

ВВЕДЕНИЕ В БИОИНФОРМАТИКУ

Лекция №17

Структурная биоинформатика

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

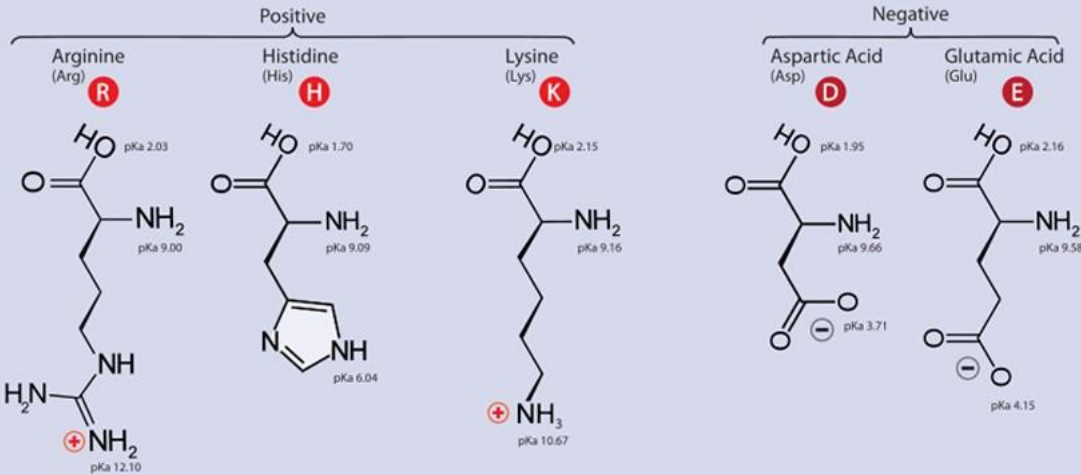
Сайт курса <http://intbio.org/bioinf2020-2021>

Структура аминокислот

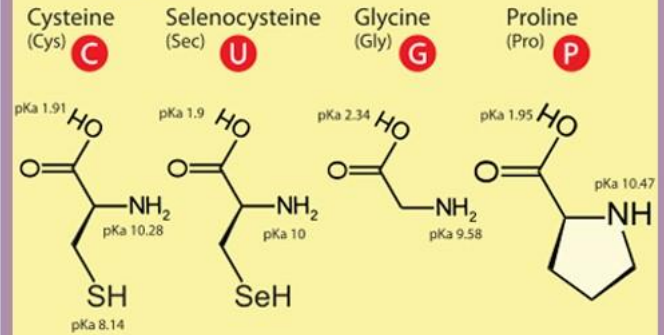
Twenty-One Amino Acids

⊕ Positive ⊖ Negative
 • Side chain charge at physiological pH 7.4

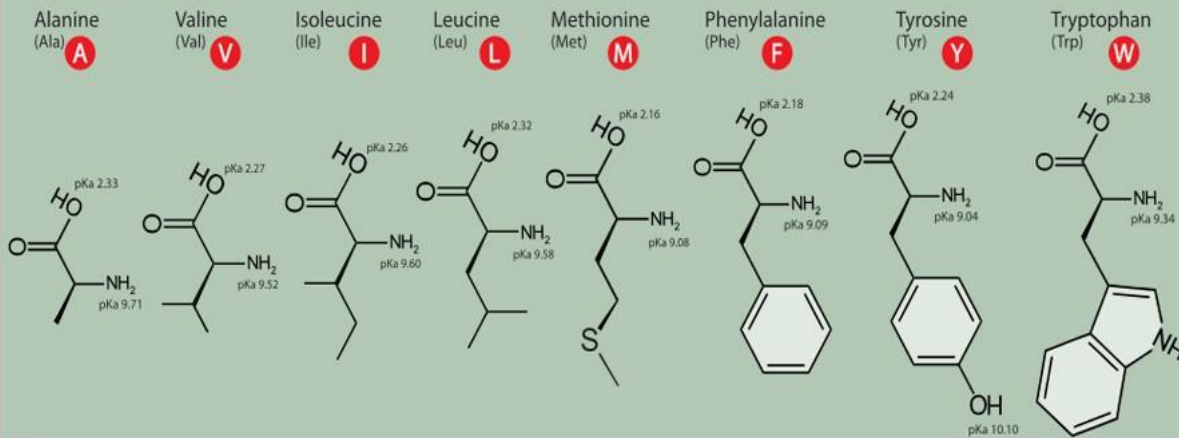
A. Amino Acids with Electrically Charged Side Chains



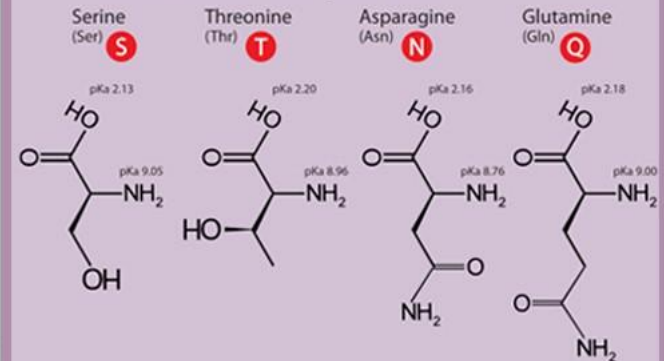
C. Special Cases



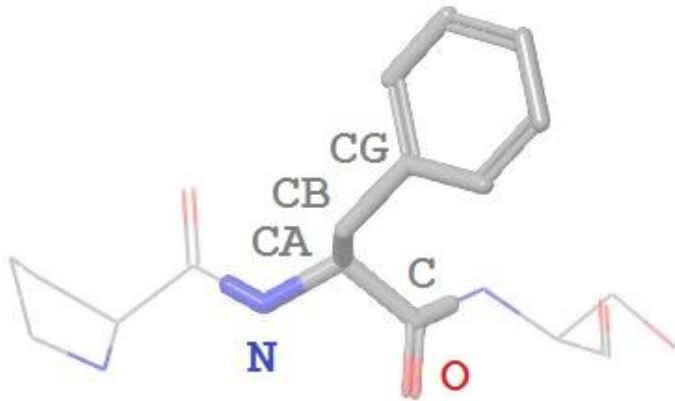
D. Amino Acids with Hydrophobic Side Chain



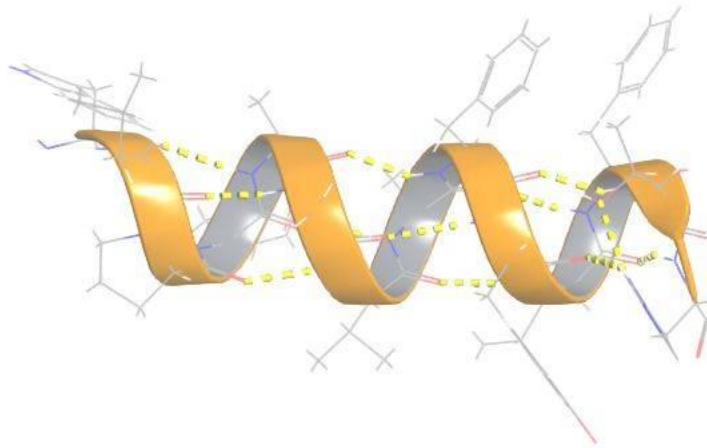
B. Amino Acids with Polar Uncharged Side Chains



Структурная организация белков



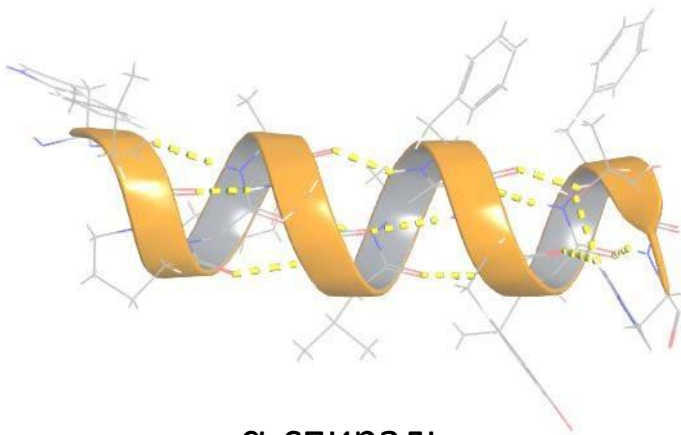
...FYV**PFS**NKT...



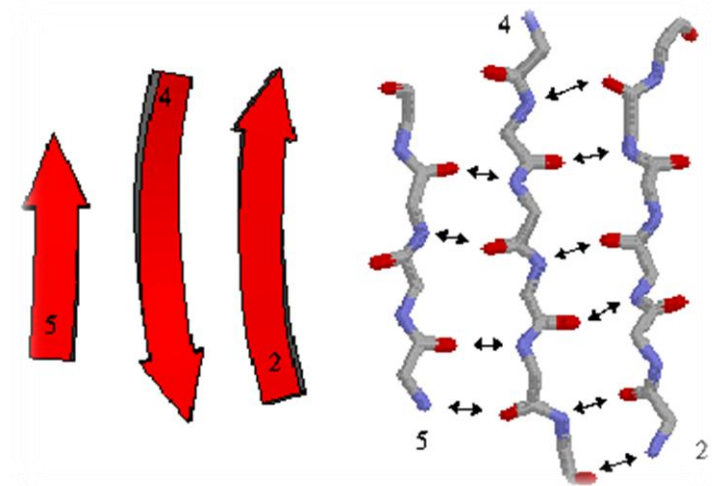
...WLPY**AGVAFYIF**TH...



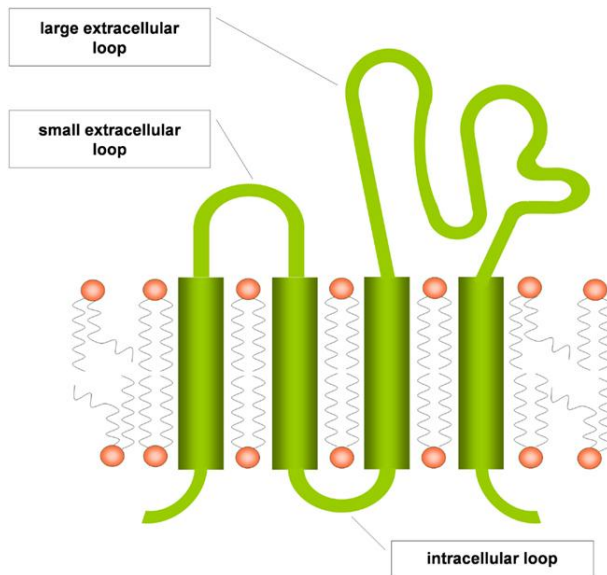
Типы вторичной структуры белков



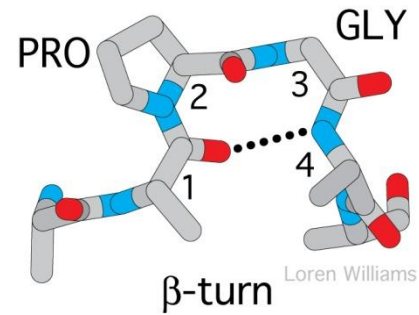
α-спираль



β-лист, состоящий из β-тяжей

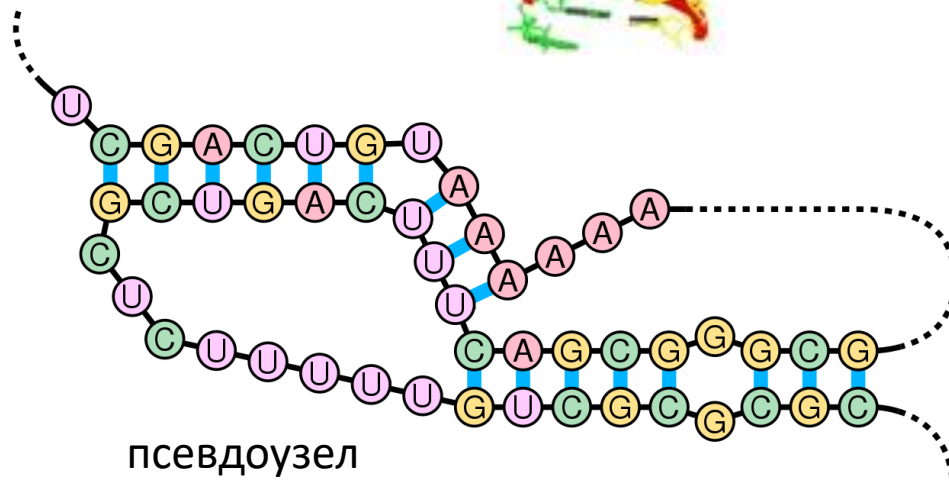
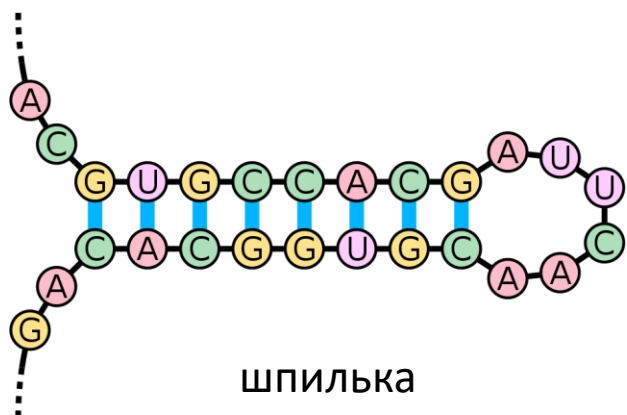
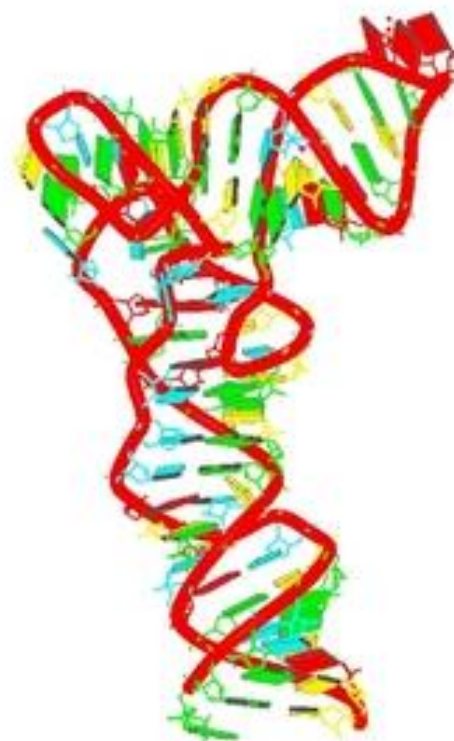
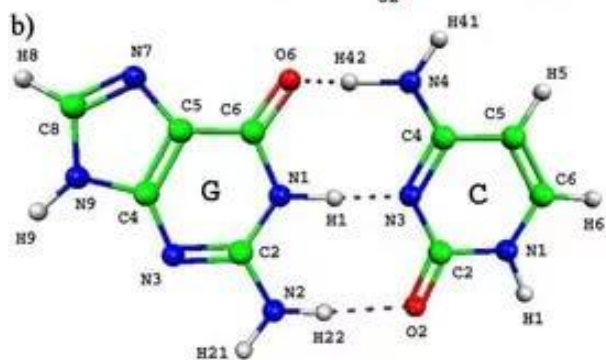
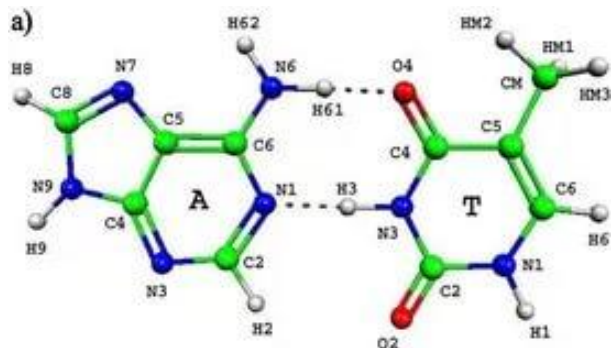


петли



поворот

Структурная организация нуклеиновых кислот



Предсказание вторичной структуры РНК



Sparks Lab

Laboratory of Structural
Bioinformatics

SPOT-RNA: RNA Secondary Structure Prediction using an Ensemble of Two-dimensional Deep Neural Networks and Transfer Learning.

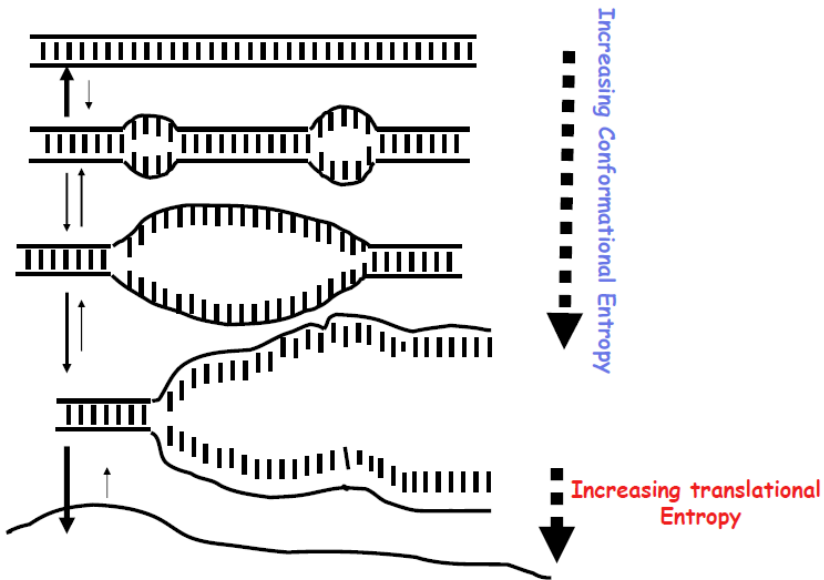
📅 2020-01-29 👤 [Jaswinder Singh](#)

Submit

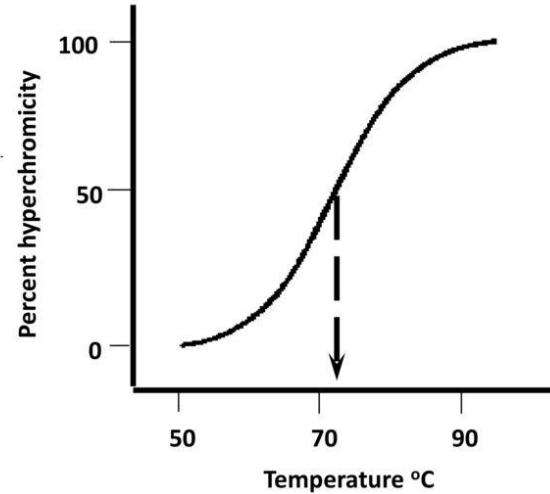
E-mail address (optional):	<input type="text"/>
Target (optional):	<input type="text"/>
Input your RNA Sequences: <input type="button" value="CLEAR"/>	Maximum: 2000 nts, Only one RNA sequence at a time
<pre>> Example sequence 4wj4_B UCCGCGAUAGCUCAGUCGGUAGAGCAAUGACUGUUAUCAUUGGGUCCUGGUU CGAGU CCAGGUCGCGGAGCCA</pre>	

Плавление нуклеиновых кислот

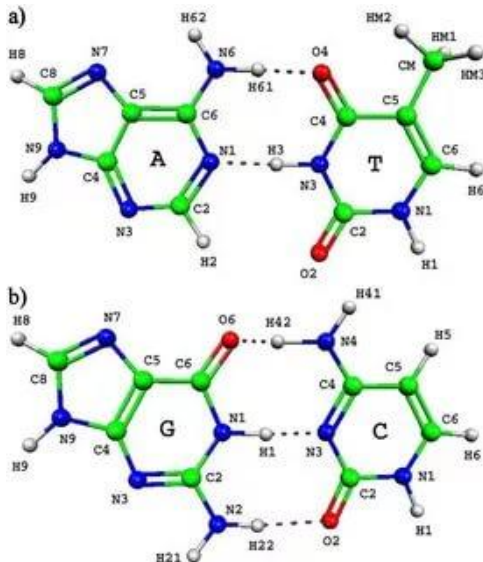
DNA melting is a cooperative process



DNA melting curve



T_m is the temperature at the midpoint of the transition •



(Wetmur and Sninsky, 1995)

$$T_m = 41p(G + C) + 16.6 \lg \left(\frac{[SALT]}{1.0 + 0.7[SALT]} \right) - \frac{500}{L} + 81.5$$

$$[SALT] = [K^+] + 4[Mg^{2+}]^{0.5}$$

Плавление нуклеиновых кислот

Oligo Calc: Oligonucleotide Properties Calculator

Enter Oligonucleotide Sequence Below
OD calculations are for single-stranded DNA or RNA

[Nucleotide base codes](#)

ATC GGT AAC GAT TAC ATT C

Reverse Complement Strand(5' to 3') is:

GAA TGT AAT CGT TAC CGA T

[5' modification](#) (if any)

[3' modification](#) (if any)

Select molecule

nM Primer

Measured Absorbance at 260 nanometers

mM Salt (Na⁺)

Calculate

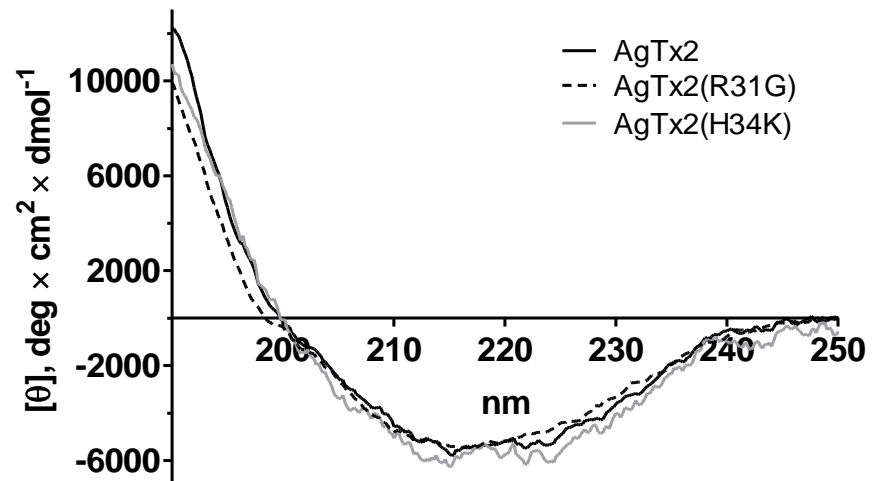
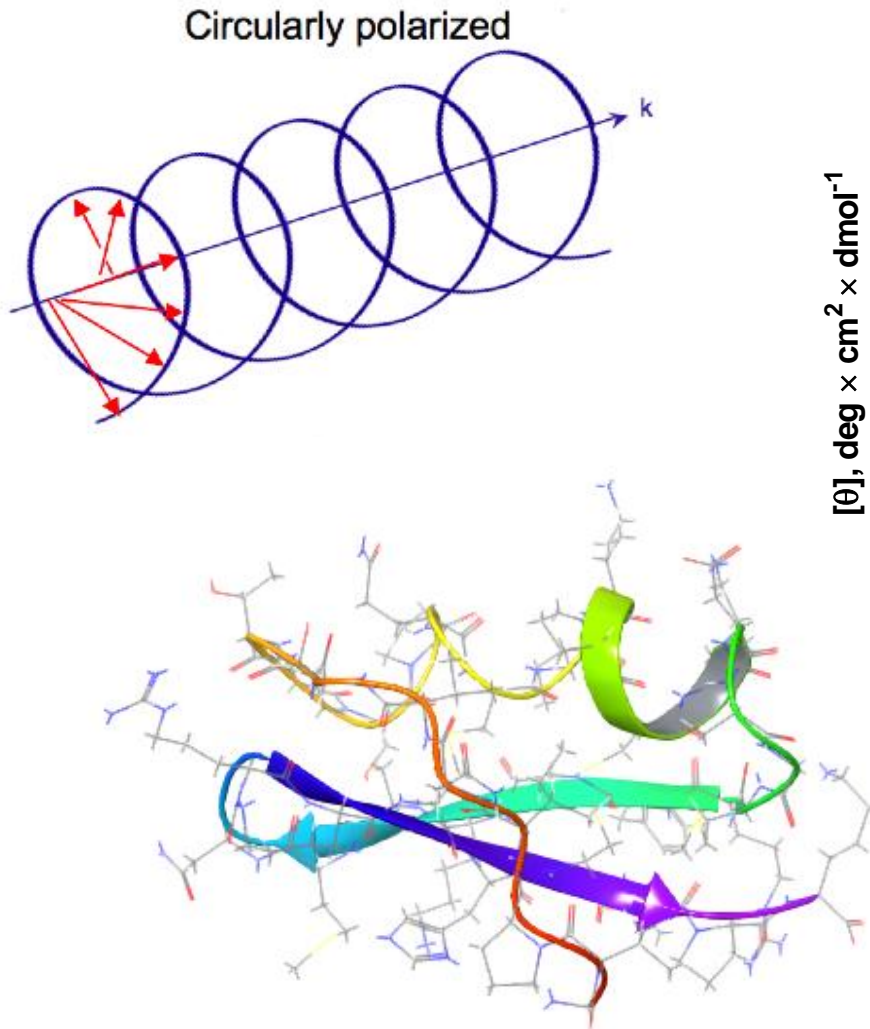
Swap Strands

BLAST

mfold

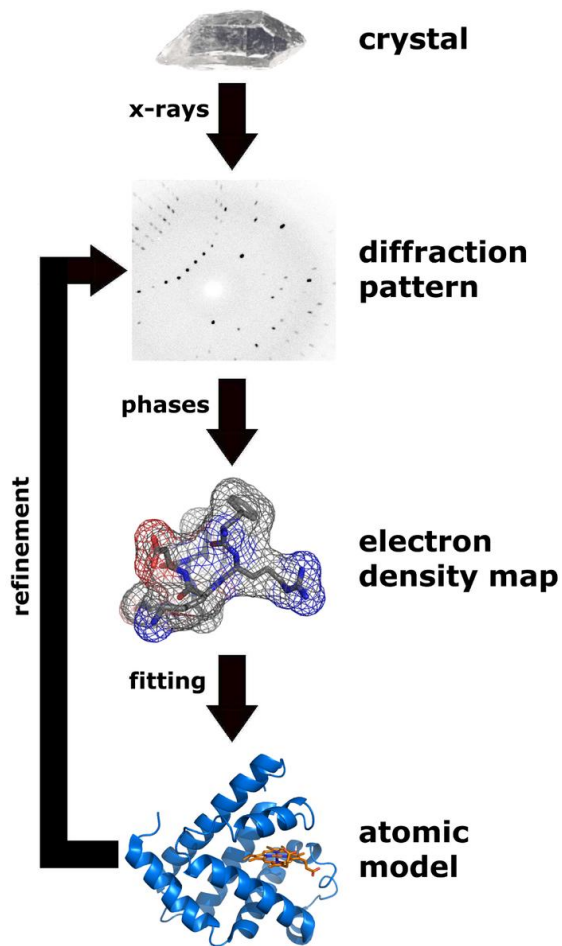
Методы изучения структуры макромолекул

Спектроскопия кругового дихроизма – экспериментальный метод изучения вторичной структуры белков

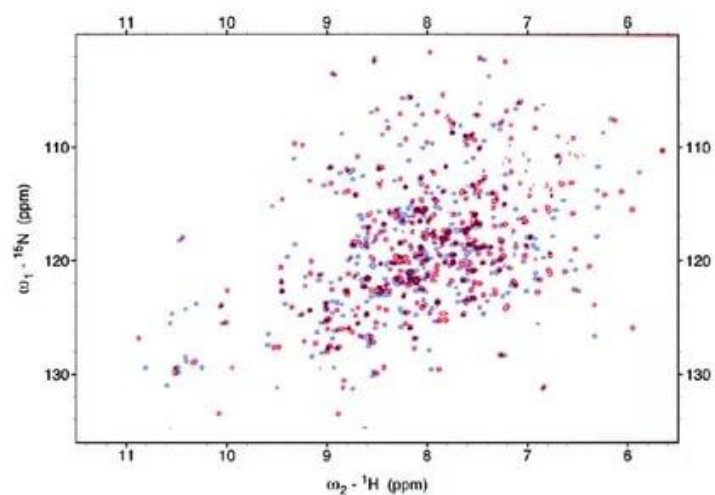


Спектр кругового дихроизма природного аджитоксина и двух его мутантных форм

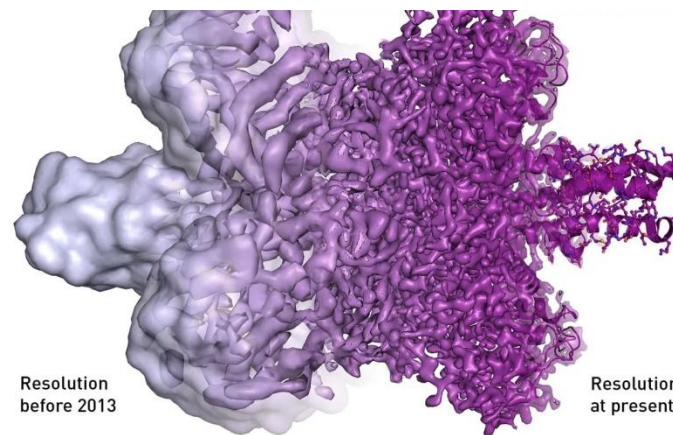
Методы изучения структуры макромолекул



Этапы определения структуры с помощью рентгеновской кристаллографии

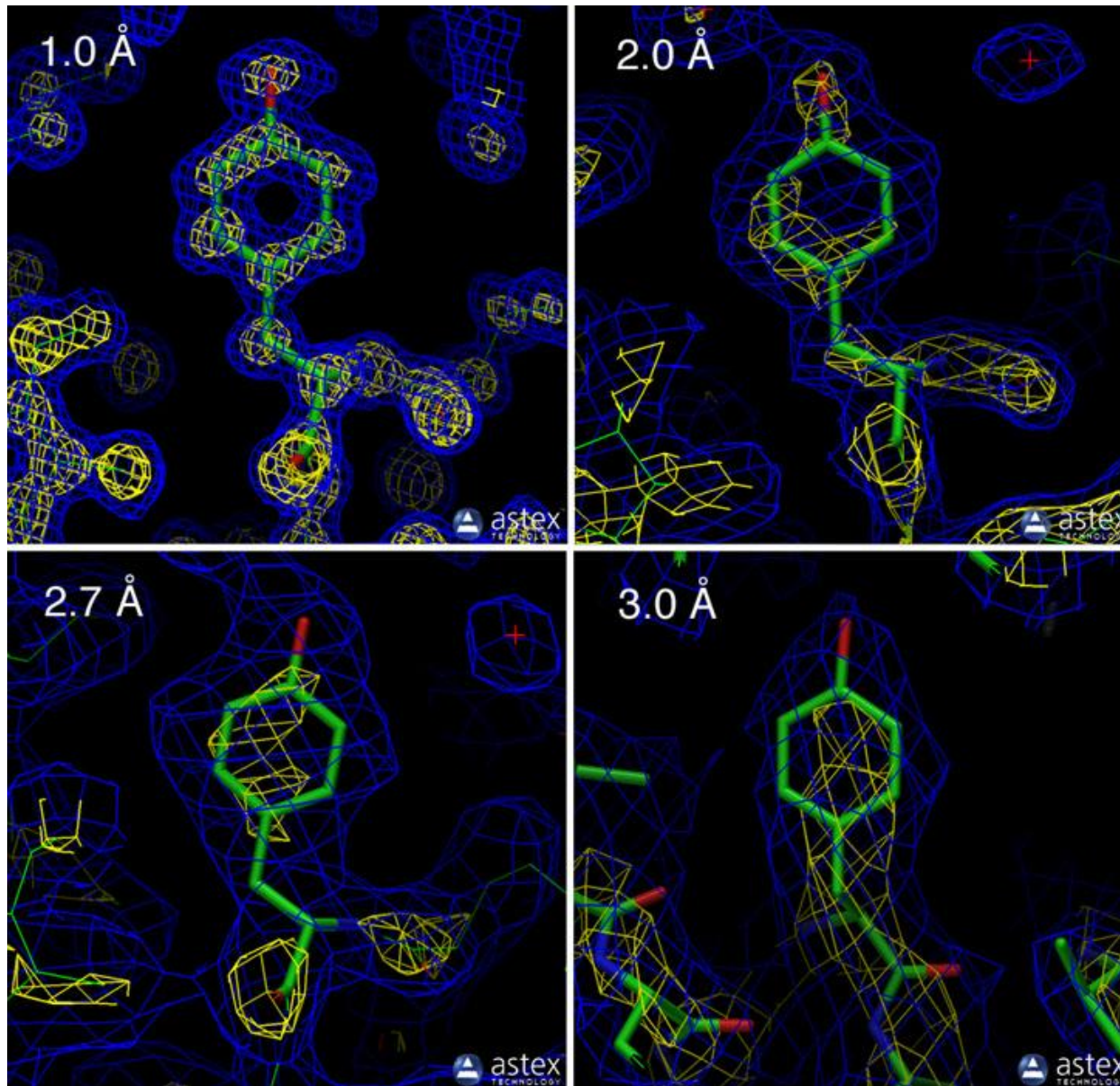


Спектроскопия ядерного магнитного резонанса (ЯМР)



(Крио) электронная микроскопия

Разрешение в РСА и ЭМ



Структурные базы данных. PDB

RCSB **PDB** PROTEIN DATA BANK 170597 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

Enter search term(s)

Advanced Search | Browse Annotations

PDB-101 WORLDWIDE PDB PROTEIN DATA BANK EMDataResource Unified Data Resource for ICM NUCLEIC ACID DATABASE Worldwide Protein Data Bank Foundation

f t y

Welcome

Deposit

Search

Visualize

Analyze

Download

Learn

A Structural View of Biology

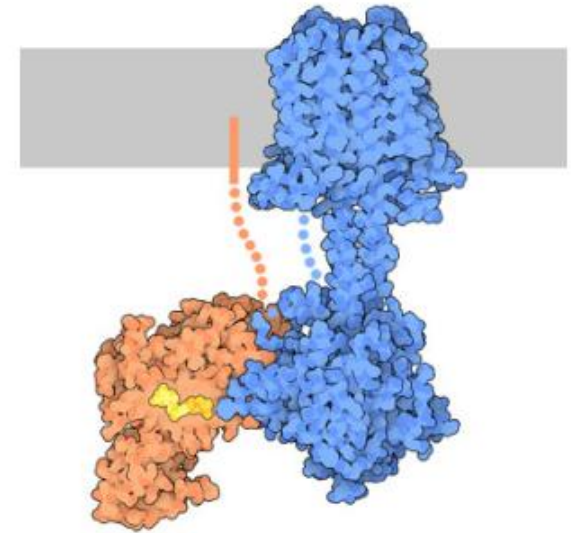
This resource is powered by the Protein Data Bank archive—information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

As a member of the wwPDB, the RCSB PDB curates and annotates PDB data.

The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.



November Molecule of the Month



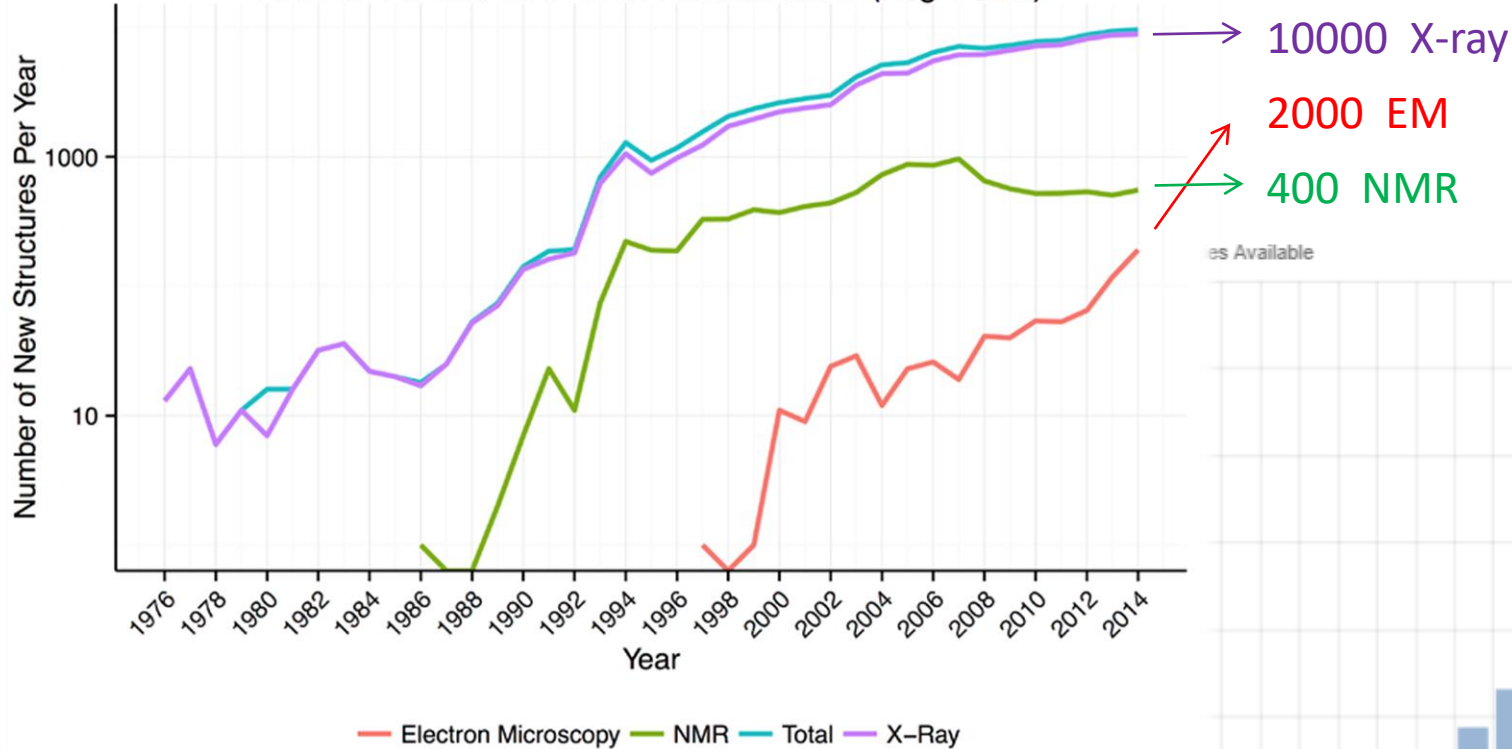
Adenylyl Cyclase

PDB (с 1971 по н.в.) – содержит информацию об экспериментально определенных структурах белков, нуклеиновых кислот и различных комплексов.

<https://www.rcsb.org/pages/about-us/history>

Структурные базы данных. PDB

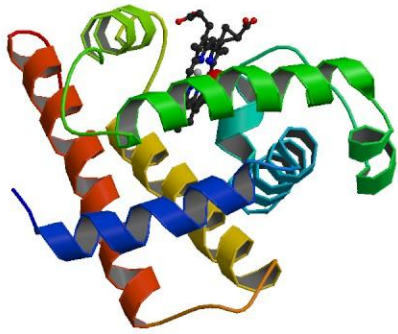
Rate of Protein Structure Determination (Log Scale)



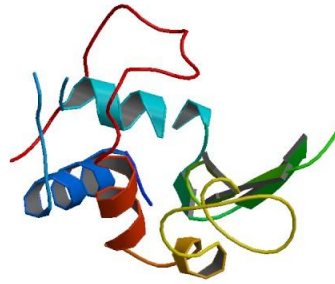
Structures Available



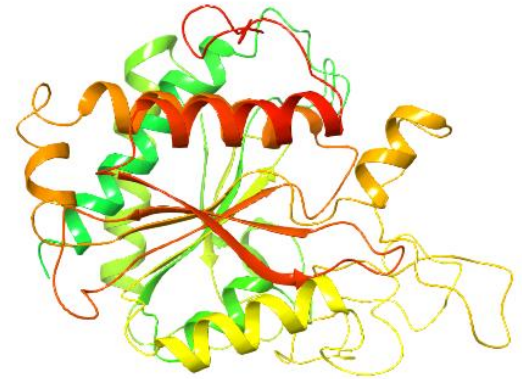
Разнообразиие белковых структур



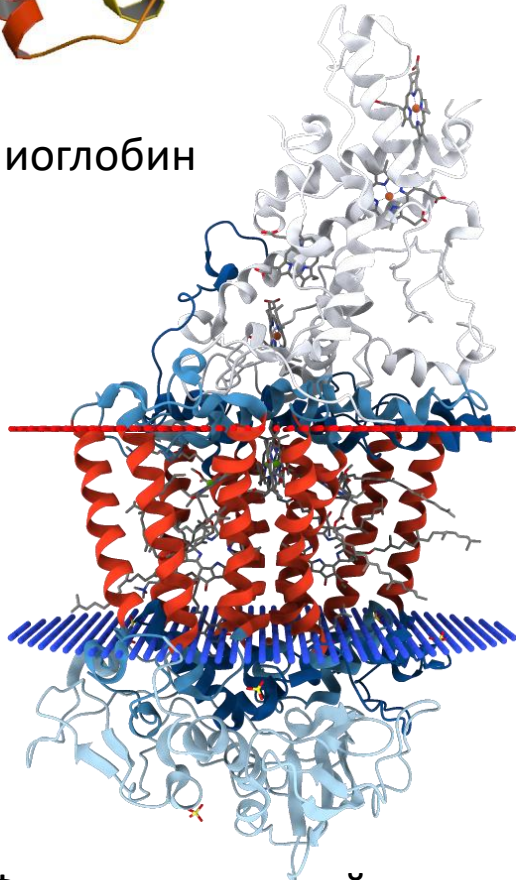
Миоглобин



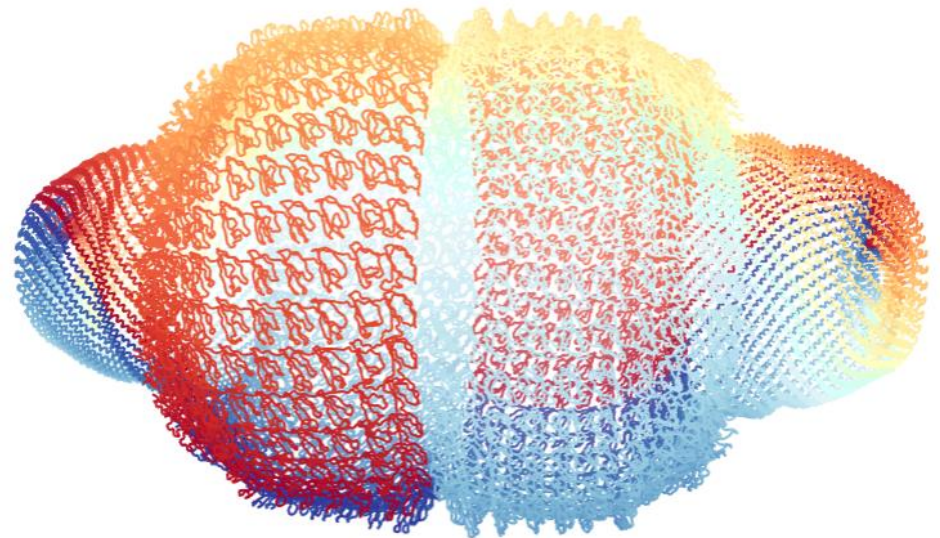
Лизоцим



Карбоксипептидаза



Фотореакционный центр



Структура органеллы vault из печени крысы

Структурные базы данных. PDB

POTASSIUM CHANNEL KCSA IN COMPLEX WITH TETRABUTYLAMMONIUM IN HIGH K

Structure Summary

3D View

Annotations

Experiment

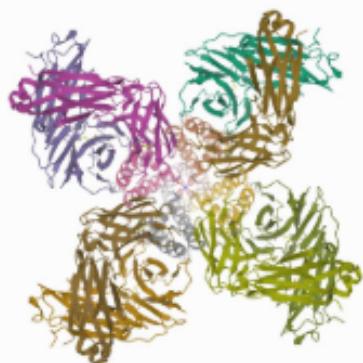
Sequence

Genome

Display Files

Download Files

Biological Assembly 1



2JK5

Potassium Channel KcsA in complex with Tetrabutylammonium in high K

DOI: 10.2210/pdb2JK5/pdb

Classification: IMMUNE SYSTEM/METAL TRANSPORT

Organism(s): *Mus musculus*, *Streptomyces lividans*

Expression System: *Mus musculus*, *Escherichia coli*

Mutation(s): No

Deposited: 2008-08-15 Released: 2009-11-17

Deposition Author(s): Lenaeus, M.J., Focia, P.J., Wagner, T., Gross, A.

Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 2.40 Å

R-Value Free: 0.249

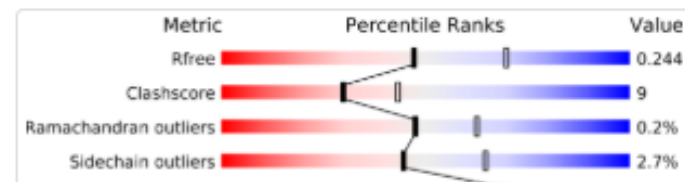
R-Value Work: 0.202

R-Value Observed: 0.205

wwPDB Validation

3D Report

Full Report



3D View: [Structure](#) | [Electron Density](#) | [Ligand Interaction](#)

Global Symmetry: Cyclic - C4 ([3D View](#))

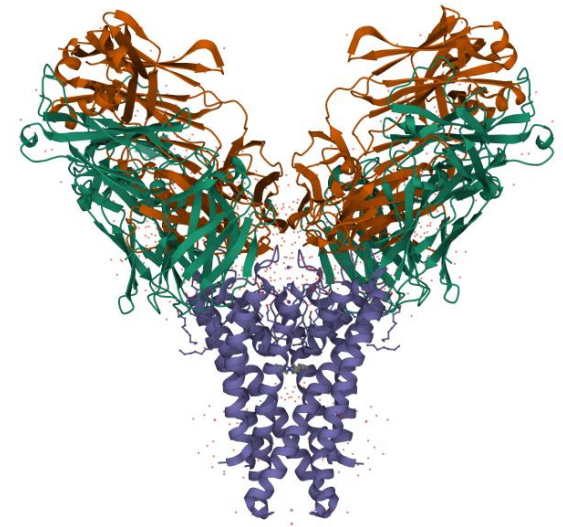
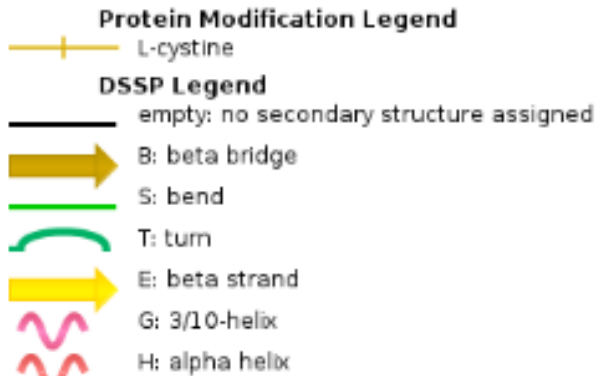
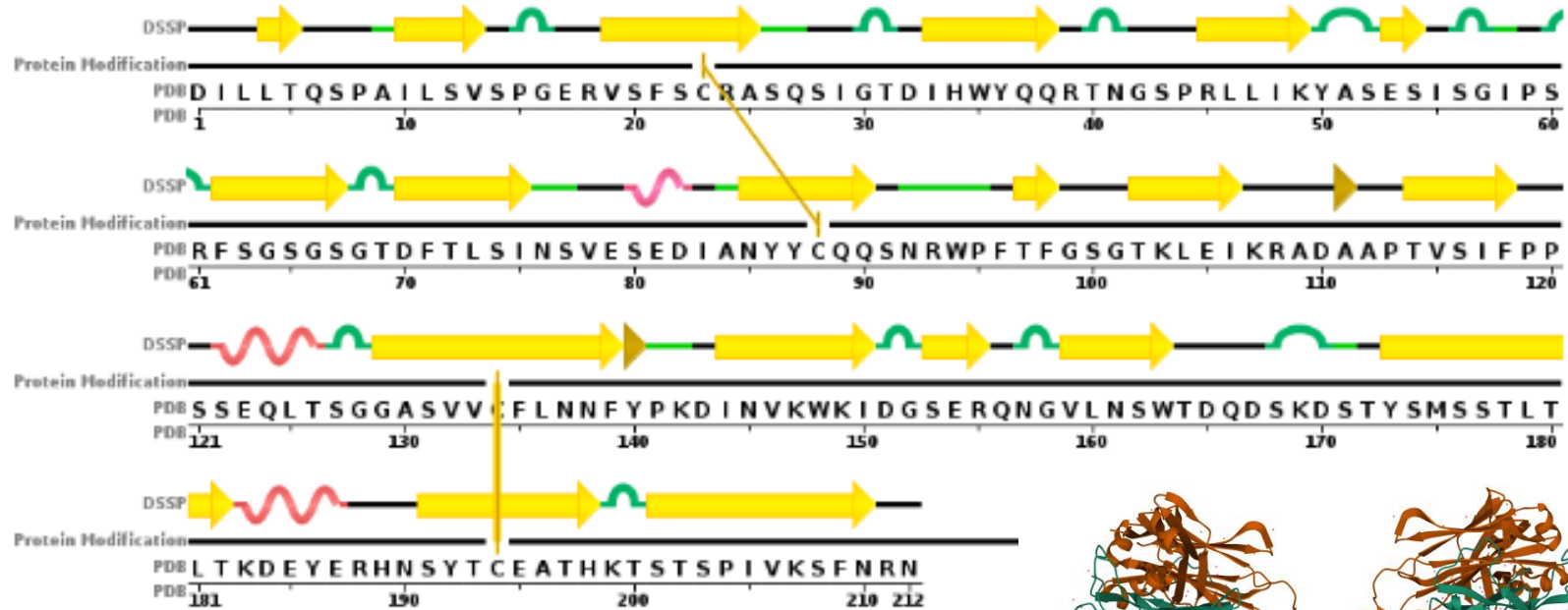
Entity ID: 3

Molecule	Chains	Sequence Length	Organism	Details	Image
VOLTAGE-GATED POTASSIUM CHANNEL	C	124	Streptomyces lividans	Mutation(s): 0 ⓘ Gene Names: kcsA , skc1	

Структурные базы данных. PDB

POTASSIUM CHANNEL KCSA IN COMPLEX WITH TETRABUTYLAMMONIUM IN HIGH K

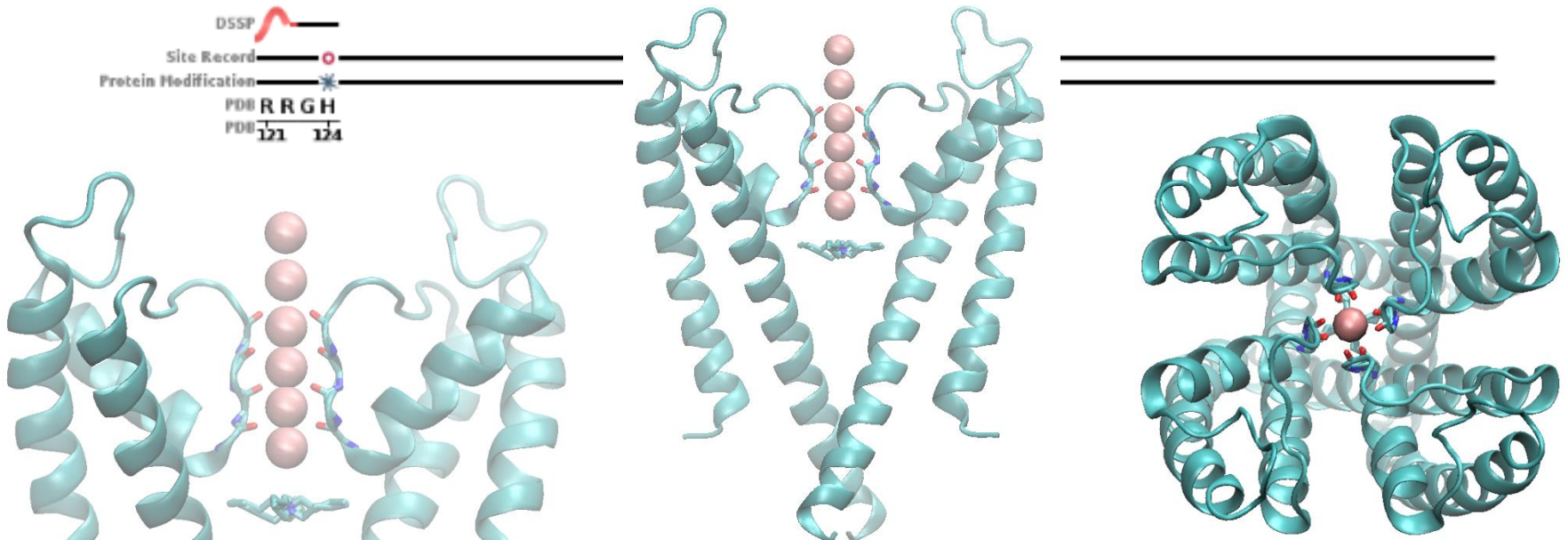
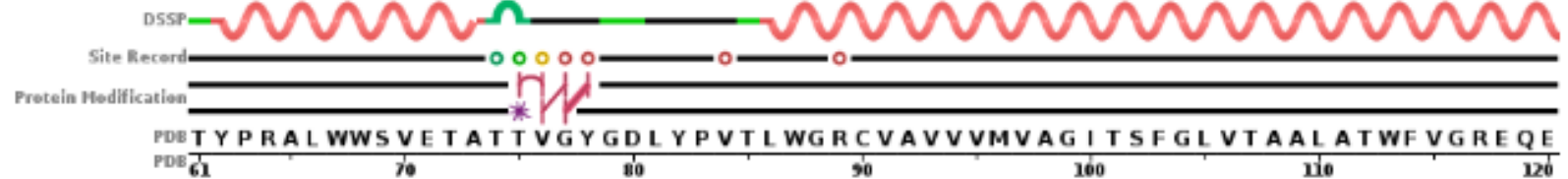
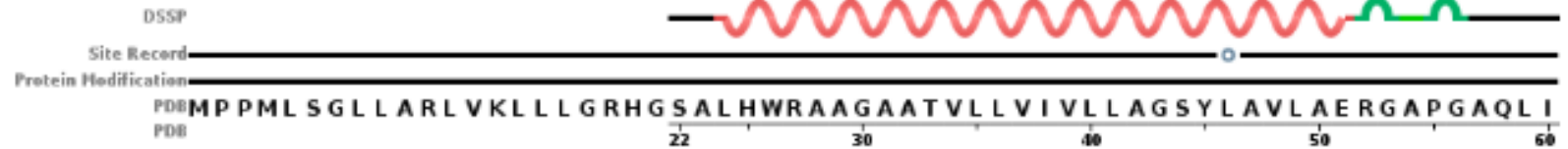
Sequence Chain View



Структурные базы данных. PDB

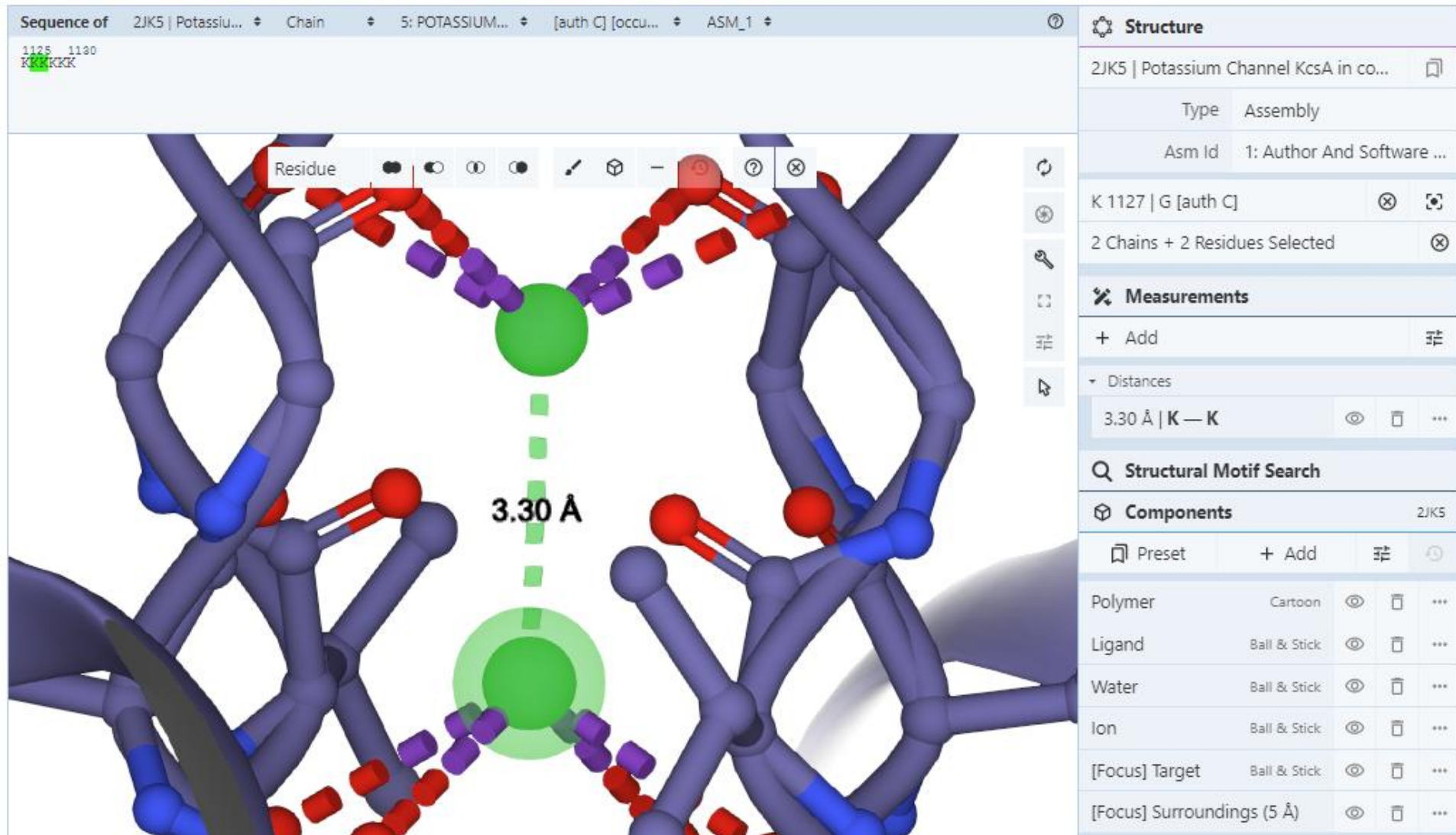
POTASSIUM CHANNEL KCSA IN COMPLEX WITH TETRABUTYLAMMONIUM IN HIGH K

Sequence Chain View



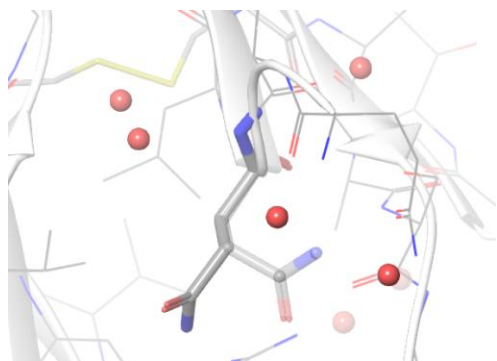
Структурные базы данных. PDB

POTASSIUM CHANNEL KCSA IN COMPLEX WITH TETRABUTYLAMMONIUM IN HIGH K



Структурные файлы. *.pdb

Descriptor		Conformation index					Atom element					
Atom number		Residue name		Chain name			Beta-factor		Occupancy			
Atom name		Chain name		Residue number								
...	ATOM	22	CG	AGLN	A	3	-30.506	25.807	16.125	0.50	43.30	C
	ATOM	23	CD	AGLN	A	3	-29.018	25.546	15.959	0.50	43.01	C
	ATOM	24	OE1	AGLN	A	3	-28.187	26.374	16.332	0.50	43.76	O
	ATOM	25	NE2	AGLN	A	3	-28.675	24.391	15.403	0.50	41.73	N
	ATOM	26	N	BGLN	A	3	-33.005	25.819	17.852	0.50	45.16	N
	ATOM	27	CA	BGLN	A	3	-32.788	24.834	16.796	0.50	44.25	C
	ATOM	28	C	BGLN	A	3	-33.529	23.510	17.024	0.50	43.75	C
...	HETATM	4140	C1	F09	C1131		-21.017	-3.092	-1.563	1.00	60.56	C
	HETATM	4141	C2	F09	C1131		-21.015	-1.597	-1.357	1.00	60.40	C



X



Y



Z coordinates in Å

Define Secondary Structure of Proteins (DSSP)

```

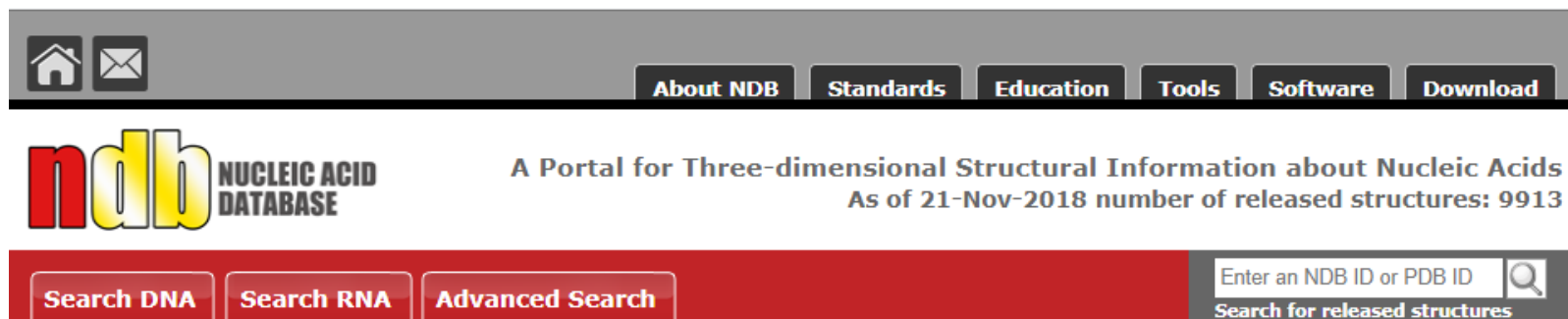
HEADER      HYDROLASE      (SERINE PROTEINASE)          17-MAY-76      1EST
...
 240  1  4  4  0 TOTAL NUMBER OF RESIDUES, NUMBER OF CHAINS,
          NUMBER OF SS-BRIDGES (TOTAL, INTRACHAIN, INTERCHAIN)
10891.0  ACCESSIBLE SURFACE OF PROTEIN (ANGSTROM**2)
 162 67.5  TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(J) ; PER 100 RESIDUES
  84 35.0  TOTAL NUMBER OF HYDROGEN BONDS IN ANTIPARALLEL BRIDGES; PER 100 RESIDUES
...
#  RESIDUE AA STRUCTURE BP1 BP2  ACC   N-H-->O  O-->H-N  N-H-->O  O-->H-N
  2   17   V  B 3   +A 182   0A   8  180,-2.5 180,-1.9   1,-0.2 134,-0.1
...;.....1.....;.....2.....;.....3.....;.....4.....;.....5.....;.....6.....;.....7..
.-- sequential resnumber, including chain breaks as extra residues
|   .-- original PDB resname, not nec. sequential, may contain letters
|   |   .-- amino acid sequence in one letter codeS
|   |   |   .-- secondary structure summary based on columns 19-38
|   |   |   |   xxxxxxxxxxxxxxxxxxxxxxxx recommend columns for secstruc details
|   |   |   |   .-- 3-turns/helix
|   |   |   |   |.-- 4-turns/helix
|   |   |   |   ||.-- 5-turns/helix
|   |   |   |   |||.-- geometrical bend
|   |   |   |   ||||.-- chiralityS
|   |   |   |   |||||.-- beta bridge label
|   |   |   |   |||||.-- beta bridge label
|   |   |   |   |||||.-- beta bridge partner resnum
|   |   |   |   |||||.-- beta bridge partner resnum
|   |   |   |   |||||.-- beta sheet label
|   |   |   |   |||||.-- solvent accessibility
|   |   |   |   |||||.
#  RESIDUE AA STRUCTURE BP1 BP2  ACC
|   |   |   |   |||||.
|   |   |   |   |||||.
35  47   I  E   +   0  0   2
36  48   R  E > S- K  0 39C 97
37  49   Q  T 3  S+  0  0  86      (example from 1EST)
38  50   N  T 3  S+  0  0  34

```

Программа для расчёта
водородных связей и
определения типов
вторичной структуры
остатков в белках
(Kabsch & Sander, 1983)

Структурные базы данных. NDB

NDB (с 1992 по н.в.) – сбор и распространение информации о структуре нуклеиновых кислот. Формат хранения данных идентичен PDB.



The screenshot shows the top navigation bar of the NDB website. It includes a home icon, an email icon, and a menu with links for 'About NDB', 'Standards', 'Education', 'Tools', 'Software', and 'Download'. Below the navigation bar is the NDB logo (stylized 'ndb' in red and yellow) and the text 'NUCLEIC ACID DATABASE'. To the right, it states 'A Portal for Three-dimensional Structural Information about Nucleic Acids' and 'As of 21-Nov-2018 number of released structures: 9913'. At the bottom of the header, there are three search buttons: 'Search DNA', 'Search RNA', and 'Advanced Search', along with a search input field containing the text 'Enter an NDB ID or PDB ID' and a magnifying glass icon, with the label 'Search for released structures' below it.

Welcome to the NDB

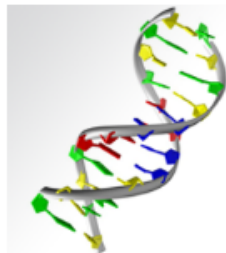
The NDB contains information about experimentally-determined nucleic acids and complex assemblies. Use the NDB to perform searches based on annotations relating to sequence, structure and function, and to download, analyze, and learn about nucleic acids.

Search Structures

[Search DNA](#)
Search DNA and its complexes

[Search RNA](#)
Search for RNA structures in the NDB archive or in the Non-Redundant list

[Advanced Search](#)
Search for structures based on structural features, chemical features, binding modes, citation and experimental information



Featured Tools

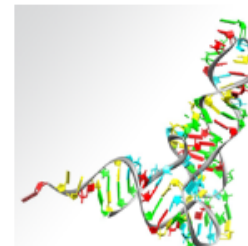
[RNA 3D Motif Atlas](#), a representative collection of RNA 3D internal and hairpin loop motifs

[Non-redundant Lists](#) of RNA-containing 3D structures

[RNA Base Triple Atlas](#), a collection of motifs consisting of two RNA basepairs

[WebFR3D](#), a webserver for symbolic and geometric searching of RNA 3D structures

[R3D Align](#), an application for detailed nucleotide to nucleotide alignments of RNA 3D structures



Структурные базы данных. NDB

COMPLEX BETWEEN NUCLEOSOME CORE PARTICLE (H3,H4,H2A,H2B) AND 146 BP LONG DNA FRAGMENT

NDB ID: PD0001 **PDB ID:** [1AOI](#) 

Title:

COMPLEX BETWEEN NUCLEOSOME CORE PARTICLE (H3,H4,H2A,H2B) AND 146 BP LONG DNA FRAGMENT

Molecular Description:

COMPLEX BETWEEN NUCLEOSOME CORE PARTICLE (H3
H4
H2A
H2B) AND 146 BP LONG DNA FRAGMENT

Structural Keywords:

B DOUBLE HELIX


Nucleic Acid Sequence:

[Click to show/hide 1 nucleic acid sequences](#)

Protein Sequence:

[Click to show/hide 4 protein sequences](#)

Primary Citation:

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., Richmond, T.J.
[Crystal structure of the nucleosome core particle at 2.8 Å resolution.](#) 
Nature, **389**, pp. 251 - 260, 1997.

Experimental Information:

X-RAY DIFFRACTION

Space Group:

P 21 21 21

Cell Constants:

a = 106.04 b = 181.78 c = 110.12 (Ångstroms)

$\alpha = 90.0$ $\beta = 90.0$ $\gamma = 90.0$ (degrees)

Refinement:

The structure was refined using the X-PLOR program. The R value is 0.0 for 51237 reflections in the resolution range 25.0 to 2.8 Ångstroms with Fobs > 0.0 sigma(Fobs) and with I > 0.0 sigma(I)

Structural Features

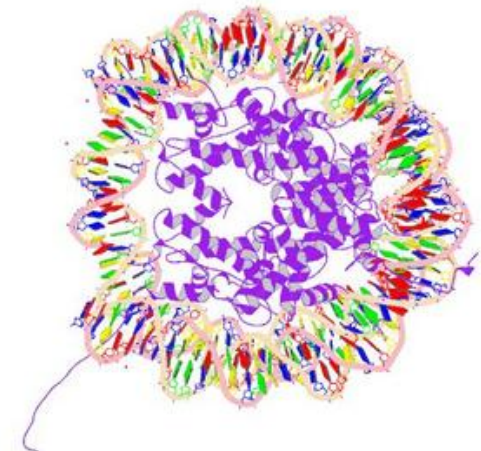
[Base Pair Hydrogen Bonding Classification](#)

[Nucleic Acid Backbone Torsions](#)

[Base Pair Morphology Parameters](#)

[Base Pair Morphology Step Parameters](#)

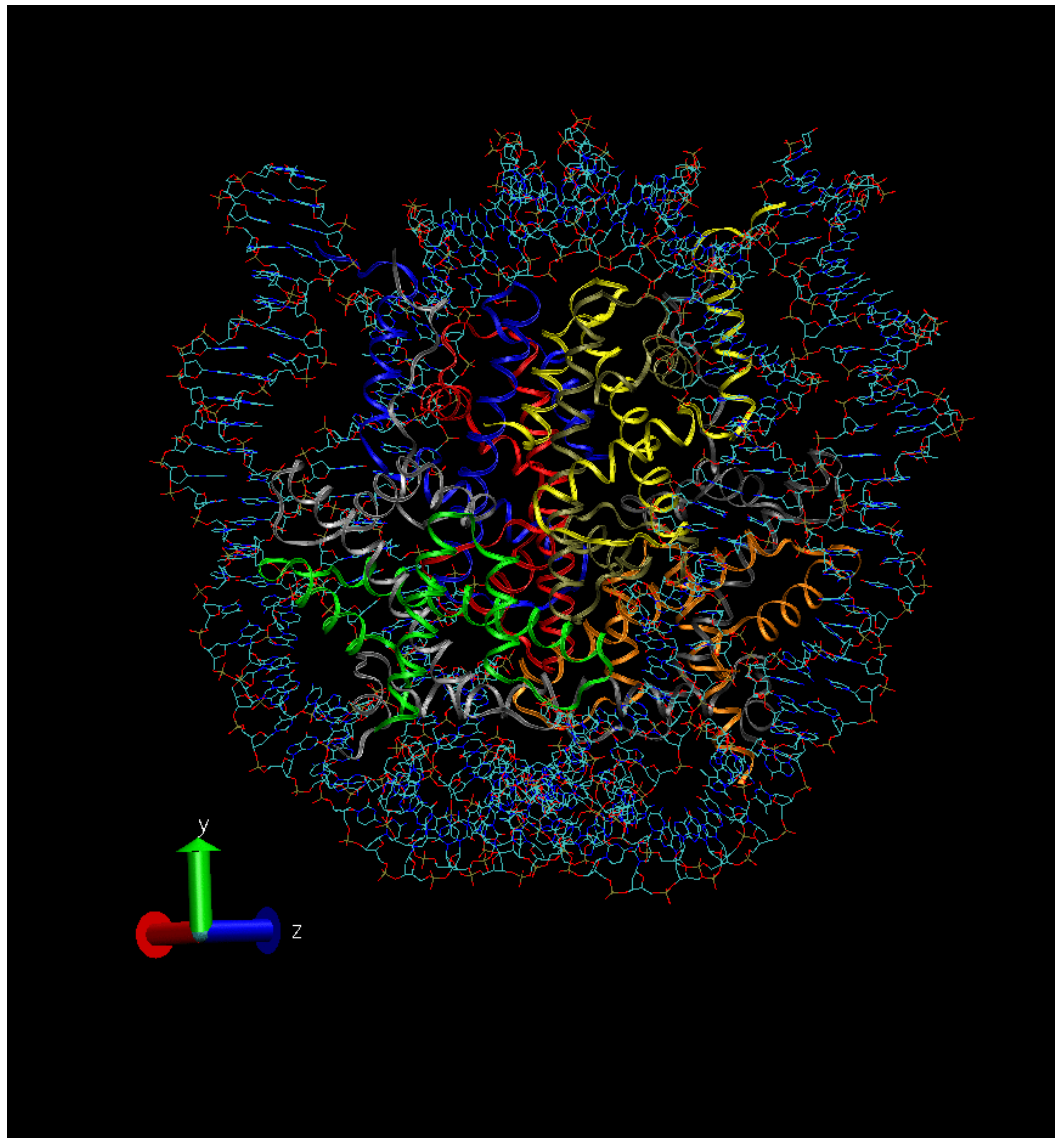
Biological Assembly 1



[More Images...](#)



Структурные базы данных. NDB

Crystal Structure of Nucleosome Core Particle Composed of the Widom 601 DNA Sequence (pdb 3LZ0)



Структурные базы данных. EMDB

EMBL-EBI [Terms of Use](#) | [Privacy](#) | [Cookies](#)

[Databases](#) | [Tools](#) | [Research](#) | [Training](#) | [Industry](#) | [About Us](#) | [Help](#) | [Site Index](#)  


[Home](#) **EMD-1367** [Mirrors](#) [Contact us](#)

Title: Three-dimensional structure of a voltage-gated potassium channel at 2.5 nm resolution.

Authors: Olga Sokolova, Ludmila Kolmakova-Partensky and Nikolaus Grigorieff

Sample: Shaker B channel

Aggregation state: Single particle (25 angstroms resolution)

 **Latest update:** 2011-05-26

- [Summary](#)
- [Experimental details](#)
- [Visualization](#)
- [Map information](#)
- [Downloads](#)

Summary

Status: Released

Deposition date: 2007-05-24

Header release date: 2007-05-30

Map release date: 2007-05-30

Primary citation: Sokolova O, Kolmakova-Partensky L, Grigorieff N: Three-dimensional structure of a voltage-gated potassium channel at 2.5 nm resolution. *STRUCTURE* (2001) **9**, pp. 215-220 [PubMed 11286888]

Sample: Shaker B channel

Resolution: 25 Å (determined by FSC at 0.5 cut-off)

Fitted PDB:	PDB Authors	PubMed Status
1bl8	Doyle, D.A., Cabral, J.M., Pfuetzner, R.A., Kuo, A., Gulbis, J.M., Cohen, S.L., Chait, B.T., Mackinnon, R.	9525859 Released
1a68	Kreusch, A., Pfaffinger, P.J., Stevens, C.F., Choe, S.	9582078 Released

Структурные базы данных. EMDB

EMDB › EMD-10656

E. coli 70S ribosome in complex with dirithromycin, and deacylated tRNA(iMet) (focused classification).


Source organism: *Escherichia coli K-12* [83333]

Fitted atomic model: [6xza](#)

Related EM entry by publication: [EMD-10655](#), [EMD-10657](#)

3Dbionotes: [available for this entry](#) 

Primary publication:

 Insights into the improved macrolide inhibitory activity from the high-resolution cryo-EM structure of dirithromycin bound to the E. coli 70S ribosome.

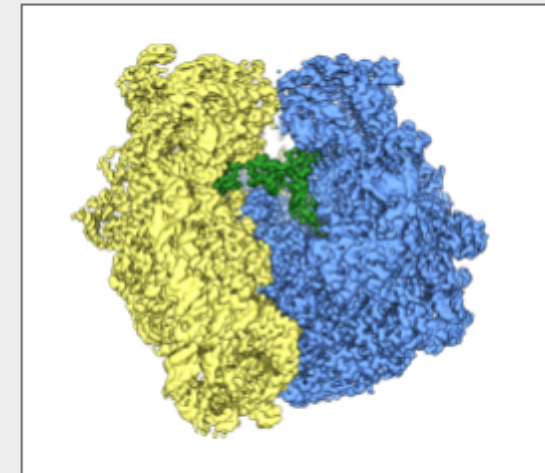
Pichkur EB, Paleskava A, Tereshchenkov AG, Kasatsky P, Komarova ES, Shiriaev DI, Bogdanov AA, Dontsova OA, Osterman IA, Sergiev PV, Polikanov YS, Myasnikov AG, Konevega AL

Rna **26** 715-723 (2020)

PMID: [32144191](#)

Single particle reconstruction
2.66Å resolution

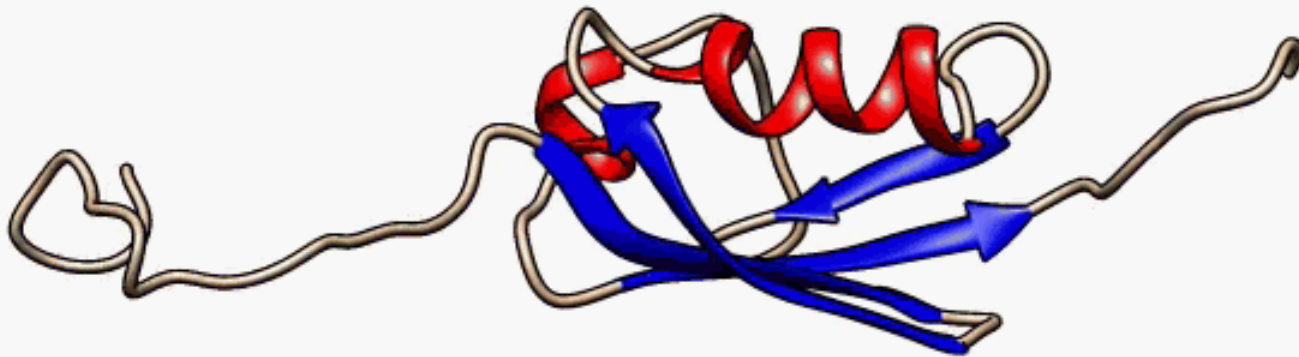
Map released: 2020-11-04



Неупорядоченные белки

Intrinsically disordered proteins - «нарушители» догмы «структура определяет функцию»

STRUCTURE DETERMINATION OF THE SMALL UBIQUITIN-RELATED MODIFIER SUMO-1,
NMR, 10 STRUCTURES (pdb 1A5R)



Задание поворота в пространстве

Матрица поворота:

$$M_x(\alpha) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & -\sin \alpha \\ 0 & \sin \alpha & \cos \alpha \end{pmatrix}$$

$$M_y(\alpha) = \begin{pmatrix} \cos \alpha & 0 & \sin \alpha \\ 0 & 1 & 0 \\ -\sin \alpha & 0 & \cos \alpha \end{pmatrix}$$

$$M_z(\alpha) = \begin{pmatrix} \cos \alpha & -\sin \alpha & 0 \\ \sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

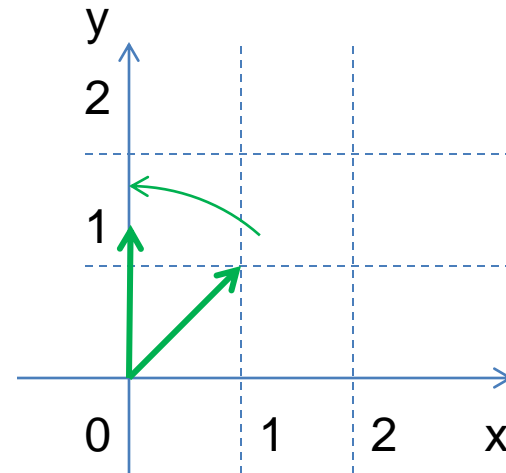
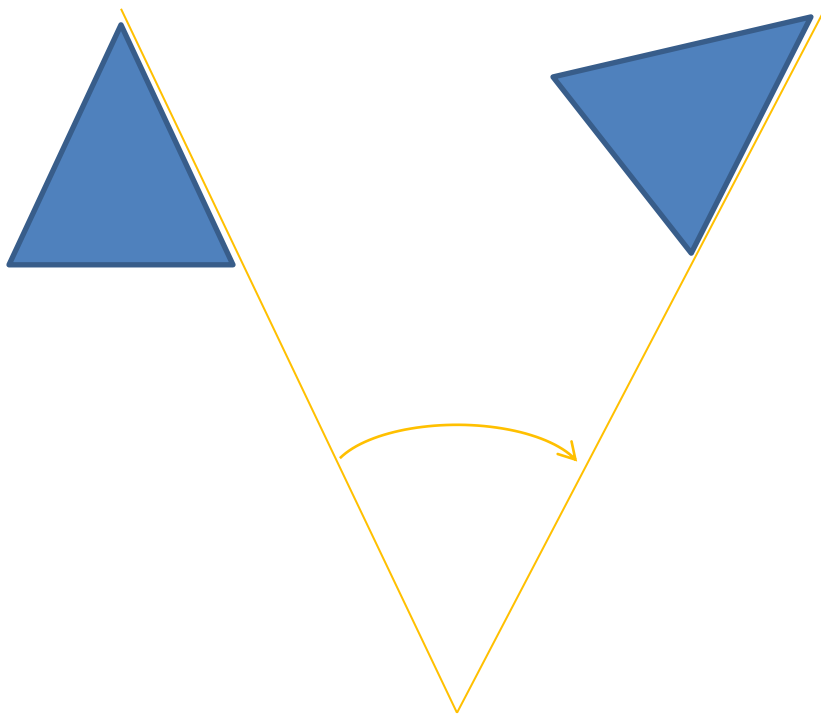
Любое вращение в трехмерном пространстве может быть представлено как композиция поворотов вокруг трех ортогональных осей.

Недостатки: 1) громоздкость вычислений
2) неустойчивость вычислений (ошибки округления)

Задание поворота в пространстве. Пример

Матрица поворота:

$$M_z(\alpha) = \begin{pmatrix} \cos \alpha & -\sin \alpha & 0 \\ \sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

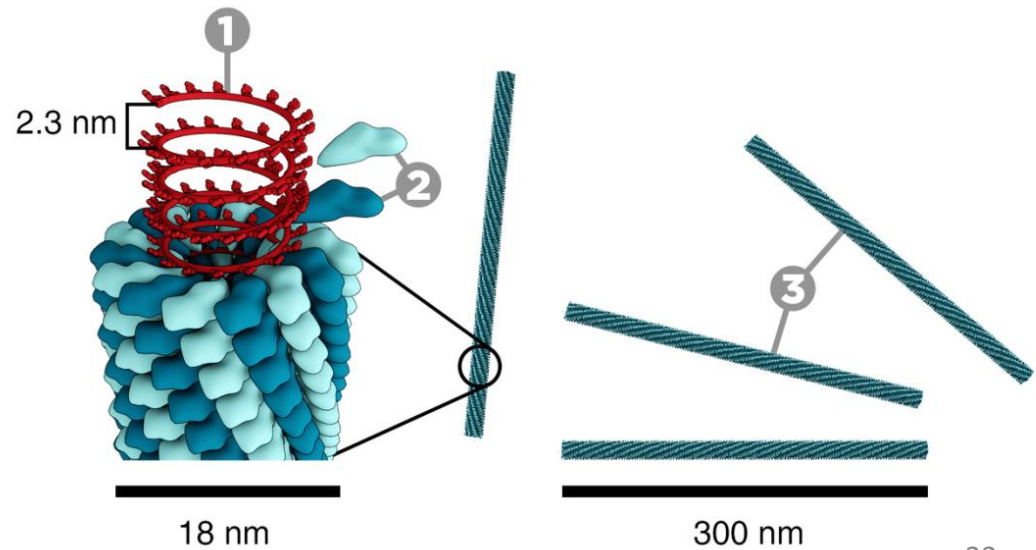
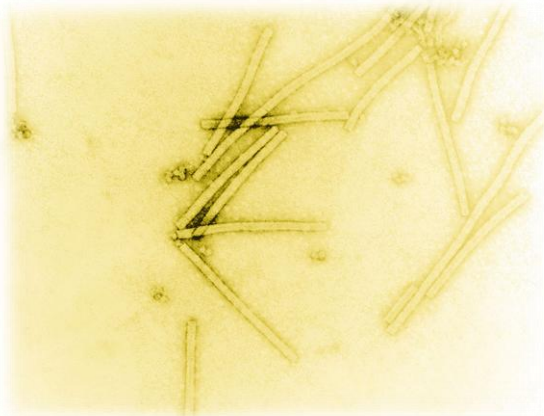
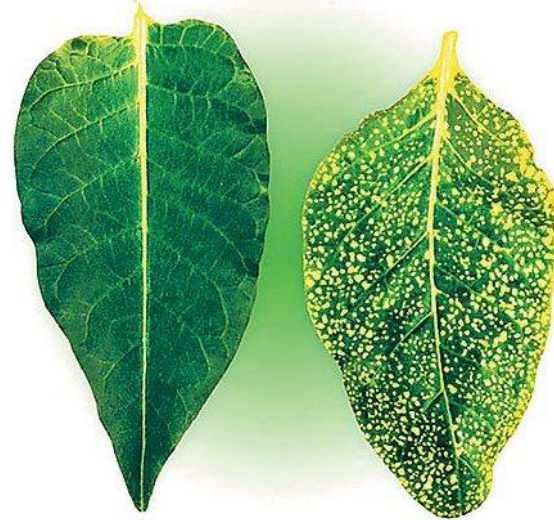


$$\begin{pmatrix} 0 \\ \sqrt{2} \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Вирус табачной мозаики



Д.И. Ивановский
(1864 — 1920)



Вирус табачной мозаики

REMARK 350 COORDINATES FOR A COMPLETE MULTIMER REPRESENTING THE KNOWN
 REMARK 350 BIOLOGICALLY SIGNIFICANT OLIGOMERIZATION STATE OF THE
 REMARK 350 MOLECULE CAN BE GENERATED BY APPLYING BIOMT TRANSFORMATIONS
 REMARK 350 GIVEN BELOW. BOTH NON-CRYSTALLOGRAPHIC AND
 REMARK 350 CRYSTALLOGRAPHIC OPERATIONS ARE GIVEN.

REMARK 350

REMARK 350 BIOMOLECULE: 1

REMARK 350 APPLY THE FOLLOWING TO CHAINS: A, R

REMARK 350 BIOMT1 1 1.000000 0.000000 0.000000

REMARK 350 BIOMT2 1 0.000000 1.000000 -0.000000

REMARK 350 BIOMT3 1 0.000000 -0.000000 1.000000

REMARK 350 BIOMT1 2 0.923478 -0.383651 0.000000

REMARK 350 BIOMT2 2 0.383651 0.923478 -0.000000

REMARK 350 BIOMT3 2 0.000000 -0.000000 1.000000

...

REMARK 350 BIOMT1 49 0.926988 0.375092 0.000000

REMARK 350 BIOMT2 49 -0.375092 0.926988 -0.000000

REMARK 350 BIOMT3 49 0.000000 -0.000000 1.000000

матрица поворота

вектор переноса

0.00000

0.00000

0.00000

-0.00000

0.00000

-67.58400

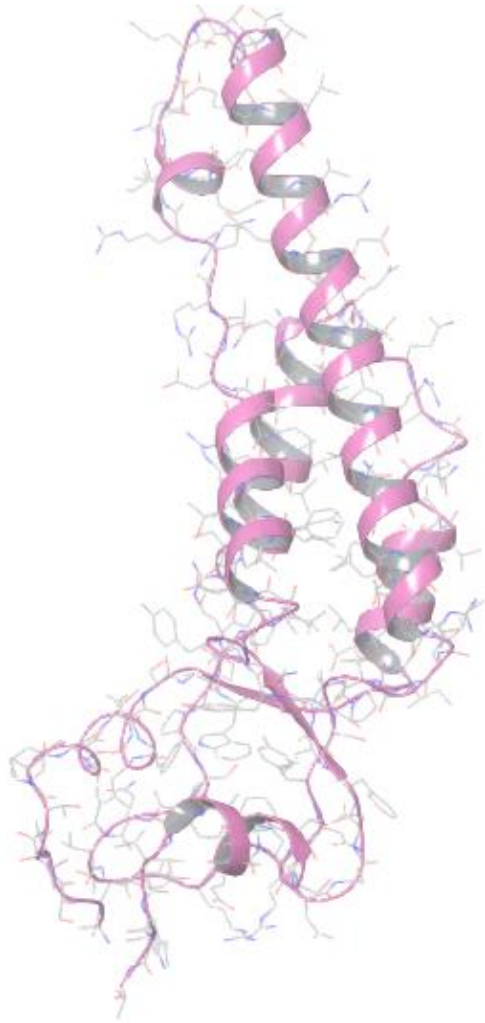
-0.00000

0.00000

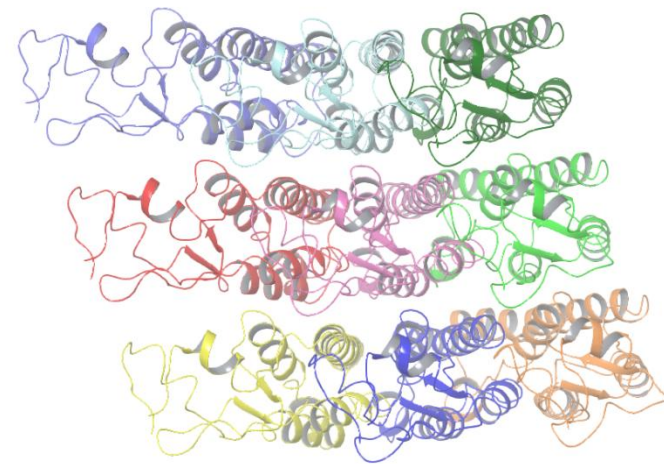
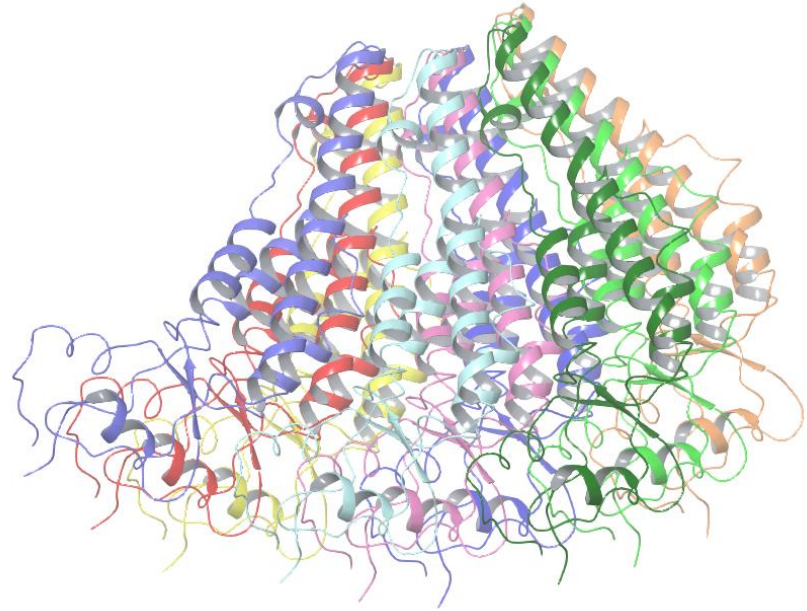
-1.40800

$$\begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} = \begin{pmatrix} BIOMT1_{2x} & BIOMT1_{2y} & BIOMT1_{2z} \\ BIOMT2_{2x} & BIOMT2_{2y} & BIOMT2_{2z} \\ BIOMT3_{2x} & BIOMT3_{2y} & BIOMT3_{2z} \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} + \begin{pmatrix} BIOMT1_{2s} \\ BIOMT2_{2s} \\ BIOMT3_{2s} \end{pmatrix}$$

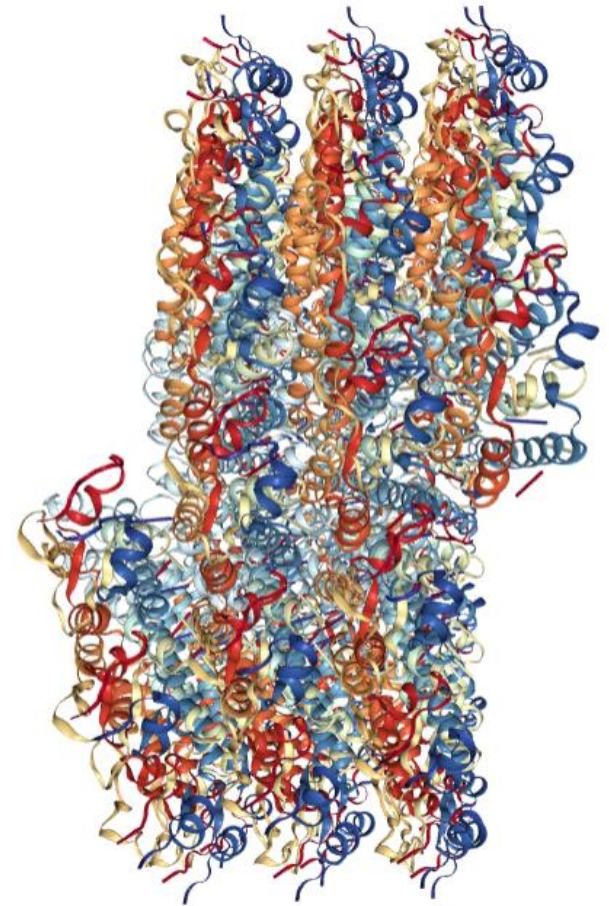
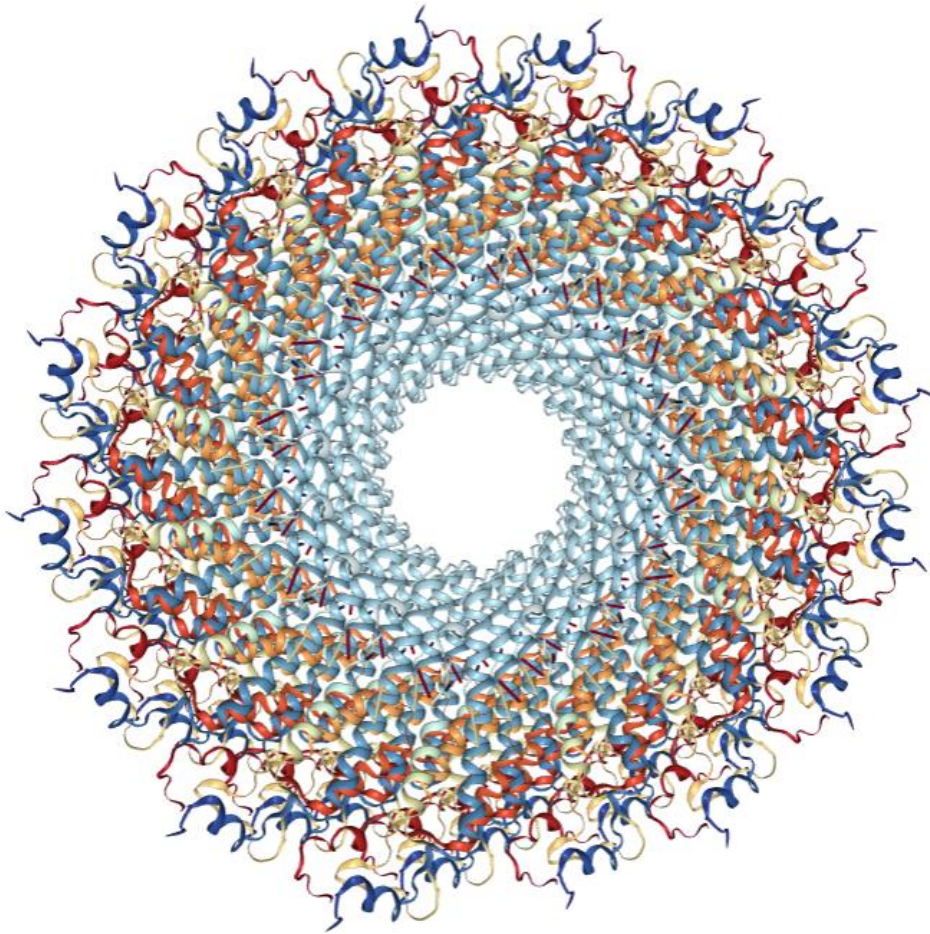
Вирус табачной мозаики



(pdb-код 4udv)



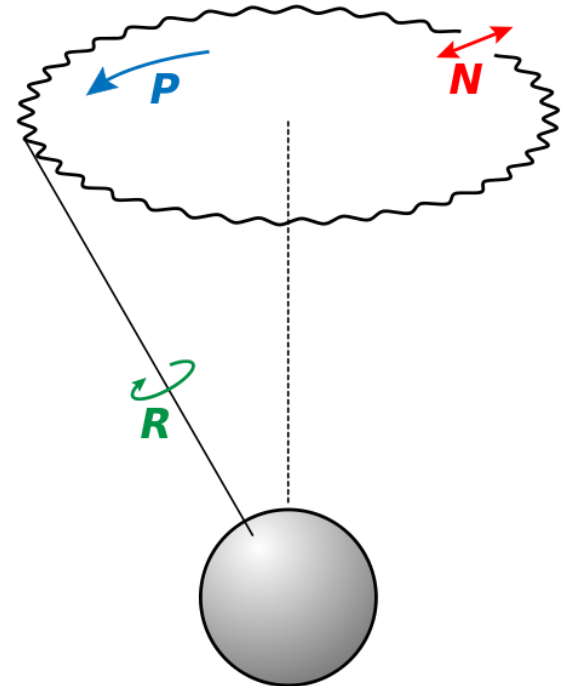
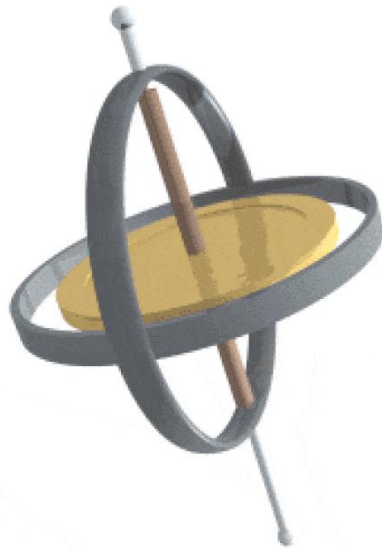
Вирус табачной мозаики



(pdb-код 4udv)

Задание поворота в пространстве

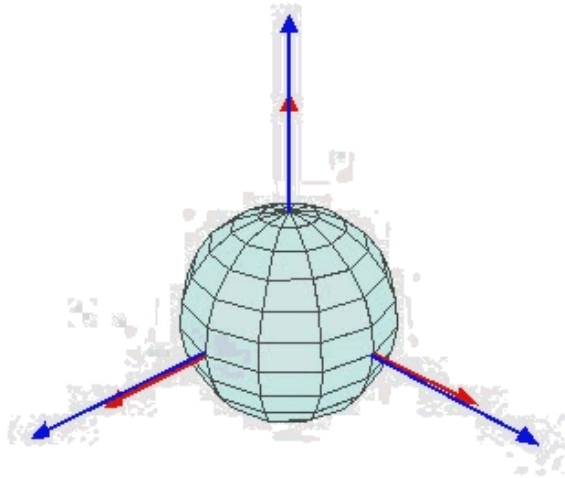
Гироскоп: прецессия, нутация и собственное вращение



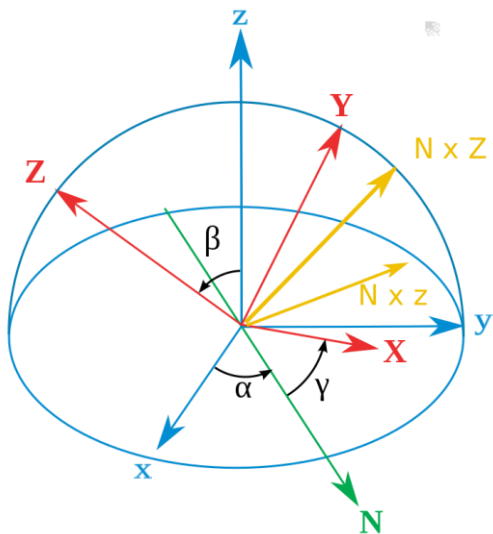
<https://www.youtube.com/watch?v=kgcM7MmBj40>

Задание поворота в пространстве

Углы Эйлера (1776):



Леонард Эйлер
(1707 – 1783)



x, y, z – исходные оси

X, Y, Z – конечные оси

N – линия узлов

Задание поворота в пространстве

Углы Эйлера (1776):

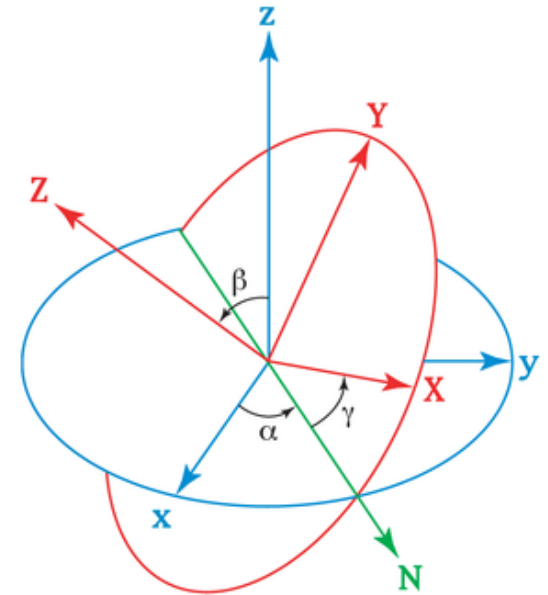
x, y, z – исходные оси

X, Y, Z – конечные оси

Угол α между осью x и линией узлов — угол прецессии (поворот вокруг оси z).

Угол β между осями z и z' — угол нутации (поворот вокруг оси $x' =$ линии узлов).

Угол γ между осью x' и линией узлов — угол собственного вращения (поворот вокруг оси $z'' = Z$).



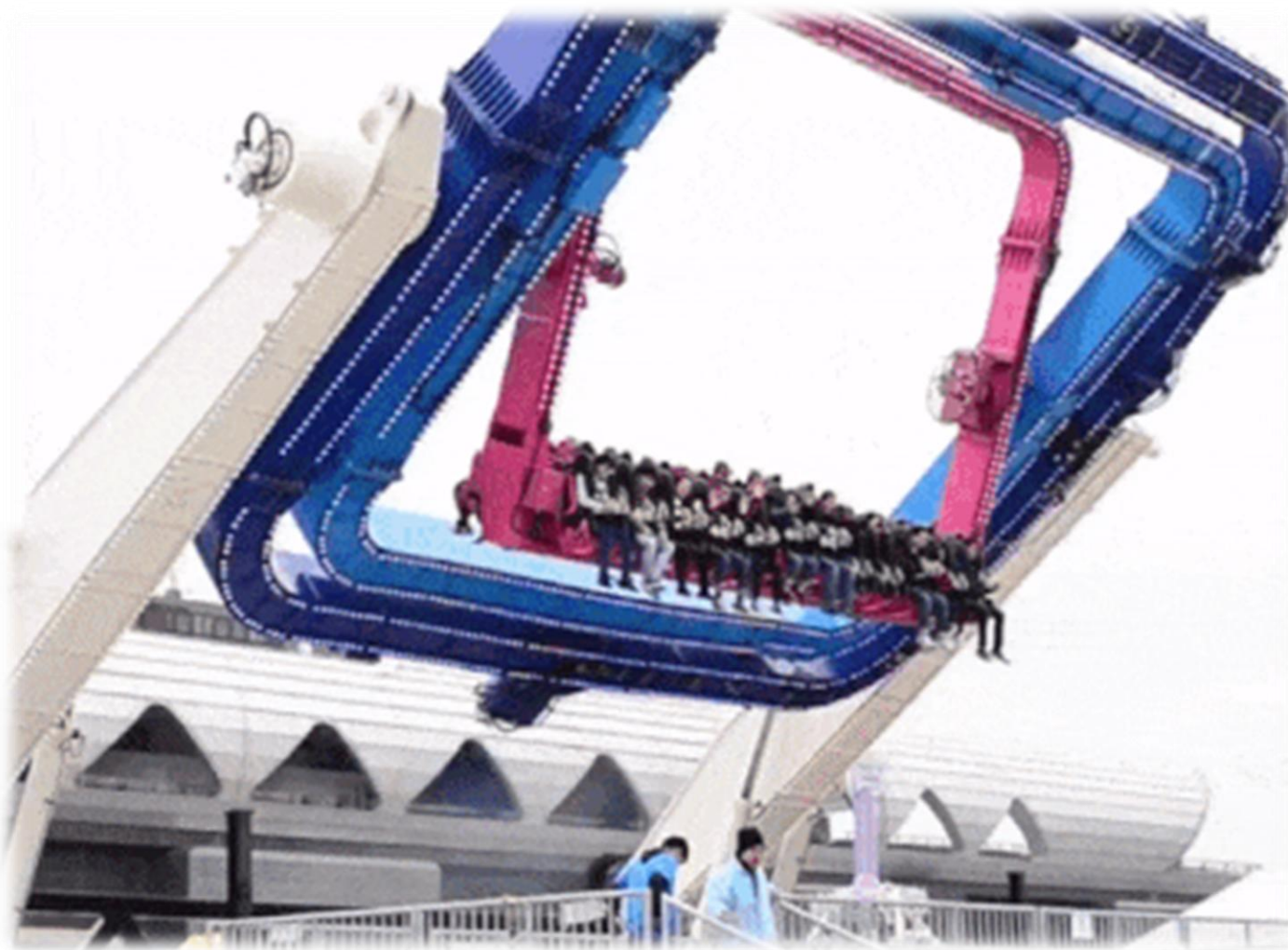
Такие повороты **некоммутативны** и конечное положение системы зависит от порядка, в котором совершаются повороты.

Недостатки: сложность вычислений при комбинации вращений, «складывание рамок карданова подвеса» ($\beta=0$).

Используются в ЭМ, ЯМР, ЭПР.

Задание поворота в пространстве

Карданов подвес



Задание поворота в пространстве

Кватернионы (1843): система гиперкомплексных чисел, образующая векторное пространство размерностью четыре над полем вещественных чисел.

$$q = a + bi + cj + dk$$

$$\bar{q} = a - bi - cj - dk$$

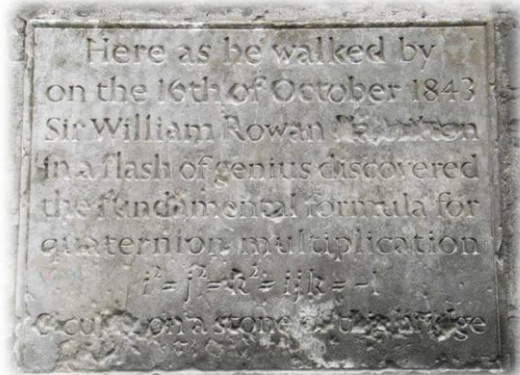
$$|q|^2 = a^2 + b^2 + c^2 + d^2$$

$$q^{-1} = \frac{\bar{q}}{|q|^2}$$

×	1	i	j	k
1	1	<i>i</i>	<i>j</i>	<i>k</i>
i	<i>i</i>	-1	<i>k</i>	- <i>j</i>
j	<i>j</i>	- <i>k</i>	-1	<i>i</i>
k	<i>k</i>	<i>j</i>	- <i>i</i>	-1



Уильям Роуэн Гамильтон
(1805 — 1865)



Задание поворота в пространстве

Описание поворота через кватернион:

$$q = w + x\mathbf{i} + y\mathbf{j} + z\mathbf{k} = w + (x, y, z) = \cos(\alpha/2) + \mathbf{u} \sin(\alpha/2)$$

$$\vec{t} = q\vec{v}q^{-1} = \frac{q\vec{v}\bar{q}}{|q|^2}$$

-результат поворота вектора \mathbf{v}
вокруг вектора \mathbf{u} на угол α

$$T = QVQ^{-1} = \frac{QVQ^T}{\sqrt{\det(Q)}}$$

- он же в матричном виде

Преимущества: а) интуитивно понятный смысл и отсутствие «складывания рамок»
б) возможность использования векторной алгебры
в) легкость выполнения интерполяции

$$q(t) = (1-t)q_1 + tq_2, \quad 0 < t < 1$$

Задание поворота в пространстве

Пример: поворот вектора \mathbf{v} $(1, 0, 0)$ вокруг вектора \mathbf{u} $(1, 1, 1)$ на угол 120° .

Кватернион поворота

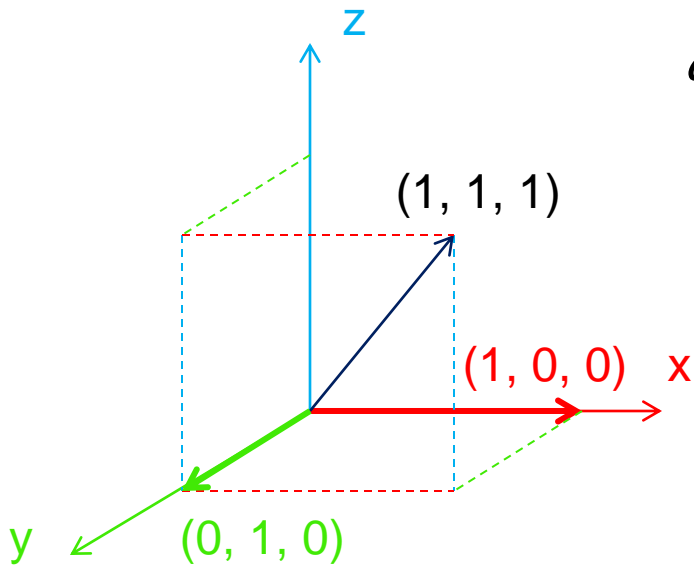
$$q = \cos\left(\frac{\pi}{3}\right) + \sin\left(\frac{\pi}{3}\right)\vec{i} + \sin\left(\frac{\pi}{3}\right)\vec{j} + \sin\left(\frac{\pi}{3}\right)\vec{k}$$

Вектор \mathbf{u} должен быть единичным, поэтому переобозначим

$$\mathbf{u} = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)$$

и для кватерниона получаем

$$q = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right) = \frac{1}{2}(1, 1, 1, 1)$$



Задание поворота в пространстве

In quaternion form

$$q = \frac{1}{2}(1, 1, 1, 1) \quad q^{-1} = \frac{1}{2}(1, -1, -1, -1) \quad \vec{v} = (0, 1, 0, 0)$$

Note that $|q|^2 = 1$

$$\vec{t} = q\vec{v}q = \frac{1}{4}(1, 1, 1, 1)(0, 1, 0, 0)(1, -1, -1, -1) = \frac{1}{4}(1 + \vec{i} + \vec{j} + \vec{k})(\vec{i})(1 - \vec{i} - \vec{j} - \vec{k}) =$$

$$= \frac{1}{4}(1 + \vec{i} + \vec{j} + \vec{k})(\vec{i} + 1 - \vec{k} + \vec{j}) =$$

$$= \frac{1}{4}((1 + \vec{i} + \vec{j} - \vec{k}) + \vec{i}(1 + \vec{i} + \vec{j} - \vec{k}) + \vec{j}(1 + \vec{i} + \vec{j} - \vec{k}) + \vec{k}(1 + \vec{i} + \vec{j} - \vec{k})) =$$

$$= \frac{1}{4}(1 + \vec{i} + \vec{j} - \vec{k} + \vec{i} - 1 + \vec{k} + \vec{j} + \vec{j} - \vec{k} - 1 - \vec{i} + \vec{k} + \vec{j} - \vec{i} + 1) =$$

$$= \frac{1}{4}(0 + 0\vec{i} + 4\vec{j} + 0\vec{k}) = (0, 0, 1, 0)$$

that corresponds to a vector $\vec{t} = (0, 1, 0)$

\times	1	i	j	k
1	1	i	j	k
i	i	-1	k	$-j$
j	j	$-k$	-1	i
k	k	j	$-i$	1

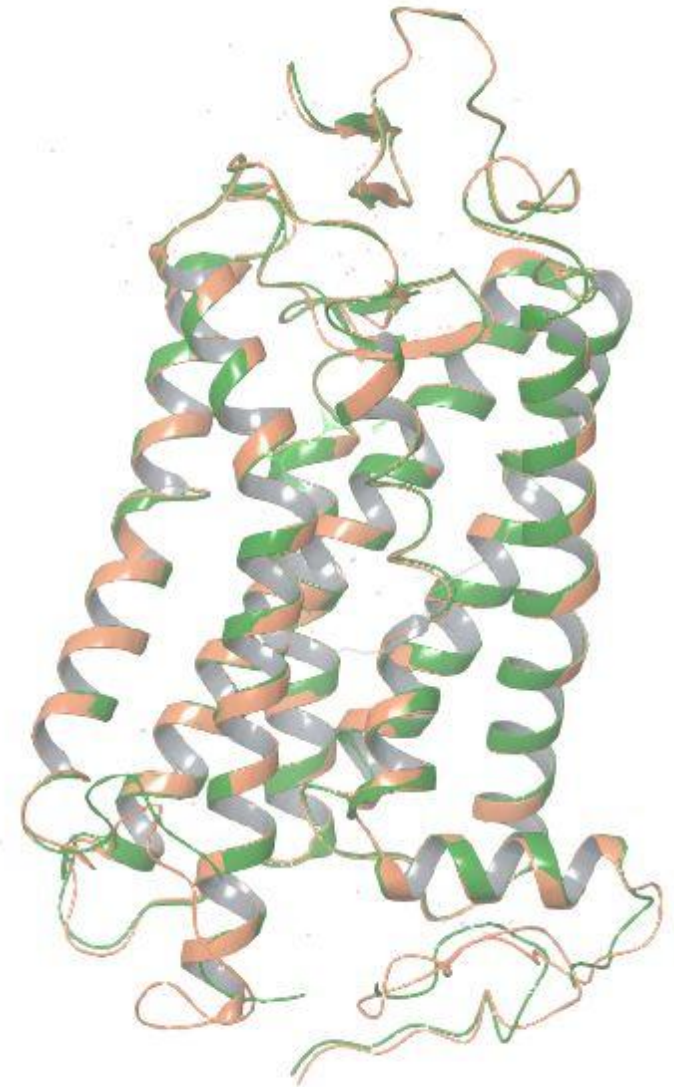
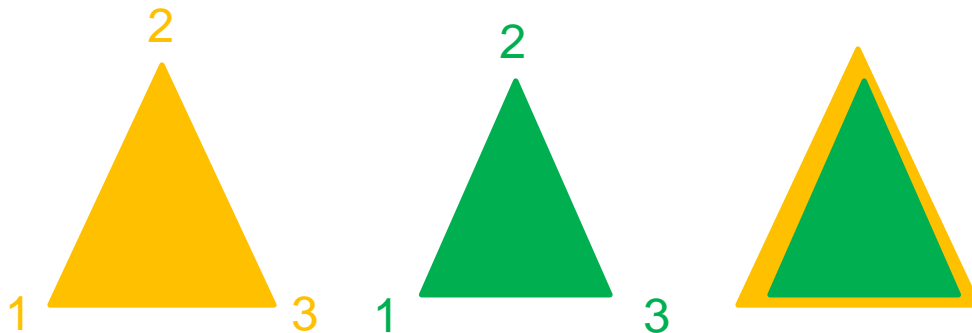
Структурные выравнивания

Варианты задачи:

1) Взаимное соответствие атомов известно заранее:

$$p_i \leftrightarrow q_i, i = 1, \dots, N$$

(различные структуры или модели одной и той же последовательности или близко гомологичных последовательностей одинаковой длины);



Совмещение структур зрительного родопсина (pdb-коды 1u19 и 1f88)

Структурные выравнивания

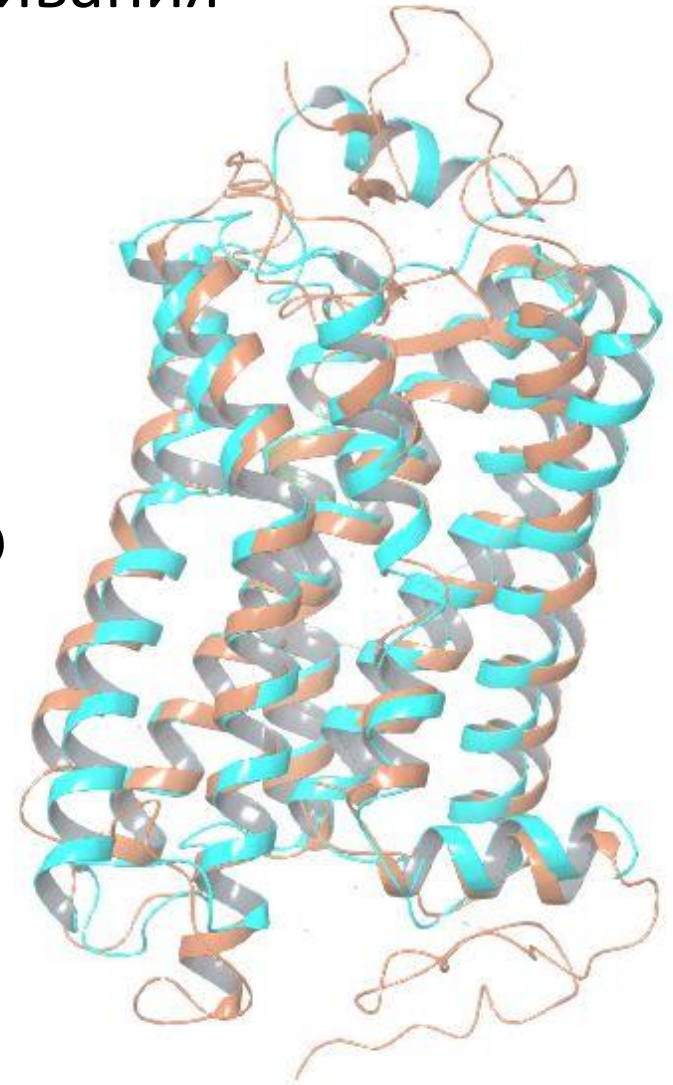
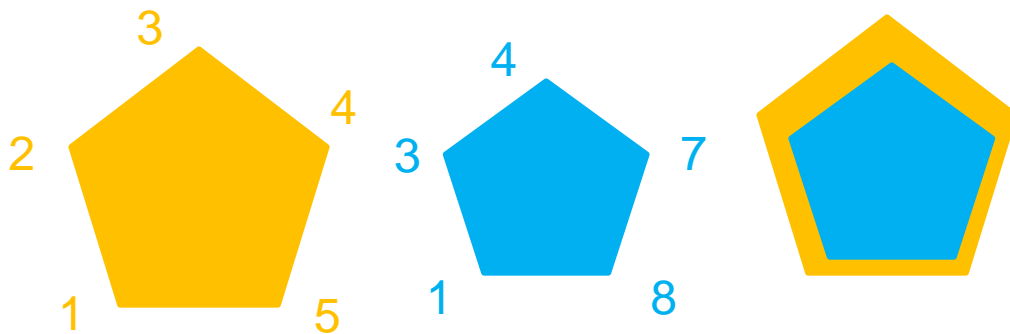
Варианты задачи:

2) Соответствие неизвестно, но известен порядок следования:

$$P_{i(k)} \leftrightarrow q_{j(k)}, i = 1, \dots, K \leq N, M,$$

$$\text{причем } k_1 > k_2 \Leftrightarrow i(k_1) > i(k_2), j(k_1) > j(k_2)$$

(структуры последовательностей разной длины, имеющие вставки в выравнивании последовательностей);



Совмещение структур зрительного родопсина (pdb-код 1u19) и β 2-адренэргического рецептора (pdb-код 2rh1)

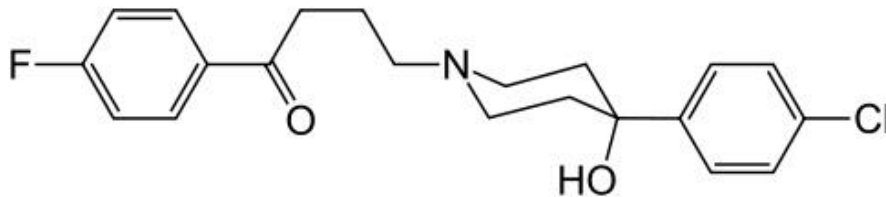
Структурные выравнивания

Варианты задачи:

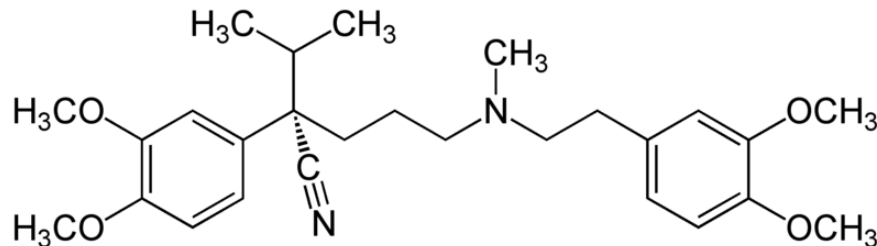
3) **Неизвестно ни соответствие, ни порядок следования:**

$$P_{i(k)} \leftrightarrow q_{j(k)}, i = 1, \dots, K \leq N, M$$

(фармакофоры)



Галоперидол
(блокатор дофаминовых
рецепторов)



Верапамил
(блокатор потенциал-зависимых
кальциевых каналов)

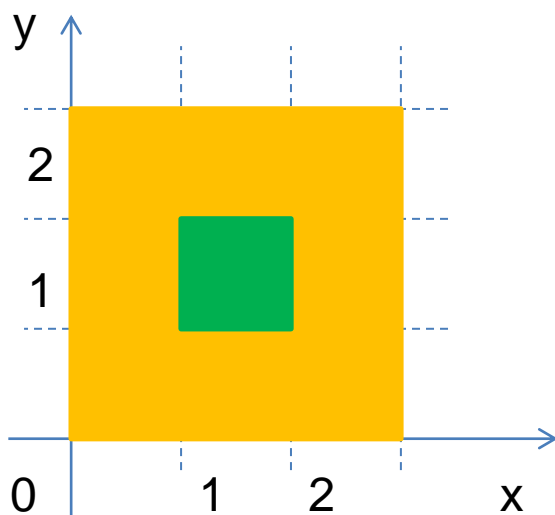
Оба проявляют активность в отношении калиевых каналов

Структурные выравнивания

Мера близости (для структур с заранее известным взаимным соответствием атомов) - **среднеквадратичное отклонение атомов (СКО)**:

$$СКО = \sqrt{\frac{\sum_i \left((x_i - x_i^0)^2 + (y_i - y_i^0)^2 + (z_i - z_i^0)^2 \right)}{n}}$$

(обычно рассчитывается для Са-атомов, но в случае очень близких гомологов – для всех тяжелых атомов)



$$СКО = \sqrt{2}$$

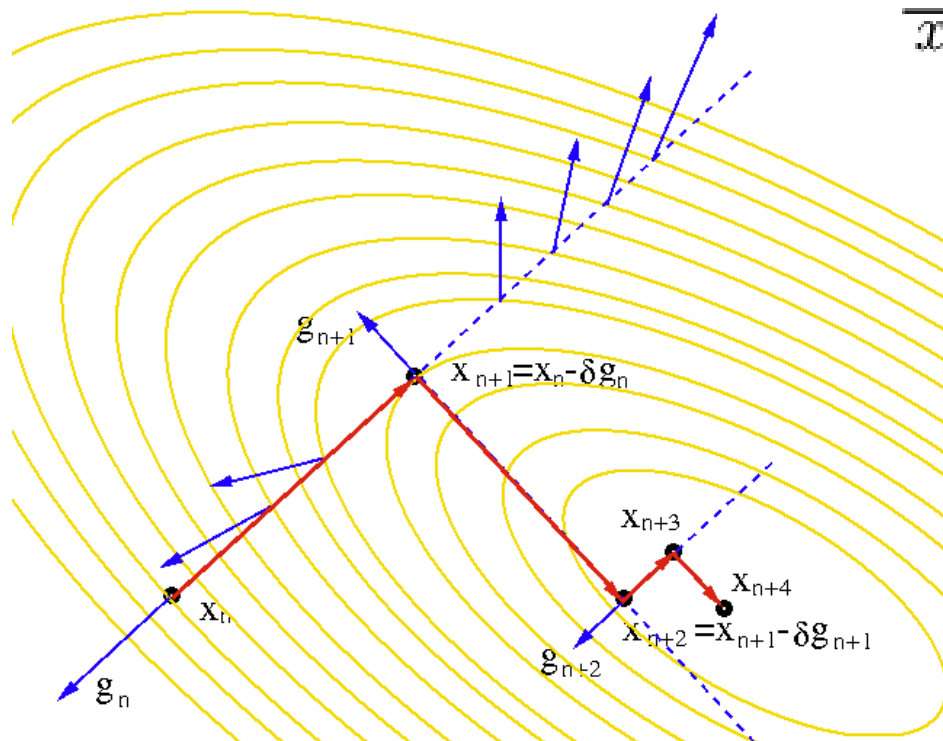
$$СКО = 2$$

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

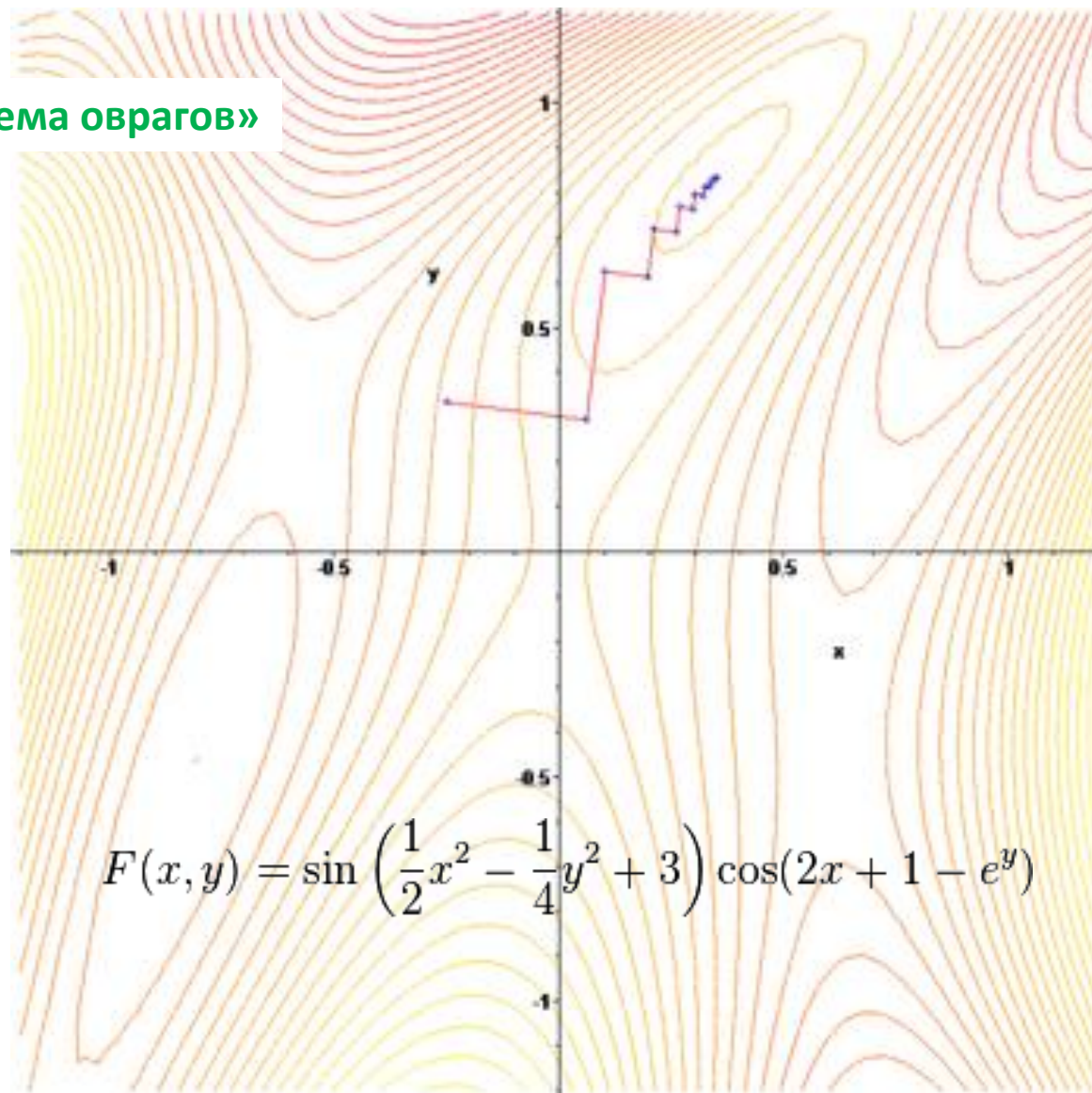
- определяемой на каждой

итерации заново

(метод наискорейшего спуска,
steepest descent)

Методы минимизации. Градиентный спуск

«Проблема оврагов»

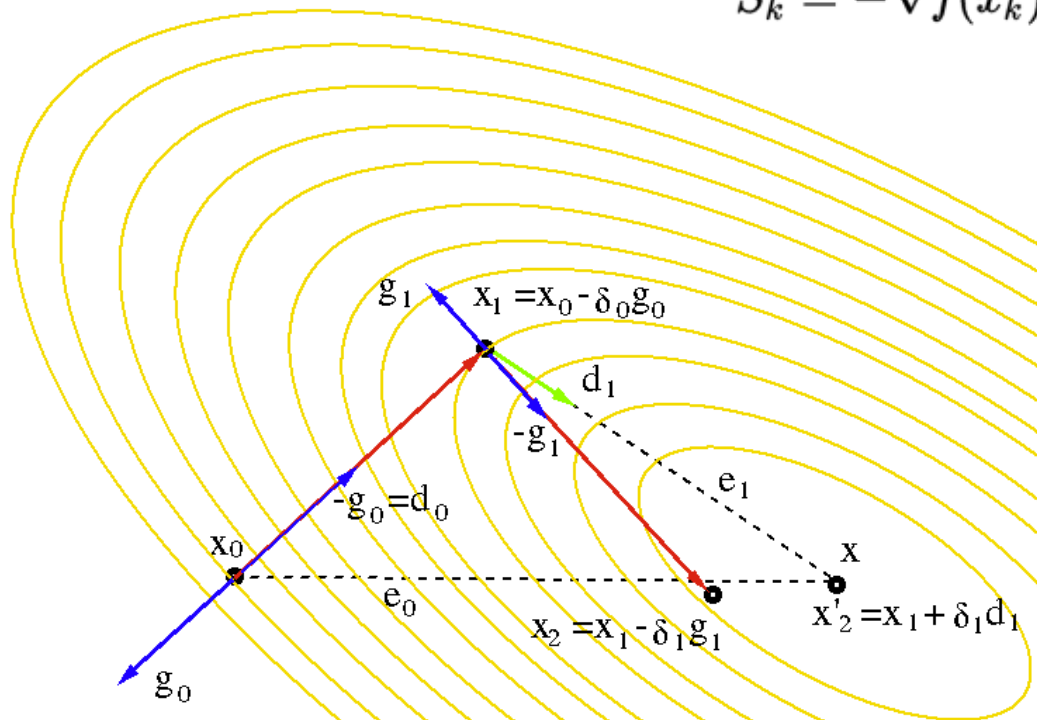


Методы минимизации. Метод сопряженных градиентов

(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.

Продолжение следует...