

# Лекция 8

## Структурная биоинформатика

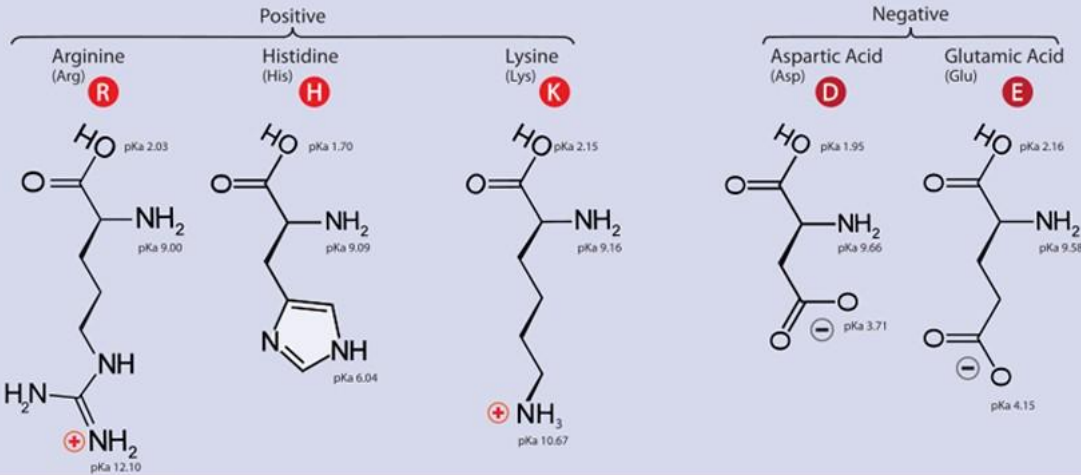
апрель 2018

# Структура аминокислот

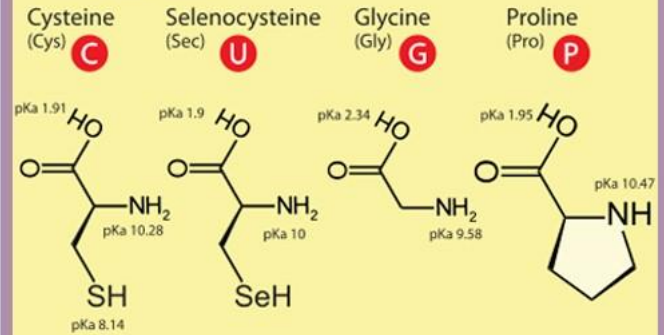
## Twenty-One Amino Acids

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

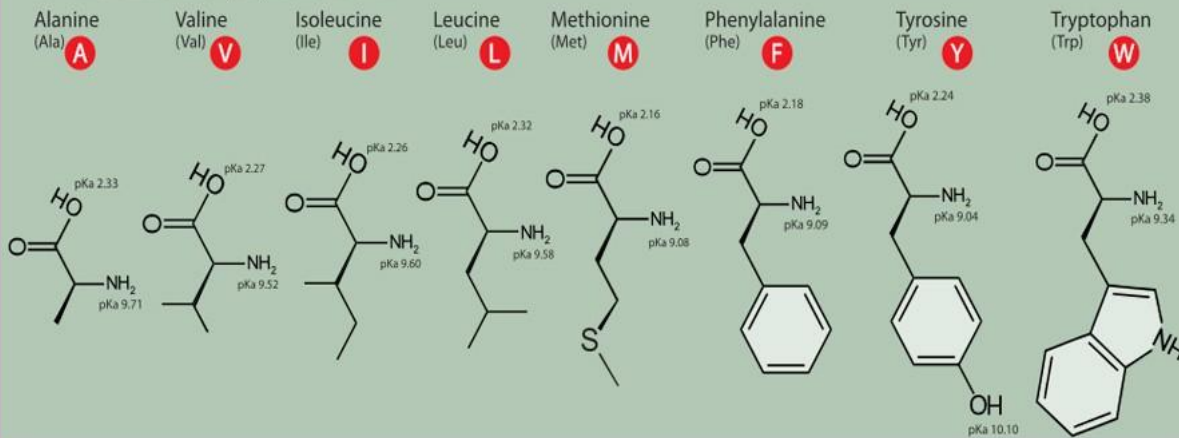
### A. Amino Acids with Electrically Charged Side Chains



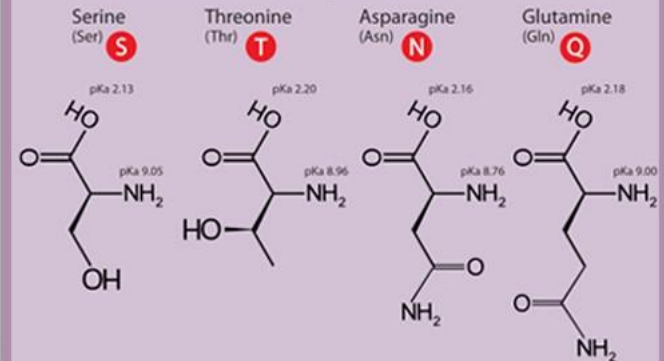
### C. Special Cases



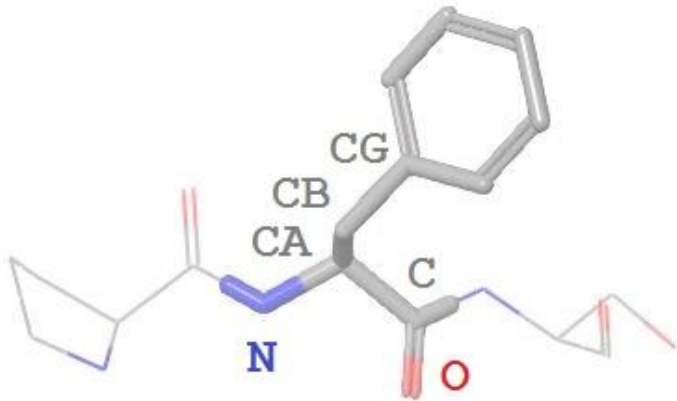
### D. Amino Acids with Hydrophobic Side Chain



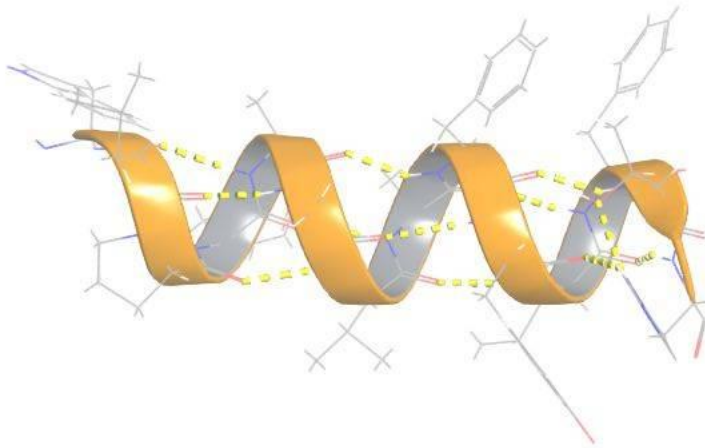
### B. Amino Acids with Polar Uncharged Side Chains



# Структурная организация белков



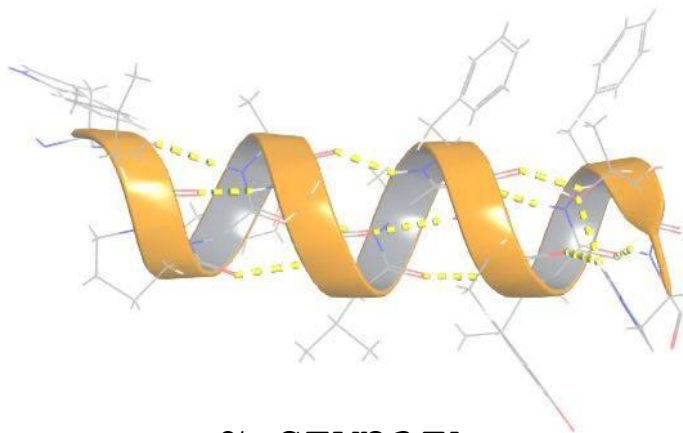
...FYV**PFS**NKT...



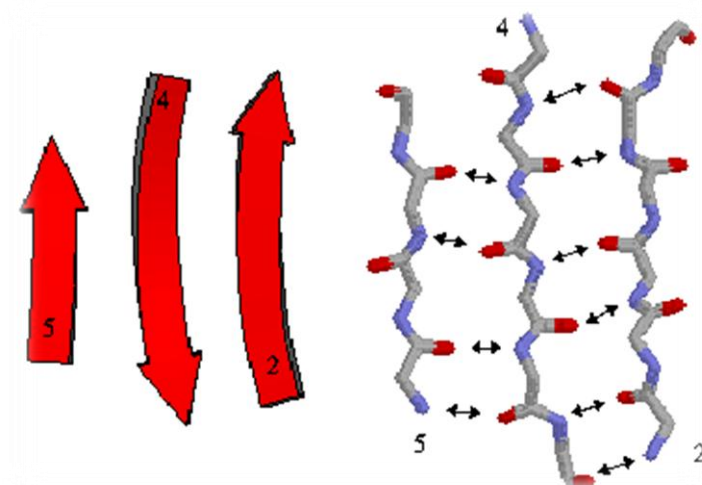
...WLPY**AGVAFYIF**TH...



# Типы вторичной структуры белков

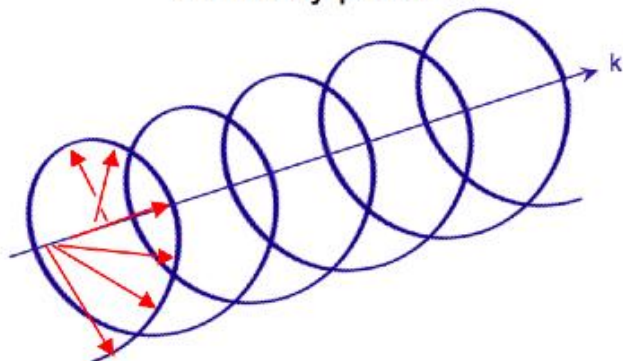


$\alpha$ -спираль

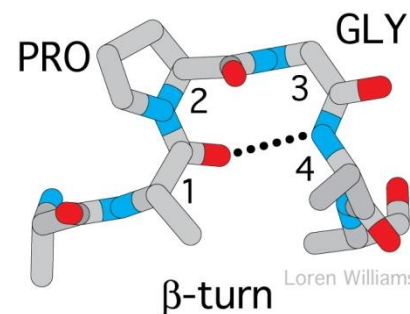


$\beta$ -лист, состоящий из  $\beta$ -тяжей

Circularly polarized



Экспериментальный метод изучения вторичной структуры – спектроскопия кругового дихроизма



$\beta$ -turn  
поворот



# Структурные базы данных

**RCSB PDB**  
PROTEIN DATA BANK

A MEMBER OF THE **PDB**

An Information Portal to Biological Macromolecular Structures

As of Tuesday Oct 12, 2010 at 5 PM PDT there are 68562 Structures | PDB Statistics

Contact Us | Print | PDB ID or Text | Search | Advanced Search

## A Resource for Studying Biological Macromolecules

The PDB archive contains information about experimentally-determined structures of proteins, nucleic acids, and complex assemblies. As a member of the **wwPDB**, the RCSB PDB curates and annotates PDB data according to agreed upon standards.

The RCSB PDB also provides a variety of tools and resources. Users can perform simple and advanced searches based on annotations relating to sequence, structure and function. These molecules are visualized, downloaded, and analyzed by users who range from students to specialized scientists.

**Hide Welcome Message**

### Featured Molecules (MoM Category View / Previous) Hide

Features: MoM | PSI

**Structural View of Biology** Protein Synthesis

**Molecule of the Month: Riboswitches**  
Why use two or more molecules when one

### Customize This Page

#### New Features Hide

**Improved Domain Display on Sequence Tab**  
Latest features released:  
Website Release Archive:

#### RCSB PDB News Hide

Weekly | Quarterly | Yearly

2010-10-12  
**Categorized View of Molecule of the Month**

Explore new Molecule of the Month feature categories.

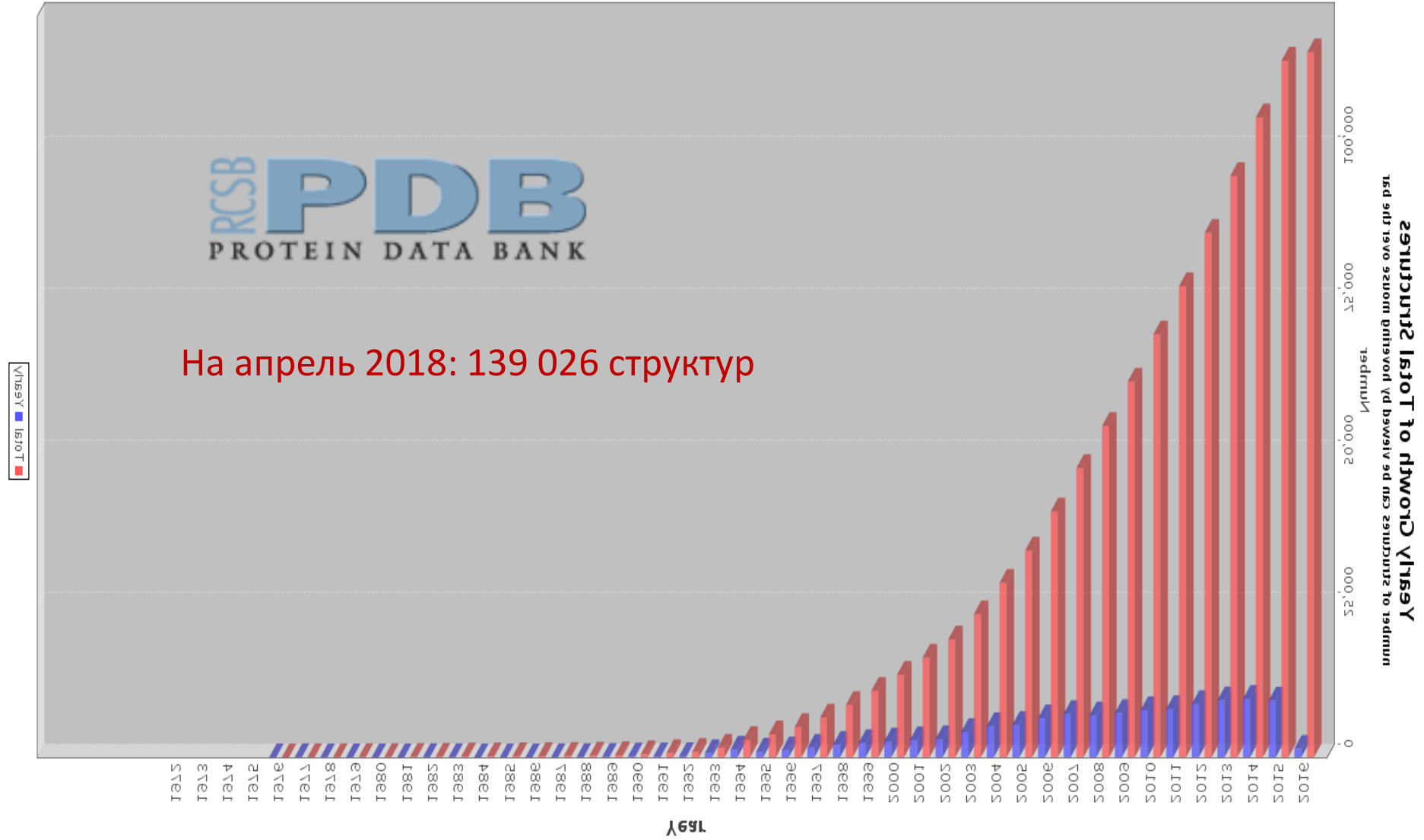
- Latest Website Release
- Poster Prize Awarded at ECM
- Analyze small molecule interactions in the PDB with

PDB – содержит информацию об экспериментально определенных структурах белков, нуклеиновых кислот и различных комплексов.

# Структурные базы данных



На апрель 2018: 139 026 структур



# Структурные базы данных

## POTASSIUM CHANNEL KCSA IN COMPLEX WITH TETRABUTYLAMMONIUM IN HIGH K

MyPDB Hide

Login to your Account  
Register a New Account

Home Hide

News & Publications  
Usage/Reference Policies  
Deposition Policies  
Website FAQ  
Deposition FAQ  
Contact Us  
About Us  
Careers  
External Links  
Sitemap  
New Website Features

Deposition Hide

All Deposit Services  
Electron Microscopy  
X-ray | NMR  
Validation Server  
BioSync Beamline  
Related Tools

Search Hide

Advanced Search  
Latest Release  
New Structure Papers  
Sequence Search  
Chemical Components  
Unreleased Entries  
Browse Database  
Help

Summary Sequence Derived Data Seq. Similarity 3D Similarity Literature Biol. & Chem. Methods Geometry Links

### POTASSIUM CHANNEL KCSA IN COMPLEX WITH TETRABUTYLAMMONIUM IN HIGH K

2JK5

Display Files ▾  
Download Files ▾  
Print this Page  
Share this Page ▸

DOI: 10.2210/pdb2jk5/pdb

Primary Citation

**Structure of a Potassium Channel in Complex with Symmetrical Quaternary Ammonium Compounds Reveals a Novel Hydrophobic Binding Site**

Lenaus, M.J., Focia, P.J., Wagner, T., Gross, A.

Journal: To be Published

Not in PubMed

Molecular Description Hide

Classification: Immune System/metal Transport  
Structure Weight: 61280.07

Molecule:	ANTIBODY FAB FRAGMENT LIGHT CHAIN	Length:	219
Polymer:	1 Type: polypeptide(L)		
Chains:	A		
Molecule:	ANTIBODY FAB FRAGMENT HEAVY CHAIN	Length:	212
Polymer:	2 Type: polypeptide(L)		
Chains:	B		
Molecule:	VOLTAGE-GATED POTASSIUM CHANNEL	Length:	124
Polymer:	3 Type: polypeptide(L)		
Chains:	C		

Biological Assembly ?



Biological Assembly Image for 2JK5

More Images...

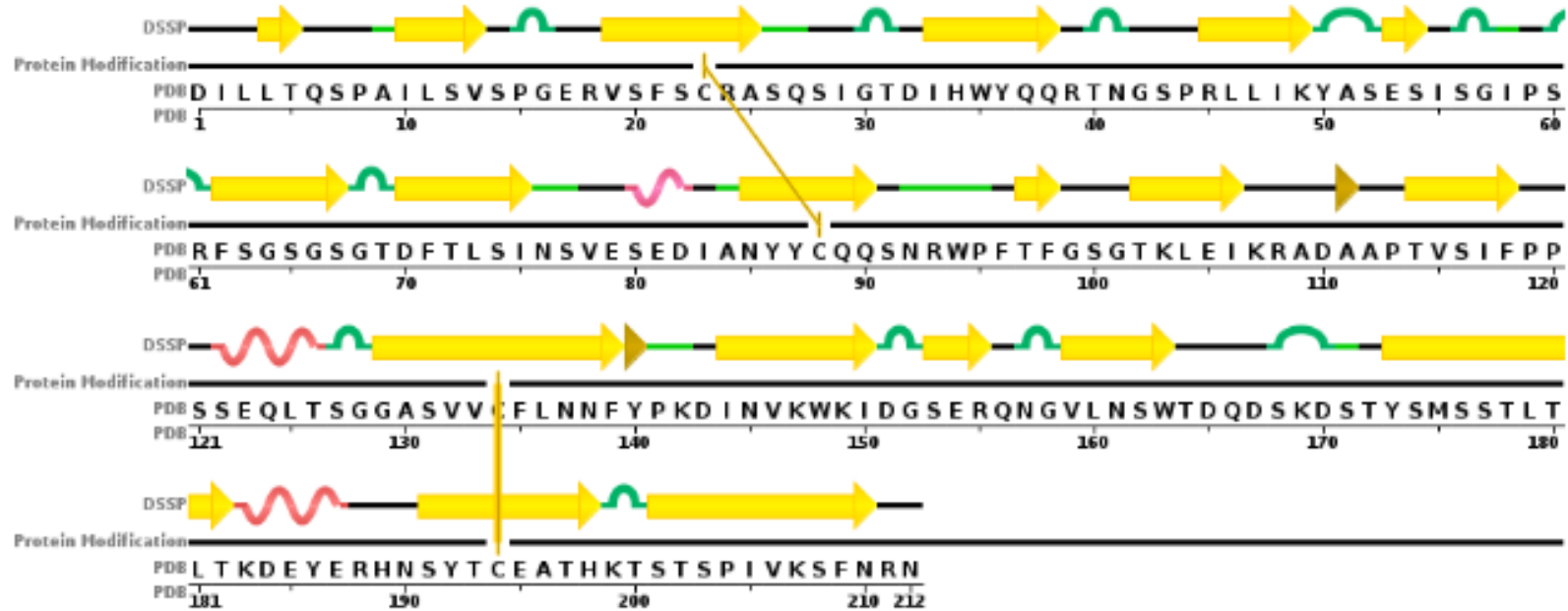
View in Jmol SimpleViewer  
Other Viewers ▾ Protein Workshop

Biological assembly generated by PISA (software)

# Структурные базы данных

## POTASSIUM CHANNEL KCSA IN COMPLEX WITH TETRABUTYLAMMONIUM IN HIGH K

### Sequence Chain View



#### Protein Modification Legend

—+— L-cystine

#### DSSP Legend

— empty: no secondary structure assigned

➔ B: beta bridge

— S: bend

⤿ T: turn

➔ E: beta strand

⤿ G: 3/10-helix

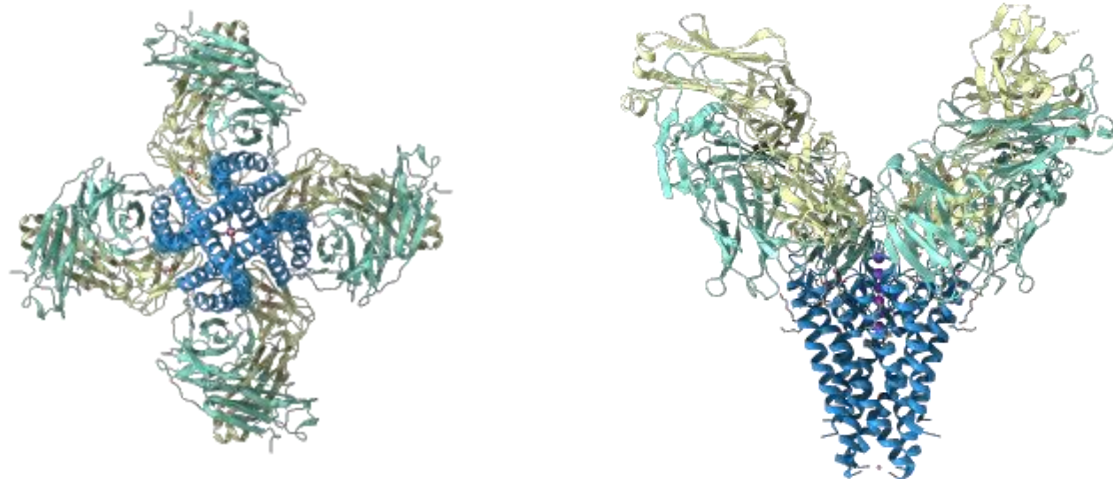
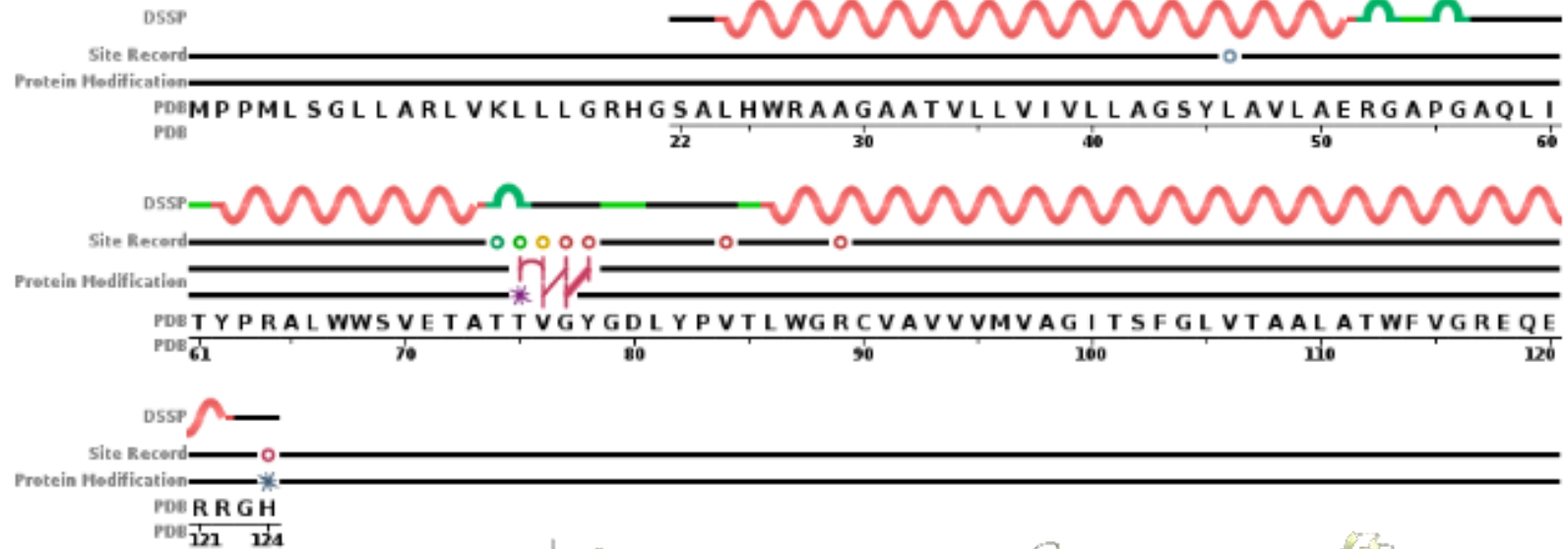
⤿ H: alpha helix



# Структурные базы данных

POTASSIUM CHANNEL KCSA IN COMPLEX WITH TETRABUTYLAMMONIUM IN HIGH K

## Sequence Chain View



# Структурные файлы. \*pdb

```
HEADER      IMMUNE SYSTEM/METAL TRANSPORT          15-AUG-08   2JK5
TITLE      POTASSIUM CHANNEL KCSA IN COMPLEX WITH TETRABUTYLAMMONIUM
TITLE      2 IN HIGH K
COMPND     MOL_ID: 1;
COMPND     2 MOLECULE: ANTIBODY FAB FRAGMENT LIGHT CHAIN;
COMPND     3 CHAIN: A;
COMPND     4 ENGINEERED: YES;
COMPND     5 MOL_ID: 2;
...
KEYWDS     IMMUNE SYSTEM METAL TRANSPORT COMPLEX, QUATERNARY AMMONIUM,
...
EXPDTA     X-RAY DIFFRACTION
AUTHOR     M.J.LENAEUS,P.J.FOCIA,T.WAGNER,A.GROSS
REVDAT     1 17-NOV-09 2JK5 0
JRNL       AUTH  M.J.LENAEUS,P.J.FOCIA,T.WAGNER,A.GROSS
JRNL       TITL  STRUCTURE OF A POTASSIUM CHANNEL IN COMPLEX WITH
JRNL       TITL 2 SYMMETRICAL QUATERNARY AMMONIUM COMPOUNDS REVEALS
JRNL       TITL 3 A NOVEL HYDROPHOBIC BINDING SITE
JRNL       REF   TO BE PUBLISHED
JRNL       REFN
REMARK     2
REMARK     2 RESOLUTION.      2.4  ANGSTROMS.
REMARK     3
REMARK     3 REFINEMENT.
REMARK     3 PROGRAM      : REFMAC 5.5.0051
...
```

# Структурные файлы. \*.pdb

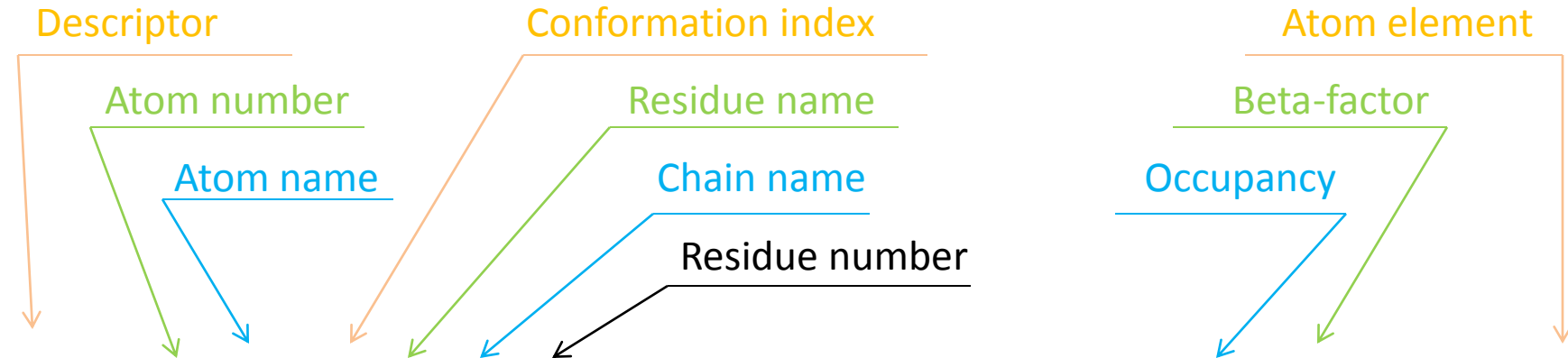
```
...
REMARK 465
REMARK 465 MISSING RESIDUES
REMARK 465 THE FOLLOWING RESIDUES WERE NOT LOCATED IN THE
REMARK 465 EXPERIMENT. (M=MODEL NUMBER; RES=RESIDUE NAME; C=CHAIN
REMARK 465 IDENTIFIER; SSSEQ=SEQUENCE NUMBER; I=INSERTION CODE.)
REMARK 465
REMARK 465      M RES C SSSEQI
REMARK 465      MET C      1
REMARK 465      PRO C      2
...
REMARK 470
REMARK 470 MISSING ATOM
REMARK 470 THE FOLLOWING RESIDUES HAVE MISSING ATOMS (M=MODEL NUMBER;
REMARK 470 RES=RESIDUE NAME; C=CHAIN IDENTIFIER; SSEQ=SEQUENCE NUMBER;
REMARK 470 I=INSERTION CODE):
REMARK 470      M RES CSSEQI  ATOMS
REMARK 470      LYS A  23    CG   CD   CE   NZ
REMARK 470      LYS A  63    CE   NZ
...
SEQRES   1 A  219  GLN VAL GLN LEU GLN GLN PRO GLY ALA GLU LEU VAL LYS
SEQRES   2 A  219  PRO GLY ALA SER VAL LYS LEU SER CYS LYS ALA SER GLY
...
HET      F09   C1131      10
...
HETNAM   F09  NONAN-1-OL
...

```

# Структурные файлы. \*.pdb

```
...
HELIX      1    1 THR A    87  SER A    91  5                                5
...
HELIX     10   10 THR C    85  GLY C   123  1                                39
SHEET      1   AA 4 LEU A    4  GLN A    5  0
...
SHEET      4   BE 4 SER B  201  ASN B  210 -1  O  SER B  201    N  HIS B  198
SSBOND     1  CYS A    22    CYS A    96                                1555  1555  2.07
...
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  OE1AGLN A    3    -28.187  26.374  16.332  0.50  43.76    O
ATOM       25  NE2AGLN 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
...
CONNECT  4140  4141
CONNECT  4141  4140  4142
...
MASTER      520    0    11    10    47    0    10    6  4420    3    101    44
END
```

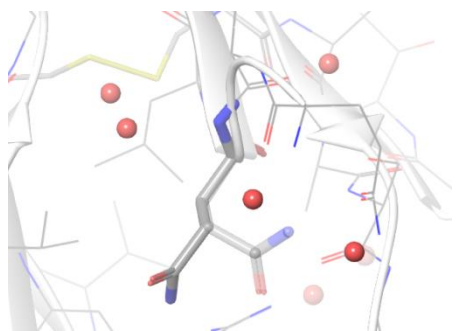
# Структурные файлы. \*.pdb



```

...
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  OE1AGLN A   3   -28.187  26.374  16.332  0.50  43.76  O
ATOM      25  NE2AGLN 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



# Задание поворота в пространстве

**Матрица поворота:**

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

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

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

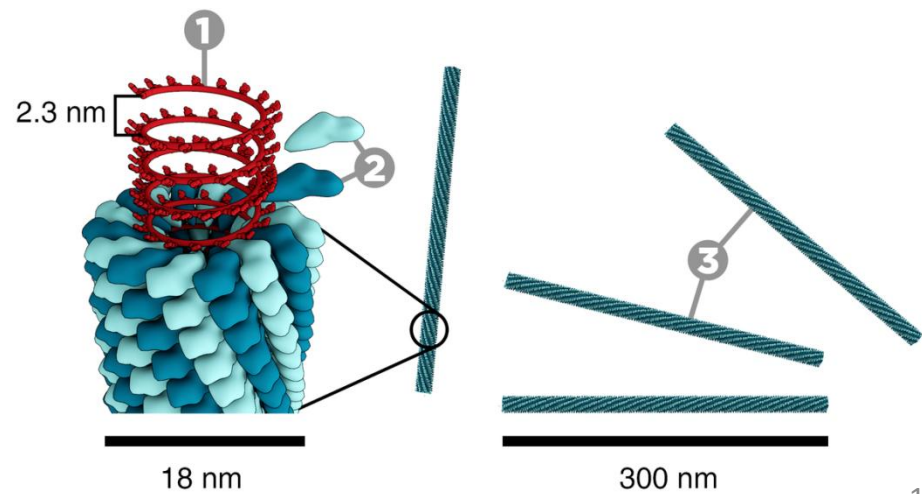
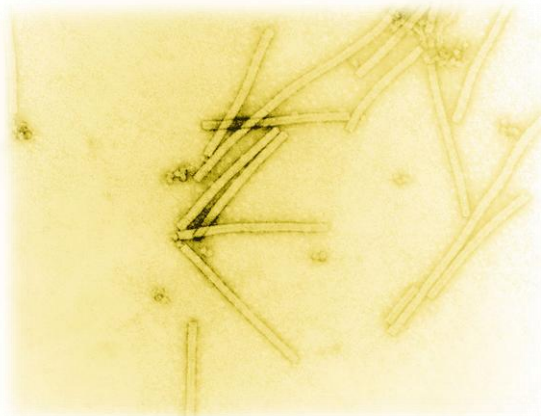
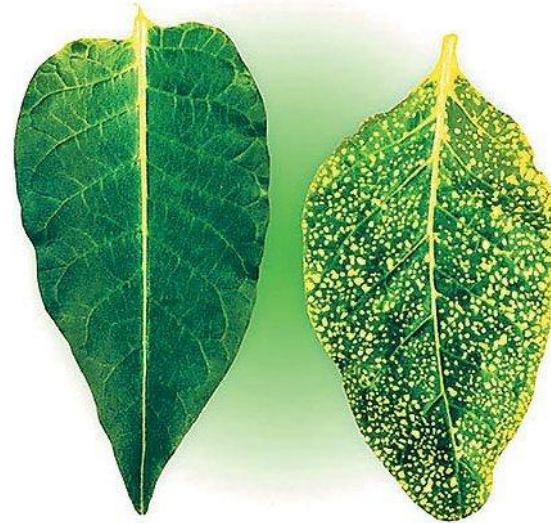
Любое вращение в трехмерном пространстве может быть представлено как композиция поворотов вокруг трех ортогональных осей.

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

# 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

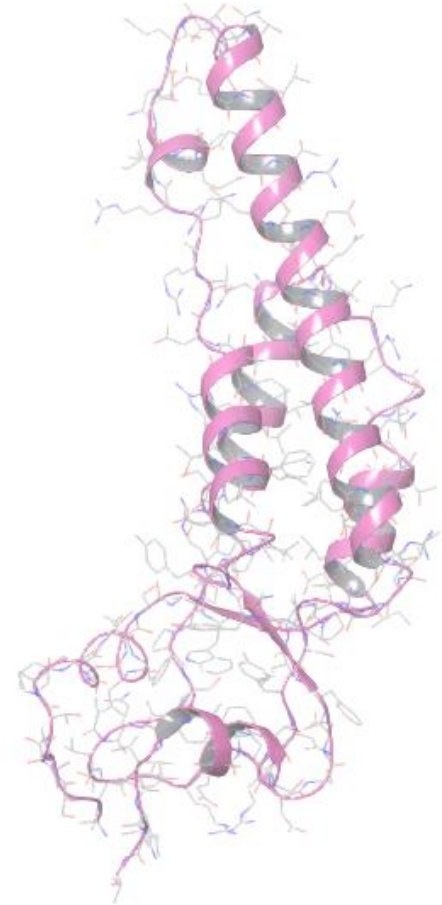
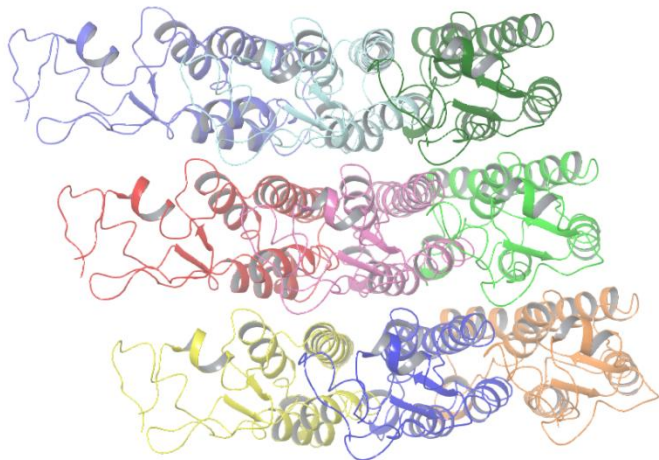
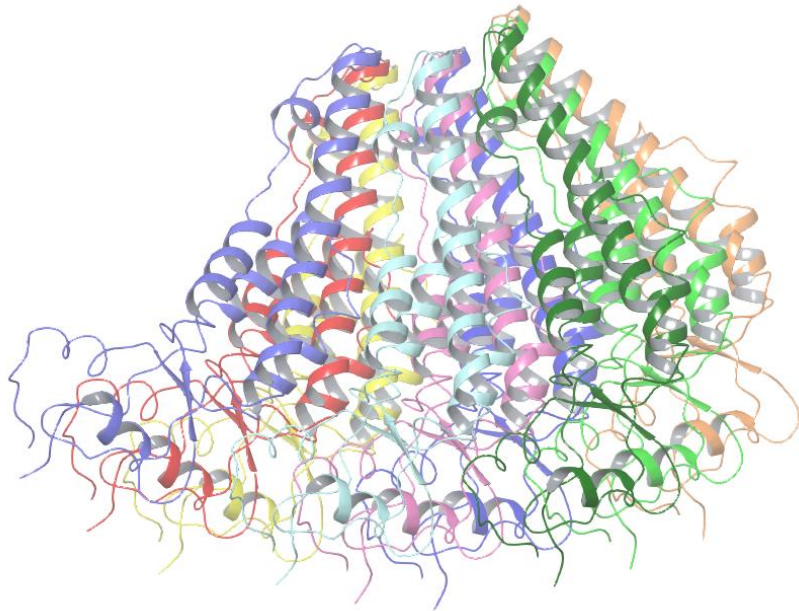
-0.00000

0.00000

-1.40800

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

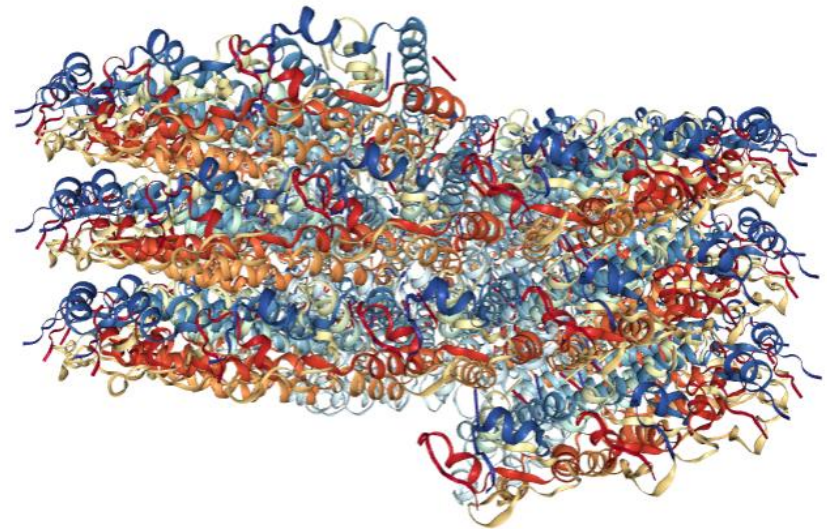
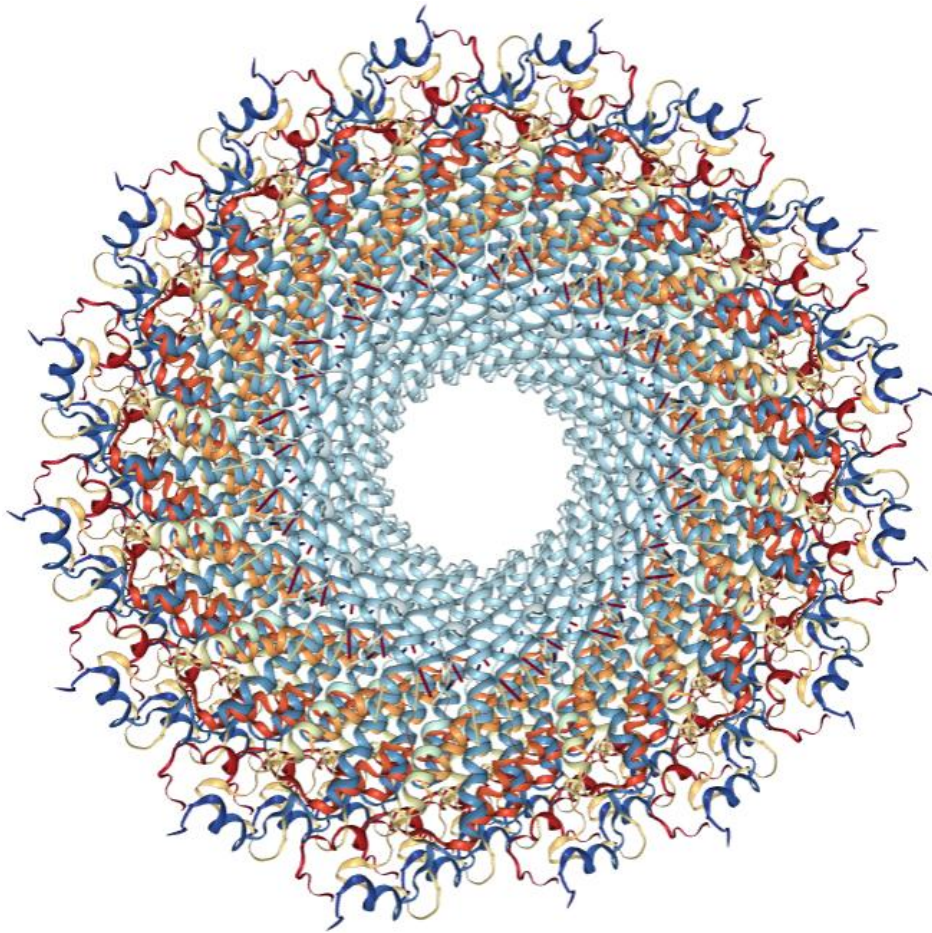
# Вирус табачной мозаики



(pdb-код 4udv)



# Вирус табачной мозаики

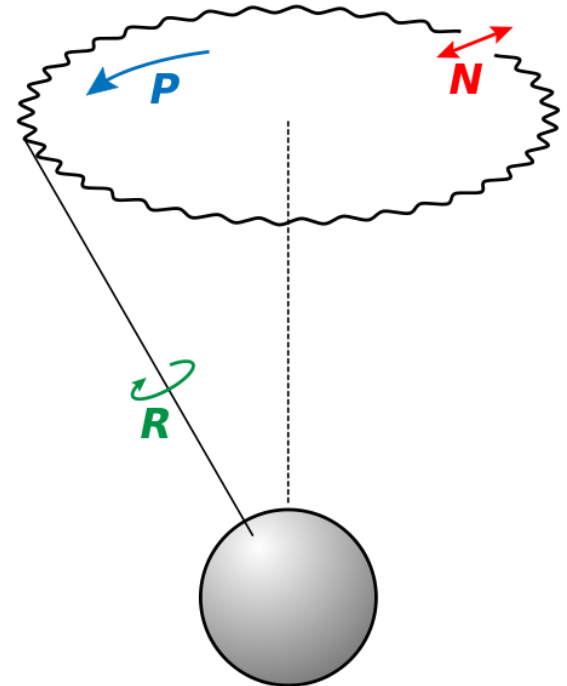
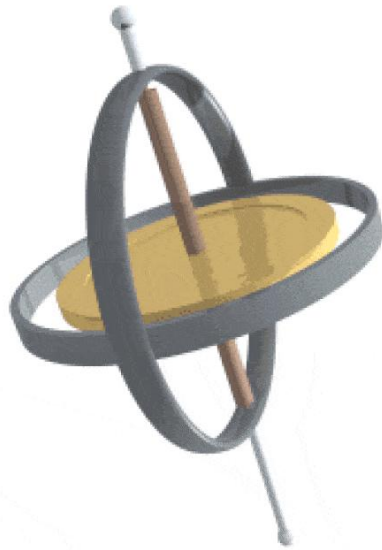


(pdb-код 4udv)



# Задание поворота в пространстве

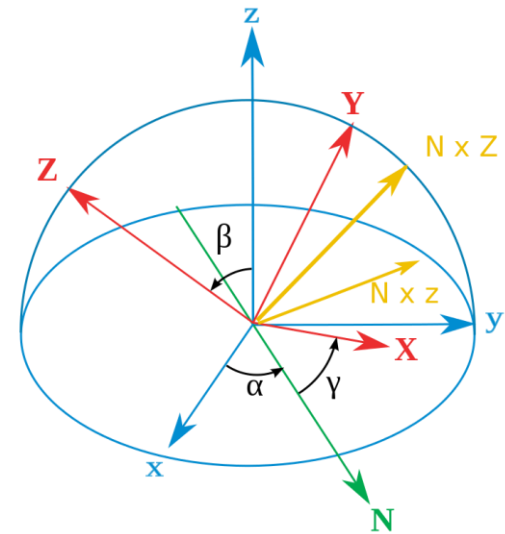
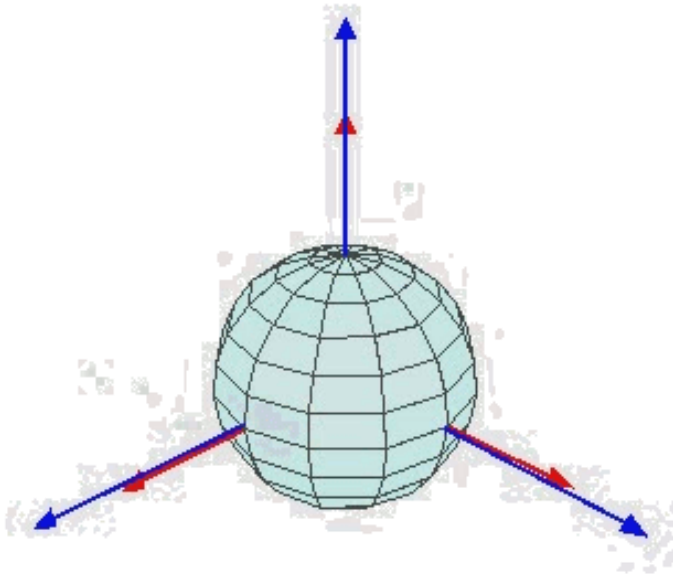
Гироскоп: прецессия, нутация и собственное вращение



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

# Задание поворота в пространстве

Прецессия гироскопа:



$x, y, z$  – исходные оси

$X, Y, Z$  – конечные оси

$N$  - линия узлов

# Задание поворота в пространстве

Углы Эйлера (1776):

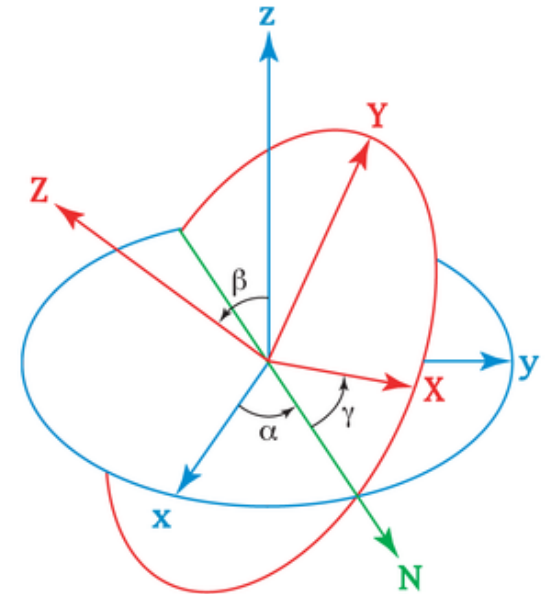
$x, y, z$  — исходные оси

$X, Y, Z$  — конечные оси

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

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

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



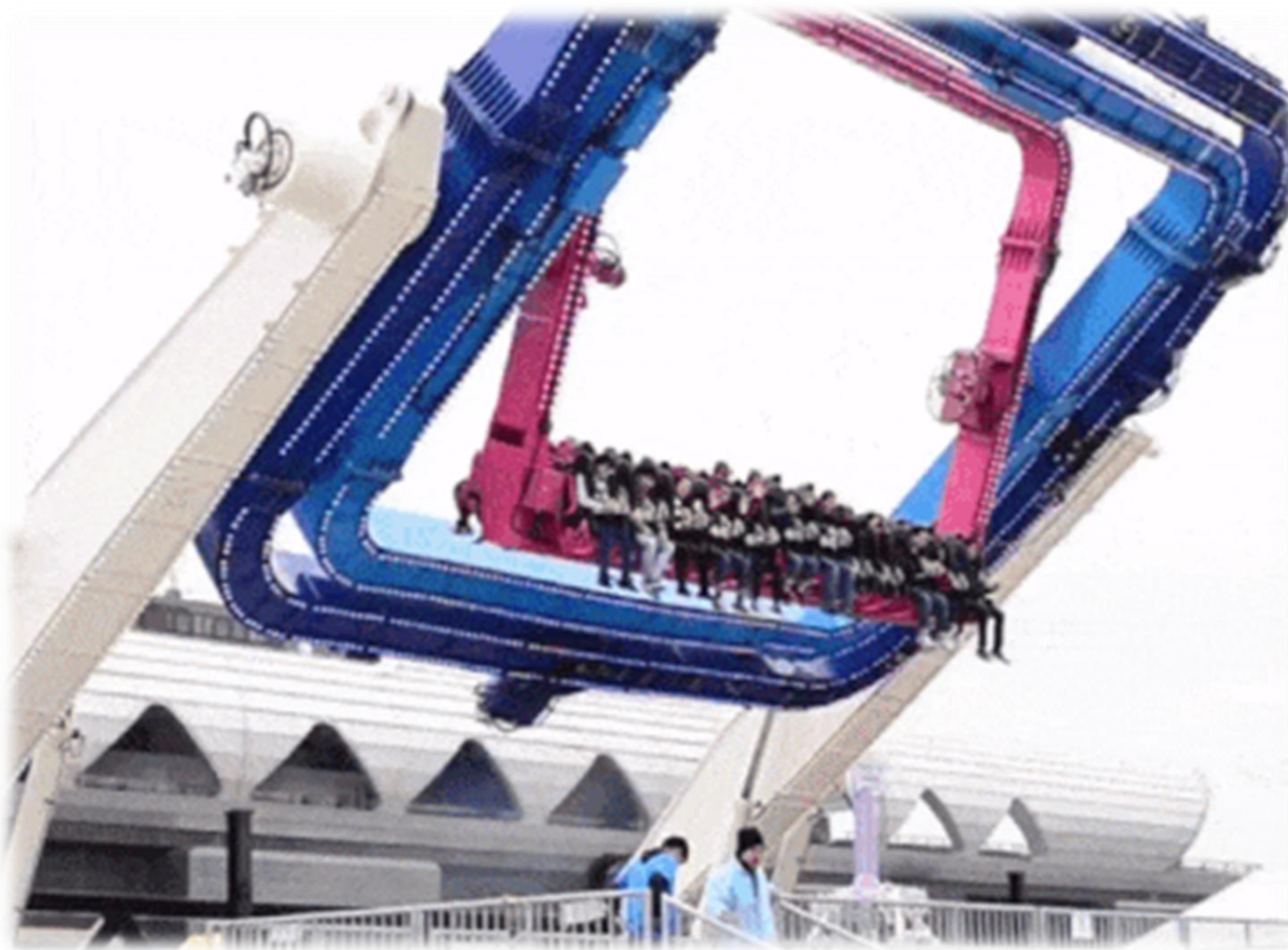
Такие повороты **некоммутативны** и конечное положение системы зависит от порядка, в котором совершаются повороты.

Недостаток: 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.

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

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

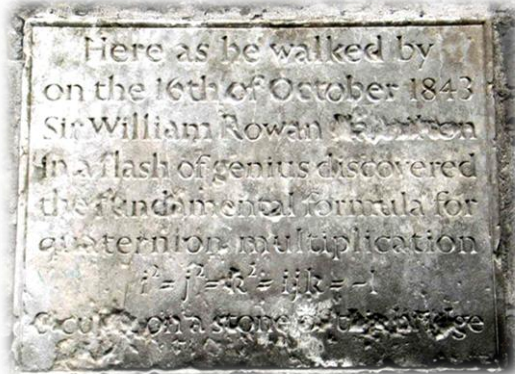
$$|q|^2 = a^2 + b^2 + c^2 + d^2$$

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

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



Уильям Роуэн Гамильтон  
(1805 — 1865)





# Задание поворота в пространстве

Описание поворота через кватернион:

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

$$\overset{\rho}{t} = q \overset{\rho}{v} q^{-1} = \frac{q \overset{\rho}{v} \bar{q}}{|q|^2} \quad \begin{array}{l} \text{-результат поворота вектора } \mathbf{v} \\ \text{вокруг вектора } \mathbf{u} \text{ на угол } \alpha \end{array}$$

$$T = QVQ^{-1} = \frac{QVQ^T}{\sqrt{\det(Q)}} \quad \text{- он же в матричном виде}$$

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

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

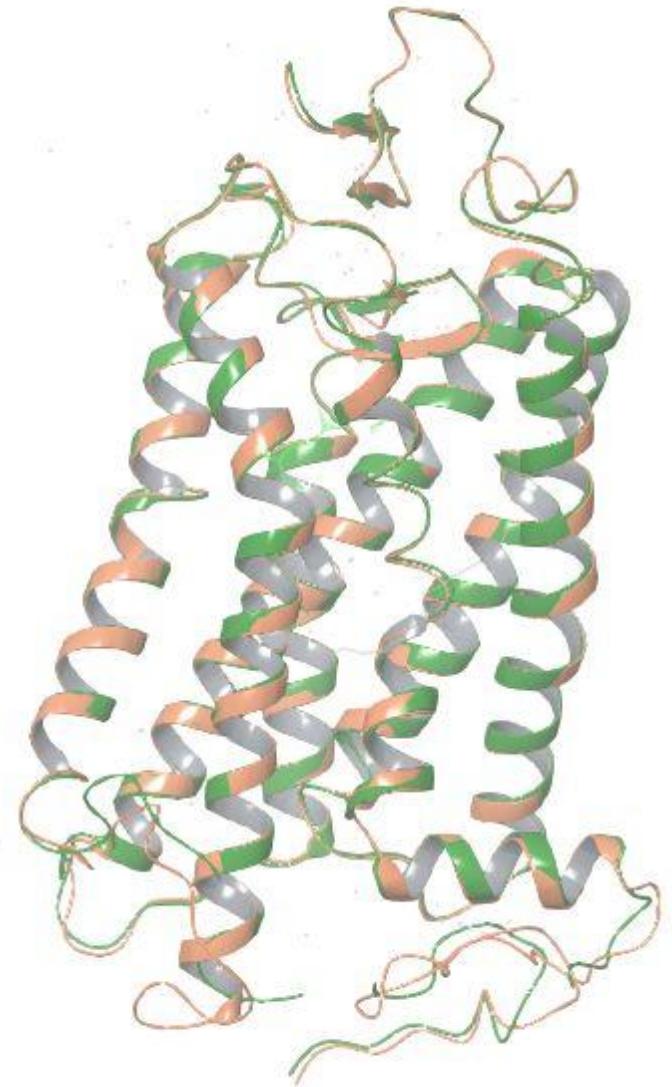
# Структурные выравнивания

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

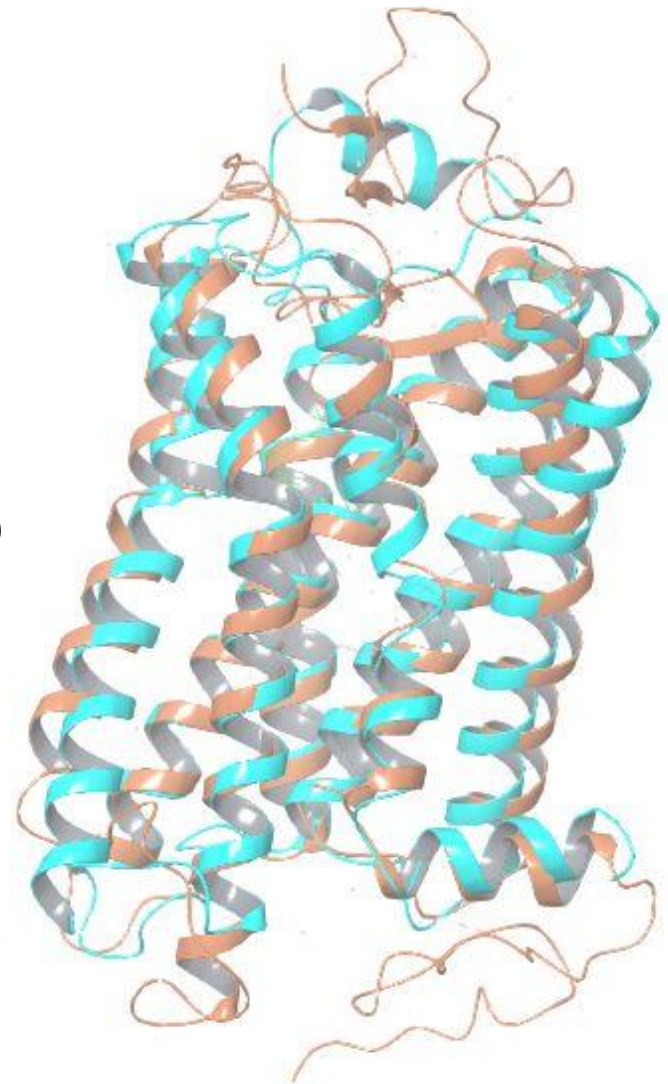
Варианты задачи:

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  $\beta$ 2-адренэргического рецептора (pdb-код 2rh1)

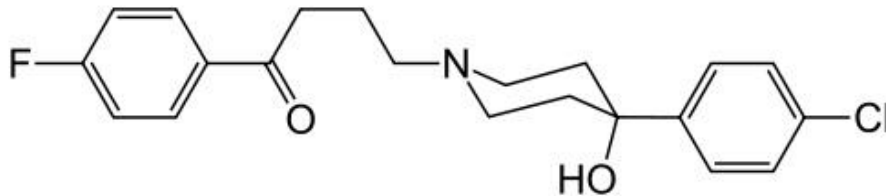
# Структурные выравнивания

Варианты задачи:

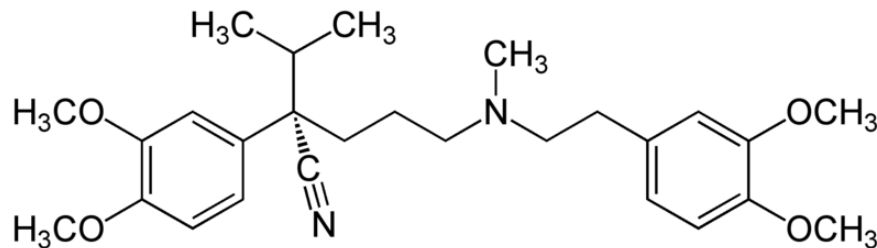
3) Неизвестно ни соответствие, ни порядок следования:

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

(фармакофоры)



Галоперидол  
(блокатор дофаминовых  
рецепторов)



Верапамил  
(блокатор потенциалзависимых  
кальциевых каналов)

Оба проявляют активность в  
отношении калиевых каналов

# Структурные выравнивания

Меры 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)

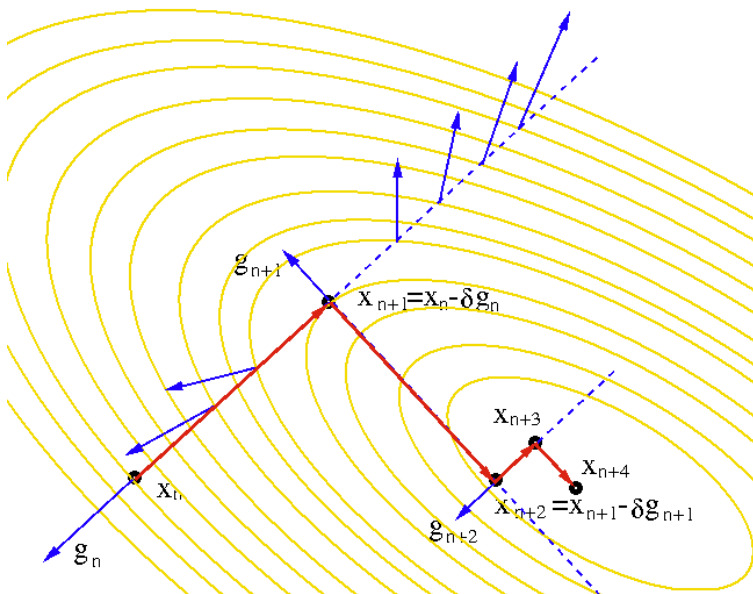


# Методы минимизации. Градиентный спуск

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  $-\nabla F$ :

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



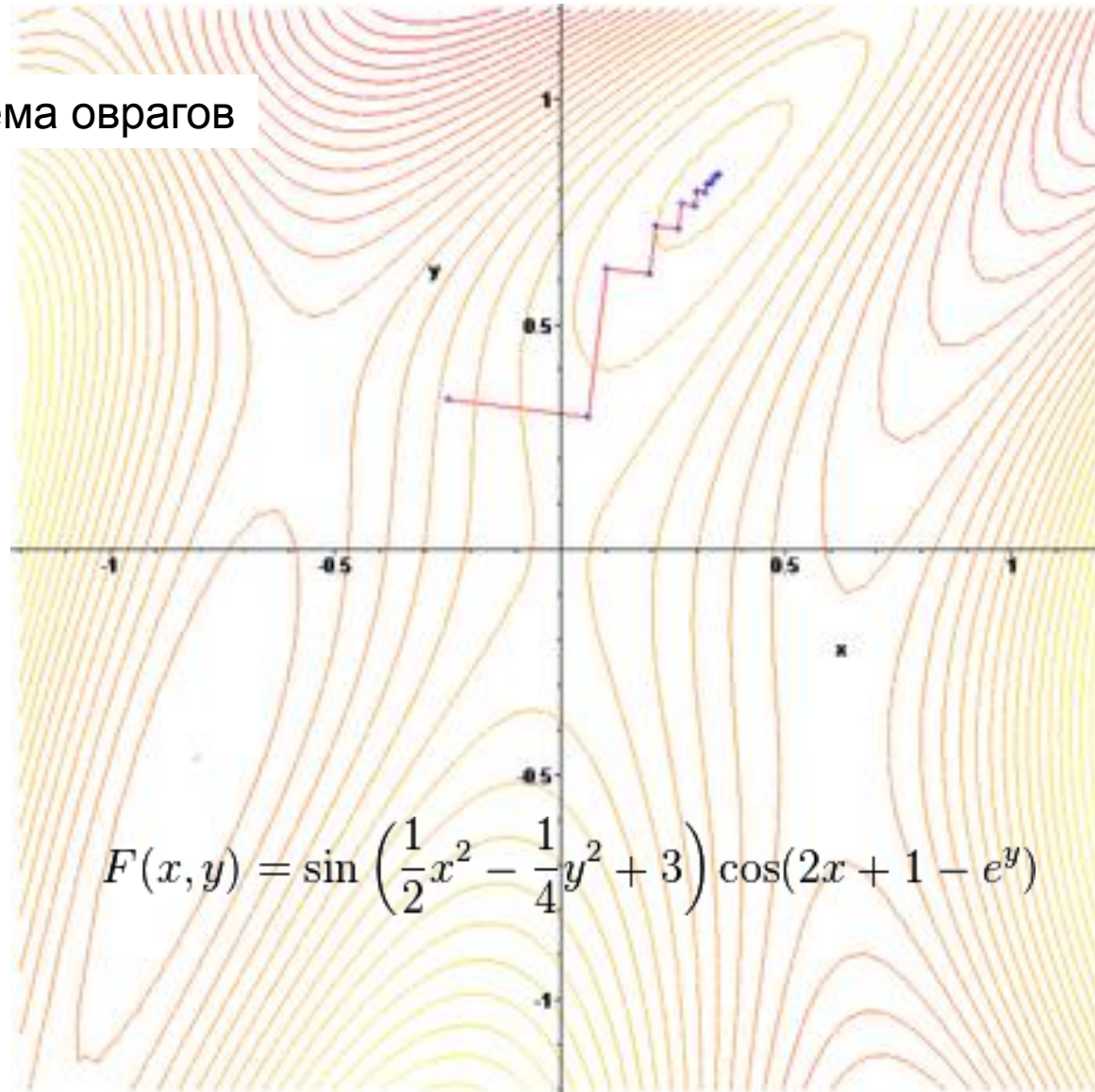
Величина шага может быть

- постоянной
- дробной
- **определяемой на каждой итерации заново**

**(метод наискорейшего спуска, steepest descent)**

# Методы минимизации. Градиентный спуск

Проблема оврагов

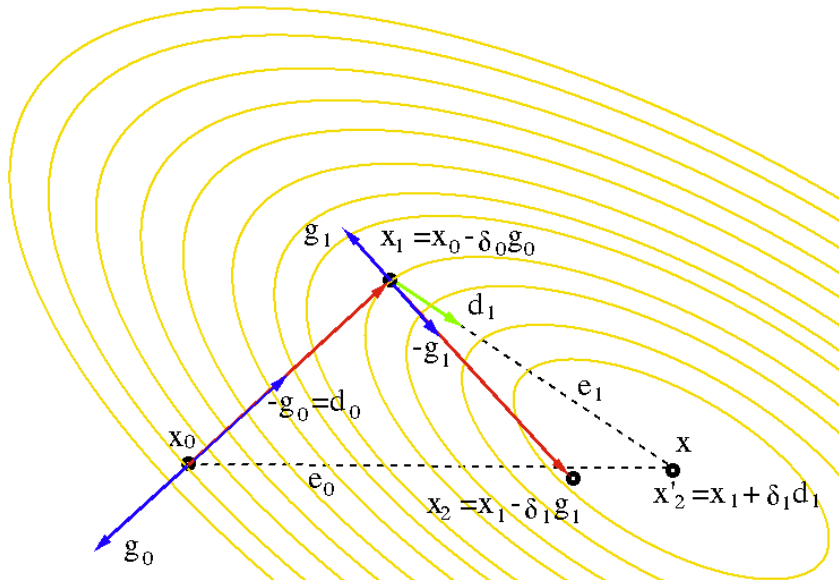


# Методы минимизации. Метод сопряженных градиентов

(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

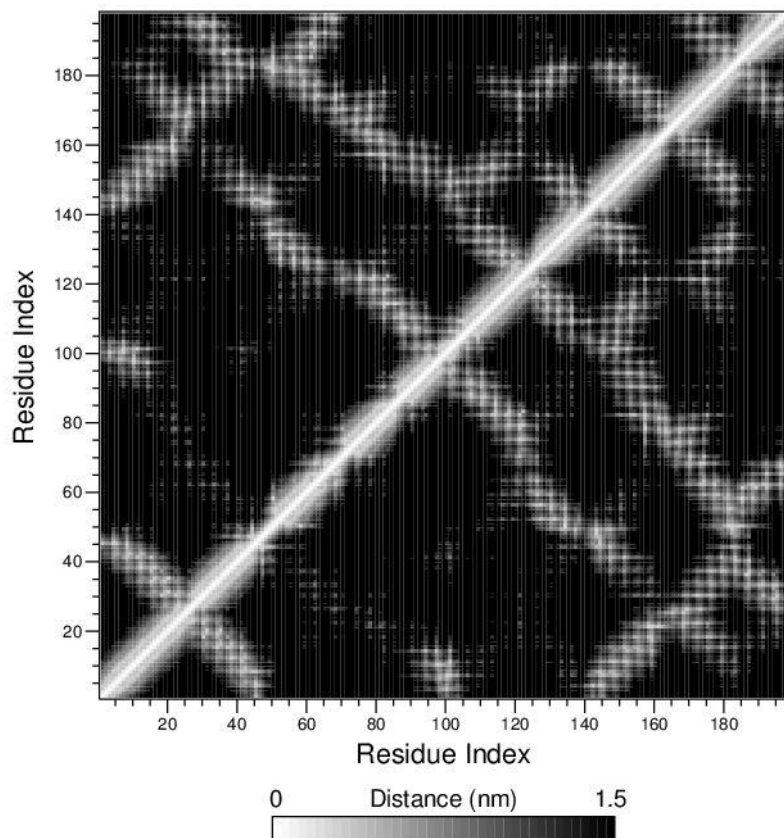
$$f(\vec{x}) = \vec{x}^T A \vec{x} + \vec{b} \cdot \vec{x} + c.$$

метод гарантированно сходится за  $n$  шагов.

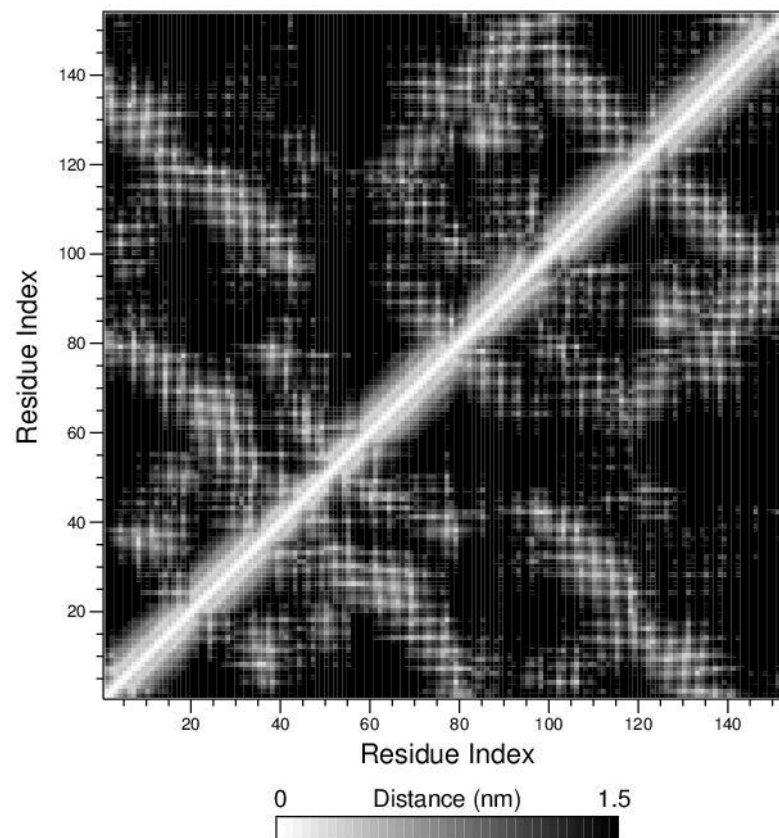
# Структурные выравнивания

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

colicin

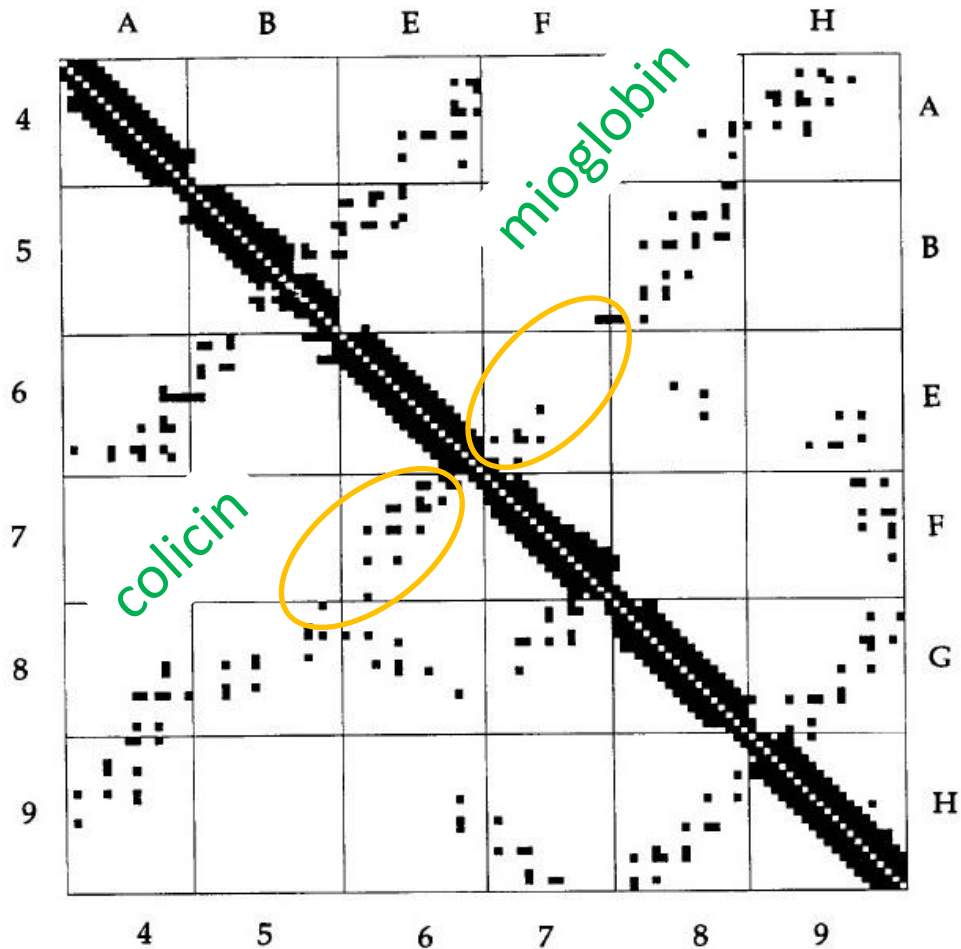


mioglobin



# Структурные выравнивания

Выравнивание матриц расстояний – программа **DALI** (Distance-matrix ALIngment) (1993)



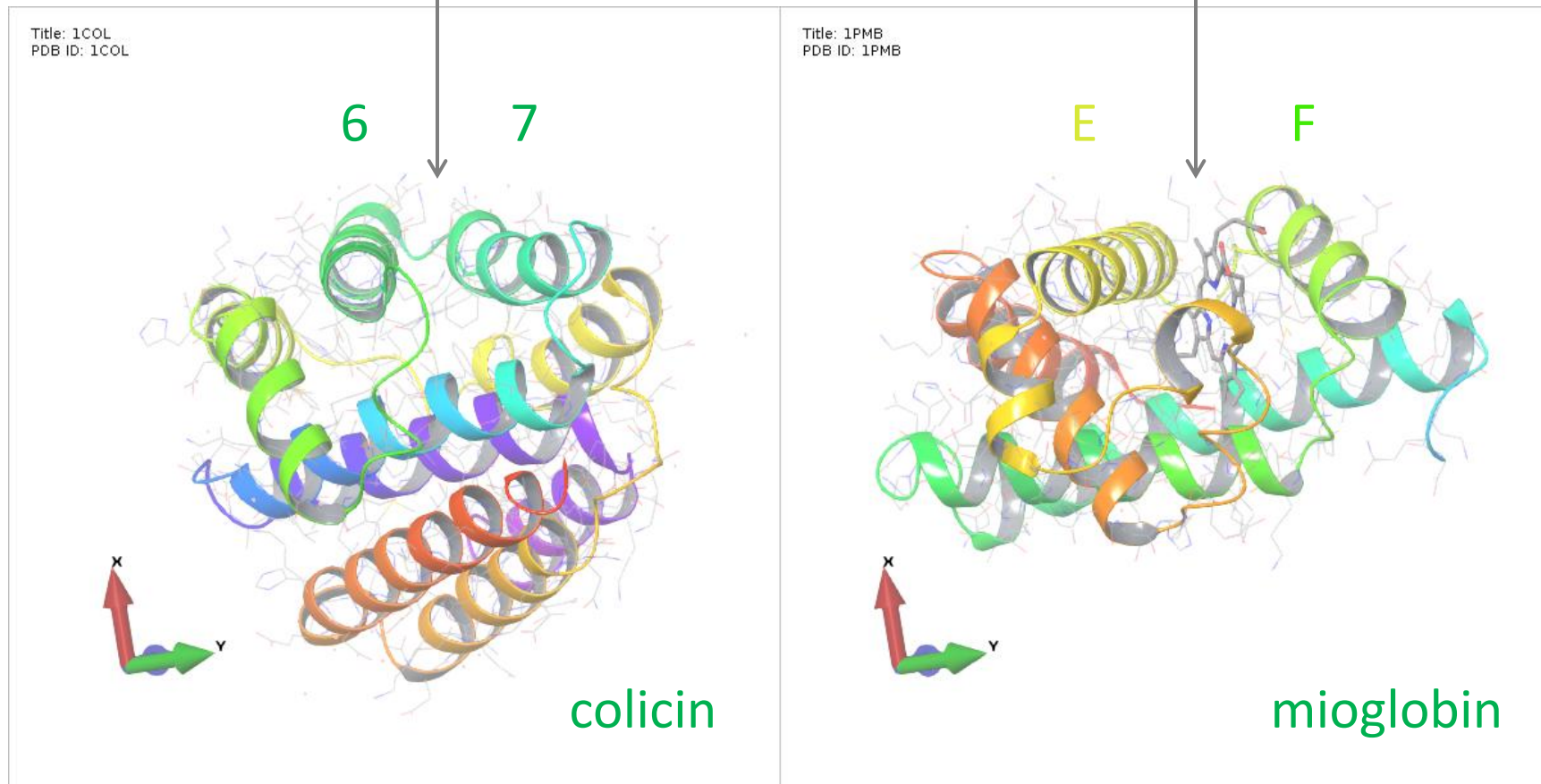
Особенности: свободная топология обнаруженных сходств в структуре, в т.ч. и «обратных» фрагментов.



# Структурные выравнивания

Есть контакт между спиралями

Гем; нет контакта между спиралями





# Структурные выравнивания

Structural alignment:

```

1COL SSA 5 CCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCHHHHHHHHHHHHHHHHHHHHCCCCCCCCHHHHHHHHHH
1PMB SSA  -----
1COL 5 AKDERELLEKTSELIAGMGDKIGEHLGDKYKAIKDIADNIKNFQGKTIRSFDDAMASLN
1PMB -----

1COL SSA 65 HHHHCCCCCCCHHHHHHHHHHHHHHHHHH--CHHHHHHHH-HHH--H-CCCCC-C-C-----
1PMB SSA 1  -----C-CHHHHHHHHHHHHHHHHHHHHHHCHHHHHHHHHHHHHHHHHHHHCHHHHHHHHCCCCCCCCH
1COL 65 KITANPAMKINKARDALVNAWKHV--DAQDMANK-LGN--L-SKAFK-V-A-----
1PMB 1  -----G-LSDGEWQLVLNVWGKVEADVAGHGQEV LIRLFKGGHPETLEKFDKFKHLKSE
              :      :      ::* * :* *          : *

1COL SSA 119 -----CHHHHHHHHHHHHHHHHHHHHHH-CHH-HHHHHHHHHHHHHHCCH-HHHHHHHHHHHH
1PMB SSA 53 HHHHHHHCHHHHHHHHHHHHHHHHHHHHHHCCCCHHHHHHHHHHHHHHHH-HC-CCC-HHH-HHHH
1COL 119 -----DVVMKVEKVVREKSIEGYET-GNWG-PLMLEVESWVLSGIAS-SVALGIFSATL
1PMB 53 DEMKASEDLKKHGNTVLTALGGILKKKGHHEAELTPLAQSHAT-KH-KIP-VKY-LEFIS
              *:      :      *          :      *          *          *

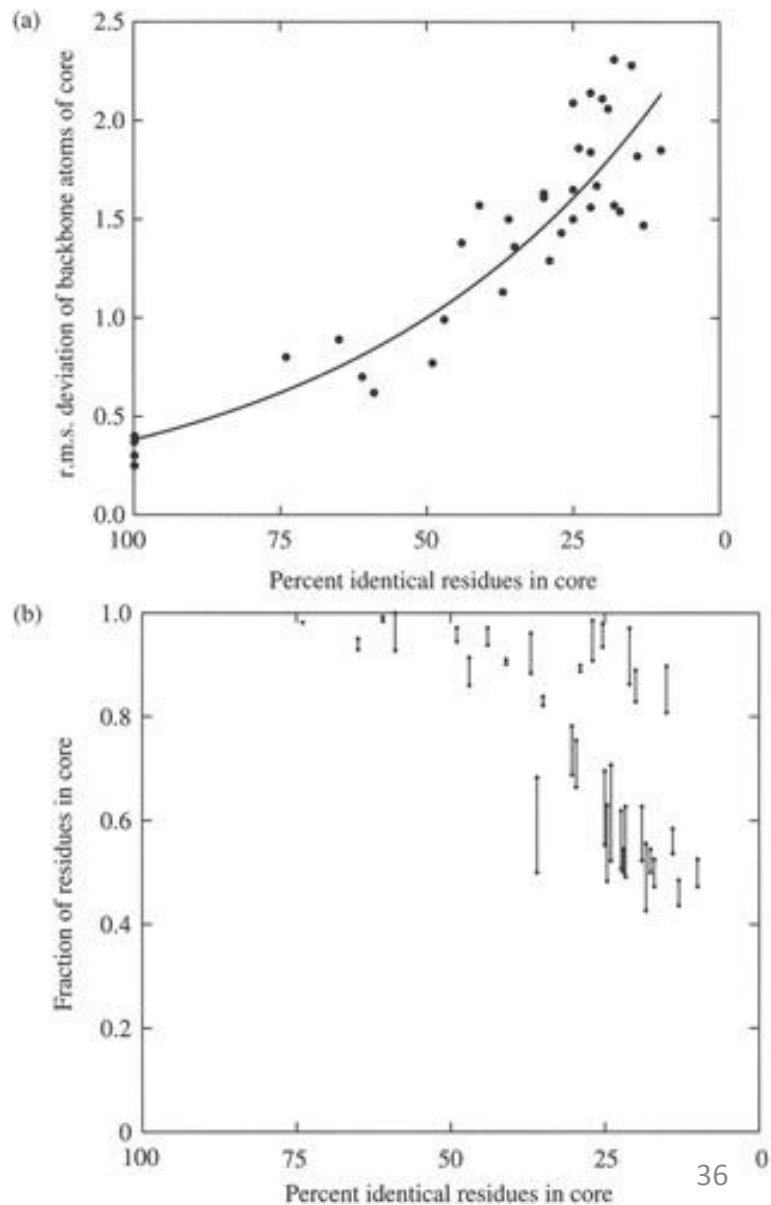
1COL SSA 179 HHHHHHH-----CCHHHHHHHHHHHHHHHHHHHHHHHH--CH-----HHHHHHHHHCCC
1PMB SSA 109 HHHHHHHHHHHCCCCC-CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCC-----
1COL 179 GAYALSL-----GVPAIAVGIAGILLA AVV GALI--DD-----KFADALNNEIIR
1PMB 109 EAI IQVLQSKHPGDF-GADAQGAMSKALELFRNDMAAKYKELGFQG-----
              *      *          * * *          *          :          :          :
    
```

RMSD: 3.827 Angstrom

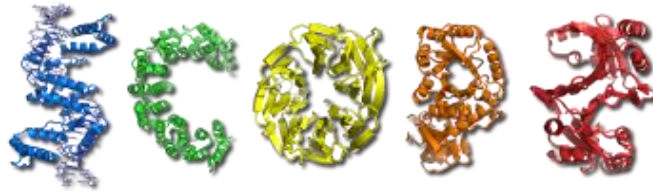
# Эволюция белковых структур

Вариации, встречающиеся в семействах гомологичных белков с одинаковой функцией, показывают, как структура приспособляется 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.



# Классификация структур белков. SCOP

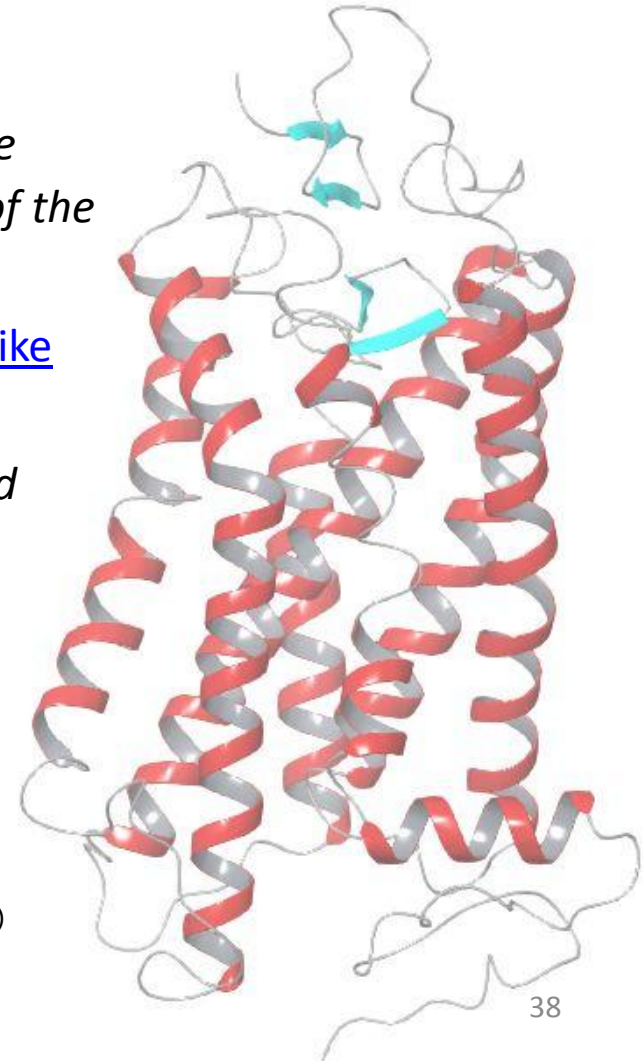


**SCOP** (Structural classification of proteins) (Murzin et al., 1994 - 2009) – организация структур для отображения их эволюционного происхождения и структурного сходства. Основные уровни организации:

- **Классы:**  $\alpha$ ,  $\beta$ ,  $\alpha+\beta$ ,  $\alpha/\beta$  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/> - что-то странное ☹

# 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»



Продолжение следует...