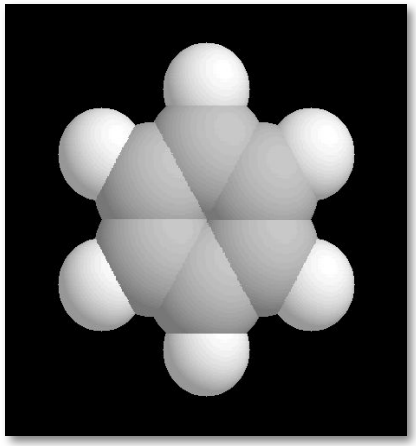
ІНФОРМАТИКА ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

МОДУЛЬ ІІ
ХІМІЧНА ІНФОРМАТИКА

Представлення хімічної структури: бензол

Бензол



ID #:
MUSE00000002

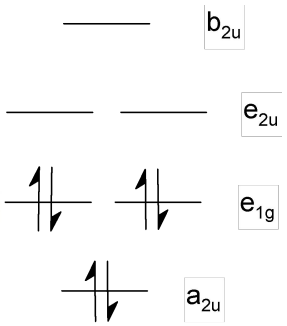
CAS #: 71-43-2

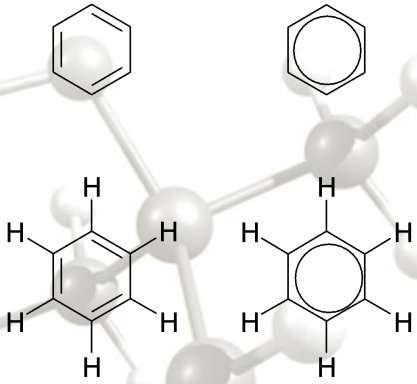
Інші назви:
Бензен
циклогекса-1,3,5-трієн

Таблиця зв'язків:

```
Benzene
 -ISIS-  08200115272D

  6  6  0  0  0  0  0  0  0  0999 V2000
   -1.0306   -1.4375    0.0000 C   0  0  0  0  0  0  0  0  0  0  0
      0
   -1.0318   -2.2648    0.0000 C   0  0  0  0  0  0  0  0  0  0  0
      0
   -0.3169   -2.6777    0.0000 C   0  0  0  0  0  0  0  0  0  0  0
      0
    0.3995   -2.2644    0.0000 C   0  0  0  0  0  0  0  0  0  0  0
      0
    0.3966   -1.4338    0.0000 C   0  0  0  0  0  0  0  0  0  0  0
      0
   -0.3187   -1.0247    0.0000 C   0  0  0  0  0  0  0  0  0  0  0
      0
  1  2  2  0  0  0  0
  3  4  2  0  0  0  0
  4  5  1  0  0  0  0
  2  3  1  0  0  0  0
  5  6  2  0  0  0  0
  6  1  1  0  0  0  0
M  END
```





Лінійна нотація

- Wiswesser: RH
- MDL LN: C-C=C-C=C-C=@1
- SMILES: c1ccccc1
- InChI InChI=1/C6H6/c1-2-4-6-5-3-1/h1-6H

Рівні представлення хімічної інформації

1D Назва, формула, лінійні коди (нотації), чисельне представлення фізико-хімічних даних

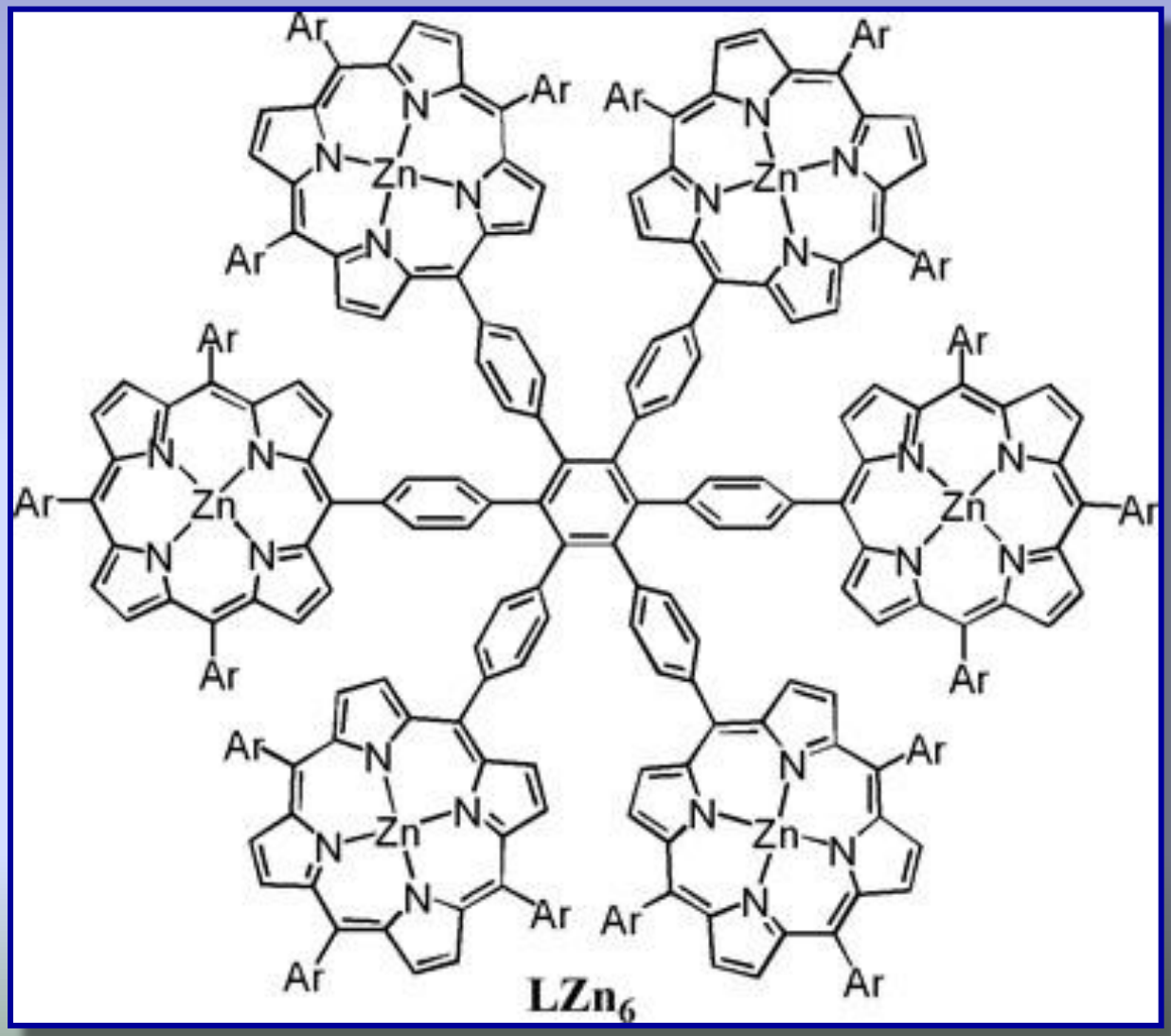
2D Графічне пласке зображення хімічної структури сполуки, представлення у вигляді двумірних матриць

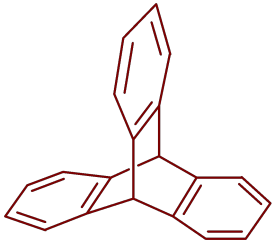
3D Визначення Декартових координат усіх атомів хімічної частинки

4D Інформація про поверхні сольватації, електростатичного потенціалу

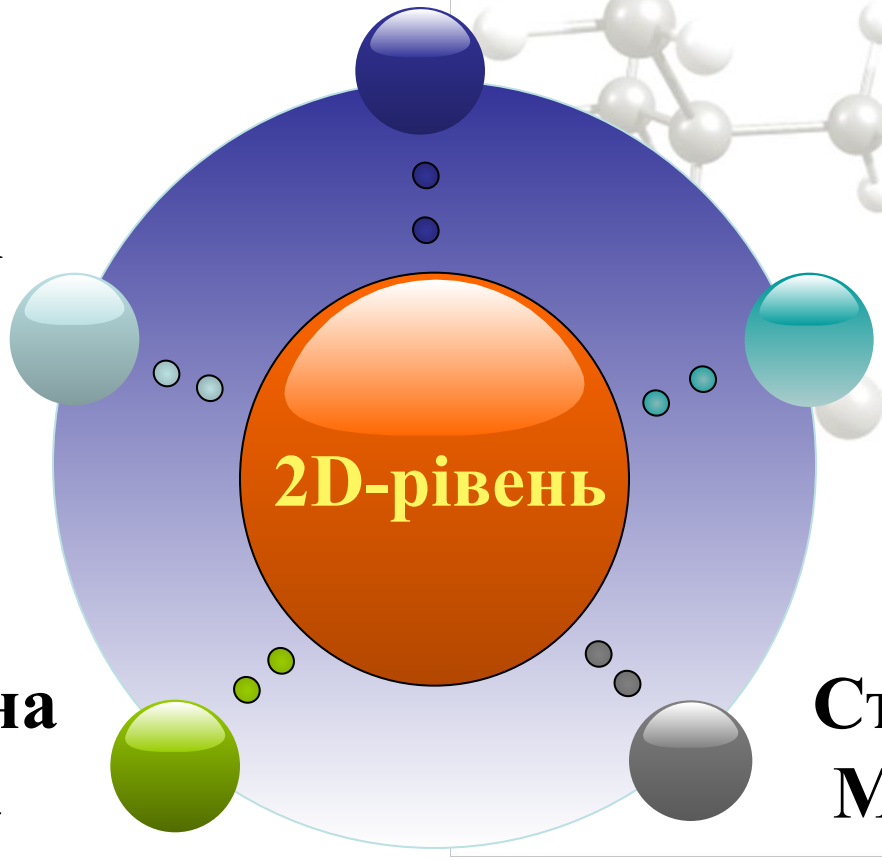
5D Інформація на рівні хімічних реакцій

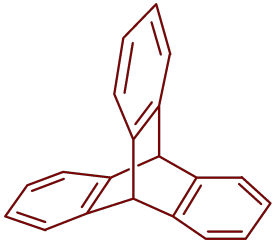
2D-РІВЕНЬ ПРЕДСТАВЛЕННЯ ХІМІЧНОЇ ІНФОРМАЦІЇ





Способи представлення структурної інформації 2D-рівня





Структурна формула

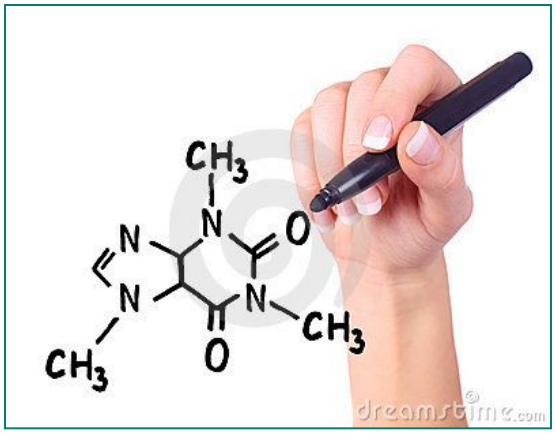
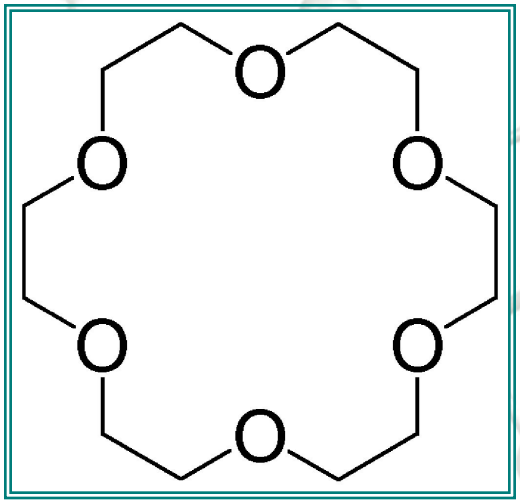


Рисунок Tif
10.7 кБ

Таблиця зв’язків
0.8 кБ

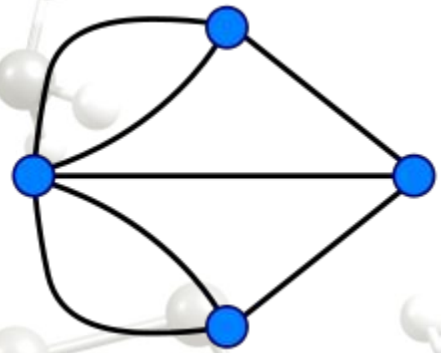


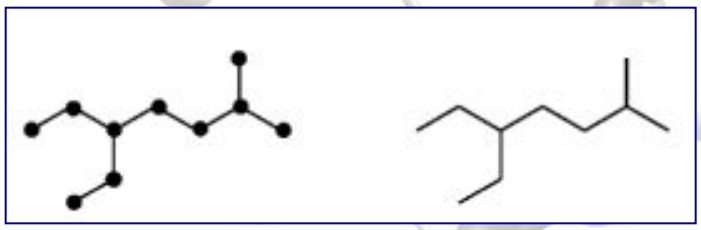
Матриця суміжності
1.7 кБ

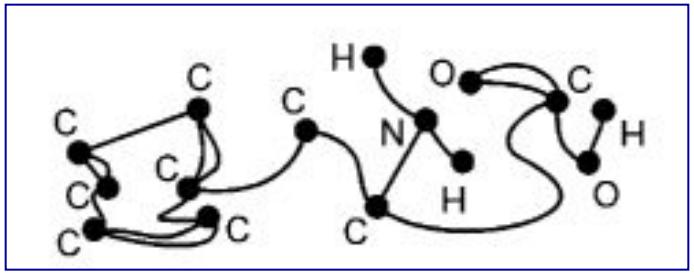
Використання теорії графів

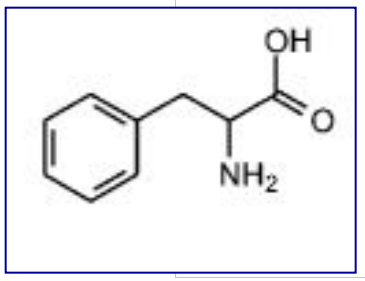
Проблема семи мостов Кёнигсберга

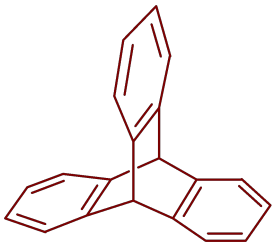
1736



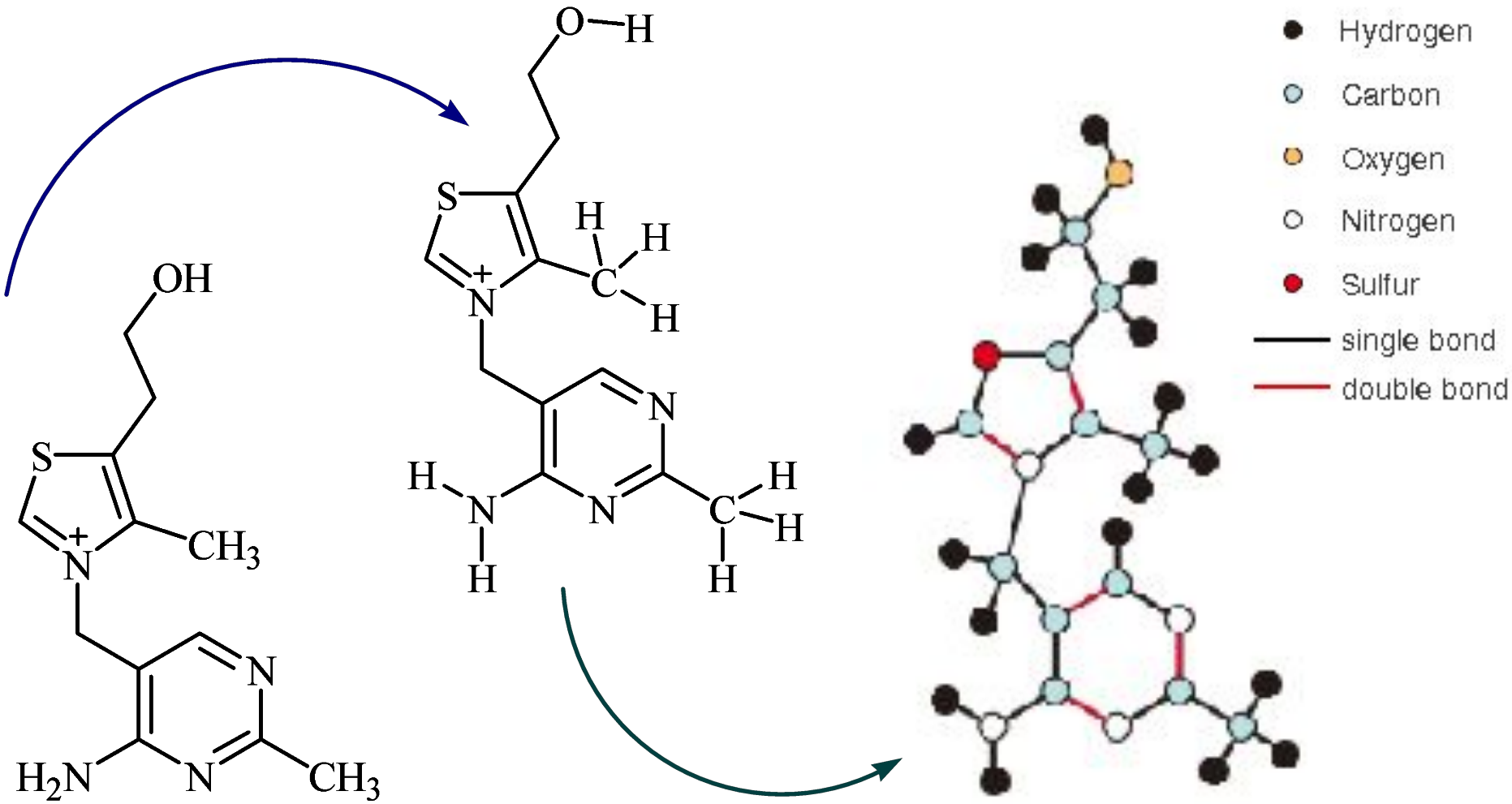


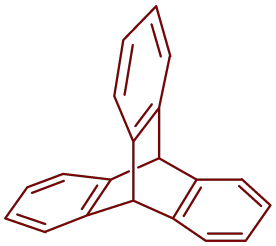




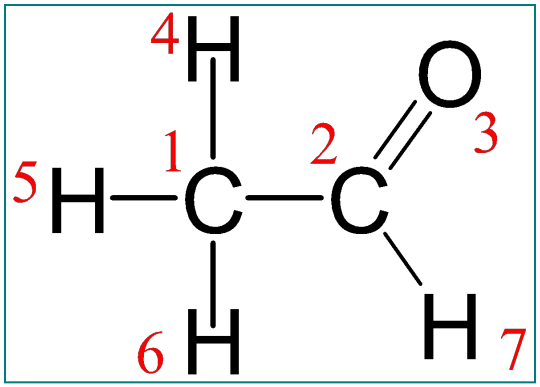


Використання теорії графів





Матриця суміжності

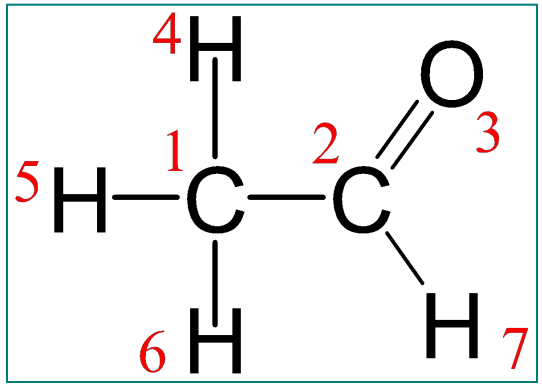


	1	2	3	4	5	6	7
1	0	1	0	1	1	1	0
2	1	0	1	0	0	0	1
3	0	1	0	0	0	0	0
4	1	0	0	0	0	0	0
5	1	0	0	0	0	0	0
6	1	0	0	0	0	0	0
7	0	1	0	0	0	0	0

a_{ij} = 1, если между атомами ***i*** и ***j*** имеется химическая связь,

a_{ij} = 0 , если между атомами ***i*** и ***j*** нет химической связи.

Матриця суміжності: етапи спрощення

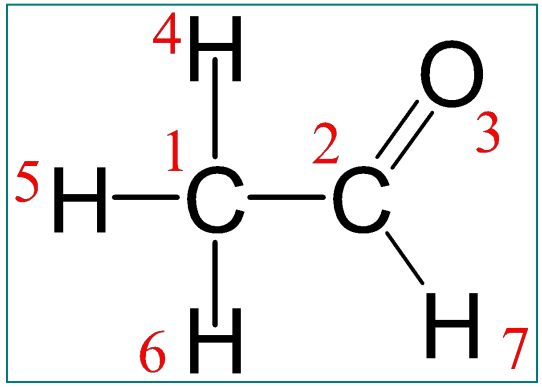


- Вилучення нулів
- Вилучення дублів
- Вилучення інформації про атоми H

	1	2	3	4	5	6	7
1	0	1	0	1	1	1	0
2	1	0	1	0	0	0	1
3	0	1	0	0	0	0	0
4	1	0	0	0	0	0	0
5	1	0	0	0	0	0	0
6	1	0	0	0	0	0	0
7	0	1	0	0	0	0	0

	1	2	3	4	5	6	7
1		1		1	1	1	
2	1		1				1
3		1					
4	1						
5	1						
6	1						
7		1					

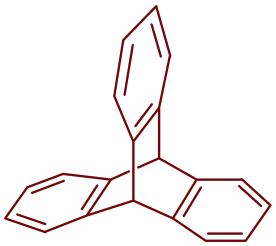
Матриця суміжності: етапи спрощення



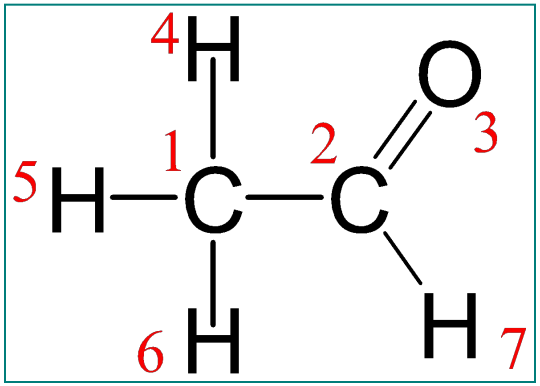
- Вилучення нулів
- Вилучення дублів
- Вилучення інформації про атоми Н

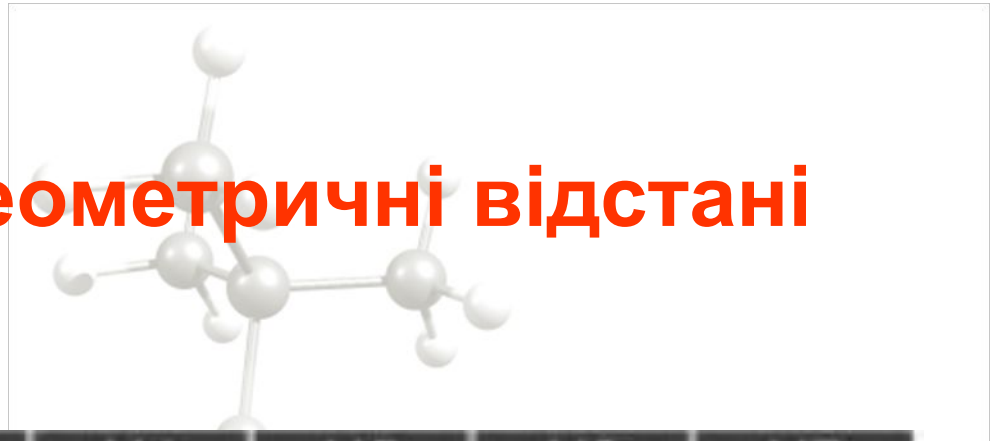
	1	2	3	4	5	6	7
1		1		1	1	1	
2			1				1
3							
4							
5							
6							
7							

	1	2	3
1		1	
2			1
3			



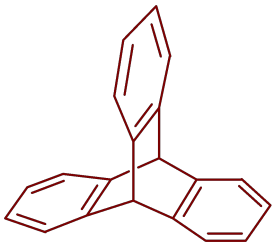
Матриця відстаней



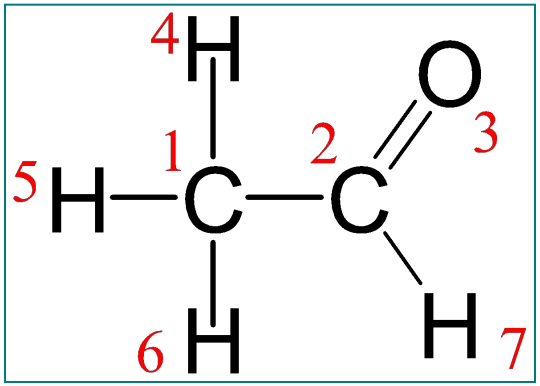


Геометричні відстані

	C1	C2	O3	H4	H5	H6	H7
C1	0	1.400	2.190	1.022	1.023	1.022	2.106
C2	1.400	0	1.123	1.999	1.982	1.999	1.022
O3	2.190	1.123	0	2.349	2.708	2.995	1.859
H4	1.022	1.999	2.349	0	1.668	1.661	2.895
H5	1.023	1.982	2.708	1.668	0	1.668	2.562
H6	1.022	1.999	2.955	1.661	1.668	0	2.336
H7	2.106	1.022	1.859	2.895	2.566	2.336	0



Матриця відстаней

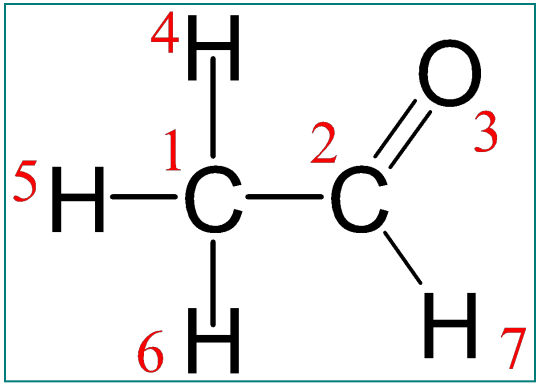


Топологічні відстані

	C1	C2	O3	H4	H5	H6	H7
C1	0	1	2	1	1	1	2
C2	1	0	1	2	2	2	1
O3	2	1	0	3	3	3	2
H4	1	2	3	0	2	2	3
H5	1	2	3	2	0	2	3
H6	1	2	3	2	2	0	3
H7	2	1	2	3	3	3	0

Таблиця зв’язків (Connection Table)

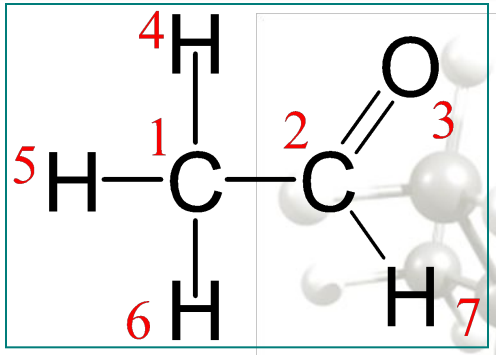
Відображення топології хімічної структури (складу та зв’язків між атомами) у формі таблиці



Список атомов	
1	C
2	C
3	O
4	H
5	H
6	H
7	H

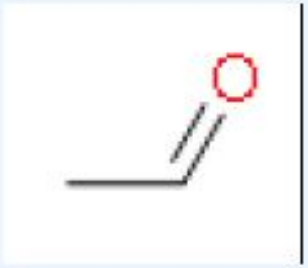
Список связей		
1-й атом	2-й атом	Порядок связи
1	2	1
2	3	2
2	7	1
1	4	1
1	5	1
1	6	1

Таблиця зв’язків (Connection Table)



№	Атом	Сосед № 1	Порядок связи	Сосед № 2	Порядок связи	Сосед № 3	Порядок связи	Сосед № 4	Порядок связи
1	C	2	1	4	1	5	1	6	1
2	C	1	1	3	2	7	1		
3	O	2	2						
4	H	1	1						
5	H	1	1						
6	H	1	1						
7	H	2	1						

Mol – файл (2D-структура)



3 атома

2 связи

список атомов

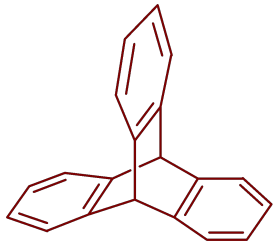
иные параметры

координаты x, y, z

```
  SMMXDraw51201020342D

  3  2  0  0  0  0  0  0  0  0999 V2000
    4.7059   -6.9865    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    5.5327   -6.9865    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    5.9461   -6.2705    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
  1  2  1  0  0  0  0
  2  3  2  0  0  0  0
M  END
```

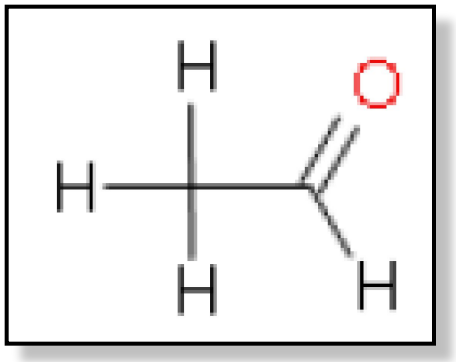
список связей



Mol – файл (2D-структура)

```
  SMMXDraw01201020342D

  7  6  0  0  0  0  0  0  0  0999 V2000
    4.9749   -5.2654    0.0000 H   0  0  0  0  0  0  0  0  0  0  0  0
    5.7945   -4.4244    0.0000 H   0  0  0  0  0  0  0  0  0  0  0  0
    5.8108   -5.9984    0.0000 H   0  0  0  0  0  0  0  0  0  0  0  0
    5.8016   -5.2654    0.0000 C   0  0  3  0  0  0  0  0  0  0  0  0
    7.0418   -5.9814    0.0000 H   0  0  0  0  0  0  0  0  0  0  0  0
    7.0418   -4.5494    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
    6.6284   -5.2654    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
  4  1  1  0  0  0  0
  4  2  1  0  0  0  0
  4  3  1  0  0  0  0
  7  4  1  0  0  0  0
  7  5  1  0  0  0  0
  7  6  2  0  0  0  0
M  END
```



Mol – файл (2D-структура)

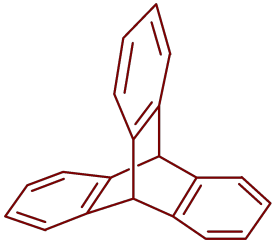
etanol.mol - AkelPad

Файл Правка Вид Настройки Справка

```
  Mrv0541 10221321302D

  6  5  0  0  0  0               999 V2000
   -2.2688    0.9429    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -3.0938    0.9429    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   -2.2688    0.1179    0.0000 H   0  0  0  0  0  0  0  0  0  0  0  0
   -1.4438    0.9429    0.0000 H   0  0  0  0  0  0  0  0  0  0  0  0
   -2.2688    1.7679    0.0000 H   0  0  0  0  0  0  0  0  0  0  0  0
   -3.5063    1.6573    0.0000 H   0  0  0  0  0  0  0  0  0  0  0  0
  1  2  1  0  0  0  0
  1  3  1  0  0  0  0
  1  4  1  0  0  0  0
  1  5  1  0  0  0  0
  2  6  1  0  0  0  0
M  END
```

1:1 Ins Win 1251 (ANSI - кириллица)



Структури Маркуша 1924

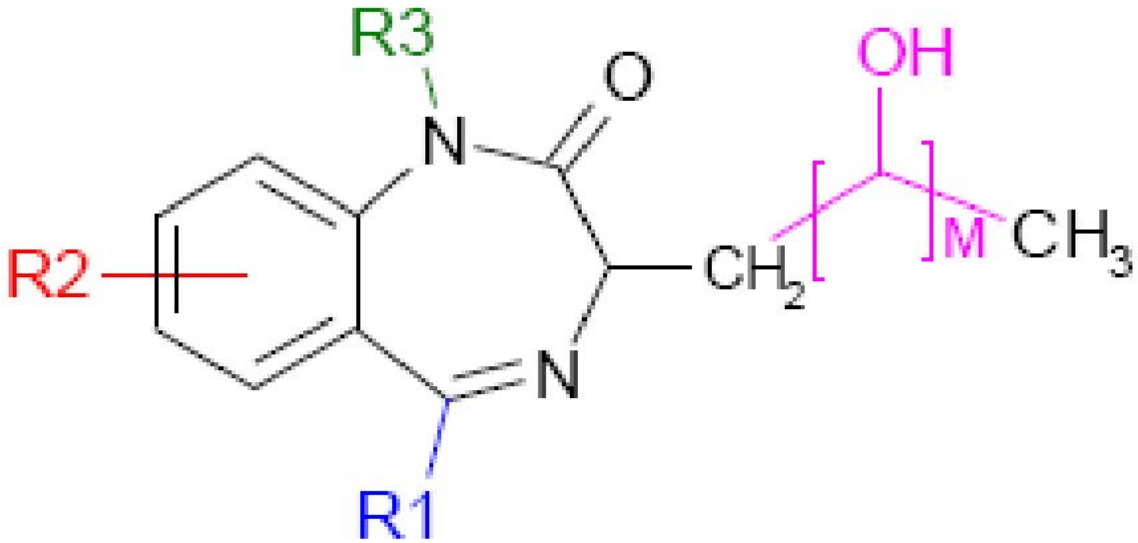
Євген А. Маркуш
1887 - 1968

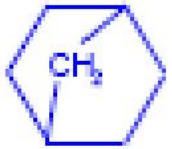
Pharma Chemical Corporation 1919

US patent 1506316

Спосіб загального відображення серії хімічних сполук, що мають спільну основну структурну частину, та відрізняються замісником чи структурним фрагментом

Структури Маркуша



R_1 = propyl / $CH_3CH(Br)CH_3$ / OMe*/ ← 

R_2 = F

M = 2-5

R3 = alkyl C1-6

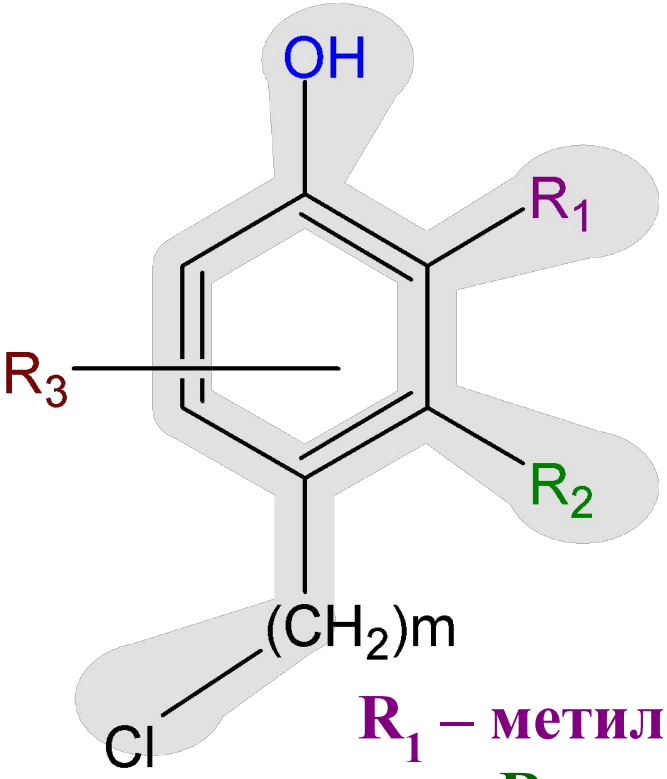
Структури Маркуша

s-варіація – варіювання замісника: різні можливі замісники наводяться у вигляді окремого переліку значень R-групи;

p-варіація – варіювання положення: використовується для позначення можливих положень замісників у структурі;

f-варіація – варіювання частоти: для позначення кількості багаторазової появи даної групи у структурі;

h-варіація – варіювання гомологів: для позначення груп, що відрізняються на гомологічну різницю.



R_1 – метил або етил;

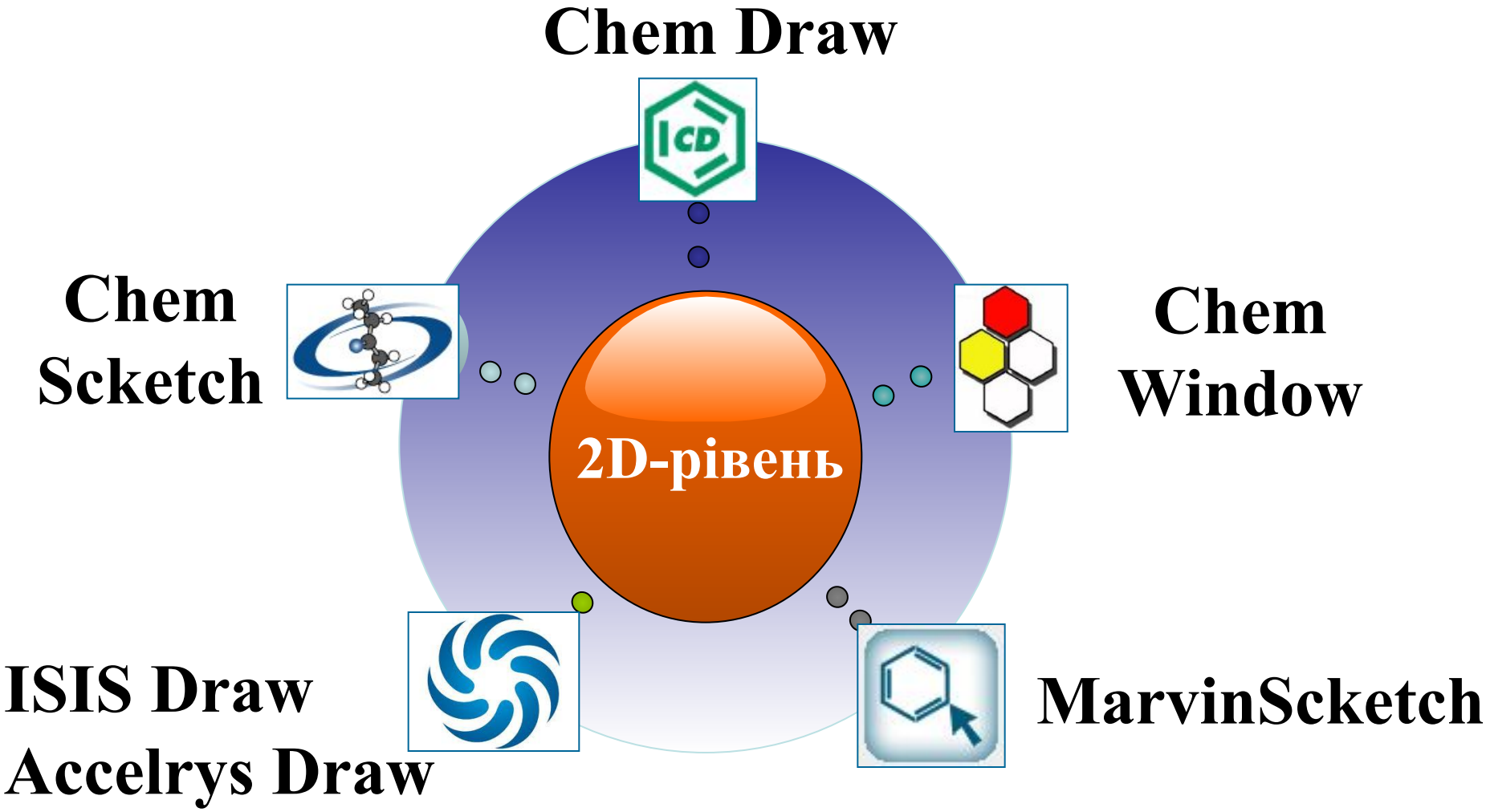
R_2 – алкіл;

R_3 – аміно-група;

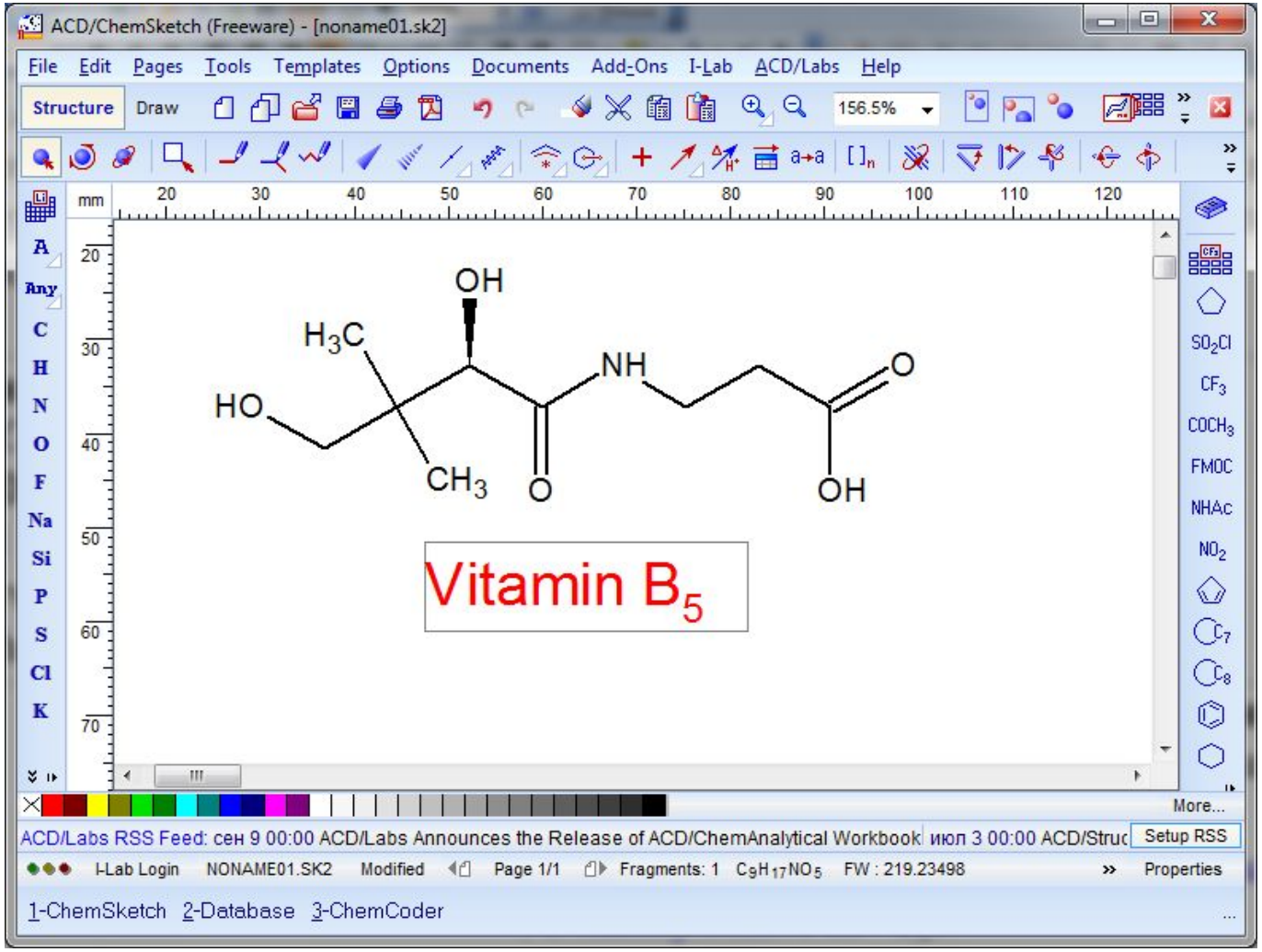
***f-варіація*:**

m = 1, 2, 3.

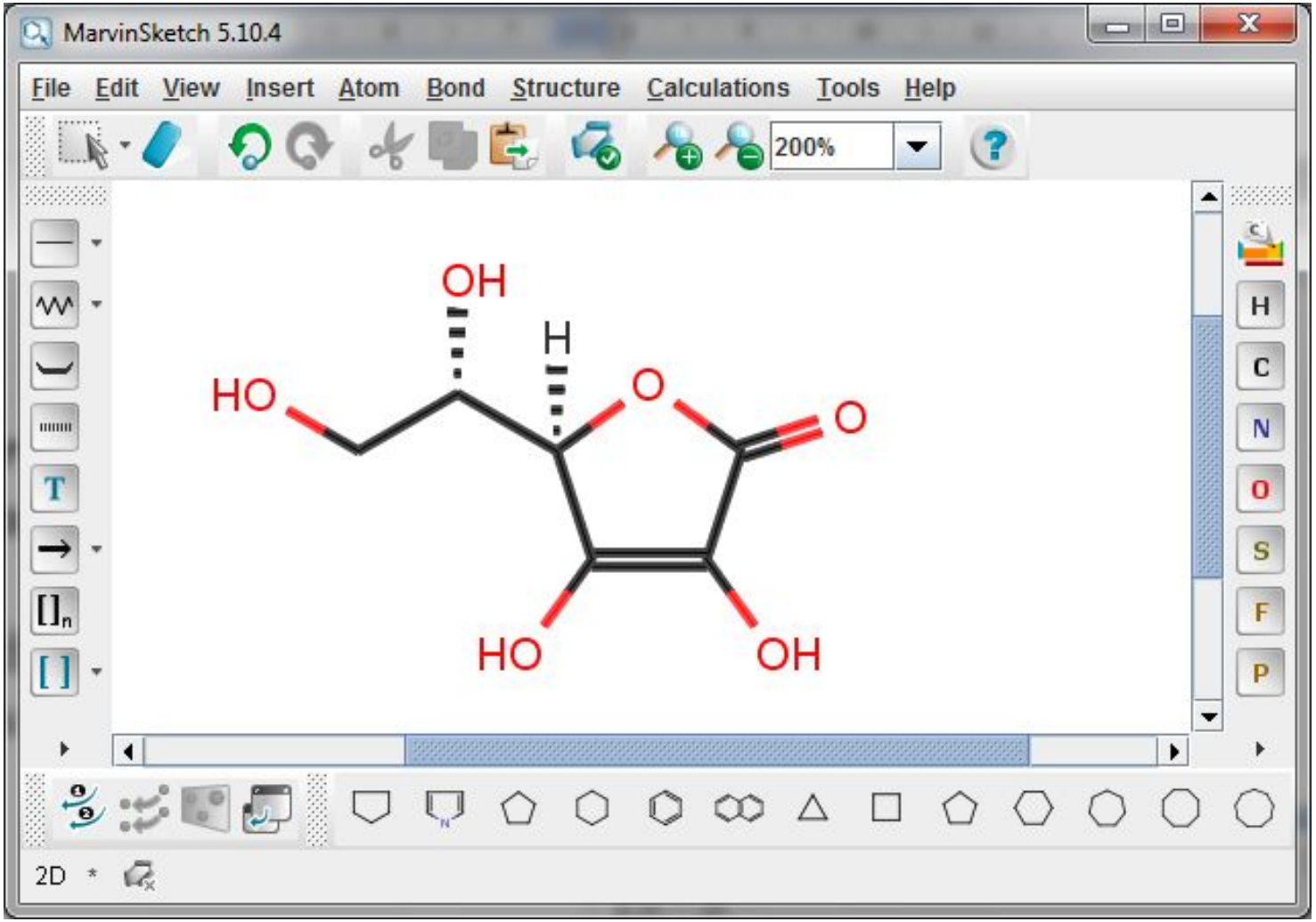
Редактори хімічних формул



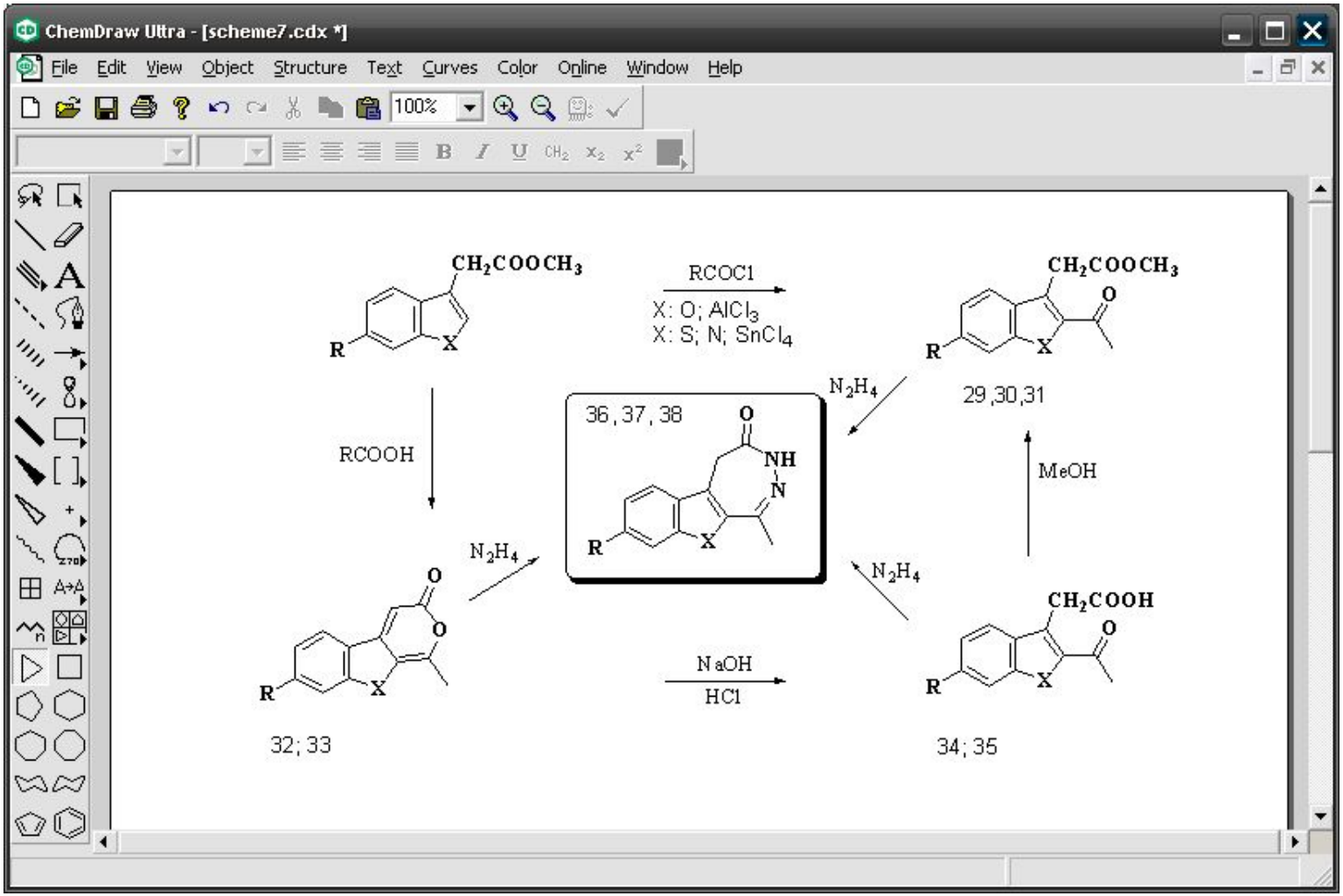
Редактори хімічних формул



Редактори хімічних формул

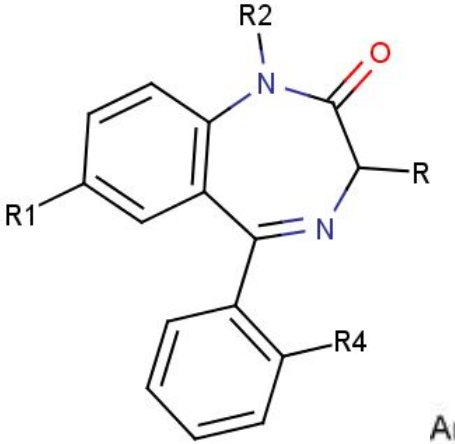


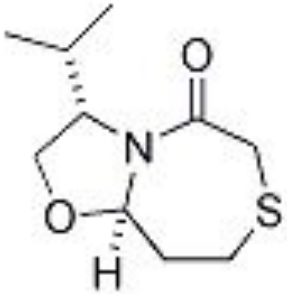
Редактори хімічних формул

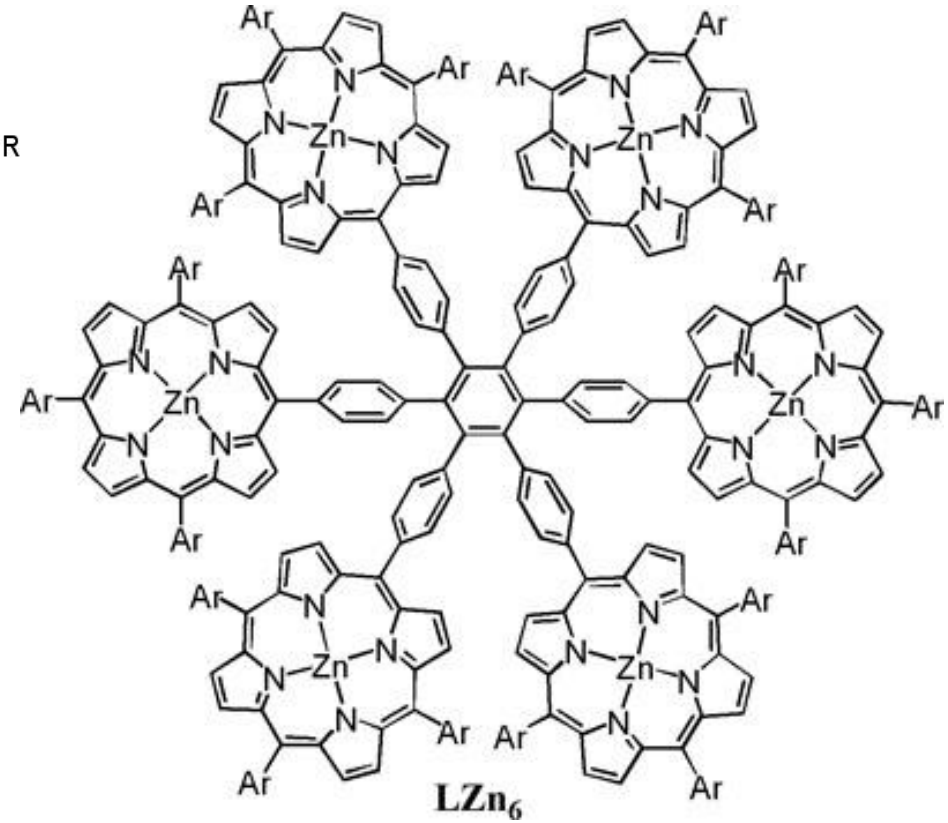


Редактори хімічних формул: основні можливості

Створення структурних формул різної складності

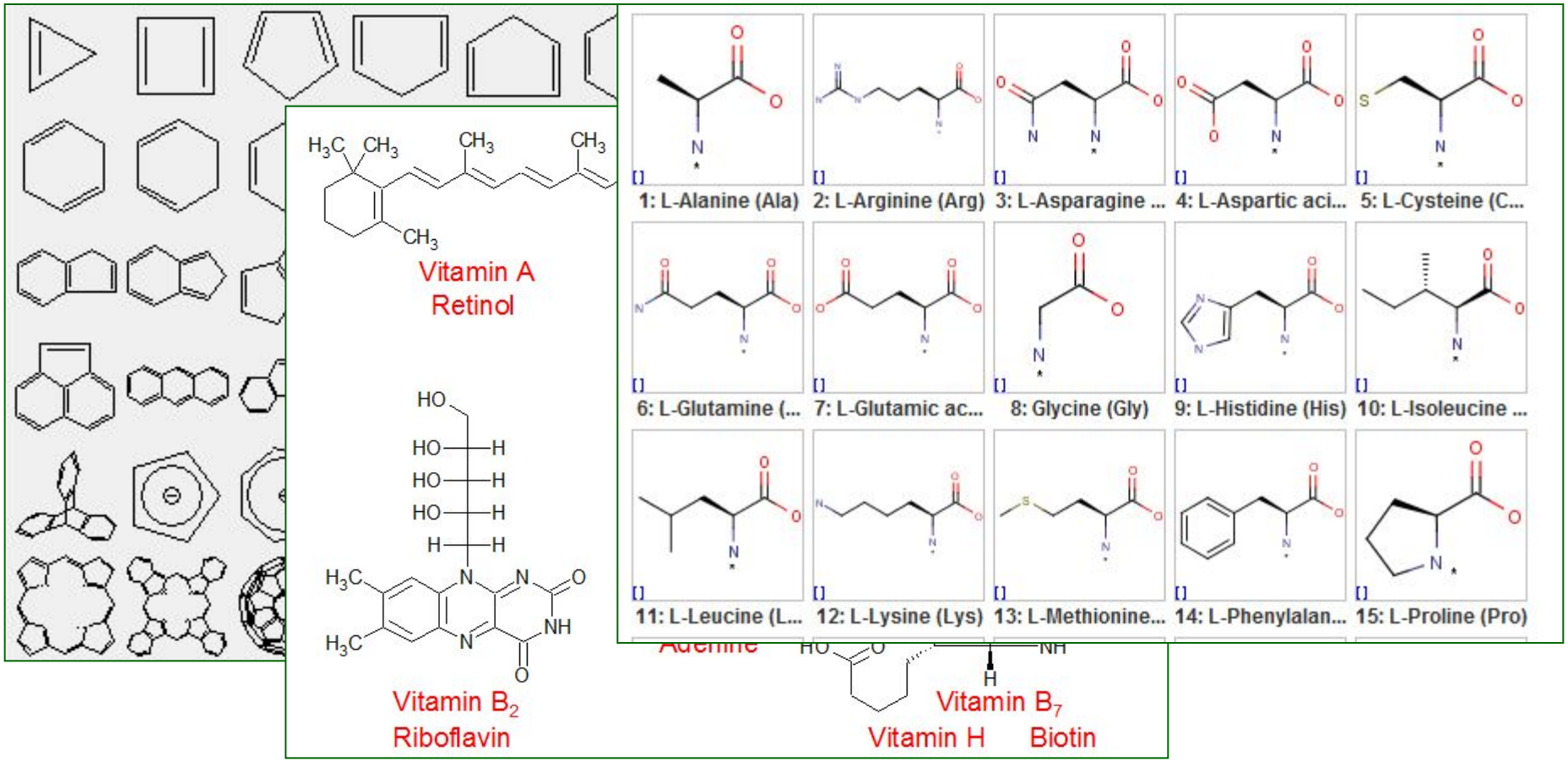






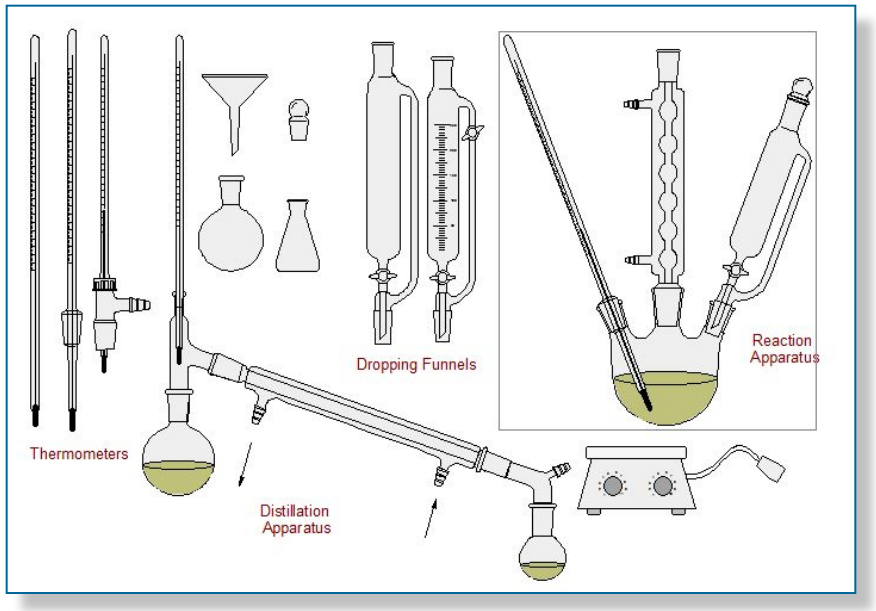
Редактори хімічних формул: основні можливості

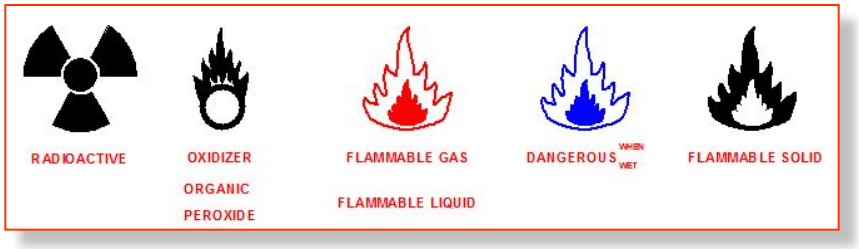
Використання шаблонів хімічних структур

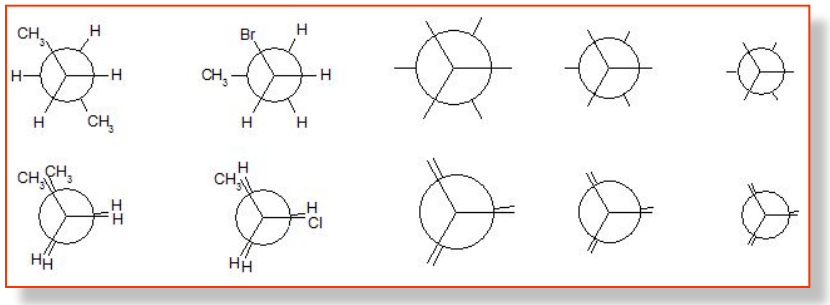


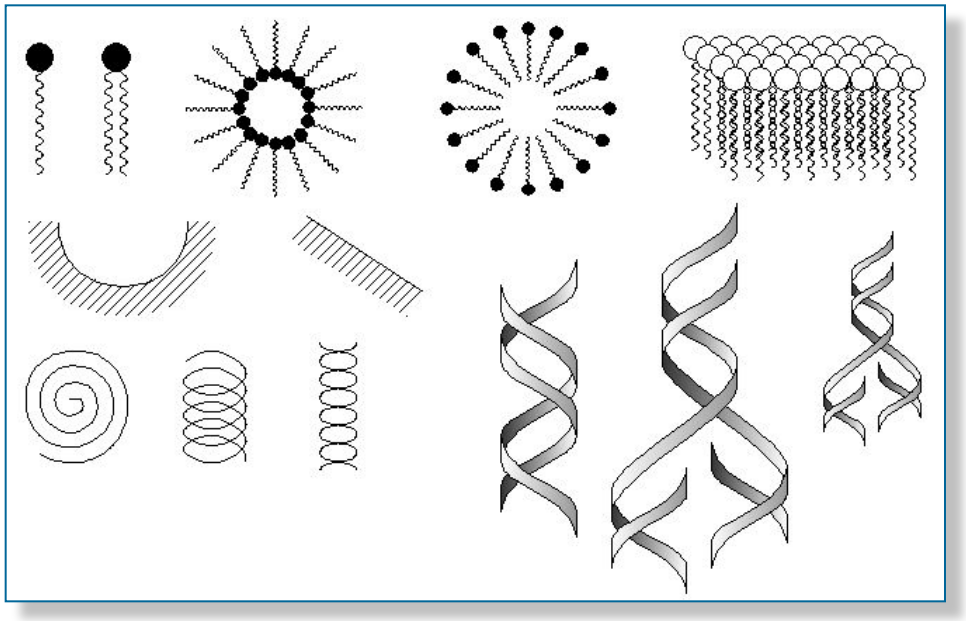
Редактори хімічних формул: основні можливості

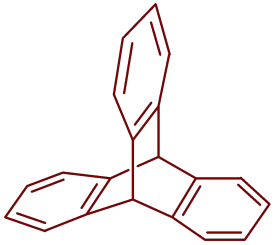
Використання шаблонів хімічної графіки





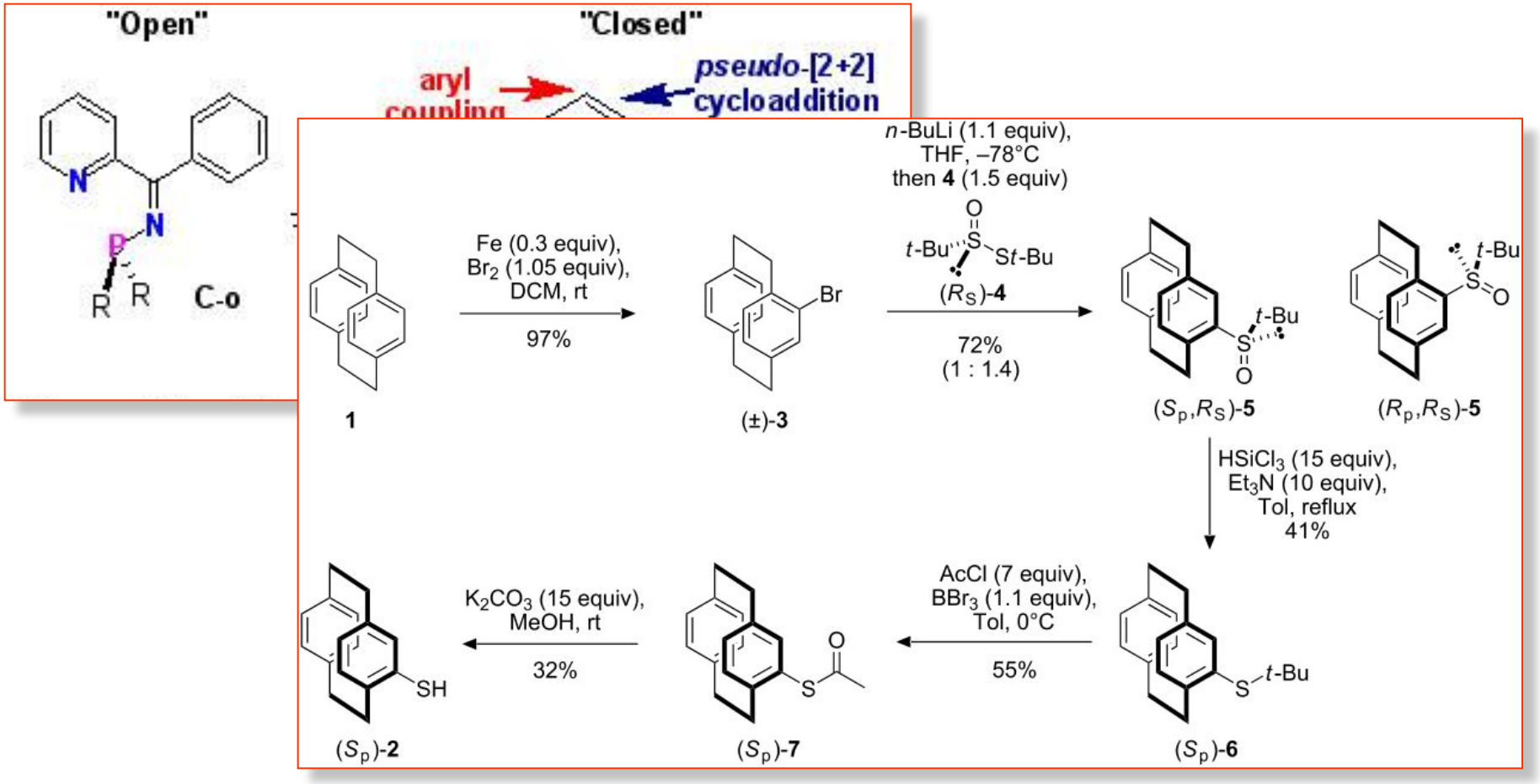


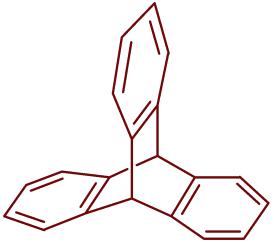




Редактори хімічних формул: основні можливості

Створення схем хімічних реакцій різної складності





Редактори хімічних формул: основні можливості

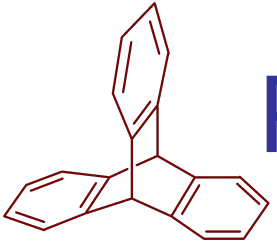
Копіювання та вбудовування об'єктів у документи інших програм
Імпорт - експорт даних в різних форматах

skc, mol, ct, wmf, rxn, sdf

Gif, bmp, png, tiff

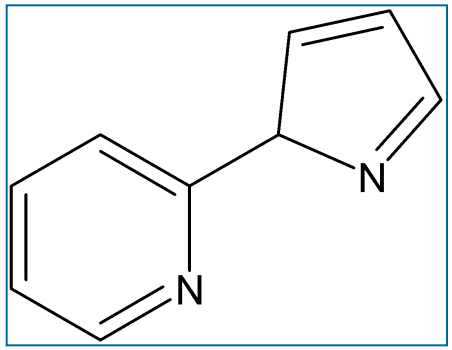
pdf

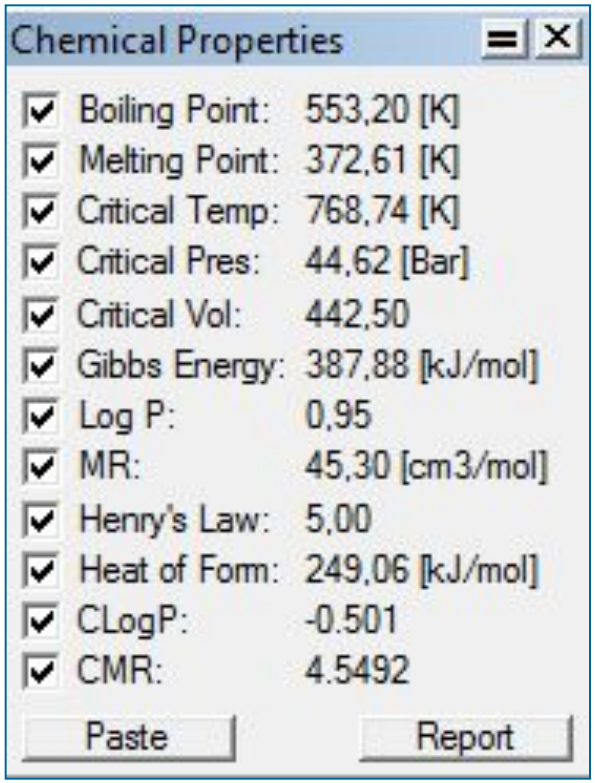
Cdx, jdx, sk2, mrv

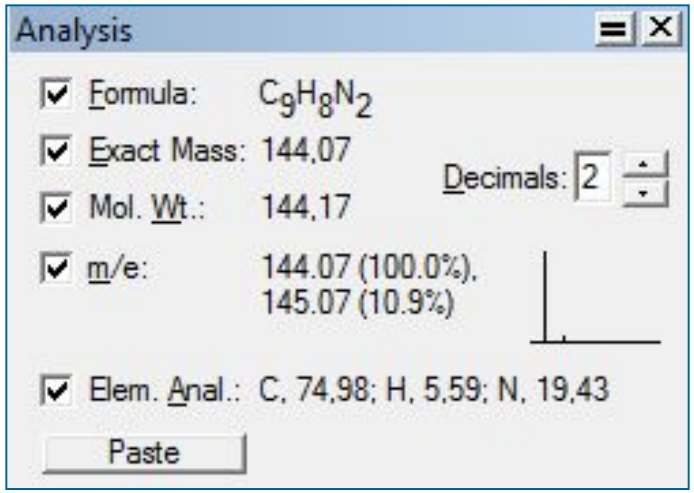


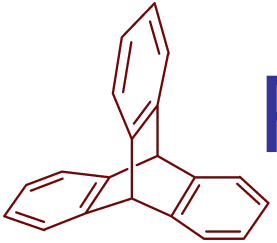
Редактори хімічних формул: основні можливості

Генерування характеристик речовин на основі їх структурних формул

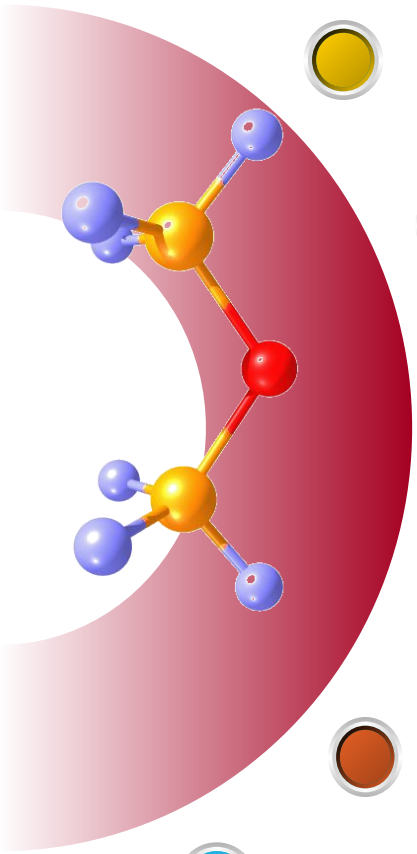








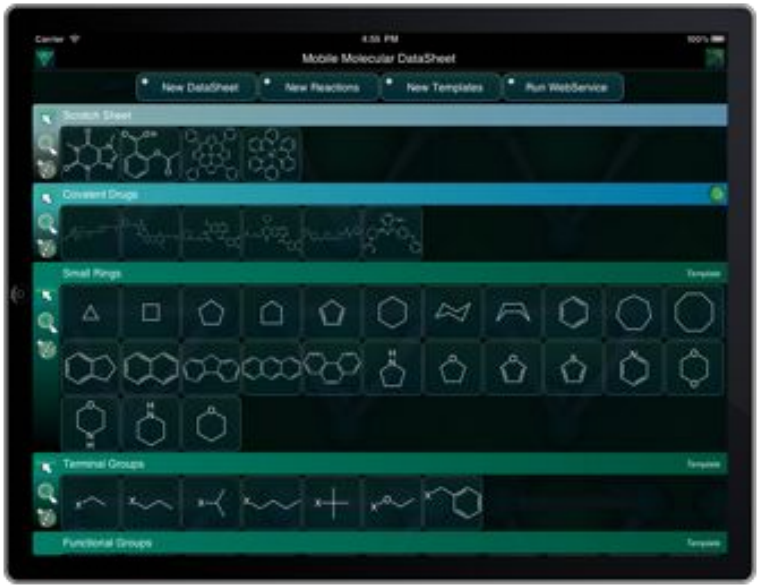
Редактори хімічних формул: додаткові можливості

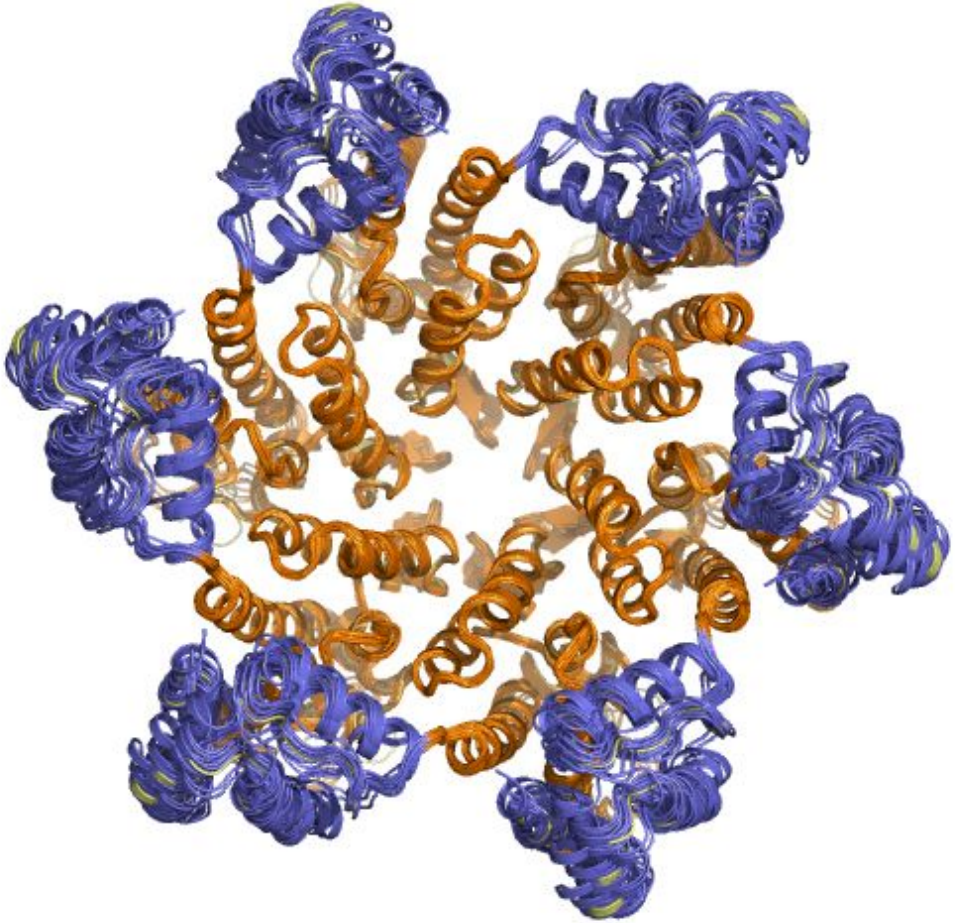


- **Генерування назви за ІЮПАК номенклатурою**
- **Генерування кодів SMILES, InChI, SLN**
- **Генерування структур таутомерів**
- **Генерування 3D моделей**
- **Генерування ЯМР спектрів**

Мобільні засоби хімічної інформатики

Mobile Molecular DataSheet



Мобільні засоби хімічної інформатики

MolPrime