

Строение белков. Структура и организация.  
Структурная Биоинформатика и моделирование лекарств  
(ВШЭ)

Головин А.В.<sup>1</sup>

<sup>1</sup>МГУ им М.В. Ломоносова, Факультет Биоинженерии и Биоинформатики

Москва, 2017

# Содержание:

Введение

Уровни организации структуры белка

Типы взаимодействий в белках

PDB

Визуализация с PyMol



# Содержание курса, лекции

- Введение, Банк структур PDB.
- Визуализация структуры биополимеров, PyMol
- Предсказание структур белков
- Молекулярная механика. Силовые поля.
- Моделирование самосборки
- Поиск новых био-активных молекул и химоинформатика
- Предсказание вторичной структуры РНК.
- Пространственные и упругие параметры ДНК как спирали.
- Макромолекулярный докинг.



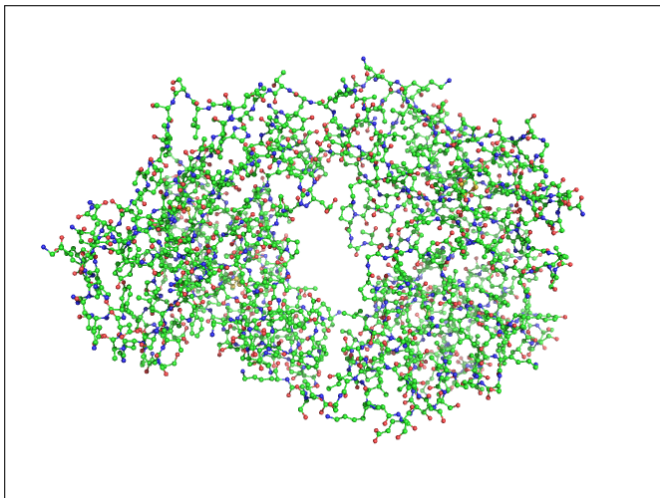
# Практические навыки

- PyMol: скриптование и анимация, визуализация.
- Modeller: гомологичное моделирование, внесение мутаций в структуру.
- Modeller: построение сложных четвертичных структур.
- Gromacs: моделирование молекулярной динамики.
- Autodock: Проводить докинг низкомолекулярных лигандов в структуру белка.
- Autodock: Проводить поиск кандидатов ингибиторов в библиотеке соединений.
- OpenBabel: Использовать химоинформатические базы данных.
- Zdock: Предсказывать структуру комплекса двух биополимеров.
- Foldit: Играть в самосборку белка.



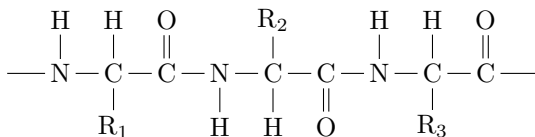


# Структура каталитического антитела



# Что такое белок?

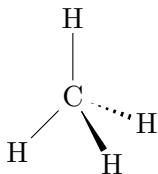
**Белки** — высокомолекулярные органические вещества, состоящие из соединённых в цепочку пептидной связью альфа-аминокислот. (wikipedia)



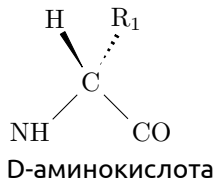
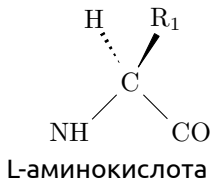
**Или:** белок это линейный полярный полимер, где мономерами является выборка из примерно 20 L-альфа-аминокислот.



# Что такое L альфа-аминокислота?

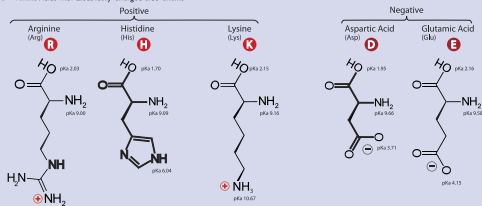


атом углерода в  $sp^3$  гибридизации имеет тетраэдрическое окружение

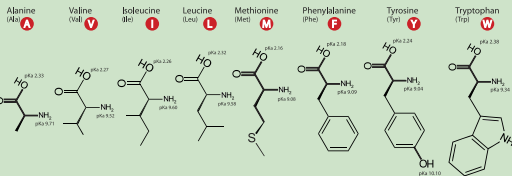
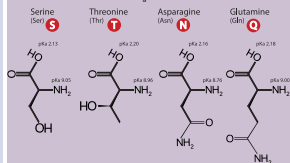


## АМИНОКИСЛОТЫ

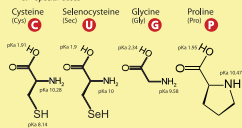
## A. Amino Acids with Electrically Charged Side Chains



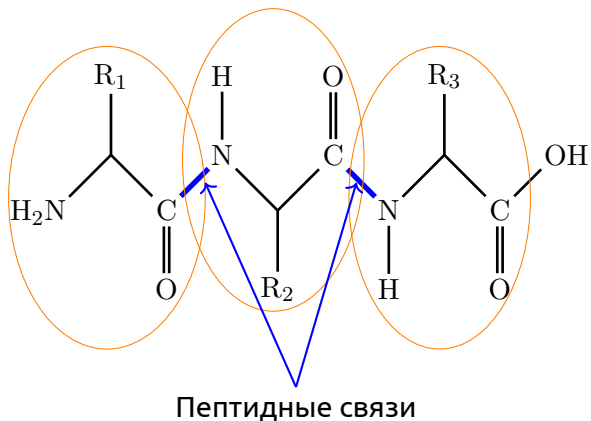
## B. Amino Acids with Polar Uncharged Side Chains



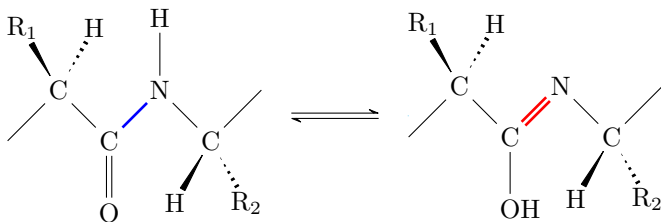
## C. Special Cases



# Пептидная связь



# Пептидная связь, таутомерия

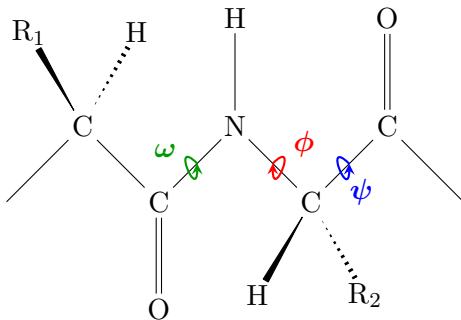


# Пептидная связь, свойства

- Пептидная связь прочнее, чем другие амиды
- Атомы пептидного звена ( $C_{\alpha}$ -C-N-  $C_{\alpha}$ ) лежат в одной плоскости
- Валентные углы у атомов C и N примерно равны  $120^{\circ}$
- Вращение вокруг связи C-N затруднено
- Возможны cis- и trans-конфигурации; в белках преобладают trans
- Карбонильный кислород – хороший акцептор водорода
- Амидный азот – хороший донор водорода

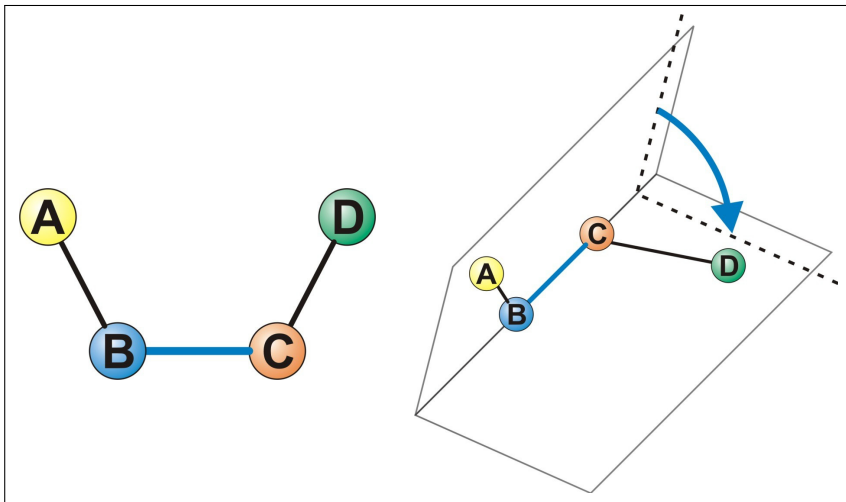


# Вращения вокруг связей в остове белка



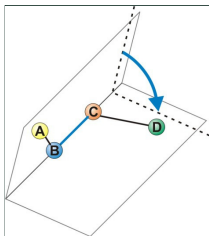


# Двугранные (торсионные) углы

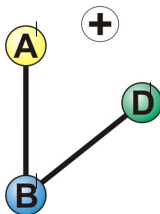


# Двугранные (торсионные) углы

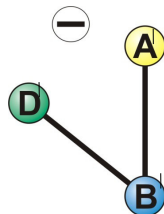
**Могут принимать значения от  $-180$  до  $+180$**



Построим проекции всех связей на плоскость, перпендикулярную связи В-С



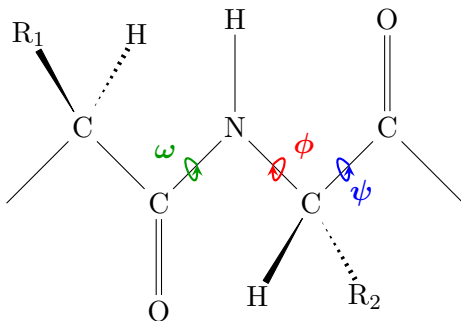
В D совмещается с В А поворотом против часовой стрелки



В D совмещается с В А поворотом по часовой стрелке



# Вращения вокруг связей в осто́ве белка



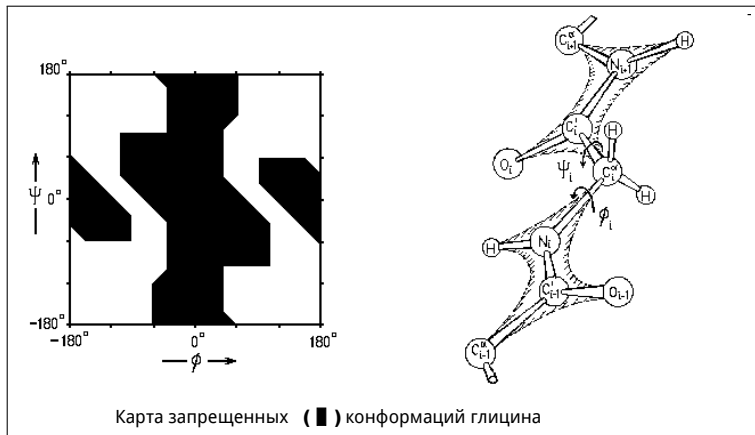
$\left. \begin{matrix} \phi \\ \psi \end{matrix} \right\}$  теоретически могут быть: от  $-180^0$  до  $+180^0$

а  $\omega$ ?



# Карта Рамачандрана

даже в полиглициновой цепи существуют стерические ограничения



# Уровни организации структуры белка

- Первичная структура
- Вторичная структура
- Укладка (fold)
- Третичная структура
- Четвертичная структура



# Первичная структура

Первичная структура – это аминокислотная последовательность:

Met-Ala-Gly-Trp-Ala-Val-Asp ...



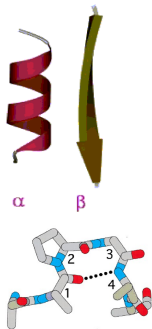
# Вторичная структура

## Вторичная структура

**белка** - это упорядоченные расположения атомов основной цепи полипептида, безотносительно к типам боковых цепей (групп) и их конформациям.

Если упорядоченность такова, что двугранные углы одинаковы у всех остатков, то говорят о регулярной вторичной структуре. Регулярными вторичными структурами являются спирали и  $\beta$ -структуры.

Пример нерегулярной вторичной структуры  $\beta$ -поворот ( $\beta$ -изгиб, реверсивный поворот).



## Вторичная структура

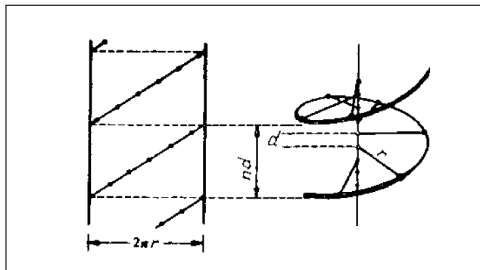
Любую регулярную структуру можно представить в виде линейной группы, т.е. спирали. Спираль можно описать с помощью следующих параметров:

$d$  – смещение вдоль оси, в расчете на 1 элемент (атом  $C_\alpha$ ),

$r$  – расстояние от  $C_\alpha$ -атома до оси,

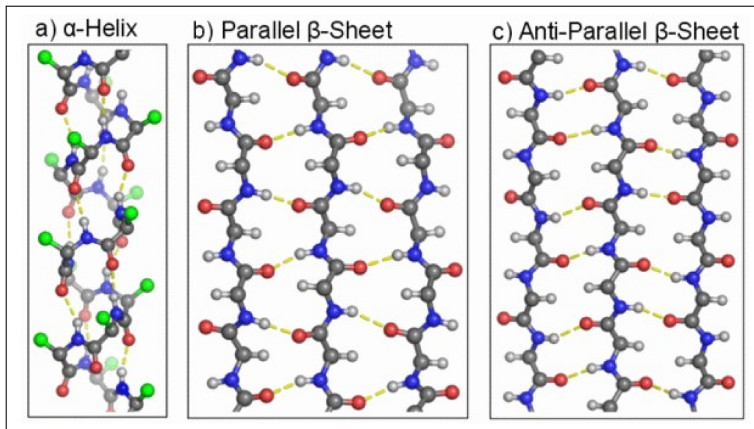
$n$  – число элементов на один виток спирали, хиральность (направление закрутки) определяется знаком, "+" - право закрученные структуры, "-" – левозакрученные.

/Г.Шульц, Р.Ширмер "Принципы структурной организации белков"/





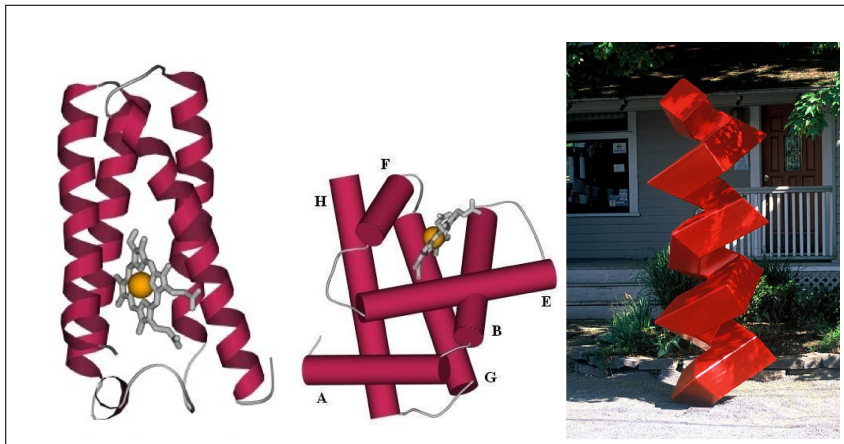
# Регулярные вторичные структуры



# Укладка (fold)

Укладкой называют организацию в пространстве элементов регулярной вторичной структуры.

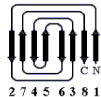
Пример:  $\alpha$ -спиральные белки



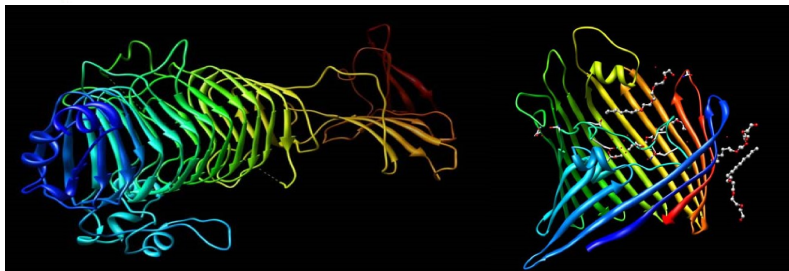
# $\beta$ структурные белки



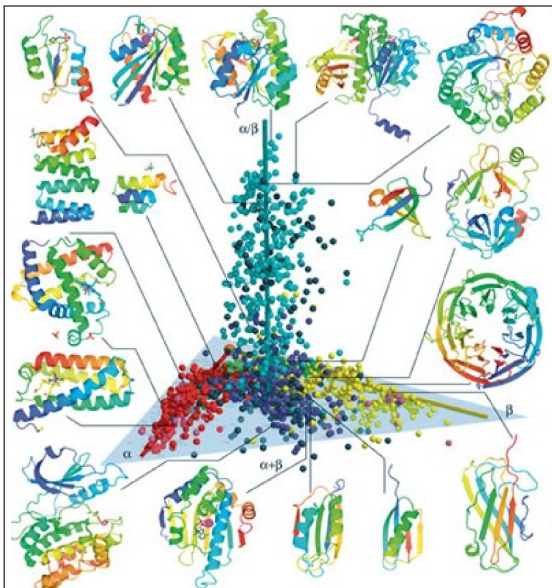
ГРЕЧЕСКИЙ КЛЮЧ



РУЛЕТ



# Распределение в природе



# Третичная структура

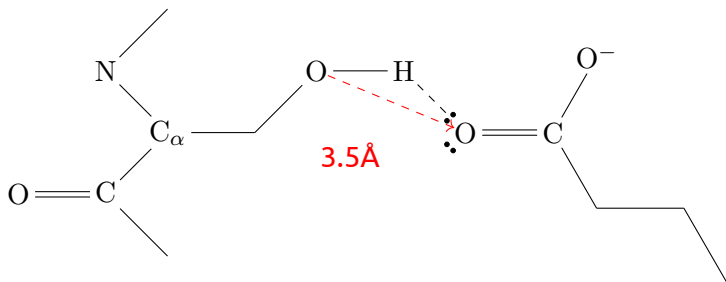
Третичной структурой называют расположение в пространстве всех атомов одной полипептидной цепи.

Т.е. описание третичной структуры включает в себя:

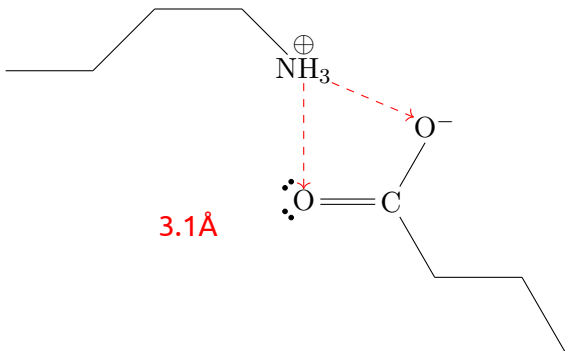
- описание элементов вторичной структуры,
- описание типа укладки,
- описание структуры петель,
- описание конформаций боковых групп всех аминокислотных остатков.



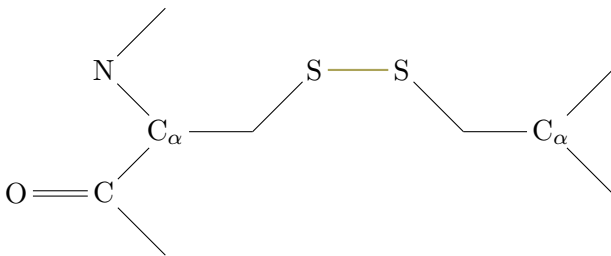
# Вспомогательные взаимодействия: водородные связи



# Ионные пары

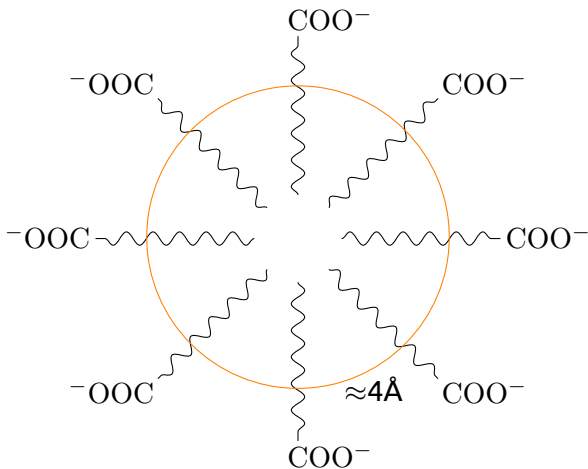


# Дисульфидные мостики характерны для секретируемых белков

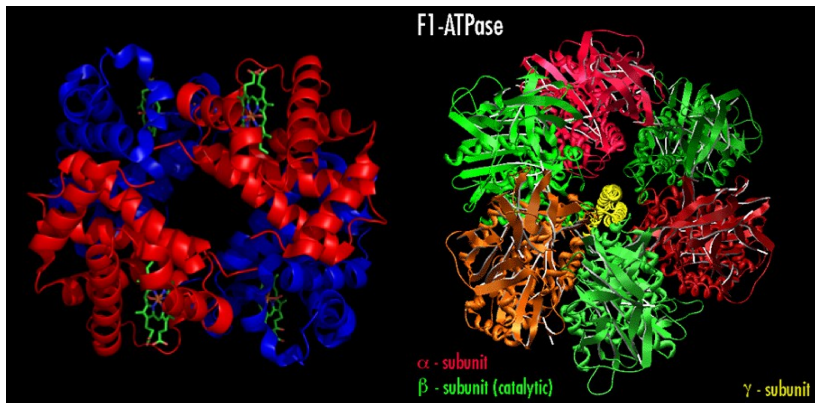


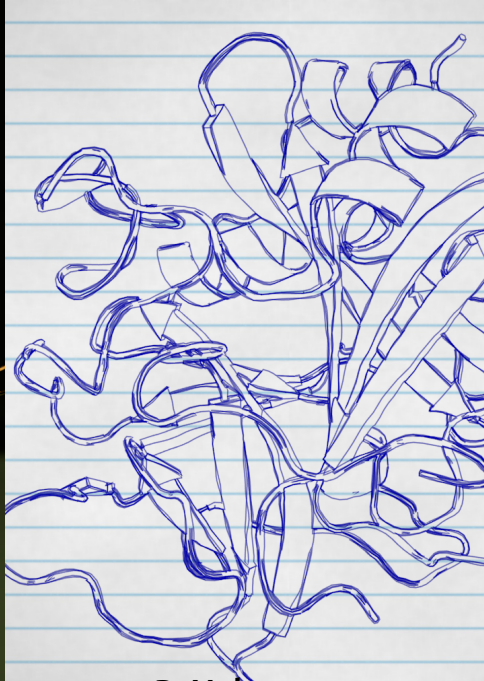
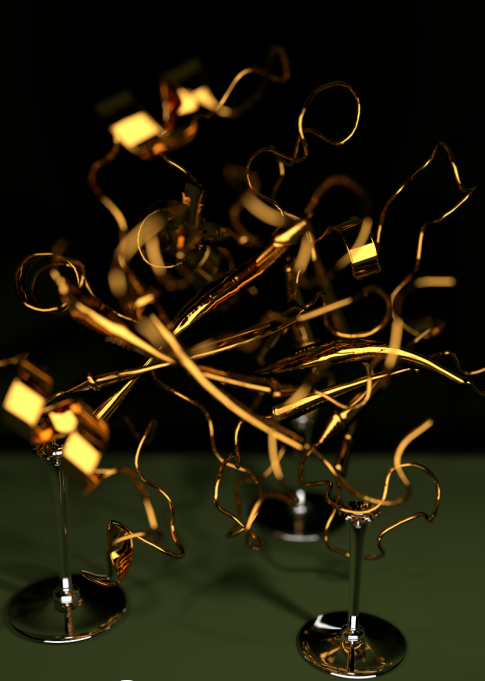


# Гидрофобные взаимодействия – главный фактор, заставляющий глобулу свертываться



# От четвертичной структуры к молекулярным машинам

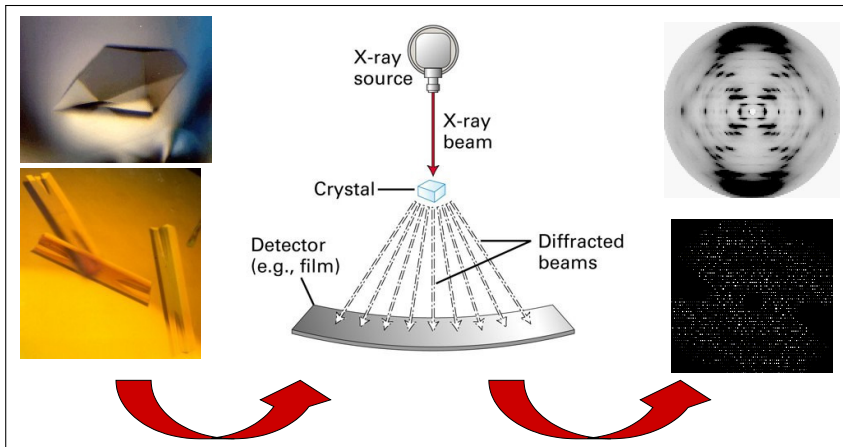




**Визуализация третичных структур, PyMol и т.д.**

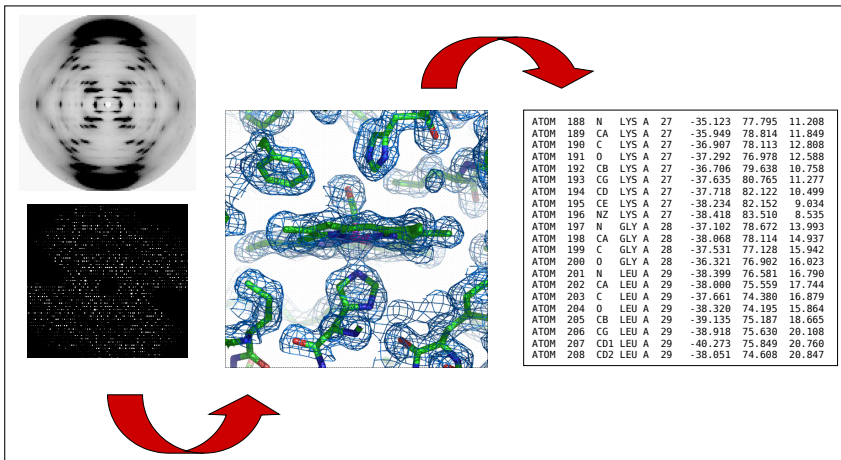
# Основные этапы расшифровки 3D-структуры

## Эксперимент



# Основные этапы расшифровки 3D-структуры

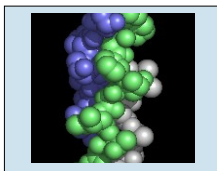
## Вычисления и моделирование



# Основные этапы расшифровки 3D-структуры

## Анализ

ATOM	188	N	LYS	A	27	-35.123	77.795	11.208
ATOM	189	CA	LYS	A	27	-35.949	78.814	11.849
ATOM	190	C	LYS	A	27	-36.907	78.113	12.808
ATOM	191	O	LYS	A	27	-37.292	76.978	12.588
ATOM	192	CB	LYS	A	27	-36.706	79.638	10.758
ATOM	193	CG	LYS	A	27	-37.635	80.765	11.277
ATOM	194	CD	LYS	A	27	-37.718	82.122	10.499
ATOM	195	CE	LYS	A	27	-38.234	82.152	9.034
ATOM	196	NZ	LYS	A	27	-38.418	83.510	8.535
ATOM	197	N	GLY	A	28	-37.102	78.672	13.993
ATOM	198	CA	GLY	A	28	-38.068	78.114	14.937
ATOM	199	C	GLY	A	28	-37.531	77.128	15.942
ATOM	200	O	GLY	A	28	-36.321	76.902	16.023
ATOM	201	N	LEU	A	29	-38.399	76.581	16.790
ATOM	202	CA	LEU	A	29	-38.000	75.559	17.744
ATOM	203	C	LEU	A	29	-37.661	74.380	16.879
ATOM	204	O	LEU	A	29	-38.320	74.195	15.864
ATOM	205	CB	LEU	A	29	-39.135	75.187	18.665
ATOM	206	CG	LEU	A	29	-38.918	75.630	20.108
ATOM	207	CD1	LEU	A	29	-40.273	75.849	20.760
ATOM	208	CD2	LEU	A	29	-38.051	74.608	20.847



**Программы для визуализации и анализа 3D —  
RasMol, PyMol, SPDBViewer, WebMol, .....**



# Брукгейвенский банк пространственных структур (PDB)

The screenshot shows the RCSB PDB website interface. At the top, the browser address bar shows the URL `http://www.pdb.org/pdb/home/home.do`. The page header includes the RCSB PDB logo and the text "A MEMBER OF THE PDB" and "An Information Portal to Biological Macromolecular Structures". The search bar is prominently displayed with the text "Search | All Categories:" and a search input field containing "e.g., PDB ID, molecule name, author".

The main content area is divided into several sections:

- Biological Macromolecular Resource**: This section includes a "Full Description" and a "Featured Molecules" list. The "Molecule of the Month" is the **Pyruvate Dehydrogenase Complex**, described as a combination of crystallography, NMR spectroscopy, and electron microscopy revealing the secrets of this complex. The **Protein Structure Initiative Featured System** is **Solute Channels**, which maintain a steady traffic of small molecules across their membranes.
- Navigation and Tools**: Includes links for "All Categories", "Author", "Macromolecule", "Sequence", "Ligand", "Browse", and "Advanced".
- Left Sidebar**: Contains a "Customize This Page" section with an "Available on the App Store" badge, a "PDB-101" link, and a "MyPDB" section for user accounts.
- Right Sidebar**: Features "New Structures", "New Features", "RCSB PDB News", and a "Head Back to School with PDB-101" section with an image of a pencil.

The footer of the page shows the URL `http://www.pdb.org/pdb/home/home.do` and the page number `28/28`.

# Protein Data Bank

- Одна запись (документ) соответствует одному эксперименту по определению пространственной структуры макромолекулы или комплекса молекул
- Архивный банк – за содержание записи отвечают авторы соответствующей работы
- Совместно поддерживается университетом Rutgers (штат Нью-Джерси); EBI (Англия) и BIRD (Institute for Bioinformatics Research and Development, Япония)
- Адреса в Интернете: <http://www.rcsb.org/pdb>, <http://www.ebi.ac.uk/msd/> , <http://www.pdbj.org/> .
- Сайты снабжены поисковыми системами
- Все записи открыты для копирования через FTP





# Что хранится в PDB?

**PDB**  
PROTEIN DATA BANK

**An Information Portal to Biological Macromolecular Structures**  
As of **Tuesday Nov 10, 2009 at 4 PM PST** there are 61418 Structures [?](#) | [PDB Statistics](#) [?](#)

**WHAT'S NEW** | [HELP](#) | [PRINT](#)  [Search](#) [?](#) | [Advanced Search](#)

**PDB Current Holdings Breakdown**

	Proteins	Nucleic Acids	Protein/NA Complexes	Other	Total
X-ray	49409	1176	2289	17	52891
NMR	7082	874	150	6	8112
<b>Exp. Method</b> Electron Microscopy	175	16	66	0	257
Hybrid	18	1	1	1	21
Other	116	4	4	13	137
<b>Total</b>	<b>56800</b>	<b>2071</b>	<b>2510</b>	<b>37</b>	<b>61418</b>



# Запись PDB

- Идентификатор записи (PDB ID, PDB-код) вида 1XYZ (цифра и три буквы/цифры) например: 1B8I, 9ANT, 10MH
- Каждая запись содержит координаты центров атомов (в некоторой произвольной системе координат) и сопровождающую информацию
- Каждая запись есть текстовый файл специального формата (PDB-формат)



# Заголовок PDB-файла

```

HEADER      COMPLEX (DNA-BINDING PROTEIN/DNA)          27-APR-97   1WET
TITLE      STRUCTURE OF THE PURR-GUANINE-PURF OPERATOR COMPLEX
COMPND     MOL_ID: 1;
COMPND     2 MOLECULE: PURINE REPRESSOR-GUANINE-PURF-OPERATOR;
COMPND     3 CHAIN: A;
COMPND     4 MOL_ID: 2;
COMPND     5 MOLECULE: DNA (AACGAAAACGTTTTTCGT);
COMPND     6 CHAIN: B;
COMPND     7 ENGINEERED: YES;
COMPND     8 BIOLOGICAL_UNIT: HOMODIMER
SOURCE     MOL_ID: 1;
SOURCE     2 ORGANISM_SCIENTIFIC: ESCHERICHIA COLI;
SOURCE     3 MOL_ID: 2;
SOURCE     4 SYNTHETIC: YES
KEYWDS     DNA-BINDING REGULATORY PROTEIN, REPRESSOR,
KEYWDS     2 COMPLEX (DNA-BINDING PROTEIN/DNA)
EXPDTA     X-RAY DIFFRACTION
AUTHOR     M.A.SCHUMACHER,A.GLASFELD,H.ZALKIN,R.G.BRENNAN
REVDAT    1 12-NOV-97 1WET 0
JRNL       AUTH  M.A.SCHUMACHER,A.GLASFELD,H.ZALKIN,R.G.BRENNAN
JRNL       TITL  THE X-RAY STRUCTURE OF THE PURR-GUANINE-PURF
JRNL       TITL 2 OPERATOR COMPLEX REVEALS THE CONTRIBUTIONS OF
JRNL       TITL 3 COMPLEMENTARY ELECTROSTATIC SURFACES AND A
JRNL       TITL 4 WATER-MEDIATED HYDROGEN BOND TO COREPRESSOR
  
```



# Координаты атомов в PDB-файле

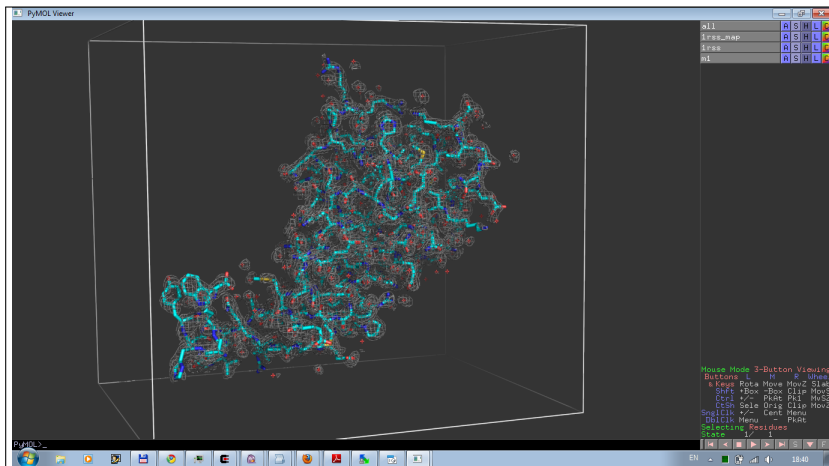
ATOM	1	N	THR	A	3	16.514	-2.279	12.062	1.00	43.86	N
ATOM	2	CA	THR	A	3	17.180	-2.102	13.371	1.00	45.43	C
ATOM	3	C	THR	A	3	16.995	-0.675	13.903	1.00	48.26	C
ATOM	4	O	THR	A	3	16.888	-0.476	15.109	1.00	56.27	O
ATOM	5	CB	THR	A	3	18.658	-2.818	13.510	1.00	61.61	C
ATOM	6	OG1	THR	A	3	18.953	-3.305	14.848	1.00	40.35	O
ATOM	7	CG2	THR	A	3	19.796	-1.970	12.934	1.00	66.69	C
ATOM	8	N	ILE	A	4	16.880	0.318	13.028	1.00	35.25	N
ATOM	9	CA	ILE	A	4	16.614	1.653	13.545	1.00	31.81	C
ATOM	10	C	ILE	A	4	15.180	1.537	14.149	1.00	36.74	C
ATOM	11	O	ILE	A	4	14.824	2.204	15.125	1.00	23.77	O
ATOM	12	CB	ILE	A	4	16.557	2.686	12.441	1.00	32.25	C
ATOM	13	CG1	ILE	A	4	16.613	4.069	13.040	1.00	32.26	C
ATOM	14	CG2	ILE	A	4	15.242	2.611	11.664	1.00	22.31	C
ATOM	15	CD1	ILE	A	4	16.468	5.127	11.966	1.00	56.11	C
ATOM	16	N	LYS	A	5	14.363	0.675	13.544	1.00	41.48	N
ATOM	17	CA	LYS	A	5	13.005	0.429	14.018	1.00	40.63	C
ATOM	18	C	LYS	A	5	13.126	-0.115	15.426	1.00	43.60	C
ATOM	19	O	LYS	A	5	12.360	0.211	16.357	1.00	45.74	O
ATOM	20	CB	LYS	A	5	12.399	-0.681	13.198	1.00	41.30	C
ATOM	21	CG	LYS	A	5	11.236	-0.268	12.361	1.00	61.61	C
ATOM	22	CD	LYS	A	5	11.427	-0.757	10.930	1.00	66.72	C
ATOM	23	CE	LYS	A	5	10.112	-0.760	10.137	1.00	90.19	C
ATOM	24	NZ	LYS	A	5	10.059	-1.825	9.080	1.00	69.42	N
ATOM	25	N	ASP	A	6	14.102	-0.973	15.597	1.00	38.66	N



# Визуализация с PyMol



## PyMol



# Для чего нужен PyMol

- Визуализация pdb и прочих файлов с координатами атомов
- Изготовление высококачественных изображений
- Начальное редактирование структур



# Системные требования

**Компьютер:** чем мощнее процессор и чем больше памяти, тем лучше

**3D монитор** не обязателен, но поддерживается

**Операционная система:** любая, под Linux проще установить, и он лучше работает с памятью.





# Как установить?

- Компиляция из исходников: <http://pymol.svn.sourceforge.net/>
- Установка бинарных пакетов в Ubuntu Linux: `sudo apt-get install pymol`
- Установка бинарных пакетов в Windows:
  - Ресурс для установки с python:  
<http://www.lfd.uci.edu/~gohlke/pythonlibs/#pymol>
  - Компиляция под Windows:  
<http://arcib.dowling.edu/~darakevn/installerpaper.pdf>



# PyMol - это GPL программа?

Да, PyMol это GPL-программа;

- исходный код доступен на [sourceforge.net](http://sourceforge.net)
- Бинарные пакеты для windows стоят денег и продаются:  
<http://pymol.org/academic.html>
- Бинарные пакеты для Linux собираются майтенерами

