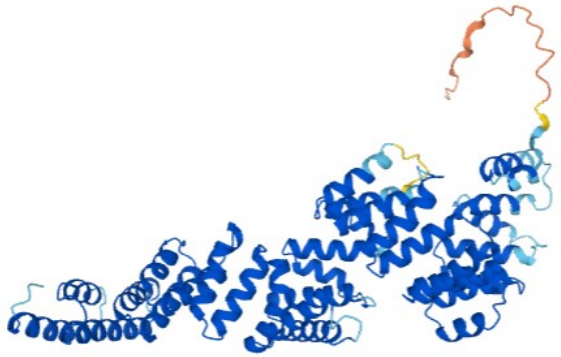
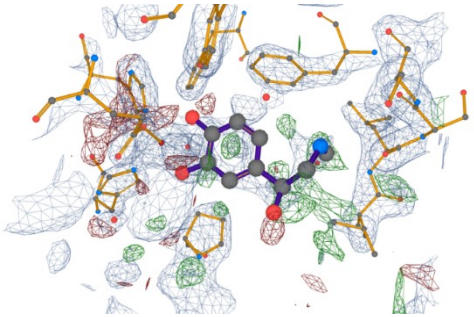
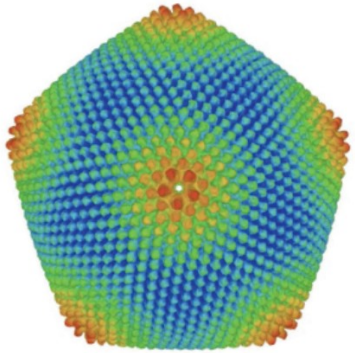
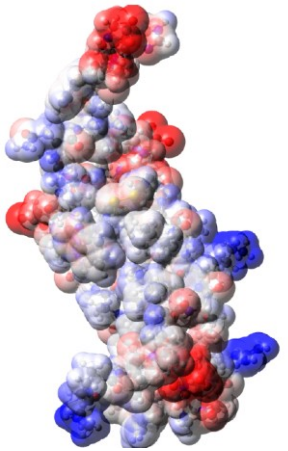
Computer model of a molecule



Radka Svobodová

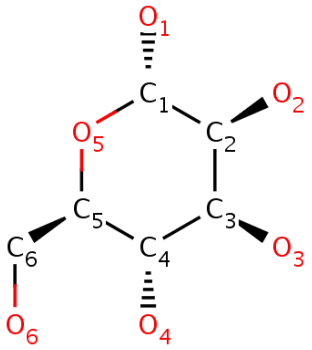


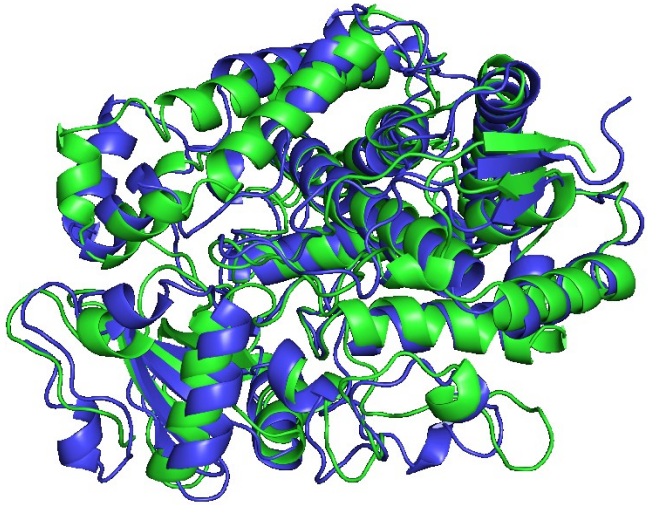




Content

- **Introduction:** concept of chemoinformatics, content of the subject, history of the field
- **Computer model of a molecule:** 1D, 2D and 3D structure, molecule representation using graph and matrix
- **2D structure (topology) of a molecule:**
  - writing a molecule using a string (SMILES, InChi, InChiKey)
  - **Molecular graphs:** Isomorphism and canonical indexing
- **3D structure (geometry) of the molecule:**
  - representation using Cartesian and internal coordinates, data formats, geometry comparison

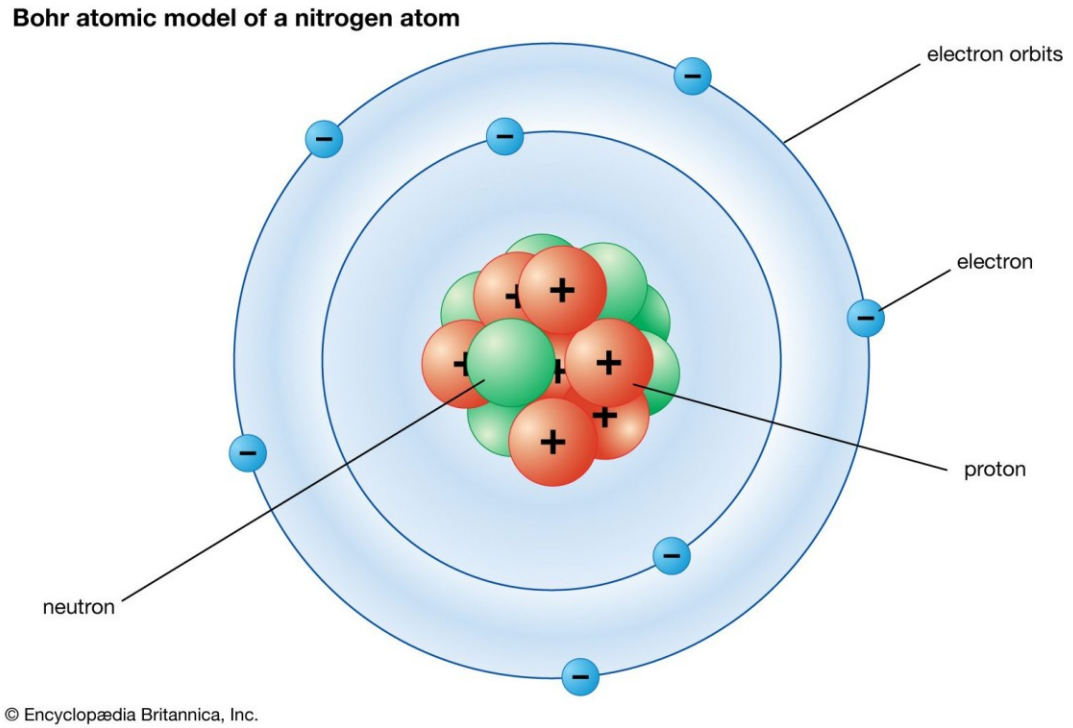




Basic chemical terms I

- Atom: basic building block from which substances are formed

- Structure of an atom:
  - Atom core: protons (positive charge), neutrons (no charge)
  - Electron shell: electrons (negative charge)

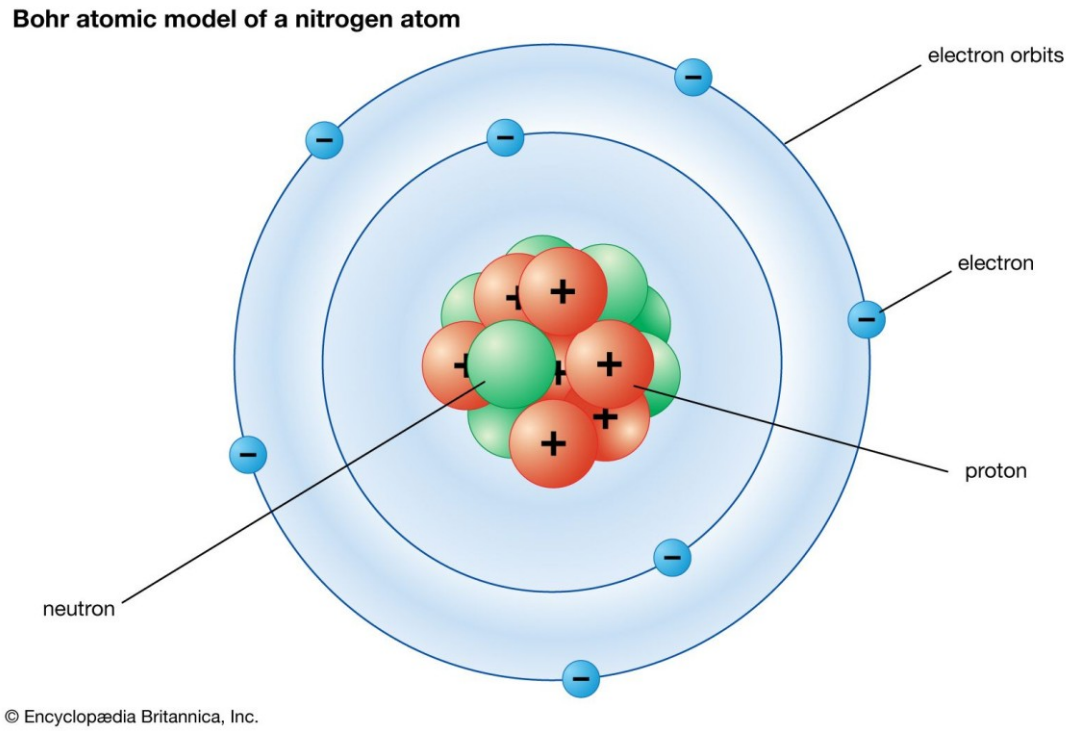


Basic chemical terms II

- All systems tend to occupy the state with the lowest possible total energy.

Basic chemical terms III

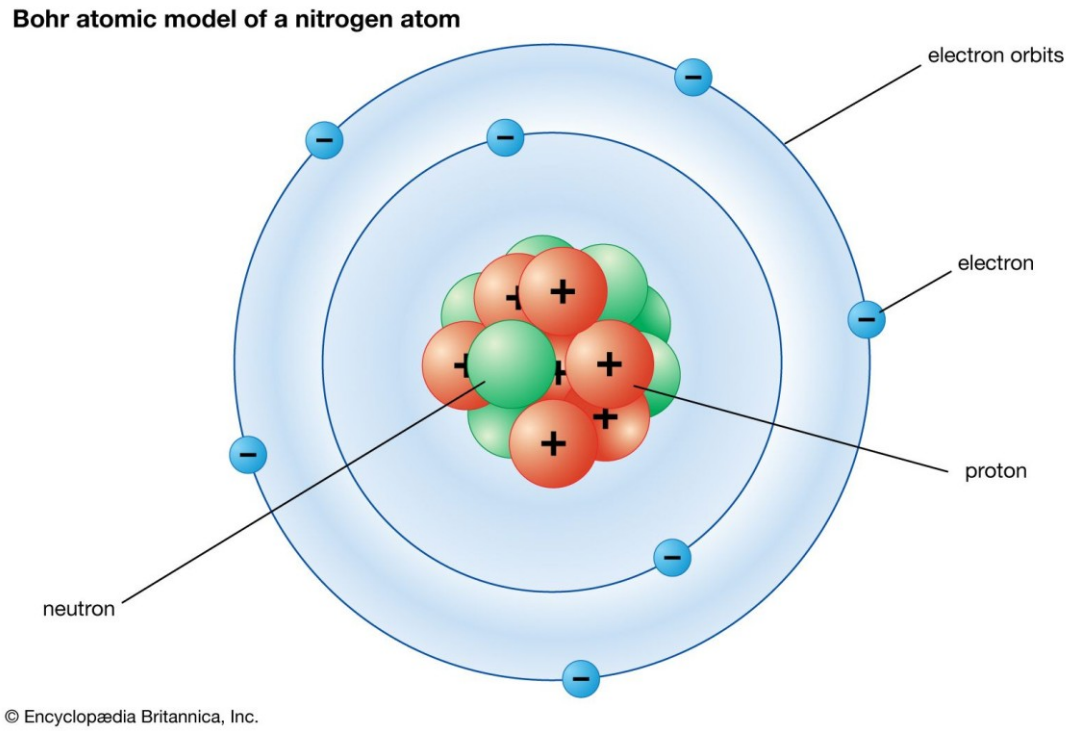
- The space of the electron shell can be divided into so-called **layers**.
- All electrons in a layer have the same energy value (this energy is characteristic of that layer).
- The further the layer is from the nucleus, the higher the energy of the electrons in it.
- The electrons in the electron shell therefore fill first the layer closest to the nucleus (the most energetically favorable), then the second closest, and so on.



Basic chemical terms IV

- The non-empty layer that is farthest from the core is called the **valence layer**.
- In this layer are the so-called **valence electrons**.

These **valence electrons** are the subject of the study of chemistry because they can participate in chemical bonding.



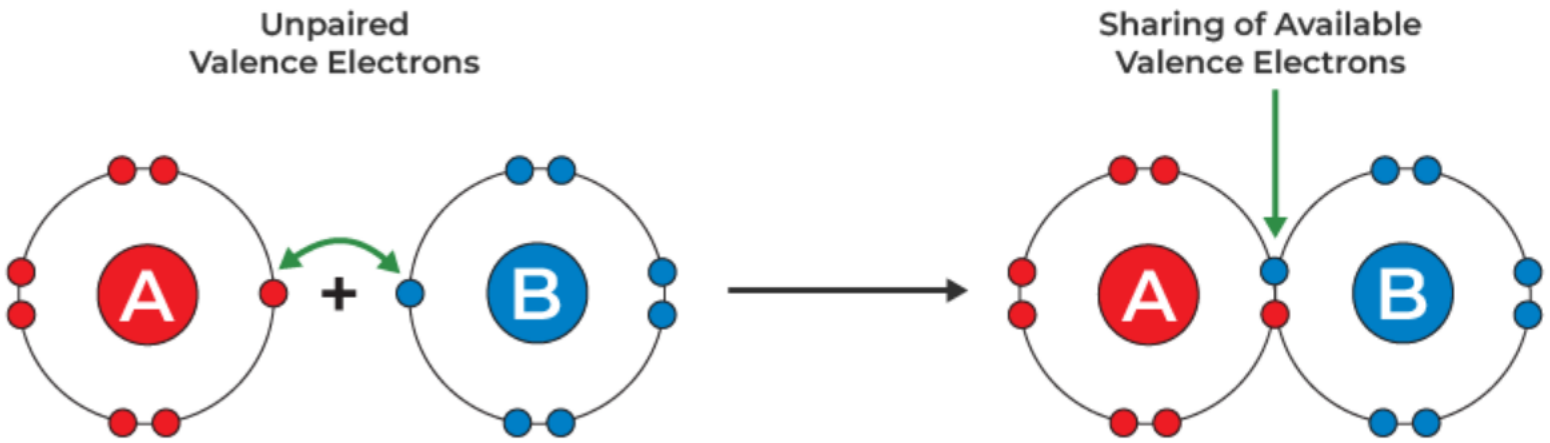
Basic chemical terms V

Chemical bond:

Two atoms come together at a sufficiently small distance (bonding distance) => overlapping of their electron shells.

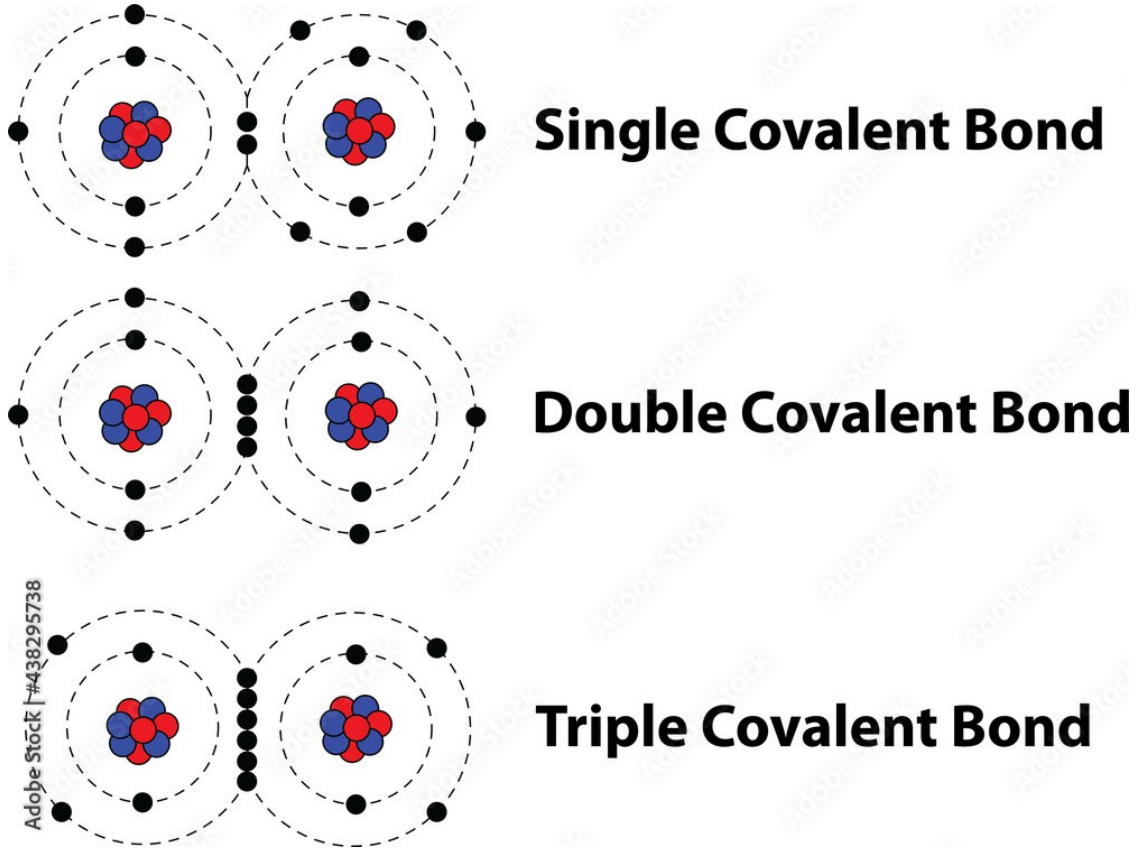
The valence electrons of both atoms change their trajectories.

If the resulting system has a lower energy than the original, the atoms remain at bonding distance => chemical bonding is formed.



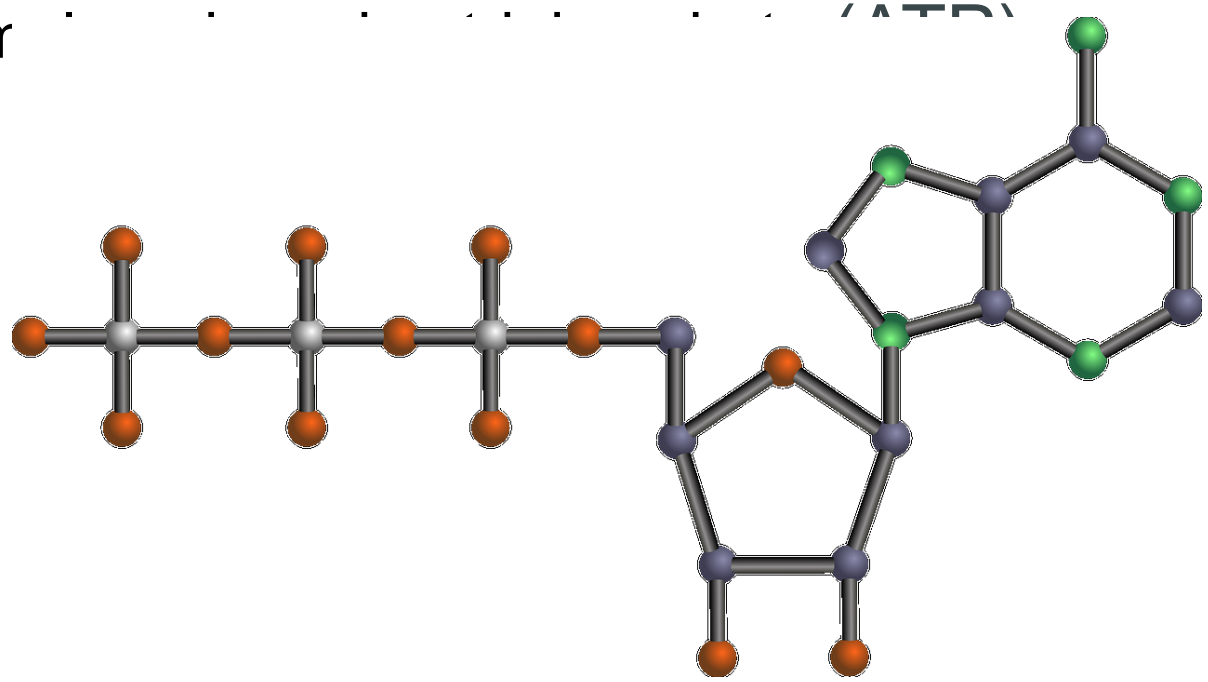
Basic chemical terms VI

- Bond order:
  - single bond: two valence electrons are involved (binding electron pair)
  - double bond: two bonding electron pairs are involved
  - Triple bond: analogous
  - higher multiplicities do not occur in real chemical environments
  - Aromatic bond: when single and double bonds alternate, the electrons are delocalised among them. These bonds have properties between single and double bonds.



Basic chemical terms VII

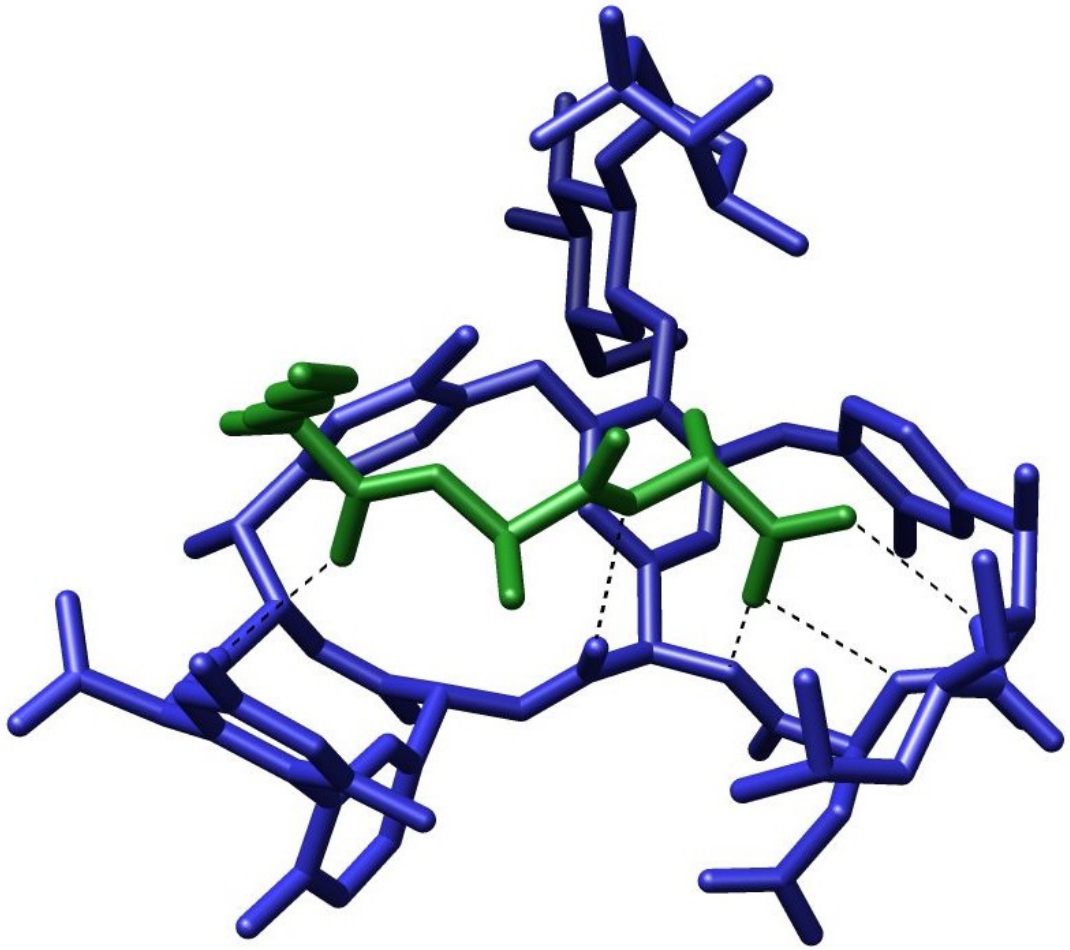
- Molecule: A system of atoms joined together by bonds to form a single unit. The basic structural unit of a substance. The carrier of the chemical properties of a substance.
- Exam



Basic chemical terms VIII

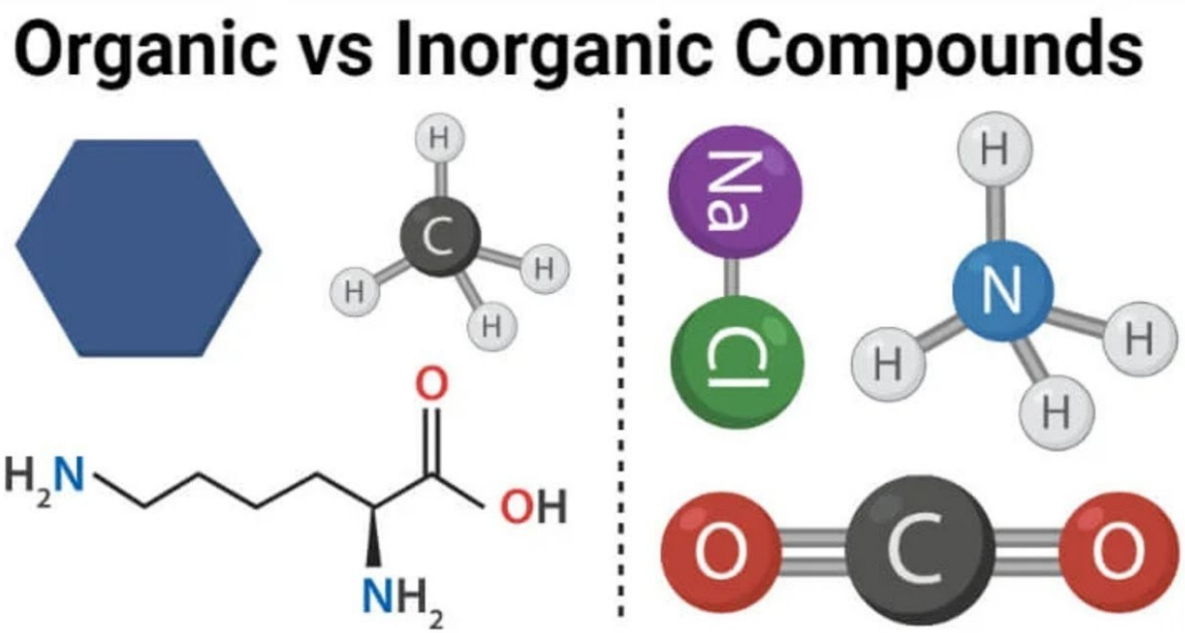
- Molecular system: A system containing one or more molecules.

Example:



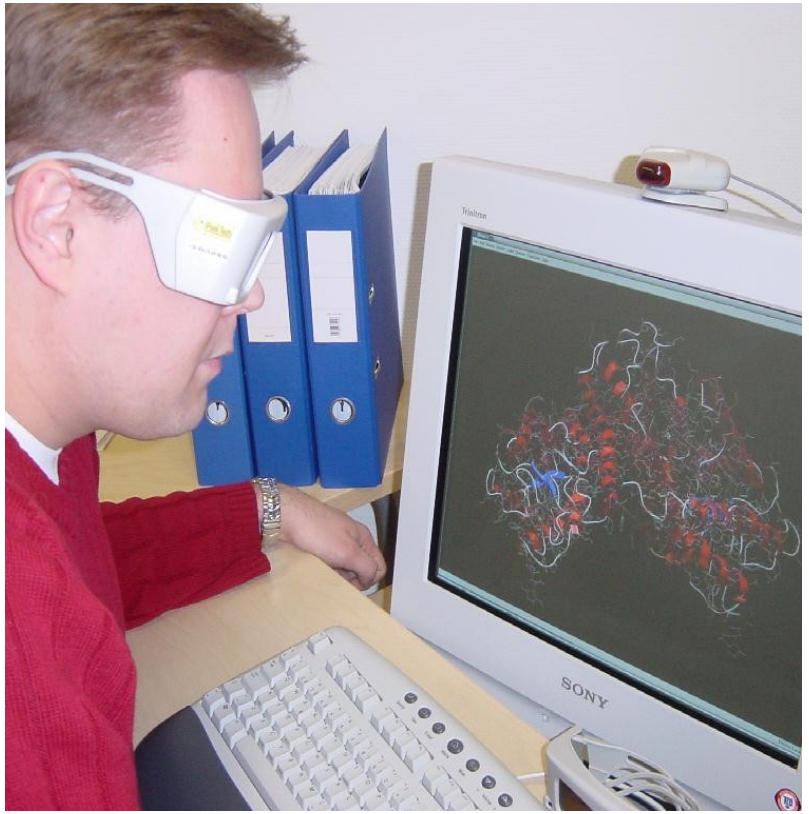
Basic chemical terms IX

- Organic molecules:
  - Their main component is carbon, the only element that is able to form longer chains of the (-C-)n type, n > 10.
  - This property of carbon allows the formation of complex molecules - the building blocks of living systems.
  - Organic molecules also contain elements: H, O, S, N, F, Cl, Br, I
- Inorganic molecules:
  - All molecules, that are not organic.



How to describe a molecule in a computer?

- **Find out which information describes the molecule**
- **Write them into the computer**



Which information describes the molecule?

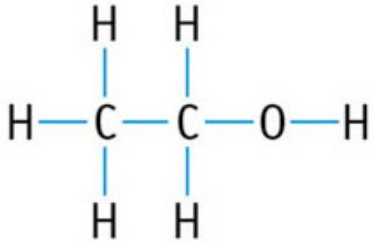
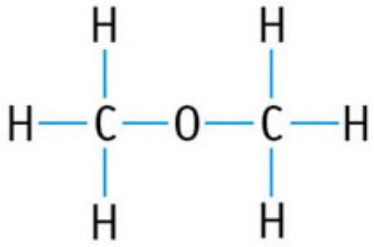
Number of atoms?

Ethanol	C_2H_6O
Dimethyl ether	C_2H_6O

Which information describes the molecule?

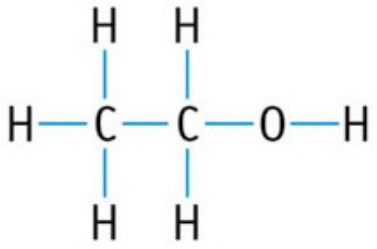
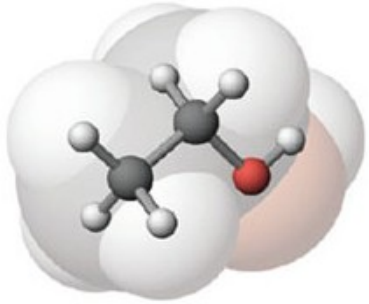
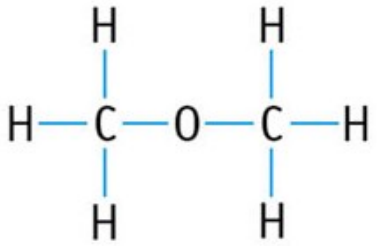
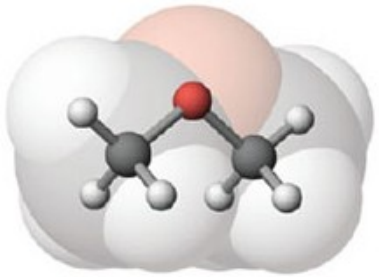
Number of atoms? **Not enough**

Number of atoms and positions of bonds?

Ethanol	C_2H_6O	CH_3CH_2OH	
Dimethyl ether	C_2H_6O	CH_3OCH_3	

Which information describes the molecule?

Number of atoms?	**Not enough**
Number of atoms and positions of bonds?	**Better**
Number of atoms, positions of bonds and positions of atoms in 3D space?	**Yes**

Ethanol	C_2H_6O	CH_3CH_2OH		
Dimethyl ether	C_2H_6O	CH_3OCH_3		

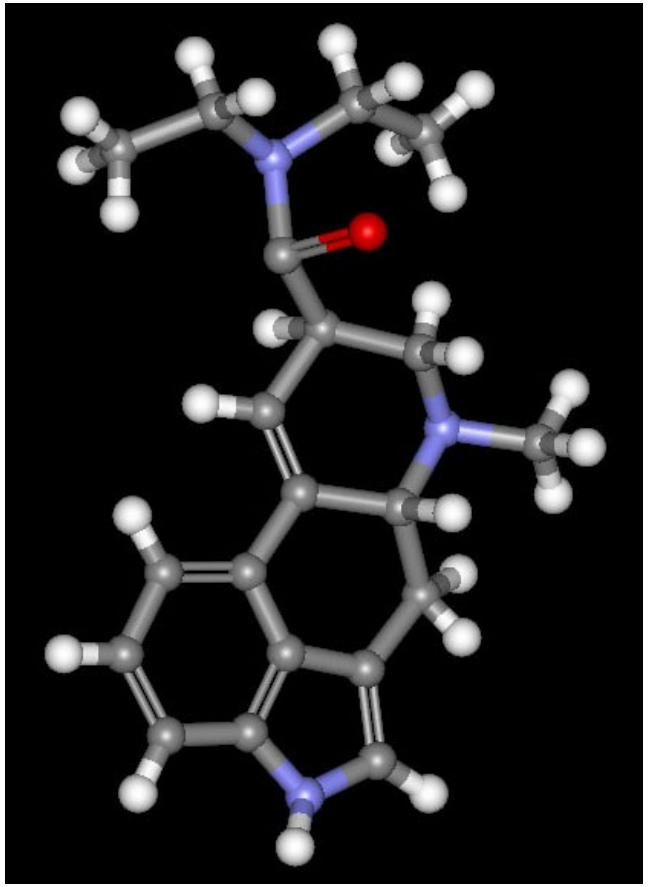
Model of molecule for computer processing

Atoms:

- Points in space
- Chemical symbol of the element listed for each

Bonds:

- Pairs of atoms that are bonded
- Bond order



Description of a molecule in a computer

number of atoms

number of bonds

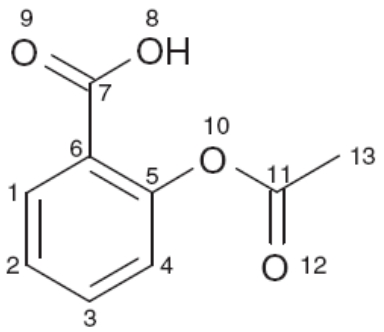
the first atom is a carbon

```
  -ISIS-  09270222202D

 13 13  0  0  0  0  0  0  0  0999 V2000
   -3.4639   -1.5375    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -3.4651   -2.3648    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -2.7503   -2.7777    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -2.0338   -2.3644    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -2.0367   -1.5338    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -2.7521   -1.1247    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -2.7545   -0.2997    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -2.0413    0.1149    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   -3.4702    0.1107    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   -1.3238   -1.1186    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   -0.6125   -1.5292    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -0.6167   -2.3542    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
    0.1000   -1.1125    0.0000 C   0  0  0  0  0  0  0  0  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
  7  8  1  0  0  0  0
  7  9  2  0  0  0  0
  4  5  1  0  0  0  0
  5 10  1  0  0  0  0
  2  3  1  0  0  0  0
 10 11  1  0  0  0  0
  5  6  2  0  0  0  0
 11 12  2  0  0  0  0
  6  1  1  0  0  0  0
 11 13  1  0  0  0  0
M END
```

the first three numbers are the *x*, *y* and *z* coordinates of the atom

the first bond is between atoms 1 and 2 and has order 2



```
  ACD/Labs08061122363D

 21 21  0  0  0  0  0  0  0  0  1 V2000
   18.7769  -15.2504   -0.1032 C   0  0  0  0  0  0  0  0  0  0  0  0
   18.7571  -16.6359   -0.1252 C   0  0  0  0  0  0  0  0  0  0  0  0
   17.5868  -14.5409   -0.1114 C   0  0  0  0  0  0  0  0  0  0  0  0
   17.5465  -17.3106   -0.1545 C   0  0  0  0  0  0  0  0  0  0  0  0
   16.3767  -15.2158   -0.1421 C   0  0  0  0  0  0  0  0  0  0  0  0
   16.3559  -16.6013   -0.1633 C   0  0  0  0  0  0  0  0  0  0  0  0
   17.6081  -13.0313   -0.0880 C   0  0  0  0  0  0  0  0  0  0  0  0
   20.0592  -14.5322   -0.0715 N   0  0  0  0  0  0  0  0  0  0  0  0
   17.5247  -18.7799   -0.1764 N   0  0  0  0  0  0  0  0  0  0  0  0
   15.1150  -14.4620   -0.1527 N   0  0  0  0  0  0  0  0  0  0  0  0
   20.0742  -13.3140   -0.0089 O   0  0  0  0  0  0  0  0  0  0  0  0
   21.1073  -15.1564   -0.0523 O   0  0  0  0  0  0  0  0  0  0  0  0
   16.4750  -19.3759    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   18.5697  -19.4030   -0.2650 O   0  0  0  0  0  0  0  0  0  0  0  0
   14.0496  -15.0560   -0.1515 O   0  0  0  0  0  0  0  0  0  0  0  0
   15.1330  -13.2425   -0.1568 O   0  0  0  0  0  0  0  0  0  0  0  0
   19.7111  -17.2054   -0.1194 H   0  0  0  0  0  0  0  0  0  0  0  0
   15.3860  -17.1427   -0.1873 H   0  0  0  0  0  0  0  0  0  0  0  0
   17.6136  -12.6451   -1.1298 H   0  0  0  0  0  0  0  0  0  0  0  0
   16.7057  -12.6567    0.4410 H   0  0  0  0  0  0  0  0  0  0  0  0
   18.5209  -12.6823    0.4410 H   0  0  0  0  0  0  0  0  0  0  0  0
  2  1  1  0  0  0  0
  3  1  2  0  0  0  0
  4  2  2  0  0  0  0
  5  3  1  0  0  0  0
  6  4  1  0  0  0  0
  6  5  2  0  0  0  0
  3  7  1  0  0  0  0
  1  8  1  0  0  0  0
  4  9  1  0  0  0  0
  5 10  1  0  0  0  0
  8 11  2  0  0  0  0
  8 12  2  0  0  0  0
  9 13  2  0  0  0  0
  9 14  2  0  0  0  0
 10 15  2  0  0  0  0
 10 16  2  0  0  0  0
 17  2  1  0  0  0  0
 18  6  1  0  0  0  0
 19  7  1  0  0  0  0
 20  7  1  0  0  0  0
 21  7  1  0  0  0  0
M  END
```

Challenge:

Draw this molecule. What is the name of the molecule?

Databases of small organic molecules

> 1 M structures of small molecules

- Small molecule: < 100 atoms
- Small molecules = “drug-like” molecules
- Experimental structures
- Predicted (computed) structures

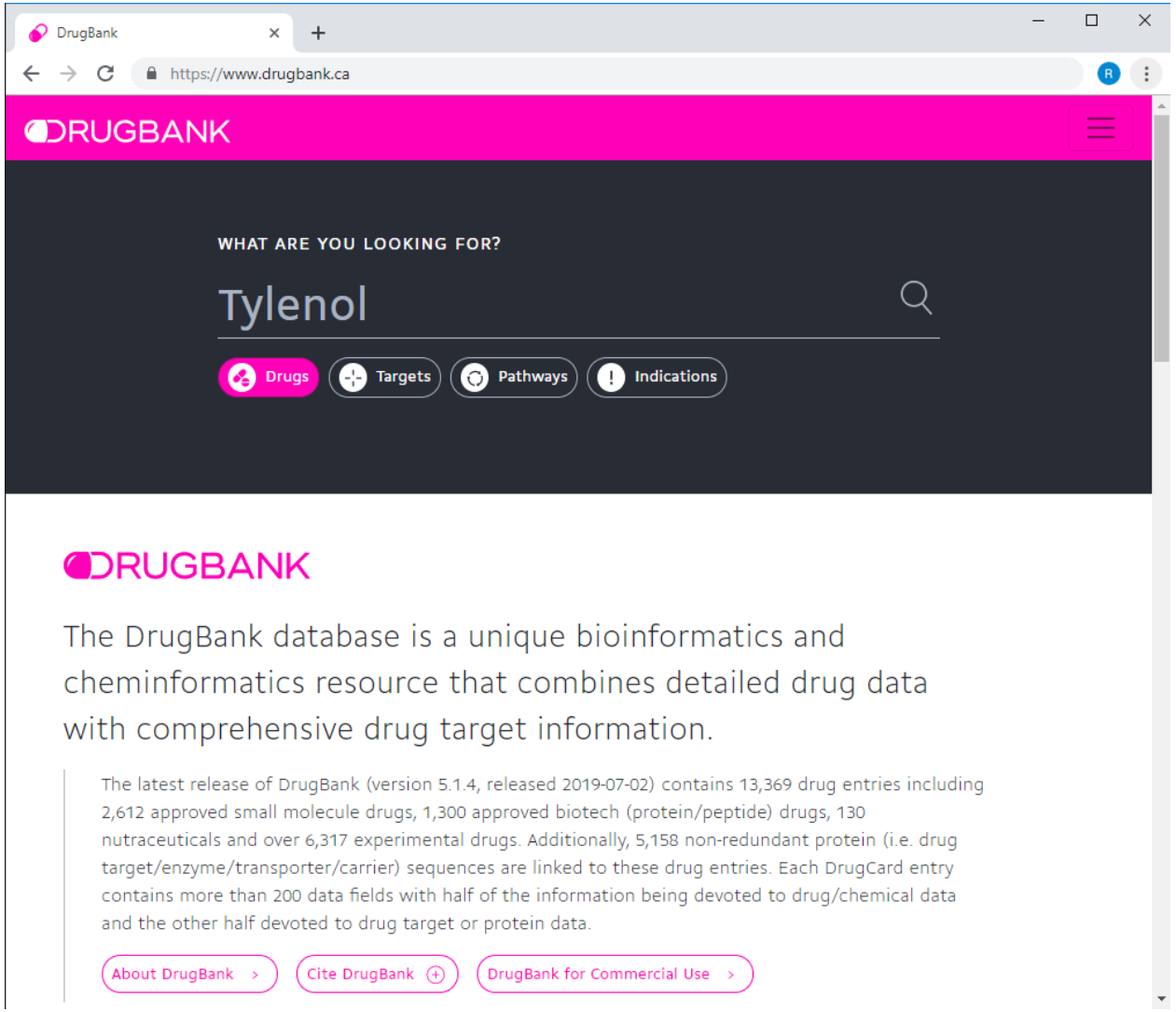
PubChem

ZINC

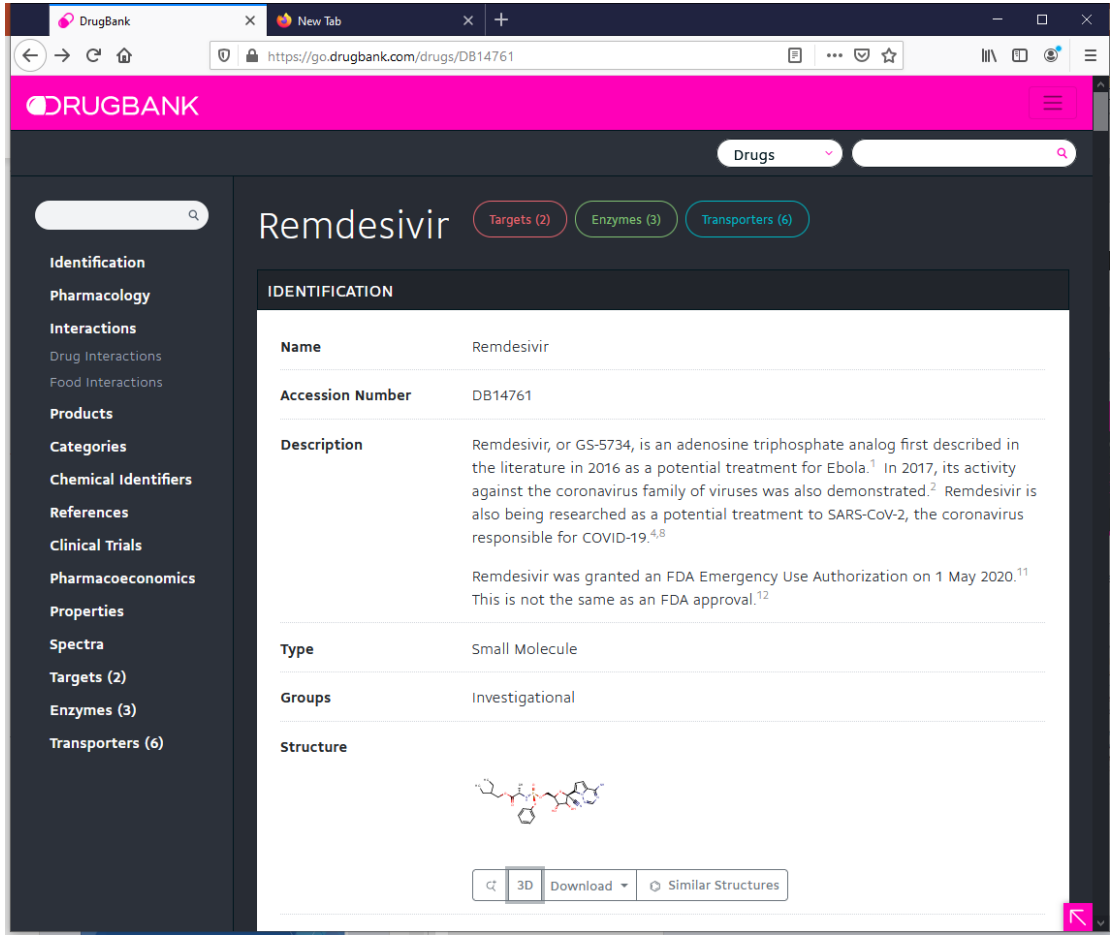
ChEMBL

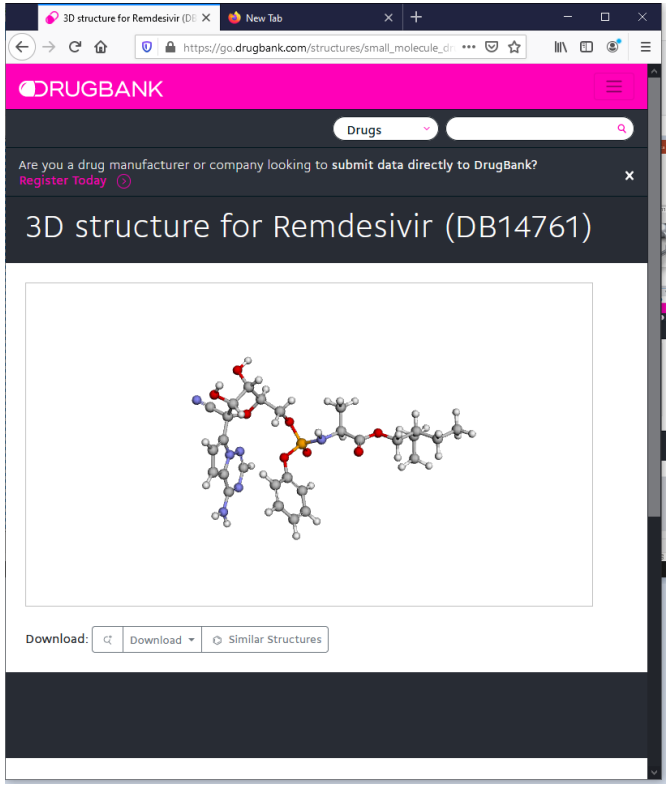
DRUGBANK

DrugBank – database of drugs

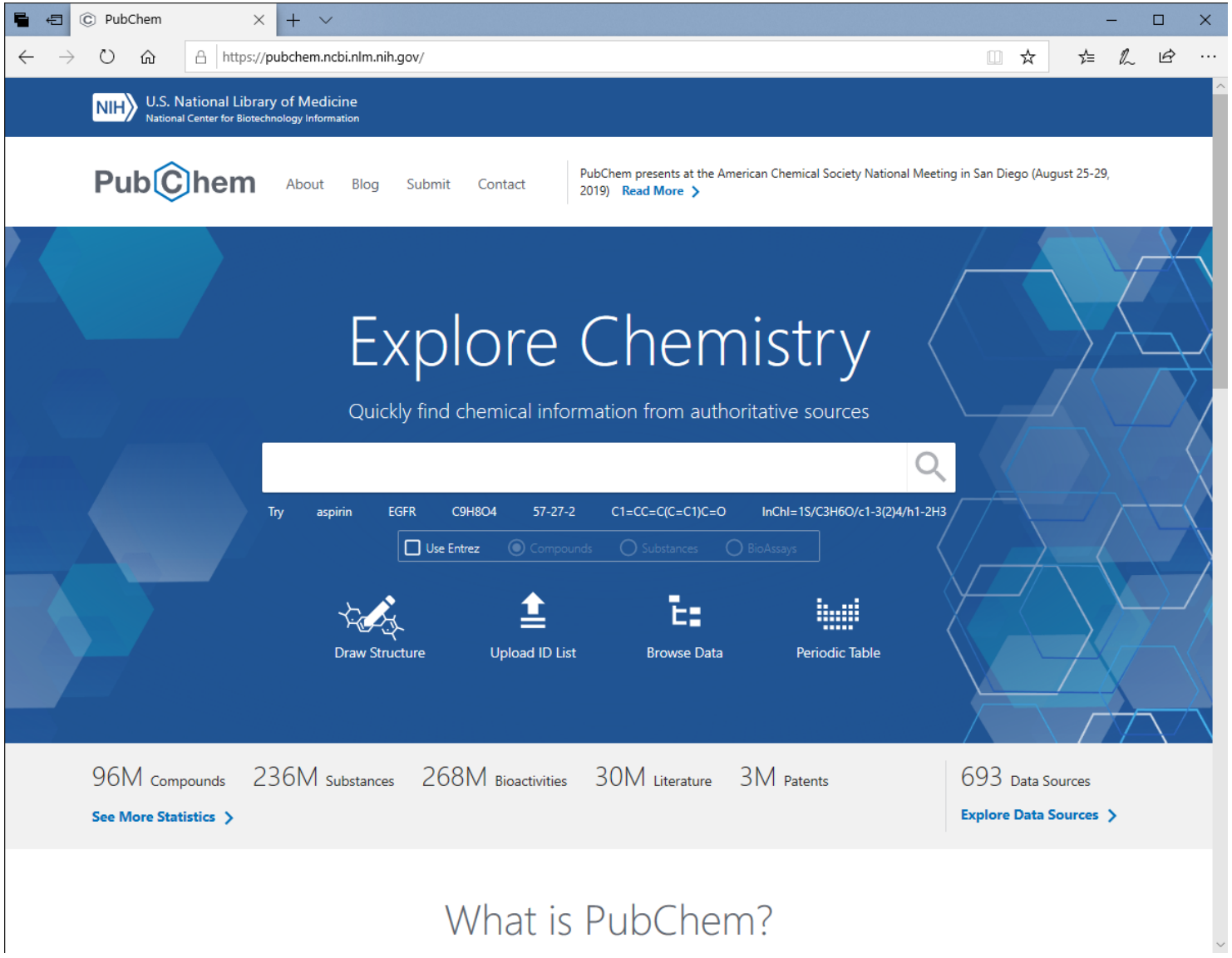


DrugBank – database of drugs

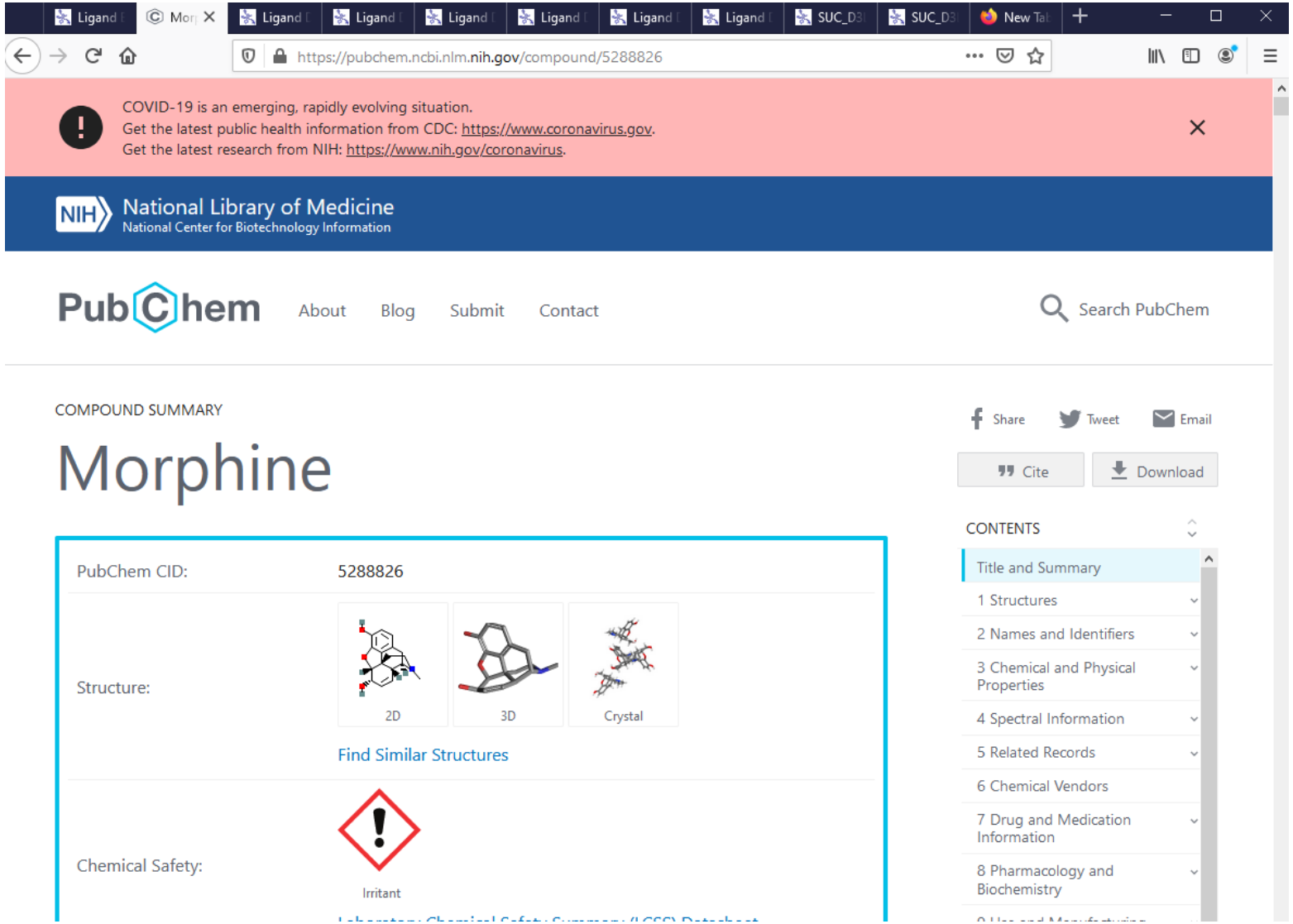




PubChem – database of organic molecules

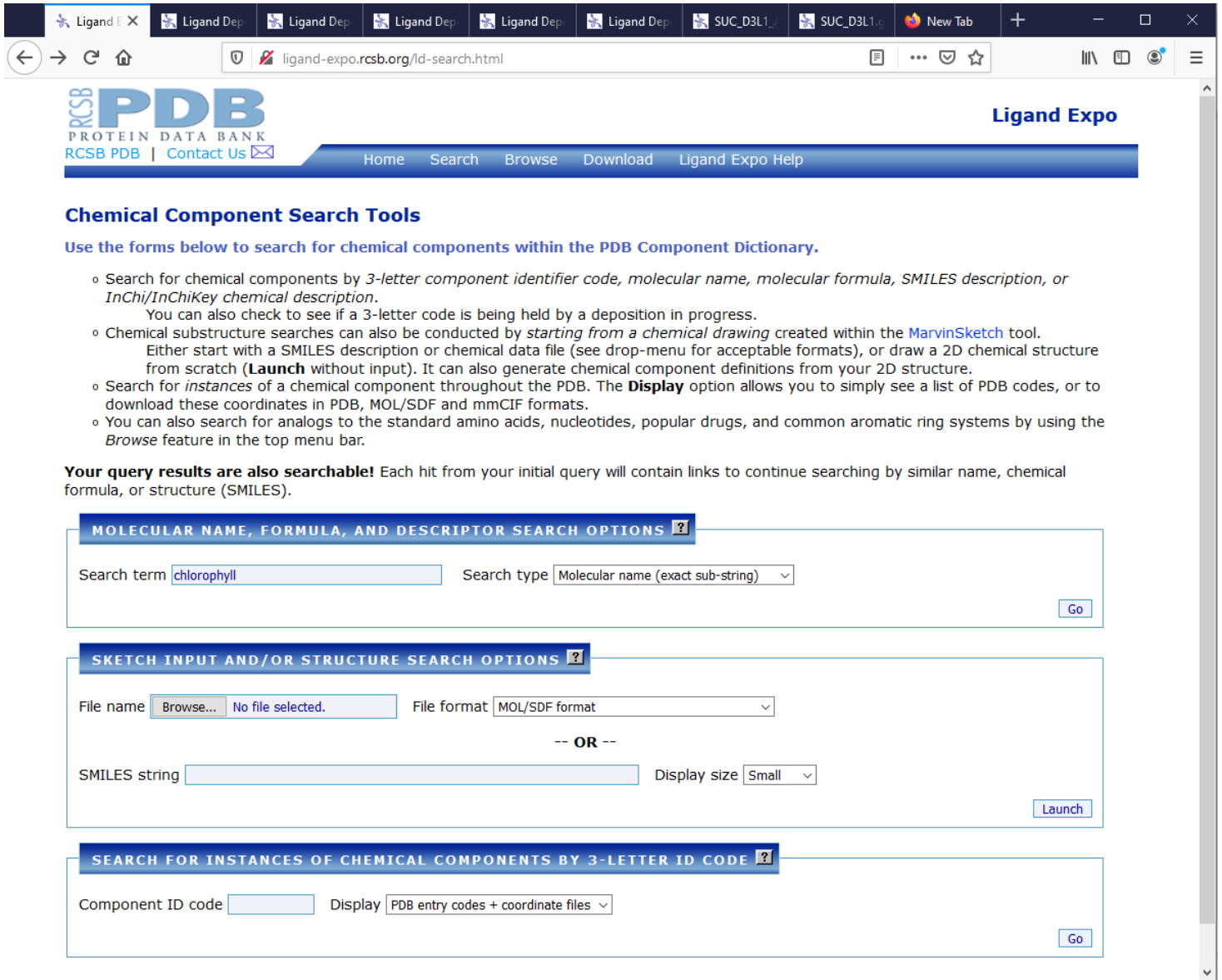


PubChem – database of organic molecules

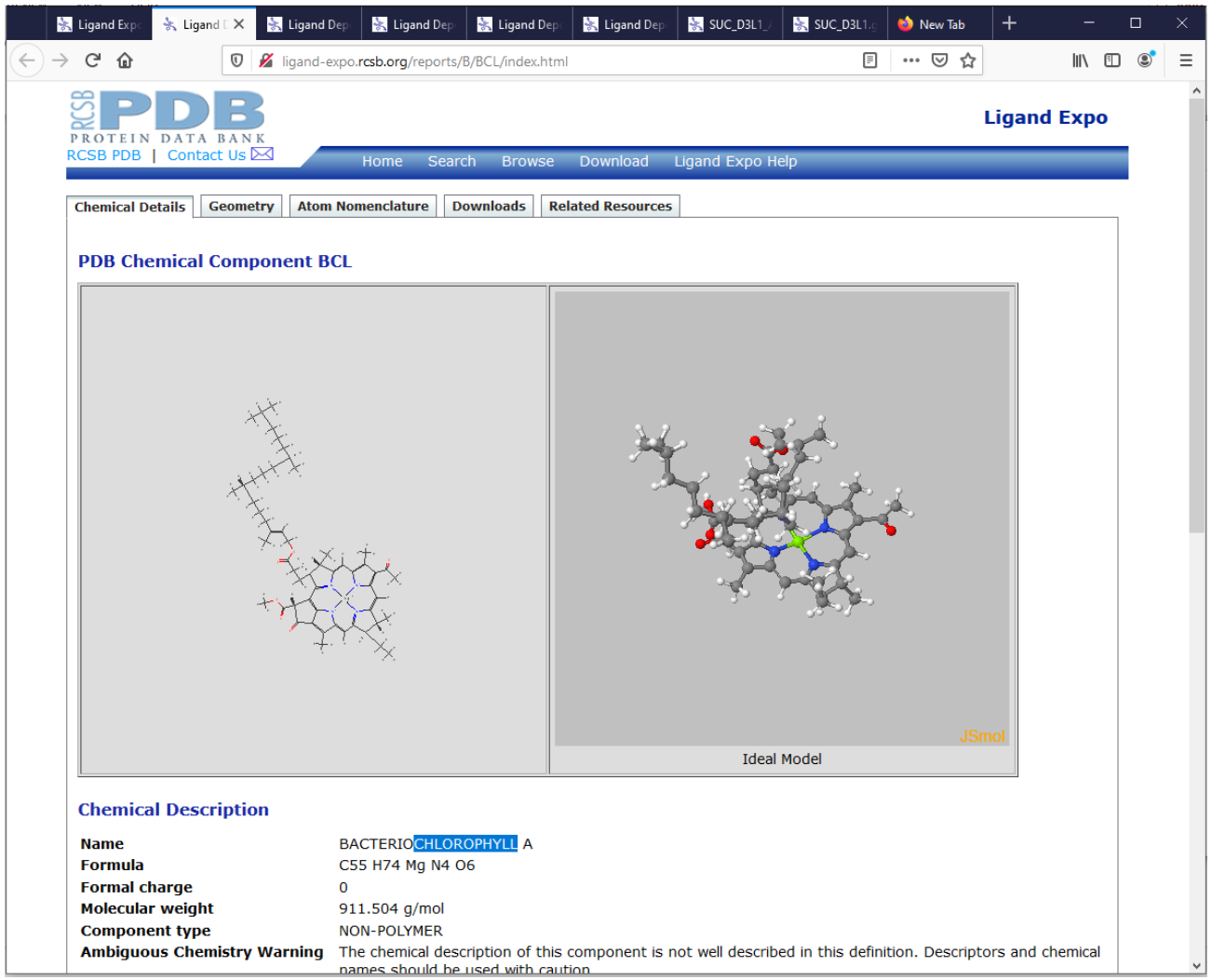


23

Ligand Expo – database of ligands
Ligand = molecule bound in a protein



Ligand Expo – database of ligands
Ligand = molecule bound in a protein



Databases of biomacromolecules

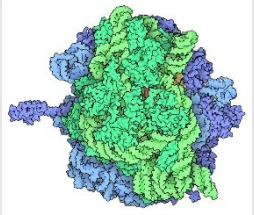
Mainly proteins
> 200 k experimental structures
> 200 M computed structures

Protein Data Bank – sources of data

89% X-ray crystallography

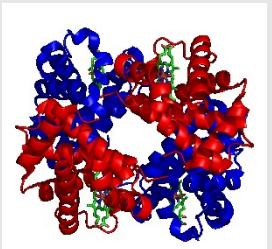
10% NMR spectroscopy

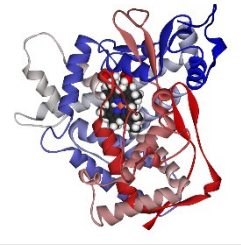
1% cryoelectron microscopy



3D struktura

```
...
ATOM    46  C   GLY A  70      51.536  23.360  40.507
ATOM    47  O   GLY A  70      50.947  22.279  40.325
ATOM    48  N   ILE A  71      50.965  24.532  40.270
ATOM    49  CA  ILE A  71      49.595  24.644  39.786
...
```

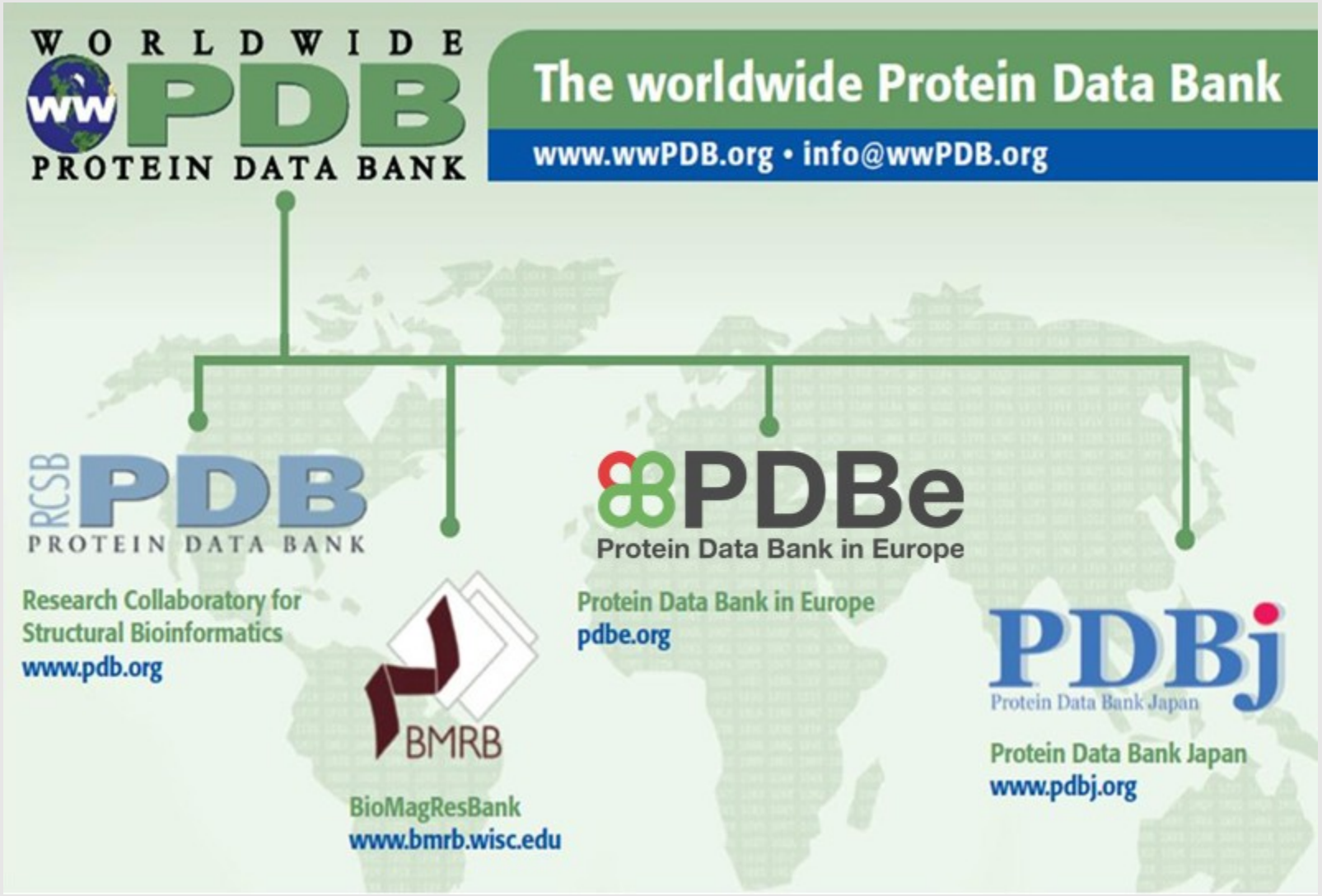




WORLDWIDE PDB PROTEIN DATA BANK

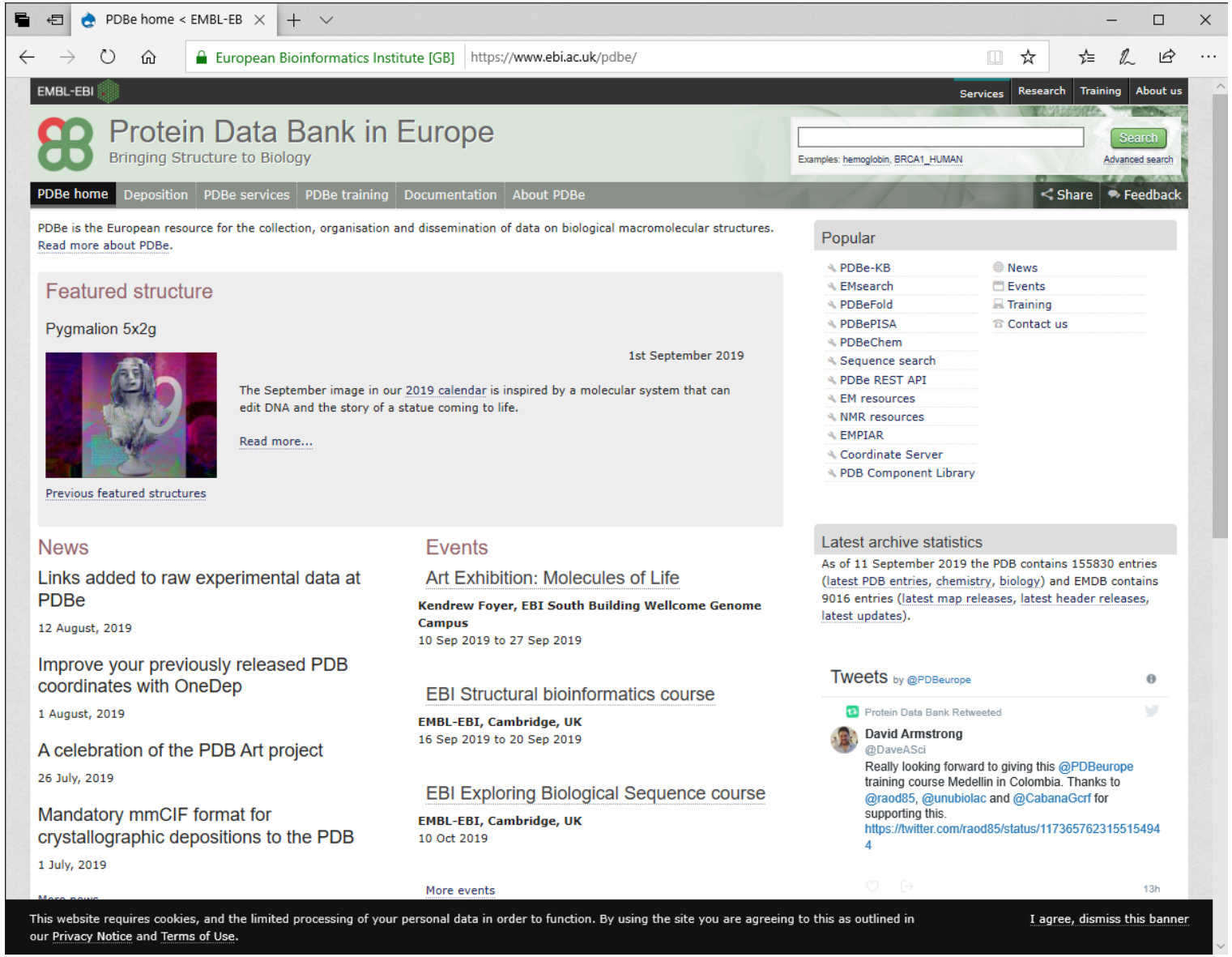
27

Protein Data Bank

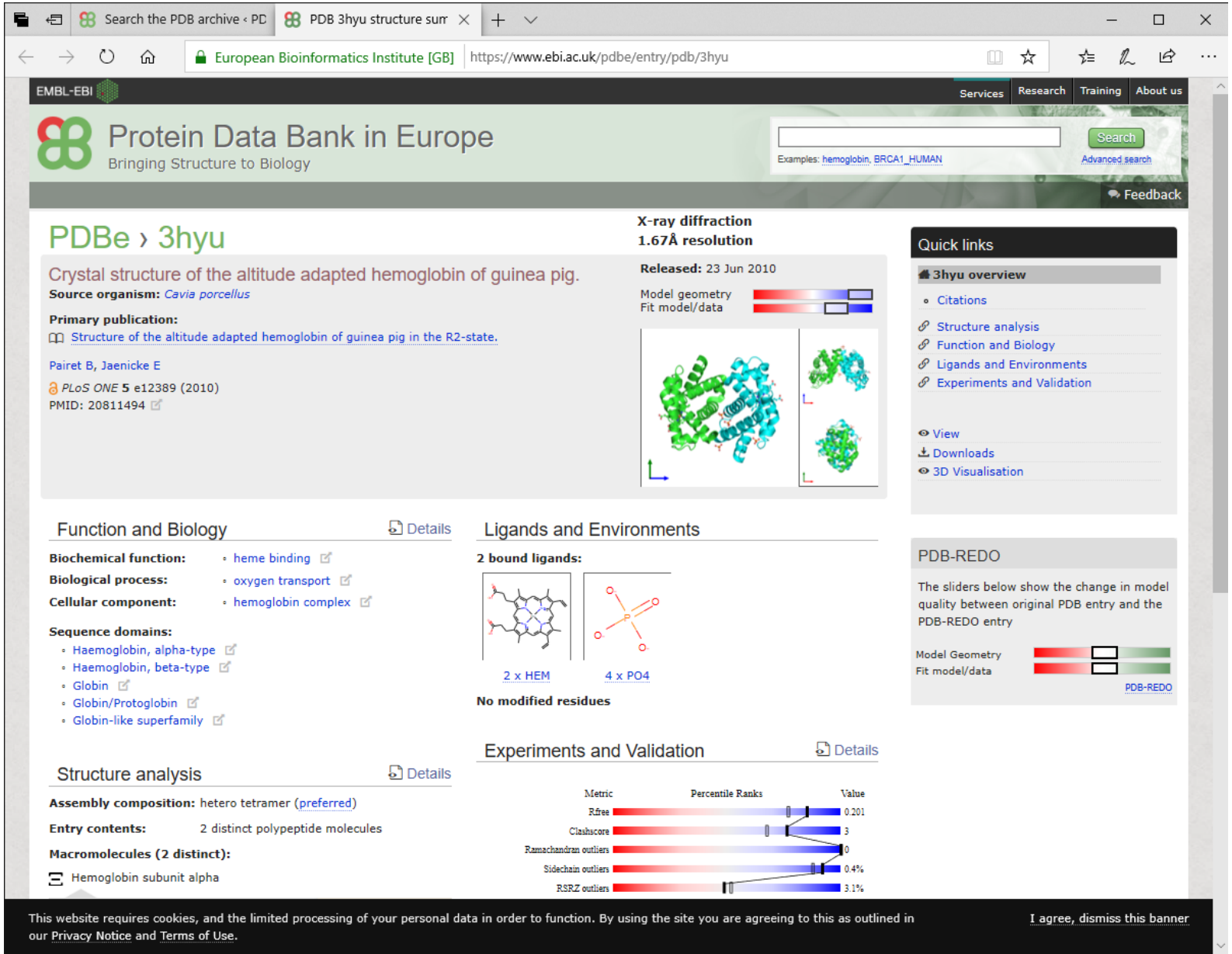


> 225 000 biomacromolecular structures

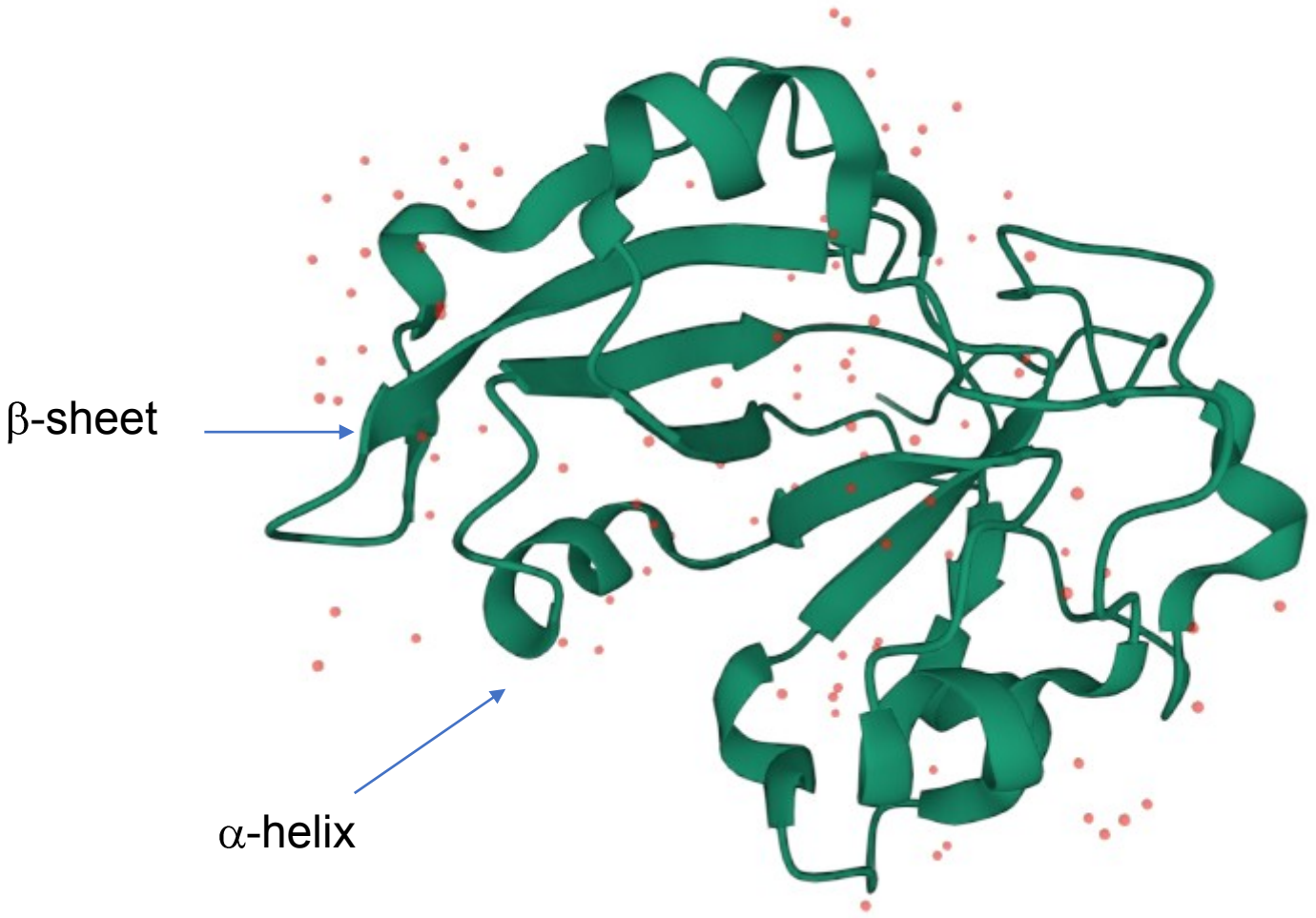
Protein Data Bank



Protein Data Bank



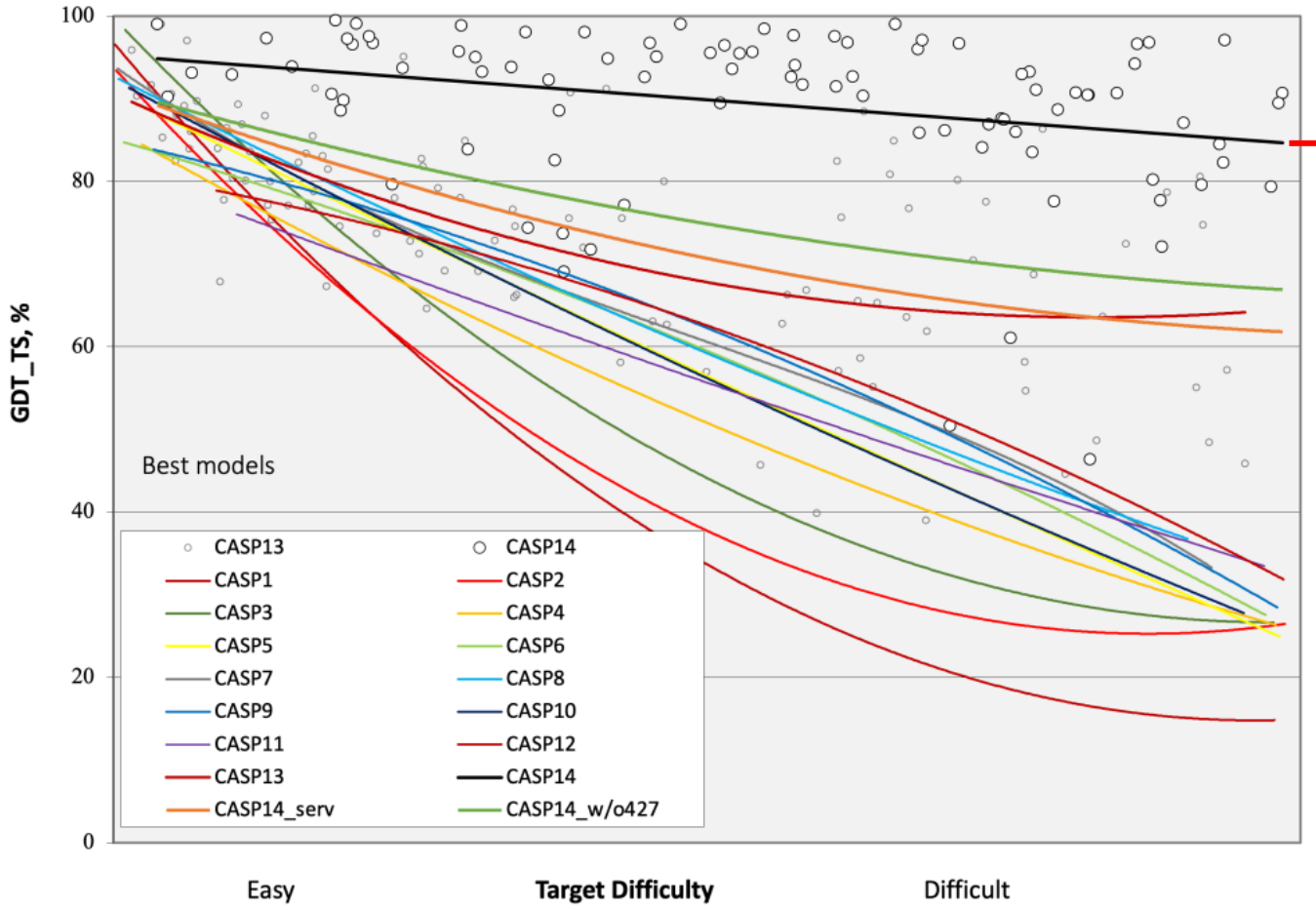
Protein Data Bank



Scabin
6vv4

Prediction of protein structures by AlphaFold

Structures generated by artificial intelligence

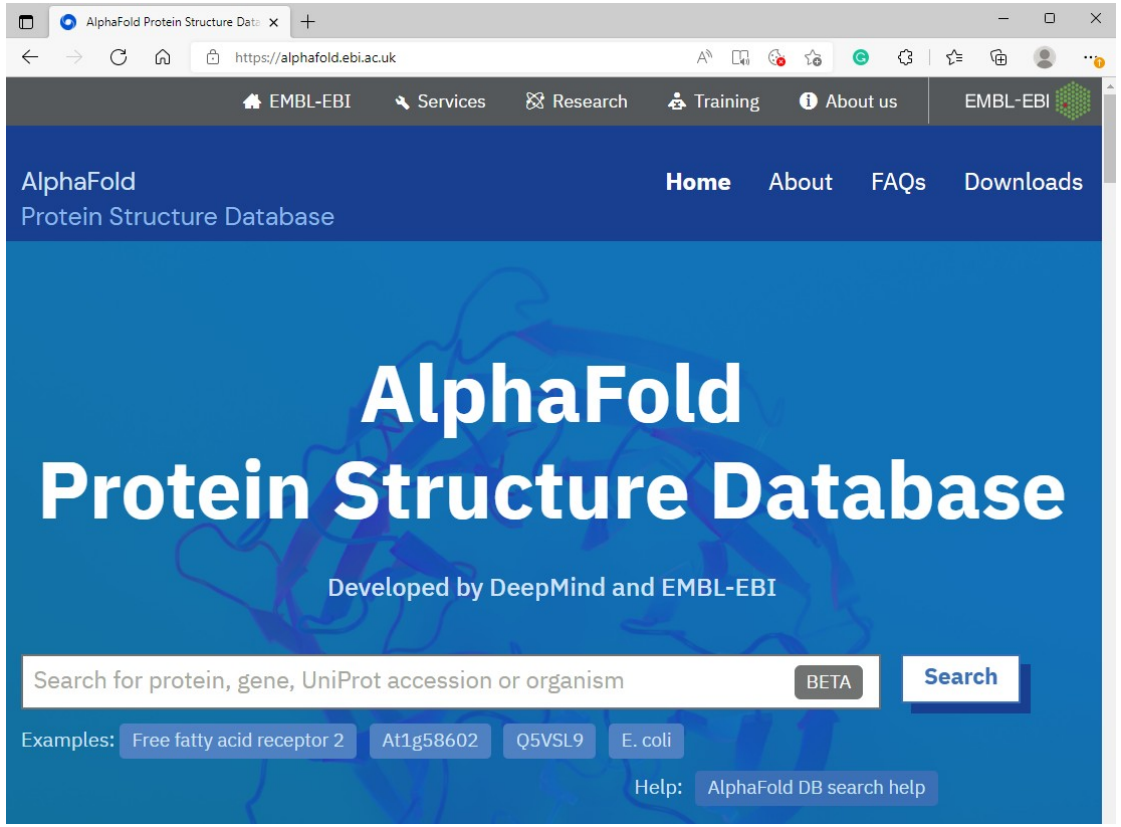


Structure prediction challenge 2020: AlphaFold2 wins

https://predictioncenter.org/casp14/doc/presentations/2020_11_30_CASP14_Introduction_Moult.pdf

Prediction of protein structures by AlphaFold

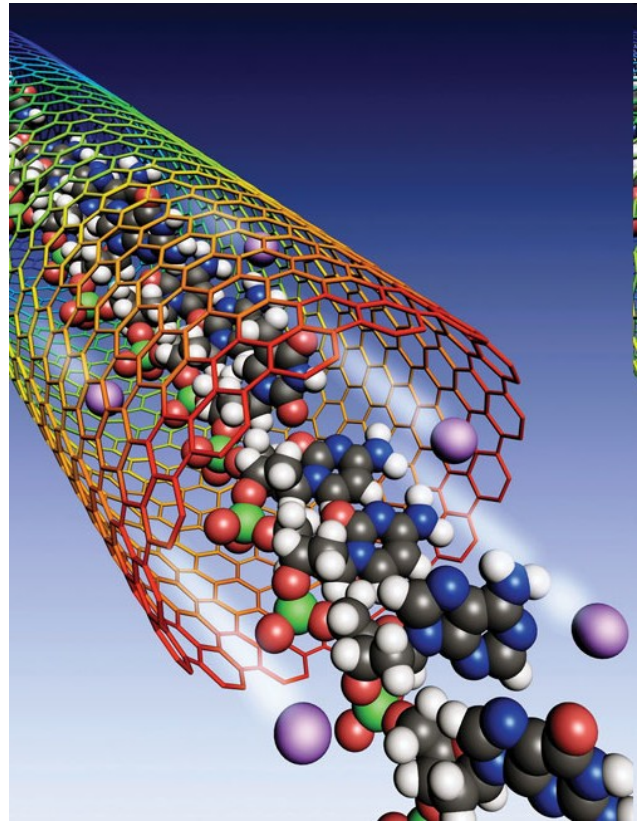
Structures generated by artificial intelligence



> 200M protein structures

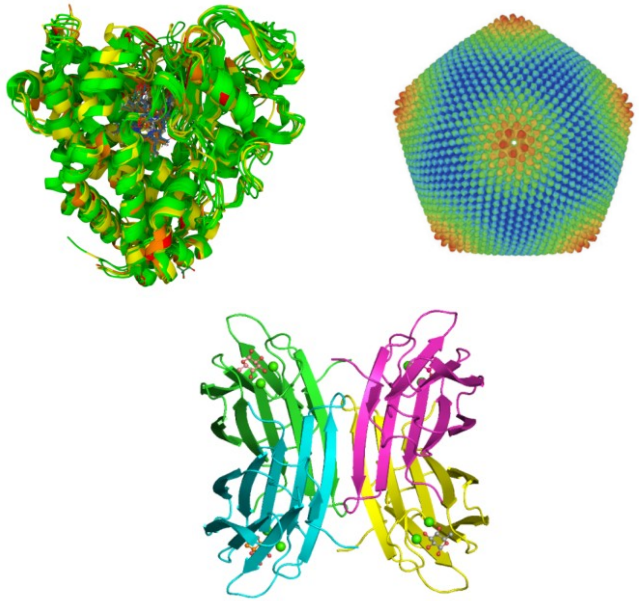
Exercises

- Search the PubChem database for a testosterone molecule:
  - See its 2D structure. How many O's does it have?
  - Look at its 3D structure. Is any of its cycles planar?
  - Look at its SDF file. What are the x, y, and z coordinates of the first atom?
- Search the DrugBank database for a penicillin molecule:
  - How many S atoms does it have?
  - Are any of its cycles planar?
  - Look at its SDF file. What 2 atoms form the first bond?
- Search the LigandExpo database for a fructose molecule:
  - How many double bonds does it has?
  - How many C's are off cycle?
  - Look at its SDF file. What are the coordinates of the first hydrogen?
- Look up the green mamba venom molecule in the Protein Data Bank:
  - How many beta-sheets does it have?
  - Look at the PDB file. Which amino acid is the first?



Thank you for your attention

Bioinformatics



Tools
Databases
AI

Chemoinformatics