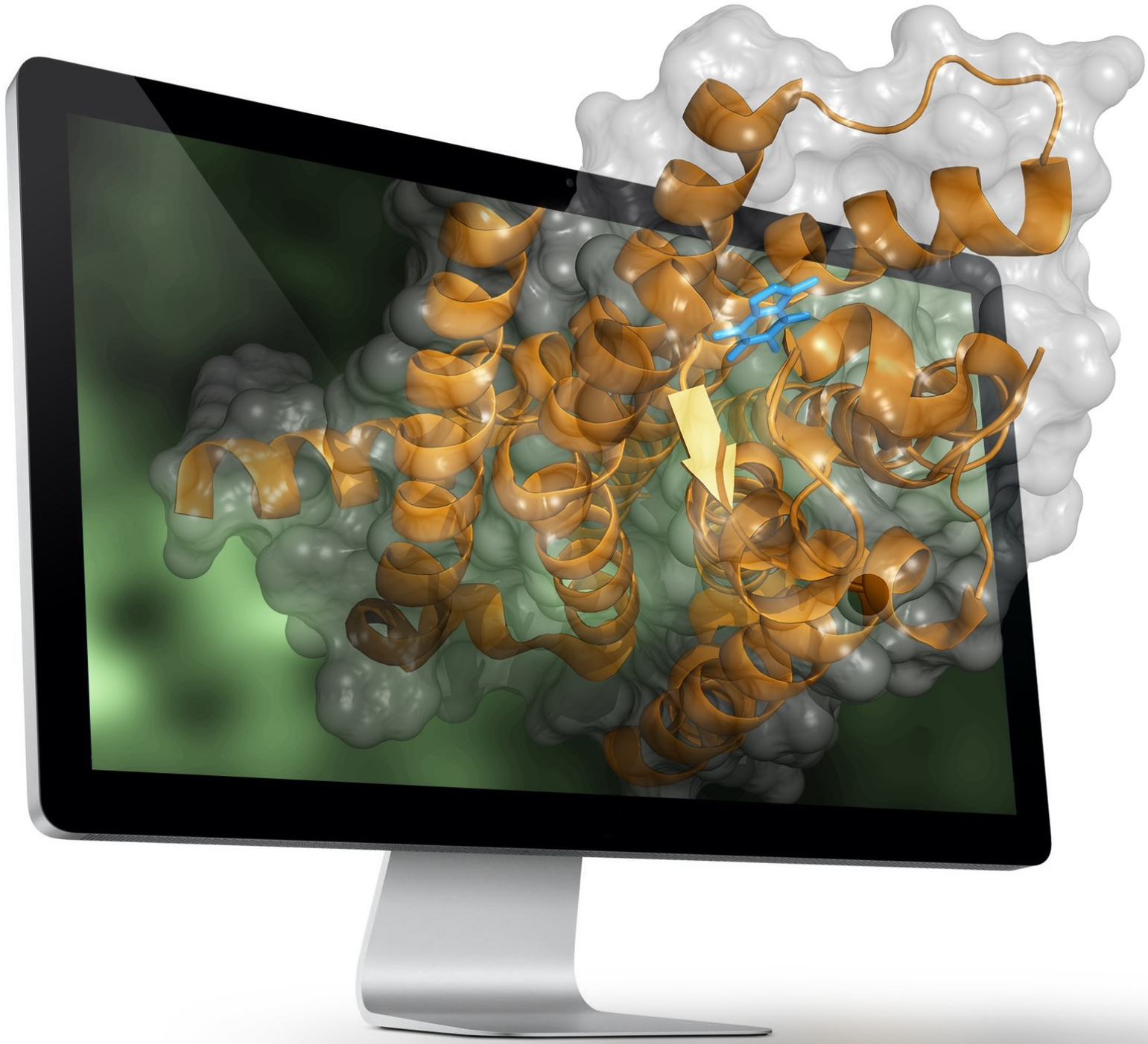


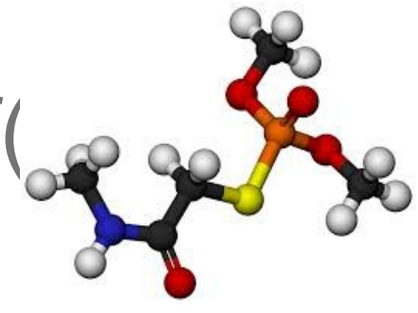
Pokročilá chemoinformatika

Úvod do QSAR/QSPR modelovaní
únor 2017

Deskriptor

- Jedná se se o “užitečné” **číslo**, které popisuje určitý aspekt struktury molekuly, toto číslo je produktem matematického, logického nebo statitického výpočtu

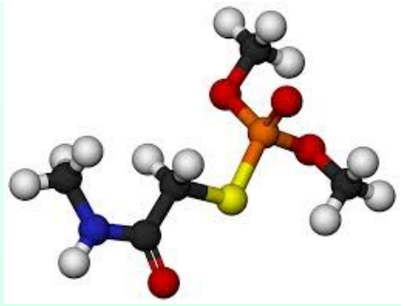
- Z pohledu statistiky se jedná o nezávislou proměnnou nebo také regresorem

- Deskriptor = f()

- V současnosti známe přes 3300 deskriptorů

Deskriptor

a real object



molecule

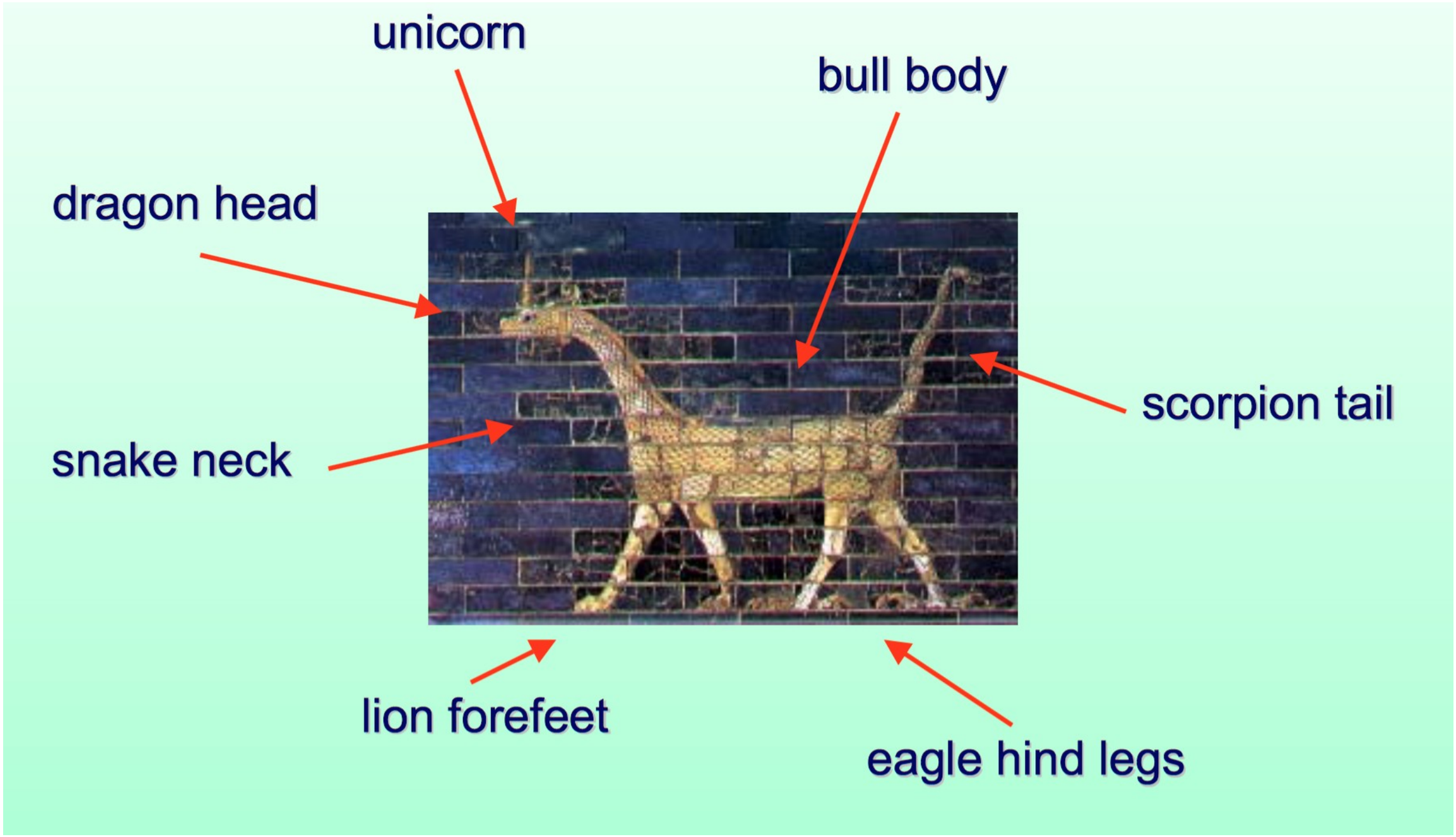
δ

molecular structure representation

molecular descriptors

numbers

Deskriptory – různý pohled na stejnou věc



Deskriptory – různý pohled na stejnou věc

symmetry

electronic aspects

branching

H - bonding

steric

hydrophobicity

size

shape

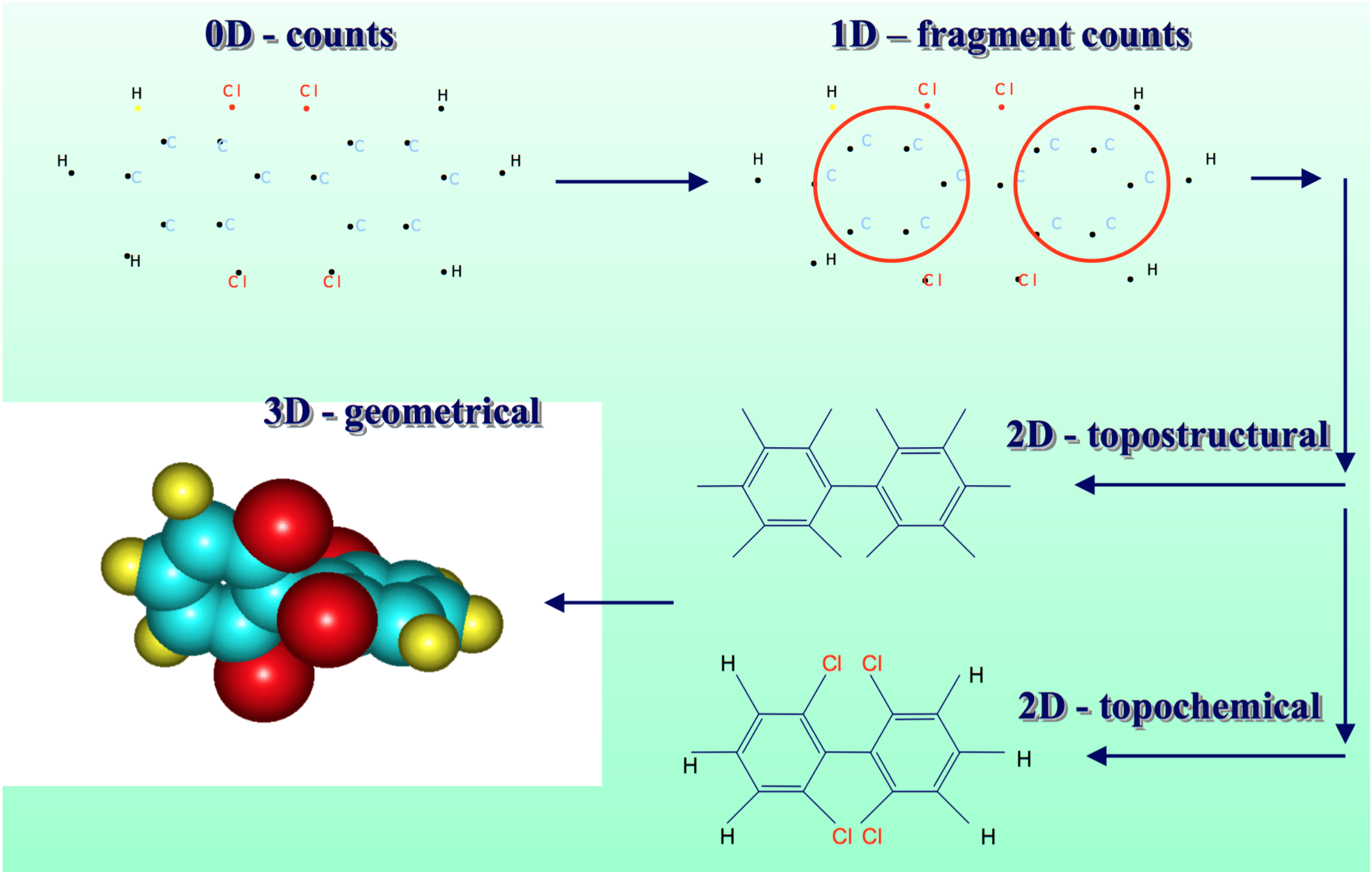
reactivity

cyclicity

Rozdělení deskriptorů

- Dle použité struktury
  - 0D,1D, 2D, 3D, 4D deskriptory
- Dle tématické skupiny
  - Početní, topologické, QM, povrchové, …

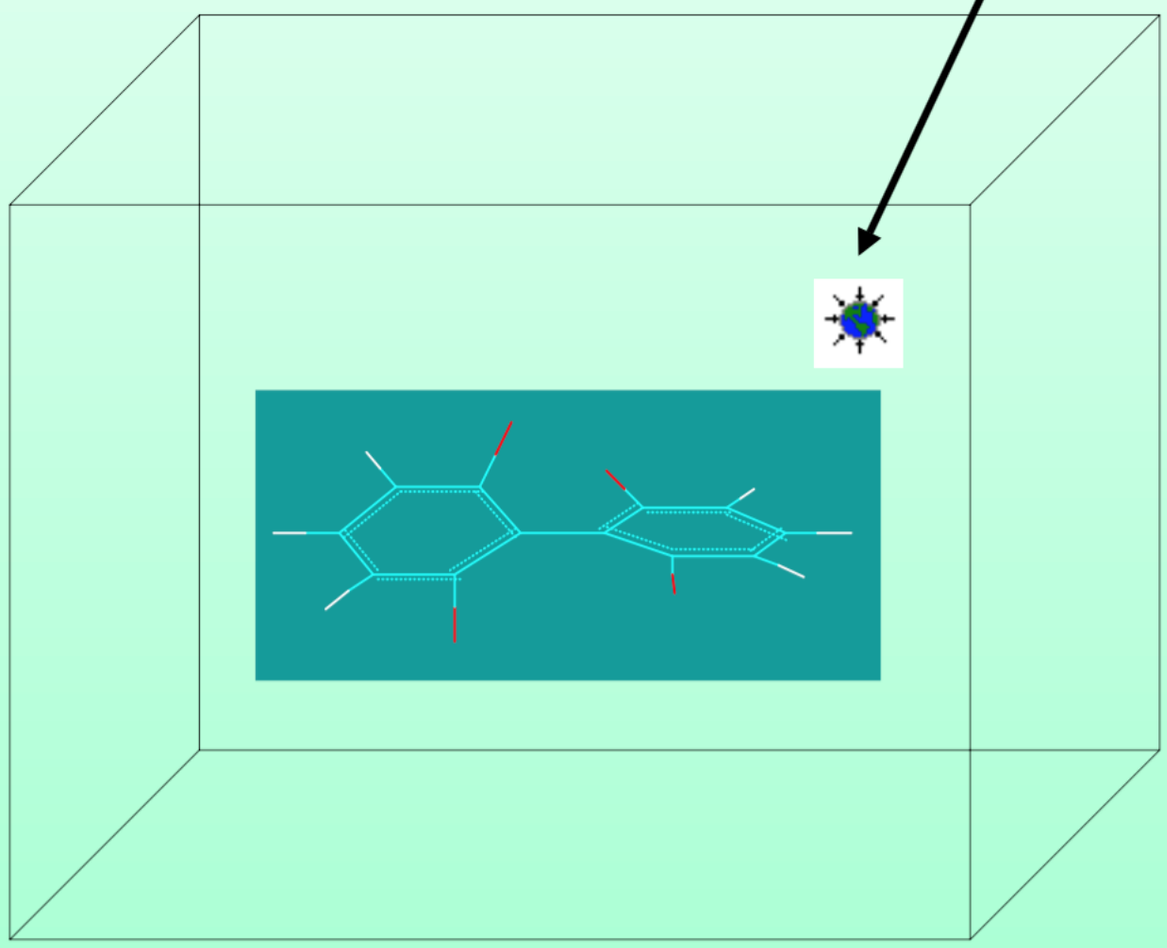
0D, 1D, 2D a 3D deskriptory



4D deskriptory

probes

- steric
- electronic
- hydrophobic



Atom list

0D

counting

summing

Substructure list

1D

counting

structural keys

molecular graph

2D

graph invariants

topographic descriptors

topostructural descriptors

topochemical descriptors

topological information indices

molecular geometry
x, y, z coordinates

3D

geometrical descriptors

bulk descriptors

quantum-chemical descriptors

molecular surface descriptors

grid-based QSAR techniques

4D

interaction energy values

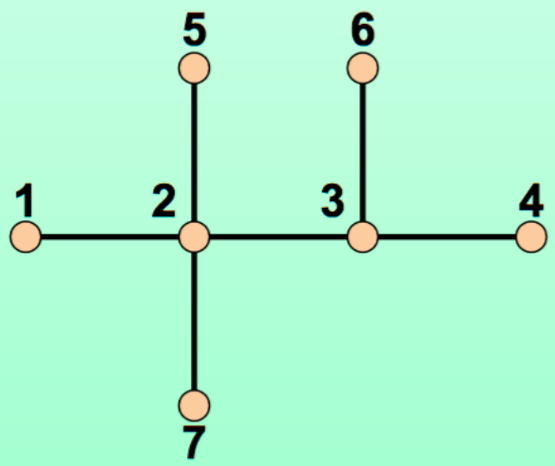
Molekulový graf

Mathematical object defined as

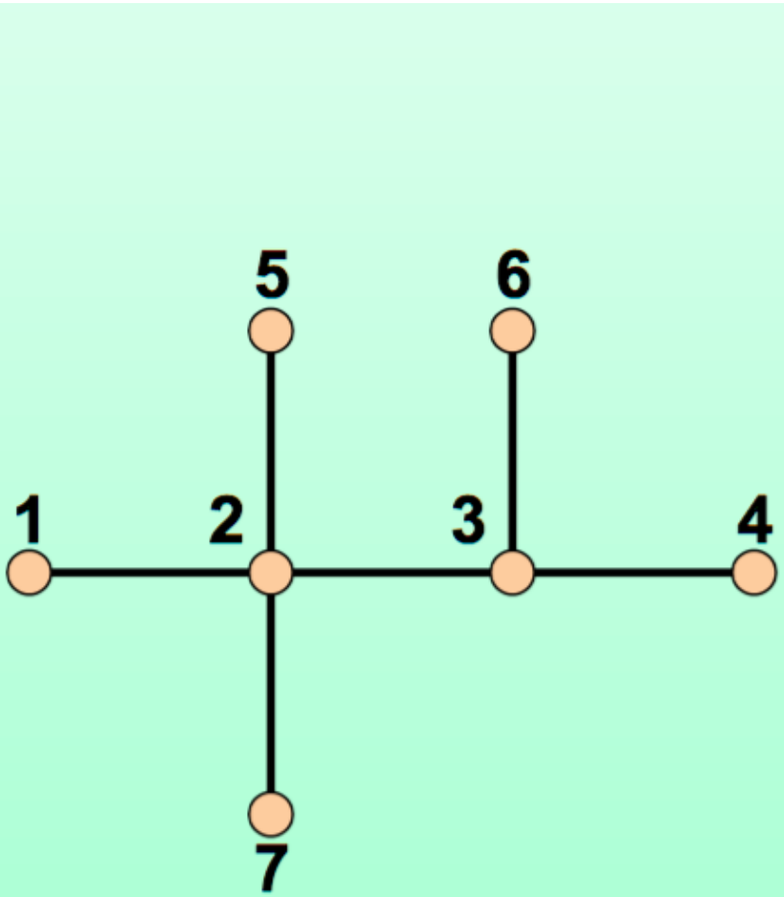
G = ($\mathcal{V}$, $\mathcal{E}$)

set $\mathcal{V}$ vertices ⟹ atoms

set $\mathcal{E}$ edges ⟹ bonds



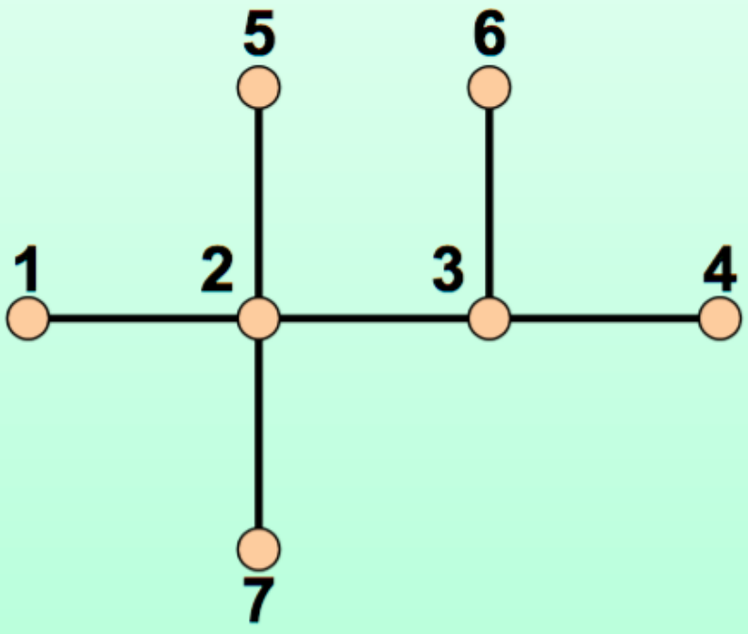
Matice sousednosti



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

Matice vzdáleností

s_i It is the row sum of the vertex distance matrix



	1	2	3	4	5	6	7	s_i	η_i
1	0	1	2	3	2	3	2	13	3
2	1	0	1	2	1	2	1	8	2
3	2	1	0	1	2	1	2	9	2
4	3	2	1	0	3	2	3	14	3
5	2	1	2	3	0	3	2	13	3
6	3	2	1	2	3	0	3	14	3
7	2	1	2	3	2	3	0	13	3

The distance d_{ij} between two vertices is the smallest number of edges between them.

s_i is high for terminal vertices and low for central vertices

Další rozdělení deskriptorů

molecular graph

graph invariants

molecular geometry
x, y, z coordinates

topostructural descriptors

topochemical descriptors

topographic descriptors

topological information indices

3D-Wiener index
3D-Balaban index
D/D index
.....

Wiener index, Hosoya Z index
Zagreb indices, Mohar indices
Randic connectivity index
Balaban distance connectivity index
Schultz molecular topological index
Kier shape descriptors
eigenvalues of the adjacency matrix
eigenvalues of the distance matrix
Kirchhoff number
detour index
topological charge indices
.....

Kier-Hall valence connectivity indices
Burden eigenvalues
BCUT descriptors
Kier alpha-modified shape descriptors
2D autocorrelation descriptors
.....

total information content on .....
mean information content on .....

molecular geometry
x, y, z coordinates

geometrical descriptors

grid-based QSAR techniques

quantum-chemical descriptors

volume descriptors

interaction energy values

charges
electronegativities
superdelocalizability
hardness
softness
E_{LUMO}
E_{HOMO}
.....

van der Waals volume
geometric volume
.....

CoMFA, GRID
G-WHIM descriptors
.....

molecular surface

gravitational indices
3D-Morse descriptors
EVA descriptors
EEVA descriptors
WHIM descriptors
GETAWAY descriptors
.....

solvent-accessible surface area
CPSA descriptors
molecular shape analysis
Mezey 3D shape analysis
.....

Software pro výpočet deskriptorů

- PaDEL – opensource (1444 1D, 2D descriptors and 431 3D descriptors)
- Dragon – komerční software (5270 molecular descriptors)

Výběr deskriptorů

- Manualní – dle intuice a korelací
- Poloautomatický
- Automatický – backward, forward nebo stepwise