

C2150

Zpracování informací a vizualizace v chemii a biochemii

5. lekce (3D vizualizace I)

Petr Kulhánek

kulhanek@chemi.muni.cz

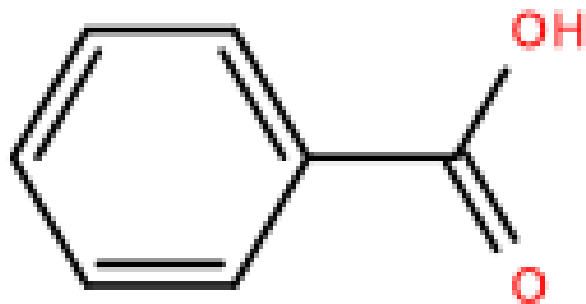
Národní centrum pro výzkum biomolekul, Přírodovědecká fakulta,
Masarykova univerzita, Kamenice 5, CZ-62500 Brno

Molekulární struktura

Molekulární struktura

Strukturní vzorec zobrazuje propojení jednotlivých atomů pomocí spojovacích čar (jednoduchých, zdvojených či ztrojených úseček, v případě potřeby i zakřivených čar), znázorňující vazby mezi atomy (jednoduché, dvojně nebo trojně vazby). Přitom délka spojovacích úseček a úhly mezi sousedními úsečkami **nevyjadřují ani skutečnou délku vazeb v molekule, ani úhly mezi vazbami**.

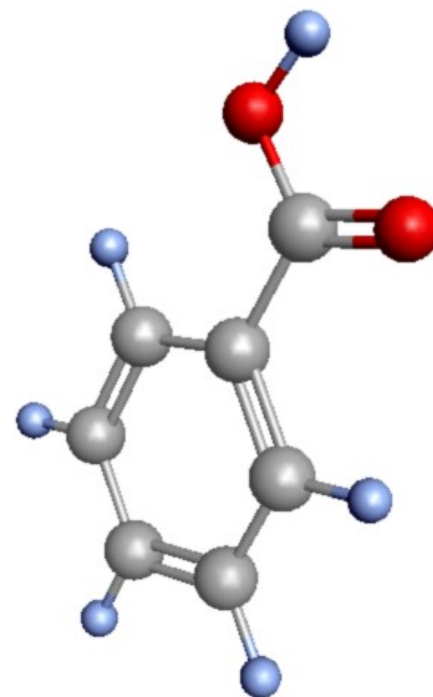
strukturní vzorec



kyselina benzoová

(3D) molekulární struktura

(3D) model struktury



(3D) Molekulární struktura popisuje prostorové uspořádání atomů. Struktura (poloha atomů) může být určena experimentálně nebo pomocí molekulového modelování.

(3D) Molekulární model je fyzikální model molekuly. Vytvářením modelů se zabývá **molekulové modelování** a jejich grafických znázorněním pak **molekulová grafika** (obě disciplíny spolu úzce souvisejí).

Geometrie molekuly

Konfigurační prostor:

$$\mathbf{R} = \{ \underbrace{x_1, y_1, z_1}_{\text{kartézské souřadnice prvního atomu}}, x_2, y_2, z_2, \dots, x_N, y_N, z_N \}_{\text{kartézské souřadnice posledního atomu}}$$

kartézské souřadnice
prvního atomu

kartézské souřadnice
posledního atomu

Jednotlivé hodnoty vektoru \mathbf{R} tvoří konfigurační prostor. **Každý bod** v konfiguračním prostoru pak představuje **unikátní strukturu** daného systému.

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

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).

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

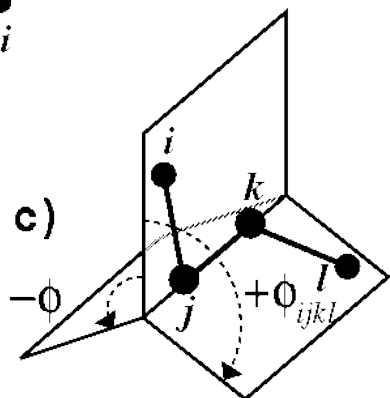
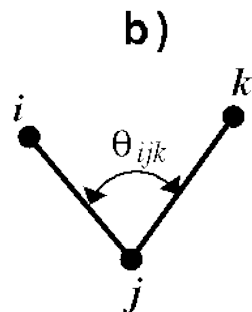
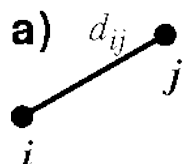
2-1
2
1

96.868054
96.868054

2

4-3-1-2

239.552651



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

Kde získat strukturu/model?

➤ Experimentální struktury

- X-ray nebo neutronová difrakce (poloha jednotlivých atomů)
- ostatní techniky poskytují pouze modely (např. NMR, CryoEM, SAXS, ...)

➤ In silico modely

- malé molekuly
- 2D -> 3D konverze (high-throughput modelování, virtuální screening)
- *ab initio* predikce biomolekulárních struktur (alphafold,...)

➤ Modelování vedené experimentálními daty

- NMR (NOE kontakty, ...)
- CryoEM, SAXS (elektronová hustota, tvar, ...)

➤ Modely vycházející z experimentálních struktur

- in silico změna experimentálních struktur
- homologní modelování (biomolekulární struktury)

Zdroje 3D struktur - experiment

Cambridge Structural Database (CSD)

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

Obsahuje struktury malých molekul určených pomocí rentgenové a neutronové difrakce.

Protein Data Bank (PDB)

<http://www.pdb.org>

Obsahuje struktury biomolekulárních systémů určených převážně pomocí rentgenové strukturní 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

Vizualizace (molekulová grafika)

https://en.wikipedia.org/wiki/Molecular_graphics

Vizualizace

vstupní geometrie
(volitelně konstituce)



umístění prvků modelu do 3D
(koule, tyčinky, čáry, povrchy)

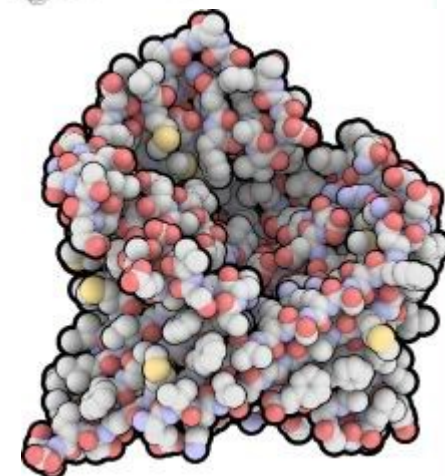
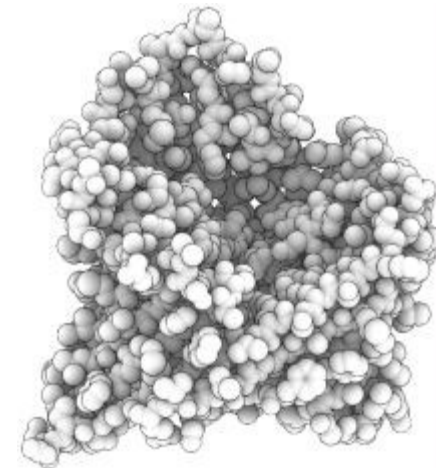
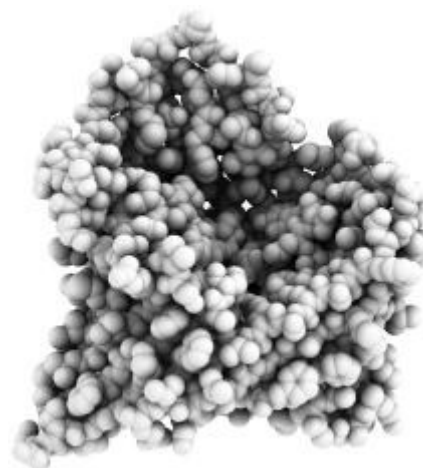


projekce do 2D
(zahrnuje natočení, posunutí, a
škálování každého bodu modelu)



zobrazení v 2D
(Z-buffer, double buffer)

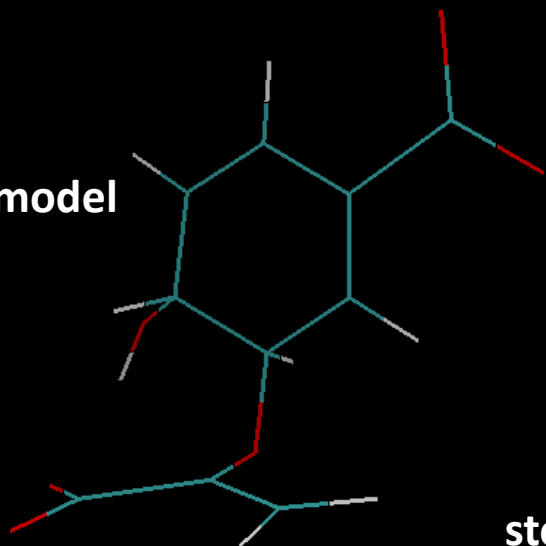
speciální techniky zobrazování pro
zdokonalení vnímání 3D modelu



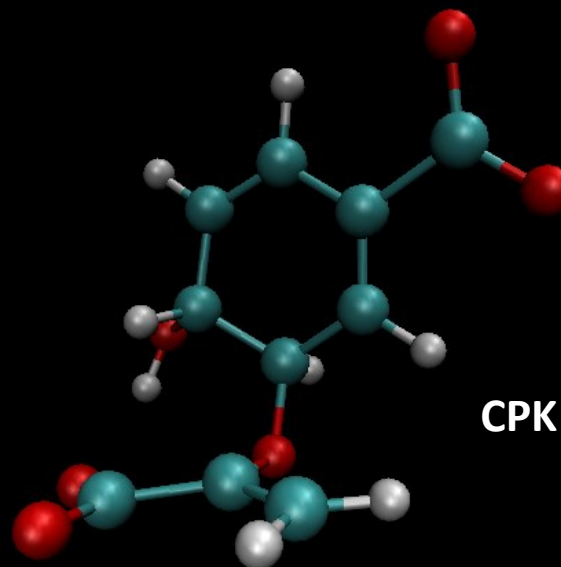
<http://qutemol.sourceforge.net>

Vizualizace – malé molekuly

čárový model

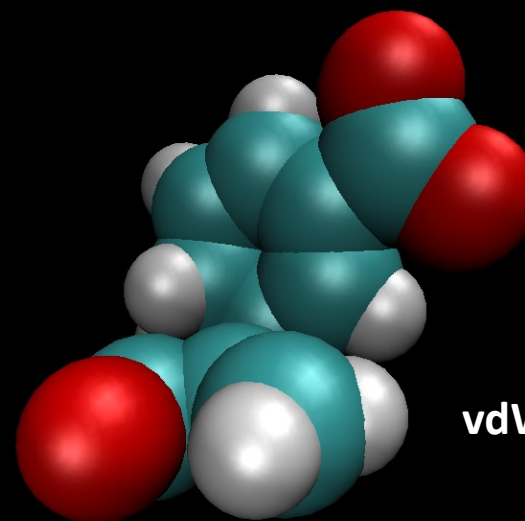
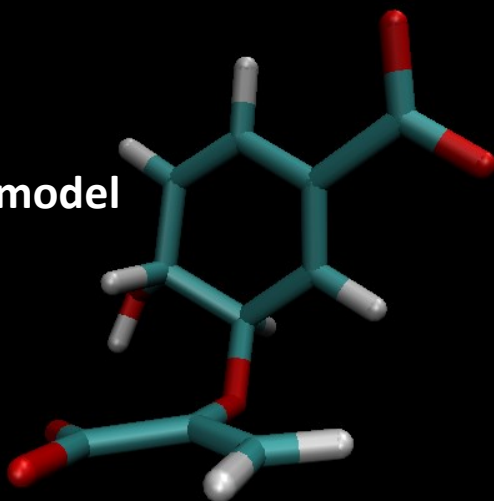


stejná struktura
jiná vizualizace



CPK model

tyčinkový model



vdW model

Cvičení

In silico modelování

Program Avogadro

Spuštění programu Avogadro

- Otevřít terminál, aktivovat modul, a spustit program

```
$ module add avogadro
```

```
$ avogadro
```



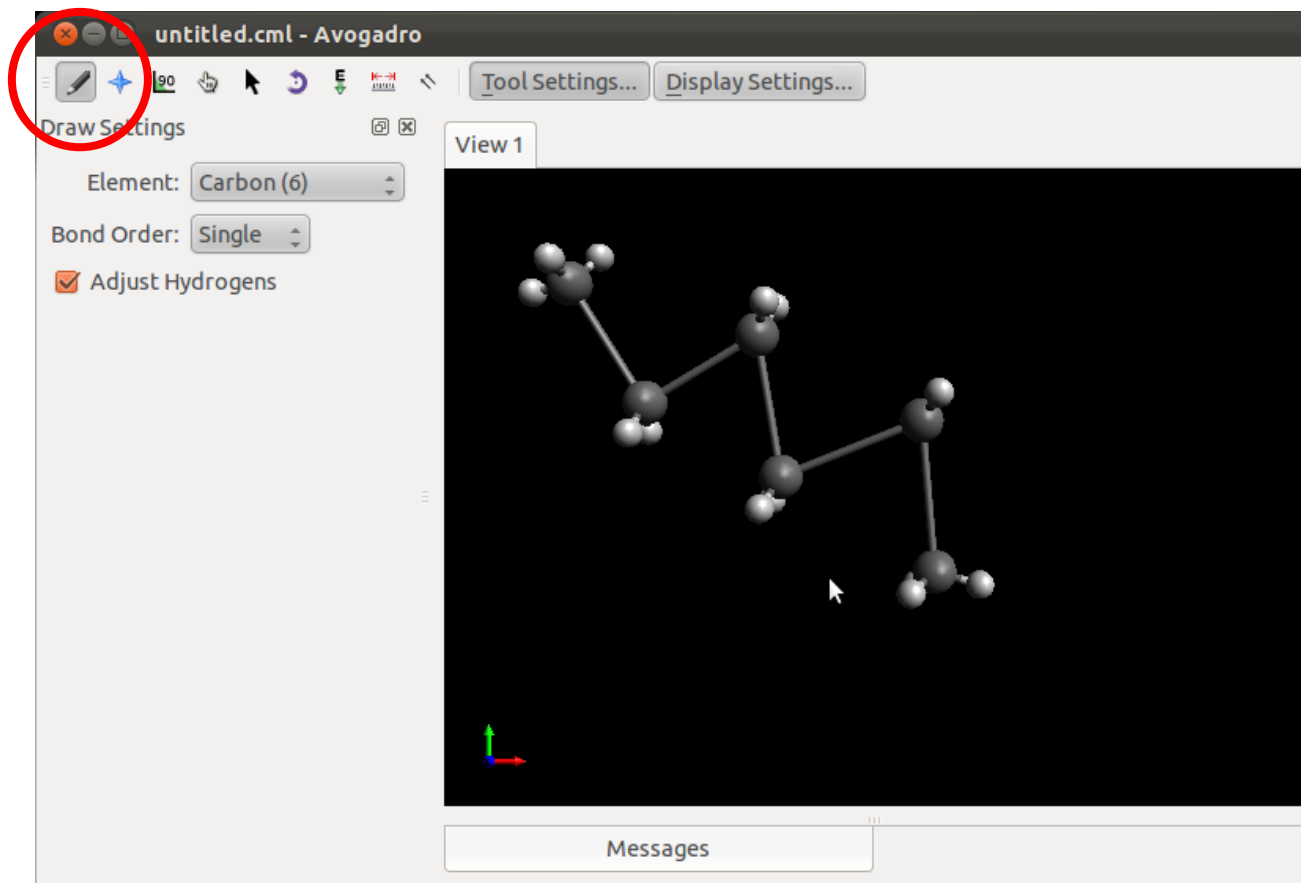
indikuje zápis do příkazové řádky (znak se nepíše)

- Novější verzi je možné spustit z nabídky standardních aplikací (nedoporučuji)
- nebo z terminálu spuštěním příkazu

```
$ avogadro2
```

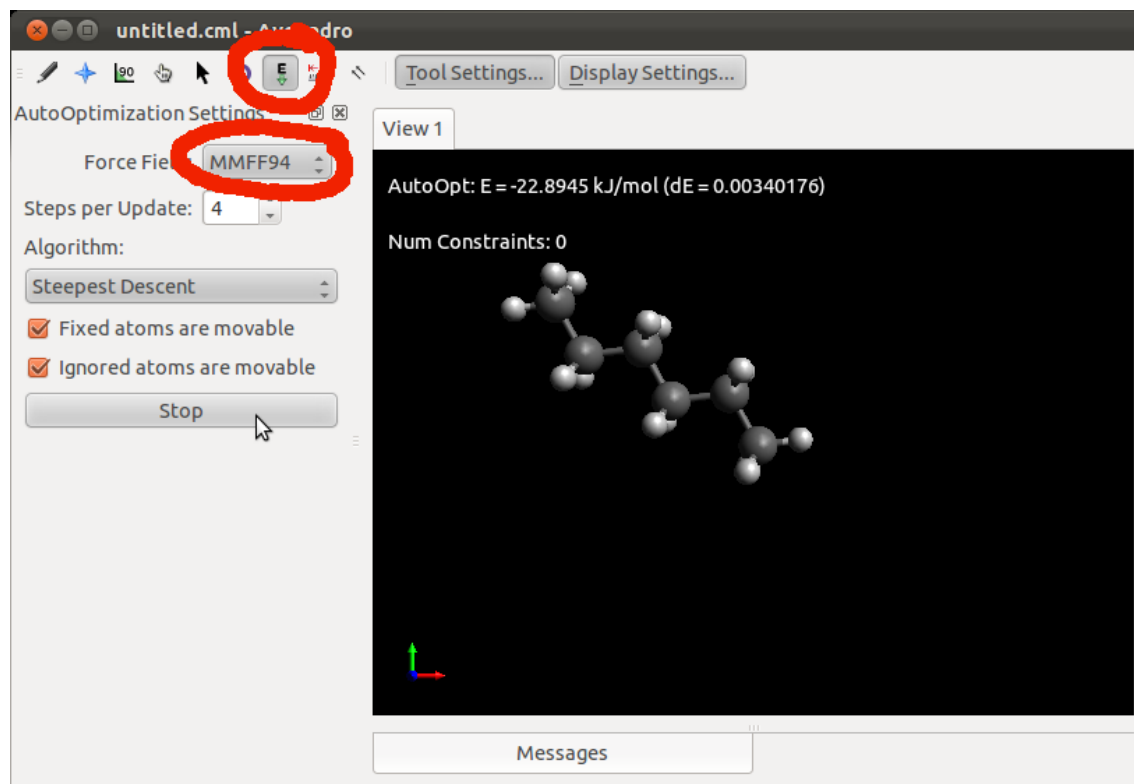
Draft modelu

- V režimu úprav navrhnete model struktury.
- Tento model nemá správné vzdálenosti mezi atomy a valenční úhly. Před dalším použitím je třeba jej optimalizovat.



Optimalizace modelu

- Avogadro používá k optimalizaci modelů molekulovou mechaniku (silové pole).
- Pro správný výsledek je nutné, aby výchozí model obsahoval správnou topologii (konstituci), tj. správně uvedené vazby včetně jejich vazebného řádu.
- Metody silového pole jsou empirické. Proto je nutné zvolit správnou parametrizaci, například MMFF94.



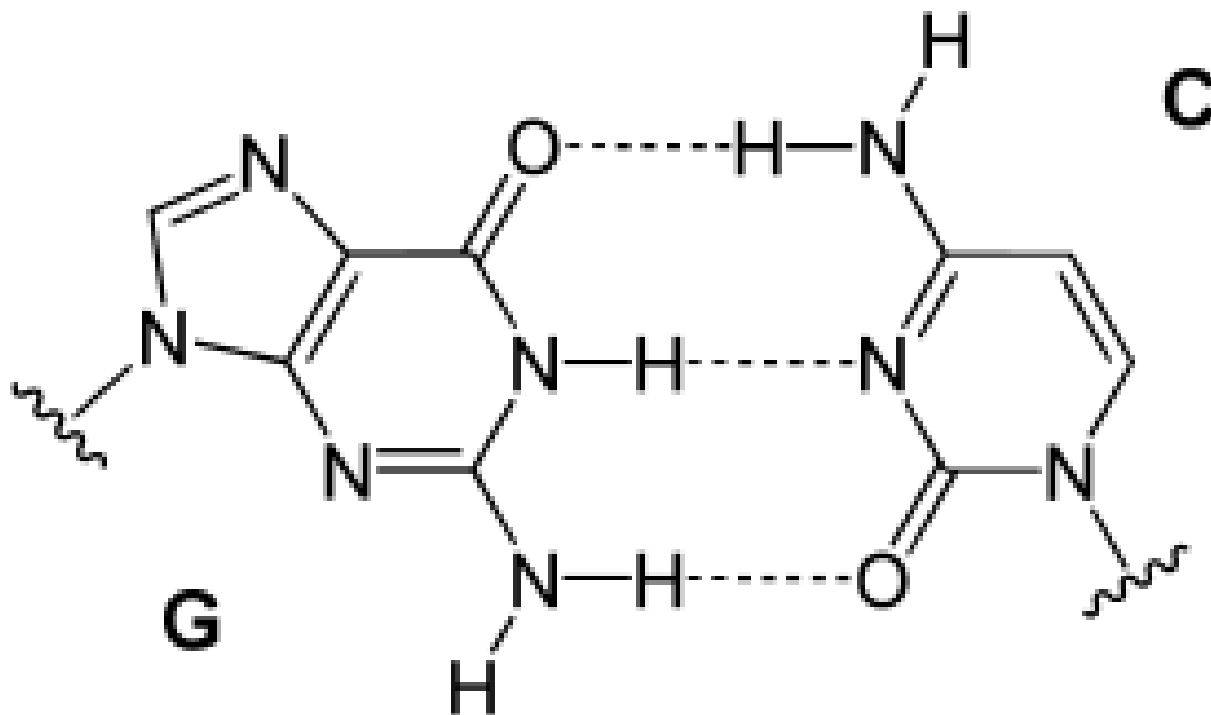
Cvičení 1

1. Vytvořte modely následujících molekul:

- methan
- ethen, ethyn
- benzen
- adamantan
- kyselina benzoová
- trinitrotoluen
- kyselina salicylová
- volitelně molekula C_{60}

Cvičení 2

1. Vytvořte model obsahující bázevý pár G:C podle níže uvedeného schématu. K nasycení volných valencí použijte atom vodíku. Pro optimalizaci geometrie použijte silové pole MMFF94.



https://en.wikipedia.org/wiki/Base_pair

In silico modelování

Program Nemesis

Spuštění programu Nemesis

- Otevřít terminál, aktivovat modul, a spustit program

```
$ module add nemesis
```

```
$ nemesis
```

Nemesis - Build Project

The screenshot shows the Nemesis Molecular Modelling Package interface. On the left, the 'Structures' panel lists 'Structure 1' with SID 1. Below it, the 'Active Profile' section shows 'Profile 1' and a list of 'Profile objects' including 'Light 1', 'Background 1', 'Standard Model 1', and 'Freezed Atoms 1'. On the right, the 'Build panel' contains various chemical symbols and buttons. The 'Optimize' button is circled in red. The 'Geometry' panel at the bottom right has tabs for 'Position', 'Distance', 'Angle', and 'Torsion'. Annotations with blue arrows point from text labels to specific parts of the interface: 'vrstvy' points to the 'Structure 1' entry; 'stavba/editace molekuly' points to the 'Build panel'; 'grafické modely' points to the 'Profile objects' list; 'měření geometrie' points to the 'Geometry' panel; and 'optimalizace modelu pomocí silového pole' points to the 'Optimize' button.

Project 1 : NEMESIS - Molecular Modelling Package

Structures

| Name | SID | Ato |
|-------------|-----|-----|
| Structure 1 | 1 | |

Number of structures: 1

Active Profile

Active profile: Profile 1

Profile objects

| Name | Type |
|------------------|---------|
| Light 1 | Light |
| Background 1 | Backgro |
| Standard Model 1 | Standar |
| Freezed Atoms 1 | Freezec |

Build panel

Basic General

Chemical symbols: C, C=, -C≡, =C=, Cl-, Br-, I-, O=, O-, S=, S-, N=, N=, N-, N=

Buttons: Delete atom, Make bond, Break bond, Delete bond, Optimize

Geometry

Position Distance Angle Torsion

Buttons: Restrain, Property, Label

Annotations:

- vrstvy
- stavba/editace molekuly
- grafické modely
- měření geometrie
- optimalizace modelu pomocí silového pole

Myš:

- levé tlačítko - selekce
- prostřední tlačítko - rotace
- pravé tlačítko - posunutí
- kolečko - zoom

Klávesy:

- Shift - XY -> Z
- Ctrl - primární/sekundární manipulátor

Nastavení silového pole: menu Geometry-> Optimizer Setup

Exercise 3

1. Vytvořte modely následujících molekul:

- metan
- ethen, ethyn
- benzen
- adamantan
- kyselina benzoová
- trinitrotoluen
- kyselina salicylová
- volitelně molekula C_{60}