Molecular Dynamics Simulations
SS 2007

Alessandra Villa
Computer Lab

How to find me:

Alessandra Villa (Room NU140/229)
www.theochem.uni-frankfurt.de/~villa
villa@theochem.uni-frankfurt.de

Computer lab:

- General introduction to start
- Display and exploration of a molecule.
- Molecular Dynamics software
- Model (force field)
- Generation of topology and coordinate files
- Generating configurations
  - Energy minimization
  - Molecular Dynamics
- Trajectory analysis
- Simulation conditions

Something about BCC computer

Useful information

user: *{user_name}*
passwd: *zzzzz*
>pwd
/home/{user_name}/

Share information:
/usr/local/courseexchange/mdcourse

Where to put data:
/tmp/{directoryname} (on each PC)

or

/usr/local/courseexchange/mdcourse/*{user_name}*

Something about Linux

Commands for navigating the Linux file system

pwd
Print Working Directory. Shows the current location in the directory tree.

mkdir {directoryname}
Make the directory with the specified directory name

cd {directoryname}
Change Directory into the specified directory name.

cd ..
Move up one directory.

ls
List all files in the current directory, in column format.

Something about Linux

Working With Files and directory

file {filename}
Find out what kind of file it is.
cat {filename}
Display the contents of a text file on the screen.
head {filename}
Display the first few lines of a text file.
tail {filename}
Display the last few lines of a text file.
cp {filename} {directoryname/filename}
Copies a file from one location to another.
mv {filename} {newfilename}
Moves a file to a new location, or renames it.
rm {filename}
Delete a file.
rmdir {directoryname}
Remove Directory.
emacs {filename}
Text editor
less {filename}
Page through a file

Something about Linux

Informational Commands

top
Displays CPU processes in a full-screen GUI. A great way to see the activity on your computer in real-time. Type `Q` to quit.
df
Report filesystem disk space usage (DiskFree is how I remember it)

Other Utilities: grep, more, history, find, ps

Help: **man {commandname}**

Useful links:
•http://www.linux-tutorial.info/
see links menu:
Table of Contents/Shell and Utilities/Commonly Used Utilities

To source the simulation package:

1) Go to your home directory:
>cd /home/{user_name}

2) Create and Open the file .bashrc
>emacs .bashrc

3) Type in the file .bashrc the following 3 lines:
#!/bin/bash
#
source /usr/local/bcc/gromacs/i686-pc-linux-gnu/bin/GMXRC.bash

4) Save and close the file .bashrc

NOTE: the *file .bashrc should be in the directory /home/***{user_name}**

To check:
Open a new xterminal and type
>grompp -h

Something about the programs

Visualization programs **file's type**
> vmd *.pdb, *.gro,*.xtc
Look at http://www-ks-uiuc.edu/Research/vmd/
or
vmd.html and vmd.pdf in mdcourse directory

> rasmol *.pdb
Look at http://www-umass.edu/mjcrobio/rasmol/rasintro.htm

>molden *.pdb
>pymol *.pdb

Visualization graphs:
>xmgrace *.xvg
Look at http://plasma-gate.weizmann.ac.il/Grace/

Simulation Software:
http://www.gromacs.org

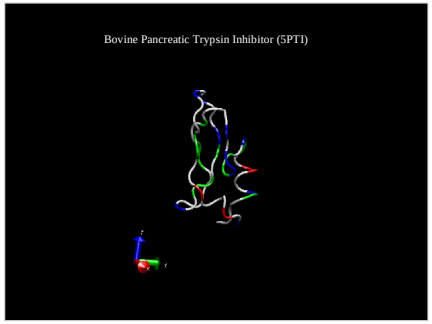
First part:

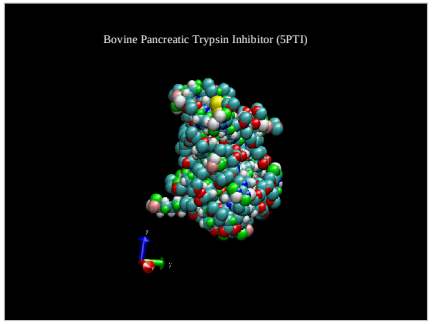
- Linux, software and Beilstein Computer Center
- Molecular structural file *.pdb *.gro
- PDB file format
- How to visualize a molecule

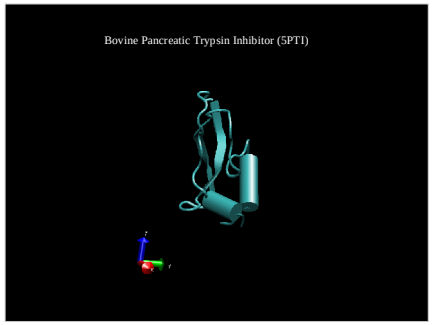
Useful abbreviations

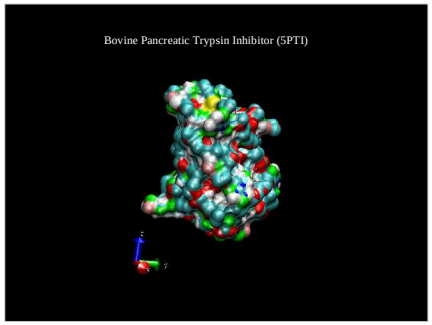
- PDB: Protein Data Bank

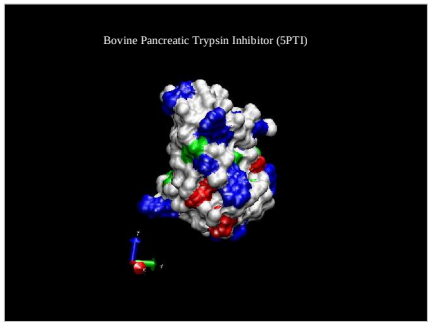
How do we represent a protein?

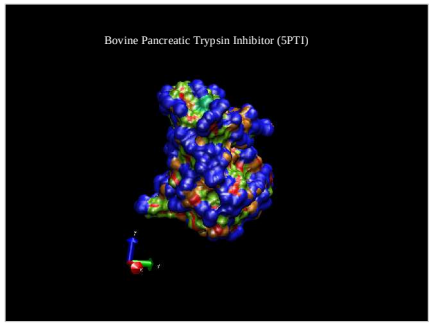












To visualize a molecule

Molecular Structure File

What does it contain?

- the name of each atom.
- the coordinates (x, y, z) of each atom.
- the number of atoms in the system (optional)
- the velocities (x, y, z) of each atom (optional).
- the dimensions of the system (optional).

How is it generated?

- automatically using predefined building parameters.
- derived from experimental data (such as X-ray or NMR study).
- by hand

How to get the experimental structure of a protein:

- Go to 'Protein Data Bank' **http://www.rcsb.org/pdb/**
- Look for a protein
- Search for and look at the trypsin-benzamidine complex from the Protein Data Bank. PDB ID: 1A0J
- Download the pdb file
- Look in the PDB file.

The PDB file

```
HEADER   HYDROLASE                              19-MAY-97   1AKI
TITLE    THE STRUCTURE OF THE ORTHORHOMBIC FORM OF HEN EGG-WHITE
TITLE   2 LYSOZYME AT 1.5 ANGSTROMS RESOLUTION
COMPND   MOL_ID: 1;
COMPND   2 MOLECULE: LYSOZYME;
...
SOURCE   4 CELL: EGG;
KEYWDS   HYDROLASE, GLYCOSIDASE
EXPDTA   X-RAY DIFFRACTION
AUTHOR   D.CARTER,J.HE,J.R.RUBLE,B.WRIGHT
REVDAT   1   19-NOV-97 1AKI    0
...
REMARK   1 REFERENCE 1
REMARK   1  AUTH   P.J.ARTYMIUK,C.C.F.BLAKE,D.W.RICE,K.S.WILSON
REMARK   1  TITL   THE STRUCTURES OF THE MONOCLINIC AND
ORTHORHOMBIC
REMARK   1  TITL 2 FORMS OF HEN EGG-WHITE LYSOZYME AT 6 ANGSTROMS
...
REMARK   1  TITL 3 RESOLUTION
REMARK   1  REF    ACTA CRYSTALLOGR.,SECT.B      V.  38   778 1982
REMARK   2 RESOLUTION. 1.5  ANGSTROMS.
REMARK   3
...
SEQRES   1   129  LYS VAL PHE GLY ARG CYS GLU LEU ALA ALA ALA MET LYS
...
SEQRES   2   129  ARG HIS GLY LEU ASP ASN TYR ARG GLY TYR SER LEU GLY
...
HELIX    1  1 ARG     5  ARG    14  1                          10
...
SHEET    1  A 2 THR    43  ARG    45  0
SHEET    2  A 2 THR    51  TYR    53  -1  N  ASP    52   O  ASN    44
CRYST1   59.062   68.452   30.017  90.00  90.00  90.00 P 21 21 21    4
...
ATOM      1  N   LYS     1      35.365  22.343 -11.883  1.00 22.28           N
ATOM      2  CA  LYS     1      35.892  21.073 -11.427  1.00 21.12           C
ATOM      3  C   LYS     1      34.741  20.264 -10.844  1.00 16.85           C
...
HETATM 1003  O   HOH     1      23.434  40.063  -6.661  1.00 19.48           O
...
CONECT  48  47  981
...
END
```

Molecular Structure File

What does it contain?

- the coordinates (x, y, z) of each atom.
- the velocities (x, y, z) of each atom (optional).
- the dimensions of the system (optional).

```
LYSOZYME
1555
   1LYSH    N   1   3.346   1.966   1.848   0.0261 -0.4172  0.0298
   1LYSH    H1  2   3.253   1.992   1.811   0.4218  1.0831 -0.6524
   1LYSH    H2  3   3.133   1.994   1.944   1.3833  1.5774 -0.3186
   1LYSH    H3  4   3.412   2.029   1.814   0.2508  0.9650  0.5365
...
 207SOL   HW2 1555   3.970   1.939   2.397  0.4891 -0.4783 -0.5275
BOX
   5.67062   6.57220   2.93002
```

Example of *.gro file

How is it generated?

- by hand

or

- automatically from a given structure file (e.g. a Protein Data Bank (PDB) file).

Exercise 1: Display and exploration of a molecule.

- Search for and download X-ray structure of the trypsin complex from the Protein Data Bank. PDB ID: 1A0J.
- Start VMD (Virtual Molecular Dynamics)
- Read trypsin into VMD.

Use VMD to display different aspects of the molecule:

- Display the different aminoacids present in Trypsin
- Display the different chains present in X-ray structure
- Show the secondary structure.
- Show the ligand (resname BEN) of the trypsin.
- What do the residues Aspartate 189 and Glycine 219 and Serine 190 define?.

Useful links:

Protein Data Bank:
http://www.rcsb.org/pdb/
Visual molecular Dynamics
http://www.ks.uiuc.edu/Research/vmd/
or
Look at vmd.html and vmd.pdf in mdcourse directory

Exercise 2: Display and exploration of a molecule.

- Search for and download the NMR structure of ubiquitin from the Protein Data Bank. PDB ID: 1D3Z.
- Start VMD (Virtual Molecular Dynamics)
- Read ubiquitin into VMD.

Use VMD to display different aspects of the molecule:

- How many structures are in the file?
- Display the hydrogen bonds and the secondary structure: what do you notice?

Visualization programs:
vmd (*.pdb, *.gro)
rasmol (*.pdb)
molden (*.pdb)
pymol (*.pdb)

Question:

After having visualized the 3 structure files (1AOJ.pdb, 1D3Z.pdb and example.gro), say (about the type of information):

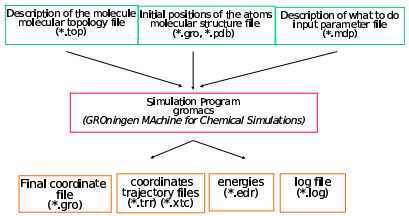
- **In what do the 3 files differ?.**
- **What do they have in common?**

Molecular Dynamics simulation package

GROMACS
http://www.gromacs.org
GROMOS
http://www.igc.ethz.ch/gromos/gromos.html
Amber
http://amber.scripps.edu/
CHARMM
http://www.charmm.org/

etc……

Running a simulation



MD Unit in GROMACS

Quantity	Symbol	Unit
distance	r	nm
mass	m	u (atomic mass unit)
time	t	ps
charge	q	e = electronic charge
temperature	T	K
energy	E, V	kJ/mol
Force	F	kJ/mol*nm
pressure	p	kJ/mol*nm^3 = 16.6054 Bar
velocity	v	nm/ps
dipole moment	μ	e*nm

GROMACS: ensamble of programs

Programs by topic:

Input

- Generating topologies and coordinates: pdb2gmx, genbox

simulation

- Running a simulation: grompp, mdrun

analysis

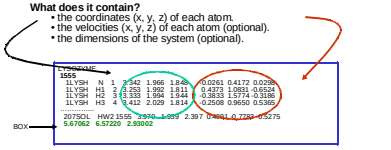
- Processing energies: g_energy
- Converting files: editconf, trjconv
- Tools: make_ndx, g_analyze
- Distances between/in structures g_rms, g_mindist, g_angle
- Structural properties: g_hbond, g_sas
- Kinetic properties: g_tcaf
- Electrostatic properties: genion

Help: **program_name -h**

Molecular Structure File

What does it contain?

- the coordinates (x, y, z) of each atom.
- the velocities (x, y, z) of each atom (optional).
- the dimensions of the system (optional).



Example of *.gro file

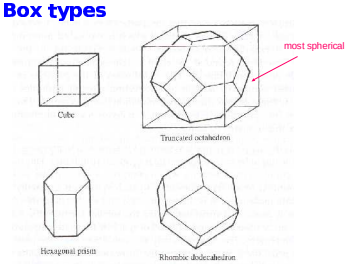
How is it generated?

- by hand

or

- automatically from a given structure file (e.g. a Protein Data Bank (PDB) file).

Box types



Define a box and solvate a molecule.

- Use editconf to define a box around the molecule in the *.gro
- Use genbox to solvate with water the molecule
- Visualize the system protein+water

Useful information

editconf edits the box and writes subgroups
genbox solvates a system

Visualization programs:

vmd (*.pdb, *.gro)
rasmol (*.pdb)
molden (*.pdb)
pymol (*.pdb)

Molecular Topology File

What does it contain?

- list of all atoms and residues including names
- a complete description of all interactions in system
- lists of all bonds, bond angles, dihedral angles etc.
- partial charges of all atoms
- Lennard-Jones parameters for all atoms
- special interactions (exclusions, 1-4 interactions, constraints)

How is it generated?

- by hand

or

- automatically from an amino acid sequence and predefined building blocks

Example of topology file (*.top)

```
#include "ffG43a1.itp"

[ moleculetype ]
; Name            nrexcl
Protein           3

[ atoms ]
;   nr       type  resnr residue  atom   cgnr     charge       mass       typeB    chargeB      massB
     1         NL      1    LYSH     N      1      0.129    14.0067 ; comments
     2          H      1    LYSH    H1      1      0.248      1.008 ;
     3          H      1    LYSH    H2      1      0.248      1.008 ;
     4          H      1    LYSH    H3      1      0.248      1.008 ;
     5        CH1      1    LYSH    CA      1      0.127     14.019 ;


[ bonds ]
;  ai    aj funct            c0            c1            c2            c3
    1     2     2      0.1000 1.8700e+07

[ angles ]
;  ai    aj    ak funct            c0            c1            c2            c3
    2     1     3     2        109.50        380.00

[ dihedrals ]
;  ai    aj    ak    al funct            c0            c1            c2            c3            c4            c5
    2     1     5    14     1         0.000         3.77             3

 ......

#include "spc.itp"

[ molecules ]
; Compound        #mols
Protein             1
SOL               78
```

Input Parameter File

What does it contain?

- Title (description of the run).
- parameters describing what run is to be performed.
  - MD or EM
  - temperature
  - nature of box
  - time step
  - length of run
  - how often to write out the results
  - etc.

How is it generated?

- by hand from a template file