

ВВЕДЕНИЕ В БИОИНФОРМАТИКУ

Лекция №19

Структурная биоинформатика

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

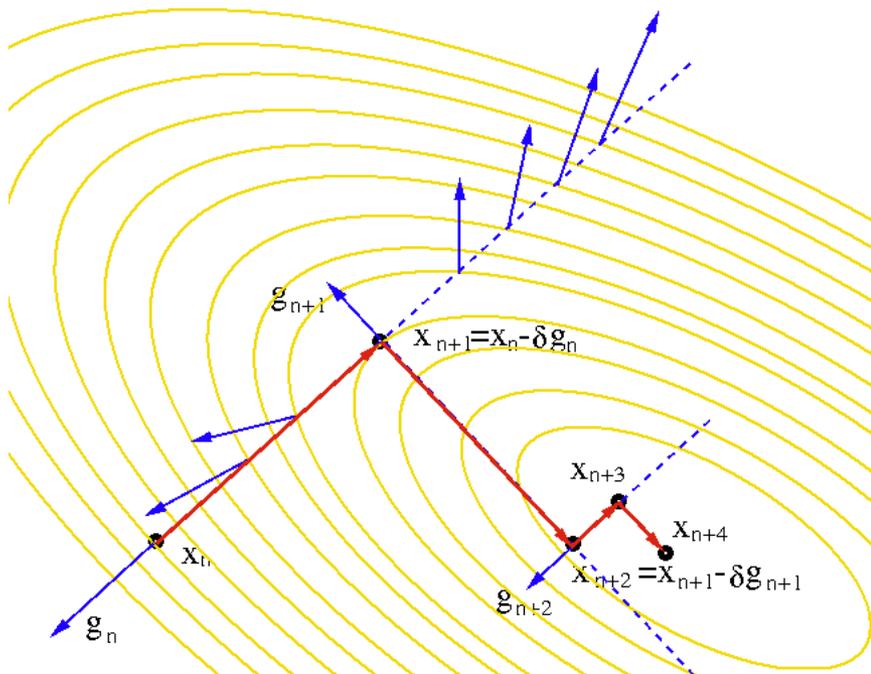
Сайт курса <http://intbio.org/bioinf2019-2020>

Методы минимизации. Градиентный спуск

Метод нахождения локального минимума функции с помощью движения вдоль градиента.

Пусть есть некая функция $F(\mathbf{x})$, для которой мы хотим найти такое значение \mathbf{x} , что функция принимает **минимальное значение**. Основная идея метода заключается в том, чтобы идти в направлении наискорейшего спуска, а это направление задаётся антиградиентом $-\nabla F$:

$$\vec{x}^{[j+1]} = \vec{x}^{[j]} - \lambda^{[j]} \nabla F(\vec{x}^{[j]})$$



Величина шага может быть

- постоянной

- дробной

- определяемой на каждой

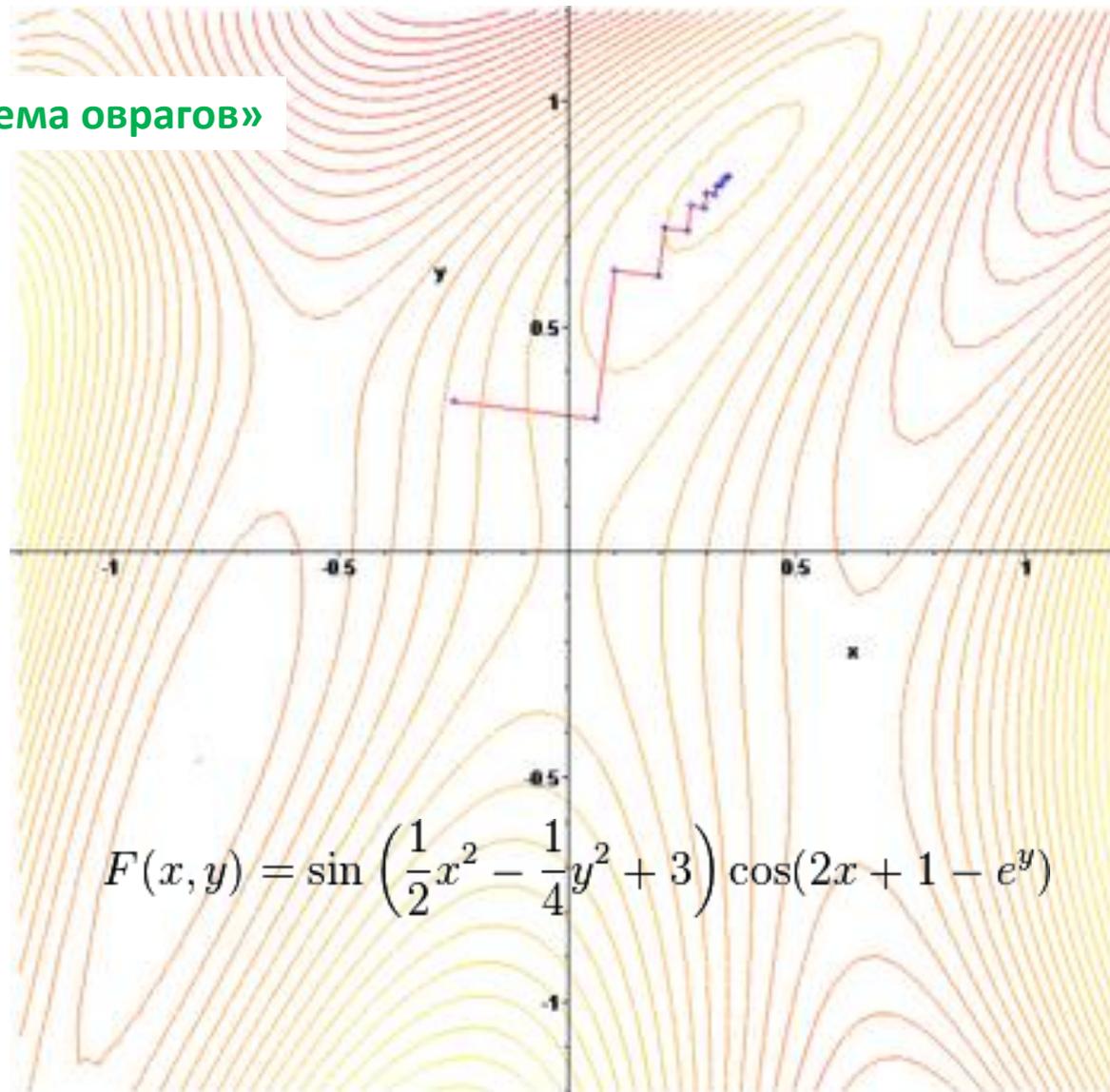
итерации заново

(метод наискорейшего спуска,

steepest descent)

Методы минимизации. Градиентный спуск

«Проблема оврагов»

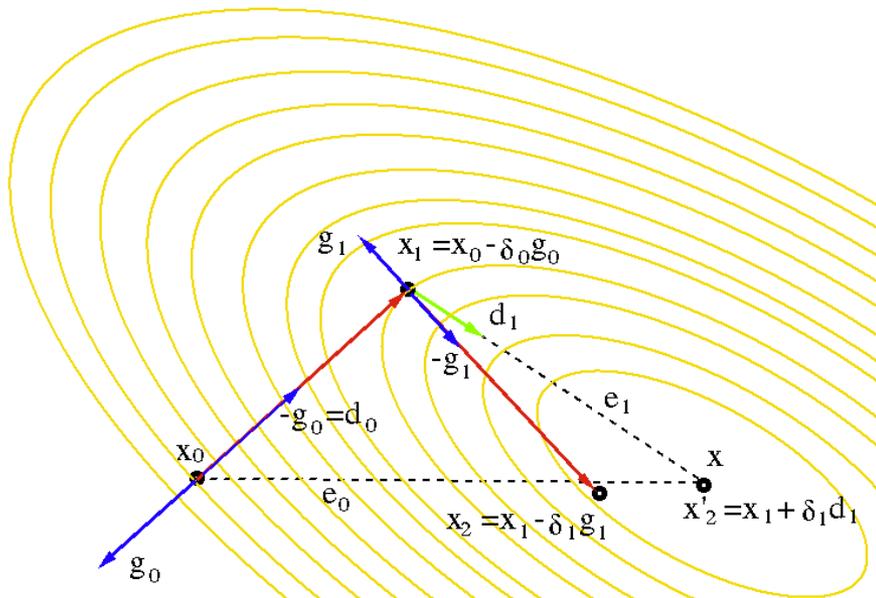


Методы минимизации. Метод сопряженных градиентов

(conjugated gradients)

Метод нахождения локального минимума функции, при котором направление и величина шага рассчитываются с учетом градиента в текущей точке и информации о предыдущем шаге:

$$\vec{S}_k = -\nabla f(\vec{x}_k) + \omega_k \vec{S}_{k-1}, \quad \omega_i = \frac{\|\nabla f(\vec{x}_i)\|^2}{\|\nabla f(\vec{x}_{i-1})\|^2}$$



В случае квадратичной функции от n переменных

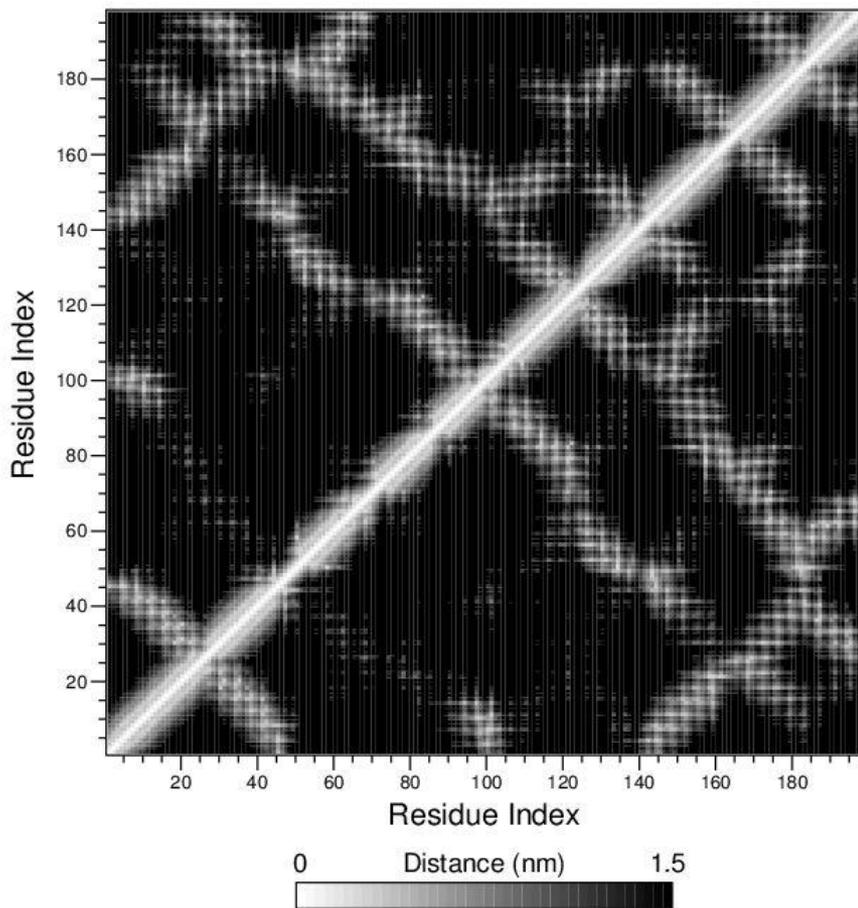
$$f(\vec{x}) = \vec{x}^T A \vec{x} + \vec{b} \cdot \vec{x} + c.$$

метод гарантированно сходится за n шагов.

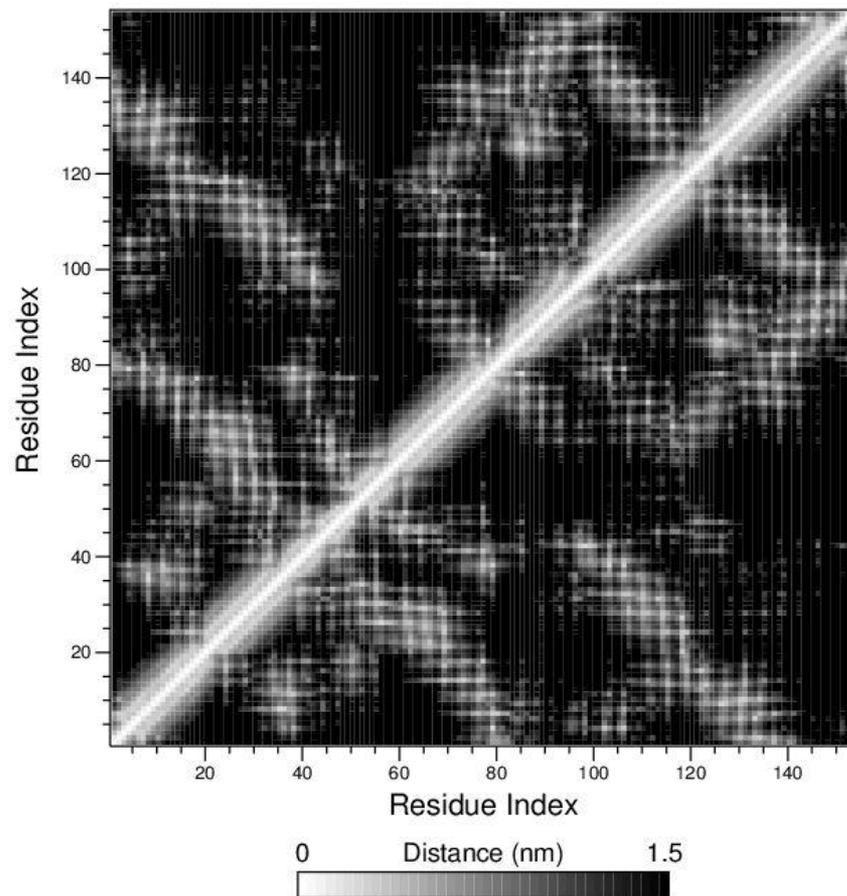
Сравнение структур

Если структуры схожи, то сохраняются паттерны контактов между остатками
⇒ анализируя матрицы расстояний (Ca-Ca), можно распознать схожие структуры

colicin



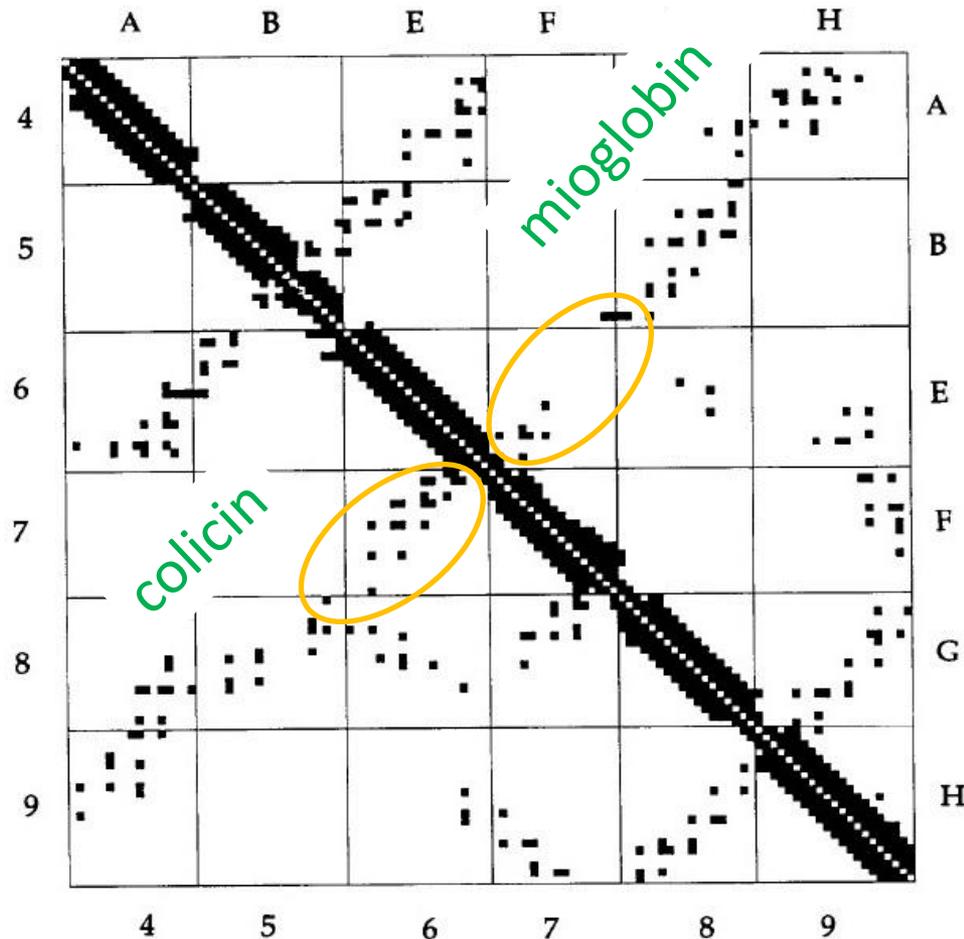
mioglobin



Сравнение структур

Выравнивание матриц расстояний – программа DALI (Distance-matrix ALIngment)

(Holm & Sander, 1993)

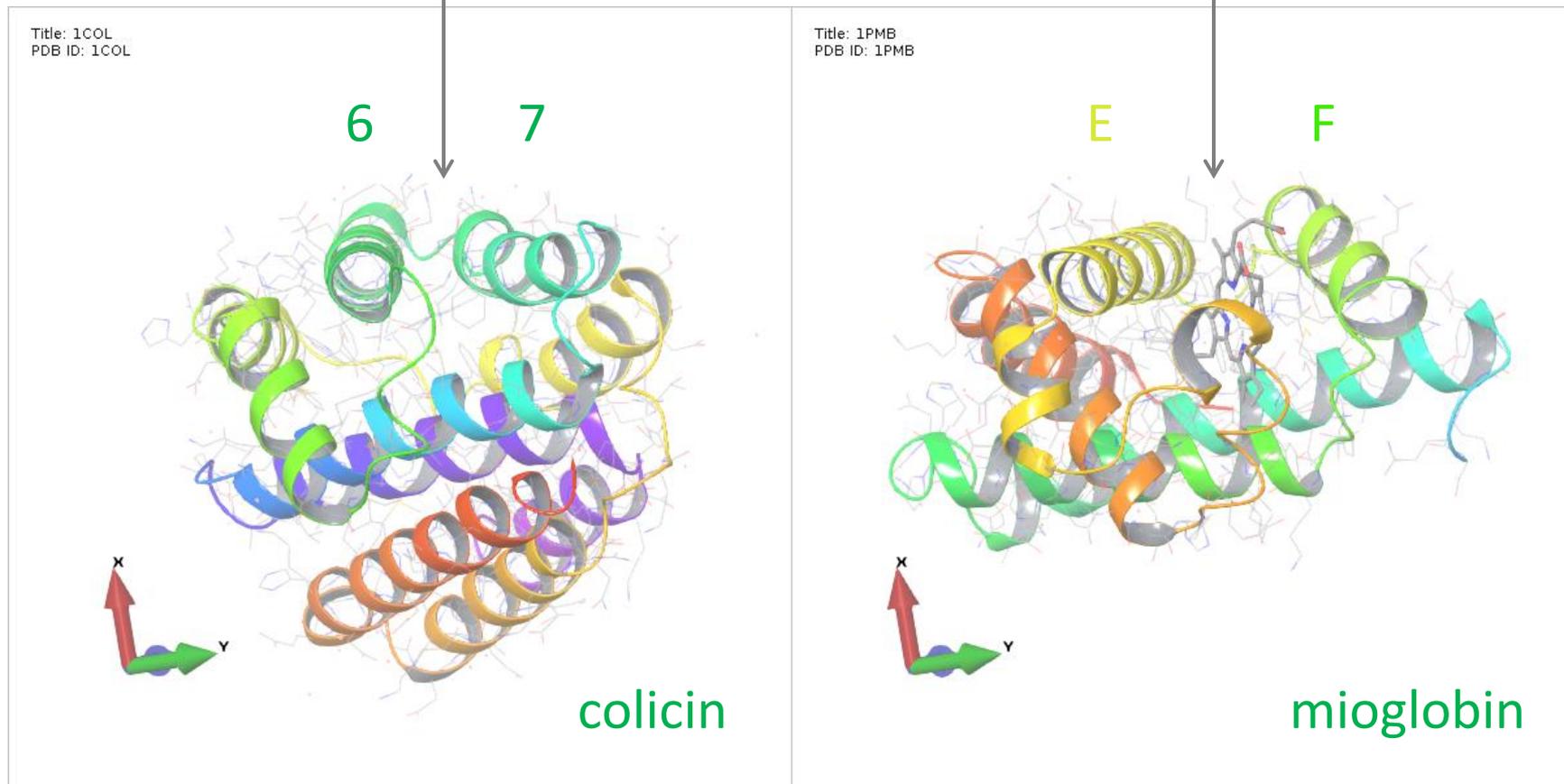


Особенности: свободная топология обнаруженных сходств в структуре, в т.ч. и «обратных» фрагментов.

Сравнение структур

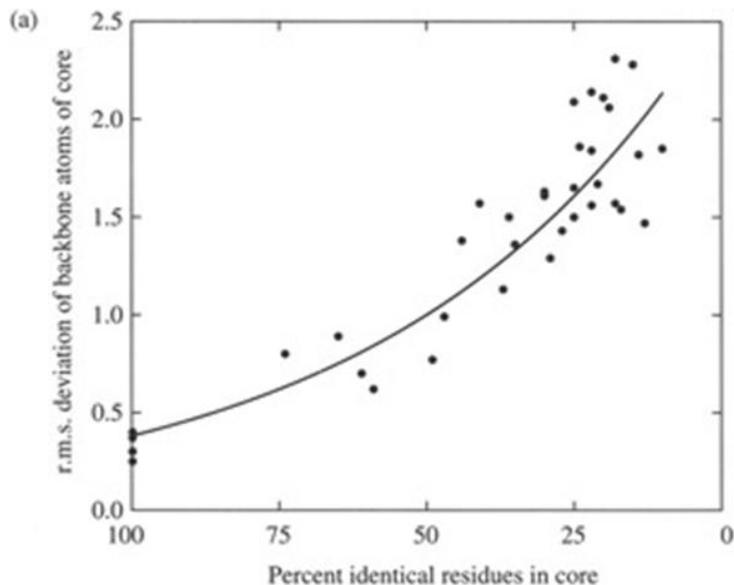
Есть контакт между спиралями

Гем; нет контакта между спиралями



Эволюция белковых структур

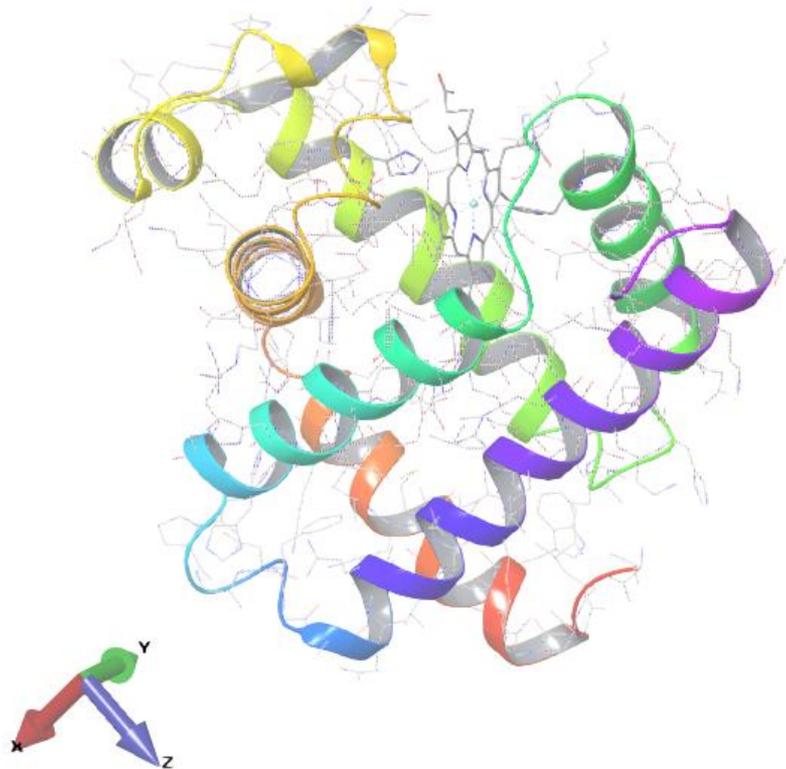
Вариации, встречающиеся в семействах гомологичных белков с одинаковой функцией, показывают, как структура приспосабливается к изменениям в последовательности: **структура устойчива к мутациям.**



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.

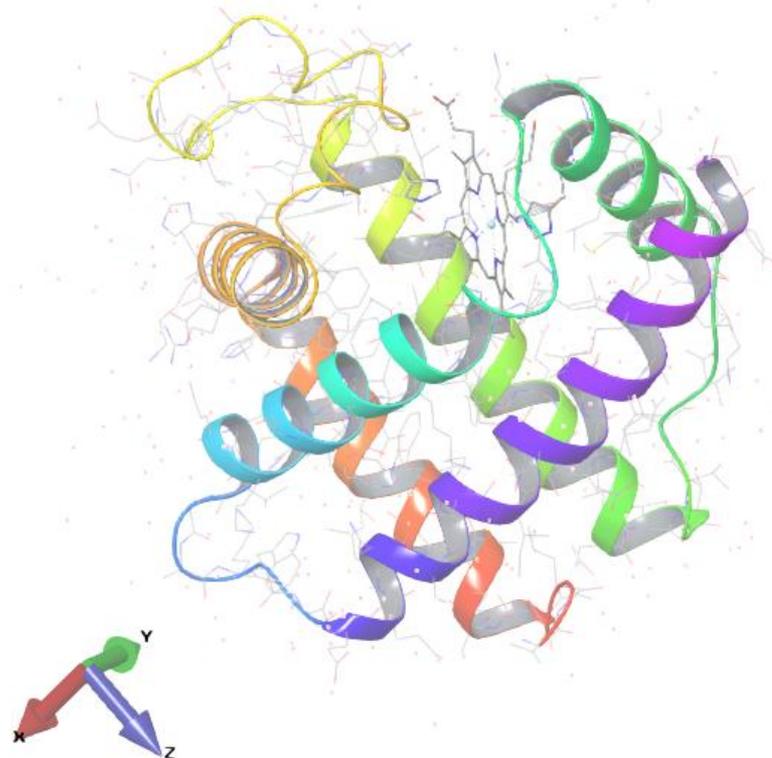
Эволюция белковых структур

Title: 1MBN
PDB ID: 1MBN



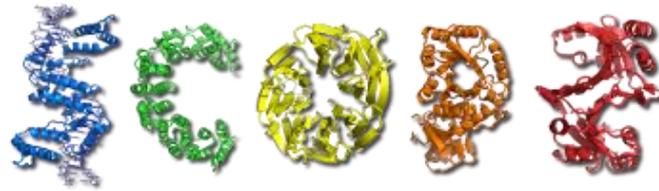
Миоглобин кашалота (1mbn, 1969)

Title: 1GDJ
PDB ID: 1GDJ



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

Классификация структур белков. SCOP

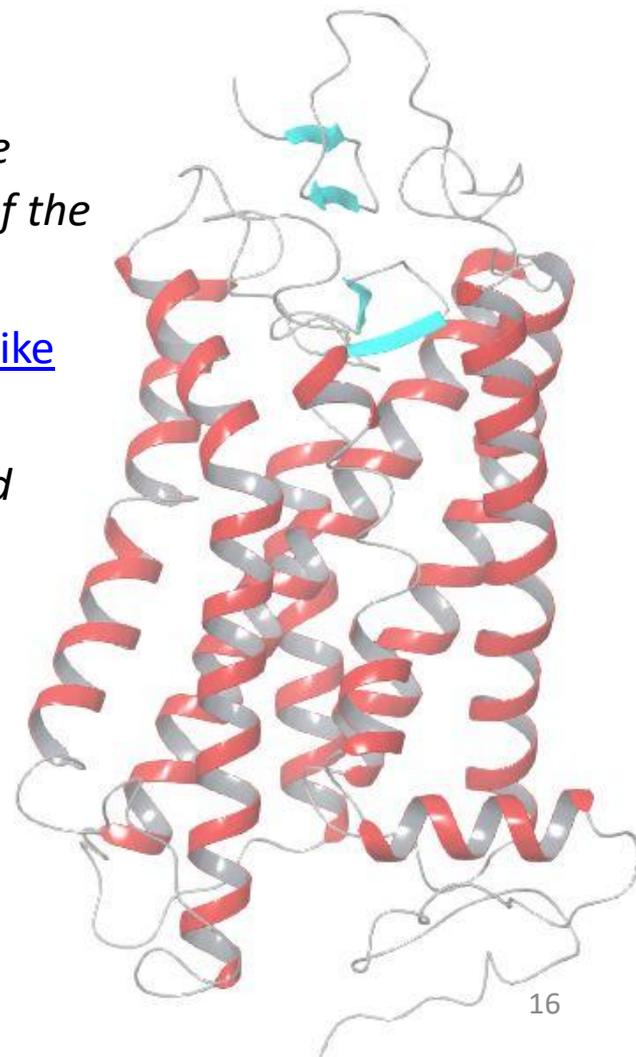


SCOP (Structural classification of proteins) (Murzin et al., 1994 - 2009) – организация структур для отображения их эволюционного происхождения и структурного сходства. Основные уровни организации:

- **Классы:** α , β , $\alpha+\beta$, 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

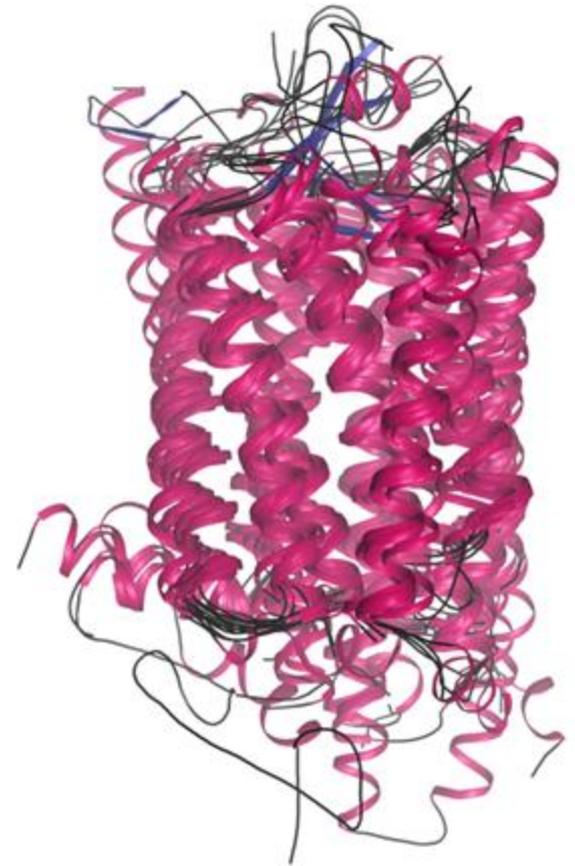
Topology – нечеткий уровень, объединяющий фолды с характерными особенностями

Homologous superfamily – эквивалентно уровню «надсемейство» в SCOP

CATH Superfamily 1.20.1070.10

Rhodopsin 7-helix transmembrane proteins

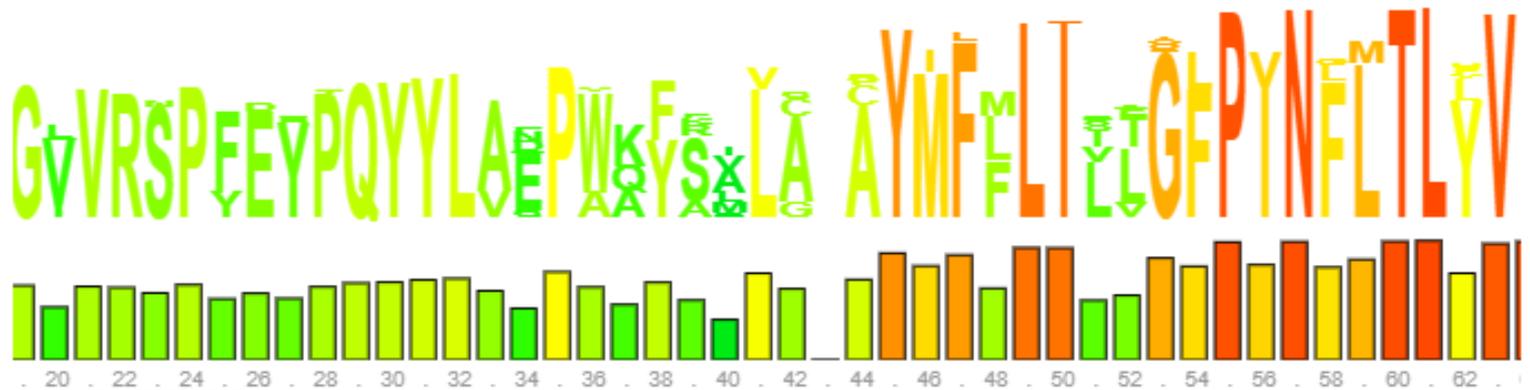
Хорошо заметно «ядро»



Классификация структур белков. CATH

CATH Superfamily 1.20.1070.10

Rhodopsin 7-helix transmembrane proteins



Label
 A0A087YRE2
 P15409
 P08100
 A0A0N4SV48
 A0A0N4SUP8
 P02699
 P51489
 P35359
 P22328
 F1P3Y2
 G1ND14
 A0A091VGC1
 O57450
 A0A093Q1T7
 D4QGQ8

G	I	V	R	S	P	Y	E	Y	P	Q	Y	Y	L	V	S	P	A	A	Y	A	C	L	G	-	A	Y	M	F	F	L	I	L	V	G	F	P	I	N	F	L	T	L	Y	V			
G	V	V	R	S	P	F	E	Q	P	Q	Y	Y	L	A	E	P	W	Q	F	S	M	L	A	-	A	Y	M	F	L	L	I	V	L	G	F	P	I	N	F	L	T	L	Y	V			
G	V	V	R	S	P	F	E	Y	P	Q	Y	Y	L	A	E	P	W	Q	F	S	M	L	A	-	A	Y	M	F	L	L	I	V	L	G	F	P	I	N	F	L	T	L	Y	V			
G	T	V	V	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
G	V	V	R	S	P	F	E	A	P	Q	Y	Y	L	A	E	P	W	Q	F	S	M	L	A	-	A	Y	M	F	L	L	I	M	L	G	F	P	I	N	F	L	T	L	Y	V			
G	V	V	R	S	P	F	E	Q	P	Q	Y	Y	L	A	E	P	W	Q	F	S	M	L	A	-	A	Y	M	F	L	L	I	V	L	G	F	P	I	N	F	L	T	L	Y	V			
G	V	V	R	S	P	Y	E	Y	P	Q	Y	Y	L	V	A	P	W	A	Y	G	L	L	A	-	A	Y	M	F	F	L	I	I	T	G	F	P	V	N	F	L	T	L	Y	V			
G	V	V	R	S	P	F	E	Y	P	Q	Y	Y	L	A	E	P	W	K	F	S	A	L	A	-	A	Y	M	F	M	L	I	L	L	G	F	P	V	N	F	L	T	L	Y	V			
G	V	V	R	S	P	F	E	Y	P	Q	Y	Y	L	A	E	P	W	K	F	S	A	L	A	-	A	Y	M	F	M	L	I	L	L	G	F	P	V	N	F	L	T	L	Y	V			
G	V	V	R	T	P	F	E	Y	P	Q	Y	Y	L	A	E	P	W	K	Y	S	G	L	A	-	A	Y	M	F	M	L	I	L	L	G	F	P	I	N	F	L	T	L	Y	V			
G	V	V	R	S	P	F	E	Y	P	Q	Y	Y	L	A	E	P	W	K	F	S	A	L	A	-	A	Y	M	F	M	L	I	L	L	G	F	P	I	N	F	L	T	L	Y	V			
G	V	V	R	S	P	F	E	Y	P	Q	Y	Y	L	A	E	P	W	K	F	S	A	L	A	-	A	Y	M	F	M	L	I	L	L	G	F	P	I	N	F	L	T	L	Y	V			

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

Поиск белков со схожей структурой

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.

##	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	<input checked="" type="checkbox"/>	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	<input checked="" type="checkbox"/>	HIGH RESOLUTION CRYSTAL STRUCTURE OF HUMAN B2-ADRENERGIC G PROTEIN- COUPLED RECEPTOR.

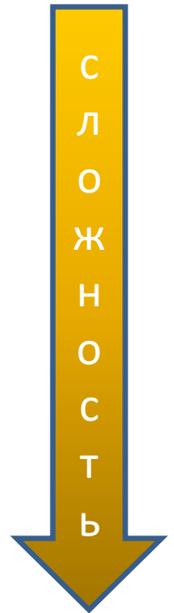
Examined 1 entry, (1 chain). Displaying Matches 1-2 of 2.

Предсказание структуры белков

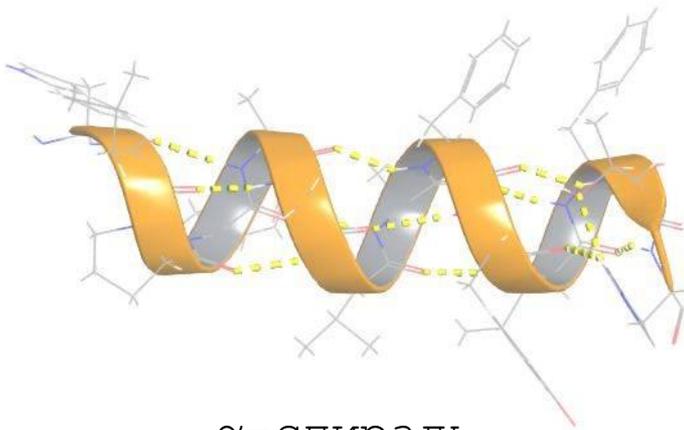
Сворачивание белка в уникальную конформацию наводит на мысль об алгоритме формирования структуры белка по его последовательности, но доказательством полноты и правильности нашего понимания могла бы стать его **реализация в виде компьютерной программы...**

Методы предсказания структуры по последовательности:

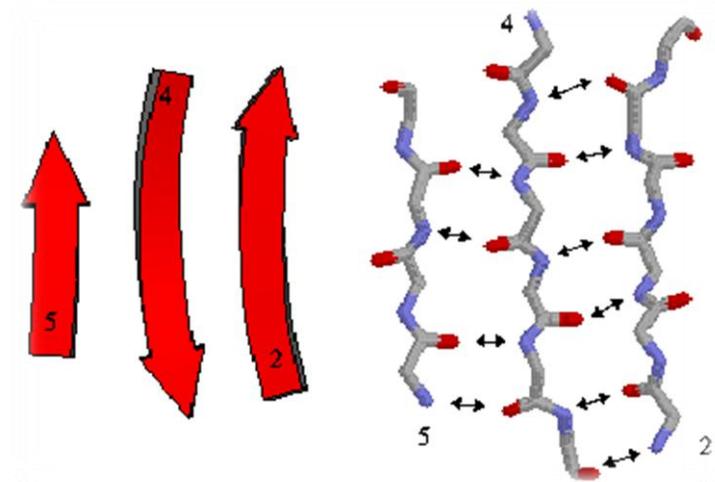
- Предсказание вторичной структуры;
- Моделирование по гомологии;
- Распознавание типов укладки (по известной библиотеке фолдов);
- Априорное предсказание новых типов укладки.



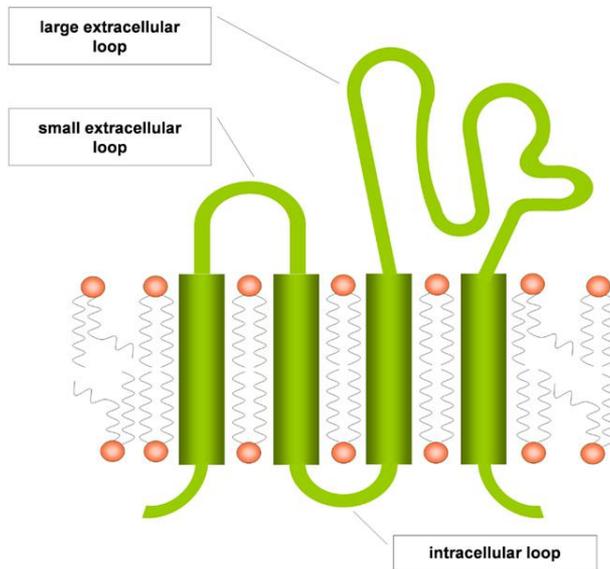
Типы вторичной структуры белков



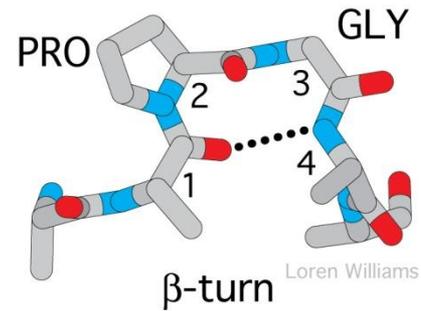
α -спираль



β -лист, состоящий из β -тяжей



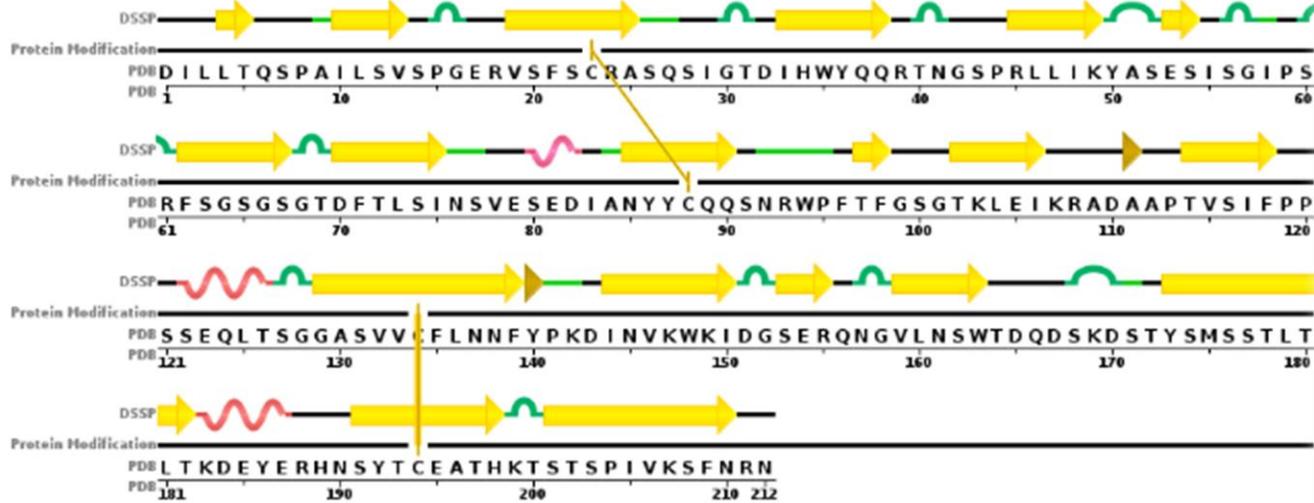
ПЕТЛЯ



ПОВОРОТ

Типы вторичной структуры белков

Sequence Chain View



Sequence Chain View



Предсказание вторичной структуры

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

	<u>10</u>	<u>20</u>	<u>30</u>	<u>40</u>	<u>50</u>
AA sequence	ALVEDPPLKVSEGGLIREGYDPDLRALRAAHREGVAYFLELEERERERTG				
Prediction	HH-----EEE-----HHHHHHHHHH-HHHHHHHHHHHHHHHHH				
Experiment	-E-----E-----HHHHHHHHHHHHHHHHHHHHHHHHHHHHHH-				
	<u>60</u>	<u>70</u>	<u>80</u>	<u>90</u>	<u>100</u>
AA sequence	IPTLVGYMAVFGYYLEVTRPYYERVPKEYRPVQTLKDRQRYTLPEMKEK				
Prediction	--EEEEEEEEEEEEEEEE-----EEEEEEEE--EEEE-HHHHHH				
Experiment	---EEEE--EEEEEEHHHHHH-----EEEE--EEEE-HHHHHH				
	<u>110</u>	<u>120</u>			
AA sequence	EREVYRLEALIRRREEEVFLEVRERAKRQ				
Prediction	HH				
Experiment	HH--				

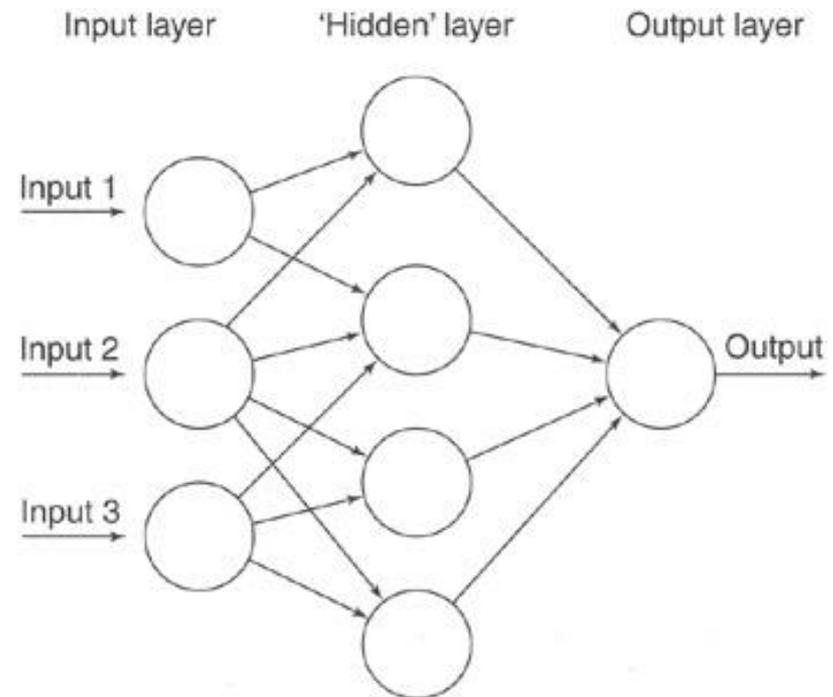
Наиболее мощные методы предсказания вторичной структуры основаны на нейронных сетях.

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.

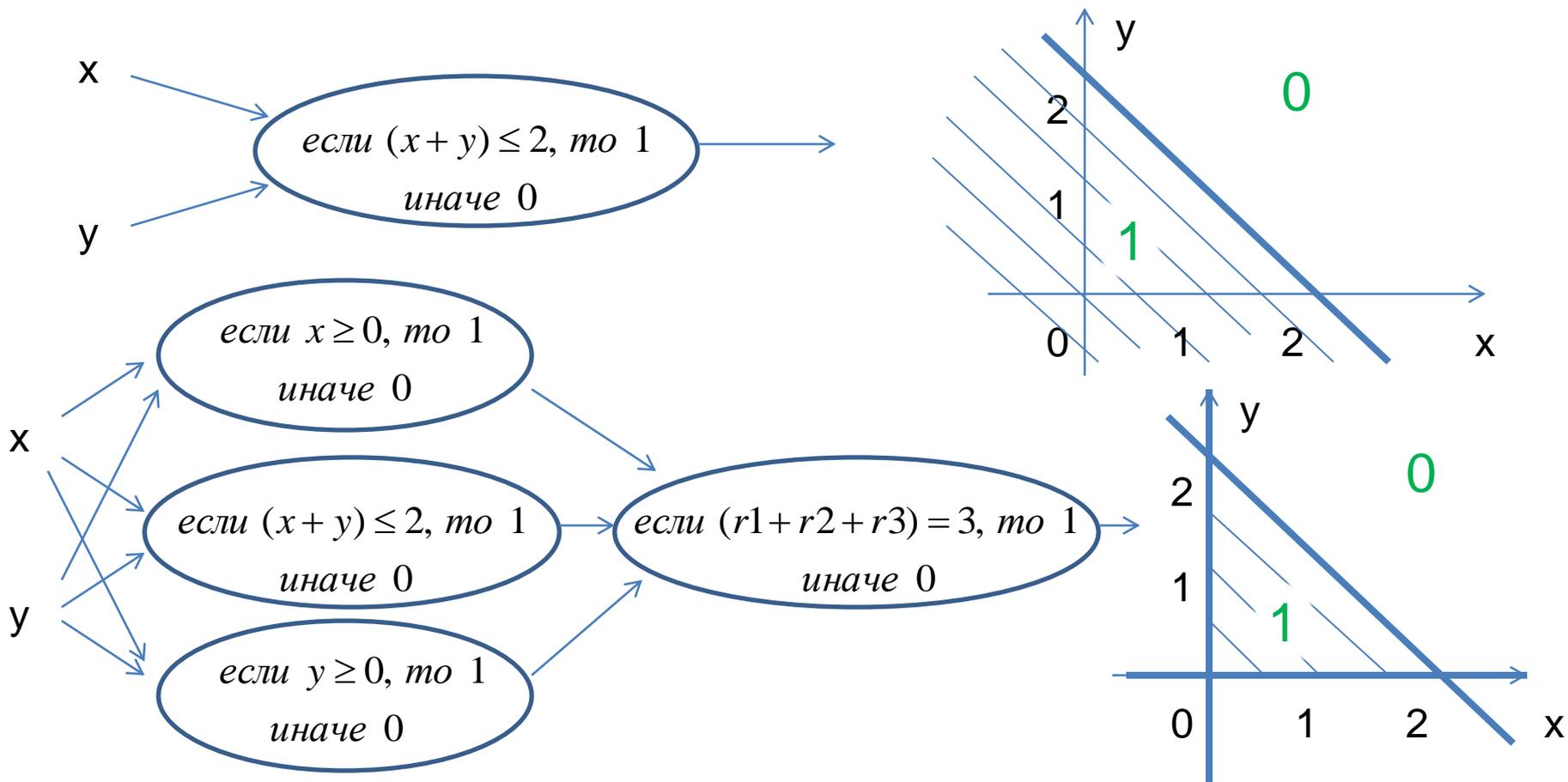
Для формирования сети необходимо соединить выходы одних нейронов со входами других.



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

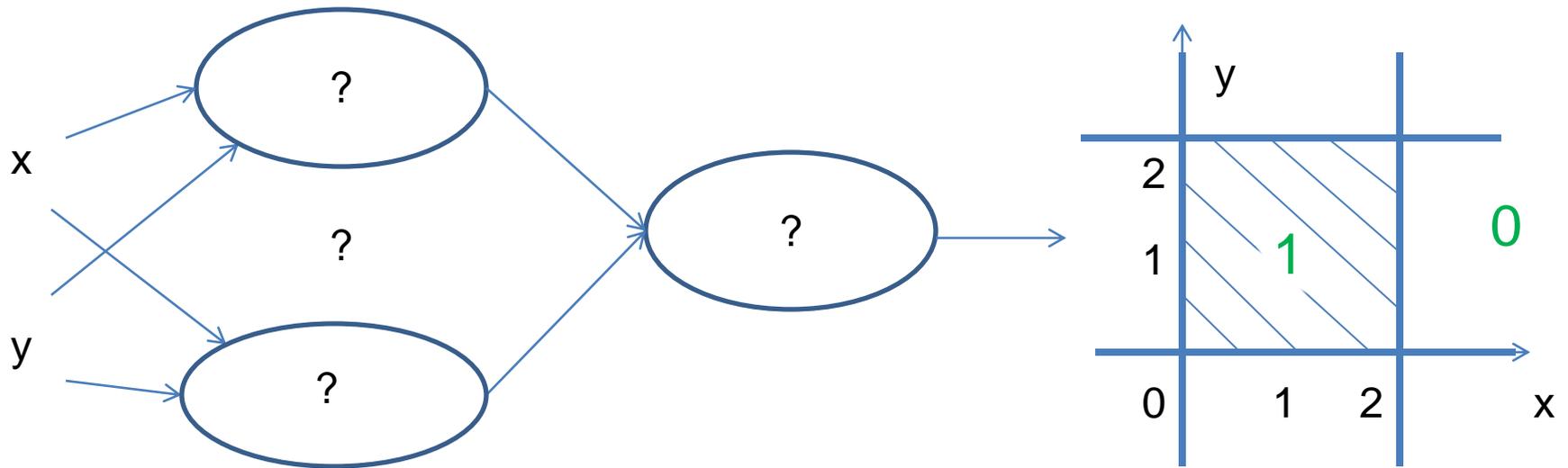
Нейронные сети. Геометрическая интерпретация

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

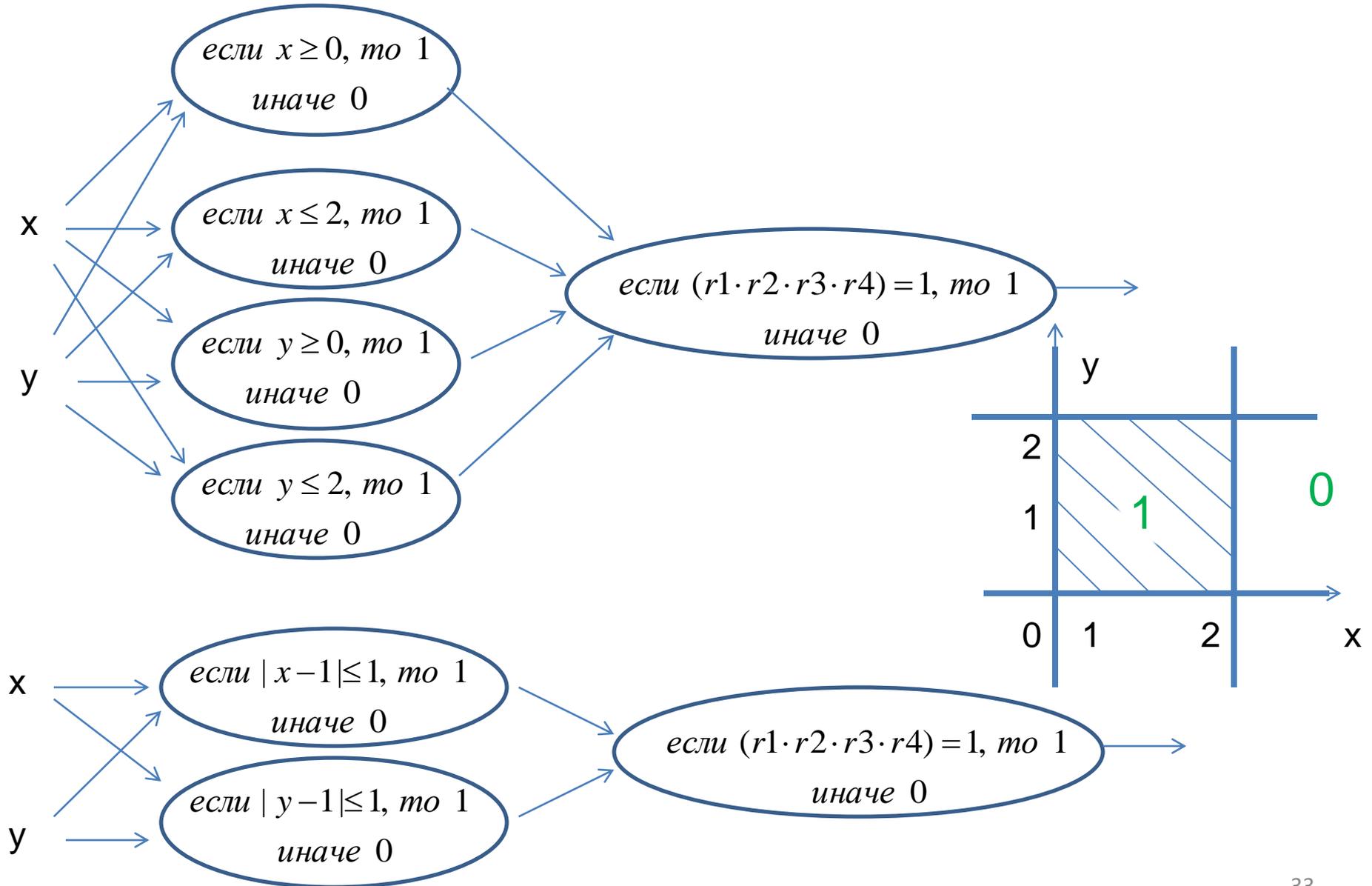


Нейронная сеть определяется топологией связей, весами и формулой принятия решения в узлах. Очевидно, сеть может принимать более сложные решения, чем один нейрон. 30

Нейронные сети. Геометрическая интерпретация

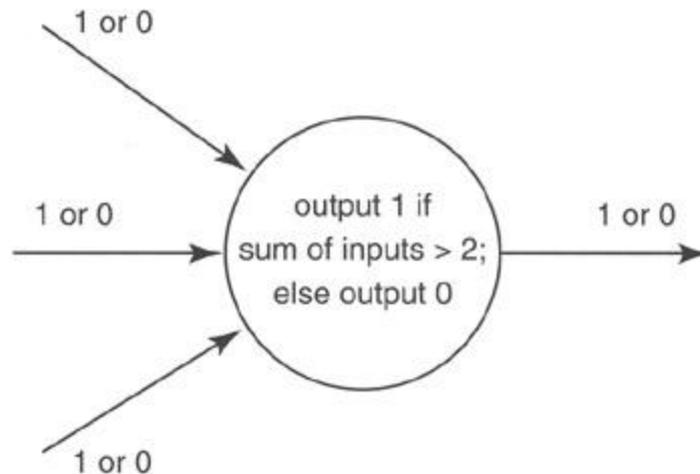


Нейронные сети. Геометрическая интерпретация



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 $i_1 + i_2 + i_3$, использовать их взвешенную сумму $10*i_1 + i_2 + 0,5*i_3$, 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.

A Neural Network Playground



Epoch
000,215

Learning rate
0.03

Activation
ReLU

Regularization
None

Regularization rate
0

Problem type
Classification

DATA

Which dataset do you want to use?



Ratio of training to test data: 50%

Noise: 0

Batch size: 9

REGENERATE

FEATURES

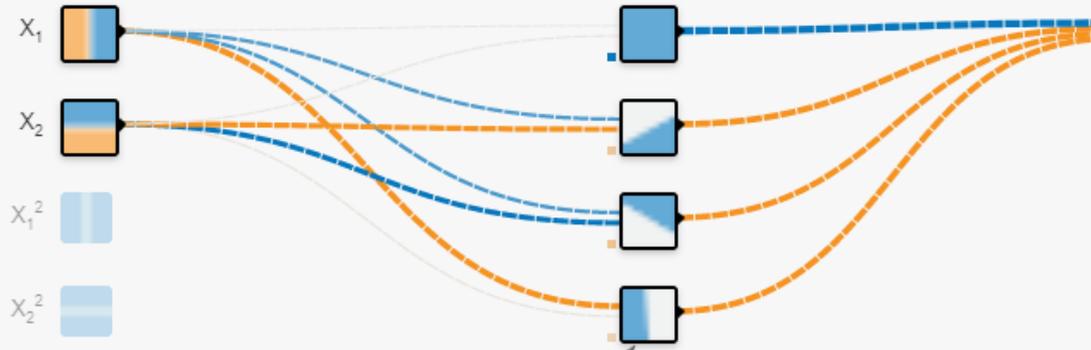
Which properties do you want to feed in?

- X_1
- X_2
- X_1^2
- X_2^2
- $X_1 X_2$
- $\sin(X_1)$
- $\sin(X_2)$

+ - 1 HIDDEN LAYER

+ -

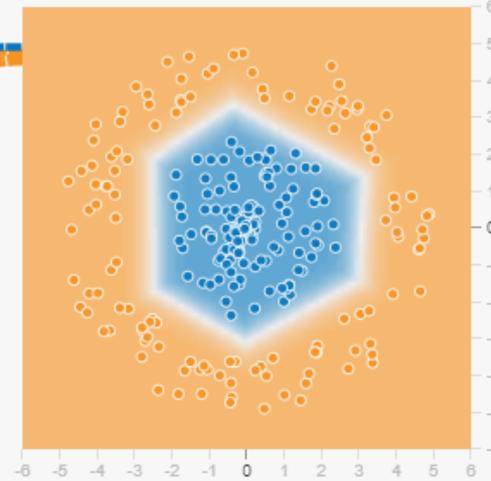
4 neurons



This is the output from one neuron. Hover to see it larger.

OUTPUT

Test loss 0.008
Training loss 0.002



Colors shows data, neuron and weight values.

A Neural Network Playground



Epoch
000,177

Learning rate
0.03

Activation
ReLU

Regularization
None

Regularization rate
0

Problem type
Classification

DATA

Which dataset do you want to use?



Ratio of training to test data: 50%



Noise: 0



Batch size: 9



REGENERATE

FEATURES

Which properties do you want to feed in?

- X_1
- X_2
- X_1^2
- X_2^2
- $X_1 X_2$
- $\sin(X_1)$
- $\sin(X_2)$

+ - 1 HIDDEN LAYER

+ -

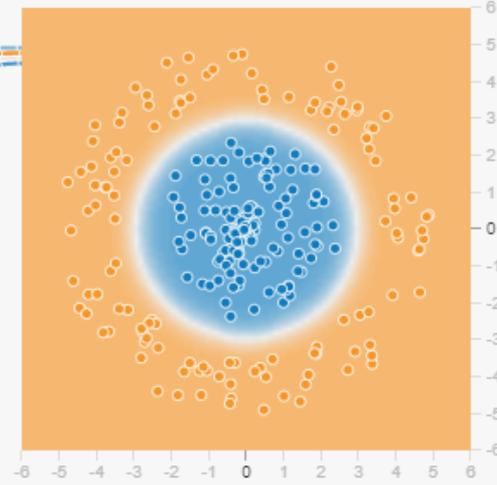
4 neurons



This is the output from one neuron. Hover to see it larger.

OUTPUT

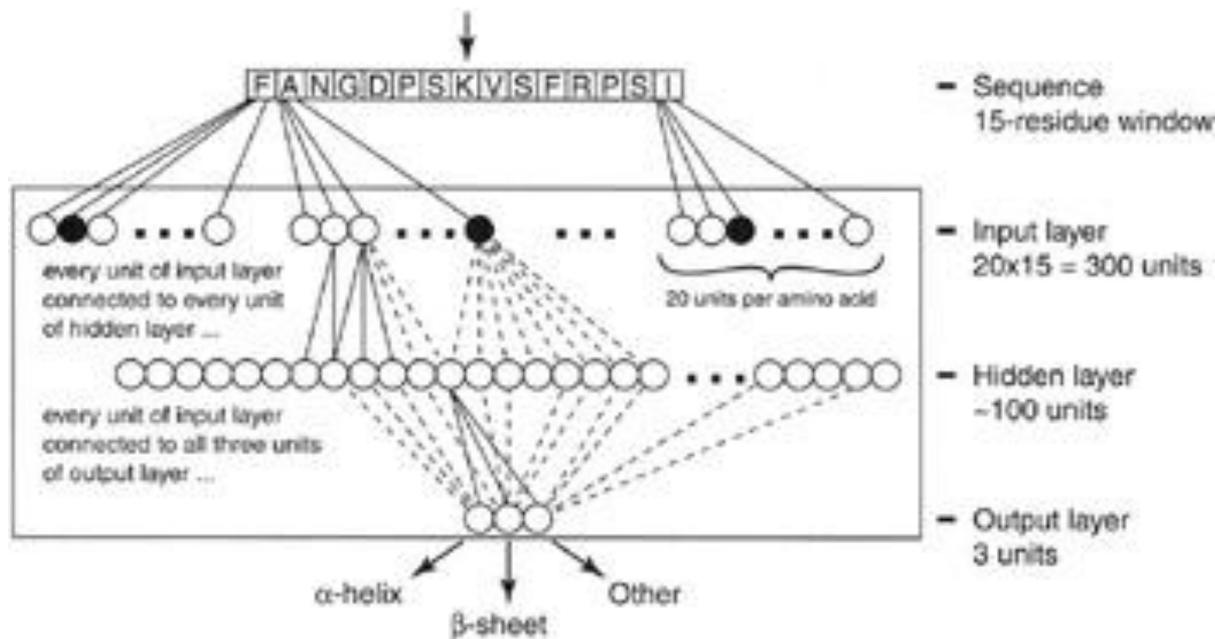
Test loss 0.001
Training loss 0.001



Colors shows data, neuron and weight values. -1 0 1

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».

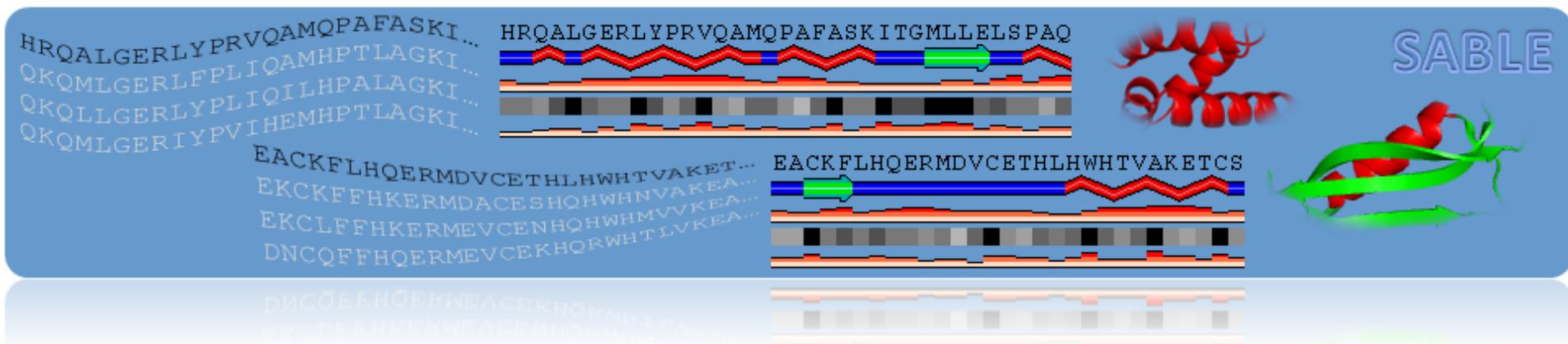
Предсказание вторичной структуры. PSIPRED и SABLE

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

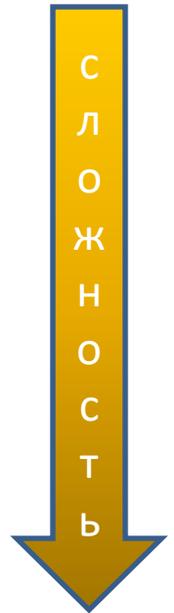


Предсказание структуры белков

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;**
- 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

Построение модели

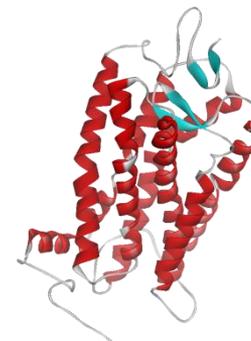
Оценка модели

Нет

Модель подходит?

Да

Ура!



Model:	FVVFVL.FAIC
	:: :
Template:	VIIMVIAFLIC

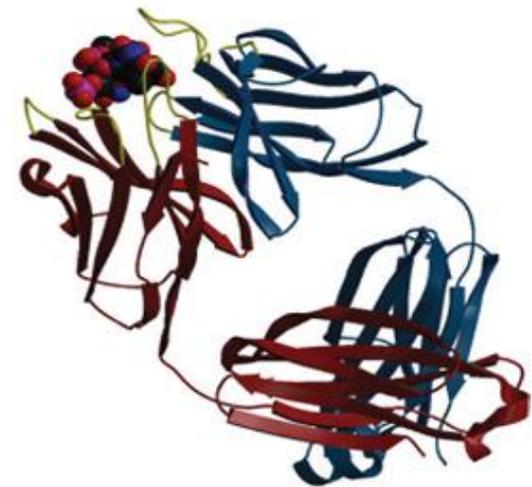
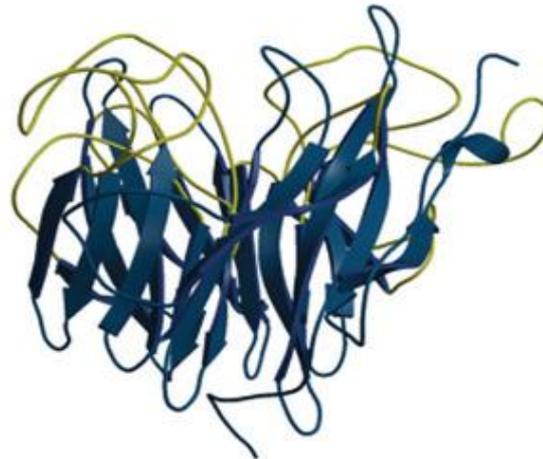
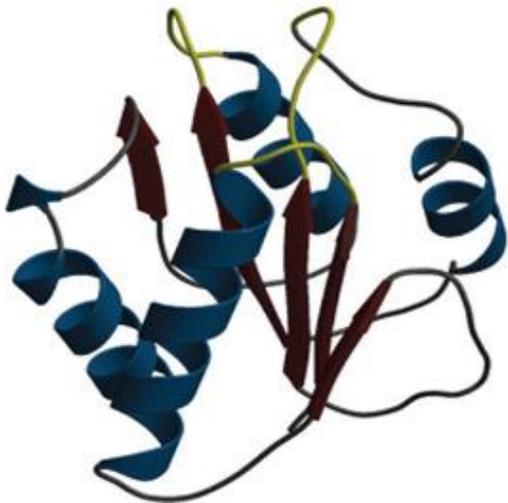


Моделирование на основании гомологии. Методы

- 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 C α -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.

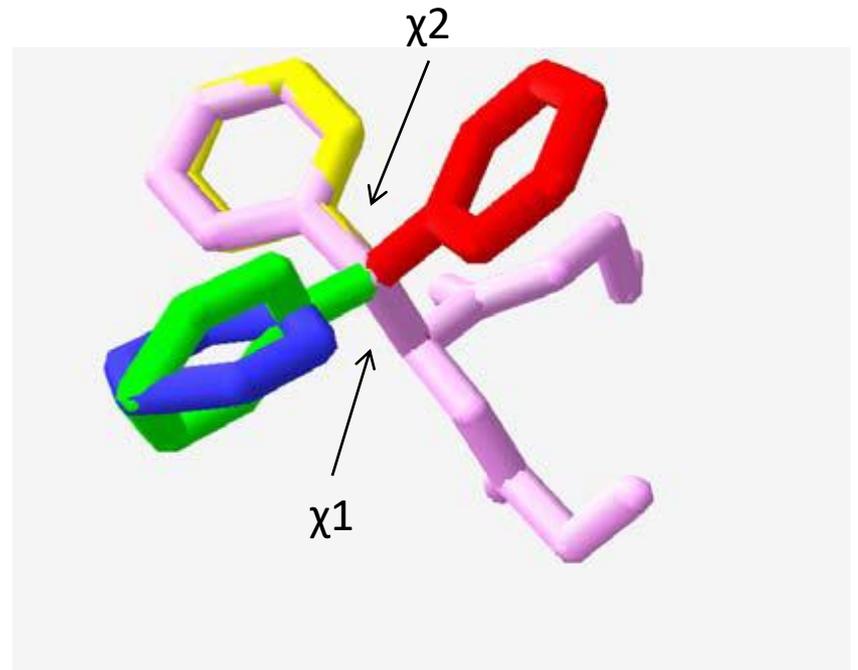
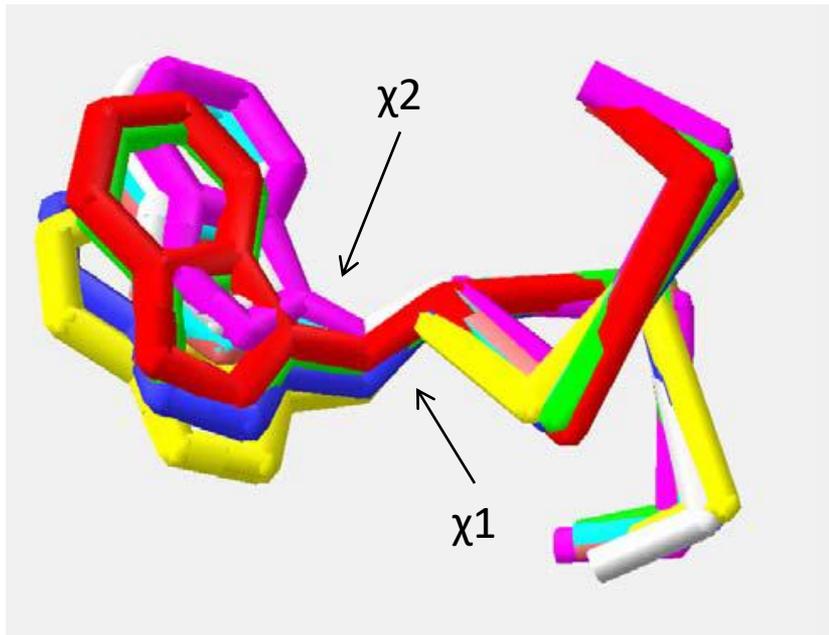


BIOZENTRUM
University of Basel
The Center for Molecular Life Sciences

SWISS-MODEL

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



Моделирование на основе пространственных ограничений. MODELLER

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

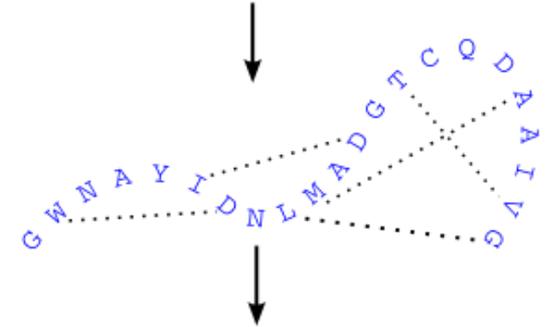
- Поиск белковых структур с последовательностями, гомологичными моделируемой.
- Выполнение выравнивания последовательностей.
- Извлечение пространственных ограничений из шаблонов.
- Построение модели с учетом этих ограничений
- Общая оптимизация геометрии.

1. Align sequence with structures

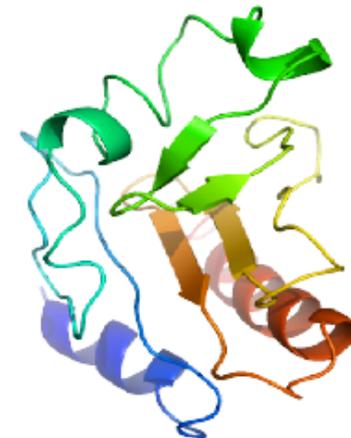
Template structure(s)
Target sequence

SWQTYVDTNLVGTGAVTQA--AI
-GWNAYIDNLMADGTCQDAAIVG

2. Extract spatial restraints

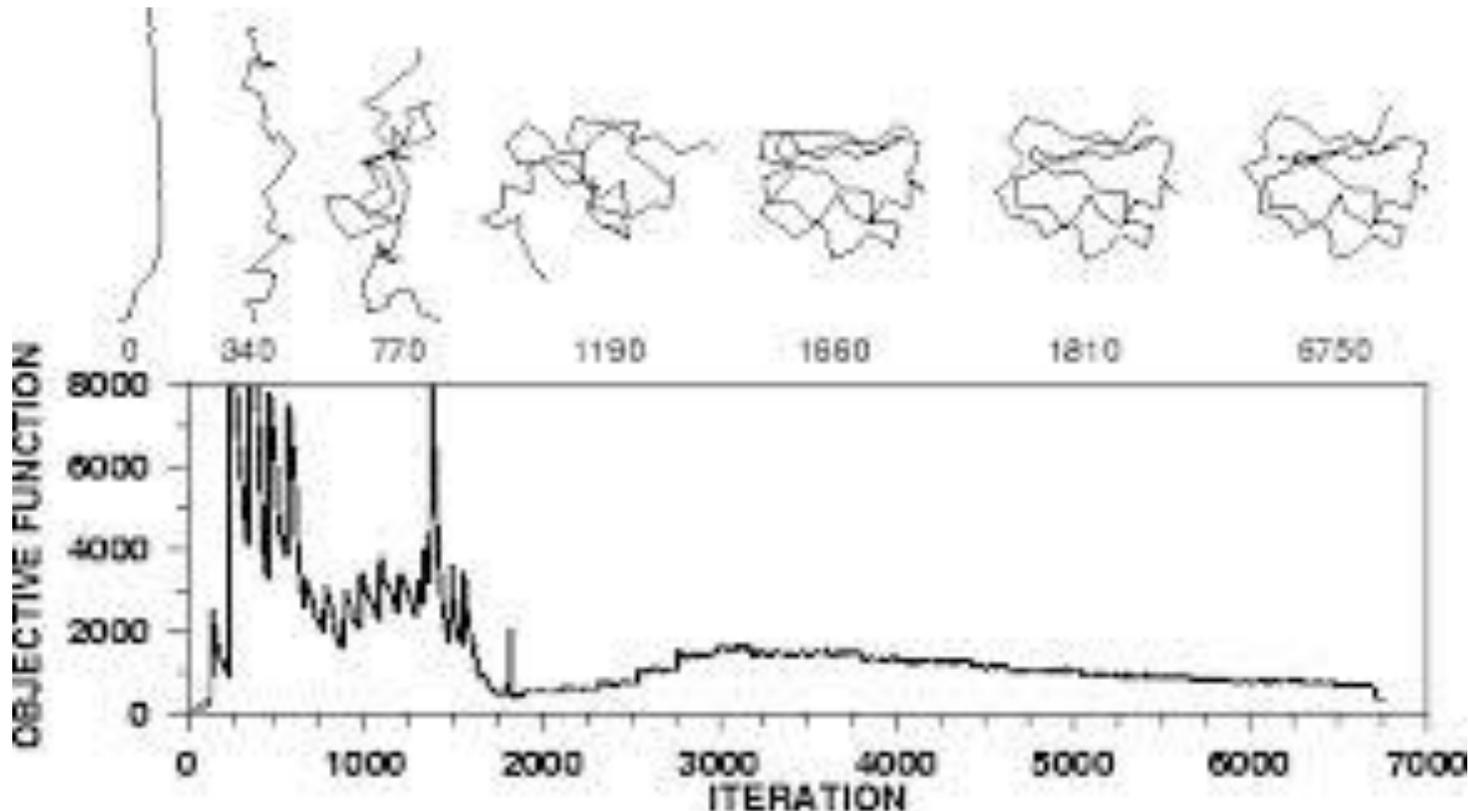


3. Satisfy spatial restraints



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:

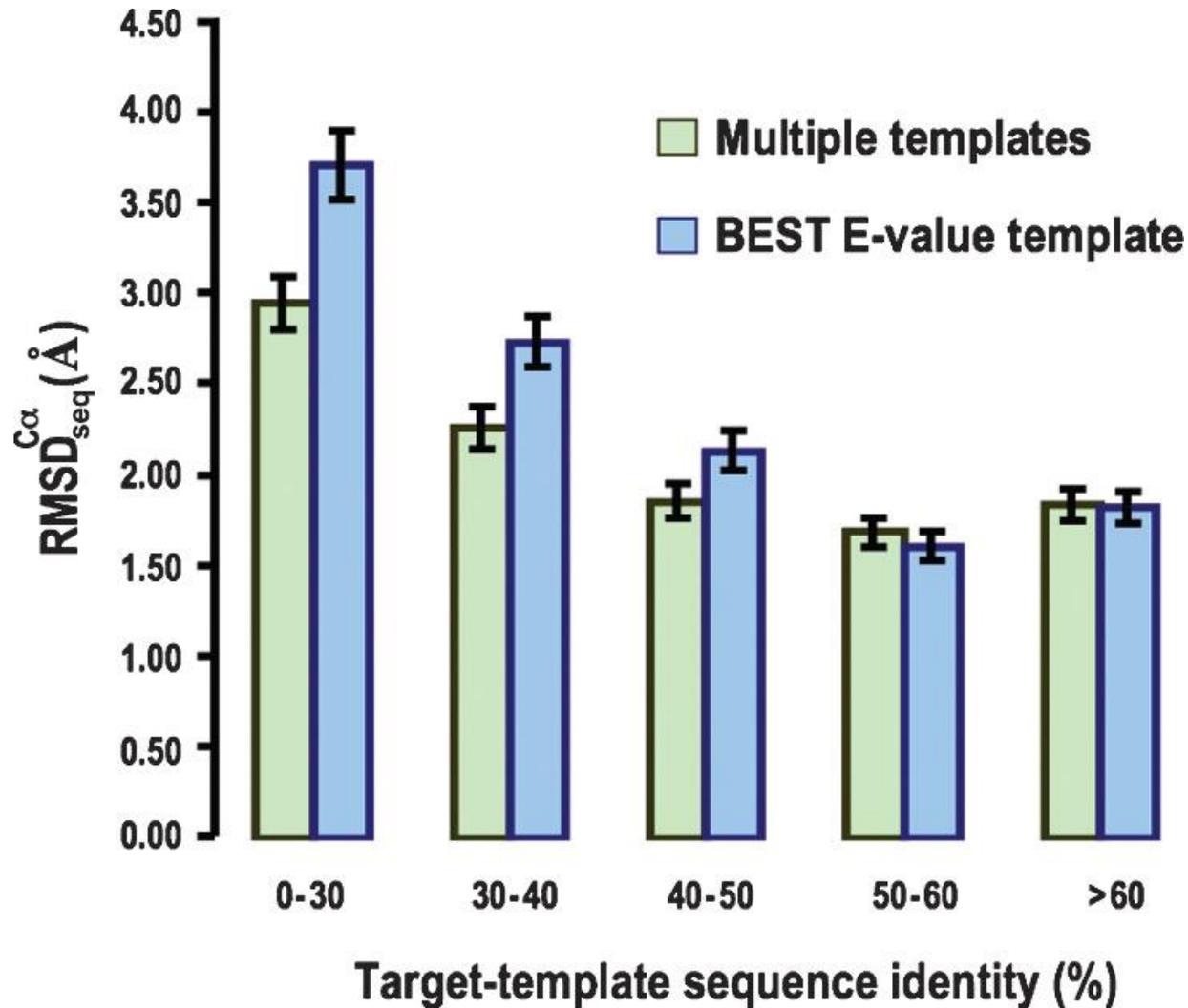
1. Парное сравнение моделируемой последовательности с каждой последовательностью из базы данных (**FASTA, BLAST**).
2. Сравнение сразу нескольких последовательностей (**Clustal, Muscle**).
3. Протягивание последовательности через библиотеку пространственных структур.

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

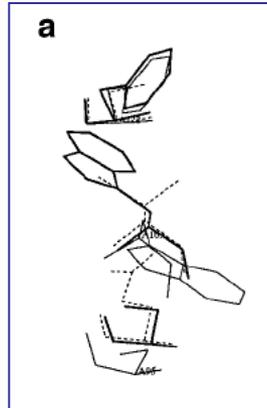
Качество модели: несколько шаблонов



Ошибки построения модели

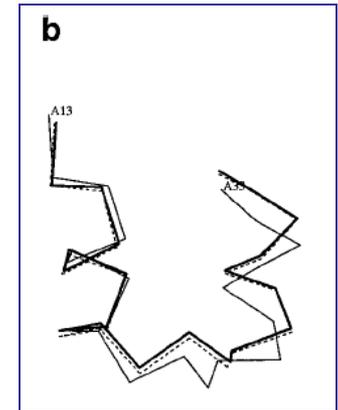
1. Ошибки в ориентации боковых цепей.

Мышиный белок, связывающий ретиноевую кислоту. Тонкая линия – кристалл, толстая линия – модель, пунктир – шаблон (мышинный липид-связывающий белок).



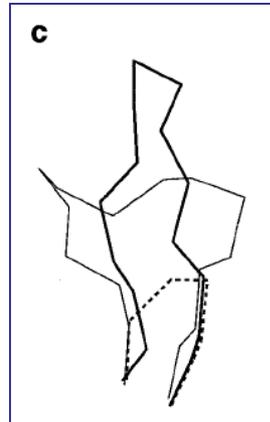
2. Сдвиги в корректно выровненных участках.

Сравнение участка кристаллической структуры мышиноного белка, связывающего ретиноевую кислоту, с его моделью и с шаблоном.



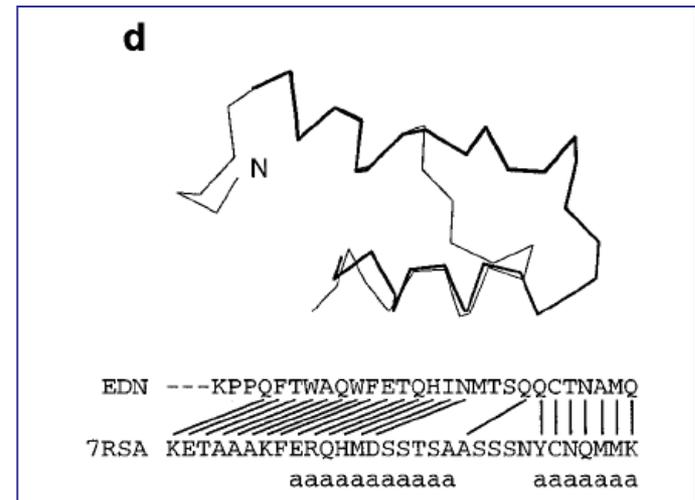
3. Ошибки в участках, для которых отсутствует шаблон.

Показан контур C α 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.

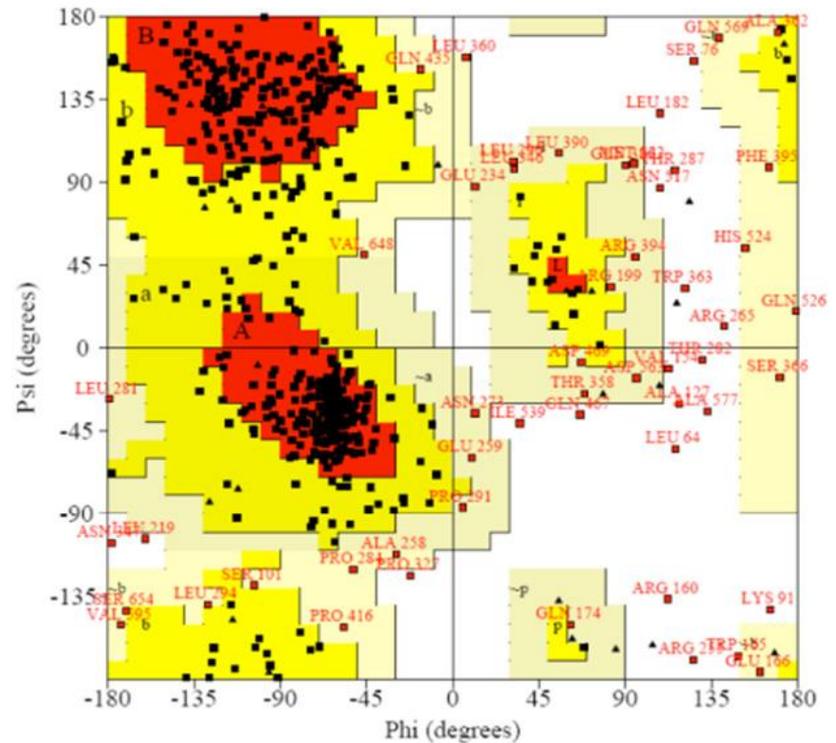
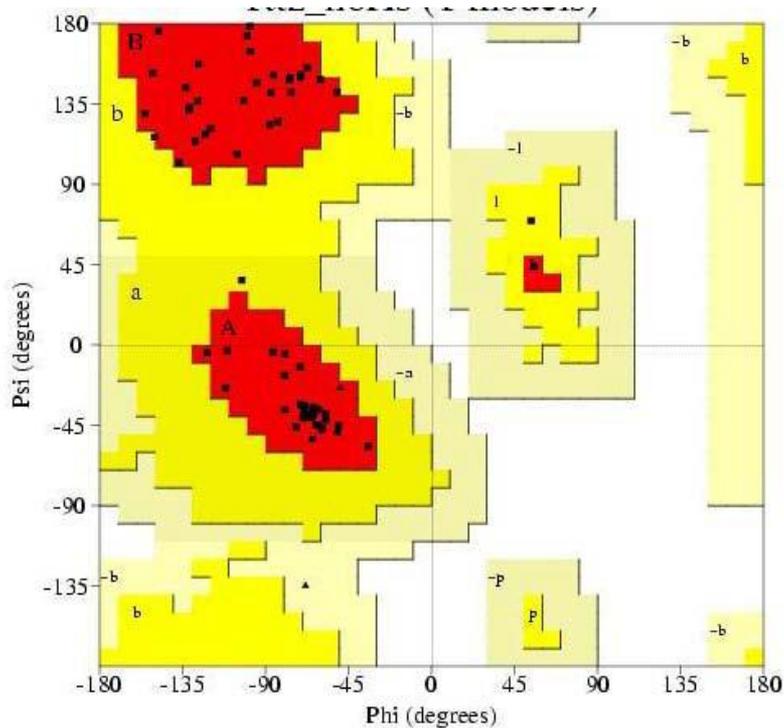
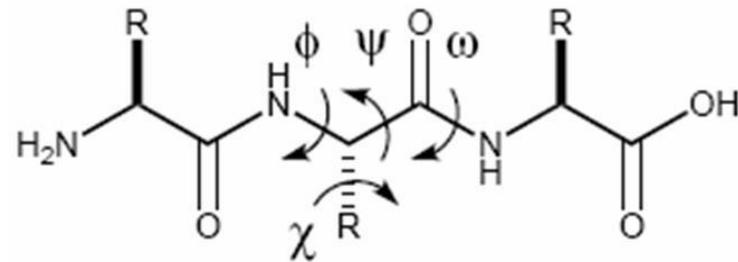
N-концевой участок токсина сравнивается с его моделью. Показан соответствующий участок выравнивания, линии показывают эквивалентные остатки.



5. Неправильно выбранный шаблон.

Оценка модели. Проверка стереохимии

Карты Рамачандрана



Пример моделирования

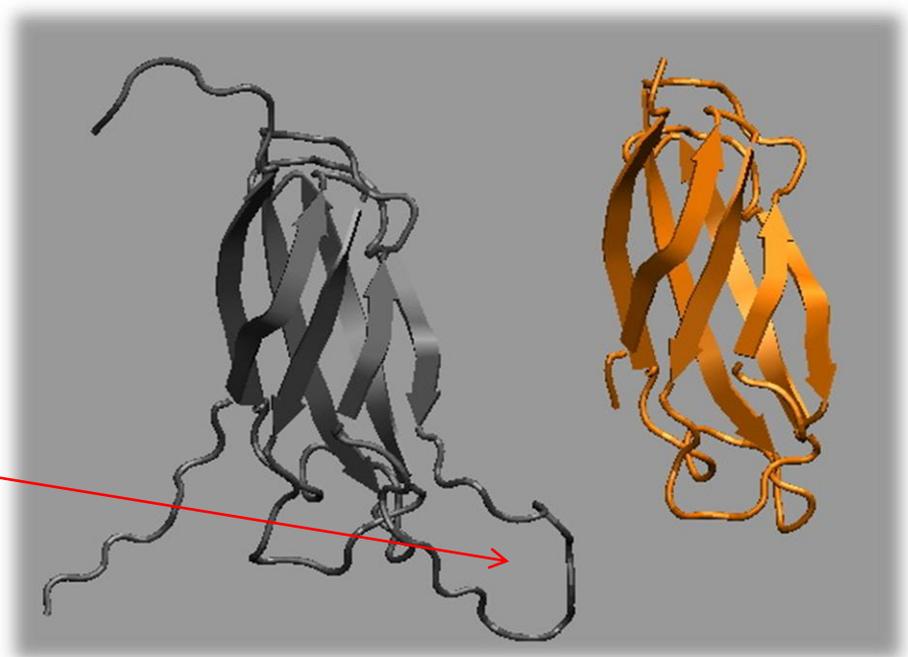
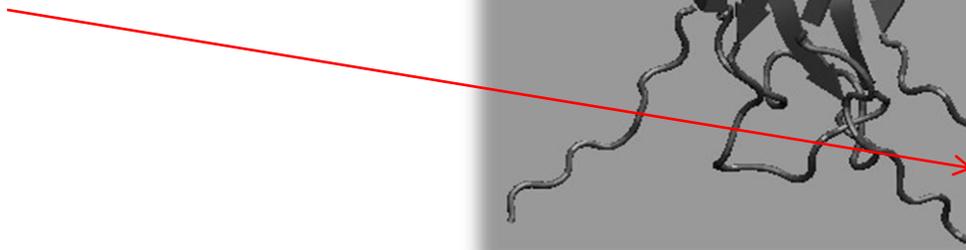
```
FN3H   -MQVSDVPTNLEVVAATPTSLLISWYTFTHYG--MNRYRITYGETGGNS 47
1FNA_A -----RDLEVVAATPTSLLISWDAP-AVT---VRYYRITYGETGGNS 38
          :***** .                               *****
```

```
FN3H   PVQEFVTPWINTYTGEPYADDFKGRFTATISGLKPGVDYTITVYAVTEF 97
1FNA_A PVQEFVTP-----GSKSTATISGLKPGVDYTITVYAVTGR 73
          *****                               *****
```

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



Вставка

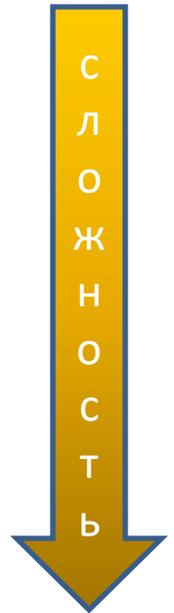


Предсказание структуры белков

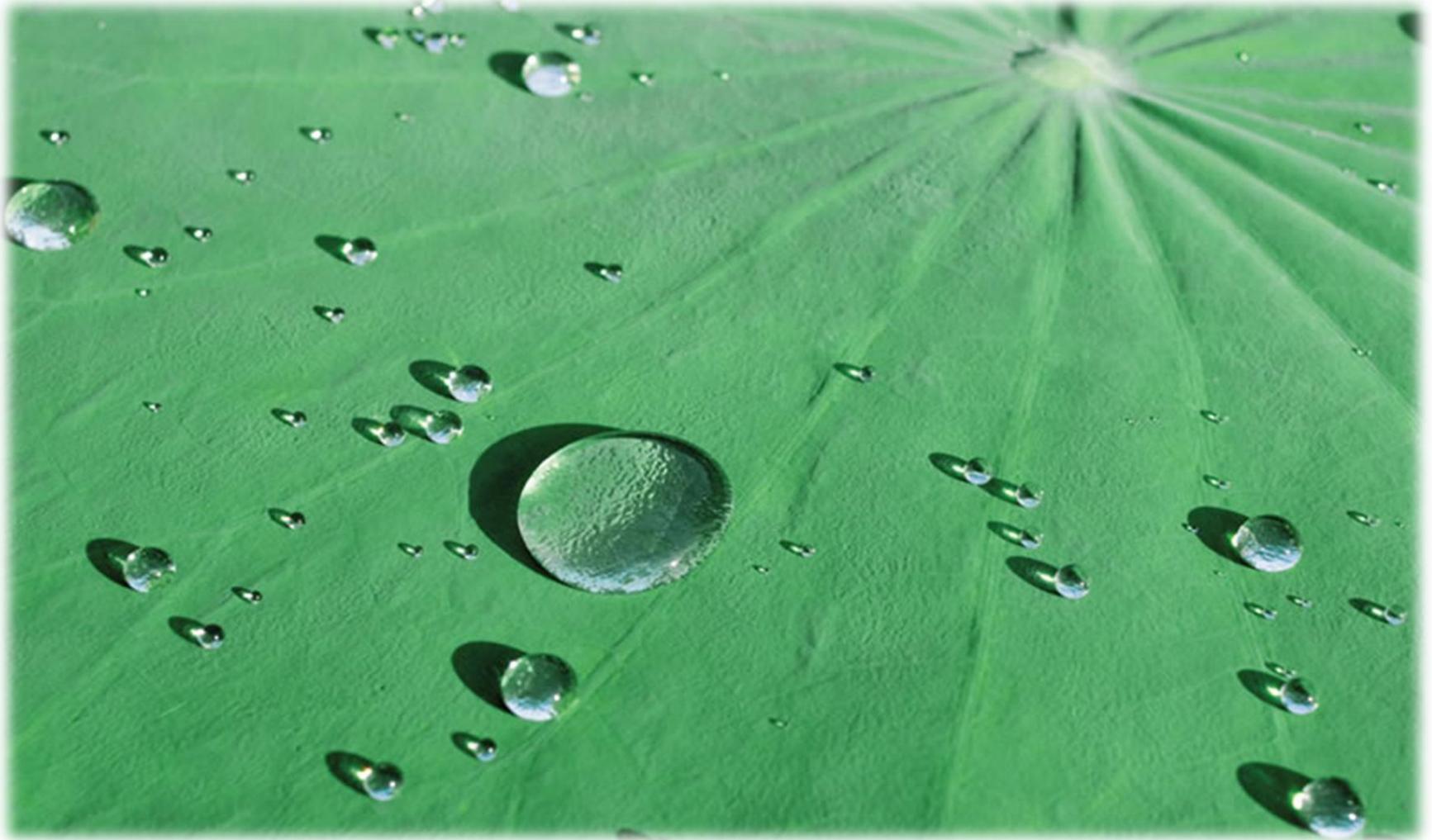
Сворачивание белка в уникальную конформацию наводит на мысль об алгоритме формирования структуры белка по его последовательности, но доказательством полноты и правильности нашего понимания могла бы стать его реализация в виде компьютерной программы...

Методы предсказания структуры по последовательности:

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

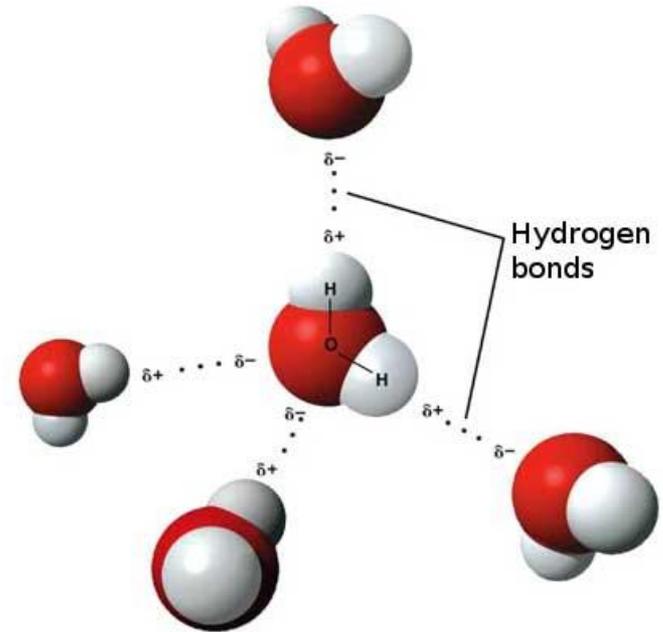


Гидрофобность



Гидрофобность

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

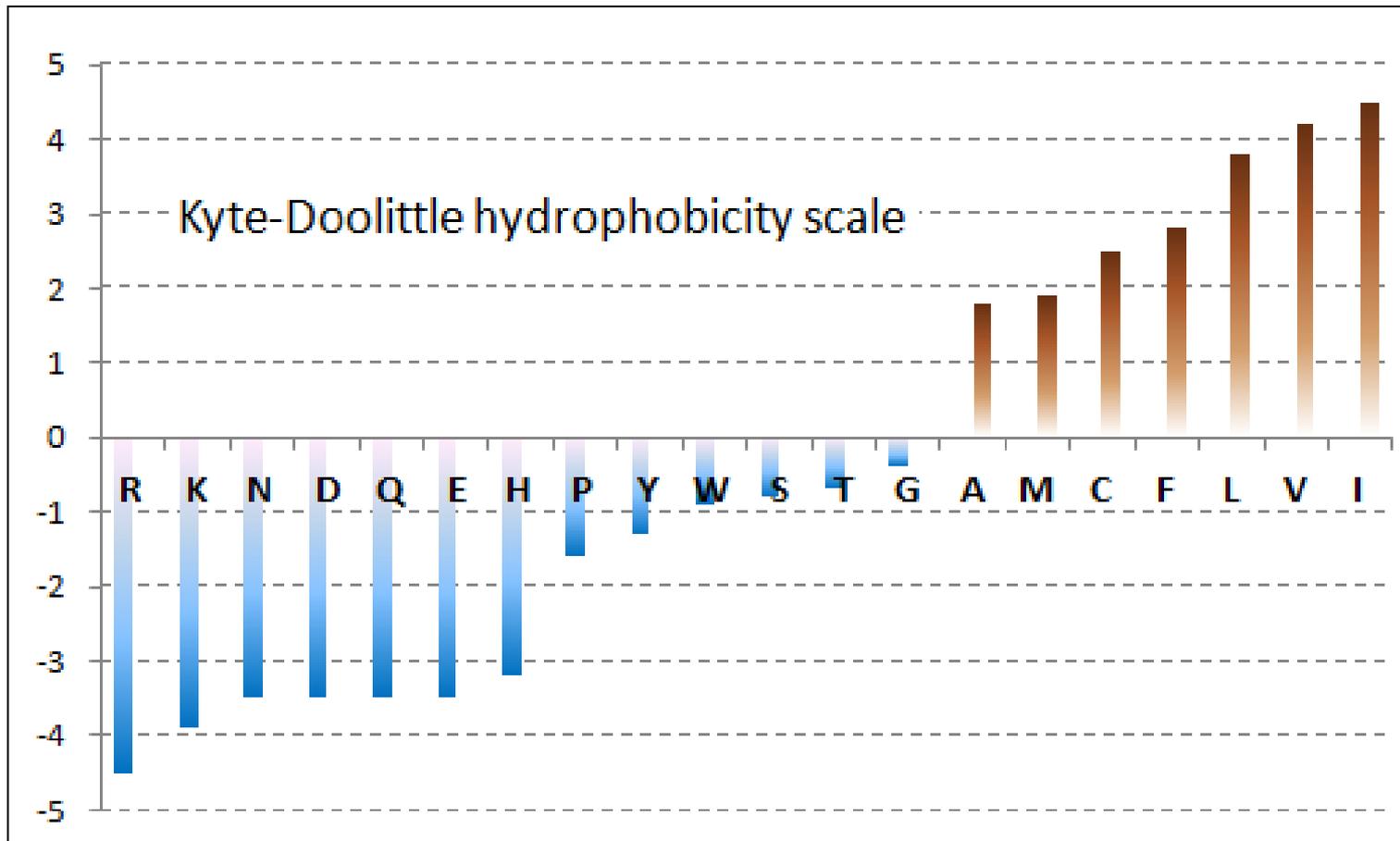


Мерой гидрофобности молекул может служить коэффициент разделения – равновесное отношение концентраций вещества в двух фазах в случае несмешивающихся растворителей:

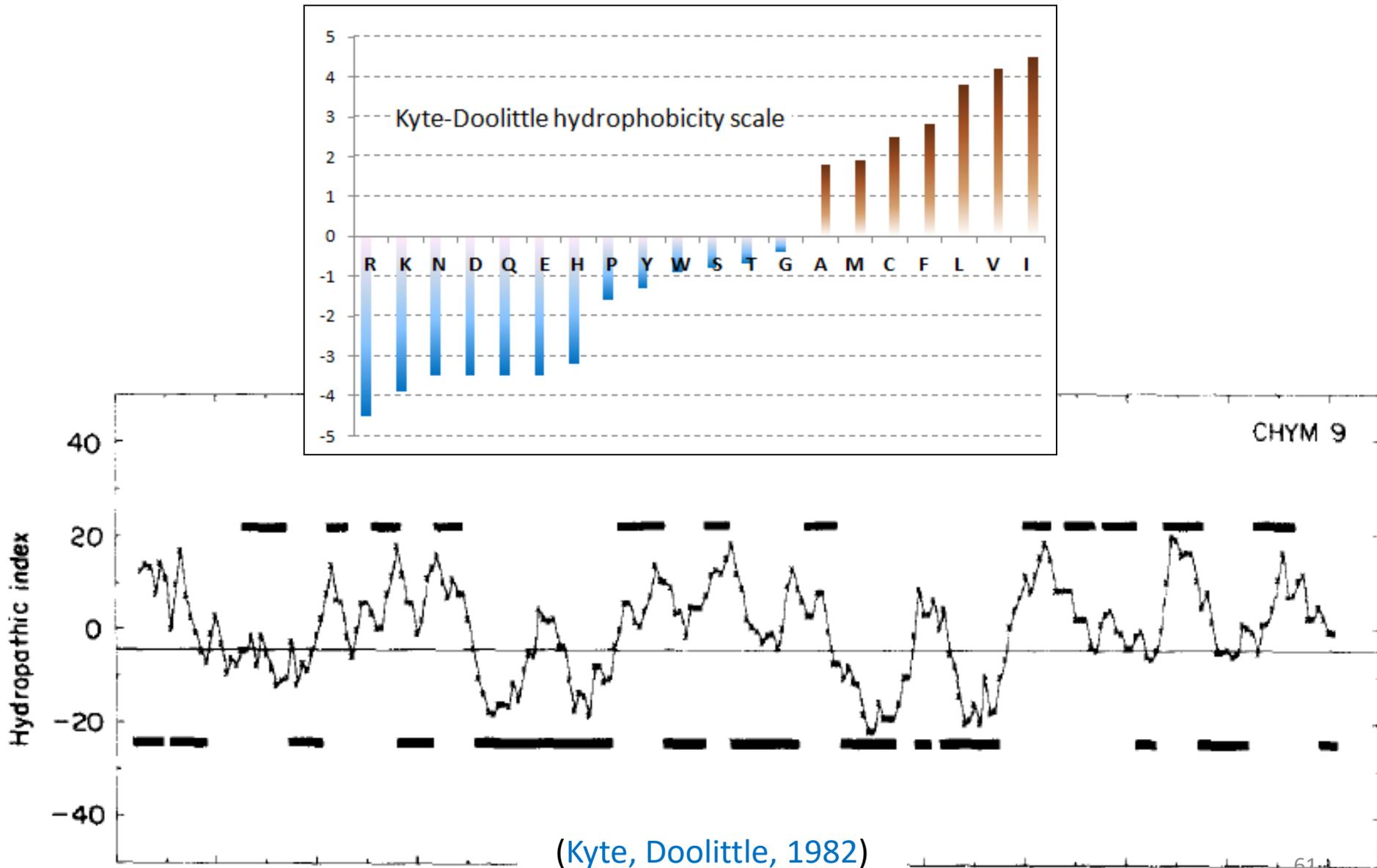
$$\log P_{oct/wat} = \log \left(\frac{[solute]_{octanol}}{[solute]_{water}^{un-ionized}} \right)$$

Гидрофобность

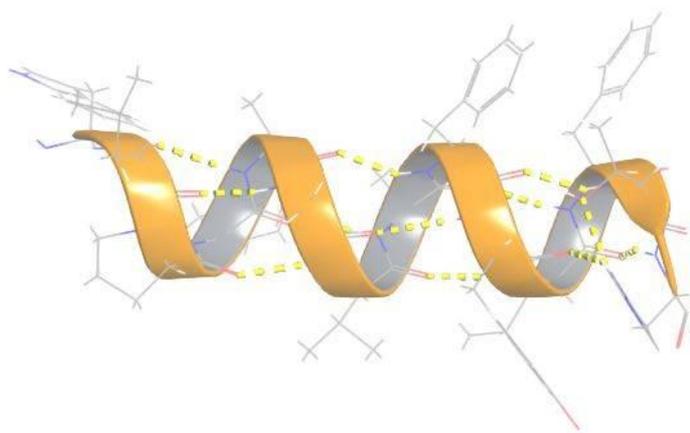
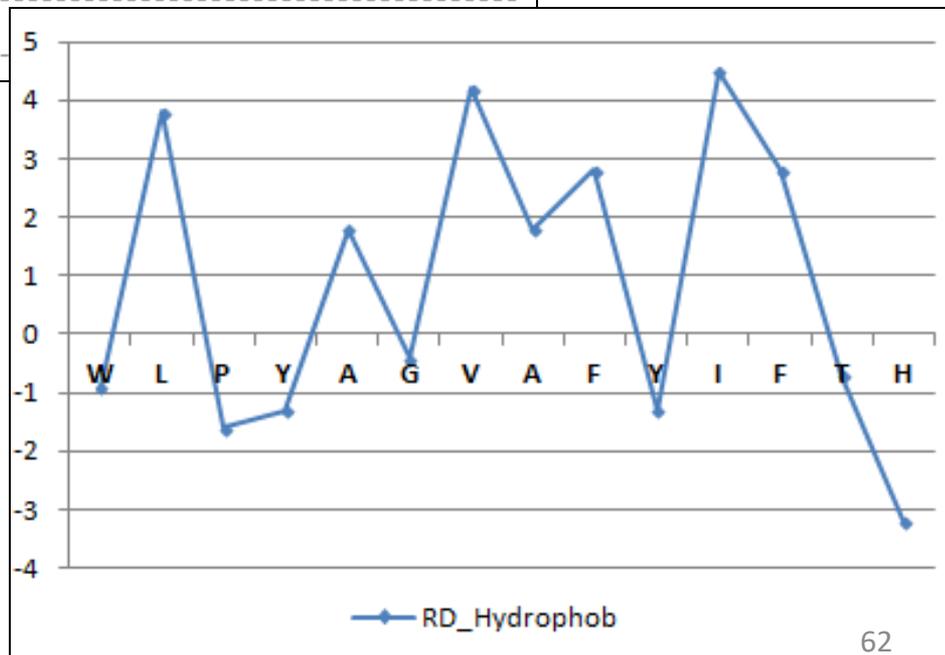
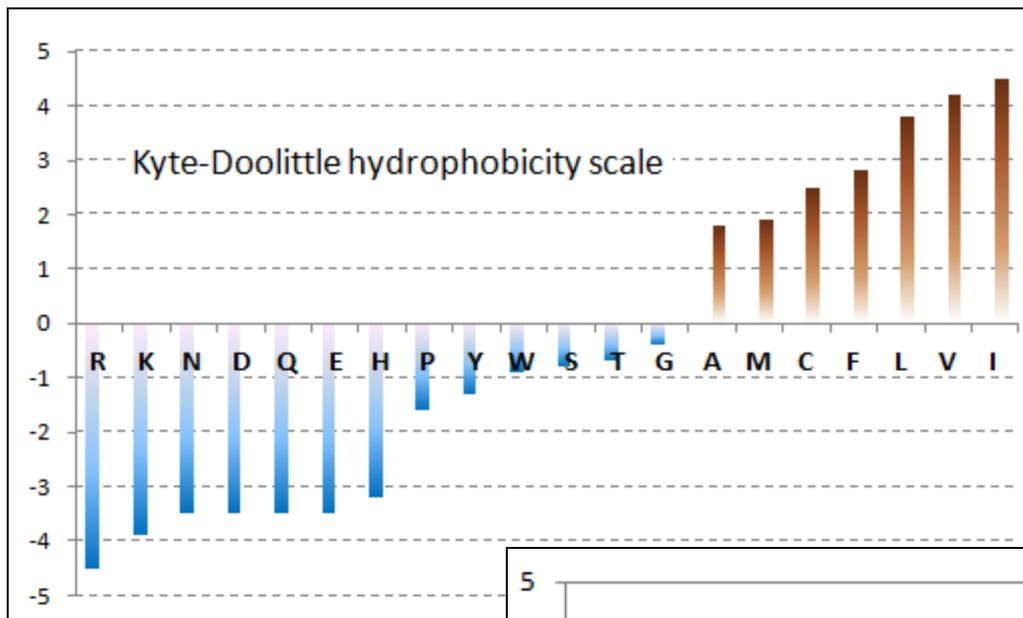
Распространенная шкала туманного происхождения ([Kyte, Doolittle, 1982](#))



Предсказание топологии. Профили гидрофобности

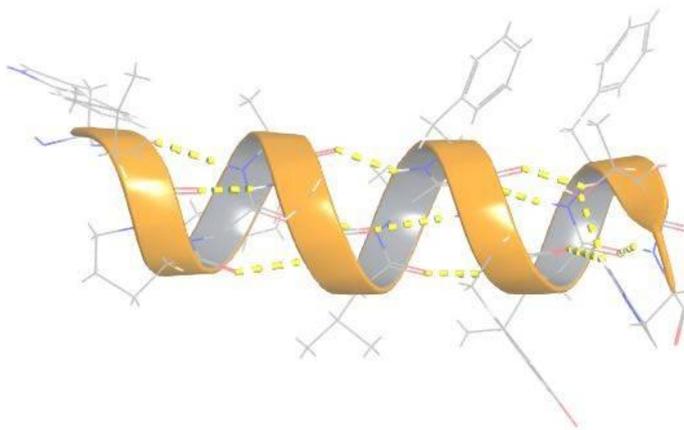
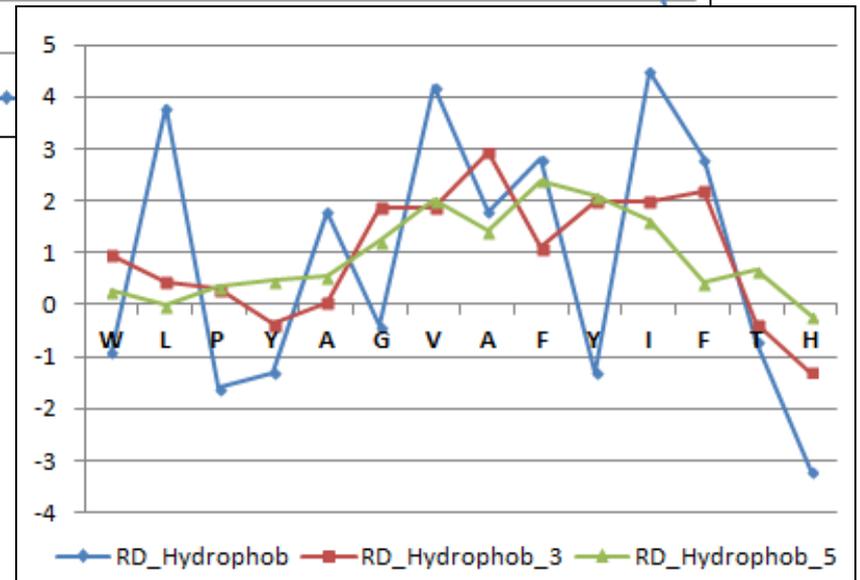
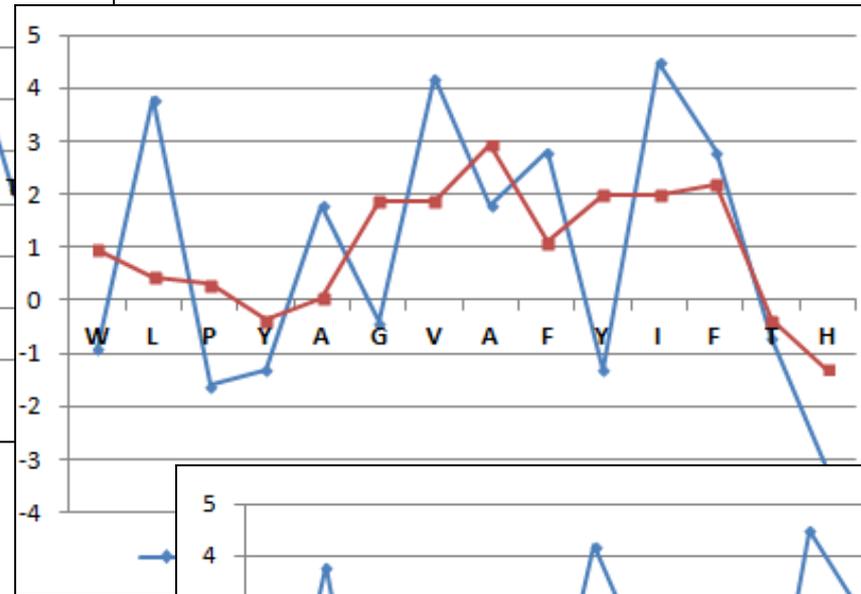
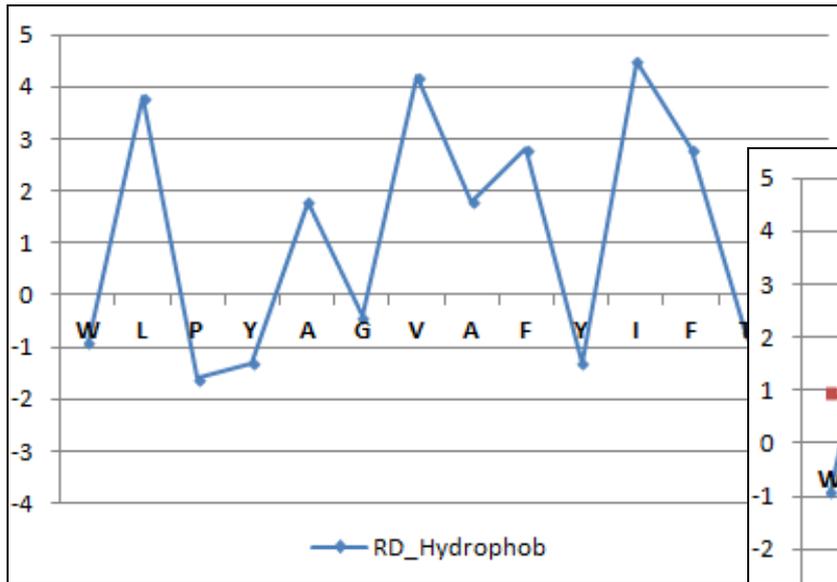


Предсказание топологии. Профили гидрофобности



...WLPY**A**GV**A**FY**I**F**T**H...

Предсказание топологии. Профили гидрофобности

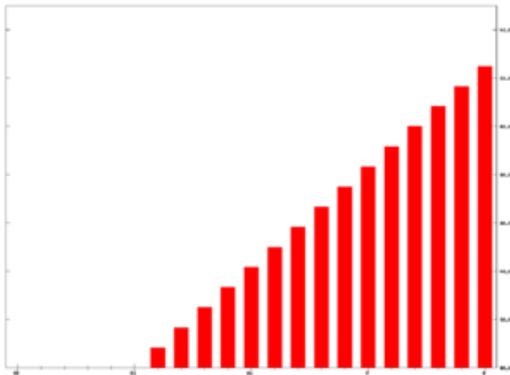
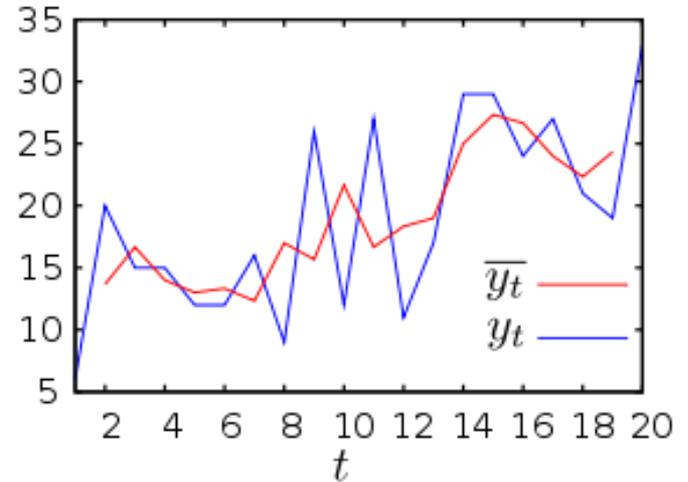


...WLPY**A**GV**A**F**Y**I**F**TH...

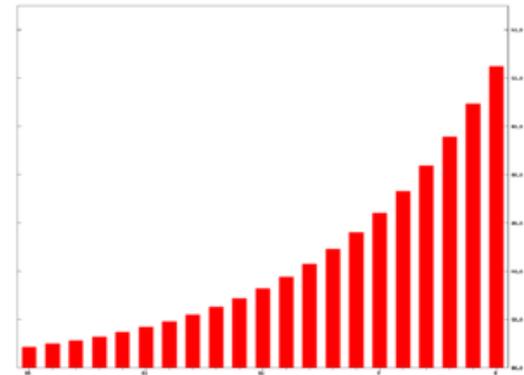
Скользящее среднее

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

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

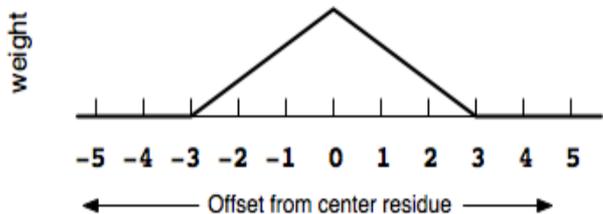
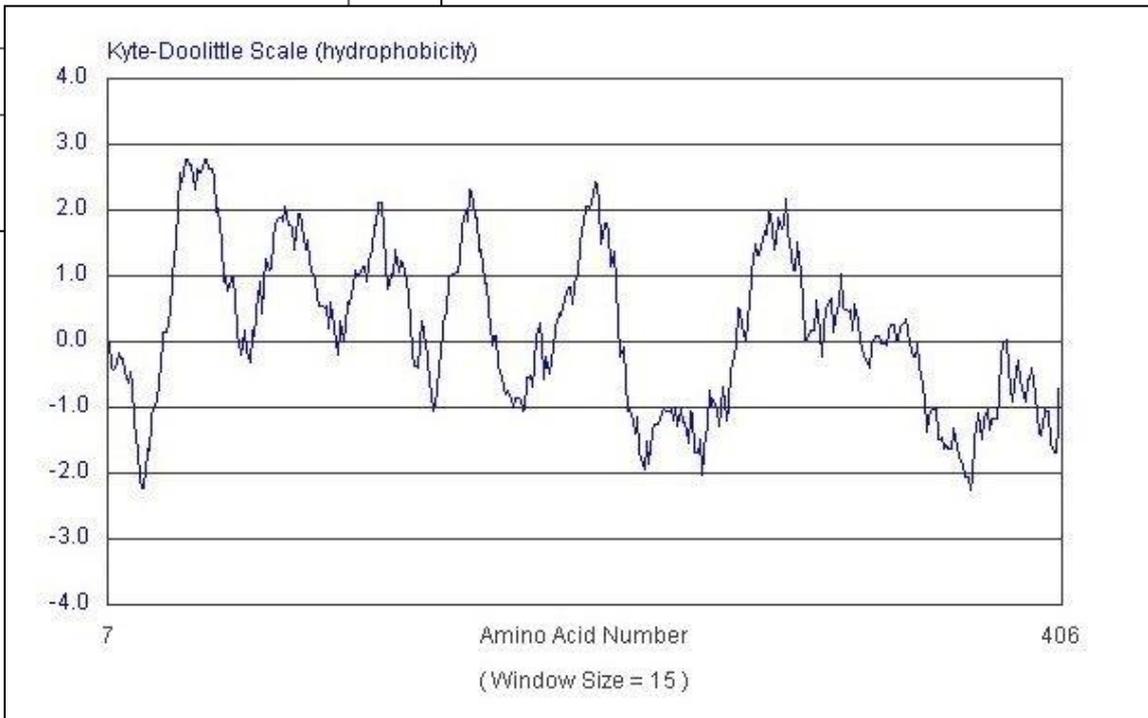
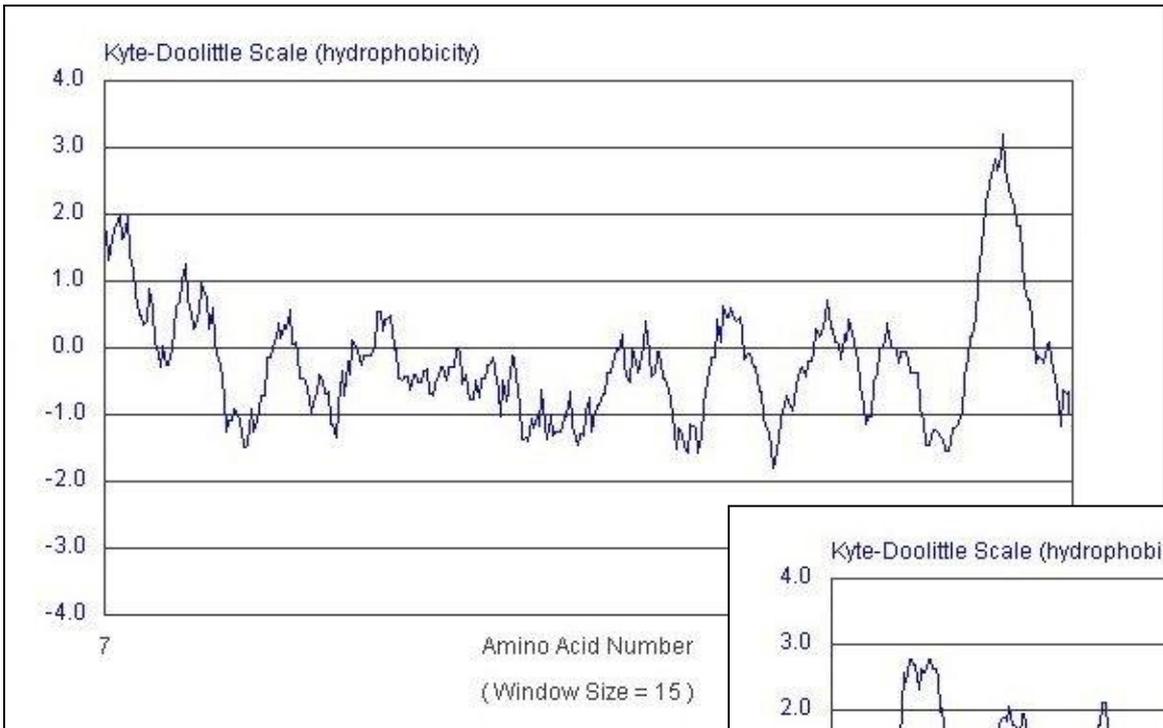


Линейное взвешивание



Экспоненциальное взвешивание

Предсказание топологии. Профили гидрофобности



Треугольное взвешивание

Профили гидрофобности

<http://www.vivo.colostate.edu/molkit/hydropathy/>



«Apologies, but this site was taken down. The programs were written in an older version of Java which is not compatible with most modern browsers and I do not have time to re-code.»

<http://gcat.davidson.edu/DGPB/kd/kyte-doolittle.htm>



Last Modified: Wednesday, 27 February 2002

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

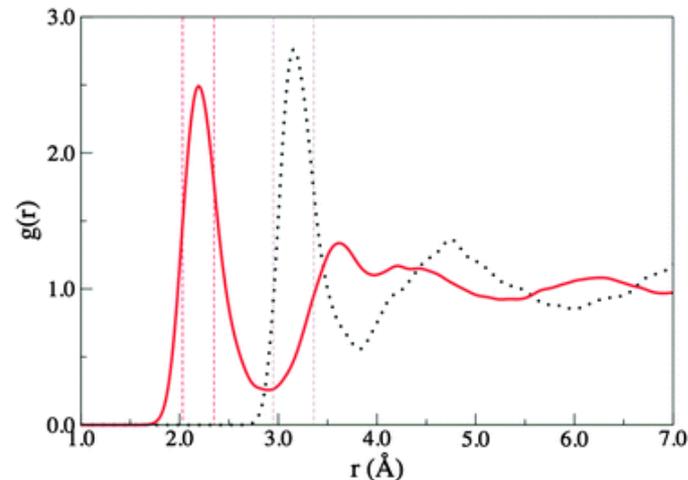
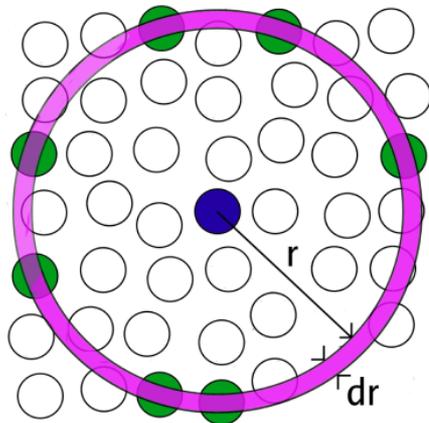
Распознавание фолда. «Протягивание»

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.**

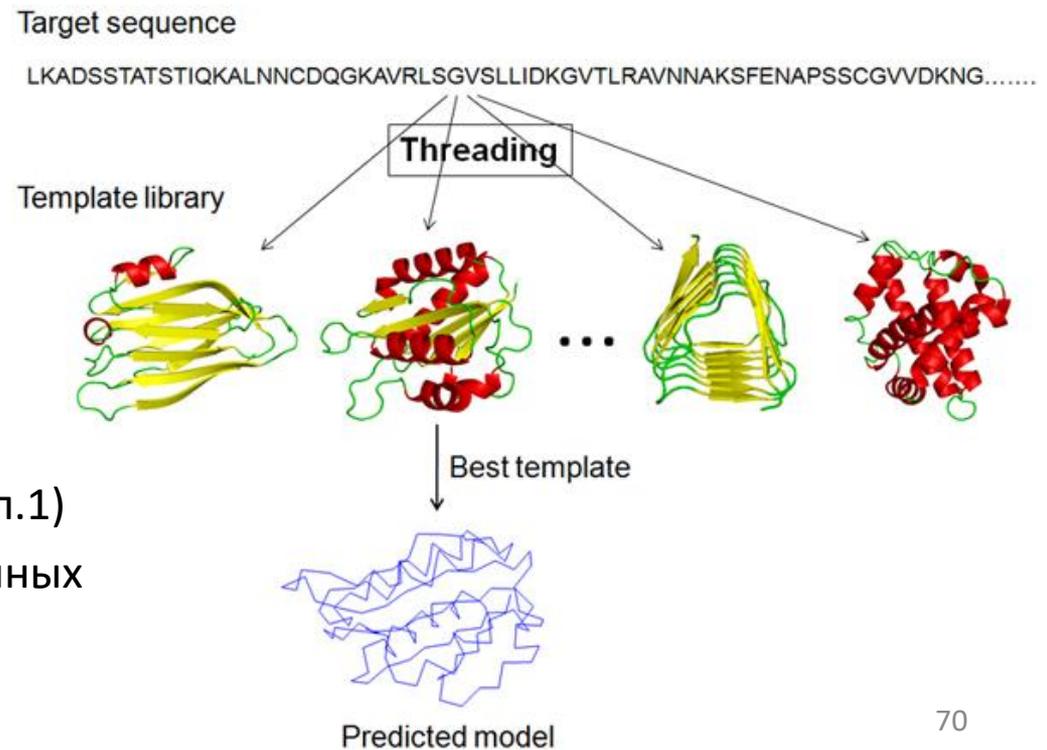
Например, функции распределения вероятности парных расстояний между остатками (например, по C β -атомам) (20x20 штук). Соотнося расстояния в моделях с этими функциями, можно оценить насколько вероятны как эти расстояния, так и модели в целом.



Распознавание фолда. «Протягивание»

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

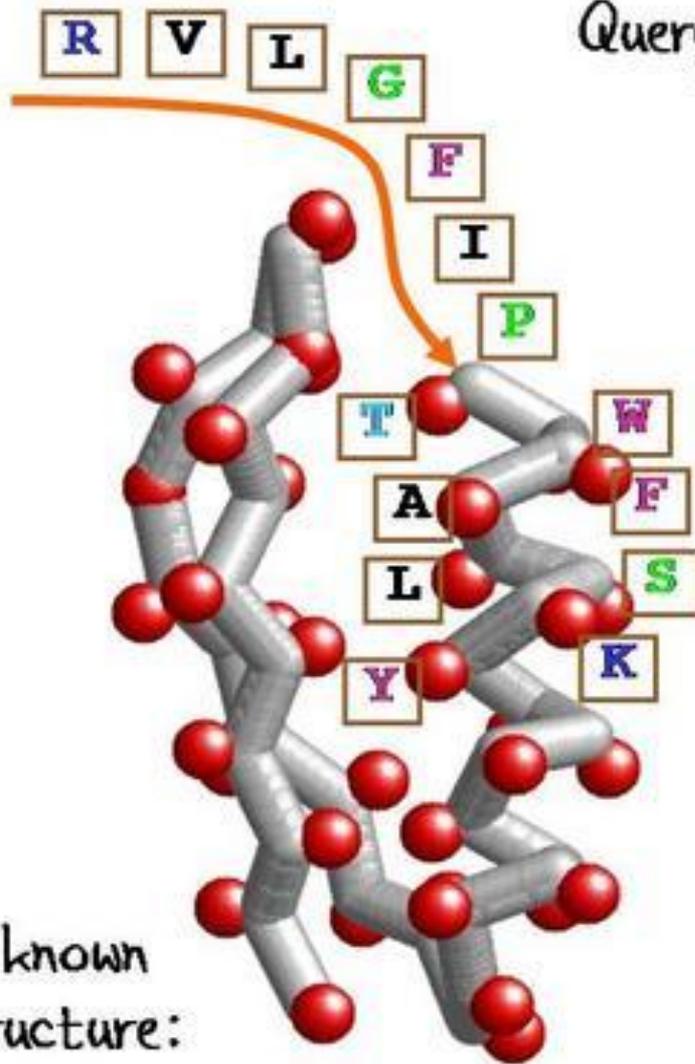
...IIAWLVKEKKVDVIV... ...IIAWLVK-EKKVDVIV... ...IIAWLVKEKKVDVIV...
...NGLELVLDSVLDATF... ...NGLELVLDSVLDATF... ...NGLELVLD-SVLDATF...
 ** * ** ** *



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

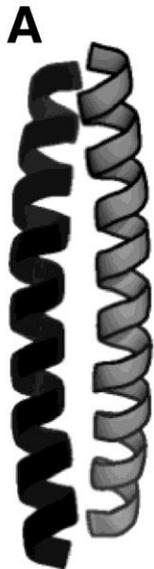
WHAT IS THREADING

Query Sequence: **R****V****L****G****F****I****P****T****W****F****A****L****S****K****Y**



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

Распознавание фолда. Суперспирали



590

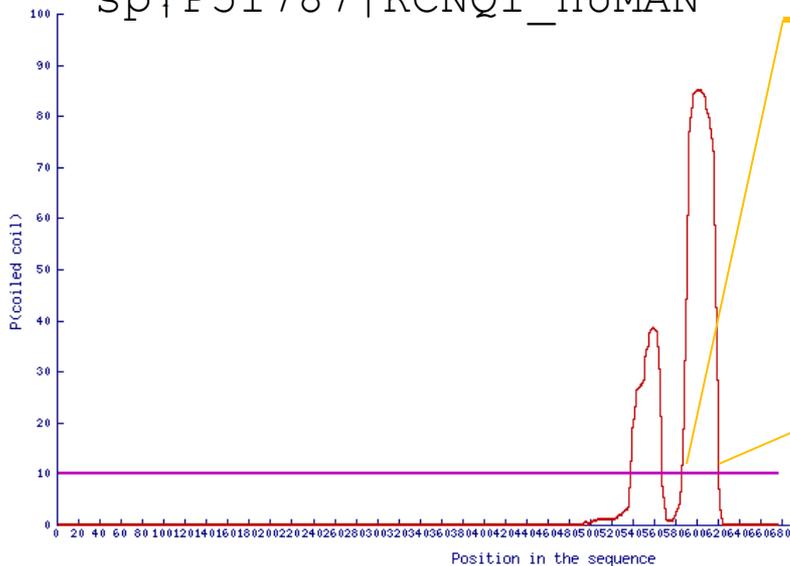
600

610

620

sp|P51787|KCNQ1_HUMAN

IGAR**L**NR**V**EDK**V**T**Q**L**D**Q**R**L**A**L**I**T**D**M**L**H**Q**L**L**S**L**H
 De fgAbcDe fgAbcDe fgAbcDe fgAbcDe fGa



LOGICOIL
 Multi-state coiled-coil oligomeric state prediction

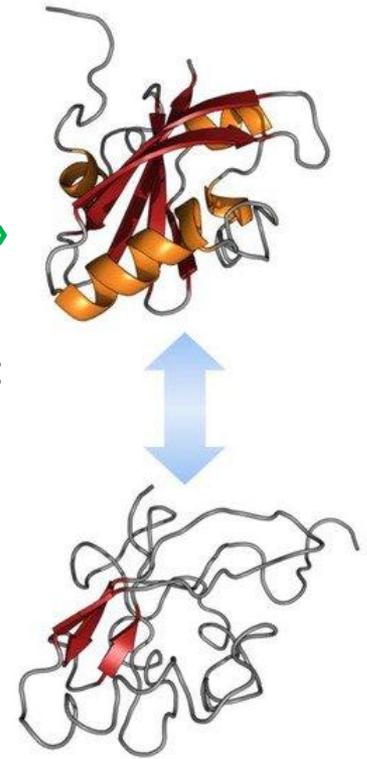
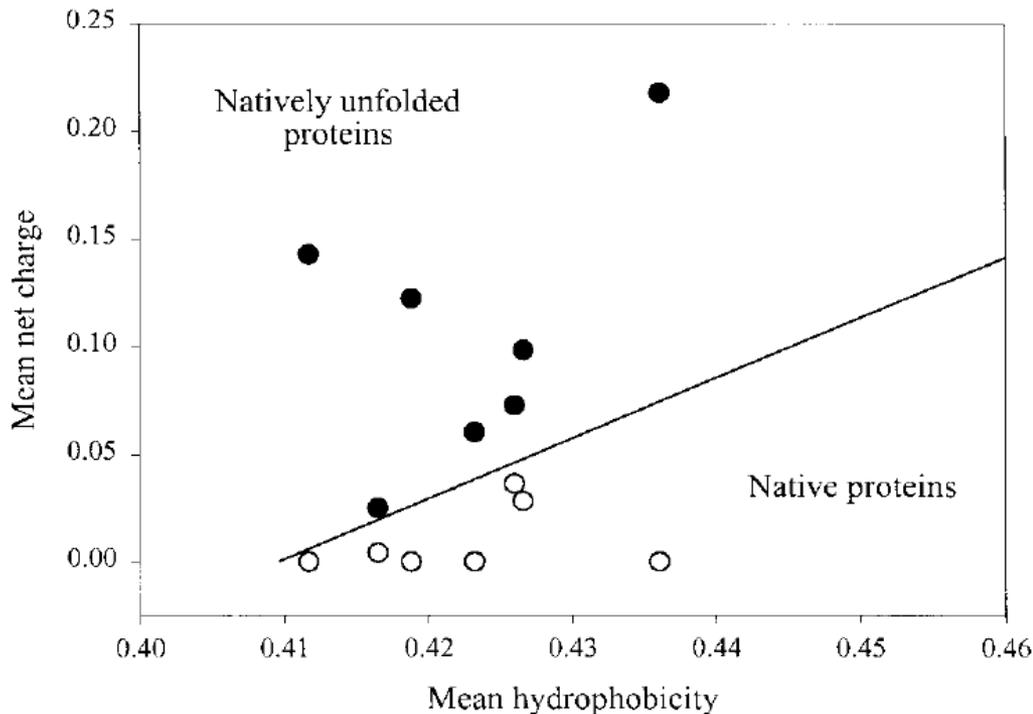
Неупорядоченные белки

Intrinsically disordered proteins –

«нарушители» догмы «структура определяет функцию»

Предсказание неупорядоченных участков по последовательности:

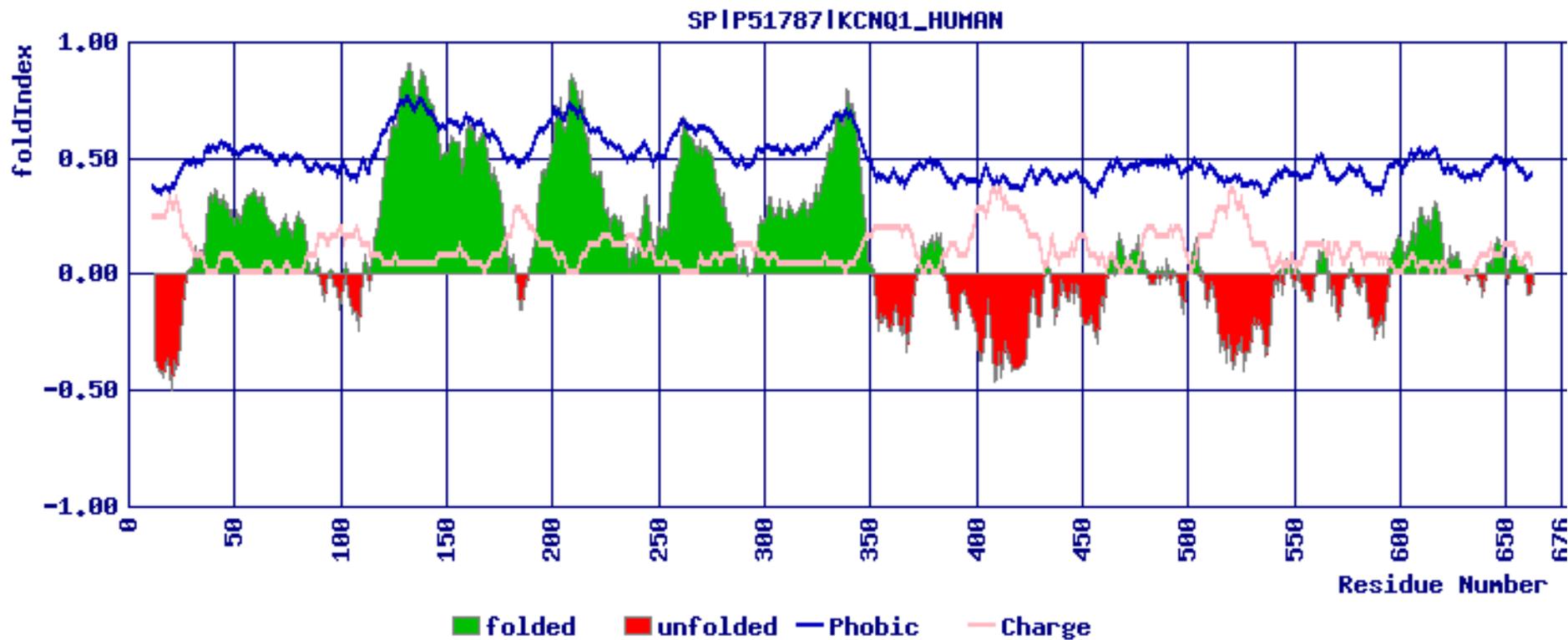
- использование структурных данных (нейронные сети)
- использование свойств аминокислот



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

(Uversky VN, et al. 2000)

Неупорядоченные белки. FoldIndex



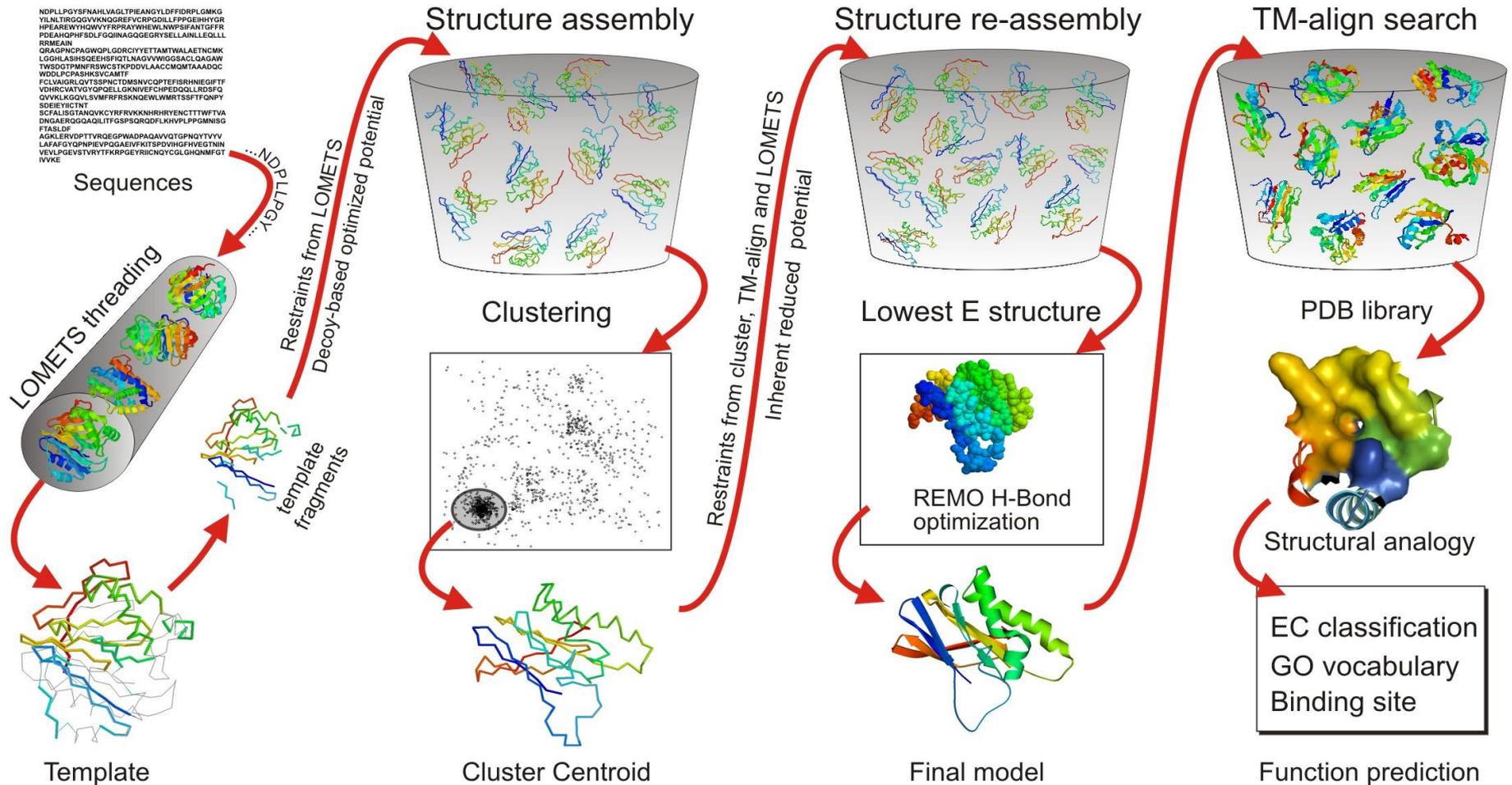
<http://bip.weizmann.ac.il/fldbin/findex>



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

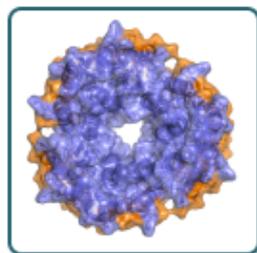
Rosetta Online Server that Includes Everyone

Welcome Queue About ChangeLog Documentation Support

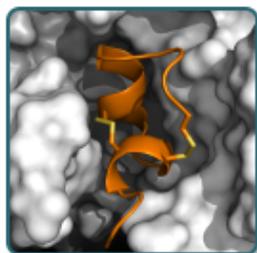
Login Create an account

f Recommend Share 5 G+ 26

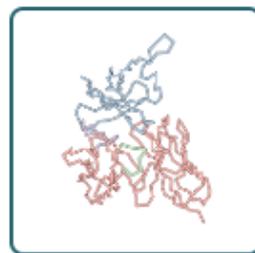
Rosetta Protocols opened for academic users:



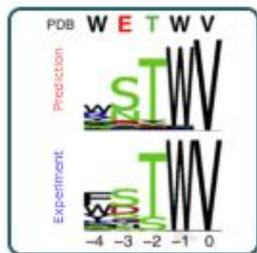
[Mp_lipid_acc]



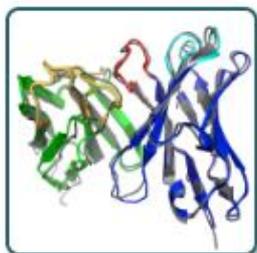
[Tox_dock]



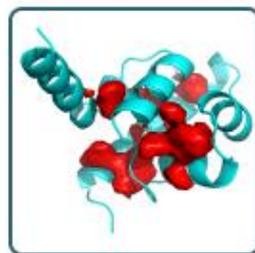
[Snug_dock]



[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

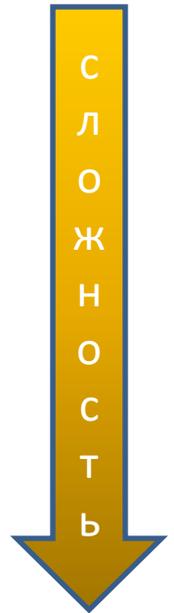
- [ROSIE Documentation](#) - Server related documentation and info.
- [Rosetta Forums](#) This is a list of forums for Rosetta users to discuss problems with running Rosetta and is monitored by Rosetta developers.

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;**
- **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

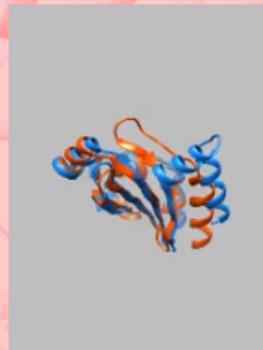
||

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 unprecedented 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.

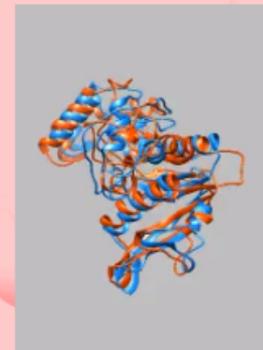
ab initio modeling



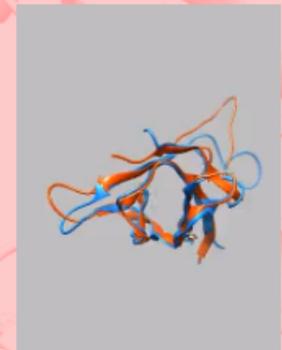
CASP7: T0283-D1
model 321_1: GDT_TS=75



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



You don't have to be a scientist to do science.

By simply running a free program, you can help advance research in medicine, clean energy, and materials science.

[Join Rosetta@home](#)



HHMI
HOWARD HUGHES MEDICAL INSTITUTE



UNIVERSITY OF WASHINGTON



How does it work?

By running **Rosetta@home** on your computer when you're not using it you will speed up and extend our efforts to design new proteins and to

Новости

Rosetta's role in fighting coronavirus

FoldIt

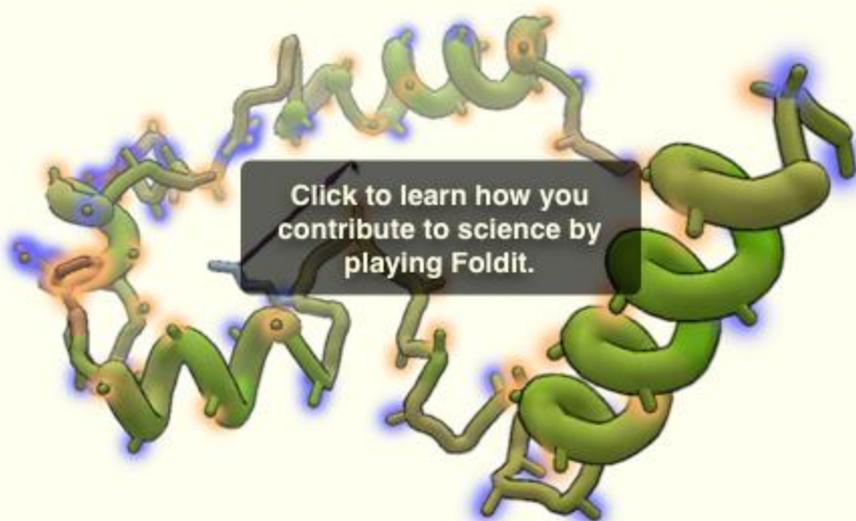


foldit BETA

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

Windows
(7/8/10)



Mac Beta

OSX
(10.7 or later)



Linux Beta

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

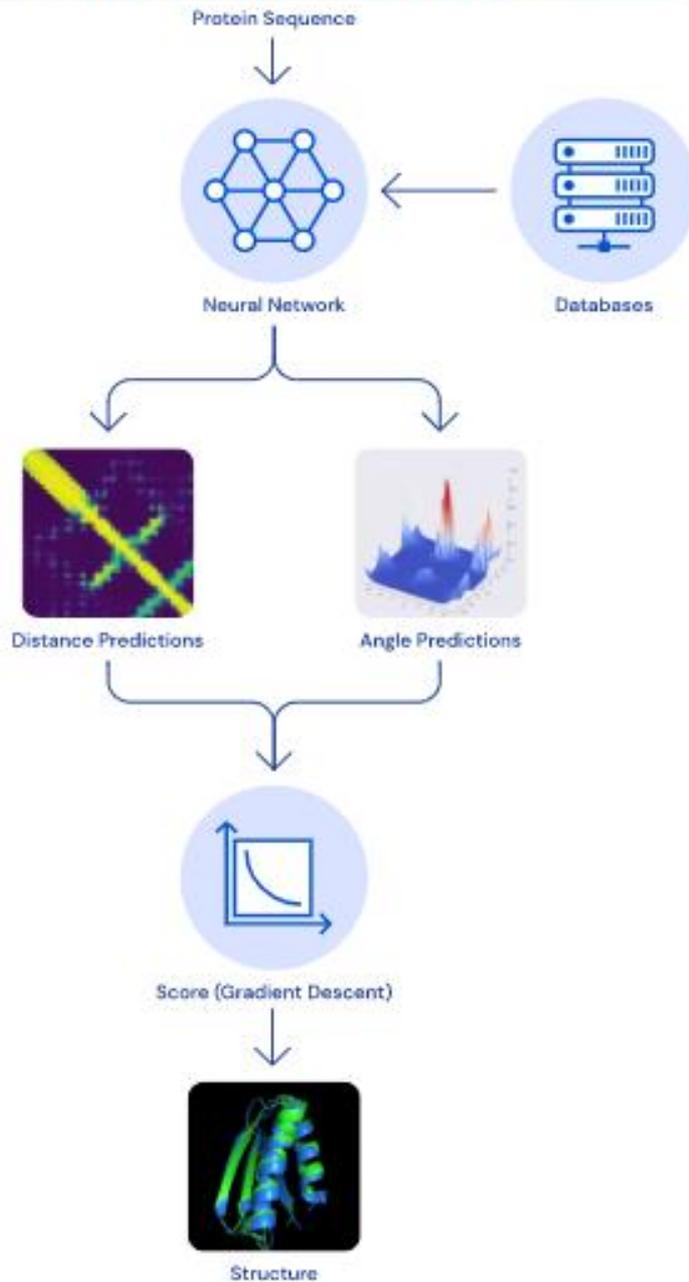


Special update on coronavirus puzzles

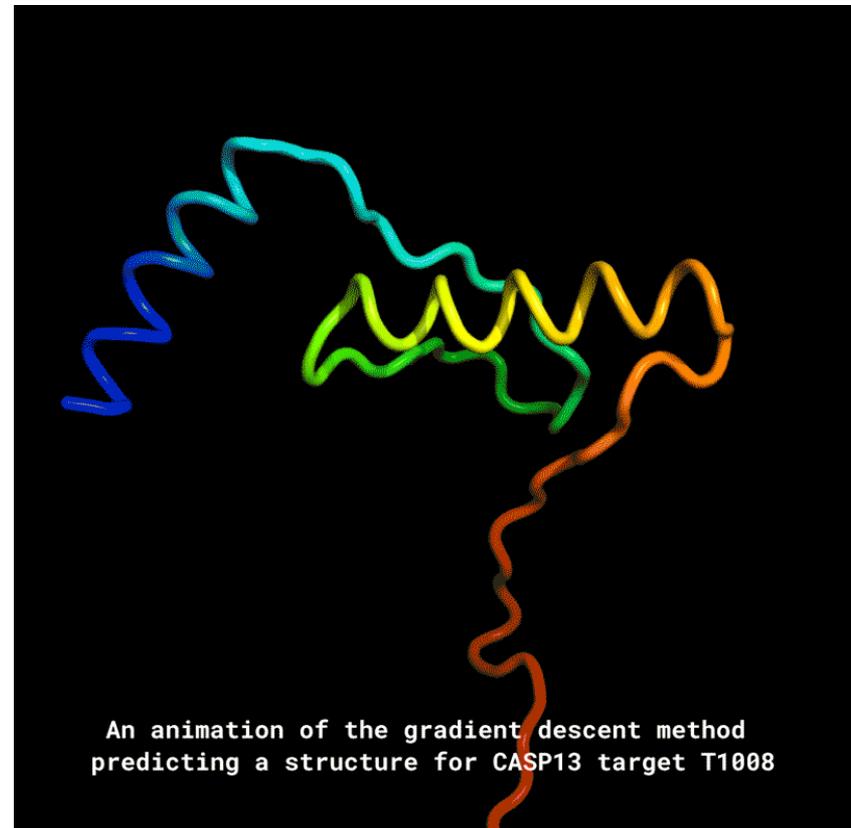
RECOMMEND FOLDIT

Send

QETRKKCTEMKKKFKNCEVRCDESNHCVEVRCSDTKYTLG



AlphaFold



An animation of the gradient descent method predicting a structure for CASP13 target T1008