

C7790

Počítačová chemie a molekulové modelování I

C7800 Počítačová chemie a molekulové modelování I - cvičení

3. Struktura

Petr Kulhánek

kulhanek@chemi.muni.cz

Národní centrum pro výzkum biomolekul, Přírodovědecká fakulta
Masarykova univerzita, Kotlářská 2, CZ-61137 Brno

Obsah - přednáška

➤ Výpočetní chemie

definice, výpočetní chemie versus experiment, přehled řešených projektů, experimentální metody s atomárním a jednomolekulárním rozlišením

➤ Kvantová mechanika I

stručný úvod, Bornova-Oppenheimerova aproximace, koncept hyperploch potenciální energie, stručný přehled metod pro výpočet potenciální energie

➤ **Struktura**

struktura, vizualizace, formáty, typy souřadnic (interní, kartézské)

➤ Plochy potenciální energie I

definice, stacionární body, jejich charakterizace a význam, optimalizační metody, lokální a globální minima

➤ Kvantová mechanika II

volná částice, tuhý rotátor, harmonický oscilátor, atom vodíku, variační a poruchové metody, Hartree-Fockova metoda, semiempirické metody

➤ Plochy potenciální energie II

reakční cesty a konformační přeměny, reakční koordináta, hledání tranzitních stavů, vztah potenciální energie k termodynamickým veličinám, primární a sekundární izotopový efekt

➤ Molekulová mechanika I

silová pole, vazebné a nevazebné interakce, dalekodosahové interakce, bodové náboje, přehled silových polí

➤ Molekulová dynamika

vývoj systému v čase, pohybové rovnice, přehled integračních metod, vlastnosti systému, termostaty, barostaty

➤ Kvantová mechanika III

post-HF metody (MP_x, CC), CBS, DFT metody, korekce disperzních interakcí, BSSE

➤ Molekulová mechanika II

dalekodosahové interakce, modelování rozpouštědel, polarizovatelná silová pole

Konfigurační prostor

$E(\mathbf{R})$

\mathbf{R} = bod v $3N$ rozměrném prostoru (N je počet atomů)

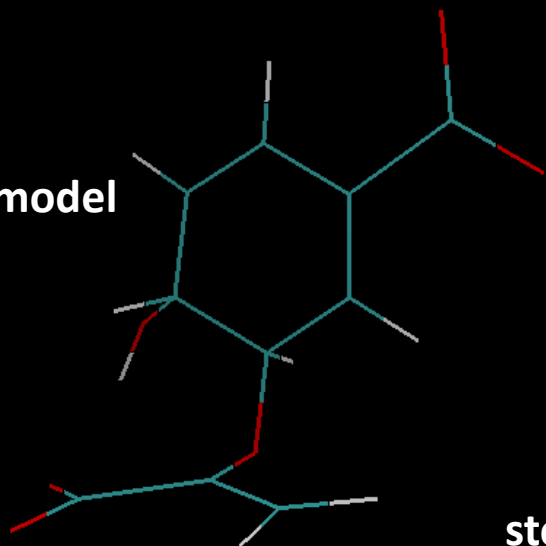
$$\mathbf{R} = \{ x_1, y_1, z_1, x_2, y_2, z_2, \dots, x_N, y_N, z_N \}$$

kartézské souřadnice
prvního atomu

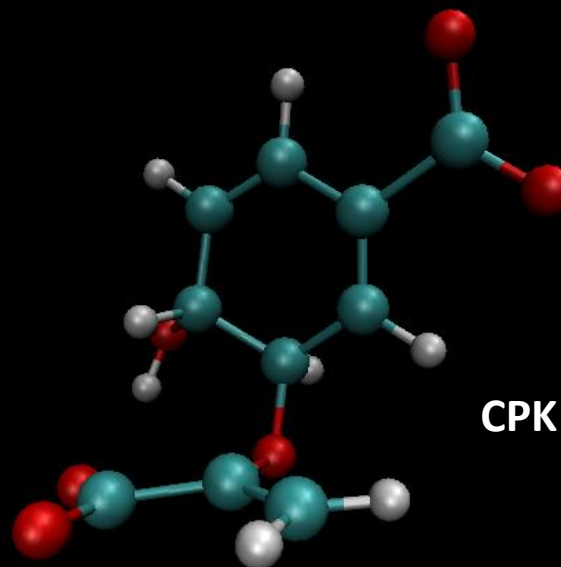
Jednotlivé body tvoří konfigurační prostor. **Každý bod** v konfiguračním prostoru pak představuje **unikátní strukturu** daného systému.

Modely – malé molekuly

čárový model

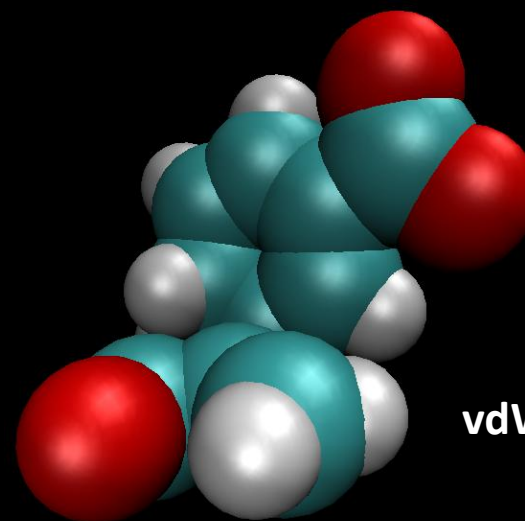
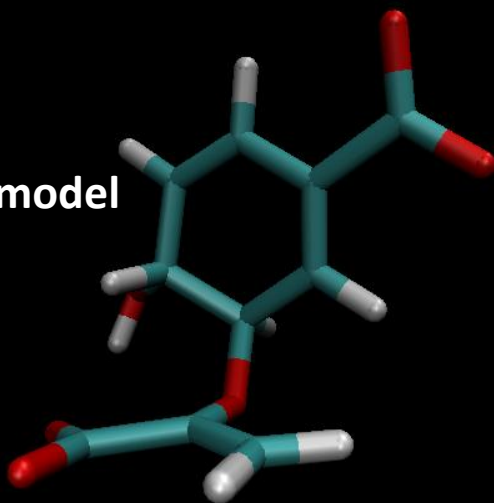


stejná struktura
jiná vizualizace



CPK model

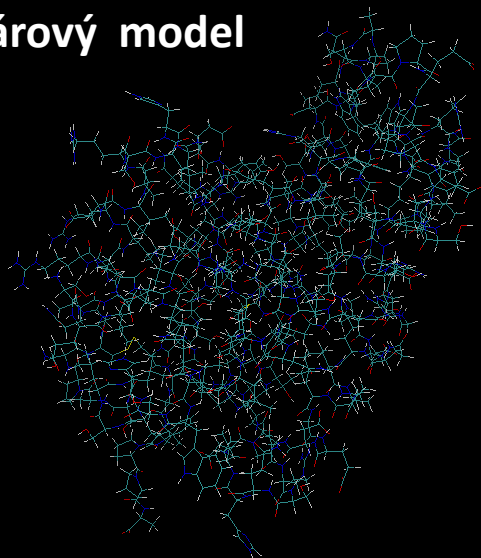
tyčinkový model



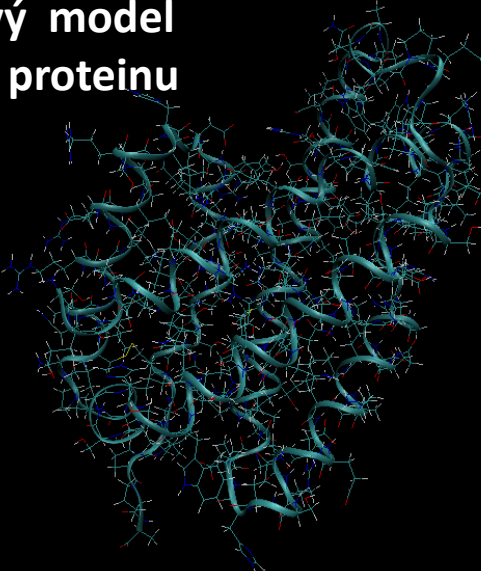
vdW model

Modely – biomolekuly

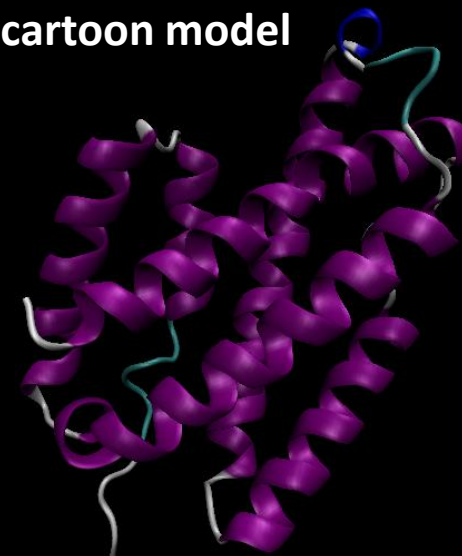
čárový model



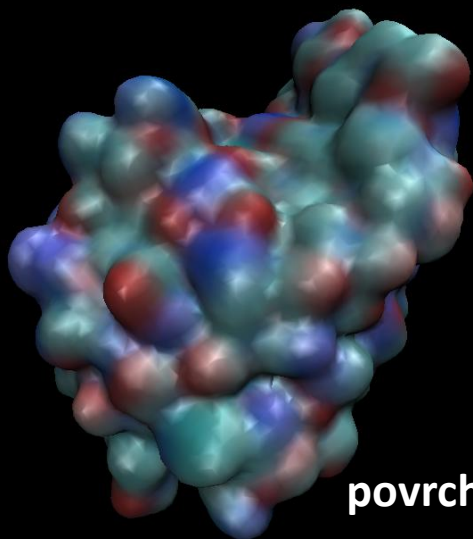
čárový model
páteř proteinu



cartoon model



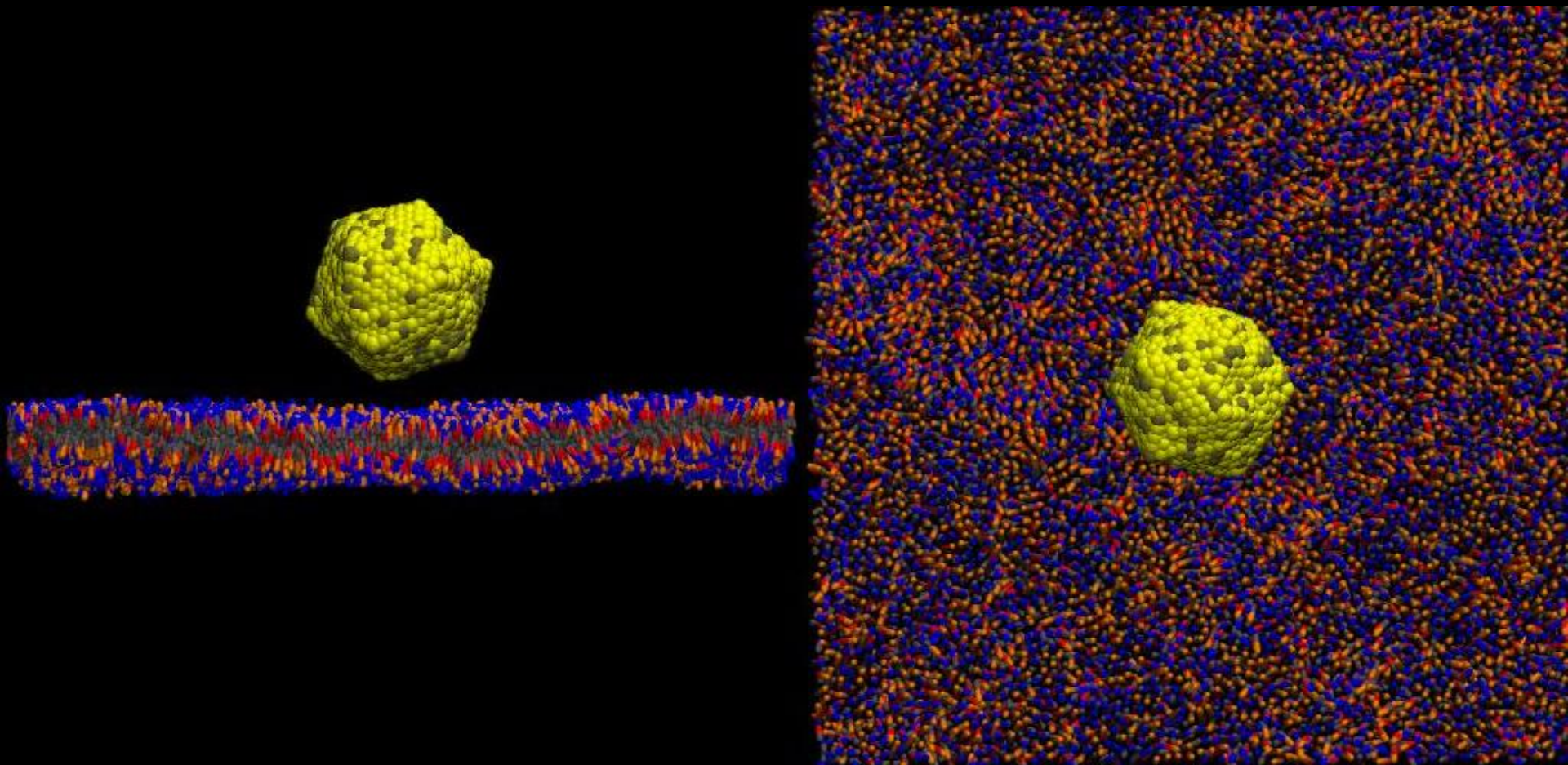
stejná struktura
jiná vizualizace



povrch biomolekuly

Různé modely slouží k zvýraznění určité strukturní informace nebo vnitřní vlastnosti molekuly či uskupení molekul, které pak usnadňuje snadnější pochopení studovaného problému.

Hrubozrné modely



Pustit video!

Počítačová reprezentace struktury

Strukturu lze reprezentovat různým způsobem. V chemii se používá více jak 100 formátů.

Formát XYZ

polohy jsou v angstroemech(Å)

	24			
	chorismate			
počet atomů	C	-1.86100	-0.57700	0.31800
komentář	O	-2.56800	0.47600	0.32600
značka x y z	O	-2.20900	-1.75300	0.64200
značka x y z	C	-0.38900	-0.41000	-0.18800
.....
značka x y z	H	-0.50900	1.67900	-0.44800

Formát **xyz** je textový soubor s volným formátováním (hodnoty ve sloupcích mohou být odděleny libovolným počtem mezer nebo jiných bílých znaků).

Džungle formátů I

acr	-- ACR format	csr	-- Accelrys/MSI Quanta CSR format
adf	-- ADF cartesian input format	cssr	-- CSD CSSR format
adfout	-- ADF output format	ct	-- ChemDraw Connection Table format
alc	-- Alchemy format	cub	-- OpenDX cube format for APBS
arc	-- Accelrys/MSI Biosym/Insight II CAR format	cube	-- OpenDX cube format for APBS
bgf	-- MSI BGF format	dmol	-- DMol3 coordinates format
box	-- Dock 3.5 Box format	dx	-- OpenDX cube format for APBS
bs	-- Ball and Stick format	ent	-- Protein Data Bank format
c3d1	-- Chem3D Cartesian 1 format	fa	-- FASTA format
c3d2	-- Chem3D Cartesian 2 format	fasta	-- FASTA format
cac	-- CAChe MolStruct format	fch	-- Gaussian formatted checkpoint file format
caccrt	-- Cacao Cartesian format	fchk	-- Gaussian formatted checkpoint file format
cache	-- CAChe MolStruct format	fck	-- Gaussian formatted checkpoint file format
cacint	-- Cacao Internal format	feat	-- Feature format
can	-- Canonical SMILES format.	fh	-- Fenske-Hall Z-Matrix format
car	-- Accelrys/MSI Biosym/Insight II CAR format	fix	-- SMILES FIX format
ccc	-- CCC format	fpt	-- Fingerprint format
cdx	-- ChemDraw binary format	fract	-- Free Form Fractional format
cdxml	-- ChemDraw CDXML format	fs	-- FastSearching
cht	-- Chemtool format	fsa	-- FASTA format
cif	-- Crystallographic Information File	g03	-- Gaussian98/03 Output
ck	-- ChemKin format	g92	-- Gaussian98/03 Output
cml	-- Chemical Markup Language	g94	-- Gaussian98/03 Output
cmlr	-- CML Reaction format	g98	-- Gaussian98/03 Output
com	-- Gaussian 98/03 Input	gal	-- Gaussian98/03 Output
copy	-- Copies raw text	gam	-- GAMESS Output
crk2d	-- Chemical Resource Kit diagram(2D)	gamin	-- GAMESS Input
crk3d	-- Chemical Resource Kit 3D format	gamout	-- GAMESS Output

Džungle formátů II

gau	-- Gaussian 98/03 Input	mopcrt	-- MOPAC Cartesian format
gjc	-- Gaussian 98/03 Input	mopin	-- MOPAC Internal
gjf	-- Gaussian 98/03 Input	mopout	-- MOPAC Output format
gpr	-- Ghemical format	mpc	-- MOPAC Cartesian format
gr96	-- GROMOS96 format	mpd	-- Sybyl descriptor format
gukin	-- GAMESS-UK Input	mpqc	-- MPQC output format
gukout	-- GAMESS-UK Output	mpqcin	-- MPQC simplified input format
gzmat	-- Gaussian Z-Matrix Input	msi	-- Accelrys/MSI Cerius II MSI format
hin	-- HyperChem HIN format	msms	-- M.F. Sanner's MSMS input format
inchi	-- InChI format	nw	-- NWChem input format
inp	-- GAMESS Input	nwo	-- NWChem output format
ins	-- ShelX format	outmol	-- DMol3 coordinates format
jin	-- Jaguar input format	pc	-- PubChem format
jout	-- Jaguar output format	pcm	-- PCModel Format
k	-- Compare molecules using InChI	pdb	-- Protein Data Bank format
mcdl	-- MCDL format	png	-- PNG files with embedded data
mcif	-- Macromolecular Crystallographic Information	pov	-- POV-Ray input format
mdl	-- MDL MOL format	pqr	-- PQR format
ml2	-- Sybyl Mol2 format	pqs	-- Parallel Quantum Solutions format
mmcif	-- Macromolecular Crystallographic Information	prep	-- Amber Prep format
mmd	-- MacroModel format	qcin	-- Q-Chem input format
mmod	-- MacroModel format	qcout	-- Q-Chem output format
mol	-- MDL MOL format	report	-- Open Babel report format
mol2	-- Sybyl Mol2 format	res	-- ShelX format
molden	-- Molden input format	rsmi	-- Reaction SMILES format
molreport	-- Open Babel molecule report	rxn	-- MDL RXN format
moo	-- MOPAC Output format	sd	-- MDL MOL format
mop	-- MOPAC Cartesian format	sdf	-- MDL MOL format

Džungle formátů III

smi	-- SMILES format
smiles	-- SMILES format
sy2	-- Sybyl Mol2 format
t41	-- ADF TAPE41 format
tdd	-- Thermo format
test	-- Test format
therm	-- Thermo format
tmol	-- TurboMole Coordinate format
txt	-- Title format
txyz	-- Tinker MM2 format
unixyz	-- UniChem XYZ format
vmol	-- ViewMol format
xed	-- XED format
xml	-- General XML format
xtc	-- XTC format
xyz	-- XYZ cartesian coordinates format
yob	-- YASARA.org YOB format
zin	-- ZINDO input format

Výše uvedené formáty obsahují většinou kromě 3D/2D struktury také doprovodné informace jako jsou konektivita, parametry silových polí, náboje, různé vlastnosti apod.

OpenBabel

Open Babel is a chemical toolbox designed to speak the many languages of chemical data. It's an open, collaborative project allowing anyone to search, convert, analyze, or store data from molecular modeling, chemistry, solid-state materials, biochemistry, or related areas.

http://openbabel.org/wiki/Main_Page

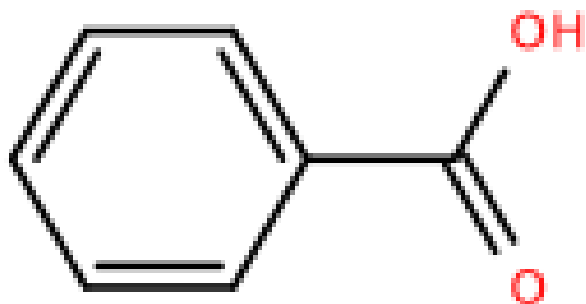
Konverze programem openbabel:

```
$ module add openbabel  
$ babel input.xyz output.mol2
```

Seznam podporovaných formátů:

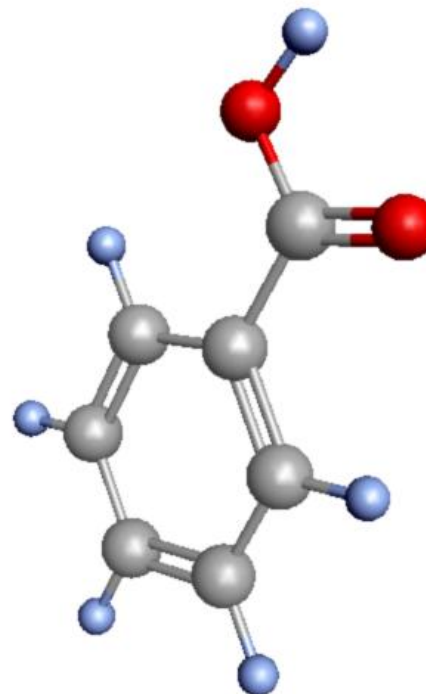
```
$ babel -L formats
```

2D versus 3D struktura



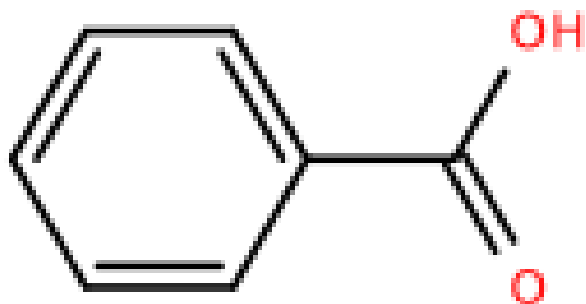
kyselina benzoová

2D struktura obsahuje informaci o atomech a vazbách, kterými jsou spojeny. Tato informace popisuje konstituci (topologii) systému.

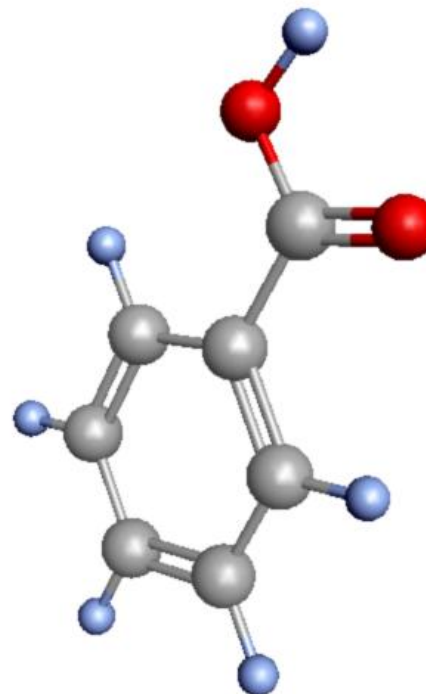


3D struktura obsahuje informaci o prostorovém rozmístění atomů. Ostatní informace (např. vazby) jsou dopočitatelné.

3D -> 2D převod



kyselina benzoová



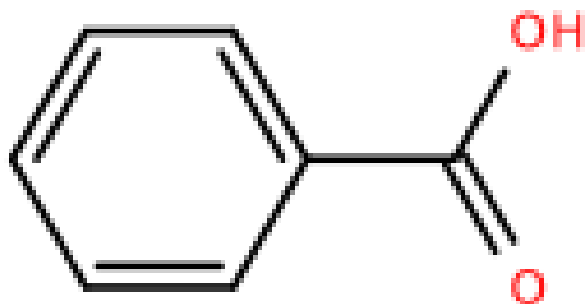
2D struktura obsahuje informaci o atomech a vazbách, kterými jsou spojeny. Tato informace popisuje konstituci (topologii) systému.

3D struktura obsahuje informaci o prostorovém rozmístění atomů. Ostatní informace (např. vazby) jsou dopočitatelné.

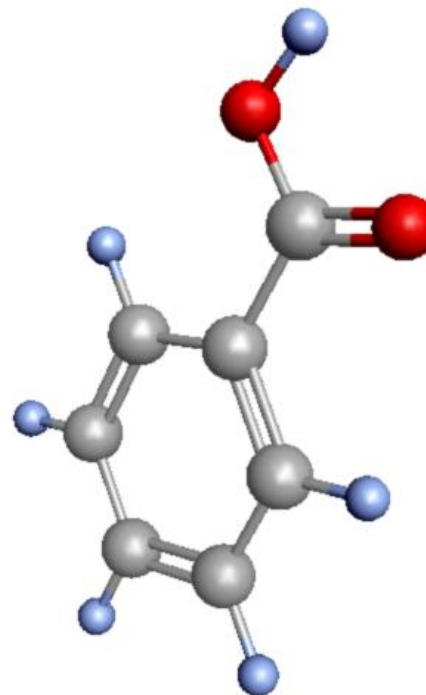


převod je snadný

2D -> 3D převod



kyselina benzoová



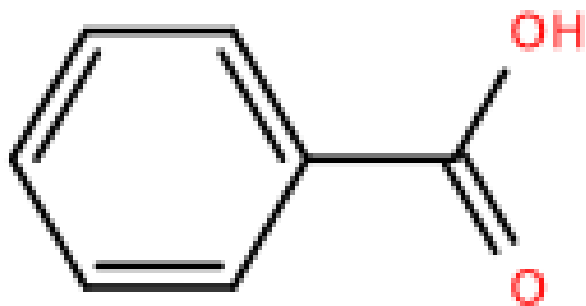
2D struktura obsahuje informaci o atomech a vazbách, kterými jsou spojeny. Tato informace popisuje konstituci (topologii) systému.

3D struktura obsahuje informaci o prostorovém rozmístění atomů. Ostatní informace (např. vazby) jsou dopočitatelné.

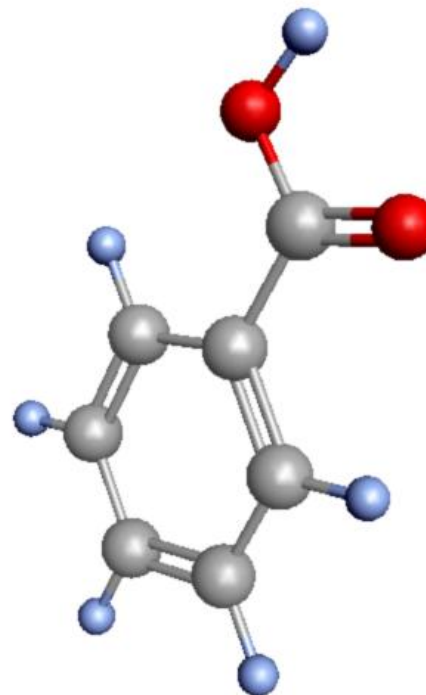
převod je komplikovaný



2D -> 3D převod



kyselina benzoová



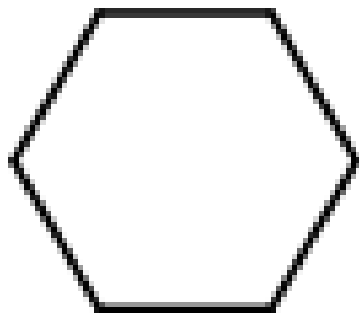
2D struktura obsahuje informaci o atomech a vazbách, kterými jsou spojeny. Tato informace popisuje konstituci (topologii) systému.

3D struktura obsahuje informaci o prostorovém rozmístění atomů. Ostatní informace (např. vazby) jsou dopočitatelné.

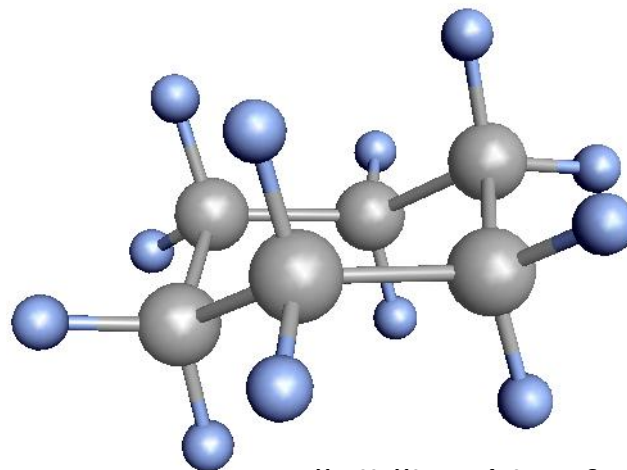
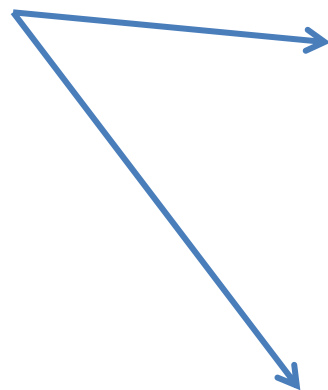


převod je komplikovaný
u velkých systémů nemusí být jednoznačný v důsledku existence více konformerů

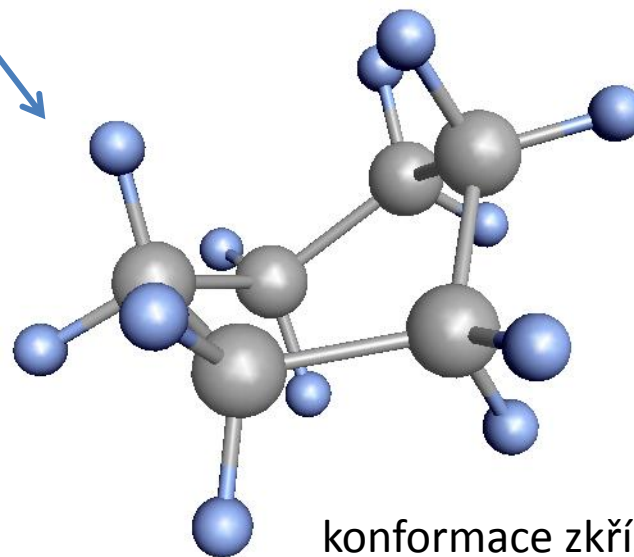
2D -> 3D převod, komplikace



cyklohexan



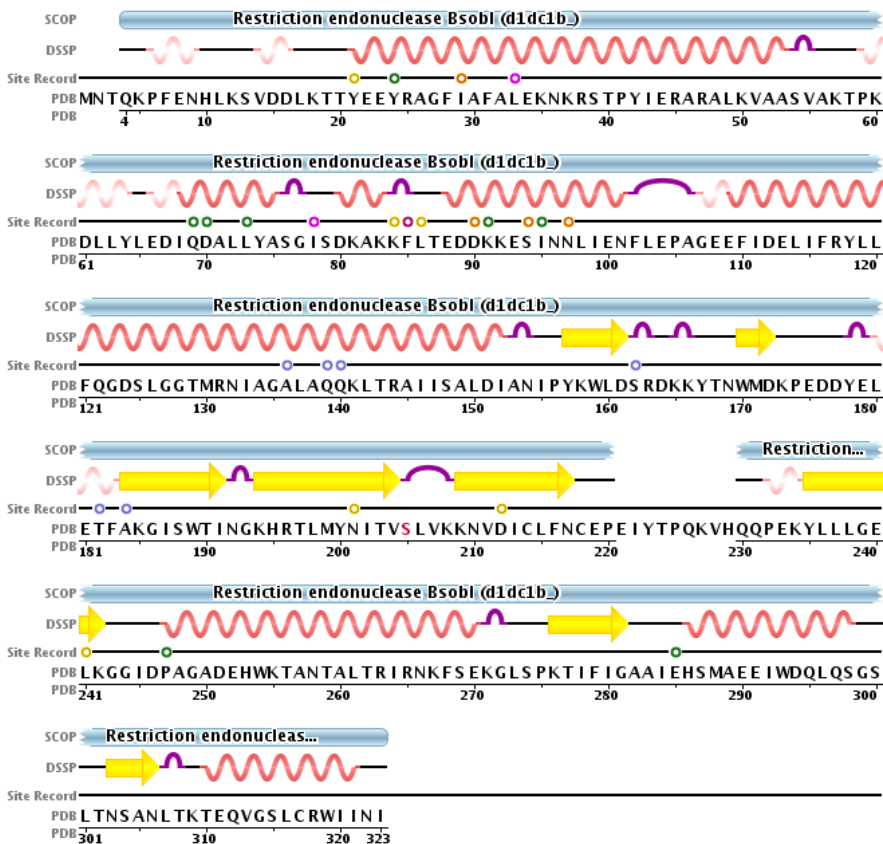
židličková konformace



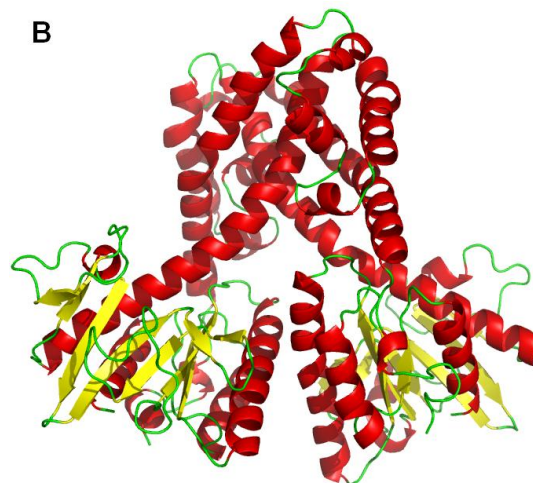
konformace zkřížená vanička

2D -> 3D převod, komplikace

Stejná primární struktura
(sekvence aminokyselin).

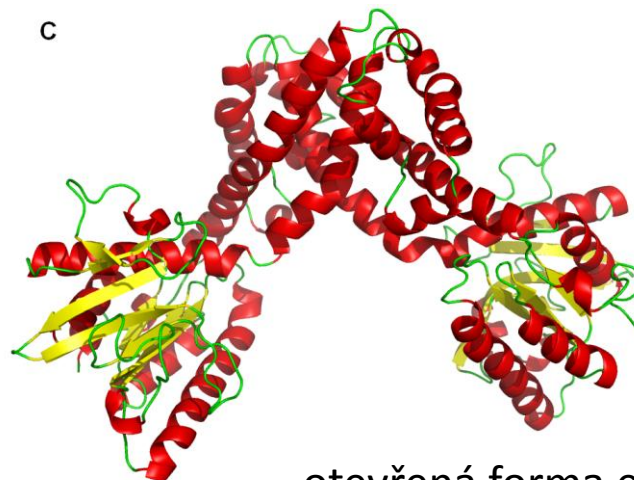


B



zavřená forma enzymu

C



otevřená forma enzymu

Využití 2D struktur

Representace molekul ve 2D formátech se využívá převážně pro ukládání informací do databází a jejich prohledávání, dále k předpovědi chemických vlastností molekul pomocí chemoinformatických přístupů.

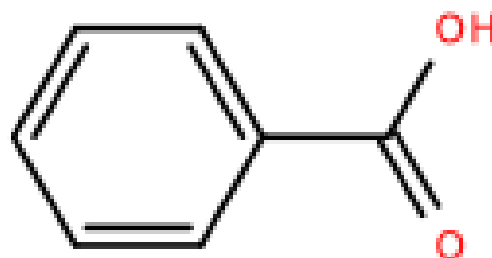
Nejrozšířenější formáty:

- **SMILES** (Simplified molecular-input line-entry system)

```
C(=O)Oc1ccccc1
```

- **InChI** (IUPAC International Chemical Identifier)

```
InChI=1S/C7H6O2/c8-7(9)6-4-2-1-3-5-6/h1-5H,(H,8,9)
```



kyselina benzoová

Zdroje 3D struktur

Experimentální metody

- struktury určení pomocí rentgenové strukturní analýzy

Cambridge Structural Database (CSD)

<http://www.ccdc.cam.ac.uk/Solutions/CSDSystem/Pages/CSD.aspx>

Protein Data Bank (PDB)

<http://www.pdb.org>

- struktury určené pomocí NMR

Protein Data Bank (PDB)

<http://www.pdb.org>

Zdroje 3D struktur

Experimentální metody

- struktury určení pomocí rentgenové strukturní analýzy

Cambridge Structural Database (CSD)

<http://www.ccdc.cam.ac.uk/Solutions/CSDSystem/Pages/CSD.aspx>

Protein Data Bank (PDB)

<http://www.pdb.org>

- struktury určené pomocí NMR

Protein Data Bank (PDB)

<http://www.pdb.org>

Výpočetní metody

- molekulové modelování
- homologní modelování

Programy pro vytváření/vizualizaci 3D struktur

http://en.wikipedia.org/wiki/List_of_molecular_graphics_systems

Avogadro

http://avogadro.openmolecules.net/wiki/Main_Page

Program pro stavbu a vizualizaci molekul. Volně dostupný pro MS Windows a Linux.

VMD

<http://www.ks.uiuc.edu/Research/vmd/>

Program pro vizualizaci molekul. Po bezplatné registraci dostupný pro MS Windows a Linux.

NEMESIS

<https://lcc.ncbr.muni.cz/whitezone/development/nemesis/>

Program pro stavbu a vizualizaci molekul. Ve vývoji.