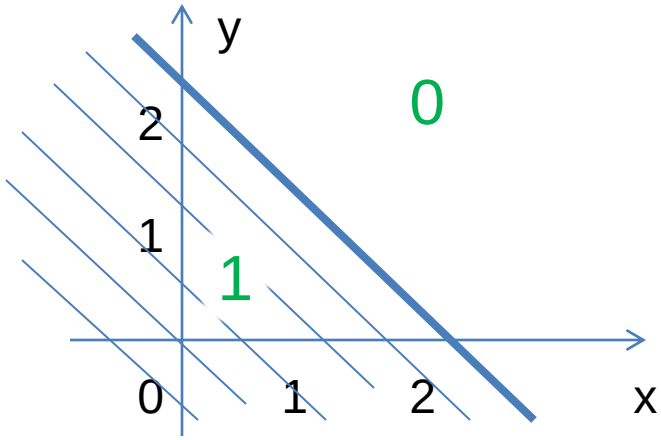


Некоторые нейроны содержат входы для всей сети, некоторые – выходы наружу, а некоторые с внешним миром не связаны (скрытые нейроны).

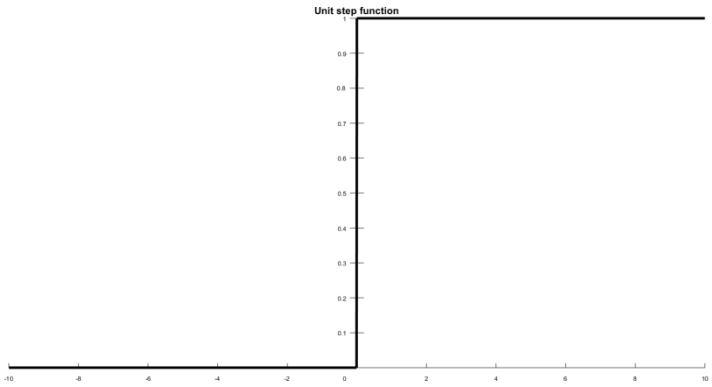
Нейронные сети. Геометрическая интерпретация

Если интерпретировать пару чисел **(x, y)** на входе как точку на плоскости, то данный нейрон принимает решение, на какой стороне от линии находится вход.



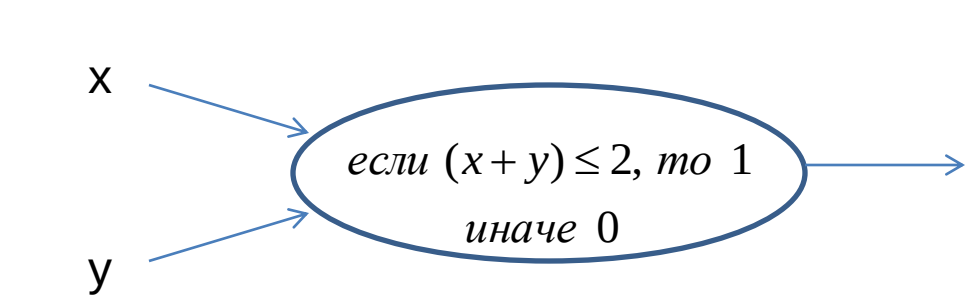


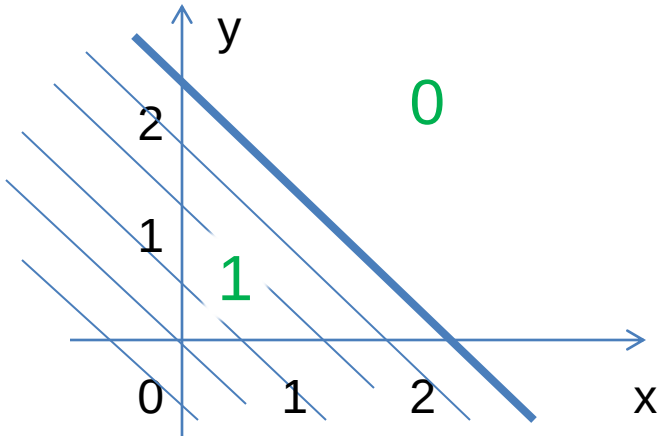
Ступенчатая функция активации

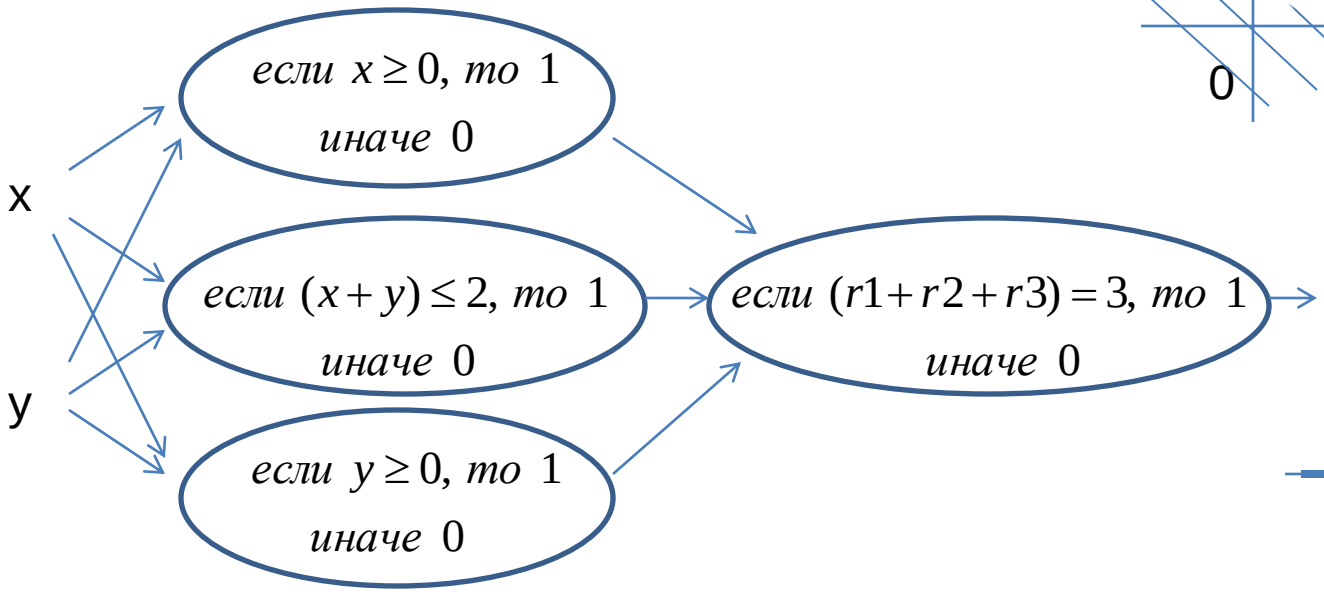


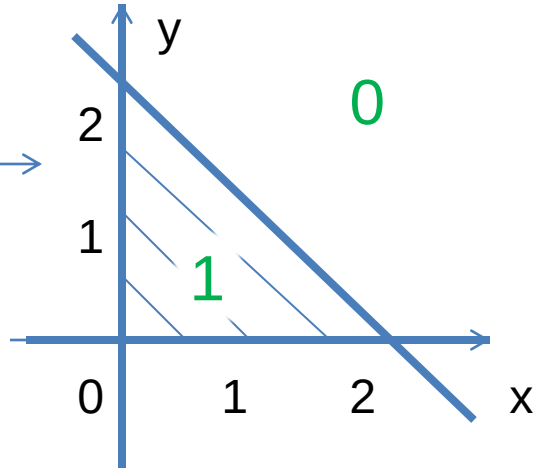
Нейронные сети. Геометрическая интерпретация

Если интерпретировать пару чисел **(x, y)** на входе как точку на плоскости, то данный нейрон принимает решение, на какой стороне от линии находится вход.



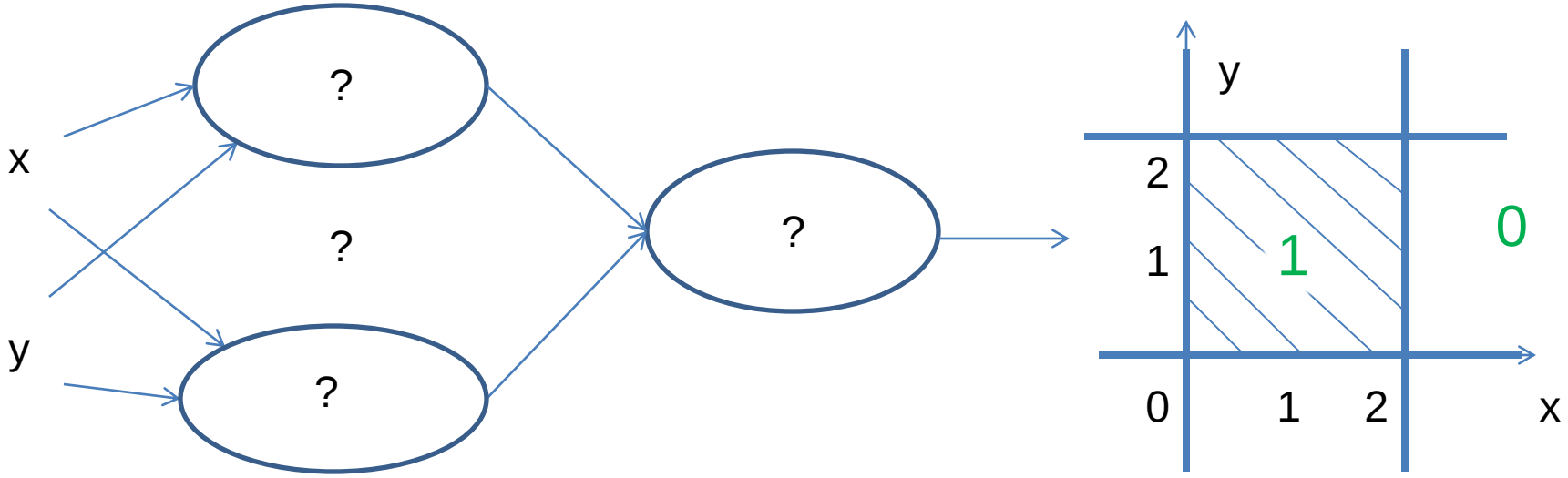




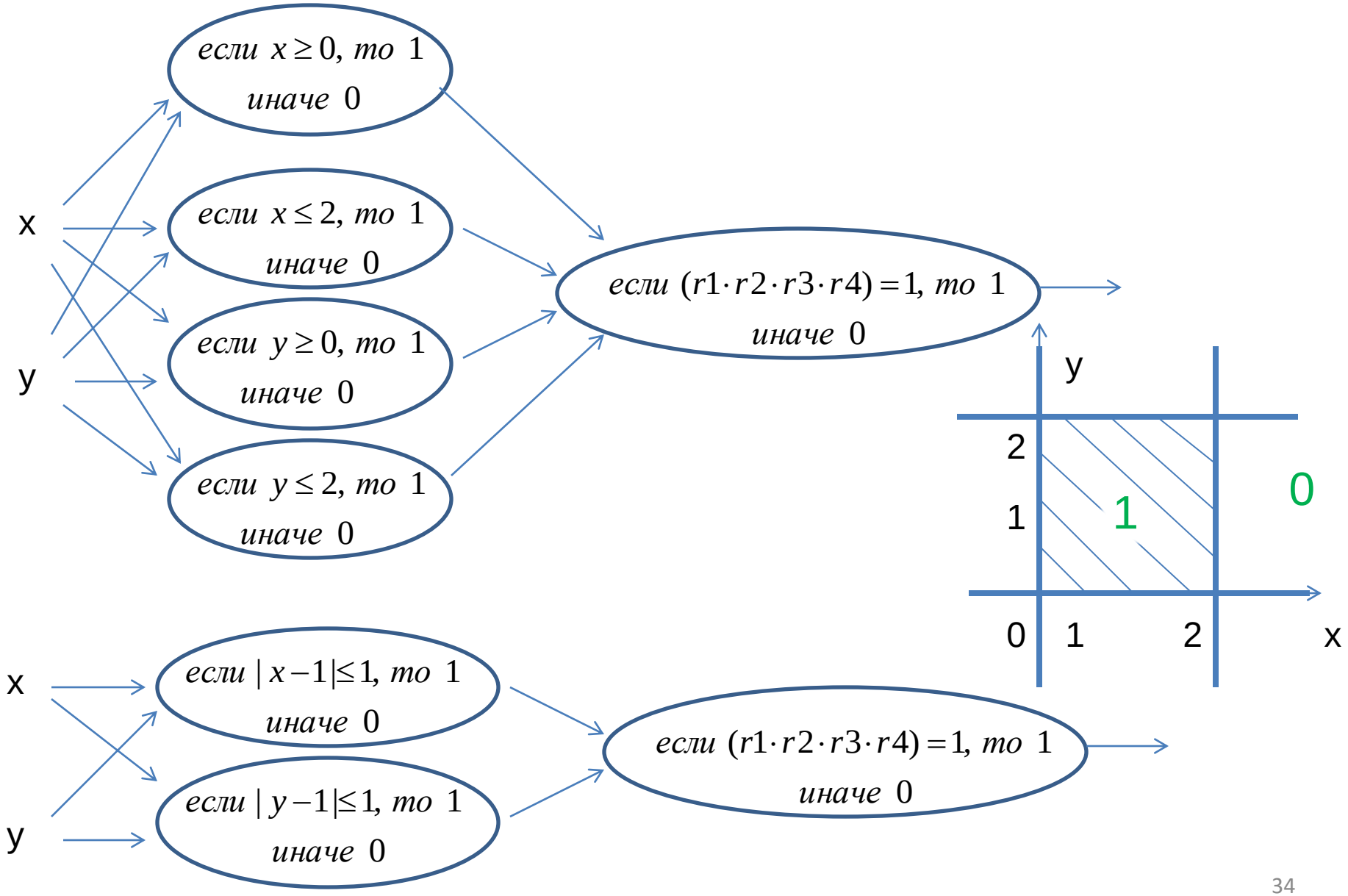


Нейронная сеть определяется топологией связей, весами и формулой принятия решения в узлах. Очевидно, сеть может принимать более сложные решения, чем один нейрон.

Нейронные сети. Геометрическая интерпретация

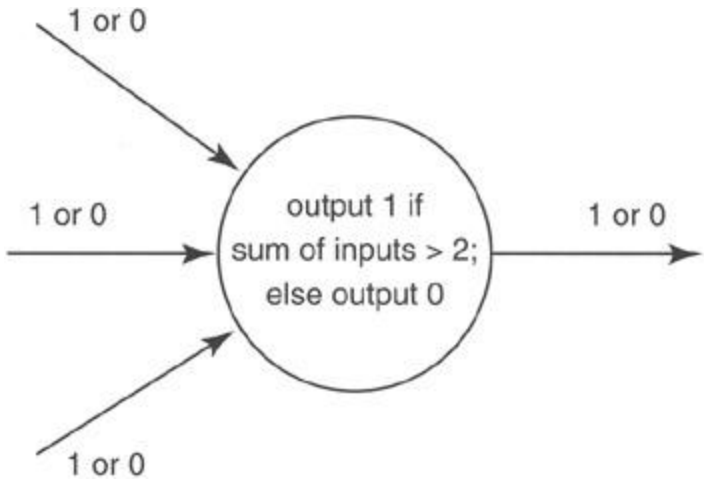


Нейронные сети. Геометрическая интерпретация



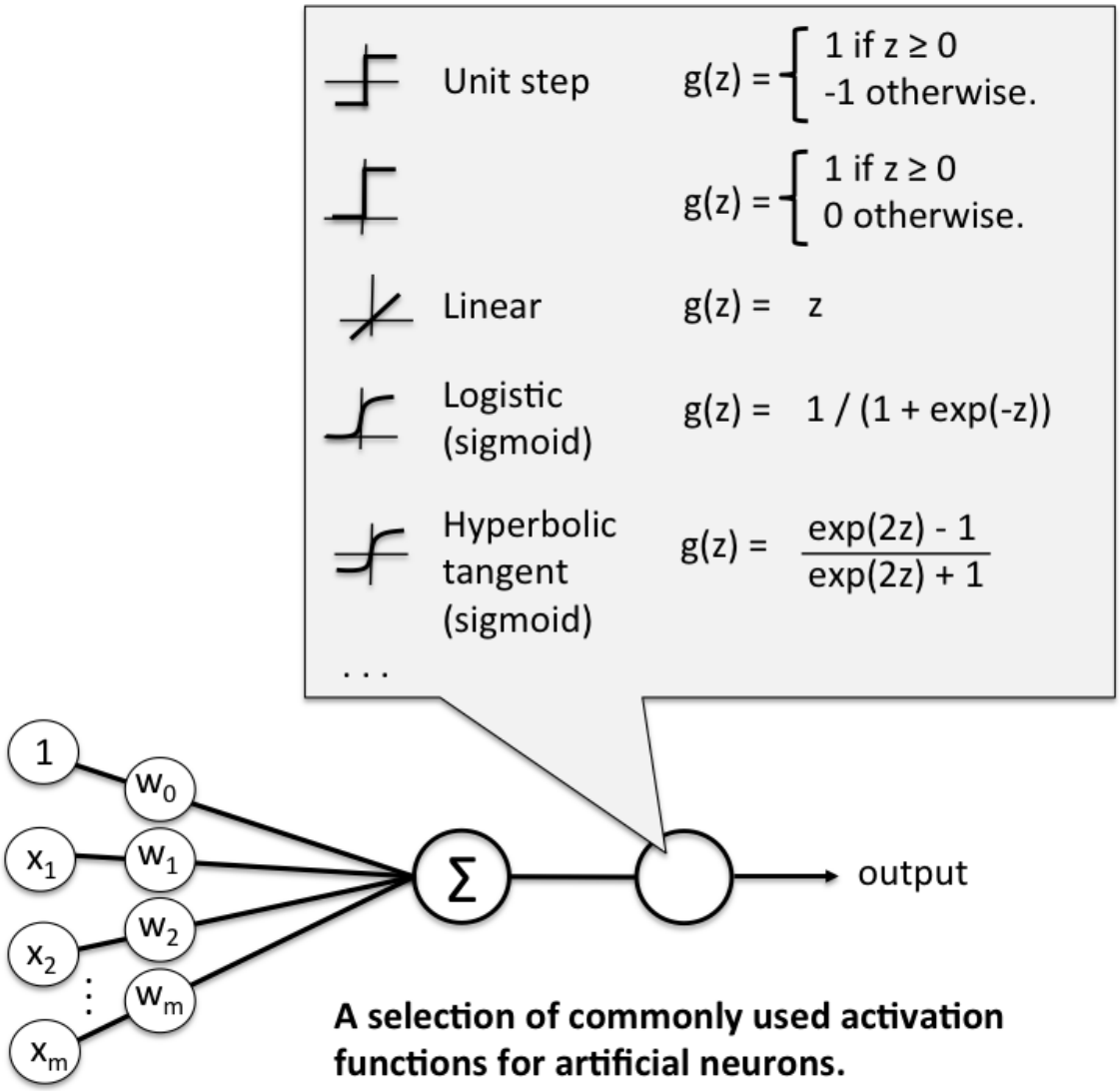
Нейронные сети. Веса связей

Неограниченная сложность возможна как при создании и соединении нейронов, так и при определении строгости связей. Если вместо простой суммы входных сигналов ***i1 + i2 + i3***, использовать их взвешенную сумму ***10*i1 + i2 + 0,5*i3***, то сеть станет более чувствительной ко входу 1 и менее ко входу 3.



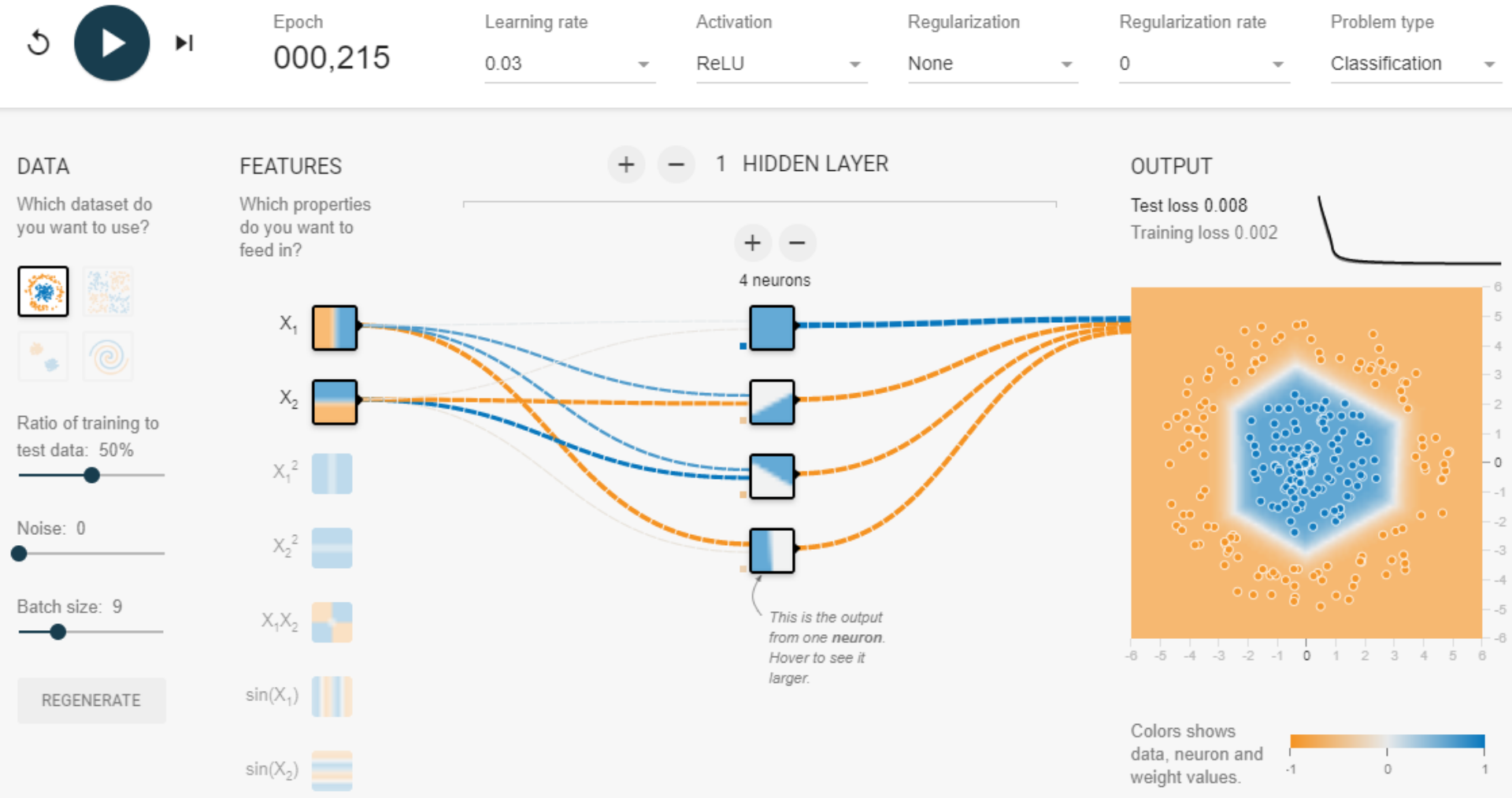
В процессе обучения происходит подбор параметров при неизменной топологии сети. Для этого применяют сеть с начальными параметрами к различным примерам и сравнивают ответ с правильным. При несовпадении производят уточнение параметров.

Нейронные сети. Функции активации



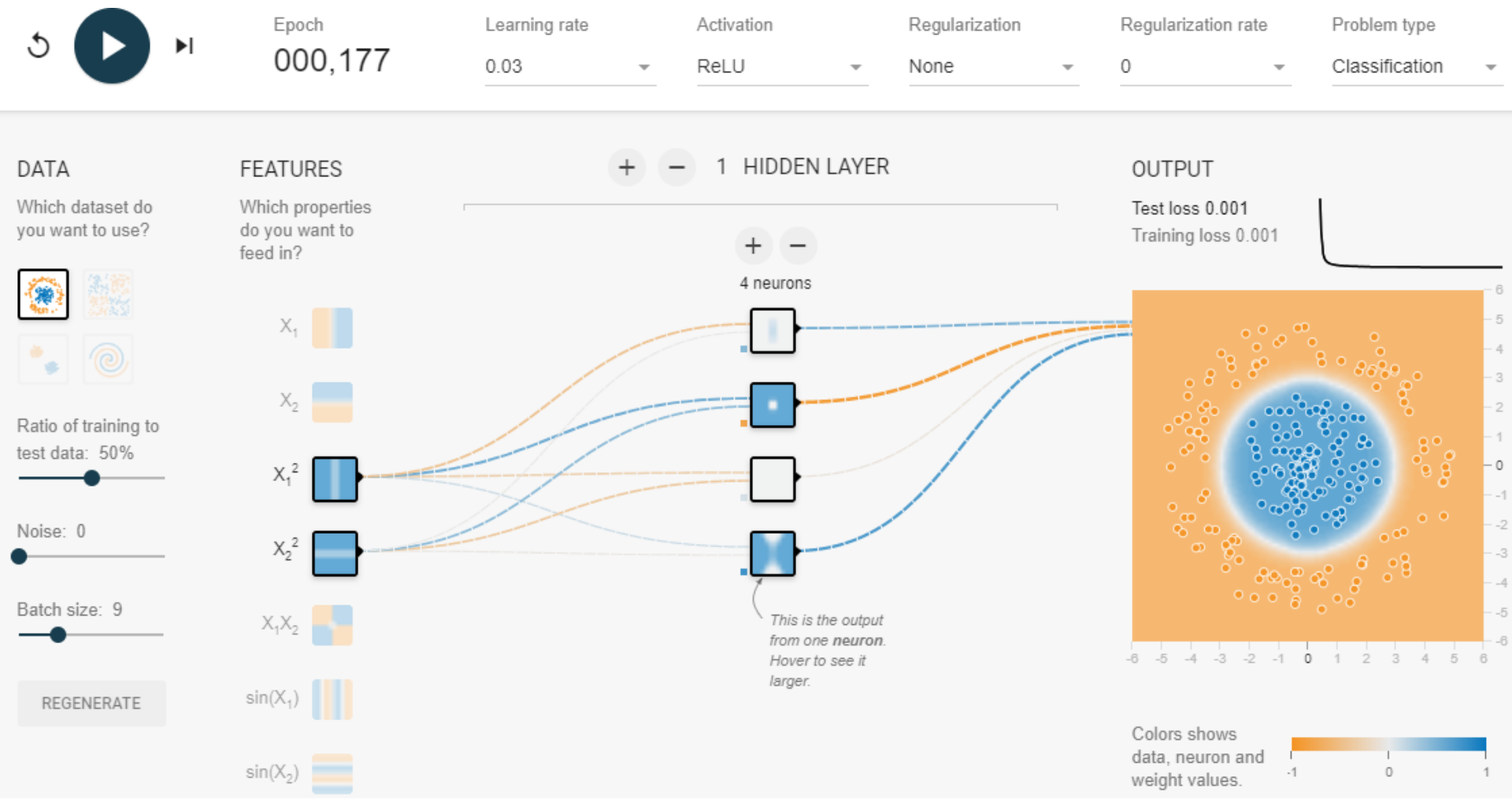
A selection of commonly used activation functions for artificial neurons.

A Neural Network Playground



http://playground.tensorflow.org

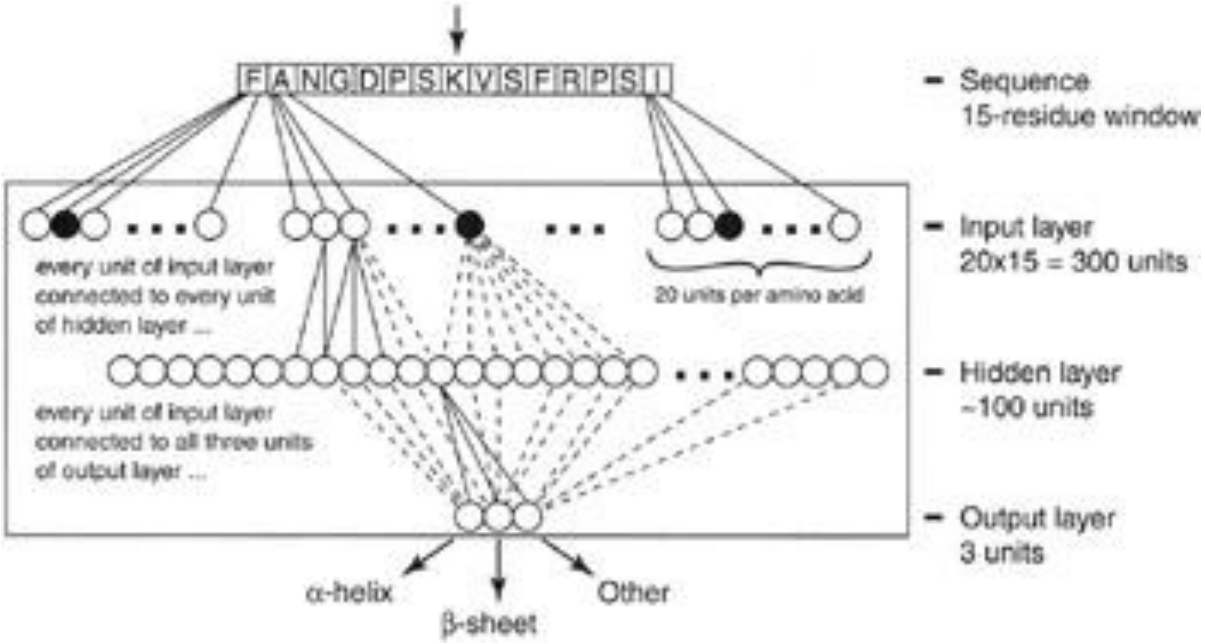
A Neural Network Playground



http://playground.tensorflow.org

Предсказание вторичной структуры. Нейронные сети

Пример нейронной сети для предсказания вторичной структуры PSIPRED (Jones, 1999).



Входная область сканирует последовательность окном шириной в 15 остатков, при этом предсказание делается для центрального остатка. Каждому из остатков соответствует 20 входных нейронов, один из которых активен.

Скрытая область состоит из ~ 100 нейронов, соединенных с каждым нейроном ввода и вывода.

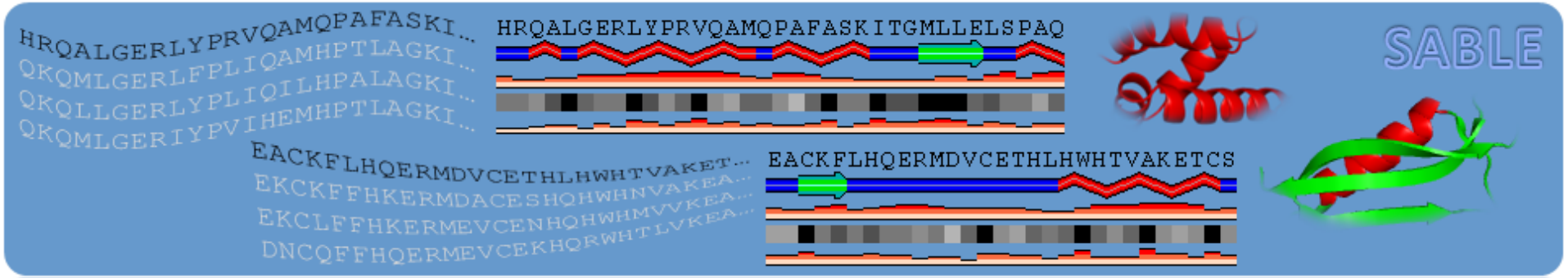
Область вывода состоит из трех нейронов, которые делают предсказание: «**спираль**», «**лист**» или «**ни то, ни другое**».

Предсказание вторичной структуры. PSIPRED и SABLE

The PSIPRED Protein Sequence Analysis Workbench

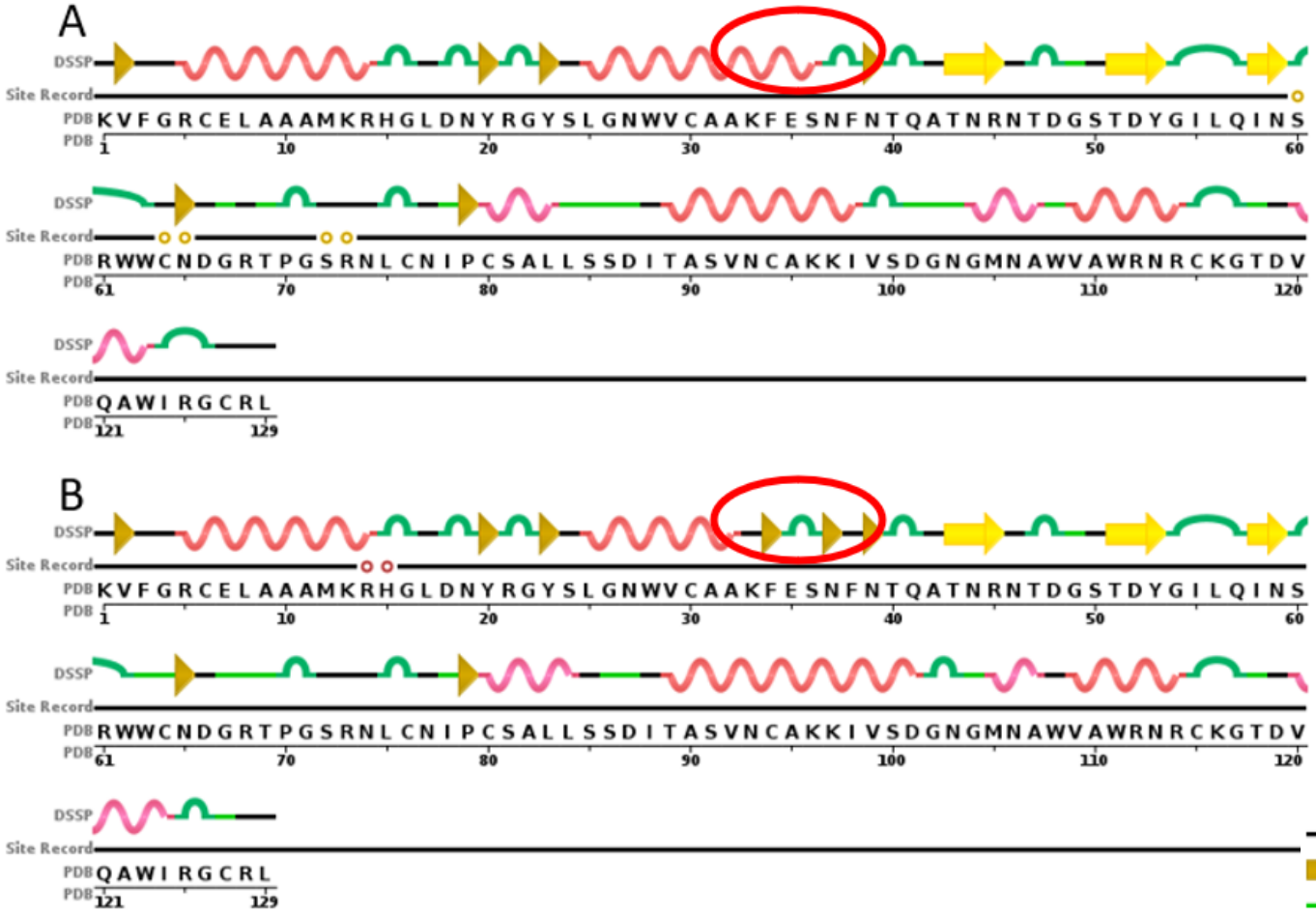
The PSIPRED Protein Sequence Analysis Workbench aggregates several UCL structure prediction methods into one location. Users can submit a protein sequence, perform the predictions of their choice and receive the results of the prediction via e-mail or the web.
For a summary of the available methods you can read More...

PSIPRED



Предсказание вторичной структуры. Ограничения

Hen lysozyme (pdb 5kxz and 6s7n)

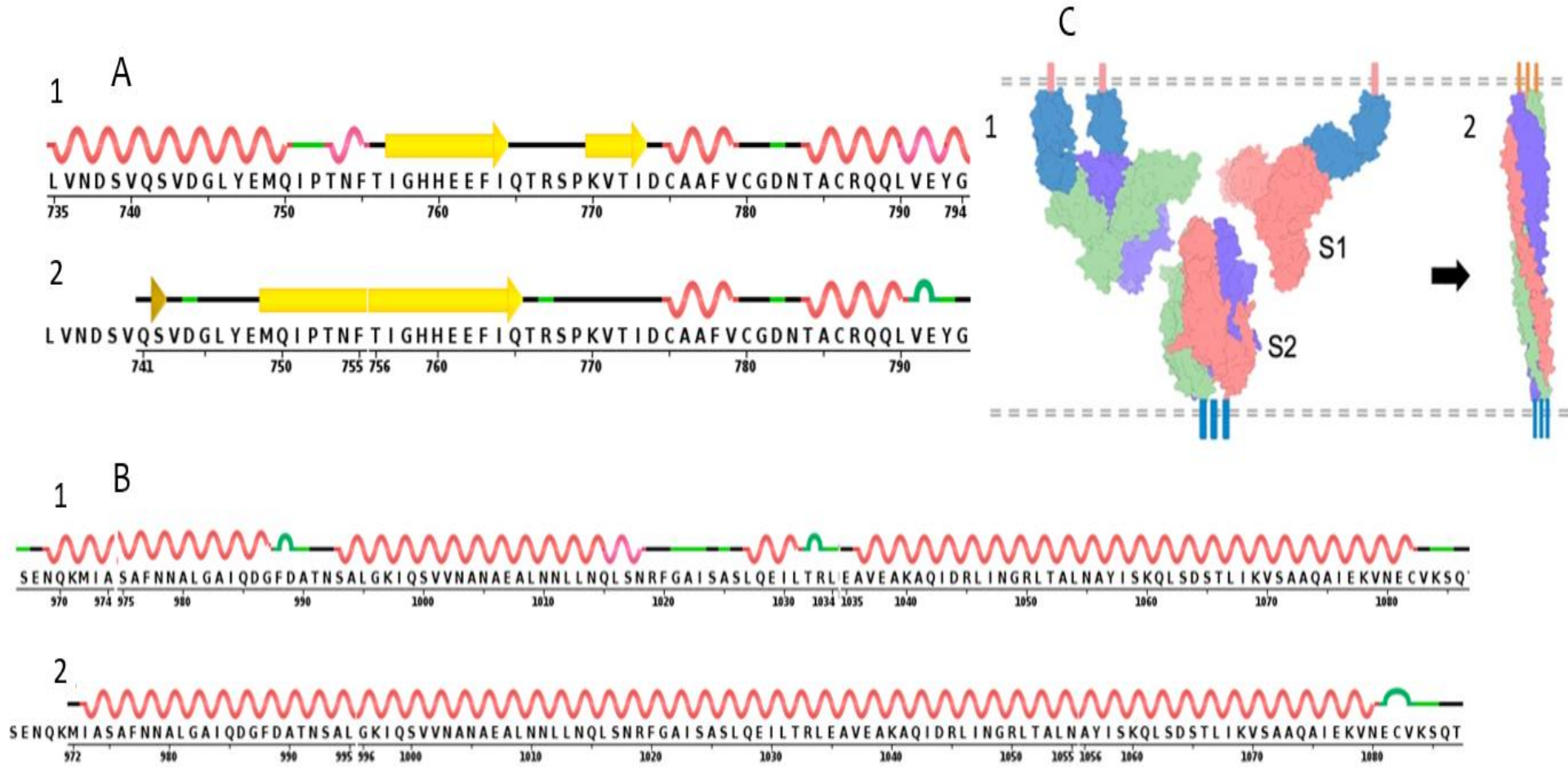


DSSP Legend

- empty: no secondary structure assigned
- B: beta bridge
- S: bend
- T: turn
- E: beta strand
- G: 3/10-helix
- H: alpha helix
- I: pi helix

Предсказание вторичной структуры. Ограничения

S2-domain of the S-protein of the murine hepatitis virus (pdb 3jcl and 6b3o)



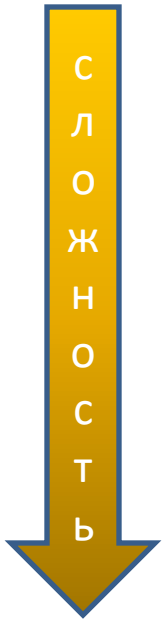
1) S-protein before dissociation to S1 and S2 fragments
2) Irreversible transition of S2 leading to membrane fusion

Предсказание структуры белков

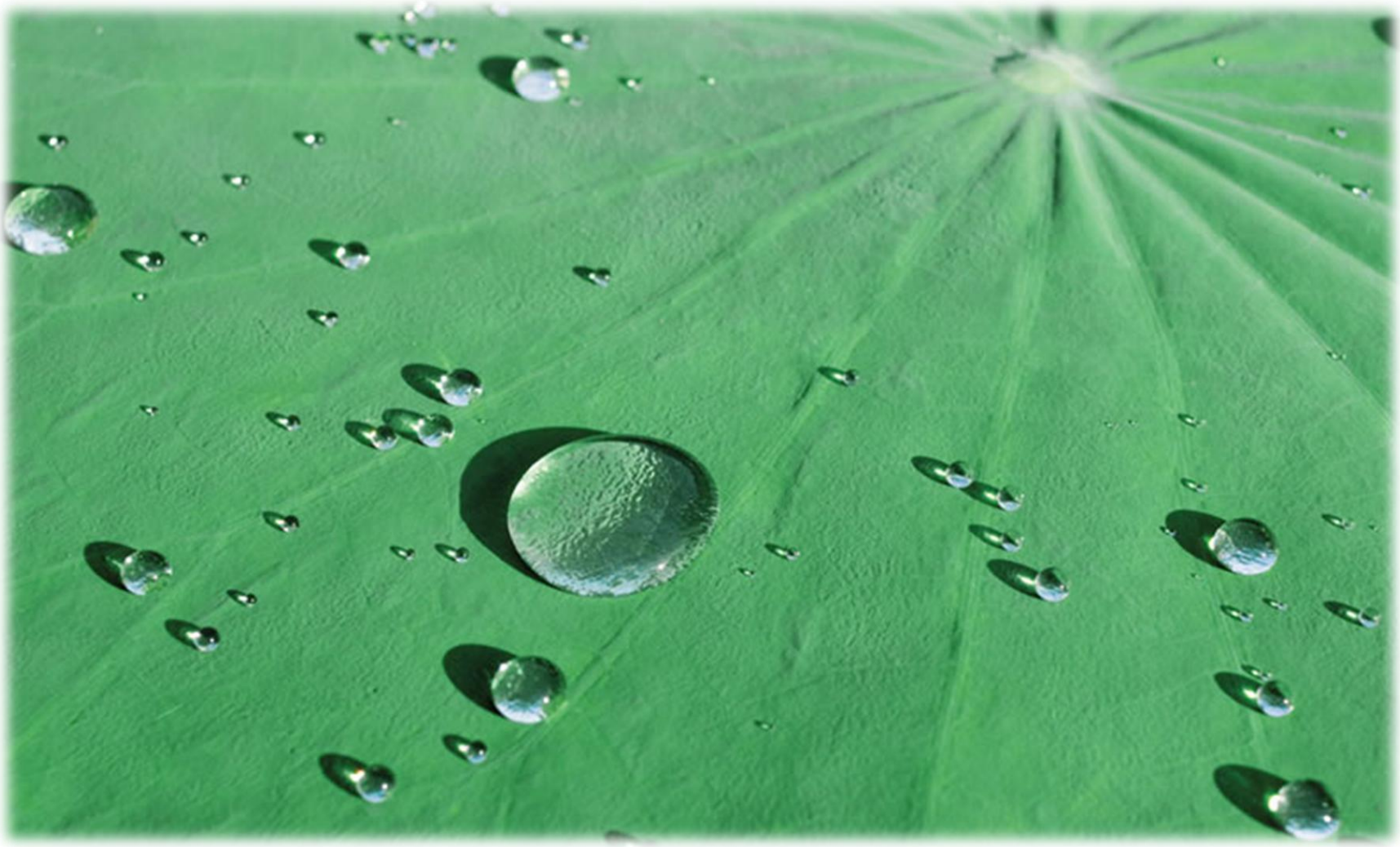
Сворачивание белка в уникальную конформацию наводит на мысль об алгоритме формирования структуры белка по его последовательности, но доказательством полноты и правильности нашего понимания могла бы стать его реализация в виде компьютерной программы...

Методы предсказания структуры по последовательности:

- **Предсказание вторичной структуры;**
- Предсказание топологии;
- Моделирование по гомологии;
- Распознавание типов укладки (по известной библиотеке фолдов);
- Априорное предсказание новых типов укладки.

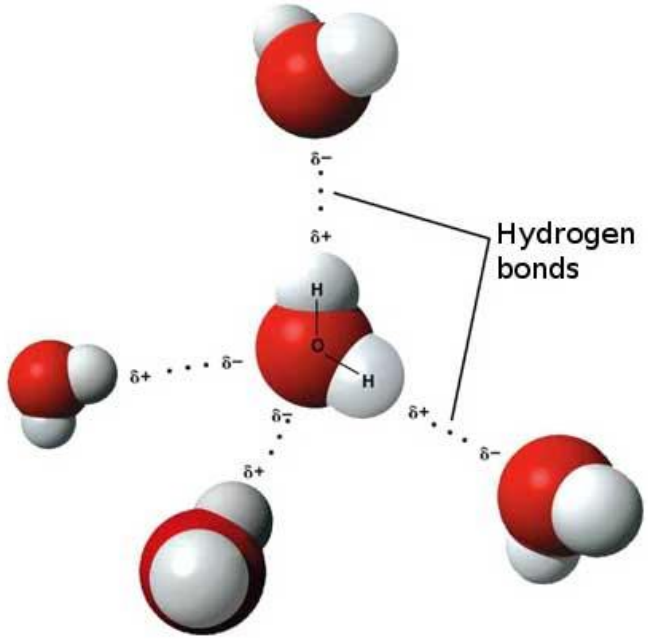


Гидрофобность



Гидрофобность

Гидрофобный эффект – следствие большей упорядоченности молекул воды вокруг неполярной молекулы.

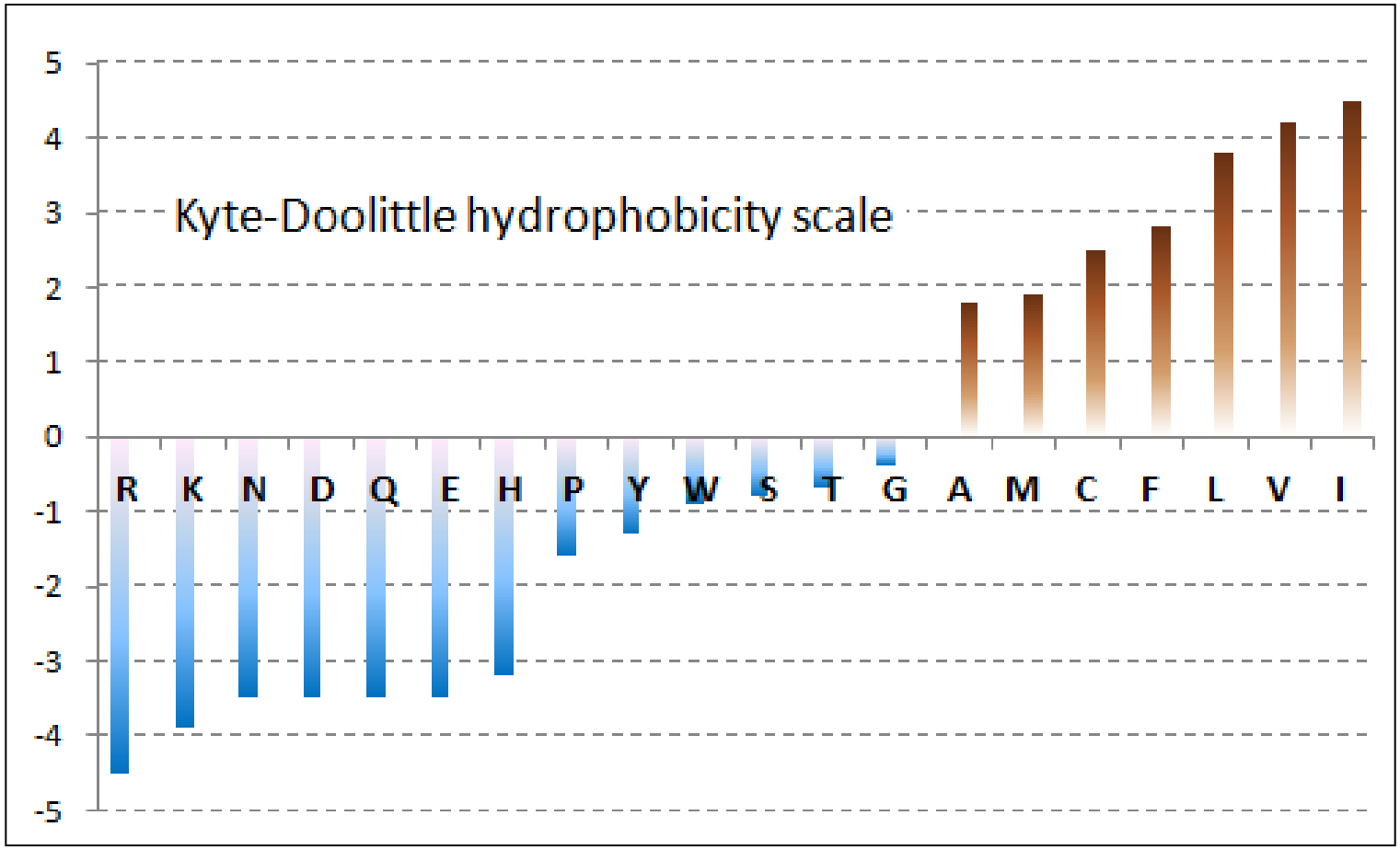


Мерой гидрофобности молекул может служить коэффициент разделения – равновесное отношение концентраций вещества в двух фазах в случае несмешивающихся растворителей:

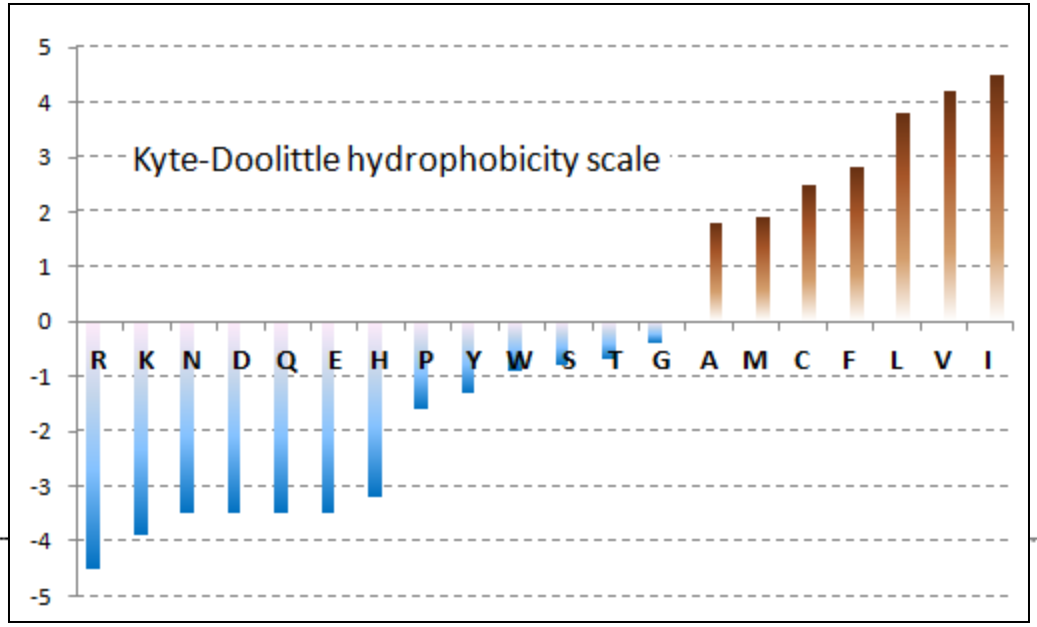
$$log\ P_{oct/wat} = log\left(\frac{[solute]_{octanol}}{[solute]^{un-ionized}_{water}}\right)$$

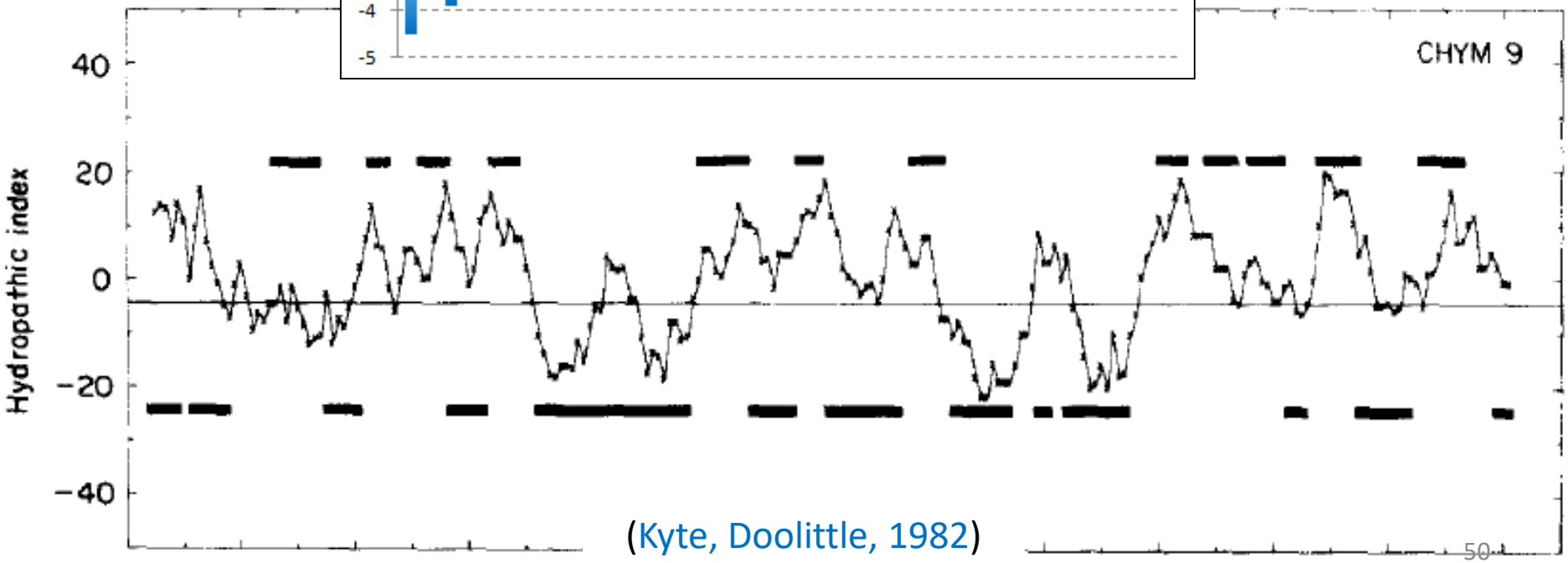
Гидрофобность

Распространенная шкала туманного происхождения (Kyte, Doolittle, 1982)



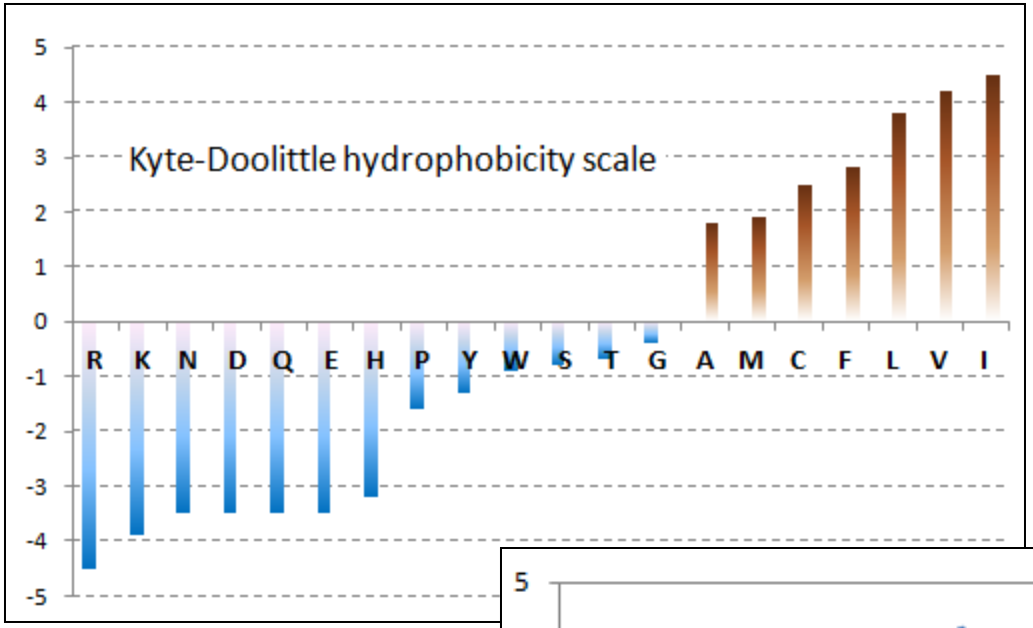
Предсказание топологии. Профили гидрофобности

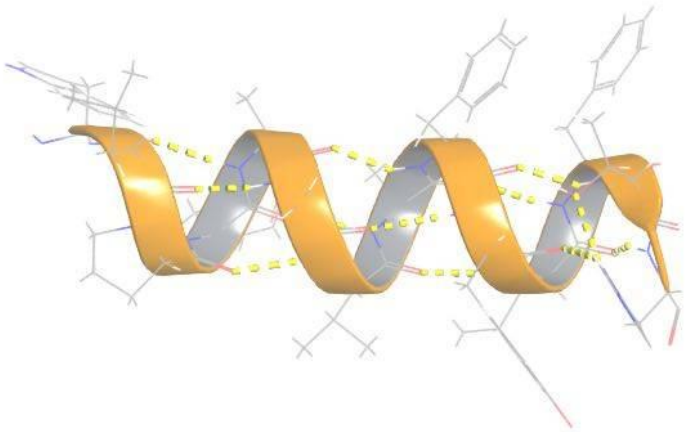




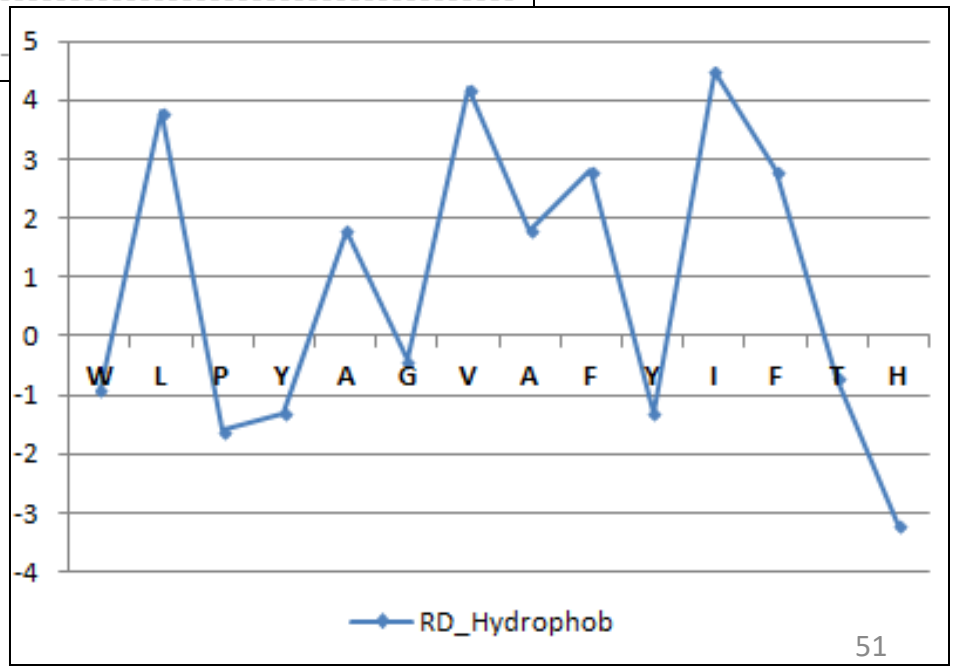
(Kyte, Doolittle, 1982)

Предсказание топологии. Профили гидрофобности

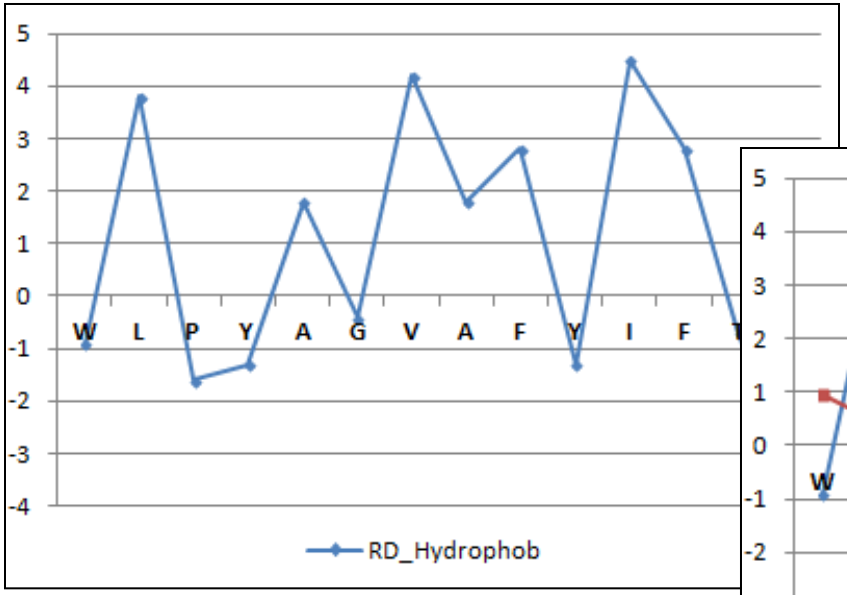


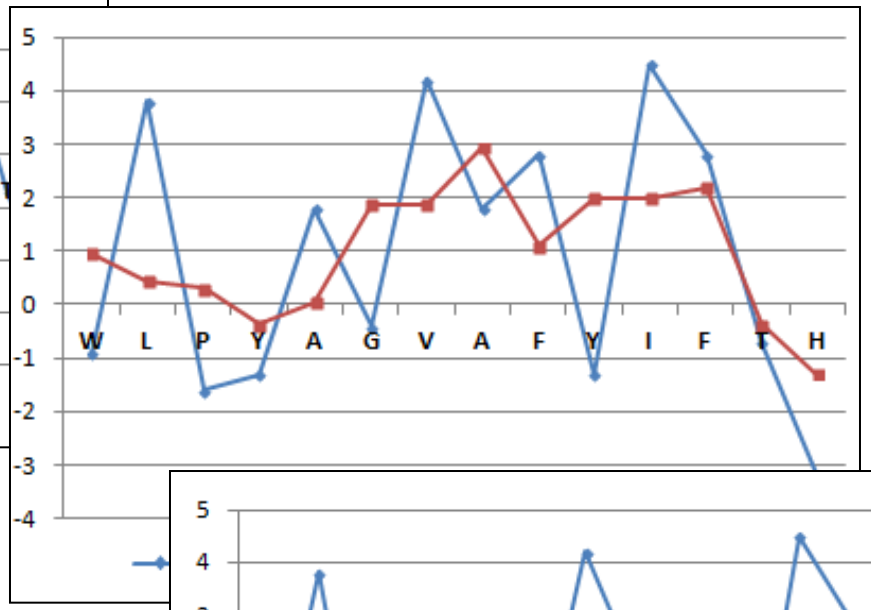


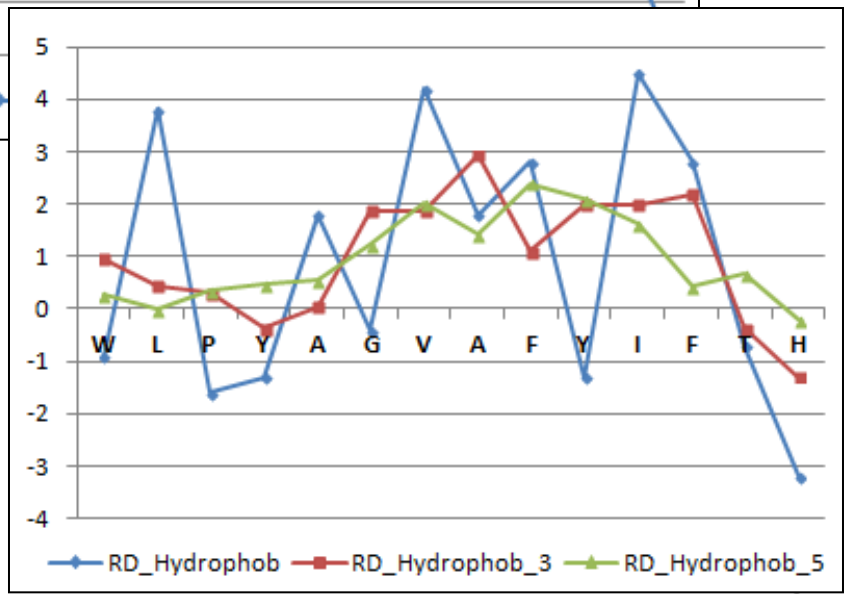
...**W**LPY**A**GVA**F**YI**F**TH...

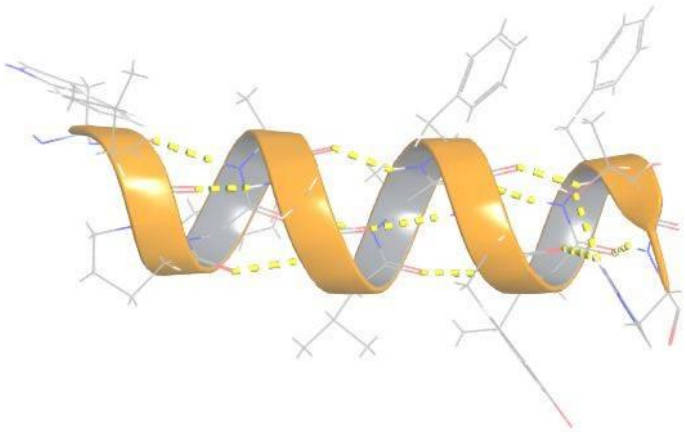


Предсказание топологии. Профили гидрофобности





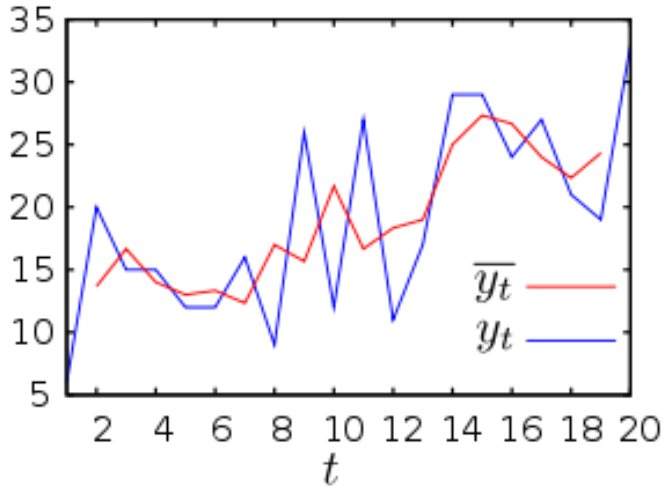




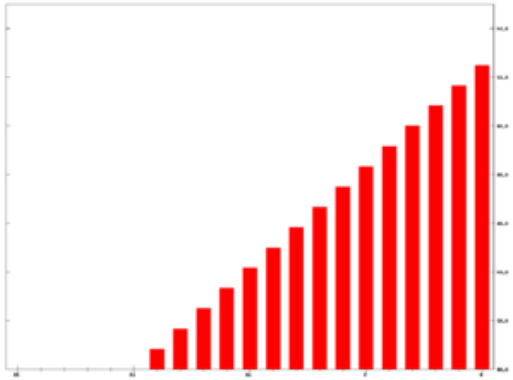
...**W**LPY**A**GVA**F**YI**F**TH...

Скользящее среднее

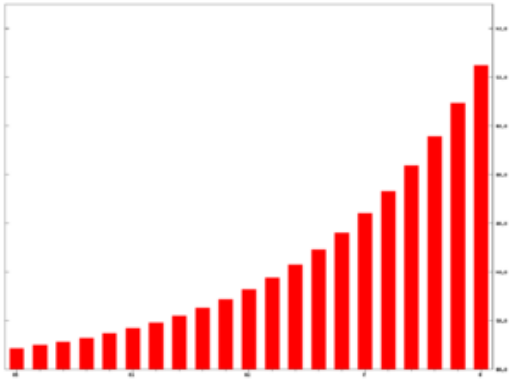
$$\bar{y}_t = \frac{y_t + y_{t-1}}{2}$$



$$WMA_t = \frac{\sum_{i=0}^{n-1} W_{t-i} \cdot p_{t-i}}{\sum_{i=0}^{n-1} W_{t-i}}$$

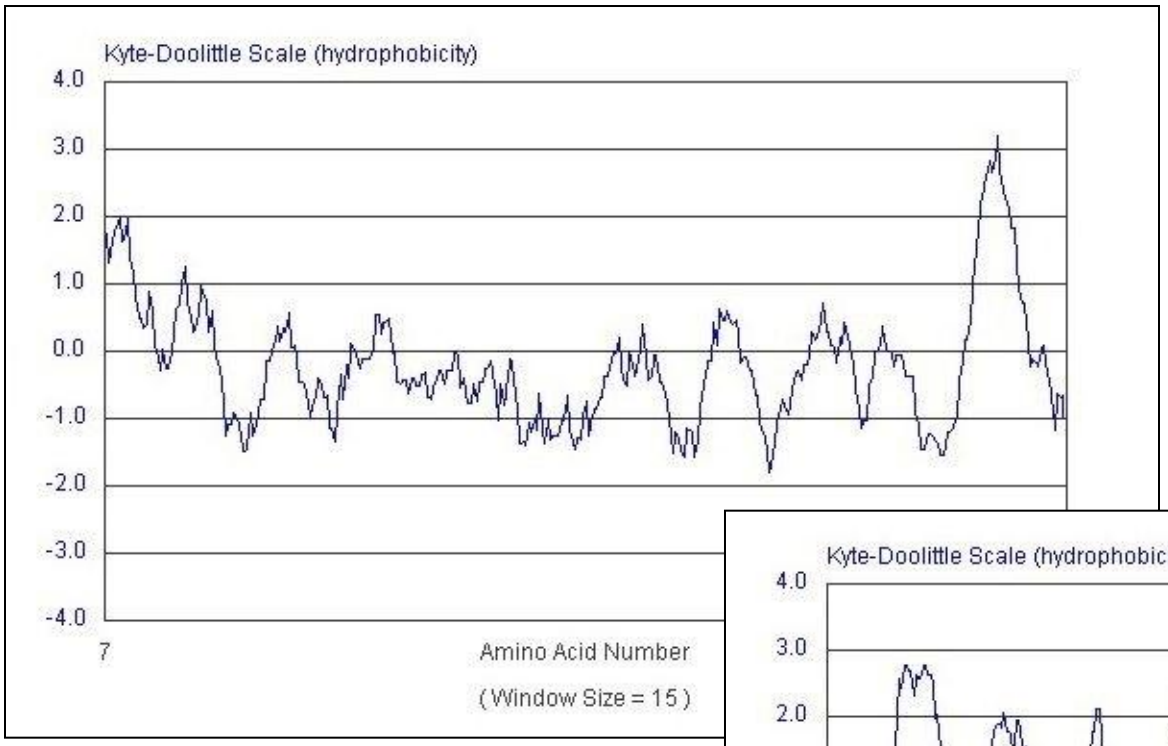


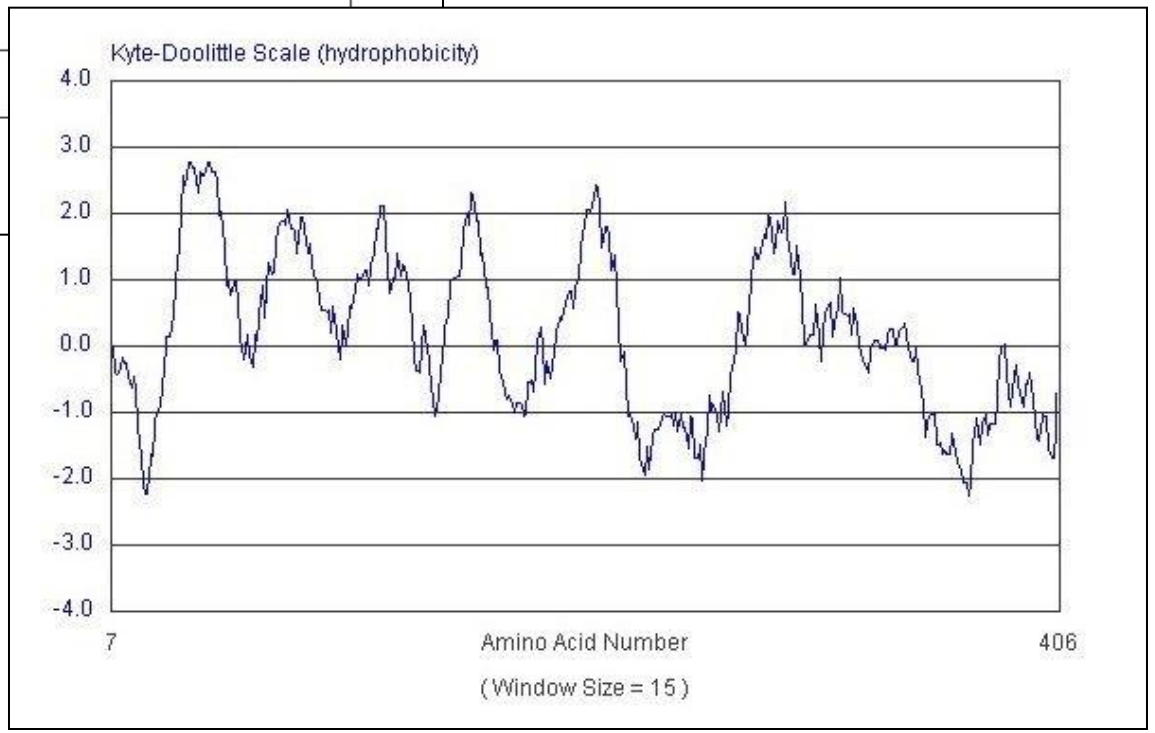
Линейное взвешивание

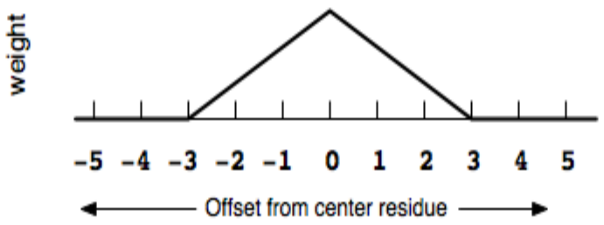


Экспоненциальное взвешивание

Предсказание топологии. Профили гидрофобности







Треугольное взвешивание

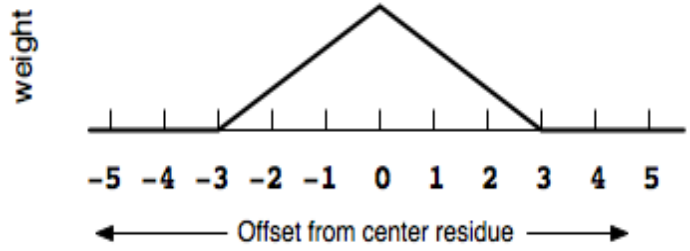
Hydrophobicity profile

ProtScale

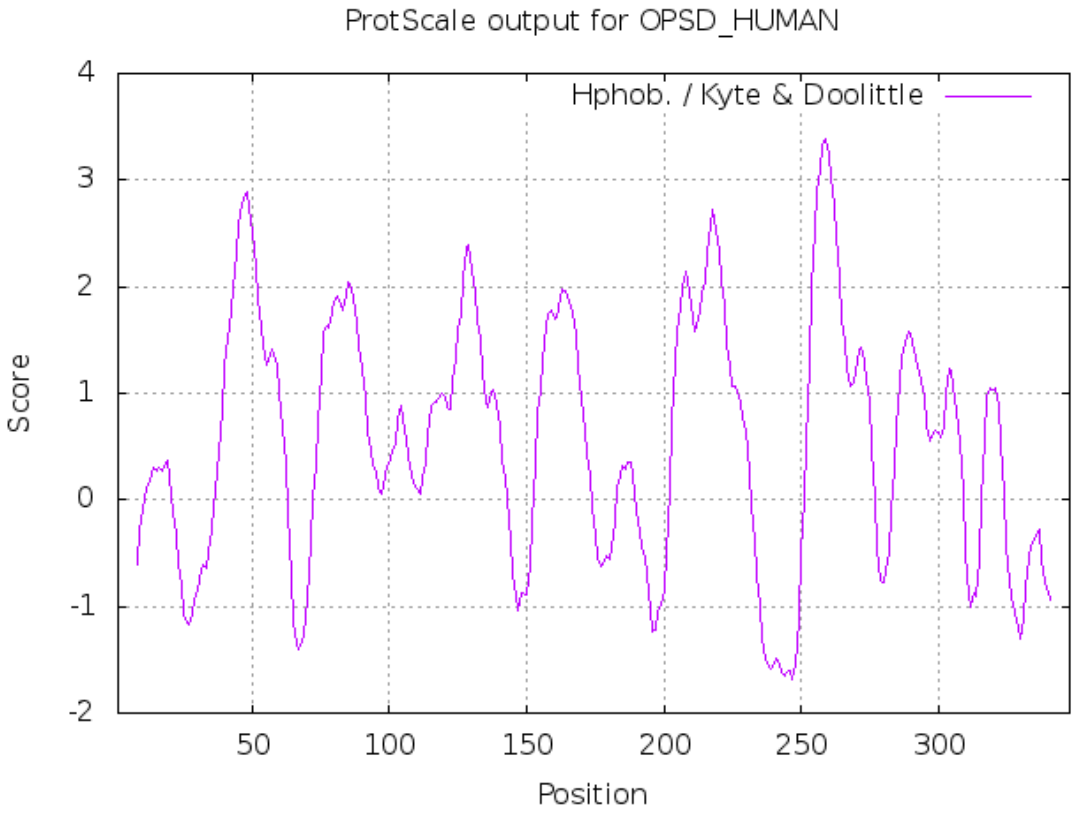
- ○ Molecular weight
- ○ Bulkiness
- ○ Polarity / Grantham
- ○ Recognition factors
- ○ Hphob. OMH / Sweet et al.
- ● Hphob. / Kyte & Doolittle
- ○ Hphob. / Abraham & Leo
- ○ Hphob. / Bull & Breese
- ○ Hphob. / Guy
- ○ Hphob. / Miyazawa et al.
- ○ Hphob. / Roseman
- ○ Hphob. / Wolfenden et al.
- ○ Hphob. HPLC / Wilson & al
- ○ Hphob. HPLC pH3.4 / Cowan
- ○ Hphob. / Rf mobility
- ○ HPLC / TFA retention
- ○ HPLC / retention pH 2.1
- ○ % buried residues
- ○ Hphob. / Chothia
- ○ Ratio hetero end/side
- ○ Average flexibility
- ○ beta-sheet / Chou & Fasman
- ○ alpha-helix / Deleage & Roux
- ○ beta-turn / Deleage & Roux
- ○ alpha-helix / Levitt
- ○ beta-turn / Levitt
- ○ Antiparallel beta-strand
- ○ A.A. composition
- ○ Relative mutability
- ○ Number of codon(s)
- ○ Polarity / Zimmerman
- ○ Refractivity
- ○ Hphob. / Eisenberg et al.
- ○ Hphob. / Hopp & Woods
- ○ Hphob. / Manavalan et al.
- ○ Hphob. / Black
- ○ Hphob. / Fauchere et al.
- ○ Hphob. / Janin
- ○ Hphob. / Rao & Argos
- ○ Hphob. / Tanford
- ○ Hphob. / Welling & al
- ○ Hphob. HPLC / Parker & al
- ○ Hphob. HPLC pH7.5 / Cowan
- ○ HPLC / HFBA retention
- ○ Transmembrane tendency
- ○ HPLC / retention pH 7.4
- ○ % accessible residues
- ○ Hphob. / Rose & al
- ○ Average area buried
- ○ alpha-helix / Chou & Fasman
- ○ beta-turn / Chou & Fasman
- ○ beta-sheet / Deleage & Roux
- ○ Coil / Deleage & Roux
- ○ beta-sheet / Levitt
- ○ Total beta-strand
- ○ Parallel beta-strand
- ○ A.A. comp. in Swiss-Prot

Hydrophobicity profile

Window size: 15 ▾
Relative weight of the window edges compared to the window center (in %): 0
Weight variation model (if the relative weight at the edges is < 100%): ◉ linear ○ exponential
Do you want to normalize the scale from 0 to 1? ○ yes ◉ no



Triangle weighting

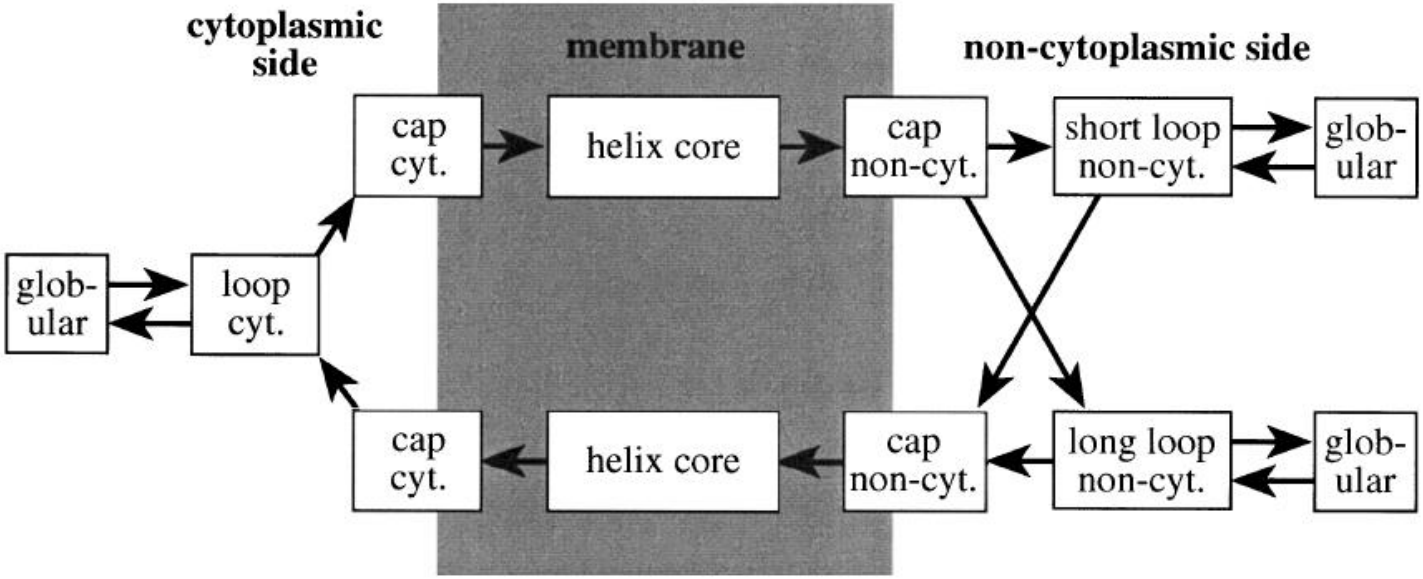


ExPASy Bioinformatics Resource Portal

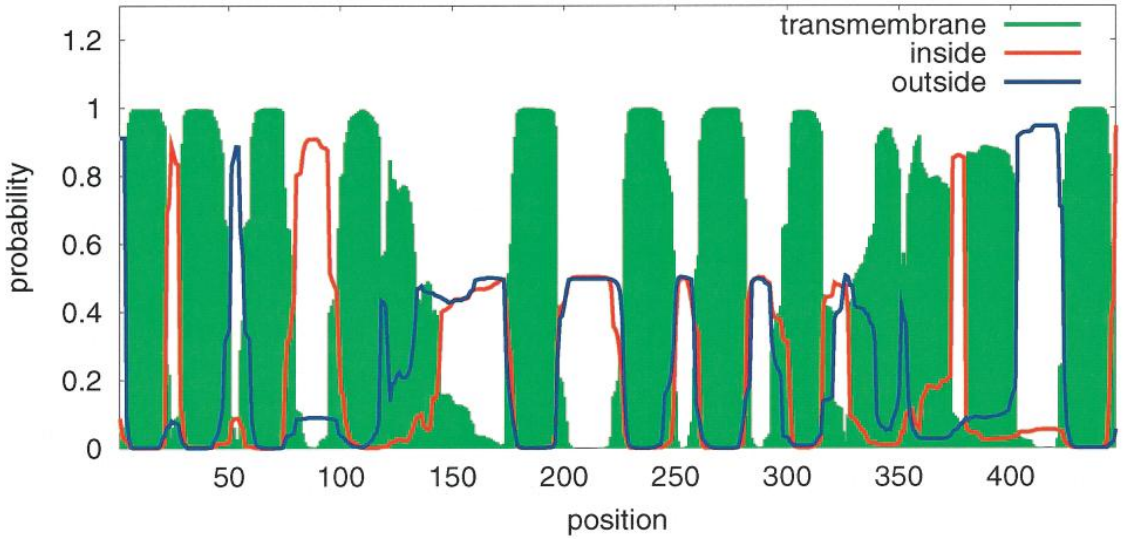
ProtScale

https://web.expasy.org/protscale/

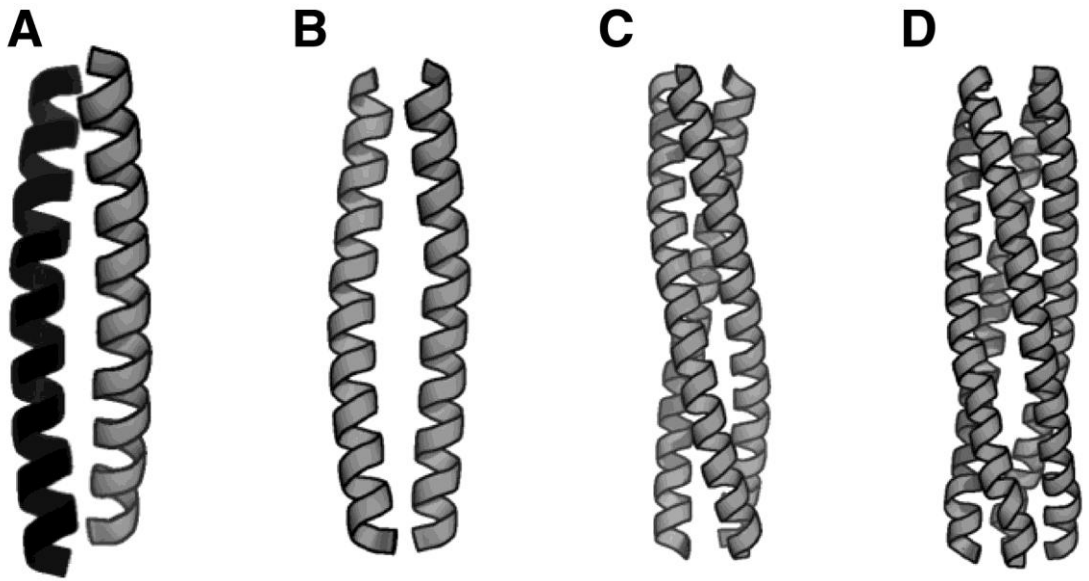
Предсказание трансмембранных сегментов



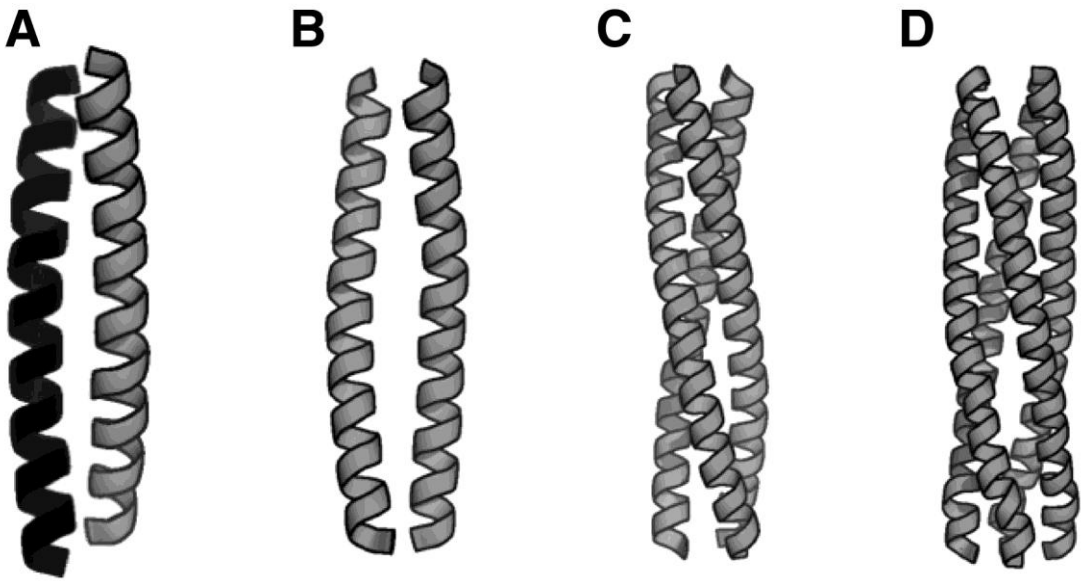
TMHMM (2001 - …)



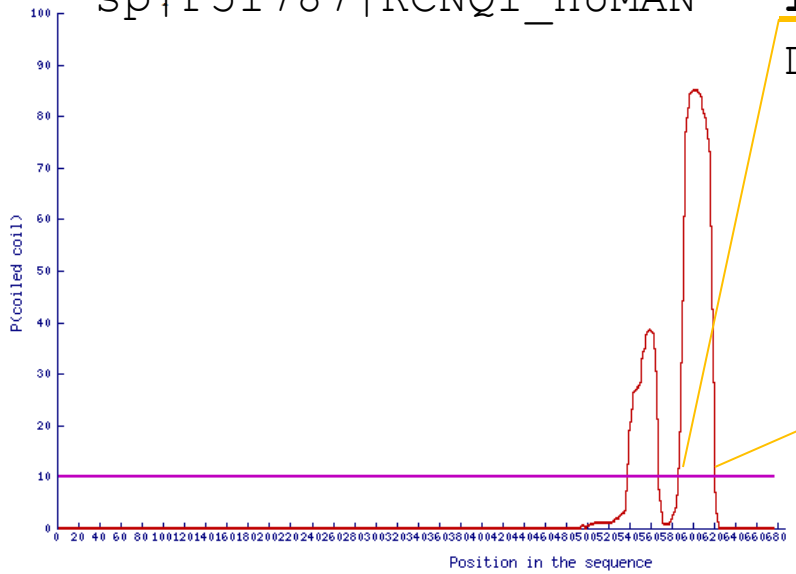
Предсказание топологии. Суперспирали



Предсказание топологии. Суперспирали




```
                                590       600       610       620
sp|P51787|KCNQ1_HUMAN           IGARLNRVEDKVTQLDQRLALITDMLHQLLSLH
                                DefgAbcDefgAbcDefgAbcDefgAbcDefGa
```



LOGICOIL

Multi-state coiled-coil oligomeric state prediction

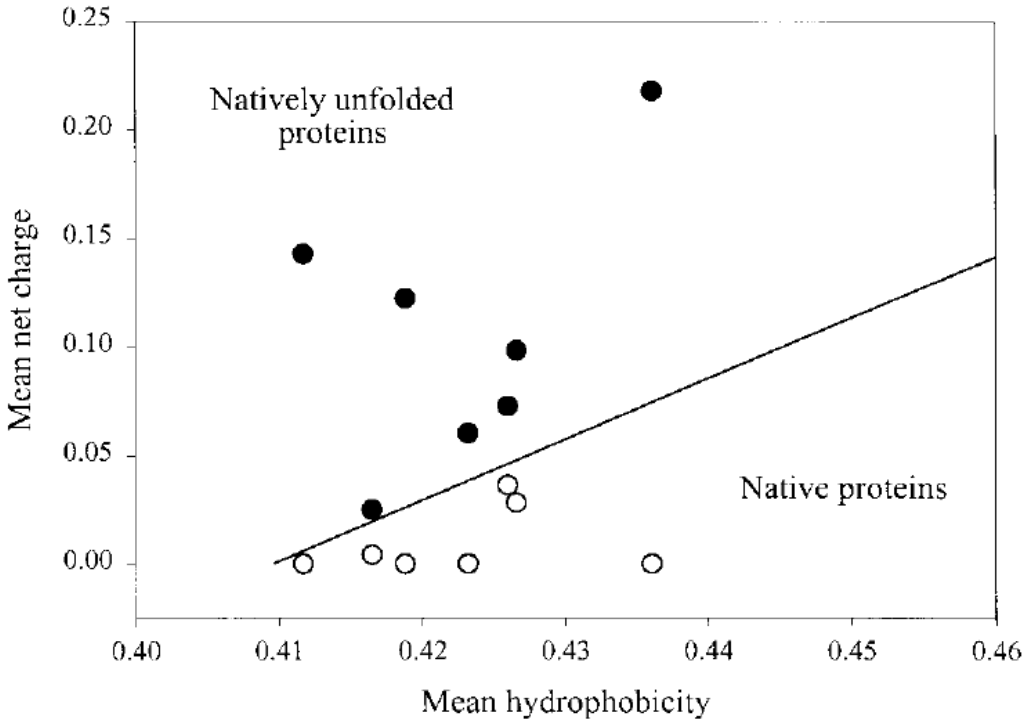
Предсказание топологии. Неупорядоченные белки

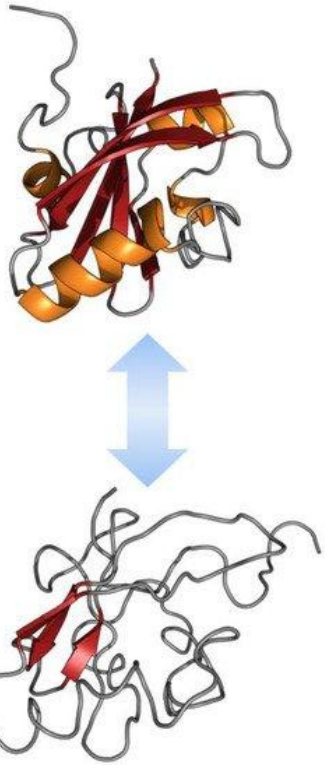
Intrinsically disordered proteins –

«нарушители» догмы «структура определяет функцию»

Предсказание неупорядоченных участков по последовательности:

- использование структурных данных (нейронные сети)
- использование свойств аминокислот

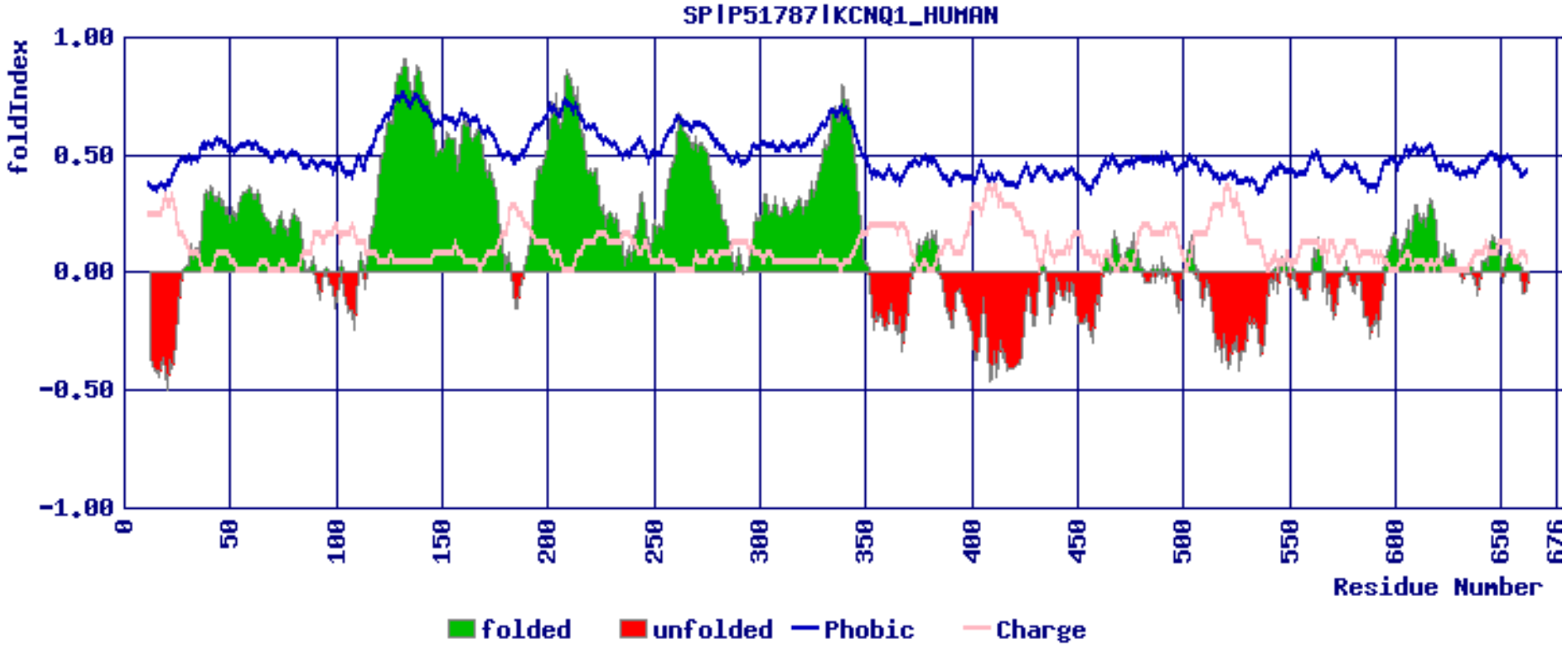




$$\langle R \rangle = 2.785 \langle H \rangle - 1.151$$

(Uversky VN, et al. 2000)

Неупорядоченные белки. FoldIndex



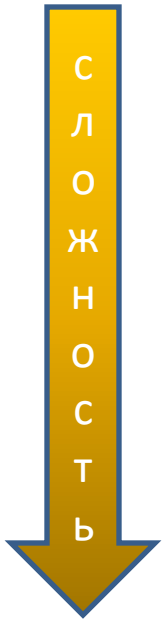
http://bip.weizmann.ac.il/fldbin/findex

Предсказание структуры белков

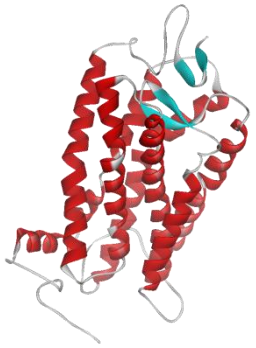
Сворачивание белка в уникальную конформацию наводит на мысль об алгоритме формирования структуры белка по его последовательности, но доказательством полноты и правильности нашего понимания могла бы стать его реализация в виде компьютерной программы...

Методы предсказания структуры по последовательности:

- **Предсказание вторичной структуры;**
- **Предсказание топологии;**
- Моделирование по гомологии;
- Распознавание типов укладки (по известной библиотеке фолдов);
- Априорное предсказание новых типов укладки.



Моделирование на основании гомологии. Алгоритм



Поиск гомологичных белков
с известной структурой (шаблоны)

Выбор подходящего шаблона

Выравнивание последовательности моделируемого белка
с последовательностью шаблона

```
Model:     FVVFVL.FAIC
             :: |: | ||
Template: VIIMVIAFLIC
```

Построение модели

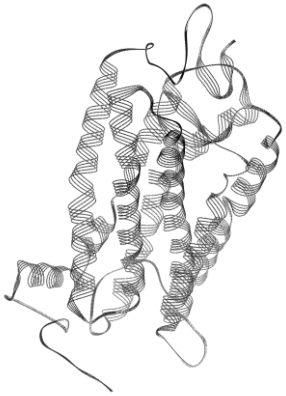
Оценка модели

Нет

Модель подходит?

Да

Ура!

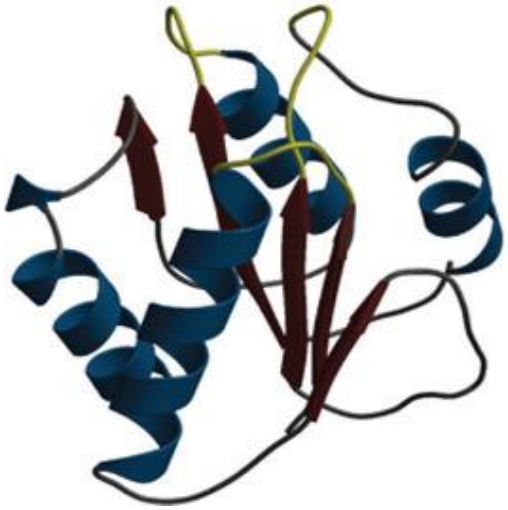


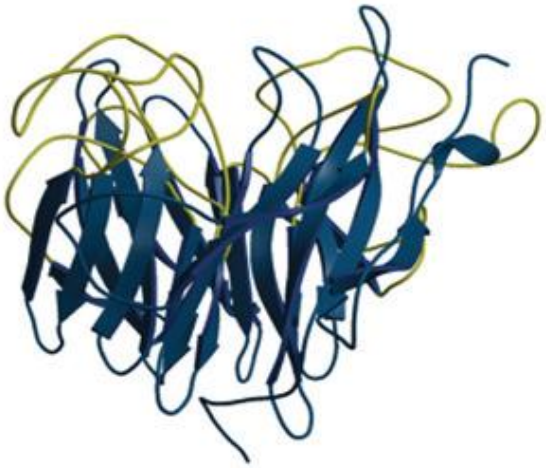
Моделирование на основании гомологии. Методы

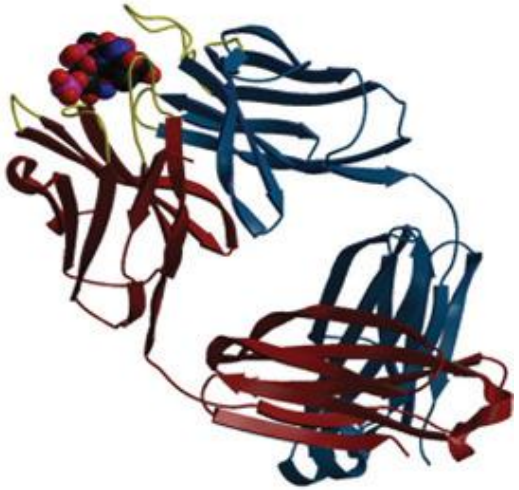
- Сборка модели из «жестких фрагментов»
- Моделирование на основе пространственных ограничений
- Моделирование путем сопоставления сегментов и другие методы

Моделирование петель

- *ab initio!*
- путем поиска в базах данных







Сборка модели из «жестких фрагментов»

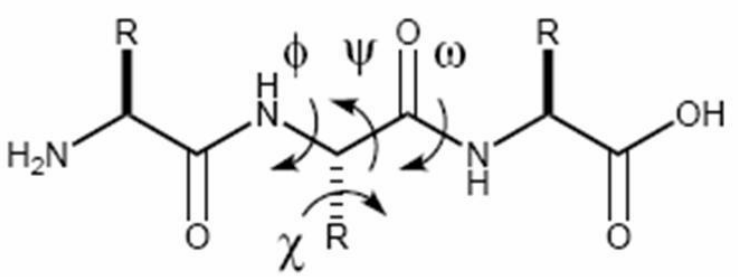
COMPOSER: исторически первый подход к моделированию

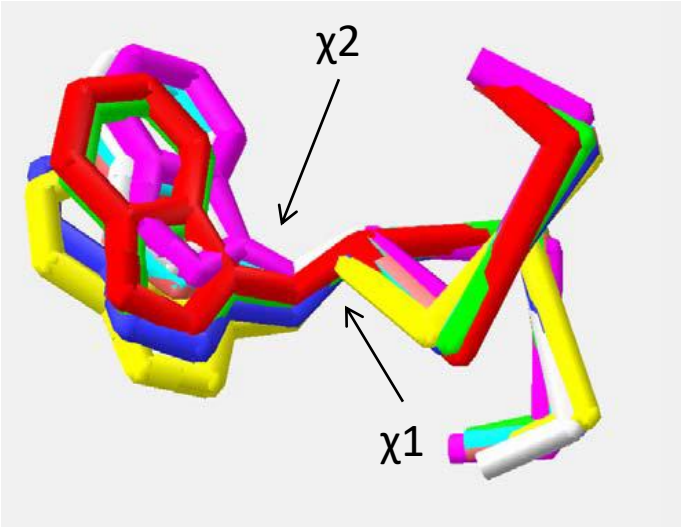
- моделирование в декартовых координатах (Sutcliffe, …, Blundell, 1987)

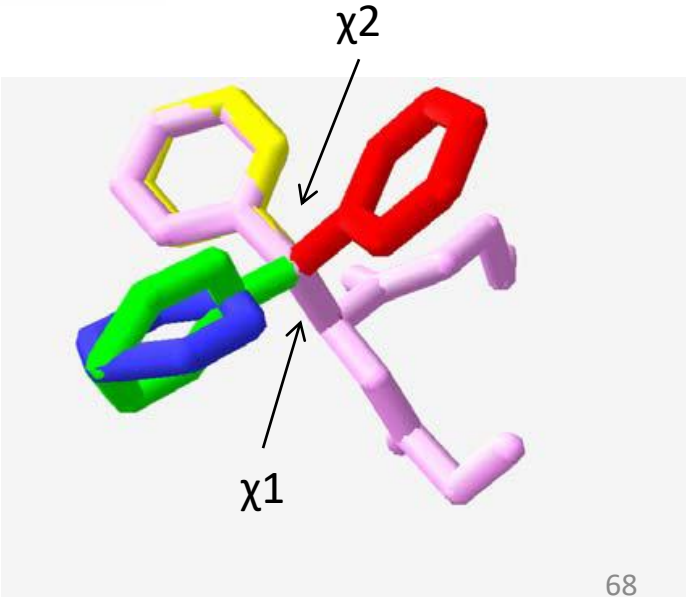
- Поиск белковых структур с последовательностями, гомологичными моделируемой. Выполнение выравнивание последовательностей, определение положения Cα-атомов консервативных остатков.
- Составление общего шаблона из перекрывающихся структурно консервативных фрагментов (при необходимости).
- Достройка боковых цепей с учетом библиотек ротамеров.
- Достройка петель путем подбора подходящих по геометрии гомологичных фрагментов среди белковых структур.
- Общая оптимизация геометрии.

Библиотеки ротамеров

- Лишь небольшая доля всех возможных конформаций боковых цепей реально наблюдается в экспериментальных структурах
- Конформация боковой цепи зависит от геометрии основной цепи
- Библиотеки ротамеров содержат наборы вероятных конформаций







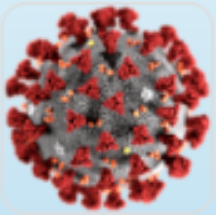
Сборка модели из «жестких фрагментов»

SWISS-MODEL Modelling Repository Tools Documentation Log in Create Account

Welcome to SWISS-MODEL

SWISS-MODEL is a fully automated protein structure homology-modelling server, accessible via the ExPASy web server, or from the program DeepView (Swiss Pdb-Viewer). The purpose of this server is to make protein modelling accessible to all life science researchers worldwide.

Start Modelling



Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is a positive-sense, single-stranded RNA coronavirus. It is a contagious virus that causes coronavirus disease 2019 (COVID-19).

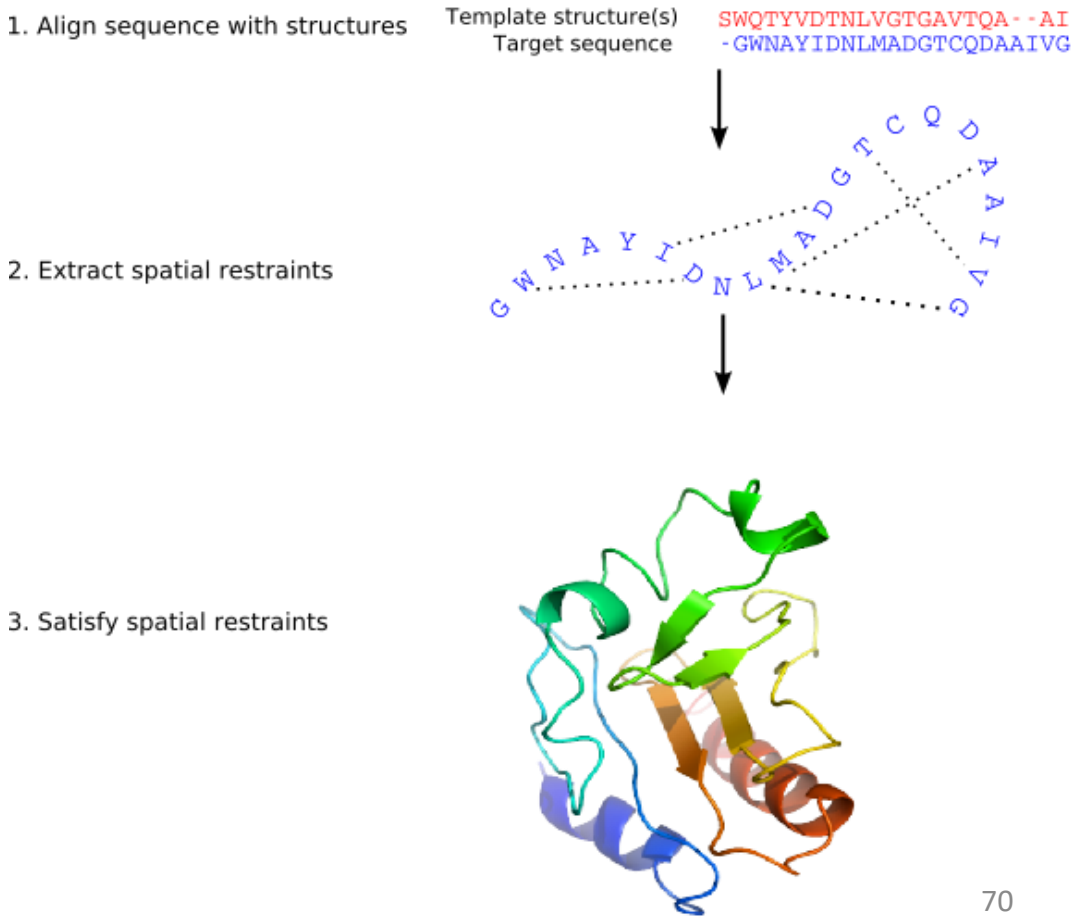
We modelled the full SARS-CoV-2 proteome based on the NCBI reference sequence NC_045512 and annotations from UniProt.

The results are available here.

Моделирование на основе пространственных ограничений. MODELLER

Наиболее распространенный подход к моделированию (Sali & Blundell, 1993) – моделирование во внутренних координатах

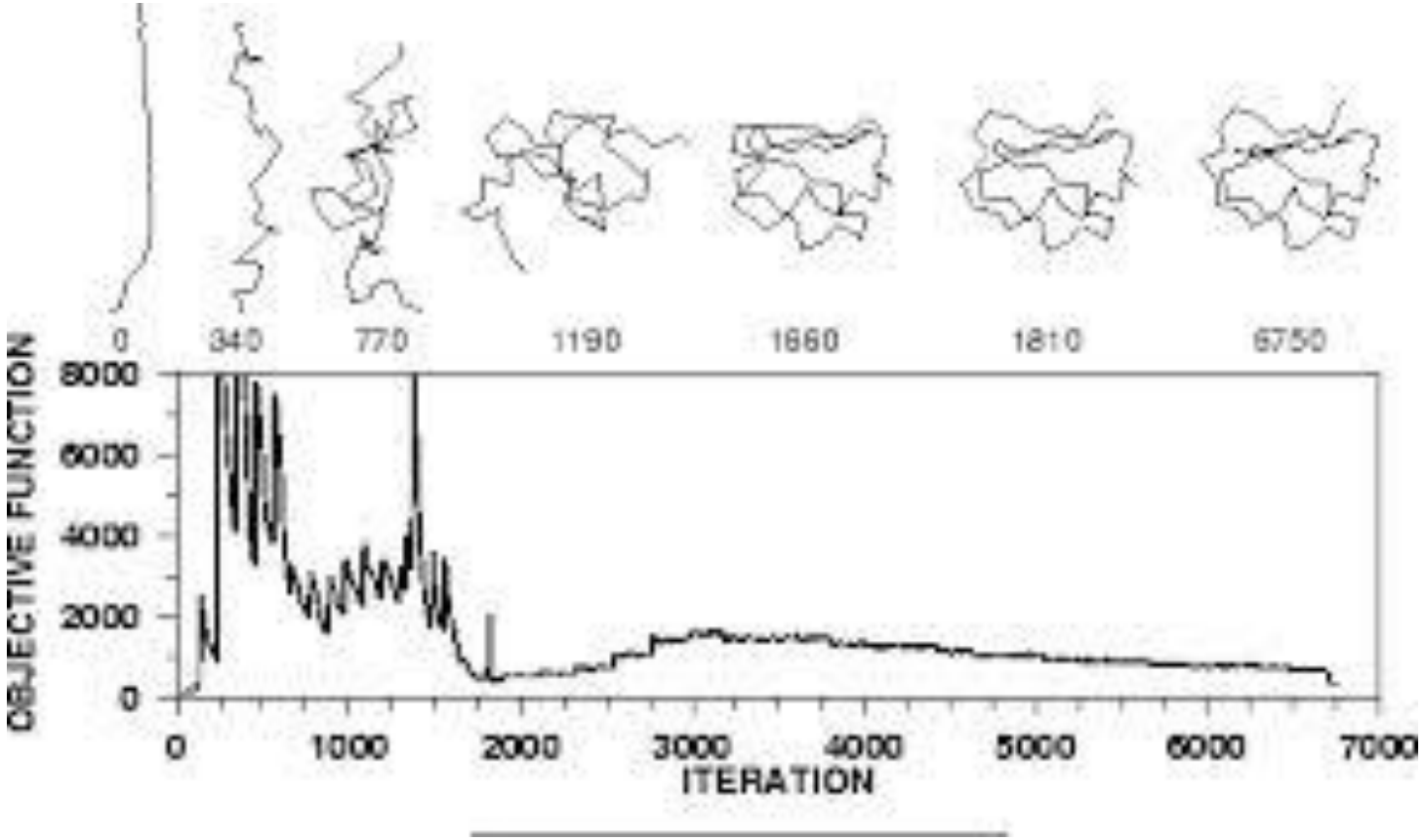
- Поиск белковых структур с последовательностями, гомологичными моделируемой.
- Выравнивание последовательностей.
- Извлечение пространственных ограничений из шаблонов.
- Построение модели с учетом этих ограничений
- Общая оптимизация геометрии.



https://salilab.org/modeller/

Моделирование на основе пространственных ограничений. MODELLER

Начиная с распрямленной конформации или конформации шаблона, выполняется учет все более далеких ограничений, чередующийся с минимизацией энергии **методом сопряженных градиентов**.



Выбор шаблона

Методы, применяемые для сравнения последовательностей:

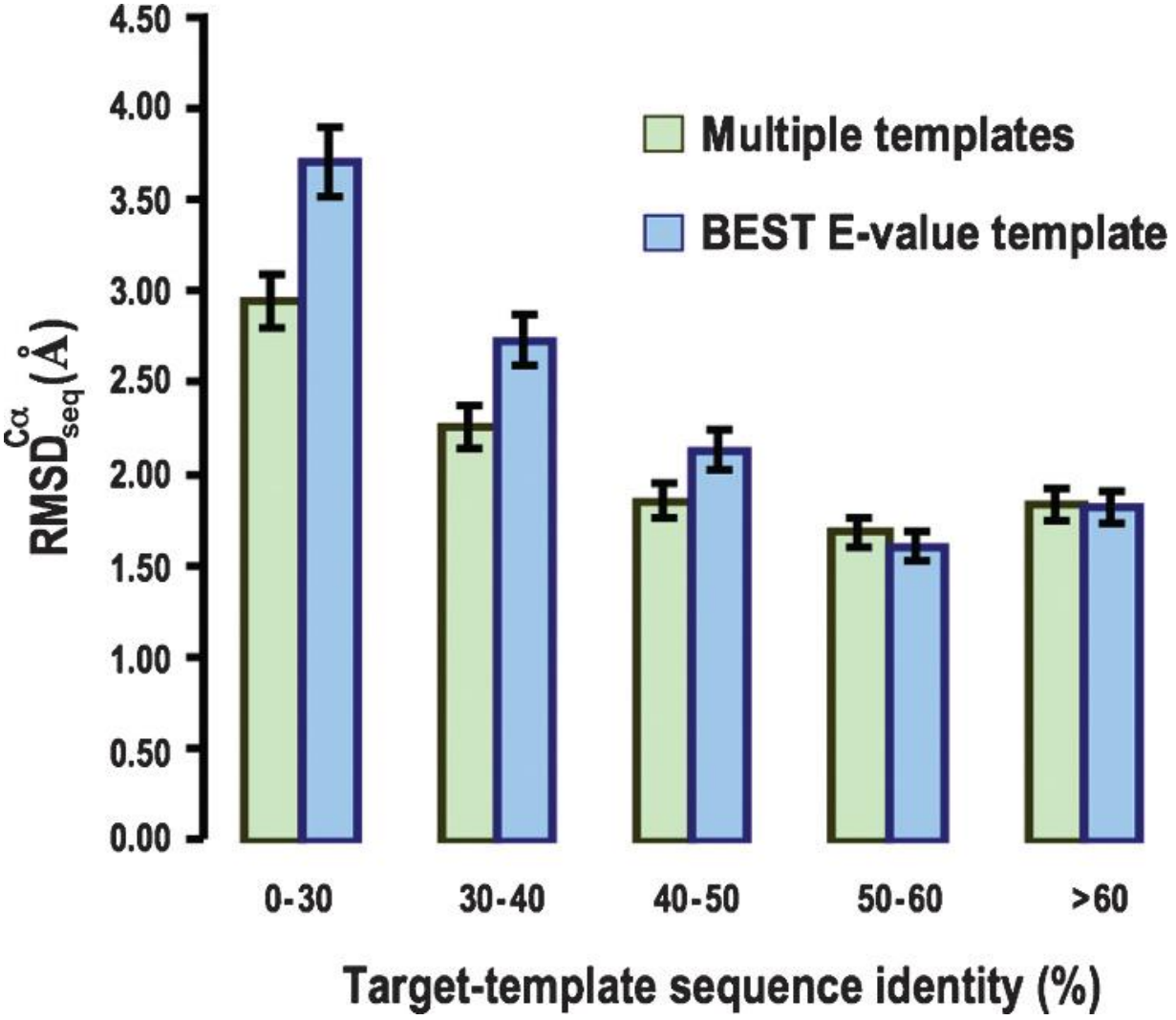
1. Попарное сравнивание моделируемой последовательности с каждой последовательностью из базы данных (**FASTA, BLAST**).
2. Сравнивание сразу нескольких последовательностей (**Clustal, Muscle**).
3. Протягивание последовательности через библиотеку пространственных структур.

Факторы, влияющие на выбор шаблона:

1. Высокая идентичность последовательностей.
2. Белки принадлежат к одному подсемейству.
3. Высокое качество экспериментальной структуры (разрешение или количество ограничений на аминокислотный остаток).

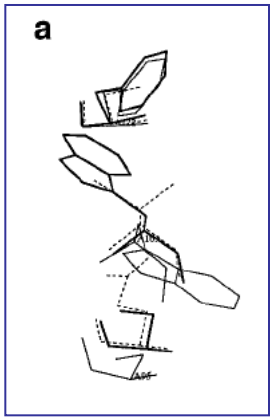
Процент идентичности	Качество выравнивания
40% <	почти всегда высокое
30-40%	«сумеречная зона»
30% >	ошибочно выровненные участки

Качество модели: несколько шаблонов

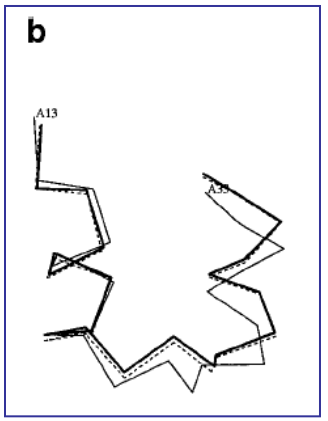


Ошибки построения модели

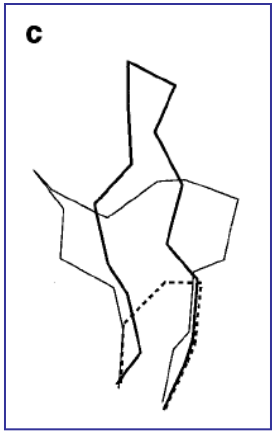
1. Ошибки в ориентации боковых цепей.
Мышиный белок, связывающий ретиноевую кислоту. Тонкая линия – кристалл, толстая линия – модель, пунктир – шаблон (мышиный липид-связывающий белок).



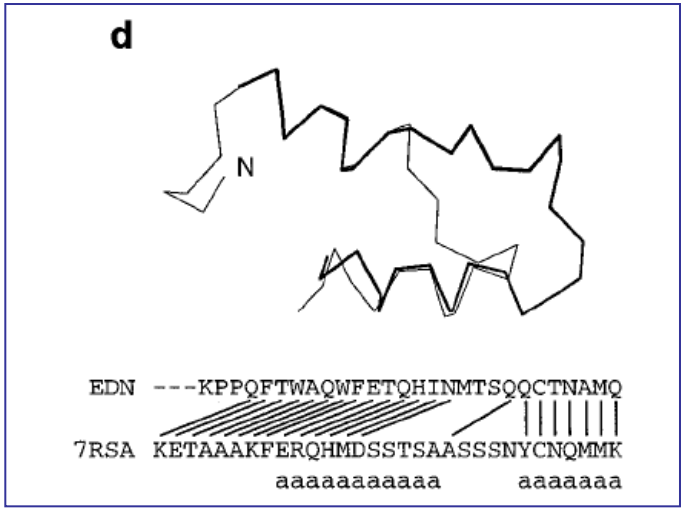
2. Сдвиги в корректно выровненных участках.
Сравнение участка кристаллической структуры мышиного белка, связывающего ретиноевую кислоту, с его моделью и с шаблоном.



3. Ошибки в участках, для которых отсутствует шаблон.
Показан контур Cα атомов остатков 112-117 кристаллографической структуры человеческого эозинофильного нейротоксина (тонкая линия), его модели (толстая линия), и шаблона – рибонуклеаза А (пунктир).



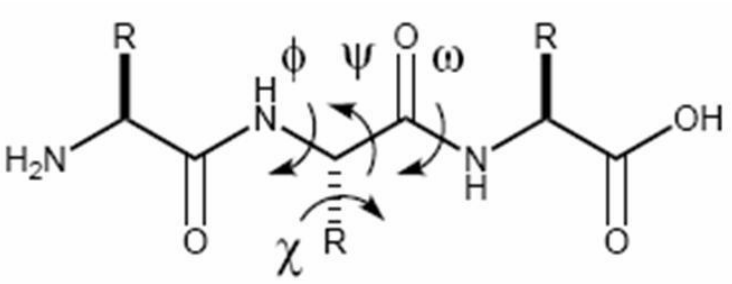
4. Ошибки из-за неправильного выравнивания.
N-концевой участок токсина сравнивается с его моделью. Показан соответствующий участок выравнивания, линии показывают эквивалентные остатки.

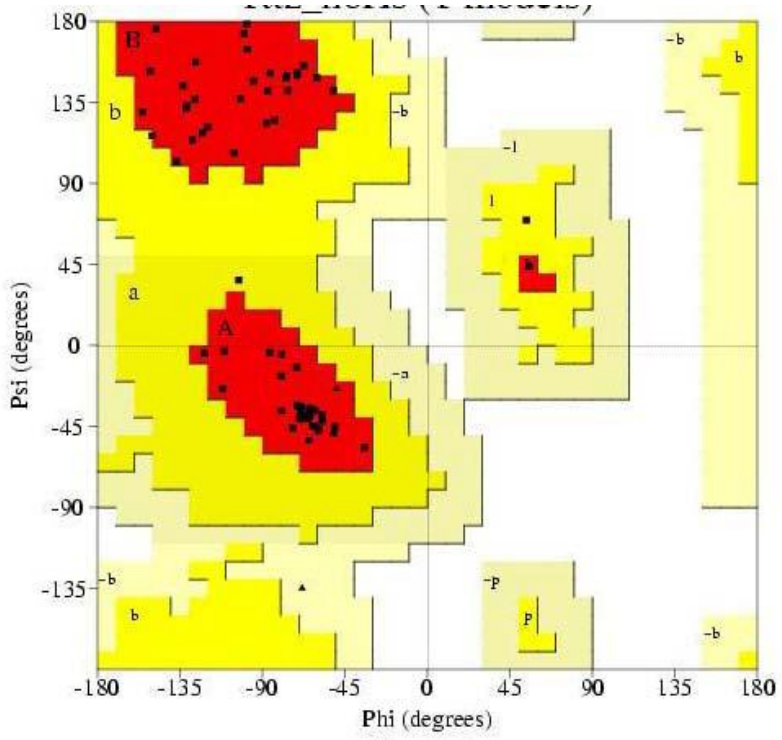


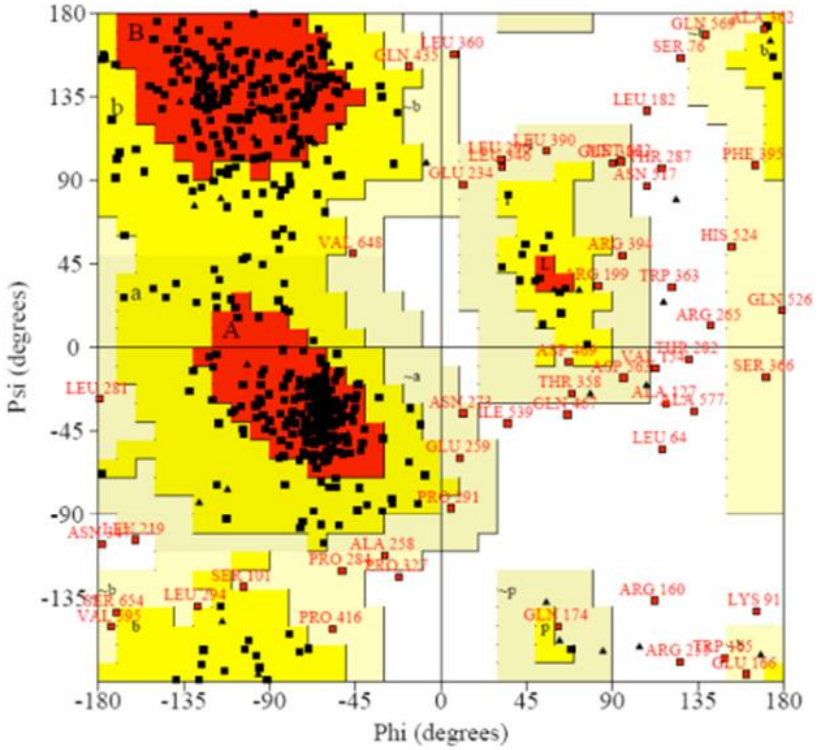
5. Неправильно выбранный шаблон.

Оценка модели. Проверка стереохимии

Карты Рамачандрана





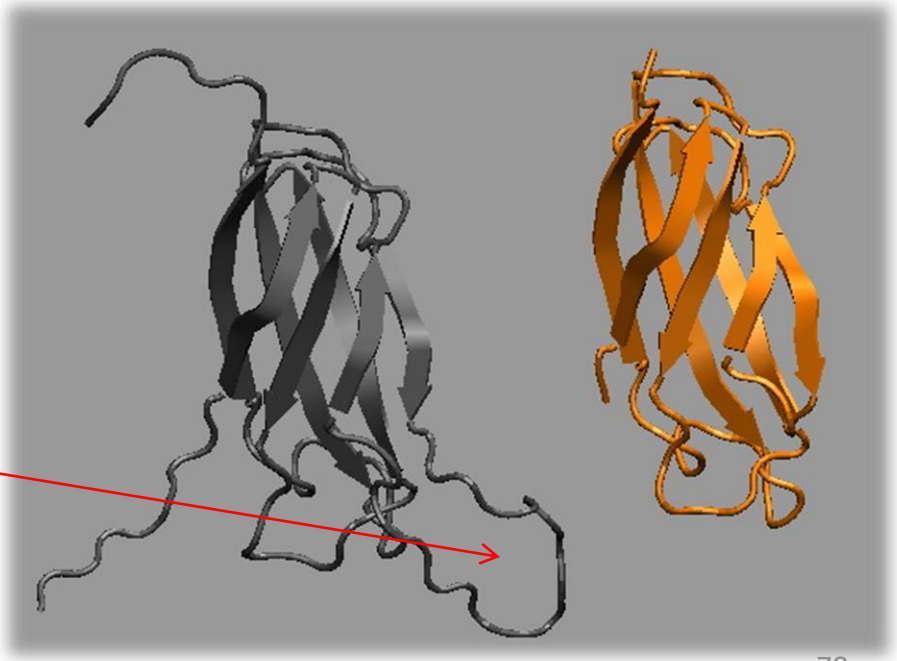


http://molprobity.biochem.duke.edu/index.php - protein structure evaluation

Пример моделирования

```
FN3H   -MQVSDVPTNLEVVAATPTSLLISWYTFTHYG--MNRYYRITYGETGGNS 47
1FNA_A --------RDLEVVAATPTSLLISWDAP-AVT---VRYYRITYGETGGNS 38
               :*************** .          **************

FN3H   PVQEFTVPWINTYTGEPTYADDFKGRFTATISGLKPGVDYTITVYAVTEF 97
1FNA_A PVQEFTVP---------------GSKSTATISGLKPGVDYTITVYAVTGR 73
       ********                ********************

FN3H   SGTGDFDYPISINYRTLEHHHHHH 121
1FNA_A GDSPASSKPISINYRTEI------ 91
              ********
```

Вставка

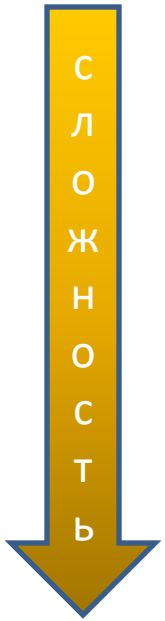


Предсказание структуры белков

Сворачивание белка в уникальную конформацию наводит на мысль об алгоритме формирования структуры белка по его последовательности, но доказательством полноты и правильности нашего понимания могла бы стать его реализация в виде компьютерной программы...

Методы предсказания структуры по последовательности:

- **Предсказание вторичной структуры;**
- **Предсказание топологии;**
- **Моделирование по гомологии;**
- Распознавание фолда;
- Априорное предсказание новых типов укладки.

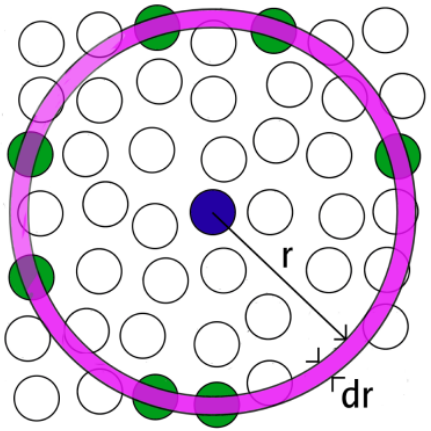


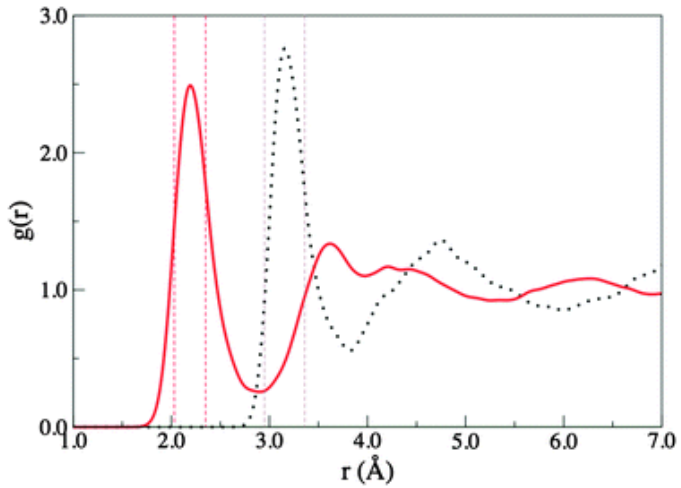
Распознавание фолда. «Протягивание»

Если для последовательности нет гомолога с известной структурой, то, возможно, есть хотя бы структура, подходящая для данной последовательности?
Фактически, нужно примерить данную последовательность на все типы укладки и выбрать наиболее подходящую - **метод «протягивания»** (threading) состоит в построении большого числа грубых моделей для данной последовательности и их последующей экспресс-оценки.

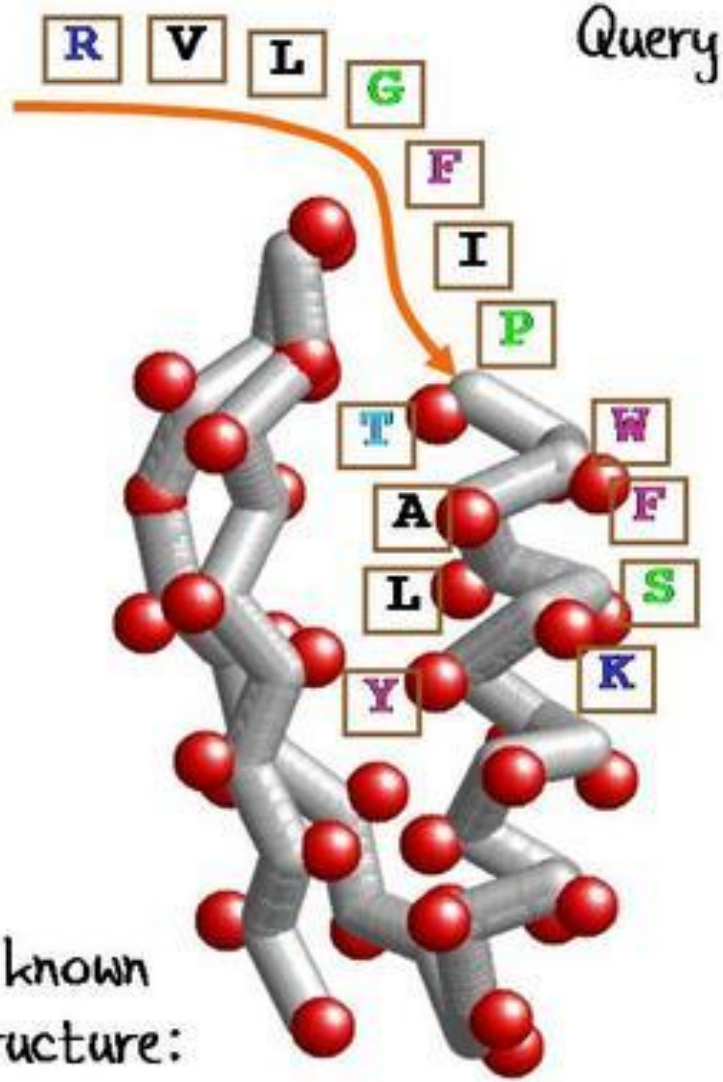
⇒ **нужна функция оценки соответствия последовательности и фолда**.

Например, функции распределения вероятности парных расстояний между остатками (например, по Cβ-атомам) (20x20 штук). Соотнося расстояния в моделях с этими функциями, можно оценить насколько вероятны как эти расстояния, так и модели в целом.





WHAT IS THREADING



A known structure:

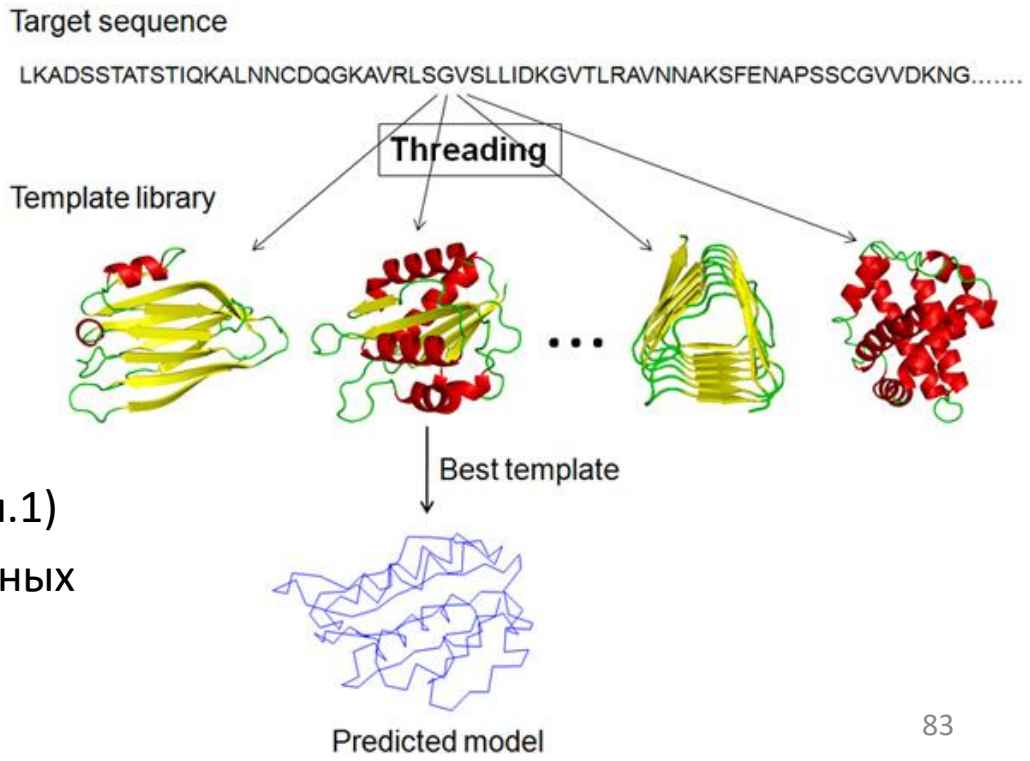
Query Sequence: RVLGFIPTWFALSKY

- Thread the sequence onto the structure.
- Use structural properties to evaluate the fit:
  - Local structure
  - Environment
  - Pairwise interactions

Распознавание фолда. «Протягивание»

1. Выбор некой известной структуры в качестве потенциального шаблона
2. Генерация всевозможных выравниваний последовательности шаблона с новой последовательностью

```
...IIAWLVKEKKVDVIV...  ...IIAWLVK-EKKVDVIV...  ...IIAWLVKEKKVDVIV...
...NGLELVLDSVLDATF...  ...NGLELVLDSVLDATF...   ...NGLELVLD-SVLDATF...
       **     *               **                       **    *
```

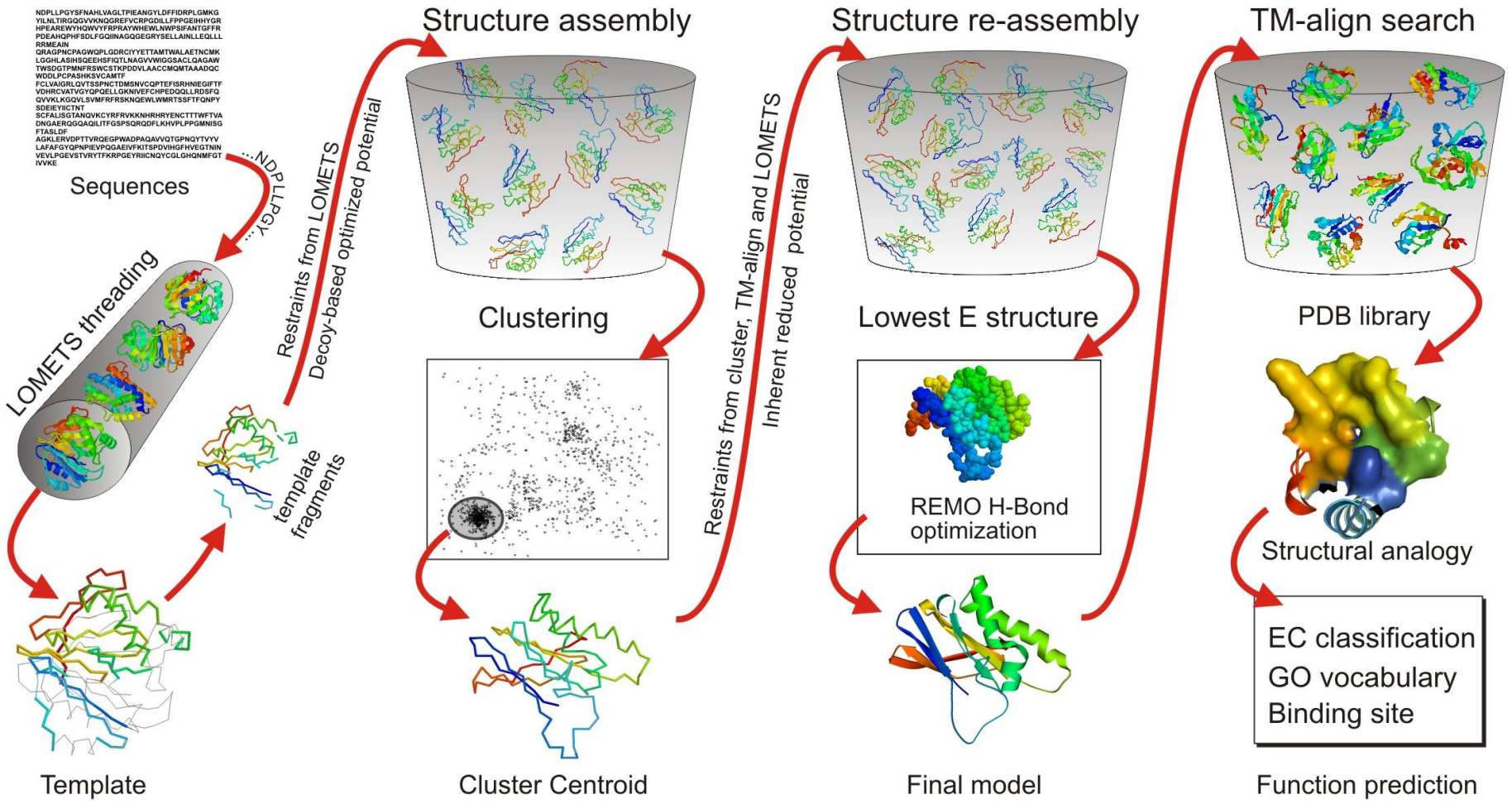


3. Построение и оценка моделей
4. Переход к следующему шаблону (п.1)
5. Сопоставление моделей, построенных по различным шаблонам, и выбор оптимальной модели

I-TASSER

Protein Structure & Function Predictions

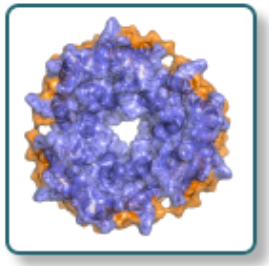
(The server completed predictions for 390328 proteins submitted by 94188 users from 138 countries)
(The template library was updated on 2018/04/02)



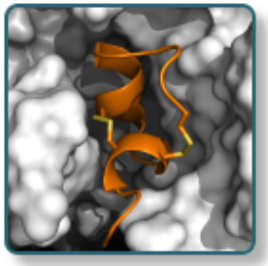
Welcome to ROSIE

Rosetta Online Server that Includes Everyone

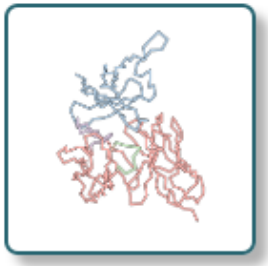
Rosetta Protocols opened for academic users:



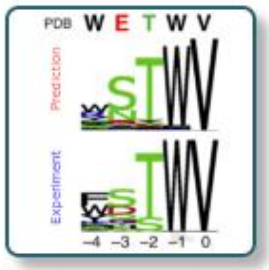
[Mp_lipid_acc]



[Tox_dock]

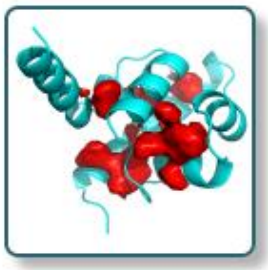


[Snug_dock]



[Sequence_tolerance]

[Antibody]



[Vip]

ROSIE stats (24hrs):

Users:	4,815	+1
Jobs:	31,163	+28
CPU hours:	3,521,032	+5,956

See more info at our About page.

Get Started with ROSIE

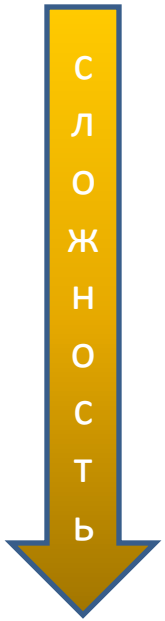
- ROSIE Documentation – Server related documentation and info.
- Rosetta Forums This is a list of forums for Rosetta users to discuss problems with running Rosetta and is monitored by Rosetta developers.

Предсказание структуры белков

Сворачивание белка в уникальную конформацию наводит на мысль об алгоритме формирования структуры белка по его последовательности, но доказательством полноты и правильности нашего понимания могла бы стать его реализация в виде компьютерной программы...

Методы предсказания структуры по последовательности:

- **Предсказание вторичной структуры;**
- **Предсказание топологии;**
- **Моделирование по гомологии;**
- **Распознавание фолда;**
- Априорное предсказание новых типов укладки.



Critical Assessment of protein Structure Prediction (CASP)

Protein Structure Prediction Center

Menu

- Home
- PC Login
- PC Registration
- CASP Experiments
  - *CASP Commons (COVID-19, 2020)*
  - CASP14 (2020)
  - CASP13 (2018)
  - CASP12 (2016)
  - CASP11 (2014)
  - CASP10 (2012)
  - CASP9 (2010)
  - CASP8 (2008)
  - CASP7 (2006)
  - CASP6 (2004)
  - CASP5 (2002)
  - CASP4 (2000)
  - CASP3 (1998)
  - CASP2 (1996)
  - CASP1 (1994)
- Initiatives
- Data Archive
- Local Services
- Proceedings

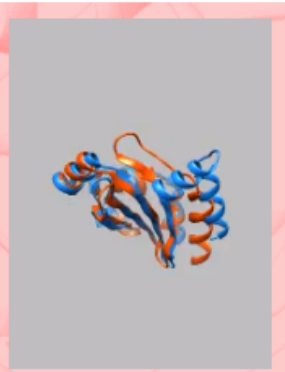
Success Stories From Recent CASPs

template-based modeling | **ab initio modeling** | contact prediction | help structural biologists | refinement | data-assisted modeling

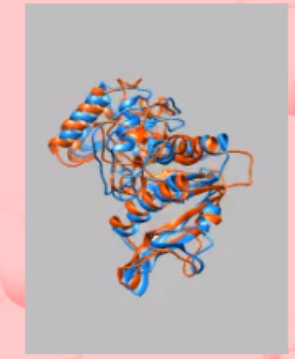
ab initio modeling

Modeling proteins with no or marginal similarity to existing structures (*ab initio, new fold, non-template* or *free* modeling) is the most challenging task in tertiary structure prediction. Probably the first ab initio model of reasonable accuracy was built in CASP4. Since then CASP witnessed sustained progress in ab initio prediction, but mainly for small proteins (120 residues or less, panels 1 and 2; models are in blue, targets in orange). In CASP11 for the first time a larger new fold protein (256 residues, sequence identity to known structures <5%) was built with unpresedented before accuracy for targets of this size (panel 3). CASP11 and CASP12 experiments (2014, 2016) also showed a new trend in building better non-template models by successful utilizing predicted contacts (panel 4) [Abriata et al, 2018]. CASP13 witnessed yet another substantial improvement in accuracy of template-free models mainly due to employing advanced deep learning artificial intelligence techniques coupled with prediction of inter-residue distances at a range of thresholds [Senior et al, 2019], [Xu and Wang, 2019]. The best models submitted on the free modeling targets showed more than 20% increase in accuracy of the backbone, with the average GDT_TS scores going up from 52.9 in CASP12 to 65.7 in CASP13.

CASP7: T0283-D1
model 321_1: GDT_TS=75



CASP9: T0581-D1
model 170_1: GDT_TS=71



CASP11: T0806-D1
model 064_1: GDT_TS=61

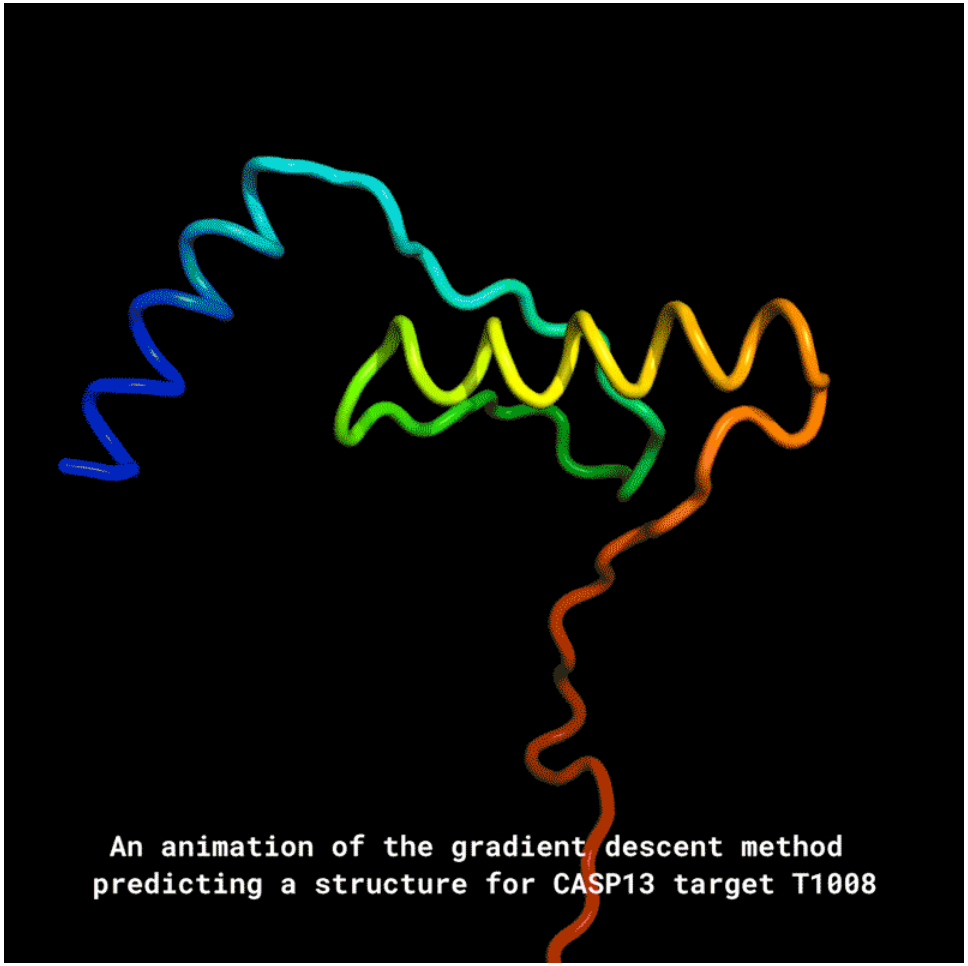
CASP12: T0866-D1
model 325_5: GDT_TS=81

Rosetta@home

FoldIt


fold it BETA
09:33:42 GMT
Solve Puzzles for Science
PUZZLES
BLOG
CATEGORIES
FEEDBACK
GROUPS
FORUM
PLAYERS
WIKI
FAQ
RECIPES
ABOUT
CONTESTS
CREDITS
Click to learn how you contribute to science by playing Foldit.
GET STARTED: DOWNLOAD
Win Beta
Mac Beta
Linux Beta
Windows (7/8/10)
OSX (10.7 or later)
Linux (64-bit)
Are you new to Foldit? Click here.
Are you a student? Click here.
Are you an educator? Click here.
SEARCH
Google Search
Only search fold.it
What's New
Special update on coronavirus puzzles
RECOMMEND FOLDIT
Send

AlphaFold



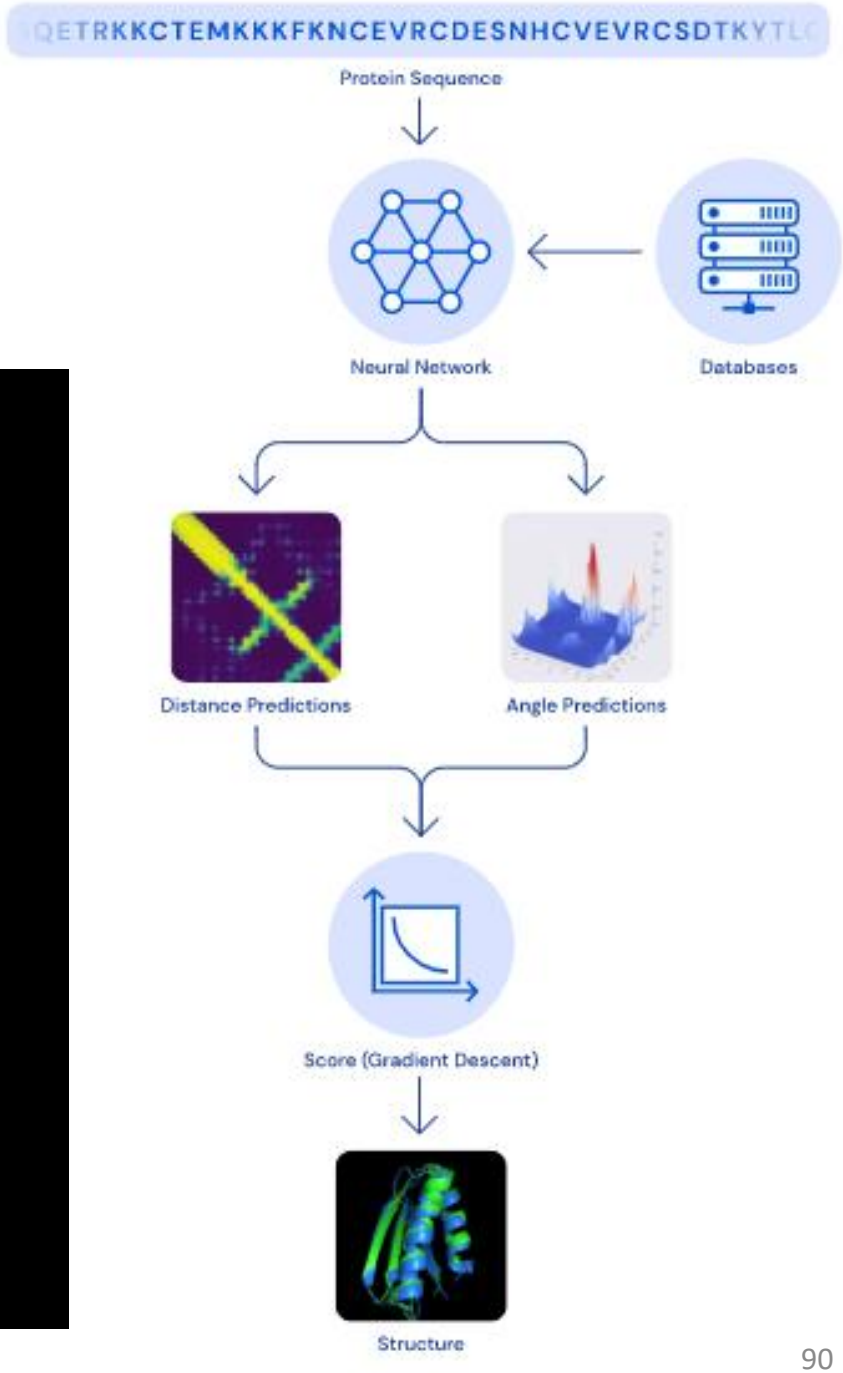
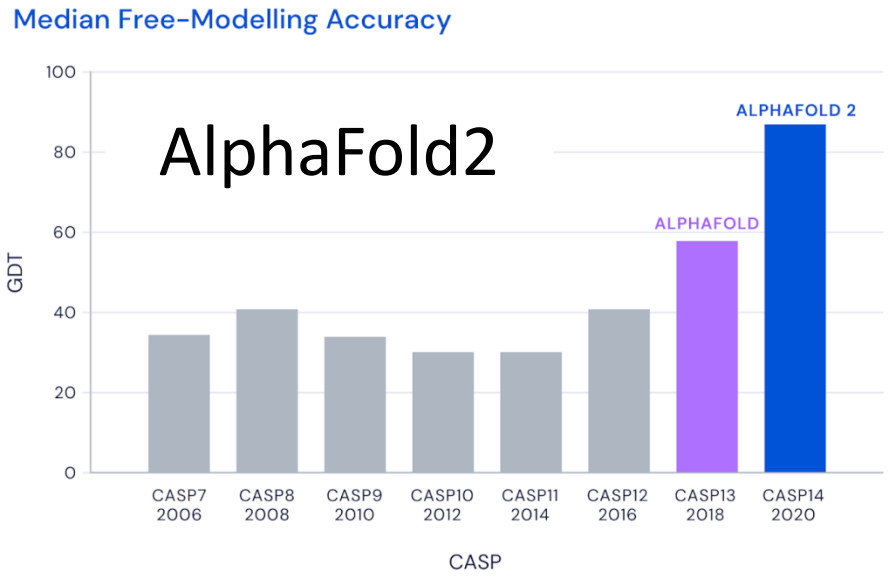
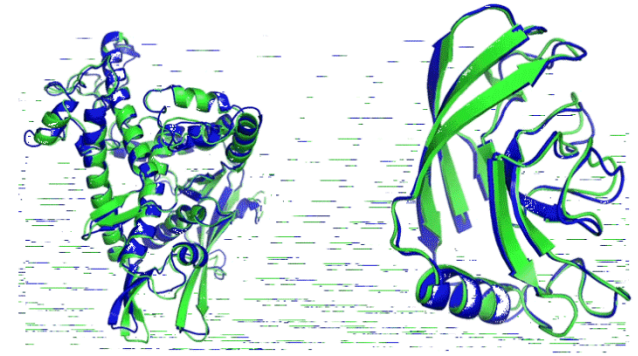
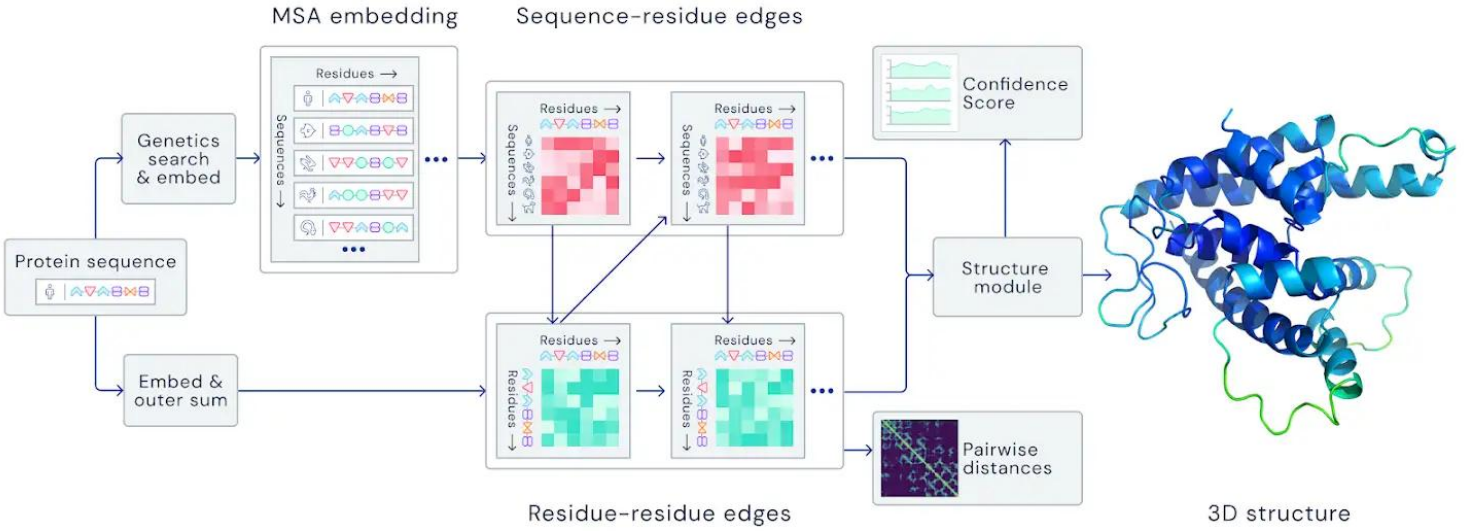


FIGURE 3: A SCHEMATIC OF THE ARCHITECTURE OF THE ALPHAFOLD SYSTEM PREDICTING STRUCTURE FROM PROTEIN SEQUENCE.







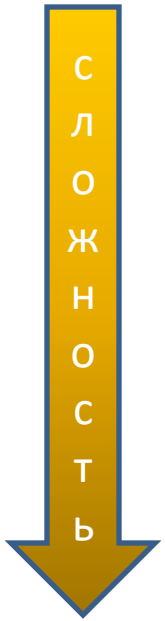
Convergence for a single sequence has been estimated to require on the order of $10,000 worth of wholesale compute time.

Предсказание структуры белков

Сворачивание белка в уникальную конформацию наводит на мысль об алгоритме формирования структуры белка по его последовательности, но доказательством полноты и правильности нашего **понимания** могла бы стать его **реализация в виде компьютерной программы...**

Методы предсказания структуры по последовательности:

- **Предсказание вторичной структуры;**
- **Предсказание топологии;**
- **Моделирование по гомологии;**
- **Распознавание фолда;**
- **Априорное предсказание новых типов укладки.**



Благодарю за внимание!