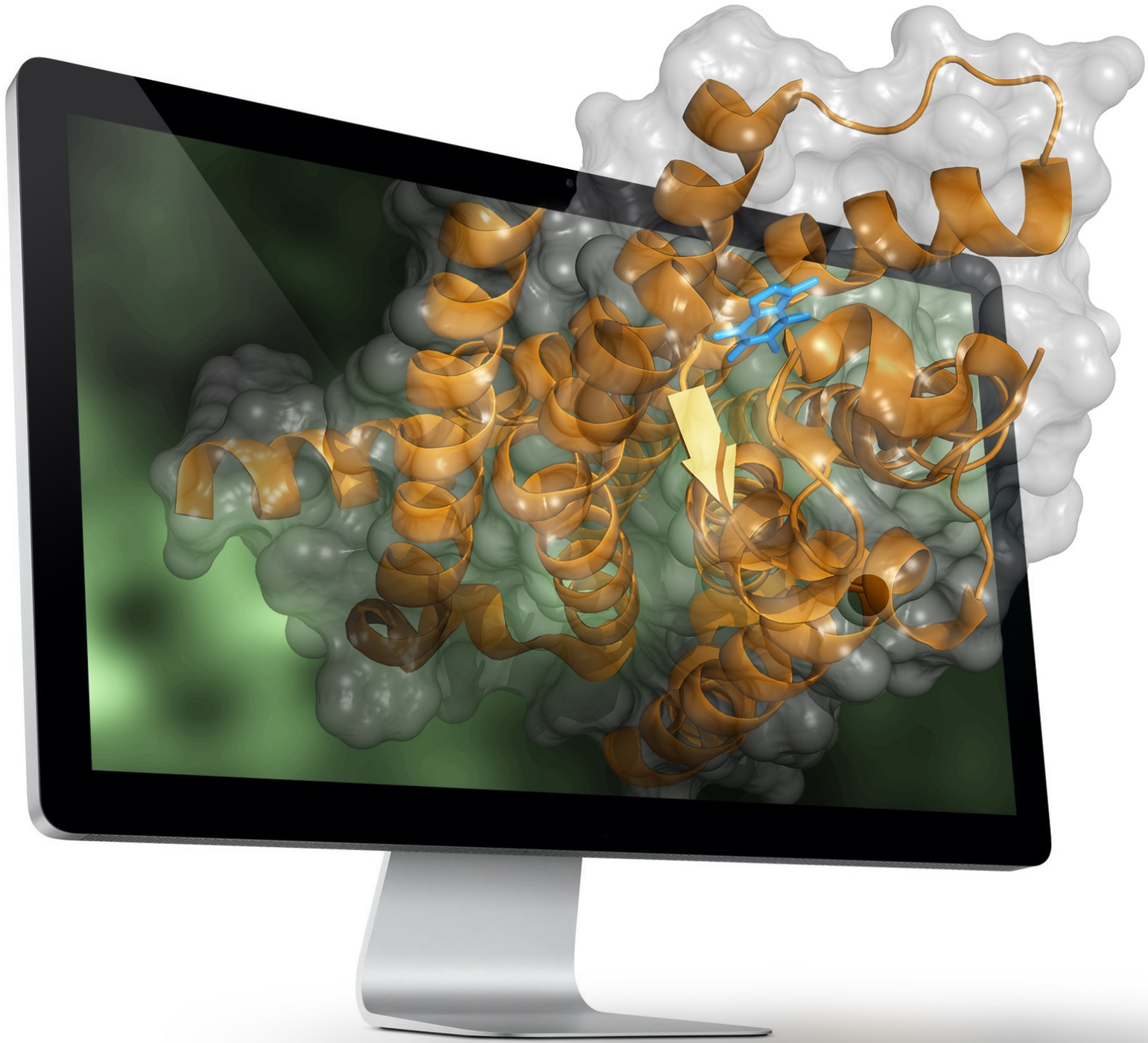


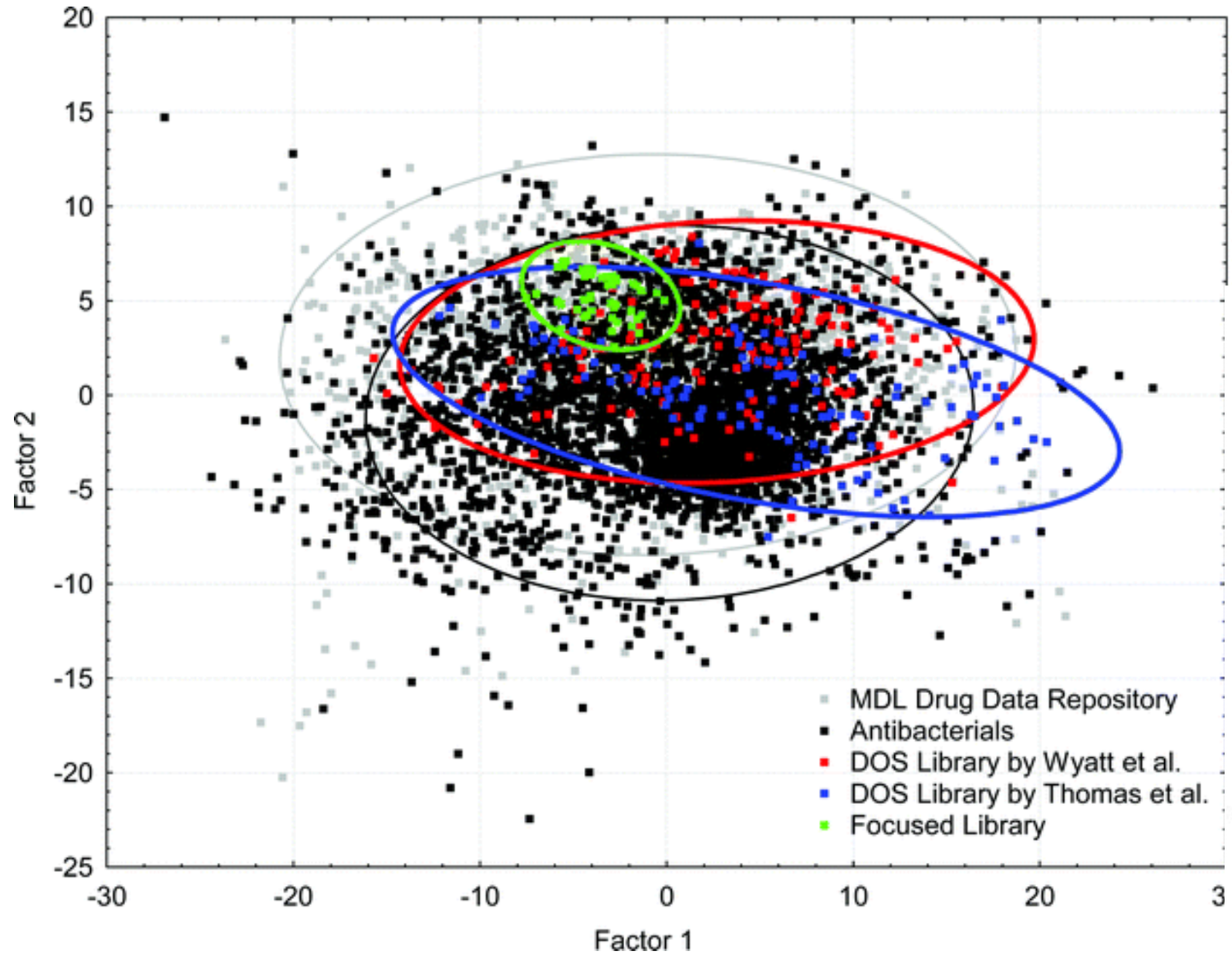
Pokročilá chemoinformatika

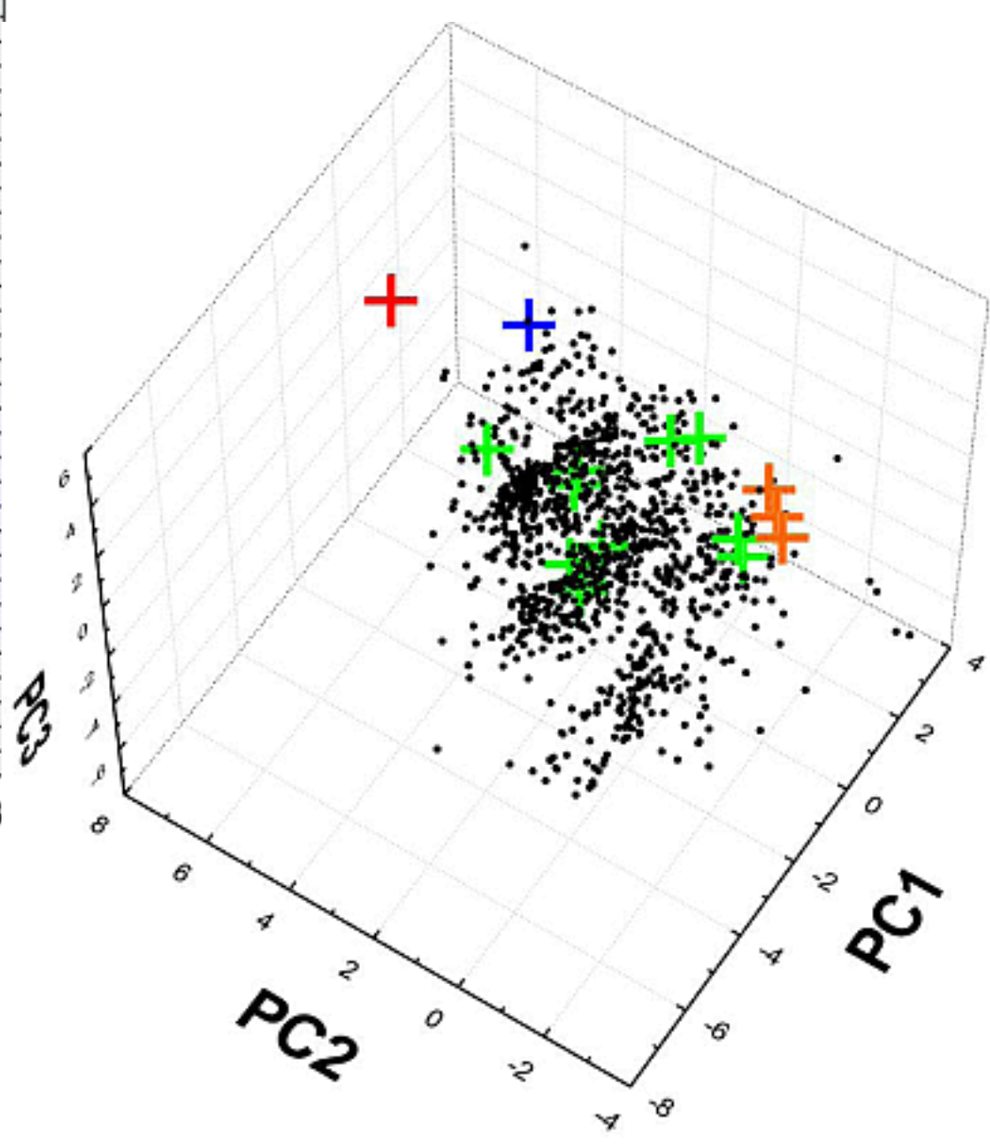
Databáze, chemický prostor
únor 2015

Chemické databáze

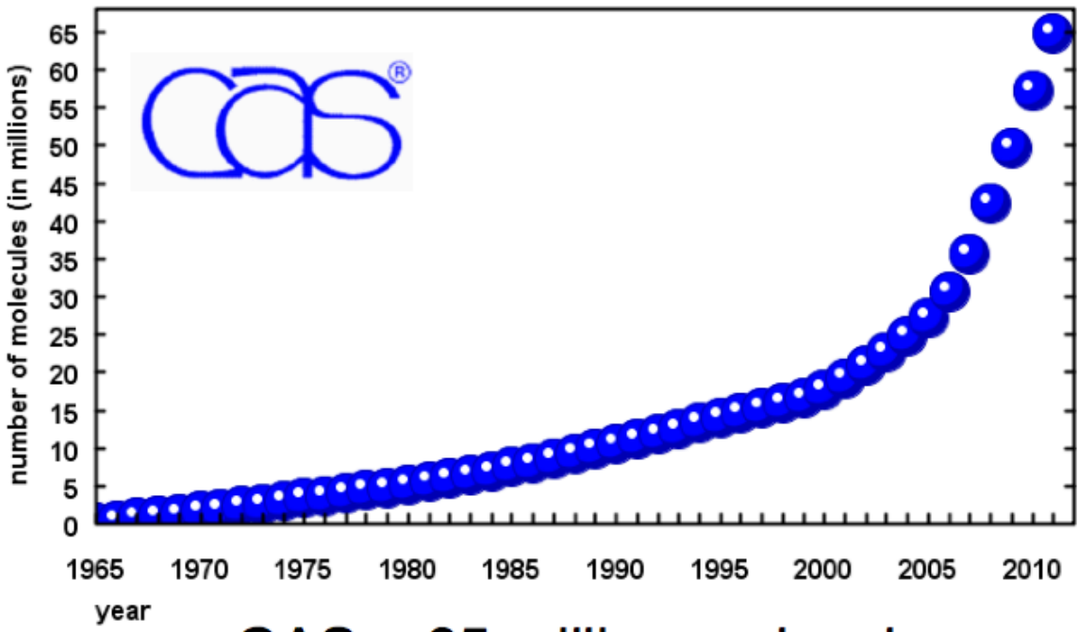
- Informace o molekulách, struktury molekul, vlastnosti, aktivity, …
- PUBCHEM
- DRUGBANK
- ZINC
- CHEMBL
- PHYSPROP
  http://esc.syrres.com/fatepointer/search.asp

Chemický prostor (chemical space)

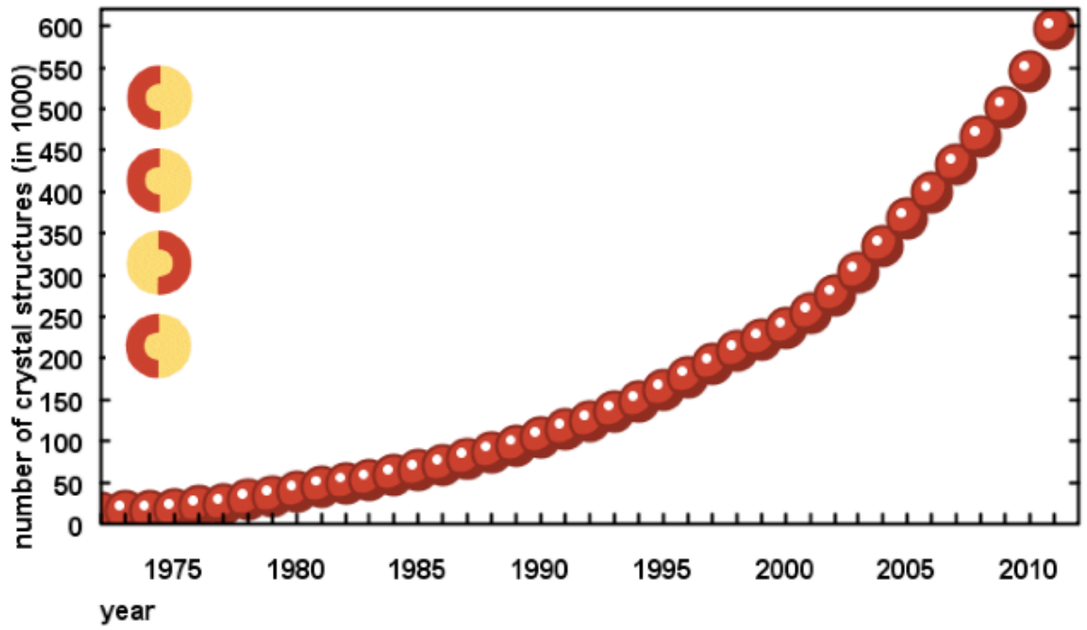




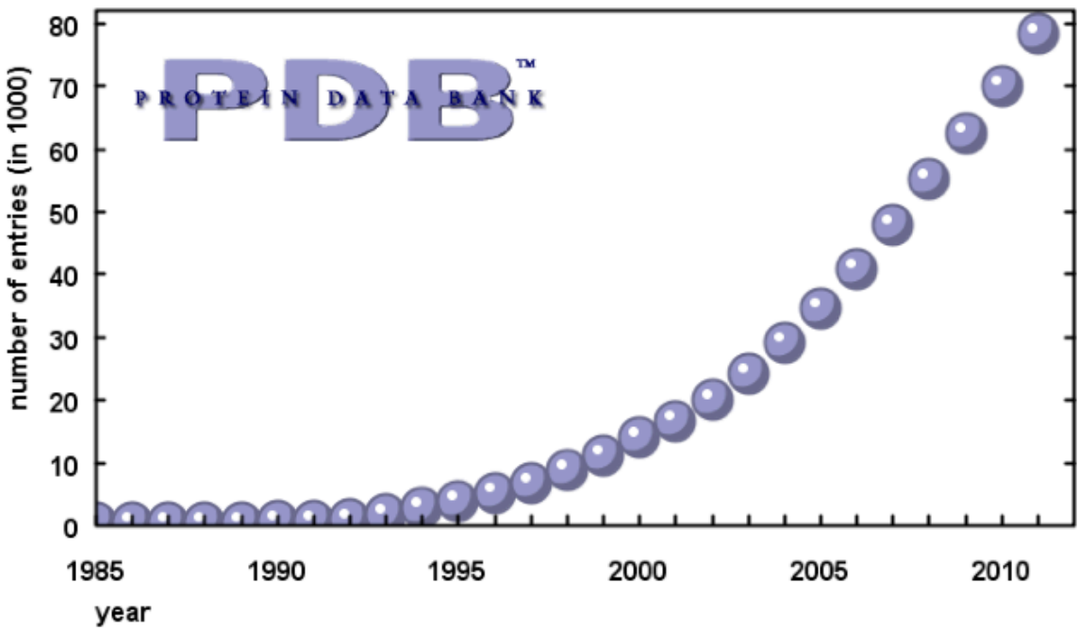
Velikost základních chemických databází



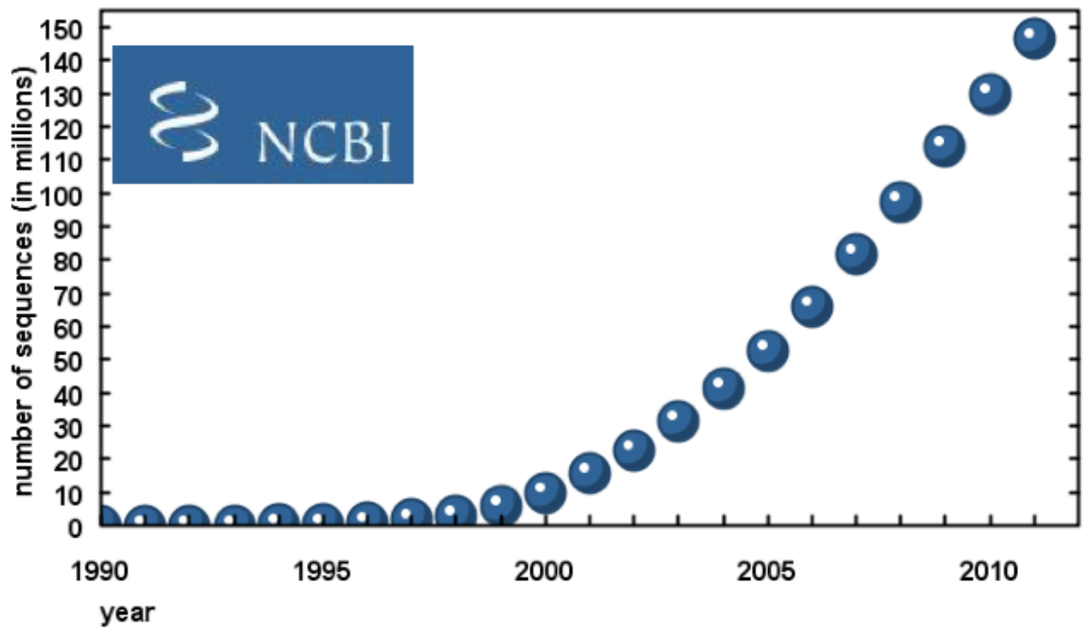
CAS – 65 million molecules



CCDC – 600’000 structures



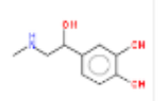
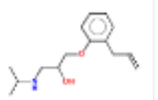
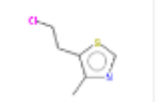
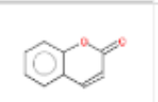
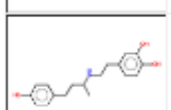
PDB – 78’000 proteins



GenBank – 145 million sequences

NOVARTIS

Práce s chemickým prostorem

Molecule	logP	PSA	natoms	MW	…
	-0.06	72.7	13	183.2	…
	2.58	41.5	18	249.3	…
	2.11	12.9	9	161.6	…
	2.01	20.2	11	146.1	…
	3.31	78.8	30	425.9	…
..					

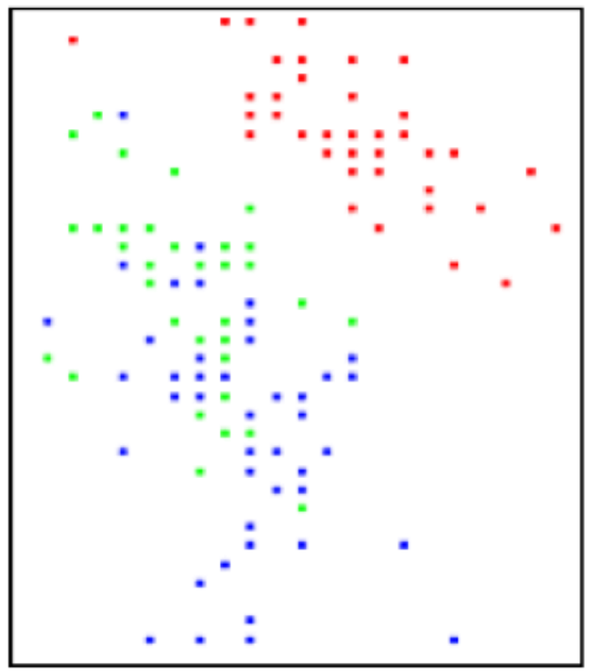


table with properties or fragments

dimensionality reduction

visualization

Fingerprinty a podobnost

Podobnostní hledání

Fingerprint

- 10010001010011110101001010001 …

Podobnost/vzdálenost

		molekula B		
		0	1	celkem
molekula A	0	*d*	*b*	*b + d*
	1	*a*	*c*	*a + c = A*
	celkem	*a + d*	*c + b = B*	*n*

a je počet „1“, které má molekula A, ale které zároveň nemá molekula B
b je počet „1“, které má molekula B, ale které zároveň nemá molekula A
c je počet „1“, které má molekula A a které má zároveň i molekula B
d je počet „0“, které má molekula A a které má zároveň i molekula B

n je počet dvojnásobný počet fragmentů, platí n = a + b + c + d
A je celkový počet „1“ v molekule A
B je celkový počet „1“ v molekule B

Chemické formáty

- MOL (V2000, V3000), SDF
  http://c4.cabrillo.edu/404/ctfile.pdf

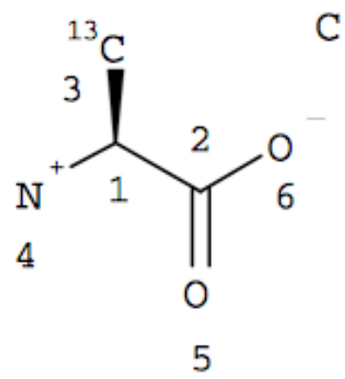
- MOL2

- PDB, mmCIF

- XYZ

MOL (V2000)

L-Alanine



Chiral

```
  6  5  0  0  1  0              3 V2000
   -0.6622    0.5342    0.0000 C   0  0  2  0  0  0
    0.6622   -0.3000    0.0000 C   0  0  0  0  0  0
   -0.7207    2.0817    0.0000 C   1  0  0  0  0  0
   -1.8622   -0.3695    0.0000 N   0  3  0  0  0  0
    0.6220   -1.8037    0.0000 O   0  0  0  0  0  0
    1.9464    0.4244    0.0000 O   0  5  0  0  0  0
  1  2  1  0  0  0
  1  3  1  1  0  0
  1  4  1  0  0  0
  2  5  2  0  0  0
  2  6  1  0  0  0
M  CHG  2   4   1   6  -1
M  ISO  1   3  13
M  END
```

Blocks not used in this Ctab: Atom List, Stext

Counts Line

Atom Block

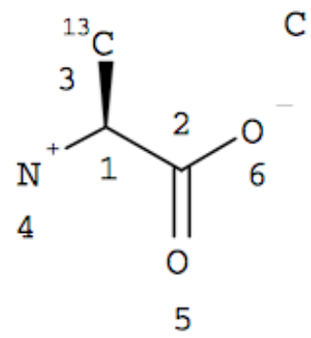
Bond Block

Properties Block

Connection Table (Ctab)

MOL (V3000)

L-Alanine



Chiral

```
L-Alanine
GSMACCS-II07189510252D 1   0.00366     0.00000     0
Figure 1, J. Chem. Inf. Comput. Sci., Vol 32, No. 3., 1992
  0  0  0     0  0            999 V3000
M  V30 BEGIN CTAB
M  V30 COUNTS 6 5 0 0 1
M  V30 BEGIN ATOM
M  V30 1 C -0.6622 0.5342 0 0 CFG=2
M  V30 2 C 0.6622 -0.3 0 0
M  V30 3 C -0.7207 2.0817 0 0 MASS=13
M  V30 4 N -1.8622 -0.3695 0 0 CHG=1
M  V30 5 O 0.622 -1.8037 0 0
M  V30 6 O 1.9464 0.4244 0 0 CHG=-1
M  V30 END ATOM
M  V30 BEGIN BOND
M  V30 1 1 1 2
M  V30 2 1 1 3 CFG=1
M  V30 3 1 1 4
M  V30 4 2 2 5
M  V30 5 1 2 6
M  V30 END BOND
M  V30 END CTAB
M  END
```

Header Block

Comment Line

Counts Line

Atom Block

Bond Block

Connection Table (Ctab)

Blocks not used in this Ctab:
Sgroup block, Rgroup block, 3D block

OpenBabel

- module add openbabel
- obabel –ixxx molecule.xxx –oyyy (-O) molecule.yyy

Úkol

1. Vyhledejte v databázi tyto látky a k něm tyto informace: MW, logP, pKa, teplotu tání, počet akceptorů, donorů, drug-like, 2D nebo 3D strukturu, smiles, obchodní název, cílový protein, podobné látky strukturně a funkčně.

a) ibuprofen

b) kofein

c) 4-Hydroxy-3-(3-oxo-1-fenylbutyl)kumarin

d) vicodin