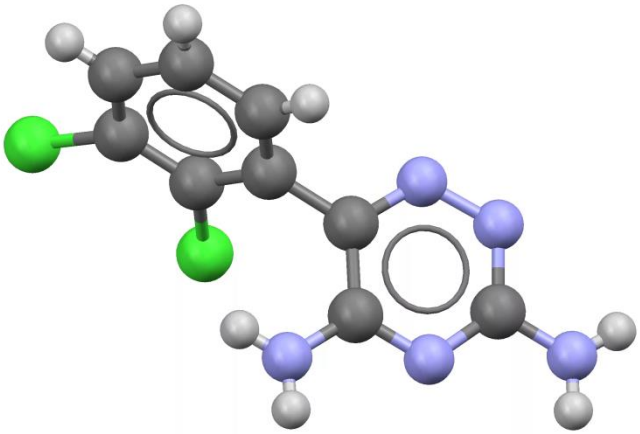
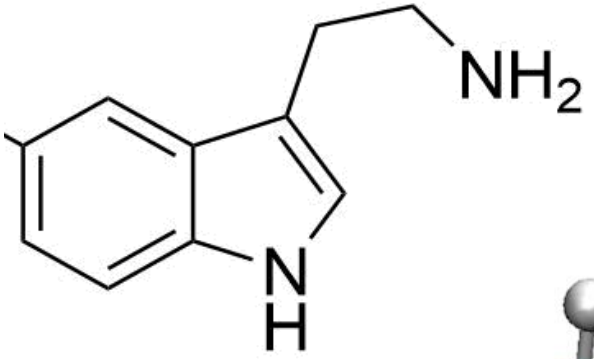
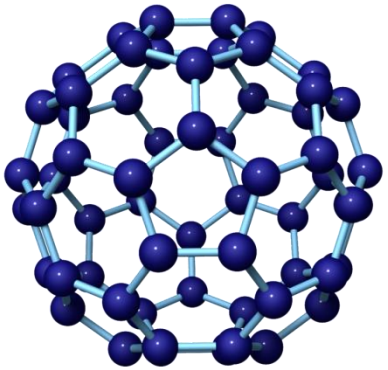
Работа с химическими структурами

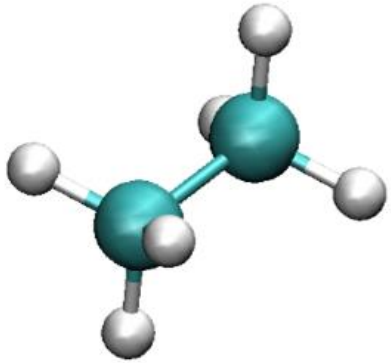
Кадцын Е.Д., Кравченко А.А.

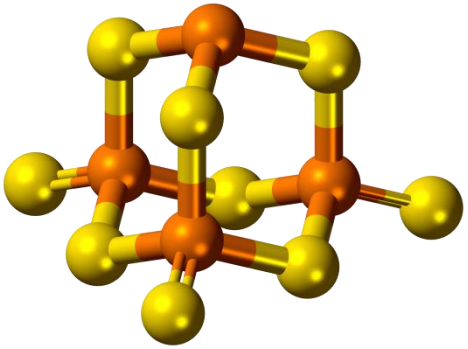
Химические структуры

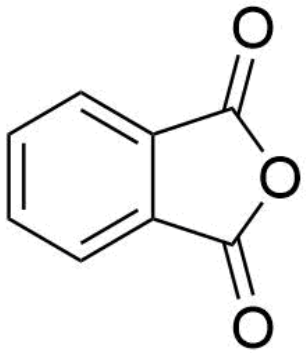


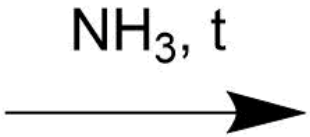


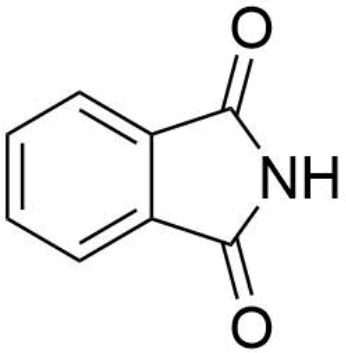












Бит и байт

Бит – основная единица представления данных. Может принимать значения 0 или 1.

Байт – элементарная ячейка компьютерной памяти, состоящая из 8 бит, обрабатывается как единое целое. Каждый байт в памяти имеет свой порядковый номер – ***адрес***. Для считывания или записи необходимо обратиться к байту по его адресу.

Представление байта в виде десятичного числа

№ байта	Биты								Десятичное значение
0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	1	1
2	0	0	0	0	0	0	1	0	2
3	0	0	0	0	0	1	0	0	4
4	0	0	0	1	0	0	0	0	16
5	0	1	0	1	0	0	0	0	80

- Каждый байт можно рассматривать как 8-разрядное двоичное число, кодирующее числа от 0 до 255 (2^8-1).
- Можно два байта, идущих подряд, рассматривать как 16-разрядное двоичное число, тогда они будут кодировать десятичные числа от 0 до 65535 (2^{16}-1).
- Можно первый бит рассматривать как знак числа (например 0 – положительное, 1 – отрицательное), тогда разрядность числа уменьшится на 1, но можно кодировать отрицательные числа.

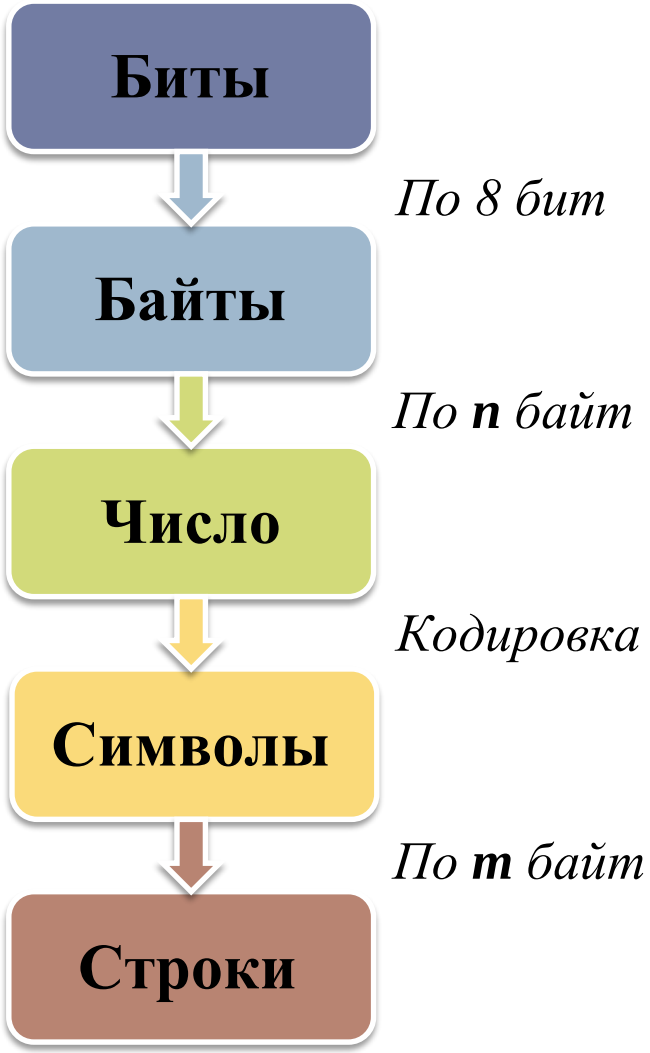
Кодировка символов и более сложные структуры

HI! –можно записать тремя байтами

№	Биты								Число	Символ
0	0	1	0	0	1	0	0	0	72	**H**
1	0	1	0	0	1	0	0	1	73	**I**
2	0	0	1	0	0	0	0	1	33	**!**

ASCII

Десятичное число	Символ	Десятичное число	Символ
33	**!**	69	E
65	A	70	F
66	B	71	G
67	C	**72**	**H**
68	D	**73**	**I**



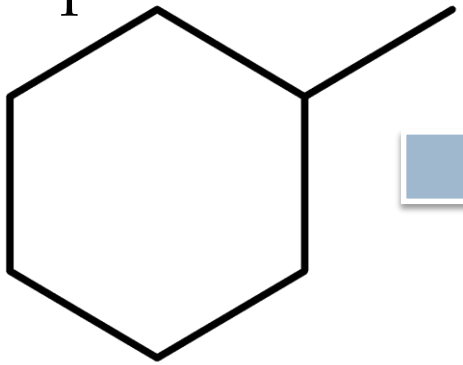
Линейные представления

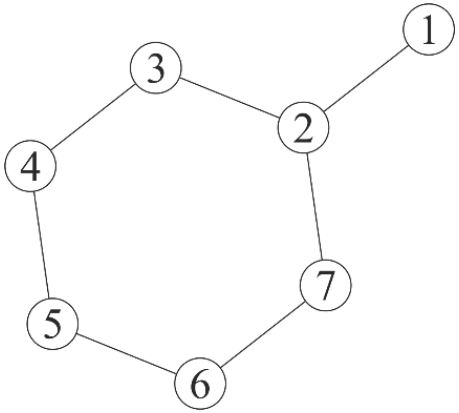
Линейные представления – способы представления структуры, позволяющие записать всю информацию в виде строки.

Форма записи линейного представления	**Тип линейного представления**
C6H10O5 или HOOCCH2CH2CH(OH)CH2COOH	*- брутто-формула*
2-гидроксибутан-1,4-дикарбоновая кислота	*- систематическое название*
OC(CCC(=O)O)CC(=O)O	*- код SMILES*
InChI=1S/C6H10O5/c7-4(3-6(10)11)1-2-5(8)9/h4,7H,1-3H2,(H,8,9)(H,10,11)	*- код InChI*

Нелинейные представления (графы)

Граф – абстрактный математический объект, представляющий собой множество вершин графа и набор рёбер, то есть соединений между парами вершин.





Связан с атомами 1, 3 и 7

2 атом

0	1	0	0	0	0	0
1	0	1	0	0	0	1
0	1	0	1	0	0	0
0	0	1	0	1	0	0
0	0	0	1	0	1	0
0	0	0	0	1	0	1
0	1	0	0	0	1	0

Матрица связности – отображает связи между атомами (вершинами графа).

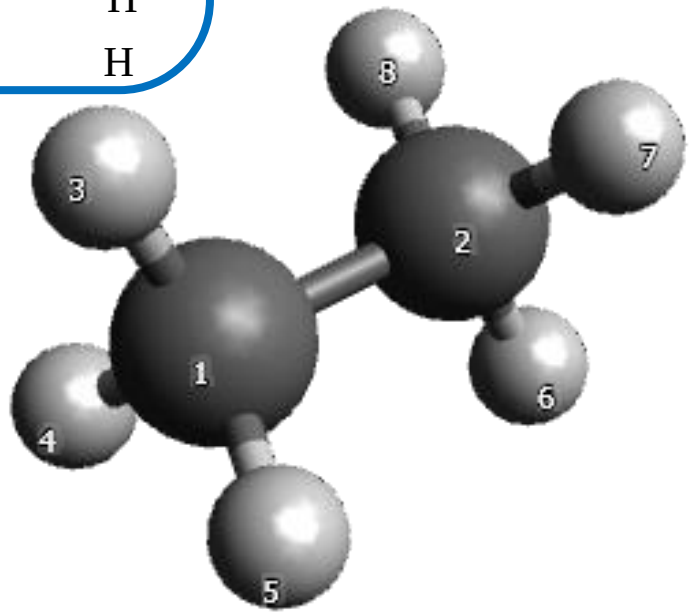
Нелинейные представления (трехмерные)

HETATM	1	C	1	-0.758	-0.072	0.000	C
HETATM	2	C	1	0.758	0.072	0.000	C
HETATM	3	H	1	-1.039	-1.149	-0.000	H
HETATM	4	H	1	-1.186	0.415	-0.905	H
HETATM	5	H	1	-1.186	0.415	0.905	H
HETATM	6	H	1	1.039	1.149	-0.000	H
HETATM	7	H	1	1.186	-0.415	-0.905	H
HETATM	8	H	1	1.186	-0.415	0.905	H

Координаты атомов

CONECT	1	2	3	4	5
CONECT	2	1	6	7	8
CONECT	3	1			
CONECT	4	1			
CONECT	5	1			
CONECT	6	2			
CONECT	7	2			
CONECT	8	2			

Таблица связности



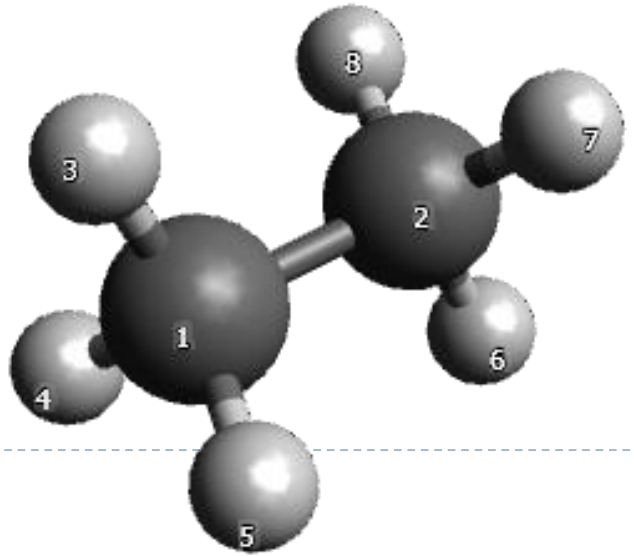
Нелинейные представления (трехмерные)

Декартовы координаты

Атом	*x, Å*	*y, Å*	*z, Å*
C(1)	-0.760	-0.045	-0.000
C(2)	0.760	0.045	-0.000
H(3)	-1.078	-1.112	0.000
H(4)	-1.171	0.456	-0.905
H(5)	-1.171	0.456	0.905
H(6)	1.078	1.112	-0.000
H(7)	1.171	-0.456	-0.905
H(8)	1.171	-0.456	0.905

Z-матрица (внутренние координаты)

Атом	*Связанный атом*	*Длина связи, Å*	*Угловой атом*	*Угол, ^{0}*	*Второй угловой атом*	*Второй угол, ^{0}*	*Тип второго угла*
C(1)							
C(2)	C(1)	1.523					
H(3)	C(1)	1.113	C(2)	110.000			
H(4)	C(1)	1.113	C(2)	110.000	H(3)	109.000	Pro-R
H(5)	C(1)	1.113	C(2)	110.000	H(3)	109.000	Pro-S
H(6)	C(2)	1.113	C(1)	110.000	H(3)	180.000	Dihedral
H(7)	C(2)	1.113	C(1)	110.000	H(6)	109.000	Pro-R
H(8)	C(2)	1.113	C(1)	110.000	H(6)	109.000	Pro-S



Химические редакторы

- Создают графическое изображение химических структур. Работают по принципу конструктора, позволяя собирать структуру из составных элементов.
- Используют различные способы представления структуры и позволяют конвертировать их друг в друга.
- Позволяют рассчитывать элементарные химические и физико-химические свойства структур.

Примеры: *ChemDraw, ISIS/Draw, BIOVIA Draw и т.д.*

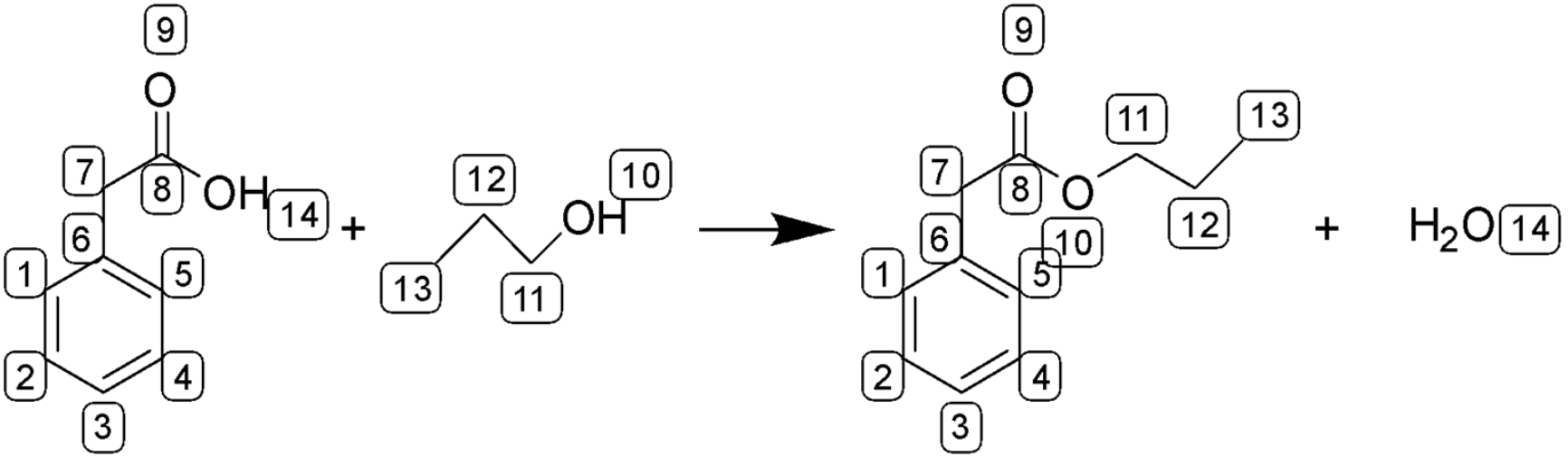
Элементарные свойства структур

- На основе брутто-формулы: элементный состав, молекулярная масса.
- На основе атомных свойств: распределение по изотопам, заряды на атомах.
- На основе связности молекулы: масс и ЯМР-спектры, топологические индексы.

Все подобные свойства рассчитываются на основании «химических» принципов и непосредственно следуют из структуры.

Способы представления реакций

- Набор продуктов и реагентов.
- Отображение атомов.



- Разность продуктов и реагентов.
- Конденсированный граф реакции.

Обменные форматы файлов

- Должны кратко описывать всю необходимую информацию.
- Должны быть универсальными (читаться и записываться любой программой).
- Должны быть текстовыми.
- Формат должен быть задокументирован и спецификация должна находиться в открытом доступе.