

А. А. Рагойша

# Информационные технологии в химии

2-й семестр

Избранные элементы хемоинформатики

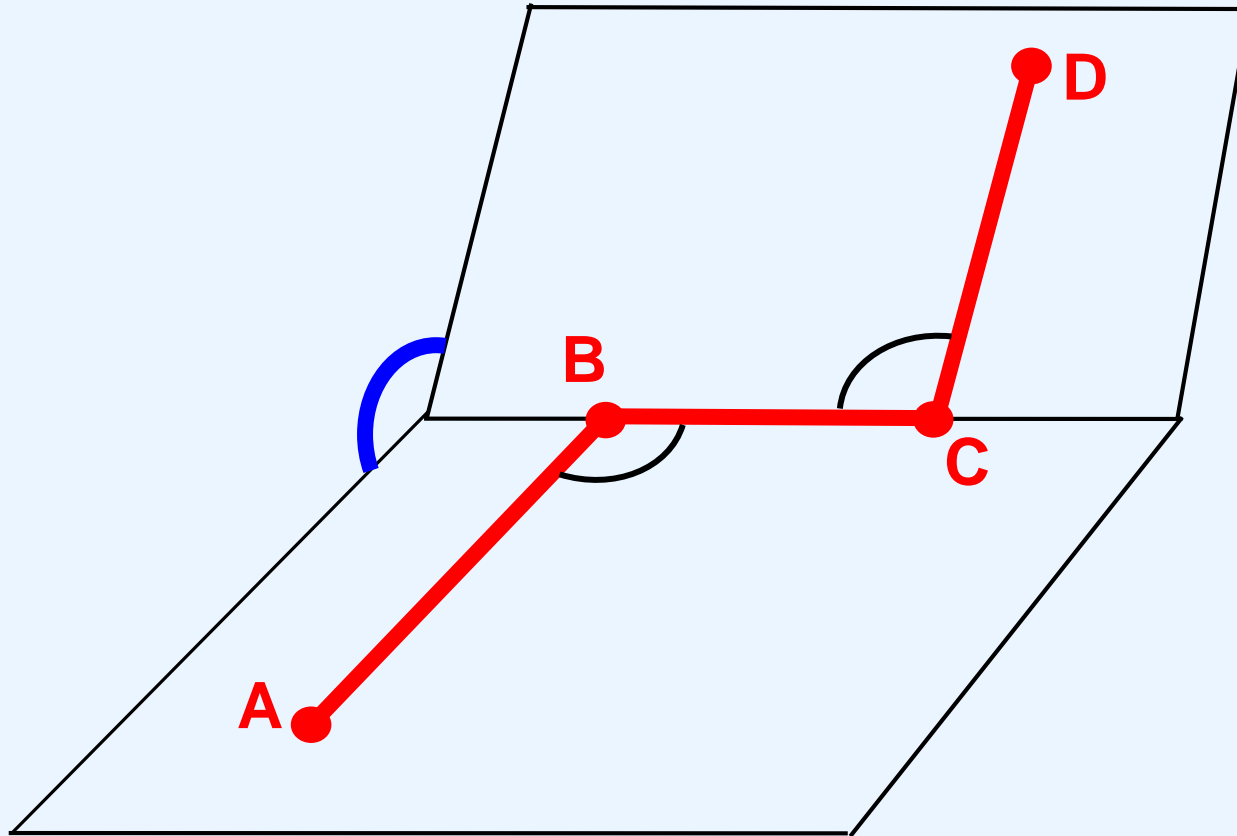
# Литература

В классе: \Материалы для студентов\

1. Andrew R. Leach, Valerie J. Gillet.  
An Introduction to Chemoinformatics. – Springer, 2007.
  2. Chemoinformatics: A Textbook.  
Edited by Johann Gasteiger and Thomas Engel. – Wiley-VCH, 2003.
- 
1. B. A. Bunin, B. Siesel, G. A. Morales, J. Bajorath.  
Chemoinformatics: Theory, Practice, & Products. – Springer, 2007.
  2. Chemical Information for Chemists: A Primer.  
Edited by Judith N. Currano and Dana L. Roth. - RSC Publishing, Cambridge, UK, 2014.

Хранение информации о  
трехмерной кристаллической структуре

# Внутренние координаты



Длина связи

Угол между связями

Двугранный угол

# Файлы в формате CIF

## Crystallographic Information File

**CIF** – стандартный формат обмена кристаллографической информацией, разработанный Международным союзом кристаллографии.

Файлы в формате CIF могут содержать в себе:

- информацию о **пространственном** расположении атомов (т.е. координаты атомов и в явной форме межатомные расстояния, значения валентных углов) – причем это данные **экспериментальные**, а не рассчитанные из моделей;
- **кристаллографические** параметры;
- **рентгенограммы**;
- **текстовый** материал.

# Структура CIF-файла

Используются только ASCII символы.

Разграничение: форма – содержание.

Каждый элемент информации  
в формате:

Имя элемента (тэг) – значение.

Файл состоит из блоков.

Табличные данные в блоках `loop_`

Структура блока:

упорядоченный список имен и  
упорядоченный список значений.

```
_chemical_formula_sum  'H10 Cl2 O10 U2'  
_chemical_formula_weight      717.04  
_symmetry_cell_setting      monoclinic  
_symmetry_space_group_name_H-M  'P 21/n'  
_symmetry_space_group_name_Hall '-P 2yn'  
loop_  
  _symmetry_equiv_pos_as_xyz  
  'x, y, z'  
  '-x+1/2, y+1/2, -z+1/2'  
  '-x, -y, -z'  
  'x-1/2, -y-1/2, z-1/2'  
_cell_length_a      10.712(2)  
_cell_length_b      6.1212(12)  
_cell_length_c      17.662(4)  
_cell_angle_alpha      90.00  
_cell_angle_beta      95.47(3)  
_cell_angle_gamma      90.00  
_cell_volume      1152.8(4)  
_cell_formula_units_Z      4
```

## Таблица в CIF-файле

```
245 loop_  
246   _geom_angle_atom_site_label_1  
247   _geom_angle_atom_site_label_2  
248   _geom_angle_atom_site_label_3  
249   _geom_angle_site_symmetry_1  
250   _geom_angle_site_symmetry_3  
251   _geom_angle  
252   _geom_angle_publ_flag  
253   O1 U1 O2 . . 177.7(4) ?  
254   O1 U1 O10 . . 91.5(4) ?  
255   O2 U1 O10 . . 87.9(4) ?  
256   O1 U1 O9 . . 88.8(4) ?  
257   O2 U1 O9 . . 88.9(4) ?
```

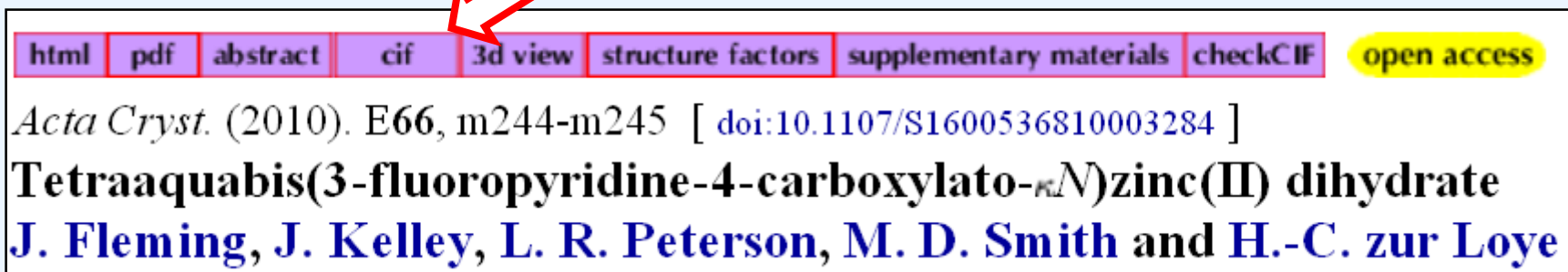
СПИСОК  
ИМЕН

СПИСОК  
ЗНАЧЕНИЙ

угол O2-U1-O9 равен 88,9°

# Примеры CIF на сайтах журналов

- В явной форме:



A screenshot of a journal article page. At the top, there is a navigation bar with several links: 'html', 'pdf', 'abstract', 'cif', '3d view', 'structure factors', 'supplementary materials', 'checkCIF', and 'open access'. A red arrow points to the 'cif' link. Below the navigation bar, the article information is displayed: 'Acta Cryst. (2010). E66, m244-m245 [ doi:10.1107/S1600536810003284 ]'. The title of the article is 'Tetraaquabis(3-fluoropyridine-4-carboxylato- $\kappa$ N)zinc(II) dihydrate'. The authors are listed as 'J. Fleming, J. Kelley, L. R. Peterson, M. D. Smith and H.-C. zur Loye'.

- В неявной форме:



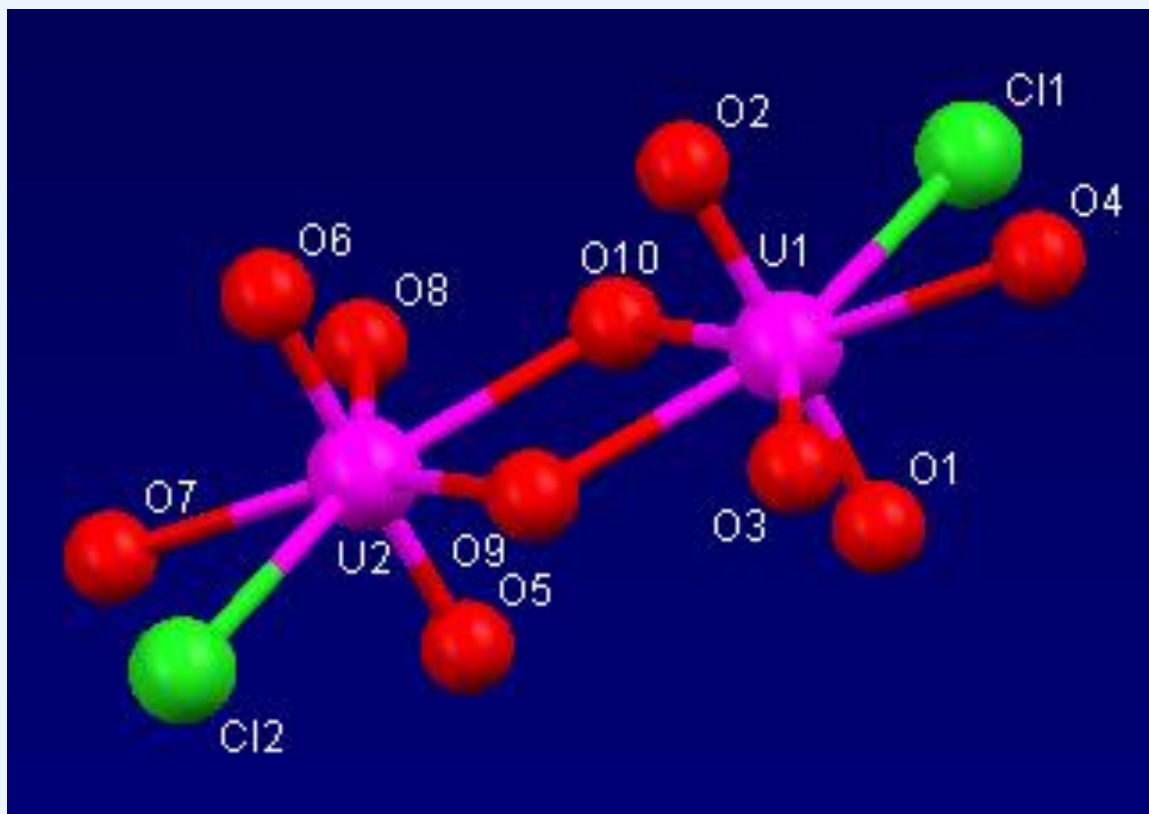
A screenshot of a journal article page. The article title is 'Crystal structure of trihydroxycopper formate,  $\text{Cu}_2(\text{OH})_3(\text{HCOO})$ '. The authors are listed as 'Harald Euler, Bruno Barbier, Armin Kirfel, Stefanie Haseloff and Gerhard Eggert'. Below the title, there are two links: 'NCS [1267-2726](#)' and 'PDF [1267-2726](#)'. A red arrow points to the 'NCS' link. A callout box below the screenshot contains the text 'NCS = New Crystal Structure'.

(Конечно же, CIF-файлы есть и в специализированных базах данных)



# Фрагмент структуры, CIF-формат

Визуализация фрагмента



Параметры структуры

Atom1	Atom2	Type	Length
U1	Cl1	Unknown	2.751(3)
U1	O1	Unknown	1.746(9)
U1	O2	Unknown	1.790(9)
U1	O3	Unknown	2.396(9)
U1	O4	Unknown	2.49(1)
U1	O9	Unknown	2.382(9)

Atom1	Atom2	Atom3	Angle
Cl1	U1	O1	91.0(3)
Cl1	U1	O2	91.1(3)
Cl1	U1	O3	141.7(2)
Cl1	U1	O4	72.9(2)
Cl1	U1	O9	143.1(2)

