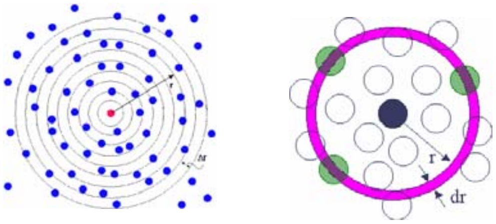
BINF 731

Protein Structure Analysis

Iosif Vaisman

2004

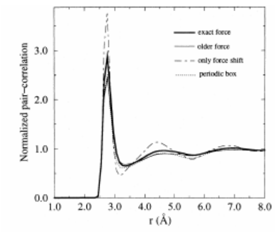
MD of proteins: radial distribution functions



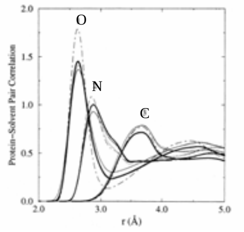
$$N_{\mathrm{ideal}}(r) = \frac{N}{V} \times V_{\mathrm{shell}}(r) = \rho \times 4\pi r^2\,dr$$

$$\mathbf{g(r)} = \frac{\mathbf{N(r)}}{\mathbf{N_i(r)}} = \frac{\mathbf{N(r)}}{(\rho \times 4\pi\, r^2 dr)}$$

MD of proteins: radial distribution functions



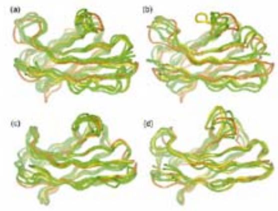
Water O-O rdf



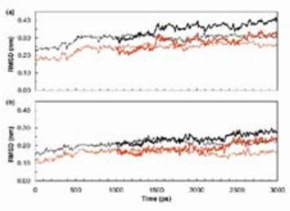
Protein-water rdf

Adopted from V.Daggett (1999)

MD of proteins: mobile regions



Snapshots of V_H domain simulation at 300 and 340 K



Adopted from W.F.VanGunsteren (2001)

---

freely available at www.gromacs.org

- Generally 3 to 10 times faster than other Molecular Dynamics programs
- Very user-friendly: issues clear error messages, no scripting language is required to run the programs, prints out the progress of the program that is running, etc.
- Allows the trajectory data to be stored in a compact way using lossy compression.
- Gromacs provides a basic trajectory data viewer; xmgr or Grace may also be used to analyze the results.
- Files from earlier versions of Gromacs may be used in the latest Gromacs, version 3.1.

File Formats

- ***.pdb**: format used by Brookhaven Protein DataBank
- *.**top**: topology file (ascii), contains all the forcefield parameters
- *.**gro**: molecular structure file in the Gromos87 format (Gromacs format)
  Information in the columns, from left to right:
  - residue number
  - residue name
  - atom name
  - atom number
  - x, y, and z position, in nm
  - x, y, and z velocity, in nm/ps
- *.**tpr**: contains the starting structure of the simulation, the molecular topology file and all the simulation parameters; binary format

File Formats

- ***.trr**: contains the trajectory data for the simulation; binary format.
  It contains all the coordinates, velocities, forces and energies as was
  indicated the mdp file.

- ***.edr**: portable file that contains the energies.

- ***.xvg**: file format that is read by Grace (formerly called Xmgr), which
  is a plotting tool for the X window system.

- ***.xtc**: portable format for trajectories which stores the information about
  the trajectories of the atoms in a compact manner (it only contains
  cartesian coordinates).

File Formats

- ***.mdp**: allows the user to set up specific parameters for all the calculations
  that Gromacs performs.

- **em.mdp file:** sets the parameters for running energy minimizations;
  allows you to specify the integrator (steepest descent or
  conjugate gradients), the number of iterations, frequency to update
  the neighbor list, constraints, etc.

- **md.mdp file:** sets the parameters for running the molecular dynamics program;
  allows you to indicate the appropriate settings depending on the
  force field used,

Generic mdp file for energy minimization

```
title                = Yo
cpp                    = /lib/cpp
include                 = -I../top
define                  =
integrator               = md
dt                   = 0.002
nsteps                  = 500000
nstxout                 = 5000
nstvout                 = 5000
nstlog                 = 5000
nstenergy                 = 250
nstxtcout                = 250
xtc_grps                 = Protein
energygrps                  = Protein  SOL
nstlist               = 10
ns_type                  = grid
rlist                = 0.8
coulombtype                  = cut-off
rcoulomb                  = 1.4
rvdw                   = 0.8
tcoupl                 = Berendsen
tc-grps                 = Protein      SOL
tau_t                 = 0.1  0.1
ref_t                = 300  300
Pcoupl                  = Berendsen
tau_p                   = 1.0
compressibility            = 4.5e-5
ref_p                  = 1.0
gen_vel                  = yes
gen_temp                   = 300
gen_seed                   = 173529
```

Force Field

- The set of equations (*potential functions*) used to generate the
  potential energies and their derivatives, the forces.

- The parameters used in this set of equations

- Gromacs provides the following force fields:

0: Gromacs Forcefield (see manual)
1: Gromacs Forcefield with all hydrogens (proteins only)
2: GROMOS96 43a1 Forcefield (official distribution)
3: GROMOS96 43b1 Vacuum Forcefield (official distribution)
4: GROMOS96 43a2 Forcefield (development) (improved alkane dihedrals)

Programs

- **pdb2gmx:**
  - reads in a pdb file and allows the user to chose a forcefield

  - reads some database files to make special bonds (i.e. Cys-Cys)

  - adds hydrogens to the protein

  - generates a coordinate file in Gromacs (Gromos) format (*.gro) and
    a topology file in Gromacs format (*.top).

  - issues a warning message if an atom is not well resolved in the
    structure.

Programs

- **editconf**:
  - converts gromacs files (*.gro) back to pdb files (*.pdb)

  - allows user to setup the box: the user can
    define the type of box (i.e. cubic, triclinic, octahedron)

    set the dimensions of the box edges relative to the molecule
    (-d 0.7 will set the box edges 0.7 nm from the molecule)

    center the molecule in the box

Programs

- **genbox**:
  - solvates the box based on the dimensions specified using editconf
  - solvates the given protein in the specified solvent (by default SPC- Simple Point Charge water)
  - water molecules are removed from the box if the distance between any atom of the solute and the solvent is less than the sum of the VanderWaals radii of both atoms (the radii are read from the database vdwradii.dat)

Programs

- **grompp** (pre-processor program):
  - reads a molecular topology file (*.top) and checks the validity of the file
  - expands the topology from a molecular description to an atomic description (*.tpr)
  - it reads the parameter file (*.mdp), the coordinate file (*.gro) and the topology file (*.top)
  - it ouputs a *.tpr file for input into the MD program **mdrun**
  - since *.tpr is a binary file, it can not be read with ‘more’ but it may be read using **gmxdump**, which prints out the input file in readable format (it also prints out the contents of a *.trr file)

Programs

- **mdrun**:
  - performs the Molecular Dynamics simulation
  - can also perform Brownian Dynamics, Langevin Dynamics, and Conjugate Gradient or Steepest Descents energy minimization
  - reads the *.tpr file, creates neighborlists from that file and calculates the forces.
  - globally sums up the forces and updates the positions and velocities.
  - outputs at least three types of files:
    (1) trajectory file (*.trr): contains coordinates, velocities, and forces
    (2) structure file (*.gro):contains coordinates and velocities of the last step
    (3) energy file (*.edr): contains energies, temperatures, pressures

Programs

- **gmxcheck**: gmxcheck reads a trajectory (*.trr) or an energy file (*.edr) and prints out useful information in them.
- **g_energy**: extracts energy components or distance restraint data from an energy file into a *.xvg file (may be read using Xmgr or Grace).
- **trjconv**: allows compression of trajectory file into a *.xtc file that can be analyzed using **ngmx**
- **ngmx**:
  - Gromacs trajectory viewer
  - plots a 3-D structure of the molecule
  - allows ratation, scaling, translation, labels on atoms, animation of trajectories, etc.