Компьютерные методы в фармакологии

В.Б.Сулимов

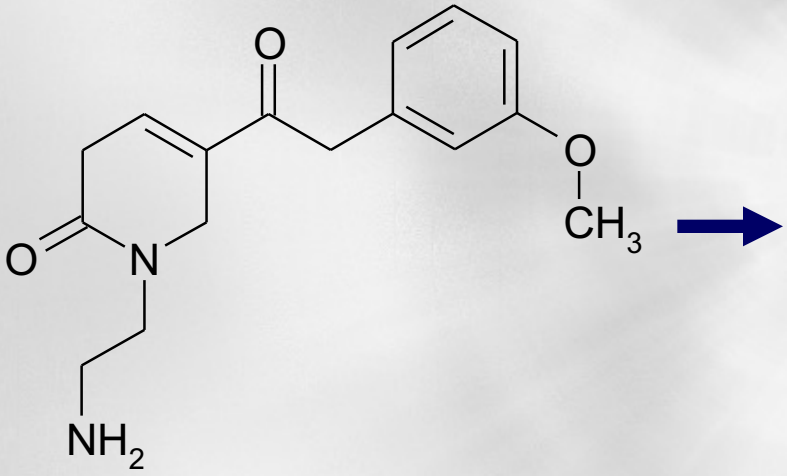
НИВЦ МГУ

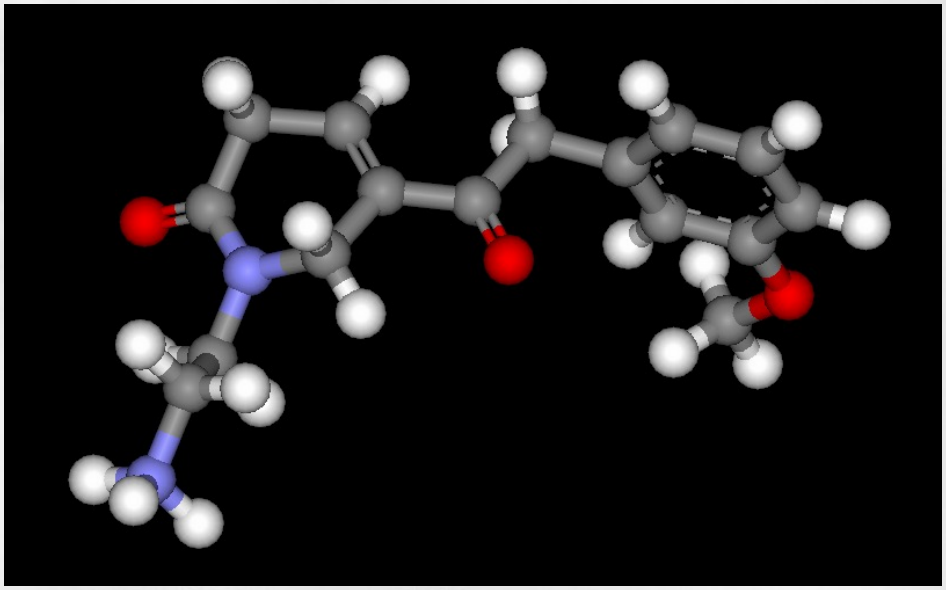
Лекция № 7

Молекулы, молекулярные группы, белки

Программы построения трехмерных структур молекул

Они способны преобразовывать «двумерные» структуры молекул в «трехмерные» (2D->3D conversion).

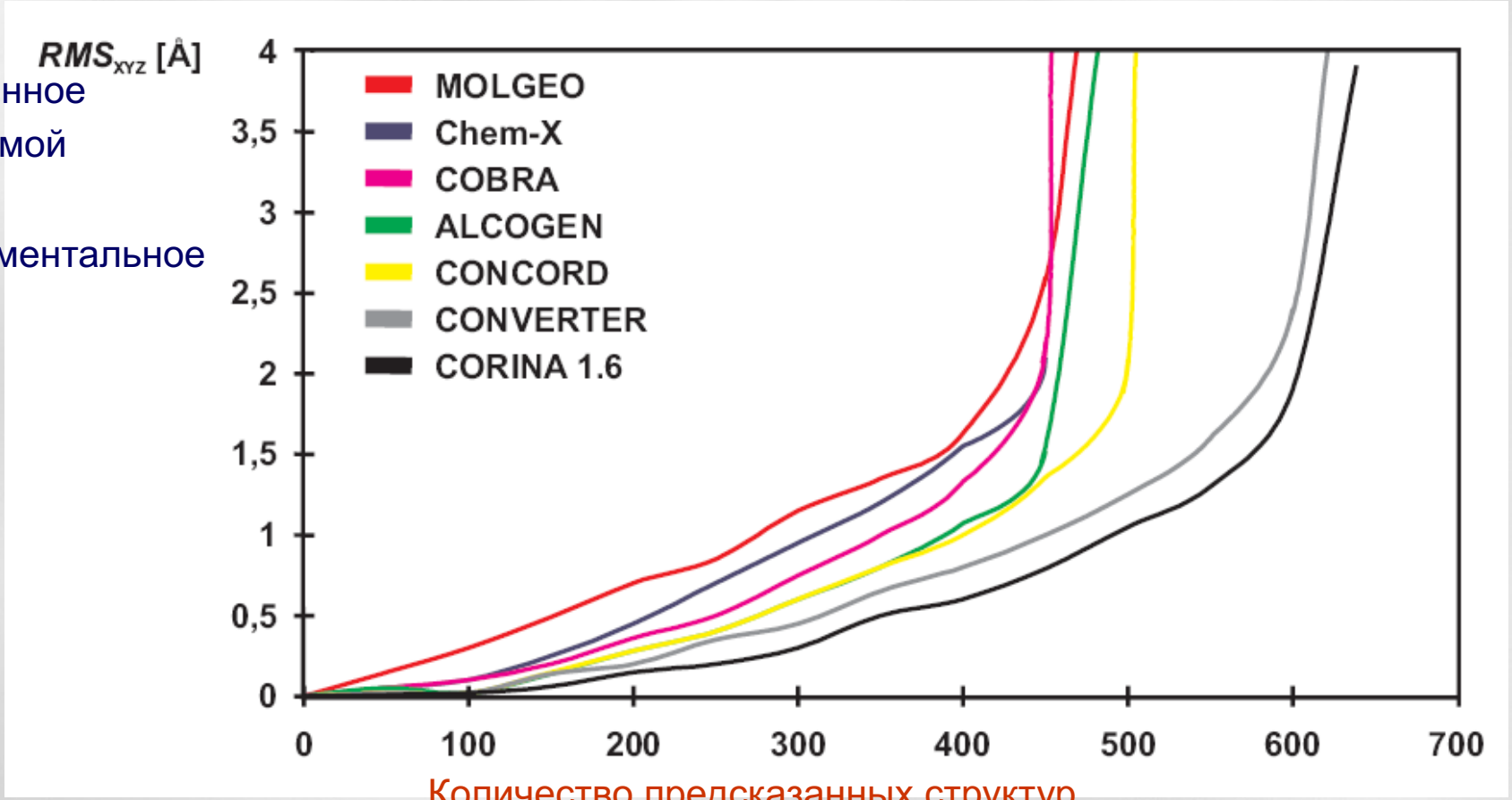




Такие программы работают путем сопоставления заданной им структуры со структурами, имеющимися у них в базе данных или на основе применения закономерностей строения молекул, полученных при анализе экспериментального материала

Сравнение «качества» работы различных программ 2D->3D преобразования (по материалам фирмы «Molecular Networks»)

Построенное программой vs. экспериментальное



Поскольку дизайн лекарственных средств – коммерческое направление, программы 2D->3D преобразования распространяются на платной основе.

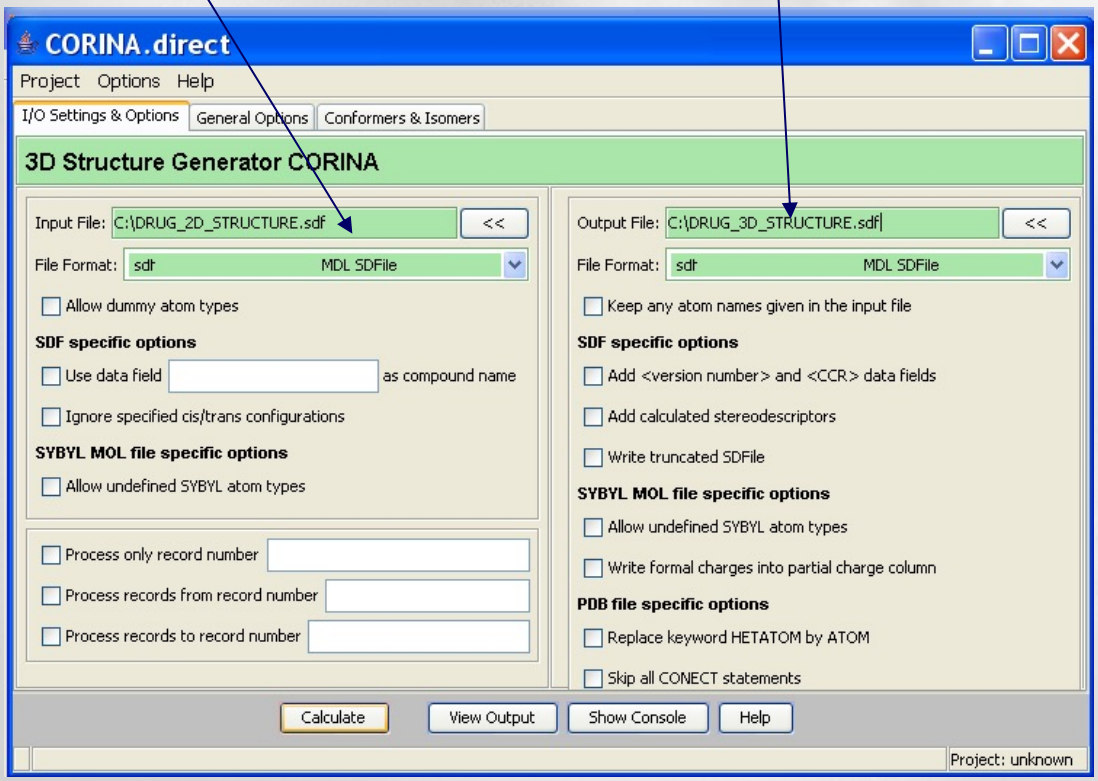
1. CORINA - http://www.molecular-networks.com/software/category/gen3dcoord.html (Molecular Networks GmbH)
2. CONCORD - http://www.tripos.com (Tripos)
3. ALCOGEN (Chemical Concepts)
4. MOLGEO - http://www.openmolgrid.org/
5. COBRA 3D Model Builder - http://www.oxmol.com/ (Oxford Molecular)
6. Chem-X - http://dtp.nci.nih.gov/docs/3d_database/chemx/chemx.html (Chemical Design Ltd.)
7. LigPrep - https://www.schrodinger.com/platform/drug-discovery

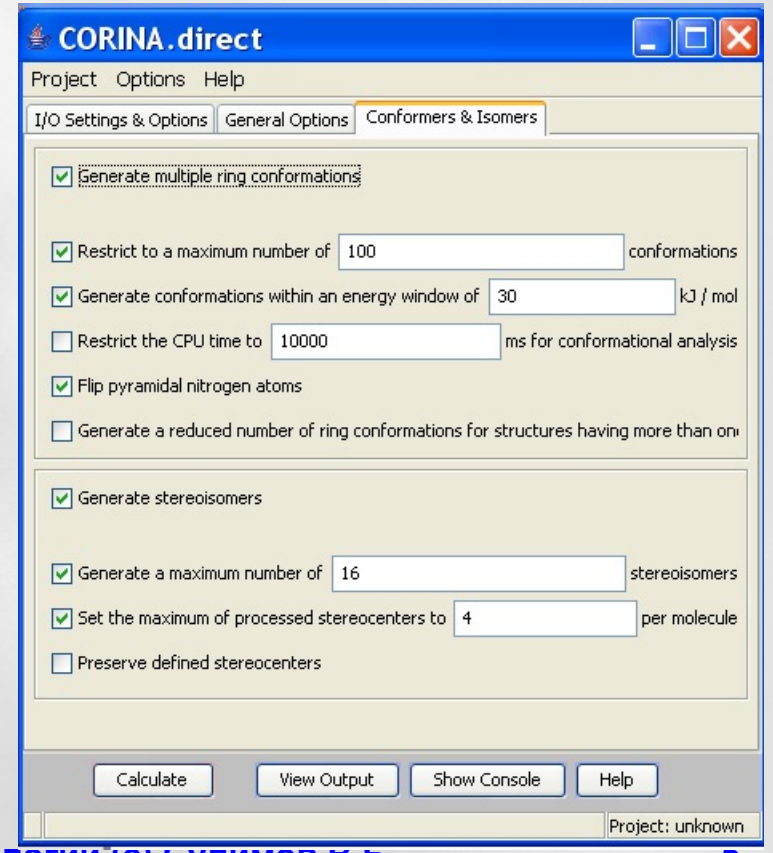
Пример интерфейса программы CORINA:

Имя структурного файла, Содержащего 2-D структуру (входной файл)

Имя структурного файла, содержащего 3-D структуру (выходной файл)

Дополнительные возможности программы CORINA – генерация множественных конформеров и стереоизомеров.





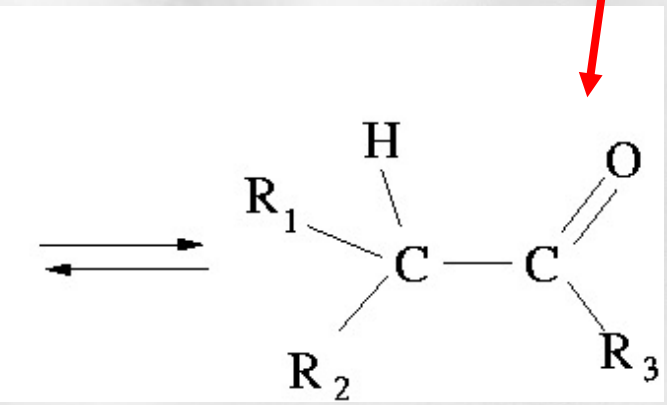
Таутомерия – обратимая изомерия. Перемещение атома водорода от одного атома в молекуле к другому

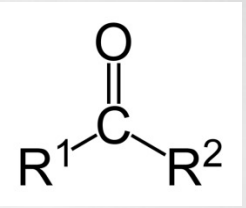
Обратимый переход из енол- в кето- форму:

Гидроксильная группа связана с атомом углерода, участвующем в двойной связи

Енол

Кетон

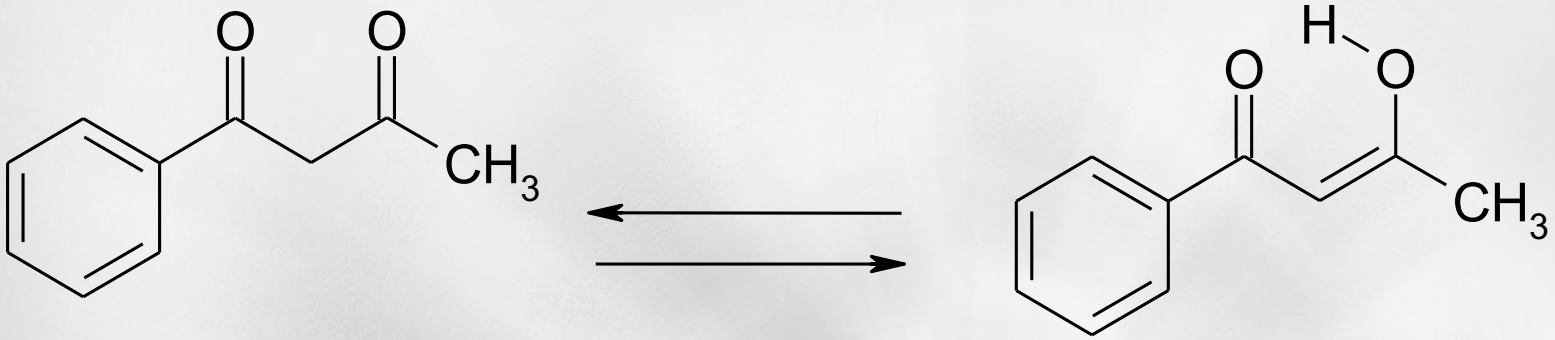


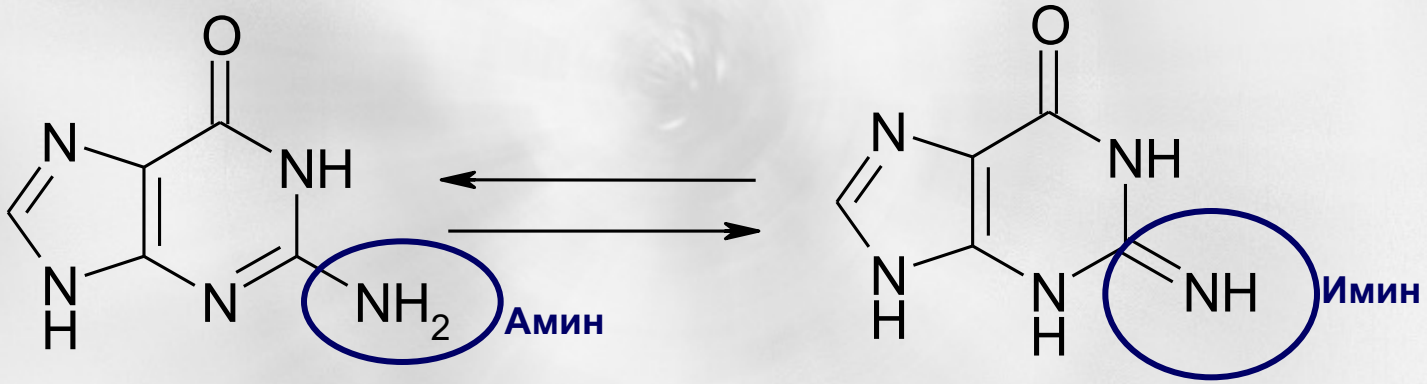


Многие вещества существуют в виде смесей таутомеров – различных соединений с одинаковым составом, но разными свойствами, способных легко превращаться друг в друга.

Таутомерия

Кето – енольная таутомерия





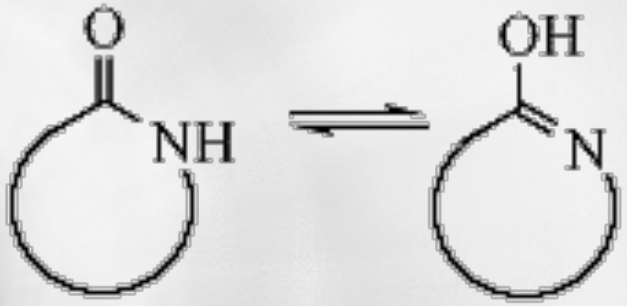
Амино – иминная таутомерия

Часто преобладание одного или другого таутомера зависит от агрегатного состояния вещества или от полярности растворителя!

Таутомерия

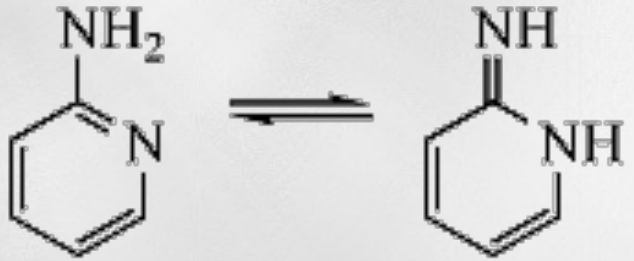
Enol form ⇌ Keto form

Lactam form ⇌ Lactim form

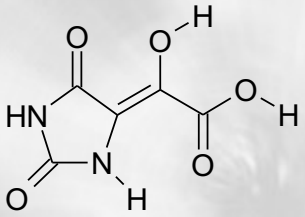
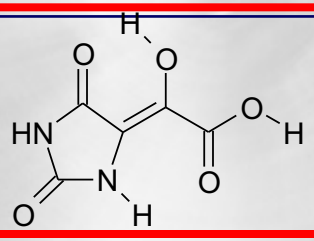
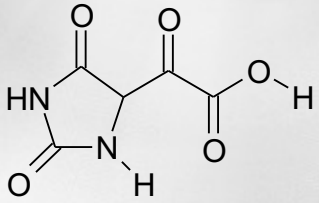


Amide form ⇌ Imidic acid form

Amine form ⇌ Imine form



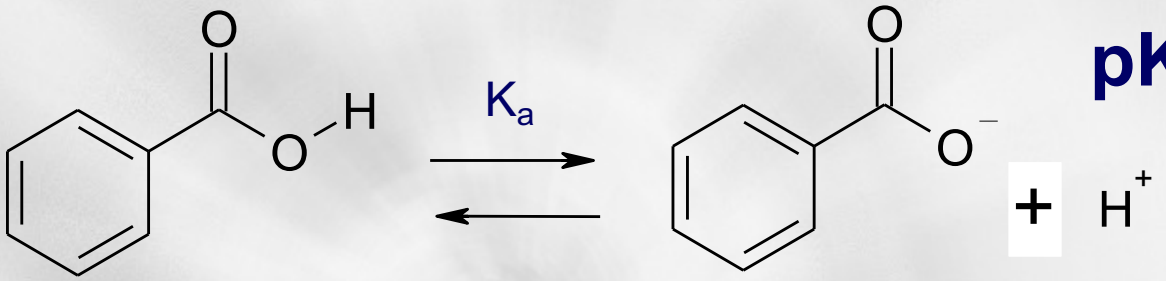
Пример кето-енольной таутомерии. Расчет энергии отдельных таутомеров

Таутомер	Структурная формула	Относительные энергии ккал/моль	
		MP2/cc-pVTZ Вакуум	**B3LYP/6-31G** Вода**
енольный таутомер 1		**6.45**	**4.31**
енольный таутомер 2		**0**	**0**
кетонный таутомер		**9.17**	**8.62**

Зарядовые состояния органических молекул

1. Многие органические вещества и лекарства в частности представляют собой сильные или слабые органические кислоты и основания. Их зарядовое состояние зависит от кислотности (pH) среды.
2. Зарядовые состояния молекул лекарственных веществ играют важнейшую роль при их связывании с местом приложения их действия (ферменты, рецепторы)
3. Зарядовое состояние молекулы играет определяющую роль в процессах ADME, так как только нейтральная форма молекулы способна проходить через неполярную фазу липидной мембраны клетки.
4. Для построения модели в правильном зарядовом состоянии нужно знать константы диссоциации функциональных групп, входящих в состав молекулы и предполагаемую кислотность среды в месте действия лекарственного вещества

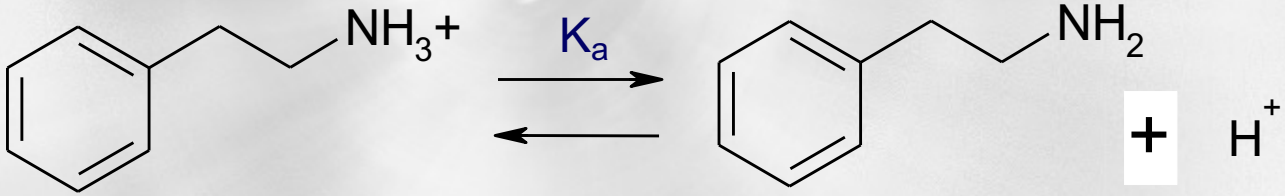
Равновесие между заряженной и нейтральной формами вещества определяется константой диссоциации K_a



pKa < 7

Нейтральная форма кислоты

Ионизированная кислота



pKa > 7

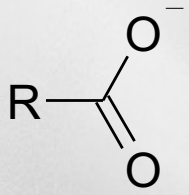
Ионизированное основание

Нейтральная форма основания

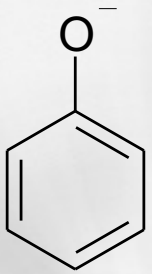
$$K_a = \frac{[H][A]}{[HA]}$$

$$pK_a = -\log_{10} K_a$$

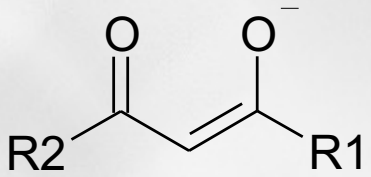
Функциональные группы, несущие **отрицательный** заряд при физиологических условиях (pH=7)



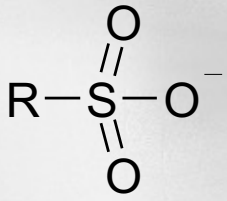
карбоновые кислоты



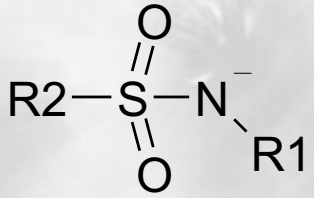
фенолы



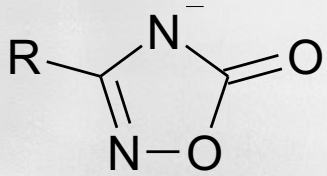
еноляты



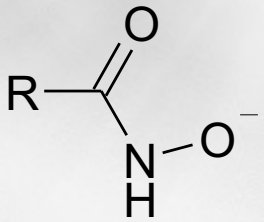
сульфокислоты



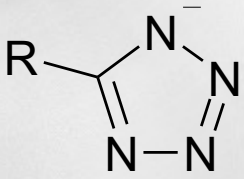
сульфамиды



производные оксадиазинона

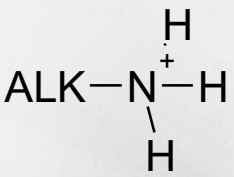


гидроксамовые кислоты

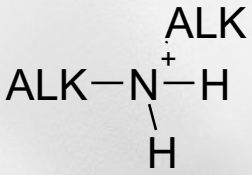


производные тетразола

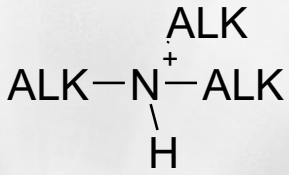
Функциональные группы, несущие **положительный** заряд при физиологических условиях (pH=7)



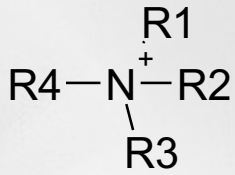
первичные алкиламины



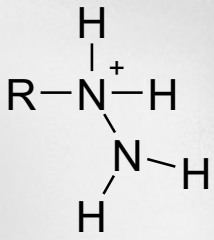
вторичные алкиламины



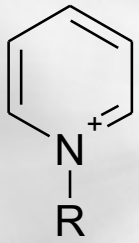
третичные алкиламины



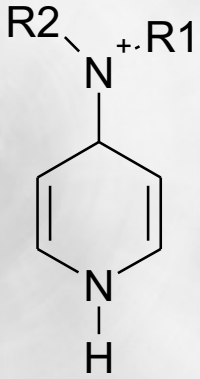
четвертичные азотные основания



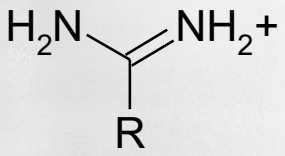
гидразины



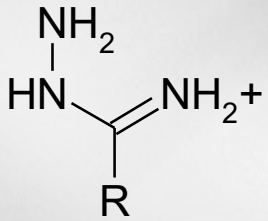
кватернизованые азосодержащие гетероциклы



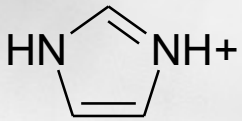
4-аминопиридины



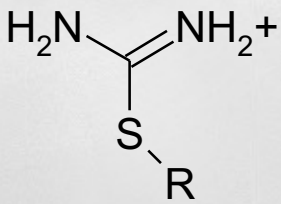
амидины



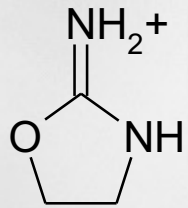
амидразоны



имидазол и его производные



соли тиоурония



2-аминооксазолин и его производные

Как в общем случае узнать, заряжена данная группа или нет?

1. Найти эту группу в составе похожих соединений в базе данных – например NIST Standard Reference Database (**www.nist.gov**)
2. Можно также воспользоваться многочисленными компиляциями данных по pKa – например **research.chem.psu.edu/brpgroup/pKa_compilation.pdf**
3. Если pKa основной группы ***превышает 7***, то вещество в физиологических условиях находится преимущественно в виде положительно заряженного протонированного иона (так, если pKa=8, то 90% вещества ионизировано, если pKa=9 – ионизировано 99%)
4. Если pKa кислотной группы ***меньше 7***, то кислота при физиологических условиях депротонирована – т.е. существует в виде отрицательно заряженного иона (если pKa=6, то 90% вещества ионизировано, если pKa=5 – ионизировано 99%)

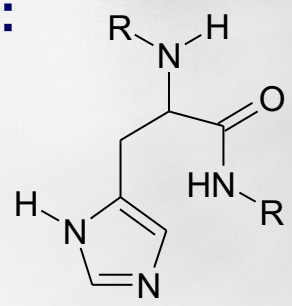
Легкие атомы водорода не рассеивают рентгеновское излучение, большинство структур белков содержит только координаты тяжелых атомов (C, N, O, S, P и др.)

Для расстановки атомов водорода используются пакеты: Sybil, Hyperchem и др.

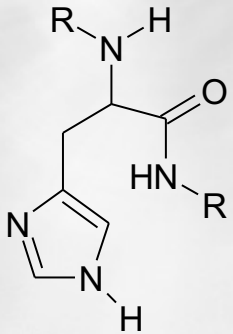
- специальные программы: **Aplite,** Babel, REDUCE, Marvin, Avogadro
- Протонирование с помощью Avogadro:
  Build > Add Hydrogens for pH... > 7.4 > OK

Протонированный лиганд сохраняется командой «Save as»

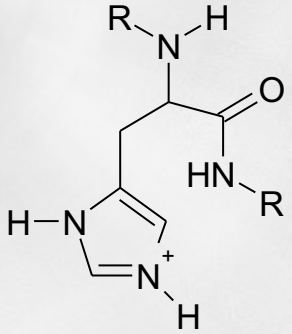
Возможны различные варианты расстановки атомов водорода, например:



D-протонирование нейтральная форма



E-протонирование нейтральная форма

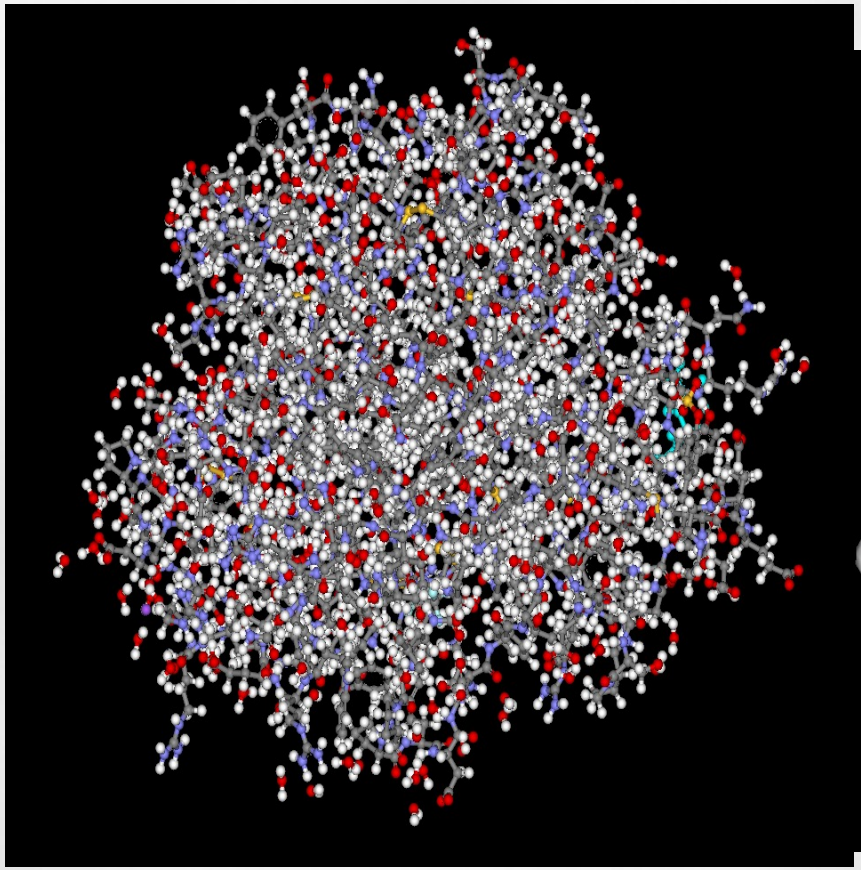


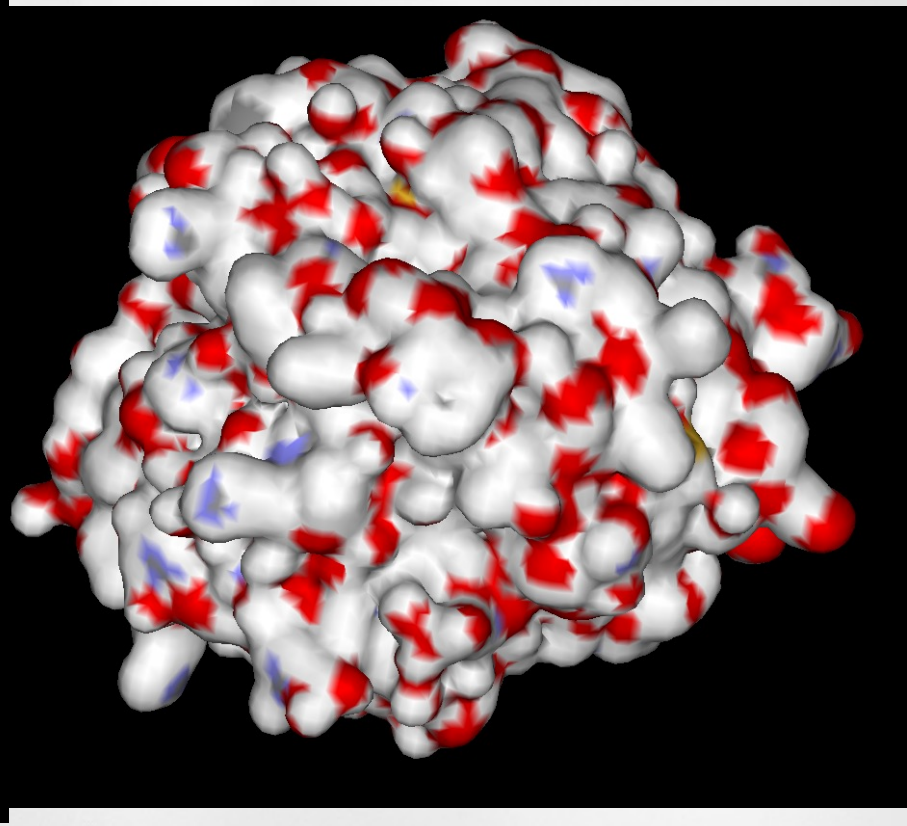
заряженная форма

Даже в одном белке в разных местах могут встречаться по-разному протонированные **гистидины.** Такое различие важно с точки зрения функционирования активного центра белка.

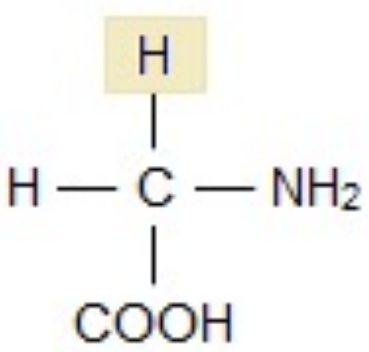
Расстановку водородов нужно выполнять тщательно и вдумчиво, анализируя окружение «проблемных» аминокислот, таких как гистидин.

Структура белка из Protein Data Bank – тромбин, файл 1AD8

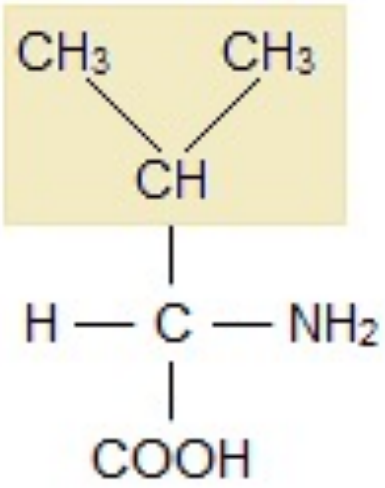




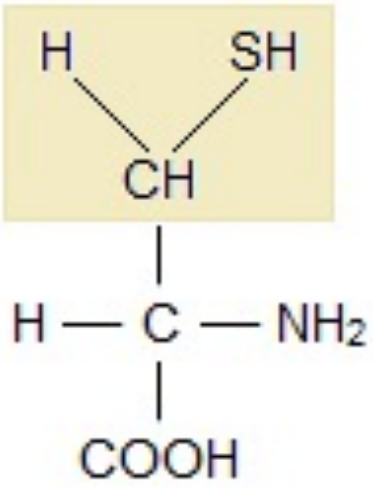
Аминокислоты: 22 шт



Глицин (гли)



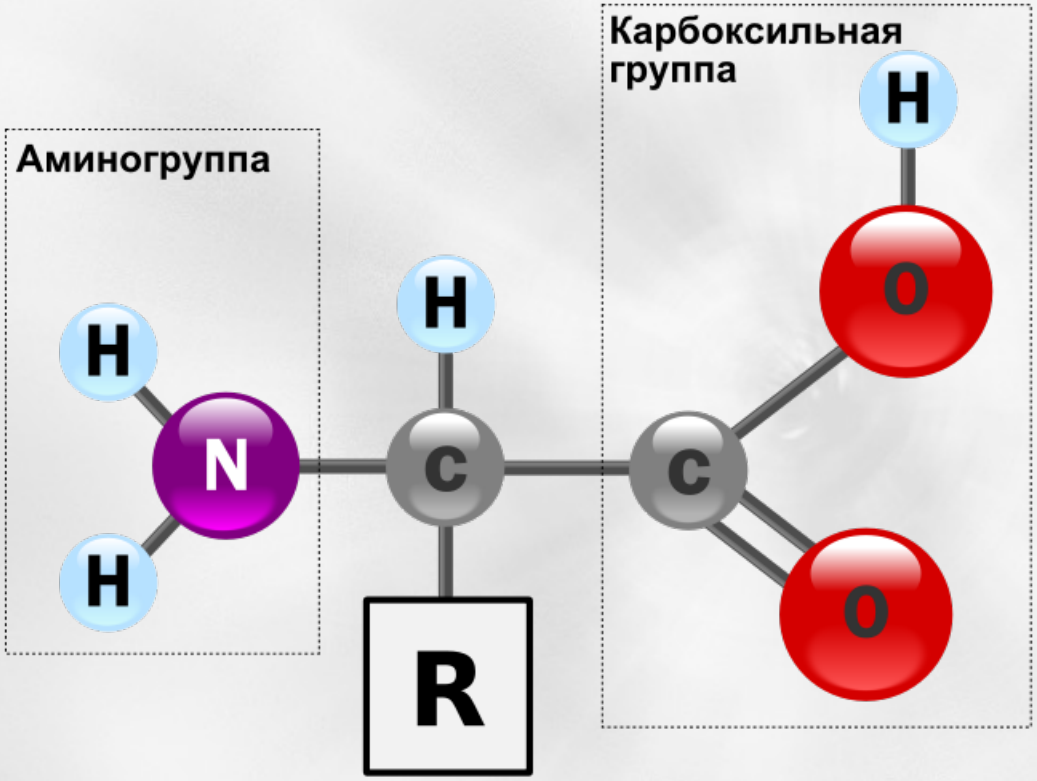
Валин (вал)

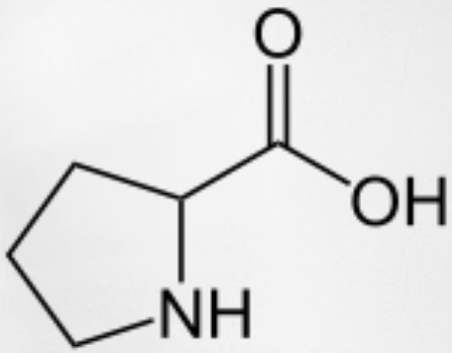


Цистеин (цис)

Рис. 1. Примеры строения аминокислот — мономеров белковых молекул (цветом обозначен радикал)

Структура аминокислоты



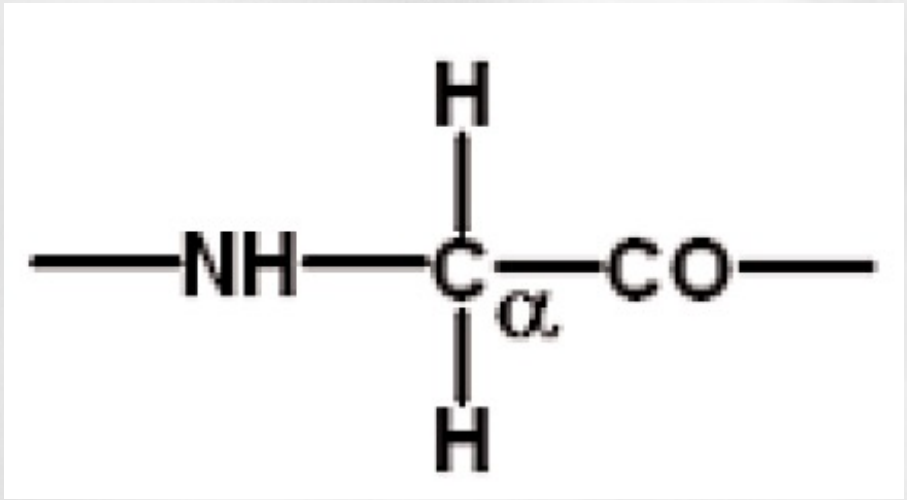


Пролин

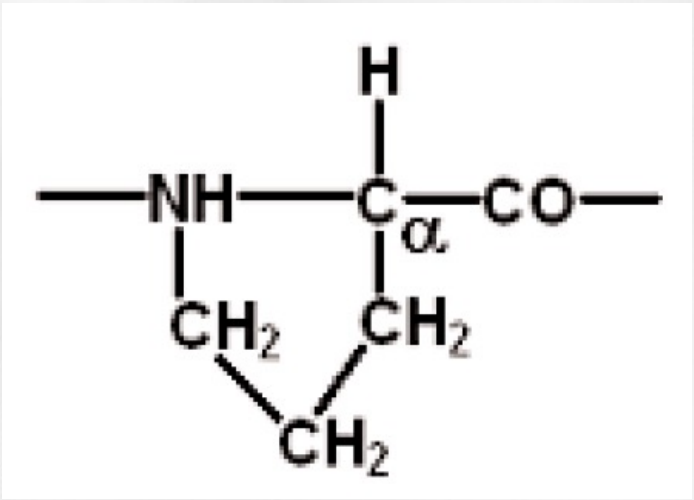
Для разных аминокислот – разные радикалы **R**

Два аминокислотных остатка, глицин, Gli и пролин Pro, играют особую роль для различных трехмерных структур белков

Глицин



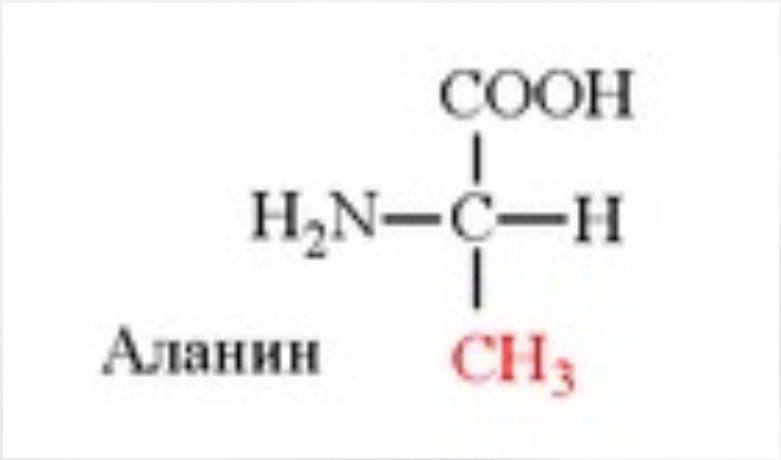
Пролин

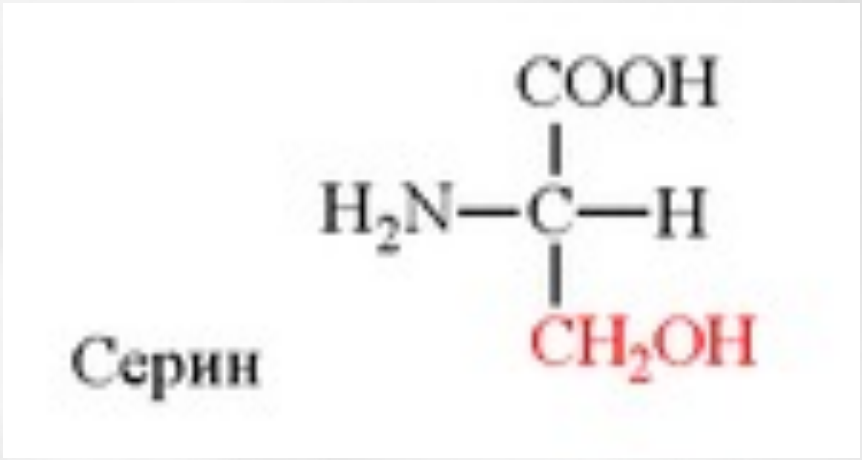


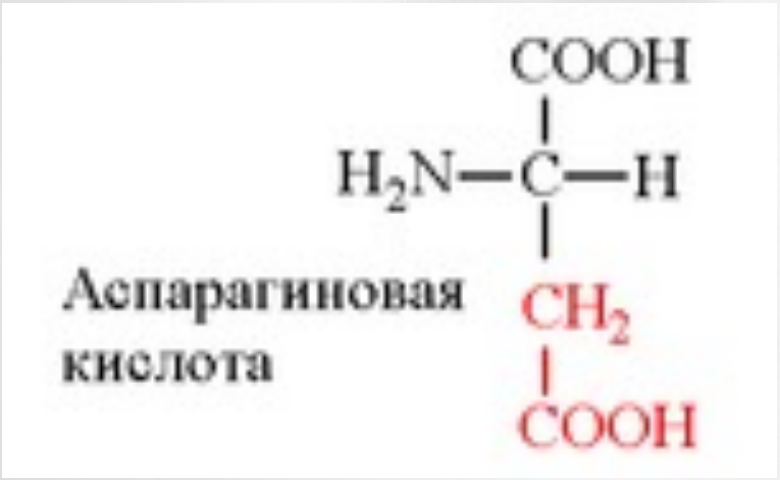
Гибкий шарнир полипептида

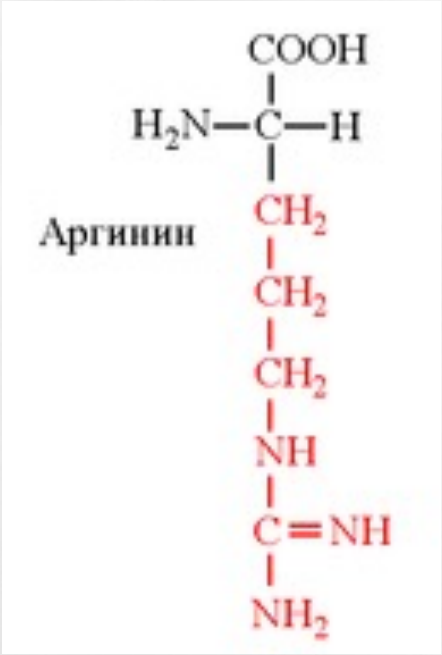
Жесткое соединение

Аминокислоты – 20 шт

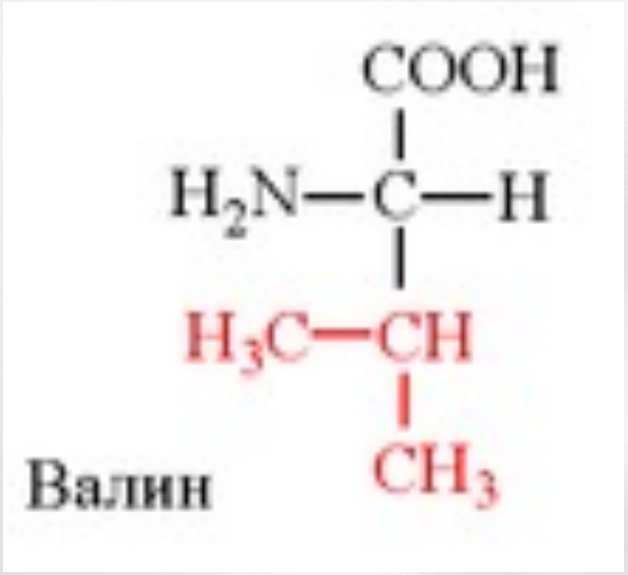


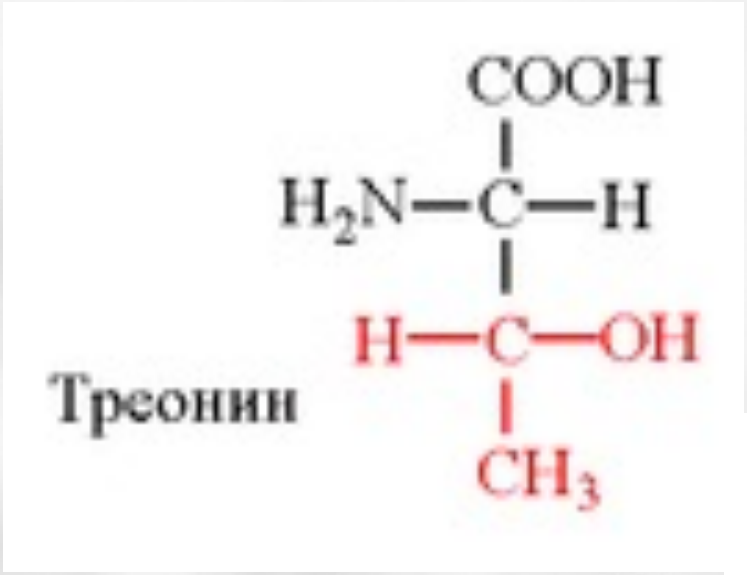


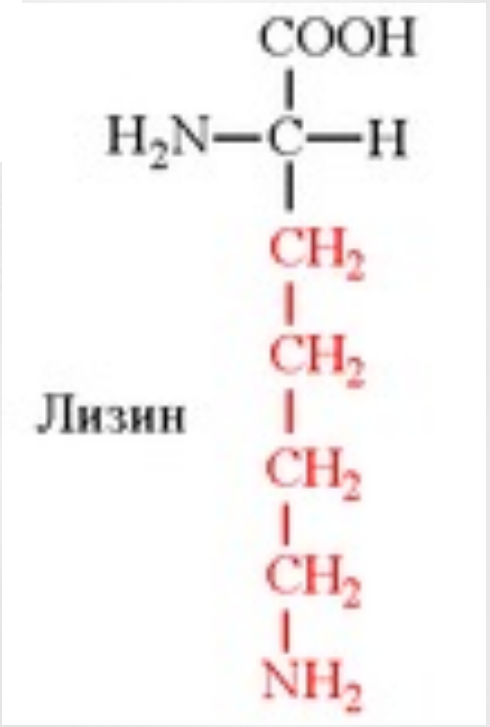


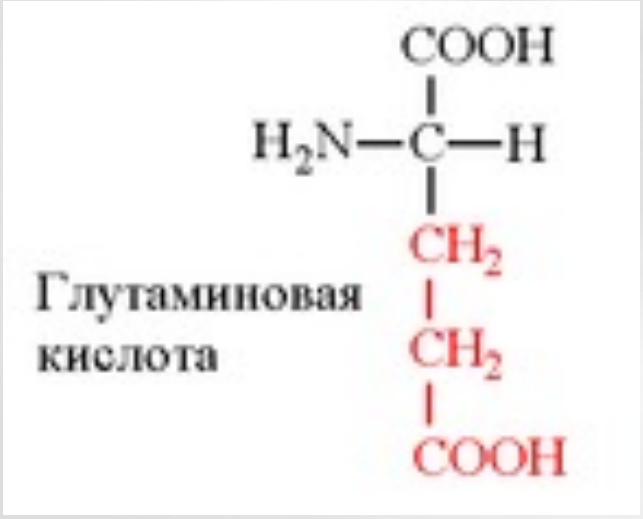


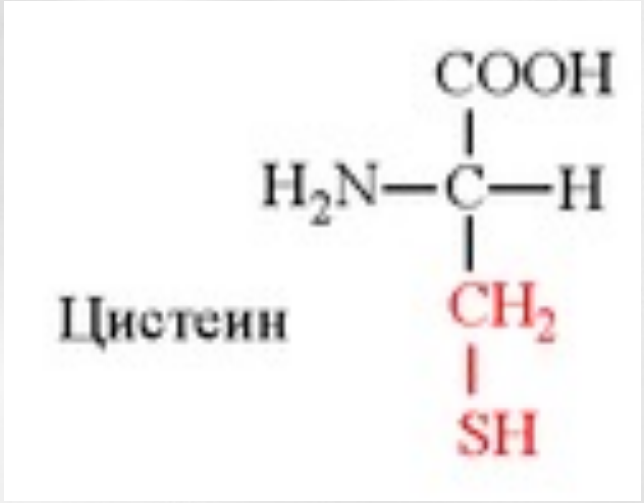
Аминокислоты – 20 шт



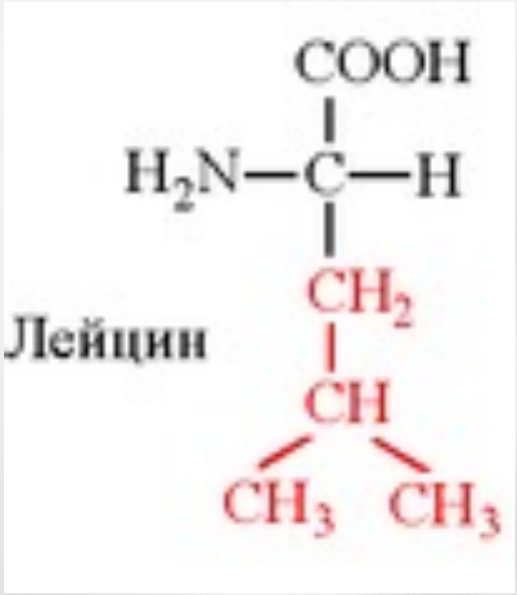




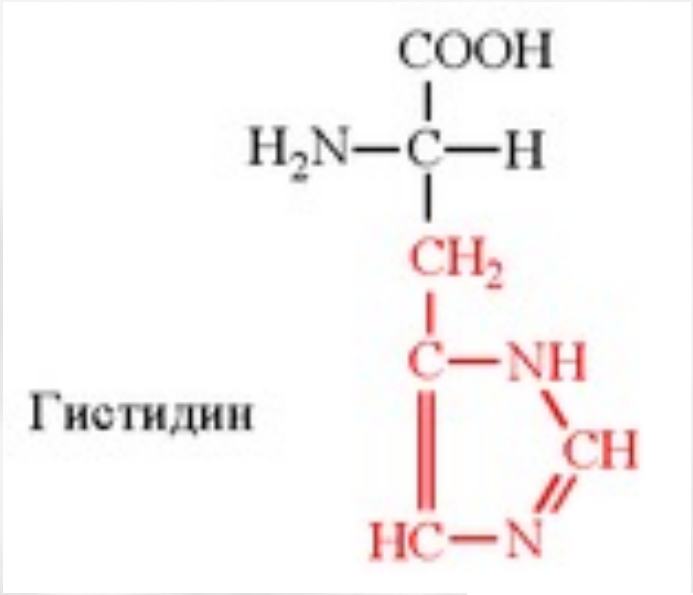




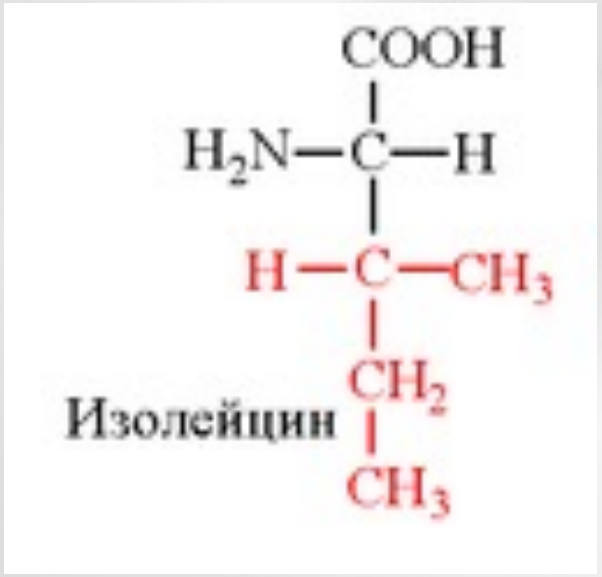
Аминокислоты – 20 шт



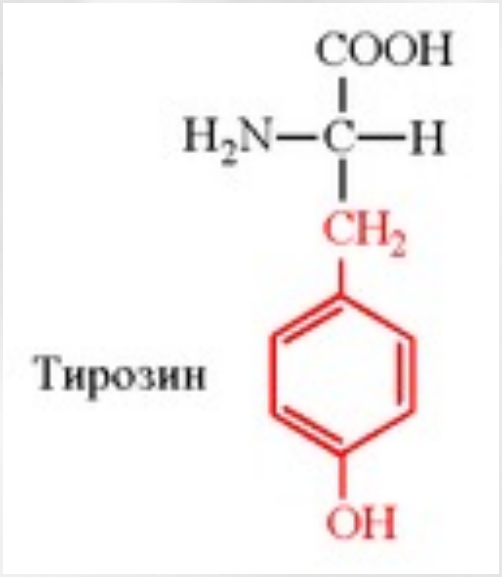
Лейцин



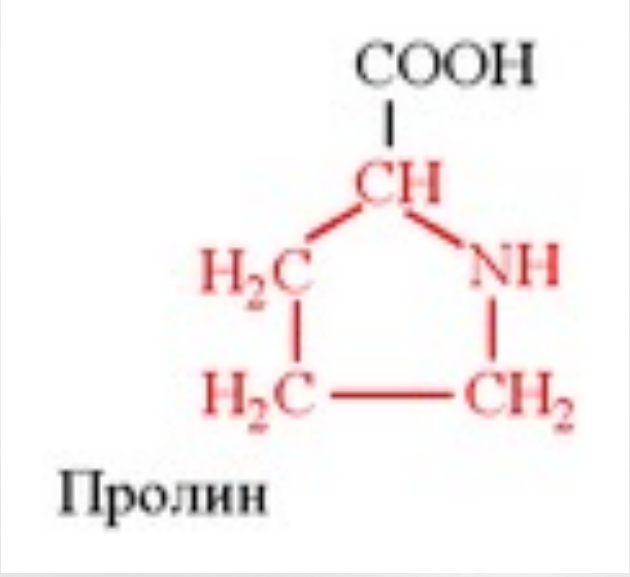
Гистидин



Изолейцин

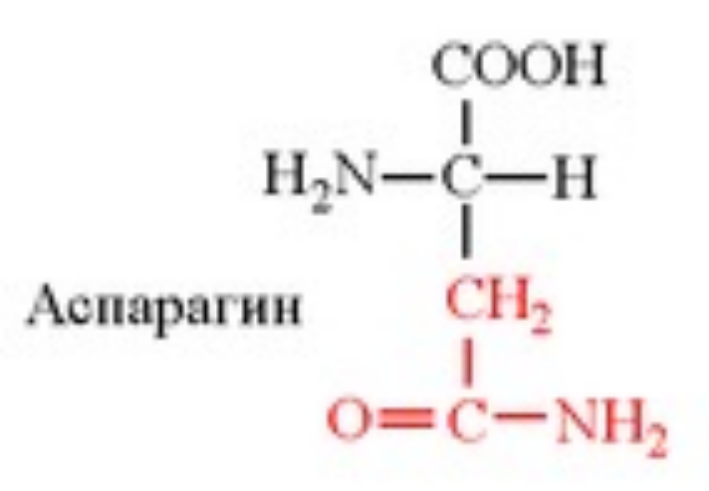


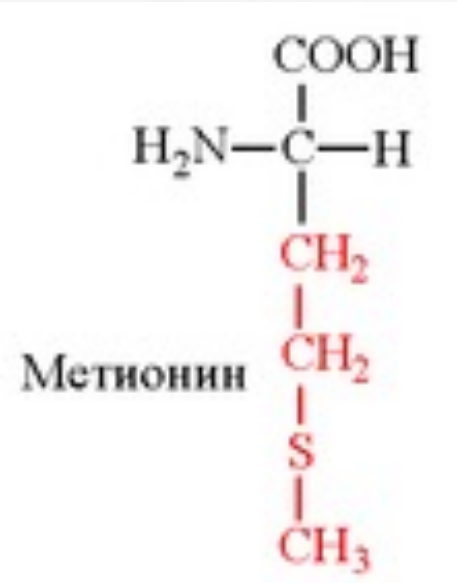
Тирозин

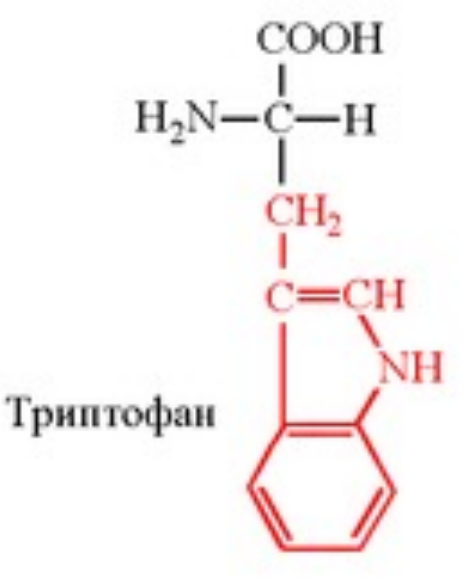


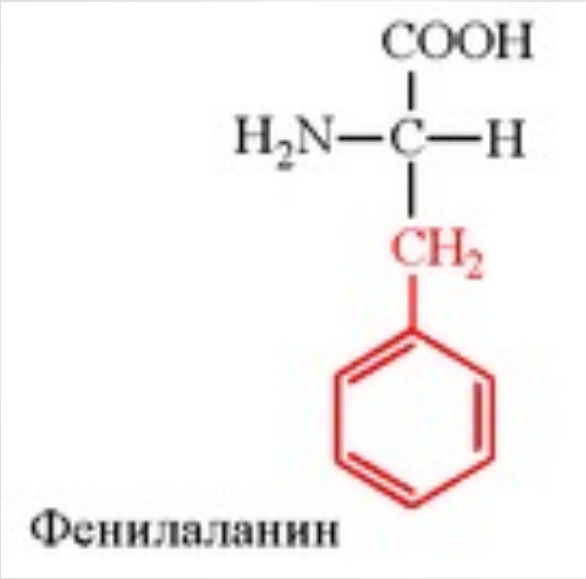
Пролин

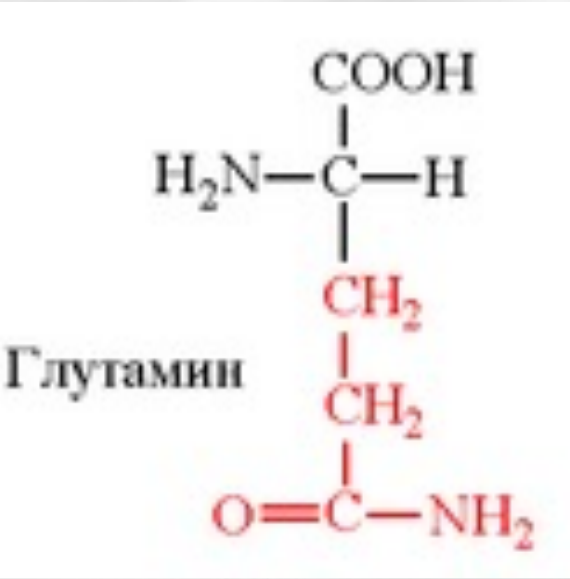
Аминокислоты – 20 шт

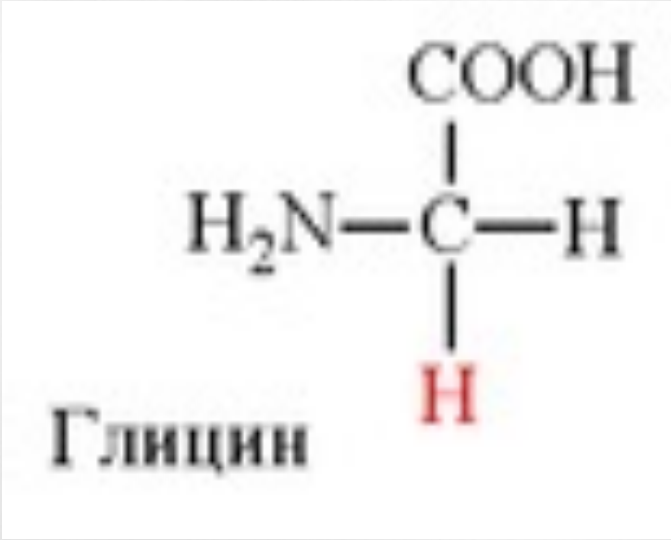




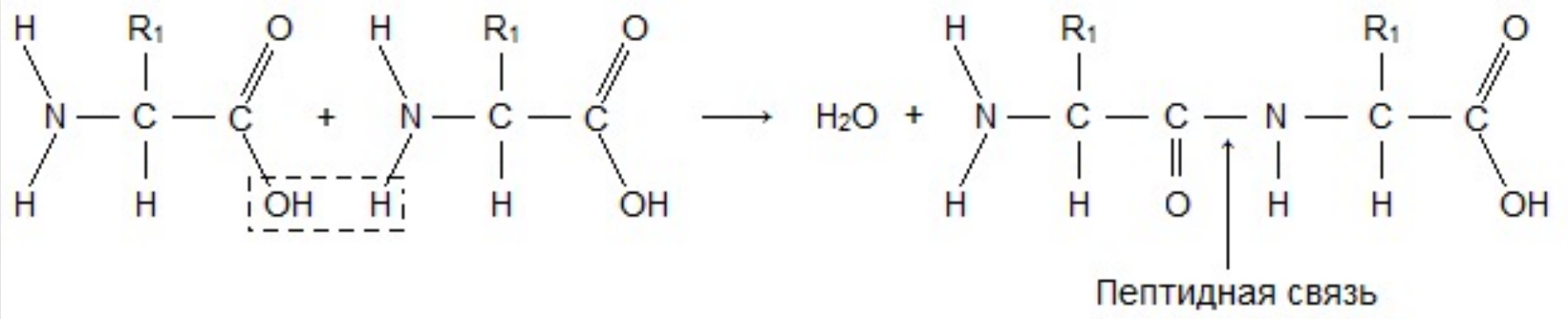




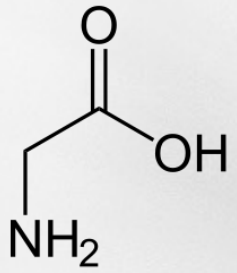




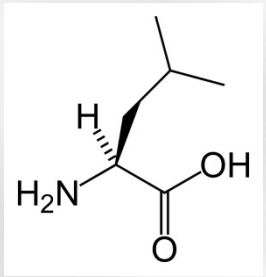
Аминокислоты вступают в химическую связь друг с другом



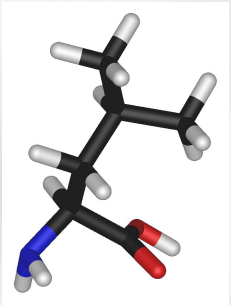
Аминокислоты

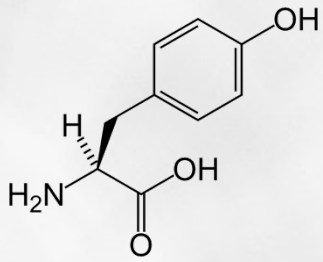


Глицин

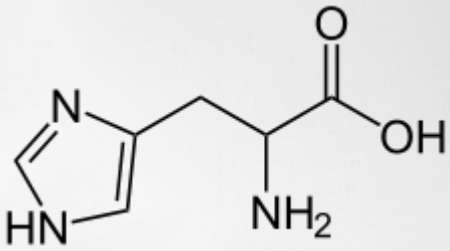


Лейцин

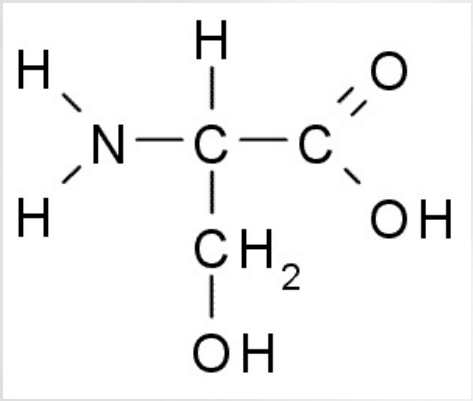




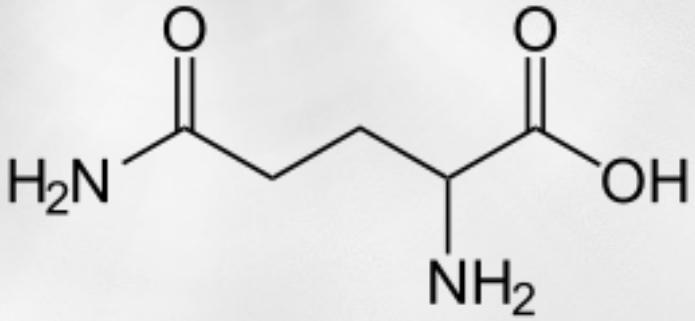
Тирозин



Гистидин



Серин

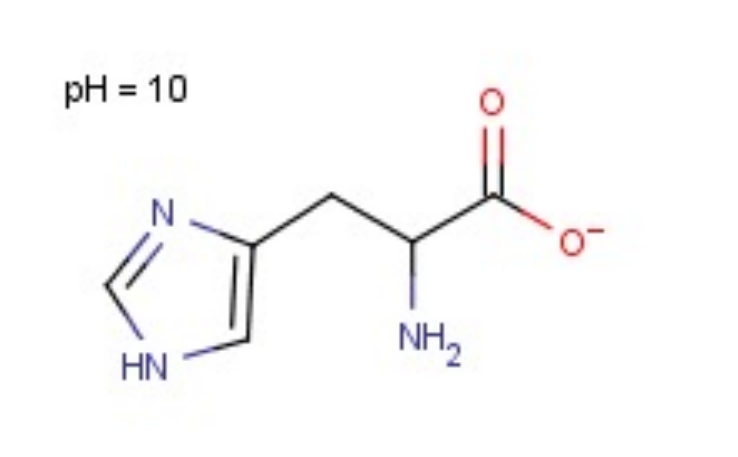


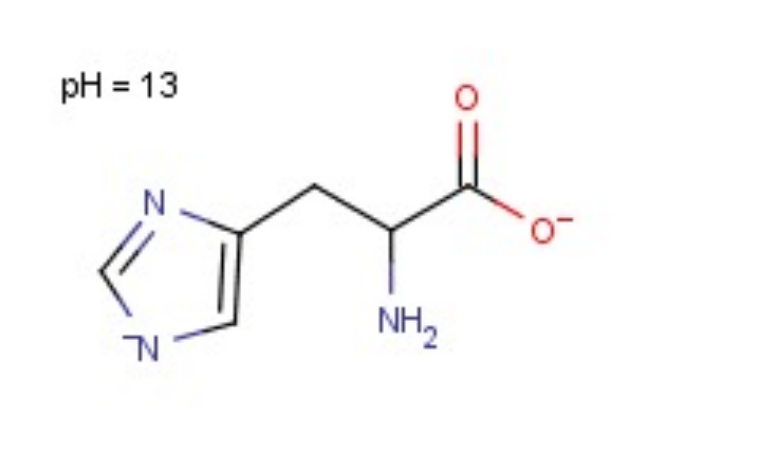
Глютамин

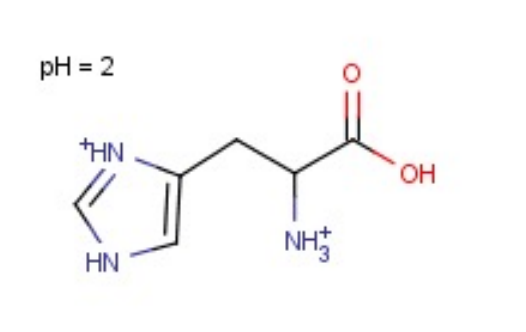
+ …

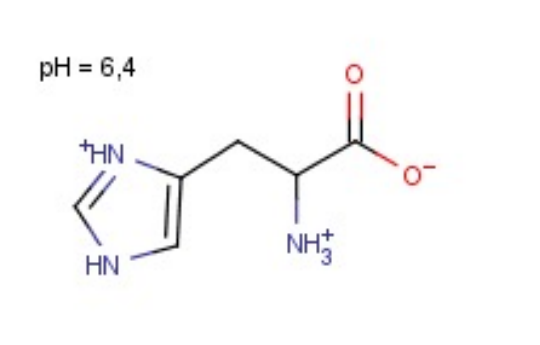
Различные протонированные состояния гистидина

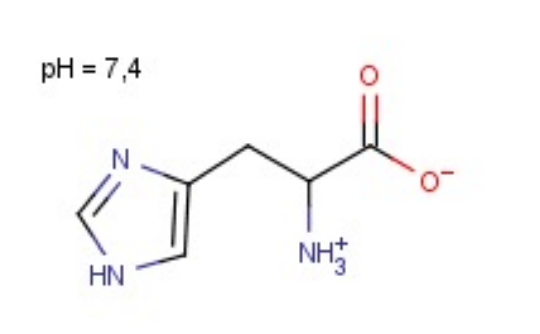
Щелочная среда











Кислая среда

Нейтральная среда

Форматы данных структурных файлов белков (pdb формат)

ATOM	20	CA	CYS	L	1	10.473	20.329	19.879	1.00	28.47
ATOM	21	C	CYS	L	1	9.336	20.079	20.788	1.00	22.76
ATOM	22	O	CYS	L	1	8.851	20.826	21.635	1.00	23.32
ATOM	23	CB	CYS	L	1	11.808	19.997	20.426	1.00	29.45
ATOM	24	SG	CYS	L	1	12.206	20.492	22.136	1.00	24.97
ATOM	25	N	GLY	L	2	8.864	18.862	20.569	1.00	18.95
ATOM	26	CA	GLY	L	2	7.827	18.338	21.429	1.00	17.85
ATOM	27	C	GLY	L	2	6.429	18.881	21.294	1.00	19.54
ATOM	28	O	GLY	L	2	5.543	18.370	21.980	1.00	20.95
ATOM	29	N	LEU	L	3	6.225	19.915	20.468	1.00	22.46
ATOM	30	CA	LEU	L	3	4.894	20.498	20.233	1.00	19.35
ATOM	31	C	LEU	L	3	4.411	20.041	18.851	1.00	22.15
ATOM	32	O	LEU	L	3	4.921	20.485	17.823	1.00	23.57
ATOM	33	CB	LEU	L	3	4.965	22.017	20.298	1.00	21.80
ATOM	34	CG	LEU	L	3	5.355	22.546	21.677	1.00	26.59
ATOM	35	CD1	LEU	L	3	5.592	24.042	21.604	1.00	30.38
ATOM	36	CD2	LEU	L	3	4.282	22.196	22.703	1.00	18.88
ATOM	37	N	ARG	L	4	3.471	19.102	18.838	1.00	21.39
ATOM	38	CA	ARG	L	4	2.967	18.520	17.587	1.00	22.20

Форматы данных структурных файлов малых молекул (mol, sdf),

Формат данных **mol** разработан фирмой MDL Information Systems Inc. Один из широко используемых на сегодняшний день. Формат **sdf** идентичен формату mol, но применяется для записи баз данных органических веществ и представляет собой записи в формате mol, идущие друг за другом.

В каждой строчке записаны координаты атома и коды, указывающие на его зарядовое состояние, стереохимию и т.д.

```
-BENZENE-   11120720222D

 12 12  0  0  0  0  0  0  0  0999 V2000
    0.9319   -21.3416    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.9307   -22.1690    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    1.6456   -22.5819    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    2.3620   -22.1685    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    2.3591   -21.3380    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    1.6438   -20.9289    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    1.6375   -20.1000    0.0000 H   0  0  0  0  0  0  0  0  0  0  0  0
    3.0708   -20.9208    0.0000 H   0  0  0  0  0  0  0  0  0  0  0  0
    3.0708   -22.5792    0.0000 H   0  0  0  0  0  0  0  0  0  0  0  0
    1.6375   -23.4042    0.0000 H   0  0  0  0  0  0  0  0  0  0  0  0
    0.2125   -22.5750    0.0000 H   0  0  0  0  0  0  0  0  0  0  0  0
    0.2125   -20.9250    0.0000 H   0  0  0  0  0  0  0  0  0  0  0  0
  6  1  1  0  0  0  0
  1  2  2  0  0  0  0
  6  7  1  0  0  0  0
  3  4  2  0  0  0  0
  5  8  1  0  0  0  0
  4  9  1  0  0  0  0
  4  5  1  0  0  0  0
  3 10  1  0  0  0  0
  2  3  1  0  0  0  0
  2 11  1  0  0  0  0
  5  6  2  0  0  0  0
  1 12  1  0  0  0  0
M  END
```

x, y, z координаты

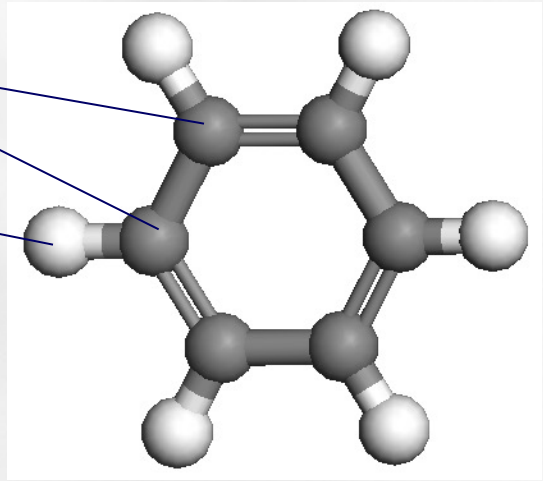


Таблица связности – какие атомы связаны друг с другом и каков порядок связи

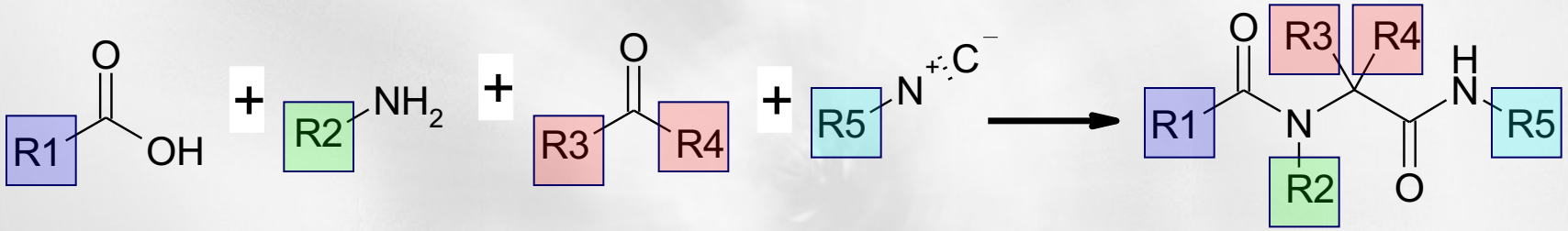
Другие форматы

1. mol2 формат – разработан фирмой Tripos и применяется в продуктах этой фирмы (Sybil)
2. hin формат – разработан фирмой Hypercube и широко используется пользователями программы HyperChem
3. mrk – формат разработан в фирме Merck, в частности, для работы с продуктами, использующими молекулярное моделирование в рамках силового поля MMFF 94 (CHARMM)
4. xyz – простейший формат, представляет собой перечисление названий атомов и их декартовых координат

Для преобразования структурных файлов из одного формата в другой можно применять специальные программы (Babel), а также пакеты молекулярного моделирования (Sybil, Hyperchem, CHARMM). Преобразования часто выполняются некорректно даже коммерческими продуктами, поэтому возникает потребность в написании собственных программ перевода форматов.

Понятие о виртуальных библиотеках химических соединений

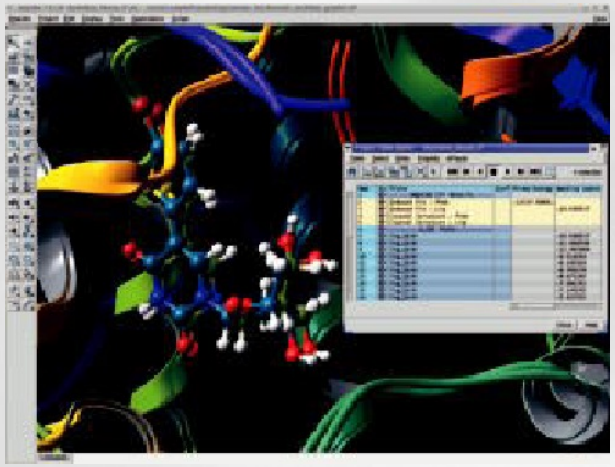
Реакции комбинаторного органического синтеза (на примере реакции Уги):



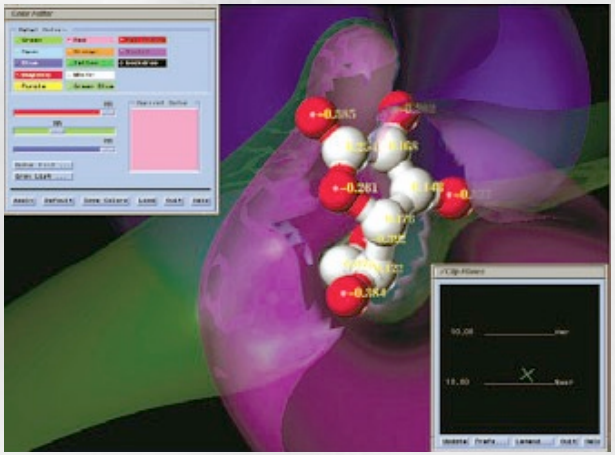
Таким образом, имея по 10 реагентов каждого типа в 4-х компонентной реакции можно получить до 10^4 различных соединений. Предположим, мы хотим предварительно исследовать методами вычислительной химии связывание этих соединений с молекулой белка. Тогда нам нужно получить файл, содержащий так называемую «виртуальную библиотеку» продуктов реакции Уги в количестве 10^4 соединений – т. е. файл, содержащий структурную информацию о всех молекулах, полученых по реакции Уги. Для записи виртуальных библиотек удобен файл формата sdf. Виртуальные библиотеки могут содержать не только структуры молекул, полученных методами комбинаторной химии. В них может заноситься информация, например о коллекции веществ, полученных в химической лаборатории, доступных в продаже, в принципе известных и т.д.

Основные программы построения молекул и их визуализации – рабочие инструменты «молекулярного дизайна»

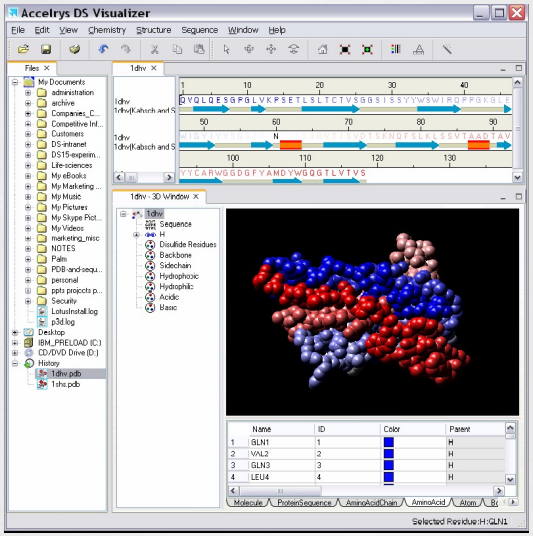
Фирмы поставляют программы визуализации и редактирования молекул в составе единых платформ, включающих и другие компоненты – программы квантовой химии, молекулярной динамики, молекулярной механики, докинга, построения виртуальных библиотек и т.д. В этом случае программы визуализации и редактирования служат универсальным интерфейсом для объединения и взаимодействия многочисленных программных компонент. Это платные, дорогие и качественные программные продукты, позволяющие осуществлять визуализацию и редактирование молекул, осуществлять чтение и запись большинства популярных молекулярных форматов. Из продуктов такого рода нужно отметить:



Среда для молекулярного моделирования “Maestro” фирмы Schrodinger. Доступна для платформ Linux и Windows http://www.schrodinger.com



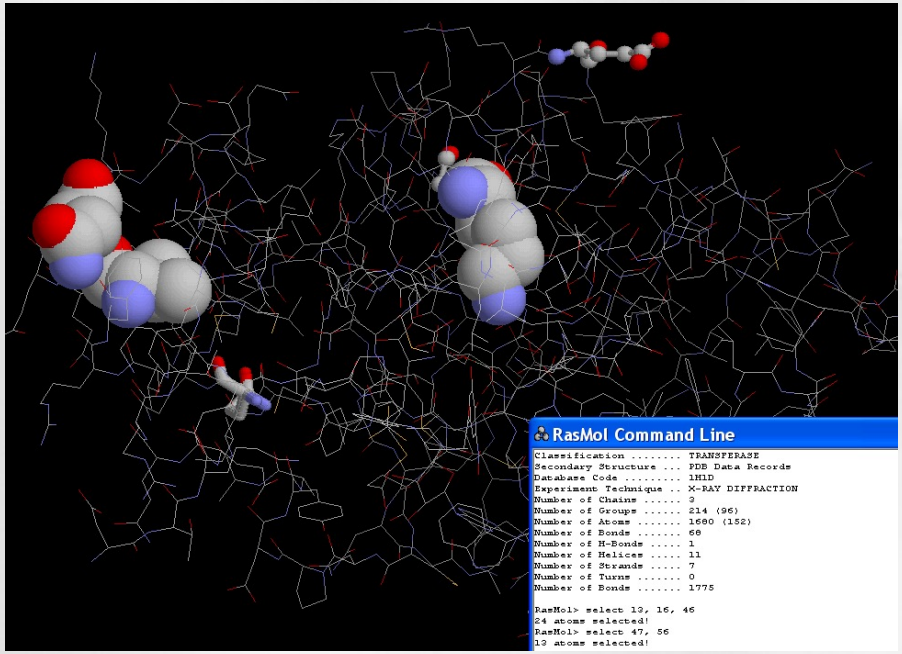
Средства визуализации и редактирования программного продукта “Sybil” фирмы Tripos. Необходимые платформы – Linux, UNIX http://www.tripos.com

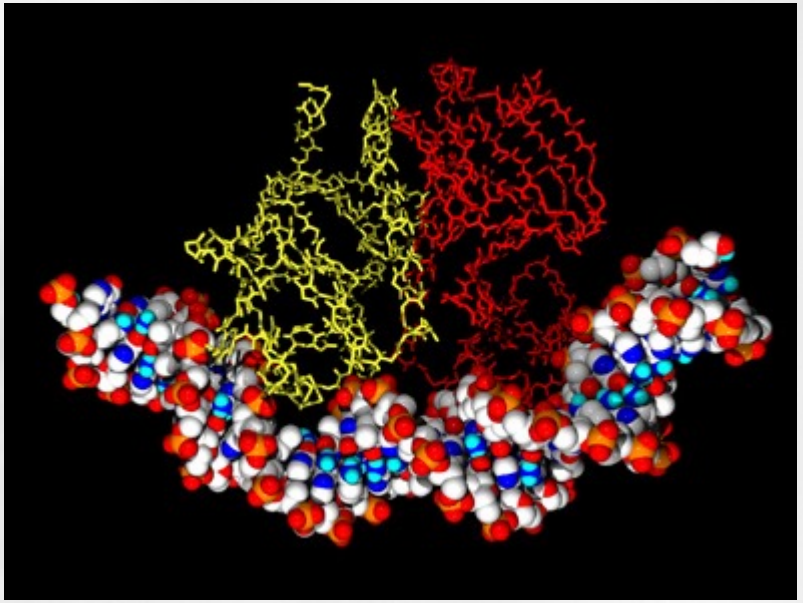


DS Visualizer фирмы Accelrys. Визуальный интерфейс для платформы Discovery Studio http://www.accelrys.com Вариант с ограниченной функциональностью доступен бесплатно

Простые визуализаторы белковых структур

Как правило имеют довольно ограниченную функциональность – позволяют просматривать файлы формата pdb и некоторых других форматов. Возможности редактирования отсутствуют или ограничены. Программы бесплатные. Ниже представлены два достаточно популярных визуализатора.

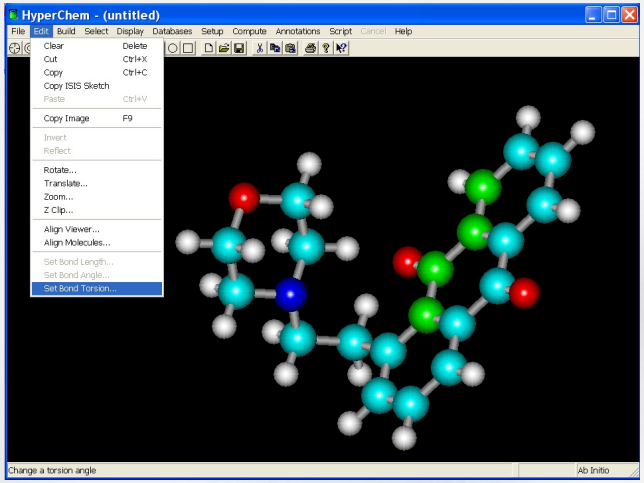


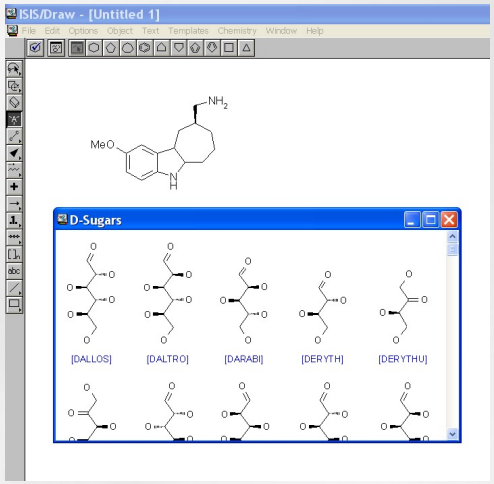


Визуализатор Rasmol (Linux) – Raswin (Windows) Удобен для просмотра, хороший пользовательский Интерфейс. Популярная, хотя и несколько устаревшая программа
http://www.umass.edu/microbio/rasmol

Swiss-PdbViewer – несколько более сложная программа, обладающая расширенной функциональностью – построение поверхностей, некоторые возможности редактирования
http://cn.expasy.org/spdbv

Еще несколько полезных программ

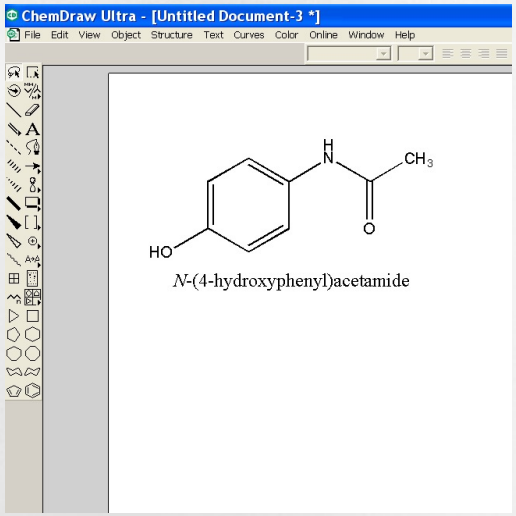




“Hyperchem” фирмы Hypercube Нельзя относиться серьезно к инструментам квантовой химии или молекулярной динамики.
В качестве визуализатора и редактора молекулярных структур программа удобна. Коммерческая, но широко распространенная у нас.
http://www.hyper.com/

“Isis-Draw” фирмы MDL. Удобная «рисовалка» 2D структур. Сохранение структур в mol формате. Удобна для трансформации в 3D структуры (например с помощью CORINA). Бесплатная программа. http://www.mdl.com

ChemDraw – «рисовалка» 2D структур. Платная, но популярная. По сравнению “Isis-Draw” имеет расширенные возможности – генерация структуры по названию, названия по структуре, предсказание ЯМР спектров и т.д.
http://www.cambridgesoft.com/

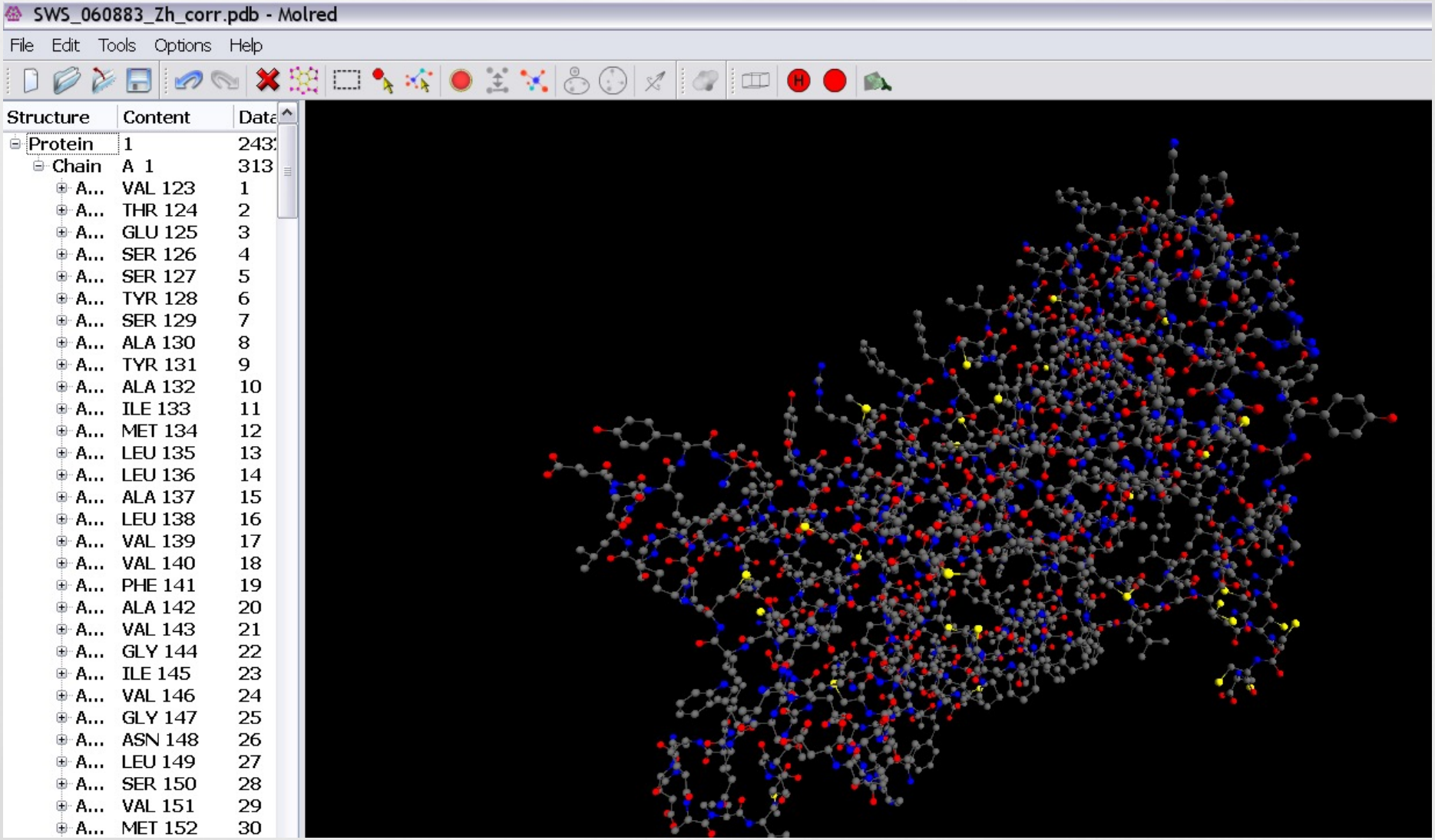


Avogadro - SourceForge, Inc.,

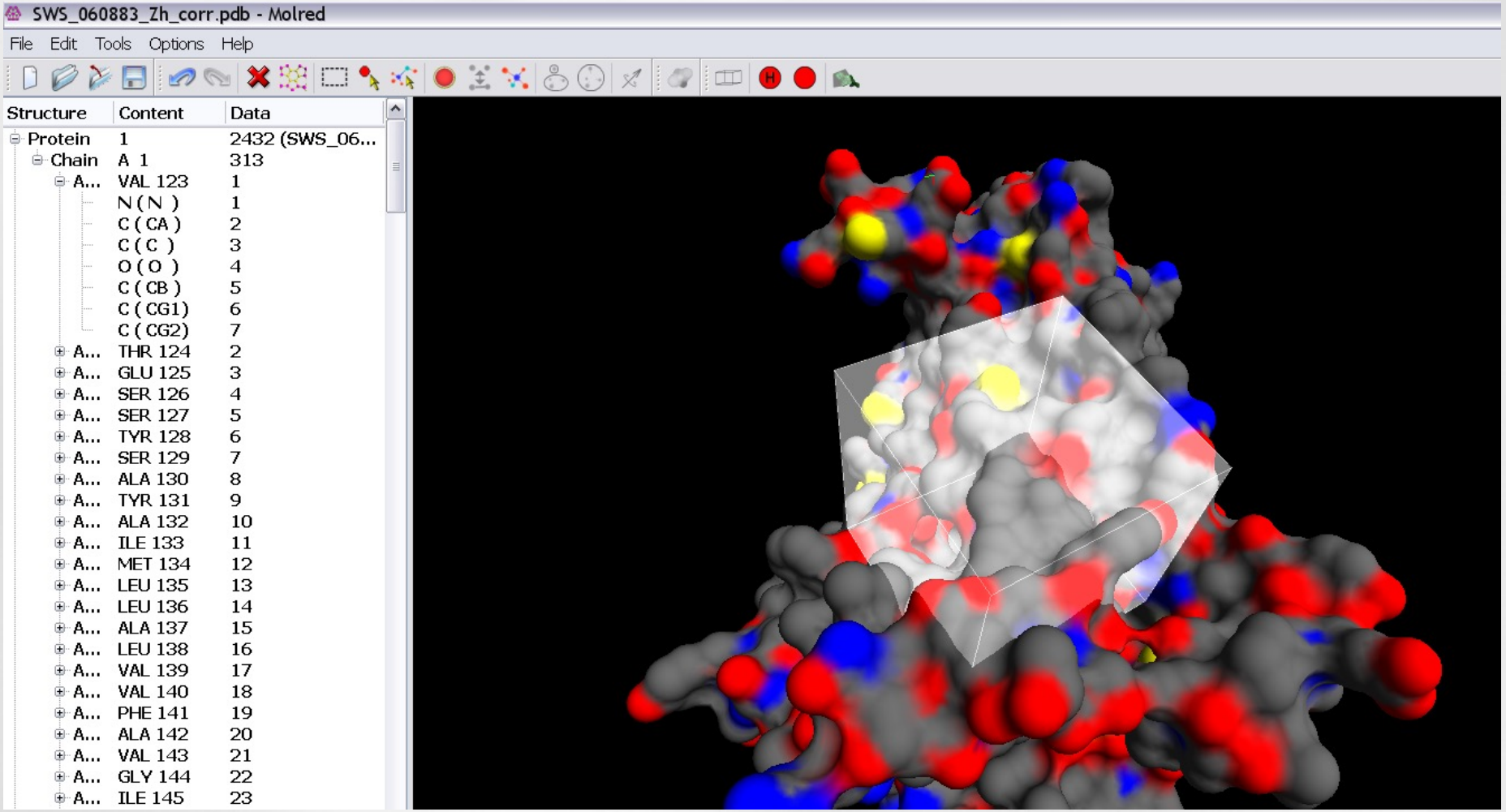
- Молекулярный редактор
- Свободный – лицензия GNU
- Сайт - **http://avogadro.openmolecules.net/wiki/Main_Page**

Marvin, PyMol

MolRed - наш


SWS_060883_Zh_corr.pdb - Molred
File Edit Tools Options Help
Structure Content Data
Protein 1 243
Chain A 1 313
A... VAL 123 1
A... THR 124 2
A... GLU 125 3
A... SER 126 4
A... SER 127 5
A... TYR 128 6
A... SER 129 7
A... ALA 130 8
A... TYR 131 9
A... ALA 132 10
A... ILE 133 11
A... MET 134 12
A... LEU 135 13
A... LEU 136 14
A... ALA 137 15
A... LEU 138 16
A... VAL 139 17
A... VAL 140 18
A... PHE 141 19
A... ALA 142 20
A... VAL 143 21
A... GLY 144 22
A... ILE 145 23
A... VAL 146 24
A... GLY 147 25
A... ASN 148 26
A... LEU 149 27
A... SER 150 28
A... VAL 151 29
A... MET 152 30

Наш MolRed


SWS_060883_Zh_corr.pdb - Molred
File Edit Tools Options Help
Structure Content Data
Protein 1 2432 (SWS_06...
Chain A 1 313
A... VAL 123 1
N (N) 1
C (CA) 2
C (C) 3
O (O) 4
C (CB) 5
C (CG1) 6
C (CG2) 7
A... THR 124 2
A... GLU 125 3
A... SER 126 4
A... SER 127 5
A... TYR 128 6
A... SER 129 7
A... ALA 130 8
A... TYR 131 9
A... ALA 132 10
A... ILE 133 11
A... MET 134 12
A... LEU 135 13
A... LEU 136 14
A... ALA 137 15
A... LEU 138 16
A... VAL 139 17
A... VAL 140 18
A... PHE 141 19
A... ALA 142 20
A... VAL 143 21
A... GLY 144 22
A... ILE 145 23

3D-структура белков и комплексов белок-лиганд

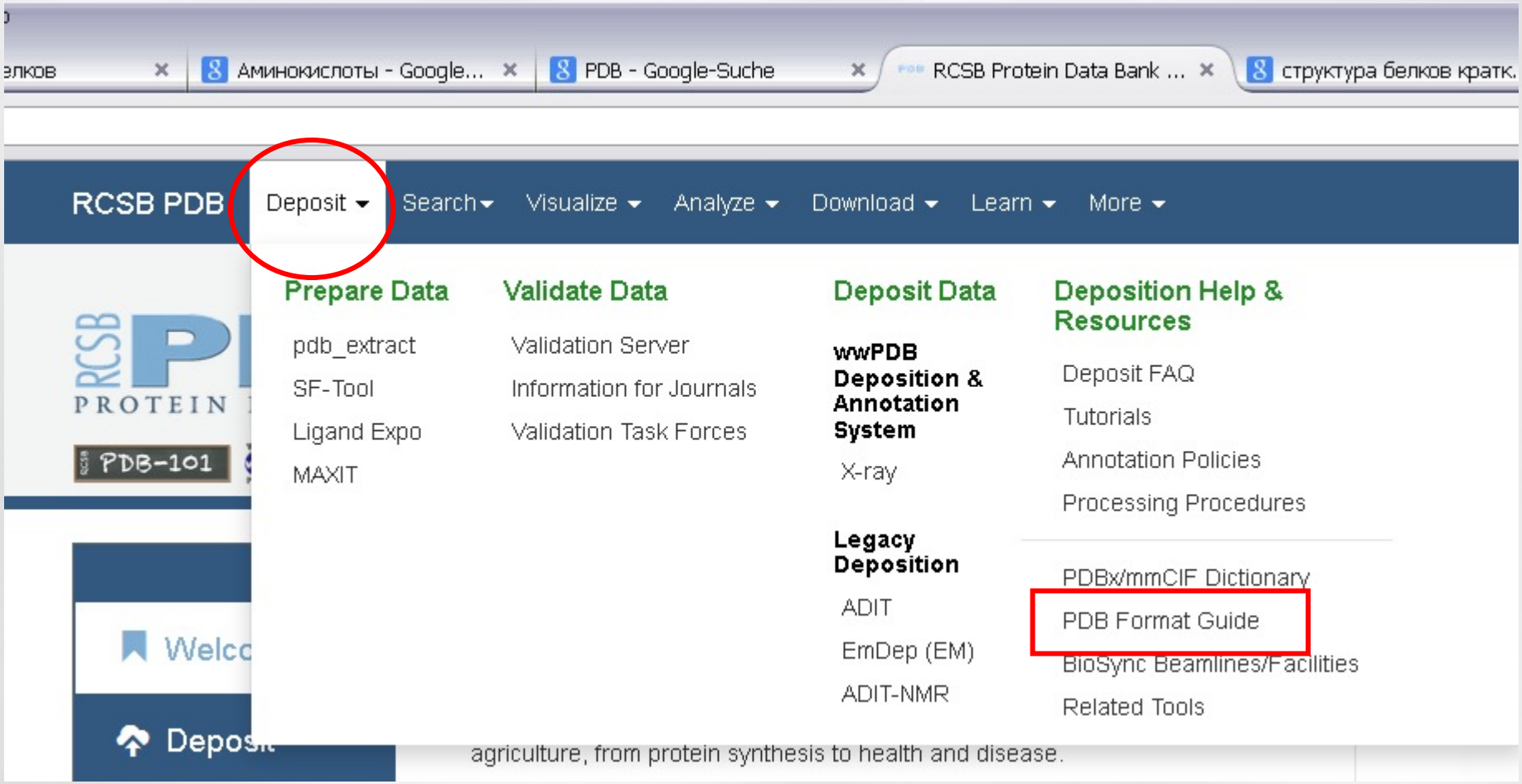
Protein Data Bank - PDB: **www.rcsb.org/pdb**

Файл белка записан в формате pdb.
Это текстовый формат

PDB ID

```
HEADER    HYDROLASE(SERINE PROTEINASE)          19-AUG-92   1DWC
TITLE     CRYSTALLOGRAPHIC ANALYSIS AT 3.0-ANGSTROMS RESOLUTION OF
TITLE    2 THE BINDING TO HUMAN THROMBIN OF FOUR ACTIVE SITE-DIRECTED
TITLE    3 INHIBITORS
COMPND    MOL_ID: 1;
COMPND   2 MOLECULE: ALPHA-THROMBIN (SMALL SUBUNIT);
COMPND   3 CHAIN: L;
COMPND   4 EC: 3.4.21.5;
COMPND   5 ENGINEERED: YES;
……..
```

Описание формата pdb



PDB-файл продолжение

```
SOURCE    MOL_ID: 1;
SOURCE   2 ORGANISM_SCIENTIFIC: HOMO SAPIENS;
SOURCE   3 ORGANISM_COMMON: HUMAN;
SOURCE   4 ORGANISM_TAXID: 9606;
SOURCE   5 ORGAN: PLASMA;
……..

KEYWDS    HYDROLASE(SERINE PROTEINASE)
EXPDTA    X-RAY DIFFRACTION
AUTHOR    D.W.BANNER,P.HADVARY
REVDAT   3   24-FEB-09 1DWC    1       VERSN
REVDAT   2   01-APR-03 1DWC    1       JRNL
REVDAT   1   31-JAN-94 1DWC    0
JRNL        AUTH   D.W.BANNER,P.HADVARY
JRNL        TITL   CRYSTALLOGRAPHIC ANALYSIS AT 3.0-A RESOLUTION OF
JRNL        TITL 2 THE BINDING TO HUMAN THROMBIN OF FOUR ACTIVE
JRNL        TITL 3 SITE-DIRECTED INHIBITORS.
JRNL        REF    J.BIOL.CHEM.                 V. 266 20085 1991
…….
```

PDB-файл продолжение

```
REMARK   1
REMARK   1 REFERENCE 1
REMARK   1  AUTH   W.BODE,D.TURK,J.STURZEBECHER
REMARK   1  TITL   GEOMETRY OF BINDING OF THE BENZAMIDINE- AND
REMARK   1  TITL 2 ARGININE-BASED INHIBITORS N
REMARK   1  TITL 3 ALPHA-(2-NAPHTHYL-SULPHONYL-GLYCYL)-DL-P-
REMARK   1  TITL 4 AMIDINOPHENYLALANYL-PIPE RIDINE (NAPAP) AND
REMARK   1  TITL 5 (2R,4R)-4-METHYL-1-[N
REMARK   1  TITL 6 ALPHA-(3-METHYL-1,2,3,4-TETRAHYDRO-8-
REMARK   1  TITL 7 QUINOLINESULPHONYL)-L-ARGINYL]-2-PIPERIDINE
REMARK   1  TITL 8 CARBOXYLIC ACID (MQPA) TO HUMAN ALPHA-THROMBIN.
REMARK   1  TITL 9 X-RAY CRYSTALLOGRAPHIC DETERMINATION OF THE
REMARK   1  TITL10 NAPAP-TRYPSIN COMPLEX AND MODELING OF
REMARK   1  TITL11 NAPAP-THROMBIN AND MQPA-THROMBIN.
REMARK   1  REF    EUR.J.BIOCHEM.              V. 193   175 1990
REMARK   1  REFN                   ISSN 0014-2956
REMARK   1  PMID   2226434
REMARK   1 REFERENCE 2
……..
```

PDB-файл продолжение

```
REMARK   2
REMARK   2 RESOLUTION.    3.00 ANGSTROMS.
REMARK   3
REMARK   3 REFINEMENT.
REMARK   3   PROGRAM     : TNT
REMARK   3   AUTHORS     : TRONRUD,TEN EYCK,MATTHEWS
REMARK   3
REMARK   3  DATA USED IN REFINEMENT.
REMARK   3   RESOLUTION RANGE HIGH (ANGSTROMS) : 3.00
REMARK   3   RESOLUTION RANGE LOW  (ANGSTROMS) : 25.00
REMARK   3   DATA CUTOFF            (SIGMA(F)) : 0.000
REMARK   3   COMPLETENESS FOR RANGE        (%) : NULL
REMARK   3   NUMBER OF REFLECTIONS             : NULL
...
REMARK   3  NUMBER OF NON-HYDROGEN ATOMS USED IN REFINEMENT.
REMARK   3   PROTEIN ATOMS            : 2352
REMARK   3   NUCLEIC ACID ATOMS       : 0
REMARK   3   HETEROGEN ATOMS          : 35
REMARK   3   SOLVENT ATOMS            : 50
REMARK   3
REMARK   3  WILSON B VALUE (FROM FCALC, A**2) : NULL
REMARK   3
REMARK   3  RMS DEVIATIONS FROM IDEAL VALUES.    RMS    WEIGHT  COUNT
REMARK   3   BOND LENGTHS                (A) : 0.014 ; NULL  ; NULL
REMARK   3   BOND ANGLES           (DEGREES) : 1.940 ; NULL  ; NULL
...
```

PDB-файл продолжение

```
REMARK   3  OTHER REFINEMENT REMARKS: PROLINES ARE TOO FLAT AS A BAD
REMARK   3  DICTIONARY WAS USED. RESIDUE ASN H 60G IS GLYCOSYLATED.
…
REMARK   4
REMARK   4 1DWC COMPLIES WITH FORMAT V. 3.15, 01-DEC-08
REMARK 100
REMARK 100 THIS ENTRY HAS BEEN PROCESSED BY BNL.
REMARK 200
REMARK 200 EXPERIMENTAL DETAILS
REMARK 200  EXPERIMENT TYPE                : X-RAY DIFFRACTION
…
REMARK 290 CRYSTALLOGRAPHIC SYMMETRY
REMARK 290 SYMMETRY OPERATORS FOR SPACE GROUP: P 43 21 2
REMARK 290
REMARK 290      SYMOP   SYMMETRY
REMARK 290     NNNMMM   OPERATOR
REMARK 290       1555   X,Y,Z
REMARK 290       2555   -X,-Y,Z+1/2
REMARK 290       3555   -Y+1/2,X+1/2,Z+3/4
REMARK 290       4555   Y+1/2,-X+1/2,Z+1/4
REMARK 290       5555   -X+1/2,Y+1/2,-Z+3/4
REMARK 290       6555   X+1/2,-Y+1/2,-Z+1/4
REMARK 290       7555   Y,X,-Z
REMARK 290       8555   -Y,-X,-Z+1/2
REMARK 290
REMARK 290     WHERE NNN -> OPERATOR NUMBER
REMARK 290           MMM -> TRANSLATION VECTOR
```

PDB-файл продолжение

REMARK 350 BIOMOLECULE: 1
REMARK 350 AUTHOR DETERMINED BIOLOGICAL UNIT: TRIMERIC
REMARK 350 SOFTWARE DETERMINED QUATERNARY STRUCTURE: TRIMERIC
REMARK 350 SOFTWARE USED: PISA
REMARK 350 TOTAL BURIED SURFACE AREA: 4190 ANGSTROM**2
REMARK 350 SURFACE AREA OF THE COMPLEX: 13380 ANGSTROM**2
REMARK 350 CHANGE IN SOLVENT FREE ENERGY: -20.0 KCAL/MOL
REMARK 350 APPLY THE FOLLOWING TO CHAINS: L, H, I
…
REMARK 400 THROMBIN IS CLEAVED BETWEEN RESIDUES 15 AND 16. CHAIN
REMARK 400 INDICATOR *L* IS USED FOR RESIDUES 1H - 15 AND CHAIN
REMARK 400 INDICATOR *H* IS USED FOR RESIDUES 16 - 247. CHAIN
REMARK 400 INDICATOR *I* IS USED FOR A C-TERMINAL HIRUDIN PEPTIDE
REMARK 400 ((DES-AMINO ASP55)HIRUDIN(55-65)). THE RESIDUES OF
REMARK 400 CHAIN I ARE NUMBERED FROM 1 - 11. IT BINDS IN THE
REMARK 400 ANION-BINDING EXOSITE.
…

PDB-файл продолжение: пропущенные а/к остатки

```
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     THR L     1H
REMARK 465     PHE L     1G
REMARK 465     GLY L     1F
REMARK 465     SER L     1E
REMARK 465     GLY L     1D
REMARK 465     GLY L    14M
REMARK 465     ARG L    14N
REMARK 465     GLU H   247
...
```

PDB-файл продолжение: пропущенные атомы

```
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    GLU L   1C   CG   CD   OE1  OE2
REMARK 470    ARG H  77A   CG   CD   NE   CZ   NH1  NH2
REMARK 470    ASN H  78    CG   OD1  ND2
REMARK 470    LYS H  81    CG   CD   CE   NZ
REMARK 470    ARG H  93    CG   CD   NE   CZ   NH1  NH2
REMARK 470    ARG H  97    CG   CD   NE   CZ   NH1  NH2
REMARK 470    ARG H 126    CG   CD   NE   CZ   NH1  NH2
REMARK 470    GLU H 127    CG   CD   OE1  OE2
REMARK 470    LYS H 149E   CG   CD   CE   NZ
REMARK 470    LYS H 169    CG   CD   CE   NZ
REMARK 470    ARG H 173    CG   CD   NE   CZ   NH1  NH2
REMARK 470    GLY H 246    CA   C    O
REMARK 470    ASP I   1    N
REMARK 470    GLU I   8    CG   CD   OE1  OE2
REMARK 470    GLN I  11    CA   C    O    CB   CG   CD   OE1
REMARK 470    GLN I  11    NE2
 …
```

PDB-файл продолжение

```
REMARK 600 MD-805 (AGRATROBAN) IS ((2R,4R)-4-METHYL-1[N==ALPHA==-
REMARK 600 [3-METHYL-1,2,3,4-TETRAHYDRO-8-QUINOLINYL)SULPHONYL]-
REMARK 600 L-ARGINYL)]-2-PIPERIDINECARBOXYLIC ACID).
REMARK 800
REMARK 800 SITE
REMARK 800 SITE_IDENTIFIER: ACT
REMARK 800 EVIDENCE_CODE: AUTHOR
REMARK 800 SITE_DESCRIPTION: CATALYTIC SITE
REMARK 800 SITE_IDENTIFIER: AC1
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE MIT H 1
REMARK 900
REMARK 900 RELATED ENTRIES
REMARK 900 RELATED ID: 1DWB   RELATED DB: PDB
REMARK 900 RELATED ID: 1DWD   RELATED DB: PDB
REMARK 900 RELATED ID: 1DWE   RELATED DB: PDB
…
```

PDB-файл продолжение

```
SEQRES   1 L   36  THR PHE GLY SER GLY GLU ALA ASP CYS GLY LEU ARG PRO
SEQRES   2 L   36  LEU PHE GLU LYS LYS SER LEU GLU ASP LYS THR GLU ARG
SEQRES   3 L   36  GLU LEU LEU GLU SER TYR ILE ASP GLY ARG
SEQRES   1 H  259  ILE VAL GLU GLY SER ASP ALA GLU ILE GLY MET SER PRO
SEQRES   2 H  259  TRP GLN VAL MET LEU PHE ARG LYS SER PRO GLN GLU LEU
SEQRES   3 H  259  LEU CYS GLY ALA SER LEU ILE SER ASP ARG TRP VAL LEU
SEQRES   4 H  259  THR ALA ALA HIS CYS LEU LEU TYR PRO PRO TRP ASP LYS
SEQRES   5 H  259  ASN PHE THR GLU ASN ASP LEU LEU VAL ARG ILE GLY LYS
SEQRES   6 H  259  HIS SER ARG THR ARG TYR GLU ARG ASN ILE GLU LYS ILE
SEQRES   7 H  259  SER MET LEU GLU LYS ILE TYR ILE HIS PRO ARG TYR ASN
SEQRES   8 H  259  TRP ARG GLU ASN LEU ASP ARG ASP ILE ALA LEU MET LYS
SEQRES   9 H  259  LEU LYS LYS PRO VAL ALA PHE SER ASP TYR ILE HIS PRO
SEQRES  10 H  259  VAL CYS LEU PRO ASP ARG GLU THR ALA ALA SER LEU LEU
SEQRES  11 H  259  GLN ALA GLY TYR LYS GLY ARG VAL THR GLY TRP GLY ASN
SEQRES  12 H  259  LEU LYS GLU THR TRP THR ALA ASN VAL GLY LYS GLY GLN
SEQRES  13 H  259  PRO SER VAL LEU GLN VAL VAL ASN LEU PRO ILE VAL GLU
SEQRES  14 H  259  ARG PRO VAL CYS LYS ASP SER THR ARG ILE ARG ILE THR
SEQRES  15 H  259  ASP ASN MET PHE CYS ALA GLY TYR LYS PRO ASP GLU GLY
SEQRES  16 H  259  LYS ARG GLY ASP ALA CYS GLU GLY ASP SER GLY GLY PRO
SEQRES  17 H  259  PHE VAL MET LYS SER PRO PHE ASN ASN ARG TRP TYR GLN
SEQRES  18 H  259  MET GLY ILE VAL SER TRP GLY GLU GLY CYS ASP ARG ASP
SEQRES  19 H  259  GLY LYS TYR GLY PHE TYR THR HIS VAL PHE ARG LEU LYS
SEQRES  20 H  259  LYS TRP ILE GLN LYS VAL ILE ASP GLN PHE GLY GLU
SEQRES   1 I   11  ASP PHE GLU GLU ILE PRO GLU GLU TYR LEU GLN
...
```

PDB-файл продолжение

```
HET    MIT  H   1      35
HETNAM     MIT ARGATROBAN
HETSYN     MIT MD-805; MITSUBISHI INHIBITOR
FORMUL   4  MIT    C23 H36 N6 O5 S
FORMUL   5  HOH   *50(H2 O)
HELIX    1   1 PHE L    7  SER L   11  5                                    5
HELIX    2   2 THR L   14B SER L   14I 1                                     8
HELIX    3   3 ALA H   55  CYS H   58  5                                    4
HELIX    4   4 PRO H   60B ASP H   60E 5                                      4
HELIX    5   5 ASP H  125  LEU H  130  1                                     9
HELIX    6   6 GLU H  164  ASP H  170  1                                     7
HELIX    7   7 LYS H  185  GLY H  186C 5                                     5
HELIX    8   8 VAL H  231  ASP H  243  1                                    13
SHEET    1   A 7 SER H  20  ASP H  21  0
SHEET    2   A 7 GLN H 156  PRO H 161 -1  N  VAL H 157   O  SER H  20
SHEET    3   A 7 LYS H 135  GLY H 140 -1  N  GLY H 136   O  LEU H 160
SHEET    4   A 7 PRO H 198  LYS H 202 -1  O  PRO H 198   N  THR H 139
SHEET    5   A 7 TRP H 207  TRP H 215 -1  N  TYR H 208   O  MET H 201
SHEET    6   A 7 GLY H 226  HIS H 230 -1  N  PHE H 227   O  TRP H 215
SHEET    7   A 7 MET H 180  ALA H 183 -1  O  PHE H 181   N  TYR H 228
…
```

PDB-файл продолжение

```
ATOM      1  N    GLU   L   1C     63.691  25.940  17.920  1.00 39.73          N
ATOM      2  CA  GLU   L   1C     64.331  26.594  16.778  1.00 41.21          C
ATOM      3  C    GLU   L   1C     63.325  27.181  15.823  1.00 43.53          C
ATOM      4  O    GLU   L   1C     63.253  28.400  15.711  1.00 48.20          O
ATOM      5  CB   GLU   L   1C     65.333  25.719  16.023  1.00 41.20          C
ATOM      6  N    ALA   L  1B     62.540  26.309  15.144  1.00 39.72          N
ATOM      7  CA  ALA   L   1B     61.528  26.767  14.192  1.00 35.59          C
ATOM      8  C    ALA   L   1B     60.677  27.829  14.830  1.00 34.53          C
ATOM      9  O    ALA   L   1B    60.267  27.670  15.965  1.00 34.72          O
ATOM     10  CB  ALA   L   1B     60.675  25.608  13.772  1.00 34.79          C
ATOM     11  N    ASP   L   1A     60.419  28.936  14.148  1.00 33.71          N
ATOM     12  CA  ASP L    1A      59.616  29.916  14.837  1.00 34.84          C
ATOM     13  C   ASP L    1A      58.180  29.473  14.803  1.00 32.52          C
ATOM     14  O   ASP L    1A      57.740  28.843  13.838  1.00 36.56          O
ATOM     15  CB  ASP L    1A      59.843  31.394  14.473  1.00 40.37          C
ATOM     16  CG  ASP L    1A      58.777  31.961  13.594  1.00 46.11          C
ATOM     17  OD1 ASP L   1A      57.587  32.151  14.206  1.00 47.53          O
ATOM     18  OD2 ASP L   1A      58.998  32.224  12.393  1.00 49.59          O
ATOM     19  N   CYS L   1      57.464  29.769  15.871  1.00 25.39          N
ATOM     20  CA  CYS L   1      56.104  29.368  15.999  1.00 19.29          C
ATOM     21  C   CYS L   1      55.465  30.177  17.051  1.00 17.84          C
ATOM     22  O   CYS L   1      56.146  30.757  17.853  1.00 18.58          O
ATOM     23  CB  CYS L   1      56.102  27.939  16.548  1.00 16.58          C
ATOM     24  SG  CYS L   1      56.982  27.807  18.150  1.00 14.20          S
ATOM     25  N   GLY L   2      54.159  30.186  17.080  1.00 15.50          N
ATOM     26  CA  GLY L   2      53.489  30.931  18.092  1.00 13.49          C
ATOM     27  C   GLY L   2      53.388  32.388  17.738  1.00 13.84          C
ATOM     28  O   GLY L   2      52.651  33.115  18.393  1.00 17.90          O
...
```

PDB-файл продолжение

```
TER    2355      GLN I  11
HETATM 2356  C1  MIT H   1      29.562  11.960  25.333  1.00 23.50           C
HETATM 2357  C2  MIT H   1      29.379  11.358  24.121  1.00 21.68           C
HETATM 2358  C3  MIT H   1      30.109  10.204  23.800  1.00 22.45           C
HETATM 2359  C4  MIT H   1      31.058   9.676  24.680  1.00 23.62           C
HETATM 2360  C5  MIT H   1      31.311  10.357  25.865  1.00 23.51           C
HETATM 2361  C6  MIT H   1      32.263   9.827  26.844  1.00 23.37           C
HETATM 2362  C7  MIT H   1      32.864  10.956  27.600  1.00 25.06           C
HETATM 2363  C8  MIT H   1      31.796  11.757  28.240  1.00 24.50           C
HETATM 2364  N9  MIT H   1      30.826  12.247  27.300  1.00 24.31           N
HETATM 2365  C10 MIT H   1      30.596  11.546  26.218  1.00 23.28           C
HETATM 2366  S11 MIT H   1      28.752  13.513  25.638  1.00 26.28           S
HETATM 2367  N12 MIT H   1      29.732  14.820  25.183  1.00 26.79           N
HETATM 2368  C13 MIT H   1      30.493  14.834  23.936  1.00 25.80           C
HETATM 2369  C14 MIT H   1      31.834  14.152  24.121  1.00 28.18           C
HETATM 2370  O15 MIT H   1      32.520  14.296  25.111  1.00 28.61           O
HETATM 2371  C16 MIT H   1      30.709  16.241  23.446  1.00 23.63           C
HETATM 2372  C17 MIT H   1      30.913  17.170  24.596  1.00 23.22           C
HETATM 2373  C18 MIT H   1      31.361  18.514  24.009  1.00 25.07           C
HETATM 2374  N19 MIT H   1      31.580  19.501  25.067  1.00 27.15           N
HETATM 2375  C20 MIT H   1      32.838  19.878  25.359  1.00 27.79           C
HETATM 2376  N21 MIT H   1      33.869  19.302  24.687  1.00 26.36           N
HETATM 2377  N22 MIT H   1      33.048  20.701  26.406  1.00 29.63           N
HETATM 2378  C23 MIT H   1      33.794  10.424  28.673  1.00 27.98           C
HETATM 2379  O24 MIT H   1      28.554  13.649  27.057  1.00 28.68           O
HETATM 2380  O25 MIT H   1      27.669  13.399  24.680  1.00 29.28           O
HETATM 2381  N26 MIT H   1      32.337  13.531  23.009  1.00 27.73           N
HETATM 2382  C27 MIT H   1      31.632  13.594  21.735  1.00 27.66           C
HETATM 2383  C28 MIT H   1      31.563  12.208  21.091  1.00 26.23           C
HETATM 2384  C29 MIT H   1      32.931  11.624  21.031  1.00 25.62           C
HETATM 2385  C30 MIT H   1      33.571  11.597  22.409  1.00 25.93           C
HETATM 2386  C31 MIT H   1      33.628  12.930  23.079  1.00 27.07           C
HETATM 2387  C32 MIT H   1      32.242  14.652  20.817  1.00 26.52           C
HETATM 2388  O33 MIT H   1      31.878  14.577  19.649  1.00 24.67           O
HETATM 2389  O34 MIT H   1      32.518  15.723  21.305  1.00 27.96           O
HETATM 2390  C35 MIT H   1      32.800  10.219  20.483  1.00 25.24           C
HETATM 2391  O   HOH L  30      45.677  42.120  11.659  1.00 42.67           O
HETATM 2392  O   HOH L  45      46.756  33.196  27.893  1.00 15.09           O
HETATM 2393  O   HOH L  48      40.609  38.026  31.933  1.00  8.90           O
```

Задание

- Взять из Protein Data Bank комплекс белок-лиганд
- Добавить атомы водорода в белок и лиганд
- Создать сетку потенциалов SOL_GRID
- Провести докинг нативного лиганда SOL
- Проанализировать результаты
- Написать отчет

Алексей Владимирович Сулимов

Спасибо за внимание

- *…Surely every medicine is an innovation; and he that will not apply new remedies, must expect new evils…*
- *…Каждый медицинский метод есть инновация; а кто не хочет применять новые средства, должен ждать новых бед…*

Sir Francis Bacon (1561-1626)

OF INNOVATIONS