

C7790

Počítačová chemie a molekulové modelování I

C7800 Počítačová chemie a molekulové modelování I - cvičení

8. Struktura

Petr Kulhánek

kulhanek@chemi.muni.cz

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

Souvislosti

makrosvět

stavy

(termodynamické vlastnosti, G , T , ...)

fenomenologická termodynamika

rovnovážná konstanta
rychlostní konstanta

volná energie
(Gibbsova/Helmholtzova)



partiční funkce

statistická termodynamika

mikrostavy

(mechanické vlastnosti, E)

mikrosvět

mikrostav \neq mikrosvět

Souvislosti

makrosvět

stavy

(termodynamické vlastnosti, G , T , ...)

fenomenologická termodynamika

rovnovážná konstanta
rychlostní konstanta

volná energie
(Gibbsova/Helmholtzova)



partiční funkce

statistická termodynamika

mikrostavy

(mechanické vlastnosti, E)

mikrosvět

fyzikální popis



mikrostav \neq mikrosvět

Opakování

$$\hat{H}\phi(\mathbf{x}, t) = i\hbar \frac{\partial \phi(\mathbf{x}, t)}{\partial t}$$

časově závislá Schrödingerova rovnice

Opakování

$$\hat{H}\phi(\mathbf{x}, t) = i\hbar \frac{\partial \phi(\mathbf{x}, t)}{\partial t}$$

časově závislá Schrödingerova rovnice

$$\phi(\mathbf{x}, t) = \psi(\mathbf{x}) f(t)$$

časově nezávislá Schrödingerova rovnice

$$\hat{H}\psi_k(\mathbf{x}) = E_k \psi_k(\mathbf{x})$$

system se může nacházet v několika kvantových stavech
stav je popsán vlnovou funkcí Ψ a má energii E

Opakování

$$\hat{H}\phi(\mathbf{x}, t) = i\hbar \frac{\partial \phi(\mathbf{x}, t)}{\partial t}$$

časově závislá Schrödingerova rovnice

$$\phi(\mathbf{x}, t) = \psi(\mathbf{x}) f(t)$$

časově nezávislá Schrödingerova rovnice

Bornova- Oppenheimerova aproximace

$$\psi(\mathbf{x}) = \Psi(\mathbf{r}, \mathbf{R}) \chi(\mathbf{R})$$

$$\hat{H}\psi_k(\mathbf{x}) = E_k \psi_k(\mathbf{x})$$

$$\hat{H}_e \Psi_m(\mathbf{r}, \mathbf{R}) = E_m(\mathbf{R}) \Psi_m(\mathbf{r}, \mathbf{R})$$

pohyb elektronů ve statickém poli jader
elektronické vlastnosti systému

$$\hat{H}_R \chi_l(\mathbf{R}) = E_{VRT,l} \chi_l(\mathbf{R})$$

pohyb jader v efektivním poli elektronů
vibrace, rotace, translace

Opakování

$$\hat{H}_e \Psi_m(\mathbf{r}, \mathbf{R}) = E_m(\mathbf{R}) \Psi_m(\mathbf{r}, \mathbf{R})$$

pohyb elektronů ve statickém poli jader
elektronické vlastnosti systému

$$\hat{H}_R \chi_l(\mathbf{R}) = E_{VRT,l} \chi_l(\mathbf{R})$$

pohyb jader v efektivním poli elektronů
vibrace, rotace, translace

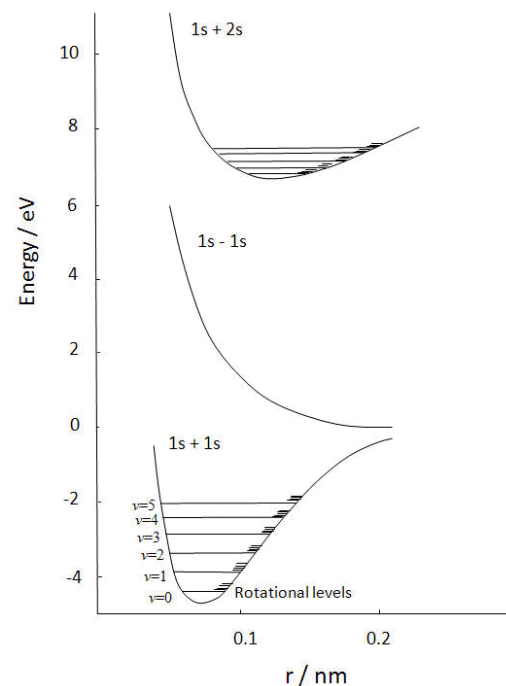
elektronická složka energie

vibračně, rotačně,
translační složka
energie

$$E_k = E_m(R_{opt,m}) + E_{VRT,l}$$

výsledná energie stavu

optimální geometrie jader,
při které je E_m minimální



Opakování

$$\hat{H}_e \Psi_m(\mathbf{r}, \mathbf{R}) = E_m(\mathbf{R}) \Psi_m(\mathbf{r}, \mathbf{R})$$

pohyb elektronů ve statickém poli jader
elektronické vlastnosti systému

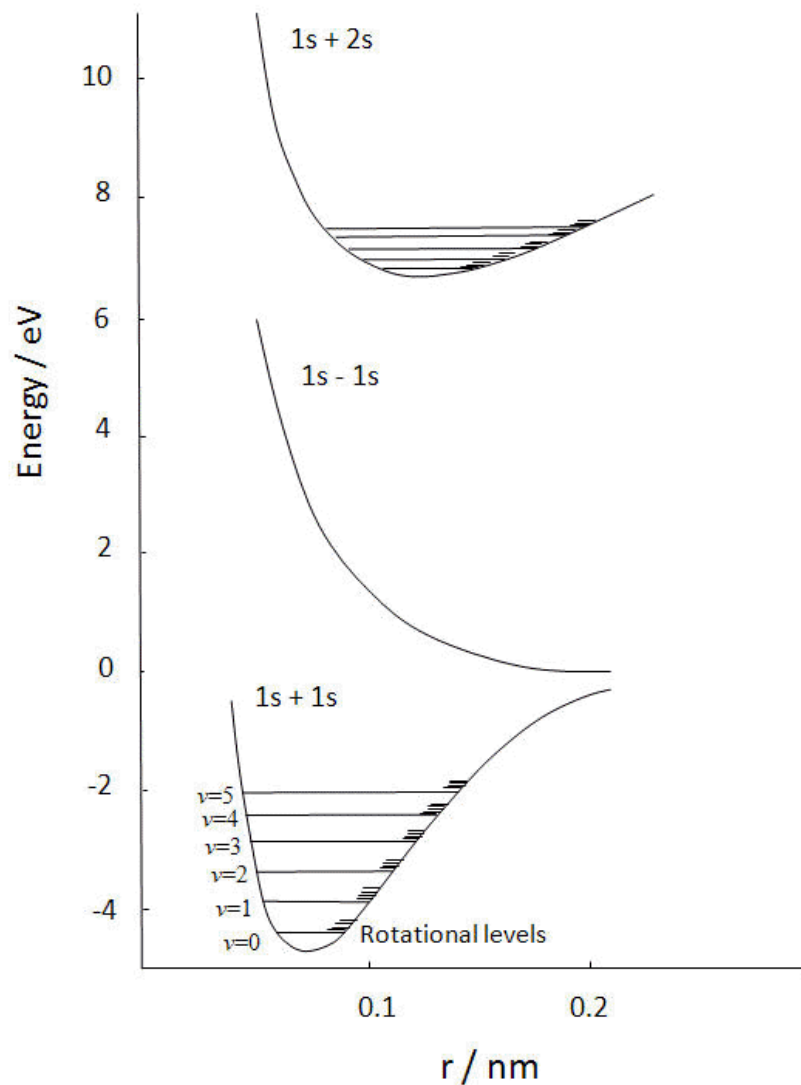
$$\hat{H}_R \chi_l(\mathbf{R}) = E_{VRT,l} \chi_l(\mathbf{R})$$

pohyb jader v efektivním poli elektronů
vibrace, rotace, translace

je možné obdobným
způsobem dále rozdělit na
samostatné příspěvky
vibrační, rotační a translační

$$E_{VRT,l} = E_{V,i} + E_{R,j} + E_{T,k}$$

Molekula vodíku



Určete základní energie stavů 1s-1s a 1s+2s

Nápověda

- vibrační energie je kvantována

$$E_V = \left(v + \frac{1}{2} \right) h\nu \quad \text{kvantové vibrační číslo } 0, 1, 2, \dots$$

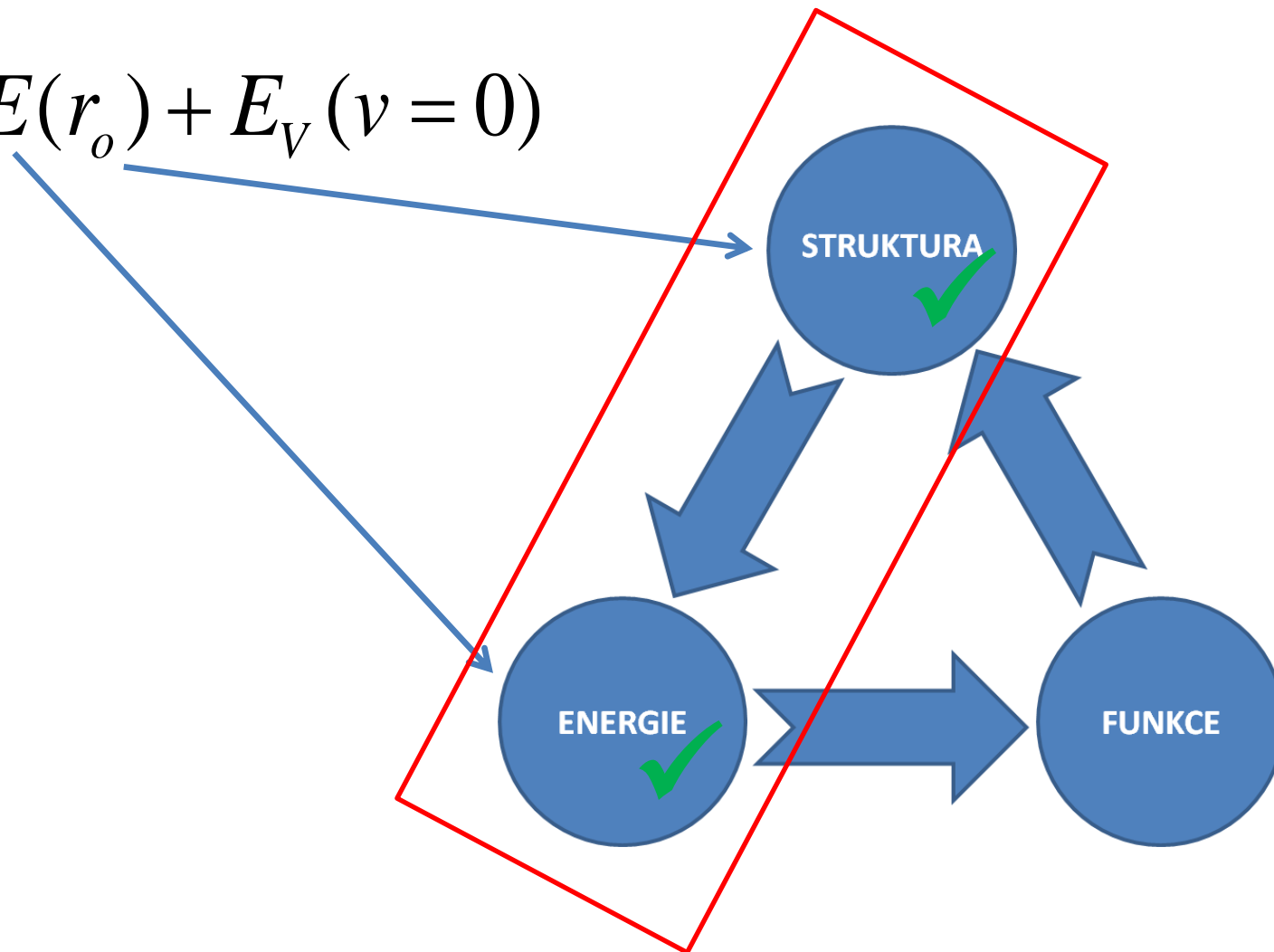
- rotační a translační stavy nebudeme uvažovat

Energie základního stavu 1s+1s:

$$E = E(r_o) + E_V (v = 0)$$

Struktura vs stav systému

$$E = E(r_o) + E_v (v = 0)$$



pouze část kvantově mechanického popisu stavu systému

Souvislosti

makrosvět

stavy

(termodynamické vlastnosti, G, T, ...)

fenomenologická termodynamika

rovnovážná konstanta
rychlostní konstanta

volná energie
(Gibbsova/Helmholtzova)

partiční funkce

statistická termodynamika

mikrostavy

(mechanické vlastnosti, E)

mikrosvět

fyzikální popis

$$\hat{H} \psi_k(\mathbf{x}) = E_k \psi_k(\mathbf{x})$$

neřešitelné pro makrosystémy

mikrostav \neq mikrosvět

Souvislosti

makrosvět

stavy

(termodynamické vlastnosti, G, T, ...)

fenomenologická termodynamika

rovnovážná konstanta
rychlostní konstanta

volná energie
(Gibbsova/Helmholtzova)

partiční funkce

statistická termodynamika

mikrostavy

(mechanické vlastnosti, E)

mikrosvět

fyzikální popis

$$\hat{H} \psi_k(\mathbf{x}) = E_k \psi_k(\mathbf{x})$$

neřešitelné pro makrosystémy

model

mikrostav \neq mikrosvět

Domácí úkol

1. Navrhněte vhodné energetické referenční stavy pro atom vodíku.
2. Navrhněte energetický referenční stav tak, aby byl stejný pro libovolný atom.
3. Odvoďte vztah pro excitační energii ze stavu n do stavu $n+1$ pro atom vodíku.
4. Navrhněte vhodný energetický referenční stav pro harmonický oscilátor.
5. Odvoďte vztah pro excitační energii ze stavu v do stavu $v+1$ pro harmonický oscilátor.
6. Navrhněte vhodný energetický referenční stav pro anharmonický oscilátor.
7. Může mít anharmonický oscilátor energii větší než D_e ?
8. Proč může mít tuhý rotátor nulovou energii a harmonický oscilátor a částice v potenciálové jámě ne?
9. Srovnejte energie pro základní stav translačního, rotačního a vibračního pohybu molekuly vodíku. V případě translačního pohybu uvažujte objem boxu, který pojme 1 mol ideálního plynu za standardních podmínek.
10. Z jakého důvodu je Bornova-Oppenheimerova aproximace použitelná?

Struktura

Konfigurační prostor

$E(\mathbf{R})$

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

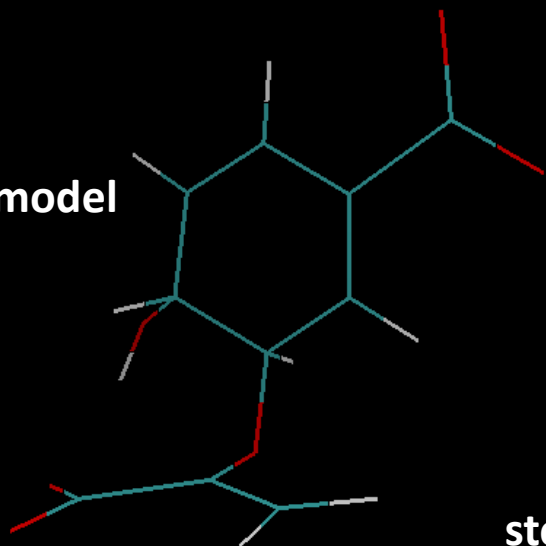
$$\mathbf{R} = \{x_1, y_1, z_1, x_2, y_2, z_2, \dots, x_N, y_N, z_N\}$$

kartézské souřadnice
prvního atomu

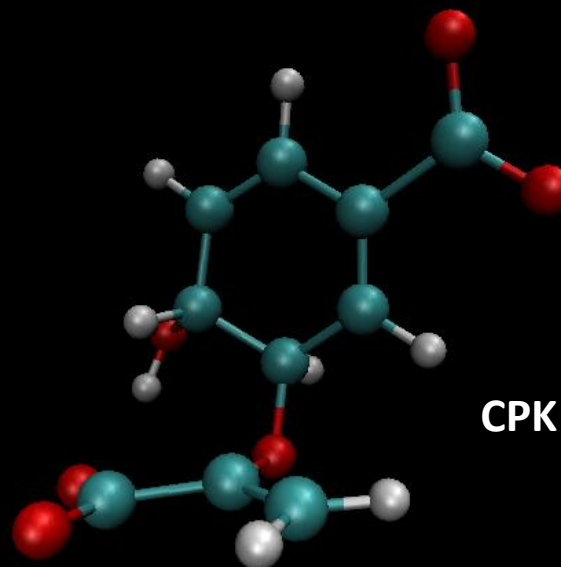
Jednotlivé body tvoří konfigurační prostor. **Každý bod** v konfiguračním prostoru pak představuje **unikátní strukturu** daného systému.

Modely – malé molekuly

čárový model

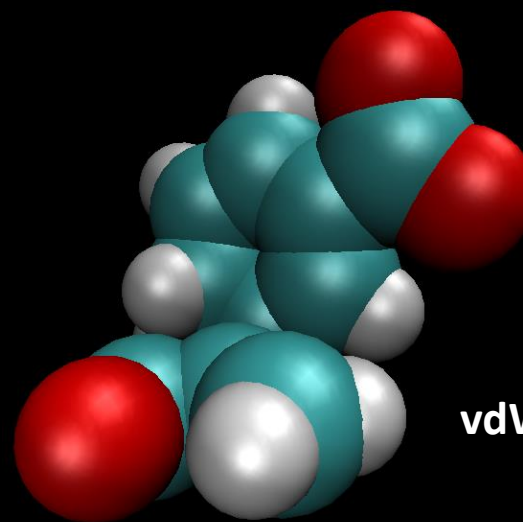
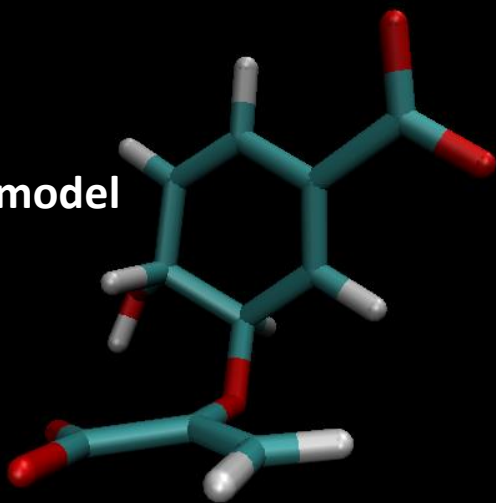


stejná struktura
jiná vizualizace



CPK model

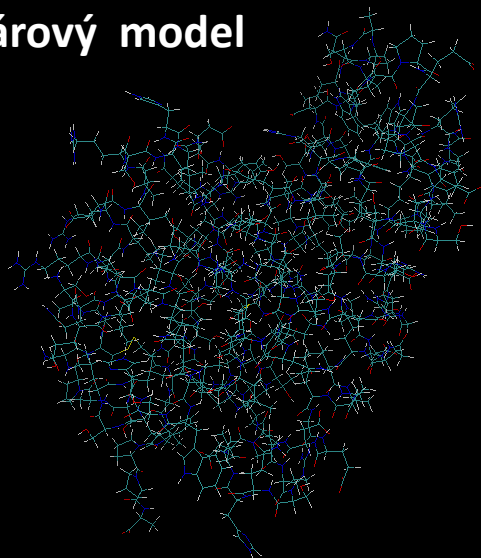
tyčinkový model



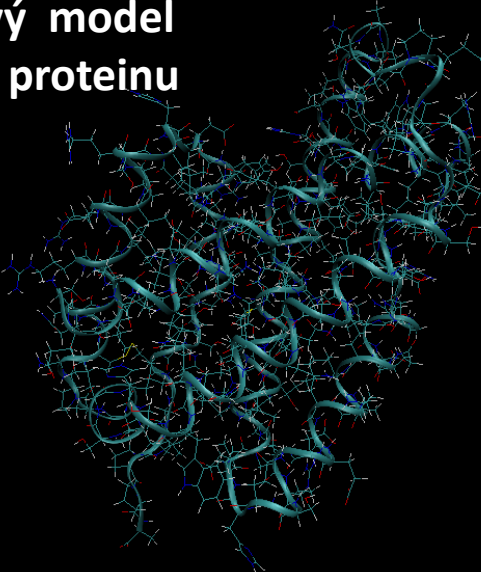
vdW model

Modely – biomolekuly

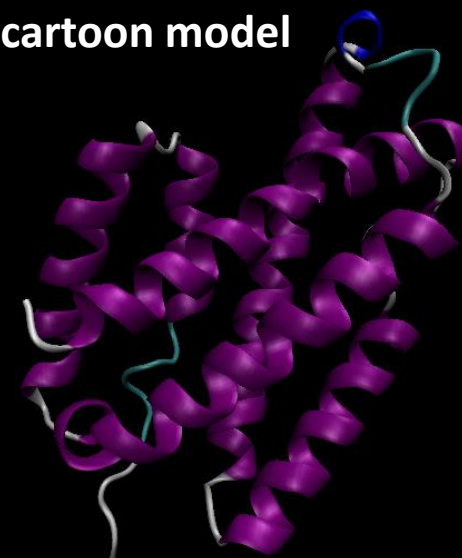
čárový model



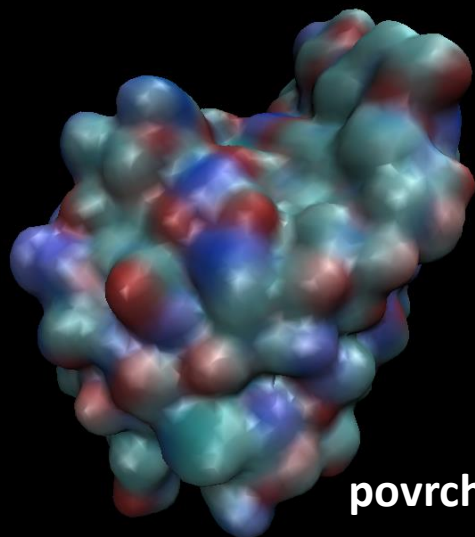
čárový model
páteř proteinu



cartoon model



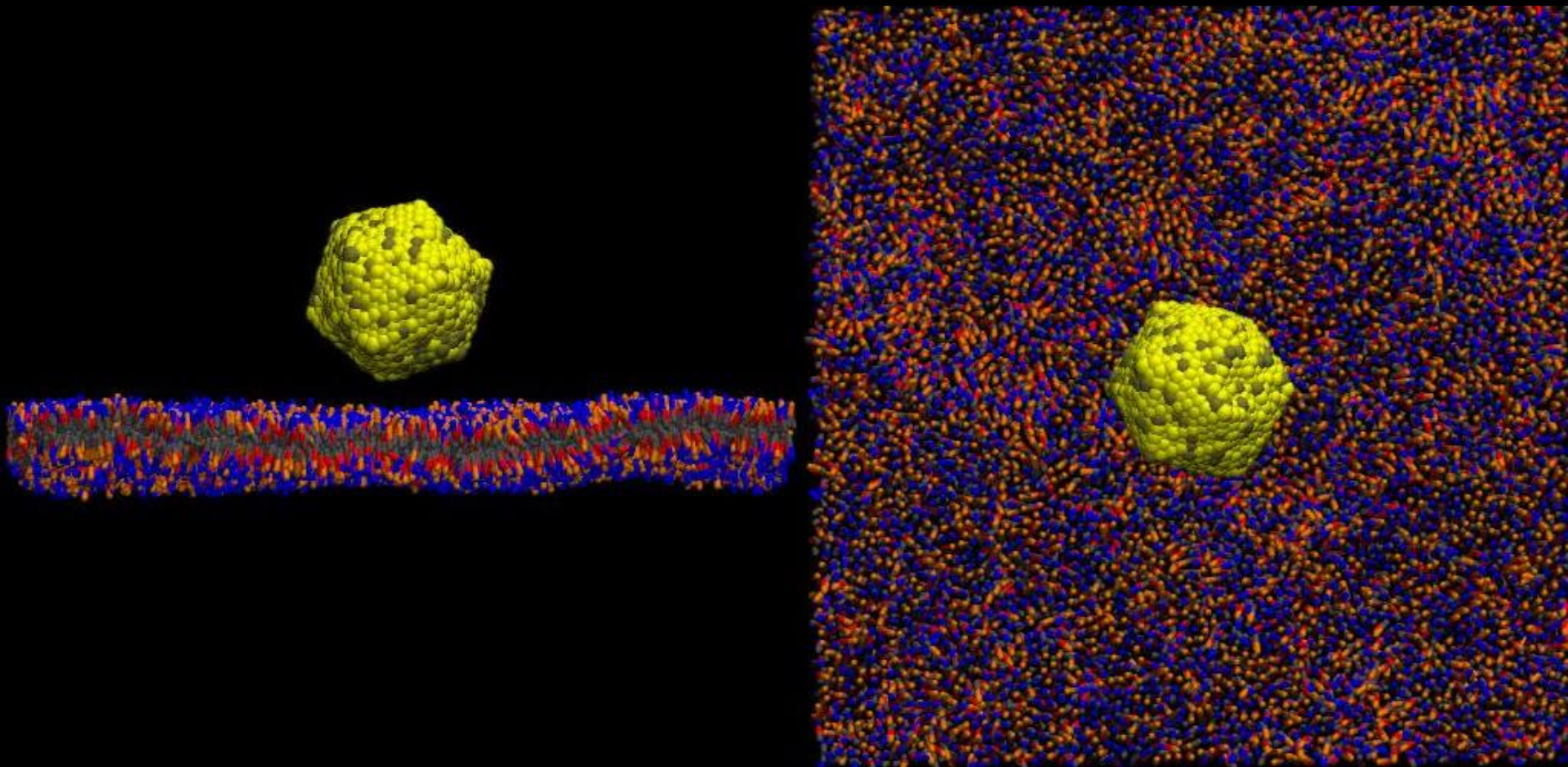
stejná struktura
jiná vizualizace



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



Počítačová reprezentace struktury

Strukturu lze reprezentovat různým způsobem. V chemii se používá více jak 100 formátů, jedná se buď o textové nebo binární soubory. Formát popisuje geometrii systému, jména atomů, skupin atomů, konektivitu mezi atomy (vazby) a další informace.

Geometrie systému může být uvedena v:

- kartézských souřadnicích
- interních souřadnicích
- varianty interních souřadnic

Kartézské vs interní souřadnice

Kartézské souřadnice

O	-0.180077	-0.046023	-0.062789
H	0.196208	-0.747659	0.498793
O	0.006537	1.047922	0.877207
H	-0.931885	1.299156	0.951390
	x	y	z

Počet stupňů volnosti:

3N

Interní souřadnice (Z-matrix)

O						
H	1	0.974298				
O	1	1.454349	2	96.868054		
H	3	0.974298	1	96.868054	2	239.552651
		vazebná délka		vazebný úhel		torzní úhel

Počet stupňů volnosti:

3N-6

3N-5 (lineární dvouatomová molekula)

Interní souřadnice

vazebná délka (a)

vazebný úhel (b)

torzní úhel (c)

1 O
2 H
3 O
4 H

1
1
3

0.974298
1.454349
0.974298

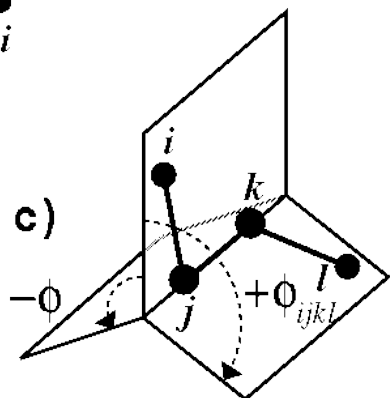
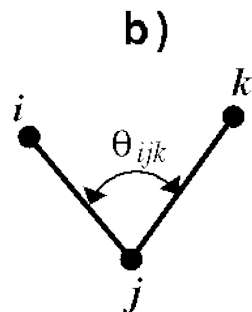
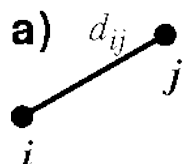
2-1
2
1

96.868054
96.868054

3-1-2
2

239.552651

4-3-1-2



<http://www.ccl.net/cca/documents/molecular-modeling/node4.html>

Domácí úkol

1. Zamyslete se nad výhodou a nevýhodou reprezentace geometrie systému pomocí kartézských a interních souřadnic.

Formát XYZ

polohy jsou v angströmech (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).

Formát popisuje pouze geometrii systému, neobsahuje informace o vazbách v systému. Program, který s formátem pracuje, musí tyto informace dopočítat (např. pomocí atomových poloměrů).

Formát PDB

Formát pdb se používá k ukládání struktur biomolekul a jejich komplexů.

klíčové slovo

číslo residua

ATOM	7	CB	SER	1	5.814	16.335	8.213	1.00	0.00
ATOM	8	HB2	SER	1	6.870	16.427	7.958	1.00	0.00
ATOM	9	HB3	SER	1	5.610	16.900	9.123	1.00	0.00
ATOM	10	OG	SER	1	5.491	14.946	8.427	1.00	0.00
ATOM	11	HG	SER	1	6.026	14.600	9.145	1.00	0.00
ATOM	12	C	SER	1	3.604	16.323	6.927	1.00	0.00
ATOM	13	O	SER	1	2.605	16.742	7.521	1.00	0.00
ATOM	14	N	GLN	2	3.567	15.251	6.134	1.00	0.00
ATOM	15	H	GLN	2	4.401	14.914	5.675	1.00	0.00
ATOM	18	CB	GLN	2	2.452	13.545	4.753	1.00	0.00
ATOM	19	HB2	GLN	2	3.006	12.696	5.153	1.00	0.00

číslo atomu

jméno atomu

jméno residua

kartézské souřadnice atomu v angströmech (Å)

Formát neobsahuje informace o vazbách v systému. Program, který s formátem pracuje, musí tyto informace dopočítat (na základě templátových struktur). U nestandardních residuích je možné použít klíčové slovo CONECT.

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

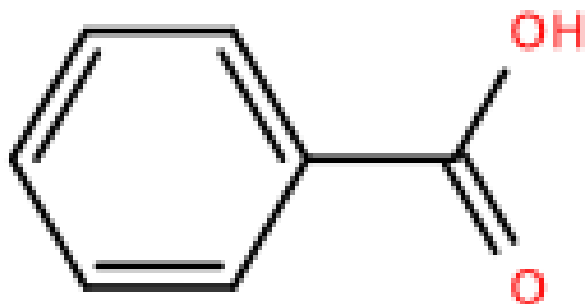
Výše uvedené formáty obsahují většinou kromě 3D/2D struktury také doprovodné informace jako jsou konektivita, parametry silových polí, náboje, různé vlastnosti apod.

OpenBabel prostředí pro konverzi mezi formáty včetně metod pro výpočet celé řady molekulárních vlastností (*chemoinformatika*)

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.

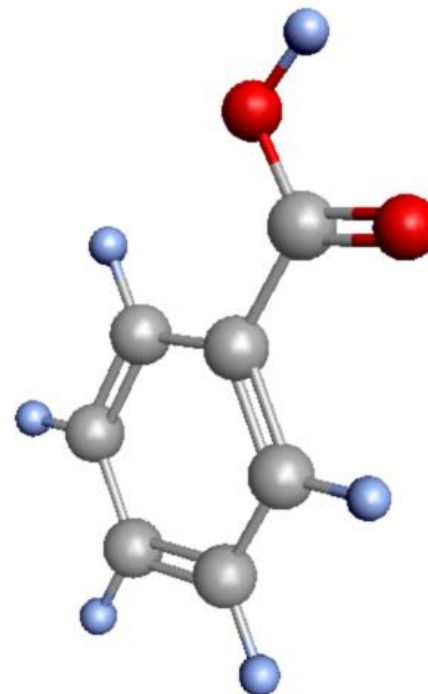
http://openbabel.org/wiki/Main_Page

2D versus 3D struktura



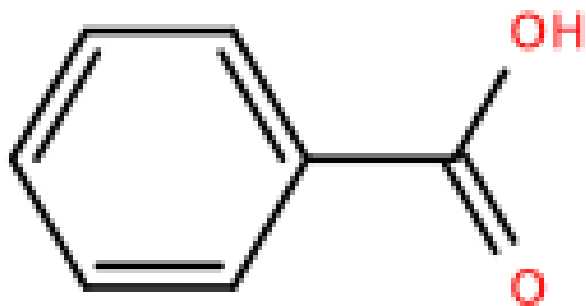
kyselina benzoová

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

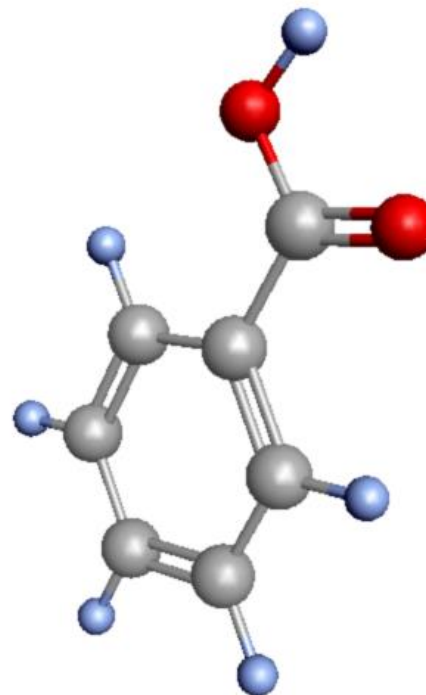


3D struktura obsahuje informaci o prostorovém rozmístění atomů. Ostatní informace (např. vazby) jsou dopočitatelné.

3D -> 2D převod



kyselina benzoová



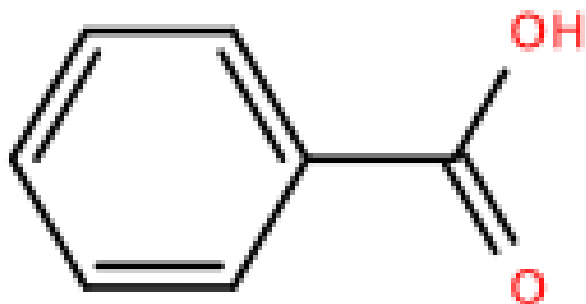
2D struktura obsahuje informaci o atomech a vazbách, kterými jsou spojeny. Tato informace popisuje konstituci (topologii) systému.

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

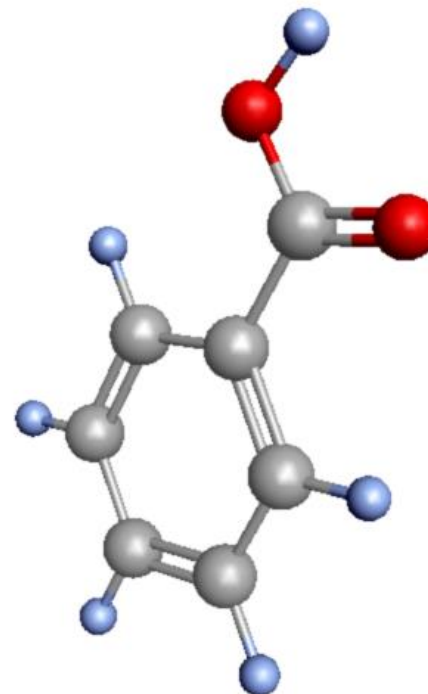


převod je snadný

2D -> 3D převod



kyselina benzoová



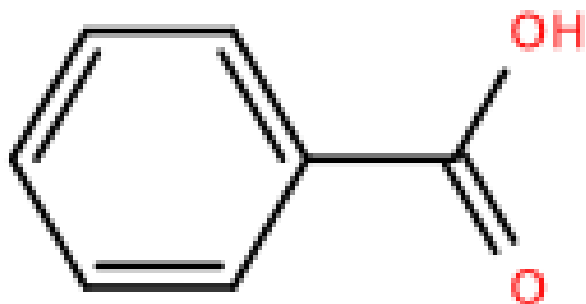
2D struktura obsahuje informaci o atomech a vazbách, kterými jsou spojeny. Tato informace popisuje konstituci (topologii) systému.

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

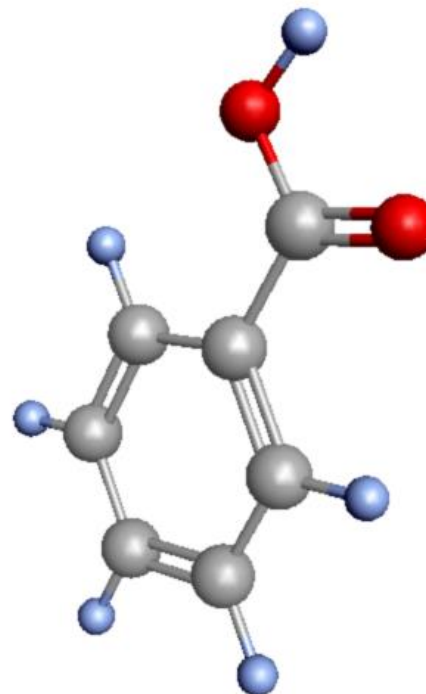
převod je komplikovaný



2D -> 3D převod



kyselina benzoová



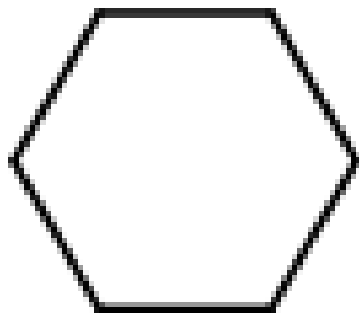
2D struktura obsahuje informaci o atomech a vazbách, kterými jsou spojeny. Tato informace popisuje konstituci (topologii) systému.

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

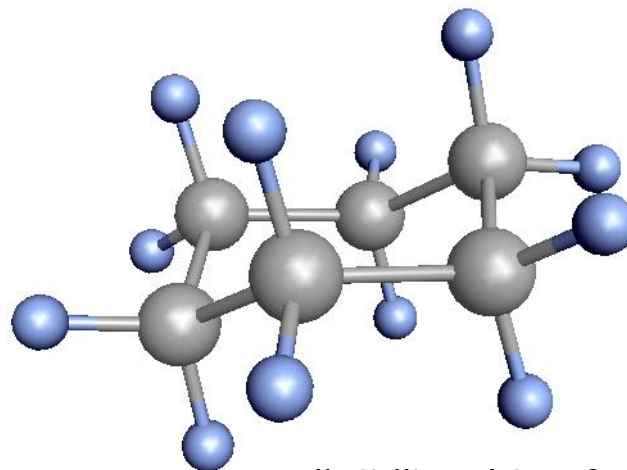
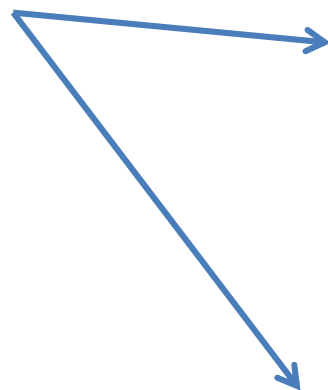


převod je komplikovaný
u velkých systémů nemusí být jednoznačný v důsledku existence více konformerů

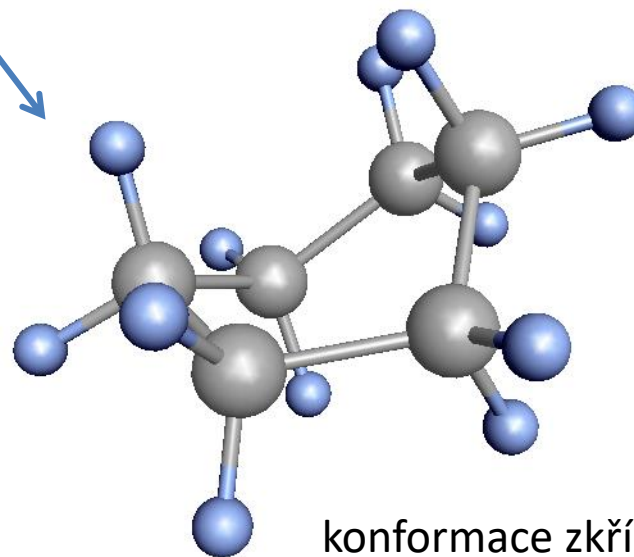
2D -> 3D převod, komplikace



cyklohexan



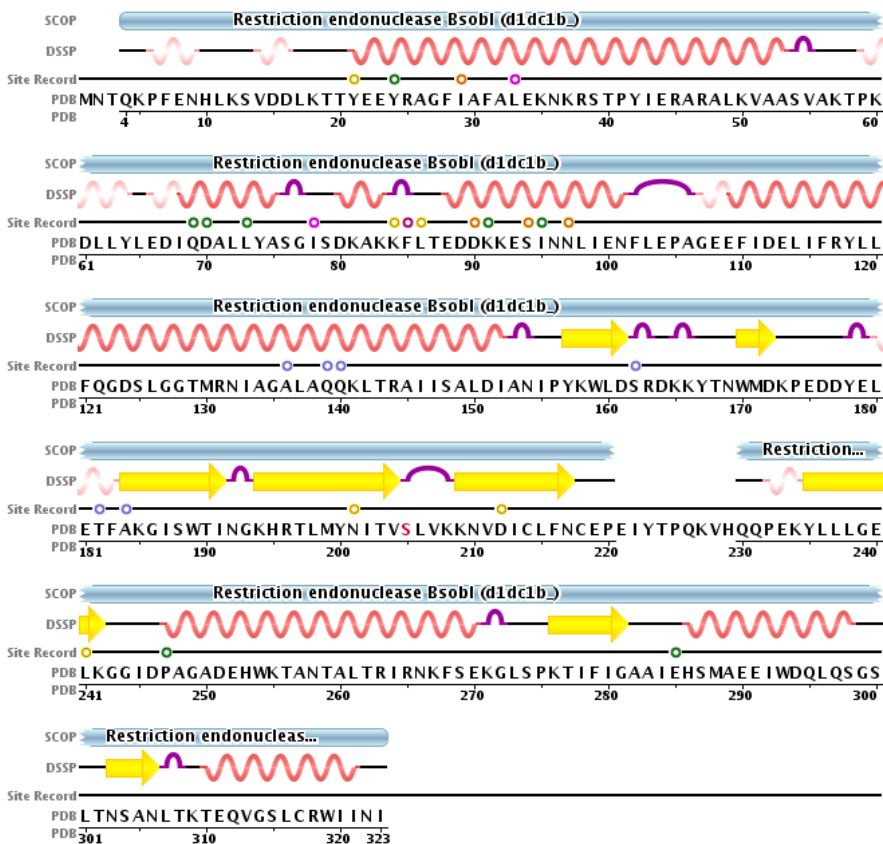
židličková konformace



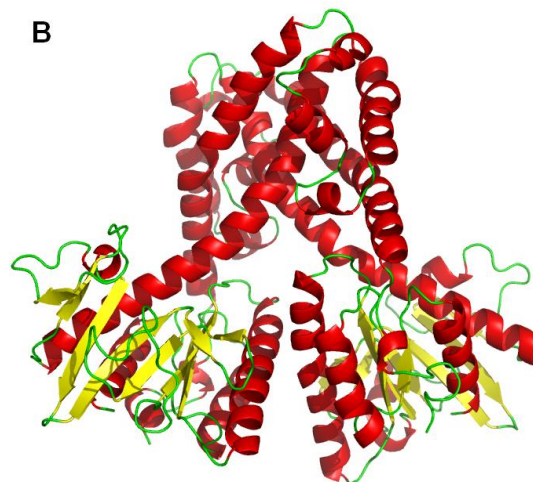
konformace zkřížená vanička

2D -> 3D převod, komplikace

Stejná primární struktura
(sekvence aminokyselin).

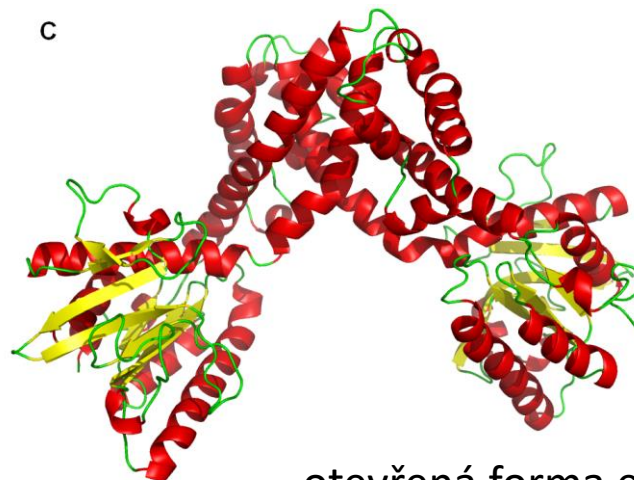


B



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.

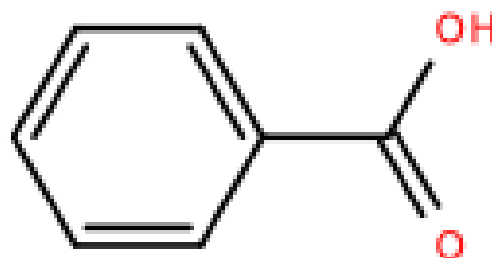
Nejrozšířenější formáty:

- **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 - experiment

Cambridge Structural Database (CSD)

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

Obsahuje zhruba půl miliónu struktur malých molekul určených pomocí rentgenové a neutronové difrakce. Software pro práci s daty: Mercury
<http://www.ccdc.cam.ac.uk/Solutions/CSDSystem/Pages/Mercury.aspx>

Protein Data Bank (PDB)

<http://www.pdb.org>

Obsahuje zhruba 94 tisíc struktur biomolekulárních systémů určených převážně pomocí rentgenostrukturní analýzy.

Experimentální metoda	Proteiny (P)	Nucleové kyseliny (NA)	P/NA komplexy	Jiné	Celkově
X-ray	77445	1481	4069	3	82998
NMR	8851	1046	193	7	10097
elektronová mikroskopie	469	45	129	0	643

stav v září 2013

Zdroje 3D struktur – *in silico*

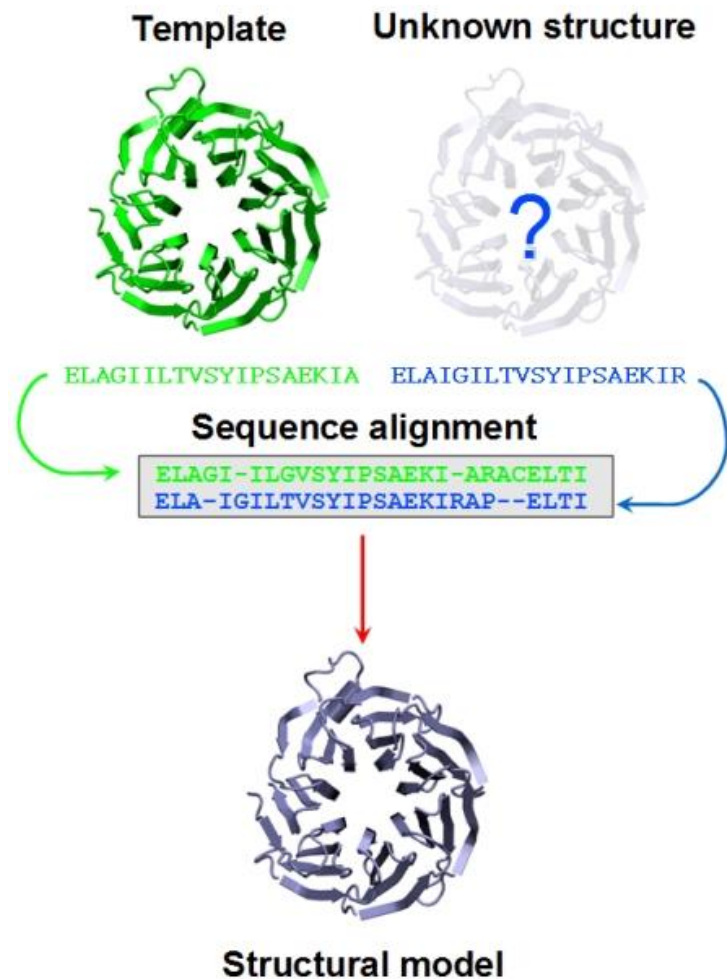
Výpočetní metody

- molekulové modelování
- homologní modelování

In silico modelování se provádí i u experimentálních struktur, které jsou neúplné:

- nedostatečné rozlišení – X-Ray
- jsou určeny jen částečné strukturní informace - NMR

Homologní modelování



<http://www.unil.ch/pmf/en/home/menuinst/technologies/homology-modeling.html>