

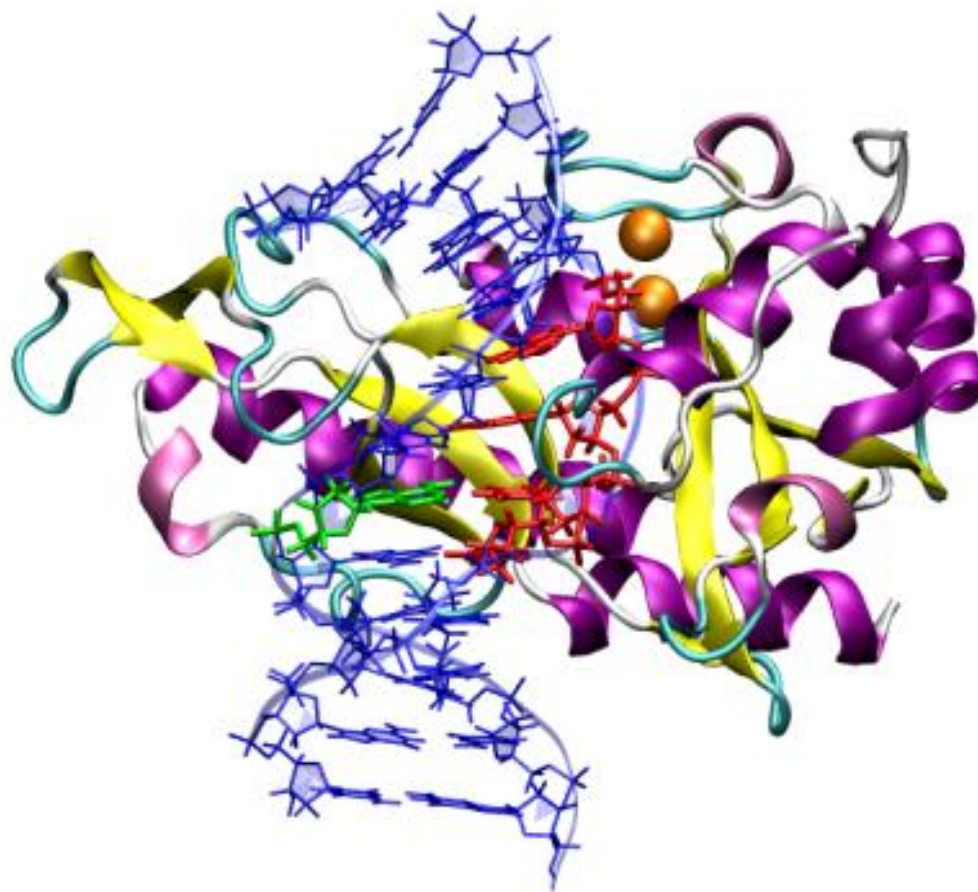
C7800

Počítačová chemie a molekulové modelování I - cvičení

Petr Kulhánek, Jakub Štěpán

kulhanek@chemi.muni.cz

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

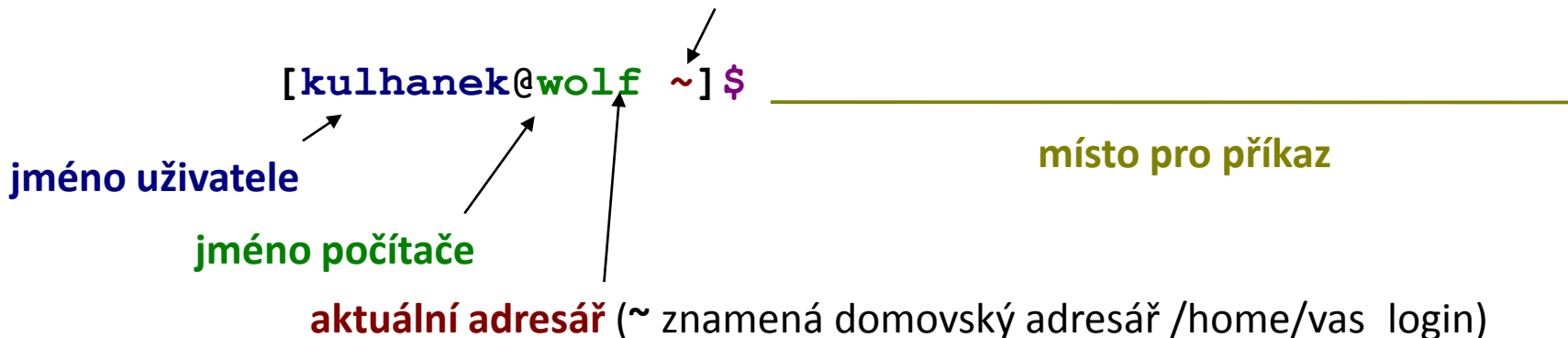


Ukázka v 3D

Opakování

Příkazová řádka

Prompt - typ uživatele / výzvy (\$ běžný uživatel, # super uživatel, další možné %, >)



Příkaz se vykoná zmáčknutím klávesy **Enter**.

Historie: pomocí kurzorových šipek nahoru a dolů lze procházet seznamem již zadaných příkazů. Příkaz z historie lze znovu použít nebo upravit a upravený použít. Historie je přístupná i příkazem **history**.

Automatické doplňování: zmáčknutím klávesy Tab (tabulátor) se interpret příkazové řádky snaží dokončit rozepsané slovo. Doplňují se jména příkazů, cesty a jména souborů (pokud jeden stisk nic nevyvolá, existuje více možností doplnění, opakovaný stisk vylistuje možnosti).

Kopírování textu: Ne pomocí Ctrl+C! Pro kopírování textu z terminálu stačí text označit, pro následné vložení stiskněte kolečko myši.

Terminály

Příkazová řádka je přístupná přímo z textových terminálů. V grafickém prostředí X11 je nutné spustit vhodnou aplikaci emulující textový terminál.

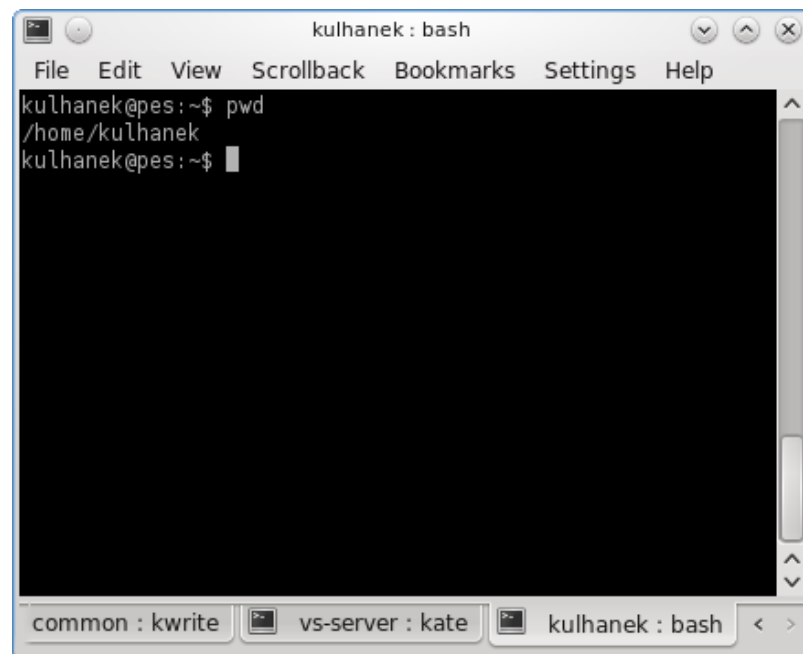
xterm



jednoduché, standard na všech UNIXových systémech

Výchozím adresářem je: **`/home/vas_login`**

konsole

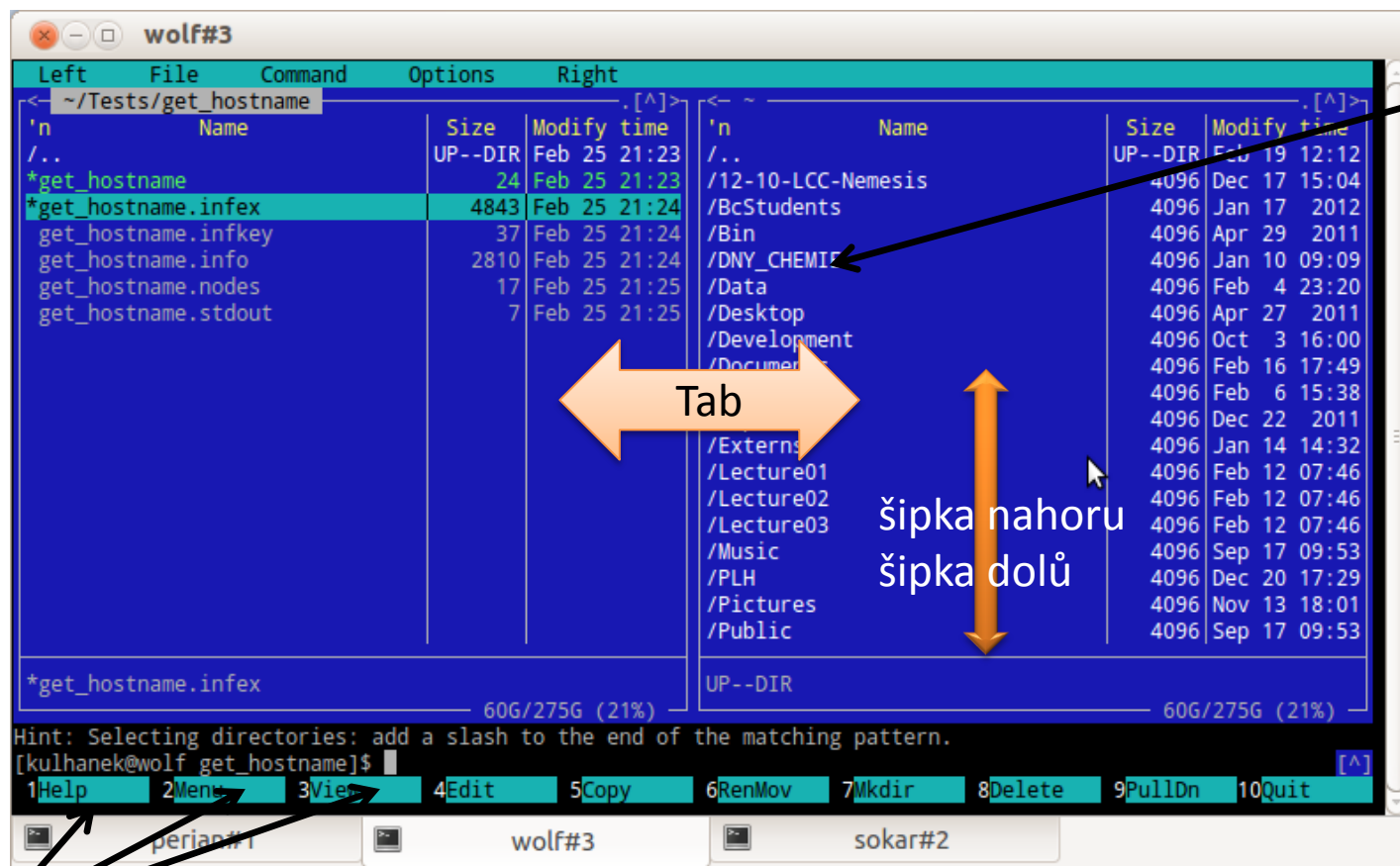


jednoduché přitom značně konfigurovatelné

Midnight Commander

Souborový manažer, který pracuje v terminálu.

<http://www.midnight-commander.org/>



označení více souborů - **Insert**

Tab

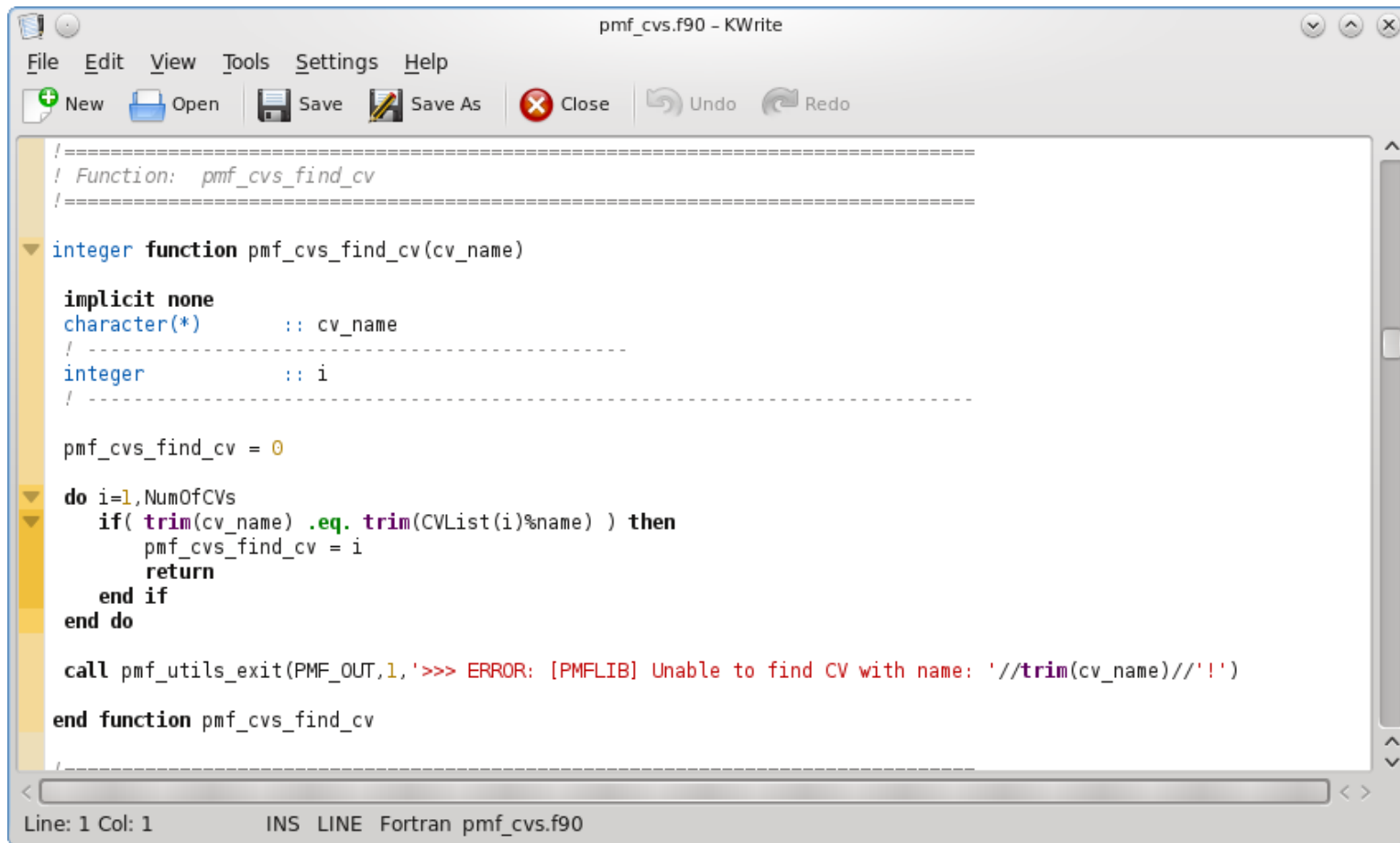
šipka nahoru
šipka dolů

funkční klávesy F1, F2, F3, ...

Skrytí obou panelů: Ctrl+O

Lze použít myš.

Textový editor - kwrite



```
pmf_cvs.f90 - KWrite
File Edit View Tools Settings Help
New Open Save Save As Close Undo Redo

!=====
! Function: pmf_cvs_find_cv
!=====
integer function pmf_cvs_find_cv(cv_name)

implicit none
character(*)      :: cv_name
!-----
integer           :: i
!-----

pmf_cvs_find_cv = 0

do i=1,NumOfCVs
  if( trim(cv_name) .eq. trim(CVList(i)%name) ) then
    pmf_cvs_find_cv = i
    return
  end if
end do

call pmf_utils_exit(PMF_OUT,1,'>>> ERROR: [PMFLIB] Unable to find CV with name: '//trim(cv_name)//'!')

end function pmf_cvs_find_cv

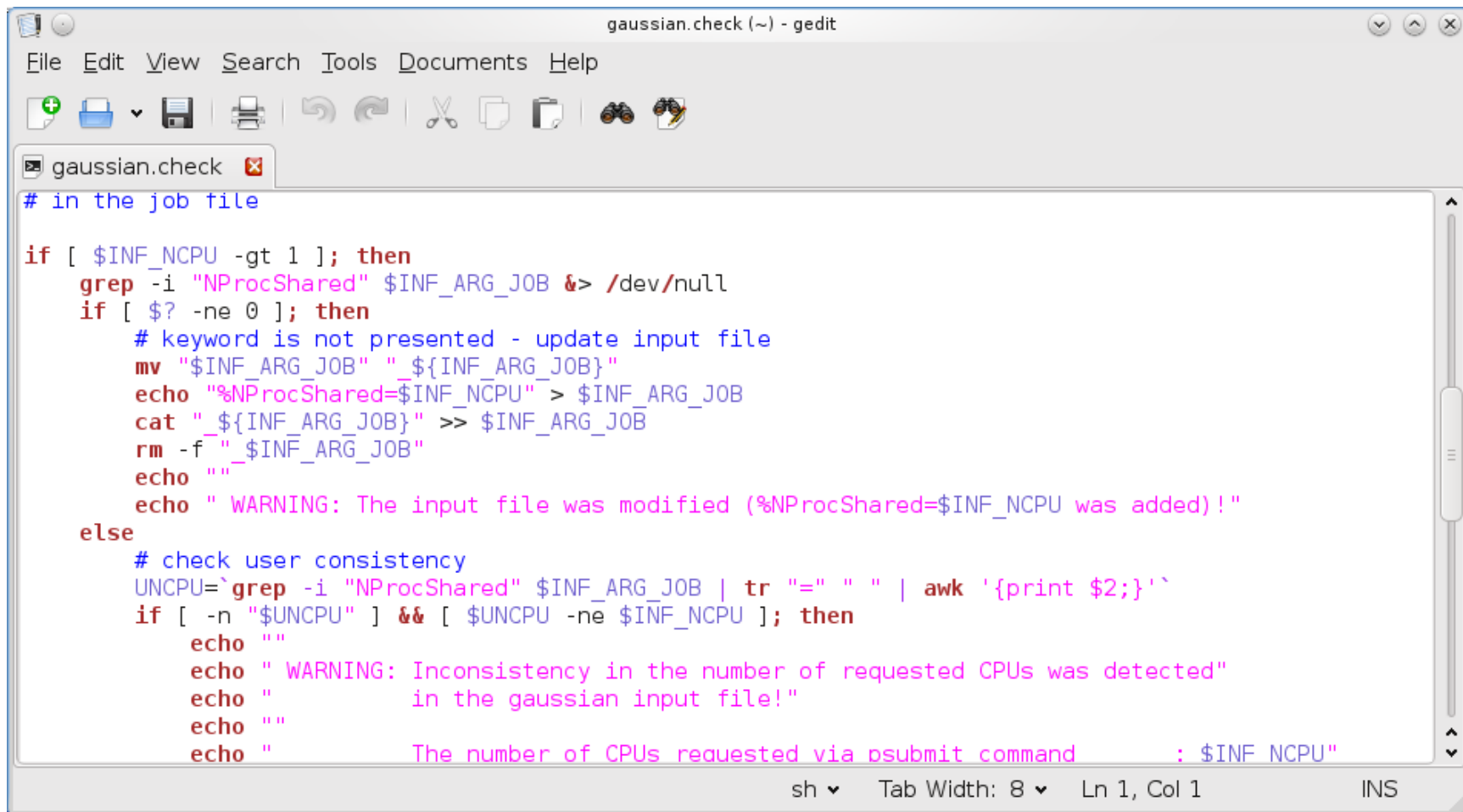
!-----

Line: 1 Col: 1      INS LINE Fortran pmf_cvs.f90
```

Spuštění:
\$ kwrite

Rozšířená funkcionlita: **kate**

Textový editor - gedit



The screenshot shows a window titled "gaussian.check (~) - gedit". The menu bar includes File, Edit, View, Search, Tools, Documents, and Help. The toolbar contains icons for new, print, save, undo, redo, cut, copy, paste, and zoom. The main text area contains the following shell script:

```
# in the job file

if [ $INF_NCPU -gt 1 ]; then
  grep -i "NProcShared" $INF_ARG_JOB &> /dev/null
  if [ $? -ne 0 ]; then
    # keyword is not presented - update input file
    mv "$INF_ARG_JOB" "_${INF_ARG_JOB}"
    echo "%NProcShared=$INF_NCPU" > $INF_ARG_JOB
    cat "_${INF_ARG_JOB}" >> $INF_ARG_JOB
    rm -f "_${INF_ARG_JOB}"
    echo ""
    echo " WARNING: The input file was modified (%NProcShared=$INF_NCPU was added)!"
  else
    # check user consistency
    UNCPU=`grep -i "NProcShared" $INF_ARG_JOB | tr "=" " " | awk '{print $2;}'`
    if [ -n "$UNCPU" ] && [ $UNCPU -ne $INF_NCPU ]; then
      echo ""
      echo " WARNING: Inconsistency in the number of requested CPUs was detected"
      echo "           in the gaussian input file!"
      echo ""
      echo "           The number of CPUs requested via psubmit command           : $INF_NCPU"
    fi
  fi
fi
```

The status bar at the bottom shows "sh", "Tab Width: 8", "Ln 1, Col 1", and "INS".

Spuštění:

\$ gedit

Programy pro molekulové modelování I

Přehled

Nemesis

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

Program pro stavbu a vizualizaci molekul. Alfa verze pro Linux. Testovací verze pro MS Windows na vyžádání.

Avogadro

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

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

Přehled funkcionality: <https://www.youtube.com/watch?v=xdmLoBILmqs>

VMD

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

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

Molekel

<http://molekel.cscs.ch/wiki/pmwiki.php>

Program pro vizualizaci molekulárních struktur.

Spouštění programů

Nemesis

```
$ module add nemesis  
$ nemesis
```

Avogadro

```
$ module add avogadro  
$ avogadro
```

VMD

```
$ module add vmd  
$ vmd
```

Molekel

```
$ module add molekel  
$ molekel
```

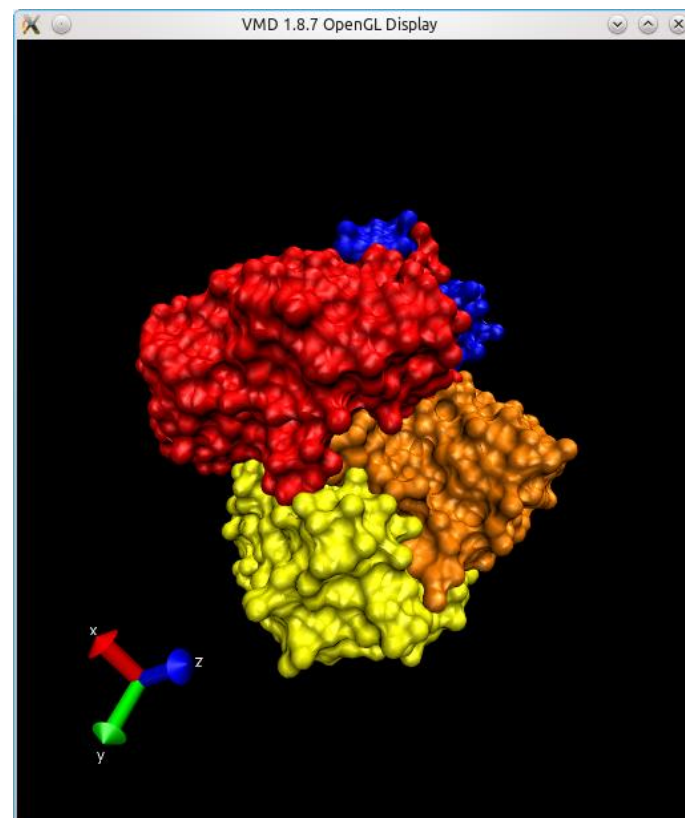
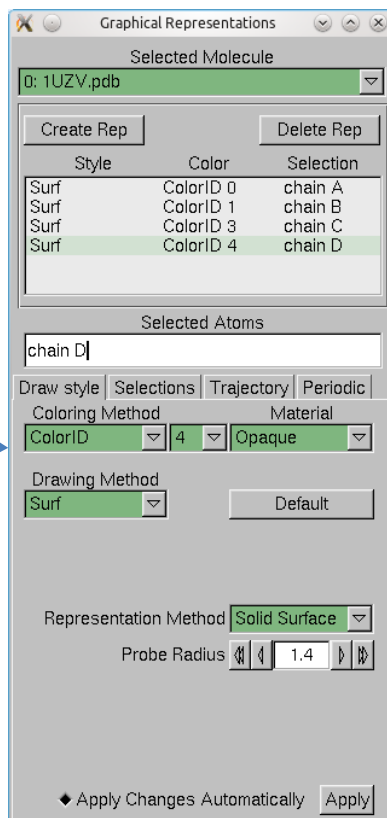
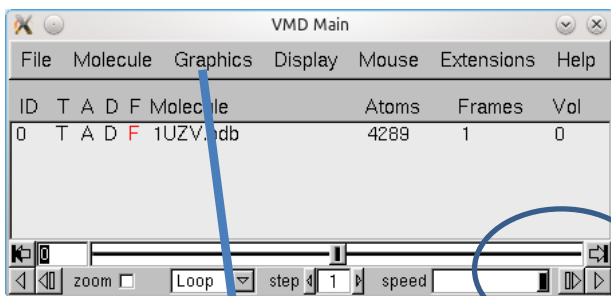


znak \$ se nezadává, značí příkazovou řádku

Program VMD

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

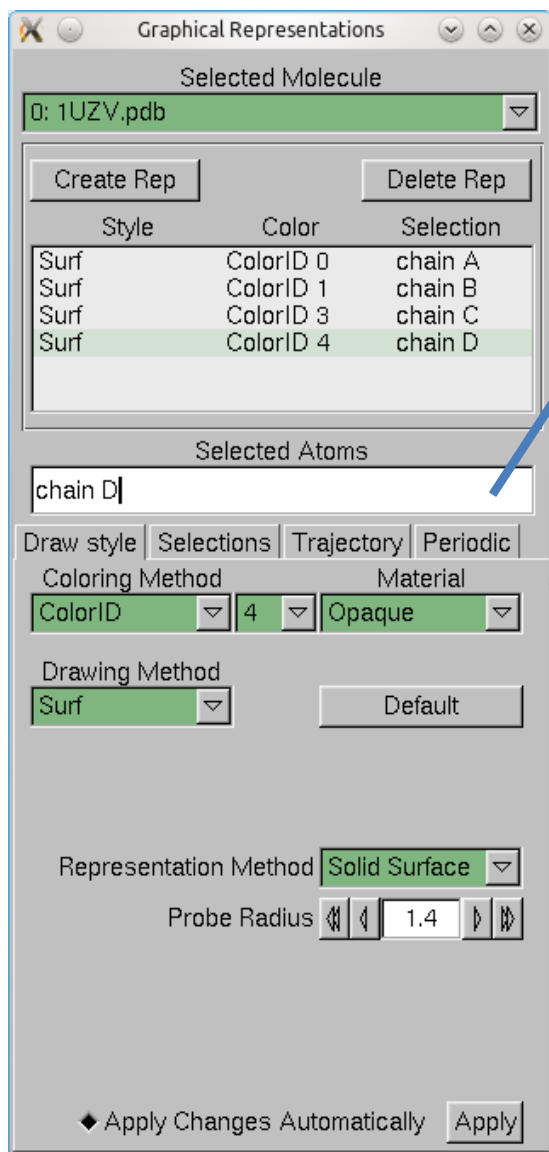
Program slouží k vizualizaci (bio)molekul a k analýze výsledků molekulárně dynamických simulací. Program je volně dostupný (vyžaduje registraci) a je dostupný i pro operační systém MS Windows.



Representation

posouvání mezi snímky

Program VMD



Selekcce (volba) části molekuly:

- protein – zvolí všechny aminokyseliny
- water – zvolí všechny molekuly vody
- chain X – zvolí řetězec X
- resname X – zvolí residuum s názvem X
- resid X – zvolí residuum s číslem X

Příklady:

chain A
chain A B C
resname ASP GLU
resid 1
resid 1 to 100

Bližší informace:

C2150 Zpracování informací a vizualizace v chemii

PDB Databáze

www.pdb.org

The screenshot displays the PDB website interface in a Mozilla Firefox browser window. The address bar shows the URL: <http://www.pdb.org/pdb/explore.do?structureId=1UZV>. The page title is "RCSB PDB - 1UZV Structure Summary".

The main content area features the structure ID "1UZV" and the title "HIGH AFFINITY FUCOSE BINDING OF PSEUDOMONAS AERUGINOSA LECTIN II: 1.0 A CRYSTAL STRUCTURE OF THE COMPLEX". The DOI is listed as 10.2210/pdb1uzv/pdb.

On the left side, there is a navigation menu with sections for "Home" and "Deposition". The "Home" section includes links for News & Publications, Usage/Reference Policies, Deposition Policies, Website FAQ, Deposition FAQ, Contact Us, About Us, Careers, External Links, Sitemap, and New Website Features. The "Deposition" section includes links for All Deposit Services, Electron Microscopy, X-ray | NMR, Validation Server, BioSync Beamlines/Facilities, and Related Tools.

The "Primary Citation" section provides the following information:
High affinity fucose binding of Pseudomonas aeruginosa lectin PA-III: 1.0 A resolution crystal structure of the complex combined with thermodynamics and computational chemistry approaches.
Mitchell, E., Sabin, C.D., Snajdrova, L., Pokorna, M., Gautier, S., Perret, C., Hofr, C., Gilboa-Garber, N., Koca, J., Imberty, M., Wimmerovaa.
Journal: (2005) Proteins: Struct., Funct., Bioinf. 58: 735
PubMed: 15573375
DOI: 10.1002/prot.20330
Search Related Articles in PubMed

On the right side, there is a "Biological Assembly" section showing a 3D ribbon diagram of the protein structure, colored in green, yellow, and blue, with a fucose molecule bound to it.

Obsahuje struktury biomolekul určené metodami rentgenové strukturní analýzy, nukleární magnetické rezonance, teoretické modely.

Cvičení VMD

Úkol I

- Najděte protein s PDB kódem **1UZV** . Kterou metodou byla struktura určena? Jakého rozlišení bylo dosaženo? Uložte strukturu ve formátu **pdb** do vašeho domovského adresáře.
- Zobrazte strukturu **1UZV** v programu **VMD**.
- Zvýrazněte jednotlivé monomerní jednotky komplexu (model: surf, selekce: chain A, chain B, ...)
- Zobrazte sekundární strukturu komplexu (NewCartoon, Color: Secondary Structure). Který strukturní element ve struktuře převládá?
- Zobrazte navázaný ligand (rename FUC). Kolik ligandů je v komplexu obsaženo?
- Zobrazte vápenaté ionty (rename CA). Kolik iontů je v komplexu obsaženo?

Úkol II

- Najděte protein s PDB kódem **2KRC**. Kterou metodou byla struktura určena? Uložte strukturu ve formátu **pdb** do vašeho domovského adresáře.
- Zobrazte strukturu **2KRC** v programu **VMD**.
- Zobrazte sekundární strukturu komplexu (NewCartoon, Color: Secondary Structure). Který strukturní element ve struktuře převládá?
- Kolik modelů bylo publikováno? Kvalitativně určete jak se od sebe modely liší.

Cvičení XYZ formát

1. V textovém editoru vytvořte soubor popisující model vody s následujícími parametry. Délka vazeb O-H bude 1 Å. Vazebný úhel H-O-H bude 90°. Uložte jej do domovského adresáře jako **water.xyz**
2. Vytvořený soubor načtěte do programu VMD.
3. Ověřte skutečnou délku vazeb a velikost úhlu H-O-H. (VMD Main >Mouse->Label, správa popisků v VMD Main >Graphics>Labels)
4. Molekulu vody zobrazte v následujících modelech: Lines, CPK, Licorice, VDW.

MS Windows jako klient

MS Windows jako klient - přehled

Přihlašování do Unixu z MS Windows (textový terminál):

putty (<http://www.chiark.greenend.org.uk/~sgtatham/putty/>)
ssh (např. z prostředí Cygwin; <http://www.cygwin.com/>)

Kopírování dat mezi Unixem a MS Windows:

WinSCP (<http://winscp.net>)
scp (např. z prostředí Cygwin; <http://www.cygwin.com/>)

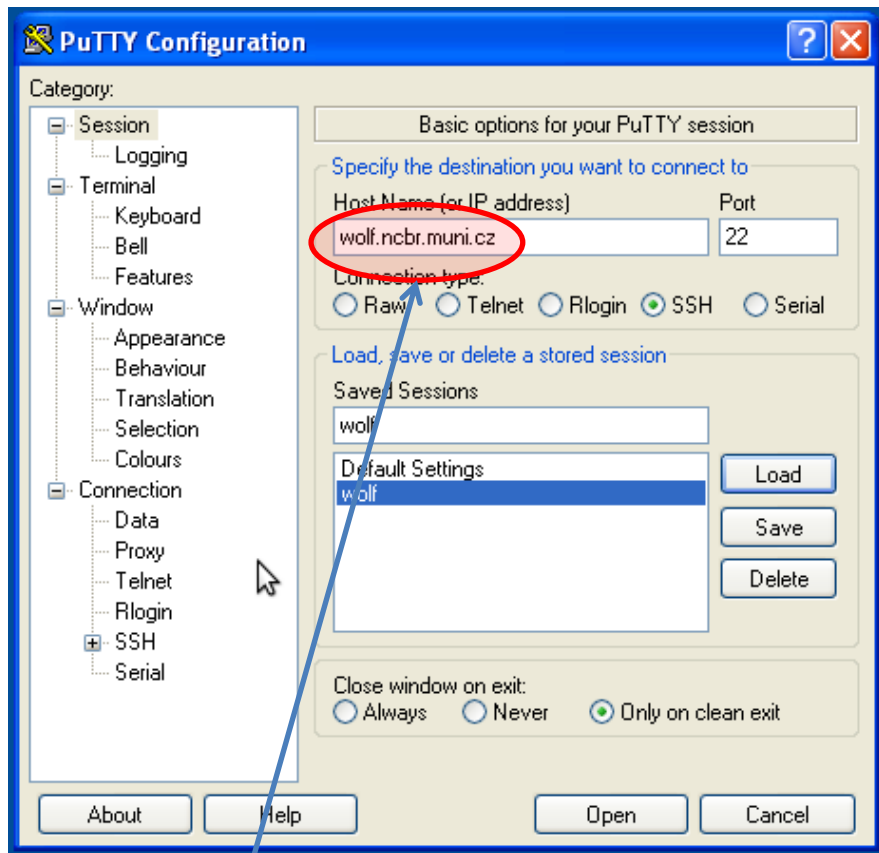
Export displeje z Unixu do MS Windows (X11 server):

Xming (<http://sourceforge.net/projects/xming/>)
cygwin (<http://www.cygwin.com/>)

Přihlašování z Unixu do MS Windows (vzdálená plocha):

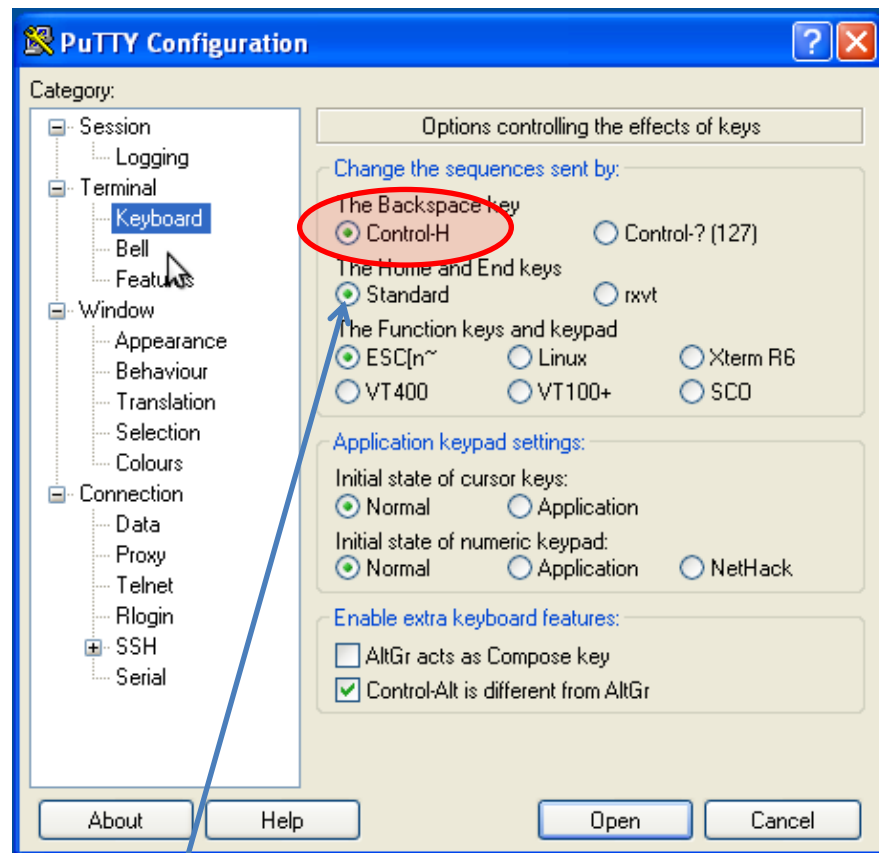
rdesktop

Putty – nastavení



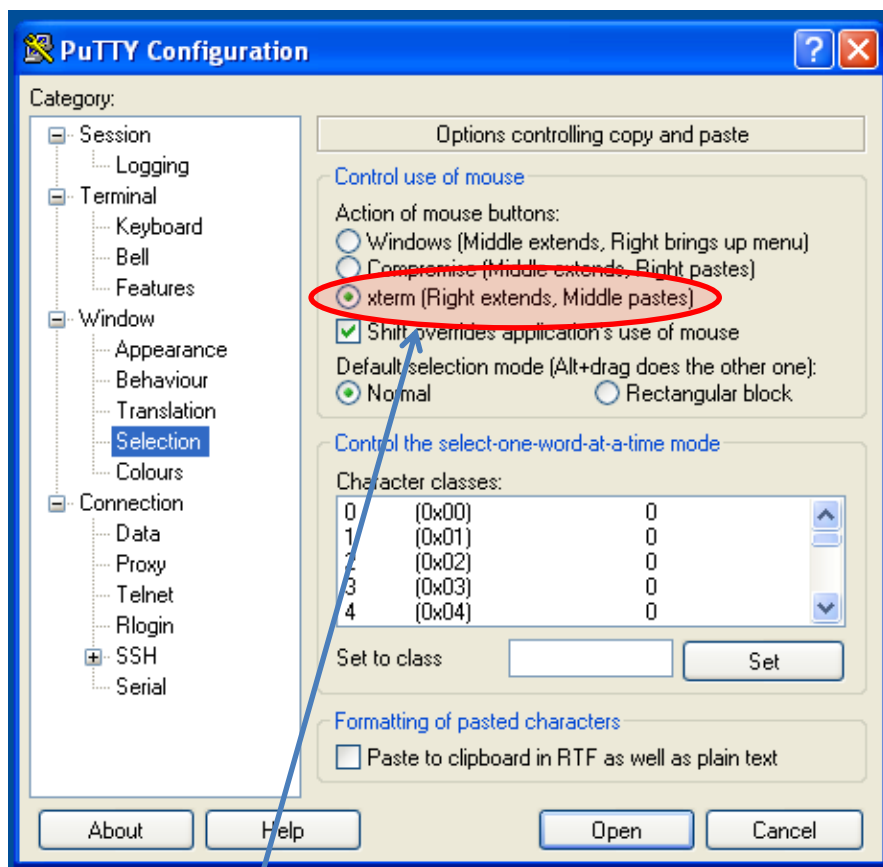
adresa vzdáleného stroje

wolf.ncbr.muni.cz

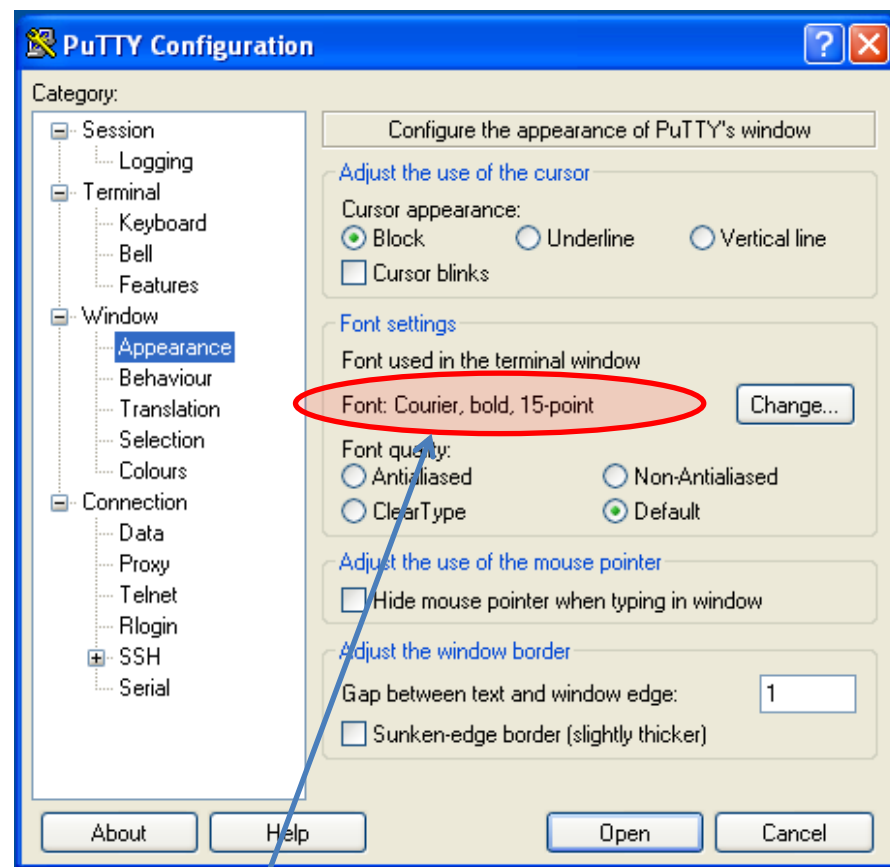


Správné fungování klávesy backspace.

Putty – nastavení II



selektce myši kompatibilní s Unixovými terminály

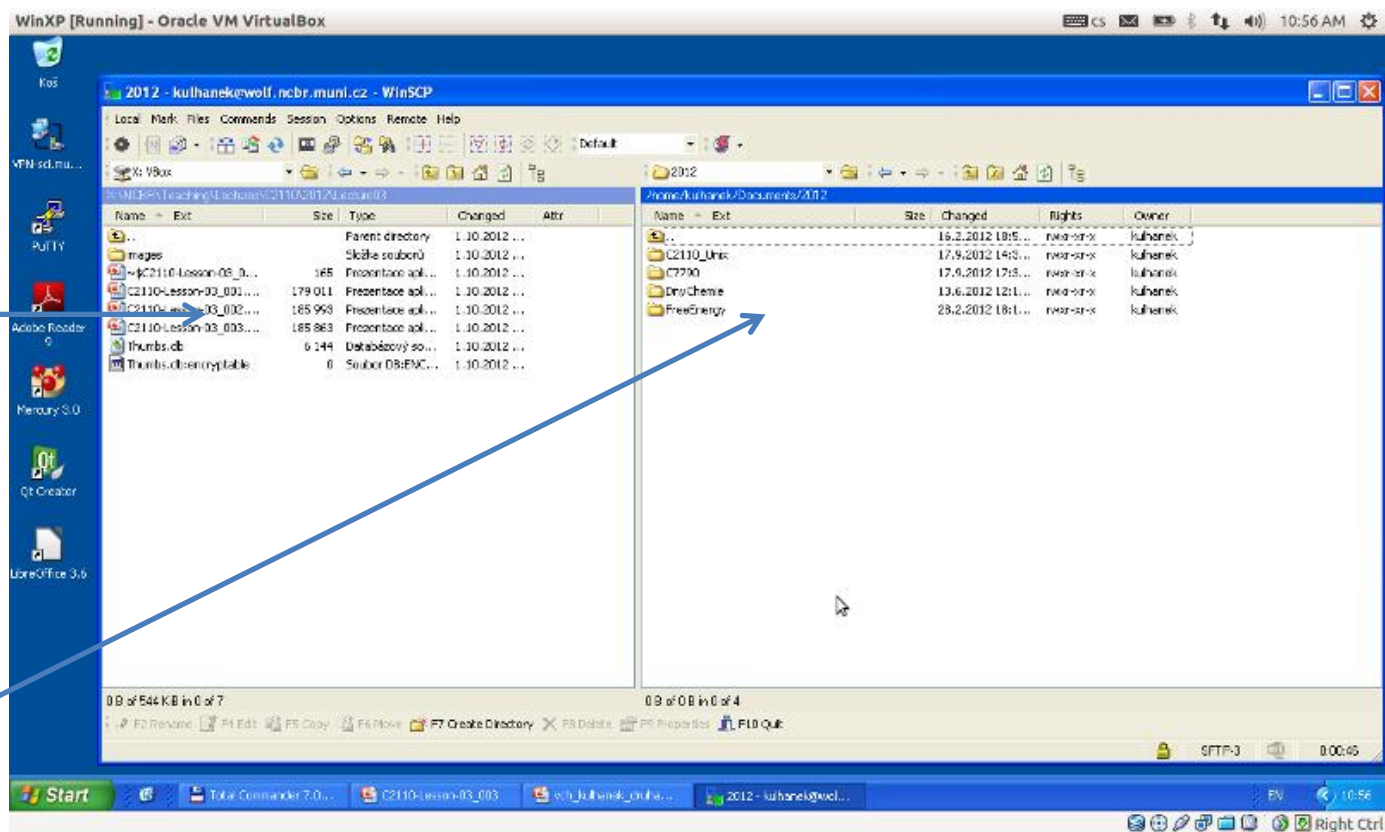


neproporcionální písmo
(všechny znaky mají stejnou šířku)

WinSCP

WinSCP <http://winscp.net/eng/docs/lang:cs>

Program pro přenos souborů mezi MS Windows a počítači podporující SFTP či SCP protokoly (převážně unixového a linuxového typu).



lokální stroj

vzdálený stroj

Textové soubory MS Win ↔ Linux

Textové soubory vytvořené pod MS Windows a Linuxem **nejsou** zcela **kompatibilní**, protože každý operační systém používá jiné kódování **konce řádku**.

Linux: \n (line feed 0x0A)

MS Windows: \r+\n (carriage return 0x0D, line feed 0x0A)

Ke konverzi souborů lze použít programy **d2u** a **u2d** (na klastru WOLF).

1) Aktivace modulu cats

```
$ module add cats
```

2) Konverze MS Windows => Linux

```
$ d2u soubor.com
```

3) Konverze Linux => MS Windows

```
$ u2d soubor.log
```

Cvičení

1. Spusťte virtuální stroj s MS Windows XP (**/win/win**).
2. Spusťte aplikaci Putty (Aplikaci stáhněte z internetu).
3. Pomocí terminálu Putty se přihlaste na stroj **wolf.ncbr.muni.cz**.
4. Monitorujte, kdo je na stroji wolf.ncbr.muni.cz přihlášen (příkazy **w** nebo **who**).
5. Zkuste v terminálu Putty spustit aplikaci **kwrite**. Proč spuštění aplikace selže?
6. Spusťte aplikaci WinSCP (Aplikaci stáhněte z internetu jako **Portable executables**).
7. Do virtuálního stroje stáhněte soubor **water.xyz**. Soubor otevřete v programu Poznámkový blok (Notepad). Zobrazí se obsah souboru správně?
8. Opravte kódování konců řádků v souboru **voda.xyz** (na straně Linuxu) a soubor znovu otevřete ve virtuálním stroji v programu Poznámkový blok.

Programy pro molekulové modelování II

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
```

Nápověda:

```
$ babel -H
```

 **velké H**

Cvičení

1. Aktivujte modul openbabel.
2. Vypište formáty, které instalovaná verze open babelu podporuje.
3. Zkonvertujte soubor **water.xyz** do formátu Sybyl Mol2 format a uložte jej pod názvem **water.mol2**
4. Otevřete soubor **water.mol2** v textovém editoru a diskutujte význam jeho částí.
5. Zkonvertujte soubor **water.xyz** do formátu InChI a ložte jej po názvem **water.txt**
6. Otevřete soubor **water.txt** v textovém editoru a diskutujte význam jeho částí.

Dávkové systémy

Dávkové zpracování

Dávkové zpracování je vykonávání série programů (tzv. dávek) na počítači bez účasti uživatele. Dávky jsou připraveny předem, takže mohou být zpracovány předány bez účasti uživatele. Všechna vstupní data jsou předem připravena v souborech (skriptech) nebo zadána pomocí parametrů na příkazovém řádku. Dávkové zpracování je opakem interaktivního zpracování, kdy uživatel až teprve za běhu programu poskytuje požadované vstupy.

Výhody dávkového zpracování

- sdílení zdrojů počítače mezi mnoha uživateli a programy
- odložení zpracování dávek do doby, kdy je počítač méně vytížen
- odstranění prodlev způsobeným čekáním na vstup od uživatele
- maximalizace využití počítače zlepšuje využití investic (zejména u dražších počítačů)

zdroj: www.wikipedia.cz, upraveno

Nástroje pro dávkové spouštění

➤ **OpenPBS**

<http://www.mcs.anl.gov/research/projects/openpbs/>

➤ **PBSPro**

<http://www.pbsworks.com>

➤ **Oracle Grid Engine**

<http://www.oracle.com/us/products/tools/oracle-grid-engine-075549.html>

➤ **Open Grid Scheduler**

<http://gridscheduler.sourceforge.net/>

open source

➤ **Torque**

<http://www.adaptivecomputing.com/products/open-source/torque/>

je použit jako dávkový systém v MetaCentrum VO, na klastrech SOKAR a WOLF

Infinity

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

Přehled příkazů

Správa software:

- site aktivace logických výpočetních zdrojů
- module aktivace/deaktivace software

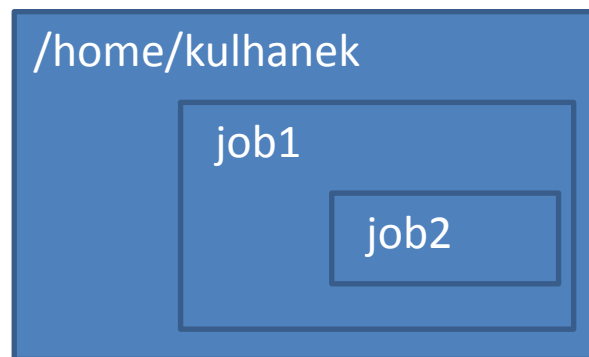
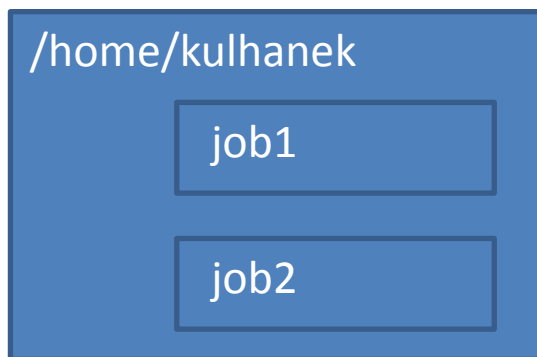
Správa úloh:

- pqueues přehled front z dávkového systému dostupných uživateli
- pnodes přehled výpočetních uzlů dostupných uživateli
- pqstat přehled všech úloh zadaných do dávkového systému
- pjobs přehled úloh uživatele zadaných do dávkového systému
- psubmit zadání úlohy do dávkového systému
- pinfo informace o úloze
- pgo přihlásí uživatele na výpočetní uzel, kde se úloha vykonává
- paliases definování aliasů

Úloha

Úloha **musí splňovat** následující podmínky:

- každá úloha se spouští v samostatném adresáři
- všechny vstupní data úlohy musí být v adresáři úlohy
- adresáře úloh nesmí být do sebe zanořené
- průběh úlohy je řízen skriptem nebo vstupním souborem (u automaticky detekovaných úloh)
- skript úlohy musí být v bashi
- ve skriptu úlohy se nesmí používat absolutní cesty, všechny cesty musí být uvedeny relativně k adresáři úlohy



Spuštění úlohy

Úlohu spouštíme **v adresáři úlohy** příkazem **psubmit**.

```
psubmit destination job [resources] [syncmode]
```

destination (kam) je buď:

- název_fronty
- název_uzlu@název_fronty

job je buď:

- název skriptu úlohy
- název vstupního souboru pro automaticky rozpoznávané úlohy

resources jsou požadované zdroje pro úlohu, pokud není uvedeno, požaduje se běh na 1 CPU

syncmode určuje způsob kopírování dat mezi adresářem úlohy a výpočetním uzlem, výchozím módem je "sync"

Monitorování běhu úlohy

K monitorování průběhu úlohy lze použít příkaz **pinfo**, který se spouští buď v adresáři úlohy nebo v pracovním adresáři na výpočetním uzlu. Dalšími možnostmi jsou příkazy **pjobs** a **pqstat**.

Pokud je úloha spuštěna na výpočetním uzlu, je možné použít příkaz **pgo**, který se naloguje na výpočetní uzel a změní aktuální adresář do pracovního adresáře úlohy.

Servisní soubory

V adresáři úlohy vznikají při zadání úlohy do dávkového systému a dále v průběhu života úlohy a po jejím ukončení servisní soubory. Jejich význam je následující:

- *.info kontrolní soubor s informacemi o průběhu úlohy
- *.infex vlastní skript (wrapper), který se spouští dávkovým systémem
- *.infout standardní výstup z běhu *.infex skriptu, **nutno analyzovat při nestandardním ukončení úlohy**
- *.nodes seznam uzlů vyhrazených pro úlohu
- *.gpus seznam GPU karet vyhrazených pro úlohu
- *.key unikátní identifikátor úlohy
- *.stdout **standardní výstup z běhu skriptu úlohy**

Cvičení I

1. Vytvořte textový soubor obsahující následující text. Soubor nazvěte **get_hostname** a uložte jej do samostatného adresáře **uloha01**.

```
#!/bin/bash
hostname
```

2. Nastavte souboru práva pro spuštění:

```
$ chmod u+x get_hostname
```

3. Vytvořený skript spusťte v terminálu. Co se stane?

```
$ ./get_hostname
```

4. Skript zadejte ke spuštění do dávkového systém. K zadání úlohy použijte frontu **short**.

```
$ psubmit short get_hostname
```

5. Průběh úlohy monitorujte příkazem **pinfo**.

6. Na kterém výpočetním uzlu se úloha spustila? Analyzujte obsah souboru **get_hostname.stdout** v textovém editoru.

Cvičení II

1. Vytvořte textový soubor obsahující následující text. Soubor nazvěte **long_job** a uložte jej do samostatného adresáře **uloha02**. Co dělá příkaz **sleep** (použijte manuálové stránky)?

```
#!/bin/bash
hostname
sleep 240
```

2. Skript zadejte ke spuštění do dávkového systém. K zadání úlohy použijte frontu **short**.

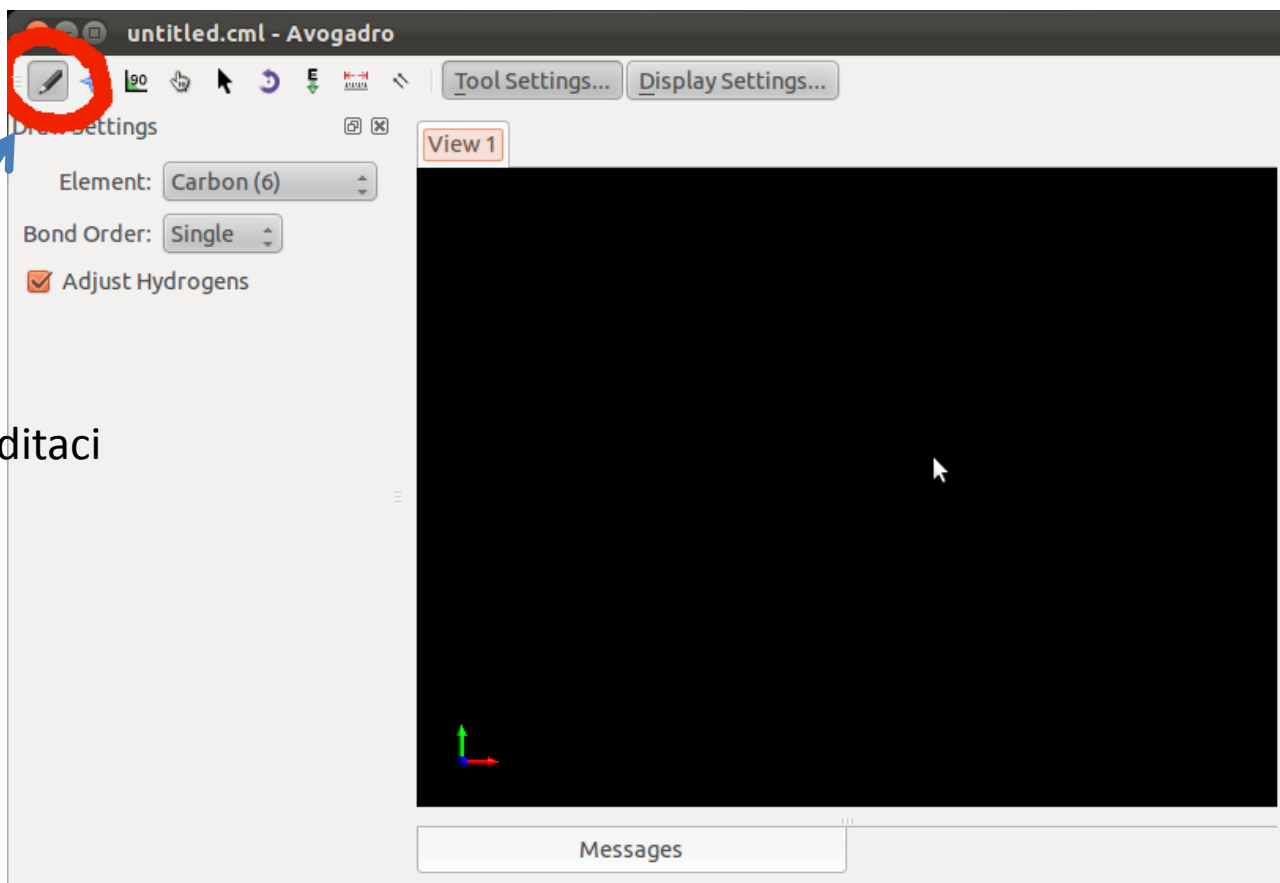
```
$ psubmit short long_job
```

3. Průběh úlohy monitorujte příkazem **pinfo**. Úlohy ostatních uživatelů příkazem **pqstat**. Obsazení jednotlivých výpočetních uzlů pak příkazem **pnodes**.

Programy pro molekulové modelování III

Avogadro

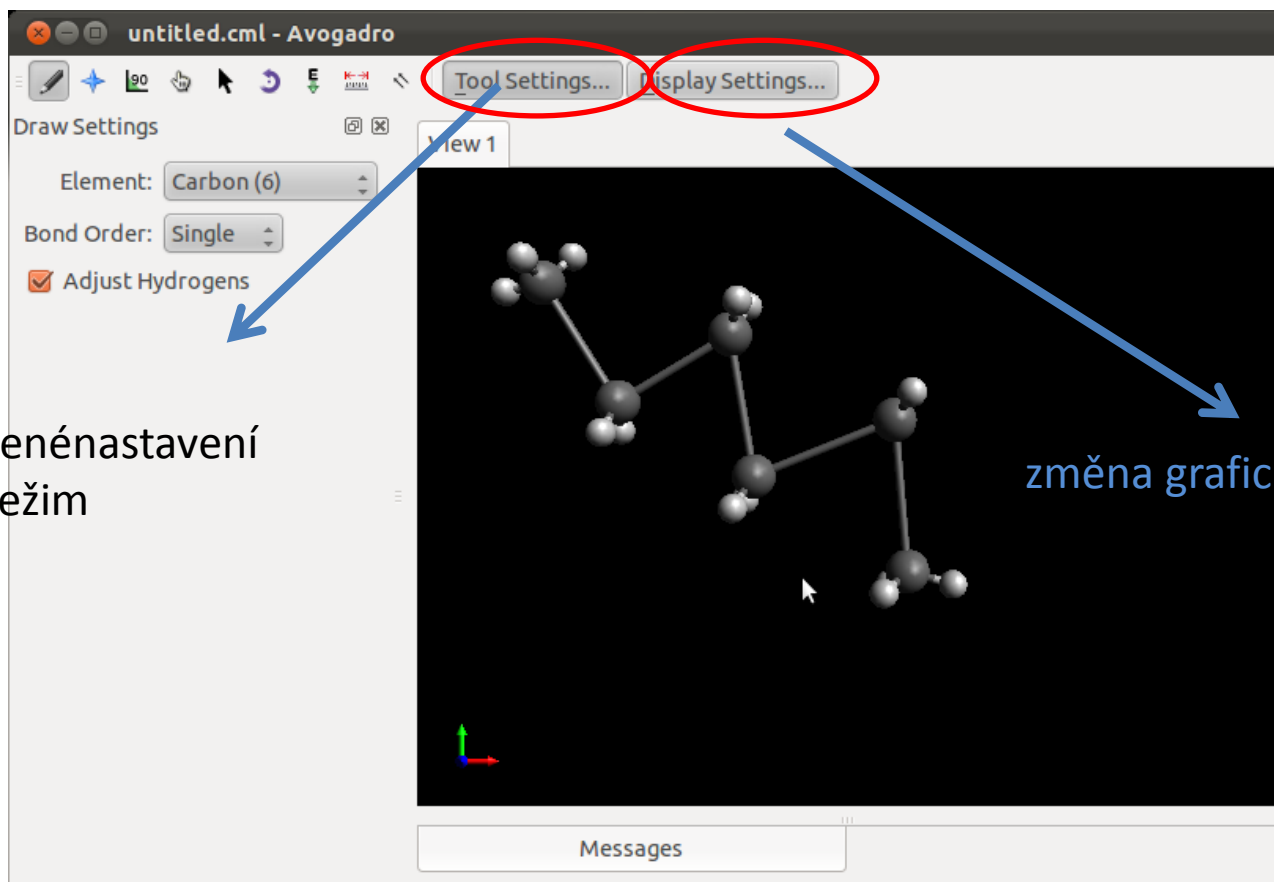
Ke stavbě 3D struktury molekul můžete použít program **Avogadro**. Jedná se o volně šiřitelný program, který lze používat jak pod operačním systémem MS Windows tak i pod Linuxovými klony (např. Ubuntu).



aktivuje editaci

Avogadro

Při stavbě molekuly nejsou délky vazeb, úhly a další parametry molekuly optimální. Je to dáno způsobem, jakým se v programu Avogadro, struktury editují. Draft struktury je proto nutné před dalším použitím upravit pomocí optimalizace geometrie.

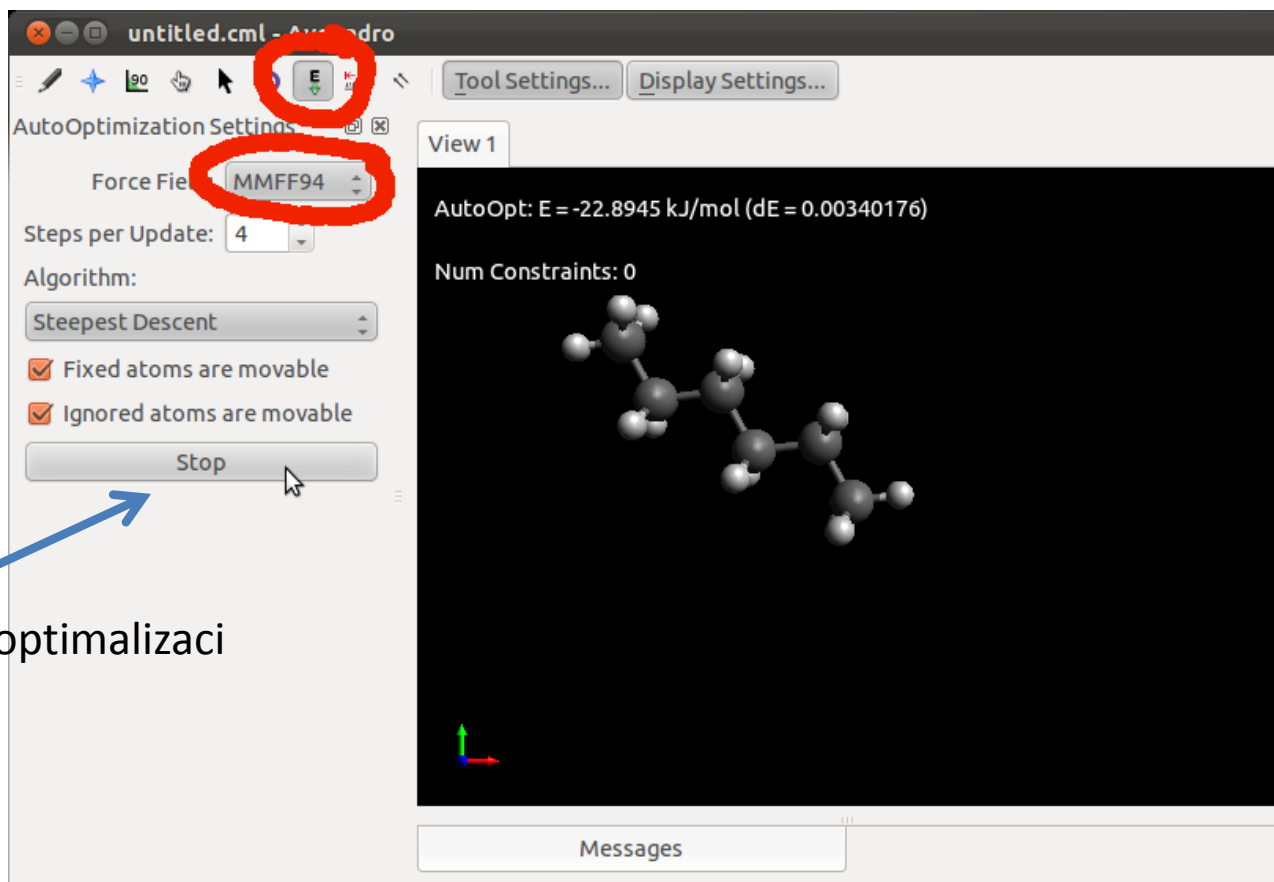


zobrazí rozšířená nastavení pro zvolený režim

změna grafické vizualizace

Avogadro

Program používá pro optimalizace geometrie metody molekulové mechaniky (MM). Pro její správnou funkci musíte ve struktuře správně uvést řady vazeb. Protože MM je empirickou metodou, musíte zvolit i typ parametrizace. V našem případě budeme používat silové pole MMFF94.



zapíná/vypíná optimalizaci geometrie

Nemesis

The screenshot shows the Nemesis Molecular Modelling Package interface. The main window is titled "Project 1: NEMESIS - Molecular Modelling Package". The interface is divided into several panels:

- Structures panel (left):** Contains a table with columns "Name", "SID", and "Ato". It lists "Structure 1" with SID "1". Below the table are controls for "Number of structures" (set to 1) and "Active Profile" (set to "Profile 1"). A "Profile objects" table lists "Light 1", "Background 1", "Standard Model 1", and "Freezed Atoms 1".
- Build panel (right):** Contains chemical symbols for building molecules (C, O, N, F, Cl, Br, I, S) and buttons for "Delete atom", "Make bond", "Break bond", and "Delete bond". The "Optimize" button is circled in red.
- Geometry panel (bottom right):** Contains buttons for "Position", "Distance", "Angle", and "Torsion".

Annotations with blue arrows point to various parts of the interface:

- "vrstvy" points to the Structures panel.
- "stavba/editace molekuly" points to the Build panel.
- "optimalizace geometrie" points to the "Optimize" button in the Build panel.
- "grafické modely" points to the Profile objects table.
- "měření" points to the Geometry panel.

Nastavení silového pole pro optimalizaci: menu Geometry-> Optimizer Setup

Nemesis

Myš:

Levé tlačítko	selekce
Prostřední tlačítko	rotace
Levé tlačítko	posun
Kolečko	zoom

Modifikátory:

Shift	XZ -> Y pohyby
Ctrl	přepíná mezi sekundárním a primárním manipulátorem

Cvičení I

1. Načtete do programu **Avogadro** molekulu ze souboru **water.xyz**
2. Provedte optimalizaci její geometrie. Jaké je optimální délka vazby a vazebný úhel?
3. Zobrazte molekulu v různých grafických reprezentacích.
4. Načtete do programu **Nemesis** molekulu ze souboru **water.xyz**
5. Zobrazte molekulu v různých grafických reprezentacích.
6. Provedte optimalizaci její geometrie. Jaké je optimální délka vazby a vazebný úhel? Srovnajte s výsledky získanými v programu Avogadro. Vysvětlete případné rozdíly.

Cvičení II

1. V programu Nemesis nakreslete strukturní vzorec molekuly benzoové v projektu **Sketch Structure**
2. Převeďte molekulu do 3D reprezentace. Ohodnoťte kvalitu převodu.
3. V programu Nemesis nakreslete strukturní vzorec molekuly cyklohexanu v projektu **"Sketch Structure"**
4. Převeďte molekulu do 3D reprezentace. Ohodnoťte kvalitu převodu.